Linear, bilinear and polynomial structures in function fields and the primes

Pierre-Yves Bienvenu

School of Mathematics

July 23, 2018

A dissertation submitted to the University of Bristol
in accordance with the requirements of the degree
of Doctor of Philosophy in the Faculty of Science

Abstract

This thesis is a contribution to arithmetic combinatorics. We present the Green-Tao method and the Green-Tao-Ziegler theorem concerning asymptotics for linear configurations of primes. Then we show extensions of our own to this theorem: first to some family of quadratic configurations, and secondly to configurations with unbounded coefficients. As a result, we are able to provide an asymptotic for configurations of primes inside the set of shifted squarefree numbers.

We then leave integers and move to vector spaces over finite fields. In this context, we prove a bidirectional additive smoothing result for sets of pairs $P \subset \mathbb{F}_p^n \times \mathbb{F}_p^n$. This is a bilinear version of Bogolyubov's theorem. We then equip these vector spaces with a multiplicative structure, that is, we consider polynomial rings over finite fields. Using the Croot-Lev-Pach method, we show that sets of polynomials of degree less than n that contain no nontrivial solution to a given polynomial equation (of some specific type) is exponentially small.

Finally, we seek to apply the Green-Tao method on polynomial rings, with the intention of deriving asymptotics for configurations of irreducible polynomials. To this aim, we bound correlations of the Möbius function with linear and quadratic phases.

Acknowledgments

First and foremost, I would like to thank my supervisor Julia Wolf. She read with an unbelievable level of attention a complete draft of this thesis (besides several individual chapters) and came up with comments which dramatically improved the exposition. She did that already for my publications, for which I am infinitely grateful. She made my PhD years pleasant as well as rich in enlightening mathematical events. She introduced me to people and articles. The amount of instruction I've received from her since my master year is enormous. She provided me with helpful advice on everything mathematical and non-mathematical.

Providing me with a fantastic PhD brother, Luka Rimanić, was certainly not her least contribution to my joy. I am awfully thankful to him, for as he knows well, my PhD years would have been considerably less fun without him.

More generally, I am thankful to the Pure Math group of the University of Bristol. This is a tremendously active group, and there were always a lot of great talks by great speakers carefully selected by our faculty. Micha Rudnev and Trevor Wooley should be particularly thanked for advice, conversations, and inspecting annually my (slow) progress. Thanks to the francophile Tim Browning for listening with some interest to my problems and suggesting things to do. I would like to thank the admin team for their great job, too.

Massive thanks are owed to the Catholic chaplaincy of Bristol. This was a fantastic place to live, great equipment, great housemates. Without them and the spiritual growth I enjoyed there, who knows how much poorer this thesis would have been.

Poi c'è Francesco, questo grande amico, con chi ho passato belli mesi a Meridian Place. Ti ringrazio molto per la tua presenza. Hai cocinato molte pizze per me, mi hai permesso d'avere una vita soziale un po diciamo; e tu mi capici al meno!

Merci à vous, mes parents. Un certain nombre de ces pages ont été rédigées (et beaucoup triturées dans tous les sens) sous votre toit. Vous vous êtes toujours bien occupés de moi quand je rentrais à la maison. Merci à la famille en général.

And last but not least, thanks to you, my dearest fiancée Katie. In spite of the distance to Bristol, you never abandoned me and took interest in my research, even accepting to proofread a draft. Thanks for your tenacity and confident support.

Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

Date: July 23, 2018

Contents

Abstract **iii**

Acknowledgments **v**

Declaration **vii**

1 Introduction **2**

1.1 Probabilistic heuristics about prime tuples 3

1.2 A higher-dimensional Siegel-Walfisz theorem 6

1.3 Polynomial configurations in the primes 8

1.4 Bilinear structures in vector spaces over finite fields 9

1.5 Polynomial equations in function fields 11

1.6 Uniformity of the Möbius function on polynomial rings 13

2 The Green-Tao method **15**

2.1 Statement of the Green-Tao-Ziegler theorem 15

2.2 The W-trick 20

2.3 Transference principle, Gowers norms, pseudorandomness 22

2.4 A pseudorandom majorant 24

2.5 Uniformity of the von Mangoldt and Möbius functions 30

3 A higher-dimensional Siegel-Walfisz theorem **31**

3.1 First reductions 33

3.2 Reduction to the case of a box 34

3.3 The W-trick 36

3.4 Reduction to a Gowers norm estimate 37

3.4.1 A pseudorandom majorant 38

3.4.2 Generalised von Neumann theorem 39
3.4.3 A Gowers norm estimate 42
3.5 Application to linear equations in a subset of the primes 45

4 Asymptotics for some polynomial patterns in the primes 50
4.1 The main theorem 50
4.2 Special cases 52
4.3 Overview of the general strategy 57
4.4 Proof of the main theorem 58
4.4.1 Elimination of a negligible set 58
4.4.2 Implementation of the W-trick 60
4.4.3 Analysis of the main term 64
4.4.4 The sum over X_2 66
4.4.5 Reduction of the main theorem 67
4.5 Majorant and uniformity of quadratic representation functions 68

5 Bilinear structures in vector spaces over finite fields 74
5.1 Preliminaries 74
5.2 The bilinear Bogolyubov theorem 79
5.3 A quick application 81
5.4 Proof of the main theorem 82
5.5 Proof of the iterative step 86
5.6 Remarks on transverse sets 91

6 The Croot-Lev-Pach method and applications 94
6.1 The Croot-Lev-Pach method 94
6.2 Polynomial equations in function fields 99
6.3 Further applications 102
6.4 Limits of the method 103
6.5 The Ellenberg-Gijswijt bound for large fields 107

7 Uniformity of the Möbius function on polynomial rings 111
7.1 Overview of the proof 115
7.2 Preliminaries 116
7.2.1 Notation and basic facts 116
7.2.2 L-functions of arithmetically distributed relations 117

7.3 Character sum estimates 120
7.4 Exponential sum estimates 123
7.5 Quadratic phases and Vaughan's identity 127
7.6 Using the polylogarithmic bilinear Bogolyubov conjecture 132
7.7 The Hankel case 136

A Volume packing arguments and local divisor density **139**

B Analysis of the local factors β_p **143**

C Verification of the linear forms condition **151**

D Digression on the Type I sum and sums of spaces of multiples **169**

E Divisor bounds **176**

Chapter 1

Introduction

One central topic in arithmetic combinatorics is the study of combinatorial structures inside large subsets of the integers or other abelian groups. Remarkably, certain structures occur almost inevitably. Szemerédi's theorem [82] from 1975, which we now state, is archetypical of this theme. Let $[N] = \{1, \ldots, N\}$.

Theorem 1.1. *Any set $A \subset \mathbb{N}$ satisfying $\limsup_{N \to +\infty} \frac{|A \cap [N]|}{N} > 0$ contains arbitrarily long nontrivial arithmetic progressions.*

The first meaningful case, the case of arithmetic progressions of length three (3-APs), had already been known since Roth [75].

A set A that satisfies the hypothesis of Theorem 1.1 is called *dense*, while sets that do not are called *sparse*. Erdős conjectured that the set of primes, a sparse set, contains arbitrarily long arithmetic progressions as well. Green and Tao famously proved this conjecture.

Theorem 1.2. *The set of primes contains arbitrarily long arithmetic progressions. In fact, the number of arithmetic progressions of length k among the primes smaller than N is asymptotic to $c_k \frac{N^2}{\log^k N}$ for some constant $c_k > 0$.*

The first part of the statement is the main result of [44]. The second part follows from [45] combined with later inputs [47, 48] by Green, Tao and Ziegler.

To pass from Theorem 1.1 to Theorem 1.2, Green and Tao devised the *transference principle*. The basic philosophy of the transference principle is that a dense subset S of a large but sparse random-looking subset $X \subset [N]$ of the integers satisfies the same asymptotics for linear configurations as a dense subset $S' \subset [N]$ of the integers. A theorem

of this type was first proven by Kohayakawa, Łuczak and Rödl [57], who found that almost surely, any dense subset of a sparse random set contains a 3-term arithmetic progression. Green and Tao's idea was to show that, in a way, primes are "pseudorandom", that is, they have a lot in common with typical random sets. For a state-of-the-art survey on the transference principle, see [22].

It turns out to be natural not to focus on arithmetic progressions, but instead to consider general linear configurations of primes, that is, tuples of the form $(\psi_1(\mathbf{n}), \ldots, \psi_t(\mathbf{n}))$ for $\mathbf{n} \in \mathbb{Z}^d$ and a given system $\Psi = (\psi_1, \ldots, \psi_t)$ of linear forms with integer coefficients, as we do in the sequel.

In the next section, we discuss heuristics concerning configurations of primes (also known as prime tuples) and state a more general version of Theorem 1.2. We then provide an overview of the thesis chapter by chapter.

1.1 Probabilistic heuristics about prime tuples

The prime number theorem says that the number of primes smaller than x is asymptotic, as x tends to infinity, to $x/\log x$. We can express it more conveniently using the von Mangoldt function Λ, which is defined on the set $\mathbb{N}$ of natural numbers by setting $\Lambda(n) = \log p$ if n is a power of a prime p and $\Lambda(n) = 0$ otherwise. The prime number theorem then states that the average of the von Mangoldt function on $[1, x]$ tends to 1 as x tends to ∞.

The prime number theorem can be reinterpreted as saying that when picking an integer in an interval near x (say $[x, 2x)$) uniformly at random, the selected number will be a prime with probability roughly $1/\log x$. We write

$$\mathbb{P}_{n\approx x}(n \text{ is a prime}) \sim \frac{1}{\log x},$$

where $\mathbb{P}_{n\approx x}$ denotes the proportion of integers $n \in [x, 2x)$ that satisfy a given property, and $a(x) \sim b(x)$ means that a is asymptotic to b, that is, a/b tends to 1 as x tends to $+\infty$. When we condition on n being coprime to a fixed prime p, we increase the probability by a factor $p/\varphi(p)$, where φ is Euler's function (in particular $\varphi(p) = p - 1$), because

$$\mathbb{P}_{n\approx x}(n \text{ is a prime}|(n,p) = 1) = \frac{\mathbb{P}_{n\approx x}(n \text{ is a prime})}{\mathbb{P}_{n\approx x}((n,p) = 1)} \sim \frac{p}{\varphi(p)} \frac{1}{\log x}.$$

Moreover these biases are multiplicative: if one knows that an integer $n \approx x$ is coprime to

both primes $p \neq q$, one should update one's estimate of the probability that n is prime by a factor $pq/\varphi(pq) = (p/\varphi(p))(q/\varphi(q))$. On the other hand, if one knows that an integer $n \approx x$ is divisible by some fixed prime p, one should update its probability of being prime to 0, because if x is large enough ($x > p$), no prime in $[x, 2x)$ is divisible by p.

Now fix two positive integers a, b. What is $\mathbb{P}_{n\approx x}(an + b \text{ is prime})$? A first guess (prior, in Bayesian statistical terms) is $1/\log x$, since when $n \approx x$, we have $an+b \in [ax+b, 2ax+b)$ and on this interval, the probability of being a prime is approximately $1/\log(ax) \sim 1/\log x$. However, $an + b$ is not any number in $[ax + b, 2ax + b)$. For instance, if a prime p divides a but not b, we know that $an + b$ is coprime to p. So we should multiply our estimate by $p/\varphi(p)$ for each prime p that divides a and not b. If a prime p divides both a and b, then $an+b$ cannot be prime and we should multiply our estimate by 0. Being of the form $an+b$ does not seem to include any other useful information for our estimate. So we might guess that

$$\mathbb{P}_{n\approx x}(an + b \text{ is prime}) \sim \frac{1}{\log x} \prod_{p|a} \frac{p}{\varphi(p)} 1_{(p,b)=1} \tag{1.1}$$

or, in terms of the von Mangoldt function, $\mathbb{E}_{n\approx x}\Lambda(an + b) \sim \prod_{p|a} \frac{p}{\varphi(p)} 1_{(p,b)=1}$. Here we denoted by $\mathbb{E}$ the averaging operator, thus $\mathbb{E}_{n\approx x} = \frac{1}{x}\sum_{n\in[x,2x)} f(n)$. For simplicity, for any integer q, we introduce the *local von Mangoldt function* $\Lambda_{\mathbb{Z}/q\mathbb{Z}}$ defined on $\mathbb{Z}$ or $\mathbb{Z}/q\mathbb{Z}$ by

$$\Lambda_{\mathbb{Z}/q\mathbb{Z}}(b) = \frac{q}{\varphi(q)} 1_{(q,b)=1}. \tag{1.2}$$

Observe that $\Lambda_{\mathbb{Z}/q\mathbb{Z}}$ depends only on the *radical* of q, defined by $\mathrm{rad}(q) = \prod_{p|q} p$.

In fact, equation (1.1) is true, and it is the content of the prime number theorem in arithmetic progressions; in other words, if a and b are fixed and coprime, then the number of integers $n \approx x$ such that $an + b$ is prime is asymptotic to $\frac{a}{\varphi(a)} \frac{x}{\log x}$. In particular, this asymptotic (and so the probability (1.1)) does not depend on b, as long as $(a, b) = 1$.

Before making this precise, we pause to introduce some standard notation. We write $X = O(Y)$ to say that the quantity $|X|$ is bounded by a constant times $|Y|$. Equivalently, we may write $X \ll Y$ or $X \gg Y$ or $Y = \Omega(X)$. We may add subscripts to indicate on which parameters the implied constant depends. The notation $X = o(Y)$ means that $|X|/|Y|$ tends to 0 when the asymptotic parameter tends to infinity.

In fact, the asymptotic (1.1) still holds if a and b are not fixed but satisfy $a = O(\log^A x)$ and $b = O(x \log^A x)$ for some $A > 0$; in that regime, the Siegel-Walfisz theorem [55,

Equation (17.3)] precisely states that

$$\mathbb{E}_{n\approx x}\Lambda(an+b) = \prod_{p|a}\Lambda_{\mathbb{Z}/p\mathbb{Z}}(b) + O_A(\log^{-A} x). \tag{1.3}$$

More generally, fix a tuple (we shall call it a system) $\Psi = (\psi_1, \ldots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ of affine-linear forms with positive integer coefficients. These coefficients may even depend on the asymptotic parameter x. Note that in the present chapter and the next one, we use bold characters to denote tuples of integers, in order to distinguish them from mere integers. One can then ask for

$$\mathbb{P}_{\mathbf{n}\approx x}(\Psi(\mathbf{n}) \text{ is prime})$$

where $\mathbf{n} \approx x$ means that $\mathbf{n} = (n_1, \cdots, n_d)$ lies in $[x, 2x)^d$, and the vector $\Psi(\mathbf{n})$ is said to be prime if all its coordinates are. Given that each $\psi_i(\mathbf{n})$ has size $O(x)$, a first, naive guess assuming that the variables $\psi_i(n)$ are pairwise independent, is that

$$\mathbb{P}_{\mathbf{n}\approx x}(\Psi(\mathbf{n}) \text{ is prime}) \sim \frac{1}{(\log x)^t}.$$

The prime number theorem in arithmetic progressions might lead one to suspect that this crude guess only needs to be updated by the biases induced by primes to be true. To understand how much the prime p affects the prior, one needs to reduce the system Ψ modulo p and consider how often the tuple $(\psi_1(\mathbf{n}), \cdots, \psi_t(\mathbf{n})) \in (\mathbb{Z}/p\mathbb{Z})^t$ has no vanishing coordinate. This gives rise to the *local factor*

$$\beta_p = \beta_p(\Psi) = \mathbb{E}_{\mathbf{a}\in(\mathbb{Z}/p\mathbb{Z})^d}\prod_{i=1}^{t}\Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(\mathbf{a})), \tag{1.4}$$

which quantifies the bias induced by the prime p. Note that β_p is the ratio between the actual probability $\mathbb{E}_{\mathbf{a}\in(\mathbb{Z}/p\mathbb{Z})^d}\prod_{i=1}^{t} 1_{(\psi_i(\mathbf{a}),p)=1}$ that each of the values $\psi_i(\mathbf{a})$ is coprime to p when $\mathbf{a}$ ranges (uniformly) over $(\mathbb{Z}/p\mathbb{Z})^d$, and the probability $(\varphi(p)/p)^t$ that each of t independent uniform variables on $\mathbb{Z}/p\mathbb{Z}$ is coprime to p.

Supposing the biases are again multiplicative, one is thus led to the (generalised) Hardy-

Littlewood conjecture [45, Conjecture 1.2]

$$\mathbb{E}_{\mathbf{n}\approx x}\prod_{i=1}^{t}\Lambda(\psi_i(\mathbf{n}))\sim\prod_{p}\beta_p, \tag{1.5}$$

which is the multivariate analogue of the original Hardy-Littlewood conjecture [49]. Under a reasonable linear-algebraic assumption that discards for instance the twin prime configuration, this is what Green, Tao and Ziegler proved [45].

Theorem 1.3. *Let L be a constant and let $\Psi=(\psi_1,\dots,\psi_t):\mathbb{Z}^d\to\mathbb{Z}^t$ be a system of affine-linear forms such that no two are affinely related. Suppose that the linear coefficients of the forms have magnitude at most L and that the constant coefficients have magnitude at most LN. Let $K\subset[-N,N]^d$ be a convex body. Then*

$$\sum_{\mathbf{n}\in\mathbb{Z}^d\cap K}\prod_{i=1}^{t}\Lambda(\psi_i(\mathbf{n}))=\beta_\infty\prod_{p}\beta_p+o_{d,t,L}(N^d),$$

where

$$\beta_\infty=\mathrm{Vol}(K\cap\Psi^{-1}(\mathbb{R}_+^t))$$

and β_p is as in equation (1.4).

We will discuss this theorem in more depth in Chapter 2, where we will give a more detailed introduction to the Green-Tao method.

1.2 A higher-dimensional Siegel-Walfisz theorem

Theorem 1.3 presents three important (and related) limitations, which we shall overcome in Chapter 3. The first concerns the convex body K: for the theorem to be nontrivial, it needs to be included and dense in $[-N,N]^d$, that is, $\mathrm{Vol}(K)\gg N^d$. To get rid of this limitation, one would need to replace the error term $o(N^d)$ by $o(\mathrm{Vol}(K))$. The second one is that the constant coefficients of Ψ are required to be $O(N)$, and the third one is that the linear coefficients have to be bounded. We shall say that a system satisfying the latter condition is *bounded*; otherwise, it is called *unbounded*. In view of the Siegel-Walfisz theorem, that is, equation (2.2), it is natural to believe that the constant coefficients could be of size $O(N\log^{O(1)}N)$ and the linear ones of size $O(\log^{O(1)}N)$. Then it is natural to

consider convex bodies K of volume potentially $O(N^d \log^{-O(1)} N)$, in order for the forms ψ_i to remain $O(N)$ across K.

The second limitation was already overcome by Green and Tao together with Ford and Konyagin [29], as they showed that the constraint on the constant coefficients can indeed be relaxed to $O(N \log^{O(1)} N)$. This relaxed condition recently allowed Tao and Ziegler [89, Theorem 1.3] to obtain an improvement of the error term $o(N^d)$ to $o(\mathrm{Vol}(K))$ in the case where $K = [1, N] \times [1, M]^{d-1}$ with $M \gg N \log^{-O(1)} N$ and $\psi_i(\mathbf{n}) = n_1 + P_i(n_2, \ldots, n_{d-1})$ for some affine-linear forms $P_1, \ldots, P_t$ whose linear coefficients are bounded; this is a first step towards removing the first limitation.

We now present our result [9, Theorem 1.3], which addresses all the aforementioned limitations. A system of affine-linear forms is called *admissible* if no two forms are affinely related and none of the local factors β_p (introduced in equation (1.4)) vanish.

Theorem 1.4. *Let d, t be positive integers. Assume that $\Psi = (\psi_1, \ldots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ is an admissible system of affine-linear forms. Suppose that the constant coefficients are $O(N \log^{O(1)} N)$ while the linear coefficients are $O(\log^{O(1)} N)$. Finally let $K \subset [-N, N]^d$ be a convex body satisfying $\mathrm{Vol}(K) \gg N^d \log^{-O(1)} N$ and $\Psi(K) \subset \mathbb{R}_+^t$. Then*

$$\sum_{\mathbf{n} \in \mathbb{Z}^d \cap K} \prod_{i=1}^{t} \Lambda(\psi_i(\mathbf{n})) = \mathrm{Vol}(K) \prod_p \beta_p (1 + o(1)).$$

Still in Chapter 3, we will show how Theorem 1.4 can be applied to extend Theorem 1.3 in yet another direction, namely to cover linear configurations inside specific subsets of the primes. More precisely, we derive asymptotics for the count of linear configurations within the primes p such that $p - 1$ is squarefree. Let $F(n) = \Lambda(n + 1)\mu^2(n)$, where μ^2 is the indicator function of the squarefree integers.

Theorem 1.5. *Let $\Psi : \mathbb{Z}^d \to \mathbb{Z}^t$ be an admissible system of affine-linear forms and let $K \subset [-N, N]^d$ be a convex body. Suppose that the linear coefficients are $O(1)$, the constants ones are $O(N)$ and that $\Psi(K) \subset \mathbb{R}_+^t$. Then there exists a constant $C(\Psi)$ (possibly equal to 0) such that*

$$\sum_{\mathbf{n} \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} F(\psi_i(\mathbf{n})) = C(\Psi)\mathrm{Vol}(K) + o(N^d).$$

1.3 Polynomial configurations in the primes

In the previous section, we saw that the arithmetic progressions of primes in Theorem 1.2 can in fact be located inside the set of shifted squarefree numbers. It is interesting to ask whether one can force the common difference of the progression to lie in a special set of integers, too. A special case of a theorem of Tao and Ziegler [88] states that the primes contain arbitrarily long arithmetic progressions whose common difference is a square. This amounts to the investigation of prime tuples inside the set

$$\{(n, n+d^2, \ldots, n+(k-1)d^2) \mid (n,d) \in \mathbb{Z}^2\}. \tag{1.6}$$

They were able to obtain a lower bound of the expected order of magnitude for the number of such configurations, but not an asymptotic (at that time).

To get an asymptotic, it is natural to increase the number of variables. Adding a degree of freedom to the set (1.6), we consider

$$\{(a, a+b^2+c^2, \ldots, a+(k-1)(b^2+c^2)) \mid (a,b,c) \in \mathbb{Z}^2\}, \tag{1.7}$$

the set of arithmetic progressions whose common difference is a sum of two squares.

In Chapter 4, we derive an asymptotic for the number of prime tuples inside (1.7), counted with multiplicity; that is, each arithmetic progression is counted as many times as the common difference is represented as a sum of two squares. This amounts to an extension of Theorem 1.3 to the case of a polynomial system Ψ, which we now state.

Theorem 1.6. *Let $k \geq 1$ be an integer and*

$$L = \{(a,b,c) \in \mathbb{R}^3 \mid 1 \leq a \leq a+(k-1)(b^2+c^2) \leq N\}.$$

Let $\Psi = (\psi_0, \cdots, \psi_{k-1}) \in \mathbb{Z}[a,b,c]^k$ be the polynomial system defined by

$$\psi_i(a,b,c) = a + i(b^2+c^2).$$

Then

$$\sum_{\mathbf{n} \in \mathbb{Z}^3 \cap L} \prod_{i=0}^{k-1} \Lambda(\psi_i(\mathbf{n})) = \beta_\infty \prod_p \beta_p + o(N^2)$$

with $\beta_\infty = \mathrm{Vol}(L)$ *and*

$$\beta_p = \mathbb{E}_{\mathbf{n}\in(\mathbb{Z}/p\mathbb{Z})^3} \prod_{i=0}^{k-1} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(\mathbf{n})).$$

This result is a special case of a much more general one (Theorem 4.1) that we will state and prove in Chapter 4.

In the next two sections, we abandon the specific topic of patterns in the primes to return to the general combinatorial topic of patterns in dense sets illustrated by Theorem 1.1. However, we will replace the set of integers $\mathbb{N}$ by a popular and useful model, namely vector spaces over finite fields.

1.4 Bilinear structures in vector spaces over finite fields

Vector spaces over finite fields are a particularly interesting frame for arithmetic combinatorics. This model arose from the desire to replace natural but untractable problems in the integers by toy problems retaining most of the flavour of the original, but is now extensively studied in its own right. It basically consists in replacing the set $[N] = \{1, \ldots, N\}$ by the group $G = \mathbb{F}_p^n$ where one thinks of the prime p as a small constant and $N = p^n$ as tending to infinity. The appeal of this so-called *finite field model* resides in the abundant supply of substructures in the group $G = \mathbb{F}_p^n$. Indeed, G has many subgroups, that is, subspaces. Given the importance of the *density increment* method in arithmetic combinatorics, an iterative method that progressively zooms in on substructures, one can imagine how convenient this model is. For more details on the model, see [38, 93].

A prime example of an additive combinatorial result in vector spaces is Meshulam's theorem [70], that is, Roth's theorem in $\mathbb{F}_3^n$. While Roth's original proof used arithmetic progressions as substructures of $[N]$ for the density increment strategy, Meshulam showed the argument was cleaner in $\mathbb{F}_3^n$ with subspaces in place of arithmetic progressions, and obtained better bounds, namely $N/\log N$ instead of $N/\log\log N$ in the integers. Meshulam's paper inspired researchers to try to achieve a similar bound in the integers [16, 77, 14].

Another related problem which benefited from translation to the vector space setting was the so-called corner problem. Posed by Erdős [28] (and Graham), it was first solved by Ajtai and Szemerédi [1]. It consists in proving that if a set $A \subset [N] \times [N]$ contains no three points of the form $(x, y), (x + d, y), (x, y + d)$ with $d > 0$, then it must have size $o(N^2)$, and in making the decay rate explicit. The distinctive feature of this problem, which

occurs again in the result discussed in this section, is its bidirectional structure: there is a horizontal direction and a vertical direction. It was brilliantly exploited by Shkredov [80], who was the first to provide an explicit decay rate. His ideas are also fruitful in the vector space setting, as shown by Green [40, 38].

Additive smoothing, which will be the core topic of Chapter 5, is a further topic of additive combinatorics that translates well to the vector space setting. The idea is that however "rough" or unstructured a set $A \subset [N]$ may be, by considering the set of sums or differences of sufficiently many elements of A, one obtains a set that inevitably contains various large structures. A good example is Bogolyubov's theorem [15], originally formulated in the integers, which we now state in the finite field setting [38]. We fix a prime p. The *density* of a subset $A \subset \mathbb{F}_p^n = V$ is the quantity $\alpha = \frac{|A|}{|V|}$.

Theorem 1.7 (Bogolyubov)**.** *If $A \subset V$ is a set of density $\alpha > 0$, then the sumset*

$$A + A - A - A := \{a_1 + a_2 - a_3 - a_4 \mid (a_1, \dots, a_4) \in A^4\}$$

contains a vector subspace of codimension $c(\alpha) = O(\alpha^{-2})$.

The notation $A + A - A - A$ is often abbreviated as $2A - 2A$. The (simple) proof of Theorem 1.7 will be provided in Chapter 5. We state, without proof, a deep improvement of the constant $c(\alpha)$ appearing in Theorem 5.1, due to Sanders [78, Theorem 11.1].

Theorem 1.8. *We can take $c(\alpha) = O(\log^4 \alpha^{-1})$ in Theorem 5.1.*

One can interprete Bogolyubov's theorem as the statement that performing a bounded number of operations $A \mapsto A \pm A$ (indeed two here) is sufficient to essentially close a set with respect to these operations.

Observe that if A is already a subspace, then the set $2A - 2A = A$ is a subspace of codimension $\log_p \alpha^{-1}$, so Sanders' result is optimal up to the exponent.

We now consider a bilinear version of Theorem 1.7. Given a set $P \subset V \times V$ of pairs (x, y), a natural smoothing operation one can perform is a vertical or horizontal sum or difference. More precisely, let

$$P \overset{V}{\pm} P = \{(x, y_1 \pm y_2) \mid (x, y_1), (x, y_2) \in P\}$$

be the set of vertical sums or differences, respectively. Note that the letter V, which stands for vertical here, is also used to denote the ambient space, but this should not create any confusion.

Then similarly define $P \overset{H}{\pm} P$ the set of horizontal sums or difference. We denote by ϕ_V the operation

$$P \mapsto (P \overset{V}{+} P) \overset{V}{-} (P \overset{V}{+} P)$$

and define the operation ϕ_H similarly. In view of Bogolyubov's theorem, it is natural to imagine that performing these operations sufficiently many times, one obtains a large closed substructure. Now what sets are closed under both horizontal and vertical operations? Reasonable examples are Cartesian products of vector subspaces as well as zero sets of bilinear forms. We say a set $P \subset V \times V$ is a *bilinear set* of codimensions (r_1, r_2, r_3) if there exist subspaces $W_1 \leq V, W_2 \leq V$ of codimension r_1, r_2, respectively, and bilinear forms $Q_1, \ldots, Q_{r_3}$ on $W_1 \times W_2$ such that

$$P = \{(x, y) \in W_1 \times W_2 \mid Q_1(x, y) = \cdots = Q_{r_3}(x, y) = 0\}.$$

We show that sets of this form inevitably appear when iterating the operations ϕ_V and ϕ_H.

Theorem 1.9. *For any $\delta > 0$, there exists a constant $c(\delta) > 0$ such that the following holds. Let $P \subset V \times V$ have density δ. Let $P' = \phi_H \phi_V \phi_H(P)$. Then P' contains a bilinear set of codimensions (r_1, r_2, r_3) where $\max(r_1, r_2, r_3) \leq c(\delta)$. Moreover, $c(\delta) = O(\exp(\exp(\exp(\log^{O(1)} 1/\delta))))$.*

It is reasonable to imagine that, like in the linear case, the bound on the codimensions should be polylogarithmic, whence the following conjecture.

Conjecture 1.10. *In Theorem 1.9, one can take $c(\delta) = O(\log^{O(1)} \delta^{-1})$.*

We call this statement the *polylogarithmic bilinear Bogolyubov conjecture*. We give some evidence for it in Chapter 5. We will rely on it in Chapter 7 in order to prove a number-theoretic result in function fields.

The next section precisely introduces function fields from the arithmetic combinatorial viewpoint, and describes the content of Chapter 6.

1.5 Polynomial equations in function fields

As observed in the previous section, problems about linear equations or linear structures in sets of integers can be conveniently phrased in the context of vector spaces over finite fields.

In this chapter, we will consider more general polynomial equations in the vector space model. As an example, let us mention Sarkőzy's theorem [79]. Originally formulated in the integers, it states that if $A \subset \mathbb{Z}$ is a dense set of integers, it must contain two elements $a \neq b$ whose difference is a perfect square. We can rephrase this problem in the realm of vector spaces over finite fields, if we view a vector $f = (f_0, \ldots, f_{n-1}) \in \mathbb{F}_q^n$ as the vector of coefficients of the polynomial $\sum_{i=0}^{n-1} f_i t^i$ in the ring $\mathbb{F}_q[t]$. The operation of squaring a polynomial then corresponds to a polynomial map $\mathbb{F}_q^n \to \mathbb{F}_q^{2n-1}$. We will refer to $\mathbb{F}_q[t]$ as a function field, although it is really a polynomial ring.

Until recently, the methods devised to tackle arithmetic-combinatorial questions in vector spaces or function fields were mostly Fourier-analytic and inspired by the solutions given to the original problems in the integers. Consequently, the best upper bounds known for subsets of $\mathbb{F}_p^n$ free of certain given configurations often looked similar to their counterparts in $\mathbb{Z}$. This is no longer the case: the algebraic method of Croot-Lev-Pach [24], exploited by Ellenberg-Gijswijt [26], brought the maximal size of a progression-free set in $\mathbb{F}_3^n$ down to $N^c = 3^{cn}$ for some constant $c < 1$. In contrast, in the set of integers up to N, progression-free sets of size $N^{1-o(1)}$ are known [5]. This method proved relevant in function fields too, where it allowed Green to considerably lower the bound in Sarkőzy's theorem [42].

In Chapter 6, we present the method of Croot-Lev-Pach, alongside a new application [11], and a discussion of its limits. We now give a precise statement of the main theorem of Chapter 6. We fix a prime power q and write $G_{q,n}$ for the set of polynomials of degree strictly less than n over $\mathbb{F}_q$, so that $|G_{q,n}| = q^n$.

Theorem 1.11. *Let r, k and d be integers satisfying $k \geq 2r^2+1$. Suppose that $a_1, \ldots, a_k$ are polynomials over $\mathbb{F}_q$ of degree at most d satisfying $\sum_{i=1}^k a_i = 0$. Then there exist constants $0 < c(r,q) < 1$ and $C = C(d,r,q)$ such that any $A \subset G_{q,n}$ satisfying $|A| \geq kCq^{c(r,q)n}$ must contain a nontrivial solution to the equation*

$$\sum_{i=1}^{k} a_i f_i^r = 0,$$

that is, there exists a solution $(f_1, f_2, \ldots, f_d) \in A^d$ which is not of the form $(f, f, \ldots, f)$ for any $f \in A$.

In the next section, we combine our interest in configurations of primes (studied in the first three sections) and in finite fields analogues (studied in the last two sections) as we

discuss configurations of irreducible polynomials over $\mathbb{F}_q[t]$.

1.6 Uniformity of the Möbius function on polynomial rings

Just like the ring $\mathbb{Z}$, the ring $\mathbb{F}_q[t]$ is euclidean, and hence a unique factorisation domain. The study of irreducible polynomials is thus analogous to the study of prime numbers.

The interval $I_N = \{x \in \mathbb{N} \mid x < N\}$ is analogous to the space $G_{q,n}$ of polynomials over $\mathbb{F}_q$ of degree less than n. Thinking of $|n|$ as the cardinality of the quotient ring $\mathbb{Z}/n\mathbb{Z}$, one is lead to write $|f| = |\mathbb{F}_q[t]/(f)| = q^{\deg f}$. Then $G_{q,n} = \{f \in \mathbb{F}_q[t] \mid |f| < q^n\}$, which underlines the analogy to the interval I_N. Thus N corresponds to the cardinality $|G_{q,n}| = q^n$, and the degree n to the logarithm $\log N$. The two units ± 1 of $\mathbb{Z}$ correspond to the set $\mathbb{F}_q^*$ of units of $\mathbb{F}_q[t]$, and accordingly positive integers correspond to monic polynomials.

The prime number theorem has an analogue [74] that says that the number of monic irreducibles of degree n is $q^n/n + O(q^{n/2}/n)$, where the implied constant depends neither on q nor n. Now observe that two distinct limit regimes exist: one where the cardinality q tends to infinity while the degree n is fixed, and one where the degree n tends to infinity while q is fixed. In the first case, a lot is known regarding configurations of irreducibles. In fact, the analogue of the Hardy-Littlewood prime-tuple conjecture (1.5) is known [4], so there is little left to do.

In Chapter 7, we focus on the second regime, so we fix a prime power $q = p^s$ (for a prime p and an integer $s \geq 1$) and discuss asymptotics when n tends to infinity. Because of the aforementioned prime number theorem in polynomial rings and of the Ellenberg-Gijswijt theorem from Section 1.5, we see that any dense (or just not too sparse) subset of the set of irreducible polynomials of degree n contains a nontrivial three-term arithmetic progressions.

However, the best known bounds for subsets of $\mathbb{F}_p^n$ without any four-term arithmetic progressions do not allow us to draw a similar conclusion for this configuration. So specific properties of the set of irreducibles have to be exploited if one wants to prove that it contains long arithmetic progressions, or other linear configurations. In this vein, Lê [60] proved the following.

Theorem 1.12. *The set of monic irreducible polynomials contains affine subspaces of arbitrary dimensions.*

He did not provide asymptotics for the number of subspaces of dimension k inside the set of polynomials of degree at most n. As in Section 1.1, we can use probabilistic heuristics to conjecture these asymptotics. It turns out that the validity of such a conjecture is essentially equivalent to a non-correlation property of the Möbius function μ, the arithmetic function defined on $\mathbb{F}_q[t]$ by

$$\mu(f) = \begin{cases} (-1)^k & \text{where } k \text{ is the number of monic irreducible factors of } f\text{, if } f \text{ is squarefree,} \\ 0 & \text{otherwise.} \end{cases}$$

The required non-correlation property is a bound of the form

$$\sum_{\deg f < n} \mu(f)\chi(P(f)) = o(q^n) \tag{1.8}$$

for any polynomial P in the coefficients $f_0, \ldots, f_{n-1}$ of $f = \sum_{i=0}^{n-1} f_i t^i$, and any additive character χ of $\mathbb{F}_q$. Chapter 7 deals with such correlations for polynomials P of degree at most 2. Here is the result we prove there.

Theorem 1.13. *The bound* (1.8) *holds uniformly in P when P is a linear polynomial or a Hankel quadratic form, i.e. $P(f) = \sum_{i,j=0}^{n-1} a_{i+j} f_i f_j$ for some coefficients $a_0, \ldots, a_{2n-2}$.*

Under Conjecture 1.10, the bound (1.8) *holds uniformly in P when P is any quadratic polynomial.*

We have concluded our overview of the main results of this thesis. The next chapter contains the necessary background material for the results in Chapters 3 and 4.

Chapter 2

The Green-Tao method

This chapter contains no original material; it is merely an exposition of the method laid out by Green and Tao [45], which enabled Green, Tao and Ziegler to prove Theorem 1.3. Compared to the original exposition, we state the various ingredients in as much generality as possible in order to prepare the ground for the extensions we prove in the next two chapters.

This chapter and the next two ones rely on the Appendices A, B and C, where we have collected and proved a number of facts that are necessary but not really enlightening.

2.1 Statement of the Green-Tao-Ziegler theorem

Let $\Psi = (\psi_1, \dots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of affine-linear forms. The vital restriction that we impose is the *finite complexity* condition [45, Definition 1.5], which we now define. For an affine-linear form ψ, let $\dot{\psi}$ be its linear part.

Definition 2.1. For $A \subset [t]$, let V_A be the set of affine forms on $\mathbb{Z}^d$ whose linear part belongs to $\mathrm{span}(\dot{\psi}_i \mid i \in A)$. Let $i \in [t]$. The system Ψ is said to be of *complexity at most k at i* if there exists a partition of $[t] \setminus \{i\}$ into at most $k+1$ parts such that $\dot{\psi}_i \notin V_A$ for each part A of the partition. It is said to be of complexity at most k if it is of complexity at most k at any $i \in [t]$. The complexity is the minimum k such that the complexity is at most k, if there is any such $k \in \mathbb{N}$. Otherwise it is said to be infinite.

Thus Ψ is of finite complexity if and only if no two of the forms are affinely dependent, i.e. for any $i \neq j$, the linear parts $\dot{\psi}_i$ and $\dot{\psi}_j$ are not proportional. If a system of t forms is of finite complexity, it is of complexity at most $t-2$ as the trivial partition into $t-1$ singletons is admissible.

We give some examples. The system defining an arithmetic progression of length $t \geq 3$ is given by $\psi_i(a,b) = a + (i-1)b$ for $i \in [t]$. It has complexity $t-2$ at each $i \in [t]$; indeed, no two forms are proportional but any two forms span any third one. On the contrary, the twin prime system $\Psi(n) = (n, n+2)$ is not of finite complexity; in general, if $d = 1$ and $t > 1$, the system is of infinite complexity.

Let N be an asymptotic parameter. The system Ψ is implicitly allowed to depend on N.

Definition 2.2. The size of Ψ at scale N is the quantity

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

where $(e_1, \ldots, e_d)$ is the canonical basis of $\mathbb{R}^d$.

As an example, let us consider the ternary Goldbach system $\Psi(a,b) = (a, b, N-a-b)$. The size $\|\Psi\|_N = 1 + 4 = 5$ is bounded. In general, the chosen normalisation means that for the system to have bounded size, we need the constant coefficients to be $O(N)$ and the linear coefficients to be $O(1)$.

We are now ready to state the theorem of Green, Tao and Ziegler [45, Main Theorem]. For convenience we extend Λ to a function on $\mathbb{Z}$ by setting $\Lambda(-n) = 0$ for any $n \in \mathbb{N}$.

Theorem 2.1. *Let L be a constant and $\Psi = (\psi_1, \ldots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of affine-linear forms of finite complexity. Suppose that $\|\Psi\|_N \leq L$. Let $K \subset [-N, N]^d$ be a convex body. Then*

$$\sum_{\mathbf{n}\in\mathbb{Z}^d\cap K} \prod_{i=1}^{t} \Lambda(\psi_i(\mathbf{n})) = \beta_\infty \prod_p \beta_p + o_{d,t,L}(N^d),$$

where

$$\beta_\infty = \mathrm{Vol}(K \cap \Psi^{-1}(\mathbb{R}_+^t))$$

and β_p is as in equation (1.4).

For each prime p, we call β_p the *local factor* modulo p. It remains to prove that $\prod_p \beta_p$ is convergent, which we proceed to do below. Throughout the thesis, the letter p is reserved for primes so $\prod_p$ is a product over primes. A prime p is called *exceptional* for Ψ (and we write $p \in P_\Psi$) if there exist $i \neq j$ such that ψ_i and ψ_j are affinely related modulo p. In particular, if a form ψ_i is a constant modulo p, then p is exceptional. We highlight that our definition of an exceptional prime is different (less restrictive) than that of Green and

Tao [45, Theorem D.3]. Even so, only finitely many primes are exceptional for any given finite complexity system as we prove now.

Lemma 2.2. *Let $\Psi = (\psi_1, \dots, \psi_t)$ be a finite-complexity system of affine-linear forms. Then P_Ψ is finite.*

Proof. If ψ_i and ψ_j are affinely related modulo p for $i \neq j \in [t]$, then all the 2×2-minors of the matrix $(\dot{\psi}_k(e_\ell))_{k \in \{i,j\}, \ell \in [d]}$ are divisible by p. However, at least one of these minors has to be nonzero because the matrix has rank 2, as the forms $\dot{\psi}_i$ and $\dot{\psi}_j$ are not proportional over $\mathbb{Z}$. Consider one of these nonzero minors. It is divisible by finitely many primes, which concludes our proof. □

We now check that $\prod_p \beta_p$ is convergent.

Lemma 2.3. *Let $\Psi = (\psi_1, \dots, \psi_t)$ be a system of affine-linear forms. Then if p is not exceptional,*

$$\beta_p = 1 + O_{d,t}(p^{-2}).$$

In particular, if Ψ is of finite complexity and no β_p is zero, then the product $\prod_p \beta_p$ is convergent and nonzero.

Proof. If two forms ψ_i and ψ_j are not affinely related modulo p, then the probability as a ranges over $(\mathbb{Z}/p\mathbb{Z})^d$ that they vanish simultaneously at a is p^{-2}, by elementary linear algebra (see Proposition A.5). Inclusion-exclusion then yields $\beta_p = 1 + O(p^{-2})$ for unexceptional p. For the second part of the lemma, it suffices to note that if Ψ is of finite complexity, only finitely many primes are exceptional thanks to Lemma 2.2. □

We now show that we can introduce a few additional hypotheses in Theorem 2.1 at no extra cost. First, upon intersecting K by the half-spaces $\{\mathbf{n} \in \mathbb{R}^d \mid \psi_i(\mathbf{n}) > 0\}$, we can suppose that $\beta_\infty = \mathrm{Vol}(K)$. In fact, it will be convenient to suppose that $\psi_i(\mathbf{n}) > N^{9/10}$ for all $i \in [t]$ (9/10 is arbitrary here). We can see that the set

$$\{\mathbf{n} \in K \cap \mathbb{Z}^d : \exists i \in [t] \quad \psi_i(\mathbf{n}) \leq N^{9/10}\}$$

contains only $O(N^{d-1/10})$ elements, and Λ is bounded by $O(\log(NL))$. So intersecting K by the half-spaces $\{\mathbf{n} \in \mathbb{R}^d \mid \psi_i(\mathbf{n}) > N^{9/10}\}$ for $i \in [t]$, we can suppose $\Psi(K) \subset [N^{9/10}, +\infty)^t$. Besides, prime powers p^k with $k \geq 2$ are so sparse that we can replace Λ by $\Lambda' = 1_{\mathcal{P}} \log$ where $\mathcal{P}$ is the set of primes. We still call Λ' the von Mangoldt function.

We will find it convenient to impose on systems of linear forms the following combinatorial condition that is stronger than that of finite complexity.

Definition 2.3. The system Ψ is in *s-normal form at* $i \in [t]$ if there exists a set $J_i \subset [t] \setminus \{i\}$ of cardinality at most $s+1$ such that $\prod_{j \in J_i} \dot{\psi}_i(e_j) \neq 0$ whereas for all $k \in [t] \setminus \{i\}$, we have $\prod_{j \in J_i} \dot{\psi}_k(e_j) = 0$. The system Ψ is in *s-normal form* if it is in s-normal form at each $i \in [t]$.

Besides, we will say that a form $(x_1, \ldots, x_d) \mapsto \psi(x_1, \ldots, x_d)$ *uses* the variable x_i if $\dot{\psi}(e_i) \neq 0$. Thus, in a system in normal form, each form ψ_i uses its own set of variables that none of the other forms uses completely. We can then split the remaining forms into $|J_i|$ classes by associating to $j \in J_i$ the set of forms that do not use e_j, that is, the forms ψ_k such that $\dot{\psi}_k(e_j) = 0$. A form may be covered by several classes, but we can then arbitrarily select a single one and get a partition.

We will want to reparametrise any system of complexity s to obtain one that is in s-normal form. We start off with an example. Recall that the system defining an arithmetic progression of length $t \geq 3$, which is given by $\psi_i(a,b) = a + (i-1)b$ for $i \in [t]$, has complexity $t-2$ at each $i \in [t]$. To reparametrise it, let $b = x_1 + \cdots + x_t$ and $a = -\sum_{j \in [t]} (j-1)x_j$. Then $\psi_i(a,b) = \sum_{j \in [t]} (i-j)x_j = \phi_i(x_1, \ldots, x_t)$ uses all variables x_j except x_i, so setting $J_i = [t] \setminus \{i\}$, we find that the new system $\Phi : \mathbb{Z}^t \to \mathbb{Z}^t$ is in $(t-2)$-normal form. Besides $\Phi(\mathbb{Z}^t) = \Psi(\mathbb{Z}^2)$. We shall provide an algorithm that "normalises" the system. An important feature of it is to keep the size of the system under control. We first define formally what it is we want to construct.

Definition 2.4. Let $\Psi : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of affine-linear forms. An *extension* of Ψ is a system $\Psi' : \mathbb{Z}^{d'} \to \mathbb{Z}^t$ of affine-linear forms such that $\Psi(\mathbb{Z}^d) = \Psi'(\mathbb{Z}^{d'})$ and Ψ is the restriction of Ψ' to $\mathbb{Z}^d$ identified with $\mathbb{Z}^d \times \{0\}^{d'-d}$.

We now prove that any finite complexity system admits a normal form extension of the same complexity, using not too many variables nor too large coefficients. We essentially reproduce the statement and the proof of [45, Lemma 4.4], but make the quantitative dependence precise, which will be useful in Chapter 3.

Proposition 2.4. *Let $\Psi : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of affine-linear forms of complexity s. Then Ψ admits an s-normal form extension $\Psi' : \mathbb{Z}^{d'} \to \mathbb{Z}^t$ where $d' = O_{d,t}(1)$ and $\|\Psi'\|_N = O(\|\Psi'\|_N^{O(1)})$.*

Proof. We produce an extension which is in s-normal form at i for any given $i \in [t]$; the proposition follows by iterating the process t times, once for each $i \in [t]$.

The system Ψ being of complexity at most s at i, we have a partition $[t] \setminus \{i\} = \bigcup_{k\in[s+1]} A_k$ such that $\dot{\psi}_i \notin \mathrm{span}(\dot{\psi}_j \mid j \in A_k)$ for any $k \in [s+1]$. As a certificate to this linear independence property, there are vectors $f_1, \dots, f_{s+1}$ in $\mathbb{Z}^d$ satisfying $\dot{\psi}_i(f_k) \neq 0$ and $\dot{\psi}_j(f_k) = 0$ if $j \in A_k$. Besides, we can take the vectors f_i for $i \in [s+1]$ to be integer valued and of norm $\|f_i\| = O(\|\Psi\|_N^{O(1)})$ (for any norm chosen on $\mathbb{R}^d$): indeed, using Cramer's rule, these vectors' coordinates can be taken to be products of determinants of matrices of size at most d whose coefficients are at most $\|\Psi\|_N$. We then introduce $s+1$ extra variables that ψ_i will be the only one to use fully. That is, we define $d' = d + s + 1$ and $\Psi' : \mathbb{Z}^{d'} \to \mathbb{Z}^t$ by $\Psi'(\mathbf{n}, m_1, \dots, m_{s+1}) = \Psi(\mathbf{n} + m_1 f_1 + \cdots + m_{s+1} f_{s+1})$ for $\mathbf{n} \in \mathbb{Z}^d$ and $(m_1, \dots, m_{s+1}) \in \mathbb{Z}^{s+1}$. It is clear that Ψ' is an extension which is in s-normal form at i, and that the complexity has not been increased anywhere else in the process. Thus iterating the procedure proves the proposition. □

The next proposition shows that the count of Ψ-configurations can be reduced to the count of Ψ'-configurations.

Proposition 2.5. *Let $\Psi : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of complexity s and $\Psi' : \mathbb{Z}^{d'} \to \mathbb{Z}^t$ be its normal form extension constructed above. Let $K \subset [-N, N]^d$ be a convex body and $M = \lfloor N/\|\Psi'\|_N \rfloor$. Then there exists a convex body $K' \subset [-N', N']^d$ where $N' = O(N)$ such that $\mathrm{Vol}(K') = (2M)^{d'-d}\,\mathrm{Vol}(K)$ and for any t functions $g_1, \dots, g_t : \mathbb{Z} \to \mathbb{R}$, we have*

$$(2M+1)^{d'-d} \sum_{\mathbf{n}\in\mathbb{Z}^d\cap K} \prod_{i\in[t]} g_i(\psi_i(\mathbf{n})) = \sum_{\mathbf{n}\in\mathbb{Z}^{d'}\cap K'} \prod_{i\in[t]} g_i(\psi_i'(\mathbf{n})).$$

Proof. Let $f_{d+1}, \dots, f_{d'}$ be the vectors of size bounded by $O(\|\Psi\|_N^{O(1)})$ produced by Proposition 2.4, so that

$$\Psi'(\mathbf{n}, m_{d+1}, \dots, m_{d'}) = \Psi(\mathbf{n} + \sum_{i=d+1}^{d'} m_i f_i).$$

The convex body

$$K' = \{(\mathbf{n}, m_{d+1}, \dots, m_{d'}) \in \mathbb{R}^d \times [-M, M]^{d'-d} \mid \mathbf{n} + \sum_{i=d+1}^{d'} m_i f_i \in K\}$$

satisfies the requirements, thanks to the change of variable $\mathbf{r} = \mathbf{n} + \sum_{i=d+1}^{d'} m_i f_i \in K$ for

$(\mathbf{n}, \mathbf{m}) \in \mathbb{Z}^{d'} \cap K$. □

We can now reduce Theorem 2.1 to the following proposition.

Proposition 2.6. *Let L be a constant and $\Psi = (\psi_1, \ldots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of affine-linear forms in s-normal form. Suppose that $\|\Psi\|_N \leq L$. Let $K \subset [-N, N]^d$ be a convex body such that $\Psi(K) \subset [N^{9/10}, +\infty)^t$. Then*

$$\sum_{\mathbf{n} \in \mathbb{Z}^d \cap K} \prod_{i=1}^{t} \Lambda'(\psi_i(\mathbf{n})) = \mathrm{Vol}(K) \prod_p \beta_p + o_{d,t,L}(N^d), \tag{2.1}$$

where β_p is as in equation (1.4).

In the next chapter, we will extend Theorem 2.1 to the case where $\|\Psi\|_N = O(\log^{O(1)} N)$, which is why we made the dependence on $\|\Psi\|_N$ explicit.

2.2 The W-trick

We want to free the von Mangoldt function from the biases induced by small primes, in order to make it a function of average $1 + o(1)$ on any congruence class to small modulus. Such a property of stability of the average upon restriction to rather structured sets like a congruence class is called *uniformity* and it is highly desirable; we shall return to it later. It is clear that the bias $p/\varphi(p)$ is significant only when p is quite small (say smaller than some threshold w), so if we remove the bias coming from small primes, we should be left with a rather uniform function. One then fixes a growth function $w(N) = O(\log\log N)$ and a parameter $\widetilde{W}$ divisible by $W = \prod_{p \leq w(N)} p = O(\log^{O(1)} N)$ and still satisfying $\widetilde{W} = O(\log^{O(1)} N)$. So $\widetilde{W} = WQ$ where $Q = O(\log^{O(1)} N)$. The exact choice of Q depends on the desired application; it may be divisible by higher powers of small primes $p \leq w$ and by a few additional larger primes. In Chapter 4, we will want $w(N) = O(\log\log\log N)$, so for the sake of definiteness we decide once and for all that $w(N) = \log\log\log N$, although for the current discussion it is not necessary.

We introduce the *tricked* von Mangoldt function defined by

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

An important property of this function is that it has average $1 + o(1)$ by the Siegel-Walfisz theorem, whenever $(b, \widetilde{W}) = 1$. More generally, for any arithmetic progression P of length

$|P| \gg N \log^{-O(1)} N$ and modulus $q = O(\log^{O(1)} N)$ satisfying $\mathrm{rad}(q) \mid \widetilde{W}$ (where $\mathrm{rad}(q)$ is the radical of q, i.e. the product of its prime factors), if the initial term of P is bounded by $O(N \log^{O(1)} N)$, we have

$$\frac{1}{|P|} \sum_{n \in P} \Lambda'_{\widetilde{W},b}(n) = 1 + o(1). \tag{2.2}$$

Theorem 2.1 can then be reduced to the following statement involving this modified von Mangoldt function.

Proposition 2.7. *Let L be a constant and $\Psi = (\psi_1, \ldots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of affine-linear forms in s-normal form. Suppose that $\|\Psi\|_N \leq L$. Let $K \subset [-N, N]^d$ be a convex body such that $\Psi(K) \subset [N^{9/10}, +\infty)^t$. Let $b_1, \ldots, b_t$ be integers in $[\widetilde{W}]$ and coprime to W. Then*

$$\sum_{\mathbf{n} \in \mathbb{Z}^d \cap K} \prod_{i=1}^{t} \Lambda'_{\widetilde{W},b_i} \psi_i(\mathbf{n}) = |K \cap \mathbb{Z}^d| + o_{d,t,L}(N^d). \tag{2.3}$$

Proof that Proposition 2.7 implies Theorem 2.1. We decompose the left-hand side of (2.1) into sums over congruence classes. We write

$$\mathbb{Z}^d \cap K = \bigcup_{\mathbf{a} \in [\widetilde{W}]^d} \mathbb{Z}^d \cap (\widetilde{W} K_{\mathbf{a}} + \mathbf{a}),$$

where

$$K_{\mathbf{a}} = \{\mathbf{x} \in \mathbb{R}^d \mid \widetilde{W}\mathbf{x} + \mathbf{a} \in K\}$$

is again a convex body. Putting

$$F(\mathbf{n}) = \prod_{i=1}^{t} \Lambda'(\psi_i(\mathbf{n}))$$

we can write the left-hand side of (2.1) as

$$\sum_{\mathbf{n} \in \mathbb{Z}^d \cap K} F(\mathbf{n}) = \sum_{\mathbf{a} \in [\widetilde{W}]^d} \sum_{\mathbf{n} \in \mathbb{Z}^d \cap K_{\mathbf{a}}} F(\widetilde{W}\mathbf{n} + \mathbf{a}). \tag{2.4}$$

Moreover, for $j \in [t]$, we can write $\psi_j(\widetilde{W}\mathbf{n} + \mathbf{a}) = \widetilde{W}\tilde{\psi}_j(\mathbf{n}) + c_j(\mathbf{a})$ where $c_j(\mathbf{a}) \in [\widetilde{W}]$ and $\widetilde{\psi_j}$ is an affine-linear form differing from ψ_j only in the constant term. We note that if $\psi_i(\mathbf{a})$ is not coprime to p for some $i \in [t]$ and some prime $p \leq w(N)$, then for each

$n \in K_{\mathbf{a}} \cap \mathbb{Z}^d$ we have $F(\widetilde{W}\mathbf{n} + \mathbf{a}) = 0$ (if $(\psi_i(\mathbf{a}), p) > 1$, the integer $\psi_i(\mathbf{a}) \geq N^{9/10} > w$ is not a prime). Thus the residues $\mathbf{a}$ which bring a nonzero contribution to the right-hand side of (2.4) are all mapped by Ψ to tuples $(b_1, \ldots, b_t)$, each entry of which is coprime to $\widetilde{W}$. We denote by $A_\Psi \subset [\widetilde{W}]^d$ the set

$$\{\mathbf{a} \in [\widetilde{W}]^d \mid \forall i \in [t] \quad (\psi_i(\mathbf{a}), \widetilde{W}) = 1\}.$$

We can then rewrite equation (2.4) as

$$\sum_{\mathbf{n} \in \mathbb{Z}^d \cap K} F(\mathbf{n}) = \left(\frac{\widetilde{W}}{\varphi(\widetilde{W})}\right)^t \sum_{\mathbf{a} \in A_\Psi} \sum_{\mathbf{n} \in \mathbb{Z}^d \cap K_{\mathbf{a}}} \prod_{i=1}^t \Lambda'_{\widetilde{W}, c_i(\mathbf{a})}(\tilde{\psi}_i(\mathbf{n})). \tag{2.5}$$

Applying Proposition 2.7 to the inner sum in the right-hand side of (2.5) for $\mathbf{a} \in A_\Psi$, we conclude that

$$\sum_{\mathbf{n} \in \mathbb{Z}^d \cap K} F(\mathbf{n}) = \left(\frac{\widetilde{W}}{\varphi(\widetilde{W})}\right)^t |A_\Psi|(\mathrm{Vol}(K_a) + o(N/\widetilde{W})^d).$$

Moreover, we see that $(\widetilde{W}/\varphi(\widetilde{W}))^t |A_\Psi| = \widetilde{W}^d \prod_{p|\widetilde{W}} \beta_p$. Lemma 2.3 implies that $\prod_{p \notin P_\Psi} \beta_p = 1 + O(1)$ and the boundedness of the coefficients of Ψ implies that P_Ψ is bounded. In particular, $\prod_{p \in P_\Psi} \beta_p$ is bounded, from which it follows that $\prod_{p|\widetilde{W}} \beta_p = \prod_p \beta_p(1 + o(1)) = \prod_p \beta_p + o(1)$. Combining this with the fact that $\widetilde{W}^d \mathrm{Vol}(K_a) = \mathrm{Vol}(K)$, we obtain the conclusion. □

2.3 The transference principle, Gowers norms and pseudorandom majorants

The transference principle or dense model theorem, first stated in [44, 88], says that if an unbounded function f is dominated by a *pseudorandom measure* ν, then when it comes to evaluating multilinear averages, f can be approximated by a 1-bounded function g; that is

$$\sum_{\mathbf{n} \in \mathbb{Z}^d \cap K} \prod_{i=1}^t f(\psi_i(\mathbf{n})) \approx \sum_{\mathbf{n} \in \mathbb{Z}^d \cap K} \prod_{i=1}^t g(\psi_i(\mathbf{n}))$$

for some function $g : \mathbb{Z} \to \mathbb{C}$ satisfying $|g| \leq 1$. For instance, equation (2.3) asserts that the functions $\Lambda'_{\widetilde{W}, b}$ can be approximated by the constant function 1.

In order to make this statement more precise, we need to define *pseudorandom measures* and the *Gowers norms*.

Definition 2.5. Let $g : \mathbb{Z} \to \mathbb{C}$ be a function and $k \geq 1$ an integer. The *Gowers norm* or U^k norm of g on $[N]$ is the expression

$$\|g\|_{U^k[N]} = \left(\mathbb{E}_{x\in[N]} \mathbb{E}_{\mathbf{h}\in[N]^k} \prod_{\boldsymbol{\omega}\in\{0,1\}^k} \mathcal{C}^{|\boldsymbol{\omega}|} g(x + \boldsymbol{\omega} \cdot \mathbf{h}) \right)^{2^{-k}},$$

where $\mathcal{C}$ is the conjugation operator and $|\boldsymbol{\omega}| = \sum_{i\in[k]} \omega_i$.

These norms originated in Gowers' new proof of Szemerédi's theorem [34].

Definition 2.6. Let $D \geq 1$. A *D-pseudorandom measure* is a sequence of functions $\nu = \nu_M : \mathbb{Z}/M\mathbb{Z} \to \mathbb{R}_+$, satisfying[1] $\mathbb{E}_{n\leq M}\nu(n) = 1+o(1)$. and the *D-linear forms conditions* defined as follows. Let $1 \leq d, t \leq D$. For every finite-complexity system of affine-linear forms $\Psi : \mathbb{Z}^d \to \mathbb{Z}^t$ with coefficients bounded by D, and any convex set $K \subset [-M, M]^d$ such that $\Psi(K) \subset [1, M]^t$, the following estimate holds

$$\mathbb{E}_{\mathbf{n}\in(\mathbb{Z}/M\mathbb{Z})^d} \prod_{i\in[t]} \nu(\psi_i(\mathbf{n})) = 1 + o(1). \tag{2.6}$$

A *D-pseudorandom majorant* of a function $f : [M] \to \mathbb{C}$ is a D-pseudorandom measure $\nu : [M] \to \mathbb{R}_+$ such that $|f| \leq c\nu$ for some constant $c > 0$ (independent of M).

Pseudorandom measures are defined on cyclic groups rather than intervals of integers, so the values of the linear forms $\psi_i(\mathbf{n})$ are understood modulo M. Similarly, some authors prefer defining the Gowers norms on cyclic groups, and then on intervals of integers by embedding them in cyclic groups, but these definitions are equivalent [45, Appendix B].

The reason why these notions are precious is the following so called *generalised von Neumann theorem* [45, Proposition 7.1], which essentially says that the U^{s+1} norm controls averages of functions bounded by a common pseudorandom measure along linear configurations of complexity at most s.

Theorem 2.8. *Let t, d, L, s be positive integer parameters. Then there are positive constants $1 \leq \Gamma$ and D, depending on t, d and L such that the following holds. Let C be a*

[1] In earlier works such as [45] or [68], there was a *correlation condition*, but it is no longer necessary due to the work of Fox, Conlon and Zhao [22], and its integration by Tao and Ziegler [87].

constant satisfying $\Gamma \leq C \leq O_{t,d,L}(1)$ *and suppose that* $M \in [CN', 2CN']$ *is a prime. Let* $\nu : \mathbb{Z}/M\mathbb{Z} \to \mathbb{R}^+$ *be a* D*-pseudorandom measure, and suppose that* $f_1, \ldots, f_t : [N'] \to \mathbb{R}$ *are functions with* $|f_i(x)| \leq \nu(x)$ *for all* $i \in [t]$ *and* $x \in [N']$*. Suppose that* $\Psi = (\psi_1, \ldots, \psi_t)$ *is a system of affine-linear forms of complexity at most* s *whose linear coefficients are bounded by* L*. Let* $K \subset [-N', N']^d$ *be a convex set such that* $\Psi(K) \subset [0, N']^t$*. Finally, suppose that*

$$\min_{1 \leq j \leq t} \|f_j\|_{U^{s+1}[N']} = o(1). \tag{2.7}$$

Then we have

$$\mathbb{E}_{\mathbf{n} \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} f_i(\psi_i(\mathbf{n})) = o(1).$$

Theorem 2.8 is proved by repeated applications of the Cauchy-Schwarz inequality. It is in this proof that the normal form parametrisation introduced above is necessary. We do not provide the proof of Theorem 2.8, but we give one later for a variant thereof, Theorem 3.6.

We highlight that this theorem actually replaces a linear system Ψ with another one, the system $(x, \mathbf{h}) \mapsto (x + \boldsymbol{\omega} \cdot \mathbf{h})_{\boldsymbol{\omega} \in \{0,1\}^{t-1}}$, so that it is not immediately obvious that we have reduced the difficulty of the problem. However, it so happens that functions that have a large average along this system can be characterised in another way: this is the *inverse theorem for the Gowers norms* [48] due to Green, Tao and Ziegler, to which we shall return in Section 2.5.

The transference principle may then be reformulated as the statement that for an unbounded function $f : \mathbb{Z} \to \mathbb{R}$ that satisfies $|f| \ll \nu$ for some pseudorandom measure, there exists a 1-bounded function $g : \mathbb{Z} \to \mathbb{C}$ such that $\|f - g\|_{U^k}$ is small. We will want to apply it to $f = \Lambda'_{\widetilde{W},b}$. Besides, we need $g = 1$ to get asymptotics; with any other g, we get at best lower or upper bounds. Finally, we need f to be bounded by a pseudorandom measure. In the next two sections, which cover the number-theoretic content of Theorem 2.1, we address the needs for a pseudorandom majorant and a Gowers norm estimate.

2.4 A pseudorandom majorant

We recall that

$$\Lambda(n) = \sum_{d|n} \mu(d) \log \frac{n}{d} \tag{2.8}$$

for any $n \geq 1$. Goldston and Yıldırım, in a preprint about small gaps between primes, that was improved to a landmark paper with Pintz [32], introduced the truncated version

$$\Lambda_R(n) = \sum_{\substack{d|n \\ d \leq R}} \mu(d) \log \frac{R}{d} \tag{2.9}$$

where $R = N^\gamma$ and $\gamma > 0$ is a sufficiently small constant. Following Green and Tao [45, Appendix D], we will rather work with the following variant

$$\Lambda_{\chi,\gamma}(n) = \log R \left(\sum_{\ell|n} \mu(\ell) \chi\left(\frac{\log \ell}{\log R}\right) \right)^2, \tag{2.10}$$

where χ is a smooth even function $\mathbb{R} \to [0,1]$ supported on $[-1,1]$ satisfying $\chi(0) = 1$. Due to the truncated nature of this sum, it is fairly easy to see that this arithmetic function has a constant average, depending only on χ, as we show in the next proposition.

Proposition 2.9. *We have the asymptotic*

$$\frac{1}{N} \sum_{n \leq N} \Lambda_{\chi,\gamma}(n) = c(\chi) + o(1) \tag{2.11}$$

for some constant $c(\chi) > 0$. More generally, for any q such that $\mathrm{rad}(q) = O(\log^{O(1)} N)$ *and b coprime to q, we have*

$$\frac{1}{N} \sum_{n \leq N} \frac{\varphi(q)}{q} \Lambda_{\chi,\gamma}(qn + b) = c(\chi) + o(1). \tag{2.12}$$

Proof. We expand the squared sum defining $\Lambda_{\chi,\gamma}$ and exchange the order of summation to obtain

$$\sum_{n \leq N} \Lambda_{\chi,\gamma}(qn + b) = \log R \sum_{\ell,\ell' \leq R} \mu(\ell)\mu(\ell') \chi\left(\frac{\log \ell}{\log R}\right) \chi\left(\frac{\log \ell'}{\log R}\right) \sum_{n \leq N} 1_{[\ell,\ell'] | qn+b}. \tag{2.13}$$

Whenever $[\ell, \ell']$ is coprime to q, the inner sum equals $\frac{N}{[\ell,\ell']} + O(R^2)$ (uniformly in q);

otherwise it is 0. Let S be the left-hand side of (2.13) divided by $N \log R$. So we have

$$S = \frac{1}{N \log R} \sum_{n \leq N} \Lambda_{\chi,\gamma}(qn+b) = \sum_{([\ell,\ell'],q)=1} \frac{\mu(\ell)\mu(\ell')}{[\ell,\ell']} \chi\left(\frac{\log \ell}{\log R}\right) \chi\left(\frac{\log \ell'}{\log R}\right) + O(N^{4\gamma-1}), \tag{2.14}$$

where the last term is $o(1)$ as soon as $\gamma < 1/4$. We now use the Fourier transform. Letting θ be the Fourier transform of the smooth compactly supported function $x \mapsto e^x \chi(x)$, it is well known that

$$\forall A > 0 \qquad \theta(\xi) \ll_A (1+|\xi|)^{-A}. \tag{2.15}$$

This allows us to reconstruct χ from θ as an integral over the compact interval[2]

$$I = \{\xi \in \mathbb{R} \mid |\xi| \leq \log^{1/2} R\}$$

at the cost of a tolerable error; more precisely, for any $A > 0$, we have

$$\chi\left(\frac{\log x}{\log R}\right) = \int_{\mathbb{R}} x^{-\frac{1+i\xi}{\log R}} \theta(\xi) d\xi = \int_I x^{-\frac{1+i\xi}{\log R}} \theta(\xi) d\xi + O(x^{-\frac{1}{\log R}} \log^{-A} R). \tag{2.16}$$

Plugging this into our equation, and neglecting the error term for the moment (a justification for that will be given in Appendix C), we find

$$S = \int_{I^2} \theta(\xi)\theta(\xi') \left(\sum_{([\ell,\ell'],q)=1} \frac{\mu(\ell)\mu(\ell')}{[\ell,\ell']} \ell^{-z} \ell'^{-z'} \right) d\xi d\xi', \tag{2.17}$$

where we have defined $z = \frac{1+i\xi}{\log R}$ and z' analagously. Thus $\Re(z) > 0$ and $|z| = O(1/\log R)$. For a fixed (ξ, ξ'), the inner sum in equation (2.17) can be rewritten, by multiplicativity, as the convergent product

$$P_q(z,z') = \prod_{p \nmid q} (1 - p^{-1-z} - p^{-1-z'} + p^{-1-z-z'}).$$

Write $E_p = 1 - p^{-1-z} - p^{-1-z'} + p^{-1-z-z'}$, omitting the dependence in z, z'. We can see that $E_p = E'_p + O(1/p^2)$ when $p \to +\infty$ (uniformly in z, z'), where $E'_p = \frac{(1-p^{-1-z})(1-p^{-1-z'})}{1-p^{-1-z-z'}}$. We want to replace P_q by $q/\varphi(q) \prod_p E'_p$. This is allowed, up to a small error, by the following lemma.

[2] We prefer integrating over a compact set, in order to be able to easily swap summation and integration using Fubini's theorem.

Lemma 2.10. *As $R \to \infty$, we have*

$$\frac{\varphi(q)}{q} P_q = (1 + O(\log^{-1/10} R)) \prod_p E'_p.$$

Proof. We split the set of primes into large primes $p > \log^{1/10} R$ and the remaining ones. We have

$$\prod_{\substack{p|q \\ p>\log^{1/10} R}} E_p = \prod_{\substack{p|q \\ p>\log^{1/10} R}} (1 + O(p^{-2}))E'_p = (1 + O(\log^{-1/10} R)) \prod_{\substack{p|q \\ p>\log^{1/10} R}} E'_p. \tag{2.18}$$

Now q has only $O(1)$ prime factors $p > \log^{1/10} R \gg \log^{1/10} N$, because $\mathrm{rad}(q) = O(\log^{O(1)} N)$. Given that $E'_p = 1 + O(p^{-1})$, we can write

$$\prod_{\substack{p|q \\ p>\log^{1/10} R}} E_p = (1 + O(\log^{-1/10} R)) \prod_p E'_p.$$

We now take care of primes $p \leq \log^{1/10} R$. In that range, we can write

$$E_p = (1 - p^{-1})(1 + O(\log p / \log^{1/2} R))$$

uniformly in p by a simple Taylor expansion in the vicinity of $z = 0$ (or equivalently, in the regime where $R \to +\infty$), and similarly

$$E'_p = (1 - p^{-1})(1 + O(\log p / \log^{1/2} R)). \tag{2.19}$$

Because $\prod_{p \leq \log^{1/10} R}(1 + O(\log p / \log^{1/2} R)) = 1 + O(\log^{-1/3} R)$, we infer

$$P_q = (1 + O(\log^{-1/10} R)) \prod_{p \nmid q} E'_p.$$

It remains to discuss the role of primes $p \mid q$. As already observed, we can suppose $p \leq \log^{1/10} R$. Then because of the bound (2.19), we have $\prod_{p|q} E'_p = \frac{\varphi(q)}{q}(1+O(\log^{-1/10} R))$. This concludes the proof of Lemma 2.10. $\square$

Now one recognises that

$$\prod_p E'_p = \frac{\zeta(1+z+z')}{\zeta(1+z)\zeta(1+z')}$$

where ζ is the Riemann zeta function, which satisfies the asymptotic $\zeta(1+z) = 1/z+o(1/z)$ as $z \to 0$. It follows that

$$\prod_p E'_p = \frac{zz'}{z+z'}(1+o(1)) = \frac{1}{\log R}\frac{(1+i\xi)(1+i\xi')}{2+i(\xi+\xi')}(1+o(1))$$

when R tends to ∞. As a result,

$$\frac{1}{N}\sum_{n\leq N}\frac{\varphi(q)}{q}\Lambda_{\chi,\gamma}(qn+b) = (1+o(1))\int_{I^2}\theta(\xi)\theta(\xi')\frac{(1+i\xi)(1+i\xi')}{2+i(\xi+\xi')}d\xi d\xi'. \tag{2.20}$$

We can now undo the truncation of the integral to I^2, causing only a multiplicative $(1+o(1))$ factor. We thus obtain the desired result with the constant in (2.11) equal to

$$c = \int_{\mathbb{R}^2}\theta(\xi)\theta(\xi')\frac{(1+i\xi)(1+i\xi')}{2+i(\xi+\xi')}d\xi d\xi'.$$

We check that it is a positive constant. Indeed, because of the identity

$$\frac{1}{2+i(\xi+\xi')} = \int_0^\infty \exp(-(2+i(\xi+\xi'))x)dx,$$

we find that

$$c = \int_0^\infty \left(\int \theta(\xi)(1+i\xi)\exp(-(1+i\xi)x)d\xi\right)^2 dx$$

where the inner integral equals $-\chi'(x)$. Hence $c = \int_{\mathbb{R}_+}|\chi'|^2$ is the desired positive constant, depending only on χ. □

Accordingly, for any integer q and any $b \in [q]$ coprime to q, we define the *Green-Tao majorant*

$$\nu_{\mathrm{GT},q,b} : n \mapsto \frac{\varphi(q)}{qc(\chi)}\Lambda_{\chi,\gamma}(qn+b).$$

The following lemma shows that this function indeed majorises the tricked von Mangoldt function. We write $N' = N/\widetilde{W}$.

Proposition 2.11. *For any $b \in [\widetilde{W}]$ coprime to W, we have*

$$\Lambda'_{\widetilde{W},b}(n) \ll \nu_{\mathrm{GT},\widetilde{W},b}(n)$$

for $n \in [R, N']$, where the implied constant depends only on γ. Moreover, for any q such that $\mathrm{rad}(q) = O(\log^{O(1)} N)$ and any b coprime to q, we have

$$\frac{1}{N}\sum_{n \le N} \nu_{\mathrm{GT},q,b}(n) = 1 + o(1). \tag{2.21}$$

Proof. To prove the upper-bound property, we need concern ourselves only with the integers $n \in [R, N']$ such that $\widetilde{W}n + b$ is prime. In this case, the left-hand side is bounded above by a constant multiple of $\frac{\varphi(\widetilde{W})}{\widetilde{W}} \log N$, while the right-hand side is $\frac{\varphi(\widetilde{W})}{\widetilde{W}} \log R$, with $\log R = \gamma \log N \gg \log N$. The second part of the statement of Proposition 2.11 is a reformulation of Propostion 2.9. □

The next proposition, originally [45, Proposition 6.4], states that $\nu_{\mathrm{GT},\widetilde{W},b}$ is a pseudo-random majorant for $\Lambda'_{\widetilde{W},b}$.

Proposition 2.12. *Fix a constant $D > 0$, and a positive integer t. Then there exists a constant $C_0(D)$ such that the following holds. For any bounded $C \ge C_0(D)$ there exists $\gamma = \gamma(C, D)$ such that if $M \in [CN', 2CN']$ is a prime, if $b_1, \dots, b_t$ are in $[\widetilde{W}]$ and coprime to $[\widetilde{W}]$, the function ν^* defined on $\mathbb{Z}/M\mathbb{Z} \supset [N']$ by*

$$\nu^*(n) = \begin{cases} \frac{1+\nu_{\mathrm{GT},\widetilde{W},b_1}(n)+\dots+\nu_{\mathrm{GT},\widetilde{W},b_t}(n)}{1+t} & \text{if } n \in [N'] \\ 1 & \text{otherwise} \end{cases}$$

is a D-pseudorandom measure that majorises each of the functions $1, \Lambda'_{\widetilde{W},b_1}, \dots, \Lambda'_{\widetilde{W},b_t}$ on $[N'^{3/5}, N']$.

The majorisation was already proven in Proposition 2.11 (we suppose $\gamma < 3/5$). The rest of the proposition follows from a slightly more general one.

Proposition 2.13. *Let $b_1, \dots, b_t$ be in $[\widetilde{W}]$ and coprime to $\widetilde{W}$. Let $\Psi = (\psi_1, \dots, \psi_t)$ be a system of linear forms whose exceptional primes all divide $\widetilde{W}$. Let $K \subset [-N, N]^d$ be such that $\mathrm{Vol}(K) \ge N^{d-o(1)}$. Then*

$$\sum_{\mathbf{n} \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} \Lambda_{\chi,\gamma}(\widetilde{W}\psi_i(\mathbf{n}) + b_i) = \mathrm{Vol}(K) \left(c(\chi) \frac{\widetilde{W}}{\varphi(\widetilde{W})} \right)^t (1 + o(1)).$$

Observe that this statement does not require any bound on the coefficients, but if the coefficients are bounded, the exceptional primes are automatically bounded so they divide $\widetilde{W}$. Proposition 2.13 can essentially be read out from [45, Theorem D.3], itself largely inspired from [32]. However, we will prove it again in the thesis as as special case of Proposition C.1.

The deduction of Proposition 2.12 from Proposition 2.13 is standard, though not entirely obvious, because of the piecewise definition of ν^*; see [44, Proposition 9.8] or [22, Proposition 8.4].

2.5 Uniformity of the von Mangoldt and Möbius functions

As mentioned at the end of Section 2.3, we need an estimate for the Gowers norms, which is given by the following theorem.

Theorem 2.14. *Let s be an integer. Let $w = O(\log\log N)$ tend to infinity with N and $\widetilde{W}$ be an integer multiple of $W = \prod_{p\leq w} p$ such that $\widetilde{W} = O(\log^{O(1)} N)$. Then*

$$\|\Lambda'_{\widetilde{W},a} - 1\|_{U^s[N]} = o(1) \tag{2.22}$$

for any $a \in [\widetilde{W}]$ coprime to $\widetilde{W}$.

This is essentially [45, Proposition 7.2], with $\widetilde{W}$ instead of W. The proof of that bound fills alone two intricate papers, each proving a difficult theorem.

The first one [48], called the *inverse theorem for the Gowers norms*, asserts that the only obstruction to uniformity is the existence of a significant correlation with some sequence from a family of structured sequences called $(s-1)$-*step nilsequences*. We shall not define these objects formally. The reader may wish to think of $n \mapsto e(\alpha_s n^s + \alpha_{s-1} n^{s-1} + \cdots + \alpha_0)$ as an example of s-step nilsequence. More generally, nilsequences are abstract generalisations of such sequences, which in turn are generalisations of the additive characters $n \mapsto e(\alpha n)$ from Fourier analysis.

The second one [47] establishes that the Möbius function, and thus $\Lambda'_{\widetilde{W},a} - 1$, does not correlate with nilsequences. In Chapter 7, we revisit these theorems in the realm of function fields.

This concludes our overview of the Green-Tao method. The next two chapters rely on it.

Chapter 3

A higher-dimensional Siegel-Walfisz theorem

This chapter, based on the author's publication [9], presents an extension of Theorem 2.1. Recall from the introduction that an *admissible* system of linear forms is a system of finite complexity whose local factors β_p (introduced in equation (1.4)) are all nonzero. Here then is the chapter's main theorem.

Theorem 3.1. *Let d, t be positive integers and A, B, L be positive constants. Assume that $\Psi = (\psi_1, \ldots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ is an admissible system. Suppose that $\|\Psi\|_N \leq L \log^B N$ and that $K \subset [-N, N]^d$ is a convex body satisfying $\mathrm{Vol}(K) \gg N^d \log^{-A} N$ and $\Psi(K) \subset \mathbb{R}_+^t$. Then*

$$\sum_{n \in \mathbb{Z}^d \cap K} \prod_{i=1}^{t} \Lambda(\psi_i(n)) = \mathrm{Vol}(K) \prod_p \beta_p (1 + o_{d,t,A,B,L}(1)). \tag{3.1}$$

Note that from now on, we abandon the convention that tuples of integers should be bold, considering that the distinction between integers and tuples thereof is now sufficiently clear without this typographic help. The product $\prod_p \beta_p$ still converges, as Lemma 2.3 still applies; the set of exceptional primes P_Ψ is finite by Lemma 2.2, even though its cardinality may tend to infinity with N.

Some special cases, Theorem 3.1 follow easily from the work of Green and Tao. For instance, Proposition 2.6 gives an asymptotic for the unbounded system $W\Psi + b$ where Ψ is a bounded system, $W = \prod_{p \leq w} p = O(\log N)$ and $b = (b_1, \ldots, b_t) \in [W]^t$ is a t-tuple of integers coprime to W. More generally, Proposition 2.6 implies that an unbounded system $q\Psi + b$ with $q = O(\log^{O(1)} N)$ and Ψ bounded is tractable, via an asymptotic for the system $\widetilde{W}\Psi + c$ where $\widetilde{W} = Wq$. By decomposing into residue classes, this method extends to

systems Ψ such that for each j, the coefficients $(\dot{\psi}_i(e_j))_{i\in[t]}$ are bounded multiples of a common coefficient q_j. We illustrate this with an example, which corresponds to the count of k-term progressions of primes whose common difference is a multiple of q. We have

$$\sum_{\substack{1\le n,d\\ n+(k-1)qd\le N}} \prod_{i=0}^{k-1} \Lambda(n+iqd) = \sum_{a\in[q]} \sum_{\substack{1\le n,d\\ n+(k-1)d\le \frac{N-a}{q}}} \prod_{i=0}^{k-1} \Lambda(q(n+id)+a)$$

and are thus left with a system of the form $q\Psi + b$ with Ψ bounded.

We now provide some less immediate examples where Theorem 3.1 applies.

Example 1. What is the proportion of arithmetic progressions $n+d\mathbb{N}$ whose q_1th,...,q_kth terms are all primes? Assume that $q_i = \lfloor \log^i N \rfloor$. The answer is given by

$$\sum_{1\le n,d\le N} \prod_{i=1}^{k} \Lambda(n+q_i d).$$

For this system, the factors β_p can be easily expressed, using the notation $h(p)$ for the number of classes modulo p occupied by $q_1,\dots,q_k$, as

$$\beta_p = \left(\frac{p}{p-1}\right)^k \frac{(p-1)(1+p-h(p))}{p^2}.$$

Example 2. We can also count k-term arithmetic progressions of primes up to N whose common difference is $q = \lfloor \log N \rfloor$ times a prime. This time the sum to consider is

$$\sum_{1\le n\le n+(k-1)qd\le N} \Lambda(d) \prod_{i=0}^{k-1} \Lambda(n+iqd).$$

To simplify the expression of the local factors, assume $\prod_{p\le k} p \mid q$. Then

$$\beta_p = \left(\frac{p}{p-1}\right)^{k+1} \frac{1}{p^2} \begin{cases} (p-1)^2 & \text{if } p \mid q \\ (p-1)(p-k) & \text{if } p \nmid q. \end{cases}$$

Example 3. We provide the asymptotic count of solutions to linear equations in the shifted squarefree primes, that is, primes p for which $p-1$ is squarefree. As it is not a direct application, we give the details in the final section of this chapter.

In view of the Siegel-Walfisz theorem (1.3), one may hope to write

$$\sum_{n\in\mathbb{Z}^d\cap K}\prod_{i=1}^{t}\Lambda(\psi_i(n)) = \mathrm{Vol}(K)\left(\prod_p \beta_p + o_{d,t,A,B,L}(1)\right)$$

instead of the estimate (3.1), but unfortunately our method does not yield this. Such an estimate is genuinely stronger than (3.1) given that $\prod_p \beta_p$ may well tend to infinity with N, if the linear coefficients do. This weaker bound is ultimately due to the ineffectiveness of the Gowers norm estimate in Proposition 2.14.

To prove Theorem 3.1, we first get rid of the convex body by decomposing it into reasonably small boxes, so that the theorem simply needs to be proven on boxes. In this context, the variables all have the same range and are independent of each other, which makes it possible, after the introduction of the W-trick, to prove a suitable von Neumann theorem. Indeed, as we shall see, the von Neumann theorem (Theorem 2.8) of Green and Tao does not apply when the linear coefficients are unbounded.

3.1 First reductions

As discussed in Section 2.1, we may assume that $\psi_i > N^{9/10}$ on K for each i and replace Λ by Λ'. In the next proposition, analogous to Proposition 2.6, we check that the normal form parametrisation goes through as usual in spite of the presence of large coefficients.

Proposition 3.2. *Let d, t be positive integers and A, B, L be positive constants. Assume $\Psi = (\psi_1, \dots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ is an admissible system in normal form. Suppose that $\|\Psi\|_N \leq L\log^B N$ and that $K \subset [-N, N]^d$ is a convex body satisfying $\mathrm{Vol}(K) \gg N^d \log^{-A} N$ and $\Psi(K) \subset [N^{9/10}, +\infty)^t$. Then*

$$\sum_{n\in\mathbb{Z}^d\cap K}\prod_{i=1}^{t}\Lambda'(\psi_i(n)) = \mathrm{Vol}(K)\prod_p \beta_p(1 + o_{d,t,A,B,L}(1)). \tag{3.2}$$

Proof. This is a straightforward application of Proposition 2.5. Indeed, with the notation of that proposition, we have $\|\Psi'\|_N = \|\Psi\|_N^{O(1)} = O(\log^{O(1)} N)$ and $K' \subset [-N', N']$ with $N' = O(N)$ and

$$\mathrm{Vol}(K') \gg N'^d \log^{-D} N'$$

for some constant D. Finally, the local factors are left unchanged by this operation. In

particular, if the system Ψ is admissible, so is Ψ', so that Proposition 3.2 can be applied to Ψ' and K', which concludes the proof of the reduction. $\square$

3.2 Reduction to the case of a box

The convex body being potentially somewhat unbalanced, it is slightly awkward to handle, as all variables do not have the same range. It is much more convenient when $K = [1, M]^d$ is a box, because in that case, the variables $n_1, \dots, n_d$ have independent and equal ranges. Fortunately, some simple geometric arguments allow one to decompose K into boxes, and so to reduce Theorem 3.1 to the following statement. We introduce the notation $\|\Psi\|_{N,b} = \frac{1}{\log^b N}\|\Psi\|_N$.

Proposition 3.3. *Let d, t be positive integers and A, b, L be positive constants. Let $\Psi = (\psi_1, \dots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ be an admissible system in normal form. Suppose that $\|\Psi\|_{M,b} \leq L$ and that $\Psi([M]^d) \subset [M^{9/10}, +\infty)^t$. Then*

$$\sum_{n \in [M]^d} \prod_{i=1}^{t} \Lambda'(\psi_i(n)) = M^d \prod_p \beta_p \left(1 + o_{d,t,b,L}(1)\right).$$

Proof that Proposition 3.3 implies Proposition 3.2. Let $K \subset [-N, N]^d$ be a convex body satisfying $\mathrm{Vol}(K) \gg N^d \log^{-A} N$. Let

$$K' = \{x \in K \mid d(x, \partial K) \geq N \log^{-A-1} N\}$$

and

$$K'' = \{x \in \mathbb{R}^d \mid d(x, K) \leq N \log^{-A-1} N\}.$$

These are two convex bodies. The arguments from elementary convex geometry contained in [45, Appendix A] allow one to infer that

$$\mathrm{Vol}(K') = \mathrm{Vol}(K) + O(N^d \log^{-A-1} N) = \mathrm{Vol}(K)(1 + o(1)),$$

and similarly for K''. Now let $M = N \log^{-A-1} N/\sqrt{d}$ and consider the grid $(M\mathbb{Z})^d$. Let $\mathcal{B} = \{c + [M]^d \mid c \in J\}$ be the collection of boxes defined by this grid that are included in K, and let $\mathcal{B}' = \{c + [M]^d \mid c \in J'\}$ be the collection of boxes defined by this grid that

meet K. Note that

$$K' \cap \mathbb{Z}^d \subset \bigcup_{B \in \mathcal{B}} B \subset K \cap \mathbb{Z}^d \subset \bigcup_{B \in \mathcal{B}'} B \subset K'' \cap \mathbb{Z}^d. \tag{3.3}$$

The first inclusion is because if a box B from the grid meets K', then it is included in K.

Now let $\Psi = (\psi_1, \ldots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of affine-linear forms of finite complexity. Suppose that $\|\Psi\|_{M,b} \leq L$ and that $\Psi([M]^d) \subset [M^{9/10}, +\infty)^t$. Equation (3.3) implies that

$$\sum_{B \in \mathcal{B}} \sum_{n \in B} \prod_{i \in [t]} \Lambda(\psi_i(n)) \leq \sum_{n \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} \Lambda(\psi_i(n)) \leq \sum_{B \in \mathcal{B}'} \sum_{n \in B} \prod_{i \in [t]} \Lambda(\psi_i(n)). \tag{3.4}$$

Now if $B = c + [M]^d$ with $c \in \mathbb{Z}^d$, letting $\Psi_c = \Psi(c) + \dot{\Psi}$, we can write

$$\sum_{n \in B} \prod_{i \in [t]} \Lambda(\psi_i(n)) = \sum_{n \in [M]^d} \prod_{i \in [t]} \Lambda(\psi_{c,i}(n)).$$

We check that the system Ψ_c satisfies $\|\Psi_c\|_{M,C} = O(1)$ for some constant C. Indeed, the linear coefficients are unchanged, and thus still of size $O(\log^b N) = O(\log^b M)$, and the constant coefficients are of size $O(N \log^b N)$, hence $O(M \log^{A+b+1} M)$. It follows that we may take $C = A + b + 1$. Moreover, we have $\Psi_c([M]^d) \subset [N^{9/10}, +\infty)^t$, so we can apply Proposition 3.3. We note that the local factor $\beta_{p,c}$ corresponding to the system Ψ_c is in fact independent of c, because the translation invariance of $\mathbb{Z}/p\mathbb{Z}$ allows one to write

$$\begin{aligned} \mathbb{E}_{a \in (\mathbb{Z}/p\mathbb{Z})^d} \prod_{i \in [t]} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(a) + \dot{\psi}_i(c)) &= \mathbb{E}_{a \in (\mathbb{Z}/p\mathbb{Z})^d} \prod_{i \in [t]} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(a + c)) \\ &= \mathbb{E}_{a \in (\mathbb{Z}/p\mathbb{Z})^d} \prod_{i \in [t]} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(a)). \end{aligned}$$

Consequently, the application of Proposition 3.3 on the rightmost and leftmost sides of equation (3.4) yields

$$|\mathcal{B}| M^d \prod_p \beta_p (1 + o(1)) \leq \sum_{n \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} \Lambda(\psi_i(n)) \leq |\mathcal{B}'| M^d \prod_p \beta_p (1 + o(1)). \tag{3.5}$$

Because of the inclusions (3.3), we see that

$$\mathrm{Vol}(K)(1 + o(1)) = \mathrm{Vol}(K') \leq |\mathcal{B}| M^d \leq |\mathcal{B}'| M^d \leq \mathrm{Vol}(K'') = \mathrm{Vol}(K)(1 + o(1)).$$

Hence, we have $|\mathcal{B}|M^d = \mathrm{Vol}(K)(1+o(1))$ as well as $|\mathcal{B}'|M^d = \mathrm{Vol}(K)(1+o(1))$. Together with the lower and upper bounds in (3.5), these asymptotics complete the proof of Proposition 3.2. □

3.3 The W-trick

In Section 2.2, we introduced $W = \prod_{p \le w} p$, for a sufficiently slowly increasing parameter $w = w(N)$, in order to deal with the biases induced by small primes. We put $\widetilde{W} = WQ$ for some parameter $Q = O(\log^{O(1)} N)$ to deal with biases potentially introduced by larger primes. A simple one-dimensional example shows that in our situation where coefficients are unbounded, we really need such a carefully chosen Q. Indeed, consider the system consisting of one form in one variable, namely $n \mapsto qn + b$, with q of size roughly $\log N$. The W-trick consists in writing

$$\sum_{n \le N} \Lambda(qn+b) = \sum_{\substack{a \in [W] \\ (qa+b,W)=1}} \frac{W}{\varphi(W)} \sum_{n \le N/W} \frac{\varphi(W)}{W} \Lambda(Wqn + qa + b).$$

But imposing that $(qa+b, W) = 1$ does not ensure that the inner sum is $N/W(1+o(1))$, because $qa+b$ could well have a common factor greater than w with q: when the coefficients are bounded, their factors are all less than w for large enough N, but this is no longer the case in our setting. Moreover the relevant average is not $W/\varphi(W)$ but $Wq/\varphi(Wq)$ which may be different if q has prime factors larger than w.[1] This suggests that the coefficients of the system have to be taken into account when determining a suitable parameter $\widetilde{W}$ instead of W.

We fix an admissible system $\Psi_1 = (\psi_1, \dots, \psi_t) : \mathbb{Z}^{d_1} \to \mathbb{Z}^{t_1}$ in normal form satisfying $\|\Psi_1\|_{M,B} \le L$ for some constants $B, L > 0$. Let

$$Q = \prod_{\substack{i \in [t_1], j \in [d_1] \\ \dot{\psi}_i(e_j) \neq 0}} \dot{\psi}_i(e_j) \times \prod_{\substack{1 \le i < k \le t_1 \\ 1 \le j < \ell \le d_1 \\ \dot{\psi}_i(e_j)\dot{\psi}_k(e_\ell) - \dot{\psi}_i(e_\ell)\dot{\psi}_k(e_j) \neq 0}} (\dot{\psi}_i(e_j)\dot{\psi}_k(e_\ell) - \dot{\psi}_i(e_\ell)\dot{\psi}_k(e_j)) \tag{3.6}$$

[1] Nevertheless, it is easy to check using Mertens' theorem that if $w = \log\log N$ and $q \le \log^B N$ that

$$Wq/\varphi(Wq) = (1 + o_B(1))W/\varphi(W).$$

be the product of the nonzero minors of size 1 and 2 in the matrix $(\dot{\psi}_i(e_j))_{i,j}$; thus $Q = O_L(\log^{O_{d,t,B}(1)} N)$. The proof of Lemma 2.2 reveals that if a prime p is exceptional for Ψ_1, then it must divide Q.

We will now state a W-tricked reduction of Theorem 3.3, in the same way that we reduced Theorem 2.1 to Proposition 2.7. The latter reduction shows that it is enough to prove

$$\sum_{n\in[M/\widetilde{W}]^{d_1}} (\prod_{i\in[t_1]} \Lambda'_{\widetilde{W},b_i}(\psi_i(n)) - 1) = o((M/\widetilde{W})^{d_1}) \tag{3.7}$$

for any $b_1, \ldots, b_t \in [\widetilde{W}]$ coprime to $\widetilde{W}$.

We then use the trivial identity $\Lambda'_{\widetilde{W},b_i}(\psi_i(n)) = (\Lambda'_{\widetilde{W},b_i}(\psi_i(n)) - 1) + 1$ for each i. This decomposes the left-hand side of equation (3.7) into 2^{t_1} sums, each sum featuring a *subsystem* of Ψ_1, that is, a system $\Psi' = (\psi_{i_1}, \ldots, \psi_{i_s})$ for some sequence $1 \leq i_1 < \ldots < i_s \leq t_1$, whence the following reduction.

Proposition 3.4. *Let $\Psi_0 = (\psi_1^0, \ldots, \psi_{t_0}^0) : \mathbb{Z}^{d_0} \to \mathbb{Z}^{t_0}$ be a subsystem of Ψ_1. Suppose that $\Psi_0([M]^{d_0}) \subset ([M^{8/10}, +\infty))^{t_0}$ and that $b_i \in [\widetilde{W}]$ is coprime to $\widetilde{W}$ for any $i \in [t_0]$. Then*

$$\sum_{n\in[M/\widetilde{W}]^{d_0}} \prod_{i\in[t_0]} (\Lambda'_{\widetilde{W},b_i}(\psi_i^0(n)) - 1) = o((M/\widetilde{W})^{d_0}). \tag{3.8}$$

3.4 Reduction to a Gowers norm estimate

Write $X = M/\widetilde{W}$ and fix a system Ψ_0 and a tuple $b_1, \ldots, b_{t_0}$ satisfying the conditions of Proposition 3.4. If $t_0 = 1$, this proposition follows directly from the one-dimensional Siegel-Walfisz theorem (1.3), so we suppose that $t_0 \geq 2$. Let Q_0 be the product of 2×2 minors for the system Ψ_0 as defined by equation (3.6). In particular, $Q_0 \mid Q$. We have to prove that

$$\sum_{n\in[X]^d} \prod_{i\in[t_0]} F_i(\psi_i^0(n)) = o(X^d)$$

for $F_i = \Lambda'_{\widetilde{W},b_i} - 1$.

We would like to apply Theorem 2.8 at this point, in order to reduce the problem to the Gowers-norm estimate (2.22). Unfortunately, Theorem 2.8 requires bounded coefficients so we cannot apply it. The purpose of this section is therefore to prove a replacement for Theorem 2.8.

3.4.1 A pseudorandom majorant

Theorem 2.8 required some notion of pseudorandomness. In our setting, it turns out to be convenient to modify this notion. Thus in this subsection, we introduce a variant of the notion of a pseudorandom measure from Definition 2.6.

Recall that we have fixed an admissible system

$$\Psi_0 = (\psi_i^0)_{i\in[t_0]} : \mathbb{Z}^{d_0} \longrightarrow \mathbb{Z}^{t_0}.$$

First, we introduce the notion of a *derived system*. It captures the important properties of those systems that arise from the original one by repeated applications of the Cauchy-Schwarz inequality.

Definition 3.1. A system $\Psi : \mathbb{Z}^d \to \mathbb{Z}^t$ of affine-linear forms is said to be *derived* from Ψ_0 if the following conditions are satisfied:

- $d \leq 2d_0$;
- $t \leq 2^{d_0} t_0$;
- $\|\Psi\|_{N,B} \ll \|\Psi_0\|_{N,B}$;
- any exceptional prime for Ψ divides Q_0.

We are now able to propose our new definition of a pseudorandom measure.

Definition 3.2. We say that a function $\nu : [Z] \to \mathbb{R}_+$ satisfies the Ψ_0-*linear forms condition* if for any system $\Psi = (\psi_1, \ldots, \psi_t)$ derived from Ψ_0 we have

$$\mathbb{E}_{n\in[Z]^d} \prod_{i\in[t]} \nu(\psi_i(n)) = 1 + o(1)$$

as $Z \to +\infty$. We also say that ν is a Ψ_0-*pseudorandom measure*.

The next proposition is about the existence of a pseudorandom majorant for a $\widetilde{W}$-tricked von Mangoldt function.

Proposition 3.5. *For any integers $b_1, \ldots, b_{t_0}$ in $[\widetilde{W}]$ coprime to $\widetilde{W}$, for $Z \gg N \log^{-O(1)} N$, there exists a Ψ_0-pseudorandom measure ν on $[Z]$ such that*

$$1 + \Lambda'_{\widetilde{W},b_1} + \cdots + \Lambda'_{\widetilde{W},b_{t_0}} \ll \nu \tag{3.9}$$

on $[Z^{3/5}, Z]$.

The construction of the majorant was explained in Section 2.3. Its Ψ_0-pseudorandomness follows from Proposition 2.13.

3.4.2 Generalised von Neumann theorem

We now prove the announced variant of Theorem 2.8. The proof of this theorem [45, Proposition 7.1] starts by embedding the convex body in a large discrete torus $(\mathbb{Z}/M\mathbb{Z})^d$ for M prime, in such a way that the convex body is dense in it and its image under Ψ involves no "wrap-around". This is impossible with unbounded coefficients.

Recall that $\Psi_0 : \mathbb{Z}^{d_0} \to \mathbb{Z}^{t_0}$ is a fixed system of affine-linear forms in s-normal form; thus without loss of generality, we can write its first form as

$$\psi_1(n_1, \dots, n_{s+1}, y) = q_1 n_1 + \cdots + q_{s+1} n_{s+1} + \psi_1(0, y)$$

with $q_i \neq 0$ for all i and $\prod_{j\in[s+1]} \dot{\psi}_i(e_j) = 0$ for all $i > 1$. Here $y = (n_{s+2}, \dots, n_d)$ is the projection of $n \in \mathbb{Z}^d = \mathbb{Z}^{s+1} \times \mathbb{Z}^{d-s-1}$ onto the second factor of the Cartesian product and will not play as important a role as the first $s+1$ variables. We have dropped the superscript 0 from the forms of the system Ψ_0 and shall always do so in the sequel. We now state our variant of the von Neumann theorem [9, Theorem 5.2].

Theorem 3.6. *Let $f_1, \dots, f_{t_0} : \mathbb{Z} \to \mathbb{R}$ be functions and ν be a Ψ_0-pseudorandom measure such that $|f_i| \leq \nu$ for all i. Then*

$$\begin{aligned} &|\mathbb{E}_{n\in[X]^d} \prod_{i\in[t_0]} f_i(\psi_i(n))| \\ &\qquad \leq |\mathbb{E}_{y\in[X]^{d-s-1}} \mathbb{E}_{n^{(0)},n^{(1)}\in[X]^{s+1}} \prod_{\omega\in\{0,1\}^{s+1}} f_1(\sum_{i=1}^{s+1} q_i n_i^{(\omega_i)} + \psi_1(0,y))|^{1/2^{s+1}} + o(1) \quad (3.10) \end{aligned}$$

We adapt the proof of Proposition 7.1" in [45]. To that aim, we need some notation. For $x \in \mathbb{Z}^{s+1}$, and $B \subset [s+1]$, write $x_B = (x_i)_{i\in B}$. For $i \in [t]$, let $\Omega(i) \subset [s+1]$ be the subset of the first $s+1$ variables that ψ_i genuinely uses, that is, $\{j \in [s+1] \mid \dot{\psi}_i(e_j) \neq 0\}$. Thus $\Omega(i) = [s+1]$ if and only if $i = 1$. For $B \subset [s+1]$ and $(x, y) \in \mathbb{Z}^{s+1} \times \mathbb{Z}^{d-s-1}$, we introduce

$$F_{B,y}(x_B) = \prod_{i:\Omega(i)=B} f_i(\psi_i(x_B, y))$$

where we naturally view ψ_i as a linear map $\mathbb{Z}^B \times \mathbb{Z}^{d-s-1} \to \mathbb{Z}$. With this notation, the left-hand side of equation (3.10) equals

$$|\mathbb{E}_{y\in[X]^{d-s-1}}\mathbb{E}_{x\in[X]^{s+1}} \prod_{B\subseteq[s+1]} F_{B,y}(x_B)|. \tag{3.11}$$

For the moment, we fix $y \in [X]^{d-s-1}$. For $B = [s+1]$, in particular, we observe that

$$F_{[s+1],y}(x_{[s+1]}) = f_1(\sum_{i\in[s+1]} q_i x_i + \psi(0,y)).$$

Similarly, we define

$$\nu_{B,y}(x_B) = \prod_{i:\Omega(i)=B} \nu(\psi_i(x_B, y)).$$

The functions $F_{B,y}$ and $G_{B,y}$ are both functions on the Cartesian product $[X]^B$. We think of them as purely set-theoretic objects, ignoring the arithmetic background. We have the bound

$$|F_{B,y}| \leq \nu_{B,y}.$$

Next we define Gowers box norms [45, Appendix B] relative to the family $(\nu_{B,y})_{B\subset[s+1]}$ by putting

$$\|G\|^{2^{|B|}}_{\square(\nu_{B,y})} = \mathbb{E}_{x^{(0)},x^{(1)}\in[X]^B} \prod_{\omega\in\{0,1\}^B} F(x^{(\omega)}) \prod_{C\subsetneq B} \nu_{C,y}((x_i^{(\omega_i)})_{i\in C})$$

for any $B \subset [s+1]$ and any function $G : [X]^B \to \mathbb{R}$. We now apply [45, Corollary B.4], with $A = [s+1]$ and $X_\alpha = [X]$ for all $\alpha \in [s+1]$, which implies that

$$|\mathbb{E}_{x\in[X]^{s+1}} \prod_{B\subseteq[s+1]} F_{B,y}(x_B)| \leq \|F_{[s+1],y}\|_{\square(\nu_{[s+1],y})} \prod_{B\subsetneq[s+1]} \|\nu_{B,y}\|^{2^{|B|-(s+1)}}_{\square(\nu_{B,y})}.$$

Averaging over y and using the triangle inequality, we bound expression (3.11) by

$$\mathbb{E}_{y\in[X]^{d-s-1}} \|F_{[s+1],y}\|_{\square(\nu_{[s+1],y})} \prod_{B\subsetneq[s+1]} \|\nu_{B,y}\|^{2^{|B|-(s+1)}}_{\square(\nu_{B,y})}. \tag{3.12}$$

By Hölder's inequality, in order to prove the bound (3.10), it suffices to show that

$$\mathbb{E}_{y\in[X]^{d-s-1}}\|F_{[s+1],y}\|^{2^{s+1}}_{\square(\nu_{[s+1],y})}$$
$$= \mathbb{E}_{y\in[X]^{d-s-1}}\mathbb{E}_{n^{(0)},n^{(1)}\in[X]^{s+1}}\prod_{\omega\in\{0,1\}^{s+1}} f_1\left(\sum_{i\in[s+1]} q_i n_i^{(\omega_i)} + \psi_1(0,y)\right) + o(1) \quad (3.13)$$

and that

$$\mathbb{E}_{y\in[X]^{d-s-1}}\|\nu_{B,y}\|^{2^{|B|}}_{\square(\nu_{B,y})} = 1 + o(1) \quad (3.14)$$

for all non empty $B \subseteq [s+1]$. To prove the latter, expand the left-hand side as

$$\mathbb{E}_{y\in[X]^{d-s-1}}\mathbb{E}_{n^{(0)},n^{(1)}\in[X]^{B}}\prod_{C\subseteq B}\prod_{i:\Omega(i)=C}\prod_{\omega\in\{0,1\}^{C}}\nu(\psi_i((n_j^{(\omega_j)})_{j\in\Omega(i)},y)), \quad (3.15)$$

which is an expression involving the average of ν on a system

$$\Psi = (\psi_{i,\omega})_{i\in[t_0],\omega\in\{0,1\}^{\Omega(i)}} : \mathbb{Z}^d \to \mathbb{Z}^t$$

of linear forms. Let us prove that this system is derived from Ψ_0 in the sense of Definition 3.1, which will allow us to apply the linear forms condition of Definition 3.2 to the expression (3.15). It easy to check that $d \leq 2d_0$ and $t \leq 2^{d_0}t_0$. It is also obvious that $\|\Psi\|_{M,B} \ll \|\Psi_0\|_{M,B}$. Let p be an exceptional prime for Ψ and let us check that it divides Q_0. Suppose that $\psi_{i,\omega} \neq \psi_{k,\alpha}$ are two forms that are affinely related modulo p. Then if $i \neq k$, we conclude that ψ_i and ψ_k are related and thus the prime is exceptional for Ψ_0, which implies that it divides Q_0. Otherwise $i = k$ and thus $\omega \neq \alpha$, in other words there exists $j \in [d_0]$ such that $\dot{\psi}_i(e_j) \neq 0$ and $\omega_j \neq \alpha_j$. Thus p must divide $\dot{\psi}_i(e_j)$ and hence also Q_0. This concludes the proof of the asymptotic (3.14).

It remains to verify (3.13). At this point, Green and Tao use the translation invariance of $\mathbb{Z}/N'\mathbb{Z}$ to perform a change of variable which is not possible here, but we make do without it. As the system is in normal form and $t \geq 2$, the form ψ_2 must also have its set of $s+1$ variables that it is the only one to use fully. In particular, ψ_1 does not use all d variables. Without loss of generality, let us thus assume that ψ_1 only uses $x_1, \ldots, x_{d-k}$ where $1 \leq k \leq d-(s+1)$, which enables us, by a slight abuse of notation, to regard ψ_1 as

a map from $\mathbb{Z}^{d-k}$ to $\mathbb{Z}$. Upon expanding the norm, the left-hand side of (3.13) becomes

$$\mathbb{E}_{\substack{x^{(0)},x^{(1)}\in[X]^{s+1}\\ y\in[X]^{d-k-s-1}}} \prod_{\omega\in\{0,1\}^{s+1}} f_1\left(\sum_{i=1}^{s+1} q_i x_i^{(\omega_i)} + \psi_1(0,y)\right) \mathbb{E}_{z\in[X]^k} \prod_{\omega\in\{0,1\}^{s+1}} \prod_{C\subsetneq[s+1]} \nu_{C,(y,z)}(x_C^{(\omega_C)})$$

where (y,z) is the vector in $\mathbb{Z}^{d-s-1}$ obtained by concatenating y and z. We want to replace the inner expectation over z, which is a function of $(x^{(0)}, x^{(1)}, y)$ of average $1+o(1)$, by the constant 1. To do that, by Cauchy-Schwarz, it is enough to prove

$$\mathbb{E}_{\substack{x^{(0)},x^{(1)}\in[X]^{s+1}\\ y\in[X]^{d-k-s-1}}} \prod_{\omega\in\{0,1\}^{s+1}} \nu(\sum_{i=1}^{s+1} q_i x_i^{(\omega_i)} + \psi_1(0,y)) = 1+o(1) = O(1),$$

which follows directly from the linear forms condition, and

$$\mathbb{E}_{\substack{x^{(0)},x^{(1)}\in[X]^{s+1}\\ y\in[X]^{d-k-s-1}}} \prod_{\omega\in\{0,1\}^{s+1}} \nu\left(\sum_{i=1}^{s+1} q_i x_i^{(\omega_i)} + \psi_1(0,y)\right) |\mathbb{E}_z W(x,y,z) - 1|^2 = o(1),$$

where $W(x,y,z) = \prod_{\epsilon\in\{0,1\}^{s+1}} \prod_{C\subsetneq[s+1]} \nu_{C,(y,z)}(x_C^{(\epsilon_C)})$. So it is enough to prove that

$$\mathbb{E}_{\substack{x^{(0)},x^{(1)}\in[X]^{s+1}\\ y\in[X]^{d-k-s-1}}} \prod_{\omega\in\{0,1\}^{s+1}} \nu\left(\sum_{i=1}^{s+1} q_i x_i^{(\omega_i)} + \psi_1(0,y)\right) (\mathbb{E}_z W(x,y,z))^j = 1+o(1)$$

for $j=0,1,2$. Let us inspect the left-hand side in the most intricate case, namely $j=2$, the other cases being similar. Upon expanding the square, we get an expectation over $x^{(0)}, x^{(1)}, y, z^{(0)}, z^{(1)}$, thus the system has at most $2d_0$ variables. There are 2^{s+1} forms arising from ψ_1 and at most $2^{s+2}(t-1)$ other forms, which means together at most $2^{d_0}t_0$ forms. Now the reasoning we used to analyse the average (3.15) also applies here and yields that the system is derived from Ψ_0. Thus the linear forms condition applies and equation (3.13) is proven, hence also Theorem 3.6.

3.4.3 A Gowers norm estimate

Together with the existence of a pseudorandom majorant provided by Proposition 3.5, Theorem 3.6 reduces Proposition 3.4 to the following.

Proposition 3.7. *Let $b \in [\widetilde{W}]$ be coprime to $\widetilde{W}$. Let $B > 0$ and $d \in \mathbb{N}$ be constants. Suppose that $q_1, \dots, q_d$ are divisors of Q satisfying $q_i = O(\log^B N)$ while $c = O(N \log^B N)$.*

Then we have

$$\mathbb{E}_{x^{(0)},x^{(1)}\in[X]^d}\prod_{\omega\in\{0,1\}^d}(\Lambda'_{\widetilde{W},b}(\sum_{i=1}^d q_i x_i^{(\omega_i)}+c)-1)=o(1). \tag{3.16}$$

Compared to Proposition 3.4, we have made progress in that each variable $x_i^{(\epsilon)}$ for $i\in[d]$ and $\epsilon\in\{0,1\}$ is affected throughout the system by one and the same coefficient q_i. We now attempt to transform the system so that all variables have the same coefficient Q'; the price we pay is that the variables will no longer have the same ranges.

To this effect, we introduce

$$Q_i=\prod_{j\neq i}q_j$$

and variables $n_i^{(\omega_i)},m_i^{(\omega_i)}$ such that $x_i^{(\omega_i)}=Q_i n_i^{(\omega_i)}+m_i^{(\omega_i)}$. Then the left-hand side of equation (3.16) decomposes as

$$\mathbb{E}_{m_i^{(\omega_i)}\in[q_i]}\mathbb{E}_{n_i^{(\omega_i)}\in[X/q_i]}\prod_{\omega\in\{0,1\}^d}(\Lambda'_{\widetilde{W},b}(\sum_{i=1}^d Q'n_i^{(\omega_i)}+q_i m_i^{(\omega_i)}+c)-1)+o(1),$$

where $Q'=q_iQ_i$ for any i.

Fix two d-tuples $(m_i^{(0)})$ and $(m_i^{(1)})$ in $\prod_{i\in[d]}[X_i]$ where $X_i=X/q_i$. To prove Proposition 3.7, it suffices to prove that

$$\mathbb{E}_{n_i^{(\omega_i)}\in[X_i]}\prod_{\omega\in\{0,1\}^d}(\Lambda'_{\widetilde{W},b}(\sum_{i=1}^d Q'n_i^{(\omega_i)}+q_i m_i^{(\omega_i)}+c)-1)=o(1). \tag{3.17}$$

We recognise the function

$$n\mapsto F_a(n)=\frac{\varphi(\widetilde{W})}{\widetilde{W}}\Lambda'(Q'\widetilde{W}n+a)=\Lambda'_{Q'\widetilde{W},a},$$

where the last equality holds because $\varphi(\widetilde{W})/\widetilde{W}=\varphi(\widetilde{W}Q')/(\widetilde{W}Q')$ as $\varphi(x)/x$ depends only on $\mathrm{rad}(x)$. The parameters a occurring are

$$a_\omega=\widetilde{W}(\sum_{i=1}^d q_i m_i^{(\omega_i)}+c)+b.$$

Given that $(b, \widetilde{W}) = 1$, we also have $(a_\omega, \widetilde{W}) = 1$ and therefore $(a_\omega, \widetilde{W}Q') = 1$. Now with this notation, the left-hand side of equation (3.17) is

$$\mathbb{E}_{n_i^{(\omega_i)} \in [X_i]} \prod_{\omega \in \{0,1\}^d} (F_{a_\omega}(\sum_{i=1}^d n_i^{(\omega_i)}) - 1).$$

We observe that for any tuple $a \in [\widetilde{W}Q']^{2^d}$ of integers coprime to $\widetilde{W}Q'$, we can create a common Ξ-pseudorandom majorant for the functions $1 + F_{a_\omega}$, where $\Xi = (\xi_\omega)_{\omega \in \{0,1\}^d}$ is defined by

$$\xi_\omega = (n_1^{(0)}, \ldots, n_d^{(0)}, n_1^{(1)}, \ldots, n_d^{(1)}) \longmapsto \sum_{i=1}^d n_i^{(\omega_i)}.$$

In fact, thanks to Proposition 2.13, we can rewrite Proposition 3.5 with $\widetilde{W}Q'$ instead of $\widetilde{W}$, because Q' still satisfies $Q' = O(\log^{O(1)} N)$.

We now prove the bound (3.17). Observe that each X_i satisfies $N \log^{-C} N \ll X_i \leq N$. Letting $Z = \max_i X_i$ and $K = \prod_i [X_i]$, we have $K \subset [Z]^d$ and $\mathrm{Vol}(K) \gg Z^d \log^{-C'} Z$. Thus we can apply the same reasoning as in Section 3.2, where we approximated such a convex body K by a set of small boxes with equal side lengths,[2] and reduce to proving that

$$\mathbb{E}_{n^{(0)}, n^{(1)} \in [Y]^d} \prod_{\omega \in \{0,1\}^d} (F_{a_\omega}(\sum_{i=1}^d n_i^{(\omega_i)}) - 1) = o(1) \tag{3.18}$$

for some $Y \gg N \log^{-D} N$. Now that the linear forms have bounded coefficients (namely 0 and 1) and the average is on a box with equal sides, there is no more objection to the use of Green-Tao's generalised von Neumann theorem [45, Proposition 7.1], as long as the functions $\Lambda'_{Q'\widetilde{W}, a_\omega} - 1$ are dominated by a pseudorandom measure in the sense of Definition 2.6, which follows from Proposition 2.12. Thus equation (3.18) follows from the claim that

$$\|F_a - 1\|_{U^k([Y])} = \|\Lambda'_{Q'\widetilde{W}, a} - 1\|_{U^k([Y])} = o(1) \tag{3.19}$$

for any $a \in [Q'\widetilde{W}]$ coprime to $Q'\widetilde{W}$. Equation (3.19) itself follows from Theorem 2.14. This completes, at last, the proof of Theorem 3.1.

[2]The reader might object that we then used the positivity of the function to average, which is not available here, but we can just as well use the majorant and the linear forms condition to bound the contribution of the few boxes included in K'' but not in K'.

3.5 Application to linear equations in the set of primes p such that $p-1$ is squarefree

Theorem 1.2 actually holds in any dense subset of the primes [44]. That is, any subset A of the set $\mathcal{P}$ of primes such that $\limsup \frac{|A\cap[N]|}{|\mathcal{P}\cap[N]|} > 0$ contains arbitrarily long arithmetic progressions. However, it is in general impossible to determine an asymptotic for the number of k-term arithmetic progressions in A, except in the special case where A is the intersection of $\mathcal{P}$ with a congruence class a mod b where $(a,b)=1$: indeed, in that case, we can apply Theorem 2.1 with the forms $\psi_i(n,d) = a(n+(i-1)d)+b$ for $i \in [k]$. In this section, as an application of Theorem 3.1, we derive an asymptotic for arithmetic progressions (or general linear configurations) inside a special subset A of the primes. The set A we shall consider is the set of squarefree shifted primes, i.e. the primes p such that $p-1$ is squarefree. By a theorem of Mirsky [71], it is a dense subset of the primes, of density $\sum_a \frac{\mu(a)}{\varphi(a^2)} = \prod_p (1-1/p(p-1))$.

For convenience, let F be the von Mangoldt function restricted to the squarefree shifted primes, that is $F(n) = \Lambda(n+1)\mu^2(n)$. We now state this section's main theorem [9, Proposition 7.1].

Theorem 3.8. *Let $\Psi : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of affine-linear forms of finite complexity and let $K \subset [-N,N]^d$ be a convex body. Suppose that the linear coefficients are $O(1)$, the constants ones are $O(N)$ and that $\Psi(K) \subset \mathbb{R}_+^t$. Then there exists a constant $C(\Psi)$ (possibly equal to 0) such that*

$$\sum_{n\in K\cap\mathbb{Z}^d} \prod_{i\in[t]} F(\psi_i(n)) = C(\Psi)\mathrm{Vol}(K) + o(N^d). \tag{3.20}$$

The constant $C(\Psi)$ will appear explicitly in the proof as a convergent series, but it is possible to write it as a product

$$C(\Psi) = \prod_p \gamma_p$$

where

$$\gamma_p = \left(\frac{p}{p-1}\right)^t \mathbb{E}_{a\in(\mathbb{Z}/p^2\mathbb{Z})^d} \prod_{i\in[t]} 1_{\psi_i(a)+1\not\equiv 0 \bmod p} 1_{\psi_i(a)\not\equiv 0 \bmod p^2}.$$

As in Chapter 2 for local factors β_p, it is easy to show that $\gamma_p = 1 + O(p^{-2})$ when p tends to infinity. We infer that $C(\Psi) \neq 0$ unless there is some prime p such that for any $a \in (\mathbb{Z}/p^2\mathbb{Z})^d$, there is $i \in [t]$ such that $\psi_i(a)+1 \equiv 0 \bmod p$ or $\psi_i(a) \equiv 0 \bmod p^2$.

Throughout the proof of this theorem, we will need the notation

$$\alpha_\Psi(k_1,\dots,k_t) = \mathbb{E}_{a\in(\mathbb{Z}/m\mathbb{Z})^d}\prod_{i\in[t]} 1_{k_i|\psi_i(a)} \tag{3.21}$$

where $m = \mathrm{lcm}(k_1,\dots,k_t)$. According to Lemma A.2, we have

$$\sum_{n\in K\cap\mathbb{Z}^d}\prod_{i\in[t]} 1_{d_i|\psi_i(n)} = \mathrm{Vol}(K)\alpha_\Psi(d_1,\dots,d_t) + O(N^{d-1}\mathrm{lcm}(d_1,\dots,d_t)). \tag{3.22}$$

We now prove Theorem 3.8.

Proof. We substitute the formula $\mu^2(n) = \sum_{a^2|n}\mu(a)$ in the left-hand side of (3.20). Thus

$$\sum_{n\in K\cap\mathbb{Z}^d}\prod_{i\in[t]} F(\psi_i(n)) = \sum_{(a_1,\dots,a_t)\in\mathbb{N}^t}\prod_{i\in[t]}\mu(a_i)\sum_{\substack{n\in K\cap\mathbb{Z}^d\\ \forall i\in[t]\ a_i^2|\psi_i(n)}}\prod_{i\in[t]}\Lambda(\psi_i(n)+1). \tag{3.23}$$

Now for any $a = (a_1,\dots,a_t)\in\mathbb{N}^t$, we introduce the set

$$L_a = \{n\in\mathbb{Z}^d : \forall i\in[t]\quad a_i^2 \mid \psi_i(n)\}.$$

Fix an a for which $L_a\neq\emptyset$ and let $n_0\in L_a$. Then

$$L_a = n_0 + \bigcap_{i=1}^t \ker g_i$$

where $g_i : \mathbb{Z}^d\to\mathbb{Z}/a_i^2\mathbb{Z}$ is the group homomorphism obtained by applying $\dot{\psi}_i$ and then reducing modulo a_i^2. It follows that $\bigcap_{i=1}^t\ker g_i$ is a subgroup of $\mathbb{Z}^d$, that is, a *lattice*. We can see that it is a lattice of full rank, because it contains $\{\prod_i a_i^2e_1,\dots,\prod_i a_i^2e_d\}$. By analogy with affine spaces, we think of L_a as an *affine sublattice* of $\mathbb{Z}^d$ of *direction* $\overrightarrow{L_a} = \bigcap_{i=1}^t\ker g_i$

As a lattice of full rank, the direction $\overrightarrow{L_a}$ of L_a has a $\mathbb{Z}$-basis: there exist $f_1,\dots,f_d$ such that

$$L_a = \{n_0 + \sum_{i=1}^d m_if_i \mid (m_1,\dots,m_d)\in\mathbb{Z}^d\}.$$

Because of a theorem of Mahler, we can assume that $\|f_i\|\le i\lambda_i$ for $i = 1,\dots,d$, where $\lambda_1\le\dots\le\lambda_d$ are the successive minima of the lattice $\overrightarrow{L_a}$ with respect to the Euclidean

unit ball. Let R^a be the affine transformation of $\mathbb{R}^d$ defined by $R^a(0) = n_0$ and $\dot{R}^a(e_i) = f_i$ for each $i \in [d]$. Note that $L_a \cap K = R^a(\mathbb{Z}^d \cap K_a)$, where K_a is also a convex body. For the notions of geometry of numbers alluded to here, see for instance the notes of Green [41] or the classic book of Cassels [20] (Chapters I and VIII).

Now if one of the a_i is larger than $\log^C N$, then K_a is small. Indeed, the set of $n \in K \cap \mathbb{Z}^d$ such that there exists $i \in [t]$ and $a_i > \log^C N$ satisfying $a_i^2 \mid \psi_i(n)$ has $O(N^d \log^{-C} N)$ elements. This follows from equation (3.22) combined with the bound $\alpha_{\psi_i}(a_i^2) \ll a_i^{-2}$, deduced from Corollary A.4 and the fact that the coefficients of ψ_i are bounded, and finally $\sum_{a>x} a^{-2} \ll x^{-1}$. Bounding the contribution to the left-hand side of (3.23) of this exceptional set of $n \in K \cap \mathbb{Z}^d$ using $F \ll \log$, and supposing that $C \geq 2t$, we obtain

$$\begin{aligned}
\sum_{n \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} F(\psi_i(n)) &= \sum_{\substack{n \in K \cap \mathbb{Z}^d \\ \forall i \in [t]\, \forall a > \log^C N\, a^2 \nmid \psi_i(n)}} \prod_{i \in [t]} F(\psi_i(n)) + O(N^d \log^{-C/2} N) \\
&= \sum_{1 \leq a_1, \ldots, a_t \leq \log^C N} \prod_{i \in [t]} \mu(a_i) \sum_{n \in K \cap L_a} \prod_{i \in [t]} \Lambda(\psi_i(n) + 1) \\
&+ O(N^d \log^{-C/2} N).
\end{aligned}$$

For each $i \in [t]$, the map $\psi_i^a : L_a \to \mathbb{Z}$ defined by

$$\psi_i^a(n) = \frac{\psi_i(n)}{a_i^2}$$

is an affine map. Let $\phi_i^a = \psi_i^a \circ R^a$. These maps define a system $\Phi^a : \mathbb{Z}^d \to \mathbb{Z}^t$ of affine-linear forms which is again of finite complexity. Thus the inner sum on the right-hand side of equation (3.23) may be written as

$$\sum_{n \in K \cap L_a} \prod_i \Lambda(\psi_i(n) + 1) = \sum_{m \in K_a \cap \mathbb{Z}^d} \prod_i \Lambda(a_i^2 \phi_i^a(m) + 1). \tag{3.24}$$

We now apply Theorem 3.1 to this expression. One can check that the linear coefficients of Φ_a have size $O(\log^{O(1)} N)$. To do this, it is enough to examine the size of the basis vectors f_j of the lattice $\overrightarrow{L_a}$. Indeed,

$$a_i^2 |\dot{\phi_i^a}(e_j)| = |\dot{\psi}_i(f_j)| \leq \|\dot{\psi}_i\| \|f_j\| \ll \|f_j\|.$$

Moreover, the constant coefficients are $O(N)$. As observed, if $n_0 \in L_a$, the lattice

$$\{n_0 + \sum_{i\in[d]} k_i a_i^2 e_i \mid k \in \mathbb{Z}^d\}$$

is a sublattice of L_a and its determinant is $\prod_i a_i^2 \leq \log^{2dC} N$. Hence using Minkowski's second theorem [20, Chapter VIII], one finds that

$$\prod_{i\in[d]} \|f_i\| \leq d! \prod_{i\in[d]} \lambda_i \ll_d |\det L_a| \leq \log^{2dC} N.$$

Similarly, we obtain the bound

$$\mathrm{Vol}(K_a) = \mathrm{Vol}(K)\det(R^a)^{-1} \geq \mathrm{Vol}(K)\log^{-2dC} N.$$

Now Theorem 3.1 tells us that the right-hand side of (3.24) is equal to $\mathrm{Vol}(K_a)\prod_p \beta_p(1+o(1))$ as soon as none of the local factors $\beta_p(a)$ corresponding to the system of forms $a_i^2\phi_i^a + 1$ vanishes. Note that if any $\beta_p(a)$ is 0, then for all m there exists $i \in [t]$ such that $p \mid a_i^2\phi_i^a(m) + 1$. Then it is easy to see that

$$\sum_{m\in K_a\cap\mathbb{Z}^d} \prod_i \Lambda(a_i^2\phi_i^a(m) + 1) = O(N^{d-1}\log^{O(1)} N).$$

Moreover, equation (3.22) reveals that

$$\begin{aligned}\mathrm{Vol}(K_a) &= |K_a \cap \mathbb{Z}^d| + O(N^{d-1}) \\ &= |K \cap L_a| + O(N^{d-1}) \\ &= \mathrm{Vol}(K)\alpha_\Psi(a_1^2, \ldots, a_t^2) + O(N^{d-1}\log^{O(1)} N).\end{aligned}$$

Thus, up to an error term of size $O(N^d \log^{-C/2} N)$, the left-hand side of equation (3.20) equals

$$\mathrm{Vol}(K)(1+o(1)) \sum_{1\leq a_1,\ldots,a_t\leq \log^C N} \alpha_\Psi(a_1^2, \ldots, a_t^2) \prod_p \beta_p(a) \prod_{i\in[t]} \mu(a_i). \tag{3.25}$$

We claim that the sum over a is absolutely convergent. To see this, first observe that the exceptional primes for the system of forms $a_i^2\phi_i^a + 1$ are divisors of a_i^2 or exceptional primes for the system Φ^a; in either case, they are divisors[3] of a parameter $Q(a) = O(\prod_i a_i^{O(1)})$.

[3]See the proof of Lemma 2.2.

For all other primes, we have $\beta_p = 1 + O(p^{-2})$ by Lemma 2.3, so that

$$\prod_p \beta_p(a) \ll \prod_{p|Q(a)} \beta_p(a) \leq \left(\frac{Q(a)}{\varphi(Q(a))}\right)^t \ll (\log\log Q(a))^t \ll (\log\log \prod_i a_i)^t.$$

Next, note that the sum

$$\sum_{a_1,\ldots,a_t} (\log\log \prod_i a_i)^t \alpha_\Psi(a_1^2,\ldots,a_t^2)$$

is convergent. Indeed, we have the bound

$$\begin{aligned} \alpha_\Psi(a_1^2,\ldots,a_t^2) = \prod_p \alpha_\Psi(p^{2v_p(a_1)},\ldots,p^{2v_p(a_t)}) &\ll \prod_p p^{-2v_p(\max_i a_i)} \\ &= \operatorname{lcm}(a_1,\ldots,a_t)^{-2}, \end{aligned}$$

where the inequality holds because the forms ψ_i have bounded linear coefficients and if $p \nmid \psi_i$, then $\alpha_{\psi_i}(p^k) \leq p^{-k}$ (this is Corollary A.4). The convergence of the sum over a in equation (3.25) then follows from a trivial bound for the number of t-tuples a of prescribed least common multiple k, namely $\tau(k)^t$. This convergence result implies that

$$\sum_{1\leq a_1,\ldots,a_t\leq \log^C N} \alpha_\Psi(a_1^2,\ldots,a_t^2) \prod_p \beta_p(a) \prod_{i\in[t]} \mu(a_i) = C(\Psi) + o(1),$$

where

$$C(\Psi) = \sum_{(a_1,\ldots,a_t)\in\mathbb{N}^t} \alpha_\Psi(a_1^2,\ldots,a_t^2) \prod_p \beta_p(a) \prod_{i\in[t]} \mu(a_i).$$

This concludes the proof.

□

Chapter 4

Asymptotics for some polynomial patterns in the primes

This chapter is based on the author's publication [10]. It draws heavily on Matthiesen's paper on linear correlations of binary quadratic forms [68]. The author is thankful to Sean Prendiville for suggesting the problem.

4.1 The main theorem

The motivation for this chapter stems from the desire to derive asymptotics for polynomial configurations. In general, the Green-Tao method of deriving asymptotics for prime tuples does not work when linear systems are replaced by polynomial ones. In spite of this, the present chapter proposes an asymptotic for very specific polynomial configurations of primes, namely arithmetic progressions whose common difference is a sum of two squares, or more generally a number represented by a given quadratic forms. We introduce some terminology before stating our result. A *binary quadratic form* is a polynomial

$$f(x, y) = ax^2 + bxy + cy^2$$

where a, b and c are integers. Its *discriminant* is $D = b^2 - 4ac$. A *positive definite* binary quadratic form (abbreviated as PDBQF) is a binary quadratic form of negative discriminant satisfying $a > 0$. The *representation function* of f is the arithmetic function defined by

$$R_f(n) = |\{(x, y) \in \mathbb{Z}^2 \mid f(x, y) = n\}|.$$

A function of the form R_f for some PDBQF f is called a *quadratic representation function.* For any integers q and β, we let

$$\rho_{f,\beta}(q) = |\{(x,y) \in [q]^2 \mid f(x,y) \equiv \beta \bmod q\}|.$$

We are now ready to state this chapter's main result [10, Theorem 1.2].

Theorem 4.1. *Let $\Psi = (\psi_1, \ldots, \psi_{t+s}) : \mathbb{Z}^d \to \mathbb{Z}^{t+s}$ be a system of affine-linear forms of finite complexity. Suppose that the coefficients of the linear part $\dot{\Psi}$ are bounded*[1] *by some constant L. Let $K \subset [-N,N]^d$ be a convex body such that $\Psi(K) \subset [0,N]^{t+s}$. Let $f_{t+1}, \ldots, f_{t+s}$ be PDBQFs of discriminants $D_j < 0$ for $j = t+1, \ldots, t+s$. Then*

$$\sum_{n \in \mathbb{Z}^d \cap K} \prod_{i=1}^{t} \Lambda(\psi_i(n)) \prod_{j=t+1}^{t+s} R_{f_j}(\psi_j(n)) = \beta_\infty \prod_p \beta_p + o(N^d),$$

where

$$\beta_\infty = \mathrm{Vol}(K) \prod_{j=t+1}^{t+s} \frac{2\pi}{\sqrt{-D_j}}$$

and

$$\beta_p = \lim_{m \to +\infty} \mathbb{E}_{a \in (\mathbb{Z}/p^m\mathbb{Z})^d} \prod_{i=1}^{t} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(a)) \prod_{j=t+1}^{t+s} \frac{\rho_{f_j,\psi_j(a)}(p^m)}{p^m}.$$

As in Theorem 2.1, the error term is non effective and the implied decaying function depends on d, t, s, L and the discriminants. For each prime p, we call β_p the *local factor* modulo p. Like in the previous chapters, the quantities β_p are called *local factors.* The existence of the limit as m tends to infinity that defines it is proven in Proposition B.1; the convergence of the infinite product $\prod_p \beta_p$ is a consequence of Lemma B.3. The technical proofs of these facts are postponed to appendices in order not to disrupt the flow of the argument.

Two important special cases arise when $s = 0$ or $t = 0$, that is, when the functions featuring are either all equal to the von Mangoldt function, or all quadratic representation functions. Then one of the products is trivial.

- When $s = 0$, one immediately recovers the result of Green and Tao [45, Main Theo-

[1] One can check that the condition that the system has bounded size at scale N is equivalent to the boundedness of the linear part together with the condition on the image of K used in the previous chapter.

rem]. Indeed, for $m \geq 1$, we have

$$\mathbb{E}_{a\in(\mathbb{Z}/p^m\mathbb{Z})^d}\prod_{i=1}^{t}\Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(a)) = \mathbb{E}_{a\in(\mathbb{Z}/p\mathbb{Z})^d}\prod_{i=1}^{t}\Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(a))$$

so that

$$\beta_p = \mathbb{E}_{a\in(\mathbb{Z}/p\mathbb{Z})^d}\prod_{i=1}^{t}\Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(a)).$$

- When $t = 0$, Theorem 3.1 boils down to the formula of Matthiesen [68, Theorem 1.1].

Sometimes one can get an asymptotic even when the system has infinite complexity, but the asymptotic takes a completely different form then. For instance it is easy to see that

$$\sum_{n\leq N}\Lambda(n)R(n) \sim 8\sum_{\substack{p\leq N\\ p\equiv 1 \bmod 4}}\log p \sim 4N$$

by Fermat's theorem on sums of two squares and the prime number theorem in arithmetic progressions. We do not address such systems in this paper.

4.2 Special cases

Our first application concerns arithmetic progressions in the primes whose common difference is required to be a sum of two squares. It shows that the Green-Tao theorem (case $s = 0$ of Theorem 3.1) holds not only for linear systems, but also for some – admittedly very specific – polynomial systems. Here R and ρ (see Section 4.1) will implicitly refer to the form $f(x, y) = x^2 + y^2$ whose discriminant is -4.

Corollary 4.2. *Let $k \geq 1$ be an integer and let*

$$L = \{(a, b, c) \in \mathbb{R}^3 \mid 1 \leq a \leq a + (k-1)(b^2 + c^2) \leq N\}.$$

Let $\Psi = (\psi_0, \ldots, \psi_{k-1}) \in \mathbb{Z}[a, b, c]^k$ be the polynomial system defined by

$$\psi_i(a, b, c) = a + i(b^2 + c^2).$$

Then

$$\sum_{n\in\mathbb{Z}^3\cap L}\prod_{i=0}^{k-1}\Lambda(\psi_i(n)) = \beta_\infty\prod_p\beta_p + o(N^2), \tag{4.1}$$

where $\beta_\infty = \mathrm{Vol}(L)$ *and*

$$\beta_p = \mathbb{E}_{n\in(\mathbb{Z}/p\mathbb{Z})^3} \prod_{i=0}^{k-1} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(n)). \tag{4.2}$$

Proof of Corollary 4.2 assuming Theorem 3.1. We note that the left-hand side of equation (4.1) can be written as

$$\sum_{(a,d)\in\mathbb{Z}^2\cap K} \Lambda(a)\Lambda(a+d)\cdots\Lambda(a+(k-1)d)R(d), \tag{4.3}$$

where $K = \{(a,d) \in \mathbb{R}^2 \mid 1 \le a \le a+(k-1)d \le N\}$ is a convex body in $\mathbb{R}^2$. Applying Theorem 3.1 to this convex body and the system $(a,d) \mapsto (a, a+d, \ldots, a+(k-1)d, d)$, which is of finite complexity, we get

$$\sum_{n\in\mathbb{Z}^3\cap L} \prod_{i=0}^{k-1} \Lambda(\psi_i(n)) = \beta_\infty \prod_p \beta_p + o(N^2), \tag{4.4}$$

with $\beta_\infty = \pi\frac{N^2}{2(k-1)}$ and

$$\beta_p = \lim_{m\to\infty} \mathbb{E}_{(a,d)\in(\mathbb{Z}/p^m\mathbb{Z})^2} \frac{\rho_d(p^m)}{p^m}\left(\frac{p}{\phi(p)}\right)^k \prod_{i=0}^{k-1} 1_{(a+id,p)=1}.$$

It is easy to see that $\mathrm{Vol}(L) = \beta_\infty$. It remains to prove that the local factors are of the form (4.2). First, observe that

$$\mathbb{E}_{(a,d)\in(\mathbb{Z}/p^m\mathbb{Z})^2} \frac{\rho_d(p^m)}{p^m} \prod_{i=0}^{k-1} 1_{(a+id,p)=1} = \mathbb{E}_{(a,b,c)\in(\mathbb{Z}/p^m\mathbb{Z})^3} \prod_{i=0}^{k-1} 1_{(a+i(b^2+c^2),p)=1}. \tag{4.5}$$

Now let $a \mapsto \tilde{a}$ be the canonical map $\mathbb{Z}/p^m\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$. We notice that it is a p^{m-1}-to-1 map and that $(a+i(b^2+c^2),p) = 1$ if and only if $(\tilde{a}+i(\tilde{b}^2+\tilde{c}^2),p) = 1$. Hence

$$\mathbb{E}_{(a,b,c)\in(\mathbb{Z}/p^m\mathbb{Z})^3} \prod_{i=0}^{k-1} 1_{(a+i(b^2+c^2),p)=1} = \mathbb{E}_{(a,b,c)\in(\mathbb{Z}/p\mathbb{Z})^3} \prod_{i=0}^{k-1} 1_{(a+i(b^2+c^2),p)=1}$$

does not depend on m, and the local factors are of the desired form. □

Corollary 4.2 now appears as a special case of a posterior result of Tao and Ziegler [88, Theorem 1.4]. However, our result is more robust in the following sense. Although we do

not formally prove it here, an adaptation of our method can deal with a variant where L is replaced by

$$[1,N]\times\{(b,c)\in\mathbb{R}^2 \mid b^2+c^2\leq N\log^{-A}N\}\subset[1,N]\times[-\sqrt{N}\log^{-A/2}N,\sqrt{N}\log^{-A/2}N]^2$$

for any constant $A>0$, thus we could allow the step of the progression to be markedly smaller than the terms of the progression. Indeed, in the proof above, this change amounts to replacing $K\cap\mathbb{Z}^d$ with $[N]\times[N\log^{-A}N]$ in equation (4.3). To handle equation (4.3) then, we can proceed as in Chapter 3. In contrast, Tao and Ziegler's method cannot restrict b and c to such a small range.

Let us now compute explicitly the local factors β_p. Suppose first that $p\geq k$. We remark that

$$\beta_p=\left(\frac{p}{p-1}\right)^k\frac{1}{p}\sum_{a\in(\mathbb{Z}/p\mathbb{Z})^*}(1-\sum_{i=1}^{k-1}\mathbb{P}_{(b,c)\in(\mathbb{Z}/p\mathbb{Z})^2}(b^2+c^2\equiv-\bar{i}a \bmod p)),$$

where $\bar{i}$ is the inverse of i modulo p. Moreover, for any $a\in(\mathbb{Z}/p\mathbb{Z})^*$, setting $e(x)=\exp(2i\pi x)$ as customary, we have

$$\begin{aligned}
|\{(b,c)\in(\mathbb{Z}/p\mathbb{Z})^2 \mid b^2+c^2\equiv a \bmod p\}| &= \sum_{(b,c)\in(\mathbb{Z}/p\mathbb{Z})^2}\frac{1}{p}\sum_{h\in\mathbb{Z}/p\mathbb{Z}}e\left(\frac{h(b^2+c^2-a)}{p}\right)\\
&= \frac{1}{p}\left(\sum_{h\in(\mathbb{Z}/p\mathbb{Z})^*}e\left(-\frac{ha}{p}\right)\left(\sum_{b\in\mathbb{Z}/p\mathbb{Z}}e\left(\frac{hb^2}{p}\right)\right)^2+p^2\right)\\
&= \begin{cases} p-1 & \text{if } p\equiv 1 \bmod 4\\ p+1 & \text{if } p\equiv -1 \bmod 4\\ p & \text{if } p=2.\end{cases}
\end{aligned}$$

The last equality follows from the classical computation of Gauss sums (see [55, 3.38]). For $p\geq k$, this leads to

$$\beta_p=\begin{cases}\left(1+\frac{1}{p-1}\right)^k\left(1-\frac{k}{p}+2\frac{k-1}{p^2}-\frac{k-1}{p^3}\right) & \text{if } p\equiv 1 \bmod 4\\ \left(1+\frac{1}{p-1}\right)^k\left(1-\frac{k}{p}+\frac{k-1}{p^3}\right) & \text{if } p\equiv -1 \bmod 4.\end{cases}$$

It is easy to compute the local factors for $p \leq k$, namely

$$\beta_p = \begin{cases} \left(\frac{p}{p-1}\right)^k \frac{(p-1)(2p-1)}{p^3} & \text{if } p \equiv 1 \bmod 4 \\ \left(\frac{p}{p-1}\right)^k \frac{p-1}{p^3} & \text{if } p \equiv -1 \bmod 4 \\ 2^{k-2} & \text{if } p = 2. \end{cases}$$

We notice that β_p is nonzero for every p and that $\beta_p = 1 + O(p^{-2})$. It follows that $\prod_p \beta_p$ is a nonzero convergent product. We prove in Lemma B.3 that the product of the local factors is always convergent for systems of finite complexity.

Corollary 4.2 counts the number of *weighted* arithmetic progressions of primes up to N whose common difference is a sum of two squares, each such arithmetic progression being weighted by the number of representations of the common difference. It would be interesting to count these progressions without the weight, but it is not possible to derive such a count from Corollary 4.2.

In general, the only polynomial patterns we are able to deal with are the ones which can be converted into linear patterns using quadratic representation functions, as in the proof of Corollary 4.2. The ability to deal with arithmetic progressions whose common difference is a sum of two squares as if they were a linear pattern is reminiscent of a result of Green [37]: he proved that if a set $A \subset [N]$ does not contain any such progressions of length 3, then $|A| \ll N(\log\log N)^{-c}$ for some $c > 0$.

Theorem 3.1 can yield many further asymptotics for the number of solutions to equations in primes and sums of squares, some of which are not covered by Tao and Ziegler [88]. In particular, one can count asymptotically (with multiplicity) progressions in the set of sums of two squares whose common difference is a prime. Such an asymptotic is given by the sum

$$\sum_{1 \leq n \leq n+(k-1)d \leq N} \prod_{i=0}^{k-1} R(n+id)\Lambda(d),$$

where R is again the representation function of sums of two squares. The system of linear forms at hand is of finite complexity, so Theorem 3.1 applies.

We claim, but do not formally prove, that our method yields a result similar to Theorem 3.1 with the divisor function τ instead of the representation functions R_{f_i}. In fact, this result is easier to prove, since the treatment of the representation function of a binary quadratic form by Matthiesen [68] relies on her earlier paper on the divisor function [66].

Theorem 4.3. *Let $\Psi = (\psi_1, \dots, \psi_{t+s}) : \mathbb{Z}^d \to \mathbb{Z}^{t+s}$ be a system of affine-linear forms of finite complexity. Suppose that the coefficients of the linear part $\dot{\Psi}$ are bounded by L. Let $K \subset [-N, N]^d$ be a convex body such that $\Psi(K) \subset [0, N]^{t+s}$. Write $\Phi = (\psi_{t+1}, \dots, \psi_{t+s})$ and $\dot{\Phi}$ for the linear part. Then*

$$\sum_{n \in \mathbb{Z}^d \cap K} \prod_{i=1}^{t} \Lambda(\psi_i(n)) \prod_{j=t+1}^{t+s} \tau(\psi_j(n)) = (\log N)^s \beta_\infty \prod_p \beta_p + o_{d,t,s,L}(N^d \log^s N),$$

where

$$\beta_\infty = \mathrm{Vol}(K)$$

and

$$\beta_p = \left(\frac{p}{p-1}\right)^{t-s} \mathbb{E}_{a \in [p]^d} \prod_{i=1}^{t} 1_{(\psi_i(a),p)=1} \sum_{(k_1,\dots,k_s) \in \mathbb{N}^s} \alpha_{\Phi_{a,p}}(p^{k_1}, \dots, p^{k_s})$$

with $\Phi_{a,p} : b \mapsto \Phi(a) + p\dot{\Phi}(b)$ and α as in Definition A.1.

This theorem provides an asymptotic for the number of triples of nonnegative integers (a, b, c) such that $a, a+bc, a+2bc$ are primes. This is again a polynomial pattern of degree 2; in fact, τ can be viewed as the representation function of the bilinear form $(x, y) \mapsto xy$. We can obtain a result similar to Corollary 4.2. We let

$$L = \{(a, b, c) \in [1, +\infty[^3 |\ a + (k-1)bc \leq N\}.$$

This is not a convex body, but we have $\mathrm{Vol}(L) \sim |L \cap \mathbb{Z}^3| \sim N^2 \log N/(k-1)$. It is not difficult to deduce from Theorem 4.3 that

$$\sum_{(a,b,c) \in L \cap \mathbb{Z}^3} \prod_{i=0}^{t-1} \Lambda(a + ibc) = \mathrm{Vol}(L) \prod_p \beta_p + o(N^2 \log N)$$

with

$$\beta_p = \prod_{i=0}^{t-1} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(a + ibc).$$

Again, this result has the same shape as the Green-Tao theorem although the configuration involved is nonlinear.

We remark that the idea of mixing Λ and τ is quite old. Titchmarsh [91] considered

sums such as

$$\sum_{p \leq N} \tau(p+a)$$

or equivalently

$$\sum_{n \leq N} \Lambda(n)\tau(n+a)$$

for $a \in \mathbb{Z}$. Assuming the Riemann hypothesis, he proved that that

$$\sum_{n \leq N} \Lambda(n)\tau(n+a) = c_1(a)N \log N + O(N \log\log N)$$

for some explicit constant $c_1(a)$. The result was proven unconditionally by Linnik [63]. Fouvry [30] proved the refined asymptotic formula

$$\sum_{n \leq N} \Lambda(n)\tau(n+a) = c_1(a)N \log N + C_2(a)\mathrm{Li}(N) + O_A(N(\log N)^{-A})$$

for any $A > 0$. Notice that this problem does not belong to the scope of our method, because the involved linear system is of infinite complexity.

We also mention that Matthiesen, together with Browning [17], was able to generalise her result about quadratic forms to norm forms originating from a number field. This implies a generalisation of Theorem 3.1, but we refrain, for the sake of simplicity, from inspecting this general case.

4.3 Overview of the general strategy

We now turn to a proof of the main theorem, Theorem 3.1. The proof follows the usual Green-Tao method, as sketched in Chapter 2. In Section 4.4.2, we perform the W-trick to mitigate the preference of the von Mangoldt function and the representation function for some residue classes. Because of the notably different behaviours of these functions with respect to arithmetic progressions, this is a delicate matter.

On the one hand, the uniformity property of the von Mangoldt function holds for congruence classes to moduli q of size $O(\log^{O(1)} N)$, which sets a bound on the size of a tolerable W. On the other hand, to "uniformise" the representation function of a quadratic form, we need to pass to congruence classes $qn + b$ where q is divisible by large prime powers and b is nonzero modulo any of these prime powers. This generates a conflict that

we carefully resolve.

Assuming some convergence properties of the local factors, which we prove in Appendix B, the implementation of the W-trick reduces the main theorem to Theorem 4.7, which is the statement that a multilinear average

$$\mathbb{E}_{n\in\mathbb{Z}^d\cap K}(F_0(\psi_0(n))-1)\prod_{i=1}^{t}F_i(\psi_i(n))$$

is asymptotically $o(1)$. Thanks to a generalised von Neumann theorem, it suffices to ensure that F_0-1 has small Gowers uniformity norm and that all the functions F_i and F_0-1 are bounded by a common enveloping sieve or pseudorandom majorant. This is another novelty of our result: while individual pseudorandom majorants for Λ and for R_f were known before [10], we needed to construct a common one that works for Λ and R_f simultaneously.

We remark that although we want to prove a result concerning quadratic and not linear patterns in the primes, we do not need the pseudorandom majorant to satisfy the polynomial forms condition introduced in [88]. This is because the polynomial character of our configurations is encapsulated in the representation functions of the quadratic forms.

4.4 Proof of Theorem 3.1

We fix some arbitrarily large integer N, so that our asymptotic results are valid in the limit where N tends to infinity. We use the notation $[N]$ for the set of the first N integers. Many of the parameters introduced in the sequel implicitly depend on N (such as the convex body K, the map $p\mapsto\iota(p)$, the numbers $w,W,\overline{W}$, the set X_0...).

4.4.1 Elimination of a negligible set

We start our proof by taking care of a technicality. We would like to eliminate slightly awkward integers from the support of the von Mangoldt and the representation functions. In fact, as already noticed in Chapter 2, it will turn out handy to exclude prime powers and small primes from the support of Λ, so we introduce $\Lambda'=1_{\mathcal{P}\setminus[N^{2\gamma}]}\log$, for some constant $\gamma\in(0,1/2)$ to be fixed later. It coincides with Λ on the bulk of its support up to N, namely large primes.

Similarly, there is a fairly sparse subset $X_0\subset[N]$, depending on some constants $C_1>0$

and $\gamma > 0$, on which the divisor function and the representation function behave abnormally, so that our process of majorising by a pseudorandom measure (carried out in Section 4.5) fails there. We recall the following definition originating from [66] and taken up in [68].

Definition 4.1. Let $\gamma = 2^{-k}$ for some $k \in \mathbb{N}$ to be decided later, and let $C_1 > 1$. We define $X_0 = X_0(\gamma, C_1, N)$ to be the set containing 0 and the set of positive integers $n \leq N$ satisfying one of the following.

1. n is excessively "rough", i.e. divisible by some large prime power $p^a > \log^{C_1} N$ with $a \geq 2$, or
2. n is excessively smooth in the sense that if $n = \prod_p p^{a_p}$ then
$$\prod_{p \leq N^{(1/\log\log N)^3}} p^{a_p} \geq N^{\gamma/\log\log N}$$
or
3. n has a large square divisor $m^2 \mid n$ that satisfies $m > N^\gamma$.

We will settle on values for γ and k later. The constant γ is the same as that introduced at the end of Chapter 2, where it had to be small enough for Proposition 2.13 to hold. We will find in the current chapter a further smallness condition it needs to satisfy.

The following lemma, which is Lemma 3.2 from [68], itself a synthesis of Lemmas 3.2 and 3.3 from [66], shows how negligible this set is.

Lemma 4.4. *For Ψ and K as in Theorem 3.1, we have*

$$\mathbb{E}_{n \in K \cap \mathbb{Z}^d} \sum_{i=t+1}^{t+s} 1_{\psi_i(n) \in X_0} \ll_{\gamma,d,s} \log^{-C_1/2} N.$$

This enables us to state the next lemma, which allows us to ignore X_0 altogether. For any PDBQF f, we use to the notation $\overline{R_f}(n)$ to denote $1_{n \notin X_0} R_f(n)$.

Lemma 4.5. *If the parameter C_1 in Definition 4.1 is large enough, and for any choice of the constant $\gamma \in (0, 1/2)$, Theorem 3.1 holds if and only if, under the same conditions, we have*

$$\sum_{n \in K \cap \mathbb{Z}^d} \prod_{i=1}^{t} \Lambda'(\psi_i(n)) \prod_{j=t+1}^{t+s} \overline{R}_{f_j}(\psi_j(n)) = \beta_\infty \prod_p \beta_p + o(N^d). \tag{4.6}$$

Proof. We first show that

$$\sum_{\substack{n\in K\cap\mathbb{Z}^d\\ \exists j\in[\![t+1\,;\,t+s]\!]:\ \psi_j(n)\in X_0}} \prod_{i=1}^{t}\Lambda(\psi_i(n))\prod_{j=t+1}^{t+s} R_{f_j}(\psi_j(n)) = o(N^d),$$

where we introduced the notation $[\![t+1\,;\,t+s]\!] = \{t+1,\dots,t+s\}$. We get rid of the von Mangoldt factors by bounding their product by $\log^t N$. Then we use the Cauchy-Schwarz inequality followed by the triangle inequality, which implies that

$$\left(\sum_{\substack{n\in K\cap\mathbb{Z}^d\\ \exists j\in[\![t+1\,;\,t+s]\!]:\ \psi_j(n)\in X_0}} \prod_{j=t+1}^{t+s} R_{f_j}(\psi_j(n))\right)^2 \leq \sum_{n\in K\cap\mathbb{Z}^d}\left(\prod_{j=t+1}^{t+s} R_{f_j}(\psi_j(n))\right)^2 \sum_{n\in K\cap\mathbb{Z}^d}\sum_{j=t+1}^{t+s} 1_{\psi_j(n)\in X_0}.$$

Finally, we use Lemma 3.1 of [68] which ensures that the first factor is $O(N^d\log^{O_s(1)} N)$, while the second one is $O(N^d\log^{-C_1/2} N)$ according to Lemma 4.4, so that taking C_1 larger than $2(t+O_s(1))$, we have the result.

To replace Λ by Λ', we remark that for each $i\in[t]$, the number of $n\in K\cap\mathbb{Z}^d$ such that $\psi_i(n)\leq N^{2\gamma}$ is $O(N^{d-1+2\gamma})$, while the number of $n\in K\cap\mathbb{Z}^d$ such that $\psi_i(n)$ is a prime power and not a prime is $O(N^{d-1}\log N\sqrt{N})$. Using Cauchy-Schwarz or even pointwise bounds such as the divisor bound $R_{f_j}(n)\ll\tau(n)\ll_\epsilon N^\epsilon$, we conclude the proof of Lemma 2.2. □

From now on we will drop the bar, so that R_f coincides with the actual representation function of f on $[N]\setminus X_0$ and is 0 on X_0.

4.4.2 Implementation of the W-trick

Recall that when performing the W-trick in Chapter 2, we had allowed some freedom in the precise choice of the modulus $\widetilde{W}$, which will come in handy. As already mentioned, the representation functions of quadratic forms require large prime powers to be incorporated into $\widetilde{W}$. We therefore introduce a new modulus

$$\overline{W} = \prod_{p\leq w} p^{\iota(p)},$$

where $\iota(p)$ is defined by

$$p^{\iota(p)-1} < \log^{C_1+1} N \leq p^{\iota(p)} \tag{4.7}$$

for some C_1 large enough as in Lemma 4.5. We observe that

$$\overline{W} \leq \prod_{p \leq w} p \log^{C_1+1} N \ll \exp((C_1+2) w \log\log N)),$$

which is less than any power of N. In particular, we can ensure that $\overline{W} < N^{\gamma} - 1$ by choosing N large enough. Notice that for N large enough and $p \leq w(N) = \log\log\log N$, we always have $\iota(p) \geq 2$.

Of course, $\overline{W}$ is larger than any power of $\log N$, hence outside of the range of the Siegel-Walfisz theorem and as a result, we cannot claim that $\|\Lambda'_{\overline{W},b} - 1\|_{U^t} = o(1)$, not even for $t = 1$. We will make do without this bound.

Following Matthiesen [68, Definition 7.2], we define

$$r'_{f,b}(m) = \frac{\sqrt{-D}}{2\pi} \frac{\overline{W}}{\rho_{f,b}(\overline{W})} R_f(\overline{W}m + b), \tag{4.8}$$

for any b such that $\rho_{f,b}(\overline{W}) > 0$, and if $\rho_{f,b}(\overline{W}) = 0$, we define $r'_{f,b}(m)$ to be 0. By construction, $R_f(n)$ equals 0 in the case where $n \in X_0$, in particular in the case where $n \equiv 0 \bmod p^{\iota(p)}$ with $p \leq w(N)$. Hence, $r'_{f,b} = 0$ if $b \equiv 0 \bmod p^{\iota(p)}$. Moreover (see [68, Definition 7.2]) for $b \not\equiv 0 \bmod p^{\iota(p)}$ and any $p \leq w(N)$ satisfying $\rho_{f,b}(\overline{W}) \neq 0$, we have

$$\mathbb{E}_{n \leq M} r'_{f,b}(n) = 1 + O(\overline{W}^3 M^{-1/2}).$$

This average in arithmetic progressions relies on elementary convex geometry and is valid uniformly in the modulus, in sharp contrast with the analogous result for primes, the Siegel-Walfisz theorem (1.3).

We now decompose the left-hand side of (4.6) into sums over congruence classes as in Proposition 2.7. Letting

$$F(n) = \prod_{i=1}^{t} \Lambda'(\psi_i(n)) \prod_{j=t+1}^{t+s} R_{f_j}(\psi_j(n)),$$

we can write the left-hand side of (4.6) as

$$\sum_{n\in\mathbb{Z}^d\cap K} F(n) = \sum_{a\in[\overline{W}]^d} \sum_{n\in\mathbb{Z}^d\cap K_a} F(\overline{W}n+a), \tag{4.9}$$

where

$$K_a = \{x \in \mathbb{R}^d \mid \overline{W}x + a \in K\}$$

is again a convex body. Moreover, for $j \in [t+s]$, we can write $\psi_j(\widetilde{W}n+a) = \widetilde{W}\tilde{\psi}_j(n)+c_j(a)$ where $c_j(a) \in [\overline{W}]$ and $\tilde{\psi}_j$ is an affine-linear form differing from ψ_j only in the constant term. We remark that if $\psi_i(a)$ is not coprime to W for $i \in [t]$, or if $\rho_{f_j,\psi_j(a)}(\overline{W}) = 0$ or $\psi_j(a) \equiv 0 \bmod p^{\iota(p)}$ for some $j \in [\![t+1\,;\,t+s]\!]$ and some prime $p \le w(N)$, then for each $n \in K_a \cap \mathbb{Z}^d$ we have $F(\overline{W}n+a) = 0$ (even if $(\psi_i(a), W) > 1$, the integer $\psi_i(a)$ could still be a prime $p \le w(N) < N^\gamma$, but given that primes smaller than N^γ are not in the support of Λ', we still have $F(\overline{W}n+a) = 0$). Thus the residues a which make a nonzero contribution to the right-hand side of (4.9) are all mapped by Ψ to tuples $(b_1, \dots, b_{t+s})$ belonging to the following set.

Definition 4.2. We denote by $B_{t,s}$ the set of residues $b \in [\overline{W}]^{t+s}$ such that

1. for any $i \in [t]$, $(b_i, W) = 1$;
2. for any $j \in [\![t+1\,;\,t+s]\!]$ and any prime $p \le w(N)$, we have $b_j \not\equiv 0 \bmod p^{\iota(p)}$;
3. for any $j \in [\![t+1\,;\,t+s]\!]$, b_j is representable by f_j modulo $\overline{W}$, that is, $\rho_{f_j,b_j}(\overline{W}) > 0$.

Moreover, for an affine-linear system $\Psi : \mathbb{Z}^d \to \mathbb{Z}^{t+s}$, we define A_Ψ to be the set of all $a \in [W]^d$ such that $(c_i(a))_{i\in[t+s]} \in B_{t,s}$. We recall that $c_i(a)$ is the reduction modulo $\overline{W}$ in $[\overline{W}]$ of $\psi_i(a)$; we also denote by $c(a)$ the vector $(c_i(a))_{i\in[t+s]}$. We usually drop the subscripts on $B_{t,s}$ and A_Ψ when there is no ambiguity.

Now we rewrite equation (4.9) as

$$\begin{aligned}\sum_{n\in\mathbb{Z}^d\cap K} F(n) &= \sum_{a\in A_\Psi} \left(\frac{W}{\varphi(W)}\right)^t \prod_{j=t+1}^{t+s} \frac{2\pi}{\sqrt{-D_j}} \frac{\rho_{f_j,\psi_j(a_j)}(\overline{W})}{\overline{W}} \sum_{n\in K_a\cap\mathbb{Z}^d} F'_a(n) \\ &= \prod_{j=t+1}^{t+s} \frac{2\pi}{\sqrt{-D_j}} \sum_{a\in[\overline{W}]^d} Q(a) \sum_{n\in K_a\cap\mathbb{Z}^d} F'_a(n),\end{aligned}$$

where

$$Q(a) = \prod_{i=1}^{t} \Lambda_{\mathbb{Z}/W\mathbb{Z}}(\psi_i(a)) \prod_{j=t+1}^{t+s} \frac{\rho_{f_j,\psi_j(a)}(\overline{W})}{\overline{W}} 1_{\forall p \leq w, \psi_j(a) \not\equiv 0 \bmod p^{\iota(p)}}, \tag{4.10}$$

and

$$F'_a(n) = \prod_{i=1}^{t} \Lambda'_{\overline{W},c_i(a)}(\tilde{\psi}_i(n)) \prod_{j=t+1}^{t+s} r'_{f_j,c_j(a)}(\tilde{\psi}_j(n)).$$

In equation (4.10), we have used the notation $\Lambda_{\mathbb{Z}/W\mathbb{Z}}$ for the local von Mangoldt function introduced in equation (1.2).

Furthermore, we use the identity

$$F'_a(n) = \prod_{i=1}^{t} \Lambda'_{\overline{W},c_i(a)}(\tilde{\psi}_i(n)) + \sum_{j=1+t}^{t+s} (r'_{f_j,c_j(a)}(\tilde{\psi}_j(n)) - 1) F'_{a,j}(n),$$

where

$$F'_{a,j}(n) = \prod_{k<j} r'_{f_k,c_k(a)}(\tilde{\psi}_k(n)) \prod_{i=1}^{t} \Lambda'_{\overline{W},c_i(a)}(\tilde{\psi}_i(n)).$$

Thus, equation (4.9) yields

$$\left(\prod_{j=t+1}^{t+s} \frac{2\pi}{\sqrt{-D_j}} \right)^{-1} \sum_{n \in \mathbb{Z}^d \cap K} F(n) = T_1 + T_2, \tag{4.11}$$

where

$$T_1 = \sum_{a \in [\overline{W}]^d} Q(a) \sum_{n \in \mathbb{Z}^d \cap K_a} \prod_{i=1}^{t} \Lambda'_{\overline{W},c_i(a)}(\tilde{\psi}_i(n)) \tag{4.12}$$

and

$$T_2 = \sum_{j=t+1}^{t+s} \sum_{a \in A} Q(a) \sum_{n \in \mathbb{Z}^d \cap K_a} (r'_{f_j,c_j(a)}(\tilde{\psi}_j(n)) - 1) F'_{a,j}(n). \tag{4.13}$$

Here, the first term is expected to be the main term, of the order of magnitude of $\mathrm{Vol}(K)$. The second one, which involves the difference of a $\overline{W}$-tricked representation function to its average 1, is expected to be negligible, that is, $o(N^d)$.

4.4.3 Analysis of the main term

To deal with the main term (4.12), we would ideally like to claim that the inner sum satisfies

$$\sum_{n\in\mathbb{Z}^d\cap K_a}\prod_{i=1}^{t}\Lambda'_{\overline{W},c_i(a)}(\tilde{\psi}_i(n)) = \mathrm{Vol}(K_a) + o((N/\overline{W})^d).$$

Unfortunately, $\overline{W}$ is too large for that, even for the case where $t = 1$.

If we were able to lower the prime powers $p^{\iota(p)} \approx \log^{C_1} N$ involved in $\overline{W}$ to smaller prime powers $p^{\eta(p)} \approx \log\log N$, the resulting $\widetilde{W}$ would be small enough for Siegel-Walfisz (and more generally Proposition 2.7) to apply. Let us then define $\eta(p)$ by

$$p^{\eta(p)-1} < \log\log N \leq p^{\eta(p)} \tag{4.14}$$

and $\widetilde{W} = \prod_{p\leq w} p^{\eta(p)} \leq \prod_{p\leq w} p\log\log N = \log^{o(1)} N$.

The reader may wonder at this point why we performed the $\overline{W}$ trick at all if we really want to deal with congruence classes modulo $\widetilde{W}$. The reason for this is that Lemma 4.5 would not hold if X_0 contained all integers smaller than N that have a prime power factor larger than $\log\log N$: this is not a sufficiently sparse set, given the possibly large values of R_f and Λ. Thus, performing the $\widetilde{W}$-trick, we could not force the residues to satisfy $c_j(a) \not\equiv 0 \bmod p^{\eta(p)}$, whereas the $\overline{W}$-trick allowed us to force $c_j(a) \not\equiv 0 \bmod p^{\iota(p)}$. Imposing such a nonzero congruence will prove crucial to ensuring that $r'_{f_j,c_j(a)}$ is dominated by a pseudorandom majorant, and thus to establishing that the term T_2 is negligible.

To reduce the size of the prime powers, we shall rely on the powerful lift-invariance property of Matthiesen [68, Lemma 6.3].

Lemma 4.6. *Let f be a PDBQF of discriminant D. Let p_0 be a prime and $\alpha \geq v_{p_0}(D)$ be an integer. Suppose that $b \not\equiv 0 \bmod p_0^\alpha$. Then for all $\beta \geq \alpha$ and $c \equiv b \bmod p_0^\alpha$, we have*

$$\rho_{f,b}(p_0^\alpha)p_0^{-\alpha} = \rho_{f,c}(p_0^\beta)p_0^{-\beta}.$$

We decompose the residue set $[\overline{W}]^d$ into X_1 and X_2, where

$$X_1 = \{a \in [\overline{W}]^d \mid \forall j \in [\![t+1\,;\, t+s]\!] \quad \forall p \leq w(N) \qquad \psi_j(a) \not\equiv 0 \bmod p^{\eta(p)}\}$$

and X_2 is the complement of X_1 in $[\overline{W}]^d$. We also introduce

$$Y_1 = \{a \in [\widetilde{W}]^d \mid \forall j \in [\![t+1\,;\, t+s]\!] \quad \forall p \leq w(N) \qquad \psi_j(a) \not\equiv 0 \bmod p^{\eta(p)}\}.$$

First we remark that, for $a \in X_1$, $Q(a)$ depends only on the reduction $\tilde{a} \in Y_1$ of a. Indeed, writing

$$\widetilde{Q}(a) = \prod_{i=1}^{t} \Lambda_W(\psi_i(a)) \prod_{j=t+1}^{t+s} \frac{\rho_{f_j,\psi_j(a)}(\widetilde{W})}{\widetilde{W}} 1_{\forall p \le w, \psi_j(a) \not\equiv 0 \bmod p^{\eta(p)}},$$

we have $\widetilde{Q}(\tilde{a}) = Q(a)$. This shows that

$$\begin{aligned}\sum_{a \in X_1} Q(a) \sum_{n \in \mathbb{Z}^d \cap K_a} \prod_{i=1}^{t} \Lambda'_{\overline{W},c_i(a)}(\tilde{\psi}_i(n)) &= \sum_{a \in Y_1} Q(a) \sum_{\substack{b \in [\overline{W}]^d \\ b \equiv a \bmod \widetilde{W}}} \sum_{n \in \mathbb{Z}^d \cap K_b} \prod_{i=1}^{t} \Lambda'_{\overline{W},c_i(b)}(\tilde{\psi}_i(n)) \\ &= \sum_{a \in Y_1} Q(a) \sum_{n \in \mathbb{Z}^d \cap K_a} \prod_{i=1}^{t} \Lambda'_{\widetilde{W},c_i(a)}(\tilde{\psi}_i(n)). \qquad (4.15)\end{aligned}$$

We admit a slight abuse of notation: in the last term, $\tilde{\psi}_i$ may be different from the other occurrences of $\tilde{\psi}_i$ (differing at most in the constant term) and $c_i(a) \equiv \psi_i(a) \bmod \widetilde{W}$ lies in $[\widetilde{W}]$. They satisfy $\psi_i(\widetilde{W}n + a) = \widetilde{W}\tilde{\psi}_i(n) + c_i(a)$. Now we can apply Proposition 2.7 to the inner sum of (4.15). Thus for any $a \in Y_1$ that has a nonzero contribution, in particular, satisfying $(c_i(a), W) = 1$ for all $i \in [t]$, we have, uniformly in a, the relation

$$\sum_{n \in \mathbb{Z}^d \cap K_a} \prod_{i=1}^{t} \Lambda'_{\widetilde{W},c_i(a)}(\tilde{\psi}_i(n)) = \mathrm{Vol}(K_a) + o((N/\widetilde{W})^d).$$

Inserting this formula in (4.15) yields

$$\sum_{a \in X_1} Q(a) \sum_{n \in \mathbb{Z}^d \cap K_a} \prod_{i=1}^{t} \Lambda'_{\overline{W},c_i(a)}(\tilde{\psi}_i(n)) = (\mathrm{Vol}(K) + o(N^d))\mathbb{E}_{a \in [\widetilde{W}]^d} Q(a) 1_{a \in Y_1}.$$

We exploit multiplicativity to write

$$\mathbb{E}_{a \in [\widetilde{W}]^d} Q(a) 1_{a \in Y_1} = \prod_{p \le w} \mathbb{E}_{a \in (\mathbb{Z}/p^{\eta(p)}\mathbb{Z})^d} \prod_{i=1}^{t} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(a)) \prod_{j=t+1}^{t+s} \frac{\rho_{f_j,\psi_j(a)}(p^{\eta(p)})}{p^{\eta(p)}} 1_{\psi_j(a) \not\equiv 0 \bmod p^{\eta(p)}},$$

and invoke results from Appendix B. Indeed, setting $m = \eta(p)$ in Lemma B.2, we find that

$$\mathbb{E}_{a \in (\mathbb{Z}/p^{\eta(p)}\mathbb{Z})^d} \prod_{i=1}^{t} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(a)) \prod_{j=t+1}^{t+s} \frac{\rho_{f_j,\psi_j(a)}(p^{\eta(p)})}{p^{\eta(p)}} 1_{\psi_j(a) \not\equiv 0 \bmod p^{\eta(p)}} = \beta_p + O((\log\log N)^{-1/3}).$$

Using Lemma B.3, we conclude that

$$\mathbb{E}_{a\in[\widetilde{W}]^d} Q(a) 1_{a\in Y_1} = \prod_p \beta_p + o(1),$$

and finally we can write

$$\sum_{a\in X_1} Q(a) \sum_{n\in\mathbb{Z}^d\cap K_a} \prod_{i=1}^t \Lambda'_{\overline{W},c_i(a)}(\tilde{\psi}_i(n)) = \beta_\infty \prod_p \beta_p + o(N^d).$$

4.4.4 The sum over X_2

We now turn to the set of bad residues X_2. We need to show that

$$\sum_{a\in X_2} Q(a) \sum_{n\in\mathbb{Z}^d\cap K_a} \prod_{i=1}^t \Lambda'_{\overline{W},c_i(a)} = o(N^d).$$

In the absence of an asymptotic for the inner sum, we shall be content with an upper bound. To that aim, we use the majorant of the von Mangoldt function described in Chapter 2, and its uniformity property given by Proposition 2.13. Indeed, we have $\mathrm{rad}(\overline{W}) = W = O(\log N)$ and the exceptional primes for the system of linear forms $\widetilde{\Psi}$ are bounded, so these propositions apply. In particular, for any $a \in X_2$, we have

$$\sum_{n\in\mathbb{Z}^d\cap K_a} \prod_{i=1}^t \Lambda'_{\overline{W},c_i(a)}(\tilde{\psi}_i(n)) \ll \sum_{n\in\mathbb{Z}^d\cap K_a} \prod_{i=1}^t \nu_{\mathrm{GT},\overline{W},c_i(a)}(\tilde{\psi}_i(n)) = \mathrm{Vol}(K_a) + o((N/\overline{W})^d).$$

From this, we infer that

$$\sum_{a\in X_2} Q(a) \sum_{n\in\mathbb{Z}^d\cap K_a} \prod_{i=1}^t \Lambda'_{\overline{W},c_i(a)}(\tilde{\psi}_i(n)) \ll N^d \mathbb{E}_{a\in[\overline{W}]^d} Q(a) 1_{a\in X_2}.$$

We use the triangle inequality to bound the expectation in the right-hand side of the above equation by

$$\sum_{\substack{p\leq w \\ j\in[\![t+1\,;t+s]\!]}} \mathbb{E}_{a\in[\overline{W}]^d} 1_{p^{\eta(p)}|\psi_j(a)} Q(a),$$

which, by multiplicativity, can be rewritten as

$$\sum_{\substack{q\leq w\\ j\in[\![t+1\,;t+s]\!]}} \mathbb{E}_{a\in(\mathbb{Z}/q^{\iota(q)}\mathbb{Z})^d} 1_{q^{\eta(q)}|\psi_j(a)} Q_q(a) \prod_{p\leq w, q\neq p} \mathbb{E}_{a\in(\mathbb{Z}/p^{\iota(p)}\mathbb{Z})^d} Q_p(a).$$

Here, as the reader may be able guess, we have written

$$Q_p(a) = \prod_{i=1}^{t} \Lambda_p(\psi_i(a)) \prod_{j=t+1}^{t+s} \frac{\rho_{f_j,\psi_j(a_j)}(p^{\iota(p)})}{p^{\iota(p)}} 1_{\psi_j(a)\not\equiv 0 \bmod p^{\iota(p)}},$$

for any prime p, so that $Q(a) = \prod_{p\leq w(N)} Q_p(a)$. Again, we invoke Appendix B. Lemmas B.2 and B.3 imply that

$$\prod_{p\leq w, q\neq p} \mathbb{E}_{a\in(\mathbb{Z}/p^{\iota(p)}\mathbb{Z})^d} Q_p(a) = O(1),$$

while the proof of Proposition B.1 shows that

$$\mathbb{E}_{a\in(\mathbb{Z}/q^{\iota(q)}\mathbb{Z})^d} 1_{q^{\eta(q)}|\psi_j(a)} Q_q(a) = O((\log\log N)^{-1/3}).$$

Because $w(N) = \log\log\log N$ is so small, we obtain the desired bound

$$\sum_{a\in X_2} Q(a) \sum_{n\in\mathbb{Z}^d\cap K_a} \prod_{i=1}^{t} \Lambda'_{\overline{W},c_i(a)}(\tilde{\psi}_i(n)) = o(N^d).$$

4.4.5 Reduction of the main theorem

Given the above discussion, the main theorem (Theorem 3.1) boils down to proving that the term T_2 defined in equation (4.13) is $o(N^d)$. This is a consequence of the next proposition.

Theorem 4.7. *Let d, t and s be nonnegative integers, and let $f_0, f_{t+1}, \ldots, f_{t+s}$ be PDBQF. Let $N' = N/\overline{W}$, and $\Phi = (\phi_0, \ldots, \phi_{t+s})$ be a system of affine-linear forms $\mathbb{Z}^d \to \mathbb{Z}^{t+s+1}$ of finite complexity whose linear coefficients are bounded by a constant. Let $L \subset [0, N']^d$ be a convex set such that $\Phi(L) \subset [1, N']^{t+s+1}$. Then for any $b \in B_{t,s+1}$, we have*

$$\sum_{n\in\mathbb{Z}^d\cap L} (r'_{f_0,b_0}(\phi_0(n)) - 1) \prod_{i\in[t]} \Lambda'_{\overline{W},b_i}(\phi_i(n)) \prod_{j=t+1}^{t+s} r'_{f_j,b_j}(\phi_j(n)) = o(N'^d).$$

The set $B = B_{t,s+1}$ was introduced in Definition 4.2. Notice the slight change of

notation with respect to the original definition, due to the fact that our quadratic forms are now labelled $f_0, f_{t+1}, f_{t+2}, \ldots, f_{t+s}$.

We prove this theorem in the next section.

4.5 Majorant and uniformity of quadratic representation functions

Here and in the remainder of the chapter, $N' = N/\overline{W}$. To prove Theorem 4.7, we have to show that the average along a linear system of a product is $o(N'^d)$, knowing that one of the factors has average $o(N'^d)$. Because of Theorem 2.8, this follows if we can prove two things:

- the uniformity estimate $\|r'_{f,b} - 1\|_{U^k[N']} = o(1)$ for any k;
- that the representation functions of quadratic forms and the von Mangoldt function are dominated by a common pseudorandom majorant.

For the first item, we can use wholesale the following result of Matthiesen, proven in [68, Sections 14-18].

Proposition 4.8. *Let f be a PDBQF, and let $b \in [\overline{W}]$ be representable by f modulo $\overline{W}$ and not divisible by any $p^{\iota(p)}$ for $p \leq w(N)$. Then the tricked representation function of f defined by* (4.8) *satisfies, for all $k \in \mathbb{N}$, the estimate*

$$\|r'_{f,b} - 1\|_{U^k[N']} = o(1).$$

The discussion of the second item will occupy the rest of this section. We use a pseudorandom majorant from Matthiesen's work. For this we need to recall some notation and facts from [68]. Given a set $\mathcal{A}$ of primes, $\langle \mathcal{A} \rangle$ stands for the set of integers whose prime factors are all in $\mathcal{A}$. Let $\tau_{\mathcal{A}}(n) = \sum_{d \in \langle \mathcal{A} \rangle} 1_{d|n}$.

Proposition 4.9. *For any integer $D \equiv 0, 1 \bmod 4$, there exists a set of primes $\mathcal{P}_D$ of density 1/2, which is a union of congruence classes modulo D, such that putting $\mathcal{P}_D^* = \mathcal{P}_D \cup \{p \in \mathcal{P} \colon p \mid D\}$ and $\mathcal{Q}_D = \mathcal{P} \setminus \mathcal{P}_D^*$, we have, for any PDBQF f of discriminant D, the bound*

$$R_f(n) \ll_D \tau_D(n) \sum_{\substack{m \in \langle \mathcal{Q}_D \rangle \\ m^2 | n}} 1_{\langle \mathcal{P}_D^* \rangle}(n/m^2).$$

To understand this result, which is the starting point of the construction of the pseudo-random majorant in [68], we recall that the number of representations of any odd number n as a sum of two squares is $4\sum_{d|n}\chi(d)$, where χ is the only nontrivial character modulo 4. By multiplicativity, this is easily seen to equal $4\tau_{\mathcal{A}}(n)\prod_{p\equiv 3 \bmod 4} 1_{v_p(n)\equiv 0 \bmod 2}$, with $\mathcal{A}$ being the set of primes congruent to 1 modulo 4, from which we derive a majorant of the desired form. This works similarly for other quadratic forms.

Thus, to majorise the function R_f it will be enough to majorise the functions τ_D and $1_{\langle \mathcal{P}_D^* \rangle}$. The heuristic to bound τ_D (or rather $\tau_D/\sqrt{\log N}$) is as follows (see [66, Lemma 4.1]). We would like to truncate the divisor sum defining it at N^γ (possibly with a smooth cut-off), just as was done earlier for the von Mangoldt function. The function defined by this truncated divisor sum is called τ_γ. Unfortunately, it turns out that the inequality $\tau \leq C\tau_\gamma$ is not entirely true, at least not true with the same constant C throughout the first N integers. Nevertheless, a heuristic of Erdős [27] says that an integer is either excessively rough or excessively smooth or has a cluster of many prime factors close together. We have excluded the first two possibilities when we took out the set X_0, so it remains to majorise $\tau(n)$ in the third case. Then the bound depends on the position of this cluster of primes and on its density. For more details on the majorant of the divisor function see [66].

To bound $1_{\langle \mathcal{P}_D^* \rangle}$ (or rather $1_{\langle \mathcal{P}_D^* \rangle}\sqrt{\log N}$), that is, the indicator function of the integers without any prime factor belonging to $\mathcal{Q}_D$, we use a sieving-type majorant, that is, a majorant similar to the one introduced above for the von Mangoldt function. Indeed, integers without any prime factor in $\mathcal{Q}_D$ are similar to prime numbers (integers without any nontrivial prime factors at all).

To formalise this heuristic, let us introduce the following definition. Recall that the constant $\gamma = 2^{-k}$ was introduced in Definition 4.1, and its exact value (or the value of k) is yet to be chosen.

Definition 4.3. Let $\xi = \gamma/2 = 2^{-k-1}$. We define sets $U(i,s)$ for integers i, s as follows. Let $\log_2$ be the base 2 logarithm. For $i = \log_2(2/\xi) - 2 = k$, we let $U(i, 2/\xi)$ be $\{1\}$ and otherwise $U(i, 2/\xi) = \emptyset$. If $s > 2/\xi$ and $i \geq \log_2 -s$, write $U(i,s)$ for the set of all products of $m_0(i,s) = \lceil \xi s(i + 3 - \log_2 s)/100 \rceil$ distinct primes from the interval $[N^{2^{-i-1}}, N^{2^{-i}}]$.

Let us fix an integer $D \equiv 0, 1 \bmod 4$. We now describe a majorant for the W-tricked representation function of a PDBQF of discriminant D, which was designed by Matthiesen [68]. We again need the smooth function χ (this should not be mistaken with a character,

as there are no more characters in the sequel) introduced for the majorant of the von Mangoldt function. We use the function

$$r_{D,\gamma}(n) = \beta'_{D,\gamma}(n)\nu'_{D,\gamma}(n), \tag{4.16}$$

where

$$\nu'_{D,\gamma} = \sum_{s=2/\xi}^{\lfloor(\log\log N)^3\rfloor} \sum_{i=\log_2 s-2}^{\lfloor 6\log\log\log N\rfloor} \sum_{u\in U(i,s)} 2^s 1_{u|n}\tau'_{D,\gamma}(n),$$

with

$$\tau'_{D,\gamma}(n) = \sum_{\substack{d\in\langle\mathcal{P}_D\rangle \\ p|d\Rightarrow p>w(N)}} 1_{d|n}\chi\left(\frac{\log d}{\log N^\gamma}\right),$$

and

$$\beta'_{D,\gamma}(n) = \sum_{\substack{m\in\langle\mathcal{Q}_D\rangle \\ p|m\Rightarrow p>w(N) \\ m<N^\gamma}} \left(\sum_{\substack{e\in\langle\mathcal{Q}_D\rangle \\ p|e\Rightarrow p>w(N)}} 1_{m^2e|n}\mu(e)\chi\left(\frac{\log e}{\log N^\gamma}\right)\right)^2.$$

As we will state in the next lemma, there exists a positive constant $C_{D,\chi}$ such that the function $r_{D,\gamma}$ has average $C_{D,\chi}+o(1)$.

We now define for any integer q and any $b\in[q]$ the function $\nu_{\mathrm{Matt},b,D}:[N']\to\mathbb{R}$ by

$$\nu_{\mathrm{Matt},q,b,D}(n) = r_{D,\gamma}(qn+b)/C_{D,\chi}. \tag{4.17}$$

When q is implicitly understood to be $\overline{W}$, it may be omitted from the subscripts of ν. The next lemma, drawn from [68, Lemma 7.5], also asserts that this function is a pseudorandom majorant for the representation function of any PDBQF of discriminant D.

Lemma 4.10. *For any PDBQF f of discriminant D and $b\in[\overline{W}]$ satisfying $b\not\equiv 0 \bmod p^{\iota(p)}$ for any $p\le w(N)$ and $\rho_{f,b}(\overline{W})>0$, the following bound holds*

$$r'_{f,b}(n) \ll \nu_{\mathrm{Matt},b,D}(n).$$

Furthermore, for some positive constant $C_{D,\chi}=O(1)$, we have $\mathbb{E}_{n\in[N']}r_{D,\gamma}=C_{D,\chi}+o(1)$ and $\nu_{\mathrm{Matt},b,D}(n)=1+o(1)$.

The crucial property of ν_{Matt} is that it is a truncated divisor sum, like ν_{GT}. Indeed, all

divisors appearing in it are constrained to be less than $R = N^{\gamma}$. It is obvious by definition of χ for the divisors called d, m, e, and less obvious, but proven by Matthiesen, for u (see Remark 3 following Proposition 4.2 in [68]). Moreover, the divisors d, m, e only have prime factors larger than $w(N)$ while u only has prime factors larger than $N^{(\log\log N)^{-3}}$.

The majorant of the divisor function we have just introduced looks extremely complicated, and other majorants are available in the literature, also of the form of truncated divisor sums. Let us mention the work of Landreau [59], revisiting a theorem of van der Corput [92], in which inequalities of the form

$$\tau(n) \leq k^{k(k-1)} \sum_{d \leq n^{1/k}} \tau(d)^k$$

for any integer $k \geq 1$ are discussed. Unfortunately, the right-hand side has an average of size $\log^{C(k)} N$ up to N, where $C(k) > 1$ for $k > 1$. This is much larger than the average of the left-hand side, which is asymptotic to $\log N$. Thus, such majorants cannot be used as pseudorandom measures, and we are left with the sixty years old idea of Erdős.

Finally, we need to produce a common majorant for the $t + s + 1$ functions occurring in Theorem 4.7, which are copies of the von Mangoldt functions and quadratic representation functions. Now each of them is bounded individually by some pseudorandom majorant defined above, so we define our common majorant by averaging all these majorants. Recall that $N' = N/\overline{W}$; we take M to be a prime satisfying $N' < M \leq O(N')$. Given a family $f_0, f_{t+1}, \ldots, f_{t+s}$ of PDBQF of discriminants $D_0, D_{t+1}, \ldots, D_{t+s}$ and a family $(b_0, \ldots, b_{t+s}) \in B$, we define a function ν^* on $[N'] \subset \mathbb{Z}/M\mathbb{Z}$ by

$$\nu^*(n) = \frac{1}{t+s+2}(1 + \sum_{i=1}^{t} \nu_{\mathrm{GT},b_i}(n) + \sum_{j=t+1}^{t+s} \nu_{\mathrm{Matt},b_j,D_j}(n) + \nu_{\mathrm{Matt},b_0,D_0}(n)), \tag{4.18}$$

where ν_{GT,b_i} is a shortcut $\nu_{\mathrm{GT},\overline{W},b_i}$. We extend ν^* to $\mathbb{Z}/M\mathbb{Z}$ by setting $\nu^*(n) = 1$ outside $[N']$. Our strategy of forming a common majorant for a family of functions by averaging a family of majorants is not unheard of. In fact, Green and Tao [45] had to combine the majorants $n \mapsto \Lambda_{\chi,\gamma}(\overline{W}n + b_j)$ for various b_j and so did Matthiesen [68]. Notice also that Lê and Wolf [62] devised a certain condition of compatibility for two pseudorandom majorants. However, in our case the majorants have rather different origins. But they have a similar structure, the structure of a truncated divisor sum, so that the proof of the linear forms condition will not be much harder than the ones in [45] or [68].

We observe that ν^* satisfies

$$1 + \sum_{i=1}^{t} \Lambda'_{\overline{W},b_i} + \sum_{j=t+1}^{t+s} r'_{f_j,b_j} + r'_{f_0,b_0} \ll \nu^*$$

and has average $1 + o(1)$ by Proposition 2.11 and Lemma 4.10. So to ensure that ν^* is a pseudorandom measure, it remains to prove the linear forms condition (2.6). This is the content of the next proposition.

Proposition 4.11. *Fix a constant $D > 0$, and positive integers t, s. Then there exists a constant $C_0(D)$ such that the following holds. For any bounded $C \geq C_0(D)$ there exists $\gamma = \gamma(C, D)$ such that if $M \in [CN', 2CN']$ is a prime, $b \in B_{t,s+1}$ and $f_0, f_{t+1}, \ldots, f_{t+s}$ are PDBQF and ν^* is defined as in equation (4.18), then ν^* satisfies the D-linear forms condition, and for any $i \in [t]$ we have*

$$\Lambda'_{\overline{W},b_i} \ll \nu^*.$$

Similarly, we have

$$|r'_{f_0,b_0} - 1| \ll \nu^*,$$

and for any $j \in [\![t+1 \,;\, t+s]\!]$, we have

$$r'_{f_j,b_j} \ll \nu^*$$

where all inequalities are valid on $[N']$.

The inequalities have already been observed above. The linear forms condition will follow from the following proposition.

Proposition 4.12. *Let $1 \leq d, t, s \leq D$, where D is the constant appearing in Theorem 2.8. Let $c_{\mathrm{GT}}(\chi)$ be the constant appearing in Proposition 2.9. For any finite complexity system $\Psi : \mathbb{Z}^d \to \mathbb{Z}^{t+s}$ whose linear coefficients are bounded by D and every convex $K \subset [0, N]^d$ such that $\Psi(K) \subset [1, N/\overline{W}]^t$, and any $b \in B$ (as in Definition 4.2), the estimate*

$$\mathbb{E}_{n \in \mathbb{Z}^d \cap K} \prod_{j=t+1}^{t+s} \nu_{\mathrm{Matt},D_j,b_j}(\psi_j(n)) \prod_{i \in [t]} \nu_{\mathrm{GT},b_i}(\psi_i(n)) = 1 + O_D\left(\frac{N^{d-1+O_D(\gamma)}}{\mathrm{Vol}(K)}\right) + o_D(1) \quad (4.19)$$

holds, provided γ is small enough.

Notice that t and s are not the same as in Proposition 4.11. The proof is postponed to Appendix C, due its utter length and complexity.

As mentioned in Section 2.4, deriving the linear forms conditions for ν^* (Proposition 4.11) from Proposition 4.12 is a standard procedure. The argument does not need any modification, so we do not reproduce it here and invite the reader to consult one of the references [44, Proposition 9.8] or [22, Proposition 8.4]. We can now prove Theorem 4.7.

Proof of Theorem 4.7 assuming Proposition 4.11. Take any integers d, t and s, and a system $\Phi : \mathbb{Z}^d \to \mathbb{Z}^{t+s+1}$ of affine-linear forms of finite complexity, where the coefficients of the linear part are bounded by L, and let $f_0, f_{t+1}, \ldots, f_{t+s}$ be any PDBQF. Let D be the constant indicated by Theorem 2.8. Fix $\gamma = 2^{-k}$ such that Proposition 4.11 holds. Take a convex set $K \subset [1, N']^d$ such that $\Phi(K) \subset [N']^{t+s+1}$. Let $b \in B$. Then Proposition 2.8 and Proposition 4.11 provide constants C_0 and Γ, of which we take the maximum $C = \max(C_0, \Gamma)$. Now take a prime $M \in [CN', 2CN']$. Such a prime exists by Bertrand's postulate. Define ν^* as in (4.18). Define $F_0 = r'_{f_0,b_0} - 1$. Set $F_i = \Lambda'_{\overline{W},b_i}$ for $i \in [t]$ and $F_j = r'_{f_j,b_j}$ for $j \in \{t+1, \ldots, t+s\}$. Then we have that $|F_j| \ll \nu^*$ for all $j \in \{0, \ldots, t+s\}$ and ν^* is a pseudorandom measure by Proposition 4.11, so that we can invoke the von Neumann theorem (Theorem 2.8). Together with the statement of Proposition 4.8 (specialised to $k = t + s$), it implies Theorem 4.7. $\square$

Chapter 5

Bilinear structures in vector spaces over finite fields

This chapter is based on a preprint of Thai Hoàng Lê and the author [12], submitted to the Journal de l'Ecole Polytechnique. Most of the mathematics and redaction was done by the author.

Thanks are due to Ben Green and Terence Tao for suggesting respectively this chapter's main result, Theorem 5.8, and a sketch of proof. Before stating our main result, we review Bogolyubov's theorem and some basic facts of discrete Fourier analysis.

5.1 Preliminaries

We fix a prime p. We work in the finite field model introduced in Section 1.4, so we let V be an $\mathbb{F}_p$-vector space of dimension n, where we think of n as tending to infinity. Recall from the introduction that the *density* of a subset $A \subset V$ is the quantity $\alpha = \frac{|A|}{|V|}$.

Theorem 5.1 (Bogolyubov)**.** *If $A \subset V$ is a set of density $\alpha > 0$, then the sumset*

$$A + A - A - A := \{a_1 + a_2 - a_3 - a_4 \mid (a_1, \ldots, a_4) \in A^4\}$$

contains a vector subspace of codimension $c(\alpha) = O(\alpha^{-2})$.

The notation $A + A - A - A$ is often abbreviated as $2A - 2A$. We shall give a (short, folklore) proof of Theorem 5.1. To this aim, and for the rest of the chapter, we need to introduce the basics of discrete Fourier analysis.

We denote by $\widehat{V}$ the *dual* of V, the set of characters on V. A character $\chi \in \widehat{V}$ takes values in the p-th roots of unity, that is, $1, \omega, \ldots, \omega^{p-1}$ where $\omega = \exp(2i\pi/p)$. The trivial character is $\chi = 1$. Let $f : V \to \mathbb{C}$ be a function. Then the Fourier transform $\hat{f}$ is defined on $\widehat{V}$ by

$$\hat{f}(\chi) = \mathbb{E}_{x \in V} f(x)\chi(x).$$

In particular, if $A \subset V$ has density α and indicator function 1_A, we have $\widehat{1_A}(1) = \alpha$. Besides, we have $\widehat{1_{-A}} = \overline{\widehat{1_A}}$.

Let W be an affine subspace of V of direction $\overrightarrow{W}$, that is, $W = a + \overrightarrow{W}$ for some $a \in V$ and some subspace $\overrightarrow{W}$ of V. If $f : W \to \mathbb{C}$ is a function, we define the function $\tilde{f}$ on the vector space $\overrightarrow{W}$ by $\tilde{f}(v) = f(a + v)$. We then define the Fourier transform of f relative to W as the Fourier transform of $\tilde{f}$ on $\overrightarrow{W}$. We will abuse notation and denote by $\widehat{W}$ the dual of $\overrightarrow{W}$. Thus the notion of Fourier transform depends on the (potentially affine) subspace W one is considering, but when no ambiguity is possible, the space considered may not be made explicit.

Besides, if $f, g : V \to \mathbb{C}$ are two functions, we define their *convolution* $f * g : V \to \mathbb{C}$ by

$$f * g(x) = \mathbb{E}_{y \in V} f(y)\overline{g(x - y)}.$$

We define the U^2 norm by

$$\|f\|^4_{U^2(V)} = \mathbb{E}_{x \in V} |f * f(x)|^2.$$

A quadruple $(x_1, x_2, x_3, x_4) \in V^4$ satisfying $x_1 + x_2 = x_3 + x_4$ is called an *additive quadruple*.

Observe that if $f = 1_A$ is the indicator function of the subset $A \subset V$, then

$$\|1_A\|^4_{U^2(V)} = \frac{|\{(x_1, x_2, x_3, x_4) \in A^4 \mid x_1 + x_2 = x_3 + x_4\}|}{|V|^3}$$

and we refer to this quantity as the *density of additive quadruples* in A. Again if W is an affine subspace of V and $f : W \to \mathbb{C}$ is a function, we will write $\|f\|_{U^2(W)} = \|\tilde{f}\|_{U^2(\overrightarrow{W})}$. Note that the connection with the additive quadruples of $A \subset W$ is preserved, because additive quadruples are invariant by translation. When it is obvious from the context which space one is considering, one will simply write $\|f\|_{U^2}$.

We recall without proof a few basic properties of the Fourier transform.

1. Parseval's identity is the statement that

$$\mathbb{E}_{x\in V}|f(x)|^2 = \sum_{\chi\in\widehat{V}}|\hat{f}(\chi)|^2.$$

In particular, for a subset $A \subset V$ of density α, we have

$$\sum_{\chi\in\widehat{V}}|\widehat{1_A}(\chi)|^2 = \alpha. \tag{5.1}$$

2. The Fourier transform of a convolution is the product of the Fourier transforms, that is

$$\widehat{f * g} = \hat{f}\hat{g}. \tag{5.2}$$

3. Combining the previous two points, we see that the U^2 norm of a function is the L_4 norm of its Fourier transform, that is

$$\|f\|_{U^2(V)} = \|\hat{f}\|_4.$$

In particular if $f = 1_A$ for a subset A of density α, Parseval's identity implies that

$$\alpha^4 \leq \|\widehat{1_A}\|_4^4 = \alpha^4 + \sum_{\chi\in\widehat{V},\,\chi\neq 1}|\widehat{1_A}(\chi)|^4 \leq \alpha^4 + \alpha \max_{\chi\in\widehat{V},\,\chi\neq 1}|\widehat{1_A}(\chi)|^2. \tag{5.3}$$

When a set $A \subset W$ of density α has about as few additive quadruples as it can, that is, $\alpha^4 \leq \|1_A\|_{U^2(W)}^4 \leq \alpha^4(1+\epsilon)$, we will call it *$\epsilon$-pseudorandom*. In particular, A is ϵ-pseudorandom in W if $\max_{\chi\in\widehat{W},\,\chi\neq 1}|\widehat{1_A}(\chi)| \leq \alpha^{3/2}\epsilon^{1/2}$.

4. The Fourier inversion formula is the statement that

$$f = \sum_{\chi\in\widehat{V}}\hat{f}(\chi)\overline{\chi}. \tag{5.4}$$

We now give the proof of Theorem 5.1. We observe that $2A - 2A$ is the support of the convolution $g = 1_A * 1_A * 1_{-A} * 1_{-A}$. So we simply need to find a large subspace W such that $g(x) > 0$ for all $x \in W$. Because of equation (5.2), we have $\widehat{g}(\chi) = |\widehat{1_A}(\chi)|^4$. Let $K = \{\chi \in \widehat{V} \mid |\widehat{1_A}(\chi)| \geq \rho\}$ for some constant ρ to be determined later, and let $W = \{x \in V \mid \forall\chi \in K \quad \chi(x) = 1\}$. It is a vector subspace of codimension at most $|K|$.

Because of Parseval's identity, or rather equation (5.1), we have $|K| \leq \rho^{-2}\alpha$. Furthermore, using equation (5.4), we have

$$g = \sum_{\chi \in \widehat{V}} |\widehat{1_A}(\chi)|^4 \chi = \sum_{\chi \in K} |\widehat{1_A}(\chi)|^4 \chi + \sum_{\chi \notin K} |\widehat{1_A}(\chi)|^4 \chi.$$

On W, the first sum is simply $\sum_{\chi \in K} |\widehat{1_A}(\chi)|^4 \geq \widehat{1_A}(0)^4 = \alpha^4$ by positivity. Meanwhile, the second sum is bounded by $\sup_{\chi \notin K} |\widehat{1_A}(\chi)|^2 \sum_{\chi \in \widehat{V}} |\widehat{1_A}(\chi)|^2 \leq \rho^2 \alpha$. By taking $\rho = \alpha^{3/2}/2$ and using the triangle inequality, we see that $g > 0$ on W, while $\mathrm{codim} W = O(\alpha^{-2})$. This concludes the proof.

In the same vein, we prove the following useful lemma. It says that if A is sufficiently pseudorandom in terms of its density then $2A - 2A$ is the whole space.

Lemma 5.2. *Let W be an affine subspace of V and $A \subset W$ have density α. If $\|1_A - \alpha\|_{U^2(W)} < \alpha$, or equivalently,*

$$\sum_{\chi \neq 1} |\widehat{1_A}(\chi)|^4 < \alpha^4, \tag{5.5}$$

then $2A - 2A = \overrightarrow{W}$. Consequently, if $\max_{\chi \in \widehat{W}, \chi \neq 1} |\widehat{1_A}(\chi)| < \alpha^{3/2}$ then $2A - 2A = \overrightarrow{W}$.

Proof. For any $x \in \overrightarrow{W}$, by the Fourier inversion formula (5.4), we have

$$1_A * 1_A * 1_{-A} * 1_{-A}(x) = \sum_{\chi \in \widehat{W}} |\widehat{1_A}(\chi)|^4 \overline{\chi(x)} \geq \alpha^4 - \sum_{\chi \neq 1} |\widehat{1_A}(\chi)|^4 > 0.$$

This implies that $x \in 2A - 2A$. □

We also need the following standard fact which relates the lack of pseudorandomness to density increment.

Lemma 5.3 ([38, Lemma 3.4])**.** *Let W be an affine subspace of V and $A \subset W$ have density α. Suppose there exists $\chi \in \widehat{W}, \chi \neq 1$ such that $|\widehat{1_A}(\chi)| \geq \beta$. Then there exists an affine subspace $H \leq W$ of codimension 1 such that the density of $A \cap H$ on H is at least $\alpha + \beta/2$.*

Our next tool is a regularity lemma.

Lemma 5.4. *Let W be an affine subspace of V and $A \subset W$ have density α. Let $\epsilon > 0$. For any t, there exists an affine subspace $H \leq W$ of codimension $O(t\epsilon^{-1} \log \alpha^{-1})$ such that $|A'| = \alpha' |H|$ (where $A' = A \cap H$) with $\alpha' \geq \alpha$ and for any affine subspace F of codimension at most t of H, $\frac{|A \cap F|}{|F|} \leq \alpha(1+\epsilon)$. Consequently, for any affine subspace F of codimension at most t of H, we also have $\frac{|A \cap F|}{|F|} \geq \alpha(1 - p^t \epsilon)$.*

Proof. Let us prove the first conclusion. If W does the trick already, we do nothing. If not, there exists a subspace H of codimension at most t such that $\frac{|A\cap H|}{|H|} > \alpha(1+\epsilon)$. We replace W by H, and A by $A\cap H$. And we iterate. We duplicate the density in at most ϵ^{-1} iterations. And we may duplicate up to $\log\alpha^{-1}$ times before hitting 1. At every iteration we may lose up to t dimensions. Whence the first conclusion. The second conclusion follows from summing the upper bound over all cosets of F. □

In particular, when $t = 1$, the following corollary says that we can always suppose that a set $A \subset W$ is pseudorandom, at the cost of passing to a subset in an affine subspace.

Corollary 5.5. *Let W be an affine subspace of V and $A \subset W$ have density α. Let $\epsilon > 0$. Then there exists an affine subspace $H \leq W$ of codimension $O((\alpha\epsilon)^{-1/2}\log\alpha^{-1})$ such that $A' = A \cap H$ has density $\geq \alpha$ and is ϵ-pseudorandom in H.*

Proof. We use Lemma 6 with $\beta = \alpha^{3/2}\epsilon^{1/2}$, and Lemma 7 with $t = 1$ and $\epsilon' = \alpha^{1/2}\epsilon^{1/2}/2$ to obtain the conclusion. □

Next we state a standard lemma, which plays a key role in the proof of the U^3 inverse theorem [43], and which results from the combination of the Balog-Szemerédi-Gowers [83, Theorem 2.29] and Freiman-Ruzsa theorems. A useful reference for this lemma is [40, Lecture 2]. We reproduce the proof as we want to incorporate the quasipolynomial bound of Sanders [78, Theorem 11.4] for the Freiman-Ruzsa theorem (see [65] for an accessible survey centered on finite fields). Note that under the polynomial Freiman-Ruzsa conjecture [93, Conjecture 2.10], this bound is polynomial.

Lemma 5.6. *Let $W \leq V$ be $\mathbb{F}$-vector spaces and $A \subset W$ have density α. Let $c > 0$ be a constant. Suppose $\xi : A \to V$ is such that are at least $c|A|^3$ additive quadruples in the graph $\Gamma = \{(y, \xi(y)) \mid y \in A\}$. Then there is a subset $S \subset A$ such that $\xi_{|S}$ coincides with an affine-linear map. Moreover, the density of S in A can be taken quasipolynomial in c, that is $|S| \gg |A|\exp(-\log^{O(1)} c^{-1})$.*

Proof. First, the Balog-Szemerédi-Gowers theorem implies that there exists a set $A' \subset A$ satisfying $|A'| \geq C|A|$ that induces a subgraph $\Gamma' \subset \Gamma$ satisfying $|\Gamma' + \Gamma'| \leq C'|\Gamma|$, where both C and C' can be taken polynomial in c. Using the Freiman-Ruzsa theorem (with Sanders' bounds from [78, Theorem 11.4]), we get a subgraph $\Gamma'' \subset \Gamma'$ corresponding to a subset $A'' \subset A'$ satisfying $|\Gamma''| \geq D|\Gamma'|$ and $|\mathrm{span}(\Gamma'')| \leq E|\Gamma''|$ with D polynomial and E quasipolynomial in c. Write $H = \mathrm{span}(\Gamma'') \leq W \times V$ and $\pi : H \to W$ the canonical projection of $W \times V$ to the first coordinate restricted to H. Then $\pi(H) \supset A''$ by definition.

Because $|H| \leq E|A''|$, the size of the kernel of π is at most E. Then we can partition H into at most E cosets of some subspace H' so that π is injective on each of them. By the pigeonhole principle, there exists such a coset that has a large intersection with Γ'', that is, an $x \in W$ such that

$$|(x + H') \cap \Gamma''| \geq |\Gamma''|/E.$$

Let now $\Delta = (x + H') \cap \Gamma''$ and S be the corresponding subset of A''. The map $\pi_{|x+H'}$ is a bijection onto its image, an affine space $M \leq V$. Its inverse function is an affine map $\psi : M \to W$ such that $(s, \psi(s)) \in \Gamma''$ for all $s \in S$, that is, $\psi(s) = \xi(s)$. Moreover,

$$|S| = |\Delta| \geq |A''|/E \geq K|A|$$

where K is quasipolynomial in c. □

A variant of Lemma 5.6 can be found in [43]. There the conclusion is that ξ is linear in a subset of polynomial (in c) density inside an affine subspace of polynomial codimension. Unfortunately, due to various losses in other places of our argument (in particular, the fact that c itself is ultimately quasipolynomial in α), we cannot make this improvement on the density bear fruit.

To conclude this section, we state, without proof, a deep improvement on the constant $c(\alpha)$ appearing in Theorem 5.1, which is due to Sanders [78, Theorem 11.1].

Theorem 5.7. *We can take $c(\alpha) = O(\log^4 \alpha^{-1})$ in Theorem 5.1.*

As noted in the introduction, the case of a subspace A shows that Sanders' result is optimal up to the exponent 4.

5.2 The bilinear Bogolyubov theorem

Our aim is to prove a bilinear version of Theorem 5.1. Let $P \subset V \times V$ be a set of pairs. As in Section 1.4, let

$$P \overset{V}{\pm} P = \{(x, y_1 \pm y_2) \mid (x, y_1), (x, y_2) \in P\}$$

be the set of vertical sums or differences. Similarly define $P \overset{H}{\pm} P$ the set of horizontal sums or differences, where V and H stand for "vertical" and "horizontal", respectively.

We denote by ϕ_V the operation

$$P \mapsto (P \overset{V}{+} P) \overset{V}{-} (P \overset{V}{+} P),$$

and define the operation ϕ_H analogously. Recall from Section 1.4 that a *bilinear set* of codimensions (r_1, r_2, r_3) is a set $P \subset V \times V$ for which there exist subspaces $W_1 \leq V, W_2 \leq V$ of codimension r_1, r_2, respectively, and bilinear forms $Q_1, \ldots, Q_{r_3}$ on $W_1 \times W_2$ such that

$$P = \{(x, y) \in W_1 \times W_2 \mid Q_1(x, y) = \cdots = Q_{r_3}(x, y) = 0\}. \tag{5.6}$$

A bilinear set satisfies $\phi_V(P) = \phi_H(P) = P$, so it is natural to imagine that iterating the operations ϕ_V and ϕ_H always produces large bilinear sets. We show that this is indeed the case.

Theorem 5.8. *For any $\delta > 0$, there exists a constant $c(\delta) > 0$ such that the following holds. Let $P \subset V \times V$ have density δ. Let $P' = \phi_H \phi_V \phi_H(P)$. Then P' contains a bilinear set of codimensions (r_1, r_2, r_3), where $\max(r_1, r_2, r_3) \leq c(\delta)$. Moreover, $c(\delta) = O(\exp(\exp(\exp(\log^{O(1)} 1/\delta))))$.*

If P is a Cartesian product $A \times B$ for some subsets $A, B \subset V$, then using Theorem 5.1 once on each coordinate, we obtain a product $A' \times B'$ of subspaces of codimension $O(\log^4 \delta^{-1})$. Also it is easy to see that $c(\delta) \gg \log \delta^{-1}$ by considering a bilinear set. It is reasonable to believe that, like in the linear case, this lower bound on δ should not be too far from the truth. Indeed, the proof of Theorem 5.8 will show that $\phi_H \phi_V \phi_H(P)$ contains a set as in (5.6) with $r_1, r_3 = O(\log^{O(1)} \delta^{-1})$ but unfortunately we do not currently have a matching bound for r_2. We state the following conjecture.

Conjecture 5.9 (polylogarithmic bilinear Bogolyubov)**.** *In Theorem 5.8, one can take $c(\delta) = O(\log^{O(1)} \delta^{-1})$.*

The conjecture remains equally interesting and useful for the application we have in mind if $O(1)$ operations ϕ_V or ϕ_H are required instead of 3.

We point out that Gowers and Milićević [35] independently proved a result very similar to Theorem 5.8. Interestingly, they obtained the bound $\exp(\exp(\log^{O(1)} 1/\delta))$ for all three parameters r_i, which is better than our bound for r_2 but worse for r_1 and r_3. Their proof is quite different and draws on ingredients from [34]. As a crucial step in their program towards a quantitative version of the inverse theorem for the Gowers norm $U^4(\mathbb{F}_p^n)$, they

also devised a variant of the bilinear Bogolyubov theorem that is based on convolutions of functions rather than sumsets [36, Section 4].

5.3 A quick application

Our first application concerns matrices of low rank. Roughly speaking, it says that if a two-parameter, bilinearly varying family of matrices is often of rank at most ϵ, then it must be of rank $O(\epsilon)$ on a whole bilinear set. We now state this application precisely. Given a matrix $A \in \mathrm{Mat}_m(\mathbb{F}_p)$, let $\mathrm{rk}A$ be its rank.

Corollary 5.10. *Suppose that we have a bilinear map $\psi : V \times V \to \mathrm{Mat}_m(\mathbb{F}_p)$. Suppose that the set*

$$P_\epsilon = \{(f, g) \in V \times V \mid \mathrm{rk}(\psi(f, g)) \leq \epsilon\}$$

has density $\delta > 0$. Then the set

$$P_{64\epsilon} = \{(f, g) \in V \times V \mid \mathrm{rk}(\psi(f, g)) \leq 64\epsilon\}$$

contains a set of the form (5.6) *of codimensions at most $c(\delta)$. Besides, if δ is large enough (in terms of n), then $P_{64\epsilon}$ contains a diagonal pair (x, x) with $x \neq 0$.*

This corollary, with the conjectured bound on $c(\delta)$ in Theorem 5.8, will play a crucial role in Chapter 7.

Proof. We apply Theorem 5.8 to P. Observe that the set P' it produces is included in $P_{64\epsilon}$ by the bilinearity of ψ and the fact that $\mathrm{rk}(A + B) \leq \mathrm{rk}A + \mathrm{rk}B$ for any two matrices A and B. The second part of the statement is a direct consequence of the following lemma of independent interest, which is reminiscent of [40, Lemma 4.2].

Lemma 5.11. *Let W be an $\mathbb{F}_p$-vector space of dimension n, and $Q_1, \ldots, Q_r$ be quadratic forms on W. Then the set of isotropic vectors*

$$X = \{x \in W | Q_1(x) = \cdots = Q_r(x) = 0\} \tag{5.7}$$

contains at least $(1 - p^{-1/2})p^{n-2r(r+1)}$ elements.

More compactly, we can write $|X| \gg p^{n-O(r^2)}$. We now prove Lemma 5.11.

Proof. The density $|X|/|W|$ of isotropic vectors is given by

$$\mathbb{E}_{x\in W}\mathbb{E}_{t_1,\ldots,t_r\in\mathbb{F}_p}\omega^{\sum_i t_iQ_i(x)} = \mathbb{E}_{t_1,\ldots,t_r}\mathbb{E}_{x\in W}\omega^{\sum_i t_iQ_i(x)}. \tag{5.8}$$

Let $m \leq n$ be a parameter to be determined later (in terms of r). Now if a quadratic form Q on $W \times W$ has rank at least m, it is well known that

$$|\mathbb{E}_{x\in W}\omega^{Q(x)}| \leq p^{-m/2}$$

(see also Lemma 7.10). Thus, if for any nonzero $(t_1, \ldots, t_r)$, the rank of $\sum_i t_iQ_i$ is at least m, we see from equation (5.8) that the density of isotropic vectors is at least $p^{-r} - p^{-m/2}$. Otherwise, there exists a form Q_i such that $Q_i = \sum_{j\neq i} t_jQ_j + R$ with $\operatorname{rk} R < m$; without loss of generality, suppose $i = r$. Let W' be the kernel of R, a subspace of codimension less than m. Then the set

$$X' = \{x \in W' | Q_1(x) = \cdots = Q_{r-1}(x) = 0\} \tag{5.9}$$

is a subset of X, and we will now count isotropic vectors in X'. Thus incurring a dimension loss of at most m, we reduce the number of quadratic forms by 1. We iterate this process until we obtain a family of quadratic forms any nontrivial linear combination of which has rank at least m (or an empty family). At that point, this is a family of at most r forms on a space of dimension at least $n - rm$. Thus it must have at least

$$p^{n-r(m+1)} - p^{n-rm-m/2}$$

isotropic vectors. Taking $m = 2r + 1$, we obtain the result. □

This concludes the proof of Corollary 5.10. □

5.4 Proof of the main theorem

To prove our theorem, we will apply the linear Bogolyubov theorem (Theorem 5.1) several times.

Write $P = \cup_{y\in V} B_y \times \{y\}$. Because P has density δ, the set A of elements $y \in V$ such that $|B_y| \geq \delta|V|/2$ has density at least $\delta/2$. Using Theorem 5.1 on each set B_y for $y \in A$, we see that $\phi_H(P)$ contains a set $P' = \cup_{y\in A} V_y \times \{y\}$ where V_y is a subspace of codimension

$O(\log^4 1/\delta)$. This argument reduces Theorem 5.8 to the following proposition.

Proposition 5.12. *Let $A \subset V$ have density $\alpha > 0$. Let $P = \cup_{y \in A} V_y \times \{y\} \subset V \times V$, where each V_y is a subspace of codimension at most $r = O(\log^4 \alpha^{-1})$. Let $P' = \phi_H \phi_V(P)$. Then there exist subspaces W_1, W_2 of codimension r_1, r_2 and bilinear forms $Q_1, \ldots, Q_{r_3}$ on $W_1 \times W_2$ such that*

$$\{(x, y) \in W_1 \times W_2 \mid Q_1(x, y) = \cdots = Q_{r_3}(x, y) = 0\} \subset P'$$

with r_3, r_1, r_2 of size $O(\exp(\exp(\exp(\log^{O(1)} \alpha^{-1}))))$.

We now prove Proposition 5.12. The theorem will be achieved through the following iteration scheme. Let V^* be the linear dual of V, that is, the set of linear forms on V. For $(x, \xi) \in V \times V^*$, we denote $x \cdot \xi = \xi(x)$. For a set $U \leq V$, we let $U^\perp = \{\xi \in V^* \mid \forall x \in U\, x \cdot \xi = 0\}$. Roughly speaking, this iteration scheme consists in finding a spanning set $(\xi_1(y), \ldots, \xi_r(y))$ of $V_y^\perp$ where more and more linear forms $\xi_i(y)$ are constant or vary linearly (or more generally affinely) with y.

Proposition 5.13. *Let V be an $\mathbb{F}$-vector space, and W be an affine subspace. Let $r \leq \dim V$ be an integer and $\alpha > 0$. Let $\epsilon = p^{-r}/256$. Then there exists a constant $c(r, \alpha)$ such that the following holds. Let $A \subset W$ be an ϵ-pseudorandom subset of density α. Let $P = \cup_{y \in A} V_y \times \{y\} \subset V \times W$ where each V_y is a subspace of codimension at most r. Suppose there exist $s \leq r$ and affine maps $\xi_1, \ldots, \xi_s$ from W to V^* and spaces $U_y \leq V^*$ for $y \in A$ of dimension at most $r - s$ such that $V_y^\perp = \operatorname{span}(\xi_j(y))_{j \in [s]} + U_y$. Then at least one of the following statements holds.*

1. *(Termination) The set $\phi_V(P)$ contains*

$$\{(x, y) \in X_3 \times W_2 \mid x \cdot \overrightarrow{\xi_1}(y) = \cdots = x \cdot \overrightarrow{\xi_s}(y) = 0\}$$

   where $\overrightarrow{\xi}$ denotes the linear part of an affine map ξ, W_2 is the direction of W and X_3 is a subset of density at least $p^{-r}/12$ in V.

2. *(Reduction of codimension) There exist a space $V' \leq V$ of codimension at most $4r$ and subspaces $V'_y \leq V'$ of codimension $r - 1$ for each $y \in A$ such that $P \supset \bigcup_{y \in A} V'_y \times \{y\}$. Besides there exist affine maps $\xi'_1, \ldots, \xi'_s$ from W to V'^* and spaces $U'_y \leq V'^*$ for $y \in A$ of dimension at most $r - s$ such that $(V'_y)^\perp = \operatorname{span}(\xi_j(y))_{j \in [s]} + U'_y$.*

3. *(Linearisation) There exist a set $S \subset A$ of density $c(r,\alpha)$ and an affine map $\xi_{s+1} : W \to V$ and a space $U'_y < U_y$ such that for all $y \in S$, we have*

$$V_y^{\perp} = \mathrm{span}(\xi_1(y), \ldots, \xi_{s+1}(y)) + U'_y.$$

Moreover $c(r,\alpha)$ can be taken quasipolynomial in αp^{-r}, that is,

$$c(r,\alpha) = O(\exp(\log^{O(1)}(\alpha^{-1}p^r))).$$

We will now use Proposition 5.13 to prove Theorem 5.8. Applying Corollary 5.5 with $\epsilon = p^{-r}/256$ and $r = O(\log^{O(1)} \alpha^{-1})$, we obtain an affine subspace $W^{(0)}$ of V of codimension

$$O((\epsilon\alpha)^{-1/2} \log \alpha^{-1}) = O(p^{r/2}\alpha^{-2/3})$$

such that the set $A_0 := A \cap W^{(0)}$ has density $\alpha_0 \geq \alpha$ and is ϵ-pseudorandom in W_0. We set $V^{(0)} = V, P_0 = \cup_{y \in A_0} V_y \times \{y\} \subset P$ and apply Proposition 5.13 with the tuple $(V^{(0)}, W^{(0)}, A_0, P_0)$ and $s_0 = 0, r_0 = r$.

If the first alternative of Proposition 5.13 holds, we stop.

Suppose the second alternative of Proposition 5.13 holds. We set $V^{(1)} \subset V^{(0)}$ to be the subspace V' given by the second alternative, of codimension $O(r)$. We are also given subspaces $V_y^{(1)} \leq V^{(1)}$ of codimension at most $r_1 \leq r-1$ such that $P \supset \bigcup_{y \in A} V_y^{(1)} \times \{y\}$.

Suppose the third alternative holds. We obtain a set $S \subset A_0$ of density $c(r,\alpha_0)$ in $W^{(0)}$, an affine map $\xi_1 : W_0 \to V^*$ and subspaces $U_y^{(1)} \leq V_y^*$ of dimension at most $r_1 \leq r-1$ such that $V_y^{\perp} = \mathrm{span}(\xi_1(y)) + U_y^{(1)}$. Then we let $V^{(1)} = V$ and $V_y^{(1)} = V_y$. We can find an affine subspace $W^{(1)} \subset W^{(0)}$ of codimension $O(p^{r/2}\alpha_0^{-2/3})$ in $W^{(0)}$ such that the set $A_1 := S \cap W^{(1)}$ is ϵ-pseudorandom and has density $\alpha_1 \geq c(r,\alpha_0)$ in $W^{(1)}$. Let $s_1 = 1$ and $r_1 = r$.

Set $P_1 = \bigcup_{y \in A_1} V_y^{(1)} \times \{y\}$. We have $P_0 \supset P_1$.

We can now apply Proposition 5.13 with $(V^{(1)}, W^{(1)}, A_1, P_1, r_1, s_1)$ and start an iterative process. This iterative process stops whenever one can apply the first item of Proposition 5.13, or when $r - s$ vanishes. When applying either of the last two alternatives, at least one of the parameters r or $r - s$ is decreased by at least one, while the other one cannot increase, so the iteration does eventually stop.

At the i-th stage, we obtain a subspace $V^{(i)} \subset V$ of codimension $O(ri)$, an affine

subspace $W^{(i)} \subset W$ of codimension

$$O(\exp(\log^{C^i} \alpha^{-1})),$$

where C is a constant (depending at most on p), an ϵ-uniform set $A_i \subset W^{(i)}$ of density

$$\alpha_i = \Omega(\exp(-\log^{C^i} \alpha^{-1}))$$

and a set

$$P_i = \cup_{y \in A_i} V_y^{(i)} \times \{y\} \subset V^{(i)} \times W^{(i)}$$

where each $V_y^{(i)} \subset V^{(i)}$ has codimension $r_i \leq r$. Besides, we have affine maps $\xi_1, \ldots, \xi_{s_i}$ from W^i to $V^{(i)*}$ and subspaces $U_y^{(i)} \leq V^{(i)*}$ of dimension at most $r_i - s_i$ such that $(V_y^{(i)})^\perp = \mathrm{span}(\xi_1(y), \cdots, \xi_{s_1}(y)) + U_y^{(i)}$. Furthermore, $P \supset P_i$.

Suppose the algorithm stops after the i-th iteration, where $i \leq 2r$. Note that we have $s_i \leq r$,

$$\mathrm{codim} V^{(i)} = O(r^2) = O(\log^{O(1)} \alpha^{-1}),$$

and

$$\mathrm{codim} W^{(i)} = O(\exp(\log^{C^r} \alpha^{-1})) = O(\exp(\exp(\exp(\log^{O(1)} \alpha^{-1})))).$$

There are two possibilities.

Case 1: $r_i = s_i$, and

$$P_i = \{(x, y) \in V^{(i)} \times A_i \mid x \cdot \xi_1(y) = \cdots = x \cdot \xi_{s_i}(y) = 0\}$$

where $\xi_1, \ldots, \xi_{s_i}$ are affine maps from $W^{(i)}$ to $V^{(i)*}$, $A_i \subset W^{(i)}$ is a set of density $\gamma := \alpha_i = \Omega\left(\exp(-\exp(\exp(\log^{O(1)} \alpha^{-1})))\right)$.

Case 2: The first alternative of Proposition 5.13 holds, and

$$\phi_V(P_i) \supset \{(x, y) \in X \times W_2 \mid x \cdot \overrightarrow{\xi_1}(y) = \cdots = x \cdot \overrightarrow{\xi_{s_i}}(y) = 0\},$$

where $\xi_1, \ldots, \xi_{s_i}$ are affine maps from $W^{(i)}$ to $V^{(i)*}$, $X \subset V^{(i)}$ is a set of density $\Omega\left(p^{-r_i}\right) = \Omega\left(\exp\left(-\log^{O(1)} \alpha^{-1}\right)\right)$ and W_2 is the direction of $W^{(i)}$.

Since the two cases are similar, we will work with Case 1. By translating P by $(0, a)$ for some $a \in A_i$ if necessary, we may assume that $W^{(i)}$ is a vector subspace of V. Let $\eta = \frac{1}{10}\gamma^{3/2} p^{-r-1}$. Applying Lemma 5.4 with $t = r + 1$, there is a subspace $H \leq W^{(i)}$ of

codimension $O(r\eta^{-1}\log\gamma^{-1})$ such that $|A'| = \gamma'|H|$ (where $A' = A_i \cap H$) with $\gamma' \geq \gamma$ and for any subspace F of codimension at most $r+1$ of H, $\frac{|A'\cap F|}{|F|} \leq \gamma(1+\eta)$.

For each $x \in V^{(i)}$, let $B_x = \{y \in H \mid x\cdot\xi_1(y) = \cdots = x\cdot\xi_{s_i}(y) = 0\}$. Then B_x is a subspace of codimension at most r inside H. Let $A_x = A' \cap B_x$. We claim that $2A_x - 2A_x = \overrightarrow{B_x}$.

By Lemma 5.2, it suffices to show $|\widehat{1_{A_x}}(\chi)| < \gamma_x^{3/2}$ for any $\chi \neq 1$, where γ_x is the density of A_x in B_x.

Suppose for a contradiction that this is not true. Then Lemma 5.3 implies that there is a hyperplane F of B_x on which the density of A is at least $\gamma_x + \gamma_x^{3/2}/2$. From Lemma 5.4 we also have $\gamma_x \geq \gamma(1-\eta p^r)$. Therefore,

$$\begin{aligned}\gamma_x + \gamma_x^{3/2}/2 &\geq \gamma(1-\eta p^{r+1}) + \frac{1}{2}\gamma^{3/2}(1-\eta p^{r+1})^{3/2}\\ &\geq \gamma(1-\eta p^{r+1}) + \frac{1}{2}\gamma^{3/2}(1-2\eta p^{r+1})\\ &\geq \gamma - \frac{1}{10}\gamma^{3/2} + \frac{2}{5}\gamma^{3/2} = \gamma + \frac{3}{10}\gamma^{3/2} > \gamma+\eta. \end{aligned} \tag{5.10}$$

This contradicts the assumption on H since F is a subspace of codimension at most $r+1$ of H. Therefore, $2A_x - 2A_x = \overrightarrow{B_x}$ and

$$\phi_V(P) \supset \cup_{x\in V^{(i)}}\{x\}\times\overrightarrow{B_x} = \{(x,y)\in V^i\times H \mid x\cdot\overrightarrow{\xi_1}(y) = \cdots = x\cdot\overrightarrow{\xi_{s_i}}(y) = 0\}.$$

Since the codimension of H in $W^{(i)}$ is $O(r\epsilon^{-1}\log\gamma^{-1})$, Theorem 5.8 follows.

5.5 Proof of the iterative step

In this section, we prove Proposition 5.13. So we take a set $P \subset V\times V$ of the form given in the hypothesis of this proposition. First, suppose there exists a nonzero $\lambda \in \mathbb{F}_p^s$ such that $\xi_0' = \sum_{j\in[s]}\lambda_j\xi_j$ satisfies $\mathrm{rk}\overrightarrow{\xi_0'} < 2r+10 \leq n$. Let $V' = \mathrm{span}(\xi_0'(y) \mid y\in A)^\perp$. By completing λ into a basis of $\mathbb{F}_p^s$, we obtain affine maps $\xi_0',\ldots,\xi_{s-1}'$ from W to V^* such that $\mathrm{span}(\xi_0'(y),\ldots,\xi_{s-1}'(y)) = \mathrm{span}(\xi_1(y),\ldots,\xi_s(y))$ for any $y\in A$. Then $P \supset \bigcup_{y\in A} V_y'\times\{y\}$ where $(V_y')^\perp = \mathrm{span}(\xi_1'(y),\ldots,\xi_{s-1}'(y)) + U_y'$ and U_y' is the projection of U_y onto V'^*. This proves the second alternative.

So let us now suppose that there exists no nonzero $\lambda\in\mathbb{F}_p^s$ such that $\xi_0' = \sum_{j\in[s]}\lambda_j\xi_j$ satisfies $\mathrm{rk}\overrightarrow{\xi_0'} < 2r+10$. We call this condition the *high rank condition.*

For $x \in V$, let $A_x = \{y \in A \mid x \in V_y\} \subset W$. Also, let

$$B_x = \{y \in W \mid x \cdot \xi_1(y) = \cdots = x \cdot \xi_s(y) = 0\};$$

this is an affine subspace of codimension at most s, and $A_x \subset A \cap B_x$. Now let us show that $\mathbb{P}_{x \in V}(\mathrm{codim} B_x < s) \leq \epsilon p^{-r}/4$.

Note that

$$\overrightarrow{B_x} = \{y \in \overrightarrow{W} \mid y \cdot \overrightarrow{\xi_1}^T(x) = \cdots = y \cdot \overrightarrow{\xi_s}^T(x)\}$$

is a subspace of codimension s unless there exists a nonzero $\lambda \in \mathbb{F}_p^s$ such that $\sum_{j \in [s]} \lambda_j \overrightarrow{\xi_j}^T(x) = 0$. For any fixed such λ, the set of x that satisfy this relation is a linear subspace, namely the kernel K_λ of $\sum_{j \in [s]} \lambda_j \overrightarrow{\xi_j}^T$ whose codimension equals the rank of $\sum_{j \in [s]} \lambda_j \overrightarrow{\xi_j}$, hence at least $2r + 10$. Hence $|K_\lambda|/|V| \leq p^{-2r-10} \leq \epsilon p^{-r}/4$. Because there are at most p^r tuples λ to consider, we conclude that $\mathbb{P}_{x \in V}(\mathrm{codim} B_x < s) \leq p^{-r}\epsilon/4$.

Let α_x be the density of A_x in B_x. Let $X = \{x \in V \mid \mathrm{codim} B_x = s\}$. Observe that

$$\begin{aligned}
\mathbb{E}_{x \in V} \alpha_x &= \frac{1}{|V|} \sum_{x \in V} \frac{|A_x|}{|B_x|} \\
&= \frac{1}{|V|} \left(p^s \sum_{x \in X} \frac{|A_x|}{|W|} + \sum_{x \in X^c} \frac{|A_x|}{|B_x|} \right) \\
&\geq p^s \mathbb{E}_{x \in V} \frac{|A_x|}{|W|} - p^s \frac{1}{|V|} \sum_{x \in X^c} \frac{|A_x|}{|W|} \\
&\geq p^s \mathbb{E}_{y \in W} 1_{y \in A} \mathbb{E}_{x \in V} 1_{x \in V_y} - \alpha p^s \frac{|X^c|}{|V|} \\
&\geq \alpha p^{s-r}(1 - \epsilon/4).
\end{aligned}$$

Proposition 5.13 will follow from Lemmas 5.14 and 5.15.

Lemma 5.14. *At least one of the following statements holds.*

1. *For at least $p^{-r}|V|/4$ elements $x \in V$, we have $2A_x - 2A_x = \overrightarrow{B_x}$ (the direction of B_x).*

2. *Among additive quadruples $y_1 + y_2 = y_3 + y_4$ in A, a proportion at least $p^{-4r}\epsilon$ has the property that* $\mathrm{codim} \bigcap_{i=1}^4 V_{y_i} < 4r - s$.

Proof. Let Q be the set of additive quadruples $\mathbf{y} = (y_1, \ldots, y_4)$ of A. Let $m = \dim W$. We

have

$$\begin{aligned}
\mathbb{E}_{x\in V}\|1_{A_x}\|^4_{U^2(B_x)} &= \mathbb{E}_{x\in V}\mathbb{E}_{\substack{y_1,\dots,y_4\in B_x\\ y_1+y_2=y_3+y_4}}\prod_{i=1}^{4}1_{y_i\in A_x}\\
&\leq \frac{1}{p^{3(m-s)}}\sum_{\substack{(y_1,\dots,y_4)\in A^4\\ y_1+y_2=y_3+y_4}}\mathbb{E}_{x\in V}1_{\forall i\, x\in V_{y_i}}\\
&= \frac{1}{p^{3(m-s)}}\sum_{\mathbf{y}\in Q}p^{-\mathrm{codim}\bigcap_i V_{y_i}}\\
&\leq \alpha^4(1+\epsilon)\big(\mathbb{E}_{\mathbf{y}\in Q}(p^{3s-\mathrm{codim}\bigcap_i V_{y_i}})\big)\\
&\leq \alpha^4 p^{-4(r-s)}(1+\epsilon)(1+p^{4r-s}\mathbb{P}_{\mathbf{y}\in Q}(\mathrm{codim}\bigcap V_{y_i}<4r-s).
\end{aligned}$$

So either

$$\mathbb{P}_{\mathbf{y}\in Q}\left(\mathrm{codim}\bigcap_{i=1}^{4}V_{y_i}<4r-s\right)\geq p^{-4r+s}\epsilon \tag{5.11}$$

or

$$\mathbb{E}_{x\in V}\|1_{A_x}\|^4_{U^2(B_x)}\leq \alpha^4 p^{-4(r-s)}(1+\epsilon)^2. \tag{5.12}$$

Equation (5.11) is exactly the second clause of Lemma 5.14, so assume instead that (5.12) holds. We infer that

$$\begin{aligned}
\mathbb{E}_{x\in V}\|1_{A_x}-\alpha_x\|^4_{U^2} &= \mathbb{E}_{x\in V}(\|1_{A_x}\|^4_{U^2}-\alpha_x^4)\\
&\leq \mathbb{E}_{x\in V}\|1_{A_x}\|^4_{U^2}-\alpha^4p^{-4(r-s)}(1-\epsilon/4)^4\\
&\leq \alpha^4p^{-4(r-s)}(2\epsilon+\epsilon^2+\epsilon)\\
&\leq 4\epsilon\alpha^4p^{-4(r-s)}=\gamma
\end{aligned}$$

where we used Jensen's inequality, the lower bound $\mathbb{E}_{x\in V}\alpha_x\geq p^{-(r-s)}(1-\epsilon p^{-r}/4)$ and the elementary inequality $(1-\epsilon/4)^4\geq 1-\epsilon$. Thus, if $X_1=\{x\in V\mid \|1_{A_x}-\alpha_x\|^4_{U^2(B_x)}\leq 4p^r\gamma\}$, by Markov's inequality, we have $|X_1|\geq |V|(1-p^{-r}/4)$. Also, because $\alpha_x\leq \alpha p^s$ for any $x\in V$, the set $X_2=\{x\in V\mid \alpha_x\geq \alpha p^{-(r-s)}/2\}$ has density at least $p^{-r}(1/2-\epsilon/4)\geq p^{-r}/3$. So $X_3=X_1\cap X_2$ must have density at least $p^{-r}/12$ by inclusion-exclusion.

Besides, if $\epsilon=1/(256p^r)$, then for $x\in X_3$ we have $\|1_{A_x}-\alpha_x\|^4_{U^2}<\alpha_x^4$ and then $2A_x-2A_x=\overrightarrow{B_x}$. □

We now prove Proposition 5.13. When the first outcome of Lemma 5.14 holds, we see

that $\phi_V(P)$ contains

$$\{(x,y) \in X_3 \times \overrightarrow{W} \mid Q_1(x,y) = \cdots = Q_s(x,y) = 0\},$$

where $Q_i(x,y) = x \cdot \overrightarrow{\xi_i}(y)$.

The real challenge lies in extracting something from the second outcome of Lemma 5.14. This is the purpose of the next lemma.

Lemma 5.15. *Suppose $r > s$ and a proportion at least κ of the additive quadruples $(y_1, \ldots, y_4)$ of A have the property that $\operatorname{codim} \bigcap_{i=1}^4 V_{y_i} < 4r - s$. Then there is a subset $S \subset A$ of density $\sigma = \sigma(r, \alpha, \kappa)$ such that one of the following holds.*

1. *There is a subspace $V' \leq V$ of codimension at most one and spaces $V'_y \leq V'$ of codimension at most $r-1$ such that $P \supset V'_y \times \{y\}$. Besides, there exist affine maps $\xi'_1, \ldots, \xi'_s$ from W to V^* and spaces $U'_y \leq V'^*$ of dimension at most $r-s$ such that $V'^{\perp}_y = \operatorname{span}(\xi'_1(y), \ldots, \xi'_s(y)) + U'_y$.*

2. *There is an affine map $\xi_{s+1} : W \to V^*$ and a subspace $U'_y \leq V^*$ of dimension at most $r-s-1$ such that $V_y^{\perp} = \operatorname{span}(\xi_j(y) \mid j \in [s+1]) + U'_y$.*

Moreover σ can be taken to be quasipolynomial[1] *in $\alpha\kappa p^{-r}$.*

Applying Lemma 5.15 with $\kappa = p^{-4r}\epsilon = p^{-5r}/256$, the first alternative implies the second statement of Proposition 5.13, while the second alternative yields the third one of Proposition 5.13.

Our goal is now to prove Lemma 5.15.

Proof of Lemma 5.15. Let $\xi_{s+1}, \ldots, \xi_r$ be maps from A to V^* such that

$$U_y = \operatorname{span}(\xi_{s+1}(y), \ldots, \xi_r(y))$$

for any $y \in A$. The number of additive quadruples in A is at least $\alpha^4|W|^3 = \alpha|A|^3$, and we assume at least $\kappa\alpha|A|^3$ of them have the property that the $4r$ vectors $\xi_j(y_i)$ satisfy at least $s+1$ linearly independent equations. For any additive quadruple in A, we already have s obvious equations

$$\xi_j(y_1) + \xi_j(y_2) = \xi_j(y_3) + \xi_j(y_4) \quad \text{for } j \in [s], \tag{5.13}$$

[1]Polynomial under the polynomial Freiman-Ruzsa conjecture.

so there needs to be one more (independent) equation. Because there are only p^{4r} possible linear equations

$$\sum_{j=1}^{r}\sum_{i=1}^{4} a_{i,j}\xi_j(y_i) = 0, \tag{5.14}$$

the pigeonhole principle implies that we can find $(a_{i,j}) \in \mathbb{F}^{4r} \setminus \{0\}$ (linearly independent from the vectors $b_{i,j} = 1_{j=j_0}$ for $j_0 \in [s]$) such that there are at least $\kappa\alpha|A|^3/p^{4r}$ quadruples $(y_1, \dots, y_4) \in W^4$ for which $y_1 + y_2 = y_3 + y_4$ and equation (5.14) holds. Let T be that set of quadruples. Write $\mathbf{a}_i = (a_{i,j})_{j=1,\dots,r}$.

If one of the four families $\mathbf{a}_1, \dots, \mathbf{a}_4$, say $\mathbf{a}_4$, satisfies $a_{4,j} = 0$ for every $j > s$, we can use the equations (5.13) to eliminate y_4 in equation (5.14). We obtain $\phi_1 + \phi_2 + \phi_3 = 0$ for some vectors $\phi_i \in V_{y_i}^{\perp}$ for $i \in [3]$. Write $r(\phi) = |\{y \in A \mid \phi \in V_y^{\perp}\}|$ for any $\phi \in V^*$. Then we have

$$p^{-4r}\kappa\alpha|A|^3 \le |T| \le \sum_{\phi_1,\phi_2\in V} r(\phi_1)r(\phi_2)r(-\phi_1-\phi_2) \le \max_{\phi\in V} r(\phi)\left(\sum_{\phi\in V} r(\phi)\right)^2.$$

A double counting argument shows that $\sum_{\phi\in V} r(\phi) \le p^r|A|$. So $\max r(\phi) \ge \kappa\alpha p^{-6r}|A|$, which implies that there exists a linear form $\phi \in V^*$ such that for a positive proportion of $y \in A$, we have $\phi \in V_y^{\perp}$. Name $S \subset A$ this set of elements $y \in A$, and $V' = \phi^{\perp}$. Projecting the maps ξ_i onto $(V')^*$, we get affine maps $\xi_i' : W \to (V')^*$ for $i \in [s]$, and letting $V_y' = \mathrm{span}(\xi_1'(y), \dots, \xi_r'(y))^{\perp}$, for any $y \in S$, we have $\mathrm{codim}_{V'} V_y' \le r - 1$. This concludes.

So now suppose none of the four families satisfies $a_{i,j} = 0$ for every $j > s$. We shall aim at linearity instead of constancy. By the probabilistic method, we can find a partition of $A = A_1 \cup A_2 \cup A_3 \cup A_4$ in four parts such that there are a lot of quadruples of T in $A_1 \times \cdots \times A_4$. We prove this claim now. Removing quadruples for which two entries are equal (there are $O(|A|^2)$ such quadruples), we still have a set T of quadruples satisfying $|T| \ge C|A|^3$ for some constant $C \gg \kappa\alpha p^{-4r}$. Now for a random partition of A where each $y \in A$ is assigned a part A_i with $i \in \{1, 2, 3, 4\}$ chosen independently, uniformly with probability $1/4$, we have

$$\mathbb{E}[|T \cap A_1 \times A_2 \times A_3 \times A_4|] = \mathbb{E}\sum_{(y_1,\dots,y_4)\in T}\prod_{i=1}^{4} 1_{y_i\in A_i} = \sum_{(y_1,\dots,y_4)\in T}\mathbb{E}\prod_{i=1}^{4} 1_{y_i\in A_i}. \tag{5.15}$$

Let $(y_1, \dots, y_4) \in T$. In particular the y_i are pairwise distinct. Then by uniform distribu-

tion and independence, for any $(m_1, \ldots, m_4) \in [4]^4$, we have

$$\mathbb{P}(y_i \in A_{m_i} \text{ for each } i \in [4]) = 4^{-4}.$$

Together with equation (5.15), this implies that

$$\mathbb{E}[|T \cap A_1 \times A_2 \times A_3 \times A_4|] = |T|/256 \geq C|A|^3/256.$$

In particular, we can pick a partition $A = A_1 \cup A_2 \cup A_3 \cup A_4$ such that the set

$$T' = T \cap A_1 \times A_2 \times A_3 \times A_4$$

satisfies $|T'| \geq C|A|^3/256$. For $i \in [4]$ and $y \in A_i$, set $\xi'_{s+1}(y) = z_i \sum_{j\in[r]} a_{i,j}\xi_j(y)$ where $z_1 = z_2 = 1$ and $z_3 = z_4 = -1$. Observe that $\xi'_{s+1}(y)$ is a nonzero vector in $V_y^\perp$. For $i \in [4]$, let $j_i > s$ be any index such that $a_{i,j_i} \neq 0$. For $y \in A_i$, let $U'_y = \operatorname{span}(\xi_{s+1}(y), \cdots, \widehat{\xi_{j_i}(y)}, \cdots, \xi_r(y))$, where the hat denotes an omitted form; this is a space of dimension at most $r - s - 1$. We have $V_y^\perp = \operatorname{span}(\xi_1(y), \ldots, \xi_s(y), \xi'_{s+1}) + U'_y$. Further, for a quadruple $\mathbf{y} \in T'$, we observe that $(y_i, \xi'_{s+1}(y_i))_{i\in[4]}$ is an additive quadruple. So there are at least $C|A|^3/256$ additive quadruples in the graph $\{(y, \xi'_{s+1}(y)) \mid y \in A\}$. We then invoke Lemma 5.6 to obtain a set $S \subset A$ whose density in A is quasipolynomial (in C), such that ξ'_{s+1} coincides with an affine map on S. This concludes the proof of Lemma 5.15, and hence also that of Proposition 5.13. □

5.6 Remarks on transverse sets

Let V_1 and V_2 be two $\mathbb{F}_p$-vector spaces. Say a set $P \subset V_1 \times V_2$ is *transverse* if it is horizontally and vertically closed, that is, $P \overset{V}{+} P = P \overset{H}{+} P = P$. Write $P_{x\cdot} = \{y \in V_2 \mid (x, y) \in P\}$ and $P_{\cdot y} = \{x \in V_1 \mid (x, y) \in P\}$, borrowing the notation from [35]; these sets are subspaces, or the empty set. Thus

$$P = \bigcup_{x\in V_1} \{x\} \times P_{x\cdot} = \bigcup_{y\in V_2} P_{\cdot y} \times \{y\}. \tag{5.16}$$

Examples of transverse sets are bilinear sets defined in equation (5.6), that is, zero-sets of bilinear forms on cartesian products of vector spaces. A special case of Theorem 5.8 is that a transverse set of density δ contains a rather large bilinear set. However, in that essentially algebraic case, it would be interesting to find a cleaner proof, with better bounds. This is

what we do in this section, in the case where the vertical fibers $P_{x\cdot}$ are hyperplanes. The proof involves a geometric, exact analogue of Lemma 5.14, but here the second alternative of that lemma always holds.

Proposition 5.16. *Suppose that* $\operatorname{codim}_{V_2} P_{x\cdot} \leq 1$ *for any* $x \in V_1$*. Then either there exist* $W \leq V_1$ *and a hyperplane* $H \leq V_2$ *such that* $P = W \times V_2 \cup V_1 \times H$*, or there exists a bilinear form* b *on* $V_1 \times V_2$ *such that* $P = \{(x, y) \in V_1 \times V_2 \mid b(x, y) = 0\}$*.*

In particular, P contains a bilinear set of codimensions $(0, 1, 0)$ or $(0, 0, 1)$.

Proof. Let $P \subset V_1 \times V_2$ be a transverse set. Then $P_{0\cdot}$ contains every $P_{x\cdot}$, because if $y \in P_{x\cdot}$, we have $(x, y) \in P$ and thus $(0, y) \in P$, hence $y \in P_{0\cdot}$. For $x \neq 0$, the set $P_{x\cdot}$ depends only on the class $[x] \in P(V_1) = V_1^*/\mathbb{F}_p^*$ of x in the projective space. Moreover, the stability under horizontal operations is equivalent to the property that if $[z]$ is on the projective line spanned by x and y, we have $P_{z\cdot} \supset P_{x\cdot} \cap P_{y\cdot}$. Because $P \subset V_1 \times P_{0\cdot}$, we may suppose that $V_2 = P_{0\cdot}$.

By hypothesis, for any $x \in V_1$, the fiber $P_{x\cdot}$ is a hyperplane or the full space V_2. Write $P_{x\cdot} = \xi(x)^\perp$ for some vector $\xi(x) \in V_2$ that is defined up to homothety. So $\xi(0) = 0$ and whenever $[z]$ is on the projective line spanned by x and y, we have $\xi(z) \in \operatorname{span}(\xi(x), \xi(y))$. It is easy to see that $\{x \in V_1 \mid P_{x\cdot} = V_2\}$ is a vector subspace which we call W. Furthermore, if $x - y = w \in W$, we have $\xi(x) \in \operatorname{span}(\xi(y), \xi(w)) = \operatorname{span}(\xi(y))$, that is, $\xi(x) = \xi(y)$ up to homothety, so that ξ descends to a map $P(V_1/W) \to P(V_2)$. Let $V_1' = V_1/W$. Thus ξ is a map $P(V_1') \to P(V_2)$ that maps aligned points to aligned points. It is easy to see that on each projective line, the map ξ is either constant or injective. Let us prove that ξ is either constant or injective on all of $P(V_1')$. If $P(V_1')$ is a line, the conclusion is exactly the previous observation, so suppose that $\dim V_1' \geq 3$. Assume that ξ is neither injective nor constant. This means that there exist two distinct points x, y such that $\xi(x) = \xi(y)$, and a third point z satisfying $\xi(z) \neq \xi(x)$. This implies that x, y, z are not (projectively) aligned, so they span a projective plane. The reader may wish to follow the proof on Figure 5.1. Take a point $w \notin \{y, z\}$ on the line spanned by y and z. Because ξ is a bijection on both lines (yz) and (xz), we can find $w' \notin \{x, z\}$ on (xz) such that $\xi(w) = \xi(w') \neq \xi(x)$. Now consider the intersection $u = (ww') \cap (xy)$ in the projective plane $\operatorname{span}(x, y, z)$. Then we have $\xi(u) = \xi(x) = \xi(y) \neq \xi(w)$, so that on the line (ww') the map ξ is neither constant nor injective, a contradiction.

If ξ is constant on $P(V_1')$, we can take $\xi(x)$ to be a nonzero constant vector $\phi \in V_2$ for all $x \in V_1/W$, viewed as the complement of W, while $\xi(x) = 0$ on W. So $P = W \times V_2 \cup V \times \phi^\perp$.

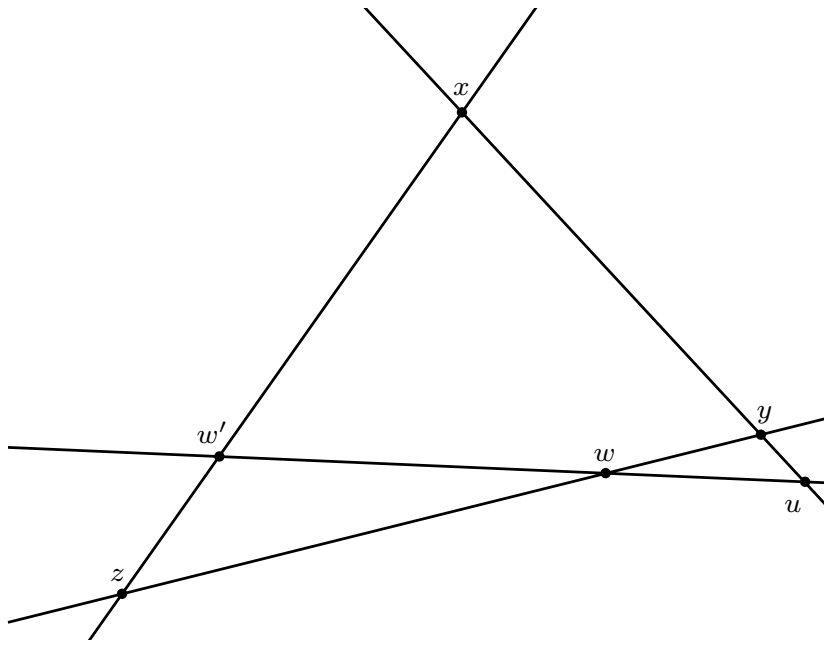

Figure 5.1: Proof of Proposition 5.16.

If ξ is injective and $\dim V_1' \geq 3$, the fundamental theorem [76, Théorème 7] of projective geometry[2] implies that it comes from an injective linear map $V_1' \to V_2$, which we extend to a linear map $\xi : V_1 \to V_2$ of kernel W. Thus P is the zero set of the bilinear form $(x, y) \mapsto \xi(x) \cdot y$, which concludes the proof of Proposition 5.16. $\square$

[2]Here we require the field $\mathbb{F}_p$ to be a prime field; on a non prime finite field $\mathbb{F}_q$, we would need to incorporate Frobenius field automorphisms.

Chapter 6

The Croot-Lev-Pach method and applications

This chapter is based on a note of the author [11] and some unpublished work. It is organised as follows. We introduce the Croot-Lev-Pach method in Section 6.1. Section 6.2 presents a result the author obtained using this method, namely a bound on the cardinality of a set of polynomials free of a certain given configuration. We compare the result obtained in this function field setting to the analogous ones in the integers, where the quantitative aspect is much weaker. We describe some further uses of the method in Section 6.3. In Section 6.4, we exhibit a serious obstacle to applying the method to other open problems of additive combinatorics, namely the problems the existence of arithmetic progressions of length 4 or corners in a given set. Finally in Section 6.5, we discuss the quality of the bounds as a function of the cardinality of the field.

6.1 The Croot-Lev-Pach method

The traditional Fourier-analytic method on $\mathbb{F}_p^n$, sketched in Section 5.1, relies on an expression for the characteristic function $1_{x=a}$, namely

$$1_{x=a} = \mathbb{E}_{t\in\mathbb{F}_p^n}\omega^{(a-x)\cdot t} = \mathbb{E}_{t\in\mathbb{F}_p^n}\omega^{a\cdot t}\omega^{-x\cdot t}, \tag{6.1}$$

where $\omega = \exp(2\pi i/p)$ is still a root of unity. Given that the family of characteristic functions of points $a \in \mathbb{F}_p^n$ is a basis of the space $\mathbb{C}^{\mathbb{F}_p^n}$, equation (6.1) shows that the functions $\chi_t : x \mapsto \omega^{-x\cdot t}$ form another basis of the same space. The coefficients $\hat{f}(t)$ of a

function f in this basis define the Fourier transform $\hat{f}$ of f.

Consequently, a 3-term arithmetic progression (or 3-AP) can be detected by the function

$$1_{x+y=2z} = \mathbb{E}_{t\in\mathbb{F}_p^n}\omega^{(x+y-2z)\cdot t} =: f(x,y,z).$$

Similarly, the basic idea of the Croot-Lev-Pach method [24] can be interpreted (see [84]) as the identity

$$1_{x=a} = \prod_{i=1}^{n}(1-(x_i-a_i)^{p-1}) \tag{6.2}$$

for any x and a in $\mathbb{F}_p^n$. Here, one can replace the prime p by a prime power q.

Equation (6.2), upon expanding the right-hand side, exhibits a basis of $\mathbb{F}_p^{\mathbb{F}_p^n}$, namely the one formed by monomials $\prod_{i=1}^n x_i^{\alpha_i}$ where $\alpha_i \in \{0,\dots,p-1\}$ for all $i \in [n]$. In fact, this identity gives rise to a ring isomorphism

$$\mathbb{F}_p^{\mathbb{F}_p^n} \simeq \mathbb{F}_p[x_1,\dots,x_n]/I,$$

where I is the ideal generated by the polynomials $x_i^p - x_i$ for $i \in [n]$.

As a result, 3-APs are detected by the expression

$$1_{x+y=2z} = \prod_{i=1}^{n}(1-(x_i+y_i-2z_i)^{p-1}) =: f_p(x,y,z). \tag{6.3}$$

While the Fourier-analytic and the Croot-Lev-Pach method start with a clever way to detect an equality, they diverge radically from this point onwards: the first one is $\mathbb{C}$-valued, while the second is $\mathbb{F}_p$-valued. As a result,

$$T(A) = \sum_{(x,y,z)\in A^3} f(x,y,z)$$

is the number of 3-APs in A (including the trivial ones (x,x,x)), whereas

$$\mathbb{F}_p \ni T_p(A) = \sum_{(x,y,z)\in A^3} f_p(x,y,z) \equiv T(A) \bmod p$$

only delivers the number of 3-APs modulo p. Say an AP is *nontrivial* if its three elements are pairwise distinct. The existence of a nontrivial 3-AP in A is equivalent to the statement

$T(A) > |A|$, whereas it cannot be phrased in terms of $T_p(A)$.[1]

Instead, if $A \subset \mathbb{F}_p^n$ is *AP-free* (meaning that it contains no nontrivial 3-APs), the identity (6.3) yields

$$(f_p)|_{A^3}(x, y, z) = \sum_{x \in A} \delta_a(x)\delta_a(y)\delta_a(z) \tag{6.4}$$

where δ_a is the Kronecker symbol (that is $\delta_a(x) = 1_{x=a}$). The crux of the Croot-Lev-Pach method is that in some sense, the left-hand side of (6.4) has "low complexity", being a polynomial of low degree in many variables, while the right-hand side has "high complexity", being a diagonal tensor.

We now make this precise. Because equation (6.2) holds with a prime power q instead of a prime p, we proceed with $\mathbb{F}_q$ instead of the prime field $\mathbb{F}_p$. Take a subset $A \subset \mathbb{F}_q^n$ and a map $P : A^k \to \mathbb{F}_q$. Let $\mathcal{M}$ be the set of functions $\mathbb{F}_q^n \to \mathbb{F}_q$. This set of functions, as already observed, is naturally in bijection with the set of polynomials in $\mathbb{F}_q[t_1, \ldots, t_n]$ in which no indeterminate is raised to a power greater than $q - 1$.

Definition 6.1. A *polynomial cover* for P is a tuple $(M_1, \ldots, M_k) \in \mathcal{M}^k$ such that for each $j \in [k]$ and $p \in M_j$, there exists a function $F_{j,p} : A^{k-1} \to \mathbb{F}_q$ such that for any $(x_1, \ldots, x_k) \in A^k$, we have

$$P(x_1, \ldots, x_k) = \sum_{j \in [k]} \sum_{p \in M_j} p(x_j) F_{j,p}(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k). \tag{6.5}$$

A function of the form $p(x_j)F_{j,p}(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k)$ is called a *slice*.

The *slice rank* $\mathrm{sr}(P)$ of P is the minimum size $\sum_{j \in [k]} |M_j|$ of a polynomial cover. In other words, it is the minimum number of slices required to write P as a linear combination of slices.

Note that this notion is a generalisation of the linear-algebraic notion of rank. Indeed, a square matrix with rows indexed by A can be viewed as a function $f(x, y)$ of two variables in A. It is well known that the rank of a matrix is the minimum number of matrices of the form $g_i(x)h_i(y)$ (a row times a column) that one needs in order to decompose $f(x, y)$ as $f(x, y) = \sum_{i=1}^k g_i(x)h_i(y)$.

Another useful observation is that the slice rank is subadditive, that is, $\mathrm{sr}(P + Q) \leq \mathrm{sr}(P) + \mathrm{sr}(Q)$. Furthermore, the slice rank cannot increase upon restriction, that is, for

[1] Note that $T(A) \not\equiv |A| \bmod p$ is a sufficient, but not necessary, condition for the existence of a nontrivial 3-AP.

$B \subset A$, we have $\mathrm{sr}(P|_B) \leq \mathrm{sr}(P)$. Finally, the trivial upper bound on $\mathrm{sr}(P)$ is $|A|$, as we always have the decomposition

$$P(x_1, \ldots, x_k) = \sum_{a \in A} \delta_a(x_1) P(a, x_2, \ldots, x_k).$$

Inspired by the linear-algebraic case, we now determine the slice rank of diagonal tensors.

Lemma 6.1 (drawn from [84]). *For any $a \in \mathbb{F}_q^n$ let $c_a \in \mathbb{F}_q$ be a coefficient. Then the slice rank of the map defined on $(\mathbb{F}_q^n)^k$ by*

$$f(x_1, \ldots, x_k) = \sum_{a \in \mathbb{F}_q^n} c_a \prod_{i=1}^{k} \delta_a(x_i) \tag{6.6}$$

equals the number of a such that $c_a \neq 0$.

Proof. Let A be the set of $a \in \mathbb{F}_q^n$ such that $c_a \neq 0$. The two-dimensional case ($k = 2$) follows from linear algebra, which is a good starting point for an inductive proof.

Let $k > 2$ and suppose that the lemma holds for $k - 1$. The right-hand side of (6.6) being a sum of $|A|$ functions of rank 1, we immediately have $\mathrm{sr}(P) \leq |A|$. Suppose for the sake of contradiction that $\mathrm{sr}(P) \leq |A| - 1$. By definition, this means that there exists a tuple $(M_1, \ldots, M_k) \in \mathcal{M}^k$ satisfying $\sum_{i=1}^{k} |M_i| \leq |A| - 1$ such that for each $j \in [k]$ and $p \in M_j$, there exists a function $F_{j,p} : A^{k-1} \to \mathbb{F}_q$ such that for any $(x_1, \ldots, x_k) \in A^k$, we have

$$\sum_{a \in \mathbb{F}_q^n} c_a \prod_{i=1}^{k} \delta_a(x_i) = \sum_{j \in [k]} \sum_{p \in M_j} p(x_j) F_{j,p}(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k). \tag{6.7}$$

Consider the set V of functions $h : A \to \mathbb{F}_q$ satisfying

$$\sum_{x \in A} F_{k,p}(x) h(x) = 0$$

for any $p \in M_k$. These $|M_k|$ equations make V a subspace of codimension at most $|M_k|$ of $\mathbb{F}_q^A$, and hence a vector space of dimension $d \geq |A| - |M_k| \geq 1$. Fix a basis of this space and consider the corresponding $d \times |A|$ coordinate matrix. It has rank d, which implies that it contains a $d \times d$ invertible submatrix. This in turn means that there is a subset $A' \subset A$ of cardinality d and a function $h \in V$ such that h does not vanish on A'.

Now if we multiply equation (6.7) by $h(x_k)$ and sum over $x_k \in A$, we obtain

$$\sum_{a \in \mathbb{F}_q^n} c_a h(a) \prod_{i=1}^{k-1} \delta_a(x_i) = \sum_{j \in [k-1]} \sum_{p \in M_j} p(x_j) \widetilde{F}_{j,p}(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_{k-1})$$

for some functions $\widetilde{F}_{j,p}$.

The right-hand side is a sum of $|A| - 1 - |M_k|$ slices, hence it has rank at most $|A| - 1 - |M_k|$. However, by the induction hypothesis and the property of h, the left-hand side has rank at least $|A'| = d \geq |A| - |M_k|$. This is the desired contradiction. □

In contrast, the next lemma is a tool to bound from above the slice rank of polynomials.

Lemma 6.2. *Fix $\epsilon \in (0, 1/2)$. Let $P : (\mathbb{F}_q^n)^k \to \mathbb{F}_q$ be a polynomial in $n \times k$ variables $(x_{j,i})_{j \in [k], i \in [n]}$. Suppose the total degree of P is at most $nk(q-1)(1/2 - \epsilon)$. Then its slice rank is at most $kq^{c(\epsilon,q)n}$ for some constant $c(\epsilon, q) \in (0, 1)$.*

Lemma 6.2 shows the importance of the parameter $\deg P/(nk(q-1))$, which we may call the *normalised degree per variable*. If this parameter is smaller than $1/2$ and bounded away from $1/2$, the polynomial has exponentially small slice rank.

Proof. Since we are interested in P as a function on $(\mathbb{F}_q^n)^k$, we reduce it modulo the ideal I generated by the polynomials $x_{j,i}^{q-1} - x_{j,i}$ for $i \in [n]$ and $j \in [k]$. We continue to use P for the only polynomial in the class P modulo I that has degree at most $q-1$ in each variable $x_{j,i}$. Further, for any integer $d \geq 0$, we denote by $\mathcal{M}_{d,n}$ the set of monomials in n variables of degree at most $q-1$ in each variable and at most d in total.

The polynomial P is a sum of monomials of the form

$$p(x_1, \ldots, x_k) = \prod_{j \in [k]} p_j(x_{j,1}, \ldots, x_{j,n}),$$

where each p_j is a monomial in n variables. For each monomial p, by the pigeonhole principle, there exists $j \in [k]$ such that

$$\deg p_j \leq (\deg P)/k \leq n(q-1)(1/2 - \epsilon) =: d.$$

In other words, $p_j \in \mathcal{M}_{d,n}$. We infer from the data above that there exist sets of monomials

$M_1, \ldots, M_k \subset \mathcal{M}_{d,n}$ and functions $F_{j,p}$ for $(j,p) \in [k] \times M_j$ such that

$$P = \sum_{j=1}^{k} \sum_{p \in M_j} p(f_j) F_{j,p}(f_1, \ldots, f_{j-1}, f_{j+1}, \ldots, f_k).$$

Now $|\mathcal{M}_{d,n}|/q^n$ may be interpreted as the probability that the sum of n independent, uniform random variables on $\{0, \ldots, q-1\}$ is at most d. To bound this probability, we use Hoeffding's concentration inequality [2, Theorem A.1.16], which implies that

$$|\mathcal{M}_{d,n}| = |\mathcal{M}_{(q-1)n(1/2-\epsilon),n}| \leq q^n e^{-\frac{n\epsilon^2}{2}} = q^{c(\epsilon,q)n},$$

where $c(\epsilon, q) = (1 - \frac{\epsilon^2}{2\log q}) \in (0,1)$. This implies that the slice rank of P is at most $kq^{c(\epsilon,q)n}$, and concludes the proof of Lemma 6.2. □

6.2 Application to the solubility of polynomial equations in function fields

As noted above, the notion of slice rank was originally devised for and applied to the polynomial representation of the indicator function of 3-APs [84]. Indeed, the AP-freeness of a set A yields a functional equality between a diagonal tensor (6.4) and a polynomial (6.3) of degree $n(q-1)$ in $3 \times n$ variables. A direct application of Lemmas 6.1 and 6.2 then gives the celebrated result of Ellenberg and Gijswijt [26, Theorem 4], which we now state.

Theorem 6.3. *Let $A \subset \mathbb{F}_q^n$ have no non trivial 3-AP. Then $|A| \leq q^{(c_q+o(1))n}$ for some $c_q \in (0,1)$.*

Ellenberg and Gijswijt found that c_3 could be taken such that $3^{c_3} \approx 2.756$. We discuss the dependence of c_q on q in Section 6.5.

As the author pointed out in [11], there are much more general equations of interest to which this method can be fruitfully applied, including polynomial equations in sufficiently many variables. We now discuss the general set-up.

Let R be a ring and $a_1, \ldots, a_k$ be elements of R which sum to 0, i.e. $\sum_{i=1}^{k} a_i = 0$. Then the equation

$$\sum_{i=1}^{k} a_i f_i^r = 0 \tag{6.8}$$

possesses a wealth of trivial solutions $(f_1, \ldots, f_k)$, namely constant tuples $(f, \ldots, f)$, even though it is not necessarily a translation-invariant equation. This suggests that if a subset $A \subset R$ is free of nontrivial solutions, then it should be small. For the ring $R = \mathbb{Z}$ and $r = 2$, this question was studied first by Smith [81], Keil [56] and Henriot [53]; they replaced the single equation by a system comprising the initial equation and an auxiliary linear equation in order to ensure invariance under translation and dilation. Recently, Browning and Prendiville [18] showed, without using the auxiliary equation, that if $k \geq 5$, and $A \subset [N]$ satisfies $|A| \gg N$ and N is large enough, then equation (6.8) necessarily admits nontrivial solutions $(f_1, \ldots, f_k)$ in A^k. Their method relies on the transference principle. Further, Chow [21] proved that any relatively dense subset of the primes contains a solution to any equation of the form (6.8), as long as $k \geq r^2 + 1$.

Similarly, one may ask whether any dense subset A of the ring $R = \mathbb{F}_q[t]$ is bound to contain a nontrivial solution to (6.8). In this chapter, we answer the question under a natural condition on the number of variables, namely $k \geq 2r^2 + 1$. In the function field setting, the polynomial method of Croot, Lev and Pach [24] can be fruitfully applied and delivers much stronger bounds than any method known in the integers. This was already noticed by Green [42] in the case of Sarkőzy's theorem.

We now give a precise statement of this chapter's main theorem. We fix a prime power q and write $G_{q,n}$ for the set of polynomials of degree strictly less than n over $\mathbb{F}_q$, so that $|G_{q,n}| = q^n$.

Theorem 6.4. *Let r, k and d be integers satisfying $k \geq 2r^2+1$. Suppose that $a_1, \ldots, a_k$ are polynomials over $\mathbb{F}_q$ of degree at most d satisfying $\sum_{i=1}^k a_i = 0$. Then there exist constants $0 < c(r,q) < 1$ and $C = C(d,r,q)$ such that any $A \subset G_{q,n}$ satisfying $|A| \geq kCq^{c(r,q)n}$ must contain a nontrivial solution to equation* (6.8).

The aforementioned paper of Chow [21] implies that $k \geq r^2 + 1$ is sufficient in the integers, but the bound on the size of A obtained by his analytic method is much weaker (we get a power saving, in contrast to his logarithmic saving).

The starting point of the proof of Theorem 6.4 is to view the map $f \mapsto f^r$ as a polynomial map in the coefficients of f. A map $\Phi : (\mathbb{F}_q^n)^k \to \mathbb{F}_q^m$ is said to *vanish on the diagonal* if $\Phi(f, \ldots, f) = 0$ for any f. A set $A \subset \mathbb{F}_q^n$ is called Φ-*free* if for any $(f_1, \ldots, f_k) \in A^k$, the equality $\Phi(f_1, \ldots, f_k) = 0$ holds if and only if $(f_1, \ldots, f_k) = (f, \ldots, f)$ for some $f \in \mathbb{F}_q^n$.

We reduce Theorem 6.4 to the following proposition, which is then tractable by the polynomial method of Croot-Lev-Pach.

Proposition 6.5. *For any $\epsilon \in (0, 1/2)$, there exists a constant $c'(\epsilon, q) \in (0, 1)$ such that the following holds. Let $\Phi : (\mathbb{F}_q^n)^k \to \mathbb{F}_q^m$ be a polynomial map of degree at most ℓ (i.e. each coordinate is a polynomial of degree at most ℓ) that vanishes on the diagonal. Suppose that $A \subset \mathbb{F}_q^n$ is Φ-free. Finally, suppose that $m\ell/k \leq (1/2 - \epsilon)n$. Then $|A| \leq kq^{c'(\epsilon,q)n}$.*

We prove that Proposition 6.5 implies Theorem 6.4. Each polynomial $f = \sum_{i=0}^{n-1} f_i t^i$ in $G_{q,n}$ can be viewed as a vector $\overrightarrow{f} = (f_0, \ldots, f_{n-1}) \in \mathbb{F}_q^n$. Now $f^r \in G_{q,(n-1)r+1}$ so we view it as the vector

$$\overrightarrow{f^r} = (f_0^r, r f_0^{r-1} f_1, \ldots, f_{n-1}^r) \in \mathbb{F}_q^{(n-1)r+1}.$$

We notice that $\overrightarrow{f^r} = Q(\overrightarrow{f})$, where Q is a polynomial map of degree r. Similarly, if $a \in \mathbb{F}_q[t]$ of degree at most d, we see that $f \mapsto \overrightarrow{af} \in \mathbb{F}_q^{n+d}$ is a polynomial map $\mathbb{F}_q^n \to \mathbb{F}_q^{n+d}$ (of degree 1). Thus,

$$\Phi : (f_1, \ldots, f_k) \mapsto \sum_{i=1}^{k} a_i f_i^r$$

induces a polynomial map of degree r

$$\overrightarrow{\Phi} : (\mathbb{F}_q^n)^k \to \mathbb{F}_q^m$$

where $m = (n-1)r + d + 1$ and

$$\overrightarrow{\Phi}(\overrightarrow{f_1}, \ldots, \overrightarrow{f_k}) = \overrightarrow{\Phi(f_1, \ldots, f_k)}.$$

We observe that if $A \subset G_{q,n}$ does not contain any nontrivial solution to (6.8), the set $\overrightarrow{A} = \{\overrightarrow{f} \mid f \in A\} \subset \mathbb{F}_q^n$ contains only trivial solutions $(\overrightarrow{f}, \ldots, \overrightarrow{f})$ to the equation $\overrightarrow{\Phi}(\overrightarrow{f_1}, \ldots, \overrightarrow{f_k}) = 0$.

Moreover, given that $k \geq 2r^2 + 1$, we have

$$\frac{mr}{kn} = \frac{(n-1)r^2 + dr + r}{kn} \leq \frac{r^2}{2r^2+1} + \frac{(d+1)r}{(2r^2+1)n}.$$

Hence, if $n \geq 4(d+1)r$, we have $mr/k \leq (1/2 - \epsilon)n$, where

$$\epsilon = \epsilon(r) = \frac{1}{4(2r^2+1)} \in (0, 1/2).$$

We can now apply Proposition 6.5 and obtain $|A| \leq kq^{c(r,q)n}$ for some constant $c(r, q) = c'(\epsilon(r), q) \in (0, 1)$. Taking care separately of the small values of n, one can find a constant

$C(d,r,q) \leq q^{4(d+1)r}$ such that the bound

$$|A| \leq kC(d,r,q)q^{c(\epsilon,q)n}$$

is valid for all n.

We now prove Proposition 6.5. As we did in equation (6.4) for 3-APs, we transform the Φ-freeness into a functional identity

$$\forall (f_1,\ldots,f_k) \in A^k \quad \prod_{i=1}^{m}(1-\Phi_i^{q-1}(f_1,\ldots,f_k)) = \sum_{f\in A}\prod_{j=1}^{k}\delta_f(f_j), \tag{6.9}$$

where $\delta_f(f_j)$ is 0 if $f \neq f_j$ and 1 otherwise. The proposition then follows from Lemmas 6.1 and 6.2.

6.3 Further applications

The method described above is rather versatile. We provide a few further examples.

Proposition 6.6. *Let A and V be two subsets of $\mathbb{F}_q^n$ and suppose that $0 \in V$. Let $\epsilon \in (0,1)$ and suppose that one of the following two hypotheses holds.*

1. *A does not contain two elements $x \neq y$ whose difference is in V and there exists a polynomial of degree at most $(1-\epsilon)n(q-1)$ in $\mathbb{F}_q[x_1,\ldots,x_n]$ whose support is V.*
2. *A does not contain any 3-AP of step $d \in V$ and there exists a polynomial of degree at most $(1/2-\epsilon)n(q-1)$ whose support is V.*

Then $|A| \leq q^{cn}$ for some $c = c(\epsilon,q) \in (0,1)$.

Proof. Under the first hypothesis, we have

$$\forall (x,y) \in A^2 \quad P(x-y) = P(0)\sum_{a\in A}\delta_a(x)\delta_a(y),$$

an equality between a polynomial in $2\times n$ variables of degree at most $(1-\epsilon)n(q-1)$ and a diagonal matrix. Under the second one we have

$$\forall (x,y,z) \in A^2 \quad f_q(x,y,z)P(y-x) = P(0)\sum_{a\in A}\delta_a(x)\delta_a(y),$$

where $f_q(x, y, z)$ is the indicator function of 3-APs viewed as in (6.3). This is an equality between a polynomial in $3 \times n$ variables of degree at most $(3/2 - \epsilon)n(q-1)$ and a diagonal tensor. Lemmas 6.1 and 6.2 conclude the proof. $\square$

Interesting examples of sets that are the supports of polynomials of low degree include $V = S^n$ for any subset $S \subset \mathbb{F}_q$ (the polynomial $\prod_{i\in[n]} \prod_{j\in\mathbb{F}_q\setminus S}(x_i - j)$ has support V and degree $n(q-|S|)$) and subspaces of low codimension. Indeed, if V is defined as the zero-set of k linear forms $\ell_1, \ldots, \ell_k$, its indicator function is $\prod_{i=1}^{k}(1 - \ell_i(x)^{q-1})$, a polynomial of degree $k(q-1)$.

A more refined example was supplied by Green [42]: he proved that the set of squares in $G_{q,n}$ (here again, we view a polynomial $f \in G_{q,n}$ as the vector of its coefficients in $\mathbb{F}_q^n$) is the support of a polynomial of degree at most $3/4n(q-1)$. More generally, he found that the set of kth powers is the support of a polynomial of degree at most $(1 - 1/k^2)n(q-1)$. This allowed him to improve Sarkőzy's theorem in function fields: a set $A \subset G_{q,n}$ that does not contain any pair of elements differing by a kth power must be exponentially small.

6.4 Limits of the method

At present, the method seems unable to bound the size of a set free of any nontrivial 4-term arithmetic progressions (or 4-APs) in $\mathbb{F}_q^n$, a problem we refer to as the 4-APs problem. We fix $q = 5$, the smallest cardinality which allows genuine 4-APs. The indicator function of 4-APs is

$$Q(x, z, y, w) = 1_{x+y=2z}1_{z+w=2y} = \prod_{i=1}^{n}(1 - (x_i + y_i - 2z_i)^4)\prod_{i=1}^{n}(1 - (z_i + w_i - 2y_i)^4), \quad (6.10)$$

a polynomial of degree $2n(p-1)$, in $4n$ variables, whence a normalised degree per variable of exactly $1/2$. Consequently the simple pigeonhole principle applied in Lemma 6.2 to bound the slice rank does not work.

Another well-known problem in additive combinatorics, already mentioned in Section 1.4, is the corners problem. It asks for a bound on the size of a set $A \subset \mathbb{F}_2^n \times \mathbb{F}_2^n$ which contains no nontrivial *corner*. Here a corner is a triple of the form $\{(x, y), (x + d, y), (x, y + d)\}$, and it is said to be nontrivial if $d \neq 0$. Equivalently, a corner is a triple $\{(x, y), (u, v), (z, w)\}$ with $x = z, y = v, x + u = y + w$. The indicator function of corners

is therefore given by

$$P((x,y),(u,v),(z,w)) = \prod_{i\in[n]} f((x_i,y_i),(u_i,v_i),(z_i,w_i)) \tag{6.11}$$

where

$$f((x,y),(u,v),(z,w)) = (1+x+z)(1+y+v)(1+x+u+y+w), \tag{6.12}$$

again a polynomial of normalised degree per variable equal to 1/2.

Our aim is to find a parameter that shows that the corners and 4-APs problems are genuinely harder than the 3-APs problem. Although it is reasonable to believe that both polynomials (6.11) and (6.10) have maximal slice ranks, that is, respectively 4^n and 5^n, proving lower bounds for the slice rank is beyond current known techniques. Instead of the slice rank, we propose the following new parameter, which we call *monomial slice rank*.

Definition 6.2. Let P be a function (that is, a polynomial) from $(\mathbb{F}_p^n)^k \to \mathbb{F}_p$. Let $\mathcal{N}$ be the set of monomials in $\mathbb{F}_p[x_1,\dots,x_n]$ that have degree at most $p-1$ in each variable. Elements of $\mathcal{N}$ are called *elementary monomials*. A *monomial cover* for P is a tuple $(N_1,\dots,N_k) \in \mathcal{N}^k$ such that for each $j \in [k]$ and $p \in N_j$, there exists a function $F_{j,p}$ from A^{k-1} to $\mathbb{F}_q$ such that for any $(x_1,\dots,x_k) \in A^k$, we have

$$P(x_1,\dots,x_k) = \sum_{j\in[k]} \sum_{p\in N_j} p(x_j) F_{j,p}(x_1,\dots,x_{j-1},x_{j+1},\dots,x_k). \tag{6.13}$$

A function of the form $p(x_j)F_{j,p}(x_1,\dots,x_{j-1},x_{j+1},\dots,x_k)$ is called a *monomial slice*.

The *monomial slice rank* $\mathrm{msr}(P)$ of P is the minimum size $\sum_{j\in[k]}|N_j|$ of a monomial cover. In other words, it is the minimum number of slices required to write P as a linear combination of monomial slices.

It is obvious that $\mathrm{sr}(P) \le \mathrm{msr}(P)$. We observe that in Lemma 6.2, we actually bounded $\mathrm{msr}(P)$ rather than $\mathrm{sr}(P)$. In contrast, we show below that the monomial slice ranks of the polynomials (6.10) and (6.11) are as large as they can be.

Proposition 6.7. *The monomial slice rank of the indicator function P of corners in $\mathbb{F}_2^n \times \mathbb{F}_2^n$ is 4^n, and that of the indicator function Q of 4-APs in $\mathbb{F}_5^n$ is 5^n.*

Although this statement does not say anything about polynomial covers, it shows that the corner and 4-APs problems are, in some sense, genuinely harder than the 3-AP problem.

We give a detailed proof for the corners problem only, as it is computationally simpler but contains all the ideas needed for the other problem.

Proof. Recall that the polynomials P and f are defined by equations (6.11) and (6.12). In fact, we are interested in the polynomial map $(\mathbb{F}_2 \times \mathbb{F}_2)^3 \to \mathbb{F}_2$ induced by f, and this map depends only on the class of f in the quotient $R = \mathbb{F}_2[x, y, u, v, z, w]/I$, where I is the ideal generated by the polynomials $t^2 + t$ for $t \in \{x, y, u, v, z, w\}$. Upon expanding and reducing modulo I, we find that f is the following sum of 33 monomials[2]

$$\begin{aligned} xyz + xyw + yzw + xyu + yzu + xyv + xzv + yzv + xwv + zwv + xuv + zuv \\ + xy + xz + yz + xw + yw + zw + xu + yu + zu + xv + yv + zv + wv + uv \\ + x + y + z + w + u + v + 1. \end{aligned} \tag{6.14}$$

Let $(\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3)$ be a monomial cover for P, that is, a triple of sets of monomials in $2n$ indeterminates such that for each $\ell \in [3]$ and $m \in \mathcal{M}_\ell$, there exists a polynomial $F_{\ell,m} : (\mathbb{F}_2^n \times \mathbb{F}_2^n)^2 \to \mathbb{F}_2$ so that

$$\begin{aligned} P = \sum_{m \in \mathcal{M}_1} m(x, y) F_{1,m}((z, w), (u, v)) + \sum_{m \in \mathcal{M}_2} m(z, w) F_{2,m}((u, v), (x, y)) \\ + \sum_{m \in \mathcal{M}_3} m(u, v) F_{3,m}((x, y), (z, w)). \end{aligned} \tag{6.15}$$

Call the three sums featuring in equation (6.15) P_1, P_2, P_3. We say that a monomial appears in a polynomial if its coefficient in the expansion of the polynomial is nonzero. Let S_n be the set of monomials that appear in P. Equation (6.11) provides a natural bijection between S_n and $S^n = S \times \cdots \times S$, where S is the set of monomials appearing in f, given in the expansion (6.14); in particular $S_1 = S$. Each $m \in S_n$ appears either in exactly one P_i or in all three. And if a monomial does not appear in P, it appears either in none or exactly two of the P_i. However, we can assume that any monomial appears exactly once or not at all. Indeed, suppose that a monomial $m_1(x, y)m_2(z, w)m_3(u, v)$ appears at least twice. We may assume without loss of generality that $m_1 \in \mathcal{M}_1$ and $m_2 \in \mathcal{M}_2$ and that $m_2(z, w)m_3(u, v)$ appears in F_{1,m_1} and that similarly $m_1(z, w)m_3(u, v)$ has a nonzero coefficient in F_{2,m_2}. Then replacing F_{1,m_1} by $F_{1,m_1} + m_2(z, w)m_3(u, v)$ and

[2]The corresponding polynomial for the 4-APs problem has 96 polynomials, whence our choice to focus on the corners problem only.

F_{2,m_2} by $F_{2,m_2} + m_1(z,w)m_3(u,v)$, we remove the redundancy, without affecting $\mathcal{M}_1$ nor $\mathcal{M}_2$.

We say that $p \in S_n$ is *covered* by $m \in \mathcal{M}_i$ if it appears in P_i. Thus each $p \in S_n$ is covered by exactly one monomial, from a single $\mathcal{M}_i$.

We shall show that $|\mathcal{M}_1|+|\mathcal{M}_2|+|\mathcal{M}_3| \geq 4^n$, but first we introduce one more definition.

Definition 6.3. Let $K = \{m^{(1)}, \ldots, \ldots, m^{(k)}\}$ be a set of monomials from S_n. Let $X_i = (x_i, y_i)$ and $Y_i = (z_i, w_i)$ as well as $Z_i = (u_i, v_i)$. Then let $X = (X_1, \ldots, X_n)$ and let Y and Z be defined analogously. Write $m^{(i)}(X,Y,Z) = a^{(i)}(X)b^{(i)}(Y)c^{(i)}(Z)$. We say that K is a *monomial matching* if the maps $i \mapsto a^{(i)}(X)$ as well as $i \mapsto b^{(i)}(Y)$ and $i \mapsto c^{(i)}(Z)$ are injective. Occasionally, we will refer to the set of triples $\{(a^{(i)}, b^{(i)}, c^{(i)})\}$ as a matching.

It is easy to see that a monomial cover and a monomial matching correspond to a vertex cover and a matching, respectively, in some hypergraph derived from P. We also observe that a monomial cover has to be at least large as any matching. Indeed, an elementary monomial $m \in \mathcal{M}_i$ can cover at most one monomial from any given matching.

We remark that if $K_1 \subset S_1$ is a monomial matching, then a monomial matching $K_n \subset S_n$ in bijection with K_1^n can be constructed[3]. Indeed, for any sequence $\mathbf{m} = (m_1, \ldots, m_n) \in K_1^n$, take

$$d_{\mathbf{m}}(X,Y,Z) = \prod_{i=1}^{n} m_i(X_i, Y_i, Z_i) = a_{\mathbf{m}}(X)b_{\mathbf{m}}(Y)c_{\mathbf{m}}(Z).$$

To check that this contruction gives rise to a matching, take two distinct sequences $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ and $\mathbf{m} = (m_1, \ldots, m_n)$ in K_1^n. We need to prove that $d_{\mathbf{m}} \neq d_{\boldsymbol{\mu}}$. Let $i \in [n]$ be such that $m_i \neq \mu_i$. Write $m_i(X_i, Y_i, Z_i) = a_i(X_i)b_i(Y_i)c_i(Z_i)$ and $\mu_i(X_i, Y_i, Z_i) = \alpha_i(X_i)\beta_i(Y_i)\gamma_i(Z_i)$. By definition of a matching, we have simultaneously $a_i \neq \alpha_i$ and $b_i \neq \beta_i$ and $c_i \neq \gamma_i$. As a result, $a_{\mathbf{m}} \neq a_{\boldsymbol{\mu}}$ and $b_{\mathbf{m}} \neq b_{\boldsymbol{\mu}}$ and $c_{\mathbf{m}} \neq c_{\boldsymbol{\mu}}$.

It therefore suffices to construct a matching of size 4 in $S = S_1$. It is easy to see that $\{zwv, xuv, yzu, xyw\}$ or $\{(1, zw, v), (x, 1, uv), (y, z, u), (xy, w, 1)\}$ will work. Finally, this gives rise to a monomial matching of size 4^n in P, which forces a monomial cover to have size at least 4^n. □

Concerning the 4-APs problem, the set

$$\{(1, y^4, 1, w^4), (x, y^3, z, w^3), (x^2, y^2, z^2, w^2), (x^3, y, z^3, w), (x^4, 1, z^4, w)\}$$

[3]One can recognise here the notion of tensor power of a hypergraph.

is a monomial matching of size 5 for the indicator polynomial (for $n = 1$) shown at equation (6.10). Because the polynomial has 96 monomials and thus can hardly be handled manually, we employed a computer algorithm to find it.

6.5 Evolution of the Ellenberg-Gijswijt bound with the cardinality

To bound the (monomial) slice rank of a polynomial (Proposition 6.2), we used a rather crude tool, namely a Chernoff-type bound. By doing so, we obtained a bound of the form q^{cn}, for some $c < 1$, which was not the best exponent the Croot-Lev-Pach method affords. Instead, Ellenberg and Gisjwijt used Cramér's theorem [23] in large deviation theory, but as pointed out by Tao [84], even an elementary analysis based on Stirling's approximation provides the optimal exponent c the method can yield. In this section we examine how this exponent varies for large q. Let us first state the Ellenberg-Gijswijt result [26].

Proposition 6.8. *For any prime power q, there is a constant $h_q < \log q$ such that the following holds. Let $a_{q,n}$ be the maximal size of a 3-AP-free set $A \subset \mathbb{F}_q^n$. Then*

$$\limsup_{n\to+\infty} \frac{\log a_{q,n}}{n} \leq h_q.$$

Furthermore, h_q can be taken to be the maximum of the entropy function

$$h(\alpha_0, \ldots, \alpha_{q-1}) = -\sum_{i=0}^{q-1} \alpha_i \log \alpha_i \tag{6.16}$$

subject to the constraints

$$\forall i \in [0, q-1] \quad \alpha_i \geq 0 \tag{6.17}$$

$$\sum_{i=0}^{q-1} \alpha_i = 1 \tag{6.18}$$

$$\sum_{i=0}^{q-1} i\alpha_i \leq \frac{q-1}{3}. \tag{6.19}$$

The constant c_q in Theorem 6.3 is related to h_q by $\exp(h_q) = q^{c_q}$. We do not include a proof here (see [84]), but simply observe that the entropy function being continuous and the domain defined by the constraints compact, a maximum does exist. In fact, the function

being strictly concave and the domain convex, this maximum is unique. Our result is the following.

Proposition 6.9. *As q tends to infinity, we have*

$$\exp h_q = q\gamma_1 + o(q)$$

where $\gamma_1 > 0$ is a constant approximately equal to 0.84.

We point out that, independently, Zeilberger in his survey [94] produced an implicit expression for h_q and supplied numerical values for not too large q. He did not compute the asymptotic, however. In that survey, one can see that although $\mathbb{F}_{p^k}^n$ is isomorphic to $\mathbb{F}_p^{kn}$ as a group, the Ellenberg-Gijswijt bound is worse in $\mathbb{F}_{p^k}^n$ than in $\mathbb{F}_p^{kn}$. In fact, c_q is an increasing function of q. For instance, the Ellenberg-Gisjwijt bound for $\mathbb{F}_3^{2n}$ is $2.756^{2n} = 7.596^n$, instead of 7.846^n for $\mathbb{F}_9^n$.

Proof. The maximum c_q is clearly not attained at a point where one coordinate is 0, because the slope of the function $x \mapsto -x \log x$ at 0 is $+\infty$, while it is finite everywhere else. The constrained extrema theorem [58, Section 2.6.2] implies that such a maximum is attained at a point $(\alpha_0, \ldots, \alpha_{q-1})$ satisfying

$$\nabla h = -(\log \alpha_0 + 1, \ldots, \log \alpha_{q-1} + 1) \in \operatorname{span}((1, \ldots, 1), (0, 1, \ldots, q-1)). \tag{6.20}$$

In other words, at the maximum, there exists a pair $(a_q, b_q) \in \mathbb{R}^2$ such that $\alpha_i = \exp(1 - a_q - ib_q)$ for all $i \in \{0, \ldots, q-1\}$. If $b_q = 0$, then α_i is constant, and equal to $1/q$ because of constraint (6.18). This is not compatible with constraint (6.19). Therefore $b_q \neq 0$, that is, constraint (6.19) is active too, in the sense that

$$\sum_{i=0}^{q-1} i\alpha_i = \frac{q-1}{3}. \tag{6.21}$$

Let $d_q = \exp(-b_q) < 1$. Inserting $\alpha_i = \exp(1 - a_q - ib_q) = \exp(1 - a_q)d_q^i$ in the constraints (6.18) and (6.21), we get

$$\exp(1 - a_q) \sum_{i=0}^{q-1} d_q^i = 1 \tag{6.22}$$

and

$$\exp(1 - a_q) \sum_{i=0}^{q-1} i d_q^i = \frac{q-1}{3}. \tag{6.23}$$

Combining equations (6.22) and (6.23) to eliminate a_q, we obtain

$$\sum_{i=0}^{q-1} i d_q^i = \frac{q-1}{3} \sum_{i=0}^{q-1} d_q^i. \tag{6.24}$$

The left-hand side of the last equality can be computed as

$$\frac{((q-1)d_q - q)d_q^q + d_q}{(d_q-1)^2}.$$

Multiplying both sides of (6.24) by $d_q - 1$ and summing the geometric series, we obtain

$$((q-1)d_q - q)d_q^q + d_q = \frac{(q-1)(d_q-1)}{3}(d_q^q - 1),$$

which upon rearranging yields

$$2(q-1)d_q^{q+1} - (2q+1)d_q^q + 3d_q + (q-1)d_q = q-1. \tag{6.25}$$

Let us prove by contradiction that $d_q \to 1$ as $q \to \infty$. Assuming the contrary, we have $\liminf d_q < 1$ and $\liminf d_q^q = 0$. Let s_q be the left-hand side of equation (6.25). Then $\liminf s_q/(q-1) = \liminf d_q < 1$, which is a contradiction to $s_q = q-1$, the statement of equation (6.25). To gain more precise information, let us write $d_q = 1 - \epsilon_q$, where $0 < \epsilon_q \to 0$ as $q \to +\infty$. We rewrite equation (6.25) as

$$q(1-d_q)(1+2d_q^q) = (1-d_q^q)(2d_q+1). \tag{6.26}$$

One can prove by contradiction that $q\epsilon_q$ cannot tend to 0. Indeed, supposing it does, one can write

$$d_q^q = \exp(q\log(1-\epsilon_q)) = \exp(-q\epsilon_q)\exp(O(q\epsilon_q^2)) = 1 - q\epsilon_q + (q\epsilon_q)^2/2 + o(q\epsilon_q)^2.$$

Inserting this equality in (6.26), we find

$$q\epsilon_q(3 - 2q\epsilon_q + (q\epsilon_q)^2 + o(q\epsilon_q)^2) = (q\epsilon_q - (q\epsilon_q)^2/2 + o(q\epsilon_q)^2)(3 - 2\epsilon_q),$$

and comparing the terms in $(q\epsilon_q)^2$ we obtain the desired contradiction.

Yet $q\epsilon_q$ needs to have a limit point $\beta \in [0, +\infty]$. In fact, the above reasoning shows

that $\beta \neq 0$. Since $d_q^q = \exp(q \log(1-\epsilon_q))$, we have $d_q^q \to \exp(-\beta)$ as q tends to $+\infty$ along some subsequence, with the obvious convention that $\exp(-\infty) = 0$. Passing to the limit along that subsequence in equation (6.26), one obtains

$$\beta(1 + 2\exp(-\beta)) = 3(1 - \exp(-\beta)). \tag{6.27}$$

A quick study of the function $\beta \mapsto \beta(1+2\exp(-\beta)) - 3(1-\exp(-\beta))$ shows that, besides 0, there is a unique solution to equation (6.27), which a computer reveals to be $\beta \approx 2.149$. Thus $d_q^q \to \exp(-\beta) \approx 0.11$.

We now determine asymptotics for a_q and b_q. By definition, $b_q = -\log d_q = -(\log d_q^q)/q = \beta/q + o(1/q)$. To find a_q, we use equation (6.22) and obtain

$$\begin{aligned} a_q &= 1 + \log \frac{1 - d_q^q}{1 - d_q} \\ &= 1 + \log \frac{1 - \exp(-\beta) + o(1)}{\epsilon_q} \\ &= \log q + \log \frac{1 - \exp(-\beta)}{\beta} + 1 + o(1). \end{aligned}$$

Then the maximum of the entropy function, according to equation (6.16) and the formula $\alpha_i = \exp(1 - a_q - ib_q)$, is

$$\begin{aligned} h_q &= -\sum_{i=0}^{q-1} \alpha_i \log \alpha_i \\ &= -\sum_{i=0}^{q-1} \alpha_i (1 - a_q - ib_q) \\ &= a_q - 1 + b_q \frac{q-1}{3} \\ &= \log q + \log \frac{1 - \exp(-\beta)}{\beta} + \beta/3 + o(1). \end{aligned}$$

Exponentiating, we find that $\exp(h_q) = \gamma_1 q + o(q)$ where

$$\gamma_1 = \frac{1 - \exp(-\beta)}{\beta} \exp(\beta/3)$$

and $\beta > 0$ was defined in equation (6.27). This concludes the proof of Proposition 6.9. $\square$

Chapter 7

Linear and quadratic uniformity of the Möbius function over $\mathbb{F}_q[t]$

This chapter is based on a paper submitted to the Proceedings of the Cambridge Philosophical Society [13], written with Thai Hoàng Lê. Except Section 7.3, most of the mathematics and the writing was done by the author.

The motivation for this chapter is the quest for asymptotics of linear configurations of irreducible polynomials. For instance, as an analogue of Theorem 1.2, we would like to estimate the number of k-term arithmetic progressions of irreducible polynomials of degree less than n over $\mathbb{F}_q$ as n tends to infinity, while $q = p^s$ is fixed and p is a prime larger than k and $s \geq 1$. Whereas for $k = 3$, the method used by Hayes [51] to solve Goldbach's ternary problem over function fields applies, the problem remains open for $k > 3$ (but lower bounds are known [60]).

Just like in the integers (see Section 2.5), one can reduce this problem to a different one involving the Möbius function on $\mathbb{F}_q[t]$, namely the problem of the *uniformity* of this function. The Möbius function on $\mathbb{F}_q[t]$ is defined, like its counterpart in the integers, by

$$\mu(f) = \begin{cases} (-1)^k & \text{where } k \text{ is the number of monic irreducible factors of } f, \text{ if } f \text{ is squarefree,} \\ 0 & \text{otherwise.} \end{cases}$$

Because q is fixed, we drop the subscript q and write G_n for $G_{q,n}$. The desired uniformity property consists in a bound for the Gowers norm of μ of the form

$$\|\mu\|_{U^k(G_n)} = o_{n\to\infty}(1).$$

The Gowers uniformity norms of a function defined on a vector space is defined analogously to the ones of a function on the integers, thus

$$\|g\|_{U^k(G_n)} = \left(\mathbb{E}_{x \in G_n} \mathbb{E}_{h \in (G_n)^k} \prod_{\omega \in \{0,1\}^k} \mathcal{C}^{|\omega|} g(x + \omega \cdot h) \right)^{2^{-k}}.$$

One can check that this is consistent with the definition of the U^2-norm from Chapter 5.

Note that these norms depend solely on the additive group structure of G_n, which is isomorphic to $\mathbb{F}_q^n$. Besides, there exists a group isomorphism $\phi : \mathbb{F}_q \to \mathbb{F}_p^s$, which makes $\mathbb{F}_q$ an s-dimensional $\mathbb{F}_p$-vector space. This map induces another map $\phi_n : \mathbb{F}_q^n \to \mathbb{F}_p^{sn}$, which is an $\mathbb{F}_p$-linear isomorphism, giving rise to a group isomorphism $G_n \cong \mathbb{F}_p^{sn}$. On the other hand, the Möbius function relies on the ring structure of $\mathbb{F}_q[t]$, which depends on the base field $\mathbb{F}_q$. Thus the question of the uniformity of the Möbius function showcases the clash of multiplicative and additive structures, one of the central themes of arithmetic combinatorics.

Analogously to the situation over the integers discussed in Section 2.5, it turns out that non-uniformity on $\mathbb{F}_p^n$ is characterised by the existence of a significant correlation with a certain family of structured functions, namely polynomial phases. This is the content of the inverse theorem for the Gowers norms [85] by Tao and Ziegler, based on a general structural result by Bergelson, Tao and Ziegler [6]. Here is the statement of their inverse theorem.

Theorem 7.1. *Let $\delta > 0$. Fix a prime p and an integer $k \leq p$. Then there exists a constant $c_{p,k}(\delta)$ such that for any function $f : \mathbb{F}_p^n \to \mathbb{C}$ satisfying $|f| \leq 1$ and $\|f\|_{U^k(\mathbb{F}_p^n)} \geq \delta$, there exists a polynomial $P \in \mathbb{F}_p[x_1, \ldots, x_n]$ such that*

$$\mathbb{E}_{x \in \mathbb{F}_p^n} f(x) \exp\left(\frac{2\pi i P(x)}{p}\right) \geq c_{p,k}(\delta).$$

In the low-characteristic case $p < k$, the statement [86] requires the introduction of *nonclassical polynomials* instead of the classical ones featuring here, but we will not need this case here.

Explicit bounds are not generally known for the constant $c_{p,k}(\delta)$. For $k = 2$, it is easily seen to be polynomial in δ, while for $k = 3$, is is conjectured to be polynomial, the best unconditional bound at the time of writing being quasipolynomial in δ [43]. Very recently, Gowers and Milićević [36] obtained a doubly exponential bound for the case $k = 4$ (and

$p > 3$).

Motivated by Theorem 7.1, we investigate correlations of μ with functions of the form $\chi(Q(f))$ for an additive character χ over $\mathbb{F}_q$ and a polynomial $Q \in \mathbb{F}_q[x_0, \ldots, x_{n-1}]$ in the coefficients $(x_0, \ldots, x_{n-1}) \in \mathbb{F}_q^n$ of $f = \sum_{i<n} x_i t^i$. We only consider polynomials Q of degree at most 2. Recall that the group $\widehat{\mathbb{F}_q}$ of additive characters is isomorphic to (the additive group of) $\mathbb{F}_q$. To express the isomorphism, let $\mathrm{Tr} : \mathbb{F}_q \to \mathbb{F}_p$ be the trace map. For $a \in \mathbb{F}_q$, let us denote

$$e_q(a) = \exp\left(\frac{2\pi i \mathrm{Tr}(a)}{p}\right).$$

Then the isomorphism $\mathbb{F}_q \to \widehat{\mathbb{F}_q}$ is given by $r \mapsto \chi_r$ where for any $r \in \mathbb{F}_q$, the character χ_r is defined by $\chi_r(x) = e_q(rx)$.

We now state our main results.

Theorem 7.2. *For any $\epsilon > 0$ and $\chi \in \widehat{\mathbb{F}_q}$, for any linear form $\ell \in \mathbb{F}_q[x_0, \ldots, x_{n-1}]$, we have*

$$\sum_{\deg f < n} \mu(f)\chi(\ell(f)) \ll_{\epsilon,q} q^{(3/4+\epsilon)n}. \tag{7.1}$$

uniformly in n and ℓ.

It suffices to prove Theorem 7.2 for $\chi = \chi_1$. In the integer case, Davenport [25] showed that for any $A > 0$, we have

$$\sum_{n=1}^{N} \mu(n)e(n\alpha) \ll_A N(\log N)^{-A}$$

uniformly in $\alpha \in \mathbb{R}/\mathbb{Z}$, where the implied constant is ineffective due to the possible existence of Siegel zeroes. Under the Generalised Riemann Hypothesis (GRH), the best result is due to Baker-Harman [3] and Montgomery-Vaughan (unpublished), who showed that for any $\epsilon > 0$,

$$\sum_{n=1}^{N} \mu(n)e(n\alpha) \ll_\epsilon N^{3/4+\epsilon} \tag{7.2}$$

uniformly in $\alpha \in \mathbb{R}/\mathbb{Z}$. Our exponent $\frac{3}{4} + \epsilon$ in (7.1) matches the one in (7.2) (though it is reasonable to conjecture that in both cases the best exponent is $\frac{1}{2} + \epsilon$). However, our proof of (7.1) differs from that of (7.2) in some respects. In particular, our proof of (7.1) uses L-functions of *arithmetically distributed relations* introduced by Hayes [50], as opposed to Dirichlet L-functions. We remark that very recently and independently of us, Sam Porritt

[72] proved a result similar to Theorem 7.2.

Regarding quadratic polynomials, we have the following similar but conditional result. It depends on the polylogarithmic bilinear Bogolyubov conjecture, Conjecture 5.9.

Theorem 7.3. *Assume $p > 2$. Let $A > 0$ and $\chi \in \widehat{\mathbb{F}_q}$. Assuming Conjecture 5.9, we have*

$$\sum_{\deg f < n} \mu(f)\chi(Q(f)) \ll_{q,A} q^n n^{-A} \tag{7.3}$$

uniformly in n and the quadratic polynomial Q in $\mathbb{F}_q[x_0, \ldots, x_{n-1}]$.

We have another result for quadratic phases that are more directly analogous to $n \mapsto e(\alpha n^2 + \beta n)$ in the integers. In this case, our bound is unconditional and gives a polynomial saving. We need some extra notation to state our result (see Section 7.2.1 for more precise definitions). Let $\mathbb{F}_q((\frac{1}{t}))$ be the ring of formal Laurent series in $1/t$. On $\mathbb{F}_q((\frac{1}{t}))$, we define the additive character $e(\alpha) = e_q(\alpha_{-1})$, where α_{-1} denotes the coefficient of t^{-1} in α.

Theorem 7.4. *There exists a constant $\epsilon > 0$ (independent of q) such that*

$$\sum_{\deg f < n} \mu(f) e(\alpha f^2 + \beta f) \ll_q q^{(1-\epsilon)n} \tag{7.4}$$

uniformly in n and $\alpha, \beta \in \mathbb{F}_q((\frac{1}{t}))$.

Note that we do not require $p > 2$ in Theorem 7.4 since when $p = 2$ the map $f \mapsto (\alpha f^2 + \beta f)_{-1}$ is linear and Theorem 7.4 follows from Theorem 7.2. When p is odd, the symmetric matrix of the quadratic form $f \mapsto (\alpha f^2)_{-1}$ is a *Hankel matrix*, i.e., a matrix whose (i, j)-entry depends only on $i + j$. Thus Theorem 7.4 can be reformulated in terms of Hankel matrices alone. We remark that in the integers, under GRH we have bounds with polynomial savings for the sum $\sum_{n=1}^{N} \mu(n) e(\alpha n^k)$ (see [52, 95]).

Note also that to prove the uniformity of the Möbius function on $\mathbb{F}_q[t]$ in the case where $q = p^s$ is not a prime, it is not enough to inspect its correlations with polynomials $P \in \mathbb{F}_q[x_1, \ldots, x_n]$ as in Theorem 7.3. Instead, for $f \in G_n \cong \mathbb{F}_p^{sn}$, write $\tilde{f} = \phi_n(f) \in \mathbb{F}_p^{sn}$ for the image of f by the isomorphism. Observe that not any $\mathbb{F}_p$-quadratic form $P(\tilde{f})$ can be realised as $\mathrm{Tr}(Q(f))$ for some $\mathbb{F}_q$-quadratic form $Q(f)$; this can be seen by simple counting. But controlling $\|\mu\|_{U^3(G_n)}$ precisely requires the control of correlations of μ with any $\mathbb{F}_p$-quadratic form $P(\tilde{f})$, whereas Theorem 7.3 only deals with $\mathbb{F}_q$-quadratic forms.

7.1 Overview of the proof

We will first focus on linear phases and prove Theorem 7.2. This will involve bounding correlations of the Möbius functions with Dirichlet characters and generalisations thereof called Hayes characters. The theory of these characters and the corresponding L-functions will be sketched in the next section, and the bounds for the character sums will be derived in Section 7.3

By analogy with the classical circle method, one can recognise in the analysis of the correlations of the Möbius function with quadratic phases a dichotomy "minor arcs/major arcs". Minor arcs correspond to very equidistributed phases, and thus to quadratic phases of high rank. For quadratic phases of rank less than $n/4$ (major arcs), we note that the saving in Theorem 7.2 allows us to conclude easily, as is shown in Corollary 7.9.

When proving Theorem 7.3, we will suppose for a contradiction that

$$\sum_{f \in G_n} \mu(f)\chi(Q(f)) \geq \epsilon q^n.$$

Let M be the $n \times n$ symmetric matrix corresponding to Q and k an integer. For any $a \in G_{k+1}$, consider the map $L_a : G_{n-k} \to G_n$ that maps f to af. We also write L_a to denote its $n \times (n-k)$ coordinate matrix in the canonical basis (i.e. the basis of monomials). For any $(a,b) \in G_{k+1}^2$, let $M_{a,b} = L_a^T M L_b + L_b^T M L_a$, which is a symmetric $(n-k) \times (n-k)$ matrix.

After exploiting Vaughan's identity in Section 7.5, we will find that for some $n \ll k \leq n$, the matrix M has the property that the set of pairs

$$P_h := \{(a,b) \in G_{k+1} \times G_{k+1} | \mathrm{rk}\, M_{a,b} \leq h\}$$

is large, that is, it contains at least δq^{2k+2} pairs for some parameters δ and h (depending on ϵ and n). We will want to convert this information about the ranks of many $M_{a,b}$ into a conclusion on the rank of M itself. However, we need these pairs to have some special structure in order to extract something useful; in particular, it would be extremely convenient if the set

$$\{(t^i, t^j) \mid (i,j) \in \{0, \ldots, k\}^2\} \tag{7.5}$$

were in P_h, because M_{t^i,t^j} is simply the symmetric part of a submatrix of M. Unfortunately, its large size alone does not force P_h to contain such a nice structure, but to boost our

chances, we are ready to do some additive smoothing, that is, adjoining to our set P_h elements such as $(a_1 - a_2, b)$ whenever (a_1, b) and (a_2, b) are in P_h; and the same on the second coordinate. The rank remains controlled under this operation, because $\operatorname{rk} M_{a_1-a_2,b} = \operatorname{rk}(M_{a_1,b} - M_{a_2,b}) \leq 2h$. We saw in Chapter 5 that such a bidirectional additive smoothing does indeed produce useful structures, namely sets cut out by a few (that is, $c(\delta)$) linear and bilinear forms.

We found in Theorem 5.8 that we can take $c(\delta)$ to be $O(\exp(\exp(\exp(\log^{O(1)} 1/\delta))))$, where the implied constants may depend on q, but unfortunately, because δ will be as small as n^{-5} say, this bound for $c(\delta)$ is too large to be of any use. Assuming Conjecture 5.9, we can take $c(\delta)$ as small as polylogarithmic in δ^{-1}. Applied with $\delta = n^{-O(1)}$, this means that the codimensions of the bilinear set identified in P_{64h} by Corollary 5.10 should be polylogarithmic in n. Further, still with Corollary 5.10 we will obtain sets of the form (7.5) inside P_{64h} in Section 7.6.

7.2 Preliminaries

7.2.1 Notation and basic facts

A useful reference for the circle method in function fields, of which the basics are sketched below, is [64]. Let $\mathbb{F}_q(t)$ be the field of fractions of $\mathbb{F}_q[t]$. On $\mathbb{F}_q(t)$ we can define a norm by $|f/g| = q^{\deg f - \deg g}$, with the convention that $\deg 0 = -\infty$. The completion of $\mathbb{F}_q(t)$ with respect to this norm is

$$\mathbb{F}_q\left(\left(\frac{1}{t}\right)\right) = \left\{\alpha = \sum_{i=-\infty}^{n} a_i t^i \mid n \in \mathbb{Z}, a_i \in \mathbb{F}_q \text{ for every } i\right\},$$

the set of formal Laurent series in $\frac{1}{t}$. It is easy to see that if α is as above and $a_n \neq 0$ then $|\alpha| = q^n$.

We observe the inclusions $\mathbb{F}_q[t] \subset \mathbb{F}_q(t) \subset \mathbb{F}_q((\frac{1}{t}))$, and think of $\mathbb{F}_q[t], \mathbb{F}_q(t)$ and $\mathbb{F}_q((\frac{1}{t}))$ are the analogs of $\mathbb{Z}, \mathbb{Q}$ and $\mathbb{R}$, respectively. Let us put $\mathbb{T} = \{\alpha \in \mathbb{F}_q((\frac{1}{t})) \mid |\alpha| < 1\}$. This is analogous to the usual torus $\mathbb{R}/\mathbb{Z}$.

For $\alpha \in \mathbb{F}_q((\frac{1}{t}))$, we write $(\alpha)_{-1}$ to denote the coefficient of t^{-1} in α and define $e(\alpha) = e_q((\alpha)_{-1})$. This is an additive character on $\mathbb{F}_q((\frac{1}{t}))$ and allows us to perform Fourier analysis on $\mathbb{F}_q[t]$. All additive characters on $\mathbb{F}_q[t]$ are given by $f \mapsto e(f\alpha)$ for some $\alpha \in \mathbb{T}$.

We denote by M the set of all monic polynomials in $\mathbb{F}_q[t]$, and by A_n the set of all monic

polynomials of degree n, while $\mathcal{I}$ denotes the set of all monic, irreducible polynomials. Remember that G_n is the set of all polynomials (not necessarily monic) of degree less than n. We use the convention that $\sum_{\deg f=l}$ means $\sum_{f\in A_l}$ (that is, a sum over monic polynomials).

The von Mangoldt function on $\mathbb{F}_q[t]$ is defined by

$$\Lambda(f) = \begin{cases} \deg P & \text{if } f = P^k \text{ for some monic irreducible } P \text{ and } k \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

The "prime number theorem" [74] on $\mathbb{F}_q[t]$ reads

$$\sum_{\deg f=l} \Lambda(f) = q^l.$$

7.2.2 L-functions of arithmetically distributed relations

To prove Theorem 7.2, we first observe that any linear form on G_n can be represented as a map $f \mapsto (\alpha f)_{-1}$ for some $\alpha \in \mathbb{T}$. Thus Theorem 7.2 can be rephrased as a bound for sums of the form

$$\sum_{f\in G_n} \mu(f)e(\alpha f),$$

or, equivalently and more conveniently, of the form

$$\sum_{f\in A_n} \mu(f)e(\alpha f).$$

Now suppose that α is approximated by a fraction a/Q of polynomials up to a remainder $\beta = \sum_{i=-\infty}^{-l} \beta_i t^i$ for some $l \geq 2$, that is, $\alpha = a/Q+\beta$. Then $e(\alpha f) = e(af/Q)e(\beta f)$ depends only on the residue class of f modulo Q and the coefficients of the terms of degrees at least $l-1$ of $f = \sum_{i=1}^{n} a_i t^{n-i} + t^n$. We refer to $a_1, \ldots, a_l$ as the first l coefficients of f (if $i > n$ then we define $a_i = 0$). We thus need to understand functions on A_n that only depend on the congruence class modulo a fixed modulus Q and the first l coefficients. Hence for $l \geq 0$ and $Q \in \mathbb{F}_q[t]$, we define an equivalence relation $R_{l,Q}$ on M as follows:

$f \equiv g \pmod{R_{l,Q}}$ if $f \equiv g \pmod{Q}$ and the first l coefficients of f and g are the same.

The above is an example of an *arithmetically distributed relation*, of which Hayes [50, Section 8] developed the theory. We briefly review it here, but the relevant facts can also be found in [54] or [19].

It is easy to check that $M/R_{l,Q}$ is a semigroup with respect to multiplication on $\mathbb{F}_q[t]$. The equivalence class of a polynomial $f \in \mathbb{F}_q[t]$ is invertible in $M/R_{l,Q}$ if and only if $(f, Q) = 1$. Set $G_{l,Q} := (M/R_{l,Q})^\times$, the set of invertible elements. This is a group of cardinality $q^l\varphi(Q)$, where $\varphi(Q) = \#(\mathbb{F}_q[t]/(Q))^\times$. Note that $G_{0,Q}$ is simply $(\mathbb{F}_q[t]/(Q))^\times$.

We can extend any character λ on $G_{l,Q}$ to all of M by setting $\lambda(f) = 0$ if $(f, Q) \neq 1$. We define the L-function associated with λ as

$$L(s, \lambda) = \sum_{f \in M} \lambda(f) \frac{1}{|f|^s},$$

which converges absolutely for $\Re(s) > 1$. It is convenient to define

$$\mathcal{L}(z, \lambda) = \sum_{f \in M} \lambda(f) z^{\deg(f)} = \sum_{n=1}^{\infty} z^n \sum_{f \in A_n} \lambda(f) \tag{7.6}$$

so that $L(s, \lambda) = \mathcal{L}(q^{-s}, \lambda)$. We have the Euler product formula

$$\mathcal{L}(z, \lambda) = \prod_{P \in \mathcal{I}} \left(1 - \lambda(P) z^{\deg P}\right)^{-1} \tag{7.7}$$

for $|z| < 1/q$.

In the same range of z, we also have

$$\frac{1}{\mathcal{L}(z, \lambda)} = \prod_{P} \left(1 - \lambda(P) z^{\deg P}\right) = \sum_{f \in M} \mu(f)\lambda(f) z^{\deg f} = \sum_{n=1}^{\infty} z^n \sum_{f \in A_n} \mu(f)\lambda(f). \tag{7.8}$$

The character constantly equal to 1 on $G_{l,Q}$ is called the *principal character*. When λ is not the principal character, $\mathcal{L}(z, \lambda)$ is a polynomial of degree $d(\lambda) < l + \deg Q$ [50, Lemma 8.2]. The *Generalised Riemann Hypothesis* states that all roots of $\mathcal{L}(z, \lambda)$ have modulus $q^{-1/2}$ or 1 for any character λ modulo an arithmetically distributed congruence relation such as $R_{l,Q}$. Weil's proof of the Riemann Hypothesis (for Dirichlet characters) was extended to these generalised characters by Rhin [73] (see in particular Chapitre 2,

Sections 4 to 6). In other words, we can write

$$\mathcal{L}(z,\lambda) = \prod_{i=1}^{d(\lambda)} (1 - \alpha_i z), \tag{7.9}$$

where $|\alpha_i| = q^{1/2}$ or 1 for $i = 1, \ldots, d(\lambda)$. In particular, $\mathcal{L}(z,\lambda)$ can be extended to an entire function and (7.8) remains valid when $|z| < q^{-1/2}$.

As an aside, we briefly show a simple bound, derived independently by the author and Porritt [72], which will be beaten by a more involved method in Theorem 7.6, but remains interesting for its simplicity and the ideas it involves.

Proposition 7.5. *Suppose λ is a non principal character on $G_{l,Q}$ and $M = l + \deg Q$. Then*

$$\sum_{f \in A_n} \mu(f)\lambda(f) \ll \binom{n+M-2}{M-2} q^{n/2}. \tag{7.10}$$

This is bound is particularly good when M is bounded, in which case the binomial coefficient just varies polynomially in n. When M is of the order of magnitude of n, the right-hand side of (7.10) can be bounded by $O(q^{(1/2+\epsilon_q)n})$ where $\epsilon_q > 0$ and $\epsilon_q \to 0$ as q tends to infinity [72].

Proof. The left-hand side of equation (7.10) is equal to the coefficient of z^n in the power series $\frac{1}{\mathcal{L}(z,\lambda)}$ because of equation (7.8). On the other hand, equation (7.9) implies that

$$\frac{1}{\mathcal{L}(z,\lambda)} = \sum_{n=0}^{+\infty} z^n \sum_{\sum_{i=1}^{M-1} n_i = n} \prod_{i=1}^{M-1} \alpha_i^{n_i}.$$

Now we use the fact that $|\alpha_i| \leq \sqrt{q}$ and the basic counting result

$$|\{(n_1, \ldots, n_k) \in \mathbb{N}^k \mid \sum_{i=1}^{k} n_i = n\}| = \binom{n+k-1}{k-1}.$$

This brings about the conclusion. □

When λ is the principal character of $G_{l,Q}$, we have

$$\begin{aligned}\mathcal{L}(z,\lambda) &= \prod_{\substack{P\in\mathcal{I},\\(P,Q)=1}} \left(1-z^{\deg P}\right)^{-1}\\ &= \prod_{\substack{P\in\mathcal{I},\\P|Q}} \left(1-z^{\deg P}\right) \prod_{P\in\mathcal{I}} \left(1-z^{\deg P}\right)^{-1}\\ &= \prod_{\substack{P\in\mathcal{I},\\P|Q}} \left(1-z^{\deg P}\right) \frac{1}{1-qz}.\end{aligned}$$

Consequently, $\mathcal{L}(z,\lambda)$ can be extended to a meromorphic function and

$$\frac{1}{\mathcal{L}(z,\lambda)} = \sum_{n=1}^{\infty} z^n \sum_{f\in A_n,(f,Q)=1} \mu(f) = (1-qz) \prod_{\substack{P\in\mathcal{I},\\P|Q}} \left(1-z^{\deg P}\right)^{-1} \tag{7.11}$$

for all $|z|\neq 1$.

7.3 Character sum estimates

In this section we prove the following.

Theorem 7.6. *Let $l\geq 0, Q\in\mathbb{F}_q[t], \deg Q = m\geq 0$ and λ be a character of $G_{l,Q}$. Then for any d, and $\epsilon>0$, we have*

$$\left|\sum_{f\in A_d}\mu(f)\lambda(f)\right| \ll_{\epsilon,q} q^{\left(\frac{1}{2}+\epsilon\right)d+\epsilon(m+l)} \tag{7.12}$$

Proof. First we assume that λ is not principal. We will prove the following more precise bound

$$\left|\sum_{f\in A_d}\mu(f)\lambda(f)\right| \leq q^{\frac{d}{2}+\frac{d\log\log(m+l)}{\log(m+l)}+O_q\left(\frac{m+l}{\log^2(m+l)}\right)}. \tag{7.13}$$

Our method is a generalization of the proof of [7, Theorem 2]. As [7], we deduce (7.13) from an estimate for $\log\mathcal{L}(z,\lambda)$ near the circle $|z|=q^{-1/2}$, which in turn is deduced from an estimate for $\frac{\mathcal{L}'(z,\lambda)}{\mathcal{L}(z,\lambda)}$. By taking the logarithmic derivatives of (7.7) and (7.9), we have

two different expressions for $\frac{\mathcal{L}'(z,\lambda)}{\mathcal{L}(z,\lambda)}$. On the one hand, we have

$$\frac{\mathcal{L}'(z,\lambda)}{\mathcal{L}(z,\lambda)} = \sum_{l=1}^{\infty} a_l z^{l-1},$$

where

$$a_l = -\sum_{i=1}^{d(\lambda)} \alpha_i^l \tag{7.14}$$

according to (7.9). On the other hand, according to (7.7), we have

$$a_l = \sum_{\deg f = l} \Lambda(f)\lambda(f). \tag{7.15}$$

It follows from (7.14) that

$$|a_l| \leq d(\lambda) q^{l/2} \tag{7.16}$$

and (7.15) implies that

$$|a_l| \leq \sum_{\deg f = l} \Lambda(f) = q^l. \tag{7.17}$$

Let $L = \lfloor 2\log_q d(\lambda) \rfloor$. For $l > L$ we use the bound (7.16) and for $l \leq L$ we use the bound (7.17). Therefore, for any z, we have

$$\left| \frac{\mathcal{L}'(z,\chi)}{\mathcal{L}(z,\chi)} \right| \leq \sum_{l=1}^{L} q^l |z|^{l-1} + \sum_{l=L+1}^{\infty} d(\lambda) q^{l/2} |z|^{l-1}. \tag{7.18}$$

Let $0 < \epsilon < 1/4$ (its precise value will be chosen later) and $R = q^{-1/2-\epsilon}$. Let w be arbitrary on the circle $|w| = R$. Integrating (7.18) along the line from 0 to w, and noting that $\mathcal{L}(0,\lambda) = 1$, we have

$$|\log \mathcal{L}(w,\lambda)| \leq \sum_{l=1}^{L} \frac{(Rq)^l}{l} + \sum_{l=L+1}^{\infty} d(\lambda) \frac{(Rq^{1/2})^l}{l}. \tag{7.19}$$

The second sum in (7.19) can be bounded by

$$\frac{d(\lambda)}{L} \sum_{l=L+1}^{\infty} (Rq^{1/2})^l \leq \frac{d(\lambda)}{L} R^L q^{\frac{L}{2}} \frac{1}{1 - Rq^{1/2}} \ll \frac{d(\lambda)^2 R^L}{L} \frac{1}{1 - Rq^{1/2}}. \tag{7.20}$$

As for the first sum in (7.19), we bound it crudely by

$$\sum_{l=1}^{L}(Rq)^l \leq (Rq)^L \sum_{k=0}^{\infty}(Rq)^{-k} \leq \frac{d(\lambda)^2 R^L}{1-(qR)^{-1}} \ll_q d(\lambda)^2 R^L \tag{7.21}$$

since $qR \geq q^{1/4}$. By combining (7.20) and (7.21), we have

$$|\log \mathcal{L}(w,\lambda)| \ll_q d(\lambda)^2 R^L \left(1 + \frac{1}{L(1-Rq^{1/2})}\right).$$

Hence,

$$\left|\frac{1}{\mathcal{L}(w,\lambda)}\right| \leq \exp\left(O_q\left(d(\lambda)^2 R^L \left(1 + \frac{1}{L(1-Rq^{1/2})}\right)\right)\right). \tag{7.22}$$

Let $C_R = \{w \in \mathbb{C} \mid |w| = R\}$. From (7.8) we see that

$$\begin{aligned}
\left|\sum_{f\in A_d} \lambda(f)\mu(f)\right| &= \left|\frac{1}{2\pi i}\int_{C_R} \frac{1}{\mathcal{L}(w,\chi)} w^{-d-1} dw\right| \\
&\leq \max_{C_R}\left|\frac{1}{\mathcal{L}(w,\lambda)}\right| R^{-d} \\
&\leq q^{d(1/2+\epsilon)+O_q\left(d(\lambda)^{1-2\epsilon}\left(1+\frac{1}{\epsilon\log d(\lambda)}\right)\right)},
\end{aligned} \tag{7.23}$$

where we have used $R = q^{-1/2-\epsilon}$ to obtain the final inequality. We now choose $\epsilon = \log\log d(\lambda)/\log d(\lambda)$. Recalling that $d(\lambda) \leq l+m-1$, (7.13) follows. The bound (7.13) is stronger than (7.12) when $\log\log(l+m)/\log(l+m)$ is smaller than the ϵ in (7.12), which is the case whenever $l+m$ is large enough. For the finitely many remaining pairs (m,l), equation (7.12) follows from (7.23) (with the same ϵ).

We now consider the case where λ is principal. From (7.11), on the circle $|z| = q^{-1/2}$, we have

$$\begin{aligned}
\left|\frac{1}{\mathcal{L}(z,\lambda)}\right| &= |1-qz| \prod_{P\in\mathcal{I}, P|Q} \left|1 - z^{\deg P}\right|^{-1} \\
&\ll \prod_{P\in\mathcal{I}, P|Q} \left(1 - q^{-\deg P/2}\right)^{-1} \\
&\leq \prod_{P\in\mathcal{I}, P|Q} \left(1 - q^{-1/2}\right)^{-1} \\
&= (1-q^{-1/2})^{-k} \leq q^{O_q\left(\frac{m}{\log m}\right)}
\end{aligned} \tag{7.24}$$

where k is the number of monic irreducible factors of Q and (7.24) follows from Lemma E.1 (a divisor bound which we postpone to an appendix). Integrating $z^{-d-1}\frac{1}{\mathcal{L}(z,\lambda)}$ along the circle $|z| = q^{-1/2}$ and using (7.24), we see that

$$\sum_{f\in A_n,(f,Q)=1} \mu(f) \ll q^{\frac{d}{2}+O_q\left(\frac{m}{\log m}\right)} \tag{7.25}$$

from which (7.12) follows. □

We remark that (7.11) readily gives a formula for $\sum_{f\in A_n,(f,Q)=1}\mu(f)$, but it is not immediate to derive (7.25) from this formula.

7.4 Exponential sum estimates

We say that a function $F : M \to \mathbb{C}$ is $R_{l,Q}$-*periodic* if it is constant on each equivalence class of $R_{l,Q}$. In other words, F is $R_{l,Q}$-periodic if $F(f)$ depends only on the residue class of f modulo Q and the first l coefficients of f. We say F is 1-*bounded* if $|F(f)| \leq 1$ for any $f \in M$. We first show that μ is orthogonal to $R_{l,Q}$-periodic functions by adapting the argument of [46, Proposition 3.2].

Proposition 7.7. *Suppose* $\deg Q = m$. *For any* $R_{l,Q}$*-periodic and 1-bounded function* $F : M \to \mathbb{C}$ *and* $\epsilon > 0$, *we have*

$$\sum_{f\in A_n} F(f)\mu(f) \ll_{\epsilon,q} q^{(1/2+\epsilon)(n+m+l)},$$

where the bound is uniform in F.

Proof. We first consider the case where $F(f) = 0$ whenever $(f, Q) \neq 1$. This means that F is a function on $G_{l,Q}$. Let $K = |G_{l,Q}| = q^l\varphi(Q) \leq q^{l+m}$ and $\lambda_1, \ldots, \lambda_K$ be the characters of $G_{l,Q}$. Define the Fourier coefficients of F by

$$\widehat{F}(\lambda) = \mathbb{E}_{f\in G_{l,Q}} F(f)\overline{\lambda(f)}$$

for any character λ of $G_{l,Q}$. Then $F(f) = \sum_{i=1}^{K} \widehat{F}(\lambda_i)\lambda_i(f)$. Plancherel's formula implies

$$\sum_{i=1}^{K} \left|\widehat{F}(\lambda_i)\right|^2 = \mathbb{E}_{f\in G_{l,Q}}|F(f)|^2 \leq 1. \tag{7.26}$$

We have

$$\begin{aligned}
\left|\sum_{f\in A_n} F(f)\mu(f)\right| &= \left|\sum_{i=1}^{K} \widehat{F}(\lambda_i) \sum_{f\in A_n} \lambda_i(f)\mu(f)\right| \\
&\ll_{\epsilon,q} q^{n/2+\epsilon(n+l+m)} \sum_{i=1}^{K} \left|\widehat{F}(\lambda_i)\right| \qquad (7.27)\\
&\leq q^{n/2+\epsilon(n+l+m)} K^{1/2} \qquad (7.28)\\
&\leq q^{n/2+(l+m)/2+\epsilon(n+l+m)}.
\end{aligned}$$

Here (7.27) follows from Theorem 7.6 and (7.28) follows from the Cauchy-Schwarz inequality and (7.26).

Next we consider the general case where $F(f)$ is not necessarily 0 when $(f,Q)=1$. If f is squarefree and $(f,Q)=D$, we can write $f=Dg$ with g squarefree and $(g,Q)=1$. Hence

$$\begin{aligned}
\sum_{f\in A_n} F(f)\mu(f) &= \sum_{\substack{D\in M, D|Q,\\ D \text{ squarefree}}} \sum_{\substack{\deg g=n-\deg D,\\ g \text{ squarefree}}} F(Dg)\mu(Dg)1_{(g,Q)=1} \\
&= \sum_{D\in M, D|Q} \mu(D) \sum_{\substack{\deg g=n-\deg D,\\ g \text{ squarefree}}} F(Dg)\mu(g)1_{(g,Q)=1} \qquad (7.29)
\end{aligned}$$

Now the function $g\mapsto F(Dg)\mu(g)1_{(g,Q)=1}$ is $R_{l,Q}$-periodic, and vanishes on elements of M that are not coprime to Q. From the above, we infer that

$$\sum_{\substack{\deg g=n-\deg D,\\ g \text{ squarefree}}} F(Dg)\mu(g)1_{(g,Q)=1} \ll_{\epsilon,q} q^{\frac{n-\deg D}{2}+\frac{l+m}{2}+\epsilon(n+m+l)}$$

for any $\epsilon>0$. Furthermore, still for any $\epsilon>0$, we observe that

$$\sum_{D|Q} q^{-(\deg D)/2} \leq \tau(Q) \ll_{\epsilon,q} |Q|^{\epsilon} = q^{\epsilon m}$$

by Lemma E.1. This completes the proof. □

We will now use Proposition 7.7 and the ideas outlined at the beginning of Section 7.2.2 to prove the following exponential sum estimate.

Theorem 7.8. *Given any $\epsilon > 0$, for all $\alpha \in \mathbb{T}$ and n, we have*

$$\sum_{f \in A_n} \mu(f) e(\alpha f) \ll_{\epsilon,q} q^{(3/4+\epsilon)n} \tag{7.30}$$

and

$$\sum_{f \in G_n} \mu(f) e(\alpha f) \ll_{\epsilon,q} q^{(3/4+\epsilon)n}. \tag{7.31}$$

The first bound implies the second, because

$$\sum_{f \in G_n} \mu(f) e(\alpha f) = \sum_{c \in \mathbb{F}_q^*} \sum_{k=0}^{n-1} \sum_{f \in A_k} \mu(f) e(\alpha c f)$$

so we only need to prove (7.30). It is easy to see that any linear form on G_n can be written as $f \mapsto (\alpha f)_{-1}$ (i.e., the coefficient of t^{-1} in αf) for some $\alpha \in \mathbb{T}$. Thus Theorem 7.2 follows immediately from Theorem 7.8.

Proof. By Dirichlet's approximation theorem, we can find a and g in $\mathbb{F}_q[t]$, where $g \neq 0$ and $\deg g \leq \lfloor \frac{n}{2} \rfloor$, such that $\left|\alpha - \frac{a}{g}\right| < \frac{1}{q^{\lfloor \frac{n}{2} \rfloor} |g|}$. Let $\beta = \alpha - \frac{a}{g}$. Then

$$\sum_{f \in A_n} \mu(f) e(\alpha f) = \sum_{f \in A_n} \mu(f) e\left(\frac{af}{g}\right) e(\beta f).$$

Since $|\beta| < q^{-\lfloor \frac{n}{2} \rfloor - \deg g}$, we see that $e(\beta f)$ depends only on the first $n - \lfloor \frac{n}{2} \rfloor - \deg g$ coefficients of f. Also, $e\left(\frac{af}{g}\right)$ depends only on the residue class of f modulo g. Applying Proposition 7.7 to $(l, Q) = (n - \lfloor \frac{n}{2} \rfloor - \deg g, g)$, for any $\epsilon > 0$, we have

$$\sum_{f \in A_n} \mu(f) e\left(\frac{af}{g}\right) e(\beta f) \ll_{\epsilon,q} q^{\frac{1+\epsilon}{2}(n+n-\lfloor \frac{n}{2} \rfloor - \deg g + \deg g)} = q^{\frac{1+\epsilon}{2}(2n - \lfloor \frac{n}{2} \rfloor)} \ll_{\epsilon,q} q^{(3/4+\epsilon)n},$$

as desired. □

As we show next, Theorem 7.2 implies that if a function is determined by the values of a few linear forms, it does not correlate with the Möbius function.

Corollary 7.9. *Let $c > 0$ be a constant. Let $F : \mathbb{F}_q^r \to \mathbb{C}$ be 1-bounded and suppose $r \leq cn$.*

Let $\ell_1, \ldots, \ell_r$ be linear forms on G_n. Then for any $\epsilon > 0$,

$$\sum_{f \in G_n} \mu(f) F(\ell_1(f), \ldots, \ell_r(f)) \ll_{\epsilon,q} q^{(3/4+c+\epsilon)n}.$$

In particular, supposing the characteristic q is odd, for a quadratic form Q of rank at most r and any $\chi \in \widehat{\mathbb{F}_q}$, we have

$$\sum_{f \in G_n} \mu(f) \chi(Q(f)) \ll_{\epsilon,q} q^{(3/4+c+\epsilon)n}.$$

Obviously, this is interesting only if $c < 1/4$.

Proof. Theorem 7.2 immediately implies that for any linear forms ℓ on G_n, we have

$$\sum_{f \in G_n} \mu(f) e_q(\ell(f)) \ll_{\epsilon,q} q^{(3/4+\epsilon)n}. \tag{7.32}$$

For any $\mathbf{a} = (a_1, \ldots, a_r) \in \mathbb{F}_q^r$, let $V_{\mathbf{a}} \leq G_n$ be the affine subspace defined by the equations $\ell_i(f) = a_i$ for $i \in [r]$. Then one can write

$$\sum_{f \in G_n} \mu(f) F(\ell_1(f), \ldots, \ell_r(f)) = \sum_{\mathbf{a} \in \mathbb{F}_q^r} F(\mathbf{a}) \sum_{f \in V_{\mathbf{a}}} \mu(f). \tag{7.33}$$

Now we observe that

$$1_{V_{\mathbf{a}}}(f) = \mathbb{E}_{\boldsymbol{\chi} = (\chi_1, \ldots, \chi_r) \in \widehat{\mathbb{F}_q}^r} \prod_{i \in [r]} \chi_i(\ell_i(f) - a_i)$$

so that

$$\sum_{f \in V_{\mathbf{a}}} \mu(f) = \mathbb{E}_{\boldsymbol{\chi} \in \widehat{\mathbb{F}_q}^r} \prod_{i \in [r]} \chi_i(-a_i) \sum_{f \in G_n} \mu(f) \prod_{i \in [r]} \chi_i(\ell_i(f)),$$

and by the triangle inequality

$$\left| \sum_{f \in V_{\mathbf{a}}} \mu(f) \right| \leq \max_{\boldsymbol{\chi} \in \widehat{\mathbb{F}_q}^r} \left| \sum_{f \in G_n} \mu(f) \prod_{i \in [r]} \chi_i(\ell_i(f)) \right|.$$

Recall from Section 7.2.1 that each χ_i is of the form $\chi_i(x) = e_q(t_i x)$, so that

$$\prod_{i\in[r]} \chi_i(\ell_i(f)) = e_q\left(\sum_{i=1}^{r} t_i \ell_i(f)\right).$$

We then apply (7.32) to the linear form $\ell = \sum_{i\in[r]} t_i \ell_i$. This shows that

$$|\sum_{f\in V_{\mathbf{a}}} \mu(f)| \ll q^{(3/4+\epsilon)n}.$$

Plugging this bound in equation (7.33) and using the fact that $|F| \leq 1$ gives the desired result.

Finally, the last affirmation of the corollary is justified by the fact that a quadratic form of rank r is a function that is determined by r linear forms (in fact, it is the sum of the squares of r linear forms). □

7.5 Quadratic phases and Vaughan's identity

From now on, we assume that the field $\mathbb{F}_q$ we work with has characteristic $p > 2$. Recall $q = p^s$ and $s \geq 1$. We call *quadratic form* on $\mathbb{F}_q^n$ a homogenous polynomial of degree 2, that is, a map of the form $F(x) = x^T M x$ where M is a symmetric matrix. The corresponding (symmetric) bilinear form is the map

$$B(x, y) = x^T M y.$$

The *rank* of F is the rank of the matrix M. It equals the codimension of the space K of vectors x such that the linear form B_x defined by $B_x(y) = B(x, y)$ satisfies $B_x = 0$. A *quadratic polynomial* on $\mathbb{F}_q^n$ is a polynomial of degree 2, that is, a quadratic form plus a linear form plus a constant. A *quadratic phase* is a map of the form $\Phi(x) = \chi(P(x))$ for a quadratic polynomial P and an additive character χ. Its *rank* is the rank of the corresponding quadratic form. Thanks to the following standard lemma, quadratic phases can be classified, depending on their rank, into major arcs and minor arcs, by analogy with the circle method.

Lemma 7.10 (Gauss sums)**.** *Let $\Phi(x) = \chi(P(x))$ be a quadratic phase of rank at least r. Then*

$$|\mathbb{E}_{x\in\mathbb{F}_q^n} \Phi(x)| \leq q^{-r/2}.$$

Thus quadratic phases of low rank correspond to major arcs, while the ones of high rank correspond to minor arcs.

Proof. We use a standard technique known as *Weyl differencing*, consisting in squaring the expectation to reduce the degree of the phase. We have

$$\begin{aligned} |\mathbb{E}_{x\in\mathbb{F}_q^n}\Phi(x)|^2 &= \mathbb{E}_{x,h}\Phi(x+h)\overline{\Phi(x)} \\ &= \mathbb{E}_{x,h}\chi(P(x+h)-P(x)) \\ &= \mathbb{E}_h\chi(P(h))\mathbb{E}_x\chi(2B_h(x)) \end{aligned}$$

where all variables range over $\mathbb{F}_q^n$. Now if $h \notin K$, the form $2B_h$ is a nonzero linear form (recall that the characteristic p is not 2), whence $\mathbb{E}_{x\in\mathbb{F}_q^n}\chi(2B_h(x)) = \mathbb{E}_{x\in\mathbb{F}_q}\chi(x) = 0$. This implies that

$$|\mathbb{E}_{x\in\mathbb{F}_q^n}\Phi(x)|^2 \leq \mathbb{E}_{h\in\mathbb{F}_q^n}1_{h\in K} = q^{-r},$$

and the claim follows. □

The rest of the section is devoted to the proof of Theorem 7.3. Let P be a quadratic polynomial on G_n and let $\Phi = \chi \circ P$ be a quadratic phase. We want to bound the sum

$$\sum_{f\in G_n}\mu(f)\Phi(f).$$

As already indicated in Section 7.1, the general strategy is the following. We first observe that when Φ is a quadratic phase of rank at most cn with $c < 1/4$, then Corollary 7.9 concludes: indeed, a quadratic form of rank r depends on r linear forms only, so a quadratic polynomial of rank r depends on $r+1$ linear forms at most. So we will show that in order for μ to correlate with a quadratic phase Φ, the corresponding quadratic form needs to be of small rank (major arc). This will imply that μ cannot correlate with a quadratic phase at all.

With the help of Vaughan's identity, a standard tool in analytic number theory, we will show the following.

Proposition 7.11. *Let $\delta > 0$. Suppose $|\sum_{f\in G_n}\mu(f)\Phi(f)| \geq \delta q^n$. Then at least one of the following two statements holds.*

1. *There exists $k \leq n/9$ such that for at least one polynomial d of degree k, the quadratic polynomial on G_{n-k} defined by*

$$w \mapsto P(dw)$$

has rank at most $O(\log(n/\delta))$.

2. *There exists* $k \in [n/18, 17n/18]$ *such that for at least* $(\delta/n)^{O(1)}q^{2k}$ *pairs of polynomials* d, d' *of degree* k*, the quadratic polynomial on* G_{n-k} *defined by*

$$w \mapsto P(d'w) - P(dw)$$

has rank at most $O(\log(n/\delta))$.

Before proving this proposition, we underline that for any $d \in G_{k+1}$, we see the map $w \mapsto dw$ as a linear map from G_{n-k} to G_n which allows one to see $w \mapsto P(dw)$ as a quadratic polynomial.

We now start the proof of Proposition 7.11. The first tool we need is Vaughan's identity [55, Proposition 13.5], which reads[1]

$$\mu(f) = -\sum_{\substack{ab|f \\ \deg a \leq u, \deg b \leq v}} \mu(a)\mu(b) + \sum_{\substack{ab|f \\ \deg a > u, \deg b > v}} \mu(a)\mu(b),$$

where the sum is over monic polynomials a and b, and $u = v = n/18$ (though in general they can be chosen arbitrarily). We shall adopt the notational convention that $N = q^n, U = q^u$ and so on. Moreover, for $f \in \mathbb{F}_q[t]$, recall the notation $|f| = q^{\deg f}$. Vaughan's identity implies that

$$\sum_{f \in G_n} \mu(f)\Phi(f) = -T_1 + T_2, \tag{7.34}$$

where

$$T_1 = \sum_{|d| \leq UV} a_d \sum_{w \in G_{n-\deg d}} \Phi(dw) \tag{7.35}$$

and

$$T_2 = \sum_{V \leq |d| \leq N/U} b_d \sum_{w \in G_{n-\deg d}} \mu(w)\Phi(dw) \tag{7.36}$$

are known as *Type I* and *Type II* sums respectively. The sums over d are over monic polynomials. The coefficients a_d are unimportant and all we need to know is that $\max(|a_d|, |b_d|) \leq \tau(d)$, where τ is the divisor function. We will need upper bounds regarding this function, the proof of which are postponed to Appendix E. In the Type I sum, we have made the change of variables $d = ab, w = f/d$, while in the Type II sum we wrote

[1]The reference we provide deals with the integer setting, but the proof is the same here.

$w = \mathrm{lc}(f)b, d = f/w$, where lc stands for leading coefficient, so that d is monic. The decomposition of the sum into two parts in (7.34) yields the following dichotomy, which we will use to prove Proposition 7.11.

Proposition 7.12. *Let $\delta > 0$. Suppose $|\sum_{f\in G_n} \mu(f)\Phi(f)| \geq \delta q^n$. Then either there exists $k \leq n/9$ such that*

$$\mathbb{E}_{d\in A_k}|\mathbb{E}_{w\in G_{n-k}}\Phi(dw)|^2 \geq \delta^2/(16n^5), \tag{7.37}$$

or there exists $k \in [n/18, 17n/18]$ such that

$$\mathbb{E}_{w,w'\in G_{n-k}}\mathbb{E}_{d,d'\in A_k}\Phi(dw)\overline{\Phi(dw')\Phi(d'w)}\Phi(d'w') \geq \delta^4/(256n^{10}). \tag{7.38}$$

Proof. If $|\sum_{f\in G_n} \mu(f)\Phi(f)| \geq \delta N$, the decomposition (7.34) implies that either $|T_1| \geq \delta N/2$ or $|T_2| \geq \delta N/2$. Suppose first $|T_1| \geq \delta N/2$. On the other hand, using the triangle inequality and equation (7.35), we bound T_1 by

$$\begin{aligned} |T_1| &\leq \sum_{k\leq u+v}\sum_{d\in A_k} \tau(d)\frac{N}{|d|}|\mathbb{E}_{w\in G_{n-k}}\Phi(dw)| \\ &\leq n\max_{k\leq u+v}\frac{N}{K}\sum_{d\in A_k}\tau(d)|\mathbb{E}_{w\in G_{n-k}}\Phi(dw)|. \end{aligned}$$

Fix an integer $k \leq u+v = n/9$ that realises the maximum in the line above. The Cauchy-Schwarz inequality then yields

$$|T_1|^2/N^2 \leq n^2\left(\mathbb{E}_{d\in A_k}\tau^2(d)\right)\left(\mathbb{E}_{d\in A_k}|\mathbb{E}_{w\in G_{n-k}}\Phi(dw)|^2\right).$$

Now Lemma E.2 ensures that

$$\mathbb{E}_{d\in A_k}\tau^2(d) \leq 4k^3 \leq 4n^3,$$

so we can affirm that

$$\delta^2N^2/4 \leq |T_1|^2 \leq 4n^5N^2\mathbb{E}_{d\in A_k}|\mathbb{E}_{w\in G_{n-k}}\Phi(dw)|^2.$$

This means that

$$\mathbb{E}_{d\in A_k}|\mathbb{E}_{w\in G_{n-k}}\Phi(dw)|^2 \geq \delta^2/(16n^5)$$

which proves equation (7.37).

Let us now suppose that $|T_2| \geq \delta N/2$. Using the triangle inequality and equation (7.36), we have

$$\begin{aligned} |T_2| &\leq \sum_{V \leq |d| \leq N/U} \tau(d) \left| \sum_{w \in G_{n-k}} \mu(w)\Phi(dw) \right| \\ &\leq nN \max_{v \leq k \leq n-u} \mathbb{E}_{d \in A_k} \tau(d) |\mathbb{E}_{w \in G_{n-k}} \mu(w)\Phi(dw)|. \end{aligned}$$

We again fix an integer k, this time $k \in [n/18, 17n/18]$, that realises the maximum, and apply Cauchy-Schwarz together with Lemma E.2, obtaining

$$|T_2|^2/N^2 \leq 4n^5\, \mathbb{E}_{d \in A_k} \mathbb{E}_{w,w' \in G_{n-k}} \mu(w)\mu(w')\Phi(dw)\overline{\Phi(dw')}.$$

It follows that

$$\mathbb{E}_{w,w' \in G_{n-k}}\, \mu(w)\mu(w') \mathbb{E}_{d \in A_k} \Phi(dw)\overline{\Phi(dw')} \geq \delta^2/(16n^5).$$

Applying Cauchy-Schwarz again to eliminate μ yields

$$\mathbb{E}_{w,w' \in G_{n-k}} \mathbb{E}_{d,d' \in A_k} \Phi(dw)\overline{\Phi(dw')\Phi(d'w)}\Phi(d'w') \geq \delta^4/(256n^{10}).$$

This is the content of clause (7.38), so the proof of Proposition 7.11 is complete. □

We now derive Proposition 7.11 using Proposition 7.12. Suppose first that equation (7.37) holds, so there is $k \leq n/9$ such that

$$\mathbb{E}_{d \in A_k} |\mathbb{E}_{w \in G_{n-k}} \Phi(dw)|^2 \geq \delta', \tag{7.39}$$

where $\delta' = \delta^2/(16n^5)$. Equation (7.39) implies that there exists $d \in A_k$ such that

$$|\mathbb{E}_{w \in G_{n-k}} \Phi(dw)|^2 \geq \delta'.$$

Fix such a $d \in A_k$. Lemma 7.10 now implies that the quadratic polynomial $w \mapsto P(dw)$ has rank at most $\log_q(\delta'^{-1}) = O(\log(n/\delta))$. This corresponds exactly to the first statement of Proposition 7.11.

Suppose instead that equation (7.38) holds. Then we have $k \in [n/18, 17n/18]$ such that

$$\mathbb{E}_{w,w' \in G_{n-k}} \mathbb{E}_{d,d' \in A_k} \Phi(dw)\overline{\Phi(dw')\Phi(d'w)}\Phi(d'w') \geq \delta',$$

where $\delta' = \delta^4/(256n^{10})$. The triangle inequality ensures that

$$\mathbb{E}_{d,d' \in A_k} |\mathbb{E}_{w \in G_{n-k}} \chi(P(dw) - P(d'w))| \geq \delta'.$$

In particular, for a proportion at least $\delta'/2$ of pairs of monic polynomials d, d' of degree k, we have

$$|\mathbb{E}_{w \in G_{n-k}} \chi(P(dw) - P(d'w))| \geq \delta'/2,$$

which implies that the rank of $w \mapsto P(dw) - P(d'w)$ is at most $-\log_q(\delta'/2) = O(\log(n/\delta))$. This is precisely the second part of Proposition 7.11. So in every case, Proposition 7.11 holds.

7.6 Using the polylogarithmic bilinear Bogolyubov conjecture

Let $A > 0$ be arbitrary, and let $\delta = n^{-A}$. To prove Theorem 7.3, it suffices to show that $|\sum_{f \in G_n} \mu(f)\Phi(f)| < \delta q^n$ for n sufficiently large. For the sake of contradiction, suppose instead that there exists an unbounded set Z of integers n such that

$$\left| \sum_{f \in G_n} \mu(f)\Phi(f) \right| \geq \delta q^n \tag{7.40}$$

whenever $n \in Z$. We then apply Proposition 7.11. Suppose the first alternative holds. Write $P(f) = B(f, f)$ for some bilinear form $B(x, y)$ on $\mathbb{F}_q^n \times \mathbb{F}_q^n$ (we may omit the linear part of P as it modifies the rank by at most 1). Then we know that the form $R_d : w \mapsto P(dw)$ on G_{n-k} has small rank for at least one d of some degree $0 \leq k \leq n/9$. Now the rank of the quadratic form R_d is simply the rank of the bilinear form B restricted to the subspace $dG_{n-k} \subset G_n$ of codimension k. Thus the rank of R_d is at least $\operatorname{rk} B - 2k$, which implies that $\operatorname{rk} B \leq 2n/9 + c \log n$. If $n \in Z$ is large enough, this bound on the rank is less than $c'n$ for some $c' < 1/4$. Corollary 7.9 now yields the desired contradiction.

In Appendix D, we show how to deal with the Type I sum for k up to $n/2 - o(n)$. As a result, we only need to consider the second alternative given by Proposition 7.11 with k in $[n/4 - o(n), 3n/4 + o(n)]$. Unfortunately, we were not able to use this shortened range, which is why we relegated this argument to an appendix.

Now let us suppose that the second alternative of Proposition 7.11 holds. Let $n/18 \leq$

$k \leq 17n/18$ be the parameter returned by this proposition. It follows that the set

$$Y = \{(d, d') \in A_k^2 \mid w \mapsto P(dw) - P(d'w) \text{ has rank at most } \gamma \log n\}$$

has size at least q^{2k+2}/k^γ for some constant $\gamma > 0$. Note that for $d, d' \in G_{k+1}$,

$$P(dw) - P(d'w) = B((d - d')w, (d + d')w)$$

is a quadratic polynomial in $w \in G_{n-k}$. For $a, b \in G_{k+1}$, let $B_{a,b}$ be the symmetric bilinear form on $\mathbb{F}_q^{n-k} \times \mathbb{F}_q^{n-k}$ (identified with $G_{n-k} \times G_{n-k}$) defined by $B_{a,b}(x, y) = (B(ax, by) + B(ay, bx))/2$. Thus we have a set

$$X = \{(a, b) \in G_{k+1} \times G_{k+1} \mid \operatorname{rk} B_{a,b} \leq \gamma \log n\}$$

of density at least $\eta = k^{-\gamma}$ in $G_{k+1} \times G_{k+1}$. As discussed in Section 7.1, we would like to replace the large set X by a more structured set, namely the zero set of a (not too large) family of bilinear forms, at the cost of slightly worsening the bounds on the rank. Corollary 5.10, an application of the bilinear Bogolyubov theorem from Chapter 5, precisely implies that

$$X' = \{(a, b) \in G_{k+1} \times G_{k+1} \mid \operatorname{rk} B_{a,b} \leq 64\gamma \log n\}$$

contains a set of the form

$$Y = \{(a, b) \in W_1 \times W_2 \mid F_1(a, b) = \ldots = F_r(a, b) = 0\},$$

where W_1, W_2 are $\mathbb{F}_p$-subspaces of G_{k+1} (itself viewed as an $\mathbb{F}_p$-vector space of dimension $s(k+1) = O(k)$) of codimension at most $r = c(\eta)$, and $F_1, \ldots, F_r$ are $\mathbb{F}_p$-bilinear forms on $W_1 \times W_2$. Under Conjecture 5.9, $c(\eta) = O(\log^{O(1)} \eta^{-1}) = O(\log^{O(1)} k)$, while the unconditional bound we have is unfortunately useless when η^{-1} is polynomial in n. Henceforth we assume that Conjecture 5.9 is true.

Now take $\epsilon = 1/10$ and consider a set of indices

$$I = \{0 = i_1 < i_2 < \cdots < i_m = \lfloor k - \epsilon k \rfloor\} \subset [0, k - \epsilon k]$$

such that $i_{j+1} - i_j < (n - k)/2$ for any j and $m = O(1)$. Such a set exists because $n - k \geq n/18 \geq k/18$. Consider $W = W_1 \cap W_2 \cap G_{\epsilon k}$, which is an $\mathbb{F}_p$-vector space of dimension at least $\epsilon sk - O(\log^{O(1)} k)$. Consider the $\mathbb{F}_p$-quadratic forms on W given by

$F_l^{i,j}(w) = F_l(t^i w, t^j w)$ for any $l \in [m]$ and $i, j \in I$, where the map $w \mapsto t^i w$ is identified with the corresponding $\mathbb{F}_p$-linear map between the vectors of coefficients. This is still a family of at most $O(\log^{O(1)} k)$ bilinear forms, so we can find at least $\Omega(p^{\epsilon sk - O(\log^{O(1)} k)})$ vectors in $G_{\epsilon k}$ which are isotropic for all of these forms, thanks to Lemma 5.11. In particular, if k (and thus n) is large enough, there is definitely at least one nonzero polynomial w of degree at most ϵk such that $F_l(t^i w, t^j w) = 0$ for all $i, j \in I$ and $l \in [m]$. Consequently, $\operatorname{rk} B_{t^i w, t^j w} \leq \kappa := 64\gamma \log n$ for all $i, j \in I$.

Consider the (symmetric) matrix M of the $\mathbb{F}_q$-bilinear form B restricted to the space of the multiples of w, written in the basis $(wt^i)_{0 \leq i < n - \deg w}$. We refer to the matrix element $B(wt^i, wt^j)$ as the *cell* (i, j) of M. The rank of B differs from the rank of M by at most $2\epsilon n$, so it suffices to bound the rank of M.

Now let us examine the (symmetric) matrix $N_{i,j}$ of the quadratic form $B_{t^i w, t^j w}$ in the canonical basis of G_{n-k}.

Observe that the map $w \mapsto t^i w$, viewed as an $\mathbb{F}_q$-linear map (between vectors of coefficients), transforms an element t^j of the canonical basis of G_{n-k} into a basis element $t^{i+j} w$. This means that its matrix in the canonical basis of G_{n-k} and the basis $(wt^i)_{0 \leq i < n - \deg w}$ is an $(n - \deg w) \times (n - k)$ matrix which we can write by block as

$$L_{t^i w} = \begin{pmatrix} 0 \\ I_{n-k} \\ 0 \end{pmatrix},$$

where the central block is an identity block of size $(n - k) \times (n - k)$ and the other blocks are 0 blocks. Here by a *block* we mean a submatrix consisting of consecutive rows and columns of a matrix. Next we observe that

$$2N_{i,j} = L_{t^i w}^T M L_{t^j w} + L_{t^j w}^T M L_{t^i w},$$

which makes it easy to see that $N_{i,j}$ is the symmetric part of the $(n-k) \times (n-k)$ block of M whose top-left corner is the (i, j) cell of M. Write $M_{i,j}$ for this block, so $2N_{i,j} = M_{i,j} + M_{i,j}^T$.

We remark that if $i = j$, then $M_{i,j}$ is a diagonal block of the symmetric matrix M, hence a symmetric matrix, so $M_{i,j} = N_{i,j}$ must have small rank itself. Hence, the matrix M contains a number of large diagonal blocks $M_{i,i}$ which have small rank. To bound the rank of M, it suffices to bound the ranks of all submatrices $M_{i,j}$ for $(i, j) \in I^2$. Indeed, the

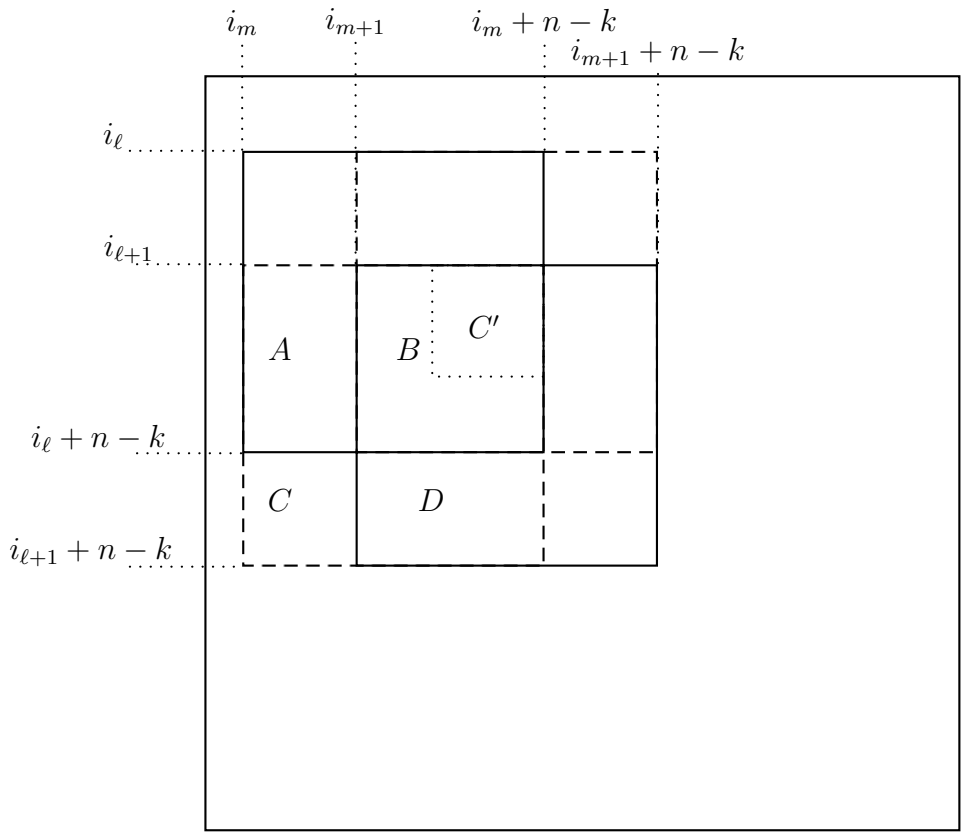

Figure 7.1: Covering M by submatrices and moving away from the diagonal

matrix M being covered by these submatrices, we have the bound

$$\operatorname{rk} M \leq \sum_{(i,j)\in I^2} \operatorname{rk} M_{i,j} \leq |I|^2 \max_{(i,j)\in I^2} \operatorname{rk} M_{i,j}.$$

The cardinality $|I|$ being bounded, bounding the ranks of these blocks suffices to bound $\operatorname{rk} M$. We now prove by induction on $\ell - m$ that M_{i_ℓ,i_m} has small rank, namely at most $5^{\ell-m}\kappa$. Because $M_{i_\ell,i_m} = M^T_{i_m,i_\ell}$, it suffices to prove the claim in the case $\ell \geq m$. When $\ell - m = 0$, as we have already seen, the corresponding block is diagonal and of rank at most κ. We now suppose that for some $\ell \geq m$ we already know that $\operatorname{rk} M_{i_\ell,i_m} \leq 5^{\ell-m}\kappa$ and we inspect $M_{i_{\ell+1},i_m}$. The reader may wish to consult Figure 7.1 while following through the proof.

In Figure 7.1 the dotted $(n-k)\times(n-k)$ block $M_{i_{\ell+1},i_m} = E$ is made up of the four blocks A, B, C, D, and is known to have a symmetric part of small rank. On the other hand, A, B and D are already known to have rank at most $5^{\ell-m}\kappa$, because they are submatrices of M_{i_ℓ,i_m} and $M_{i_{\ell+1},i_{m+1}}$ respectively. Now the symmetric part $E + E^T$ admits as bottom-left square block of the size of C the matrix $C + C'^T$, where C' is the top-right block of B (here it is crucial that $i_{\ell+1} - i_\ell < (n-k)/2$). As a submatrix of a matrix of small rank, $C + C'^T$ must have small rank. But C' has iself small rank as a submatrix of B, whence it follows that $C = (C + C'^T) - C'^T$ has small rank, namely a rank at most $2 \cdot 5^{\ell-m}\kappa$. Hence

$$\operatorname{rk} M_{i_{\ell+1},i_m} = \operatorname{rk} E \leq \operatorname{rk} A + \operatorname{rk} B + \operatorname{rk} C + \operatorname{rk} D \leq 5^{\ell+1-m}\kappa.$$

This completes the inductive proof, and implies that $\operatorname{rk} M = O(\kappa) = O(\log n)$.

Finally, as already noted, the rank of B is at most the rank of M plus $2\epsilon n$. In particular, given that $2\epsilon = 1/5$, it is surely less than $c'n$ for some $c' < n/4$, if $n \in Z$ is large enough. Again invoking Corollary 7.9, we obtain the desired contradiction with the hypothesis (7.40). This concludes the proof of Theorem 7.3.

7.7 The Hankel case

We prove Theorem 7.4, again assuming $p > 2$. If $\alpha = \sum_{j=-\infty}^{m} a_j t^j$ then the matrix of the quadratic form $f \mapsto (\alpha f^2)_{-1}$ in the canonical basis of G_n is

$$M = M(\alpha) = \begin{pmatrix} a_{-1} & a_{-2} & \cdots & a_{-n} \\ a_{-2} & \iddots & \iddots & a_{-n-1} \\ \vdots & \iddots & \iddots & \vdots \\ a_{-n} & \alpha_{-n-1} & \cdots & a_{-2n+1} \end{pmatrix}.$$

We will follow the same strategy as in Sections 7.5 and 7.6 with $\Phi(f) = e(\alpha f^2 + \beta f)$. Suppose for a contradiction that, for arbitrarily large n, we have

$$\sum_{f \in G_n} \mu(f)\Phi(f) > \delta q^n \tag{7.41}$$

with $\delta = q^{-\epsilon' n}$ for some $\epsilon' > 0$ to be determined later. Applying Proposition 7.11, we may discard the first alternative, because in that case the reasoning of Section 7.6 goes

through without Conjecture 5.9. The parameter $\delta' = (\delta/n)^{O(1)}$ is still at least $q^{-\epsilon n}$ for some $\epsilon = O(\epsilon')$, if n is large enough. Thus we find a $k \in [n/18, 17n/18]$ such that for at least $q^{(2-\epsilon)(k+1)}$ pairs of polynomials (d, d') of degree k, the quadratic phase on G_{n-k} defined by

$$w \mapsto e(\alpha(d^2 - d'^2)w^2)$$

has rank at most $O(\epsilon n)$. Write $d - d' = a$ and $d + d' = b$. We infer that for at least $q^{(2-\epsilon)(k+1)}$ pairs of polynomials a, b of degree at most k, the quadratic phase

$$w \mapsto e(\alpha abw^2)$$

has rank at most $c\epsilon n$ for some constant $c = O(1)$.

With the notation of the previous section, the relevant symmetric matrix is

$$M_{a,b} = L_a^T M(\alpha) L_b = L_b^T M(\alpha) L_a = M(\alpha ab).$$

In contrast to the general case, $M_{a,b}$ is here a product involving M and not a sum of two products, which makes it much easier to analyse. As in the proof of Theorem 7.3, we will show that M has low rank by covering it by submatrices of low rank.

By Markov's inequality, there exists a set $X \subset G_{k+1}$ of size $q^{(1-\epsilon)(k+1)}/2$ such that for any $a \in X$, the set

$$B_a := \{b \in G_{k+1} \mid \operatorname{rk} M_{a,b} \leq c\epsilon n\}$$

has size at least $q^{(1-\epsilon)(k+1)}/2$.

Let $\eta = 2\epsilon$. For any $i \in \{0, \ldots, k - \eta k\}$ and $a \in X$, by the pigeonhole principle, there exist two distinct $b \neq b'$ in B_a such that $f = b' - b = \sum_{m=i}^{i+\eta k} c_m t^m$ for some coefficients c_m. Moreover, we have $\operatorname{rk} M_{a,f} \leq 2c\epsilon n$. Write $f = f_{a,i}$ to emphasize the dependence. Fix $(i, j) \in \{0, \ldots, k - 2\eta k\}^2$. Again, the pigeonhole principle implies that there exist $a \neq a' \in X$ such that $g = a - a' \in \operatorname{span}(t^j, \ldots, t^{j+2\eta k})$ and $f_{a,i} = f_{a',i}$. If f is this common value, we have $\operatorname{rk} M_{g,f} = O(\epsilon n)$. Observe that for such a pair (g, f) we have

$$L_g = \begin{pmatrix} 0 \\ C_g \\ 0 \end{pmatrix},$$

where the central block is a $(n - k + 2\eta k) \times (n - k)$ matrix of rank $n - k$ and the other

blocks are 0 blocks. The same decomposition holds for L_f, with a central block C_f. So if N is the $(n-k+2\eta k)\times(n-k+2\eta k)$ block of M whose top-left cell is (j,i), then $M_{g,f}=C_g^T N C_f$, and thus $\operatorname{rk} M_{g,f}\geq \operatorname{rk} N-4\eta k$. As a result, $\operatorname{rk} N=O(\epsilon n)$.

Covering M by a bounded number of blocks of size $(n-k+2\eta k)\times(n-k+2\eta k)$, we find that $\operatorname{rk} M=O(\epsilon n)$. By taking ϵ small enough, the bound $O(\epsilon n)$ is constrained to be smaller than, say, $n/5$, for sufficiently large n. Thus if ϵ is small enough (that is, if ϵ' is small enough), we get a contradiction between the hypothesis (7.41) and Corollary 7.9. Theorem 7.4 follows.

Appendix A

Volume packing arguments and local divisor density

In this appendix, we shall collect some frequently used facts concerning the number of integral solutions to a system of linear equations in a convex set of $\mathbb{R}^d$ and in $(\mathbb{Z}/m\mathbb{Z})^d$. The first lemma we state is borrowed from Green and Tao [45, Appendix A].

Lemma A.1. *Let $K \subset [0, N]^d$ be a convex body of $\mathbb{R}^d$. Then*

$$|K \cap \mathbb{Z}^d| = \sum_{n \in K \cap Z^d} 1 = \mathrm{Vol}(K) + O_d(N^{d-1}).$$

We recall the definition of the local divisor density from equation (3.21) (already present in [45]) and we mention some useful properties.

Definition A.1. For a given system of affine-linear forms $\Psi = (\psi_1, \ldots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$, positive integers $d_1, \ldots, d_t$ of lcm m, define the *local divisor density* by

$$\alpha_\Psi(d_1, \ldots, d_t) = \mathbb{E}_{n \in (\mathbb{Z}/m\mathbb{Z})^d} \prod_{i=1}^{t} 1_{\psi_i(n) \equiv 0 \bmod d_i}.$$

The following lemma is borrowed from Matthiesen [68, Lemma 9.3].

Lemma A.2. *Let $K \subset [-B, B]^d$ be a convex body and Ψ a system of affine-linear forms, and let $d_1, \ldots, d_t$ be integers of lcm m. Then*

$$\sum_{n \in \mathbb{Z}^d \cap K} \prod 1_{d_i | \psi_i(n)} = \mathrm{Vol}(K) \alpha_\Psi(d_1, \ldots, d_t) + O(B^{d-1} m).$$

We shall try to bound $\alpha_\Psi(p^{a_1}, \dots, p^{a_t})$. To this aim, we state a version of Hensel's lemma in several variables.

Lemma A.3. *Let $Q \in \mathbb{Z}[X_1, \dots, X_d]$, p be a prime and $k \geq 1$ an integer and $x \in (\mathbb{Z}/p^k\mathbb{Z})^d$ such that $Q(x) \equiv 0 \bmod p^k$ and*

$$\nabla Q(x) = \left(\frac{\partial Q}{\partial x_1}, \dots, \frac{\partial Q}{\partial x_d}\right)(x) \not\equiv 0 \bmod p.$$

Then there exist precisely p^{d-1} vectors $y \in (\mathbb{Z}/p^{k+1}\mathbb{Z})^d$ such that $x \equiv y \bmod p^k$ and $Q(y) \equiv 0 \bmod p^{k+1}$.

Proof. Let $y \in (\mathbb{Z}/p^{k+1}\mathbb{Z})^d$ satisfy $x \equiv y \bmod p^k$; in other words, $y = x + p^k z$ for some uniquely determined $z \in (\mathbb{Z}/p\mathbb{Z})^d$. Here, by abuse of notation, we replace $x \in (\mathbb{Z}/p^k\mathbb{Z})^d$ by some fixed lift in $(\mathbb{Z}/p^{k+1}\mathbb{Z})^d$. We then treat $Q(x)$ as an element of $\mathbb{Z}/p^{k+1}\mathbb{Z}$ congruent to $0 \bmod p^k$ and put $Q(x) = p^k a$ with $a \in \mathbb{Z}/p\mathbb{Z}$. Then Taylor's formula ensures that

$$Q(y) \equiv Q(x) + p^k \nabla Q(x) \cdot z \equiv p^k(a + \nabla Q(x) \cdot z) \bmod p^{k+1}.$$

So $Q(y) \equiv 0 \bmod p^{k+1}$ is equivalent to $a + \nabla Q(x) \cdot z \equiv 0 \bmod p$. As $\nabla Q(x)$ is not zero modulo p, this imposes a nontrivial affine equation on z in the vector space $\mathbb{F}_p^d$, so z is constrained to lie in a $(d-1)$-dimensional affine $\mathbb{F}_p$-subspace, which has p^{d-1} elements, hence the conclusion. □

As an application, we prove the following statement.

Corollary A.4. *Let ψ be an affine-linear form in d variables, and let p be a prime such that ψ is not the trivial form modulo p. Then for any $m \geq 1$*

$$\alpha_m = \alpha_\psi(p^m) = \mathbb{E}_{n \in (\mathbb{Z}/p^m\mathbb{Z})^d} 1_{p^m | \psi(n)} = \mathbb{P}_{n \in (\mathbb{Z}/p^m\mathbb{Z})^d}(\psi(n) = 0) \leq p^{-m}.$$

Proof. If $n \in (\mathbb{Z}/p^m\mathbb{Z})^d$ satisfies $\psi(n) \equiv 0 \bmod p^m$, then in particular $\tilde{\psi}(\tilde{n}) \equiv 0 \bmod p$, where $\tilde{\cdot}$ is the reduction modulo p, which imposes that $\tilde{n}$ lies in $\ker \tilde{\psi}$. By assumption, $\tilde{\psi} \neq 0$. If its linear part is 0, then its constant part is nonzero, thus $\ker \tilde{\psi} = \emptyset$ and $\alpha_m = 0$. Otherwise, the linear part is nonzero modulo p, and then $\ker \tilde{\psi}$ is an affine $\mathbb{F}_p$-hyperplane, thus has p^{d-1} elements. Let us prove the proposition by induction on m. For $m = 1$, we have just proved the result. Suppose now that $\alpha_m \leq p^{-m}$ for some $m \geq 1$. Because of the assumption above, $\nabla\psi$ is a constant vector which is nonzero modulo p. Applying Lemma

A.3 for $k = m$, we find that each zero modulo p^m of ψ gives rise to exactly p^{d-1} zeros modulo p^{m+1}, which proves that $\alpha_{m+1} \leq p^{-(m+1)}$. This concludes the induction step and the proof. $\square$

Exploiting this corollary, we can now prove a bound on more general local densities.

Proposition A.5. *Let $\Psi = (\psi_1, \ldots, \psi_t)$ be a system of integral affine linear forms in d variables and p be a prime so that the system reduced modulo p is of finite complexity, i.e. no two of the forms are affinely related modulo p. Then*

$$\alpha = \alpha_\Psi(p^{a_1}, \ldots, p^{a_t}) \leq p^{-\max_{i \neq j}(a_i + a_j)}.$$

Proof. If all a_i are zero, the result is trivial, so let $m = \max a_i$ and suppose $m \geq 1$; let $i \neq j$ be such that $a_i + a_j$ is maximal (in particular, it is at least m). Suppose first that either a_i or a_j is 0. Without loss of generality, suppose $a_i = 0$ and $a_j \neq 0$. Then for $n \in (\mathbb{Z}/p^m\mathbb{Z})^d$ to satisfy $\psi_k(n) \equiv 0 \bmod p^{a_k}$ for all $k = 1, \ldots, t$, we must have in particular $\tilde{\psi_j}(\tilde{n}) \equiv 0 \bmod p^{a_j}$, and using Corollary A.4, we find that

$$\alpha = \mathbb{E}_{n \in (\mathbb{Z}/p^m\mathbb{Z})^d} \prod_{i \in [t]} 1_{p^{a_i} | \phi_i(n)} \leq \mathbb{E}_{n \in (\mathbb{Z}/p^{a_j}\mathbb{Z})^d} 1_{p^{a_j} | \phi_j(n)} \leq p^{-a_j} = p^{-\max_{i \neq j}(a_i + a_j)}.$$

Now suppose $1 \leq a_i \leq a_j$. Then for $n \in (\mathbb{Z}/p^m\mathbb{Z})^d$ to satisfy $\psi_k(n) \equiv 0 \bmod p^{a_k}$ for all $k = 1, \ldots, t$, we must have in particular $\tilde{\psi_i}(\tilde{n}) \equiv \tilde{\psi_j}(\tilde{n}) \equiv 0 \bmod p$. This imposes that $\tilde{n}$ lies in the intersection of two affine $\mathbb{F}_p$-subspaces, namely $\ker \tilde{\psi_i}$ and $\ker \tilde{\psi_j}$, which are two nonparallel hyperplanes because these forms are affinely independent by assumption. Now we use induction on $m \geq 1$ to show that

$$\beta_m = \mathbb{P}_{n \in (\mathbb{Z}/p^m\mathbb{Z})^d}(\psi_i(n) \equiv \psi_j(n) \equiv 0 \bmod p^m) = p^{-2m}.$$

For $m = 1$, what we have seen above implies that $\beta_1 = p^{-2}$ (the intersection of two nonparallel affine hyperplanes of $\mathbb{F}_p^d$ is an affine subspace of dimension $d-2$, so its cardinality is p^{d-2}), so the statement is true. Suppose now that for some $m \geq 1$ we have $\beta_m = p^{-2m}$. If $x \in (\mathbb{Z}/p^m\mathbb{Z})^d$ satisfies $\psi_i(x) \equiv \psi_j(x) \equiv 0 \bmod p^m$ and if $y = x + p^m z \in (\mathbb{Z}/p^{m+1}\mathbb{Z})^d$ for some $z \in (\mathbb{Z}/p\mathbb{Z})^d$ satisfies $\psi_i(y) \equiv \psi_j(y) \equiv 0 \bmod p^{m+1}$, then following the proof of Lemma A.3, we infer that z has to satisfy two affine equations

$$a + \nabla\psi_i \cdot z \equiv 0 \bmod p \quad \text{and} \quad a + \nabla\psi_j \cdot z \equiv 0 \bmod p.$$

This forces z to lie in the intersection of two nonparallel affine $\mathbb{F}_p$-hyperplanes of $\mathbb{F}_p^d$ (they are nonparallel because we supposed that the gradients were not proportional). Hence for a fixed x as above, there are p^{d-2} such y, so finally $\beta_{m+1} = p^{d-2}\beta_m$ whence the conclusion. In particular, putting $m = a_i$, we have that $\mathbb{E}_{n\in(\mathbb{Z}/p^{a_i}\mathbb{Z})^d} 1_{\phi_i(n)\equiv\phi_j(n)\equiv 0 \bmod p^{a_i}} \leq p^{-2a_i}$. It remains to induct on $a_j - a_i \geq 0$ using Lemma A.3 in order to find that

$$\mathbb{E}_{n\in(\mathbb{Z}/p^{a_j}\mathbb{Z})^d} 1_{p^{a_i}|\phi_i(n)} 1_{p^{a_j}|\phi_j(n)} \leq p^{-(a_i+a_j)},$$

which implies the desired result. □

We prove another statement which is helpful during the proof of the linear forms conditions (Proposition C.1).

Proposition A.6. *Let $\Phi : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of affine-linear forms. Let p be a prime such that the reduction modulo p of the system is of finite complexity. Let $K \subset [-B, B]^d$ be a convex body. Then*

$$\sum_{n\in K\cap\mathbb{Z}^d} 1_{p^2|\prod_{i\in[t]}\phi_i(n)} \ll_t p^{-2}\mathrm{Vol}(K) + B^{d-1}p^2.$$

Proof. First, we observe that $p^2 \mid \prod_{i\in[t]}\phi_i(n)$ implies that either there exists $i \in [t]$ such that $p^2 \mid \phi_i(n)$ or there exist $i \neq j$ such that $p \mid \phi_i(n)$ and $p \mid \phi_j(n)$. Hence

$$\sum_{n\in K\cap\mathbb{Z}^d} 1_{p^2|\prod_{i\in[t]}\phi_i(n)} \leq \sum_{i\in[t]} \sum_{n\in K\cap\mathbb{Z}^d} 1_{p^2|\phi_i(n)} + \sum_{i\neq j} \sum_{n\in K\cap\mathbb{Z}^d} 1_{\phi_i(n)\equiv\phi_j(n)\equiv 0 \bmod p}.$$

Now for any $i \in [t]$ we apply Lemma A.2 which implies

$$\sum_{n\in K\cap\mathbb{Z}^d} 1_{p^2|\phi_i(n)} = \mathrm{Vol}(K)\alpha_{\phi_i}(p^2) + O(B^{d-1}p^2)$$

and for any $i \neq j$

$$\sum_{n\in K\cap\mathbb{Z}^d} 1_{\phi_i(n)\equiv\phi_j(n)\equiv 0 \bmod p} = \mathrm{Vol}(K)\alpha_{\phi_i,\phi_j}(p,p) + O(B^{d-1}p).$$

But the assumption of finite complexity modulo p means that we may invoke Proposition A.5, which implies that $\alpha_{\phi_i}(p^2) \leq p^{-2}$ and that $\alpha_{\phi_i,\phi_j}(p,p) \leq p^{-2}$. The result then follows. □

Appendix B

Analysis of the local factors β_p

This appendix deals with the local factors appearing in Chapter 4.5. First, we check that the limit defining β_p in Theorem 4.1 exists. We fix integers $d, t, s \geq 1$ and a system of linear forms $\Psi : \mathbb{Z}^d \to \mathbb{Z}^{t+s}$ of finite complexity, and we suppose its linear coefficients are bounded by L.

We also fix PDBQFs $f_{t+1}, \ldots, f_{t+s}$ of discriminants $D_{t+1}, \ldots, D_{t+s}$; these notions and the notation ρ_{f_j} were defined in the introduction. Let p be a fixed prime and $M_0 = \max_j v_p(D_j)$. For $m \geq 1$ an integer and $a \in (\mathbb{Z}/p^m\mathbb{Z})^d$, let

$$P_m(a) = \prod_{i=1}^{t} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(a)) \prod_{j=t+1}^{t+s} \frac{\rho_{f_j,\psi_j(a)}(p^m)}{p^m}. \tag{B.1}$$

Finally, let $\beta_p(m) = \mathbb{E}_{a\in(\mathbb{Z}/p^m\mathbb{Z})^d} P_m(a)$. Thus we want to prove that $\beta_p(m)$ is convergent as m tends to ∞. This is a consequence of the following proposition

Proposition B.1. *The sequence $(\beta_p(m))_{m\in\mathbb{N}}$ is a Cauchy sequence. More precisely, there exists $M_0 = M_0(D_{t+1}, \ldots, D_{t+s})$ so that for all integers $m_0 \geq M_0$ and $m, n \geq m_0$, we have*

$$\beta_p(m) - \beta_p(n) = O(m_0^s p^{-m_0/2}).$$

In particular, this sequence has a limit β_p and we have

$$\beta_p(m) = \beta_p + O(m^s p^{-m/2}).$$

Proof. Let $m_0 \geq M_0$ and $m, n \geq m_0$. We split $(\mathbb{Z}/p^m\mathbb{Z})^d$ into two parts

$$A_1 = A_1(m, m_0) = \{a \in (\mathbb{Z}/p^m\mathbb{Z})^d \mid \forall j \in [\![t+1\,;\, t+s]\!] \quad \psi_j(a) \not\equiv 0 \bmod p^{m_0}\}$$

and

$$A_2 = A_2(m, m_0) = \{a \in (\mathbb{Z}/p^m\mathbb{Z})^d \mid \exists j \in [\![t+1\,;\, t+s]\!] \quad \psi_j(a) \equiv 0 \bmod p^{m_0}\}.$$

Thus

$$\beta_p(m) = \mathbb{E}_{a\in(\mathbb{Z}/p^m\mathbb{Z})^d} P_m(a) 1_{A_1(m,m_0)}(a) + \mathbb{E}_{a\in(\mathbb{Z}/p^m\mathbb{Z})^d} P_m(a) 1_{A_2(m,m_0)}(a). \tag{B.2}$$

For the first term, we use the lift-invariance property [68, Corollary 6.4] already stated in Lemma 4.6. It implies that

$$\mathbb{E}_{a\in(\mathbb{Z}/p^m\mathbb{Z})^d} P_m(a) 1_{A_1(m,m_0)} = \mathbb{E}_{a\in(\mathbb{Z}/p^{m_0}\mathbb{Z})^d} P_{m_0}(a) 1_{A_1(m_0,m_0)}$$

thus the first term on the right-hand side of (B.2) does not depend on m. For the second term, we invoke the following general bound from [68] (see Lemma 6.3 and the proof of Lemma 8.2)

$$\frac{\rho_{f_j,\psi_j(a)}(p^m)}{p^m} \ll \sum_{k=0}^{m} 1_{\psi_j(a)\equiv 0 \bmod p^k}.$$

We also use the trivial bound $\Lambda_{\mathbb{Z}/p\mathbb{Z}} \leq 2$ to infer the inequalities

$$\begin{aligned}
P_m(a) 1_{A_2}(a) &\ll 2^t 1_{A_2}(a) \prod_{j=t+1}^{t+s} \sum_{k=0}^{m} 1_{\psi_j(a)\equiv 0 \bmod p^k} \\
&\ll m_0^s 1_{A_2}(a) + 1_{A_2}(a) \sum_{\substack{0\leq k_{t+1},\dots,k_{t+s}\leq m \\ \max k_i \geq m_0}} \prod_{j=t+1}^{t+s} 1_{\psi_j(a)\equiv 0 \bmod p^{k_j}} \\
&\leq (m_0^s + 1) \sum_{\substack{0\leq k_{t+1},\dots,k_{t+s}\leq m \\ \max k_i \geq m_0}} \prod_{j=t+1}^{t+s} 1_{\psi_j(a)\equiv 0 \bmod p^{k_j}}.
\end{aligned}$$

Here the factor m_0^s appears as the number of s-tuples whose entries are all in $[\![0\,;\, m_0 - 1]\!]$; moreover, the 2^t is merged with the implied constant, which crucially remains independent of m or m_0. The third line follows from the fact that if $a \in A_2$, then the sum over tuples

k_i whose maximum is at least m_0 is at least 1. We then average over a and let

$$Z = (\zeta_1, \dots, \zeta_s) = (\psi_{t+1}, \dots, \psi_{t+s}) \tag{B.3}$$

be the system of the last s linear forms of Ψ, obtaining

$$\mathbb{E}_{a\in(\mathbb{Z}/p^m\mathbb{Z})^d} P_m(a) 1_{A_2}(a) \ll m_0^s \sum_{\substack{0\le k_1,\dots,k_s\le m\\ M:=\max k_i\ge m_0}} \mathbb{E}_{a\in(\mathbb{Z}/p^M\mathbb{Z})^d} \prod_{i=1}^{s} 1_{p^{k_i}|\zeta_i}. \tag{B.4}$$

We recognise the local divisor density α_Z on the right-hand side, so we put

$$\delta_p = \sum_{\substack{0\le k_1,\dots,k_s\le m\\ M:=\max k_i\ge m_0}} \alpha_Z(p^{k_1}, \dots, p^{k_s}),$$

which enables us to rewrite (B.4) as

$$\mathbb{E}_{a\in(\mathbb{Z}/p^m\mathbb{Z})^d} P_m(a) 1_{A_2}(a) \ll m_0^s \delta_p.$$

Since the linear coefficients of Z are bounded and none of its forms is the trivial form, we see that for any $i \in [s]$, the maximal k such that ζ_i is the trivial form modulo p^k is bounded. Write then $\psi_i = p^k \psi_i'$ where ψ_i' is not the trivial form modulo p. Thus applying Corollary A.4 to the form ψ_i, we find that $\alpha_{\psi_i}(p^{k_i}) \le p^{k-k_i} \ll p^{-k_i}$ as k_i tends to ∞ while p and k are bounded. Further,

$$\alpha_Z(p^{k_1}, \dots, p^{k_s}) \le \min_j \alpha_{\psi_j}(p^{k_j}) \ll p^{-\max_j k_j},$$

and thus

$$\delta_p \ll \sum_{\substack{0\le k_{t+1},\dots,k_{t+s}\le m\\ M:=\max k_i\ge m_0}} p^{-M}.$$

Bounding the number of tuples $(k_1, \dots, k_s)$ satisfying $\max k_i = M$ crudely by $(M+1)^s$,

we conclude that

$$\begin{aligned}\delta_p &\ll \sum_{M\geq m_0} p^{-M}M^s \\ &\ll \sum_{M\geq m_0} p^{-M/2} \\ &\ll_p p^{-m_0/2}.\end{aligned}$$

Finally, this means that for $m \geq m_0$, we have

$$\beta_p(m) = \mathbb{E}_{a\in(\mathbb{Z}/p^{m_0}\mathbb{Z})^d} P(a) 1_{A_1(m_0,m_0)} + O(m_0^s p^{-m_0/2}). \tag{B.5}$$

The same holds for $\beta_p(n)$, hence

$$\beta_p(m) - \beta_p(n) = O(m_0^s p^{-m_0/2})$$

and the conclusion follows. □

We record a useful byproduct of the above proof.

Lemma B.2. *As $m \geq M_0$ tends to infinity, we have*

$$\mathbb{E}_{a\in(\mathbb{Z}/p^m\mathbb{Z})^d} \prod_{i=1}^{t} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(a)) \prod_{j=t+1}^{t+s} \frac{\rho_{f_j,\psi_j(a)}(p^m)}{p^m} 1_{\psi_j(a)\not\equiv 0 \bmod p^m} = \beta_p + O(p^{-m/3}).$$

Proof. We simply use equation (B.5) with $m = m_0$ and the bound $m^s p^{-m/2} \ll p^{-m/3}$, where the implied constant is independent of m and p. Together with the conclusion of Proposition B.1 that $\beta_p = \beta_p(m) + O(m^s p^{-m/2})$, this yields the desired result. □

We now analyse the behaviour of β_p as p tends to infinity.

Lemma B.3. *For primes p tending to infinity,*

$$\beta_p = 1 + O(p^{-2}).$$

Thus $\prod_p \beta_p$ is convergent and

$$\prod_{p\leq w(N)} \beta_p = \left(1 + O\left(\frac{1}{w(N)}\right)\right) \prod_p \beta_p.$$

Proof. Assume p is large enough so that p does not divide the product $D_{t+1} \cdots D_{t+s}$ of the (negative) discriminants of our quadratic forms.

Recall the notation $P_m(a)$ from (B.1) and the sets $A_1 = A_1(m,m)$ and $A_2 = A_2(m,m)$ introduced during the proof of Proposition B.1. As m tends to ∞, we have

$$\begin{aligned}
\beta_p + o(1) &= \mathbb{E}_{a\in(\mathbb{Z}/p^m\mathbb{Z})^d} P_m(a) \\
&= \mathbb{E}_{a\in(\mathbb{Z}/p^m\mathbb{Z})^d} P_m(a) 1_{A_1}(a) + \mathbb{E}_{a\in(\mathbb{Z}/p^m\mathbb{Z})^d} P_m(a) 1_{A_2}(a) \\
&= \frac{1}{p^{md}} \sum_{a\in A_1} P_m(a) + 2^t O(sm^s p^{-m}).
\end{aligned}$$

To get this error term, we used Corollary A.4 and the triangle inequality to bound $|A_2|$, and the fact that $\rho_{f_j,\beta}(p^m)/p^m \ll m$ [68, Lemma 6.3(c)] to bound $P_m(a)$. This error term tends to 0 as m tends to infinity, and thus merges with the $o(1)$ of the left-hand side. Let us now consider the main term. Thanks to the choice of p and the fact that the forms do not vanish at a mod p^m, we can use Lemma 6.3 from [68] which states that if f is a PDBQF of discriminant D, and if p is a prime which does not divide D, and if $\beta \not\equiv 0$ mod p^m, then

$$\frac{\rho_{f,\beta}(p^m)}{p^m} = (1 - \chi_D(p)p^{-1}) \sum_{k=0}^{m} 1_{p^k|m} \chi_D(p^k).$$

Here χ_D is a real character modulo p, namely the Kronecker symbol [68, Lemma 2.1]. Thus

$$\beta_p = \lim_{m\to\infty} \mathbb{E}_{a\in(\mathbb{Z}/p^m\mathbb{Z})^d} \prod_{i=1}^{t} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(a)) \prod_{j=t+1}^{t+s} \left((1 - \chi_{D_j}(p)p^{-1}) \sum_{k=0}^{m} 1_{p^k|\psi_j(a)} \chi_{D_j}(p^k) \right)$$

where we have obviously reintegrated the once excluded $a \in A_2$, because their sparsity ensures that they do not affect the limit. For $a \in (\mathbb{Z}/p^m\mathbb{Z})^d$, we then write $a = a' + pb$ with $b \in (\mathbb{Z}/p^{m-1}\mathbb{Z})^d$ and $a' \in [p]^d$. Thus the average $\mathbb{E}_a$ becomes

$$\prod_{j=t+1}^{t+s} \left(1 - \chi_{D_j}(p)p^{-1}\right) \mathbb{E}_{a\in[p]^d} \prod_{i=1}^{t} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(a)) \mathbb{E}_{b\in(\mathbb{Z}/p^{m-1}\mathbb{Z})^d} \prod_{j=t+1}^{t+s} \sum_{k=0}^{m} 1_{p^k|\psi_j(a+pb)} \chi_{D_j}(p^k). \tag{B.6}$$

We expand the product of sums as follows

$$\prod_{j=t+1}^{t+s}\sum_{k=0}^{m}1_{p^k|\psi_j(a+pb)}\chi_{D_j}(p^k)$$
$$= 1+\sum_{j}\sum_{k_j=1}^{m}1_{p^{k_j}|\psi_j(a+pb)}\chi_{D_j}(p^{k_j})+\sum_{\substack{0\le k_{t+1},\dots,k_{t+s}\le m\\ \text{at least two } k_i>0}}\prod 1_{p^{k_j}|\psi_j(a+pb)}\chi_{D_j}(p^{k_j})$$

according to whether we take no, one or several nonzero k. The expectation over a from (B.6) then splits into three terms. The first one is

$$\mathbb{E}_{a\in(\mathbb{Z}/p\mathbb{Z})^d}\prod_{i=1}^{t}\Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(a)),$$

and the second one is

$$\sum_{j=t+1}^{t+s}\sum_{k_j=1}^{m}\chi_{D_j}(p^{k_j})\mathbb{E}_{a\in[p]^d}\prod_{i=1}^{t}\Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(a))\mathbb{E}_{b\in(\mathbb{Z}/p^{m-1}\mathbb{Z})^d}1_{p^{k_j}|\psi_j(a+pb)}. \tag{B.7}$$

Now we decompose $\psi_j(a+pb)=\psi_j(a)+p\dot{\psi}_j(b)$, where $\dot{\psi}$ is the linear part of ψ. If p^{k_j} is to divide $\psi_j(a)+p\dot{\psi}_j(b)$, we need $p\mid\psi_j(a)$. Thus we can write, for each such a fixed, $\psi_j(a+pb)=p\tilde{\psi}_j(b)$, where $\tilde{\psi}_j$ is again an affine-linear form whose linear part is $\dot{\psi}_j$. We then need $p^{k_j-1}\mid\tilde{\psi}_j(b)$. Because of Corollary A.4,

$$\mathbb{E}_{b\in(\mathbb{Z}/p^{m-1}\mathbb{Z})^d}1_{p^{k_j-1}|\tilde{\psi}_j(b)}=p^{-k_j+1}$$

so the expression (B.7) equals

$$\sum_{j=t+1}^{t+s}\sum_{k_j=1}^{m}\chi_{D_j}(p^{k_j})p^{-k_j}\mathbb{E}_{a\in[p]^d}\prod_{i=1}^{t}\Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(a))p1_{p|\psi_j(a)}$$

To deal with the last term, which is

$$\mathbb{E}_{a\in(\mathbb{Z}/p^m\mathbb{Z})^d}\prod_{i=1}^{t}\Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(a))\sum_{\substack{0\le k_{t+1},\dots,k_{t+s}\le m\\ \text{at least two } k_i>0}}\prod_{j=t+1}^{t+s}1_{p^{k_j}|\psi_j(a)}\chi_{D_j}(p^{k_j}), \tag{B.8}$$

we crudely bound $\Lambda_{\mathbb{Z}/p\mathbb{Z}}$ by 2 and χ_{D_j} by 1. Recall the notation Z from (B.3). Thus as m

tends to infinity, the expression (B.8) is bounded above by a constant times

$$O\left(\sum_{\substack{k_1,\dots,k_s \\ \text{at least two } k_i>0}} \alpha_Z(p^{k_1},\dots,p^{k_s})\right)$$

To bound this expression, we remember that Z is a system of finite complexity. In particular, it has finitely many exceptional primes by Lemma 2.2. This implies, thanks to Proposition A.5, that for p large enough depending on s, d, L, we have

$$\alpha_Z(p^{k_1},\dots,p^{k_s}) \leq p^{-\max_{i\neq j}(k_i+k_j)} \leq p^{-1-\max(k_i)}$$

whenever at least two k_i are nonzero. For any $k \geq 1$, there are at most $s(k+1)^{s-1}$ s-tuples that satisfy $\max k_i = k$. Thus

$$\sum_{\substack{k_1,\dots,k_s \\ \text{at least two } k_i>0}} \alpha_Z(p^{k_1},\dots,p^{k_s}) = O(\sum_{k\geq 1} sk^{s-1}p^{-k-1}) = O_s(p^{-2}).$$

Putting these three terms together and letting m tend to infinity, we get

$$\begin{aligned}\beta_p = \prod_{j=t+1}^{t+s} \left(1-\chi_{D_j}(p)p^{-1}\right) \Bigg(&\mathbb{E}_{a\in[p]^d} \prod_{i=1}^{t} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(a)) \\ &+ \sum_{j=t+1}^{t+s} \mathbb{E}_{a\in[p]^d} p 1_{\psi_j(a)=0} \prod_{i=1}^{t} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(a)) \sum_{k=1}^{+\infty} \chi_{D_j}(p^k)p^{-k}\Bigg) + O_{s,t}(p^{-2}). \quad \text{(B.9)}\end{aligned}$$

Lemma 2.3 proved that $\mathbb{E}_{a\in[p]^d} \prod_{i=1}^{t} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(a)) = 1 + O_{d,t}(p^{-2})$. Similarly, for any $j \in [\![t+1\,;\, t+s]\!]$, we have

$$\begin{aligned}\mathbb{E}_{a\in[p]^d} p 1_{p|\psi_j(a)} \prod_{i=1}^{t} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(a)) &= p\left(\frac{p}{p-1}\right)^t \mathbb{P}((\prod_{i=1}^{t} \psi_i(a), p) = 1 \text{ and } p \mid \psi_j(a)) \\ &= 1 + O(p^{-2})\end{aligned}$$

because the probability is $p^{-1}(1 - t/p + O(p^{-2}))$ by linear independence. Moreover,

$$\prod_{j=t+1}^{t+s} \left(1-\chi_{D_j}(p)p^{-1}\right) \left(1 + \sum_{j=t+1}^{t+s} \sum_{k_j>0} \chi_{D_j}(p^{k_j})p^{-k_j}\right) = 1 + O_s(p^{-2})$$

so that finally, plugging these estimates in (B.9), we obtain

$$\beta_p = \prod_{j=t+1}^{t+s} \left(1 - \chi_{D_j}(p)p^{-1}\right) \left(1 + \sum_{j=t+1}^{t+s} \sum_{k_j>0} \chi_{D_j}(p^{k_j})p^{-k_j} + O_{s,t}(p^{-2})\right) = 1 + O(p^{-2}).$$

Here the implied constant depends on t, d, s, L and the discriminants only. This last equation is exactly the claimed result. $\square$

Appendix C

Verification of the linear forms condition

This appendix is dedicated to the lengthy and technical proof of Proposition 4.12 and Proposition 2.13; observe that the former does not exactly imply the latter, because the latter allows unbounded coefficients, which the former does not. We actually prove the following proposition, which proves both aforementioned propositions.

Proposition C.1. *Let $\Psi = (\psi_1, \dots, \psi_{t+s})$ be a system of affine-linear forms on $\mathbb{Z}^d$. Let $D_1, \dots, D_s$ be some discriminants of PDBQFs. Let $c_{\mathrm{GT}}(\chi)$ be the constant appearing in Proposition 2.9. Let $\widehat{W}$ be divisible by W and every exceptional prime for the system Ψ. For $i \in [t]$, let $\nu_i = \nu_{\mathrm{GT},\widehat{W},b_i}$ and for $i \in [\![t+1 \,;\, t+s]\!]$, let $\nu_i = \nu_{\mathrm{Matt},\widehat{W},b_i,D_i}$. Suppose that $K \subset [0,N]^d$ satisfies $\Psi(K) \subset \mathbb{R}_+^t$. Let $(b_1, \dots, b_{t+s}) \in [\widehat{W}]^{t+s}$ be such that for any prime p, we have $(b_i, W) = 1$ for any $i \in [t]$ and $b_j \not\equiv 0 \bmod p^{v_p(\widehat{W})}$ for any $j \in [t+1, \dots, t+s]$. Suppose that no exceptional prime larger than w for Ψ divides any of the integers b_i. Further, suppose that*

$$\mathbb{E}_{n \in [N]} \nu_i(n) = 1 + o(1) \tag{C.1}$$

for any $i \in [t+s]$. Then

$$\mathbb{E}_{n \in K \cap \mathbb{Z}^d} \prod_{i \in [t+s]} \nu_i(\psi_i(n)) = 1 + O\left(\frac{N^{d-1+O(\gamma)}}{\mathrm{Vol}(K)}\right) + o(1). \tag{C.2}$$

Observe that if $\mathrm{rad}(\widehat{W}) = O(\log^{O(1)} N)$, hypothesis (C.1) is satisfied for any $i \in [t]$ by Proposition 2.11, so Proposition C.1 does imply Proposition 2.13. Further, if $\widehat{W} = \overline{W}$, hypothesis (C.1) is satisfied for any $i \in [t+s]$, so Proposition C.1 does imply Proposition

4.12 as well.

We loosely follow Matthiesen's proof in [68], taking inspiration from the more recent paper [17]. However, there is some flaw there, as the author overlooked the possibility that u and $dm^2\epsilon$ may not be coprime; we provide, based on the earlier paper [66], a corrected version of these computations.

Compared to Matthiesens's articles, the presence of the majorant for the von Mangoldt function introduces factors of $\log R$ which will be cancelled out during the Fourier transformation step. It also introduces factors of $\frac{\varphi(W)}{W}$ which remain untouched throughout the proof. And in the core of the calculation, it adds to the variables d, m, e, u another variable ℓ also ranging among the integers whose prime factors are all greater than $w(N)$, which shall interact nicely with the other ones. The aim of the game is to dissociate the factors, that is, to transform the average of the product into the product of averages. This way, we will reduce the problem to the case where $t = 1, s = 0$, that is, Proposition 2.11, and the case where $t = 0, s = 1$, which corresponds to Lemma 4.10.

Notational conventions for the proof. In order to somewhat lighten the formidable notation, we will not always specify the range on sums, products or integrals. In principle, the name of the variable alone should tell the reader what its range is. We list a few important conventions.

- The integer vector n will always range in $\mathbb{Z}^d \cap K$.
- We put $\phi_j(n) = \widehat{W}\psi_j(n) + b_j$, for $j \in [t+s]$. Let $\Phi = (\phi_1, \ldots, \phi_{t+s})$.
- For $i = 1, \ldots, t$ and $k = 1, 2$, the variable $\ell_{i,k}$ is a positive integer. Because it will always be a divisor of $\phi_i(n)$ which satisfies $\phi_i(n) \equiv b_i \bmod W$ and $(b_i, W) = 1$ by definition of B, the prime factors of $\ell_{i,k}$ are all greater than $w(N)$.
- For $j = t+1, \ldots, t+s$ and $k = 1, 2$, the variable $e_{j,k}$ is a positive integer in $\langle \mathcal{Q}_j \rangle$, where $\mathcal{Q}_j = \mathcal{Q}_{D_j}$. All its prime factors are greater than $w(N)$.
- For $j = t+1, \ldots, t+s$, the variable s_j will range from $2/\gamma$ to $(\log\log N)^3$ and i_j from $\log_2 s - 2$ to $6 \log\log\log N$, while u_j ranges in $U(s_j, i_j)$. The s_j should not be confused with s, the number of factors of the form $\nu_{\mathrm{Matt},b}$. Notice that i is also the standard name of the index ranging in $[t]$ but this should not cause any ambiguity.
- Occasionally we may want to write e_j for $e_{j,1}$ and $e'_j = e_{j,2}$; similarly $\ell_i = \ell_{i,1}$ and $\ell'_i = \ell_{i,2}$. Moreover ϵ_j will be the least common multiple (lcm) of e_j and e'_j, while λ_i will be the lcm of ℓ_i and ℓ'_i.

- For $j = t+1, \dots, t+s$, the integer d_j only has prime factors greater than $w(N)$ and lying in $\mathcal{P}_j$ where $\mathcal{P}_j = \mathcal{P}_{D_j}$.
- For $j = t+1, \dots, t+s$, the integer m_j only has prime factors greater than $w(N)$ and lying in $\mathcal{Q}_j$.
- A bold character denotes a vector; thus $\mathbf{e} = (e_{j,k})_{\substack{j\in[\![t+1\,;\,t+s]\!]\\ k=1,2}}$ and again the range of such indices i, k will frequently be omitted.
- For $i \in [t]$, let $c_i = c(\chi)$ be the constant appearing in Proposition 2.11, and for $j \in [t+1, t+s]$, let $c_j = c(D_{t+j}, \chi)$ be the constant appearing in Lemma 4.10.

With these conventions, recalling the definitions (2.10) of $\Lambda_{\chi,R}$ and (4.16) of $r_{D,\gamma}$, we expand the left-hand side of (4.19) as

$$\Omega = H\Omega'$$

where

$$\Omega' = \mathbb{E}_{n\in\mathbb{Z}^d\cap K} \prod_{i\in[t]} \sum_{\ell_i,\ell_i'} \mu(\ell_i)\mu(\ell_i')\chi\left(\frac{\log \ell_i}{\log R}\right)\chi\left(\frac{\log \ell_i'}{\log R}\right) 1_{\lambda_i|\phi_i(n)}$$
$$\prod_{j=t+1}^{t+s} \sum_{s_j,i_j,u_j} 2^{s_j} 1_{u_j|\phi_j(n)} \sum_{d_j,m_j,e_j,e_j'} 1_{d_j m_j^2 \epsilon_j|\phi_j(n)} \mu(e_j)\mu(e_j')\chi\left(\frac{\log e_j}{\log R}\right)\chi\left(\frac{\log e_j'}{\log R}\right)\chi\left(\frac{\log d_j}{\log R}\right)\chi\left(\frac{\log m_j}{\log R}\right). \tag{C.3}$$

and H is defined by

$$H = \left(\log R \frac{\varphi(\widehat{W})}{\widehat{W}}\right)^t \prod_{j=1}^{t+s} c_j^{-1}.$$

To prove Proposition 4.12, we have to prove that

$$\Omega = 1 + O\left(\frac{N^{d-1+O(\gamma)}}{\mathrm{Vol}(K)}\right) + o(1).$$

Notice that $H = O((\log R)^t) = O((\log N)^t)$. We now work on Ω'. It is an average over n of $t+s$ products, and we aim at transforming it into a product of $t+s$ averages. We will remember to multiply the error terms obtained for Ω' during the transformation of this average by $(\log N)^t$ to obtain error terms for Ω.

We observe that when u_j, d_j, m_j, e_j, e_j' divide $\phi_j(n)$ and u_j satisfies $\gcd(u_j, \phi_j(n)/u_j) = 1$,

there exists, for x equal to any of the symbols e, e', d, m, a unique decomposition

$$x_j = x_j^{(1)} v_{j,x} \quad \text{with} \quad \gcd(x_j^{(1)}, u_j) = 1 \quad \text{and} \quad v_{j,x} \mid u_j. \tag{C.4}$$

We would very much like to perform this decomposition, but not every term satisfies the required coprimality condition. However, the following claim shows that we can pretend it does at a small cost. In fact it shows more.

Claim 1. The summands in (C.3) satisfying $\gcd(u_j, \phi_j(n)/u_j) > 1$ for some j or $\gcd(u_j, \phi_i(n)) > 1$ for some $i \neq j$ contribute only $O(N^{-(\log\log N)^{-3}/8})$ to Ω.

Proof. Bounding μ and χ by 1, we find that the contribution S of these summands to Ω' satisfies

$$|S| \leq \sum_{\mathbf{i},\mathbf{s}} \left(\prod_{j=t+1}^{t+s} 2^{s_j} \right) \mathbb{E}_n a_n,$$

where

$$a_n = a_{n,\mathbf{i},\mathbf{s}} = \sum_{\mathbf{u}} 1_{\substack{\exists j \mid \gcd(u_j, \phi_j(n)/u_j) > 1 \\ \text{or } \exists i \neq j \mid \gcd(u_j, \phi_i(n)) > 1}} \quad \textstyle\sum_{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}} \prod_{i=1}^{t} 1_{\lambda_i \mid \phi_i(n)} \prod_{j=t+1}^{t+s} 1_{\Delta_j \mid \phi_j(n)}$$

with the notation $\Delta_j = \operatorname{lcm}(u_j, d_j m_j^2 \epsilon_j)$. To bound $\mathbb{E}_n a_n$, we apply the simple rule, based on Cauchy-Schwarz, that

$$(\mathbb{E}_{n \in \mathbb{Z}^d \cap K} a_n)^2 \leq \mathbb{P}_n(a_n \neq 0) \mathbb{E}_n a_n^2.$$

Now if $a_n \neq 0$ then either the value of one of the last s linear forms $\phi_i(n)$ has a repeated prime factor, or the values of two of the $t+s$ linear forms have a common prime factor. Such a prime p is a factor of some u_i, which, by Definition 4.3, only has prime factors larger than $N^{1/(\log\log N)^3}$ and satisfies $u_i \leq N^\gamma$ (see [66, Proposition 4.2]). Thus p certainly lies between $N^{1/(\log\log N)^3}$ and N^γ. Using the triangle inequality, we get

$$\mathbb{P}_n(a_n \neq 0) \leq \sum_{N^{1/(\log\log N)^3} \leq p \leq N^\gamma} \mathbb{P}_n(p^2 \mid \prod_{i=1}^{t+s} \phi_i(n)).$$

Let $N^{1/(\log\log N)^3} \leq p \leq N^\gamma$ be a prime. In particular, we have $p > w(N)$. We use the hypothesis of Proposition C.1 regarding exceptional primes. If $p \mid \widehat{W}$, then for any $i \in [t+s]$, the form ϕ_i mod p is constantly equal to the nonzero residue b_i. So $\mathbb{P}_n(p^2 \mid \prod_{i=1}^{t+s} \phi_i(n)) = 0$.

Otherwise, p is not exceptional for Ψ nor Φ, whence

$$\mathbb{P}_n(p^2 \mid \prod_i \phi_i(n)) \ll p^{-2} + O\left(p^2 \frac{N'^{d-1}}{\mathrm{Vol}(K)}\right) = p^{-2} + O\left(p^2 \frac{N^{d-1}}{\mathrm{Vol}(K)}\right),$$

according to Proposition A.6 and Lemma A.1, i.e. the fact[1] that $|K \cap \mathbb{Z}^d| \sim \mathrm{Vol}(K)$. Hence

$$\begin{aligned}
\mathbb{P}(a_n \neq 0) &\leq \sum_{N^{1/(\log\log N)^3} \leq p \leq N^\gamma} \mathbb{P}(p^2 \mid \prod_i \phi_i(n)) \\
&\ll \sum_{p \geq N^{1/(\log\log N)^3}} p^{-2} + \frac{N^{d-1}}{\mathrm{Vol}(K)} \sum_{p \leq N^\gamma} p^2 \\
&\ll N^{-1/(\log\log N)^3} + \frac{N^{3\gamma+d-1}}{\mathrm{Vol}(K)}.
\end{aligned}$$

Assuming that γ is small enough (less than $1/3$), the second term is $O(N^{-c})$ with $c > 0$ so it is negligible with respect to the first one.

We then bound $\mathbb{E}_n a_n^2$ quite crudely as follows

$$\begin{aligned}
\mathbb{E}_n a_n^2 &\leq \mathbb{E}_n \left(\sum_{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell},\mathbf{u}} \prod_{i=1}^{t} 1_{\lambda_i | \phi_i(n)} \prod_{j=t+1}^{t+s} 1_{\Delta_j | \phi_j(n)} \right)^2 \\
&\ll \prod_{i=1}^{t} \left(\mathbb{E}_n \left(\sum_{\ell_i, \ell_i'} 1_{\lambda_i | \phi_i(n)} \right)^{2(t+s)} \right)^{1/(t+s)} \prod_{j=t+1}^{t+s} \left(\mathbb{E}_n \left(\sum_{d_j, m_j, e_j, e_j', u_j} 1_{\Delta_j | \phi_j(n)} \right)^{2(t+s)} \right)^{1/(t+s)} \\
&\ll (\log N)^{O_{t,s}(1)}.
\end{aligned}$$

The second inequality is Hölder's. The last one follows from bounds of Matthiesen [66, Lemma 3.1] on moments of the divisor function, and the observation that for instance $\sum_{\ell_i, \ell_i'} 1_{\lambda_i | \phi_i(n)} \leq \tau(\phi_i(n))^2$. Thus $|\mathbb{E}_n a_n| \ll N^{-(\log\log N)^{-3}/4}$. Summing now over $\mathbf{i}, \mathbf{s}$ and multiplying by H, we get $H|S| \leq N^{-(\log\log N)^{-3}/8}$ as desired. This concludes the proof of Claim 1. □

Thus to evaluate (C.3), we shall pretend all summands satisfy the coprimality condition, transform them under this hypothesis, and then reintegrate the formerly excluded terms, which generates an error term of size $O(N^{-(\log\log N)^{-3}/8})$. So from now on, the vectors

[1] Here, we assume that $\mathrm{Vol}(K) \gg N'^d$ or at least that $N'^{d-1} = o(\mathrm{Vol}(K))$. Indeed, in the statement of the main theorem, we could also add the assumption that $\mathrm{Vol}(K) \gg N^d$ because otherwise the error term is not smaller than the main term.

$\mathbf{d}, \mathbf{e}, \mathbf{m}$ will be assumed to be entrywise coprime to the vector $\mathbf{u}$. Under this convention, and up to an error term of size $O(N^{-(\log\log N)^{-3}/8})$, the expression Ω' of equation (C.3) is equal to

$$\sum_{\mathbf{i},\mathbf{s}} \sum_{\mathbf{u}}{}' \mathbb{E}_n \prod_{i\in[t],k=1,2} \left(\sum_{\ell_{i,k}} \mu(\ell_{i,k}) \chi\left(\frac{\log \ell_{i,k}}{\log R}\right) 1_{\lambda_i | \phi_i(n)} \right) \prod_{j=t+1}^{t+s} 2^{s_j} \sum_{d_j, e_j, e'_j, m_j \text{ coprime to } u_j}$$
$$\sum_{\substack{v_{j,d}, v_{j,m}, v_{j,e}, v_{j,e'} \\ \text{divisors of } u_j}} \prod_{x_j \in \{d_j, e_j, e'_j, m_j\}} \chi\left(\frac{\log x_j v_{j,x}}{\log R}\right) \mu(e_j v_{j,e}) \mu(e'_j v_{j,e'}) 1_{u_j d_j \epsilon_j m_j^2 | \phi_j(n)} \quad \text{(C.5)}$$

where the dashed sum indicates a sum over vectors whose entries are coprime.

By the coprimality condition, we can perform the decomposition (C.4). The vector $\mathbf{v}$ stands for $(v_{j,x})_{x\in\{d,e,e',m\}, j\in[\![t+1\,;\,t+s]\!]}$ where we impose for every j the conditions $v_{j,x} \mid u_j$ and $v_{j,d} \in \langle \mathcal{P}_j \rangle, v_{j,m} \in \langle \mathcal{Q}_j \rangle, v_{j,e} \in \langle \mathcal{Q}_j \rangle$. Furthermore, we shall use the notation

$$q_j = \begin{cases} \lambda_j & \text{if } j \in [t] \\ d_j \epsilon_j m_j^2 & \text{if } j \in [\![\, t+1\,;\, t+s \,]\!]. \end{cases}$$

Claim 2. The main term of (C.5) is equal to

$$\sum_{\mathbf{i},\mathbf{s}} \sum_{\mathbf{u}} \sum_{\mathbf{d},\mathbf{e},\mathbf{m},\boldsymbol{\ell}} \alpha(q_1, \ldots, q_{t+s})$$
$$\sum_{\mathbf{v}} \prod_{i\in[t],k=1,2} \mu(\ell_{i,k}) \chi\left(\frac{\log \ell_{i,k}}{\log R}\right) \prod_{j\in[\![t+1\,;\,t+s]\!]} \frac{2^{s_j}}{u_j} \mu(e'_j v_{j,e'}) \mu(e_j v_{j,e}) \prod_{x_j \in \{d_j, e_j, e'_j, m_j\}} \chi\left(\frac{\log x_j v_{j,x}}{\log R}\right) \quad \text{(C.6)}$$

up to an error of size $O\left(N^{d-1+O(\gamma)}/\mathrm{Vol}(K)\right)$.

We note that this error term, after multiplication by the initial factor $H = O((\log N)^t)$, is still of the same magnitude.

Proof. First, we apply Lemma A.2

$$\mathbb{E}_{n\in\mathbb{Z}^d\cap K} \prod_{i=1}^{t} 1_{\lambda_i|\phi_i(n)} \prod_{j=t+1}^{t+s} 1_{u_j d_j m_j^2 \epsilon_j | \phi_j(n)} = \alpha((q_i)_{i\in[t]}, (u_j q_j)_{j\in[\![t+1\,;\,t+s]\!]}) + O(N^{d-1+O(\gamma)}/\mathrm{Vol}(K)).$$

To explain the error term, observe that for any set of tuples bringing a nonzero contribution,

for any $j \in [t+s]$, we have $u_j \leq N^\gamma$ and $q_j = N^{O(\gamma)}$ because $d_j, m_j, e_j, e'_j, \ell_j, \ell'_j \leq N^\gamma$. To bound the contribution of this error term to the sum defining the main term of (C.5), we simply notice that the number of terms is $N^{O(\gamma)}$ anyway, that the μ and χ factors are 1-bounded, and that 2^{s_j} is always $o(N^\gamma)$ because $s_j \leq (\log\log N)^3$.

Notice that we can also exclude summands for which $\gcd(\lambda_i, u_j) > 1$ for some $i \in [t]$ and $j \in [\![t+1\,;\, t+s]\!]$ because of Claim 1. For summands satisfying on the contrary $\gcd(\lambda_i, u_j) = 1$, by multiplicativity of α and because of the other implicit coprimality conditions, we can write

$$\alpha((q_i)_{i\in[t]}, (u_j q_j)_{j\in[\![t+1\,;\,t+s]\!]}) = \frac{\alpha(q_1, \ldots, q_{t+s})}{\prod_j u_j}.$$

This concludes the proof of this claim with a dashed sum on u instead of the normal sum, and a sum on $\boldsymbol{\ell}$ restricted to tuples satisfying $\gcd(\lambda_i, u_j) = 1$ for all i and j. We can reintegrate now the formerly excluded terms because they have a negligible contribution anyway, so Claim 2 is proven. □

From now on, we fix vectors $\mathbf{i}, \mathbf{s}$ in their usual ranges, and consider the individual terms

$$\sum_{\mathbf{u}} \sum_{\mathbf{d},\mathbf{e},\mathbf{m},\boldsymbol{\ell}} \alpha(q_1, \ldots, q_{t+s}) \prod_{i\in[t],k=1,2} \mu(\ell_{i,k})\chi\left(\frac{\log \ell_{i,k}}{\log R}\right)$$
$$\sum_{\mathbf{v}} \prod_{j\in[\![t+1\,;\,t+s]\!]} \frac{2^{s_j}}{u_j}\mu(e_j v_{j,e})\mu(e'_j v_{j,e'}) \prod_{x_j\in\{d_j,e_j,e'_j,m_j\}} \chi\left(\frac{\log x_j v_{j,x}}{\log R}\right). \quad \text{(C.7)}$$

Recall that we introduced the Fourier transform θ of the function $x \mapsto e^x\chi(x)$ during the proof of Proposition 2.11. When plugging the Fourier transforms into our sum, we need $4s+2t$ real variables $\xi_{j,k}$ with $k = 1, \ldots, 4$ for $j = t+1, \ldots, t+s$ and $k = 1, 2$ for $j = 1, \ldots, t$. Collectively, they form the vector $\boldsymbol{\Xi}$. Furthermore, we write $z_{j,k} = (1 + i\xi_{j,k})/(\log R)$. We sometimes allow, for a function f, the slight abuse of notation

$$\prod_{j,k} f(\xi_{j,k}) = \prod_{i\in[t],k\in[2]} f(\xi_{i,k}) \prod_{j\in[t+s]\setminus[t],k\in[4]} f(\xi_{j,k}),$$

and write

$$\theta(\boldsymbol{\Xi}) = \prod_{j,k} \theta(\xi_{j,k}).$$

We introduce the notation $\tilde{x}_j = x_j v_{j,x}$ for x equal to any of the symbols e, e', d, m, and $\mathbf{v_i} = (v_{i,d}, v_{i,e}, v_{i,e'}, v_{i,m})$. For any fixed values of the tuples $\mathbf{s}, \mathbf{i}, \mathbf{u}, \mathbf{v}, \mathbf{d}, \mathbf{m}, \mathbf{e}, \boldsymbol{\ell}$ we write

$$M = \prod_{i\in[t],k=1,2} \mu(\ell_{i,k}) \prod_{j\in[\![t+1\,;\,t+s]\!]} \frac{2^{s_j}}{u_j} \mu(e_j v_{j,e}) \mu(e'_j v_{j,e'}).$$

Observe that $\mu(e_j v_{j,e}) = \mu(e_j)\mu(v_{j,e})$ by coprimality, and the same with e'. Finally, we introduce

$$F_{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}}(\boldsymbol{\Xi}) = F(\boldsymbol{\Xi}) = \theta(\boldsymbol{\Xi}) \prod_{j>t} \tilde{e}_{j,1}^{-z_{j,1}} \tilde{e}_{j,2}^{-z_{j,2}} \tilde{d}_j^{-z_{j,3}} \tilde{m}_j^{-z_{j,4}} \prod_{i\in[t]} \ell_{i,1}^{-z_{i,1}} \ell_{i,2}^{-z_{i,2}}. \tag{C.8}$$

We now insert (2.16) into the expression (C.7) to get

$$\sum_{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}} \alpha(q_1,\dots,q_{t+s}) \sum_{\mathbf{u},\mathbf{v}} M\Big(\int_{I^{4s+2t}} F(\boldsymbol{\Xi}) d\boldsymbol{\Xi} + O((\log R)^{-A} (\prod_{i,j,k} \tilde{e}_{j,k} \ell_{i,k} \tilde{d}_j \tilde{m}_j)^{-1/\log R}) \Big).$$

Above we abused notation slightly and wrote

$$\prod_{i,j,k} \tilde{e}_{j,k} \ell_{i,k} \tilde{d}_j \tilde{m}_j = \prod_{i\in[t],k=1,2} \ell_{i,k} \prod_{j>t,k'=1,2} \tilde{e}_{j,k'} \tilde{m}_j \tilde{d}_j.$$

We shall use this notation again in the sequel.

Now the term arising from the big oh will not matter too much, thanks to the following claim.

Claim 3. For $A > 0$ large enough,

$$H \sum_{\mathbf{s},\mathbf{i}} \sum_{\mathbf{u},\mathbf{v}} \sum_{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}} \alpha(q_1,\dots,q_{t+s}) |M| \log^{-A} R \left(\prod_{i,j,k} \tilde{e}_{j,k} \ell_{i,k} \tilde{d}_j \tilde{m}_j \right)^{-1/\log R} = o(1).$$

Proof. Matthiesen [66, Proposition 4.2] showed that

$$\sum_{\mathbf{s},\mathbf{i}} \prod_{j=t+1}^{t+s} \sum_{u_j \in U(s_j,i_j)} \frac{2^{s_j}}{u_j} = O(1).$$

On the other hand, we can suppress the sum over $\mathbf{v}$ by reintegrating into the sum over $\mathbf{d}, \mathbf{m}, \mathbf{e}$ the summands not termwise coprime to $\mathbf{u}$. We can then drop the $\tilde{\cdot}$ on the variables.

We put $q'_j = \ell_j \ell'_j$ for $j \in [t]$ and $q'_j = e_j e'_j d_j m_j$ for $j \in [t+s] \setminus [t]$. By multiplicativity,

$$\begin{aligned}
\sum_{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}} \alpha(q_1,\dots,q_{t+s}) \left(\prod_{i,j,k} e_{j,k}\ell_{i,k} d_j m_j\right)^{-\frac{1}{\log R}} &= \sum_{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}} \prod_{p^{a_i}\|q_i} \alpha(p^{a_1},\dots,p^{a_{t+s}}) \prod_{\substack{j\in[t+s]\\ p^{a'_j}\|q'_j}} p^{-\frac{a'_j}{\log R}} \\
&\leq \sum_{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}} \prod_{p^{a_i}\|q_i} p^{-\max a_i (1+(2\log R)^{-1})} \\
&\leq \prod_p (1 - p^{-(1+(2\log R)^{-1})})^{-O(t+s)} \\
&\ll \log^{O(t+s)} N.
\end{aligned}$$

Here we used $a'_j \geq a_j/2$, Corollary A.4 and a crude bound $k^{O(t+s)}$ for the number of tuples a_i satisfying $\max_i a_i = k$. The last inequality follows from a well-known estimate for the Riemann zeta function near 1, namely

$$\zeta(x) = O\left(\frac{1}{x-1}\right).$$

Given that $H = O(\log^t N)$, the claim follows for A large enough depending on t and s only. □

We are left to deal with

$$\sum_{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}} \alpha(q_1,\dots,q_{t+s}) \sum_{\mathbf{u},\mathbf{v}} M \int_{I^{4s+2t}} F(\boldsymbol{\Xi}) d\boldsymbol{\Xi}. \tag{C.9}$$

We now swap the summation $\sum_{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}}$ and the integration over the compact set I^{4s+2t}, using Fubini's theorem. This causes no problem because the sum is absolutely convergent; this absolute convergence is a byproduct of the proof of Claim 3.

We also continue swapping summation and multiplication, by enforcing at little cost an extra coprimality condition: we show we can restrict to tuples where $(q_i, q_j) = 1$ for all $i \neq j$. We need another, more subtle argument to impose this coprimality compared to the coprimality condition involving the variables u_j in Claim 1, because a crucial ingredient of the proof of that claim was that the prime factors involved were all at least $N^{(\log\log N)^{-3}}$, an assumption we do not have for d, m, e.

Claim 4. Let $\mathbf{s}, \mathbf{i}, \mathbf{u}, \mathbf{v}$ be fixed vectors of integers satisfying the usual conditions. Then

we have

$$\sum_{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}} \alpha(q_1,\dots,q_{t+s})F(\boldsymbol{\Xi})\prod_{i,j}\mu(\ell_{i,1})\mu(\ell_{i,2})\mu(\tilde{e}_{j,1})\mu(\tilde{e}_{j,2})$$
$$= (1+O(w(N)^{-1/2}))\sideset{}{'}\sum_{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}} \alpha(q_1,\dots,q_{t+s})F(\boldsymbol{\Xi})\prod_{i,j}\mu(\ell_{i,1})\mu(\ell_{i,2})\mu(\tilde{e}_{j,1})\mu(\tilde{e}_{j,2}),$$

where the dashed sum is restricted to tuples satisfying $(q_i,q_j)=1$ for all $i\neq j$.

Proof. The goal is to bound the contribution of the entries failing the coprimality conditions. To achieve this, we observe that each summand is a product of $\theta(\boldsymbol{\Xi})$, a term depending only on the fixed tuple $\mathbf{v}$ and a term $T(\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell})$ of the form

$$\alpha(q_1,\dots,q_{t+s})\prod_{i=1}^{t}\ell_{i,1}^{-z_{i,1}}\ell_{i,2}^{-z_{i,2}}\mu(\ell_{i,1})\mu(\ell_{i,2})\prod_{j=t+1}^{t+s}e_{j,1}^{-z_{j,1}}e_{j,2}^{-z_{j,2}}d_j^{-z_{j,3}}m_j^{-z_{j,4}}\mu(e_{j,1})\mu(e_{j,2}), \quad \text{(C.10)}$$

whose multiplicativity we will exploit, in order to write it as a product over primes; only primes greater than $w(N)$ need be considered, as smaller ones have no chance of dividing any of the parameters. We can even partition the primes p into two classes C_1 and C_2, according to whether p divides a single q_j or at least two of them. Thus, the term $T(\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell})$ can be written as

$$\prod_{p\in C_1}\alpha((p^{v_p(q_j)})_j)A_p(\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}))\prod_{p\in C_2}\alpha((p^{v_p(q_j)})_j)A_p(\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell})$$

where A_p is a complex number of modulus at most one and v_p is the p-adic valuation. For any given tuples $\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}$ and $j\in[s+t]$, we write $\kappa_j=\prod_{p\in C_2}p^{v_p(q_j)}$. Thus

$$p\mid\kappa_i\Rightarrow p\mid\prod_{j\neq i}\kappa_j.$$

We now arrange the terms $T(\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell})$ according to their tuples $(\kappa_1,\dots,\kappa_{t+s})$. Let us fix such a tuple $(\kappa_1,\dots,\kappa_{t+s})$. Let κ be the radical of $\kappa_1\dots\kappa_{t+s}$, that is, the product of its prime factors. Thus a number n is coprime to $\prod_i\kappa_i$ if and only if it is coprime to κ. The

sum of terms T corresponding to this tuple is equal to

$$S_{\kappa_1,\ldots,\kappa_{t+s}} = E_\kappa \sum_{\substack{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell} \\ \forall j\, q_j=\kappa_j}} T(\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}), \tag{C.11}$$

where

$$E_\kappa = \sideset{}{'}\sum_{\substack{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell} \\ \forall j\, (q_j,\kappa)=1}} T(\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell})$$

is absolutely convergent, as a subsum of the unrestricted sum which was shown during the proof of Claim 3 to be convergent and less than $\log^{O(t+s)} N$. The second sum in the right-hand side of equation (C.11) is a finite sum. Note that the coprimality condition denoted by the dashed sum defining E_κ implies that $\alpha(q_1,\ldots,q_{t+s}) = (q_1\cdots q_{t+s})^{-1}$. Now we write

$$E = \sideset{}{'}\sum_{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}} T(\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}) = \sum_{\delta|\kappa} \sideset{}{'}\sum_{\substack{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell} \\ (\prod_j q_j,\kappa)=\delta}} T(\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}). \tag{C.12}$$

Fix a divisor δ of κ. By the coprimality condition, for $\delta = p_1\cdots p_r$ with $p_1,\ldots,p_r$ pairwise distinct primes, we have

$$\sideset{}{'}\sum_{\substack{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell} \\ (\kappa,\prod_j q_j)=\delta}} T(\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}) = \sum_{f:[r]\to[t+s]} \sideset{}{'}\sum_{\substack{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell} \\ \forall i\, (q_i,\kappa)=\prod_{f(j)=i} p_j}} T(\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}).$$

Fix a map $f:[r]\to[t+s]$ (there are $(t+s)^r$ choices) and write $A_i = f^{-1}(\{i\})$ and $\delta_i = \prod_{j\in A_i} p_j$, thus $\delta = \prod_i \delta_i$ and the δ_i are pairwise coprime. For $i\in[t]$, we have $(q_i,\kappa)=\delta_i$ if and only if there is a (unique) pair of sets $A_{i,1}, A_{i,2}$ such that $A_i = A_{i,1}\cup A_{i,2}$ satisfying $(\ell_{i,k},\kappa) = \prod_{j\in A_{i,k}} p_j = \delta_{i,k}$. For $i>t$ we similarly need four sets with $A_{i,1}\cup A_{i,2}\cup A_{i,3}\cup A_{i,4} = A_i$ satisfying $(e_{i,k},\kappa) = \prod_{j\in A_{i,k}} p_j$ and $(d_i,\kappa) = \prod_{j\in A_{i,3}} p_j$ as well as $(m_i,\kappa) = \prod_{j\in A_{i,4}} p_j$. We shall use the obvious notation $\delta_{i,k} = \prod_{j\in A_{i,k}} p_j$. Fix now sets $A_{i,k}$ as described and consider the sum of the terms $T(\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell})$ over tuples $\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}$ satisfying the gcd conditions corresponding to these sets $A_{i,k}$ (as well as the coprimality conditions). This sum equals

$$E_\kappa \prod_{i\in[t+s]} \delta_i^{-1} \prod_{k=1,2} \delta_{i,k}^{-z_{i,k}} (-1)^{|A_{i,k}|} \prod_{j>t} \delta_{j,3}^{-z_{j,3}} \delta_{j,4}^{-z_{j,4}}$$

where the factor after E_κ has modulus at most δ^{-1}. Now the number of choices for the collection of sets $A_{i,j}$ is $\exp(O(r))$. Thus equation (C.12) can be rewritten as

$$E = E_\kappa \prod_{p|\kappa} (1 + O(p^{-1})),$$

an equation we can invert to get

$$E_\kappa = E \prod_{p|\kappa} (1 + O(p^{-1})).$$

Plugging the last line into equation (C.11), we obtain

$$S_{\kappa_1,\ldots,\kappa_{t+s}} = E S'_{\kappa_1,\ldots,\kappa_{t+s}}$$

where

$$S'_{\kappa_1,\ldots,\kappa_{t+s}} = \sum_{\substack{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell} \\ \forall j\, q_j = \kappa_j}} T(\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}) \prod_{p|\prod_j \kappa_j} (1 + O(p^{-1})). \tag{C.13}$$

What is left to do is to bound

$$S = \sum_{\substack{\kappa_1,\ldots,\kappa_{t+s} \\ \exists j\, \kappa_j > 1}} S'_{\kappa_1,\ldots,\kappa_{t+s}}.$$

We observe that in equation (C.13), we have $|T(\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell})| \leq \alpha(\kappa_1,\ldots,\kappa_{t+s})$. Using multiplicativity, we can then crudely bound S by

$$\prod_{p>w(N)} \left(1 + \sum_{\substack{a_1,\ldots,a_{t+s} \\ \text{at least two } a_i>0}} O(a_1^2 + \cdots + a_t^2 + a_{t+1}^4 + \cdots + a_{t+s}^4)\alpha((p^{a_i}))(1 + O(p^{-1})) \right) - 1$$

where we have used the simple bound $\tau_k(p^{a_i}) \ll a_i^{k-1}$ for $k = 3$ (because of $\lambda_i = \lambda_i/\ell_i \cdot \lambda_i/\ell'_i \cdot \ell_i\ell'_i/\lambda_i$, hence the number of occurrences of λ_i is bounded by the number of decompositions of it into three factors) and for $k = 5$ (because of $d_j m_j^2 \epsilon_j = d_j \cdot m_j^2 \cdot \epsilon_j/e_j \cdot \epsilon_j/e'_j \cdot e_j e'_j/\epsilon_j$). The requirement that at least two a_i be positive comes from the very definition of κ_i. Notice that the -1 is here to remove the 1 arising from $\alpha(1,\ldots,1)$. To further bound this expression, we first bound a_i^2 by a_i^4 and recall that the number of tuples $(a_1,\ldots,a_{t+s})$ satisfying $\max a_i = k$ is at most $t'(k+1)^{t'-1}$ (with $t' = t+s$). For such tuples, we have

$\sum_i a_i^4 \leq t'k^4$ and because of the hypothesis on primes larger than w and the fact that at least two a_i are nonzero, $\alpha((p^{a_i})_{i\in[t+s]}) \leq p^{-k-1}$ according to Proposition A.5. Thus

$$\sum_{\substack{a_1,\ldots,a_{t+s} \\ \text{at least two } a_i>0}} O(a_1^2+\cdots+a_t^2+a_{t+1}^4+\cdots+a_{t+s}^4)\alpha((p^{a_i}))(1+O(p^{-1}))$$

is bounded by

$$\sum_{k\geq 1} p^{-k-1}t'^2k^{t'+3} \ll \sum_{k\geq 1} p^{-3k/4-1} \ll p^{-3/2}$$

the first inequality being provided by obvious growth comparisons valid for large p (we may assume N to be large enough for $p > w(N)$ to satisfy automatically this condition). Since

$$\prod_{p>w(N)} (1+p^{-3/2}) - 1 \leq \sum_{n>w(N)} n^{-3/2} \ll w(N)^{-1/2},$$

Claim 4 follows. □

The extra coprimality condition that Claim 4 allows us to assume enables us to write α as the product of the reciprocals of its arguments, resulting in

$$\begin{aligned}\sum_{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}}' \alpha(q_1,\ldots,q_{t+s}) \prod_{j=t+1}^{t+s} \mu(e_j)\mu(e_j')e_{j,1}^{-z_{j,1}}e_{j,2}^{-z_{j,2}}d_j^{-z_{j,3}}m_j^{-z_{j,4}} \prod_{i=1}^{t} \mu(\ell_i)\mu(\ell_i')\ell_{i,1}^{-z_{i,1}}\ell_{i,2}^{-z_{i,2}} \\ = \sum_{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}}' \prod_{j=t+1}^{t+s} \frac{\mu(e_j)\mu(e_j')}{\epsilon_j} e_{j,1}^{-z_{j,1}}e_{j,2}^{-z_{j,2}}d_j^{-1-z_{j,3}}m_j^{-2-z_{j,4}} \prod_{i=1}^{t} \frac{\mu(\ell_i)\mu(\ell_i')}{\lambda_i}\ell_{i,1}^{-z_{i,1}}\ell_{i,2}^{-z_{i,2}}.\end{aligned}$$

Notice that the above is an equation without tildes. We will in the sequel avoid them, observing that for any fixed $\mathbf{u}$, we have

$$\begin{aligned}\sum_{\mathbf{d},\mathbf{m},\mathbf{e},\mathbf{v},\boldsymbol{\ell}}' \prod_{j=t+1}^{t+s} \frac{\mu(\tilde{e_j})\mu(\tilde{e_j}')}{\epsilon_j} \tilde{e}_{j,1}^{-z_{j,1}}\tilde{e}_{j,2}^{-z_{j,2}}\tilde{d}_j^{-1-z_{j,3}}\tilde{m}_j^{-2-z_{j,4}} \prod_{i=1}^{t} \frac{\mu(\ell_i)\mu(\ell_i')}{\lambda_i}\ell_{i,1}^{-z_{i,1}}\ell_{i,2}^{-z_{i,2}} \\ = \sum_{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}}' \prod_{j=t+1}^{t+s} \frac{\mu(e_j)\mu(e_j')}{\epsilon_j} e_{j,1}^{-z_{j,1}}e_{j,2}^{-z_{j,2}}d_j^{-1-z_{j,3}}m_j^{-2-z_{j,4}} \prod_{i=1}^{t} \frac{\mu(\ell_i)\mu(\ell_i')}{\lambda_i}\ell_{i,1}^{-z_{i,1}}\ell_{i,2}^{-z_{i,2}} \\ \times \sum_{\mathbf{v}} \mu(v_{j,e})\mu(v_{j,e'})v_{j,e}^{-z_{j,1}}v_{j,e'}^{-z_{j,2}}v_{j,d}^{-z_{j,3}}v_{j,m}^{-z_{j,4}},\end{aligned}$$

where the sum over $\mathbf{v}$ is as usual over vectors $(v_{j,x})$ where $v_{j,x} \mid u_j$ and $v_{j,x}$ satisfies the same condition on its prime factors as x (all in $\mathcal{P}_j$ for d and e, all in $\mathcal{Q}_j$ for m).

Next we claim that we can remove the dash on the sum.

Claim 5. The following equality holds, for any choice of the family $\xi_{j,k}$ in $I = [-\sqrt{\log R}, \sqrt{\log R}]$.

$$\sum_{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}}{}' \prod_{j=t+1}^{t+s} \frac{\mu(e_j)\mu(e'_j)}{\epsilon_j} e_{j,1}^{-z_{j,1}} e_{j,2}^{-z_{j,2}} d_j^{-1-z_{j,3}} m_j^{-2-z_{j,4}} \prod_{i=1}^{t} \frac{\mu(\ell_i)\mu(\ell'_i)}{\lambda_i} \ell_i^{-z_{i,1}} \ell_i'^{-z_{i,2}}$$

$$(1+O(w(N))^{-1/2}) \sum_{\mathbf{d},\mathbf{m},\mathbf{e},\boldsymbol{\ell}} \prod_{j=t+1}^{t+s} \frac{\mu(e_j)\mu(e'_j)}{\epsilon_j} e_{j,1}^{-z_{j,1}} e_{j,2}^{-z_{j,2}} d_j^{-1-z_{j,3}} m_j^{-2-z_{j,4}} \times \prod_{i\in[t]} \frac{\mu(\ell_i)\mu(\ell'_i)}{\lambda_i} \ell_{i,1}^{-z_{i,1}} \ell_{i,2}^{-z_{i,2}}.$$

Proof. The justification is basically the same as for Claim 4, because the claim simply consists in replacing the dashed sum by a complete sum, at the same small cost. □

Let us introduce for any $i \in [t]$ and $\boldsymbol{\ell}, \boldsymbol{\Xi}$ the notation

$$V_i = V_i(\boldsymbol{\ell}, \boldsymbol{\Xi}) = \frac{\mu(\ell_i)\mu(\ell'_i)}{\lambda_i} \ell_i^{-z_{i,1}} \ell_i'^{-z_{i,2}}$$

and

$$V(\boldsymbol{\ell}, \boldsymbol{\Xi}) = \prod_{i\in[t]} V_i(\boldsymbol{\ell}, \boldsymbol{\Xi}).$$

Similarly, for any $j \in [\![t+s\,;\, t+s]\!]$ and tuples $\mathbf{u}, \mathbf{v}, \mathbf{d}, \mathbf{m}, \mathbf{e}$ we define

$$\begin{aligned} S_j(\mathbf{u}, \mathbf{v}, \boldsymbol{\Xi}) &= \frac{2^{s_j}}{u_j} \mu(v_{j,e})\mu(v_{j,e'}) v_{j,e}^{-z_{j,1}} v_{j,e'}^{-z_{j,2}} v_{j,d}^{-z_{j,3}} v_{j,m}^{-z_{j,4}} \\ T_j(\mathbf{d}, \mathbf{m}, \mathbf{e}, \boldsymbol{\Xi}) &= \frac{\mu(e_j)\mu(e'_j)}{\epsilon_j} e_j^{-z_{j,1}} e_j'^{-z_{j,2}} d_j^{-1-z_{j,3}} m_j^{-2-z_{j,4}}. \end{aligned}$$

Finally we put

$$S(\mathbf{u}, \mathbf{v}, \boldsymbol{\Xi}) = \prod_{j=t+1}^{t+s} S_j \qquad \text{and} \qquad T(\mathbf{d}, \mathbf{m}, \mathbf{e}, \boldsymbol{\Xi}) = \prod_{j=t+1}^{t+s} T_j.$$

With this notation, one can rewrite (C.9) as

$$(1 + O(w^{-1/2})) \int_{I^{4s+2t}} \theta(\boldsymbol{\Xi}) \sum_{\mathbf{u},\mathbf{v}} S(\mathbf{u}, \mathbf{v}, \boldsymbol{\Xi}) \sum_{\mathbf{d},\mathbf{m},\mathbf{e}} T(\mathbf{d}, \mathbf{m}, \mathbf{e}, \boldsymbol{\Xi}) \sum_{\boldsymbol{\ell}} V(\boldsymbol{\ell}, \boldsymbol{\Xi}) d\boldsymbol{\Xi}. \tag{C.14}$$

Now we show that the error arising from the $O(w^{-1/2})$ term in (C.14) is indeed negligible:

we must ensure that

$$w^{-1/2}H\sum_{\mathbf{s},\mathbf{i}}\int_{I^{4s+2t}}\theta(\Xi)\sum_{\mathbf{u},\mathbf{v}}S(\mathbf{u},\mathbf{v},\Xi)\sum_{\mathbf{d},\mathbf{m},\mathbf{e}}T(\mathbf{d},\mathbf{m},\mathbf{e},\Xi)\sum_{\boldsymbol{\ell}}V(\boldsymbol{\ell},\Xi)d\Xi=o(1). \quad \text{(C.15)}$$

This follows because on the one hand

$$\left|\sum_{\mathbf{v_j}}\mu(v_{j,e})\mu(v_{j,e'})v_{j,e}^{-z_{j,1}}v_{j,e'}^{-z_{j,2}}v_{j,d}^{-z_{j,3}}v_{j,m}^{-z_{j,4}}\right|\leq\tau(u_j)^4$$

and

$$\sum_{\mathbf{s},\mathbf{i},\mathbf{u}}\prod_{j=t+1}^{t+s}\frac{2^{s_j}\tau(u_j)^4}{u_j}=O(1)$$

by similar calculations[2] to the ones of Matthiesen [66, Proof of Proposition 4.2]. And on the other hand, the next claim provides a fitting bound.

Claim 6. We have

$$\int\left|\theta(\Xi)\sum_{\mathbf{d},\mathbf{m},\mathbf{e}}T(\mathbf{d},\mathbf{m},\mathbf{e},\Xi)\sum_{\boldsymbol{\ell}}V(\boldsymbol{\ell},\Xi)\right|d\Xi=O(1/(\log R)^t), \quad \text{(C.16)}$$

where the integral is over I^{4s+2t}.

Given that $H=O(\log R)^t$, the bound (C.15) follows from this claim.

Proof. We first replace the sum over ℓ_i,ℓ_i', for any $i\in[t]$, by a product over primes, using multiplicativity, to get

$$\sum_{\ell_i,\ell_i'}V_i=\sum_{\ell_i,\ell_i'}\frac{\mu(\ell_i)\mu(\ell_i')}{\lambda_j}\ell_i^{-z_{i,1}}\ell_i'^{-z_{i,2}}=\prod_{s\in\mathbb{P}}(1-s^{-1-z_{i,1}}-s^{-1-z_{i,2}}+s^{-1-z_{i,1}-z_{i,2}}).$$

Then we notice that for large primes s and complex numbers z,z' of positive real part

$$1-s^{-1-z}-s^{-1-z'}+s^{-1-z-z'}=\frac{(1-s^{-1-z})(1-s^{-1-z'})}{1-s^{-1-z-z'}}+O(s^{-2}),$$

[2]The main ingredients are the easy observation that any $u\in U(i,s)$ has $2^{m_0(i,s)}$ divisors and the bound $\sum_{u\in U(i,s)}u^{-1}\leq(\sum_{p\in I_i}p^{-1})^{m_0}\ll(\log 2)^{m_0}$, where $I_i=[N^{2^{-i-1}},N^{2^{-i}}]$.

so that

$$\prod_{s\in\mathcal{P}}(1-s^{-1-z_{j,1}}-s^{-1-z_{j,2}}+s^{-1-z_{j,1}-z_{j,2}}) \ll \prod_{s\in\mathcal{P}}\frac{(1-s^{-1-z})(1-s^{-1-z'})}{1-s^{-1-z-z'}}.$$

Finally we recall that the Riemann zeta function is defined for $\Re z > 1$ by

$$\zeta(z) = \sum_{n\geq 1} n^{-z} = \prod_p (1-p^{-z})^{-1}$$

and satisfies

$$\zeta(z) = \frac{1}{z-1} + O(1)$$

for values of z near 1. From this fact, a quick computation yields

$$\prod_{s\in\mathcal{P}}\frac{(1-s^{-1-z})(1-s^{-1-z'})}{1-s^{-1-z-z'}} \ll \frac{zz'}{z+z'},$$

whence the bound

$$\prod_{s\in\mathcal{P}}(1-s^{-1-z_{i,1}}-s^{-1-z_{i,2}}+s^{-1-z_{i,1}-z_{i,2}}) \ll \frac{z_{i,1}z_{i,2}}{z_{i,1}+z_{i,2}}. \tag{C.17}$$

for any $i \in [t]$ and $\xi_{i,k} \in I$ (for $k = 1, 2$) and the corresponding $z_{i,k}$. Similarly, for any $j \in \{t+1, \ldots, t+s\}$

$$\sum_{d_j,m_j,e_j,e'_j}\frac{\mu(e_j)\mu(e'_j)}{\epsilon_j}e_j^{-z_{j,1}}e_j'^{-z_{j,2}}d_j^{-1-z_{j,3}}m_j^{-2-z_{j,4}} = \prod_{q\in\mathcal{Q}_j}(1-q^{-1-z_{j,1}}-q^{-1-z_{j,2}}+q^{-1-z_{j,1}-z_{j,2}})$$
$$\prod_{r\in\mathcal{Q}_j}(1-r^{-2-z_{j,4}})^{-1}\prod_{p\in\mathcal{P}_j}(1-p^{-1-z_{j,3}})^{-1}. \tag{C.18}$$

Notice that the product in r is a convergent product, bounded by a constant when $z_{j,4}$ varies in the permitted range.

Given that $\mathcal{P}_j$ and $\mathcal{Q}_j$ each have density 1/2 among the primes, we can write[3]

$$\sum_{q\in\mathcal{P}_j} q^{-1-z} = \frac{1}{2}\log\frac{1}{z} + O(1)$$

[3]This amounts to saying that if a set of primes has a natural density, it has a Dirichlet density which is equal to its natural density.

for $\Re z > 0$. This provides a bound for the product (C.18), similar to the one in (C.17), namely

$$\begin{aligned}\prod_{q\in\mathcal{Q}_j} & (1-q^{-1-z_{j,1}}-q^{-1-z_{j,2}}+q^{-1-z_{j,1}-z_{j,2}})\prod_{r\in\mathcal{Q}_j}(1-r^{-2-z_{j,4}})^{-1}\prod_{p\in\mathcal{P}_j}(1-p^{-1-z_{j,3}})^{-1}\\ \ll & \quad |z_{j,1}|^{1/2}|z_{j,2}|^{1/2}|z_{j,1}+z_{j,2}|^{-1/2}|z_{j,3}|^{-1/2}.\end{aligned}$$

Recall that $z_{j,k}=(1+\xi_{j,k})(\log R)^{-1}$, thus $|z_{j,k}|\le(1+|\xi_{j,k}|)(\log R)^{-1}$ by triangle inequality, and $|z_{j,1}+z_{j,2}|^{-1}\le\log R$ for any $j\in[t+s]$. Moreover, (2.15) yields

$$\theta(\Xi)=O_A\left(\prod_{j,k}(1+|\xi_{j,k}|)^{-A}\right).$$

Multiplying all these bounds, we find that the integrand in (C.16) is bounded by

$$\begin{aligned}\prod_{i=1}^{t} & |z_{i,2}||z_{i,1}||z_{i,1}+z_{i,2}|^{-1}\prod_{j=t+1}^{t+s}|z_{j,1}|^{1/2}|z_{j,2}|^{1/2}|z_{j,1}+z_{j,2}|^{-1/2}|z_{j,3}|^{-1/2}\prod_{j,k}(1+|\xi_{j,k}|)^{-A}\\ \ll & \quad (\log R)^{-t}\left(\prod_{i=1}^{t}(1+|\xi_{i,1}|)(1+|\xi_{i,2}|)\right)^{1-A}\left(\prod_{j=t+1}^{t+s}(1+|\xi_{j,1}|)(1+|\xi_{j,2}|)\right)^{1/2-A}\\ \ll & \quad (\log R)^{-t}\prod_{j,k}(1+|\xi_{j,k}|)^{-A/2}\end{aligned}$$

when A is large enough (for the last step). This last product is certainly integrable as soon as $A>2$, so the final expression is $O((\log R)^{-t})$ as claimed. □

We now study the main term of (C.14). We can again swap summation and integration using Fubini's theorem. Using separation of variables, we transform the main term of (C.14) into

$$\sum_{\mathbf{u},\mathbf{v},\mathbf{d},\mathbf{e},\mathbf{m},\boldsymbol{\ell}}\prod_{i=1}^{t}\int_{I^2}V_i\theta(\xi_{i,1})\theta(\xi_{i,2})d\xi_{i,1}d\xi_{i,2}\prod_{j=t+1}^{t+s}\int_{I^4}S_jT_j\prod_{k\in[4]}\theta(\xi_{j,k})d\xi_{j,k}. \quad \text{(C.19)}$$

It is now time to undo the truncation to I in these integrals, in order to be able to collapse them into factors of χ. The error term arising from the removal of this truncation is the same as the one introduced by the truncation, so it can be subsumed into the $o(1)$ of (4.19). Thus, up to an error term $E_{\mathbf{i},\mathbf{s}}$ satisfying $(\log R)^t\sum_{\mathbf{i},\mathbf{s}}E_{\mathbf{i},\mathbf{s}}=o(1)$, the expression

(C.19) is equal to

$$\sum_{\mathbf{u},\mathbf{v},\mathbf{d},\mathbf{e},\mathbf{m},\boldsymbol{\ell}} \prod_{i=1}^{t} \frac{\mu(\ell_{i,1})\mu(\ell_{i,2})}{\lambda_i} \prod_{k=1,2} \chi\left(\frac{\log \ell_{i,k}}{\log R}\right)$$
$$\prod_{j=t+1}^{t+s} \frac{2^{s_j}\tau(u_j)}{u_j} \frac{\mu(e_j v_{j,e})\mu(e'_j v_{j,e'})}{d_j m_j^2 \epsilon_j} \chi\left(\frac{\log d_j v_{j,d}}{\log R}\right) \chi\left(\frac{\log m_j v_{j,m}}{\log R}\right) \prod_{k=1,2} \chi\left(\frac{\log e_{j,k} v_{j,e_k}}{\log R}\right). \tag{C.20}$$

Interchanging summation and multiplication, we find that, up to error terms of the desired magnitude ($O_D\left(\frac{N^{d-1+O_D(\gamma)}}{\mathrm{Vol}(K)}\right)$ in Claims 1 and 2, various $o(1)$ throughout the proof), Ω equals

$$\prod_{j=t+1}^{t+s} C_{D_j,\gamma}^{-1} \sum_{s_j,i_j,u_j,v_j} \sum_{d_j,m_j,e_j,e'_j} \frac{2^{s_j}}{u_j} \frac{\mu(e_j v_{j,e})\mu(e'_j v_{j,e'})}{d_j m_j^2 \epsilon_j} \prod_{x\in\{d,m,e,e'\}} \chi\left(\frac{\log x_j v_{j,x}}{\log R}\right)$$
$$\times \prod_{i\in[t]} \left(\log R \frac{\varphi(\widehat{W})}{\widehat{W}} \sum_{\ell_i,\ell'_i} \frac{\mu(\ell_i)\mu(\ell'_i)}{\lambda} \prod_{x\in\{\ell_i,\ell'_i\}} \chi\left(\frac{\log x}{\log R}\right)\right),$$

which is a product of $t+s$ factors, independent of the system of linear forms. It follows that the jth factor, for $j \in [t+s]$, is also the main term of the average of the jth pseudorandom majorant for the one-variable system $\Psi : \mathbb{Z} \to \mathbb{Z}, n \mapsto n$. Now because of the hypothesis (C.1) on $\widehat{W}$, each of these averages is $1 + o(1)$, whence the result.

Appendix D

Digression on the Type I sum and sums of spaces of multiples

In this appendix, which pertains to Chapter 7, we give a treatment of the Type I sum which is valid for k as large as $n/2$. As a result, in order to prove Theorem 7.3, it would be enough to understand the Type II sum for k in a shortened range such as $[n/4 - o(n), 3n/4 + o(n)]$. We thought constraining k to this range might enable us to find an alternative, unconditional argument for the Type II sum, but we were not successful so far. Nevertheless, we present this method, because of its independent interest and of its potential use. The argument is much deeper than the one we used for $k \leq n/9$ in Section 7.6.

We present the statement we will prove in this section. The hypothesis is essentially the result we get when the Type I sum is large in Proposition 7.11.

Theorem D.1. *Let $c > 0$ be some arbitrarily small constant. Let P be a quadratic form on G_n. Let $n/9 \leq k \leq n/2(1-c) - 1$. Suppose that there is a set $X \subset A_k$ of size at least $q^{(1-\epsilon)k}$ such that for any $d \in X$, the rank of the quadratic form $w \mapsto P(dw)$ is at most ϵn on G_{n-k}. Then the rank of P is $O((n\sqrt{\epsilon}/c))$.*

Observe that the form $w \mapsto P(dw)$ on G_{n-k} is equivalent to the restriction $P|_{(d)_n}$ of the form P to the subspace $(d)_n = dG_{n-k} \leq G_n$ of multiples of d, because the map $w \mapsto dw$ is a linear isomorphism $G_{n-k} \to (d)_n$. So if $d \in X$, we know that the restriction $P|_{(d)}$ is of small rank. Let B the symmetric bilinear form underlying P. Thus for any $x \in (d)_n$, the bilinear form B has small rank on $(d)_n \times (d_n)$, that is, the rank of the restriction of B to $(d)_n \times (d_n)$ is small. And if d' is any other element of X, the bilinear form B has rank at

most $2\epsilon n$ on $((d)_n \cap (d')_n) \times ((d)_n + (d')_n)$. Note that because $k < n/2$, the intersection $(d)_n \cap (d')_n = (\mathrm{lcm}(d, d'))_n \supset (dd')_n$ is not empty, its codimension being at most $2k$.

We see here that spaces of multiples and their sums will play a crucial role for the discussion. Because of Bézout's theorem, if $\gcd(d, d') = 1$, then the sum of the two ideals (d) and (d') in $\mathbb{F}_q[t]$ equals the whole algebra $\mathbb{F}_q[t]$, that is, $(d) + (d') = \mathbb{F}_q[t]$, but when we replace (d) by $(d)_n = (d) \cap G_n$, the situation may change. If $\deg d = \deg d' = k > n/2$, one cannot have $(d)_n + (d')_n = G_n$ because $\dim(d)_n = \dim(d')_n = n - k < n/2$. On the contrary, if $k \leq n/2$, one has

$$\dim\big((d)_n + (d')_n\big) = 2(n-k) - \dim\big((d)_n \cap (d')_n\big) = 2(n-k) - \dim(\mathrm{lcm}(d, d'))_n = n$$

so that $(d)_n + (d')_n = G_n$. The same argument shows that if d and d' are "almost coprime" in the sense that (d, d') has degree at most γ, then $(d)_n + (d')_n$ almost equals G_n in the sense that it has codimension at most γ in G_n.

We now prove Theorem D.1. We first claim that for at least $|X|^2(1 - O(q^{-\epsilon k}))$ pairs $(a, b) \in X^2$, we have $\deg \gcd(a, b) \leq 3\epsilon k$. Indeed, for a given $m \in (3\epsilon k, k)$ and a polynomial $d \in A_m$, there exist at most $q^{2(k-m)}$ pairs of monic polynomials of degree k divisible by d. As a result, the number of pairs of monic polynomials of degree k whose gcd has degree more than $3\epsilon k$ is bounded by $\sum_{3\epsilon k < m \leq k} q^m q^{2k-2m} = O(q^{2k-3\epsilon k})$, which proves the claim.

Let $D = \{(d, d') \in X^2 \mid \deg \gcd(d, d') \leq 3\epsilon k\}$. By the above, we have $|D| \gg q^{2(1-\epsilon)k}$. Consider $E = \{dd' \mid (d, d') \in D\}$. The Cauchy-Schwarz inequality and a bound on the second moment of the divisor function (Lemma E.2) shows that $|E| \geq D^4/(q^{2k}k^3) \geq q^{2k(1-2\epsilon)}$. Besides for any $e = dd'$ in E (with $(d, d') \in X$), the form B has rank at most $2\epsilon n$ on $(e)_n \times ((d)_n + (d')_n)$, and hence at most $\leq 5\epsilon n$ on $(e)_n \times G_n$.

The next proposition shows that because of the size of $E \subset A_{2k}$ (where $2k$ is bounded away from n), for t sufficiently large, for most t-tuples $(e_1, \ldots, e_t)$ from E^t, the sum $\sum_{i=1}^t (e_i)_n$ covers almost all the space G_n. In the next lemma, for any $d \in G_\ell$, we abbreviate $(d)_n = dG_{n-\ell}$ into (d). With this convention, (d) is a subspace of G_{n-1} and thus of G_n.

Proposition D.2. *Let $c > 0$ and $\ell \leq n(1-c)$. Suppose $E \subset G_\ell$ contains at least $q^{(1-\eta)\ell}$ elements. Let $\eta < \eta' < 1$ be another constant. Then there exists $t = O(1/(c\eta'))$ and $(e_1, \ldots, e_t) \in E^t$ such that* $\mathrm{codim}_{G_n} \sum_{i=1}^t (e_i) = O(\eta' n)$.

The sumspace is in fact inside G_{n-1}, so we could consider $\mathrm{codim}_{G_{n-1}}$ instead of codim_{G_n}, but they differ by 1 only.

Proof. Note that if $\ell \leq n/2$, the arguments at the beginning of the section yield the conclusion, taking $t = 2$ and e_1, e_2 approximately coprime. So we suppose instead that $\ell > n/2$. Let $(e_1, \dots, e_t) \in E^t$, for some t to be determined later. Let $S = \sum_{i=1}^t (e_i)$, where (e_i) is a shortcut for $(e_i)_n = e_i G_{n-\ell}$. We introduce the functions $f_i = 1_{(e_i)}$ and the convolution

$$g(x) = f_1 * \cdots * f_t(x) = \mathbb{E}_{x=\sum_i y_i} \prod_{i=1}^t f_i(y_i).$$

Thus

$$q^{t(n-\ell)} g(x) = |\{(c_1, \dots, c_t) \in G_{n-\ell}^t \mid x = \sum_{i=1}^t c_i e_i\}|.$$

The aim is to bound the codimension of the subspace S in G_n. Observe that for any $x \in S$, we have $g(x) = g(0)$. Besides, the solutions $(c_1, \dots, c_t) \in G_{n-\ell}^t$ to the equation $\sum_{i=1}^t c_i e_i = 0$ form a linear subspace, so that $g(0) = q^{-s}$ for some $s \in \mathbb{N}$. Thus $|S| q^{-s} = 1$ and so we need to show that $s = n - O(\eta' n)$.

We then use the circle method, which basically consists in the identity

$$1_{x=\sum_{i=1}^t x_i} = \int_\alpha e(\alpha(x - \sum_{i=1}^t x_i)) d\alpha$$

where the integral is with respect to the Haar measure on $\mathbb{T}$. Thus

$$g(x) = \int_\alpha e(-\alpha x) \sum_{(c_1,\dots,c_t) \in G_{n-\ell}^t} \prod_{i=1}^t e(\alpha c_i e_i) d\alpha.$$

We need to pause to introduce some notation. Let $\beta = \sum_{i=-\infty}^m \beta_i t^i \in \mathbb{F}_q((t^{-1}))$ for some $m \in \mathbb{Z}$. We write $\{\beta\} = \sum_{i=-\infty}^{\min(m,-1)} \in \mathbb{T}$ for the fractional part of β, so that $\beta - \{\beta\} \in \mathbb{F}_q[t]$, and $\|\beta\| = |\{\beta\}| \leq q^{-1}$ for the distance of β to the closest polynomial. Now we observe that

$$\sum_{c \in G_m} e(\beta c) = 1_{\|\beta\| < q^{-m}}$$

for any $m \in \mathbb{N}$ and $\beta \in \mathbb{F}_q((1/t))$. As a result,

$$g(0) = \int_\alpha \prod_{i=1}^t 1_{|\{\alpha e_i\}| < q^{-n+\ell}} \, d\alpha.$$

Observe that this integral is at least q^{-n}, because the integrand is constantly 1 on the set

$t^{-n}\,\mathbb{T} \subset \mathbb{T}$ whose measure is q^{-n}. Our aim is to show that on average over $(e_1, \ldots, e_t)$, the integral is not much larger, that is, at most $q^{-n+O(\eta' n)}$. So we now consider

$$\mathbb{E}_{(e_1,\ldots,e_t)\in G_\ell^t} g(0) = \mathbb{E}_{(e_1,\ldots,e_t)\in G_\ell^t} \int_\alpha \prod_{i=1}^{t} 1_{|\{\alpha e_i\}|<q^{-n+\ell}}\, d\alpha = \int \left(\mathbb{E}_{e\in G_\ell} 1_{|\{\alpha e\}|<q^{-n+\ell}}\right)^t d\alpha.$$

Fix $\alpha = \sum_{i=-\infty}^{-1} \alpha_i t^i$. The map $e \mapsto \{\alpha e\}$ is linear, and for $\beta = \sum_{i=-\infty}^{m} \beta_i t^i$, the condition $|\beta| < q^{-m}$ is linear. Consequently, we note that $\mathbb{E}_{e\in G_\ell} 1_{|\{\alpha e\}|<q^{-n+\ell}} = q^{-\operatorname{rk} M_\alpha}$, where M_α is the rectangular $(n-\ell)\times \ell$ Hankel matrix

$$M_\alpha = \begin{pmatrix} \alpha_{-1} & \cdots & \alpha_{-\ell} \\ \alpha_{-2} & \cdots & \alpha_{-\ell-1} \\ \vdots & \vdots & \vdots \\ \alpha_{-n+\ell} & \cdots & \alpha_{-n+1} \end{pmatrix}.$$

Now we aim at showing that this matrix is "almost surely" (in the sense of the Haar probability measure) of large rank. To do that, we provide a characterisation of the rank of Hankel matrices.

Lemma D.3. *If* $\operatorname{rk} M_\alpha = r < n-\ell$*, then there exists a decomposition* $r = i + h$ *such that the first* i *rows are independent, the next* $n-\ell-r$ *rows are a linear combination of the first* i *rows, and the minor formed of the first* i *and last* h *rows and columns is nonzero.*

This statement is an easy consequence of Lemma 2 and Theorem 23 from [31, Chapter X, Paragraph 10]. We provide a further characterisation of the rank based on Diophantine properties of α.

Lemma D.4. *If* $\operatorname{rk} M_\alpha = r < n-\ell$*, then there exist a decomposition* $r = i+h$ *and* $d \in A_i$ *such that*

$$\|d\alpha\| < q^{-n+1+h}.$$

Proof. Let $L_1, \ldots, L_{n-\ell}$ be the rows of M_α. We invoke Lemma D.3. Let $r = i+h$ be the decomposition in its conclusion. In particular, we have $i < n-\ell$. For any $m \in [n-\ell]$, write $L_m = (L'_m, a_m)$ where a_m is the last coefficient of L_m and L'_m the row of the first $\ell-1$ coefficients. By Lemma D.3, we have a relation $L_{i+1} = \sum_{m=1}^{i} c_m L_m$. In fact, we shall

show that for any $j = 0, \dots, n - r - \ell - 1$, we have

$$L_{i+1+j} = \sum_{m=1}^{i} c_m L_{m+j}. \tag{D.1}$$

This equation holds for $j = 0$. So we argue by induction and assume equation (D.1) holds for any $j' \leq j$ for some $j < n-r-\ell-1$ and prove it for $j+1$. To start with, equation (D.1) implies that $L'_{i+1+j+1} = \sum_{m=1}^{i} c_m L'_{m+j+1}$. Applying the induction hypothesis iteratively, we find coefficients $c_m^{(k)}$ for $m = 1, \dots, i$ and $k \leq j+1$ such that

$$L_{i+1+k} = \sum_{m=1}^{i} c_m^{(k)} L_m$$

and

$$L'_{i+1+j+1} = \sum_{m=1}^{i} c_m^{(j+1)} L'_m. \tag{D.2}$$

These coefficients satisfy the initial condition $c_m^{(0)} = c_m$ and the recurrence relations $c_m^{(k+1)} = c_{m-1}^{(k)} + c_i c_m^{(k)}$ for $m > 1$ and $c_1^{(k)} = c_h^k c_1$.

On the other hand, we know that there exist coefficients $d_1, \dots, d_i$ such that

$$L_{i+1+j+1} = \sum_{m=1}^{i} d_m L_m. \tag{D.3}$$

Comparing equations (D.3) and (D.2), and using the linear independence of the first i rows, we find that $d_m = c_m^{(j+1)}$.

Thus $a_{i+1+j+1} = \sum_{m=1}^{i} c_m^{(j+1)} a_m$. But $\sum_{m=1}^{i} c_m a_{m+j+1} = \sum_{m=1}^{i} c_m^{(j+1)} a_m$ by definition of the coefficients $c_m^{(k)}$. We infer that $a_{i+1+j+1} = \sum_{m=1}^{i} c_m a_{m+j+1}$, which concludes the inductive argument.

To see the connexion with Diophantine properties of α, notice that L_i is the row of the first n coefficients of $\{t^{i-1}\alpha\}$. Thus the validity of the identity (D.1) for all $j = 0, \dots, n - r - \ell - 1$ implies that $\{t^i \alpha\} = \{\sum_{m=1}^{i} c_m t^{m-1} \alpha\} + \beta$ where $\beta \in \mathbb{T}$ satisfies $|\beta| < q^{-n+1+h}$. That is, using the polynomial $P = t^i - \sum_{m=1}^{i} c_m t^{m-1}$, we find that $\|P\alpha\| < q^{-n+1+h}$. This concludes the proof of the lemma. □

Now we want to infer from Lemma D.4 that Hankel matrices of low rank are very

rare. Suppose $\operatorname{rk} M_\alpha \leq r$ and let i, h be as in the conclusion of the lemma. Let $d \in A_i$ be such that $\|d\alpha\| < q^{-n+1+h}$. We note that the map $\alpha \mapsto \{d\alpha\}$ is linear. As a result, for $\|d\alpha\| < q^{-m}$ to hold, the vector $\overrightarrow{\alpha} = (\alpha_{-1}, \ldots, \alpha_{-i-m})$ has to be in the kernel of the $m \times (m+i)$ matrix

$$L_d = \begin{pmatrix} d_0 & \cdots & d_{i-1} & 1 & 0 & \cdots & \cdots & 0 \\ 0 & d_0 & \cdots & d_{i-1} & 1 & 0 & \cdots & 0 \\ & & \ddots & & & \ddots & & \\ & & & \ddots & & & \ddots & \\ 0 & & & & d_0 & \cdots & d_{i-1} & 1 \end{pmatrix},$$

whose rank is m. Applying this observation with $m = n-1-h$, we see that the probability that α lies in the kernel of L_d is q^{-n+h+1}. The map $\alpha \mapsto \overrightarrow{\alpha}$ being a measure preserving operation from $\mathbb{T}$ to $\mathbb{F}_q^{m+i}$, the probability for $\alpha \in \mathbb{T}$ to satisfy $\|\alpha d\| < q^{-n+1+h}$ is q^{-n+h+1}. By the triangle inequality, the probability that α satisfies $\|\alpha d\| < q^{-n+1+h}$ for at least one $d \in A_i$ is at most q^{-n+1+r}. But then the probability that $\operatorname{rk} M_\alpha \leq r$ is bounded by $r^2 q^{-n+1+r}$. On other other hand, when $\operatorname{rk} M_\alpha \geq r$, we have $\mathbb{E}_{e \in G_\ell} 1_{|\{\alpha e\}| < q^{-n+\ell}} \leq q^{-r}$. All in all, this implies that

$$\int \left(\mathbb{E}_{e \in G_\ell} 1_{|\{\alpha e\}| < q^{-n+\ell}}\right)^t d\alpha \leq r^2 q^{-n+1+r} + q^{-tr}. \tag{D.4}$$

We take r of the form $r = \eta'(n-\ell)/4 \geq c\eta' n/4$, and $t = 4/(c\eta')$. This way, both terms are $O(q^{-(1-\eta')n})$. So we use Markov's identity to infer that

$$\mathbb{P}_{e_1, \ldots, e_t \in G_\ell} \left(\int_\alpha \prod_i 1_{\|\alpha e_i\| < q^{-n+\ell}} d\alpha > q^{-n+2\eta' n} \right) = O(q^{-\eta' n}).$$

In other words,

$$\mathbb{P}_{e_1, \ldots, e_t \in G_\ell} \left(\int_\alpha \prod_i 1_{\|\alpha e_i\| < q^{-n+\ell}} d\alpha \leq q^{-n+2\eta' n} \right) > 1 - O(q^{-\eta' n}).$$

Because $\eta < \eta'$, one can ensure that there exists $(e_1, \ldots, e_t) \in E^t$ such that

$$g(0) = \int_\alpha \prod_i 1_{\|\alpha e_i\| < q^{-n+\ell}} d\alpha \leq q^{-n+2\eta' n}.$$

Finally, $g(0) = q^{n-O(\eta' n)}$, which concludes the proof of Proposition D.2. $\square$

We finish the proof of Theorem D.1. Write $\ell = 2k+1 \leq (1-2c)n$. The observations at the beginning of this section imply that there exists a set $E \subset G_\ell$ of cardinality $q^{(1-2\epsilon)\ell-1}$ such that for any $e \in E$, the rank of B is at most $5\epsilon n$ on $(e) \times G_n$. We then apply Proposition D.2 for some value of $\eta' > 2\epsilon$ to determine later, which provides us with t elements $e_1, \ldots, e_t$ of E such that $\operatorname{codim}_{G_n} \sum_{i=1}^{t}(e_i) = O(\eta' n)$. In particular, we infer that the total rank of B is at most $(5t\epsilon + O(\eta'))n$. Besides, we have $t = O((\eta' c)^{-1})$, so selecting $\eta' = \sqrt{\epsilon} + 2\epsilon$, we conclude that the rank of B is $O(\sqrt{\epsilon} n/c)$, which concludes the proof.

Appendix E

Divisor bounds

We list some facts regarding the divisor function in $\mathbb{F}_q[t]$ which we will need in the sequel. Let $\tau(f)$ denote the number of monic divisors of $f \in \mathbb{F}_q[t]$. We first give a pointwise bound.

Lemma E.1 ([60, Lemma 8])**.** *If* $\deg f = n > 1$*, then*

$$\tau(f) \leq \exp\left(O_q\left(\frac{n}{\log n}\right)\right).$$

Consequently, the number of monic irreducible factors of f *is* $O_q\left(\frac{n}{\log n}\right)$.

The next result is a bound for the second moment of τ.

Lemma E.2. *We have*

$$\mathbb{E}_{\deg d=n}\tau(d)^2 \leq 4n^3.$$

Proof. We observe that for any irreducible P and any integer k, we have $\tau(P^k)^2 = (k+1)^2$. Thus the Dirichlet series $D = \sum_{n=0}^{+\infty}\sum_{f\in A_n}\frac{\tau(f)^2}{|f|^s}$ of the function τ^2 can be written as an Euler product as

$$D = \prod_P \sum_{k=0}^{+\infty}(k+1)^2|P|^{-ks}. \tag{E.1}$$

Next we note the following relations between formal power series

$$\sum_{k=0}^{+\infty}(k+1)^2x^k = \sum_{k=0}^{+\infty}(k+2)(k+1)x^k - \sum_{k=0}^{+\infty}(k+1)x^k = 2(1-x)^{-3} - (1-x)^{-2} = \frac{1+x}{(1-x)^3}$$

so finally

$$\sum_{k=0}^{+\infty}(k+1)^2x^k = \frac{1-x^2}{(1-x)^4}. \tag{E.2}$$

Combining equations (E.1) and (E.2) yields

$$D = \prod_P \frac{1-|P|^{-2s}}{(1-|P|^{-s})^4}.$$

We can then express this Euler product in terms of the zeta function of $\mathbb{F}_q[t]$. Letting $u = q^{-s}$ we obtain

$$D = \zeta(s)^4/\zeta(2s) = (1-q^{1-2s})(1-q^{1-s})^{-4} = (1-qu^2)(1-qu)^{-4}.$$

This is a power series $S(u)$ in u, and $S(u) = \sum_n a_n u^n = \sum_n \frac{S^{(n)}(0)}{n!}u^n$ where $a_n = \sum_{\deg d=n}\tau(d)^2$. Now for $n \geq 3$, deriving n times using Leibniz' formula, we find that

$$\begin{aligned} S^{(n)}(u) &= (1-qu^2)q^n(4\times\cdots\times(n+3))(1-qu)^{-4-n} \\ &- 2qunq^{n-1}(4\times\cdots\times(n+2))(1-qu)^{-3-n} \\ &- 2q\binom{n}{2}q^{n-2}(4\times\cdots\times(n+1))(1-qu)^{-2-n} \end{aligned}$$

Evaluating in $u = 0$ gives

$$\frac{S^{(n)}(0)}{q^n n!} = (n+3)(n+2)(n+1)/6 - q^{-1}n(n+1)^2/6 \leq 4n^3,$$

where the left-hand side is exactly $\mathbb{E}_{\deg d=n}\tau(d)^2$. $\square$

Bibliography

[1] M. Ajtai and E. Szemerédi. *Sets of lattice points that form no squares.* Stud. Sci. Math. Hungar., 9:9–11, 1974.

[2] N. Alon and J. Spencer. *The probabilistic method.* Fourth edition. Wiley Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, 2016.

[3] R. C. Baker and G. Harman. *Exponential sums formed with the Möbius Function.* J. London Math. Soc. (2), 43(2):193–198, 1991.

[4] L. Bary-Soroker. *Hardy-Littlewood tuple conjecture over large finite fields.* Int. Math. Res. Not., rns 249, 8 pp, 2012.

[5] F.A. Behrend. *On sets of integers which contain no three terms in arithmetical progression.* Proc. Nat. Acad. Sci. U. S. A., 32:331–332, 1946.

[6] V. Bergelson, T. Tao, and T. Ziegler, *An inverse theorem for the uniformity seminorms associated with the action of* $\mathbb{F}_p^\infty$, Geom. Funct. Anal., 19(6):1539–1596, 2010.

[7] A. Bhowmick, T. H. Lê and Y-R. Liu. *A note on character sums in finite fields.* Finite Fields Appl. 46:247–254, 2017.

[8] P.-Y. Bienvenu. *Comments on the Ellenberg-Gijswijt bound for caps in higher characteristic.* Blog post, `https://blogderbeweise.wordpress.com/2016/07/29/`.

[9] P.-Y. Bienvenu. *A higher-dimensional Siegel–Walfisz theorem.* Acta Arith., 179(1):79–100, 2017.

[10] P.-Y. Bienvenu. *Asymptotics for some polynomial patterns in the primes.* Proc. Roy. Soc. Edin., 2018, to appear.

[11] P.-Y. Bienvenu. *Polynomial equations in* $\mathbb{F}_q[t]$. Q. J. Math., 68(4):1395–1398, 2017.

[12] P.-Y. Bienvenu and T. H. Lê. *A bilinear Bogolyubov theorem.* `arXiv:1711.05349`.

[13] P.-Y. Bienvenu and T. H. Lê. *Linear and quadratic uniformity of the Möbius function over* $\mathbb{F}_q[t]$. `arXiv:1711.05358`.

[14] T. F. Bloom. *A quantitative improvement for Roth's theorem on arithmetic progressions.* J. Lond. Math. Soc. (2), 93(3):643–663, 2016.

[15] N. Bogoliuboff. *Sur quelques propriétés arithmétiques des presque-périodes.* Ann. Chaire Phys. Math. Kiev, 4:185–205, 1939

[16] J. Bourgain. *Roth's theorem on progressions revisited.* J. Anal. Math., 104:155–192, 2008.

[17] T. D. Browning and L. Matthiesen. *Norm forms for arbitrary number fields as products of linear polynomials.* Annales scientifiques de l'ENS 50(6): 1383-1446, 2017.

[18] T. D. Browning and S. M. Prendiville. *A transference approach to a Roth-type theorem in the squares.* Int. Math. Res. Not., (7):2219–2248, 2017.

[19] M. Car. *Distribution des polynômes irréductibles dans* $\mathbb{F}_q[T]$. Acta Arith., 88(2):141–153, 1999.

[20] J.W.S. Cassels. *An introduction to the geometry of numbers.* Classics in Mathematics, Springer-Verlag Berlin Heidelberg, 1957.

[21] S. Chow. *Roth–Waring–Goldbach.* Int. Math. Res. Not., rnw 307, `https://doi.org/10.1093/imrn/rnw307`, 2017.

[22] D. Conlon, J. Fox, and Y. Zhao. *The Green-Tao theorem: an exposition.* EMS Surv. Math. Sci., 1(2):249–282, 2014.

[23] H. Cramér. *Sur un nouveau théorème limite de la théorie des probabilités.* Actual. Sci. Indust., 736:5–23, 1938.

[24] E. Croot, V.F. Lev, and P.P. Pach. *Progression-free sets in* $\mathbb{Z}_4^n$ *are exponentially small.* Ann. Math., 185(1):331–337, 2017.

[25] H. Davenport. *On some infinite series involving arithmetical functions.* Quart. J. Math., 8:8–13, 1937.

[26] J. Ellenberg and D Gijswijt. *On large subsets of $\mathbb{F}_3^n$ with no three-term arithmetic progression.* Ann. Math., 185(1):339–343, 2017.

[27] P. Erdős. *On the sum $\sum_{k=1}^{x} d(f(k))$.* J. London Math. Soc., 27:7–15, 1952.

[28] P. Erdős. *Problems and results in combinatorial number theory.* Proc. Internat. Sympos., Colorado State Univ., Fort Collins, Colo., pp. 117–138. Astérisque, Nos. 24–25, 1971.

[29] K. Ford, B. Green, S. Konyagin, and T. Tao. *Large gaps between consecutive prime numbers.* Ann. of Math. (2), 183(3):935 – 974, 2016.

[30] E. Fouvry. *Sur le problème des diviseurs de Titchmarsh.* J. Reine Angew. Math., 357:51–76, 1984.

[31] F. Gantmacher. *The theory of matrices.* Translated by K. A. Hirsch. Chelsea Publishing Co., New York 1959.

[32] D. A. Goldston, J. Pintz, and C. Y. Yıldırım. *Primes in tuples. I.* Ann. of Math. (2), 170(2):819–862, 2009.

[33] D. A. Goldston, J. Pintz, and C. Y. Yıldırım. *Primes in tuples IV: Density of small gaps between consecutive primes.* Acta Arith., 160(1):37–53, 2013.

[34] W. T. Gowers. *A new proof of Szemerédi's theorem.* Geom. Funct. Anal., 11(3):465–588, 2001.

[35] W. T. Gowers and L. Milićević. *A bilinear version of Bogolyubov's theorem.* `arXiv:1712.00248`.

[36] W. T. Gowers and L. Milićević. *A quantitative inverse theorem for the U^4 norm over finite fields.* `arXiv:1712.00241` .

[37] B. Green. *On arithmetic structures in dense sets of integers.* Duke Math. J., 114(2):215–238, 2002.

[38] B. Green. *Finite field models in additive combinatorics.* In *Surveys in combinatorics 2005*, volume 327 of London Math. Soc. Lecture Note Ser., pages 1–27. Cambridge Univ. Press, Cambridge, 2005.

[39] B. Green. *An argument of Gowers in the finite field setting.* `http://people.maths.ox.ac.uk/greenbj/papers/corners.pdf`.

[40] B. Green. *Montréal notes on quadratic Fourier analysis.* In *Additive combinatorics*, volume 43 of CRM Proc. Lecture Notes, pages 69–102. Amer. Math. Soc., Providence, RI, 2007.

[41] B. Green. *Notes on progressions and convex geometry.* `http://people.maths.ox.ac.uk/greenbj/papers/convexnotes.pdf`.

[42] B. Green. *Sárközy's theorem in function fields.* Q. J. Math., 68(1):237–242, 2017.

[43] B. Green and T. Tao. *An inverse theorem for the Gowers $U^3(G)$ norm.* Proc. Edinb. Math. Soc. (2), 51(1):73–153, 2008.

[44] B. Green and T. Tao. *The primes contain arbitrarily long arithmetic progressions.* Ann. of Math. (2), 167(2):481—547, 2008.

[45] B. Green and T. Tao. *Linear equations in primes.* Ann. of Math. (2), 171(3):1753–1850, 2010.

[46] B. Green and T. Tao. *Quadratic uniformity of the Möbius function.* Ann. Inst. Fourier, 58(6):1863–1935, 2008.

[47] B. Green and T. Tao. *The Möbius function is strongly orthogonal to nilsequences.* Ann. of Math. (2), 175(2):541–566, 2012.

[48] B. Green, T. Tao, and T. Ziegler. *An inverse theorem for the Gowers $U^{s+1}[N]$-norm.* Ann. of Math. (2), 176(2):1231–1372, 2012.

[49] G.H. Hardy and J.E. Littlewood. *Some problems of partitio numerorum: on the expression of a number as a sum of primes.* Acta Math., 44:1–70, 1923.

[50] D. R. Hayes. *The distribution of irreducibles in $GF[q, x]$.* Transactions of the AMS, 117:101–127, 1965.

[51] D. R. Hayes. *The expression of a polynomial as a sum of three irreducibles.* Acta Arith., 11(4):461–488, 1966.

[52] X. He and B. Huang. *Exponential sums involving the Möbius function.* Acta Arith., 175(3):201–209, 2016.

[53] K. Henriot. *Logarithmic bounds for translation-invariant equations in squares.* Int. Math. Res. Not., (23):12540–12562, 2015.

[54] C-N. Hsu. *The distribution of irreducible polynomials in* $\mathbb{F}_q[t]$. Journal of Number Theory, 61:85–96, 1996.

[55] H. Iwaniec and E. Kowalski. *Analytic number theory*, volume 53 of *American Mathematical Society Colloquium Publications.* American Mathematical Society, Providence, RI, 2004.

[56] E. Keil. *On a diagonal quadric in dense variables.* Glasg. Math. J., 56(3):601–628, 2014.

[57] Y. Kohayakawa, T. Łuczak, and V. Rödl. *Arithmetic progressions of length three in subsets of a random set.* Acta Arith., 75(2):133–163, 1996.

[58] J. Lafontaine. *An introduction to differential manifolds* Based on the 2010 French second edition. Springer, Cham, 2015.

[59] B. Landreau. *A new proof of a theorem of van der Corput.* Bull. London Math. Soc., 21:366–368, 1989.

[60] T. H. Lê. *Green-Tao theorem in function fields.* Acta Arith., 147(2):129–152, 2011.

[61] T. H. Lê and Y.-R. Liu. *On sets of polynomials whose difference set contains no squares.* Acta Arith., 161(2):127–143, 2013.

[62] T.H. Lê and J. Wolf. *Polynomial configurations in the primes.* Int. Math. Res. Not., (23):6448–6473, 2014.

[63] Ju. V. Linnik. *The dispersion method in binary additive problems.* Translated by S. Schuur. American Mathematical Society, Providence, R.I., 1963.

[64] Y.-R. Liu and T. D. Wooley. *Waring's problem in function fields.* J. Reine Angew. Math., 638:1–67, 2010.

[65] S. Lovett. *An Exposition of Sanders' Quasi-Polynomial Freiman-Ruzsa Theorem.* Theory of Computing Library. Graduate surveys 6, 1–14, 2015.

[66] L. Matthiesen. *Correlations of the divisor function.* Proc. Lond. Math. Soc. (3), 104(4):827–858, 2012.

[67] L. Matthiesen. *Linear correlations of multiplicative functions.* `arXiv:1606.04482`.

[68] L. Matthiesen. *Linear correlations amongst numbers represented by positive definite binary quadratic forms.* Acta Arith., 154(3):235–306, 2012.

[69] J. Maynard. *Almost-prime k-tuples.* Mathematika, 60(1):108–138, 2014.

[70] R. Meshulam. *On subsets of finite abelian groups with no 3-term arithmetic progressions.* J. Combin. Theory Ser. A, 71(1):168–172, 1995.

[71] L. Mirsky. *The number of representations of an integer as the sum of a prime and a k-free integer.* Amer. Math. Monthly, 56:17–19, 1949.

[72] S. Porritt. *A note on exponential-Möbius sums over $\mathbb{F}_q[t]$.* `arXiv:1711.08729`.

[73] G. Rhin. *Répartition modulo 1 dans un corps de séries formelles sur un corps fini.* Dissertationes Math. (Rozprawy Mat.), 95:75, 1972.

[74] M. Rosen. *Number theory in function fields.* Graduate Text in Mathematics (210), Springer-Verlag New York.

[75] K. Roth. *On certain sets of integers.* J. London Math. Soc., 28:104–109, 1953.

[76] P. Samuel. *Géométrie projective.* Presses Universitaires de France, Paris, 1986.

[77] T. Sanders. *On Roth's theorem on progressions.* Ann. of Math. (2), 174(1):619–636, 2011.

[78] T. Sanders. *On the Bogolyubov-Ruzsa lemma.* Anal. PDE, 5(3):627–655, 2012.

[79] A. Sárközy. *On difference sets of sequences of integers. III.* Acta Math. Acad. Sci. Hungar., 31(3-4):355–386, 1978.

[80] I. D. Shkredov. *On a problem of Gowers.* Izv. Ross. Akad. Nauk Ser. Mat., 70(2):179–221, 2006.

[81] M. L. Smith. *On solution-free sets for simultaneous quadratic and linear equations.* J. Lond. Math. Soc. (2), 79(2):273–293, 2009.

[82] E. Szemerédi. *On sets of integers containing no k elements in arithmetic progressions.* Acta Arith., 27:199–245, 1975

[83] T.Tao and V. Vu. *Additive combinatorics.* Cambridge Studies in Advanced Mathematics, 105. Cambridge University Press, Cambridge, 2010.

[84] T. Tao. *A symmetric formulation of the Croot-Lev-Pach-Ellenberg-Gijswijt capset bound.* Personal blog post, `https://terrytao.wordpress.com/2016/05/18/`.

[85] T. Tao and T. Ziegler. *The inverse conjecture for the Gowers norm over finite fields via the correspondence principle.* Anal. PDE, 3(1):1–20, 2010.

[86] T. Tao and T.Ziegler. *The inverse conjecture for the Gowers norm over finite fields in low characteristic.* Ann. Comb. 16(1), 121–188, 2012.

[87] T. Tao and T. Ziegler. *Narrow progressions in the primes.* In *Analytic number theory*, In Honor of Helmut Maier's 60th Birthday, pages 357–379. Springer, Cham, 2015.

[88] T. Tao and T. Ziegler. *The primes contain arbitrarily long polynomial progressions.* Acta Math., 201(2):213–305, 2008.

[89] T. Tao and T. Ziegler. *Polynomial patterns in the primes.* Forum of Mathematics, Pi, 6:e1, 2018.

[90] T. Tao and T. Ziegler. *Concatenation Theorems for Anti-Gowers-Uniform Functions and Host-Kra Characteristic Factors.* Discrete Anal. 2016:13, 61 pp, 2016.

[91] E.C. Titchmarsh. *A divisor problem.* Rend. di Palermo, 54:414–429, 1931.

[92] J.G. van der Corput. *Une inégalité relative au nombre de diviseurs.* Koniklijke Nederlandsche Akad. Wet. Proceedings 42:547-553, 1939.

[93] J. Wolf. *Finite field models in arithmetic combinatorics—ten years on.* Finite Fields Appl., 32:233–274, 2015.

[94] D. Zeilberger. *A Motivated Rendition of the Ellenberg-Gijswijt Gorgeous proof that the Largest Subset of $\mathbb{F}_3^n$ with No Three-Term Arithmetic Progression is $O(c^n)$, with* $c = \sqrt[3]{(5589 + 891\sqrt{33})}/8 = 2.755104613023633000221 27\ldots$. `arXiv:1607.01804`.

[95] T. Zhan, J.-Y. Liu. *Exponential sums involving the Möbius function*. Indag. Math. (N.S.), 7(2):271–278, 1996.