Correlations of multiplicative functions and factoring algorithms



Cédric Pilatte

St John's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2026

Acknowledgements

First and foremost, I wish to express my sincere gratitude to my advisors, James Maynard and Ben Green, for their invaluable guidance, enlightening discussions, and continuous support throughout my research. I will always remember with great fondness the time I spent under their supervision.

I would also like to thank Jon Keating and Oleksiy Klurman for agreeing to examine this thesis, and Louis-Pierre Arguin, Thomas Bloom, Ofir Gorodetsky, and Tom Sanders for their involvement during the earlier stages of my DPhil.

My time at the Mathematical Institute has been greatly enhanced by interactions with current and former students and postdoctoral researchers. In particular, I thank Akshat, Aleksandra, Aleksander, Alex, Alp, Arun, Benjamin, Catherine, Dan, Håvard, Jad, Jakub, Jared, Jeck, Jori, Joshua, Julia, Kate, Lambert, Lasse, Matt, Michael, Michał, Mira, Nathan, Noah, Sun Kai, Vishal, Yifan, and Zach.

I am deeply grateful to Joni Teräväinen, with whom I look forward to collaborating in Cambridge. I also wish to thank Régis de la Bretèche, Emmanuel Breuillard, Cécile Dartyge, Andrew Granville, Adam Harper, Harald Helfgott, Florent Jouve, Youness Lamzouri, Kaisa Matomäki, James Newton, Mayank Pandey, Sarah Peluse, Maksym Radziwiłł, Oliver Riordan, and Kannan Soundararajan for enriching discussions and, in several cases, for their hospitality. Furthermore, I am indebted to my professors in Mons and Paris, with special thanks to Timothy Gowers for supervising my research internship during my time at the ENS.

Beyond the department, I am grateful for the many good moments shared with my friends in Oxford, including Sebastian, Gabriel, Gustavo, Isaac, Anya, and Martin, as well as my friends outside the UK.

I wish to express my warmest thanks to my family, particularly Cécile, Marc and Nicole, my brothers Benoît and Nicolas, and my grandparents André and Christiane, for their unwavering encouragement. I also owe much to Christian, a true friend whose wise counsel has guided me for over a decade. Finally, I am deeply grateful to Madeleine, for her love, her support, and the joy she brings to my life.

My doctoral studies were supported by the Mathematical Institute, St John's College, a Saven Scholarship, a Jane Street Graduate Research Fellowship and a prize from the Vocatio Foundation.

Abstract

This thesis is devoted to the study of classical problems in analytic number theory, with a particular focus on multiplicative functions. A fundamental example is the Liouville function $\lambda : \mathbb{N} \to \{\pm 1\}$, which satisfies $\lambda(p) = -1$ for every prime p and is conjectured to behave pseudorandomly in many respects.

Our first result is motivated by Chowla's conjecture, which predicts that correlations between shifts of λ exhibit substantial cancellation. Improving on previous work of Helfgott and Radziwiłł, we prove that

$$\sum_{n \leqslant x} \frac{1}{n} \lambda(n)\lambda(n+1) \ll (\log x)^{1-c}$$

for some absolute constant $c > 0$.

Our second result concerns the Fourier uniformity conjecture, which predicts that λ has negligible correlation with linear phases in almost all short intervals. Improving a result of Walsh, we show that

$$\sum_{X \leqslant x < 2X} \sup_{\alpha \in \mathbb{R}/\mathbb{Z}} \left| \sum_{x \leqslant n < x+H} \lambda(n) e(n\alpha) \right| = o(HX)$$

as $X \to \infty$, in the regime $H = H(X) \geqslant \exp((\log X)^{2/5+\varepsilon})$.

Finally, we apply analytic number theory techniques to the study of quantum algorithms. Specifically, we prove a variant of Regev's conjecture on the distribution of short products of small primes in $(\mathbb{Z}/N\mathbb{Z})^\times$. This yields unconditional correctness and efficiency guarantees for a new family of quantum algorithms, including algorithms for factoring integers and computing discrete logarithms that use fewer gates than Shor's algorithm.

Statement of AI usage. I declare that my use of Generative AI in the preparation of this thesis complies with University, divisional, and departmental guidance on the responsible use of Generative AI. Generative AI tools were used only for identifying minor typographical and grammatical issues. The Generative AI tool used for this purpose was Gemini 3.1 Pro by Google. All suggested changes were subject to my critical review, and I retain full responsibility for the final content of the thesis.

Contents

1 Introduction **1**

1.1 Prime numbers and pseudorandomness 1

1.2 Chowla's conjecture 4

1.3 The Fourier uniformity conjecture 7

1.4 Quantum factoring algorithms 13

1.5 Notation 20

2 On Chowla's conjecture for two-point correlations **21**

2.1 Structure of the chapter 21

2.2 Main technical theorem and consequences 23

2.3 Reformulations of the main theorem 27

2.4 Approach via non-backtracking operators 30

2.5 High trace bound for the non-backtracking operator 38

2.6 Overview of the proof of the high trace bound 40

2.7 Definition of the set Y_L 43

2.8 Single and repeated primes 45

2.9 Cancellation from single primes 51

2.10 Predictable walks 57

2.11 Triangular systems of divisibility constraints 63

2.12 Unpredictable walks 67

2.13 Prohibited progressions and the combinatorial sieve 81

2.14 Applying the Matomäki-Radziwiłł-Tao theorem 92

3 On the Fourier uniformity conjecture **96**

3.1 Definitions and structure of the chapter 96

3.2 Construction of a tower of configurations 101

3.3 Analysis of the top configuration 104
3.4 Proof of the global structure theorem 112
3.5 Local patterns and clusters 119
3.6 Covering lemmas for clusters 122
3.7 Proof of the local structure theorem 125
3.8 Recovering Walsh's conditional theorem 134

4 Improved quantum algorithms **142**
4.1 Structure of the chapter 142
4.2 An illustrative special case 143
4.3 Proof of the existence of a short lattice basis 146
4.4 Consequences for quantum algorithms 160

A Non-backtracking operators **165**

B Frequency relations via Elliott's inequality **167**

C An adaptation of the Matomäki-Radziwiłł theorem **171**

References **190**

Chapter 1

Introduction

1.1 Prime numbers and pseudorandomness

Despite their fundamental role in number theory and their widespread real-world applications, prime numbers remain among the most mysterious objects in mathematics. Basic questions, such as whether there exist infinitely many twin primes, remain open.

One of the central aims of analytic number theory is to improve our understanding of the distribution of prime numbers. Although primes appear to be irregularly distributed, their overall behaviour can be described with considerable accuracy.

Around 1800, Gauss and Legendre observed, based on extensive numerical evidence, that the density of primes near a large number x is approximately $1/\log x$. In other words, the number of primes up to x is well approximated by the function

$$\mathrm{Li}(x) := \int_2^x \frac{dt}{\log t}.$$

Table 1.1 compares $\pi(x)$, the number of primes up to x, with the approximating function $\mathrm{Li}(x)$.[1]

x	10^3	10^6	10^{12}	10^{24}
$\pi(x)$	168	78498	37607912018	18435599767349200867866
$\mathrm{Li}(x)$	176.5	78626.5	37607950279.8	18435599767366347775143.1

Table 1.1: Approximation of $\pi(x)$ by $\mathrm{Li}(x)$.

After important partial results by Chebyshev [68, 69] and an influential eight-page paper by Riemann [52], this prediction was finally established in 1896, independently by Hadamard [22] and de la Vallée Poussin [11]. The resulting statement, known as the *Prime Number Theorem*, is a cornerstone of number theory. In simple terms,

[1] One should keep in mind that Gauss and Legendre only had tables of primes up to a few million.

it guarantees that the prime count $\pi(x)$ and its approximation $\mathrm{Li}(x)$ share an ever-increasing number of leading digits as $x \to \infty$.

The classical proof of the Prime Number Theorem relies on properties of the Riemann zeta function $\zeta(s)$. This is a meromorphic function on the complex plane, given for $\mathrm{Re}(s) > 1$ by the absolutely convergent series

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

According to the *Riemann Hypothesis*, one of the most famous conjectures in mathematics, all zeroes of $\zeta(s)$ in the strip $0 < \mathrm{Re}(s) < 1$ lie on the vertical line $\mathrm{Re}(s) = 1/2$.

The Riemann Hypothesis has far-reaching consequences for the distribution of primes. In particular, it would imply that about half of the digits of $\pi(x)$ and $\mathrm{Li}(x)$ must match, as can be observed in Table 1.1. More precisely, the Riemann Hypothesis is equivalent to the strong estimate[2]

$$\pi(x) = \mathrm{Li}(x) + O(x^{1/2} \log x).$$

This would be essentially best possible (up to logarithmic factors), since $\pi(x) - \mathrm{Li}(x)$ is known to exhibit fluctuations of order at least $x^{1/2} \frac{\log\log\log x}{\log x}$ infinitely often, in both the positive and negative directions [34].

Primes form a deterministic sequence, yet many of their features resemble those of a random sequence. One way to observe this is through the gaps between consecutive primes. By the Prime Number Theorem, the average gap between consecutive primes up to x is about $\log x$. However, the actual gaps between primes can be much smaller or much larger than $\log x$. To understand the distribution of these gaps, it is natural to compare the primes with a simple random model.

Suppose one selects $N \sim x/\log x$ real numbers independently and uniformly at random from the interval $[1, x]$. After ordering these points as $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_N$, one finds that the proportion of gaps $x_{i+1} - x_i$ greater than $A \log x$ converges to e^{-A} as $x \to \infty$, for every fixed constant $A > 0$. That is, in the limit, the normalised gaps in this random model follow an exponential distribution. It is conjectured that the gaps between primes exhibit the same behaviour, but such a precise statement remains out of reach of existing techniques.

Another useful probabilistic model, which we now describe, concerns the parity of the number of prime factors of integers. Let $\lambda : \mathbb{N} \to \{-1, +1\}$ be the Liouville function, defined by $\lambda(n) = 1$ if n has an even number of prime factors, counted with multiplicity, and $\lambda(n) = -1$ otherwise. While it might seem that λ encodes

[2]We refer the reader to Section 1.5 for the notation used in this thesis, including asymptotic notation.

limited arithmetic information, its statistical properties are in fact deeply tied to the distribution of prime numbers, via the identity

$$\sum_{n=1}^{\infty} \frac{\lambda(n)}{n^s} = \frac{\zeta(2s)}{\zeta(s)}$$

for $\mathrm{Re}(s) > 1$.

The *Liouville pseudorandomness principle* is a heuristic suggesting that λ should behave statistically like a sequence of independent random variables taking the values -1 and $+1$ with equal probability. In particular, a random integer $n \in \{1, \dots, x\}$ should have an even or odd number of prime factors with probability approaching $1/2$ as $x \to \infty$, so that

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

This is indeed the case; in fact, the estimate (1.1) is equivalent to the Prime Number Theorem by elementary arguments. Moreover, the Riemann Hypothesis is equivalent to the square-root cancellation bound

$$\sum_{n \leqslant x} \lambda(n) \ll_{\varepsilon} x^{1/2+\varepsilon}. \tag{1.2}$$

The estimate (1.2) is consistent with the Liouville pseudorandomness principle, since a simple random walk of length x with steps ± 1 grows like $O(x^{1/2+\varepsilon})$ with high probability. We note, however, that the fine-scale behaviour of these Liouville partial sums is expected to deviate slightly from that of a simple random walk [43]. A plot of the function $x \mapsto \sum_{n \leqslant x} \lambda(n)$ is given in Fig. 1.1. The oscillating pattern can be explained by the existence of a formula for these partial sums in terms of the zeroes of the Riemann zeta function.

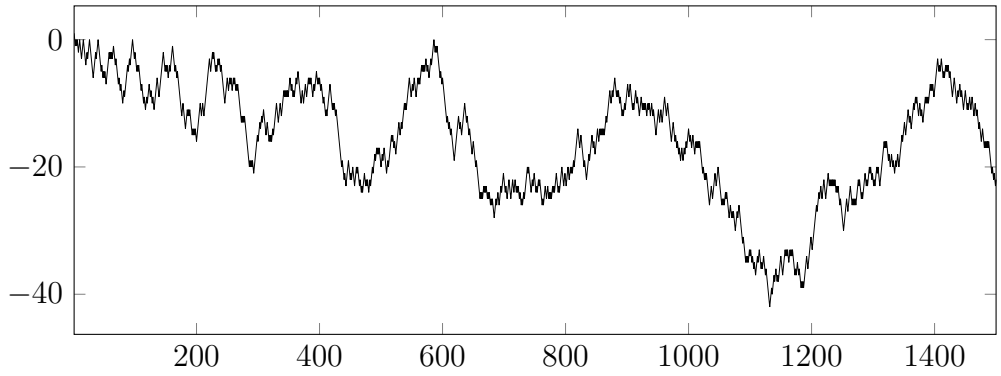


Figure 1.1: Plot of the function $x \mapsto \sum_{n \leqslant x} \lambda(n)$.

A substantial portion of this thesis is devoted to the study of *correlations* of the Liouville function, which are also predicted to exhibit significant cancellation.

Chapter 2 is dedicated to two-point correlations. Roughly speaking, the objective is to establish that, for a randomly chosen integer $n \in \{1, \dots, x\}$, the values $\lambda(n)$ and $\lambda(n+1)$ are asymptotically independent. In Chapter 3, we examine a different aspect of the pseudorandomness of λ, namely its Fourier uniformity in short intervals, which is closely related to the study of higher-order correlations.

Finally, Chapter 4 presents applications of analytic number theory techniques to the study of quantum algorithms. The problems addressed in this chapter are intimately connected to the Generalised Riemann Hypothesis, which extends Riemann's original conjecture to Dirichlet L-functions. Although current knowledge about the zeroes of these L-functions remains limited, it turns out to be just sufficient for our purposes. In particular, we establish the correctness and efficiency guarantees of the relevant quantum algorithms unconditionally.

1.2 Chowla's conjecture

1.2.1 Background and main results

A well-known conjecture of Chowla [10] asserts that, for any $k \geqslant 1$ and distinct integers $h_1, \dots, h_k$, one has

$$\frac{1}{x} \sum_{n \leqslant x} \lambda(n+h_1)\lambda(n+h_2) \cdots \lambda(n+h_k) = o_{x\to\infty}(1). \tag{1.3}$$

This can be regarded as a multiplicative analogue of the Hardy-Littlewood prime k-tuple conjecture, which predicts an asymptotic formula for correlations of the von Mangoldt function Λ. Chowla's conjecture is subject to the *parity problem*, a major obstacle in analytic number theory (see [18, Section 16.4] for more details). It remains wide open for $k \geqslant 2$.

Yet, in recent years, remarkable progress has been made on weaker variants of Chowla's conjecture.

In 2015, Matomäki, Radziwiłł and Tao proved that (1.3) holds *on average* over $h_1, \dots, h_k$, for every fixed $k \geqslant 2$ [38]. A crucial ingredient in their proof was the groundbreaking work of Matomäki and Radziwiłł [35] on sums of multiplicative functions over short intervals.

One year later, Tao proved a *logarithmic* version of Chowla's conjecture for $k = 2$ [61]. This means that the regular average $\frac{1}{x}\sum_{n\leqslant x} f(n)$ is replaced by the logarithmic average $\frac{1}{\log x}\sum_{n\leqslant x} \frac{1}{n} f(n)$. Fixing $(h_0, h_1) = (0, 1)$ for simplicity, Tao's result thus reads

$$\frac{1}{\log x} \sum_{n \leqslant x} \frac{1}{n} \lambda(n)\lambda(n+1) = o_{x\to\infty}(1). \tag{1.4}$$

Tao's proof [61], which used a novel *entropy decrement argument*, was a key step in his resolution of the Erdős discrepancy problem [60]. From his paper [61], it is possible (see [27]) to extract the explicit bound

$$\sum_{n\leqslant x}\frac{1}{n}\lambda(n)\lambda(n+1)\ll\frac{\log x}{(\log\log\log\log x)^{1/5}}. \tag{1.5}$$

The logarithmic version of Chowla's conjecture (1.3) was later proved for all *odd* $k\geqslant 3$ by Tao and Teräväinen [65]. The two authors gave a different proof of that result in [63]. For *even* $k\geqslant 4$, the logarithmically averaged Chowla conjecture is still open. The methods of their paper [63] can be used to obtain the following quantitative refinement of (1.5): for some small absolute constant $c>0$,

$$\sum_{n\leqslant x}\frac{1}{n}\lambda(n)\lambda(n+1)\ll\frac{\log x}{(\log\log\log x)^{c}}. \tag{1.6}$$

In 2021, Helfgott and Radziwiłł [26] proved the substantial quantitative improvement

$$\sum_{n\leqslant x}\frac{1}{n}\lambda(n)\lambda(n+1)\ll\frac{\log x}{(\log\log x)^{1/2}}. \tag{1.7}$$

They used a very different combinatorial approach, studying the eigenvalues of a certain weighted graph defined in terms of divisibility by small primes. A high-level exposition of their proof is given by Helfgott [25].

In Chapter 2, we improve the approach of Helfgott and Radziwiłł [26] to prove the following bound.

Theorem 1.2.1 (Logarithmic two-point Chowla correlations)**.** *For some absolute constant* $c>0$*,*

$$\sum_{n\leqslant x}\frac{1}{n}\lambda(n)\lambda(n+1)\ll(\log x)^{1-c}.$$

It appears that saving a fixed power of the logarithm is the best that is achievable with current techniques. Ultimately, our proof relies on the work of Matomäki and Radziwiłł [35] on multiplicative functions in short intervals, where the current state of the art only allows one to save a small power of $\log x$. The exploitation of multiplicativity using an idea of Tao [61] also separately appears to limit our saving to a small power of $\log x$, because a typical integer has $O(\log x)$ divisors.

The methods of this thesis should generalise to a wider class of multiplicative functions through appropriate modifications. The complete multiplicativity of λ is only used in Proposition 2.3.1 and in the proof that Theorem 2.2.4 implies Theorem 1.2.1. The only other property of λ we use is that it is 1-bounded (for Sections 2.2 and 2.3), but a weaker ℓ^p bound would suffice.

Our proof also yields an improved bound for the *unweighted* two-point correlations (i.e. without logarithmic averaging) at *almost all scales*; see Remark 2.2.8.

The present work has recently been used by Tao and Teräväinen [66] to resolve problems of independent interest, including the proof of the irrationality of the series

$$\sum_{n=1}^{\infty} \frac{\omega(n)}{2^n},$$

where $\omega(n)$ is the number of distinct prime factors of n. This confirms a conjecture of Erdős [15].

1.2.2 Proof outline

In this section, we give a very short description of the overall strategy. Fuller explanations are given along the way, at various points in Chapter 2.

Using the multiplicativity of λ, Tao [61] showed that the problem of bounding $\sum_{n\leqslant x} \frac{1}{n}\lambda(n)\lambda(n+1)$ reduces to bounding

$$\mathbb{E}_{p\in P} \sum_{n\leqslant x} \lambda(n)\lambda(n+p)\left(\mathbf{1}_{p|n} - \frac{1}{p}\right),$$

where $P \subset [1, \exp(\sqrt{\log x})]$ is a set of primes.

Helfgott and Radziwiłł [26] interpreted the above expression as the matrix product $\boldsymbol{\lambda}^\top A\boldsymbol{\lambda}$ where $\boldsymbol{\lambda} := (\lambda(1), \dots, \lambda(x))^\top$ and A is the matrix with entries

$$A_{mn} := \begin{cases} \mathbf{1}_{p|n} - \frac{1}{p} & \text{if } |m-n| = p \in P, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, it is sufficient to bound the eigenvalues of the matrix A, or the eigenvalues of its restriction $A|_X$ to some very dense subset $X \subset \{1, \dots, x\}$. Using a high trace method, Helfgott and Radziwiłł [26] managed to obtain the bound $\left(\sum_{p\in P} 1/p\right)^{1/2+o(1)}$ for the largest eigenvalue of such a matrix, which is essentially the best possible. Since $\sum_{p\in P} 1/p \ll \log\log x$, this approach cannot yield a saving better than a power of $\log\log x$ over the trivial bound for two-point Chowla correlations.

In our new approach, we replace the average over primes $p \in P$ with an average over integers $d = p_1 \cdots p_k$ that are products of k primes, where $k \asymp \log\log x$. By Tao's argument, we need to bound

$$\mathbb{E}_{d\in D_k} \sum_{n\leqslant x} \lambda(n)\lambda(n+d) \prod_{p|d}\left(\mathbf{1}_{p|n} - \frac{1}{p}\right)$$

where D_k is a set of integers with k prime factors. Following the strategy of Helfgott and Radziwiłł [26], it is sufficient to bound the eigenvalues of the matrix $\widetilde{A}|_X$ where $\widetilde{A}$ is the matrix defined by

$$\widetilde{A}_{mn} := \begin{cases} \prod_{p|d}\left(\mathbf{1}_{p|n} - \frac{1}{p}\right) & \text{if } |m-n| = d \in D_k, \\ 0 & \text{otherwise;} \end{cases}$$

and X is a large subset of $\{1, \dots, x\}$. We prove that all eigenvalues of $\widetilde{A}|_X$ are bounded by

$$\Big(\sum_{d \in D_k} \frac{1}{d}\Big)^{1/2+o(1)}$$

in absolute value. Since $k \gg \log\log x$, we can choose D_k so that $\sum_{d\in D_k} 1/d \gg (\log x)^c$, which produces the exponential improvement in Theorem 1.2.1.

Unfortunately, working with products of multiple primes rather than single primes introduces new difficulties throughout the argument. It is handling all of these new difficulties which is the key new contribution of our work. We are forced to rework and generalise all the arguments of [26] with the result that Chapter 2 is essentially self-contained.

One particular new difficulty is in Section 2.12 where we wish to bound the number of solutions to systems of divisibility constraints. In the prior work this was a linear system, and so could be bounded by a simple lattice point argument. In our situation this now becomes a polynomial system, and to handle it we require a much more involved argument based on the structure of what we call *unpredictable words*. Another novel feature of our work is the use of *non-backtracking operators*; these considerably simplify certain technical aspects of the proof.

1.3 The Fourier uniformity conjecture

1.3.1 Background and main results

We now turn to the behaviour of the Liouville function λ in short intervals.

The celebrated Matomäki-Radziwiłł theorem [35] shows that λ exhibits cancellation in almost all short intervals. More precisely, for any function $H = H(X)$ tending to infinity with X, we have

$$\sum_{X \leqslant x < 2X} \Big| \sum_{x \leqslant n < x+H} \lambda(n) \Big| = o(HX)$$

as $X \to \infty$. Remarkably, this estimate holds for arbitrarily slowly growing H.

For applications to Chowla's conjecture and related problems (see Section 1.3.2), a stronger form of pseudorandomness is required: the *Fourier uniformity* of the Liouville function in short intervals.

Conjecture 1.3.1 (Fourier uniformity conjecture)**.** *For any function $H = H(X)$ tending to infinity with X, we have*

$$\sum_{X \leqslant x < 2X} \sup_{\alpha \in \mathbb{R}/\mathbb{Z}} \left| \sum_{x \leqslant n < x+H} \lambda(n) e(n\alpha) \right| = o(HX) \tag{1.8}$$

as $X \to \infty$.

Informally, Conjecture 1.3.1 asserts that λ has negligible correlations with linear phases in almost all short intervals. Conjecture 1.3.1 appears to have first been suggested by Matomäki, Radziwiłł and Tao [38, p.3].

A crucial feature of Conjecture 1.3.1 is that the frequency α is permitted to depend on the interval $[x, x+H)$ in an arbitrary way. A weaker version of (1.8), in which the supremum over α is taken outside the sum over x, was obtained by Matomäki, Radziwiłł, and Tao [38] shortly after the original Matomäki-Radziwiłł theorem.

As one might expect, Conjecture 1.3.1 becomes more difficult to establish for shorter interval lengths H. The conjecture was initially proved for $H \geqslant X^{\varepsilon}$ by Matomäki, Radziwiłł, and Tao [39], with a simpler proof later given by Walsh [71]. This threshold was lowered to $H \geqslant \exp((\log X)^{5/8+\varepsilon})$ by Matomäki, Radziwiłł, Tao, Teräväinen, and Ziegler [40], and further improved to $H \geqslant \exp((\log X)^{1/2+\varepsilon})$ by Walsh [72].

As explained in [40] and [72], a severe bottleneck emerges at the scale

$$H \approx \exp((\log X)^{1/2}). \tag{1.9}$$

Indeed, all known approaches require an iterative argument consisting of k steps, where $k \asymp \frac{\log X}{\log H}$ is the diameter of a specific underlying graph. The most natural ways to perform such an iteration incur $k!$-type losses, which become overwhelming once H drops below the threshold (1.9).

Assuming the Generalised Riemann Hypothesis, Walsh [73] was able to bypass this obstacle, proving Conjecture 1.3.1 for intervals of length $H \geqslant (\log X)^{\psi(X)}$, where $\psi(X)$ tends to infinity arbitrarily slowly.

In Chapter 3, we overcome this barrier unconditionally. Our main result is the following.

Theorem 1.3.2. *Let $\varepsilon > 0$. For any function $H = H(X)$ with $H \geqslant \exp((\log X)^{2/5+\varepsilon})$, we have*

$$\sum_{X \leqslant x < 2X} \sup_{\alpha \in \mathbb{R}/\mathbb{Z}} \left| \sum_{x \leqslant n < x+H} \lambda(n) e(n\alpha) \right| = o(HX)$$

as $X \to \infty$.

In fact, we prove a more general version of this result for arbitrary non-pretentious multiplicative functions, with explicit quantitative savings on the order of $\sqrt{\log\log X}$; see Theorem 1.3.3 for the full technical statement.

1.3.2 Relation to Chowla's conjecture

The primary motivation for studying the Fourier uniformity conjecture is its close connection to Chowla's conjecture (see Section 1.2), and in particular to the logarithmic version of that conjecture.

Recall that the *logarithmic Chowla conjecture* asserts that, for any distinct positive integers $h_1, \dots, h_k$,

$$\sum_{n \leqslant X} \frac{\lambda(n+h_1)\cdots\lambda(n+h_k)}{n} = o(\log X) \tag{1.10}$$

as $X \to \infty$. It is known for $k = 2$ and for all odd $k \geqslant 3$, but remains open for even $k \geqslant 4$.

The proof of (1.10) for two-point correlations fundamentally relies on the Matomäki-Radziwiłł-Tao theorem [38] on short exponential sums of the Liouville function. As it turns out, the odd order cases are amenable to a parity trick that avoids such deep number-theoretic inputs [63, 65].

In [62], Tao proved the equivalence of three different statements:

(i) the logarithmic Chowla conjecture (for all $k \geqslant 1$);

(ii) the logarithmic Sarnak conjecture;

(iii) the logarithmic higher-order Fourier uniformity conjecture.

Here, the logarithmic Sarnak conjecture is the assertion that λ has negligible logarithmic correlation with any deterministic sequence.[3] The *higher-order* Fourier uniformity conjecture is a natural generalisation of Conjecture 1.3.1 in which the linear phases $e(n\alpha)$ are replaced by nilsequences of a given step, and (iii) is a logarithmically weighted version of that statement. The deduction of (i) from (iii) relies on the *entropy decrement argument*, first introduced by Tao in [61].

Following this equivalence, a potential path towards the general logarithmic Chowla conjecture would be to prove the higher-order Fourier uniformity conjecture. Crucially, it is not necessary to consider arbitrarily slowly growing H: it was shown in [40, Proposition 1.7] that establishing the conjecture for $H \geqslant (\log X)^{\varepsilon}$ for every $\varepsilon > 0$ is sufficient to deduce the logarithmic Chowla conjecture. Walsh's result under GRH [73]

[3] That is, any sequence arising from a topological dynamical system of zero entropy; see [55] or [62] for the relevant definitions.

comes remarkably close to this threshold, as it proves the Fourier uniformity conjecture for $H \geqslant (\log X)^{\psi(X)}$ with $\psi(X)$ tending to infinity arbitrarily slowly. Lowering this bound further, even conditionally, remains a significant open problem.

An alternative approach towards the logarithmic Chowla conjecture proceeds via expander graph techniques. As explained in Section 1.2, this method has been successfully applied to the $k = 2$ case of (1.10), yielding substantially stronger quantitative bounds than the entropy method. Furthermore, the required number-theoretic input is less demanding: the proof only uses the Matomäki-Radziwiłł-Tao theorem for intervals of length $H = \exp((\log X)^{1/2-c})$, for an arbitrarily small constant $c > 0$.

The expander graph approach relies on trace methods that currently lack suitable generalisations to hypergraphs. Nevertheless, if a higher-order analogue of the two-point decoupling estimate could be established,[4] such a result would reduce the logarithmic Chowla conjecture to the higher-order Fourier uniformity conjecture at the $H = \exp((\log X)^{1/2-c})$ threshold. Prior to the present work, the best unconditional bounds for the Fourier uniformity conjecture, due to Walsh [72], were limited to the range $H \geqslant \exp((\log X)^{1/2+\varepsilon})$. By lowering this exponent to $2/5$, Theorem 1.3.2 would now provide the necessary number-theoretic ingredient, at least for linear phases.

It would then remain to generalise Theorem 1.3.2 to arbitrary nilsequences. However, the main conceptual difficulties are generally expected to manifest already in the linear setting. Although technically intricate, the task of transferring Fourier uniformity results from linear phases to nilsequences has been successfully accomplished in previous works. In particular, Matomäki, Radziwiłł, Tao, Teräväinen, and Ziegler [40] have proved the higher-order Fourier uniformity conjecture in the range $H \geqslant X^{\varepsilon}$ for any $\varepsilon > 0$. See also [37, 41] for further results on higher-order uniformity of arithmetic functions in short intervals.

1.3.3 Proof outline and comparison with previous work

The general framework of our proof follows that of all previous approaches to the Fourier uniformity conjecture. Assuming that Theorem 1.3.2 fails, there are many short intervals $[x, x + H)$ on which λ has a large correlation with some linear phase $e(n\alpha_x)$, where α_x depends on x in an uncontrolled manner. Still, by exploiting the multiplicativity of λ at small primes, these frequencies α_x can be shown to exhibit certain weak dependencies. The heart of the proof is to show that, together, these dependencies are rigid enough to force a global structure on the α_x. Once this is established, the Matomäki-Radziwiłł theorem can be applied to obtain a contradiction.

[4] See [26, Section 9.5] for a more extensive discussion.

The central task is to analyse the following setup. Let $A \subset [Y, 2Y]$ be a well-spaced set of size $|A| \asymp Y$, where $Y := X/H$, with a frequency α_x assigned to each $x \in A$. Let $\mathcal{P}$ be a set of primes $p \asymp P$, where the scale P is a small power of H. In this setting, the weak dependencies amount to the existence of $\gg Y|\mathcal{P}|^2$ quadruples $(x, y, p, q) \in A^2 \times \mathcal{P}^2$ satisfying $px \approx qy$ and $q\alpha_x \approx p\alpha_y \pmod 1$.

The graph formed by interpreting these dependencies as edges on the vertex set A is highly sparse, with $\asymp Y$ vertices and an average degree of $\asymp |\mathcal{P}|^2$. Propagating this local information to extract a global structure therefore requires an iterative argument with $\asymp \frac{\log Y}{\log P}$ steps. As in [71, 72, 73], our iterative procedure involves the construction of auxiliary frequencies that are 'lifts' of the original α_x and satisfy a similar system of relations. In previous unconditional approaches, the density of relations satisfied by the auxiliary frequencies decays by at least a constant factor at every step; such accumulated losses become fatal below the $H \approx \exp((\log X)^{1/2})$ barrier.

We take the conditional approach of Walsh [73] as our starting point. Working under GRH, Walsh introduced a 'single-prime lifting' argument in which the auxiliary frequencies retain almost all of the relations satisfied by the original α_x. This is achieved via a *local structure theorem* that partitions A into a bounded number of components, all but one of which can be successfully lifted. To control this partition, he employs an expansion lemma based on RH, which ensures that no two subsets of A share significantly more edges than the random model predicts. This expansion property forces the exceptional component to be negligibly small. After iterating this procedure $\asymp \frac{\log Y}{\log P}$ times, GRH is applied once more to extract the required global structure from the relations between the constructed frequencies.

Our main new ingredient is a stronger local structure theorem (Theorem 3.1.8) that is robust enough to allow for an unconditional iterative argument. It differs from Walsh's local structure theorem in two key respects:

- It is *quantitative*, yielding a decomposition of A where the number of parts can be chosen to be almost as large as $\log P$, rather than a fixed constant.
- It is *relative*, requiring only that the number of relations be large relative to the size of A (rather than an absolute density), which allows the theorem to apply even to very sparse sets.

Inspired by Walsh's methods, the proof of Theorem 3.1.8 adopts a different viewpoint centred on the quantitative analysis of substructures called 'clusters', and requires new tools such as an *ad hoc* covering lemma for these clusters.

Iterating this local structure theorem allows us to construct the auxiliary frequencies while working with potentially very sparse sets throughout. To analyse the resulting construction, we substitute GRH with an unconditional expansion estimate derived

from Vinogradov-Korobov type bounds for Dirichlet L-functions.[5] This expansion estimate is significantly weaker than its conditional counterpart, yielding non-trivial information only for very sparse sets. Nonetheless, our combinatorial framework is robust enough that this input suffices to establish the required global structure.

This conclusion is the content of our *global structure theorem* (Theorem 3.1.9), which roughly states that, on a large subset $A' \subset A$, the frequencies α_x satisfy an approximate formula of the form $\alpha_x \approx a_0/q_0 + T_0/x \pmod 1$ for some fixed a_0, q_0 and T_0 of controlled size. However, a strong version of the Matomäki-Radziwiłł theorem shows that the Liouville function can only correlate with such linear phases on a set of density at most H^{-c} for some small $c > 0$. By contrast, our Vinogradov-Korobov expansion estimate guarantees that A' has density at least $\exp(-(\log X)^{1-3\theta/2+o(1)})$ when $H = \exp((\log X)^\theta)$. This yields a contradiction for any $\theta > 2/5$.

The above argument requires a variant of the Matomäki-Radziwiłł theorem that allows for large values of q_0 while retaining a power-saving bound for the exceptional set. As the precise quantitative statement we need does not appear to be recorded in the literature, we provide a proof in Chapter C, drawing entirely on ideas from [35, 36, 40].

These methods naturally extend to arbitrary non-pretentious multiplicative functions, yielding the following general result.

Theorem 1.3.3 (Fourier uniformity estimate)**.** *Let $f : \mathbb{N} \to \mathbb{C}$ be a 1-bounded multiplicative function.*

Let $0 < \varepsilon < 1/5$. Let X be a parameter chosen sufficiently large in terms of ε, and let $H := \exp((\log X)^{2/5+\varepsilon})$.

Assume that

$$\sum_{X \leqslant x < 2X} \sup_{\alpha \in \mathbb{R}/\mathbb{Z}} \left| \sum_{x \leqslant n < x+H} f(n) e(n\alpha) \right| \geqslant \delta H X,$$

where $\delta \geqslant C_0(\log\log X)^{-1/2}$ for some sufficiently large absolute constant $C_0 > 0$.

Then, f is pretentious in the sense that

$$\sum_{p \leqslant X} \frac{1 - \mathrm{Re}(f(p)\chi(p)p^{it})}{p} \ll \log \delta^{-1}$$

for some Dirichlet character χ of conductor $O(\delta^{-O(1)})$ and some real number t with $|t| \leqslant X^2/H^{2-o(1)}$.

For simplicity, we have stated Theorem 1.3.3 for a specific choice of H near the threshold $\exp((\log X)^{2/5})$; the regime of larger H is treated in [72]. The deduction of

[5] See [30, Section 8.5, p.216-228] for an exposition of the Vinogradov-Korobov bound for the Riemann zeta function.

Theorem 1.3.2 from Theorem 1.3.3 follows from standard Vinogradov-Korobov type non-pretentiousness estimates for the Liouville function.

We note that, unlike previous results, Theorem 1.3.3 is not restricted to the dense regime $\delta \asymp 1$, as it allows for quantitative savings of order $\sqrt{\log\log X}$. This rate of decay is the best possible for an approach based on Elliott's inequality, which serves as the starting point for our approach and most earlier work.

Our quantitative methods allow us to recover a version of Walsh's Fourier uniformity estimate under GRH [73] with explicit savings; we do so in Section 3.8. The proof is self-contained; no prior familiarity with [73] is needed to follow the exposition.

Finally, we remark that the exponent $2/5$ in Theorem 1.3.3 represents the natural limit of our unconditional methods. For H below this threshold, we do not know how to rule out the possibility that all connected components in the graph of relations are extremely sparse (say, of density less than $1/H$). In such a scenario, one cannot hope to establish a global structure without extracting more information from the Liouville correlations than the current Elliott inequality argument provides.

1.4 Quantum factoring algorithms

1.4.1 Background and main results

Public key cryptography has become a crucial element of our global digital communication infrastructure. Notable examples include the Diffie-Hellman key exchange [13] and the RSA (Rivest-Shamir-Adleman) cryptosystem [53], which rely on the difficulty of finding discrete logarithms and factoring large numbers, respectively.

However, in 1994, Peter Shor [57] developed an algorithm capable of efficiently solving these problems using a quantum computer.

Theorem (Shor[6]). *There is a quantum circuit having $O(n^2 \log n)$ quantum gates and $O(n \log n)$ qubits with the following property. There is a classical randomised polynomial-time algorithm that solves the factoring problem*

$$\textbf{Input} : \textit{a composite integer } N \leqslant 2^n$$
$$\textbf{Output} : \textit{a non-trivial divisor of } N$$

using $O(1)$ calls to this quantum circuit, and succeeds with probability $\asymp 1$.

[6]The version of Shor's algorithm stated here incorporates some small improvements from [58] (bounded number of calls) and [24] (fast integer multiplication). The number of qubits in Shor's algorithm can be brought down to $O(n)$, though this typically requires a larger number of quantum gates (see e.g. [20, 32]).

Shor's original article [57] also includes a similar algorithm to solve the discrete logarithm problem. This advancement prompted the development of post-quantum cryptography, such as lattice-based cryptography, to ensure the security of our communications in the face of potential quantum computing breakthroughs [6].

Recently, Regev [50] devised a multidimensional variant of Shor's factoring algorithm that reduces the size of the quantum circuit, i.e. the number of quantum gates, to $O(n^{3/2}\log n)$. A distinctive feature of Regev's algorithm is that the quantum circuit must be called $O(\sqrt{n})$ times, rather than a constant number of times for Shor's algorithm. This is not considered to be a serious drawback as the various complexity parameters of the quantum circuit are far more relevant metrics in quantum computing.

However, this remarkable work of Regev initially came with two main limitations.

First, the original version of Regev's algorithm requires $O(n^{3/2})$ qubits – many more than Shor's algorithm. The reason for this ultimately lies in the difficulty of performing classical computations on quantum computers in a space-efficient manner, due to the need for any quantum computation to be *reversible*. In a subsequent paper, Ragavan and Vaikuntanathan [49] improved Regev's algorithm to use only $O(n\log n)$ qubits, while maintaining the circuit size at $O(n^{3/2}\log n)$ quantum gates. This work, inspired by ideas of Kaliski [33], essentially matches the space cost of Shor's algorithm.

The second issue, which we address in this thesis, is that Regev's algorithm [50] does not have theoretical guarantees, unlike Shor's algorithm. The correctness of Regev's algorithm is based on an *ad hoc* number-theoretic conjecture which we describe below. This unproven assumption cannot be avoided as it lies at the core of Regev's improvement on the circuit size. The variant of Ragavan and Vaikuntanathan [49] also crucially relies on this conjecture.

In Chapter 4, we prove a version of Regev's conjecture. This allows us to unconditionally prove the correctness of (slightly modified versions of) the algorithms of Regev [50] and Ragavan and Vaikuntanathan [49]. More precisely, we obtain the following algorithmic result.

Theorem 1.4.1. *There is a quantum circuit having $O(n^{3/2}\log^3 n)$ quantum gates and $O(n\log^3 n)$ qubits with the following property. There is a classical randomised polynomial-time algorithm that solves the factoring problem*

$$\textbf{Input} : \textit{a composite integer } N \leqslant 2^n$$
$$\textbf{Output} : \textit{a non-trivial divisor of } N$$

using $O(\sqrt{n})$ calls to this quantum circuit, and succeeds with probability $\asymp 1$.

This unconditionally establishes the results in [49, 50] up to logarithmic factors.

Regev's algorithm was adapted by Ekerå and Gärtner [14] to the discrete logarithm problem. Their paper [14] also uses the space-saving arithmetic of Ragavan and Vaikuntanathan [49] to obtain a quantum circuit with $O(n^{3/2} \log n)$ gates and $O(n \log n)$ qubits for computing discrete logarithms.

Once more, the correctness of the algorithm by Ekerå and Gärtner [14] relies on an unproven hypothesis, which can be viewed as a stronger form of Regev's conjecture. Our methods also apply to this stronger statement (again, with minor technical adjustments). Thus, we get an analogue of Theorem 1.4.1 for the discrete logarithm problem, i.e. an unconditional proof of the results in [14] up to logarithmic factors.

Theorem 1.4.2. *There is a quantum circuit having $O(n^{3/2} \log^3 n)$ quantum gates and $O(n \log^3 n)$ qubits with the following property. There is a classical randomised polynomial-time algorithm that solves the discrete logarithm problem*

> **Input** : *an integer $N \leqslant 2^n$ and elements $g, y \in (\mathbb{Z}/N\mathbb{Z})^\times$ such that $y \in \langle g \rangle$*
> **Output** : *an integer x such that $g^x \equiv y \pmod{N}$*

using $O(\sqrt{n})$ calls to this quantum circuit, and succeeds with probability $\asymp 1$.

Moreover, the slight technical modifications that we need to introduce to make all these quantum algorithms unconditional are compatible with the error-correction results of [14, 49]. Finally, our results extend to further variants of Regev's algorithm, such as computing discrete logarithms or multiplicative orders modulo N for several elements simultaneously (see [14]).

Remark 1.4.3. By being slightly more careful in the space usage of our quantum algorithms, it is possible to reduce the number of qubits to $O(n \log^2 n)$ for Theorems 1.4.1 and 1.4.2. We will not show this here in order to keep the quantum computing prerequisites to a minimum.

1.4.2 An overview of Regev's algorithm

Let us start by recalling the basic idea behind Shor's factoring algorithm [57].

Let N be the integer to be factored, which can be assumed to be odd and with at least two distinct prime factors. Shor proved that there is an efficient quantum algorithm to find the multiplicative order of any invertible element modulo N. With this in hand, factoring N is straightforward. The first step is to pick an integer $1 < a < N$ uniformly at random. If $\gcd(a, N) > 1$, the factoring task is trivial. Otherwise, with probability at least $1/2$, the order of a modulo N, denoted r, is even and satisfies $a^{r/2} \not\equiv \pm 1$ (i.e. $a^{r/2}$ is a non-trivial square root of 1 modulo N). Whenever this is the case, we get a non-trivial divisor of N, namely $\gcd(N, a^{r/2} - 1)$.

The reason that Shor's algorithm uses $O(n^2 \log n)$ quantum gates, where $n := \lceil \log_2 N \rceil$, comes from the part of the quantum circuit that performs modular exponentiation. Consider the task of computing a power $a^M \pmod N$ on a classical computer, where $1 < a < N-1$ and $M \leqslant N$. This computation can be performed efficiently using the well-known square-and-multiply method, which involves $O(\log M)$ multiplications of two integers modulo N. Since two n-bit integers can be multiplied in time $O(n \log n)$ by [24], the complexity of this modular exponentiation problem is thus $O(n^2 \log n)$.[7] For Shor's algorithm, a similar modular exponentiation needs to be performed quantumly, which requires $O(n^2 \log n)$ quantum gates.

Regev's improvement of Shor's algorithm is made possible by combining two key ideas.

The first idea in Regev's algorithm is to work in *higher-dimensional* space and replace the random parameter a by several integers $b_1, \ldots, b_d$ (chosen in a specific way, as we explain below). Eventually, the optimal choice of dimension turns out to be $d \asymp \sqrt{\log N}$. Similarly to Shor's algorithm, the problem of factoring N easily reduces to the task of finding a vector $(e_1, \ldots, e_d) \in \mathbb{Z}^d$ such that $\prod_{i=1}^d b_i^{e_i}$ is a non-trivial square root of 1 modulo N.

Using additional tools such as the LLL lattice reduction algorithm, Regev [50] generalised Shor's quantum algorithm to efficiently find all such vectors $(e_1, \ldots, e_d)$ in a ball of radius $N^{O(1/d)}$ centred at the origin.

Regev's algorithm only succeeds if there indeed exists a vector $v = (e_1, \ldots, e_d) \in \mathbb{Z}^d$ such that

(i) $\|v\|_2 \leqslant N^{O(1/d)}$, and

(ii) $\prod_{i=1}^d b_i^{e_i}$ is a non-trivial square root of 1 modulo N.

If the parameters $b_1, \ldots, b_d$ are "sufficiently multiplicatively independent" modulo N, one might heuristically expect the existence of a vector $v = (e_1, \ldots, e_d)$ satisfying (i) and (ii). This is, roughly speaking, what Regev needs to assume. To state his conjecture properly, it remains to specify how the parameters $b_1, \ldots, b_d$ are chosen. This is a crucial point, as the idea of working in dimension d does not, on its own, offer any advantage over Shor's algorithm.

The second key insight of Regev is to choose $b_1, \ldots, b_d$ to be *very small* compared to N. We mentioned that the most costly part of Shor's quantum circuit lies in the modular exponentiation step. Similarly, the number of gates of Regev's circuit is

[7] Note that, even if a is a small integer, say $a = 2$, this procedure still takes time $O(n^2 \log n)$, because most multiplications will involve n-bit integers.

dominated by the cost of computing an expression of the form

$$\prod_{i=1}^{d} b_i^{M_i} \pmod{N} \tag{1.11}$$

in the quantum setting, where the exponents $M_1, \ldots, M_d$ are $\leqslant N^{O(1/d)}$. Regev observed that (1.11) can be computed classically in time $O(n^{3/2} \log n)$ if $|b_i| \leqslant (\log N)^{O(1)}$ for all i (for $d \asymp \sqrt{n} \asymp \sqrt{\log N}$). This can be achieved by cleverly ordering the intermediate multiplications so that most of them involve integers with much fewer than n bits. This classical procedure can then be turned into a quantum version that uses $O(n^{3/2} \log n)$ gates.

Shor's algorithm corresponds to the $d = 1$ case of Regev's algorithm where the parameter b_1 is an element of $(\mathbb{Z}/N\mathbb{Z})^\times$ chosen uniformly at random. In this case, simple considerations from elementary number theory immediately imply the existence of an integer e_1 satisfying (i) and (ii) with probability $\asymp 1$. This would hold more generally for $d \geqslant 1$ if $b_1, \ldots, b_d$ were independent, uniformly distributed random elements of $(\mathbb{Z}/N\mathbb{Z})^\times$. However, for Regev's algorithm, the b_i's are far from uniformly distributed in $(\mathbb{Z}/N\mathbb{Z})^\times$ as they are constrained to lie in a very small subset of this group.

1.4.3 The number-theoretic conjecture behind Regev's algorithm

Regev's algorithm [50] and its space-efficient variant [49] crucially rely on the possibility of choosing very small integers $b_1, \ldots, b_d$ such that $\prod_{i=1}^{d} b_i^{e_i}$ is a non-trivial square root of 1 modulo N for some $e_1, \ldots, e_d$ with $|e_i| \leqslant N^{O(1/d)}$. We remind the reader that $n \asymp \log N$ and $d \asymp \sqrt{n}$.

The least restrictive bound[8] on the b_i's that still allows for a quantum circuit with $\tilde{O}(n^{3/2})$ gates is to have $|b_i| \leqslant \exp(\tilde{O}(d))$, where the $\tilde{O}(\cdot)$ notation possibly hides a factor $(\log n)^{O(1)}$. Moreover, it is natural to set the b_i's to be prime numbers (as in [14, 49, 50]) in order to avoid obvious multiplicative relations between them.

In this work, we choose $b_1, \ldots, b_d$ to be independent random prime numbers less than $d^{10^3 d}$. With these parameters, we can now state a version of Regev's conjecture[9] that follows from our results.

Corollary 1.4.4. *Let $N > 2$ be an integer. Let $d := \lceil \sqrt{\log N} \rceil$ and $X := d^{10^3 d}$.*

Let $\mathbf{b}_1, \ldots, \mathbf{b}_d$ be i.i.d. random variables, each uniformly distributed in the set of primes $\leqslant X$ not dividing N.

[8] The bound stated in Regev's paper [50] is $|b_i| \leqslant (\log N)^{O(1)}$, but this condition can be relaxed somewhat, as observed by Ragavan (private communication).

[9] Compare with [50, Theorem 1.1] or [49, Conjecture 1].

Then, with high probability, every $x \in \langle \mathbf{b}_1, \dots, \mathbf{b}_d \rangle$ *can be expressed as*

$$x \equiv \prod_{i=1}^{d} \mathbf{b}_i^{e_i} \pmod{N}$$

for some integers e_i *with* $|e_i| \leqslant e^{O(d)}$ *for all* $1 \leqslant i \leqslant d$.

The analogue of Regev's algorithm for the discrete logarithm problem, proposed by Ekerå and Gärtner [14], relies on a stronger version of Regev's conjecture. It concerns the geometry of a certain lattice $\mathcal{L}$ which encodes all multiplicative dependencies between the $\mathbf{b}_i$'s (modulo N). We prove the following version of it.[10]

Corollary 1.4.5. *Let* $N > 2$ *be an integer. Let* $d := \lceil \sqrt{\log N} \rceil$ *and* $X := d^{10^3 d}$.

Let $\mathbf{b}_1, \dots, \mathbf{b}_d$ *be i.i.d. random variables, each uniformly distributed in the set of primes* $\leqslant X$ *not dividing* N. *Let* $\mathcal{L}$ *be the random lattice defined by*

$$\mathcal{L} := \left\{ (e_1, \dots, e_d) \in \mathbb{Z}^d \ : \ \prod_{i=1}^{d} \mathbf{b}_i^{e_i} \equiv 1 \pmod{N} \right\}. \tag{1.12}$$

Then, with high probability, $\mathcal{L}$ *admits a basis consisting of vectors with Euclidean norm* $\leqslant e^{O(d)}$.

The main technical result of Chapter 4 is Theorem 4.3.18, of which Corollary 1.4.5 is a special case. In turn, Corollary 1.4.4 is a direct consequence of Corollary 1.4.5.

1.4.4 Subgroup obstructions

Let N, d and $\mathbf{b}_1, \dots, \mathbf{b}_d$ be as in Corollary 1.4.4. This corollary shows that, with high probability, every element x in the subgroup generated by $\mathbf{b}_1, \dots, \mathbf{b}_d$ can be written as a *short* product of these $\mathbf{b}_i$ modulo N. To obtain the full strength of Regev's assumption [49, Conjecture 1], it would be necessary to show that this subgroup $\langle \mathbf{b}_1, \dots, \mathbf{b}_d \rangle$ contains a non-trivial square root of 1 modulo N.

Unfortunately, it is not possible to prove this in full generality without considerable advances on the well-known *least quadratic non-residue problem* in number theory. For any prime number p, write $n(p)$ for the smallest positive integer a which is a quadratic non-residue (i.e. not a square) modulo p. The best known asymptotic upper bound for $n(p)$ is Burgess's classical result [9] that $n(p) \ll_\varepsilon p^{\frac{1}{4\sqrt{e}}+\varepsilon}$.

For simplicity, suppose that the integer N to be factored is a product of two equally-sized primes $p_1, p_2 \equiv 3 \pmod 4$. Recall that, for Regev's algorithm, the parameters $\mathbf{b}_i$ are chosen to be less than $\exp(\tilde{O}(d))$. With current techniques, we cannot rule out the possibility that both $n(p_1)$ and $n(p_2)$ are larger than this threshold. If this is the

[10] Compare with [14, Assumption 1], or Conjecture E.1 in the full version of [49].

case, $\langle \mathbf{b}_1, \ldots, \mathbf{b}_d \rangle$ is necessarily contained in the subgroup of squares modulo N. In particular, since $p_1, p_2 \equiv 3 \pmod 4$, the subgroup $\langle \mathbf{b}_1, \ldots, \mathbf{b}_d \rangle$ does not contain any non-trivial square roots of 1 modulo N.

This issue can be approached from several perspectives.

1. Assuming the Generalised Riemann Hypothesis, Ankeny proved the much stronger bound $n(p) \ll (\log p)^2$ [2]. This suggests that quadratic residues may no longer be an obstruction in this case, and indeed, under GRH, it is straightforward to show that $\langle \mathbf{b}_1, \ldots, \mathbf{b}_d \rangle$ contains a non-trivial square root of 1 modulo N with high probability. Assuming GRH would also considerably simplify the proof of Theorem 4.3.18 and remove a logarithmic factor for the gate and qubit costs in Theorems 1.4.1 and 1.4.2. However, in this thesis we seek fully unconditional results.

2. It is possible to prove the following partial result unconditionally: for *almost all* odd N with at least two distinct prime factors, the subgroup $\langle \mathbf{b}_1, \ldots, \mathbf{b}_d \rangle$ contains a non-trivial square root of 1 modulo N with high probability (with $d = d(N)$ and $\mathbf{b}_i = \mathbf{b}_i(N)$ as in Corollary 1.4.4). By Corollary 1.4.4, this implies that Regev's original algorithm (with these parameters $\mathbf{b}_i$) finds a non-trivial divisor of N with high probability, for *almost all* N. More precisely, the set $E \subset \mathbb{N}$ of exceptional values of N for which the above does not hold can be shown to satisfy
$$|E \cap [1, x]| \leqslant \exp\big((\log x)^{1/2+o(1)}\big)$$
for $x \geqslant 1$. This partial result can be proved using character sums, the Landau-Page theorem [42, Corollary 11.10] and a grand zero-density estimate [31, Theorem 1 (1.8)], by means of a case analysis depending on the prime factorisation of N.

3. In the present work, we follow a different approach to obtain a completely unconditional result that applies to *all* N, by slightly modifying the algorithm itself. While the key to the efficiency of Regev's algorithm is that the $\mathbf{b}_i$'s are small, one can tolerate a bounded number of large $\mathbf{b}_i$'s. In fact, this observation was already needed to adapt Regev's algorithm to the discrete logarithm problem [14]. We use this extra flexibility to overcome the subgroup obstructions. The simplest way to proceed is to allow for one of the parameters, say $\mathbf{b}_1$, to be uniformly distributed in $(\mathbb{Z}/N\mathbb{Z})^\times$. As with Shor's algorithm, this ensures that the subgroup $\langle \mathbf{b}_1 \rangle$, and hence also $\langle \mathbf{b}_1, \ldots, \mathbf{b}_d \rangle$, contains a non-trivial square root of 1 with probability $\asymp 1$.

Remark 1.4.6. In his paper [50], Regev proposed to select the parameters $b_1, \ldots, b_d$ deterministically, for example by setting b_i to be the i-th smallest prime for $1 \leqslant i \leqslant d$.

While such a deterministic choice of parameters is likely to work for *almost all* N, this seems to be very difficult to prove.

Remark 1.4.7. In our work, the parameters $\mathbf{b}_i$ are chosen to be random primes (not dividing N) under a certain threshold X. Another natural choice would have been to let $\mathbf{b}_1, \dots, \mathbf{b}_d$ be i.i.d. random variables uniformly distributed in the set of all integers up to X that are coprime to N. Our proofs would carry over to this framework if an analogue of Proposition 4.3.9 for short character sums $\sum_{n \leqslant x} \chi(n)$ were available. Assuming the Generalised Riemann Hypothesis, the work of Granville and Soundararajan [21, Theorem 2] gives a bound of the required strength. However, this approach does not seem sufficient to obtain an unconditional result (the bounds in [21] become too weak when $L(s, \chi)$ has zeroes close to the line $\operatorname{Re} s = 1$).

1.5 Notation

We write $f \ll g$ or $f = O(g)$ if $|f| \leqslant Cg$ for some absolute constant $C > 0$. If, instead, C depends on a parameter δ, we write $f \ll_\delta g$ or $f = O_\delta(g)$. The notation $f \asymp g$ means that $f \ll g$ and $g \ll f$. For a finite set S, we let $\mathbb{E}_{x \in S} f(x)$ denote the average $\frac{1}{|S|} \sum_{x \in S} f(x)$.

If $a, b \in \mathbb{Z}$, we write $[\![a, b]\!] := \mathbb{Z} \cap [a, b]$ (which is empty if $a > b$), and we call a set of this form a *discrete interval*. The length, or size, of a discrete interval is its cardinality. For $n \in \mathbb{N}$, we write $[\![n]\!] := \{1, 2, \dots, n\}$. If $A \subset \mathbb{R}$, we write $\pm A$ for $\{\sigma a : \sigma \in \{\pm 1\}, a \in A\}$. A set $A \subset \mathbb{R}$ is *t-separated* if $|x - y| \geqslant t$ for all distinct $x, y \in A$. For $\alpha \in \mathbb{R}$ and $X \subset \mathbb{R}$, we define $\operatorname{dist}(\alpha, X) := \inf_{x \in X} |\alpha - x|$, $\|\alpha\| := \operatorname{dist}(\alpha, \mathbb{Z})$ and $e(\alpha) := \exp(2\pi i \alpha)$.

Euler's totient function, the prime counting function, and the divisor sum function are denoted by φ, π, and σ_1, respectively. For $n \in \mathbb{Z}$, we write $\omega(n)$ for the number of distinct prime factors of n. If $\mathcal{P}$ is a set of primes, we let $\omega_{\mathcal{P}}(n)$ be the number of primes in $\mathcal{P}$ that divide n. The p-adic valuation of $n \in \mathbb{N}$ is denoted by $\nu_p(n)$. We write log for the natural logarithm and $\log_2$ for the logarithm in base 2.

For a finite abelian group G, we denote its character group by $\widehat{G}$. The order of a character $\chi \in \widehat{G}$ is written as $\operatorname{ord}(\chi)$. For elements $g_1, \dots, g_k \in G$, we let $\langle g_1, \dots, g_k \rangle$ denote the subgroup they generate.

We write $\|\cdot\|_2$ for the Euclidean norm and $\|\cdot\|_\infty$ for the sup norm. If $A = (a_{m,n})_{m,n \in I}$ is a square matrix with complex entries indexed by a set I, and $J \subset I$, we write $A|_J$ for the restriction of A to the rows and columns indexed by J. The conjugate transpose of A is denoted by A^*. The operator norm of A is $\|A\|_{\mathrm{op}} := \sup_{\|x\|_2 = 1} \|Ax\|_2$. The spectral radius of A is the largest absolute value of an eigenvalue of A. The eigenvalues of a matrix are always counted with multiplicity.

Chapter 2

On Chowla's conjecture for two-point correlations

2.1 Structure of the chapter

This chapter is devoted to the proof of our two-point logarithmic Chowla estimate, Theorem 1.2.1, which states that for some absolute constant $c > 0$,

$$\sum_{n \leqslant x} \frac{1}{n} \lambda(n)\lambda(n+1) \ll (\log x)^{1-c}.$$

We start by giving a broad overview of the proof structure. The reader may wish to refer to Fig. 2.1, which depicts the main propositions along with their logical dependencies. This chapter is designed to be as self-contained as possible. In particular, no prior familiarity with [26] is needed.

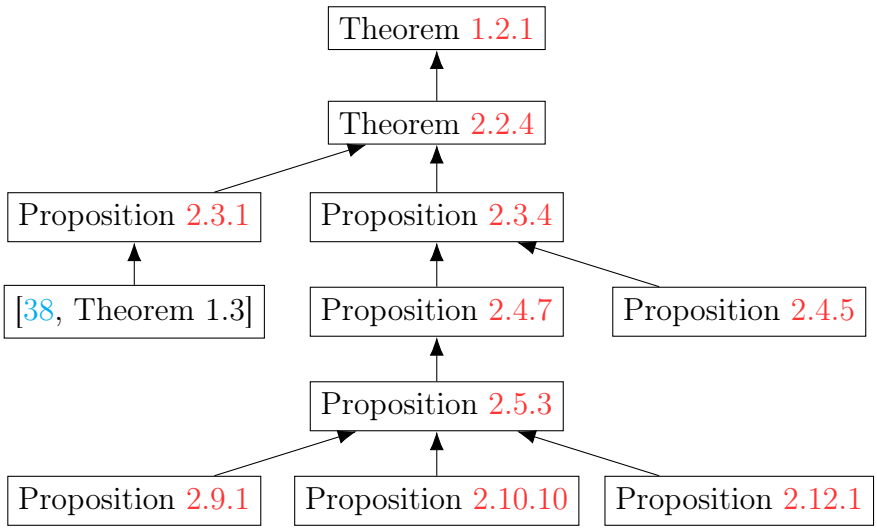


Figure 2.1: Dependency graph for the proof of Theorem 1.2.1 (main results only).

We begin by reducing Theorem 1.2.1 to Theorem 2.2.4, an estimate that involves an additional average over products of many primes. This reduction, carried out in Section 2.2, follows the clever manipulations of Tao [61].

The proof of Theorem 2.2.4 begins with Proposition 2.3.1, which replaces the double sum in Theorem 2.2.4 with a more convenient 'centred' version. The proof of Proposition 2.3.1 relies on an exponential sum estimate of Matomäki, Radziwiłł and Tao [38]. In order to proceed rapidly to the heart of our argument, we defer the proof of Proposition 2.3.1 to Section 2.14.

Following the approach of Helfgott and Radziwiłł [26], our task reduces to proving Proposition 2.3.4, a bound on the eigenvalues of a certain symmetric matrix (denoted $\widetilde{A}|_X$ in Section 1.2.2).

The high trace method cannot be applied directly to this matrix, due to the excessive contribution of 'backtracking walks' that repeatedly visit integers with a very large number of prime factors. Helfgott and Radziwiłł [26] overcame this obstacle by introducing smooth weights to discard such integers. Unfortunately, this method creates considerable technical difficulties, which would be even more severe in our setting.

Instead, we replace our matrix with a related *non-backtracking matrix*, which is no longer symmetric. We relate its spectral properties to those of the original matrix using a fundamental identity for non-backtracking operators, known as the *Ihara-Bass formula* (Proposition 2.4.5).

As a result, it suffices to prove Proposition 2.4.7, a tail estimate on the eigenvalues of the non-backtracking matrix. In turn, this spectral tail estimate follows from Proposition 2.5.3, which bounds the trace of a high power of the non-backtracking matrix.

The proof of this high trace bound is technically involved, occupying Sections 2.7 to 2.13. To guide the reader, we provide a detailed overview of this part of the argument in Section 2.6.

For ease of reading, we provide a table showing the main parameters, their size and a reference to where they are introduced.

Parameter	Size properties	First appearance
ε_1	$\varepsilon_1 > 0$ sufficiently small	Notation 2.2.1
H	$H = H(\varepsilon_1)$ sufficiently large	Notation 2.2.1
H_0	$H_0 = \exp((\log H)^{1-\varepsilon_1})$	Notation 2.2.1
J	$1 \leqslant J \leqslant \varepsilon_1^2 \log\log H$	Definition 2.2.2
V_j	$V_j = \sum_{p \in \mathcal{P}_j} 1/p \leqslant V$	Definition 2.2.2
V	$\varepsilon_1^{-1} \ll V \ll \varepsilon_1 \log\log H$	Definition 2.2.2
N	$N \geqslant \exp((\log H)^{2.001})$	Theorem 2.2.4
K	$K = \lfloor \log H \rfloor$	Notation 2.5.2
L	$L = K^{1-10\varepsilon_1}$	Notation 2.7.1

The following table contains most of the other symbols used repeatedly in the paper.

Notation	Description	Reference
$\mathcal{P}_j$	disjoint sets of primes $\subset (H_0, H)$	Definition 2.2.2
$\mathcal{P}$	$\mathcal{P} := \bigcup_j \mathcal{P}_j$	Notation 2.2.7
$\mathcal{D}$	set of all products $p_1 p_2 \cdots p_J$ with each $p_j \in \mathcal{P}_j$	Notation 2.2.7
I_N	$I_N := \mathbb{N} \cap (N, 2N]$	Notation 2.2.7
w_Y	explicit function $I_N \times I_N \to \mathbb{R}$ depending on $Y \subset \mathbb{Z}$	Definition 2.3.2
G_Y	graph on I_N with edge weights given by w_Y	Definition 2.4.6
A_Y	weighted adjacency matrix of G_Y	Definitions 2.3.2 and 2.4.6
M_Y	non-backtracking matrix associated to G_Y	Definition 2.4.6
I_N^{typ}	set of $n \in I_N$ with a moderate number of prime factors in $\mathcal{P}$	Definition 2.4.9
b_i	partial sums $b_i := \sum_{1 \leqslant k \leqslant i} d_k$	Lemma 2.5.4
$\mathcal{Y}$	set of all prohibited progressions	Definition 2.7.4
Y_L	integers not contained in a prohibited progression	Definition 2.7.4
d_{ij}	unique prime in $\mathcal{P}_j$ dividing d_i	Definition 2.8.1
$\boldsymbol{d}_*$	product of all d_{ij}	Definition 2.8.1
$\boldsymbol{d}_I$	product of all d_{ij} with $(i,j) \in I$	Definition 2.8.1
$\mathcal{S}$	set of single indices	Definition 2.8.2
$\mathcal{L}, \mathcal{U}$	sets of lit and unlit indices	Lemma 2.8.6
$\mathbf{D}_{\mathcal{S},\mathcal{L},r}$	set of vectors $\boldsymbol{d} \in (\pm\mathcal{D})^{2K}$ of interest	Definition 2.8.4
$E_{\mathcal{S},\mathcal{L},\mathcal{U},r}$	quantity to be bounded	Definition 2.8.5
$A_{\mathcal{L},\boldsymbol{d}}$	arithmetic progression determined by the lit indices	Lemma 2.9.6
$\boldsymbol{d}^{(k)}$	subsequence of $\boldsymbol{d}$, for $k = 1, 2, 3$	Definition 2.8.4
$v_{j,\boldsymbol{d}}^{(k)}, w_{j,\boldsymbol{d}}^{(k)}$	words associated to $\boldsymbol{d}$ with letters in $\mathcal{P}_j$	Definition 2.10.9
$\mathbf{Prd}_{\mathcal{S},\mathcal{L},r}$	vectors $\boldsymbol{d}$ such that all $w_{j,\boldsymbol{d}}^{(k)}$ are predictable	Definition 2.10.9
$\mathbf{Unp}_{\mathcal{S},\mathcal{L},r}$	vectors $\boldsymbol{d}$ such that some $w_{j,\boldsymbol{d}}^{(k)}$ is unpredictable	Definition 2.10.9
$\mathbf{Red}_{R,\mathcal{S},\mathcal{L}}$	reduced vectors $\boldsymbol{d} \in (\pm\mathcal{D})^R$ compatible with $\mathcal{S}$ and $\mathcal{L}$	Definition 2.12.12

2.2 Main technical theorem and consequences

2.2.1 Main parameters

To state the main theorem of this chapter, we need to introduce some parameters.

Notation 2.2.1 (ε_1, H, H_0)**.** Let $\varepsilon_1 > 0$ be a sufficiently small constant. Let $H \in \mathbb{N}$ be sufficiently large in terms of ε_1. Let $H_0 := \exp((\log H)^{1-\varepsilon_1})$.

Definition 2.2.2 (J, $\mathcal{P}_i$, V_i, V)**.** Let J be an integer satisfying $1 \leqslant J \leqslant \varepsilon_1^2 \log\log H$.

For $1 \leqslant j \leqslant J$, let $\mathcal{P}_j$ be the set of all primes p such that

$$\rho^{2j-2} < \frac{\log p}{\log H_0} < \rho^{2j-1},$$

where $\rho := \exp\big(\varepsilon_1(\log\log H)/(2J)\big)$.

For each $1 \leqslant j \leqslant J$, we define $V_j := \sum_{p \in \mathcal{P}_j} 1/p$, and we let $V := \max_{1 \leqslant j \leqslant J} V_j$.

The next lemma collects some properties of the sets $\mathcal{P}_j$ and parameters V_j that will be used throughout the chapter.

Lemma 2.2.3. *The following hold:*

(a) $\mathcal{P}_1, \dots, \mathcal{P}_J$ *are disjoint subsets of* (H_0, H)*;*

(b) if $(p_1, \dots, p_J) \in \mathcal{P}_1 \times \cdots \times \mathcal{P}_J$*, then* $p_1 p_2 \cdots p_J < H$ *and* $p_1 p_2 \cdots p_i < p_{i+1}^{1/10}$ *for all* $1 \leqslant i < J$*;*

(c) $\varepsilon_1^{-1} \ll V \asymp \dfrac{\varepsilon_1 \log\log H}{J} \ll \varepsilon_1 \log\log H$*;*

(d) $V_1 V_2 \cdots V_J \asymp V^J \ll (\log H)^{c_0(\varepsilon_1)}$ *where* $c_0(\varepsilon_1) := \varepsilon_1^2 \log(\frac{1}{2\varepsilon_1})$*. Moreover, in the special case where* $J = \lfloor \varepsilon_1^2 \log\log H \rfloor$*, we have* $V^J \asymp_{\varepsilon_1} (\log H)^{c_0(\varepsilon_1)}$*.*

Proof. Let $\rho := \exp\big(\varepsilon_1(\log\log H)/(2J)\big)$; note that $\rho \geqslant 20$. Since $\mathcal{P}_j$ is the set of all primes in the interval

$$\Big(\exp\big(\rho^{2j-2}\log H_0\big),\ \exp\big(\rho^{2j-1}\log H_0\big)\Big),$$

property (a) is clear. If $(p_1, \dots, p_J) \in \mathcal{P}_1 \times \cdots \times \mathcal{P}_J$, then for all $i \in [\![J]\!]$ we have

$$p_1 p_2 \cdots p_i < \exp\Bigg(\sum_{1 \leqslant j \leqslant i} \rho^{2j-1} \log H_0\Bigg) \leqslant \exp\Big(2\rho^{2i-1}\log H_0\Big) \leqslant \exp\Big(\frac{\rho^{2i}}{10}\log H_0\Big).$$

The right-hand side is $\leqslant p_{i+1}^{1/10}$ if $i < J$, and equals $H^{1/10}$ if $i = J$. This proves (b).

By Mertens' second estimate, we have

$$V_j = \frac{\varepsilon_1 \log\log H}{2J} + O\Big(\frac{1}{\log H_0}\Big),$$

which implies property (c) and the bounds

$$V_1 V_2 \cdots V_J \asymp V^J \asymp \Big(\frac{\varepsilon_1 \log\log H}{2J}\Big)^J.$$

Finally, since $J \leqslant \varepsilon_1^2 \log\log H$ and the function $x \mapsto (A/x)^x$ is increasing on $[0, A/e]$, property (d) follows. □

2.2.2 Statement of the main theorem

Our bound for the two-point logarithmic Chowla correlations is a consequence of the following key estimate.

Theorem 2.2.4. *Let* N *be an integer such that* $\log N \geqslant (\log H)^{2.001}$*. Then*

$$\Bigg|\sum_{(p_1,\dots,p_J)\in\mathcal{P}_1\times\cdots\times\mathcal{P}_J} \sum_{\substack{n\in(N,2N]\\ p_1\cdots p_J \mid n}} \lambda(n)\lambda(n+p_1\cdots p_J)\Bigg| \leqslant e^{O(J)}(V_1\cdots V_J)^{1/2}N. \tag{2.1}$$

Remark 2.2.5. Theorem 2.2.4 can be compared with the trivial bound $\ll V_1 \cdots V_J N$.

For $J = 1$, we recover the result of Helfgott and Radziwiłł [26]. Crucially, however, our approach allows J to be as large as a small multiple of $\log\log H$. By Lemma 2.2.3, this flexibility enables us to save a power of $\log H$.

Remark 2.2.6. In fact, our methods yield a slightly stronger estimate with absolute values:

$$\sum_{(p_1,\dots,p_J)\in\mathcal{P}_1\times\cdots\times\mathcal{P}_J} \left| \sum_{\substack{n\in(N,2N]\\ p_1\cdots p_J\mid n}} \lambda(n)\lambda(n+p_1\cdots p_J)\right| \leqslant e^{O(J)}(V_1\cdots V_J)^{1/2}N. \tag{2.2}$$

This follows by repeating the proof with arbitrary signs $c_{p_1,\dots,p_J} \in \{\pm 1\}$, but for clarity we focus on the version in Theorem 2.2.4, which suffices for our main application.

2.2.3 Proof of the two-point logarithmic Chowla bound

Notation 2.2.7 ($\mathcal{D}$, I_N, $\mathcal{P}$)**.** To shorten the expressions, we define $\mathcal{D} \subset (H_0, H)$ to be the set of all products $p_1p_2\cdots p_J$ with $p_j \in \mathcal{P}_j$ for all $j \in [\![J]\!]$. We also write $I_N := \mathbb{N}\cap(N, 2N]$ and $\mathcal{P} := \bigsqcup_{j=1}^{J}\mathcal{P}_j$.

With this notation, we can rewrite the double sum in (2.1) as

$$\sum_{d\in\mathcal{D}}\sum_{\substack{n\in I_N\\ d\mid n}} \lambda(n)\lambda(n+d).$$

We now show how a bound on this expression implies a bound on the two-point logarithmic Chowla conjecture. This step is due to Tao [61], and crucially relies on the multiplicativity of λ. Together with the proof of Proposition 2.3.1, this is the only place where the multiplicativity of λ is used – the rest of the proof will only use that λ is 1-bounded.

Proof of Theorem 1.2.1, assuming Theorem 2.2.4. Let ε_1, H and H_0 be as in Notation 2.2.1. In particular, $\varepsilon_1 > 0$ is a sufficiently small absolute constant, and H is chosen sufficiently large in terms of ε_1. Choose $J := \lfloor \varepsilon_1^2 \log\log H\rfloor$. Let $(\mathcal{P}_j)$, (V_j) and V be as in Definition 2.2.2. Thus, by Lemma 2.2.3,

$$V^J \asymp_{\varepsilon_1} (\log H)^{\varepsilon_1^2\log(\frac{1}{2\varepsilon_1})}. \tag{2.3}$$

Finally, let $x := \exp((\log H)^6)$.

For any $d \in \mathcal{D}$, since λ is completely multiplicative and $\lambda^2 = 1$, we may rewrite the logarithmic Chowla average as

$$\sum_{n\leqslant x}\frac{1}{n}\lambda(n)\lambda(n+1) = \sum_{n\leqslant x}\frac{1}{n}\lambda(dn)\lambda(dn+d) = d\sum_{\substack{m\leqslant dx\\ d\mid m}}\frac{1}{m}\lambda(m)\lambda(m+d).$$

Dividing by d and summing over all $d \in \mathcal{D}$ yields

$$V_1 \cdots V_J \sum_{n \leqslant x} \frac{1}{n} \lambda(n)\lambda(n+1) = \sum_{d \in \mathcal{D}} \sum_{\substack{m \leqslant dx \\ d \mid m}} \frac{1}{m} \lambda(m)\lambda(m+d). \tag{2.4}$$

We now apply Theorem 2.2.4 in a dyadic fashion to bound the right-hand side of (2.4).

For every N such that $\log N \geqslant (\log H)^3 = \sqrt{\log x}$, by Theorem 2.2.4, we know that

$$\left| \sum_{d \in \mathcal{D}} \sum_{\substack{n \in I_N \\ d \mid n}} \lambda(n)\lambda(n+d) \right| \leqslant e^{O(J)} V^{J/2} N. \tag{2.5}$$

Moreover, when $\log N \leqslant \sqrt{\log x}$, trivially bounding $|\lambda| \leqslant 1$ we have

$$\sum_{d \in \mathcal{D}} \sum_{\substack{n \in I_N \\ d \mid n}} \lambda(n)\lambda(n+d) \ll V^J N. \tag{2.6}$$

By a suitable dyadic decomposition, (2.5) and (2.6) together give, for all $M \geqslant 1$, the bound

$$\left| \sum_{d \in \mathcal{D}} \sum_{\substack{m \leqslant M \\ d \mid m}} \lambda(m)\lambda(m+d) \right| \ll V^J \min\bigl(M, e^{\sqrt{\log x}}\bigr) + e^{O(J)} V^{J/2} M. \tag{2.7}$$

By partial summation (and using (2.3)), the estimate (2.7) implies that

$$\left| \sum_{d \in \mathcal{D}} \sum_{\substack{m \leqslant x \\ d \mid m}} \frac{1}{m} \lambda(m)\lambda(m+d) \right| \leqslant e^{O(J)} V^{J/2} \log x. \tag{2.8}$$

This is almost the expression in (2.4), up to an error

$$\left| \sum_{d \in \mathcal{D}} \sum_{\substack{x < m \leqslant dx \\ d \mid m}} \frac{1}{m} \lambda(m)\lambda(m+d) \right| \leqslant \sum_{d \in \mathcal{D}} \sum_{\substack{x < m \leqslant dx \\ d \mid m}} \frac{1}{m} \ll \sum_{d \in \mathcal{D}} \frac{\log(d)}{d} \ll V^J \log H. \tag{2.9}$$

Hence, by (2.4), (2.8) and (2.9), we conclude that

$$\sum_{n \leqslant x} \frac{1}{n} \lambda(n)\lambda(n+1) \ll \frac{1}{V^J} \bigl(e^{O(J)} V^{J/2} \log x + V^J \log H \bigr).$$

By (2.3), this is $\ll_{\varepsilon_1} (\log x)^{1-c}$ for some constant $c > 0$ depending only on ε_1, as desired. □

Remark 2.2.8 (Two-point Chowla at almost all scales)**.** Our main result also implies an improved quantitative version of Chowla's conjecture for two-point correlations at *almost all scales*. Namely, for all $2 < w \leqslant X$, we have

$$\frac{1}{\log w} \int_{X/w}^{X} \left| \frac{1}{x} \sum_{n \leqslant x} \lambda(n)\lambda(n+1) \right| \frac{dx}{x} \ll \frac{1}{(\log w)^c}, \tag{2.10}$$

where $c > 0$ is an absolute constant. In particular, we get

$$\frac{1}{x}\sum_{n\leqslant x}\lambda(n)\lambda(n+1) \ll \frac{1}{(\log X)^{c/2}}$$

for all $x \in [1, X]$ outside of a set E_X of logarithmic density $O((\log X)^{-c/2})$.[1]

This almost all scales result (2.10) follows from (2.2) by a straightforward adaptation of the proof in [26, Section 8] (which in turn was inspired by [64]) to our setting.

2.3 Reformulations of the main theorem

2.3.1 Centred sum

The bound in Theorem 2.2.4 turns out to be easier to prove if the divisibility term $\mathbf{1}_{p_1|n}\cdots\mathbf{1}_{p_J|n}$ is replaced by the 'centred' product $\left(\mathbf{1}_{p_1|n}-\frac{1}{p_1}\right)\cdots\left(\mathbf{1}_{p_J|n}-\frac{1}{p_J}\right)$. The following proposition shows that working with either of these two expressions is equivalent up to a small error.

Proposition 2.3.1. *Let N be an integer such that $\log N \geqslant (\log H)^2$. Let*

$$S_1 := \sum_{n\in I_N}\sum_{\substack{d\in\mathcal{D}\\ d|n}}\lambda(n)\lambda(n+d) \quad \text{and} \quad S_2 := \sum_{n\in I_N}\sum_{d\in\mathcal{D}}\lambda(n)\lambda(n+d)\prod_{p|d}\left(\mathbf{1}_{p|n}-\frac{1}{p}\right).$$

Then

$$|S_1 - S_2| \ll \frac{N}{(\log H)^{1/2500}}.$$

Proposition 2.3.1 is proved in Section 2.14, using the circle method and crucially relying on an estimate of Matomäki, Radziwiłł and Tao [38]. We will now focus on the centred expression S_2.

2.3.2 Reduction to an eigenvalue bound

To prove our main theorem, we analyse the spectral properties of a matrix whose entries correspond to the coefficients $\prod_{p|d}\left(\mathbf{1}_{p|n}-\frac{1}{p}\right)$ appearing in S_2. This key step effectively suppresses the role of the Liouville function in the problem.

Definition 2.3.2 (w_Y, A_Y)**.** For any set $Y \subset \mathbb{Z}$, let $w_Y : I_N \times I_N \to \mathbb{R}$ be the function defined by

$$w_Y(m,n) := \begin{cases} \prod_{p|d}\left(\mathbf{1}_{p|n}-\frac{1}{p}\right) & \text{if } |m-n| = d \in \mathcal{D} \text{ and } m,n \in Y,\\ 0 & \text{otherwise.}\end{cases}$$

[1] This means that $\frac{1}{\log X}\int_1^X \mathbf{1}_{E_X}(x)\frac{dx}{x} \ll (\log X)^{-c/2}$.

Define A_Y to be the real symmetric matrix[2] $A_Y := (w_Y(m,n))_{m,n\in I_N}$.

In this section, we take $Y = \mathbb{Z}$ (so the condition $m, n \in Y$ can be omitted). The general definition of A_Y will be useful later: by analysing A_Y for a different, more technical choice of Y (introduced in Definition 2.7.4), we will obtain spectral information that ultimately applies to $A_{\mathbb{Z}}$.

As a consequence of this analysis, almost all eigenvalues of $A_{\mathbb{Z}}$ are bounded by $e^{O(J)}V^{J/2}$ in absolute value.[3] In fact, we prove a stronger result (Proposition 2.3.4) which, combined with Proposition 2.3.1, yields our main theorem. To state it, we recall the following notation.

Notation 2.3.3. Given a matrix $A = (a_{m,n})_{m,n\in I_N}$ and a subset $X \subset I_N$, we write $A|_X$ for the principal submatrix $(a_{m,n})_{m,n\in X}$ obtained by deleting all rows and columns at indices not in X.

Proposition 2.3.4 (Spectral radius of restriction)**.** *Let N be an integer such that $\log N \geqslant (\log H)^{2.001}$.*

Then, there is a subset $X \subset I_N$ with $|I_N \setminus X| \ll e^{-JV/3}N$ such that all eigenvalues of $(A_{\mathbb{Z}})|_X$ are bounded in absolute value by $e^{O(J)}V^{J/2}$.

Remark 2.3.5. The restriction to a dense subset $X \subset I_N$ is essential. Indeed, the full matrix $A_{\mathbb{Z}}$ does have very large eigenvalues, which arise from integers $n \in I_N$ having an unusually large number of prime factors from $\mathcal{P}$. For instance, if $n \in I_N$ is such that $\omega_{\mathcal{P}_j}(n) \geqslant CV$ for every $1 \leqslant j \leqslant J$, and $\boldsymbol{v}$ is the standard basis vector corresponding to n, then

$$\|A_{\mathbb{Z}}\boldsymbol{v}\|_2 \geqslant \left(\sum_{\substack{d\in\mathcal{D}\\ d|n}}\prod_{p|d}\left(1-\frac{1}{p}\right)^2\right)^{1/2} \gg \left(\prod_{j=1}^{J}\omega_{\mathcal{P}_j}(n)\right)^{1/2} \gg C^{J/2}V^{J/2}.$$

Consequently, $A_{\mathbb{Z}}$ possesses an eigenvalue of size at least $\gg C^{J/2}V^{J/2}$ in absolute value, exceeding the desired bound if C is too large. This large eigenvalue directly reflects the exceptional factorisation of n. The previous computation also shows that the eigenvalue bound in Proposition 2.3.4 is optimal up to the $e^{O(J)}$ factor.

Proof of Theorem 2.2.4, assuming Proposition 2.3.4. Let S_1 and S_2 be as in Proposition 2.3.1. Our goal is to show that $|S_1| \leqslant e^{O(J)}V^{J/2}N$. By Proposition 2.3.1, it suffices to prove the same bound for S_2.

[2]Note that the entries of A_Y are indexed by the elements of I_N, rather than $\{1,\dots,N\}$.
[3]This follows from Proposition 2.3.4 via Cauchy's interlacing theorem.

Let $\boldsymbol{\lambda} := (\lambda(n))_{n\in I_N}$ be the vector of Liouville function values. Up to a negligible error, we can express S_2 in terms of the matrix $A_{\mathbb{Z}}$ as:

$$S_2 = \sum_{\substack{m,n\in I_N\\ m<n}} \lambda(m)\lambda(n)w_{\mathbb{Z}}(m,n) + O(H^2) = \frac{1}{2}\boldsymbol{\lambda}^{\top}A_{\mathbb{Z}}\,\boldsymbol{\lambda} + O(H^2).$$

To exploit Proposition 2.3.4, we restrict $A_{\mathbb{Z}}$ to the subset $X \subset I_N$ and write $\boldsymbol{\lambda}|_X := (\lambda(n))_{n\in X}$. Since $|\lambda| \leqslant 1$, we have

$$S_2 = \frac{1}{2}(\boldsymbol{\lambda}|_X)^{\top}(A_{\mathbb{Z}})|_X(\boldsymbol{\lambda}|_X) + O\left(H^2 + \sum_{n\in I_N\setminus X}\sum_{d\in\mathcal{D}}\prod_{p|d}\left(\mathbf{1}_{p|n}+\frac{1}{p}\right)\right). \tag{2.11}$$

By Proposition 2.3.4, the operator norm of $(A_{\mathbb{Z}})|_X$ is bounded by $e^{O(J)}V^{J/2}$, as this is a real symmetric matrix. Therefore, by Cauchy-Schwarz,

$$\left|(\boldsymbol{\lambda}|_X)^{\top}(A_{\mathbb{Z}})|_X(\boldsymbol{\lambda}|_X)\right| \leqslant \|\boldsymbol{\lambda}|_X\|_2\|(A_{\mathbb{Z}})|_X(\boldsymbol{\lambda}|_X)\|_2 \leqslant e^{O(J)}V^{J/2}\|\boldsymbol{\lambda}|_X\|_2^2 \leqslant e^{O(J)}V^{J/2}N,$$

which gives the desired bound for the main term in (2.11).

It remains to treat the error term. By the inequality of arithmetic and geometric means (AM-GM inequality),

$$\sum_{d\in\mathcal{D}}\prod_{p|d}\left(\mathbf{1}_{p|n}+\frac{1}{p}\right) = \prod_{j=1}^{J}\left(\omega_{\mathcal{P}_j}(n)+V_j\right) \leqslant \left(\frac{\omega_{\mathcal{P}}(n)}{J}+V\right)^{J}.$$

Hence, by Cauchy-Schwarz,

$$\sum_{n\in I_N\setminus X}\sum_{d\in\mathcal{D}}\prod_{p|d}\left(\mathbf{1}_{p|n}+\frac{1}{p}\right) \leqslant |I_N\setminus X|^{1/2}\left(V^{2J}\sum_{n\in I_N}\left(\frac{\omega_{\mathcal{P}}(n)/V+J}{J}\right)^{2J}\right)^{1/2}.$$

Since $(a/k)^k \leqslant e^a$ for any $a \geqslant 0$ and $k\in\mathbb{N}$, we have

$$\sum_{n\in I_N}\left(\frac{\omega_{\mathcal{P}}(n)/V+J}{J}\right)^{2J} \leqslant e^{O(J)}\sum_{n\in I_N}e^{2\omega_{\mathcal{P}}(n)/V} \leqslant e^{O(J)}N$$

using [45, Lemma (3.10)] for the last inequality. Recalling that $|I_N\setminus X| \ll e^{-JV/3}N$, we conclude that

$$\sum_{n\in I_N\setminus X}\sum_{d\in\mathcal{D}}\prod_{p|d}\left(\mathbf{1}_{p|n}+\frac{1}{p}\right) \ll e^{O(J)}V^{J}e^{-JV/6}N.$$

By part (c) of Lemma 2.2.3, and since ε_1 is assumed to be sufficiently small, this is $\ll e^{-JV/8}N$. Thus, the error term in (2.11) is negligible compared to our main term bound. □

2.4 Approach via non-backtracking operators

In the previous section, we reduced the proof of our main theorem to understanding the spectral properties of the matrix $A_{\mathbb{Z}}$ from Definition 2.3.2. This matrix can be viewed as the weighted adjacency matrix of a certain weighted graph. To make the analysis more accessible to the trace method, we instead consider a related object: the *non-backtracking matrix* of this graph. We then use the *Ihara-Bass formula* to transfer spectral information back to the original adjacency matrix.

Remark 2.4.1. The use of non-backtracking operators in conjunction with the high trace method has found various applications in other areas of mathematics and computer science, see e.g. [5, 8, 16, 19]. In particular, Friedman [19] and Bordenave [8] employed this approach in their proofs of Alon's second eigenvalue conjecture, which states that random d-regular graphs are almost Ramanujan (i.e., near-optimal spectral expanders).

2.4.1 Adjacency and non-backtracking operators

We now introduce the necessary definitions and notation for working with weighted graphs and their associated linear operators.

Definition 2.4.2. A *weighted graph* G is a triple (V, E, w) where

- V is a finite set of vertices,
- $E \subseteq \binom{V}{2}$ is the set of edges,
- $w : E \to \mathbb{R}$ is a real-valued weight function.

The corresponding set of *directed edges* is defined as

$$\vec{E} := \{(x, y) \in V^2 : \{x, y\} \in E\}.$$

For any $(x, y) \in \vec{E}$, we define its weight as $w(x, y) := w(\{x, y\})$. For pairs $(x, y) \notin \vec{E}$, we set $w(x, y) := 0$. Note that this implies $w(x, y) = w(y, x)$ and $w(x, x) = 0$.

Notation 2.4.3. For any set S, we write $L(S)$ for the $\mathbb{C}$-vector space freely generated by S. Thus, $L(V)$ and $L(\vec{E})$ are $\mathbb{C}$-vector spaces with bases V and $\vec{E}$ respectively, and we denote their corresponding basis vectors by $\mathtt{x}$ for $x \in V$ and $\overrightarrow{\mathtt{xy}}$ for $(x, y) \in \vec{E}$.

Definition 2.4.4. Let $G = (V, E, w)$ be a weighted graph. The *adjacency operator* A associated to G is the linear map $A : L(V) \to L(V)$ defined for each basis vector $\mathtt{x}$ of $L(V)$ by

$$A(\mathtt{x}) := \sum_{y \in V} w(x, y)\, \mathtt{y}.$$

The *non-backtracking operator* M associated to G is the linear map $M : L(\vec{E}) \to L(\vec{E})$ acting on a basis vector $\overrightarrow{\mathtt{xy}}$ (where $(x, y) \in \vec{E}$) as

$$M(\overrightarrow{\mathtt{xy}}) := \sum_{\substack{(y,z)\in\vec{E} \\ z\neq x}} w(y, z)\, \overrightarrow{\mathtt{yz}}$$

and extended linearly to $L(\vec{E})$.

We will sometimes identify these operators with their matrices in the natural bases of $L(V)$ and $L(\vec{E})$, respectively. Note that A is symmetric, while M is usually not.

2.4.2 The Ihara-Bass formula

The *Ihara-Bass formula* establishes a remarkable connection between the characteristic polynomial of the non-backtracking operator and a deformed characteristic polynomial of the adjacency operator. This formula originates from the work of Ihara [28], who proved it in an algebraic setting corresponding to regular graphs, and was later extended by Bass [4] to the general unweighted case. The paper by Stark and Terras [59] is a recommended source for a transparent and elegant proof.

In our setting, we require a version of the Ihara-Bass formula applicable to weighted graphs. We have identified three papers that seem to have derived such an extension independently [12, 16, 74]. We state below a formulation adapted to our notation.

Proposition 2.4.5 (Ihara-Bass formula)**.** *Let $G = (V, E, w)$ be a weighted graph, with adjacency and non-backtracking operators A and M. Define the linear maps $D_u, O_u : L(V) \to L(V)$ for each basis vector $\mathtt{x}$ by*

$$D_u(\mathtt{x}) := \sum_{y\in V} \frac{w(x, y)^2}{1 - u^2 w(x, y)^2}\mathtt{x} \quad \textit{and} \quad O_u(\mathtt{x}) := u \sum_{y\in V} \frac{w(x, y)^3}{1 - u^2 w(x, y)^2}\mathtt{y}, \tag{2.12}$$

where $u \in \mathbb{C}$ is a parameter chosen such that all denominators are non-zero. Then,[4]

$$\det(I - uM) = \det\big(I - uA + u^2(D_u - O_u)\big) \prod_{\{x,y\}\in E} \big(1 - u^2 w(x, y)^2\big). \tag{2.13}$$

We detail in Chapter A how Proposition 2.4.5 can be derived from the results in [12], where the notation differs slightly from ours.

The Ihara-Bass formula will be applied to the graph whose edges are weighted by the function w_Y from Definition 2.3.2.

[4] In the formula (2.13), the same letter I denotes the identity operator on $L(\vec{E})$ on the left-hand side, and the identity operator on $L(V)$ on the right-hand side.

Definition 2.4.6 (G_Y, M_Y, $D_{Y,u}$, $O_{Y,u}$)**.** For any set $Y \subset \mathbb{Z}$, we define G_Y to be the weighted graph with vertex set I_N and weight function w_Y.[5]

We write A_Y and M_Y for the adjacency and non-backtracking operators associated to G_Y.[6] In addition, we denote by $D_{Y,u}$ and $O_{Y,u}$ the operators D_u and O_u from Proposition 2.4.5, specialised to the weighted graph G_Y.

We now state a central proposition, which provides a bound for almost all eigenvalues of the non-backtracking operator M_Y, for some well-chosen set Y. Recall that eigenvalues are always counted with multiplicity.

Proposition 2.4.7 (Spectral tail bound for M_Y)**.** *Let N be an integer such that $\log N \geqslant (\log H)^{2.001}$.*

Then, there exists a set $Y \subset \mathbb{Z}$ with the following properties.

1. $|I_N \setminus Y| \ll H_0^{-1/3} N$.
2. *All but $\leqslant N/H^3$ eigenvalues of M_Y have absolute value at most $e^{O(J)} V^{J/2}$.*

The plan for the rest of this section is to use the Ihara-Bass formula to relate the spectrum of M_Y to that of the adjacency operator A_Y. This will allow us to prove Proposition 2.3.4 assuming Proposition 2.4.7. The proof of Proposition 2.4.7 will occupy the rest of the chapter.

Remark 2.4.8. Several previous works have also used the Ihara-Bass formula to relate the spectra of the adjacency and non-backtracking operators [5, 8, 19].

For d-regular graphs (the setting of [8, 19]), this relation is immediate: the nontrivial eigenvalues λ of the non-backtracking operator are obtained explicitly as the roots of the quadratic equations $\lambda^2 - \mu\lambda + (d-1) = 0$, where μ ranges over the eigenvalues of the adjacency operator.

The random graphs considered in [5] have a maximum degree that is, with high probability, very close to the average degree (a consequence of concentration of measure). This near-regularity allows the perturbative term in the Ihara-Bass formula to be bounded uniformly.

In our setting, the weighted graph G_Y is highly irregular, with some vertices having very large weighted degrees. Moreover, we require a quantitative analysis that controls not only the magnitude, but also the number of large eigenvalues.

[5] The edge set of G_Y is defined to be the set of all $\{m, n\} \subset I_N$ with $w_Y(m, n) \neq 0$.

[6] Note that this definition of A_Y is consistent with the one in Definition 2.3.2.

2.4.3 Integers with many prime factors

As explained in Remark 2.3.5, the integers $n \in I_N$ with an unusually large number of prime factors in $\mathcal{P}$ are responsible for large eigenvalues of the adjacency operator A_Y (which disappear when passing to the non-backtracking operator M_Y). To handle these large eigenvalues, we study the restriction of A_Y, $D_{Y,u}$, and $O_{Y,u}$ to the integers not having much more than the typical number of prime factors from $\mathcal{P}$.

Definition 2.4.9. Let I_N^{typ} be the set of integers $n \in I_N$ such that $\omega_{\mathcal{P}}(n) \leqslant 2JV$.

Lemma 2.4.11 shows that the matrices $D_{Y,u}$ and $O_{Y,u}$, when restricted to I_N^{typ}, have small operator norms, uniformly in $u \in [-1/2, 1/2]$. It relies on the following elementary bound.

Lemma 2.4.10 ($\ell_1 - \ell_\infty$ bound)**.** *Let $A \in M_n(\mathbb{C})$ be a matrix with entries $(a_{ij})_{1 \leqslant i,j \leqslant n}$. Then*

$$\|A\|_{\mathrm{op}}^2 \leqslant \left(\max_i \sum_{j=1}^n |a_{ij}|\right)\left(\max_j \sum_{i=1}^n |a_{ij}|\right).$$

Proof. By Cauchy-Schwarz, for any $v \in \mathbb{C}^n$ we have

$$\|Av\|_2^2 \leqslant \sum_{i=1}^n \left(\sum_{j=1}^n |a_{ij}|\right)\left(\sum_{j=1}^n |a_{ij}||v_j|^2\right) \leqslant \left(\max_i \sum_{j=1}^n |a_{ij}|\right)\left(\max_j \sum_{i=1}^n |a_{ij}|\right)\|v\|_2^2,$$

which gives the claimed operator norm bound. □

Lemma 2.4.11. *Let $Y \subset \mathbb{Z}$ be any set. Then, for any $u \in [-1/2, 1/2]$,*

$$\left\|D_{Y,u}|_{I_N^{\mathrm{typ}}}\right\|_{\mathrm{op}}, \left\|O_{Y,u}|_{I_N^{\mathrm{typ}}}\right\|_{\mathrm{op}} \leqslant e^{O(J)}V^J.$$

Proof. We apply the simple $\ell_1 - \ell_\infty$ bound given by Lemma 2.4.10. Uniformly in $u \in [-1/2, 1/2]$, using that $\|w_Y\|_\infty \leqslant 1$, we get

$$\left\|D_{Y,u}|_{I_N^{\mathrm{typ}}}\right\|_{\mathrm{op}}, \left\|O_{Y,u}|_{I_N^{\mathrm{typ}}}\right\|_{\mathrm{op}} \ll \max_{m \in I_N^{\mathrm{typ}}} \sum_{n \in I_N^{\mathrm{typ}}} |w_Y(m,n)|.$$

By the AM-GM inequality, for any $m \in I_N^{\mathrm{typ}}$, we have

$$\sum_{n \in I_N^{\mathrm{typ}}} |w_Y(m,n)| \ll \sum_{d \in \mathcal{D}} \prod_{p \mid d}\left(\mathbf{1}_{p \mid m} + \frac{1}{p}\right) = \prod_{j=1}^J \left(\omega_{\mathcal{P}_j}(m) + V_j\right) \leqslant \left(\frac{\omega_{\mathcal{P}}(m)}{J} + V\right)^J.$$

Since $\omega_{\mathcal{P}}(m) \leqslant 2JV$ by definition of I_N^{typ}, the claimed bound follows. □

Intuitively, Lemma 2.4.11 indicates that the terms $D_{Y,u}$ and $O_{Y,u}$ in the Ihara-Bass formula (2.13) should constitute only a small perturbation to the characteristic polynomial of A_Y, once we restrict to I_N^{typ}. Moreover, this restriction should only affect a few eigenvalues, since the condition $\omega_{\mathcal{P}}(n) \leqslant 2JV$ is satisfied by most integers $n \in I_N$.

To make these ideas precise, we employ a topological argument together with Cauchy's interlacing theorem. The following lemma will be used to handle eigenvalues occurring with multiplicity greater than one.

Lemma 2.4.12. *Let $P \in \mathbb{C}[X]$ be a non-zero polynomial. Suppose that, for $x \in [0,1]$,*

$$P(x) = f_1(x)f_2(x)\cdots f_n(x)g(x)$$

where $f_i : [0,1] \to \mathbb{C}$ are Lipschitz continuous functions and $g : [0,1] \to \mathbb{C}$ is a continuous function. If each f_i has a root in $(0,1)$, then P has at least n roots in $(0,1)$, counting multiplicities.

Proof. Let $x_0 \in (0,1)$ be a point such that $f_i(x_0) = 0$ for all i in a subset $I \subset \{1,\dots,n\}$. It suffices to show that $P^{(k)}(x_0) = 0$ for all $0 \leqslant k < |I|$. We prove this by induction on k. Suppose that $P(x_0) = P'(x_0) = \cdots = P^{(k-1)}(x_0) = 0$ for some $0 \leqslant k < |I|$. By Taylor expansion,

$$P^{(k)}(x_0) = k! \lim_{x\to x_0} \frac{P(x)}{(x-x_0)^k} = k!g(x_0)\prod_{i\notin I} f_i(x_0) \cdot \lim_{x\to x_0}(x-x_0)^{|I|-k}\prod_{i\in I}\frac{f_i(x)}{x-x_0}.$$

Since each f_i is Lipschitz continuous, the last product remains bounded as $x \to x_0$, and thus $P^{(k)}(x_0) = 0$ as desired. □

Using Lemmas 2.4.11 and 2.4.12, we can now relate the eigenvalues of the restricted adjacency matrix $A_Y|_{I_N^{\mathrm{typ}}}$ to those of the non-backtracking matrix M_Y via the Ihara-Bass formula.

To do this, we require two classical facts on eigenvalues of Hermitian matrices. For any Hermitian matrix $M \in M_n(\mathbb{C})$, write $\lambda_1(M) \geqslant \ldots \geqslant \lambda_n(M)$ for the eigenvalues of M in descending order.

Lemma 2.4.13 (Weyl inequalities)**.** *Let $A, B \in M_n(\mathbb{C})$ be Hermitian matrices. Then*

$$\lambda_{i+h}(A) + \lambda_{n-h}(B) \leqslant \lambda_i(A+B) \leqslant \lambda_{i-h}(A) + \lambda_{h+1}(B)$$

for all $1 \leqslant i \leqslant n$ and $0 \leqslant h \leqslant \min(i-1, n-i)$.

Proof. This is [7, Theorem III.2.1]. □

Lemma 2.4.14 (Cauchy's interlacing theorem)**.** *Let $A \in M_n(\mathbb{C})$ be a Hermitian matrix and let B be the matrix obtained by deleting r rows of A and the r corresponding columns, for some $r < n$. Then*

$$\lambda_{i+r}(A) \leqslant \lambda_i(B) \leqslant \lambda_i(A)$$

for all $1 \leqslant i \leqslant n - r$.

Proof. This is [7, Corollary III.1.5]. □

Lemma 2.4.15. *Let $C \geqslant 1$ be a sufficiently large constant. Let $Y \subset \mathbb{Z}$ be any set and let $k \geqslant 1$.*

If the restricted adjacency matrix $A_Y|_{I_N^{\mathrm{typ}}}$ has at least k eigenvalues larger than $2C^J V^{J/2}$ in absolute value, then the non-backtracking matrix M_Y must have at least k eigenvalues larger than $C^J V^{J/2}$ in absolute value.

Proof. Since $A_Y|_{I_N^{\mathrm{typ}}}$ is real and symmetric, its eigenvalues are real. We treat the case of positive eigenvalues, the negative case being similar.

Suppose that $A_Y|_{I_N^{\mathrm{typ}}}$ has at least k positive eigenvalues larger than $2C^J V^{J/2}$. By the Weyl inequalities (Lemma 2.4.13), the k largest eigenvalues of the perturbed matrix

$$\big(A_Y - u(D_{Y,u} - O_{Y,u})\big)\big|_{I_N^{\mathrm{typ}}}$$

are bounded below by

$$2C^J V^{J/2} - u\big\|D_{Y,u}|_{I_N^{\mathrm{typ}}}\big\|_{\mathrm{op}} - u\big\|O_{Y,u}|_{I_N^{\mathrm{typ}}}\big\|_{\mathrm{op}}. \tag{2.14}$$

The operator norm estimate, Lemma 2.4.11, ensures that the expression (2.14) is greater than $C^J V^{J/2}$ for $u \in [-V^{-J/2}, V^{-J/2}]$, provided C is sufficiently large. Hence, by Cauchy's interlacing theorem (Lemma 2.4.14), the unrestricted matrix

$$A_Y - u(D_{Y,u} - O_{Y,u})$$

has at least k eigenvalues greater than $C^J V^{J/2}$, for every $u \in [-V^{-J/2}, V^{-J/2}]$.

Let $\mathcal{H}_N$ denote the space of $N \times N$ Hermitian matrices, equipped with the operator norm. For $A \in \mathcal{H}_N$, let $\lambda_1(A) \geqslant \ldots \geqslant \lambda_N(A)$ be the eigenvalues of A. The Weyl inequalities (Lemma 2.4.13) imply that each λ_i is a 1-Lipschitz function on $\mathcal{H}_N$. Since the map $u \mapsto A_Y - u(D_{Y,u} - O_{Y,u}) \in \mathcal{H}_N$ is smooth over $-1 < u < 1$ (as $|w_Y| \leqslant 1$), the function

$$f_i(u) := 1 - u\lambda_i\big(A_Y - u(D_{Y,u} - O_{Y,u})\big)$$

is Lipschitz on $[-1/2, 1/2]$, for every $1 \leqslant i \leqslant N$.

For each i, we have $f_i(0) = 1$. Moreover, for $1 \leqslant i \leqslant k$, we have $f_i(C^{-J}V^{-J/2}) < 0$ by our lower bound on the k largest eigenvalues of $A_Y - u(D_{Y,u} - O_{Y,u})$. Therefore, the functions $f_1, \dots, f_k$ each have a root in the interval $(0, C^{-J}V^{-J/2})$.

By the Ihara-Bass formula (Proposition 2.4.5), for $|u| < 1$ we have the identity

$$\det(I - uM_Y) = \prod_{i=1}^{N} f_i(u) \cdot g(u),$$

where g is a continuous function that can be written down explicitly. Applying Lemma 2.4.12, we conclude that $\det(I - uM_Y)$ has at least k roots in $(0, C^{-J}V^{-J/2})$, counting multiplicities. In other words, M_Y has at least k eigenvalues greater than $C^J V^{J/2}$, as desired. □

2.4.4 Eigenvalues of localised matrices

A key feature of our adjacency matrix A_Y is that its non-zero entries lie close to the diagonal. The next lemma states that, if such a Hermitian matrix has a very small number of large eigenvalues, it is possible to delete a small number of rows and columns to obtain a matrix without any large eigenvalue.

Lemma 2.4.16. *Let $N \geqslant 10H^3$. Let $A = (a_{m,n})_{m,n \in I_N}$ be a Hermitian matrix such that $a_{m,n} = 0$ whenever $|m - n| \geqslant H$.*

Let $\alpha > 0$, and suppose that A has $\leqslant N/H^3$ eigenvalues with absolute value $\geqslant \alpha$. Then, there is a subset $E \subset I_N$ with $|I_N \setminus E| \ll N/H$ such that every eigenvalue of $A|_E$ has absolute value $\leqslant \alpha$.

Proof. Consider a sequence of disjoint discrete intervals

$$B_1, \dots, B_q \subset I_N,$$

each of length $|B_i| = H^2$. We define this sequence such that for each i, B_i is separated from B_{i+1} by a gap of size exactly H. Choosing q to be maximal with these properties, we have $q \asymp N/H^2$ and

$$\left| I_N \setminus \bigsqcup_{i=1}^{q} B_i \right| \ll qH + H^2 \ll N/H. \tag{2.15}$$

Since $B_1, \dots, B_q$ have pairwise distance $\geqslant H$, the property of A in the statement implies that the principal submatrix $A' := A|_{\bigsqcup_{1 \leqslant i \leqslant q} B_i}$ is block-diagonal, with blocks $(A|_{B_i})_{1 \leqslant i \leqslant q}$.

Let $\mathcal{I}$ be the set of all indices i such that $A|_{B_i}$ has an eigenvalue whose absolute value is $> \alpha$. Then, defining

$$E := \bigsqcup_{\substack{1 \leqslant i \leqslant q \\ i \notin \mathcal{I}}} B_i,$$

the matrix $A|_E$ has no eigenvalue with absolute value $> \alpha$.

By (2.15), we have

$$|I_N \setminus E| \ll |\mathcal{I}|H^2 + N/H.$$

Observe moreover that $|\mathcal{I}| \leqslant N/H^3$. Indeed, A has $\leqslant N/H^3$ eigenvalues with absolute value $\geqslant \alpha$, so by Cauchy's interlacing theorem (Lemma 2.4.14) the same holds for its principal submatrix A'. This implies the desired bound $|I_N \setminus E| \ll N/H$. □

We can now prove that Proposition 2.4.7 implies Proposition 2.3.4. We remark that Lemma 2.4.16 will not be applied directly to A_Y, but rather to its restriction to I_N^{typ}. This is because A_Y admits large eigenvalues arising from integers outside I_N^{typ} (see Remark 2.3.5), in greater number than allowed by Lemma 2.4.16.

Proof of Proposition 2.3.4, assuming Proposition 2.4.7. Let $Y \subset \mathbb{Z}$ be the set provided by Proposition 2.4.7. In particular,

$$|I_N \setminus Y| \ll H_0^{-1/3} N. \tag{2.16}$$

In addition, all but $\leqslant N/H^3$ eigenvalues of the non-backtracking matrix M_Y have absolute value at most $e^{O(J)}V^{J/2}$. By Lemma 2.4.15, the same holds for the restricted adjacency matrix $A_Y|_{I_N^{\mathrm{typ}}}$.

Applying Lemma 2.4.16 to $A_Y|_{I_N^{\mathrm{typ}}}$,[7] we obtain a subset $E \subset I_N$ with

$$|I_N \setminus E| \ll N/H \tag{2.17}$$

such that all eigenvalues of $A_Y|_{I_N^{\mathrm{typ}} \cap E}$ are $\leqslant e^{O(J)}V^{J/2}$. In other words, defining $X := I_N^{\mathrm{typ}} \cap E \cap Y$, all eigenvalues of $A_{\mathbb{Z}}|_X$ are $\leqslant e^{O(J)}V^{J/2}$.

By [45, Eq. (1.11)], we have

$$\left|I_N \setminus I_N^{\mathrm{typ}}\right| = \sum_{n \in I_N} \mathbf{1}_{\omega_{\mathcal{P}}(n) > 2JV} \ll e^{-JV/3} N. \tag{2.18}$$

Combining (2.16) to (2.18), we conclude that

$$|I_N \setminus X| \ll \left(H_0^{-1/3} + H^{-1} + e^{-JV/3}\right) N \ll e^{-JV/3} N,$$

as $J, V \ll \log\log H \asymp \log\log H_0$ by part (c) of Lemma 2.2.3. This proves Proposition 2.3.4. □

[7] Strictly speaking, Lemma 2.4.16 is applied to the matrix obtained from A_Y by setting the entries with indices outside I_N^{typ} to zero.

2.5 High trace bound for the non-backtracking operator

The high trace method is a standard technique to control the eigenvalues of square matrices: if M is Hermitian and $\mathrm{Tr}(M^k) \leqslant C$ for some $k \in 2\mathbb{N}$, then every eigenvalue of M satisfies $|\lambda_i(M)| \leqslant C^{1/k}$, because $\mathrm{Tr}(M^k)$ is the k-th moment of the eigenvalues of M. A similar idea can be used to control the tail of the spectrum of general square matrices.

Lemma 2.5.1 (Schur's inequality)**.** *Let $A \in M_n(\mathbb{C})$ be a matrix with eigenvalues $\lambda_1(A), \dots, \lambda_n(A)$. Then*

$$\sum_{i=1}^{n} |\lambda_i(A)|^2 \leqslant \mathrm{Tr}(A^*A). \tag{2.19}$$

Proof. Observe that $\mathrm{Tr}(A^*A) = \sum_{1 \leqslant i,j \leqslant n} |a_{ij}|^2$, where (a_{ij}) are the entries of A. As a result, the inequality (2.19) holds when A is triangular. Since every square matrix is unitarily similar to a triangular matrix, the general case follows. □

2.5.1 Statement of the high trace bound

In Proposition 2.5.3, we state a suitable high trace bound for the non-backtracking operator M_Y. We then deduce the spectral tail bound (Proposition 2.4.7) from it. To effectively bound the number of large eigenvalues of M_Y, the power k in the high trace method must be sufficiently large. In our setting, this requires taking a power that is logarithmic in H.

Notation 2.5.2. We define $K := \lfloor \log H \rfloor$.

Proposition 2.5.3 (High trace bound for M_Y)**.** *Let N be an integer such that $\log N \geqslant (\log H)^{2.001}$.*

Then, there exists a set $Y \subset \mathbb{Z}$ with $|I_N \setminus Y| \ll H_0^{-1/3} N$, such that the non-backtracking operator M_Y associated to the weighted graph G_Y satisfies the high trace bound

$$\mathrm{Tr}\big[\big((M_Y)^K\big)^* (M_Y)^K\big] \leqslant e^{O(JK)} V^{JK} N. \tag{2.20}$$

Proof of Proposition 2.4.7, assuming Proposition 2.5.3. Let $Y \subset \mathbb{Z}$ be the set given in Proposition 2.5.3. Let $(\lambda_i)_{i \in I}$ be the (complex) eigenvalues of M_Y. By the high trace bound (2.20) and Schur's inequality (Lemma 2.5.1), these satisfy

$$\sum_{i \in I} |\lambda_i|^{2K} \leqslant \mathrm{Tr}\big[\big((M_Y)^K\big)^* (M_Y)^K\big] \leqslant e^{O(JK)} V^{JK} N,$$

using that the eigenvalues of $(M_Y)^K$ are the K-th powers of the eigenvalues of M_Y. By Markov's inequality, this implies that the number of eigenvalues λ_i with $|\lambda_i| \geqslant C^J V^{J/2}$ is at most

$$e^{O(JK)} C^{-2JK} N.$$

Since $K \asymp \log H$, this number can be made smaller than N/H^3, provided that C is chosen to be a sufficiently large absolute constant. □

The rest of this chapter is devoted to the proof of Proposition 2.5.3. Henceforth, we fix an integer N such that

$$\log N \geqslant (\log H)^{2.001}. \tag{2.21}$$

In particular, we have the useful bound

$$H^{JK} \ll_\varepsilon N^\varepsilon, \tag{2.22}$$

which follows from (2.21) and the estimates $J \leqslant \log\log H$ and $K \asymp \log H$.

2.5.2 Expanding the trace as a sum over walks

To prove Proposition 2.5.3, we work on the physical side, computing the trace directly by expanding the matrix product.

Lemma 2.5.4. *Let $Y \subset \mathbb{Z}$, and let M_Y be the non-backtracking operator associated to G_Y. Then*

$$\mathrm{Tr}\big[\big((M_Y)^K\big)^* (M_Y)^K\big] \ll H \sum_{\substack{d_1,\dots,d_{2K} \in \pm\mathcal{D} \\ d_{i+1} = -d_i \iff i = K}} \Bigg| \sum_{\substack{n \in I_N \\ n + b_i \in Y \text{ for } 0 \leqslant i \leqslant 2K}} \prod_{i=1}^{2K} \prod_{p \mid d_i} \Big(\mathbf{1}_{p \mid n + b_i} - \frac{1}{p} \Big) \Bigg| + N, \tag{2.23}$$

where $b_i := \sum_{1 \leqslant k \leqslant i} d_k$ for $0 \leqslant i \leqslant 2K$.

Proof. We use the formula for $\mathrm{Tr}[(M^K)^* M^K]$ given in Lemma A.2.2, which applies to the non-backtracking operator M of a general weighted graph. This gives

$$\mathrm{Tr}\big[\big((M_Y)^K\big)^* (M_Y)^K\big] = \sum_{\substack{n_{-1}, n_0, n_1, \dots, n_{2K} \in I_N \\ (n_{-1}, n_0, \dots, n_K) \in \mathrm{nbw}(G_Y) \\ (n_K, n_{K+1}, \dots, n_{2K}, n_{-1}) \in \mathrm{nbw}(G_Y) \\ (n_0, n_{K-1}) = (n_{2K}, n_{K+1})}} \prod_{i=1}^{2K} w_Y(n_{i-1}, n_i), \tag{2.24}$$

where $\mathrm{nbw}(G_Y)$ denotes the set of non-backtracking walks in G_Y (see Definition A.2.1).

We rewrite this expression in terms of n_0 and the increments $d_i := n_i - n_{i-1}$ for $0 \leqslant i \leqslant 2K$. Since we only need an upper bound, we can apply the triangle inequality and keep only the sum over n_0 inside the absolute value:

$$\mathrm{Tr}\big[\big((M_Y)^K\big)^*(M_Y)^K\big] \leqslant \sum_{\substack{d_0,\dots,d_{2K}\in\pm\mathcal{D}\\ d_{i+1}\neq -d_i \text{ for } 0\leqslant i<K\\ d_{i+1}\neq -d_i \text{ for } K<i<2K\\ d_{K+1}=-d_K}} \Bigg| \sum_{\substack{n_0\in I_N\\ n_0-d_0\in I_N}} \prod_{i=1}^{2K} w_Y\Big(n_0+\sum_{j=1}^{i-1} d_j,\, n_0+\sum_{j=1}^{i} d_j\Big)\Bigg|, \tag{2.25}$$

with the convention that $w_Y(m,n) = 0$ if $m \notin I_N$ or $n \notin I_N$. In passing from (2.24) to (2.25), we have dropped the conditions $n_0 = n_{2K}$ and $n_{2K-1} \neq n_{-1}$, which can be expressed in terms of the increments d_i alone (independently of n_0).

If $n_0 \in I_N$, the requirement that $n_0 - d_0 \in I_N$ is automatically satisfied unless n_0 lies within distance $O(H)$ of the boundary of I_N. Likewise, for each $1 \leqslant i \leqslant 2K$, the implicit condition $n_0 + \sum_{j=1}^{i} d_j \in I_N$ can only fail if n_0 lies within distance $O(KH)$ of the boundary. Discarding all these conditions trivially, we incur an error of size $\ll KH(2H)^{2K+1}$, which is at most $O(N)$ by (2.22).

Expanding the formula for w_Y (see Definition 2.3.2), the trace (2.25) is thus bounded by

$$\sum_{\substack{d_0,\dots,d_{2K}\in\pm\mathcal{D}\\ d_{i+1}=-d_i \iff i=K}} \Bigg| \sum_{n_0\in I_N} \prod_{i=1}^{2K} \prod_{p\mid d_i} \Big(\mathbf{1}_{p\mid n_0+\sum_{j=1}^{i-1} d_j} - \frac{1}{p}\Big) \prod_{i=0}^{2K} \mathbf{1}_Y\Big(n_0 + \sum_{j=1}^{i} d_j\Big)\Bigg| + O(N).$$

Notice that the summand is independent of the variable d_0. Since d_0 ranges over $\ll |\mathcal{D}| \ll H$ values, the lemma follows. □

2.6 Overview of the proof of the high trace bound

This section serves purely as motivation and is separate from the actual proof. Our aim here is to outline the strategy for bounding the right-hand side of (2.23). We will also explain why the two technical assumptions that appear there – the non-backtracking condition $d_{i+1} \neq -d_i$ (for almost all i) and the restriction of the vertices $n + b_i$ to a well-chosen set Y – are essential to our arguments.

Let us first examine the situation without these hypotheses. For each pair $(i,j) \in [\![2K]\!] \times [\![J]\!]$, let d_{ij} be the unique prime in $\mathcal{P}_j$ dividing d_i. Ignoring the two technical hypotheses, the right-hand side of (2.23) essentially reduces to

$$\sum_{d_1,\dots,d_{2K}\in\pm\mathcal{D}} \Bigg| \sum_{n\in I_N} \prod_{i=1}^{2K} \prod_{j=1}^{J} \Big(\mathbf{1}_{d_{ij}\mid n+b_i} - \frac{1}{d_{ij}}\Big)\Bigg|. \tag{2.26}$$

The basic idea is that the factors $\mathbf{1}_{d_{ij}|n+b_i} - \frac{1}{d_{ij}}$ should generate significant cancellation in the inner sum because, for a random integer $n \in I_N$, the event $d_{ij} \mid n+b_i$ occurs with probability almost exactly $1/d_{ij}$. One must be careful, however: these factors are not independent whenever they involve the same prime, and these dependencies can prevent cancellation.

With this in mind, we call a prime $p \in \mathcal{P}$ a *single prime* for the tuple $(d_1, \dots, d_{2K})$ if there is exactly one pair $(i,j) \in [\![2K]\!] \times [\![J]\!]$ such that $p = d_{ij}$. If a tuple has at least one single prime, then the inner sum in (2.26) exhibits almost perfect cancellation, by the Chinese Remainder Theorem (since $N > d_1 \cdots d_{2K}$). Thus, we only need to consider the tuples for which every prime appearing in the array (d_{ij}) is repeated at least twice.

In this situation, we cannot hope to obtain cancellation. We may therefore use the triangle inequality and trivially bound $|\mathbf{1}_{d_{ij}|n+b_i} - \frac{1}{d_{ij}}| \leqslant \mathbf{1}_{d_{ij}|n+b_i} + \frac{1}{d_{ij}}$. Expanding the resulting product, we are reduced to bounding

$$\sum_{\substack{d_1,\dots,d_{2K} \in \pm\mathcal{D} \\ p \mid \prod d_i \Rightarrow p^2 \mid \prod d_i}} \sum_{n \in I_N} \prod_{(i,j)\in\mathcal{L}} \mathbf{1}_{d_{ij}|n+b_i} \prod_{(i,j)\in\mathcal{U}} \frac{1}{d_{ij}}, \tag{2.27}$$

for every partition $[\![2K]\!] \times [\![J]\!] = \mathcal{L} \sqcup \mathcal{U}$ of the index set into two sets $\mathcal{L}$ and $\mathcal{U}$. The pairs $(i,j) \in \mathcal{L}$ are called *lit indices*, while those in $\mathcal{U}$ are called *unlit indices*.

Heuristically, we might expect each factor $\mathbf{1}_{d_{ij}|n+b_i}$ to behave like $1/d_{ij}$ on average, in which case we would obtain a strong bound from the fact that each prime occurs at least twice. Indeed, the common contribution of all factors corresponding to a repeated prime p would then be at most $O(1/p^2)$ on average, which is very small since all primes in $\mathcal{P}$ are larger than H_0. A more precise implementation of this idea shows that the expression (2.27) is negligible whenever the set $\mathcal{U}$ of unlit indices is somewhat large. We may thus assume that nearly all pairs $(i,j) \in [\![2K]\!] \times [\![J]\!]$ lie in $\mathcal{L}$.

The 'enemy scenario' is therefore when the factors $\mathbf{1}_{d_{ij}|n+b_i}$ conspire so that their product is, on average over $d_1, \dots, d_{2K}$ and n, significantly larger than the product of $1/d_{ij}$. To rule it out, we must analyse the joint behaviour of these indicator functions.

Recall that $b_i := \sum_{1\leqslant k\leqslant i} d_k$. If a prime p is repeated twice at lit indices, say $p = d_{i_1 j} = d_{i_2 j}$ with $i_1 < i_2$, then the congruences $p \mid n + b_{i_1}$ and $p \mid n + b_{i_2}$ imply that

$$p \,\Big|\, \sum_{i_1<k<i_2} d_k = \sum_{i_1<k<i_2} \sigma_k \prod_{j=1}^{J} d_{kj}, \tag{2.28}$$

where $\sigma_k \in \{\pm 1\}$ is the sign of d_k. By considering all pairs of lit indices corresponding to each repeated prime, we obtain a large system of such divisibility relations. Intuitively, this system of constraints should considerably limit the number of admissible choices for

the primes d_{ij}, and hence the number of tuples $(d_1, \dots, d_{2K})$ contributing significantly to (2.27).[8]

The shape of the divisibility constraints is determined by the partition of the index set $[\![2K]\!] \times [\![J]\!]$ according to repetitions of the same prime, and the signs σ_k. Therefore, to organise our analysis, we first fix the shape, and then consider the number of possible assignments of prime values to the variables which satisfy that system of divisibility constraints.[9] The number of possible shapes is very large (more than exponential in JK). We might hope that this is balanced by there being very few assignments that satisfy the associated divisibility system, and therefore that the total contribution is acceptably small.

Unfortunately, certain shapes do yield systems with many solutions, so we cannot expect a strong uniform bound that applies to every shape. Our two-step strategy is therefore as follows. First, obtain a good bound that holds for most shapes. Second, show that the remaining shapes are rare enough that their total contribution is acceptable.

It is not necessary to obtain an optimal bound for the number of solutions to divisibility systems. Roughly speaking, it suffices to show that the system determines the values of at least $K^{0.01}$ of the primes d_{ij} based on the values of the remaining primes. We do so by extracting a reasonably large 'triangular' subsystem of divisibility constraints which is clearly non-degenerate, via a careful combinatorial analysis.

This combinatorial argument, however, fails for certain degenerate shapes. This failure becomes critical because some of these shapes are not sufficiently rare to be handled by the second step of our strategy. Our two technical assumptions are introduced precisely to preclude these problematic configurations.

The first such issue arises from backtracking walks. When $d_{i+1} = -d_i$ for many i, many sums of the form $\sum_{i_1<k<i_2} d_k$ vanish, rendering the corresponding divisibility constraints (2.28) trivial. Moreover, the contribution of these backtracking walks to the trace is genuinely excessive (reflecting the existence of large eigenvalues of the adjacency operator), meaning that the simplistic expression (2.26) cannot possibly satisfy the desired bound. This explains the use of the non-backtracking operator throughout, which eliminates this class of degenerate shapes from the outset.

[8] Note that the techniques of Helfgott and Radziwiłł [26] cannot be applied here as the right-hand side of (2.28) is a polynomial expression in the primes d_{ij}, rather than a linear combination with small coefficients, as soon as $J > 1$.

[9] That is, if we have a partition of $[\![2K]\!] \times [\![J]\!]$, we look for values d_{ij} satisfying the constraints

$$d_{i_1 j_1} = d_{i_2 j_2} \mid \sum_{i_1<k<i_2} d_k$$

for all $(i_1, j_1), (i_2, j_2) \in \mathcal{L}$ in the same part of the partition.

The second class of problematic shapes roughly corresponds to walks that involve too few distinct primes. In this case, we can only extract a small triangular subsystem, which is insufficient to compensate for the large number of combinatorial arrangements of those primes into long walks. To discard these shapes, we restrict the vertices $n + b_i$ to lie in a well-chosen set Y. The set Y (defined precisely in Section 2.7) is chosen so that every such pathological walk visits at least one vertex outside of Y. Although working with Y introduces significant technicalities,[10] this device removes the second type of degeneracies that our combinatorial methods cannot handle.

With the non-backtracking condition and the restriction to Y in place, a large triangular subsystem can be extracted for all shapes satisfying a generic *unpredictability* assumption. The first step of the strategy is then complete: the contribution from these unpredictable shapes is negligible.

Finally, we analyse the remaining *predictable* shapes, for which no large triangular subsystem can be found. We show that, while these shapes yield the main term in the trace, they are sufficiently rare that their contribution is admissible, which concludes the second step.

In summary, we use cancellation to show that walks containing (a sufficient number of) single primes contribute negligibly to the trace; this is done in Section 2.9. The only remaining walks that are not easily handled are those with many repeated primes at lit indices. For these walks, no cancellation is available. Instead, we analyse the associated system of divisibility constraints. For most shapes of such systems (the *unpredictable* case, treated in Section 2.12), we can extract a sufficiently large triangular subsystem, which yields enough savings to render their contribution negligible. Our ability to perform this step for most shapes relies crucially on the non-backtracking condition and the restriction to the set Y, which remove the worst degeneracies. The residual walks (the *predictable* case, treated in Section 2.10) are sufficiently rare that their total contribution is also suitably bounded. Putting all these estimates together yields the desired high trace bound, which is essentially sharp. The remainder of the chapter is devoted to turning this sketch into a rigorous argument.

2.7 Definition of the set Y_L

We finally define a set $Y \subset \mathbb{Z}$ satisfying Proposition 2.5.3: our specific choice is denoted by Y_L. Note that this definition will not be used in earnest until Section 2.12.

The role of Y_L is to exclude certain tuples $(d_1, \dots, d_{2K})$ from the expression in Lemma 2.5.4 that our methods cannot handle. The definition of Y_L is quite delicate:

[10] Notably, using single primes to obtain cancellation over Y rather than over I_N requires a highly technical analysis based on a *combinatorial sieve* for composite moduli.

even a slight modification can disrupt key parts of the argument and render the high trace bound intractable.

The set Y_L is obtained from $\mathbb{Z}$ by removing certain sparse arithmetic progressions, called *prohibited progressions*. These are defined in terms of *prohibited sequences*, which are sequences of integers that satisfy a special type of divisibility relation.

To define prohibited sequences, we first fix a maximum length parameter L. This parameter is fundamental to the definition of our set Y_L, which is why we retain it in the notation. It is chosen to be a power of K slightly smaller than one.

Notation 2.7.1. We define $L := K^{1-10\varepsilon_1}$.

Definition 2.7.2. Let $\ell \geqslant 1$. A sequence $(d_1, \dots, d_\ell) \in (\pm\mathcal{D})^\ell$ is said to be *reduced* if $d_{i+1} \neq -d_i$ for all $1 \leqslant i < \ell$.

Definition 2.7.3. A *prohibited sequence* is a reduced sequence $(d_1, \dots, d_\ell)$ of ℓ elements of $\pm\mathcal{D}$, for some $2 < \ell \leqslant L$, with the following properties:

1. (consecutiveness) for every prime q, the set $\{i \in [\![\ell]\!] : q \mid d_i\}$ is a discrete interval, and
2. (prohibited pattern) there is a prime p and some $1 < \ell_0 < \ell$ such that $p \mid d_1$, $p \nmid d_\ell$ and
$$p \Big| \sum_{\ell_0 \leqslant i \leqslant \ell} d_i. \tag{2.29}$$

A prohibited sequence $(d_1, \dots, d_\ell)$ is *primitive* if there is no consecutive[11] subsequence of $(d_1, \dots, d_\ell)$ or of $(d_\ell, \dots, d_1)$, of length $< \ell$, which is also prohibited.

A key difference with [26] is that, in their situation, the authors can restrict themselves to the case $\ell_0 = 1$. This is not possible here, and leads to additional complications in the proof of Lemma 2.9.6 (due to the fact that the constraint (2.29) only involves a subset of the prime factors of the d_i). We may now define the set Y_L.

Definition 2.7.4 ($\mathcal{Y}$, Y_L)**.** The *prohibited progression* associated to a primitive prohibited sequence $(d_1, \dots, d_\ell)$ is the set of all integers $n \in \mathbb{Z}$ such that
$$d_1 \mid n, \quad d_2 \mid n + d_1, \quad \dots, \quad d_\ell \mid n + d_1 + \cdots + d_{\ell-1}.$$
It is an arithmetic progression of square-free modulus $\operatorname{lcm}(d_1, \dots, d_\ell)$ (the consecutiveness property in Definition 2.7.3 ensures that it is non-empty).

[11] By 'consecutive subsequence of $(d_1, \dots, d_\ell)$', we mean a sequence of the form $(d_{k_1}, d_{k_1+1}, \dots, d_{k_2})$ for some $1 \leqslant k_1 < k_2 \leqslant \ell$.

Let $\mathcal{Y}$ be the set of all prohibited progressions associated with some primitive prohibited sequence. We define

$$Y_L := \mathbb{Z} \setminus \cup \mathcal{Y},$$

the set of all integers that do not belong to any prohibited progression.

To apply Proposition 2.5.3 with $Y = Y_L$, we need to show that Y_L covers almost all of I_N; specifically, that $|I_N \setminus Y_L| \ll H_0^{-1/3} N$. Although this is not difficult, we postpone the proof to Section 2.13 (see Lemma 2.13.2), where we establish a broader class of results concerning prohibited sequences and progressions.

2.8 Single and repeated primes

In this section, we introduce the notation that we will use to study the high trace of the non-backtracking matrix M_{Y_L} and make some simple observations.

2.8.1 Single, lit and unlit indices

The purpose of the following definitions is to characterise the tuples $\boldsymbol{d} = (d_1, \dots, d_{2K})$ that contribute to the right-hand side of (2.23). We will show in Lemma 2.8.6 that every such tuple belongs to one of the sets $\mathbf{D}_{\mathcal{S},\mathcal{L},r}$ introduced in Definition 2.8.4.

Definition 2.8.1 (d_{ij}, $\boldsymbol{d}_I$, $\boldsymbol{d}_*$)**.** Let $\boldsymbol{d} = (d_i)_{i \in [\![2K]\!]}$ be a vector with coordinates $d_i \in \pm\mathcal{D}$.

We write d_{ij} for the unique prime in $\mathcal{P}_j$ that divides d_i. Thus, $|d_i| = \prod_{j \in [\![J]\!]} d_{ij}$.

For any subset $I \subset [\![2K]\!] \times [\![J]\!]$, we set

$$\boldsymbol{d}_I := \prod_{(i,j) \in I} d_{ij}.$$

In the special case $I = [\![2K]\!] \times [\![J]\!]$, we will write $\boldsymbol{d}_*$ instead of $\boldsymbol{d}_{[\![2K]\!] \times [\![J]\!]}$ to shorten the notation.

Definition 2.8.2 (Single indices)**.** Let $\boldsymbol{d} = (d_i)_{i \in [\![2K]\!]}$ be a vector with coordinates $d_i \in \pm\mathcal{D}$.

We say that a pair $(i,j) \in [\![2K]\!] \times [\![J]\!]$ is a *single index* if $d_{ij}^2 \nmid \boldsymbol{d}_*$, i.e. the prime d_{ij} does not appear at any pair other than (i,j). In that case, we also say that d_{ij} is a *single prime* of $\boldsymbol{d}$.

Definition 2.8.3. Let $R \geqslant 1$ and let $\mathcal{L} \subset [\![R]\!] \times [\![J]\!]$.

A vector $\boldsymbol{d} \in (\pm\mathcal{D})^R$ is called *$\mathcal{L}$-admissible* if, for all $1 \leqslant k_1 < k_2 \leqslant R$ such that $[\![k_1, k_2]\!] \times [\![J]\!] \subset \mathcal{L}$, neither $(d_{k_1}, d_{k_1+1}, \dots, d_{k_2})$ nor $(d_{k_2}, d_{k_2-1}, \dots, d_{k_1})$ is a prohibited sequence (see Definition 2.7.3).[12]

[12] Recall that a prohibited sequence has length $\leqslant L$ by definition, so this condition is empty when $k_2 - k_1 + 1 > L$.

We now define the sets $\mathbf{D}_{\mathcal{S},\mathcal{L},r}$. The tuples $\boldsymbol{d} = (d_1, \dots, d_{2K})$ arising from the trace expansion in Lemma 2.5.4 are non-reduced at exactly one position, since $d_{K+1} = -d_K$ while $d_{i+1} \neq -d_i$ for $i \neq K$. This structure is a consequence of multiplying the K-th power of the non-backtracking matrix by its conjugate transpose, a necessary step for controlling the eigenvalues. While this isolated backtracking requires technical adjustments (in parts (3) and (4) of Definition 2.8.4), the resulting complexity is largely notational rather than conceptual.

Definition 2.8.4 ($\mathbf{D}_{\mathcal{S},\mathcal{L},r}$)**.** Let $\mathcal{S}$ and $\mathcal{L}$ be disjoint subsets of $[\![2K]\!] \times [\![J]\!]$ and let $1 \leqslant r \leqslant K$.

We define $\mathbf{D}_{\mathcal{S},\mathcal{L},r}$ to be the set of all vectors $\boldsymbol{d} = (d_i)_{i \in [\![2K]\!]}$ with coordinates $d_i \in \pm\mathcal{D}$ having the following properties.

1. The set $\mathcal{S}$ is precisely the set of single indices for $\boldsymbol{d}$.
2. Whenever two indices $(i,j), (i',j) \in \mathcal{L}$ are such that $d_{ij} = d_{i'j}$, we have
$$d_{ij} \,\Big|\, \sum_{i<k<i'} d_k.$$
3. The integer r satisfies $(d_{K+1}, d_{K+2}, \dots d_{K+r}) = -(d_K, d_{K-1}, \dots, d_{K-r+1})$.
4. Let
$$\begin{cases} I^{(1)} := [\![1, K]\!] \\ I^{(2)} := [\![K+1, 2K]\!] \\ I^{(3)} := [\![1, K-r]\!] \cup [\![K+r+1, 2K]\!], \end{cases}$$
and let $\iota^{(k)} : I^{(k)} \to \{1, \dots, |I^{(k)}|\}$ be the unique order-preserving bijection. Then, for $k = 1, 2, 3$, the vector
$$\boldsymbol{d}^{(k)} := (d_i)_{i \in I^{(k)}}$$
is reduced[13] and $\mathcal{L}^{(k)}$-admissible, where $\mathcal{L}^{(k)} := \{(\iota^{(k)}(i), j) \mid (i,j) \in \mathcal{L}\}$.

Note that if $r = K$, then $I^{(3)} = \emptyset$ and the conditions on $\boldsymbol{d}^{(3)}$ can be ignored.

Our goal is to bound the expression in Lemma 2.5.4, which includes a double sum over $\boldsymbol{d} \in (\pm\mathcal{D})^{2K}$ and $n \in I_N$. Lemma 2.8.6 below reduces this problem to analysing the contribution of the pairs $(\boldsymbol{d}, n)$ such that

- $\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}$,
- $d_{ij} \mid n + b_i$ for $(i,j) \in \mathcal{L}$, and
- $d_{ij} \nmid n + b_i$ for $(i,j) \in \mathcal{U}$,

[13] See Definition 2.7.2 for the definition of reduced tuples.

where $\mathcal{S}$, $\mathcal{L}$, and $\mathcal{U}$ are sets such that $\mathcal{S} \sqcup \mathcal{L} \sqcup \mathcal{U} = [\![2K]\!] \times [\![J]\!]$, and $1 \leqslant r \leqslant K$. Thus, $\mathcal{S}$ is the set of single indices for $\boldsymbol{d}$, and the complement $([\![2K]\!] \times [\![J]\!]) \setminus \mathcal{S}$ is divided into two sets according to whether the corresponding divisibility condition holds or not. Following the terminology of Helfgott and Radziwiłł [26], we call $\mathcal{L}$ the set of *lit indices* and $\mathcal{U}$ the set of *unlit indices*.

The contribution of these pairs $(\boldsymbol{d}, n)$ is captured by the quantity $E_{\mathcal{S},\mathcal{L},\mathcal{U},r}$ defined below.

Definition 2.8.5 ($E_{\mathcal{S},\mathcal{L},\mathcal{U},r}$)**.** Let $\mathcal{S}$, $\mathcal{L}$, and $\mathcal{U}$ be sets such that $\mathcal{S} \sqcup \mathcal{L} \sqcup \mathcal{U} = [\![2K]\!] \times [\![J]\!]$ and let $1 \leqslant r \leqslant K$.

We write $E_{\mathcal{S},\mathcal{L},\mathcal{U},r}$ for the quantity

$$E_{\mathcal{S},\mathcal{L},\mathcal{U},r} := \sum_{\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}} \left| \sum_{\substack{n \in I_N \\ n+b_i \in Y_L \text{ for } 0 \leqslant i \leqslant 2K}} \mathbf{1}_{\substack{d_{ij} \mid n+\mathbf{b}_i \, \forall (i,j) \in \mathcal{L} \\ d_{ij} \nmid n+\mathbf{b}_i \, \forall (i,j) \in \mathcal{U}}} \prod_{(i,j) \in [\![2K]\!] \times [\![J]\!]} \left(\mathbf{1}_{d_{ij} \mid n+b_i} - \frac{1}{d_{ij}} \right) \right|, \tag{2.30}$$

where $b_i := \sum_{1 \leqslant k \leqslant i} d_k$ for $0 \leqslant i \leqslant 2K$.

Lemma 2.8.6. *We have*

$$\operatorname{Tr}\left[\left((M_Y)^K \right)^* (M_Y)^K \right] \ll e^{O(JK)} \max_{\substack{\mathcal{S} \sqcup \mathcal{L} \sqcup \mathcal{U} = [\![2K]\!] \times [\![J]\!] \\ 1 \leqslant r \leqslant K}} E_{\mathcal{S},\mathcal{L},\mathcal{U},r} \; + \; N.$$

Proof. For any $\boldsymbol{d} \in (\pm\mathcal{D})^{2K}$ with $d_{K+1} = -d_K$, let $r(\boldsymbol{d})$ be the largest integer $1 \leqslant r \leqslant K$ such that

$$(d_{K+1}, d_{K+2}, \dots d_{K+r}) = -(d_K, d_{K-1}, \dots, d_{K-r+1}).$$

By Lemma 2.5.4 and the triangle inequality, we have

$$\operatorname{Tr}\left[\left((M_Y)^K \right)^* (M_Y)^K \right] \ll H \sum_{\substack{\mathcal{S} \sqcup \mathcal{L} \sqcup \mathcal{U} = [\![2K]\!] \times [\![J]\!] \\ 1 \leqslant r \leqslant K}} \sum_{\substack{\boldsymbol{d} \in (\pm\mathcal{D})^{2K} \\ d_{i+1} = -d_i \iff i = K \\ d_{ij}^2 \mid \boldsymbol{d}_* \iff (i,j) \in \mathcal{S} \\ r(\boldsymbol{d}) = r}} |S_{\mathcal{S},\mathcal{L},\mathcal{U}}(\boldsymbol{d})| \; + \; N, \tag{2.31}$$

where $S_{\mathcal{S},\mathcal{L},\mathcal{U}}(\boldsymbol{d})$ denotes the inner sum on the right-hand side of (2.30), i.e.

$$S_{\mathcal{S},\mathcal{L},\mathcal{U}}(\boldsymbol{d}) := \sum_{\substack{n \in I_N \\ n+b_i \in Y_L \text{ for } 0 \leqslant i \leqslant 2K}} \mathbf{1}_{\substack{d_{ij} \mid n+\mathbf{b}_i \, \forall (i,j) \in \mathcal{L} \\ d_{ij} \nmid n+\mathbf{b}_i \, \forall (i,j) \in \mathcal{U}}} \prod_{(i,j) \in [\![2K]\!] \times [\![J]\!]} \left(\mathbf{1}_{d_{ij} \mid n+b_i} - \frac{1}{d_{ij}} \right).$$

Up to a multiplicative factor $e^{O(JK)}$, the sum over $\mathcal{S}$, $\mathcal{L}$, $\mathcal{U}$ and r in (2.31) can be replaced by a maximum. The factor H can be absorbed in this new $e^{O(JK)}$ term as $K \asymp \log H$.

Hence, to prove Lemma 2.8.6, it suffices to show that any vector $\boldsymbol{d} \in (\pm\mathcal{D})^{2K}$ satisfying the properties

(i) $d_{i+1} = -d_i \iff i = K$;

(ii) $d_{ij}^2 \nmid \boldsymbol{d}_* \iff (i,j) \in \mathcal{S}$;

(iii) $r(\boldsymbol{d}) = r$; and

(iv) $S_{\mathcal{S},\mathcal{L},\mathcal{U}}(\boldsymbol{d}) \neq 0$

is contained in $\mathbf{D}_{\mathcal{S},\mathcal{L},r}$.

Let $\boldsymbol{d}$ be such a vector. We need to show that it satisfies the four properties of Definition 2.8.4. Some are immediate: property (1) of Definition 2.8.4 is equivalent to (ii) and, by definition of $r(\boldsymbol{d})$, property (3) of Definition 2.8.4 follows from (iii). In addition, (iii) implies that

$$d_{K+r+1} \neq -d_{K-r} \tag{2.32}$$

in the case $r < K$. From (2.32) and (i), we deduce that the vectors $\boldsymbol{d}^{(1)}$, $\boldsymbol{d}^{(2)}$ and $\boldsymbol{d}^{(3)}$ from Definition 2.8.4 are reduced.

It only remains to prove property (2) of Definition 2.8.4 and the $\mathcal{L}^{(k)}$-admissibility of $\boldsymbol{d}^{(k)}$ for $k = 1, 2, 3$. This is where assumption (iv) is used.

Let $(i,j), (i',j) \in \mathcal{L}$ be such that $d_{ij} = d_{i'j}$, where we assume without loss of generality that $i < i'$. By (iv), there is at least one non-zero term in the sum defining $S_{\mathcal{S},\mathcal{L},\mathcal{U}}(\boldsymbol{d})$, which means that there is at least one $n \in I_N$ such that $d_{ij} \mid n + b_i$ and $d_{i'j} \mid n + b_{i'}$. Since $d_{ij} = d_{i'j}$, we can subtract these two divisibility relations to get $d_{ij} \mid b_{i'} - b_i$. Recalling that $b_l := \sum_{1 \leqslant k \leqslant l} d_k$, we get $d_{ij} \mid \sum_{i < k \leqslant i'} d_k$, which implies property (2) of Definition 2.8.4 as $d_{ij} = d_{i'j} \mid d_{i'}$.

Finally, suppose by contradiction that $\boldsymbol{d}^{(k)}$ is not $\mathcal{L}^{(k)}$-admissible, for some $k \in \{1,2,3\}$. Let $\tau : \{1, \dots, |I^{(k)}|\} \to I^{(k)}$ be the inverse of $\iota^{(k)}$. Then, there are $a, \ell \in \mathbb{N}$ and $\sigma \in \{\pm 1\}$ such that

- $\{\tau(a), \tau(a+\sigma), \dots, \tau(a+\ell\sigma)\} \times [\![J]\!] \subset \mathcal{L}$,[14] and
- $(d_{\tau(a)}, d_{\tau(a+\sigma)}, \dots, d_{\tau(a+\ell\sigma)})$ is a prohibited sequence.

By passing to a suitable subsequence, we can assume without loss of generality that this prohibited sequence is primitive. By (iv), there exists $n \in I_N$ such that the term in $S_{\mathcal{S},\mathcal{L},\mathcal{U}}(\boldsymbol{d})$ corresponding to n is non-zero. In particular, $n + b_i \in Y_L$ for all $0 \leqslant i \leqslant 2K$. Furthermore, we have $d_{ij} \mid n + b_i$ for all $(i,j) \in \mathcal{L}$, and thus

$$d_{\tau(a)} \mid n + b_{\tau(a)}, \quad d_{\tau(a+\sigma)} \mid n + b_{\tau(a+\sigma)}, \quad \dots, \quad d_{\tau(a+\ell\sigma)} \mid n + b_{\tau(a+\ell\sigma)}. \tag{2.33}$$

For any $1 < i \leqslant |I^{(k)}|$, we have

$$b_{\tau(i)} - b_{\tau(i-1)} = d_{\tau(i)}. \tag{2.34}$$

[14] In particular, $\{a, a+\sigma, \dots, a+\ell\sigma\} \subset \{1, \dots, |I^{(k)}|\}$.

While this is obvious for $k \in \{1,2\}$, it is also true for $k = 3$ because $b_{K-r} = b_{K+r}$ by property (3) of Definition 2.8.4, which we have already shown.

In the case $\sigma = 1$, using (2.34), we can rewrite (2.33) as

$$\begin{cases} d_{\tau(a)} \mid n + b_{\tau(a-1)} \\ d_{\tau(a+1)} \mid n + b_{\tau(a-1)} + d_{\tau(a)} \\ d_{\tau(a+2)} \mid n + b_{\tau(a-1)} + d_{\tau(a)} + d_{\tau(a+1)} \\ \vdots \\ d_{\tau(a+\ell)} \mid n + b_{\tau(a-1)} + d_{\tau(a)} + d_{\tau(a+1)} + \ldots + d_{\tau(a+\ell-1)}. \end{cases}$$

Since $(d_{\tau(a)}, d_{\tau(a+1)}, \ldots, d_{\tau(a+\ell)})$ is a primitive prohibited sequence, this means that $n + b_{\tau(a-1)}$ belongs to a prohibited progression (see Definition 2.7.4), contradicting the fact that $n + b_{\tau(a-1)} \in Y_L$.

Likewise, in the case $\sigma = -1$, we can use (2.34) to rewrite (2.33) as

$$\begin{cases} d_{\tau(a)} \mid n + b_{\tau(a)} \\ d_{\tau(a-1)} \mid n + b_{\tau(a)} - d_{\tau(a)} \\ d_{\tau(a-2)} \mid n + b_{\tau(a)} - d_{\tau(a)} - d_{\tau(a-1)} \\ \vdots \\ d_{\tau(a-\ell)} \mid n + b_{\tau(a)} - d_{\tau(a)} - d_{\tau(a-1)} - \ldots - d_{\tau(a-\ell+1)}. \end{cases}$$

Since $(d_{\tau(a)}, d_{\tau(a-1)}, \ldots, d_{\tau(a-\ell)})$ is a primitive prohibited sequence, so is its negative. Similarly to the previous case, these divisibilities imply that $n + b_{\tau(a)}$ belongs to a prohibited progression, contradicting the fact that $n + b_{\tau(a)} \in Y_L$. This concludes the proof that $\boldsymbol{d}^{(k)}$ is $\mathcal{L}^{(k)}$-admissible. □

2.8.2 Walks with many unlit indices

The next lemma shows that the expression $E_{\mathcal{S},\mathcal{L},\mathcal{U},r}$ from Definition 2.8.5 is negligible once the number of unlit indices exceeds a modest threshold, using that the primes in $\mathcal{P}$ are larger than H_0.

Lemma 2.8.7. *Let $\mathcal{S}$, $\mathcal{L}$, $\mathcal{U}$ be sets such that $[\![2K]\!] \times [\![J]\!] = \mathcal{S} \sqcup \mathcal{L} \sqcup \mathcal{U}$ and let $1 \leqslant r \leqslant K$. Suppose that $|\mathcal{U}| \geqslant K^{2\varepsilon_1}$. Then*

$$E_{\mathcal{S},\mathcal{L},\mathcal{U},r} \ll N.$$

To prove Lemma 2.8.7, we need the following basic estimate.

Lemma 2.8.8. *We have*

$$\sum_{\boldsymbol{d} \in (\pm\mathcal{D})^{2K}} \prod_{p \mid \boldsymbol{d}_*} \frac{1}{p} \ll K^{O(JK)}.$$

Proof. Any $\boldsymbol{d} \in (\pm\mathcal{D})^{2K}$ induces a partition of $[\![2K]\!] \times [\![J]\!]$, where (i,j) and (i',j') are in the same class if and only if $d_{ij} = d_{i'j'}$. Since the sets $\mathcal{P}_j$ are disjoint, this partition has the property that every class α is contained in $[\![2K]\!] \times \{j_\alpha\}$ for some $j_\alpha \in [\![J]\!]$.

Observe that $\boldsymbol{d}$ can be fully reconstructed from a sequence of $2K$ signs (the signs of the d_i), a partition of $[\![2K]\!] \times [\![J]\!]$ with the above property, and the assignment of a prime $p_\alpha \in \mathcal{P}_{j_\alpha}$ to every class α of this partition. Recall that $\sum_{p\in\mathcal{P}_j} 1/p = V_j \leqslant V$ for every j. Summing over all sequences of signs σ, suitable partitions Π of $[\![2K]\!] \times [\![J]\!]$ and primes $(p_\alpha)_{\alpha\in\Pi}$, we get

$$\sum_{\boldsymbol{d}\in(\pm\mathcal{D})^{2K}} \prod_{p\mid \boldsymbol{d}_*} \frac{1}{p} \leqslant \sum_{\sigma\in\{\pm1\}^{2K}} \sum_{\Pi} \prod_{\alpha\in\Pi} \sum_{p_\alpha\in\mathcal{P}_{j_\alpha}} \frac{1}{p_\alpha} \leqslant 2^{2K}(2JK)^{2JK} V^{2JK} \leqslant K^{O(JK)},$$

where we used that the number of partitions of a finite set S is $\leqslant |S|^{|S|}$ and that $V, J \ll \log K$. □

Proof of Lemma 2.8.7. By the triangle inequality,

$$\begin{aligned} E_{\mathcal{S},\mathcal{L},\mathcal{U},r} &\leqslant \sum_{\mathcal{S}_1\sqcup\mathcal{S}_2=\mathcal{S}} \sum_{\boldsymbol{d}\in\mathbf{D}_{\mathcal{S},\mathcal{L},r}} \sum_{n\in I_N} \Big(\prod_{(i,j)\in\mathcal{L}\sqcup\mathcal{S}_1} \mathbf{1}_{d_{ij}\mid n+\mathbf{b}_i}\Big)\Big(\prod_{(i,j)\in\mathcal{U}\sqcup\mathcal{S}_2} \frac{1}{d_{ij}}\Big) \\ &\ll N \sum_{\mathcal{S}_1\sqcup\mathcal{S}_2=\mathcal{S}} \sum_{\boldsymbol{d}\in\mathbf{D}_{\mathcal{S},\mathcal{L},r}} \Big(\prod_{p\mid \boldsymbol{d}_{\mathcal{L}\sqcup\mathcal{S}_1}} \frac{1}{p}\Big)\Big(\prod_{(i,j)\in\mathcal{U}\sqcup\mathcal{S}_2} \frac{1}{d_{ij}}\Big), \end{aligned}$$

using that the arithmetic progression defined by the divisibility conditions on n has modulus $\prod_{p\mid \boldsymbol{d}_{\mathcal{L}\sqcup\mathcal{S}_1}} p \leqslant H^{JK}$, which is $N^{o(1)}$ by (2.22).

We now examine the number of occurrences of the primes $p \mid \boldsymbol{d}_*$ in the two products

$$P_1 := \prod_{p\mid \boldsymbol{d}_{\mathcal{L}\sqcup\mathcal{S}_1}} \frac{1}{p} \quad \text{and} \quad P_2 := \prod_{(i,j)\in\mathcal{U}\sqcup\mathcal{S}_2} \frac{1}{d_{ij}}.$$

Every prime $p \mid \boldsymbol{d}_{\mathcal{S}}$ appears exactly once, either in P_1 or in P_2. Every prime $p \mid \boldsymbol{d}_{\mathcal{L}}$ with $p \nmid \boldsymbol{d}_{\mathcal{U}}$ appears exactly once in P_1 and does not appear in P_2. If instead $p \mid \boldsymbol{d}_{\mathcal{U}}$ and $p \nmid \boldsymbol{d}_{\mathcal{L}}$, then p does not appear in P_1 but appears $\nu_p(\boldsymbol{d}_{\mathcal{U}}) \geqslant 2$ times in P_2. Finally, every prime p dividing both $\boldsymbol{d}_{\mathcal{L}}$ and $\boldsymbol{d}_{\mathcal{U}}$ appears exactly once in P_1 and $\nu_p(\boldsymbol{d}_{\mathcal{U}}) \geqslant 1$ times in P_2. Therefore,

$$P_1P_2 = \prod_{p\mid \boldsymbol{d}_*} \frac{1}{p} \cdot \prod_{p\mid \boldsymbol{d}_{\mathcal{U}}} \frac{1}{p^{\nu_p(\boldsymbol{d}_{\mathcal{U}})-\mathbf{1}_{p\nmid \boldsymbol{d}_{\mathcal{L}}}}} \leqslant \prod_{p\mid \boldsymbol{d}_*} \frac{1}{p} \cdot \prod_{p\mid \boldsymbol{d}_{\mathcal{U}}} \frac{1}{p^{\nu_p(\boldsymbol{d}_{\mathcal{U}})/2}} \leqslant H_0^{-|\mathcal{U}|/2} \prod_{p\mid \boldsymbol{d}_*} \frac{1}{p}.$$

Hence, by Lemma 2.8.8,

$$E_{\mathcal{S},\mathcal{L},\mathcal{U},r} \ll N2^{|\mathcal{S}|} H_0^{-|\mathcal{U}|/2} \sum_{\boldsymbol{d}\in(\pm\mathcal{D})^{2K}} \prod_{p\mid \boldsymbol{d}_*} \frac{1}{p} \ll NK^{O(JK)} H_0^{-|\mathcal{U}|/2}.$$

Recall that $\log H_0 \gg K^{1-\varepsilon_1}$, while $|\mathcal{U}| \geqslant K^{2\varepsilon_1}$ by assumption. Thus

$$K^{O(JK)} H_0^{-|\mathcal{U}|} \ll K^{O(K\log K)} e^{-K^{1+\varepsilon_1}} \ll 1,$$

and the bound $E_{\mathcal{S},\mathcal{L},\mathcal{U},r} \ll N$ follows. □

2.9 Cancellation from single primes

In Lemma 2.8.7, we proved that $E_{\mathcal{S},\mathcal{L},\mathcal{U},r}$ is small whenever the number of unlit indices is moderately large. In this section, we will show that the same is true when the number of single indices is large.

Proposition 2.9.1. *Let $\mathcal{S}$, $\mathcal{L}$, $\mathcal{U}$ be sets such that $[\![2K]\!] \times [\![J]\!] = \mathcal{S} \sqcup \mathcal{L} \sqcup \mathcal{U}$ and let $1 \leqslant r \leqslant K$.*

Suppose that $|\mathcal{S}| \geqslant K^{1-\varepsilon_1}$ and $|\mathcal{U}| \leqslant K^{2\varepsilon_1}$. Then

$$E_{\mathcal{S},\mathcal{L},\mathcal{U},r} \ll e^{O(JK)}N.$$

To prove Proposition 2.9.1, we need to exhibit cancellation in a sum ranging over the intersection of I_N with shifted copies of Y_L. If we did not have to deal with Y_L, this would be a simple task. However, the presence of Y_L requires us to develop a more sophisticated approach. We will use a combinatorial sieve to approximate the indicator of Y_L by a linear combination of arithmetic progressions, which are easier to handle.

2.9.1 Cancellation over arithmetic progressions

In this chapter, the term *arithmetic progression* always refers to a strictly two-sided infinite arithmetic progression of the form $a + q\mathbb{Z}$ for some $a \in \mathbb{Z}$ and $q \in \mathbb{N}$.

Notation 2.9.2. The modulus q of an arithmetic progression $R = a + q\mathbb{Z}$ is denoted by q_R.

We first prove cancellation over arithmetic progressions in the presence of many single indices.

Lemma 2.9.3. *Let $\mathcal{S}$, $\mathcal{L}$, $\mathcal{U}$ be sets such that $[\![2K]\!] \times [\![J]\!] = \mathcal{S} \sqcup \mathcal{L} \sqcup \mathcal{U}$. Let $1 \leqslant r \leqslant K$ and $\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}$.*

Let $R \subset \mathbb{Z}$ be an arithmetic progression with square-free modulus q_R not divisible by $\boldsymbol{d}_{\mathcal{S}}$. Then

$$\sum_{n \in I_N \cap R} \mathbf{1}_{\substack{d_{ij} | n+\mathbf{b}_i \, \forall (i,j) \in \mathcal{L} \\ d_{ij} \nmid n+\mathbf{b}_i \, \forall (i,j) \in \mathcal{U}}} \prod_{(i,j) \in [\![2K]\!] \times [\![J]\!]} \left(\mathbf{1}_{d_{ij} | n+b_i} - \frac{1}{d_{ij}} \right) \ll H^{2K} q_R. \tag{2.35}$$

Proof. Since $\boldsymbol{d}_{\mathcal{S}}$ does not divide q_R, there exists an index $(i_0, j_0) \in \mathcal{S}$ such that $p_0 := d_{i_0 j_0} \nmid q_R$.

Introducing the variable

$$Q := \prod_{\substack{p | \boldsymbol{d}_* q_R \\ p \neq p_0}} p,$$

we may express the sum (2.35) as

$$\sum_{a \,(\mathrm{mod}\, Q)} c(a) \sum_{\substack{n \in I_N \\ n \equiv a \,(\mathrm{mod}\, Q)}} \left(\mathbf{1}_{p_0 | n + b_{i_0}} - \frac{1}{p_0} \right)$$

for some 1-bounded coefficients $c(a)$. Since $p_0 \nmid Q$, the inner sum exhibits near-perfect cancellation:

$$\sum_{\substack{n \in I_N \\ n \equiv a \,(\mathrm{mod}\, Q)}} \left(\mathbf{1}_{p_0 | n + b_{i_0}} - \frac{1}{p_0} \right) = \frac{N}{p_0 Q} + O(1) - \frac{1}{p_0} \left(\frac{N}{Q} + O(1) \right) \ll 1.$$

Thus, the left-hand side of (2.35) is bounded by $\ll Q \leqslant \boldsymbol{d}_* q_R \leqslant H^{2K} q_R$, as desired. □

2.9.2 Cancellation over Y_L

Recall that $\mathcal{Y}$ is the set of all prohibited progressions, and Y_L is the complement of the union of these prohibited progressions. We need to use a suitable version of the inclusion-exclusion principle to express $\mathbf{1}_{Y_L}$ as a linear combination of indicators of intersections of prohibited progressions.

The exact inclusion-exclusion formula

$$\mathbf{1}_{n \in Y_L} = \mathbf{1}_{n \notin P \ \forall P \in \mathcal{Y}} = 1 - \sum_{P_1 \in \mathcal{Y}} \mathbf{1}_{n \in P_1} + \sum_{\substack{P_1, P_2 \in \mathcal{Y} \\ \text{distinct}}} \mathbf{1}_{n \in P_1 \cap P_2} - \sum_{\substack{P_1, P_2, P_3 \in \mathcal{Y} \\ \text{distinct}}} \mathbf{1}_{n \in P_1 \cap P_2 \cap P_3} + \cdots . \tag{2.36}$$

has too many terms to be useful. We require a truncated version, also known as a combinatorial sieve. The combinatorial sieve we will use was developed by Helfgott and Radziwiłł [26], using ideas from the theory of the Möbius function of partially ordered sets. Its two main features are the following.

- Because the progressions $P \in \mathcal{Y}$ have composite (square-free) moduli, several intersections of progressions in $\mathcal{Y}$ can yield the same result. For example,

  $$5\mathbb{Z} \cap 6\mathbb{Z} \cap 7\mathbb{Z} = 14\mathbb{Z} \cap 30\mathbb{Z} = 2\mathbb{Z} \cap 6\mathbb{Z} \cap 15\mathbb{Z} \cap 21\mathbb{Z}.$$

  Let R be a progression. In the right-hand side of (2.36), all of the terms $\pm \mathbf{1}_{n \in P_1 \cap \ldots \cap P_i}$ with $i \geqslant 1$ and $P_1 \cap \ldots \cap P_i = R$ can be combined, and simplify to $c_R \mathbf{1}_{n \in R}$ for some integer coefficient c_R. However, if the modulus q_R of R has k prime factors, there can be close to 2^{2^k} ways of expressing R as an intersection of distinct arithmetic progressions. The most naive bound would thus give $|c_R| \leqslant 2^{2^k}$. This is much larger than what we can allow. Fortunately, the combinatorial interpretation[15] of this coefficient c_R shows that there is a vast

[15] In combinatorial language, c_R is a value of the Möbius function of the partially ordered set consisting of all possible intersections of prohibited progressions.

amount of cancellation from the ± 1 signs, and the much more reasonable bound $|c_R| \leqslant 2^k$ holds.[16]

- A classical way to approximate the inclusion-exclusion formula is by means of the *Bonferroni inequalities*. These imply that, for any $t \geqslant 1$,

$$\mathbf{1}_{n\in Y_L} = \sum_{i=0}^{t-1} \sum_{\substack{P_1,\dots,P_i\in\mathcal{Y}\\ \text{distinct}}} (-1)^i \mathbf{1}_{n\in P_1\cap\dots\cap P_i} + O\Bigg(\sum_{\substack{P_1,\dots,P_t\in\mathcal{Y}\\ \text{distinct}}} \mathbf{1}_{n\in P_1\cap\dots\cap P_t} \Bigg).$$

  In this simple version, the terms $(-1)^i \mathbf{1}_{n\in P_1\cap\dots\cap P_i}$ with $i < t$ are kept in the main term, those with $i = t$ constitute the remainder term and the others can be discarded. We require a more flexible truncation method, not just based on the number i of sets in the intersection, but on specific properties of the progression $P_1 \cap \dots \cap P_i$. For Helfgott and Radziwiłł [26], this cut-off was determined by the number of prime factors of the modulus of the intersection $P_1 \cap \dots \cap P_i$. In our setting, the truncation and its analysis are significantly more technical.

We can now state the Helfgott-Radziwiłł combinatorial sieve for composite moduli with a general cut-off.

Proposition 2.9.4. *Let $\mathcal{Y}$ be a finite set of arithmetic progressions in $\mathbb{Z}$ with square-free moduli. Let*

$$\mathcal{Y}^\cap := \{\cap S : S \subset \mathcal{Y}\}.$$

Fix a non-empty subset $X \subset \mathcal{Y}^\cap$ that is closed under containment, i.e. if a progression $P \in \mathcal{Y}^\cap$ is an element of X, then so are all $P' \in \mathcal{Y}^\cap$ with $P' \supset P$. Let $\mathcal{X}$ be the collection of subsets of $\mathcal{Y}$ defined by[17]

$$\mathcal{X} = \{S \subset \mathcal{Y} : \cap S \in X\}.$$

Then

$$\mathbf{1}_{n\notin P \ \forall P\in\mathcal{Y}} = \sum_{S\in\mathcal{X}} (-1)^{|S|} \mathbf{1}_{n\in\cap S} + O\Bigg(\sum_{R\in\partial X} 3^{\omega(q_R)} \mathbf{1}_{n\in R} \Bigg) \tag{2.37}$$

where

$$\partial X := \{R \in \mathcal{Y}^\cap : R \notin X \text{ and } R = P \cap P' \text{ for some } P \in X \text{ and } P' \in \mathcal{Y}\}.$$

Moreover, the first sum can be rewritten as

$$\sum_{S\in\mathcal{X}} (-1)^{|S|} \mathbf{1}_{n\in\cap S} = \sum_{P\in X} c_P \mathbf{1}_{n\in P}$$

for some coefficients $c_P \in \mathbb{Z}$ satisfying $|c_P| \leqslant 2^{\omega(q_P)}$.

[16] Optimal bounds for c_R are due to Sagan, Yeh and Ziegler (see [54, after Corollary 2.5]). Helfgott and Radziwiłł [26] gave a one-line proof of the slightly weaker bound $|c_R| \leqslant 2^k$.

[17] Note that $\emptyset \in \mathcal{X}$ since $\cap\emptyset = \mathbb{Z}$ by convention.

A self-contained proof of Proposition 2.9.4 is provided in [47, Appendix B] (which is a shortened version of the proof in [26]). We will apply this combinatorial sieve to rewrite the term $\mathbf{1}_{\forall i,\, n+\mathbf{b}_i \in Y_L}$ as a suitable combination of arithmetic progressions. To do so, we need to introduce some notation.

Notation 2.9.5. Let $\mathcal{S}$, $\mathcal{L}$, $\mathcal{U}$ be sets such that $[\![2K]\!] \times [\![J]\!] = \mathcal{S} \sqcup \mathcal{L} \sqcup \mathcal{U}$ and let $1 \leqslant r \leqslant K$. Let $\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}$ and let $\boldsymbol{b} := (b_i)_{0 \leqslant i \leqslant 2K}$ be the associated vector of partial sums $b_i := \sum_{1 \leqslant k \leqslant i} d_k$. We write

$$\mathcal{Y} - \boldsymbol{b} := \{P - \mathbf{b}_i : P \in \mathcal{Y},\, 0 \leqslant i \leqslant 2K\}.$$

We also define

$$(\mathcal{Y} - \boldsymbol{b})^{\cap} := \left\{ \bigcap_{P \in X} P : X \subset \mathcal{Y} - \boldsymbol{b} \right\},$$

the set of all possible intersections of such shifted progressions (with the convention $\bigcap_{P \in \emptyset} P := \mathbb{Z}$).

The next lemma captures our application of the combinatorial sieve. It is rather technical, and we defer its proof to Section 2.13.2. The statement of Lemma 2.9.6 can be understood as follows. In (3), the approximate inclusion-exclusion formula is given, with a main term and a remainder term. The main term is a sum over all progressions with small *rank*. The rank of a progression can be thought of as a measure of its complexity. It is a quantity depending on $\boldsymbol{d}$, but its precise definition is not immediately needed and hence will only be given later, in Definition 2.13.3. Two simple properties of the rank are given in (1) and (2). Finally, the bound (4) will be used to control the remainder term in the combinatorial sieve. The proof of (4) is particularly intricate.

Lemma 2.9.6. *Let $\mathcal{S}$, $\mathcal{L}$, $\mathcal{U}$ be sets such that $[\![2K]\!] \times [\![J]\!] = \mathcal{S} \sqcup \mathcal{L} \sqcup \mathcal{U}$ and let $1 \leqslant r \leqslant K$.*

For every $\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}$, there exists a function

$$\mathrm{rank}_{\boldsymbol{d}} : (\mathcal{Y} - \boldsymbol{b})^{\cap} \to \mathbb{Z}^{\geqslant 0} \cup \{+\infty\}$$

satisfying the conditions below.

Define the arithmetic progression $A_{\mathcal{L},\boldsymbol{d}} := \{n \in \mathbb{Z} : \forall (i,j) \in \mathcal{L},\ d_{ij} \mid n + \mathbf{b}_i\}$.

Let $X_{\boldsymbol{d}}$ be the set of all $R \in (\mathcal{Y} - \boldsymbol{b})^{\cap}$ such that $\mathrm{rank}_{\boldsymbol{d}}(R) < K^{5\varepsilon_1}$. Let $\partial X_{\boldsymbol{d}}$ be the set of all $R \in (\mathcal{Y} - \boldsymbol{b})^{\cap} \setminus X_{\boldsymbol{d}}$ of the form $R = R' \cap P$ for some $R' \in X_{\boldsymbol{d}}$ and $P \in \mathcal{Y} - \boldsymbol{b}$.

Then, the following properties hold.

1. *(Primes dividing the modulus and $\boldsymbol{d}_{\mathcal{S}}$) For every $R \in (\mathcal{Y} - \boldsymbol{b})^{\cap}$,*

$$|\{p : p \mid (q_R, \boldsymbol{d}_{\mathcal{S}})\}| \leqslant JL \,\mathrm{rank}_{\boldsymbol{d}}(R).$$

2. *(Primes dividing the modulus) For every* $R \in (\mathcal{Y} - \boldsymbol{b})^{\cap}$,

$$\omega(q_R) \ll JL \operatorname{rank}_{\boldsymbol{d}}(R) + JK.$$

3. *(Combinatorial sieve) For all* $n \in \mathbb{Z}$, *we have*

$$\mathbf{1}_{\forall i,\, n+\mathbf{b}_i \in Y_L} \mathbf{1}_{A_{\mathcal{L},\boldsymbol{d}}}(n) = \sum_{R \in X_{\boldsymbol{d}}} c_{R,\boldsymbol{d}} \mathbf{1}_{n \in R \cap A_{\mathcal{L},\boldsymbol{d}}} + O\left(e^{O(JK)} \sum_{\substack{R \in \partial X_{\boldsymbol{d}} \\ R \cap A_{\mathcal{L},\boldsymbol{d}} \neq \emptyset}} \mathbf{1}_{n \in R \cap A_{\mathcal{L},\boldsymbol{d}}} \right),$$

where the coefficients $c_{R,\boldsymbol{d}}$ *are independent of* n *and satisfy* $|c_{R,\boldsymbol{d}}| \leqslant e^{O(JK)}$.

4. *(Remainder term bound) Suppose that* $|\mathcal{U}| \leqslant K^{2\varepsilon_1}$. *Then*

$$\sum_{\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}} \sum_{\substack{R \in \partial X_{\boldsymbol{d}} \\ R \cap A_{\mathcal{L},\boldsymbol{d}} \neq \emptyset}} \prod_{p \mid q_R \boldsymbol{d}_*} \frac{1}{p} \ll 1.$$

We now have all the necessary ingredients to prove Proposition 2.9.1.

Proof of Proposition 2.9.1, assuming Lemma 2.9.6. We use the combinatorial sieve to rewrite the term involving Y_L in terms of arithmetic progressions: by part (3) of Lemma 2.9.6 and the triangle inequality, we have

$$E_{\mathcal{S},\mathcal{L},\mathcal{U},r} \leqslant e^{O(JK)}(F_1 + F_2) \tag{2.38}$$

where

$$F_1 := \sum_{\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}} \sum_{R \in X_{\boldsymbol{d}}} \left| \sum_{n \in I_N \cap R} \mathbf{1}_{\substack{d_{ij} \mid n+\mathbf{b}_i \, \forall (i,j) \in \mathcal{L} \\ d_{ij} \nmid n+\mathbf{b}_i \, \forall (i,j) \in \mathcal{U}}} \prod_{(i,j) \in [\![2K]\!] \times [\![J]\!]} \left(\mathbf{1}_{d_{ij} \mid n+b_i} - \frac{1}{d_{ij}} \right) \right|$$

and

$$F_2 := \sum_{\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}} \sum_{\substack{R \in \partial X_{\boldsymbol{d}} \\ R \cap A_{\mathcal{L},\boldsymbol{d}} \neq \emptyset}} \sum_{n \in I_N \cap R} \mathbf{1}_{\substack{d_{ij} \mid n+\mathbf{b}_i \, \forall (i,j) \in \mathcal{L} \\ d_{ij} \nmid n+\mathbf{b}_i \, \forall (i,j) \in \mathcal{U}}} \prod_{(i,j) \in [\![2K]\!] \times [\![J]\!]} \left| \mathbf{1}_{d_{ij} \mid n+b_i} - \frac{1}{d_{ij}} \right|.$$

We wish to apply Lemma 2.9.3 to bound the innermost sum in F_1. To do so, we must check that the modulus of the arithmetic progression R is square-free and is not divisible by $\boldsymbol{d}_{\mathcal{S}}$. The first condition is immediate since $R \in (\mathcal{Y} - \boldsymbol{b})^{\cap}$ and all prohibited progressions have square-free moduli. To verify the second condition, observe that the number of primes dividing both q_R and $\boldsymbol{d}_{\mathcal{S}}$ is bounded by

$$\leqslant JL \operatorname{rank}_{\boldsymbol{d}}(R) \leqslant JK^{1-10\varepsilon_1} K^{5\varepsilon_1} \ll K^{1-4\varepsilon_1},$$

using part (1) of Lemma 2.9.6 and the definition of $X_{\boldsymbol{d}}$. On the other hand, $\boldsymbol{d}_{\mathcal{S}}$ has $|\mathcal{S}| \geqslant K^{1-\varepsilon_1}$ distinct prime factors by assumption. Since K is sufficiently large, these bounds ensure that $\boldsymbol{d}_{\mathcal{S}} \nmid q_R$.

Therefore, by Lemma 2.9.3, for every $R \in X_{\boldsymbol{d}}$ we have

$$\sum_{n \in I_N \cap R} \mathbf{1}_{d_{ij} | n + \mathbf{b}_i \, \forall (i,j) \in \mathcal{L}} \mathbf{1}_{d_{ij} \nmid n + \mathbf{b}_i \, \forall (i,j) \in \mathcal{U}} \prod_{(i,j) \in [\![2K]\!] \times [\![J]\!]} \left(\mathbf{1}_{d_{ij} | n + b_i} - \frac{1}{d_{ij}} \right) \ll H^{2K} q_R.$$

All primes dividing q_R are in $\mathcal{P}$, so by part (2) of Lemma 2.9.6 and the definition of $X_{\boldsymbol{d}}$, we have

$$q_R \leqslant H^{\omega(q_R)} \leqslant H^{O(JL \operatorname{rank}_{\boldsymbol{d}}(R) + JK)} \ll H^{O(JK)}. \tag{2.39}$$

To complete the estimation of F_1, we only need a crude bound for the sums over $\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}$ and $R \in X_{\boldsymbol{d}}$. Since an arithmetic progression R is uniquely determined by its modulus q_R and a residue class modulo q_R, by (2.39), there are $\ll H^{O(JK)}$ choices for $R \in X_{\boldsymbol{d}}$. Moreover, there are at most $(2H)^{2K}$ choices for $\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}$. We thus have

$$F_1 \ll H^{O(JK)}. \tag{2.40}$$

We turn to the second term F_2. By the triangle inequality, we have

$$F_2 \leqslant \sum_{\mathcal{S}_1 \sqcup \mathcal{S}_2 = \mathcal{S}} \sum_{\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}} \sum_{\substack{R \in \partial X_{\boldsymbol{d}} \\ R \cap A_{\mathcal{L},\boldsymbol{d}} \neq \emptyset}} \sum_{n \in I_N \cap R} \left(\prod_{(i,j) \in \mathcal{L} \sqcup \mathcal{S}_1} \mathbf{1}_{d_{ij} | n + \mathbf{b}_i} \right) \left(\prod_{(i,j) \in \mathcal{U} \sqcup \mathcal{S}_2} \frac{1}{d_{ij}} \right). \tag{2.41}$$

For each $R \in \partial X_{\boldsymbol{d}}$, by definition of $\partial X_{\boldsymbol{d}}$, there are progressions $R' \in X_{\boldsymbol{d}}$ and $P \in \mathcal{Y} - \boldsymbol{b}$ such that $R = R' \cap P$. This implies that $q_R \leqslant q_{R'} q_P$. Note that $\omega(q_P) \leqslant JL$ by definition of a prohibited progression, so that $q_P \leqslant H^{JL}$. In addition, we have $q_{R'} \ll H^{O(JK)}$ by (2.39). This proves that

$$q_R \ll H^{O(JK)} \tag{2.42}$$

for all $R \in \partial X_{\boldsymbol{d}}$.

The estimate (2.42), the simple bound $\boldsymbol{d}_* \leqslant H^{2K}$ and (2.22) imply that

$$\sum_{n \in I_N \cap R} \prod_{(i,j) \in \mathcal{L} \sqcup \mathcal{S}_1} \mathbf{1}_{d_{ij} | n + \mathbf{b}_i} \ll N \prod_{p | q_R \boldsymbol{d}_{\mathcal{L} \sqcup \mathcal{S}_1}} \frac{1}{p}. \tag{2.43}$$

Substituting (2.43) into (2.41), we obtain

$$F_2 \ll 2^{|S|} N \sum_{\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}} \sum_{\substack{R \in \partial X_{\boldsymbol{d}} \\ R \cap A_{\mathcal{L},\boldsymbol{d}} \neq \emptyset}} \prod_{p | q_R \boldsymbol{d}_*} \frac{1}{p}.$$

Here, we recognise the expression appearing in part (4) of Lemma 2.9.6. Applying this lemma, we get

$$F_2 \ll e^{O(JK)} N. \tag{2.44}$$

Finally, substituting (2.40) and (2.44) into (2.38), we obtain the claimed bound $E_{\mathcal{S},\mathcal{L},\mathcal{U},r} \ll e^{O(JK)} N$. □

2.10 Predictable walks

By Lemma 2.8.7 and Proposition 2.9.1, it remains to bound the expression $E_{\mathcal{S},\mathcal{L},\mathcal{U},r}$ in the case where there are few single indices and very few unlit indices. In this regime, the inner sum appearing in the definition of $E_{\mathcal{S},\mathcal{L},\mathcal{U},r}$ yields no meaningful cancellation, so we begin by applying the triangle inequality.

Lemma 2.10.1. *Let $\mathcal{S}$, $\mathcal{L}$, $\mathcal{U}$ be sets such that $[\![2K]\!] \times [\![J]\!] = \mathcal{S} \sqcup \mathcal{L} \sqcup \mathcal{U}$ and let $1 \leqslant r \leqslant K$. Then*

$$E_{\mathcal{S},\mathcal{L},\mathcal{U},r} \ll e^{O(JK)} N \sum_{\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}} \prod_{p \mid \boldsymbol{d}_*} \frac{1}{p}.$$

Proof. By definition of $E_{\mathcal{S},\mathcal{L},\mathcal{U},r}$ and the triangle inequality,

$$E_{\mathcal{S},\mathcal{L},\mathcal{U},r} \leqslant \sum_{\mathcal{S}_1 \sqcup \mathcal{S}_2 = \mathcal{S}} \sum_{\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}} \sum_{n \in I_N} \Bigg(\prod_{(i,j) \in \mathcal{L} \sqcup \mathcal{S}_1} \mathbf{1}_{d_{ij} \mid n + \mathbf{b}_i} \Bigg) \Bigg(\prod_{(i,j) \in \mathcal{U} \sqcup \mathcal{S}_2} \frac{1}{d_{ij}} \Bigg).$$

Since $\boldsymbol{d}_* \leqslant H^{2K} \ll N$ by (2.22), we have

$$\sum_{n \in I_N} \prod_{(i,j) \in \mathcal{L} \sqcup \mathcal{S}_1} \mathbf{1}_{d_{ij} \mid n + \mathbf{b}_i} \ll N \prod_{p \mid \boldsymbol{d}_{\mathcal{L} \sqcup \mathcal{S}_1}} \frac{1}{p}.$$

Simplifying gives the lemma. □

By Lemma 2.10.1, our task is reduced to giving a good bound for

$$\sum_{\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}} \prod_{p \mid \boldsymbol{d}_*} \frac{1}{p}. \tag{2.45}$$

A similar sum over the larger domain $\boldsymbol{d} \in (\pm\mathcal{D})^{2K}$ was analysed in Lemma 2.8.8, where we obtained an optimal bound of order $K^{O(JK)}$. However, such a bound is far too large for our purposes here, as we require a bound of the form $e^{O(JK)} V^{JK}$. To achieve this, we must exploit the structure of the set $\mathbf{D}_{\mathcal{S},\mathcal{L},r}$. This marks the beginning of a purely combinatorial phase of the argument, where we analyse how the conditions defining $\mathbf{D}_{\mathcal{S},\mathcal{L},r}$ limit the complexity of the sum.

In the present setting, most indices are lit. The central condition is property (2) of Definition 2.8.4, which imposes a divisibility relation for every pair of lit indices corresponding to the same prime. Obtaining sufficient control over the dependencies among these divisibility conditions requires a detailed combinatorial analysis.

2.10.1 Predictable words

For our combinatorial work, we adopt the language of words and letters. The general statements we prove in this setting will later be applied to sequences of primes from a fixed set $\mathcal{P}_j$, to track repeated occurrences of the same prime along a walk.

Definition 2.10.2. Let $\mathcal{A}$ be a finite set (the alphabet). Let $\mathcal{W}_n$ be the set of all n-letter words on $\mathcal{A}$, where no two consecutive letters are the same. Let $\mathcal{W}_n^{\neq} \subset \mathcal{W}_n$ be the set of all n-letter words on $\mathcal{A}$ with distinct letters. Let $\mathcal{W} = \bigcup_{n \geqslant 1} \mathcal{W}_n$ and $\mathcal{W}^{\neq} = \bigcup_{n \geqslant 1} \mathcal{W}_n^{\neq}$.

For $w \in \mathcal{W}_n$ and $1 \leqslant k \leqslant n$, we write $w[k]$ for the k-th letter of w. We denote by $w[*]$ the set of all letters of w.

We denote the set of all positions of the letter $\mathtt{A}$ in w by $\mathrm{Pos}(\mathtt{A}; w)$ (which is the set of all $k \in [\![n]\!]$ such that $w[k] = \mathtt{A}$). For $l \in [\![n]\!]$, we write $\mathrm{Pos}(l; w)$ for the set $\{k \in [\![n]\!] : w[k] = w[l]\}$ (instead of '$\mathrm{Pos}(w[l]; w)$').

The notation $v \sqsubset w$ means that v is a *substring* of w, i.e. a sequence of consecutive letters of w.

We write $\overleftarrow{w}$ for the word obtained by writing the letters of w in the reversed order.

The *concatenation* of two words w_1 and w_2 is the word obtained by appending the letters of w_2 at the end of w_1. We denote it by $w_1 w_2$.

We now introduce a measure of the amount of structure in a word. We do so by counting the number of letters with *constant neighbours*. These are letters for which every occurrence is always surrounded by the same set of letters. When most letters in a word have constant neighbours, their repetition patterns are jointly well-behaved.

Definition 2.10.3. Let $w \in \mathcal{W}$ and $\mathtt{A} \in \mathcal{A}$. We say that $\mathtt{A}$ has *variable neighbours in* w if there exist two occurrences of $\mathtt{A}$ in w, such that the letters immediately adjacent to $\mathtt{A}$ at these two positions form two distinct sets. Otherwise, $\mathtt{A}$ is said to have *constant neighbours in* w.

For example,

w	set of neighbours for every $\mathtt{A}$ in w	neighbours of $\mathtt{A}$ in w
XAYZYAXAY	{X,Y}, {X,Y}, {X,Y}	constant
AXYXAXZY	{X}, {X}	constant
XAYZYAYZXAY	{X,Y}, {Y}, {X,Y}	variable
YAXYZAXA	{X,Y}, {X,Z}, {X}	variable

.

Definition 2.10.4. A word $w \in \mathcal{W}$ is said to be *t-predictable* if the following conditions both hold.

1. Every letter appears $\leqslant t$ times in w.

2. There are $\leqslant t$ letters with variable neighbours in w.

Otherwise w is called *t-unpredictable*.

2.10.2 Counting predictable words

We show in Lemma 2.10.7 that, up to letter relabelling, there are $\leqslant n^{O(t^3)}$ words of length n which are t-predictable. For moderate values of t, this improves on the trivial bound given by the number of partitions of $[\![n]\!]$, which is $n^{(1+o(1))n}$.

For this section, we could have used the language of partitions since our primary focus is on the positions of the letters, and not the letters themselves. However, we found it more convenient to use words for Section 2.12, so we will use them here as well.

Definition 2.10.5. Two words w_1 and w_2 are said to be *equivalent* if one can be obtained from the other by letter relabelling. That is, they have the same length n and there is a bijection $\varphi : w_1[*] \to w_2[*]$ such that $w_2[i] = \varphi(w_1[i])$ for all $i \in [\![n]\!]$.

The next lemma states that the sets of positions of a certain subset L_w of the letters of any word $w \in \mathcal{W}$ uniquely determine the sets of positions of all the other letters.

Lemma 2.10.6. *Let $n \geqslant 2$ and let $w_1, w_2 \in \mathcal{W}_n$. For $i \in \{1, 2\}$, let $L_i \subset w_i[*]$ be defined by*

$$\begin{aligned} L_i := \{w_i[1], w_i[2]\} &\cup \{\mathtt{A} \in w_i[*] : \mathtt{A} \textit{ has variable neighbours in } w_i\} \\ &\cup \{\mathtt{B} \in w_i[*] : \mathtt{B} \textit{ appears in } w_i \textit{ next to a letter} \\ &\qquad \textit{having variable neighbours in } w_i\}. \end{aligned}$$

Suppose that $\big\{\mathrm{Pos}(\mathtt{A}; w_1) : \mathtt{A} \in L_1\big\} = \big\{\mathrm{Pos}(\mathtt{A}; w_2) : \mathtt{A} \in L_2\big\}$. Then w_1 and w_2 are equivalent.

Proof. Suppose that the conclusion does not hold, and let $k \geqslant 1$ be minimal with the property that $\mathrm{Pos}(k; w_1) \neq \mathrm{Pos}(k; w_2)$. Hence, $w_1[k] \notin L_1$ and $w_2[k] \notin L_2$ by the assumption in the statement. In particular, $k \geqslant 3$ since $w_i[1], w_i[2] \in L_i$.

Note that $w_1[k] \neq w_1[k-2]$. Indeed, if $w_1[k] = w_1[k-2]$, we would have

$$k \in \mathrm{Pos}(k; w_1) = \mathrm{Pos}(k-2; w_1) = \mathrm{Pos}(k-2; w_2),$$

where the last equality follows from the minimality of k. This would imply that

$$\mathrm{Pos}(k; w_2) = \mathrm{Pos}(k-2; w_2) = \mathrm{Pos}(k; w_1),$$

contradicting our assumption.

By definition of L_1, both $w_1[k]$ and $w_1[k-1]$ have constant neighbours in w_1. This means that every occurrence of the letter $w_1[k-1]$ in w_1 is surrounded by the (distinct) letters $w_1[k-2]$ and $w_1[k]$, in any order. In addition, every appearance of $w_1[k]$ is adjacent to an occurrence of $w_1[k-1]$. Thus, we may describe $\mathrm{Pos}(k; w_1)$ exactly as

$$\mathrm{Pos}(k; w_1) = \big\{ l \in [\![n]\!] \ : \ \{l-1, l+1\} \cap \mathrm{Pos}(k-1; w_1) \neq \emptyset, \ l \notin \mathrm{Pos}(k-2; w_1) \big\}. \quad (2.46)$$

The same reasoning with w_2 shows that

$$\mathrm{Pos}(k; w_2) = \big\{ l \in [\![n]\!] \ : \ \{l-1, l+1\} \cap \mathrm{Pos}(k-1; w_2) \neq \emptyset, \ l \notin \mathrm{Pos}(k-2; w_2) \big\}. \quad (2.47)$$

However, $\mathrm{Pos}(k-2; w_1) = \mathrm{Pos}(k-2; w_2)$ and $\mathrm{Pos}(k-1; w_1) = \mathrm{Pos}(k-1; w_2)$ by minimality of k, so (2.46) and (2.47) imply that $\mathrm{Pos}(k; w_1) = \mathrm{Pos}(k; w_2)$, a contradiction. □

We can now prove our bound for the number of inequivalent t-predictable words of length n.

Lemma 2.10.7. *Let $n, t \geqslant 1$. There are $\leqslant n^{O(t^3)}$ t-predictable words in $\mathcal{W}_n$, up to equivalence.*

Proof. By Lemma 2.10.6, it suffices to bound the number of possibilities for the set

$$\big\{ \mathrm{Pos}(\mathtt{A}; w) : \mathtt{A} \in L_w \big\} \quad (2.48)$$

where w ranges over the set of t-predictable words in $\mathcal{W}_n$, and $L_w \subset w[*]$ is the set defined in Lemma 2.10.6 (with w instead of w_i).

If w is t-predictable, there are $\leqslant t$ letters with variable neighbours. Moreover, every letter appears $\leqslant t$ times, so for every letter $\mathtt{A}$ there are $\leqslant 2t$ letters adjacent to an occurrence of $\mathtt{A}$. Thus, by definition of L_w we have

$$|L_w| \leqslant 2 + t + t \cdot 2t \leqslant 5t^2.$$

For every $\mathtt{A} \in L_w$, the set $\mathrm{Pos}(\mathtt{A}; w)$ is a subset of $[\![n]\!]$ of size $\leqslant t$, and there are $\leqslant n^t$ such sets. Hence, there are $\leqslant (n^t)^{5t^2}$ possibilities for the set in (2.48), which concludes the proof. □

2.10.3 Contribution of predictable walks

We define a subset $\mathbf{Prd}_{\mathcal{S},\mathcal{L},r} \subset \mathbf{D}_{\mathcal{S},\mathcal{L},r}$ of tuples $\boldsymbol{d}$ for which three words defined in terms of the primes d_{ij} are predictable. By Lemma 2.10.7, this combinatorial restriction allows us to bound the contribution of these $\boldsymbol{d}$ to (2.45). The remaining tuples will be handled in Section 2.12.

Definition 2.10.8 ($v_{j,\boldsymbol{d}}^{(k)}$, $w_{j,\boldsymbol{d}}^{(k)}$)**.** Let $\mathcal{S}$ and $\mathcal{L}$ be disjoint subsets of $[\![2K]\!] \times [\![J]\!]$ and let $1 \leqslant r \leqslant K$.

Let $\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}$ and $j \in [\![J]\!]$. For each $1 \leqslant k \leqslant 3$, we define a word $v_{j,\boldsymbol{d}}^{(k)}$ on the alphabet $\mathcal{P}_j$, whose letters are the primes in $\mathcal{P}_j$ corresponding to the tuple $\boldsymbol{d}^{(k)}$ from Definition 2.8.4. Explicitly,

$$\begin{cases} v_{j,\boldsymbol{d}}^{(1)} := d_{1j}d_{2j}\cdots d_{Kj}, \\ v_{j,\boldsymbol{d}}^{(2)} := d_{(K+1)j}d_{(K+2)j}\cdots d_{(2K)j}, \\ v_{j,\boldsymbol{d}}^{(3)} := d_{1j}d_{2j}\cdots d_{(K-r)j}d_{(K+r+1)j}d_{(K+r+2)j}\cdots d_{(2K)j}. \end{cases}$$

We define $w_{j,\boldsymbol{d}}^{(k)}$ to be the *compression* of $v_{j,\boldsymbol{d}}^{(k)}$, meaning the word formed by replacing any string of consecutive occurrences of a letter with a single instance of that letter. Thus, $w_{j,\boldsymbol{d}}^{(k)} \in \mathcal{W}$.

Definition 2.10.9 ($\mathbf{Prd}_{\mathcal{S},\mathcal{L},r}$, $\mathbf{Unp}_{\mathcal{S},\mathcal{L},r}$)**.** Let $\mathcal{S}$ and $\mathcal{L}$ be disjoint subsets of $[\![2K]\!]\times[\![J]\!]$ and let $1 \leqslant r \leqslant K$.

We define $\mathbf{Prd}_{\mathcal{S},\mathcal{L},r}$ to be the set of all tuples $\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}$ such that, for every $j \in [\![J]\!]$, the three words $w_{j,\boldsymbol{d}}^{(1)}$, $w_{j,\boldsymbol{d}}^{(2)}$ and $w_{j,\boldsymbol{d}}^{(3)}$ are $K^{1/4}$-predictable.

We denote its complement by $\mathbf{Unp}_{\mathcal{S},\mathcal{L},r} := \mathbf{D}_{\mathcal{S},\mathcal{L},r} \setminus \mathbf{Prd}_{\mathcal{S},\mathcal{L},r}$.

We conclude this section by bounding the contribution of the tuples $\boldsymbol{d} \in \mathbf{Prd}_{\mathcal{S},\mathcal{L},r}$ to (2.45).

Proposition 2.10.10. *Let $\mathcal{S}$, $\mathcal{L}$, $\mathcal{U}$ be sets such that $[\![2K]\!] \times [\![J]\!] = \mathcal{S} \sqcup \mathcal{L} \sqcup \mathcal{U}$ and let $1 \leqslant r \leqslant K$. Then*

$$\sum_{\boldsymbol{d}\in\mathbf{Prd}_{\mathcal{S},\mathcal{L},r}} \prod_{p\mid \boldsymbol{d}_*} \frac{1}{p} \ll e^{O(JK)} V^{|\mathcal{S}|+(|\mathcal{L}|+|\mathcal{U}|)/2}.$$

Proof. The proof resembles that of Lemma 2.8.8. The key difference is that we restrict to partitions coming from $K^{1/4}$-predictable words, thereby avoiding combinatorial explosion.

We first identify the combinatorial data needed to uniquely reconstruct a tuple $\boldsymbol{d} \in \mathbf{Prd}_{\mathcal{S},\mathcal{L},r}$. Any $\boldsymbol{d} \in (\pm\mathcal{D})^{2K}$ is uniquely determined by a sequence $\sigma \in \{\pm 1\}^{2K}$ of signs and, for each $j \in [\![J]\!]$, a word u_j of length $2K$, along with a sequence of distinct primes $(p_{\mathtt{A}})_{\mathtt{A}\in u_j[*]}$ with each $p_{\mathtt{A}} \in \mathcal{P}_j$. Indeed, given this data, we can reconstruct $\boldsymbol{d}$ by setting $\operatorname{sign}(d_i) = \sigma_i$ and $d_{ij} = p_{\mathtt{A}}$ if $u_j[i] = \mathtt{A}$.

Fix some $j \in [\![J]\!]$. We claim that there are at most $e^{O(K)}$ possibilities for the word u_j up to equivalence, as $\boldsymbol{d}$ ranges over $\mathbf{Prd}_{\mathcal{S},\mathcal{L},r}$. To prove this, it suffices to show that:

(i) there are $\leqslant e^{O(K)}$ possible words $\left(v_{j,\boldsymbol{d}}^{(k)}\right)_{1\leqslant k\leqslant 3}$ up to equivalence, and

(ii) the equivalence class of u_j is uniquely determined by the equivalence classes of $\big(v_{j,\boldsymbol{d}}^{(k)}\big)_{1\leqslant k\leqslant 3}$.

Let $1 \leqslant k \leqslant 3$, and denote by n_k the length of $w_{j,\boldsymbol{d}}^{(k)}$. Since $n_k \leqslant 2K$, by Lemma 2.10.7, there are

$$\leqslant 2K \cdot (2K)^{O(K^{3/4})} \ll e^K$$

possibilities for the $K^{1/4}$-predictable word $w_{j,\boldsymbol{d}}^{(k)}$ up to equivalence. To recover the equivalence class of $v_{j,\boldsymbol{d}}^{(k)}$ from that of the compressed word $w_{j,\boldsymbol{d}}^{(k)}$, we need to specify a vector $(c_1, c_2, \dots, c_{n_k})$ of positive integers, where c_i records the number of consecutive repetitions of the i-th letter of $w_{j,\boldsymbol{d}}^{(k)}$ in $v_{j,\boldsymbol{d}}^{(k)}$. Since the length of $v_{j,\boldsymbol{d}}^{(k)}$ is $\leqslant 2K$, the number of such integer vectors is $\leqslant e^{O(K)}$. Hence, the total number of possible $v_{j,\boldsymbol{d}}^{(k)}$, modulo equivalence, is also $\leqslant e^{O(K)}$, which proves (i).

We turn to (ii). Fix $\big(v_{j,\boldsymbol{d}}^{(k)}\big)_{1\leqslant k\leqslant 3}$ modulo equivalence, and let u and u' be words of length $2K$ that are possible candidates for u_j. We need to show that u and u' are equivalent. By Definition 2.10.8, the compatibility of u and u' with $\big(v_{j,\boldsymbol{d}}^{(k)}\big)_{1\leqslant k\leqslant 3}$ means that

$$\begin{cases} u|_{[\![K]\!]} \sim u'|_{[\![K]\!]} \\ u|_{[\![K+1,2K]\!]} \sim u'|_{[\![K+1,2K]\!]} \\ u|_{[\![K-r]\!]\cup[\![K+r+1,2K]\!]} \sim u'|_{[\![K-r]\!]\cup[\![K+r+1,2K]\!]} \end{cases} \tag{2.49}$$

where $\sim$ denotes equivalence of words and $u|_I$ denotes the subword of u restricted to the indices in I. By part (3) of Definition 2.8.4, we also know that

$$u[K+\ell] = u[K-\ell+1] \quad \text{and} \quad u'[K+\ell] = u'[K-\ell+1] \tag{2.50}$$

for all $\ell \in [\![r]\!]$.

Suppose that there are $i_1 < i_2$ such that $u[i_1] = u[i_2]$, and let us show that $u'[i_1] = u'[i_2]$.

- If $i_2 \leqslant K$, the first equation in (2.49) implies that $u'[i_1] = u'[i_2]$.
- If $i_1 > K$, the second equation in (2.49) implies that $u'[i_1] = u'[i_2]$.
- If $i_1, i_2 \in [\![K-r]\!] \cup [\![K+r+1, 2K]\!]$, the third equation in (2.49) implies that $u'[i_1] = u'[i_2]$.
- If $i_1 \leqslant K$ and $i_2 = K+\ell$ for some $\ell \in [\![r]\!]$, then $u[i_1] = u[i_2] = u[K-\ell+1]$ by (2.50). The first equation in (2.49) implies that $u'[i_1] = u'[K-\ell+1]$, and thus $u'[i_1] = u'[i_2]$ by (2.50).
- If $i_1 = K-\ell+1$ for some $\ell \in [\![r]\!]$ and $i_2 > K$, then $u[i_2] = u[i_1] = u[K+\ell]$ by (2.50). The second equation in (2.49) implies that $u'[i_2] = u'[K+\ell]$, and thus $u'[i_1] = u'[i_2]$ by (2.50).

Therefore, $u[i_1] = u[i_2]$ implies that $u'[i_1] = u'[i_2]$ in all cases. Of course, the converse is also true by symmetry. This shows that u and u' are equivalent, proving (ii).

We can now finish the proof of the proposition. For every $\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}$, the number of distinct primes dividing $\boldsymbol{d}_*$ is $\leqslant |\mathcal{S}| + \frac{1}{2}(|\mathcal{L}| + |\mathcal{U}|)$, as every d_{ij} with $(i,j) \notin \mathcal{S}$ appears at least twice. Summing over all sequences of signs σ, possible words $(u_j)_{j\in[\![J]\!]}$ modulo equivalence, and primes $(p_{\mathtt{A}})_{\mathtt{A}\in u_j[*]}$ for every j, we get

$$\sum_{\boldsymbol{d}\in\mathbf{Prd}_{\mathcal{S},\mathcal{L},r}} \prod_{p\mid\boldsymbol{d}_*} \frac{1}{p} \leqslant 2^{2K}\big(e^{O(K)}\big)^J V^{|\mathcal{S}|+(|\mathcal{L}|+|\mathcal{U}|)/2},$$

using that $\sum_{p\in\mathcal{P}_j} 1/p = V_j \leqslant V$. Simplifying, we obtain the desired bound. □

2.11 Triangular systems of divisibility constraints

To bound the contribution of unpredictable walks and to treat prohibited sequences, we will need to count vectors $\boldsymbol{d}$ with coordinates $d_i \in \pm\mathcal{D}$ satisfying certain divisibility relations. The precise form of these relations will vary. In this section, we introduce a general framework for them that encompasses all the cases that will arise.

2.11.1 Constraints

We begin by giving a precise definition of a *constraint* on a vector $\boldsymbol{d}$.

Notation 2.11.1. For $R \geqslant 1$ and $d \in (\pm\mathcal{D})^R$, we write d_{ij} for the unique prime in $\mathcal{P}_j$ dividing d_i.

Definition 2.11.2. Let $R \geqslant 1$ and let $\boldsymbol{d} \in (\pm\mathcal{D})^R$. A *constraint* on $\boldsymbol{d}$ is any predicate of the form

$$d_{i_0j_0} \;\Big|\; \sum_{i\in I} d_i + \kappa \tag{2.51}$$

for some $I \subset [\![R]\!]$, $(i_0, j_0) \in [\![R]\!] \times [\![J]\!]$ and $\kappa \in \mathbb{Z}$, where I is the union of at most 10 discrete intervals.[18] The integer κ is called the *shift* of the constraint.

This constraint (2.51) should be viewed as a polynomial divisibility condition on the primes d_{ij}. In many cases, the shift κ will be zero.

We now define what it means for a prime to be *absent* from a constraint, and *involved* in a constraint.

Definition 2.11.3. A prime $p \in \mathcal{P}$ is *absent* from a constraint '$d_{i_0j_0} \mid \sum_{i\in I} d_i + \kappa$' if $p \neq d_{i_0j_0}$ and $p \nmid d_i$ for all $i \in I$.

[18] This condition on the combinatorial complexity of I ensures that the number of possible I is polynomial in R. It will always hold in practice.

The definition of a prime p being *involved* in a constraint is not merely the negation of being absent, because we want to make sure that the constraint is not 'degenerate' when viewed as a condition on p. For instance, consider the constraint $d_{11} \mid d_2 + d_3$. If it happens that $d_{11} = d_{21} = d_{31}$ and $J = 2$, the condition becomes $d_{11} \mid d_{11}(d_{22} + d_{32})$, which is always true. Such a degenerate case provides no useful information about any of the primes, so we wish to exclude it. This motivates the following definition.

Definition 2.11.4. We say that $p \in \mathcal{P}$ is *involved* in a constraint '$d_{i_0 j_0} \mid \sum_{i \in I} d_i + \kappa$' if

$$\sum_{\substack{i \in I \\ p \mid d_i}} d_i \not\equiv 0 \pmod{d_{i_0 j_0}}.$$

That is, p appears in the sum $\sum_{i \in I} d_i$ and its contribution to this sum does not vanish modulo $d_{i_0 j_0}$. In particular, $p \neq d_{i_0 j_0}$.

Definition 2.11.4 is by no means the most general possible, but it is well adapted to the cases we will encounter.

2.11.2 Triangular systems of constraints

Our applications require bounding the number of vectors $\boldsymbol{d} \in (\pm\mathcal{D})^R$ satisfying systems with many constraints. Handling arbitrary systems of constraints is often intractable, as constraints are non-linear divisibility conditions to very large, possibly distinct moduli $d_{i_0 j_0}$. However, we can effectively derive meaningful bounds for systems of constraints that have a certain 'triangular' structure. Specifically, after a suitable ordering of the variables, each variable is essentially determined by its corresponding constraint together with the preceding variables.

Definition 2.11.5. A *triangular system with shift* κ on $\boldsymbol{d}$ is a sequence

$$C_1(\boldsymbol{d}), \dots, C_T(\boldsymbol{d})$$

of constraints on $\boldsymbol{d}$ with shift κ such that, for each $t \in [\![T]\!]$, there exists a prime p_t that is involved in $C_t(\boldsymbol{d})$ and absent from $C_1(\boldsymbol{d}), C_2(\boldsymbol{d}), \dots, C_{t-1}(\boldsymbol{d})$.

Our main task in the forthcoming sections will be to exhibit such triangular systems of constraints. Once we achieve this, we will be able to bound the weighted number of solutions to these systems using the following lemma.

Lemma 2.11.6. *Let $1 \leqslant T \leqslant R \leqslant 20K$. Let $B \geqslant 1$. Let $\mathbf{T} \subset (\pm\mathcal{D})^R$ be a set such that each $\boldsymbol{d} \in \mathbf{T}$ satisfies a triangular system of T constraints with shift $\kappa \in [-B, B]$ (thus, the system and its shift may depend on $\boldsymbol{d}$). Then*

$$\sum_{\boldsymbol{d} \in \mathbf{T}} \prod_{p \mid d_1 \cdots d_R} \frac{1}{p} \ll B K^{O(JR)} H_0^{-T/2}.$$

The proof consists of simple iterated substitutions, but is quite heavy on the notational side. The key takeaway is that each constraint in a triangular system produces a saving of size $\sqrt{H_0}$.

Proof. Let $\mathcal{C}$ denote the data consisting of:

- a sequence of signs $\sigma \in \{\pm 1\}^R$;
- a partition Π of $[\![R]\!] \times [\![J]\!]$;
- for each $t \in [\![T]\!]$, a set $I_t \subset [\![R]\!]$ which is a union of at most 10 discrete intervals;
- for each $t \in [\![T]\!]$, a pair $(i_t, j_t) \in [\![R]\!] \times [\![J]\!]$;
- an integer $\kappa \in [-B, B]$;
- a function $f : [\![T]\!] \to \Pi$.

We define $\mathbf{T}_{\mathcal{C}}$ to be the set of all $\boldsymbol{d} \in \mathbf{T}$ such that

- $\mathrm{sign}(d_i) = \sigma_i$ for $i \in [\![R]\!]$;
- for all $(i,j), (i',j') \in [\![R]\!] \times [\![J]\!]$, $d_{ij} = d_{i'j'}$ iff (i,j) and (i',j') are in the same class in Π;
- for every $t \in [\![T]\!]$, the constraint $C_t(\boldsymbol{d})$ defined by
$$d_{i_t j_t} \Big| \sum_{i \in I_t} d_i + \kappa$$
is satisfied by $\boldsymbol{d}$;
- for $t \in [\![T]\!]$, the prime $d_{f(t)}$ is involved in $C_t(\boldsymbol{d})$ but absent from $C_s(\boldsymbol{d})$ for $s < t$.[19]

Thus, each $\boldsymbol{d} \in \mathbf{T}$ is contained in $\mathbf{T}_{\mathcal{C}}$ for at least one choice of $\mathcal{C}$ as above.

We will show that, for each such choice of $\mathcal{C}$, we have

$$\sum_{\boldsymbol{d} \in \mathbf{T}_{\mathcal{C}}} \prod_{p | d_1 \cdots d_R} \frac{1}{p} \ll V^{RJ} H_0^{-T/2}. \tag{2.52}$$

This is enough to prove Lemma 2.11.6. Indeed, to bound the sum over $\boldsymbol{d} \in \mathbf{T}$, it suffices to multiply the right-hand side of (2.52) by the number of possibilities for $\mathcal{C}$. There are 2^R choices for σ. The number of partitions of $[\![R]\!] \times [\![J]\!]$ is $\leqslant (RJ)^{RJ}$. For $t \in [\![T]\!]$, since I_t is a union of at most 10 discrete intervals, it is uniquely determined by 20 elements of $[\![R]\!]$. Thus, the number of choices for $(I_t, i_t, j_t)_{t \in [\![T]\!]}$ is $\leqslant (R^{20} RJ)^T$. There are $\leqslant 2B + 1$ choices for $\kappa \in [-B, B] \cap \mathbb{Z}$. Any function $f : [\![T]\!] \to \Pi$ induces

[19] For $\alpha \in \Pi$, we write d_α for the prime d_{ij}, where (i,j) is any element of α; this is well-defined by construction.

a function $[\![T]\!] \to [\![R]\!] \times [\![J]\!]$ which uniquely determines f, so there are $\leqslant (RJ)^T$ possibilities for f. Therefore, assuming (2.52), we have

$$\sum_{\boldsymbol{d}\in\mathbf{T}} \prod_{p|d_1\cdots d_R} \frac{1}{p} \ll B\, 2^R (RJ)^{RJ+22T} V^{RJ} H_0^{-T/2}.$$

By Lemma 2.2.3, we have $V^J \ll K$. Using $T \leqslant R \ll K$, we can simplify the above to obtain

$$\sum_{\boldsymbol{d}\in\mathbf{T}} \prod_{p|d_1\cdots d_R} \frac{1}{p} \ll BK^{O(JR)} H_0^{-T/2},$$

as desired.

It remains to prove (2.52). Fix $\mathcal{C} = (\sigma, \Pi, (I_t), (i_t), (j_t), \kappa, f)$ such that $\mathbf{T}_{\mathcal{C}}$ is non-empty. Note that every class α of Π is contained in $[\![R]\!] \times \{j(\alpha)\}$ for some $j(\alpha) \in [\![J]\!]$, which is the unique integer such that $d_\alpha \in \mathcal{P}_{j(\alpha)}$ for all $\boldsymbol{d} \in \mathbf{T}_{\mathcal{C}}$.

Any $\boldsymbol{d} \in \mathbf{T}_{\mathcal{C}}$ is uniquely determined by the sequence of primes $(d_\alpha)_{\alpha\in\Pi}$.

We define a nested sequence $(\Pi_t)_{0\leqslant t\leqslant T}$ of subsets of Π that reveals the classes $f(t) \in \Pi$ one at a time, by setting

$$\Pi_0 := \Pi \setminus \{f(t) : t \in [\![T]\!]\} \quad \text{and, for } t \in [\![T]\!], \quad \Pi_t := \Pi_{t-1} \cup \{f(t)\}.$$

Let W_0 be the set of all sequences $\boldsymbol{p}_0 := (p_\alpha)_{\alpha\in\Pi_0}$ with $p_\alpha \in \mathcal{P}_{j(\alpha)}$ for all $\alpha \in \Pi_0$.

For any $t \in [\![T]\!]$ and any sequence of primes $\boldsymbol{p}_{<t} := (p_\alpha)_{\alpha\in\Pi_{t-1}}$, we define $W_t[\boldsymbol{p}_{<t}]$ to be the set of all primes $p \in \mathcal{P}_{j(f(t))}$ for which there is some $\boldsymbol{d} \in \mathbf{T}_{\mathcal{C}}$ such that $d_\alpha = p_\alpha$ for all $\alpha \in \Pi_{t-1}$ and $d_{f(t)} = p$.

Then, we have

$$\sum_{\boldsymbol{d}\in\mathbf{T}_{\mathcal{C}}} \prod_{p|d_1\cdots d_R} \frac{1}{p} = \sum_{\boldsymbol{p}_0\in W_0} \Big(\prod_{\alpha\in\Pi_0} \frac{1}{p_\alpha}\Big) \sum_{p_{f(1)}\in W_1} \frac{1}{p_{f(1)}} \sum_{p_{f(2)}\in W_2} \frac{1}{p_{f(2)}} \cdots \sum_{p_{f(T)}\in W_T} \frac{1}{p_{f(T)}}, \tag{2.53}$$

writing W_t instead of $W_t[\boldsymbol{p}_{<t}]$ to shorten notation.

Fix some $t \in [\![T]\!]$ and some sequence $\boldsymbol{p}_{<t} = (p_\alpha)_{\alpha\in\Pi_{t-1}}$. We claim that

$$\sum_{p\in W_t} \frac{1}{p} \ll \frac{\log H}{H_0}. \tag{2.54}$$

Recall that, for any $p \in W_t$, there is some $\boldsymbol{d} \in \mathbf{T}_{\mathcal{C}}$ with $d_\alpha = p_\alpha$ for all $\alpha \in \Pi_{t-1}$ and $d_{f(t)} = p$.

In particular, p is involved in the constraint $C_t(\boldsymbol{d})$ defined earlier, which means that

$$d_{i_t j_t} \mid Ap + B + \kappa \tag{2.55}$$

where

$$A := \frac{1}{p} \sum_{\substack{i\in I_t \\ p|d_i}} d_i \not\equiv 0 \pmod{d_{i_t j_t}} \qquad \text{and} \qquad B := \sum_{\substack{i\in I_t \\ p\nmid d_i}} d_i.$$

Observe that A and B are explicit expressions in the primes $\boldsymbol{p}_{<t}$. Indeed, $f(t)$ is of the form $f(t) = I'_t \times \{j(f(t))\}$ for some $I'_t \subset [\![R]\!]$, and we may rewrite

$$A = \sum_{i \in I_t \cap I'_t} \prod_{j \in [\![J]\!] \setminus \{j(f(t))\}} d_{ij} \qquad \text{and} \qquad B = \sum_{i \in I_t \setminus I'_t} \prod_{j \in [\![J]\!]} d_{ij}.$$

By definition of $f(t)$ and I'_t, the prime $p = d_{f(t)}$ does not appear in A or B. Furthermore, by construction, the primes $d_{f(t+1)}, \dots, d_{f(T)}$ are absent from $C_t(\boldsymbol{d})$, which means that $d_{f(t+1)}, \dots, d_{f(T)}$ cannot be any of the primes d_{ij} occurring in A or B either. Hence, A and B are fully determined by the primes $\boldsymbol{p}_{<t} = (p_\alpha)_{\alpha \in \Pi_{t-1}}$.

Thus, this divisibility condition (2.55) uniquely determines the congruence class of p modulo $d_{i_t j_t}$. Using that $\mathcal{P} \subset (H_0, H)$, we have, for any x,

$$\sum_{\substack{p \in \mathcal{P} \\ p \equiv x \,(\mathrm{mod}\, d_{i_t j_t})}} \frac{1}{p} \leqslant \max_a \sum_{\substack{H_0 \leqslant n \leqslant H \\ n \equiv a \,(\mathrm{mod}\, \lfloor H_0 \rfloor)}} \frac{1}{n} \ll \frac{\log H}{H_0},$$

which proves (2.54).

Applying (2.54) in (2.53) successively for $t = T, T-1, \dots, 1$ yields

$$\sum_{\boldsymbol{d} \in \mathbf{T}_{\mathcal{C}}} \prod_{p \mid d_1 \cdots d_R} \frac{1}{p} \leqslant \left(\frac{O(\log H)}{H_0} \right)^T \sum_{\boldsymbol{p}_0 \in W_0} \prod_{\alpha \in \Pi_0} \frac{1}{p_\alpha} \ll H_0^{-T/2} V^{|\Pi_0|}.$$

Equation (2.52) follows, which concludes the proof of Lemma 2.11.6. □

2.12 Unpredictable walks

The goal of this section is to prove the following proposition, which states that the contribution of tuples $\boldsymbol{d}$ arising from unpredictable words is negligible.

Proposition 2.12.1. *Let $\mathcal{S}$, $\mathcal{L}$, $\mathcal{U}$ be sets such that $[\![2K]\!] \times [\![J]\!] = \mathcal{S} \sqcup \mathcal{L} \sqcup \mathcal{U}$ and $|\mathcal{U}| \leqslant K^{2\varepsilon_1}$. Let $1 \leqslant r \leqslant K$. Then*

$$\sum_{\boldsymbol{d} \in \mathbf{Unp}_{\mathcal{S},\mathcal{L},r}} \prod_{p \mid \boldsymbol{d}_*} \frac{1}{p} \ll 1.$$

Our strategy is as follows. Whenever a prime is repeated at lit indices, part (2) of Definition 2.8.4 yields a divisibility condition on $\boldsymbol{d}$. These conditions restrict the possibilities for $\boldsymbol{d}$: generically, we might hope to win a factor $\approx H_0$ from each such condition, which would be more than sufficient. Unfortunately, the multitude of dependencies between these conditions makes it hard to rule out highly degenerate systems.

However, since H_0 is much larger than K, it is enough to win a moderate number of factors H_0 to beat the trivial bound of Lemma 2.8.8. To do so, we extract from

the original system of lit conditions a subset that is trivially non-singular, namely a *triangular system*. The existence of such a triangular system will ultimately follow from the unpredictability of the words associated to $\boldsymbol{d} \in \mathbf{Unp}_{\mathcal{S},\mathcal{L},r}$.

2.12.1 Structure of unpredictable words

Our intermediate aim is to prove Proposition 2.12.10, which states that t-unpredictable words must contain one of several special patterns. These patterns will allow us to extract large triangular systems from the divisibility conditions associated to the lit indices, when $\boldsymbol{d} \in \mathbf{Unp}_{\mathcal{S},\mathcal{L},r}$.

Recall that $\mathcal{W}$ denotes the set of words on the alphabet $\mathcal{A}$ with no two consecutive equal letters, and $\mathcal{W}^{\neq} \subset \mathcal{W}$ is the set of words with distinct letters.

Definition 2.12.2. A word $w \in \mathcal{W}$ contains n *separated repetitions* if it has a substring of the form

$$\mathtt{A}_1 \ldots \mathtt{A}_1 \ldots \mathtt{A}_2 \ldots \mathtt{A}_2 \ldots \cdots \ldots \mathtt{A}_n \ldots \mathtt{A}_n,$$

for some not necessarily distinct letters $\mathtt{A}_1, \ldots, \mathtt{A}_n$. The three dots $\ldots$ represent a string of letters of arbitrary length (possibly empty). In other words, there are $k_1 < l_1 < k_2 < \ldots < k_n < l_n$ such that $w[k_i] = w[l_i]$ for all i.

Lemma 2.12.3. *Let $m \geqslant 10$. Let $k_1 < k_2 < \ldots < k_m$ be positive integers. Let $w \in \mathcal{W}$ be a word of length $\geqslant k_m$. Then, either w contains $\gg m^{1/2}$ separated repetitions, or there are $i, j \in [\![m]\!]$ with $j - i \gg m^{1/2}$ such that the substring*

$$w[k_i]w[k_i + 1] \cdots w[k_j]$$

of w has distinct letters.

Proof. Let $n = \lfloor m^{1/2} \rfloor$. If the second conclusion does not hold, there must be a repeated letter in the substring $w[k_{rn+1}]w[k_{rn+1} + 1] \cdots w[k_{(r+1)n}]$, for each $0 \leqslant r \leqslant n - 1$. This implies that w contains n separated repetitions. □

Lemma 2.12.4. *Let $\mathtt{A}, \mathtt{B}, \mathtt{C} \in \mathcal{A}$. Let $w_1, w_2 \in \mathcal{W}^{\neq}$ be two words of the form*

$$\mathtt{A} \ldots \mathtt{B} \ldots \mathtt{C}.$$

Suppose that $\mathtt{B}$ has variable neighbours in the concatenation $w_1 w_2$.[20]

Then, there exist substrings $v_1 \sqsubset w_1$ and $v_2 \sqsubset w_2$, both of the form $\mathtt{A} \ldots \mathtt{Y}$ for some letter $\mathtt{Y}$ (possibly equal to $\mathtt{B}$ or $\mathtt{C}$), with distinct sets of letters, i.e. $v_1[*] \neq v_2[*]$.

[20] This just means that the two letters adjacent to $\mathtt{B}$ in w_1 are not the same as the two letters adjacent to $\mathtt{B}$ in w_2

Proof. If $w_1[*] \neq w_2[*]$, we can just take $v_1 := w_1$, $v_2 := w_2$ and $\mathtt{Y} := \mathtt{C}$.

Otherwise, w_1 and w_2 have the same sets of letters, and thus the same length, as $w_1, w_2 \in \mathcal{W}^{\neq}$. Let $k \geqslant 2$ be minimal such that $w_1[k] \neq w_2[k]$. We know that k exists, since $w_1 \neq w_2$. Let $\mathtt{X} := w_1[k]$ and $\mathtt{Y} := w_2[k]$. The letter $\mathtt{Y}$ is present in w_1 as both words have the same letters. By minimality of k, we must have $\mathtt{Y} = w_1[l]$ for some $l > k$. Set $v_1 := w_1[1]w_1[2]\cdots w_1[l] = \mathtt{A}\ldots\mathtt{X}\ldots\mathtt{Y}$ and $v_2 := w_2[1]w_2[2]\cdots w_2[k] = \mathtt{A}\ldots\mathtt{Y}$. Then $v_1[*] \neq v_2[*]$ as $\mathtt{X} \in v_1[*] \setminus v_2[*]$. □

Notation 2.12.5. Let $w \in \mathcal{W}^{\neq}$, and suppose that w is of the form $\mathtt{A}\ldots\mathtt{X}\ldots\mathtt{Y}\ldots\mathtt{B}$. We write $w|_{\mathtt{X}\ldots\mathtt{Y}}$ for the unique substring of w of the form $\mathtt{X}\ldots\mathtt{Y}$. This is well-defined as w has distinct letters.

Lemma 2.12.6. *Let $m \geqslant 1$. Let $\mathtt{A}_0, \ldots, \mathtt{A}_{2m} \in \mathcal{A}$. Let $w_1, w_2 \in \mathcal{W}^{\neq}$ be two words of the form*

$$\mathtt{A}_0\ldots\mathtt{A}_1\ldots\mathtt{A}_2\ldots\cdots\ldots\mathtt{A}_{2m}$$

such that, for all $1 \leqslant i \leqslant 2m-1$, the letter $\mathtt{A}_i$ has variable neighbours in the concatenation w_1w_2.

Then, there are substrings $v_1 \sqsubset w_1$ and $v_2 \sqsubset w_2$, both of the form

$$\mathtt{Y}_0\ldots\mathtt{Y}_1\ldots\mathtt{Y}_2\ldots\cdots\ldots\mathtt{Y}_m,$$

for some letters $\mathtt{Y}_0, \ldots, \mathtt{Y}_m$ (possibly equal to some of the $\mathtt{A}_i$) such that, for all $j \in [\![m]\!]$, the sets of letters of $v_1|_{\mathtt{Y}_{j-1}\ldots\mathtt{Y}_j}$ and $v_2|_{\mathtt{Y}_{j-1}\ldots\mathtt{Y}_j}$ are distinct.

Proof. Set $\mathtt{Y}_0 := \mathtt{A}_0$. Let $w_1^1 := w_1|_{\mathtt{Y}_0\ldots\mathtt{A}_1\ldots\mathtt{A}_2}$ and $w_2^1 := w_2|_{\mathtt{Y}_0\ldots\mathtt{A}_1\ldots\mathtt{A}_2}$. Applying Lemma 2.12.4 to these words w_1^1 and w_2^1, we find two further substrings $v_1^1 \sqsubset w_1^1$ and $v_2^1 \sqsubset w_2^1$ of the form $\mathtt{Y}_0\ldots\mathtt{Y}_1$ for some common ending letter $\mathtt{Y}_1$, such that $v_1^1[*] \neq v_2^1[*]$. Notice that w_1 and w_2 are of the form

$$\mathtt{Y}_0\ldots\mathtt{Y}_1\ldots\mathtt{A}_3\ldots\mathtt{A}_4\ldots\cdots\ldots\mathtt{A}_{2m}.$$

We may thus define the substrings $w_1^2 := w_1|_{\mathtt{Y}_1\ldots\mathtt{A}_3\ldots\mathtt{A}_4}$ and $w_2^2 := w_2|_{\mathtt{Y}_1\ldots\mathtt{A}_3\ldots\mathtt{A}_4}$. Applying Lemma 2.12.4 again with w_1^2 and w_2^2, we obtain two substrings $v_1^2 \sqsubset w_1^2$ and $v_2^2 \sqsubset w_2^2$ of the form $\mathtt{Y}_1\ldots\mathtt{Y}_2$, with $v_1^2[*] \neq v_2^2[*]$. In particular, w_1 and w_2 can now be written as

$$\mathtt{Y}_0\ldots\mathtt{Y}_1\ldots\mathtt{Y}_2\ldots\mathtt{A}_5\ldots\mathtt{A}_6\ldots\cdots\ldots\mathtt{A}_{2m}.$$

We can repeat this process; after m applications of Lemma 2.12.4, we obtain substrings of w_1 and w_2 of the form $\mathtt{Y}_0\ldots\mathtt{Y}_1\ldots\mathtt{Y}_2\ldots\cdots\ldots\mathtt{Y}_m$ with the required properties. □

Lemma 2.12.7. *Let $m \geqslant 1$. Let $\mathtt{Y}_0, \dots, \mathtt{Y}_{4m} \in \mathcal{A}$. Let $w_1, w_2 \in \mathcal{W}^{\neq}$ be two words of the form*

$$\mathtt{Y}_0 \dots \mathtt{Y}_1 \dots \mathtt{Y}_2 \dots \cdots \dots \mathtt{Y}_{4m}.$$

Suppose that, for all $j \in [\![4m]\!]$, the words $w_1|_{\mathtt{Y}_{j-1}\dots\mathtt{Y}_j}$ and $w_2|_{\mathtt{Y}_{j-1}\dots\mathtt{Y}_j}$ have distinct sets of letters.

Then, there is a pair of words $(w_1', w_2') \in \{(w_1, w_2), (w_2, w_1), (\overleftarrow{w_1}, \overleftarrow{w_2}), (\overleftarrow{w_2}, \overleftarrow{w_1})\}$ with the following properties.

There are letters $\mathtt{X}_1, \dots, \mathtt{X}_m, \mathtt{Z}_0, \mathtt{Z}_1, \dots, \mathtt{Z}_m, \mathtt{Z}_{m+1}$ (possibly equal to some of the $\mathtt{Y}_j$) such that w_1' is of the form

$$\mathtt{Z}_0 \dots \mathtt{X}_1 \dots \mathtt{Z}_1 \dots \mathtt{X}_2 \dots \mathtt{Z}_2 \dots \cdots \dots \mathtt{Z}_{m-1} \dots \mathtt{X}_m \dots \mathtt{Z}_m \dots \mathtt{Z}_{m+1},$$

w_2' is of the form

$$\mathtt{Z}_0 \dots \mathtt{Z}_1 \dots \mathtt{Z}_2 \dots \cdots \dots \mathtt{Z}_m \dots \mathtt{Z}_{m+1},$$

and, for all $j \in [\![m]\!]$, the letter $\mathtt{X}_j$ does not appear in $w_2'|_{\mathtt{Z}_0\dots\mathtt{Z}_j}$.

Proof. Let J_1 be the set of all $j \in [\![4m]\!]$ such that $w_1|_{\mathtt{Y}_{j-1}\dots\mathtt{Y}_j}$ contains a letter not appearing in $w_2|_{\mathtt{Y}_{j-1}\dots\mathtt{Y}_j}$. Similarly, let J_2 be the set of all $j \in [\![4m]\!]$ such that $w_2|_{\mathtt{Y}_{j-1}\dots\mathtt{Y}_j}$ has a letter that is not present in $w_1|_{\mathtt{Y}_{j-1}\dots\mathtt{Y}_j}$. By assumption, $J_1 \cup J_2 = [\![4m]\!]$, so one of J_1 and J_2 has size $\geqslant 2m$. Without loss of generality, assume that $|J_1| \geqslant 2m$, swapping w_1 and w_2 if necessary.

Let $j \in J_1$, and let $\mathtt{X}$ be a letter present in $w_1|_{\mathtt{Y}_{j-1}\dots\mathtt{Y}_j}$ but not in $w_2|_{\mathtt{Y}_{j-1}\dots\mathtt{Y}_j}$. The letter $\mathtt{X}$ could possibly appear in $w_2|_{\mathtt{Y}_0\dots\mathtt{Y}_{j-1}}$ or in $w_2|_{\mathtt{Y}_j\dots\mathtt{Y}_{4m}}$, but not in both as $w_2 \in \mathcal{W}^{\neq}$.

We define $J_1^<$ to be the set of all $j \in J_1$ for which there exists a letter $\mathtt{X}$ present in $w_1|_{\mathtt{Y}_{j-1}\dots\mathtt{Y}_j}$ but not in $w_2|_{\mathtt{Y}_0\dots\mathtt{Y}_j}$. Similarly, we define $J_1^>$ to be the set of all $j \in J_1$ for which there exists a letter $\mathtt{X}$ of $w_1|_{\mathtt{Y}_{j-1}\dots\mathtt{Y}_j}$ not appearing in $w_2|_{\mathtt{Y}_{j-1}\dots\mathtt{Y}_{4m}}$. By the previous observation, we have $J_1^< \cup J_1^> = J_1$, so one of $J_1^<$ and $J_1^>$ has size $\geqslant m$. Considering the reversed words if necessary, we may assume without loss of generality that $|J_1^<| \geqslant m$.

Let $j_1 < j_2 < \dots < j_m$ be elements of $J_1^<$. For $i \in [\![m]\!]$, let $\mathtt{X}_i$ be a letter of $w_1|_{\mathtt{Y}_{j_i-1}\dots\mathtt{Y}_{j_i}}$ not appearing in the substring $w_2|_{\mathtt{Y}_0\dots\mathtt{Y}_{j_i}}$ of w_2. Then w_1 is of the form

$$\mathtt{Y}_0 \dots \mathtt{X}_1 \dots \mathtt{Y}_{j_1} \dots \mathtt{X}_2 \dots \mathtt{Y}_{j_2} \dots \cdots \dots \mathtt{Y}_{j_{m-1}} \dots \mathtt{X}_m \dots \mathtt{Y}_{j_m} \dots \mathtt{Y}_{4m}.$$

The lemma follows, defining $\mathtt{Z}_0 := \mathtt{Y}_0$, $\mathtt{Z}_{m+1} := \mathtt{Y}_{4m}$ and $\mathtt{Z}_i := \mathtt{Y}_{j_i}$ for all $i \in [\![m]\!]$. □

Combining Lemma 2.12.6 and Lemma 2.12.7, we immediately obtain the following.

Lemma 2.12.8. *Let $m \geqslant 1$. Let $\mathtt{A}_0, \dots, \mathtt{A}_{8m} \in \mathcal{A}$. Let $w_1, w_2 \in \mathcal{W}^{\neq}$ be two words of the form*

$$\mathtt{A}_0 \dots \mathtt{A}_1 \dots \mathtt{A}_2 \dots \cdots \dots \mathtt{A}_{8m}.$$

Suppose that, for $1 \leqslant i \leqslant 8m-1$, the letter $\mathtt{A}_i$ has variable neighbours in the concatenation $w_1 w_2$.

Then, the following holds, after possibly replacing (w_1, w_2) by an element of $\{(w_1, w_2), (w_2, w_1), (\overleftarrow{w_1}, \overleftarrow{w_2}), (\overleftarrow{w_2}, \overleftarrow{w_1})\}$.

For some letters $\mathtt{X}_1, \dots, \mathtt{X}_m, \mathtt{Z}_0, \mathtt{Z}_1, \dots, \mathtt{Z}_m$ (possibly equal to some of the $\mathtt{A}_j$), there are words $v_1 \sqsubset w_1$ and $v_2 \sqsubset w_2$, with v_1 of the form

$$\mathtt{Z}_0 \dots \mathtt{X}_1 \dots \mathtt{Z}_1 \dots \mathtt{X}_2 \dots \mathtt{Z}_2 \dots \cdots \dots \mathtt{Z}_{m-1} \dots \mathtt{X}_m \dots \mathtt{Z}_m$$

and v_2 of the form

$$\mathtt{Z}_0 \dots \mathtt{Z}_1 \dots \mathtt{Z}_2 \dots \cdots \dots \mathtt{Z}_m,$$

such that, for all $j \in [\![m]\!]$, the letter $\mathtt{X}_j$ does not appear in the substring $v_2|_{\mathtt{Z}_0 \dots \mathtt{Z}_j}$.

It is a well-known combinatorial fact that from any sequence of n distinct real numbers one can always extract an increasing or decreasing subsequence of length $\gg \sqrt{n}$. We will use a similar result about *pairs* of real numbers.

Lemma 2.12.9. *Let S be a set of n pairs of real numbers, such that*

- *if $(a, b) \in S$ then $a < b$, and*
- *if $(a, b), (c, d) \in S$ are two distinct pairs, then $\{a, b\} \cap \{c, d\} = \emptyset$.*

Then, there exists $S' \subset S$ of size $n' \geqslant n^{1/4}$ such that one of the following holds.[21]

(i) $S' = \{(a_1, b_1), (a_2, b_2), \dots, (a_{n'}, b_{n'})\}$ *for some* $a_1 < b_1 < a_2 < b_2 < \cdots < b_{n'}$.

(ii) $S' = \{(a_1, b_1), (a_2, b_2), \dots, (a_{n'}, b_{n'})\}$ *for some*

$$a_1 < a_2 < \cdots < a_{n'} < b_1 < b_2 < \cdots < b_{n'}.$$

(iii) $S' = \{(a_1, b_1), (a_2, b_2), \dots, (a_{n'}, b_{n'})\}$ *for some*

$$a_1 < a_2 < \cdots < a_{n'} < b_{n'} < b_{n'-1} < \cdots < b_1.$$

Proof. Define a strict partial order $\prec^1$ on S by setting $(a, b) \prec^1 (c, d)$ iff $b < c$. A well-known consequence of Dilworth's theorem states that any partially ordered set on n elements contains a chain or an antichain[22] of size $\geqslant n^{1/2}$ (see [56, Proposition 2.5.9]).

[21] The bound $n' \geqslant n^{1/4}$ can be improved, but that is not relevant for us.

[22] Recall that a *chain* is a totally ordered subset of a partially ordered set, and an *antichain* is a subset in which no two elements are comparable.

If S contains a chain of size $\geqslant n^{1/2}$ for $\prec^1$, we are in case (i). Suppose that S contains an antichain A of size $\geqslant n^{1/2}$. We introduce another strict partial order $\prec^2$ on A by defining $(a,b) \prec^2 (c,d)$ iff $a < c < d < b$. By the same combinatorial fact, either A contains a chain for $\prec^2$ of size $\geqslant n^{1/4}$, and case (iii) applies, or A contains an antichain A' for $\prec^2$ of size $\geqslant n^{1/4}$. Suppose that the latter possibility occurs. Let $(a_1,b_1),\dots,(a_{n'},b_{n'})$ be the elements of A', with $a_1 < a_2 < \cdots < a_{n'}$. Since A' is an antichain for $\prec^1$, all the b_i are greater than $a_{n'}$. Since A' is also an antichain for $\prec^2$, we deduce that $b_1 < b_2 < \cdots < b_{n'}$, and we are in case (ii). □

We will combine the previous lemmas to extract useful substructures in unpredictable words.

Proposition 2.12.10. *There is an absolute constant $c_1 > 0$ such that the following holds.*

Let $n, t \geqslant 10$. Let $w \in \mathcal{W}_n$ be a t-unpredictable word. Then, for some $m \gg t^{c_1}$, at least one of the properties below is satisfied.

1. *w has m separated repetitions.*
2. *There are words v_1, v_2 with all of the following properties:*
   - *(i) $v_1 \in \mathcal{W}^{\neq}$ and $v_2 \in \mathcal{W}^{\neq}$;*
   - *(ii) $v_1 \sqsubset w$ or $\overleftarrow{v_1} \sqsubset w$;*
   - *(iii) $v_2 \sqsubset w$ or $\overleftarrow{v_2} \sqsubset w$;*
   - *(iv) there are letters $\mathtt{X}_1,\dots,\mathtt{X}_m,\mathtt{Z}_0,\dots,\mathtt{Z}_m$ such that v_1 is of the form*

$$\mathtt{Z}_0\dots\mathtt{X}_1\dots\mathtt{Z}_1\dots\mathtt{X}_2\dots\mathtt{Z}_2\dots\cdots\dots\mathtt{Z}_{m-1}\dots\mathtt{X}_m\dots\mathtt{Z}_m$$

   and v_2 is of the form

$$\mathtt{Z}_0\dots\mathtt{Z}_1\dots\mathtt{Z}_2\dots\cdots\dots\mathtt{Z}_m.$$

   Moreover, for all $j \in [\![m]\!]$, the letter $\mathtt{X}_j$ does not appear in $v_2|_{\mathtt{Z}_0\dots\mathtt{Z}_j}$.

Proof. By definition of unpredictability, either w contains a letter repeated $> t$ times, or it has $> t$ letters with variable neighbours. In the first case, we immediately see that w has $\lfloor t/2 \rfloor$ separated repetitions.

Suppose now that there are $> t$ letters with variable neighbours in w. Let E be the set of all these letters, with the possible exception of the first and last letters of w which are discarded (to simplify the notation below). Thus, $|E| \geqslant t-2$. For every letter $\mathtt{A} \in E$, there are two positions $1 < k_{\mathtt{A}} < l_{\mathtt{A}} < n$ such that $w[k_{\mathtt{A}}] = w[l_{\mathtt{A}}] = \mathtt{A}$, and the sets of letters adjacent to these two occurrences of $\mathtt{A}$ are different, i.e.

$$\{w[k_{\mathtt{A}}-1], w[k_{\mathtt{A}}+1]\} \neq \{w[l_{\mathtt{A}}-1], w[l_{\mathtt{A}}+1]\}.$$

We apply Lemma 2.12.9 to the set $S = \{(k_{\mathtt{A}}, l_{\mathtt{A}}) : \mathtt{A} \in E\}$. If case (i) occurs, we can immediately conclude that w has $\gg t^{1/4}$ separated repetitions and we are done.

Suppose that case (ii) of Lemma 2.12.9 applies. This implies that, for some $c \gg 1$, there exists a subset

$$F = \{\mathtt{A}_1, \dots, \mathtt{A}_{|F|}\} \subset E$$

of size $|F| \geqslant t^c$ such that

$$1 < k_{\mathtt{A}_1} < k_{\mathtt{A}_2} < \cdots < k_{\mathtt{A}_{|F|}} < l_{\mathtt{A}_1} < l_{\mathtt{A}_2} < \cdots < l_{\mathtt{A}_{|F|}} < n.$$

By Lemma 2.12.3, either w has $\gg t^{c/2}$ separated repetitions, and the first conclusion holds, or we can find a 'large' substring of $w[k_{\mathtt{A}_1}]w[k_{\mathtt{A}_1}+1]\cdots w[k_{\mathtt{A}_{|F|}}]$ with distinct letters. Without loss of generality (by replacing F with a smaller subset, c with a smaller absolute constant and relabelling the letters), we may thus assume that the word $w[k_{\mathtt{A}_1}]w[k_{\mathtt{A}_1}+1]\cdots w[k_{\mathtt{A}_{|F|}}]$ itself has distinct letters. By a further application of Lemma 2.12.3, we may also assume that the word $w[l_{\mathtt{A}_1}]w[l_{\mathtt{A}_1}+1]\cdots w[l_{\mathtt{A}_{|F|}}]$ has distinct letters.

We apply Lemma 2.12.8 with the words $w_1 := w[k_{\mathtt{A}_1}]w[k_{\mathtt{A}_1}+1]\cdots w[k_{\mathtt{A}_{|F|}}]$ and $w_2 := w[l_{\mathtt{A}_1}]w[l_{\mathtt{A}_1}+1]\cdots w[l_{\mathtt{A}_{|F|}}]$. Since both w_1 and w_2 are words in $\mathcal{W}^{\neq}$ of the form

$$\mathtt{A}_1 \dots \mathtt{A}_2 \dots \mathtt{A}_3 \dots \cdots \dots \mathtt{A}_{|F|},$$

the assumptions of Lemma 2.12.8 are satisfied (of course, we may assume without loss of generality that $|F| \equiv 1 \pmod 8$). The conclusion of Lemma 2.12.8 provides us with two words v_1 and v_2 precisely satisfying the second conclusion of Proposition 2.12.10.

The treatment of case (iii) of Lemma 2.12.9 is similar. For some $c \gg 1$, there exists a subset

$$F = \{\mathtt{A}_1, \dots, \mathtt{A}_{|F|}\} \subset E$$

of size $|F| \geqslant t^c$ such that

$$1 < k_{\mathtt{A}_1} < k_{\mathtt{A}_2} < \cdots < k_{\mathtt{A}_{|F|}} < l_{\mathtt{A}_{|F|}} < l_{\mathtt{A}_{|F|-1}} < \cdots < l_{\mathtt{A}_1} < n.$$

By two successive applications of Lemma 2.12.3, we may assume, without loss of generality, that $w_1 := w[k_{\mathtt{A}_1}]w[k_{\mathtt{A}_1}+1]\cdots w[k_{\mathtt{A}_{|F|}}]$ and $w_2 := w[l_{\mathtt{A}_{|F|}}]w[l_{\mathtt{A}_{|F|}}-1]\cdots w[l_{\mathtt{A}_1}]$ each have distinct letters. Then, applying Lemma 2.12.8 with these two substrings w_1 and w_2 produces two words v_1 and v_2 with the required properties. □

2.12.2 Triangular systems for unpredictable walks

We now show that tuples $\boldsymbol{d}$ whose associated words are unpredictable (and thus satisfy the conclusion of Proposition 2.12.10) admit triangular systems with many constraints.

For the rest of this section, we use the following notation.

Notation 2.12.11. Let $R \geqslant 1$ and $\boldsymbol{d} \in (\pm\mathcal{D})^R$. For $I \subset [\![R]\!]$, we write $\boldsymbol{d}_I := \prod_{i\in I} d_i$.

Definition 2.12.12 ($\mathbf{Red}_{R,\mathcal{S},\mathcal{L}}$). Let $R \geqslant 1$ and $\mathcal{S}, \mathcal{L}$ be disjoint subsets of $[\![R]\!] \times [\![J]\!]$.

We define $\mathbf{Red}_{R,\mathcal{S},\mathcal{L}}$ to be the set of all $\mathcal{L}$-admissible reduced tuples[23] $\boldsymbol{d} \in (\pm\mathcal{D})^R$ such that:

1. For all $(i,j) \in \mathcal{S}$, we have $d_{ij}^2 \nmid \boldsymbol{d}_{[\![R]\!]}$.
2. Whenever two indices $(i,j), (i',j) \in \mathcal{L}$ satisfy $d_{ij} = d_{i'j}$, we have
$$d_{ij} \,\Big|\, \sum_{i<k<i'} d_k.$$

Definition 2.12.13. Let $\boldsymbol{d} \in \mathbf{Red}_{R,\mathcal{S},\mathcal{L}}$. We say that $(x,y,p) \in [\![R]\!]^2 \times \mathcal{P}$ is a *divisibility triple* for $\boldsymbol{d}$ if $p \mid d_x$, $p \mid d_y$, and there exists $x < i < y$ such that $p \nmid d_i$.[24]

Lemma 2.12.14. *Let $1 \leqslant R < H_0$ and $\mathcal{S}, \mathcal{L}$ be disjoint subsets of $[\![R]\!] \times [\![J]\!]$. Let $\boldsymbol{d} \in \mathbf{Red}_{R,\mathcal{S},\mathcal{L}}$.*

Let (x,y,p) be a divisibility triple for $\boldsymbol{d}$. Then, one of the following holds.

1. *There are $x', y' \in [\![x,y]\!]$ and $p' \in \mathcal{P}$ such that (x',y',p') is a divisibility triple for $\boldsymbol{d}$ and $|y'-x'| < |y-x|$.*
2. *There exists $q \in \mathcal{P}$ dividing d_{y-1}, such that*
$$\sum_{\substack{x<i<y \\ q\mid d_i}} d_i \not\equiv 0 \pmod{p}.$$

Proof. Suppose that case (1) does not hold. In particular, this implies that $p \nmid d_{y-1}$.

Let k be the smallest integer with $x < k < y$ such that
$$|d_k| = |d_{k+1}| = \ldots = |d_{y-1}|.$$
Since $\boldsymbol{d}$ is reduced, we in fact have $d_k = d_{k+1} = \ldots = d_{y-1}$.

If $k = x+1$, we can directly compute, for any prime $q \mid d_{y-1}$,
$$\sum_{\substack{x<i<y \\ q\mid d_i}} d_i = (y-x-1)d_{y-1},$$
which is not divisible by p since $p \nmid d_{y-1}$ and $0 < y-x-1 \leqslant R < H_0 \leqslant p$. Hence, case (2) holds.

[23] See Definition 2.7.2 and Definition 2.8.3 for the definitions of reduced tuples and $\mathcal{L}$-admissibility.
[24] In particular, $y \geqslant x+2$.

Suppose now that $k > x+1$. Then, since $|d_{k-1}| \neq |d_k|$, there is a prime q dividing $d_k = d_{y-1}$ but not d_{k-1}. Note that q cannot divide any d_i with $x < i < k-1$, since otherwise (i, k, q) would be a divisibility triple satisfying case (1). Therefore,

$$\sum_{\substack{x<i<y \\ q|d_i}} d_i = (y-k)d_{y-1}.$$

As in the previous case, this is not divisible by p, so case (2) holds. □

The following lemma corresponds to the tuples $\boldsymbol{d}$ that satisfy the first conclusion of Proposition 2.12.10, i.e. tuples having an associated word with m separated repetitions.

Lemma 2.12.15. *Let $1 \leqslant R < H_0$ and let $\mathcal{S}, \mathcal{L}$ be disjoint subsets of $[\![R]\!] \times [\![J]\!]$.*

Let $\boldsymbol{d} \in \mathbf{Red}_{R,\mathcal{S},\mathcal{L}}$. Suppose that, for some $m \geqslant 4R/L$, there exist

$$1 \leqslant x_1 < y_1 < x_2 < y_2 < \ldots < x_m < y_m \leqslant R$$

and $p_1, \ldots, p_m \in \mathcal{P}$ such that (x_i, y_i, p_i) is a divisibility triple for $\boldsymbol{d}$, for all $i \in [\![m]\!]$.

Assume that $\mathcal{S} \sqcup \mathcal{L} \supset [\![x_1, y_m]\!] \times [\![J]\!]$.

Then, $\boldsymbol{d}$ satisfies a triangular system of $\gg m$ constraints with shift $\kappa = 0$.

Proof. Without loss of generality, we may assume that, for each $i \in [\![m]\!]$, the divisibility triple (x_i, y_i, p_i) is minimal, in the sense that there is no divisibility triple (x', y', p') with $x_i \leqslant x' < y' \leqslant y_i$ and $|y' - x'| < |y_i - x_i|$.

Thus, by Lemma 2.12.14, for each $i \in [\![m]\!]$, there is a prime $q_i \in \mathcal{P}$ dividing d_{y_i-1} such that

$$\sum_{\substack{x_i<k<y_i \\ q_i|d_k}} d_k \not\equiv 0 \pmod{p_i}. \tag{2.56}$$

Each (x_i, y_i, p_i) is a divisibility triple for $\boldsymbol{d}$, which implies that $p_i = d_{x_i j_i} = d_{y_i j_i}$ for some $j_i \in [\![J]\!]$. Since every element of $[\![x_1, y_m]\!] \times [\![J]\!]$ is either in $\mathcal{S}$ or in $\mathcal{L}$, and all elements of $\mathcal{S}$ correspond to primes that appear only once, it must be that $(x_i, j_i), (y_i, j_i) \in \mathcal{L}$. Therefore, by Definition 2.12.12, we have

$$p_i = d_{x_i j_i} \Bigg| \sum_{x_i<k<y_i} d_k \tag{2.57}$$

for every $i \in [\![m]\!]$.

The divisibility relation (2.57) defines a constraint C_i on $\boldsymbol{d}$ with shift $\kappa = 0$ (see Definition 2.11.2). Moreover, q_i is involved in this constraint by (2.56) (see Definition 2.11.4). We now perform a case analysis to extract a triangular system of constraints.

Let I_0 be the set of all $i \in [\![m]\!]$ such that q_i does not divide $\prod_{k<i} \boldsymbol{d}_{[\![x_k, y_k]\!]}$.

Suppose that $|I_0| \geqslant m/4$. If $i \in I_0$, we know that q_i is absent from the constraints C_k with $k < i$. Therefore, $(C_i)_{i\in I_0}$ forms a triangular system of $\gg m$ constraints satisfied by $\boldsymbol{d}$, so we are done. Henceforth, we assume that $|I_0| < m/4$.

Let I_1 be the set of all $i \in [\![m]\!]$ such that $[\![x_i + 1, y_i - 1]\!] \times [\![J]\!]$ contains a pair $(s_i, t_i) \in \mathcal{S}$.

Suppose that $|I_1| \geqslant m/4$. Notice that $d_{s_it_i}$ is involved in the constraint C_i, as

$$\sum_{\substack{x_i<k<y_i \\ d_{s_it_i}|d_k}} d_k = d_{s_i} \not\equiv 0 \pmod{p_i}.$$

Here we used that $(s_i, t_i) \in \mathcal{S}$ for the first equality and the minimality of (x_i, y_i, p_i) to say that $p_i \nmid d_{s_i}$. In addition, $d_{s_it_i}$ is absent from all other constraints $(C_{i'})_{i'\neq i}$, since $(s_i, t_i) \in \mathcal{S}$. Thus, $(C_i)_{i\in I_1}$ is a triangular system of $\gg m$ constraints satisfied by $\boldsymbol{d}$, as desired. We now assume that $|I_1| < m/4$.

Let I_2 be the set of all $i \in [\![m]\!]$ such that $\{x_i\} \times [\![J]\!]$ contains a pair $(s_i, t_i) \in \mathcal{S}$ (thus $s_i = x_i$).

Suppose that $|I_2| \geqslant m/2$. Then $|I_2 \setminus I_0| \geqslant m/4$. This time, we will use a different sequence of constraints. Let $i \in I_2 \setminus I_0$. By definition of I_0, we know that there exists

$$u_i \in \bigcup_{i'<i} [\![x_{i'}, y_{i'}]\!]$$

such that $q_i \mid d_{u_i}$. We also know that $q_i \mid d_{v_i}$ for some $x_i < v_i < y_i$ by (2.56). Hence, by Definition 2.12.12, we obtain

$$q_i \,\Big|\, \sum_{u_i<k<v_i} d_k,$$

which is a constraint on $\boldsymbol{d}$ (with shift $\kappa = 0$) that we call D_i. Since

$$u_i < x_i = s_i < v_i < y_i \tag{2.58}$$

and $(s_i, t_i) \in \mathcal{S}$, we have

$$\sum_{\substack{u_i<k<v_i \\ d_{s_it_i}|d_k}} d_k = d_{s_i}. \tag{2.59}$$

We claim that $d_{s_i} = d_{x_i}$ is not divisible by q_i, which by (2.59) implies that $d_{s_it_i}$ is involved in the constraint D_i. Indeed, if q_i were to divide d_{x_i}, recalling that q_i divides d_{y_i-1}, we would have $q_i \mid d_k$ for all $x_i \leqslant k \leqslant y_i - 1$ by minimality of the divisibility triple (x_i, y_i, p_i). This is incompatible with (2.56) and (2.57).

Therefore, $d_{s_it_i}$ is involved in D_i for all $i \in I_2 \setminus I_0$. Moreover, for all $i', i \in I_2 \setminus I_0$ with $i' < i$, equation (2.58) and the fact that $(s_i, t_i) \in \mathcal{S}$ show that $d_{s_it_i}$ is absent from $D_{i'}$. Thus, $(D_i)_{i\in I_2\setminus I_0}$ is a triangular system of $\gg m$ constraints satisfied by $\boldsymbol{d}$. We may assume henceforth that $|I_2| < m/2$.

We have reached the final case of the proof. Let $I_3 = [\![m]\!] \setminus (I_1 \cup I_2)$, so that $|I_3| \geqslant m/4$. For $i \in I_3$, by definition of I_1 and I_2, the set $[\![x_i, y_i - 1]\!] \times [\![J]\!]$ has empty intersection with $\mathcal{S}$, so that

$$[\![x_i, y_i - 1]\!] \times [\![J]\!] \subset \mathcal{L}. \tag{2.60}$$

By the pigeonhole principle, since $|I_3| \geqslant m/4 \geqslant R/L$, there exists $i \in I_3$ such that

$$|y_i - x_i| \leqslant L. \tag{2.61}$$

We claim that $(d_{x_i}, d_{x_i+1}, \dots, d_{y_i-1})$ is a prohibited sequence. Indeed, it satisfies all the assumptions of Definition 2.7.3:

- $(d_{x_i}, d_{x_i+1}, \dots, d_{y_i-1})$ is reduced as $\boldsymbol{d} \in \mathbf{Red}_{R,\mathcal{S},\mathcal{L}}$;
- for every prime q, the set $\{i \in [\![x_i, y_i - 1]\!] : q \mid d_i\}$ is a discrete interval by minimality of the divisibility triple (x_i, y_i, p_i);
- p_i divides d_{x_i} and satisfies the divisibility condition (2.57);
- p_i does not divide d_{y_i-1} by minimality of (x_i, y_i, p_i) again;
- the tuple $(d_{x_i}, d_{x_i+1}, \dots, d_{y_i-1})$ has length $|y_i - x_i| \leqslant L$ by (2.61).

By (2.60), this contradicts the assumption that $\boldsymbol{d}$ is $\mathcal{L}$-admissible (see Definitions 2.8.3 and 2.12.12). Thus, this final case cannot occur, which concludes the proof. □

The next lemma deals with tuples $\boldsymbol{d}$ having an associated word that satisfies the second conclusion of Proposition 2.12.10.

Lemma 2.12.16. *Let $1 \leqslant R < H_0$ and let $\mathcal{S}, \mathcal{L}$ be disjoint subsets of $[\![R]\!] \times [\![J]\!]$.*

Let $\boldsymbol{d} \in \mathbf{Red}_{R,\mathcal{S},\mathcal{L}}$ and let $j \in [\![J]\!]$. Suppose that, for some $m \geqslant 16R^2/L^2$, there exist

$$1 \leqslant x_0 < z_1 < x_1 < z_2 < x_2 < \ldots < x_{m-1} < z_m < x_m < y_0 < y_1 < y_2 < \ldots < y_m \leqslant R$$

and $p_0, \dots, p_m \in \mathcal{P}_j$ such that (x_i, y_i, p_i) is a divisibility triple for $\boldsymbol{d}$, for all $0 \leqslant i \leqslant m$.

Define $q_i := d_{z_i j}$ for $i \in [\![m]\!]$. Suppose that $q_i \nmid \boldsymbol{d}_{[\![x_0, x_{i-1}]\!]} \boldsymbol{d}_{[\![y_0, y_i]\!]}$ and $p_i \nmid \boldsymbol{d}_{[\![x_0, z_i]\!]}$ for all $i \in [\![m]\!]$.

Finally, assume that $\mathcal{S} \sqcup \mathcal{L} \supset \big([\![x_0, x_m]\!] \cup [\![y_0, y_m]\!]\big) \times [\![J]\!]$.

Then, the concatenation[25] of $\boldsymbol{d}$ and $-\boldsymbol{d}$ satisfies a triangular system of $\gg m^{1/2}$ constraints with shift $\kappa = O(RH)$.

[25] We work with the concatenation of $\boldsymbol{d}$ and $-\boldsymbol{d}$ to allow for negative signs in the constraints. The reason for this will be apparent in the proof.

Proof. Call an integer $i \in [\![m]\!]$ *unsuitable* if $|x_i - x_{i-1}| \geqslant L$, $|y_i - y_{i-1}| \geqslant L$, or there exists a divisibility triple (x, y, p) with $x_{i-1} < x < y \leqslant x_i$. Otherwise, we call i *suitable*.

The first two scenarios can only happen for $\leqslant R/L \leqslant m^{1/2}$ values of $i \in [\![m]\!]$. If the third one applies to $\geqslant m^{1/2} \geqslant 4R/L$ values of i, then we are done by Lemma 2.12.15. Therefore, we can assume that there are at most $O(m^{1/2})$ unsuitable integers $i \in [\![m]\!]$.

We will henceforth work with a subinterval of $[\![m]\!]$ containing only suitable integers: there exists such a subinterval $[\![m_1, m_2]\!] \subset [\![m]\!]$ with $m_2 - m_1 \gg m^{1/2}$.

Let $i \in [\![m_1, m_2]\!]$. Since $p_i = d_{x_i j} = d_{y_i j}$, we have $(x_i, j), (y_i, j) \notin \mathcal{S}$, and thus $(x_i, j), (y_i, j) \in \mathcal{L}$. Therefore, by Definition 2.12.12, we have

$$p_i \Big| \sum_{x_i < k < y_i} d_k.$$

We rewrite this as

$$p_i = d_{x_i j} \Big| \kappa + \sum_{x_0 \leqslant k < x_i} -d_k + \sum_{y_0 < k < y_i} d_k, \tag{2.62}$$

with $\kappa := \sum_{x_0 \leqslant z \leqslant y_0} d_z$. This is a constraint with shift κ on the concatenation of $\boldsymbol{d}$ and $-\boldsymbol{d}$; we call this constraint C_i. Note that the same shift κ is used for all constraints C_i, and it satisfies $|\kappa| \leqslant RH$. We will show that an appropriate subset of these constraints forms a triangular system.

We define I to be the set of all $i \in [\![m_1, m_2]\!]$ such that $\big([\![x_{i-1}, x_i]\!] \times [\![J]\!]\big) \cap \mathcal{S} \neq \emptyset$.

Suppose first that $|I| \geqslant \frac{1}{2}(m_2 - m_1)$. Then one of the sets

$$I_0 := \{i \in I : i \equiv 0 \pmod 2\} \quad \text{and} \quad I_1 := \{i \in I : i \equiv 1 \pmod 2\}$$

has size $\gg m^{1/2}$. Without loss of generality, suppose that $|I_0| \gg m^{1/2}$. For each $i \in I_0$, there is some $(s_i, t_i) \in \mathcal{S}$ such that $s_i \in [\![x_{i-1}, x_i]\!]$. We claim that the constraints $(C_{i+1})_{i \in I_0 \setminus \{m\}}$ form a triangular system. Indeed, the prime $d_{s_i t_i}$ is involved in C_{i+1} as

$$\sum_{\substack{x_0 \leqslant k < x_{i+1} \\ d_{s_i t_i} | d_k}} -d_k + \sum_{\substack{y_0 < k < y_{i+1} \\ d_{s_i t_i} | d_k}} d_k = -d_{s_i} \not\equiv 0 \pmod{p_{i+1}},$$

using that $(s_i, t_i) \in \mathcal{S}$, and the assumption $p_{i+1} \nmid \boldsymbol{d}_{[\![x_0, z_{i+1}]\!]}$ in the statement. Furthermore, $d_{s_i t_i}$ is absent from $C_{i'+1}$ for all even $i' < i$, as for such an i' we have $x_{i'+1} \leqslant s_i < y_0$. Thus, the concatenation of $\boldsymbol{d}$ and $-\boldsymbol{d}$ satisfies a triangular system of $\gg m^{1/2}$ constraints with shift κ, as required.

Hence, we may assume that $|I| < \frac{1}{2}(m_2 - m_1)$. Let $i \in [\![m_1, m_2]\!] \setminus I$. We will now use the integers z_i from the statement. We claim that the prime $q_i = d_{z_i j}$ is involved in the constraint C_i defined by (2.62).

Suppose for contradiction that q_i is not involved in C_i. Recalling the assumption $q_i \nmid \boldsymbol{d}_{[\![x_0,x_{i-1}]\!]}\boldsymbol{d}_{[\![y_0,y_i]\!]}$, this means that

$$\sum_{\substack{x_{i-1}<k<x_i \\ q_i \mid d_k}} d_k \equiv 0 \pmod{p_i}. \tag{2.63}$$

Since i is suitable, the set $\{x_{i-1} < k < x_i : q_i \mid d_k\}$ is a discrete interval, say $[\![r,t]\!]$ (otherwise, there would be a divisibility triple (x,y,q_i) for some $x_{i-1} < x < y < x_i$).

We claim that the vector $(d_{x_i}, d_{x_i-1}, d_{x_i-2}, \dots, d_r)$ is a prohibited sequence. Indeed, it satisfies all the assumptions of Definition 2.7.3:

- $(d_{x_i}, d_{x_i-1}, d_{x_i-2}, \dots, d_r)$ is reduced as $\boldsymbol{d} \in \mathbf{Red}_{R,\mathcal{S},\mathcal{L}}$;
- for every prime q, the set $\{i \in [\![r, x_i]\!] : q \mid d_i\}$ is a discrete interval, since there is no divisibility triple (x,y,p) with $x_{i-1} < x < y \leqslant x_i$ as i is suitable;
- p_i divides d_{x_i} and satisfies the divisibility condition

$$p_i \Bigg| \sum_{\substack{x_{i-1}<k<x_i \\ q_i \mid d_k}} d_k = \sum_{r \leqslant k \leqslant t} d_k$$

  by (2.63);
- p_i does not divide d_r, as otherwise p_i would equal $d_{rj} = d_{z_i j} = q_i$, contradicting the assumption $p_i \nmid \boldsymbol{d}_{[\![x_0,z_i]\!]}$ in the statement;
- the tuple $(d_{x_i}, d_{x_i-1}, d_{x_i-2}, \dots, d_r)$ has length $|x_i - r + 1| \leqslant |x_i - x_{i-1}| < L$ as i is suitable.

However, since $i \notin I$, we have

$$[\![x_{i-1}, x_i]\!] \times [\![J]\!] \subset \mathcal{L}.$$

This contradicts the $\mathcal{L}$-admissibility of $\boldsymbol{d}$.

As a consequence, we deduce that q_i is involved in the constraint C_i. Since $q_i \nmid \boldsymbol{d}_{[\![x_0,x_{i-1}]\!]}\boldsymbol{d}_{[\![y_0,y_i]\!]}$ by assumption, it follows that $q_i = d_{z_i j}$ is absent from $C_{i'}$ for all $i' < i$. Therefore, the concatenation of $\boldsymbol{d}$ and $-\boldsymbol{d}$ satisfies the triangular system $(C_i)_{i \in [\![m_1,m_2]\!] \setminus I}$, which is made of $\gg m^{1/2}$ constraints with shift $\kappa = O(RH)$. This concludes the proof. □

We also state a variant of Lemma 2.12.16 where the y_i come in the opposite order. The proof is completely analogous and left to the reader.

Lemma 2.12.17. *Let $1 \leqslant R < H_0$ and let $\mathcal{S}, \mathcal{L}$ be disjoint subsets of $[\![R]\!] \times [\![J]\!]$.*

Let $\boldsymbol{d} \in \mathbf{Red}_{R,\mathcal{S},\mathcal{L}}$ *and let* $j \in [\![J]\!]$. *Suppose that, for some* $m \geqslant 16R^2/L^2$, *there exist*

$$1 \leqslant x_0 < z_1 < x_1 < z_2 < x_2 < \ldots < x_{m-1} < z_m < x_m < y_m < y_{m-1} < \ldots < y_0 \leqslant R$$

and $p_0, \ldots, p_m \in \mathcal{P}_j$ *such that* (x_i, y_i, p_i) *is a divisibility triple for* $\boldsymbol{d}$, *for all* $0 \leqslant i \leqslant m$.

Define $q_i := d_{z_i j}$ *for* $i \in [\![m]\!]$. *Suppose that* $q_i \nmid \boldsymbol{d}_{[\![x_0, x_{i-1}]\!]} \boldsymbol{d}_{[\![y_i, y_0]\!]}$ *and* $p_i \nmid \boldsymbol{d}_{[\![x_0, z_i]\!]}$ *for all* $i \in [\![m]\!]$.

Finally, assume that $\mathcal{S} \sqcup \mathcal{L} \supset \big([\![x_0, x_m]\!] \cup [\![y_m, y_0]\!]\big) \times [\![J]\!]$.

Then, the concatenation of $\boldsymbol{d}$ *and* $-\boldsymbol{d}$ *satisfies a triangular system of* $\gg m^{1/2}$ *constraints with shift* $\kappa = O(RH)$.

2.12.3 Contribution of unpredictable walks

We can finally prove Proposition 2.12.1, which we restate here for convenience.

Proposition 2.12.1. *Let* $\mathcal{S}$, $\mathcal{L}$, $\mathcal{U}$ *be sets such that* $[\![2K]\!] \times [\![J]\!] = \mathcal{S} \sqcup \mathcal{L} \sqcup \mathcal{U}$ *and* $|\mathcal{U}| \leqslant K^{2\varepsilon_1}$. *Let* $1 \leqslant r \leqslant K$. *Then*

$$\sum_{\boldsymbol{d} \in \mathbf{Unp}_{\mathcal{S},\mathcal{L},r}} \prod_{p \mid \boldsymbol{d}_*} \frac{1}{p} \ll 1.$$

Proof of Proposition 2.12.1. Let $\boldsymbol{d} \in \mathbf{Unp}_{\mathcal{S},\mathcal{L},r}$. By Definition 2.10.9, there exist $j \in [\![J]\!]$ and $k \in \{1,2,3\}$ such that the word $w_{j,\boldsymbol{d}}^{(k)}$ introduced in Definition 2.10.8 is $K^{1/4}$-unpredictable. Let $\boldsymbol{d}^{(k)}$, $I^{(k)}$, $\iota^{(k)}$ and $\mathcal{L}^{(k)}$ be as in Definition 2.8.4, and define

$$\mathcal{S}^{(k)} := \{(\iota^{(k)}(i), j) \mid (i,j) \in \mathcal{S}\} \quad \text{and} \quad \mathcal{U}^{(k)} := \{(\iota^{(k)}(i), j) \mid (i,j) \in \mathcal{U}\}.$$

We claim that $\boldsymbol{d}^{(k)} \in \mathbf{Red}_{R,\mathcal{S}^{(k)},\mathcal{L}^{(k)}}$ where $R := |I^{(k)}|$. To prove this, we need to check the properties of Definition 2.12.12. The fact that $\boldsymbol{d}^{(k)}$ is reduced and $\mathcal{L}^{(k)}$-admissible follows from part (4) of Definition 2.8.4. Property (1) of Definition 2.12.12 follows from the fact that $\mathcal{S}$ is the set of single indices for $\boldsymbol{d}$. Finally, property (2) of Definition 2.12.12 follows from part (2) of Definition 2.8.4, together with the identity $\sum_{K-r<i\leqslant K+r} d_i = 0$ in the case $k = 3$ (which follows from part (3) of Definition 2.8.4).

Therefore, $\boldsymbol{d}^{(k)} \in \mathbf{Red}_{R,\mathcal{S}^{(k)},\mathcal{L}^{(k)}}$, which will enable us to apply Lemma 2.12.15, Lemma 2.12.16 or Lemma 2.12.17 to $\boldsymbol{d}^{(k)}$, depending on the structure of the word $w_{j,\boldsymbol{d}}^{(k)}$, to exhibit a triangular system of constraints.

By Proposition 2.12.10, since $w_{j,\boldsymbol{d}}^{(k)}$ is $K^{1/4}$-unpredictable, there exists $m \gg K^{c_2}$ for some absolute constant $c_2 > 0$, such that one of the conclusions of Proposition 2.12.10 holds with this m.

The first possibility is that the word $w_{j,\boldsymbol{d}}^{(k)}$ has m separated repetitions (see Definition 2.12.2). Since $w_{j,\boldsymbol{d}}^{(k)}$ is the compression of the word $v_{j,\boldsymbol{d}}^{(k)}$ (see Definition 2.10.8), this means that there are

$$1 \leqslant x_1 < y_1 < x_2 < y_2 < \ldots < x_m < y_m \leqslant R$$

such that (x_i, y_i, p_i) is a divisibility triple for $\boldsymbol{d}^{(k)}$ for all $i \in [\![m]\!]$, where p_i is the unique prime in $\mathcal{P}_j$ dividing the x_i-th coordinate of $\boldsymbol{d}^{(k)}$. The assumptions of Lemma 2.12.15 are satisfied, except possibly for the condition that $[\![x_1, y_m]\!] \times [\![J]\!]$ does not intersect $\mathcal{U}^{(k)}$. Since $\left|\mathcal{U}^{(k)}\right| \leqslant |\mathcal{U}| \leqslant K^{2\varepsilon_1}$, $m \gg K^{c_2}$ and ε_1 is sufficiently small, we can find a discrete interval $[\![a, b]\!] \subset [\![m]\!]$ with $b - a \gg m^{1/2}$ such that $[\![x_a, y_b]\!] \times [\![J]\!]$ does not intersect $\mathcal{U}^{(k)}$. Then, applying Lemma 2.12.15 to $\boldsymbol{d}^{(k)}$ with the divisibility triples $(x_i, y_i, p_i)_{a \leqslant i \leqslant b}$, we conclude that $\boldsymbol{d}^{(k)}$ satisfies a triangular system of $\gg K^{c_2/2}$ constraints with shift $\kappa = 0$.

Otherwise, the second conclusion of Proposition 2.12.10 holds. Unpacking the notation, this means that either $\boldsymbol{d}^{(k)}$ or its reverse admits divisibility triples arranged as in Lemma 2.12.16 or Lemma 2.12.17. As in the previous paragraph, the condition that $\big([\![x_0, x_m]\!] \cup [\![y_0, y_m]\!]\big) \times [\![J]\!]$ (for Lemma 2.12.16) or $\big([\![x_0, x_m]\!] \cup [\![y_m, y_0]\!]\big) \times [\![J]\!]$ (for Lemma 2.12.17) is disjoint from $\mathcal{U}^{(k)}$ may not be initially satisfied, but it can be guaranteed by passing to a consecutive subsequence of $\gg m^{1/2}$ divisibility triples. Then by Lemma 2.12.16 or Lemma 2.12.17, we conclude that the concatenation of $\boldsymbol{d}^{(k)}$, $-\boldsymbol{d}^{(k)}$ and their reverses,[26] satisfies a triangular system of $\gg K^{c_2/4}$ constraints with shift $\kappa = O(KH)$.

In summary, we have shown that, for every $\boldsymbol{d} \in \mathbf{Unp}_{\mathcal{S},\mathcal{L},r}$, the concatenation of all $\boldsymbol{d}^{(k)}$, $-\boldsymbol{d}^{(k)}$ and their reverses satisfies a triangular system of $\gg K^{c_2/4}$ constraints with shift $\kappa = O(KH)$.

By Lemma 2.11.6 (and since every $\boldsymbol{d}$ is uniquely determined by $\boldsymbol{d}^{(1)}$ and $\boldsymbol{d}^{(2)}$), we obtain the bound

$$\sum_{\boldsymbol{d} \in \mathbf{Unp}_{\mathcal{S},\mathcal{L},r}} \prod_{p \mid \boldsymbol{d}_*} \frac{1}{p} \ll KHK^{O(JK)} H_0^{-cK^{c'}} \ll K^{O(K \log K)} H_0^{-cK^{c'}}$$

for some absolute constants $c, c' > 0$. Since $\log H_0 \gg K^{1-\varepsilon_1}$ and ε_1 is assumed to be sufficiently small, this is $\ll 1$, as desired. □

2.13 Prohibited progressions and the combinatorial sieve

In this section, we establish several results on prohibited sequences and their associated progressions. We first bound the size of $I_N \setminus Y_L$ in Lemma 2.13.2. We also prove

[26] Including the reverses of $\boldsymbol{d}^{(k)}$ and $-\boldsymbol{d}^{(k)}$ is not strictly necessary.

Lemma 2.9.6, which constructs a suitable cut-off function for the combinatorial sieve and was stated without proof in Section 2.9.

We then prove the high trace bound for the non-backtracking operator of G_{Y_L} (Proposition 2.5.3), bringing together the results from Section 2.5 onwards.

2.13.1 Primitive prohibited sequences

In this section, we prove a technical lemma that allows us to find constraints and involved primes in primitive prohibited sequences.

The divisibility condition in the definition of prohibited sequences (see Definition 2.7.3) only brings up a subset of the prime factors of the d_i. Even the primes that do appear in that constraint might not be involved in the sense of Definition 2.11.4. Lemma 2.13.1 is a useful tool to circumvent this problem: it allows us to pass from an arbitrary prime to a (possibly different) involved prime.

Lemma 2.13.1. *Let $\boldsymbol{d} = (d_1, \dots, d_\ell)$ be a primitive prohibited sequence. Let Γ be the set of all constraints with shift $\kappa = 0$ satisfied by $\boldsymbol{d}$.*

For every prime $p \mid d_1 \cdots d_\ell$, one of the following holds:

1. *p is involved in a constraint $C \in \Gamma$;*
2. *there is another prime q involved in a constraint of Γ, such that $q \mid d_k$ for some $k \in [\![\ell]\!]$, and*
$$\sum_{\substack{1 \leqslant i < k \\ p \mid d_i}} d_i \not\equiv 0 \pmod{q}.$$

Proof. By definition of a prohibited sequence (Definition 2.7.3), there are $1 < \ell_0 < \ell$ and $j_0 \in [\![J]\!]$ such that $d_{1j_0} \nmid d_\ell$ and

$$d_{1j_0} \Big| \sum_{\ell_0 \leqslant i \leqslant \ell} d_i. \tag{2.64}$$

Observe that this is a constraint with shift $\kappa = 0$ satisfied by $\boldsymbol{d}$; we denote it by C_1.

Let $h \in [\![\ell]\!]$ be the smallest positive integer such that

$$|d_h| = |d_{h+1}| = \ldots = |d_\ell|.$$

Since prohibited sequences are reduced, we must in fact have $d_h = d_{h+1} = \ldots = d_\ell$. Note that $h \geqslant 2$ since $d_{1j_0} \nmid d_\ell$.

Let $p_1 \in \mathcal{P}$ be any prime divisor of d_h but not d_{h-1}. We claim that p_1 is involved in C_1. Indeed,

$$\sum_{\substack{i \in [\![\ell_0, \ell]\!] \\ p_1 \mid d_i}} d_i = \sum_{i \in [\![h, \ell]\!]} d_i = (\ell - h + 1) d_\ell, \tag{2.65}$$

which is not divisible by d_{1j_0} since $\ell - h + 1 < L < H_0 < d_{1j_0}$. To obtain the first equality in (2.65), we used that, for any prime $p' \in \mathcal{P}$, the set

$$I(p') := \{i \in [\![\ell]\!] : p' \mid d_i\}$$

is a discrete interval, by definition of a prohibited sequence.

We are now ready to start the proof of Lemma 2.13.1 in earnest. Let $p \mid d_1 \cdots d_\ell$ be a prime.

We can assume that $I(p) \setminus [\![h, \ell]\!]$ is non-empty: otherwise, p divides d_h but not d_{h-1}, which implies that p is involved in C_1 by the previous paragraph, and we are in case (1).

Hence, $I(p) \setminus [\![h, \ell]\!]$ is a non-empty discrete interval, which we denote by $[\![a_1, a_2]\!]$.

Suppose first that

$$\sum_{\substack{i \in [\![\ell]\!] \\ p \mid d_i}} d_i \not\equiv 0 \pmod{p_1}. \tag{2.66}$$

Then, we are immediately done, as case (2) holds with $q = p_1$ and $k = h$.

Suppose now that (2.66) fails, which means that

$$p_1 \Big| \sum_{i \in [\![a_1, a_2]\!]} d_i. \tag{2.67}$$

This divisibility condition (2.67) implies that $\big(d_h, d_{h-1}, \ldots, d_{a_1+1}, d_{a_1}\big)$ is a prohibited sequence. Since $\boldsymbol{d}$ is primitive, this is only possible if $a_1 = 1$ and $h = \ell$. In particular, $a_2 \geqslant 2$ (otherwise, we would have $p_1 \mid d_1$ by (2.67)).

We will now exhibit another prime p_2 for which case (2) of the lemma holds with $q = p_2$.

Let p_2 be the prime d_{2j_0}. Note that $p_2 \nmid d_1$, or else we would have $p_2 = d_{1j_0}$, and thus $p_2 \mid \sum_{\ell_0 \leqslant i \leqslant \ell} d_i$ by (2.64). This would imply that $(d_2, d_3, \ldots, d_\ell)$ is a prohibited sequence, which is impossible since $\boldsymbol{d}$ is primitive.

We claim that p_2 is involved in the constraint C_2 defined by the divisibility relation (2.67). Suppose for contradiction that this is not the case. This means that

$$p_1 \Big| \sum_{\substack{i \in [\![1, a_2]\!] \\ p_2 \mid d_i}} d_i = \sum_{\substack{i \in [\![2, a_2]\!] \\ p_2 \mid d_i}} d_i, \tag{2.68}$$

using that $p_2 \nmid d_1$. Note that $\{i \in [\![2, a_2]\!] : p_2 \mid d_i\}$ is a discrete interval containing 2 and not containing h. Thus, (2.68) implies that $\big(d_h, d_{h-1}, \ldots, d_2\big)$ is a prohibited sequence, contradicting that $\boldsymbol{d}$ is primitive. Hence, p_2 is involved in C_2.

To check that case (2) holds with $q = p_2$, it only remains to verify that

$$\sum_{\substack{1 \leqslant i < k \\ p \mid d_i}} d_i \not\equiv 0 \pmod{p_2}$$

for some index k such that $p_2 \mid d_k$. It suffices to take $k = 2$. □

We can use the previous lemma (in fact, a much weaker version would suffice) to bound $|I_N \setminus Y_L|$.

Lemma 2.13.2. *Whenever* $\log N \geqslant (\log H)^2$, *we have* $|I_N \setminus Y_L| \ll H_0^{-1/3} N$.

Proof. Recall that $\mathbb{Z} \setminus Y_L$ is the union of all prohibited progressions $P \in \mathcal{Y}$. For any $P \in \mathcal{Y}$, there is a primitive prohibited sequence $\boldsymbol{d}$ of length at most $L = K^{1-10\varepsilon_1}$ such that P is the prohibited progression associated to $\boldsymbol{d}$. In particular, the modulus of P satisfies

$$q_P \leqslant H^L \ll \exp\big(O\big(K^{2-10\varepsilon_1}\big)\big) \ll N,$$

using that $\log N \gg K^2$. By the union bound, we deduce that

$$|I_N \setminus Y_L| \leqslant \sum_{P \in \mathcal{Y}} |I_N \cap P| \ll N \sum_{P \in \mathcal{Y}} \frac{1}{q_P}.$$

By Lemma 2.13.1, there exists a constraint C satisfied by $\boldsymbol{d}$, which has zero shift and involves at least one prime. This constraint alone can be viewed as a triangular system. Applying Lemma 2.11.6 with $T = 1$ and $R = \ell$ for each $2 < \ell \leqslant L$, we obtain

$$\sum_{P \in \mathcal{Y}} \frac{1}{q_P} \leqslant \sum_{2 < \ell \leqslant L} \sum_{\boldsymbol{d} \in \mathbf{T}_\ell} \prod_{p \mid d_1 \cdots d_\ell} \frac{1}{p} \ll L K^{O(JL)} H_0^{-1/2} \ll \exp\big(O\big(K^{1-10\varepsilon_1} (\log K)^2\big)\big) H_0^{-1/2},$$

where $\mathbf{T}_\ell$ denotes the set of all primitive prohibited sequences of length ℓ. The conclusion follows as $\log H_0 \gg K^{1-\varepsilon_1}$. □

2.13.2 Cut-off function for the combinatorial sieve

We turn to the proof of Lemma 2.9.6.

Recall that $(\mathcal{Y} - \boldsymbol{b})^\cap$ is the set defined in Notation 2.9.5. In the next definition, we introduce the function $\mathrm{rank}_{\boldsymbol{d}} : (\mathcal{Y} - \boldsymbol{b})^\cap \to \mathbb{Z}^{\geqslant 0} \cup \{+\infty\}$ which is used as a cut-off for the combinatorial sieve (or rather, a family of such functions, one for every $\boldsymbol{d}$).

Definition 2.13.3. Let $\mathcal{S}$, $\mathcal{L}$, $\mathcal{U}$ be sets such that $[\![2K]\!] \times [\![J]\!] = \mathcal{S} \sqcup \mathcal{L} \sqcup \mathcal{U}$ and let $1 \leqslant r \leqslant K$.

Let $\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}$ and let $A_{\mathcal{L},\boldsymbol{d}} \subset \mathbb{Z}$ be the arithmetic progression

$$A_{\mathcal{L},\boldsymbol{d}} := \{n \in \mathbb{Z} : \forall (i,j) \in \mathcal{L},\ d_{ij} \mid n + \mathbf{b}_i\}.$$

Let $R \in (\mathcal{Y} - \boldsymbol{b})^\cap$. If $R \cap A_{\mathcal{L},\boldsymbol{d}} = \emptyset$, we set $\mathrm{rank}_{\boldsymbol{d}}(R) := +\infty$. Otherwise, we define $\mathrm{rank}_{\boldsymbol{d}}(R)$ to be the largest integer $T \geqslant 0$ for which there exist progressions $Q_1, \dots, Q_T \in \mathcal{Y} - \boldsymbol{b}$ containing R such that, for each $t \in [\![T]\!]$, the modulus q_{Q_t} does not divide $q_{A_{\mathcal{L},\boldsymbol{d}}} \prod_{s \in [\![T]\!] \setminus \{t\}} q_{Q_s}$.

We need to show that these rank functions satisfy the four properties of Lemma 2.9.6. We will be able to quickly derive the first three properties from the following simple fact.

Lemma 2.13.4. *Let $R \in (\mathcal{Y} - \boldsymbol{b})^{\cap}$ be such that $\mathrm{rank}_{\boldsymbol{d}}(R) = T < +\infty$. Then, there are progressions $Q_1, \dots, Q_T \in \mathcal{Y} - \boldsymbol{b}$ such that*

$$\emptyset \neq R \cap A_{\mathcal{L},\boldsymbol{d}} = \bigcap_{t \in [\![T]\!]} Q_t \cap A_{\mathcal{L},\boldsymbol{d}}.$$

Proof. By definition of $(\mathcal{Y} - \boldsymbol{b})^{\cap}$, we may write $R = \bigcap_{i \in I} Q_i$ for some finite set I and some $Q_i \in \mathcal{Y} - \boldsymbol{b}$. Let I_0 be a minimal subset of I such that

$$R \cap A_{\mathcal{L},\boldsymbol{d}} = \bigcap_{i \in I_0} Q_i \cap A_{\mathcal{L},\boldsymbol{d}}. \tag{2.69}$$

Note that the modulus of a non-empty intersection of arithmetic progressions is the least common multiple of the moduli of these progressions. There is no $i_0 \in I_0$ such that $q_{Q_{i_0}}$ divides $q_{A_{\mathcal{L},\boldsymbol{d}}} \prod_{i \in I_0 \setminus \{i_0\}} q_{Q_i}$, for otherwise $\bigcap_{i \in I_0 \setminus \{i_0\}} Q_i \cap A_{\mathcal{L},\boldsymbol{d}}$ and $\bigcap_{i \in I_0} Q_i \cap A_{\mathcal{L},\boldsymbol{d}}$ would have the same modulus, so these progressions would be equal, contradicting the minimality of I_0. This shows that $|I_0| \leqslant \mathrm{rank}_{\boldsymbol{d}}(R)$. Thus, (2.69) means that we have been able to write $R \cap A_{\mathcal{L},\boldsymbol{d}}$ as an intersection of at most $\mathrm{rank}_{\boldsymbol{d}}(R)$ progressions $Q_i \cap A_{\mathcal{L},\boldsymbol{d}}$. Repeating some Q_i if necessary, we can make it an intersection of exactly $\mathrm{rank}_{\boldsymbol{d}}(R)$ sets. □

Proof of parts (1) and (2) of Lemma 2.9.6. Let $\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}$ and let $R \in (\mathcal{Y} - \boldsymbol{b})^{\cap}$ be a progression with $\mathrm{rank}_{\boldsymbol{d}}(R) = T < +\infty$. By Lemma 2.13.4, there are progressions $Q_1, \dots, Q_T \in \mathcal{Y} - \boldsymbol{b}$ such that

$$\emptyset \neq R \cap A_{\mathcal{L},\boldsymbol{d}} = \bigcap_{t \in [\![T]\!]} Q_t \cap A_{\mathcal{L},\boldsymbol{d}}.$$

Property (2) follows, since

$$\omega(q_R) \leqslant \omega(q_{R \cap A_{\mathcal{L},\boldsymbol{d}}}) \leqslant \omega(q_{A_{\mathcal{L},\boldsymbol{d}}}) + \sum_{t \in [\![T]\!]} \omega(q_{Q_t}) \leqslant 2JK + JLT.$$

For property (1), write $\omega_{\mathcal{S}}(n) := \sum_{p \mid \boldsymbol{d}_{\mathcal{S}}} \mathbf{1}_{p \mid n}$. We similarly obtain

$$\omega_{\mathcal{S}}(q_R) \leqslant \omega_{\mathcal{S}}(q_{R \cap A_{\mathcal{L},\boldsymbol{d}}}) \leqslant \omega_{\mathcal{S}}(q_{A_{\mathcal{L},\boldsymbol{d}}}) + \sum_{t \in [\![T]\!]} \omega_{\mathcal{S}}(q_{Q_t}) \leqslant 0 + JLT$$

as $q_{A_{\mathcal{L},\boldsymbol{d}}}$ is only divisible by the primes $p \mid \boldsymbol{d}_{\mathcal{L}}$. □

For part (3) of Lemma 2.9.6, namely the combinatorial sieve, we just need to use Proposition 2.9.4, checking that the hypotheses are satisfied.

Proof of part (3) of Lemma 2.9.6. We use Proposition 2.9.4 with the initial set of arithmetic progressions being $\mathcal{Y}-\boldsymbol{b}$, and with $X = X_{\boldsymbol{d}}$ being the set of all $R \in (\mathcal{Y}-\boldsymbol{b})^{\cap}$ such that $\mathrm{rank}_{\boldsymbol{d}}(R) < K^{5\varepsilon_1}$. Note that $X_{\boldsymbol{d}} \neq \emptyset$ as $\mathbb{Z} \in X_{\boldsymbol{d}}$.

For any $R, R' \in (\mathcal{Y}-\boldsymbol{b})^{\cap}$ with $R \subset R'$, it is clear from Definition 2.13.3 that $\mathrm{rank}_{\boldsymbol{d}}(R') \leqslant \mathrm{rank}_{\boldsymbol{d}}(R)$. Therefore $X_{\boldsymbol{d}}$ is closed under containment. Furthermore, $\omega(q_R) \ll JK$ for all $R \in X_{\boldsymbol{d}}$, by property (2) of Lemma 2.9.6, as $L = K^{1-10\varepsilon_1}$. For elements $R \in \partial X_{\boldsymbol{d}} \setminus \{\emptyset\}$, we also have $\omega(q_R) \ll JK$, as any $P \in \mathcal{Y}-\boldsymbol{b}$ has $\omega(q_P) \leqslant JL \leqslant JK$ by definition of a prohibited progression. The conclusion follows from Proposition 2.9.4, observing that '$n \notin P$ for all $P \in \mathcal{Y}-\boldsymbol{b}$' is equivalent to '$n + \mathbf{b}_i \in Y_L$ for all $i \in [\![2K]\!]$'. □

To prove part (4) of Lemma 2.9.6, we will need two technical lemmas: Lemma 2.13.5 and Lemma 2.13.6.

Lemma 2.13.5. *Let $\mathcal{S}$, $\mathcal{L}$, $\mathcal{U}$ be sets such that $[\![2K]\!] \times [\![J]\!] = \mathcal{S} \sqcup \mathcal{L} \sqcup \mathcal{U}$ and let $1 \leqslant r \leqslant K$.*

Let $\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}$. Let $\mathcal{P}' \subset \mathcal{P}$ be a set of size $\leqslant 3JK$ containing the prime divisors of $q_{A_{\mathcal{L},\boldsymbol{d}}}$. Then

$$\sum_{\substack{R \in \partial X_{\boldsymbol{d}} \\ R \cap A_{\mathcal{L},\boldsymbol{d}} \neq \emptyset}} \prod_{\substack{p \mid q_R \\ p \notin \mathcal{P}'}} \frac{1}{p} \ll e^{O(JK)}.$$

Proof. Let $\mathcal{X}$ be the set of all $R_1 \in (\mathcal{Y}-\boldsymbol{b})^{\cap} \setminus \{\emptyset\}$ of the form $R_1 = R \cap A_{\mathcal{L},\boldsymbol{d}}$ for some $R \in \partial X_{\boldsymbol{d}}$. Since the prime factors of $q_{A_{\mathcal{L},\boldsymbol{d}}}$ are in $\mathcal{P}'$, we may rewrite the expression in the statement as

$$\sum_{\substack{R \in \partial X_{\boldsymbol{d}} \\ R \cap A_{\mathcal{L},\boldsymbol{d}} \neq \emptyset}} \prod_{\substack{p \mid q_R \\ p \notin \mathcal{P}'}} \frac{1}{p} = \sum_{R_1 \in \mathcal{X}} \prod_{\substack{p \mid q_{R_1} \\ p \notin \mathcal{P}'}} \frac{1}{p} \sum_{\substack{R \in \partial X_{\boldsymbol{d}} \\ R \cap A_{\mathcal{L},\boldsymbol{d}} = R_1}} 1.$$

To bound the inner sum, we use the following fact: for any arithmetic progression R_1 and any $n \mid q_{R_1}$, there is a unique arithmetic progression $R \supset R_1$ with $q_{R_1}/q_R = n$; moreover, all progressions $R \supset R_1$ are obtained in this way. Therefore, the inner sum is bounded by the number of divisors of q_{R_1}. For every $R_1 \in \mathcal{X}$, we have $\omega(q_{R_1}) \ll JK$. This follows from part (2) of Lemma 2.9.6, using that $R_1 = R_0 \cap P \cap A_{\mathcal{L},\boldsymbol{d}}$ for some $R_0 \in X_{\boldsymbol{d}}$ and $P \in \mathcal{Y}-\boldsymbol{b}$. Therefore, q_{R_1} has $e^{O(JK)}$ divisors, and hence the inner sum is $e^{O(JK)}$.

It remains to show that

$$\sum_{R_1 \in \mathcal{X}} \prod_{\substack{p \mid q_{R_1} \\ p \notin \mathcal{P}'}} \frac{1}{p} \ll e^{O(JK)}. \tag{2.70}$$

This is a simple counting problem, similar to Lemma 2.8.8 or Proposition 2.10.10, but the notation is much heavier in this setting.

Let $T := \lfloor K^{5\varepsilon_1} \rfloor + 1$. By definition of $X_{\boldsymbol{d}}$ and Lemma 2.13.4, for any $R_0 \in X_{\boldsymbol{d}}$, the intersection $R_0 \cap A_{\mathcal{L},\boldsymbol{d}}$ can be written as the intersection of $T-1$ progressions of the form $Q \cap A_{\mathcal{L},\boldsymbol{d}}$ where $Q \in \mathcal{Y} - \boldsymbol{b}$. Hence, the same is true for any $R_0 \in \partial X_{\boldsymbol{d}}$ with T instead of $T-1$, by definition of $\partial X_{\boldsymbol{d}}$.

Let $R_1 \in \mathcal{X}$. By definition of $\mathcal{X}$ and the previous paragraph, we can write

$$R_1 = \bigcap_{t \in [\![T]\!]} \left(Q_t - \mathbf{b}_{k_t}\right) \cap A_{\mathcal{L},\boldsymbol{d}} \tag{2.71}$$

for some $Q_t \in \mathcal{Y}$ and $k_t \in [\![0, 2K]\!]$. For $t \in [\![T]\!]$, let $\boldsymbol{d}^{[t]}$ be a primitive prohibited sequence having Q_t as its associated prohibited progression. Let ℓ_t be the length of $\boldsymbol{d}^{[t]}$ and let $\sigma_t \in \{\pm 1\}^{\ell_t}$ be the sequence of signs of the coordinates of $\boldsymbol{d}^{[t]}$. As usual, for $(i,j) \in [\![\ell_t]\!] \times [\![J]\!]$ we write $d_{ij}^{[t]}$ for the unique prime in $\mathcal{P}_j$ dividing $d_i^{[t]}$. Let $\sim$ be the equivalence relation on $\bigsqcup_{t \in [\![T]\!]} \left(\{t\} \times [\![\ell_t]\!] \times [\![J]\!]\right)$ defined by

$$(t_1, i_1, j_1) \sim (t_2, i_2, j_2) \quad \Longleftrightarrow \quad d_{i_1 j_1}^{[t_1]} = d_{i_2 j_2}^{[t_2]}.$$

For any equivalence class α of $\sim$, we write p_α for the prime $d_{ij}^{[t]}$, where (t,i,j) is any element of α. This definition does not depend on the choice of representative, by definition of $\sim$. Let E be the set of all equivalence classes α for $\sim$ such that $p_\alpha \in \mathcal{P}'$. Let $\varphi : E \to \mathcal{P}'$ be the map defined by $\varphi(\alpha) = p_\alpha$. We call the tuple $\theta := ((k_t)_{t \in [\![T]\!]}, (\ell_t)_{t \in [\![T]\!]}, (\sigma_t)_{t \in [\![T]\!]}, \sim, E, \varphi)$ a *template* for R_1. Thus, to every progression $R_1 \in \mathcal{X}$ we may associate a template θ (there may not be a canonical choice for the template associated to R_1, as it depends on the choice of a representation of R_1 as in (2.71)).

Let Θ denote the set of all tuples θ which are the template of some progression $R_1 \in \mathcal{X}$.

Fix some $\theta = ((k_t)_{t \in [\![T]\!]}, (\ell_t)_{t \in [\![T]\!]}, (\sigma_t)_{t \in [\![T]\!]}, \sim, E, \varphi) \in \Theta$ and let $\mathcal{X}_\theta$ be the set of all $R_1 \in \mathcal{X}$ for which θ is a template. Observe that any $R_1 \in \mathcal{X}_\theta$ is uniquely determined by the sequence of primes $(p_\alpha)_{\alpha \in E'}$, where E' is the set of all equivalence classes α of $\sim$ such that $\alpha \notin E$. Thus

$$\sum_{R_1 \in \mathcal{X}_\theta} \prod_{\substack{p \mid q_{R_1} \\ p \notin \mathcal{P}'}} \frac{1}{p} \leqslant \sum_{(p_\alpha)_{\alpha \in E'}} \prod_{\alpha \in E'} \frac{1}{p_\alpha} \leqslant V^{|E'|} \leqslant V^{TLJ} \leqslant e^{O(JK)}, \tag{2.72}$$

where we used that $T \ll K^{5\varepsilon_1}$ and $L < K^{1-10\varepsilon_1}$ for the last inequality.

We proceed to sum (2.72) over all possible choices of $\theta \in \Theta$. We will be done provided that the number of such templates is $e^{O(JK)}$. The number of choices for $(k_t)_{t \in [\![T]\!]}$, $(\ell_t)_{t \in [\![T]\!]}$ and $(\sigma_t)_{t \in [\![T]\!]}$ is at most $(2K+1)^T$, L^T and $(2^L)^T$ respectively. Since $\sim$ is an equivalence relation on a set of size $\leqslant TLJ$, there are $\leqslant (TLJ)^{TLJ}$ choices for $\sim$. There are $\leqslant 2^{TLJ}$ choices for E. Finally, φ is a map from a set of size $\leqslant TLJ$

to a set of size $\leqslant 3JK$ (using the assumption on $\mathcal{P}'$ in the statement), so there are $\leqslant (3JK)^{TLJ}$ possibilities for φ. In summary, the number of templates is

$$\leqslant (2K+1)^T \cdot L^T \cdot (2^L)^T \cdot (TLJ)^{TLJ} \cdot 2^{TLJ} \cdot (3JK)^{TLJ} = e^{O(JK)}.$$

This proves (2.70) and concludes the proof of Lemma 2.13.5. □

The next technical lemma relates the remainder term of the combinatorial sieve to triangular systems of constraints. It relies on Lemma 2.13.1 and is the key ingredient in the proof of part (4) of Lemma 2.9.6.

Lemma 2.13.6. *Let $\mathcal{S}$, $\mathcal{L}$, $\mathcal{U}$ be sets such that $[\![2K]\!] \times [\![J]\!] = \mathcal{S} \sqcup \mathcal{L} \sqcup \mathcal{U}$ and $|\mathcal{U}| \leqslant K^{2\varepsilon_1}$. Let $1 \leqslant r \leqslant K$.*

Let $\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}$. Let $T \geqslant 10K^{2\varepsilon_1}$ and let $k_1 \leqslant k_2 \leqslant \cdots \leqslant k_T$ be elements of $[\![0, 2K]\!]$. Let $\boldsymbol{d}^{[1]}, \dots, \boldsymbol{d}^{[T]}$ be primitive prohibited sequences, and let $Q_1, \dots, Q_T \in \mathcal{Y}$ be the associated prohibited progressions. Suppose that, for each $t \in [\![T]\!]$, the modulus q_{Q_t} does not divide $q_{A_{\mathcal{L},\boldsymbol{d}}} \prod_{s \in [\![T]\!] \setminus \{t\}} q_{Q_s}$, and that

$$\bigcap_{t \in [\![T]\!]} (Q_t - \mathbf{b}_{k_t}) \cap A_{\mathcal{L},\boldsymbol{d}} \neq \emptyset.$$

Let $\boldsymbol{v}$ be the sequence obtained by concatenation of $\boldsymbol{d}, \boldsymbol{d}^{[1]}, \dots, \boldsymbol{d}^{[T]}, -\boldsymbol{d}^{[1]}, \dots, -\boldsymbol{d}^{[T]}$. Then, $\boldsymbol{v}$ satisfies a triangular system of $\gg T$ constraints with shift $\kappa = 0$.

Proof. Fix some $n \in \bigcap_{t=1}^{T} (Q_t - \mathbf{b}_{k_t}) \cap A_{\mathcal{L},\boldsymbol{d}}$.

For every $t \in [\![T]\!]$, let ℓ_t be the length of $\boldsymbol{d}^{[t]}$, and let Γ_t be the set of all constraints with shift $\kappa = 0$ satisfied by $\boldsymbol{d}^{[t]}$.

Suppose first that there is a set $I \subset [\![T]\!]$ of size $\geqslant \frac{1}{10}T$ such that, for every $t \in I$, there is a constraint $C_t \in \Gamma_t$ and a prime p_t which is involved in C_t and does not divide $\prod_{s<t} q_{Q_s}$. Thus, p_t is absent from C_s, for every $s \in I$ with $s < t$. This means that the constraints $(C_t)_{t \in I}$ form a triangular system, so we are done. The same conclusion holds if there is a set $I \subset [\![T]\!]$ of size $\geqslant \frac{1}{10}T$ such that, for every $t \in I$, there is a constraint $C_t \in \Gamma_t$ and a prime p_t which is involved in C_t and does not divide $\prod_{s>t} q_{Q_s}$. We may thus assume that, for $\geqslant \frac{8}{10}T$ values of $t \in [\![T]\!]$, every prime involved in some constraint of Γ_t divides both $\prod_{s<t} q_{Q_s}$ and $\prod_{s>t} q_{Q_s}$.

For every $t \in [\![T]\!]$, fix a prime p_t dividing q_{Q_t} but not dividing $q_{A_{\mathcal{L},\boldsymbol{d}}} \prod_{s \in [\![T]\!] \setminus \{t\}} q_{Q_s}$. This is possible by the assumption in the statement. We apply Lemma 2.13.1 with this prime p_t. Note that the first case of Lemma 2.13.1 (i.e. p_t being involved in some constraint of Γ_t) can only occur for $< \frac{2}{10}T$ values of $t \in [\![T]\!]$ by definition of p_t and

the previous paragraph. Let I_1 be the set of $t \in [\![T]\!]$ such that the second case holds, i.e. for $t \in I_1$ there is a prime q_t involved in a constraint $C_t \in \Gamma_t$ such that

$$\sum_{\substack{1 \leqslant i < i_t \\ p_t \mid d_i^{[t]}}} d_i^{[t]} \not\equiv 0 \pmod{q_t}, \tag{2.73}$$

where $i_t \in [\![\ell_t]\!]$ is such that $q_t \mid d_{i_t}^{[t]}$. Thus $|I_1| \geqslant \frac{8}{10}T$.

By our earlier arguments, there is a subset $I_2 \subset I_1$ of size $|I_2| \geqslant \frac{6}{10}T$ such that, for all $t \in I_2$, there are $1 \leqslant s_1(t) < t < s_2(t) \leqslant T$ with $q_t \mid q_{Q_{s_1(t)}}$ and $q_t \mid q_{Q_{s_2(t)}}$.

By definition of p_t, we know that $p_t \nmid q_{A_{\mathcal{L},\boldsymbol{d}}}$, i.e. $p_t \nmid \boldsymbol{d}_{\mathcal{L}}$. Moreover, there are at most $\frac{1}{10}T$ values of $t \in I_2$ such that $p_t \mid \boldsymbol{d}_{\mathcal{U}}$, since $|\mathcal{U}| \leqslant K^{2\varepsilon_1} \leqslant \frac{1}{10}T$ and all p_t are distinct. We may thus find a subset $I_3 \subset I_2$ of size $|I_3| \geqslant \frac{1}{2}T$ such that $p_t \nmid \boldsymbol{d}_{\mathcal{L}\sqcup\mathcal{U}}$ for every $t \in I_3$.

Let I_4 be the set of all $t \in I_3$ such that $p_t \nmid \boldsymbol{d}_{[\![1,k_t]\!]\times[\![J]\!]}$. Let I_5 be the set of all $t \in I_3$ such that $p_t \nmid \boldsymbol{d}_{[\![k_t+1,2K]\!]\times[\![J]\!]}$. By definition of I_3, we know that for every $t \in I_3$ there is at most one pair $(i,j) \in [\![2K]\!] \times [\![J]\!]$ (which lies in $\mathcal{S}$) such that $p_t = d_{ij}$. In particular, $I_4 \cup I_5 = I_3$, so one of I_4 and I_5 has size $\geqslant \frac{1}{4}T$. We will only treat the case where $|I_4| \geqslant \frac{1}{4}T$; the proof for the case $|I_5| \geqslant \frac{1}{4}T$ is the same up to symmetry.

Let $t \in I_4$. By definition of n, we have $n + \mathbf{b}_{k_t} \in Q_t$ and thus, by definition of Q_t being the prohibited progression associated to $\boldsymbol{d}^{[t]}$,

$$q_t \,\Big|\, n + \mathbf{b}_{k_t} + \sum_{1 \leqslant i < i_t} d_i^{[t]},$$

with i_t as defined earlier. Since $q_t \mid q_{Q_{s_1(t)}}$, the same reasoning shows that

$$q_t \,\Big|\, n + \mathbf{b}_{k_{s_1(t)}} + \sum_{1 \leqslant i < i'_t} d_i^{[s_1(t)]},$$

where $i'_t \in [\![\ell_{s_1(t)}]\!]$ is any index such that $q_t \mid d_{i'_t}^{[s_1(t)]}$. Subtracting the two divisibility relations, we obtain

$$q_t \,\Big|\, \sum_{k_{s_1(t)} < i \leqslant k_t} d_i + \sum_{1 \leqslant i < i_t} d_i^{[t]} - \sum_{1 \leqslant i < i'_t} d_i^{[s_1(t)]}.$$

This is now a genuine constraint on $\boldsymbol{v}$ (with shift $\kappa = 0$), which we call C'_t. By (2.73), and since $p_t \nmid q_{Q_{s_1(t)}}$ (by definition of p_t) and $p_t \nmid \boldsymbol{d}_{[\![1,k_t]\!]\times[\![J]\!]}$ (by definition of I_4), we see that p_t is involved in this constraint C'_t. In addition, for $t_1, t_2 \in I_4$ with $t_1 < t_2$, the prime p_{t_2} is absent from C'_{t_1} since none of the integers $\boldsymbol{d}_{[\![1,k_{t_1}]\!]\times[\![J]\!]}$, $q_{Q_{t_1}}$ and $q_{Q_{s_1(t_1)}}$ are divisible by p_{t_2}. Therefore, the family $(C'_t)_{t\in I_4}$ forms a triangular system of $\gg T$ constraints, and we are done.

The case $|I_5| \geqslant \frac{1}{4}T$ is analogous, where this time $s_2(t)$ takes the role of $s_1(t)$. □

We can finally prove part (4) of Lemma 2.9.6.

Proof of part (4) of Lemma 2.9.6. Let $T := \lceil K^{5\varepsilon_1} \rceil$. Let $\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}$ and let $R \in \partial X_{\boldsymbol{d}}$ be such that $R \cap A_{\mathcal{L},\boldsymbol{d}} \neq \emptyset$. Observe that R satisfies $T \leqslant \mathrm{rank}_{\boldsymbol{d}}(R) < +\infty$ by definition of $\partial X_{\boldsymbol{d}}$. Thus, by Definition 2.13.3, we can find $Q_1, \dots, Q_T \in \mathcal{Y} - \boldsymbol{b}$ containing R such that, for each $t \in [\![T]\!]$, the modulus q_{Q_t} does not divide $q_{A_{\mathcal{L},\boldsymbol{d}}} \prod_{s \in [\![T]\!] \setminus \{t\}} q_{Q_s}$. We will first sum over all possibilities for $R_1 := \bigcap_{t \in [\![T]\!]} Q_t \cap A_{\mathcal{L},\boldsymbol{d}}$.

Let $\mathcal{X}$ be the set of all $R_1 \in (\mathcal{Y} - \boldsymbol{b})^{\cap}$ which are of the form

$$R_1 = \bigcap_{t \in [\![T]\!]} Q_t \cap A_{\mathcal{L},\boldsymbol{d}} \neq \emptyset \tag{2.74}$$

for some $Q_1, \dots, Q_T \in \mathcal{Y} - \boldsymbol{b}$ such that q_{Q_t} does not divide $q_{A_{\mathcal{L},\boldsymbol{d}}} \prod_{s \in [\![T]\!] \setminus \{t\}} q_{Q_s}$ for all $t \in [\![T]\!]$. Summing over $\mathcal{X}$ first, we have

$$\sum_{\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}} \sum_{\substack{R \in \partial X_{\boldsymbol{d}} \\ R \cap A_{\mathcal{L},\boldsymbol{d}} \neq \emptyset}} \prod_{p \mid q_R \boldsymbol{d}_*} \frac{1}{p} \leqslant \sum_{\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}} \sum_{R_1 \in \mathcal{X}} \prod_{p \mid q_{R_1} \boldsymbol{d}_*} \frac{1}{p} \sum_{\substack{R \in \partial X_{\boldsymbol{d}} \\ R \cap A_{\mathcal{L},\boldsymbol{d}} \neq \emptyset}} \prod_{\substack{p \mid q_R \\ p \nmid q_{R_1} \boldsymbol{d}_*}} \frac{1}{p}. \tag{2.75}$$

For fixed R_1, we bound the innermost sum using Lemma 2.13.5 with the set of primes $\mathcal{P}' := \{p : p \mid q_{R_1} \boldsymbol{d}_*\}$. The assumptions on $\mathcal{P}'$ are satisfied since $\mathcal{P}'$ contains the prime divisors of $q_{A_{\mathcal{L},\boldsymbol{d}}}$ and

$$|\mathcal{P}'| \leqslant \omega(\boldsymbol{d}_*) + \big|\{p : p \mid q_{R_1},\, p \nmid q_{A_{\mathcal{L},\boldsymbol{d}}}\}\big| \leqslant 2JK + TJL \leqslant 3JK$$

in view of the representation (2.74) of elements of $\mathcal{X}$. Thus, applying Lemma 2.13.5, we obtain

$$\sum_{\substack{R \in \partial X_{\boldsymbol{d}} \\ R \cap A_{\mathcal{L},\boldsymbol{d}} \neq \emptyset}} \prod_{\substack{p \mid q_R \\ p \nmid q_{R_1} \boldsymbol{d}_*}} \frac{1}{p} = e^{O(JK)}. \tag{2.76}$$

It remains to bound the expression

$$\sum_{\boldsymbol{d} \in \mathbf{D}_{\mathcal{S},\mathcal{L},r}} \sum_{R_1 \in \mathcal{X}} \prod_{p \mid q_{R_1} \boldsymbol{d}_*} \frac{1}{p}. \tag{2.77}$$

For every non-decreasing sequence $\boldsymbol{k} = (k_t)_{t \in [\![T]\!]}$ of elements of $[\![0, 2K]\!]$, let $\mathbf{T}(\boldsymbol{k})$ be the set of all pairs $(\boldsymbol{d}, R_1) \in \mathbf{D}_{\mathcal{S},\mathcal{L},r} \times \mathcal{X}$ such that

$$\emptyset \neq R_1 = \bigcap_{t \in [\![T]\!]} \big(Q_t - \mathbf{b}_{k_t}\big) \cap A_{\mathcal{L},\boldsymbol{d}}$$

for some prohibited progressions $Q_t \in \mathcal{Y}$ with $q_{Q_t} \nmid q_{A_{\mathcal{L},\boldsymbol{d}}} \prod_{s \in [\![T]\!] \setminus \{t\}} q_{Q_s}$ for all t.

By Lemma 2.13.6, for any pair $(\boldsymbol{d}, R_1) \in \mathbf{T}(\boldsymbol{k})$ and any choice $\boldsymbol{d}^{[1]}, \dots, \boldsymbol{d}^{[T]}$ of prohibited sequences used in the definition of R_1, the concatenation of the vectors $\boldsymbol{d}, \boldsymbol{d}^{[1]}, \dots, \boldsymbol{d}^{[T]}, -\boldsymbol{d}^{[1]}, \dots, -\boldsymbol{d}^{[T]}$ satisfies a triangular system of $\gg T$ constraints with

shift $\kappa = 0$. This concatenation has length $\leqslant 2K+2TL \leqslant 3K$. Applying Lemma 2.11.6, we get

$$\sum_{(\boldsymbol{d},R_1)\in\mathbf{T}(\boldsymbol{k})} \prod_{p\mid q_{R_1}\boldsymbol{d}_*} \frac{1}{p} \ll K^{O(JK)} H_0^{-cT}$$

for some absolute constant $c > 0$. Summing over all $K^{O(T)}$ choices for $\boldsymbol{k} = (k_t)_{t\in[\![T]\!]}$, we obtain that the expression (2.77) is bounded by $K^{O(JK+T)} H_0^{-cT}$.

By (2.75), and recalling our bound (2.76) for the inner sum, we conclude that

$$\sum_{\boldsymbol{d}\in\mathbf{D}_{\mathcal{S},\mathcal{L},r}} \sum_{\substack{R\in\partial X_{\boldsymbol{d}}\\ R\cap A_{\mathcal{L},\boldsymbol{d}}\neq\emptyset}} \prod_{p\mid q_R\boldsymbol{d}_*} \frac{1}{p} \ll e^{O(JK)} K^{O(JK+T)} H_0^{-cT},$$

which is $\ll 1$ since $J \ll \log K$, $\log H_0 \gg K^{1-\varepsilon_1}$ and $T \asymp K^{5\varepsilon_1}$. □

2.13.3 Proof of the high trace bound

Having assembled all the necessary ingredients over the course of the preceding sections, we now reach the final step: proving the high-trace estimate for the non-backtracking operator M_{Y_L}.

We restate Proposition 2.5.3 here for convenience.

Proposition 2.5.3 (High trace bound for M_Y)**.** *Let N be an integer such that $\log N \geqslant (\log H)^{2.001}$.*

Then, there exists a set $Y \subset \mathbb{Z}$ with $|I_N \setminus Y| \ll H_0^{-1/3} N$, such that the non-backtracking operator M_Y associated to the weighted graph G_Y satisfies the high trace bound

$$\mathrm{Tr}\big[\big((M_Y)^K\big)^*(M_Y)^K\big] \leqslant e^{O(JK)} V^{JK} N. \tag{2.20}$$

Proof of Proposition 2.5.3. We prove Proposition 2.5.3 with $Y = Y_L$, the set defined in Definition 2.7.4. This choice is acceptable as $|I_N \setminus Y_L| \ll H_0^{-1/3} N$ by Lemma 2.13.2.

We turn to the high trace bound. By Lemma 2.8.6, we have

$$\mathrm{Tr}\big[\big((M_Y)^K\big)^*(M_Y)^K\big] \ll e^{O(JK)} \max_{\substack{\mathcal{S}\sqcup\mathcal{L}\sqcup\mathcal{U}=[\![2K]\!]\times[\![J]\!]\\ 1\leqslant r\leqslant K}} E_{\mathcal{S},\mathcal{L},\mathcal{U},r} + N$$

with $E_{\mathcal{S},\mathcal{L},\mathcal{U},r}$ as in Definition 2.8.5. Hence, it suffices to prove that

$$E_{\mathcal{S},\mathcal{L},\mathcal{U},r} \ll e^{O(JK)} V^{JK} N \tag{2.78}$$

for any choice of $\mathcal{S}$, $\mathcal{L}$, $\mathcal{U}$ and r.

By Lemma 2.8.7, we have $E_{\mathcal{S},\mathcal{L},\mathcal{U},r} \ll N$ when $|\mathcal{U}| \geqslant K^{2\varepsilon_1}$, which is stronger than (2.78). We now assume that $|\mathcal{U}| < K^{2\varepsilon_1}$.

By Proposition 2.9.1, we have $E_{\mathcal{S},\mathcal{L},\mathcal{U},r} \ll e^{O(JK)} N$ when $|\mathcal{S}| \geqslant K^{1-\varepsilon_1}$, which is also stronger than (2.78). We may thus assume that $|\mathcal{S}| < K^{1-\varepsilon_1}$.

In the remaining case, we apply first the trivial bound given by Lemma 2.10.1 to obtain

$$E_{\mathcal{S},\mathcal{L},\mathcal{U},r} \ll e^{O(JK)} N \sum_{\boldsymbol{d}\in \mathbf{D}_{\mathcal{S},\mathcal{L},r}} \prod_{p\mid \boldsymbol{d}_*} \frac{1}{p}.$$

We then split the sum over $\boldsymbol{d}$ into two parts: the predictable part $\mathbf{Prd}_{\mathcal{S},\mathcal{L},r}$ and the unpredictable part $\mathbf{Unp}_{\mathcal{S},\mathcal{L},r}$ (see Definition 2.10.9).

By Proposition 2.10.10, the predictable contribution is

$$\sum_{\boldsymbol{d}\in \mathbf{Prd}_{\mathcal{S},\mathcal{L},r}} \prod_{p\mid \boldsymbol{d}_*} \frac{1}{p} \ll e^{O(JK)} V^{|\mathcal{S}|+(|\mathcal{L}|+|\mathcal{U}|)/2}.$$

Since $|\mathcal{S}| < K^{1-\varepsilon_1}$ and $V \ll K$, we have $V^{|\mathcal{S}|} \ll e^{O(K)}$, which can be absorbed in the $e^{O(JK)}$ factor. Therefore, since $|\mathcal{L}| + |\mathcal{U}| \leqslant 2JK$, we obtain

$$\sum_{\boldsymbol{d}\in \mathbf{Prd}_{\mathcal{S},\mathcal{L},r}} \prod_{p\mid \boldsymbol{d}_*} \frac{1}{p} \ll e^{O(JK)} V^{JK}. \tag{2.79}$$

Finally, by Proposition 2.12.1, the unpredictable contribution satisfies

$$\sum_{\boldsymbol{d}\in \mathbf{Unp}_{\mathcal{S},\mathcal{L},r}} \prod_{p\mid \boldsymbol{d}_*} \frac{1}{p} \ll 1. \tag{2.80}$$

Combining (2.79) and (2.80), we obtain the bound (2.78) in the remaining case where $|\mathcal{S}| < K^{1-\varepsilon_1}$ and $|\mathcal{U}| < K^{2\varepsilon_1}$. This finishes the proof of Proposition 2.5.3. □

2.14 Applying the Matomäki-Radziwiłł-Tao theorem

In this final section, we prove Proposition 2.3.1, which quickly follows from the next proposition.

Proposition 2.14.1. *Let $N \geqslant \frac{1}{H}\exp((\log H)^2)$ and let $I_N := \mathbb{N} \cap (N, 2N]$.*

For $\mathcal{I} \subset [\![J]\!]$, define $\mathcal{D}_{\mathcal{I}}$ to be the set of all products $\prod_{i\in\mathcal{I}} p_i$ with $p_i \in \mathcal{P}_i$ for all i. Then, for all non-empty $\mathcal{I} \subset [\![J]\!]$, we have

$$\sum_{n\in I_N} \sum_{d\in\mathcal{D}_{\mathcal{I}}} \frac{1}{d}\lambda(n)\lambda(n+d) \ll \frac{V^J N}{(\log H)^{1/2000}}.$$

Proof of Proposition 2.3.1 assuming Proposition 2.14.1. We can expand the difference $S_2 - S_1$ as

$$\sum_{\emptyset\neq\mathcal{I}\subset[\![J]\!]} S(\mathcal{I}),$$

where

$$S(\mathcal{I}) := \sum_{n\in I_N} \sum_{(p_1,\dots,p_J)\in\mathcal{P}_1\times\cdots\times\mathcal{P}_J} \left(\prod_{i\in\mathcal{I}} \frac{1}{p_i}\right)\left(\prod_{i\in[\![J]\!]\setminus\mathcal{I}} \mathbf{1}_{p_i\mid n}\right)\lambda(n)\lambda(n+p_1\cdots p_J).$$

Changing variables $n = md$ with $d = \prod_{i\in[\![J]\!]\setminus\mathcal{I}} p_i$ gives

$$S(\mathcal{I}) = \sum_{d\in\mathcal{D}_{[\![J]\!]\setminus\mathcal{I}}} \lambda(d)^2 \sum_{\frac{N}{d}<m\leqslant\frac{2N}{d}} \sum_{d'\in\mathcal{D}_{\mathcal{I}}} \frac{1}{d'}\lambda(m)\lambda(m+d').$$

By Proposition 2.14.1, the double sum over m and d' is

$$\ll \frac{V^J N/d}{(\log H)^{1/2000}}.$$

Hence,

$$S(\mathcal{I}) \ll \frac{V^J N}{(\log H)^{1/2000}} \sum_{d\in\mathcal{D}_{[\![J]\!]\setminus\mathcal{I}}} \frac{1}{d} \ll \frac{V^{2J} N}{(\log H)^{1/2000}}$$

for every non-empty $\mathcal{I} \subset [\![J]\!]$. Therefore

$$|S_2 - S_1| \ll 2^J V^{2J} \frac{N}{(\log H)^{1/2000}}.$$

Note that $2^J V^{2J} \ll (\log H)^{\varepsilon_1}$ by Lemma 2.2.3, so $|S_2 - S_1| \ll N(\log H)^{-1/2500}$ if ε_1 is sufficiently small. □

Lemma 2.14.2. *Fix a non-empty $\mathcal{I} \subset [\![J]\!]$ and let $\mathcal{D}_{\mathcal{I}}$ be as in Proposition 2.14.1. Let $M \in [H_0, H]$. Define*

$$Q(\alpha) := \sum_{\substack{d\in\mathcal{D}_{\mathcal{I}}\\ d\in(M/2,M]}} \frac{e(\alpha d)}{d}.$$

Then,

$$\int_0^1 |Q(\alpha)|^4\, d\alpha \ll \frac{V^{4J}}{M(\log M)^4}.$$

Proof. By Parseval's identity, we can expand the fourth moment $\int_0^1 |Q(\alpha)|^4 d\alpha$ as

$$\int_0^1 \left| \sum_{|m|\leqslant M} \left(\sum_{\substack{d_1,d_2\in\mathcal{D}_{\mathcal{I}}\\ d_1,d_2\in(M/2,M]\\ d_1-d_2=m}} \frac{1}{d_1 d_2} \right) e(m\alpha) \right|^2 d\alpha = \sum_{|m|\leqslant M} \left| \sum_{\substack{d_1,d_2\in\mathcal{D}_{\mathcal{I}}\\ d_1,d_2\in(M/2,M]\\ d_1-d_2=m}} \frac{1}{d_1 d_2} \right|^2.$$

For $m = 0$, the inner sum is trivially $\ll 1/M$.

Fix $m > 0$. Let $N(m, b, \mathcal{I}, M)$ denote the number of pairs $(d_1, d_2) \in \mathcal{D}_{\mathcal{I}} \times \mathcal{D}_{\mathcal{I}}$ such that $d_1 - d_2 = m$, $d_1 \in (M/2, M]$ and $\gcd(d_1, d_2, m) = b$. Note that $N(m, b, \mathcal{I}, M) = 0$ unless $b \mid m$ and $b \in \mathcal{D}_{\mathcal{I}'}$ for some $\mathcal{I}' \subset \mathcal{I}$, in which case we have

$$N(m, b, \mathcal{I}, M) = N\big(\tfrac{m}{b}, 1, \mathcal{I}\setminus\mathcal{I}', \tfrac{M}{b}\big).$$

We thus are led to bound the number of coprime solutions $(e_1, e_2) \in \mathcal{D}_{\mathcal{I}\setminus\mathcal{I}'} \times \mathcal{D}_{\mathcal{I}\setminus\mathcal{I}'}$ to the equation $e_1 - e_2 = m/b$ with $e_1 \in (M/2b, M/b]$. Let i_+ be the largest element of

$\mathcal{I} \setminus \mathcal{I}'$. We can rewrite $e_i = n_i p_i$ where $n_i \in \mathcal{D}_{(\mathcal{I}\setminus\mathcal{I}')\setminus\{i_+\}}$ and $p_i \in \mathcal{P}_{i_+}$ for $i \in \{1,2\}$. For fixed n_1, n_2, the number of solutions $(p_1, p_2) \in \mathcal{P}_{i_+} \times \mathcal{P}_{i_+}$ to the linear equation

$$n_1 p_1 - n_2 p_2 = \frac{m}{b}$$

with $n_1 p_1 \in (M/2b, M/b]$ is

$$\ll \frac{M/b}{\varphi(n_1)\varphi(n_2)(\log M/b)^2} \cdot \frac{m/b}{\varphi(m/b)}$$

by classical sieve theoretic methods, such as [18, Proposition 6.22]. To apply this particular result, we used the fact that $\max(n_1, n_2) \leqslant (M/b)^{1/10}$, which holds by property (b) of Lemma 2.2.3.

Note that $\varphi(n) \gg n$ if n is a product of $\leqslant J$ primes, each $\geqslant H_0$. This is the case for n_1 and n_2. Therefore,

$$N(m, b, \mathcal{I}, M) \ll \sum_{n_1, n_2 \in \mathcal{D}_{(\mathcal{I}\setminus\mathcal{I}')\setminus\{i_+\}}} \frac{M/b}{n_1 n_2 (\log M/b)^2} \cdot \frac{m/b}{\varphi(m/b)} \ll V^{2J} \frac{M/b}{(\log M/b)^2} \cdot \frac{m}{\varphi(m)}.$$

We conclude that the total number of solutions (d_1, d_2) to the equation $d_1 - d_2 = m$ with $d_1, d_2 \in (M/2, M]$ is

$$\ll \sum_{\substack{b \mid m \\ b < m}} V^{2J} \frac{M/b}{(\log M/b)^2} \cdot \frac{m}{\varphi(m)} \ll \frac{V^{2J} M}{(\log M)^2} \cdot \frac{\sigma_1(m)}{\varphi(m)}.$$

We thus obtain

$$\int_0^1 |Q(\alpha)|^4 d\alpha \ll \frac{1}{M^2} + \left(\frac{V^{2J}}{M(\log M)^2}\right)^2 \sum_{m=1}^{M} \left(\frac{\sigma_1(m)}{\varphi(m)}\right)^2 \ll \frac{V^{4J}}{M(\log M)^4},$$

where we used the elementary estimate [70, Corollary 3.6] in the last inequality. □

Proof of Proposition 2.14.1. Let $V_{[M]} := \sum_{d \in \mathcal{D}_{\mathcal{I}} \cap (M/2, M]} 1/d$. It suffices to show that

$$T_M := \sum_{n \in I_N} \sum_{\substack{d \in \mathcal{D}_{\mathcal{I}} \\ d \in (M/2, M]}} \frac{1}{d} \lambda(n)\lambda(n+d) \ll \left(\frac{V^J}{\log M} + V_{[M]}\right) \frac{N}{(\log H)^{1/1750}} \tag{2.81}$$

holds for all $M \in [H_0, H]$. Indeed, summing this inequality for all M in the set $\{H2^{-j} : j \geqslant 0\} \cap [H_0, H]$ gives the desired upper bound

$$\sum_{n \in I_N} \sum_{d \in \mathcal{D}_{\mathcal{I}}} \frac{1}{d} \lambda(n)\lambda(n+d) \ll \left(V^J (\log\log H) + V^J\right) \frac{N}{(\log H)^{1/1750}} \ll \frac{V^J N}{(\log H)^{1/2000}}.$$

To prove (2.81), we start by introducing a new average over shifts $m \leqslant M$ and use the circle method:

$$T_M = \frac{1}{M}\sum_{m\leqslant M}\sum_{n\in I_N}\sum_{\substack{d\in\mathcal{D}_{\mathcal{I}}\\ d\in(M/2,M]}}\frac{1}{d}\lambda(n+m)\lambda(n+m+d)+O(MV^J)$$
$$= \frac{1}{M}\sum_{n\in I_N}\int_0^1 Q(\alpha)F_n(\alpha)G_n(\alpha)d\alpha + O(MV^J),$$

with $F_n(\alpha) := \sum_{m\leqslant M}\lambda(n+m)e(\alpha m)$, $G_n(\alpha) := \sum_{k\leqslant 2M}\lambda(n+k)e(-\alpha k)$ and $Q(\alpha)$ as in Lemma 2.14.2. The error term $O(MV^J)$ is clearly negligible.

Let $\varepsilon > 0$ be a parameter to be fixed later, and let $E_\varepsilon := \{\alpha \in [0,1] : |Q(\alpha)| > \varepsilon\}$. Outside of E_ε, the function $|Q|$ is small and we have

$$\sum_{n\in I_N}\int_{[0,1]\setminus E_\varepsilon}|Q||F_n||G_n| \leqslant \varepsilon N\|F_n\|_2\|G_n\|_2 \ll \varepsilon MN.$$

On the other hand, the Lebesgue measure of E_ε is $\ll V^{4J}/(\varepsilon^4 M(\log M)^4)$, by Lemma 2.14.2 and Markov's inequality. Hence

$$\sum_{n\in I_N}\int_{E_\varepsilon}|Q||F_n||G_n| \ll \frac{V^{4J}\|Q\|_\infty}{\varepsilon^4 M(\log M)^4}\Big\|\sum_{n\in I_N}|F_n||G_n|\Big\|_\infty \ll \frac{V^{4J}\,V_{[M]}}{\varepsilon^4(\log M)^4}\Big\|\sum_{n\in I_N}|F_n|\Big\|_\infty.$$

We now make crucial use of [38, Theorem 1.3] to obtain

$$\Big\|\sum_{n\in I_N}|F_n|\Big\|_\infty = \sup_{\alpha\in\mathbb{R}}\sum_{n\in I_N}\Big|\sum_{n\leqslant n'\leqslant n+M}\lambda(n')e(\alpha n')\Big| \ll \big((\log M)^{-1/2}+(\log N)^{-1/700}\big)MN.$$

Since $M \geqslant H_0$ and $\log N \gg (\log H)^2$, this upper bound is $\ll (\log H)^{-c'}MN$ where $c' = 1/350$.

Putting everything together, we conclude that

$$T_M \ll \frac{1}{M}\left(\varepsilon + \frac{V^{4J}\,V_{[M]}}{\varepsilon^4(\log M)^4(\log H)^{c'}}\right)MN.$$

Choosing $\varepsilon = V^J(\log M)^{-1}(\log H)^{-c'/5}$, we deduce the claimed bound (2.81). □

Chapter 3

On the Fourier uniformity conjecture

3.1 Definitions and structure of the chapter

This chapter is devoted to the proof of our main result towards the Fourier uniformity conjecture, Theorem 1.3.3. A proof outline is provided in Section 1.3.3.

We start this section by introducing the main definitions and notation used throughout the chapter. With these definitions in hand, we state the two key intermediate results: our *local* and *global* structure theorems (Theorems 3.1.8 and 3.1.9). We then outline the structure of the chapter and deduce Theorem 1.3.3 from our global structure theorem. This work is based on the preprint [46].

3.1.1 Configurations and lifts

We fix the following global parameters for the rest of the chapter.

Definition 3.1.1 (Global parameters)**.** Let $P \geqslant 10^3$, $Y \geqslant P^3$ and let $\mathcal{P}$ be the set of primes in $[P, 2P]$.

The basic objects of study in this chapter are discrete subsets $A \subset [Y, 2Y)$, where each point $x \in A$ carries a frequency $\alpha_x \in \mathbb{R}/\mathbb{Z}$. We package this data into the following definition.

Definition 3.1.2 (Configuration)**.** A *configuration* is a triple $(A, \alpha_\bullet, H)$ where:

- $A \subset [Y, 2Y)$ is a non-empty 1-separated set;[1]
- $\alpha_\bullet := (\alpha_x)_{x \in A}$ is a family of elements in $\mathbb{R}/\mathbb{Z}$ indexed by A; and
- H is a real number satisfying $H \geqslant P^3$.

[1] Contrary to the convention in [71, 73], our setup is organised so that the ambient interval $[Y, 2Y)$ of any configuration remains the same throughout the proof, while the new parameter H is permitted to vary.

We refer to the elements $x \in A$ as the *points* of the configuration, and the associated $\alpha_x \in \mathbb{R}/\mathbb{Z}$ as the *frequencies*. The parameter H in the definition of a configuration will be used in Section 3.1.2 to quantify the strength of certain *local relations* between the frequencies α_x.

Definition 3.1.3 (Sub-configuration)**.** Let $\mathcal{A} = (A, \alpha_\bullet, H)$ be a configuration. For any non-empty subset $B \subset A$, we define the restriction

$$\mathcal{A}|_B := (B, (\alpha_x)_{x\in B}, H).$$

Any configuration of this form is called a *sub-configuration* of $\mathcal{A}$.

A key part of the proof is to construct, given a configuration $\mathcal{A}$ with many local relations, a new configuration $\mathcal{B}$ that retains many of these relations, and is a 'lift' of $\mathcal{A}$ in the following sense.

Definition 3.1.4 (Lift)**.** Let $\mathcal{A} = (A, \alpha_\bullet, H)$ and $\mathcal{B} = (B, \beta_\bullet, H')$ be configurations. For any $p \in \mathcal{P}$, we define

$$p\mathcal{B} := (B, (p\beta_x)_{x\in B}, H'/p).$$

We say $\mathcal{B}$ is a *lift* of $\mathcal{A}$ if there exists a prime $p \in \mathcal{P}$ such that $p\mathcal{B}$ is a sub-configuration of $\mathcal{A}$; that is, $B \subset A$, $H' = Hp$ and $p\beta_x = \alpha_x \pmod 1$ for all $x \in B$. We call p the *lifting prime*.

The proof relies on constructing a sequence of such lifts, which we call a tower.

Definition 3.1.5 (Tower)**.** A *tower of configurations* of *height* k is a sequence $(\mathcal{A}_i)_{0\leqslant i\leqslant k}$ of configurations such that $\mathcal{A}_i$ is a lift of $\mathcal{A}_{i-1}$ for all $1 \leqslant i \leqslant k$.

3.1.2 Local relations

Roughly speaking, two points $x, y \in A$ satisfy a local relation if there exist primes $p, q \in \mathcal{P}$ such that $px \approx qy$ and $q\alpha_x \approx p\alpha_y \pmod 1$. We measure the error terms in these approximations as follows.

Definition 3.1.6 (Strength of a local relation)**.** Let $\mathcal{A} = (A, \alpha_\bullet, H)$ be a configuration. For $x, y \in A$ and $p, q \in \mathcal{P}$, we define

$$d_{p,q;\mathcal{A}}(x, y) := \max\Big(\frac{1}{P}|px - qy|,\ H\|q\alpha_x - p\alpha_y\|\Big). \tag{3.1}$$

When $\mathcal{A}$ is clear from the context, we simply write $d_{p,q}(x, y)$.

We collect the quadruples (x, y, p, q) satisfying strong local relations in the following set.

Definition 3.1.7 (Strong local relations)**.** We define

$$Q(\mathcal{A}) := \big\{(x,y,p,q) \in A^2 \times \mathcal{P}^2 \,:\, d_{p,q;\mathcal{A}}(x,y) \leqslant \tfrac{1}{10}\big\}. \tag{3.2}$$

Thus, a quadruple (x,y,p,q) lies in $Q(\mathcal{A})$ if and only if $|px-qy| \leqslant \frac{P}{10}$ and $\|q\alpha_x - p\alpha_y\| \leqslant \frac{1}{10H}$.[2]

3.1.3 Local structure theorem

The central tool of our iterative argument is the following local structure theorem for configurations with many local relations.

Theorem 3.1.8 (Local structure theorem)**.** *There exists an absolute constant $C_0 \geqslant 1$ such that the following holds. Let $Y, P, \mathcal{P}$ be as in Definition 3.1.1. Let $\mathcal{A} = (A, \alpha_\bullet, H)$ be a configuration such that*

$$|Q(\mathcal{A})| \geqslant \delta |A||\mathcal{P}|^2$$

for some $0 < \delta < \frac{1}{2}$. Let $L \in \mathbb{N}$ and $0 < \varepsilon < \frac{1}{2}$ be parameters such that $\varepsilon^{-C_0 L} \leqslant P$ and $\varepsilon\delta^{-5}L \leqslant 1$.

Then, there exists a decomposition $A = A_0 \sqcup \ldots \sqcup A_L \sqcup A_{L+1}$, together with a lift $\mathcal{B} = (A \setminus A_{L+1},\, \beta_\bullet, Hp^*)$ of $\mathcal{A}$, with the following properties.

(I) $|A_0| \gg \delta^5 |A|$.

(II) For all but $\ll \varepsilon|A||\mathcal{P}|^2$ quadruples $(x,y,p,q) \in Q(\mathcal{A})$,

- *(a) if $x \in A_\ell$ then $y \in A_{\ell+e}$ for some $e \in \{-1,0,1\}$, and*
- *(b) if $x, y \notin A_{L+1}$, then $(x,y,p,q) \in Q(\mathcal{B})$.*

The local structure theorem provides a partition of A into (possibly empty) sets $A_0, A_1, \ldots, A_L$, along with an exceptional set A_{L+1}. On the non-exceptional part $A \setminus A_{L+1}$, the frequencies can be lifted to form a configuration $\mathcal{B}$ that preserves almost all local relations not involving A_{L+1}. Moreover, this decomposition heavily restricts the combinatorial structure of $\mathcal{A}$: apart from a negligible proportion of exceptions, the local relations in $\mathcal{A}$ must either be internal to a single set A_ℓ, or connect adjacent sets A_ℓ and $A_{\ell+1}$.

We defer the technical proof of Theorem 3.1.8 to Sections 3.5 to 3.7 to proceed directly to the main consequences on Fourier uniformity.

[2] The constant $1/10$ in Definition 3.1.7 is not important, but this choice ensures that y is uniquely determined by x, p, q whenever $(x,y,p,q) \in Q(\mathcal{A})$, since A is 1-separated.

3.1.4 Global structure theorem

The ultimate goal of our analysis is to establish a global formula for the frequencies of a configuration $\mathcal{A}$ with many local relations.

Theorem 3.1.9 (Global structure theorem)**.** *Let $\varepsilon_0 > 0$ and let Y be sufficiently large in terms of ε_0. Let $P := \exp((\log Y)^{2/5+\varepsilon})$ for some $\varepsilon_0 < \varepsilon < \frac{1}{5}$.*

Let $\mathcal{A} = (A, \alpha_\bullet, H)$ be a configuration such that $|Q(\mathcal{A})| \geqslant \delta|A||\mathcal{P}|^2$ for some $\delta \geqslant (\log\log Y)^{-10}$.

Let $\kappa := \exp((\log Y)^{2/5-\varepsilon})$. Then, there is a subset $A' \subset A$ of size

$$|A'| \geqslant \kappa^{-1}|A|,$$

such that, for all $x \in A'$, we have the approximate formula

$$\alpha_x = \frac{a_0}{q_0} + \frac{T_0}{x} + O\left(\frac{\kappa}{HP}\right) \pmod 1,$$

where a_0, q_0 are coprime integers with $1 \leqslant q_0 \leqslant \kappa$ and $T_0 \in \mathbb{R}$ satisfies $|T_0| \leqslant \kappa\frac{Y^2}{HP}$.

We prove Theorem 3.1.9 in Sections 3.2 to 3.4, using the local structure theorem (Theorem 3.1.8) as a black box. The argument proceeds in three main stages.

In Section 3.2, we iteratively apply Theorem 3.1.8 to construct a tower of configurations above $\mathcal{A}$. If the exceptional set A_{L+1} has negligible size, Theorem 3.1.8 provides a lift $\mathcal{B}$ that is almost an exact copy of $\mathcal{A}$; unfortunately, this cannot be guaranteed in general. We circumvent this by exploiting two key features of Theorem 3.1.8. First, its quantitative strength allows for a logarithmically large chain length L, imposing more rigid constraints on the local relations than [73, Proposition 4.1]. Second, its validity on arbitrarily sparse sets allows the iteration to proceed despite severe losses in the cardinality of A.

In Section 3.3, we analyse the resulting construction. The lack of GRH-type expansion prevents us from adopting the strategy of [73], where one fixes a single point x_0 in the top configuration and traces its 'descendants' down the tower to deduce a global formula on a positive density subset of A. Instead, we work solely within the top configuration $\mathcal{A}'$. The local relations in $\mathcal{A}'$ have exceptionally small error terms on the frequency side (these errors having contracted by a factor $\asymp P$ at each step of the iteration), making them amenable to more direct graph-theoretic arguments.

Finally, in Section 3.4, we combine the preceding analysis with an unconditional expansion estimate based on Vinogradov-Korobov type bounds for Dirichlet L-functions to establish the global structure theorem.

3.1.5 Deduction of the Fourier uniformity estimate

We now prove our main result on Fourier uniformity, Theorem 1.3.3. The main ingredients are Elliott's inequality (via Proposition B.2.1), our global structure theorem (Theorem 3.1.9), and a variant of the Matomäki-Radziwiłł theorem with a power-saving upper bound for the number of exceptional intervals (Theorem C.5.1).

Proof of Theorem 1.3.3. The assumption in the statement implies the existence of an H-separated subset $S \subseteq [X, 2X)$ of size $\gg \delta X/H$, such that for each $x \in S$, there exists $\theta_x \in \mathbb{R}/\mathbb{Z}$ satisfying

$$\left| \sum_{x \leqslant n < x+H} f(n) e(n\theta_x) \right| \gg \delta H.$$

By Proposition B.2.1, there exists a scale P with $H^{e^{-O(\delta^{-2})}} \leqslant P \leqslant H^{1/10}$ such that, writing $\mathcal{P}$ for the set of primes in $[P, 2P)$, there are

$$\gg \delta^7 |S||\mathcal{P}|^2$$

quadruples $(x, y, p, q) \in S^2 \times \mathcal{P}^2$ satisfying $|px - qy| \leqslant \frac{1}{10}PH$ and $\|q\theta_x - p\theta_y\| \leqslant \frac{P}{10H}$. Defining $Y := X/H$, $A := \frac{1}{H}S \subset [Y, 2Y)$ and $\alpha_x := \theta_{Hx}$ for $x \in A$, we see that $\mathcal{A} := (A, \alpha_\bullet, H/P)$ is a configuration with

$$|Q(\mathcal{A})| \gg \delta^7 |A||\mathcal{P}|^2.$$

Let $\kappa := \exp((\log Y)^{2/5-\varepsilon/2})$. By our global structure theorem, Theorem 3.1.9, there exists a subset $A' \subseteq A$ of size $|A'| \geqslant \kappa^{-1}|A|$ such that, for each $x \in A'$, we have the approximate formula

$$\alpha_x = \frac{a_0}{q_0} + \frac{T_0}{x} + O\left(\frac{\kappa}{H}\right) \pmod 1$$

for some a_0, q_0, T_0 independent of x, where a_0 and q_0 are coprime integers with $1 \leqslant q_0 \leqslant \kappa$ and T_0 is a real number with $|T_0| \leqslant \kappa Y^2/H$.

To conclude, we apply Theorem C.5.1 to the dilated set $S' := H \cdot A' \subset [X, 2X)$, the frequencies $(\theta_x)_{x \in S'}$, and $T := HT_0$. The required bound (C.24) simplifies to

$$\exp((\log X)^{2/5-\varepsilon/2})^{\delta^{-C}} \leqslant H \leqslant X^{\delta^C}$$

for some absolute constant $C > 0$, which holds by our choices of δ and H provided that X is sufficiently large in terms of ε. This yields the desired pretentiousness estimate for f. □

3.2 Construction of a tower of configurations

In this section, we use our local structure theorem, Theorem 3.1.8, to construct a tower of configurations above a given configuration $\mathcal{A}$ with many local relations, ensuring that each intermediate lift maintains a large relative density of local relations. We show that this construction succeeds unless $\mathcal{A}$ contains a very sparse, highly connected subconfiguration, a scenario we rule out in Section 3.4.

To accomplish this, we exploit the full strength of Theorem 3.1.8 to decompose $\mathcal{A}$ into L parts, where L can be chosen to be almost as large as $\log P$. We will use the following simple lemma to analyse the resulting partition.

Lemma 3.2.1. *Let $L \in \mathbb{N}$ and $1 \leqslant C \leqslant L$. Let $a_1, \dots, a_L \geqslant 0$ be such that, for all $i \in \{1, \dots, L\}$,*

$$a_i \geqslant \frac{C}{L} \sum_{j>i} a_j.$$

Then,

$$\min_{1 \leqslant i \leqslant L} a_i \ll \frac{e^{-C/3}}{L} \sum_{i=1}^{L} a_i.$$

Proof. We first consider a continuous analogue of the problem. Let $f : [0,1] \to \mathbb{R}_{\geqslant 0}$ be a non-negative integrable function such that, for some $\gamma \geqslant 1$ and all $x \in [0,1]$,

$$f(x) \geqslant \gamma \int_x^1 f(y)\, dy. \tag{3.3}$$

We claim that

$$\min_{x \in [0,1]} f(x) \leqslant \gamma e^{1-\gamma} \int_0^1 f(y)\, dy. \tag{3.4}$$

To see this, let $F(x) := \int_x^1 f(y)\, dy$. Then F is absolutely continuous, and $F' = -f$ almost everywhere. The bound (3.3) translates to $F' + \gamma F \leqslant 0$, i.e. $(e^{\gamma x} F(x))' \leqslant 0$, almost everywhere. Therefore, the function $e^{\gamma x} F(x)$ (which is absolutely continuous) is non-increasing, so that

$$(1-x) \min_{y \in [0,1]} f(y) \leqslant F(x) \leqslant e^{-\gamma x} F(0)$$

for all $x \in [0,1]$. Setting $x = 1 - \frac{1}{\gamma}$, we obtain the claim (3.4).

To deduce the lemma from this continuous version, we define a piecewise constant function f on $[0,1]$ by setting $f(x) := a_i$ for $x \in \left(\frac{i-1}{L}, \frac{i}{L}\right]$, for $1 \leqslant i \leqslant L$. Then, for all $x \in \left(\frac{i-1}{L}, \frac{i}{L}\right]$, we have

$$\int_x^1 f(y)\, dy \leqslant \frac{1}{L} \sum_{j \geqslant i} a_j \leqslant \left(\frac{1}{L} + \frac{1}{C}\right) a_i \leqslant \frac{2}{C} f(x),$$

so that f satisfies (3.3) with $\gamma := \max(\frac{C}{2}, 1)$. Applying (3.4) concludes the proof. □

We can now prove a dichotomy: if $\mathcal{A}$ has many local relations, then either it admits a suitable lift, or it contains a small subconfiguration whose relative density of local relations only decreases by a factor of the shape $1 - \frac{1}{\log P}$.

Lemma 3.2.2. *Let $\mathcal{A} = (A, \alpha_\bullet, H)$ be a configuration such that $|Q(\mathcal{A})| \geqslant \delta |A||\mathcal{P}|^2$, where $\delta \geqslant 1/\log Y$. Then, one of the following holds:*

(i) There exists a lift $\mathcal{B} = (B, \beta_\bullet, Hp^*)$ of $\mathcal{A}$ such that

$$|Q(\mathcal{B})| \geqslant \left(1 - O\left(\frac{1}{(\log Y)^{10}}\right)\right)\delta |B||\mathcal{P}|^2.$$

(ii) There is a subset $B \subset A$ with $|B| \leqslant \frac{1}{2}|A|$, such that

$$|Q(\mathcal{A}|_B)| \geqslant \left(1 - O\left(\frac{\delta^{-6}(\log\log Y)^2}{\log P}\right)\right)\delta |B||\mathcal{P}|^2.$$

Proof. Without loss of generality, we may assume that every proper subset $A' \subset A$ satisfies

$$|Q(\mathcal{A}|_{A'})| \leqslant \delta |A'||\mathcal{P}|^2, \tag{3.5}$$

or else we can start over with $\mathcal{A}|_{A'}$ in place of $\mathcal{A}$.

We apply Theorem 3.1.8 with $\varepsilon := (\log Y)^{-20}$ and $L := \lfloor (\log P)/(20C_0 \log\log Y)\rfloor$. Note that the conditions on ε and L in Theorem 3.1.8 are satisfied if Y is sufficiently large and $L \geqslant 1$, which we may assume or else the lemma holds for trivial reasons.

Write $A = A_0 \sqcup A_1 \sqcup \ldots \sqcup A_{L+1}$ and $\mathcal{B} = (A \setminus A_{L+1}, \beta_\bullet, Hp^*)$ for the decomposition and the lift given by Theorem 3.1.8.

We claim that one of the following holds:

(a) $|A_{L+1}| \ll |A|/(\log Y)^{15}$,

(b) $|A_{L+1}| \geqslant |A|/(\log Y)^{15}$ and there exists $1 \leqslant i \leqslant L$ such that

$$|A_i| \ll \frac{\log\log Y}{L} \sum_{j=i+1}^{L+1} |A_j|.$$

Indeed, suppose that $|A_{L+1}| \geqslant |A|/(\log Y)^{15}$ and that $|A_i| \geqslant \frac{60\log\log Y}{L}\sum_{j=i+1}^{L+1}|A_j|$ for all $1 \leqslant i \leqslant L$. Then, by Lemma 3.2.1, we have $\min_{1\leqslant i\leqslant L+1}|A_i| \ll |A|/(\log Y)^{20}$. If this minimum is attained at $i = L+1$, we are in case (a); otherwise, we are in case (b).

We now treat the two cases separately.

Suppose first that case (a) holds. By part (IIb) of Theorem 3.1.8, we have

$$|Q(\mathcal{B})| \geqslant \delta |A||\mathcal{P}|^2 - O\big(|A_{L+1}||\mathcal{P}|^2 + \varepsilon |A||\mathcal{P}|^2\big).$$

Recalling that $B = A \setminus A_{L+1}$ and inserting our bound on $|A_{L+1}|$, we obtain

$$|Q(\mathcal{B})| \geqslant \delta|A||\mathcal{P}|^2 - O\Big(\frac{|A||\mathcal{P}|^2}{(\log Y)^{15}}\Big),$$

which yields conclusion (i) of the lemma.

Now, suppose that case (b) holds. Define $B_1 := \bigsqcup_{j=0}^{i-1} A_j$ and $B_2 := \bigsqcup_{j=i+1}^{L+1} A_j$. By property (IIa) of Theorem 3.1.8, we have

$$|Q(\mathcal{A})| \leqslant |Q(\mathcal{A}|_{B_1})| + |Q(\mathcal{A}|_{B_2})| + O\big(|A_i||\mathcal{P}|^2 + \varepsilon|A||\mathcal{P}|^2\big).$$

To simplify the error term, we note that $|A_i| \ll \frac{\log\log Y}{L}|B_2|$ and $|A| \ll (\log Y)^{15}|B_2|$ by assumption. Thus,

$$|Q(\mathcal{A})| \leqslant |Q(\mathcal{A}|_{B_1})| + |Q(\mathcal{A}|_{B_2})| + O\Big(\frac{\log\log Y}{L}|B_2||\mathcal{P}|^2\Big). \tag{3.6}$$

Since

$$|Q(\mathcal{A})| \geqslant \delta|A||\mathcal{P}|^2 \geqslant \delta|B_1||\mathcal{P}|^2 + \delta|B_2||\mathcal{P}|^2$$

and $|Q(\mathcal{A}|_{B_k})| \leqslant \delta|B_k||\mathcal{P}|^2$ for $k = 1, 2$ by (3.5), the estimate (3.6) implies that

$$|Q(\mathcal{A}|_{B_k})| \geqslant \delta|B_k||\mathcal{P}|^2 - O\Big(\frac{\log\log Y}{L}|B_2||\mathcal{P}|^2\Big) \tag{3.7}$$

for $k = 1, 2$. Moreover, by part (I) of Theorem 3.1.8, we have

$$|B_2| \leqslant |A| \ll \delta^{-5}|A_0| \leqslant \delta^{-5}|B_1|,$$

so that (3.7) also holds with $\delta^{-5}|B_1|$ in place of $|B_2|$ in the error term. Hence, choosing B to be the smaller of the two sets B_1 and B_2, we obtain conclusion (ii) of the lemma. □

Iterating Lemma 3.2.2, we either obtain the desired tower of configurations, or we can isolate a very sparse subconfiguration with a large relative density of local relations. Crucially, in the latter case, this density parameter only decreases by a constant factor, while the size of the underlying set drops by almost a power of P.

Proposition 3.2.3. *Let $\mathcal{A} = (A, \alpha_\bullet, H)$ be a configuration with $|Q(\mathcal{A})| = \delta|A||\mathcal{P}|^2$, where $\delta \geqslant 2/\log Y$. Then, one of the following holds:*

1. *There exists a tower of configurations $(\mathcal{A}_i)_{0\leqslant i\leqslant k}$ of height $k \gg (\log Y)^{10}$, such that $\mathcal{A}_0 = \mathcal{A}$ and $|Q(\mathcal{A}_k)| \gg \delta|A_k||\mathcal{P}|^2$ (where A_k is the set of points of the configuration $\mathcal{A}_k$).*
2. *There is $B \subset A$ such that $|Q(\mathcal{A}|_B)| \gg \delta|B||\mathcal{P}|^2$ and $|B| \ll P^{-c_1\delta^6/(\log\log Y)^2}|A|$, where $c_1 > 0$ is an absolute constant.*

Proof. We first dispose of the trivial case where $\delta^6 \log P \leqslant C_0(\log\log Y)^2$ for a sufficiently large absolute constant $C_0 > 0$. In this regime, conclusion (2) is trivially satisfied for $B = A$. Hence, we may assume that $\delta^6 \log P > C_0(\log\log Y)^2$.

By induction, we construct a sequence of triples $(\mathcal{B}_i, B_i, \delta_i)$ where $\mathcal{B}_i$ is a configuration with point set B_i satisfying $|Q(\mathcal{B}_i)| = \delta_i |B_i||\mathcal{P}|^2$. This sequence is initialised at $(\mathcal{B}_0, B_0, \delta_0) := (\mathcal{A}, A, \delta)$.

Suppose that $\mathcal{B}_0, \dots, \mathcal{B}_{i-1}$ have been constructed. If $\delta_{i-1} < \delta/2$, we stop the procedure. Otherwise, we apply Lemma 3.2.2 to $\mathcal{B}_{i-1}$:

- *Lifting step.* If case (i) of Lemma 3.2.2 holds, define $\mathcal{B}_i$ to be the lift of $\mathcal{B}_{i-1}$ produced by the lemma; the corresponding parameter δ_i satisfies
$$\delta_i \geqslant \left(1 - O\left(\frac{1}{(\log Y)^{10}}\right)\right)\delta_{i-1}.$$
- *Halving step.* If, instead, case (ii) of Lemma 3.2.2 holds, there is a subset $B_i \subset B_{i-1}$ with $|B_i| \leqslant \frac{1}{2}|B_{i-1}|$ such that, defining $\mathcal{B}_i := \mathcal{B}_{i-1}|_{B_i}$, we have
$$\delta_i \geqslant \left(1 - O\left(\frac{\delta^{-6}(\log\log Y)^2}{\log P}\right)\right)\delta_{i-1}.$$

Our initial assumption ensures that $\delta_i \geqslant \frac{99}{100}\delta_{i-1}$ in either case. For any $k, m \in \mathbb{N}$, if the first $k+m$ steps of the procedure consist of k lifting steps and m halving steps (in any order), we have
$$\delta_{k+m} \geqslant \exp\left(-O\left(k \cdot \frac{1}{(\log Y)^{10}} + m \cdot \frac{\delta^{-6}(\log\log Y)^2}{\log P}\right)\right)\delta.$$
Hence, the stopping condition $\delta_i < \delta/2$ is never triggered before either $\gg (\log Y)^{10}$ lifting steps or $\gg \delta^6 \log P/(\log\log Y)^2$ halving steps have been performed.

If there are $k \gg (\log Y)^{10}$ lifting steps, defining $\mathcal{A}_0 := \mathcal{A}$ and letting $\mathcal{A}_i$ be the configuration obtained after the i-th lifting step, we obtain a tower of configurations as required for conclusion (1) of the proposition.

Otherwise, there are $m \gg \delta^6 \log P/(\log\log Y)^2$ halving steps. In this case, the configuration $\mathcal{B} = (B, \beta_\bullet, H')$ obtained after these m halving steps satisfies $|B| \leqslant 2^{-m}|A|$ and $|Q(\mathcal{B})| \geqslant \frac{\delta}{2}|B||\mathcal{P}|^2$. Since $Q(\mathcal{B}) \subset Q(\mathcal{A}|_B)$ by definition of a lift, we obtain conclusion (2) of the proposition. □

3.3 Analysis of the top configuration

Having constructed high towers of configurations in the previous section, we now study the configurations that lie at the top of these towers. For such a configuration

$\mathcal{A} = (A, \alpha_\bullet, H)$, the parameter H is exceptionally large, meaning that the local relations in $\mathcal{A}$ hold with much smaller error terms on the frequency side. This rigidity allows us to employ graph-theoretic exploration arguments to extract an initial approximate formula for the frequencies α_x on a highly connected subset of A.

3.3.1 Paths in the graph of local relations

Definition 3.3.1 (Graph of a configuration)**.** Let $\mathcal{A} = (A, \alpha_\bullet, H)$ be a configuration. We define $G(\mathcal{A})$ to be the graph with vertex set A, and with an edge between two distinct points $x, y \in A$ whenever $(x, y, p, q) \in Q(\mathcal{A})$ for some $p, q \in \mathcal{P}$.

Remark 3.3.2. Given a vertex $x \in A$ and $p, q \in \mathcal{P}$, there is at most one neighbour y of x such that $(x, y, p, q) \in Q(\mathcal{A})$, by Definition 3.1.6 and the 1-separation of A.

Similarly, if two distinct points $x, y \in A$ are neighbours in $G(\mathcal{A})$, then the pair $(p, q) \in \mathcal{P}^2$ with $(x, y, p, q) \in Q(\mathcal{A})$ is unique. Indeed, if $(x, y, p, q), (x, y, p', q') \in Q(\mathcal{A})$, then $\left|\frac{p}{q} - \frac{p'}{q'}\right| \leqslant \frac{1}{5Y}$, and thus $(p, q) = (p', q')$, as the left-hand side is either zero or at least $\frac{1}{qq'}$.

Definition 3.3.3 (Prime sequences and prime products of a path)**.** A *path of length* k in $G(\mathcal{A})$ is a sequence $x_\bullet = (x_i)_{i=0}^k$ of vertices of $G(\mathcal{A})$ such that $\{x_{i-1}, x_i\}$ is an edge of $G(\mathcal{A})$ for all $1 \leqslant i \leqslant k$.

The *prime sequences* of the path $x_\bullet$ are the unique sequences $(p_i)_{i=1}^k, (q_i)_{i=1}^k$ of primes in $\mathcal{P}$ such that $(x_{i-1}, x_i, p_i, q_i) \in Q(\mathcal{A})$ for all $1 \leqslant i \leqslant k$. The *prime products* of the path $x_\bullet$ are defined to be the numbers $\prod_{i=1}^k p_i$ and $\prod_{i=1}^k q_i$.

We have the following simple estimate relating the endpoints of a path and the corresponding frequencies, which is similar to [73, Lemma 6.3].

Lemma 3.3.4. *Let $\mathcal{A} = (A, \alpha_\bullet, H)$ be a configuration. Let $x_\bullet = (x_i)_{i=0}^k$ be a path of length $k \leqslant \log Y$ in $G(\mathcal{A})$, with prime products R and S. Then*

$$\left|\frac{R}{S} x_0 - x_k\right| \leqslant k \tag{3.8}$$

and

$$\|S\alpha_{x_0} - R\alpha_{x_k}\| \leqslant \frac{kR}{HP}. \tag{3.9}$$

Proof. We introduce the notation $R_{i\to j} := \prod_{i<\ell\leqslant j} p_\ell$ and $S_{i\to j} := \prod_{i<\ell\leqslant j} q_\ell$ (and use the convention $R_{i\to i} = S_{i\to i} = 1$). For each $1 \leqslant \ell \leqslant k$, we have $(x_{\ell-1}, x_\ell, p_\ell, q_\ell) \in Q(\mathcal{A})$, so that

$$\left|x_\ell - \frac{p_\ell}{q_\ell} x_{\ell-1}\right| \leqslant \frac{1}{10}.$$

By the triangle inequality, this implies that, for all $i < j$,

$$\left| x_j - \frac{R_{i\to j}}{S_{i\to j}} x_i \right| \leqslant \frac{1}{10} \sum_{i<\ell\leqslant j} \frac{R_{\ell\to j}}{S_{\ell\to j}}. \tag{3.10}$$

In particular, using that $\mathcal{P} \subset [P, 2P]$, we obtain the crude bound

$$\left| x_j - \frac{R_{i\to j}}{S_{i\to j}} x_i \right| \leqslant \frac{1}{10} \sum_{i<\ell\leqslant j} 2^{j-\ell} \leqslant \frac{2^{j-i}}{10}.$$

Since $x_i, x_j \in [Y, 2Y]$ and $k \leqslant \log Y$, this implies that

$$\frac{R_{i\to j}}{S_{i\to j}} \in \left[\frac{1}{2} - \frac{2^{j-i}}{10Y}, 2 + \frac{2^{j-i}}{10Y} \right] \subset \left[\frac{1}{3}, 3 \right] \tag{3.11}$$

for all $i < j$. Plugging this estimate back into (3.10) yields

$$\left| x_k - \frac{R_{0\to k}}{S_{0\to k}} x_0 \right| \leqslant \frac{3k}{10},$$

which proves (3.8).

The proof of (3.9) is very similar. For each $1 \leqslant j \leqslant k$, we have

$$\left\| R_{0\to j} S_{j\to k} \alpha_{x_j} - R_{0\to j-1} S_{j-1\to k} \alpha_{x_{j-1}} \right\| \leqslant R_{0\to j-1} S_{j\to k} \left\| p_j \alpha_{x_j} - q_j \alpha_{x_{j-1}} \right\| \leqslant \frac{R_{0\to j} S_{j\to k}}{10HP}.$$

By the triangle inequality, we obtain

$$\| R_{0\to k} \alpha_{x_k} - S_{0\to k} \alpha_{x_0} \| \leqslant \frac{1}{10HP} \sum_{j=1}^{k} R_{0\to j} S_{j\to k}.$$

Using the bound $S_{j\to k} \leqslant 3R_{j\to k}$ given in (3.11), the second conclusion follows. □

Lemma 3.3.5. *Let $\mathcal{A} = (A, \alpha_\bullet, H)$ be a configuration whose graph $G(\mathcal{A})$ has minimum degree at least $\delta|\mathcal{P}|^2$. Suppose that $|\mathcal{P}| \geqslant (\log Y)^4$ and $\delta \geqslant 4/\log|\mathcal{P}|$.*

Then, for any $x \in A$, there exists a path of length

$$\frac{\log Y}{\log|\mathcal{P}|} + O\left(\frac{(\log Y)\log\log Y}{(\log|\mathcal{P}|)^2} + 1 \right)$$

starting and ending at x, whose prime products R and S are distinct.

Proof. Call a path $(x_i)_{i=0}^k$ in $G(\mathcal{A})$ *typical* if its prime products R, S are coprime.

Let $x \in A$. By induction, the number of typical paths of length k with starting vertex x is at least $(\delta|\mathcal{P}|^2/2)^k$, for all $1 \leqslant k \leqslant \delta|\mathcal{P}|/4$. Indeed, a typical path $(x_i)_{i=0}^\ell$ of length $\ell < k$ starting at x, with prime sequences $(p_i)_{i=1}^\ell$ and $(q_i)_{i=1}^\ell$, can be extended to a typical path of length $\ell+1$ by choosing an oriented edge (x_ℓ, y) whose corresponding primes $p, q \in \mathcal{P}$ satisfy $p \notin \{q_1, \dots, q_\ell\}$ and $q \notin \{p_1, \dots, p_\ell\}$, and there are at least

$$\geqslant \delta|\mathcal{P}|^2 - 2\ell|\mathcal{P}| \geqslant \tfrac{1}{2}\delta|\mathcal{P}|^2$$

ways to do so.

A path of length k starting at x is uniquely determined by its prime sequences, and each prime sequence is a permutation of the k prime factors of the corresponding prime product. Hence, given any pair (R, S) of integers, there are $\leqslant (k!)^2$ paths of length k starting at x with prime products R and S.

Therefore, for any positive integer k such that

$$k \leqslant \delta|\mathcal{P}|/4 \quad \text{and} \quad \left(\frac{\delta|\mathcal{P}|^2}{2}\right)^k > (k!)^2|A|, \tag{3.12}$$

there exists some $y \in A$ and two typical paths $(x_i)_{i=0}^k$ and $(y_i)_{i=0}^k$ from x to y not having the same pair of prime products. Let R_1, S_1 and R_2, S_2 be the prime products of $(x_i)_{i=0}^k$ and $(y_i)_{i=0}^k$, respectively. Since these two paths are typical, we have $\gcd(R_1, S_1) = \gcd(R_2, S_2) = 1$, and thus

$$\frac{R_1}{S_1} \neq \frac{R_2}{S_2}. \tag{3.13}$$

Hence, the path obtained by concatenating $(x_i)_{i=0}^k$ with the reverse of $(y_i)_{i=0}^k$ is a path of length $2k$ starting and ending at x. The prime products of this closed path are R_1S_2 and R_2S_1, which are distinct by (3.13).

It only remains to show that the size conditions (3.12) are verified for some $k \in \mathbb{N}$ of the form

$$k = \frac{\log Y}{2\log|\mathcal{P}|} + O\left(\frac{(\log Y)\log\log Y}{(\log|\mathcal{P}|)^2} + 1\right). \tag{3.14}$$

By the intermediate value theorem, there exists a real number $t \in \left[\frac{\log Y}{2\log|\mathcal{P}|}, \frac{\log Y}{\log|\mathcal{P}|}\right]$ such that

$$t - \frac{\log Y}{2\log|\mathcal{P}|} - \frac{4t\log(4\delta^{-1}t)}{2\log|\mathcal{P}|} = 0.$$

Indeed, the left-hand side is clearly negative at $t = \frac{\log Y}{2\log|\mathcal{P}|}$, and the non-negativity at $t = \frac{\log Y}{\log|\mathcal{P}|}$ is equivalent to

$$\log|\mathcal{P}| \geqslant 4\log(4\delta^{-1}\log Y/\log|\mathcal{P}|),$$

which holds since $\delta \geqslant 4/\log|\mathcal{P}|$ and $|\mathcal{P}| \geqslant (\log Y)^4$.

Define $k = \lceil t \rceil$. Recalling that $Y \geqslant P^3$ (see Definition 3.1.1), we have $t \geqslant 3/2$, so that $k < 2t$. The bound $t \leqslant \log Y/\log|\mathcal{P}|$ and the inequalities $\delta \geqslant 4/\log|\mathcal{P}|$ and $|\mathcal{P}| \geqslant (\log Y)^4$ readily imply (3.14) and the first condition in (3.12). For the second condition in (3.12), we observe that

$$2^k\delta^{-k}(k!)^2 \leqslant (2\delta^{-1}k)^{2k} < (4\delta^{-1}t)^{4t} = \frac{|\mathcal{P}|^{2t}}{Y} \leqslant \frac{|\mathcal{P}|^{2k}}{|A|}.$$

Hence, this choice of k satisfies all the required conditions. □

3.3.2 Extraction of a global formula

We now consider a configuration $\mathcal{A} = (A, \alpha_\bullet, H)$ with many local relations and a very large parameter H. Using the preceding lemmas, we show that many frequencies α_x can be approximated by rationals a_x/q_0 with a common denominator q_0. Our control over the size of q_0 is inevitably quite weak at this stage; we will improve it in Section 3.4 using a 'modular' expansion estimate.

Lemma 3.3.6. *Let $d, n, \delta > 0$. Let G be a graph on n vertices, with maximum degree d and $\geqslant \delta dn$ edges. Then, there exists a subgraph G' of G with minimum degree $\geqslant \delta d/2$ and $\geqslant \delta dn/2$ edges.*

Proof. Let G' be the graph obtained by iteratively removing all vertices of degree $< \delta d/2$. This procedure removes at most $\delta dn/2$ edges in total. Hence, G' has the desired properties. □

Lemma 3.3.7. *Let $\mathcal{A} = (A, \alpha_\bullet, H)$ be a configuration such that $|Q(\mathcal{A})| \geqslant \delta |A||\mathcal{P}|^2$, where $H \geqslant Y^{C_1}$ and $\delta \geqslant C_1/\log|\mathcal{P}|$ for some sufficiently large absolute constant $C_1 > 0$. Assume that $|\mathcal{P}| \geqslant (\log Y)^4$.*

Then, there exists a subset $A_2 \subset A$ and an integer

$$1 \leqslant q_0 \leqslant \exp\left(O\left(\frac{(\log Y)\log\log Y}{\log|\mathcal{P}|} + \log|\mathcal{P}|\right)\right)$$

with the following properties.

1. *For each $x \in A_2$, there exist $a_x \in \mathbb{Z}$ and $\beta_x \in \mathbb{R}$ such that $(a_x, q_0) = 1$, $|\beta_x| \leqslant Y^{O(1)}/H$ and*
$$\alpha_x = \frac{a_x}{q_0} + \beta_x \pmod 1.$$
2. *For each $x \in A_2$, there are $\gg \delta|\mathcal{P}|^2$ quadruples $(x, y, p, q) \in Q(\mathcal{A}|_{A_2})$ such that $(pq, q_0) = 1$, $p \neq q$,*
$$qa_x \equiv pa_y \pmod{q_0} \quad \text{and} \quad |q\beta_x - p\beta_y| \leqslant \frac{1}{10H}.$$

Proof. Since $|Q(\mathcal{A})| \geqslant \delta|A||\mathcal{P}|^2$, the graph $G(\mathcal{A})$ has

$$\geqslant \tfrac{1}{2}(\delta|A||\mathcal{P}|^2 - |A||\mathcal{P}|) \geqslant \tfrac{1}{3}\delta|A||\mathcal{P}|^2$$

edges (as every edge corresponds to exactly two quadruples $(x, y, p, q) \in Q(\mathcal{A})$ with $p \neq q$). By Lemma 3.3.6, there is a subset $A_1 \subset A$ such that $G(\mathcal{A}|_{A_1})$ has minimum degree $\geqslant \frac{1}{6}\delta|\mathcal{P}|^2$.

Let $x \in A_1$. Let κ denote the quantity $\kappa := (\log Y)\log\log Y/\log|\mathcal{P}|$. By Lemma 3.3.5, there exists a path of length ℓ with

$$\ell = \frac{\log Y}{\log|\mathcal{P}|} + O\bigg(\frac{\kappa}{\log|\mathcal{P}|} + 1\bigg), \tag{3.15}$$

starting and ending at x, whose prime products R_x and S_x are distinct. Applying Lemma 3.3.4 to this path, we get

$$\bigg|\frac{R_x}{S_x}x - x\bigg| \leqslant \ell \quad \text{and} \quad \|S_x\alpha_x - R_x\alpha_x\| \leqslant \frac{\ell R_x}{HP}.$$

Since $R_x, S_x \leqslant (2P)^\ell$, we may rewrite these bounds as

$$|R_x - S_x| \leqslant \frac{\ell 2^\ell P^\ell}{Y} \quad \text{and} \quad \|(R_x - S_x)\alpha_x\| \leqslant \frac{\ell 2^\ell P^{\ell-1}}{H}.$$

The last inequality implies that

$$\alpha_x = \frac{a_x}{q_x} + \beta_x \pmod 1 \tag{3.16}$$

for some reduced fraction a_x/q_x with denominator $q_x \mid R_x - S_x$ (with $a_x, q_x \in \mathbb{N}$), and some $\beta_x \in \mathbb{R}$ satisfying $|\beta_x| \leqslant \ell 2^\ell P^{\ell-1}/H$. In particular, $q_x \leqslant |R_x - S_x| \leqslant \ell 2^\ell P^\ell/Y$. By Chebyshev's estimate and our estimate (3.15) for ℓ, we have

$$\frac{P^\ell}{Y} \leqslant \frac{|\mathcal{P}|^\ell}{Y} \cdot (O(\log Y))^\ell \leqslant e^{O(\kappa + \log|\mathcal{P}|)}.$$

Thus, $q_x \leqslant e^{O(\kappa+\log|\mathcal{P}|)}$ and $|\beta_x| \leqslant Y^{O(1)}/H$.

The next step is to pass to a suitable subset $A_2 \subset A_1$ on which all q_x are equal. Let

$$q_0 := \max\{q_x : x \in A_1\}.$$

Let G_0 be the subgraph of $G(\mathcal{A}|_{A_1})$ obtained by discarding all $(x, y, p, q) \in Q(\mathcal{A}|_{A_1})$ with $(pq, q_0) > 1$. Recall that $G(\mathcal{A}|_{A_1})$ has minimum degree $\geqslant \frac{1}{6}\delta|\mathcal{P}|^2$. Therefore, the subgraph G_0 has minimum degree

$$\geqslant \tfrac{1}{6}\delta|\mathcal{P}|^2 - 2|\mathcal{P}|\sum_{p\in\mathcal{P}} \mathbf{1}_{p|q_0} \geqslant \tfrac{1}{6}\delta|\mathcal{P}|^2 - 2|\mathcal{P}|\log q_0 \gg \delta|\mathcal{P}|^2,$$

where we used that $\log q_0 \ll \kappa + \log|\mathcal{P}| \ll \log Y$ and $\delta \geqslant C_1/\log|\mathcal{P}| \geqslant C_1 \log Y/|\mathcal{P}|$ for C_1 a sufficiently large constant.

Let $x_0 \in A_1$ be any point such that $q_{x_0} = q_0$. Let A_2 be the connected component of x_0 in G_0.

Suppose that $x, y \in A_2$ are neighbours in G_0, meaning that $(x, y, p, q) \in Q(\mathcal{A}|_{A_2})$ for some $p, q \in \mathcal{P}$ not dividing q_0. Then $\|q\alpha_x - p\alpha_y\| \leqslant (10H)^{-1}$, which by (3.16) implies that

$$\bigg\|\frac{qq_ya_x - pq_xa_y}{q_xq_y} + q\beta_x - p\beta_y\bigg\| \leqslant \frac{1}{10H}. \tag{3.17}$$

Since $|q\beta_x - p\beta_y| \leqslant 2P(|\beta_x| + |\beta_y|) \leqslant Y^{O(1)}/H$, we get

$$\left\| \frac{qq_y a_x - pq_x a_y}{q_x q_y} \right\| \leqslant \frac{Y^{O(1)}}{H}.$$

The left-hand side is either zero or a fraction with denominator at most $q_x q_y \leqslant Y^{O(1)}$. The latter case is impossible by our assumption that $H \geqslant Y^{C_1}$ for some large enough constant C_1. Therefore, it must be the case that

$$qq_y a_x \equiv pq_x a_y \pmod{q_x q_y}. \tag{3.18}$$

Plugging this back into (3.17) yields $\|q\beta_x - p\beta_y\| \leqslant (10H)^{-1}$. This modulo 1 bound can be upgraded to $|q\beta_x - p\beta_y| \leqslant (10H)^{-1}$, using that $|q\beta_x - p\beta_y| \leqslant Y^{O(1)}/H \leqslant 1/2$ (provided C_1 is sufficiently large).

Now, suppose that $x, y \in A_2$ are neighbours in G_0 such that $q_x = q_0$. By (3.18), we have

$$q_0 \mid qq_y a_x.$$

Since $(a_x, q_x) = 1$ (by definition of a_x, q_x) and $(q_0, q) = 1$ (by definition of G_0), we deduce that $q_0 \mid q_y$. By maximality of q_0, we conclude that $q_y = q_0$. Thus, the property of having $q_x = q_0$ propagates through the edges of G_0. Since A_2 is connected in G_0, and $x_0 \in A_2$ satisfies $q_{x_0} = q_0$, we obtain that $q_x = q_0$ for all $x \in A_2$. In particular, (3.18) simplifies to $qa_x \equiv pa_y \pmod{q_0}$. This concludes the proof of Lemma 3.3.7. □

The next technical lemma will be used to derive an approximate formula for the frequencies β_x in the conclusion of Lemma 3.3.7.

Lemma 3.3.8. *Let $\mathcal{B} = (B, \beta_\bullet, H)$ be a configuration whose frequencies satisfy $|\beta_x| \leqslant P^{-2\log Y}$ for all $x \in B$.*[3] *Suppose that the minimum degree of the graph $G(\mathcal{B})$ is $\geqslant \delta|\mathcal{P}|^2 \geqslant 1$.*

Then, there exists $B' \subset B$ with $|Q(\mathcal{B}|_{B'})| \gg \delta|B'||\mathcal{P}|^2$ and $x_0 \in B'$ such that, for all $x \in B'$,

$$\beta_x = \frac{x_0\beta_{x_0}}{x} + O\left(\frac{|\beta_{x_0}|\log Y}{Y} + \frac{\log Y}{HP}\right).$$

Furthermore, there is a multiset $V \subset \left[\frac{1}{3}Y, 6Y\right]$ with $O(\log Y)$ elements in any unit interval, such that $\gg \delta|V||\mathcal{P}|^2$ quadruples $(v_1, v_2, p, q) \in V^2 \times \mathcal{P}^2$ satisfy

$$|\beta_{x_0}||pv_1 - qv_2| \ll \frac{Y\log Y}{H}.$$

[3] Here we slightly abuse notation by identifying each $\beta_x \in \mathbb{R}/\mathbb{Z}$ with its unique representative in $[-1/2, 1/2)$.

Proof. We first pass to a subgraph of $G(\mathcal{B})$ of reasonable diameter, while maintaining a large edge concentration.

Let $x_0 \in B$. For $i \geqslant 0$, let B_i be the set of all vertices $x \in B$ at distance at most i from x_0 in $G(\mathcal{B})$. Since $|B| \leqslant Y$, there exists an integer $1 \leqslant r \leqslant \log Y$ such that $|B_r| \leqslant 3|B_{r-1}|$. Then, by the minimum degree assumption on $G(\mathcal{B})$, we have

$$|Q(\mathcal{B}|_{B_r})| \geqslant |B_{r-1}|\delta|\mathcal{P}|^2 \geqslant \tfrac{1}{3}\delta|B_r||\mathcal{P}|^2. \tag{3.19}$$

We set $B' := B_r$. For any $x \in B'$, there exists a path of length $\leqslant \log Y$ from x_0 to x in $G(\mathcal{B})$. Hence, by Lemma 3.3.4, there are $R_x, S_x \leqslant (2P)^{\log Y}$ such that

$$\left|\frac{R_x}{S_x}x_0 - x\right| \leqslant \log Y \quad \text{and} \quad \|S_x\beta_{x_0} - R_x\beta_x\| \leqslant \frac{(\log Y)R_x}{HP}. \tag{3.20}$$

In particular, $R_x/S_x \in \left[\frac{1}{3}, 3\right]$. Hence, we may write the first part of (3.20) as

$$\left|\frac{S_x}{R_x} - \frac{x_0}{x}\right| \leqslant \frac{3\log Y}{Y}. \tag{3.21}$$

The second part of (3.20) can be simplified using the assumption on the size of the frequencies of $\mathcal{B}$. Indeed, since $|\beta_{x_0}|, |\beta_x| \leqslant P^{-2\log Y}$ and $R_x, S_x \leqslant (2P)^{\log Y}$, we have $|S_x\beta_{x_0} - R_x\beta_x| \leqslant 1/2$, which implies that $\|S_x\beta_{x_0} - R_x\beta_x\| = |S_x\beta_{x_0} - R_x\beta_x|$. Therefore,

$$\left|\beta_x - \frac{S_x}{R_x}\beta_{x_0}\right| \leqslant \frac{\log Y}{HP}. \tag{3.22}$$

Combining the estimates (3.21) and (3.22), we see that every $x \in B'$ satisfies the approximate formula

$$\beta_x = \frac{x_0\beta_{x_0}}{x} + O\left(\frac{|\beta_{x_0}|\log Y}{Y} + \frac{\log Y}{HP}\right).$$

This proves the first part of the lemma.

For the second part, we define the multiset V consisting of the values $v_x := \frac{S_x}{R_x}Y$ for $x \in B'$. Since $S_x/R_x \in \left[\frac{1}{3}, 3\right]$, it is immediate that $V \subset \left[\frac{1}{3}Y, 6Y\right]$. Moreover, for any $x, y \in B'$ we have $|v_x - v_y| \geqslant \frac{1}{4}|x-y| - 6\log Y$ by (3.21), so that V has $O(\log Y)$ elements in any unit interval.

Finally, for any quadruple $(x, y, p, q) \in Q(\mathcal{B}|_{B'})$, we have

$$\frac{|\beta_{x_0}|}{Y}|qv_x - pv_y| = \left|q\frac{S_x}{R_x}\beta_{x_0} - p\frac{S_y}{R_y}\beta_{x_0}\right| \leqslant |q\beta_x - p\beta_y| + O\left(\frac{\log Y}{H}\right)$$

by (3.22). Since $|q\beta_x - p\beta_y| \ll 1/H$ by definition of $Q(\mathcal{B}|_{B'})$, the desired bound follows. □

3.4 Proof of the global structure theorem

We start this section by proving an unconditional expansion estimate, Proposition 3.4.4. Combining this with the results of the previous sections, we then establish the global structure theorem, Theorem 3.1.9, from which we deduce our main Fourier uniformity result, Theorem 1.3.3.

3.4.1 Vinogradov-Korobov expansion estimate

To circumvent the need for GRH, we employ a large value estimate for Dirichlet polynomials over primes. Specifically, we bound the number of pairs (χ, t) for which $\left|\sum_{p\in\mathcal{P}} \chi(p)p^{it}\right|$ is exceptionally large.

Lemma 3.4.1. *Let $T \geqslant 3$ and $q \in \mathbb{N}$. Let $\mathcal{T}$ be a set of pairs (χ, t) where χ is a Dirichlet character modulo q and $t \in [-T, T]$. Suppose that $|t_1 - t_2| \geqslant 1$ for distinct pairs $(\chi, t_1), (\chi, t_2) \in \mathcal{T}$.*

Let $P \geqslant 3$ and $k \in \mathbb{N}$ be such that $P^k \leqslant T^2$. Let (c_p) be a sequence of 1-bounded complex numbers supported on the primes $p \in [P, 2P]$.

Then, uniformly for $0 < \eta < 1/2$,

$$\sum_{(\chi,t)\in\mathcal{T}} \left|\sum_p \chi(p)c_p p^{it}\right|^{2k} \leqslant k^{O(k)}P^{2k}\Big(1 + |\mathcal{T}|q^{\eta}P^{-\eta k}\big(T^{5\eta^{3/2}}\log T + \eta^{-1}\big)\Big).$$

Lemma 3.4.1 is a straightforward adjustment of [40, Lemma 6.6], which treats the case $k = 1$ (and achieves a more refined bound). For completeness,[4] we provide a detailed proof of Lemma 3.4.1 below.

The key ingredient is the following Vinogradov-Korobov type upper bound for Dirichlet L-functions.

Lemma 3.4.2. *Let $q \geqslant 1$ and let χ be a Dirichlet character modulo q. For $1/2 \leqslant \sigma < 1$ and $|t| \geqslant 3$, we have*

$$|L(\sigma + it, \chi)| \ll q^{1-\sigma}|t|^{5(1-\sigma)^{3/2}}(\log|t|)^{2/3} + \frac{q^{1-\sigma}}{1-\sigma}.$$

Proof. As explained in [40, p.92], this is a direct consequence of the bound [17, Theorem 1] (see also [51]) on the Hurwitz zeta function $\zeta(s, u)$, using the identity $L(s, \chi) = q^{-s}\sum_{m=1}^{q}\chi(m)\zeta(s, m/q)$. □

[4] The proof of [40, Lemma 6.6] is not explicitly provided in [40], but instead suggested as a modification of [36, Lemma 4.4].

Proof of Lemma 3.4.1. By multiplicativity, we have

$$\Big(\sum_p \chi(p)c_p p^{it}\Big)^k = \sum_n a_n \chi(n) n^{it}$$

for some complex coefficients $|a_n| \leqslant k!$ supported on $n \in [P^k, (2P)^k]$. Let $W : \mathbb{R} \to \mathbb{R}$ be a smooth function such that $\mathbf{1}_{[1,2]} \leqslant W \leqslant \mathbf{1}_{[0.5,2.5]}$ pointwise. By repeated integration by parts, its Mellin transform satisfies $|\widetilde{W}(\sigma+it)| \ll_A (1+|t|)^{-A}$ for $|\sigma| \leqslant 2$ and $t \in \mathbb{R}$, for every $A > 0$. Fix some scale $P^k \leqslant N \leqslant (2P)^k$. By duality (see [30, Section 7.1]) and summing over dyadic scales, it suffices to prove the estimate

$$\sum_n W\Big(\frac{n}{N}\Big)\Big|\sum_{(\chi,t)\in\mathcal{T}} b_{\chi,t}\chi(n)n^{it}\Big|^2 \ll N\Big(1+|\mathcal{T}|q^\eta N^{-\eta}\big(T^{5\eta^{3/2}}\log T+\eta^{-1}\big)\Big)\sum_{(\chi,t)\in\mathcal{T}}|b_{\chi,t}|^2 \tag{3.23}$$

for arbitrary coefficients $b_{\chi,t} \in \mathbb{C}$.

Expanding the square, the left-hand side of (3.23) becomes

$$\sum_{(\chi_1,t_1),(\chi_2,t_2)\in\mathcal{T}} b_{\chi_1,t_1}\overline{b_{\chi_2,t_2}}\sum_n W\Big(\frac{n}{N}\Big)\chi_1\overline{\chi_2}(n)n^{i(t_1-t_2)}. \tag{3.24}$$

By Mellin inversion,

$$\sum_n W\Big(\frac{n}{N}\Big)\chi_1\overline{\chi_2}(n)n^{i(t_1-t_2)} = \frac{1}{2\pi i}\int_{2-i\infty}^{2+i\infty} N^s\widetilde{W}(s)L\big(s-i(t_1-t_2),\chi_1\overline{\chi_2}\big)ds.$$

We can restrict the range of integration to $|\mathrm{Im}(s)| \leqslant 3T$, at the cost of a negligible error term $O_A(T^{-A})$, using that $N \leqslant (2P)^k \leqslant T^{O(1)}$. We then shift the contour to $\mathrm{Re}(s) = 1-\eta$, obtaining a contribution

$$\mathbf{1}_{\chi_1=\chi_2}N^{1-i(t_1-t_2)}\widetilde{W}(1-i(t_1-t_2)) \ll_A \mathbf{1}_{\chi_1=\chi_2}N(1+|t_1-t_2|)^{-A}$$

from the possible pole at $s = 1-i(t_1-t_2)$. Appealing to Lemma 3.4.2, the contribution of the remaining integrals is bounded by

$$\ll_A N^{1-\eta}q^\eta\Big(T^{5\eta^{3/2}}\log T+\eta^{-1}\Big)+T^{-A}.$$

Combining these bounds, we get that (3.24) is

$$\ll \sum_{(\chi_1,t_1),(\chi_2,t_2)\in\mathcal{T}} |b_{\chi_1,t_1}||b_{\chi_2,t_2}|\Big(\frac{\mathbf{1}_{\chi_1=\chi_2}N}{(1+|t_1-t_2|)^2}+N^{1-\eta}q^\eta\Big(T^{5\eta^{3/2}}\log T+\eta^{-1}\Big)\Big).$$

By the inequality $2|b_1||b_2| \leqslant |b_1|^2+|b_2|^2$ and the separation assumption on $\mathcal{T}$, we obtain (3.23), which concludes the proof of Lemma 3.4.1. □

Lemma 3.4.3. *Let T be sufficiently large.*

Let $P = \exp((\log T)^\theta)$ for some $\frac{1}{1000} \leqslant \theta \leqslant \frac{2}{3} - \frac{1}{1000}$ and let $\mathcal{P}$ be the set of all primes in $[P, 2P]$. Let $1 \leqslant q \leqslant \exp((\log T)^{1-\theta/2})$ be an integer.

Let $\mathcal{T}$ be a set of pairs (χ, t) such that

$$\left|\sum_{p\in\mathcal{P}} \chi(p)p^{it}\right| \geqslant \frac{|\mathcal{P}|}{(\log T)^{100}},$$

where χ is a Dirichlet character modulo q and $t \in [-T, T]$. Suppose that $|t_1 - t_2| \geqslant 1$ for distinct pairs $(\chi, t_1), (\chi, t_2) \in \mathcal{T}$.

Then

$$|\mathcal{T}| \leqslant \exp\big(O\big((\log T)^{1-\frac{3\theta}{2}}(\log\log T)^3\big)\big).$$

Proof. Applying Lemma 3.4.1 with

$$\begin{cases} \eta = (\log T)^{-\theta}(\log\log T)^2 \\ k = \lceil (\log T)^{1-\frac{3\theta}{2}}(\log\log T)^2 \rceil, \end{cases}$$

and all coefficients $c_p = 1$ (the condition $P^k \leqslant T^2$ is satisfied for large T), we obtain

$$|\mathcal{T}|\left(\frac{|\mathcal{P}|}{(\log T)^{100}}\right)^{2k} \leqslant k^{O(k)}P^{2k}\Big(1 + |\mathcal{T}|q^\eta P^{-\eta k}\big(T^{5\eta^{3/2}}\log T + \eta^{-1}\big)\Big).$$

By our choices of parameters, this simplifies to

$$|\mathcal{T}| \leqslant (\log T)^{O(k)}\Big(1 + |\mathcal{T}|q^\eta T^{5\eta^{3/2}}P^{-\eta k}\Big).$$

Thus, either

$$|\mathcal{T}| \leqslant (\log T)^{O(k)} \leqslant \exp\big(O\big((\log T)^{1-\frac{3\theta}{2}}(\log\log T)^3\big)\big)$$

and we are done, or

$$P^{\eta k} \leqslant (\log T)^{O(k)}q^\eta T^{5\eta^{3/2}}. \tag{3.25}$$

However, in view of the estimates

$$\begin{cases} q^\eta \leqslant \exp\big((\log T)^{1-3\theta/2}(\log\log T)^2\big) \\ T^{\eta^{3/2}} = \exp\big((\log T)^{1-3\theta/2}(\log\log T)^3\big) \\ P^{\eta k} \geqslant \exp\big((\log T)^{1-3\theta/2}(\log\log T)^4\big), \end{cases}$$

the inequality (3.25) cannot hold for large enough T. □

We can now establish our expansion estimate, Proposition 3.4.4, by passing to the Fourier side and applying the preceding large value estimate.

Proposition 3.4.4 (Expansion Estimate)**.** *Let Y be sufficiently large.*

Let $P = \exp((\log Y)^\theta)$ for some $\frac{1}{100} \leqslant \theta \leqslant \frac{2}{3} - \frac{1}{100}$ and let $\mathcal{P}$ be the set of primes in $[P, 2P]$.

Let $A \subset \left[\frac{1}{10}Y, 10Y\right]$ be a multiset with $O(\log Y)$ elements in any unit interval.

Let $1 \leqslant q_0 \leqslant \exp((\log Y)^{1-\theta/2})$ be an integer, let $\mathcal{P}' \subset \mathcal{P}$ consist of all primes not dividing q_0, and let $(a_x)_{x\in A}$ be a sequence of integers coprime to q_0.

Fix $1/Y \leqslant \varepsilon \leqslant 1$. Suppose that there are $\geqslant (\log Y)^{-100}|A||\mathcal{P}|^2$ quadruples $(x, y, p, q) \in A^2 \times (\mathcal{P}')^2$ such that

$$|px - qy| \leqslant \varepsilon P \quad \text{and} \quad qa_x \equiv pa_y \pmod{q_0}. \tag{3.26}$$

Then

$$\varepsilon^{-1} q_0 \frac{Y}{|A|} \leqslant \exp\big(O\big((\log Y)^{1-\frac{3\theta}{2}}(\log\log Y)^3\big)\big).$$

Taking $q_0 = \varepsilon = 1$ in Proposition 3.4.4, we immediately deduce the following.

Corollary 3.4.5. *Suppose that $P = \exp\big((\log Y)^\theta\big)$ for some $\frac{1}{100} \leqslant \theta \leqslant \frac{2}{3} - \frac{1}{100}$.*

Let $\mathcal{A} = (A, \alpha_\bullet, H)$ be a configuration such that $|Q(\mathcal{A})| \geqslant \delta|A||\mathcal{P}|^2$ for some $\delta \geqslant (\log Y)^{-100}$. Then,

$$\frac{Y}{|A|} \leqslant \exp\big(O\big((\log Y)^{1-\frac{3\theta}{2}}(\log\log Y)^3\big)\big).$$

Proof of Proposition 3.4.4. By standard rounding and multiplicity arguments, we may assume that $A \subset \left[\frac{1}{10}Y, 10Y\right]$ is a set of integer multiples of ε (rather than a multiset of arbitrary real numbers), such that there are $\gg (\log Y)^{-102}|A||\mathcal{P}|^2$ quadruples $(x, y, p, q) \in A^2 \times (\mathcal{P}')^2$ satisfying

$$|px - qy| \leqslant 5\varepsilon P \quad \text{and} \quad qa_x \equiv pa_y \pmod{q_0}. \tag{3.27}$$

Let $\Phi : \mathbb{R} \to \mathbb{R}$ be a Schwartz function such that $\Phi \geqslant \mathbf{1}_{[-50,\,50]}$ and whose Fourier transform is supported on $[-1, 1]$. If $|px - qy| \leqslant 5\varepsilon P$, then $|\log px - \log qy| \leqslant 50\varepsilon/Y$ by the mean value theorem. Thus, the number of quadruples $(x, y, p, q) \in A^2 \times (\mathcal{P}')^2$ satisfying (3.27) is

$$\leqslant \sum_{x,y\in A} \sum_{p,q\in\mathcal{P}'} \Phi\left(\frac{Y}{\varepsilon}\log\frac{px}{qy}\right)\mathbf{1}_{qa_x\equiv pa_y \,(\mathrm{mod}\, q_0)}.$$

By orthogonality of characters and Fourier inversion, this expression can be rewritten as

$$\frac{1}{\varphi(q_0)} \sum_{\chi \,(\mathrm{mod}\, q_0)} \frac{\varepsilon}{Y} \int_{-Y/\varepsilon}^{Y/\varepsilon} \widehat{\Phi}\left(\frac{\varepsilon t}{Y}\right) \left|\sum_{x\in A} \chi(a_x) x^{it}\right|^2 \left|\sum_{p\in\mathcal{P}} \overline{\chi(p)} p^{it}\right|^2 dt. \tag{3.28}$$

We bound (3.28) by separating the contribution of those (χ, t) for which the prime Dirichlet polynomial is small from those for which it is large. By the mean-value theorem for Dirichlet polynomials [30, Theorem 9.1], recalling that $\varepsilon^{-1}A \subset \mathbb{N}$, we have

$$\frac{\varepsilon}{Y}\int_{-Y/\varepsilon}^{Y/\varepsilon}\left|\sum_{x\in A}\chi(a_x)x^{it}\right|^2 dt = \frac{\varepsilon}{Y}\int_{-Y/\varepsilon}^{Y/\varepsilon}\left|\sum_{x\in A}\chi(a_x)(\varepsilon^{-1}x)^{it}\right|^2 dt \ll |A|.$$

Thus, for any $w > 0$, (3.28) is bounded by

$$\ll (w|\mathcal{P}|)^2|A| + \frac{\varepsilon|A|^2|\mathcal{P}|^2}{Y\varphi(q_0)}\sum_{\chi \,(\mathrm{mod}\, q_0)}\int_{-Y/\varepsilon}^{Y/\varepsilon}\mathbf{1}_{\{|\sum_{p\in\mathcal{P}}\overline{\chi(p)}p^{it}|>w|\mathcal{P}|\}}dt.$$

Choose $w := (\log(Y/\varepsilon))^{-100}$. Discretising and applying Lemma 3.4.3 with $T = Y/\varepsilon$ (in particular, $T \in [Y, Y^2]$), we have

$$\sum_{\chi \,(\mathrm{mod}\, q_0)}\int_{-Y/\varepsilon}^{Y/\varepsilon}\mathbf{1}_{\{|\sum_{p\in\mathcal{P}}\overline{\chi(p)}p^{it}|>w|\mathcal{P}|\}}dt \leqslant \exp\big(O\big((\log Y)^{1-\frac{3\theta}{2}}(\log\log Y)^3\big)\big).$$

Combining the above estimates, we obtain

$$\frac{|A||\mathcal{P}|^2}{(\log Y)^{102}} \ll \frac{|A||\mathcal{P}|^2}{(\log Y)^{200}} + \frac{\varepsilon|A|^2|\mathcal{P}|^2}{Y\varphi(q_0)}\exp\big(O\big((\log Y)^{1-\frac{3\theta}{2}}(\log\log Y)^3\big)\big).$$

Rearranging and using that $\varphi(n) \gg n^{1-o(1)}$, the result follows for Y large enough. □

3.4.2 Proof of the global structure theorem

Combining this expansion estimate with the results of Section 3.3 yields the following approximate formula for the frequencies of configurations having a very large parameter H.

Lemma 3.4.6. *Suppose that Y is sufficiently large and $P = \exp((\log Y)^\theta)$ for some $\frac{1}{100} \leqslant \theta \leqslant \frac{2}{3} - \frac{1}{100}$.*

Let $\mathcal{A} = (A, \alpha_\bullet, H)$ be a configuration with $|Q(\mathcal{A})| \geqslant \delta|A||\mathcal{P}|^2$, where $H \geqslant P^{3\log Y}$ and $\delta \geqslant C_1/\log|\mathcal{P}|$ for some sufficiently large absolute constant $C_1 > 0$.

Then, writing $f(Y) := (\log Y)^{1-\frac{3\theta}{2}}(\log\log Y)^3$, there exist

- *a subset $A' \subset A$ with $|Q(\mathcal{A}|_{A'})| \gg \delta|A'||\mathcal{P}|^2$,*
- *an integer $q_0 \leqslant e^{O(f(Y))}$, and*
- *a real number t with $|t| \leqslant e^{O(f(Y))}\frac{Y^2}{HP}$,*

such that, for every $x \in A'$, there is an integer a_x coprime to q_0 such that

$$\alpha_x = \frac{a_x}{q_0} + \frac{t}{x} + O\left(\frac{e^{O(f(Y))}}{HP}\right) \pmod 1. \tag{3.29}$$

Proof. Applying Lemma 3.3.7, we obtain a subset $A_2 \subset A$, an integer q_0 and sequences $(a_x)_{x\in A_2}$ and $(\beta_x)_{x\in A_2}$ satisfying the properties given in the lemma. In particular, q_0 satisfies

$$q_0 \leqslant \exp\big(O\big((\log Y)^{1-\theta}\log\log Y + (\log Y)^{\theta}\big)\big). \tag{3.30}$$

Given the range of θ, this implies that $q_0 \leqslant \exp((\log Y)^{1-\theta/2})$ for Y sufficiently large, which is the condition needed to apply the expansion estimate, Proposition 3.4.4.

We apply it to the set A_2; all the other assumptions are satisfied since, by property (2) of Lemma 3.3.7, there are $\gg \delta|A_2||\mathcal{P}|^2$ quadruples $(x,y,p,q) \in (A_2)^2 \times \mathcal{P}^2$ such that $(pq, q_0) = 1$, $|px - qy| \leqslant P/10$ and $qa_x \equiv pa_y \pmod{q_0}$. Hence, Proposition 3.4.4 gives $q_0 \leqslant e^{O(f(Y))}$, thus sharpening our initial bound (3.30).

Let $\mathcal{B} := (A_2, \beta_\bullet, H)$. This configuration satisfies the assumptions of Lemma 3.3.8. Indeed, $G(\mathcal{B})$ has minimum degree $\gg \delta|\mathcal{P}|^2$ by property (2) of Lemma 3.3.7, and the frequencies $(\beta_x)_{x\in A_2}$ satisfy $|\beta_x| \leqslant P^{-2\log Y}$ for Y large enough. Lemma 3.3.8 then provides a subset $B' \subset A_2$ with $|Q(\mathcal{B}|_{B'})| \gg \delta|B'||\mathcal{P}|^2$ and an element $x_0 \in B'$ such that, for every $x \in B'$, we have

$$\beta_x = \frac{x_0\beta_{x_0}}{x} + O\left(\frac{|\beta_{x_0}|\log Y}{Y} + \frac{\log Y}{HP}\right).$$

Combining this with the formula for α_x from Lemma 3.3.7, we obtain that, for every $x \in B'$,

$$\alpha_x = \frac{a_x}{q_0} + \frac{x_0\beta_{x_0}}{x} + O\left(\frac{|\beta_{x_0}|\log Y}{Y} + \frac{\log Y}{HP}\right) \pmod 1. \tag{3.31}$$

It remains to estimate $|\beta_{x_0}|$. To do so, we use the final property in Lemma 3.3.8, which states that there is a multiset $V \subset \left[\frac{1}{3}Y, 6Y\right]$ with $O(\log Y)$ elements in any unit interval, such that $\gg \delta|V||\mathcal{P}|^2$ quadruples $(v_1, v_2, p, q) \in V^2 \times \mathcal{P}^2$ satisfy

$$|\beta_{x_0}||pv_1 - qv_2| \ll \frac{Y\log Y}{H}.$$

If $|\beta_{x_0}| > e^{Cf(Y)}\frac{Y}{HP}$ for some large enough constant $C > 0$, this contradicts the expansion estimate, Proposition 3.4.4, applied to the set V with $q_0 = 1$. Thus, we must have $|\beta_{x_0}| \leqslant e^{O(f(Y))}\frac{Y}{HP}$ which, when plugged back into the formula (3.31) for the frequencies, yields the desired conclusion. □

We now have all the necessary ingredients to prove the global structure theorem, which we restate for convenience.

Theorem 3.1.9 (Global structure theorem)**.** *Let $\varepsilon_0 > 0$ and let Y be sufficiently large in terms of ε_0. Let $P := \exp((\log Y)^{2/5+\varepsilon})$ for some $\varepsilon_0 < \varepsilon < \frac{1}{5}$.*

Let $\mathcal{A} = (A, \alpha_\bullet, H)$ be a configuration such that $|Q(\mathcal{A})| \geqslant \delta|A||\mathcal{P}|^2$ for some $\delta \geqslant (\log\log Y)^{-10}$.

Let $\kappa := \exp((\log Y)^{2/5-\varepsilon})$. *Then, there is a subset* $A' \subset A$ *of size*

$$|A'| \geqslant \kappa^{-1}|A|,$$

such that, for all $x \in A'$, *we have the approximate formula*

$$\alpha_x = \frac{a_0}{q_0} + \frac{T_0}{x} + O\Big(\frac{\kappa}{HP}\Big) \pmod 1,$$

where a_0, q_0 *are coprime integers with* $1 \leqslant q_0 \leqslant \kappa$ *and* $T_0 \in \mathbb{R}$ *satisfies* $|T_0| \leqslant \kappa \frac{Y^2}{HP}$.

Proof of Theorem 3.1.9. We start by applying Proposition 3.2.3 to construct a tower of configurations above $\mathcal{A}$.

Suppose first that we are in case (2) of Proposition 3.2.3. Then, there is a subset $B \subset A$ such that $|Q(\mathcal{A}|_B)| \gg \delta|B||\mathcal{P}|^2$ and, writing $\theta := 2/5 + \varepsilon$, we have

$$|B| \ll \exp\Big(-\frac{(\log Y)^\theta}{(\log\log Y)^{O(1)}}\Big)|A|.$$

On the other hand, by Corollary 3.4.5, we have

$$|B| \geqslant \exp\Big(-(\log Y)^{1-\frac{3\theta}{2}}(\log\log Y)^{O(1)}\Big)|A|.$$

Since $1 - \frac{3\theta}{2} < \theta - \varepsilon$, this is a contradiction for Y large enough in terms of ε_0.

Therefore, we must be in case (1) of Proposition 3.2.3: there is a tower $(\mathcal{A}_i)_{0\leqslant i\leqslant k}$ of height $k \gg (\log Y)^{10}$, such that $\mathcal{A}_0 = \mathcal{A}$ and, if $\mathcal{A}_k = (A_k, \widetilde{\alpha}_\bullet, H_k)$, we have $|Q(\mathcal{A}_k)| \gg \delta|A_k||\mathcal{P}|^2$.

By definition of a lift, we have $H_k \geqslant HP^k$, which is comfortably larger than $P^{3\log Y}$ for Y large enough. We can thus apply Lemma 3.4.6 to $\mathcal{A}_k$, obtaining a subset $A' \subset A_k$ with $|Q(\mathcal{A}_k|_{A'})| \gg \delta|A'||\mathcal{P}|^2$, an integer $q_0 \leqslant e^{O(f(Y))}$ and a real number $|t| \leqslant e^{O(f(Y))}\frac{Y^2}{H_kP}$, such that, for every $x \in A'$,

$$\widetilde{\alpha}_x = \frac{a_x}{q_0} + \frac{t}{x} + O\Big(\frac{e^{O(f(Y))}}{H_kP}\Big) \pmod 1$$

for some integer a_x coprime to q_0. Here, we recall that $f(Y) := (\log Y)^{1-\frac{3\theta}{2}}(\log\log Y)^3$.

Applying our expansion estimate, Corollary 3.4.5, to the configuration $\mathcal{A}_k|_{A'}$, we obtain the lower bound $|A'| \geqslant e^{-O(f(Y))}Y$. By pigeonholing on the values of $a_x \pmod{q_0}$, we can find a further subset $A'' \subset A'$ with $|A''| \geqslant e^{-O(f(Y))}|A|$ such that $a_x = a_0$ is constant for $x \in A''$.

Let $p_1^*, \dots, p_k^*$ be the lifting primes for the tower $(\mathcal{A}_i)_{0\leqslant i\leqslant k}$. By the definition of a lift, we have $H_k = Hp_1^*\cdots p_k^*$ and $\alpha_x = p_1^*\cdots p_k^*\widetilde{\alpha}_x \pmod 1$ for every $x \in A_k$. Thus, for every $x \in A''$, we have

$$\alpha_x = \frac{p_1^*\cdots p_k^*a_0}{q_0} + \frac{p_1^*\cdots p_k^*t}{x} + O\Big(\frac{e^{O(f(Y))}}{HP}\Big) \pmod 1.$$

Relabelling the parameters gives the desired conclusion. □

3.5 Local patterns and clusters

Throughout Sections 3.5 to 3.7, we fix an arbitrary configuration $\mathcal{A} = (A, \alpha_\bullet, H)$. All results and definitions in these sections implicitly refer to this configuration. Unless specified otherwise, the parameters Y, P, and H are only assumed to satisfy the default bounds in Definitions 3.1.1 and 3.1.2.

We begin by establishing some basic properties of the function $d_{p,q}(x, y)$ from Definition 3.1.6. These elementary facts appear in previous works under different notation; we include short proofs here to keep our exposition self-contained.

3.5.1 Triangle-like inequalities

Importantly, the function $d_{p,q}(x, y)$ does *not* satisfy a triangle-like inequality of the form

$$d_{p,r}(x, z) \leqslant d_{p,q}(x, y) + d_{q,r}(y, z),$$

due to the second argument of the maximum in (3.1). Yet, a useful substitute is available: we prove it in Lemma 3.5.2 below.

Lemma 3.5.1. *Let $m, n \geqslant 1$ be coprime integers, $\alpha \in \mathbb{R}$ and $0 < \varepsilon \leqslant 1$. Suppose that $\|m\alpha\| < \frac{\varepsilon}{2n}$ and $\|n\alpha\| < \frac{1}{2m}$. Then $\|\alpha\| < \frac{\varepsilon}{2mn}$.*

Proof. The assumption $\|n\alpha\| < \frac{1}{2m}$ means that $\alpha = \frac{k}{n} + \frac{\theta}{2mn} \pmod 1$ for some $|\theta| < 1$ and $k \in \mathbb{Z}$. Thus

$$m\alpha = \frac{km}{n} + \frac{\theta}{2n} \pmod 1,$$

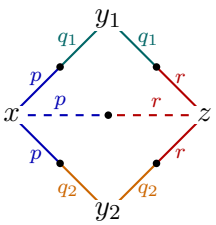
but since $\|m\alpha\| < \frac{\varepsilon}{2n} \leqslant \frac{1}{2n}$ and $|\theta| < 1$, it must be that $n \mid km$, i.e. $n \mid k$ as $(m, n) = 1$. Therefore, $\alpha = \frac{\theta}{2mn} \pmod 1$. Recalling that $\|m\alpha\| < \frac{\varepsilon}{2n}$, we see that $|\theta| < \varepsilon$, so that $\|\alpha\| < \frac{\varepsilon}{2mn}$. □

Lemma 3.5.2 ([71, Lemma 2.3])**.** *Let $0 < C \leqslant \frac{H}{30P^2}$. Let $x, y_1, y_2, z \in A$ and $p, q_1, q_2, r \in \mathcal{P}$. Suppose that*

$$d_{p,q_i}(x, y_i), d_{q_i,r}(y_i, z) \leqslant C$$

for $i = 1, 2$. If $q_1 \neq q_2$, then

$$d_{p,r}(x, z) \leqslant 4C.$$



Proof. From the assumptions $\|p\alpha_{y_i} - q_i\alpha_x\| \leqslant C/H$ and $\|q_i\alpha_z - r\alpha_{y_i}\| \leqslant C/H$, we have

$$\big\|q_i(p\alpha_z - r\alpha_x)\big\| \leqslant \big\|r(p\alpha_{y_i} - q_i\alpha_x)\big\| + \big\|p(q_i\alpha_z - r\alpha_{y_i})\big\| \leqslant \frac{4CP}{H}$$

for each $i = 1, 2$. Since $q_1 \neq q_2$, we can apply Lemma 3.5.1 to deduce that $\|p\alpha_z - r\alpha_x\| \leqslant 4C/H$. It remains to prove that $|px - rz| \leqslant 4CP$, but the stronger bound $|px - rz| \leqslant 2CP$ follows directly from the standard triangle inequality. □

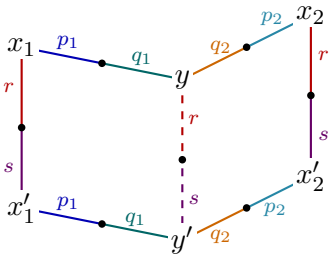
We also require a version of Lemma 3.5.2 for 'parallelogram patterns'.

Lemma 3.5.3 ([73, Lemma 3.5])**.** *Let $0 < C \leqslant \frac{H}{30P^2}$. Let $x_1, x_1', x_2, x_2', y, y' \in A$ and $p_1, p_2, q_1, q_2, r, s \in \mathcal{P}$. Suppose that*

$$d_{p_i,q_i}(x_i, y), d_{p_i,q_i}(x_i', y'), d_{r,s}(x_i, x_i') \leqslant C$$

for $i = 1, 2$. If $p_1 \neq p_2$, then

$$d_{r,s}(y, y') \leqslant 6C.$$



Proof. By the triangle inequality, for $i = 1, 2$, we have

$$\left\|p_i(s\alpha_y - r\alpha_{y'})\right\| \leqslant \left\|s(p_i\alpha_y - q_i\alpha_{x_i})\right\| + \left\|r(p_i\alpha_{y'} - q_i\alpha_{x_i'})\right\| + \left\|q_i(s\alpha_{x_i} - r\alpha_{x_i'})\right\| \leqslant \frac{6CP}{H}.$$

Since $p_1 \neq p_2$, we conclude that $\|s\alpha_y - r\alpha_{y'}\| \leqslant 6C/H$ by Lemma 3.5.1. On the other hand,

$$|q_1(ry - sy')| \leqslant |r(p_1x_1 - q_1y)| + |s(p_1x_1' - q_1y')| + |p_1(rx_1 - sx_1')| \leqslant 6CP^2,$$

so that $|ry - sy'| \leqslant 6CP$. Therefore, $d_{r,s}(y, y') \leqslant 6C$. □

3.5.2 Clusters of local relations

To prove the local structure theorem, we will need to show the existence of certain patterns of local relations in $\mathcal{A}$, which we call *clusters*.

Definition 3.5.4 (Cluster)**.** Let $C > 0$. A *cluster of diameter at most C* is a subset $S \subset A \times \mathcal{P}$ such that $d_{r,s}(x, x') \leqslant C$ for all $(x, r), (x', s) \in S$.

A cluster S of diameter $\leqslant C$ has size $|S| \leqslant (2C + 1)|\mathcal{P}|$. This is an immediate consequence of the following basic observation.

Lemma 3.5.5. *Let $C > 0$. For any $x \in A$ and $p, q \in \mathcal{P}$, there are at most $2C + 1$ elements $y \in A$ such that $d_{p,q}(x, y) \leqslant C$.*

Proof. Fix x, p, q. If $d_{p,q}(x, y) \leqslant C$ for some $y \in A$, then $|px/q - y| \leqslant C$. Since A is 1-separated, this interval contains at most $2C + 1$ points of A. □

We now give examples of clusters.

Example 3.5.6. Let $z \in [PY, 4PY]$ and $\beta \in \mathbb{R}/\mathbb{Z}$. Let S be the set of all pairs $(x,p) \in A \times \mathcal{P}$ such that $|px - z| \leqslant CP$ and $\|p\beta - \alpha_x\| \leqslant C(PH)^{-1}$. Then S is a cluster of diameter at most $4C$. Indeed, for any two pairs $(x,r),(x',s) \in S$, we have

$$|rx - sx'| \leqslant |rx - z| + |sx' - z| \leqslant 2CP$$

and

$$\|r\alpha_{x'} - s\alpha_x\| \leqslant \|r(\alpha_{x'} - s\beta)\| + \|s(r\beta - \alpha_x)\| \leqslant \frac{4C}{H},$$

which implies that $d_{r,s}(x,x') \leqslant 4C$.

The next lemma shows that any cluster involving at least two distinct primes is of the form given in Example 3.5.6, up to constants. Moreover, the approximation $p\beta \approx \alpha_x \pmod 1$ can be made exact for one pair (x,p) of our choice in the cluster.

Lemma 3.5.7 ([71, Lemmas 2.1 and 2.2])**.** *Let $0 < C \leqslant \frac{H}{30P}$. Let $S \subset A \times \mathcal{P}$ be a cluster of diameter $\leqslant C$ and let $(x,p) \in S$. Suppose that S is not entirely contained in $A \times \{p\}$. Then, there exists $\beta \in \mathbb{R}/\mathbb{Z}$ such that*

(i) $\|p\beta - \alpha_x\| = 0$, and

(ii) for all $(x',r) \in S$, we have $\|r\beta - \alpha_{x'}\| \leqslant C(Hp)^{-1}$.

Proof. By assumption, there exists $(y,q) \in S$ with $q \neq p$. By definition of a cluster, we have $\|q\alpha_x - p\alpha_y\| \leqslant C/H$. Since the linear combinations $qm - pn$ with $m,n \in \mathbb{Z}$ attain all integer values, we can find representatives $\widetilde{\alpha}_x, \widetilde{\alpha}_y \in \mathbb{R}$ of $\alpha_x, \alpha_y \in \mathbb{R}/\mathbb{Z}$ such that $|q\widetilde{\alpha}_x - p\widetilde{\alpha}_y| \leqslant C/H$.

We set $\beta := \widetilde{\alpha}_x/p \pmod 1$; this choice clearly verifies (i). Moreover,

$$\|q\beta - \widetilde{\alpha}_y\| = \left\| \frac{q\widetilde{\alpha}_x - p\widetilde{\alpha}_y}{p} \right\| \leqslant \left| \frac{q\widetilde{\alpha}_x - p\widetilde{\alpha}_y}{p} \right| \leqslant \frac{C}{Hp}. \tag{3.32}$$

Now, let $(x',r) \in S$ be arbitrary. On the one hand, applying the cluster property to (x',r) and (x,p), we get

$$\|p(r\beta - \alpha_{x'})\| = \|r\alpha_x - p\alpha_{x'}\| \leqslant \frac{C}{H}.$$

On the other hand, using (3.32) and the cluster property applied to (x',r) and (y,q), we have

$$\|q(r\beta - \alpha_{x'})\| \leqslant \|r(q\beta - \alpha_y)\| + \|r\alpha_y - q\alpha_{x'}\| \leqslant \frac{3C}{H}.$$

By Lemma 3.5.1, we conclude that $\|r\beta - \alpha_{x'}\| \leqslant C(Hp)^{-1}$, as desired. □

In Lemma 3.5.9, we show that two clusters can be combined into a larger cluster provided their intersection involves at least two distinct primes.

Definition 3.5.8. Two sets $S_1, S_2 \subset A \times \mathcal{P}$ are called *almost disjoint* if

$$S_1 \cap S_2 \subset A \times \{p\}$$

for some $p \in \mathcal{P}$. Otherwise, we say that S_1, S_2 *strongly intersect*.

Lemma 3.5.9. *Let $k \in \mathbb{N}$ and $0 < C \leqslant \frac{4^{-k}H}{30P^2}$. Let $S_1, \dots, S_k$ be clusters of diameter $\leqslant C$, where S_i and S_{i+1} strongly intersect for all $1 \leqslant i < k$. Then, $\bigcup_{i=1}^k S_i$ is a cluster of diameter $\leqslant 4^{k-1}C$.*[5]

Proof. It suffices to prove the case $k = 2$, as the general case follows by induction.

Consider arbitrary elements $(x, r) \in S_1$ and $(x', s) \in S_2$. Since S_1 and S_2 strongly intersect, $S_1 \cap S_2$ contains two elements (y_1, q_1) and (y_2, q_2) with $q_1 \neq q_2$. By Definition 3.5.4, we have $d_{r,q_i}(x, y_i) \leqslant C$ and $d_{q_i,s}(y_i, x') \leqslant C$ for $i = 1, 2$. By Lemma 3.5.2, we conclude that $d_{r,s}(x, x') \leqslant 4C$. □

3.6 Covering lemmas for clusters

In this section, we prove several lemmas on global arrangements of clusters. We begin with a technical lemma that bounds the size of the overlaps for a family of almost disjoint clusters.

Lemma 3.6.1. *Let $C \geqslant 1$ and $\eta > 0$. Let $(F_i)_{i \in I}$ be a collection of pairwise almost disjoint clusters of diameter $\leqslant C$ and size $\geqslant \eta|\mathcal{P}|$. For each i, let $\overline{F_i}$ be a cluster of diameter $\leqslant 20C$ containing F_i. If $J \subset I \times I$ is the set of all pairs (i, j) such that $\overline{F_i}$ and $\overline{F_j}$ are almost disjoint, then*

$$\sum_{(i,j) \in J} \left|\overline{F_i} \cap \overline{F_j}\right| \ll C^9 \eta^{-5} |A|.$$

Proof. For every cluster S of diameter $\leqslant C$, the set $\{px : (x, p) \in S\}$ is contained in an interval of length $\leqslant CP$. For $z \in [PY, 4PY]$, define

$$V_z := \left\{(x, p) \in A \times \mathcal{P} \,:\, |px - z| \leqslant 100CP\right\}$$

and consider the multiplicity function

$$m(z) := \sum_{i \in I} \mathbf{1}_{F_i \subset V_z}.$$

We first prove a pointwise upper bound for $m(z)$. By Lemma 3.5.5, for every $i \in I$ and $p \in \mathcal{P}$,

$$|\{x \in A \,:\, (x, p) \in F_i\}| \leqslant \left|\{x \in A \,:\, (x, p) \in \overline{F_i}\}\right| \ll C. \tag{3.33}$$

[5]The precise constant is not important as we will only use Lemma 3.5.9 with bounded values of k.

In particular, each F_i admits a subset F_i' of size $\gg |F_i|/C$ such that no two elements of F_i' share the same prime. Let $\mathcal{T}_i$ be the collection of all two-element sets $\{u, v\}$ with $u, v \in F_i'$, $u \neq v$. For all $i \neq j$, since F_i and F_j are almost disjoint, we have $|F_i' \cap F_j'| \leqslant 1$, and thus $\mathcal{T}_i \cap \mathcal{T}_j = \emptyset$.

Note that $|V_z| \ll C|\mathcal{P}|$ as A is 1-separated. Since the sets $(\mathcal{T}_i)_{i\in I}$ are pairwise disjoint, we have

$$\sum_{\substack{i\in I\\ F_i\subset V_z}} |\mathcal{T}_i| \leqslant \binom{|V_z|}{2} \ll C^2|\mathcal{P}|^2.$$

Recalling that $|\mathcal{T}_i| \gg |F_i'|^2 \gg C^{-2}|F_i|^2$ and $|F_i| \geqslant \eta|\mathcal{P}|$, we get

$$m(z) \ll C^4\eta^{-2}. \tag{3.34}$$

Next, we bound the measure of the support of $m(z)$. For each $i \in I$, fix some $z_i \in \{px : (x, p) \in F_i\}$ and let I' be a maximal subset of I such that the points $(z_i)_{i\in I'}$ are $(100CP)$-separated. Then, the clusters $(F_i)_{i\in I'}$ are pairwise disjoint, so that

$$|I'|\eta|\mathcal{P}| \leqslant \sum_{i\in I'} |F_i| \leqslant |A||\mathcal{P}|.$$

This yields $|I'| \leqslant \eta^{-1}|A|$. The support of $m(z)$ is clearly contained in the union of the intervals $[z_i - 200CP, z_i + 200CP]$ for $i \in I'$. Hence, the support of $m(z)$ has measure at most

$$\leqslant 400CP \cdot |I'| \ll C\eta^{-1}|A|P. \tag{3.35}$$

For $(i, j) \in J$, since the clusters $\overline{F_i}$ and $\overline{F_j}$ are almost disjoint, we must have $\left|\overline{F_i} \cap \overline{F_j}\right| \ll C$ by (3.33). In addition, whenever $\overline{F_i} \cap \overline{F_j} \neq \emptyset$, the set

$$\{px : (x, p) \in \overline{F_i} \cup \overline{F_j}\}$$

is contained in an interval of length $\leqslant 40CP$. Therefore, for all $(i, j) \in J$,

$$\left|\overline{F_i} \cap \overline{F_j}\right| \ll \frac{1}{P}\int_{PY}^{4PY} \mathbf{1}_{\overline{F_i}\cup\overline{F_j}\subset V_z}\, dz \leqslant \frac{1}{P}\int_{PY}^{4PY} \mathbf{1}_{F_i\subset V_z}\mathbf{1}_{F_j\subset V_z}\, dz.$$

Summing over all $(i, j) \in J$, we get

$$\sum_{(i,j)\in J} \left|\overline{F_i} \cap \overline{F_j}\right| \ll \frac{1}{P}\int_{PY}^{4PY} m(z)^2\, dz.$$

The claimed estimate now follows from (3.34) and (3.35). □

With the previous technical estimate in hand, we can prove a covering result reminiscent of the Vitali covering lemma in metric spaces: Lemma 3.6.2 allows us to replace an arbitrary collection of clusters with a more manageable collection while only losing a constant factor in the diameter.

Lemma 3.6.2 (Efficient covering)**.** *Let $1 \leqslant C \leqslant |\mathcal{P}|^{1/4}$ and $\eta \geqslant |\mathcal{P}|^{-1/4}$. Let $\mathcal{U} \subset A \times \mathcal{P}$ be the union of all clusters of diameter $\leqslant C$ and size $\geqslant \eta|\mathcal{P}|$.*

There exists a collection $(K_i)_{i\in I}$ of clusters of diameter $\leqslant 16C$ with the following properties.

(i) For every cluster S of diameter $\leqslant C$ and size $\geqslant \eta|\mathcal{P}|$, and every $(x,p) \in S$, there is a cluster K_i such that $(x,p) \in K_i$ and $|K_i| \geqslant |S|$.

(ii) $\sum_{i\in I} |K_i| \ll |A||\mathcal{P}| + C^9\eta^{-5}|A|$.

Proof. If $\mathcal{U} = \emptyset$, the result is trivial. Otherwise, we proceed as follows.

We start by constructing an auxiliary collection of clusters $(F_i)_{i\in I}$ using a greedy algorithm. Let F_1 be a cluster of diameter $\leqslant C$ of maximal size. If $F_1, \dots, F_{i-1}$ have been chosen, let F_i be a cluster of maximal cardinality among all clusters of diameter $\leqslant C$ and size $\geqslant \eta|\mathcal{P}|$ that are almost disjoint from each of $F_1, \dots, F_{i-1}$. This algorithm clearly terminates with a maximal collection $(F_i)_{i\in I}$ of pairwise almost disjoint clusters of diameter $\leqslant C$ and size $\geqslant \eta|\mathcal{P}|$, where $I \subset \mathbb{N}$.

These clusters may not cover all of $\mathcal{U}$, so we shall expand them. For each $(x,p) \in \mathcal{U}$, let $S_{(x,p)}$ be a cluster of diameter $\leqslant C$ containing (x,p) of maximal size, and define $i(x,p)$ to be the smallest i such that $S_{(x,p)}$ and F_i strongly intersect. Note that $|S_{(x,p)}| \leqslant |F_{i(x,p)}|$ by the maximality assumption in the greedy procedure.

For every $i \in I$, define

$$K_i := F_i \cup \big\{(x,p) \in \mathcal{U} \,:\, i(x,p) = i\big\}.$$

We claim that K_i is a cluster of diameter $\leqslant 16C$. To see this, consider arbitrary elements $(x,r), (x',s) \in K_i$. By definition of K_i, there are clusters S, S' of diameter $\leqslant C$ (possibly equal to F_i) such that $(x,r) \in S$, $(x',s) \in S'$, and both S and S' strongly intersect F_i. By Lemma 3.5.9, $S \cup F_i \cup S'$ is a cluster of diameter $\leqslant 16C$, so that $d_{r,s}(x,x') \leqslant 16C$.

Property (i) now follows: if $(x,p) \in \mathcal{U}$ lies in a cluster S of diameter $\leqslant C$, then $(x,p) \in K_{i(x,p)}$ and

$$|S| \leqslant |S_{(x,p)}| \leqslant |F_{i(x,p)}| \leqslant |K_{i(x,p)}|.$$

We turn to property (ii). By definition of K_i and truncated inclusion-exclusion (i.e. the Bonferroni inequalities), we have

$$\sum_{i\in I} |K_i| \leqslant \sum_{i\in I} |F_i| + |\mathcal{U}| \ll \Big|\bigcup_{i\in I} F_i\Big| + \sum_{\substack{i,j\in I\\ i\neq j}} |F_i \cap F_j| + |\mathcal{U}|.$$

The middle term of the right-hand side is $\ll C^9\eta^{-5}|A|$ by Lemma 3.6.1 (with $\overline{F_i} = F_i$). The other terms are $\ll |A||\mathcal{P}|$, so we obtain (ii). □

Finally, we establish a result that can be viewed as a strengthening of Lemma 3.5.9. While both results show that a union of clusters is again a cluster of slightly larger diameter, the conditions under which this holds are now significantly relaxed. Instead of requiring the clusters to strongly intersect, we consider families of clusters containing a given pair (x, p), even if those clusters share no other points. The conclusion is not guaranteed to hold for every pair (x, p), but Lemma 3.6.4 shows that it does hold for almost all pairs.

Definition 3.6.3. Let $C \geqslant 1$ and $\eta > 0$. A pair $(x, p) \in A \times \mathcal{P}$ is called (C, η)-*regular* if the union of all clusters of diameter $\leqslant C$ and size $\geqslant \eta|\mathcal{P}|$ containing (x, p) is either empty, or itself a cluster of diameter $\leqslant 64C$.

Lemma 3.6.4 (Regularity)**.** *Let $1 \leqslant C \leqslant |\mathcal{P}|^{1/4}$ and $\eta \geqslant |\mathcal{P}|^{-1/4}$. Then, all but $\ll C^9\eta^{-5}|A|$ pairs $(x, p) \in A \times \mathcal{P}$ are (C, η)-regular.*

Proof. Let $\mathcal{U}$ be the union of all clusters of diameter $\leqslant C$ and size $\geqslant \eta|\mathcal{P}|$. Let $(F_i)_{i\in I}$ be a maximal collection of pairwise almost disjoint clusters of diameter $\leqslant C$ and size $\geqslant \eta|\mathcal{P}|$.

For each $i \in I$, define $\overline{F_i}$ to be the union of F_i and all clusters of diameter $\leqslant C$ that strongly intersect F_i. By Lemma 3.5.9, each $\overline{F_i}$ is a cluster of diameter $\leqslant 16C$ (as in the proof of Lemma 3.6.2).

Consider a pair $(x, p) \in \mathcal{U}$ that is not (C, η)-regular. Then, there are clusters S_1, S_2 of diameter $\leqslant C$ and size $\geqslant \eta|\mathcal{P}|$, both containing (x, p), such that $S_1 \cup S_2$ is not a cluster of diameter $\leqslant 64C$. By definition of (F_i), there exist indices $j_1, j_2 \in I$ such that S_i strongly intersects F_{j_i} for each $i = 1, 2$. Hence, $S_i \subset \overline{F_{j_i}}$. Observe that $\overline{F_{j_1}}$ and $\overline{F_{j_2}}$ must be almost disjoint, as otherwise their union (and hence also $S_1 \cup S_2$) would be a cluster of diameter $\leqslant 64C$ by Lemma 3.5.9. We have thus shown that for every pair (x, p) that is not (C, η)-regular, there are j_1, j_2 such that $(x, p) \in \overline{F_{j_1}} \cap \overline{F_{j_2}}$, where $\overline{F_{j_1}}$ and $\overline{F_{j_2}}$ are almost disjoint. The conclusion now follows from Lemma 3.6.1. □

3.7 Proof of the local structure theorem

The goal of this section is to prove Theorem 3.1.8, which describes the structure of configurations having many local relations relative to their size, and constructs lifts preserving a large proportion of these relations.

3.7.1 Initial clusters

In the next lemma, we construct many large disjoint clusters, which will later help us define the first set A_0 of the decomposition of A given in Theorem 3.1.8. This

lemma is a variant of [73, Corollary 4.10] that only assumes the existence of many local relations *relative to* $|A|$, and thus also applies to sparse sets A.

Lemma 3.7.1 (Large clusters)**.** *Suppose that* $|Q(\mathcal{A})| \geqslant \delta|A||\mathcal{P}|^2$*, where* $\delta \geqslant C_1|\mathcal{P}|^{-1/2}$ *for some large absolute constant* C_1*. Then, there exists a collection of* $\gg \delta|A|$ *disjoint clusters, each of diameter* $\leqslant 1/2$ *and size* $\gg \delta^4|\mathcal{P}|$*.*

Proof. For $z \in [PY, 4PY]$, let

$$V_z := \big\{(x,p) \in A \times \mathcal{P} \,:\, |px - z| \leqslant \tfrac{1}{4}P\big\}$$

and

$$Q_z := \big\{((x,p),(y,q)) \in V_z^2 \,:\, (x,y,p,q) \in Q(\mathcal{A})\big\}.$$

We have the simple bound $|Q_z| \leqslant |V_z|^2 \leqslant |\mathcal{P}|^2$ as A is 1-separated.

Suppose that the reverse inequality $|Q_z| \geqslant \varepsilon|\mathcal{P}|^2$ holds for some $\varepsilon \geqslant 10/|\mathcal{P}|$. We claim that V_z contains a cluster of diameter $\leqslant 1/2$ and size $\gg \varepsilon^2|\mathcal{P}|$. To see this, consider the graph G_z with vertex set V_z and edge set $E_z := Q_z \setminus \Delta$ where $\Delta = \{((x,p),(x,p)) : (x,p) \in V_z\}$ is the diagonal. Since $|Q_z| \geqslant \varepsilon|\mathcal{P}|^2$ and $|\Delta| = |V_z| \leqslant |\mathcal{P}|$, we have $|E_z| \gg \varepsilon|\mathcal{P}|^2$ (as $\varepsilon \geqslant 10/|\mathcal{P}|$). By convexity of the function $n \mapsto \binom{n}{2}$, we get

$$\frac{1}{|V_z|}\sum_{\substack{v_1,v_2 \in V_z \\ v_1 \neq v_2}} \sum_{u \in V_z} \mathbf{1}_{(u,v_1),(u,v_2) \in E_z} = \mathbb{E}_{u \in V_z}\binom{\deg(u)}{2} \geqslant \binom{2|E_z|/|V_z|}{2} \gg \varepsilon^2|\mathcal{P}|^2.$$

Hence, there exist two distinct vertices $v_1, v_2 \in V_z$ having $\gg \varepsilon^2|\mathcal{P}|$ common neighbours in G_z. Write $v_1 = (y_1, q_1)$, $v_2 = (y_2, q_2)$. We know that $q_1 \neq q_2$ as $v_1 \neq v_2$ and A is 1-separated. We claim that the set N_z of common neighbours of v_1 and v_2 is a cluster of diameter $\leqslant 1/2$. Indeed, if $(x_1,p_1),(x_2,p_2) \in N_z$, then $(x_i, y_j, p_i, q_j) \in Q(\mathcal{A})$ for all $i,j \in \{1,2\}$, and hence $d_{p_1,p_2}(x_1,x_2) \leqslant 4/10$ by Lemma 3.5.2.

To finish the proof, it therefore suffices to find a P-separated set $Z \subset [PY, 4PY]$ of size $|Z| \gg \delta|A|$ such that $|Q_z| \gg \delta^2|\mathcal{P}|^2$ for every $z \in Z$; the separation assumption guarantees that the sets $(V_z)_{z \in Z}$ will be pairwise disjoint.

To achieve this, it is enough to show that

$$\mathbb{P}\big(|Q_{\mathbf{z}}| \geqslant c\delta^2|\mathcal{P}|^2\big) \gg \frac{\delta|A|}{Y} \tag{3.36}$$

where $\mathbf{z}$ is a random variable uniformly distributed in $[PY, 4PY]$, and $c > 0$ is an absolute constant. On the one hand, since $|Q(\mathcal{A})| \geqslant \delta|A||\mathcal{P}|^2$ and $|px - qy| \leqslant \frac{1}{10}P$ whenever $(x,y,p,q) \in Q(\mathcal{A})$, we have

$$\mathbb{E}[|Q_{\mathbf{z}}|] \geqslant \sum_{(x,y,p,q) \in Q(\mathcal{A})} \mathbb{P}\big(|px - \mathbf{z}| \leqslant \tfrac{1}{10}P\big) \gg \frac{\delta|A||\mathcal{P}|^2}{Y}.$$

On the other hand, for any $\varepsilon > 0$,

$$\mathbb{E}[|Q_{\mathbf{z}}|] \leqslant |\mathcal{P}|^2\,\mathbb{P}\big(|Q_{\mathbf{z}}| \geqslant \varepsilon|\mathcal{P}|^2\big) + \mathbb{E}\big[|Q_{\mathbf{z}}|\mathbf{1}_{|Q_{\mathbf{z}}|\leqslant\varepsilon|\mathcal{P}|^2}\big],$$

and

$$\mathbb{E}\big[|Q_{\mathbf{z}}|\mathbf{1}_{|Q_{\mathbf{z}}|\leqslant\varepsilon|\mathcal{P}|^2}\big] \leqslant \varepsilon^{1/2}|\mathcal{P}|\,\mathbb{E}\Big[|Q_{\mathbf{z}}|^{1/2}\Big] \leqslant \varepsilon^{1/2}|\mathcal{P}|\,\mathbb{E}[|V_{\mathbf{z}}|] \ll \varepsilon^{1/2}\frac{|A||\mathcal{P}|^2}{Y}.$$

Rearranging and choosing $\varepsilon = c\delta^2$ for some small constant $c > 0$ yields (3.36). □

3.7.2 Rigidity

Exploiting the properties of $d_{p,q}(x,y)$, we show that certain patterns of local relations are rare.

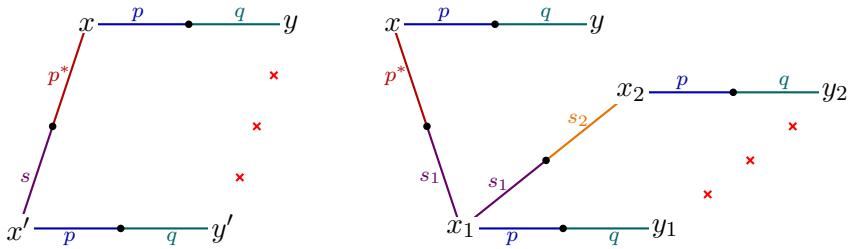


Figure 3.1: Patterns considered in Definition 3.7.2 (where x, y, p, q and p^* are fixed).

Definition 3.7.2. Let $(x,y,p,q) \in Q(\mathcal{A})$, $p^* \in \mathcal{P}$ and $C > 0$. Let $N_{\square^\times}(x,y,p,q,p^*;C)$ be the number of triples $(x',y',s) \in A^2 \times \mathcal{P}$ such that

$$d_{p,q}(x',y'), d_{p^*,s}(x,x') \leqslant C \quad \text{and} \quad d_{p^*,s}(y,y') > 6C.$$

Also, let $N_{\square^\times_\diagdown}(x,p,q,p^*;C)$ be the number of tuples $(x_1,x_2,y_1,y_2,s_1,s_2) \in A^4 \times \mathcal{P}^2$ such that

$$d_{p,q}(x_1,y_1), d_{p,q}(x_2,y_2), d_{p^*,s_1}(x,x_1), d_{s_1,s_2}(x_1,x_2) \leqslant C \quad \text{and} \quad d_{s_1,s_2}(y_1,y_2) > 6C.$$

We have the trivial bound $N_{\square^\times}(x,y,p,q,p^*;C) \ll_C |\mathcal{P}|$, as fixing s leaves only $O_C(1)$ possibilities for x' and y' (by Lemma 3.5.5). Similarly, we have the trivial bound $N_{\square^\times_\diagdown}(x,p,q,p^*;C) \ll_C |\mathcal{P}|^2$. The next lemma shows that, *on average* over $(x,y,p,q) \in Q(\mathcal{A})$, we can gain a factor of $|\mathcal{P}|$ in each case.

Lemma 3.7.3. *Let $1 \leqslant C \leqslant c_1|\mathcal{P}|^{c_1}$ for some small absolute constant $c_1 > 0$.*

Fix $p^* \in \mathcal{P}$. Then

$$\sum_{(x,y,p,q)\in Q(\mathcal{A})} N_{\square^\times}(x,y,p,q,p^*;C) \ll C^3|A||\mathcal{P}|^2$$

and

$$\sum_{(x,y,p,q)\in Q(\mathcal{A})} N_{\square^\times_\diagdown}(x,p,q,p^*;C) \ll C^5|A||\mathcal{P}|^3.$$

Proof. The first part is a relative version of [73, Lemma 4.8]. To shorten notation, write

$$\mathbf{1}_{\square^*}(x,x',y,y';p,q,r,s) := \mathbf{1}_{\{d_{p,q}(x,y),d_{p,q}(x',y'),d_{r,s}(x,x')\leqslant C\}}\mathbf{1}_{d_{r,s}(y,y')>6C},$$

i.e. the indicator function of the pattern represented in Fig. 3.1 (left side). Swapping the order of summation, we have

$$\sum_{(x,y,p,q)\in Q(\mathcal{A})} N_{\square^*}(x,y,p,q,p^*;C) \leqslant \sum_{(y,y',s)\in A^2\times\mathcal{P}} \sum_{(x,x',p,q)\in A^2\times\mathcal{P}^2} \mathbf{1}_{\square^*}(x,x',y,y';p,q,p^*,s).$$

We may restrict the outer sum to the triples (y,y',s) satisfying $|p^*y - sy'| \leqslant 6CP$, as otherwise $\mathbf{1}_{\square^*}(x,x',y,y',p,q,p^*,s) = 0$ (regardless of x,x',p,q). Thus, there are $\ll C|A||\mathcal{P}|$ choices for y,y' and s. Now, the key observation is that, by Lemma 3.5.3, there is at most one prime $p = p(y,y',s) \in \mathcal{P}$ giving a non-zero contribution to the inner sum. There are $|\mathcal{P}|$ choices for q. Once y,y',s,p and q are fixed, there are $\ll C$ choices for each of x and x', by Lemma 3.5.5. Hence,

$$\sum_{(x,y,p,q)\in Q(\mathcal{A})} N_{\square^*}(x,y,p,q,p^*;C) \ll C|A||\mathcal{P}|\cdot|\mathcal{P}|\cdot C^2 = C^3|A||\mathcal{P}|^2$$

as claimed.

Similarly, the $N_{\square\!\!\!\!\diagup^*}$ sum is bounded by

$$\sum_{(y_1,y_2,s_1,s_2)\in A^2\times\mathcal{P}^2} \sum_{(x_1,x_2,p,q)\in A^2\times\mathcal{P}^2} \mathbf{1}_{\square^*}(x_1,x_2,y_1,y_2;p,q,s_1,s_2) \sum_{x,y\in A} \mathbf{1}_{\{d_{p^*,s_1}(x,x_1),d_{p,q}(x,y)\leqslant C\}}.$$

The inner sum over x,y is always $\ll C^2$ by Lemma 3.5.5, so that the bound becomes

$$\ll C^2 \sum_{(y_1,y_2,s_1,s_2)\in A^2\times\mathcal{P}^2} \sum_{(x_1,x_2,p,q)\in A^2\times\mathcal{P}^2} \mathbf{1}_{\square^*}(x_1,x_2,y_1,y_2;p,q,s_1,s_2).$$

Once s_1 is fixed, we are left with the same expression as in the previous case, after relabelling. Since there are $|\mathcal{P}|$ choices for $s_1 \in \mathcal{P}$, we obtain a final bound of $\ll C^2|\mathcal{P}|\cdot C^3|A||\mathcal{P}|^2$, as required. □

3.7.3 Replication of clusters

The following lemma provides a mechanism to generate new clusters from existing ones, at the cost of multiplying the diameter by a fixed constant. This operation is possible under certain technical conditions, including bounds on the functions $N_{\square^*}$ and $N_{\square\!\!\!\!\diagup^*}$ introduced in Definition 3.7.2.

Lemma 3.7.4. *Let $1 \leqslant C \leqslant c_1|\mathcal{P}|^{c_1}$ for some small absolute constant $c_1 > 0$. Let $(x,y,p,q) \in Q(\mathcal{A})$ and $p^* \in \mathcal{P}$. Let S be a cluster of diameter $\leqslant C$ and size $\geqslant |\mathcal{P}|^{1/2}$ containing (x,p^*).*

Suppose that, for all $(x', r) \in S$, there exists $y' \in A$ such that $d_{p,q}(x', y') \leqslant 1/10$. Furthermore, assume that

$$N_{\square}(x, y, p, q, p^*; C) \leqslant \tfrac{1}{100}|S| \quad \text{and} \quad N_{\square\!\!\!/}(x, p, q, p^*; C) \leqslant \tfrac{1}{100}|S|^2. \tag{3.37}$$

Then, there exists a cluster S' of diameter $\leqslant 24C$ and size $\geqslant |S|/2$ containing (y, p^*). Moreover, for any $(y', r) \in S'$, there is $(x', r) \in S$ such that $d_{p,q}(x', y') \leqslant C$.

Proof. Let V be the set of all triples $(x', y', r) \in A^2 \times \mathcal{P}$ such that $(x', r) \in S$ and $d_{p,q}(x', y') \leqslant 1/10$. Note that $|V| = |S|$ as, by assumption, every $(x', r) \in S$ corresponds to some unique $(x', y', r) \in V$.

Let G be the graph on V where two distinct vertices $(x_1, y_1, r_1), (x_2, y_2, r_2) \in V$ are joined by an edge whenever $d_{r_1,r_2}(y_1, y_2) \leqslant 6C$.

We restrict to a highly connected subset of V containing (x, y, p^*) (which lies indeed in V) as follows. By the first bound in (3.37), there are $\leqslant \frac{1}{100}|V|$ vertices $(x', y', s) \in V$ which are not neighbours of (x, y, p^*) in G. In addition, by the second bound in (3.37), there are $\leqslant \frac{1}{100}|V|^2$ pairs of vertices of G that are not connected by an edge. Defining V' to be the set of all vertices $(x', y', r) \in V$ which have $\geqslant \frac{3}{4}|V|$ neighbours in G, we see that $(x, y, p^*) \in V'$ and $|V'| \geqslant |V|/2$.

We can now construct the required cluster. Let

$$S' := \big\{(y', r) \in A \times \mathcal{P} \,:\, \exists x' \in A, (x', y', r) \in V'\big\}.$$

Observe that $(y, p^*) \in S'$ and $|S'| = |V'| \geqslant |S|/2$, as required. Moreover, for every $(y', r) \in S'$, there is $x' \in A$ such that $(x', r) \in S$ and $d_{p,q}(x', y') \leqslant 1/10$. Thus, it only remains to prove that S' is a cluster of diameter $\leqslant 24C$.

Let $(y_1, r_1), (y_2, r_2) \in S'$ and let $(x_1, y_1, r_1), (x_2, y_2, r_2) \in V'$ be the corresponding vertices of G. These two vertices each have $\geqslant \frac{3}{4}|V|$ neighbours in G. Hence, they have $\geqslant \frac{1}{2}|V|$ common neighbours. We can find two of these common neighbours, say (x_1', y_1', r_1') and (x_2', y_2', r_2'), such that $r_1' \neq r_2'$ (otherwise S would contain $\geqslant \frac{1}{2}|\mathcal{P}|^{1/2}$ pairs with the same prime, contradicting Lemma 3.5.5). By definition of G we have

$$d_{r_i,r_j'}(y_i, y_j) \leqslant 6C$$

for all $i, j \in \{1, 2\}$. By Lemma 3.5.2, this implies $d_{r_1,r_2}(y_1, y_2) \leqslant 24C$, which concludes the proof. □

3.7.4 Proof of the local structure theorem

We are now ready to prove the main result of this section, which we recall for convenience.

Theorem 3.1.8 (Local structure theorem)**.** *There exists an absolute constant $C_0 \geqslant 1$ such that the following holds. Let $Y, P, \mathcal{P}$ be as in Definition 3.1.1. Let $\mathcal{A} = (A, \alpha_\bullet, H)$ be a configuration such that*

$$|Q(\mathcal{A})| \geqslant \delta |A||\mathcal{P}|^2$$

for some $0 < \delta < \frac{1}{2}$. Let $L \in \mathbb{N}$ and $0 < \varepsilon < \frac{1}{2}$ be parameters such that $\varepsilon^{-C_0 L} \leqslant P$ and $\varepsilon \delta^{-5} L \leqslant 1$.

Then, there exists a decomposition $A = A_0 \sqcup \ldots \sqcup A_L \sqcup A_{L+1}$, together with a lift $\mathcal{B} = (A \setminus A_{L+1}, \beta_\bullet, Hp^*)$ of $\mathcal{A}$, with the following properties.

(I) $|A_0| \gg \delta^5 |A|$.

(II) For all but $\ll \varepsilon |A||\mathcal{P}|^2$ quadruples $(x, y, p, q) \in Q(\mathcal{A})$,

- *(a) if $x \in A_\ell$ then $y \in A_{\ell+e}$ for some $e \in \{-1, 0, 1\}$, and*
- *(b) if $x, y \notin A_{L+1}$, then $(x, y, p, q) \in Q(\mathcal{B})$.*

Proof of Theorem 3.1.8. The assumptions $\varepsilon^{-C_0 L} \leqslant P$ and $\varepsilon \delta^{-5} L \leqslant 1$ imply that $\delta \geqslant \varepsilon \geqslant P^{-1/C_0}$.

We will use an iterative argument, working with clusters of increasing diameter. To this end, we set up a hierarchy of scales

$$\mathcal{C} = \{500^\ell : 0 \leqslant \ell \leqslant L + 1\}.$$

For each $C \in \mathcal{C}$, we apply Lemma 3.6.2 with the parameters C and $\eta = |\mathcal{P}|^{-1/100}$ to obtain a collection $(K_{i;C})_{i \in I_C}$ of clusters of diameter $\leqslant 16C$ satisfying the two properties in that lemma. These auxiliary clusters will be used in the construction of the sets A_ℓ.

By Lemma 3.7.1, there exist $\asymp \delta |A|$ disjoint clusters of diameter $\leqslant 1/2$ and size $\gg \delta^4 |\mathcal{P}|$. By an averaging argument, there is a set $\mathcal{P}^* \subset \mathcal{P}$ of size $\gg \delta^5 |\mathcal{P}|$ such that every $p \in \mathcal{P}^*$ appears in $\gg \delta^5 |A|$ of these clusters. That is, for each $p \in \mathcal{P}^*$, there are $\gg \delta^5 |A|$ elements $x \in A$ such that (x, p) belongs to a cluster of diameter $\leqslant 1/2$ and size $\gg \delta^4 |\mathcal{P}|$.

We now select a prime $\mathbf{p}^*$ uniformly at random from $\mathcal{P}^*$. We will show that Theorem 3.1.8 holds with positive probability with this prime as the lifting prime.

To construct the required decomposition of A, we will define a subset $\mathcal{R} \subset Q(\mathcal{A})$ of quadruples satisfying various technical properties, which we now describe.

- *Regularity.* Let A_{reg} be the random set of $x \in A$ such that $(x, \mathbf{p}^*)$ is $(C, |\mathcal{P}|^{-1/100})$-regular for all $C \in \mathcal{C}$. By Lemma 3.6.4,

$$\mathbb{E}_{\mathbf{p}^* \in \mathcal{P}^*} |A \setminus A_{\mathrm{reg}}| \ll \frac{1}{|\mathcal{P}^*|} \sum_{C \in \mathcal{C}} C^9 |\mathcal{P}|^{1/20} |A| \ll \frac{e^{O(L)}}{\delta^5 |\mathcal{P}|^{1-1/20}} |A|.$$

By Markov's inequality and our choice of parameters (choosing C_0 to be sufficiently large), with probability at least 99%, we have

$$|A \setminus A_{\mathrm{reg}}| \ll \frac{|A|}{|\mathcal{P}|^{1-1/10}}. \tag{3.38}$$

- *Rigidity.* By Lemma 3.7.3, for every $p_0 \in \mathcal{P}$, the number of $(x,y,p,q) \in Q(\mathcal{A})$ satisfying $N_{\square}(x,y,p,q,p_0;C) > |\mathcal{P}|^{19/20}$ or $N_{\square\!\!\!/}(x,p,q,p_0;C) > |\mathcal{P}|^{39/20}$ for some $C \in \mathcal{C}$ is bounded by

$$\ll \sum_{C\in\mathcal{C}} \left(\frac{C^3|A||\mathcal{P}|^2}{|\mathcal{P}|^{19/20}} + \frac{C^5|A||\mathcal{P}|^3}{|\mathcal{P}|^{39/20}} \right) \ll e^{O(L)}|A||\mathcal{P}|^{1+1/20} \ll |A||\mathcal{P}|^{1+1/10}. \tag{3.39}$$

- *Connectivity.* Let Q_{bad} be the random set of quadruples $(x,y,p,q) \in Q(\mathcal{A})$ such that one of the previously defined clusters $K_{i;C}$ has the following properties: $K_{i;C}$ contains $(x,\mathbf{p}^*)$ and there are fewer than $\varepsilon^2|K_{i;C}|$ elements $(x',r) \in K_{i;C}$ for which $d_{p,q}(x',y') \leqslant 1/10$ for some $y' \in A$.

  By the union bound, we have

$$\mathbb{E}_{\mathbf{p}^*\in\mathcal{P}^*}|Q_{\mathrm{bad}}| \leqslant \frac{1}{|\mathcal{P}^*|} \sum_{C\in\mathcal{C}} \sum_{i\in I_C} \sum_{p_0\in\mathcal{P}^*} \sum_{\substack{(x,y,p,q)\in Q(\mathcal{A}) \\ (x,p_0)\in K_{i;C}}} I(p,q;i,C)$$

  where $I(p,q;i,C)$ is the indicator function that the number of triples $(x',y',r) \in A^2 \times \mathcal{P}$ with $(x',y',p,q) \in Q(\mathcal{A})$ and $(x',r) \in K_{i;C}$ is less than $\varepsilon^2|K_{i;C}|$. For any $p,q \in \mathcal{P}$, by definition of $I(p,q;i,C)$, we have

$$I(p,q;i,C) \sum_{\substack{x,y\in A \\ (x,y,p,q)\in Q(\mathcal{A})}} \sum_{\substack{p_0\in\mathcal{P} \\ (x,p_0)\in K_{i;C}}} 1 \leqslant \varepsilon^2|K_{i;C}|.$$

  Inserting this bound into the previous expression, we obtain

$$\mathbb{E}_{\mathbf{p}^*\in\mathcal{P}^*}|Q_{\mathrm{bad}}| \leqslant \frac{1}{|\mathcal{P}^*|} \sum_{C\in\mathcal{C}} \sum_{i\in I_C} \sum_{p,q\in\mathcal{P}} \varepsilon^2|K_{i;C}| \ll \varepsilon^2\delta^{-5}|\mathcal{P}| \sum_{C\in\mathcal{C}} \sum_{i\in I_C} |K_{i;C}|.$$

  Since $(K_{i;C})_{i\in I_C}$ satisfies the second property in Lemma 3.6.2 with $\eta = |\mathcal{P}|^{-1/100}$, we have

$$\sum_{i\in I_C} |K_{i;C}| \ll |A||\mathcal{P}| + C^9|\mathcal{P}|^{1/20}|A|.$$

  We conclude that

$$\mathbb{E}_{\mathbf{p}^*\in\mathcal{P}^*}|Q_{\mathrm{bad}}| \ll \varepsilon^2\delta^{-5}\big(L|A||\mathcal{P}|^2 + e^{O(L)}|A||\mathcal{P}|^{1+1/20}\big).$$

  By Markov's inequality and our choice of parameters (recall that $\varepsilon\delta^{-5}L \leqslant 1$), with probability at least 99%, we have

$$|Q_{\mathrm{bad}}| \ll \varepsilon|A||\mathcal{P}|^2. \tag{3.40}$$

We can now define the random set $\mathcal{R}$ to be the collection of all $(x,y,p,q) \in Q(\mathcal{A})$ satisfying the following three properties:

(a) $x, y \in A_{\mathrm{reg}}$.

(b) For all $C \in \mathcal{C}$,

$$\begin{cases} N_{\square}(x,y,p,q,\mathbf{p}^*;C) \leqslant |\mathcal{P}|^{19/20} \\ N_{\square\!\!\!\!\diagup}(x,p,q,\mathbf{p}^*;C) \leqslant |\mathcal{P}|^{39/20} \end{cases} \quad \text{and} \quad \begin{cases} N_{\square}(y,x,q,p,\mathbf{p}^*;C) \leqslant |\mathcal{P}|^{19/20} \\ N_{\square\!\!\!\!\diagup}(y,q,p,\mathbf{p}^*;C) \leqslant |\mathcal{P}|^{39/20}. \end{cases}$$

(c) Neither (x,y,p,q) nor (y,x,q,p) belongs to Q_{bad}.

By (3.38) to (3.40), with positive probability, we have $|A \setminus A_{\mathrm{reg}}| \ll |A||\mathcal{P}|^{-1+1/10}$ and

$$|Q(\mathcal{A}) \setminus \mathcal{R}| \ll \varepsilon |A||\mathcal{P}|^2. \tag{3.41}$$

For the rest of this proof, we fix a realisation $p^* \in \mathcal{P}^*$ of $\mathbf{p}^*$ for which this holds. By a slight abuse of notation, we will henceforth use A_{reg}, Q_{bad} and $\mathcal{R}$ to refer to the deterministic sets obtained by evaluating these random sets at the fixed prime p^*.

We turn to the construction of the decomposition $A = A_0 \sqcup \ldots \sqcup A_L \sqcup A_{L+1}$.

By definition of $\mathcal{P}^*$ and (3.38), there is a subset $A_0 \subset A_{\mathrm{reg}}$ of size $\gg \delta^5|A|$ such that, for each $x \in A_0$, the pair (x,p^*) is contained in a cluster of diameter $\leqslant 1/2$ and size $\gg \delta^4|\mathcal{P}|$. In particular, property (I) of Theorem 3.1.8 holds.

Suppose that $A_0, \ldots, A_{\ell-1}$ have been constructed, for some $1 \leqslant \ell \leqslant L$. We define

$$A_\ell := \Big\{ y \in A_{\mathrm{reg}} \setminus \bigsqcup_{k<\ell} A_k \ : \ \exists (x,y,p,q) \in \mathcal{R} \text{ with } x \in A_{\ell-1} \Big\}. \tag{3.42}$$

Finally, we define $A_{L+1} := A \setminus \bigsqcup_{\ell \leqslant L} A_\ell$. Property (IIa) follows directly from the definition of the sets A_ℓ, the symmetry of $\mathcal{R}$ (i.e. the fact that $(x,y,p,q) \in \mathcal{R}$ if and only if $(y,x,q,p) \in \mathcal{R}$) and (3.41).

We shall now iterate Lemma 3.7.4 to construct, for each $x \in A \setminus A_{L+1}$, a cluster containing (x,p^*).

Claim 1. *Let $(x,y,p,q) \in \mathcal{R}$, where $x \in A_k$ for some $k \leqslant L$. Suppose that (x,p^*) is contained in a cluster S of diameter $\leqslant 500^k$ and size $\geqslant |\mathcal{P}|^{99/100}$. Then, there is a cluster S' of diameter $\leqslant 500^{k+1}$ and size $\geqslant \frac{1}{2}\varepsilon^2|S|$ containing (y,p^*).*

Proof of Claim 1. Let $C := 500^k$. By definition of the collection $(K_{i;C})_{i \in I_C}$, there exists $i \in I_C$ such that $(x,p^*) \in K_{i;C}$ and $|K_{i;C}| \geqslant |S|$ (see Lemma 3.6.2). Let S_1 be the set of all $(x',r) \in K_{i;C}$ such that $d_{p,q}(x',y') \leqslant 1/10$ for some $y' \in A$. By property (c) in the definition of $\mathcal{R}$, we have $(x,y,p,q) \notin Q_{\mathrm{bad}}$, which means that $|S_1| \geqslant \varepsilon^2|K_{i;C}| \geqslant \varepsilon^2|S|$.

We are now in a position to apply Lemma 3.7.4 to S_1 (which is a cluster of diameter $\leqslant 16C$ as it is contained in $K_{i;C}$). The assumption (3.37) in Lemma 3.7.4 holds by

property (b) in the definition of $\mathcal{R}$ and our choice of parameters. Thus, Lemma 3.7.4 yields a cluster S' of diameter $\leqslant 24 \cdot 16C \leqslant 500^{k+1}$ and size $\geqslant |S_1|/2 \geqslant \frac{1}{2}\varepsilon^2|S|$ containing (y, p^*), as required. ◁

Claim 2. *For every $k \leqslant L$ and $x \in A_k$, the pair (x, p^*) is contained in a cluster S of diameter $\leqslant 500^k$ and size $\geqslant \max(|\mathcal{P}|^{99/100}, c_0(\frac{1}{2}\varepsilon^2)^k\delta^4|\mathcal{P}|)$ where $c_0 > 0$ is an absolute constant.*

Proof of Claim 2. This follows from Claim 1 by induction on k. For $k = 0$, this is true by definition of A_0. For $1 \leqslant k \leqslant L$, we know that each $y \in A_k$ is connected to some $x \in A_{k-1}$ by a quadruple $(x, y, p, q) \in \mathcal{R}$. By the induction hypothesis, (x, p^*) is contained in a cluster of diameter $\leqslant 500^{k-1}$ and size $\geqslant c_0(\frac{1}{2}\varepsilon^2)^{k-1}\delta^4|\mathcal{P}|$. This size is at least $|\mathcal{P}|^{99/100}$ by our assumption that $\varepsilon^{-C_0 L} \leqslant P$ and $\delta^{-C_0} \leqslant P$ for a sufficiently large constant C_0. Applying Claim 1 completes the proof. ◁

We remark that any cluster S as in Claim 2 contains at least one pair (x', r) with $r \neq p^*$, by Lemma 3.5.5 and the lower bound $|S| \geqslant |\mathcal{P}|^{99/100}$. This will be useful in the next step.

Now, we define the frequencies β_x for $x \in A \setminus A_{L+1}$. Let $k \leqslant L$ and $x \in A_k$. Let $S(x)$ be the union of all clusters of diameter $\leqslant 500^k$ and size $\geqslant |\mathcal{P}|^{99/100}$ containing (x, p^*); note that $S(x)$ is non-empty by Claim 2. By construction, $A_k \subset A_{\mathrm{reg}}$, which implies that $S(x)$ is a cluster of diameter $\leqslant 64 \cdot 500^k$. We define β_x to be the real number given by applying Lemma 3.5.7 to this cluster $S(x)$ and the pair $(x, p^*) \in S(x)$. The first property in Lemma 3.5.7 ensures that $\mathcal{B} := (A \setminus A_{L+1}, \beta_\bullet, Hp^*)$ is a lift of $\mathcal{A}$.

It only remains to prove (IIb) of Theorem 3.1.8. In view of our bound (3.41) on $|Q(\mathcal{A}) \setminus \mathcal{R}|$, it suffices to show that, for every $(x, y, p, q) \in \mathcal{R}$ with $x \in A_k$ for some $k \leqslant L$, we have $(x, y, p, q) \in Q(\mathcal{B})$. Consider such a quadruple, and let S be a cluster of diameter $\leqslant 500^k$ and size $\geqslant |\mathcal{P}|^{99/100}$ containing (x, p^*) (which exists by Claim 2). Let $K_{i;C}$, S_1 and S' be the clusters constructed in the proof of Claim 1. In particular, $(y, p^*) \in S'$. Let $(y', r) \in S'$ be any pair such that $r \neq p^*$. Recall that S' was obtained from S_1 by applying Lemma 3.7.4. By the last property in Lemma 3.7.4, there exists $(x', r) \in S_1$ such that $d_{p,q}(x', y') \leqslant 500^k$, so that $\|q\alpha_{x'} - p\alpha_{y'}\| \leqslant 500^k/H$. Therefore, by the triangle inequality,

$$\|r(q\beta_x - p\beta_y)\| \leqslant \|q(r\beta_x - \alpha_{x'})\| + \|q\alpha_{x'} - p\alpha_{y'}\| + \|p(r\beta_y - \alpha_{y'})\| \ll \frac{500^L}{H},$$

using the second property in Lemma 3.5.7. Moreover,

$$\|p^*(q\beta_x - p\beta_y)\| = \|q\alpha_x - p\alpha_y\| \leqslant \frac{1}{10H}$$

as $(x, y, p, q) \in Q(\mathcal{A})$. Since $p^* \neq r$, we get $\|q\beta_x - p\beta_y\| \leqslant (10Hp^*)^{-1}$ by Lemma 3.5.1. Thus, $(x, y, p, q) \in Q(\mathcal{B})$, which completes the proof of Theorem 3.1.8. □

3.8 Recovering Walsh's conditional theorem

In this section, we show how Walsh's Fourier uniformity result under GRH [73] can be derived from our global structure theorem, yielding explicit quantitative savings (see Corollary 3.8.7).

3.8.1 Conditional expansion estimate

Assuming the Generalised Riemann Hypothesis for Dirichlet L-functions, we can replace Proposition 3.4.4 with a much stronger expansion estimate.

Lemma 3.8.1 (GRH expansion estimate)**.** *Assume GRH. Suppose that $P \geqslant (\log Y)^C$ for some sufficiently large absolute constant $C > 0$.*

Let $A, B \subset \left[\frac{1}{10}Y, 10Y\right]$ be multisets with at most M elements in any unit interval.

Let $1 \leqslant q_0 \leqslant Y^{O(1)}$ be an integer, let $\mathcal{P}' \subset \mathcal{P}$ consist of all primes not dividing q_0, and let $(a_x)_{x\in A}$ and $(b_y)_{y\in B}$ be sequences of integers coprime to q_0.[6]

Let $1/Y \leqslant \varepsilon \leqslant 1$. Then, the number of quadruples $(x, y, p, q) \in A \times B \times (\mathcal{P}')^2$ such that

$$|px - qy| \leqslant \varepsilon P \quad \text{and} \quad qa_x \equiv pb_y \pmod{q_0}$$

is

$$\ll M^2\left(|A|^{1/2}|B|^{1/2}P^{2-c} + \frac{\varepsilon|A||B||\mathcal{P}|^2}{Y\varphi(q_0)}\right),$$

where $c > 0$ is an absolute constant.

Proof. Repeating the arguments in the proof of Proposition 3.4.4, we see that the number of quadruples under consideration is bounded by a constant times

$$M^2\left((w|\mathcal{P}|)^2|A|^{1/2}|B|^{1/2} + \frac{\varepsilon|A||B|}{Y\varphi(q_0)}\sum_{\chi \,(\mathrm{mod}\, q_0)}\int_{-Y/\varepsilon}^{Y/\varepsilon}\left|\sum_{p\in\mathcal{P}}\chi(p)p^{it}\right|^2 \mathbf{1}_{\{|\sum_{p\in\mathcal{P}}\overline{\chi(p)}p^{it}|>w|\mathcal{P}|\}}dt\right),$$

for any $w > 0$. Setting $w := P^{-c}$ for some small absolute constant $c > 0$ and bounding $\sum_{p\in\mathcal{P}}\chi(p)p^{it}$ using GRH (see [73, Section 2.2] for details), we obtain the claimed bound. □

We remark that the Riemann Hypothesis is sufficient to obtain Lemma 3.8.1 in the case $q_0 = 1$. Specialising to $M = q_0 = \varepsilon = 1$ yields the following corollary.

Corollary 3.8.2. *Assume RH. Suppose that $P \geqslant (\log Y)^C$ for some sufficiently large constant $C > 0$.*

Then, any configuration $\mathcal{A} = (A, \alpha_\bullet, H)$ such that $|Q(\mathcal{A})| \geqslant \delta|A||\mathcal{P}|^2$ for some $\delta \geqslant (\log Y)^{-100}$ satisfies

$$|A| \gg \delta Y.$$

[6] We assume that C is sufficiently large in terms of the implied constant in the upper bound for q_0.

The conclusion of Corollary 3.8.2 is essentially as strong as what the random graph heuristic would predict.

3.8.2 Construction of a tower of configurations

For P as small as in Walsh's conditional setting, we cannot directly apply Proposition 3.2.3 to obtain a tower of configurations. To construct a suitable tower, we instead combine our local structure theorem, Theorem 3.1.8, with the strong expansion estimate given by Lemma 3.8.1.

We start by proving a variant of Lemma 3.2.2. The main observation is that cases (ii) and (iii) of Lemma 3.8.3 pass to a subset of A that retains more local relations than would be expected from a random subset of the same size. Unlike in the unconditional setting, losing the constant factor $\frac{99}{100}$ in case (iii) is acceptable here since the RH expansion estimate is strong enough to handle a loss of this magnitude.

Lemma 3.8.3. *Assume RH. Let $\delta > 0$. Let $\mathcal{A} = (A, \alpha_\bullet, H)$ be a configuration such that $|Q(\mathcal{A})| \geqslant \delta|A||\mathcal{P}|^2$, and suppose that $P \geqslant (\log Y)^{C\delta^{-1}}$ for some sufficiently large absolute constant $C > 0$.*

Then, one of the following is true.

(i) There exists a lift $\mathcal{B} = (B, \beta_\bullet, Hp^*)$ of $\mathcal{A}$ such that

$$|Q(\mathcal{B})| \geqslant \Big(1 - O\Big(\frac{1}{(\log Y)^{10}}\Big)\Big)\delta|B||\mathcal{P}|^2.$$

(ii) There is a proper subset $B \subset A$ such that $|B| \geqslant \frac{1}{2}|A|$ and $|Q(\mathcal{A}|_B)| \geqslant \delta|B||\mathcal{P}|^2$.

(iii) There is a subset $B \subset A$ such that $|B| \leqslant \frac{1}{2}|A|$ and $|Q(\mathcal{A}|_B)| \geqslant \frac{99}{100}\delta|B||\mathcal{P}|^2$.

Proof. Let $\varepsilon := (\log Y)^{-20}$ and let L be the largest integer satisfying the conditions $\varepsilon^{-C_0 L} \leqslant P$ and $\varepsilon\delta^{-5}L \leqslant 1$ of Theorem 3.1.8. Note that $\delta \geqslant 1/\log Y$ as $Y \geqslant P^3 \geqslant (\log Y)^{C\delta^{-1}}$. Thus, the parameter L is only limited by the condition $\varepsilon^{-C_0 L} \leqslant P$, so we have $L \asymp \log P/\log\log Y$.

Let $A = A_0 \sqcup A_1 \sqcup \ldots \sqcup A_{L+1}$ and $\mathcal{B} = (A \setminus A_{L+1}, \beta_\bullet, Hp^*)$ be the decomposition and the lift given by Theorem 3.1.8.

Case 1. If $|A_{L+1}| \leqslant |A|/(\log Y)^{11}$, then by part (IIb) of Theorem 3.1.8, we have

$$|Q(\mathcal{B})| \geqslant \delta|A||\mathcal{P}|^2 - O\big(\varepsilon|A||\mathcal{P}|^2 + |A_{L+1}||\mathcal{P}|^2\big) \geqslant \delta|B||\mathcal{P}|^2 - O\Big(\frac{|A||\mathcal{P}|^2}{(\log Y)^{11}}\Big),$$

so that conclusion (i) holds. We may therefore assume that $|A_{L+1}| \geqslant |A|/(\log Y)^{11}$.

Case 2. Suppose that there exists $1 \leqslant i \leqslant L$ such that $|A_i| \leqslant |A|/(\log Y)^{13}$. Let $B_1 := \bigsqcup_{j=0}^{i} A_j$ and $B_2 := \bigsqcup_{j=i+1}^{L+1} A_j$. Then, by part (IIa) of Theorem 3.1.8,

$$\begin{aligned}|Q(\mathcal{A}|_{B_1})| + |Q(\mathcal{A}|_{B_2})| &\geqslant \delta|A||\mathcal{P}|^2 - O\big(\varepsilon|A||\mathcal{P}|^2 + |A_i||\mathcal{P}|^2\big)\\ &\geqslant \delta|A||\mathcal{P}|^2 - O\bigg(\frac{|A||\mathcal{P}|^2}{(\log Y)^{13}}\bigg).\end{aligned}$$

Since $|A_0| \gg |A|/(\log Y)^5$ by part (I) of Theorem 3.1.8, and $|A_{L+1}| \geqslant |A|/(\log Y)^{11}$, we have $\min(|B_1|, |B_2|) \gg |A|/(\log Y)^{11}$. Hence, the previous bound can be rewritten as

$$|Q(\mathcal{A}|_{B_1})| + |Q(\mathcal{A}|_{B_2})| \geqslant \delta|B_1||\mathcal{P}|^2 + \delta|B_2||\mathcal{P}|^2 - O\bigg(\frac{\delta\min(|B_1|, |B_2|)|\mathcal{P}|^2}{\log Y}\bigg).$$

For Y sufficiently large, this estimate implies that one of conclusions (ii) or (iii) holds for some choice of $B \in \{B_1, B_2\}$: if $|B_1| \geqslant |A|/2$, then either conclusion (ii) holds for $B = B_1$, or conclusion (iii) holds for $B = B_2$; the case $|B_2| \geqslant |A|/2$ is treated symmetrically.

Case 3. It only remains to treat the case where $|A_i| \gg |A|/(\log Y)^{13}$ for all $0 \leqslant i \leqslant L+1$. By averaging, there exists $1 < i < L$ such that $|A_i| + |A_{i+1}| \ll |A|/L$.

Let $B_1 := \bigsqcup_{j=0}^{i} A_j$ and $B_2 := \bigsqcup_{j=i+1}^{L+1} A_j$. By the conditional expansion estimate, Lemma 3.8.1, the number of quadruples $(x, y, p, q) \in Q(\mathcal{A})$ with $x \in A_i$ and $y \in A_{i+1}$ is bounded by

$$\ll |A_i|^{1/2}|A_{i+1}|^{1/2}P^{2-c} + \frac{|A_i||A_{i+1}||\mathcal{P}|^2}{Y} \ll \frac{|A_i||A_{i+1}||\mathcal{P}|^2}{Y},$$

using that $P \geqslant (\log Y)^C$ for some sufficiently large constant $C > 0$ in the last inequality. Therefore, by part (IIa) of Theorem 3.1.8, we get

$$|Q(\mathcal{A}|_{B_1})| + |Q(\mathcal{A}|_{B_2})| \geqslant \delta|A||\mathcal{P}|^2 - O\bigg(\varepsilon|A||\mathcal{P}|^2 + \frac{|A_i||A_{i+1}||\mathcal{P}|^2}{Y}\bigg).$$

Since $|A|/(\log Y)^{13} \ll |A_i|, |A_{i+1}| \ll |A|/L$ and $\varepsilon = (\log Y)^{-20}$, the error term can be bounded by $O(\min(|B_1|, |B_2|)|\mathcal{P}|^2/L)$, so that

$$|Q(\mathcal{A}|_{B_1})| + |Q(\mathcal{A}|_{B_2})| \geqslant \delta|B_1||\mathcal{P}|^2 + \delta|B_2||\mathcal{P}|^2 - O\bigg(\frac{\delta\min(|B_1|, |B_2|)|\mathcal{P}|^2}{\delta L}\bigg).$$

As before, this estimate implies that one of conclusions (ii) or (iii) holds for some $B \in \{B_1, B_2\}$, provided that δL is sufficiently large. Since

$$L \asymp \frac{\log P}{\log\log Y} \geqslant C\delta^{-1},$$

this is indeed the case if C is chosen to be sufficiently large. □

Iterating the preceding lemma, we can construct a tower of configurations with a large relative density of local relations at the top level.

Lemma 3.8.4. *Assume RH. Let $\delta > 0$. Let $\mathcal{A} = (A, \alpha_\bullet, H)$ be a configuration such that $|Q(\mathcal{A})| = \delta|A||\mathcal{P}|^2$, and suppose that $P \geqslant (\log Y)^{C\delta^{-2}}$ for some sufficiently large absolute constant $C > 0$.*

Then, there exists a tower of configurations $(\mathcal{A}_i)_{0\leqslant i\leqslant k}$ of height $k \gg (\log Y)^{10}$, such that $\mathcal{A}_0 = \mathcal{A}$ and $|Q(\mathcal{A}_k)| \gg \delta^2|A_k||\mathcal{P}|^2$ (where A_k is the set of points of $\mathcal{A}_k$).

Proof. We imitate the proof of Proposition 3.2.3, using Lemma 3.8.3 instead of Lemma 3.2.2.

That is, by induction, we construct a sequence of triples $(\mathcal{B}_i, B_i, \delta_i)$ where $\mathcal{B}_i$ is a configuration with point set B_i satisfying $|Q(\mathcal{B}_i)| = \delta_i|B_i||\mathcal{P}|^2$. We start with $(\mathcal{B}_0, B_0, \delta_0) := (\mathcal{A}, A, \delta)$ and construct the subsequent triples by repeatedly applying Lemma 3.8.3. Thus, for each $i \geqslant 1$, one of the following alternatives holds:

- *Lifting step:* $\mathcal{B}_i$ is a lift of $\mathcal{B}_{i-1}$ and $\delta_i \geqslant (1 - O((\log Y)^{-10}))\delta_{i-1}$.
- *Lossless step:* $\mathcal{B}_i = \mathcal{B}_{i-1}|_{B_i}$ for some proper subset $B_i \subset B_{i-1}$ and $\delta_i \geqslant \delta_{i-1}$.
- *Halving step.* $\mathcal{B}_i = \mathcal{B}_{i-1}|_{B_i}$ for some $B_i \subset B_{i-1}$ with $|B_i| \leqslant \frac{1}{2}|B_{i-1}|$, and $\delta_i \geqslant \frac{99}{100}\delta_{i-1}$.

We run this iterative procedure until either $\delta_i < c\delta^2$ or we have encountered $\lfloor c(\log Y)^{10}\rfloor$ lifting steps, for some small parameter $c > 0$ to be chosen later.

Suppose that, for some $i_1 \geqslant 1$, we have $\delta_{i_1} < c\delta^2$, but there have been fewer than $\lfloor c(\log Y)^{10}\rfloor$ lifting steps up to time i_1. Then, if $c > 0$ is sufficiently small, we have $\delta_{i_1} \asymp (99/100)^m\delta$, where m is the number of halving steps up to time i_1. This is because lossless steps do not decrease δ_i, and lifting steps decrease δ_i by at most a factor $1 - O((\log Y)^{-10})$. Thus, we see that

$$\delta_{i_1} \asymp \left(\tfrac{99}{100}\right)^m\delta \asymp c\delta^2 \quad \text{and} \quad |B_{i_1}| \leqslant 2^{-m}|A|.$$

However, since $|Q(\mathcal{B}_{i_1})| = \delta_{i_1}|B_{i_1}||\mathcal{P}|^2$, the expansion estimate, Corollary 3.8.2, implies that

$$|B_{i_1}| \gg \delta_{i_1}Y.$$

Combining these estimates, we get

$$c\delta^2 \ll 2^{-m} \leqslant \left(\tfrac{99}{100}\right)^{2m} \asymp c^2\delta^2,$$

which is a contradiction if $c > 0$ is a sufficiently small absolute constant.

As a result, the algorithm must terminate after $\gg (\log Y)^{10}$ lifting steps have been performed. Defining $\mathcal{A}_i$ to be the configuration obtained after the i-th lifting step, we obtain the desired tower of configurations. □

3.8.3 Derivation of Walsh's conditional theorem

We now prove a stronger version of Lemma 3.4.6 assuming GRH, and then deduce Walsh's global structure theorem.

Lemma 3.8.5. *Assume GRH. There is an absolute constant $C > 0$ such that the following holds.*

Suppose that $P \geqslant (\log Y)^C$. Let $\mathcal{A} = (A, \alpha_\bullet, H)$ be a configuration such that $|Q(\mathcal{A})| \geqslant \delta |A||\mathcal{P}|^2$, where $H \geqslant P^{3\log Y}$ and $\delta \geqslant C/\log|\mathcal{P}|$.

Then, there exist

- *a subset $A' \subset A$ with $|Q(\mathcal{A}|_{A'})| \gg \delta|A'||\mathcal{P}|^2$,*
- *an integer $q_0 \ll \delta^{-2}$, and*
- *a real number t with $|t| \ll \delta^{-1}(\log Y)^3 \frac{Y^2}{HP}$,*

such that, for every $x \in A'$, there is an integer a_x coprime to q_0 such that

$$\alpha_x = \frac{a_x}{q_0} + \frac{t}{x} + O\left(\frac{\delta^{-1}(\log Y)^4}{HP}\right) \pmod 1. \tag{3.43}$$

Proof. The proof is the same as that of Lemma 3.4.6, with the GRH expansion lemma Lemma 3.8.1 replacing Proposition 3.4.4. We quickly reproduce the main steps for the reader's convenience.

After applying Lemma 3.3.7, we obtain an initial bound on q_0 of the form

$$q_0 \leqslant \exp\left(O\left(\frac{(\log Y)\log\log Y}{\log|\mathcal{P}|} + \log|\mathcal{P}|\right)\right) \leqslant Y^{O(1)}.$$

We also obtain a subset $A_2 \subset A$ such that, for all $x \in A_2$,

$$\alpha_x = \frac{a_x}{q_0} + \beta_x \pmod 1,$$

where $(a_x, q_0) = 1$ and $|\beta_x| \leqslant P^{-2\log Y}$. Moreover, for each $x \in A_2$, there are $\gg \delta|\mathcal{P}|^2$ quadruples $(x, y, p, q) \in Q(\mathcal{A}|_{A_2})$ such that $(pq, q_0) = 1$, $p \neq q$, $qa_x \equiv pa_y \pmod{q_0}$ and $|q\beta_x - p\beta_y| \leqslant 1/H$.

The GRH expansion estimate, Lemma 3.8.1, applied with $A = B = A_2$ and $M = \varepsilon = 1$, implies that

$$q_0 \ll \varphi(q_0)^{1+o(1)} \ll \delta^{-1+o(1)}.$$

Applying Lemma 3.3.8 to the configuration $\mathcal{B} := (A_2, \beta_\bullet, H)$ then produces a subset $B' \subset A_2$ with $|Q(\mathcal{B}|_{B'})| \gg \delta|B'||\mathcal{P}|^2$ and an element $x_0 \in B'$ such that, for every $x \in B'$,

$$\alpha_x = \frac{a_x}{q_0} + \frac{x_0\beta_{x_0}}{x} + O\left(\frac{|\beta_{x_0}|\log Y}{Y} + \frac{\log Y}{HP}\right) \pmod 1.$$

As in the proof of Lemma 3.4.6, we bound $|\beta_{x_0}|$ by applying an expansion estimate to the multiset V given in Lemma 3.3.8. By Lemma 3.8.1, the number of quadruples $(v_1, v_2, p, q) \in V^2 \times \mathcal{P}^2$ such that $|pv_1 - qv_2| \leqslant \varepsilon P$ is

$$\ll (\log Y)^2 \left(|V| P^{2-c} + \frac{\varepsilon |V|^2 |\mathcal{P}|^2}{Y} \right)$$

for any $1/Y \leqslant \varepsilon \leqslant 1$. Comparing this with the final conclusion of Lemma 3.3.8, we deduce that

$$|\beta_{x_0}| \ll \delta^{-1} (\log Y)^3 \frac{Y}{HP},$$

which gives the claimed result. □

Finally, we can prove Walsh's global structure theorem under GRH.

Theorem 3.8.6. *Assume GRH. There is an absolute constant $C > 0$ such that the following holds.*

Let $\delta > 0$. Let $\mathcal{A} = (A, \alpha_\bullet, H)$ be a configuration with $|Q(\mathcal{A})| \geqslant \delta |A| |\mathcal{P}|^2$, where $\delta \geqslant C(\log P)^{-1/2}$.

Suppose that $P \geqslant (\log Y)^{C\delta^{-2}}$. Then, there is a subset $A' \subset A$ of size

$$|A'| \gg \delta^6 |A|,$$

such that, for all $x \in A'$, we have the approximate formula

$$\alpha_x = \frac{a_0}{q_0} + \frac{T_0}{x} + O\left(\frac{\delta^{-2} (\log Y)^4}{HP} \right) \pmod 1,$$

where a_0, q_0 are coprime integers with $1 \leqslant q_0 \ll \delta^{-4}$ and T_0 is a real number satisfying $|T_0| \leqslant \delta^{-2} (\log Y)^3 \frac{Y^2}{HP}$.

Proof. The proof of Theorem 3.8.6 closely follows that of Theorem 3.1.9, with Lemma 3.8.4 and Lemma 3.8.5 replacing Proposition 3.2.3 and Lemma 3.4.6, respectively.

By Lemma 3.8.4, we can construct a tower of configurations $(\mathcal{A}_i)_{0 \leqslant i \leqslant k}$ of height $k \gg (\log Y)^{10}$, such that $\mathcal{A}_0 = \mathcal{A}$ and $|Q(\mathcal{A}_k)| \gg \delta^2 |A_k| |\mathcal{P}|^2$, where $\mathcal{A}_k = (A_k, \widetilde{\alpha}_\bullet, H_k)$.

Applying Lemma 3.8.5 to $\mathcal{A}_k$, we find $A' \subset A_k$ with $|Q(\mathcal{A}_k|_{A'})| \gg \delta^2 |A'| |\mathcal{P}|^2$, along with an integer $q_0 \ll \delta^{-4}$ and a real number $|t| \ll \delta^{-2} (\log Y)^3 \frac{Y^2}{H_k P}$, such that, for every $x \in A'$,

$$\widetilde{\alpha}_x = \frac{a_x}{q_0} + \frac{t}{x} + O\left(\frac{\delta^{-2} (\log Y)^4}{H_k P} \right) \pmod 1$$

for some integer a_x coprime to q_0.

By Corollary 3.8.2, we have the lower bound $|A'| \gg \delta^2 Y$. By pigeonholing on the values of $a_x \pmod{q_0}$, there exists a subset $A'' \subset A'$ with $|A''| \gg \delta^6 Y$ such that a_x is constant for $x \in A''$. The conclusion then follows from the definition of a lift, as in the proof of Theorem 3.1.9. □

Combining Theorem 3.8.6 with Proposition B.2.1 and Theorem C.5.1 (as in the proof of Theorem 1.3.3), we recover a quantitative form of Walsh's conditional Fourier uniformity estimate [73].

Corollary 3.8.7. *Assume GRH. Let $H = H(X) = (\log X)^{\psi(X)}$, where $\psi(X) \to \infty$ as $X \to \infty$, and $\psi(X) \leqslant (\log X)^{1-c}$ for some fixed $c > 0$.*

Then

$$\sum_{X \leqslant x < 2X} \sup_{\alpha \in \mathbb{R}/\mathbb{Z}} \left| \sum_{x \leqslant n < x+H} \lambda(n) e(n\alpha) \right| \ll \frac{HX}{\sqrt{\log \psi(X)}}.$$

Proof. Let X be sufficiently large, and suppose that the conclusion fails. Then, there exists an H-separated subset $S \subseteq [X, 2X)$ of size $\gg \delta X/H$, such that for each $x \in S$, there exists $\theta_x \in \mathbb{R}/\mathbb{Z}$ satisfying

$$\left| \sum_{x \leqslant n < x+H} \lambda(n) e(n\theta_x) \right| \gg \delta H,$$

where

$$\delta := \frac{C_1}{\sqrt{\log \psi(X)}} \tag{3.44}$$

for some large constant $C_1 > 0$ to be chosen later.

By Proposition B.2.1 (the assumption on δ holds by (3.44) if C_1 is sufficiently large), there exists a scale P with $H^{e^{-O(\delta^{-2})}} \leqslant P \leqslant H^{1/10}$ such that, writing $\mathcal{P}$ for the set of primes in $[P, 2P)$, the number of quadruples $(x, y, p, q) \in S^2 \times \mathcal{P}^2$ such that $|px - qy| \leqslant \frac{1}{10} PH$ and $\|q\theta_x - p\theta_y\| \leqslant \frac{P}{10H}$ is $\gg \delta^7 |S||\mathcal{P}|^2$. Defining $Y := X/H$, $A := \frac{1}{H} S \subset [Y, 2Y)$ and $\alpha_x := \theta_{Hx}$ for $x \in A$, we see that $\mathcal{A} := (A, \alpha_\bullet, H/P)$ is a configuration satisfying $|Q(\mathcal{A})| \gg \delta^7 |A||\mathcal{P}|^2$.

We now apply Theorem 3.8.6 to $\mathcal{A}$. Again, note that the required lower bound on P is satisfied, using (3.44) and choosing C_1 sufficiently large. Thus, by Theorem 3.8.6, there exists a subset $A' \subseteq A$ of size $|A'| \gg \delta^{O(1)} |A|$, such that, for each $x \in A'$, we have the approximate formula

$$\alpha_x = \frac{a_0}{q_0} + \frac{T_0}{x} + O\left(\frac{\delta^{O(1)} (\log X)^4}{H} \right) \pmod 1$$

where $(a_0, q_0) = 1$, $q_0 \ll \delta^{-O(1)}$ and $|T_0| \ll \delta^{-O(1)} (\log X)^3 Y^2/H$.

We conclude by applying Theorem C.5.1. In the present setting, the condition (C.24) takes the form $(\log X)^{\delta^{-C}} \leqslant H \leqslant X^{\delta^C}$ for some large constant $C > 0$, which holds by definition of H and δ, for X sufficiently large. Hence, Theorem C.5.1 yields

$$\sum_{p \leqslant X} \frac{1 + \mathrm{Re}(\chi(p) p^{it})}{p} \ll \log(\delta^{-1})$$

for some Dirichlet character χ of conductor $\ll \delta^{-O(1)} \ll (\log\log X)^{O(1)}$ and some real number t with $|t| \ll (\delta^{-1}\log X)^{O(1)} X^2/H^2 \ll X^2$. This contradicts classical estimates for twisted character sums under GRH (see e.g. [42, Chapter 13]) when X is sufficiently large. $\square$

Chapter 4

Improved quantum algorithms

4.1 Structure of the chapter

This final chapter is devoted to the proof of a variant of Regev's conjecture (Theorem 4.3.18), and to the applications of that result to quantum algorithms for factoring and computing discrete logarithms. The chapter is organised as follows.

We begin in Section 4.2 with a proof of a simple special case of our main number-theoretic result. This is intended to illustrate our overall strategy and motivate the need for the more technical approach required in the general case. The purpose of this section is to build intuition; it is not formally required for the main argument.

Our main technical result is Theorem 4.3.18, which shows that lattices $\mathcal{L}$ similar to that in Corollary 1.4.5 have a basis of short vectors with high probability. Using simple geometry of numbers (see Section 4.3.4), we reduce this problem to estimating the number of lattice points in balls of growing radii. Unfortunately, we are unable to obtain a suitable lattice point count for $\mathcal{L}$ directly. We resolve this by considering a different lattice $\mathcal{L}_M$ from the start of the argument (using the lemmas in Section 4.3.1). In Section 4.3.2, we expand the lattice point count for $\mathcal{L}_M$ in terms of Dirichlet characters modulo N. This produces a main term, which can be estimated precisely, and an error term. The heart of the proof lies in using a zero-density estimate for Dirichlet characters modulo N to bound this error term unconditionally. Finally, we prove our quantum algorithmic applications (Theorems 1.4.1 and 1.4.2) in Section 4.4.

In this chapter, random variables are typically written in bold font. An event occurs *with high probability* if it holds with probability tending to 1 as $N \to \infty$, where N is the modulus we are considering. A non-trivial square root of 1 modulo N is an element $a \in (\mathbb{Z}/N\mathbb{Z})^\times$ such that $a^2 \equiv 1 \pmod{N}$ and $a \not\equiv \pm 1 \pmod{N}$.

We refer the reader to [44] for a detailed textbook on quantum computing, and [6, Chapter 2] for a brief overview.

4.2 An illustrative special case

In this section, we prove a weak version of our main theorem in a simplified setting. This special case can be used to obtain an unconditional factoring algorithm for integers N of a special form: RSA moduli that are products of two "safe primes".[1]

Definition 4.2.1 (RSA moduli with safe primes)**.** Let $\mathcal{M}_{\mathrm{RSA}}$ be the set of integers N of the form $N = P \cdot Q$, where $P, Q \geqslant N^{1/4}$ are distinct primes such that $(P-1)/2$ and $(Q-1)/2$ are also prime.

For $N \in \mathcal{M}_{\mathrm{RSA}}$, the multiplicative group $(\mathbb{Z}/N\mathbb{Z})^\times$ has a very convenient structure, allowing us to give a relatively short proof of the following proposition.

Proposition 4.2.2. *Let $N \in \mathcal{M}_{\mathrm{RSA}}$ be a sufficiently large integer. Let $d := \lceil \sqrt{\log N} \rceil$ and $X := d^d$.*

Let $\mathbf{b}_1, \dots, \mathbf{b}_d$ be i.i.d. random variables, each uniformly distributed in the set of primes less than X. Let $\mathbf{b}_0$ be a random variable uniformly distributed in $(\mathbb{Z}/N\mathbb{Z})^\times$, independent of the $\mathbf{b}_i$'s.

Let $x \in (\mathbb{Z}/N\mathbb{Z})^\times$. Then, with probability $\asymp 1$, there are nonnegative integers $e_0, \dots, e_d \leqslant e^{O(d)}$ such that

$$\prod_{i=0}^{d} \mathbf{b}_i^{e_i} \equiv x \pmod{N}.$$

Applying Proposition 4.2.2 (with x being a non-trivial square root of 1 modulo N) shows that Regev's factoring algorithm, with random parameters $\mathbf{b}_0, \dots, \mathbf{b}_d$ as in that proposition[2], admits an unconditional proof of correctness for integers $N \in \mathcal{M}_{\mathrm{RSA}}$.

It should be noted that Theorem 4.3.18 not only removes the restriction $N \in \mathcal{M}_{\mathrm{RSA}}$, it also provides a stronger geometric conclusion required for applications to the discrete logarithm problem.

Notation 4.2.3. For the rest of this section, we fix N, d, X, $\mathbf{b}_0, \dots, \mathbf{b}_d$ and x as in the statement of Proposition 4.2.2. In particular, $N \in \mathcal{M}_{\mathrm{RSA}}$ is assumed to be sufficiently large. Moreover, there are primes P, Q, P' and Q' such that $N = P \cdot Q$, $P = 2P' + 1$ and $Q = 2Q' + 1$. Hence,

$$G := (\mathbb{Z}/N\mathbb{Z})^\times \cong (\mathbb{Z}/2\mathbb{Z})^2 \times \mathbb{Z}/P'\mathbb{Z} \times \mathbb{Z}/Q'\mathbb{Z} \cong \widehat{G}. \tag{4.1}$$

To prove Proposition 4.2.2, we require the following lemma from analytic number theory, that ultimately relies on a zero-density estimate for Dirichlet L-functions.

[1] Note that this restrictive assumption is generally not satisfied by RSA moduli used in practice.

[2] Again, the presence of one potentially large parameter $\mathbf{b}_0$ circumvents the subgroup obstructions discussed in the introduction, while not significantly affecting the efficiency of the quantum algorithm.

Lemma 4.2.4. *For every $j \geqslant 0$, the number of characters $\psi \in \widehat{G}$ such that*

$$\Big|\mathop{\mathbb{E}}_{p\leqslant X} \psi(p)\Big| \geqslant e^{-j}$$

is at most $e^{O(d)}N^{O(j/\log X)}$. Here, $\mathbb{E}_{p\leqslant X}$ denotes the average over primes $p \leqslant X$.

We omit the proof of Lemma 4.2.4, as it is somewhat tangential to the aim of this explanatory section.[3]

Proof of Proposition 4.2.2. Let $H = \lfloor e^{Cd} \rfloor$, where C is a large absolute constant to be chosen later.

We will show that, with probability $\asymp 1$, there exist integers $0 \leqslant h_1, \dots, h_d < H$ such that

$$x \equiv \mathbf{b}_0 \mathbf{b}_1^{h_1} \cdots \mathbf{b}_d^{h_d} \pmod N,$$

which clearly implies the desired conclusion. By orthogonality of characters modulo N, we have

$$\sum_{0\leqslant h_1,\dots,h_d<H} \mathbf{1}_{\mathbf{b}_0\mathbf{b}_1^{h_1}\cdots\mathbf{b}_d^{h_d}\equiv x \,(\mathrm{mod}\, N)} = \frac{1}{\varphi(N)} \sum_{\chi\in\widehat{G}} \chi(\mathbf{b}_0 x^{-1}) \prod_{i=1}^{d} \sum_{0\leqslant h<H} \chi^h(\mathbf{b}_i). \tag{4.2}$$

Let χ_0 be the trivial character, and χ_1, χ_2, χ_3 be the three characters of order 2 in $\widehat{G}$. Each of these four characters χ_j satisfies $\chi_j(\mathbf{b}_i) \in \{\pm 1\}$ for all i, so the term $\sum_{0\leqslant h<H} \chi_j^h(\mathbf{b}_i)$ is either equal to H or lies in $\{0, 1\}$. Hence, the contribution of these four characters to (4.2) is

$$\frac{H^d}{\varphi(N)} + \frac{1}{\varphi(N)} \sum_{j=1}^{3} \chi_j(\mathbf{b}_0 x^{-1})\, C_j(\mathbf{b}_1, \dots, \mathbf{b}_d) \tag{4.3}$$

for some quantities $C_j(\mathbf{b}_1, \dots, \mathbf{b}_d) \geqslant 0$. Since $\mathbf{b}_0$ is uniformly distributed in G, by (4.1) we have

$$\mathbb{P}\big(\chi_1(\mathbf{b}_0 x^{-1}) = \chi_2(\mathbf{b}_0 x^{-1}) = \chi_3(\mathbf{b}_0 x^{-1}) = 1\big) = \frac{1}{4},$$

as this is the probability that $\mathbf{b}_0 x^{-1}$ lies in the subgroup of squares modulo N, which has index 4 in G. Therefore, with probability $\geqslant 1/4$, the expression (4.3) is bounded below by $H^d/\varphi(N)$.

As a result, Proposition 4.2.2 reduces to showing that the contribution of the remaining characters $\chi \in \widehat{G} \setminus \{\chi_0, \chi_1, \chi_2, \chi_3\}$ to (4.2) is, with high probability, at most $H^d/(2\varphi(N))$ in absolute value. By Chebyshev's inequality, it suffices to prove that

$$\mathbb{E}\big[|\mathbf{z}|^2\big] = o\big(H^{2d}\big), \tag{4.4}$$

[3] This lemma is a consequence of Propositions 4.3.9 and 4.3.10; see Section 4.3.3 and [48, Appendix A] for further details.

where $\mathbf{z}$ is the random variable defined by

$$\mathbf{z} := \sum_{\substack{\chi \in \widehat{G} \\ \mathrm{ord}(\chi) > 2}} \left| \prod_{i=1}^{d} \sum_{0 \leqslant h < H} \chi^h(\mathbf{b}_i) \right|.$$

By the triangle inequality for the L^2-norm and the definition of the random variables $\mathbf{b}_i$, we have

$$\mathbb{E}\big[|\mathbf{z}|^2\big]^{1/2} \leqslant \sum_{\substack{\chi \in \widehat{G} \\ \mathrm{ord}(\chi) > 2}} \left(\mathbb{E}\left[\left| \prod_{i=1}^{d} \sum_{0 \leqslant h < H} \chi^h(\mathbf{b}_i) \right|^2 \right] \right)^{1/2} = \sum_{\substack{\chi \in \widehat{G} \\ \mathrm{ord}(\chi) > 2}} \left(\mathop{\mathbb{E}}_{p \leqslant X} \left| \sum_{0 \leqslant h < H} \chi^h(p) \right|^2 \right)^{d/2}.$$

Expanding the square and exchanging the order of summation, we obtain

$$\mathop{\mathbb{E}}_{p \leqslant X} \left| \sum_{0 \leqslant h < H} \chi^h(p) \right|^2 \leqslant \sum_{0 \leqslant h_1, h_2 < H} \left| \mathop{\mathbb{E}}_{p \leqslant X} \chi^{h_1 - h_2}(p) \right| \leqslant H \sum_{|h| < H} \left| \mathop{\mathbb{E}}_{p \leqslant X} \chi^h(p) \right|.$$

Therefore,

$$\mathbb{E}\big[|\mathbf{z}|^2\big]^{1/2} \leqslant H^{d/2} \sum_{\substack{\chi \in \widehat{G} \\ \mathrm{ord}(\chi) > 2}} \left(\sum_{|h| < H} \left| \mathop{\mathbb{E}}_{p \leqslant X} \chi^h(p) \right| \right)^{d/2}. \tag{4.5}$$

To estimate the right-hand side of (4.5), we partition the set of characters as follows. For $j \geqslant 0$, let

$$E_j := \left\{ \chi \in \widehat{G} \setminus \{\chi_0, \chi_1, \chi_2, \chi_3\} \ : \ \max_{0 < |h| < H} \left| \mathop{\mathbb{E}}_{p \leqslant X} \chi^h(p) \right| \in \left(e^{-j-1}, e^{-j} \right] \right\}.$$

Notice that, by (4.1) and Definition 4.2.1, every character χ of order greater than 2 has order at least $\min(P', Q') \gg N^{1/4}$. Hence, provided N is sufficiently large, for any χ of order greater than 2, the characters χ^h for $0 < |h| < H$ are all distinct. In addition, for any fixed $\psi \in \widehat{G} \setminus \{\chi_0, \chi_1, \chi_2, \chi_3\}$ and $0 < |h| < H$, there are at most four characters χ such that $\chi^h = \psi$. By Lemma 4.2.4, these observations imply that

$$|E_j| \ll H e^{O(d)} N^{O(j/\log X)} \ll H e^{O(d)} e^{O(jd/\log d)} \tag{4.6}$$

for all $j \geqslant 0$.[4]

Let $j \geqslant 0$ and let $\chi \in E_j$. By Lemma 4.2.4, there are at most $e^{O(d)}$ integers $0 \leqslant |h| < H$ such that $\left| \mathbb{E}_{p \leqslant X} \chi^h(p) \right| > d^{-2}$. By definition of E_j, all other integers $0 < |h| < H$ satisfy

$$\left| \mathop{\mathbb{E}}_{p \leqslant X} \chi^h(p) \right| \leqslant \min(d^{-2}, e^{-j}).$$

[4]It is here that we crucially relied on the particular structure of $(\mathbb{Z}/N\mathbb{Z})^\times$ for $N \in \mathcal{M}_{\mathrm{RSA}}$. These arguments can be extended somewhat, but additional ideas are required to handle the general case (the most difficult case being when N has many prime factors).

Hence,

$$\Big(\sum_{|h|<H}\Big|\mathop{\mathbb{E}}_{p\leqslant X}\chi^h(p)\Big|\Big)^{d/2} \leqslant \big(e^{O(d)}+2H\min(d^{-2},e^{-j})\big)^{d/2}$$
$$\leqslant e^{O(d^2)}+e^{O(d)}H^{d/2}\min(d^{-d},e^{-jd/2}).$$

Summing this over all $\chi\in E_j$ and $j\geqslant 0$, we can bound (4.5) by

$$\mathbb{E}\big[|\mathbf{z}|^2\big]^{1/2}\leqslant e^{O(d^2)}\varphi(N)H^{d/2}+e^{O(d)}H^d\sum_{j\geqslant 0}|E_j|\min(d^{-d},e^{-jd/2}).$$

Note that $\varphi(N)\leqslant N\leqslant e^{d^2}$. Bounding $|E_j|$ using (4.6) and summing the resulting geometric series (using that N is sufficiently large), we obtain

$$\mathbb{E}\big[|\mathbf{z}|^2\big]^{1/2}\leqslant \big(e^{O(d^2)}H^{-d/2}+e^{O(d)}Hd^{-d}\big)H^d.$$

Recalling that $H=\lfloor e^{Cd}\rfloor$ we see that, if C is a sufficiently large absolute constant, the right-hand side is $o(H^d)$ as $N\to\infty$. This establishes (4.4) and concludes the proof of Proposition 4.2.2. □

Remark 4.2.5. For Theorem 4.3.18, we obtain (with high probability) an asymptotic formula for quantities similar to (4.2), which is then combined with techniques from geometry of numbers to prove the existence of a short lattice basis.

This requires a more careful analysis of the small-order characters, which may be more numerous than in the special case treated above. Moreover, the analysis of large-order characters is much more delicate, as an exceptional character ψ (in the sense of Lemma 4.2.4) may have more representations of the form $\psi=\chi^h$ (for some $\chi\in\widehat{G}$ and $|h|<H$) than the previous argument can handle. This issue is resolved by performing Fourier analysis on a well-chosen subgroup of $(\mathbb{Z}/N\mathbb{Z})^\times$, introduced in Section 4.3.1.

4.3 Proof of the existence of a short lattice basis

4.3.1 Restricting to a convenient subgroup of $\widehat{(\mathbb{Z}/N\mathbb{Z})^\times}$

The goal of this preliminary section is to define a subgroup of the character group of $(\mathbb{Z}/N\mathbb{Z})^\times$ that does not have too many elements of small order. This will be crucial for Section 4.3.3, to reduce the influence of potential counterexamples to the Generalised Riemann Hypothesis. For example, we need to avoid having many characters $\chi_1,\dots,\chi_k$ modulo N such that $\chi_i^2=\psi$ for all i, where ψ is an exceptional character (in the sense that $L(s,\psi)$ has a zero very close to the line $\operatorname{Re}s=1$). The definition of the suitable subgroup depends on the precise structure of $(\mathbb{Z}/N\mathbb{Z})^\times$.

Notation 4.3.1. If G is a finite abelian group (written multiplicatively) and $M \geqslant 1$, we write $G^M := \{x^M : x \in G\}$ for the subgroup of Mth powers in G.[5]

Lemma 4.3.2. *Let G be a finite abelian group (written multiplicatively) and $M \geqslant 1$. There is an isomorphism*

$$\iota : \widehat{G^M} \xrightarrow{\sim} \widehat{G}^M$$

such that, for any $\chi \in \widehat{G^M}$, $\iota(\chi) = \widetilde{\chi}^M$ where $\widetilde{\chi} \in \widehat{G}$ is an arbitrary extension of χ to the full group G.

Proof. Pontryagin duality for finite abelian groups is an equivalence of categories, and therefore an exact functor. In simple terms, this implies that injections turn into surjections when switching to the dual, and vice versa. Hence, the Mth power map $G \twoheadrightarrow G^M$, $g \mapsto g^M$, induces an injection $\iota : \widehat{G^M} \hookrightarrow \widehat{G}$ defined by

$$\iota(\chi)(g) := \chi\big(g^M\big)$$

for every $\chi \in \widehat{G^M}$ and $g \in G$. Moreover, the restriction map $\widehat{G} \to \widehat{G^M}$ is surjective, using exactness of Pontryagin duality again. Thus, every $\chi \in \widehat{G^M}$ can be extended to a character $\widetilde{\chi}$ on G, and

$$\chi\big(g^M\big) = \widetilde{\chi}(g^M) = \widetilde{\chi}^M(g).$$

This implies that the image of ι is contained in $\widehat{G}^M$. Since $\big|\widehat{G^M}\big| = \big|G^M\big| = \big|\widehat{G}^M\big|$, the lemma follows. □

Definition 4.3.3. Let $N \geqslant 1$. For every $h \geqslant 1$ we define

$$K(h) := \frac{|(\mathbb{Z}/N\mathbb{Z})^\times|}{|((\mathbb{Z}/N\mathbb{Z})^\times)^h|}.$$

In other words, $K(h)$ is the size of the kernel of the hth power map in $(\mathbb{Z}/N\mathbb{Z})^\times$.

Lemma 4.3.4. *Let $N > 2$ be an integer and $d := \lceil \sqrt{\log N} \rceil$.*

For every prime p, define m_p to be the largest non-negative integer such that

$$K(p^{m_p}) \geqslant p^{dm_p/10}.$$

Let $M_* = M_*(N) := \prod_p p^{m_p}$. Then the following holds.

1. *We have $M_* \leqslant e^{10d}$.*
2. *For every $h \geqslant 1$, we have $K(M_* h)/K(M_*) \leqslant h^{d/10}$.*

[5] In particular, G^M is *not* the M-fold direct product of G.

Proof. 1. For every prime p, since $K(p^{m_p})$ is a power of p dividing $|(\mathbb{Z}/N\mathbb{Z})^\times|$, we have

$$\prod_p K(p^{m_p}) \mid \varphi(N).$$

By definition of m_p, we know that $K(p^{m_p}) \geqslant p^{dm_p/10}$ for every p, and thus

$$M_*^{d/10} = \prod_p p^{dm_p/10} \leqslant \prod_p K(p^{m_p}) \leqslant \varphi(N) \leqslant N.$$

Recalling that $d = \lceil\sqrt{\log N}\rceil$, it follows that $M_* \leqslant e^{10d}$.

2. In any finite abelian group (written multiplicatively) and for any coprime integers a and b, there is an isomorphism

$$\ker\big(g \mapsto g^a\big) \times \ker\big(g \mapsto g^b\big) \cong \ker\big(g \mapsto g^{ab}\big).$$

Thus, if $h = \prod_p p^{e_p}$ is the prime factorisation of h, we obtain the identities $K(M_*) = \prod_p K(p^{m_p})$ and $K(M_*h) = \prod_p K(p^{m_p+e_p})$, where all products are finite. Whenever $e_p \geqslant 1$, we have

$$K(p^{m_p+e_p}) < p^{d(m_p+e_p)/10}, \quad K(p^{m_p}) \geqslant p^{dm_p/10}$$

by definition of m_p. Thus $K(p^{m_p+e_p})/K(p^{m_p}) \leqslant p^{de_p/10}$, and this last inequality also holds when $e_p = 0$. We conclude that

$$\frac{K(M_*h)}{K(M_*)} = \prod_p \frac{K(p^{m_p+e_p})}{K(p^{m_p})} \leqslant \prod_p p^{de_p/10} = h^{d/10}$$

as claimed. □

Lemma 4.3.5. *Let $N > 2$, $d := \lceil\sqrt{\log N}\rceil$, $G := (\mathbb{Z}/N\mathbb{Z})^\times$ and let $M_* \geqslant 1$ be the integer defined in the statement of Lemma 4.3.4.*

Let $h \geqslant 1$. Let $\chi_1, \ldots, \chi_n \in \widehat{G^{M_*}}$ be distinct characters such that

$$\chi_1^h = \chi_2^h = \ldots = \chi_n^h.$$

Then $n \leqslant h^{d/10}$.

Proof. Let $f_h : G^{M_*} \to G^{M_*}$ be the hth power map, $f_h(\chi) := \chi^h$. By assumption, the n distinct characters $\chi_1\chi_n^{-1}, \chi_2\chi_n^{-1}, \ldots, \chi_n\chi_n^{-1}$ lie in the kernel of f_h. Hence,

$$n \leqslant \big|\ker(f_h)\big| = \frac{|G^{M_*}|}{|(G^{M_*})^h|} = \frac{|((\mathbb{Z}/N\mathbb{Z})^\times)^{M_*}|}{|((\mathbb{Z}/N\mathbb{Z})^\times)^{M_*h}|} = \frac{K(M_*h)}{K(M_*)},$$

which is $\leqslant h^{d/10}$ by Lemma 4.3.4. □

4.3.2 Lattice point counting via characters

To prove Theorem 4.3.18, we will need to study the following quantities.

Definition 4.3.6. Let $N \geqslant 1$, $G = (\mathbb{Z}/N\mathbb{Z})^\times$ and $\chi \in \widehat{G}$. For all $d \geqslant 1$ and $b_1, \dots, b_d \in G$, we define the quantity

$$F_{\chi,H}(b_1, \dots, b_d) := \prod_{i=1}^{d} \sum_{|h| \leqslant H} \chi^h(b_i).$$

The following lemma relates these expressions to a lattice point counting problem.

Lemma 4.3.7. *Let $N, M, d, H \geqslant 1$ be integers. Let $b_1, \dots, b_d \in G := (\mathbb{Z}/N\mathbb{Z})^\times$. The number of vectors $(e_1, \dots, e_d) \in \mathbb{Z}^d \cap [-H, H]^d$ such that*

$$\prod_{i=1}^{d} b_i^{Me_i} \equiv 1 \pmod{N}$$

is equal to

$$\frac{1}{|\widehat{G}^M|} \sum_{\chi \in \widehat{G}^M} F_{\chi,H}(b_1, \dots, b_d). \tag{4.7}$$

Proof. By orthogonality of characters in the abelian group G^M, we have

$$\mathbf{1}_{\prod_i b_i^{Me_i} \equiv 1 \,(\mathrm{mod}\, N)} = \frac{1}{|\widehat{G^M}|} \sum_{\psi \in \widehat{G^M}} \psi\left(\prod_{i=1}^{d} b_i^{Me_i}\right) = \frac{1}{|\widehat{G^M}|} \sum_{\psi \in \widehat{G^M}} \prod_{i=1}^{d} \psi\big(b_i^{Me_i}\big).$$

Observe that $\psi\big(b_i^{Me_i}\big) = \iota(\psi)(b_i^{e_i})$, where ι is the isomorphism $\widehat{G^M} \xrightarrow{\sim} \widehat{G}^M$ defined in Lemma 4.3.2. We may thus rewrite this equality as

$$\mathbf{1}_{\prod_i b_i^{Me_i} \equiv 1 \,(\mathrm{mod}\, N)} = \frac{1}{|\widehat{G}^M|} \sum_{\chi \in \widehat{G}^M} \prod_{i=1}^{d} \chi\big(b_i^{e_i}\big).$$

Therefore, the number of vectors $(e_1, \dots, e_d) \in \mathbb{Z}^d \cap [-H, H]^d$ satisfying the congruence $\prod_{i=1}^{d} b_i^{Me_i} \equiv 1 \pmod{N}$ is

$$\frac{1}{|\widehat{G}^M|} \sum_{\chi \in \widehat{G}^M} \sum_{\substack{(e_1,\dots,e_d) \in \mathbb{Z}^d \\ \max_i |e_i| \leqslant H}} \prod_{i=1}^{d} \chi^{e_i}(b_i) = \frac{1}{|\widehat{G}^M|} \sum_{\chi \in \widehat{G}^M} F_{\chi,H}(b_1, \dots, b_d),$$

as claimed. □

It is fairly straightforward to estimate $F_{\chi,H}(b_1, \dots, b_d)$ when χ is a character of small order in $\widehat{G}$. The contribution of these small-order characters gives the expected main term for (4.7).

Lemma 4.3.8 (Main term). *Let $N, M, d \geqslant 1$ be integers. Let $b_1, \dots, b_d \in G$, where $G := (\mathbb{Z}/N\mathbb{Z})^\times$. Uniformly for all integers $H \geqslant e^{31d}$, we have*

$$\frac{1}{\left|\widehat{G}^M\right|} \sum_{\chi \in \widehat{G}^M} F_{\chi,H}(\underline{b}) = \left(1 + O\left(\frac{e^{31d}}{H}\right)\right) \frac{(2H+1)^d}{\left|\langle b_1^M, \dots, b_d^M \rangle\right|} + O\left(\frac{1}{\left|\widehat{G}^M\right|} \sum_{\substack{\chi \in \widehat{G}^M \\ \mathrm{ord}(\chi) \geqslant e^{10d}}} |F_{\chi,H}(\underline{b})|\right), \tag{4.8}$$

where $F_{\chi,H}(\underline{b})$ is short for $F_{\chi,H}(b_1, \dots, b_d)$.

Proof. Define

$$\langle b_1, \dots, b_d \rangle^\perp := \{\chi \in \widehat{G} \ : \ \forall i \in [d],\ \chi(b_i) = 1\}.$$

Observe that

$$\left|\langle b_1, \dots, b_d \rangle^\perp \cap \widehat{G}^M\right| = \left|\left\{\chi \in \widehat{G^M} \ : \ \forall i \in [d],\ \chi\left(b_i^M\right) = 1\right\}\right| = \frac{\left|G^M\right|}{\left|\langle b_1^M, \dots, b_d^M \rangle\right|} \tag{4.9}$$

where the first equality follows from Lemma 4.3.2 and the second from the canonical isomorphism $H_1^\perp \cong \widehat{G_1/H_1}$ for any subgroup H_1 of a finite abelian group G_1.

Clearly, if $\chi \in \langle b_1, \dots, b_d \rangle^\perp \cap \widehat{G}^M$, we have $F_{\chi,H}(\underline{b}) = (2H+1)^d$. Hence, by (4.9), those characters contribute

$$\frac{1}{\left|\widehat{G}^M\right|} \sum_{\chi \in \langle b_1, \dots, b_d \rangle^\perp \cap \widehat{G}^M} F_{\chi,H}(\underline{b}) = \frac{(2H+1)^d}{\left|\langle b_1^M, \dots, b_d^M \rangle\right|}$$

to the sum (4.8).

For any character $\chi \in \widehat{G}^M \setminus \langle b_1, \dots, b_d \rangle^\perp$ of order $\leqslant e^{10d}$, and any $i \in [d]$ such that $\chi(b_i) \neq 1$, we can bound

$$\left|\sum_{|h| \leqslant H} \chi(b_i)^h\right| \leqslant \mathrm{ord}(\chi) \leqslant e^{10d}$$

by the geometric series formula. Thus, defining

$$I_\chi := \{i \in [d] : \chi(b_i) \neq 1\},$$

we have

$$|F_{\chi,H}(\underline{b})| \leqslant e^{10d|I_\chi|}(2H+1)^{d-|I_\chi|}.$$

Consequently,

$$\sum_{\substack{\chi \in \widehat{G}^M \setminus \langle b_1, \dots, b_d \rangle^\perp \\ \mathrm{ord}(\chi) \leqslant e^{10d}}} |F_{\chi,H}(\underline{b})| \ll \sum_{\substack{I \subset [d] \\ I \neq \emptyset}} e^{10d|I|}(2H+1)^{d-|I|} \sum_{\substack{\chi \in \widehat{G}^M \\ \mathrm{ord}(\chi) \leqslant e^{10d} \\ I_\chi = I}} 1. \tag{4.10}$$

Fix some non-empty set $I \subset [d]$ and complex numbers $(z_i)_{i\in I}$. Let $C_{I,(z_i)}$ be the set of all characters $\chi \in \widehat{G}^M$ such that

$$\chi(b_i) = \begin{cases} z_i & \text{if } i \in I \\ 1 & \text{if } i \in [d] \setminus I. \end{cases}$$

Note that $C_{I,(z_i)}$ is either the empty set, or a coset of $\langle b_1, \dots, b_d\rangle^\perp \cap \widehat{G}^M$ in $\widehat{G}^M$. Moreover, if χ has order at most e^{10d}, this set $C_{I,(z_i)}$ can only be non-empty if all z_i are roots of unity of order $\leqslant e^{10d}$, and there are $\leqslant e^{20d}$ such roots of unity. We conclude that

$$\sum_{\substack{\chi\in\widehat{G}^M \\ \mathrm{ord}(\chi)\leqslant e^{10d} \\ I_\chi = I}} 1 \leqslant \sum_{(z_i)_{i\in I}} \big|C_{I,(z_i)}\big| \leqslant e^{20d|I|}\big|\langle b_1, \dots, b_d\rangle^\perp \cap \widehat{G}^M\big|.$$

Thus, we can bound the right-hand side of (4.10) by

$$\ll (2H+1)^d\big|\langle b_1, \dots, b_d\rangle^\perp \cap \widehat{G}^M\big| \sum_{\substack{I\subset[d] \\ I\neq\emptyset}} \big(e^{30d}H^{-1}\big)^{|I|}.$$

Since $H \geqslant e^{30d}$, the last sum is $\ll 2^d e^{30d} H^{-1}$. By (4.9), this means that

$$\frac{1}{\big|\widehat{G}^M\big|} \sum_{\substack{\chi\in\widehat{G}^M\setminus\langle b_1,\dots,b_d\rangle^\perp \\ \mathrm{ord}(\chi)\leqslant e^{10d}}} |F_{\chi,H}(\underline{b})| \ll e^{31d}H^{-1}\frac{(2H+1)^d}{\big|\langle b_1^M, \dots, b_d^M\rangle\big|},$$

which completes the proof. □

4.3.3 Bounding the contribution of large-order characters

In this section, we bound the error term coming from large-order characters, on average over primes $b_1, \dots, b_d$ in a short interval.

We will need the following ingredients from classical analytic number theory.

Proposition 4.3.9 (Character sums over primes)**.** *Let $1/2 \leqslant \alpha \leqslant 1$. Let $q, x \geqslant 2$. Let χ be a non-principal character modulo q whose Dirichlet L-function $L(s,\chi)$ has no zero in the rectangle*

$$\big\{s \in \mathbb{C} \ : \ \alpha < \mathrm{Re}\, s \leqslant 1, \ |\mathrm{Im}\, s| \leqslant x^{1-\alpha}\big\}.$$

Then

$$\frac{1}{\pi(x)}\sum_{p\leqslant x}\chi(p) \ll x^{-(1-\alpha)}\log^3(qx).$$

Proposition 4.3.9 is a standard consequence of the explicit formula for $L(s,\chi)$; the details can be found in [48, Appendix A]. The logarithmic factors can be somewhat improved, but such refinements are irrelevant here.

The Generalised Riemann Hypothesis is the claim that, for every Dirichlet character χ, $L(s,\chi)$ has no zero ρ with $\frac{1}{2} < \operatorname{Re}\rho < 1$. Assuming GRH, Proposition 4.3.9 thus implies almost square-root cancellation for character sums over primes. Proposition 4.3.10 will serve as an unconditional substitute for GRH.

Proposition 4.3.10 (Zero-density estimate for a fixed modulus)**.** *Uniformly for $\frac{4}{5} \leqslant \alpha \leqslant 1$ and $q, T \geqslant 1$, we have*

$$\sum_{\chi \,(\mathrm{mod}\, q)} \mathcal{N}(\alpha, T, \chi) \ll_\varepsilon (qT)^{(2+\varepsilon)(1-\alpha)},$$

where $\mathcal{N}(\alpha,T,\chi)$ denotes the number of zeros (with multiplicity) of $L(s,\chi)$ in the rectangle

$$\{s \in \mathbb{C} \ : \ \alpha < \operatorname{Re} s \leqslant 1, \ |\operatorname{Im} s| \leqslant T\}.$$

Proof. This is [31, Theorem 1 (1.7)]. □

Our goal is to control the total contribution of all large-order characters in (4.8). We will do so in Proposition 4.3.14 (restricting to a suitable subgroup of $\widehat{G}$). We first prove the following L^2 bound for a single large-order character.

Lemma 4.3.11. *Let $N > 2$ be an integer and let $G = (\mathbb{Z}/N\mathbb{Z})^\times$. Let $d = \lceil \sqrt{\log N} \rceil$ and $X = d^{10^3 d}$.*

Let $\mathbf{b}_1, \dots, \mathbf{b}_d$ be i.i.d. random variables, each uniformly distributed in the set of primes less than X not dividing N. Let $\chi \in \widehat{G}$ be a character of order $\geqslant e^{10d}$. Then, for every $H \geqslant e^{10d}$,

$$\mathbb{E}\big[|F_{\chi,H}(\mathbf{b}_1, \dots, \mathbf{b}_d)|^2\big] \ll d^{-10d} H^{2d}.$$

Proof. By Proposition 4.3.10 (the zero-density estimate), the number of Dirichlet characters modulo N whose associated L-function has a zero in the region

$$\left\{ s \in \mathbb{C} \ : \ 1 - \frac{1}{10d} < \operatorname{Re} s \leqslant 1, \ |\operatorname{Im} s| \leqslant X \right\} \tag{4.11}$$

is bounded by

$$\ll (NX)^{(2+1)/(10d)} \ll e^d. \tag{4.12}$$

If a non-principal character ψ modulo N has no zero in the region (4.11), then by Proposition 4.3.9 we have

$$\mathbb{E}[\psi(\mathbf{b}_1)] \ll X^{-1/(10d)} (\log(NX))^3 \ll d^{-100} d^6 \ll d^{-94}. \tag{4.13}$$

We now expand $\mathbb{E}[|F_{\chi,H}(\mathbf{b}_1,\dots,\mathbf{b}_d)|^2]$ as

$$\begin{aligned}\mathbb{E}\left[\left|\prod_{i=1}^{d}\sum_{|h|\leqslant H}\chi^h(\mathbf{b}_i)\right|^2\right] &= \prod_{i=1}^{d}\mathbb{E}\left[\left|\sum_{|h|\leqslant H}\chi^h(\mathbf{b}_i)\right|^2\right]\\ &= \left(\sum_{|h_1|\leqslant H}\sum_{|h_2|\leqslant H}\mathbb{E}\big[\chi^{h_1-h_2}(\mathbf{b}_1)\big]\right)^d\\ &\leqslant (2H+1)^d\left(\sum_{|h|\leqslant 2H}\big|\,\mathbb{E}\big[\chi^h(\mathbf{b}_1)\big]\,\big|\right)^d.\end{aligned}$$

We can use (4.13) to bound the term $\mathbb{E}\big[\chi^h(\mathbf{b}_1)\big]$, unless χ^h is principal or $L(s,\chi^h)$ has a zero in the region (4.11). By (4.12), the number of values of h with $|h|\leqslant 2H$ for which one of these two situations occurs is

$$\ll \left(1+\frac{H}{\mathrm{ord}(\chi)}\right)e^d \ll e^{-10d}e^dH \ll e^{-d}H.$$

By (4.13), we deduce that, for all but $O(e^{-d}H)$ values of $|h|\leqslant 2H$, the bound $\mathbb{E}\big[\chi^h(\mathbf{b}_1)\big]\ll d^{-94}$ holds. Hence

$$\begin{aligned}\mathbb{E}\big[|F_{\chi,H}(\mathbf{b}_1,\dots,\mathbf{b}_d)|^2\big] &\leqslant (2H+1)^d\Big(O(e^{-d}H)+O(d^{-94}H)\Big)^d\\ &\ll e^{O(d)}d^{-94d}H^{2d} \ll d^{-10d}H^{2d},\end{aligned}$$

as desired. □

Lemma 4.3.11 applies to all characters χ of order at least e^{10d}, but gives a relatively weak upper bound. In Lemma 4.3.13, we will prove that a stronger bound can be obtained if a small set of exceptions is allowed. We begin by describing the exceptional characters in the following lemma.

Lemma 4.3.12. *There is some absolute constant $t_0\geqslant 1$ such that the following holds.*

Let N,G,d,X and $\mathbf{b}_i$ be as in Lemma 4.3.11. Let $H\geqslant e^{10d}$.

Let $\chi\in\widehat{G}$ be a character of order $\geqslant e^{10d}$ such that

$$\mathbb{E}\big[|F_{\chi,H}(\mathbf{b}_1,\dots,\mathbf{b}_d)|^2\big] > e^{-2td}H^{2d}$$

for some t with $t_0\leqslant t\leqslant 2d$.

Then there are integers $-2H\leqslant h_1,h_2\leqslant 2H$ with $0<h_2-h_1\leqslant e^{3t}$ such that both $L(s,\chi^{h_1})$ and $L(s,\chi^{h_2})$ have a zero in the region

$$\left\{s\in\mathbb{C}\ :\ 1-\frac{t}{100d}<\mathrm{Re}\,s\leqslant 1,\ |\mathrm{Im}\,s|\leqslant X\right\}. \tag{4.14}$$

Proof. As in the proof of Lemma 4.3.11, we have

$$\mathbb{E}\big[|F_{\chi,H}(\mathbf{b}_1,\dots,\mathbf{b}_d)|^2\big] \leqslant (2H+1)^d \bigg(\sum_{|h|\leqslant 2H} \big|\mathbb{E}\big[\chi^h(\mathbf{b}_1)\big]\big|\bigg)^d. \tag{4.15}$$

Let I_1 be the set of all $-2H \leqslant h \leqslant 2H$ such that χ^h is principal. Since χ has order $\geqslant e^{10d}$, we have

$$|I_1| \ll 1 + e^{-10d}H \ll e^{-10d}H.$$

Let I_2 be the set of all $-2H \leqslant h \leqslant 2H$ such that $L(s,\chi^h)$ has a zero in the region (4.14). By contradiction, suppose that the conclusion of Lemma 4.3.12 does not hold. Then any sub-interval of $[-2H, 2H]$ of length e^{3t} contains at most one element of I_2. This implies that

$$|I_2| \ll 1 + e^{-3t}H \ll e^{-3t}H.$$

Moreover, for any integer $h \in [-2H, 2H]\setminus(I_1 \cup I_2)$, the character χ^h is non-principal and has no zero in the region (4.14), which by Proposition 4.3.9 implies that

$$\mathbb{E}\big[\chi^h(\mathbf{b}_1)\big] \ll X^{-t/(100d)}(\log(NX))^3 \ll d^{-10t}d^6 \ll e^{-3t}.$$

Therefore, we can bound

$$\sum_{|h|\leqslant 2H} \big|\mathbb{E}\big[\chi^h(\mathbf{b}_1)\big]\big| \ll |I_1| + |I_2| + e^{-3t}H \ll e^{-10d}H + e^{-3t}H \ll e^{-3t}H.$$

Hence, by (4.15) we get

$$\mathbb{E}\big[|F_{\chi,H}(\mathbf{b}_1,\dots,\mathbf{b}_d)|^2\big] \leqslant e^{O(d)}e^{-3td}H^{2d},$$

which contradicts the assumption in the statement if t_0 is chosen to be sufficiently large. □

We wish to use a zero-density estimate again to show that there are few characters satisfying the conclusion of Lemma 4.3.12. For this step to work, we need to restrict to the subgroup $\widehat{G}^{M_*}$ of the full character group $\widehat{G}$, where M_* is the integer defined in Lemma 4.3.4.

Lemma 4.3.13. *Let $t_0 \geqslant 1$ be the constant from Lemma 4.3.12. Let N, G, d, X and $\mathbf{b}_i$ be as in Lemma 4.3.11. Let $H \geqslant e^{10d}$ and let $M_* \geqslant 1$ be the integer defined in Lemma 4.3.4.*

For every $t \geqslant t_0$,

$$\Big|\big\{\chi \in \widehat{G}^{M_*} \;:\; \mathrm{ord}(\chi) \geqslant e^{10d},\ \mathbb{E}\big[|F_{\chi,H}(\mathbf{b}_1,\dots,\mathbf{b}_d)|^2\big] > e^{-2td}H^{2d}\big\}\Big| \ll e^{td/2}. \tag{4.16}$$

Proof. If $t \geqslant 2d$, the bound (4.16) is trivially satisfied as the right-hand side is $\gg N$. We may thus assume that $t_0 \leqslant t \leqslant 2d$, in which case Lemma 4.3.12 applies.

Let E_t be the set of all characters ψ modulo N such that $L(s,\psi)$ has a zero in the rectangle defined in (4.14). By Proposition 4.3.10, this set E_t has size

$$|E_t| \ll (NX)^{(2+1/2)t/(100d)} \ll e^{td/20}. \tag{4.17}$$

For every $\chi \in \widehat{G}$ of order $\geqslant e^{10d}$, if $\mathbb{E}[|F_{\chi,H}(\mathbf{b}_1,\dots,\mathbf{b}_d)|^2] > e^{-2td}H^{2d}$ then by Lemma 4.3.12 we can write

$$\chi^h = \psi_1\overline{\psi_2}$$

for some integer $0 < h < e^{3t}$ and some characters $\psi_1, \psi_2 \in E_t$. Hence, the left-hand side of (4.16) is

$$\leqslant \sum_{\psi_1,\psi_2\in E_t} \sum_{0<h<e^{3t}} \Big|\big\{\chi \in \widehat{G^{M_*}} \,:\, \chi^h = \psi_1\overline{\psi_2}\big\}\Big|. \tag{4.18}$$

Since $\widehat{G}^{M_*}$ and $\widehat{G^{M_*}}$ are isomorphic, Lemma 4.3.5 implies that

$$\Big|\big\{\chi \in \widehat{G}^{M_*} \,:\, \chi^h = \psi_1\overline{\psi_2}\big\}\Big| \leqslant h^{d/10} \leqslant e^{3td/10}. \tag{4.19}$$

Inserting (4.17) and (4.19) into (4.18), we conclude that the left-hand side of (4.16) is

$$\leqslant |E_t|^2 e^{3t} e^{3td/10} \leqslant e^{3t} e^{4td/10} \ll e^{td/2}$$

as claimed. □

Combining the previous lemmas, we obtain a suitable bound for the sum of second moments of $F_{\chi,H}(\mathbf{b}_1,\dots,\mathbf{b}_d)$ over all large-order characters in $\widehat{G}^{M_*}$.

Proposition 4.3.14 (Large-order characters)**.** *Let N, G, d, X and $\mathbf{b}_i$ be as defined in Lemma 4.3.11. Let $M_* \geqslant 1$ be the integer introduced in Lemma 4.3.4. For $H \geqslant e^{10d}$, we have*

$$\sum_{\substack{\chi\in\widehat{G}^{M_*}\\ \mathrm{ord}(\chi)\geqslant e^{10d}}} \mathbb{E}\big[|F_{\chi,H}(\mathbf{b}_1,\dots,\mathbf{b}_d)|^2\big]^{1/2} \ll d^{-2d}H^d.$$

Proof. By Lemma 4.3.11, we know that $\mathbb{E}[|F_{\chi,H}(\underline{\mathbf{b}})|^2]^{1/2} \leqslant Cd^{-5d}H^d$ for all characters $\chi \in \widehat{G}$ of order $\geqslant e^{10d}$, where $C > 0$ is an absolute constant (as before, $\underline{\mathbf{b}}$ stands for $\mathbf{b}_1,\dots,\mathbf{b}_d$). Therefore,

$$\sum_{\substack{\chi\in\widehat{G}^{M_*}\\ \mathrm{ord}(\chi)\geqslant e^{10d}}} \mathbb{E}\big[|F_{\chi,H}(\underline{\mathbf{b}})|^2\big]^{1/2} \leqslant \sum_{m=m_0}^{+\infty} \sum_{\substack{\chi\in\widehat{G}^{M_*}\\ \mathrm{ord}(\chi)\geqslant e^{10d}}} e^{-m+1}H^d \mathbf{1}_{e^{-m}<\mathbb{E}\left[|F_{\chi,H}(\underline{\mathbf{b}})|^2\right]^{1/2}H^{-d}\leqslant e^{-m+1}}(\chi)$$

where $m_0 := \lfloor 5d \log d - \log C + 1 \rfloor$. We may assume that $m_0 \geqslant \max(t_0 d, 4d \log d)$, since otherwise $d \ll 1$ and Proposition 4.3.14 is trivially satisfied (given that $|\widehat{G}^{M_*}| \leqslant N \ll 1$ when $d \ll 1$).

Applying Lemma 4.3.13 with $t := m/d$, we obtain that for every $m \geqslant t_0 d$,

$$\left| \left\{ \chi \in \widehat{G}^{M_*} \ : \ \mathrm{ord}(\chi) \geqslant e^{10d}, \ \ \mathbb{E}\big[|F_{\chi,H}(\underline{\mathbf{b}})|^2\big]^{1/2} > e^{-m} H^d \right\} \right| \ll e^{m/2}.$$

Therefore,

$$\sum_{\substack{\chi \in \widehat{G}^{M_*} \\ \mathrm{ord}(\chi) \geqslant e^{10d}}} \mathbb{E}\big[|F_{\chi,H}(\underline{\mathbf{b}})|^2\big]^{1/2} \ll \sum_{m=m_0}^{+\infty} e^{m/2} e^{-m+1} H^d \ll e^{-m_0/2} H^d$$

which is $O(d^{-2d} H^d)$ as $m_0 \geqslant 4d \log d$. □

4.3.4 Geometry of numbers

In this section, we show how to pass from good estimates on the number of lattice points in certain regions to the existence of a short basis for the lattice.

Lemma 4.3.15. *Let $L \geqslant 1$ be an integer. Cover the cube $[-L, L]^d$ by $(2L)^d$ cubes of side length 1 in the obvious way. Label these unit cubes $C_1, \ldots, C_{(2L)^d}$ (in any order). Let $V \subset \mathbb{R}^d$ be a hyperplane through the origin. Then the number of unit cubes C_i intersecting V is at most $(d+1)(2L)^{d-1}$.*

Proof. Let $e_1, \ldots, e_d$ be the standard basis for $\mathbb{R}^d$. Let $v \in \mathbb{R}^d$ be a unit vector orthogonal to V. Without loss of generality, since $\|v\|_2 = 1$, we may assume that $\langle v, e_1 \rangle \geqslant 1/\sqrt{d}$.

Suppose that V intersects two unit cubes C_i and C_j where $C_i = C_j + k e_1$ for some integer $k \geqslant 0$. Then, there is some $p \in C_i$ and some $w \in \mathbb{R}^d$ with $\|w\|_\infty \leqslant 1$ such that $p \in V$ and $p + k e_1 + w \in V$. Therefore, $k e_1 + w \in V$ and thus $\langle k e_1 + w, v \rangle = 0$. Noting that

$$\langle k e_1 + w, v \rangle \geqslant k \langle e_1, v \rangle - \|w\|_2 \|v\|_2 \geqslant \frac{k}{\sqrt{d}} - \sqrt{d},$$

we deduce that $k \leqslant d$.

We have thus proved that, for any cube C_i, there are at most $d+1$ cubes of the form $C_i + k e_1$ for some $k \in \mathbb{Z}$ which intersect V. The lemma follows. □

The next lemma allows us to convert information about the number of lattice points in cubes into the existence of short linearly independent lattice vectors.

Lemma 4.3.16. *Let $d \geqslant 1$. Let $\Lambda \subset \mathbb{R}^d$ be a full-rank lattice. Let $2 \leqslant H_0 < H_1$ be real numbers such that H_1/H_0 is an integer. Suppose that, for $i \in \{0,1\}$,*

$$\big|\Lambda \cap [-H_i, H_i]^d\big| = \theta_i \frac{(2H_i+1)^d}{\mathrm{vol}(\mathbb{R}^d/\Lambda)} \tag{4.20}$$

where $\theta_0, \theta_1 > 0$ satisfy $\frac{H_1}{H_0} > \frac{\theta_0}{\theta_1} d\big(\frac{5}{2}\big)^d$. Then $\Lambda \cap [-H_1, H_1]^d$ contains d linearly independent vectors.

Proof. By contradiction, suppose that there exists a linear hyperplane V containing all the points $v \in \Lambda \cap [-H_1, H_1]^d$.

Let $L = H_1/H_0$. The large cube $[-H_1, H_1]^d$ can be covered by $(2L)^d$ axis-parallel cubes of side length H_0 in the natural way. By Lemma 4.3.15, we can bound

$$\big|\Lambda \cap [-H_1, H_1]^d\big| \leqslant (d+1)(2L)^{d-1} \sup_C |\Lambda \cap C|,$$

where the supremum runs over all (not necessarily centred) axis-parallel cubes $C \subset \mathbb{R}^d$ of side length H_0. If C is such a cube and $v \in \Lambda \cap C$, then $C \subset v + [-H_0, H_0]^d$, which implies that

$$|\Lambda \cap C| \leqslant \big|\Lambda \cap \big(v + [-H_0, H_0]^d\big)\big| = \big|\Lambda \cap [-H_0, H_0]^d\big|.$$

Thus,

$$\big|\Lambda \cap [-H_1, H_1]^d\big| \leqslant (d+1)(2L)^{d-1}\big|\Lambda \cap [-H_0, H_0]^d\big|.$$

Plugging in our lattice point estimate (4.20), we get

$$\theta_1(2H_1+1)^d \leqslant (d+1)(2L)^{d-1}\theta_0(2H_0+1)^d.$$

Using $2H_1+1 \geqslant 2LH_0$ and $d+1 \leqslant 2d$, this implies that $L \leqslant \frac{\theta_0}{\theta_1} d\big(2+\frac{1}{H_0}\big)^d$, contradicting the inequality in the statement of the lemma. □

The linearly independent vectors given by Lemma 4.3.16 can be upgraded to a genuine basis for Λ by standard geometry of numbers, namely Mahler's theorem. We state this fact in a slightly more general situation.

Lemma 4.3.17. *Let $d \geqslant 1$. Let $\Lambda_1, \Lambda_2 \subset \mathbb{R}^d$ be full-rank lattices such that $M\Lambda_1 \subset \Lambda_2$ for some integer $M \geqslant 1$. Suppose that Λ_1 contains d linearly independent vectors in $[-H, H]^d$ for some $H > 0$. Then, Λ_2 admits a basis where each basis vector has Euclidean norm $\leqslant d^{3/2}MH$.*

Proof. By assumption, there exist d linearly independent vectors $v_1, \dots, v_d$ in the intersection $\Lambda_1 \cap [-H, H]^d$. Then $Mv_1, \dots, Mv_d$ are linearly independent vectors of Λ_2 such that $\max_{i \in [d]} \|Mv_i\|_2 \leqslant \sqrt{d}MH$.

Let $B \subset \mathbb{R}^d$ be the unit ball for the Euclidean norm. Let $0 < \lambda_1 \leqslant \cdots \leqslant \lambda_d$ be the successive minima[6] of B with respect to Λ_2. Since $Mv_1, \ldots, Mv_d \in \Lambda_2$ are linearly independent, we deduce from the above that $\lambda_d \leqslant \sqrt{d}MH$.

By Mahler's theorem (see [67, Theorem 3.34]), we conclude that Λ_2 admits a basis of vectors of Euclidean norm at most $d\lambda_d \leqslant d^{3/2}MH$, which is what we needed to show. □

4.3.5 Short basis vectors

We can now prove our main technical result, which may be of independent interest.

Theorem 4.3.18. *Let $N > 2$ be an integer. Let $d := \lceil \sqrt{\log N} \rceil$ and $X := d^{10^3 d}$.*

Let $\mathbf{b}_1, \ldots, \mathbf{b}_d$ be i.i.d. random variables, each uniformly distributed in the set of primes $\leqslant X$ not dividing N. Let $r \geqslant 0$ and let $\mathbf{x}_1, \ldots, \mathbf{x}_r$ be arbitrary[7] random variables taking values in $(\mathbb{Z}/N\mathbb{Z})^\times$.

Then, with probability $1 + O\big(d^{-d}\big)$, the lattice

$$\mathcal{L} := \Big\{ (e_1, \ldots, e_d, f_1, \ldots, f_r) \in \mathbb{Z}^{d+r} \ : \ \prod_{i=1}^{d} \mathbf{b}_i^{e_i} \prod_{i=1}^{r} \mathbf{x}_i^{f_i} \equiv 1 \pmod N \Big\}$$

has a basis consisting of vectors of Euclidean norm $\ll e^{42(d+r)}$.

Remark 4.3.19. The constant 42 in the exponent is by no means the limit of our techniques. For example, using smooth cutoffs in Definition 4.3.6 would significantly reduce the error term in Lemma 4.3.8 and hence lower this constant. We have not performed such optimisations in order to keep the arguments as simple as possible.

Proof of Theorem 4.3.18. Let $M := M_*(N)$ be the integer defined in Lemma 4.3.4. We introduce the auxiliary lattice

$$\mathcal{L}_M := \Big\{ (e_1, \ldots, e_d, f_1, \ldots, f_r) \in \mathbb{Z}^{d+r} \ : \ \prod_{i=1}^{d} \mathbf{b}_i^{Me_i} \prod_{i=1}^{r} \mathbf{x}_i^{Mf_i} \equiv 1 \pmod N \Big\}.$$

By Lemmas 4.3.7 and 4.3.8, we have the lattice point estimate

$$\big|\mathcal{L}_M \cap [-H, H]^{d+r}\big| = \frac{\big(1 + O\big(e^{31(d+r)} H^{-1}\big)\big)(2H+1)^{d+r}}{\big|\langle \mathbf{b}_1^M, \ldots, \mathbf{b}_d^M, \mathbf{x}_1^M, \ldots, \mathbf{x}_r^M \rangle\big|} + O\Bigg(\frac{1}{\big|\widehat{G}^M\big|} \sum_{\substack{\chi \in \widehat{G}^M \\ \mathrm{ord}(\chi) \geqslant e^{10d}}} |F_{\chi,H}(\underline{\mathbf{b}}, \underline{\mathbf{x}})| \Bigg) \tag{4.21}$$

[6] See [67, Definition 3.29] for the definition of successive minima.

[7] In particular, the $\mathbf{x}_i$ are not assumed to be independent or identically distributed.

for every integer $H \geqslant e^{31(d+r)}$, where

$$F_{\chi,H}(\underline{\mathbf{b}},\underline{\mathbf{x}}) := F_{\chi,H}(\mathbf{b}_1,\dots,\mathbf{b}_d,\mathbf{x}_1,\dots,\mathbf{x}_r) = \left(\prod_{i=1}^{d}\sum_{|h|\leqslant H}\chi^h(\mathbf{b}_i)\right)\left(\prod_{i=1}^{r}\sum_{|h|\leqslant H}\chi^h(\mathbf{x}_i)\right).$$

Since the $\mathbf{x}_i$ have unknown distributions, we will use the trivial bound

$$|F_{\chi,H}(\underline{\mathbf{b}},\underline{\mathbf{x}})| \leqslant (2H+1)^r|F_{\chi,H}(\mathbf{b}_1,\dots,\mathbf{b}_d)|.$$

Applying Proposition 4.3.14, we deduce that, for all $H \geqslant e^{10d}$,

$$\sum_{\substack{\chi\in\widehat{G}^{M*}\\ \mathrm{ord}(\chi)\geqslant e^{10d}}} \mathbb{E}\big[|F_{\chi,H}(\underline{\mathbf{b}},\underline{\mathbf{x}})|^2\big]^{1/2} \ll d^{-2d}(2H+1)^{d+r}.$$

By the L^2 triangle inequality and Chebyshev's inequality, this implies that, for fixed $H \geqslant e^{10d}$,

$$\mathbb{P}\Bigg(\sum_{\substack{\chi\in\widehat{G}^{M}\\ \mathrm{ord}(\chi)\geqslant e^{10d}}} |F_{\chi,H}(\underline{\mathbf{b}},\underline{\mathbf{x}})| > d^{-d}(2H+1)^{d+r}\Bigg) \ll d^{-2d}. \tag{4.22}$$

Let $H_0 := d\lceil e^{31(d+r)}\rceil$ and $H_1 := \lceil d(d+r)(5/2)^{d+r}\rceil H_0$. We apply (4.21) and (4.22) twice, once with $H = H_0$ and once with $H = H_1$ to obtain the following: with probability $1+O\big(d^{-d}\big)$, the two estimates

$$\big|\mathcal{L}_M \cap [-H_0,H_0]^{d+r}\big| = (1+O(1/d))\frac{(2H_0+1)^{d+r}}{\big|\langle \mathbf{b}_1^M,\dots,\mathbf{b}_d^M,\mathbf{x}_1^M,\dots,\mathbf{x}_r^M\rangle\big|}$$

and

$$\big|\mathcal{L}_M \cap [-H_1,H_1]^{d+r}\big| = (1+O(1/d))\frac{(2H_1+1)^{d+r}}{\big|\langle \mathbf{b}_1^M,\dots,\mathbf{b}_d^M,\mathbf{x}_1^M,\dots,\mathbf{x}_r^M\rangle\big|}$$

simultaneously hold.

We now apply Lemma 4.3.16. By our choice of H_0 and H_1, the inequality in the statement is satisfied provided that d is sufficiently large. We can assume that d is large enough as, for $d \ll 1$, we have $N \ll 1$ and Theorem 4.3.18 is trivially true. Thus, Lemma 4.3.16 implies that $\mathcal{L}_M \cap [-H_1,H_1]^{d+r}$ contains $d+r$ linearly independent vectors, with probability $1+O\big(d^{-d}\big)$. By Lemma 4.3.17, since $M\mathcal{L}_M \subset \mathcal{L}$, we conclude that, with probability $1+O\big(d^{-d}\big)$, $\mathcal{L}$ admits a basis of vectors of Euclidean norm at most

$$\ll d^{3/2}MH_1 \ll d^{3/2}e^{10d}d^2(d+r)(5/2)^{d+r}e^{31(d+r)} \ll e^{42(d+r)}$$

using the bound $M \leqslant e^{10d}$ given by Lemma 4.3.4. This completes the proof. □

4.4 Consequences for quantum algorithms

In this section, we prove the correctness of efficient quantum algorithms for factoring and for the discrete logarithm problem by applying our version of Regev's number-theoretic conjecture, Theorem 4.3.18.

4.4.1 Preparatory lemmas

Lemma 4.4.1. *Let $N, d, m \geqslant 2$ be integers. There is a classical algorithm that, given integers $0 \leqslant a_1, \ldots, a_d \leqslant 2^m$ and exponents $t_1, \ldots, t_d \in \{0, 1\}$, computes the product*

$$\prod_{i=1}^{d} a_i^{t_i} \pmod N$$

in time $O\big(md(\log d)\log(md)\big)$.

Proof. This is similar to [50, p5] or [49, Lemma 12] but for general values of m. Without loss of generality, assume that d is a power of 2, say $d = 2^l$. We proceed to compute this product in a binary tree fashion. Let $T(k)$ be the complexity of multiplying any k of these numbers $a_i^{t_i}$ modulo N. Note that $T(2k) \leqslant 2T(k) + O(M(mk))$ where $M(x)$ is the time needed to multiply two integers having at most x bits. By the work of Harvey and van der Hoeven [24], it is known that $M(x) = O(x \log x)$, which leads to the bound

$$T(d) \ll \sum_{j=0}^{l} 2^j M(md/2^{j+1}) \ll \sum_{j=0}^{l} md \log(md/2^{j+1}) \ll md(\log d)\log(md),$$

as claimed. □

We can turn this into a quantum circuit with the following well-known fact.

Lemma 4.4.2. *Any classical circuit can be "compiled" into a reversible quantum circuit that carries out the same computations, with the number of gates and qubits used being proportional to the size of the classical circuit.*

Proof. This well-known fact is explained in Section A.1 of the full version of [49]. □

Lemma 4.4.3. *Let N be a sufficiently large integer. Let $d := \lceil \sqrt{\log N} \rceil$ and $X = d^{10^3 d}$. Let $k = d^4$.*

Let $\mathbf{n}_1, \ldots, \mathbf{n}_k$ be i.i.d. random variables uniformly distributed in $\{1, \ldots, X\}$. Then, the probability that at least d of these $\mathbf{n}_i$ are primes not dividing N is $1 + O(1/N)$.

Proof. Since N has $\ll \log N$ prime factors, and since the number of primes $\leqslant X$ is $\gg X/\log X$, we have

$$\mathbb{P}(\mathbf{n}_1 \text{ is prime and } \mathbf{n}_1 \nmid N) \geqslant \frac{c}{\log X}$$

for some absolute constant $c > 0$. Thus, if E_i is the event that $\mathbf{n}_i$ is not prime or divides N, then by the union bound and independence,

$$\mathbb{P}\Big(\bigcup_{\substack{I\subset[k]\\|I|>k-d}}\bigcap_{i\in I} E_i\Big) \leqslant \sum_{l<d}\binom{k}{l}\Big(1-\frac{c}{\log X}\Big)^{k-l} \leqslant dk^d e^{-c(k-d)/\log X} \ll e^{-d^2}$$

as needed. □

4.4.2 The discrete logarithm problem

We now prove Theorem 1.4.2. For convenience, we restate it here.

Theorem 1.4.2. *There is a quantum circuit having $O(n^{3/2}\log^3 n)$ quantum gates and $O(n\log^3 n)$ qubits with the following property. There is a classical randomised polynomial-time algorithm that solves the discrete logarithm problem*

> **Input** : *an integer $N \leqslant 2^n$ and elements $g, y \in (\mathbb{Z}/N\mathbb{Z})^\times$ such that $y \in \langle g\rangle$*
>
> **Output** : *an integer x such that $g^x \equiv y \pmod N$*

using $O(\sqrt{n})$ calls to this quantum circuit, and succeeds with probability $\asymp 1$.

Proof of Theorem 1.4.2. Suppose we are given $N > 2$ and elements $g, y \in (\mathbb{Z}/N\mathbb{Z})^\times$ such that y lies in the subgroup generated by g in $(\mathbb{Z}/N\mathbb{Z})^\times$. Let $d := \lceil\sqrt{\log N}\rceil$ and $X := d^{10^3 d}$. Let $\mathbf{b}_1, \dots, \mathbf{b}_d$ be i.i.d. random variables, each uniformly distributed in the set of primes $\leqslant X$ not dividing N.

Consider the random lattice

$$\mathcal{L}_{\underline{\mathbf{b}},g,y} := \Big\{(e_1,\dots,e_d,f_1,f_2)\in\mathbb{Z}^{d+2} \;:\; \Big(\prod_{i=1}^d \mathbf{b}_i^{e_i}\Big) g^{f_1} y^{f_2} \equiv 1 \pmod N\Big\}.$$

By Theorem 4.3.18, with probability $1 + O(d^{-d})$, this lattice has a basis consisting of vectors of Euclidean norm $\ll e^{42d}$.

Computing the discrete logarithm of y with respect to the base g reduces (with a polynomial-time classical algorithm) to computing a short basis for $\mathcal{L}_{\underline{\mathbf{b}},g,y}$. To see why this is the case, suppose that we managed to compute a basis $v_1, \dots, v_{d+2}$ for $\mathcal{L}_{\underline{\mathbf{b}},g,y}$ with $\max_i \|v_i\|_2 \ll e^{42d}$. It is then easy to find a vector in $\mathcal{L}_{\underline{\mathbf{b}},g,y}$ of the form $(0,\dots,0,x,-1)$ for some integer x (note that such a vector exists since we assume that $y \in \langle g\rangle$). Indeed, this amounts to solving a linear system with integer coefficients.

The complexity of solving an integer linear system is polynomial in the dimensions of the matrix and the number of bits of the coefficients, which are both $O(d)$ since $\max_i \|v_i\|_2 \ll e^{42d}$ (see [3] for an early example of such an algorithm). This yields an integer x such that $g^x \equiv y \pmod{N}$.

The algorithm for solving the discrete logarithm problem thus proceeds as follows.

1. Generate d primes $b_1, \ldots, b_d$ independently and uniformly at random in the set of primes $\leqslant X$ not dividing N. To do this, it suffices to generate d^4 independent random integers $\leqslant X$. By Lemma 4.4.3, the required conditions will be satisfied for at least d of those with probability $1 + O(1/N)$. This first step takes polynomial time on a classical computer using the AKS primality test [1].
2. Use the procedure described by Ekerå and Gärtner [14] to obtain a basis for the lattice $\mathcal{L}_{\underline{b},g,y}$. This involves making $O(d)$ calls to a quantum circuit, followed by polynomial-time classical post-processing. This step is now guaranteed to succeed with probability $\asymp 1$, using the fact that $\mathcal{L}_{\underline{b},g,y}$ has a basis of vectors of Euclidean norm $\ll e^{42d}$ with probability $1 + O(d^{-d})$ (Theorem 4.3.18).
3. Compute the discrete logarithm of y in classical polynomial time using this basis, as explained above.

It remains to analyse the gate and space costs of the quantum part of this algorithm.

The only modification needed to the analysis of the quantum circuit in [14] comes from the fact that $b_1, \ldots, b_d$ are not quite as small as in [14]. In [14] (and more generally in [49, 50]), the primes b_i are assumed to have $O(\log d)$ bits. In the present situation, we instead have the bound $b_i \leqslant X = d^{10^3 d}$, i.e. each b_i has $O(d \log d)$ bits.

The only place in [14] where the assumption on the size of the b_i's is used is in [14, Lemma 3]. In turn, the only point in the proof of [14, Lemma 3] where this assumption is needed is when [49, Lemma 5] is invoked (as a black box). The proof of [49, Lemma 5] is given in [49, Section 5], and the size assumption on the b_i's only comes up in [49, Lemma 12].

The result [49, Lemma 12] essentially[8] states that there is a quantum circuit using $O(d \log^3 d)$ gates and $O(d \log^3 d)$ qubits to perform the computation of $\prod_{i=1}^{d} a_i^{t_i}$ where $t_i \in \{0, 1\}$ and a_i are integers with $O(\log d)$ bits. This is a special case of Lemma 4.4.1 (used together with Lemma 4.4.2) applied with $m = O(\log d)$. In our case, we just apply Lemma 4.4.1 with $m = O(d \log d)$, together with Lemma 4.4.2 to convert the classical circuit into a quantum one. This yields a quantum circuit having $O(d^2 \log^3 d)$ gates and $O(d^2 \log^3 d)$ qubits to compute $\prod_{i=1}^{d} b_i^{t_i}$.

[8] There are extra details pertaining to the precise use of qubits (e.g. restoring the ancilla qubits to $|0\rangle$), but these will be the same in our setup.

The number of gates of the quantum circuit [49, Lemma 5] is

$$O\big(d(n\log n + d\log^3 d)\big) = O(n^{3/2}\log n)$$

(because [49, Lemma 12] is used $O(d)$ times, see [49, Algorithm 5.2] and the surrounding explanations). Note that this is the same as for Regev's algorithm (see [50, p5]). In our case, the number of gates is

$$O\big(d(n\log n + d^2\log^3 d)\big) = O(n^{3/2}\log^3 n).$$

The space optimisations of Ragavan and Vaikuntanathan keep the total number of qubits for their circuit [49, Lemma 5] under

$$O\big(n\log n + d\log^3 d\big) = O(n\log n),$$

due to the way the qubits are used and restored in the main loop of [49, Algorithm 5.2]. In our situation, the number of qubits is

$$O\big(n\log n + d^2\log^3 d\big) = O(n\log^3 n).$$

In summary, our final quantum circuit has $O(n^{3/2}\log^3 n)$ gates and $O(n\log^3 n)$ qubits. As noted in [14], the fact that the two elements g and y are not small (g and y can be as large as N, as opposed to the b_i's) does not affect these complexity bounds. □

Remark 4.4.4. In the proof of Theorem 1.4.2, we used the naive Lemma 4.4.2 to convert a classical circuit into a quantum one. As mentioned in Remark 1.4.3, it is possible to save a factor of $\log n$ in the space cost for Theorems 1.4.1 and 1.4.2 by reusing certain qubits when performing the quantum computation corresponding to Lemma 4.4.1.

4.4.3 Factoring integers

In this section, we prove Theorem 1.4.1.

Lemma 4.4.5. *Let $N > 1$ be an odd integer with at least two distinct prime factors. Let $\mathbf{x}$ be a random variable, uniformly distributed in $(\mathbb{Z}/N\mathbb{Z})^\times$. Then $\langle\mathbf{x}\rangle$ contains a non-trivial square root of 1 modulo N with probability $\geqslant 1/2$.*

Proof. This elementary number-theoretic fact is well-known – it was already needed for Shor's algorithm [57]. See [44, Appendix A4.3] for a detailed proof. □

For the reader's convenience, we reproduce Theorem 1.4.1 below.

Theorem 1.4.1. *There is a quantum circuit having $O(n^{3/2}\log^3 n)$ quantum gates and $O(n\log^3 n)$ qubits with the following property. There is a classical randomised polynomial-time algorithm that solves the factoring problem*

$$\textbf{Input} : \textit{a composite integer } N \leqslant 2^n$$
$$\textbf{Output} : \textit{a non-trivial divisor of } N$$

using $O(\sqrt{n})$ calls to this quantum circuit, and succeeds with probability $\asymp 1$.

Proof of Theorem 1.4.1. We can assume that N is odd and not a perfect prime power, as otherwise it is easy to factor N in polynomial time with a classical computer (see [44, Exercise 5.17]).

Let $d := \lceil\sqrt{\log N}\rceil$ and $X := d^{10^3 d}$. The probabilistic algorithm to find a non-trivial divisor of N goes as follows.

1. Generate d primes $b_1, \ldots, b_d$ independently and uniformly at random in the set of primes $\leqslant X$ not dividing N. Sample another integer x uniformly chosen in $(\mathbb{Z}/N\mathbb{Z})^\times$. As in the proof of Theorem 1.4.2, this can be done classically in polynomial time with probability $1 + O(1/N)$.
2. Use the algorithm of Ekerå and Gärtner [14] to obtain a basis for the lattice
$$\mathcal{L}_{\underline{b},x} := \left\{(e_1, \ldots, e_d, f) \in \mathbb{Z}^{d+1} \ : \ \left(\prod_{i=1}^{d} b_i^{e_i}\right) x^f \equiv 1 \pmod N\right\}.$$
This involves making $O(d)$ calls to a quantum circuit, followed by polynomial-time classical post-processing. By Theorem 4.3.18, $\mathcal{L}_{\underline{b},x}$ has a basis of vectors of Euclidean norm $\ll e^{42d}$ with probability $1 + O(d^{-d})$, which means that this step is guaranteed to succeed with probability $\asymp 1$.
3. Using the short basis for $\mathcal{L}_{\underline{b},x}$ computed in the previous step, find the vector of the form $(0, \ldots, 0, r)$ in $\mathcal{L}_{\underline{b},x}$ with $r \geqslant 1$ as small as possible. This involves solving a linear system with integer coefficients, which can be done efficiently as in the proof of Theorem 1.4.2 (using e.g. [3]). This integer r is the order of x in $(\mathbb{Z}/N\mathbb{Z})^\times$. By Lemma 4.4.5, with probability $\geqslant 1/2$, this order r will be even and the element $x^{r/2}$ will be a non-trivial square root of 1 modulo N. Hence, N divides the product $(x^{r/2} - 1)(x^{r/2} + 1)$ but neither term individually, which implies that $\gcd(N, x^{r/2} - 1 \bmod N)$ is a non-trivial divisor of N.

The analysis of this algorithm is identical to the corresponding part of the proof of Theorem 1.4.2. □

Appendix A

Non-backtracking operators

A.1 A weighted Ihara-Bass formula

We begin by showing how the results in Deitmar [12] combine to yield the weighted Ihara-Bass formula stated in Proposition 2.4.5. Note that Deitmar's setting is slightly different, as it encompasses directed weighted graphs that may be infinite but have summable weights. Alternatively, one could derive the same identity from [16, Lemma 5.2] or [74, Theorem 2].

Proof of Proposition 2.4.5. Let $u \in \mathbb{C}$ be sufficiently small. The weighted *Ihara zeta function* $Z(u)$ of G is defined, in analogy to the Riemann zeta function, as an Euler product over regular prime cycles in G (see [12, Definition 1.4]). As stated in [12, Theorem 1.6], it is related to the non-backtracking operator M by the identity

$$\det(I - uM) = Z(u)^{-1}. \tag{A.1}$$

Furthermore, [12, Theorem 2.10] provides an alternative expression for $Z(u)^{-1}$:

$$Z(u)^{-1} = \det(B(-u)) \prod_{\{x,y\}\in E} \big(1 - u^2 w(x,y)^2\big), \tag{A.2}$$

where $B(-u)$ is an operator on $L(V)$ defined in [12, Lemma 2.2]. Expanding this definition, we get

$$B(-u) = I - uB_1 + \sum_{n\geqslant 1} u^{2n} B_{2n} - \sum_{n\geqslant 1} u^{2n+1} B_{2n+1} \tag{A.3}$$

where the linear maps $B_j : L(V) \to L(V)$ are defined on the canonical basis $(\mathtt{x})_{x\in V}$ by

$$B_{2n-1}(\mathtt{x}) = \sum_{y\in V} w(x,y)^{2n-1}\mathtt{y} \quad \text{and} \quad B_{2n}(\mathtt{x}) = \sum_{y\in V} w(x,y)^{2n}\mathtt{x}$$

for $n \geqslant 1$. In particular, B_1 is the adjacency operator A, and the two sums in (A.3) evaluate to $u^2 D_u$ and $u^2 O_u$ respectively, by the geometric series formula.

Comparing (A.1) and (A.2) yields the Ihara-Bass formula (2.13) for small u. A straightforward analytic continuation argument extends this equality for all $u \in \mathbb{C}$ for which D_u and O_u are defined, and also shows that any positivity assumption on the weights in [12] can be disregarded for this result. □

A.2 Trace expansion and non-backtracking walks

Definition A.2.1. Let $G = (V, E, w)$ be a weighted graph. A *walk* in G is a finite sequence $(v_1, v_2, \dots, v_k)$ of vertices of G such that $\{v_i, v_{i+1}\} \in E$ for all $1 \leqslant i < k$. A walk $(v_1, v_2, \dots, v_k)$ is *non-backtracking* if $v_{i-1} \neq v_{i+1}$ for all $1 < i < k$. We write $\mathrm{nbw}(G)$ for the set of non-backtracking walks in G (of any length).

Lemma A.2.2. *Let $G = (V, E, w)$ be a weighted graph with non-backtracking operator M. Then, for any $k \geqslant 1$,*

$$\mathrm{Tr}\big[(M^k)^* M^k\big] = \sum_{\substack{x_{-1},x_0,x_1,\dots,x_{2k} \in V \\ (x_{-1},x_0,\dots,x_k) \in \mathrm{nbw}(G) \\ (x_k,x_{k+1},\dots,x_{2k},x_{-1}) \in \mathrm{nbw}(G) \\ (x_0,x_{k-1}) = (x_{2k},x_{k+1})}} \prod_{i=1}^{2k} w(x_{i-1}, x_i).$$

Proof. By definition of the non-backtracking operator M, for any basis vector $\overrightarrow{\mathtt{x}_{-1}\mathtt{x}_0}$ of $L(\vec{E})$,

$$M^k(\overrightarrow{\mathtt{x}_{-1}\mathtt{x}_0}) = \sum_{\substack{x_1,\dots,x_k \in V \\ (x_{-1},x_0,\dots,x_k) \in \mathrm{nbw}(G)}} \prod_{i=1}^{k} w(x_{i-1}, x_i)\ \overrightarrow{\mathtt{x}_{k-1}\mathtt{x}_k}.$$

Similarly, for any basis vector $\overrightarrow{\mathtt{y}_{-1}\mathtt{y}_0}$ of $L(\vec{E})$,

$$(M^k)^*(\overrightarrow{\mathtt{y}_{-1}\mathtt{y}_0}) = \sum_{\substack{y_{-2},\dots,y_{-k-1} \in V \\ (y_0,y_{-1},\dots,y_{-k-1}) \in \mathrm{nbw}(G)}} \prod_{i=0}^{k-1} w(y_{-i}, y_{-i-1})\ \overrightarrow{\mathtt{y}_{-k-1}\mathtt{y}_{-k}}.$$

Therefore, the diagonal entry at $\overrightarrow{\mathtt{x}_{-1}\mathtt{x}_0}$ in the standard matrix for $(M^k)^* M^k$ is

$$\sum_{\substack{x_1,\dots,x_k \in V \\ (x_{-1},x_0,\dots,x_k) \in \mathrm{nbw}(G)}} \sum_{\substack{y_0,y_{-1},\dots,y_{-k-1} \in V \\ (y_0,y_{-1},\dots,y_{-k-1}) \in \mathrm{nbw}(G) \\ (y_{-1},y_0) = (x_{k-1},x_k) \\ (y_{-k-1},y_{-k}) = (x_{-1},x_0)}} \prod_{i=1}^{k} w(x_{i-1}, x_i) \prod_{i=0}^{k-1} w(y_{-i}, y_{-i-1}).$$

Performing the change of variables $x_{k+i} := y_{-i}$ for $1 \leqslant i \leqslant k$, this can be rewritten as

$$\big[(M^k)^* M^k\big]_{\overrightarrow{\mathtt{x}_{-1}\mathtt{x}_0},\overrightarrow{\mathtt{x}_{-1}\mathtt{x}_0}} = \sum_{\substack{x_1,\dots,x_{2k} \in V \\ (x_{-1},x_0,\dots,x_k) \in \mathrm{nbw}(G) \\ (x_k,x_{k+1},\dots,x_{2k},x_{-1}) \in \mathrm{nbw}(G) \\ (x_0,x_{k-1}) = (x_{2k},x_{k+1})}} \prod_{i=1}^{2k} w(x_{i-1}, x_i).$$

Summing over all diagonal entries gives the formula for the trace. □

Appendix B

Frequency relations via Elliott's inequality

In this appendix, we reproduce the arguments in [71, Section 3] to show that the frequencies θ_x associated to large Fourier coefficients of a multiplicative function in short intervals must satisfy certain local relations, making the quantitative dependence on δ explicit.

B.1 Ingredients

Lemma B.1.1. *Let $g : I \to \mathbb{C}$ be a 1-bounded function, where $I \subset \mathbb{Z}$ is an interval of size $|I| = H$. For every $\tau > 0$, there is an exceptional set of primes $\mathcal{E}$ with*

$$\sum_{p\in\mathcal{E}} \frac{1}{p} \ll \tau^{-2}$$

such that, for every prime $p \leqslant H$ not in $\mathcal{E}$, we have

$$\left|\frac{1}{H}\sum_{n\in I} g(n) - \frac{p}{H}\sum_{\substack{n\\ pn\in I}} g(pn)\right| \leqslant \tau.$$

Proof. This is (essentially) Elliott's inequality [70, Theorem 3.13, p.462]. We reproduce the proof for the interested reader.

It is enough to show the second moment bound

$$\sum_{p\leqslant H}\frac{1}{p}\left|\frac{1}{H}\sum_{n\in I} g(n) - \frac{p}{H}\sum_{\substack{n\\ pn\in I}} g(pn)\right|^2 \ll 1.$$

Writing $F(p) := \sum_{n\in I}\left(\mathbf{1}_{p|n} - \frac{1}{p}\right)p^{1/2}g(n)$, this is equivalent to

$$\sum_{p\leqslant H}|F(p)|^2 \ll H^2. \tag{B.1}$$

By Cauchy-Schwarz, we have

$$\sum_{p\leqslant H}|F(p)|^2 \leqslant H^{1/2}\Big(\sum_{n\in I}\Big|\sum_{p\leqslant H}\big(\mathbf{1}_{p|n}-\tfrac{1}{p}\big)p^{1/2}\overline{F(p)}\Big|^2\Big)^{1/2}.$$

Expanding the square and swapping the order of summation, we get

$$\sum_{n\in I}\Big|\sum_{p\leqslant H}\big(\mathbf{1}_{p|n}-\tfrac{1}{p}\big)p^{1/2}\overline{F(p)}\Big|^2 \ll H\sum_{p\leqslant H}|F(p)|^2+\sum_{\substack{p_1,p_2\leqslant H\\ p_1\neq p_2}}p_1^{1/2}p_2^{1/2}|F(p_1)||F(p_2)|, \quad \text{(B.2)}$$

using that $\sum_{n\in I}\big(\mathbf{1}_{p_1|n}-\frac{1}{p_1}\big)\big(\mathbf{1}_{p_2|n}-\frac{1}{p_2}\big)\ll 1$ for $p_1\neq p_2$. By Cauchy-Schwarz, the off-diagonal contribution in (B.2) is negligible. Simplifying, we obtain (B.1). □

Lemma B.1.2. *Let $g: I\to\mathbb{C}$ be a 1-bounded function, where $I\subset\mathbb{Z}$ is an interval of size $|I|=H$. Suppose that $S\subset\mathbb{R}/\mathbb{Z}$ is a set of $1/H$-separated frequencies such that*

$$\Big|\frac{1}{H}\sum_{n\in I}g(n)e(n\alpha)\Big|\geqslant\tau$$

for all $\alpha\in S$. Then $|S|\ll\tau^{-3}$.

Proof. By the triangle inequality, the lower bound $\big|\frac{1}{H}\sum_{n\in I}g(n)e(n\alpha)\big|\gg\tau$ persists for all α in a $(c\tau/H)$-neighbourhood of S, for some small constant $c>0$. Since S is $1/H$-separated, we have

$$|S|\frac{\tau}{H}\tau^2\ll\int_0^1\Big|\frac{1}{H}\sum_{n\in I}g(n)e(n\alpha)\Big|^2d\alpha\leqslant\frac{1}{H},$$

which rearranges to $|S|\ll\tau^{-3}$. □

B.2 Obtaining the local relations

Proposition B.2.1. *Let $f:\mathbb{N}\to\mathbb{C}$ be a 1-bounded multiplicative function.*

Let $C_0\leqslant H\leqslant X$ and $\delta\geqslant C_0(\log\log H)^{-1/2}$ for some sufficiently large absolute constant C_0.

Let $A\subset[X,2X]$ be an H-separated set. Suppose that, for each $x\in A$, there exists a frequency $\theta_x\in\mathbb{R}/\mathbb{Z}$ satisfying

$$\Big|\sum_{x\leqslant n<x+H}f(n)e(n\theta_x)\Big|\geqslant\delta H.$$

Then, there exists a scale P with $H^{e^{-O(\delta^{-2})}}\leqslant P\leqslant H^{1/10}$ such that, writing $\mathcal{P}$ for the set of primes in $[P,2P]$, there are

$$\gg\delta^6\frac{|A|^2}{X/H}|\mathcal{P}|^2$$

quadruples $(x,y,p,q)\in A^2\times\mathcal{P}^2$ with $|qx-py|\leqslant\frac{1}{10}PH$ and $\|p\theta_x-q\theta_y\|\leqslant\frac{P}{10H}$.

Proof. Let W be the set of all pairs $(x,p) \in A \times \{p \text{ prime} : p \leqslant H\}$ such that

$$\left| \frac{p}{H} \sum_{\substack{n \\ x \leqslant pn < x+H}} f(pn) e(pn\theta_x) \right| \geqslant \frac{\delta}{2}. \tag{B.3}$$

The set W can be shown to be large using Elliott's inequality. Indeed, for each $x \in A$, applying Lemma B.1.1 with $g(n) := f(n)e(n\theta_x)$, $I := [x, x+H) \cap \mathbb{Z}$ and $\tau := \delta/2$, we have

$$\sum_{p \leqslant H} \frac{1 - \mathbf{1}_{(x,p)\in W}}{p} \ll \delta^{-2}.$$

Let $P_1 := H^{1/10}$ and $P_0 := e^{-k}P_1$ for some parameter $k \in \mathbb{N}$ such that $10 \leqslant P_0 \leqslant P_1^{1/2}$. Dropping the primes $p \notin [P_0, P_1)$ and summing over $x \in A$, we get

$$\sum_{i=0}^{k-1} \frac{1}{e^{-i}P_1} \sum_{e^{-i-1}P_1 \leqslant p < e^{-i}P_1} \sum_{x \in A} \left(1 - \mathbf{1}_{(x,p)\in W}\right) \ll \delta^{-2}|A|.$$

By averaging, there exists $0 \leqslant i \leqslant k$ such that

$$\frac{\log(e^{-i}P_1)}{e^{-i}P_1} \sum_{e^{-i-1}P_1 \leqslant p < e^{-i}P_1} \sum_{x \in A} \left(1 - \mathbf{1}_{(x,p)\in W}\right) \ll \frac{\delta^{-2}}{\sum_{i=0}^{k-1} \frac{1}{\log(e^{-i}P_1)}}|A| \ll \frac{\delta^{-2}}{\log \frac{\log P_1}{\log P_0}}|A|.$$

Let C be a large constant to be chosen later. We assume throughout that the constant C_0 in the statement of the proposition is sufficiently large with respect to C. Select k so that $P_0 \asymp H^{e^{-C\delta^{-2}}}$. By the prime number theorem, for $P' := e^{-i-1}P_1 \in [P_0, P_1]$, we have

$$\frac{1}{|\mathcal{P}'|} \sum_{p \in \mathcal{P}'} \sum_{x \in A} \mathbf{1}_{(x,p)\in W} \gg |A|$$

provided C is sufficiently large, where $\mathcal{P}'$ denotes the set of primes in $[P', eP')$. It will be convenient to restrict to a slightly shorter interval of primes. By the pigeonhole principle, there exists $P \in [P', eP')$ such that, writing $\mathcal{P}$ for the set of primes in $[P, (1+\delta/C)P]$, we have

$$\sum_{x \in A} \sum_{p \in \mathcal{P}} \mathbf{1}_{(x,p)\in W} \gg \frac{\delta}{C}|A||\mathcal{P}'|. \tag{B.4}$$

For every $(x,p) \in W \cap (A \times \mathcal{P})$, since f is multiplicative (but not necessarily completely multiplicative), the estimate (B.3) implies that

$$\left| \frac{p}{H} \sum_{\substack{n \\ x \leqslant pn < x+H}} f(n) e(np\theta_x) \right| \geqslant \frac{\delta}{2} - O(P^{-1}) \geqslant \frac{\delta}{4}, \tag{B.5}$$

where the last inequality holds provided C_0 is large enough in terms of C, using the assumption $\delta \geqslant C_0(\log\log H)^{-1/2}$ and the definition of P_0.

Let Z be a maximal $C^{-1}\delta H/P$-separated subset of $[\frac{X}{2P}, \frac{2X}{P}]$. For each $z \in Z$, let S_z be a maximal $\frac{P}{20H}$-separated set of frequencies $\beta \in \mathbb{R}/\mathbb{Z}$ with the property that

$$\left|\frac{P}{H}\sum_{z\leqslant n<z+\frac{H}{P}} f(n)e(n\beta)\right| \geqslant \frac{\delta}{8}.$$

By Lemma B.1.2, we know that $|S_z| \ll \delta^{-3}$.

For each pair $(x,p) \in W \cap (A\times\mathcal{P})$, there is at least one $z \in Z$ such that $|x/p - z| \leqslant C^{-1}\delta H/P$; by (B.5), we then have

$$\left|\frac{P}{H}\sum_{z\leqslant n<z+\frac{H}{P}} f(n)e(np\theta_x)\right| \geqslant \left|\frac{p}{H}\sum_{\frac{x}{p}\leqslant n<\frac{x+H}{p}} f(n)e(np\theta_x)\right| - O(C^{-1}\delta) \geqslant \frac{\delta}{8}$$

for C large enough, using that $|p-P| \leqslant C^{-1}\delta P$. Therefore, $\|p\theta_x - \beta\| \leqslant \frac{P}{20H}$ for some $\beta \in S_z$. By Cauchy-Schwarz, this implies that the number of tuples (x,y,p,q,z,β) in $A^2\times\mathcal{P}^2\times Z\times\mathbb{R}/\mathbb{Z}$ such that $(x,p),(y,q)\in W$, $\beta\in S_z$,

$$|x/p - z|, |y/q - z| \leqslant C^{-1}\delta H/P \quad \text{and} \quad \|p\theta_x - \beta\|, \|q\theta_y - \beta\| \leqslant \tfrac{P}{20H}$$

is at least

$$\frac{(\sum_{x\in A}\sum_{p\in\mathcal{P}}\mathbf{1}_{(x,p)\in W})^2}{\sum_{z\in Z}|S_z|} \gg \frac{(C^{-1}\delta|A||\mathcal{P}'|)^2}{C\delta^{-4}X/H} = C^{-3}\delta^6\frac{|A|^2}{X/H}|\mathcal{P}'|^2,$$

using (B.4) and the bound $|S_z| \ll \delta^{-3}$. Hence, by the triangle inequality, the number of quadruples $(x,y,p,q)\in A^2\times\mathcal{P}^2$ with $|qx-py| \ll C^{-1}\delta PH$ and $\|p\theta_x - q\theta_y\| \leqslant \frac{P}{10H}$ is at least

$$\gg C^{-3}\delta^6\frac{|A|^2}{X/H}|\mathcal{P}'|^2.$$

The conclusion follows upon choosing C sufficiently large. □

Appendix C

An adaptation of the Matomäki-Radziwiłł theorem

In this appendix, we prove Theorem C.5.1, a structural result characterising multiplicative functions that correlate with linear phases $e(n\alpha_x)$ on many short intervals, where $x \mapsto \alpha_x$ is a smoothly varying function of the form

$$\alpha_x \approx \frac{a}{q} + \frac{T}{x} \pmod 1.$$

While the underlying arguments are well known to experts, they are scattered across [35, 36, 40], and no result of the required generality is proved in the literature. We fill in this gap by providing a detailed proof.

C.1 Preliminary lemmas and technical setup

The following lemma is a restatement of [36, Lemma 8.1], whose proof is not explicitly provided in [36]. Instead, the reader is referred to the proof of [35, Lemma 14] for justification. Since [36, Lemma 8.1] differs slightly from [35, Lemma 14], we include a proof for completeness.

Lemma C.1.1. *Let $X \geqslant y \geqslant 1$. Let $f : \mathbb{R} \to \mathbb{C}$ be in $L^1 \cap L^2$. Then*

$$\frac{1}{X}\int_X^{2X}\left|\int_{\mathbb{R}} f(t)\frac{(x+y)^{1+it}-x^{1+it}}{1+it}dt\right|^2 dx \ll \max_{T\geqslant X/y}\frac{Xy}{T}\int_{-T}^{T}|f(t)|^2dt. \tag{C.1}$$

Proof. If f is supported on $[-X/y, X/y]$, using the trivial bound

$$\left|\frac{(x+y)^{1+it}-x^{1+it}}{1+it}\right| = \left|\int_x^{x+y} u^{it}du\right| \leqslant y, \tag{C.2}$$

we immediately obtain that the left-hand side of (C.1) is $\leqslant y^2\int_{-X/y}^{X/y}|f(t)|^2dt$, which is acceptable.

Hence, it suffices to treat the case where the support of f is contained in the complement of $[-X/y, X/y]$; the general case then follows by the triangle inequality.

It turns out to be enough to prove the inequality

$$\frac{1}{X}\int_X^{2X}\left|\int_{\mathbb{R}} f(t)x^{1+it}\frac{(1+u)^{1+it}-1}{1+it}dt\right|^2 dx \ll \max_{T\geqslant X/y}\frac{Xy}{T}\int_{-T}^{T}|f(t)|^2dt \tag{C.3}$$

where $u \ll y/X \leqslant 1$ is a *fixed* real number (this is the analogue of (C.1) where the x-dependent quantity y/x has been replaced by u). This reduction from (C.1) to (C.3) follows from a trick, which consists of applying the identity

$$\frac{(x+y)^s - x^s}{s} = \frac{1}{2y}\left(\int_y^{3y}\frac{(x+w)^s-x^s}{s}dw - \int_y^{3y}\frac{(x+w)^s-(x+y)^s}{s}dw\right)$$

and pulling the integral over w outside via Cauchy-Schwarz (see [35, Proof of Lemma 14] for further details).

Introducing a smooth weight function $\mathbf{1}_{[1,2]} \leqslant W \leqslant \mathbf{1}_{[1/2,5/2]}$ and changing the order of summation, the left-hand side of (C.3) is

$$\ll \int_{\mathbb{R}}\int_{\mathbb{R}}\left|\frac{f(t_1)f(t_2)}{t_1t_2}\right|\left|\frac{1}{X}\int_{\mathbb{R}} W\Big(\frac{x}{X}\Big)x^{2+i(t_1-t_2)}dx\right|dt_1dt_2. \tag{C.4}$$

A standard Mellin transform computation shows that

$$\frac{1}{X}\int_{\mathbb{R}} W\Big(\frac{x}{X}\Big)x^{2+i(t_1-t_2)}dx \ll \frac{X^2}{1+|t_1-t_2|^2}.$$

Using $|f(t_1)f(t_2)| \leqslant |f(t_1)|^2 + |f(t_2)|^2$, the expression (C.4) simplifies to

$$\ll X^2\int_{\mathbb{R}}\left|\frac{f(t)}{t}\right|^2 dt.$$

By a dyadic decomposition, and recalling that f is zero on $[-X/y, X/y]$, we obtain (C.3). This proves Lemma C.1.1. □

Following Matomäki and Radziwiłł [35, 36], we shall work with a sequence (a_n) supported on integers n having prime factors in prescribed ranges, and obeying a suitable factorability assumption.

Notation C.1.2. Let $X \geqslant 10^{100}$. Let $10 \leqslant P_1 < Q_1 \leqslant P_2 < Q_2 \leqslant P_3 < Q_3 \leqslant X^{1/3}$. Let $(a_n)_{n\geqslant 1}$ be a sequence of 1-bounded complex numbers with the following properties.

1. The support of (a_n) is contained in the set of $X/4 < n \leqslant 8X$ having at least one prime factor in each of the three intervals $(P_i, Q_i]$, and no repeated prime factor in those intervals.

2. There are 1-bounded sequences (b_m) and (c_p) such that, whenever $n \in (X/4, 8X]$ is of the form $n = mp_1p_2p_3$ for some integer m and primes $p_i \in (P_i, Q_i]$ not dividing m, we have $a_n = b_m c_{p_1} c_{p_2} c_{p_3}$.

Notation C.1.3. Let W_1, W_2 be smooth functions satisfying the pointwise bounds

$$\mathbf{1}_{[1,5/2]} \leqslant W_1 \leqslant \mathbf{1}_{[1/2,4]} \leqslant W_2 \leqslant \mathbf{1}_{[1/4,8]},$$

and whose Mellin transforms $\widetilde{W_j}(s) := \int_0^\infty W_j(x)x^{s-1}dx$ obey the decay estimate

$$\widetilde{W_j}(it) \ll \exp(-c|t|^{1/2}) \tag{C.5}$$

for $j = 1, 2$ (use e.g. the construction in [29]). We define

$$F(s) := \sum_{n\geqslant 1} \frac{a_n}{n^s} W_1\Big(\frac{n}{X}\Big). \tag{C.6}$$

Lemmas C.1.4 and C.1.5 give alternative expressions for $F(s)$ via Mellin inversion, with Lemma C.1.5 also relying on the multiplicativity properties of the sequence (a_n).

Lemma C.1.4. *For $s \in \mathbb{C}$ with $\mathrm{Re}(s) \geqslant 1$,*

$$F(s) = \frac{1}{2\pi} \int_{-(\log X)^3}^{(\log X)^3} \Big(\sum_{n\geqslant 1} \frac{a_n}{n^{s+iu}} W_2\Big(\frac{n}{X}\Big)\Big) \widetilde{W_1}(iu) X^{iu} du + O(X^{-10}).$$

Proof. By Mellin inversion of W_1, for any $s \in \mathbb{C}$, we have

$$F(s) = \sum_{n\geqslant 1} \frac{a_n}{n^s} W_2\Big(\frac{n}{X}\Big) W_1\Big(\frac{n}{X}\Big) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \sum_{n\geqslant 1} \frac{a_n}{n^{s+iu}} W_2\Big(\frac{n}{X}\Big) \widetilde{W_1}(iu) X^{iu} du.$$

The integral can then be truncated using the estimate (C.5). □

Lemma C.1.5. *For $s \in \mathbb{C}$ with $\mathrm{Re}(s) \geqslant 1$, the function $F(s)$ can be written as*

$$\sum_{A,B,C} \frac{1}{2\pi} \int_{-(\log X)^3}^{(\log X)^3} Q_{2,A}(s+iu) Q_{3,B}(s+iu) R_C(s+iu) \widetilde{W_1}(iu) X^{iu} du + F_2(s) + O(X^{-10}),$$

where A, B and C range over the powers of two such that $A \in (\frac{1}{2}P_2, Q_2]$, $B \in (\frac{1}{2}P_3, Q_3]$ and $ABC \in [\frac{1}{16}X, 4X]$, and where

$$\begin{cases} Q_{j,D}(s) := \displaystyle\sum_{p\in(D,2D]\cap(P_j,Q_j]} \frac{c_p}{p^s}, \\ R_D(s) := \displaystyle\sum_{D<m\leqslant 2D} \frac{r_m}{m^s}, \\ F_2(s) := \displaystyle\sum_{X/2<n\leqslant 4X} \frac{e_n}{n^s} \end{cases}$$

for some complex coefficients r_m, e_n satisfying $|r_m| \leqslant 1$ and $\sum_n |e_n|^2 \ll X(\log X)^4/P_2$.

Proof. Write $\omega_{(P_j,Q_j]}(n) := \sum_{p\in(P_j,Q_j]} \mathbf{1}_{p|n}$. The assumptions on (a_n) (see Notation C.1.2) imply that

$$F(s) = \sum_{p\in(P_2,Q_2]} \sum_{q\in(P_3,Q_3]} \sum_{\substack{m\geqslant 1\\(m,pq)=1}} \frac{b'_m c_p c_q}{(mpq)^s \omega_{(P_2,Q_2]}(mpq)\omega_{(P_3,Q_3]}(mpq)} W_1\Big(\frac{mpq}{X}\Big)$$

for some 1-bounded complex sequence (b'_m). Thus, letting

$$r_m := \frac{b'_m}{(\omega_{(P_2,Q_2]}(m)+1)(\omega_{(P_3,Q_3]}(m)+1)},$$

we have

$$F(s) = \sum_{p\in(P_2,Q_2]} \sum_{q\in(P_3,Q_3]} \sum_{\substack{m\geqslant 1\\(m,pq)=1}} \frac{r_m c_p c_q}{(mpq)^s} W_1\Big(\frac{mpq}{X}\Big).$$

The condition $(m,pq)=1$ can be dropped, at the expense of an extra term of the form

$$\sum_{n\geqslant 1} \frac{e_n}{n^s}$$

for some coefficients $|e_n| \ll (\log X)^2$ supported on the integers $n \in (\frac{1}{2}X, 4X]$ having a repeated prime factor in $(P_2,Q_2]\cup(P_3,Q_3]$. Note that the number of such integers n is

$$\ll \sum_{P_2<p\leqslant Q_3} \sum_{X/2<n\leqslant 4X} \mathbf{1}_{p^2|n} \ll X \sum_{P_2<p\leqslant Q_3} \frac{1}{p^2} \ll \frac{X}{P_2}.$$

Therefore, by dyadic partitioning,

$$F(s) = \sum_{A,B,C} \sum_{\substack{p\in(P_2,Q_2]\\p\in(A,2A]}} \sum_{\substack{q\in(P_3,Q_3]\\q\in(B,2B]}} \sum_{m\in(C,2C]} \frac{r_m c_p c_q}{(mpq)^s} W_1\Big(\frac{mpq}{X}\Big) + \sum_{n\geqslant 1} \frac{e_n}{n^s}$$

where A,B,C range over the set of powers of two such that $ABC \in [\frac{1}{16}X, 4X]$. The conclusion follows by Mellin inversion and truncation, as in the proof of Lemma C.1.4. □

C.2 Main part of the Matomäki-Radziwiłł proof

Proposition C.2.1 below is the version of [36, Proposition 8.3] that we require. As noted in Remark C.2.3, treating [36, Proposition 8.3] as a black box would be insufficient for our purposes (even though it is more general than Proposition C.2.1 in other respects). By extracting from its proof only those ideas relevant to our purposes, we avoid some technical complications.

Proposition C.2.1. *Let $X, P_j, Q_j, (a_n), (b_m)$ and (c_p) be as in Notation C.1.2.*

Let $1 \leqslant y \leqslant X^{1/2}$. Let $\mathcal{U} \subset [-X, X]$ be any measurable set and $\nu > 0$. Then, there exists a measurable function $g : [X, 2X] \to \mathbb{C}$ such that

$$\frac{1}{X}\int_X^{2X} \Big| g(x) - \sum_{x<n\leqslant x+y} a_n \Big|^2 dx \ll E_1 + E_2 + E_3$$

where

$$\begin{cases} E_1 := (\log X)^2 + \dfrac{y^2(\log X)^5}{P_2}, \\ E_2 := \max\limits_{X/y\leqslant T\leqslant X} \dfrac{Xy}{T} \displaystyle\int_{\substack{[-T,T] \\ \mathrm{dist}(t,\mathcal{U})\geqslant X^{1/10}}} \Big| \sum_{n\geqslant 1} \frac{a_n}{n^{1+it}} W_2\Big(\frac{n}{X}\Big)\Big|^2 dt, \\ E_3 := \dfrac{y^2(\log X)^{11}}{X^{2\nu}} \Big(1 + \dfrac{Q_2 \cdot \big|\{t \in [-X, X] \,:\, \mathrm{dist}(t, \mathcal{U}) \leqslant 4X^{1/10}\}\big|}{(Xy)^{1/2}}\Big), \end{cases}$$

and g satisfies

$$\|g\|_\infty \ll y \sum_{A,B} \int_{\mathcal{V}} |Q_{2,A}(1+it)Q_{3,B}(1+it)|dt \tag{C.7}$$

with

$$\mathcal{V} := \Big\{ t \in [-X, X] \,:\, \max_A |Q_{2,A}(1+it)| \geqslant X^{-\nu} \Big\}. \tag{C.8}$$

In (C.7) and (C.8), the variables A and B range over the sets of powers of two in $(\frac{1}{2}P_2, Q_2]$ and $(\frac{1}{2}P_3, Q_3]$, respectively. The Dirichlet polynomials $Q_{2,A}$ and $Q_{3,B}$ are those defined in Lemma C.1.5.

Remark C.2.2. Note that E_2 and E_3 depend on $\mathcal{U}$, while E_3 and $\|g\|_\infty$ depend on ν. We will later choose $\mathcal{U}$ and ν (as well as the intervals $(P_j, Q_j]$) to ensure that all of these quantities are suitably small.

Remark C.2.3. A key difference between Proposition C.2.1 and [36, Proposition 8.3] is that we do not apply a Halász-Montgomery type estimate to bound $\|g\|_\infty$ at this stage. This allows us to exploit an additional averaging over Dirichlet characters later on.

The following lemma is the first step towards Proposition C.2.1.

Lemma C.2.4. *Let $X, P_j, Q_j, (a_n), (b_m), (c_p), y$ and $\mathcal{U}$ be as in Proposition C.2.1.*

Let $J \subset [-X, X]$ be any measurable set. We have

$$\frac{1}{X}\int_X^{2X} \Big| g(x) - \sum_{x<n\leqslant x+y} a_n \Big|^2 dx \ll E_1 + E_2 + E_3^*$$

with E_1, E_2 as in Proposition C.2.1 and

$$E_3^* := \max_{\substack{A,B,C \\ X/y \leqslant T \leqslant X}} \frac{X(\log X)^{10} y}{T} \int_{\substack{[-T,T]\setminus J \\ \mathrm{dist}(t,\mathcal{U}) \leqslant 4X^{1/10}}} |Q_{2,A}(1+it)Q_{3,B}(1+it)R_C(1+it)|^2 dt,$$

for some measurable function g satisfying the uniform bound

$$\|g\|_\infty \ll y \sum_{A,B} \int_J |Q_{2,A}(1+it)Q_{3,B}(1+it)| dt.$$

In these expressions, the variables A, B, C and the Dirichlet polynomials $Q_{2,A}, Q_{3,B}$ and R_C are defined to be as in Lemma C.1.5.

Proof. By Perron's formula (e.g. take $\alpha = 0$, $c = 1$, $T = X/2$ and $s \to 0^+$ in [23, Lemma 1.1 (p.11)]),

$$\sum_{x<n\leqslant x+y} a_n = \sum_{x<n\leqslant x+y} a_n W_1\Big(\frac{n}{X}\Big) = \frac{1}{2\pi}\int_{-X/2}^{X/2} F(1+it)\frac{(x+y)^{1+it} - x^{1+it}}{1+it} dt + O(\log X). \tag{C.9}$$

Our goal is to write

$$\sum_{x<n\leqslant x+y} a_n = g(x) + h(x)$$

with g small in $L^\infty([X,2X])$ and h small in $L^2([X,2X])$ (more precisely, we mean that $\frac{1}{X}\int_X^{2X} |h(x)|^2 dx \ll E_1 + E_2 + E_3^*$). We will establish this decomposition by successively extracting terms from the formula (C.9) and demonstrating that their contribution in $L^2([X,2X])$ is acceptable. These terms will be absorbed into $h(x)$, while what remains after this process will define $g(x)$.

First of all, the error term $O(\log X)$ from (C.9) can clearly be absorbed into $h(x)$.

Let $I := \big\{t \in [-X/2, X/2] : \mathrm{dist}(t,\mathcal{U}) \geqslant 2X^{1/10}\big\}$. By Lemma C.1.1 applied to the function $f(t) := F(1+it)\mathbf{1}_I(t)$, we have

$$\frac{1}{X}\int_X^{2X} \left|\int_I F(1+it)\frac{(x+y)^{1+it} - x^{1+it}}{1+it} dt\right|^2 dx \ll \max_{\frac{X}{y}\leqslant T\leqslant \frac{X}{2}} \frac{Xy}{T} \int_{I\cap[-T,T]} |F(1+it)|^2 dt.$$

By Lemma C.1.4, Cauchy-Schwarz and (C.5), for $X/y \leqslant T \leqslant X/2$ we have

$$\int_{I\cap[-T,T]} |F(1+it)|^2 dt \ll \int_{I\cap[-T,T]} \int_{-(\log X)^3}^{(\log X)^3} \left|\sum_{n\geqslant 1} \frac{a_n W_2(n/X)}{n^{1+it+iu}}\right|^2 \left|\widetilde{W_1}(iu)\right| du\, dt + \frac{1}{X^{19}}.$$

Swapping the order of integration, performing the change of variables $r := t + u$ and integrating over u, this is

$$\ll \int_{\substack{[-2T,2T] \\ \mathrm{dist}(r,\mathcal{U})\geqslant X^{1/10}}} \left|\sum_{n\geqslant 1} \frac{a_n}{n^{1+ir}} W_2\Big(\frac{n}{X}\Big)\right|^2 dr + \frac{1}{X^{19}}$$

where we used the definition of I and the fact that $(\log X)^3 < \min(X^{1/10}, T)$. This is an acceptable contribution (see the definition of E_2 in Proposition C.2.1). In other words, the contribution corresponding to integration over I in (C.9) can be absorbed into $h(x)$.

We now use the formula for $F(1+it)$ given in Lemma C.1.5. The error term $O(X^{-10})$ of that formula can trivially be absorbed into $h(x)$.

Let us consider the term $F_2(1+it)$ appearing in Lemma C.1.5. By Lemma C.1.1,

$$\frac{1}{X}\int_X^{2X}\left|\frac{1}{2\pi}\int_{[-X/2,X/2]\setminus I}F_2(1+it)\frac{(x+y)^{1+it}-x^{1+it}}{1+it}dt\right|^2 dx \\ \ll \max_{T\geqslant X/y}\frac{Xy}{T}\int_{-T}^{T}|F_2(1+it)|^2dt.$$

By the mean-value theorem [30, Theorem 9.1], this is

$$\ll \max_{T\geqslant X/y}\frac{Xy}{T}(T+X)\sum_n\left|\frac{e_n}{X}\right|^2\log T \ll \max_{T\geqslant X/y}\frac{Xy}{T}(T+X)\frac{(\log X)^5}{XP_2} \ll \frac{y^2(\log X)^5}{P_2}.$$

Thus, the contribution of $F_2(1+it)$ can be absorbed into $h(x)$, by definition of E_1.

Hence, up to some terms that are suitably bounded in $L^2([X,2X])$, the sum $\sum_{x<n\leqslant x+y}a_n$ equals

$$\frac{1}{2\pi}\sum_{A,B,C}\int_{-(\log X)^3}^{(\log X)^3}\int_{-X/2}^{X/2}G_{A,B,C}(t,u)\frac{(x+y)^{1+it}-x^{1+it}}{1+it}dt\,du \tag{C.10}$$

where the range of A, B, C is the same as in Lemma C.1.5 and $G_{A,B,C}(t,u)$ is the function defined by

$$\mathbf{1}_{\{\mathrm{dist}(t,\mathcal{U})<2X^{1/10}\}}Q_{2,A}(1+it+iu)Q_{3,B}(1+it+iu)R_C(1+it+iu)\widetilde{W_1}(iu)X^{iu}.$$

We split (C.10) based on whether $t+u\in J$ or not (recall that J is the set given in the statement of Lemma C.2.4).

The contribution of $t+u\notin J$ can be absorbed into $h(x)$. Indeed, by Cauchy-Schwarz and Lemma C.1.1 we have

$$\frac{1}{X}\int_X^{2X}\left|\sum_{A,B,C}\int_{-(\log X)^3}^{(\log X)^3}\int_{-X/2}^{X/2}\mathbf{1}_{t+u\notin J}G_{A,B,C}(t,u)\frac{(x+y)^{1+it}-x^{1+it}}{1+it}dt\,du\right|^2 dx \\ \ll (\log X)^5\sum_{A,B,C}\int_{-(\log X)^3}^{(\log X)^3}\max_{X/y\leqslant T\leqslant X/2}\frac{Xy}{T}\int_{-T}^{T}|\mathbf{1}_{t+u\notin J}G_{A,B,C}(t,u)|^2dt\,du.$$

Using the trivial bound $\widetilde{W_1}(iu)X^{iu}\ll 1$ and changing variables $r:=t+u$, this becomes

$$\ll (\log X)^{10}\max_{\substack{A,B,C\\ X/y\leqslant T\leqslant X/2}}\frac{Xy}{T}\int_{-2T}^{2T}\mathbf{1}_{\{\mathrm{dist}(r,\mathcal{U})\leqslant 4X^{1/10}\}}\mathbf{1}_{r\notin J}|(Q_{2,A}Q_{3,B}R_C)(1+ir)|^2dr,$$

which is an acceptable contribution of the form E_3^*.

We can finally define $g(x)$ to be the remaining expression

$$g(x) := \frac{1}{2\pi} \sum_{A,B,C} \int_{-(\log X)^3}^{(\log X)^3} \int_{-X/2}^{X/2} \mathbf{1}_{t+u\in J} G_{A,B,C}(t,u) \frac{(x+y)^{1+it} - x^{1+it}}{1+it} dt\, du.$$

By the triangle inequality, along with the trivial bounds $R_C(1+it+iu) \ll 1$ and (C.2), we have

$$\|g\|_\infty \ll y \sum_{A,B} \int_{-(\log X)^3}^{(\log X)^3} \int_{-X/2}^{X/2} \mathbf{1}_{t+u\in J} \Big| Q_{2,A}(1+it+iu) Q_{3,B}(1+it+iu) \widetilde{W_1}(iu) \Big| dt\, du.$$

Changing variables $r := t+u$ and integrating over u yields

$$\|g\|_\infty \ll y \sum_{A,B} \int_{-X}^{X} \mathbf{1}_{r\in J} |Q_{2,A}(1+ir) Q_{3,B}(1+ir)| dr,$$

which is the claimed bound for $\|g\|_\infty$. □

We can now prove Proposition C.2.1. To do so, we need to select an appropriate set J in Lemma C.2.4; we choose it to be the set of all $t \in [-X, X]$ where one of the Dirichlet polynomials $Q_{2,A}(1+it)$ is large.

Proof of Proposition C.2.1. We apply Lemma C.2.4 with J being the set $\mathcal{V}$ defined in (C.8). The announced bound for $\|g\|_\infty$ is immediate.

It remains to treat the term E_3^*. Using the bound for $Q_{2,A}$ in the definition of $\mathcal{V}$, we have

$$E_3^* \ll \frac{X(\log X)^{10} y}{X^{2\nu}} \max_{\substack{A,B,C \\ X/y \leqslant T \leqslant X}} \frac{1}{T} \int_{\substack{[-T,T] \\ \operatorname{dist}(t,\mathcal{U}) \leqslant 4X^{1/10}}} |Q_{3,B}(1+it) R_C(1+it)|^2 dt. \tag{C.11}$$

For any 1-separated set $S \subset [-T,T]$, the Halász-Montgomery inequality [30, Theorem 9.6] gives

$$\sum_{t\in S} |Q_{3,B}(1+it) R_C(1+it)|^2 \ll \frac{1}{BC} \big(BC + |S| T^{1/2}\big) \log T.$$

Applying this to bound the integral in (C.11) (after discretising), we deduce that

$$E_3^* \ll \frac{X(\log X)^{11} y}{X^{2\nu}} \max_{\substack{A,B,C \\ X/y \leqslant T \leqslant X}} \frac{1}{T} \left(1 + \frac{\big|\{t \in [-T,T] \,:\, \operatorname{dist}(t,\mathcal{U}) \leqslant 4X^{1/10}\}\big| \cdot T^{1/2}}{BC}\right).$$

Observing that $(BC)^{-1} \asymp A/X \ll Q_2/X$ and treating T trivially, we conclude that $E_3^* \ll E_3$. □

Next, we choose the set $\mathcal{U}$ in Proposition C.2.1 to obtain a good bound for the quantities E_2 and E_3. This step corresponds to [36, Theorem 9.2], but is technically much simpler in our context.

Proposition C.2.5. *Let $\nu > 0$ be sufficiently small.*

Let $X \geqslant 10^{100}$ and $10 \leqslant P_1 < Q_1 \leqslant P_2 < Q_2 \leqslant P_3 < Q_3 \leqslant X^{1/3}$. Let $(a_n), (b_m)$ and (c_p) be as in Notation C.1.2.

Let $1 \leqslant y \leqslant X^{1/2}$. Suppose that $P_1 \geqslant (\log X)^{1/\nu}$. Then

$$\frac{1}{X}\int_X^{2X} \Big| g(x) - \sum_{x<n\leqslant x+y} a_n \Big|^2 dx \ll \frac{y(y+Q_1)(\log X)^{11}}{P_1^{2\nu}} + (\log X)^2$$

for some measurable function g obeying the uniform bound (C.7).

Proof. We apply Proposition C.2.1. Recall that

$$E_2 := \max_{X/y\leqslant T\leqslant X} \frac{Xy}{T} \int_{\substack{[-T,T]\\ \operatorname{dist}(t,\mathcal{U})\geqslant X^{1/10}}} \Big| \sum_{n\geqslant 1} \frac{a_n}{n^{1+it}} W_2\Big(\frac{n}{X}\Big)\Big|^2 dt. \tag{C.12}$$

Taking out one prime from the interval $(P_1, Q_1]$ exactly as in the proof of Lemma C.1.5, we obtain the identity

$$\sum_{n\geqslant 1} \frac{a_n}{n^s} W_2\Big(\frac{n}{X}\Big) = \sum_{D,E} \frac{1}{2\pi} \int_{-(\log X)^3}^{(\log X)^3} (Q_{1,D} R'_E)(s+iu) \widetilde{W}_2(iu) X^{iu} du + F'_2(s) + O(X^{-10}),$$

where D and E range over the powers of two such that $\frac{1}{16}X \leqslant DE \leqslant 8X$ and $D \in (\frac{1}{2}P_1, Q_1]$, and where

$$\begin{cases} Q_{1,D}(s) := \displaystyle\sum_{p\in(D,2D]\cap(P_1,Q_1]} \frac{c_p}{p^s}, \\ R'_E(s) := \displaystyle\sum_{E<m\leqslant 2E} \frac{r'_m}{m^s}, \\ F'_2(s) := \displaystyle\sum_{X/4<n\leqslant 8X} \frac{e'_n}{n^s} \end{cases}$$

for some complex coefficients r'_m, e'_n satisfying $|r'_m| \leqslant 1$ and $\sum_n |e'_n|^2 \ll X(\log X)^2/P_1$.

As in the proof of Lemma C.2.4, the error term $O(X^{-10})$ is negligible, and the contribution of $F'_2(1+it)$ to (C.12) is

$$\ll \max_{T\geqslant X/y} \frac{Xy}{T}(T+X) \sum_n \Big|\frac{e'_n}{X}\Big|^2 \log T \ll \max_{T\geqslant X/y} \frac{Xy}{T}(T+X)\frac{(\log X)^3}{XP_1} \ll \frac{y^2(\log X)^3}{P_1}.$$

by the mean-value theorem [30, Theorem 9.1].

For the main term of the identity, we use Cauchy-Schwarz and change variables as in the proof of Lemma C.2.4. This gives

$$E_2 \ll \max_{\substack{X/y\leqslant T\leqslant X\\ D,E}} \frac{X(\log X)^8 y}{T}\int_{[-2T,2T]\setminus\mathcal{U}} |(Q_{1,D}R'_E)(1+ir)|^2 dr + \frac{y^2(\log X)^3}{P_1}. \tag{C.13}$$

Defining

$$\mathcal{U} := \Big\{t\in[-2X,2X] \,:\, \max_D |Q_{1,D}(1+it)| \geqslant P_1^{-\nu}\Big\}, \tag{C.14}$$

we have

$$\int_{[-2T,2T]\setminus\mathcal{U}} |Q_{1,D}(1+ir)R'_E(1+ir)|^2 dr \leqslant P_1^{-2\nu}\int_{[-2T,2T]} |R'_E(1+ir)|^2 dr.$$

By the mean-value theorem and the bound $E \asymp X/D \gg X/Q_1$,

$$\begin{aligned}\int_{[-2T,2T]} |R'_E(1+ir)|^2 dr &\ll (T+E)\sum_m \Big|\frac{r'_m}{E}\Big|^2 \log T\\ &\ll \Big(\frac{T}{E}+1\Big)\log X \ll \Big(\frac{TQ_1}{X}+1\Big)\log X.\end{aligned}$$

Hence, the estimate (C.13) becomes

$$E_2 \ll \max_{X/y\leqslant T\leqslant X} \frac{X(\log X)^9 y}{P_1^{2\nu}}\Big(\frac{Q_1}{X}+\frac{1}{T}\Big) + \frac{y^2(\log X)^3}{P_1} \ll \frac{y(Q_1+y)(\log X)^9}{P_1^{2\nu}},$$

which is acceptable.

It remains to give a suitable bound for E_3, and for this it is enough to show that

$$\frac{Q_2\cdot\big|\{t\in[-X,X] \,:\, \operatorname{dist}(t,\mathcal{U})\leqslant 4X^{1/10}\}\big|}{(Xy)^{1/2}} \ll 1. \tag{C.15}$$

Let S be an arbitrary 1-separated subset of $\mathcal{U}$. We bound the size of S using the large value estimate [35, Lemma 8] (which is a simple consequence of the mean-value theorem, see also [30, Eq. (9.30), p.236]). We get

$$|S| \ll \sum_D \exp\Big(2\nu\frac{(\log X)(\log P_1)}{\log D} + 2\nu\log P_1 + 2\frac{\log X}{\log D}\log\log X\Big),$$

where D ranges over the powers of two in $(\frac{1}{2}P_1, Q_1]$. Using our assumption that $\log X \leqslant P_1^{\nu}$, we have $\log\log X \ll \nu\log D$ for any such D. Therefore, we have the simple bound

$$|S| \ll X^{O(\nu)}\log X.$$

Choosing S to be a maximal 1-separated subset of $\mathcal{U}$, we conclude that

$$\big|\{t\in[-X,X] \,:\, \operatorname{dist}(t,\mathcal{U})\leqslant 4X^{1/10}\}\big| \ll X^{1/10+O(\nu)}\log X.$$

Since $Q_2 \leqslant X^{1/3}$, this establishes (C.15) (in a strong form) if ν is sufficiently small, which concludes the proof of Proposition C.2.5. □

C.3 Average over Dirichlet characters

As mentioned in Remark C.2.3, our application of Proposition C.2.5 involves an extra average over Dirichlet characters. The following estimate will then be used to control the $g(x)$ term on average (compare with (C.7)).

Lemma C.3.1. *Let X be sufficiently large, and let $(\log X)^{-1} \leqslant \nu \leqslant 10^{-8}$.*

Let $X^{\nu^{1/8}} \leqslant P_2 < Q_2 < P_3 < Q_3 \leqslant X^{1/3}$. Let q be a positive integer such that $q \leqslant X^{\nu}$.

For a Dirichlet character $\chi \,(\mathrm{mod}\, q)$, $D > 0$ and $j \in \{2, 3\}$, define

$$Q_{j,D}(s, \chi) := \sum_{p \in (D, 2D] \cap (P_j, Q_j]} \frac{\chi(p) c_p}{p^s},$$

where (c_p) is a sequence of 1-bounded complex coefficients supported on primes.

Then

$$\sum_{\chi \,(\mathrm{mod}\, q)} \sum_{A,B} \int_{\mathcal{V}(\chi)} |Q_{2,A}(1+it, \chi) Q_{3,B}(1+it, \chi)| dt \ll \left(1 + \log \frac{\log Q_3}{\log P_2}\right)^2, \tag{C.16}$$

where

$$\mathcal{V}(\chi) := \left\{ t \in [-X, X] \,:\, \max_{A} |Q_{2,A}(1+it, \chi)| \geqslant X^{-\nu} \right\}$$

and, in these expressions, A and B range over the powers of two in $(\frac{1}{2}P_2, Q_2]$ and $(\frac{1}{2}P_3, Q_3]$, respectively.

The proof of Lemma C.3.1 follows [40, p.92], which itself generalises some ideas from the proof of [36, Proposition 8.3] to the case $q > 1$.

Proof. It suffices to prove that

$$\sum_{\chi \,(\mathrm{mod}\, q)} \sum_{A,B} \sum_{t \in S_\chi} |Q_{2,A}(1+it, \chi) Q_{3,B}(1+it, \chi)| \ll \left(1 + \log \frac{\log Q_3}{\log P_2}\right)^2 \tag{C.17}$$

where, for every $\chi \,(\mathrm{mod}\, q)$, S_χ is a 1-separated subset of $\mathcal{V}(\chi)$. Define

$$M_{j,D} := \sum_{\chi \,(\mathrm{mod}\, q)} \sum_{t \in S_\chi} |Q_{j,D}(1+it, \chi)|^2$$

for $j = 2, 3$ and $D > 10$. By Cauchy-Schwarz, the left-hand side of (C.17) is

$$\leqslant \left(\sum_A \frac{1}{\log A}\right)^{1/2} \left(\sum_A (\log A) M_{2,A}\right)^{1/2} \left(\sum_B \frac{1}{\log B}\right)^{1/2} \left(\sum_B (\log B) M_{3,B}\right)^{1/2}.$$

We will show that each of these terms is $\ll (1 + \log \frac{\log Q_3}{\log P_2})^{1/2}$. This is clear for the first and third terms.

To bound the other terms, we use [40, Lemma 6.6][1] which implies that, uniformly for $\eta \in (0, 1/2)$,

$$M_{j,D} \ll \frac{1}{(\log D)^2} + \frac{q^\eta D^{-\eta/2}}{\log D}\Big(X^{5\eta^{3/2}} \log X + \eta^{-1}\Big) \sum_{\chi \,(\mathrm{mod}\, q)} |S_\chi|$$

for $j = 2, 3$ and $D \geqslant 10$. In particular,

$$\max\Big(\sum_A (\log A) M_{2,A},\ \sum_B (\log B) M_{3,B}\Big) \ll 1 + \log\frac{\log Q_3}{\log P_2} + q^\eta P_2^{-\eta/2} X^{5\eta^{3/2}} (\log X)^2 \eta^{-1} \sum_{\chi \,(\mathrm{mod}\, q)} |S_\chi|.$$

On the other hand, by definition of $\mathcal{V}(\chi)$, we have

$$\sum_A (\log A) M_{2,A} \geqslant \sum_A M_{2,A} \geqslant X^{-2\nu} \sum_{\chi \,(\mathrm{mod}\, q)} |S_\chi|.$$

We choose $\eta := \nu^{1/2}$. Since $(\log X)^{-1} \leqslant \nu \leqslant 10^{-8}$, this ensures that

$$X^{-2\nu} > q^\eta P_2^{-\eta/2} X^{5\eta^{3/2}} (\log X)^3 \eta^{-1}$$

if X is sufficiently large, using the bounds for P_2 and q in the statement (the term $P_2^\eta \geqslant X^{\nu^{5/8}}$ dominates). Thus, with this choice of η, we must have

$$\max\Big(\sum_A (\log A) M_{2,A},\ \sum_B (\log B) M_{3,B}\Big) \ll 1 + \log\frac{\log Q_3}{\log P_2}$$

as desired. □

C.4 Rewriting correlations in terms of characters

Suppose that a multiplicative function $f(n)$ correlates with a linear phase $e(n\alpha_x)$ over many short intervals $[x, x + H)$, where $\alpha_x \approx \frac{a}{q} + \frac{T}{x}$. We show that this forces f to correlate with $\chi(n)n^{2\pi iT}$ for many Dirichlet characters $\chi \,(\mathrm{mod}\, q)$, over many slightly shorter intervals. This is done in two steps, Lemmas C.4.1 and C.4.2, which are heavily inspired by [40, p.40] and [40, p.90-91] respectively.

Lemma C.4.1. *Let $1 \leqslant H' \leqslant H \leqslant X$, $\delta > 0$ and $\kappa \geqslant 1$. Assume that $H' \leqslant c_0 \delta H/\kappa$ for some sufficiently small absolute constant $c_0 > 0$.*

[1] The inequality [40, Lemma 6.6] corresponds to the case $k = 1$ of Lemma 3.4.1, except that it accounts for the sparsity of the primes. It can be obtained by inserting a linear sieve upper bound in the proof of Lemma 3.4.1, as in [36, Proof of Lemma 4.4].

Let (b_n) be a 1-bounded sequence of complex numbers. Suppose that there exists an H-separated set $S \subset [X, 2X]$ such that, for every $x \in S$,

$$\Big| \sum_{x \leqslant n < x+H} b_n e(n\beta_x) \Big| \geqslant \delta H, \tag{C.18}$$

where the frequencies β_x satisfy $\big\|\beta_x - \frac{T}{x}\big\| \leqslant \frac{\kappa}{H}$ for some $T \in \mathbb{R}$ with $|T| \leqslant \kappa X^2/H^2$.

Then, there is an H'-separated set $S' \subset [X, 2X]$ of size $\gg \delta \frac{H}{H'}|S|$ such that, for every $x \in S'$,

$$\Big| \sum_{x \leqslant n < x+H'} b_n n^{2\pi i T} \Big| \gg \delta H'.$$

Proof. Fix $x \in S$. For any integer $0 \leqslant h < H'$, we have

$$\sum_{x \leqslant n < x+H} b_{n+h} e((n+h)\beta_x) = \sum_{x \leqslant n < x+H} b_n e(n\beta_x) + O(H').$$

Averaging over h and applying the triangle inequality gives

$$\frac{1}{H'} \sum_{x \leqslant n < x+H} \Big| \sum_{0 \leqslant h < H'} b_{n+h} e((n+h)\beta_x) \Big| \geqslant \delta H - O(H') \geqslant \tfrac{1}{2}\delta H,$$

provided c_0 is chosen sufficiently small. It follows that there are $\gg \delta H$ integers $n \in [x, x+H)$ such that

$$\Big| \sum_{0 \leqslant h < H'} b_{n+h} e((n+h)\beta_x) \Big| \geqslant \tfrac{1}{4}\delta H'. \tag{C.19}$$

From these integers, one may select an H'-separated subset of $[x, x+H)$ of size $\gg \delta H/H'$ on which the above inequality holds. Since this is true for all $x \in S$ and S is H-separated, we deduce that there exists an H'-separated set $S' \subset [X, 2X]$ of size $\gg \delta \frac{H}{H'}|S|$ such that (C.19) holds for every $n \in S'$.

Note that for $0 \leqslant h < H'$ and $n \in [x, x+H)$, we have

$$e(h\beta_x) = e\Big(h\Big(\beta_x - \frac{T}{x}\Big)\Big) e\Big(h\Big(\frac{T}{x} - \frac{T}{n}\Big)\Big) e\Big(\frac{hT}{n}\Big) = \Big(1 + O\Big(\frac{\kappa H'}{H}\Big)\Big) e\Big(\frac{hT}{n}\Big)$$

using that $|T| \leqslant \kappa X^2/H^2$. Since $\kappa H' \leqslant c_0 \delta H$ for sufficiently small $c_0 > 0$, we deduce from (C.19) that

$$\Big| \sum_{0 \leqslant h < H'} b_{n+h} e\Big(\frac{hT}{n}\Big) \Big| \geqslant \tfrac{1}{8}\delta H'$$

for every $n \in S'$. By Taylor expansion,

$$(n+h)^{2\pi i T} = n^{2\pi i T}\Big(1 + \frac{h}{n}\Big)^{2\pi i T} = n^{2\pi i T} e\Big(\frac{hT}{n}\Big) + O\Big(\frac{\kappa (H')^2}{H^2}\Big),$$

and since $\kappa(H')^2/H^2 \ll \delta$, it follows that

$$\left| \sum_{0 \leqslant h < H'} b_{n+h} (n+h)^{2\pi i T} \right| \gg \delta H'$$

for every $n \in S'$, as required. □

Lemma C.4.2. *Let $1 \leqslant H \leqslant X$. Let $q \in \mathbb{N}$ and $a \in \mathbb{Z}$ be such that $(a,q) = 1$. Let $f : \mathbb{N} \to \mathbb{C}$ be a 1-bounded function such that $f(mn) = f(m)f(n)$ whenever $(m,n) = 1$ and $\mathrm{rad}(m) \mid q$.*

Let $\varepsilon, \delta > 0$ and assume that $C_0^{\delta^{-1}} q^2 \leqslant H$ for some large enough constant C_0 depending on ε. Suppose that

$$\left| \sum_{x \leqslant n < x+H} f(n) e\Big(\frac{na}{q}\Big) \right| \geqslant \delta H \tag{C.20}$$

for all x in an H-separated set $S \subset [X, 2X]$.

Then, there exists a positive integer $d \ll_\varepsilon \delta^{-2} q^\varepsilon$ and an H/d-separated subset $S' \subset [X/d, 2X/d]$ of size $|S'| \gg_\varepsilon \delta |S|$ such that

$$\sum_{\chi \,(\mathrm{mod}\, q)} \left| \sum_{x \leqslant n < x+H/d} f(n)\chi(n) \right| \gg_\varepsilon \delta^3 q^{1/2-\varepsilon} \frac{H}{d}$$

for all $x \in S'$.

Proof. We begin by excluding integers with unusually large prime power divisors coming from primes dividing q. Let $A := C(\varepsilon)\delta^{-1}$, where $C(\varepsilon) \geqslant 1$ is a sufficiently large constant to be chosen later.

Let E be the set of integers divisible by p^k for some prime $p \mid q$ and exponent $k \geqslant 2$ such that $p^k \geqslant A^2$. For any $x \in [X, 2X]$, we have

$$|E \cap [x, x+H)| \leqslant \sum_{p \mid q} \min_{\substack{k \geqslant 2 \\ p^k \geqslant A^2}} \left(\frac{H}{p^k} + 1 \right) \ll q + H \left(\sum_{p \leqslant A} \frac{1}{A^2} + \sum_{p > A} \frac{1}{p^2} \right) \ll q + \frac{H}{A}.$$

If $C(\varepsilon)$ is chosen sufficiently large, then for every $x \in S$, the hypothesis (C.20) implies

$$\left| \sum_{\substack{x \leqslant n < x+H \\ n \notin E}} f(n) e\Big(\frac{na}{q}\Big) \right| \gg \delta H. \tag{C.21}$$

Every $n \geqslant 1$ can be written uniquely as $n = dm$ with $(m,q) = 1$ and $\mathrm{rad}(d) \mid q$. Using the multiplicativity assumption on f, we may rewrite (C.21) as

$$\left| \sum_{\substack{d \notin E \\ \mathrm{rad}(d) \mid q}} f(d) \sum_{\substack{x/d \leqslant m < (x+H)/d \\ (m,q)=1}} f(m) e\Big(\frac{mad}{q}\Big) \right| \gg \delta H \tag{C.22}$$

for every $x \in S$.

We first bound the contribution of large d. By definition of E, for every $d \in \mathbb{N} \setminus E$ with $\mathrm{rad}(d) \mid q$, we have the upper bound

$$d \leqslant q \prod_{p \leqslant A} A^2 \leqslant q e^{O(A)},$$

and thus $d \leqslant H$ by the assumption $C_0^{\delta^{-1}} q^2 \leqslant H$ in the statement. Hence, the contribution of those d with $d \geqslant A^2 q^\varepsilon$ to the left-hand side of (C.22) is trivially bounded by

$$\sum_{\substack{A^2 q^\varepsilon \leqslant d \leqslant H \\ \mathrm{rad}(d) \mid q}} \frac{H}{d} \leqslant \frac{H}{A q^{\varepsilon/2}} \sum_{\substack{d \geqslant 1 \\ \mathrm{rad}(d) \mid q}} \frac{1}{d^{1/2}} = \frac{H}{A q^{\varepsilon/2}} \prod_{p \mid q} \frac{1}{1 - p^{-1/2}} \ll_\varepsilon \frac{H}{A}.$$

This is negligible by our choice of A if $C(\varepsilon)$ is sufficiently large.

Therefore, for every $x \in S$,

$$\left| \sum_{\substack{1 \leqslant d < A^2 q^\varepsilon \\ \mathrm{rad}(d) \mid q \\ d \notin E}} f(d) \sum_{\substack{x/d \leqslant m < (x+H)/d \\ (m,q)=1}} f(m) e\Big(\frac{mad}{q}\Big) \right| \gg \delta H.$$

Since the inner sum is trivially $\ll \frac{\varphi(q)}{q} \frac{H}{d}$ and

$$\sum_{\substack{d \geqslant 1 \\ \mathrm{rad}(d) \mid q}} \frac{1}{d} = \prod_{p \mid q} \frac{p}{p-1} = \frac{q}{\varphi(q)},$$

an averaging argument shows that there exists an integer d (with $1 \leqslant d < A^2 q^\varepsilon$) such that

$$\left| \sum_{\substack{x/d \leqslant m < (x+H)/d \\ (m,q)=1}} f(m) e\Big(\frac{mad}{q}\Big) \right| \gg \delta \frac{\varphi(q)}{q} \frac{H}{d} \tag{C.23}$$

for $\gg \delta |S|$ values of $x \in S$. Since S is H-separated, the corresponding values of x/d form an H/d-separated set $S' \subset [X/d, 2X/d]$ with $|S'| \gg \delta |S|$.

Expanding the sum into congruence classes modulo q and then into Dirichlet characters, we have

$$\sum_{\substack{y \leqslant m < y + H/d \\ (m,q)=1}} f(m) e\Big(\frac{mad}{q}\Big) = \frac{1}{\varphi(q)} \sum_{\chi \,(\mathrm{mod}\, q)} c_{\overline{\chi}}(ad) \sum_{y \leqslant m < y + H/d} f(m) \chi(m),$$

where

$$c_\chi(x) := \sum_{u \in (\mathbb{Z}/q\mathbb{Z})^\times} \chi(u) e\Big(\frac{ux}{q}\Big).$$

By [42, Theorem 9.12], writing $q' := q/(q,d)$, we have

$$|c_{\overline{\chi}}(ad)| \leqslant \frac{\varphi(q)\sqrt{q'}}{\varphi(q')}.$$

Combining this with (C.23), we obtain, for all $y \in S'$,

$$\sum_{\chi \,(\mathrm{mod}\, q)} \Big| \sum_{y \leqslant m < y+H/d} f(m)\chi(m) \Big| \gg \delta \frac{\varphi(q)}{q} \frac{\varphi(q')}{\sqrt{q'}} \frac{H}{d}.$$

Recalling that $d \ll_\varepsilon \delta^{-2} q^\varepsilon$, we see that $q' \gg_\varepsilon \delta^2 q^{1-\varepsilon}$. Since $\varphi(n) \gg n/\log\log 5n$ for $n \geqslant 1$, we deduce that

$$\sum_{\chi \,(\mathrm{mod}\, q)} \Big| \sum_{y \leqslant m < y+H/d} f(m)\chi(m) \Big| \gg_\varepsilon \delta^3 \frac{q^{(1-\varepsilon)/2}}{(\log\log 5q)^2} \frac{H}{d}$$

for all $y \in S'$, giving the result. □

C.5 From correlations in short intervals to pretentiousness

Finally, we bring together the results from the previous sections to prove Theorem C.5.1. We emphasize that all proof ideas originate from [36] and [40]; our objective here is to provide a comprehensive derivation, explicitly stating the quantitative aspects that were previously scattered or left to the reader to adapt.

Theorem C.5.1. *Let $f : \mathbb{N} \to \mathbb{C}$ be a 1-bounded multiplicative function.*

Let $10 \leqslant H \leqslant X$. Let $a, q \in \mathbb{N}$ be coprime integers. Let $T \in \mathbb{R}$ with $|T| \leqslant \kappa X^2/H^2$ for some $\kappa \geqslant 1$.

Let $0 < \delta < 1/2$. Suppose that

$$\Big| \sum_{x \leqslant n < x+H} f(n) e(n\alpha_x) \Big| \geqslant \delta H$$

for all x in an H-separated set $S \subset [X, 2X]$, where $\alpha_x \in \mathbb{R}/\mathbb{Z}$ satisfy the approximate formula

$$\alpha_x = \frac{a}{q} + \frac{T}{x} + O\Big(\frac{\kappa}{H}\Big) \pmod 1.$$

Assume that

$$\kappa^C \Big(q \frac{|S|}{X/H} \log X \Big)^{\delta^{-C}} \leqslant H \leqslant X^{\delta^C} \tag{C.24}$$

where C is a sufficiently large absolute constant.

Then, $q \ll \delta^{-10}$ and f is pretentious in the sense that

$$\sum_{p \leqslant X} \frac{1 - \mathrm{Re}(f(p)\chi(p)p^{it})}{p} \ll \log(\delta^{-1})$$

for some Dirichlet character $\chi \pmod q$ and some real number $t = O(\kappa X^2/H^2)$.

Theorem C.5.1 mainly relies on Proposition C.2.5, which requires restricting to integers n with prime factors in certain ranges. To do this, we use the following result from [36].

Lemma C.5.2. *Let $X > 0$ and $10 \leqslant P < Q \leqslant X^{3/4}$. Let $2 \leqslant H \leqslant \frac{1}{2}X^{1/6}$. Let $S \subset [X, 2X]$ be an H-separated set. Then, for all but*

$$\ll \frac{\log Q}{\log P} \cdot \frac{X(\log H)^2}{H^2}$$

elements $x \in S$, the number of integers $n \in [x, x + H)$ without any prime factor in $(P, Q]$ is

$$\ll \frac{\log P}{\log Q} H.$$

Proof. Let $E \subset \mathbb{N}$ be the set of integers without any prime factor in $(P, Q]$.

We apply [36, Proposition 10.4] with interval length $h := 2H$, taking f to be the constant function 1 and $\Delta := \log P/\log Q$. Note that $h_1 = H(f; X) = 1$ (see [36, p.7] for the definition of $H(f; X)$). We obtain that, for all but

$$\ll \frac{X(\log H)^2}{\Delta H}$$

integers $y \in [X, 2X]$, it holds that $|[y, y + 2H) \cap E| \ll \Delta H$.

Noting that $[x, x + H) \cap E \subset [y, y + 2H) \cap E$ for any $x \in S$ and $y \in [x - H, x]$, and recalling that S is H-separated, the conclusion follows. □

Proof of Theorem C.5.1. We first restrict to integers n with suitably sized prime factors. Let $c > 0$ be a small absolute constant to be chosen later. The constant C in the statement of the theorem is assumed to be sufficiently large in terms of c. Define $Q_1 = H^{1/2}$, $Q_2 = X^{c\delta/3}$, $Q_3 = X^{1/3}$ and, for $j \in \{1, 2, 3\}$, let $P_j := Q_j^{c\delta}$. Let $\mathcal{N}$ be the set of positive integers n having at least one prime factor in $(P_j, Q_j]$ for each $j \in \{1, 2, 3\}$, and no repeated prime factor in these ranges.

By Lemma C.5.2, for each $j \in \{1, 2, 3\}$, there are $\ll X(\log X)^3/H^2$ elements $x \in S$ such that

$$\sum_{x \leqslant n < x+H} \mathbf{1}_{p|n \Rightarrow p \notin (P_j, Q_j]} \geqslant \tfrac{1}{8}\delta H,$$

provided that c is sufficiently small.

Furthermore, the number of integers $n \in [X, 3X]$ divisible by p^2 for some prime $P_1 < p \leqslant Q_3$ is $\ll X/P_1$. Hence, at most $O(\delta^{-1}P_1^{-1}X/H)$ elements $x \in S$ have the property that $[x, x+H)$ contains $\geqslant \frac{1}{8}\delta H$ such integers n, since S is H-separated.

We have thus shown the existence of a subset $S_1 \subset S$ of size

$$|S_1| \geqslant |S| - O\Big(\frac{X(\log X)^3}{H^2} + \frac{X}{\delta P_1 H}\Big) \gg |S|,$$

(using (C.24)) such that all $x \in S_1$ satisfy

$$\Big|\sum_{x \leqslant n < x+H} \mathbf{1}_{n\in\mathcal{N}} f(n) e(n\alpha_x)\Big| \geqslant \tfrac{1}{2}\delta H.$$

Let $H' := \lfloor c\delta\kappa^{-1}H \rfloor$. By Lemma C.4.1 applied with $b_n := \mathbf{1}_{n\in\mathcal{N}} f(n)e(na/q)$ and $\beta_x := \alpha_x - \frac{a}{q}$, there is a subset $S_2 \subset S_1$ of size $|S_2| \gg \delta\frac{H}{H'}|S_1|$ such that, for every $x \in S_2$,

$$\Big|\sum_{x \leqslant n < x+H'} \mathbf{1}_{n\in\mathcal{N}} f(n) e\Big(\frac{na}{q}\Big) n^{2\pi i T}\Big| \gg \delta H'.$$

We apply Lemma C.4.2 with $f(n)$ replaced by $\mathbf{1}_{n\in\mathcal{N}} f(n) n^{2\pi i T}$, H replaced by H' and $\varepsilon := 1/6$. Note that the weak multiplicativity assumption of Lemma C.4.2 holds since $q < P_1$ (by (C.24)) and the definition of $\mathcal{N}$ only involves primes larger than P_1. Thus, there is an integer $1 \leqslant d \ll \delta^{-2}q^{1/6}$ and an H'/d-separated set $S_3 \subset [X/d, 2X/d]$ of size $|S_3| \gg \delta|S_2|$ such that

$$\sum_{\chi \,(\mathrm{mod}\, q)} \Big|\sum_{x \leqslant n < x+H'/d} \mathbf{1}_{n\in\mathcal{N}} f(n)\chi(n) n^{2\pi i T}\Big| \gg \delta^3 q^{1/3}\frac{H'}{d} \tag{C.25}$$

for all $x \in S_3$.

Let $\widetilde{X} := X/d$ and $\widetilde{H} := H'/d$. Define $\nu := (c\delta/3)^{16}$. For every $\chi \,(\mathrm{mod}\, q)$, the sequence

$$(a_n) := \big(\mathbf{1}_{n\in\mathcal{N}} f(n)\chi(n) n^{2\pi i T}\big)_{\widetilde{X}/4 < n \leqslant 8\widetilde{X}},$$

satisfies the properties in Notation C.1.2, by definition of $\mathcal{N}$ and multiplicativity of $f(n)\chi(n)n^{2\pi iT}$ (with $c_p = f(p)\chi(p)p^{2\pi iT}$). We can thus apply Proposition C.2.5 and Lemma C.3.1 to obtain a family of functions $g_\chi : [\widetilde{X}, 2\widetilde{X}] \to \mathbb{C}$ such that

$$\frac{1}{\widetilde{X}}\int_{\widetilde{X}}^{2\widetilde{X}} \Big|g_\chi(x) - \sum_{x < n \leqslant x+\widetilde{H}} \mathbf{1}_{n\in\mathcal{N}} f(n)\chi(n) n^{2\pi i T}\Big|^2 dx \ll \frac{\widetilde{H}(\widetilde{H}+Q_1)(\log X)^{11}}{P_1^{2\nu}} + (\log X)^2 \tag{C.26}$$

and

$$\sum_{\chi \,(\mathrm{mod}\, q)} \|g_\chi\|_\infty \ll \Big(1 + \log\frac{\log Q_3}{\log P_2}\Big)^2 \widetilde{H} \ll \big(\log\delta^{-1}\big)^2 \widetilde{H}. \tag{C.27}$$

Observe that the conditions $P_1 \geqslant (\log \widetilde{X})^{1/\nu}$, $P_2 \geqslant \widetilde{X}^{\nu^{1/8}}$, $q \leqslant \widetilde{X}^{\nu}$ and $\nu \geqslant 1/\log \widetilde{X}$ of Proposition C.2.5 and Lemma C.3.1 are all satisfied by the assumption (C.24) for our choice $\nu = (c\delta/3)^{16}$, if C is sufficiently large in terms of c.

Noting that $\widetilde{H} \gg c\delta^3 q^{-1/6}\kappa^{-1}H \gg Q_1$ by (C.24), the right-hand side of (C.26) is bounded by $O(\widetilde{H}^2(\log X)^{11}/P_1^{2\nu})$. By Cauchy-Schwarz, this implies that

$$\frac{1}{\widetilde{X}}\int_{\widetilde{X}}^{2\widetilde{X}} \sum_{\chi \,(\mathrm{mod}\, q)} \left| g_\chi(x) - \sum_{x<n\leqslant x+\widetilde{H}} \mathbf{1}_{n\in\mathcal{N}} f(n)\chi(n)n^{2\pi iT}\right| dx \ll q\frac{\widetilde{H}(\log X)^{11/2}}{P_1^{\nu}} \ll \frac{\widetilde{H}}{P_1^{\nu/2}}.$$

Consequently, we have

$$\sum_{\chi \,(\mathrm{mod}\, q)} \left| g_\chi(x) - \sum_{x<n\leqslant x+\widetilde{H}} \mathbf{1}_{n\in\mathcal{N}} f(n)\chi(n)n^{2\pi iT}\right| \leqslant \widetilde{H}$$

for all but $\ll P_1^{-\nu/2}\widetilde{X}$ integers $x \in [\widetilde{X}, 2\widetilde{X}]$. By (C.27), we conclude that

$$\sum_{\chi \,(\mathrm{mod}\, q)} \left| \sum_{x<n\leqslant x+\widetilde{H}} \mathbf{1}_{n\in\mathcal{N}} f(n)\chi(n)n^{2\pi iT}\right| \ll \big(\log \delta^{-1}\big)^2 \widetilde{H} \tag{C.28}$$

for all but $\ll P_1^{-\nu/2}\widetilde{X}$ integers $x \in [\widetilde{X}, 2\widetilde{X}]$.

On the other hand, by the triangle inequality, the lower bound (C.25) continues to hold whenever x is at distance at most $c\delta^3 q^{-2/3}\widetilde{H}$ from an element of S_3, if c is sufficiently small. Since S_3 is $\widetilde{H}$-separated, we conclude that (C.25) holds for

$$\gg |S_3|\delta^3 q^{-2/3}\widetilde{H} \gg \delta^5 q^{-2/3}\frac{|S|}{X/H}\widetilde{X}$$

integers $x \in [\widetilde{X}, 2\widetilde{X}]$. By our assumption (C.24), this contradicts (C.28) if C is sufficiently large, unless $\delta^3 q^{1/3} \ll (\log \delta^{-1})^2$. We have thus proved the key estimate $q \ll \delta^{-10}$.

Applying Lemma C.4.1 and Lemma C.4.2 as in the beginning of this proof without first restricting to $\mathcal{N}$, we obtain the following analogue of (C.25): for some integer $1 \leqslant d' \ll \delta^{-2}q^{1/6} \ll \delta^{-4}$, we have

$$\sum_{\chi \,(\mathrm{mod}\, q)} \left| \sum_{x\leqslant n<x+H'/d'} f(n)\chi(n)n^{2\pi iT}\right| \gg \delta^3 q^{1/3}\frac{H'}{d'}$$

for all x in an H'/d'-separated set $S' \subset [X/d', 2X/d']$ of size $|S'| \gg \delta^2\frac{H}{H'}|S|$. By the pigeonhole principle, there is a character $\chi \,(\mathrm{mod}\, q)$ such that

$$\frac{1}{H'/d'}\left| \sum_{x\leqslant n<x+H'/d'} f(n)\chi(n)n^{2\pi iT}\right| \gg \frac{\delta^3 q^{1/3}}{q} \gg \delta^{10} \tag{C.29}$$

for $\gg |S'|/q \gg \delta^{12}\frac{H}{H'}|S|$ elements $x \in S'$. By the triangle inequality and the H'/d'-separation of S', the estimate (C.29) holds for

$$\gg \delta^{10}\frac{H'}{d'} \cdot \delta^{12}\frac{H}{H'}|S| \gg \delta^{22}\frac{|S|}{X/H}\frac{X}{d'}$$

integers $x \in [X/d', 2X/d']$.

We can now conclude the proof using the complex-valued Matomäki-Radziwiłł theorem with power savings and Halász's theorem. By the Matomäki-Radziwiłł theorem [36, Theorem 1.7], there is an absolute constant $C_1 \geqslant 1$ and a real number $t_0 \in [-X/d', X/d']$ such that the following holds: for any $0 < \delta_1 < 1/1000$, we have

$$\frac{1}{X/d'}\left|\sum_{X/d'<n\leqslant 2X/d'} f(n)\chi(n)n^{2\pi i(T-t_0)}\right| \gg \delta^{10} - O\left(\delta_1 + \frac{\log\log(H'/d')}{\log(H'/d')} + \frac{1}{(\log(X/d'))^{3/1000}}\right)$$

provided that

$$C_1\left(\frac{1}{(H'/d')^{\delta_1/15}} + \frac{1}{(X/d')^{\delta_1^4/10^{16}}}\right) < \delta^{22}\frac{|S|}{X/H}. \tag{C.30}$$

Using the assumption (C.24), we see that the condition (C.30) is satisfied if δ_1 is chosen to be a sufficiently small multiple of δ^{10}. Hence, we obtain

$$\frac{1}{X/d'}\left|\sum_{X/d'<n\leqslant 2X/d'} f(n)\chi(n)n^{2\pi i(T-t_0)}\right| \gg \delta^{10}. \tag{C.31}$$

For $A, B > 0$, define

$$m(A,B) := \min_{|\tau|\leqslant B}\sum_{p\leqslant A}\frac{1-\mathrm{Re}(f(p)\chi(p)p^{2\pi i(T-t_0)}p^{-i\tau})}{p}.$$

By Halász's theorem [70, Corollary 4.12, p.494], for every $B \geqslant 2$ we have

$$\frac{1}{X/d'}\left|\sum_{X/d'<n\leqslant 2X/d'} f(n)\chi(n)n^{2\pi i(T-t_0)}\right| \ll \frac{1+m(X/d',B)}{e^{m(X/d',B)}} + \frac{1}{B}. \tag{C.32}$$

Setting $B := X$, we deduce from (C.31) and (C.32) that $m(X/d', X) \ll \log(\delta^{-1})$. Since

$$\sum_{X/d'\leqslant p\leqslant X}\frac{1}{p} \ll 1,$$

we get $m(X,X) \ll \log(\delta^{-1})$, which concludes the proof of Theorem C.5.1. □

References

[1] Manindra Agrawal, Neeraj Kayal, and Nitin Saxena. Primes is in P. *Annals of Mathematics*, 160:781–793, 2004.

[2] Nesmith C. Ankeny. The least quadratic non residue. *Annals of mathematics*, pages 65–72, 1952.

[3] Erwin H. Bareiss. Sylvester's identity and multistep integer-preserving Gaussian elimination. *Mathematics of Computation*, 22(103):565–578, 1968.

[4] Hyman Bass. The Ihara-Selberg zeta function of a tree lattice. *International Journal of Mathematics*, 3(6):717–797, 1992.

[5] Florent Benaych-Georges, Charles Bordenave, and Antti Knowles. Spectral radii of sparse random matrices. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 56(3):2141–2161, 2020.

[6] Daniel J. Bernstein, Johannes Buchmann, and Erik Dahmen. *Post-quantum cryptography*. Springer, 2009.

[7] Rajendra Bhatia. *Matrix analysis*, volume 169. Springer, 1997.

[8] Charles Bordenave. A new proof of Friedman's second eigenvalue theorem and its extension to random lifts. *Annales Scientifiques de l'École Normale Supérieure. Quatrième Série*, 53(6):1393–1439, 2020.

[9] David A. Burgess. The distribution of quadratic residues and non-residues. *Mathematika*, 4(2):106–112, 1957.

[10] Sarvadaman Chowla. *The Riemann hypothesis and Hilbert's tenth problem*, volume 4 of *Mathematics and Its Applications*. Gordon and Breach Science Publishers, New York-London-Paris, 1965.

[11] Charles-Jean de la Vallée Poussin. Recherches analytiques sur la théorie des nombres premiers. *Annales de la Société scientifique de Bruxelles*, 20:183–256, 1896.

[12] Anton Deitmar. Ihara zeta functions of infinite weighted graphs. *SIAM Journal on Discrete Mathematics*, 29(4):2100–2116, 2015.

[13] Whitfield Diffie and Martin E. Hellman. New directions in cryptography. *IEEE Transactions on Information Theory*, 22(6):644–654, 1976.

[14] Martin Ekerå and Joel Gärtner. Extending Regev's factoring algorithm to compute discrete logarithms. In *Post-quantum cryptography. Part II*, volume 14772 of *Lecture Notes in Computer Science*, pages 211–242. Springer, 2024.

[15] Paul Erdös. On the irrationality of certain series. *Indagationes Mathematicae*, 19:212–219, 1957. Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, Series A **60**.

[16] Zhou Fan and Andrea Montanari. How well do local algorithms solve semidefinite programs? In *STOC 2017: Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 604–614. ACM, New York, 2017.

[17] Kevin Ford. Vinogradov's integral and bounds for the Riemann zeta function. *Proceedings of the London Mathematical Society*, 85(3):565–633, 2002.

[18] John B. Friedlander and Henryk Iwaniec. *Opera de Cribro*, volume 57. American Mathematical Society, 2010.

[19] Joel Friedman. A proof of Alon's second eigenvalue conjecture and related problems. *Memoirs of the American Mathematical Society*, 195(910):1–100, 2008.

[20] Craig Gidney. Asymptotically efficient quantum Karatsuba multiplication. arXiv:1904.07356, 2019.

[21] Andrew Granville and Kannan Soundararajan. Large character sums. *Journal of the American Mathematical Society*, 14(2):365–397, 2001.

[22] Jacques Hadamard. Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques. *Bulletin de la Société Mathématique de France*, 24:199–220, 1896.

[23] Glyn Harman. *Prime-detecting sieves*, volume 33 of *London Mathematical Society Monographs Series*. Princeton University Press, Princeton, NJ, 2007.

[24] David Harvey and Joris van der Hoeven. Integer multiplication in time $O(n \log n)$. *Annals of Mathematics*, 193(2):563–617, 2021.

[25] Harald A. Helfgott. Expansion, divisibility and parity: an explanation. In *Combinatorial and Additive Number Theory, New York Number Theory Seminar*, pages 199–237. Springer, 2021.

[26] Harald A. Helfgott and Maksym Radziwiłł. Expansion, divisibility and parity. arXiv:2103.06853, 2021.

[27] Harald A. Helfgott and Adrián Ubis. Primos, paridad y análisis. arXiv:1812.08707, 2018.

[28] Yasutaka Ihara. On discrete subgroups of the two by two projective linear group over $\mathfrak{p}$-adic fields. *Journal of the Mathematical Society of Japan*, 18:219–235, 1966.

[29] Albert Edward Ingham. A note on Fourier transforms. *Journal of the London Mathematical Society*, 9(1):29–32, 1934.

[30] Henryk Iwaniec and Emmanuel Kowalski. *Analytic number theory*, volume 53. American Mathematical Society, 2004.

[31] Matti Jutila. On Linnik's constant. *Mathematica Scandinavica*, 41(1):45–62, 1977.

[32] Gregory D. Kahanamoku-Meyer and Norman Y. Yao. Fast quantum integer multiplication with zero ancillas. arXiv:2403.18006, 2024.

[33] Burton S. Kaliski Jr. Targeted Fibonacci exponentiation. arXiv:1711.02491, 2017.

[34] John Edensor Littlewood. Sur la distribution des nombres premiers. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, 158:1869–1872, 1914.

[35] Kaisa Matomäki and Maksym Radziwiłł. Multiplicative functions in short intervals. *Annals of Mathematics*, 183(3):1015–1056, 2016.

[36] Kaisa Matomäki and Maksym Radziwiłł. Multiplicative functions in short intervals II. arXiv:2007.04290, 2020.

[37] Kaisa Matomäki, Maksym Radziwiłł, Xuancheng Shao, Terence Tao, and Joni Teräväinen. Higher uniformity of arithmetic functions in short intervals II. Almost all intervals. *Inventiones mathematicae*, pages 1–125, 2026.

[38] Kaisa Matomäki, Maksym Radziwiłł, and Terence Tao. An averaged form of Chowla's conjecture. *Algebra Number Theory*, 9(9):2167–2196, 2015.

[39] Kaisa Matomäki, Maksym Radziwiłł, and Terence Tao. Fourier uniformity of bounded multiplicative functions in short intervals on average. *Inventiones Mathematicae*, 220(1):1–58, 2020.

[40] Kaisa Matomäki, Maksym Radziwiłł, Terence Tao, Joni Teräväinen, and Tamar Ziegler. Higher uniformity of bounded multiplicative functions in short intervals on average. *Annals of Mathematics*, 197(2):739–857, 2023.

[41] Kaisa Matomäki, Xuancheng Shao, Terence Tao, and Joni Teräväinen. Higher uniformity of arithmetic functions in short intervals I. All intervals. *Forum of Mathematics, Pi*, 11:e29, 2023.

[42] Hugh L. Montgomery and Robert C. Vaughan. *Multiplicative Number Theory I: Classical Theory*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2006.

[43] Nathan Ng. The distribution of the summatory function of the Möbius function. *Proceedings of the London Mathematical Society*, 89(2):361–389, 2004.

[44] Michael A. Nielsen and Isaac L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, 2010.

[45] Karl K. Norton. On the number of restricted prime factors of an integer. I. *Illinois Journal of Mathematics*, 20(4):681–705, 1976.

[46] Cédric Pilatte. Improved bounds for the Fourier uniformity conjecture. arXiv:2604.26564, 2026.

[47] Cédric Pilatte. Improved bounds for the two-point logarithmic Chowla conjecture. To appear in the *Journal of the American Mathematical Society* (arXiv:2310.19357v2), 2026.

[48] Cédric Pilatte. Unconditional correctness of recent quantum algorithms for factoring and computing discrete logarithms. *Forum of Mathematics. Pi*, 14:e5, 2026.

[49] Seyoon Ragavan and Vinod Vaikuntanathan. Space-efficient and noise-robust quantum factoring. In *Advances in Cryptology – CRYPTO 2024. Part VI*, volume 14925 of *Lecture Notes in Computer Science*, pages 107–140. Springer, 2024.

[50] Oded Regev. An efficient quantum factoring algorithm. *Journal of the ACM*, 72(1):Art. 10, 1–13, 2025.

[51] Hans-Egon Richert. Zur Abschätzung der Riemannschen Zetafunktion in der Nähe der Vertikalen $\sigma = 1$. *Mathematische Annalen*, 169:97–101, 1967.

[52] Bernhard Riemann. Ueber die Anzahl der Primzahlen unter einer gegebenen Grösse. *Monatsberichte der Berliner Akademie*, pages 671–680, 1859.

[53] Ronald L. Rivest, Adi Shamir, and Leonard Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2):120–126, 1978.

[54] Bruce E. Sagan, Yeong-Nan Yeh, and Günter M. Ziegler. Maximizing Möbius functions on subsets of Boolean algebras. *Discrete Mathematics*, 126(1-3):293–311, 1994.

[55] Peter Sarnak. Möbius randomness and dynamics. *Notices of the South African Mathematical Society*, 43(2):89–97, 2012.

[56] Bernd S. W. Schröder. *Ordered sets: an introduction*, volume 29. Springer, 2003.

[57] Peter W. Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pages 124–134. IEEE, 1994.

[58] Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, 1997.

[59] Harold M. Stark and Audrey A. Terras. Zeta functions of finite graphs and coverings. *Advances in Mathematics*, 121(1):124–165, 1996.

[60] Terence Tao. The Erdős discrepancy problem. *Discrete Analysis*, (1):1–27, 2016.

[61] Terence Tao. The logarithmically averaged Chowla and Elliott conjectures for two-point correlations. *Forum of Mathematics, Pi*, 4:e8, 2016.

[62] Terence Tao. Equivalence of the logarithmically averaged Chowla and Sarnak conjectures. In *Number theory – Diophantine problems, uniform distribution and applications*, pages 391–421. Springer, Cham, 2017.

[63] Terence Tao and Joni Teräväinen. Odd order cases of the logarithmically averaged Chowla conjecture. *Journal de Théorie des Nombres de Bordeaux*, 30(3):997–1015, 2018.

[64] Terence Tao and Joni Teräväinen. The structure of correlations of multiplicative functions at almost all scales, with applications to the Chowla and Elliott conjectures. *Algebra & Number Theory*, 13(9):2103–2150, 2019.

[65] Terence Tao and Joni Teräväinen. The structure of logarithmically averaged correlations of multiplicative functions, with applications to the Chowla and Elliott conjectures. *Duke Mathematical Journal*, 168(11):1977–2027, 2019.

[66] Terence Tao and Joni Teräväinen. Quantitative correlations and some problems on prime factors of consecutive integers. arXiv:2512.01739, 2025.

[67] Terence Tao and Van H. Vu. *Additive combinatorics*, volume 105. Cambridge University Press, 2006.

[68] Pafnouti Tchebichef. Mémoire sur les nombres premiers. *Journal de Mathématiques Pures et Appliquées*, 17:366–390, 1852.

[69] Pafnouti Tchebichef. Sur la fonction qui détermine la totalité des nombres premiers inférieurs à une limite donnée. *Journal de Mathématiques Pures et Appliquées*, 17:341–365, 1852.

[70] Gérald Tenenbaum. *Introduction to analytic and probabilistic number theory*, volume 163. American Mathematical Society, third edition, 2015.

[71] Miguel Nicolás Walsh. Local uniformity through larger scales. *Geometric and Functional Analysis*, 31(4):981–991, 2021.

[72] Miguel Nicolás Walsh. Phase relations and pyramids. *Journal of the European Mathematical Society*, page published online first, 2025.

[73] Miguel Nicolás Walsh. Stability under scaling in the local phases of multiplicative functions. *Inventiones Mathematicae*, 241(1):325–362, 2025.

[74] Yusuke Watanabe and Kenji Fukumizu. Graph zeta function in the Bethe free energy and loopy belief propagation. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 22, pages 2017–2025, 2009.