UNIVERSITY OF CALIFORNIA
Los Angeles

Dynamical Methods for the Sarnak and Chowla Conjectures

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Mathematics

by

Redmond McNamara

2021

© Copyright by
Redmond McNamara
2021

ABSTRACT OF THE DISSERTATION

Dynamical Methods for the Sarnak and Chowla Conjectures

by

Redmond McNamara
Doctor of Philosophy in Mathematics
University of California, Los Angeles, 2021
Professor Terence Tao, Chair

This thesis concerns the Liouville function, the prime number theorem, the Erdős discrepancy problem and related topics. We prove the logarithmic Sarnak conjecture for sequences of subquadratic word growth. In particular, we show that the Liouville function has at least quadratically many sign patterns. We deduce this theorem from a variant which bounds the correlations between multiplicative functions and sequences with subquadratically many words which occur with positive logarithmic density. This allows us to actually prove that our multiplicative functions do not locally correlate with sequences of subquadratic word growth. We also prove a conditional result which shows that if the $\kappa - 1$-Fourier uniformity conjecture holds then the Liouville function does not correlate with sequences with $O(n^{t-\varepsilon})$ many words of length n where $t = \kappa(\kappa+1)/2$. We prove a variant of the 1-Fourier uniformity conjecture where the frequencies are restricted to any set of box dimension < 1. We give a new proof of the prime number theorem. We show how this proof can be interpreted in a dynamical setting. Along the way we give a new and improved version of the entropy decrement argument. We give a quantitative version of the Erdős discrepancy problem. In particular, we show that for any N and any sequence f of plus and minus ones, for some

$n \leq N$ and $d \leq \exp(N)$ that $|\sum_{i \leq n} f(id)| \geq (\log\log N)^{\frac{1}{484}-o(1)}$.

The dissertation of Redmond McNamara is approved.

Rowan Killip

Bill Duke

Tim Austin

Terence Tao, Committee Chair

University of California, Los Angeles

2021

TABLE OF CONTENTS

1 Introduction **1**

2 Sarnak's Conjecture for Sequences of Almost Quadratic Word Growth . **9**

2.1 Introduction to Chapter 1 9

2.1.1 Background and notation 21

2.1.2 Acknowledgments 28

2.2 Main Argument 28

2.2.1 The Entropy Decrement Argument 46

2.2.2 Nilsystems and Algebraic Structure 54

2.3 Proof of Theorem 2.1.11 74

2.4 Frantzikinakis-Host and dynamical models 75

2.5 Reduction to the completely multiplicative case 79

3 A Dynamical Proof of the Prime Number Theorem **84**

3.1 Introduction to Chapter 2 84

3.1.1 A comment on notation 89

3.1.2 Acknowledgments 90

3.2 Proof of the prime number theorem 90

3.3 In what ways is this a dynamical proof? 109

4 A Quantitative Erdős Discrepancy Theorem **118**

4.1 Introduction to Chapter 3 118

4.2 Multiplicative Fourier Reduction 125

4.3 The Nonpretentious Case 130

4.4 The Pretentious Case 135

References **151**

ACKNOWLEDGMENTS

I would like to thank Terence Tao for his help and guidance. I would also like to thank Tim Austin, Carlos Kenig and Amie Wilkinson for their generous assistance. Finally, I would like to thank my family, especially my parents Bruce McNamara and Daria Walsh for their love and support.

VITA

2016	B.A. (Mathematics), University of Chicago
2018	M.A. (Mathematics), University of California, Los Angeles
2020	Beckenbach Fellowship, Department of Mathematics, University of California, Los Angeles

PUBLICATIONS

McNamara, Redmond, *Sarnak's Conjecture for Sequences of Almost Quadratic Word Growth*, Ergodic Theory and Dynamical Systems (2020).

McNamara, Redmond, *A Dynamical Proof of the Prime Number Theorem*, Hardy Ramanujan Journal (to appear).

CHAPTER 1

Introduction

Arguably the oldest known algorithm for finding primes is the sieve of Eratosthenes, first recorded by Nicomachus in the second century A.D. but attributed to the third century B.C. mathematician. Up to cosmetic modifications, the algorithm works as follows. Let x be a natural number. Then write a list of all the numbers between x and $2x$. Cross off all the multiples of 2, then all the multiples of 3, then all the multiples of 5 and so on. What remains after everything else has been crossed off (say up to multiples of $\sqrt{2x}$) are the primes. This naturally gives a way to count primes. The number of primes between x and $2x$ is x minus the number of numbers divisible by 2, minus the number of numbers divisible by 3 minus the number of numbers divisible by 5 and so on. Except now we have twice subtracted off the number of numbers divisible by 6 because we subtracted them off once when we subtracted off the multiples of 2 and once when we subtracted off the multiples of 3. Similarly with 10, 15 et cetera. So we add on all the multiples of 6, all the multiples of 10, all the multiples of 15 and so on. But now again we have over counted the multiples of 30. Altogether, we get the formula

$$\begin{aligned} \#\text{ of primes between } x \text{ and } 2x \; = x &- \Big(\ \#\text{ of \#s divisible by } 2\ \Big) - \Big(\ \#\text{ of \#s divisible by } 3\ \Big) \\ &- \Big(\ \#\text{ of \#s divisible by } 5\ \Big) - \cdots \\ &+ \Big(\ \#\text{ of \#s divisible by } 6\ \Big) + \Big(\ \#\text{ of \#s divisible by } 10\ \Big) \\ &+ \Big(\ \#\text{ of \#s divisible by } 15\ \Big) + \cdots \\ &- \Big(\ \#\text{ of \#s divisible by } 30\ \Big) - \cdots \end{aligned}$$

or to rewrite this using variables

$$\#\text{ of primes between } x \text{ and } 2x \;=\; \sum_{\substack{d\le\sqrt{2x}\\ d \text{ squarefree}}} \pm\Big(\;\#\text{ of \#s divisible by } d\;\Big).$$

How do you know whether you are supposed to add or subtract the multiples of d? The coefficient of the (# of #s divisible by d) is precisely $(-1)^{\#\text{ of prime factors of } d}$. This is the Liouville function

$$\lambda(d) = (-1)^{\#\text{ of prime factors of } d},$$

and its central role in the sieve of Eratosthenes is the first of its many important roles in number theory. Another hint that the Liouville function might be an important function in the study of the primes is the identity

$$\frac{\zeta(2s)}{\zeta(s)} = \sum_{n\in\mathbb{N}} \frac{\lambda(n)}{n^s}$$

where ζ is the famous Riemann zeta function defined by

$$\begin{aligned}\zeta(s) &= \sum_{n\in\mathbb{N}} \frac{1}{n^s} \\ &= \prod_{p} \left(1 - \frac{1}{p^s}\right)^{-1}.\end{aligned}$$

Therefore the famous zeros of the zeta function correspond to poles of the Dirichlet series for λ. Yet another hint that the Liouville function might be central to analytic number theory is given by the convolution formula

$$\Lambda = \lambda * \log * \phi$$

where $\phi(d^2) = \lambda(d)$ if d is squarefree and 0 otherwise i.e. $\phi(d^2) = \mu(d)$ and Λ is the von Mangoldt function, which acts sort of like the normalized indicator function of the primes.

All of this perhaps makes the following fact less surprising: the prime number theorem, which states that the number of primes less than x is $\frac{x}{\log x}(1 + \text{error})$ where the error term goes to 0 as x tends to infinity, is equivalent to the fact that

$$\lim_{n\to\infty} \mathbb{E}_{n\le N}\lambda(n) = 0,$$

where by definition

$$\mathbb{E}_{n\leq N} f(n) = \frac{1}{N}\sum_{n\leq N} f(n).$$

There are many interesting ways of thinking about this theorem, which in turn lead to their own interesting generalizations.

First, we can think of

$$\lim_{n\to\infty} \mathbb{E}_{n\leq N}\lambda(n) = 0$$

as telling us that $\lambda(n) = +1$ roughly 50% of the time and $\lambda(n) = -1$ roughly 50% of the time. One might naturally then ask what about $(\lambda(n), \lambda(n+1))$? Is that $(+1,+1)$, $(+1,-1)$, $(-1,+1)$ and $(-1,-1)$ with equal probability as well? What about $(\lambda(n), \lambda(n+1), \lambda(n+2))$? It turns out that the answer to this question is still unknown and would represent major progress in analytic number theory. In fact, this is a famous conjecture. To state it precisely, define a word of length k of a sequence f as a string of k consecutive values of f i.e. ϵ is a word of f if there exists n such that

$$f(n+i) = \epsilon_i$$

for all $i \leq k$. Then Chowla's conjecture states

Conjecture (Chowla)**.** *All 2^k possible words of length k of the Liouville function occur with equal probability.*

This is open. In fact, we do not even know whether all 2^k possible words of length k occur at all. Previously [Hil86a] showed that all 8 words of length 3 occur infinitely often. [MRT16] showed all 8 words of length 3 occur with positive density. [Tao16b] proved that all 4 words of length 2 occur with equal probability if one uses logarithmic weights rather than the normal uniform weights. (We will define logarithmic weights later on). [TT17b] proved that all 16 sign patterns of length 4 occur with positive density using an argument communicated to them by Matomäki and Sawin. [TT17b] showed that all 8 words of length 3 occur with equal probability again using logarithmic weights. [TT17b] also showed the

number of words of length k is at least $2k+8$ for $k \geq 4$. [FH18b] showed that the number of words is super linear. In this thesis (see Chapter 2), based on work in [McN18], we show

Theorem. *There is a constant c such that at least ck^2 many words of length k occur with positive upper logarithmic density.*

Since [McN18] was published, it was proved in [MRT$^+$] that actually super polynomially many words occur, but not necessarily with positive density.

Another way to generalize

$$\lim_{n\to\infty} \mathbb{E}_{n\leq N}\lambda(n) = 0$$

is to ask about multiple correlations. Is it true that

$$\lim_{n\to\infty} \mathbb{E}_{n\leq N}\lambda(n)\lambda(n+1) = 0?$$

What about the more general question

$$\lim_{n\to\infty} \mathbb{E}_{n\leq N}\lambda(n+h_1)\lambda(n+h_2)\cdots\lambda(h+h_k) = 0 \tag{1.1}$$

where none of the h_i's are equal? Clearly this would follow from the Chowla conjecture as stated earlier. But actually, it is equivalent, since

$$\text{Probability } \lambda(n+i) = \epsilon_i \text{ for all } i \leq k \ = \pm\frac{1}{2^k}\mathbb{E}(\lambda(n+1)-\epsilon_1)\cdot\cdots\cdot(\lambda(n+k)-\epsilon_k)$$

for any possible word ϵ of length k. (Expanding out the product and applying (1.1) to each term completes the proof). For all $k > 1$, this is again open. However, Terence Tao proved an exciting averaged version of 1.1 in the case $k = 2$. In particular, define the logarithmic average of a function f over a set A by the formula,

$$\mathbb{E}^{\log}_{a\in A} f(n) = \left(\sum_{a\in A}\frac{1}{a}\right)^{-1}\sum_{a\in A}\frac{f(a)}{a}.$$

Then Tao proved

Theorem ([Tao16b]). *For all $h \neq 0$*

$$\lim \mathbb{E}_{n \leq N} \lambda(n)\lambda(n+h) = 0$$

Inspired by this proof and the proof of [TT17b], I managed to give a new proof of the prime number theorem which will be given in Chapter 3. Tao then used [Tao16b] to resolve a longstanding and important problem in combinatorics called the Erős discrepancy problem.

Theorem ([Tao16a]). *For any function f from $\mathbb{N}$ to $\{\pm 1\}$,*

$$\sup_{d,n} \left| \sum_{i \leq n} f(id) \right| = +\infty.$$

One could try to generalize these results in further by asking for a quantitative version: how quickly does $\mathbb{E}_{n \leq N} \lambda(n)$ go to 0? how quickly does $\mathbb{E}^{\log}_{n \leq N} \lambda(n)\lambda(n+h)$ go to 0? how quickly does $\sup_{d,n} \left| \sum_{i \leq n} f(id) \right|$ go to infinity? The first and second questions turn out to be extremely important. For instance, the famous Riemann hypothesis is equivalent to the fact that

$$|\mathbb{E}_{n \leq N} \lambda(n)| \leq C_\varepsilon N^{-\frac{1}{2}+\varepsilon}.$$

The twin primes conjecture would probably follow from an estimate of the form

$$|\mathbb{E}_{n \leq N} \lambda(an+b)\lambda(cn+d)| \leq C_A \log^{-A} N$$

for all a, b, c and $d \leq N$. (See [SS] for an exciting result showing that the twin primes conjecture is true using this strategy over function fields). [HR21] also made significant progress on this second question recently, showing essentially that

$$|\mathbb{E}^{\log}_{n \leq N} \lambda(n)\lambda(n+h)| \leq C|h| \cdot (\log \log N)^{-\frac{1}{2}}.$$

Using their result, I was able to give something of an answer to the third question.

Theorem. *Let f be a function from $\mathbb{N}$ to $\{\pm 1\}$. Then*

$$\sup_{n \leq N, d \leq e^N} \left| \sum_{i \leq n} f(id) \right| \geq (\log \log N)^{\frac{1}{484} - o(1)}.$$

The proof of this theorem is contained in Chapter 4.

Another way to think about

$$\lim_{n\to\infty} \mathbb{E}_{n\leq N}\lambda(n) = 0$$

is that it is saying λ does not correlate with the constant function 1. (We will say two sequences a and b correlate if $\mathbb{E}a(n)b(n)$ does not tend to 0). Then one could ask what other sequences does λ not correlate with? what does that tell us about the prime numbers? These turn out to be very fruitful questions. For example, the fact that λ does not correlate with periodic functions i.e. for any periodic function f

$$\lim_{N\to\infty} \mathbb{E}_{n\leq N} f(n)\lambda(n) = 0$$

is equivalent to Dirichlet's theorem and the prime number theorem in arithmetic progressions which says that 25% of primes have 1 as their last digit, 25% of primes end in a 3, 25% of primes end in a 7 and 25% of primes end in a 9 and the same thing happens in every other base. Using the discrete Fourier transform, we see that this is equivalent to saying that for any rational angle α

$$\lim_{N\to\infty} \mathbb{E}_{n\leq N} e^{n\alpha}\lambda(n) = 0.$$

One could try to generalize this further: does the previous estimate hold when α is any (possible irrational) real number? The answer is yes. This is a theorem of Davenport which can be implicitly found already as the key ingredient in Vinogradov's proof that any large odd number can be written as the sum of three primes. Green and Tao proved a version where the function $n \mapsto n\alpha$ is replaced by an arbitrary polynomial in n. This was an ingredient in their proof of the Hardy-Littlewood conjecture in the case where one has at least two degrees of freedom. For instance, this lets them show that the number of arithmetic progressions in the primes of length d with starting point and jump size $\leq N$ is

$$\frac{N^2}{\log^d N}(C_d + o(1)).$$

for some positive constant C_d independent of N. In particular, this implies the Green-Tao theorem. This motivated Sarnak to propose the following conjecture:

Conjecture. *Let b a sequence with* zero entropy*. Then b does not correlate with λ i.e.*

$$\lim_{n\leq N} \mathbb{E}_{n\leq N} b(n)\lambda(n) = 0.$$

The condition that b has zero entropy is a bit technical. Perhaps the most illuminating thing one can say in a few lines is that it is implied by the following condition: if the number of words of length k in b is less than $(1+\varepsilon)^k$ for all $\varepsilon > 0$ and all k sufficiently large depending on ε then b has zero entropy. The zero entropy condition is actually a quantitative version of the previous condition: for instance if a word shows up with probability zero, then it does not count toward the entropy. However, it will do no harm if the unacquainted reader chooses to think of the zero entropy condition as being essentially equivalent to subexponential word growth. Tao also introduced a logarithmically averaged version of Sarnak's conjecture.

Conjecture. *Let b a sequence with zero entropy. Then b does not correlate with λ i.e.*

$$\lim_{n\leq N} \mathbb{E}^{\log}_{n\leq N} b(n)\lambda(n) = 0.$$

The reader may chose to think of the word growth rate as a measure of the complexity of the sequence b. Sarnak's conjecture claims that for any sequence with subexponential complexity that sequence does not correlate with λ. So what is the best complexity rate we know how to prove Sarnak's conjecture for? [FH18b] showed the logarithimically averaged Sarnak conjecture holds for any sequence of linear complexity i.e. any sequence where the number of words of length k is bounded by $C \cdot k$ for some constant k. Later in Chapter 2, we will show that

Theorem. *Let b a sequence with subquadratic complexity i.e. the number of words of length k is smaller than εk^2 for any $\varepsilon > 0$ for some k sufficiently large depending on ε. Then b does not correlate with λ i.e.*

$$\lim_{n\leq N} \mathbb{E}^{\log}_{n\leq N} b(n)\lambda(n) = 0.$$

In [Tao17a], Tao showed that the logarithmically averaged Sarnak and Chowla conjectures are equivalent. He did this by introducing a third conjecture which both the logarithmically averaged Sarnak and Chowla conjectures are equivalent to.

Conjecture. *Let G be a k-step nilpotent Lie group, Γ a cocompact subgroup, F a continuous function on G/Γ and x a point in G/Γ. Then*

$$\lim_{H\to\infty}\limsup_{N\to\infty}\mathbb{E}^{\log}_{n\leq N}\sup_{g\in G}|\mathbb{E}_{h\leq H}F(g^h)\lambda(n+h)|=0.$$

The definition of a k-step nilpotent Lie group will be given later but it is not too important for the moment. For the moment, it suffices to know that when $k=1$, this reduces to the conjecture

Conjecture.

$$\lim_{H\to\infty}\limsup_{N\to\infty}\mathbb{E}^{\log}_{n\leq N}\sup_{\alpha\in[0,1]}|\mathbb{E}_{h\leq H}e(\alpha h)\lambda(n+h)|=0.$$

This is still open but some exciting recent progress has been made in [MRT20] and [MRT$^+$] handling the case when H is a small power of N. Another way to make progress on this conjecture is by proving the conjecture when we restrict our frequencies to lie in some subset C of the interval i.e.

$$\lim_{H\to\infty}\limsup_{N\to\infty}\mathbb{E}^{\log}_{n\leq N}\sup_{\alpha\in C}|\mathbb{E}_{h\leq H}e(\alpha h)\lambda(n+h)|=0. \tag{1.2}$$

In Chapter 2, we prove this for every subset C of the interval of box dimension <1. (If we knew this for all sets of box dimension 1 that would imply the conjecture). This theorem is only theorem I know of the form (1.2) for any infinite set C.

We remark that Chapters 2 and 3 are based upon [McN18] and [McN20] respectively with very few changes. As such, these chapters more or less read independently. Although the introductory material may be slightly repetitive, we hope that the reader still finds it enjoyable. In contrast, Chapter 4 is original work first appearing in this thesis.

CHAPTER 2

Sarnak's Conjecture for Sequences of Almost Quadratic Word Growth

2.1 Introduction to Chapter 1

The prime number theorem states that

$$\lim_{N\to\infty} \mathbb{E}_{n\leq N}\Lambda(n) = 1,$$

where $\Lambda(n) = \log p$ if n is a power of a prime p and 0 otherwise is the von Mangoldt function. (We refer the reader to Section 2.1.1 for an explanation of the $\mathbb{E}$ notation). This is equivalent to the estimate

$$\lim_{N\to\infty} \mathbb{E}_{n\leq N}\lambda(n) = 0,$$

where $\lambda(n) = (-1)^{\#\text{ of prime factors of n}}$ is the Liouville function. Dirichlet's theorem on prime numbers in arithmetic progressions morally follows from the estimate

$$\lim_{N\to\infty} \mathbb{E}_{n\leq N}\mathbb{1}_{n\equiv r \bmod d}\lambda(n) = 0,$$

for any d and r. Taking linear combinations, we find that for any periodic function f,

$$\lim_{N\to\infty} \mathbb{E}_{n\leq N} f(n)\lambda(n) = 0.$$

Equivalently, for any function $F\colon S^1 \to \mathbb{C}$ and any rational angle α,

$$\lim_{N\to\infty} \mathbb{E}_{n\leq N} F(\alpha^n)\lambda(n) = 0.$$

The analogous estimate when α is irrational and F is a continuous function was proved by Vinogradov and was a key ingredient in his proof that any sufficiently large odd number is the sum of three primes. Green and Tao proved that

$$\lim_{N\to\infty} \mathbb{E}_{n\leq N} F(g^n\Gamma)\lambda(n) = 0.$$

where G is a nilpotent Lie group, g is an element of G, Γ is a cocompact lattice and F is a continuous function $F\colon G/\Gamma \to \mathbb{C}$. A version of this statement was a key ingredient in their proof with Tamar Ziegler that counts the solutions to almost any system of linear equations over the primes. This motivates the following conjecture, due to Sarnak:

Conjecture 2.1.1 (Sarnak, see [Sar12])**.** *For any topological dynamical system (X,T) with zero entropy, any continuous function $F\colon X \to \mathbb{C}$ and any point x in X,*

$$\lim_{N\to\infty} \mathbb{E}_{n\leq N} F(T^n x)\lambda(n) = 0.$$

Tao introduced the following variant,

Conjecture 2.1.2 (Logarithmically Averaged Sarnak Conjecture)**.** *For any topological dynamical system (X,T) with zero entropy, any continuous function $F\colon X \to \mathbb{C}$ and any point x in X,*

$$\lim_{N\to\infty} \mathbb{E}^{\log}_{n\leq N} F(T^n x)\lambda(n) = 0.$$

Many instances of Sarnak's conjecture have been proven. We give a few examples but stress that this is an incomplete list: [Bou13a], [BSZ13], [Bou13b], [DK15], [EAKL16], [EAKPLdlR17], [FJ18], [HLSY17], [LS15], [MMR14], [Mül17], [Pec18], [Vee17], [Wan17].

Definition 2.1.3. *A word ϵ of length k is an element of $\mathbb{C}^k$. Let k be a natural number, let $\epsilon \in \mathbb{C}^k$ and let $b\colon \mathbb{N} \to \mathbb{C}$. We say that ϵ occurs as a word of b if there exists a natural number n such that $b(n+h) = \epsilon_h$ for all $h \leq k$. We say that ϵ occurs with (upper) logarithmic density $\delta \in \mathbb{R}$ if*

$$\limsup_{N\to\infty} \mathbb{E}^{\log}_{n\leq N} \mathbb{1}_{\epsilon_h = b(n+h) \text{ for all } h\leq k} = \delta.$$

In this chapter, when we refer to log*-density we mean upper logarithmic density. A word* ϵ *whose entries are all* ± 1 *is called a sign pattern. We say that* b *has subquadratic word growth if* b *takes finitely many possible values and the number of words of length* k *that occur with positive upper logarithmic density is* $o(k^2)$*.*

Then a particular case of Sarnak's conjecture predicts that for any bounded sequence $b\colon \mathbb{N} \to \mathbb{C}$ with subexponential word growth that

$$\lim_{N\to\infty} \mathbb{E}^{\log}_{n\leq N} b(n)\lambda(n) = 0.$$

Because λ correlates with itself, this in particular implies that the number of sign patterns of λ of length k is exponential in k. [FH18b] proved the special case where b has linear word growth. In this chapter, we prove the following special case:

Theorem 2.1.4. *Let* b *be a bounded sequence with subquadratic word growth. Then*

$$\lim_{N\to\infty} \mathbb{E}^{\log}_{n\leq N} b(n)\lambda(n) = 0.$$

Previously [Hil86a] showed that all 8 sign patterns of length 3 occur infinitely often. [MRT16] showed all 8 sign patterns of length 3 occur with positive density. [TT17b] proved that all 16 sign patterns of length 4 occur with positive density using an argument communicated to them by Matomäki and Sawin. [TT17b] also showed the number of sign patterns of length k is at least $2k + 8$ for $k \geq 4$. [FH18b] showed that the number of sign patterns is super linear. In particular, Theorem 2.1.4 implies that λ does not have subquadratically many sign patterns. We actually prove something slightly stronger.

Theorem 2.1.5. *There is a constant* $\delta > 0$ *such that* λ *has at least* δk^2 *many sign patterns of length* k*.*

[Tao17a] showed that the log Sarnak conjecture is equivalent to the following Fourier uniformity conjecture for every natural number t.

Conjecture 2.1.6 (t-Fourier uniformity)**.** *Let G be a nilpotent Lie group of step t, Γ a cocompact lattice and $F\colon G/\Gamma \to \mathbb{C}$ a continuous function. Then*

$$\lim_{H\to\infty}\lim_{N\to\infty}\mathbb{E}^{\log}_{n\leq N}\sup_{g\in G}|\mathbb{E}_{h\leq H}\lambda(n+h)F(g^h\Gamma)| = 0.$$

[Tao17a] also showed that this is equivalent to the log-Chowla conjecture for every t.

Conjecture 2.1.7 (Logarithmic Chowla Conjecture)**.** *For every natural number t and every distinct natural numbers $h_1,\dots,h_t$, we have*

$$\lim_{N\to\infty}\mathbb{E}^{\log}_{n\leq N}\lambda(n+h_1)\cdots\lambda(n+h_t) = 0.$$

A function $a\colon \mathbb{N}\to\mathbb{C}$ is said to be unpretentious, nonpretentious or strongly aperiodic if there exists a function ϕ from $\mathbb{N}$ to $\mathbb{N}$ such that, for all natural numbers A, for all Dirichlet characters χ of period at most A we have, for all natural numbers N sufficiently large and for all real numbers $|t|\leq AN$ we have

$$\sum_{p\leq N}\frac{1-\mathrm{Re}(a(p)\chi(p)p^{-it})}{p}\geq\phi(A),$$

and $\phi(A)\to\infty$ as $A\to\infty$. The main goal of this chapter is to prove the following theorems.

Theorem 2.1.8. *Let $a\colon\mathbb{N}\to S^1$ be an unpretentious completely multiplicative function taking values in the unit circle. Let $b\colon\mathbb{N}\to\mathbb{C}$ be a finite-valued 1-bounded function. Suppose further that for any $\delta>0$ there are infinitely many k such that the number of words of b of length k that occur with positive upper logarithmic density is at most δk^2. Then*

$$\lim_{N\to\infty}\left|\mathbb{E}^{\log}_{n\leq N}a(n)b(n)\right| = 0.$$

We also obtain a conditional version of this result.

Theorem 2.1.9. *Let κ be a natural number. Set $t=\binom{\kappa+1}{2}$. Let $a\colon\mathbb{N}\to S^1$ be an unpretentious completely multiplicative function taking values in the unit circle so that the local $\kappa-1$-Fourier uniformity conjecture holds for a. Let $b\colon\mathbb{N}\to\mathbb{C}$ be a finite-valued 1-bounded*

function. Suppose further that for some $\epsilon > 0$ there are infinitely many k such that the number of words of b of length k that occur with positive upper logarithmic density is at most $k^{t-\varepsilon}$. Then

$$\lim_{N\to\infty} \left|\mathbb{E}^{\log}_{n\leq N} a(n)b(n)\right| = 0.$$

We note that this result matches the numerology in [Saw20] and may be almost the best possible result one can obtain with purely dynamical methods. We also note that even the $1-$Fourier uniformity conjecture is still unknown and so this theorem currently has no unconditional content. We also obtain a version of the theorem where b need not take only finitely many values and we only have information about the number of "approximate" words.

Definition 2.1.10. *We say a sequence b has at most h words of length k up to ε rounding if there exists a set Σ of words of length k such that for all $n \in \mathbb{N}$ there is an ϵ in Σ such that $|b(n+j) - \epsilon_j| \leq \varepsilon$ for all $j \leq k$ and the cardinality of Σ is at most h. We say b has at most h words of length k that occur with positive logarithmic density up to ε rounding if we only require $|b(n+j) - \epsilon_j| \leq \varepsilon$ for a set of n of lower logarithmic density 1.*

Theorem 2.1.11. *Let $c > 0$ and $\varepsilon > 0$. Then if ε is sufficiently small depending on c then the following holds: Let $a\colon \mathbb{N} \to S^1$ be an unpretentious completely multiplicative function taking values in the unit circle. Let $b\colon \mathbb{N} \to \mathbb{C}$ be a 1-bounded function with entropy zero. Suppose further that for every $\delta > 0$ there are infinitely many k such that the number of words of b of length k that occur with positive logarithmic density up to ε rounding is at most δk. Then*

$$\limsup_{N\to\infty} \left|\mathbb{E}^{\log}_{n\leq N} a(n)b(n)\right| \leq c.$$

In fact, this works for any ε satisfying $c^2 > 2\varepsilon$.

We list a few new applications of this theorem.

Proof of Theorem 2.1.5. Apply Theorem 2.1.8 to $a = b = \lambda$. □

Theorem 2.1.12. *If S is a finite set of sequences of subquadratic word growth and a is an unpretentious completely multiplicative function taking values in the unit circle then*

$$\lim_{H\to\infty}\lim_{N\to\infty}\mathbb{E}^{\log}_{n\leq N}\sup_{\phi\in S}|\mathbb{E}_{h\leq H}a(n+h)\phi(h)| = 0.$$

Remark 2.1.13. *We remark that since the set S is finite, it is enough to show that any one function does not locally correlate with a. However, we also remark that it is generally harder to show that a does not locally correlate with b than it is to show that a does not correlate with b. For Theorem 2.1.12, we need to use that Theorem 2.1.8 allows us to handle the case where b may have many words which occur with* 0 log*-density but still only subquadratically many words which occur with positive* log*-density. Theorem 2.1.12 in the linear word growth case seems to follow implicitly from* [GLdLR19].

Proof. For convenience, we will assume that 0 is in S. Let $\varepsilon > 0$. We aim to show that

$$\limsup_{H\to\infty}\limsup_{N\to\infty}\mathbb{E}^{\log}_{n\leq N}\sup_{\phi\in S}|\mathbb{E}_{h\leq H}a(n+h)\phi(h)| = O(\varepsilon).$$

Suppose not. We will now use an argument of [Tao17a] (see Section 5 of that paper) to show that a must be correlate with a "ticker tape" function. We define S_ε to be the set of sequences of the form $\phi'(n) = e(\alpha)\phi(n)$ where ϕ is an element of S and α is a rational number with denominator $O(\varepsilon)$. By the pigeonhole principle, for any ϕ in S and any natural numbers H and n in $\mathbb{N}$ there exists α a rational number with denominator $O(\varepsilon)$ such that

$$\mathrm{Re}\Big(|\mathbb{E}^{\log}_{h\leq H}\phi(h)a(n+h)| - \mathbb{E}^{\log}_{h\leq H}e(\alpha)\phi(h)a(n+h)\Big) = O(\varepsilon).$$

Therefore, we may assume for the sake of contradiction that for some $\phi_{n,H}$ in S_ε

$$\limsup_{H\to\infty}\limsup_{N\to\infty}\mathrm{Re}\ \Big(\mathbb{E}^{\log}_{n\leq N}\mathbb{E}_{h\leq H}a(n+h)\phi_{n,H}(h)\Big) \gg \varepsilon.$$

By a diagonalization argument, we may find a sequence H_i and N_i of natural numbers both tending to infinity and functions $\phi_{n,i} = \phi_{n,H_i}$ such that $N_{i+1} \gg N_i \gg H_i$ and

$$\lim_{i\to\infty}\mathrm{Re}\ \Big(\mathbb{E}^{\log}_{n\leq N_i}\mathbb{E}_{h\leq H_i}a(n+h)\phi_{n,i}(h)\Big) \gg \varepsilon.$$

Since the functions ϕ in S_ε and a are bounded, for i sufficiently large there exists a set A_i of natural numbers of lower logarithmic density $\gg \varepsilon$ in the interval $[1, N_i]$ such that for n in A_i,

$$\mathrm{Re}\left(\mathbb{E}_{h\leq H_i} a(n+h)\phi_{n,i}(h)\right) \gg \varepsilon.$$

By a greedy algorithm, we can select a subset B_i of A_i of upper logarithmic density at least $\frac{\varepsilon}{H_i}$ in $[1, N_i]$ that is at least H_i separated (meaning distinct points of B_i differ by at least H_i). Now define the "ticker tape" function ψ as follows:

$$\psi(n+h) = \phi_{n,i}(h),$$

for all n in B_i between N_{i-1} and N_i and $h \leq H_i$. If m is not of the form $n+h$ for n in B_i between N_{i-1} and N_i and $h \leq H_i$ then we set $\psi(m) = 0$. Thus,

$$\limsup_{N\to\infty} \mathrm{Re}\left(\mathbb{E}^{\log}_{n\leq N} a(n)\psi(n)\right) \gg \varepsilon^2.$$

Now we aim to show that ψ has subquadratically many words of length k that occur with positive upper logarithmic density. Let k be a natural number and let ϵ be a word of length k which occurs in ψ with positive upper logarithmic density. Consider the set C of natural numbers m such that m is within k of an element n of B_i or $B_i + H_i$ for some i. Then since elements of B_i are at least H_i separated, the upper logarithmic density of C in $[N_{i-1}, N_i]$ is at most $\frac{2k}{H_i}$ which clearly tends to 0 as i tends to infinity. Since $N_i \gg N_{i-1}$, we may assume that the log-density of $[1, N_{i-1}]$ in the interval $[1, N_i]$ is also $o(1)$. Thus, C has log-density 0. Therefore, if ϵ occurs with positive log-density then $(\psi(n+h))_{h=1}^k = \epsilon$ for a positive density set of n not in C. Since S_ε has only finitely many members, we get that there exists ϕ in S_ε such that for a positive upper logarithmic density set of n, $\psi(n+h) = \phi(n+h) = \epsilon_h$ for all $h \leq k$. Thus, ψ has subquadratic word growth and a does not correlate with ψ by Theorem 2.1.8, which gives a contradiction. □

Theorem 2.1.14. *Let C be a subset of $[0,1]$ of upper box dimension < 1. Then if a is an unpretentious completely multiplicative function taking values in the unit circle*

$$\lim_{H\to\infty} \lim_{N\to\infty} \mathbb{E}^{\log}_{n\leq N} \sup_{\alpha\in C} |\mathbb{E}_{h\leq H} a(n+h)e(h\alpha)| = 0.$$

Remark 2.1.15. *In particular, this implies that if C is the middle thirds Cantor set then*

$$\lim_{H\to\infty}\lim_{N\to\infty}\mathbb{E}^{\log}_{n\leq N}\sup_{\alpha\in C}|\mathbb{E}_{h\leq H}a(n+h)e(h\alpha)|=0.$$

Of course, the result also applies to a large family of other fractals. The author does not know of any results in the literature where this is established for any infinite set. He does not know of any proof for any set of positive box dimension which does not use Theorem 2.1.11.

Proof. Suppose the upper box dimension of $C\subset S^1$ is $<d<1$. Let $\varepsilon>0$. As in the proof of Theorem 2.1.12, we assume that

$$\limsup_{H\to\infty}\limsup_{N\to\infty}\mathbb{E}^{\log}_{n\leq N}\sup_{\alpha\in C}|\mathbb{E}_{h\leq H}a(n+h)e(h\alpha)|\gg\varepsilon,$$

and derive a contradiction. As before, there is a ticker tape function $\psi\colon\mathbb{N}\to\mathbb{C}$ such that

$$\limsup_{N\to\infty}\mathrm{Re}\left(\mathbb{E}^{\log}_{n\leq N}\mathbb{E}_{h\leq H}a(n+h)\psi(h)\right)\gg\varepsilon^2,$$

of the following form: there exists sequences of natural numbers N_i and H_i tending to infinity with $N_{i+1}\gg N_i\gg H_i$, a sequence of H_i-separated sets B_i, and $\psi(n+h)=e(\beta_n)e(\alpha_n h)$ for some rational β_n of denominator at most $O(\varepsilon)$, some α_n in C and for all n in some set $B_i\cap[N_{i-1},N_i]$ and $h\leq H_i$. We set $\psi(m)=0$ for all natural numbers m not of this form. As before, for any natural number k, the natural numbers m that are within k of a number n in B_i or B_i+H_i has log-density 0.

Let k be a natural number sufficiently large depending on C and ε. Let $\varepsilon'=\varepsilon^2$. Then because C has upper box dimension $<d$ there exists a collection of at most $(\frac{k}{\varepsilon'})^d$ intervals of length $\frac{\varepsilon'}{k}$ covering C. If two points on the circle α and α' differ by at most $\frac{\varepsilon'}{k}$ then by the triangle inequality, for all $h\leq k$, we have that $|e(h\alpha)-e(h\alpha')|\leq\varepsilon'$. Therefore, the number of sign patterns of ψ that occur with positive log-density up to ε' rounding is sublinear. In particular, for any $\delta>0$, there are fewer than δk many sign patterns that occur with positive log-density up to ε^2 rounding. By Theorem 2.1.11, we get a contradiction. □

On the surface, this argument appears to be very close to the t-Fourier uniformity conjecture, which Tao introduced in [Tao17a] and proved was equivalent to the log-Chowla and log-Sarnak conjectures. (For recent significant progress on the Fourier uniformity conjecture, see [MRT18]). If you wanted to prove the Fourier uniformity conjecture in the case $d = 1$, namely that

$$\lim_{H\to\infty} \limsup_{N\to\infty} \mathbb{E}^{\log}_{n\leq N} \sup_{\alpha\in\mathbb{R}} |\mathbb{E}_{h\leq H} a(n+h)e(h\alpha)| = 0,$$

the ticker tape functions that you would need λ to be orthogonal to have $\sim \varepsilon^{-1}k$ many sign patterns of length k up to ε rounding. Thus, one might hope that a simple argument could adjust the constants in Theorem 2.1.11 and thereby prove the Fourier uniformity conjecture. However, there is a major theoretical obstacle to further progress. [FH18b] introduced the dynamical system $(S^1 \times S^1, dx, T, \mathcal{B})$ where $T(\alpha, \beta) = (\alpha, \alpha\beta)$. [Saw20] showed that this dynamical system with some additional structure is a dynamical model for the Liouville function (a notion which we will precisely define later). This is an obstruction to solving the Fourier uniformity conjecture purely with dynamical methods and without any new input from number theory. [Saw20] further showed that there are dynamical models for the Liouville function which have only polynomially many sign patterns. Explicitly, consider the following function $\tilde{a}$ which behaves almost like a multiplicative function: we partition the natural numbers into intervals with the length of the intervals slowly tending to infinity. For instance, we could split all the numbers between 10^{10^n} and $10^{10^{n+1}}$ into blocks of length $\sim n$. Then on each interval I we pick a random phase α_I in S^1 uniformly and indepently. Then we set $\tilde{a}$ to be the function obtained by rounding the function which sends $n \mapsto e(\alpha_I n)$ for n in I. In formulas, we set $\tilde{a}(n) = 2\mathbb{1}_{\mathrm{Re}\, e(\alpha_I n)>0} - 1$ for n in I. We remark that the dynamical model for this sequence is isomorphic to the product of the dynamical system introduced by [FH18b] with $\widehat{\mathbb{Z}}$ (again, we defer the precise definition until later). Clearly, $\tilde{a}$ is not multiplicative. However, it is "statistically" multiplicative in the sense that, with high probability, for any sign pattern ϵ of length k, for any m and for large N

$$\mathbb{E}^{\log}_{n\leq N} \mathbb{1}_{\tilde{a}_{n+h}=\epsilon_h \text{ for all } h\leq k} \approx \mathbb{E}^{\log}_{n\leq N} m\mathbb{1}_{m|n}\mathbb{1}_{\tilde{a}_{n+mh}=-\epsilon_h \text{ for all } h\leq k}.$$

This function clearly does not satisfy the 1-Fourier uniformity conjecture and [Saw20] showed that it has quadratically many sign patterns that occur with positive upper logarithmic density even though it does satisfy the a version of [MRT15]. If we had used a random $\kappa-1$-degree polynomial instead of a random linear polynomial, we would get a function which is again statistically multiplicative but which fails the $\kappa-1$-Fourier uniformity conjecture and [Saw20] showed that it has $\lesssim k^{\binom{\kappa+2}{2}}$ many sign patterns of length k. However, the author is unaware of any "dynamical" techiniques that distinguish these statistically multiplicative functions from the Liouville function. This is made precise with Definition 2.1.18.

We give one last application.

Theorem 2.1.16. *Again, suppose that a is an unpretentious completely multiplicative function taking values in the unit circle. There is a set $C \subset [0,1]$ of Hausdorff dimension 1 such that*

$$\lim_{H\to\infty} \limsup_{N\to\infty} \mathbb{E}^{\log}_{n\leq N} \sup_{\alpha\in C} |\mathbb{E}_{h\leq H} a(n+h)e(h\alpha)| = 0.$$

Proof. The main idea is to combine Theorem 2.1.14 with a diagonalization argument. For a disjoint collection of intervals $\mathcal{J} = \{J\}$ and a natural number n we define $D_n(\mathcal{J})$ to be the set of intervals obtained by taking each J, removing a ball of diameter $\frac{|J|}{n}$ around the center of the interval J, taking the two remaining intervals, then taking the union over all J in $\mathcal{J}$.

We construct C inductively as follows. Start with any interval I and set $\mathcal{J}_2 = \{I\}$. Assume inductively that we have constructed $\mathcal{J}_{n-1}$ Then we apply D_n again and again. Let

$$C_n = \cap_{m\in\mathbb{N}} \cup_{J\in D^m \mathcal{J}_{n-1}} J.$$

Since C_n has box dimension $\frac{\log n-1}{\log n}$, we know by Theorem 2.1.14 that there exists a natural number H_n such that if $H \geq H_n$ then

$$\limsup_{N\to\infty} \mathbb{E}^{\log}_{n\leq N} \sup_{\alpha\in C} |\mathbb{E}_{h\leq H} a(n+h)e(h\alpha)| \leq \frac{1}{n}.$$

Then define

$$\mathcal{J}_n = D_n^{nH_n} \mathcal{J}_{n-1}.$$

We set

$$C = \cap_{n\geq 2} \cup_{J\in\mathcal{J}_n} J.$$

Clearly, the Hausdorff dimension of C is at least $\frac{-\log 2}{\log(\frac{n-1}{2n})}$ for every n and therefore the Hausdorff dimension is precisely 1. Now we verify that C has the desired property. For each natural number m, by enlarging the set we are maximizing over, we have that

$$\begin{aligned}&\limsup_{H\to\infty}\limsup_{N\to\infty}\mathbb{E}^{\log}_{n\leq N}\sup_{\alpha\in C}|\mathbb{E}_{h\leq H}a(n+h)e(h\alpha)|\\ \lesssim&\limsup_{N\to\infty}\mathbb{E}^{\log}_{n\leq N}\sup_{\alpha\in J\in\mathcal{J}_m}|\mathbb{E}_{h\leq H_m}a(n+h)e(h\alpha)|.\end{aligned}$$

Since every element $\alpha \in J \in \mathcal{J}_m$ is in $D_m^{mH_m}(\mathcal{J}_{m-1})$ there exists $\beta = \beta_\alpha$ depending on α such that β is in C_m and the distance from α to β is no more than $\frac{1}{H_m m}$. Therefore, for all $h \leq H_m$, αh is within $\frac{1}{m}$ of βh. Thus,

$$\lesssim\limsup_{N\to\infty}\mathbb{E}^{\log}_{n\leq N}\sup_{\beta\in C_m}|\mathbb{E}_{h\leq H_m}a(n+h)e(h\beta)|+\frac{1}{m}.$$

However, by our choice of H_m, we have

$$\lesssim\frac{1}{m}.$$

Since m was arbitrary, we obtain the desired result. □

Remark 2.1.17. *We have stated our main theorems in the case that a is completely multiplicative and takes values in the unit circle. We remark that these assumptions can be weakened to include all multiplicative functions taking values in the unit disk. The reduction from multiplicative functions taking values in the unit disk to multiplicative functions taking values in the unit circle is essentially due to Tao (see* [Tao16b]*, Proposition 2.1). The reduction from multiplicative functions to completely multiplicative functions (say both taking values in the unit circle) is carried out in section 2.5. The argument is rather short and was essentially communicated to me by Tao. However, it may be more broadly known and I make no claim of originality.*

We now sketch an outline of an argument that is morally very similar to the main argument in this chapter. However, for the moment we will work in a more concrete setting. To make this argument rigorous, it is much easier to pass to the dynamical context. Suppose that b is a sequence with quadratic word growth rate and that

$$\limsup_{N\to\infty} |\mathbb{E}^{\log}_{n\leq N}\lambda(n)b(n)| > c > 0.$$

Then we can fix a natural number k and average over translates,

$$\limsup_{N\to\infty} |\mathbb{E}^{\log}_{n\leq N}\mathbb{E}_{h\leq k}\lambda(n+h)b(n+h)| > c.$$

Fix a large natural number P with $N \gg P \gg k$. Because λ also has a multiplicative symmetry, we can average over dilates

$$\limsup_{N\to\infty} |\mathbb{E}^{\log}_{n\leq N}\mathbb{E}_{P/2<p\leq P}\mathbb{E}_{h\leq k}\lambda(pn+ph)b(n+h)| > c.$$

Moving the absolute values inside and crudely replacing b by the worst word of length k, we get

$$\limsup_{N\to\infty} \mathbb{E}^{\log}_{n\leq N}\sup_{\epsilon}\mathbb{E}_{P/2<p\leq P}|\mathbb{E}_{h\leq k}\lambda(pn+ph)\epsilon_h| > c,$$

where the supremum is taken over all words ϵ of b. Tao's entropy decrement argument, introduced in [Tao16b], allows us to replace pn by n.

$$\limsup_{N\to\infty} \mathbb{E}^{\log}_{n\leq N}\sup_{\epsilon}\mathbb{E}_{P/2<p\leq P}|\mathbb{E}_{h\leq k}\lambda(n+ph)\epsilon_h| > c.$$

Now if λ behaves randomly, then we already know that λ is orthogonal to b. Therefore, if λ correlates with b it must have some structure. Morally, [FH18b] says we can break up λ into a structured part and a random part, and that all the correlation comes from the structured part. [HK05] proves that the structured part must take the form of a nilsequence. For the purposes of this sketch, we will focus on the case that there exists α_n and β_n irrational such that

$$\limsup_{N\to\infty} \mathbb{E}^{\log}_{n\leq N}\sup_{\epsilon}\mathbb{E}_{P/2<p\leq P}|\mathbb{E}_{h\leq k}e(\alpha_n(ph)^2+\beta_n ph)\epsilon_h| > c.$$

By Hölder's inequality,

$$\limsup_{N\to\infty} \mathbb{E}^{\log}_{n\leq N} \sup_{\epsilon} \mathbb{E}_{P/2<p\leq P}|\mathbb{E}_{h\leq k} e(\alpha_n (ph)^2 + \beta_n ph)\epsilon_h|^4 > c^4.$$

By the pigeonhole principle, since there are only δk^2 many sign patterns, there is a sign pattern ϵ such that,

$$\limsup_{N\to\infty} \mathbb{E}^{\log}_{n\leq N} \mathbb{E}_{P/2<p\leq P}|\mathbb{E}_{h\leq k} e(\alpha_n (ph)^2 + \beta_n ph)\epsilon_h|^4 > \delta^{-1}k^{-2}c^4.$$

Expanding everything out and using that $\delta \leq \frac{c^4}{2}$

$$\limsup_{N\to\infty} \mathbb{E}^{\log}_{n\leq N}|\mathbb{E}_{P/2<p\leq P}\mathbb{E}_{j\in[k]^4} e(\alpha_n p^2(j_1^2 + j_2^2 - j_3^2 - j_4^2) + \beta_n p(j_1 + j_2 - j_3 - j_4))| > 2k^{-2}.$$

When $j_1^2 + j_2^2 - j_3^2 - j_4^2 \neq 0$ or $j_1 + j_2 - j_3 - j_4 \neq 0$ then for P large, by the circle method

$$\mathbb{E}_{P/2<p\leq P} e(\alpha_n p^2(j_1^2 + j_2^2 - j_3^2 - j_4^2) + \beta_n p(j_1 + j_2 - j_3 - j_4)) \approx 0.$$

The analogue of the circle method for more general nilpotent Lie groups was introduced in [GT12a], [GT10] and [GTZ12]. The analogue of the step where we conclude that the sums of powers is 0 for more general nilpotent Lie groups is an argument of [Fra17]. Thus the only contribution is from the terms where $j_1^2 + j_2^2 - j_3^2 - j_4^2 = 0$ and $j_1 + j_2 - j_3 - j_4 = 0$. But it is easily seen from Newton's identities for symmetric polynomials that this only happens for the $2k^2$ "diagonal" terms. Thus, we get

$$2k^{-2} > 2k^{-2},$$

which of course provides a contradiction. For the proof of Theorem 2.1.9, we need to not only use the theory of symmetric polynomials but also use [BDG16].

2.1.1 Background and notation

Suppose $a(n)$ is a 1-bounded, unpretentious multiplicative function with $|a(n)| = 1$ for all n. Let $b(n)$ a sequence where only $o(k^2)$ or $O(k^{t-\varepsilon})$ many sign patterns occur with positive

log-density. The usual construction of a Furstenberg system (see [FKO82]) for (a, b) proceeds as follows: consider the point (a, b) in the space of pairs of sequences. Then apply a random shift to this deterministic variable, $(T^n a, T^n b)$. This gives a random variable in the space of pairs of sequences. The distribution of this random variable is then a shift invariant measure on the space of pairs of sequences. Furthermore, if f is the function on the space of pairs of sequences that evaluates the first sequence at 1 and f' is the function which evaluates the second sequence at 1 then

$$f(T^n a, T^n b) f'(T^n a, T^n b) = a(n+1)b(n+1),$$

which is the sequence whose average value we care about. Therefore, if the average of $a(n)b(n)$ is greater than c in absolute value then

$$\left| \int f \cdot f' \right| > c,$$

as well. Of course, it does not really make sense to take a random natural number. Instead, one must shift by a random natural number in a large but finite interval whose length tends to infinity, then find a subsequence of the random variables that converges in distribution. This corresponds to taking a weak-$*$ limit of the corresponding measures.

However, we take a slightly modified approach. The reason is that the function a has some additional symmetry, namely $a(nm) = a(n)a(m)$. As such, the probability that some word occurs i.e., that $a(n+h) = \epsilon_h$ for $h = 1, \ldots, k$ and for n randomly chosen between 1 and N is the same as the probability that $a(pn + ph) = a(p) \cdot \epsilon_h$ for $h = 1, \ldots, k$ and for n chosen randomly between 1 and N. That's the same as p times the probability that for a randomly chosen n between 1 and pN one has $a(n + ph) = a(p) \cdot \epsilon_h$ for $h = 1, \ldots, k$ and p divides n. Just flipping everything around, the probability that a random n between 1 and pN satisfies $a(n + ph) = a(p) \cdot \epsilon_h$ and is divisible by p is $\frac{1}{p}$ times the probability that a random n between 1 and N satisfies $a(n + h) = \epsilon_h$. We want our dynamical system to capture this symmetry. There are two difficulties which arise when we want to translate this symmetry to our dynamical system. The first is that the interval keeps changing: the

distribution of $T^n a$ might be very different on the intervals from 1 to N and from 1 to pN so when we take a weak limit along a subsequence of intervals, the distribution $T^n a$ for shifts in one interval might approximate our invariant measure while shifts along the other interval might not. The fix for this problem is to use log-averaging. After we weight each natural number n by $\frac{1}{n}$, the probability that a random n will be between N and pN is $\sim \frac{\log p}{\log N}$ which tends to 0 as N tends to infinity. Therefore, the distribution of $T^n a$ for a random n between 1 and N is very close to the distribution of $T^n a$ for a random n between 1 and pN as long as we choose n randomly using logarithmic weights. The other problem is that our dynamical system does not have a good notion of "being divisible" by a number. To remedy this, we make use of the profinite completion of the integers

$$\widehat{\mathbb{Z}} = \prod_p \mathbb{Z}_p,$$

where p is always restricted to be a prime and $\mathbb{Z}_p$ is the p-adic integers $\mathbb{Z}_p = \lim_{\leftarrow} \mathbb{Z}/p^k\mathbb{Z}$ i.e. the inverse limit of $\mathbb{Z}/p^k\mathbb{Z}$ for all k. For each natural number n, we get an element of $\widehat{\mathbb{Z}}$ by reducing n mod p^k for every prime p and every natural number k. Then to build our dynamical system, we take the space of triples consisting of two sequences and a profinite integer and for a logarithmically randomly chosen integer n we consider the random variable $(T^n a, n, T^n b)$ in this space. The distribution of this random variable is a shift invariant measure. Furthermore, we have the following symmetry: let $Y = \overline{\{T^n b \colon n \in \mathbb{N}\}}$ and $X = (S^1)^{\mathbb{N}} \times \widehat{\mathbb{Z}}$. Define the function

$$M \colon X \to \widehat{\mathbb{Z}},$$

by projecting onto the $\widehat{\mathbb{Z}}$ coordinate in X,

$$M \colon (\alpha, r) \mapsto r.$$

Define the function

$$I_p \colon M^{-1}(p\widehat{\mathbb{Z}}) \to X,$$

by "zooming in" by a factor of p and multiplying by $\overline{a(p)}$ on the first factor and dividing by p on the second,

$$I_p \colon (\alpha(n), r) \mapsto (\overline{a(p)}\alpha(pn), r/p),$$

where r/p is the unique element of $\widehat{\mathbb{Z}}$ such that $p \cdot (r/p) = r$. Then if ν is our invariant measure on $X \times Y$ and μ is its first marginal then I_p pushes forward μ restricted to $M^{-1}(p\widehat{\mathbb{Z}})$ to $\frac{1}{p}\mu$. Formally, we make the following definition:

Definition 2.1.18. *Let (X, μ, T) be a dynamical system, let $f \colon X \to \mathbb{C}$ be a measurable function, let $M \colon X \to \widehat{\mathbb{Z}}$ be a measurable function and for each m let $I_m \colon M^{-1}(m\widehat{\mathbb{Z}}) \to X$ be a measurable function. We say (X, μ, T, f, M, I_m) is a dynamical model for a if,*

- *$M \circ T = M + 1$ almost everywhere.*
- *$I_m \circ T^m = T \circ I_m$ almost everywhere in $M^{-1}(m\widehat{\mathbb{Z}})$.*
- *I_m pushes forward the measure μ restricted to $M^{-1}(m\widehat{\mathbb{Z}})$ to $\frac{1}{m}\mu$. Symbolically, for any function ϕ in $L^1(\mu)$ we have*

$$\int_X \phi(x)\mu(dx) = \int_X m\mathbb{1}_{x \in M^{-1}(m\widehat{\mathbb{Z}})}\phi(I_m(x))\mu(dx).$$

- *$f \circ I_m = \overline{a(m)} \cdot f$ almost everywhere in $M^{-1}(m\widehat{\mathbb{Z}})$.*
- *For all m and n, $I_{nm} = I_n \circ I_m$ almost everywhere in $M^{-1}(mn\widehat{\mathbb{Z}})$.*

We also ask for the following property that [Saw20] *does not impose.*

- *For any natural number m and any measurable subset A of $\mathbb{C}^m$,*

$$\mu\{x \in X \colon (f(T^1x), \ldots, f(T^mx)) \in A\} \leq$$
$$d^{\log}\{n \leq N \colon (a(n+1), \ldots, a(n+m)) \in A\},$$

where $d^{\log}$ denotes upper logarithmic density. We remark that we can also fix a Banach limit $p - \lim$ extending the usual limit functional and require that equality holds in the

previous equation holds for any limit taken with respect to that Banach limit. For more details, see [Tao17b].

Let $(X \times Y, \nu, T)$ be a joining of two dynamical systems X and Y. Suppose that μ is the first marginal and (X, μ, T, f, M, I_m) is a dynamical model for a. Let f' be a measurable function on $X \times Y$ which is Y measurable. We say $(X \times Y, \nu, T, f, f', M, I_m)$ is a joining of a dynamical model of a with b if we also have that, for any natural number m and any measurable subset A of $\mathbb{C}^m$,

$$\nu\{(x, y) \in X \times Y : (f'(T^1 y), \dots, f'(T^m y)) \in A\} \leq$$
$$d^{\log}\{n \leq N : (b(n+1), \dots, b(n+m)) \in A\}$$

where $d^{\log}$ denotes upper logarithmic density. We could also require that the joint statistics of (f, f') agree with the joint statistics of (a, b) but this is not necessary for our argument.

Remark 2.1.19. *The preceding definition was used first in* [Tao17b] *and generalized in* [Saw20].

We abuse notation and denote all transformations by the letter T. We also remark that for the proof of Theorems 2.1.8 and 2.1.9 that f' only takes finitely many values.

We now specify some notation used in the main argument:

- We fix an unpretentious 1-bounded multiplicative function a. (For the definition of unpretentious, see [MRT15]; we will only really use that a is unpretentious in Theorem 2.2.1; we remark that the Liouville function is unpretentious). We fix constants $t \in \mathbb{N}$, $c > 0$ and $\delta > 0$. We fix a 1-bounded function b with at most $o(k^2)$ or $k^{t-\varepsilon}$ many words of length k occurring with positive upper logarithmic density for all $k \in \mathcal{K}$ where $\mathcal{K}$ is some fixed infinite set. We suppose that

$$\limsup_{N \to \infty} |\mathbb{E}^{\log}_{n \leq N} a(n) b(n)| > c.$$

We fix $\eta > 0$ such that

$$\limsup_{N\to\infty} |\mathbb{E}^{\log}_{n\leq N} a(n)b(n)| > c + \eta.$$

- We use the following theorem of [FH18a].

  Theorem 2.2.8 ([FH18a] Theorem 1.5)**.** *There exists a joining of a dynamical model for a with b,* $(X \times Y, T, \nu, f, f', M, I_m)$ *satisfying*

  $$\left|\int_{X\times Y} f(x,y)f'(x,y)\nu(dxdy)\right| > c + \eta,$$

  and if μ *is the first marginal then the ergodic components* (X, μ_ω, T) *are isomorphic to products of Bernoulli systems with the Host-Kra factor of* (X, μ_ω, T)*.*

  Because the statement here is slightly different than Theorem 1.5 in [FH18a], we will go through the details in section 2.4. We fix such a system. We will always denote by μ the first marginal of ν. We also fix ergodic decompositions $\nu = \int_\Omega \nu_\omega d\omega$ and $\mu = \int_\Omega \mu_\omega d\omega$. We define the words of length k of f' to be those words ϵ of length k such that the set of (x, y) such that $f'(T^h x, T^h y) = \epsilon_h$ for all $h \leq k$ has positive measure. We note that the set of words of f' is a subset of the set of words of b that occur with positive upper log density: after all, if $f'(T^h x, T^h y) = \epsilon_h$ then by definition of a joining of a dynamical model a with b,

  $$\begin{aligned} 0 <& \mu\{(x,y) \in X \times Y \colon f'(T^h y) = \epsilon_h \text{ for all } h \leq k\} \\ \leq& \mathrm{d}^{\log}\{n \in \mathbb{N} \colon b(n+h) = \epsilon_h \text{ for all } h \leq k\}, \end{aligned}$$

  where $\mathrm{d}^{\log}$ denotes upper logarithmic density.

- G will always refer to a nilpotent Lie group. G_s will always refer to the s^{th} step in the lower central series. Γ will always refer to a cocompact lattice in G, meaning that G_s/Γ_s is compact for every s. g, σ and τ will always refer to group elements. $\mathcal{B}$ will always refer to the Borel sigma algebra. We will fix a particular G, Γ and g following Corollary 2.2.20. For more on this see [GT12b].

- For a nonempty, finite set A and $\phi\colon A \to \mathbb{C}$, we denote $\mathbb{E}_{n\in A}\phi(n) = \frac{1}{\#A}\sum_{n\in A}\phi(n)$. For $A \subset \mathbb{N}$, we denote
$$\mathbb{E}^{\log}_{n\in A}\phi(n) = \frac{1}{\sum_{n\in A}\frac{1}{n}}\sum_{n\in A}\frac{\phi(n)}{n}.$$
This notation is due to Frantzikinakis (see [Fra17]). We always restrict p to be prime. By definition a nilsystem is a dynamical system $(G/\Gamma, dx, T, \mathcal{B})$ where G is a nilpotent Lie group, Γ is a cocompact subgroup, dx is Haar measure, there exists g such that $T(x) = gx$ and $\mathcal{B}$ is the Borel sigma algebra. A nilsequence is a sequence of the form $F(g^n\Gamma)$ where G is a nilpotent Lie group, Γ is a cocompact lattice in G, g is an element in G and and $F\colon G/\Gamma \to \mathbb{C}$ is a continuous function. Suppose G is an s-step nilpotent Lie group so that G_s is an abelian group and G_s/Γ_s is a compact abelian group. Then a nilcharacter Φ is a function $G/\Gamma \to \mathbb{C}$ such that there exists a character $\xi\colon G_s/\Gamma_s \to S^1$ called the frequency of Φ such that, for all x in G/Γ and u in G_s we have $\Phi(ux) = \xi(u\Gamma_s)\Phi(x)$. We will abuse notation and identify ξ with the function on G_s that maps $u \mapsto \xi(u\Gamma_s)$. We say ξ is nontrivial if there exists u in G_s such that $\xi(u) \neq 1$. We say ξ is nontrivial on the identity component if we can find a u in the identity component of G_s such that $\xi(u) \neq 1$.

- For Theorem 2.2.14, we will adopt conventions from the theory of Shannon entropy. In particular, $H(x)$ will denote the Shannon entropy of x and $I(x, y)$ will denote the mutual information between x and y. For more details, see [Tao16b].

- We will always denote by $\mathcal{Z}$ the smallest sigma algebra on X generated by the union of the sigma algebras corresponding each of the Host-Kra factors. We will denote
$$B = \{(x, y) \in X \times Y\colon f'(T^n y) \text{ is eventually periodic as a function of } n\}.$$
Since whether (x, y) is in B depends only y, we will abuse notation and also use $B = \{y \in Y\colon f'(T^n y) \text{ is eventually periodic as a function of } n\}$.

- For a complex numbers z, a set A and a real number w we say $z = O_A(w)$ and $z \lesssim_A w$ if there exists a constant C depending on A but not z and w such that $|z| \leq Cw$. If

there are more subscripts we mean that the constant may depend on more parameters. For instance, by $\lesssim_{A,u,K}$ we mean that the implied constant can depend on A, u and K.

2.1.2 Acknowledgments

Special thanks to Terence Tao for sharing many ideas on an earlier version of this paper and for his many helpful comments. Also, special thanks to Nikos Frantzikinakis for pointing out a number of ways to strengthen the main theorem of this chapter. I would also like to thank Tim Austin, Björn Bringmann, Alex Dobner, Asgar Jamneshan, Bernard Host, Gyu Eun Lee, Zane Li, Adam Lott, Maksym Radziwiłł, Bar Roytman, Chris Shriver, Joni Teräväinen and Alex Wertheim for many helpful discussions. Thanks to Will Sawin for suggesting I use Vinogradov's mean value theorem to improve an earlier version of Theorem 2.1.9. I would lastly like to thank the anonymous reviewer for their extremely helpful comments. Some of this work was completed while the author was at the American Institute for Mathematics workshop on Sarnak's conjecture.

2.2 Main Argument

In this section, we prove Theorem 2.1.8 and Theorem 2.1.9. In Section 2.3, we explain how to adapt the proof to handle Theorem 2.1.11.

We remark that much of the notation, including X, Y, μ, ν, f, f', a, and b was defined in Subsection 2.1.1.

We start off with a theorem by [MRT15], relying on work in [MR16]. This is a special case of our theorem, so it is no surprise that we need this result.

Theorem 2.2.1 ([MRT15] Theorem 1.7; see also [MR16])**.** *Let a be a bounded, non-pretentious*

multiplicative function. Let θ be a periodic sequence. Then

$$\lim_{H\to\infty}\limsup_{N\to\infty}\mathbb{E}^{\log}_{n\leq N}|\mathbb{E}_{h\leq H}a(n+h)\theta(h)|=0.$$

This theorem says that a does not locally correlate with periodic functions. Eventually, we plan to use a local argument. In particular, our argument will only work for those points where f' does not behave locally like a periodic function. Therefore, we need to exclude any contribution to the integral coming from points where f' behaves like a periodic function. That is the content of the following corollary.

Corollary 2.2.2. *Let $B=\{(x,y)\in X\times Y\colon f'(T^ny)$ is eventually periodic as a function of $n\}$. Then*

$$\int_B f(x)f'(y)\nu(dxdy)=0.$$

Proof. In this proof, we introduce some notation which will not be used in the rest of the chapter. Because T preserves ν and because B is T-invariant, we can average over shifts:

$$\int_B f(x)f'(y)\nu(dxdy)=\lim_{H\to\infty}\int_B\mathbb{E}_{h\leq H}f(T^hx)f'(T^hy)\nu(dxdy).$$

We know f' takes only finitely many values. There are only countably many different periodic sequences taking values in a finite alphabet. Therefore, it suffices to prove that if B_θ is the set of points (x,y) on which $f'(T^hy)$ is eventually equal to the periodic function θ that

$$0=\lim_{H\to\infty}\int_{B_\theta}\mathbb{E}_{h\leq H}f(T^hx)f'(T^hy)\nu(dxdy).$$

Let $\varepsilon>0$. By Theorem 2.2.1, for H sufficiently large

$$\varepsilon^3\gg\limsup_{N\to\infty}\mathbb{E}^{\log}_{n\leq N}\sup_{j\in\mathbb{N}}|\mathbb{E}_{h\leq H}a(n+h)\theta(h+j)|. \tag{2.1}$$

We claim that, translating this to the dynamical world using the definition of a dynamical model for a,

$$\nu\{(x,y)\,:\,\limsup_{H\to\infty}|\mathbb{E}_{h\leq H}f(T^hx)\theta(h)|>\varepsilon\}\leq\varepsilon.$$

After all, by Chebyshev's inequality, for any H such that 2.1 holds,

$$\mathrm{d}^{\log}\{n \in \mathbb{N} : \sup_{j \in \mathbb{N}} |\mathbb{E}_{h \leq H} a(n+h)\theta(h+j)| \geq \varepsilon\} \ll \varepsilon^2,$$

where $\mathrm{d}^{\log}$ denotes upper logarithmic density. Fix such an H for the moment and fix a natural number $H' > H$. In fact, more is true. Let S be the subset of the natural numbers such that n is in S if and only if there exists a natural number $H' \geq H_n \geq H$ such that

$$|\mathbb{E}_{h \leq H_n} a(n+h)\theta(h)| \geq \varepsilon.$$

Let S' denote the union of all the intervals $[n+1, n+H_n]$ for n in S. We claim there is a subcollection $\mathcal{I}$ of these intervals which covers S' and such that each natural number is contained in at most two intervals in $\mathcal{I}$. This is a somewhat standard covering lemma, but we include the details for the interested reader. For instance, consider the following construction. Let $\mathcal{I}_0$ denote the empty set. Then assuming we have constructed $\mathcal{I}_\ell$ for some natural number ℓ, let m denote the smallest natural number in S' not contained in the union of the intervals in $\mathcal{I}_\ell$. (If no such m exists, then just set $\mathcal{I}_{\ell+1} = \mathcal{I}_\ell$). Let n be a natural number maximizing $n+H_n$ subject to the constraints that n is in S and m is in $[n+1, n+H_n]$. Such an n exists because m is in S'. Then let $\mathcal{I}_{\ell+1} = \mathcal{I}_\ell \cup \{[n+1, n+H_n]\}$. Now we check that $\mathcal{I} = \cup \mathcal{I}_\ell$ has the desired property. First, for any m in S, m is clearly contained in the union of the intervals in $\mathcal{I}_m$. Thus, $\mathcal{I}$ covers S'. Second, suppose that m in S' is contained in I_1, I_2 and I_3 with I_ℓ chosen before $I_{\ell+1}$ for $\ell = 1, 2$. Suppose that $\mathcal{I}_{\ell_i}$ is the first set of the form $\mathcal{I}_\ell$ where I_i is contained in $\mathcal{I}_{\ell_i}$. Then the union of the intervals in $\mathcal{I}_{\ell_1}$ contains m. Thus, there exists m' in S' such that $I_2 = [n+1, n+H_n]$ was chosen to maximize $n+H_n$ subject to the constraint that m' is in I_2. Since we assumed m was contained in I_2, we have that $n < m$. Now let $I_3 = [n'+1, n'+H_{n'}]$. Since I_3 also contains m, n' is also less than m which is in turn less than m'. But I_2 maximized $n+H_n$ over all intervals containing m' and if $n'+H_{n'}$ were larger than $n+H_n$ which is larger than m', then I_3 would contained m' as well. Thus $n'+H_{n'} \leq n+H_n$ and therefore any point contained in I_3 is already contained in the union of the intervals in $\mathcal{I}_{\ell_2}$. Therefore, I_3 should not have been selected for $\mathcal{I}_{\ell_3}$ which

leads to a contradiction. Thus, every natural number is covered at most twice by the union of the intervals in $\mathcal{I}$. If

$$|\mathbb{E}_{h\leq H_n} a(n+h)\theta(h)| \gtrsim \varepsilon$$

then

$$\mathbb{E}_{h\leq H_n} |\mathbb{E}_{h'\leq H} a(n+h+h')\theta(h+h')| \gtrsim \varepsilon.$$

Therefore, for at least $\varepsilon \cdot H_n$ many points $n+h'$ in the interval $[n+1, n+H_n]$,

$$\sup_{j\in\mathbb{N}} |\mathbb{E}_{h\leq H} a(n+h+h')\theta(h+j)| \gtrsim \varepsilon.$$

However, we know that

$$\mathrm{d}^{\log}\{n\in\mathbb{N} : \sup_{j\in\mathbb{N}} |\mathbb{E}_{h\leq H} a(n+h)\theta(h+j)| \geq \varepsilon\} \ll \varepsilon^2,$$

and that each such natural number is contained in at most two intervals of the form $[n+1, n+H_n]$ in $\mathcal{I}$. We conclude that, by Chebyshev's inequality, the logarithmic density of S' is at most ε. Therefore, the logarithmic density of S is at most ε. This precisely means

$$\mathrm{d}^{\log}\{n\in\mathbb{N} : \sup_{L\in[H,H']} |\mathbb{E}_{h\leq L} a(n+h)\theta(h)| \geq \varepsilon\} \leq \varepsilon.$$

The condition $\sup_{L\in[H,H']} |\mathbb{E}_{h\leq L} a(n+h)\theta(h)| \geq \varepsilon$ depends measurably on $(a(n+1), \ldots, a(n+H'))$ so by definition of a dynamical model for a,

$$\nu\{(x,y) : \sup_{L\in[H,H']} |\mathbb{E}_{h\leq L} f(T^h x)\theta(h)| \geq \varepsilon\} \leq \varepsilon.$$

Since this is true for all H', we get that

$$\nu\{(x,y) : \sup_{L\geq H} |\mathbb{E}_{h\leq L} f(T^h x)\theta(h)| \geq \varepsilon\} \leq \varepsilon.$$

Since this is true for all $\varepsilon > 0$, for all (x,y) outside a set of measure 0, we have

$$\lim_{H\to\infty} \mathbb{E}_{h\leq H} f(T^h x)\theta(h) = 0.$$

For $(x, y) \in B_\theta$, we know that $f'(T^h y) = \theta(h)$ for h sufficiently large so for $(x, y) \in B_\theta$ outside a set of measure 0, we know $\mathbb{E}_{h \leq H} f(T^h x) f'(T^h y) \to 0$. By the dominated convergence theorem, we have

$$0 = \lim_{H \to \infty} \int_{B_\theta} \mathbb{E}_{h \leq H} f(T^h x) f'(T^h y) \nu(dxdy)$$

as desired. □

We will also need the following result later. It states that f does not correlate locally with periodic functions.

Corollary 2.2.3. *Let $\mu = \int_\Omega \mu_\omega d\omega$ be an ergodic decomposition of μ. For almost every ω, for all 1-bounded function $\phi \colon X \to \mathbb{C}$ such that, for μ_ω almost every x, $\phi(T^h x)$ is periodic in h we have*

$$\int_X f(x)\phi(x)\mu_\omega(dx) = 0.$$

Proof. Let d be a natural number. Then we claim that,

$$\lim_{H \to \infty} \lim_{N \to \infty} \mathbb{E}^{\log}_{n \leq N} \sup_{\theta \in S_d} |\mathbb{E}_{h \leq H} a(n+h)\theta(h)| = 0,$$

where S_d is the set of $d!$-periodic, 1-bounded functions. Since the supremum is over a finite set, this directly follows from Theorem 2.2.1. Let $\varepsilon > 0$. For H sufficiently large,

$$\limsup_{N \to \infty} \mathbb{E}^{\log}_{n \leq N} \sup_{\theta \in S_d} |\mathbb{E}_{h \leq H} a(n+h)\theta(h)| \leq \varepsilon^3.$$

Therefore, as in the proof of Proposition 2.2.2

$$\nu\{(x, y) \,:\, \limsup_{H \to \infty} \sup_{\theta \in S_d} |\mathbb{E}_{h \leq H} f(T^h x)\theta(h)| > \varepsilon\} \leq \varepsilon.$$

Since this is true for all ε, we get that

$$\int_X \limsup_{H \to \infty} \sup_{\theta \in S_d} |\mathbb{E}_{h \leq H} f(T^h x)\theta(h)| \mu(dx) = 0.$$

Therefore, there exists $\Omega_d \subset \Omega$ of full measure such that for ω in Ω_d, we have

$$\int_X \limsup_{H \to \infty} \sup_{\theta \in S_d} |\mathbb{E}_{h \leq H} f(T^h x)\theta(h)| \mu_\omega(dx) = 0.$$

Now let ω be an element of Ω_d for all d and let ϕ be a 1-bounded function such that $\phi(T^h x)$ is periodic in h for μ_ω-almost every x. Suppose that there exists $\varepsilon > 0$ such that

$$\Big| \int_X f(x)\phi(x)\mu_\omega(dx) \Big| > \varepsilon.$$

Then by translation invariance, we know

$$\Big| \limsup_{H\to\infty} \int_X \mathbb{E}_{h\leq H} f(T^h x)\phi(T^h x)\mu_\omega(dx) \Big| \geq \varepsilon.$$

Let X_d be the set of all points x such that $\phi(T^h x)$ is periodic with period at most d. Note by assumption that $\mu_\omega(\cup X_d) = 1$. Then by dominated convergence, there exists d such that

$$\int_{X_d} \limsup_{H\to\infty} |\mathbb{E}_{h\leq H} f(T^h x)\phi(T^h x)|\mu_\omega(dx) > .5\varepsilon.$$

Since $\phi(T^h x)$ is $d!$ periodic for every x in X_d, this integral is bounded by

$$\int_X \limsup_{H\to\infty} \sup_{\theta\in S_d} |\mathbb{E}_{h\leq H} f(T^h x)\theta(h)|\mu_\omega(dx),$$

which gives a contradiction. □

For the proof of Theorem 2.1.9, we also need an upgraded version of Corollary 2.2.3 under the assumption that the $\kappa - 1$-Fourier uniformity conjecture holds.

Proposition 2.2.4. *Suppose that the $\kappa-1$-Fourier uniformity conjecture holds i.e., for every nilpotent Lie group G of step $< \kappa$, every cocompact lattice Γ and every continuous function $F\colon G/\Gamma \to \mathbb{C}$*

$$\lim_{H\to\infty} \limsup_{N\to\infty} \mathbb{E}^{\log}_{n\leq N} \sup_{g\in G} |\mathbb{E}_{h\leq H} a(n+h)F(g^h\Gamma)| = 0.$$

Then for almost every ω, we have the following property: for every nilpotent Lie group G of step $< \kappa$, every cocompact lattice Γ, every continuous function $F\colon G/\Gamma \to \mathbb{C}$ and every function ϕ on X such that for μ_ω almost every x there exists x' in G/Γ and g in G we have $\phi(T^h x) = F(g^h x')$ for all h in $\mathbb{N}$ we have that

$$\int_X f(x)\phi(x)\mu_\omega(dx) = 0.$$

Proof. In the proof of this proposition, we will introduce some notation which will not be used in the rest of the chapter. By, for instance, [HK18, Chapter 10, Theorem 28] there are only countably many pairs (G, Γ) up to isomorphism of G/Γ. Thus, we can fix a sequence (G_i, Γ_i) of nilpotent Lie groups of step $< \kappa$ and cocompact lattices such that, for any nilpotent Lie group G of step $< \kappa$ and for any cocompact lattice Γ there exists a natural number i and a Lie group isomorphism $\psi\colon G_i \to G$ such that $\psi(\Gamma_i) = \Gamma$. By Stone-Weierstass, there exists a countable, uniformly dense subset of the continuous functions on G_i/Γ_i. Fix such a subset and call it $\mathcal{F}_i$. We are assuming the $\kappa - 1$-Fourier uniformity conjecture:

$$\lim_{H\to\infty} \limsup_{N\to\infty} \mathbb{E}^{\log}_{n\leq N} \sup_{g\in G_i} |\mathbb{E}_{h\leq H} a(n+h)F(g^h\Gamma_i)| = 0,$$

for all i and all F a continuous function on G/Γ. By [Fra17, Section 4.5, Step 4] we also get that, for all i and F as before,

$$\lim_{H\to\infty} \limsup_{N\to\infty} \mathbb{E}^{\log}_{n\leq N} \sup_{\substack{g\in G_i \\ x\in G_i/\Gamma_i}} |\mathbb{E}_{h\leq H} a(n+h)F(g^h x)| = 0.$$

Fix a natural number i for the moment and a function F in $\mathcal{F}_i$. For each $\varepsilon > 0$ there exists H_ε such that

$$\limsup_{N\to\infty} \mathbb{E}^{\log}_{n\leq N} \sup_{\substack{g\in G_i \\ x\in G_i/\Gamma_i}} |\mathbb{E}_{h\leq H_\varepsilon} a(n+h)F(g^h x)| \ll \varepsilon^3.$$

Therefore, by Chebyshev's inequality,

$$\mathrm{d}^{\log}\{n \in \mathbb{N}\colon \sup_{\substack{g\in G_i \\ x\in G_i/\Gamma_i}} |\mathbb{E}_{h\leq H_\varepsilon} a(n+h)F(g^h x)| \geq \varepsilon\} \ll \varepsilon^2,$$

where $\mathrm{d}^{\log}$ denotes the upper logarithmic density. Note that

$$\sup_{\substack{g\in G_i \\ x\in G_i/\Gamma_i}} |\mathbb{E}_{h\leq H_\varepsilon} a(n+h)F(g^h x)|$$

depends measurably on $(a(n+1), \ldots, a(n+H_\varepsilon))$. Thus, there exists some set A in $\mathbb{C}^{H_e}$ such that

$$\sup_{\substack{g\in G_i \\ x\in G_i/\Gamma_i}} |\mathbb{E}_{h\leq H_\varepsilon} a(n+h)F(g^h x)| \geq \varepsilon$$

if and only if $(a(n+1),\dots,a(n+H_\varepsilon))$ are in A. Therefore, we know that

$$\mathrm{d}^{\log}\{n\in\mathbb{N}\colon (a(n+1),\dots,a(n+H_\varepsilon))\in A\}\ll\varepsilon^2.$$

By the definition of a dynamical model for a,

$$\mu\{x'\in X\colon (f(T^1x'),\dots,f(T^{H_\varepsilon}x'))\in A\}\ll\varepsilon^2.$$

Unpacking definitions, we get

$$\mu\{x'\in X\colon \sup_{\substack{g\in G_i\\ x\in G_i/\Gamma_i}}|\mathbb{E}_{h\le H_\varepsilon}f(T^hx')F(g^hx)|\ge\varepsilon\}\ll\varepsilon^2.$$

We call this set

$$\{x'\in X\colon \sup_{\substack{g\in G_i\\ x\in G_i/\Gamma_i}}|\mathbb{E}_{h\le H_\varepsilon}f(T^hx')F(g^hx)|\ge\varepsilon\}=S_\varepsilon.$$

Remember that S_ε implicitly depends on i and F. By the definition of the ergodic decomposition, we have that

$$\mu(S_\varepsilon)=\int_\Omega\mu_\omega(S_\varepsilon)d\omega.$$

Therefore, by another application of Chebyshev's inequality, we find that

$$|\{\omega\in\Omega\colon \mu_\omega(S_\varepsilon)\le\varepsilon\}|\ge 1-\varepsilon.$$

We call this set $K_\varepsilon=\{\omega\in\Omega\colon \mu_\omega(S_\varepsilon)\le\varepsilon\}$. Of course K_ε depends on i and F. Define

$$\Omega_{i,F}=\bigcap_{m\in\mathbb{N}}\bigcup_{r\ge m}K_{\frac{1}{r}},$$

and define,

$$\Omega'=\bigcap_{i\in\mathbb{N}}\bigcap_{F\in\mathcal{F}}\Omega_{i,F}.$$

Since $|K_\varepsilon|\ge 1-\varepsilon$, we know that for any m, we have $\left|\bigcup_{r\ge m}K_{\frac{1}{r}}\right|=1$ and therefore $|\Omega'|=1$.

Now we check that Ω' has the desired properties. Thus, fix ω in Ω', ϕ a measurable function on X, G a nilpotent Lie group of step $< \kappa$, Γ a cocompact lattice and F' a function on G/Γ. Suppose that for μ_ω almost every x in X, there exists x' in G/Γ such that $\phi(T^h x) = F'(g^h x')$ for some g in G. Fix $\varepsilon > 0$. We aim to show

$$\left| \int_X f(x)\phi(x)\mu_\omega(dx) \right| \lesssim \varepsilon \cdot (||F'||_{L^\infty} + 1) .$$

Fix i in the natural numbers such that (G, Γ) is isomorphic to (G_i, Γ_i). Fix $\psi \colon G_i \to G$ an isomorphism such that $\psi(\Gamma_i) = \Gamma$. Fix F in $\mathcal{F}_i$ such that $||F \circ \psi - F'||_{L^\infty} \leq \varepsilon$. Then ω is in Ω' so ω is in $\Omega_{i,F}$ and therefore there exists $r > \frac{1}{\varepsilon}$ such that ω is in $K_{\frac{1}{r}}$. Therefore, for some $H = H_{\frac{1}{r}}$,

$$\mu_\omega\{x' \in X \colon \sup_{\substack{g \in G_i \\ x \in G_i/\Gamma_i}} |\mathbb{E}_{h \leq H} f(T^h x') F(g^h x)| \geq \varepsilon\} \leq \varepsilon.$$

By the triangle inequality,

$$\mu_\omega\{x' \in X \colon \sup_{\substack{g \in G \\ x \in G/\Gamma}} |\mathbb{E}_{h \leq H} f(T^h x') F'(g^h x)| \geq 2\varepsilon\} \leq \varepsilon.$$

Next, we use that ϕ locally looks like F':

$$\mu_\omega\{x' \in X \colon |\mathbb{E}_{h \leq H} f(T^h x') \phi(T^h x')| \geq 2\varepsilon\} \leq \varepsilon.$$

Bounding the exceptional points by the L^∞ norm, we get that:

$$\int_X \left| \mathbb{E}_{h \leq H} f(T^h x) \phi(T^h x) \right| \mu_\omega(dx) \lesssim \varepsilon \cdot (||F'||_{L^\infty} + 1) .$$

By the triangle inequality,

$$\left| \int_X \mathbb{E}_{h \leq H} f(T^h x) \phi(T^h x) \mu_\omega(dx) \right| \lesssim \varepsilon \cdot (||F'||_{L^\infty} + 1) .$$

By translation invariance,

$$\left| \int_X f(x)\phi(x)\mu_\omega(dx) \right| \lesssim \varepsilon \cdot (||F'||_{L^\infty} + 1) .$$

This completes the proof. □

Proposition 2.2.5. *Let (X, μ, T) be a (topologically) compact, invertible, not necessarily ergodic dynamical system. Let $\mu = \int_\Omega \mu_\omega d\omega$ be an ergodic decomposition. Recall that, for each ω, the Host-Kra factor $\mathcal{Z}_\omega$ is defined up to sets of μ_ω-measure 0. For each ω, fix such a Host-Kra factor. For instance, one could use any definition of the Host-Kra factor and then add all sets of μ_ω-measure 0 to obtain the complete Host-Kra factor. Then there exists a sigma algebra $\mathcal{Z}$ on X such that, for any measurable set A, A is $\mathcal{Z}$ measurable if and only if there exists a full measure subset $\Omega' \subset \Omega$ such that for all ω in Ω', A is $\mathcal{Z}_\omega$ measurable. This implies that a function f in $L^\infty(\mu)$ is $\mathcal{Z}$ measurable if and only if there exists a full measure subset $\Omega' \subset \Omega$ such that f is $\mathcal{Z}_\omega$ measurable for every ω in Ω'.*

Proof. Let $\mathcal{Z}$ be the set of measurable subsets of X such that there exists a full measure set $\Omega_A \subset \Omega$ such that for all ω in Ω_A, A is $\mathcal{Z}_\omega$ measurable. For each such set, fix such an Ω_A. Let $A_1, A_2, A_3, \ldots$ be a countable list of sets in $\mathcal{Z}$. Consider

$$\Omega' = \bigcap_{i \in \mathbb{N}} \Omega_{A_i}.$$

Because Ω' is the intersection of countably many full measure sets, it has full measure. Let ω be an element of Ω'. Then for every natural number i, A_i is $\mathcal{Z}_\omega$ measurable. Because $\mathcal{Z}_\omega$ is a sigma algebra, that implies the countable intersection and countable union of the sets A_i are also $\mathcal{Z}_\omega$ measurable. Thus the intersection $\cap A_i$ and union $\cup A_i$ are both $\mathcal{Z}_\omega$ measurable for a full measure subset $\Omega' \subset \Omega$ and thus, by definition of $\mathcal{Z}$, $\mathcal{Z}$ is closed under countable unions and intersections. If A is in $\mathcal{Z}$, then A is in $\mathcal{Z}_\omega$ for every ω in Ω_A. Since $\mathcal{Z}_\omega$ is a sigma algebra, the complement A^c is also in $\mathcal{Z}_\omega$ for every ω in Ω_A. By definition of $\mathcal{Z}$, we conclude that $\mathcal{Z}$ is closed under complements. Obviously X and $\emptyset$ are in $\mathcal{Z}$ so $\mathcal{Z}$ is a sigma algebra.

Lastly, we check that a function f in $L^\infty(\mu)$ is $\mathcal{Z}$ measurable if and only if it is $\mathcal{Z}_\omega$ measurable for a full measure set of ω. First, suppose there exists a full measure subset $\Omega' \subset \Omega$ such that, for ω in Ω', f is $\mathcal{Z}_\omega$ measurable. Let A be a measurable subset of $\mathbb{C}$. Then since f is $\mathcal{Z}_\omega$ measurable for any ω in Ω', $f^{-1}(A)$ is in $\mathcal{Z}_\omega$ for any ω in Ω'. Therefore, by definition of $\mathcal{Z}$, $f^{-1}(A)$ is in $\mathcal{Z}$ so f is $\mathcal{Z}$ measurable.

Now suppose f is $\mathcal{Z}$ measurable. We approximate f by simple functions f_i. For instance, we can take $f_i(x) = k \cdot 2^{-i}$ if $f(x)$ is between $k \cdot 2^{-i}$ and $(k+1) \cdot 2^{-i}$ for any natural number k. Then $f_i \to f$ in $L^1(\mu)$ and also in $L^1(\mu_\omega)$ for any ω by the dominated convergence theorem. For each i, the function f_i has only finitely many distinct level sets. Because f is $\mathcal{Z}$ measurable, the level sets of f_i are $\mathcal{Z}$ measurable. Therefore, there exists a full measure subset $\Omega_i \subset \Omega$ such that f_i is $\mathcal{Z}_\omega$ measurable for all ω in Ω_i. Let

$$\Omega' = \bigcap_{i \in \mathbb{N}} \Omega_i.$$

Then since Ω' is the intersection of sets of full measure, Ω' has full measure. For each ω in Ω', f_i is $\mathcal{Z}_\omega$ measurable for all natural numbers i. But $f_i \to f$ in $L^1(\mu_\omega)$ so the limit f is also $\mathcal{Z}_\omega$ measurable for all ω in Ω'. □

Definition 2.2.6. *By Proposition 2.2.5, there exists a sigma algebra $\mathcal{Z}$ such that a $L^\infty(\mu)$ function f is $\mathcal{Z}$ measurable if and only if it is $\mathcal{Z}_\omega$ measurable for almost every ω in Ω. We fix such a sigma algebra and call it the Host-Kra sigma algebra for (X, μ, T).*

Proposition 2.2.7. *Let f be a function in $L^\infty(\mu)$. Then there exists a set Ω' of full measure in Ω such that for ω in Ω',*

$$\mathbb{E}^\mu[f|\mathcal{Z}] = \mathbb{E}^{\mu_\omega}[f|\mathcal{Z}_\omega],$$

μ_ω almost everywhere.

Proof. First, we need the following quick ergodic theoretic fact. The space X can be essentially partitioned into pieces where each piece carries all the mass of an ergodic component. More precisely, there exists a map $\omega' \colon X \to \Omega$ such that

$$\int_\Omega \int_X \phi(x) \mu_\omega(dx) d\omega = \int_\Omega \int_X \phi(x) \mathbb{1}_{\omega = \omega'(x)} \mu_\omega(dx) d\omega$$

for any integrable ϕ. For instance, in the usual construction of an ergodic decomposition, one can take Ω to be the set of atoms of X with respect to the invariant sigma algebra $\mathcal{I}$.

Then let $\int \psi(x')\mu_{[x]}(dx') = \mathbb{E}[\psi|\mathcal{I}](x)$ where x is any point in the atom $[x]$. In this case the map ω' just sends x to the atom containing x.

By Proposition 2.2.5, there is a set Ω_0 of full measure such that $\mathbb{E}^{\mu}[f|\mathcal{Z}]$ is $\mathcal{Z}_\omega$ measurable for every ω in Ω_0. We also ask that for ω in Ω_0 that

$$||f||_{L^\infty(\mu_\omega)} \leq ||f||_{L^\infty(\mu)}$$

which holds for a full measure set of ω. Fix such an Ω_0. Since X is compact, there exists a countable uniformly dense subset of the space of continuous functions. Fix such a subset and fix an order on that subset $f_1, f_2, f_3, f_4, \ldots$. Again by Proposition 2.2.5, there exists a full measure subset Ω_i of Ω such that for ω in Ω_i, the function $\mathbb{E}^{\mu}[f_i|\mathcal{Z}]$ is $\mathcal{Z}_\omega$ measurable. Let

$$\Omega' = \bigcap_{i \geq 0} \Omega_i.$$

Since each Ω_i has full measure and there are only countably many choices of i, we conclude that Ω' has full measure. Now let ω be an element of Ω' and suppose for the sake of contradiction that

$$\mathbb{E}^{\mu}[f|\mathcal{Z}] \neq \mathbb{E}^{\mu_\omega}[f|\mathcal{Z}_\omega],$$

meaning equality does not hold up to sets of μ_ω measure 0. The conditional expectation is uniquely defined by two properties, namely that $\mathbb{E}^{\mu_\omega}[f|\mathcal{Z}_\omega]$ is $\mathcal{Z}_\omega$ measurable and that

$$\int_X \mathbb{E}^{\mu_\omega}[f|\mathcal{Z}_\omega](x)\phi(x)\mu_\omega(dx) = \int_X f(x)\phi(x)\mu_\omega(dx),$$

for any $\mathcal{Z}_\omega$ measurable function ϕ in $L^\infty(\mu_\omega)$. If $\mathbb{E}^{\mu}[f|\mathcal{Z}]$ satisfies the same properties then $\mathbb{E}^{\mu}[f|\mathcal{Z}]$ equals $\mathbb{E}^{\mu_\omega}[f|\mathcal{Z}_\omega]$ μ_ω-almost everywhere. We know since ω is in Ω' which is contained in Ω_0 that $\mathbb{E}^{\mu}[f|\mathcal{Z}]$ is $\mathcal{Z}_\omega$ measurable. Therefore, there exists ϕ in $L^\infty(\mu_\omega)$ which is $\mathcal{Z}_\omega$ measurable such that

$$\int_X \mathbb{E}^{\mu}[f|\mathcal{Z}](x)\phi(x)\mu_\omega(dx) \neq \int_X f(x)\phi(x)\mu_\omega(dx).$$

By subtracting off the appropriate multiple of $\mathbb{E}^{\mu}[f|\mathcal{Z}]$, we may assume that ϕ is μ_ω-orthogonal to $\mathbb{E}^{\mu}[f|\mathcal{Z}]$. Multiplying by a scalar we may assume that $\langle f, \phi\rangle_{L^2(\mu_\omega)}$ is a positive real number greater than 1.

For each ω in Ω' such that $\mathbb{E}^{\mu}[f|\mathcal{Z}] \neq \mathbb{E}^{\mu_\omega}[f|\mathcal{Z}_\omega]$, we showed there exists ϕ a $\mathcal{Z}_\omega$ measurable function such that $\langle \mathbb{E}^{\mu}[f|\mathcal{Z}], \phi\rangle_{L^2(\mu_\omega)} = 0$ and $\langle \phi, f\rangle_{L^2(\mu_\omega)} > 1$. Let ϕ be such a function. Suppose for the moment that $||\phi||_{L^\infty(\mu_\omega)} < C$. Since $f_1, f_2, f_3, \ldots$ are dense in $L^2(\mu_\omega)$, for any ε and for any power $p < \infty$ we can find an i such that $||\phi - f_i||_{L^p(\mu_\omega)} \leq \varepsilon$. This implies, by Cauchy-Schwarz, that

$$\langle \mathbb{E}^{\mu}[f|\mathcal{Z}], f_i\rangle_{L^2(\mu_\omega)} \leq \varepsilon ||f||_{L^\infty(\mu)} \tag{2.2}$$

and

$$\langle \phi, f_i\rangle_{L^2(\mu_\omega)} > 1 - \varepsilon. \tag{2.3}$$

We also need a quantitative way of saying that f_i is close to being $\mathcal{Z}_\omega$ measurable. One option is to use the Host-Kra norms defined for an ergodic system in [HK05] section 3.5. Let $|||\phi|||_{k,\omega}$ denote the k^{th} Host-Kra norm. The key feature of the Host-Kra norms is that a function ϕ is $\mathcal{Z}_\omega$ measurable if and only if $|||\phi|||_{k,\omega} = 0$ for all k (see [HK05] Lemma 4.3). We claim that $|||\phi|||_{k,\omega}$ is a measurable function of ω. After all, by definition $|||\phi|||_{k,\omega}^{2^k}$ is the integral of some fixed function on X^{2^k}, namely $(x_1, \ldots, x_{2^k}) \mapsto \phi(x_1) \ldots \phi(x_{2^k})$ with respect to some measure (namely $\mu_\omega^{[k]}$ defined in [HK05] section 3.1) which depends measurably on ω. Thus, we can find also f_i with

$$|||f_i|||_{k,\omega} \leq 2\varepsilon, \tag{2.4}$$

for all $k \leq \frac{1}{\varepsilon}$. Since we also know that

$$||\phi||_{L^p(\mu_\omega)} < C$$

for some constant C then also by the triangle inequality,

$$||f_i||_{L^p(\mu_\omega)} \leq C + \varepsilon. \tag{2.5}$$

Fix a constant C. Now we define a function $i\colon \Omega \times \mathbb{R}_{>0} \to \mathbb{N} \cup \{\infty\}$ as follows: Let $i(\omega, \varepsilon)$ be the first index such that all four inequalities 2.2 - 2.5 are satisfied with $p = \frac{1}{\varepsilon}$ if such an i exists and $+\infty$ otherwise. Note that i implicitly depends on C. Let E denote the set of ω such that $i(\omega, \varepsilon)$ is finite for all ε. In particular, if $\mathbb{E}^{\mu}[f|\mathcal{Z}] \neq \mathbb{E}^{\mu_\omega}[f|\mathcal{Z}_\omega]$ then ω is in this set for some choice of C. Thus, we may assume for the sake of contradiction that the measure of E is positive. Let

$$\psi_\varepsilon(x) = \begin{cases} f_{i(\omega,\varepsilon)}(x) & \omega \in E, \omega'(x) = \omega \\ 0 & \text{otherwise} \end{cases}$$

Since, for all $1 < p < \frac{1}{\varepsilon}$

$$||\psi_\varepsilon||_{L^p(\mu)} = \int_\Omega ||\psi_\varepsilon||_{L^p(\mu_\omega)} d\omega \leq C + \varepsilon,$$

we can take an $L^p(\mu)$ weak-$*$ limit (which is the same as a weak limit in this case) $\psi_\varepsilon \to \psi$ for some subsequence of epsilons tending to 0. By a diagonalization argument, we can ensure that this weak-$*$ limit exists for all $1 < p < \infty$. By 2.4, we conclude that ψ is $\mathcal{Z}_\omega$ measurable for each ω in E. If ω is not in E, then $\psi = 0$ on a set of μ_ω full measure so ψ is measurable with respect to $\mathcal{Z}_\omega$ for a full measure set of ω in Ω so by definition ψ is $\mathcal{Z}$ measurable. Futhermore, by 2.3

$$\langle \phi, \psi \rangle_{L^2(\mu_\omega)} \geq 1$$

so we conclude that

$$\langle \phi, \psi \rangle_{L^2(\mu)} \geq |E|$$

by integrating in ω. On the other hand, by 2.2

$$\langle \mathbb{E}^{\mu}[f|\mathcal{Z}], \psi \rangle_{L^2(\mu)} = 0.$$

This contradicts the definition of $\mathbb{E}^{\mu}[f|\mathcal{Z}]$. Thus,

$$\mathbb{E}^{\mu}[f|\mathcal{Z}] = \mathbb{E}^{\mu_\omega}[f|\mathcal{Z}_\omega]$$

for almost every ω in Ω. □

A crucial input is the following theorem of [FH18a]. This theorem says that if a correlates with b then it does so for some algebraic reason. In particular, any correlation between f and f' is due solely to some locally algebraic structure in f.

Theorem 2.2.8 ([FH18a] Theorem 1.5; see also section 2.4)**.** *Let μ be the first marginal of ν corresponding to the factor X. Then the ergodic components (X, μ_ω, T) of μ are isomorphic to the product of a Bernoulli system with the Host-Kra factor of (X, μ_ω, T).*

To use this theorem, we need the following result, which essentially appears in [FH18b]:

Lemma 2.2.9 ([FH18b]; see the proof of Lemma 6.2)**.** *Suppose that $(X, \mu_\omega, T) \cong (W, dw, T) \times (Z, dz, T)$ where W is a Bernoulli system, Z is a zero entropy system and μ_ω is the first marginal of ν_ω. Then for any function $\phi\colon X \to \mathbb{C}$ and any function $\psi\colon Y \to \mathbb{C}$ we have*

$$\int_{X\times Y} \phi(x)\psi(y)\nu_\omega(dxdy) = \int_{X\times Y} \mathbb{E}^{\nu_\omega}[\phi|Z](x)\psi(y)\nu_\omega(dxdy)$$

where $\mathbb{E}^{\nu_\omega}[\phi|Z]$ denotes the conditional expectation of ϕ with respect to the measure ν_ω and the sigma algebra of Z-measurable functions.

Proof. By density, it suffices to consider the case $\phi(w, z) = \phi_W(w)\phi_Z(z)$. Because any joining of the Bernoulli system W and the zero entropy system $Z \times Y$ is trivial i.e. is equipped with the product measure, we can break up the the integral

$$\begin{aligned}\int_{W\times Z\times Y} \phi_W(w)\phi_Z(z)\psi(y)\nu_\omega(dwdzdy) &= \int_W \phi_W(w)\nu_\omega(dwdzdy) \cdot \int_{Z\times Y} \phi_Z(z)\psi(y)\nu_\omega(dwdzdy) \\ &= \int_{X\times Y} \mathbb{E}^{\nu_\omega}[\phi|Z](z)\psi(y)\nu_\omega(dwdzdy).\end{aligned}$$

□

We also need the following result, which says that conditional expectation is essentially local.

Corollary 2.2.10. *Let X, Y, ν, f, f', c and η be as in Subsection 2.1.1. Let B be as in Corollary 2.2.2. Then*

$$\left|\int_{B^c} \mathbb{E}_{h\leq k} T^h(\mathbb{E}^\nu[f|\mathcal{Z}]\cdot f')d\nu\right| > c+\eta.$$

Proof. Recall that

$$\left|\int_{X\times Y} f(x)\cdot f'(y)\nu(dxdy)\right| > c+\eta.$$

By Corollary 2.2.2, we have that

$$\left|\int_{B^c} f(x)\cdot f'(y)\nu(dxdy)\right| > c+\eta.$$

Since B^c is T invariant and ν is T invariant, we can average over shifts

$$\left|\int_{B^c} \mathbb{E}_{h\leq k} f(T^h x)\cdot f'(T^h y)\nu(dxdy)\right| > c+\eta.$$

Next, we disintegrate the measure ν,

$$\left|\int_\Omega\int_{B^c} \mathbb{E}_{h\leq k} f(T^h x)\cdot f'(T^h y)\nu_\omega(dxdy)d\omega\right| > c+\eta.$$

Notice, for each h, $f'(T^h y)\mathbb{1}_{y\notin B}$ is a function on Y. By Theorem 2.2.8, (X,μ_ω,T) is isomorphic to a product of a Bernoulli factor with the Host-Kra factor for almost every ω. Since the Host-Kra factor has entropy zero, the hypotheses of Lemma 2.2.9 are satsified. Thus, by Lemma 2.2.9,

$$\left|\int_\Omega\int_{B^c} \mathbb{E}_{h\leq k}\mathbb{E}^{\mu_\omega}[f|\mathcal{Z}_\omega](T^h x)\cdot f'(T^h y)\nu_\omega(dxdy)d\omega\right| > c+\eta.$$

By Proposition 2.2.7,

$$\left|\int_\Omega\int_{B^c} \mathbb{E}_{h\leq k}\mathbb{E}^{\mu}[f|\mathcal{Z}](T^h x)\cdot f'(T^h y)\nu_\omega(dxdy)d\omega\right| > c+\eta.$$

By definition of the ergodic decomposition,

$$\left|\int_{B^c} \mathbb{E}_{h\leq k}\mathbb{E}^{\mu}[f|\mathcal{Z}](T^h x)\cdot f'(T^h y)\nu(dxdy)\right| > c+\eta.$$

This completes the proof. □

Now we forget everything about the joining of X and Y and reduce to the worst case scenario, where we choose the worst possible y in Y for each x in X.

Corollary 2.2.11. *Let X, Y, ν, μ, f, f', c and η be as in Subsection 2.1.1. Let $\mathcal{Z}$ be as in Definition 2.2.6. Let B be as in Corollary 2.2.2. Since whether $(x, y) \in B$ only depends on y, we abuse notation and write $y \in B$ to mean $(x, y) \in B$ for some x. Then,*

$$\int_X \sup_{y \notin B} \left| \mathbb{E}_{h \leq k} \mathbb{E}^{\mu}[f|\mathcal{Z}](T^h x) \cdot f'(T^h y) \right| \mu(dx) > c + \eta,$$

where the supremum is an essential supremum taken with respect to the second marginal of ν.

We will need the following lemma, which states that conditioning with respect to a conditional measure is essentially the same as conditioning with respect to the original measure.

Lemma 2.2.12. *Let A be a positive measure set in $\mathcal{Z}$ and denote $\mu_A(S) = \mu(S|A)$. Then for any measurable function f,*

$$\mathbb{E}^{\mu_A}[f|\mathcal{Z}] = \mathbb{E}^{\mu}[f|\mathcal{Z}],$$

μ_A almost everywhere i.e. for μ-almost every point in A.

Proof. Let C be another set in $\mathcal{Z}$. Then

$$\int_C \mathbb{E}^{\mu}[f|\mathcal{Z}](x)\mu_A(dx) = \int_{C \cap A} \frac{1}{\mu(A)} \mathbb{E}^{\mu}[f|\mathcal{Z}](x)\mu(dx)$$

Since A is in $\mathcal{Z}$, we know that $A \cap C$ is in $\mathcal{Z}$. By definition of conditional expectation, this is

$$\begin{aligned} &= \frac{1}{\mu(A)} \int_{C \cap A} f\mu(dx) \\ &= \int_{C \cap A} f\mu_A(dx). \end{aligned}$$

This is the defining property of $\mathbb{E}^{\mu_A}[f|\mathcal{Z}]$. Since conditional expectation is well defined up to sets of measure 0, we obtain the result. □

The system X possesses an extra symmetry that most dynamical systems do not have, a dilation symmetry. In fact, it possesses a whole family of dilation symmetries. It is not obvious which dilation makes the problem easiest. Therefore, instead of choosing a particular dilation, we use a random dilation.

Proposition 2.2.13. *Let P be any natural number. Then*

$$\mathbb{E}_{P/2<p\leq P}\int_X p\mathbb{1}_{M^{-1}(p\hat{\mathbb{Z}})}(x)\sup_{y\notin B}\Big|\mathbb{E}_{h\leq k}\mathbb{E}^{\mu}[f|\mathcal{Z}](T^{ph}x)\cdot f'(T^h y)\Big|\mu(dx)>c+\eta,$$

where p is always restricted to be prime.

Proof. By Corollary 2.2.11 we have

$$\int_X \sup_{y\notin B}\Big|\mathbb{E}_{h\leq k}\mathbb{E}^{\mu}[f|\mathcal{Z}](T^h x)\cdot f'(T^h y)\Big|\mu(dx)>c+\eta.$$

Now we use that I_p pushes forward $p\mathbb{1}_{M^{-1}(p\hat{\mathbb{Z}})}\mu$ to μ for every p and average in p.

$$\mathbb{E}_{P/2<p\leq P}\int_X p\mathbb{1}_{M^{-1}(p\hat{\mathbb{Z}})}\sup_{y\notin B}\Big|\mathbb{E}_{h\leq k}\mathbb{E}^{\mu}[f|\mathcal{Z}](T^h I_p x)\cdot f'(T^h y)\Big|\mu(dx)>c+\eta.$$

Because $I_p\circ T^{hp}(x)=T^h\circ I_p(x)$ for almost every x in $M^{-1}(p\hat{\mathbb{Z}})$ we have that,

$$\mathbb{E}_{P/2<p\leq P}\int_X p\mathbb{1}_{M^{-1}(p\hat{\mathbb{Z}})}\sup_{y\notin B}\Big|\mathbb{E}_{h\leq k}\mathbb{E}^{\mu}[f|\mathcal{Z}](I_p T^{ph}x)\cdot f'(T^h y)\Big|\mu(dx)>c+\eta.$$

Next we use the standard fact that

$$\mathbb{E}^{\mu}[f|\mathcal{Z}]\circ I_p=\mathbb{E}^{I_{p*}\mu}[f\circ I_p|I_p^{-1}(\mathcal{Z})],$$

$I_{p*}\mu$-almost everywhere, where $I_{p*}\mu$ is the pushforward of μ. Since $I_{p*}\mu=\frac{1}{p}\mu$, we can replace $I_{p*}\mu$ by μ. Note that I_p defines a factor map between $(M^{-1}(p\widehat{\mathbb{Z}}),p\mu,T^p)$ and (X,μ,T). Since Host-Kra factors are functorial, the Host-Kra factor for $(M^{-1}(p\widehat{\mathbb{Z}},p\mu,T^p))$ factors onto the Host-Kra factor for (X,μ,T). Thus, $I_p^{-1}(\mathcal{Z})$ is contained in the Host-Kra factor of some dynamical system and thus corresponds to an inverse limit of nilsystems. This is all we actually need for our purposes. However, for the sake of avoiding notation, we also prove that

$$\mathbb{1}_{M^{-1}(p\widehat{\mathbb{Z}})}\mathbb{E}^{\mu}[f\circ I_p|I_p^{-1}(\mathcal{Z})]=\mathbb{1}_{M^{-1}(p\widehat{\mathbb{Z}})}\overline{a(p)}\mathbb{E}^{\mu}[f|\mathcal{Z}].$$

That $f \circ I_p = \overline{a(p)} f$ follows from the definition of I_p. If $\mathcal{Z}_i(T^p)$ denotes the i^{th} Host-Kra factor for T^p and $\mathcal{Z}_i(T)$ denotes the i^{th} Host-Kra factor for T, then any T^p invariant subset of the cube X^{2^i} is an element of the Konecker factor i.e. the first Host-Kra factor for $(X^{2^i}, T, \mu^{[i]})$ (where $\mu^{[i]}$ is the measure on the cube defined in section 3 of [HK05]). Since the Host-Kra factor of an ergodic system is the smallest sigma algebra generating the invariant factor on the cube, we conclude that $\mathcal{Z}_i(T^p) \subset \mathcal{Z}_{i+1}(T)$ so $I_p^{-1}(\mathcal{Z}) \subset \mathcal{Z} \cap M^{-1}(p\widehat{\mathbb{Z}})$. In fact, as in section 2.4, the Host-Kra factor for X is a joining of the Host-Kra factor on the space of sequences $D^{\mathbb{Z}}$ and $\widehat{\mathbb{Z}}$. On the second factor, I_p acts by division by p. On the first factor, $I_p \circ T^p = T \circ I_p$ and so on each ergodic component of the first factor, I_p acts by multiplication by p up to a possible translation. Multiplication by p is a local isomorphism of any nilmanifold that does not contain p torsion. However, by Corollary 2.2.3, f is already orthogonal to all p torsion. Thus,

$$\mathbb{1}_{M^{-1}(p\widehat{\mathbb{Z}})} \mathbb{E}^{\mu}[f \circ I_p | I_p^{-1}(\mathcal{Z})] = \mathbb{1}_{M^{-1}(p\widehat{\mathbb{Z}})} \overline{a(p)} \mathbb{E}^{\mu}[f|\mathcal{Z}].$$

Combined with Lemma 2.2.12 and the fact that $M^{-1}(p\hat{\mathbb{Z}})$ is T^p invariant and therefore an element of $\mathcal{Z}$ we get,

$$\mathbb{E}_{P/2<p\leq P} \int_X p \mathbb{1}_{M^{-1}(p\hat{\mathbb{Z}})} \sup_{y \notin B} \left| \mathbb{E}_{h \leq k} \mathbb{E}^{\mu} \overline{a(p)} [f|\mathcal{Z}](T^{ph}x) \cdot f'(T^h y) \right| \mu(dx) > c + \eta.$$

Recall that $|a(p)| = 1$ for all p. Thus, $\overline{a(p)}$ merely gets absorbed into the absolute value.

$$\mathbb{E}_{P/2<p\leq P} \int_X p \mathbb{1}_{M^{-1}(p\hat{\mathbb{Z}})} \sup_{y \notin B} \left| \mathbb{E}_{h \leq k} \mathbb{E}^{\mu} [f|\mathcal{Z}](T^{ph}x) \cdot f'(T^h y) \right| \mu(dx) > c + \eta.$$

□

2.2.1 The Entropy Decrement Argument

Next, we use the entropy decrement method to replace $p\mathbb{1}_{M^{-1}(\hat{\mathbb{Z}})}$ by its average, 1. This is essentially due to Tao but because our statement is slightly different we reproduce the argument. For the definitions of entropy, conditional entropy, mutual information and conditional mutual information see [Tao16b].

Let x' be a random variable distributed according to μ and fix a natural number P. From this, we get the following two random variables. Set $X_P = (x_1, \dots, x_{(k+1)P})$ where $x_i = \mathbb{E}^\mu[f|\mathcal{Z}](T^i x')$ and set Y_P in $\prod_{P/2<p\leq P} \mathbb{Z}/p\mathbb{Z}$ by $Y_P = (M(x') \bmod p)_{P/2<p\leq P}$. Denote $Y_P \bmod p = y_p$ so that $Y_P = (y_p)_{P/2<p\leq P}$. Note that Y_P is uniformly distributed in $\prod_{P/2<p\leq P} \mathbb{Z}/p\mathbb{Z}$ and that the distribution of X_P is the same as the distribution of $T^i X_P$ for any i because μ is translation invariant. Technically, if $\mathbb{E}^\mu[f|\mathcal{Z}]$ takes infinitely many values then we will have to round $\mathbb{E}^\mu[f|\mathcal{Z}](T^i x')$ so that each x_i takes values in a finite set but this slightly annoying detail may be delayed for the moment. We want to study the following integral:

$$\mathbb{E}_{P/2<p\leq P} \int_X p\mathbb{1}_{M^{-1}(p\hat{\mathbb{Z}})} \sup_{y\notin B} \left|\mathbb{E}_{h\leq k}\mathbb{E}^\mu[f|\mathcal{Z}](T^{ph}x)\cdot f'(T^h y)\right|\mu(dx)$$

By translation invariance, this is equal to

$$\mathbb{E}_{P/2<p\leq P}\mathbb{E}_{i\leq P} \int_X pT^i\mathbb{1}_{M^{-1}(p\hat{\mathbb{Z}})} \sup_{y\notin B} \left|\mathbb{E}_{h\leq k}\mathbb{E}^\mu[f|\mathcal{Z}](T^{ph+i}x)\cdot f'(T^h y)\right|\mu(dx).$$

Notice that this is the expected value of some function of X_P and Y_P. In particular, we are interested in

$$\mathbb{E}\left(\mathbb{E}_{P/2<p\leq P} f_{P,p}(X_P, y_p)\right)$$

where $f_{P,p}\colon \mathbb{C}^{(k+1)P} \times \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ is defined by the formula

$$f_{P,p}(X_P, y_p) = \mathbb{E}_{i\leq P}\ p\mathbb{1}_{y_p=i} \sup_{y\notin B} \left|\mathbb{E}_{h\leq k} x_{hp+i} f(Ty)\right|.$$

Define

$$f_P(X_P, Y_P) = \mathbb{E}_{P/2<p\leq P} f_{P,p}(X_P y_P).$$

Thus, we are interested in

$$\mathbb{E} f_P(X_P, Y_P).$$

We would like to say that X_P and Y_P are very close to independent for some large choice of P. Let W_P be a random variable with the same distribution as Y_P but which is independent of X_P. We would like to say that

$$\mathbb{E}[f_P(X_P, Y_P)] \approx \mathbb{E}[f_P(X_P, W_P)].$$

A property like this actually holds in a more general setting, which we take the liberty of stating now.

Theorem 2.2.14 ([Tao16b] Section 3; see also [Mor18],[TT17b] Lemma 3.4 and Proposition 3.5 and [TT17a] Section 4)**.** *Let A be a finite set and let C be a natural number. For each power of two P, let $X_P = (x_1, \dots, x_{CP})$ be a sequence of random variables with x_i taking values in A and let Y_P be a random variable that is uniformly distributed in $\prod_{P/2<p\leq P} \mathbb{Z}/p\mathbb{Z}$. We write $Y_P = (y_p)_{P/2<p\leq P}$ where $y_p = Y_P \bmod p$. We further assume that for different values of P, the random variables Y_P are jointly independent meaning $(y_p)_{p\leq P}$ is uniformly distributed in $\prod_{p\leq P} \mathbb{Z}/p\mathbb{Z}$ for all powers of two P. Suppose that, for any natural numbers i and m such that $i + m \leq CP$ we have that the distribution of $(x_1, \dots, x_m)$ is equal to the distribution of $(x_{i+1}, \dots, x_{i+m})$. Furthermore, suppose that for any P and any element b in $\prod_{p\leq P} \mathbb{Z}/p\mathbb{Z}$ and any S a measurable subset of $\mathbb{C}^m$,*

$$\mathbb{P}((x_1, \dots, x_m) \in S \mid (y_p)_{p\leq P} = b) = \mathbb{P}((x_{i+1}, \dots, x_{i+m}) \in S \mid (y_p)_{p\leq P} = b + i).$$

For each p with $P/2 < p \leq P$, let $f_{P,p}$ be a 1-bounded function $A^{CP} \times \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ and let $f_P(X_P, Y_P) = \mathbb{E}_{P/2<p\leq P} f_{P,p}(X_P, y_p)$. Let W_P be a random variable with the same distribution as Y_P but which is independent of X_P. Then

$$\liminf_{P\to\infty} \mathbb{E}[|f_P(X_P, Y_P) - f_P(X_P, W_P)|] = 0.$$

Proof. Fix a large power of two P and $\varepsilon > 0$. By replacing $f_{P,p}(a, b)$ by $f_{P,p}(a, b) - f_{P,p}(a, W_P)$ we may assume that $f_{P,p}(a, W_P) = 0$ for all a. To prove the theorem, first we need a very good understanding of the case when X_P and Y_P are independent. In that case, even if we know the exact value of X_P, f_P is still a sum of independent random variables $f_{P,p}(a, W_P)$ and therefore exhibits concentration. This is formalized in Hoeffding's inequality, which says that large collections of independent random variables exhibit concentration.

Lemma 2.2.15 (Hoeffding's Inequality)**.** *Suppose $Z_1, \dots, Z_n$ are independent random vari-*

ables taking values in $[-2,2]$. *Then*

$$\mathbb{P}(|Z_1+\cdots+Z_n-\mathbb{E}[Z_1+\cdots+Z_n]|>t)\leq \exp(-nt^2/16).$$

Let a be an element of A^{CP}. We apply Hoeffding's inequality to the random variables $f_{P,p}(a,Y_P)$. (We remind the reader that there are roughly $\frac{P}{2\log P}$ many such terms, by the prime number theorem).

$$\mathbb{P}(|f_P(a,W_P)|>\varepsilon)\leq \exp(-\frac{\varepsilon^2 P}{40\log P}). \tag{2.6}$$

Next, we aim to show that if Y_P is not necessarily independent of X_P but nearly independent of Y_P, we still can obtain a good bound. To do this, we use a Pinsker-type inequality.

Lemma 2.2.16. *[*[TT17a] *Lemma 3.4] Let Y be a random variable taking values in a finite set, let W be a uniformly distributed random variable on the same set and let E be a set. Then*

$$\mathbb{P}(Y\in E)\leq -\frac{H(W)-H(Y)+\log 2}{\log \mathbb{P}(W\in E)}.$$

Let a be an element of A^{CP}. Let E be the set of b in $\prod_{P/2<p\leq P}\mathbb{Z}/p\mathbb{Z}$ such that $|f_P(a,b)|>\varepsilon$. By 2.6, we know

$$\mathbb{P}(W_P\in E)\leq \exp(-\frac{\varepsilon^2 P}{40\log P}).$$

Applying Lemma 2.2.16 to $\mathbb{P}(\,\cdot\,|X_P=a)$, we find

$$\mathbb{P}(|f(a,Y_P)|>\varepsilon|X_P=a)\leq -\frac{(H(W_P)-H(Y_P|X_P=a)+\log 2)40\log P}{\varepsilon^2 P}.$$

Note that

$$\sum_a \mathbb{P}(X_P=a)(H(W_P)-H(Y_P|X_P=a))=H(W_P)-H(Y_P|X_P)=I(X_P,Y_P),$$

where the last equality follows since $H(W_P)=H(Y_P)$ since the two random variables have the same distribution. Therefore, summing over a, we get

$$\mathbb{P}(|f(X_P,Y_P)|>\varepsilon)\leq -\frac{(I(X_P,Y_P)+\log 2)40\log P}{\varepsilon^2 P}.$$

If

$$I(X_P, Y_P) \lesssim \frac{\varepsilon P}{\log P}$$

then

$$\mathbb{P}(|f(X_P, Y_P)| > \varepsilon) \lesssim \varepsilon.$$

This would complete the proof. Let $Y_{\leq P/2} = (y_p)_{p \leq P/2}$. Fix b' an element of $\prod_{p \leq P/2} \mathbb{Z}/p\mathbb{Z}$. Then we may repeat the previous argument with $\mathbb{P}(\,\cdot\,|Y_{\leq P/2} = b')$ to conclude:

$$\mathbb{P}(|f(X_P, Y_P)| > \varepsilon) \leq -\frac{(I(X_P, Y_P|Y_{\leq P/2}) + \log 2)40 \log P}{\varepsilon^2 P}.$$

and therefore if

$$I(X_P, Y_P|Y_{\leq P/2}) \lesssim \frac{\varepsilon^3 P}{\log P}$$

then

$$\mathbb{P}(|f(X_P, Y_P)| > \varepsilon) \lesssim \varepsilon$$

and therefore

$$\mathbb{E}|f(X_P, Y_P)| \lesssim \varepsilon.$$

Let P_0 be a power of two. We will try to show that there exists $P \geq P_0$ such that

$$\mathbb{E}|f(X_P, Y_P)| \lesssim \varepsilon.$$

This would complete the proof. Suppose not. Then

$$I(X_P, Y_P|Y_{\leq P/2}) \gg \frac{\varepsilon^3 P}{\log P},$$

for all $P \geq P_0$. By definition of mutual information,

$$H(X_P|Y_{\leq P}) = H(X_P|Y_{\leq P/2}) - I(X_P, Y_P|Y_{\leq P/2})$$

where $Y_{\leq P} = (y_p)_{p \leq P}$. By assumption, we have a lower bound for the mutual information

$$\leq H(X_P|Y_{\leq P/2}) - \frac{\varepsilon^3 P}{\log P}$$

By subadditivity of entropy,

$$\leq H(X_{P/2}|Y_{\leq P/2}) + H(x_{CP/2+1}, \ldots, x_{CP}|Y_{\leq P/2}) - \frac{\varepsilon^3 P}{\log P} \tag{2.7}$$

where $X_{P/2} = (x_1, \ldots, x_{CP/2})$. Since $(x_1, \ldots, x_{CP/2})$ has the same distribution as $(x_1, \ldots, x_{CP/2})$, for any set S in $\mathbb{C}^{CP/2}$ and for any b' in $\prod_{p \leq P/2} \mathbb{Z}/p\mathbb{Z}$

$$\mathbb{P}((x_1, \ldots, x_{CP/2}) \in S|Y_{\leq P/2} = b') = \mathbb{P}((x_{CP/2+1}, \ldots, x_{CP}) \in S|Y_{\leq P/2} = b' + CP/2).$$

Since the entropy of a random variable only depends on its distribution, we conclude that, for all b'

$$H(x_1, \ldots, x_{CP/2}|Y_{\leq P/2} = b') = H(x_{CP/2+1}, \ldots, x_{CP}|Y_{\leq P/2} = b' + CP/2)$$

Since $Y_{\leq P/2}$ is uniformly distributed, for all b',

$$\mathbb{P}(Y_{P/2} = b') = \mathbb{P}(Y_{P/2} = b' + CP/2).$$

Therefore, summing in b',

$$\begin{aligned} & H(x_1, \ldots, x_{CP/2}|Y_{\leq P/2}) \\ =& \sum_{b'} \mathbb{P}(Y_{P/2} = b') H(x_1, \ldots, x_{CP/2}|Y_{\leq P/2} = b') \\ =& \sum_{b'} \mathbb{P}(Y_{P/2} = b' + CP/2) H(x_{CP/2+1}, \ldots, x_{CP}|Y_{\leq P/2} = b' + CP/2) \\ =& H(x_{CP/2+1}, \ldots, x_{CP}|Y_{\leq P/2}). \end{aligned} \tag{2.8}$$

Applying 2.8 to 2.7,

$$H(X_P|Y_{\leq P}) \leq 2H(X_{P/2}|Y_{\leq P/2}) - \frac{\varepsilon^3 P}{\log P}.$$

We just obtained an upper bound for $H(X_P|Y_{\leq P})$. We can apply the same argument to obtain an upper bound for $H(X_{P/2}|Y_{\leq P/2})$.

$$\leq 4H(X_{P/4}|Y_{\leq P/4}) - 2\frac{\varepsilon P/2}{\log P/2} - \frac{\varepsilon^3 P}{\log P}$$

where $X_{P/2} = (x_1, \ldots, x_{CP/4})$ and where $Y_{\leq P/4} = (y_p)_{p \leq P/4}$. Applying this argument inductively, if $P = 2^m \cdot P_0$ then

$$\leq 2^m \left(H(X_{P_0}|Y_{\leq P_0}) - \varepsilon \sum_{j=1}^{m} \frac{P_0}{j} \right). \tag{2.9}$$

However, $\sum_{m \leq \log_2 P/P_0} \frac{1}{m} \sim \log\log P$ so for large P

$$\varepsilon \sum_{j=1}^{m} \frac{P_0}{j} \gg CP_0 \log|A| \geq H(X_{P_0}|Y_{\leq P_0}).$$

Combining this with 2.7,

$$H(X_P, Y_{\leq P}) < 0$$

which is impossible.

□

Applying the Theorem 2.2.14 to our situation yields,

Corollary 2.2.17. *Let X, Y, μ, f, f', M, I_p, c and η be as in Subsection 2.1.1. Let $\mathcal{Z}$ be as in Definition 2.2.6. Let B be as in Corollary 2.2.2. We have*

$$\limsup_{P \to \infty} \mathbb{E}_{P/2 < p \leq P} \int_X \sup_{y \notin B} \left| \mathbb{E}_{h \leq k} \mathbb{E}^{\mu}[f|\mathcal{Z}](T^{ph}x) \cdot f'(T^h y) \right| \mu(dx) > c.$$

Proof. Recall that, for all natural numbers P,

$$\mathbb{E}_{P/2 < p \leq P} \int_X p \mathbb{1}_{M^{-1}(p\hat{\mathbb{Z}})} \sup_{y \notin B} \left| \mathbb{E}_{h \leq k} \mathbb{E}^{\mu}[f|\mathcal{Z}](T^{ph}x) \cdot f'(T^h y) \right| \mu(dx) > c + \eta.$$

By translation invariance, for all natural numbers P,

$$\mathbb{E}_{P/2 < p \leq P} \mathbb{E}_{i \leq P} \int_X p \mathbb{1}_{M^{-1}(p\hat{\mathbb{Z}} + i)} \sup_{y \notin B} \left| \mathbb{E}_{h \leq k} \mathbb{E}^{\mu}[f|\mathcal{Z}](T^{ph+i}x) \cdot f'(T^h y) \right| \mu(dx) > c + \eta.$$

Let x' be a random variable with distribution μ. Fix $\varepsilon > 0$ small. We will ask that $\varepsilon < 10 \cdot \eta$. Let ϕ be a measurable function on X which uniformly approximates $\mathbb{E}^{\mu}[f|\mathcal{Z}]$ i.e.

$$||\phi - \mathbb{E}^{\mu}[f|\mathcal{Z}]||_{L^\infty} < \varepsilon.$$

For instance, $\phi(x)$ could be obtained by rounding $\mathbb{E}^{\mu}[f|\mathcal{Z}](x)$ to the closest element of $\frac{\varepsilon}{10}\cdot\mathbb{Z}[i]$. By the triangle inequality

$$\mathbb{E}_{P/2<p\leq P}\mathbb{E}_{i\leq P}\int_X p\mathbb{1}_{M^{-1}(p\hat{\mathbb{Z}}+i)}\sup_{y\notin B}\left|\mathbb{E}_{h\leq k}\mathbb{E}^{\mu}\phi(T^{ph+i}x)\cdot f'(T^h y)\right|\mu(dx) > c+\eta-\varepsilon.$$

For each natural number P, let $X_P = (x_1,\ldots,x_{CP})$ where $x_i = \phi(T^i x')$ and where $C = k+1$. Let $Y_P = (y_p)_{P/2<p\leq P}$ where $y_p = M(x') \bmod p$. For each natural number P, let

$$f_{P,p}(X_P, y_p) = \mathbb{E}_{i\leq P}\ p\mathbb{1}_{y_p=i}\sup_{y\notin B}\left|\mathbb{E}_{h\leq k}x_{hp+i}f(Ty)\right|.$$

Define

$$f_P(X_P, Y_P) = \mathbb{E}_{P/2<p\leq P}f_{P,p}(X_P y_P).$$

Unpacking definitions, for every natural number P,

$$\mathbb{E}f_P(X_P, Y_P) > c+\eta-\varepsilon.$$

Now we check the hypotheses of Theorem 2.2.14. Because ϕ takes only finitely many values, x_i takes values in a finite set. For all natural numbers P, since the distribution of $(y_p)_{p\leq P}$ is a $+1$ invariant measure on $\prod_{p\leq P}\mathbb{Z}/p\mathbb{Z}$, it must be the uniform distribution. Since μ is translation invariant, for any natural numbers i and m and any subset E of X^m

$$\mathbb{P}((Tx',\ldots,T^m x')\in E) = \mathbb{P}((T^{i+1}x',\ldots,T^{i+m}x')\in E).$$

Applying this to the preimage under $(\phi,\ldots,\phi)$ of an arbitrary subset S of $\mathbb{C}^m$ reveals that the distribution of $(x_1,\ldots,x_m)$ is the same as the distribution of $(x_{i+1},\ldots,x_{i+m})$. Similarly, if b is an element in $\prod_{p\leq P}\mathbb{Z}/p\mathbb{Z}$ if E is the preimage under $(\phi,\ldots,\phi)$ of an arbitrary set S intersected with the set of points z in X such that $M(z) = b \bmod \prod_{p\leq P}p$ then we conclude

$$\mathbb{P}((x_1,\ldots,x_m)\in S \mid (y_p)_{p\leq P} = b) = \mathbb{P}((x_{i+1},\ldots,x_{i+m})\in S \mid (y_p)_{p\leq P} = b+i).$$

For each natural number P and each prime $P/2 < p \leq P$, for at most two values of $i \leq P$ is it true that $y_p = i \bmod p$. Therefore, at most two terms in the sum

$$\mathbb{E}_{i\leq P}\ p\mathbb{1}_{y_p=i}\sup_{y\notin B}\left|\mathbb{E}_{h\leq k}x_{hp+i}f(Ty)\right|$$

are nonzero. Therefore $f_{P,p}$ is bounded by 2. Let W_P be a random variable with the same distribution as Y_P. Then by Theorem 2.2.14

$$\liminf_{P\to\infty} \mathbb{E}[|f_P(X_P, Y_P) - f_P(X_P, W_P)|] = 0.$$

Since, for any natural number P,

$$\mathbb{E} f_P(X_P, Y_P) > c + \eta - \varepsilon.$$

We conclude that

$$\limsup_{P\to\infty} \mathbb{E} f_P(X_P, W_P) > c + \eta - \varepsilon.$$

Unpacking definitions, this proves

$$\limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P}\mathbb{E}_{i\leq P} \int_X \sup_{y\notin B} \left|\mathbb{E}_{h\leq k}\phi(T^{ph+i}x)\cdot f'(T^h y)\right|\mu(dx) > c + \eta - \varepsilon.$$

By the triangle inequality

$$\limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P}\mathbb{E}_{i\leq P} \int_X \sup_{y\notin B} \left|\mathbb{E}_{h\leq k}\mathbb{E}^\mu[f|\mathcal{Z}](T^{ph+i}x)\cdot f'(T^h y)\right|\mu(dx) > c + \eta - 2\varepsilon.$$

Since ε was arbitrary

$$\limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P}\mathbb{E}_{i\leq P} \int_X \sup_{y\notin B} \left|\mathbb{E}_{h\leq k}\mathbb{E}^\mu[f|\mathcal{Z}](T^{ph+i}x)\cdot f'(T^h y)\right|\mu(dx) > c.$$

By translation invariance

$$\limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P} \int_X \sup_{y\notin B} \left|\mathbb{E}_{h\leq k}\mathbb{E}^\mu[f|\mathcal{Z}](T^{ph}x)\cdot f'(T^h y)\right|\mu(dx) > c.$$

This completes the proof. □

2.2.2 Nilsystems and Algebraic Structure

Now we want to use [HK05] to show that $\mathbb{E}^\mu[f|\mathcal{Z}]$ has some local algebraic structure. This algebraic structure makes $\mathbb{E}^\mu[f|\mathcal{Z}]$ much easier to understand than f.

Proposition 2.2.18. *Let ω be an element of Ω such that*

$$\limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P} \int_X \sup_{y\notin B} \left|\mathbb{E}_{h\leq k}\mathbb{E}^{\mu}[f|\mathcal{Z}](T^{ph}x)\cdot f'(T^h y)\right|\mu_\omega(dx) > c.$$

Then for almost all such choices for ω, there exists a collection of nilsystems $(G(j)/\Gamma(j), dx, g(j), \mathcal{B})$, 1-bounded functions F_j and factor maps $\psi_j\colon X\to G(j)/\Gamma(j)$ so that F_j is a nilcharacter on $G(j)/\Gamma(j)$ with frequency nontrivial on the identity component and such that, after identifying F_j with a function on X, we have that $\sum F_j = F$ satisfies $||F||_\infty \leq 1$ and

$$\limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P} \int_X \sup_{y\notin B} \left|\mathbb{E}_{h\leq k}F(T^{ph}x)\cdot f'(T^h y)\right|\mu_\omega(dx) > c.$$

Proof. We are given that

$$\limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P} \int_X \sup_{y\notin B} \left|\mathbb{E}_{h\leq k}\mathbb{E}^{\mu}[f|\mathcal{Z}](T^{ph}x)\cdot f'(T^h y)\right|\mu_\omega(dx) > c.$$

Recall that by Lemma 2.2.7, we know that

$$\limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P} \int_X \sup_{y\notin B} \left|\mathbb{E}_{h\leq k}\mathbb{E}^{\mu_\omega}[f|\mathcal{Z}_\omega](T^{ph}x)\cdot f'(T^h y)\right|\mu_\omega(dx) > c.$$

By [HK05] Theorem 10.1, $(X, \mu_\omega, T, \mathcal{Z})$ is isomorphic to an inverse limit of nilsystems. Therefore, there exists $(G/\Gamma, dx, g, \mathcal{B})$ a nilsystem, $\psi\colon X\to G/\Gamma$ a factor map such that

$$\limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P} \int_X \sup_{y\notin B} \left|\mathbb{E}_{h\leq k}\mathbb{E}^{\mu_\omega}[f|\psi^{-1}(\mathcal{B})](T^{ph}x)\cdot f'(T^h y)\right|\mu_\omega(dx) > c.$$

We denote $F = \mathbb{E}^{\mu_\omega}[f|\psi^{-1}(\mathcal{B})]$. By a Fourier decomposition, we may write F as a sum of nilcharacters, $F = \sum_\xi F_\xi$. For each ξ, either ξ is nontrivial on the identity component of G/Γ or ξ is trivial on the identity component. If ξ is trivial on the identity component and the step s of G is >1, then ξ is actually trivial on G_s. That is because, for any σ in G, the multiplication by σ map $\sigma\colon G/\Gamma\to G/\Gamma$ is continuous so it takes components to components. Let σ_*: components of $G\to$ components of G be the induced map on components and let τ be any other element of G. Then if σ and τ are in the same component of G then for any σ' in G, multiplication by σ' on the right is also continuous, so $\sigma\sigma'$ is in the same component as $\tau\sigma'$ so $\sigma_* = \tau_*$. We return to the general case where σ and τ are not necessarily in the

same component. Also note that, for any element γ in Γ, $(\gamma\sigma)_* = [\gamma, \sigma]_*\sigma_*\gamma_* = [\gamma, \sigma]_*\sigma_*$. Pick n, m, γ and γ' such that $g^n\gamma$ is in the same component as σ and $g^m\gamma'$ is in the same component as τ. Thus $[\sigma, \tau]_* = [g^n\gamma, g^m\gamma']_* = \pi_*[g^n, g^m]_*$ where π is an element of higher order. Of course $[g^n, g^m] = e$ and by induction we get that $[\sigma, \tau]_*$ is the identity and therefore $[\sigma, \tau]\gamma$ is in the identity component for some γ. Therefore, if $s > 1$, the function F_ξ descends to a function on $(G/G_s)/(\Gamma/\Gamma_s)$. By induction, we can almost prove the theorem, namely we can find a collect of nilsystems $(G(j)/\Gamma(j), dx, g(j), \mathcal{B})$ and functions F_j and factor maps $\psi_j \colon X \to G(j)/\Gamma(j)$ so that F_j is a nilcharacter on $G(j)/\Gamma(j)$ with frequency nontrivial on the identity component or $G(j)$ is abelian and such that, after identifying F_j with a function on X, we have that $\sum F_j = F$ satisfies $||F||_\infty \leq 1$ and

$$\limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P} \int_X \sup_{y\notin B} \left|\mathbb{E}_{h\leq k} F(T^{ph}x) \cdot f'(T^h y)\right| \mu_\omega(dx) > c.$$

It remains to observe that the case of a locally constant function on an abelian group cannot occur by Corollary 2.2.3 as follows: we can think of the $F_j's$ as all functions on the group G with some additional equivariance properties; by construction the different F_j's have different frequencies so if F_r is a locally constant function on an abelian group and thus is locally periodic, meaning $F_r(T^h x)$ is a periodic function of h. Then by Corollary 2.2.3,

$$0 = \int_X f \cdot \overline{F_r}\mu_\omega(dx).$$

Since F_r is $\psi^{-1}(\mathcal{B})$ measurable,

$$= \int_{G/\Gamma} F \cdot \overline{F_r} dx.$$

Since all the F_j's have different frequencies, they are all orthogonal to each other.

$$= \int_{G/\Gamma} F_r \cdot \overline{F_r} dx.$$

□

Remark 2.2.19. *Note that if we also know the $\kappa - 1$-Fourier uniformity conjecture then the step of all nilpotent Lie groups is $\geq \kappa$ by Proposition 2.2.4 (plugging in $\overline{F_r} = \phi$ in the statement of that proposition).*

Corollary 2.2.20. *There exists a natural number L independent of k, a nilpotent Lie group G of step s, a cocompact subgroup Γ, an ergodic element g in G and a nilcharacter Φ with nontrivial frequency even when restricted to the identity component,*

$$\limsup_{k\in\mathcal{K}}\limsup_{P\to\infty}\mathbb{E}_{P/2<p\leq P}\int_X\sup_{y\notin B}\Big|\mathbb{E}_{h\leq k}\Phi(g^{ph}x)\cdot f'(T^hy)\Big|dx>\frac{c}{L},$$

where $\mathcal{K}$ as defined in Subsection 2.1.1 is an infinite set such that for k in $\mathcal{K}$, the number of words of length k of f' is $o(k^2)$ if $t=2$ or $O(k^{t-\varepsilon})$ if $t\neq 2$ for some ε.

Proof. By Corollary 2.2.17,

$$\limsup_{P\to\infty}\mathbb{E}_{P/2<p\leq P}\int_X\sup_{y\notin B}\Big|\mathbb{E}_{h\leq k}\mathbb{E}^\mu[f|\mathcal{Z}](T^{ph}x)\cdot f'(T^hy)\Big|\mu(dx)>c.$$

Thus, for a positive measure set of ω,

$$\limsup_{P\to\infty}\mathbb{E}_{P/2<p\leq P}\int_X\sup_{y\notin B}\Big|\mathbb{E}_{h\leq k}\mathbb{E}^\mu[f|\mathcal{Z}](T^{ph}x)\cdot f'(T^hy)\Big|\mu_\omega(dx)>c.$$

By Proposition 2.2.18, we know that for almost every ω,

$$\limsup_{k\in\mathcal{K}}\limsup_{P\to\infty}\mathbb{E}_{P/2<p\leq P}\int_X\sup_{y\notin B}\Big|\mathbb{E}_{h\leq k}\sum_j F_j(T^{ph}x)\cdot f'(T^hy)\Big|\mu_\omega(dx)>c,$$

where F_j is as in Proposition 2.2.18. Fix such an ω.

Since the sum $\sum_j F_j$ converges in $L^2(\mu_\omega)$, there exists a natural number L independent of k such that $||\sum_{j\leq\frac{L}{2}}F_j-\sum_j F_j||_{L^2(\mu_\omega)}<\frac{c}{2}$. By the triangle inequality

$$\begin{aligned}\limsup_{k\in\mathcal{K}}\limsup_{P\to\infty}\mathbb{E}_{P/2<p\leq P}\int_X\sup_{y\notin B}\Big|\mathbb{E}_{h\leq k}\sum_{j\leq\frac{L}{2}}F_j(T^{ph}x)\cdot f'(T^hy)\Big|\mu_\omega(dx)\\ +\int_X\sup_{y\notin B}\mathbb{E}_{h\leq k}\Big|\sum_{j>\frac{L}{2}}F_j(T^{ph}x)\cdot f'(T^hy)\Big|\mu_\omega(dx)>c.\end{aligned}$$

The second term is bounded by

$$\int_X\Big|\sum_{j>\frac{L}{2}}F_j(x)\Big|\mu_\omega(dx),$$

using that μ_ω is shift invariant and f' is 1-bounded. By Cauchy-Schwarz, this term is bounded by $||\sum_{j\leq\frac{L}{2}} F_j - \sum_j F_j||_{L^2(\mu_\omega)} < \frac{c}{2}$. Thus by the triangle inequality.

$$\limsup_{k\in\mathcal{K}} \limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P} \int_X \sup_{y\notin B} \left|\mathbb{E}_{h\leq k} \sum_{j\leq\frac{L}{2}} F_j(T^{ph}x)\cdot f'(T^h y)\right| \mu_\omega(dx) > \frac{c}{2}.$$

By the pigeonhole principle, there exists some F_j such that

$$\limsup_{k\in\mathcal{K}} \limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P} \int_X \sup_{y\notin B} \left|\mathbb{E}_{h\leq k} F_j(T^{ph}x)\cdot f'(T^h y)\right| \mu_\omega(dx) > \frac{c}{L}.$$

Renaming everything gives the conclusion. We remark that the corollary just stated that such an L, G, Γ, g and Φ exist and therefore the statement of the corollary allows L to depend on G, Γ and all the other data that comes from ω. The remainder of the argument essentially takes place inside a single ergodic component and so how the constants vary from component to component is not important for our purposes. □

For the remainder of the proof, we fix G, Γ, g and Φ. We let $c_0 = \frac{c}{L}$. For the next few pages, we fix an integer k in $\mathcal{K}$ such that

$$\limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P} \int_X \sup_{y\notin B} \left|\mathbb{E}_{h\leq k} \Phi(g^{ph}x)\cdot f'(T^h y)\right| dx > c_0.$$

We will later send k to infinity. The following lemma does two things: First, it uses Hölder's inequality to raise the exponent of $\left|\mathbb{E}_{h\leq k}\Phi(g^{ph}x)\cdot f'(T^h y)\right|$. We want this term raised to an even power because we want to expand out the product and get rid of the absolute values which are less "algebraic" and therefore harder to understand directly using the theory of nilpotent Lie groups. We also want this even power to be larger the more oscillatory our function Φ is. This is because the more Φ oscillates, the more cancellation we expect in larger and larger products. The larger the power we use, the smaller the fraction of terms which do not exhibit cancellation is. Second, we use the pigeonhole principle. This lemma and the following lemma are where we make essential use of our bound on the word growth rate of b.

Lemma 2.2.21. *Recall that b had at most $k^{t-\varepsilon}$ words of length k occuring with positive upper logarithmic density for k in $\mathcal{K}$ or b has $o(k^2)$ many words of length k that occur with positive upper logarithmic density if $t = 2$. Fix δ a constant that is small even when compared to c_0. Then for each k in $\mathcal{K}$ there is a word $\epsilon = (\epsilon_1, \ldots, \epsilon_k)$ of length k such that*

$$\limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P} \int_X \left|\mathbb{E}_{h\leq k}\Phi(g^{ph}x)\cdot\epsilon_h\right|^{2t} dx > k^{-t+\varepsilon}c_0^{2t},$$

for $t \neq 2$ and when $t = 2$, we have

$$\limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P} \int_X \left|\mathbb{E}_{h\leq k}\Phi(g^{ph}x)\cdot\epsilon_h\right|^{2t} dx > \delta^{-1}k^{-t}c_0^{2t}.$$

Proof. We know that

$$\limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P} \int_X \sup_{y\notin B} \left|\mathbb{E}_{h\leq k}\Phi(g^{ph}x)\cdot f'(T^h y)\right| dx > c_0.$$

By Holder's inequality, we have

$$\limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P} \int_X \sup_{y\notin B} \left|\mathbb{E}_{h\leq k}\Phi(g^{ph}x)\cdot f'(T^h y)\right|^{2t} dx > c_0^{2t}.$$

Because each term is nonnegative, we can replace the essential sup by a sum over words that occur with positive log-density.

$$\sum_{\epsilon} \limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P} \int_X \left|\mathbb{E}_{h\leq k}\Phi(g^{ph}x)\cdot\epsilon_h\right|^{2t} dx > c_0^{2t}.$$

We assumed that the number of words occuring with positive logarithmic density and therefore the number of terms in the sum is at most δk^t when $t = 2$ or $k^{t-\varepsilon}$ when $t \neq 2$. By the pigeonhole principle, when $t = 2$ there is a word such that

$$\limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P} \int_X \left|\mathbb{E}_{h\leq k}\Phi(g^{ph}x)\cdot\epsilon_h\right|^{2t} dx > \delta^{-1}k^{-t}c_0^{2t},$$

and similarly for $t \neq 2$, we have

$$\limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P} \int_X \left|\mathbb{E}_{h\leq k}\Phi(g^{ph}x)\cdot\epsilon_h\right|^{2t} dx > k^{-t+\varepsilon}c_0^{2t},$$

which completes the proof. □

We need a slightly different estimate for the abelian case. The key to the next lemma is the idea that if $e(\alpha h)$ correlates with ϵ_h for $h \leq k$ then $e(\alpha h)$ also must correlate with translates of ϵ of size $\sim k$. Thus, in the abelian case, the previous lemma is rather lossy. When we replace the sup by a sum, we should gain an extra power of k.

Lemma 2.2.22. *For $t = 2$, for all k in $\mathcal{K}$, there is a word ϵ such that*

$$\limsup_{P\to\infty} \int_X \mathbb{E}_{P/2<p\leq P} \sup_{\ell\in\mathbb{N}} \left|\mathbb{E}_{h\leq k}\Phi(g^{ph+\ell}x)\cdot\epsilon_h\right|^{2t-2} dx > \frac{c_0^{2t-2}}{9}\cdot\left\lfloor\frac{c_0^{2t-2}k}{6}\right\rfloor\delta^{-1}k^{-t}.$$

Proof. Again, we know that

$$\limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P}\int_X \sup_{y\notin B}\left|\mathbb{E}_{h\leq k}\Phi(g^{ph}x)\cdot f'(T^h y)\right|dx > c_0.$$

Again, by Holder's inequality

$$\limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P}\int_X \sup_{y\notin B}\left|\mathbb{E}_{h\leq k}\Phi(g^{ph}x)\cdot f'(T^h y)\right|^{2t-2}dx > c_0^{2t-2}.$$

Again we want to replace $F'(T^h y)$ by a sum over words. Let P be a number satisfying

$$\mathbb{E}_{P/2<p\leq P}\int_X \sup_{y\notin B}\left|\mathbb{E}_{h\leq k}\Phi(g^{ph}x)\cdot f'(T^h y)\right|^{2t-2}dx > c_0^{2t-2}.$$

Let A be the set of x such that

$$\mathbb{E}_{P/2<p\leq P}\sup_{y\notin B}\left|\mathbb{E}_{h\leq k}\Phi(g^{ph}x)\cdot f'(T^h y)\right|^{2t-2} > \frac{2c_0^{2t-2}}{3}.$$

Therefore, the measure of A is at least $\frac{c^{2t-2}}{3}$. We want to show that for μ-almost every x in A, there are at least $\left\lfloor\frac{c^{2t-2}k}{6}\right\rfloor$ many distinct words of f' such that

$$\mathbb{E}_{P/2<p\leq P}\sup_{\ell\in\mathbb{N}}\sup_{y\notin B}\left|\mathbb{E}_{h\leq k}\Phi(g^{ph+\ell}x)\cdot f'(T^h y)\right|^{2t-2} > \frac{c_0^{2t-2}}{3}.$$

Let y be an element of B^c such that the words of $f'(T^h y)$ are words of Y and such that

$$\mathbb{E}_{P/2<p\leq P}\left|\mathbb{E}_{h\leq k}\Phi(g^{ph}x)\cdot f'(T^h y)\right|^{2t-2} > \frac{2c_0^{2t-2}}{3}.$$

Denote by $\epsilon(m)$ the word of length k whose h^{th} entry is $\epsilon(m)_h = f'(T^{m+h}y)$. If the words $\epsilon(m)$ are distinct for $m = 1, \ldots, \left\lfloor \frac{c_0^{2s}k}{6} \right\rfloor$ then by the triangle inequality

$$\begin{aligned}&\mathbb{E}_{P/2<p\leq P} \sup_{y\notin B} \left|\mathbb{E}_{h\leq k}\Phi(g^{ph+pm}x)\cdot\epsilon(m)_h\right|^{2t-2}\\ =&\mathbb{E}_{P/2<p\leq P}\left|\mathbb{E}_{h\leq k}\Phi(g^{ph+pm}x)\cdot f'(T^{h+pm}y)\right|^{2t-2}.\end{aligned}$$

We note that all but $2mk$ many terms in the average are the same if we replace $\Phi(g^{ph+pm}x)\cdot f'(T^{h+pm}y)$ with $\Phi(g^{ph}x)\cdot f'(T^h y)$. Thus

$$\begin{aligned}\geq&\mathbb{E}_{P/2<p\leq P}\left|\mathbb{E}_{h\leq k}\Phi(g^{ph}x)\cdot f'(T^h y)\right|^{2t-2} - \frac{2m}{k}\\ \geq&\mathbb{E}_{P/2<p\leq P}\left|\mathbb{E}_{h\leq k}\Phi(g^{ph}x)\cdot f'(T^h y)\right|^{2t-2} - \frac{c_0^{2t-2}}{3}.\end{aligned}$$

Suppose for a moment that instead the words $\epsilon(m)$ are not distinct for $m = 1, \ldots, \left\lfloor \frac{c_0^{2t-2}k}{6} \right\rfloor$. Then there exist a minimum j such that $\epsilon(1), \ldots, \epsilon(j)$ are not distinct. Fix such a j for the remainder of Section 2.2. Thus, there exists some $1 \leq d < j$ such that $\epsilon(j) = \epsilon(j-d)$. We claim that $\epsilon(j-d)$ is d-periodic: that's because $\epsilon(j-d)_h = \epsilon(j)_h = \epsilon(j-d)_{h+d}$. Furthermore, if $\epsilon(j-d-1)_1 = \epsilon(j-1)_1$ then since $\epsilon(j-d-1)_h = \epsilon(j-d)_{h-1} = \epsilon(j)_{h-1} = \epsilon(j-1)_h$ for all $h > 1$, we clearly have $\epsilon(j-d-1) = \epsilon(j-1)$ and j is not minimal. For the rest of the proof, let r be the minimum number such that $r \geq j-d$ and $\epsilon(r)$ is not d periodic. For y not in B, we can find such an r because $f'(T^h y)$ is not eventually periodic. Since $\epsilon(r)$ is not d periodic but $\epsilon(r-1)$ is d periodic and is equal to $\epsilon(q)$ for some q between $j-d$ and $j-1$, we have that $\epsilon(r)_k \neq \epsilon(r)_{k-d}$ but $\epsilon(r)_h = \epsilon(r)_{h-d}$ for all other $h \leq k$. We claim that the words $1, \ldots, j-1$ and $r, \ldots, r + \left\lfloor \frac{c_0^{2t-2}k}{6} \right\rfloor - j + 1$ are all distinct. The reason is that for all m between 1 and $j-d$, we have that $\epsilon(m)_h = \epsilon(j-d)_{h+m-j+d}$ for all $h > m-j+d$ and precisely no larger range of h and for all m between r and $r + \left\lfloor \frac{c_0^{2t-2}k}{6} \right\rfloor - j + 1$ we have that $\epsilon(m)_h = \epsilon(j-d)_{h+m-r}$ for all $1 \leq h < k-m+r$ and precisely no larger range of h. For m between $j-d$ and $j-1$, $\epsilon(m)$ is d periodic but because j was the minimal natural number such that $\epsilon(1), \ldots, \epsilon(j)$ are not distinct, we have that the $\epsilon(m)$ for m between $j-d$ and $j-1$ are still distinct. For m between 1 and $j-d$ and m' between r and $r + \left\lfloor \frac{c_0^{2t-2}k}{6} \right\rfloor - j + d$ we

have that the intervals $h > m-j+d$ and $h < k-m+r$ meet so the previous argument shows that $\epsilon(m) \neq \epsilon(m')$. A similar triangle inequality computation shows that for m between r and $r + \left\lfloor \frac{c_0^{2t-2}k}{6} \right\rfloor - j + 1$ we still have

$$\begin{aligned}
&\mathbb{E}_{P/2<p\leq P} \sup_{\ell\in\mathbb{N}} \left|\mathbb{E}_{h\leq k}\Phi(g^{ph+\ell}x)\cdot\epsilon(m)_h\right|^{2t-2} \\
=&\mathbb{E}_{P/2<p\leq P} \sup_{\ell\in\mathbb{N}} \left|\mathbb{E}_{h\leq k}\Phi(g^{ph+\ell}x)\cdot f'(T^{h+m}y)\right|^{2t-2} \\
\geq&\mathbb{E}_{P/2<p\leq P} \sup_{\ell\in\mathbb{N}} \left|\mathbb{E}_{h\leq k}\Phi(g^{ph+\ell}x)\cdot f'(T^{h+r-1}y)\right|^{2t-2} - \frac{2(m-r+1)}{k}
\end{aligned}$$

Next, we use that $\epsilon(r-1) = \epsilon(q)$ for some q between $j-d$ and $j-1$.

$$\begin{aligned}
\geq&\mathbb{E}_{P/2<p\leq P} \sup_{\ell\in\mathbb{N}} \left|\mathbb{E}_{h\leq k}\Phi(g^{ph+\ell}x)\cdot f'(T^{h+q}y)\right|^{2t-2} - \frac{2(m-r+1)}{k} \\
\geq&\mathbb{E}_{P/2<p\leq P}\left|\mathbb{E}_{h\leq k}\Phi(g^{ph}x)\cdot f'(T^{h}y)\right|^{2t-2} - \frac{c_0^{2t-2}}{3}.
\end{aligned}$$

This proves the claim that for x in A there are at least $\left\lfloor \frac{c_0^{2t-2}k}{6} \right\rfloor$ many distinct words ϵ of f' such that

$$\mathbb{E}_{P/2<p\leq P} \sup_{\ell\in\mathbb{N}} \left|\mathbb{E}_{h\leq k}\Phi(g^{ph+\ell}x)\cdot\epsilon_h\right|^{2t-2} > \frac{c_0^{2t-2}}{3}.$$

Summing over words we get that for almost every x in A,

$$\begin{aligned}
&\sum_{\epsilon \text{ a word of } f'} \mathbb{E}_{P/2<p\leq P} \sup_{\ell\in\mathbb{N}} \left|\mathbb{E}_{h\leq k}\Phi(g^{ph+\ell}x)\cdot\epsilon_h\right|^{2t-2} \\
>&\left\lfloor \frac{c^{2t-2}k}{6} \right\rfloor \cdot \frac{c_0^{2t-2}}{3}.
\end{aligned}$$

Next, we use that $\mu(A) > \frac{c_0^{2t-2}}{3}$.

$$\begin{aligned}
&\int_X \sum_{\epsilon \text{ a word of } f'} \sup_{\ell\in\mathbb{N}} \mathbb{E}_{P/2<p\leq P}\left|\mathbb{E}_{h\leq k}\Phi(g^{ph+\ell}x)\cdot\epsilon_h\right|^{2t-2} \\
>&\frac{c_0^{2t-2}}{3}\left\lfloor \frac{c_0^{2t-2}k}{6} \right\rfloor \cdot \frac{c_0^{2t-2}}{3}.
\end{aligned}$$

Sending P to infinity and using the pigeonhole principle, we deduce that for some word ϵ,

$$\begin{aligned}
&\limsup_{P\to\infty} \int_X \sup_{\ell\in\mathbb{N}} \mathbb{E}_{P/2<p\leq P}\left|\mathbb{E}_{h\leq k}\Phi(g^{ph+\ell}x)\cdot\epsilon_h\right|^{2t-2} dx \\
>&\frac{c_0^{2t-2}}{9} \cdot \left\lfloor \frac{c_0^{2t-2}k}{6} \right\rfloor \delta^{-1}k^{-t}.
\end{aligned}$$

□

Remark 2.2.23. *For G abelian and therefore Φ a character, we have*

$$\begin{aligned}&\limsup_{P\to\infty}\int_X \mathbb{E}_{P/2<p\leq P}\sup_{\ell\in\mathbb{N}}\left|\mathbb{E}_{h\leq k}\Phi(g^{ph+\ell}x)\cdot\epsilon_h\right|^{2t-2}dx\\ =&\limsup_{P\to\infty}\int_X \mathbb{E}_{P/2<p\leq P}\left|\mathbb{E}_{h\leq k}\Phi(g^{ph}x)\cdot\epsilon_h\right|^{2t-2}dx\end{aligned}$$

Therefore, by choosing δ sufficiently small, in the abelian case, we get

$$\limsup_{k\in\mathcal{K}}\limsup_{P\to\infty}\mathbb{E}_{P/2<p\leq P}\int_X|\mathbb{E}_{h\leq k}\epsilon_h\Phi(g^{ph}x)|^2dx\gg k^{-1}.$$

The next theorem contradicts the previous two lemmas and proves Theorem 2.1.8. In its proof, we rely heavily on [BDG16], [Fra17], [GT10], [GT12a] and [GTZ12].

Theorem 2.2.24. *Recall that, after Corollary 2.2.20, we fixed a nilpotent Lie group G, a cocompact lattice Γ, a nilcharacter Φ with with nontrivial frequency on the identity component ξ and an element g which acts ergodically on G/Γ, such that $||\Phi||_{L^\infty_x}=1$. Recall that the step s of G is at least κ where $t=\binom{\kappa+1}{2}$. Let ϵ be a sequence of words implicitly depending on k. Let $\varepsilon>0$. Then*

$$\limsup_{k\in\mathcal{K}}\limsup_{P\to\infty}\int_X\mathbb{E}_{P/2<p\leq P}|\mathbb{E}_{h\leq k}\epsilon_h\Phi(g^{ph}x)|^{2t}dx\cdot k^{t-\varepsilon}=0,$$

If $t=2$ then we do not need the epsilon loss and instead get the estimate

$$\limsup_{k\in\mathcal{K}}\limsup_{P\to\infty}\int_X\mathbb{E}_{P/2<p\leq P}|\mathbb{E}_{h\leq k}\epsilon_h\Phi(g^{ph}x)|^{2t}dx\cdot k^{t}\leq C_s.$$

This contradicts Lemmas 2.2.21 and 2.2.22 as follows. When G is abelian and thus $t=2$, Lemma 2.2.22 states that there is a word ϵ of length k such that

$$\begin{aligned}&\limsup_{P\to\infty}\int_X \mathbb{E}_{P/2<p\leq P}\sup_{\ell\in\mathbb{N}}\left|\mathbb{E}_{h\leq k}\Phi(g^{ph+\ell}x)\cdot\epsilon_h\right|^{2t-2}dx\\ >&\frac{c_0^{2t-2}}{9}\cdot\left\lfloor\frac{c_0^{2t-2}k}{6}\right\rfloor\delta^{-1}k^{-t},\end{aligned}$$

for any δ we choose so long as k is chosen from the set $\mathcal{K}$ of natural numbers such that f' has fewer than δk^2 many words of length k. Thus, picking δ small, (in particular, smaller than say $C_1^{-1}100c_0^4$), we find that

$$\limsup_{P\to\infty}\int_X \mathbb{E}_{P/2<p\leq P}\left|\mathbb{E}_{h\leq k}\Phi(g^{ph}x)\cdot\epsilon_h\right|^{2t-2}dx > k^{-t}C_1,$$

contradicting Theorem 2.2.24. (Note that we have replaced $|\mathbb{E}_{h\leq k}\Phi(g^{ph+\ell})\epsilon_h|$ by the same expression without the shift in ℓ as in Remark 2.2.23). Similarly, if the group is not abelian, Lemma 2.2.21 states that there exists a word ϵ of length k such that

$$\limsup_{P\to\infty}\mathbb{E}_{P/2<p\leq P}\int_X \left|\mathbb{E}_{h\leq k}\Phi(g^{ph}x)\cdot\epsilon_h\right|^{2t}dx > k^{-t+\varepsilon}c_0^{2t},$$

for $t\neq 2$ and when $t=2$, we have

$$\limsup_{P\to\infty}\mathbb{E}_{P/2<p\leq P}\int_X \left|\mathbb{E}_{h\leq k}\Phi(g^{ph}x)\cdot\epsilon_h\right|^{2t}dx > \delta^{-1}k^{-t}c_0^{2t}.$$

When $t=2$, again by picking δ small, this time smaller than $C_s^{-1}c_0^{2t}$, proves that

$$\limsup_{P\to\infty}\mathbb{E}_{P/2<p\leq P}\int_X \left|\mathbb{E}_{h\leq k}\Phi(g^{ph}x)\cdot\epsilon_h\right|^{2t}dx > C_s k^{-t},$$

again contadicting Theorem 2.2.24. Finally, when $t\neq 2$,

$$\limsup_{P\to\infty}\mathbb{E}_{P/2<p\leq P}\int_X \left|\mathbb{E}_{h\leq k}\Phi(g^{ph}x)\cdot\epsilon_h\right|^{2t}dx > k^{-t+\varepsilon}c_0^{2t},$$

contradicts

$$\limsup_{k\in\mathcal{K}}\limsup_{P\to\infty}\int_X \mathbb{E}_{P/2<p\leq P}|\mathbb{E}_{h\leq k}\epsilon_h\Phi(g^{ph}x)|^{2t}dx\cdot k^{t-\varepsilon}=0,$$

from Theorem 2.2.24.

Thus, the rest of this section will be devoted to showing that Theorem 2.2.24 is true. Suppose not and for the moment fix k in $\mathcal{K}$ such that

$$\limsup_{P\to\infty}\mathbb{E}_{P/2<p\leq P}\int_X \left|\mathbb{E}_{h\leq k}\epsilon_h\Phi(g^{ph}x)\right|^{2t}dx \gg k^{-t+\varepsilon}.$$

The first step is to replace averages over primes by uniform averages over natural numbers. To do this, we need the machinery of Green-Tao [GT12a] [GT10] and Green-Tao-Ziegler

[GTZ12]. By the triangle inequality, we may replace averages over primes by averages weighted by the von Mangoldt function.

$$\limsup_{P\to\infty} \mathbb{E}_{P/2<n\leq P} \int_X \Lambda(n) \left| \mathbb{E}_{h\leq k} \epsilon_h \Phi(g^{ph}x) \right|^{2t} dx \gg k^{-t+\varepsilon}.$$

We denote $\psi_x(m) = \Phi(g^m x)$. We expand:

$$\limsup_{P\to\infty} \mathbb{E}_{P/2<n\leq P} \int_X \Lambda(n) \mathbb{E}_{J\in[k]^{2t}} \epsilon_J \psi_x(nj_1)\cdot\dots\cdot\psi_x(nj_t)\cdot \overline{\psi_x}(nj_{t+1})\cdot\dots\cdot\overline{\psi_x}(nj_{2t})dx \gg k^{-t+\varepsilon},$$

where ϵ_j is a phase given by the formula $\epsilon_J = \epsilon_{j_1}\cdots\epsilon_{j_t}\cdot\overline{\epsilon}_{j_{t+1}}\cdots\overline{\epsilon}_{j_{2t}}$.

We say $J \in [k]^{2t}$ is diagonal if $\#\{m \leq t\colon j_m = h\} = \#\{m > t\colon j_m = h\}$ for all $h \leq k$. We say J solves the Vinogradov mean value problem if, for all m between 1 and s, we have

$$j_1^m + \cdots + j_t^m = j_{t+1}^m + \cdots + j_{2t}^m.$$

Every diagonal J also solves Vinogradov's mean value problem. We rely on the following Theorem due to Bourgain, Demeter and Guth which says that those account for "most" solutions, up to a constant.

Theorem 2.2.25 ([BDG16] Theorem 1.1)**.** *For all ε and s there exists a constant $C_{s,\varepsilon}$ such that the number of solutions to the Vinogradov mean value problem is less than $C_{s,\varepsilon}k^{t+\varepsilon}$ where $t \leq \binom{s+1}{2}$.*

We will show that if J does not solve Vinogradov's mean value problem then J does not contribute to the sum. Thus, fix J which does not solve Vinogradov's mean value theorem and suppose that

$$\limsup_{P\to\infty} \left| \mathbb{E}_{P/2<n\leq P} \int_X \Lambda(n)\psi_x(nj_1)\cdot\dots\cdot\psi_x(nj_t)\cdot\overline{\psi_x}(nj_{t+1})\cdot\dots\cdot\overline{\psi_x}(nj_{2t})dx \right| \gtrsim_{k,s,c} 1.$$

We denote

$$\Psi_x(n) = \psi_x(nj_1)\cdot\dots\cdot\psi_x(nj_t)\cdot\overline{\psi_x}(nj_{t+1})\cdot\dots\cdot\overline{\psi_x}(nj_{2t}).$$

Fix a subsequence such that

$$\lim_{P\in I}\left|\mathbb{E}_{P/2<n\leq P}\int_X \Lambda(n)\Psi_x(n)dx\right| = \limsup_{P\to\infty}\left|\mathbb{E}_{P/2<n\leq P}\int_X \Lambda(n)\Psi_x(n)dx\right| \gtrsim_{k,s,c} 1,$$

where I is some infinite subset of the natural numbers and where the implied constant may depend on Ψ_x. Fix a large number W, a product of many small primes. We will later choose exactly how large W must be. We pass to a subsequence where the following limit exists for each $b \leq W$,

$$\lim_{P\in I'}\left|\mathbb{E}_{P/2<Wn\leq P}\int_X \Lambda(Wn+b)\Psi_x(Wn+b)dx\right|,$$

where I' is an infinite subset of I. We may do this by a diagonalization argument. By the triangle inequality,

$$\mathbb{E}_{b<W}\lim_{P\in I'}\left|\mathbb{E}_{P/2<Wn+b\leq P}\int_X \Lambda(Wn+b)\Psi_x(Wn+b)dx\right| \gtrsim_{k,s,c} 1,$$

where the implied constant does not depend on W. Note that because $b < W$, we miss at most one term by changing the bounds of the sum from $P/2 < Wn+b \leq P$ to $P/2 < Wn \leq P$. Since W is much smaller than P, this is an acceptable error. Note that if b is not coprime to W, then

$$\lim_{P\in I'}\mathbb{E}_{P/2<Wn\leq P}\int_X \Lambda(Wn+b)\Psi_x(Wn+b)dx = 0,$$

because $Wn+b$ is never prime. By the pigeonhole principle, there exists $b < W$ such that

$$\lim_{P\in I'}\left|\mathbb{E}_{P/2<Wn\leq P}\int_X \frac{W}{\varphi(W)}\Lambda(Wn+b)\Psi_x(Wn+b)dx\right| \gtrsim_{k,s,c} 1,$$

where again the implied constant does not depend on W and where $\varphi(W)$ is Euler's torient function, the function which counts the number of residue classes mod W that are coprime to W. Denote $\frac{W}{\varphi(W)}\Lambda(Wn+b) = \Lambda_{b,W}$. Then we can write our expression as a sum of two terms

$$\begin{aligned}&\lim_{P\in I'}\left|\mathbb{E}_{P/2<Wn\leq P}\int_X \Lambda_{b,W}(n)\Psi_x(Wn+b)dx\right| \\ =&\lim_{P\in I'}\left|\mathbb{E}_{P/2<Wn\leq P}\int_X (\Lambda_{b,W}(n)-1)\Psi_x(Wn+b)+\Psi_x(Wn+b)dx\right|\end{aligned}$$

To handle the first term, we need the following theorems of Green-Tao and Green-Tao-Ziegler.

Theorem 2.2.26 ([GT10] Proposition 11.2)**.** *Let G/Γ be a degree s filtered nilmanifold, and let $M > 0$. Suppose that $F(g^n x)_{n=1}^{\infty}$ is a bounded nilsequence on G/Γ with Lipschitz constant at most M, where F is a function on G/Γ, g is an element of G and x is a point in G/Γ. Let $\varepsilon \in (0,1)$ and P a large natural number. Then we may decompose*

$$F(g^n x) = F_1(n) + F_2(n),$$

where $F_1 \colon \mathbb{N} \to [-1,1]$ is a sequence with Lipschitz constant $O_{M,\varepsilon,G/\Gamma}(1)$ and obeying the dual norm bound

$$||F_1||_{U^{s+1}[P/2<Wn\leq P]^*} = O_{M,\varepsilon,G/\Gamma}(1),$$

while $F_2 \colon \mathbb{N} \to \mathbb{R}$ obeys the uniform bound

$$||F_2||_{\infty} \leq \varepsilon.$$

Note that the bound $||F_1||_{U^{s+1}[N]^*} = O_{M,\varepsilon,G/\Gamma}(1)$ is uniform in the element g. We also need the following theorem of Green-Tao-Ziegler. The proof of this theorem is spread out over [GTZ12], [GT10] and [GT12a], making it somewhat hard to give a specific theorem number. Essentially, if the Gowers norm were big then the Inverse Conjecture for the Gowers Norms would imply that the Mobius function correlates with a nilsequence which it does not by the Mobius-Nilsequence Conjecture. In [GT10], Theorem 7.2 states the theorem follows from the Mobius-Nilsequence Conjecture and the Inverse Conjecture for the Gowers Norms. The first of these conjectures is an immediate consequence of Theorem 1.1 in [GT12a]. The second of these conjectures is Theorem 1.3 in [GTZ12].

Theorem 2.2.27 ([GTZ12]; see also [GT12a] and [GT10])**.** *With all the notation as before,*

$$||\Lambda_{b,W} - 1||_{U^{s+1}[P/2<Wn\leq P]} = o_{W\to\infty}(1).$$

Thus, our nilsequence Ψ_x can be written as a sum $\Psi_x = F_1 + F_2$ where F_1 and F_2 implicitly depend on x and enjoy the following properties. F_2 is uniformly small so

$$\limsup_{P\in I'} \left|\mathbb{E}_{P/2<Wn\leq P}(\Lambda_{b,W}(n) - 1)F_1(Wn+b)\right|$$

can be estimated by simply moving the absolute values inside. The remaining term is bounded in dual norm so

$$(\Lambda_{b,W}-1)(n)\cdot F_1(n)\leq ||\Lambda_{b,W}-1||_{U^{s+1}[P/2<Wn\leq P]}\cdot ||F_1||_{U^{s+1}[P/2<Wn\leq P]^*}$$

which tends to 0. For a similar argument, see the proof of Proposition 10.2 in [GT10]. It may also be possible to circumvent the use of [GTZ12] by using Theorem 7.1 in [GT12a]. Putting this together, we get that

$$\limsup_{P\in I'}\left|\mathbb{E}_{P/2<Wn\leq P}\int_X(\Lambda_{b,W}(n)-1)\Psi_x(Wn+b)dx\right|=o_{W\to\infty}(1).$$

As such for W sufficiently large, by the triangle inequality

$$\liminf_{P\in I'}\left|\mathbb{E}_{P/2<Wn\leq P}\int_X\Psi_x(Wn+b)dx\right|\gtrsim_{k,s,c}1.$$

So far we exploited cancellation in the $\Lambda_{b,W}(n)-1$ term and simply boundedness in the $\Psi_x(n)$ term. Next, we will try to exploit cancellation in Ψ_x to obtain a contradiction. To exploit this cancellation we interpret the average as an integral over a complicated nilmanifold, then use the fact that the frequency of Φ is nontrivial on the identity component of G/Γ and therefore nontrivial on every component of G/Γ. Let $G^{2t}=G\times\cdots\times G$ be the product of G with itself $2t$ many times and let $\mathbf{g}=(g^{Wj_1},g^{Wj_2},\ldots,g^{Wj_{2t}})$ be the element of G^{2t} whose ℓ^{th} coordinate is g^{Wj_ℓ}. For any σ in G let $\Delta\sigma=(\sigma,\ldots,\sigma)$ be the element of G^{2t} whose entries are all σ and let G_Δ be the set of all the elements of the form $\Delta\sigma$. Define

$$\mathbf{G}=\overline{\langle\mathbf{g},G_\Delta,\Gamma\times\cdots\times\Gamma\rangle},$$

the closure of the group generated by $\mathbf{g}$, G_Δ and Γ^{2t} inside G^{2t}. Our sequence Ψ_x is a nilsequence on $\mathbf{G}/\Gamma^{2t}$. Consider the sequence of "empirical" measures on G^{2t}/Γ^{2t},

$$\rho_P=\mathbb{E}_{P/2<Wn\leq P}(g^{(Wn+b)j_1},\ldots,g^{(Wn+b)j_{2t}})_*(\Delta_*dx),$$

where Δ_*dx is the Haar measure on $G_\Delta/(\Gamma^{2s}\cap G_\Delta)$ and where $*$ denotes the pushforward. By construction, if $\Xi\colon G^{2t}/\Gamma^{2t}\to\mathbb{C}$ is defined by

$$\Xi(x_1,\ldots,x_{2t})=\prod_{j=1}^{t}\Phi(x_j)\cdot\prod_{j=t+1}^{2t}\overline{\Phi}(x_j),$$

then

$$\mathbb{E}_{P/2<Wn\leq P}\int_X \Psi_x(Wn+b)dx = \int_{G^{2t}/\Gamma^{2t}} \Xi \ \ \rho_P(dx).$$

By the Banach-Alaoglu theorem, there is a further subsequence along which the empirical measures converge weakly,

$$\lim_{P\in I''} \rho_P \overset{*}{\rightharpoonup} \rho,$$

where I'' is an infinite subset of I'. Note that, by summation by parts, ρ_P is almost invariant by $\mathbf{g}$ in the following sense:

$$\mathbf{g}_*\rho_P = \rho_P + O\left(\frac{W}{P}\right)$$

Therefore ρ is actually $\mathbf{g}$ invariant. Since ρ is an average of G_Δ invariant measures, ρ is a also G_Δ invariant. Of course ρ is also Γ^{2t} invariant because Γ^{2t} acts trivially on G^{2t}/Γ^{2t}. Since stabilizers of measures are closed, ρ is invariant under $\mathbf{G}$. By the classification of invariant measures, we know that ρ is actually (a translate of) Haar measure on some nilmanifold $\mathbf{X}$. Next we need the following result essentially due to Frantzikinakis [Fra17].

Lemma 2.2.28 ([Fra17]; see section 5.7 and especially the proof of Proposition 5.7)**.** *With all the notation as before, for any $u \in G_s$ and $m \leq s$, we have $(u^{(Wj_\ell)^m})_{\ell=1}^{2t} \in \boldsymbol{G}$.*

We include the proof for completeness and because our result differs very slightly from the way it was stated in [Fra17].

Proof. We split Lemma 2.2.28 into three claims:

Claim 2.2.29. *Let m and ℓ be natural numbers. If g_1 is in $\boldsymbol{G}_m$ and g_2 and g_3 are in $\boldsymbol{G}_r$ then there exists σ in $\boldsymbol{G}_{m+r+1}$ such that*

$$[g_1, g_2]\cdot[g_1, g_3] = [g_1, g_2\cdot g_3]\cdot\sigma.$$

Moreover, σ depends continuously on g_1, g_2 and g_3. In fact, this holds for any nilpotent Lie group, not just $\boldsymbol{G}$.

Claim 2.2.30. *For any m between 1 and s and any element τ in G_m and in the identity component (which is automatic for $m > 1$), there exists an element σ in $(G_{m+1})^{2t}$ such that*

$$(\tau^{(Wj_\ell)^m})_{\ell=1}^{2t} \cdot \sigma \in \boldsymbol{G}.$$

Claim 2.2.31. *For any natural number r between 1 and s and any natural number m between 1 and r and for any τ in G_r there exists an element σ in $(G_{r+1})^{2s}$ such that*

$$(\tau^{(Wj_\ell)^m})_{\ell=1}^{2t} \cdot \sigma \in \boldsymbol{G}.$$

We remark that taking $\tau = u$ in Claim 2.2.31 gives Lemma 2.2.28.

Proof of Claim 2.2.29. The proof is simply a computation. For any g_1, g_2 and g_3 as above

$$\begin{aligned}
[g_1, g_2] \cdot [g_1, g_3] =& g_1 g_2 g_1^{-1} g_2^{-1} [g_1, g_3] \\
=& g_1 g_2 g_1^{-1} [g_1, g_3] g_2^{-1} \bmod G_{m+r+1} \\
=& g_1 g_2 g_3 g_1^{-1} g_3^{-1} g_2^{-1} \bmod G_{m+r+1} \\
=& [g_1, g_2 g_3] \bmod G_{m+r+1}
\end{aligned}$$

□

Proof of Claim 2.2.30. We prove Claim 2.2.30 by induction on m. First, suppose $m = 1$. Consider the torus $Z = G/G_2\Gamma$. Let π be the projection map $\pi\colon G \to Z$. Then since g acts ergodically on G/Γ, we know $\pi(g)$ is an ergodic element in Z. Therefore, for any $\pi(\tau)$ in G/G_2, note that $\pi(\tau)$ is in the orbit of $\pi(g)$. By the definition of $\mathbf{G}$, $(\pi(g^{Wj_\ell}))_{\ell=1}^{2t}$ is an element of $\pi^{2t}(\mathbf{G})$. Thus, for any τ in G, $(\pi(\tau^{Wj_\ell}))_{\ell=1}^{2t}$ is an element of $\pi^{2t}(\mathbf{G})$ so by definition of the quotient

$$(\tau^{Wj_\ell})_{\ell=1}^{2t} \cdot \sigma \in \mathbf{G}$$

for some σ in G_2^{2t}.

Next, assume by induction that Claim 2.2.30 holds for m. We will try to prove the claim for $m+1$. We begin with the case where τ is the commutator of two elements of the following

form. Suppose that there exists g_1 in G and g_2 in G_m such that $[g_1, g_2] = \tau$. By assumption, there exists σ_1 in G_2^{2t} and σ_2 in G_{m+1}^{2t} such that

$$(g_1^{Wj_\ell})_{\ell=1}^{2t} \cdot \sigma_1 \in \mathbf{G} \text{ and } (g_2^{(Wj_\ell)^m})_{\ell=1}^{2t} \cdot \sigma_2 \in \mathbf{G}.$$

Since $\mathbf{G}$ is a group, we conclude that the commutator is in $\mathbf{G}$.

$$[(g_1^{Wj_\ell})_{\ell=1}^{2t} \cdot \sigma_1, (g_2^{(Wj_\ell)^m})_{\ell=1}^{2t} \cdot \sigma_2] \in \mathbf{G}.$$

Using Claim 2.2.29 repeatedly, this is

$$([g_1, g_2]^{Wj_\ell^{m+1}})_{\ell=1}^{2t} \sigma \in \mathbf{G},$$

for some σ in G_{m+2}^{2t}.

Finally, we note that commutators generate G_{m+1} so it suffices to show that if τ_1 and τ_2 are elements of G_{m+1} that satisfy Claim 2.2.30 then so does their product. After all, if

$$(\tau_1^{(Wj_\ell)^{m+1}})_{\ell=1}^{2t} \cdot \sigma_1 \in \mathbf{G} \text{ and } (\tau_2^{(Wj_\ell)^{m+1}})_{\ell=1}^{2t} \cdot \sigma_2 \in \mathbf{G},$$

where σ_1 and σ_2 are in G_{m+2}^{2t} then

$$\begin{aligned}
&(\tau_1^{(Wj_\ell)^{m+1}})_{\ell=1}^{2t} \cdot \sigma_1 \cdot (\tau_2^{(Wj_\ell)^{m+1}})_{\ell=1}^{2t} \cdot \sigma_2 \\
=&(\tau_1^{(Wj_\ell)^{m+1}})_{\ell=1}^{2t} \cdot (\tau_2^{(Wj_\ell)^{m+1}})_{\ell=1}^{2t} \cdot \sigma \\
=&((\tau_1\tau_2)^{(Wj_\ell)^{m+1}})_{\ell=1}^{2t} \cdot \sigma' \\
=&(\tau^{(Wj_\ell)^{m+1}})_{\ell=1}^{2t} \cdot \sigma',
\end{aligned}$$

where σ and σ' are in G_{m+2}^{2t}. This completes the proof of Claim 2.2.30. □

Proof of Claim 2.2.31. First, if $m = r$ then we are done by Claim 2.2.30. Thus, we will assume $m < r$.

Second, we check that if τ_1 and τ_2 are in G_r and satisfy Claim 2.2.31 then so does their product. By assumption, we may write

$$(\tau_i^{(Wj_\ell)^m})_{\ell=1}^{2t} \cdot \sigma_i \in \mathbf{G},$$

where σ_i is an element of $(G_{r+1})^{2s}$ and $i = 1, 2$. Then the product is given by

$$\begin{aligned}&(\tau_1^{(Wj_\ell)^m})_{\ell=1}^{2t} \cdot \sigma_1 \cdot (\tau_2^{(Wj_\ell)^m})_{\ell=1}^{2t} \cdot \sigma_2 \\ =&(\tau_1^{(Wj_\ell)^m})_{\ell=1}^{2t} \cdot (\tau_2^{(Wj_\ell)^m})_{\ell=1}^{2t} \cdot \sigma_1 \cdot [\sigma_1^{-1}, (\tau_2^{-(Wj_\ell)^m})_{\ell=1}^{2t}] \cdot \sigma_2.\end{aligned}$$

Then we use that $\tau_1^{(Wj_\ell)^m} \cdot \tau_2^{(Wj_\ell)^m} = (\tau_1\tau_2)^{(Wj_\ell)^m}$ up to higher order terms.

$$=((\tau_1\tau_2)^{(Wj_\ell)^m})_{\ell=1}^{2t} \bmod (G_{r+1})^{2s}.$$

Therefore, it suffices to prove Claim 2.2.31 in the case that $\tau = [g_1, g_2]$ where g_1 is in G_m and g_2 is in G_{r-m} because such commutators generate G_r as a group up to higher order corrections.

By Claim 2.2.30, there exists σ in $(G_{m+1})^{2s}$ such that

$$(g_1^{(Wj_\ell)^m})_{\ell=1}^{2t} \cdot \sigma \in \mathbf{G}.$$

We also know, because $\mathbf{G}$ contains diagonal elements, that $(g_2)_{\ell=1}^{2t}$ is an element of $\mathbf{G}$. We conclude that

$$[(g_1^{(Wj_\ell)^m})_{\ell=1}^{2t} \cdot \sigma, (g_2)_{\ell=1}^{2t}] \in \mathbf{G}.$$

By Claim 2.2.29, this is given by

$$(\tau^{(Wj_\ell)^m})_{\ell=1}^{2t} \cdot \sigma' \in \mathbf{G},$$

for some σ' in G_{r+1}. □

This completes the proof of Lemma 2.2.28 by plugging in $r = s$. □

Since the frequency ξ of Φ is nontrivial on the identity component, there exists an element u in the identity component of G_s such that $\frac{1}{2\pi i} \log \xi(u)$ is irrational. Fix such a u. Now since J does not solve Vinogradov's mean value problem there exists $m \leq s$ such that $j_1^m + \cdots + j_t^m - j_{t+1}^m - \cdots - j_{2t}^m \neq 0$. Fix such an m. Then the map $G_s \to G_s$ given by $v \mapsto v^{(j_1^m + \cdots + j_t^m - j_{t+1}^m - \cdots - j_{2t}^m)W^m}$ has image both open and closed so u is in the image. For more

details, see [Fra17]. Fix a v such that $v \mapsto u$. Then by Lemma 2.2.28, $(v^{(Wj_\ell)^m})_{\ell=1}^{2t} \in \mathbf{G}$. As such

$$\begin{aligned}\int_{\mathbf{X}} \Xi(x)\ \rho(dx) &= \int_{\mathbf{X}} \Xi(vx)\ \rho(dx)\\ &= \int_{\mathbf{X}} \xi(u)\Xi(x)\ \rho(dx)\\ &= 0.\end{aligned}$$

This gives a contradiction. We conclude that the terms which do not solve Vinogradov's mean value problem do not contribute to our sum.

For every $2t$-tuple $j_1, \dots, j_{2t}$ in $[k]^{2t}$, we have that for all p

$$\left|\Phi(g^{pj_1}x) \cdot \dots \cdot \Phi(g^{pj_t}x) \cdot \overline{\Phi}(g^{pj_{t+1}}x) \cdot \dots \cdot \overline{\Phi}(g^{pj_{2t}}x)\right| \leq 1,$$

simply using a trivial L^∞ bound. For every $2t$-tuple which does not solve Vinogradov's mean value problem we have

$$\limsup_{k\in\mathcal{K}} \limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P} \int_X \Phi(g^{pj_1}x) \cdot \dots \cdot \Phi(g^{pj_t}x) \cdot \overline{\Phi}(g^{pj_{t+1}}x) \cdot \dots \cdot \overline{\Phi}(g^{pj_{2t}}x) = 0.$$

Therefore, the average is bounded by the fraction of terms which solve Vinogradov's mean value problem. There are no more than $C_{s,\varepsilon}k^{t+.5\varepsilon}$ such solutions by Bourgain-Demeter-Guth (Theorem 2.2.25). Thus

$$\limsup_{k\in\mathcal{K}} \limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P} \int_X |\mathbb{E}_{h\leq k}\epsilon_h \Phi(g^{ph}x)|^{2t} dx \cdot k^{t-\varepsilon} = 0,$$

in the case $t \neq 2$ and

$$\limsup_{k\in\mathcal{K}} \limsup_{P\to\infty} \mathbb{E}_{P/2<p\leq P} \int_X |\mathbb{E}_{h\leq k}\epsilon_h \Phi(g^{ph}x)|^{2t} dx \cdot k^{t} \leq C,$$

in the case $t = 2$. After all, since diagonal solutions are the only solutions to Vinogradov's mean value problem in the case of two variables and one equation i.e. $j_1 = j_2$, there is no ε loss when $t = 2$. Thus, we obtain Theorem 2.2.24 and in turn Theorem 2.1.8 and Theorem 2.1.9.

2.3 Proof of Theorem 2.1.11

The proof of Theorem 2.1.11 is essentially the proof of Theorem 2.1.8 with a few minor simplifications. As before suppose not. Then as before, we can find a joining such that

$$\left|\int_{X\times Y} f(x)f'(y)\nu(dxdy)\right| > c.$$

As before, we can apply [FH18a] such that

$$\left|\int_{X\times Y} \mathbb{E}^{\mu}[f|\mathcal{Z}](x)f'(y)\nu(dxdy)\right| > c.$$

Unlike before, we do not need to restrict the integral to B. As before, we can average over translates

$$\int_{X\times Y} |\mathbb{E}_{h\leq k}\mathbb{E}^{\mu}[f|\mathcal{Z}](T^h x)f'(T^h y)|\nu(dxdy) > c.$$

As before, we can take an essential supremum over y

$$\int_X \sup_{y\in Y} |\mathbb{E}_{h\leq k}\mathbb{E}^{\mu}[f|\mathcal{Z}](T^h x)f'(T^h y)|\mu(dx) > c.$$

As before, we can apply the entropy decrement argument, for some $P \gg k$, we have

$$\mathbb{E}_{P/2<p\leq P}\int_X \sup_{y\in Y} |\mathbb{E}_{h\leq k}\mathbb{E}^{\mu}[f|\mathcal{Z}](T^{ph} x)f'(T^h y)|\mu(dx) > c.$$

We can use the Cauchy-Schwarz inequality

$$\mathbb{E}_{P/2<p\leq P}\int_X \sup_{y\in Y} |\mathbb{E}_{h\leq k}\mathbb{E}^{\mu}[f|\mathcal{Z}](T^{ph} x)f'(T^h y)|^2\mu(dx) > c^2.$$

This time, would like to replace f' by a sum over words of length k up to ε rounding. In the no-rounding case, we knew that words of f' were words of b. We double check that a similar result holds for words up to constant rounding. In particular, fix k such that there are at most δk words of length k that occur with positive log density up to ε rounding. Thus, we can fix a set Σ of words of length k such that $\#\Sigma \leq \delta k$ and for all n outside a set of 0 log density there exists an ϵ in Σ such that $|b(n+h)-\epsilon_h| \leq \varepsilon$. Translating this to the dynamical setting,

$$\nu\{(x,y)\colon \text{there exists } \epsilon \text{ in } \Sigma \text{ such that } |f'(T^h y)-\epsilon_h| \leq \varepsilon\} = 0.$$

Therefore, we can replace f' by a sum over words as before.

$$\sum_{\epsilon} \mathbb{E}_{P/2<p\leq P} \int_X |\mathbb{E}_{h\leq k} \mathbb{E}^{\mu}[f|\mathcal{Z}](T^{ph}x)\epsilon_h|^2 \mu(dx) > c^2 - 2\varepsilon.$$

Notice that this time, when we replace $f'(T^h y)$ by a word, we incur an error of ε. Now the rest of the argument runs exactly the same as before. In fact, after pigeonholing, any dependence on b completely drops out of the argument.

2.4 Frantzikinakis-Host and dynamical models

[Tao17b] shows that there is a joining $(X_0 \times Y, \nu_0, T, f, f', M, I_m)$ of a dynamical model for a with b where $X_0 = D^{\mathbb{Z}} \times \widehat{\mathbb{Z}}$ is the space of sequences in the unit disk with the product topology, T is the shift map on $D^{\mathbb{Z}}$ and $+1$ on $\widehat{\mathbb{Z}}$, f is the evaluation at 0 map, M is projection onto the second factor and $I_m((x(n))_{n\in\mathbb{Z}}, r) = ((\overline{a(m)}x(mn))_{n\in\mathbb{Z}}, \frac{r}{m})$ whenever r is in $m\widehat{\mathbb{Z}}$. Call μ_0 the pushforward of ν_0 onto X_0. Of course, X_0 factors onto $D^{\mathbb{Z}}$ by projection onto the first factor. Call ρ the pushforward of ν onto $D^{\mathbb{Z}}$. [FH18a] (Proposition 4.2 in that paper) showed that $(D^{\mathbb{Z}}, T, \rho)$ is a factor of a system $(\tilde{X}, \tilde{\rho}, T)$ where $\tilde{X} = (D^{\mathbb{Z}})^{\mathbb{Z}}$, T is the shift map and there exists a natural number d so that if $\mathbb{P}_d$ is the set of primes which are 1 mod d then

$$\int_{\tilde{X}} \prod_{j=-K}^{K} F_j(T^j x)\tilde{\rho}(dx) = \lim_{N\to\infty} \mathbb{E}_{p\in\mathbb{P}_d\cap[N]} \int_{D^{\mathbb{Z}}} \prod_{j=-K}^{K} F_j(T^{pj}x)\rho(dx),$$

where K is any natural number, the functions F_j are any bounded measurable functions depending only on the 0^{th} coordinate and by [FH18a] the limit always exists. We fix such a d. By [FH18a] (see Theorem 4.5 in that paper), each ergodic component of $\tilde{X}$ is isomorphic to a product of a Bernoulli system with an inverse limit of nilsystems. Thus, we get a joining of $\tilde{X}$ with X_0 over their common factor $D^{\mathbb{Z}}$. Call this joining (X, μ, T). We also get a joining of $X_0 \times Y$ and X over their common factor X_0, which we call $(X \times Y, \nu, T)$. Explicitly, this joining is defined as follows. A point in $(X \times Y, \nu, T)$ can be thought of as a triple of points (x_1, x_2, y) with x_1 in $\tilde{X}$ and (x_2, y) in $X_0 \times Y$. Since $X_0 = D^{\mathbb{Z}} \times \widehat{\mathbb{Z}}$, we have that $x_2 = (x_3, r)$ for some x_3 in $D^{\mathbb{Z}}$ and r in $\widehat{\mathbb{Z}}$. The measure is supported on triples where $\pi(x_1) = x_3$ so we

will often forget x_1 and simply write a point in $X \times Y$ as a triple (x, r, y) with x in $\tilde{X}$, r in $\widehat{\mathbb{Z}}$ and y in Y. The measure is given explicitly by the following formula: if K is a natural number, F_j are bounded measurable functions on $D^{\mathbb{Z}}$ depending only on the 0^{th} coordinate, ϕ is a bounded measurable function on $\widehat{\mathbb{Z}}$ and ψ is a bounded measurable function on Y then

$$\int_{X\times Y} \prod_{j=-K}^{K} F_j(T^j x) \cdot \phi(r) \cdot \psi(y) \nu(dxdrdy)$$
$$= \lim_{N\to\infty} \mathbb{E}_{p\in\mathbb{P}_d\cap[N]} \int_{X_0\times Y} \prod_{j=-K}^{K} F_j(T^{pj} x) \cdot \phi(r) \cdot \psi(y) \nu_0(dxdrdy).$$

We will proceed to check that $X \times Y$ has all the desired properties. We define $M\colon X \to \widehat{\mathbb{Z}}$ by taking an element (x, r) with x in $\tilde{X}$ and r in $\widehat{\mathbb{Z}}$ to r. Let x be an element of $\tilde{X}$. We will write $x = (x_n)_{n\in\mathbb{Z}}$ for a sequence of elements x_n in $D^{\mathbb{Z}}$ and write $x_n(k) \in D$ for the k^{th} element of the sequence x_n. Let $\iota_m((x(k))_{k\in\mathbb{Z}}) = \overline{a(m)}(x(mk))_{k\in\mathbb{Z}}$ We define $I_m(x, r) = (\iota_m((x_{nm})_{n\in\mathbb{Z}}), \frac{r}{m})$. Explicitly

$$I_m(x, r) = \left((\overline{a(m)} x_{nm}(km)_{k\in\mathbb{Z}})_{n\in\mathbb{Z}}, \frac{r}{m}\right)$$

whenever r is in $m\widehat{\mathbb{Z}}$. We define $f\colon X \times Y \to \mathbb{C}$ by the formula $f(x, r, y) = x_0(0)$. This is just the pullback of $f\colon X_0 \times Y \to \mathbb{C}$ under the factor map $X \times Y \to X_0 \times Y$. We define $f'\colon X \times Y \to \mathbb{C}$ by pulling back $f'\colon X_0 \times Y \to \mathbb{C}$ under the same factor map i.e. $f'(x, r, y) = y$. Now we check

- $M(T(x, r)) = M(Tx, r + 1) = r + 1 = M(x, r) + 1$
- We have

$$\begin{aligned} I_m \circ T^m(x, r) &= \left((\overline{a(m)} x_{nm+m}(km)_{k\in\mathbb{Z}})_{n\in\mathbb{Z}}, \frac{r+m}{m}\right) \\ &= \left((\overline{a(m)} x_{(n+1)m}(km)_{k\in\mathbb{Z}})_{n\in\mathbb{Z}}, \frac{r}{m} + 1\right) \\ &= T\left((\overline{a(m)} x_{nm}(km)_{k\in\mathbb{Z}})_{n\in\mathbb{Z}}, \frac{r}{m}\right) \\ &= T \circ I_m(x, r). \end{aligned}$$

for any m and whenever r is in $m\widehat{\mathbb{Z}}$.

- Let K be a natural number and $F_j\colon \tilde{X} \to \mathbb{C}$ be a sequence of bounded measurable functions depending only on 0. Let ϕ be a function which is measurable with respect to $\widehat{\mathbb{Z}}$. Then for any m,

$$\int_X \mathbb{1}_{M^{-1}(m\widehat{\mathbb{Z}})}(x)\phi(I_m x)\cdot\prod_{j=-K}^{K} F_j(T^j I_m x)\mu(dx)$$

$$= \lim_{N\to\infty} \mathbb{E}_{p\in\mathbb{P}_d\cap[N]} \int_{X_0} \mathbb{1}_{r\in m\widehat{\mathbb{Z}}}\phi\left(\frac{r}{m}\right)\cdot\prod_{j=-K}^{K} F_j\circ\iota_m(T^{pmj}x)\mu_0(dxdr),$$

by definition of μ. Next, we use that $\iota_m\circ T^{pmj} = T^{pj}\circ\iota_m$.

$$= \lim_{N\to\infty} \mathbb{E}_{p\in\mathbb{P}_d\cap[N]} \int_{X_0} \mathbb{1}_{r\in m\widehat{\mathbb{Z}}}\phi\left(\frac{r}{m}\right)\cdot\prod_{j=-K}^{K} F_j\circ T^{pj}\circ\iota_m(x)\mu_0(dxdr).$$

Because X_0 is a dynamical model for a, it possesses a dilation symmetry,

$$= \lim_{N\to\infty} \mathbb{E}_{p\in\mathbb{P}_d\cap[N]} \int_{X_0} \frac{1}{m}\phi(r)\cdot\prod_{j=-K}^{K} F_j(T^{pj}x)\mu_0(dxdr).$$

Finally, we apply the definition of μ one more time,

$$= \int_X \frac{1}{m}\phi(x)\cdot\prod_{j=-K}^{K} F_j(T^j x)\mu(dx).$$

- For any natural number m and any r in $m\widehat{\mathbb{Z}}$, we have

$$f(I_m(x,r)) = f(\Big((\overline{a(m)}x_{nm}(km)_{k\in\mathbb{Z}})_{n\in\mathbb{Z}}, \frac{r}{m}\Big)) = \overline{a(m)}x_0(0) = \overline{a(m)}f(x,r).$$

- Clearly, for any natural numbers m and h,

$$I_h I_m(x,r) = \left((\overline{a(mh)}x_{nmh}(kmh)_{k\in\mathbb{Z}})_{n\in\mathbb{Z}}, \frac{r}{mh}\right) = I_m I_h(x,r),$$

for any r in $hm\widehat{\mathbb{Z}}$.

- Since f and f' are pulled back from $X_0\times Y$, the "statistics" of $f\colon X\times Y\to\mathbb{C}$ will be the same as the statistics of $f\colon X_0\times Y\to\mathbb{C}$ and similarly for f'.

Therefore, $(X \times Y, \nu, T, f, f', I_m, M)$ is a joining of a dynamical model for a with b.

Let (X, μ_ω, T) be an ergodic component of (X, μ, T) which joins the corresponding ergodic component $(\tilde{X}, \tilde{\rho}_\omega, T)$ of $(\tilde{X}, \tilde{\rho}, T)$ with $\widehat{\mathbb{Z}}$. Note that $\widehat{\mathbb{Z}}$ is already an ergodic inverse limit of nilsystems: after all it is an inverse limit of the ergodic systems of the form $\mathbb{Z}/m\mathbb{Z}$ and the inverse limit of ergodic systems is ergodic. By [FH18a], there is a Bernoulli system (W, dw, T) and an inverse limit of nilsystems (Z_0, dz, T) such that $(\tilde{X}, \tilde{\rho}_\omega, T) \cong (W, dw, T) \times (Z_0, dz, T)$. Therefore (X, μ_ω, T) is isomorphic to $(W \times Z_0 \times \widehat{\mathbb{Z}}, \mu', T)$ where μ' is some mystery measure and where T is just the product transformation. We can think of this system as a joining of $(W \times Z_0, dw \times dz, T)$ with $(\widehat{\mathbb{Z}}, dz, T)$ or we can think of this system as a joining of (W, dw, T) with $(Z_0 \times \widehat{\mathbb{Z}}, \zeta, T)$ where ζ is some unknown measure given by pushing forward μ' onto $Z_0 \times \widehat{\mathbb{Z}}$. Next, we claim that any ergodic joining of two inverse limits of nilsystems is in fact isomorphic to an inverse limit of nilsystems. After all, if Z_1 and Z_2 are two nilsystems and ζ is an ergodic invariant measure on $Z_1 \times Z_2$, then ζ is a translate of Haar measure on some closed subgroup by measure classification for nilsystems. Thus $(Z_1 \times Z_2, \zeta, T) \cong (Z_3, dz, T)$ for some nilsystem Z_3. Taking inverse limits, $(Z_0 \times \widehat{\mathbb{Z}}, \zeta, T)$ is isomorphic to an inverse limit of nilsystems (Z, dz, T). Because (Z, dz, T) is an inverse limit of nilsystems, it has zero entropy so the only possible joining of (Z, dz, T) with the Bernoulli system (W, dw, T) is the trivial joining i.e. μ' is the product measure $dw \times dz$. Lastly, we claim that Z is isomorphic to the Host-Kra factor of (X, μ_ω, T). Since the Host-Kra factor $\mathcal{Z}(X)$ is isomorphic to an inverse limit of nilsystems, it has zero entropy, so any factor map from $W \times Z$ to $\mathcal{Z}(X)$ where W is Bernoulli necessarily factors through Z. Thus Z factors onto $\mathcal{Z}(X)$. Of course, since X factors onto Z, the Host-Kra factor for X factors onto the Host-Kra factor for Z. Implicitly in [HK05] and explicitly, for instance, in [HK18] chapter 12, for any nilsystem (Z_1, dz, T) the Host-Kra factor of Z_1 is Z_1. Thus, taking inverse limits gives that the Host-Kra factor of Z is Z so $\mathcal{Z}(X) \cong Z$. This completes the proof.

2.5 Reduction to the completely multiplicative case

We have stated our main theorems in the case that a is completely multiplicative. In this section, we show that these assumptions can be weakened to include all multiplicative functions. For example, we will show that Theorem 2.1.8 holds in this generality. The same argument works for Theorem 2.1.9 and Theorem 2.1.11 (although in this last case, the way that c depends on ε gets worse). The argument here will be entirely formal, using nothing of the proof of Theorem 2.1.8 and only the result. However, we remark that the interested reader could check that the proof we give can be adapted to the more general case of multiplicative functions. The main difference is that now the dynamical model for a does not satisfy the identity that the push forward of μ restricted to $M^{-1}(m\widehat{\mathbb{Z}})$ is $\frac{1}{m}\mu$ but instead we incur a $\frac{1}{m}$ error i.e. for all ϕ in satisfying $||\phi||_{L^\infty(\mu)} \leq 1$ we have

$$\int_X \phi(x)\mu(dx) = \int_X m\mathbb{1}_{x\in M^{-1}(m\widehat{\mathbb{Z}})}\phi(I_m(x))\mu(dx) + O\left(\frac{1}{m}\right).$$

This introduces an error term of size $O\left(\frac{1}{P}\right)$ in Corollary 2.2.17 which tends to 0 as P tends to infinity.

However, here we proceed just using the statement of Theorem 2.1.8. Famously, we can write

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

where λ is Liouville function and μ is the Möbius function which agree with the Liouville function on squarefree numbers and vanishes on numbers which are not squarefree. Of course,

$$\lambda\left(\frac{n}{d^2}\right) = \lambda(n),$$

but we write it this way to suggest that the convolution identity

$$\mu = \lambda * \phi$$

where $\phi(d^2) = \mu(d)$ may be generalized. In fact, for any multiplicative function a taking values on the unit circle, we may write

$$a = a_1 * a_2$$

where a_1 is some completely multiplicative function taking values on the unit circle and a_2 is a (possibly unbounded) multiplicative function supported on numbers of the form d^k for some natural numbers d and k with $k \geq 2$. To prove this is possible, it suffices to check it is possible on prime powers since both sides are multiplicative. For any prime p, we define

$$a_1(p) = a(p)$$

and so

$$a_1(p) \cdot a_2(1) + a_1(1) \cdot a_2(p) = a(p) \cdot 1 + 1 \cdot 0 = a(p).$$

We also want,

$$a(p^2) = a_1(p^2) \cdot 1 + 1 \cdot a_2(p^2)$$

so we choose

$$\begin{aligned} a_2(p^2) =& a(p^2) - a_1(p^2) \\ =& a(p^2) - a(p)^2. \end{aligned}$$

Iteratively, we may define

$$\begin{aligned} a_2(p^k) =& a(p^k) - \sum_{0 \leq i < k} a_1(p^{k-i}) a_2(p^i) \\ =& a(p^k) - \sum_{0 \leq i < k} a(p)^{k-i} a_2(p^i). \end{aligned}$$

Since whether a is unpretentious or not depends only on the behavior of a at primes, clearly if a is unpretentious then so is a_1.

Informally, the probability that a random number is divisible by d is roughly $\frac{1}{d}$. Thus, the expected number of times that any number of the form d^k for $k \geq 2$ divides a random natural number is at most

$$\sum_{d \geq 2} \sum_{k \geq 2} \frac{1}{d^k}$$

which is summable. Thus the tails

$$\sum_{d\geq C}\sum_{k\geq 2}\frac{1}{d^k}$$

and

$$\sum_{d\geq 2}\sum_{k\geq C}\frac{1}{d^k}$$

tend to zero as C tends to infinity. Let S be the set of natural numbers n for which d^k divides n for $d, k \geq 2$ implies $d, k \leq C$. The previous analysis says most numbers are in S. Fix a function b as in the statement of Theorem 2.1.8, that is a bounded function such that for any $\delta > 0$ there are infinitely many k such that the number of words of b of length k that occur with positive upper logarithmic density is at most δk^2. Our goal will be to show that for N large,

$$|\mathbb{E}^{\log}_{n\leq N} a(n)b(n)|$$

is small, say less than a constant times some small positive number ε. If C is sufficiently large depending on ε but still very small compared to N, we may modify a on the set of numbers outside S. In particular, a is given by the formula

$$a(n) = \sum_{\ell|n} a_1\left(\frac{n}{\ell}\right) a_2(\ell)$$

For most numbers, this is the same as

$$= \sum_{\substack{\ell|n \\ \ell\leq C^C}} a_1\left(\frac{n}{\ell}\right) a_2(\ell).$$

That formula works as long as n is not divisible by a number of the form d^k where either d or k is greater than C. In that exceptional case when n is divisble by a number of the form d^k with d or k greater than C, we can write n as $i \cdot j$ where i is not divisible by any number greater than C^C and is as large as possible given that constraint. We conclude that if $\ell \leq C^C$ and $\ell | n$ then $\ell | i$. Thus, expanding the definitions and using multiplicativity,

$$\begin{aligned} a_1(j) \cdot a(i) =& a_1(j) \sum_{\substack{\ell | i \\ \ell \leq C^C}} a_1\left(\frac{i}{\ell}\right) a_2(\ell) \\ &= \sum_{\substack{\ell | i \\ \ell \leq C^C}} a_1\left(\frac{ij}{\ell}\right) a_2(\ell) \\ &= \sum_{\substack{\ell | n \\ \ell \leq C^C}} a_1\left(\frac{n}{\ell}\right) a_2(\ell) \end{aligned}$$

We conclude that the formula

$$\sum_{\ell | n} a_1\left(\frac{n}{\ell}\right) a_2(\ell)$$

is bounded and agrees with $a(n)$ all but at most

$$\sum_{d \geq C} \sum_{k \geq 2} \frac{1}{d^k} + \sum_{d \geq 2} \sum_{k \geq C} \frac{1}{d^k}$$

of the time. Thus, it suffices to show

$$\left| \mathbb{E}^{\log}_{n \leq N} \sum_{\substack{m\ell = n \\ \ell \leq C^C}} a_1(m) a_2(\ell) b(n) \right|$$

is small. By changing variables and applying Fubini,

$$= \left| \sum_{\ell \leq C^C} a_2(\ell) \mathbb{E}^{\log}_{m \leq N/\ell} a_1(m) b(m\ell) \right|.$$

Fix a natural number ℓ. Notice that every word of length k of the function $m \mapsto b(m\ell)$ embeds in a word of b of length $k \cdot \ell$. Thus, it is easy to check that $m \mapsto b(m\ell)$ still satisfies

the conditions of Theorem 2.1.8. Therefore, as N tends to infinity, the previous expression tends to 0.

CHAPTER 3

A Dynamical Proof of the Prime Number Theorem

3.1 Introduction to Chapter 2

The prime number theorem states that

$$\pi(N) = (1 + o_{N\to\infty}(1))\frac{N}{\log N},$$

where $\pi(N)$ denotes the number of primes of size at most N. In some sense, the result was first publicly conjectured by Legendre in 1798 who suggested that

$$\pi(N) = \frac{N}{A\log N + B + o_{N\to\infty}(1)},$$

for some constants A and B. Legendre specifically conjectured $A = 1$ and $B = -1.08366$. Gauss conjectured the same formula and stated he was not sure what the constant B might turn out to be. Gauss' conjecture was based on millions of painstaking calculations first obtained in 1792 and 1793 which were never published but nonetheless predate Legendre's work on the subject. It is worth noting that later in his 1849 letter to Encke Gauss conjectured that $\pi(N) \approx \mathrm{Li}(N)$, which in particular implies the correct values for A and B. The first major breakthrough on the problem was due to Chebyshev who showed that

$$c + o_{N\to\infty}(1) \leq \frac{\pi(N)\log N}{N} \leq C + o_{N\to\infty}(1)$$

for some explicit constants c and C with $c > 0$. There is a long history of improvements to these explicit constants for which we refer to Goldstein [Gol73] and Goldfeld [Gol04]. The prime number theorem was important motivation for Riemann's seminal work on the zeta function.

The first proofs of the prime number theorem were given independently by Hadamard and de la Vallée Poussin in 1896. The key step in their proof is a difficult argument showing that the Riemann zeta function does not have a zero on the line $\mathrm{Re}(z) = 1$. Their proof was later substantially simplified by many mathematicians. In 1930, Wiener found a "Fourier analytic" proof of the prime number theorem. In 1949, Erdös [Erd49] and Selberg [Sel50] discovered an elementary proof of the prime number theorem, where here elementary is used in the technical sense that the proof involves no complex analysis and does not necessarily mean that the proof is easy reading. The bitter battle over credit for this result is the subject of an informative note by Goldfeld [Gol04]. Other proofs are due to Daboussi [Dab89] and Hildebrand [Hil86b]. In a blog post from 2014, Tao proves the prime number theorem using the theory of Banach algebras [Taoc]. A published version of this theorem can be found in a book by Einsiedler and Ward [EW17]. In an unpublished book from 2014, Granville and Soundarajan prove the prime number theorem using pretentious methods (see, for instance, [GHS19]). A note by Zagier [Zag97] from 1997 contains perhaps the quickest proof of the prime number theorem using a tauberian argument in the spirit of the Erdös-Selberg proof combined with complex analysis in the form of Cauchy's theorem. Zagier attributes this proof to Newman.

The goal of this note is to present a new proof of the prime number theorem. Florian Richter and I discovered similar proofs concurrently and independently. His proof can be found in [Ric]. Terence Tao wrote up a version of this argument on his blog following personal communication from the author which can be found in [Taoa].

The proof proceeds as follows. To prove the prime number theorem, it suffices to prove that

$$\frac{1}{N}\sum_{n\leq N}\Lambda(n) = 1 + o(1),$$

where $\Lambda(n)$ is the von Mangoldt function which is $\log p$ if n is a power of a prime p and 0 otherwise. The reader may think of Λ as the normalized indicator function of the primes.

The von Mangoldt function is related to the Möbius function via the formula

$$\Lambda = \mu * \log,$$

where the Möbius function $\mu(n)$ is 0 if n has a repeated factor, -1 if n has an odd number of distinct prime factors, $+1$ if n has an even number of distinct prime factors. This formula, sometimes called the Möbius inversion formula, encodes the fundamental theorem of arithmetic. Thus, there is a dictionary between properties of the von Mangoldt function Λ and the Möbius function μ. Landau observed that cancellation in the Möbius function is equivalent to the prime number theorem i.e. the prime number theorem is equivalent to the statement

$$\frac{1}{N}\sum_{n\leq N}\mu(n) = o_{N\to\infty}(1).$$

This is what we actually try to prove.

The next observation is that, if one wants to compute a sum, it suffices to sample only a small number of terms. Typically (for instance for an i.i.d. randomly chosen sequence) the average value

$$\frac{1}{N}\sum_{n\leq N} a(n)$$

is approximately the same as the average over only the even terms

$$\approx \frac{2}{N}\sum_{n\leq N} a(n)\mathbb{1}_{2|n}.$$

However, for certain sequences, like $a = (-1, +1, -1, +1, \ldots)$, the averages do not agree. Still for this sequence, if we instead sample every third point or every fifth point or every p^{th} point for any other prime then the averages are approximately equal. It turns out, this is a rather general phenomenon: for any sequence, for most primes p, the average of the sequence is the same as the average along only those numbers divisible by p.

Applying this to the Möbius function, for each N, for most primes p

$$\frac{1}{N}\sum_{n\leq N}\mu(n) \approx \frac{1}{N}\sum_{n\leq N}\mu(n)p\mathbb{1}_{p|n}.$$

For the purposes of this introduction, we will "cheat" and pretend that this equation is true for any prime p. By changing variables

$$\frac{1}{N}\sum_{n\leq N}\mu(n)p\mathbb{1}_{p|n} = \frac{p}{N}\sum_{n\leq N/p}\mu(pn).$$

But $\mu(pn) = -\mu(n)$ for most numbers n since μ is multiplicative. Combining the last two equations gives

$$\frac{1}{N}\sum_{n\leq N}\mu(n) \approx -\frac{p}{N}\sum_{n\leq N/p}\mu(n).$$

The plan is to use this identity three times. Suppose we can find primes p_1, p_2 and p such that $\frac{p_1p_2}{p} \approx 1$. Then by applying the previous identity

$$\frac{1}{N}\sum_{n\leq N}\mu(n) \approx -\frac{p}{N}\sum_{n\leq N/p}\mu(n)$$

and also

$$\begin{aligned}\frac{1}{N}\sum_{n\leq N}\mu(n) &\approx -\frac{p_1}{N}\sum_{n\leq N/p_1}\mu(n)\\ &\approx +\frac{p_1p_2}{N}\sum_{n\leq N/p_1p_2}\mu(n).\end{aligned}$$

But since $\frac{p_1p_2}{p} \approx 1$, we know that

$$\frac{p_1p_2}{N}\sum_{n\leq N/p_1p_2}\mu(n) \approx \frac{p}{N}\sum_{n\leq N/p}\mu(n).$$

Putting everything together we conclude that

$$\frac{1}{N}\sum_{n\leq N}\mu(n) \approx -\frac{1}{N}\sum_{n\leq N}\mu(n)$$

which implies

$$\frac{1}{N}\sum_{n\leq N}\mu(n) \approx 0.$$

This implies the prime number theorem.

Thus, the main difficulty in the proof is finding primes p, p_1 and p_2 lying outside some exceptional set for which $\frac{p_1 p_2}{p} \approx 1$. We give a quick sketch of the argument. The Selberg symmetry formula (Theorem 3.2.5) roughly tells us that, even if we do not know how many primes there are at a certain scale (say in the interval from x to $x(1+\varepsilon)$) and we do not know how many semiprimes (products of two primes) there are at that scale, the weighted sum of the number of primes and semiprimes is as we would expect. In particular, if there are no semiprimes between x and $x(1+\varepsilon)$ there are twice as many primes as one would expect (meaning $2 \cdot \varepsilon \frac{x}{\log x}$ many primes). Let x be a large number. If there are both primes and semiprimes between x and $x(1+\varepsilon)$ then we can find p, p_1 and p_2 such that $\frac{p_1 p_2}{p} \approx 1 + O(\varepsilon)$ and we are done. Thus, assume that there are either only primes or only semiprimes in the interval $[x, x(1+\varepsilon)]$. For the sake of our exposition, we will assume there are only primes between x and $x(1+\varepsilon)$. By the Selberg symmetry formula, there are twice as many primes in this interval as expected. Now if there is a semiprime $p_1 p_2$ in the interval $[x(1+\varepsilon), x(1+\varepsilon)^2]$ then picking any prime p in the interval $[x, x(1+\varepsilon)]$ we conclude that there exists p, p_1 and p_2 such that $\frac{p_1 p_2}{p} \approx 1 + O(\varepsilon)$. Thus, either we win (and the prime number theorem is true) or there are again twice as many primes in the interval $[x(1+\varepsilon), x(1+\varepsilon)^2]$ as one would expect. Running this argument again shows that there are again only primes and no semiprimes in the interval $[x(1+\varepsilon)^2, x(1+\varepsilon)^3]$. Iterating this argument using the connectedness of the interval, we find large intervals $[x, 100x]$ where there are twice as many primes as predicted by the prime number theorem. But this contradicts Chebyshev's theorem: Chebyshev's theorem gives a lower bound on the number of primes, which in turn gives a lower bound on the number of semiprimes; alternately, we remark that one could use Erdös's version of Chebyshev's theorem that the number of primes less than x is at most $\log 4 \frac{x}{\log x}$ and because $\log 4 < 2$ this gives a contradiction. This completes the proof.

3.1.1 A comment on notation

Throughout this chapter, we will use asymptotic notation. Since number theory, dynamics and analysis sometime use different conventions, we take a moment here to fix notation. We will write

$$x = O(y)$$

to mean that there exists a constant C such that

$$|x| \leq Cy.$$

When we adorn these symbols with subscripts, the subscripts specify which variables the constants are allowed to depend on. Thus

$$x = O_{A,B}(y)$$

means that there exists a constant C which is allowed to depend on A and B such that

$$|x| \leq Cy.$$

We write

$$x = y + O(z)$$

to mean that

$$x - y = O(z).$$

We also adopt little o notation:

$$x = o_{n\to\infty}(y)$$

means that

$$\lim_{n\to\infty} \frac{x}{y} = 0.$$

Occasionally, when the variable with respect to which the limit is being taken is clear from context, we may simply write

$$x = o(y).$$

As before, we write

$$x = y + o_{n\to\infty}(z)$$

to mean

$$x - y = o_{n\to\infty}(z).$$

If the expression x depends on more than one variables, say n, m and k, we may use subscripts to make explicit that the rate of convergence implicit in the little o notation is allowed to depend on more variables. Thus,

$$x = o_{n\to\infty,m,k}(y)$$

means that $\frac{x}{y}$ tends to zero with n at a rate which may depend on m and k.

3.1.2 Acknowledgments

I would like to thank Florian Richter for his patience and conversation. I would also like to thank Terence Tao for including this proof in his class on number theory and for many helpful discussions. I would like to thank Gergley Harcos and Joni Teräväinen for comments on an earlier draft. I would also like to thank Tim Austin, Will Baker, Bjorn Bringmann, Asgar Jamneshan, Gyu Eun Lee, Adam Lott, Clark Lyons, Bar Roytman, Chris Shriver and Will Swartworth for helpful conversations.

3.2 Proof of the prime number theorem

From number theory, we will use Mertens' Theorem, in particular the version which states,

$$\sum_{p\leq x} \frac{1}{p} = \log\log x + M + O\left(\frac{1}{\log x}\right)$$

for some constant M; we will also use Chebyshev's Theorem, the Selberg Symmetry Formula, Landau's formulation of the prime number theorem (i.e. that the prime number theorem is

equivalent to $\sum_{n\leq N}\mu(n) = o_{N\to\infty}(N)$) and a slightly modified version of the Turán-Kubilius inequality which we will prove using the following Bombieri-Halász-Montgomery inequality.

Proposition 3.2.1 (Bombieri-Halász-Montgomery inequality [Bom71])**.** *Let w_i be a sequence of nonnegative real numbers. Let u and v_i be vectors in a Hilbert space. Then*

$$\sum_{i=1}^{n} w_i|\langle u, v_i\rangle|^2 \leq ||u||^2 \cdot \left(\sup_i \sum_{j=1}^{n} w_j|\langle v_i, v_j\rangle|\right).$$

Proof. By duality, there exists c_i such that

$$\sum_{i=1}^{n} w_i|c_i|^2 = 1$$

and

$$\sum_{i=1}^{n} w_i|\langle u, v_i\rangle|^2 = \left(\sum_{i=1}^{n} w_i c_i\langle u, v_i\rangle\right)^2$$

and therefore by conjugate bilinearity of the inner product

$$= \left\langle u, \sum_{i=1}^{n} w_i\overline{c}_i v_i\right\rangle^2.$$

By Cauchy-Schwarz, this is at most

$$\leq ||u||^2\left|\left|\sum_{i=1}^{n} w_i\overline{c}_i v_i\right|\right|^2.$$

By the pythagorean theorem this is given by

$$= ||u||^2 \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j c_i\overline{c}_j\langle v_i, v_j\rangle.$$

The geometric mean is dominated by the arithmetic mean.

$$\leq ||u||^2 \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \frac{1}{2}(|c_i| + |c_j|)|\langle v_i, v_j\rangle|.$$

By symmetry this is

$$= ||u||^2 \sum_{i=1}^{n} w_i|c_i|^2 \sum_{j=1}^{n} w_j|\langle v_i, v_j\rangle|.$$

Because everything is nonnegative, we may replace the inner term with a supremum

$$\leq ||u||^2 \sum_{i=1}^{n} w_i |c_i|^2 \sup_k \sum_{j=1}^{n} w_j |\langle v_k, v_j \rangle|.$$

Using that $\sum w_i |c_i|^2 = 1$ completes the proof. □

The next proposition applies the previous proposition in order to show that, for any bounded sequence, the average of the sequence is the same as the average over the p^{th} terms in the sequence for most prime p.

Proposition 3.2.2 (Turán-Kubilius [Kub64])**.** *Let S denote a set of primes less than some natural number P. Let N be a natural number which is at least P^3. Let f be a 1-bounded function from $\mathbb{N}$ to $\mathbb{C}$. Then*

$$\sum_{p \in S} \frac{1}{p} \left| \frac{1}{N} \sum_{n \leq N} f(n)(1 - p\mathbb{1}_{p|n}) \right|^2 = O(1).$$

Proof. We will apply Proposition 3.2.1: our Hilbert space is L^2 on the space of function on the integers $\{1, \ldots, N\}$ equipped with normalized counting measure; set $w_p = \frac{1}{p}$; set $v_p = (n \mapsto 1 - p\mathbb{1}_{p|n})$ and $u = f$; thus, by Proposition 3.2.1

$$\begin{aligned} &\sum_{p \in S} \frac{1}{p} \left| \frac{1}{N} \sum_{n \leq N} f(n)(1 - p\mathbb{1}_{p|n}) \right|^2 \\ \leq &\frac{1}{N} \sum_{n \leq N} |f(n)|^2 \cdot \sup_{p \in S} \sum_{q \in S} \frac{1}{q} \left| \frac{1}{N} \sum_{n \leq N} (1 - p\mathbb{1}_{p|n})(1 - q\mathbb{1}_{q|n}) \right|. \end{aligned}$$

Since f is 1-bounded, we may bound the L^2 norm of f by 1. Thus,

$$\leq \sup_{p \in S} \sum_{q \in S} \frac{1}{q} \left| \frac{1}{N} \sum_{n \leq N} (1 - p\mathbb{1}_{p|n})(1 - q\mathbb{1}_{q|n}) \right|. \tag{3.1}$$

For primes p and q,

$$\frac{1}{N} \sum_{n \leq N} (1 - p\mathbb{1}_{p|n})(1 - q\mathbb{1}_{q|n})$$

can be expanded into a signed sum of four terms

$$\frac{1}{N}\sum_{n\leq N} 1 - p\mathbb{1}_{p|n} - q\mathbb{1}_{q|n} + pq\mathbb{1}_{p|n}\mathbb{1}_{q|n}.$$

When $p \neq q$, we claim that each term is $1 + O\left(\frac{P^2}{N}\right)$. The trickiest term is the last term

$$\frac{1}{N}\sum_{n\leq N} pq\mathbb{1}_{p|n}\mathbb{1}_{q|n}.$$

When $p \neq q$, we have that

$$\mathbb{1}_{p|n}\mathbb{1}_{q|n} = \mathbb{1}_{pq|n}.$$

Of course, for any natural number m,

$$\#\text{ of } n \leq N \text{ such that } m \text{ divides } n \ = \frac{N}{m} + O(1),$$

where the $O(1)$ term comes from the fact that m need not perfectly divide N. Thus,

$$\frac{1}{N}\sum_{n\leq N} pq\mathbb{1}_{p|n}\mathbb{1}_{q|n} = pq\left(\frac{1}{pq} + O\left(\frac{1}{N}\right)\right),$$

which is $1 + O\left(\frac{P^2}{N}\right)$ as claimed. A similar argument handles the three other terms. Altogether, we conclude that

$$\frac{1}{N}\sum_{n\leq N}(1 - p\mathbb{1}_{p|n})(1 - q\mathbb{1}_{q|n}) = O\left(\frac{P^2}{N}\right),$$

when $p \neq q$. Inserting this bound into 3.1 and remembering that there are at most P terms in the sum over q in S, we find

$$\begin{aligned}
&\sum_{p\in S}\frac{1}{p}\left|\frac{1}{N}\sum_{n\leq N} f(n)(1 - p\mathbb{1}_{p|n})\right|^2 \\
\leq &\sup_{p\in S}\sum_{q\in S}\frac{1}{q}\left|\frac{1}{N}\sum_{n\leq N}(1 - p\mathbb{1}_{p|n})(1 - q\mathbb{1}_{q|n})\right| \\
\leq &\sup_{p\in S}\frac{1}{p}\left|\frac{1}{N}\sum_{n\leq N}(1 - p\mathbb{1}_{p|n})(1 - p\mathbb{1}_{p|n})\right| + O\left(\frac{P^3}{N}\right)
\end{aligned}$$

Expanding out the product, the main term is

$$\sup_{p\in S}\frac{1}{p}\left|\frac{1}{N}\sum_{n\leq N}p^2\mathbb{1}_{p|n}\right|.$$

By the same trick as before, we may replace the average of $\mathbb{1}_{p|n}$ by $\frac{1}{p}$ plus a small error dominated by the main term. Cancelling factors of p as appropriate, we are left with

$$=\sup_{p\in S}\frac{1}{p}\left|\frac{1}{N}\sum_{n\leq N}p^2\frac{1}{p}\right|=O(1).$$

Of course, all the smaller terms can be bounded by the triangle inequality. This completes the proof. □

Note that,

$$\sum_{p\in S}\frac{1}{p}\frac{1}{N}\sum_{n\leq N}1=\sum_{p\in S}\frac{1}{p}.$$

For instance, if S is the set of all primes less than P, Euler proved that

$$\sum_{p\leq P}\frac{1}{p}\to\infty$$

as P tends to infinity. In fact, Mertens' theorem states that this sum is approximately $\log\log P$. Thus, Proposition 3.2.2 represents a real improvement over the trivial bound. Therefore, for S, P, N and f as in the statement of Proposition 3.2.2

$$\left|\frac{1}{N}\sum_{n\leq N}f(n)(1-p\mathbb{1}_{p|n})\right|^2$$

is small for "most" primes. This shows that most primes are "good" in the sense that

$$\frac{1}{N}\sum_{n\leq N}f(n)\approx\frac{1}{N}\sum_{n\leq N}f(n)p\mathbb{1}_{p|n}$$

This notion is captured in the following definition.

Definition 3.2.3. *Let ε be a positive real number, let P be a natural number which is sufficiently large depending on ε and let N be a natural number sufficiently large depending*

on P. Denote by $\ell(N)$ the quantity

$$\ell(N) = \sum_{n \leq N} \frac{1}{n}.$$

Denote by $S(N)$ the set of primes $p \leq P$ such that

$$\frac{1}{N} \left| \sum_{n \leq N} \mu(n) - \sum_{n \leq N} \mu(n) p \mathbb{1}_{p|n} \right| \geq \varepsilon.$$

Then we say a prime p is good if

$$\frac{1}{\ell(N)} \sum_{n \leq N} \frac{1}{n} \mathbb{1}_{p \in S(n)} \leq \varepsilon.$$

Otherwise, we say p is bad.

From Proposition 3.2.2, we obtain the following corollary.

Corollary 3.2.4. *Let ε be a positive real number, let P be a natural number which is sufficiently large depending on ε and let N be a natural number sufficiently large depending on P. Then the set of bad primes is small in the sense that*

$$\sum_{p \text{ bad} \leq P} \frac{1}{p} = O(\varepsilon^{-3}).$$

Proof. By Proposition 3.2.2, for each n sufficiently large,

$$\sum_{p \leq P} \frac{1}{p} \mathbb{1}_{p \notin S(n)} = O(\varepsilon^{-2}).$$

Summing in n gives,

$$\sum_{p \leq P} \frac{1}{p} \frac{1}{\ell(N)} \sum_{n \leq N} \frac{1}{n} \mathbb{1}_{p \notin S(n)} = O(\varepsilon^{-2}) + o_{N \to \infty, P}(1).$$

We remark that for N sufficiently large depending on P, this second error term may be absorbed into the first term. By definition, the set of bad primes is the set of primes such that

$$\frac{1}{\ell(N)} \sum_{n \leq N} \frac{1}{n} \mathbb{1}_{p \notin S(n)} \geq \varepsilon.$$

But then by Chebyshev's inequality (i.e. not his theorem on counting primes),

$$\sum_{p \text{ bad } \leq P} \frac{1}{p} = O(\varepsilon^{-3}).$$

as desired. □

Next, we turn to the Selberg symmetry formula. To state Selberg's symmetry formula, we need to introduce the following function. Let $\Lambda_2 = \log \cdot \Lambda + \Lambda * \Lambda$ i.e.

$$\Lambda_2(n) = \log(n)\Lambda(n) + \sum_{d|n} \Lambda(d)\Lambda\left(\frac{n}{d}\right),$$

where the von Mangoldt function $\Lambda(n)$ when $\log p$ is n is a power of a prime p and 0 otherwise. Thus, we remark that Λ_2 is supported on prime powers and products of two prime powers. It is not too hard to show that Λ_2 is "mostly" supported on primes and semiprimes. Recall that the prime number theorem is the statement that

$$\frac{1}{N} \sum_{n \leq N} \Lambda(n) = 1 + o_{N \to \infty}(1)$$

and thus

$$\frac{1}{N} \sum_{n \leq N} \Lambda(n) \log n = \log(N)(1 + o_{N \to \infty}(1)).$$

We are now ready to state the Selberg symmetry formula.

Theorem 3.2.5 (Selberg symmetry formula)**.** *The average of the second von Mangoldt function defined above is*

$$\frac{1}{N} \sum_{n \leq N} \Lambda_2(n) = 2 \log N(1 + o_{N \to \infty}(1)).$$

We will refer the reader to, for instance, [Taoc] section 1 for the proof. The next proposition says that, at each scale, there are either many primes or many semiprimes.

Proposition 3.2.6. *Let $\varepsilon > 0$ be a sufficiently small number. Suppose that k_0 is sufficiently large depending on ε and let I_k denote the interval $[(1+\varepsilon)^k, (1+\varepsilon)^{k+1}]$. Then for every $k \geq k_0$,*

$$\sum_{p \in I_k} \frac{1}{p} \geq \frac{1}{k}$$

or

$$\sum_{\substack{p_1p_2\in I_k\\ p_i\geq \exp(\varepsilon^3 k)}} \frac{1}{p_1p_2} \geq \frac{1}{k}.$$

Proof. This follows from the Selberg symmetry formula (Theorem 3.2.5): after all, by the Selberg symmetry formula, for k_0 sufficiently large, for all $k \geq k_0$,

$$\frac{1}{(1+\varepsilon)^k} \sum_{n\leq (1+\varepsilon)^k} \Lambda_2(n) = 2\log(1+\varepsilon)^k(1+O(\varepsilon^2)).$$

The same holds for k replaced by $k+1$.

$$\frac{1}{(1+\varepsilon)^{k+1}} \sum_{n\leq (1+\varepsilon)^{k+1}} \Lambda_2(n) = 2\log(1+\varepsilon)^{k+1}(1+O(\varepsilon^2)).$$

Taking differences, and using that $k\log(1+\varepsilon) = (k+1)\log(1+\varepsilon)(1+O(\varepsilon^2))$, for $k \geq k_0$ sufficiently large, we find that

$$\frac{1}{\varepsilon(1+\varepsilon)^k} \sum_{n\in I_k} \Lambda_2(n) = 2\log(1+\varepsilon)^k(1+O(\varepsilon)). \tag{3.2}$$

We aim to show that prime powers do not contribute very much to this sum. Notice that, if a prime power contributes to the sum, then the corresponding prime must be at most the square root of $(1+\varepsilon)^{k+1}$ and there is at most one power of any prime in the interval I_k (because $\varepsilon < 1$). Also, notice that $\Lambda_2(p^a) \leq 2\Lambda(p^a)\log p^a$. Thus, we bound

$$\begin{aligned}\frac{1}{\varepsilon(1+\varepsilon)^k} \sum_{\substack{n=p^a,a>1\\ n\in I_k}} \Lambda(n)\log n &= \frac{1}{\varepsilon(1+\varepsilon)^k} \sum_{\substack{n=p^a,a>1\\ n\in I_k}} \log p \log p^a \\ &\leq \frac{1}{\varepsilon(1+\varepsilon)^k} \sum_{p\leq (1+\varepsilon)^{(k+1)/2}} \log p \log(1+\varepsilon)^{k+1}.\end{aligned}$$

Now the number of primes less than $(1+\varepsilon)^{(k+1)/2}$ is certainly less than $(1+\varepsilon)^{(k+1)/2}$, so

$$\begin{aligned}&\leq \frac{1}{\varepsilon(1+\varepsilon)^k}(1+\varepsilon)^{(k+1)/2}\log(1+\varepsilon)^{(k+1)/2}\log(1+\varepsilon)^{k+1}. \\ &= o_{k\to\infty,\varepsilon}(1).\end{aligned}$$

For instance, by choosing k_0 large depending on ε, we can make this quantity

$$=O(\varepsilon)$$

Similarly for products of a natural number m and a prime power,

$$\begin{aligned}\frac{1}{\varepsilon(1+\varepsilon)^k}\sum_{\substack{n=p^a m, a>1\\ n\in I_k}}\Lambda(p)\Lambda(m) \leq & \frac{1}{\varepsilon(1+\varepsilon)^k}\sum_{\substack{n=p^a m, a>1\\ n\in I_k}}\log p\log m\\ \leq & \frac{1}{\varepsilon(1+\varepsilon)^k}\sum_{p\leq(1+\varepsilon)^{(k+1)/2}}\log p\sum_{1<a\leq\log_p(1+\varepsilon)^k}\sum_{mp^a\in I_k}\Lambda(m).\end{aligned}$$

Now the inner most sum is bounded by Chebyshev's inequality. We simplify slightly using the factor of $\frac{1}{\varepsilon(1+\varepsilon)^k}$ out front.

$$\begin{aligned}\leq & C\sum_{p\leq(1+\varepsilon)^{(k+1)/2}}\log p\sum_{1<a\leq\log_p(1+\varepsilon)^k}\frac{1}{p^a}.\\ \leq & C\sum_{p\leq(1+\varepsilon)^{(k+1)/2}}\frac{\log p}{p^2}.\\ = & O(1).\end{aligned}$$

Finally, we claim that when one of the prime factors of a semiprime is less than $\exp(\varepsilon^3 k)$ then that semiprime does not contribute very much to the sum. Indeed,

$$\frac{1}{\varepsilon(1+\varepsilon)^k}\sum_{\substack{p_1p_2\in I_k\\ p_1\leq\exp(\varepsilon^3 k)}}\Lambda(p_1)\Lambda(p_2)=\frac{\varepsilon}{(1+\varepsilon)^k}\sum_{\substack{p_1p_2\in I_k\\ p_1\leq\exp(\varepsilon^3 k)}}\log p_1\log p_2.$$

Now we use that p_1 is at most $\exp(\varepsilon^3 k)$ and p_2 is at most $(1+\varepsilon)^{k+1}$.

$$\leq\frac{1}{\varepsilon(1+\varepsilon)^k}\sum_{\substack{p_1p_2\in I_k\\ p_1\leq\exp(\varepsilon^3 k)}}\varepsilon^3 k\cdot\log(1+\varepsilon)^{k+1}.$$

Summing over scales,

$$\leq\frac{1}{\varepsilon(1+\varepsilon)^k}\varepsilon^3 k\cdot\log(1+\varepsilon)^{k+1}\sum_{m\leq k\varepsilon^3}\sum_{\substack{m\leq\log p_1\leq m-1\\ p_1p_2\in I_k}}1.$$

The number of terms in the inner sum is can be estimated using Chebyshev's theorem. The outersum has roughly $k\varepsilon^3$ many terms. Thus, for some constant C,

$$\leq C\frac{1}{\varepsilon(1+\varepsilon)^k}\cdot\varepsilon^3 k\log(1+\varepsilon)^{k+1}\cdot\varepsilon^3 k\cdot\frac{\exp(\varepsilon^3 k)}{\varepsilon^3 k}\frac{(1+\varepsilon)^{k+1}}{\log(1+\varepsilon)^{k+1}}$$

Simplifying, this is

$$=O(\varepsilon^2 k)$$
$$=O(\varepsilon\cdot\log(1+\varepsilon)^k).$$

Altogether, we find that we can restrict 3.2 to primes and semiprimes where neither factor is too small.

$$\frac{1}{\varepsilon(1+\varepsilon)^k}\sum_{\substack{n\in I_k\\ n=p \text{ or } n=p_1p_2\\ p_i\geq\exp(\varepsilon^3 k)}}\Lambda_2(n)=2\log(1+\varepsilon)^k(1+O(\varepsilon)).$$

For any two numbers n and m in I_k, $\frac{1}{n}=\frac{1}{m}\cdot(1+O(\varepsilon))$, so

$$\varepsilon^{-1}\sum_{\substack{n\in I_k\\ n=p \text{ or } n=p_1p_2\\ p_i\geq\exp(\varepsilon^3 k)}}\frac{\Lambda_2(n)}{n}=2\log(1+\varepsilon)^k(1+O(\varepsilon)).$$

By the pigeonhole principle, either

$$\varepsilon^{-1}\sum_{p\in I_k}\frac{\Lambda_2(p)}{p}\geq\log(1+\varepsilon)^k(1+O(\varepsilon))$$

or

$$\varepsilon^{-1}\sum_{\substack{p_1p_2\in I_k\\ p_i\geq\exp(\varepsilon^3 k)}}\frac{\Lambda_2(p_1p_2)}{p_1p_2}\geq\log(1+\varepsilon)^k(1+O(\varepsilon)).$$

In the first case, moving the ε and $\log p\approx\log(1+\varepsilon)^k$ terms to the other side

$$\sum_{p\in I_k}\frac{1}{p}\geq\frac{\varepsilon}{k\log(1+\varepsilon)}\cdot(1+O(\varepsilon)).$$

Taylor explanding the logarithm gives

$$\geq \frac{1}{k} \cdot (1 + O(\varepsilon)),$$

as desired. In the second case,

$$\varepsilon^{-1} \sum_{\substack{p_1 p_2 \in I_k \\ p_i \geq \exp(\varepsilon^3 k)}} \frac{1}{p_1 p_2} k^2 \log^2(1+\varepsilon) \geq \log(1+\varepsilon)^k (1 + O(\varepsilon)).$$

Rearranging terms gives

$$\sum_{\substack{p_1 p_2 \in I_k \\ p_i \geq \exp(\varepsilon^3 k)}} \frac{1}{p_1 p_2} \geq \frac{\varepsilon}{k \log(1+\varepsilon)} (1 + O(\varepsilon)).$$

Taylor expanding the logarithm again completes the proof. □

Next, we show that we can actually find two nearby scales where both inequalities from Proposition 3.2.6 hold. The key idea is to use the connectedness of the interval.

Proposition 3.2.7. *Let $\varepsilon > 0$ be a number sufficiently small. Suppose that k_0 is sufficiently large depending on ε and let I_k denote the interval $(1+\varepsilon)^k$ to $(1+\varepsilon)^{k+1}$. Then there exists k and k' such that $|k - k'| \leq 1$ with k and k' in $[k_0, \varepsilon^{-2} + k_0]$ and such that*

$$\sum_{p \in I_k} \frac{1}{p} \geq \frac{1}{2k}$$

and

$$\sum_{\substack{p_1 p_2 \in I_{k'} \\ p_i \geq \exp(\varepsilon^3 k')}} \frac{1}{p_1 p_2} \geq \frac{1}{2k'}$$

Proof. Suppose not. Then by Proposition 3.2.6, for each k in $[k_0, \varepsilon^{-2} + k_0]$ either

$$\sum_{p \in I_k} \frac{1}{p} \geq \frac{1}{2k}$$

or

$$\sum_{\substack{p_1 p_2 \in I_k \\ p_i \geq \exp(\varepsilon^3 k)}} \frac{1}{p_1 p_2} \geq \frac{1}{2k}.$$

If both hold for some k, then by choosing $k = k'$, we could conclude that Proposition 3.2.7 holds. Thus, we will assume that exactly one of

$$\sum_{p \in I_k} \frac{1}{p} \geq \frac{1}{2k}$$

or

$$\sum_{\substack{p_1 p_2 \in I_k \\ p_i \geq \exp(\varepsilon^3 k)}} \frac{1}{p_1 p_2} \geq \frac{1}{2k}$$

hold for any choice of k. Whichever holds for k_0 must also hold for $k_0 + 1$ since otherwise we may choose $k = k_0$ and $k' = k_0 + 1$. Inductively, we may assume that for every k in $[k_0, \varepsilon^{-2} + k_0]$ either

$$\sum_{p \in I_k} \frac{1}{p} < \frac{1}{2k}$$

or

$$\sum_{\substack{p_1 p_2 \in I_k \\ p_i \geq \exp(\varepsilon^3 k)}} \frac{1}{p_1 p_2} < \frac{1}{2k}.$$

Summing in k, we eventually obtain a contradiction with Mertens' theorem: either

$$\sum_{(1+\varepsilon)^{k_0} \leq p \leq (1+\varepsilon)^{k_0 + \varepsilon^{-2}}} \frac{1}{p} < \frac{1}{10} \cdot \left(\log(k_0 + \varepsilon^{-2}) - \log k_0 + O\left(\frac{1}{k_0}\right) \right) \tag{3.3}$$

or

$$\sum_{k \in [k_0, \varepsilon^{-2} + k_0]} \sum_{\substack{p_1 p_2 \in I_k \\ p_i \geq \exp(\varepsilon^3 k)}} \frac{1}{p_1 p_2} < \frac{1}{2} \cdot \left(\log(k_0 + \varepsilon^{-2}) - \log k_0 + O\left(\frac{1}{k_0}\right) \right). \tag{3.4}$$

We remark that a Taylor expansion could simplify

$$\log(k_0 + \varepsilon^{-2}) - \log k_0 + O\left(\frac{1}{k_0}\right) = O\left(\frac{1}{\varepsilon^2 k_0}\right).$$

Note that Mertens' theorem implies that

$$\sum_{p \leq x} \frac{1}{p} = \log\log x + M + O\left(\frac{1}{\log x}\right),$$

for some constant M. Taking differences,

$$\sum_{(1+\varepsilon)^{k_0}\leq p\leq(1+\varepsilon)^{k_0+\varepsilon^{-2}}}\frac{1}{p}=\log\log(1+\varepsilon)^{k_0+\varepsilon^{-2}}-\log\log(1+\varepsilon)^{k_0}+O\left(\frac{1}{k_0\log(1+\varepsilon)}\right)$$
$$=\log(k_0+\varepsilon^{-2})-\log(k_0)+O\left(\frac{1}{k_0\log(1+\varepsilon)}\right).$$

But 3.3 says that the sum on the left is 2 times smaller than that which gives a contradiction.

□

In the next proposition, we show that this implies there are nearby primes and semiprimes which are good.

Proposition 3.2.8. *Let $\varepsilon>0$. Let P be a natural number which is sufficiently large depending on ε. Let N be a natural number which is sufficiently large depending on P. Then there exists p_1, p_2 and p such that*

$$\frac{p_1p_2}{p}=1+O(\varepsilon)$$

with p_1, p_2 and p good in the sense of Definition 3.2.3 meaning p_1,p_2 and p are not in $S(n)$ for "most" $n\leq N$ (see Definition 3.2.3 for details). Furthermore, we can require that p_1, p_2 and p are all greater than $\frac{1}{\varepsilon}$.

Proof. By Proposition 3.2.7, it suffices to show that, for some k_0 sufficiently large depending on ε with the property that $(1+\varepsilon)^{k_0+\varepsilon^{-2}}\leq P$, we have

$$\sum_{\substack{p\in[(1+\varepsilon)^{k_0},(1+\varepsilon)^{\varepsilon^{-2}+k_0}]\\ p\text{ bad}}}\frac{1}{p}\leq\frac{1}{10k_0}\tag{3.5}$$

and that

$$\sum_{\substack{p_1p_2\in[(1+\varepsilon)_0^k,(1+\varepsilon)^{\varepsilon^{-2}+k_0}]\\ p_1\text{ bad}\\ p_1^{\varepsilon^3}\leq p_2\leq p_1^{\varepsilon^{-3}}}}\frac{1}{p_1p_2}\leq\frac{1}{10k_0}.\tag{3.6}$$

After all, once we have shown this, we can argue as follows: by Proposition 3.2.7 there exists an interval of the form k and k' in $[k_0, k_0 + \varepsilon^{-2}]$ with $|k - k'| \leq 1$ for which

$$\sum_{p \in [(1+\varepsilon)^k, (1+\varepsilon)^{k+1}]} \frac{1}{p} > \frac{1}{k}$$

and

$$\sum_{\substack{p_1 p_2 \in I_{k'} \\ p_i \geq \exp(\varepsilon^3 k')}} \frac{1}{p_1 p_2} \geq \frac{1}{k'},$$

where $I_k = [(1+\varepsilon)^k, (1+\varepsilon)^{k+1}]$ and similarly for $I_{k'}$. By 3.5

$$\sum_{\substack{p \in [(1+\varepsilon)^k, (1+\varepsilon)^{k+1}] \\ p \text{ good}}} \frac{1}{p} > 0$$

and by 3.6

$$\sum_{\substack{p_1 p_2 \in I_{k'} \\ p_i \geq \exp(\varepsilon^3 k') \\ p_1, p_2 \text{ good}}} \frac{1}{p_1 p_2} > 0.$$

Now any good p in I_k and any good $p_1 p_2$ in $I_{k'}$ suffices to prove the result.

Now, for the sake of contradiction, suppose first that

$$\sum_{\substack{p \in [(1+\varepsilon)_0^k, (1+\varepsilon)^{\varepsilon^{-2}+k_0}] \\ p \text{ bad}}} \frac{1}{p} \geq \frac{1}{10k_0}.$$

Summing in $k_0 \leq \log\log P$, for P large enough we get that

$$\sum_{\substack{p \leq \log N \\ p \text{ bad}}} \frac{1}{p} \geq \frac{1}{20} \log\log\log P$$

which contradicts Corollary 3.2.4. Second, suppose that

$$\sum_{\substack{p_1 p_2 \in [(1+\varepsilon)_0^k, (1+\varepsilon)^{\varepsilon^{-2}+k_0}] \\ p_1 \text{ bad} \\ p_1^{\varepsilon^3} \leq p_2 \leq p_1^{\varepsilon^{-3}}}} \frac{1}{p_1 p_2} \geq \frac{1}{10k_0}.$$

Summing in $k_0 \leq \log\log P$ gives, for P large enough

$$\sum_{\substack{p_1p_2\leq \log N \\ p_1 \text{ bad} \\ p_1^{\varepsilon^3}\leq p_2\leq p_1^{\varepsilon^{-3}}}} \frac{1}{p_1p_2} \geq \frac{1}{20}\log\log\log P.$$

For each p_1, by Mertens' theorem,

$$\sum_{p_1^{\varepsilon^3}\leq p_2\leq p_1^{\varepsilon^{-3}}} \frac{1}{p_2} \leq -10\log\varepsilon.$$

By Corollary 3.2.4, this implies

$$\sum_{\substack{p_1p_2\leq \log N \\ p_1 \text{ bad} \\ p_1^{\varepsilon^3}\leq p_2\leq p_1^{\varepsilon^{-3}}}} \frac{1}{p_1p_2} = O\left(\varepsilon^{-3}|\log\varepsilon|\right)$$

which yields a contradiction since for P large enough, $\frac{1}{20}\log\log\log P \gg \varepsilon^{-3}|\log\varepsilon|$. □

Finally, we show this implies the prime number theorem.

Theorem 3.2.9. *The prime number theorem holds, i.e.*

$$\frac{1}{N}\sum_{n\leq N}\Lambda(n) = 1 + o_{N\to\infty}(1)$$

Proof. Let ε be a positive real number, let P be a natural number which is sufficiently large depending on ε and let N be a natural number sufficiently large depending on P. By Proposition 3.2.8, there exist primes p_1, p_2 and p all good and greater than $\frac{1}{\varepsilon}$ such that

$$\frac{p_1p_2}{p} = 1 + O(\varepsilon).$$

By definition of a good prime,

$$\frac{1}{M}\left|\sum_{n\leq M}\mu(n) - \sum_{n\leq M}\mu(n)p\mathbb{1}_{p|n}\right| \geq \varepsilon,$$

for at most a small set of M (exactly how small will be spelled out shortly). In particular, let $S(M)$ denote the set of primes such that

$$\frac{1}{M}\left|\sum_{n\leq M}\mu(n)-\sum_{n\leq M}\mu(n)p\mathbb{1}_{p|n}\right|\geq\varepsilon.$$

Then by definition of a good prime,

$$\frac{1}{\ell(N)}\sum_{M\leq N}\frac{1}{M}\mathbb{1}_{p_1\in S(M)}\mathbb{1}_{p_2\in S(M)}\mathbb{1}_{p\in S(M)}=O(\varepsilon).$$

Thus, we may conclude that

$$\frac{1}{\ell(N)}\sum_{M\leq N}\frac{1}{M}\frac{1}{M}\left|\sum_{n\leq M}\mu(n)-\sum_{n\leq M}\mu(n)p\mathbb{1}_{p|n}\right|=O(\varepsilon).$$

Since $\mu(np)=-\mu(n)$ for most n (including all but those $O(\frac{1}{p})=O(\varepsilon)$ fraction of n which are not divisible by p), we conclude that

$$\frac{1}{\ell(N)}\sum_{M\leq N}\frac{1}{M}\left|\frac{1}{M}\sum_{n\leq M}\mu(n)+\frac{p}{M}\sum_{n\leq M/p}\mu(n)\right|=O(\varepsilon).$$

Similarly, since p_1 is good,

$$\frac{1}{\ell(N)}\sum_{M\leq N}\frac{1}{M}\left|\frac{1}{M}\sum_{n\leq M}\mu(n)+\frac{p_1}{M}\sum_{n\leq M/p_1}\mu(n)\right|=O(\varepsilon).$$

By change of variables,

$$\frac{1}{\ell(N)}\sum_{M\leq N}\frac{1}{M}\left|\frac{p_1}{M}\sum_{n\leq M/p_1}\mu(n)+\frac{p_1p_2}{M}\sum_{n\leq M/p_1p_2}\mu(n)\right|=O(\varepsilon)+O\left(\frac{\log p_1}{\log N}\right).$$

By the triangle inequality and since N is much larger than p_1,

$$\frac{1}{\ell(N)}\sum_{M\leq N}\frac{1}{M}\left|\frac{p}{M}\sum_{n\leq M/p}\mu(n)+\frac{p_1p_2}{M}\sum_{n\leq M/p_1p_2}\mu(n)\right|=O(\varepsilon).$$

But since $\frac{p_1p_2}{p}=1+O(\varepsilon)$,

$$\frac{1}{\ell(N)}\sum_{M\leq N}\frac{1}{M}\left|\frac{p}{M}\sum_{n\leq M/p}\mu(n)\right|=O(\varepsilon).$$

and therefore, again using that p is good,

$$\frac{1}{\ell(N)}\sum_{M\leq N}\frac{1}{M}\left|\frac{1}{M}\sum_{n\leq M}\mu(n)\right| = O(\varepsilon).$$

This is an averaged version on the equation we want. We want that

$$\left|\frac{1}{N}\sum_{n\leq N}\mu(n)\right| = O(\varepsilon),$$

for all N sufficiently large. Thus, we just need to prove

Lemma 3.2.10. *Let $\varepsilon > 0$, let N be sufficiently large depending on ε and suppose that*

$$\frac{1}{\ell(N)}\sum_{M\leq N}\frac{1}{M}\left|\frac{1}{M}\sum_{n\leq M}\mu(n)\right| = O(\varepsilon).$$

Then

$$\left|\frac{1}{N}\sum_{n\leq N}\mu(n)\right| = O(\varepsilon),$$

To prove this we use the identity

$$\mu\cdot\log = -\mu * \Lambda.$$

Summing both sides up to N gives

$$\sum_{n\leq N}\mu(n)\log n = -\sum_{n\leq N}\sum_{d|n}\mu\left(\frac{n}{d}\right)\Lambda(d).$$

Now by switching the order of summation

$$= -\sum_{d\leq N}\Lambda(d)\left(\sum_{n\leq N/d}\mu(n)\right).$$

If it were not for the factor of $\Lambda(d)$, this would be exactly what we want. Each $\sum_{n\leq M}\mu(n)$ for an integer M occurs in this sum the number of times that $\left\lfloor\frac{N}{d}\right\rfloor = M$ where $\lfloor\cdot\rfloor$ denotes the floor which is proportional to $\frac{N}{M^2}$. The factor of $\Lambda(d)$ can be removed using the Brun-Titchmarsh inequality as follows. First, we break up the sum into different scales

$$= -\sum_{a\in(1+\varepsilon)^{\mathbb{N}}}\sum_{\substack{d\leq N\\ a\leq d<(1+\varepsilon)a}}\Lambda(d)\left(\sum_{n\leq N/d}\mu(n)\right).$$

For all d between a and $(1+\varepsilon)a$, the sums $\sum_{n\leq N/d}\mu(n)$ all give roughly the same value. Therefore

$$\begin{aligned}
&= -\sum_{a\in(1+\varepsilon)^{\mathbb{N}}}\sum_{\substack{d\leq N\\ a\leq d<(1+\varepsilon)a}}\Lambda(d)\left(\sum_{n\leq N/a}\mu(n)\right)\\
&+O\left(\sum_{\substack{a\in(1+\varepsilon)^{\mathbb{N}}\\ a\leq N}}\sum_{a\leq d<(1+\varepsilon)a}\Lambda(d)\sum_{\frac{N}{a(1+\varepsilon)}\leq n\leq\frac{N}{a}}1\right)
\end{aligned}$$

First, we focus on the error term.

$$\begin{aligned}
&O\left(\sum_{\substack{a\in(1+\varepsilon)^{\mathbb{N}}\\ a\leq N}}\sum_{a\leq d<(1+\varepsilon)a}\Lambda(d)\sum_{\frac{N}{a(1+\varepsilon)}\leq n\leq\frac{N}{a}}1\right)\\
=&O\left(\sum_{\substack{a\in(1+\varepsilon)^{\mathbb{N}}\\ a\leq N}}\sum_{a\leq d<(1+\varepsilon)a}\Lambda(d)\left(\frac{N}{a}-\frac{N}{a(1+\varepsilon)}\right)\right)\\
=&O\left(\sum_{\substack{a\in(1+\varepsilon)^{\mathbb{N}}\\ a\leq N}}\sum_{a\leq d<(1+\varepsilon)a}\Lambda(d)\frac{N\varepsilon}{a(1+\varepsilon)}\right)
\end{aligned}$$

By the Brun-Titchmarsh inequality

$$\begin{aligned}
=&O\left(\sum_{\substack{a\in(1+\varepsilon)^{\mathbb{N}}\\ a\leq N}}\varepsilon a\frac{N\varepsilon}{a(1+\varepsilon)}\right)\\
=&O\left(\sum_{\substack{a\in(1+\varepsilon)^{\mathbb{N}}\\ a\leq N}}N\varepsilon^2\right)\\
=&O\left(N\varepsilon^2\log_{1+\varepsilon}N\right)\\
=&O\left(\varepsilon N\log N\right)
\end{aligned}$$

where the last step involves Taylor expanding $\log(1+\varepsilon)$ near $\varepsilon = 0$. Next, we turn our attention to the main term. We begin by pulling out the sum over $\mu(n)$ which no longer depends on d.

$$\begin{aligned} &- \sum_{a \in (1+\varepsilon)^{\mathbb{N}}} \sum_{\substack{d \leq N \\ a \leq d < (1+\varepsilon)a}} \Lambda(d) \left(\sum_{n \leq N/a} \mu(n) \right) \\ =& - \sum_{a \in (1+\varepsilon)^{\mathbb{N}}} \left(\sum_{n \leq N/a} \mu(n) \right) \left(\sum_{\substack{d \leq N \\ a \leq d < (1+\varepsilon)a}} \Lambda(d) \right). \end{aligned}$$

By the Brun-Titchmarsh inequality, this is bounded in absolute value by

$$\leq \sum_{\substack{a \in (1+\varepsilon)^{\mathbb{N}} \\ a \leq N}} \left| \sum_{n \leq N/a} \mu(n) \right| \cdot (10\varepsilon a).$$

Earlier, we replaced a sum indexed by $n \leq N/d$ by a sum indexed by $n \leq N/a$, showing these two sums were close up to an error of size $O(\varepsilon N \log N)$. Undoing this process, we find

$$=10 \sum_{d \leq N} \left| \sum_{n \leq N/d} \mu(n) \right| + O(\varepsilon N \log N).$$

Now we let $M = \frac{N}{d}$. The number of values of d such that $\lfloor \frac{N}{d} \rfloor$ is the number of values of d such that $M \leq \frac{N}{d} < M+1$ and therefore $\frac{N}{M+1} < d \leq \frac{N}{M}$. The number of such d's is bounded by $\frac{N}{M} - \frac{N}{M+1} = \frac{N}{M(M+1)}$. Thus

$$\leq 10 \sum_{M \leq N} \frac{N}{M^2} \left| \sum_{n \leq M} \mu(n) \right| + O(\varepsilon N \log N).$$

But we already showed that this sum is bounded by

$$\begin{aligned} &= O(\varepsilon N \ell(N)) \\ &= O(\varepsilon N \log N). \end{aligned}$$

Thus,

$$\sum_{n \leq N} \mu(n) \log(n) = O(\varepsilon N \log N).$$

Since $\log n = \log N(1+O(\varepsilon))$ for n between $\varepsilon \frac{N}{\log N}$ and N and ε sufficiently small we conclude that

$$\sum_{n \leq N} \mu(n) = O(\varepsilon N).$$

But this classically implies the prime number theorem. □

3.3 In what ways is this a dynamical proof?

To begin the argument, we showed that for all N, for most p i.e. all p outside a bad set where

$$\sum_{p \text{ bad}} \frac{1}{p} \leq C_\varepsilon$$

we have that

$$\sum_{n \leq N} \mu(n) = \sum_{n \leq N} \mu(n) p \mathbb{1}_{p|n} + O(\varepsilon).$$

We did this using an L^2 orthogonality argument (Propositions 3.2.1 and 3.2.2). Alternately, we can argue using a variant of Tao's entropy decrement argument (the first version of this argument appeared in [Tao16b]; a different version of the entropy decrement argument appeared in [TT17b] and [TT17a]; the version presented here is somewhat different from what appeared in those papers). Let $\mathbf{n}$ be a random integer less than N. Let $\mathbf{x}_i = \mu(\mathbf{n}+i)$ and let $\mathbf{y}_p = \mathbf{n} \bmod p$. In probability and dynamics, a stochastic process is a sequence of random variables $(\ldots, \xi_{-2}, \xi_{-1}, \xi_0, \xi_1, \xi_2, \ldots)$ such that

$$\mathbb{P}((\xi_1, \ldots \xi_k) \in A) = \mathbb{P}((\xi_{1+m}, \ldots \xi_{k+m}) \in A)$$

for any set A and for any m. In our setting $(\ldots, \mathbf{x}_{-2}, \mathbf{x}_{-1}, \mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots)$ is approximately stationary in the sense that

$$\mathbb{P}((\mathbf{x}_1, \ldots \mathbf{x}_k) \in A) \approx \mathbb{P}((\mathbf{x}_{1+m}, \ldots \mathbf{x}_{k+m}) \in A)$$

where the two terms differ by some small error which is $o_{N\to\infty,m}(1)$. A stationary process is the same as a random variable in a measure preserving system where ξ_{i+1} is the transfor-

mation applied to ξ_i. A key invariant of a stationary process is thus the Kolmogorov-Sinai entropy:

$$h(\xi) = \lim_{n\to\infty} \frac{1}{n} H(\xi_1, \ldots, \xi_n)$$

where

$$H(\xi_1, \ldots, \xi_n)$$

is the Shannon entropy of $(\xi_1, \ldots, \xi_n)$. This limit exists because

$$\frac{1}{n} H(\xi_1, \ldots, \xi_n) = \frac{1}{n} \sum_{i\leq n} H(\xi_i | \xi_1 \ldots, \xi_{i-1})$$

by the chain rule for entropy, which is equal to

$$= \frac{1}{n} \sum_{i\leq n} H(\xi_0 | \xi_{-1} \ldots, \xi_{-i+1})$$

by stationarity. This is a Caesaró average of a decreasing sequence which is therefore decreasing. Since entropy is nonnegative, we can conclude that the limit exists. In our case, because $(\ldots, \mathbf{x}_{-1}, \mathbf{x}_0, \mathbf{x}_1, \ldots)$ is almost stationary, we can conclude that

$$\frac{1}{n} H(\mathbf{x}_1, \ldots, \mathbf{x}_n)$$

is almost decreasing in the sense that, for $m > n$,

$$\frac{1}{m} H(\mathbf{x}_1, \ldots, \mathbf{x}_m) \leq \frac{1}{n} H(\mathbf{x}_1, \ldots, \mathbf{x}_n) + o_{N\to\infty,n}(1).$$

The same is true for the relative entropy

$$\frac{1}{n} H(\mathbf{x}_1, \ldots, \mathbf{x}_n | \mathbf{y}_{p_1}, \ldots, \mathbf{y}_{p_k})$$

for any fixed set of primes $p_1, \ldots, p_k$.

We define the mutual information between two random variables $\mathbf{x}$ and $\mathbf{y}$ as

$$I(\mathbf{x}; \mathbf{y}) = H(\mathbf{x}) - H(\mathbf{x}|\mathbf{y})$$

and more generally the conditional mutual information

$$I(\mathbf{x};\mathbf{y}|\mathcal{Z}) = H(\mathbf{x}|\mathcal{Z}) - H(\mathbf{x}|\mathbf{y},\mathcal{Z}).$$

We assume for the rest of the explanation that all random variables take only finitely many values. Mutual information measures how close two random variables are to independent. Two random variables $\mathbf{x}$ and $\mathbf{y}$ are independent if and only if

$$I(\mathbf{x};\mathbf{y}) = 0.$$

Intuitively, we think of $\mathbf{x}$ and $\mathbf{y}$ as close to independent if the mutual information is small. The crux of the entropy decrement argument is that we can find primes p such that $(\mathbf{x}_1,\ldots,\mathbf{x}_p)$ is close to independent of $\mathbf{y}_p$. The argument is as follows. Let $p_1 < p_2 < \ldots < p_k$ be a sequence of primes. Consider the relative entropy

$$\begin{aligned}&\frac{1}{p_k}H(\mathbf{x}_1,\ldots,\mathbf{x}_{p_k}|\mathbf{y}_{p_1},\ldots,\mathbf{y}_{p_k})\\ =&\frac{1}{p_k}H(\mathbf{x}_1,\ldots,\mathbf{x}_{p_k}|\mathbf{y}_{p_1},\ldots,\mathbf{y}_{p_{k-1}}) - \frac{1}{p_k}I(\mathbf{x}_1,\ldots,\mathbf{x}_{p_k};\mathbf{y}_{p_k}|\mathbf{y}_{p_1},\ldots,\mathbf{y}_{p_{k-1}})\end{aligned}$$

and because the relative entropy is almost decreasing

$$=\frac{1}{p_{k-1}}H(\mathbf{x}_1,\ldots,\mathbf{x}_{p_{k-1}}|\mathbf{y}_{p_1},\ldots,\mathbf{y}_{p_{k-1}}) - \frac{1}{p_k}I(\mathbf{x}_1,\ldots,\mathbf{x}_{p_k};\mathbf{y}_{p_k}|\mathbf{y}_{p_1},\ldots,\mathbf{y}_{p_{k-1}}) + o(1).$$

Inductively, we find

$$\leq H(\mathbf{x}_1) - \sum_{j\leq k}\frac{1}{p_j}I(\mathbf{x}_1,\ldots,\mathbf{x}_{p_j};\mathbf{y}_{p_j}|\mathbf{y}_{p_1},\ldots,\mathbf{y}_{p_{j-1}}) + o(1)$$

We conclude that the set of bad primes p_j for which

$$I(\mathbf{x}_1,\ldots,\mathbf{x}_{p_j};\mathbf{y}_{p_j}|\mathbf{y}_{p_1},\ldots,\mathbf{y}_{p_{j-1}}) \geq \varepsilon$$

satisfies

$$\sum_{p_j \text{ bad}}\frac{1}{p_j} \leq \varepsilon^{-1}H(\mathbf{x}_1) + o(1) < \infty.$$

Thus, for most primes,

$$I(\mathbf{x}_1, \ldots, \mathbf{x}_{p_j}; \mathbf{y}_{p_j} | \mathbf{y}_{p_1}, \ldots, \mathbf{y}_{p_{j-1}}) < \varepsilon.$$

In a slight abuse of terminology, we say such primes are good. Although this definition is apparently different from Definition 3.2.3, we will show that this notion of good meaning small mutual information essentially implies the "random sampling" version defined in Definition 3.2.3.

Intuitively, if p is good then $\mathbf{x}_1, \ldots, \mathbf{x}_p$ and $\mathbf{y}_p$ are nearly independent. This is formalized by Pinsker's inequality. Pinsker's inequality states that

$$d_{TV}(\mathbf{x}, \mathbf{y}) \leq D(\mathbf{x}||\mathbf{y})^{1/2}$$

where d_{TV} is the total variation distance and D is the Kullback-Leibler divergence. For our purposes, the important thing about the Kullback-Liebler divergence is that if $\mathbf{y}'$ is a random variable with the same distribution as $\mathbf{y}$ which is independent of $\mathbf{x}$ then

$$D((\mathbf{x}, \mathbf{y})||(\mathbf{x}, \mathbf{y}')) = I(\mathbf{x}; \mathbf{y}).$$

Therefore, we conclude that

$$d_{TV}((\mathbf{x}, \mathbf{y}), (\mathbf{x}, \mathbf{y}')) \leq I(\mathbf{x}; \mathbf{y})^{1/2}.$$

Similarly, there is a relative version

$$d_{TV}((\mathbf{x}, \mathbf{y}, \mathcal{Z}), (\mathbf{x}, \mathbf{y}', \mathcal{Z})) \leq I(\mathbf{x}; \mathbf{y}|\mathcal{Z})^{1/2},$$

where now $\mathbf{y}'$ has the same distribution as $\mathbf{y}$ but is relatively independent of $\mathbf{x}$ over $\mathcal{Z}$ meaning that

$$\mathbb{P}(\mathbf{x} \in A, \mathbf{y} \in B | \mathcal{Z} = c) = \mathbb{P}(\mathbf{x} \in A | \mathcal{Z} = c)\mathbb{P}(\mathbf{y} \in B | \mathcal{Z} = c).$$

Thus, for bounded function F,

$$\mathbb{E}F(\mathbf{x}, \mathbf{y}, \mathcal{Z}) = \mathbb{E}F(\mathbf{x}, \mathbf{y}', \mathcal{Z}) + O(I(\mathbf{x}; \mathbf{y})^{1/2}),$$

where again $\mathbf{y}'$ is relatively independent of $\mathbf{x}$ over $\mathcal{Z}$ and $\mathbb{E}$ denotes the expectation. In our case, for a good prime p where

$$I(\mathbf{x}_1, \dots, \mathbf{x}_p; \mathbf{y}_p | (y_q)_{q<p}) < \varepsilon$$

we note that

$$\mathbb{E}F(\mathbf{x}_1, \dots, \mathbf{x}_p, \mathbf{y}_p) = \mathbb{E}F(\mathbf{x}_1, \dots \mathbf{x}_p, \mathbf{y}'_p) + O(\varepsilon^{1/2}).$$

for any bounded function F where $\mathbf{y}'_p$ is relatively independent of $(\mathbf{x}_1, \dots, \mathbf{x}_p)$ over $(\mathbf{y}_q)_{q<p}$. Since $\mathbf{y}_p$ and $(\mathbf{y}_q)_{q<p}$ are already very nearly independent by the Chinese remainder theorem (and in fact if N is a multiple of the product of primes less than p, then $\mathbf{y}_p$ and $(\mathbf{y}_q)_{q<p}$ are genuinely independent) we can conclude that

$$\mathbb{E}F(\mathbf{x}_1, \dots, \mathbf{x}_p, \mathbf{y}_p) = \mathbb{E}F(\mathbf{x}_1, \dots \mathbf{x}_p, \mathbf{y}'_p) + O(\varepsilon^{1/2}),$$

where now $\mathbf{y}'_p$ is genuinely independent of $(\mathbf{x}_1, \dots, \mathbf{x}_p)$. For example, if we want to evaluate

$$\frac{1}{N} \sum_{n \leq N} \mu(n)$$

we could interpret this as

$$\mathbb{E}F(\mathbf{x}_0)$$

where $F(x) = x$. Alternately, we can average

$$\frac{1}{N} \sum_{n \leq N} \mu(n) \approx \frac{1}{p} \sum_{i \leq p} \mu(n+i),$$

which is

$$\mathbb{E}F(\mathbf{x}_1, \dots, \mathbf{x}_p)$$

where now $F(x_1, \dots, x_p) = \frac{1}{p} \sum_{i \leq p} x_i$. Now let $\mathbf{y}'_p$ as before be independent of $(\mathbf{x}_1, \dots, \mathbf{x}_p)$ and uniformly distributed among residue classes mod p. Then this is also

$$\mathbb{E}F(\mathbf{x}_1, \dots, \mathbf{x}_p, \mathbf{y}'_p)$$

where

$$F(x_1, \dots, x_p, y_p) = \frac{1}{p} \sum_{i \leq p} x_i p \mathbb{1}_{y_p = -i}.$$

As we noted, for p a good prime, this is approximately,

$$\mathbb{E}F(\mathbf{x}_1, \dots, \mathbf{x}_p, \mathbf{y}'_p) \approx \mathbb{E}F(\mathbf{x}_1, \dots, \mathbf{x}_p, \mathbf{y}_p)$$

and unpacking definitions this is

$$\mathbb{E}F(\mathbf{x}_1, \dots, \mathbf{x}_p, \mathbf{y}_p) = \frac{1}{N} \sum_{n \leq N} \frac{1}{p} \sum_{i \leq p} \mu(n+i) p \mathbb{1}_{n = -1 \bmod p}.$$

Undoing the averaging in i gives

$$\approx \frac{p}{N} \sum_{n \leq N} \mu(n) \mathbb{1}_{p|n}.$$

Thus, the analogue of Corollary 3.2.4 can be proved using the entropy decrement argument, which can be interpreted in the dynamical setting.

The rest of the proof can also be translated to the dynamical setting. The Furstenberg system corresponding to the Möbius function can be constructed as follows. The underlying space is the set of functions from $\mathbb{Z}$ to $\{-1, 0, 1\}$. We construct a random variable on this space. Consider a random shift of the Möbius function. Formally, let $\mathbf{n}$ be a uniformly chosen random integer between 1 and N and let $\mathbf{X}_N$ denote the function μ (say extended by 0 to the left) shifted by $\mathbf{n}$ i.e. $\mathbf{X}_N(i) = \mu(i + \mathbf{n})$. Since the underlying space of functions from $\mathbb{Z}$ to $\{-1, 0, 1\}$ is compact, there is a subsequence of $(\mathbf{X}_N)_N$ which converges weakly to a random variable $\mathbf{X}$. Since the distribution of each random variable $\mathbf{X}_N$ is "approximately" shift invariant, the distribution of the limit $\mathbf{X}$ is actually shift invariant. Thus, we obtain a shift invariant measure ν on the space of functions from $\mathbb{Z}$ to $\{-1, 0, 1\}$ with the property that if f is the "evaluation at zero" map

$$f((a_n)_{n \in \mathbb{Z}}) = a_0$$

then

$$\int f(x) \nu(dx) = \mathbb{E}f(\mathbf{X})$$

is a subsequential limit of terms of the form

$$\frac{1}{N}\sum_{n\leq N}\mu(n).$$

Thus, we can encode questions about the average of μ or more generally shifts like $\mu(n)\mu(n+1)$ in a dynamical way.

In order to take advantage of the fact that μ is multiplicative, we need to impose extra structure on the dynamical systems we associate to μ. This extra structure is implicit in [TT17b] and [TT17a] and is explicitly described first in [Tao17b]. See also [Saw20] and [McN18]. One key feature of multiplicative functions is that they are statistically multiplicative in the sense that for any $\epsilon_1,\dots,\epsilon_k$ in $\{-1,0,1\}$,

$$\begin{aligned}&\frac{p}{N}\#\{n\leq N\colon \mu(n+pi)=\epsilon_i \text{ for all } i \text{ and } p|n\}\\ =&\frac{p}{N}\#\{n\leq N/p\colon \mu(n+i)=-\epsilon_i \text{ for all } i\}+O\left(\frac{1}{p}\right).\end{aligned}$$

(This holds simply by changing variables and using that μ is multiplicative). For N in some subsequence, we can think of the right hand side as

$$\frac{p}{N}\#\{n\leq N/p\colon \mu(n+pi)=\epsilon_i \text{ for all } i\}\approx \nu\{x\colon f(T^{ip}x)=\epsilon_i\}.$$

We would like a way of encoding this identity in our dynamical system. One solution is to use logarithmic averaging. Now let $\mathbf{n}$ denote a random integer between 1 and N which is not uniformly distributed but which is logarithmically distributed meaning the probability that $\mathbf{n}=m$ is proportional to $\frac{1}{m}$ for $m\leq N$. Let $\mathbf{X}_N(i)=\mu(n+i)$ be a random translate of the Möbius function. Consider the pair $(\mathbf{X}_N,\mathbf{n})$ in the space of pairs of functions from $\mathbb{Z}$ to $\{-1,0,1\}$ and profinite integers. This product space is compact so there is a weak limit $(\mathbf{X},\mathbf{y})$ where $\mathbf{X}$ is a functions from $\mathbb{Z}$ to $\{-1,0,1\}$ and $\mathbf{y}$ is a profinite integer. Let $T(x,y)=(n\mapsto x(n+1),y+1)$. Let ρ be the distribution of $(\mathbf{X},\mathbf{y})$ which is a T-invariant measure on our space. Consider the map I_p on pairs of functions and profinite integers which are 0 mod p which dilates the function by p, multiplies the function by -1 and divides the

profinite integer by p i.e.

$$I_p(x, y) = (n \mapsto -x(pn), y/p).$$

For a point (x, y) in our space, let M denote the projection onto the second factor

$$M(x, y) = y.$$

Let f be the "evaluation of the function at 0" function i.e.

$$f(x, y) = x(0).$$

Then the dynamical system has the following properties, where x is always a function from $\mathbb{Z}$ to $\{-1, 0, 1\}$, p and q are primes and y is a profinite integer:

1. For all p, for all x and y such that $M(x, y) = 0 \bmod p$,

$$I_p(T^p(x, y)) = T(I_p(x, y)).$$

2. For all p and q, for all x and y where $M(x, y)$ is 0 mod pq, we have

$$I_p(I_q(x, y)) = I_q(I_p(x, y)).$$

3. For all p, and for all measurable functions on our space ϕ,

$$\int \phi(x, y)\rho(dxdy) = \int p\mathbb{1}_{M(x,y)=0 \bmod p}\phi(I_p(x, y))\rho(dxdy) + O\left(\frac{1}{p}\right).$$

4. For all p and for all x and y such that $M(x, y) = 0 \bmod p$ we have that

$$f(I_p(x, y)) = -f(x, y).$$

A tuple $(X, \rho, T, f, M, (I_p)_p)$ where (X, ρ, T) is a measure preserving system and satisfying (1) through (4) is a called a dynamical model for μ. Translating our argument over to the dynamical context, there exists some p such that

$$\int f(x, y)\rho(dxdy) \approx \int f(x, y) \cdot p\mathbb{1}_{M(x,y)=0 \bmod p},$$

with an error term which we may make arbitrarily small by increasing p. On the other hand,

$$\begin{aligned}\int f(x,y)\cdot p\mathbb{1}_{M(x,y)=0 \bmod p} &= \int -f(I_p(x,y))\cdot p\mathbb{1}_{M(x,y)=0 \bmod p}\\ &= -\int f(x,y).\end{aligned}$$

We conclude that

$$\int f = 0,$$

for any dynamical model for μ.

In [Tao17b], Tao constructs a dynamical model where

$$\int f \approx \frac{1}{\log N}\sum_{n\leq N}\frac{1}{n}\mu(n)$$

i.e. using logarithmic averaging and the Furstenberg correspondence principle. However using either Corollary 3.2.4 or a version of the entropy decrement argument, we can argue as follows. Let ρ_N denote the distribution of $(\mathbf{X}_N, \mathbf{n})$ in the space of pairs of functions $\mathbb{Z}\to\{-1,0,1\}$ and profinite integers and where $\mathbf{n}$ is a uniformly distributed random integer between 1 and N and $\mathbf{X}_N(i)=\mu(\mathbf{n}+i)$. For any ϵ in S^1 and ϕ, define $\epsilon_*\rho_n$ by

$$\int \phi(x,y)\epsilon_*\rho_N(dxdy) = \int \phi(\epsilon\cdot x, y)\rho_N(dxdy).$$

Choose ϵ_N so that

$$\nu_m = \left(\sum_{n\leq m}\frac{1}{n}\right)^{-1}\sum_{N\leq m}\frac{1}{N}(\epsilon_N)_*\rho_N,$$

satisfies

$$\int f(x,y)\nu_m(x,y) = \left(\sum_{n\leq m}\frac{1}{n}\right)^{-1}\sum_{N\leq M}\frac{1}{N}\left|\frac{1}{N}\sum_{n\leq N}\mu(n)\right|,$$

i.e. ϵ_N is the sign of $\sum_{n\leq N}\mu(n)$. Using a version of Corollary 3.2.4 or the entropy decrement argument, one can prove that for most p (except for a set of logarithmic size at most a constant depending on ε),

$$(I_p)_*(p\mathbb{1}_{M=0 \bmod p}\ \nu_m) \approx \nu_m + O\left(\varepsilon + \frac{\log p}{\log m}\right).$$

By the argument from before (see the proof of Theorem 3.2.9), this is enough to conclude the prime number theorem.

CHAPTER 4

A Quantitative Erdős Discrepancy Theorem

4.1 Introduction to Chapter 3

Does there exist a sequence $f\colon \mathbb{N} \to \{\pm 1\}$ such that

$$\sup_{d,n} \left| \sum_{i \leq n} f(id) \right| < \infty?$$

This was apparently one of Paul Erdős' favorite problems, mentioned, for instance, in [Erd57] and [Erd90]. He proposed the problem in the 1930s and offered a $500 prize to any mathematician who could solve it. In 2015, it was solved by Terence Tao in [Tao16a]. The purpose of this chapter is to show the following quantitative version of Tao's theorem.

Theorem 4.1.1. *For any sequence f taking values in the unit sphere of a Hilbert space, for all natural numbers N,*

$$\sup_{\substack{n \leq N \\ d \leq e^{(\log\log N)^{\frac{-1}{242}} \cdot N}}} \left|\left| \sum_{i \leq n} f(id) \right|\right| \gtrsim \frac{(\log\log N)^{\frac{1}{484}}}{(\log\log\log\log N)^{\frac{1}{4}} \, (\log\log\log\log\log N)}.$$

We begin with a review of Tao's proof. Tao's proof uses an idea from the [Pol] which shows that it suffices to settle the Erdős discrepancy problem in the case of a random multiplicative function. Tao then splits up his multiplicative functions into two cases: a random case and a structured case.

First, let us talk about the random case. If f is a random string of independent plus or

minus ones, then expanding the square shows

$$\mathbb{E}\left|\sum_{n\leq N} f(n)\right|^2 = \mathbb{E}\sum_{n_i\leq N} f(n_1)f(n_2).$$

By interchanging the sum and the expectation,

$$= \sum_{n_i\leq N} \mathbb{E}f(n_1)f(n_2).$$

Now if $n_1 \neq n_2$ then $f(n_1)$ and $f(n_2)$ are independent. Thus the only contribution comes from the N different terms where $n_1 = n_2$. We conclud that

$$\mathbb{E}\left|\sum_{n\leq N} f(n)\right|^2 = N.$$

This implies that much of the time, the sum $\left|\sum_{n\leq N} f(n)\right|$ is at least $\gg N^{\frac{1}{2}}$. For this argument to work, we needed that $f(n_1)$ and $f(n_2)$ were uncorrelated even when n_1 and n_2. However, this is too much to ask of most multiplicative functions. After all, if this kind of logic worked for λ we could use it to imply the Riemann hypothesis. However, Tao in [Tao16b] was able to show that for a large class of unstructured multiplicative functions g that $g(n)$ and $g(n+h)$ were in some sense uncorrelated for small h. Running this same style of argument

$$\begin{aligned}\mathbb{E}\left|\sum_{h\leq H} g(n+h)\right|^2 &= \mathbb{E}\sum_{h_i\leq N} g(n+h_1)\overline{g(n+h_2)}. \\ &= \sum_{h_i\leq H} \mathbb{E}\ g(n+h_1)\overline{g(n+h_2)}. \\ &\approx H + \text{ small error.}\end{aligned} \tag{4.1}$$

As long as H tends to infinity, this style of argument shows that $\mathbb{E}\left|\sum_{h\leq H} g(n+h)\right|^2$ tends to infinity as well.

If our multiplicative function g is not random and unstructured, then it must have some algebraic structure. For Tao, that means g *pretends* to be an algebraic multiplicative function

namely a Dirichlet character times n^{it} for some t which is not too big. For instance, let χ_3 denote the multiplicative function for which $\chi_3(n) = 1$ if n is 1 mod 3, $\chi_3(n) = -1$ when $n = -1$ mod 3, and $\chi_3(n) = 0$ when $n = 0$ mod 3. Then χ_3 is almost a counter example to the Erdős discrepancy problem. After all, the partial sums of χ_3 are all either 0 or 1 and since χ_3 is multiplicative, summing over multiples of d changes nothing about the size of the partial sums (in absolute value). The reason χ_3 is not a counter example is that χ_3 is sometimes 0, so it does not fulfill the hypotheses of the Erdős discrepancy problem. To remedy this, one can define

$$\tilde{\chi}(3^a n) = \chi_3(n)$$

for any n which is not divisible by 3 and any natural number a. Now $\tilde{\chi}$ always takes values ± 1. However, it is not too hard to see that for any natural number k

$$\sum_{i \leq 3^k + 3^{k-1} + \cdots + 3 + 1} \tilde{\chi}(i) = k.$$

Thus the partial sums

$$\lim_{N \to \infty} \sup_{n \leq N} \left| \sum_{i \leq n} \tilde{\chi}(i) \right|$$

do tend to infinity; they just do so slowly at a rate of $\sim \log_3 N$. Note the differences between the random case and the structured case. In the structured case, there were clearly correlations between $g(n)$ and $g(n+h)$ even when h was small which meant we could not apply the techniques that worked in the random setting. However, there were advantages to being in the structured case. For example, we could explicitly describe for which values of n was

$$\left| \sum_{i \leq n} \tilde{\chi}(i) \right|$$

big. Since the two cases have different advantages and disadvantages and since some techniques which work in one case do not work in the other, it makes sense to split our proof into cases: when g is random and when g is structured.

In order to split into cases, we have to have a way to decide if g is close to a structured, algebraic multiplicative function like $\chi(n)n^{it}$ for some Dirichlet character χ. Thus, we need

a way of deciding if two multiplicative functions are close. Note that for any prime p, the fraction of numbers divisible by p is roughly $\frac{1}{p}$. This means that if P is a set of primes and

$$\sum_{p \in P} \frac{1}{p} < \varepsilon$$

then average number of divisors of a number in the set P is ε. In particular, if g and f are multiplicative functions which differ only on the set P then they agree on all but ε share of all natural numbers. Even if g and f agree on all but a set of primes P for which

$$\sum_{p \in P} \frac{1}{p} \sim 1$$

I claim that we should think of g and f as being "close". After all, since this is a convergent sum, it means the tails are eventually less than ε, which means that once we restrict our attention to some specific residue class say $b \bmod W$ then for those n which are $b \bmod W$ at most ε fraction of the terms fail to satisfy

$$g(n) = \text{ const } \cdot f(n)$$

for some constant independent of n. This motivates the following notion of distance.

$$\mathbb{D}(f, g; N)^2 = \sum \frac{1 - \text{Re } f(n)\overline{g(n)}}{p}.$$

Here the distance between $f(p)\overline{g(p)}$ and 1 is weighted according to which fraction of numbers are divisible by by p and therefore "care" about the difference. When

$$\mathbb{D}(g, n^{it}\chi(n); N)$$

is small for some $|t| \leq N$ and χ a Dirichlet character of small modulus, say $\leq \log\log\log N$, then we will say that g is pretentious. Otherwise, we will say that g is nonpretentious. This terminology might seem a little funny but it is standard in the field. In the case that our function g is pretentious, we will try to use structured methods to understand g. In the case where g is nonpretentious, we will use "random" methods to understand g.

Our proof of Theorem 4.1.1 is similar to Tao's proof in [Tao16a], adding in quantitative details as we go along. The proof of Theorem 4.1.1 is based on the following five results.

Lemma 4.2.1. *Let M and N be natural numbers and let $p_1, \dots, p_r$ be the primes less than N. Let f be a function taking values on the unit sphere in some Hilbert space H. Let A be a real number. Suppose that, for all $N' \leq N$,*

$$\frac{1}{M^r} \sum_{b_1,\dots,b_r \leq M} \left|\left| \sum_{n \leq N'} f(p_1^{b_1} \cdot \dots \cdot p_r^{b_r} n) \right|\right|^2 \leq A^2.$$

Then there exists a random multiplicative function g_ξ (meaning for each ξ a point in a probability space there exists a multiplicative function g_ξ) such that

$$\mathbb{E}_\xi \left| \sum_{n \leq N'} g_\xi(n) \right|^2 \leq A^2 + O\left(\frac{N}{M}\right).$$

Lemma 4.3.1. *Suppose that for some function g taking values in the unit circle and for natural numbers H and N, for all $1 \leq |h| \leq H$, we have*

$$\frac{1}{\log N} \sum_{n \leq N} \frac{g(n)\overline{g(n+h)}}{n} \leq \frac{1}{2H}.$$

Then

$$\frac{1}{\log N} \sum_{n \leq N} \frac{1}{n} \left| \sum_{h \leq H} g(n+h) \right|^2 \geq \frac{H}{2} - \frac{H+10}{\log N}.$$

Corollary 4.3.5. *Suppose g is a completely multiplicative function taking values on the unit circle. Let N be a sufficiently large natural number. Suppose further that*

$$\inf_{\substack{|t| \leq 10N \\ q \leq \log^{20} H}} \sum_{p \leq N} \frac{1 - Re\ (g(p)\overline{\chi(p)}p^{-it})}{p} \geq B.$$

Then for all $h \leq \log^{\frac{1}{4}} \log N$,

$$\left| \sum_{n \leq N} g(n)g(n+h) \right| \lesssim \left(e^{-\frac{B}{60}} + \log^{\frac{-1}{4}} \log N \right) \cdot N.$$

Corollary 4.4.3. *Suppose that g is a multiplicative function, χ and χ' are characters of modulus q and q' respectively and t and t' are real numbers such that $\log|t - t'| + \log qq'$ lies between 0 and $\log^{1.4} N$. Then either*

$$(\log\log N)^{\frac{1}{2}} \lesssim qq'\mathbb{D}(g, n^{it}\chi(n); N) \text{ or } (\log\log N)^{\frac{1}{2}} \lesssim qq'\mathbb{D}(g, n^{it'}\chi'(n); N).$$

Lemma 4.4.4. *Suppose that g is a multiplicative function satisfying, for some H, N, $A \geq 1$, χ, q, t and B that*

$$\sum_{H' \sim H} \sum_{n \in \mathbb{N}} \frac{\left|\sum_{m \leq H'} g(n+m)\right|^2}{n^{1+c}} \leq A^2 H \log N$$

and

$$\sum_{p \leq N} \frac{1 - Re\ g(p)\overline{\chi}(p)p^{-it}}{p} \leq B$$

where χ is a primitive, non-principal Dirichlet character mod q and $c = \frac{1}{\log N}$. Then for any k,

$$\min\left(k-1, \frac{\log H}{\log q}\right) \cdot e^{-2B-20} \cdot \frac{1}{40 \log\log q} - \ Error\ \leq A^2$$

where the error term smaller than the main term as long as

1. $|t| \leq \frac{1}{H^3} N^{\frac{1}{H^2}}$.
2. $\frac{\log N}{q^k} \geq 20$.
3. $e^{10} q^k \log q^k e^{B^{\frac{1}{2}}} \leq e^{(\log\log N)^{\frac{1}{2}}}$.
4. $H^4 \leq 2^k$.
5. *N and H are sufficiently large i.e. larger than some constant.*

In particular, for $q \leq (\log\log\log N)^{20}$, $|t| \leq \frac{1}{\log N} e^{\log^{\frac{1}{2}} N}$, $k = \left(\frac{\log\log N}{2\log\log\log\log N}\right)^{\frac{1}{2}}$, $\log H = \frac{1}{4}k$ and $B = \frac{60}{242} \cdot \log\log\log N + \log\log\log\log\log\log N$ we get that

$$\frac{\log^{\frac{1}{242}} \log N}{80 e^{20} \log^{\frac{1}{2}} \log\log\log N \log^2 \log\log\log\log N} \leq A^2$$

Now we show how to use these five results to conclude Theorem 4.1.1. Suppose that

$$\sup_{\substack{n \leq N \\ d \leq e^{\log^{\frac{1}{242}} \log N \cdot N}}} \left|\left|\sum_{i \leq n} f(id)\right|\right| \leq A.$$

For the sake of contradiction we may assume A is a small multiple of

$$\frac{(\log\log N)^{\frac{1}{484}}}{(\log\log\log\log N)^{\frac{1}{4}} (\log\log\log\log\log N)^{\frac{1}{2}}}.$$

By squaring both sides, we find

$$\sup_{\substack{n\leq N\\ d\leq e^{\log^{\frac{-1}{242}}\log N\cdot N}}}\left|\left|\sum_{i\leq n}f(id)\right|\right|^2\leq A^2.$$

By Lemma 4.2.1,

$$\mathbb{E}_\xi\left|\sum_{n\leq N'}g_\xi(n)\right|^2\lesssim A^2,$$

for all $N'\leq N$. Taking differences, we find that for any $H\ll N$ and any $n\leq N-H$,

$$\mathbb{E}_\xi\left|\sum_{h\leq H}g_\xi(n+h)\right|^2\lesssim A^2.$$

Now summing in n gives

$$\mathbb{E}_\xi\sum_{n\leq N'}\frac{\left|\sum_{h\leq H}g_\xi(n+h)\right|^2}{n}\lesssim A^2\log N' \tag{4.2}$$

for all $N'\leq N$ and similarly

$$\mathbb{E}_\xi\sum_{n\leq N'}\frac{\left|\sum_{h\leq H}g_\xi(n+h)\right|^2}{n^{1+\frac{1}{\log N}}}\lesssim A^2\log N'+O\left(\frac{1}{\log N}\right). \tag{4.3}$$

The error term can be absorbed into the main term because we assumed that A was not too small. In particular we can assume that (4.2) holds when $N'=N$ and $H=(\log\log N)^{\frac{1}{4}}$ and when $N'=\frac{1}{\log N}e^{(\log N)^{\frac{1}{2}}}$ and $H=(\log\log N')^{\frac{1}{4}}$ and (4.3) holds when $N'=N$ and $H=\exp\left(\frac{\log^{\frac{1}{2}}\log N}{8\log\log\log\log N}\right)$. By the pigeonhole principle, there exists at least one multiplicative function for which all three hold, possibly after worsening the implicit constant by $\frac{1}{3}$. Fix such a g. If g is nonpretentious then Lemma 4.3.1 and Corollary 4.3.5 will provide a contradiction: after all, if

$$\sum_{p\leq\frac{1}{\log N}e^{(\log N)^{\frac{1}{2}}}}\frac{1-\operatorname{Re}\,g(p)\overline{\chi(p)}p^{-it}}{p}\geq\frac{60}{242}\log\log\log N+O(1)$$

for all characters χ of modulus $\lesssim\log\log\log N$ then by Corollary 4.3.5, for all $h\leq\log^{\frac{1}{4}}\log N$,

$$\left|\sum_{n\leq\frac{1}{\log N}e^{(\log N)^{\frac{1}{2}}}}g(n)g(n+h)\right|\lesssim(\log\log N)^{\frac{-1}{242}}\cdot N;$$

then by Lemma 4.3.1

$$A^2 \geq (\log\log N)^{\frac{1}{242}}.$$

A similar thing happens for N. Thus there exists t, t' χ and χ' such that

$$\sum_{p \leq \frac{1}{\log N} e^{(\log N)^{\frac{1}{2}}}} \frac{1 - \text{Re } g(p)\overline{\chi(p)}p^{-it}}{p} \geq \frac{60}{242}\log\log\log N + O(1)$$

and

$$\sum_{p \leq N} \frac{1 - \text{Re } g(p)\overline{\chi'(p)}p^{-it'}}{p} \geq \frac{60}{242}\log\log\log N + O(1).$$

By Corollary 4.4.3, $|t - t'| \leq 1$. In particular, we may conclude that $|t'| \leq \frac{1}{\log N} e^{(\log N)^{\frac{1}{2}}} + 1$. Replacing χ' with a primitive character (by possibly reducing the modulus if necessary) only changes the value of $\chi'(p)$ for primes p dividing q, of which there are at most $\log\log\log\log N$. The sum of the reciprocals of the primes up to $\log\log\log\log N$ is at most $\log\log\log\log\log\log N$ by Mertens' theorem. Thus, the hypotheses for Lemma 4.4.4 are satisfied and we conclude that

$$\frac{\log^{\frac{1}{242}} \log N}{80e^{20}\log^{\frac{1}{2}}\log\log\log N \log^2\log\log\log\log N} \leq A^2$$

4.2 Multiplicative Fourier Reduction

The goal of this section is to prove the following lemma, which we need to reduce the proof of the Erdős discrepancy problem to the multiplicative case.

Lemma 4.2.1. *Let M and N be natural numbers and let $p_1, \dots, p_r$ be the primes less than N. Let f be a function taking values on the unit sphere in some Hilbert space H. Let A be a real number. Suppose that, for all $N' \leq N$,*

$$\frac{1}{M^r} \sum_{b_1,\dots,b_r \leq M} \left|\left| \sum_{n \leq N'} f(p_1^{b_1} \cdot \dots \cdot p_r^{b_r} n) \right|\right|^2 \leq A^2.$$

Then there exists a random multiplicative function g_ξ (meaning for each ξ a point in a probability space there exists a multiplicative function g_ξ) such that

$$\mathbb{E}_\xi \left| \sum_{n \leq N'} g_\xi(n) \right|^2 \leq A^2 + O\left(\frac{N}{M}\right).$$

Proof. We begin with a quick overview of the general strategy. Completely multiplicative functions on the interval $[1, N]$ are functions which satisfy

$$g(p_1^{a_1+b_1} \cdot \cdots \cdot p_r^{a_r+b_r}) = g(p_1^{a_1} \cdot \cdots \cdot p_r^{a_r}) g(p_1^{b_1} \cdot \cdots \cdot p_r^{b_r}).$$

We can think about a multiplicative function as a homomorphism from $(\mathbb{N}, \times)$ to $(\mathbb{C}, \times)$. However, $(\mathbb{N}, \times)$ is not a group since there is no inverse operation. Our goal is to use Fourier analysis so we want to artificially introduce inverses. Thus, we want to think of multiplicative functions as associated to functions

$$G \colon (\mathbb{Z}/M\mathbb{Z})^r \to \mathbb{C}$$

under the identification

$$G(a_1, \ldots, a_r) = g(p_1^{a_1} \cdot \cdots \cdot p_r^{a_r}).$$

With this language, the condition that g is multiplicative corresponds to the condition that G is a homomorphism now from $((\mathbb{Z}/M\mathbb{Z})^r, +)$ to $(\mathbb{N}, \times)$ i.e.

$$G(a_1 + b_1, \ldots, a_r + b_r) = G(a_1, \ldots, a_r) G(b_1, \ldots, b_r).$$

Thus, with this new language, multiplicative functions correspond to group characters. Fourier analysis lets us write any function as a weighted sum of characters. Therefore, we hope to write our function as a weighted sum of multiplicative functions. We are given an L^2 condition on our original function f. The other good reason to use Fourier analysis in this situation is that L^2 conditions on a function transform to L^2 conditions on the Fourier transform. Analytic number theory gives us many tools to recast the same problem in many different languages (e.g. Dirichlet characters, Dirichlet series, Möbius inverse, all different

kinds of Laplace transforms, etc.). However, we hope that the use of a "multiplicative Fourier transform" in this context makes sense for the reasons stated above: we want to write our function as a sum of multiplicative functions which correspond to characters in this setting and we have an L^2 estimate which works well with the Fourier transform. To this end, we begin by rewriting

$$\frac{1}{M^r} \sum_{0 \leq b_1, \ldots, b_r \leq M-1} \left| \left| \sum_{n \leq N'} f(p_1^{b_1} \cdot \cdots \cdot p_r^{b_r} n) \right| \right|^2 \leq A^2$$

in terms of a product of primes

$$\frac{1}{M^r} \sum_{0 \leq b_1, \ldots, b_r \leq M-1} \left| \left| \sum_{p_1^{a_1} \cdots p_r^{a_r} \leq N'} f(p_1^{a_1+b_1} \cdot \cdots \cdot p_r^{a_r+b_r}) \right| \right|^2 \leq A^2. \tag{4.4}$$

As promised, we want to think about f as a function on $(\mathbb{Z}/M\mathbb{Z})^r$. Therefore, we define $F \colon N \to H$ by the formula

$$F(a_1, \ldots, a_r) = f(p_1^{a_1} \cdot \cdots \cdot p_r^{a_r}),$$

whenever $0 \leq a_1, \ldots, a_r \leq M-1$. Therefore

$$F(a_1 + b_1, \ldots, a_r + b_r) = f(p_1^{a_1+b_1} \cdot \cdots \cdot p_r^{a_r+b_r}),$$

unless $a_i + b_i \geq M$ for some i i.e. unless there is some "wrap around". Our next goal is to show that this wrap around does not happen very often. Note that if $p_i^{a_i} \leq N' \leq N$ then $a_i = O(\log N)$. Therefore, for only $O(\log N)$ many choices of $b_i \leq M-1$ is $a_i + b_i \geq M$ or put another way this only happens $O\left(\frac{\log N}{M}\right)$ percent of the time for each i. Therefore,

$$\frac{1}{M^r} \cdot \Big(\text{ \# of } b_1, \ldots, b_r \leq M-1 \text{ such that } a_i + b_i \geq M \text{ for some } i \Big) = O\left(r \frac{\log N}{M} \right),$$

which is $O\left(\frac{N}{M}\right)$ by Chebyshev's theorem or the prime number theorem. Going back to (4.4), we find

$$\frac{1}{M^r} \sum_{0 \leq b_1, \ldots, b_r \leq M-1} \left| \left| \sum_{p_1^{a_1} \cdots p_r^{a_r} \leq N'} F(a_1 + b_1, \ldots, a_r + b_r) \right| \right|^2 \leq A^2 + O\left(\frac{N}{M}\right)$$

The Fourier transform allows us to rewrite F as a sum of characters,

$$F(a_1, \ldots, a_r) = \sum_{\xi_1, \ldots, \xi_r \leq M} \widehat{F}(\xi) e(\xi_1 a_1 + \cdots + \xi_r a_r),$$

where $e(x) = e^{2\pi i x}$. We will also use the notation $\xi \cdot a = \xi_1 a_1 + \cdots + \xi_r a_r$ in which case we have

$$F(a_1, \ldots, a_r) = \sum_{\xi_1, \ldots, \xi_r \leq M} \widehat{F}(\xi) e(\xi \cdot a).$$

Explicitly, this is achieved by the formula

$$\widehat{F}(\xi) = \frac{1}{M^r} \sum_{b_1, \ldots, b_r \leq M} F(b_1, \ldots, b_r) e(-\xi \cdot b).$$

By Plancherel, the Fourier transform of any function Φ satisfies

$$\sum_{\xi_1, \ldots, \xi_r \leq M} \left|\left| \widehat{\Phi}(\xi_1, \ldots, \xi_r) \right|\right|^2 = \frac{1}{M^r} \sum_{b_1, \ldots, b_r \leq M} ||\Phi(b_1, \ldots, b_r)||^2 .$$

We apply this identity with

$$\Phi(b_1, \ldots, b_r) = \sum_{p_1^{a_1} \cdot \cdots \cdot p_r^{a_r} \leq N'} F(a_1 + b_1, \ldots, a_r + b_r).$$

Plugging this in yields

$$\sum_{\xi_1, \ldots, \xi_r \leq M} \left|\left| \widehat{\Phi}(\xi_1, \ldots, \xi_r) \right|\right|^2 \leq A^2 + O\left(\frac{N}{M}\right),$$

with Φ as above. Of course translation Fourier transforms to modulation; if $\tau_a F$ denotes the translate $\tau_a F(b) = F(a + b)$ then

$$\begin{aligned}
\widehat{\tau_a F}(\xi) =& \frac{1}{M^r} \sum_{b_1, \ldots, b_r \leq M} F(b_1 + a_1, \ldots, b_r + a_r) e(-\xi \cdot b) \\
=& \frac{1}{M^r} \sum_{b_1, \ldots, b_r \leq M} F(b_1 + a_1, \ldots, b_r + a_r) e(-\xi \cdot (a + b)) e(\xi \cdot a) \\
=& \widehat{F}(\xi) e(\xi \cdot a),
\end{aligned}$$

where the last step just involves a change of variables replacing $a+b$ with b. Also, the Fourier transform is linear. Therefore, we get

$$\sum_{\xi_1,\ldots,\xi_r\leq M}\left|\left|\widehat{F}(\xi_1,\ldots,\xi_r)\sum_{p_1^{a_1}\cdot\cdots\cdot p_r^{a_r}\leq N'}e(a\cdot\xi)\right|\right|^2\leq A^2+O\left(\frac{N}{M}\right).$$

We notice that $\widehat{F}(\xi_1,\ldots,\xi_r)$ does not depend on a and so we may factor

$$\sum_{\xi_1,\ldots,\xi_r\leq M}\left|\left|\widehat{F}(\xi_1,\ldots,\xi_r)\right|\right|^2\cdot\left|\sum_{p_1^{a_1}\cdot\cdots\cdot p_r^{a_r}\leq N'}e(a\cdot\xi)\right|^2\leq A^2+O\left(\frac{N}{M}\right).$$

For each frequency ξ, define the completely multiplicative function g_ξ by the formula

$$g_\xi(p_i)=e(\xi_i).$$

With this definition,

$$\sum_{p_1^{a_1}\cdot\cdots\cdot p_r^{a_r}\leq N'}e(a\cdot\xi)=\sum_{n\leq N'}g_\xi(n).$$

Therefore,

$$\sum_{\xi_1,\ldots,\xi_r\leq M}\left|\left|\widehat{F}(\xi_1,\ldots,\xi_r)\right|\right|^2\cdot\left|\sum_{n\leq N'}g_\xi(n)\right|^2\leq A^2+O\left(\frac{N}{M}\right).\tag{4.5}$$

We claim that this is actually a weighted *average* of sums of multiplicative functions. By Plancherel again,

$$\sum_{\xi_1,\ldots,\xi_r\leq M}\left|\left|\widehat{F}(\xi_1,\ldots,\xi_r)\right|\right|^2=\frac{1}{M^r}\sum_{b_1,\ldots,b_r\leq M}||F(b_1,\ldots,b_r)||^2.$$

Now recall that f and therefore F takes values on the unit sphere by assumption.

$$=1.$$

So (4.5) tells us that the weighted average of sums of the form

$$\left|\sum_{n\leq N'}g_\xi(n)\right|^2$$

is at most $\leq A^2+O\left(\frac{N}{M}\right)$. □

4.3 The Nonpretentious Case

We begin with a lemma which shows that the Erős discrepancy problem can be solved when g does not have local correlations. This is the rigorous, quantitative analogue of the analysis preformed in (4.1).

Lemma 4.3.1. *Suppose that for some function g taking values in the unit circle and for natural numbers H and N, for all $1 \leq |h| \leq H$, we have*

$$\frac{1}{\log N} \sum_{n \leq N} \frac{g(n)\overline{g(n+h)}}{n} \leq \frac{1}{2H}. \tag{4.6}$$

Then

$$\frac{1}{\log N} \sum_{n \leq N} \frac{1}{n} \left| \sum_{h \leq H} g(n+h) \right|^2 \geq \frac{H}{2} - \frac{H+10}{\log N}. \tag{4.7}$$

Proof. We expand the square

$$\frac{1}{\log N} \sum_{n \leq N} \frac{1}{n} \left| \sum_{h \leq H} g(n+h) \right| = \frac{1}{\log N} \sum_{n \leq N} \frac{1}{n} \sum_{h_1, h_2 \leq H} g(n+h_1)\overline{g(n+h_2)}.$$

Changing variables and shifting the sum by h_1 gives

$$\geq \left(\frac{1}{\log N} \sum_{n \leq N} \frac{1}{n} \sum_{h_1, h_2 \leq H} g(n)\overline{g(n+h_2-h_1)} \right) - \frac{H}{\log N}.$$

When $h_1 = h_2$ we get H different terms all of size 1.

$$\frac{1}{\log N} \sum_{n \leq N} \frac{1}{n} \geq 1 - \frac{10}{\log N}.$$

There are at most H^2 many remaining terms each of which can be bounded by (4.6), for a total error of

$$\#\text{ of terms } \cdot \text{ size of each term } = H^2 \cdot \frac{1}{2H} = \frac{H}{2}.$$

Putting everything together gives (4.7). □

We just showed that we could settle the Erdős discrepancy problem if we could control local correlations. The following result was explicitly used in [HR21] to control correlations of $\lambda(n)\lambda(n+1)$ with a good error term and they mention that this result can be used to control local correlations more generally.

Theorem 4.3.2. *Let N, H and H_0 be real numbers. Let P be a set of primes between H_0 and H with $H_0 \leq H$ and $\log^{\frac{2}{3}} H \cdot \log^2 \log H \leq H_0$. Set*

$$\mathscr{L} = \sum_{p \in P} \frac{1}{p}$$

and suppose $\mathscr{L} > e$ and $\log H \leq \left(\frac{\log N}{\mathscr{L}}\right)^{\frac{1}{2}}$. Let f and g be 1 bounded functions from $\mathbb{N}$ to $\mathbb{C}$. Then

$$\frac{1}{N\mathscr{L}} \left| \sum_{n \sim N} \sum_{\sigma = \pm 1} \sum_{\substack{p \in P \\ p | n}} f(n) g(n + \sigma p) - \sum_{n \sim N} \sum_{\sigma = \pm 1} \sum_{p \in P} \frac{1}{p} f(n) g(n + \sigma p) \right| \lesssim \frac{1}{\sqrt{\mathscr{L}}}.$$

Now we apply Theorem 4.3.2 to the graph where we connect two numbers n and m if $n + h \cdot p = m$.

Corollary 4.3.3. *Let h be a natural number. Let N, H and H_0 be real numbers. Let P be a set of primes between H_0 and H with $H_0 \leq H$ and $\log^{\frac{2}{3}} H \cdot \log^2 \log H \leq H_0$ which does not contain h. Set*

$$\mathscr{L} = \sum_{p \in P} \frac{1}{p}$$

and suppose $\mathscr{L} > e$ and $\log H \leq \left(\frac{\log \frac{N}{h}}{\mathscr{L}}\right)^{\frac{1}{2}}$. Let f and g be 1 bounded functions from $\mathbb{N}$ to $\mathbb{C}$. Then

$$\frac{1}{N\mathscr{L}} \left| \sum_{n \sim N} \sum_{\sigma = \pm 1} \sum_{\substack{p \in P \\ p | n}} f(n) g(n + \sigma p h) - \sum_{n \sim N} \sum_{\sigma = \pm 1} \sum_{p \in P} \frac{1}{p} f(n) g(n + \sigma p h) \right| \lesssim \frac{h}{\sqrt{\mathscr{L}}}$$

One can deduce Corollary 4.3.3 from Theorem 4.3.2 in one of two more or less equivalent ways. Firstly, for each residue class $a \bmod h$, define $f_a(n) = f(h \cdot n + a)$ and similarly $g_a(n) = g(h \cdot n + a)$. Now by the triangle inequality we can bound

$$\frac{1}{N\mathscr{L}} \left| \sum_{n \sim N} \sum_{\sigma = \pm 1} \sum_{\substack{p \in P \\ p|n}} f(n) g(n + \sigma p h) - \sum_{n \sim N} \sum_{\sigma = \pm 1} \sum_{p \in P} \frac{1}{p} f(n) g(n + \sigma p h) \right|$$

by

$$\frac{1}{N\mathscr{L}} \sum_{a \bmod h} \left| \sum_{\substack{n \sim N \\ n = a \bmod h}} \sum_{\sigma = \pm 1} \sum_{\substack{p \in P \\ p|n}} f(n) g(n + \sigma p h) \right.$$
$$\left. - \sum_{\substack{n \sim N \\ n = a \bmod h}} \sum_{\sigma = \pm 1} \sum_{p \in P} \frac{1}{p} f(n) g(n + \sigma p h) \right|.$$

By changing variables, we can write this as

$$= \frac{1}{N\mathscr{L}} \sum_{a \bmod h} \left| \sum_{n \sim \frac{N}{h}} \sum_{\sigma = \pm 1} \sum_{\substack{p \in P \\ p|n}} f(h \cdot n + a) g(h \cdot n + a + \sigma p h) \right.$$
$$\left. - \sum_{n \sim \frac{N}{h}} \sum_{\sigma = \pm 1} \sum_{p \in P} \frac{1}{p} f(h \cdot n + a) g(h \cdot n + a + \sigma p h) \right|.$$

By definition of f_a and g_a this is

$$= \frac{1}{N\mathscr{L}} \sum_{a \bmod h} \left| \sum_{n \sim \frac{N}{h}} \sum_{\sigma = \pm 1} \sum_{\substack{p \in P \\ p|n}} f_a(n) g_a(n + \sigma p h) - \sum_{n \sim \frac{N}{h}} \sum_{\sigma = \pm 1} \sum_{p \in P} \frac{1}{p} f_a(n) g_a(n + \sigma p h) \right|.$$

We are now in a position to apply Theorem 4.3.2 for each a:

$$\lesssim \sum_{a \bmod h} \frac{1}{\sqrt{\mathscr{L}}} = \frac{h}{\sqrt{\mathscr{L}}}.$$

This completes the first possible proof.

The second proof is not so different: one can simply argue that the graph on the natural numbers that one gets by connecting n of size $\sim N$ with $n + h \cdot p$ is isomorphic to h disjoint

copies (one for every residue class) of the graph one gets by connecting every n of size $\sim \frac{N}{h}$ with $n+p$. One can then deduce the appropriate bound on the eigenvalues just using abstract graph theory which Matomäki and Helfgott in turn use to deduce Theorem 4.3.2 in the first place.

In order to use Corollary 4.3.3 to understand local correlations, we need the following theorem. (We remark to the reader that the approach given here also works to circumvent the use of the circle method explicitly in [Tao16b], resulting in a somewhat easier proof).

Theorem 4.3.4 ([MRT15], Theorem 5.1)**.** *Let N and H be natural numbers with $H \lesssim \log N$. Suppose that g_1 and g_2 be 1-bounded functions from $\mathbb{N}$ to $\mathbb{C}$. Suppose g_1 is multiplicative and*

$$\inf_{\substack{|t|\leq 10N \\ q\leq \log^{20} H}} \sum_{p\leq N} \frac{1 - Re\ (g(p)\overline{\chi(p)}p^{-it})}{p} \geq B.$$

Suppose finally that $H \leq e^{100\cdot B}$. Then there exists a set S such that

$$\sum_{\substack{n\notin S \\ n\leq N}} 1 \lesssim \left(e^{-\frac{B}{60}} + \frac{\log\log H}{\log H}\right)\cdot N,$$

and such that

$$\sum_{h\leq H}\left|\sum_{n\leq N} g_1(n)g_2(n+h)\right| \leq H^{1-\frac{1}{500}}N.$$

We remark that the original result in [MRT15] is stated with $e^{-\frac{B}{80}}$ but a cursory glance at the proof of Theorem 5.1 in that paper shows that all that was need was that $80 \geq 3\cdot 20$. There are probably other improvements to the exponent one could find going carefully through [MRT15], but we stick with this for now. Finally, we can combine Theorem 4.3.4 with Corollary 4.3.3 to control local correlations.

Corollary 4.3.5. *Suppose g is a completely multiplicative function taking values on the unit circle. Let N be a sufficiently large natural number. Suppose further that*

$$\inf_{\substack{|t|\leq 10N \\ q\leq \log^{20} H}} \sum_{p\leq N} \frac{1 - Re\ (g(p)\overline{\chi(p)}p^{-it})}{p} \geq B.$$

Then for all $h \leq \log^{\frac{1}{4}} \log N$,

$$\left| \sum_{n \leq N} g(n) g(n+h) \right| \lesssim \left(e^{-\frac{B}{60}} + \log^{\frac{-1}{4}} \log N \right) \cdot N.$$

Proof. By multiplicativity and since g takes values on the unit circle, for any prime p,

$$\begin{aligned} &\mathbb{E}^{\log}_{n \leq N} g(n) \overline{g(n+h)} \\ =&\mathbb{E}^{\log}_{n \leq N} g(pn) \overline{g(pn+ph)}. \end{aligned}$$

By changing variables, this is

$$=\mathbb{E}^{\log}_{n \leq p \cdot N} p \mathbb{1}_{p|n} g(n) \overline{g(n+ph)}.$$

By the almost dilation invariance of logarithmic averages this is

$$=\mathbb{E}^{\log}_{n \leq p \cdot N} p \mathbb{1}_{p|n} g(n) \overline{g(n+ph)} + O\left(\frac{\log p}{\log N} \right).$$

This holds for any prime p. We want to average over some set of primes described in the set up of Corollary 4.3.3. Thus, let $H = \log^{.49} N$ and $H_0 = \exp(\log^{\frac{3}{4}} H)$. Let P be the set of all primes between H_0 and H. Then

$$=\frac{1}{\mathscr{L}} \sum_{p \in P} \mathbb{E}^{\log}_{n \leq p \cdot N} \mathbb{1}_{p|n} g(n) \overline{g(n+ph)} + O\left(\frac{\log \log N}{\log N} \right).$$

Now we can apply Corollary 4.3.3 to the previous sum to conclude that

$$=\frac{1}{\mathscr{L}} \sum_{p \in P} \mathbb{E}^{\log}_{n \leq p \cdot N} \frac{1}{p} g(n) \overline{g(n+ph)} + O\left(\log^{\frac{-1}{4}} \log N \right),$$

where here we have used the fact that $h \lesssim \log^{\frac{1}{4}} \log N$ and Mertens' thereom. Now we are summing over all primes between H_0 and H. By the prime number theorem, the primes have density at least $\gtrsim \frac{1}{\log H}$ in that interval. But by Theorem 4.3.4, outside a small set S, one has a power savings of size $\sim H^{\frac{-1}{500}}$. Thus, by Theorem 4.3.4,

$$=O\left(e^{-\frac{B}{60}} + \log^{\frac{-1}{4}} \log N \right).$$

□

This completes our treatment of the nonpretentious case.

4.4 The Pretentious Case

Now we turn to the pretentious or structured case. Our first goal is to prove that if the Erdős discrepancy problem fails at multiple different scales and therefore (by the previous section) g pretends to be a function of the form $\chi(n)n^{it}$ at many different scales, we can use that information to lower t. We begin with a lemma which can be found on [Taob].

Lemma 4.4.1 ([Taob] Proposition 19)**.** *Suppose that N is a natural number and t is a real number such that $\log|t|$ lies between 0 and $\log^{1.4} N$. Then*

$$\mathbb{D}(1, n^{it}; N)^2 \sim \log\log N$$

Next, we use Lemma 4.4.1 to show that 1 is not too close to $\chi(n)n^{it}$ unless t is small.

Corollary 4.4.2. *Suppose that χ is a Dirichlet character of modulus q, N is a natural number and t is a real number such that $\log|t| + \log q$ lies between 0 and $\log^{1.4} N$. Then*

$$\log\log N \lesssim q^2\mathbb{D}(1, n^{it}\chi(n); N)^2 \tag{4.8}$$

Proof. By the pretentious triangle inequality,

$$\mathbb{D}(1^q, (n^{it}\chi(n))^q; N) \leq q\mathbb{D}(1, n^{it}\chi(n); N).$$

Squaring both sides gives

$$\mathbb{D}(1^q, (n^{it}\chi(n))^q; N)^2 \leq q^2\mathbb{D}(1, n^{it}\chi(n); N)^2.$$

The right hand side is now just as in (4.8). Now the left hand side is

$$\mathbb{D}(1, n^{iqt}; N)$$

to which we apply Lemma 4.4.1, which finishes the proof. □

Finally, we use Corollary 4.4.2 to show that if g is close to both $\chi(n)n^{it}$ and $\chi'(n)n^{it'}$ then t and t' are close together.

Corollary 4.4.3. *Suppose that g is a multiplicative function, χ and χ' are characters of modulus q and q' respectively and t and t' are real numbers such that $\log|t-t'|+\log qq'$ lies between 0 and $\log^{1.4} N$. Then either*

$$(\log\log N)^{\frac{1}{2}} \lesssim qq'\mathbb{D}(g, n^{it}\chi(n); N) \text{ or } (\log\log N)^{\frac{1}{2}} \lesssim qq'\mathbb{D}(g, n^{it'}\chi'(n); N).$$

Proof. By the pretentious triangle inequality,

$$\mathbb{D}(g, n^{it}\chi(n); N) + \mathbb{D}(g, n^{it'}\chi'(n); N) \geq \mathbb{D}(1, \chi(n)\overline{\chi'(n)}n^{i(t-t')}; N).$$

Applying Corollary 4.4.2 finishes the proof. □

Now we arrive at the hardest part of the proof. We have to show that pretentious multiplicative functions actually do satisfy the Erdős discrepancy problem.

Lemma 4.4.4. *Suppose that g is a multiplicative function satisfying, for some H, N, $A \geq 1$, χ, q, t and B that*

$$\sum_{H'\sim H}\sum_{n\in\mathbb{N}} \frac{\left|\sum_{m\leq H'} g(n+m)\right|^2}{n^{1+c}} \leq A^2 H\log N \tag{4.9}$$

and

$$\sum_{p\leq N} \frac{1-Re\ g(p)\overline{\chi}(p)p^{-it}}{p} \leq B \tag{4.10}$$

where χ is a primitive, non-principal Dirichlet character mod q and $c=\frac{1}{\log N}$. Then for any k,

$$\min\left(k-1, \frac{\log H}{\log q}\right)\cdot e^{-2B-20}\cdot\frac{1}{40\log\log q} - \ Error\ \leq A^2$$

where the error term smaller than the main term as long as

1. $|t| \leq \frac{1}{H^3}N^{\frac{1}{H^2}}$.
2. $\frac{\log N}{q^k} \geq 20$.
3. $e^{10}q^k\log q^k e^{B^{\frac{1}{2}}} \leq e^{(\log\log N)^{\frac{1}{2}}}$.
4. $H^4 \leq 2^k$.

5. N and H are sufficiently large i.e. larger than some constant.

In particular, for $q \leq (\log\log\log N)^{20}$, $|t| \leq \frac{1}{\log N} e^{\log^{\frac{1}{2}} N}$, $k = \left(\frac{\log\log N}{2\log\log\log\log N}\right)^{\frac{1}{2}}$, $\log H = \frac{1}{4}k$ *and* $B = \frac{60}{242} \cdot \log\log\log N + \log\log\log\log\log\log N$ *we get that*

$$\frac{\log^{\frac{1}{242}} \log N}{80e^{20} \log^{\frac{1}{2}} \log\log\log N \log^2 \log\log\log\log N} \leq A^2$$

Proof. Define $\tilde{\chi}$ to be the completely multiplicative function such that $\tilde{\chi}(d) = d$ for $(d, q) = 1$ and $\tilde{\chi}(p) = g(p)p^{-it}$ for $p|q$. Set $h(n) = \overline{\tilde{\chi}(n)} n^{-it} g(n)$ i.e. define h such that

$$g(n) = \tilde{\chi}(n) n^{it} h(n).$$

(It is perhaps worth remarking that with this definition $h(p) = 1$ for p dividing q). Now plugging this into (4.9) yields

$$\sum_{H' \sim H} \sum_{n \in \mathbb{N}} \frac{\left|\sum_{m \leq H'} (n+m)^{it} \tilde{\chi}(n+m) h(n+m)\right|^2}{n^{1+c}} \leq A^2 H \log N. \tag{4.11}$$

By Taylor expansion

$$(n+m)^{it} = n^{it} + itn^{it} \cdot \left(\frac{m}{n}\right) + O\left(\left(\frac{m}{n}\right)^2\right).$$

Since $m \leq H' \leq 2H$, we can see that

$$\left|(n+m)^{it} \tilde{\chi}(n+m) h(n+m) - n^{it} \tilde{\chi}(n+m) h(n+m)\right| \leq 2\frac{H \cdot |t|}{n} + O\left(\frac{H^2 t^2}{n^2}\right).$$

Plugging this into (4.11), we get

$$\sum_{H' \sim H} \sum_{n \in \mathbb{N}} \frac{\left|\sum_{m \leq H'} \tilde{\chi}(n+m) h(n+m)\right|^2}{n^{1+c}} \leq A^2 H \log N + E_1 \tag{4.12}$$

where

$$\begin{aligned} E_1 &= \sum_{H' \sim H} \sum_{n \in \mathbb{N}} \frac{4H^2 \min(1, 2H \cdot \frac{|t|}{n} + O\left(\frac{H^2 t^2}{n^2}\right))}{n^{1+c}} \\ &= \sum_{n \in \mathbb{N}} \frac{8H^3 \min(1, 2H \cdot \frac{|t|}{n} + O\left(\frac{H^2 t^2}{n^2}\right))}{n^{1+c}}. \end{aligned}$$

We remark that so long as $|t| \leq \frac{1}{H^3} N^{1/H^2}$ thenby splitting up the sum we find

$$\leq \sum_{n \leq N^{\frac{1}{H^2}}} \frac{8H^3}{n^{1+c}} + \sum_{n > N^{\frac{1}{H^2}}} 16H \frac{|t|}{n^{2+c}} + O\left(\frac{H^5 t^2}{n^{3+c}}\right).$$

By the integral test this is at most,

$$\leq 8H^3 \cdot \log N^{\frac{1}{H^2}} + 16H^4 |t| N^{-\frac{1}{H^2}} + O\left(H^5 t^2 N^{-\frac{2}{H^2}}.\right)$$

which is no larger than $A^2 H \log N$ for N large enough.

Next, we say that a residue class a mod q^k is good if $a + m$ is not divisible by p^k for any $m \leq 2H$ and for any p dividing q. For any residue class b mod q^k and for any natural number $n = b$ mod q^k we check that

$$\tilde{\chi}(b) = \tilde{\chi}(n).$$

To see this, set $d = (n, q^k)$. By the Euclidean algorithm, we also have $d = (b, q^k)$. If p divides q then if p^j is the highest power of p dividing n, because $n = b$ mod q^k and p^k does not divide bm we conclude that $j < k$. Since p^k divides q^k, we conclude that p^j divides d by definition of the greatest common divisor. Thus, since j was the highest power of p dividing n, p does not divide $\frac{n}{d}$. Since p was an arbitrary prime dividing q, we conclude that $\frac{n}{d}$ is coprime to q. Thus by multiplicativity

$$\tilde{\chi}(n) = \tilde{\chi}(d) \cdot \tilde{\chi}\left(\frac{n}{d}\right). \tag{4.13}$$

Since $\frac{n}{d}$ is coprime to q, by definition of $\tilde{\chi}$,

$$= \tilde{\chi}(d) \cdot \chi\left(\frac{n}{d}\right).$$

Since χ is a Dirichlet character, it is q periodic so this none other than

$$= \tilde{\chi}(d) \cdot \chi\left(\frac{b}{d}\right).$$

Since the right hand side no longer depends on n, this yields

$$=\tilde{\chi}(b).$$

By definition of a good residue class, for any $m \leq 2H$ and for any $n = a \bmod q^k$,

$$\tilde{\chi}(n+m) = \tilde{\chi}(a+m). \tag{4.14}$$

Returning for (4.12), because the summand is nonnegative, we may restrict our sum to only good residue classes

$$\sum_{H' \sim H} \sum_{a \text{ good}} \sum_{n=a \bmod q^k} \frac{\left|\sum_{m \leq H'} \tilde{\chi}(n+m)h(n+m)\right|^2}{n^{1+c}} \leq A^2 H \log N + E_1.$$

Applying (4.14) yields

$$\sum_{H' \sim H} \sum_{a \text{ good}} \sum_{n=a \bmod q^k} \frac{\left|\sum_{m \leq H'} \tilde{\chi}(a+m)h(n+m)\right|^2}{n^{1+c}} \leq A^2 H \log N + E_1.$$

By Cauchy Schwartz

$$\begin{aligned} &\sum_{H' \sim H} \sum_{a \text{ good}} \left| \sum_{n=a \bmod q^k} \frac{\sum_{m \leq H'} \tilde{\chi}(a+m)h(n+m)}{n^{1+c}} \right|^2 \\ &\leq \sum_{H' \sim H} \sum_{a \text{ good}} \left(\sum_{n=a \bmod q^k} \frac{\left|\sum_{m \leq H'} \tilde{\chi}(n+m)h(n+m)\right|^2}{n^{1+c}} \right) \cdot \left(\sum_{n=a \bmod q^k} \frac{1}{n^{1+c}} \right) \\ &\leq \left(HA^2 \log N + E_1\right) \cdot \left(\frac{\log N}{q^k} + 10 \right). \end{aligned} \tag{4.15}$$

Set

$$E_1' = E_1 \cdot \left(\frac{\log N}{q^k} + 10 \right)$$

and

$$E_2 = 10 \cdot \left(HA^2 \log N + E_1\right).$$

Then plugging this notation into (4.15) gives

$$\sum_{H' \sim H} \sum_{a \text{ good}} \left| \sum_{n=a \bmod q^k} \frac{\sum_{m \leq H'} \tilde{\chi}(a+m)h(n+m)}{n^{1+c}} \right|^2 \leq HA^2 \frac{\log^2 N}{q^k} + E_1' + E_2. \tag{4.16}$$

It is worth remarking that if $\frac{\log N}{q^k} \geq 20$ then

$$H\frac{\log^2 N}{q^k} \geq E_2$$

so it is appropriate to think of E_2 as a small error term. By applying Fubini to (4.16), we obtain

$$\sum_{H' \sim H} \sum_{a \text{ good}} \left| \sum_{m \leq H'} \tilde{\chi}(a+m) \sum_{n=a \bmod q^k} \frac{h(n+m)}{n^{1+c}} \right|^2 \leq HA^2 \frac{\log^2 N}{q^k} + E_1' + E_2. \tag{4.17}$$

We turn our attention to the sum

$$\sum_{n \in \mathbb{N}} \mathbb{1}_{n=a \bmod q^k} \cdot \frac{h(n+m)}{n^{1+c}},$$

for some (momentarily) fixed choice of a and m. By a change of variables, this is

$$= \sum_{n>m} \mathbb{1}_{n=a+m \bmod q^k} \cdot \frac{h(n)}{(n-m)^{1+c}}.$$

By Taylor expansion,

$$= \sum_{n>m} \left(\mathbb{1}_{n=a+m \bmod q^k} \cdot \frac{h(n)}{n^{1+c}} + r_1 \right). \tag{4.18}$$

for some r_1 of size at most

$$|r_1| \leq \frac{m}{n^2} + O\left(\frac{m^2}{n^3}\right).$$

Summing over n yields

$$\sum_{n>m} |r_1| \leq 1 + O\left(\frac{1}{m}\right).$$

The first m terms,

$$\sum_{n \leq m} \mathbb{1}_{n=a+m \bmod q^k} \cdot \frac{h(n)}{n^{1+c}}$$

also contribute no more than $1 + \frac{\log m + 10}{q^k}$ to the total sum. Therefore, (4.18) equals

$$= \left(\sum_{n \in \mathbb{N}} \mathbb{1}_{n=a+m \bmod q^k} \cdot \frac{h(n)}{n^{1+c}} \right) + r_1' \tag{4.19}$$

for some r_1' satisfying

$$|r_1'| \leq 2 + \frac{\log m + 10}{q^k} + O\left(\frac{1}{m}\right).$$

Now if $a+m$ is coprime to q then by the theory of the Fourier transform applied to the group $(\mathbb{Z}/q^k\mathbb{Z})^\times$, we can write

$$\mathbb{1}_{n=a+m \bmod q^k} = \frac{1}{\varphi(q^k)}\chi_0 + \sum_{\chi_1} c_{\chi_1} \cdot \chi_1(n)$$

where χ_0 is the principal character, φ is Euler's totient function, the sum is over non-principal characters χ_1 and the c_{χ_1} are coefficients with the property that

$$\sum_{\chi_1} |c_{\chi_1}|^2 \leq \frac{1}{\varphi(q^k)} \tag{4.20}$$

by Bessel's inequality. If $a+m$ is not coprime to q, say if $d = (a+m, q^k)$ then by change of variables, (4.19) redues to

$$\frac{h(d)}{d^{1+c}} \sum_{n\in\mathbb{N}} \mathbb{1}_{n=\frac{a+m}{d} \bmod \frac{q^k}{d}} \cdot \frac{h(n)}{n^{1+c}}.$$

Now the function

$$\mathbb{1}_{n=\frac{a+m}{d} \bmod \frac{q^k}{d}}$$

can similarly be written as a weighted sum of Dirichlet characters. This motivates our study of sums of the form

$$\sum_{n\in\mathbb{N}} \frac{h(n)\chi_1(n)}{n^{1+c}}. \tag{4.21}$$

Using the usual Euler product trick, which applies because we have a sum of completely multiplicative functions, we may write (4.21) as a product

$$= \prod_p \left(1 + \frac{h(p)\chi_1(p)}{p^{1+c}} + \left(\frac{h(p)\chi_1(p)}{p^{1+c}}\right)^2 + \cdots\right)$$

which, by the geometric series formula yields

$$= \prod_p \left(1 - \frac{h(p)\chi_1(p)}{p^{1+c}}\right)^{-1}. \tag{4.22}$$

Multiplying and dividing by $L(\chi_1, 1+c)$ gives

$$=L(\chi_1, 1+c)\prod_p \left(1-\frac{h(p)\chi_1(p)}{p^{1+c}}\right)^{-1}\left(1-\frac{\chi_1(p)}{p^{1+c}}\right). \tag{4.23}$$

This step makes sense because h pretends to be the function 1 so $L(\chi_1, 1+c)$ should be kind of like the main term in (4.22) and in analysis it is often profitable to to subtract off, or in this case divide off the main term and then try to bound the remainder. it is classical that

$$|L(\chi_1, 1+c)| \leq \log q^k + 10.$$

Thus (4.23) is in absolute value at most

$$\leq \log q^k \left|\prod_p \left(1-\frac{h(p)\chi_1(p)}{p^{1+c}}\right)^{-1}\left(1-\frac{\chi_1(p)}{p^{1+c}}\right)\right|. \tag{4.24}$$

Applying the definition of the logarithm to (4.24), we get

$$=\log q^k \left|\prod \exp\left(\log\left(1-\frac{h(p)\chi_1(p)}{p^{1+c}}\right)^{-1}\left(1-\frac{\chi_1(p)}{p^{1+c}}\right)\right)\right|$$

which by the log rules simplifies to

$$=\log q^k \left|\exp\left(\sum_p \log\left(\frac{\chi_1(p)}{p^{1+c}}\right) - \log\left(1-\frac{h(p)\chi_1(p)}{p^{1+c}}\right)\right)\right|. \tag{4.25}$$

By Taylor's theorem, (4.25) is at most

$$\leq \log q^k \exp\left(\sum_p \frac{|\chi - 1(p)\cdot(1-h(p))|}{p^{1+c}} + r_2\right) \tag{4.26}$$

for some

$$|r_2| \leq 5\sum_p \frac{1}{p^2}.$$

The sum of the reciprocals of the squares of the primes is clearly some finite absolute constant less than $\frac{\pi^2}{6}$, so (4.26) is at most

$$\leq e^9 \log q^k \exp\left(\sum_p \frac{|1-h(p)|}{p^{1+c}}\right) \tag{4.27}$$

where we have also used that $\chi_1(p)$ is a unit or 0 which is therefore absorbed by the absolute values. Since $h(p)$ takes values on the unit circle and because, for any z near 1, $(1-\text{Re } z)^{\frac{1}{2}} \sim \text{Im } z$ essentially because $1-\cos\theta$ does not vanish to 2^{nd} order, we have

$$|1-h(p)| \leq 2 \cdot (1-\text{Re } h(p))^{\frac{1}{2}}.$$

Plugging this into (4.27) is at most

$$e^9 \log q^k \exp\left(2\sum_p \frac{(1-\text{Re } h(p))^{\frac{1}{2}}}{p^{1+c}}\right).$$

By Cauchy Schwarz and Mertens' theorem,

$$\leq e^9 \log q^k \exp\left(2\left(\sum_p \frac{1}{p^{1+c}}\right)^{\frac{1}{2}}\left(\sum_p \frac{1-\text{Re } h(p)}{p^{1+c}}\right)^{\frac{1}{2}}\right). \tag{4.28}$$

The sum over the reciprocals of the primes is bounded by Mertens' theorem by $\log\log N + 1$ for N large enough which means (4.28) is bounded by

$$\leq e^{10} \log q^k \exp\left(2(\log\log N)^{\frac{1}{2}}\left(\sum_p \frac{1-\text{Re } h(p)}{p^{1+c}}\right)^{\frac{1}{2}}\right). \tag{4.29}$$

We can replace the sum

$$\sum_p \frac{1-\text{Re } h(p)}{p^{1+c}}$$

by a sum over only those primes less than N

$$\sum_{p\leq N} \frac{1-\text{Re } h(p)}{p^{1+c}}$$

at a cost of at most

$$\sum_{p>N} \frac{1}{p^{1+c}} \leq \frac{10}{\log N}$$

by the integral test. By assumption (4.10),

$$\sum_{p\leq N} \frac{1-\text{Re } h(p)}{p^{1+c}} \leq B.$$

Thus, (4.29) is bounded by

$$\leq e^{10} \log q^k \exp\left(2(\log\log N)^{\frac{1}{2}}(B + \frac{10}{\log N})^{\frac{1}{2}}\right).$$

Altogether, this says

$$\sum_{n\in\mathbb{N}} \frac{\chi_1(n)h(n)}{n^{1+c}} \leq e^{10} \log q^k \exp\left(2(\log\log N)^{\frac{1}{2}}(B + \frac{10}{\log N})^{\frac{1}{2}}\right) \tag{4.30}$$

for any non-principal character χ_1 of modulus dividing q^k.

Alternatively, we can lower bound the similar sum where χ_1 is replaced by a principal character χ_0 of modulus d dividing q^k:

$$\sum_{n\in\mathbb{N}} \frac{\chi_0(n)h(n)}{n^{1+c}}. \tag{4.31}$$

Again by the Euler product trick we used in transforming (4.21) to (4.22), we find (4.31) is equal to

$$= \prod_p \left(1 - \chi_0(p)h(p)p^{1+c}\right)^{-1}. \tag{4.32}$$

Recall that $\chi_0(p) = 0$ for p divding d and $\chi_0(p) = 1$ for p not dividing d by definition of a principal character. Thus, we may rewrite (4.32) as

$$= \prod_{p\nmid d} \left(1 - \frac{h(p)}{p^{1+c}}\right)^{-1}. \tag{4.33}$$

Define the singular series $\mathfrak{S}$ by the formula

$$\mathfrak{S} = \prod_p \left(1 - \frac{h(p)}{p^{1+c}}\right)^{-1}. \tag{4.34}$$

Using this definition, we may rewrite (4.33) as

$$\mathfrak{S} \cdot \prod_{p|d} \left(1 - \frac{h(p)}{p^{1+c}}\right).$$

Note that $h(p) = 1$ for p dividing q by definition of h so this simplifies to

$$\sum_{n\in\mathbb{N}} \frac{\chi_0(n)h(n)}{n^{1+c}} = \mathfrak{S} \prod_{p|d} \left(1 - \frac{1}{p^{1+c}}\right). \tag{4.35}$$

Next, we lower bound $|\mathfrak{S}|$. To get a lower bound on $|\mathfrak{S}|$, we attempt to upper bound

$$\left|\frac{\zeta(1+c)}{\mathfrak{S}}\right|.$$

We remark that this analysis will be similar to the argument (4.21). By the Euler product expansion, this is

$$\left|\prod_p \left(1-\frac{1}{p^{1+c}}\right)^{-1}\left(1-\frac{h(p)}{p^{1+c}}\right)\right|.$$

Applying the definition of the logarithm yields

$$\left|\exp\left(\sum_p \log\left(1-\frac{h(p)}{p^{1+c}}\right)-\log\left(1-\frac{1}{p^{1+c}}\right)\right)\right|.$$

By Taylor expanding the logarithms just as in (4.26), we find

$$\leq e^9\left|\exp\left(\sum_p \frac{1-h(p)}{p^{1+c}}\right)\right|. \tag{4.36}$$

Now comes the crucial difference from how we bounded (4.21): since there is no absolute value on $1-h(p)$, we can simply note that $|e^z| = e^{\mathrm{Re}\,(z)}$ for any complex number z. Applying this to (4.36) yields

$$= e^9 \exp\left(\sum_p \frac{1-\mathrm{Re}\ h(p)}{p^{1+c}}\right).$$

As before, we can crudely bound the sum over $p \geq N$ to conclude that for N large enough

$$\leq e^{10} \exp\left(\sum_{p\leq N} \frac{1-\mathrm{Re}\ h(p)}{p^{1+c}}\right).$$

By assumption (4.10), this is bounded by

$$\leq e^{B+10}.$$

Since $\zeta(1+c) \geq \log N - 10$,

$$|\mathfrak{S}| \geq e^{-B-10}(\log N - 10). \tag{4.37}$$

Plugging (4.37) into (4.35), we find

$$\left|\sum_{n\in\mathbb{N}}\frac{\chi_0(n)h(n)}{n^{1+c}}\right| \geq e^{-B-10}(\log N-10)\cdot\prod_{p|d}\left(1-\frac{1}{p^{1+c}}\right). \tag{4.38}$$

By Taylor expansion,

$$\frac{1}{p^{1+c}} = \frac{1}{p}+O\left(\frac{1}{\log N}\right).$$

Plugging this into (4.38) produces

$$\left|\sum_{n\in\mathbb{N}}\frac{\chi_0(n)h(n)}{n^{1+c}}\right| \geq e^{-B-10}(\log N-10)\cdot\prod_{p|d}\left(1-\frac{1}{p}\right)+r_3. \tag{4.39}$$

for some r_3 satisfying

$$|r_3| = O\left(e^{-B}q\right).$$

Recall that φ, Euler's totient function, is a multiplicative and $\varphi(p)=p-1$. Thus, we may simplify the product over prime divisors of d as

$$\begin{aligned}\prod_{p|d}\left(1-\frac{1}{p}\right) =& \prod_{p|d}\frac{p}{p-1}\\ =& \prod_{p|d}\frac{p}{\varphi(p)}\\ =& \frac{d}{\varphi(d)}.\end{aligned} \tag{4.40}$$

Plugging (4.40) into (4.39) gives

$$\left|\sum_{n\in\mathbb{N}}\frac{\chi_0(n)h(n)}{n^{1+c}}\right| \geq e^{-B-10}(\log N-10)\cdot\frac{d}{\varphi(d)}+r_3. \tag{4.41}$$

By plugging (4.41), (4.30) and the simple bound (4.20) into (4.17) we find

$$\sum_{H'\sim H}\sum_{a \text{ good}}\left|\sum_{m\leq H'}\tilde{\chi}(a+m)\right|^2\cdot e^{-2B-20}\frac{\log^2 N}{q^{2k}} \leq HA^2\frac{\log^2 N}{q^k}+E_1'+E_2+E_3$$

where

$$\begin{aligned}E_3 =& \left(e^{10}\log q^k\exp\left(2(\log\log N)^{\frac{1}{2}}B^{\frac{1}{2}}\right)+r_3\right)\cdot e^{-B-10}\frac{\log N}{q^k}H\\ &+q^k\left(2e^{10}\log q^k\exp\left((\log\log N)^{\frac{1}{2}}B^{\frac{1}{2}}\right)\right)^2.\end{aligned}$$

Note that for E_3 to be smaller than the main term it suffices to ask that

$$e^{10}q^k \log q^k \exp(2B^{\frac{1}{2}}) \leq \exp\left((\log\log N)^{\frac{1}{2}}\right)$$

We conclude that

$$\sum_{H'\sim H} \sum_{a \text{ good}} \left| \sum_{m_i \leq H'} \tilde{\chi}(a+m) \right|^2 \leq e^{2B+20} H A^2 q^k + E_4 \tag{4.42}$$

where

$$E_4 = \frac{E_1' + E_2 + E_3}{\log^2 N} q^{2k} e^{2B+20}.$$

Now if $q = 1$ then $\tilde{\chi} = 1$ and we obtain the bound

$$H^3 \leq e^{2B+20} H A^2 + E_4$$

or

$$\left((H^2 - E_4) \cdot e^{-2B-20}\right)^{\frac{1}{2}} \leq A.$$

In the rest of the argument we turn out attention to the case $q \neq 1$. Returning to (4.42) and expanding the square, we find

$$\sum_{H'\sim H} \sum_{a \text{ good}} \sum_{m_i \leq H'} \tilde{\chi}(a+m_1)\overline{\tilde{\chi}(a+m_2)} \leq e^{2B+20} H A^2 q^k + E_4.$$

Setting $d_i = (a+m_i, q^k)$, which since a is good is also $(a+m_i, q^{k-1})$ yields

$$\sum_{H'\sim H} \sum_{a \text{ good}} \sum_{\substack{m_i \leq H' \\ d_i = (a+m_i, q^k)}} \tilde{\chi}(a+m_1)\overline{\tilde{\chi}(a+m_2)} \leq e^{2B+20} H A^2 q^k + E_4. \tag{4.43}$$

As in (4.13), we can rewrite $\tilde{\chi}(a+m_i)$ as $\tilde{\chi}(d_i)\tilde{\chi}\left(\frac{a+m_i}{d_i}\right)$ and as before $\tilde{\chi}\left(\frac{a+m_i}{d}\right) = \chi\left(\frac{a+m_i}{d}\right)$. Plugging this into (4.43) produces

$$\sum_{H'\sim H} \sum_{a \text{ good}} \sum_{\substack{m_i \leq H' \\ d_i = (a+m_i, q^k)}} \tilde{\chi}(d_1)\overline{\tilde{\chi}(d_2)}\, \chi\left(\frac{a+m_1}{d_1}\right) \overline{\chi\left(\frac{a+m_2}{d_2}\right)}$$

$$\leq e^{2B+20} H A^2 q^k + E_4.$$

Pulling out the sum in d_i gives

$$\sum_{d_i|q^{k-1}} \tilde{\chi}(d_1)\overline{\tilde{\chi}(d_2)} \sum_{H'\sim H} \sum_{a \text{ good}} \sum_{\substack{m_i\leq H' \\ d_i=(a+m_i,q^k)}} \chi\left(\frac{a+m_1}{d_1}\right)\overline{\chi\left(\frac{a+m_2}{d_2}\right)} \tag{4.44}$$
$$\leq e^{2B+20}HA^2q^k + E_4.$$

Right now, our sum over residue classes mod q^k is restricted to only those good residue classes. Now we wish to add in all those "bad" residue classes a for which there exists $m \leq 2H$ such that $a+m$ is divisible by p^k for some p dividing q. For each prime p, there $\frac{q^k}{p^k}$ residue classes which are divisible by p^k and therefore there are at most $2H\frac{q^k}{p^k}$ many residue classes a such that there exists $m \leq 2H$ such that $a+m$ is divisible by p^k. Therefore, the number of bad residue classes is at most

$$\begin{aligned} 2H \cdot \sum_{p|q} &\leq 2H\frac{q^k}{2^k}\sum_{n\in\mathbb{N}}\left(\frac{2}{n}\right)^k \\ &\leq 20H\frac{q^k}{2^k}. \end{aligned}$$

Therefore, we can bound

$$\begin{aligned} \sum_{d_i|q^{k-1}} \sum_{H'\sim H} \sum_{a \text{ bad}} \sum_{\substack{m_i\leq H' \\ d_i=(a+m_i,q^k)}} 1 &\leq 20H^2 \cdot \frac{q^k}{2^k} \cdot \sum_{d|q^{k-1}} \sum_{\substack{m_i\leq H' \\ d_i=(a+m_i,q^k)}} 1 \\ &\leq 20H^4 \cdot \frac{q^k}{2^k}, \end{aligned}$$

where in the last step we used that for each (m_1, m_2) there is at most one choice of (d_1, d_2) satisfying $d_i = (a+m_i, q^k)$. Plugging this into (4.44) yields

$$\sum_{d_i|q^{k-1}} \tilde{\chi}(d_1)\overline{\tilde{\chi}(d_2)} \sum_{H'\sim H} \sum_{a \bmod q^k} \sum_{\substack{m_i\leq H' \\ d_i=(a+m_i,q^k)}} \chi\left(\frac{a+m_1}{d_1}\right)\overline{\chi\left(\frac{a+m_2}{d_2}\right)}$$
$$\leq e^{2B+20}H^2A^2q^k + E_4 + E_5$$

where

$$E_5 \leq 20H^4 \cdot \frac{q^k}{2^k}.$$

We remark that this is smaller than the main term if $H^4 \leq 2^k$ i.e. $\log H \ll k$. We also remark that unless $\frac{a+m_i}{d_i}$ is coprime to q that $\chi\left(\frac{a+m_i}{d_i}\right)$ is zero so instead of summing over those m_i such that $(a+m_i, q^k) = d_i$ we can instead sum over all m_i such that $d_i | a+m$; after all, the only time $\chi\left(\frac{a+m_i}{d_i}\right)$ contributes to the sum for any choice of $d_i | a+m_i$ is when $\frac{a+m_i}{d}$ is coprime to q which implies $(a+m_i, q^k) = d_i$. Thus

$$\sum_{d_i | q^{k-1}} \tilde{\chi}(d_1)\overline{\tilde{\chi}(d_2)} \sum_{H' \sim H} \sum_{a \bmod q^k} \sum_{\substack{m_i \leq H' \\ d_i | a+m_i}} \chi\left(\frac{a+m_1}{d_1}\right)\overline{\chi\left(\frac{a+m_2}{d_2}\right)} \tag{4.45}$$
$$\leq e^{2B+20} H A^2 q^k + E_4 + E_5.$$

Momentarily fix some choice d_1, d_2, H', m_1 and m_2. We will try to show that if $d_1 \neq d_2$ then

$$\sum_{\substack{a \bmod q^k \\ d_i | a+m_i}} \chi\left(\frac{a+m_1}{d_1}\right)\overline{\chi\left(\frac{a+m_2}{d_2}\right)} = 0. \tag{4.46}$$

By the theory of the Fourier transform applied to the group $(\mathbb{Z}/q\mathbb{Z}, +)$, we can write

$$\chi(n) = \sum_{\xi \leq q} c_\xi e\left(\frac{\xi}{q} \cdot n\right),$$

and since χ is primitive, we claim the c_ξ are zero unless ξ is a unit in $(\mathbb{Z}/q\mathbb{Z}, \times)$ (see, for instance, Theorem 4.16 of [Eve13] or equation 3.9 and the remarks after equation 3.12 in [IK04]). Therefore, by substitution,

$$\mathbb{1}_{d_i | a+m_i} \cdot \chi\left(\frac{a+m_i}{d_i}\right) = \mathbb{1}_{d_i | a+m_i} \cdot \sum_{\xi \leq q} c_\xi e\left(\frac{\xi}{q} \cdot \frac{a+m_i}{d_i}\right).$$

Thinking of a as the "variable" and m_i and d_i as constants, we could also write

$$\mathbb{1}_{d_i | a+m_i} \cdot \chi\left(\frac{a+m_i}{d_i}\right) = \mathbb{1}_{d_i | a+m_i} \cdot \sum_{\xi \leq q} c_{\xi, m_i, d_i} e\left(\frac{\xi}{q d_i} \cdot a\right),$$

by the rules for exponents where

$$c_{\xi, m_i, d_i} = c_\xi \cdot e\left(\frac{\xi}{q} \cdot \frac{m_i}{d_i}\right).$$

In particular, we still have that $c_{\xi,m_i,d_i} = 0$ unless $(q,\xi) = 1$. Thus, if d_1 and d_2 are different then $\mathbb{1}_{d_1|a+m_1} \cdot \chi\left(\frac{a+m_1}{d_1}\right)$ and $\mathbb{1}_{d_2|a+m_2} \cdot \chi\left(\frac{a+m_2}{d_2}\right)$ can be written as sums of additive characters with different frequencies. Since q^k is a multiple of both qd_1 and qd_2, we conclude that (4.46) holds.

Naively combining (4.46) and (4.45), we find that

$$\sum_{d|q^{k-1}} \tilde{\chi}(d)\overline{\tilde{\chi}(d)} \sum_{H' \sim H} \sum_{a \bmod q^k} \sum_{\substack{m_i \leq H' \\ d|a+m_i}} \chi\left(\frac{a+m_1}{d}\right) \overline{\chi\left(\frac{a+m_2}{d}\right)}$$
$$\leq e^{2B+20} H A^2 q^k + E_4 + E_5.$$

Now we cancel the $\tilde{\chi}(d)$ and $\overline{\tilde{\chi}(d)}$ and write the innermost sum as a square.

$$\sum_{d|q^{k-1}} \sum_{H' \sim H} \sum_{a \bmod q^k} \left| \sum_{\substack{m \leq H' \\ d|a+m}} \chi\left(\frac{a+m}{d}\right) \right|^2 \leq e^{2B+20} H A^2 q^k + E_4 + E_5.$$

Since the summand is now nonnegative, we may restrict our attention those values of d for which d is less than $\frac{H}{4}$.

$$\sum_{\substack{d|q^{k-1} \\ d < \frac{H}{4}}} \sum_{H' \sim H} \sum_{a \bmod q^k} \left| \sum_{\substack{m \leq H' \\ d|m}} \chi\left(\frac{a+m}{d}\right) \right|^2 \leq e^{2B+20} H A^2 q^k + E_4 + E_5. \tag{4.47}$$

Fix, for the moment, some value of $d < \frac{H}{4}$ diving q^{k-1}, some a mod q^k and some value of H' between H and $2H$. Then usually the sums

$$\sum_{\substack{m \leq H' \\ d|a+m}} \chi\left(\frac{a+m}{d}\right) \tag{4.48}$$

and

$$\sum_{\substack{m \leq H'+d \\ d|a+m}} \chi\left(\frac{a+m}{d}\right) \tag{4.49}$$

differ by a complex number on the unit circle because there is precisely one m between H' and $H'+d$ such that $d|a+m$ and for that value of m, we know that $\chi\left(\frac{a+m}{d}\right)$ will appear in

the second sum but not the first. Occasionally, $\chi\left(\frac{a+m}{d}\right)$ is not a unit length complex number but that happens only when $\left(q, \frac{a+m}{d}\right) \neq 1$. Classically (for instance see [RS62] Theorem 15), we can bound

$$\varphi(q) \geq \frac{q}{10 \log \log q}.$$

Therefore, for each $H'$and d, for at least $\frac{q^k}{10 \log \log q}$ many values of a, (4.48) and (4.49) do differ by a unit length complex number. By the triangle inequality, if z and w are complex numbers,

$$|z| + |w| \geq |z - w|.$$

Squaring both sides, one finds that

$$|z|^2 + |w|^2 \geq |z - w|^2.$$

Therefore, if (4.48) and (4.49) do differ by a unit complex number,

$$\left| \sum_{\substack{m \leq H' \\ d | a+m}} \chi\left(\frac{a+m}{d}\right) \right|^2 + \left| \sum_{\substack{m \leq H'+d \\ d | a+m}} \chi\left(\frac{a+m}{d}\right) \right|^2 \geq 1.$$

Plugging everything in to (4.47),

$$\sum_{\substack{d | q^{k-1} \\ d < \frac{H}{4}}} \frac{q^k}{10 \log \log q} \cdot \left(\frac{H}{2} - d\right) \leq e^{2B+20} H A^2 q^k + E_4 + E_5.$$

Since there are at least $\min\left(k-1, \log_q \frac{H}{4}\right)$ many valid choices for d, this simplifies to

$$\min\left(k-1, \frac{\log H}{\log q}\right) \cdot e^{-2B-20} \cdot \frac{1}{40 \log \log q} - E_4' - E_5' - E_6 \leq A^2$$

where

$$E_4' = E_4 \cdot q^{-k} H^{-1} e^{-2B-20}, \quad E_5' = E_5 \cdot q^{-k} H^{-1} e^{-2B-20}$$
$$\text{and } E_6 = \frac{\log 4}{\log q} \cdot e^{-2B-20} \cdot \frac{1}{40 \log \log q}.$$

As long as $H \gg 1$, the main term dominates the E_6 error term. □

This completes the proof of Theorem 4.1.1.

BIBLIOGRAPHY

[Bom71] E. Bombieri, *A note on the large sieve*, Acta Arith. **18** (1971), 401–404, available at https://doi.org/10.4064/aa-18-1-401-404. MR286773

[Bou13a] J. Bourgain, *Möbius-Walsh correlation bounds and an estimate of Mauduit and Rivat*, J. Anal. Math. **119** (2013), 147–163, available at https://doi.org/10.1007/s11854-013-0005-2. MR3043150

[Bou13b] ———, *On the correlation of the Moebius function with rank-one systems*, J. Anal. Math. **120** (2013), 105–130, available at https://doi.org/10.1007/s11854-013-0016-z. MR3095150

[BDG16] Jean Bourgain, Ciprian Demeter, and Larry Guth, *Proof of the main conjecture in Vinogradov's mean value theorem for degrees higher than three*, Ann. of Math. (2) **184** (2016), no. 2, 633–682, available at https://doi.org/10.4007/annals.2016.184.2.7. MR3548534

[BSZ13] J. Bourgain, P. Sarnak, and T. Ziegler, *Disjointness of Moebius from horocycle flows*, From Fourier analysis and number theory to Radon transforms and geometry, 2013, pp. 67–83, available at https://doi.org/10.1007/978-1-4614-4075-8_5. MR2986954

[Dab89] H. Daboussi, *On the prime number theorem for arithmetic progressions*, J. Number Theory **31** (1989), no. 3, 243–254, available at https://doi.org/10.1016/0022-314X(89)90071-1. MR993901

[DK15] Tomasz Downarowicz and Stanisław Kasjan, *Odometers and Toeplitz systems revisited in the context of Sarnak's conjecture*, Studia Math. **229** (2015), no. 1, 45–72. MR3459905

[EW17] Manfred Einsiedler and Thomas Ward, *Functional analysis, spectral theory, and applications*, Graduate Texts in Mathematics, vol. 276, Springer, Cham, 2017. MR3729416

[EAKL16] El Houcein El Abdalaoui, Stanisław Kasjan, and Mariusz Lemańczyk, *0-1 sequences of the Thue-Morse type and Sarnak's conjecture*, Proc. Amer. Math. Soc. **144** (2016), no. 1, 161–176, available at https://doi.org/10.1090/proc/12683. MR3415586

[EAKPLdlR17] El Houcein El Abdalaoui, Joanna Kułaga Przymus, Mariusz Lemańczyk, and Thierry de la Rue, *The Chowla and the Sarnak conjectures from ergodic theory point of view*, Discrete Contin. Dyn. Syst. **37** (2017), no. 6, 2899–2944, available at https://doi.org/10.3934/dcds.2017125. MR3622068

[Erd49] P. Erdős, *On a Tauberian theorem connected with the new proof of the prime number theorem*, J. Indian Math. Soc. (N.S.) **13** (1949), 131–144. MR33309

[Erd57] ———, *Some unsolved problems*, Michigan Math. J. **4** (1957), 291–300.

[Erd90] ———, *Some of my favourite unsolved problems*, Cambridge Univ. Press, Cambridge, 1990.

[Eve13] Jan-Hendrik Evertse, *Analytic Number Theory* (2013), available at `http://www.math.leidenuniv.nl/~evertse/ant13-4.pdf`.

[FJ18] Ai-Hua Fan and Yunping Jiang, *Oscillating sequences, MMA and MMLS flows and Sarnak's conjecture*, Ergodic Theory Dynam. Systems **38** (2018), no. 5, 1709–1744, available at `https://doi.org/10.1017/etds.2016.121`. MR3819999

[Fra17] Nikos Frantzikinakis, *Ergodicity of the Liouville system implies the Chowla conjecture*, Discrete Anal. (2017), Paper No. 19, 41. MR3742396

[FH18a] Nikos Frantzikinakis and Bernard Host, *Furstenberg systems of bounded multiplicative functions and applications*, International Mathematics Research Notices (2018), available at `https://arxiv.org/pdf/1804.08556.pdf`.

[FH18b] ———, *The logarithmic Sarnak conjecture for ergodic weights*, Ann. of Math. (2) **187** (2018), no. 3, 869–931, available at `https://doi.org/10.4007/annals.2018.187.3.6`. MR3779960

[FKO82] H. Furstenberg, Y. Katznelson, and D. Ornstein, *The ergodic theoretical proof of Szemerédi's theorem*, Bull. Amer. Math. Soc. (N.S.) **7** (1982), no. 3, 527–552, available at `https://doi.org/10.1090/S0273-0979-1982-15052-2`. MR670131

[Gol04] D. Goldfeld, *The elementary proof of the prime number theorem: an historical perspective*, Number theory (New York, 2003), 2004, pp. 179–192. MR2044518

[Gol73] L. J. Goldstein, *Correction to: "A history of the prime number theorem" (Amer. Math. Monthly **80** (1973), 599–615)*, Amer. Math. Monthly **80** (1973), 1115, available at `https://doi.org/10.2307/2318546`. MR330016

[GLdLR19] Alexander Gomilko, Mariusz Lemańczyk, and Thierry de La Rue, *Möbius orthogonality in density for zero entropy dynamical systems* (2019), available at `https://arxiv.org/abs/1905.06563`.

[GHS19] Andrew Granville, Adam J. Harper, and K. Soundararajan, *A new proof of Halász's theorem, and its consequences*, Compos. Math. **155** (2019), no. 1, 126–163, available at https://doi.org/10.1112/s0010437x18007522. MR3880027

[GT10] Ben Green and Terence Tao, *Linear equations in primes*, Ann. of Math. (2) **171** (2010), no. 3, 1753–1850, available at https://doi.org/10.4007/annals.2010.171.1753. MR2680398

[GT12a] ———, *The Möbius function is strongly orthogonal to nilsequences*, Ann. of Math. (2) **175** (2012), no. 2, 541–566, available at https://doi.org/10.4007/annals.2012.175.2.3. MR2877066

[GT12b] ———, *The quantitative behaviour of polynomial orbits on nilmanifolds*, Ann. of Math. **175** (2012), 465–540, available at https://doi.org/10.4007/annals.2012.175.2.2.

[GTZ12] Ben Green, Terence Tao, and Tamar Ziegler, *An inverse theorem for the Gowers $U^{s+1}[N]$-norm*, Ann. of Math. (2) **176** (2012), no. 2, 1231–1372, available at https://doi.org/10.4007/annals.2012.176.2.11. MR2950773

[HR21] Harald Helfgott and Maksym Radziwiłł, *Expansion, Divisibility and Parity* (2021), available at https://arxiv.org/pdf/2103.06853.pdf.

[Hil86a] Adolf Hildebrand, *On consecutive values of the Liouville function*, Enseign. Math. (2) **32** (1986), no. 3-4, 219–226. MR874689

[Hil86b] ———, *The prime number theorem via the large sieve*, Mathematika **33** (1986), no. 1, 23–30, available at https://doi.org/10.1112/S002557930001384X. MR859495

[HK05] Bernard Host and Bryna Kra, *Nonconventional ergodic averages and nilmanifolds*, Ann. of Math. (2) **161** (2005), no. 1, 397–488, available at https://doi.org/10.4007/annals.2005.161.397. MR2150389

[HK18] ———, *Nilpotent structures in ergodic theory*, American Mathematical Society, 2018.

[HLSY17] Wen Huang, Zhengxing Lian, Song Shao, and Xiangdong Ye, *Sequences from zero entropy noncommutative toral automorphisms and Sarnak conjecture*, J. Differential Equations **263** (2017), no. 1, 779–810, available at https://doi.org/10.1016/j.jde.2017.02.046. MR3631324

[IK04] Henryk Iwaniec and Emmanuel Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications, vol. 53, American Mathematical Society, Providence, RI, 2004.

[Kub64] J. Kubilius, *Probabilistic methods in the theory of numbers*, Translations of Mathematical Monographs, Vol. 11, American Mathematical Society, Providence, R.I., 1964. MR0160745

[LS15] Jianya Liu and Peter Sarnak, *The Möbius function and distal flows*, Duke Math. J. **164** (2015), no. 7, 1353–1399, available at `https://doi.org/10.1215/00127094-2916213`. MR3347317

[MMR14] Bruno Martin, Christian Mauduit, and Joël Rivat, *Théoréme des nombres premiers pour les fonctions digitales*, Acta Arith. **165** (2014), no. 1, 11–45, available at `https://doi.org/10.4064/aa165-1-2`. MR3263939

[MR16] Kaisa Matomäki and Maksym Radziwiłł, *Multiplicative functions in short intervals*, Ann. of Math. (2) **183** (2016), no. 3, 1015–1056, available at `https://doi.org/10.4007/annals.2016.183.3.6`. MR3488742

[MRT16] Kaisa Matomäki, Maksym Radziwiłł, and Terence Tao, *Sign patterns of the Liouville and Möbius functions*, Forum Math. Sigma **4** (2016), e14, 44, available at `https://doi.org/10.1017/fms.2016.6`. MR3513734

[MRT18] ———, *Fourier uniformity of bounded multiplicative functions in short intervals on average*, Inventiones Mathematicae (2018), available at `https://arxiv.org/pdf/1812.01224.pdf`.

[MRT15] ———, *An averaged form of Chowla's conjecture*, Algebra Number Theory **9** (2015), no. 9, 2167–2196, available at `https://doi.org/10.2140/ant.2015.9.2167`. MR3435814

[MRT20] Kaisa Matomäki, Maksym Radziwiłł, and Terence Tao, *Fourier Uniformity of Bounded Multiplicative Functions in Short Intervals on Average*, Invent. Math. **220** (2020), no. 1, 1–58, DOI 10.1007/s00222-019-00926-W.

[MRT+] Kaisa Matomäki, Maksym Radziwiłł, Terence Tao, Joni Teräväinen, and Tamar Ziegler, *Higher uniformity of bounded multiplicative functions in short intervals on average.*

[McN18] Redmond McNamara, *Sarnak's conjecture for sequences of almost quadratic word growth*, Ergodic Theory and Dynamical Systems (2018), available at `https://arxiv.org/abs/1901.06460`.

[McN20] ———, *A Dynamical Proof of the Prime Number Theorem*, Hardy Ramanujan Journal (2020), available at `https://arxiv.org/abs/2002.04007`.

[Mor18] Joel Moreira, *Tao's proof of (logarithmically averaged) chowla's conjecture for two point correlations* (2018), available at `https://joelmoreira.wordpress.com/2018/11/04/taos-proof-of-logarithmically-averaged-chowlas-conjecture-for-two-point-correlations/`.

[Mül17] Clemens Müllner, *Automatic sequences fulfill the Sarnak conjecture*, Duke Math. J. **166** (2017), no. 17, 3219–3290, available at https://doi.org/10.1215/00127094-2017-0024. MR3724218

[Pec18] Ryan Peckner, *Möbius disjointness for homogeneous dynamics*, Duke Math. J. **167** (2018), no. 14, 2745–2792, available at https://doi.org/10.1215/00127094-2018-0026. MR3859364

[Pol] D.H.J. Polymath, *The Erdős discrepancy problem*, available at https://asone.ai/polymath/index.php?title=The_Erd%C5%91s_discrepancy_problem.

[Ric] Florian Richter, *A new elementary proof of the prime number theorem*, available at https://arxiv.org/abs/2002.03255.

[RS62] J. Barkley Rosser and Lowell Schoenfeld, *Approximate formulas for some functions of prime numbers*, Illinois J. Math. **6** (1962), 64–94.

[Sar12] Peter Sarnak, *Mobius randomness and dynamics*, Not. S. Afr. Math. Soc. **43** (2012), no. 2, 89–97. MR3014544

[Saw20] Will Sawin, *Dynamical models for Liouville and obstructions to further progress on sign patterns*, J. Number Theory **213** (2020), 1–15, available at https://doi.org/10.1016/j.jnt.2020.01.012. MR4091952

[SS] Will Sawin and Mark Shusterman, *On the Chowla and twin primes conjectures over* $\mathbb{F}_q[tT$, available at https://arxiv.org/abs/1808.04001.

[Sel50] Atle Selberg, *An elementary proof of the prime-number theorem for arithmetic progressions*, Canad. J. Math. **2** (1950), 66–78, available at https://doi.org/10.4153/cjm-1950-007-5. MR33306

[Taoa] Terence Tao, *254a, notes 9 – second moment and entropy methods.*

[Taob] ———, *254a, notes 10 – mean values of nonpretentious multiplicative functions*, available at https://terrytao.wordpress.com/2019/12/17/254a-notes-10-mean-values-of-nonpretentious-multiplicative-functions/.

[Taoc] ———, *A banach algebra proof of the prime number theorem*, available at https://terrytao.wordpress.com/2014/10/25/a-banach-algebra-proof-of-the-prime-number-theorem/.

[Tao16a] ———, *The Erdős discrepancy problem*, Discrete Anal., posted on 2016, Paper No. 1, 29, DOI 10.19086/da.609.

[Tao16b] ———, *The logarithmically averaged Chowla and Elliott conjectures for two-point correlations*, Forum Math. Pi **4** (2016), e8, 36, available at `https://doi.org/10.1017/fmp.2016.6`. MR3569059

[Tao17a] ———, *Equivalence of the logarithmically averaged Chowla and Sarnak conjectures*, Number theory—Diophantine problems, uniform distribution and applications, 2017, pp. 391–421, available at `https://arxiv.org/abs/1605.04628`. MR3676413

[Tao17b] ———, *Furstenberg limits of the liouville function* (2017), available at `https://terrytao.wordpress.com/2017/03/05/furstenberg-limits-of-the-liouville-function/`.

[TT17a] Terence Tao and Joni Teräväinen, *Odd order cases of the logarithmically averaged chowla conjecture*, J. Numb. Thy. Bordeaux (2017), available at `https://arxiv.org/abs/1710.02112`.

[TT17b] ———, *The structure of logarithmically averaged correlations of multiplicative functions, with applications to the chowla and elliott conjectures*, Duke Mathematical Journal (2017), available at `https://arxiv.org/pdf/1708.02610.pdf`.

[Vee17] William A. Veech, *Möbius orthogonality for generalized Morse-Kakutani flows*, Amer. J. Math. **139** (2017), no. 5, 1157–1203, available at `https://doi.org/10.1353/ajm.2017.0031`. MR3702497

[Wan17] Zhiren Wang, *Möbius disjointness for analytic skew products*, Invent. Math. **209** (2017), no. 1, 175–196, available at `https://doi.org/10.1007/s00222-016-0707-z`. MR3660308

[Zag97] D. Zagier, *Newman's short proof of the prime number theorem*, Amer. Math. Monthly **104** (1997), no. 8, 705–708, available at `https://doi.org/10.2307/2975232`. MR1476753