THE LOGARITHMICALLY AVERAGED CHOWLA AND ELLIOTT CONJECTURES FOR TWO-POINT CORRELATIONS

TERENCE TAO

Abstract. Let λ denote the Liouville function. The Chowla conjecture, in the two-point correlation case, asserts that

$$\sum_{n \leqslant x} \lambda(a_1 n + b_1)\lambda(a_2 n + b_2) = o(x)$$

as $x \to \infty$, for any fixed natural numbers a_1, a_2, b_1, b_2 with $a_1 b_2 - a_2 b_1 \neq 0$. In this paper we establish the logarithmically averaged version

$$\sum_{x/\omega(x) < n \leqslant x} \frac{\lambda(a_1 n + b_1)\lambda(a_2 n + b_2)}{n} = o(\log \omega(x))$$

of the Chowla conjecture as $x \to \infty$, where $1 \leqslant \omega(x) \leqslant x$ is an arbitrary function of x that goes to infinity as $x \to \infty$, thus breaking the "parity barrier" for this problem. Our main tools are the multiplicativity of the Liouville function at small primes, a recent result of Matomäki, Radziwiłł, and the author on the averages of modulated multiplicative functions in short intervals, concentration of measure inequalities, the Hardy-Littlewood circle method combined with a restriction theorem for the primes, and a novel "entropy decrement argument". Most of these ingredients are also available (in principle, at least) for the higher order correlations, with the main missing ingredient being the need to control short sums of multiplicative functions modulated by local nilsequences.

Our arguments also extend to more general bounded multiplicative functions than the Liouville function λ, leading to a logarithmically averaged version of the Elliott conjecture in the two-point case. In a subsequent paper we will use this version of the Elliott conjecture to affirmatively settle the Erdős discrepancy problem.

1. Introduction

Let λ denote the Liouville function, thus λ is the completely multiplicative function such that $\lambda(p) = -1$ for all primes p. We have the following well known conjecture of Chowla [2]:

Conjecture 1.1 (Chowla conjecture). *Let $k \geqslant 1$, let $a_1, \dots, a_k$ be natural numbers and let $b_1, \dots, b_k$ be distinct nonnegative integers such that $a_i b_j - a_j b_i \neq 0$ for $1 \leqslant i < j \leqslant k$. Then*

$$\sum_{n \leqslant x} \lambda(a_1 n + b_1) \dots \lambda(a_k n + b_k) = o(x)$$

as $x \to \infty$.

Thus for instance the $k = 2$ case of the Chowla conjecture implies that

$$\sum_{n \leqslant x} \lambda(n)\lambda(n+1) = o(x) \tag{1.1}$$

as $x \to \infty$. This can be compared with the twin prime conjecture, which is equivalent to the assertion that

$$\sum_{n \leqslant x} \theta(n)\theta(n+2) \to \infty \tag{1.2}$$

as $x \to \infty$, where $\theta(n) := \log p$ when n is equal to a prime p, and $\alpha(n) = 0$ otherwise.

The $k = 1$ case of the Chowla conjecture is equivalent to the prime number theorem. The higher k cases are open, although there are a number of partial results available if one allows for some averaging in the $b_1, \dots, b_k$ parameters; see [17], [6] for some recent results in this direction.

The first main result of this paper is to obtain a different averaged form of the Chowla conjecture in the first nontrivial case $k = 2$, in which one averages in x rather than in $b_1, \dots, b_k$. More precisely, we show

Theorem 1.2 (Logarithmically averaged Chowla conjecture)**.** *Let a_1, a_2 be natural numbers, and let b_1, b_2 be integers such that $a_1b_2 - a_2b_1 \neq 0$. Let $1 \leqslant \omega(x) \leqslant x$ be a quantity depending on x that goes to infinity as $x \to \infty$. Then one has*

$$\sum_{x/\omega(x) < n \leqslant x} \frac{\lambda(a_1n+b_1)\lambda(a_2n+b_2)}{n} = o(\log \omega(x)) \tag{1.3}$$

as $n \to \infty$.

Thus for instance this theorem implies (after setting $\omega(x) := x$, $a_1 = a_2 = b_2 = 1$ and $b_1 = 0$) that

$$\sum_{n \leqslant x} \frac{\lambda(n)\lambda(n+1)}{n} = o(\log x) \tag{1.4}$$

as $x \to \infty$; this can be deduced from (1.1) by a routine summation by parts argument, but is a strictly weaker estimate. From this and the elementary estimate $\sum_{n \leqslant x} \frac{\lambda(n)}{n} = o(\log x)$ we see that for any sign pattern $(\epsilon_1, \epsilon_2) \in \{-1, +1\}^2$, the set $\{n : (\lambda(n), \lambda(n+1)) = (\epsilon_1, \epsilon_2)\}$ occurs with logarithmic density $1/4$, that is to say

$$\frac{1}{\log x} \sum_{n \leqslant x : (\lambda(n), \lambda(n+1)) = (\epsilon_1, \epsilon_2)} \frac{1}{n} = \frac{1}{4} + o(1)$$

as $x \to \infty$.

More generally, one can deduce Theorem 1.2 from the $k = 2$ case of Conjecture 1.1 by summation by parts; we leave the details to the interested reader. Conversely, the $k = 2$ case of Conjecture 1.1 is equivalent to the limiting case of Theorem 1.2 in which ω is fixed rather than going to infinity. The logarithmic averaging is unfortunately needed in our method in order to obtain an approximate dilation invariance in the n variable; we do not know how to modify our argument to remove this averaging. However, the logarithmic averaging can be tolerated in some applications (for instance to the Erdös discrepancy problem, discussed below).

Estimates such as (1.1), (1.2), (1.3), (1.4) are well known to be subject to the parity problem obstruction (see e.g. [9, Chapter 16]), and thus cannot be resolved purely by existing sieve-theoretic techniques that rely solely on "linear" estimates for the Liouville function. We avoid the parity obstacle here by using a new "bilinear" estimate[1] for the Liouville function, which comes from the multiplicativity property $\lambda(pn) = -\lambda(n)$ of the Liouville function at small primes p, together with the expansion properties of a certain random graph, closely related to one recently introduced in [18]. To describe this strategy in somewhat informal terms, let us specialise to the case of establishing (1.4) for simplicity. Suppose for contradiction that the left-hand side of (1.4) was large and (say) positive. Using the multiplicativity $\lambda(pn) = -\lambda(n)$, we conclude that

$$\sum_{n \leqslant x} \frac{\lambda(n)\lambda(n+p)1_{p|n}}{n}$$

is also large and positive for all primes p that are not too large; note here how the logarithmic averaging allows us to leave the constraint $n \leqslant x$ unchanged. Summing in p, we conclude that

$$\sum_{n \leqslant x} \frac{\sum_{p \in \mathcal{P}} \lambda(n)\lambda(n+p)1_{p|n}}{n}$$

is large and positive for any given set $\mathcal{P}$ of medium-sized primes. By a standard averaging argument, this implies that

$$\frac{1}{H}\sum_{j=1}^{H}\sum_{p \in \mathcal{P}} \lambda(n+j)\lambda(n+p+j)1_{p|n+j} \tag{1.5}$$

is large for many choices of n, where H is a medium-sized parameter at our disposal to choose, and we take $\mathcal{P}$ to be some set of primes that are somewhat smaller than H. To obtain the required contradiction, one thus wants to demonstrate significant cancellation in the expression (1.5). As in [18], we view n as a random variable, in which case (1.5) is essentially a bilinear sum of the random sequence $(\lambda(n+1), \dots, \lambda(n+H))$ along a random graph $G_{n,H}$ on $\{1, \dots, H\}$, in

[1] Bilinear estimates have been used to get around the parity obstacle in previous works, most notably in the Friedlander-Iwaniec result [8] on primes of the form $a^2 + b^4$.

which two vertices $j, j+p$ are connected if they differ by a prime p in $\mathcal{P}$ that divides $n+j$. A key difficulty in controlling this sum is that for randomly chosen n, the sequence $(\lambda(n+1), \dots, \lambda(n+H))$ and the graph $G_{n,H}$ need not be independent. To get around this obstacle we introduce a new argument which we call the "entropy decrement argument" (in analogy with the "density increment argument" and "energy increment argument" that appear in the literature surrounding Szemerédi's theorem on arithmetic progressions (see e.g. [22]), and also reminiscent of the "entropy compression argument" of Moser and Tardos [20]). This argument, which is a simple consequence of the Shannon entropy inequalities, can be viewed as a quantitative version of the standard subadditivity argument that establishes the existence of Kolmogorov-Sinai entropy in topological dynamical systems; it allows one to select a scale parameter H (in some suitable range $[H_-, H_+]$) for which the sequence $(\lambda(n+1), \dots, \lambda(n+H))$ and the graph $G_{n,H}$ exhibit some weak independence properties (or more precisely, the mutual information between the two random variables is small). With this additional property, one can use standard concentration of measure results such as the Hoeffding inequality to approximate (1.5) by the significantly simpler expression

$$\frac{1}{H}\sum_{j=1}^{H}\sum_{p\in\mathcal{P}}\frac{\lambda(n+j)\lambda(n+p+j)}{p}.$$

This latter expression can then be controlled in turn by an application of the Hardy-Littlewood circle method and an estimate for short sums of a modulated Liouville function established recently by Matomäki, Radziwiłł and the author in [17].

The arguments in this paper extend to other bounded multiplicative functions than the Liouville function, though as they rely in an essential fashion on multiplicativity at small primes, they unfortunately do not appear to have any bearing as yet on twin prime-type sums such as (1.2). More precisely, we have the following logarithmically averaged and nonasymptotic version of the Elliott conjecture [3] (in the "corrected" form introduced in [17]):

Theorem 1.3 (Logarithmically averaged nonasymptotic Elliott conjecture). *Let a_1, a_2 be natural numbers, and let b_1, b_2 be integers such that $a_1b_2 - a_2b_1 \neq 0$. Let $\varepsilon > 0$, and suppose that A is sufficiently large depending on $\varepsilon, a_1, a_2, b_1, b_2$. Let $x \geqslant \omega \geqslant A$, and let $g_1, g_2 : \mathbb{N} \to \mathbb{C}$ be multiplicative functions with $|g_1(n)|, |g_2(n)| \leqslant 1$ for all n, with g_1 "non-pretentious" in the sense that*

$$\sum_{p\leqslant x}\frac{1-\operatorname{Re} g_1(p)\overline{\chi(p)}p^{-it}}{p} \geqslant A \tag{1.6}$$

for all Dirichlet characters χ of period at most A, and all real numbers t with $|t| \leqslant Ax$. Then

$$\left| \sum_{x/\omega < n \leqslant x} \frac{g_1(a_1 n + b_1) g_2(a_2 n + b_2)}{n} \right| \leqslant \varepsilon \log \omega. \tag{1.7}$$

Remark 1.4. Our arguments are in principle effective, and would yield an explicit value of A as a function of $\varepsilon, a_1, a_2, b_1, b_2$ if one went through all the arguments carefully, however we did not do so here as we expect the bounds to be rather poor.

Theorem 1.3 clearly implies the following asymptotic version:

Corollary 1.5 (Logarithmically averaged Elliott conjecture)**.** *Let a_1, a_2 be natural numbers, and let b_1, b_2 be integers such that $a_1 b_2 - a_2 b_1 \neq 0$. Let $g_1, g_2 : \mathbb{N} \to \mathbb{C}$ be multiplicative functions bounded in magnitude by one, with g_1 "non-pretentious" in the sense that*

$$\inf_{|t| \leqslant Ax} \sum_{p \leqslant x} \frac{1 - \operatorname{Re} g_1(p) \overline{\chi(p)} p^{-it}}{p} \to \infty \tag{1.8}$$

as $x \to \infty$ for all Dirichlet characters χ and all $A \geqslant 1$. Then for any $1 \leqslant \omega(x) \leqslant x$ which goes to infinity as $x \to \infty$, one has

$$\sum_{x/\omega(x) < n \leqslant x} \frac{g_1(a_1 n + b_1) g_2(a_2 n + b_2)}{n} = o(\log \omega(x)) \tag{1.9}$$

as $x \to \infty$.

Remark 1.6. If one replaced the conclusion (1.9) with the stronger, non-logarithmically-averaged estimate

$$\sum_{n \leqslant x} g_1(a_1 n + b_1) g_2(a_2 n + b_2) = o(x), \tag{1.10}$$

(say with $b_1, b_2 \geqslant 0$ to avoid the linear forms $a_1 n + b_1, a_2 n + b_2$ leaving the domain of g_1, g_2) then this is the $k = 2$ version of the corrected Elliott conjecture introduced in [17]. The original Elliott conjecture in [3] replaced the condition (1.8) with the weaker condition

$$\sum_{p} \frac{1 - \operatorname{Re} g_1(p) \overline{\chi(p)} p^{-it}}{p} = +\infty$$

for all real numbers $t \in \mathbb{R}$, but it was shown in [17] that this hypothesis was insufficient to establish (1.10) (and it is not difficult to adapt the counterexample to also show that (1.9) fails). On the other hand, in [17], it was shown that the corrected Elliott conjecture held if one averaged in the $b_1, \dots, b_k$ parameters (rather than in the x parameter as is done here).

Using the prime number theorem in arithmetic progressions with Vinogradov-Korobov error term (see [19, §9.5]), it is not difficult to establish (1.8) when g is the Liouville function; thus Corollary 1.5 implies Theorem 1.2. Some condition of the form (1.8) must be needed in order to derive the conclusion (1.9), as one can see by considering examples such as $g(n) := \chi(n)n^{2ix}$.

Corollary 1.5 also implies the asymptotic

$$\sum_{n \leqslant x} \frac{g_1(n)g_2(n+1)}{n} = o(\log x)$$

as $x \to \infty$ when g_1, g_2 are multiplicative functions bounded by 1, and at least one of g_1, g_2 is equal to the Möbius function μ. Thus for instance one has

$$\sum_{n \leqslant x} \frac{\mu(n)\mu(n+1)}{n}, \sum_{n \leqslant x} \frac{\mu^2(n)\mu(n+1)}{n}, \sum_{n \leqslant x} \frac{\mu(n)\mu^2(n+1)}{n} = o(\log x).$$

The latter two estimates can be easily deduced from the prime number theorem in arithmetic progressions, but the first estimate is new. Combining this with the computations in [18, §2] (using logarithmic density in place of asymptotic probability), we conclude

Corollary 1.7 (Sign patterns of the Möbius function)**.** *Let*

$$c := \prod_p \left(1 - \frac{2}{p^2}\right) = 0.3226\ldots$$

and let $(\epsilon_1, \epsilon_2) \in \{-1, 0, +1\}^2$. *Then the set* $\{n : (\mu(n), \mu(n+1)) = (\epsilon_1, \epsilon_2)\}$ *has logarithmic density*

- $1 - \frac{2}{\zeta(2)} + c = 0.1067\ldots$ *when* $(\epsilon_1, \epsilon_2) = (0, 0)$;
- $\frac{1}{2}\left(\frac{1}{\zeta(2)} - c\right) = 0.1426\ldots$ *when* $(\epsilon_1, \epsilon_2) = (+1, 0), (-1, 0), (0, +1), (0, -1)$; *and*
- $\frac{c}{4} = 0.0806\ldots$ *when* $(\epsilon_1, \epsilon_2) = (+1, +1), (+1, -1), (-1, +1), (-1, -1)$.

Again, the first two cases here could already be treated using the prime number theorem in arithmetic progressions, but the last case is new. One can also use similar arguments to give an alternate proof of [18, Theorem 1.9] (that is to say, that all nine of the above sign patterns for the Möbius function occur with positive lower density); we leave the details to the interested reader.

In a subsequent paper [23], we will combine Corollary 1.5 with some arguments arising from the `Polymath5` project [21] to obtain an affirmative answer to the Erdős discrepancy problem [4]:

Theorem 1.8. *Let* $f : \mathbb{N} \to \{-1, +1\}$ *be a function. Then*

$$\sup_{d,n \in \mathbb{N}} \left| \sum_{j \leqslant n} f(jd) \right| = +\infty.$$

1.1. **Notation.** We adopt the usual asymptotic notation of $X \ll Y$, $Y \gg X$, or $X = O(Y)$ to denote the assertion that $|X| \leqslant CY$ for some constant C. If we need C to depend on an additional parameter we will denote this by subscripts, e.g. $X = O_\varepsilon(Y)$ denotes the bound $|X| \leqslant C_\varepsilon Y$ for some C_ε depending on Y. Similarly, we use $X = o_{A\to\infty}(Y)$ to denote the bound $|X| \leqslant c(A)Y$ where $c(A)$ depends only on A and goes to zero as $A \to \infty$.

If E is a statement, we use 1_E to denote the indicator, thus $1_E = 1$ when E is true and $1_E = 0$ when E is false.

Given a finite set S, we use $|S|$ to denote its cardinality.

For any real number α, we write $e(\alpha) := e^{2\pi i\alpha}$; this quantity lies in the unit circle $S^1 := \{z \in \mathbb{C} : |z| = 1\}$. By abuse of notation, we can also define $e(\alpha)$ when α lies in the additive unit circle $\mathbb{R}/\mathbb{Z}$.

All sums and products will be over the natural numbers $\mathbb{N} = \{1, 2, \dots\}$ unless otherwise specified, with the exception of sums and products over p which is always understood to be prime.

We use $d|n$ to denote the assertion that d divides n, and $n\ (d)$ to denote the residue class of n modulo d. We use (a, b) to denote the greatest common divisor of a and b.

We will frequently use probabilistic notation such as the expectation $\mathbb{E}\mathbf{X}$ of a random variable $\mathbf{X}$ or a probability $\mathbb{P}(E)$ of an event E; later we will also need the Shannon entropy $\mathbb{H}(\mathbf{X})$ of a discrete random variable, as well as related quantities such as relative entropy $\mathbb{H}(\mathbf{X}|\mathbf{Y})$ or mutual information $\mathbb{I}(\mathbf{X}, \mathbf{Y})$, the definitions of which we review in Section 3. We will use boldface symbols such as $\mathbf{X}$, $\mathbf{Y}$ or $\mathbf{n}$ to refer to random variables.

1.2. **Acknowledgments.** The author is supported by NSF grant DMS-0649473 and by a Simons Investigator Award. The author also thanks Andrew Granville, Ben Green, Kaisa Matomäki, Maksym Radziwiłł, and Will Sawin for helpful discussions and comments.

2. Preliminary reductions

In this section we make a number of basic reductions, in particular reducing matters to a probabilistic problem involving a random graph, somewhat similar to one considered in [18]. Readers who are interested just in the case of the Liouville function (Theorem 1.2) can skip the initial reductions and move directly to Theorem 2.3 below.

As mentioned in the introduction, Theorem 1.2 is a special case of Corollary 1.5, which is in turn a corollary of Theorem 1.3. Thus it will suffice to establish Theorem 1.3.

We first reduce to the case when g_1 takes values on the unit circle S^1:

Proposition 2.1. *In order to establish Theorem 1.3, it suffices to do so in the special case where* $|g_1(n)| = 1$ *for all* n.

Proof. Suppose that g_1 takes values in the unit disk. Then we may factorise $g_1 = g_1' g_1''$ where g_1', g_1'' are multiplicative, with g_1' taking values in $[0,1]$ and g_1'' taking values in the unit circle S^1.

Let A_0 be a large quantity (depending on $a_1, a_2, b_1, b_2, \varepsilon$) to be chosen later; we assume that A is sufficiently large depending on $a_1, a_2, b_1, b_2, \varepsilon, A_0$. Suppose first that

$$\sum_{p \leqslant x} \frac{1 - g_1'(p)}{p} \geqslant A_0.$$

By Mertens' theorem and the largeness of A_0 and x, this implies that

$$\sum_{p \leqslant y} \frac{1 - g_1'(p)}{p} \geqslant \frac{A_0}{2}$$

for every $x^{1/A_0} \leqslant y \leqslant x$ (say). Applying the Halasz inequality (see e.g. [24] or [10, Corollary 1]) we conclude that

$$\frac{1}{y} \sum_{n \leqslant y} g_1'(n) \ll A_0 \exp(-A_0/2)$$

for all $x^{1/A_0} \leqslant y \leqslant x$ (assuming $x \geqslant A$ and A is sufficiently large depending on A_0). From this and the nonnegativity and boundedness of $g_1'(n)$ it is easy to see that

$$\sum_{x/\omega \leqslant n \leqslant x} \frac{g_1'(a_1 n + b_1)}{n} = o_{A_0 \to \infty}(\log \omega)$$

since $x \geqslant \omega \geqslant A$ and A is large compared to A_0, and A_0 is large compared to a_1, b_1. Since $g_1(a_1 n + b_1) g_2(a_2 n + b_2)$ is bounded in magnitude by $g_1'(a_1 n_1 + b_1)$, the claim (1.7) now follows from the triangle inequality (taking A_0 large enough).

It remains to treat the case when

$$\sum_{p \leqslant x} \frac{1 - g_1'(p)}{p} < A_0.$$

We now use the probabilistic method to model g_1' by a multiplicative function of unit magnitude. Since $g_1'(p^j)$ takes values in the convex hull of $\{-1, +1\}$ for every prime power p^j, we can construct a random multiplicative function $\mathbf{g}_1'$ taking values in $\{-1, +1\}$, such that the values $\mathbf{g}_1'(p^j)$ at prime powers are jointly independent and have mean $\mathbb{E}\mathbf{g}_1'(p^j) = g_1'(p^j)$. By multiplicativity and joint independence, we thus have $\mathbb{E}\mathbf{g}_1'(n) = g_1'(n)$ for arbitrary n. By linearity of expectation we have

$$\mathbb{E} \sum_{p \leqslant x} \frac{1 - \mathbf{g}_1'(p)}{p} < A_0.$$

so by Markov's inequality we see with probability $1 - O(1/A_0)$ that

$$\sum_{p \leqslant x} \frac{1 - \mathbf{g}_1'(p)}{p} < A_0^2.$$

Let us restrict to this event, and set $\mathbf{g}_1 := \mathbf{g}_1' g_1''$, thus $\mathbf{g}_1$ is a random multiplicative function taking values in S^1 whose mean is g_1. By the triangle inequality we have

$$\mathbf{g}_1(p) = g(p) + O(1 - g_1'(p)) + O(1 - \mathbf{g}_1'(p))$$

and hence by (1.6) and the triangle inequality again we have

$$\sum_{p \leqslant x} \frac{1 - \operatorname{Re} \mathbf{g}_1(p)\overline{\chi(p)} p^{-it}}{p} \geqslant A/2$$

for all Dirichlet characters χ of period at most A and all t with $|t| \leqslant Ax$, if A is large enough. Using the hypothesis that Theorem 1.3 holds when g_1 has unit magnitude, we conclude (again taking A large enough) that

$$\left| \sum_{x/\omega < n \leqslant x} \frac{\mathbf{g}_1(a_1 n + b_1) g_2(a_2 n + b_2)}{n} \right| \leqslant \frac{\varepsilon}{2} \log \omega \tag{2.1}$$

with probability $1 - O(1/A_0)$. In the exceptional event that this fails, we can still bound the left-hand side of (2.1) by $O(\log \omega)$. Taking expectations, we obtain (1.7) as desired (for A_0 large enough). □

A similar argument allows one to also reduce to the case where $|g_2(n)| = 1$ for all n (indeed, the argument is slightly simpler as (1.6) is unaffected by changes in g_2).

Next, we upgrade the functions g_1, g_2 from being multiplicative to being completely multiplicative.

Proposition 2.2. *In order to establish Theorem 1.3, it suffices to do so in the special case where* $|g_1(n)| = |g_2(n)| = 1$ *for all* n*, and* g_1 *is completely multiplicative.*

Proof. By the previous reductions we may already assume that $|g_1(n)| = |g_2(n)| = 1$ for all n. If g_1 is not completely multiplicative, we can introduce the completely multiplicative function $\tilde{g}_1$ with $\tilde{g}_1(p) = g_1(p)$ for all p. Clearly, $\tilde{g}_1$ takes values in S^1. From Möbius inversion (twisted by $\tilde{g}_1$) we can factor g_1 as a Dirichlet convolution $g_1 = \tilde{g}_1 * h$ for a multiplicative function h with $h(p) = 0$ and $|h(p^j)| \leqslant 2$ for all $j \geqslant 2$. The left-hand side of (1.7) can then be rewritten as

$$\left| \sum_d h(d) \sum_{x/\omega < n \leqslant x : d | a_1 n + b_1} \frac{\tilde{g}_1(\frac{a_1 n + b_1}{d}) g_2(a_2 n + b_2)}{n} \right|.$$

As in the previous proposition, we choose a quantity A_0 that is sufficiently large depending on $a_1, a_2, b_1, b_2, \varepsilon$, and assume A is sufficiently large depending on $A_0, a_1, a_2, b_1, b_2, \varepsilon$. We consider first the contribution to the above sum of a single value of d with $d \leqslant A_0$. We crudely bound $|h(d)|$ by (say) A_0. The constraint $d|a_1n + b_1$ constrains n to at most d residue classes modulo d. Making an appropriate change of variables and using the hypothesis that Theorem 1.3 holds for completely multiplicative g_1, we thus have

$$\left| \sum_{x/\omega < n \leqslant x : d | a_1 n + b_1} \frac{\tilde{g}_1(\frac{a_1n+b_1}{d}) g_2(a_2n+b_2)}{n} \right| \leqslant \frac{\varepsilon}{2A_0^2} \log \omega$$

if A is large enough. Thus the total contribution of those d with $d \leqslant A_0$ is at most $\frac{\varepsilon}{2} \log \omega$.

Now we turn to the contribution where $d > A_0$. Here, we can use the triangle inequality to bound $\sum_{x/\omega < n \leqslant x : d | a_1 n + b_1} \frac{\tilde{g}_1(\frac{a_1n+b_1}{d}) g_2(a_2n+b_2)}{n}$ by $O(\frac{\log \omega}{d})$, so the net contribution of this case is $O(\log \omega \sum_{d > A_0} \frac{|h(d)|}{d})$. However, from taking Euler products one sees that

$$\sum_d \frac{|h(d)|}{d^{2/3}} = O(1)$$

(say), and thus

$$\sum_{d > A_0} \frac{|h(d)|}{d} = O(A_0^{-1/3}).$$

Taking A_0 large enough, we obtain the claim. □

A similar argument allows one to also reduce to the case where g_2 is completely multiplicative. As g_1, g_2 are now multiplicative and take values in S^1, we have

$$g_1(a_1n+b_1)g_2(a_2n+b_2) = \overline{g_1}(a_2)\overline{g_2}(a_1)g_1(a_1a_2n+a_2b_1)g_2(a_1a_2n+a_1b_2)$$

so by replacing a_1, a_2, b_1, b_2 with $a_1a_2, a_1a_2, b_1a_2, b_2a_1$ respectively, we may assume that $a_1 = a_2 = a$, $b_1 = b$, and $b_2 = b + h$ for some natural number a, integer b, and nonzero integer h.

Finally, we observe that we can strengthen the condition $\omega \leqslant x$ slightly to $\omega \leqslant \frac{x}{\log x}$, since for $\frac{x}{\log x} < \omega \leqslant x$, the contribution of those n for which $n \leqslant \log x$ can be seen to be negligible.

Putting all these reductions together, we see that Theorem 1.3 will be a consequence of the following theorem.

Theorem 2.3 (Logarithmically averaged nonasymptotic Elliott conjecture)**.** *Let a be a natural number, and let b, h be integers with $h \neq 0$. Let $\varepsilon > 0$, and suppose that A is sufficiently large depending on ε, a, b, h. Let $x \geqslant \frac{x}{\log x} \geqslant \omega \geqslant A$, and let $g_1, g_2 : \mathbb{N} \to S^1$ be completely multiplicative functions such that* (1.6) *holds for all Dirichlet characters χ*

of period at most A, and all real numbers t with $|t| \leqslant Ax$. Then

$$\left| \sum_{x/\omega < n \leqslant x} \frac{g_1(an+b) g_2(an+b+h)}{n} \right| \leqslant \varepsilon \log \omega.$$

Let a, b, h, ε be as in the above theorem[2]. Suppose for sake of contradiction that Theorem 2.3 fails for this set of parameters. By shrinking ε, we may assume that ε is sufficiently small depending on a, b, h. Thus for instance any quantity of the form $O_{a,b,h}(\varepsilon)$ can be assumed to be much smaller than 1, any quantity of the form $O_{a,b,h}(\varepsilon^2)$ can be assumed to be much smaller than ε, and so forth. We will also need a number of large quantities, chosen in the following order:

- We choose a natural number H_- that is sufficiently large depending on a, b, h, ε.
- Then, we choose a natural number H_+ that is sufficiently large depending on $H_-, a, b, h, \varepsilon$.
- Finally, we choose a quantity $A > 0$ that is sufficiently large depending on $H_+, H_-, a, b, h, \varepsilon$.

The quantity A is of course the one we will use in Theorem 2.3. The intermediate parameters H_-, H_+ will be the lower and upper ranges for a certain medium-sized scale $H \in [H_-, H_+]$ which we willlater select using a pigeonholing argument which we call the "entropy decrement argument".

We will implicitly take repeated advantage of the above relative size assumptions between the parameters $A, H_+, H_-, a, b, h, \varepsilon$ in the sequel to simplify the estimates; in particular, we will repeatedly absorb lower order error terms into higher order error terms when the latter would dominate the former under the above assumptions. Thus for instance $O_{H_+,H_-,a,b,h,\varepsilon}(1) \times o_{A\to\infty}(1)$ can be simplified to just $o_{A\to\infty}(1)$ by the assumption that A is sufficiently large depending on all previous parameters, and $o_{A\to\infty}(1) + o_{H_-\to\infty}(1)$ can similarly be simplified to $o_{H_-\to\infty}(1)$. The reader may wish to keep the hierarchy

$$a, b, h \ll \frac{1}{\varepsilon} \ll H_- \ll p \ll H \ll H_+ \ll A \leqslant \omega \leqslant \frac{x}{\log x} \leqslant x$$

and also

$$x \geqslant n \geqslant x/\omega \geqslant \log x \geqslant \log A \gg H_+$$

in mind in the arguments that follow.

By hypothesis, there exists real numbers

$$x \geqslant \omega \geqslant A \tag{2.2}$$

[2] The reader may initially wish to restrict to the model case $a = 1, b = 0, h = 1$ in what follows to simplify the notation and arguments slightly.

and completely multiplicative functions $g_1, g_2 : \mathbb{N} \to S^1$ such that

$$\sum_{p \leqslant x} \frac{1 - \operatorname{Re} g_1(p)\overline{\chi(p)}p^{-it}}{p} \geqslant A \tag{2.3}$$

for all Dirichlet characters χ of period at most A, and all real numbers t with $|t| \leqslant Ax$, but such that

$$\left| \sum_{x/\omega < n \leqslant x} \frac{g_1(an+b)g_2(an+b+h)}{n} \right| > \varepsilon \log \omega. \tag{2.4}$$

To use the hypothesis (2.3), we apply the results in [17] to control short sums of g_1 modulated by Fourier characters.

Proposition 2.4. *Let the notation be as above. For all $H_- \leqslant H \leqslant H_+$, one has*

$$\sup_\alpha \sum_{x/\omega < n \leqslant x} \frac{1}{Hn} \left| \sum_{j=1}^H g_1(n+j)e(j\alpha) \right| = o_{H_- \to \infty}(\log \omega). \tag{2.5}$$

Proof. Let $\alpha \in \mathbb{R}$. Applying [17, Lemma 2.2, Theorem 2.3] (with $W := \log^5 H$, noting that this is much less than A or $(\log X)^{1/125}$), we see that

$$\frac{1}{X} \sum_{X \leqslant n \leqslant 2X} \left| \frac{1}{H} \sum_{j=1}^H g_1(n+j)e(\alpha j) \right| = o_{H_- \to \infty}(1)$$

for all $\frac{x}{2\omega} \leqslant X \leqslant 2x$, noting from the hypotheses that $X \geqslant \frac{x}{2\omega} \geqslant \frac{\log x}{2} \geqslant \frac{\log A}{2}$. Averaging this from X between $x/2\omega$ and $2x$, we obtain (2.5). □

It will be convenient to interpret these estimates in probabilistic language (particularly when we start using the concept of Shannon entropy in the next section). We introduce a (discrete) random variable $\mathbf{n}$ in the interval $\{n \in \mathbb{N} : x/\omega < n \leqslant x\}$ by setting

$$\mathbb{P}(\mathbf{n} = n) = \frac{1/n}{\sum_{n \in \mathbb{N}: x/\omega < n \leqslant x} \frac{1}{n}}$$

whenever n lies in this interval.

From (2.2) we see that

$$\sum_{n \in \mathbb{N}: x/\omega < n \leqslant x} \frac{1}{n} = (1 + o_{A \to \infty}(1)) \log \omega.$$

We conclude from (2.4) that

$$|\mathbb{E} g_1(a\mathbf{n}+b)g_2(a\mathbf{n}+b+h)| \gg \varepsilon \tag{2.6}$$

while from (2.5) we conclude that

$$\sup_\alpha \mathbb{E} \left| \sum_{j=1}^H g_1(\mathbf{n}+j)e(\alpha j) \right| = o_{H_- \to \infty}(H) \tag{2.7}$$

for all $H_- \leqslant H \leqslant H_+$. By Fourier expansion we may insert the constraint $1_{a|\mathbf{n}}$ in the left-hand side of (2.7), and thus by Lemma 2.5 we also have

$$\sup_\alpha \mathbb{E}\left|\sum_{j=1}^{H} g_1(a\mathbf{n}+j)e(\alpha j)\right| = o_{H_-\to\infty}(H) \tag{2.8}$$

The logarithmic averaging in the n variable gives an approximate affine invariance to these probabilities and expectations (cf. [18, Proposition 2.4]), which is of fundamental importance to our approach:

Lemma 2.5 (Approximate affine invariance)**.** *Let q be a natural number bounded by H_+, and let r be a fixed integer with $|r| \leqslant H_+$. Then for any event $P(\mathbf{n})$ depending on $\mathbf{n}$, one has*

$$\mathbb{P}(P(\mathbf{n}) \text{ and } \mathbf{n} = r\ (q)) = \frac{1}{q}\mathbb{P}(P(q\mathbf{n}+r)) + o_{A\to\infty}(1).$$

More generally, for any complex-valued random variable $X(\mathbf{n})$ depending on $\mathbf{n}$ and bounded in magnitude by $O(1)$, one has

$$\mathbb{E}(X(\mathbf{n})1_{\mathbf{n}=r\ (q)}) = \frac{1}{q}\mathbb{E}(X(q\mathbf{n}+r)) + o_{A\to\infty}(1).$$

Note in particular that this lemma implies the approximate translation invariance $\mathbb{P}(P(\mathbf{n}+r)) = \mathbb{P}(P(\mathbf{n})) + o_{A\to\infty}(1)$ and $\mathbb{E}(X(\mathbf{n}+r)) = \mathbb{E}(X(\mathbf{n})) + o_{A\to\infty}(1)$ for any $r = O(H_+)$. If we did not perform a logarithmic averaging, then we would still have approximate translation invariance, but we would not necessarily have the more general approximate *affine* invariance, which causes the remainder of our arguments to break down.

Proof. It suffices to prove the latter claim. The left-hand side can be written as

$$\frac{1+o_{A\to\infty}(1)}{\log\omega}\sum_{x/\omega<n\leqslant x: n=r\ (q)}\frac{X(n)}{n}.$$

Making the change of variables $n = qn'+r$, noting that $\frac{1}{n}$ is equal to $\frac{1}{q}\frac{1}{n'} + o_{A\to\infty}(\frac{1}{n'})$ uniformly in n', we can write the previous expression as

$$\frac{1+o_{A\to\infty}(1)}{\log\omega}\sum_{x/\omega<qn'+r\leqslant x}\left(\frac{1}{q}\frac{X(qn'+r)}{n'} + o_{A\to\infty}\left(\frac{1}{n'}\right)\right).$$

The net contribution of the $o_{A\to\infty}(\frac{1}{n'})$ term can be seen to be $o_{A\to\infty}(1)$ (recall that A is assumed large compared to H_+ and hence with q). The constraint $x/\omega < qn'+r \leqslant x$ can be replaced with $x/\omega < n' \leqslant x$ while incurring an error of $O(\frac{1+o_{A\to\infty}(1)}{\log\omega}O(1)) = o_{A\to\infty}(1)$. The claim follows. □

From Lemma 2.5 and (2.6) we have

$$|\mathbb{E}1_{\mathbf{n}=b\ (a)}g_1(\mathbf{n})g_2(\mathbf{n}+h)| \gg \varepsilon. \tag{2.9}$$

Crucially, we can exploit the multiplicativity of g_1, g_2 at medium-sized primes to average this lower bound by further application of Lemma 2.5:

Proposition 2.6. *Let $H_- \leqslant H \leqslant H_+$. Let $\mathcal{P}_H$ denote the set of primes between $\frac{\varepsilon^2}{2}H$ and $\varepsilon^2 H$. For each prime p, let $c_p \in S^1$ denote the coefficient $c_p := \overline{g_1}(p)\overline{g_2}(p)$. Then one has*

$$\left|\mathbb{E} \sum_{p \in \mathcal{P}_H} \sum_{j: j, j+ph \in [1,H]} c_p 1_{a\mathbf{n}+j=pb\ (ap)} g_1(a\mathbf{n}+j) g_2(a\mathbf{n}+j+ph)\right| \gg \varepsilon \frac{H}{\log H}.$$

We remark that in the Liouville case $g_1 = g_2 = \lambda$, we have $c_p = 1$. This leads to some minor simplification in the arguments (in particular, we only need to apply Proposition 2.4 for "major arc" values of α, allowing one to replace [17, Lemma 2.2, Theorem 2.3] by the simpler [17, Theorem A.1]), however it turns out that existing results in the literature (in particular, the restriction theorem for the primes in [11]) allow us to handle the extension to more general c_p without much additional difficulty.

Proof. Write

$$X := \mathbb{E} 1_{\mathbf{n}=b\ (a)} g_1(\mathbf{n}) g_2(\mathbf{n}+h),$$

thus (2.9) tells us that $|X| \gg \varepsilon$. From complete multiplicativity and the definition of c_p we see that

$$1_{\mathbf{n}=b\ (a)} g_1(\mathbf{n}) g_2(\mathbf{n}+h) = c_p 1_{p\mathbf{n}=pb\ (ap)} g_1(p\mathbf{n}) g_2(p\mathbf{n}+ph)$$

and thus

$$\mathbb{E} c_p 1_{p\mathbf{n}=pb\ (ap)} g_1(p\mathbf{n}) g_2(p\mathbf{n}+ph) = X$$

for any $p \in \mathcal{P}_H$. Applying Lemma 2.5, we conclude that

$$\mathbb{E} c_p 1_{\mathbf{n}+j=pb\ (ap)} g_1(\mathbf{n}+j) g_2(\mathbf{n}+j+ph) = \frac{1}{p} X + o_{A\to\infty}(1)$$

for any $1 \leqslant j \leqslant H$ and any $p \in \mathcal{P}_H$. Summing over $j = 1, \ldots, H$, we have

$$\mathbb{E} c_p \sum_{j=1}^{H} 1_{\mathbf{n}+j=pb\ (ap)} g_1(\mathbf{n}+j) g_2(\mathbf{n}+j+ph) = \frac{1}{p} HX + o_{A\to\infty}(1).$$

Thus the quantities

$$\mathbb{E} c_p \sum_{j=1}^{H} 1_{\mathbf{n}+j=pb\ (ap)} g_1(\mathbf{n}+j) g_2(\mathbf{n}+j+ph) 1_{\mathbf{n}=s\ (a)} \tag{2.10}$$

have an average value of $\frac{1}{p}HX + o_{A\to\infty}(1)$ if one averages s over $\mathbb{Z}/a\mathbb{Z}$.

Suppose one increments s to $s+1$ in the expression (2.10). By Lemma 2.5, this is equivalent (up to an error of $o_{A\to\infty}(1)$) of letting j range from 2 to $H+1$, rather than from 1 to H. This alters the sum $\sum_{j=1}^{H} 1_{\mathbf{n}+j=pb\ (ap)} g_1(\mathbf{n}+j) g_2(\mathbf{n}+j+ph) 1_{\mathbf{n}=s\ (a)}$ by $O(1)$ with probability

$O(1/p)$, and leaves the sum unchanged otherwise. Thus the expression (2.10) only varies by $O(1/p)$ when incrementing s to $s+1$. As such, one can estimate (2.10) by its average and conclude in particular that

$$\mathbb{E} c_p \sum_{j=1}^{H} 1_{\mathbf{n}+j=pb\ (ap)} g_1(\mathbf{n}+j) g_2(\mathbf{n}+j+ph) 1_{\mathbf{n}=0\ (a)} = \frac{HX}{ap} + O_a\left(\frac{1}{p}\right)$$

(here we have absorbed errors of the form $O_a(1) \times o_{A\to\infty}(1)$ into the $O_a(\frac{1}{p})$ term). Summing over $\mathcal{P}_H$, we conclude that

$$\mathbb{E} \sum_{j=1}^{H} \sum_{p\in\mathcal{P}_H} c_p 1_{\mathbf{n}+j=pb\ (ap)} g_1(\mathbf{n}+j) g_2(\mathbf{n}+j+ph) 1_{\mathbf{n}=0\ (a)} = \left(\frac{HX}{a} + O_a(1)\right) \sum_{p\in\mathcal{P}} \frac{1}{p}$$

and hence by the prime number theorem and the lower bound $|X| \gg \varepsilon$, one has

$$\left| \mathbb{E} \sum_{j=1}^{H} \sum_{p\in\mathcal{P}_H} c_p 1_{\mathbf{n}+j=pb\ (ap)} g_1(\mathbf{n}+j) g_2(\mathbf{n}+j+ph) 1_{\mathbf{n}=0\ (a)} \right| \gg \varepsilon \frac{H}{a \log H}.$$

Applying Lemma 2.5, we obtain

$$\left| \mathbb{E} \sum_{j=1}^{H} \sum_{p\in\mathcal{P}_H} c_p 1_{a\mathbf{n}+j=pb\ (ap)} g_1(a\mathbf{n}+j) g_2(a\mathbf{n}+j+ph) \right| \gg \varepsilon \frac{H}{\log H}.$$

If $j+ph$ lies outside of the interval $[1,H]$, then j lies in either $[1, |h|\varepsilon^2 H]$ or $[(1-|h|\varepsilon^2)H, H]$. The contribution of these values of j can be easily estimated to be $O(\sum_{p\in\mathcal{P}_H} \frac{|h|\varepsilon^2 H}{p}) = O_h(\varepsilon^2 \frac{H}{\log H})$, so from the smallness of ε we may discard these intervals and conclude the claim. □

We will shortly need to deploy the theory of Shannon entropy, at which point we encounter the inconvenient fact that g could potentially take an infinite number of values and thus have unbounded Shannon entropy. To get around this, we perform a standard discretisation. Namely, define $g_{i,\varepsilon^2}(n)$ for $i = 1,2$ to be $g_i(n)$ rounded to the nearest element of the lattice $\varepsilon^2\mathbb{Z}]i]$, where $\mathbb{Z}[i]$ denotes the Gaussian integers. (We break ties arbitrarily.) This function is no longer multiplicative, but it takes at most $O_\varepsilon(1)$ values, it is bounded in magnitude by $O(1)$, and we have $g_{i,\varepsilon^2} = g_i + O(\varepsilon^2)$ for $i = 1,2$. Thus from the above proposition and the triangle inequality, we have

$$\left| \mathbb{E} \sum_{p\in\mathcal{P}_H} c_p \sum_{j: j, j+ph \in [1,H]} 1_{a\mathbf{n}+j=pb\ (ap)} g_{1,\varepsilon^2}(a\mathbf{n}+j) g_{2,\varepsilon^2}(a\mathbf{n}+j+ph) \right| \gg \varepsilon \frac{H}{\log H}$$

since the error incurred by replacing g_i with g_{i,ε^2} can be computed to be $O_a(\varepsilon^2 \sum_{p\in\mathcal{P}_H} \frac{H}{p}) = O_a(\varepsilon^2 \frac{H}{\log H})$. We rewrite this inequality as

$$|\mathbb{E} F(\mathbf{X}_H, \mathbf{Y}_H)| \gg \varepsilon \frac{H}{\log H} \qquad (2.11)$$

where $\mathbf{X}_H$ is the discrete random variable

$$\mathbf{X}_H := (g_{1,\varepsilon^2}(a\mathbf{n} + j))_{i=1,2;j=1,\dots,H}$$

(taking values in $(\varepsilon^2\mathbb{Z}[i])^{2H}$), $\mathbf{Y}_H$ is the random variable

$$\mathbf{Y}_H := \mathbf{n}\ (P_H)$$

(taking values in $\mathbb{Z}/P_H\mathbb{Z}$) where $P_H := \prod_{p\in\mathcal{P}_H} p$, and $F : (\varepsilon^2\mathbb{Z}[i])^{2H} \times \mathbb{Z}/P_H\mathbb{Z} \to \mathbb{C}$ is the function

$$F((x_{i,j})_{i=1,2;j=1,\dots,H}, y\ (P_H)) := \sum_{p\in\mathcal{P}_H} c_p \sum_{j: j, j+ph\in[1,H]} 1_{ay+j=pb\ (ap)} x_{1,j} x_{2,j+ph}. \tag{2.12}$$

(Note that the residue class $ay\ (ap)$ is well defined for $p \in \mathbb{Z}/P_H\mathbb{Z}$ and $p \in \mathcal{P}_H$, noting that P_H is coprime to a.)

It is thus of interest to try to calculate the typical value of $F(\mathbf{X}_H, \mathbf{Y}_H)$. One can interpret $F(\mathbf{X}_H, \mathbf{Y}_H)$ as a “bilinear” expression of the components of $\mathbf{X}_H$ along a certain random graph determined by $\mathbf{Y}_H$. A key difficulty is that the random variables $\mathbf{X}_H$ and $\mathbf{Y}_H$ are not independent, and could potentially be coupled together in an adversarial fashion. In this worst case, this would require one to establish a suitable “expander” property for the random graph associated to $\mathbf{Y}_H$ that would ensure cancellation in the sum regardless of what values that $\mathbf{X}_H$ will take. It may well be that such an expansion property holds (with high probability, of course). However, we can avoid having to establish such a strong expansion property by taking advantage of an “entropy decrement argument” (basically an application of the pigeonhole principle) to give some weak independence between $\mathbf{X}_H$ and $\mathbf{Y}_H$ for at least one choice of H between H_- and H_+. Once one obtains such a weak independence, it turns out that one only needs to show that for a typical choice of $\mathbf{X}_H$, that $F(\mathbf{X}_H, \mathbf{Y}_H)$ is small for *most* choices of $\mathbf{Y}_H$, where we allow a (nearly) exponentially small failure set for the $\mathbf{Y}_H$. This turns out to be much easier to establish than the expander graph property, being obtainable from standard concentration of measure inequalities (such as Hoeffding’s inequality), and an application of the Hardy-Littlewood circle method.

The entropy decrement argument can be viewed as a quantitative variant of the construction of the Kolmogorov-Sinai entropy of a topological dynamical system (see e.g. [1]), but we will not explicitly use the language of topological dynamics here.

3. The entropy decrement argument

We now finish the proof of Theorem 1.3. We begin by briefly reviewing the basic Shannon inequalities from information theory.

Recall that if $\mathbf{X}$ is a discrete random variable (taking at most countably many values), the *Shannon entropy* $\mathbb{H}(\mathbf{X})$ is defined[3] by the formula

$$\mathbb{H}(\mathbf{X}) := \sum_x \mathbb{P}(\mathbf{X} = x) \log \frac{1}{\mathbb{P}(\mathbf{X} = x)}$$

where x takes values in the essential range of $\mathbf{X}$ (that is to say, those x for which $\mathbb{P}(\mathbf{X} = x)$ is nonzero). A standard computation then gives the identity

$$\mathbb{H}(\mathbf{X}, \mathbf{Y}) = \mathbb{H}(\mathbf{X}|\mathbf{Y}) + \mathbb{H}(\mathbf{Y}) = \mathbb{H}(\mathbf{X}) + \mathbb{H}(\mathbf{Y}|\mathbf{X}) \tag{3.1}$$

for the joint entropy $\mathbb{H}(\mathbf{X}, \mathbf{Y})$ of the random variable $(\mathbf{X}, \mathbf{Y})$, where the *relative entropy* $\mathbb{H}(\mathbf{X}|\mathbf{Y})$ is defined by the formulae

$$\mathbb{H}(\mathbf{X}|\mathbf{Y}) := \sum_y \mathbb{P}(\mathbf{Y} = y) \mathbb{H}(\mathbf{X}|\mathbf{Y} = y) \tag{3.2}$$

(with y ranging over the essential range of $\mathbf{Y}$) and

$$\mathbb{H}(\mathbf{X}|\mathbf{Y} = y) := \sum_x \mathbb{P}(\mathbf{X} = x|\mathbf{Y} = y) \log \frac{1}{\mathbb{P}(\mathbf{X} = x|\mathbf{Y} = y)}$$

with $\mathbb{P}(E|F) := \mathbb{P}(E \wedge F)/\mathbb{P}(F)$ being the conditional probability of E relative to F, and the sum is over the essential range of $\mathbf{X}$ conditioned to $\mathbf{Y} = y$. From the concavity of the function $x \mapsto x \log \frac{1}{x}$ and Jensen's inequality we have

$$\mathbb{H}(\mathbf{X}|\mathbf{Y}) \leqslant \mathbb{H}(\mathbf{X}) \tag{3.3}$$

so we conclude the subadditivity of entropy

$$\mathbb{H}(\mathbf{X}, \mathbf{Y}) \leqslant \mathbb{H}(\mathbf{X}) + \mathbb{H}(\mathbf{Y}). \tag{3.4}$$

If we define the *mutual information*

$$\mathbb{I}(\mathbf{X}, \mathbf{Y}) := \mathbb{H}(\mathbf{X}) + \mathbb{H}(\mathbf{Y}) - \mathbb{H}(\mathbf{X}, \mathbf{Y}) = \mathbb{H}(\mathbf{X}) - \mathbb{H}(\mathbf{X}|\mathbf{Y}) = \mathbb{H}(\mathbf{Y}) - \mathbb{H}(\mathbf{Y}|\mathbf{X}) \tag{3.5}$$

between two discrete random variables $\mathbf{X}, \mathbf{Y}$, we thus see that $\mathbb{I}(\mathbf{X}, \mathbf{Y}) = \mathbb{I}(\mathbf{Y}, \mathbf{X}) \geqslant 0$.

Conditioning the random variables $\mathbf{X}, \mathbf{Y}$ to an auxiliary discrete random variable $\mathbf{Z}$, we conclude the relative subadditivity of entropy

$$\mathbb{H}(\mathbf{X}, \mathbf{Y}|\mathbf{Z}) \leqslant \mathbb{H}(\mathbf{X}|\mathbf{Z}) + \mathbb{H}(\mathbf{Y}|\mathbf{Z}). \tag{3.6}$$

Finally, a further application of Jensen's inequality gives the bound

$$\mathbb{H}(\mathbf{X}) \leqslant \log N \tag{3.7}$$

whenever $\mathbf{X}$ takes on at most N values.

[3] In the information theory literature, the logarithm to base 2 is often used to define entropy, rather than the natural logarithm, in which case $\mathbb{H}(\mathbf{X})$ can be interpreted as the number of bits needed to describe $\mathbf{X}$ on the average. One could use this choice of base in the arguments below if desired, but ultimately the choice of base is a normalisation which has no impact on the final bounds.

Recall the discrete random variables $\mathbf{X}_H, \mathbf{Y}_H$ defined previously. From (3.7), (3.4), and the fact that each component of $\mathbf{X}_H$ takes on only $O_\varepsilon(1)$ values, we have the upper bound

$$0 \leqslant \mathbb{H}(\mathbf{X}_H) \ll_\varepsilon H. \tag{3.8}$$

Note that $\mathbf{Y}_H$ is within $o_{A\to\infty}(1)$ (in any reasonable metric) of being uniformly distributed on $\mathbb{Z}/P_H\mathbb{Z}$, thus

$$\mathbb{H}(\mathbf{Y}_H) = \log P_H - o_{A\to\infty}(1). \tag{3.9}$$

In particular, from the prime number theorem we have the crude bound

$$\mathbb{H}(\mathbf{Y}_H) \ll H \tag{3.10}$$

for all $H_- \leqslant H \leqslant H_+$.

Let us temporarily define the variant

$$\mathbf{X}_{H_1,H_1+H_2} := \mathbb{H}((g_{i,\varepsilon^2}(\mathbf{n}+j))_{i=1,2;j=H_1+1,\dots,H_1+H_2}$$

of $\mathbf{X}_H$, where H_1, H_2 are natural numbers. From the approximate translation invariance provided by Lemma 2.5, we see that

$$\mathbb{H}(\mathbf{X}_{H_1,H_1+H_2}) = \mathbb{H}(\mathbf{X}_{H_2}) + o_{A\to\infty}(1)$$

for any $H_1, H_2 \leqslant H_+$; applying (3.4), and noting that $\mathbf{X}_{H_1+H_2}$ is the concatenation of $\mathbf{X}_{H_1}$ and $\mathbf{X}_{H_1,H_1+H_2}$, we obtain the subadditivity property

$$\mathbb{H}(\mathbf{X}_{H_1+H_2}) \leqslant \mathbb{H}(\mathbf{X}_{H_1}) + \mathbb{H}(\mathbf{X}_{H_2}) + o_{A\to\infty}(1)$$

for any natural numbers $H_1, H_2 \leqslant H_+$.

We can improve this inequality if X_H shares some mutual information with Y_H, as Y_H does not generate any entropy upon translation. Indeed, from Lemma 2.5 again, we see for any natural numbers H, H_1, H_2 between H_- and H_+ that

$$\mathbb{H}(\mathbf{X}_{H_1,H_1+H_2}|\mathbf{n}+H_1\ (P_H)) = \mathbb{H}(\mathbf{X}_{H_2}|\mathbf{n}\ (P_H)) + o_{A\to\infty}(1).$$

But $\mathbf{n}+H_1\ (P_H)$ conveys exactly the same information as $\mathbf{n}\ (P_H)$ (they generate exactly the same finite σ algebra of events), so

$$\mathbb{H}(\mathbf{X}_{H_1,H_1+H_2}|\mathbf{n}+H_1\ (P_H)) = \mathbb{H}(\mathbf{X}_{H_2}|\mathbf{n}\ (P_H)).$$

Inserting these identities into (3.6) and recalling that $\mathbf{Y}_H = \mathbf{n}\ (P_H)$, we obtain the relative subadditivity property

$$\mathbb{H}(\mathbf{X}_{H_1+H_2}|\mathbf{Y}_H) \leqslant \mathbb{H}(\mathbf{X}_{H_1}|\mathbf{Y}_H) + \mathbb{H}(\mathbf{X}_{H_2}|\mathbf{Y}_H) + o_{A\to\infty}(1)$$

for any H, H_1, H_2 between H_- and H_+. Iterating this, we conclude in particular that

$$\mathbb{H}(\mathbf{X}_{kH}|\mathbf{Y}_H) \leqslant k\mathbb{H}(\mathbf{X}_H|\mathbf{Y}_H) + o_{A\to\infty}(1)$$

for any natural numbers k, H with $H_- \leqslant H \leqslant kH \leqslant H_+$ (note that the number of iterations here is at most H_+, so that the $o_{A\to\infty}(1)$ error stays under control). From this and (3.1), (3.3), (3.5) we see that

$$\begin{aligned}\mathbb{H}(\mathbf{X}_{kH}) &\leqslant k\mathbb{H}(\mathbf{X}_H|\mathbf{Y}_H) + \mathbb{H}(\mathbf{Y}_H) + o_{A\to\infty}(1)\\ &= k\mathbb{H}(\mathbf{X}_H) - k\mathbb{I}(\mathbf{X}_H, \mathbf{Y}_H) + +\mathbb{H}(\mathbf{Y}_H) + o_{A\to\infty}(1)\end{aligned}$$

which on dividing by kH and using (3.10) gives

$$\frac{\mathbb{H}(\mathbf{X}_{kH})}{kH} \leqslant \frac{\mathbb{H}(\mathbf{X}_H)}{H} - \frac{\mathbb{I}(\mathbf{X}_H, \mathbf{Y}_H)}{H} + O\left(\frac{1}{k}\right), \tag{3.11}$$

whenever $H_- \leqslant H \leqslant kH \leqslant H_+$ (note that we can absorb the $o_{A\to\infty}(1)$ term in the $O(1/k)$ term since $k \leqslant H_+$).

We can iterate this inequality and use an “entropy decrement argument” to get a non-trivial upper bounds on the mutual information $\mathbb{I}(\mathbf{X}_H, \mathbf{Y}_H)$ for some large H:

Lemma 3.1 (Entropy decrement argument)**.** *There exists a natural number H between H_- and H_+, which is a multiple of a, and such that*

$$\mathbb{I}(\mathbf{X}_H, \mathbf{Y}_H) \leqslant \frac{H}{\log H \log\log\log H}.$$

As we shall see later, the key point here is that this bound is not only better than the trivial bound of $O(H)$ coming from (3.10), but is (barely!) smaller than $H/\log H$ in the limit as $H \to \infty$; in particular, the mutual information between $\mathbf{X}_H$ and $\mathbf{Y}_H$ is smaller than the number $|\mathcal{P}_H|$ of primes one is using to define $F(\mathbf{X}_H, \mathbf{Y}_H)$. One may think of this lemma as providing a weak independence between $\mathbf{X}_H$ and $\mathbf{Y}_H$ for certain large H.

Proof. Suppose for sake of contradiction that one has

$$\mathbb{I}(\mathbf{X}_H, \mathbf{Y}_H) > \frac{H}{\log H \log\log\log H}$$

for all $H_- \leqslant H \leqslant H_+$ that are multiples of a. Let C_0 be a sufficiently large natural number depending on H_-, and let J be a sufficiently large natural number depending on C_0, H_-, ε. We may assume that H_+ is sufficiently large depending on H_-, C_0, J.

Let us recursively define the natural numbers $H_- \leqslant H_1 \leqslant H_2 \leqslant \dots \leqslant H_J$ by setting $H_1 := aH_-$ and

$$H_{j+1} := H_j \lfloor C_0 \log H_j \log\log\log H_j \rfloor$$

for all $1 \leqslant j < J$. Note that if H_+ is sufficiently large depending on H_-, C_0, J, then all the H_j will lie between H_- and H_+ and are multiples of a. For C_0 large enough, we see from (3.11), (3.10) with H, k replaced by H_j and $\lfloor C_0 \log H_j \log\log\log H_j \rfloor$ respectively that

$$\frac{\mathbb{H}(\mathbf{X}_{H_{j+1}})}{H_{j+1}} \leqslant \frac{\mathbb{H}(\mathbf{X}_{H_j})}{H_j} - \frac{1}{2\log H_j \log\log\log H_j}$$

for all $1 \leqslant j < J$. (The $o_{A\to\infty}(1)$ error may be absorbed as we are assuming A to be large.) On the other hand, an easy induction shows that there exists $B \geqslant 10^{10}$ (depending on C_0, H_-) such that

$$H_j \leqslant \exp(Bj \log j)$$

for all $2 \leqslant j \leqslant J$. Thus we have

$$\frac{\mathbb{H}(\mathbf{X}_{H_{j+1}})}{H_{j+1}} \leqslant \frac{\mathbb{H}(\mathbf{X}_{H_j})}{H_j} - \frac{1}{2Bj\log j \log\log(Bj\log j)}$$

for all $2 \leqslant j \leqslant J$, which on telescoping using (3.8) gives the bound

$$\sum_{j=2}^{J} \frac{1}{2Bj\log j\log\log(Bj\log j)} \ll_\varepsilon 1.$$

But the sum on the left-hand side diverges (very slowly!) in the limit $J \to \infty$, and so we obtain a contradiction by choosing J (and then H_+) large enough. □

From the above lemma we can find an H between H_- and H_+ that is a multiple of a, such that

$$\mathbb{I}(\mathbf{X}_H, \mathbf{Y}_H) = o_{H_-\to\infty}\left(\frac{H}{\log H}\right). \tag{3.12}$$

Fix this value of H. From (3.5) and (3.12) we have

$$\sum_x \mathbb{P}(\mathbf{X}_H = x)\left(\mathbb{H}(\mathbf{Y}_H) - \mathbb{H}(\mathbf{Y}_H|\mathbf{X}_H = x)\right) = o_{H_-\to\infty}\left(\frac{H}{\log H}\right).$$

By (3.3), the summands are non-negative. Thus, by Markov's inequality, we see that with probability $1 - o_{H_-\to\infty}(1)$, the random variable $\mathbf{X}_H$ attains a value x for which

$$\mathbb{H}(\mathbf{Y}_H) - \mathbb{H}(\mathbf{Y}_H|\mathbf{X}_H = x) = o_{H_-\to\infty}\left(\frac{H}{\log H}\right). \tag{3.13}$$

Informally, this estimate asserts that $\mathbf{Y}_H$ remains somewhat uniformly distributed across $\mathbb{Z}/P_H\mathbb{Z}$ even after one conditions $\mathbf{X}_H$ to equal x, in the sense that this conditioned random variable cannot concentrate too much mass into a small region. More precisely, we have

Lemma 3.2 (Weak uniform distribution). *Let x be a value in the range of $\mathbf{X}_H$ that obeys the bound* (3.13)*. Let E_x be a subset of $\mathbb{Z}/P_H\mathbb{Z}$ (which can depend on x) of cardinality*

$$|E_x| \leqslant \exp\left(-\varepsilon^7 \frac{H}{\log H}\right) P_H.$$

Then one has

$$\mathbb{P}(\mathbf{Y}_H \in E_x|\mathbf{X}_H = x) = o_{H_-\to\infty}(1).$$

The quantity ε^7 here could be replaced by any other function of ε, but we use this particular choice to match with Lemma 3.3 below.

Proof. Applying (3.3), (3.1) (conditioned to the event $\mathbf{X}_H = x$) we have

$$\mathbb{H}(\mathbf{Y}_H|\mathbf{X}_H = x, 1_{E_x}(\mathbf{Y}_H)) \geqslant \mathbb{H}(\mathbf{Y}_H|\mathbf{X}_H = x) - \mathbb{H}(1_{E_x}(\mathbf{Y}_H)|\mathbf{X}_H = x)$$

and thus by (3.13), (3.2)

$$\begin{aligned}&\mathbb{P}(\mathbf{Y}_H \in E_x|\mathbf{X}_H = x)\mathbb{H}(\mathbf{Y}_H|\mathbf{X}_H = x, \mathbf{Y}_H \in E_x)\\ &\quad + \mathbb{P}(\mathbf{Y}_H \notin E_x|\mathbf{X}_H = x)\mathbb{H}(\mathbf{Y}_H|\mathbf{X}_H = x, \mathbf{Y}_H \notin E_x)\\ &\qquad \geqslant \mathbb{H}(\mathbf{Y}_H) - \mathbb{H}(1_{E_x}(\mathbf{Y}_H)|\mathbf{X}_H = x) - o_{H_-\to\infty}\left(\frac{H}{\log H}\right).\end{aligned}$$

By (3.7), $\mathbb{H}(1_{E_x}(\mathbf{Y}_H)|\mathbf{X}_H = x)$ is bounded by $\log 2$ and so this term can be absorbed in the $o_{H_-\to\infty}(H/\log H)$ error. From (3.3) we have

$$\mathbb{H}(\mathbf{Y}_H|\mathbf{X}_H = x, \mathbf{Y}_H \notin E_x) \leqslant \mathbb{H}(\mathbf{Y}_H)$$

and hence

$$\mathbb{P}(\mathbf{Y}_H \in E_x|\mathbf{X}_H = x)\left(\mathbb{H}(\mathbf{Y}_H) - \mathbb{H}(\mathbf{Y}_H|\mathbf{X}_H = x, \mathbf{Y}_H \in E_x)\right) \leqslant o_{H_-\to\infty}\left(\frac{H}{\log H}\right).$$

But from (3.7) one has

$$\mathbb{H}(\mathbf{Y}_H|\mathbf{X}_H = x, \mathbf{Y}_H \in E_x) \geqslant \log|E_x| \geqslant \log P_H - \varepsilon^7\frac{H}{\log H}$$

and the claim then follows from (3.9) (recalling that H_- is large depending on ε). □

We can use this weak uniform distribution to show that $F(\mathbf{X}_H, \mathbf{Y}_H)$ concentrates as a function of $\mathbf{Y}_H$. We first observe

Lemma 3.3 (Hoeffding inequality). *Let x lie in the range of X_H. Let E_x denote the set of all $y \in \mathbb{Z}/P_H\mathbb{Z}$ such that*

$$\left|F(x,y) - \frac{1}{P_H}\sum_{y'\in\mathbb{Z}/P_H\mathbb{Z}} F(x,y')\right| \geqslant \varepsilon^2\frac{H}{\log H}.$$

Then

$$|E_x| \leqslant \exp\left(-\varepsilon^7\frac{H}{\log H}\right)P_H.$$

Proof. We interpret this inequality probabilistically. Let $\mathbf{y}$ be drawn uniformly at random from $\mathbb{Z}/P_H\mathbb{Z}$, then our task is to show that

$$\mathbb{E}\left(|F(x,\mathbf{y}) - \mathbb{E}F(x,\mathbf{y})| \geqslant \varepsilon^2\frac{H}{\log H}\right) \leqslant \exp\left(-\varepsilon^7\frac{H}{\log H}\right).$$

We can write

$$F(x,\mathbf{y}) = \sum_{p\in\mathcal{P}} F_p(x,\mathbf{y})$$

where

$$F_p(x,\mathbf{y}) := c_p \sum_{j:j,j+ph\in[1,H]} 1_{a\mathbf{y}+j=pb\ (ap)} x_{1,j}x_{2,j+ph}. \tag{3.14}$$

From the Chinese remainder theorem that the residue classes $a\mathbf{y}$ (ap) are jointly independent in p, thus the $F_p(x,\mathbf{y})$ are also jointly independent in p. On the other hand, since all $p \in \mathcal{P}_H$ lie in the interval $\frac{\varepsilon^2}{2}H \leqslant p \leqslant \varepsilon^2 H$, we have the deterministic bound $|F_p(x,\mathbf{y})| \leqslant C/\varepsilon^2$ for some absolute constant C. Applying the Hoeffding inequality [14], we conclude that

$$\mathbb{E}\left(|F(x,\mathbf{y}) - \mathbb{E}F(x,\mathbf{y})| \geqslant \varepsilon^2\frac{H}{\log H}\right) \ll \exp\left(-\frac{2(\varepsilon^2\frac{H}{\log H})^2}{(C/\varepsilon^2)^2|P_H|}\right).$$

From the prime number theorem we have $|P_H| \ll \varepsilon^2\frac{H}{\log H}$, and the claim follows (as ε is small and H is large). □

Combining this lemma with Lemma 3.2, we conclude that with probability $1 - o_{H_-\to\infty}(1)$, $\mathbf{X}_H$ attains a value x for which

$$\mathbb{P}\left(\left|F(x,\mathbf{Y}_H) - \frac{1}{P_H}\sum_{y\in\mathbb{Z}/P_H\mathbb{Z}} F(x,y)\right| \geqslant \varepsilon^2\frac{H}{\log H}\right) = o_{H_-\to\infty}(1).$$

By Fubini's theorem, this implies that

$$F(\mathbf{X}_H,\mathbf{Y}_H) = \frac{1}{P_H}\sum_{y\in\mathbb{Z}/P_H\mathbb{Z}} F(\mathbf{X}_H,y) + O\left(\varepsilon^2\frac{H}{\log H}\right)$$

with probability $1 - o_{H_-\to\infty}(1)$. On the other hand, from the triangle inequality, (2.12), and the prime number theorem we have

$$F(x,y) \ll \frac{H}{\log H}.$$

We can thus take expectations and conclude that

$$\mathbb{E}F(\mathbf{X}_H,\mathbf{Y}_H) = \mathbb{E}\frac{1}{P_H}\sum_{y\in\mathbb{Z}/P_H\mathbb{Z}} F(\mathbf{X}_H,y) + O\left(\varepsilon^2\frac{H}{\log H}\right),$$

and hence by (2.11) we have

$$\left|\mathbb{E}\frac{1}{P_H}\sum_{y\in\mathbb{Z}/P_H\mathbb{Z}} F(\mathbf{X}_H,y)\right| \gg \varepsilon\frac{H}{\log H}. \tag{3.15}$$

The advantage here is that we have decoupled the x and y variables, and the y average is now easy to compute. Indeed, from the Chinese remainder theorem and (3.14) we see that

$$\frac{1}{P_H}\sum_{y\in\mathbb{Z}/P_H\mathbb{Z}} F_p(x,y) = \frac{c_p}{p}\sum_{j:j,j+ph\in[1,H]} 1_{j=pb\ (a)} x_{1,j}x_{2,j+ph}$$

for any x and any $p \in \mathcal{P}$, and on summing in $\mathcal{P}$ and inserting into (3.15), we conclude that

$$\left| \mathbb{E} \sum_{p \in \mathcal{P}_H} \frac{c_p}{p} \sum_{j: j, j+ph \in [1,H]} 1_{j = pb\ (a)} g_{1,\varepsilon^2}(a\mathbf{n} + j) g_{2,\varepsilon^2}(a\mathbf{n} + j + ph) \right| \gg \varepsilon \frac{H}{\log H}.$$

Since $g_i = g_{i,\varepsilon^2} + O(\varepsilon^2)$ and $g_i, g_{i,\varepsilon^2} = O(1)$ for $i = 1, 2$, we we can replace g_{i,ε^2} by g_i on the left-hand side at the cost of an error of $O(\varepsilon^2 \sum_{p \in \mathcal{P}_H} \frac{H}{p}) = O(\varepsilon^2 \frac{H}{\log H})$. We thus have

$$\left| \mathbb{E} \sum_{p \in \mathcal{P}_H} \frac{c_p}{p} \sum_{j: j, j+ph \in [1,H]} 1_{j = pb\ (a)} g_1(a\mathbf{n} + j) g_2(a\mathbf{n} + j + ph) \right| \gg \varepsilon \frac{H}{\log H}. \tag{3.16}$$

On the other hand, by using the Hardy-Littlewood circle method, we can obtain the following deterministic estimate for the expression inside the expectation.

Lemma 3.4 (Circle method estimate). *For any $\alpha \in \mathbb{R}/\mathbb{Z}$, let $S_H(\alpha)$ denote the exponential sum*

$$S_H(\alpha) := \sum_{p \in \mathcal{P}_H} \frac{c_p}{p} e(\alpha p) \tag{3.17}$$

and let Ξ_H denote the elements $\xi \in \mathbb{Z}/H\mathbb{Z}$ for which

$$\left| S_H\left(-\frac{(b+h)\eta}{a} - \frac{h\xi}{H} \right) \right| \geqslant \frac{\varepsilon^2}{\log H}$$

for some $\eta \in \mathbb{Z}/a\mathbb{Z}$. For $j = 1, \dots, H$, let $x_{1,j}, x_{2,j}$ be complex numbers bounded in magnitude by one. Then

$$\begin{aligned} &\sum_{p \in \mathcal{P}_H} \frac{c_p}{p} \sum_{j: j, j+ph \in [1,H]} 1_{j = pb\ (a)} x_{1,j} x_{2,j+ph} \\ &\ll_{a,h} \frac{H}{\log H} \left(\varepsilon^2 + \sum_{\xi \in \Xi_H} \frac{1}{H} \left| \sum_{j=1}^H x_{1,j} e(-j\xi/H) \right| \right). \end{aligned} \tag{3.18}$$

Proof. We extend $x_{1,j}$, $x_{2,j}$ periodically with period H. If we remove the constraint that $j + ph \in [1, H]$, we incur an error of $O(\sum_{p \in \mathcal{P}_H} \frac{1}{p} |h| p) = O(|h| \varepsilon^2 \frac{H}{\log H})$ which is acceptable. Thus, viewing j now as an element of $\mathbb{Z}/H\mathbb{Z}$, we may replace the left-hand side of (3.18) by

$$\sum_{p \in \mathcal{P}_H} \frac{c_p}{p} \sum_{j \in \mathbb{Z}/H\mathbb{Z}} 1_{j = pb\ (a)} x_{1,j} x_{2,j+ph}. \tag{3.19}$$

We perform a Fourier expansion

$$x_{i,j} = \sum_{\xi \in \mathbb{Z}/H\mathbb{Z}} G_i(\xi) e(j\xi/H)$$

for $i = 1, 2$, where

$$G_i(\xi) := \frac{1}{H} \sum_{j \in \mathbb{Z}/H\mathbb{Z}} x_{i,j} e(-j\xi/H).$$

We can thus expand (3.19) as

$$\sum_{\xi,\xi' \in \mathbb{Z}/H\mathbb{Z}} G_1(\xi) G_2(\xi') \sum_{p \in \mathcal{P}_H} \frac{c_p}{p} \sum_{j \in \mathbb{Z}/H\mathbb{Z}} 1_{j=pb\ (a)} e\left(\frac{j\xi}{H} - \frac{(j+ph)\xi'}{H} \right).$$

The inner sum vanishes unless $\xi' = \xi + \frac{H}{a}\eta$ for some $\eta \in \mathbb{Z}/a\mathbb{Z}$, in which case one has

$$\sum_{j \in \mathbb{Z}/H\mathbb{Z}} 1_{j=pb\ (a)} e\left(\frac{j\xi}{H} - \frac{(j+ph)\xi'}{H} \right) = \frac{H}{a} e\left(-\frac{p(b+h)\eta}{a} - \frac{ph\xi}{H} \right)$$

(recall that H was chosen to be a multiple of a), and thus by (3.17) we can write (3.19) as

$$\sum_{\eta \in \mathbb{Z}/a\mathbb{Z}} \sum_{\xi \in \mathbb{Z}/H\mathbb{Z}} G_1(\xi) G_2(\xi + \frac{H}{a}\eta) S_H\left(-\frac{(b+h)\eta}{a} - \frac{h\xi}{H} \right) \gg \frac{a\varepsilon}{\log H}.$$

From the Plancherel identity and the Cauchy-Schwarz inequality, one has

$$\sum_{\xi \in \mathbb{Z}/H\mathbb{Z}} |G_1(\xi)||G_2(\xi + \frac{H}{a}\eta)| \ll 1,$$

so those $\xi \notin \Xi_H$ give an acceptable contribution. For the remaining ξ, we bound $G_2(\xi + \frac{H}{a}\eta)$ crudely by $O(1)$ and $S_H(-\frac{(b+h)\eta}{a} - \frac{h\xi}{H})$ by $O(\frac{1}{\log H})$ and use the triangle inequality to obtain the claim. □

Combining this lemma with (3.16), we conclude that

$$\sum_{\xi \in \Xi_H} \mathbb{E} \frac{1}{H} \left| \sum_{j=1}^{H} g_1(a\mathbf{n} + j) e(-j\xi/H) \right| \gg_{a,h} \varepsilon.$$

By (2.8) we thus have

$$\varepsilon \ll_{a,h} o_{H_- \to \infty}(|\Xi_H|).$$

To conclude the desired contradiction, it thus suffices (by taking H_- large enough) to show

Lemma 3.5 (Restriction theorem for the primes)**.** *We have* $|\Xi_H| \ll_{a,h,\varepsilon} 1$.

Proof. Applying [11, Theorem 4.7] with $p = 4$, $N := H$, $R := N^{1/10}$, and a_n set equal to $\frac{c_p}{p\beta_R(p)}$ when n is a prime in $\mathcal{P}_H$, and $a_n = 0$ otherwise, we have[4]

$$\sum_{k\in\mathbb{Z}/aH\mathbb{Z}} \left|S_H\left(\frac{k}{H}\right)\right|^4 \ll_{\varepsilon,a} \frac{1}{\log^4 H}$$

and thus by Markov's inequality we have $|S_H(\frac{k}{H})| \geqslant \frac{\varepsilon^2}{H}$ for at most $O_{\varepsilon,a}(1)$ values of $k \in \mathbb{Z}/H\mathbb{Z}$. The claim follows. □

Remark 3.6. In the special case $g_1 = g_2 = \lambda$ (or more generally when g_2 is the complex conjugate of g_1, we have $c_p = 1$, and the exponential sum $S_H(\alpha)$ can then be handled by the Vinogradov estimates for exponential sums over primes (see e.g. [15, §13.5]). In that case, one can compute Ξ_H fairly explicitly; it basically consists of those frequencies ξ which are "major arc" in the sense that ξ/H is close to a rational a/q of bounded denominator q. As remarked previously, this allows for a slight simplification in the arguments in that the exponential sum estimates in [17, Lemma 2.2, Theorem 2.3] can be replaced with the simpler estimate in [17, Theorem A.1]; also, the quantitative bounds in Theorem 1.2 should improve if one uses this approach. However, for more general choices of g_1, g_2, the coefficients c_p are essentially arbitrary unit phases, and the frequency set Ξ_H need not be contained within major arcs.

4. Further remarks

It is natural to ask if the arguments can be extended to higher point correlations than the $k = 2$ case, for instance to bound sums such as the three-point correlation

$$\sum_{x/\omega<n\leqslant x} \frac{\lambda(n)\lambda(n+1)\lambda(n+2)}{n}.$$

Most of the above arguments carry through to this case. However, the "bilinear" left-hand side of (3.19) will be replaced by a "trilinear" expression such ash

$$\left|\sum_{p\in\mathcal{P}_H} \frac{1}{p} \sum_{j\in\mathbb{Z}/H\mathbb{Z}} x_{1,j}x_{2,j+p}x_{3,j+2p}\right|.$$

These sorts of sums have been studied in the ergodic theory literature [7], [25]. Roughly speaking, the analysis there shows that these sums

[4] As an alternative proof of this estimate, one can use standard Fourier-analytic manipulations to rewrite the left-hand side as $H\sum_{p_1.p_2,p_3,p_4\in\mathcal{P}:p_1+p_2=p_3+p_4} \frac{c_{p_1}c_{p_2}\overline{c_{p_3}c_{p_4}}}{p_1p_2p_3p_4}$, and the estimate then follows from the triangle inequality and a standard upper bound sieve. We thank Ben Green for this observation.

are small unless one has a large Fourier coefficient $G_1(\xi)$ for some $\xi \in \mathbb{Z}/H\mathbb{Z}$. However, in contrast to the previous argument in which ξ was restricted to a small set Ξ_H (which, crucially, was independent of $\mathbf{n}$), one now has no control whatsoever on the location of ξ. As such, one would now need to control maximal averaged exponential sums such as

$$\frac{1}{X}\int_X^{2X} \sup_\alpha \left| \frac{1}{H}\sum_{j=1}^H \lambda(n+j)e(\alpha j) \right| \ dx, \tag{4.1}$$

which (as pointed out in [17]) is not currently covered by the existing literature (note carefully that the supremum in α is *inside* the integral over x). However, this appears to be the only significant obstacle to extending the results of this paper to the $k = 3$ case, and so it would certainly be of interest to obtain non-trivial estimates on (4.1).

For even higher values of k, one has to now control quartilinear and higher expressions in place of (3.19). Using the literature from higher order Fourier analysis (in particular the inverse theorem in [13], together with transference arguments from [7], [12], or [25]), one is now faced with the task of controlling sums even more complicated than (4.1), in which the linear phases $j \mapsto e(\alpha j)$ are now replaced by more general nilsequences of higher order (which one then has to take the supremum over, before performing the integral); this task can be viewed as a local version of the machinery in [5], [6]. Of course, since satisfactory control on (4.1) is not yet available, it is not feasible at present to control higher step analogues of (4.1) either. However, one can hope that if a technique is found to give good bounds on (4.1), it could also extend (in principle at least) to higher step sums.

References

[1] P. Billingsley, Ergodic theory and information. Reprint of the 1965 original. Robert E. Krieger Publishing Co., Huntington, N.Y., 1978.

[2] S. Chowla, The Riemann hypothesis and Hilbert's tenth problem, Gordon and Breach, New York, 1965.

[3] P. D. T. A. Elliott, *On the correlation of multiplicative functions*, Notas Soc. Mat. Chile, Notas de la Sociedad de Matemática de Chile, **11** (1992), 1–11.

[4] P. Erdős, *Some unsolved problems*, Michigan Math. J. **4** (1957), 299–300.

[5] N. Frantzikinatkis, B. Host, *Higher order Fourier analysis of multiplicative functions and applications*, preprint. `1403.0945`.

[6] N. Frantzikinakis, B. Host, *Asymptotics for multilinear averages of multiplicative functions*, preprint. `arXiv:1502.02646`.

[7] N. Frantzikinakis, B. Host, B. Kra, *Multiple recurrence and convergence for sequences related to the prime numbers*, J. Reine Angew. Math. **611** (2007), 131–144.

[8] J. Friedlander, H. Iwaniec, *The polynomial X^2+Y^4 captures its primes*, Ann. of Math. (2) **148** (1998), no. 3, 945–1040.

[9] J. Friedlander, H. Iwaniec, Opera de cribro. American Mathematical Society Colloquium Publications, 57. American Mathematical Society, Providence, RI, 2010.

[10] A. Granville, K. Soundararajan, *Decay of mean values of multiplicative functions*, Canad. J. Math. **55** (2003), no. 6, 1191–1230.

[11] B. Green, T. Tao, *Restriction theory of the Selberg sieve, with applications*, J. Théor. Nombres Bordeaux **18** (2006), no. 1, 147–182.

[12] B. Green, T. Tao, *Linear equations in primes*, Ann. of Math. (2) **171** (2010), no. 3, 1753–1850.

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

[14] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Amer. Stat. Assoc. **58** (1963), 13–30.

[15] H. Iwaniec, E. Kowalski, Analytic number theory. American Mathematical Society Colloquium Publications, 53. American Mathematical Society, Providence, RI, 2004.

[16] K. Matomäki, M. Radziwiłł, *Multiplicative functions in short intervals*, preprint. `arXiv:1501.04585`.

[17] K. Matomäki, M. Radziwiłł, T. Tao, *An averaged form of Chowla's conjecture*, preprint. `arXiv:1503.05121`.

[18] K. Matomäki, M. Radziwiłł, T. Tao, *Sign patterns of the Möbius and Liouville functions*, preprint. `arXiv:1509.01545`.

[19] H. Montgomery, Ten lectures on the interface between analytic number theory and harmonic analysis, volume 84 of CBMS Regional Conference Series in Mathematics. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 1994

[20] R. Moser, G. Tardos, *A constructive proof of the general Lovász local lemma*, J. ACM **57** (2010), no. 2, Art. 11, 15 pp.

[21] D.H.J. Polymath, `michaelnielsen.org/polymath1/index.php?title=The_Erd%C5%91s_discrepancy_problem`

[22] T. Tao, *The ergodic and combinatorial approaches to Szemerédi's theorem*, Additive combinatorics, 145193, CRM Proc. Lecture Notes, 43, Amer. Math. Soc., Providence, RI, 2007.

[23] T. Tao, *The Erdős discrepancy problem*, preprint. `arXiv:1509.05363`

[24] G. Tenenbaum, Introduction to analytic and probabilistic number theory, Translated from the second French edition (1995) by C. B. Thomas. Cambridge Studies in Advanced Mathematics, 46. Cambridge University Press, Cambridge, 1995.

[25] T. Wooley, T. Ziegler, *Multiple recurrence and convergence along the primes*, Amer. J. Math. **134** (2012), no. 6, 1705–1732.

Department of Mathematics, UCLA, 405 Hilgard Ave, Los Angeles CA 90095, USA

E-mail address: `tao@math.ucla.edu`