LINEAR INEQUALITIES IN PRIMES

By

ALED WALKER

Abstract. In this paper we prove an asymptotic formula for the number of solutions in prime numbers to systems of simultaneous linear inequalities with algebraic coefficients. For m simultaneous inequalities we require at least \( m + 2 \) variables, improving upon existing methods, which generically require at least \( 2m + 1 \) variables. Our result also generalises the theorem of Green–Tao–Ziegler on linear equations in primes. Many of the methods presented apply for arbitrary coefficients, not just for algebraic coefficients, and we formulate a conjecture concerning the pseudorandomness of sieve weights which, if resolved, would remove the algebraicity assumption entirely.

Contents

1 Introduction 31
2 The structure of the argument 41

Part 1 Preliminaries 43
3 Smooth functions 43
4 Notation and conventions 47

Part 2 Linear algebra 49
5 Dimension reduction 50
6 Normal form 55
Part 3  Pseudorandomness 56
7 The $W$-trick and Gowers norms 56
8 Inequalities in lattices 61
9 The linear inequalities condition 67

Part 4  The structure of inequalities 80
10 An alternative formulation 80
11 Variation in parameters 86

Part 5  The main argument 94
12 Controlling by Gowers norms 94
13 Transferring from $\mathbb{Z}$ to $\mathbb{R}$ 96
14 Parametrising the kernel 98
15 Gowers–Cauchy–Schwarz argument 102
16 Combining the lemmas 112

Part 6  Final deductions 113
17 Removing Lipschitz cut-offs 113

Part 7  Appendices 116
Appendix A Estimating integrals 116
Appendix B An analytic argument 121
References 126
1 Introduction

Fourier analysis is a vital tool in the study of diophantine problems. In recent years, however, new tools have been developed which can prove asymptotic formulae for the number of solutions to certain systems even when the Fourier-analytic approach is not known to succeed. In particular, in [13] Green and Tao established an asymptotic formula for the number of prime solutions to generic systems of \( m \) simultaneous linear equations in at least \( m+2 \) variables. Their result was conditional on various conjectures, but these conjectures were later proved by the same authors and Ziegler, in the series of papers [14], [15] and [16].

**Theorem 1.1** ([13, Theorem 1.8], Green–Tao–Ziegler). Let \( N, m, \) and \( d \) be natural numbers, with \( d \geq m+2 \), and let \( C \) be a positive constant. Let \( L = (\lambda_{ij})_{i \leq m,j \leq d} \) be an \( m \)-by-\( d \) matrix with integer coefficients, with rank \( m \), and assume the non-degeneracy condition that the only element of the row-space of \( L \) over \( \mathbb{Q} \) with two or fewer non-zero entries is the zero vector. Let \( b \in \mathbb{Z}^m \), and suppose that \( \|b\|_\infty \leq CN \) and that \( |\lambda_{ij}| \leq C \) for all \( i \) and \( j \). Let \( K \subset [-N,N]^d \) be a convex set. Then

\[
|\{ p \in K : Lp = b \}| = \left( \alpha_\infty \prod_p \alpha_p \right) (\log N)^{-d} + o_{C,d,m}(N^{d-m}(\log N)^{-d}),
\]

where the local densities \( \alpha_p \) are given, for each prime \( p \), by

\[
\alpha_p := \lim_{M \to \infty} \frac{1}{(2M)^d} \sum_{n \in [-M,M]^d \atop (n,p) = 1} \left( 1 + \frac{1}{p-1} \right)^d
\]

and the global factor \( \alpha_\infty \) is given by

\[
\alpha_\infty := |\{ n \in \mathbb{Z}^d : n \in K, Ln = b, n_i \geq 0 \text{ for all } i \}|.
\]

Here and throughout, \( p \) denotes a prime, \( \mathbf{p} \) denotes a vector all of whose coordinates are prime, and \( \mathbf{n} \) denotes a vector all of whose coordinates are integers \( n_i \). The expression \( (n_i, p) \) denotes the greatest common divisor of \( n_i \) and \( p \).

To give a concrete example to which this result may be applied, by considering

\[
L = \begin{pmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{pmatrix}, \quad b = 0
\]

one may deduce an asymptotic formula for the number of four-term arithmetic progressions of primes that are less than \( N \).
For \( m \geq 2 \), Theorem 1.1 is stronger than any similar statement that may be proved using the Fourier transform alone. Indeed, notwithstanding Balog’s example [2, Corollary 3] of a certain non-generic class of \( m \) equations in \( m + \lceil \sqrt{2m} \rceil \) prime variables, generically the Fourier transform approach needs at least \( 2m + 1 \) prime variables in order to succeed. The proof of Theorem 1.1 rests on many creative innovations, in particular the authors’ use of Gowers norms and their inverse theory, which is a subject that is now referred to as ‘higher order Fourier analysis’. The object of the present paper is to use certain aspects of this machinery to establish, in a related setting, an analogous reduction in the number of variables that are required to prove an asymptotic formula.

We will be concerned with diophantine inequalities, a topic that we first considered in [21]. Before giving our first main result (Theorem 1.7) let us briefly review some previous results concerning diophantine inequalities in the primes. Consider the following classical theorem of Baker.

**Theorem 1.2** ([1], Baker). Let \( \varepsilon > 0 \), and let \( \lambda_1, \lambda_2, \lambda_3 \in \mathbb{R} \setminus \{0\} \) be three non-zero reals that are not all of the same sign. Furthermore, suppose that for all \( \alpha \in \mathbb{R} \setminus \{0\} \) the relation \( (a\lambda_1, a\lambda_2, a\lambda_3) \notin \mathbb{Z}^3 \) holds. Then there exist infinitely many triples of primes \( (p_1, p_2, p_3) \) satisfying

\[
|\lambda_1 p_1 + \lambda_2 p_2 + \lambda_3 p_3| \leq \varepsilon.
\]

**Remark 1.3.** The condition concerning the signs of \( \lambda_1, \lambda_2, \lambda_3 \) is clearly a necessary one, as otherwise there exist only finitely many solutions to (1.2) in the positive integers (and so certainly there exist only finitely many solutions in the primes). Regarding the other condition, the conclusion of Theorem 1.2 may hold even if there exists some \( \alpha \in \mathbb{R} \setminus \{0\} \) for which

\[
(a\lambda_1, a\lambda_2, a\lambda_3) = (q_1, q_2, q_3) \in \mathbb{Z}^3.
\]

But then one is required to solve

\[
|q_1 p_1 + q_2 p_2 + q_3 p_3| \leq \varepsilon \alpha,
\]

which, if \( \varepsilon \) is small enough, is equivalent to solving

\[
q_1 p_1 + q_2 p_2 + q_3 p_3 = 0.
\]

Theorem 1.1 can then affirm that there are infinitely many solutions, provided that \( q_1, q_2, \) and \( q_3 \) satisfy certain local properties. This issue, of when an inequality can

\[1\]In fact Baker proved a slightly different result, writing in the cited paper that the result we quote here followed easily from the then existing methods. Vaughan proved a similar result in [20].
encode a certain equation with rational coefficients, will be an important theme of the paper.

The classical approach to proving results such as Theorem 1.2 involves Fourier analysis over \( \mathbb{R} \), after having replaced the characteristic function of the interval \( [-\varepsilon, \varepsilon] \) with a smoother cut-off function. This approach is known as the Davenport–Heilbronn method, it having originated in a paper [5] of those two authors. For a variety of technical reasons this method was, until relatively recently, unable to give an asymptotic formula for the number of solutions to (1.2) that satisfied \( 1 \leq p_1, p_2, p_3 \leq N \), or even give a lower bound of the expected order of magnitude (at least for arbitrary \( N \)). However, certain advances of Freeman [6, 7] enabled Parsell to achieve the second of these two goals.

**Theorem 1.4** ([18, Theorem 1], Parsell). *Let \( \varepsilon > 0 \), and let \( \lambda_1, \lambda_2, \lambda_3 \in \mathbb{R} \setminus \{0\} \) be three non-zero reals that are not all of the same sign. Furthermore, suppose that for all \( \alpha \in \mathbb{R} \setminus \{0\} \) the relation \( (\alpha \lambda_1, \alpha \lambda_2, \alpha \lambda_3) \notin \mathbb{Z}^3 \) holds. Then the number of prime triples \( (p_1, p_2, p_3) \) satisfying \( 1 \leq p_1, p_2, p_3 \leq N \) and

\[
|\lambda_1 p_1 + \lambda_2 p_2 + \lambda_3 p_3| \leq \varepsilon
\]

is \( \Omega_{\lambda_1, \lambda_2, \lambda_3}(\varepsilon N^2 (\log N)^{-3}) \).

Since [18] was published, it has been understood that a very minor modification to Parsell’s analytic method can be used to obtain an asymptotic expression for the number of solutions to (1.3), namely

\[
C_{\lambda_1, \lambda_2, \lambda_3} \varepsilon N^2 \log^{-3} N + o_{\lambda_1, \lambda_2, \lambda_3}(N^2 \log^{-3} N),
\]

for some positive constant \( C_{\lambda_1, \lambda_2, \lambda_3} \). Furthermore, in the case of \( m \) simultaneous (rationally independent) inequalities of the form (1.3), Parsell’s method can calculate an asymptotic formula for the number of solutions in primes provided the number of variables is at least \( 2m + 1 \). In Appendix B we take the opportunity to record the details of both the statement and the proof of this result.

In the main theorems of this paper (Theorem 1.7 and Theorem 1.16) we specialise to the case of algebraic coefficients and reduce the number of variables that are required from \( 2m + 1 \) to \( m + 2 \). Our first result does not concern the most general type of diophantine inequality, but nonetheless it enjoys several applications. To state it, we recall the notion of the dual degeneracy variety, which we defined in Definition 2.3 of [21] in order to manipulate the non-degeneracy conditions more succinctly.
Definition 1.5 (Dual degeneracy variety, [21]). Let \( m, d \) be natural numbers satisfying \( d \geq m + 2 \). Let \( V^\ast_{\text{degen}}(m, d) \) denote the set of all \( m \)-by-\( d \) matrices with real coefficients that contain a non-zero row-vector in their row-space over \( \mathbb{R} \) that has two or fewer non-zero coordinates. We call \( V^\ast_{\text{degen}}(m, d) \) the dual degeneracy variety.

For example, the matrix
\[
\begin{pmatrix}
1 & -2 & 1 & 0 \\
2 & -4 & 0 & \sqrt{3}
\end{pmatrix}
\]
is in \( V^\ast_{\text{degen}}(2, 4) \), since the vector \((0, 0, -2, \sqrt{3})\) lies in its row space. As is explained at length in [21], if one wishes to count solutions to an inequality given by \( L \) using a method involving Gowers norms, then one can only possibly succeed if \( L \not\in V^\ast_{\text{degen}}(m, d) \). Returning to Theorem 1.1, we observe that the non-degeneracy condition in the statement of that theorem is exactly the condition that \( L \not\in V^\ast_{\text{degen}}(m, d) \). If \( d = m + 2 \), non-degeneracy in this sense is easy to detect. Indeed, \( L \not\in V^\ast_{\text{degen}}(m, d) \) if and only if the determinants of all the \( m \)-by-\( m \) submatrices of \( L \) are non-vanishing.

Remark 1.6. The above notion is ‘dual’ to the notion of finite Cauchy–Schwarz complexity (see Definition 5.5), in the sense that \( L \) is in the dual degeneracy variety if and only if \( \ker L \) may be parametrised by a system of linear forms with finite Cauchy–Schwarz complexity. In [21] we also introduced a degeneracy variety in order to manipulate quantitative versions of this fact, but this will not be necessary here. For more on these issues, we invite the reader to consult Sections 4 and 7 of [21].

We are now ready to state our first main result. In the statement below, \([N]\) refers to the set \( \mathbb{N} \cap [1, N] \) and the function \( 1_{[-\epsilon, \epsilon]^m} \) refers to the indicator function of the set \([-\epsilon, \epsilon]^m\).

Theorem 1.7 (Main theorem, purely irrational version). Let \( N, m, d \) be natural numbers, with \( d \geq m + 2 \), and let \( C, \epsilon \) be positive constants. Let \( L \) be an \( m \)-by-\( d \) real matrix with algebraic coefficients and rank \( m \). Suppose that \( L \not\in V^\ast_{\text{degen}}(m, d) \). Suppose further that for all \( \alpha \in \mathbb{R}^m \setminus \{0\} \) one has \( L^T \alpha \notin \mathbb{Z}^d \), i.e., suppose that \( L \) is purely irrational in the sense of Definition 2.2 of [21]. Let \( v \in \mathbb{R}^m \) be any vector satisfying \( \|v\|_\infty \leq CN \). Then
\[
\sum_{p \in [N]^d} 1_{[-\epsilon, \epsilon]^m}(Lp + v) = \frac{1}{\log^d N} \int_{x \in [0, N]^d} 1_{[-\epsilon, \epsilon]^m}(Lx + v) \, dx + o_{C,L,\epsilon}(N^{d-m}(\log N)^{-d})
\]
as \( N \to \infty \).
Remark 1.8. One notes that in the asymptotic formula (1.4) there is not a contribution from any non-archimedean local factors. In Theorem 1.16 below, we will remove the supposition that there does not exist any non-zero vector $\alpha \in \mathbb{R}^m \setminus \{0\}$ for which $L^T \alpha \in \mathbb{Z}^d$. Once these potential rational relations are permitted, one does indeed observe a contribution from local factors.

Remark 1.9. When $v = 0$, it is straightforward to show (see Lemma A.2) that the main term in (1.4) is equal to

$$C_L e^{mN^{d-m}(\log N)^{-d}} + o_{L,v}(N^{d-m}(\log N)^{-d}),$$

where $C_L$ is a constant depending only on $L$. The positivity of $C_L$ may be determined in practice.

Remark 1.10. The reader may note that Theorem 1.7 insists upon a fixed matrix $L$, rather than a matrix $L$ with bounded coefficients (as appeared in Theorem 1.1). In our previous work [21, Theorem 2.12], performed in the context of linear inequalities weighted by bounded functions, we proved a result that enabled $L$ to vary, as long as the coefficients of $L$ were bounded and $L$ was bounded away from $V^*_\text{degen}(m,d)$. In the present paper there are many auxiliary linear equalities $L'$, which will also need to enjoy such a quantitative non-degeneracy. We found keeping track of these features throughout the whole argument to be extremely complicated, but in principle it should be possible to do so.

Remark 1.11. Theorem 1.7 strengthens Theorem B.1 of Parsell, in the sense that the number of variables has been reduced (from $2m + 1$ to $m + 2$). But unfortunately this has been achieved at the cost of imposing an algebraicity assumption on the coefficients of $L$. The situation is regrettable as, under this assumption, the classical Davenport–Heilbronn method alone is adequate to count the number of prime solutions to $m$ simultaneous linear inequalities in $2m + 1$ variables, without needing the developments of Parsell. We should stress that most of our method does not rely on the algebraicity assumption. Indeed, the conclusions of Theorems 1.7 and 1.16 do in fact hold for some explicit set of matrices $L$ that has full Lebesgue measure (see Remark 9.7). Unfortunately, owing to the intricacy of the linear-algebraic manipulations in Section 15, we have not been able to formulate a clean or enlightening characterisation of this full-measure set. We have decided to clarify the exposition of the paper by working with algebraic coefficients throughout.

Let us give a concrete example of a linear inequality to which Theorem 1.7 applies but the Davenport–Heilbronn method does not.
**Example 1.12.** Let $\varepsilon > 0$. Then the number of prime quadruples $(p_1, p_2, p_3, p_4) \in [N]^4$ satisfying

\[
\begin{align*}
|p_1 + p_3 \sqrt{2} - p_4 \sqrt{3}| & \leq \varepsilon \\
|p_2 + p_3 \sqrt{5} - p_4 \sqrt{7}| & \leq \varepsilon
\end{align*}
\]

(1.5)

is equal to $C \varepsilon^2 N^2 (\log N)^{-4} + o_\varepsilon(N^2(\log N)^{-4})$, for some positive constant $C$.

**Proof.** Taking

\[ L = \begin{pmatrix} 1 & 0 & \sqrt{2} & -\sqrt{3} \\ 0 & 1 & \sqrt{5} & -\sqrt{7} \end{pmatrix}, \]

$L$ certainly satisfies the hypotheses of Theorem 1.7, since all the 2-by-2 submatrices have non-zero determinant and surds of primes are rationally independent. Taking $v = 0$, one may therefore apply Theorem 1.7.

This yields an asymptotic expression for the number of solutions to (1.5) with the main term in the form of an integral. Since $v = 0$, by Remark 1.9 we may express the main term as $C_L \varepsilon^2 N^2 (\log N)^{-4}$ for some constant $C_L$. Explicitly, from Lemma A.2 and expression (A.6) therein,

\[
\frac{C_L}{4} = \int_{0 \leq x_1, x_2 \leq 1} 1 \, dx_1 \, dx_2,
\]

where

\[ v^{(1)} = \begin{pmatrix} -\sqrt{2} \\ -\sqrt{5} \end{pmatrix}, \quad v^{(2)} = \begin{pmatrix} \sqrt{3} \\ \sqrt{7} \end{pmatrix}. \]

By a computation, we satisfy ourselves that $C_L \approx 1.394 \ldots$ is positive. \( \square \)

Theorem 1.7 may also be used to count prime solutions to other systems.

**Corollary 1.13.** Let $(\theta_1, \ldots, \theta_d)^T = \theta \in \mathbb{R}^d$ be a real vector with algebraic coefficients. Suppose that there does not exist any $k \in \mathbb{Z}^d \setminus \{0\}$ that satisfies $k \cdot \theta \in \mathbb{Z}$. Let $\mathcal{P}$ denote the set of primes. Then

\[
\sum_{p_1, p_2 \leq N} \prod_{j=1}^{d} \mathbb{1}_{p \in [N]}(|p_1 + p_2 \theta_j|) = C_{\theta} \frac{N^2}{\log^d N} + o_{\theta}\left(\frac{N^2}{\log^d N}\right),
\]

(1.6)

for some positive constant $C_{\theta}$.

Here $\lfloor x \rfloor$ denotes the floor function of $x$, i.e., the greatest integer that is at most $x$.

**Proof.** We can expand the left-hand side of (1.6) as

\[
\sum_{p_1, p_2 \leq N} \sum_{p_{d+2} \leq N} \prod_{j=3}^{d+2} \mathbb{1}_{[0,1]}(p_1 + p_2 \theta_{j-2} - p_j).
\]
Observe that the equation \( p_1 + p_2 \theta_{j-2} - p_j = 1 \) has no solutions, since \( \theta_{j-2} \) is irrational by assumption. So the above is equal to

\[
\sum_{p_1, p_2 \leq N} \prod_{j=3}^{d+2} 1_{[0,1]}(p_1 + p_2 \theta_{j-2} - p_j),
\]

and this in turn is equal to

\[
(1.7) \sum_{p_1, p_2 \leq N} \prod_{j=3}^{d+2} 1_{[0,1]}(p_1 1 + p_2 \theta - p_3^{d+2}),
\]

where \( 1 \in \mathbb{R}^d \) is the vector with every coordinate equal to 1, and

\[ p_3^{d+2} := (p_3, \ldots, p_{d+2})^T. \]

Let \( L \) be the \( d \)-by-\((d + 2)\) matrix

\[
L = \begin{pmatrix} 1 & \theta & -I \end{pmatrix}.
\]

Then (1.7) is equal to

\[
\sum_{p \in [N]^{d+2}} 1_{[-1/2, 1/2]^d}(Lp + v),
\]

where \( v := (-1/2, \ldots, -1/2)^T \).

One sees that \( L \) satisfies the hypotheses of Theorem 1.7. Indeed, note first that if there exists some \( \alpha \in \mathbb{R}^d \setminus \{0\} \) for which \( L^T \alpha \in \mathbb{Z}^{d+2} \), then by considering the final \( d \) coordinates of \( L^T \alpha \) it follows such an \( \alpha \) must have integer coordinates. But by considering the second coordinate of \( L^T \alpha \) it follows that \( \alpha \cdot \theta \in \mathbb{Z} \), which is a contradiction to our assumptions on \( \theta \). Secondly, if \( L \) were in \( V_{\text{degen}}(d, d + 2) \) then either \( \theta_i = 0 \) for some index \( i \), or \( \theta_i = \theta_j \) for two different indices \( i \) and \( j \). Both of these possibilities are precluded by the assumptions on \( \theta \).

Therefore we may apply Theorem 1.7, and by Remark 1.9 we get a main term of the form \( C_\theta N^2 (\log N)^{-d} \). Explicitly, using Lemma A.2 and expression (A.6) as above, we have

\[
C_\theta = \int_{0 \leq x_1, x_2 \leq 1} 1_{0 \leq x_1 + \theta x_2 \leq 1 \text{ for all } i} \, dx_1 \, dx_2.
\]

For any vector \( \theta \) this integral is positive, and so the corollary is proved.

Let us now present a theorem which does not require \( L \) to be purely irrational. This is Theorem 1.16 below, and we consider it to be our main result.

For ease of notation, we introduce the following definition.
**Definition 1.14.** Let $N, m, d$ be natural numbers, and let $L : \mathbb{R}^d \to \mathbb{R}^m$ be a linear map. Let $F : \mathbb{R}^d \to \mathbb{R}$ and $G : \mathbb{R}^m \to \mathbb{R}$ be functions with compact support. Let $v \in \mathbb{R}^m$. Then, for functions $f_1, \ldots, f_d : \mathbb{Z} \to \mathbb{R}$, we define

$$T_{E,G,N}^{L,v}(f_1, \ldots, f_d) := \frac{1}{Nd^m} \sum_{n \in \mathbb{Z}^d} \left( \prod_{j=1}^d f_j(n_j) \right) F(n/N)G(Ln + v).$$

It will be convenient to introduce a logarithmic weighting to the primes. To this end, following [13], we define the function $\Lambda' : \mathbb{Z} \to \mathbb{R}$ by

$$\Lambda'(n) := \begin{cases} \log n, & n \text{ is prime}, \\ 0, & \text{otherwise}. \end{cases}$$

The von Mangoldt function $\Lambda$ will not be needed in this paper.

Another notion from [13] will be useful.

**Definition 1.15 (Local von Mangoldt function).** For $q \geq 2$, the local von Mangoldt function $\Lambda_{\mathbb{Z}/q\mathbb{Z}} : \mathbb{Z} \to \mathbb{R}$ is the $q$-periodic function defined by

$$\Lambda_{\mathbb{Z}/q\mathbb{Z}}(n) = \begin{cases} \frac{q}{\varphi(q)} & (n, q) = 1 \\ 0 & \text{otherwise}. \end{cases}$$

We let $\Lambda_{\mathbb{Z}/q\mathbb{Z}}^+ : \mathbb{Z} \to \mathbb{R}$ denote the restriction of $\Lambda_{\mathbb{Z}/q\mathbb{Z}}$ to the non-negative reals, namely the function $\Lambda_{\mathbb{Z}/q\mathbb{Z}}^+ : [0, \infty)$.

The local von Mangoldt function, when $q$ is the product of small primes, can be viewed as a model for the function $\Lambda'$. This model\(^2\) is intimately connected to a technical device known as the $W$-trick, which we recall in Section 7.

For a function $F : \mathbb{R}^d \to \mathbb{R}^m$ we define the Lipschitz constant of $F$ to be

$$\sup_{x, y \in \mathbb{R}^d} \frac{\|F(x) - F(y)\|_\infty}{\|x - y\|_\infty},$$

and call $F$ Lipschitz if this value is finite.

We may now state the main theorem.

**Theorem 1.16 (Main theorem).** Let $N, m, d$ be natural numbers with $d \geq m + 2$, and let $C, \varepsilon, \sigma$ be positive real parameters. Let $L : \mathbb{R}^d \to \mathbb{R}^m$ be a surjective linear map with algebraic coefficients, and suppose that $L \notin V_{\text{degen}}^*(m, d)$. Let $v \in \mathbb{R}^m$ be any vector that satisfies $\|v\|_\infty \leq CN$. Let $F : \mathbb{R}^d \to [0, 1]$ and $G : \mathbb{R}^m \to [0, 1]$ be compactly supported Lipschitz functions with Lipschitz constants at most $\sigma^{-1}$.

\(^2\)This is essentially the modified Cramér random model.
and assume that $F$ is supported on $[-1, 1]^d$ and $G$ is supported on $[-\varepsilon, \varepsilon]^m$. Let $w = w(N) := \log \log \log N$, assuming that $N$ is large enough for this function to be well defined, and let $W = W(N) := \prod_{p \leq w} p$. Then

$$(1.9) \quad T^{L, v}_{F, G, N}(\Lambda', \ldots, \Lambda') = T^{L, v}_{F, G, N}(\Lambda_{Z/WZ}^+, \ldots, \Lambda_{Z/WZ}^+) + o_{C, L, \varepsilon, \sigma}(1)$$

as $N \to \infty$.

**Remark 1.17.** If $F$ is supported on $[0, 1]^d$, we have

$$T^{L, v}_{F, G, N}(\Lambda_{Z/WZ}^+, \ldots, \Lambda_{Z/WZ}^+) = T^{L, v}_{F, G, N}(\Lambda_{Z/WZ}^+, \ldots, \Lambda_{Z/WZ})$$

We will prove an asymptotic formula for $T^{L, v}_{F, G, N}(\Lambda_{Z/WZ}^+, \ldots, \Lambda_{Z/WZ})$ later, in Lemma 9.11 and Remark 9.12. For example, if $v = 0$ and

$$L = \begin{pmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -\sqrt{3} & 1 \end{pmatrix},$$

say, and $F$ and $G$ are smooth functions supported on $[0, 1]^d$ and $[-1/2, 1/2]^2$ respectively, one may use Lemma 9.11 and Remark 9.12 to show that

$$T^{L, v}_{F, G, N}(\Lambda_{Z/WZ}^+, \ldots, \Lambda_{Z/WZ}) = \mathcal{G}J + o_{F, G, L}(1),$$

where

$$\mathcal{G} = \frac{1}{2} \prod_{p \geq 3} \left( 1 - \frac{2}{p} \right) \left( \frac{p}{p-1} \right)^2$$

and

$$J = \frac{1}{N^2} \int_{x \in \mathbb{R}^3} F(\Xi(x)/N)G(L, \Xi(x)) \, dx,$$

where

$$\Xi(x_1, x_2, x_3) = (x_1, x_1 + x_2, x_1 + 2x_2, x_3).$$

The constant $\mathcal{G}$ is in fact equal to

$$\prod_p \frac{1}{p^3} \sum_{x_1, x_2, x_3 \leq p} \prod_{j=1}^4 \Lambda_{Z/pZ}(\tilde{\xi}(x_1, x_2, x_3)),$$

where $(\tilde{\xi}_1, \tilde{\xi}_2, \tilde{\xi}_3, \tilde{\xi}_4)$ are the coordinate maps for $\Xi$.

It takes some effort to establish precisely what the map $\Xi$ should be for a given $L$. What’s more, the asymptotic formula in the general case is not just a product of a local factor and a global factor but rather a finite sum of products of local factors and global factors, and we will need to introduce an abundance of additional notation in order to be able to state these terms properly. Thus, in the interests of readability, we choose not to include this formula as part of the statement of Theorem 1.16.
Remark 1.18. If $L$ has rational coefficients\footnote{or more generally if $L$ has rational dimension $m$, see Definition 5.2 below.}, then Theorem 1.16 reduces to a statement on linear equations in primes (a reduction which we will make precise in Remark 5.7 below). In this sense, our work is a generalisation of Green–Tao–Ziegler.

Remark 1.19. We have phrased Theorem 1.16 with Lipschitz cut-offs $F$ and $G$. In Section 17 we will demonstrate how these cut-offs may be removed when $L$ is ‘purely irrational’, and in doing so will demonstrate how Theorem 1.16 implies Theorem 1.7. The same methods may be applied when $L$ is not purely irrational, but they will not always succeed, due to the rational degeneracy introduced in those cases. Unfortunately we have not been able to formulate what we regard to be a satisfactory general condition for saying when (1.9) holds with sharp cut-offs $F$ and $G$. Note in particular how the proof of Lemma A.2 relies heavily on the convex sets $[-\varepsilon, \varepsilon]^m$ and $[0, N]^d$ being axis-parallel boxes. Therefore we do not present a version of the theorem in which summation is over a general convex set $K$, as is done in Theorem 1.1. However, if the reader wishes to apply a specific instance of Theorem 1.16 with sharp cut-offs, the methods of Section 17 and Appendix A will almost certainly suffice for the purpose.

Remark 1.20. The reader will observe that, as in Theorem 1.7, we do not determine the nature of the dependence of the error term in (1.9) on the map $L$. We discussed this feature in Remark 1.10.

We conjecture that the conclusion of Theorem 1.16 holds for all $L \notin V_{\text{deg}}^*(m, d)$, provided $w$ grows slowly enough in terms of $L$.

Conjecture 1.21 (Transcendental case). \textit{Let $L$, $v$, $F$, and $G$ be as in the statement of Theorem 1.16, but do not assume that $L$ necessarily has algebraic coefficients. Then there is some function $w : \mathbb{N} \to \mathbb{R}_{\geq 0}$, with $w(N) \to \infty$ as $N \to \infty$, such that (1.9) holds with $W = \prod_{p \leq w} p$.}

In Section 9 we will formulate a statement involving smoothed sieve weights (namely Conjecture 9.6) which, if resolved, would settle Conjecture 1.21.

Acknowledgments. During the writing of this paper we benefited greatly from the supervision of Ben Green, and had helpful conversations with Sam Chow, Trevor Wooley, Yufei Zhao, Joni Teräväinen and Kaisa Matomäki. We would like to thank an anonymous referee for an exceptionally detailed reading of the manuscript and for many helpful corrections and comments. The majority of the work was
carried out while the author was supported by EPSRC grant no. EP/M50659X/1, continued while the author was a Program Associate at the Mathematical Sciences Research Institute in Berkeley, and finished while the author was supported by a Junior Research Fellowship at Trinity College Cambridge.

2 The structure of the argument

In this section we discuss our approach to proving Theorem 1.16, and describe the geography of the paper as a whole.

Initially, one might hope that Theorem 1.16 could be proved by replacing the coefficients of $L$ with some rational approximations, by considering the corresponding linear equation with rational coefficients, and then by appealing directly to Theorem 1.1 on linear equations in primes. However, unless the coefficients of $L$ are extremely well-approximable by rationals (and in particular are transcendental), such an approach does not seem to succeed. Indeed, let $L = (\lambda_{ij})_{i \leq m, j \leq d}$ and let $\lambda'_{ij}$ be a rational approximation to $\lambda_{ij}$, with $L'$ being the corresponding approximation to $L$. In order for the comparison of $L$ with $L'$ to be meaningful, we will need $\|Ln - L'n\|_\infty = O(1)$ for all relevant $n$, and in the general situation where all coordinates of $n$ have magnitude $\Omega(N)$ this requires $|\lambda'_{ij} - \lambda_{ij}|$ to be $O(N^{-1})$. Hence the numerator and denominator of $\lambda'_{ij}$ must grow rapidly with $N$, unless $\lambda_{ij}$ is extremely well-approximable. Yet Theorem 1.1 requires the coefficients of the associated affine linear equations to have height $O(1)$ (excepting the constant term, which may be $O(N)$). In [3] Bienvenu offers a slight improvement, but even with this refinement it does not seem that we can apply an existing result on linear equations in primes as a black box.

Instead, we will follow a similar approach to that which we used in our work [21], a paper that considered diophantine inequalities in the setting of bounded functions. Namely, we replace the function $\Lambda' : \mathbb{Z} \to \mathbb{R}$ by a suitable convolution $\Lambda' * \chi : \mathbb{R} \to \mathbb{R}$, designed to ensure the validity of the approximation

$$\sum_{n \in \mathbb{Z}^d} \left( \prod_{j=1}^d \Lambda'(n_j) \right) F(n/N)G(Ln + v)$$

(2.1)

$$\approx \int_{x \in \mathbb{R}^d} \left( \prod_{j=1}^d (\Lambda' * \chi)(x_j) \right) F(x/N)G(Lx + v) \, dx.$$

The integral may be manipulated by certain reparametrisations (Lemma 14.3),
yielding expressions of the form
\[ \int_{\mathbf{y} \in \mathbb{R}^{d-m}} \left( \prod_{j=1}^{d} g_j(\psi_j(\mathbf{y})) \right) F(\Psi(\mathbf{y})/N) \, d\mathbf{y}, \]
where \((\psi_1, \ldots, \psi_d) = \Psi : \mathbb{R}^{d-m} \rightarrow \mathbb{R}^{d}\) parametrises \(\ker L\) and \(g_1, \ldots, g_d\) are certain functions. By applying the Gowers–Cauchy–Schwarz inequality, in a manner strongly resembling [13, Appendix C], such expressions may be bounded by the Gowers norm \(\|\Lambda' - \Lambda_{\mathbb{Z}/WZ}\|_{U^s[N]}\), for some \(s = O(1)\). A qualitative bound on this Gowers norm is known by the work of Green-Tao-Ziegler (see Lemma 7.5), and so Theorem 1.16 follows.

The novel aspect of this manipulation, over the work of [13] and [21], is the appearance of various auxiliary linear inequalities, weighted by upper bound sieve weights. These enter in a manner that is somewhat analogous to the way in which the so-called ‘linear forms condition’ arises in [13]. Asymptotics for the number of solutions to these auxiliary inequalities underpin the argument, and this leads to a ‘linear inequalities condition’
\[ T_{L,\mathbf{v}}(\nu, \ldots, \nu) \approx T_{L,\mathbf{v}}(\Lambda_{\mathbb{Z}/WZ}, \ldots, \Lambda_{\mathbb{Z}/WZ}) \]
for a sieve weight \(\nu\), which is our corresponding notion of pseudorandomness (made precise in Definition 9.1). We are unable to verify this pseudorandomness condition in full generality, but we succeed in the case when \(L\) has algebraic coefficients. Our key technical tool is a bound for the number of solutions to a diophantine inequality restricted to a lattice, which we prove using the Davenport–Heilbronn method. This is the only part of the entire argument that uses the fact that the coefficients are assumed to be algebraic.

There is a final technical manoeuvre that we employ, one which has no direct analogue in [13] or [21]. It will transpire that passing to the local von Mangoldt function \(\Lambda_{\mathbb{Z}/WZ}\) introduces certain singular expressions, which arise from the fact that we are dealing with inequalities rather than equations. To circumvent this issue we find it necessary to work at two different ‘local scales’, introducing functions \(\Lambda_{\mathbb{Z}/W^*Z}\) and \(\Lambda_{\mathbb{Z}/WZ}\). By careful manoeuvring one can ensure that the singular expressions are only introduced by the \(W^*\) scale, and so, provided \(W^*\) grows slowly enough compared to \(W\), these singularities may be offset by the decay in the Gowers norm expressions involving \(W\). This further complicates the analysis of the expressions, and in fact our final choice of function \(W^*\) will be non-effective.

The structure of the paper is as follows. The main elements of the proof of Theorem 1.16 take place in Part 5, and the reader may wish to begin with this
section. It is here that we reduce matters to bounding certain systems by Gowers norms (Section 12), prove the approximation (2.1) (Section 13), and apply the Gowers–Cauchy–Schwarz inequality (Section 15).

However, the arguments of this part rely heavily on lemmas that are proved earlier in the paper, and these lemmas split naturally into four types. There are those results that are standard properties of smooth functions, and these are recorded in Section 3. We also have lemmas whose proofs involve manipulation of a purely linear algebraic nature, in order to reduce inequalities to ones that are ‘purely irrational’ or to put linear equations into ‘normal form’. We describe these notions in Part 2. The definition of pseudorandomness for an enveloping sieve weight is contained in Part 3, as is our proof that a certain weight satisfies this pseudorandomness condition. Also in this part one may find Conjecture 9.6, which, if resolved, would remove the algebraicity assumptions. Part 4 is reserved for those lemmas that involve the (somewhat tedious) manipulation of integrals into more pleasant forms. One of these lemmas is Lemma 11.1, which is the lemma that introduces the second local scale $W^*$ that we mentioned above.

The first appendix is concerned with elementary estimates relating to the integral that appears in the global factor of Theorem 1.16. As we have already said, Appendix B presents a Fourier-analytic argument which is essentially due to Parsell.

Finally, let us mention that, to help to streamline the statements of various propositions and lemmas in the paper as a whole, we have found it useful to introduce certain notational conventions that are unique to this paper. We describe these in Section 4.

Part 1

Preliminaries

3 Smooth functions

Smooth functions will play a significant role in the paper, and in this section we collect together those notions and lemmas that will be necessary for our forthcoming manipulations.

Following [17, Section 2], given a natural number $d$ and a compactly supported smooth function $F : \mathbb{R}^d \rightarrow \mathbb{R}$, we define $d(F)$ to be the corresponding value of $d$, $\text{Rad}(F)$ to be the smallest $R$ such that $F$ is supported on $[-R, R]^d$, and for every
non-negative integer \( j \) we define
\[
d_j(F) := \max \left\{ \left| \frac{\partial^{j_1+\cdots+j_d} F}{\partial x_1 \cdots \partial x_d} \right| : a \in \mathbb{R}^d, \sum_{i=1}^d j_i = j \right\}.
\]

Then, if \( P \) is any set, we shall define \( C(P) \) to be the set of those smooth functions \( F \) for which
\[
d(F), \text{Rad}(F), d_0(F), d_1(F), d_2(F), \ldots
\]
can be bounded above by quantities that depend only on the elements of \( P \). For example, let \( g : \mathbb{R} \to \mathbb{R} \) be the function given by
\[
g(x) := \begin{cases} 
\exp(-1 - x^2)^{-1} & |x| \leq 1 \\
0 & |x| > 1,
\end{cases}
\]
and then for a positive parameter \( \delta \) let \( g_\delta : \mathbb{R} \to \mathbb{R} \) be defined by \( g_\delta(x) := g(x/\delta) \). Then \( g_\delta \in \mathcal{C}(\delta) \), as is proved rather succinctly in [4, Lemma 9], say.

In order to shorten some of the statements in the main part of the paper, it will be convenient to consider all functions on \( \mathbb{R}^0 \) to be smooth (with derivatives equal to 0).

Let us record a standard proposition on smooth majorants and minorants.

**Lemma 3.1.** Let \( \delta \) be a real number in the range \( 0 < \delta < 1 \). Then there exist two smooth functions \( f^+, f^- : \mathbb{R} \to [0, 1] \), with \( f^+, f^- \in \mathcal{C}(\delta) \), satisfying
\[
1_{[\delta, 1-\delta]}(x) \leq f^-\delta(x) \leq 1_{[0,1]}(x) \leq f^+\delta(x) \leq 1_{[-\delta, 1+\delta]}(x)
\]
for all \( x \in \mathbb{R} \).

**Proof.** Let \( g \) be as above, and let \( C := \int g(x) \, dx \). Then one may define
\[
f^-\delta(x) := 4 \frac{\delta}{C} \int_{y \in \mathbb{R}} 1_{[\delta/2, 1-\delta/2]}(y) g(4(x-y)/\delta) \, dy
\]
and
\[
f^+\delta(x) := 4 \frac{\delta}{C} \int_{y \in \mathbb{R}} 1_{[-\delta/2, 1+\delta/2]}(y) g(4(x-y)/\delta) \, dy.
\]
The fact that \( f^+, f^- \in \mathcal{C}(\delta) \) follows from differentiating under the integral (which is easily justified by the mean value theorem).

**Lemma 3.2 (Smooth partition of unity).** Let \( \delta \) be a real number in the range \( 0 < \delta < 1 \). Then there exists a natural number \( t \), satisfying \( t = O(\delta^{-1}) \), and functions \( f_1, \ldots, f_t : \mathbb{R} \to [0, 1] \) such that
(1) for each \( i \leq t \), \( f_i \in \mathcal{C}(\delta) \);
(2) for each \( i \leq t \), \( f_i \) is supported on an interval of length at most \( 2\delta \);
(3) for all \( x \in \mathbb{R} \), \( 1_{[-1+\delta,1-\delta]}(x) \leq \sum_{i=1}^t f_i(x) \leq 1_{[-1-\delta,1+\delta]}(x) \);
(4) for all \( x \in \mathbb{R} \), \( x \) is contained in the support of at most 2 of the functions \( f_i \).

Proof. Let \( t = \lceil 4\delta^{-1} \rceil \), and write

\[
1_{[-1,-1+\delta/2)} = \sum_{i=1}^t I_i,
\]

where

\[
I_i := [-1 + (i - 1)\delta/2, -1 + i\delta/2).
\]

Then define

\[
f_i(x) := \frac{4}{\delta \mathcal{C}} \int_{y \in \mathbb{R}} 1_{I_i}(y) g(4(x-y)/\delta) \, dy.
\]

The desired properties are immediate. \( \square \)

Lemma 3.3 (Approximating Lipschitz functions by smooth boxes). Let \( \delta, \sigma, N \) be positive real parameters, with \( \delta, \sigma \) in the range \( 0 < \delta, \sigma < 1/2 \). Let \( d \) be a natural number, and let \( F : \mathbb{R}^d \to [0,1] \) be a Lipschitz function supported on \( [-N,N]^d \) with Lipschitz constant at most \( (\sigma N)^{-1} \). Then there exists a natural number \( k \), satisfying \( k = O(\delta^{-d}) \), and functions \( F_1, \ldots, F_k : \mathbb{R}^d \to [0,1] \) such that

(1) \( \| F - \sum_{i=1}^k F_i \|_\infty = O(\delta \sigma^{-1}) \);
(2) for each \( i \leq k \), \( F_i \) is supported on a box with side length \( O(\delta) \);
(3) there is a natural number \( t \), satisfying \( t = O(\delta^{-1}) \), and functions \( f_1, \ldots, f_t : \mathbb{R} \to [0,1] \), satisfying \( f_1, \ldots, f_t \in \mathcal{C}(\delta) \), such that

\[
F_i(x) = c_{i,F} \prod_{j=1}^d f_{S^{(i)}_j}(x_j/2N)
\]

for each \( i \leq k \), for some element \( S^{(i)} \in [t]^d \) and some constant \( c_{i,F} \in [0,1] \).

Proof. We have

\[
F(x) = F(x) 1_{[-1,1]^d}(x/2N)
\]

(3.1)

\[
= F(x) \left( \prod_{j=1}^d 1_{[1-\delta,1]}(x_j/2N) \right)
\]

\[
= F(x) \left( \prod_{j=1}^d \sum_{i=1}^t f_i(x_j/2N) \right),
\]
where the functions $f_1, \ldots, f_t$ are those constructed by applying Lemma 3.2 with this value of $\delta$. This manipulation is indeed valid, since $F(x) = 0$ for any $x$ for which
\[
\prod_{j=1}^d \sum_{i=1}^t f_i(x_j/2N) \neq \prod_{j=1}^d 1_{[-1,1]}(x_j/2N).
\]
Swapping the product and summation, (3.1) equals
\[
\sum_{S \in [t]^d} F(x) \left( \prod_{j=1}^d f_{S_j}(x_j/2N) \right).
\]

Let $x^{(S)} \in \mathbb{R}^d$ be any point at which $\prod_{j=1}^d f_{S_j}(x^{(S)}_j/2N)$ is non-zero. Then the above is equal to
\[
\sum_{S \in [t]^d} \left( F(x^{(S)}) + O(\delta \sigma^{-1}) \right) \left( \prod_{j=1}^d f_{S_j}(x_j/2N) \right),
\]
by the Lipschitz properties of $F$ and the limited support of the functions $f_1, \ldots, f_t$ (which was part (2) of Lemma 3.2).

Define
\[
F_S(x) := F(x^{(S)}) \left( \prod_{j=1}^d f_{S_j}(x_j/2N) \right).
\]
These functions satisfy properties (2) and (3) of Lemma 3.3. Finally note that, by part (4) of Lemma 3.2, each $x \in \mathbb{R}$ is contained in the support of at most $O(1)$ of the functions $F_S$, and hence $\|F - \sum_{S \in [t]^d} F_S\|_\infty = O(\delta \sigma^{-1})$, as required.

The Fourier transform of smooth functions will be an important tool in Section 8. We choose the following convention. If $F : \mathbb{R}^d \to \mathbb{R}$ is a compactly supported smooth function, we define the Fourier transform $\hat{F} : \mathbb{R}^d \to \mathbb{C}$ by the formula
\[
\hat{F}(\alpha) := \int_{x \in \mathbb{R}^d} F(x) e(-\alpha \cdot x) \, dx.
\]

**Lemma 3.4.** Let $P$ be a set of parameters and suppose $F \in \mathcal{C}(P)$. Then for every $\alpha$ and every non-negative integer $K$ one has
\[
|\hat{F}(\alpha)| \ll_{P,K} (1 + \|\alpha\|_\infty)^{-K}.
\]

**Proof.** This follows from integration by parts. 

Finally, we recall the definition of dual lattices and the version of the Poisson summation formula that we will use.
Definition 3.5 (Dual lattice). Let $h$ be a natural number and let $\Gamma \leq \mathbb{R}^h$ be a lattice of rank $h$. Then the dual lattice $\Gamma^*$ is defined by

$$\Gamma^* := \{ y \in \mathbb{R}^h : \langle y, x \rangle \in \mathbb{Z} \text{ for all } x \in \Gamma \}.$$ 

It is easily seen that if $M$ is an $h$-by-$h$ matrix whose columns are a lattice basis for $\Gamma$, then $(M^{-1})^T$ is an $h$-by-$h$ matrix whose columns are a lattice basis for $\Gamma^*$.

Lemma 3.6 (Poisson summation). Let $h$ be a natural number and let $\hat{\Gamma} \leq \mathbb{R}^h$ be a lattice of rank $h$. Let $F : \mathbb{R}^h \rightarrow \mathbb{C}$ be a smooth compactly supported function. Then

$$\sum_{x \in \hat{\Gamma}} F(x) = \frac{1}{\text{vol}(\mathbb{R}^h/\Gamma)} \sum_{y \in \Gamma^*} \hat{F}(y).$$

Proof. This is a standard result. The version in which $\Gamma = \mathbb{Z}^h$ appears as [8, Theorem 3.1.17], with the extension to general full-rank lattices following from a change of variables. \qed

4 Notation and conventions

For the most part the notation used in this paper is very standard, and any usage that could be viewed as somewhat unusual will be introduced as and when it is required. However, there are a few particular points that will apply to the paper as a whole which we believe to be important to address now.

We will use the Bachmann–Landau asymptotic notation $O$, $o$, and $\Omega$, but we do not, as is sometimes the convention, for a function $f$ and a positive function $g$ choose to write $f = O(g)$ if there exists a constant $C$ such that $|f(N)| \leq Cg(N)$ for $N$ sufficiently large. Rather we require the inequality to hold for all $N$ in some pre-specified range. If $N$ is a natural number, the range is always assumed to be $\mathbb{N}$ unless otherwise specified. It will be a convenient shorthand to use these symbols in conjunction with minus signs, whenever they appear in exponents. For example, $N^{-\Omega(1)}$ refers to a term $N^{-c}$, where $c$ is some positive quantity bounded away from 0 as the asymptotic parameter tends to infinity.

The Vinogradov symbol $\ll$ will be used, where for a function $f$ and a positive function $g$ we write $f \ll g$ if and only if $f = O(g)$. We write $f \asymp g$ if $f \ll g$ and $g \ll f$. If an implied constant or a $o(1)$ term depends on other parameters, we will denote these by subscripts, e.g., $O_{c,d}(1)$, or $f \asymp_{c,d} g$. However, if the implied constants depend on the underlying dimensions (denoted by $m, d$, and occasionally by $h, s, t$) we will not record this fact explicitly, as this would render most of the expressions unreadable.
The notation $\text{Rad}(F)$, which was introduced in the previous section for compactly supported smooth functions $F$, will also be used when $F$ is not smooth.

In order to keep track of which variables are scalars and which are vectors, we will use boldface $\mathbf{x}$ to denote any $\mathbf{x} \in \mathbb{R}^d$ where $d$ could be at least 2. In order to describe certain integrals over many variables, the following notational convention will be useful. If $\mathbf{x} \in \mathbb{R}^d$ and if $a$ and $b$ are two subscripts with $1 \leq a \leq b \leq d$, we use $\mathbf{x}_a^b$ to denote the vector $(x_a, x_{a+1}, \ldots, x_b)^T \in \mathbb{R}^{b-a+1}$.

With a view to trying to shorten some of the statements and proofs to follow, there are certain functions that we will fix throughout the paper, namely $w$, $W$, $\rho$, and $\chi$. From now on, the function $w : \mathbb{N} \rightarrow \mathbb{R}_{\geq 0}$ will always be defined by

$$w(N) := \max(1, \log \log \log N).$$

Whenever $N$ is a quantity that we have defined, we write $w$ for $w(N)$ and let

$$W = W(N) = \prod_{p \leq w(N)} p.$$

The empty product is considered to be equal to 1. Whenever other functions $w_1, \ldots, w_d, w^* : \mathbb{N} \rightarrow \mathbb{R}_{\geq 0}$ occur, and a natural number $N$ is given, we will define $W_1, \ldots, W_d, W^*$ analogously.

The following definition (a smooth version of [21, Definition 6.2]) will be a useful way to control certain functions that are required in the argument.

**Definition 4.1 ($\eta$-supported).** Let $\chi : \mathbb{R} \rightarrow [0, 1]$ be a smooth function, and let $\eta$ be a positive parameter. We say that $\chi$ is $\eta$-supported if $\chi$ is supported on $[-\eta, \eta]$ and $\chi(x) \equiv 1$ for all $x \in [-\eta/2, \eta/2]$.

It follows from Lemma 3.1 that 1-supported functions exist. From now on we fix a smooth function $\rho : \mathbb{R} \rightarrow [0, 1]$ that is 1-supported. We think of $\rho$ as an element of $\mathcal{C}(\emptyset)$. Whenever a positive parameter $\eta$ is defined we also define

$$\chi : \mathbb{R} \rightarrow [0, 1]$$

by the relation $\chi(x) := \rho(x/\eta)$. The function $\chi$ is $\eta$-supported, and satisfies $\chi \in \mathcal{C}(\eta)$.

We finish this section with some pieces of notation of a more standard nature. If $X, Y \subset \mathbb{R}^d$ for some $d$, we define

$$\text{dist}(X, Y) := \inf_{x \in X, y \in Y} \|x - y\|_\infty.$$
If $X$ is the singleton $\{x\}$, we write $\text{dist}(x, Y)$ for $\text{dist}(\{x\}, Y)$. We let $\partial(X)$ denote the topological boundary of $X$ (though the symbol $\partial$ will also be used for partial differentiation, as usual). If $A$ and $B$ are two sets with $A \subseteq B$, we let $1_A : B \longrightarrow \{0, 1\}$ denote the indicator function of $A$. The relevant set $B$ will usually be obvious from context. If $E$ is some event, e.g., a divisor condition, we will also use $1_E$ for the indicator function of this event. For $\theta \in \mathbb{R}$ we adopt the standard shorthand $e^{2\pi i \theta}$. The Möbius function will be denoted by $\mu$, though in Section 15 the symbol $\mu$ will also be used to denote a measure. In Section 9 we will use $\phi$ for Euler’s $\phi$-function, and for two natural numbers $a$ and $b$ we use the shorthand $(a, b)$ to denote their greatest common divisor.

Part 2

Linear algebra

In [21] we developed an armoury of linear-algebraic methods, which enabled us to manipulate linear inequalities into certain desired forms. The same manipulation is necessary here. We have chosen not to consign this material to an appendix, nor simply to cite [21], since the result of Lemma 5.6 below will be very important during subsequent sections. We will also need a few results (on the vector $\tilde{r}$ below) that were not required in our previous work, and so citing [21] won’t quite do.

Fortunately, as we do not seek to determine exactly how the error term in Theorem 1.16 depends on $L$, we can offer a significant simplification over the work that was presented in [21]. This is another reason to include this material.

Before starting, we remind the reader of some of the central definitions from the theory of dual vector spaces and dual linear maps, which will be used liberally throughout. Let $V$ be a finite-dimensional vector space over a field $\mathbb{F}$. Then $V^*$ denotes the dual vector space, i.e. the vector space of all linear maps $\omega : V \longrightarrow \mathbb{F}$ under pointwise addition and scalar multiplication. If $L : V \longrightarrow W$ is a linear map between two finite-dimensional vector spaces, the dual map $L^* : W^* \longrightarrow V^*$ is defined by the relation $(L^*(\omega))(\mathbf{v}) := \omega(L(\mathbf{v}))$ for all $\omega \in W^*$ and $\mathbf{v} \in V$. Given a basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ for $V$, the dual basis $\mathbf{e}_1^*, \ldots, \mathbf{e}_n^*$ for $V^*$ is defined by extending linearly the relations

$$e_i^*(e_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Finally, given a set $S \subset V$ the annihilator $S^0 \subset V^*$ is defined by

$$S^0 := \{ \omega \in V^* : \omega(\mathbf{v}) = 0 \text{ for all } \mathbf{v} \in S \}.$$
5 Dimension reduction

We begin with a generalisation of Definition 1.8. Note that the case $m = 0$ is permitted below.

**Definition 5.1.** Let $N, d, h$ be natural numbers, and let $m$ be a non-negative integer. Let $L : \mathbb{R}^h \to \mathbb{R}^m$ be a linear map, and let $(\xi_1, \ldots, \xi_d) = \Xi : \mathbb{R}^h \to \mathbb{R}^d$ be a linear map with integer coefficients. Let $F : \mathbb{R}^d \to \mathbb{R}$ and $G : \mathbb{R}^m \to \mathbb{R}$ be functions with compact support. Let $v \in \mathbb{R}^m$ and $\tilde{r} \in \mathbb{Z}^d$. Then for $f_1, \ldots, f_d : \mathbb{Z} \to \mathbb{R}$ we define

$$T_{L,v,\Xi,\tilde{r}}(f_1, \ldots, f_d) := \frac{1}{N^{h-m}} \sum_{n \in \mathbb{Z}^h} \left( \prod_{j=1}^d f_j(\xi_j(n) + \tilde{r}_j) \right) F\left( \frac{\Xi(n) + \tilde{r}}{N} \right) G(Ln + v),$$

where $\tilde{r}_j$ is the $j$th coordinate of $\tilde{r}$.

The reader might notice that this definition is subtly different from the similar definition that appeared in [21], namely Definition 5.3 of that paper, in which the function $n \mapsto F((\Xi(n) + \tilde{r})/N)$ was treated as an arbitrary function $F_1 : \mathbb{R}^h \to [0, 1]$. When dealing with quantitative aspects of smooth functions (a feature of this paper that is not required in [21]) it is convenient to preserve the internal structure of this particular function, and so we have modified Definition 5.1 accordingly.

Recall the notion of rational maps from [21].

**Definition 5.2** (Rational dimension, rational map, purely irrational). Let $m$ and $d$ be natural numbers, with $d \geq m$. Let $L : \mathbb{R}^d \to \mathbb{R}^m$ be a surjective linear map. Let $u$ denote the largest integer for which there exists a surjective linear map $\Theta : \mathbb{R}^m \to \mathbb{R}^u$ for which $\Theta(L(\mathbb{Z}^d)) \subseteq \mathbb{Z}^u$. We call $u$ the rational dimension of $L$, and we call any map $\Theta$ with the above property a rational map for $L$. We say that $L$ is purely irrational if $u = 0$.

**Remark 5.3.** If (the matrix of) $L$ has algebraic coefficients, then there exists a rational map for $L$ that also has algebraic coefficients.

Purely irrational linear maps are those that we may analyse most easily using the Davenport-Heilbronn method (see Section 8). However, even when proving Theorem 1.7, whose statement concerns only purely irrational linear maps, we will be forced to consider auxiliary linear maps that are not purely irrational. It is necessary therefore to develop a rudimentary theory of these maps. Readers
desiring more detail and motivating examples concerning rational maps and rational dimension may consult Sections 2 and 5 of [21].

Our key tool will be Lemma 5.6, which is a version of Lemma 5.10 from [21]. This lemma will enable us to ‘quotient out’ the rational relations that are present in a diophantine inequality, leaving behind a purely irrational linear map between spaces of a lower dimension. In particular, we will show that

\[ T_{F,G,N}^{L',\nu}(f_1, \ldots, f_d) = \sum_{\tilde{r} \in \tilde{R}} T_{F,G\tilde{r},N}^{L',\nu,\tilde{r}}(f_1, \ldots, f_d), \]

where \( L' \) is purely irrational, and the vectors \( \nu' \) and \( \tilde{r} \), the linear map \( \tilde{\Xi} \) and the function \( G_{\tilde{r}} \) are objects that we may control.

To state the lemma we need to recall explicitly the notion from [13] that was mentioned in Remark 1.6, namely finite Cauchy–Schwarz complexity for linear maps.\(^4\)

**Definition 5.4** (Finite Cauchy–Schwarz complexity). Let \( d, h \) be natural numbers, and let \((\xi_1, \ldots, \xi_d) := \Xi : \mathbb{R}^h \rightarrow \mathbb{R}^d\) be a linear map. We say that \( \Xi \) has infinite Cauchy–Schwarz complexity if there are two distinct indices \( i \) and \( j \), and some \( \lambda \in \mathbb{R} \), for which \( \xi_i = \lambda \xi_j \). If no such \( i \) and \( j \) exist we say that \( \Xi \) has finite Cauchy–Schwarz complexity.

There is an equivalent definition, which will be more convenient for algebraic manipulations.

**Definition 5.5** (Finite Cauchy–Schwarz complexity, equivalent definition). Let \( d, h \) be natural numbers. Let \( e_1, \ldots, e_d \) denote the standard basis vectors of \( \mathbb{R}^d \), and let \( e_1^*, \ldots, e_d^* \) denote the dual basis of \((\mathbb{R}^d)^*\). Then let \( V_{\text{degen}}(h, d) \) denote the set of all linear maps

\[ \Xi : \mathbb{R}^h \rightarrow \mathbb{R}^d \]

for which there exist two indices \( i, j \leq d \), and some real number \( \lambda \), such that \( e_i - \lambda e_j \) is non-zero and \( e_i^* - \lambda e_j^* \in \ker(\Xi^*) \). If \( \Xi \notin V_{\text{degen}}(h, d) \), we say that \( \Xi \) has infinite Cauchy–Schwarz complexity.

The equivalence of these definitions is elementary.

For more background on the notion of finite Cauchy–Schwarz complexity, the reader may consult Section 1 of [13] or Section 4 of [21].

Now we may state and prove the important lemma, which provides the ‘dimension reduction’ of the section title.

\(^4\)In [21] a notion of degeneracy for pairs of linear maps was useful, but we have structured the present paper in such a way as to avoid requiring this complicated notion.
**Lemma 5.6** (Generating a purely irrational map). Let $m$, $d$ be natural numbers, with $d \geq m+2$, and let $C$, $\eta$ be positive parameters. Let $L: \mathbb{R}^d \to \mathbb{R}^m$ be a surjective linear map with algebraic coefficients. Let $u$ be the rational dimension of $L$. Let $F : \mathbb{R}^d \to [0,1]$ and $G : \mathbb{R}^m \to [0,1]$ be compactly supported functions. Assume that $G$ is smooth, $\text{Rad}(G) \leq \eta$, and moreover that $G \in \mathcal{C}(P, \eta)$ for some set of parameters $P$. Let $v \in \mathbb{R}^m$ be a vector with $\|v\|_\infty \leq CN$. Then there exists a surjective linear map $\Theta : \mathbb{R}^m \to \mathbb{R}^u$, a surjective linear map $L' : \mathbb{R}^{d-u} \to \mathbb{R}^{m-u}$, an injective linear map $\Xi : \mathbb{R}^{d-u} \to \mathbb{R}^d$, a finite subset $\tilde{R} \subset \mathbb{Z}^d$, a vector $v' \in \mathbb{R}^{m-u}$, and, for each $\tilde{r} \in \tilde{R}$, a compactly supported function $G_{\tilde{r}} : \mathbb{R}^{m-u} \to [0,1]$, such that

1. $\Theta$ is a rational map for $L$ with algebraic coefficients;
2. $\Xi$ has integer coefficients, depends only on $L$, and satisfies $\text{Im } \Xi = \ker \Theta L$ and $\Xi(\mathbb{Z}^{d-u}) = \mathbb{Z}^d \cap \text{Im } \Xi$;
3. $\tilde{R}$ satisfies $|\tilde{R}| = O_{L,\eta}(1)$, and $\|\tilde{r}\|_\infty = O_{C,L,\eta}(N)$ for all $\tilde{r} \in \tilde{R}$;
4. for all $\tilde{r} \in \tilde{R}$, the function $G_{\tilde{r}}$ is smooth, $\text{Rad}(G_{\tilde{r}}) = O_L(\eta)$, and $G_{\tilde{r}} \in \mathcal{C}(L, P, \eta)$;
5. $v'$ satisfies $\|v'\|_\infty = O_{C,L}(N)$;
6. for all natural numbers $N$, and for all functions $f_1, \ldots, f_d : \mathbb{Z} \to \mathbb{R}$, one has

$$T_{F,G,N}^{L,L'}(f_1, \ldots, f_d) = \sum_{\tilde{r} \in \tilde{R}} T_{F,G,N}^{L',v'}(\tilde{r})(f_1, \ldots, f_d);$$

7. $L'$ is purely irrational, depends only on $L$, and has algebraic coefficients;
8. if $L \notin V_{\text{degen}}(m, d)$ then $\Xi$ has finite Cauchy–Schwarz complexity.

The above properties suffice for Section 9, but three additional properties also hold. We will need these additional properties in Section 11.

9. Letting $e_1, \ldots, e_{d-u}$ denote the standard basis of $\mathbb{R}^{d-u}$, there is a set

$$\{ x_i : i \leq u \} \subset \mathbb{R}^d$$

for which

$$B := \{ x_i : i \leq u \} \cup \{ \Xi(e_j) : j \leq d - u \}$$

is a basis for $\mathbb{R}^d$ and a lattice basis for $\mathbb{Z}^d$. Furthermore, $\bar{R} \subset \text{span}(x_i : i \leq u)$ and $\{ \Theta Lx_i : i \leq u \}$ is a lattice basis for $\Theta L \mathbb{Z}^d$;

10. if $\eta$ is small enough in terms of $L$, and if $v = L\alpha$ for some $\alpha \in \mathbb{R}^d$, then $|\bar{R}| = 1$ and $\bar{r} \in \bar{R}$ is a vector that minimises $\|\Theta L(\bar{r} + \alpha)\|_\infty$ over all $\bar{r} \in \mathbb{Z}^d$;

11. for all $\bar{r} \in \bar{R}$ and $x \in \mathbb{R}^{d-u}$ one has

$$G_{\bar{r}}(L'x + v') = G(L \Xi(x) + L\bar{r} + v).$$

**Proof.** Parts (1) and (2). Choose $\Theta : \mathbb{R}^m \to \mathbb{R}^u$ to be a rational map for $L$ that has algebraic coefficients. By rank-nullity $\ker(\Theta L)$ is a $d - u$ dimensional subspace of $\mathbb{R}^d$, and also the matrix of $\Theta L$ has integer coefficients. Combining
these two facts, we see that \( \ker(\Theta L) \cap \mathbb{Z}^d \) is a \( d-u \) dimensional lattice, and (by the standard algorithms) one can find a lattice basis \( \mathbf{v}_1, \ldots, \mathbf{v}_{d-u} \in \mathbb{Z}^d \) that satisfies \( \| \mathbf{v}_i \|_\infty = O_L(1) \) for every \( i \).

Let \( \mathbf{e}_1, \ldots, \mathbf{e}_{d-u} \) denote the standard basis of \( \mathbb{R}^{d-u} \), and then define

\[
\Xi : \mathbb{R}^{d-u} \rightarrow \mathbb{R}^d
\]

by

\[
\Xi(\mathbf{e}_i) := \mathbf{v}_i
\]

for all \( i \leq d-u \). Then \( \Xi \) satisfies part (2) of the lemma.

**Parts (3), (9) and (10).** There is a set of vectors \( \{ \mathbf{a}_1, \ldots, \mathbf{a}_u \} \subset \mathbb{Z}^u \) that is an integer basis for the lattice \( \Theta L(\mathbb{Z}^d) \) and for which \( \| \mathbf{a}_i \|_\infty = O_L(1) \) for each \( i \). Furthermore there exists a set of vectors \( \{ \mathbf{x}_1, \ldots, \mathbf{x}_u \} \subset \mathbb{Z}^d \) such that \( \Theta L(\mathbf{x}_i) = \mathbf{a}_i \) for each \( i \), and \( \| \mathbf{x}_i \|_\infty = O_L(1) \). By Lemma 5.8 of [21],

\[
(5.3) \quad \mathcal{B} := \{ \mathbf{x}_i : i \leq u \} \cup \{ \Xi(\mathbf{e}_j) : j \leq d-u \}
\]

is a basis for \( \mathbb{R}^d \) and a lattice basis for \( \mathbb{Z}^d \).

Now, if \( \mathbf{z} \in \mathbb{R}^m \) and \( \Theta(\mathbf{z}) = \mathbf{r} \) then \( \| \mathbf{z} \|_\infty = \Omega_L(\| \mathbf{r} \|_\infty) \). Recall that \( \text{Rad}(G) \leq \eta \) and that \( \Theta L(\mathbb{Z}^d) \subseteq \mathbb{Z}^u \). It follows that there are at most \( O_{L,\eta}(1) \) possible vectors \( \mathbf{r} \in \mathbb{Z}^u \) for which there exists a vector \( \mathbf{n} \in \mathbb{Z}^d \) for which both \( G(L\mathbf{n} + \mathbf{v}) \neq 0 \) and \( \Theta L \mathbf{n} = \mathbf{r} \). Let \( \mathcal{R} \) denote the set of all such vectors \( \mathbf{r} \). Observe that, for all \( \mathbf{r} \in \mathcal{R} \),

\[
\| \mathbf{r} \|_\infty = O_{C, L, \eta}(N).
\]

For each \( \mathbf{r} \in \mathcal{R} \), there exists a unique vector \( \tilde{\mathbf{r}} \in \text{span}(\mathbf{x}_i : i \leq u) \) such that \( \Theta L \tilde{\mathbf{r}} = \mathbf{r} \). Note that \( \| \tilde{\mathbf{r}} \|_\infty = O_{C, L, \eta}(N) \). Letting \( \mathcal{R} \) denote the set of these \( \tilde{\mathbf{r}} \), we see that \( \mathcal{R} \) satisfies part (3).

If \( \eta \) is small enough in terms of \( L \), then \( \mathcal{R} \) has size at most 1. Indeed, if \( \mathbf{r}^{(1)} \) and \( \mathbf{r}^{(2)} \) are two different vectors in \( \mathcal{R} \), with respective \( \tilde{\mathbf{r}}^{(1)} \) and \( \tilde{\mathbf{r}}^{(2)} \), then \( G(L\tilde{\mathbf{r}}^{(1)} + \mathbf{v}) \neq 0 \) and \( G(L\tilde{\mathbf{r}}^{(2)} + \mathbf{v}) \neq 0 \). Hence \( \| L(\mathbf{r}^{(1)} - \mathbf{r}^{(2)}) \|_\infty \ll \eta \). Yet

\[
\| \Theta L(\tilde{\mathbf{r}}^{(1)} - \tilde{\mathbf{r}}^{(2)}) \|_\infty = \| \mathbf{r}^{(1)} - \mathbf{r}^{(2)} \|_\infty \gg 1
\]

(which is a contradiction). In this instance, writing \( \mathbf{v} \) in the form \( L\mathbf{a} \), we may pick \( \tilde{\mathbf{r}} \in \mathbb{Z}^d \) to be an element in \( \text{span}(\mathbf{x}_i : i \leq u) \) that minimises \( \| \Theta L(\tilde{\mathbf{r}} + \mathbf{a}) \|_\infty \) over all \( \tilde{\mathbf{r}} \in \mathbb{Z}^d \).

**Parts (4), (5), (6) and (11):** By the definition of \( \mathcal{R} \), and the fact that

\[
\Xi(\mathbb{Z}^{d-u}) = \mathbb{Z}^d \cap \ker(\Theta L),
\]

we have that \( T^{L, \mathbf{v}}_{F, G, N}(f_1, \ldots, f_d) \) is equal to

\[
(5.4) \quad \sum_{\tilde{\mathbf{r}} \in \mathcal{R}} \frac{1}{N^{d-m}} \sum_{\mathbf{n} \in \mathbb{Z}^{d-u}} \left( \prod_{j=1}^d f_j(\zeta_j^m(\mathbf{n}) + \tilde{r}_j) \right) F \left( \frac{\Xi(\mathbf{n}) + \tilde{\mathbf{r}}}{N} \right) G(L\Xi(\mathbf{n}) + L\tilde{\mathbf{r}} + \mathbf{v}).
\]
This is very close to being of the form required for part (6), and indeed it can be massaged into exactly the required form.

To do this, note that

\[ \mathbb{R}^m = \text{span}(Lx_i : i \leq u) \oplus \ker \Theta \]

and so there exists an invertible linear map \( Q : \mathbb{R}^m \rightarrow \mathbb{R}^m \) with algebraic coefficients such that

\[
Q((\text{span}(Lx_i : i \leq u))) = \mathbb{R}^u \times \{0\}^{m-u}, \\
Q(\ker \Theta) = \{0\}^u \times \mathbb{R}^{m-u}.
\]

For all \( x \in \mathbb{R}^{d-u} \) we have

\[
G(L\Xi(x) + L\bar{r} + v) = (G \circ Q^{-1})(QL\Xi(x) + QL\bar{r} + Qv).
\]

We also note that \( QL\Xi(x) \in \{0\}^u \times \mathbb{R}^{m-u} \), and that \( QL\bar{r} \in \mathbb{R}^u \times \{0\}^{m-u} \).

Now, write \( \pi_{m-u} : \mathbb{R}^m \rightarrow \mathbb{R}^{m-u} \) for the projection map onto the final \( m-u \) coordinates. Define \( G_{\bar{r}} : \mathbb{R}^{m-u} \rightarrow [0,1] \) by

\[
(5.5) \quad G_{\bar{r}}(x) := (G \circ Q^{-1})(x_0 + QL\bar{r} + Qv - (\pi_{m-u}Qv)_0),
\]

where \( x_0 \) is the extension of \( x \) by 0 in the first \( u \) coordinates. Then \( G_{\bar{r}} \) satisfies the desired properties of part (3), since \( x_0 \) and \( QL\bar{r} + Qv - (\pi_{m-u}Qv)_0 \) are orthogonal.

Then (5.4) is equal to

\[
(5.6) \quad \sum_{\bar{r} \in \tilde{R}} \frac{1}{N^{d-m}} \sum_{n \in \mathbb{Z}^{d-u}} \left( \prod_{j=1}^d f_j(\xi_j(n) + \bar{r}_j) \right) \\
\times F\left( \frac{\Xi(n) + \bar{r}}{N} \right) G_{\bar{r}}(\pi_{m-u}QL\Xi(n) + \pi_{m-u}Qv).
\]

Let

\[
(5.7) \quad L' := \pi_{m-u}QL\Xi, \quad v' = \pi_{m-u}Qv.
\]

Then \( L' : \mathbb{R}^{d-u} \rightarrow \mathbb{R}^{m-u} \) is surjective, and

\[
T_{F,G,N}^{L',v',\Xi,\bar{r}}(f_1, \ldots, f_d) = \sum_{\bar{r} \in \tilde{R}} T_{F,G_{\bar{r}},N}^{L',v,\Xi,\bar{r}}(f_1, \ldots, f_d).
\]

This resolves parts (5) and (6). But furthermore, by the construction of \( G_{\bar{r}} \), part (10) is also satisfied.
**Part (7).** This is immediate from Lemma 5.10 of [21]. To spell it out, suppose for contradiction that there exists some surjective linear map $\varphi : \mathbb{R}^{m-u} \to \mathbb{R}$ with $\varphi L'(\mathbb{Z}^{d-u}) \subseteq \mathbb{Z}$, i.e. with $\varphi \pi_{m-u}QL\Xi(\mathbb{Z}^{d-u}) \subseteq \mathbb{Z}$. Then define the map $\Theta' : \mathbb{R}^m \to \mathbb{R}^{u+1}$ by

$$\Theta'(x) := (\Theta(x), \varphi \pi_{m-u}Q(x)).$$

Then $\Theta'$ is surjective, and $\Theta' L(\mathbb{Z}^d) \subseteq \mathbb{Z}^{u+1}$. This second fact is immediately seen by writing $\mathbb{Z}^d$ with respect to the lattice basis $B$ from (5.3). This contradicts the assumption that $L$ has rational dimension $u$. So $L'$ is purely irrational.

**Part (8).** Suppose $L \notin V_{\text{degen}}^*(m, d)$ and suppose for contradiction that $\Xi$ has infinite Cauchy-Schwarz complexity. Letting $e_1, \ldots, e_d$ denote the standard basis of $\mathbb{R}^d$, this means there exists $i, j \leq d$ and a non-zero vector $e_i - \lambda e_j$ such that $e_i^* - \lambda e_j^* \in \ker(\Xi^*)$. But $\ker(\Xi^*) = (\text{Im} \Xi)^0 = (\ker \Theta L)^0 = \text{Im}(L^* \Theta^*)$. Hence $e_i - \lambda e_j \in \text{Im} L^*$, which implies that $L \in V_{\text{degen}}^*(m, d)$, contradicting our hypothesis.

The lemma is proved. □

**Remark 5.7.** Applying Lemma 5.6 with $f_j = \Lambda'$ for all $j$, and when $L$ has rational dimension $m$, it is evident that estimating $T_{F,G,N}^{L,\lambda}(\Lambda', \ldots, \Lambda')$ is equivalent to counting solutions to $|\tilde{R}|$ systems of linear equations given by $\Xi$. This is handled by the Main Theorem of [13]. In this sense, one may see how our work in this paper generalises Green–Tao’s work in [13] to the cases in which the rational dimension is not equal to $m$.

### 6 Normal form

In this section we describe, very briefly, what it means for a linear map

$$(\psi_1, \ldots, \psi_t) = \Psi : \mathbb{R}^d \to \mathbb{R}^t$$

to be in $s$-normal form. For a more complete discussion we refer the reader to [13] and [21].

**Definition 6.1** (Normal form). Let $d, t$ be natural numbers, let $s$ be a non-negative integer, and let $(\psi_1, \ldots, \psi_t) = \Psi : \mathbb{R}^d \to \mathbb{R}^t$ be a linear map. We say that $\Psi$ is in $s$-normal form if for every $i \in [t]$ there exists a collection $J_i \subseteq \{e_1, \ldots, e_d\}$ of basis vectors of cardinality $|J_i| \leq s + 1$ such that $\prod_{e \in J_i} \psi_i(e)$ is non-zero for $i' = i$ and vanishes otherwise.

The notion of normal form is intimately connected with the notion of finite Cauchy–Schwarz complexity (Definition 5.5). The key proposition was proved\(^5\) in [13].

---

\(^5\)In [21] we were forced to prove a delicate quantitative version, but this will not be necessary here.
Lemma 6.2 (Normal form extensions). Let $d, t$ be natural numbers, and let $(\psi_1, \ldots, \psi_t) = \Psi : \mathbb{R}^d \rightarrow \mathbb{R}^t$ be a linear map with finite Cauchy–Schwarz complexity. Then there is a linear map $\Psi' : \mathbb{R}^{d'} \rightarrow \mathbb{R}^t$ such that:

- $d' = O(1)$;
- for some vectors $f_k \in \mathbb{R}^d$ that satisfy $\|f_k\|_\infty = O(1)$ for every $k$, the map $\Psi'$ is of the form

$$\Psi'(u, x_1, \ldots, x_{d'-d}) = \Psi(u + x_1 f_1 + \cdots + x_{d'-d} f_{d'-d})$$

for all $u \in \mathbb{R}^d$;
- $\Psi'$ is in $s$-normal form, for some $s = O(1)$.

Proof. In [13, Lemma 4.4] this lemma was proved for a linear map over a $\mathbb{Q}$-vector space. The proof over $\mathbb{R}$ is identical. Alternatively one can iterate [21, Proposition 4.8] over all $i \leq t$. □

Remark 6.3. In Lemma 6.2 one may take $s$ to be the Cauchy–Schwarz complexity of $\Psi$. This notion will not be used in this paper, save for the ‘finite versus infinite’ dichotomy already given in Definition 5.5.

Part 3

Pseudorandomness

Notions of pseudorandomness are crucial to the theory of higher order Fourier analysis. A small Gowers norm is one such notion, as is satisfying the ‘linear forms condition’ of [11] and [13]. In this part we review what is known about Gowers norms in relation to the primes, and then formulate a ‘linear inequalities condition’, which will be the analogous notion of pseudorandomness for this paper.

7 The W-trick and Gowers norms

To begin with, let us recall the definition of the Gowers norm over a cyclic group and over $[N]$. Given a function $f : \mathbb{Z}/N\mathbb{Z} \rightarrow \mathbb{C}$, and a natural number $d$, one defines the Gowers $U^d$ norm $\|f\|_{U^d([N])}$ to be the unique non-negative solution to the equation

$$\|f\|_{U^d([N])}^{2^d} = \frac{1}{N^{d+1}} \sum_{x, h_1, \ldots, h_d \in \{0, 1\}^d} \prod_{\omega \in \{0, 1\}^d} G[\omega] f(x + h \cdot \omega),$$

(7.1)
where $|\omega| = \sum_i\omega_i$, $h = (h_1, \ldots, h_d)$, $C$ is the complex-conjugation operator, and the summation is over $x, h_1, \ldots, h_d \in \mathbb{Z}/N\mathbb{Z}$. It is not immediately obvious why the right-hand side of (7.1) is always a non-negative real, nor why the $U^d$ norms are genuine norms if $d \geq 2$, but both facts are true. There are many expositions of the standard theory of these norms available in the literature, for example [19, Chapter 11] and [10]. For the most general treatment, the reader may consider Appendices B and C of [13].

In the sequel we will be considering functions defined on $[N]$ rather than on $\mathbb{Z}/N\mathbb{Z}$. However, the Gowers norm of such functions may be easily defined by reference to the cyclic group case. Indeed, if $f : [N] \to \mathbb{C}$, and $d$ is a natural number, one chooses a natural number $N' > N$ and then considers $[N]$ as an initial segment of $\mathbb{Z}/N'\mathbb{Z}$ (viewing $[N]$ as a set of representative classes for $\mathbb{Z}/N'\mathbb{Z}$). One then defines

$$
\|f\|_{U^d[N]} := \frac{\|f[N]\|_{U^d(N')}}{\|1[N]\|_{U^d(N')}},
$$

which is independent of $N'$ provided $N'/N$ is large enough in terms of $d$.

This is as much background as we will give here, and the reader is invited to consult the aforementioned references for more detail. A Gowers norm over $\mathbb{R}$ will also appear later on in this paper, but will be introduced in Section 13 as and when it is needed.

We move our consideration to the primes. Given some fixed modulus $q$ the primes are not uniformly distributed across arithmetic progressions modulo $q$ (as almost all the primes are coprime to $q$), and this lack of uniformity is an obstacle when trying to count solutions to equations in primes. Fortunately, there is a technical device, known as the $W$-trick, that has long been used to manage this difficulty.

This device is usually introduced via the following function.

**Definition 7.1.** Let $N$ be a natural number, and let $W$ be as in Section 4. For any natural number $b$ with $(b, W) = 1$, let $\Lambda_{b,W} : \mathbb{Z} \to \mathbb{R}_{\geq 0}$ be defined by

$$
\Lambda_{b,W}(n) = \begin{cases} 
\frac{\varphi(W)}{W} \Lambda'(Wn + b), & n \geq 1, \\
0, & \text{otherwise}.
\end{cases}
$$

The idea from [13], going back to [9] and [11], is that the function

$$
\frac{1}{\varphi(W)} \sum_{b \leq W \atop (b,W)=1} \Lambda_{b,W}
$$
should act as a proxy for $\Lambda'$, while each $\Lambda'_{b,W}$ enjoys strong pseudorandomness properties. For example we have the following deep result, which is a crucial component of the proof of Theorem 1.1 on linear equations in primes.

**Theorem 7.2** ([13, Theorem 7.2]). Let $N, s$ be natural numbers, and let $w^* : \mathbb{N} \to \mathbb{R}_{\geq 0}$ be any function that satisfies $w^*(n) \to \infty$ as $n \to \infty$ and $w^*(n) \leq \frac{1}{2} \log \log n$ for all $n$. Let $b = b(N)$ be a natural number that satisfies $b \leq W^*$ and $(b, W^*) = 1$. Then

$$\|\Lambda'_{b,W^*} - 1\|_{U^{s+1}[N]} = o(1)$$

as $N \to \infty$, where the $o(1)$ term may depend on the function $w^*$ chosen (but is independent of the choice of $b$).

We remind the reader that $s$ is a dimension parameter, and so dependence on $s$ is not denoted explicitly in our implied constants.

**Remark 7.3.** In [13], Theorem 7.2 is proved conditionally, relying on two other conjectures. But, as we intimated in the introduction, these conjectures were later settled in joint work of Green–Tao and Green–Tao–Ziegler [14, 15, 16].

**Remark 7.4.** We will use Theorem 7.2 to prove Theorem 1.16. Unfortunately it seems that this cannot be done in the same manner as in [13], i.e., by splitting $[N]$ into arithmetic progressions modulo $W$ at an early stage and then performing subsequent manipulations with the functions $\Lambda'_{b,W}$. As a heuristic, instead of considering an inequality such as

$$\|Ln + v\|_{\infty} \leq \varepsilon,$$

for some $L$ with irrational coefficients and some positive $\varepsilon$, [13] considers (7.4) for some $L$ with rational coefficients and sets $\varepsilon$ equal to 0. Under those assumptions one may rescale the variables $n$ by a factor of $W$, as required in Definition 7.1, without fundamentally altering the problem. However, in the more general scenario of Theorem 1.16, where $\varepsilon$ is strictly positive, rescaling the variable $n$ by a factor of $W$ means we must replace $\varepsilon$ by $\varepsilon/W$, and we cannot afford this loss, as the manipulations in Section 13 lose some powers of $\varepsilon$. As far as we have been able to tell, this means that we cannot perform the $W$-trick in this manner.

To circumvent this issue of scaling, we will manipulate with the local von Mangoldt functions $\Lambda_{Z/WZ}$ throughout, saving our rescaling for the very end of the argument. Regarding the control on Gowers norms, the following lemma is therefore the more appropriate bound.
Lemma 7.5. Let $N, s$ be natural numbers. Then
\[
\|\Lambda' - \Lambda_{Z/WZ}\|_{U^{s+1}[N]} = o(1)
\]
as $N \to \infty$.

The proof is a standard deduction from results of [13], achieved by splitting into arithmetic progressions modulo $W$. We would however like to thank the anonymous referee for suggesting a simplification to our original argument.

Proof. Let $(\psi_\omega)_{\omega \in \{0,1\}^{s+1}} = \Psi : \mathbb{R}^{s+2} \longrightarrow \mathbb{R}^{2^{s+1}}$ denote the linear map giving the Gowers norm, i.e., where each $\psi_\omega$ is of the form $\psi_\omega(x, h) = x + \omega \cdot h$. From expression (7.2), we then have
\[(7.5) \quad \|\Lambda' - \Lambda_{Z/WZ}\|^{2^{s+1}}_{U^{s+1}[N]} = \frac{1}{|Z|} \sum_{n \in \mathbb{Z}} \prod_{\omega \in \{0,1\}^{s+1}} (\Lambda' - \Lambda_{Z/WZ})(\psi_\omega(n)) \]
where
\[Z = \{ n \in \mathbb{Z}^{s+2} : \Psi(n) \in [1, N]^{2^{s+1}} \}.\]
It is immediate that $|Z| \asymp N^{s+2}$.

We now split into arithmetic progressions modulo $W$. To this end let $A \subset [W]^{s+2}$ be the set defined by
\[A = \{ a \in [W]^{s+2} : (\psi_\omega(a), W) = 1 \text{ for all } \omega \in \{0,1\}^{s+1} \}.\]
Then the right-hand side of (7.5) is
\[(7.6) \quad \asymp \frac{1}{N^{s+2}} \sum_{a \in A} \sum_{m \in \mathbb{Z}^{s+2}} \prod_{\omega \in \{0,1\}^{s+1}} (\Lambda' - \Lambda_{Z/WZ})(\psi_\omega(Wm + a)),\]
plus an error of magnitude at most
\[\frac{\log^{2^{s+1}}}{N^{s+2}} \sum_{n \in \mathbb{Z}} \sum_{\omega \in \{0,1\}^{s+1}} 1_{[0,W]}(\psi_\omega(n)).\]
This error is $o(1)$.

By the linearity of $\Psi$, and recalling the definition of $\Lambda'_b, W$ from Definition 7.1, we have that expression (7.6) is equal to
\[\frac{1}{N^{s+2}} \sum_{a \in A} \left( \frac{W}{\varphi(W)} \right)^{2^{s+1}} \sum_{m \in \mathbb{Z}^{s+2}} \prod_{\omega \in \{0,1\}^{s+1}} (\Lambda'_{\psi_\omega(a), W} - 1)(\psi_\omega(m)).\]
Observe that
\[|A| = \beta_W W^{s+2} \varphi(W)^{2^{s+1}} W^{2^{s+1}},\]
where
\[ \beta_W := \frac{1}{W^{s+2}} \sum_{m \in [W]^{s+2}} \prod_{\omega \in \{0, 1\}^{s+1}} \Lambda_{\mathbb{Z}/W\mathbb{Z}}(\psi_\omega(m)) \]
is the local factor associated to the system of forms \( \Psi \). Since \( \Psi \) has finite Cauchy–Schwarz complexity, we have the bound \( \beta_W = O(1) \) (by [13, Lemma 1.3]). This means that the lemma would follow from the bound

\[
\frac{1}{(N/W)^{s+2}} \sum_{m \in [Z/W]^{s+2}} \prod_{\omega \in \{0, 1\}^{s+1}} (\Lambda'_{\psi_\omega(a), W} - 1)(\psi_\omega(m)) = o(1)
\]

for each fixed \( a \in A \). What’s more, expression (7.7) is an immediate consequence of the Gowers–Cauchy–Schwarz inequality when combined with Theorem 7.2.

To spell out some of the details, let \( M := \lfloor N/W \rfloor \) and let \( M' > M \) be a natural number with \( M'/M \) large enough in terms of \( s \). Then, recalling the definition of the set \( Z \), the left-hand side of (7.7) is equal to

\[
\frac{1}{M^{s+2}} \sum_{m \in [Z/m/W]^{s+2}} \prod_{\omega \in \{0, 1\}^{s+1}} (\Lambda'_{\psi_\omega(a), W} - 1)(\psi_\omega(m)) + o(1).
\]

Taking the \( o(1) \) term as read, this is

\[
\ll \frac{1}{(M')^{s+2}} \sum_{m \in \mathbb{Z}/M'W} \prod_{\omega \in \{0, 1\}^{s+1}} (\Lambda'_{\psi_\omega(a), W} - 1)(\psi_\omega(m)).
\]

Now, by the Gowers–Cauchy–Schwarz inequality (see [19, Expression (11.6)]), expression (7.9) is at most

\[
\max_{\omega \in \{0, 1\}^{s+1}} \| \Lambda'_{\psi_\omega(a), W} \mathbf{1}_{[M]} - \mathbf{1}_{[M]}(\psi_\omega(m)) \|_{U^{s+1}(M')}^2.
\]

By expression (7.2), this is bounded above by a constant times

\[
\max_{1 \leq b \leq (s+2)W} \| \Lambda'_{b,W} - \mathbf{1} \|_{U^{s+1}[M]}^2.
\]

Expression (7.10) is directly amenable to Theorem 7.2, with the only wrinkle being the fact that Theorem 7.2 only applied to functions \( \Lambda'_{b,W} \) with \( b \leq W \). But this is easy to deal with. Indeed, for natural numbers \( n \) and \( k \) we have the identity

\[ \Lambda'_{b+kW,W}(n) = \Lambda'_{b,W}(n+k), \]

and so one establishes that if \( b \) is in the range \( 1 \leq b \leq (s+2)W \) then

\[ \| \Lambda'_{b,W} - \mathbf{1} \|_{U^{s+1}[M]}^2 = \| \Lambda'_{b',W} - \mathbf{1} \|_{U^{s+1}[M]}^2 + E, \]
where $b' \in [W]$ and $b' \equiv b \pmod W$, and where the error term $E$ is at most a constant times

$$\frac{\log^{x+1} M}{M^{x+2}} \sum_{m \in [M]^{x+2}} \sum_{a \in [0,1]^{x+1}} (1_{[s+2]}(\psi_{\omega}(m)) + 1_{[M+1,\ldots,M+s+2]}(\psi_{\omega}(m))).$$

We have $E = o(1)$ and therefore, by Theorem 7.2, expression (7.10) is $o(1)$. The lemma follows. □

8 Inequalities in lattices

This section will be devoted to proving the following technical lemma. This is the only part of the paper in which we pay especial attention to the quantitative aspects of the smooth cut-off functions, as the lemma will be applied in contexts where the functions $F$ and $G$ depend on the asymptotic parameter $N$ (albeit tamely).

**Lemma 8.1 (Inequalities in lattices).** Let $N, m, d, h$ be natural numbers, with $d \geq h \geq m + 1$, and let $\gamma$ be a positive constant. Suppose that $N > 2^\frac{\gamma}{2}$. Let $P$ be an additional set of parameters. Let $(\zeta_1, \ldots, \zeta_d) = \Xi : \mathbb{R}^h \rightarrow \mathbb{R}^d$ be an injective linear map with integer coefficients and let $L : \mathbb{R}^h \rightarrow \mathbb{R}^m$ be a purely irrational surjective linear map with algebraic coefficients. Let $v \in \mathbb{R}^m$ and let $\tilde{r} \in \mathbb{Z}^d$. Let $e_1, \ldots, e_d \in \mathbb{N}$ and suppose that $e_j < N^\gamma$ for all $j$. Let $F : \mathbb{R}^h \rightarrow [0,1]$ and $G : \mathbb{R}^m \rightarrow [0,1]$ be functions in $C^*(P)$. Then, provided that $\gamma$ is small enough in terms of $L$, for all positive $K$ we have

$$\frac{1}{N^{h-m}} \sum_{n \in \mathbb{Z}^h} F(n/N)G(Ln + v) = \frac{\alpha_{e,\tilde{r}}}{N^{h-m}} \int_{\mathbb{R}^h} F(x/N)G(Lx + v) \, dx + O_{K,L,P}(N^{-K}),$$

(8.1)

where $\alpha_{e,\tilde{r}}$ is the local factor

$$\alpha_{e,\tilde{r}} := \lim_{M \rightarrow \infty} \frac{1}{M^h} \sum_{n \in [M]^h} \prod_{j=1}^d 1_{e_j|\zeta_j(n)+\tilde{r}_j}.$$  

(8.2)

**Remark 8.2.** If $h = d$ and if $\Xi : \mathbb{R}^h \rightarrow \mathbb{R}^d$ is the identity map, then the Chinese Remainder Theorem guarantees that $\alpha_{e,\tilde{r}} = (e_1 \ldots e_d)^{-1}$. In the general case, the local factors $\alpha_{e,\tilde{r}}$ are the same objects as those factors $\alpha_{m_1,\ldots,m_t}$ considered in [13, Page 1831].

**Proof of Lemma 8.1.** We assume throughout that $\gamma$ is small enough in terms of $L$, and that $K$ is large enough in terms of the dimensions $m, d,$ and $h$. 

By applying Fourier inversion to $G$, we see that the left-hand side of (8.1) is equal to

\begin{equation}
\frac{1}{N^{h-m}} \int_{\alpha \in \mathbb{R}^m} \hat{G}(\alpha) \sum_{\substack{n \in \mathbb{Z}^h \ni e_j|\xi_j(n) + \tilde{r}_j \forall j \leq d}} F(n/N)e((L^T \alpha) \cdot n + a \cdot v) \, d\alpha.
\end{equation}

To bound this integral, we split $\mathbb{R}^m$ into three ranges. Let $\eta$ be a small positive parameter to be chosen later, which we assume to be small enough in terms of $L$. We then define the so-called ‘trivial arc’ by

\[ t := \{ \alpha \in \mathbb{R}^m : \| \alpha \|_\infty \geq N^\eta \}, \]

the ‘minor arc’ by

\[ m := \{ \alpha \in \mathbb{R}^m : N^{-1+\eta} \leq \| \alpha \|_\infty < N^\eta \}, \]

and the ‘major arc’ by

\[ \mathcal{M} := \{ \alpha \in \mathbb{R}^m : \| \alpha \|_\infty < N^{-1+\eta} \}. \]

**Trivial arc.** By Lemma 3.4, $|\hat{G}(\alpha)| \ll_{K,P} \| \alpha \|_\infty^K$. Therefore, applying the trivial $O_P(N^h)$ bound to the inner sum, we have

\[ \frac{1}{N^{h-m}} \int_{\alpha \in t} \hat{G}(\alpha) \sum_{\substack{n \in \mathbb{Z}^h \ni e_j|\xi_j(n) + \tilde{r}_j \forall j \leq d}} F(n/N)e((L^T \alpha) \cdot n + a \cdot v) \, d\alpha \ll_{K,P} N^{-\eta K + O(1)}. \]

**Minor arc.** Choose $x \in \mathbb{Z}^h$ to satisfy the simultaneous divisor conditions $e_j|\xi_j(x) + \tilde{r}_j$ for every $j \leq d$. If there is no such $x \in \mathbb{Z}^h$ then (8.1) is trivially true. Further, we may assume that $x$ satisfies $\| x \|_\infty \leq e_1 \ldots e_d$. Let $\Gamma_{\Xi,e}$ denote the lattice

\[ \Gamma_{\Xi,e} := \{ n \in \mathbb{Z}^h : e_j|\xi_j(n) \text{ for every } j \leq d \}. \]

Then

\begin{equation}
\sum_{\substack{n \in \mathbb{Z}^h \ni e_j|\xi_j(n) + \tilde{r}_j \forall j \leq d}} F(n/N)e((L^T \alpha) \cdot n) = \sum_{n \in \Gamma_{\Xi,e}} F((x + n)/N)e((L^T \alpha) \cdot (x + n)).
\end{equation}

Using this reformulation, we apply Poisson summation (Lemma 3.6) to the inner sum of (8.3). Then the contribution to (8.3) from the minor arc $m$ is equal to

\begin{equation}
\frac{N^{d-h+m}}{\text{vol}(\mathbb{R}^h/\Gamma_{\Xi,e})} \int_{\alpha \in m} \hat{G}(\alpha)e(a \cdot v) \sum_{c \in \Gamma^*_\Xi,e} \hat{F}(N(c - L^T \alpha))e(x \cdot c) \, d\alpha,
\end{equation}

where $\Gamma^*_\Xi,e$ is the lattice that is dual to $\Gamma_{\Xi,e}$ (see Definition 3.5).
We need the following obvious lemma.

**Lemma 8.3.** There is a natural number \( A \), of size at most \( O(N^{O(\gamma)}) \), such that \( A \Gamma^*_{\Xi,e} \subset \mathbb{Z}^h \).

**Proof.** There is an \( h \)-dimensional sublattice of \( \Gamma^*_{\Xi,e} \), namely \((e_1 \ldots e_d \mathbb{Z})^h\). Therefore, we may choose a lattice basis for \( \Gamma^*_{\Xi,e} \) all of whose elements \( b \) satisfy \( \|b\|_\infty = O(N^{O(\gamma)}) \). Let \( M \) be the \( h \)-by-\( h \) matrix that has these basis vectors as its columns. Then the columns of the matrix \((M^T)^{-1}\) are a lattice basis for the dual lattice \( \Gamma^*_{\Xi,e} \). The entries in \((M^T)^{-1}\) are rational numbers with numerator and denominator at most \( O(N^{O(\gamma)}) \). Clearing denominators, the lemma follows. \( \square \)

Let \( \langle L^T \alpha \rangle \) denote some \( c \in \Gamma^*_{\Xi,e} \) that minimises the expression \( \|c - L^T \alpha\|_\infty \).

We claim that the only term in (8.5) that cannot be easily absorbed into the error term comes from \( c = \langle L^T \alpha \rangle \).

Indeed, let \( A \) be the quantity provided by Lemma 8.3, and let \( \langle L^T \alpha \rangle_2 \) denote the second closest point to \( L^T \alpha \) in the lattice \( \Gamma^*_{\Xi,e} \). If more than one such point exists, choose arbitrarily. Then

\[
\|\langle L^T \alpha \rangle_2 - L^T \alpha\|_\infty \geq aA^{-1},
\]

where \( a \) is some positive constant which depends only on \( h \). By the triangle inequality and dyadic pigeonholing, one then has

\[
\left| \sum_{c \in \Gamma^*_{\Xi,e} \atop c \neq \langle L^T \alpha \rangle} \hat{F}(N\langle c - L^T \alpha \rangle)e(x \cdot c) \right| \leq \sum_{k=0}^{\infty} \sum_{2^k aA^{-1} \leq \|c - L^T \alpha\|_\infty \leq 2^{k+1} aA^{-1}} |\hat{F}(N\langle c - L^T \alpha \rangle)| .
\]

By Lemma 8.3 we also have the estimate

\[
\sum_{c \in \Gamma^*_{\Xi,e} \atop \|c - L^T \alpha\|_\infty \leq R} 1 \ll \max(1, R^h A^h) ,
\]

which holds for all \( R > 0 \). Using (8.8), Lemma 3.4, and the bound \( A = O(N^{O(\gamma)}) \), the quantity (8.7) is seen to be

\[
\ll_{K,P} N^{-K+O(K\gamma)} .
\]

This implies that the contribution from these lattice points to (8.5) is at most

\[
\ll_{K,P} N^{-K+O(K\gamma)+O(1)} \int_{\alpha \in m} |\hat{F}(\alpha)| \, d\alpha
\]

\[
\ll_{K,P} N^{-K+O(K\gamma)+O(1)+O(\eta)} .
\]
Since \( \gamma \) and \( \eta \) are small enough, (8.10) is
\[
\ll_{K, \mathcal{P}} N^{-K/2},
\]
which may be absorbed into the error term of (8.1) after adjusting the implied constant appropriately.

It remains to estimate
\[
N^{d-h+m} \int_{a \in m} \hat{G}(a)e(a \cdot \nu)\hat{F}(N(\langle L^T \alpha \rangle - L^T \alpha))e(x \cdot \langle L^T \alpha \rangle) \, da.
\]
We have the following key lemma.

**Lemma 8.4.** Under the assumption that \( \eta \) and \( \gamma \) are suitably small in terms of \( L \),
\[
\sup_{a \in m} |\hat{F}(N(\langle L^T \alpha \rangle - L^T \alpha))| \ll_{K, \mathcal{L}, \mathcal{P}} N^{-K\eta}.
\]

**Remark 8.5.** The proof of this lemma uses the algebraicity of the coefficients of \( L \). One should note that the bound (8.14) below, which holds for matrices with algebraic coefficients, also holds for almost all matrices. It is this fact which ultimately leads to our observation in the introduction that the main theorems of this paper hold for almost all matrices (as well as for matrices with algebraic coefficients, as stated).

**Proof.** Certainly, by rescaling \( \alpha \) and using Lemmas 3.4 and 8.3,
\[
\sup_{a \in m} |\hat{F}(N(\langle L^T \alpha \rangle - L^T \alpha))| \ll_{K, \mathcal{P}} N^{-K} \left( \inf_{a \in m} \| \langle L^T \alpha \rangle - L^T \alpha \|_\infty \right)^{-K}
\]
\[
\ll_{K, \mathcal{P}} A_K N^{-K} \left( \inf_{\beta \in \mathbb{R}^m} \inf_{n \in \mathbb{Z}^h} \| n - L^T \beta \|_\infty \right)^{-K}
\]
\[
\ll_{K, \mathcal{P}} A_K N^{-1+\eta} \left( \inf_{\beta \in \mathbb{R}^m} \inf_{n \in \mathbb{Z}^h} \| n - L^T \beta \|_\infty \right)^{-K}
\]

The quantity
\[
\inf_{\beta \in \mathbb{R}^m} \inf_{n \in \mathbb{Z}^h} \| n - L^T \beta \|_\infty
\]
encodes information about diophantine approximations to the coefficients of \( L \). For example, since \( L \) is purely irrational, by definition\(^6\) we have \( L^T \beta \not= \mathbb{Z} \) for any \( \beta \not= 0 \). Therefore, since the function
\[
\beta \mapsto \inf_{n \in \mathbb{Z}^h} \| L^T \beta - n \|_\infty
\]

\(^6\)The reader may consult Definition 5.2.
is continuous, (8.13) is always non-zero. We will need a quantitative refinement of this fact.

Fortunately, in [21] we extensively analysed expressions such as (8.13). Consider Definition 2.7 of [21] in particular, in which we defined the approximation function\(^7\) \(A_L\). In this language, (8.13) is equal to

\[
A_L(AN^{-1+\eta}, A^{-1}N^{-\eta}).
\]

Therefore, since \(L\) is purely irrational and has algebraic coefficients, Lemma E.1 of [21] tells us that

\[
\inf_{\beta \in \mathbb{R}^m} \inf_{\mathbf{n} \in \mathbb{Z}^d} \|L^T \beta - \mathbf{n}\|_\infty \gg_L \min(AN^{-1+\eta}, A^{-O_L(1)}N^{-O_L(\eta)}).
\]

Since \(\eta\) and \(\gamma\) are small enough in terms of \(L\), and since \(A = O(N^{O(\gamma)})\), (8.12) implies that

\[
\sup_{\alpha \in \mathcal{M}} |\hat{F}(\langle L^T \alpha \rangle - L^T \alpha)| \ll_{K,L,P} N^{-\eta},
\]
as claimed.

The lemma above implies that (8.11) has size

\[
O_{K,L,P}(N^{-K\eta+O(1)}),
\]
which is thus our bound for the total contribution from the minor arc \(m\).

**Major arc.** Performing the same Poisson summation argument as in the minor arc case, the main term on the left-hand side of (8.1) is equal to

\[
Nd - h + m \int_{\alpha \in \mathcal{M}} \hat{G}(\alpha)(\alpha \cdot \mathbf{v})\hat{F}(N(\langle L^T \alpha \rangle - L^T \alpha))e(\mathbf{x} \cdot (L^T \alpha))\ d\alpha.
\]

For \(\alpha \in \mathcal{M}\) one has \(\|L^T \alpha\|_\infty \ll_L N^{-1+\eta}\), and so \(\langle L^T \alpha \rangle = 0\). Therefore (8.16) is equal to

\[
Nd - h + m \int_{\alpha \in \mathcal{M}} \hat{G}(\alpha)(\alpha \cdot \mathbf{v})\hat{F}(-NL^T \alpha)\ d\alpha.
\]

Since \(L^T : \mathbb{R}^m \rightarrow \mathbb{R}^h\) is injective, one has \(\|L^T \alpha\|_\infty \gg_L \|\alpha\|_\infty\). Therefore (8.17) is equal to

\[
Nd - h + m \int_{\alpha \in \mathcal{M}} \hat{G}(\alpha)(\alpha \cdot \mathbf{v})\hat{F}(-NL^T \alpha)\ d\alpha + O_{K,L,P}(N^{-\eta K+O(1)}),
\]

\(^7\)We stress that the notation \(A_L\) is unrelated to the parameter \(A\) from this section.
which, after the obvious manipulations, equals

\[
\frac{1}{N^{h-m}} \frac{1}{\text{vol}(\mathbb{R}^h/\Gamma_{\xi,e})} \left( \int_{x \in \mathbb{R}^h} F(x/N)G(Lx + v) \, dx \right) + O_{K,L,P}(N^{-\eta K + O(1)}).
\]

Fixing suitably small \( \eta \) and \( \gamma \), and combining the contribution from the trivial, minor, and major arc, we deduce that

\[
\frac{1}{N^{h-m}} \sum_{\substack{n \in \mathbb{Z}^h \in \mathbb{R}^h \cap \Gamma_{\xi,e} \ \forall j}} F(n/N)G(Ln + v)
\]

(8.18)

\[= \frac{1}{N^{h-m}} \frac{1}{\text{vol}(\mathbb{R}^h/\Gamma_{\xi,e})} \int_{x \in \mathbb{R}^h} F(x/N)G(Lx + v) + O_{K,L,P}(N^{-\eta K + O(1)}).
\]

By adjusting the implied constant appropriately, the error term from (8.18) is \( O_{K,L,P}(N^{-K}) \) for all positive real \( K \). The final observation is that, considering the definition of the local factor \( a_{e,\tilde{r}} \) in (8.2) and the fact that we assumed \( a_{e,\tilde{r}} \) is non-zero,

\[
\frac{1}{\text{vol}(\mathbb{R}^h/\Gamma_{\xi,e})} = a_{e,\tilde{r}}.
\]

The lemma follows. \( \square \)

The following estimate will also be needed.

**Lemma 8.6.** Under the same hypotheses as Lemma 8.1, for all positive \( K \)

\[
\frac{1}{N^h} \sum_{\substack{n \in \mathbb{Z}^h \cap \mathbb{R}^h \cap \Gamma_{\xi,e} \ \forall j \leq d}} F(n/N) = \frac{a_{e,\tilde{r}}}{N^h} \int_{x \in \mathbb{R}^h} F(x/N) \, dx + O_{K,L,P}(N^{-K}),
\]

(8.19)

where \( a_{e,\tilde{r}} \) is as in (8.2).

**Proof.** By applying Poisson summation, the left-hand side of (8.19) is equal to

\[
\frac{N^{d-h}}{\text{vol}(\mathbb{R}^h/\Gamma_{\xi,e})} \sum_{c \in \Gamma_{\xi,e}^*} \hat{F}(Nc)e(x \cdot c),
\]

(8.20)

where \( x \) and \( \Gamma_{\xi,e}^* \) are as in (8.5). By applying estimates (8.7) and (8.9), one shows that the main term of (8.19) comes from the \( c = 0 \) term above. After the obvious manipulations, this concludes the lemma. \( \square \)
9 The linear inequalities condition

In [13], the key notion of pseudorandomness is the so-called ‘linear forms condition’ (see Definition 6.2 of that paper). The upshot is that in order to understand the number of solutions to a particular linear equation in primes, it is enough to understand the number of solutions to certain auxiliary linear equations weighted by a sieve weight \( \nu \). In this paper an analogous philosophy holds. Indeed we will show that, in order to understand the number of solutions to a particular linear inequality in primes, it is enough to understand the number of solutions to certain auxiliary linear inequalities weighted by a sieve weight \( \nu \).

Let us proceed with the formal definition. The reader is reminded that

\[
W_j = \prod_{p \leq w_j(N)} p
\]

(see Section 4).

**Definition 9.1** (Linear inequalities condition). Let \( m, d \) be natural numbers, and let \( L : \mathbb{R}^d \rightarrow \mathbb{R}^m \) be a linear map. For each natural number \( N \), let \( \nu_N : \mathbb{Z} \rightarrow \mathbb{R} \) be a function. We say that the family of functions \( (\nu_N)_{N=1}^{\infty} \) is \((L, \nu)\)-pseudorandom if the following holds: For all positive constants \( C \) and for all sets of parameters \( P \), for all compactly supported smooth functions \( F : \mathbb{R}^d \rightarrow [0, 1] \) and \( G : \mathbb{R}^m \rightarrow [0, 1] \) such that \( F, G \in \mathcal{C}(P) \), and for all functions \( w_1, \ldots, w_d : \mathbb{N} \rightarrow [0, \infty] \) that each satisfy \( w_j(n) \rightarrow \infty \) as \( n \rightarrow \infty \) and \( w_j(n) \leq w(n) \) for all \( n \), for all \( v \in \mathbb{R}^m \) satisfying \( \|v\|_{\infty} \leq CN \), and for functions \( f_1, \ldots, f_d : \mathbb{Z} \rightarrow \mathbb{R} \) such that each \( f_j \) equals either \( \nu_N \) or \( \Lambda_{Z/W_j} \),

\[
T_{F,G,N}(f_1, \ldots, f_d) = T_{F,G,N}(\Lambda_{Z/W_1}, \ldots, \Lambda_{Z/W_d}) + o(1)
\]

(9.1)

as \( N \rightarrow \infty \), where the \( o(1) \) term may depend on the family \( (\nu_N)_{N=1}^{\infty} \), on \( C, L, P \), and on the functions \( w_1, \ldots, w_d \).

**Remark 9.2.** Equation (9.1) might seem to be a slightly curious formulation of a pseudorandomness principle, as it does not claim that the weight \( \nu_N \) behaves like the constant 1 function but rather behaves like the local von Mangoldt function. However, referring to Remark 7.4, let us reiterate the comment that we are not performing the \( W \)-trick in the same manner as [13].

The aim of this section is to introduce a sieve weight \( \nu^\gamma_{N,w} \), and to prove that it is \((L, \nu)\)-pseudorandom for a large class of linear maps \( L \). We begin by introducing the sieve weight from [13, Appendix D].
Definition 9.3 (Smooth sieve weight). Let $N$ be a natural number, $\gamma$ be a positive real, and define $R := N^\gamma$. Let $\rho \in \mathcal{C}(\emptyset)$ be the smooth 1-supported function fixed in Section 4. Define the function $\Lambda_{\rho,R,2} : \mathbb{Z} \to \mathbb{R}_{\geq 0}$ by the formula

$$
\Lambda_{\rho,R,2}(n) := (\log R)^2 \left( \sum_{d|n} \mu(d) \rho \left( \frac{\log d}{\log R} \right)^2 \right),
$$

for non-negative integers $n$, and then by the obvious extension to negative integers.

We now define the family of majorants themselves.

Definition 9.4 (Pseudorandom majorant). Let $N$ be a natural number, let $\gamma$ be a positive real, and let $R := N^\gamma$. Define the constant $c_{\rho,2} := \int_0^\infty |\rho'(x)|^2 \, dx$. Then define the weight $\nu_{N,w}^\gamma : \mathbb{Z} \to \mathbb{R}_{\geq 0}$ by

$$
\nu_{N,w}^\gamma(n) := \frac{1}{2c_{\rho,2}} \Lambda_{\rho,R,2}(n) + \frac{1}{2} \Lambda_{\mathbb{Z},/WZ}(n).
$$

Note that $\nu_{N,w}^\gamma$ also depends on $\rho$, but we suppress that dependence from the notation (as we fixed $\rho$ in Section 4).

We now state our main new result on the pseudorandomness of this sieve weight.

Theorem 9.5 (Pseudorandomness of sieve weights). Let $m, d$ be natural numbers, with $d \geq m + 2$. Let $L : \mathbb{R}^d \to \mathbb{R}^m$ be a surjective linear map, and suppose that $L \notin V^*_\text{degen}(m, d)$ and that the coefficients of $L$ are algebraic. Assume that $\gamma$ is a positive parameter that is small enough in terms of $L$. Then $\nu_{N,w}^\gamma$ is $(L, \omega)$-pseudorandom.

Temporarily dropping the convention that $\omega(N) = \max(1, \log \log \log N)$, we speculate that the following general result holds.

Conjecture 9.6 (Pseudorandomness conjecture). Let $m, d$ be natural numbers, with $d \geq m + 2$. Let $L : \mathbb{R}^d \to \mathbb{R}^m$ be a surjective linear map, and suppose that $L \notin V^*_\text{degen}(m, d)$. Then there is some value of $\gamma$ and some function $\omega : \mathbb{N} \to \mathbb{R}_{\geq 0}$, satisfying $\omega(N) \to \infty$ as $N \to \infty$, for which $\nu_{N,w}^\gamma$ is $(L, \omega)$-pseudorandom.

Unfortunately we have not been able to resolve Conjecture 9.6, but we strongly believe it to be true. If $d$ is large enough in terms of $m$, then the analytic methods of Parsell (see [18] and Appendix B) can be used to show that $\nu_{N,w}^\gamma$ is $(L, \omega)$-pseudorandom without any algebraicity assumptions. But these methods seem
harder to apply in the range \( d \geq m + 2 \), and we have not been able to establish
the appropriate mean value estimate. Resolving Conjecture 9.6 would, after a
straightforward adaptation of the methods of this paper, enable one to remove the
algebraicity assumption from Theorem 1.7 and Theorem 1.16.

**Remark 9.7.** The proof of Theorem 9.5 is the only moment during the proof
of the main theorems Theorem 1.7 and Theorem 1.16 when we use the fact that
the coefficients of the original linear map \( L \) are algebraic. Furthermore, we will
ultimately only ever appeal to the linear inequalities condition for a certain fi-
nite collection of linear maps, which includes the original linear map \( L \) itself as
well as some auxiliary linear maps that are generated from applications of the
Cauchy–Schwarz inequality. Since only the diophantine approximation properties
of algebraic numbers are used (witness Lemma 8.4 and [21, Lemma E.1]), and
since these properties are satisfied by almost all real numbers, one may show that
Theorems 1.7 and 1.16 remain true for some explicit set of maps \( L \) that has full
Lebesgue measure.

To demonstrate our approach to proving Theorem 9.5, we first give the argu-
ment under the simplifying additional assumption that \( L \) is purely irrational (see
Definition 5.2).

**Lemma 9.8.** Suppose that \( F, G, L, v \) and the functions \( w_1, \ldots, w_d \) all satisfy
the conditions in Definition 9.1. Suppose in addition that \( L \) is surjective, purely
irrational, and has algebraic coefficients. Then for all positive \( K \) we have

\[
T_{F,G,N}^{L,v}(\Lambda_{Z/W_1 Z}, \ldots, \Lambda_{Z/W_d Z}) = J + O_{K,L,P}(N^{-K})
\]

where \( J \) is the singular integral

\[
J := \frac{1}{N^{d-m}} \int_{x \in \mathbb{R}^d} F(x/N) G(Lx + v) \, dx.
\]

**Proof.** We have the identity

\[
\Lambda_{Z/W_j Z}(n) = \frac{W_j}{\varphi(W_j)} \sum_{e_j \mid n, \, e_j \mid W_j} \mu(e_j).
\]

Then the expression \( T_{F,G,N}^{L,v}(\Lambda_{Z/W_1 Z}, \ldots, \Lambda_{Z/W_d Z}) \) is equal to

\[
\left( \prod_{j=1}^d \frac{W_j}{\varphi(W_j)} \right) \frac{1}{N^{d-m}} \sum_{e_1, \ldots, e_d, \, e_j \mid W_j, \forall j \leq d} \left( \prod_{j=1}^d \mu(e_j) \right) \sum_{n \in \mathbb{Z}^d, \, e_j \mid n, \forall j \leq d} F(n/N) G(Ln + v)
\]

\[
= \left( \prod_{j=1}^d \frac{W_j}{\varphi(W_j)} \right) \sum_{e_1, \ldots, e_d, \, e_j \mid W_j, \forall j \leq d} \left( \prod_{j=1}^d \mu(e_j) \right) \left( J \left( \prod_{j=1}^d e_j \right)^{-1} + O_{K,L,P}(N^{-K}) \right),
\]
by applying Lemma 8.1 to the inner sum, where in the statement of that lemma we take \( h = d \), the map \( \Xi : \mathbb{R}^h \to \mathbb{R}^d \) to be the identity, and \( \mathbf{r} = \mathbf{0} \). The local factor \( \alpha_{e, \mathbf{r}} \) is equal to \( \prod_{j \leq d} e_j^{-1} \) in this instance.

Sum the error term in (9.6) over all \( e_j \). The bound \( W_j = O(\log \log N) \) that comes from the prime number theorem controls the resulting error term (with room to spare), and the main term of (9.3) follows from the identity

\[
\left( \prod_{j=1}^d \frac{W_j}{\varphi(W_j)} \right) \sum_{e_1, \ldots, e_d \in \mathbb{N}} \mu(e_j) \frac{1}{e_j(W_j)} = 1. \tag{9.7} \]

To finish the proof of Theorem 9.5 (in the case when \( L \) is purely irrational, that is) it now suffices to show that

\[
T_{L, v}^{L,v}(f_1, \ldots, f_d) = J + O_{L, P, \gamma}(\log^{-\Omega(1)} N), \tag{9.8}
\]

where each \( f_j \) is either \( \nu_{N, w}^y \) or \( \Lambda_{\mathbb{Z}, W_j} \). By multiplying out the left-hand side of (9.8), we see that it is sufficient to prove that

\[
\frac{1}{N^{d-m}} \sum_{n \in \mathbb{Z}^d} \left( \prod_{j=1}^d v_j(n_j) \right) F(n/N)G(Ln + v) = J + O_{L, P, \gamma}(\log^{-\Omega(1)} N) \tag{9.9}
\]

where each \( v_j \) equals either \( c_{\rho, 2}^{-1} \Lambda_{\rho, R, 2} \) or \( \Lambda_{\mathbb{Z}, W_j} \) (recall \( R := N^y \)).

After our analysis in Lemma 8.1, it turns out that the estimate (9.9) will follow almost immediately from the sieve calculation performed in [13, Theorem D.3]. To describe the details, it will be useful to introduce the following notation. Let

\[
S := \{ j \leq d : v_j = c_{\rho, 2}^{-1} \Lambda_{\rho, R, 2} \}
\]

and

\[
S' := \{ j \leq d : v_j = \Lambda_{\mathbb{Z}, W_j} \}.
\]

We may assume that \( S \neq \emptyset \), as otherwise the estimate (9.9) follows from the estimate (9.3).

Each \( v_j \) may be expressed as a divisor sum, using either Definition 9.3 or expression (9.5). Doing this, and swapping orders of summations, we have that the left-hand side of expression (9.9) is equal to

\[
\left( \prod_{j \in S} \frac{W_j}{\varphi(W_j)} \right) c_{\rho, 2}^{-|S|}(\log R)^{|S|} \sum_{(e, k) \in [N^y \times \mathbb{N}_k]} \left( \prod_{j \in S} \mu(e_j, k) \right) \left( \frac{\log e_j}{\log R} \right) \left( \prod_{j \in S'} \mu(e_j) \right) \frac{1}{N^{d-m}} \sum_{n \in \mathbb{Z}^d} \frac{F(n/N)G(Ln + v)}{e_j | n_j | j \leq d} \tag{9.10}
\]
where \( R = N^\gamma \), and if \( j \in S \) we write \( e_j \) for the least common multiple \([e_{j,1}, e_{j,2}]\). Using the compact support of the function \( \rho \), when analysing the inner sum one may assume that each \( e_j \) is at most \( N^{2\gamma} \).

We apply Lemma 8.1. Therefore, provided \( \gamma \) is small enough,

\[
(9.11) \quad \frac{1}{N^{d-m}} \sum_{\substack{n \in \mathbb{Z}^d \cap \mathbb{R}^d \ni [e_j] \forall j \leq d}} F(n/N)G(Ln + v) = J\left( \prod_{j=1}^{d} e_j \right)^{-1} + O_{L,P,\gamma}(N^{-20}),
\]

as in (9.6). By the bounds on \( W_j \) and \( e_j \), the error term from (9.11) may be summed over all \( e_j \) and remain acceptable. We also have the identity

\[
(9.12) \quad \left( \prod_{j \in S} W_j \right) \sum_{(e_j)_{j \in \mathbb{N}^d} \ni [e_j] \forall j \in S} \left( \prod_{j \in S} \mu(e_j) \right) = 1.
\]

Therefore expression (9.9) would follow from the asymptotic

\[
(9.13) \quad \sum_{(e_{j,k})_{j,k \in \mathbb{N}^d \times [2]} \ni [e_{j,k}] \forall k \in [2]} \left( \prod_{j \in S} \mu(e_{j,k}) \rho \left( \frac{\log e_{j,k}}{\log R} \right) \right) \left( \prod_{j \in S} e_j \right)^{-1} = c_{\rho,2}^{[S]} + O(\log^{-1/20} R).
\]

But this is just expression D.4 of [13], applied to the identity map \( \Psi : \mathbb{R}^{[S]} \rightarrow \mathbb{R}^{[S]} \). Note that the quantity \( X \) in expression D.4 of [13] is zero, as if \( \psi_1, \ldots, \psi_{|S|} : \mathbb{R}^{[S]} \rightarrow \mathbb{R} \)

are the linear maps given by \( \psi_j(x) := x_j \) for all \( j \leq |S| \) then there are no primes \( p \) for which there exist two forms \( \psi_i \) and \( \psi_j \) that are linearly dependent modulo \( p \). This proves (9.9), and hence resolves Theorem 9.5 in the case when \( L \) is purely irrational.

We now present the detailed proof of Theorem 9.5 in full generality.

**Proof of Theorem 9.5.** Let \( u \) be the rational dimension of \( L \) (see Definition 5.2). Apply Lemma 5.6 to both the expression \( T_{F,G,N}^L(f_1, \ldots, f_d) \) and the expression \( T_{F,G,N}^L(\Lambda_{\mathbb{Z}/W_1\mathbb{Z}}, \ldots, \Lambda_{\mathbb{Z}/W_d\mathbb{Z}}) \). Writing \( h := d - u \), where \( u \) is the rational dimension of \( L \), and renaming \( m - u \) as \( m \), \( L' \) as \( L \), \( \nu' \) as \( v \), and \( G_F \) as \( G \), we see that it suffices to prove the following theorem.

**Theorem 9.9.** Let \( N, d, h \) be natural numbers, and let \( m \) be a non-negative integer. Suppose that \( d \geq h \geq m + 2 \). Let \( C, \gamma \) be positive parameters, and let \( P \) be a set of additional parameters. Let \( L : \mathbb{R}^h \rightarrow \mathbb{R}^m \) be a surjective purely
irrational linear map with algebraic coefficients, and let $\Xi : \mathbb{R}^h \rightarrow \mathbb{R}^d$ be an injective linear map with integer coefficients. Assume that $\gamma$ is small enough in terms of $L$. Let $v \in \mathbb{R}^m$ be a vector with $\|v\|_{\infty} \leq CN$, and let $\tilde{r} \in \mathbb{Z}^d$ be a vector with $\|\tilde{r}\|_{\infty} \leq CN$. Let $F : \mathbb{R}^d \rightarrow [0, 1]$ and $G : \mathbb{R}^m \rightarrow [0, 1]$ be in $C(P)$. Let $w_1, \ldots, w_d : \mathbb{N} \rightarrow \mathbb{R}_{\geq 0}$ be functions that each satisfy $w_j(n) \rightarrow \infty$ as $n \rightarrow \infty$ and $w_j(n) \leq w(n)$ for all $n$.

These conditions will be referred to as ‘the hypotheses of Theorem 9.9’.

Then, if $\Xi$ has finite Cauchy–Schwarz complexity,

\begin{equation}
T_{F,G,N}(f_1, \ldots, f_d)
= T_{F,G,N}(\Lambda_{Z/W_1Z}, \ldots, \Lambda_{Z/W_dZ}) + O_{C,L,P,\Xi,\gamma}(\min_j w_j(N)^{-1/2})
\end{equation}

where each $f_j$ equals either $\nu^\gamma_{N,w}$ or $\Lambda_{Z/W_jZ}$.

**Proof of Theorem 9.9.** Let $\Xi : \mathbb{R}^h \rightarrow \mathbb{R}^d$ have coordinate maps $(\xi_1, \ldots, \xi_d)$. Let

\begin{equation}
\mathcal{S}_{\tilde{r}} := \prod_p \frac{1}{p^h} \sum_{m \in [p]^h} \prod_{j=1}^d \Lambda_{Z/pZ}(\xi_j(m) + \tilde{r}_j)
\end{equation}

be the singular series, where $\tilde{r}_j$ denotes the $j^{th}$ coordinate of $\tilde{r}$. Let

\begin{equation}
J_{\tilde{r}} := \frac{1}{Nh-m} \int_{x \in \mathbb{R}^h} F((\Xi(x) + \tilde{r})/N)G(Lx + v) \, dx
\end{equation}

be the singular integral.

**Lemma 9.10.** Under the hypotheses of Theorem 9.9, if $\Xi$ has finite Cauchy–Schwarz complexity then the singular series and singular integral satisfy the bounds

\[ \mathcal{S}_{\tilde{r}} \ll_{\Xi} 1 \]

and

\[ J_{\tilde{r}} \ll_{C,L,\Xi} \text{Rad}(F)^{h-m} \text{Rad}(G)^m \|G\|_{\infty}. \]

The reader may find the definition of $\text{Rad}(F)$ and $\text{Rad}(G)$ in Section 3.

**Proof.** Since $\Xi$ has finite Cauchy–Schwarz complexity, no two of the forms $\xi_1, \ldots, \xi_d$ are parallel. Hence by [13, Lemma 1.3] the singular series $\mathcal{S}_{\tilde{r}}$ converges, and the size may be bounded by a constant depending only on $\Xi$.

The bound on $J_{\tilde{r}}$ follows directly from Lemma A.1. □

We continue with the following lemma, which is a more general version of Lemma 9.8.
Lemma 9.11. Under the hypotheses of Theorem 9.9 we have, for every positive real $K$,

\[
T_{F,G,N}^{L,v,\Xi,\bar{r}}(\Lambda_{Z/W_1Z}, \ldots, \Lambda_{Z/W_dZ})
\]

\begin{equation}
= \left( \frac{1}{(\max W_j)^h} \sum_{m \in [\max W_j]^h} \prod_{j=1}^d \Lambda_{Z/W_jZ}(\xi_j(m) + \bar{r}_j) \right) J_{\bar{r}} + O_{C,K,L,P,\Xi}(N^{-K}).
\end{equation}

If $\Xi$ has finite Cauchy–Schwarz complexity, then

\begin{equation}
T_{F,G,N}^{L,v,\Xi,\bar{r}}(\Lambda_{Z/W_1Z}, \ldots, \Lambda_{Z/W_dZ}) = \mathcal{S}_{\bar{r}} J_{\bar{r}} + O_{C,L,P,\Xi}((\min_{j} w_j(N))^{-1}).
\end{equation}

Proof. We proceed similarly to the proof of Lemma 9.8. First we have

\[
T_{F,G,N}^{L,v,\Xi,\bar{r}}(\Lambda_{Z/W_1Z}, \ldots, \Lambda_{Z/W_dZ})
\]

is equal to

\begin{equation}
\left( \prod_{j=1}^d \frac{W_j}{\varphi(W_j)} \right) \frac{1}{N^{h-m}} \sum_{e_1, \ldots, e_d \in \mathbb{Z}_+^d \setminus \emptyset} \left( \prod_{j=1}^d \mu(e_j) \right) \times \sum_{\substack{n \in \mathbb{Z}^h \\
 \text{ such that } e_j\xi_j(n) + \bar{r}_j \forall j \leq d}} F((\Xi(n) + \bar{r})/N) G(Ln + v)
\end{equation}

\begin{equation}
= \left( \prod_{j=1}^d \frac{W_j}{\varphi(W_j)} \right) \sum_{e_1, \ldots, e_d \in \mathbb{Z}_+^d \setminus \emptyset} \left( \prod_{j=1}^d \mu(e_j) \right) \alpha_{e,\bar{r}} J_{\bar{r}} + O_{C,K,L,P,\Xi}(N^{-K}),
\end{equation}

by applying Lemma 8.1 to the inner sum, where

\[
\alpha_{e,\bar{r}} = \lim_{M \to \infty} \frac{1}{M^h} \sum_{n \in [M]^h} \prod_{j=1}^d 1_{e_j(\xi_j(n) + \bar{r}_j)}.
\]

If $m = 0$, one should apply Lemma 8.6 in place of Lemma 8.1.

By using the identity (9.5) again one obtains

\begin{equation}
= \lim_{M \to \infty} \frac{1}{M^h} \sum_{n \in [M]^h} \prod_{j=1}^d \Lambda_{Z/W_jZ}(\xi_j(n) + \bar{r}_j)
\end{equation}

\begin{equation}
= \frac{1}{(\max W_j)^h} \sum_{n \in [\max W_j]^h} \prod_{j=1}^d \Lambda_{Z/W_jZ}(\xi_j(n) + \bar{r}_j).
\end{equation}

This settles the first part of the lemma.
For the second part, by the Chinese Remainder Theorem we have that (9.20) is equal to
\[
(9.21) \quad \left( \prod_{p \leq \min_j w_j} \frac{1}{p^h} \sum_{m \in [p^h]} \Lambda_{Z/pZ}(\zeta_j(m) + \tilde{r}_j) \right) \times \left( \prod_{\min_j w_j < p \leq \max_j w_j} \frac{1}{p^h} \sum_{m \in [p^h]} \prod_{j \leq d} \Lambda_{Z/pZ}(\tilde{\zeta}_j(m) + \tilde{r}_j) \right),
\]
where \(\prod^*\) denotes the product over those \(j \leq d\) for which \(p \leq w_j\).

Since \(\Xi\) has finite Cauchy–Schwarz complexity there is no pair of forms \(\zeta_i\) and \(\zeta_j\) that are parallel. Therefore we may apply the analysis of local factors in [13, Lemma 1.3] to conclude that the first bracket in (9.21) is equal to \(\mathcal{S} \tilde{r}_j (1 + O_{\Xi_1}(\min_j w_j(N))^{1/2})\), and that the second bracket is equal to \((1 + O_{\Xi_1}(\min_j w_j(N))^{1/2})\). Combining these bounds with Lemma 9.10 gives the second part of the present lemma.

**Remark 9.12.** As we intimated earlier, in Remark 1.17, one can use Lemma 9.11 to establish an asymptotic expression for \(T_{L^y,F,G,N}(\Lambda_{Z/WZ}, \ldots, \Lambda_{Z/WZ})\) in the general case. Indeed, one applies the rational parametrisation process of Lemma 5.6 and then the asymptotic in Lemma 9.11 to obtain
\[
T_{L^y,F,G,N}(\Lambda_{Z/WZ}, \ldots, \Lambda_{Z/WZ}) = \sum_{\tilde{r} \in \tilde{R}} \mathcal{S} \tilde{r} J_{\tilde{r}} + o_{C,L,p}(1).
\]

Now, Theorem 9.9 will be settled if we can show that the left-hand side of (9.14) enjoys the same asymptotic expression as the one present in (9.18). By multiplying out the left-hand side of (9.14), we see that it is sufficient to prove the following lemma.

**Lemma 9.13.** Under the hypotheses of Theorem 9.9, if \(\Xi\) has finite Cauchy–Schwarz complexity then
\[
(9.22) \quad \frac{1}{N^{h-m}} \sum_{n \in \mathbb{Z}^d} \left( \prod_{j=1}^d \nu_j(n_j) \right) F((\Xi(n) + \tilde{r})/N) G(Ln + v)
= \mathcal{S} \tilde{r} J_{\tilde{r}} + O_{C,L,p,\Xi,\gamma}(\min_j w_j(N))^{-1/2},
\]
where each \(\nu_j\) equals either \(c^{-1}_{\rho,1} \Lambda_{\rho,12}\) or \(\Lambda_{Z/w_jZ}\).

**Proof of Lemma 9.13.** The first half of the proof of this lemma comprises manipulations that are very similar to those that have appeared previously in this section. Indeed, as before, it will be useful to let
\[
S := \{ j \in [d] : \nu_j = c^{-1}_{\rho,2} \Lambda_{\rho,22} \}.
\]
and

\[ S' := \{ j \leq d : v_j = \Lambda_{\mathbb{Z}/w_j\mathbb{Z}} \}. \]

We may assume that \( S \neq \emptyset \), as otherwise the estimate (9.22) follows from Lemma 9.11.

Considering (9.5) again, and expressing each \( v_j \) as a divisor sum, we have that the left-hand side of expression (9.22) is equal to

\[
\left( \prod_{j \in S'} \frac{W_j}{\varphi(W_j)} \right) c^{-|S|} (\log R)^{|S|}
\times \sum_{(e_{j,k})_{j \in S' \subseteq 1 \leq j \leq [2]}} \left( \prod_{j \in S' \subseteq 1 \leq j \leq [2]} \mu(e_{j,k}) \rho \left( \frac{\log e_{j,k}}{\log R} \right) \right) \left( \prod_{j \in S} \mu(e_{j}) \right)
\times \frac{1}{N^{h-m}} \sum_{e_j \in \mathbb{Z}^d \atop e_j | \xi_{(n) + \bar{r}} \forall j \leq d} F((\Xi(n) + \bar{r})/N)G(Ln + v)
\]

where if \( j \in S \), we write \( e_j \) for the least common multiple \([e_{j,1}, e_{j,2}]\). Using the compact support of the function \( \rho \), when analysing the inner sum one may assume that each \( e_j \) is at most \( N^{2h} \).

We apply Lemma 8.1 (or, if \( m = 0 \), we apply Lemma 8.6). Therefore

\[
\frac{1}{N^{h-m}} \sum_{e_j \in \mathbb{Z}^d \atop e_j | \xi_{(n) + \bar{r}} \forall j \leq d} F((\Xi(n) + \bar{r})/N)G(Ln + v)
\]

\[ = \alpha_{e, \bar{r}} J_{\bar{r}} + O_{C, L, P, \Xi}(N^{-20}), \]

where

\[
\alpha_{e, \bar{r}} := \lim_{M \to \infty} \frac{1}{M^h} \sum_{m \in [M]^d} \prod_{j=1}^d \frac{1}{1 + \log e_{j}((\zeta(m) + \bar{r}_j)).}
\]

Therefore expression (9.22) (and hence the entirety of Theorem 9.9) would follow from the asymptotic expression

\[
\left( \prod_{j \in S'} \frac{W_j}{\varphi(W_j)} \right) (\log R)^{|S|}
\times \sum_{(e_{j,k})_{j \in S' \subseteq 1 \leq j \leq [2]}} \left( \prod_{j \in S' \subseteq 1 \leq j \leq [2]} \mu(e_{j,k}) \rho \left( \frac{\log e_{j,k}}{\log R} \right) \right) \left( \prod_{j \in S} \mu(e_{j}) \right) \alpha_{e, \bar{r}}
\]

\[ = c_{\rho, 2} \mathcal{E}_{\bar{r}} + O((\min w_j(N))^{-1/2}). \]
Note that this expression concerns linear forms with integer coefficients. We have removed the irrational information entirely.

Expression (9.26) follows from the sieve calculation [13, Theorem D.3], after restricting to suitable arithmetic progressions. Indeed, let

$$A := \{ a \in [W]^h : (\zeta_j(a) + \tilde{r}_j), W_j = 1 \forall j \in S' \}.$$ 

Then the left-hand side of (9.26) is equal to

$$\left( \prod_{j \in S} \frac{W_j}{\varphi(W_j)} \right) \frac{1}{Wh} \sum_{a \in A} (\log R)^{|S|} \sum_{(e_j,k) \in S \times [2]} \left( \prod_{j \in S} \mu(e_{j,k}) \rho \left( \frac{\log e_{j,k}}{\log R} \right) \right) \times \lim_{M \to \infty} \frac{1}{M^h} \sum_{m \in [M]^h} \prod_{j \in S} \lambda_j(\zeta_j(Wm + a) + \tilde{r}_j).$$

The expression following the summation in $a$ is amenable to the estimate (D.4) from [13], applied with $t = |S|$ and affine linear forms

$$\psi_j(m) := \zeta_j(Wm + a) + \tilde{r}_j, \quad j \in S.$$ 

In order to apply this estimate we note first that $t \neq 0$ (since we have previously assumed that $S' \neq \emptyset$). We also note again that, by the finite Cauchy–Schwarz complexity assumption, no two of the forms $\psi_j$ are rational multiples of each other.

So, applying the estimate (D.4) from [13] we have that the expression in (9.27) following the summation in $a$ is equal to

$$c_{\rho,2}^{[|S|]} \prod_p \beta_{p,a} + O_{C,\gamma}(e^{O(1)} \log^{-\frac{1}{20}} R),$$

where

$$\beta_{p,a} = \frac{1}{p^h} \sum_{m \in [p]^h} \prod_{j \in S} \Lambda_{p^h/p}(\zeta_j(Wm + a) + \tilde{r}_j),$$

and

$$X := \sum_{p \in P_\Xi} p^{-1/2},$$

where $P_\Xi$ is the set of ‘exceptional’ primes, i.e., those primes $p$ for which there exist $i$ and $j$ for which the forms $\zeta_i(Wm + a) + \tilde{r}_i$ and $\zeta_j(Wm + a) + \tilde{r}_j$ are affinely related modulo $p$.

**Remark 9.14.** The reader may have noticed that expression (9.28) is not exactly what was proved in estimate (D.4) of [13]. Rather than having an error
term depending on $\Xi$ and $C$, that expression has an error term depending on the linear maps

$$m \mapsto \xi_j(Wm + a) + \tilde{r}_j$$

which, one notes, have coefficients that depend on $W$ and that are therefore unbounded. Fortunately, the dependence of the error term on the size of the coefficients is only polynomial, and so any contribution from powers of $W$ may be absorbed into the $\log^{-\pi} R$ factor.

This technical manoeuvre is also required in [13] (in the application of Theorem D.3 that follows expression (D.24)), although it is not explicitly stated by the authors.

Following on from (9.28) and assuming that $N$ is large enough in terms of $\Xi$, we see that any $p \in \mathcal{P}_{\Xi}$ satisfies $p \leq w$ (as $\Xi$ has finite Cauchy–Schwarz complexity). Since

$$w(N) = \max(1, \log \log \log N),$$

the error in (9.28) is therefore $O_{C;\Xi,\gamma}(\log \log \log N)$. Furthermore, by [13, Lemma 1.3] we have $\beta_{p,a} = 1 + O(p^{-2})$, and so

$$\prod_{p > w} \beta_{p,a} = 1 + O(w^{-1}).$$

Finally, if $p \leq w$ then

$$\beta_{p,a} = \prod_{j \in S} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(\xi_j(a) + \tilde{r}_j).$$

Therefore expression (9.27), up to an error term of $O_{C;\Xi,\gamma}(w^{-1/2})$, is equal to

$$c_{\rho,2}' \left( \prod_{j \in S} \frac{W_j}{\varphi(W_j)} \right) \frac{1}{W_h} \sum_{a \in A} \prod_{p \leq w} \beta_{p,a}$$

$$= c_{\rho,2}' \left( \prod_{j \in S} \frac{W_j}{\varphi(W_j)} \right) \frac{1}{W_h} \sum_{a \in A} \prod_{j \in S} \Lambda_{\mathbb{Z}/w\mathbb{Z}}(\xi_j(a) + \tilde{r}_j)$$

$$= c_{\rho,2}' \left( \prod_{j \in S} \frac{W_j}{\varphi(W_j)} \right) \frac{1}{W_h} \sum_{a \in \mathcal{W}^h} \prod_{j \in S} \Lambda_{\mathbb{Z}/w\mathbb{Z}}(\xi_j(a) + \tilde{r}_j) \prod_{j \in S} \Lambda_{\mathbb{Z}/w\mathbb{Z}}(\xi_j(a) + \tilde{r}_j)$$

$$= c_{\rho,2}' \left( \prod_{p \leq \min_j w_j} \frac{1}{p^h} \sum_{m \in |p|} \prod_{j \in S} \Lambda_{\mathbb{Z}/w\mathbb{Z}}(\xi_j(m) + \tilde{r}_j) \times \prod_{\min_j w_j < p \leq w} \tilde{\beta}_p, \right.$$}

where

$$\tilde{\beta}_p := \frac{1}{p^h} \sum_{m \in |p|} \prod_{j \in S} \Lambda_{\mathbb{Z}/w\mathbb{Z}}(\xi_j(m) + \tilde{r}_j),$$

where $\prod^*$ denotes the product over all $j \in S'$ for which $p \leq w_j$. 
By invoking [13, Lemma 1.3] again we conclude that \( \tilde{\beta}_p = 1 + O(p^{-2}) \) and also that the first part of expression (9.29) is equal to \( c^{\rho, 2}_p \Sigma_f (1 + O, z, \min_j w_j^{-1}) \). Hence, as in the conclusion of the proof of Lemma 9.11, we conclude that expression (9.29) is equal to \( c^{\rho, 2}_p \Sigma_f + O, z, j, \min_j w_j^{-1/2} \). This establishes expression (9.26), and so Lemma 9.13 is proved.

Therefore Theorem 9.9 is resolved.

Hence Theorem 9.5 is settled as well, i.e. we conclude that the weight \( \nu_{\gamma, N, w} \) is \((L, w)\)-pseudorandom.

We finish this section by noting a corollary of the theorems above, which will be useful in its own right.

**Corollary 9.15 (Upper bound for linear inequalities).** Let \( N, m, d \) be natural numbers, with \( d \geq m + 2 \), and let \( C, \epsilon, \gamma \) be positive reals. Let \( L : \mathbb{R}^d \rightarrow \mathbb{R}^m \) be a surjective linear map with algebraic coefficients, and suppose that \( \lambda_{\text{deg}}(m, d) \) and that the coefficients of \( L \) are algebraic. Let \( u \) be the rational dimension of \( L \).

Let \( w_1, \ldots, w_d : N \rightarrow \mathbb{R}_{\geq 0} \) be functions that satisfy \( w_j(n) \rightarrow \infty \) as \( n \rightarrow \infty \) for all \( j \) and satisfy \( w_j(n) \leq w(n) \) for all \( j \) and for all \( n \). If \( \gamma \) is small enough in terms of \( L \), then for all functions \( F : \mathbb{R}^d \rightarrow [0, 1] \) supported on \([-C, C]^d\), for all functions \( G : \mathbb{R}^m \rightarrow [0, 1] \) supported on \([-\epsilon, \epsilon]^m\), and for all \( v \in \mathbb{R}^m \) satisfying \( \|v\|_\infty \leq CN \), one has

\[
T_{L, v, N}^{1, \gamma}(f_1, \ldots, f_d) \ll C, L \|G\|_\infty \epsilon^{m-u} + o(1)
\]
as \( N \rightarrow \infty \), where each \( f_j \) equals either \( \nu_{\gamma, N, w} \) or \( \Lambda_{Z, W, \gamma} \). The \( o(1) \) term may depend on \( C, L, \epsilon, \gamma \), and the choice of functions \( w_1, \ldots, w_d \).

**Proof.** Using Lemma 3.1, replace both \( F \) and \( G \) by compactly supported smooth majorants \( F_1 \) and \( G_1 \) for which

\[
1_{[-C, C]^d} \leq F_1 \leq 1_{[-2C, 2C]^d}
\]
and

\[
1_{[-\epsilon, \epsilon]^m} \leq G_1 \leq 1_{[-2\epsilon, 2\epsilon]^m}.
\]
We have \( F_1 \in \mathcal{C}(C_1) \) and \( G_1 \in \mathcal{C}(\epsilon) \). Then, by Theorem 9.5,

\[
T_{F, G, N}^{L, v}(f_1, \ldots, f_d) \ll T_{F_1, G_1, N}^{L, v}(f_1, \ldots, f_d) = T_{F_1, G_1, N}(\Lambda_{Z, W, \gamma}^1, \ldots, \Lambda_{Z, W, \gamma}^d) + o(1),
\]
where the error term may depend on \( C, L, \epsilon, \gamma \), and the functions \( w_1, \ldots, w_d \).
In Remark 9.12, we noted that
\[ T_{F,G,N}^{L,v}(\Lambda_{\mathbb{Z}/w_1\mathbb{Z}}, \ldots, \Lambda_{\mathbb{Z}/w_d\mathbb{Z}}) = \sum_{r \in \tilde{R}} \tilde{\mathcal{S}} r \tilde{J} r + o(1), \]
where the error term depends on the parameters mentioned above, and where \( \tilde{\mathcal{S}} r \) and \( \tilde{J} r \) are of the form (9.15) and (9.16). The corollary then follows from the bounds in Lemma 9.10. □

This result is to be compared with the following statement.

**Lemma 9.16 (Weak upper bound).** Let \( N, m, d \) be natural numbers, with \( d \geq m \), and let \( C, \varepsilon \) be positive parameters. Let
\[ L : \mathbb{R}^d \rightarrow \mathbb{R}^m \]
be a surjective linear map. Then, for all functions
\[ F : \mathbb{R}^d \rightarrow [0, 1] \]
supported on \([-C, C]^d\), for all functions
\[ G : \mathbb{R}^m \rightarrow [0, 1] \]
supported on \([-\varepsilon, \varepsilon]^m\), for all \( \mathbf{v} \in \mathbb{R}^m \), and for all functions \( f_1, \ldots, f_d : \mathbb{Z} \rightarrow \mathbb{R} \),
\[ T_{F,G,N}^{L,v}(f_1, \ldots, f_d) \ll_{C,L,\varepsilon} \| G \|_{\infty} \sup_{j \leq d, |n| \leq \text{Rad}(F)N} |f_j(n)|^d. \]

The bound in Lemma 9.16 is weaker than the bound in Corollary 9.15, but has the advantage of holding for all surjective maps \( L \), which is a situation that will be needed later.

**Proof.** This is essentially identical to Lemma 3.2 of [21]. Indeed, one sees immediately that
\[ T_{F,G,N}^{L,v}(f_1, \ldots, f_d) \ll \frac{1}{N^{d-m}} \left( \sup_{j \leq d, |n| \leq \text{Rad}(F)N} |f_j(n)|^d \right) \times \sum_{n \in [-CN,CN]^d, \|Ln+v\|_{\infty} \leq \varepsilon} 1. \]

Since \( L \) is surjective, without loss of generality we may assume that the first \( m \) columns of \( L \) form an invertible matrix. If the variables \( n_{m+1} \) to \( n_d \) are fixed, there are only \( O_{C,\varepsilon,L}(1) \) possible choices for \( n_1, \ldots, n_m \) for which the inequality \( \|Ln+v\|_{\infty} \leq \varepsilon \) is satisfied. Summing over \( n_{m+1} \) to \( n_d \), the lemma follows. □
Part 4
The structure of inequalities

Before embarking upon this part of the argument, we remind the reader of the following basic notion from functional analysis. A linear map

$$L : (V, \| \cdot \|_V) \rightarrow (W, \| \cdot \|_W)$$

between two normed spaces will be called a bounded operator if there exists a constant $C_L$ such that for all $v \in V$ one has $\|Lv\|_W \leq C_L\|v\|_V$. It is a standard fact that all linear maps between two finite-dimensional normed spaces are bounded.

10 An alternative formulation

So far all of our theorems and lemmas have been phrased in terms of linear inequalities that are written in the form $T_{F,G,N}(f_1, \ldots, f_d)$. In Section 14 the auxiliary inequalities will appear in a different form, but, as is shown in Lemma 10.1 below, these different forms are more-or-less equivalent. The statement of this lemma is unfortunately rather technical, but the proof is straightforward. The reader may wish in the first instance to consider the special case in which $l = 0$ and $\Phi$ is injective.

Lemma 10.1 (Alternative formulation). Let $m, d, l$ be natural numbers, with $d \geq m$, and let $C, \sigma, \eta$ be positive parameters. Let $P$ be another set of parameters. Let $k$ be a non-negative integer, and suppose that $\eta$ is small enough in terms of $m, d, k$ and $l$. Let $\Phi : \mathbb{R}^{d-k} \rightarrow \mathbb{R}^d$ be a linear map, and suppose that $k = \dim \ker \Phi$. Let

$$I : \mathbb{R}^d \rightarrow [0, 1] \quad \text{and} \quad H : \mathbb{R}^{d-k+l} \rightarrow [0, 1]$$

be smooth functions, where $\text{Rad}(I) \leq \eta$ and $\text{Rad}(H) \leq C$. Assume that the Lipschitz constant of $H$ is at most $\sigma^{-1}$ and assume further that $H, I \in \mathcal{C}(P)$. Then:

1. There exists a surjective linear map $L : \mathbb{R}^d \rightarrow \mathbb{R}^m$ such that $\ker L = \text{Im} \Phi$ and $\|L\|_\infty = O_\Phi(1)$. If $\Phi$ has algebraic coefficients then $L$ can be chosen to have algebraic coefficients.

2. For any $L$ satisfying part (1), if $\Phi$ has finite Cauchy–Schwarz complexity then $L \notin V^*_\text{degen}(m, d)$.

3. For any $L$ satisfying part (1), if $\eta$ is small enough in terms of $L$ and $\Phi$ then there exist smooth functions

$$F : \mathbb{R}^{d+l} \rightarrow \mathbb{R}_{\geq 0} \quad \text{and} \quad G : \mathbb{R}^m \rightarrow \mathbb{R}_{\geq 0},$$
with \( F \in \mathcal{C}(P, \Phi), G \in \mathcal{C}(L, P, \Phi) \) and \( \text{Rad}(G) \ll_L \eta \), such that for all \( v \in \mathbb{R}^l \), \( z \in \mathbb{R}^d \), and natural numbers \( N \),

\[
\frac{1}{N^k} \int_{x \in \mathbb{R}^{d-m+k}} I(z - \Phi(x))H((v, x)/N) \, dx = F((v, z)/N)G(Lz) + E(z, N),
\]

(10.1)

where \( E(z, N) \) is an error term of size at most

\[
O_{C, \Phi}(\eta \sigma^{-1} N^{-1} 1_{[0, O_{C, \Phi}(N)]}(\|z\|_{\infty}) 1_{[0, O_L(\eta)]}(\|Lz\|_{\infty})).
\]

**Proof.** Part (1) of the lemma is immediate. Indeed, one has the quotient map \( \pi : \mathbb{R}^d \rightarrow \mathbb{R}^d/\text{Im} \Phi \). Choosing an isomorphism \( \iota : \mathbb{R}^d/\text{Im} \Phi \cong \mathbb{R}^m \), we may define \( L := \iota \circ \pi \). If \( \Phi \) has algebraic coefficients, then choosing such an \( \iota \) with algebraic coefficients gives a suitable \( L \) with algebraic coefficients.

For part (2), suppose that \( \Phi \) has finite Cauchy–Schwarz complexity. If \( L \) were in \( V_{\text{deg}}(m, d) \), then there would exist \( i, j \leq d \) and a real number \( \lambda \) for which \( e_i^* - \lambda e_j^* \) is non-zero and \( e_i^* - \lambda e_j^* \in L^*((\mathbb{R}^m)^*), \) which would imply \( e_i^* - \lambda e_j^* \in (\text{Im} \Phi)^0 \), which would imply that \( \Phi \) has infinite Cauchy–Schwarz complexity, contradicting the hypothesis.

It remains to prove part (3). Let \( \{u^{(1)}, \ldots, u^{(k)}\} \subset \mathbb{R}^{d-m+k} \) be an orthonormal basis for \( \text{ker} \Phi \), and extend this to an orthonormal basis \( \{u^{(1)}, \ldots, u^{(d-m+k)}\} \) for \( \mathbb{R}^{d-m+k} \). Then define the linear map \( \Psi : \mathbb{R}^{d-m+k} \rightarrow \mathbb{R}^{d-m+k} \) by

\[
\Psi(y) = \sum_{j=1}^{d-m+k} y_j u^{(j)}.
\]

By changing variables, we have that the left-hand side of (10.1) is equal to

\[
\frac{1}{N^k} \int_{y \in \mathbb{R}^{d-m+k}} I(z - \Phi(\Psi(y)))H((v, \Psi(y))/N) \, dy,
\]

which equals

\[
\frac{1}{N^k} \int_{y \in \mathbb{R}^{d-m+k}} I(z - \Phi(\Psi(0, y^{d-m+k}_{k+1})))
\times H((v, (\Psi(y^k_1, 0) + \Psi(0, y^d_{k+1}))) / N) \, dy.
\]

(10.2)

Recall, from Section 4, that we use the notation \( y^k \) to refer to the vector \( (y_1, \ldots, y_k)^T \in \mathbb{R}^k \), etcetera.

We make some observations. Firstly, we observe that (10.2) is equal to 0 unless \( \|z\|_{\infty} = O_{C, \Phi}(N) \). Indeed, if \( \|z\|_{\infty} \geq C_1 N \), then for all \( y_{k+1}, \ldots, y_{d-m+k} \) that give a
non-zero contribution to (10.2) we have

\[ \| \Phi(\Psi(0, y_{k+1}^{d-m+k})) \|_\infty \geq \frac{1}{2} C_1 N, \]

if \( \eta \) is small enough. This means that

\[ \| \Psi(0, y_{k+1}^{d-m+k}) \|_\infty \gg \Phi C_1 N, \]

which if \( C_1 \) is large enough in terms of \( C \) and \( \Phi \) means that

\[ H((v, (\Psi(y_k^k, 0) + \Psi(0, y_{k+1}^{d-m+k}))/N) = 0 \]

for all \( y_1, \ldots, y_k \). [Note that \( \Psi(y_k^k, 0) \) and \( \Psi(0, y_{k+1}^{d-m+k}) \) are orthogonal.]

Secondly, we observe that

\[ \| z - \Phi(\Psi(0, y_{k+1}^{d-m+k})) \|_\infty \ll \eta \]

for all \( y_{k+1}, \ldots, y_{d-m+k} \) that give a non-zero contribution to the integral (10.2). Write \( z = z_1 + z_2 \), where \( z_1 \in \text{Im } \Phi \) and \( z_2 \in (\text{Im } \Phi)^\perp \). By orthogonality, we conclude that

\[ \| z_1 - \Phi(\Psi(0, y_{k+1}^{d-m+k})) \|_\infty \ll \eta. \]

Since \( (\Phi|_{(\ker \Phi)^\perp})^{-1} : \text{Im } \Phi \rightarrow (\ker \Phi)^\perp \) is a bounded linear map, this in turn means that

\[ \| (\Phi|_{(\ker \Phi)^\perp})^{-1}(z_1) - \Psi(0, y_{k+1}^{d-m+k}) \|_\infty \ll \Phi \eta. \]

Since \( H \) is Lipschitz, with Lipschitz constant at most \( \sigma^{-1} \), this all means that (10.2) is equal to

\[
(10.3) \quad \left( \int_{y_{k+1}^{d-m+k} \in \mathbb{R}^{d-m}} I(z - \Phi(\Psi(0, y_{k+1}^{d-m+k}))) \, dy_{k+1}^{d-m+k} \right) \\
\times \frac{1}{N^k} \left( \int_{y_1^k \in \mathbb{R}^d} H((v, (\Psi(y_k^k, 0) + (\Phi|_{(\ker \Phi)^\perp})^{-1}(z_1)))/N) \, dy_1^k \right),
\]

plus an error of size at most

\[ O_{C, \Phi}(\eta \sigma^{-1} N^{-1} \mathbf{1}_{[0, O_C(N)]}(\|z\|_\infty) \mathbf{1}_{[0, O(\eta)]}(\text{dist}(z, \text{Im } \Phi))). \]

We proceed to analyse the terms of (10.3) separately. Firstly, by shifting the variables \( y_{k+1}, \ldots, y_{d-m+k} \) we see that the first bracket of (10.3) is equal to

\[
(10.4) \quad \int_{y_{k+1}^{d-m+k} \in \mathbb{R}^{d-m}} I(z_2 - \Phi(\Psi(0, y_{k+1}^{d-m+k}))) \, dy_{k+1}^{d-m+k}.
\]
Now let $L : \mathbb{R}^d \to \mathbb{R}^m$ be any surjective linear map that satisfies $\ker L = \text{Im } \Phi$. Note that $L|_{\text{Im } \Phi^\perp}$ is an injective linear map, and thus (10.4) is equal to

$$
\int_{y_k+1}^{d-m+k} \left( L(L|_{\text{Im } \Phi^\perp})^{-1} - \Phi(\psi(0, y_k^{d-m+k})) \right) dy_k^{d-m+k}.
$$

Differentiating inside the integral, one sees that this expression is equal to $G(Lz_2)$ for some smooth compactly supported function $G : \mathbb{R}^m \to \mathbb{R}_{\geq 0}$ satisfying $G \in \mathcal{C}(L, P, \Phi)$. Moreover, $G$ is supported on $[O_L(\eta), O_L(\eta)]^m$, since $(L|_{\text{Im } \Phi^\perp})^{-1}Lz_2$ and $\Phi(\psi(0, y_k^{d-m+k}))$ are orthogonal. Note that $Lz_2 = Lz$, so the expression is equal to $G(Lz)$. We move to the second term of (10.3). Choose $\iota : \text{Im } \Phi \to \mathbb{R}^{d-m}$ to be an isomorphism with $\|\iota(x)\|_\infty \approx \Phi \|x\|_\infty$. Then the second term of (10.3) is equal to $F_1((v, \iota(z_1))/N)$ for some smooth function $F_1 : \mathbb{R}^{d-m+l} \to \mathbb{R}_{\geq 0}$ satisfying $F_1 \in \mathcal{C}(P, \Phi)$. Note that $F_1$ is indeed compactly supported, since $(\Phi|_{\ker \Phi^\perp})^{-1}(z_1)$ and $\psi(y_k^{1}, 0)$ are orthogonal vectors.

In summary, we have shown that (10.1) is equal to

$$
F_1((v, \iota(z_1))/N)G(Lz),
$$

plus an error of size

$$
O_{C, \Phi}(\eta \sigma^{-1}N^{-1})_{[0, O_{C, \Phi}(N)]}(\|z\|_\infty)_{[0, O_L(\eta)]}(\text{dist}(z, \text{Im } \Phi)).
$$

By the construction of $L$, this error is bounded by

$$
O_{C, \Phi}(\eta \sigma^{-1}N^{-1})_{[0, O_{C, \Phi}(N)]}(\|z\|_\infty)_{[0, O_L(\eta)]}(\|Lz\|_\infty).
$$

The term $F_1((v, \iota(z_1))/N)G(Lz)$ is not quite of the required form, since $F_1((v, \iota(z_1))/N)$ is not compactly supported as a function of $z$. However, it may be easily massaged into this form. Indeed, from the above discussion we know that $G(Lz) \neq 0$ implies that $\|z_2\|_\infty \leq C_1 \eta$, for some constant $C_1$ that satisfies $C_1 = O_L(\Phi(1))$. Let $b : \mathbb{R} \to [0, 1]$ be a 1/2-supported function (in the sense of Definition 4.1), and let $B : \mathbb{R}^d \to [0, 1]$ be defined by $B(x) = \prod_{j=1}^d b(x_j)$. Then let $F : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ be defined by

$$
F(v, z) := F_1((v, \iota(z_1))B(z_2).
$$

Then $F \in \mathcal{C}(P, \Phi)$, and if $\eta \leq 1/2C_1$ we have

$$
F_1((v, \iota(z_1))/N)G(Lz) = F((v, z)/N)G(Lz).
$$

The lemma is proved.
This reformulation allows us to deduce Corollary 10.3 below. This is a corollary of Theorem 9.5 and is the result on inequalities and sieve weights that we will actually use in Section 15. In order to state this inequality, we introduce the following convention.

**Definition 10.2 (Convolution).** If

\[ f : \mathbb{Z} \rightarrow \mathbb{R} \]

has finite support, and

\[ g : \mathbb{R} \rightarrow [0, 1] \]

is a measurable function, we may define the convolution \((f * g)(x) : \mathbb{R} \rightarrow \mathbb{R}\) by

\[(f * g)(x) := \sum_{n \in \mathbb{Z}} f(n)g(x - n).\]

Recall from Section 4 that, for some positive parameter \(\eta\), the function \(\chi : \mathbb{R} \rightarrow [0, 1]\) denotes a fixed \(\eta\)-supported function.

**Corollary 10.3 (Switching functions).** Let \(N, m, d\) be natural numbers, with \(d \geq m + 2\), and let \(k\) be a non-negative integer. Let \(C, \gamma, \eta\) be positive parameters, and let \(P\) be a set of further parameters. Suppose that \(\eta\) is small enough in terms of \(m, d,\) and \(k\). Let

\[(\varphi_1, \ldots, \varphi_d) = \Phi : \mathbb{R}^{d-m+k} \rightarrow \mathbb{R}^d\]

be a linear map with algebraic coefficients, and suppose that \(k = \text{dim ker} \Phi\). Suppose that \(\Phi\) has finite Cauchy–Schwarz complexity. Let

\[H : \mathbb{R}^{d-m+k} \rightarrow [0, 1]\]

be a smooth function in \(C(P)\). For \(j \leq d\), let \(w_1, \ldots, w_d\) be any functions with \(w_j(n) \leq w(n)\) for all \(n\) and for which \(w_j(n) \rightarrow \infty\) as \(n \rightarrow \infty\). For each \(j \leq d\) let the function \(f_j : \mathbb{Z} \rightarrow \mathbb{R}_{\geq 0}\) be equal to either \(v_{N, w_j}^\gamma\) or \(\Lambda_{\mathbb{Z}/w_j\mathbb{Z}}\). Let \(r \in \mathbb{R}^d\) be any vector satisfying \(\|r\|_{\infty} \leq CN\).

Then, if \(\gamma\) is small enough in terms of \(\Phi\), the expression

\[
\frac{1}{N^{d-m+k}} \int_{x \in \mathbb{R}^{d-m+k}} \left( \prod_{j=1}^{d} (f_j * \chi)(\varphi_j(x) - r_j) \right) H(x/N) \, dx
\]

is independent of the choices of the functions \(f_j\), up to an error of size \(o(1)\) as \(N \rightarrow \infty\). This \(o(1)\) term may depend on \(C, P, \Phi, \eta, \gamma\), and on the functions \(w_1, \ldots, w_d\).
Proof. Expanding out the definition of $f_j \ast \chi$, one observes that the left-hand side of (10.6) is equal to

$$\frac{1}{N^{d-m+k}} \sum_{n_1, \ldots, n_d \in \mathbb{Z}} \left( \prod_{j=1}^{d} f_j(n_j) \right)$$

(10.7)

$$\times \int_{x \in \mathbb{R}^{d-m+k}} \left( \prod_{j=1}^{d} \chi(\varphi_j(x) - n_j - r_j) \right) H(x/N) \, dx.$$

By applying Lemma 10.1 to the inner integral, we get a surjective linear map $L : \mathbb{R}^d \to \mathbb{R}^m$ with algebraic coefficients, and smooth functions $F : \mathbb{R}^d \to \mathbb{R}^\geq0$ and $G : \mathbb{R}^m \to \mathbb{R}^\geq0$ supported on $[-O_P, \Phi(1), O_P, \Phi(1)]^d$ and $[-O_\Phi(\eta), O_\Phi(\eta)]^m$ respectively and with $F, G \in C(P, \Phi, \eta)$, such that (10.7) is equal to

$$\frac{1}{N^{d-m}} \sum_{n_1, \ldots, n_d \in \mathbb{Z}} \left( \prod_{j=1}^{d} f_j(n_j) \right) F(n/N) G(Ln + Lr)$$

(10.8)

plus an error of size

$$O_{P, \Phi} \left( \frac{1}{N^{d-m+1}} \sum_{n_1, \ldots, n_d \ll \epsilon \Phi N} \prod_{j=1}^{d} f_j(n_j) \right).$$

(10.9)

Furthermore, $L \not\in V^*_{\text{degen}}(m, d)$.

Now apply Theorem 9.5 to the main term (10.8). As written this theorem applies to functions $F$ and $G$ that take values in $[0, 1]$, but by the obvious rescaling we may nonetheless apply the theorem to the present functions $F$ and $G$. This shows immediately that (10.8) is independent of the particular choices of $f_1, \ldots, f_d$, up to an error of size $o(1)$. The $o(1)$ term has the appropriate dependencies.

For the error term (10.9), we apply the upper bound in Corollary 9.15. This shows that (10.9) is $o(1)$, so may be absorbed into the $o(1)$ term above. Corollary 10.3 is proved. □

An upper bound in this setting will also be convenient.

Corollary 10.4. Under the same hypotheses as Corollary 10.3,

$$\frac{1}{N^{d-m+k}} \int_{x \in \mathbb{R}^{d-m+k}} \left( \prod_{j=1}^{d} (f_j \ast \chi)(\varphi_j(x) + r_j) \right) H(x/N) \, dx \ll 1,$$

(10.10)

where the implied constant may depend on $C, P, \Phi, \eta, \gamma$, and on the functions $w_1, \ldots, w_d$.

Proof. Proceed as in the previous proof to get to expression (10.8). Then apply the upper bound in Corollary 9.15. □
11 Variation in parameters

This section will be devoted to proving Lemma 11.1 below. This technical lemma shows that the number of solutions to certain inequalities, weighted by the local von Mangoldt function, is a quantity that behaves well when the underlying parameters are perturbed. The slightly esoteric notation, in which we introduce a dimension $d$ only to consider $x \in \mathbb{R}^{d-1}$, is designed to correspond to the moment in Section 15 in which this lemma will be applied.

**Lemma 11.1.** Let $d, l, N, s$ be natural numbers, with $d \geq 2$, and let $C, \eta$ be positive parameters. Let $(\phi_1, \ldots, \phi_l) = \Phi : \mathbb{R}^{d-1} \to \mathbb{R}^l$ and $(\psi_1, \ldots, \psi_l) = \Psi : \mathbb{R}^{s+2} \to \mathbb{R}^l$ be linear maps with algebraic coefficients. Let $P$ be a set of parameters, and let $b \in \mathcal{C}(C, P, \eta, \Phi, \Psi)$ be an arbitrary smooth function. Let $w^* : \mathbb{N} \to \mathbb{R}$ be a function such that $w^*(n) \to \infty$ as $n \to \infty$ and $w^*(n) \leq w(n)$ for all $n$. Let $\alpha \in \mathbb{R}^l$ be a vector satisfying $\|\alpha\|_{\infty} \leq CN$. For $y \in \mathbb{R}^{s+1}$, define

$$Q_{\alpha, N}(y) := \frac{1}{N^{d-1}} \int_{x \in \mathbb{R}^{d-1}} \left( \prod_{j \leq l} (\Lambda_{\mathbb{Z}/W^* \mathbb{Z}} \ast \chi)(\phi_j(x) + \Psi(y) + a_j) \right) b((x, y)/N) \, dx,$$

where $a_j$ is the $j^{th}$ coordinate of $\alpha$. Then, if $\eta$ is sufficiently small in terms of $\Phi$ and $\Psi$, there is a function $f_1 : \mathbb{Z}^l \to \mathbb{C}$, satisfying $\|f_1\|_{\infty} \ll (\log \log W^*)^{O(1)}$, such that

$$Q_{\alpha, N}(y) = b_{\alpha, N}(y/N) \sum_{\|k\|_{\infty} \leq (\log \log W^*)^{O(1)}} f_1(k) e\left( \frac{k \cdot (\Psi(y) + \alpha)}{W^*} \right) + o_{C, P, \eta, \Phi, \Psi}(1).$$

Here $b_{\alpha, N} \in \mathcal{C}(C, P, \eta, \Phi, \Psi)$, though it may also depend on $\alpha$ and $N$.

None of the methods required to prove this lemma will be particularly deep, but the technical manoeuvres will be a little intricate. In particular, we will need to apply the approximation in Lemma 10.1 multiple times within the same argument.

The proof of Lemma 11.1 will require the preliminary result below, namely Lemma 11.2. To state this lemma, we define a metric on $\mathbb{R}^d / K \mathbb{Z}^d$ by the formula

$$\|x\|_{\mathbb{R}^d / K \mathbb{Z}^d} := \min_{n \in K \mathbb{Z}^d} \|x - n\|_{\infty}.$$ 

Lipschitz constants of functions $\overline{\Phi} : \mathbb{R}^d / K \mathbb{Z}^d \to \mathbb{R}$ will be considered with respect to this metric.
Lemma 11.2. Let \( d, m, K \) be natural numbers, and let \( \eta, \sigma \) be positive parameters. Let \( S : \mathbb{R}^d \rightarrow \mathbb{R}^m \) be a surjective linear map with integer coefficients, and let \( G : \mathbb{R}^m \rightarrow [0, 1] \) be a Lipschitz function supported on \([-\eta, \eta]^m\), with Lipschitz constant at most \( \sigma^{-1} \). Let \( \mathfrak{F} : S\mathbb{Z}^d \rightarrow \mathbb{R} \) be any function for which
\[
\mathfrak{F}(Sx) = \mathfrak{F}(Sx + Sn)
\]
for all \( x \in \mathbb{Z}^d \) and all \( n \in K\mathbb{Z}^d \).

For each \( a \in \mathbb{R}^d \), define \( \tilde{a} \in \mathbb{Z}^d \) to be some vector with integer coordinates for which
\[
\|S(\tilde{a} - a)\|_\infty = \min_{n \in \mathbb{Z}^d} \|S(n - a)\|_\infty.
\]

Then, provided \( \eta \) is small enough in terms of \( S \), the function
\[
(11.2) \quad a \mapsto \mathfrak{F}(S\tilde{a})G(S(\tilde{a} - a))
\]

- depends only on the value of \( a \) modulo \( K\mathbb{Z}^d \);
- is Lipschitz when viewed as a function on \( \mathbb{R}^d / K\mathbb{Z}^d \), with Lipschitz constant at most
\[
O_S(\|\mathfrak{F}\|_\infty (\eta^{-1} + \sigma^{-1})).
\]

Remark 11.3. The expression \( \min_{n \in \mathbb{Z}^d} \|S(n - a)\|_\infty \) is well-defined, since \( S\mathbb{Z}^d \) is a lattice.

Proof. To prove the first part of the lemma, let \( a \in \mathbb{R}^d \) and first suppose that there is a unique vector \( x \in S\mathbb{Z}^d \) for which
\[
\|x - Sa\|_\infty = \min_{n \in \mathbb{Z}^d} \|Sn - Sa\|_\infty.
\]
In this case, by the uniqueness of \( x \), we have \( x = S\tilde{a} \). By translation, we know that
\[
S(a + n) - Sn = x
\]
for all \( n \in \mathbb{Z}^d \), and hence
\[
S(a + n) - Sn = S\tilde{a}
\]
for all \( n \in \mathbb{Z}^d \). Hence
\[
G(S(\tilde{a} - a)) = G(S(a + n - (a + n))),
\]
and so the function
\[
a \mapsto G(S(\tilde{a} - a))
\]
depends only on the value of $a$ modulo $\mathbb{Z}^d$. Furthermore, if $n \in K\mathbb{Z}^d$, 
$$\tilde{\mathcal{F}}(S(\tilde{a} + n)) = \tilde{\mathcal{F}}(S\tilde{a} + Sn) = \tilde{\mathcal{F}}(S\tilde{a}),$$
by the invariance properties of $\tilde{\mathcal{F}}$. Hence the function (11.2) only depends on the value of $a$ modulo $K\mathbb{Z}^d$.

Now suppose that there were two distinct vectors $x_1, x_2 \in S\mathbb{Z}^d$ for which
$$\|x_i - Sa\|_\infty = \min_{n \in \mathbb{Z}^d} \|Sn - Sa\|_\infty$$
for $i = 1, 2$. Then in fact $G(S(\tilde{a} - a)) = 0$. Indeed, if this were not the case then we would have $\|x_1 - x_2\|_\infty \leq O(\eta)$, which is impossible if $\eta$ is small enough, since $x_1$ and $x_2$ are two distinct elements of $\mathbb{Z}^m$. By translation, we may also conclude that
$$G(S(a + n - (a + n))) = 0 \quad \text{for all } n \in \mathbb{Z}^d.$$ So again, the function (11.2) depends only on the value of $a$ modulo $K\mathbb{Z}^d$.

Regarding the second part of the lemma, the idea of the proof is similar to the above. Indeed, the only aspect of the function (11.2) that could lead to a large Lipschitz constant is of course the term $S\tilde{a}$, which could, one fears, jump sharply for small changes in $a$. However, when such jumps occur, the function $G(S(\tilde{a} - a))$ is always equal to zero.

Let us proceed with the full proof. Indeed, let $a_0, a_1 \in \mathbb{R}^d$ and suppose first that
$$\|a_0 - a_1\|_{\mathbb{R}^d/K\mathbb{Z}^d} \leq \eta.$$ By choosing suitable coset representatives, without loss of generality we may assume that
$$\|a_0 - a_1\|_{\mathbb{R}^d/K\mathbb{Z}^d} = \|a_0 - a_1\|_\infty.$$ Then either $S\tilde{a}_0 = S\tilde{a}_1$ or $S\tilde{a}_0 \neq S\tilde{a}_1$. If $S\tilde{a}_0 = S\tilde{a}_1$ then
$$|G(S(\tilde{a}_0 - a_0)) - G(S(\tilde{a}_1 - a_1))| \leq \sigma^{-1} \|Sa_0 - Sa_1\|_\infty \ll S \sigma^{-1} \|a_0 - a_1\|_\infty$$
(11.3)
$$\ll S \sigma^{-1} \|a_0 - a_1\|_{\mathbb{R}^d/K\mathbb{Z}^d}.$$ Therefore
$$|\tilde{\mathcal{F}}(S\tilde{a}_0)G(S(\tilde{a}_0 - a_0)) - \tilde{\mathcal{F}}(S\tilde{a}_1)G(S(\tilde{a}_1 - a_1))|$$
$$= |\tilde{\mathcal{F}}(S\tilde{a}_0)||G(S(\tilde{a}_0 - a_0)) - G(S(\tilde{a}_1 - a_1))|$$
$$\ll S \|\tilde{\mathcal{F}}\|_\infty \sigma^{-1} \|a_0 - a_1\|_{\mathbb{R}^d/K\mathbb{Z}^d}.$$ That resolves the lemma in this case.
If on the other hand $S\tilde{a}_0 \neq S\tilde{a}_1$, we may conclude that both

\begin{equation}
\|S\tilde{a}_0 - Sa_0\|_\infty \geq 10\eta
\end{equation}

and

\begin{equation}
\|S\tilde{a}_1 - Sa_1\|_\infty \geq 10\eta.
\end{equation}

Indeed, if $\|S\tilde{a}_0 - Sa_0\|_\infty \leq 10\eta$, say, then

$$\|S\tilde{a}_0 - Sa_1\|_\infty \leq 10\eta + \|Sa_0 - Sa_1\|_\infty \ll \eta.$$ 

If $\eta$ is small enough, this implies that $S\tilde{a}_0$ must be the unique element of $SZ^d$ for which

$$\|S\tilde{a}_0 - Sa_1\|_\infty = \min_{n \in Z^d} \|Sn - Sa_1\|_\infty,$$

and hence that

$$S\tilde{a}_0 = S\tilde{a}_1,$$

contradicting the assumption.

If $\eta$ is small enough, expressions (11.4) and (11.5) imply that

\begin{equation}
G(S\tilde{a}_0 - Sa_0) = G(S\tilde{a}_1 - Sa_1) = 0,
\end{equation}

and so

$$|\tilde{\Phi}(S\tilde{a}_0)G(S(a_0 - a_0)) - \tilde{\Phi}(S\tilde{a}_1)G(S(a_1 - a_1))| = 0.$$

That resolves the lemma in this case.

The only remaining case to consider is when

$$\|a_0 - a_1\|_{R^d/KZ^d} \geq \eta.$$ 

In this case we bound the Lipschitz constant very crudely, as $O(\eta^{-1}\|\tilde{\Phi}\|_\infty\|G\|_\infty)$, which is $O(\|\tilde{\Phi}\|_\infty\eta^{-1})$, since $\|G\|_\infty \leq 1$. This settles the lemma. \hfill \Box

We are now ready to prove Lemma 11.1.

**Proof of Lemma 11.1.** For this proof we make the following conventions. Any implied constant may depend on $C$, $\Phi$ and $\Psi$, and we will use the notation $b$, $b_1$, $b_2$ etc. to denote a function in $\mathcal{C}(C, P, \eta, \Phi, \Psi)$, that may change from line to line.

The first part of the proof will involve establishing an asymptotic formula for $Q_{a,N}(y)$, namely the expression $Q_{a,N}(y) = \mathcal{G}_{a,N}(y)I_{a,N}(y) + o_{P,N}(1)$ in (11.11)
below. Indeed, expanding out the definition of \( \Lambda_{\mathbb{Z}/W^*\mathbb{Z}} \) (see Definition 10.2) we have

\[
Q_{a,N}(y) = \frac{1}{N^{d-1}} \sum_{n \in \mathbb{Z}^l} \left( \prod_{j \leq l} \Lambda_{\mathbb{Z}/W^*\mathbb{Z}}(n_j) \right) 
\times \int_{x \in \mathbb{R}^{d-1}} \chi(n - \Phi(x) - \Psi(y) - a) b((x, y)/N) \, dx,
\]

where \( \chi : \mathbb{R}^l \to [0, 1] \) is defined by \( \chi(z) := \prod_{j \leq l} \chi(z_j) \). Let \( k := \dim \text{Im} \Phi \), and note that \( k \leq d - 1 \).

The inner integral of (11.7) may be analysed using Lemma 10.1. The following table indicates which objects in (11.7) play which role in Lemma 10.1:

| Notation of Lemma 10.1 | Objects in (11.7) |
|------------------------|------------------|
| \( z \)               | \( n - \Psi(y) - a \) |
| \( v \)               | \( y \)          |
| \( \Phi \)             | \( \Phi \)       |
| \( H \)               | \( b \)          |
| \( I \)               | \( \chi \)       |

So, applying Lemma 10.1, one sees that (11.7) is equal to

\[
Q_{a,N}(y) = \frac{1}{N^{d-1}} \sum_{n \in \mathbb{Z}^l} \left( \prod_{j \leq l} \Lambda_{\mathbb{Z}/W^*\mathbb{Z}}(n_j) \right) 
\times b_1((y, n)/N)b_2(L(n - \Psi(y) - a)) + E.
\]

Here, \( L : \mathbb{R}^l \to \mathbb{R}^{l-k} \) is a surjective linear map with algebraic coefficients, that depends only on \( \Phi \), \( \text{Rad}(b_2) = O(\eta) \), and the error term \( E \) may be bounded above by

\[
\ll_P \left| \frac{1}{N^d} \sum_{n \in \mathbb{Z}^l} \left( \prod_{j \leq l} \Lambda_{\mathbb{Z}/W^*\mathbb{Z}}(n_j) \right) \right|,
\]

where the summation \( \sum^* \) denotes summation over the set

\[
\{ n \in \mathbb{Z}^l : \|n\|_\infty \ll_P N, \|Ln - L(\Psi(y)) - La\|_\infty \ll_P \eta \}.
\]

The error term \( E \) is easy to bound. Indeed, by Lemma 9.16, expression (11.9) may be bounded by \( O_P(N^{k-d}(\log \log W^*)^d) \). Since \( k \leq d - 1 \), this is an \( o_P(1) \) error.

It remains to analyse the main term in (11.8), which we will do with the help of Lemma 5.6. The reader is invited to consult Section 5 for the statement of this result, and for the definitions of rational map, rational dimension, etcetera.
Now, let \( u \) be the rational dimension of \( L \), and let \( \Theta : \mathbb{R}^{l-k} \to \mathbb{R}^u \) be a rational map for \( L \) with algebraic coefficients. Then, there exists an injective linear map \((\xi_1, \ldots, \xi_l) = \Xi : \mathbb{R}^{l-u} \to \mathbb{R}^l\) with integer coefficients, satisfying \( \Xi \mathbb{Z}^{l-u} = \mathbb{Z}^l \cap \ker \Theta L \), and a vector \( \tilde{r}(a, y) \in \mathbb{Z}^l \), such that the main term of (11.8) is equal to

\[
(11.10) \quad \frac{1}{N^{d-1}} \sum_{m \in \mathbb{Z}^{l-u}} \left( \prod_{j=1}^{l} \Lambda_{\mathbb{Z} \times \mathbb{Z}}(\xi_j(n) + \tilde{r}(a, y), j) \right) \times b_1((\mathbf{y}, \Xi(n) + \tilde{r}(a, y))/N) b_2(L(\Xi(n) + \tilde{r}(a, y) - \Psi(y) - a)),
\]

where \( \tilde{r}(a, y)_j \) is the \( j \)th coordinate of \( \tilde{r}(a, y) \). Note how we’ve appealed to part (11) of Lemma 5.6 for the particular form of the argument of \( b_2 \). Note also how, since \( \eta \) is sufficiently small, we have been able to apply part (10) of the lemma to establish that \( \tilde{R} \) consists of a single element \( \tilde{r}(a, y) \).

Moreover, from part (10) of the lemma again, we have that \( \tilde{r}(a, y) \) is an element of \( \mathbb{Z}^l \) for which

\[
\|\Theta L(\tilde{r}(a, y) - \Psi(y) - a)\|_\infty = \min_{m \in \mathbb{Z}^l} \|\Theta L(m - \Psi(y) - a)\|_\infty.
\]

From part (9) of Lemma 5.6, letting \( \{e_1, \ldots, e_{l-u}\} \) be the standard basis vectors of \( \mathbb{R}^{l-u} \), we have a set

\[
\mathcal{B} = \{x_i : i \leq u\} \cup \{\Xi(e_j) : j \leq l - u\}
\]

which is a lattice basis for \( \mathbb{Z}^l \) and for which \( \{\Theta L e_i : i \leq u\} \) is a lattice basis for \( \Theta L \mathbb{Z}^l \). Letting \( U = \text{span}(\{x_i : i \leq u\}) \), we have that \( \tilde{r}(a, y) \in U \).

By applying the first part of Lemma 9.11 to expression (11.10), one immediately derives

\[
(11.11) \quad Q_{a,N}(y) = \mathcal{S}_{a,N}(y) I_{a,N}(y) + o_{P,\eta}(1),
\]

where

\[
(11.12) \quad \mathcal{S}_{a,N}(y) := \frac{1}{(W^*)^{l-u}} \sum_{m \in [W^*]^{l-u}} \prod_{j=1}^{l} \Lambda_{\mathbb{Z} \times \mathbb{Z}}(\xi_j(m) + \tilde{r}(a, y)_j)
\]

and \( I_{a,N}(y) \) is equal to

\[
(11.13) \quad \frac{1}{N^{d-1}} \int_{x \in \mathbb{R}^{l-u}} b_1((\mathbf{y}, \Xi(x) + \tilde{r}(a, y))/N) \times b_2(L\Xi(x) + L\tilde{r}(a, y) - L(\Psi(y)) - La) \, dx.
\]
Note that

\[(11.14) \quad \dim \ker L \Xi \leq \dim \ker L = k \leq d - 1.\]

The remainder of the proof of Lemma 11.1 will consist of analysing expressions (11.12) and (11.13) for \(G_{a,N}(y)\) and \(I_{a,N}(y)\).

We begin with \(I_{a,N}(y)\), aiming for expression (11.18). Letting \(V = \text{Im} \Xi\), we have that \(\mathbb{R}^L = U \oplus V\). For any vector \(v \in \mathbb{R}^L\) let \(v|_U\) and \(v|_V\) be the components in \(U\) and \(V\) respectively. Then we have that

\[(11.15) \quad \| \tilde{r}(a, y) - \Psi(y) - a |_U \|_\infty = O(1),\]

since

\[\| \Theta L (\tilde{r}(a, y) - \Psi(y) - a) \|_\infty = O(1).\]

By the bound on the Lipschitz constant of \(b_1\), we may replace \(b_1((y, \Xi(x) + \tilde{r}(a, y))/N)\) with \(b_1((y, \Xi(x) + \Psi(y)|_U + a|_U)/N)\) in (11.13), up to an error of \(O_{P, \eta}(N^{-1})\). Also, note that

\[\frac{1}{N^{d-1}} \int_{x \in \mathbb{R}^L \subset \mathbb{R}} b_2(L \Xi(x) + L \tilde{r}(a, y) - L(\Psi(y)) - La) \, dx = O_{P, \eta}(N^{d \dim \ker L \Xi - d + 1}),\]

by Lemma A.1. This is \(O_{P, \eta}(1)\), since \(\dim \ker L \Xi \leq d - 1\) by (11.14). Therefore we may replace (11.13) by the expression

\[(11.16) \quad \frac{1}{N^{d-1}} \int_{x \in \mathbb{R}^L \subset \mathbb{R}} b_1((y, \Xi(x) + \Psi(y)|_U + a|_U)/N) \times b_2(L \Xi(x) + L \tilde{r}(a, y) - L(\Psi(y)) - La) \, dx,\]

plus an error of size \(o_{P, \eta}(1)\).

The expression (11.16) is in a form that is amenable to Lemma 10.1. The following table indicates which objects from our present discussion play which role in the notation of Lemma 10.1:

| Notation of Lemma 10.1 | Objects related to (11.16) |
|------------------------|---------------------------|
| \(z\)                  | \(L \tilde{r}(a, y) - L(\Psi(y)) - La\) |
| \(v\)                  | \(y\)                      |
| \(\Phi\)               | \(-L \Xi\)                 |
| \(L\)                  | \(\Theta\)                 |
| \((v, x) \mapsto H(v, x)\) | \((y, x) \mapsto b_1(y, \Xi(x) + \frac{\Psi(y)}{N} + \frac{a}{N})\) |
| \(I\)                  | \(b_2\)                    |
This is a valid application of Lemma 10.1, since ker \( \Theta = \mathrm{Im} L \Xi \) and the final two functions in the right-hand column are compactly supported smooth functions of their arguments (as \( \Xi \) is injective, \( \Xi(x) \in V \), and \( V \) is an algebraic complement to \( U \)). Recalling that \( \Theta \) has algebraic coefficients, by the third part of Lemma 10.1 we may therefore replace (11.16) by an expression of the form

\[
(11.17) \quad b_1((y, L(\tilde{r}(a, y) - \Psi(y) - a))/N)b_2(\Theta L(\tilde{r}(a, y) - \Psi(y) - a)) + o_{P, \eta}(1)
\]

where \( \mathrm{Rad}(b_2) = O(\eta) \).

The argument of the function \( b_1 \) above doesn’t depend smoothly on \( y \), but this may be easily rectified. Indeed, by (11.15) and the fact that \( b_1 \) is Lipschitz and \( b_2 \) is bounded, (11.17) is equal to

\[
b_1((y, -L(\Psi(y)|_V + a|_V))/N)b_2(\Theta L(\tilde{r}(a, y) - \Psi(y) - a)) + o_{P, \eta}(1),
\]

e.g., is equal to

\[
(11.18) \quad b_{a, N, 1}(y/N)b_2(\Theta L(\tilde{r}(a, y) - \Psi(y) - a)) + o_{P, \eta}(1),
\]

where \( \mathrm{Rad}(b_2) = O(\eta) \).

In summary then, since \( \mathcal{G}_{a, N}(y) \ll (\log \log W^*)^{O(1)} \) we have shown that

\[
Q_{a, N}(y) = \mathcal{G}_{a, N}(y)b_{a, N, 1}(y/N)b_2(\Theta L(\tilde{r}(a, y) - \Psi(y) - a)) + o_{P, \eta}(1).
\]

The function

\[
(11.20) \quad (\Psi(y) + a) \rightarrow \mathcal{G}_{a, N}(y)b_2(\Theta L(\tilde{r}(a, y) - \Psi(y) - a))
\]

is of the form considered in Lemma 11.2 in expression (11.2). Indeed, one first notes that (11.20) is a well-defined mapping, since \( \tilde{r}(a, y) \) is determined only by \( \Psi(y) + a \) and \( \mathcal{G}_{a, N}(y) \) depends on \( a \) and \( y \) only through the value of \( \tilde{r}(a, y) \) (see (11.12)). Then, one takes the map \( S \) from Lemma 11.2 to be the map \( \Theta L : \mathbb{R}^l \rightarrow \mathbb{R}^u \) here, ones takes \( K \) from that lemma to be \( W^* \) here, and one takes the map \( G \) from that lemma to be \( b_2 \) here, and one takes the map \( \tilde{\mathcal{G}} : \Theta L \mathbb{Z}^l \rightarrow \mathbb{R} \) from that lemma to be

\[
\tilde{\mathcal{G}}(x) = \frac{1}{(W^*)^{l-u}} \sum_{m \in [W^*]^{l-u}} \prod_{j=1}^{l} \Lambda_{\mathbb{Z}^l/W^* \mathbb{Z}}(\tilde{\zeta_j}(m) + (\Theta L|_U)^{-1}(x)_j)
\]

here. The definition of \( \tilde{\mathcal{G}} \) is valid since \( \Theta L|_U : U \rightarrow \mathbb{R}^u \) is indeed a bijection, and by part (9) of Lemma 5.6 we have \( (\Theta L|_U)^{-1}(\Theta L(\mathbb{Z}^l)) = \mathbb{Z}^l \cap U \). Consulting expression (11.12) for \( \mathcal{G}_{a, N}(y) \), one sees that

\[
\tilde{\mathcal{G}}(\Theta L \tilde{r}(a, y)) = \mathcal{G}_{a, N}(y)
\]
and so (11.20) is indeed of the form (11.2) as we have claimed. The only hypothesis of Lemma 11.2 that we have not already verified is the invariance of $\Phi$ under translation by elements of $\Theta L(W^*Z^l)$, but this is immediate from the definition of $\Phi$, since $(\Theta L|U)^{-1} : \mathbb{R}^d \to U$ is linear and $A_{Z/W^*Z}$ is $W^*$-periodic. Therefore, by applying Lemma 11.2, we conclude that the function (11.20) is Lipschitz on $\mathbb{R}^l/W^*Z^l$, with Lipschitz constant $O_{p,\eta}((\log \log W^*)^{O(1)})$.

The proof of Lemma 11.1 is nearly complete, since Lipschitz functions enjoy good approximation by short exponential sums. Indeed, by Lemma A.9 of [12], for all $X > 2$ there exists a function $f_1 : \mathbb{Z} \to \mathbb{C}$ such that $\|f_1\|_\infty \ll (\log \log W^*)^{O(1)}$ and

$$
\mathcal{S}_{a,N}(y)b_2(\Theta L(\bar{r}(a, y) - \Psi(y) - a))
$$

equals

$$
\sum_{\|k\|_\infty \leq X} f_1(k)e\left(\frac{k \cdot (\Psi(y) + a)}{W^*}\right) + O_{p,\eta}((\log \log W^*)^{O(1)}(\log X/X)).
$$

Then, picking $X$ to be a suitably large power of $\log \log W^*$, Lemma 11.1 follows. \qed

Part 5

The main argument

Having completed all the preparatory material, the main thrust of the proof can begin in earnest.

12 Controlling by Gowers norms

In this section we state a type of result that has become known as a ‘generalised von Neumann theorem’, which uses Gowers norms to bound the number of solutions to a diophantine inequality. For readers familiar with [13], the procedure is routine. We will then show that this result implies the main theorem (Theorem 1.16).

Theorem 12.1 (Generalised von Neumann Theorem). Let $N, m, d$ be natural numbers, satisfying $d \geq m + 2$, and let $C, \gamma, \varepsilon$ be positive parameters. Let $L : \mathbb{R}^d \to \mathbb{R}^m$ be a surjective linear map with algebraic coefficients, and assume that $L \notin V^*_{\text{deg}}(m, d)$ and that $\gamma$ is small enough (depending on $L$). Let $v \in \mathbb{R}^m$ satisfy $\|v\|_\infty \leq CN$. Let $F : \mathbb{R}^d \to [0, 1]$ and $G : \mathbb{R}^m \to [0, 1]$ be functions with Lipschitz constants at most $\sigma^{-1}$, and suppose that $F$ is supported on $[-1, 1]^d$ and $G$ is supported on $[-\varepsilon, \varepsilon]^m$. Let $f_1, \ldots, f_d : [N] \to \mathbb{R}$ be arbitrary functions, satisfying $|f_j(n)| \leq v_{N, \varepsilon}^\gamma(n)$ for all $j \leq d$ and for all $n \leq N$.\n
Then there exists an $s = O(1)$ such that, if

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

as $N \to \infty$, then

$$|T_{F,G,N}^L(f_1, \ldots, f_d)| = o(1)$$

as $N \to \infty$. The second $o(1)$ term may also depend on $C, L, \gamma, \varepsilon, \sigma$, and the rate of decay of the first $o(1)$ term.

**Proof of Theorem 1.16 assuming Theorem 12.1.** Assume the hypotheses of Theorem 1.16. By telescoping we have that

$$T_{F,G,N}^L(\Lambda', \ldots, \Lambda') - T_{F,G,N}^L(\Lambda^+_{Z/WZ}, \ldots, \Lambda^+_{Z/WZ})$$

is equal to

$$T_{F,G,N}^L(\Lambda', \ldots, \Lambda') - T_{F,G,N}^L(\Lambda^+_{Z/WZ}, \ldots, \Lambda^+_{Z/WZ}) + T_{F,G,N}^L(\Lambda', \Lambda' - \Lambda^+_{Z/WZ}, \Lambda^+_{Z/WZ}, \ldots, \Lambda^+_{Z/WZ})$$

(12.1)

$$\vdots$$

$$+ T_{F,G,N}^L(\Lambda', \ldots, \Lambda', \Lambda' - \Lambda^+_{Z/WZ}).$$

Since $F$ is supported on $[-1, 1]^d$, we may restrict the functions $\Lambda'$ and $\Lambda^+_{Z/WZ}$ to $[N]$ without altering the size of expression (12.1).

By the construction of the sieve weight $v_{N,w}'$, we have

$$|\Lambda'(n) + \Lambda^+_{Z/WZ}(n)| \ll_\gamma v_{N,w}'(n)$$

for all $n \leq N$. Therefore, after rescaling, we may apply Theorem 12.1 in this setting.

Recall that, by Lemma 7.5,

$$\|\Lambda' - \Lambda^+_{Z/WZ}\|_{U^{s+1}[N]} = \|\Lambda' - \Lambda_{Z/WZ}\|_{U^{s+1}[N]} = o(1)$$

as $N \to \infty$, for all $s \leq d - 2$. So, applying Theorem 12.1 to each term of (12.1) separately, we derive

$$|T_{F,G,N}^L(\Lambda', \ldots, \Lambda') - T_{F,G,N}^L(\Lambda^+_{Z/WZ}, \ldots, \Lambda^+_{Z/WZ})| = o_{C,L,\gamma,\varepsilon,\sigma}(1)$$

as $N \to \infty$. By fixing a suitably small value of $\gamma$, we conclude Theorem 1.16. □
13 Transferring from \( \mathbb{Z} \) to \( \mathbb{R} \)

In this section we begin the proof of Theorem 12.1. Following the programme set out in [21], our first step will be to transfer the problem from the setting of functions on \( \mathbb{Z} \) to functions on \( \mathbb{R} \).

**Definition 13.1.** Let \( N, m, d \) be natural numbers. Let \( L : \mathbb{R}^d \to \mathbb{R}^m \) be a linear map, let \( v \in \mathbb{R}^m \), and let \( F : \mathbb{R}^d \to [0, 1] \) and \( G : \mathbb{R}^m \to [0, 1] \) be compactly supported measurable functions. Then, for all bounded measurable functions \( g_1, \ldots, g_d : \mathbb{R} \to \mathbb{R} \) we define

\[
\tilde{T}_{L,v}^{F,G,N}(g_1, \ldots, g_d) := \frac{1}{N^{d-m}} \int_{x \in \mathbb{R}^d} \left( \prod_{j=1}^d g_j(x_j) \right) F(x/N)G(Lx + v) \, dx.
\]

We now state the key lemma. For the definition of \( f \ast \chi \), where \( \chi \) is the function we determined in Section 4, the reader may consult Definition 10.2.

**Lemma 13.2** (Transfer). Let \( N, m, d \) be natural numbers, with \( d \geq m + 2 \), and let \( C, \varepsilon, \gamma, \eta, \sigma \) be positive constants. Let \( L : \mathbb{R}^d \to \mathbb{R}^m \) be a surjective linear map, and let \( v \in \mathbb{R}^m \) be a vector satisfying \( \|v\|_\infty \leq CN \). Let \( F : \mathbb{R}^d \to [0, 1] \) and \( G : \mathbb{R}^m \to [0, 1] \) be compactly supported Lipschitz functions, with Lipschitz constants at most \( \sigma^{-1} \). Suppose that \( F \) is supported on \([-1, 1]^d\), and \( G \) is supported on \([-\varepsilon, \varepsilon]^m\). Then there exists some positive real number \( C_\chi \), satisfying \( C_\chi \approx 1 \), such that the following holds. Let \( f_1, \ldots, f_d : [N] \to \mathbb{R} \) be arbitrary functions that satisfy \( |f_j(n)| \leq v_j^N(n) \) for all \( j \leq d \) and for all \( n \leq N \). Assume that \( \eta \leq \min(1, \varepsilon) \) and that \( \gamma \) is small enough depending on \( L \). Then

\[
T_{L,v}^{F,G,N}(f_1, \ldots, f_d) = \frac{1}{C_\chi \gamma^d} \tilde{T}_{L,v}^{F,G,N}(f_1 \ast \chi, \ldots, f_d \ast \chi) + O(\eta \sigma^{-1}) + o(1)
\]

as \( N \to \infty \). The implied constant in the \( O(\eta \sigma^{-1}) \) term may depend on \( C, L, \) and \( \varepsilon \), and the \( o(1) \) term may depend on all these parameters together with \( \gamma \) and \( \sigma \).

**Proof.** The proof is very similar to the proof of Lemma 6.4 in [21], although we do have to insert various estimates that are only proved in this paper.

Indeed, let \( \chi : \mathbb{R}^d \to [0, 1] \) denote the function \( x \mapsto \prod_{j=1}^d \chi(x_j) \). We choose

\[
C_\chi := \frac{1}{\eta^d} \int_{x \in \mathbb{R}^d} \chi(x) \, dx.
\]

Since \( \chi \) is \( \eta \)-supported, \( C_\chi \approx 1 \). Then, expanding the definition of the convolutions \( f_i \ast \chi \),

\[
\frac{1}{C_\chi \gamma^d} \tilde{T}_{L,v}^{F,G,N}(f_1 \ast \chi, \ldots, f_d \ast \chi)
\]
equals

\[(13.3) \quad \frac{1}{N^{d-m}} \sum_{n \in \mathbb{Z}^d} \left( \prod_{j=1}^d f_j(n_j) \right) \frac{1}{C \chi \eta^d} \int_{y \in \mathbb{R}^d} F(y/N)G(Ly+v)\chi(y-n) \, dy.\]

This is equal to

\[(13.4) \quad \frac{1}{N^{d-m}} \sum_{n \in \mathbb{Z}^d \atop \|n\|_\infty \leq 2N} \left( \prod_{j=1}^d f_j(n_j) \right) \left( F(n/N) + O(\eta^{-1}N^{-1}) \right)G(Ly+v)\chi(y-n) \, dy.\]

Indeed, the inner integrand is only non-zero when \(\|y-n\|_\infty \leq \eta\), and \(F\) has Lipschitz constant \(O(\sigma^{-1})\).

Continuing, expression (13.4) is equal to

\[(13.5) \quad \frac{1}{N^{d-m}} \sum_{n \in \mathbb{Z}^d \atop \|n\|_\infty \leq 2N} \left( \prod_{j=1}^d f_j(n_j) \right) F(n/N)H(Ln + v) + E\]

where

\[H(x) = \frac{1}{C \chi \eta^d} \int_{y \in \mathbb{R}^d} \chi(y)G(x+Ly) \, dy\]

and \(E\) is a certain error, which may be bounded above by a constant times

\[(13.6) \quad \frac{\eta}{\sigma N} \frac{1}{N^{d-m}} \sum_{n \in \mathbb{Z}^d \atop \|n\|_\infty \leq 2N} \left( \prod_{j=1}^d v^{\gamma}_{N,w}(n_j) \right) H(Ln + v).\]

Let us deal with the first term of (13.5), in which we wish to replace \(H\) with \(G\).

We therefore consider

\[\left| \frac{1}{N^{d-m}} \sum_{n \in \mathbb{Z}^d} \left( \prod_{j=1}^d f_j(n_j) \right) F(n/N)(G(Ln + v) - H(Ln + v)) \right|,\]

which is

\[(13.7) \quad \leq \frac{1}{N^{d-m}} \sum_{n \in \mathbb{Z}^d} \left( \prod_{j=1}^d v^{\gamma}_{N,w}(n_j) \right) F(n/N)|G - H|(Ln + v).\]

Observe that \(\|G - H\|_\infty = O(\eta^{-1})\). Indeed,

\[G(x) - \frac{1}{C \chi \eta^d} \int_{y \in \mathbb{R}^d} G(x+Ly)\chi(y) \, dy = G(x) - \frac{1}{C \chi \eta^d} \int_{y \in \mathbb{R}^d} (G(x) + O(\eta^{-1}))\chi(y) \, dy = O(\eta^{-1}),\]
by the definition of $C_\chi$ and the Lipschitz property of $G$. The function $|G - H|$ is compactly supported, with $\text{Rad}(|G - H|) \ll \epsilon + \eta \ll \epsilon$.

Of course $|G - H|$ needn’t be smooth, but we may nonetheless apply Corollary 9.15, concluding that expression (13.7) is at most

$$O_{C, L, e}(\eta^{-1} \sigma) + o_{C, L, e, \sigma}(1).$$

Turning to the error $E$ from (13.5), we’ve already remarked that it may be bounded above by expression (13.6). Applying Corollary 9.15 again, expression (13.6) is $o(1)$ (with the appropriate dependencies on $C, L$, etc.).

The lemma then follows.

We will need to show that the operation of replacing $f$ by $f \ast \chi$ is compatible with Gowers norms.

Firstly, if $g : [-N, N] \rightarrow \mathbb{R}$ is a bounded measurable function, we define the Gowers norm over the reals $\|g\|_{U^d(\mathbb{R}, N)}$ by

$$\|g\|_{U^d(\mathbb{R}, N)} := \frac{1}{(2N)^{d+1}} \int_{(x, h) \in \mathbb{R}^{d+1}} \prod_{\omega \in \{0, 1\}^d} C_{[\omega]} g \left( x + \sum_{i=1}^{d} h_i \omega_i \right) \, dx \, dh.$$

More detail about this quantity may be found in Appendix A of [21].

Secondly, we note that $\|f\|_{U^s[N]}$ and $\|f \ast \chi\|_{U^s(\mathbb{R}, 2N)}$ may be related.

**Lemma 13.3** (Relating different Gowers norms). Let $s$ be a natural number, and assume that $\eta$ is a positive parameter that is small enough in terms of $s$. Let $N$ be a natural number, and let $f : [N] \rightarrow \mathbb{R}$ be an arbitrary function. Then we have

$$\|f \ast \chi\|_{U^s(\mathbb{R}, 2N)} \ll \eta^{\frac{d}{2s+2}} \|f\|_{U^{s+1}[N]}.$$

**Proof.** This is Lemma 6.5 of [21].

### 14 Parametrising the kernel

In this section we will convert the expression $T_{F, G, N}^{L, V}(f_1, \ldots, f_d)$ into an expression that is tailored to the subsequent manipulations. We begin with a lemma that is very similar to Proposition 8.2 of [21].

**Lemma 14.1** (Separating out the kernel). Let $N, m, d$ be natural numbers, with $d \geq m + 2$, and let $C, \epsilon, \sigma$ be positive constants. Let $L : \mathbb{R}^d \rightarrow \mathbb{R}^m$ be a surjective linear map with algebraic coefficients, and assume further that $L \notin V^*_\text{deg}(m, d)$. Let $F : \mathbb{R}^d \rightarrow [0, 1]$ be a Lipschitz function supported on $[-1, 1]^d$, with Lipschitz constant at most $\sigma^{-1}$, and let $G : \mathbb{R}^m \rightarrow [0, 1]$. 


be any function supported on $[-\varepsilon, \varepsilon]^m$. Then there exists an injective linear map $(\psi_1, \ldots, \psi_d) = \Psi : \mathbb{R}^{d-m} \to \mathbb{R}^d$ with algebraic coefficients (depending only on $L$), and a Lipschitz function $F_1 : \mathbb{R}^{d-m} \to [0, 1]$ with Lipschitz constant $O_L(\sigma^{-1})$ and with $\text{Rad}(F_1) = O_{C,L,\varepsilon}(1)$, such that, if $g_1, \ldots, g_d : \mathbb{R} \to \mathbb{R}$ are arbitrary bounded measurable functions,

$$|\tilde{T}_{F,G,N}^{L^v}(g_1, \ldots, g_d)| \ll_{L,\varepsilon} \left| \frac{1}{N^{d-m}} \int_{x \in \mathbb{R}^{d-m}} F_1(x/N) \left( \prod_{j=1}^d g_j(\psi_j(x) + a_j) \right) dx \right|,$$

where, for each $j$, $a_j$ is some real number that satisfies $a_j = O_{C,L,\varepsilon}(N)$. Furthermore, $\Psi$ has finite Cauchy–Schwarz complexity (see Definition 5.5).

**Proof of Lemma 14.1.** For ease of notation, let

$$\beta := \tilde{T}_{F,G,N}^{L^v}(g_1, \ldots, g_d).$$

Noting that $\ker L$ is a vector space of dimension $d - m$, define

$$\{u^{(1)}, \ldots, u^{(d-m)}\} \subset \mathbb{R}^d$$

to be an orthonormal basis for $\ker L$ consisting of vectors with algebraic coordinates. Then the map $(\psi_1, \ldots, \psi_d) = \Psi : \mathbb{R}^{d-m} \to \mathbb{R}^d$, defined by

$$\Psi(x) := \sum_{i=1}^{d-m} x_i u^{(i)},$$

is an injective map that parametrises $\ker L$. Furthermore $\Psi$ has finite Cauchy–Schwarz complexity, since otherwise there would exist $i \neq j \leq d$ and a real number $\lambda$ such that $e_i^* - \lambda e_j^* \in (\text{Im} \Psi)^0$, i.e. $e_i^* - \lambda e_j^* \in (\ker L)^0$. This implies that $e_i^* - \lambda e_j^* \in L((\mathbb{R}^m)^*)$, which, by definition, implies that $L \in V^*_{\text{degen}}(m, d)$, contradicting our hypotheses.

Now, extend the orthonormal basis $\{u^{(1)}, \ldots, u^{(d-m)}\}$ for $\ker L$ to an orthonormal basis $\{u^{(1)}, \ldots, u^{(d)}\}$ for $\mathbb{R}^d$. By implementing a change of basis, we may rewrite

$$\beta = \frac{1}{N^{d-m}} \int_{x \in \mathbb{R}^d} F \left( \sum_{i=1}^d x_i u^{(i)}/N \right) G \left( L \left( \sum_{i=1}^d x_i u^{(i)} \right) + v \right)$$

$$\times \left( \prod_{j=1}^d g_j(\psi_j(x)) + \sum_{i=d-m+1}^d x_i u^{(i)}_j \right) dx,$$

where $u^{(i)}_j$ is the $j^{th}$ coordinate of $u^{(i)}$. 

We wish to remove the presence of the variables \(x_{d,m+1}, \ldots, x_d\). To set this up, note that, by the choice of the vectors \(u^{(i)}\),

\[
G(L(\sum_{i=1}^{d} x_i u^{(i)}) + v) = G(L(\sum_{i=d-m+1}^{d} x_i u^{(i)}) + v).
\]

The vector \(\sum_{i=d-m+1}^{d} x_i u^{(i)}\) is in \((\ker L)^\perp\) and so, since \(L|_{(\ker L)^\perp}\) is a bounded invertible operator, \(G(L(\sum_{i=d-m+1}^{d} x_i u^{(i)}) + v)\) is equal to zero unless \((x_{d-m+1}, \ldots, x_d)^T \in D\), for some domain \(D \subseteq \mathbb{R}^m\) of diameter \(O_{L,\varepsilon}\) and satisfying

\[
\sup_{x \in D} \|x\|_\infty = O_{C,L}(N + \varepsilon).
\]

We can use this observation to bound the right-hand side of (14.3). Indeed, we have

\[
\beta \ll \text{vol } D \times \sup_{x^{d-m+1} \in D} \frac{1}{N^{d-m}} \left| \int_{(x^{d-m})_1 \in \mathbb{R}^{d-m}} F\left(\sum_{i=1}^{d} x_i u^{(i)} / N\right) \left(G\left(L\left(\sum_{i=d-m+1}^{d} x_i u^{(i)}\right) + v\right) \right. \right.
\]

\[
\times \prod_{j=1}^{d} g_j \left(\psi_j(x^{d-m}_1) + \sum_{i=d-m+1}^{d} x_i u^{(i)}_j\right) d(x^{d-m}_1) \right|.
\]

See Section 4 for explanation of \(x^{d-m}\) notation. So there exists some fixed vector \((x_{d-m+1}, \ldots, x_d)^T\) in \(D\) such that

\[
\beta \ll L, \varepsilon \frac{1}{N^{d-m}} \left| \int_{(x^{d-m})_1 \in \mathbb{R}^{d-m}} F\left(\sum_{i=1}^{d} x_i u^{(i)} / N\right) \left(G\left(L\left(\sum_{i=d-m+1}^{d} x_i u^{(i)}\right) + v\right) \right. \right.
\]

\[
\times \prod_{j=1}^{d} g_j \left(\psi_j(x^{d-m}_1) + \sum_{i=d-m+1}^{d} x_i u^{(i)}_j\right) d(x^{d-m}_1) \right|.
\]

Define the function \(F_1 : \mathbb{R}^{d-m} \rightarrow [0, 1]\) by

\[
F_1(x^{d-m}_1) := F\left(\psi(x^{d-m}_1) + \left(\sum_{i=d-m+1}^{d} x_i u^{(i)} / N\right)\right)
\]

and for each \(j\) at most \(d\) a shift

\[
a_j := \sum_{i=d-m+1}^{d} x_i u^{(i)}_j.
\]
Then

\begin{equation}
\beta \ll L, c \left| \frac{1}{N^{d-m}} \int_{x \in \mathbb{R}^{d-m}} F_1(x/N) \left( \prod_{j=1}^{d} g_j(y_j(x) + a_j) \right) dx \right|,
\end{equation}

and $F_1$ and $a_j$ satisfy the conclusions of the proposition. \(\square\)

The next proposition is essentially identical to an argument that appears in [21] at the end of Section 8 of that paper. Unfortunately that argument is not in an easily citable form, and so we have found it necessary to state and prove the precise version that we need here. For readers unfamiliar with the notion of normal form, we included a brief summary in Section 6.

**Lemma 14.2 (Parametrising by normal form).** Following on from above, there exists a \(d' = O(1)\), a linear map \((\psi_1', \ldots, \psi_d') : \mathbb{R}^d \rightarrow \mathbb{R}^d\) with algebraic coefficients that is in \(s\)-normal form for some \(s = O(1)\), and a Lipschitz function \(F_2 : \mathbb{R}^d \rightarrow [0, 1]\) with Lipschitz constant \(O_L(\sigma^{-1})\) and with \(\text{Rad}(F_2) = O_{C, L, \varepsilon}(1)\) such that

\begin{equation}
\left| \frac{1}{N^{d-m}} \int_{x \in \mathbb{R}^{d-m}} F_1(x/N) \left( \prod_{j=1}^{d} g_j(y_j(x) + a_j) \right) dx \right|
\end{equation}

is bounded above by a constant times

\begin{equation}
\ll \left| \frac{1}{N^{d'}} \int_{x \in \mathbb{R}^{d'}} F_2(x/N) \left( \prod_{j=1}^{d} g_j'(y'_j(x) + a_j) \right) dx \right|.
\end{equation}

**Proof.** We apply Lemma 6.2 to \(\Psi\). Therefore, there is a natural number \(k = O(1)\) such that, for any real numbers \(y_1, \ldots, y_k\), (14.7) is equal to

\begin{equation}
\left| \frac{1}{N^{d-m}} \int_{x \in \mathbb{R}^{d-m}} F_1 \left( \left( x + \sum_{i=1}^{k} y_i f_i \right) / N \right) \left( \prod_{j=1}^{d} g_j(y'_j(y, x) + a_j) \right) dx \right|
\end{equation}

where

- \(f_1, \ldots, f_k \in \mathbb{R}^{d-m}\) are some vectors that satisfy \(\|f_i\|_{\infty} = O_{\Psi}(1)\) for each \(i\) at most \(k\);
- for each \(j\) at most \(d\),
  \[\psi'_j : \mathbb{R}^k \times \mathbb{R}^{d-m} \rightarrow \mathbb{R}\]
  is linear, and \((\psi'_1, \ldots, \psi'_d) := \Psi' : \mathbb{R}^k \times \mathbb{R}^{d-m} \rightarrow \mathbb{R}^d\) is defined by
    \[\Psi'(y, x) := \psi(x + \sum_{i=1}^{k} y_i f_i);\]
- \(\Psi'\) is in \(s\)-normal form, for some \(s = O(1)\).
We remark that the right-hand side of expression (14.9) is independent of $y$, as it was obtained by applying the change of variables $x \mapsto x + \sum_{i=1}^{k} y_i f_i$.

Now, with $\rho$ as fixed in Section 4, let $P : \mathbb{R}^k \to [0, 1]$ be defined by

$$P(y) := \prod_{i=1}^{k} \rho(y_i).$$

Integrating over $y$, we have that (14.9) is at most a constant times

$$\frac{1}{N^{d-m+k}} \int_{y \in \mathbb{R}^k} P(y/N) \times \left| \int_{x \in \mathbb{R}^{d-m}} F_1 \left( \left( x + \sum_{i=1}^{k} y_i f_i \right)/N \right) \prod_{j=1}^{d} g_j(\psi'_j(y, x) + a_j) \, dx \right| \, dy \ll \left| \frac{1}{N^{d-m+k}} \int_{y \in \mathbb{R}^k} F_2((y, x)/N) \prod_{j=1}^{d} g_j(\psi'_j(y, x) + a_j) \, dx \, dy \right|,$$

where the function $F_2 : \mathbb{R}^{d-m+k} \to [0, 1]$ is defined by

$$F_2(y, x) := F_1 \left( x + \sum_{i=1}^{k} y_i f_i \right) P(y).$$

Notice in (14.10) that we were able to move the absolute value signs outside the integral, as $P$ is positive and the integral over $x$ is independent of $y$ (so in particular has constant sign).

Letting $d' := d - m + k$, the lemma is proved. \qed

## 15 Gowers–Cauchy–Schwarz argument

This section will be devoted to proving the following theorem, which lies at the heart of the proof of our main results.

**Theorem 15.1** (Gowers–Cauchy–Schwarz argument). Let $N, t, d, s$ be natural numbers, and let $\gamma, \eta, \sigma, C$ be positive constants. Let $a_1, \ldots, a_d$ be fixed real numbers that satisfy $|a_j| \leq CN$ for all $j$. Let $(\psi_1, \ldots, \psi_t) = \Psi : \mathbb{R}^d \to \mathbb{R}^t$ be a linear map with algebraic coefficients, which is in $s$-normal form. Let $F : \mathbb{R}^{d} \to [0, 1]$ be a Lipschitz function supported on $[-1, 1]^d$ and with Lipschitz constant at most $\sigma^{-1}$. Let

$$g_1, \ldots, g_t : [-2N, 2N] \to \mathbb{R}$$

be any bounded measurable functions that satisfy $|g_j(x)| \leq (\nu^\gamma_{N, \omega} * \chi)(x + a_j)$ for all $x$. Suppose that

$$\min_{j \leq d} \|g_j\|_{L^{1+s}(\mathbb{R}, 2N)} = o(1)$$
as $N \to \infty$. Then if $\eta$ and $\gamma$ are small enough in terms of $\Psi$ and the dimensions $t$, $d$, and $s$,

$$
\frac{1}{N^d} \int_{x \in \mathbb{R}^d} \left( \prod_{j=1}^{t} g_j(\psi_j(x)) \right) F(x/N) \, dx = o(1)
$$

as $N \to \infty$, where the error term can depend on $C$, $\sigma$, $\eta$, $\gamma$, $\Psi$, and the first $o(1)$ term.

For the definition of $\|g_j\|_{U^d(\mathbb{R}, 2N)}$, the reader may consult expression (13.8).

Theorem 15.1 is closely analogous to [13, Proposition 7.1], and the first half of our proof will follow the proof of that proposition closely (and in particular will contain no new ideas). However, new technicalities will become apparent as the argument progresses. In particular, it will become important to understand the structure of a function that we will come to denote by $Q_{a, N}(z, h)$, and this will not be easy, in that we will have to appeal to the highly technical Lemma 11.1. This observation and the subsequent analysis constitute the main new elements of the proof of Theorem 15.1.

**Proof.** We begin by replacing $F$ with a cut-off function that will be easier to work with during the subsequent manipulations. Indeed, let us pick a positive parameter $\delta \in (0, 1]$. By Lemma 3.3 there is some parameter $k = O(\delta^{-d})$ and some smooth functions $F_1, \ldots, F_k : \mathbb{R}^d \to [0, 1]$ such that

$$
\left\| F - \sum_{i=1}^{k} F_i \right\|_{\infty} = O(\delta \sigma^{-1})
$$

and each $F_i$ is of the form

$$
F_i(x) = c_{i,F} \prod_{j=1}^{d} b_j^i(x_j/N),
$$

where $|c_{i,F}| \leq 1$ and the functions $b_j^i : \mathbb{R} \to [0, 1]$ are smooth, supported on $[-2, 2]$, and satisfy $b_j^i \in C(\delta)$.

Therefore, we may write the left-hand side of (15.1) as the sum of $O(\delta^{-d})$ expressions of the form

$$
(15.2) \quad c_{i,F} \frac{1}{N^d} \int_{x \in \mathbb{R}^d} \prod_{l=1}^{t} g_l(\psi_l(x)) \prod_{j=1}^{d} b_j^i(x_j/N) \, dx,
$$

plus an error of size at most

$$
(15.3) \quad \delta \sigma^{-1} \frac{1}{N^d} \int_{x \in \mathbb{R}^d} \prod_{l=1}^{t} (\nu_{\Psi, a_l} * \chi)(\psi_l(x) + a_l).$$
Since $\Psi$ is in $s$-normal form, for some finite $s$, it follows that $\Psi$ has finite Cauchy–Schwarz complexity (see Definition 5.5). Therefore, by Corollary 10.4, expression (15.3) has size $O_{C,\eta,\gamma}(\delta \sigma^{-1})$.

We now arrange our notation for the rest of the proof, in part to mimic the notation that is used in the proof of [13, Proposition 7.1′′]. This will hopefully increase the readability for those who are familiar with [13]. Indeed, without loss of generality we may assume that
\[ \min_{j \leq d}(\|g_j\|_{L^{s+1}[N]}) = \|g_1\|_{L^{s+1}[N]} \]
Since $\Psi$ is in $s$-normal form, there is a set $J \subset \{e_1, \ldots, e_d\}$ of standard basis vectors with $|J| \leq s + 1$ and for which $\prod_{i \in J} \psi_i(e_j)$ vanishes for $i \neq 1$ and is non-zero for $i = 1$. By the nested property of Gowers norms we may assume that $|J| = s + 1$, and by reordering the variables we can assume without loss of generality that
\[ \prod_{j=1}^{s+1} \psi_i(e_j) \]
vanishes for $i \neq 1$ and is non-zero for $i = 1$. It will be useful to rename the first $s + 1$ variables $x$ and the remainder as $y$. If $d = s + 1$ then the variable $y$ is trivial. Note that the coefficients $\psi_1(e_j)$ are non-zero for all $j \in [s+1]$, so, by rescaling the variables $x$, we may assume that
\[ \psi_1(x, y) = x_1 + \cdots + x_{s+1} + \psi_1(0, y). \]

For $i \leq t$, let $\Omega(i)$ denote\footnote{This is the notation used in [13]. In this paper it will never risk being confused with the meaning of $\Omega$ in asymptotic notation.} the set
\[ \Omega(i) := \{ j \in [s + 1] : \psi_i(e_j) \neq 0 \}. \]
Note that $\Omega(1) = [s + 1]$ and $\Omega(i) \neq [s + 1]$ for $i = 2, \ldots, t$.

Now, for any set $B \subseteq [s + 1]$ and vector $x \in \mathbb{R}^{s+1}$, we define the vector $x_B$ to be the restriction of $x$ to the coordinates in $B$. Then, for any set $B \subseteq [s + 1]$ and vector $y \in \mathbb{R}^{d-s-1}$, we define
\[ G_{B,y}(x_B) := \prod_{i \in [t]: \Omega(i) = B} g_i(\psi_i(x_B, y)), \]
where we have abused notation slightly in viewing $\psi_i$ only as a function of those variables $x_j$ on which it depends.

We also use $b : \mathbb{R}^d \rightarrow \mathbb{R}$ (for some implied dimension parameter $a$) to denote a smooth function in $C(C, \delta, \eta, \gamma, \Psi)$. The exact function may change from line to line.
With this notation, by picking $\delta$ to be a suitably slowly decaying function of $N$ we see that Theorem 15.1 would follow from the upper bound

$$
\frac{1}{N^{d-s-1}} \int_{y \in \mathbb{R}^{d-s-1}} \frac{1}{N^{s+1}} \times \int_{x \in \mathbb{R}^{s+1}} \prod_{B \subseteq [s+1]} G_{B,y}(x_B) \prod_{j=1}^{s+1} b_j(x_j/N) \prod_{k=s+2}^{d} b_k(y_{k-s-1}/N) \, dx \, dy
$$

$$
= o_{C, \delta, \eta, \gamma, \psi}(1).
$$

Our entire task is now to establish (15.4). From this point onwards, we will allow any error term or implied constant to depend on $C, \delta, \eta, \gamma, \psi$, without noting so explicitly.

We proceed by considering the following version of [13, Corollary B.4].

**Proposition 15.2** (The weighted generalised von Neumann theorem). Let $A$ be a finite set, and let $(\mu_a)_{a \in A}$ be a finite collection of compactly supported Borel probability measures on $\mathbb{R}$. For every $B \subseteq A$, let $\mu_B$ denote the product measure $\bigotimes_{a \in B} \mu_a$ on $\mathbb{R}^B$, and let $f_B : \mathbb{R}^B \rightarrow \mathbb{C}$ and $\theta_B : \mathbb{R}^B \rightarrow \mathbb{R}_{\geq 0}$ be integrable functions such that $|f_B(x_B)| \leq \theta_B(x_B)$ for all $x_B \in \mathbb{R}^B$. Then

$$
\left| \int_{x_A \in \mathbb{R}^A} \left( \prod_{B \subseteq A} f_B(x_B) \right) \, d\mu_A(x_A) \right| \leq \|f_A\|_{\square^\alpha(\theta, \mu_A)} \prod_{B \subseteq A} \|\theta_B\|_{\square^\beta(\theta, \mu_B)},
$$

where for any $B \subseteq A$ and $h_B : \mathbb{R}^B \rightarrow \mathbb{C}$ we define $\|h_B\|_{\square^\gamma(\theta, \mu_B)}$ to be the unique non-negative real number satisfying

$$
\|h_B\|_{\square^\gamma(\theta, \mu_B)} := \int_{x_B(0), x_B(1) \in \mathbb{R}^B} \left( \prod_{\omega_B \in \{0, 1\}^B} \phi_{\omega_B}^{\left[\omega_B\right]} h_B(x_B^{(\omega_B)}) \right) \times \prod_{C \subseteq B} \prod_{\omega_C \in \{0, 1\}^C} \theta_C(x_C^{(\omega_C)}) \, d\mu_B(x_B^{(0)}) \, d\mu_B(x_B^{(1)}).
$$

Here, as before, we use $x_C$ to denote the restriction of $x_B$ to $\mathbb{R}^C$.

**Proof.** The proof is identical to the proof of [13, Corollary B.4], replacing all summations with integrals, and is a consequence of the Gowers–Cauchy–Schwarz inequality. 

We now apply this proposition to the left-hand side of (15.4) above. Observe that we have the pointwise bounds $|G_{B,y}(x_B)| \ll \theta_{B,y}(x_B)$, where

$$
\theta_{B,y}(x_B) := \prod_{i \in [r]: \Omega(i) = B} (\nu_{B,y}^{i} * \check{\chi})(\psi_i(x_B, y) + a_i).
$$
Therefore, applying Proposition 15.2 by taking \( A \) to be the set \([s + 1]\), each \( \mu_a \) to be proportional to \( (1/N) b_j(x_j/N) dx_j \), and \( \theta_B \) to be the function \( \theta_{B,y} \), we establish that the left-hand side of (15.4) is

\[
\leq \frac{1}{N^{d-s-1}} \int_{y \in \mathbb{R}^{d-s-1}} \| G_{[s+1],y} \|_{\nu(\theta_{[s+1]} \mu_{[s+1]})} \prod_{B \subseteq [s+1]} \| \theta_{B,y} \|_{\nu(\theta_{[s+1]} \mu_{[s+1]})} \prod_{k=s+2}^d b_k(y_{k-s-1}/N) \, dy.
\]

Observe that

\[
G_{[s+1],y}(x_{[s+1]}) = g_1(x_1 + \cdots + x_{s+1} + \psi_1(0, y)),
\]

and so all the functions \( g_j \) other than \( g_1 \) have been eliminated. Experienced readers will note that, so far, we have been following [13, Appendix C] almost verbatim.

After applying Hölder’s inequality to (15.7), we see that to establish (15.4) it suffices to prove

\[
\frac{1}{N^{d-s-1}} \int_{y \in \mathbb{R}^{d-s-1}} \| G_{[s+1],y} \|_{\nu(\theta_{[s+1]} \mu_{[s+1]})} \prod_{k=1}^{d-s-1} b_k(y_k/N) \, dy = o(1),
\]

and, for all \( B \subseteq [s + 1] \),

\[
\frac{1}{N^{d-s-1}} \int_{y \in \mathbb{R}^{d-s-1}} \| \theta_{B,y} \|_{\nu(\theta_{[s+1]} \mu_{[s+1]})} \prod_{k=1}^{d-s-1} b_k(y_k/N) \, dy \ll 1.
\]

These two expressions correspond respectively to expressions (C.10) and (C.11) of [13].

Establishing (15.9) is straightforward. Indeed, we expand the left-hand side, yielding (up to a multiplicative constant factor) the expression

\[
\frac{1}{N^{d-s-1}} \int_{y \in \mathbb{R}^{d-s-1}} \frac{1}{N^{2\|B\|}} \times \prod_{x^{(0)}_B, x^{(1)}_B \in \mathbb{R}^B} \prod_{C \subseteq B, \omega_C \in \{0,1\}^C} \prod_{i \in \bar{r}, \Omega(i)=C} (v_{N,10} \ast \chi)(\psi_i(x^{(oC)}_C, y) + a_i)
\]

\[
\times \prod_{j \in B} b_j(x^{(0)}_j/N) b_j(x^{(1)}_j/N) \prod_{k=1}^{d-s-1} b_{k+s+1}(y_k/N) \, dx^{(0)}_B \, dx^{(1)}_B \, dy.
\]

As noted in [13, p. 1824], the system of forms given by

\[
(y, x^{(0)}_B, x^{(1)}_B) \mapsto \psi_i(x^{(oC)}_C, y),
\]

for each \( C \subseteq B, \omega_C \in \{0,1\}^C \) and \( i \in \bar{r} \) such that \( \Omega(i) = C \), has finite Cauchy–Schwarz complexity (since \( \Psi \) does). We may therefore apply the upper bound in Corollary 10.4 to expression (15.10), and this immediately yields (15.9).
It remains to prove (15.8), which will be a much more major undertaking. We introduce some space-saving notation, namely for any subset $B \subseteq [s+1]$ we define the indexing set

$$I_B := \{(C, \omega_C, i) : C \subseteq B, \omega_C \in \{0, 1\}^C, \Omega(i) = C\}.$$  

If a product is taken over triples $t \in I_B$, we interpret $C, \omega_C$ and $i$ as coming from the triple $t = (C, \omega_C, i)$. For notational expedience we will also identify the space $\mathbb{R}^{|I_B|}$ with the space $\mathbb{R}^{|I|}$. 

With this notation, the left-hand side of (15.8) expands as

$$\frac{1}{N^{d+s+1}} \int_{\mathbb{R}^{d-s-1}} \frac{1}{\prod_{\omega \in \{0, 1\}^{s+1}}} \left( \prod_{\omega \in \{0, 1\}^{s+1}} g_1 \left( \sum_{j=1}^{s+1} x_j^{(\omega)} + \psi_1(0, y) \right) \right)$$

$$\times \left( \prod_{t \in I_B} (\nu_{N, \omega}^* \chi)(\psi(C, y) + a_i) \right)$$

$$\times b((x_{[s+1]}^{(0)}, x_{[s+1]}^{(1)}), y)/N) \, dx_{[s+1]}^{(0)} \, dx_{[s+1]}^{(1)} \, dy.$$

We make the substitution $h := x_{[s+1]}^{(1)} - x_{[s+1]}^{(0)}$ and $z := x_1^{(0)} + \cdots + x_{s+1}^{(0)} + \psi_1(0, y)$. Given $h, z, x_{[s]}^{(0)}$ and $y$ one can recover $x_{[s]}^{(0)}, x_{[s+1]}^{(1)}$ and $y$, so the change of variables is invertible. Therefore we may bound (15.11) above by a constant (the Jacobian of the change of variables) times

$$\frac{1}{N^{d+2}} \sum_{\omega \in \{0, 1\}^{s+1}} P_{a, N}(z, h) \left( \prod_{\omega \in \{0, 1\}^{s+1}} g_1(z + \sum_{j=1}^{s+1} \omega_j h_j) \right) dz \, dh,$$

where $P_{a, N}(z, h)$ is equal to

$$\frac{1}{N^{d-1}} \int_{(x_{[s]}^{(0)}, y) \in \mathbb{R}^{d-1}} \left( \prod_{t \in I_B} (\nu_{N, \omega}^* \chi)(\varphi_t(z, h, x_{[s]}^{(0)}, y) + a_i) \right)$$

$$\times b((z, h, x_{[s]}^{(0)}, y)/N) \, dx_{[s]}^{(0)} \, dy$$

for some linear functions $\varphi_t : \mathbb{R}^{d+s+1} \rightarrow \mathbb{R}$.

To be precise, if $t = (C, \omega, i)$ then the expression $\varphi_t(z, h, x_{[s]}^{(0)}, y)$ is equal to

$$\sum_{k=1}^{s} (\psi_j(e_k) - \psi_j(e_{s+1})) x_k^{(0)} - \psi_j(e_{s+1}) \psi_1(0, y) + \psi_1(0, y) + c(z, h)_t,$$

where

$$c(z, h)_t = \psi_j(e_{s+1}) z + \sum_{k=1}^{s+1} \psi_j(e_k) \omega_k h_k.$$
This expression is analogous to expression (C.14) of [13]. We let \( c(z, h) \in \mathbb{R}^{I_{[s+1]}} \) denote the vector \((c(z, h))_{t \in I_{[s+1]}}\). Most fortunately, the exact structure of the linear maps \( \varphi_t \), save for the fact that they form a system with finite Cauchy-Schwarz complexity, will be unimportant.

Following the philosophy of [11] and [13], our next manoeuvre will be to replace \( P_{a,N}(z, h) \) with a simpler function. To that end, let \( w^*: \mathbb{N} \longrightarrow \mathbb{R}_{>0} \) be a function for which \( w^*(N) \rightarrow \infty \) as \( N \rightarrow \infty \) and \( w^*(n) \leq w(n) \) for all \( n \). Recall from Section 4 that \( W^* = W^*(N) = \prod_{p \leq w^*(N)} p \).

**Lemma 15.3** (Comparing \( P_{a,N}(z, h) \) and \( Q_{a,N}(z, h) \)). Define \( Q_{a,N}(z, h) \) to be equal to

\[
(15.14) \quad \frac{1}{N^{d-1}} \int_{(x_{[s]}, y) \in \mathbb{R}^{d-1}} \left( \prod_{t \in I_{[s+1]}} (\Lambda_{I_{[s]}^{w^{*}}} \ast \chi)(\varphi_t(z, h, x_{[s]}^0, y) + a_t) \right) \\
\times b((z, h, x_{[s]}^0, y)/N) \, dx_{[s]}^0 \, dy,
\]

where \( b((z, h, x_{[s]}^0, y)/N) \) here denotes the same function as is present in (15.13).

Then expression (15.12) is equal to

\[
(15.15) \quad \frac{1}{N^{s+2}} \int_{(z, h) \in \mathbb{R}^{s+2}} Q_{a,N}(z, h) \left( \prod_{\omega \in \{0, 1\}^{s+1}} g_1(z + \sum_{j=1}^{s+1} \omega_j h_j) \right) \, dz \, dh + o(1),
\]

where the \( o(1) \) may depend on the function \( w^* \).

**Proof.** Considering the upper bound \(|g_1(x)| \leq (v_{N,w}^* \ast \chi)(x + a_1)|\), it suffices to show that

\[
\frac{1}{N^{s+2}} \int_{(z, h) \in \mathbb{R}^{s+2}} |P_{a,N}(z, h) - Q_{a,N}(z, h)| \left( \prod_{\omega \in \{0, 1\}^{s+1}} (v_{N,w}^* \ast \chi)(z + \sum_{j=1}^{s+1} \omega_j h_j + a_1) \right) \, dz \, dh
\]

is \( o(1) \). By Cauchy–Schwarz, it then suffices to show that both

\[
(15.16) \quad \frac{1}{N^{s+2}} \int_{(z, h) \in \mathbb{R}^{s+2}} \left( \prod_{\omega \in \{0, 1\}^{s+1}} (v_{N,w}^* \ast \chi)(z + \sum_{j=1}^{s+1} \omega_j h_j + a_1) \right) \, dz \, dh \ll 1
\]

and

\[
(15.17) \quad \frac{1}{N^{s+2}} \int_{(z, h) \in \mathbb{R}^{s+2}} (P_{a,N}(z, h) - Q_{a,N}(z, h))^2 \\
\times \left( \prod_{\omega \in \{0, 1\}^{s+1}} (v_{N,w}^* \ast \chi)(z + \sum_{j=1}^{s+1} \omega_j h_j + a_1) \right) \, dz \, dh = o(1).
\]
The bound (15.16) is immediate from Corollary 10.4. To prove (15.17), expanding out the square we must consider three expressions. One of them is

\[
\frac{1}{N^{s+2}} \int_{(z,h) \in \mathbb{R}^{s+2}} P_{\alpha,N}(z,h)^2 \prod_{\omega \in \{0,1\}^{s+1}} (v_{N,w}^\gamma \ast \chi)(z + \sum_{j=1}^{s+1} \omega_j h_j + a_1) \, dz \, dh.
\]

When multiplied out, (15.18) is equal to the large expression

\[
\frac{1}{N^{2d+2}} \int_{(z,h) \in \mathbb{R}^{s+2}} \prod_{t \in I_{s+1}} (v_{N,w}^\gamma \ast \chi) \times (\varphi_t(z,h,x_{[s]}^{(0)},y) + a_i)(v_{N,w}^\gamma \ast \chi)(\varphi_t(z,h,x_{[s]}^{(0)},y) + a_i)
\]

\[
\times \prod_{\omega \in \{0,1\}^{s+1}} (v_{N,w}^\gamma \ast \chi)(z + \sum_{j=1}^{s+1} \omega_j h_j + a_1)
\]

\[
\times b((z,h,x_{[s]}^{(0)},y)/N) b((z,h,x_{[s]}^{(0)},y)/N) dx_{[s]}^{(0)} dy dx_{[s]}^{(0)} dy dz dh.
\]

By applying Corollary 10.3 to the above expression, we may replace the functions $v_{N,w}^\gamma \ast \chi$ with $\Lambda_{Z/W \ast Z} \ast \chi$, up to an $o(1)$ error.

It is worth noting why the application of Corollary 10.3 is valid. Indeed, the underlying set of linear forms is given by (for each $t \in I_{s+1}$)

\[
(z,h,x_{[s]}^{(0)},y,x_{[s]}^{(0)},y) \mapsto (\varphi_t(z,h,x_{[s]}^{(0)},y), \varphi_t(z,h,x_{[s]}^{(0)},y)).
\]

We need this linear map to have algebraic coefficients and to have finite Cauchy–Schwarz complexity. Algebraicity follows by the assumptions in the statement of Theorem 15.1. Establishing finite Cauchy–Schwarz complexity is rather involved, but fortunately this has already been done by Green and Tao, on pages 1826 and 1827 of [13], in the analysis of expression (C. 14).

Replacing (15.18) with one of the other two terms that arises from expanding out the square in (15.17), and performing the same estimation, the lemma follows. \hfill \square

Let us take stock. As a reminder, we are trying to establish that (15.8) holds. Lemma 15.3 above reduces matters to choosing some function $u^*$ that tends to infinity for which the bound

\[
\frac{1}{N^{s+2}} \left| \int_{(z,h) \in \mathbb{R}^{s+2}} Q_{\alpha,N}(z,h) \prod_{\omega \in \{0,1\}^{s+1}} g_1(z + \sum_{k=1}^{s+1} \omega_k h_k + a_1) \, dz \, dh \right| = o(1)
\]

holds. If $Q_{\alpha,N}(z,h)$ were identically equal to 1, then expression (15.20) would be of the order of $\|g_1\|_{U^{s+1}(\mathbb{R},2N)}$, and hence be $o(1)$ by the hypotheses of Theorem 15.1.
Of course $Q_{a,N}(z, h)$ is not identically equal to 1, but we do observe that $Q_{a,N}(z, h)$ is a function of the form considered in Lemma 11.1. Indeed, consulting the definition of $Q_{a,N}(z, h)$ in (15.14), the following table shows which objects in Lemma 11.1 correspond to which objects concerning the definition of $Q_{a,N}(z, h)$.

| Lemma 11.1 | (15.14) |
|-----------|---------|
| $a$       | $a$     |
| $x$       | $(x^{(0)}, y)$ |
| $y$       | $(z, h)$ |
| $\varphi_j(x)$ | $\varphi_1(0, 0, x^{(0)}, y)$ |
| $\Psi(y)$ | $c(z, h)$ |

From Lemma 11.1, we therefore know that there exists some function $f_1 : \mathbb{Z}^{[s+1]} \rightarrow \mathbb{C}$ satisfying $\|f_1\|_{\infty} \ll (\log \log W^*)^{O(1)}$ for which

$$Q_{a,N}(z, h) = b_{a,N}((z, h)/N) \sum_{\|\mathbf{k}\|_{\infty} \leq (\log \log W^*)^{O(1)}} f_1(\mathbf{k}) e\left(\frac{\mathbf{k} \cdot (c(z, h) + a)}{W^*}\right) + o(1).$$

Therefore, one gets an upper bound for the left-hand side of (15.20), namely

$$\frac{(\log \log W^*)^{O(1)}}{N^{s+2}} \times \left(\sup_{\mathbf{k} \in \mathbb{Z}^{[s+1]}} \left| \int_{(z, h) \in \mathbb{R}^{s+2}} e\left(\frac{\mathbf{k} \cdot (c(z, h))}{W^*}\right) b_{a,N}((z, h)/N) \right. \right.$$

$$\left. \times \prod_{\omega \in \{0, 1\}^{s+1}} g_1\left(z + \sum_{j=1}^{s+1} \omega_j h_j + a_1\right) dz \, dh \right| \right.$$ 

plus an error of size

$$o(1) \times \frac{1}{N^{s+2}} \int_{(z, h) \in \mathbb{R}^{s+2}} \prod_{||z, h||_{\infty} \ll N \omega \in \{0, 1\}^{s+1}} (v_{N, \omega}^{(z, h)} \ast \chi)\left(z + \sum_{k=1}^{s+1} \omega_k h_k + a_1\right) dz \, dh.$$

By Corollary 10.4, the size of term (15.22) is $o(1)$. To analyse (15.21) we apply Lemma B.4 of [21]. Since the function $b_{a,N}$ is Lipschitz, this means that for all $Y > 2$ there exists a complex valued function $f_{a,N,2}$ such that $\|f_{a,N,2}\|_{\infty} \ll 1$ and for all $(z, h)$ one has

$$b_{a,N}((z, h)/N) = \int_{\|y\|_{\infty} \leq Y} f_{a,N,2}(y) e\left(\frac{y \cdot (z, h)}{N}\right) dx + O((\log Y)/Y).$$
Choosing $Y$ to be a suitably large power of $\log \log W^*$, \eqref{e:15.21} may be bounded above by

$$\frac{(\log \log W^*)^{O(1)}}{N^{s+2}} \sup_{k,y} \left| \int_{(z,h) \in \mathbb{R}^{s+2}} e\left( \frac{k \cdot c(z,h)}{W^*} \right) e\left( \frac{y \cdot (z,h)}{N} \right) \prod_{\omega \in \{0,1\}^{s+1}} g_1 \left( z + \sum_{j=1}^{s+1} \omega_j h_j + a_1 \right) \, dz \, dh \right|$$

plus an error of size

$$\frac{1}{N^{s+2}} \int_{(z,h) \in \mathbb{R}^{s+2}} \prod_{\omega \in \{0,1\}^{s+1}} (\nu_{N,w^*}^* \chi)(z + \sum_{k=1}^{s+1} \omega_k h_k + a_1) \, dz \, dh.$$  \hfill \eqref{e:15.24}

Using Corollary 10.4 as above, expression \eqref{e:15.24} is $o(1)$.

The term \eqref{e:15.23} may be analysed using the standard methods. Indeed, by shifting the variable $z$ (and noting that $c(z,h)$ is a linear function of $z$ and $h$) we may assume that $a_1 = 0$. Then, by spreading the exponential functions across the different instances of $g_1$, we see it suffices to show that

$$\frac{(\log \log W^*)^{O(1)}}{N^{s+2}} \left| \int_{(z,h) \in \mathbb{R}^{s+2}} \prod_{\omega \in \{0,1\}^{s+1}} g_\omega \left( z + \sum_{k=1}^{s+1} \omega_k h_k \right) \, dz \, dh \right| = o(1),$$

where each function $g_\omega$ is of the form

$$g_\omega(x) := g_1(x)e(\lambda_\omega x),$$

for some $\lambda_\omega \in \mathbb{R}$.

The argument is nearly complete. Considering expression \eqref{e:13.8}, for each $\omega$ we observe that

$$\|g_\omega\|_{U^{s+1}(\mathbb{R},2N)} = \|g_1\|_{U^{s+1}(\mathbb{R},2N)}.$$ 

So, by the Gowers–Cauchy–Schwarz inequality (recorded in this setting as Proposition A.4 of [21]), the left-hand side of expression \eqref{e:15.25} is

$$O((\log \log W^*)^{O(1)} \|g_1\|_{U^{s+1}(\mathbb{R},2N)}^{O(1)}).$$

If $w^*$ grows slowly enough, this expression is $o(1)$.

We have therefore established the upper bound \eqref{e:15.20}, and so, by our long sequence of deductions, Theorem 15.1 is finally proved. \hfill \square
16 Combining the lemmas

With all the previous lemmas in hand, we may finally prove Theorem 12.1 (and hence prove Theorem 1.16).

Proof of Theorem 12.1. Assume the hypotheses of the theorem, fixing a suitably small value of $\gamma$.

By applying Proposition 14.1 and Proposition 14.2, we conclude that there is some $s = O(1)$ and some $d' = O(1)$ for which $|\tilde{T}_{F,G,N}^{L,x}(f_1 \ast \chi, \ldots, f_d \ast \chi)|$ is

\begin{equation}
\ll_{L,\varepsilon} \left| \frac{1}{N^{d'}} \int_{x \in \mathbb{R}^d} F_2(x/N) \prod_{j=1}^{d}(f_j \ast \chi)(\psi_j'(x) + a_j) \, dx \right|,
\end{equation}

where $(\psi_1', \ldots, \psi_d') = \Psi' : \mathbb{R}^{d'} \rightarrow \mathbb{R}^{d}$ is in $s$-normal form,

$F_2 : \mathbb{R}^{d'} \rightarrow [0, 1]$ has Lipschitz constant $O_L(\sigma^{-1})$ and $\text{Rad}(F_2) = O_{C,L,\varepsilon}(1)$, and each $a_j$ satisfies $|a_j| = O_{C,L,\varepsilon}(N)$. Taking this value of $s$ in the hypotheses of Theorem 12.1, without loss of generality we may assume that

\begin{equation}
\|f_1\|_{U^{s+1}[N]} = o(1)
\end{equation}
as $N \rightarrow \infty$.

Then we may apply Theorem 15.1 to expression (16.1). Indeed, by rescaling the variable $x$ we may assume that $F_2$ is supported on $[-1, 1]^{d'}$. For each $j \in [d]$ we set

$g_j := f_j \ast \chi$.

Provided $\eta$ is small enough, by combining (16.2) and Lemma 13.3 we deduce that

\begin{equation}
\|g_1\|_{U^{s+1}(\mathbb{R}, 2N)} = o_{\eta}(1)
\end{equation}
as $N \rightarrow \infty$. So Theorem 15.1 may indeed be applied, which yields

\begin{equation}
|\tilde{T}_{F,G,N}^{L,x}(f_1 \ast \chi, \ldots, f_d \ast \chi)| = o_{C,L,\gamma,\varepsilon,\eta,\sigma}(1)
\end{equation}
as $N \rightarrow \infty$.

But then, combining the estimate (16.3) with Lemma 13.2, one derives the bound

\begin{equation}
|T_{F,G,N}^{L,x}(f_1, \ldots, f_d)| = O_{C,L,\gamma,\varepsilon,\sigma}(\eta) + o_{C,L,\gamma,\varepsilon,\eta,\sigma}(1).
\end{equation}
Choosing $\eta = \eta(N)$ to be a function tending to zero suitably slowly with $N$, we conclude that

$$|T_{F,G,N}^{L,v}(f_1, \ldots, f_d)| = o_{C,L,\gamma,\varepsilon,\sigma}(1).$$

This is the conclusion of Theorem 12.1, and we are done. □

From the work in Section 12, this means that Theorem 1.16, the main result of this paper, is finally settled. □

Part 6
Final deductions

17 Removing Lipschitz cut-offs

In this section we assume Theorem 1.16, and deduce Theorem 1.7. This deduction will be a routine matter of removing Lipschitz cut-offs.

**Lemma 17.1.** Assume the hypotheses of Theorem 1.7. Let $\delta$ be a real number in the range $0 < \delta < 1/2$ and let $I \subset [0, 1]$ be an interval of length $\delta$. Then

$$\frac{1}{N^{d-m}} \sum_{n \in [N]^d} \left( \prod_{j=1}^d \Lambda'(n_j) \right) 1_{[-\varepsilon, \varepsilon]^m}(L_n + v) \ll L \delta \varepsilon^m + o_{C,L,\delta,\varepsilon}(1).$$

The reader will note that this lemma is a slight refinement of Corollary 9.15.

**Proof.** Fix some $i \leq d$. Let $F : \mathbb{R}^d \rightarrow [0, 1]$ be a smooth function in $C(\delta)$, supported on $\{ x \in [-1, 2]^d : x_i \in I + [-\delta, \delta] \}$, that majorises the indicator function of the set $\{ x \in [0, 1]^d : x_i \in I \}$. Let $G : \mathbb{R}^m \rightarrow [0, 1]$ be some smooth function in $C(\varepsilon)$, supported on $[-2\varepsilon, 2\varepsilon]^m$, that majorises $1_{[-\varepsilon, \varepsilon]^m}$. Let $\gamma$ be small enough in terms of $L$. Then, by Theorem 9.5 and Lemma 9.11,

$$\frac{1}{N^{d-m}} \sum_{n \in [N]^d} \left( \prod_{j=1}^d \Lambda'(n_j) \right) 1_{[-\varepsilon, \varepsilon]^m}(L_n + v) \ll \gamma \frac{1}{N^{d-m}} \sum_{n \in [N]^d} \left( \prod_{j=1}^d \Lambda'(n_j) \right) 1_{[-\varepsilon, \varepsilon]^m}(L_n + v)$$

$$\leq T_{F,G,N}^{L,v}(v_{N,\omega}^\gamma, \ldots, v_{N,\omega}^\gamma)$$

$$= T_{F,G,N}(A_{Z/W}^\gamma, \ldots, A_{Z/W}^\gamma) + o_{C,L,\gamma,\delta,\varepsilon}(1)$$

$$= \int_{x \in \mathbb{R}^d} F(x/N) G(Lx + v) \, dx + o_{C,L,\gamma,\delta,\varepsilon}(1).$$
Since \( L \notin V^\text{degen}(m, d) \), for all \( d \) of the coordinate subspaces \( U \leq \mathbb{R}^d \) of dimension \( d - 1 \) the map \( L|_U : U \to \mathbb{R}^m \) is surjective. We may therefore apply Lemma A.3, and conclude that expression (17.1) is \( O_L(\delta^m) + o_{C, L, \delta, \epsilon, \gamma}(1) \). The lemma is proved, after having fixed a suitable \( \gamma \). \( \square \)

**Lemma 17.2.** Under the hypotheses of Theorem 1.7,

\[
\frac{1}{N^{d-m}} \sum_{n \in [N]^d} \left( \prod_{j=1}^{d} \Lambda'(n_j) \right) 1_{[-\epsilon, \epsilon]^m}(Ln + v) = \frac{1}{N^{d-m}} \int_{x \in [0, N]^d} 1_{[-\epsilon, \epsilon]^m}(Lx + v) \, dx + o_{C, L, \epsilon}(1).
\]

**Proof.** Let \( \delta \) be a positive parameter in the range \((0, 1/2)\), to be chosen later. Let us first consider

\[
\frac{1}{N^{d-m}} \sum_{n \in [dN, (1-\delta)N]^d} \left( \prod_{j=1}^{d} \Lambda'(n_j) \right) 1_{[-\epsilon, \epsilon]^m}(Ln + v).
\]

Let \( F^{\pm \delta} : \mathbb{R}^d \to [0, 1] \) be two Lipschitz functions satisfying

\[
1_{[3\delta/2, 1-3\delta/2]^d} \leq F^{-\delta} \leq 1_{[\delta, 1-\delta]^d} \leq F^{+\delta} \leq 1_{[\delta/2, 1-\delta/2]^d},
\]

with Lipschitz constants depending only on \( \delta \). Let \( G^{\pm \delta} : \mathbb{R}^m \to [0, 1] \) be two Lipschitz functions satisfying

\[
1_{[-\epsilon(1-\delta), \epsilon(1-\delta)]^m} \leq G^{-\delta} \leq 1_{[-\epsilon, \epsilon]^m} \leq G^{+\delta} \leq 1_{[-\epsilon(1+\delta), \epsilon(1+\delta)]^m},
\]

with Lipschitz constants\(^9\) depending only on \( \delta \). Then we have

\[
\sum_{n \in [dN, (1-\delta)N]^d} \left( \prod_{j=1}^{d} \Lambda(n_j) \right) F^{-\delta}(n/N) G^{-\delta}(Ln + v) \leq \sum_{n \in [dN, (1-\delta)N]^d} \left( \prod_{j=1}^{d} \Lambda'(n_j) \right) 1_{[-\epsilon, \epsilon]^m}(Ln + v) \leq \sum_{n \in [dN, (1-\delta)N]^d} \left( \prod_{j=1}^{d} \Lambda(n_j) \right) F^{+\delta}(n/N) G^{+\delta}(Ln + v).
\]

By Theorem 1.16, the lower bound in (17.2) is equal to

\[
\sum_{n \in [dN, (1-\delta)N]^d} \left( \prod_{j=1}^{d} \Lambda_{Z/W}(n_j) \right) F^{-\delta}(n/N) G^{-\delta}(Ln + v) + o_{C, L, \delta, \epsilon}(N^{d-m}),
\]

---

\(^9\)The existence of such functions is immediate by interpolating linearly, or by appealing to the results of Section 3.
since we may replace $\Lambda_{Z/W}^+$ with $\Lambda_{Z/W}$ as $F$ is supported on $[0, 1]^d$. By Lemma 9.11, and the properties of the support of $F^{-\delta}$ and $G^{-\delta}$, this is at least

$$
\int_{x \in [3\delta N/2, N(1-3\delta/2)]^d} 1_{[-\varepsilon(1-\delta), \varepsilon(1+\delta)]^m}(Lx + v) \, dx + o_{C, L, \delta, \varepsilon}(N^{d-m}).
$$

Note that the singular series $S$ is equal to 1 in this instance, since $L$ is purely irrational. By Lemma A.4, expression (17.3) is at least

$$
\int_{x \in [0, N]^d} 1_{[-\varepsilon, \varepsilon]^m}(Lx + v) \, dx - O(\delta \varepsilon^m N^{d-m}) + o_{C, L, \delta, \varepsilon}(N^{d-m}).
$$

By performing an analogous manipulation with the upper bound, we may conclude that

$$
\sum_{n \in \mathbb{Z}^d} \left( \prod_{j=1}^d \Lambda'(n_j) \right) 1_{[-\varepsilon, \varepsilon]^m}(Ln + v)
$$

is equal to

$$
\int_{x \in [0, N]^d} 1_{[-\varepsilon, \varepsilon]^m}(Lx + v) \, dx + o_{C, L, \delta, \varepsilon}(N^{d-m}).
$$

Therefore, by Lemma 17.1, we have that

$$
\frac{1}{N^{d-m}} \sum_{n \in [N]^d} \left( \prod_{j=1}^d \Lambda'(n_j) \right) 1_{[-\varepsilon, \varepsilon]^m}(Ln + v)
$$

is equal to

$$
\frac{1}{N^{d-m}} \int_{x \in [0, N]^d} 1_{[-\varepsilon, \varepsilon]^m}(Lx + v) \, dx + O(\delta \varepsilon^m) + o_{C, L, \delta, \varepsilon}(1).
$$

Letting $\delta$ be a function of $N$, tending to zero suitably slowly as $N$ tends to infinity, the lemma follows.

To establish Theorem 1.7 as given, i.e., to establish Lemma 17.2 without the log weighting, is standard. To spell it out, Lemma 17.2 implies that, for any $\delta$ in the range $0 < \delta < 1/2$,

$$
\sum_{p \in [\delta N, N]^d} 1_{[-\varepsilon, \varepsilon]^m}(Lp + v)
$$

is at least

$$
\frac{1}{(\log(\delta) + \log N)^d} \sum_{n \in [\delta N, N]^d} \left( \prod_{j=1}^d \Lambda'(n_j) \right) 1_{[-\varepsilon, \varepsilon]^m}(Ln + v)
$$

(17.5)

$$
\leq \left( \frac{1 + o_{\delta}(1)}{(\log N)^d} \right) \left( \int_{x \in [0, N]^d} 1_{[-\varepsilon, \varepsilon]^m}(Lx + v) \, dx + o_{C, L, \varepsilon}(N^{d-m}) \right).$$
But also, from expression (17.4)

\[
\sum_{p \in [\delta N, N]^d} 1_{[-\epsilon, \epsilon]^m}(Lp + v) \\
\geq \frac{1}{(\log N)^d} \sum_{n \in [\delta N, (1-\delta)N]^d} \left( \prod_{j=1}^{d} \Lambda'(n_j) \right) 1_{[-\epsilon, \epsilon]^m}(Ln + v) \\
\geq \frac{1}{(\log N)^d} \times \left( \int_{x \in [0, N]^d} 1_{[-\epsilon, \epsilon]^m}(Lx + v) \, dx + O(\delta e \sqrt{N^{d-m}}) + o_{C,L,\delta,\epsilon}(N^{d-m}) \right).
\]

(17.6)

By Lemma A.1,

\[
\int_{x \in [0, N]^d} 1_{[-\epsilon, \epsilon]^m}(Lx + v) \, dx = O_{L,\epsilon}(N^{d-m}).
\]

Hence, choosing \(\delta\) to be a function of \(N\) tending to zero suitably slowly, combining bounds (17.6) and (17.5) establishes Theorem 1.7.

\[
\square
\]

Part 7

Appendices

Appendix A Estimating integrals

In this appendix we include the lemmas that help us estimate the ‘global factor’ from Theorem 1.7, namely

\[
\int_{x \in [0, N]^d} 1_{[-\epsilon, \epsilon]^m}(Lx + v) \, dx.
\]

**Lemma A.1** (Upper bound). Let \(h\) be a natural number, let \(m\) be a non-negative integer, and let \(C, K\) be positive constants. Let \(L : \mathbb{R}^h \rightarrow \mathbb{R}^m\) be a surjective linear map. Let \(F : \mathbb{R}^h \rightarrow [0, 1]\) and \(G : \mathbb{R}^m \rightarrow [0, 1]\) be any compactly supported measurable functions, and assume that \(F\) is supported on a box of the form \(x^{(0)} + [-C, C]^h\) and \(G\) is supported on a box of the form \(y^{(0)} + [-K, K]^m\). Then

\[
(A.1) \quad \int_{x \in \mathbb{R}^h} F(x)G(Lx) \, dx \ll_L C^{h-m}K^m \|F\|_\infty \|G\|_\infty.
\]

**Proof.** Split \(\mathbb{R}^h\) as a direct sum \((\ker L) \oplus (\ker L)^\perp\). Observe that \(L|_{(\ker L)^\perp}\) is an injective linear map, so has bounded inverse. Hence the integrand in (A.1) is
zero unless $x_{l(\ker L)}$ is contained within a region which has volume $O_L(K^m)$. The integrand is also zero unless $x_{l(\ker L)}$ is contained within a region which has volume $O_L(C^{m-n})$. Together, these observations combine to give the required bound. \hfill \Box

**Lemma A.2.** Let $N$, $m$, $d$ be natural numbers, with $d \geq m + 1$, and let $\varepsilon$ be a positive parameter. Let $L : \mathbb{R}^d \rightarrow \mathbb{R}^m$ be a surjective purely irrational linear map. Let $v \in \mathbb{R}^m$ be any vector. Then there exists a parameter $C_{L,v/N}$ satisfying $|C_{L,v/N}| = O_L(1)$ such that

(A.2) \[ \int_{x \in [0,N]^d} 1_{[-\varepsilon,\varepsilon]^m}(Lx + v)\,dx = C_{L,v/N} \varepsilon^m N^{d-m} + O_L(\varepsilon^{m+1} N^{d-m-1}). \]

Furthermore, if $\|v\|_\infty = o(N)$ then there exists a constant $C_L$, independent of $v$ and $N$, for which

$$C_{L,v/N} = C_L + o_L(1).$$

**Proof.** Since $L$ has rank $m$, without loss of generality we may assume that the first $m$ columns of $L$ form an invertible submatrix $M$. So

$$\int_{x \in [0,N]^d} 1_{[-\varepsilon,\varepsilon]^m}(Lx + v)\,dx = N^d \int_{x \in [0,1]^d} 1_{[-\varepsilon,\varepsilon]^m}(M^{-1}Lx + M^{-1}v/N)\,dx.$$ 

The first $m$ columns of $M^{-1}L$ form the identity matrix, and so this expression is equal to

(A.3) \[ N^d \int_{x_{m+1} \in [0,1]^{d-m}} \operatorname{vol}\left( \left( M^{-1} \left[ \begin{array}{c} -\varepsilon \\ N \\ \varepsilon \\ N \end{array} \right]^m \right) - \sum_{j=m+1}^d x_j a^{(j)} - M^{-1}v/N \right) \cap [0, 1]^m \right) \, dx_{m+1}^d, \]

where $a^{(j)} \in \mathbb{R}^m$ is the $j$th column of the matrix $M^{-1}L$.

If the vector

$$\sum_{j=m+1}^d x_j a^{(j)} - M^{-1}v/N$$

lies in $[0, 1]^m$, then unless it lies close to the boundary of $[0, 1]^m$ we have

$$\operatorname{vol}\left( \left( M^{-1} \left[ \begin{array}{c} -\varepsilon \\ N \\ \varepsilon \\ N \end{array} \right]^m \right) - \sum_{j=m+1}^d x_j a^{(j)} - M^{-1}v/N \right) \cap [0, 1]^m \right) \]

$$= N^{-m} \varepsilon^m \operatorname{vol}(M^{-1}([-1, 1]^m)).$$

More precisely, letting $C_1$ be a constant that is sufficiently large in terms of $L$, we have that expression (A.3) is equal to

(A.4) \[ N^{d-m} \varepsilon^m \operatorname{vol}(M^{-1}([-1, 1]^m)) \int_{x_{m+1} \in [0,1]^{d-m}} 1_{-\sum_{j=m+1}^d x_j a^{(j)} - M^{-1}v/N \in [0,1]^m} \, dx_{m+1}^d, \]
plus an error term of size at most

\[(A.5) \ll_L N^{d-m} \varepsilon^m \int^* 1 \, dx_{m+1}^d,\]

where \(\int^*\) indicates integration over those \(x_{m+1}^d \in [0, 1]^{d-m}\) for which

\[\text{dist} \left(- \sum_{j=m+1}^d x_j a^{(j)} - M^{-1} v/N, \partial([0, 1]^m)\right) \leq C_1 \varepsilon/N.\]

We remind the reader that \(\partial\) refers to the topological boundary.

Define

\[C_{L, v/N} := \text{vol}(M^{-1}([-1, 1]^m)) \int_{-\sum_{j=m+1}^d x_j a^{(j)} - M^{-1} v/N \in [0, 1]^m} 1 \, dx_{m+1}^d,\]

and

\[(A.6) \quad C_L = \text{vol}(M^{-1}([-1, 1]^m)) \int_{-\sum_{j=m+1}^d x_j a^{(j)} \in [0, 1]^m} 1 \, dx_{m+1}^d.\]

Then certainly \(|C_{L, v/N}| = O_L(1)\). To prove the first part of the lemma it then suffices to control the error term \((A.5)\). Let \(\Phi : \mathbb{R}^{d-m} \longrightarrow \mathbb{R}^m\) denote the map

\[x_{m+1}^d \mapsto \sum_{j=m+1}^d x_j a^{(j)}.\]

For all \(i \leq m\), let \(\pi_i : \mathbb{R}^m \longrightarrow \mathbb{R}\) denote projection onto the \(i^{th}\) coordinate. Then the size of the error term \((A.5)\) is

\[\ll_L N^{d-m} \varepsilon^m \sup_{I \subseteq \mathbb{R}, |I| \leq 2C_1 \varepsilon/N} \left| \int_{x_{m+1}^d \in [0, 1]^{d-m}} 1_I(\pi_i(\Phi(x_{m+1}^d))) \, dx_{m+1}^d \right|,\]

where the supremum is over intervals \(I\).

Since \(L\) is purely irrational (and so \(M^{-1} L\) is also purely irrational), for all \(i \leq m\) the linear map

\[\pi_i \circ \Phi : \mathbb{R}^{d-m} \longrightarrow \mathbb{R}\]

is non-zero. From this we conclude that \((A.5)\) has size at most

\[\ll_L N^{d-m-1} \varepsilon^{m+1},\]

from which the first part of the lemma follows.
For the second part, assume that \( \|v\|_\infty = o(N) \). Then note that

\[
(A.7) \quad |C_{L,v/N} - C_L| \ll_L \int_{m+1}^{\infty} 1 \, dx_{m+1},
\]

where \( \int^{**} \) indicates integration over those \( x_{m+1}^d \in [0, 1]^{d-m} \) for which

\[
\text{dist} \left( - \sum_{j=m+1}^d x_j a^{(j)}, \partial([0, 1]^m) \right) = o_L(1).
\]

One can estimate (A.7) by exactly the same procedure as was used to estimate (A.5), and thereby conclude that (A.7) is \( o_L(1) \). This settles the second part of the lemma. \( \square \)

The next lemma concerns the global factor when one of the variables is restricted to a short interval.

**Lemma A.3.** Let \( N, m, d \) be natural numbers, with \( d \geq m + 1 \), and let \( \varepsilon, \delta \) be positive parameters. Let \( L : \mathbb{R}^d \longrightarrow \mathbb{R}^m \) be a surjective linear map, and assume that for all \( d \) of the coordinate subspaces,\(^{10}\) \( U \subseteq \mathbb{R}^d \) of dimension \( d - 1 \), the map \( L|_U : U \longrightarrow \mathbb{R}^m \) is also surjective. Let \( v \in \mathbb{R}^m \) be any vector, and let \( I \subset \mathbb{R} \) be an interval of length \( \delta N \). Fix a coordinate \( j \in [d] \). Then

\[
\left| \int_{x \in [-N,N]^d} 1_{[-\varepsilon,\varepsilon]^m}(Lx + v) \, dx \right| \ll_L \delta \varepsilon^m N^{d-m}.
\]

**Proof.** By the assumptions of the surjectivity of the restrictions of \( L \), without loss of generality we may assume that \( j = d \) and that the first \( m \) columns of \( L \) form an invertible matrix \( M \). Then, by integrating over \( x_{m+1}, \ldots, x_d \), one has

\[
\int_{x \in [-N,N]^d} 1_{[-\varepsilon,\varepsilon]^m}(Lx + v) \, dx \ll \delta N^{d-m} \sup_{a \in \mathbb{R}^m} \int_{x \in \mathbb{R}^m} 1_{[-\varepsilon,\varepsilon]^m}(Mx + a) \, dx \ll_L \delta \varepsilon^m N^{d-m},
\]

since \( M \) is invertible. \( \square \)

The final lemma of this section details what occurs when one permutes the parameters in the global factor.

**Lemma A.4.** Let \( N, m, d \) be natural numbers, with \( d \geq m + 1 \), and let \( \delta, \varepsilon \) be positive parameters. Assume further that \( \delta < 1/2 \). Let \( L : \mathbb{R}^d \longrightarrow \mathbb{R}^m \) be a surjective purely irrational linear map, and assume that for all \( d \) of the coordinate

\(^{10}\)A coordinate subspace of \( \mathbb{R}^d \) is a subspace generated by a subset of the standard basis vectors.
subspaces $U \subseteq \mathbb{R}^d$ of dimension $d - 1$, the map $L|_U : U \rightarrow \mathbb{R}^m$ is also surjective. Let $v \in \mathbb{R}^m$ be any vector. Then

$$\int_{x \in [-\varepsilon(1\pm\delta), \varepsilon(1\mp\delta)]^m} 1_{[-\varepsilon(1\pm\delta), \varepsilon(1\mp\delta)]^n} (Lx + v) \, dx$$

(A.8)

$$= \int_{x \in [0, N]^d} 1_{[-\varepsilon, \varepsilon]^m} (Lx + v) \, dx + O_L(\delta e^m N^{d-m}) + O_L(e^{m+1} N^{d-m-1}).$$

The proof of this lemma is very similar to the proof of Lemma A.2. We merely sketch the relevant changes.

**Proof.** By Lemma A.3 one may replace the left-hand side of (A.8) by

$$\int_{x \in [0, N]^d} 1_{[-\varepsilon(1\pm\delta), \varepsilon(1\mp\delta)]^m} (Lx + v) \, dx.$$

Let $C_1$ be a suitably large constant (depending on $L$). Following the procedure in the proof of Lemma A.2, and with the same notation for $M$ and $a^{(j)}$, one establishes that the left-hand side of (A.8) is equal to

(A.9)  

$$N^{d-m} e^m (1 + O(\delta)) \operatorname{vol}(M^{-1}([-1, 1]^m)) \int \sum_{j=m+1}^d x_{m+1} \in [0, 1]^d \int_{x \in [-\varepsilon, \varepsilon]^m} 1 \, d x_{m+1},$$

plus an error of size at most

(A.10)  

$$\ll L N^{d-m} e^m \int_{x \in [-\varepsilon, \varepsilon]^m} 1 \, d x_{m+1},$$

where $\int$ indicates integration over those $x_{m+1} \in [-1, 2]^{d-m}$ for which

$$\operatorname{dist}(- \sum_{j=m+1}^d x_j a^{(j)} - M^{-1} v / N, \partial([0, 1]^m)) \leq C_1 (\delta + \varepsilon / N).$$

The error (A.10) may be bounded above by $O_L((\delta + \varepsilon / N) e^m N^{d-m})$, by the same method as we used to bound (A.5).

The main term (A.9), from the work in Lemma A.2, is equal to

$$(1 + O(\delta)) \left( \int \sum_{x \in [0, N]^d} 1_{[-\varepsilon, \varepsilon]^m} (Lx + v) \, dx + O_L(e^{m+1} N^{d-m-1}) \right).$$

Bounding this integral using Lemma A.1, the present lemma follows. □

**Remark A.5.** In the proofs above, we used, in a critical way, the fact that the convex domains $[-\varepsilon, \varepsilon]^m$ and $[0, N]^d$ are axis-parallel boxes.
Appendix B  An analytic argument

We take the opportunity to record a rather more direct argument which yields an asymptotic formula for expressions of the form

\[ \sum_{\mathbf{n} \in [N]^d} \left( \prod_{j=1}^{d} \Lambda'(n_j) \right) 1_{[-\varepsilon, \varepsilon]}(\mathbf{Lx} + \mathbf{v}) \]

in the case when \( \mathbf{L} : \mathbb{R}^d \rightarrow \mathbb{R}^m \) is a linear map with \( d \) at least \( 2m + 1 \) (and certain irrationality conditions hold). This method is a simple elaboration on Parsell’s ideas [18], and can handle more general coefficients than Theorem 1.7 (although it requires more variables, of course). We suspect that this result has been obvious to the experts for fifteen years or more, but we feel that it should appear explicitly in the literature.

**Theorem B.1.** Let \( N, m, d \) be natural numbers, with \( d \geq 2m + 1 \), and let \( \varepsilon \) be a positive parameter. Let \( \mathbf{L} : \mathbb{R}^d \rightarrow \mathbb{R}^m \) be a surjective linear map. Assume further that, when written as a matrix with respect to the standard bases, all the \( m \)-by-\( m \) sub-matrices of \( \mathbf{L} \) have non-zero determinant. Assume also that there does not exist a vector \( \mathbf{a} \in \mathbb{R}^m \setminus \{0\} \) such that \( \mathbf{L}^T \mathbf{a} \in \mathbb{Z}^d \) (i.e., in the language of Definition 5.2 assume that \( \mathbf{L} \) is purely irrational). Then, for all vectors \( \mathbf{v} \in \mathbb{R}^m \),

\[
\sum_{\mathbf{n} \in [N]^d} \left( \prod_{j=1}^{d} \Lambda'(n_j) \right) 1_{[-\varepsilon, \varepsilon]}(\mathbf{Lx} + \mathbf{v}) \tag{B.1}
\]

is equal to

\[
\int_{\mathbf{x} \in [0, N]^d} 1_{[-\varepsilon, \varepsilon]}(\mathbf{Lx} + \mathbf{v}) \, d\mathbf{x} + o_{\mathbf{L}, \varepsilon}(N^{d-m}).
\]

**Remark B.2.** The asymptotic size of the main term may easily be established using Lemma A.2.

**Sketch of proof.** We sketch the argument, referring heavily to estimates from [7] and [18]. Define

\[
f(\theta) = \sum_{n \leq N} \Lambda'(n)e(\theta n).
\]

Let \( T(N) \) be a function that tends to infinity as \( N \) tends to infinity, to be defined later, and let \( \delta = \delta(N) \) be a function (depending on the function \( T \)) that tends to zero suitably slowly as \( N \) tends to infinity.

By Lemma 3.1, there exist smooth functions \( G_{\pm} : \mathbb{R}^m \rightarrow [0, 1] \) for which \( G_{\pm} \in \mathcal{C}(\delta) \) and

\[
1_{[-\varepsilon(1-\delta), \varepsilon(1-\delta)]} \leq G_- \leq 1_{[-\varepsilon, \varepsilon]} \leq G_+ \leq 1_{[-\varepsilon(1+\delta), \varepsilon(1+\delta)]}.
\]
By Fourier inversion we see that
\[
\int_{\alpha \in \mathbb{R}^m} \left( \prod_{j=1}^d f \left( \sum_{i=1}^m \alpha_i \lambda_{ij} \right) \right) \left( \prod_{i=1}^m \hat{G}_{-}(a_i) \right) e(\alpha \cdot \mathbf{v}) \, d\alpha 
\]
(B.2)
\[
\leq \sum_{n \in [N]^d} \left( \prod_{j=1}^d \Lambda'(n_j) \right) 1_{[-\varepsilon, \varepsilon]^d}(L\mathbf{x}) \nonumber
\]
\[
\leq \int_{\alpha \in \mathbb{R}^m} \left( \prod_{i=1}^m f \left( \sum_{i=1}^m \alpha_i \lambda_{ij} \right) \right) \left( \prod_{i=1}^m \hat{G}_+(a_i) \right) e(\alpha \cdot \mathbf{v}) \, d\alpha,
\]
where \( L = (\lambda_{ij})_{i \leq m, j \leq d} \). We estimate the integrals by splitting the range of integration in three regions.

\[
t := \{ \alpha : \|\alpha\|_{\infty} \geq T(N) \} \quad \text{trivial arc}
\]
\[
m := \{ \alpha : \|\alpha\|_{\infty} \in [N^{-1} \log B N, T(N)) \} \quad \text{minor arc}
\]
\[
\mathcal{M} := \{ \alpha : \|\alpha\|_{\infty} \leq N^{-1} \log B N \} \quad \text{major arc}
\]
for some large constant \( B \). See Section 8 for another instance of this technique.

Much of the estimation relies on the following tight mean value bound.

**Lemma B.3.** Let \( U \subset \mathbb{R}^m \) be a domain. Let \( d > 2m \) and let \( l \) be a positive real number satisfying
\[
\left( \frac{d}{m} \right) l > 2 \left( \frac{d-1}{m-1} \right).
\]
Then
\[
\int_{\alpha \in U} \left| \prod_{j=1}^d f \left( \sum_{i=1}^m \alpha_i \lambda_{ij} \right) \right|^l \, d\alpha \ll_L (\text{diam } U + 1)^{O(1)} N^{dl-m}.
\]

**Proof.** We write the left-hand side of (B.3) as
\[
\int_{\alpha \in U} \prod_{S \subseteq [d]} \prod_{|S| = m} \left| \prod_{j \in S} f \left( \sum_{i=1}^m \alpha_i \lambda_{ij} \right) \right|^{l/(d-1)} \, d\alpha
\]
which is
\[
\ll \prod_{S \subseteq [d]} \left( \int_{\alpha \in U} \left| \prod_{j \in S} f \left( \sum_{i=1}^m \alpha_i \lambda_{ij} \right) \right|^l \, d\alpha \right)^{1/(d-1)}
\]
(B.4)

by Hölder’s inequality. Note that \( dl/m = \left( \frac{d}{m} \right) l/ \left( \frac{d-1}{m-1} \right) > 2 \).
Now recall the bound from [18, Lemma 3], namely
\[
\int_0^1 |f(\alpha)|^u \, d\alpha \ll N^{u-1} \quad \text{if } u > 2.
\]

Note that for each fixed \( S \) the \( m \times m \) submatrix of \( L \) given by \((\lambda_{i,j})_{i \leq m, j \in S}\) is invertible. Therefore, by applying an invertible change of variables and splitting \( U \) into boxes, (B.4) is

\[
\ll_L (\text{diam } U + 1)^{O(1)} N^{(d/m-1)m} = (\text{diam } U + 1)^{O(1)} N^{d/m-1}.
\]

This implies the lemma. \( \square \)

**Trivial arc.** The estimation on the trivial arc proceeds very similarly to page 153 of [18]. Indeed, by the bound in Lemma 3.4 we have

\[
\left| \int_{\alpha \in t} \left( \prod_{j=1}^d f \left( \sum_{i=1}^m \alpha_i \lambda_{i,j} \right) \right) \left( \prod_{i=1}^m \hat{G}_\pm(\alpha_i) \right) e(\alpha \cdot \mathbf{v}) \, d\alpha \right| \ll \delta \int_{\alpha \in t} \left| \prod_{j=1}^d f \left( \sum_{i=1}^m \alpha_i \lambda_{i,j} \right) \right| \|\alpha\|_{\infty}^{-m-1} \, d\alpha
\]

which is

\[
\ll \delta \sum_{n \geq T(N)} n^{-m-1} \int_{\alpha : n \leq \|\alpha\|_{\infty} \leq n+1} \left| \prod_{j=1}^d f \left( \sum_{i=1}^m \alpha_i \lambda_{i,j} \right) \right| \, d\alpha
\]

\[
\ll \delta \sum_{n \geq T(N)} n^{-2} N^{d-m}
\]

\[
\ll \delta T(N)^{-1} N^{d-m} = o_L(N^{d-m})
\]

by Lemma B.3, provided \( \delta(N) \) decays slowly enough.

**Minor arc.** The following is the natural higher-dimensional version of the argument in [18].

**Lemma B.4.** For any positive \( A \) and \( B \),

\[
\sup_{\log^B N \leq \|\alpha\|_{\infty} \leq A} \left| \prod_{j=1}^d f \left( \sum_{i=1}^m \alpha_i \lambda_{i,j} \right) \right| = o_{A,B,L}(N^d).
\]

**Proof.** Note that by the prime number theorem one has the trivial bound \(|f(\alpha)| \ll N\). Assuming for contradiction that the lemma is false, there exists
some $A$, some $B$, and some positive $\varepsilon$ such that, for infinitely many $N$, there exists a vector $\alpha^{(N)} \in \mathbb{R}^m$ satisfying $\frac{\log^B N}{N} \leq \|\alpha^{(N)}\|_\infty \leq A$ and

$$f\left(\sum_{i=1}^m \alpha_i^{(N)} \lambda_{ij}\right) \geq \varepsilon N \quad \forall j \in [d].$$

Then by [18, Lemma 1] it follows that for each such $N$, and for all $j \in [d]$, there exist integers $q_j^{(N)}$ and $a_j^{(N)}$ such that $1 \leq q_j^{(N)} \ll_{\varepsilon} 1$ and

$$q_j^{(N)} \left(\sum_{i=1}^m \alpha_i^{(N)} \lambda_{ij}\right) = a_j^{(N)} + \theta_j^{(N)}, \quad \forall j \in [d],$$

where $|\theta_j^{(N)}| \ll_{\varepsilon} N^{-1}$. We observe that, if $N$ is large enough, we have the bound $d_j^{(N)} \ll_{A,\varepsilon} 1$. Since $A$ and $\varepsilon$ are fixed we may (by taking a subsequence of $N$) assume that both $d_j^{(N)}$ and $a_j^{(N)}$ are independent of $N$. We call these integers $q_j$ and $a_j$ respectively.

Now, suppose that $a_j = 0$ for all $j$. Then

$$\left|\sum_{i=1}^m \alpha_i^{(N)} \lambda_{ij}\right| \ll_{\varepsilon} N^{-1}, \quad \forall j \in [d].$$

Since the map $L^T : \mathbb{R}^m \to \mathbb{R}^d$ is injective, there exists an inverse linear map

$$M : (\text{Im} L^T, \| \cdot \|_\infty) \to (\mathbb{R}^m, \| \cdot \|_\infty),$$

which must necessarily be bounded. Hence $\|\alpha^{(N)}\|_\infty \ll_{\varepsilon} N^{-1}$, which is a contradiction for large enough $N$, since $\|\alpha^{(N)}\|_\infty \gg (\log^B N)/N$. Therefore there exists some $j$ for which $a_j \neq 0$.

Finally, the sequence $\alpha_i^{(N)}$ is contained in a compact domain, and so it must have a convergent subsequence with limit $\alpha_i$, say. Taking this limit, we observe that

$$q_j \sum_{i=1}^m \alpha_i \lambda_{ij} = a_j$$

and so

$$\sum_{i=1}^m \left(\alpha_i \prod_{k=1}^d q_k\right) \lambda_{ij} = a_j \prod_{k \leq d, k \neq j} q_k.$$ 

Hence there exists a vector $\beta \in \mathbb{R}^m \setminus \{0\}$ such that $L^T \beta \in \mathbb{Z}^d$, contradicting the assumptions of Theorem B.1. This proves the lemma. \qed
In the usual fashion, one may use Lemma B.4 to deduce that there is some slowly growing function \( T(N) \) such that \( T(N) \to \infty \) as \( N \to \infty \) such that

\[
\sup_{\|a\|_{\infty} \leq T(N)} \left| \prod_{j=1}^{d} f \left( \sum_{i=1}^{m} a_{i} \lambda_{ij} \right) \right| = o_{B,L}(N^{d}),
\]

the details being given in Section 3 of [18]. Defining the minor arc \( m \) using this function \( T(N) \), we have exactly

\[
\sup_{a \in m} \left| \prod_{j=1}^{d} f \left( \sum_{i=1}^{m} a_{i} \lambda_{ij} \right) \right| = o_{B,L}(N^{d}).
\]

Therefore, picking some positive parameter \( \eta \) that is small enough such that taking \( l = 1 - \eta \) satisfies the hypotheses of Lemma B.3,

\[
\left| \int_{a \in m} \left( \prod_{j=1}^{d} f \left( \sum_{i=1}^{m} a_{i} \lambda_{ij} \right) \right) \left( \prod_{i=1}^{m} \hat{G}_{\pm}(a_{i}) \right) e(a \cdot \nu) \, da \right|
\]

is

\[
\ll \sup_{a \in m} \left| \prod_{j=1}^{d} f \left( \sum_{i=1}^{m} a_{i} \lambda_{ij} \right) \right|^{\eta} \int_{a \in m} \left| \prod_{j=1}^{d} f \left( \sum_{i=1}^{m} a_{i} \lambda_{ij} \right) \right|^{1-\eta} \, da
\]

(B.6)

\[
\ll o_{B,L}(N^{nd})T(N)^{O(1)}N^{d(1-\eta)-m}
\]

\[
\ll o_{B,L}(N^{d-m})
\]

if \( T(N) \) grows slowly enough.

**Major arc.** The analysis of the contribution from the major arc is routine, given the lemmas we established in Appendix A. Let \( c \) be a small positive constant whose exact value may change between each line. By the estimate (7) from [18] one has, for \( a \in M \),

\[
f \left( \sum_{i=1}^{m} a_{i} \lambda_{ij} \right) = \nu \left( \sum_{i=1}^{m} a_{i} \lambda_{ij} \right) + O(N \exp(-c \sqrt{\log N}))
\]

where

\[
\nu(\beta) = \int_{0}^{N} e(\beta x) \, dx.
\]

Since the measure of \( M \) is \( O((\log m N)^{N-m}) \), we have

\[
\int_{a \in M} \left( \prod_{j=1}^{d} f \left( \sum_{i=1}^{m} a_{i} \lambda_{ij} \right) \right) \hat{G}_{\pm}(a) e(a \cdot \nu) \, da
\]

\[
= \int_{a \in M} \left( \prod_{j=1}^{d} \nu \left( \sum_{i=1}^{m} a_{i} \lambda_{ij} \right) \right) \hat{G}_{\pm}(a) e(a \cdot \nu) \, da + O_{B}(N^{d-m} \exp(-c \sqrt{\log N})).
\]
Since
\[ |v(\beta)| \ll N(1 + |\beta|N)^{-1}, \]
we may extend the above integral to all of \( \mathbb{R}^m \) at the cost of an error of \( O(N^{d-m} \log^{-B} N) \). In other words, the contribution from the major arcs is
\[
\int_{x \in [0,N]^d} \int_{\alpha \in \mathbb{R}^m} \left( \prod_{j=1}^d \mathcal{F} \left( \sum_{i=1}^m \alpha_i x_i \lambda_{ij} \right) \right) \mathcal{G}_{\pm}^{\pm}(\alpha) e(\alpha \cdot v) \, d\alpha \, dx + O_B(N^{d-m} \log^{-B} N),
\]
which is
\[
\int_{x \in [0,N]^d} G_{\pm}(Lx + v) \, dx + O_B(N^{d-m} \log^{-B} N).
\]
Fixing a large value of \( B \), since \( \delta = o(1) \) this expression is equal to
\[
\int_{x \in [0,N]^d} 1_{[-\varepsilon,\varepsilon]^m}(Lx + v) \, dx + o_{L,\varepsilon}(N^{d-m}),
\]
by Lemma A.4. This completes the theorem. \( \square \)

**REFERENCES**

[1] A. Baker, *On some diophantine inequalities involving primes*, J. Reine Angew. Math. **228** (1967), 166–181.
[2] A. Balog, *Linear equations in primes*, Mathematika **39** (1992), 367–378.
[3] P.-Y. Bienvenu, *A higher-dimensional Siegel–Walfisz theorem*. Acta Arith. **179** (2017), 79–100.
[4] E. Bombieri, J. B. Friedlander and H. Iwaniec, *Primes in arithmetic progressions to large moduli. II*, Math. Ann. **277** (1987), 361–393.
[5] H. Davenport and H. Heilbronn, *On indefinite quadratic forms in five variables*, J. London Math. Soc. **21** (1946), 185–193.
[6] D. E. Freeman, *Asymptotic lower bounds for Diophantine inequalities*, Mathematika **47** (2000), 127–159.
[7] D. E. Freeman, *Asymptotic bounds for formulas for Diophantine inequalities*, in *Number Theory for the Millennium, II* (Urbana, IL, 2000), A K Peters, Natick, MA, 2002, pp. 57–74.
[8] L. Grafakos, *Classical Fourier Analysis*, Springer, New York, 2008.
[9] B. Green, *Roth’s theorem in the primes*. Ann. of Math. (2) **161** (2005), 1609–1636.
[10] B. Green, *Montréal notes on quadratic Fourier analysis*, in *Additive Combinatorics*, American Mathematical Society, Providence, RI, 2007, pp. 69–102.
[11] B. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math.(2) **167** (2008), 481–547.
[12] B. Green and T. Tao, *Quadratic uniformity of the Möbius function*, Ann. Inst. Fourier (Grenoble) **58** (2008), 1863–1935.
[13] B. Green and T. Tao, *Linear equations in primes*, Ann. of Math. (2) **171** (2010), 1753–1850.
[14] B. Green and T. Tao, *The Möbius function is strongly orthogonal to nilsequences*, Ann. of Math. (2) **175** (2012), 541–566.
[15] B. Green and T. Tao, *The quantitative behaviour of polynomial orbits on nilmanifolds*, Ann. of Math. (2)**175** (2012), 465–540.
[16] B. Green, T. Tao and T. Ziegler, *An inverse theorem for the Gowers $U^{s+1}[N]$-norm*, Ann. of Math. (2) **176** (2012), 1231–1372.

[17] D. R. Heath-Brown, *A new form of the circle method, and its application to quadratic forms*, J. Reine Angew. Math. **481** (1996), 149–206.

[18] S. T. Parsell, *Irrational linear forms in prime variables*, J. Number Theory **97** (2002), 144–156.

[19] T. Tao and V. H. Vu, *Additive Combinatorics*, Cambridge University Press, Cambridge, 2010.

[20] R. C. Vaughan, *Diophantine approximation by prime numbers. I*, Proc. London Math. Soc. (3) **28** (1974), 373–384.

[21] A. Walker, *Gowers norms control diophantine inequalities*, Algebra Number Theory **14** (2020), 1457–1536.

*Aled Walker*

**TRINITY COLLEGE**

**CAMBRIDGE, CB2 1TQ, UNITED KINGDOM**

email: aw530@cam.ac.uk

(Received March 25, 2019 and in revised form October 8, 2019)