Quadratic patterns in the primes

Pierre-Yves Bienvenu

July 30, 2015
Abstract

This article provides asymptotics for sums of the form

\[ \sum_{n \in \mathbb{Z}^d \cap K} \prod_{i=1}^{t} F_i(\psi_i(n)) \]

where the \( F_i \)'s can be the von Mangoldt function or the representation function of a quadratic form, and the \( \psi_i \)'s are affine-linear forms and \( K \) is a convex body. As an application, we provide the asymptotic frequency of \( k \)-term arithmetic progressions where the common difference is a sum of two squares. To the best of our knowledge, this might be the first time an asymptotic of the frequency of some polynomial pattern in the primes is derived.

The article combines the works of Green and Tao on linear equations in primes and Matthiesen on linear correlations between primes. To make the cohabitation of the von Mangoldt function with the representation function possible, we provide a pseudorandom majorant, which is simply an average of the already well-known majorants for each of the functions; the fact that this average still has the randomness required is not obvious and is a novelty of this article.
1 History and introduction

1.1 Early results on patterns in the primes

When interested in the frequency of patterns in the primes, the very first question is to ask how many primes (of which the set is denoted by \( \mathcal{P} \)) there are up to \( N \). The answer is the famous prime-number theorem, already hinted at by Legendre and Gauss at the end of the XVIIIth and fully proven at the turn of the XIXth century, after the works of Hadamard and De la Vallée-Poussin, which states that

\[
\sum_{n \leq N} 1_{\mathcal{P}}(n) \sim \frac{N}{\log N}
\]

or, equivalently and more conveniently, that

\[
\sum_{n \leq N} \Lambda(n) \sim N,
\]

where \( \Lambda \), the von Mangoldt function, is defined by \( \Lambda(n) = \log p \) if \( n \) is a power of some prime \( p \), and \( \Lambda(n) = 0 \) otherwise. It is not much harder to obtain the asymptotic

\[
\sum_{q_1 n + q_0 \leq N} \Lambda(q_1 n + q_0) \sim \frac{\varphi(q_1)}{q_1} N,
\]

for \((q_1, q_0)\) coprimes, which is the definitive asymptotic result for the number of prime values taken by an affine-linear form \( \mathbb{Z} \rightarrow \mathbb{Z} \).

Notational conventions. We need to stop to explain our conventions for asymptotic statements. First, the asymptotic parameter going to infinity will always be \( N \). The parameter \( N' \) introduced later goes to infinity with \( N \) so that the limit when \( N \) goes to infinity is the same as the limit when \( N' \) goes to infinity. We used and will use the symbols \( X \sim Y \) to say that \( X/Y \) tends to 1 as \( N \) tends to infinity. We shall use \( X = O(Y) \) so say that \( X/Y \) is bounded and \( X = o(Y) \) to say that \( X/Y \) tends to 0. Both \( O \) and \( o \) can be completed with a subscript indicating the dependence of the implied constant or the implied decaying function. We also use \( X \ll Y \) which is synonymous to \( X = O(Y) \) and can be completed by subscripts as well.

Hardy and Littlewood and Dickson (see [4]) conjectured an asymptotic for sums involving affine forms in one variable

\[
\sum_{n \leq N} \prod_{i=1}^{t} \Lambda(a_i n + b_i)
\]

but it turned out that such one-dimensional problems are difficult. Hardy and Littlewood considered only the case where all \( a_i = 1 \), and their problem is often referred to as the prime \( t \)-tuple conjecture. This type of problem includes the notoriously hard twin prime conjecture (put \( t = 2, a_1 = a_2 = 1, b_1 = 0, b_2 = 2 \)). We still do not even know if there are infinitely many twin primes \((n, n + 2)\), but we have a good upper bound for (1) in this case. They remain nowadays out of reach (although recent results on small gaps
betwin primes, especially since the articles of Goldston-Pintz-Yıldırım like [7], and the breakthroughs of Maynard from [15] shed a new light on the prime \( t \)-tuple conjecture), unlike some multi-dimensional problems, as we shall see later.

Early results, obtained in the twenties and thirties, include Vinogradov’s three primes theorem, an asymptotic result for the expression

\[
S_N = \sum_{1 \leq n_1 \leq n_1 + n_2 \leq N} \Lambda(n_1)\Lambda(n_2)\Lambda(N - n_1 - n_2)
\]

and van der Corput’s result about three-term arithmetic progressions (abbreviated as APs) in the primes, dealing with the sum

\[
\sum_{1 \leq n + 2d \leq N} \Lambda(n)\Lambda(n + d)\Lambda(n + 2d).
\]

Both results can be obtained by using the classical method of Hardy and Littlewood, also called circle method. The basic principle is to test the von Mangoldt function \( \Lambda \) against exponentials \( n \mapsto e^{\alpha n} \) for \( \alpha \) in \([0, 1]\) or in fact in the circle \( \mathbb{R}/2\pi\mathbb{Z} \), where \( e(x) \) stands for \( \exp(2i\pi x) \). Thus it involves considering the function

\[
F(\alpha) = \sum_{n \leq N} \Lambda(n)e(\alpha n).
\]

The circle method produced a wealth of other results, but we refrain from reviewing them all here and instead pass on a more recent and powerful method.

### 1.2 Green and Tao’s results and method

We recall a definition inspired from [9].

**Definition 1.1.** Let \( \Psi = (\psi_1, \cdots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t \) be a system of affine-linear forms. We say it has *finite complexity* if no two of the \( \psi_i \)'s are affinely related.

In this article, we will not need to quantify the complexity of a system properly so we refrain from defining it. Moreover the phrase “affine-linear forms” stands exclusively for forms with integer coefficients throughout the article. In [9], Green and Tao were able, conditionally on two conjectures they later proved completely (in the follow-up articles [11] and [10]), to formulate a vast generalisation of the Vinogradov and van der Corput theorems.

**Theorem 1.1.** Let \( \Psi = (\psi_1, \cdots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t \) 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 \( \Psi(K) \subset [0, N]^t \). Then

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

where

\[
\beta_\infty = \text{Vol}(K)
\]
and

\[ \beta_p = E_{a \in (\mathbb{Z}/p\mathbb{Z})^t} \prod_{i=1}^t \Lambda_p(\psi_i(a)) \]

with \( \Lambda_p(n) = \frac{p}{\phi(p)} 1_{(n,p)=1} \).

Their method, which can legitimately be described as the Hardy-Littlewood nilpotent method (term used in [14] for instance), involves testing \( \Lambda \) not only against exponentials, which are sequences arising from the abelian group \( \mathbb{R}/2\pi\mathbb{Z} \) known as the circle, but also against more general sequences called nilsequences because they arise from nilpotent groups. We will rely on this method in the course of the proof of our theorem, and hence detail the machinery.

1.3 Polynomial patterns in the primes

Again one-dimensional patterns are too hard to say anything meaningful about. For instance, it is not even known whether there exist infinitely many primes of the form \( n^2 + 1 \). The old conjecture of Bateman and Horn, which generalises the Hardy-Littlewood conjecture to a system \( P_1, \cdots, P_t \) of polynomials, suggests an asymptotic for the sum

\[ \sum_{n \leq N} \prod_{i=1}^t \Lambda(P_i(n)) \]

but is far beyond the reach of current methods.

Thus it is more reasonable to consider prime values of polynomials in several variables. For quadratic polynomials, Fermat had already characterised primes \( p \) of the form \( p = x^2 + y^2 \) (these are 2 and the primes congruent to 1 modulo 4), but also \( p = x^2 + 2y^2 \) (these are the primes congruent to 1 modulo 3) and \( p = x^2 + 3y^2 \) (these are the primes congruent to 1 or 3 modulo 8): these polynomial constraints happen to be equivalent to congruence conditions. Together with the result for primes in congruence class seen earlier, this yields an asymptotic for the number of primes up to \( N \) represented by these quadratic forms. More generally, primes of the form \( x^2 + ny^2 \) are characterized by Cox in the book [3]. Let us also remark that the circle method has been employed to evaluate the frequency of patterns where primes and powers mingle; for example Estermann evaluated asymptotically in [5] the number of triples \( (p_1, p_2, m) \) such that \( N = p_1 + p_2 + m^2 \), where \( p_1, p_2 \) are primes, for large \( N \).

We also mention that Friedlander and Iwaniec proved in [6] that there exist infinitely many primes of the form \( a^2 + b^4 \) and gave an asymptotic for the frequency of this pattern.

In 2008, Tao and Ziegler, pushing the ideas of Green and Tao from [8] and [9], were able to obtain the existence, of infinitely many non-trivial polynomial progressions of the form

\( (n + P_1(d), \cdots, n + P_t(d)) \)

in the primes, for any system of polynomials \( P_1, \cdots, P_t \) all vanishing in 0, with a lower bound for the number of occurrences of this pattern in the primes up to \( N \) of the same order of magnitude as the known upper bound from sieve theory.
1.4 Linear correlations amongst numbers represented by positive definite quadratic forms

The function $\Lambda$, or just as well $1_p \log$, is a weight concentrated on the sparse set of primes, which has an average of 1. Similarly the representation function $R$ of the sums of two squares, defined by

$$R(n) = \left| \{(x, y) \in \mathbb{Z}^2 \mid x^2 + y^2 = n \} \right|$$

is a weight supported on the sparse set of sums of two squares, which has a bounded non-zero average, namely $\pi$, as $\sum_{n \leq N} R(n)$ is simply the number of integral points in the disk $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq N\}$, which is asymptotic to its volume (see Lemma C.1).

So it was reasonable to imagine, and Matthiesen proved in [14], a result similar to Theorem 1.1 with $\Lambda$ being replaced by $R_f$ where $f$ is an arbitrary positive definite binary quadratic form (in short: PDBQF) and $R_f(n) = \left| \{(x, y) \in \mathbb{Z}^2 \mid f(x, y) = n \} \right|$. Her result, implying as an application an asymptotic for the number of simultaneous solutions to certain systems of quadratic equations, is the following.

**Theorem 1.2.** Let $\Psi = (\psi_1, \cdots, \psi_t) : \mathbb{Z}^d \rightarrow \mathbb{Z}^t$ 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 $\Psi(K) \subset [0, N]^t$. Let $f_1, \cdots, f_t$ be positive definite quadratic forms, of discriminants $D_i < 0$ for $i = 1, \cdots, t$. Then

$$\sum_{n \in \mathbb{Z}^d \cap K} \prod_{i=1}^{t} R_{f_i}(\psi_i(n)) = \beta_\infty \prod_{p \text{ prime}} \beta_p + o(N^d),$$

where

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

and

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

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

**Remark 1.1.** We remark that the local factors $\beta_p$ here involve considering the representation function of sums of two squares modulo arbitrarily large high powers of $p$ and hence taking a limit of expectations on $(\mathbb{Z}/p^m\mathbb{Z})^d$ as $m$ tends to infinity, while in the Green-Tao theorem 1.1, no such limit is necessary. This is because there, the local factor is about coprimality to $p$, which is the same as coprimality to $p^m$ for any $m$. Here, the limit would be superfluous if we had for any $m \geq 1$ and $d \in \mathbb{Z}$ the following lift-invariance property

$$\rho_d(p^m) = \rho_{d+kp^m}(p^{m+1})$$

but unfortunately it is not true for $d \equiv 0 \mod p^m$ (it is true for other $d$, see Corollary 6.4 of [14] which we will use later). Thus the limit cannot be ignored, and will again feature in our main theorem (Theorem 2.1).
Her proof relies on Green and Tao’s method, so a reasonable amount of compatibility between both proofs is expected, paving the way for our result.

We also notice that Matthiesen, together with Browning, was able to generalise in [1] her result from quadratic forms to norm forms originating from a number field. This implies a generalisation of our theorem, stated below, but we refrain, for the sake of simplicity, from inspecting this general case.

2 The main theorem and applications

Our theorem consists in merging Theorems 1.1 and 1.2 and its principal interest is to provide asymptotics for some particular quadratic configurations in the primes.

**Theorem 2.1.** Let $\Psi = (\psi_1, \cdots, \psi_{t+s}) : \mathbb{Z}^d \rightarrow \mathbb{Z}^{t+s}$ 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 $\Psi(K) \subset [0, N]^{t+s}$. Let $f_{t+1}, \cdots, f_{t+s}$ be positive definite quadratic forms, of discriminants $D_i < 0$ for $i = t+1, \cdots, t+s$. 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

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

and

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

with $\Lambda_p(n) = \frac{1}{\phi(p)} \sum_{(n, p) = 1} 1$ and $\rho_{f, \beta}(q) = \lfloor (x, y) \in [q]^2 \mid f(x, y) \equiv \beta \mod q \rfloor$.

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

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

by Fermat’s and Dirichlet’s theorems mentioned in Section 1.1.

We now show two attractive applications of this result. Here $R$ and $\rho$ will implicitly refer to the form $f(x, y) = x^2 + y^2$ whose discriminant is $-4$. The first application is a result concerning $k$-APs in primes where the common difference is required to be a sum of two squares and $k$ is arbitrarily large but fixed. Using the system $(a, d) \mapsto (a, a+d, \cdots, a+(k-1)d, d)$, which is of finite complexity, the theorem reads

$$\sum_{1 \leq a \leq a+(k-1)d \leq N} \Lambda(a) \Lambda(a+d) \cdots \Lambda(a+(k-1)d) R(d) \sim \pi \frac{N^2}{2(k-1)} \prod_p \beta_p + o(N^2) \quad (3)$$
where
\[ \beta_p = \lim_{m \to \infty} \mathbb{E}_{(a,d) \in (\mathbb{Z}/p^m\mathbb{Z})^2} \frac{\rho_d(p^m)}{p^m} \left( \frac{p}{\phi(p)} \right)^k \prod_{i=1}^{k} 1_{(a+(i-1)d,p) = 1}. \]

Notice that the sum to be estimated can also be written as
\[ \sum_{1 \leq a \leq a+(k-1)(b^2+c^2) \leq N} \prod_{i=1}^{k} \Lambda(a + (i-1)(b^2+c^2)). \]

Recall we can easily convert a statement involving \( \Lambda \) to a statement involving the indicator function of primes, yielding the following theorem which we will deduce from Theorem 2.1.

**Theorem 2.2. The size of the set**
\[ \{(a,b,c) \in \mathbb{Z}^3 \mid a \geq 0, \text{ and } a,a+b^2+c^2, \cdots, a+(k-1)(b^2+c^2) \text{ are all primes not larger than } N\} \]

is asymptotic to
\[ \frac{\pi N^2}{2(k-1) \log N} \prod_p \beta_p. \]

Notice here that we obtain an asymptotic for the number of prime values of a polynomial system \((\psi_1, \cdots, \psi_k) \in (\mathbb{Z}[a,b,c]^3 \text{ defined by } \psi_i(a,b,c) = a + (i-1)(b^2+c^2). \)

Notice also that in words, this cardinality is the number of \( k \)-APs in the primes where the common difference is a sum of two squares, and where each such AP is counted as many times as the common difference is represented as a sum of two squares.

**Proof (of Theorem 2.2 assuming Theorem 2.1).** Here we imitate [9], page 6. Let \( \epsilon \) be a small positive parameter, to be determined later. The contribution to (3) of the terms with \( a \leq N^{1-\epsilon} \) is at most
\[ \sum_{d \leq N/2} R(d) \sum_{a \leq N^{1-\epsilon}} \prod_{i \in [t]} \Lambda(a + (i-1)d) \leq N^{1-\epsilon} \log N \sum_{d \leq N/2} R(d) = O(N^{2-\epsilon/2}) \]
and the contribution of these terms to the size of the set (4) is at most \( N^{2-\epsilon} \), which is in particular \( o(N^{2-\epsilon/2}) \).

Now the pairs \((a,d)\) such that one of the \( \psi_i(a,d) \) is a non prime prime power, bring also a negligible contribution to (3); for instance the terms where \( a \) is of the form \( p^m \) with \( m \geq 2 \) bring a contribution which is smaller than
\[ \sum_{d \leq N/2} R(d) \sum_{p \leq \sqrt{N}} \sum_{m \leq \log_p N} \, (\log N)^k = O(N^{3/2} \log^{k+1} N) = o(N^2). \]

Finally for the remaining non-zero contributions, we have
\[ \prod_{i \in [t]} \Lambda(a + (i-1)d) = (1 + O(\epsilon))^k \log_k N. \]
This shows in particular that the size of the set (4) is
\[
(1 + O(\epsilon)) \frac{\pi N^2/(2(k-1))}{(\log N)^k} \prod_p \beta_p + o\left(\frac{N^2}{\log N}\right) + o(N^{2-\epsilon/2})
\]
and choosing \(\epsilon\) to be a sufficiently slowly decaying function of \(N\), the last term will be \(o(N^2/\log N)\) while the \(O(\epsilon)\) will become an \(o(1)\), which completes the proof. \(\blacksquare\)

Let’s fix \(k = 3\). The \(\pi N^2/4\) has the nice property of being the volume of the convex body
\[
L = \{(a, b, c) \in \mathbb{R}^3 \mid 1 \leq a \leq a + 2(b^2 + c^2) \leq N\}
\]
but unfortunately, as seen in Remark 1.1,
\[
\beta_p \neq \mathbb{E}_{(a,b,c)\in(\mathbb{Z}/p\mathbb{Z})^3} \Lambda_p(a)\Lambda_p(a + (b^2 + c^2))\Lambda_p(a + 2(b^2 + c^2)) = \mathbb{E}_{n\in(\mathbb{Z}/p\mathbb{Z})^3} \prod_{i=[k]} \Lambda_p(\psi_i(n))
\]
so the form the local factors take for this polynomial system is not exactly the form they take for linear systems.

Let’s now permute \(\Lambda\) and \(R\) in (3); here the theorem yields
\[
\sum_{1 \leq a \leq a+(k-1)d \leq N} R(a)R(a+d)\cdots R(a+(k-1)d)\Lambda(d+1) = \frac{\pi^k N^2}{2(k-1)} \prod_p \beta_p + o(N^2) \tag{5}
\]
with
\[
\beta_p = \lim_{m \to \infty} \mathbb{E}_{(a,d)\in(\mathbb{Z}/p^m\mathbb{Z})^2} \frac{p}{p-1}\frac{1_{(d+1,p)=1}}{\prod_{i=1}^k \rho_{a+(i-1)(d+1)}(p^m)/p^m}.
\]
Notice that this is basically counting \(k\)-APs in the set of sums of two squares where the common difference is a prime minus one. If the common difference had to be a prime, for \(k \geq 4\), there would hardly be any AP like this, because a prime, except 2, has to be congruent to 1 or -1 modulo 4, while a sum of two squares can be only 0, 1 or 2 modulo 4; that’s why we put \(\Lambda(d+1)\) instead of \(\Lambda(d)\). The value of this sum for \(k = 3\) for instance is intimately related to the number of solutions to the system of equations
\[
\begin{align*}
\begin{cases}
  c - b & = b - a & = p - 1 \\
  a & = \alpha_1^2 + \alpha_2^2 \\
  b & = \beta_1^2 + \beta_2^2 \\
  c & = \gamma_1^2 + \gamma_2^2
\end{cases}
\end{align*}
\]
with \(p\) prime, \(1 \leq a, b, c \leq N\).

Another application is the number of quadruples of the form
\[
(p, p + a^2 + b^2, p + c^2 + d^2, p + a^2 + b^2 + c^2 + d^2)
\]
in the primes up to \(N\). Here the theorem reads
\[
\sum_{(n,h,g)\in K\cap\mathbb{Z}^3} \Lambda(n)\Lambda(n+h)\Lambda(n+g)\Lambda(n+g+h)R(g)R(h) = \beta_\infty \prod_p \beta_p + o(N^3) \tag{6}
\]
where $K = \{(n, g, h) \in \mathbb{R}^3 \mid n, g, h \geq 0, n + g + h \leq N\}$ and $\beta_\infty = \pi^2 N^3/6$ is again the volume of the convex body

$$L = \{(n, a, b, c, d) \in \mathbb{R}^3 \mid 0 \leq n \leq n + a^2 + b^2 + c^2 \leq N\}$$

and

$$\beta_p = \lim_{m \to \infty} \mathbb{E}_{(n, g, h) \in (\mathbb{Z}/p^m\mathbb{Z})^3} \Lambda_p(n) \Lambda_p(n + g) \Lambda_p(n + g + h) \frac{\rho_p(p^m)}{p^m} \frac{\rho_h(p^m)}{p^m}$$

We claim, but we do not formally prove, that a result similar to Theorem 2.1 is possible with the divisor function $\tau$ instead of the representation functions $R_i$. In fact, the result would certainly be easier to prove, since the treatment of the representation function of a binary quadratic form, done in [14] and reused here, involves a reduction to a restricted divisor function. So the aforementioned theorem, inspired from [9] and [13], would be the following.

**Theorem 2.3.** Let $\Psi = (\psi_1, \cdots, \psi_{t+s}) : \mathbb{Z}^d \to \mathbb{Z}^{t+s}$ 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 $\Phi = (\psi_{t+1}, \cdots, \psi_{t+s})$ and $\dot{\Phi}$ for the linear part. Then

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

where

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

and

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

with $\Phi_{a,p} : b \mapsto \Phi(a) + p\Phi(b)$ (so that $\Phi(a + pb) = \Phi_{a,p}(b)$) and $\alpha$ as in Definition A.1.

**Example 2.1.** This theorem provides an asymptotic for the number of triples of non-negative integers $(a, b, c)$ such that $a, a + bc, a + 2bc$ are primes. This is again a quadratic pattern; in fact $\tau$ can be considered as the representation function of the quadratic form $(x, y) \mapsto xy$.

Now we turn to a proof of the main theorem, Theorem 2.1. The philosophy of the proof is the typical philosophy governing the Green-Tao method: to prove that a multilinear average

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

is asymptotically 1, knowing that all functions involved have average 1, we want to show that each $F_i$ can be replaced by the constant function 1 without too much loss. So we want to show that $F_i - 1$ is negligible in any multilinear average. Thanks to a *generalised von Neumann theorem*, we know that it suffices to ensure that $F_i - 1$ has
small *Gowers uniformity norms* (this uniformity condition is already provided in [9] and [14]) and that all the functions $F_i - 1$ are bounded by a common *enveloping sieve* or *pseudorandom majorant*. This is precisely where the novelty is. What is already known is a pseudorandom majorant for $\Lambda$ and another one for $R_f$, and we show that we can combine these two majorants into a common majorant.

This possibility of combining $\Lambda$ with other suitably uniform arithmetic functions could open the door to the determination of the asymptotic frequency of various patterns in the future.

## 3 Proof of Theorem 2.1

We fix some arbitrarily large integer $N$, so that our asymptotic results are valid in the limit where $N$ tends to $\infty$. Many of the parameters introduced in what follows implicitly depend on $N$ (like the convex bodies $K$, the map $p \mapsto \alpha(p)$, the numbers $w, W, \overline{W}$, the set $X_0$...).

### 3.1 Elimination of a negligible set

It will be necessary to start our proof by taking care of a technicality. It turns out that there is a quite skinny subset of the integers $X_0 \subset [N]$, depending on some numeric constants $C_1 > 0$ and $\gamma$, on which the divisor function, and also the representation function, behaves abnormally, so that our process of majorisation by a pseudorandom measure (done in section 5) fails there. So we recall the following definition originating from [13] and taken up in [14].

**Definition 3.1.** Let $\gamma = 2^{-k}$ for some $k \in \mathbb{N}$ 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 either

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 $m^2 \mid n$, which satisfies $m > N^{\gamma}$.

The following lemma, which is Lemma 3.2 from [14], itself a synthesis of Lemmas 3.2 and 3.3 from [13], shows how negligible this set is.

**Lemma 3.1.** For $\Psi$ and $K$ as in Theorem 2.1, 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 altogether \(X_0\). For any PDBQF \(f\), we use to the notation \(\overline{R}_f(n)\) to denote \(1_{n \not\in X_0} R_f(n)\).

**Lemma 3.2.** If the parameter \(C_1\) in Definition 3.1 is large enough, and for any choice of the constant \(\gamma\), Theorem 2.1 holds if and only if, under the same conditions and with the same notations:

\[
\sum_{n \in K \cap \mathbb{Z}^d} \prod_{i=1}^t \Lambda(\psi_i(n)) \prod_{j=t+1}^{t+s} \overline{R}_f(\psi_j(n)) = \beta_\infty \prod_p \beta_p + o(N^d).
\]

**Proof.** We have to show that

\[
\sum_{n \in K \cap \mathbb{Z}^d} \prod_{i=1}^t \Lambda(\psi_i(n)) \prod_{j=t+1}^{t+s} R_f(\psi_j(n)) = o(N^d).
\]

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

\[
\left( \sum_{n \in K \cap \mathbb{Z}^d} \prod_{j=t+1}^{t+s} R_f(\psi_j(n)) \right)^2 \leq \sum_{n \in K \cap \mathbb{Z}^d} \left( \prod_{j=t+1}^{t+s} R_f(\psi_j(n)) \right) \sum_{n \in K \cap \mathbb{Z}^d} \sum_{j=t+1}^{t+s} 1_{\psi_j(n) \in X_0}
\]

and then we use Lemma 3.1 of [14] which ensures that the first factor is \(N^d \log^{O(1)} N\) while the second is \(N^d \log^{-C_1/2} N\) according to Lemma 3.1, so that taking \(C_1\) bigger than \(2(t + O_s(1))\), we have the result. ■

**Remark 3.1.** From now on, we will drop the bar, thus \(R_f\) coincides with the actual representation function of \(f\) on \([N] \setminus X_0\) and is 0 on \(X_0\).

### 3.2 Implementation of the W-trick

The idea of the W-trick, which is by now fairly standard (see [8], [9] for its implementations by Green and Tao, see also [13] and [14], which we are going to follow more closely), is to eliminate the obvious bias of the actual primes, like the strong preference for odd numbers, to produce a more uniform set. The representation function of a PDBQF is also biased (it does not have the same average on every residue class), so this has to be corrected. To do this we introduce, for some slowly growing function of \(N\) like \(w(N) = \log \log \log N\), the numbers

\[
W = \prod_{p \leq w} p \quad \text{and} \quad \overline{W} = \prod_{p \leq w} p^{\alpha(p)},
\]

where \(\alpha(p)\) appears in Matthiesen’s paper [14] defined by

\[
p^{\alpha(p)-1} < \log^{C_1+1} N \leq p^{\alpha(p)}
\]
for some \( C_1 \) large enough as in Lemma 3.2.

Green and Tao did not need prime powers in their \( W \), but in the case of a representation function of a PDBQF, it turns out to be necessary. This is of course linked to the Remark 1.1. Notice that for \( N \) big enough, for \( p \leq w(N) = \log \log \log N \), we have always \( \alpha(p) \geq 1 \). We also introduce for \( b \in [\overline{W}] \), the function \( \Lambda'_{b,\overline{W}} \) defined by

\[
\Lambda'_{b,\overline{W}}(n) = \frac{\phi(W)}{W} \Lambda'(\overline{W}n + b),
\]

where \( \Lambda' = 1_p \log \) (we recall that \( \phi(W)/W = \phi(\overline{W})/\overline{W} \)). The advantage of this function, which of course also works with \( W \) instead of \( \overline{W} \), is that it has average 1, and even average 1 in any arithmetic progression of \( w(N) \)-smooth modulus. Indeed, let \( q_1 \) be an integer whose prime factors are all at most \( w(N) \) (what we call \( w(N) \)-smooth or friable), and let \( q_0 \in [q_1] \) and \( P = \{q_0 + mq_0 | m \leq M\} \) and \( b \in [\overline{W}] \) coprime to \( W \). Then

\[
\mathbb{E}_{n \in P} \Lambda'_{b,\overline{W}}(n) = \frac{1}{M} \sum_{m \leq M} \frac{\phi(W)}{W} \Lambda'(\overline{W}q_1m + \overline{W}q_0 + b) \sim \frac{\phi(W)}{W} \frac{\overline{W}q_1}{\phi(\overline{W}q_1)} = 1
\]

by the prime number theorem in APs and the fact that \( \phi(q)/q \) only depends on the radical of \( q \).

We must derive a function with this property from \( R_f \) as well, for any PDBQF \( f \), of discriminant \( D \). Here, following Matthiesen in [14], we define

\[
r'_{f,b}(m) := \frac{\sqrt{-D}}{\pi} \frac{\overline{W}}{\rho_{f,b}(\overline{W})} R_f(\overline{W}m + b),
\]

where we set

\[
\rho_{f,a}(q) = \left| \{(x, y) \in [q]^2 | a = f(x, y) \mod q\} \right|
\]

for any \( b \) such that \( \rho_{f,b}(\overline{W}) > 0^1 \). Recall Remark 3.1, namely that \( R_f(n) \) equals 0 in the case where \( n \in X_0 \), in particular in the case where \( n \equiv 0 \mod p^{\alpha(p)} \) with \( p \leq w(N) \). Hence, \( r'_{f,b} = 0 \) if \( b \equiv 0 \mod p^{\alpha(p)} \). This function will have the desired property. In fact (see [14], Definition 7.2) for \( b \neq 0 \mod p^{\alpha(p)} \) for any \( p \leq w(N) \) satisfying \( \rho_{f,b}(\overline{W}) = 0 \), we have

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

We now decompose the sum appearing in 3.2 into sums over congruence classes. Schematically, we want to write

\[
\sum_{n \in \mathbb{Z}^d \cap K} = \sum_{a \in [\overline{W}]^d} \sum_{n \in K_a},
\]

where

\[
K_a = \{x \in \mathbb{Z}^d | \overline{W}x + a \in K\}
\]

is again a convex body. Moreover, for \( j \in [1, t + s] \), \( \psi_j(\overline{W}n + a) \) can be written as \( \overline{W} \psi_j(n) + c_j(a) \) with \( c_j(a) \in [\overline{W}] \) and \( \psi_j \) differs from \( \psi_j \) only in the constant term. We remark that if \( \psi_j(a) \) is not coprime to \( W \) for \( i \in [t] \) or if \( \rho_{f_j,\psi_j(a)}(\overline{W}) = 0 \) or \( \psi_j(a) \equiv 0 \mod p^{\alpha(p)} \) for some \( j \in [t + 1, t + s] \) and \( p \leq w(N) \) a prime, then \( a \) make a zero contribution. So the residues \( a \) which make make a non-zero contribution are all sent by \( \Psi \) on tuples \((b_1, \cdots, b_{t+s})\) belonging to the following set.

\(^1\)Even if \( \rho_{f,b}(\overline{W}) = 0 \), we will use the notation \( r'_{f,b,\overline{W}}(m) \) with the understanding that it means 0.
Definition 3.2. We denote by \( B = B_{t,s} \) (the subscripts being usually dropped when no ambiguity is possible) the set of residues \( b \in \mathbb{W}^{t+s} \) such that

1. for any \( i \in [t] \), \( (b_i, W) = 1 \)
2. for any \( j \in [t+1, t+s] \) and any prime \( p \leq w(N) \), we have \( b_j \neq 0 \mod p^{\alpha(p)} \)
3. for any \( j \in [t+1, t+s] \), \( b_j \) is representable by \( f_j \) modulo \( W \), that is \( \rho_{f_j,b_j(W)} > 0 \)

Moreover, for an affine-linear system \( \Psi : \mathbb{Z}^d \to \mathbb{Z}^{t+s} \), we define \( A = A_\Psi \) (the subscript will again be often dropped) to be the set of all \( a \in [W]^d \) such that \( (c_i(a))_{i \in [t+s]} \in B \). We recall that \( c_i(a) \) is the reduction modulo \( W \) in \( [W] \) of \( \psi_i(a) \); we will also denote by \( c(a) \) the vector \( (c_i(a))_{i \in [t+s]} \). As seen above, any \( a \in [W]^d \setminus A \) brings no contribution.

3.3 Reduction of the main theorem

We are going to derive the main theorem (Theorem 2.1) from the following.

Theorem 3.3. For \( N' = N/W \), for any system \( \Phi = (\phi_1, \ldots, \phi_{t+s}) \) of affine-linear forms \( \mathbb{Z}^d \to \mathbb{Z}^{t+s} \) of finite complexity whose linear coefficients are bounded, for any convex set \( L \subset [0,N]^d \) such that \( \Phi(L) \subset [1,N']^{t+s} \) and for any \( b \in B \) the following asymptotic holds

\[
\sum_{n \in \mathbb{Z}^d \cap L} \prod_{i \in [t]} \Lambda'_b,\mathbb{W}(\phi_i(n)) \prod_{i=t+1}^{t+s} r'_{f_i,b_i}(\phi_i(n)) = \text{Vol}(L) + o(N^d)
\]

We will use this theorem with a system \( \Phi \) differing from \( \Psi \) only in the constant part, \( L = K_a \) and \( b = c(a) \) for some \( a \in A \). Notice that \( \text{Vol}(K_a) = \text{Vol}(K)W^{-d} \) for any \( a \) by dilation, so it is expected to be of the order of magnitude of \( N^d \) as \( \text{Vol}(K) \) is expected to be of the order of magnitude of \( N^d \).

Proof (Theorem 3.3 implies Theorem 2.1). We proceed to the already sketched decomposition of the sum.

\[
\sum_{n \in K \cap \mathbb{Z}^d} \prod_{i=1}^{t} \Lambda'(\psi_i(n)) \prod_{j=t+1}^{t+s} R_{f_j}(\psi_j(n))
\]

\[
= \sum_{a \in [W]^d \cap \mathbb{Z}^d} \sum_{n \in \mathbb{Z}^d \cap K} \prod_{i=1}^{t} \Lambda'(\psi_i(Wn + a)) \prod_{j=t+1}^{t+s} R_{f_j}(\psi_j(Wn + a)).
\]

But if \( a \) does not lie in \( A \), its contribution is 0, so the sum can be restricted to \( a \in A \); the last sum becomes

\[
\sum_{a \in A} \left( \frac{W}{\phi(W)} \right)^t \prod_{j=t+1}^{t+s} \frac{2\pi}{\sqrt{-D_j}} \frac{\rho_{f_j,\psi_j(a)}(W)}{W} \sum_{n \in K_a \cap \mathbb{Z}^d} \prod_{i=1}^{t} \Lambda'_{c_i(a),W}(\psi_i(n)) \prod_{j=t+1}^{t+s} r'_{f_j,c_j(a)}(\tilde{\psi}_j(n)).
\]

(7)
The inner sum is now tractable by Theorem 3.3, and its value is
\[
\sum_{n \in K_a \cap \mathbb{Z}^d} \prod_{i=1}^t \Lambda'_{c_i(a)}(\tilde{\psi}_i(n)) \prod_{j=t+1}^{t+s} r'_{f_j,c_j(a)}(\tilde{\psi}_j(n)) = \text{Vol}(K_a) + o(N^{td}) = \text{Vol}(K)W^{-d} + o(N^{td}).
\]

We inject this estimate into (7), which then becomes
\[
\sum_{a \in A} \left( \frac{W}{\phi(W)} \right)^t \prod_{j=t+1}^{t+s} \frac{2\pi}{\sqrt{-D_j}} \frac{\rho_{f_j,\psi_j(a)}(W)}{W} \left( \text{Vol}(K)W^{-d} + o(N^{td}) \right).
\]

This can be rewritten as
\[
(\text{Vol}(K) + o(N^{td})) \mathbb{E}_{a \in [W]^d} \left( \frac{W}{\phi(W)} \right)^t \prod_{j=t+1}^{t+s} \frac{2\pi}{\sqrt{-D_j}} \frac{\rho_{f_j,\psi_j(a)}(W)}{W}.
\]

This expectation over \( a \in [W]^d \) is in fact an expectation over \( a \in (\mathbb{Z}/W\mathbb{Z})^d \). By the Chinese remainder theorem, we can then write the expectation as a product over primes, so that
\[
\mathbb{E}_{a \in [W]^d} \left( \frac{W}{\phi(W)} \right)^t \prod_{j=t+1}^{t+s} \frac{2\pi}{\sqrt{-D_j}} \frac{\rho_{f_j,\psi_j(a)}(W)}{W} = \prod_{p \leq w(N)} \mathbb{E}_{a \in (\mathbb{Z}/p^{\alpha(p)})^d} \left( \frac{W}{\phi(W)} \right)^t \prod_{i=1}^t \Lambda_p(\tilde{\psi}_i(a)) \prod_{j=t+1}^{t+s} \frac{\rho_{f_j,\psi_j(a)}(W)}{W},
\]

Now we invoke our results on the local factors from Appendix A, first Lemma A.1, which imply that the last term in (9) is \( \prod_{p \leq w(N)} \beta_p + O(\log^{-C_1/5} N) \) which is then, according to Lemma A.2, equal to \( \prod_p \beta_p + o(1) \). This finally means that (8) is \( \beta_{\infty} \prod_p \beta_p + o(N^{td}) \), which is exactly the claim of Theorem 2.1.

We can make one more minor reduction of the theorem. We use the following simple volume packing lemma from [9], Appendix A.

**Lemma 3.4.** Let \( K \subset [0, N]^d \) be a convex body of \( \mathbb{R}^d \). Then
\[
|K \cap \mathbb{Z}^d| = \sum_{n \in K \cap \mathbb{Z}^d} 1 = \text{Vol}(K) + O_d(N^{d-1}).
\]

Thus, the result of Theorem 3.3 can be written as
\[
\sum_{n \in L \cap \mathbb{Z}^d} \left( \prod_{i=1}^t \Lambda'_{c_i(a)}(\tilde{\psi}_i(n)) \prod_{j=t+1}^{t+s} r'_{f_j,c_j(a)}(\tilde{\psi}_j(n)) - 1 \right) = o(N^{td}).
\]

And then writing each factor \( F_i \) as \( (F_i - 1) + 1 \), we find that the following theorem would imply Theorem 2.1.
Theorem 3.5. For any integers $d, t$ and $s$, and any $s$-tuple of PDBQF $f_{t+1}, \ldots, f_{t+s}$, for $N' = N/W$, for any system $\Phi = (\phi_1, \ldots, \phi_{t+s})$ of affine-linear forms $\mathbb{Z}^d \to \mathbb{Z}^{t+s}$ of finite complexity, the coefficients of the linear part being bounded, for any convex set $K \in [N]^d$ such that $\Phi(K) \subset [N']^{t+s}$ and for any $b \in B$ the following asymptotic holds
\[
\sum_{n \in \mathbb{Z}^d \cap K} \prod_{i \in [t]} (\Lambda'_{b_i W}(\phi_i(n)) - 1) \prod_{j=t+1}^{t+s} (r'_{f_j b_j}(\phi_j(n)) - 1) = o(N'^d).
\]

We prove this theorem in the next section.

4 Proof of Theorem 3.5

4.1 Generalized von Neumann theorem and uniformity

To prove Theorem 3.5, we have to show that the average along a linear system of a product is $o(N'^d)$, knowing that each factor has average $o(N'^d)$. To do so, we reduce to a family of standard linear systems, the ones which underlie the definition of Gowers norms which we now introduce.

Definition 4.1. Let $g : \mathbb{Z} \to \mathbb{R}$ be a function and $k \geq 1$ an integer. The Gowers $U^k$ norm of $g$ on $[N]$ is the expression
\[
\|g\|_{U^k[N]} = \left( E_{x \in [N]} E_{h \in [N]^k} \prod_{\omega \in \{0,1\}^k} g(x + \omega \cdot h) \right)^{2^{-k}}.
\]

In order to prove that the multilinear average of Theorem 3.5 is small, the idea is to admit, for any integer $k$, that both $\|\Lambda'_{b W} - 1\|_{U^k[N]} = o(1)$ (for any $b$ coprime to $W$) and $\|r'_{f b} - 1\|_{U^k[N]} = o(1)$ for any PDBQF $f$ and $b$ representable by $f$ modulo $W$ and not congruent to 0 modulo $p^{\omega(p)}$ for any $p \leq w(N)$ (these are the results of [9] and of [14], respectively). Then we need to ensure that these $W$-tricked functions are also bounded by a pseudorandom measure; this was again done in the same references. Finally we invoke the generalized von Neumann theorem which shows that the conditions on the Gowers norms and the domination by a pseudorandom measure guarantee that any finite complexity multilinear average is arbitrarily small. So we first need to define the notion of pseudorandom measure.

Definition 4.2. A pseudorandom measure is a function $\nu$ or rather a sequence of functions $\nu_M : \mathbb{Z}/M\mathbb{Z} \to \mathbb{R}_+$ (but the dependence on $M$ will not appear explicitly later on) satisfying
1. $E_{n \leq M} \nu(n) = 1 + o(1)$.
2. $(D$-linear forms conditions) 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 $\Psi(K) \subset [M]^t$, the following estimate holds
\[
E_{n \in \mathbb{Z}^d \cap K} \prod_{i \in [t]} \nu(\psi_i(n)) = 1 + o(1).
\]
Remark 4.1. Pseudorandom measures are defined on cyclic groups rather than intervals of integers, so the values of the linear forms $\psi_i(n)$ are understood modulo $M$. On the other hand, our functions to majorise, of the form $\Lambda'_{f,b} - 1$ and $r'_{f,b} - 1$, are naturally defined on the interval $[N']$. It involves some technicalities to pass from intervals to cyclic groups (to avoid torsion and wrap-around issues, we try to embed an interval into a sufficiently big cyclic group of prime order).

Remark 4.2. There used to be a correlation condition, but it is not necessary any more since the work [2] of Fox, Conlon and Zhao, and its integration by Tao and Ziegler in [16]. It used to crop up in the derivation of the generalized Gowers inverse theorem (i.e. applied to unbounded functions) from the standard Gowers inverse theorem.

Armed with this definition, and following [9] (Proposition 7.1), we state a suitable generalized von Neumann theorem.

Theorem 4.1. Let $t, d, L$ be positive integer parameters. Then there are positive constants $1 \leq C_1$ and $D$, depending on $t, d$ and $L$ such that the following holds. Let $C_1 \leq C \leq O_{t,d,L}(1)$ be arbitrary and suppose that $M \in [CN', 2CN']$ is a prime. Let $\nu : \mathbb{Z}/MZ \to \mathbb{R}^+$ be a $D$-pseudorandom measure, and suppose that $f_1, \cdots, f_t : [N'] \to \mathbb{R}$ are functions with $|f_i(x)| \leq \nu(x)$ for all $i \in [t]$ and $x \in [N']$. Suppose that $\Psi = (\psi_1, \cdots, \psi_t)$ is system of affine-linear forms of finite complexity whose linear coefficients are bounded by $L$. Let $K \subset [-M,M]^d$ be a convex set such that $\Psi(K) \subset [M]^t$. Finally, suppose that

$$\min_{1 \leq j \leq t} \|f_j\|_{U^t[N]} = o(1).$$

Then we have

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

The constant $C_1$ is not to be mistaken with the one in Definition 3.1! The quite distinct contexts should discard any ambiguity.

We highlight that this theorem actually replaces a linear system $\Psi$ with another one, the system of the $(x + \omega \cdot h)_{\omega \in \{0,1\}^{t-1}}$, so that it is not immediately obvious that we have reduced the difficulty. However, it happens that uniformity with respect to this system can be characterised in another way: this is the Gowers inverse theorem (see for instance [11]).

Remark 4.3. Here we demand for at least one of the functions a very high uniformity, indeed a uniformity of order $t - 1$ where $t$ is the number of linear forms. We sometimes do not need this much uniformity; the notion of complexity of a system aims at providing the degree of uniformity actually needed. In this paper, we do not go into these subtleties.

So now we must provide the uniformity condition (11) for our functions.

Proposition 4.2. 1. Let $b \in [W]$ coprime to $W$. Then the $\overline{W}$-tricked von Mangoldt function $\Lambda'_{b,\overline{W}} : n \mapsto \Lambda(\overline{W}n + b)$ satisfies:

$$\forall k \in \mathbb{N}, \left\| \Lambda'_{b,\overline{W}} - 1 \right\|_{U^k[N']} = o(1).$$
2. Let \( f \) be a PDBQF of discriminant \( D < 0 \), and \( b \in \overline{W} \) be representable by \( f \) modulo \( W \) and not divisible by any \( p^{\alpha(p)} \) for \( p \leq w(N) \). Then the tricked representation function of \( f \), defined by

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

satisfies

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

Both points rely on Gowers inverse theorem, and hence their proofs, done in [9] and [14], consist in evaluating the correlation of \( \Lambda'_{b,W} - 1 \) or \( r'_{f,b} - 1 \) with nilsequences.

There’s a slight difference between the first point of the above proposition and the original statement of Green and Tao: they proved it with \( W \) instead of \( \overline{W} \). However, it is easy to check in the proof of \( \|\Lambda'_{b,W} - 1\|_{U^k[N']} = o(1) \) in [9] (section 12, and appendix D) that the role of the \( W \)-trick is exclusively to bound the exceptional primes by \( w(N) \), which is done equally well by tricking with \( W \) or with \( \overline{W} \).

### 4.2 Construction of a pseudorandom majorant

We provide here a pseudorandom measure which dominates both the \( W \)-tricked von Mangoldt and representation functions. We recall for each an already well-studied pseudorandom majorant, and later try to combine them.

#### 4.2.1 The pseudorandom majorant of the von Mangoldt function.

We take a pseudorandom majorant from the Green-Tao machinery. Let us introduce

\[
\Lambda_{\chi,R}(n) = \log R \left( \sum_{l \mid n} \mu(l) \chi \left( \frac{\log l}{\log R} \right) \right)^2
\]

where \( R = N^\gamma \) and \( \gamma \) will be chosen later, and \( \chi \) is a smooth even function \( \mathbb{R} \to [0,1] \) supported on \([-1,1]\) satisfying \( \chi(0) = 1 \) and \( \int_0^1 \chi'(x)^2 dx = 1 \). Finally, for any \( b \in \overline{W} \) coprime to \( \overline{W} \), we define \( \nu_{GT,b} : n \mapsto 2^\Theta(W) \Lambda_{\chi,R}(\overline{W}n + b) \). We recall that this function majorises the \( W \)-tricked von Mangoldt function and that it has average 1 thanks to the following lemma, reproducing results from [9].

**Lemma 4.3.** 1. For any \( b \in \overline{W} \) coprime to \( W \), we have

\[
\Lambda'_{b,W}(n) \lesssim \frac{\phi(W)}{W} \Lambda_{\chi,R}(\overline{W}n + b)
\]

for \( n \in [R,N'] \), where the implied constant depends only on \( \gamma \).

2. If \( \gamma \) is small enough, for any \( b \in \overline{W} \) coprime to \( \overline{W} \), we have

\[
\mathbb{E}_{n \in [N']} \nu_{GT,b}(n) = \log R \frac{\phi(W)}{W} \sum_{l, l'} \mu(l) \mu(l') \chi \left( \frac{\log l}{\log R} \right) \chi \left( \frac{\log l'}{\log R} \right) \mathbb{E}_{n \in [N']} 1_{l \mid \overline{W}n + b} 1_{l' \mid \overline{W}n + b}
\]

\[= 1 + o(1).\]
Proof. 1. We have to do something only for \( n \in [R, N'] \) such that \( Wn + b \) is prime. In this case the left hand side is bounded above by a constant multiple of \( \frac{\phi(W)}{W} \log N \) while the right-hand side is \( \phi(W)W \log R \), but \( \log R = \gamma \log N \gg \log N \).

2. The first equality is obvious. The second one is a very special case of Theorem D.3 from [9], where in fact a result similar to Theorem 1.1 is proven with \( \Lambda_{\chi,R} \) instead of \( \Lambda \).

\[\quad\]

Remark 4.4 (important). From now on, \( \gamma \) is supposed to be less than 2/5 and \( \Lambda \) is defined to be 0 on \( [N^{3/5}] \), thus for \( N \) big enough, for any \( b \in [W] \) and \( n \in [N^{2/5}] \), we have \( \Lambda_{\chi,R}(n) = 0 \). Notice that this change bears no impact at all on the main theorems, because the contribution of the integers smaller \( N^{3/5} \) fit in the error terms \( o(N^d) \), as shows the reasoning done in the proof of Theorem 2.2. Thanks to this convention the majoration \( \Lambda_{\chi,R} \ll \nu \) now holds on all of \( [N'] \).

4.2.2 The pseudorandom majorant of the representation function

We take also a pseudorandom majorant from Matthiesen’s work. For this we need to recall a few facts from [14].

Proposition 4.4. For an integer \( D \), there exists a set of primes \( P_D \) of density 1/2, which is a union of congruence classes modulo \( D \), such that putting \( P_D^* = P_D \cup \{ p \in P : p \mid D \} \) and \( Q_D = P \setminus P_D^* \) we have for any \( PDBQF f \) the bound

\[ R_f(n) \ll D \tau_D(n) \sum_{m \in (Q_D)} 1_{n/m \in (P_D^*)}, \]

where for a set of primes \( A \), \( \langle A \rangle \) stands for the set of integers whose prime factors are all in \( A \) and \( \tau_D(n) = \sum_{d \in (P_D)} 1_{d \mid n} \).

To heuristically understand this result, which is in [14] the starting point of the construction of the pseudorandom majorant, we recall that the number of representation of any number \( n \) as a sum of two squares is \( 4 \sum_{d \mid n} \chi(d) \) where \( \chi \) is the only non-trivial character modulo 4. By multiplicativity, this is easily seen to become \( \tau_4(n)1_{n \equiv 3 \bmod{4}} \), from which we get easily a majorant of the given form, which must be extended to other quadratic forms.

Thus to majorize the function \( R_f \), it will be enough to majorize the functions \( \tau_D \) and \( 1_{P_D^*} \). The heuristic to bound \( \tau_D \) (or rather \( \tau_D/\sqrt{\log N} \)) is as follows (see Lemma 4.1 of [13]). We would like to truncate the divisor sum defining it at \( N^\gamma \) (possibly with a smooth cut-off), just as done earlier for the von Mangoldt function. However it turns out that the majoration \( \tau \leq C\tau_\gamma \) is not entirely true, at least not true with the same constant \( C \) throughout the \( N \) first integers. However, a heuristic of Erdős says that either an integer is excessively rough or excessively smooth or has a cluster of many prime factors close together. Moreover we have excluded the two first possibilities when we took out the set \( X_0 \), so the majoration has to be done only in the third case. But the majoration depends on the position of this cluster of primes and on its density.
To bound 1\(_{(\mathbb{P}_D)}\) (or rather 1\(_{(\mathbb{P}_D)}\)\(\sqrt{\log N}\)), that is the indicator of the numbers without any \(\mathbb{Q}_D\) prime factor, we use a sieving type majorant, that is a majorant similar to the one introduced above for the von Mangoldt function. Indeed, numbers without any \(\mathbb{Q}_D\) prime factor are similar to prime numbers (numbers without any prime factor at all).

To formalise this heuristic, let us introduce the following definitions.

**Definition 4.3.** Let \(\xi = 2^{-m}\) for some \(m \in \mathbb{N}\), put \(\gamma = 2\xi\). We define sets \(U(i, s)\) for integers \(i, s\) as follows. For \(i = \log_2(\frac{2}{\xi}) - 2 = m - 1\) we let \(U(i, 2/\xi) = \{1\}\) and otherwise \(U(i, 2/\xi) = \emptyset\). If \(s > 2/\xi\) write \(U(i, s)\) for the set of all products of \(m_0\) distinct primes from the interval \([N/2 - i - 1, N/2 - i]\).

Using the idea from Proposition 4.4, but without much more motivation (motivation shall be found in [14] and [13]), let us introduce a majorant for the W-tricked representation function. We need again the smooth function \(\chi\) (this should not be mistaken with a character, as there are no more characters in what follows) introduced for the majorant of the von Mangoldt function. We use the function

\[
r_{D,\gamma}(n) = \frac{\beta'_{D,\gamma}(n)\nu'_{D,\gamma}(n)}{C_{D,\gamma}},
\]

where

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

with

\[
\tau'_{D,\gamma}(n) = \sum_{d \in \mathbb{P}_D} 1_{d | n} \chi\left( \frac{\log d}{\log N\gamma} \right),
\]

and

\[
\beta'_\gamma(n) = \sum_{m_0 \in \mathbb{Q}_D} \left( \sum_{e \in \mathbb{Q}_D} 1_{m_0 e | n} \mu(e) \chi\left( \frac{\log e}{\log N\gamma} \right) \right)^2.
\]

The constant \(C_{D,\gamma}\) is the one which ensures that the function \(r_{D,\gamma}\) has average 1; such a constant does indeed exist, as was proved in Lemma 7.5 from [14]. Moreover, \(\nu_{\text{Matt},b}\) indeed majorises the \(\overline{W}\)-tricked function: this is again Lemma 7.5 from [14], which we recall below.

**Lemma 4.5.** Let \(\nu_{\text{Matt},b}(n) = r_{D,\gamma}(\overline{W} n + b)\).

1. For any PDBQF \(f\) of discriminant \(D\) and \(b \in B_{0,1}\) (i.e. \(b \in [\overline{W}]\) with \(b \neq 0 \mod p^{r(p)}\) for any \(p \leq w(N)\) and satisfying \(\rho_{f,b}(\overline{W}) > 0\)) the following bound holds

\[
r'_{f,b}(n) \ll \nu_{\text{Matt},b}(n).
\]

2. For some \(C_{D,\gamma} = O(1)\), we have \(\mathbb{E}_{n \in [N]} \nu_{\text{Matt},b}(n) = 1 + o(1)\).
4.2.3 Combination of both majorants

To be able to use the von Neumann theorem (Theorem 4.1), and thus establish Theorem 3.5, we need to bound all \( t + s \) functions by the same majorant. Now all of them are bounded by some pseudorandom majorants defined above, so we define our common majorant by averaging all these majorants. Recall that \( N' = N/W \); we take \( M \) to be a prime satisfying \( N' < M \leq O(N') \). Given a family \( (b_1, \cdots, b_{t+s}) \in B \), we define \( \nu^* \) as follows

\[
\nu^*(n) = \begin{cases} 
\frac{1}{t+s+1}(1 + \frac{\phi(W)}{W} \sum_{i=1}^{t} \nu_{GT, b_i}(n) + \sum_{j=t+1}^{t+s} \nu_{Matt, b_j}(n)) & \text{if } n \leq N' \\
1 & \text{if } N' < n \leq M
\end{cases}
\]

We remark that this function satisfies

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

and has average 1 thanks to Lemmas 4.3 and 4.5. So to ensure that \( \nu^* \) is a pseudorandom measure, there remains only to prove the linear forms condition (10). To this aim, we first state the following proposition.

**Proposition 4.6.** Let \( 1 \leq d, t, s \leq D \) where \( D \) is the constant appearing in 4.1 and suppose \( t \leq D \). For any finite complexity system \( \Psi : \mathbb{Z}^d \to \mathbb{Z}^{t+s} \) with coefficients bounded by \( D \) and every convex \( K \subset [0, N]^d \) such that \( \Psi(K) \subset [1, N/W]^t \), and any \( b \in B \), the estimate

\[
\mathbb{E}_{n \in \mathbb{Z}^d \cap K} \prod_{j=t+1}^{t+s} r_{D_j, \gamma}(W\psi_j(n) + b_j) \prod_{i \in [t]} \frac{\phi(W)}{W^{\Lambda_{x_i, R}(W\psi_i(n) + b_i)}} = 1 + O_D \left( \frac{N^{d-1+O_D(\gamma)}}{\text{Vol}(K)} \right) + o_D(1) \tag{12}
\]

holds, provided \( \gamma \) was small enough.

The proof is postponed to Appendix B. We can now state the important theorem of this section.

**Proposition 4.7.** 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, \( b \in B \) and \( f_{t+1}, \cdots, f_{t+s} \) are PDBQF of discriminants \( D_{t+1}, \cdots, D_{t+s} \) and \( \nu^* \) is defined as above, then \( \nu^* \) satisfies the \( D \)-linear condition and for any \( i \in [t] \) we have

\[
\left| \Lambda'_{W, b_i} - 1 \right| \leq \Lambda'_{W, b_i} + 1 \ll \nu^*
\]

and for any \( j \in \{t+1, \cdots, t+s\} \) we have

\[
\left| r'_{f_j, b_j} - 1 \right| \ll \nu^*
\]

where both inequalities are valid on \([N']\).
The majorisation has already been observed earlier. Deriving the linear forms conditions for $\nu^*$ from the above Proposition 4.6 requires some extra work, because of the piecewise definition of $\nu^*$. This work was done in [8], Proposition 9.8 for instance. See also [2], Proposition 8.4, where the same “localisation argument” is employed. Matthiesen also relies on it in [14]. It has a geometric flavour and does not need any modification, so we do not reproduce it here and invite the reader to consult one of the references. We can now prove Theorem 3.5.

Proof (of Theorem 3.5 assuming Proposition 4.7). Take any integers $d, t$ and $s$, and a system $\Phi : \mathbb{Z}^d \to \mathbb{Z}^{t+s}$ of affine-linear forms of finite complexity, where the coefficients of the linear part are bounded by $L$ and $f_{t+1}, \ldots, f_{t+s}$ any PDBQF. Take a convex set $K \subset [1, N']^d$ such that $\Phi(K) \subset [N']^{t+s}$. Let $b \in B$. Then Proposition 4.1 and Proposition 4.7 provide constants $C_0$ and $C_1$, of which we take the maximum $C = \max(C_0, C_1)$. Now take a prime $M \in [CN', 2CN']$. Such a prime exists by Bertrand’s postulate. Define $\nu^*$ as above. Put $f_i = \Lambda'_{b_i, \overline{W}} - 1$ for $i \in [t]$ and $f_j = r'_{f_j, b_j} - 1$ for $j \in \{t + 1, \ldots, t + s\}$. Then we have that $|f_j| \ll \nu^*$ for all $j$ and $\nu^*$ is a pseudorandom measure by Proposition 4.7, so that we can invoke the von Neumann theorem (Theorem 4.1), which, together with the statements of Proposition 4.2 (specialized to $k = t + s - 1$), implies Theorem 3.5. ■

Remark 4.5. We remark that although we want to prove a result concerning quadratic, and not linear, patterns in the primes, we do not need the polynomial forms condition introduced in [17]. This is because the “quadracticity” of our configurations is hidden in a linear input into the representation functions of quadratic forms.

Remark 4.6. Our program to combine two pseudorandom majorants is not really unheard of. In fact, Green-Tao in [9] had to combine the majorants $\Lambda'_{x, R}(\overline{Wn} + b_j)$ for various $b_j$ and so did Matthiesen [14]. Notice also that Wolf and Le [12] devised a certain condition of compatibility for two pseudorandom majorants. However in our case, the majorants have deeply different origins. But they have a similar structure, the structure of truncated divisor sum, so that the proof of the linear forms condition will not be much harder than the ones in [9] or [14].

We have completed the proof our main theorem, conditionally on the following quite technical appendices. Appendix A aims at filling the gap in the proof that Theorem 3.3 implies Theorem 2.1. Appendix B aims at verifying the linear forms condition for the majorant introduced above, that is, at proving Proposition 4.7. Appendix C provides elementary justifications to some statements made in Appendices A and B.

A Analysis of the local factors $\beta_p$

In this section, our aim is to provide the tools to recognise in (9) the product of local factors $\beta_p$. First, we ought to introduce a convenient notation present in both [9] and [14].

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 densities by

$$\alpha_{\psi}(d_1, \ldots, d_t) = \mathbb{E}_{n \in \mathbb{Z}/m\mathbb{Z}^d} \prod_{i=1}^t 1_{\psi_i(n) \equiv 0 \mod d_i}.$$
Remark A.1. It can be thought of as the density of zeros of the system modulo the parameters $d_i$ per unit volume.

From now on we fix such a system $\Psi$, and we suppose its coefficients are bounded by $L$. This will allow the error terms in the following calculations to be uniformly bounded (i.e. the bounds will depend on $L$ but not on the system). We copy and adapt two lemmas of Matthiesen.

Lemma A.1. Let $p$ be a prime. Then

$$\mathbb{E}_{a \in (\mathbb{Z}/p^n\mathbb{Z})^d} \prod_{i=1}^{t} \Lambda_p(\psi_i(a)) \prod_{j=t+1}^{t+s} \frac{\rho_{f_j,\psi_j(a)}(p^{\alpha(p)})}{p^{\alpha(p)}} 1_{\psi_j(a) \not\equiv 0 \mod p^{\alpha(p)}} = \beta_p + O(\log^{-C_1/5} N).$$

Proof. Let $m > \alpha(p)$. We split the sum $\mathbb{E}_{a \in (\mathbb{Z}/p^n\mathbb{Z})^d}$ defining $\beta_p$ into two parts.

1. A sum over $a$ such that for all $j \in [t+1, \ldots, t+s]$ we have $\psi_j(a) \not\equiv 0 \mod p^{\alpha(p)}$. Notice that this property depends only on the projection on $\mathbb{Z}/p^n\mathbb{Z}$, just as $\Lambda_p(\psi_i(a))$, and that the lift-invariance property (Corollary 6.4 of [14]) ensures $\rho_{f_j,\psi_j(a)}(p^m)p^{-m}$ depends only on this projection of $a$ as well, so that

$$\mathbb{E}_{a \in (\mathbb{Z}/p^n\mathbb{Z})^d} \prod_{i=1}^{t} \Lambda_p(\psi_i(a)) \prod_{j=t+1}^{t+s} \frac{\rho_{f_j,\psi_j(a)}(p^m)}{p^{\alpha(p)}} 1_{\psi_j(a) \not\equiv 0 \mod p^{\alpha(p)}}$$

$$= \mathbb{E}_{a \in (\mathbb{Z}/p^n\mathbb{Z})^d} \prod_{i=1}^{t} \Lambda_p(\psi_i(a)) \prod_{j=t+1}^{t+s} \frac{\rho_{f_j,\psi_j(a)}(p^{\alpha(p)})}{p^{\alpha(p)}} 1_{\psi_j(a) \not\equiv 0 \mod p^{\alpha(p)}}.$$

2. And a sum over $a$ such that there exists $j \in [t+1, \ldots, t+s]$ satisfying $\psi_j(a) \equiv 0 \mod p^{\alpha(p)}$. To bound the contribution of such an $a$, we use the following general bound from [14] (see Lemma 6.3 and proof of lemma 8.2)

$$\frac{\rho_{f_j,\psi_j(a)}(p^{\alpha(p)})}{p^{\alpha(p)}} \ll \sum_{k=0}^{m} 1_{\psi_j(a) \equiv 0 \mod p^k}$$

so for any $a$, the contribution satisfies

$$\prod_{i=1}^{t} \Lambda_p(\psi_i(a)) \prod_{j=t+1}^{t+s} \frac{\rho_{f_j,\psi_j(a)}(p^{\alpha(p)})}{p^{\alpha(p)}} \leq O(\alpha(p)^s) \prod_{i=1}^{t} \Lambda_p(\psi_i(a)) \sum_{0 \leq k_1, \ldots, k_t \leq \alpha(p)} \prod_{j=t+1}^{t+s} 1_{\psi_j(a) \equiv 0 \mod p^{k_j}},$$

the $\alpha(p)^s$ factor being there to account for the quantity of $s$-tuple $(k_j)$ where all the $k_j$ are in $[0, \alpha(p) - 1]$.

The total contribution of this average is thus

$$O(\alpha(p)^s) \sum_{0 \leq k_1, \ldots, k_t \leq \alpha(p)} \mathbb{E}_{a \in (\mathbb{Z}/p^n\mathbb{Z})^d} \prod_{i=1}^{t} \Lambda_p(\psi_i(a)) \prod_{j=t+1}^{t+s} 1_{\psi_j(a) \equiv 0 \mod p^{k_j}}.$$
We aim to bound the latter expression, to discard it as a small error term. We do it quite brutally. We bound \( \prod_{i=1}^{t} \Lambda_p(\psi_i(a)) \) by \( 2^t \). Then we just copy Matthiesen’s treatment. The term to bound is now

\[
O(\alpha(p)^s 2^t) \sum_{0 \leq k_{t+1}, \ldots, k_{t+s} \leq m} \sum_{M := \max k_i \geq \alpha(p)} \prod_{j=t+1}^{t+s} 1_{\psi_j(a) \equiv 0 \mod p^j}
\]

where we recognise the divisor density \( \alpha(p^{k_{t+1}}, \ldots, p^{k_{t+s}}) \) for the relevant system of affine forms. We note \( \delta_p \) the sum

\[
\delta_p := \sum_{0 \leq k_{t+1}, \ldots, k_{t+s} \leq m} \alpha(p^{k_{t+1}}, \ldots, p^{k_{t+s}}).
\]

Now since the coefficients of the linear part of the system are bounded and none of the form is 0, we see that except for a bounded number of primes \( p \), none of the forms is the 0 form mod \( p \), which implies, according to Proposition C.5, that

\[
\alpha(p^{k_{t+1}}, \ldots, p^{k_{t+s}}) \ll p^{-\max k_i}
\]

so that

\[
\delta_p \ll \sum_{0 \leq k_{t+1}, \ldots, k_{t+s} \leq m} p^{-M}.
\]

Now recall that

\[
\alpha(p) = v_p(W) = (C_1 + 1) \frac{\log \log N}{\log p} + O(1)
\]

for some integer \( C_1 \). Bounding the number of tuples \( (k_{t+1}, \ldots, k_{t+s}) \) satisfying \( \max k_i = j \) crudely by \( (j + 1)^s \), we conclude that for \( p \leq w(N) = \log \log \log N \)

\[
\delta_p \ll \sum_{j \geq C_1 \log \log N/2 \log p} p^{-j} j^s
\ll \sum_{j \geq C_1 \log \log N/2 \log p} p^{-j/2}
\ll (\log N)^{-C_1/4}.
\]

Hence \( (\alpha(p))^t \delta_p \ll (\log N)^{-C_1/5} \).

\[\blacksquare\]

**Lemma A.2.** For primes \( p \) tending to infinity,

\[
\beta_p = 1 + O(p^{-2}).
\]

Thus the product of the \( \beta_p \) is convergent (though it could be constantly 0 from a certain rank on, because some \( \beta_p \) may be 0 for a small \( p \)) and

\[
\prod_{p \leq w(n)} \beta_p = \left(1 + O\left(\frac{1}{w(N)}\right)\right) \prod_p \beta_p.
\]

\[\text{22}\]
Proof. Assume $p$ is large enough so that $p \nmid D_{t+1} \cdots D_{t+s}$ (recall that the $D_i$’s are the (negative) discriminants of our quadratic forms).

Letting $P(a)$ denote $\prod_{i=1}^{t} \Lambda_p(\psi_i(a)) \prod_{j=t+1}^{t+s} \frac{\rho_{f_j,\psi_j(a)}(p^m)}{p^m}$, we have, when $m$ tends to $\infty$

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

$$= \frac{1}{p^m} \left( \sum_{a \in (\mathbb{Z}/p^m\mathbb{Z})^d} \prod_{i=1}^{t} \Lambda_p(\psi_i(a)) \prod_{j=t+1}^{t+s} \frac{\rho_{f_j,\psi_j(a)}(p^m)}{p^m} P(a) + \sum_{a \in (\mathbb{Z}/p^m\mathbb{Z})^d} \prod_{i=1}^{t} \Lambda_p(\psi_i(a)) \prod_{j=t+1}^{t+s} \frac{\rho_{f_j,\psi_j(a)}(p^m)}{p^m} P(a) \right)$$

$$= \frac{1}{p^m} \sum_{a \in (\mathbb{Z}/p^m\mathbb{Z})^d} \prod_{i=1}^{t} \Lambda_p(\psi_i(a)) \prod_{j=t+1}^{t+s} \frac{\rho_{f_j,\psi_j(a)}(p^m)}{p^m} P(a) + 2^t O(sm^3p^{-m}).$$

To explain the error term, notice that if $\psi$ is an affine linear form in $d$ variable whose reduction modulo $p$ is not the trivial form, then the number of zeros of its reduction $\hat{\psi} : (\mathbb{Z}/p^m\mathbb{Z})^d \to \mathbb{Z}/p^m\mathbb{Z}$ is at most $p^{nd}/p^n$ (this is linear algebra in the case $m = 1$, and Hensel’s lemma allows one to extend to higher powers $m$; this is proven in Corollary C.4); we then use the triangle inequality to bound the number of summands in the second sum to $sp^{nd}p^{-m}$. We also require the fact that $\rho_{f_j,\psi}(p^m)/p^m \ll m$ (this is proven in [14], section 6). Finally, the product of the $\Lambda_p$ factors is not larger than $2^t$. So this error term tends to $0$ as $m$ tends to infinity. Let us consider the main term. Thanks to the choice of $p$ and the non-vanishing of the forms in $a \mod p^m$, we can use Lemma 6.3 from [14] which states that if $f$ is a PDBQF, of discriminant $D$, and if $p$ is a prime which does not divide $D$, and if $\beta \neq 0 \mod p^m$, then

$$\frac{\rho_{f,\beta}(p^m)}{p^m} = (1 - \chi_D(p)p^{-1}) \sum_{k=0}^{m} 1_{p^k|D} p^k$$

so that

$$\beta_p = \lim_{m \to \infty} \mathbb{E}_{a \in \mathbb{Z}/p^m\mathbb{Z}} \prod_{i=1}^{t} \Lambda_p(\psi_i(a)) \prod_{j=t+1}^{t+s} (1 - \chi_D_j(p)p^{-1}) \sum_{k=0}^{m} 1_{p^k|\psi_j(a)} \chi_D_j(p^k)$$

where we have obviously reintegrated the once excluded $a$, because their sparsity ensures that they do not disturb 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 $a' \in (\mathbb{Z}/p\mathbb{Z})^d$. Thus the average $\mathbb{E}_a$ becomes

$$\prod_{j=t+1}^{t+s} (1 - \chi_D_j(p)p^{-1}) \mathbb{E}_{a \in (\mathbb{Z}/p\mathbb{Z})^d} \prod_{i=1}^{t} \Lambda_p(\psi_i(a)) \prod_{a' \in (\mathbb{Z}/p^{m-1}\mathbb{Z})^d} \prod_{j=t+1}^{t+s} \sum_{k=0}^{m} 1_{p^k|\psi_j(a'+pb)} \chi_D_j(p^k).$$

We expand the product of sums as follows

$$\prod_{j=t+1}^{t+s} \sum_{k=0}^{m} 1_{p^k|\psi_j(a'+pb)} \chi_D_j(p^k) = 1 + \sum_{j} \sum_{k_j=1}^{m} 1_{p^{k_j}|\psi_j(a'+pb)} \chi_D_j(p^{k_j}) + \sum_{k_{t+1}, \ldots, k_{t+s} \geq 2} \prod_{k_i=1}^{t+s} 1_{p^{k_i}|\psi_j(a'+pb)} \chi_D_j(p^{k_i})$$

at least two $k_i > 0$. 

23
according to whether we take no, one or several non-zero \(k\). The expectation over \(a\) from (13) then splits into three terms; the first one is

\[
\mathbb{E}_{a \in (\mathbb{Z}/p\mathbb{Z})^d} \prod_{i=1}^t \Lambda_p(\psi_i(a)).
\]

The second one is

\[
\sum_{j=t+1}^{t+s} \sum_{k_j=1}^m \chi_{D_j}(p^{k_j}) \mathbb{E}_{a \in (\mathbb{Z}/p\mathbb{Z})^d} \prod_{i=1}^t \Lambda_p(\psi_i(a)) \mathbb{E}_{b \in (\mathbb{Z}/p^{m-1}\mathbb{Z})^d} 1_{p^{k_j}|\psi_j(a+pb)}. \tag{14}
\]

Now we use \(\psi_j(a+pb) = \psi_j(a) + p\tilde{\psi}_j(b)\) where \(\tilde{\psi}\) is the linear part of \(\psi\). If \(p^{k_j}\) is to divide \(\psi_j(a) + p\tilde{\psi}_j(b)\), we need \(\psi_j(a) = 0\) and \(p^{k_j-1} | \tilde{\psi}_j(b)\). Because of Corollary C.5,

\[
\mathbb{E}_{b \in (\mathbb{Z}/p^{m-1}\mathbb{Z})^d} 1_{p^{k_j-1} | \tilde{\psi}_j(b)} = p^{-k_j+1}
\]

so we get for (14) the expression

\[
\sum_{j=t+1}^{t+s} \sum_{k_j=1}^m \chi_{D_j}(p^{k_j}) p^{-k_j} \mathbb{E}_{a \in (\mathbb{Z}/p\mathbb{Z})^d} \prod_{i=1}^t \Lambda_p(\psi_i(a)) p 1_{\psi_j(a)=0}
\]

To deal with the last term which is

\[
\mathbb{E}_{a \in (\mathbb{Z}/p^m\mathbb{Z})^d} \prod_{i=1}^t \Lambda_p(\psi_i(a)) \sum_{k_{t+1}, \ldots, k_{t+s}} \prod_{i=k_{t+1}}^{k_{t+s}} 1_{p^{k_i} | \psi_j(a)} \chi_{D_j}(p^{k_j}) \tag{15}
\]

we bound crudely \(\prod_{i=1}^t \Lambda_p(\psi_i(a))\) by \(2^t\). Thus (15) becomes an

\[
O \left( \sum_{k_{t+1}, \ldots, k_{t+s}} \alpha(p^{k_{t+1}}, \ldots, p^{k_{t+s}}) \right)
\]

To bound again this term, we remember that the \(\psi_j\) form a system of finite complexity: no two forms are affinely related. This implies, thanks to Proposition C.5 that for \(p\) large enough \(^2\) depending on \(s, d, L\), we have

\[
\alpha(p^{k_{t+1}}, \ldots, p^{k_{t+s}}) \leq p^{-\max_i(k_i + k_j)} \leq p^{-1 - \max(k_i)}
\]

whenever at least two \(k_i\) are non-zero. For any \(k \geq 1\), there are at most \(sk^{s-1}\) \(s\)-tuples that satisfy \(\max k_i = k\). Thus

\[
\sum_{k_{t+1}, \ldots, k_{t+s}} \alpha(p^{k_{t+1}}, \ldots, p^{k_{t+s}}) = O(\sum_{k \geq 1} sk^{s-1}p^{-k-1}) = O_s(p^{-2})
\]

\(^2\)We need \(p\) to be large because for some small \(p\), there could be two forms that, though affinely independent, become dependent when reduced modulo \(p\). Such primes are called \textit{exceptional}. The same need for large \(p\) will appear again later.
and this bound does not depend on \( a \in (\mathbb{Z}/p^m\mathbb{Z})^d \) any more.

So we get

\[
\beta_p = \prod_{j=t+1}^{t+s} \left( 1 - \chi_{D_j}(p)p^{-1} \right) \\
\left( \mathbb{E}_{a \in (\mathbb{Z}/p\mathbb{Z})^d} \prod_{i=1}^{t} \Lambda_p(\psi_i(a)) + \sum_{j=t+1}^{t+s} \mathbb{E}_{a \in (\mathbb{Z}/p\mathbb{Z})^d} p1_{\psi_j(a)=0}^{t} \prod_{i=1}^{t} \Lambda_p(\psi_i(a)) \sum_{k=1}^{+\infty} \chi_{D_j}(p^k)p^{-k} \right) + O_{s,t}(p^{-2})
\]

(16)

But \( \mathbb{E}_{a \in (\mathbb{Z}/p\mathbb{Z})^d}^{t} \prod_{i=1}^{t} \Lambda_p(\psi_i(a)) = 1 + O_t(p^{-2}) \), as appeared in [9]. To prove it, write this expression as

\[
\left( \frac{p}{p-1} \right)^t \mathbb{P}(\prod_{i=1}^{t} \psi_i(a), p = 1)
\]

and notice that the probability, because of the affine independence and by some linear algebra and inclusion-exclusion, is \( 1 - t/p + O(p^{-2}) \) for \( p \) large enough depending on \( d, t, L \) while the front factor is \( 1 + t/p + O(p^{-2}) \) whence the result. Similarly for any \( j \in \{ t+1, \ldots, t+s \} \), we have

\[
\mathbb{E}_{a \in (\mathbb{Z}/p\mathbb{Z})^d} p1_{\psi_j(a)=0}^{t} \prod_{i=1}^{t} \Lambda_p(\psi_i(a)) = p \left( \frac{p}{p-1} \right)^t \mathbb{P}(\prod_{i=1}^{t} \psi_i(a), p = 1) \text{ and } \psi_j(a) = 0 = 1+O(p^{-2})
\]

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 (16),

\[
\beta_p = \prod_{j=t+1}^{t+s} \left( 1 - \chi_{D_j}(p)p^{-1} \right) \left( 1 + \sum_{j=t+1}^{t+s} \sum_{k_j > 0} \chi_{D_j}(p^{k_j})p^{-k_j} + O_s,t(p^{-2}) \right) = 1 + O(p^{-2}).
\]

This is the claimed result (the implied constant depends on \( t, d, s, L \)).

\[\blacksquare\]

**B Verification of the linear forms condition**

This section is dedicated to the lengthy and technical proof of Proposition 4.6, that is, the verification that our majorant, introduced in Subsection 4.2.3, satisfies the linear forms condition.

**Remark B.1.** We will thereafter basically reproduce Matthiesen’s proof in [14], also inspiring ourselves of and the more recent paper [1]. However, there is some flaw there, as she oversaw the possibility that \( u \) and \( dm^2 \epsilon \) are not coprime; we provide, based in the earlier paper [14], a correction of these computations.

25
Compared to Matthiesen’s articles, the introduction of the majorant for the von Mangoldt function adds $\log R$ factors which will be cancelled during the Fourier transformation step. It also adds $\frac{\phi(W)}{W}$ factors which remains untouched throughout the proof. And in the core of the calculation, it adds to the variables $d, m, c, u$ another variable $l$ also ranging among the integers whose prime factors are all greater than $w$, which shall mingle 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.

Notational conventions for the proof: in order to somewhat lighten the formidable involved notations, we will not always specify the range on sums, product of integrals. In principle, the name of the variable should suffice to the reader to know what its range is. For instance,

1. The integer vector $n$ will always range in $\mathbb{Z}^d \cap K$.
2. We put $\phi_j(n) = W \psi_j(n) + b_j$, for $j \in [t + s]$, where $b_j$ is coprime to $W$. Thus $\phi_j(n)$ is always coprime to $W$, in other words all its prime factors are greater than $w(N)$.
3. For $k = 1, \cdots, t$ and $j = 1, 2$, $l_{k,j}$ is a positive integer.
4. For $k = t + 1, \cdots, t + s$ and $j = 1, 2$, $e_{k,j}$ is a positive integer in $\langle Q_k \rangle$ where $Q_k = Q_{D_k}$.
5. For $k = t + 1, \cdots, t + s$, $s_k$ will range from $2/\gamma$ to $(\log \log N)^3$ and $i_k$ from $\log_2 s - 2$ to $6 \log \log N$, while $u_k$ ranges in $U(s_k, i_k)$. We believe no $s_k$ shall provoke an ambiguity with $s$, the number of factors of the form $\nu_{\text{Matt}, b}$.
6. Occasionally we may want to write $e_k$ for $e_{k,1}$ and $e'_k = e_{k,2}$; similarly $l_k = l_{k,1}$ and $l'_k = l_{k,2}$. Moreover $e_k$ will be the least common multiple (lcm) of $e_k$ and $e'_k$; while $\lambda_k$ will be the lcm of $l_k$ and $l'_k$.
7. For $k = t + 1, \cdots, t + s$, the integer $d_k$ have all its prime factors in $\mathcal{P}_k$ where $\mathcal{P}_k = \mathcal{P}_{D_k}$.
8. For $k = t + 1, \cdots, t + s$, $m_k$ will range in $\langle Q_{D_k} \rangle$.
9. A bold character denotes a vector; thus $l = (l_{j,k})_{j \in [t]}_{k=1,2}$ and again the range of such indices $j, k$ will frequently be omitted.
10. Moreover, we will use the shortcut $\prod \mu$ to denote the product of all relevant Moebius factors, generally $\mu(e_i) \mu(e'_i) \mu(l_j) \mu(l'_j)$ for instance. Similarly we will use $\prod \chi$ to denote the product of relevant $\chi$ factors.

With these notations, we expand the first term of (12) as

$$
(\log R \frac{\phi(W)}{W})^t \prod_{j = t+1}^{t+s} C_{D_j, \gamma}^{-1} \sum_{n \in \mathbb{Z}^d \cap K} \sum_{i \in [t]} \sum_{l, l'_i} \mu(l_i) \mu(l'_i) \chi \left( \frac{\log l_i}{\log R} \right) \chi \left( \frac{\log l'_i}{\log R} \right) 1_{\lambda_i | \phi_j(n)} \\
\prod_{j = t+1}^{t+s} \sum_{s_j, i_j, u_j} 2^{\nu_j} 1_{u_j | \phi_j(n)} \sum_{d_j, m_j, e'_j} 1_{d_j m_j \in \phi_j(n) \mu(e_j) \mu(e'_j)} \chi \left( \frac{\log e_j}{\log R} \right) \chi \left( \frac{\log e'_j}{\log R} \right) \chi \left( \frac{\log d_j}{\log R} \right) \chi \left( \frac{\log m_j}{\log R} \right)
$$

(17)
We will drop the initial factor \((\log R)^t\frac{\phi(W)}{w}\prod_{j=t+1}^{t+s} C_{D_j,\gamma}^{-1} = O((\log R)^t) = O((\log N)^t)\) and work on the sum. It is basically a sum over \(n\) of \(t + s\) products, and we aim at transforming it into a product of \(t + s\) sums. We will remember to multiply the error terms obtained during the transformation of this sum by \((\log R)^t\).

We remark that when \(u_j, d_j, m_j, e_j, e_j'\) divide \(\phi_j(n)\) and when \(u_j\) satisfies \(\gcd(u_j, \phi_j(n) / u_j) = 1\), we can write, for \(x\) any of the symbols \(e, e', d, m, a\) unique decomposition \(x_j = \bar{x}_j v_{j,x}\) with \(\gcd(\bar{x}_j, u_j) = 1\) and \(v_{j,x} | u_j\). We would very much like to do this decomposition, but not every term satisfies this coprimality condition. However, the following claim shows that we can pretend it is so at a small cost.

**Claim 1.** The summands in (17) satisfying \(\gcd(u_j, \phi_j(n) / u_j) > 1\) for some \(j \in [t + 1; t + s]\) or \(\gcd(u_j, u_i) > 1\) for some \(i \neq j\) contribute only \(O(N^{-(\log \log N)^{-3/8}})\).

**Proof.** We write the contribution of these summands as

\[
S = \sum_{i,s} \prod_{j=t+1}^{t+s} 2^{s_j} \mathbb{E}_n a_n
\]

where

\[
a_n = a_{n,i,s} = \sum_u 1_{j \not\mid \gcd(u_j, \phi_j(n) / u_j)} \prod_{i=1}^{t} \mu(l_i) \mu(l_i') \chi\left(\frac{\log l_i}{\log R}\right) \chi\left(\frac{\log l_i'}{\log R}\right) 1_{i | \phi_i(n)} \prod_{j=t+1}^{t+s} \mu(d_j) \mu(l_j) \mu(l_j') \chi\left(\frac{\log e_i}{\log R}\right) \chi\left(\frac{\log e_i'}{\log R}\right) 1_{\Delta_j | \phi_j(n)}
\]

with the notation \(\Delta_j = \gcd(u_j, d_j m_j^2 e_j)\). We remark that for any non zero summand in \(S\), there exists a prime \(p \geq N^{1/(\log \log N)^3}\) satisfying \(p^2 | \prod_j \phi_j(n)\). We basically apply the simple rule based on Cauchy-Schwarz

\[
(\mathbb{E}_{n \in \mathbb{Z}^d \cap K} a_n)^2 \leq \mathbb{P}(a_n \neq 0) \mathbb{E} a_n^2.
\]

Now if \(a_n \neq 0\) then either the value in \(n\) of one of the \(s\) last linear forms \(\phi_i(n)\) has a repeated prime factor, or the values in \(n\) of two of the \(t + s\) linear forms have a common prime factor, these prime factors being factors of \(u_i\) which by definition only has prime factors bigger than \(N^{1/(\log \log N)^3}\); moreover because of the support of \(\chi\) we can also assume that its prime factors are at most \(N^7\). Thus

\[
\mathbb{P}(a_n \neq 0) \leq \sum_{p > N^{1/(\log \log N)^3}} \mathbb{P}(p^2 | \prod_i \phi_i(n))
\]

but the \(p\) in this range are not exceptional primes, i.e. primes modulo which the linear form are affinely dependent (such primes, thanks to the \(W\)-trick and the fact that no two of the original linear forms \(\psi_i\) were linearly dependent, are all at most \(w(N)\)). Thus

\[
\mathbb{P}(p^2 | \prod_i \phi_i(n)) \ll p^{-2} + O\left(\frac{N^{d-1}}{\text{Vol}(K)}\right) = O\left(\frac{N^{-\omega}}{\text{Vol}(K)}\right)
\]

27
, according to Proposition C.6 and the fact that \(|K \subset \mathbb{Z}^d| \sim \text{Vol}(K)^3\), so that

\[
\mathbb{P}(a_n \neq 0) \leq \sum_{N^{1/(\log \log N)^3} < p \leq N^{\gamma}} \mathbb{P}(p^2 \mid \prod_i \phi_i(n)) \ll \sum_{p > N^{1/(\log \log N)^3}} p^{-2 + N^{-1}/\text{Vol}(K)} \sum_{p \leq N^{\gamma \text{perm}}} p^2 \ll N^{-1/(\log \log N)^3}.
\]

Assuming that \(\gamma\) 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 other one. Concerning the expectation of the \(a_n^2\), we bound quite crudely as follows, noticing that the \(\mu\) and \(\chi\) are 1-bounded and using Hölder’s inequality

\[
\mathbb{E}_n a_n^2 \leq \mathbb{E}_n \left( \sum_{d,m,e,l} \prod_{i=1}^t 1_{\lambda_i|\phi_i(n)} \prod_{j=t+1}^{t+s} 1_{\Delta_j|\phi_j(n)} \right)^2
\]

\[
\ll \prod_{i=1}^t \left( \mathbb{E}_n \left( \sum_{l_i,j_i} 1_{\lambda_i|\phi_i(n)} \right) \right)^{2(t+s)+1/(t+s)} \prod_{j=t+1}^{t+s} \left( \mathbb{E}_n \left( \sum_{d_j,m_j,e_j,e_j',u_j} 1_{\Delta_j|\phi_j(n)} \right) \right)^{2(t+s)}
\]

\[
\ll (\log N)^{O_1,\gamma(1)}
\]

where the last term follows from Lemma 3.1 of [13], bounding moments of the divisor function, and the observation that for instance \(\sum_{l_i,j_i} 1_{\lambda_i|\phi_i(n)} \leq \tau(\phi(n))^2\). Thus, it is clear that \(|\mathbb{E}_n a_n| \ll N^{-(\log \log N)^3/4}\). Summing now over \(i,s\) in their ranges, we certainly get that \(|S| \leq N^{-(\log \log N)^3/8}\) as desired. \(\square\)

Thus to evaluate (17), we shall pretend all summands satisfy the coprimality condition, transform them under this hypothesis, and then reintegrate once excluded terms, at the same cost \(O(N^{-(\log \log N)^3/8})\). So (17) is equal to

\[
\sum_{i,s} \sum_{u} \mathbb{E}_n \prod_{i \in [t], k=1,2} \mu(i,k) \chi \left( \frac{\log l_{i,k}}{\log R} \right) 1_{\lambda_i|\phi_i(n)} \prod_{j \in \ll [t+1; t+s]} 2^x_j \sum_{d_j,m_j,e_j,e_j',m_j \text{ coprime to } u_j} \mu(e_j v_j, e_j' m_j) + O(N^{-(\log \log N)^3/8})
\]

\[
\sum_{v_{j,d,e_{j,m},e_{j,e'},v_{j,e'}} \mid u_j} \prod_{x \in \{d,e,e',m\}} \chi \left( \frac{\log x v_j}{\log R} \right) \mu(v_j v_j, e_j \chi e_j) 1_{u_j v_j, e_j \chi e_j} + O(N^{-(\log \log N)^3/8})
\]

where the dashed sum indicates a sum over vectors whose entries are coprime.

Above and from now on, the vectors \(d,e,m\) are automatically supposed to be entry-wise coprime to the vector \(u\), and the vector \(v\) 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 \(v_{j,d,e,m} \in \langle P_j \rangle\), \(v_{j,e} \in \langle Q_j \rangle\), \(v_{j,e'} \in \langle P_j \rangle\).

\[\text{(Here, we make the hypothesis that } \text{Vol}(K) \gg N^d \text{ or at least that } N^{d-1} = o(\text{Vol}(K)). \text{ Indeed, in the statement of the main theorem, we could also add the assumption that } \text{Vol}(K) \gg N^d \text{ because otherwise the error term is not smaller than the main term.)} \]
Claim 2. The main term of (18) is equal to

\[
\sum_{i,s} \sum_{u \in \mathbb{d}, m \in \mathbb{m}} \alpha((\lambda_i)_{i \in [t]}, (d_j m_j^2 \epsilon_j)_{j \in [t+1; t+s]}) \\
\sum_{v \in \mathbb{v}, k=1, 2} \mu(i, k) \chi \left( \frac{\log l_{i,k}}{\log R} \right) \chi \left( \frac{\log x_j v_j x}{\log R} \right)
\]

up to an error term of the form \( O \left( N^{d-1+O(\gamma)} / \text{Vol}(K) \right) \).

We remark that this error term, after multiplication by the initial factor \( O((\log N)^4) \), is still of the same form.

**Proof.** Here, we first have to simply replace

\[
\mathbb{E}_{n \in \mathbb{Z} \cap K} \prod_{i=1}^{t} 1_{\lambda_i \mid \phi_i(n)} \prod_{j=t+1}^{t+s} 1_{u_j d_j m_j^2 \epsilon_j \mid \phi_j(n)}
\]

by

\[
\alpha((\lambda_i)_{i \in [t]}, (u_j d_j m_j^2 \epsilon_j)_{j \in [t+1; t+s]}).
\]

The error term inherent to such a simplification is as small as \( O \left( N^{d-1+O(\gamma)} / \text{Vol}(K) \right) \). To see this, the important point is that for any set of tuples bringing a non-zero contribution, for any \( i \in [t] \) and any \( j \in [t+1, t+s] \), we have that \( \lambda_i \leq N^\gamma \) and that \( u_j d_j m_j^2 \epsilon_j = N^{O(\gamma)} \)

because \( d_j, m_j, \epsilon_j, \epsilon_j' \leq N^\gamma \) because of the support of \( \chi \) and \( u_j \leq N^\gamma \) because of the definition of \( U(i_j, s_j) \) (cf the remark (3) after the Lemma 4.2 of [14]). Thus we can use Lemma C.2, which immediately implies that

\[
\mathbb{E}_{n} \prod_{j=t+1}^{t+s} 1_{u_j d_j m_j^2 \epsilon_j \mid \phi_j(n)} \prod_{i \in [t]} 1_{\lambda_i \mid \phi_i(n)}
\]

\[
= \alpha((\lambda_i)_{i \in [t]}, (u_j d_j m_j^2 \epsilon_j)_{j=t+1, \ldots, t+s}) + O(N^{d-1+O(\gamma)} / \text{Vol}(K)).
\]

To bound the contribution of this error term to the sum defining the main term of (18), we simply notice that the number of terms is \( O(N^{\gamma}) \) anyway, that the \( \mu \) and \( \chi \) factors are 1-bounded, and that \( 2^{s_j} \) is always \( o(N^{\gamma}) \) because \( s \leq (\log \log N)^3 \).

Notice we can also exclude summands for which \( \gcd(\lambda_i, u_j) > 1 \) for some \( i \in [t], j \in [t+1; t+s] \) because of the very same argument as in Claim 1. For summands satisfying to the contrary \( \gcd(\lambda_i, u_j) = 1 \), by multiplicativity of \( \alpha \) and because of the other implicit coprimality conditions, we can write that

\[
\alpha((\lambda_i)_{i \in [t]}, (u_j d_j m_j^2 \epsilon_j)_{j \in [t+1; t+s]}) = \frac{\alpha((\lambda_i)_{i \in [t]}, (d_j m_j^2 \epsilon_j)_{j \in [t+1; t+s]})}{\prod_j u_j}.
\]

This concludes the proof of this claim with a dashed sum on \( u \) instead of the normal sum (and strictly speaking with some vectors \( \mathbf{1} \) excluded), but we can reintegrate now the once 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 we consider the corresponding term, that is

$$\sum_u \sum_{d,e,m,l} \alpha((\lambda_i)_{i \in [t]}, (d_j m_j^2 \xi_j)_{j \in [t+1:t+s]}) \prod_{i \in [t], k=1,2} \mu(l,k) \chi \left( \frac{\log l,k}{\log R} \right)$$

$$\sum_v \prod_{j \in [t+1:t+s]} \frac{2^{s_j}}{u_j} \mu(e_j v_j, e') \prod_{x_j \in \{d_j, e_j, e'_j, m_j\}} \chi \left( \frac{\log x_j v_j, e'}{\log R} \right)$$

We use the integral representation of $\chi$ now (in other words, the Fourier transform). Letting $\theta$ be the Fourier transform of the smooth compactly supported function $x \mapsto e^x \chi(x)$, we learn from the theory of Fourier transforms that

$$\forall A > 0 \quad \theta(\xi) \ll_A (1 + |\xi|)^{-A}. \quad (21)$$

This allows us to reconstruct $\chi$ from $\theta$ through an integral over the compact interval\footnote{We prefer integrating over a compact set, in order to swap summation and integration using Fubini’s theorem without any worry.}

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

at the cost of a tolerable error term, more precisely for any $A > 0$:

$$\chi \left( \frac{\log x}{\log R} \right) = \int_I x^{-\frac{1+i\xi}{\log R}} \theta(\xi) d\xi \quad (20)$$

$$= \int_I x^{-\frac{1+i\xi}{\log R}} \theta(\xi) d\xi + O(x^{-\frac{1}{\log R} \log^{-A} R})$$

When plugging this into our sum, we need $4s + 2t$ real variables $\xi_{j,k}$ with $k = 1, \cdots , 4$ for $j = t+1, \cdots , t+s$ and $k = 1, 2$ for $j = 1, \cdots , t$. And we write $z_{j,k} = (1 + i\xi_{j,k})/(\log R)$.

We shall sometimes allow, for some function $f$, the slight abuse of notation

$$\prod_{j,k} f(\xi_{j,k}) = \prod_{i \in [t], k \in [2]} f(\xi_{i,k}) \prod_{j \in [t+s] \setminus [t], k \in [4]} f(\xi_{j,k}).$$

Thus the main term is now

$$\sum_{d,m,e,l} \alpha((\lambda_i)_{i \in [t]}, (d_j m_j^2 \xi_j)_{j > t}) \sum_{i > t, u_i, v_i} \frac{2^{s_i}}{u_i} \times \left( \prod_{j > t} \mu \prod_{j > t} \theta(\xi_{j,1}) \theta(\xi_{j,2}) \theta(\xi_{j,3}) \theta(\xi_{j,4}) \right)$$

$$\prod_{j \in [t]} \left( \prod_{\xi_{j,k}} \theta(\xi_{j,k}) d\xi_{j,k} \right) + O((\log R)^{-A} \prod_{j,k} \theta(\xi_{j,k}) d\xi_{j,k})$$

where we have introduced the notation $\tilde{x}_j = x_j v_j, x$ where $x$ is any of the symbols $e, e', d, m,$ and the other notation $\mathbf{v}_1 = (v_{i,d}, v_{i,e}, v_{i,e'}, v_{i,m})$.

Here the term arising from the big oh will not matter too much, even after summation over $\mathbf{s, i}$, because on the one hand, we have the useful bound

$$\sum_{s, i} \prod_{j = t+1}^{t+s} \sum_{u_j \in [\ell(s, i_j)]} \frac{2^{s_j}}{u_j} = O(1)$$
originating from [13], Proposition 4.2. On the other hand, we can suppress the sum over \(v\) by reintegrating in the sum over \(d, m, e\) the summands not termwise coprime to \(u\). But

\[
\sum_{d, m, e} \alpha((\lambda_i)_{i \in [t]}, (d_i m_i^2 \epsilon_i)_{i \in [t+1:t+s]}) \left( \prod_{i \in [t], k=1:2} l_i,k \prod_{j>t} e_j,1 e_j,2 m_j d_j \right)^{-1/\log R} \ll \log^{O(t+s)} N
\]

after the computations of [14] (page 33). Thus when choosing \(A > 0\) large enough we find that

\[
\left( \log R \phi(W) \right)^t \sum_{s, i, d, m, e, l} \left( \alpha((\lambda_i)_{i \in [t]}, (d_i m_i^2 \epsilon_i)_{i \in [t+t']) \Omega(\log^{-A} N^\gamma (\prod_{j,k} e_j,k l_j,k d_j m_j)^{-1/\log R}) \right) = \log^{-A/2} N^\gamma
\]

which is \(o(1)\). So incurring only an error \(o(1)\), which is tolerable for (12), we can replace (20) by

\[
\sum_{d, m, e} \alpha((\lambda_i)_{i \in [t]}, (d_j m_j^2 \epsilon_j)_{j \geq t}) \prod_{i \geq t} \sum_{u_i} \frac{2^{s_i}}{u_i} \\
\times \int_{\prod_{l=1}^{l+s} \prod_{j \geq t} \prod_{i \in [t]} \frac{e_j}{e_{j,1}} \frac{e_j}{e_{j,2}} \frac{d_j}{d_j} \frac{m_j}{m_j} \prod_{j \leq t} \frac{l_j,1}{l_j,1} \frac{l_j,2}{l_j,2} \prod_k \theta(\xi_j,k) d\xi_j,k.
\]

Now we swap the summation \(\sum_{d, m, e, l}\) and the integration over a compact \(\int_{\prod_{l=1}^{l+s} \prod_{j \geq t} \prod_{i \in [t]} \frac{e_j}{e_{j,1}} \frac{e_j}{e_{j,2}} \frac{d_j}{d_j} \frac{m_j}{m_j} \prod_{j \leq t} \frac{l_j,1}{l_j,1} \frac{l_j,2}{l_j,2} \prod_k \theta(\xi_j,k) d\xi_j,k\), using Fubini’s theorem. This causes no problem because the sum is absolutely convergent; we are not explicitly going to prove the absolute convergence, but this results from the bounds we are going to prove in the proof of the next claim. We also continue swapping summation and multiplication, by forcing at little cost an extra coprimality condition: we show we can restrict to tuples where \((d_j m_j \epsilon_j, \lambda_i) = 1\) for all \(i, j\) and \((d_i m_i \epsilon_i, d_j m_j \epsilon_j) = 1\) for all \(i \neq j\).

**Claim 3.** Let \(i, s, u, v\) be fixed vectors of integers satisfying the usual conditions. Then the following equality holds

\[
\sum_{d, m, e} \alpha((\lambda_i)_{i \in [t]}, (d_j m_j^2 \epsilon_j)_{j \geq t}) \prod_{j=t+1}^{t+s} \prod_{i \in [t]} \frac{e_j}{e_{j,1}} \frac{e_j}{e_{j,2}} \frac{d_j}{d_j} \frac{m_j}{m_j} \prod_{j \leq t} \frac{l_j,1}{l_j,1} \frac{l_j,2}{l_j,2}
\]

\[
= (1 + o(w(N))^{-1/2}) \sum_{d, m, e} \alpha((\lambda_i)_{i \in [t]}, (d_j m_j^2 \epsilon_j)_{j \geq t}) \prod_{j=t+1}^{t+s} \prod_{i \in [t]} \frac{e_j}{e_{j,1}} \frac{e_j}{e_{j,2}} \frac{d_j}{d_j} \frac{m_j}{m_j} \prod_{j \leq t} \frac{l_j,1}{l_j,1} \frac{l_j,2}{l_j,2}
\]

where the dashed sum is retracted to tuples where the above coprimality conditions hold.

**Remark B.2.** We need an other, more subtle argument to impose this coprimality compared to the coprimality condition involving the \(u_i\) of Claim 1, because a crucial ingredient of the proof of this claim was that the prime factors involved were all at least \(N^{(\log \log N)^{-3}}\), an assumption we don’t have for \(d, m, e\).

**Proof (of Claim 3).** The mission is to bound the contribution of the entries failing the coprimality conditions. To achieve this, we exploit the multiplicativity of each such
The requirement that at least two \( \alpha_k \), which is a term of the form
\[
\alpha((\lambda_i)_{i \in [t]}, (d_j m_j^2 \epsilon_j)_{j > t}) \prod_{i=1}^{t} \sum_{l_i}^{T_{z_i}} T_{l_i}^{z_i} \mu(l_i) \mu(l_{i+1}) \prod_{j=t+1}^{t+s} \epsilon_j \prod_{i \in [t]} \mu(\epsilon_i) \mu(\epsilon_{i+2}),
\]
(23) in order to write it as a product over primes; only primes greater than \( w(N) \) need be considered, as other ones have no chance to divide any of the parameters. Reusing the notation \( v_i \) employed in Claim 2, we can even separate primes into those which divide a single \( v_i \) and those which divide at least two of them. Thus, a typical summand is factored into

1. a factor arising from the primes dividing at least two of the \( v_i \), i.e. a factor of the form \( \alpha(k_1, \cdots, k_{t+s}) \) times a complex factor of modulus at most one (this is because \( z_j, k \) has a positive real part) where for any \( i, p | k_i \Rightarrow p | \prod_{j \neq i} k_j \).

2. a factor arising from the primes dividing a single \( v_i \), which is again of the form of (23) but with parameters satisfying the coprimality condition.

A factor of the first kind, corresponding to some tuple \( k_1, \cdots, k_{t+s} \) with
\[
p | k_i \Rightarrow p | \prod_{j \neq i} k_j,
\]
can of course appear in several summands. Its total contribution to the sum is then bounded by
\[
\alpha(k_1, \cdots, k_{t+s}) \left| \sum_{\substack{d \text{ prime} \mid \prod_{i | k_i} k_i}} \alpha((\lambda_i)_{i \in [t]}, (d_j m_j^2 \epsilon_j)_{j > t}) \prod_{i \in [t]} \mu(l_i) \mu(l_{i+1}) \prod_{j=t+1}^{t+s} \epsilon_j \prod_{i \in [t]} \mu(\epsilon_i) \right| \leq \alpha((k_i)) \prod_{p \mid \prod_{i | k_i} k_i} (1 + O(p^{-1})) \left| \sum_{\substack{d \text{ prime} \mid \prod_{i | k_i} k_i}} \alpha((\lambda_i), (d_j m_j^2 \epsilon_j)) \prod_{i \in [t]} \mu(l_i) \mu(l_{i+1}) \prod_{j=t+1}^{t+s} \epsilon_j \prod_{i \in [t]} \mu(\epsilon_i) \right|
\]
by Moebius inversion and the triangle inequality. Next we try evaluate the sum of all occurring terms of the form \( \alpha(k_1, \cdots, k_{t+s}) \prod_{p \mid \prod_{i | k_i} k_i} (1 + O(p^{-1})) \). Using multiplicativity, we can write them as products over primes, these primes being all larger than \( w(N) \) because \( \phi_j(n) \equiv b_j \mod W \) is coprime to \( W \). Thus, a crude bound is
\[
\prod_{p > w(N)} \left( 1 + \sum_{a_1, \cdots, a_{t+s} \geq 0} a_1^2 + \cdots + a_t^2 + a_{t+1}^4 + \cdots + a_{t+s}^4 \alpha((p^{a_1}))) (1 + O(p^{-1})) \right) - 1
\]
where we used the simple bound \( \tau_k(p^{a_1}) \leq a_k^{k-1} \) for \( k = 3 \) (because of \( \lambda_i = \lambda_i / l_i \cdot \lambda_i / l_i' \cdot l_i / \lambda_i \), hence the number of occurrences of \( \lambda_i \) is bounded by the number of decompositions of it into three factors) and for \( k = 5 \) (because of \( d_j m_j^2 \epsilon_j = d_j \cdot m_j^2 \cdot \epsilon_j / \epsilon_j \cdot \epsilon_j / \epsilon_j' \cdot \epsilon_j' / \epsilon_j' \)). The requirement that at least two \( a_i \) be positive comes from the fact that these factors \( \alpha(k_1, \cdots, k_{t+s}) \) arose from the first type of factors (see above). 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 we 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 since the system is of finite complexity and at least two \( a_i \) are nonzero, \( \alpha((p^a_i)_{i \in [t+s]}) \leq p^{-k-1} \) according to Proposition C.5. Thus

\[
\sum_{a_1, \ldots, a_{t+s} \geq 0, \text{at least two } a_i > 0} O(a_1^2 + \cdots + a_{t+s}^2 + a_{t+1}^4 + \cdots + a_{t+s}^4) \alpha((p^a_i)) (1 + O(p^{-1})) \ll \sum_{k \geq 1} p^{k-1}k^{t'3} \ll \sum_{k \geq 1} p^{-3k/4-1}
\]

the last bound 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). Now the final bound is \( p^{-3/2} \), and since

\[
\prod_{p > w(N)} (1 + p^{-3/2}) - 1 \leq \sum_{n > w(N)} n^{-3/2} \ll w(N)^{-1/2},
\]

the Claim 3 follows.

The advantage of this coprimality is that this way, the \( \alpha \) ‘dissolves’ itself, or becomes the product of the reciprocals of its arguments, as follows.

\[
\sum_{d,m,e,l} \alpha((\lambda_i)_{i \in [t]}, (d_jm_j^2 \epsilon_j)_{j \geq t}) \prod_{j=t+1}^{t+s} \frac{\epsilon_j - \epsilon_{j,1}^{-z_{j,1}} \epsilon_{j,2}^{-z_{j,2}} d_j^{-z_{j,3}} m_j^{-z_{j,4}} \prod_{i=t}^{t+s} l_i^{-1} \epsilon_j \epsilon_{j,1}^{-z_{j,1}} \epsilon_{j,2}^{-z_{j,2}} \epsilon_j \epsilon_{j,1}^{-z_{j,1}} \epsilon_{j,2}^{-z_{j,2}}}{\lambda_i}
\]

We remark that we wrote here an equation without the tilded versions of the parameters. We will in the following avoid them, thanks to the following easy remark, valid for any fixed \( u \)

\[
\sum_{d,m,e,l} \prod_{j=t+1}^{t+s} \mu(e_{j}) \mu(e'_{j}) \alpha((\lambda_i)_{i \in [t]}, (d_jm_j^2 \epsilon_{j} \epsilon_{j,1}^{-z_{j,1}} \epsilon_{j,2}^{-z_{j,2}} d_j^{-1} - z_{j,3} m_j^{-2} z_{j,4}) \prod_{i=t}^{t+s} \lambda_i)
\]

\[
= \sum_{d,m,e,l} \prod_{j=t+1}^{t+s} \mu(e_{j}) \mu(e'_{j}) \alpha((\lambda_i)_{i \in [t]}, (d_jm_j^2 \epsilon_{j} \epsilon_{j,1}^{-z_{j,1}} \epsilon_{j,2}^{-z_{j,2}} d_j^{-1} - z_{j,3} m_j^{-2} z_{j,4}) \prod_{i=t}^{t+s} \lambda_i)
\]

where the sum over \( v \) is as usual over vectors \((v_{j,x} \mid u_j \) and \( v_{j,x} \) satisfies the same condition on its prime factors as \( x \) (all in \( P_j \) for \( d \) and \( e \), all in \( Q_j \) for \( m \)).

Next we claim we can with good precision swap again a sum and a product signs.
Claim 4. The following equality holds, for any choice of the family \( \xi_{j,k} \) in \( I = [-\sqrt{\log R}, \sqrt{\log R}] \).

\[
\sum_{d \text{ m.e.} \ell} \prod_{j=t+1}^{t+s} \frac{\mu(e_j)\mu(e'_j)}{e_j} e_j^{t-z_{j,1}} e_j^{r-z_{j,2}} d_{j}^{1-z_{j,3}} m_{j}^{2-z_{j,4}} \prod_{i=1}^{t} \frac{\mu(l_i)\mu(l'_i)}{\lambda_i} l_i^{-z_{i,1}} l'_i^{-z_{i,2}} = (1 + o(w(N)^{-1/2})) \prod_{j=t+1}^{t+s} \sum_{d_j, m_j, e_j, e'_j} \frac{\mu(e_j)\mu(e'_j)}{e_j} e_j^{t-z_{j,1}} e_j^{r-z_{j,2}} d_{j}^{1-z_{j,3}} m_{j}^{2-z_{j,4}} \prod_{i=1}^{t} \sum_{l_i, l'_i} \frac{\mu(l_i)\mu(l'_i)}{\lambda_i} l_i^{-z_{i,1}} l'_i^{-z_{i,2}}.
\]

Proof. The justification is exactly the same as for Claim 3, because the claim consists simply in replacing back the dashed sum by a complete sum, at the same small cost. \(\blacksquare\)

So the integral can be rewritten as

\[
\int_{I^{4+2t}} \left( \prod_{j=t+1}^{t+s} \sum_{u_j} \frac{2^{s_j}}{u_j} (1 + o(w^{-1/2})) \sum_{d_j, m_j, e_j, e'_j} \frac{\mu(e_j)\mu(e'_j)}{e_j} e_j^{t-z_{j,1}} e_j^{r-z_{j,2}} d_{j}^{1-z_{j,3}} m_{j}^{2-z_{j,4}} \times \sum_{v} \mu(v_j, e)\mu(v_{j', e'}) v_j^{-z_{j,1}} v_{j'}^{-z_{j,2}} v_{j, d}^{-z_{j,3}} v_{j, m}^{-z_{j,4}} \prod_{i=1}^{t} \sum_{l_i, l'_i} \frac{\mu(l_i)\mu(l'_i)}{\lambda_i} l_i^{-z_{i,1}} l'_i^{-z_{i,2}} \right) \prod_{j,k} \theta(\xi_{j,k}) d\xi_{j,k}
\]

Now we claim that the effect of the error term arising from the \( o(w^{-1/2}) \) is indeed negligible. This is because on the on the one hand

\[
\left| \sum_{v} \mu(v_j, e)\mu(v_{j', e'}) v_j^{-z_{j,1}} v_{j'}^{-z_{j,2}} v_{j, d}^{-z_{j,3}} v_{j, m}^{-z_{j,4}} \right| \leq \tau(u_j)^4
\]

and

\[
\sum_{s_{i+1}, \ldots, s_{t'}} \prod_{i=t+1}^{t+s} \sum_{u_j \in U(i_j, s_j)} \frac{2^{s_j}}{u_j} \tau(u_j)^4 = O(1)
\]

thanks again to the proof of Proposition 4.2 of [13], already mentioned between the Claims 2 and 3, and on the other hand we have the following Claim.

Claim 5. The following bound holds, where the integral is over \( I^{4+2t} \).

\[
\int \left| \prod_{i \in [t]} \sum_{l_i, l'_i} \frac{\mu(l_i)\mu(l'_i)}{\lambda_i} l_i^{-z_{i,1}} l'_i^{-z_{i,2}} \times \prod_{j>t} \sum_{d_j, m_j, e_j, e'_j} \frac{\mu(e_j)\mu(e'_j)}{e_j} e_j^{t-z_{j,1}} e_j^{r-z_{j,2}} d_{j}^{1-z_{j,3}} m_{j}^{2-z_{j,4}} \prod_{j,k} \theta(\xi_{j,k}) d\xi_{j,k} = O(1/(\log R)^t). \right. \quad (24)
\]

These two bounds, together with the \( (\log R)^t \) factor at the front and the \( o(w(N)^{-1/2}) \) yield assuredly an \( o(1) \).

Proof (of Claim 5). To prove this claim, we first replace the sum over \( l_i, l'_i \), for any \( i \in [t] \), by a product over primes, by multiplicativity

\[
\sum_{l_i, l'_i} \frac{\mu(l_i)\mu(l'_i)}{\lambda_i} l_i^{-z_{i,1}} l'_i^{-z_{i,2}} = \prod_{s \in \mathcal{P}} (1 - s^{-1-z_{j,1}} - s^{-1-z_{j,2}} + s^{-1-z_{j,1}-z_{j,2}}).
\]
Then we notice that for large primes \( s \), and complex numbers \( z = z_{j,1}, z' = z_{j,2} \) of real part greater than 0

\[
1 - s^{-1-z} - s^{-1-z'} + s^{-1-z-z'} = \frac{(1 - s^{-1-z})(1 - s^{-1-z'})}{1 - s^{-1-z-z'}} + O(s^{-2})
\]
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 \( \zeta \) function is defined for \( \Re s > 1 \) by

\[
\zeta(s) = \sum_{n \geq 1} n^{-s} = \prod_p (1 - p^{-s})^{-1}
\]

and satisfies for \( s \) near 1

\[
\zeta(s) = \frac{1}{s-1} + O(1).
\]

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 for any \( j \in [t] \) and \( \xi_{j,k} \in I \) (for \( k = 1, 2 \)) and the corresponding \( z_{j,k} \)

\[
\prod_{s \in \mathcal{P}} (1 - s^{-1-z_{j,1}} - s^{-1-z_{j,2}} + s^{-1-z_{j,1}-z_{j,2}}) \ll \frac{z_{j,1}z_{j,2}}{z_{j,1} + z_{j,2}}.
\] (25)

Similarly, for any \( j \in \{t + 1, \cdots, t + s\} \)

\[
\sum_{d_j, m_j, e_j, e'_j} \frac{\mu(e_j)\mu(e'_j)}{e_j} e_j^{-z_{j,1}} e'_j^{-z_{j,2}} d_j^{-1-z_{j,3}} m_j^{-2-z_{j,4}} = \prod_{q \in \mathcal{Q}_j} (1 - q^{-1-z_{j,1}} - q^{-1-z_{j,2}} + q^{-1-z_{j,1}-z_{j,2}})
\] (26)

\[
\prod_{r \in \mathcal{Q}_j} (1 - r^{-2-z_{j,4}})^{-1} \prod_{p \in \mathcal{P}_j} (1 - p^{-1-z_{j,3}})^{-1}.
\] (27)

Notice that the product in \( r \) is a convergent product and it is bounded by a constant when \( z_{j,4} \) varies in the permitted range. Moreover, because of (21), the product of the theta factors is bounded by

\[
O_A \left( \prod_{j,k} (1 + |\xi_{j,k}|)^{-A} \right).
\]

Given that \( \mathcal{P}_j \) and \( \mathcal{Q}_j \) have each density \( 1/2 \) among the primes, we can write that

\[
\sum_{q \in \mathcal{P}_j} q^{-1-z} = \frac{1}{2} \log \frac{1}{z} + O(1)
\]
for \( Rz > 0 \). This provides again a bound for the product (26), similar to the one in (25) (but with exponents 1/2). Now we use \( z_{j,k} = (1 + \xi_{j,k})(\log R)^{-1} \) and the triangle inequality. Finally the integrand in (24) is bounded by

\[
\prod_{i=1}^{t} \left| z_{i,2} \right| \left| z_{i,1} \right| \left| z_{i,1} + z_{i,2} \right|^{-1} \prod_{j=t+1}^{t+s} \left| z_{j,1} \right|^{1/2} \left| z_{j,2} \right|^{1/2} \left| z_{j,1} + z_{j,2} \right|^{-1/2} \left| z_{j,3} \right|^{-1/2} \prod_{j,k} (1 + |\xi_{j,k}|)^{-A}
\]

\[
\ll (\log R)^{-t} \prod_{i=1}^{t} (1 + |\xi_{i,1}|) (1 + |\xi_{i,2}|) \prod_{k=1,2} \prod_{j=t+1}^{t+s} (1 + |\xi_{j,k}|)^{1/2} \prod_{k \in [4]} (1 + |\xi_{j,k}|)^{-A}
\]

\[
\ll (\log R)^{-t} \prod_{j,k} (1 + |\xi_{j,k}|)^{-A/2}
\]

when \( A \) is large enough (for the last step). This last product is certainly integrable as soon as \( A > 0 \), so the final expression is as expected \( O((\log R)^{-t}) \) as claimed.

Thus for fixed vectors \( i,s \), the integral from (24) is equal, up to an error term \( E_{i,s} \), satisfying \( (\log R)^t \sum_{i,s} E_{i,s} = o(1) \), to the following integral

\[
\int_{I^{s+2t}} \sum_{u,v,d,e,m} \prod_{i=1}^{t} \frac{\mu(l_i)\mu(l'_i)}{\lambda_i} l_i^{-z_{i,1}} l'_i^{-z_{i,2}} \prod_{k=1,2} \theta(\xi_{i,k}) d\xi_{i,k}
\]

\[
\prod_{j=t+1}^{t+s} \frac{2^{s_j} \tau(u_j) \mu(e_j v_{j,e}) \mu(e'_j v_{j,e'})}{\epsilon_j} e_j^{-z_{j,1}} e'_j^{-z_{j,2}} \epsilon_j^{-z_{j,2}} d_j^{-1-z_{j,3}} m_j^{-2-z_{j,4}} v_{j,e}^{-z_{j,1}} v_{j,e'}^{-z_{j,2}} v_{j,d}^{-z_{j,3}} v_{j,m}^{-z_{j,4}} \prod_{k \in [4]} \theta(\xi_{j,k}) d\xi_{j,k}.
\]

(28)

We can now swap again the summation and integration using Fubini’s theorem: this is because the sum can be shown to be absolutely convergent (by basically the same argument as the one used to prove Claim 5) and the interval is compact. But then using the separation of the variables, one can even prove that (28) is equal to

\[
\sum_{u,v,d,e,m} \prod_{i=1}^{t} \frac{\mu(l_i)\mu(l'_i)}{\lambda_i} \prod_{k=1,2} l_i^{-z_{i,k}} \theta(\xi_{i,k}) d\xi_{i,k} \prod_{j=t+1}^{t+s} \frac{2^{s_j} \mu(e_j) \mu(e'_j)}{\epsilon_j} e_j^{-z_{j,1}} e'_j^{-z_{j,2}} \epsilon_j^{-z_{j,2}} \prod \theta(\xi_{j,k}) d\xi_{j,k}.
\]

(29)

It is now time to undo the truncation to \( I \) in these integrals, in order to be able to collapse them into \( \chi \) factors. The error term provoked by the removal of this truncation is the same as the one provoked earlier by the truncation, so it will fit, all summation done, into the \( o(1) \) of (12). So up to an error term \( E_{i,s}^{(3)} \) satisfying again \( (\log R)^t \sum_{i,s} E_{i,s}^{(3)} = o(1) \),

---

\(^5\)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.
the expression (28) is equal to
\[
\sum_{u,v,d,e,m} \prod_{i=1}^{t} \frac{\mu(l_{i,1})\mu(l_{i,2})}{\lambda_i} \prod_{k=1,2} \chi \left( \frac{\log l_{i,k}}{\log R} \right) \\
\prod_{j=t+1}^{t+s} \frac{2^{s_j}(u_j) \mu(e_jv_{j,e})\mu(e'_jv_{j,e'})}{u_j d_j m_j^2 \epsilon_j} \chi \left( \frac{\log d_j v_{j,d}}{\log R} \right) \chi \left( \frac{\log m_j v_{j,m}}{\log R} \right) \prod_{k=1,2} \chi \left( \frac{\log e_{j,k} v_{j,e,e'}}{\log R} \right).
\]

So this is what we get for (12), up to error terms of the desired form \(O_D \left( \frac{N^d-1+O_D(\gamma)}{\text{Vol}(K)} \right)\) in Claims 1 and 2, various \(o(1)\) throughout the proof:
\[
\prod_{j=t+1}^{t+s} C_{D_j,\gamma}^{-1} \sum_{s_j,i,j,u_j,v_j,d_j,m_j,e_j,e'_j} \frac{2^{s_j}(u_j) \mu(e_jv_{j,e})\mu(e'_jv_{j,e'})}{u_j d_j m_j^2 \epsilon_j} \chi \left( \frac{\log d_j v_{j,d}}{\log R} \right) \chi \left( \frac{\log m_j v_{j,m}}{\log R} \right) \prod_{x \in \{d,m,e,e'\}} \chi \left( \frac{\log x v_{j,x}}{\log R} \right) \\
\times \prod_{i \in [t]} \left( \frac{\phi(W)}{W} \sum_{l_i,l'_i} \frac{\mu(l_i)\mu(l'_i)}{\lambda} \prod_{x \in \{l_i,l'_i\}} \chi \left( \frac{\log x}{\log R} \right) \right)
\]

which is an expression which does not depend on the system \(\Phi\) anymore. Hence the \(j\)-th factor, for \(j = 1, \ldots, t+s\), is also the main term of the average of the \(j\)-th pseudorandom majorant on the trivial system \(\Phi : \mathbb{Z} \to \mathbb{Z}, n \mapsto n\). Now because of the properties of the Green-Tao majorant and of the Matthiesen majorant, from subsections respectively 4.2.1 and 4.2.2, these averages are \(1+o(1)\), whence the result.

\[\text{C Volume packing arguments and local divisor density}\]

In this appendix, we shall collect some statements frequently used concerning the number of solutions to a system of linear equations in a convex set and in \((\mathbb{Z}/m\mathbb{Z})^d\). We first recall a lemma already stated earlier but particularly relevant here, originating from [9], Appendix A.

\textbf{Lemma C.1.} Let \(K \subset [0,N]^d\) be a convex body of \(\mathbb{R}^d\). Then
\[
|K \cap \mathbb{Z}^d| = \sum_{n \in K \cap \mathbb{Z}^d} = \text{Vol}(K) + O_d(N^{d-1}).
\]

We recall the definition of local divisor density and we provide some useful properties.

\textbf{Definition C.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 densities by
\[
\alpha_{\Psi}(d_1, \ldots, d_t) = \mathbb{E}_{n \in (\mathbb{Z}/m\mathbb{Z})^d} \prod_{i=1}^{t} 1_{\psi_i(n) \equiv 0 \mod d_i}.
\]
Lemma C.2. Let $K \subset [-B, B]^d$ be a convex body and $\Psi$ a system of $t$ linear forms, and $d_1, \ldots, d_t$ of lcm $m$. Then
\[ \sum_{n \in \mathbb{Z}^d \cap K} \prod_{1 \leq i \leq t} 1_{d_i | \psi(n)} = \text{Vol}(K) \alpha(d_1, \ldots, d_t) + O(B^{d-1}m). \]

Then we try to bound $\alpha_\Psi(p^{a_1}, \ldots, p^{a_t})$. To this aim, we state a version of Hensel’s lemma in several variables.

Lemma C.3. Let $Q \in \mathbb{Z}[X_1, \ldots, X_d]$, $p$ be a prime and $k \geq 1$ an integer and $x \in (\mathbb{Z}/p^k \mathbb{Z})^d$ such that $Q(x) \equiv 0 \pmod{p^k}$ and $\nabla Q(x) \neq 0 \pmod{p}$. Then there exists precisely $p^{d-1}$ vectors $y \in (\mathbb{Z}/p^k \mathbb{Z})^d$ such that $x \equiv y \pmod{p^k}$ and $Q(y) \equiv 0 \pmod{p^{k+1}}$.

Proof. Let $y \in (\mathbb{Z}/p^k \mathbb{Z})^d$ satisfy $x \equiv y \pmod{p^k}$; in other words $y = x + p^k z$ for some uniquely determined $z \in (\mathbb{Z}/p \mathbb{Z})^d$. Here we treat $x \in (\mathbb{Z}/p^k \mathbb{Z})^d$ as an element of $(\mathbb{Z}/p^{k+1} \mathbb{Z})^d$ by using the canonical injection. We then treat $Q(x)$ as an element of $\mathbb{Z}/p^{k+1} \mathbb{Z}$ congruent to 0 mod $p^k$ and we put $Q(x) = p^k a$ with $a \in \mathbb{Z}/p \mathbb{Z}$. Then Taylor’s formula ensures that
\[ Q(y) = Q(x) + p^k \nabla Q(x) \cdot z \pmod{p^{k+1}} = p^k (a + \overline{\nabla Q(x)} \cdot z). \]

So $Q(y) \equiv 0 \pmod{p^{k+1}}$ is equivalent to $a + \overline{\nabla Q(x)} \cdot z \equiv 0 \pmod{p}$. As $\nabla Q(x)$ is not zero modulo $p$, this imposes a non-trivial affine equation on $z$ in the vector space $\mathbb{F}_p^n$, 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 C.4. Let $\psi$ be an affine-linear form in $d$ variables, and $p$ a prime such that $p \nmid \psi$. Then for any $m \geq 1$
\[ \alpha = \alpha_\psi(p^m) = E_{n \in (\mathbb{Z}/p^m \mathbb{Z})^d} 1_{p^m | \psi(n)} = \mathbb{P}_{n \in (\mathbb{Z}/p^m \mathbb{Z})^2} (p^m | \psi(n)) \leq p^{-m}. \]

Proof. If $n \in (\mathbb{Z}/p^m \mathbb{Z})^d$ satisfies $\psi(n) \equiv 0 \pmod{p^m}$, then in particular $\tilde{\psi}(n) \equiv 0 \pmod{p}$, where $\tilde{\psi}$ is the reduction modulo $p$, which imposes that $\tilde{n}$ lies in ker $\tilde{\psi}$ (our use of the notation ker may be slightly improper, because $\tilde{\psi}$ is an affine form, and not a linear one, on the space $\mathbb{F}_p^n$). By hypothesis, $\psi \neq 0$. 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. So let us prove the proposition by induction on $m$.

1. For $m = 1$, we have just proved the result.

2. Suppose $\alpha_m \leq p^{-m}$ for some $m \geq 1$. Because of the assumption above, $\nabla \psi$ is a constant vector which is nonzero modulo $p$. Applying Lemma C.3 for $k = m$, we find that each zero modulo $p^m$ of $\psi$ gives rise to exactly $p^{d-1}$ zeros modulo $p^{m+1}$, which proves that $\alpha_{m+1} \leq p^{-(m+1)}$.

Exploiting this corollary, we can now show a bound on more general local densities.
Proposition C.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

$$\alpha := \alpha_\Psi(p^{a_1}, \ldots, p^{a_t}) \leq p^{-\max_i \phi_i(a_i + a_j)}.$$ 

Proof. If all $a_i$ are zeros, the result is trivial, so let $m$ be the maximum of the $a_i$ and suppose $m \geq 1$; let $i < j$ 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 $\psi_k(n) \equiv 0 \mod p^{a_k}$ for all $k = 1, \ldots, t$, we must have in particular $\psi_j(n) \equiv 0 \mod p^{a_j}$, and using Corollary C.4, we find that

$$\alpha = E_{n \in (\mathbb{Z}/p^m \mathbb{Z})^d} \prod_{i \in [t]} 1_{p^{\alpha_i} | \phi_i(n)} \leq E_{n \in (\mathbb{Z}/p^{a_j} \mathbb{Z})^d} 1_{p^{\alpha_j} | \phi_j(n)} = p^{-a_j} = p^{-\max_i \phi_i(a_i + a_j)}.$$ 

Now suppose both $a_i$ and $a_j$ are non-zero. Without loss of generality, $1 \leq a_i \leq a_j$. Then for $n \in (\mathbb{Z}/p^m \mathbb{Z})^d$ to satisfy $\psi_k(n) \equiv 0 \mod p^{a_k}$ for all $k = 1, \ldots, t$, we must have in particular $\psi_j(n) \equiv \tilde{\psi}_j(n) \equiv 0 \mod p$ We can again suppose that $\tilde{\psi}_j$ as well as $\tilde{\psi}_i$ have linear parts which are not modulo $p$, otherwise $\alpha = 0$. This imposes that $\tilde{n}$ lies in the intersection of two affine $\mathbb{F}_p$-hyiperplanes, namely ker $\tilde{\psi}_i$ and ker $\tilde{\psi}_j$ which are distinct, because these forms are affinely independent modulo $p$. This intersection is empty (and then $\alpha = 0$) if and only if these hyperplanes are parallel (hence the linear parts of $\tilde{\psi}_j$ and $\tilde{\psi}_i$ are proportional modulo $p$). So let us suppose that the linear parts are not proportional modulo $p$, which amounts to saying that the constant vectors $\overrightarrow{\tilde{\psi}_j}, \overrightarrow{\tilde{\psi}_i} \in (\mathbb{Z}/p^2 \mathbb{Z})^d$ are not proportional. Now we use induction on $m \geq 1$ to show that

$$\beta_m = E_{n \in (\mathbb{Z}/p^m \mathbb{Z})^d} (\psi_i(n) \equiv \psi_j(n) \equiv 0 \mod p^m) \leq p^{-2m}.$$ 

1. For $m = 1$, what we have seen above implies that $\beta_1 = 0$ or $\beta_1 = p^{-2}$ (the intersection of two non-parallel 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.

2. Suppose that for some $m \geq 1$ we have $\beta_m \leq p^{-2m}$. Now if $x \in (\mathbb{Z}/p^m \mathbb{Z})^d$ satisfies $\psi_i(x) \equiv \psi_j(x) \equiv 0 \mod p^m$ (there are $\beta_m$ such $x$) 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 $\psi_i(y) \equiv \psi_j(y) \equiv 0 \mod p^{m+1}$, then following the proof of Lemma C.3, we infer that $z$ has to satisfy two affine equations

$$\begin{align*}
\begin{aligned}
\alpha + \overrightarrow{\psi_i} \cdot z &\equiv 0 \mod p \\
\beta + \overrightarrow{\psi_j} \cdot z &\equiv 0 \mod p,
\end{aligned}
\end{align*}$$

which forces $z$ to be in the intersection of two non-parallel affine $\mathbb{F}_p$-hyperplanes of $\mathbb{F}_p^d$ (they are non parallel because we supposed that the gradients were not proportional). Hence for a fixed $x$ as above, there are $p^{d-2}$ such $y$, so in total $p^{d-2} \beta_m$ such $y$, and finally $\beta_{m+1} = p^{d-2} \beta_m$ whence the conclusion.

In particular, putting $m = a_i$, we have that $E_{n \in (\mathbb{Z}/p^{a_i} \mathbb{Z})^d} 1_{\phi_i(n) \equiv \phi_j(n) \equiv 0 \mod p^{a_i}} \leq p^{-2a_i}$. There remains to induct on $a_j - a_i \geq 0$ using Lemma C.3 in order to find that

$$E_{n \in (\mathbb{Z}/p^{a_j} \mathbb{Z})^d} 1_{p^{a_i} | \phi_i(n)} 1_{p^{a_j} | \phi_j(n)} \leq p^{-\max_i \phi_i(a_i + a_j)}$$

which implies the desired result.  


We prove another statement which is helpful during the proof of the linear forms conditions (Appendix B).

**Proposition C.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

$$\sum_{n \in K \cap \mathbb{Z}^d} 1_{p^2 \mid \prod_{i \in [t]} \phi_i(n)} \ll p^{-2} \text{Vol}(K) + B^{d-1} p^2.$$

**Proof.** First, we remark 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

$$\sum_{n \in K \cap \mathbb{Z}^d} 1_{p^2 \mid \prod_{i \in [t]} \phi_i(n)} \leq \sum_{i \in [t]} \sum_{n \in K \cap \mathbb{Z}^d} 1_{p^2 \mid \phi_i(n)} + \sum_{i \neq j} \sum_{n \in K \cap \mathbb{Z}^d} 1_{\phi_i(n) \equiv \phi_j(n) \equiv 0 \mod p}.$$

Now for any $i \in [t]$ we apply Lemma C.2 which provides

$$\sum_{n \in K \cap \mathbb{Z}^d} 1_{p^2 \mid \phi_i(n)} = \text{Vol}(K) \alpha_{\phi_i}(p^2) + O(B^{d-1} p^2)$$

and for any $i \neq j$

$$\sum_{n \in K \cap \mathbb{Z}^d} 1_{\phi_i(n) \equiv \phi_j(n) \equiv 0 \mod p} = \text{Vol}(K) \alpha_{\phi_i, \phi_j}(p, p) + O(B^{d-1} p).$$

But the hypothesis of finite complexity modulo $p$ means that we may invoke Proposition C.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.

**References**


40


