GOWERS NORMS CONTROL DIOPHANTINE INEQUALITIES

ALED WALKER

Abstract. A central tool in the study of systems of linear equations with integer coefficients is the Generalised von Neumann Theorem of Green and Tao. This theorem reduces the task of counting the weighted solutions of these equations to that of counting the weighted solutions for a particular family of forms, the Gowers norms $\|f\|_{U^{m+1}[N]}$ of the weight $f$. In this paper we consider systems of linear inequalities with real coefficients, and show that the number of solutions to such weighted diophantine inequalities may also be bounded by Gowers norms. Further, we provide a necessary and sufficient condition for a system of real linear forms to be governed by Gowers norms in this way. In a forthcoming paper we will discuss the case in which the weights are unbounded but suitably pseudorandom, with applications to counting the number of solutions to diophantine inequalities over the primes.

Contents

1. Introduction
2. Historical background and the main theorem
3. Upper bounds
4. Dimension reduction
5. Transfer from $\mathbb{Z}$ to $\mathbb{R}$
6. Normal form
7. Degeneracy relations
8. A Generalised von Neumann Theorem
9. Constructions
Appendix A. Gowers norms
Appendix B. Lipschitz functions
Appendix C. Rank matrix and normal form: proofs
Appendix D. Additional linear algebra
Appendix E. The approximation function in the algebraic case
References

1. Introduction

Diophantine inequalities are a vast and varied topic in analytic number theory (see [2], say). We will focus on a particular class of problems, which are of the following general form. Let $A$ be a set of integers, let $\varepsilon$ be a positive parameter, and let $L$ be an $m$-by-$d$ real matrix, with $d \geq m + 1$. One may ask whether there are infinitely many solutions to

$$\|La\|_{\infty} \leq \varepsilon$$

with all the coordinates of $a$ lying in $A$. Further, letting $N$ be a natural number, one might seek an asymptotic formula (as $N$ tends to infinity) for the number of such solutions that satisfy $\|a\|_{\infty} \leq N$. Much is known about this problem for certain special sets $A$ (see [1, 7, 13, 20, 21, 22]), in particular for the image sets of polynomials. This
The other focus of this paper will be Gowers norms. These norms were introduced around twenty years ago, as part of Gowers’ proof of Szemeredi’s Theorem [10], and since then they have become a fundamental tool in additive combinatorics and in analytic number theory. Their basic theory is covered in [12] and [24], and we recall the main definitions and properties in Appendix A of this paper.

One particular application of Gowers norms is to the study of linear equations with rational coefficients. Indeed, the study of such systems was greatly enhanced by the introduction, by Green and Tao in [13, 15], of a powerful and wide-ranging technique, known as a ‘Generalised von Neumann Theorem’, which can be used to show that Gowers norms are, in some sense, ‘universal’ over all such linear systems: this is [15, Theorem 7.1], and we recall a similar version in Theorem 1.1. It was using this technique, in combination with a deep study of the inverse theory of Gowers norms, that those authors and Ziegler managed to prove that, generically, \( m+2 \) prime variables are adequate to obtain an asymptotic formula for the number of prime solutions to \( m \) linear equations with rational coefficients, rather than the \( 2m+1 \) variables required by the circle method.

Our motivation for trying to combine Gowers norms and diophantine inequalities was the potential of using these ideas to understand (1.1) when \( A \) is the set of primes. However, the purpose of this paper is rather to develop a theory for diophantine inequalities weighted by bounded functions (as opposed to inequalities weighted by the von Mangoldt function). Many additional technical difficulties occur for the primes, and we choose to present a separate paper on these issues.

**Notation:** We will use standard asymptotic notation \( O \), \( o \), and \( \Omega \). 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 \( |f(N)| \leq C g(N) \) 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. (For us, \( 0 \not\in \mathbb{N} \)).

It will be a convenient shorthand to use these symbols in conjunction with minus signs. So, by convention, we determine that expressions such as \(-O(1), -o(1), -\Omega(1)\) are negative, e.g. \( N^{-\Omega(1)} \) refers to a term \( N^{-c} \), where \( c \) is some positive quantity bounded away from 0 as the asymptotic parameter tends to infinity. It will also be convenient to use the Vinogradov symbol \( \ll \), 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 \( g \ll f \). We also adopt the \( \kappa \) notation from [15]: \( \kappa(x) \) denotes any quantity that tends to zero as \( x \) tends to zero, with the exact value being permitted to change from line to line.

All the implied constants may depend on the dimensions of the underlying spaces. These will be obvious in context, and will always be denoted by \( m, d, h, \) or \( s \) (or, in the case of Proposition 6.7 by \( n \)). If an implied constant depends on other parameters, we will denote these by subscripts, e.g. \( O_{C,C,\epsilon}(1) \), or \( f \asymp_{\epsilon} g \).

If \( N \) is a natural number, we use \([N]\) to denote \( \{ n \in \mathbb{N} : n \leq N \} \), whereas \([1,N]\) will be reserved for the closed real interval. For \( x \in \mathbb{R} \), we write \([x] := \lfloor x + \frac{1}{2} \rfloor \) for the nearest integer to \( x \), and \( \| x \| \) for \( | x - [x] | \). This means that there is slight overloading of the notation \([N]\), but the sense will always be obvious in context. When other
norms are present, we may write $\|x\|_{\mathbb{R}/\mathbb{Z}}$ for $\|x\|$ to avoid confusion. For $x \in \mathbb{R}^m$, we let $\|x\|_{\mathbb{R}^m/\mathbb{Z}^m}$ denote $\sup_i |x_i - [x_i]|$.

We always assume that the vector space $\mathbb{R}^d$ is written with respect to the standard basis. 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 $\text{dist}(\{x\}, Y)$. By identifying sets of $m$-by-$d$ matrices with subsets of $\mathbb{R}^{md}$ (by identifying the coefficients of the matrices with coordinates in $\mathbb{R}^{md}$), we may also define $\text{dist}(X, Y)$ when $X$ and $Y$ are sets of matrices of the same dimensions. We will consider a linear map $L : \mathbb{R}^d \to \mathbb{R}^m$ to be synonymous with the $m$-by-$d$ matrix that represents $L$ with respect to the standard bases. The norm $\|L\|_\infty$ will refer to the maximum absolute value of the coefficients of this matrix.

We let $\partial(X)$ denote the topological boundary of a set $X \subset \mathbb{R}^d$. If $A$ and $B$ are two sets with $A \subseteq B$, we let $1_A : B \to \{0, 1\}$ denote the indicator function of $A$. (The relevant set $B$ will usually be obvious from context). The notation for logarithms, log, will always denote the natural log. For $\theta \in \mathbb{R}$ we also adopt the standard shorthand $e^{i\theta}$ to mean $e^{2\pi i \theta}$.

Theorem 1.1 (Generalised von Neumann Theorem for rational forms (non-quantitative)). Let $N, m, d$ be natural numbers, satisfying $d \geq m + 2$. Let $L$ be an $m$-by-$d$ real matrix with integer coefficients, with rank $m$. Suppose that there does not exist any non-zero row-vector in the row-space of $L$ that has two or fewer non-zero coordinates. Then there is some natural number $s$ at most $d - 2$ that satisfies the following. Let $f_1, \cdots, f_d : [N] \to [-1, 1]$ be arbitrary functions, and suppose that

$$\min_j \|f_j\|_{U^{s+1}[N]} \leq \rho$$

for some parameter $\rho$ in the range $0 < \rho \leq 1$. Then

$$\frac{1}{N^{d-m}} \sum_{n \in [N]^d} \prod_{j=1}^d f_j(n_j) \ll_L \rho^\Omega(1) + o_\rho(1).$$

Theorem 1.1 is implicit in [15] but it is not explicitly stated in that paper, the authors’ focus being on results over primes. We will later require a quantitative version (Theorem 4.2), at which point we will describe fully how to extract these statements from [15].

At first sight the non-degeneracy condition in the statement of Theorem 1.1 concerning the row-space of $L$, may seem a little unnatural. However, it is actually a necessary condition for Gowers norms to be used in this way (as we show later in Theorem 2.12).
The main result of this paper (Theorem 2.10) will generalise Theorem 1.1 to the setting of diophantine inequalities. Because we take care to record the quantitative dependencies of the error terms, Theorem 2.10 is rather technical to state. Fortunately, it admits a corollary that is much more transparent. This corollary is strong enough to give our main application (an application to cancellation of the Möbius function, see Corollary 1.11).

**Corollary 1.2.** Let $N, m, d$ be natural numbers, satisfying $d \geq m + 2$, and let $\varepsilon$ be a positive parameter. Let $L : \mathbb{R}^d \rightarrow \mathbb{R}^m$ be an $m$-by-$d$ real matrix, with rank $m$. Suppose that there does not exist any non-zero row-vector in the row-space of $L$ that has two or fewer non-zero coordinates. Then there is some natural number $s$ at most $d - 2$, independent of $\varepsilon$, such that the following is true. Let $f_1, \ldots, f_d : [N] \rightarrow [-1, 1]$ be arbitrary functions, and suppose that

$$\min_j \|f_j\|_{U^{s+1}[N]} \leq \rho,$$

for some parameter $\rho$ in the range $0 < \rho \leq 1$. Then

$$\left| \frac{1}{N^{d-m}} \sum_{n \in [N]^d \atop \|Ln\|_\infty \leq \varepsilon} \left( \prod_{j=1}^d f_j(n_j) \right) \right| \ll_{L, \varepsilon} \rho^{\Omega(1)} + o_{\rho, L}(1).$$

We can provide detailed information about how the implied constant and the $o_{\rho, L}(1)$ term depend on $L$, but we defer those technicalities to Theorem 2.10.

Note that if $L$ has integer coefficients then, by picking $\varepsilon$ small enough, Corollary 1.2 immediately implies Theorem 1.1.

Also note that, due to the nested property of Gowers norms (see Appendix A), Corollary 1.2 may be fruitfully applied under the hypothesis $\min_j \|f_j\|_{U^{d-s}[N]} \leq \rho$.

Let us illustrate Corollary 1.2 with some examples.

**Example 1.3** (Three-term irrational AP). The first example could have been proved by Davenport and Heilbronn using the methods of [7], but we include it here to demonstrate the simplest case in which Corollary 1.2 applies. Let

$$L := \begin{pmatrix} 1 & -\sqrt{2} & -1 + \sqrt{2} \end{pmatrix}.$$ 

Then $m = 1$ and $d = 3$, and manifestly there does not exist any non-zero row-vector in the row-space of $L$ that has two or fewer non-zero coordinates.

Therefore Corollary 1.2 applies, and so, if $f_1, f_2, f_3 : [N] \rightarrow [-1, 1]$ are three functions satisfying $\min_j \|f_j\|_{U^2[N]} \leq \rho$ for some $\rho$ in the range $0 < \rho \leq 1$, we have

$$\left| \frac{1}{N^2} \sum_{n_1, n_2, n_3 \leq N \atop |n_1 - \sqrt{2}n_2 + (-1 + \sqrt{2})n_3| \leq \varepsilon} f_1(n_1)f_2(n_2)f_3(n_3) \right| \ll_{\varepsilon} \rho^{\Omega(1)} + o_{\rho}(1). \quad (1.2)$$

The statement (1.2) admits a different interpretation, which some readers may find more natural, that of counting the number of occurrences of a certain irrational pattern: a ‘three-term irrational arithmetic progression’. Indeed, recall that for $\theta \in \mathbb{R}$ we let $[\theta]$ denote $[\theta + \frac{1}{2}]$, i.e. the nearest integer to $\theta$. Then for any three functions $f_1, f_2, f_3 : [N] \rightarrow [-1, 1]$, we make the definition

$$T(f_1, f_2, f_3) := \frac{1}{N^2} \sum_{x, d \in \mathbb{Z}} f_3(x)f_2(x + d)f_1([x + \sqrt{2}d]). \quad (1.3)$$
Informally speaking, $T$ counts the number of near-occurrences of the pattern $(x, x + d, x + \sqrt{2}d)$, weighted by the functions $f_j$. By a simple change of variables $n_1 = [x + \sqrt{2}d], n_2 = x + d, n_3 = x$, and noting that $x + \sqrt{2}d \notin \frac{1}{2}\mathbb{Z}$, we see

$$T(f_1, f_2, f_3) = \frac{1}{N^2} \sum_{n_1, n_2, n_3 \leq N} f_1(n_1) f_2(n_2) f_3(n_3).$$

(1.4)

By (1.2), this means

$$|T(f_1, f_2, f_3)| \leq \rho^R(1) + o_{\rho}(1),$$

(1.5)

provided $\min_j \|f_j\|_{U^2[N]} \leq \rho$.

One can use these results to count the number of near-occurrences of the pattern $(x, x + d, x + \sqrt{2}d)$ in a Fourier-uniform set, which we do in Corollary 1.4 below. Indeed, suppose that $A$ is a subset of $[N]$ with $|A| = \alpha N$. Let

$$f_A := 1_A - \alpha 1_{[N]}$$

(1.6)

be its so-called ‘balanced function’. By the usual telescoping trick, $T(1_A, 1_A, 1_A)$ is equal to

$$T(\alpha 1_{[N]}, \alpha 1_{[N]}, \alpha 1_{[N]}) + T(f_A, \alpha 1_{[N]}, \alpha 1_{[N]}) + T(1_A, f_A, \alpha 1_{[N]}) + T(1_A, 1_A, f_A).$$

(1.7)

Bounding the final three terms using $\|f_A\|_{U^2[N]}$, and using the relation (1.4), one may establish that

$$\frac{1}{N^2} \sum_{x, d \in \mathbb{Z}} 1_A(x) 1_A(x + d) 1_A([x + \sqrt{2}d])$$

is equal to

$$\frac{\alpha^3}{N^2} \sum_{x, d \in \mathbb{Z}} 1_{[N]}(x) 1_{[N]}(x + d) 1_{[N]}([x + \sqrt{2}d]) + O(\rho^R(1)) + o_{\rho}(1),$$

(1.8)

provided $\|f_A\|_{U^2[N]} \leq \rho$. If $\|f_A\|_{U^2[N]} = o(1)$ then, by picking $\rho = \rho(N)$ to be a quantity that tends to zero suitably slowly as $N$ tends to infinity, the error term in (1.8) can be made to be $o(1)$.

For bounded functions, the $U^2$-norm is closely related to the Fourier transform. Indeed, we say that $A$ is Fourier-uniform if its balanced function $f_A$ satisfies

$$\sup_{\theta \in [0, 1]} \left| \frac{1}{N} \sum_{n \leq N} f_A(n) e(n \theta) \right| = o(1),$$

and it is a standard result (see [23 Exercise 1.3.18]) that $A$ is Fourier uniform if and only if $\|f_A\|_{U^2[N]} = o(1)$. Therefore expression (1.8), and the remarks following it, imply the following corollary.

**Corollary 1.4 (Fourier uniform set).** Let $N$ be a natural number, and let $\beta \in \mathbb{R} \setminus \mathbb{Q}$. If $A$ is a Fourier-uniform subset of $[N]$, with $|A| = \alpha N$, then

$$\frac{1}{N^2} \sum_{x, d \in \mathbb{Z}} 1_A(x) 1_A(x + d) 1_A([x + \beta d])$$

is equal to

$$\frac{\alpha^3}{N^2} \sum_{x, d \in \mathbb{Z}} 1_{[N]}(x) 1_{[N]}(x + d) 1_{[N]}([x + \beta d]) + o_{\beta}(1).$$
Example 1.5. Let
\[ L := \begin{pmatrix} 1 & 0 & -\sqrt{2} & -1 + \sqrt{2} \\ 0 & 1 & -\sqrt{3} & -1 + \sqrt{3} \end{pmatrix} \] (1.9)
We verify that the non-degeneracy condition from Corollary 1.2 is satisfied. Indeed, when \( L \) is an \( m \)-by-\( m + 2 \) matrix, elementary linear algebra shows that there exists a non-zero row-vector in the row-space of \( L \) that has two or fewer non-zero coordinates if and only if there exists some \( m \)-by-\( m \) submatrix of \( L \) that has determinant zero.

With \( L \) as in (1.9), we see that none of the 6 determinants of the 2-by-2 submatrices are zero, and hence Corollary 1.2 applies.

Let \( N \) be a natural number, and let \( f_1, f_2, f_3, f_4 : [N] \to [-1, 1] \) be arbitrary functions. Then
\[
\frac{1}{N^2} \sum_{n \in [N]^4} \left( \prod_{j=1}^{4} f_j(n_j) \right) = \frac{1}{N^2} \sum_{x, d \in \mathbb{Z}} f_1(x) f_2(x + d) f_3([x + \sqrt{2}d]) f_4([x + \sqrt{3}d]).
\]
Corollary 1.2 controls the left-hand side of this expression.

We can summarise the previous two examples in the following general corollary.

Corollary 1.6. Let \( s \) be a natural number, and let \( \theta_1, \ldots, \theta_s \in \mathbb{R} \) be pairwise distinct irrational numbers. Let \( N \) be a natural number, and let \( A \) be a subset of \([N]\) with \( |A| = \alpha N \). Suppose that \( \|f_A\|_{U^{s+1}[N]} = o(1) \). Then
\[
\frac{1}{N^2} \sum_{x, d \in \mathbb{Z}} 1_A(x) 1_A(x + d) \left( \prod_{i=1}^{s} 1_A([x + \theta_i d]) \right)
\]
is equal to
\[
\frac{\alpha^{s+2}}{N^2} \sum_{x, d \in \mathbb{Z}} 1_{[N]}(x) 1_{[N]}(x + d) \left( \prod_{i=1}^{s} 1_{[N]}([x + \theta_i d]) \right) + o(1),
\]
where the \( o(1) \) error term may depend on \( \theta_1, \ldots, \theta_s \).

Proof. Apply Corollary 1.2 to the \( s \)-by-\( s + 2 \) matrix
\[ L = \begin{pmatrix} I & -\theta & -1 + \theta \end{pmatrix}, \]
where \( I \) denotes the identity matrix and \( \theta \) denotes the vector \((\theta_1, \cdots, \theta_s)^T \in \mathbb{R}^s\).

We comment that the infinitary theory of patterns such as (1.10) was previously considered in [19], albeit in the different language of ergodic theory. In particular, an easy deduction from [19] Theorem B shows that all sets of natural numbers with positive upper Banach density contain infinitely many copies of the pattern \((x, x+d, [x+\sqrt{2}d], [x+\sqrt{3}d])\). Yet from [19] one cannot recover any statement that has the generality of Corollary 1.2 nor any asymptotic formula that holds in the Gowers uniform cases (as we deduced above).

Corollary 1.2 has immediate consequences for counting solutions to diophantine inequalities weighted by explicit bounded pseudorandom functions. In particular there is the following natural analogue of [15] Proposition 9.1.

Corollary 1.7 (Möbius orthogonality). Let \( N, m, d \) be natural numbers satisfying \( d \geq m + 2 \), and let \( \varepsilon \) be a positive parameter. Let \( L : \mathbb{R}^d \to \mathbb{R}^m \) be an \( m \)-by-\( d \) real matrix, with rank \( m \). Suppose that there does not exist any non-zero row-vector in the
row-space of $L$ that has two or fewer non-zero coordinates. Let $\mu$ denote the Möbius function. Then
\[ \sum_{n \in [N]^d, \|n\|_\infty \leq \varepsilon} \mu(n_j) \left( \prod_{j=2}^d f_j(n_j) \right) = o_{L,\varepsilon}(N^{d-m}) \]
for any bounded functions $f_2, \cdots, f_d : [N] \to [-1,1]$. The same is true with $\mu$ replaced by the Liouville function $\lambda$.

**Proof.** This follows immediately from Corollary 1.2 and the deep facts (stated in [15], proved in [16] and [17]) that $\|\mu\|_{U^{s+1}[N]} = o_s(1)$ and $\|\lambda\|_{U^{s+1}[N]} = o_s(1)$.

For example, Corollary 1.7 implies that
\[ \sum_{n \in [N]^4, n_1 - n_2 = n_3 - n_4} \mu(n_1) \mu(n_2) \mu(n_3) \mu(n_4) = o(N^2). \] (1.11)

There are of course many such examples; we chose (1.11) to emphasise that one can choose configurations that combine rational and irrational relations.

**Acknowledgements:** We would like to thank an anonymous referee, for their very careful reading of an earlier version of this work, and Ben Green, for his advice and comments. We also benefited from conversations with Sam Chow and Trevor Wooley. Parts of the paper were completed while the author was a Programme Associate at the Mathematical Sciences Research Institute in Berkeley, who provided excellent working conditions. During part of the work the author was supported by EPSRC grant no. EP/M50659X/1.

## 2. Historical background and the main theorem

The aim of this section is to state Theorem 2.10, our main result, and a fully quantitative version of Corollary 1.2. We will also state a partial converse to this theorem; this will be Theorem 2.12.

Before doing this, let us take the opportunity to recall some of the main classical results in the area. As we have already remarked, much is known about the inequality (1.1) for certain special sets $A$, particularly when $m = 1$. If $A$ is the set of squares, say, it was shown by Davenport and Heilbronn in [7] that there are infinitely many solutions to (1.1) for $m = 1$ and $d = 5$, i.e. infinitely many solutions to
\[ |\lambda_1 n_1^2 + \lambda_2 n_2^2 + \lambda_3 n_3^2 + \lambda_4 n_4^2 + \lambda_5 n_5^2| \leq \varepsilon, \]
provided the coefficients $\lambda_i$ are non-zero, not all of the same sign, and not all in pairwise rational ratio. Their work also proves the same result for $k^{th}$ powers, provided that the number of variables is at least $2^k + 1$. The method is Fourier-analytic, replacing the interval $[-\varepsilon, \varepsilon]$ with a smooth cut-off and expressing the solution count via the inversion formula. See [6, Chapter 20], [26, Chapter 11]. Freeman [9] refined the minor-arc analysis from [7] to obtain asymptotic formulas for the number of solutions where $n_i \leq N$ for every $i$. The number of variables required was subsequently reduced by Wooley in [28].

Of course there is much more work on such polynomial questions, only tangentially related to this paper, i.e. Margulis’ solution to the Oppenheim Conjecture [18], and the
subsequent quantitative versions given by Bourgain [3]. Regarding questions with \( m \geq 2 \), Parsell [22] considered the case of \( A \) being the \( k^{th} \) powers, with Müller [20] developing a refined result in the case of inequalities for general real quadratics. Müller’s main result imposes a technical hypothesis on the so-called ‘real pencil’ of the quadratic forms under consideration: we will return to this issue when considering the technical details of our main theorem (Theorem 2.10).

These questions have also been asked when \( A \) is the set of prime numbers, and may be tackled using similar analytic tools. A result first claimed in [1] by Baker states that for any fixed positive \( \varepsilon \) there exist infinitely many triples of primes \( (p_1, p_2, p_3) \) satisfying

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

assuming again that the coefficients \( \lambda_i \) are non-zero, not all of the same sign, and not all in pairwise rational ratio. Parsell [21] then used a similar refinement to that of Freeman to prove a lower bound on the number of solutions to (2.1) satisfying \( p_1, p_2, p_3 \leq N \).

This bound had the expected order of magnitude, namely \( \varepsilon N^2 (\log N)^{-3} \).

These analytic approaches ultimately rely on establishing tight mean-value estimates for certain exponential sums, and thus require a large enough number of variables for such estimates to hold. In the case of primes, say, for \( m \) inequalities the method of Parsell will yield an asymptotic for the number of solutions to (1.1) in prime variables provided \( d \geq 2m + 1 \) (at least for generic \( L \)). In preparation, we have a paper [27] that reaches the same conclusion under the weaker hypothesis \( d \geq m + 2 \) (provided \( L \) has algebraic coefficients).

Having introduced the background of this work, we can begin to build up the necessary notation required in order to state the main theorem (Theorem 2.10). First, let us introduce a multilinear form that will count solutions to a general version of (1.1).

**Definition 2.1.** Let \( N, m, d \) be natural numbers, and let \( L : \mathbb{R}^d \rightarrow \mathbb{R}^m \) be a linear map. Let \( F : \mathbb{R}^d \rightarrow [0, 1] \) and \( G : \mathbb{R}^m \rightarrow [0, 1] \) be two functions, with \( F \) supported on \([-N, N]^d\) and \( G \) compactly supported. Let \( f_1, \cdots, f_d : [N] \rightarrow [-1, 1] \) be arbitrary functions. We define

\[
T_{F,G,N}^L(f_1, \cdots, f_d) := \frac{1}{N^{d-m}} \sum_{n \in \mathbb{Z}^d} \left( \prod_{j=1}^d f_j(n_j) \right) F(n) G(Ln).
\]

The normalisation factor of \( N^{d-m} \) is appropriate; we will show in Lemma 3.2 that \( T_{F,G,N}^L(f_1, \cdots, f_d) \ll_{G, F} 1 \).

Now let us introduce the appropriate notions of ‘non-degeneracy’. These will be needed in order to appropriately quantify the Gowers norm relations in the main theorem (Theorem 2.10).

**Definition 2.2 (Rank varieties).** Let \( m, d \) be natural numbers satisfying \( d \geq m + 1 \). Let \( V_{\text{rank}}(m, d) \) denote the set of all linear maps \( L : \mathbb{R}^d \rightarrow \mathbb{R}^m \) whose rank is less than \( m \). We call \( V_{\text{rank}}(m, d) \) the rank variety.

---

1In fact Baker proved a slightly different result, writing in [1] that the result we quote here followed easily from the then-existing methods. A proof does not seem to have been written down until Parsell [21].

2An asymptotic formula for the number of solutions follows very easily from Parsell’s work, though does not appear to be present in the literature.
Let $V_{\text{rank}}^{\text{global}}(m,d)$ denote the set of all linear maps $L : \mathbb{R}^d \to \mathbb{R}^m$ for which there exists a standard basis vector of $\mathbb{R}^d$, say $e_i$, for which $L|_{\text{span}(e_j; j \neq i)}$ has rank less than $m$. We call $V_{\text{rank}}^{\text{global}}(m,d)$ the global rank variety.

We remark that $V_{\text{rank}}^{\text{global}}(m,d)$ contains $V_{\text{rank}}(m,d)$.

The next notion is a rephrasing of the non-degeneracy condition that appeared in Corollary 1.2 and Theorem 1.1.

**Definition 2.3** (Dual degeneracy variety). Let $m,d$ be natural numbers satisfying $d \geq m + 2$. Let $e_1, \cdots, e_d$ denote the standard basis vectors of $\mathbb{R}^d$, and let $e_1^*, \cdots, e_d^*$ denote the dual basis of $(\mathbb{R}^d)^*$. Then let $V_{\text{degen}}^*(m,d)$ denote the set of all linear maps $L : \mathbb{R}^d \to \mathbb{R}^m$ 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 $(e_i^* - \lambda e_j^*) \in L^*((\mathbb{R}^m)^*)$. We call $V_{\text{degen}}^*(m,d)$ the dual degeneracy variety.

Though defined as sets of linear maps, by fixing bases we can view $V_{\text{rank}}(m,d)$ and $V_{\text{degen}}^*(m,d)$ as sets of matrices. In that language, one can easily verify that an $m$-by-$d$ matrix $L$ is in $V_{\text{degen}}^*(m,d)$ precisely when there exists a non-zero row-vector in the row-space of $L$ that has two or fewer non-zero coordinates. The formulation in terms of dual spaces will be particularly convenient for some of the algebraic manipulations in section 4, however. We remark that $V_{\text{degen}}^*(m,d)$ contains $V_{\text{rank}}^{\text{global}}(m,d)$.

If $L : \mathbb{R}^d \to \mathbb{R}^m$ is a surjective linear map, it is certainly true that $\text{span}(L(\mathbb{Z}^d)) = \mathbb{R}^m$. But $L(\mathbb{Z}^d)$ needn’t be dense in $\mathbb{R}^m$, as it may satisfy some rational relations.

**Definition 2.4** (Rational dimension, rational map, purely irrational). Let $m$ and $d$ be natural numbers, with $d \geq m + 1$. 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 $\Theta : \mathbb{R}^m \to \mathbb{R}^n$ for which $\Theta L(\mathbb{Z}^d) \subseteq \mathbb{Z}^n$. 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 $u = 0$.

For example, suppose that $L : \mathbb{R}^4 \to \mathbb{R}^2$ is the linear map represented by the matrix

$$L := \begin{pmatrix} 1 & 0 & -\sqrt{2} & -\sqrt{3} + 1 \\ 0 & 1 & 5\sqrt{2} & 5\sqrt{3} \end{pmatrix}.$$ 

If $\Theta : \mathbb{R}^2 \to \mathbb{R}$ is given by the matrix

$$\Theta := \begin{pmatrix} 5 \\ 1 \end{pmatrix},$$

then $\Theta L(\mathbb{Z}^4) \subseteq \mathbb{Z}$, and in fact $\Theta L(\mathbb{Z}^4) = \mathbb{Z}$. So the rational dimension of $L$ is at least 1. But the rational dimension of $L$ cannot be 2, as if there were a surjective map $\Theta : \mathbb{R}^2 \to \mathbb{R}^2$ such that $\Theta L(\mathbb{Z}^4) \subseteq \mathbb{Z}^2$ then $L(\mathbb{Z}^4)$ would be the subset of a 2-dimensional lattice, which it is not. So the rational dimension of $L$ is equal to 1.

Earlier in this section we remarked that Müller, in the work [20], imposed a technical hypothesis on the so-called ‘real pencil’ of the quadratic forms under consideration. In our language, Müller was trying find conditions for when $T^{\mathbb{L}}_{F,G,N}(f_1, \cdots, f_d) > 0$ in the case where the functions $f_j$ are supported on the image of quadratic monomials. In this language, one of the hypotheses he imposes on $L$ is exactly that $L$ should be purely irrational. We work in a more general framework, considering all $L$, including
Lemma 2.5. Let $m$ and $d$ be natural numbers, with $d \geq m + 1$. Let $L : \mathbb{R}^d \to \mathbb{R}^m$ be a surjective linear map, and let $u$ be the rational dimension of $L$. Then if $\Theta_1, \Theta_2 : \mathbb{R}^m \to \mathbb{R}^u$ are two rational maps for $L$, $\ker \Theta_1 = \ker \Theta_2$.

Proof. Suppose that $\Theta_1, \Theta_2 : \mathbb{R}^m \to \mathbb{R}^u$ are two rational maps for $L$ for which $\ker \Theta_1 \neq \ker \Theta_2$. Then consider the map $(\Theta_1, \Theta_2) : \mathbb{R}^m \to \mathbb{R}^{2u}$. The kernel of this map has dimension at most $m - u - 1$, as it is the intersection of two different subspaces of dimension $m - u$. Therefore the image has dimension at least $u + 1$.

Also, $((\Theta_1, \Theta_2) \circ L)(\mathbb{Z}^d) \subseteq \mathbb{Z}^{2u}$. Let $\Phi$ be any surjective map from $\text{im}((\Theta_1, \Theta_2))$ to $\mathbb{R}^{u+1}$ for which $\Phi(\mathbb{Z}^{2u} \cap \text{im}((\Theta_1, \Theta_2))) \subseteq \mathbb{Z}^{u+1}$. Then $\Phi \circ (\Theta_1, \Theta_2) : \mathbb{R}^m \to \mathbb{R}^{u+1}$ is surjective and $(\Phi \circ (\Theta_1, \Theta_2) \circ L)(\mathbb{Z}^d) \subseteq \mathbb{Z}^{u+1}$. This contradicts the definition of $u$ as the rational dimension. \hfill \Box

The quantitative aspects of such relations will be required in order to properly state the main theorem (Theorem 2.10). Recall that for all linear maps between vector spaces of the form $\mathbb{R}^a$, we identify them with their matrix representation with respect to the standard bases. Also recall that for a linear map $\Theta : \mathbb{R}^m \to \mathbb{R}^u$, we use $\|\Theta\|_\infty$ to denote the maximum absolute value of the coefficients of its matrix.

Definition 2.6 (Rational complexity). Let $m$ and $d$ be natural numbers, with $d \geq m + 1$. Let $L : \mathbb{R}^d \to \mathbb{R}^m$ be a surjective linear map, and let $u$ denote the rational dimension of $L$. We say that $L$ has rational complexity at most $C$ if there exists a map $\Theta$ that is a rational map for $L$ and for which $\|\Theta\|_\infty \leq C$.

If $L$ is purely irrational, then $L$ has rational complexity 0.

A linear map with maximal rational dimension is equivalent to a linear map with integer coefficients, in the following sense:

Lemma 2.7. Let $m$ and $d$ be natural numbers, with $d \geq m + 1$. Let $L : \mathbb{R}^d \to \mathbb{R}^m$ be a surjective linear map, and suppose that $L$ has rational dimension $m$ and rational complexity at most $C$. Then there exists an invertible $m$-by-$m$ matrix $\Theta$ and an $m$-by-$d$ matrix $S$ with integer coefficients such that, as matrices, $\Theta L = S$. Furthermore, $\|\Theta\|_\infty \leq C$.

Proof. Let $\Theta : \mathbb{R}^m \to \mathbb{R}^m$ be a rational map for $L$ for which $\|\Theta\|_\infty \leq C$. \hfill \Box

We will use this lemma in section 4 to reduce the study of maps $L$ with maximal rational dimension to the study of maps $L$ with integer coefficients, which were considered in [15] (see Theorems 1.1 and 4.2).

We must quantify the rational relations in a second way. Indeed, $L$ might have rational dimension $u$ but be extremely close to having rational dimension at least $u + 1$, in the sense that there might exist some surjective linear map $\Theta : \mathbb{R}^m \to \mathbb{R}^{u+1}$ such that the matrix of $\Theta L$ is very close to having integer coefficients. This phenomenon, essentially a notion of diophantine approximation, will also have a quantitative effect on our final bounds. We introduce the following definition:
Definition 2.8 (Approximation function). Let $m$ and $d$ be natural numbers, with $d \geq m+1$. Let $L : \mathbb{R}^d \to \mathbb{R}^m$ be a surjective linear map, and let $u$ denote the rational dimension of $L$. Let $\Theta : \mathbb{R}^m \to \mathbb{R}^u$ be any rational map for $L$. Suppose that $u \leq m-1$. We then define the approximation function of $L$, denoted $A_L : [0,1] \times (0,1] \to (0,\infty)$ by

$$A_L(\tau_1, \tau_2) := \inf_{\varphi \in (\mathbb{R}^m)^*} \text{dist}(L^*\varphi, (\mathbb{Z}^d)^T),$$

where $(\mathbb{Z}^d)^T$ denotes the set of those $\varphi \in (\mathbb{R}^d)^*$ that have integer coordinates with respect to the standard dual basis.

If $u = m$, we define $A_L(\tau_1, \tau_2)$ to be identically equal to $\tau_1$.

Let us unpack this definition, before giving some examples. Firstly, note that the definition is independent of the choice of $\Theta$. Indeed, $\Theta^\circ((\mathbb{R}^u)^*) = (\ker \Theta)^\circ$ which, by Lemma 2.5, is independent of $\Theta$. Regarding the notion ‘dist’, we remind the reader that we consider $a$-by-$b$ matrices $M$ as elements of $\mathbb{R}^{ab}$, simply by identifying the coefficients of $M$ with coordinates in $\mathbb{R}^{ab}$. The $\ell^\infty$ norm and the dist operator may then be defined on matrices, i.e. if $V$ is a set of $a$-by-$b$ matrices, and $L$ is an $a$-by-$b$ matrix, then

$$\text{dist}(L, V) := \inf_{L' \in V} \|L - L'\|_\infty.$$ 

In this instance we are working with 1-by-$d$ matrices, i.e. elements of $(\mathbb{R}^d)^*$.

Let us consider a simple example. Suppose that, as a matrix,

$$L := \begin{pmatrix} 1 & -\sqrt{2} & -1 + \sqrt{2} \end{pmatrix},$$

as in Example 1.3. Then $A_L(\tau_1, \tau_2)$ is equal to

$$\inf_{k \in \mathbb{R} : \tau_1 \leq |k| \leq \tau_2^{-1}} \max(\|k\|_\mathbb{R}/\mathbb{Z}, \| - k\sqrt{2}\|_\mathbb{R}/\mathbb{Z}, \| - k + \sqrt{2}\|_\mathbb{R}/\mathbb{Z}).$$

We claim that

$$A_L(\tau_1, \tau_2) \gg \min(\tau_1, \tau_2).$$

Indeed, we know that, for all $q \in \mathbb{N}$, $\|q\sqrt{2}\|_\mathbb{R}/\mathbb{Z} \geq 1/(10q)$. This is the statement that $\sqrt{2}$ is a badly approximable irrational. The proof is straightforward: if there were some natural number $p$ for which $|q\sqrt{2} - p| < 1/(10q)$, then

$$1 \leq |2q^2 - p^2| < \frac{\sqrt{2}}{10} + \frac{p}{10q} < \frac{\sqrt{2}}{5} + \frac{1}{10},$$

which is a contradiction.

Suppose first that $\|k\|_\mathbb{R}/\mathbb{Z} \leq \tau_2/100$ and $1/2 \leq |k| \leq \tau_2^{-1}$. Then, replacing $k$ by $\lfloor k \rfloor$ (the nearest integer to $k$), we can conclude that

$$\max(\| - k\sqrt{2}\|_\mathbb{R}/\mathbb{Z}, \| - k + \sqrt{2}\|_\mathbb{R}/\mathbb{Z}) \geq \|[k]\sqrt{2}\|_\mathbb{R}/\mathbb{Z} - \frac{\tau_2}{50}$$

$$\geq \frac{1}{10\lfloor k \rfloor} - \frac{\tau_2}{50}$$

$$\geq \frac{1}{10\tau_2^{-1} + 10} - \frac{\tau_2}{50} \gg \tau_2.$$ 

Otherwise, one has

$$\|k\|_\mathbb{R}/\mathbb{Z} \gg \min(\tau_1, \tau_2).$$
Therefore, 
\[ A_L(\tau_1, \tau_2) \gg \min(\tau_1, \tau_2) \]
as claimed.

Such a function is clearly rather tame. In fact, it is not too difficult to show that if \( L \) is an \( m \)-by-\( d \) matrix with rank \( m \) and with algebraic coefficients, then 
\[ A_L(\tau_1, \tau_2) \gg L \min(\tau_1, \tau_2^{O_L(1)}), \]
where the \( O_L(1) \) term in the exponent depends on the algebraic degree of the coefficient \( L \) of \( L \). We shall sketch a proof of this statement in section \( \ref{sec:proof} \). In general, however, \( A_L(\tau_1, \tau_2) \) could tend to zero arbitrarily quickly as \( \tau_1 \) tends to zero, for example in the case when \( L = (1 \ -\lambda \ -1 + \lambda) \) and \( \lambda \) is a Liouville number (an irrational number that may be very well approximated by rationals).

Yet, however fast \( A_L(\tau_1, \tau_2) \) decays, we have the following critical claim:

**Claim 2.9.** For all permissible choices of \( L, \tau_1 \) and \( \tau_2 \) in Definition \( \ref{def:AL} \), \( A_L(\tau_1, \tau_2) \) is positive.

**Proof.** Let \( u \) be the rational dimension of \( L \). Without loss of generality we may assume that \( u \leq m - 1 \). Then, for all \( \varphi \in (\mathbb{R}^m)^* \setminus \Theta^*((\mathbb{R}^u)^*) \) we have that \( \text{dist}(L^*\varphi, (\mathbb{Z}^d)^T) > 0 \). (If this were not the case then the map \( (\Theta, \varphi) : \mathbb{R}^m \rightarrow \mathbb{R}^{u+1} \) would contradict the definition of \( u \).) Therefore, as the definition of \( A_L(\tau_1, \tau_2) \) involves taking the infimum of a positive continuous function over a compact set, \( A_L(\tau_1, \tau_2) \) is positive. \( \square \)

One might ask why we chose to formulate Definition \( \ref{def:AL} \) in terms of a general \( \varphi \in (\mathbb{R}^m)^* \), instead of one with integer coordinates, when in practice the calculation of \( A_L(\tau_1, \tau_2) \) quickly reduces to considering those \( \varphi \) with integer coordinates. This will certainly be true in the one lemma of this paper where \( A_L \) plays a significant role, namely Lemma \( \ref{lem:one} \). Our first reason is that we find the definition as stated more natural, in that it does not presuppose that any of the coordinates of \( L \) are integers; our second reason is that, when one comes to apply these ideas to the setting of the primes, one is drawn to estimate certain sieve expressions using the Davenport-Heilbronn method. This method involves estimating an integral over \( \varphi \in (\mathbb{R}^m)^* \), where one wishes to control the minor arc contribution by \( A_L(\tau_1, \tau_2) \), and so it is natural that the variable \( \varphi \) should be allowed to vary continuously. More details will appear in \( \ref{sec:applications} \).

Having laid the necessary groundwork, we may now state the main theorem of this paper.

**Theorem 2.10 (Main Theorem).** Let \( N, m, d \) be natural numbers, satisfying \( d \geq m+2 \), and let \( \varepsilon, c, C, C' \) be positive reals. Let \( L = L(N) : \mathbb{R}^d \rightarrow \mathbb{R}^m \) be a surjective linear map that satisfies \( \|L\|_\infty \leq C \). Let \( A_L : (0, 1] \times (0, 1] \rightarrow (0, \infty) \) be the approximation function of \( L \). Suppose further that \( \text{dist}(L, V_{\text{deg}}^*(m, d)) \geq c \), and that \( L \) has rational complexity at most \( C' \). Then there exists a natural number \( s \) at most \( d-2 \), independent of \( \varepsilon \), such that the following is true. Let \( F : \mathbb{R}^d \rightarrow [0, 1] \) be the indicator function of \( [1, N]^d \), and let \( G : \mathbb{R}^m \rightarrow [0, 1] \) be the indicator function of a convex domain.

---

8One could perhaps remove this dependence by using the Schmidt subspace theorem, though as there are power losses throughout the rest of the argument there does not seem to be a great advantage in doing so.
contained in \([-\varepsilon, \varepsilon]^m\). Let \(f_1, \ldots, f_d : [N] \rightarrow [-1, 1]\) be arbitrary functions, and suppose that

\[
\min_j \|f_j\|_{U^{s+1}[N]} \leq \rho,
\]

for some parameter \(\rho\) in the range \(0 < \rho \leq 1\). Then

\[
T_{F,G,N}^L(f_1, \ldots, f_d) \ll_{c,c,c',x} \rho^{O(1)} + o_{\rho,A_L,c,C',C'}(1)
\]

as \(N \rightarrow \infty\). \(o_{\rho,A_L,c,C',C'}(1)\) term may be bounded above by

\[
N^{-O(1)} \rho^{-O(1)} A_L(\Omega_cC',C')(1, \rho)^{-1}.
\]

We remind the reader that the implied constants may depend on the dimensions \(m\) and \(d\). Also note that in the above statement one may replace \(C\) and \(C'\) by a single constant \(C\), without weakening the conclusion. We proceed with this assumption. Observe also that the non-degeneracy condition \(\text{dist}(L, V^*_{\text{degen}}(m, d)) \geq c\) is a quantitative refinement of the non-degeneracy condition on the row-space of \(L\) in Theorem \ref{thm:2.10} and Corollary \ref{cor:1.2}.

Since \(A_L(\Omega_cC,1,\rho)^{-1}\) is finite (by Claim \ref{claim:2.9}), Theorem \ref{thm:2.10} immediately implies Corollary \ref{cor:1.2} from the start of this paper. From (2.3), or rather our full quantitative version Lemma \ref{lem:E.1}, we also have the following corollary:

**Corollary 2.11.** Assume the same hypotheses as Theorem \ref{thm:2.10}, and assume further that \(L\) has algebraic coefficients with algebraic degree at most \(k\). Let \(H\) denote the maximum absolute value of all of the coefficients of all of the minimal polynomials of the coefficients of \(L\). Then

\[
T_{F,G,N}^L(f_1, \ldots, f_d) \ll_{c,c,c,H} \rho^{O(1)} + N^{-O(1)} \rho^{-O(1)}.
\]

At this juncture, it might not be clear why so many quantitative non-degeneracy conditions were required in the statement of Theorem \ref{thm:2.10}. To try to illuminate this issue, we will also prove the following partial converse to Theorem \ref{thm:2.10}, demonstrating that the non-degeneracy condition involving \(V^*_{\text{degen}}(m, d)\) is necessary in order to use Gowers norms in this way.

**Theorem 2.12.** Let \(m, d\) be natural numbers, satisfying \(d \geq m + 2\), and let \(\varepsilon, c, C\) be positive constants. For each natural number \(N\), let \(L = L(N) : \mathbb{R}^d \rightarrow \mathbb{R}^m\) be a linear map satisfying \(|L|_\infty \leq C\). Let \(F : \mathbb{R}^d \rightarrow [0, 1]\) denote the indicator function of \([1, N]^d\) and \(G : \mathbb{R}^m \rightarrow [0, 1]\) denote the indicator function of \([-\varepsilon, \varepsilon]^m\). Assume further that \(\text{dist}(L, V^*_{\text{rank}}(m, d)) \geq c\) and that \(T_{F,G,N}^L(1, \ldots, 1) \gg_{c,c,\varepsilon} 1\) for large enough \(N\).

Suppose that

\[
\liminf_{N \rightarrow \infty} \text{dist}(L, V^*_{\text{degen}}(m, d)) = 0.
\]

Let \(s\) be a natural number, and let \(H : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}\) be any function satisfying \(H(\rho) = o(\rho)\), and let \(E_\rho(N)\) denote some error term depending on a parameter \(\rho\) and satisfying \(E_\rho(N) = o_\rho(1)\). Then one can find infinitely many natural numbers \(N\) such that there exist functions \(f_1, \ldots, f_d : [N] \rightarrow [-1, 1]\) and some \(\rho\) at most 1 such that both

\[
\min_j \|f_j\|_{U^{s+1}[N]} \leq \rho
\]

and

\[
|T_{F,G,N}^L(f_1, \ldots, f_d)| > H(\rho) + E_\rho(N).
\]
In other words, the conclusion of Theorem 2.10 cannot possibly hold if \( \text{dist}(L, V^*_\text{degen}(m, d)) \) is arbitrarily close to 0, even if one replaces the \( \rho^{\Omega(1)} \) dependence on \( \rho \) with a function \( H(\rho) \) that could potentially decay to zero arbitrarily slowly as \( \rho \) tends to zero.

**Example 2.13.** Suppose
\[
L = \begin{pmatrix} 1 + N^{-1/2} & 2 & \pi & -\pi + \sqrt{2} \\ N^{-1/2} & 2 & \sqrt{3} + N^{-1/2} & -\sqrt{3} + N^{-1/2} & e \end{pmatrix}.
\]
Then \( L \) has rank 2 and \( L \notin V^*_\text{degen}(2, 4) \). If one considers Theorem 1.1, one might therefore hope to apply the theory of Gowers norms to bound the number of solutions to inequalities given by \( L \). However, by considering perturbations of the first two columns, we see that \( \text{dist}(L, V^*_\text{degen}(2, 4)) = o(1) \). (Indeed, one may perturb \( L \) by \( O(N^{-1/2}) \) such that there is a vector \((0, 0, x_3, x_4)\) in the row space). Therefore Theorem 2.10 does not apply in this case, despite the fact that \( L \notin V^*_\text{degen}(2, 4) \). Furthermore, Theorem 2.12 shows that, in fact, we cannot possibly use Gowers norms to control inequalities given by such an \( L \). This example is informative, as it shows us that whatever methods we use to prove Theorem 2.10, these methods must break down when applied to such an \( L \), despite the fact that \( L \notin V^*_\text{degen}(2, 4) \).

The proof of Theorem 2.10 will be rather involved. It is tempting to think that the result would follow more easily from taking rational approximations of the coefficients of \( L \), and then using the existing Generalised von Neumann Theorem (a quantitative version of Theorem 1.1) as a black box. Though of course we cannot completely rule out an alternative approach to that of this paper, it seems that such an argument will only quickly succeed if the coefficients of \( L \) are all extremely well-approximable, else the height of the rational approximations becomes too great to apply [15, Theorem 7.1]. One must find an alternative method for other maps \( L \).

To finish this introduction, and to assist the reader, we now describe the overall structure of the paper, and also indicate our proof strategy.

If in the statement of Theorem 2.10 one replaces the convex cut-offs \( F \) and \( G \) with Lipschitz cut-offs, then the expression \( T_{F,G,N}(f_1, \ldots, f_d) \) may be bounded by Gowers norms using a relatively straightforward argument, which we present in sections 5 through 8. In section 5 we introduce a new approximation argument, in which we replace the solution count \( T_{F,G,N}(f_1, \ldots, f_d) \) by a related solution count \( \widetilde{T}_{F,G,N}(\tilde{f}_1, \ldots, \tilde{f}_d) \), which, rather than being a discrete summation over \( \mathbb{Z}^d \), is an integral over \( \mathbb{R}^d \). The expression \( \widetilde{T}_{F,G,N}(\tilde{f}_1, \ldots, \tilde{f}_d) \) may be analysed using the Cauchy-Schwarz inequality in a way that is almost identical to the proof of the usual Generalised von Neumann Theorem [15, Theorem 7.1]. We perform this manipulation in section 8 using results of sections 6 (in which we recap the notion of normal form from [15]) and 7 (in which we relate certain notions of linear-algebraic degeneracy). This argument makes no mention of the approximation function \( A_L \).

So it remains to reduce Theorem 2.10 to the version with Lipschitz cut-offs (we explicitly state this version in Theorem 4.6). Unfortunately, if \( L \) is not purely irrational then there are substantial technical difficulties in replacing \( G \) with a Lipschitz cut-off. To circumvent these difficulties, in section 4 we give an intricate (though ultimately
elementary) linear algebraic argument that reduces Theorem 2.10 to the case where \( L \) is purely irrational, at which point one may easily replace the functions \( F \) and \( G \) with Lipschitz cut-offs, using the upper bounds for \( T_{F,G,N}(1, \cdots, 1) \) that are proved in section 3. This argument thus resolves the main theorem (Theorem 2.10), and all the associated corollaries.

Section 9 deals solely with the proof of the partial converse, namely Theorem 2.12, and may be read largely independently of the other sections. Using a semi-random method, we explicitly construct functions \( f_1, \cdots, f_d \) that satisfy (2.5).

The appendices contain information on Gowers norms; some material on Lipschitz functions; the proofs of the statements from section 6; an assortment of other short linear algebraic lemmas that we require; and finally an illustration of how one may control the approximation function \( A_L \) in the case when \( L \) has algebraic coefficients.

Remark 2.14. Many of the implied constants throughout the paper will depend on the parameter \( \varepsilon \) from the statement of Theorem 2.10. Ultimately, the implied constant in (2.4) tends to infinity as \( \varepsilon \) tends to zero, as our approximation argument in section 5 will not be efficient in powers of \( \varepsilon \). Yet, to save our notation from becoming unreadable, we choose not to keep track of the precise behaviour of implied constants involving \( \varepsilon \).

3. Upper bounds

This section is devoted to proving three upper bounds on the expression \( T_{F,G,N}(1, \cdots, 1) \). (For the definition of this quantity, the reader may refer to Definition 2.1).

The following proposition, which represents a quantitative version of the ‘row-rank equals column-rank’ principle, will be useful throughout.

**Proposition 3.1 (Rank matrix).** Let \( m, d \) be natural numbers, with \( d \geq m + 1 \). Let \( c, C \) be positive constants. For a natural number \( N \), let \( L = L(N) : \mathbb{R}^d \rightarrow \mathbb{R}^m \) be a surjective linear map, denoted by matrix \( (\lambda_{ij})_{i \leq m, j \leq d} \), and assume that \( \|L\|_\infty \leq C \) and \( \text{dist}(L, V_{\text{rank}}(m, d)) \geq c \). Then there exists a matrix \( M \) that is an \( m \)-by-\( m \) submatrix of \( L \) and enjoys the following properties:

1. \( |\det M| = \Omega_{c,C}(1) \);
2. \( \|M^{-1}\|_\infty = O_{c,C}(1) \).

We call such a matrix \( M \) a rank matrix of \( L \). Furthermore,

3. Let \( \mathbf{v} \in \mathbb{R}^d \) be a vector such that \( \mathbf{v}^T \) is in the row-space of \( L \), and suppose that \( \|\mathbf{v}\|_\infty \leq C_1 \) for some positive constant \( C_1 \). Then for all \( i \) in the range \( 1 \leq i \leq m \), there exist coefficients \( a_i \) satisfying \( |a_i| = O_{c,C,C_1}(1) \) such that \( \sum_{i=1}^{m} a_i \lambda_{ij} = v_j \) for all \( j \) in the range \( 1 \leq j \leq d \).

Finally,
(4) If $L$ satisfies the stronger hypothesis $\text{dist}(L, V_{\text{rank}}^{\text{global}}(m, d)) \geq c$, then, for each $j$, there exists a rank matrix of $L$ that doesn’t include the $j^{th}$ column of $L$.

We defer the proof to Appendix C.

Our first upper bound is exceptionally crude, but will nonetheless be useful in section 5.

**Lemma 3.2.** Let $N, m, d$ be natural numbers, satisfying $d \geq m + 1$, and let $c, C, \varepsilon$ be positive constants. Let $L : \mathbb{R}^d \rightarrow \mathbb{R}^m$ be a surjective linear map, and suppose that $\|L\|_{\infty} \leq C$ and $\text{dist}(L, V_{\text{rank}}(m, d)) \geq c$. Let $F : \mathbb{R}^d \rightarrow [0, 1]$ and $G : \mathbb{R}^m \rightarrow [0, 1]$ be two functions, with $F$ supported on $[-N, N]^d$ and $G$ supported on $[-\varepsilon, \varepsilon]^m$. Then

$$T_{F,G,N}^L(1, \ldots, 1) \ll_{c,C,\varepsilon} \|G\|_{\infty}.$$ 

**Proof.** Let $M$ be a rank matrix of $L$ (Proposition 3.1), and suppose without loss of generality that $M$ consists of the first $m$ columns of $L$. For $j$ in the range $m + 1 \leq j \leq d$, let the vector $v_j \in \mathbb{R}^m$ be the $j^{th}$ column of the matrix $M^{-1}L$. Then $N^{d-m}T_{F,G,N}^L(1, \ldots, 1) \leq \|G\|_{\infty} \cdot Z$, where $Z$ is the number of solutions to

$$\begin{pmatrix} n_1 \\ \vdots \\ n_m \end{pmatrix} + \sum_{j=m+1}^d \sigma_j n_j \in M^{-1}([-\varepsilon, \varepsilon]^m)$$

in which $n_1, \ldots, n_d$ are integers satisfying $|n_1|, \ldots, |n_d| \leq N$. Fixing a choice of the variables $n_{m+1}, \ldots, n_d$ forces the vector $(n_1, \ldots, n_m)^T$ to lie in a convex region of diameter $O_{c,C,\varepsilon}(1)$. There are at most $O_{c,C,\varepsilon}(1)$ such points, so $Z \ll_{c,C,\varepsilon} N^{d-m}$. The claimed bound follows.

Our second estimate is a slight strengthening of the above, albeit under stronger hypotheses.

**Lemma 3.3.** Let $N, m, d$ be natural numbers, with $d \geq m + 1$, and let $c, C, \varepsilon$ be positive constants. Let $L : \mathbb{R}^d \rightarrow \mathbb{R}^m$ be a surjective linear map, and suppose that $\|L\|_{\infty} \leq C$ and $\text{dist}(L, V_{\text{rank}}^{\text{global}}(m, d)) \geq c$. Let $\sigma$ be a real number in the range $0 < \sigma < 1/2$. Let $F : \mathbb{R}^d \rightarrow [0, 1]$ and $G : \mathbb{R}^m \rightarrow [0, 1]$ be two functions, with $F$ supported on

$$\{x \in \mathbb{R}^d : \text{dist}(x, \partial([-1, 1]^d)) \leq \sigma N\}$$

and $G$ supported on $[-\varepsilon, \varepsilon]^m$. Then

$$T_{F,G,N}^L(1, \ldots, 1) \ll_{c,C,\varepsilon} \sigma \|G\|_{\infty}.$$ 

**Proof.** Without loss of generality, we may assume that $F$ is supported on

$$\{x \in \mathbb{R}^d : \|x\|_{\infty} \leq 2N, |x_d - 1| \leq \sigma N\}$$

or

$$\{x \in \mathbb{R}^d : \|x\|_{\infty} \leq 2N, |x_d - N| \leq \sigma N\}$$

Consider the first case. By Proposition 3.1 there exists a rank matrix $M$ that does not contain the column $d$. By reordering columns, we can assume without loss of generality that $M$ consists of the first $m$ columns of $L$. Continuing as in the proof of Lemma 3.2 for $j$ in the range $m + 1 \leq j \leq d$, let the vector $v_j \in \mathbb{R}^m$ be the
jth column of the matrix $M^{-1}L$. Then the expression $N^{d-m}T_{F,G,N}^L(1,\cdots,1)$ may be bounded above by $\|G\|_\infty$ times the number of solutions to

$$
\begin{pmatrix}
\ell_1 \\
\vdots \\
\ell_{n_m}
\end{pmatrix} + \sum_{j=m+1}^d v_j n_j \in M^{-1}([-\varepsilon,\varepsilon]^m)
$$

satisfying $|n_1|, \cdots, |n_{d-1}| \leq 2N$ and $|n_d| \leq \sigma N$. We conclude as in the previous proof.

In the second case, the relevant equation is

$$
\begin{pmatrix}
\ell_1 \\
\vdots \\
\ell_{n_m}
\end{pmatrix} + \sum_{j=m+1}^d v_j n_j + (N-1)v_d \in M^{-1}([-\varepsilon,\varepsilon]^m),
$$

in which we count solutions satisfying $|n_1|, \cdots, |n_{d-1}| \leq 2N$ and $|n_d-1| \leq \sigma N$. We conclude as in the previous proof. \hfill \square

Our third estimate is more refined, and will be needed in section 2 when we deduce the main result (Theorem 2.10) from Theorem 4.6. It will help us to replace the sharp cut-off $1_{[-\varepsilon,\varepsilon]^m}$ with a Lipschitz cut-off. For the definition of the approximation function $A_L$, we refer the reader to Definition 2.8.

**Lemma 3.4.** Let $N,m,d$ be natural numbers, with $d \geq m + 1$. Let $c,C,\varepsilon$ be positive constants, and let $\sigma_G$ be a parameter in the range $0 < \sigma_G < 1/2$. Suppose that $L : \mathbb{R}^d \rightarrow \mathbb{R}^m$ is a purely irrational surjective linear map, satisfying $\|L\|_\infty \leq C$ and $\text{dist}(L,V_{\text{rank}}(m,d)) \geq c$. Let $L$ denote the approximation function of $L$. Let $F : \mathbb{R}^d \rightarrow [0,1]$ be supported on $[-N,N]^d$, and let $G : \mathbb{R}^m \rightarrow [0,1]$ be a Lipschitz function, with Lipschitz constant $O(1/\sigma_G)$, supported on $[-\varepsilon,\varepsilon]^m$. Assume further that $\int x G(x) \, dx = O_{c} (\sigma_G)$. Then for all $\tau_2$ in the range $0 < \tau_2 \leq 1$,

$$
T_{F,G,N}^L (1,\cdots,1) \ll_{c,C,\varepsilon} c^{-1} \frac{1}{\sigma_G} + \frac{\tau_2^{-O(1)} A_L(\Omega_{c,C}(1),\tau_2)^{-1}}{N}, \quad (3.1)
$$

Proof. Following the proof of Lemma 3.2 verbatim, we arrive at the bound

$$
T_{F,G,N}^L (1,\cdots,1) \ll_{c,C,\varepsilon} \frac{1}{N^{d-m}} \sum_{n_{m+1},\cdots,n_d \in \mathbb{Z}} \tilde{G} \left( \sum_{j=m+1}^d v_j n_j \right), \quad (3.2)
$$

where $v_j$ denotes the $j$th column of the matrix $M^{-1}L$, and $\tilde{G} : \mathbb{R}^m \rightarrow [0,1]$ denotes the function

$$
\tilde{G}(x) = \sum_{a \in \mathbb{Z}^m} (G \circ M)(a + x).
$$

It remains to estimate the right-hand side of (3.2).

We may consider $\tilde{G}$ as a function on $\mathbb{R}^m/\mathbb{Z}^m$. With respect to the metric $\|x\|_{\mathbb{R}^m/\mathbb{Z}^m}$, $\tilde{G}$ is Lipschitz with Lipschitz constant $O_{c,C,\varepsilon}(1/\sigma_G)$. Also,

$$
\int_{x \in [0,1]^m} \tilde{G}(x) \, dx = \int_{x \in \mathbb{R}^m} (G \circ M)(x) \, dx = O_{c,C,\varepsilon}(\sigma_G).
$$
By \cite{14} Lemma A.9, which we recall in Appendix \cite{14} for any $X$ at least 2 we may write

$$
\tilde{G}(x) = \sum_{k \in \mathbb{Z}^m} b_X(k) e(k \cdot x) + O_{c,C,\varepsilon}\left(\frac{\log X}{\sigma_G X}\right), \quad (3.3)
$$

where $b_X(k) \in \mathbb{C}$ and satisfies $|b_X(k)| = O(1)$. Moreover $b_X(0) = \int_{x \in [0,1)^m} \tilde{G}(x) \, dx$.

Returning to (3.2), we see that for any $X$ at least 2 we may write

$$
T_{F,G,N}^L(1, \cdots, 1) \ll_{c,C,\varepsilon} \sigma_G + \frac{\log X}{\sigma_G X} + X^{0(1)} \max_{0 < ||k||_\infty \leq X} \left( \prod_{j=m+1}^d \min(1, N^{-1} ||k \cdot \nu_j||_{\mathbb{R}/\mathbb{Z}})^{-1}\right), \quad (3.4)
$$

where the final error term comes from summing over the arithmetic progressions $[-N, N] \cap \mathbb{Z}$.

It remains to relate the final error term of (3.4) to the approximation function $A_L$. Since $L$ is purely irrational,

$$
A_L(\tau_1, \tau_2) = \inf_{\varphi \in (\mathbb{R}^m)^*} \inf_{\tau_1 \leq ||\varphi||_\infty \leq \tau_2^{-1}} \text{dist}(L^* \varphi, (\mathbb{Z}^d)^T).
$$

Let $\tau_2$ be in the range $0 < \tau_2 \leq 1$. Then there is a suitable choice of parameter $X$, which satisfies $X \ll_{c,C} \tau_2^{-1}$, such that

$$
A_L(\Omega_{c,C}(1), \tau_2) \ll \inf_{k \in \mathbb{Z}^m} \text{dist}(k^T M^{-1} L, (\mathbb{Z}^d)^T)
$$

$$
\leq \min_{k \in \mathbb{Z}^m} \max(\{||k \cdot \nu_j||_{\mathbb{R}/\mathbb{Z}} : m + 1 \leq j \leq d\})
$$

$$
\leq \min_{k \in \mathbb{Z}^m} \max(\{||k \cdot \nu_j||_{\mathbb{R}/\mathbb{Z}} : m + 1 \leq j \leq d\}). \quad (3.5)
$$

Substituting this bound into (3.4), one derives

$$
T_{F,G,N}^L(1, \cdots, 1) \ll_{c,C,\varepsilon} \sigma_G + \frac{\tau_1^{1-o(1)}}{\sigma_G} + \frac{\tau_2^{-O(1)} A_L(\Omega_{c,C}(1), \tau_2)^{-1}}{N}
$$

as required. \hfill \Box

In Lemma 3.4 it was vitally important that $L$ was assumed to be purely irrational. This was manifested in the relations (3.5), when one could upper-bound $A_L(\Omega_{c,C}(1), \tau_2)$ by a minimum taken over all $k \in \mathbb{Z}^m$ of a certain size. Although one can attempt such an estimate when $L$ is not purely irrational, the integral $\int_{x \in \mathbb{R}^m} G(x) \, dx$ is no longer the relevant object. Rather, one must take some rational map for $L$, denoted $\Theta$, and consider $\int_{x \in \ker (\Theta + y)} G(x) \, dx$ for some shift $y$ (where $\ker (\Theta + y)$ receives the natural Lebesgue measure). It could be that $\int_{x \in \ker (\Theta + y)} G(x) \, dx$ is controlled but $\int_{x \in \ker (\Theta + y)} G(x) \, dx$ is not (consider the case where $G$ is the indicator of thin domain that has a flat side parallel to $\ker (\Theta)$, say). We opt to avoid these technicalities, creating instead a dimension reduction argument, that reduces all cases to the purely irrational case.

\footnote{This final fact is not given explicitly in the statement of \cite{14} Lemma A.9, although it is given in the proof. In any case, it may be immediately deduced from (3.3), by letting $X$ tend to infinity and integrating (3.3) over all $x \in \mathbb{R}^m / \mathbb{Z}^m$.}
4. Dimension reduction

In this section we reduce the main result (Theorem 2.10) to a different result, namely Theorem 4.6. This theorem will be simpler in one key respect: the replacement of sharp cut-offs by Lipschitz cut-offs.

We begin by dismissing the case of maximal rational dimension.

**Proposition 4.1.** Theorem 2.10 holds under the additional assumption that \( L \) has rational dimension \( m \).

**Proof.** We appeal to a quantitative version of Theorem 1.1.

**Theorem 4.2 (Generalised von Neumann Theorem for rational forms (quantitative version)).** Let \( N, m, d \) be natural numbers, satisfying \( d \geq m + 2 \), and let \( C_1, C_2 \) be positive constants. Let \( S = S(N) \) be an \( m \times d \) matrix with integer coefficients, satisfying \( \|S\|_\infty \leq C_1 \), and let \( \mathbf{r} \in \mathbb{Z}^m \) be some vector with \( \|\mathbf{r}\|_\infty \leq C_2 N \). Suppose \( S \) has rank \( m \), \( S \notin V_{\text{degen}}^*(m, d) \). Let \( K \subseteq [-N, N]^d \) be convex. Then there exists some natural number \( s \) at most \( d - 2 \) that satisfies the following. Let \( f_1, \ldots, f_d : [N] \rightarrow \mathbb{C} \) be arbitrary functions with \( \|f_j\|_{U^{s+1}[N]} \leq 1 \) for all \( j \), and assume that

\[
\min_j \|f_j\|_{U^{s+1}[N]} \leq \rho
\]

for some \( \rho \) in the range \( 0 < \rho \leq 1 \). Then

\[
\frac{1}{N^{d-m}} \sum_{\mathbf{n} \in \mathbb{Z}^d \cap K} \prod_{j=1}^d f_j(n_j) \ll C_1 C_2 \rho^{O(1)} + o_\rho(1).
\]

Furthermore, the \( o_\rho(1) \) term may be bounded above by \( \rho^{-O(1)} N^{-\Omega(1)} \).

In the proof, a certain familiarity with the methods and terminology of [15] will be assumed.

**Proof Sketch.** This theorem may be proved by following the proof of Theorem 1.8 of [15]. (In our language, the non-degeneracy condition in the statement of Theorem 1.8 of [15] is exactly \( S \notin V_{\text{degen}}^*(m, d) \)). One follows the same linear algebraic reductions as those used in section 4 of [15] to reduce Theorem 1.8 to Theorem 7.1 of the same paper (the Generalised von Neumann Theorem).

Theorem 7.1 may then be considered solely in the case of bounded functions \( f_j \), as in [24, Exercise 1.3.23], rather than in the more general case of functions bounded by a pseudorandom measure. It is clear from the proof that, in this more restricted setting, the \( \kappa(\rho) \) term that appears in the statement may be replaced by a polynomial dependence, and the \( o_\rho(1) \) term may be bounded above by \( \rho^{-O(1)} N^{-\Omega(1)} \).

This settles Theorem 4.2 where \( s \) is the Cauchy-Schwarz complexity of some system of forms \( (\psi_1, \ldots, \psi_d) \) that parametrises \( \ker S \). But \( s \) is at most \( d - 2 \), as any system of \( d \) forms with finite Cauchy-Schwarz complexity has Cauchy-Schwarz complexity at most \( d - 2 \). Therefore Theorem 4.2 is proved. \( \Box \)

Now let us use Theorem 4.2 to resolve Proposition 4.1. Indeed, let \( L \) be as in Theorem 2.10 and assume that \( L \) has rational dimension \( m \) and rational complexity at most \( C \). Let \( \Theta : \mathbb{R}^m \rightarrow \mathbb{R}^m \) be some linear isomorphism satisfying \( \Theta L(\mathbb{Z}^d) \subseteq \mathbb{Z}^m \) and \( \|\Theta\|_{\infty} \leq C \). Let \( M \) be a rank matrix of \( L \) (Proposition 3.1). Then the matrix \( M^{-1}L \) satisfies \( \|M^{-1}L\|_{\infty} \ll_{c,C} 1 \) and has rational dimension \( m \), since...
That \( S / N \), \( S \) columns of \( R \), and \( Ξ \) by Theorem 4.2. The main result (Theorem 2.10).

Although this is no great subtlety, we should emphasise that in the above definition we only consider perturbations to \( Ξ \), and not perturbations to \( L \) as well.

We are now ready to state Theorem 4.6, the theorem to which we will reduce the main result (Theorem 2.10).
Theorem 4.6 (Lipschitz case). Let $N, m, d, h$ be natural numbers, with $d \geq h \geq m + 2$, and let $c, C, \varepsilon$ be positive constants. Let $\Xi = \Xi(N) : \mathbb{R}^h \rightarrow \mathbb{R}^d$ be an injective linear map with integer coefficients, and assume that $\Xi(\mathbb{Z}^h) = \mathbb{Z}^d \cap \text{im} \, \Xi$. Let $L = L(N) : \mathbb{R}^h \rightarrow \mathbb{R}^m$ be a surjective linear map. Assume that $\|\Xi\|_\infty \leq C$, $\|L\|_\infty \leq C$, $\text{dist}(L, V_{\text{rank}}(m, d)) \geq c$ and $\text{dist}((\Xi, L), V_{\text{deg,2}}(m, d, h)) \geq c$. Then there exists a natural number $s$ such that for every $i \leq s$, independent of $\varepsilon$, such that the following holds. Let $\sigma_F, \sigma_G$ be any two parameters in the range $0 < \sigma_F, \sigma_G < 1/2$. Let $F : \mathbb{R}^h \rightarrow [0, 1]$ be a Lipschitz function supported on $[-N, N]^h$ with Lipschitz constant $O(1/\sigma_F N)$, and let $G : \mathbb{R}^m \rightarrow [0, 1]$ be a Lipschitz function supported on $[-\varepsilon, \varepsilon]^m$ with Lipschitz constant $O(1/\sigma_G)$. Let $\tilde{r}$ be a fixed vector in $\mathbb{Z}^d$, satisfying $\|\tilde{r}\|_\infty = O_{c, C, \varepsilon}(1)$. Suppose that $f_1, \ldots, f_d : [N] \rightarrow [-1, 1]$ are arbitrary bounded functions satisfying
\[
\min_j \|f_j\|_{U^{s+1}[N]} \leq \rho,
\]
for some $\rho$ in the range $0 < \rho \leq 1$. Then
\[
T_{F, G, N}^N(f_1, \ldots, f_d) \ll_{c, C, \varepsilon} \rho^{O(1)}(\sigma_F^{O(1)} + \sigma_G^{O(1)}) + \sigma_F^{O(1)} N^{-\Omega(1)}.
\]

Theorem 4.6

Although the above theorem contains more technical conditions than even Theorem 2.10 did, it does represent a significant reduction in complexity from the original problem. Note in particular that the approximation function $A_L$ does not feature in the estimate (4.3).

The remainder of this section will be devoted to proving the main theorem (Theorem 2.10), assuming the result of Theorem 4.6.

We begin with two lemmas: one concerning lattices, and the other concerning a quantitative decomposition of the dual space $(\mathbb{R}^d)^*$. Their proofs are entirely standard, but we state them prominently, as we will need to refer to them often in the dimension reduction argument of Lemma 4.10.

Lemma 4.7 (Parametrising the image lattice). Let $u, d$ be integers with $d \geq u + 1$. Let $S : \mathbb{R}^d \rightarrow \mathbb{R}^u$ be a surjective linear map with $S(\mathbb{Z}^d) \subseteq \mathbb{Z}^u$, and suppose that $\|S\|_\infty \leq C$. Then there exists a set $\{a_1, \ldots, a_u\} \subset \mathbb{Z}^u$ that is a basis for the lattice $S(\mathbb{Z}^d)$ and for which $\|a_i\|_\infty = O_C(1)$ for every $i$. Furthermore there exist $x_1, \ldots, x_u \in \mathbb{Z}^d$ such that, for every $i$, $S(x_i) = a_i$ and $\|x_i\|_\infty = O_C(1)$.

Proof. The lattice $S(\mathbb{Z}^d)$ is $u$ dimensional, as $S$ is surjective. If $\{e_j : j \leq d\}$ denotes the standard basis of $\mathbb{R}^d$ then integer combinations of elements from the set $\{S(e_j) : j \leq d\}$ span $S(\mathbb{Z}^d)$. Since $\|S\|_\infty \leq C$, these vectors also satisfy $\|S(e_j)\|_\infty = O_C(1)$. Therefore the $u$ successive minima of the lattice $S(\mathbb{Z}^d)$ are all $O_C(1)$, and so, by Mahler’s theorem (25 Theorem 3.34) the lattice $S(\mathbb{Z}^d)$ has a basis $\{a_1, \ldots, a_u\}$ of the required form.

Note that $S$ has integer coefficients. The construction of suitable $x_1, \ldots, x_u$ may be achieved by applying any of the standard algorithms. For example, using Gaussian elimination one may find a basis for ker $S$ that, by inspection of the algorithm, consists of vectors with rational coordinates of naive height $O_C(1)$. By clearing denominators, one gets vectors $v_1, \ldots, v_{d-u} \in \mathbb{Z}^d$ whose integer span is a full-dimensional sublattice of the $d-u$ dimensional lattice $\mathbb{Z}^d \cap \text{ker} \, S$, and that satisfy $\|v_i\|_\infty = O_C(1)$ for all $i$. Now given some $a_i$, by its construction there must be some $x_i \in \mathbb{Z}^d$ that satisfies $S(x_i) = a_i$. Write $x_i = x_{i|_{\text{ker} \, S}} + x_{i|_{\text{ker} \, S}^\perp}$ as the sum of the obvious projections. By adding a suitable integer combination of the vectors $v_1, \ldots, v_{d-u}$ to $x_i$ one may find such an $x_i$ that satisfies $\|x_i|_{\text{ker} \, S}\|_\infty = O_C(1)$. Furthermore, $\text{dist}(S, V_{\text{rank}}(m, d)) = O_C(1)$, since
$S$ has integer coordinates, and so (by Lemma D.1) $\|x_i(\ker S)\|_\infty = O_C(1)$. Hence $\|x_i\|_\infty = O_C(1)$, as desired. 

Having established that such a lattice basis $\{a_1, \cdots, a_u\}$ exists, we can now use it to quantitatively decompose $(\mathbb{R}^d)^*$. 

**Lemma 4.8** (Dual space decomposition). Let $u, d$, be integers with $d \geq u + 1$, and let $C, \eta$ be constants. Let $S : \mathbb{R}^d \rightarrow \mathbb{R}^u$ be a surjective linear map with $S(\mathbb{Z}^d) \subseteq \mathbb{Z}^u$, and suppose that $|S|^{\infty} \leq C$. Let $\{a_1, \cdots, a_u\}$ be a basis for the lattice $S(\mathbb{Z}^d)$ that satisfies $\|a_i\|_\infty = O_C(1)$ for every $i$. Let $x_1, \cdots, x_u \in \mathbb{Z}^d$ be vectors such that, for every $i$, $S(x_i) = a_i$ and $\|x_i\|_\infty = O_C(1)$. Suppose that $\Xi : \mathbb{R}^{d-u} \rightarrow \mathbb{R}^d$ is an injective linear map such that $\text{im } \Xi = \ker S$ and such that $\Xi(\mathbb{Z}^{d-u}) = \mathbb{Z}^d \cap \text{im } \Xi$. Suppose further that $\|\Xi\|_\infty \leq C$.

Let $w_1, \cdots, w_{d-u}$ denote the standard basis vectors in $\mathbb{R}^{d-u}$. Then

1. the set $B := \{x_i : 1 \leq i \leq u\} \cup \{\Xi(w_j) : 1 \leq j \leq d-u\}$ is a basis for $\mathbb{R}^d$, and a lattice basis for $\mathbb{Z}^d$;
2. writing $B^* := \{x_i^* : 1 \leq i \leq u\} \cup \{\Xi(w_j)^* : 1 \leq j \leq d-u\}$ for the dual basis, both the change of basis matrix between the standard dual basis and $B^*$ and the inverse of this matrix have integer coordinates. The coefficients of both of these matrices are bounded in absolute value by $O_C(1)$.

Write $V := \text{span}(x_i^* : 1 \leq i \leq u)$ and $W := \text{span}(\Xi(w_j)^* : j \leq d-u)$. Then

3. $V = S^*((\mathbb{R}^u)^*)$;
4. Suppose that $\varphi \in (\mathbb{R}^d)^*$ satisfies $\|\Xi^*(\varphi)\|_\infty \leq \eta$. Then, writing $\varphi = \varphi_V + \varphi_W$ with $\varphi_V \in V$ and $\varphi_W \in W$, we have $\|\varphi_W\|_\infty = O_C(\eta)$.

**Proof.** For part (1), the fact that $B$ is a basis for $\mathbb{R}^d$ is just a manifestation of the familiar principle $\mathbb{R}^d \cong \ker S \oplus \text{im } S$. To show that $B$ is a lattice basis for $\mathbb{Z}^d$, let $n \in \mathbb{Z}^d$ and write

$$n = \sum_{i=1}^{u} \lambda_i x_i + \sum_{j=1}^{d-u} \mu_j \Xi(w_j)$$

for some $\lambda_i, \mu_j \in \mathbb{R}$. Applying $S$, we see $S(n) = \sum_{i=1}^{u} \lambda_i a_i$, and hence $\lambda_i \in \mathbb{Z}$ for all $i$, as $\{a_1, \cdots, a_u\}$ is a basis for the lattice $S(\mathbb{Z}^d)$. But this implies $\sum_{j=1}^{d-u} \mu_j \Xi(w_j) \in \mathbb{Z}^d \cap \text{im } \Xi$. Therefore, as $\Xi(\mathbb{Z}^{d-u}) = \mathbb{Z}^d \cap \ker S$, $\mu_j \in \mathbb{Z}$ for all $j$.

Part (2) follows immediately from part (1). Part (3) is immediate from the definitions.

For part (4), let $j$ be at most $d-u$. Then the assumption $\|\Xi^*(\varphi)\|_\infty \leq \eta$ means that $|\Xi^*(\varphi)(w_j)| \leq \eta$. Hence $|\varphi(\Xi(w_j))| \leq \eta$. But, writing $\varphi_W = \sum_{j=1}^{d-u} \mu_j \Xi(w_j)^*$, this implies that $|\mu_j| \leq \eta$. Since the coefficients of the change of basis matrix between $B^*$ and the standard dual basis are bounded in absolute value by $O_C(1)$, this implies that $\|\varphi_W\|_\infty \leq O_C(\eta)$.

We now begin the attack on Theorem 2.10 in earnest. Assume the hypotheses of Theorem 2.10. As a reminder, we have natural numbers $m, d$ satisfying $d \geq m + 2$, and positive reals $\varepsilon, c, C$. For a natural number $N$, we have $L = L(N) : \mathbb{R}^d \rightarrow \mathbb{R}^m$ being a surjective linear map with approximation function $A_L$, with $\text{dist}(L, V_{\text{deg}}^*(m, d)) > c$, and with rational complexity at most $C$. We have $F : \mathbb{R}^d \rightarrow [0, 1]$ being the indicator function of $[1, N]^d$ and $G : \mathbb{R}^m \rightarrow [0, 1]$ being the indicator function of a convex domain contained in $[-\varepsilon, \varepsilon]^m$, and functions $f_1, \cdots, f_d : [N] \rightarrow [-1, 1]$ that satisfy $\min_j \|f_j\|_{U^{+1}[N]} \leq \rho$ for some $\rho$ in the range $0 < \rho \leq 1$. 

Lemma 4.9 (Replacing variable cut-off). Assume the hypotheses of Theorem 2.10 (in particular let $F$ be the indicator function $1_{[1,N]^d}$, and let $\sigma_F$ be any parameter in the range $0 < \sigma_F < 1/2$. Then there exists a Lipschitz function $F_{1,\sigma_F} : \mathbb{R}^d \to [0,1]$, supported on $[-2N,2N]^d$ and with Lipschitz constant $O(1/\sigma_F N)$, such that

$$ |T_{F_{1,\sigma_F},G,N}(f_1, \cdots, f_d)| \ll |T_{L,F_1,\sigma_F,G,N}(f_1, \cdots, f_d)| + O_c(C(\sigma_F)). $$

Proof. By Lemma 3.2, for any parameter $\sigma_F$ in the range $0 < \sigma_F < 1/2$ we may write

$$ 1_{[1,N]^d} = F_{1,\sigma_F} + O(F_{2,\sigma_F}), $$

where $F_{1,\sigma_F}, F_{2,\sigma_F}$ are Lipschitz functions supported on $[-2N,2N]^d$, with Lipschitz constants $O(1/\sigma_F N)$, and with $\int_{\mathbb{R}^d} F_{2,\sigma_F}(x) \, dx = O(\sigma_F N^d)$. Moreover, $F_{2,\sigma_F}$ is supported on

$$ \{ x \in \mathbb{R}^d : \text{dist}(x, \partial([1,N]^d)) = O(\sigma_F N) \}. $$

Therefore

$$ T_{F,\sigma_F,G,N}(f_1, \cdots, f_d) \ll |T_{F_{1,\sigma_F},G,N}(f_1, \cdots, f_d)| + |T_{F_{2,\sigma_F},G,N}(1, \cdots, 1)|. $$

Recall, from the remark after Definition 2.3, that $V_{\text{deg}}(m,d)$ contains $V^\text{glob}(m,d)$. Therefore, since we assume that $\text{dist}(L,V_{\text{deg}}(m,d)) \geq c$, we have

$$ \text{dist}(L,V^\text{glob}_{\text{rank}}(m,d)) \geq c. $$

Hence, by Lemma 3.3,

$$ |T_{F_{2,\sigma_F},G,N}(f_1, \cdots, f_d)| = O_{c,C}(\sigma_F). $$

This gives the lemma. \qed

Next comes the critical lemma, in which we successfully replace the map $L$ by a purely irrational map $L'$. For the definition of the approximation function $A_L$, one may consult Definition 2.8.

Lemma 4.10 (Generating a purely irrational map). Let $\sigma_F$ be a parameter in the range $0 < \sigma_F < 1/2$. Assume the hypotheses of Theorem 2.10, with the exception that $F : \mathbb{R}^d \to [0,1]$ now denotes a Lipschitz function supported on $[-2N,2N]^d$ and with Lipschitz constant $O(1/\sigma_F N)$. Let $u$ be the rational dimension of $L$, and assume that $u \leq m - 1$. Then there exists 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 $\mathcal{R} \subset \mathbb{Z}^d$, and, for each $\mathbf{r} \in \mathcal{R}$, functions $F'_{\mathbf{r}} : \mathbb{R}^{d-u} \to [0,1]$ and $G'_{\mathbf{r}} : \mathbb{R}^{m-u} \to [0,1]$, that together satisfy the following properties:

1. $\Xi$ has integer coefficients, $\|\Xi\|_\infty = O_c(C(1))$, and $\Xi(\mathbb{Z}^{d-u}) = \mathbb{Z}^d \cap \text{im} \Xi$;

2. $|\mathcal{R}| = O_{c,C}(1)$, and $\|\mathcal{R}\|_\infty = O_{c,C}(1)$ for all $\mathbf{r} \in \mathcal{R}$;

3. $F'_{\mathbf{r}}$ is supported on $[-O_c(C(N)),O_c(C(N))]^{d-u}$, with Lipschitz constant $O_c(C(1/\sigma_F N)$, and $G'_{\mathbf{r}}$ is the indicator function of a convex domain contained in $[-O_{c,C}(1),O_{c,C}(1)]^{m-u}$;

4. $T_{L,F,G,N}^I(f_1, \cdots, f_d) = \sum_{\mathbf{r} \in \mathcal{R}} T_{F'_r,G'_r,N}(f_1, \cdots, f_d)$;

5. $L'$ is purely irrational;
Then, \( \operatorname{dist}((\Xi, L'), V^*_{\text{deg}, 2}(m - u, d - u)) = \Omega_{c,C}(1); \)
(7) \( \operatorname{dist}((\Xi, L'), V^*_{\text{deg}, 2}(m - u, d - u)) = \Omega_{c,C}(1); \)
(8) for all \( \tau_1, \tau_2 \in (0, 1], A_L(\tau_1, \tau_2) \gg_{c,C} A_L(\Omega_{c,C}(\tau_1), \Omega_{c,C}(\tau_2)); \)
(9) for all \( \tau_1, \tau_2 \in (0, 1], A_L(\tau_1, \tau_2) \ll_{c,C} A_L(\Omega_{c,C}(\tau_1), \Omega_{c,C}(\tau_2)). \)

The fundamental aspect of this lemma is part (4), of course, as this directly concerns how we control the number of solutions to the diophantine inequality itself when passing from \( L \) to \( L' \). However, we do need to establish parts (1) - (8), in order to be able to ensure that the hypotheses of Lemma 3.4 and Theorem 4.6 are satisfied. Part (9) is included for completeness, and to assist the calculations in section 4.

Before giving the full details of the proof, we sketch the idea. Let \( \Theta : \mathbb{R}^m \rightarrow \mathbb{R}^u \) be a rational map for \( L \). The space \( \ker(\Theta L) \) has dimension \( d - u \), and so we may parametrise it by some injective map \( \Xi : \mathbb{R}^{d-u} \rightarrow \ker(\Theta L) \). Without too much difficulty, \( \Xi \) can be chosen to satisfy \( \Xi(\mathbb{Z}^{d-u}) = \mathbb{Z}^d \cap \im \Xi \). Then

\[
L\Xi : \mathbb{R}^{d-u} \rightarrow \ker \Theta,
\]

is a map from a \( d - u \) dimensional space to an \( m - u \) dimensional space, and it turns out that \( L\Xi \) is purely irrational, and \( L' = L\Xi \) may be used in Lemma 4.10.

Of course this isn’t quite possible, as we only defined the notion of purely irrational maps between vector spaces of the form \( \mathbb{R}^n \). But it is true after choosing a judicious isomorphism from \( \ker \Theta \) to \( \mathbb{R}^{m-u} \) (though this does complicate the notation).

Let us complete the details.

**Proof.** First we note that the lemma is obvious when \( u = 0 \), since one may take \( \Xi : \mathbb{R}^d \rightarrow \mathbb{R}^d \) to be the identity map, \( \tilde{r} \) to be \( \mathbf{0} \), and \( L' \) to be \( L \). So assume that \( u > 1 \).

We proceed with a general reduction, familiar from our proof of Proposition 4.1 in which we may assume that the first \( m \) columns of \( L \) form the identity matrix.

Indeed, let \( \Theta : \mathbb{R}^m \rightarrow \mathbb{R}^u \) be a rational map for \( L \) with \( \|\Theta\|_\infty \leq C \). Now let \( \tilde{L} := M^{-1}L \), where \( M \) is a rank matrix of \( L \) (Proposition 3.1), which, without loss of generality, consists of the first \( m \) columns of \( L \). Let \( \tilde{\Theta} := \Theta M \) and let \( \tilde{G} := G \circ M \). Then

\[
T_{F,G,N}^L(f_1, \ldots, f_d) = T_{F,G,N}^\tilde{L}(f_1, \ldots, f_d),
\]

and, considering \( \tilde{\Theta}, \tilde{L} \) has rational complexity \( O_{c,C}(1) \). Furthermore, \( \tilde{G} \) is the indicator function of a convex domain contained in \( [-O_{c,C}(\varepsilon), O_{c,C}(\varepsilon)]^m \). We also have \( \operatorname{dist}(\tilde{L}, V^*_{\text{deg}}(m, d)) = \Omega_{c,C}(1) \). Finally, for all \( \tau_1, \tau_2 \in (0, 1], \) we have that \( A_L(\tau_1, \tau_2) \ll_{c,C} A_L(\Omega_{c,C}(\tau_1), \Omega_{c,C}(\tau_2)). \)

Therefore, by replacing \( L \) with \( \tilde{L} \) and \( G \) with \( \tilde{G} \), we may assume throughout the proof of Lemma 4.10 that the first \( m \) columns of \( L \) form the identity matrix. This is at the cost of replacing \( \varepsilon \) by \( O_{c,C}(\varepsilon) \), \( C \) by \( O_{c,C}(1) \), and \( \varepsilon \) by \( \Omega_{c,C}(1) \).

Now let \( \Theta : \mathbb{R}^m \rightarrow \mathbb{R}^u \) be a rational map for \( L \) with \( \|\Theta\|_\infty = O_{c,C}(1) \). Since the first \( m \) columns of \( L \) form the identity matrix, \( \Theta \) must have integer coefficients.

**Part (1):** By rank-nullity \( \ker(\Theta L) \) is a \( d - u \) dimensional subspace of \( \mathbb{R}^d \). The matrix of \( \Theta L \) has integer coefficients and \( \|\Theta L\|_\infty = O_{c,C}(1) \). 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 \( \|\mathbf{v}^{(i)}\|_{\infty} = O_{c,C}(1) \) for every \( i \). Define \( \Xi : \mathbb{R}^{d-u} \rightarrow \mathbb{R}^d \) by

\[
\Xi(w) := \sum_{i=1}^{d-u} w_i \mathbf{v}^{(i)}.
\]

Then \( \Xi \) satisfies property (1) of the lemma. Note that image of the map \( L\Xi : \mathbb{R}^{d-u} \rightarrow \mathbb{R}^m \) is exactly \( \ker \Theta \).

**Part (2):** Since \( \|\Theta\|_{\infty} = O_{c,C}(1) \), if \( y \in \mathbb{R}^m \) and \( \Theta(y) = r \) then \( \|y\|_{\infty} \gg c,C \|r\|_{\infty} \). Recall that the support of \( G \) is contained within \([-O_{c,C,c}(1),O_{c,C,c}(1)]^m \), and that \( \Theta L(\mathbb{Z}^d) \subseteq \mathbb{Z}^u \). It follows that there are at most \( O_{c,C,c}(1) \) possible vectors \( r \in \mathbb{Z}^u \) for which there exists a vector \( n \in \mathbb{Z}^d \) for which both \( G(Ln) \neq 0 \) and \( \Theta Ln = r \). Let \( R \) denote the set of all such vectors \( r \).

For each \( r \in R \), there exists a vector \( \tilde{r} \in \mathbb{Z}^d \) such that \( \Theta L \tilde{r} = r \) and \( \|\tilde{r}\|_{\infty} = O_{c,C,c}(1) \). Let \( \tilde{R} \) denote the set of these \( \tilde{r} \). Then \( \tilde{R} \) satisfies part (2).

Before proceeding to prove part (3) of the lemma, we pause to apply Lemmas 4.7 and 4.8. Indeed, applying these lemmas to the map \( S := \Theta L \), there exists a set \( \{\mathbf{a}_1, \ldots, \mathbf{a}_u\} \subset \mathbb{Z}^u \) that is a basis for the lattice \( \Theta L(\mathbb{Z}^d) \) and for which \( \|a_i\|_{\infty} = O_{c,C}(1) \) for each \( i \). Also, 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 \( \|\mathbf{x}_i\|_{\infty} = O_{c,C}(1) \). By Lemma 4.8, \( \mathcal{B} := \{\mathbf{x}_i : i \leq u\} \cup \{\Xi(w_j) : j \leq d - u\} \) (4.4) is a basis for \( \mathbb{R}^d \) and a lattice basis for \( \mathbb{Z}^d \), where \( \mathbf{w}_1, \ldots, \mathbf{w}_{d-u} \) denotes the standard basis of \( \mathbb{R}^{d-u} \).

**Part (3):** By the definition of \( \tilde{R} \), and the fact that \( \Xi(\mathbb{Z}^{d-u}) = \mathbb{Z}^d \cap \ker(\Theta L) \), we have

\[
T_{E,G,N}^d(f_1, \ldots, f_d) = \sum_{\tilde{r} \in \tilde{R}} \frac{1}{Nd-m} \sum_{n \in \mathbb{Z}^{d-u}} \left( \prod_{j=1}^d f_j(\xi_j(n) + \tilde{r}_j) \right) F(\Xi(n) + \tilde{r}) G(L\Xi(n) + L\tilde{r}),
\]

where \( \tilde{r}_j \) denotes the \( j \)th coordinates of \( \tilde{r} \). Now by an easy linear algebraic argument (recorded in Lemma 4.4),

\[
\mathbb{R}^m = \text{span}(L\mathbf{x}_i : i \leq u) \oplus \ker \Theta \tag{4.6}
\]

as an algebraic direct sum, and there exists an invertible linear map \( P : \mathbb{R}^m \rightarrow \mathbb{R}^m \) such that

\[
P((\text{span}(L\mathbf{x}_i : i \leq u))) = \mathbb{R}^u \times \{0\}^{m-u}, \tag{4.7}
\]

\[
P(\ker \Theta) = \{0\}^u \times \mathbb{R}^{m-u}, \tag{4.8}
\]

and both \( \|P\|_{\infty} = O_{c,C}(1) \) and \( \|P^{-1}\|_{\infty} = O_{c,C}(1) \).

We have

\[
G(L\Xi(n) + L\tilde{r}) = (G \circ P^{-1})(PL\Xi(n) + PL\tilde{r}),
\]

and we note that \( PL\Xi(n) \in \{0\}^u \times \mathbb{R}^{m-u} \) for every \( n \in \mathbb{Z}^{d-u} \). Define \( G_{\tilde{r}} : \mathbb{R}^{m-u} \rightarrow [0,1] \) by

\[
G_{\tilde{r}}(x) := (G \circ P^{-1})(x_0 + PL\tilde{r}),
\]
where \( x_0 \) is the extension of \( x \) by 0 in the first \( u \) coordinates. Then the function \( G_\tilde{r} \) is the indicator function of a convex set contained in \([-O_{c,C,\varepsilon}(1), O_{c,C,\varepsilon}(1)]^{m-u}\).

Define
\[
F_\tilde{r} (n) := F(\Xi(n) + \tilde{r}).
\]
Then \( F_\tilde{r} \) has Lipschitz constant \( O_{c,C}(1/\alpha_F,N) \) and \( F_\tilde{r} \) is supported on \([-O_{c,C,\varepsilon}(N), O_{c,C,\varepsilon}(N)]^{d-u} \). (For a full proof of this fact, apply Lemma D.3 to the map \( \Xi \)). So \( F_\tilde{r} \) and \( G_\tilde{r} \) satisfy part (3).

**Part (4):** Writing \( \pi_{m-u} : \mathbb{R}^m \rightarrow \mathbb{R}^{m-u} \) for the projection onto the final \( m-u \) coordinates, expression (4.5) is equal to
\[
\sum_{\tilde{r} \in \tilde{R}} \frac{1}{N^{d-m}} \sum_{n \in \mathbb{Z}^{d-u}} \left( \prod_{j=1}^{d} f_j(\xi_j(n) + \tilde{r}_j) \right) F_\tilde{r}(n) G_\tilde{r}(\pi_{m-u} PL \Xi(n)).
\]
(4.9)

Let
\[
L' := \pi_{m-u} PL \Xi.
\]
Then \( L' : \mathbb{R}^{d-u} \rightarrow \mathbb{R}^{m-u} \) is surjective, and
\[
T_{F,G,N}^L(f_1, \ldots, f_d) = \sum_{\tilde{r} \in \tilde{R}} T_{F,G,N}^{L', \Xi, \tilde{r}}(f_1, \ldots, f_d).
\]
This resolves part (4).

**Part (5):** We wish to show that \( L' \) is purely irrational. Suppose for contradiction that there exists some surjective linear map \( \varphi : \mathbb{R}^{m-u} \rightarrow \mathbb{R} \) with \( \varphi L'(\mathbb{Z}^{d-u}) \subseteq \mathbb{Z} \), i.e. with \( \varphi \pi_{m-u} PL \Xi(\mathbb{Z}^{d-u}) \subseteq \mathbb{Z} \). Then define the map \( \Theta' : \mathbb{R}^m \rightarrow \mathbb{R}^{u+1} \) by
\[
\Theta'(x) := (\Theta(x), \varphi \pi_{m-u} P(x)).
\]
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 (1.1)). This contradicts the assumption that \( L \) has rational dimension \( u \). So \( L' \) is purely irrational.

**Part (6):** The bound \( \|L'\|_{\infty} = O_{c,C}(1) \) follows immediately from the bounds on the coefficients of \( \Xi, L, P, \) and \( \pi_{m-u} \) separately.

We wish to prove that \( \text{dist}(L', V_{\text{rank}}(m-u, d-u)) \gg_{c,C} 1 \), i.e. that \( \text{dist}(\pi_{m-u} PL \Xi, V_{\text{rank}}(m-u, d-u)) \gg_{c,C} 1 \). Suppose for contradiction that, for a small parameter \( \eta \), there exists a linear map \( Q : \mathbb{R}^{d-u} \rightarrow \mathbb{R}^{m-u} \) such that \( \|Q\|_{\infty} < \eta \) and \( \pi_{m-u} PL \Xi + Q \) has rank less than \( m-u \). Recall that \( PL \Xi(\mathbb{R}^{d-u}) = \{0\}^u \times \mathbb{R}^{m-u} \). So, extending \( Q \) by zeros to a map \( \tilde{Q} : \mathbb{R}^{d-u} \rightarrow \{0\}^u \times \mathbb{R}^{m-u} \), and applying \( P^{-1} \), there is a map \( Q' : \mathbb{R}^{d-u} \rightarrow \mathbb{R}^m \) such that \( \|Q'\|_{\infty} = O_{c,C}(\eta) \) and \( L \Xi + Q' \) has rank less than \( m-u \).

We may factorise \( Q' = H \Xi \) for some \( m \)-by-\( d \) matrix \( H \). Indeed let
\[
B := \{ x_i : i \leq u \} \cup \{ \Xi(w_j) : j \leq d-u \}
\]
be the basis of \( \mathbb{R}^d \) from (1.4), i.e. the basis formed by applying Lemma 4.8 to the map \( S := \Theta L \). Define the linear map \( H \) by \( H(\Xi(w_j)) := Q'(w_j) \) for each \( j \) and \( H(x_i) := 0 \) for each \( i \). Since the change of basis matrix between \( B \) and the standard basis of \( \mathbb{R}^d \) has integer coefficients with absolute values at most \( O_{c,C}(1) \), it follows that the matrix representing \( H \) with respect to the standard bases satisfies \( \|H\|_{\infty} = O_{c,C}(\eta) \).

So we know that \( (L + H) \Xi \) has rank less than \( m-u \). But \( \Xi : \mathbb{R}^{d-u} \rightarrow \mathbb{R}^d \) is injective, so this implies that the rank of \( L + H \) is less than \( m \). Hence \( \text{dist}(L, V_{\text{rank}}(m,d)) = \)
that where dual basis. Expanding out the definition of $L\phi_i,j$ in other words, we suppose there exist two indices $i, j \leq d$, and a real number $\lambda$, such that $e_i^* - \lambda e_j^*$ is non-zero and

$$(\Xi + Q)^* (e_i^* - \lambda e_j^*) \in (L')^* (\mathbb{R}^m - u)^*,$$

where $\{e_1, \ldots, e_d\}$ denotes the standard basis of $\mathbb{R}^d$ and $\{e_1^*, \ldots, e_d^*\}$ denotes the dual basis. Expanding out the definition of $L'$, this means that there exists some $\varphi \in (\mathbb{R}^m - u)^*$ such that

$$\Xi^* (e_i^* - \lambda e_j^* - L^* (P^* \pi_{m-u}^* (\varphi))) = -Q^* (e_i^* - \lambda e_j^*).$$

Because $\|Q\|_\infty \leq \eta$, this means that

$$\|\Xi^* (e_i^* - \lambda e_j^* - L^* (P^* \pi_{m-u}^* (\varphi)))\|_\infty = O(\eta). \quad (4.11)$$

Let

$$B^* := \{x_i^* : i \leq u\} \cup \{\Xi (w_j)^* : j \leq d - u\} \quad (4.12)$$

denote the basis of $(\mathbb{R}^d)^*$ that is dual to the basis $B$ from (4.4). It follows from part (4) of Lemma 4.8 and (4.11) that

$$e_i^* - \lambda e_j^* - L^* (P^* \pi_{m-u}^* (\varphi)) = \omega_V + \omega_W,$$

where $\omega_V \in L^* \Theta^* ((\mathbb{R}^u)^*)$, $\omega_W \in \text{span}(\Xi (w_j)^* : j \leq d - u)$, and $\|\omega_W\|_\infty = O_{c,C}(\eta)$. So therefore

$$e_i^* - \lambda e_j^* = L^*(\alpha) + \omega_W,$$

for some $\alpha \in (\mathbb{R}^m)^*$.

This is enough to derive a contradiction. Indeed, without loss of generality one may assume that $\|e_i^* - \lambda e_j^*\|_\infty \geq 1$ (this is obvious if $i \neq j$, and if $i = j$ we may just pick $\lambda = 0$ at the outset). Therefore $\|e_i^* - \lambda e_j^* - \omega_W\|_\infty \geq 1/2$, provided $\eta$ is small enough. Since $\|L^*\|_\infty = O_{c,C}(1)$, we conclude that $\|\alpha\|_\infty = O_{c,C}(1)$.

This means that there exists a linear map $E : \mathbb{R}^d \rightarrow \mathbb{R}^m$ with $\|E\|_\infty = O_{c,C}(\eta)$ for which $E^*(\alpha) = \omega_W$. Then

$$e_i^* - \lambda e_j^* \in (L + E)^* ((\mathbb{R}^m)^*),$$

and hence $\text{dist}(L, V_{\text{deg},d}(m, d)) = O_{c,C}(\eta)$. This is a contradiction to the hypotheses of Theorem 2.10 provided $\eta$ is small enough, and hence $\text{dist}(\Xi, L'), V_{\text{deg},d}(m - u, d, d - u)) = O_{c,C}(1)$.

Part (8): Let $\tau_1, \tau_2 \in (0, 1]$. We desire to prove the relationship

$$A_{L'}(\tau_1, \tau_2) \gg_{c,C} A_L(\Omega_{c,C}(\tau_1), \Omega_{c,C}(\tau_2)), \quad (4.13)$$

where $L'$ is as in (4.10).

We have already proved that $L'$ is purely irrational (that was part (5) of the lemma). So, if $A_{L'}(\tau_1, \tau_2) < \eta$, for some $\eta$, there exists some $\varphi \in (\mathbb{R}^{m-u})^*$ for which $\tau_1 \leq \|\varphi\|_\infty \leq \tau_2^{-1}$ and for which

$$\text{dist}(\pi_{m-u}^* P L \Xi)^* (\varphi), (\mathbb{Z}^{d-u})^T ) < \eta,$$

where, one recalls, we use $(\mathbb{Z}^{d-u})^T$ to denote the set of those functions in $(\mathbb{R}^{d-u})^*$ that have integer coordinates with respect to the standard dual basis.
We claim that
\[ \text{dist}(L^*(P^*\pi^*_{m-u}(\varphi)), (\mathbb{Z}^d)^T) \ll c, C\eta; \tag{4.14} \]
\[ ||P^*\pi^*_{m-u}(\varphi)||_{\infty} \ll c, C \tau_2^{-1}; \tag{4.15} \]
\[ \text{dist}(P^*\pi^*_{m-u}(\varphi), \Theta^*((\mathbb{R}^u)^*)) \gg_{c, C} \tau_1; \tag{4.16} \]
from which \((4.13)\) immediately follows.

Let us prove \((4.14)\). Indeed, we already know that
\[ \text{dist}(\Xi^*L^*P^*\pi^*_{m-u}(\varphi), (\mathbb{Z}^d-u)^T) < \eta, \]
i.e. that
\[ ||\Xi^*L^*P^*\pi^*_{m-u}(\varphi) - \alpha||_{\infty} < \eta, \tag{4.17} \]
for some \(\alpha \in (\mathbb{Z}^d-u)^T\). Let us write \(\alpha = \sum_{j=1}^{d-u} \lambda_j w_j^*\) for some \(\lambda_j \in \mathbb{Z}\), where \(w_1, \ldots, w_{d-u}\) denotes the standard basis for \(\mathbb{R}^{d-u}\) and \(w_1^*, \ldots, w_{d-u}^*\) denotes the dual basis. Let \(B^*\) be as in \((4.12)\). Then \(w_j^* = \Xi^*(\Xi(w_j)^*)\), and so
\[ \alpha = \Xi^*\left(\sum_{j=1}^{d-u} \lambda_j \Xi(w_j)^*\right). \]

So from \((4.17)\) and the final part of Lemma \(4.8\)
\[ L^*P^*\pi^*_{m-u}(\varphi) - \sum_{j=1}^{d-u} \lambda_j \Xi(w_j)^* = \omega_V + \omega_W, \tag{4.18} \]
where \(\omega_V \in \text{span}(x^*_i: i \leq u), \omega_W \in \text{span}(\Xi(w_j)^*: j \leq d - u)\), and \(||\omega_W||_{\infty} = O_{c, C}(\eta)\).

But \(L^*P^*\pi^*_{m-u}(\varphi) \in \text{span}(\Xi(w_j)^*: j \leq d - u)\) too. Indeed, for every \(i\) at most \(d - u\),
\[ L^*P^*\pi^*_{m-u}(\varphi)(x_i) = \varphi(\pi_{m-u}(P\xi_i)) = \varphi(0) = 0, \]
by the properties of \(P\) (see \((4.17)\)). Therefore \(\omega_V = 0\), and so
\[ ||L^*P^*\pi^*_{m-u}(\varphi) - \sum_{j=1}^{d-u} \lambda_j \Xi(w_j)^*||_{\infty} = O_{c, C}(\eta). \]

Since \(\sum_{j=1}^{d-u} \lambda_j \Xi(w_j)^* \in (\mathbb{Z}^d)^T\), this implies \((4.14)\) as claimed.

The bound \((4.15)\) is immediate from the bounds on the coefficients of \(P^*\) and \(\pi^*_{m-u}\), so it remains to prove \((4.16)\). Suppose for contradiction that, for some small parameter \(\delta\),
\[ P^*\pi^*_{m-u}(\varphi) = \alpha_1 + \alpha_2, \]
where \(\alpha_1 \in \Theta^*((\mathbb{R}^u)^*)\) and \(||\alpha_2||_{\infty} \leq \delta \tau_1\), which means that there is some standard basis vector \(f_k \in \mathbb{R}^m\) for which \(|\varphi(f_k)| \geq \tau_1\). Let \(b_{k+u}\) be the standard basis vector of \(\mathbb{R}^{m-u}\) for which \(\pi_{m-u}(b_{k+u}) = f_k\). Recall the properties of \(P\) (given in \((4.7)\) and \((4.8)\)), in particular recall that \(P: \ker \Theta \rightarrow \{0\}^u \times \mathbb{R}^{m-u}\) is an isomorphism. Then
\[ |P^*\pi^*_{m-u}(\varphi)(P^{-1}(b_{k+u}))| = |\pi^*_{m-u}(\varphi)(b_{k+u})| = |\varphi(f_k)| \geq \tau_1. \]

Note that \(\Theta^*((\mathbb{R}^u)^*) = (\ker \Theta)^o\), and so
\[ |P^*\pi^*_{m-u}(\varphi)(P^{-1}(b_{k+u}))| = ||(\alpha_1 + \alpha_2)(P^{-1}(b_{k+u}))|| = |\alpha_2(P^{-1}(b_{k+u}))| \ll_{c, C} \delta \tau_1, \]
as \(P^{-1}(b_{k+u}) \in \ker \Theta\) and satisfies \(||P^{-1}(b_{k+u})||_{\infty} = O_{c, C}(1)\). This is a contradiction if \(\delta\) is small enough, and so \((4.16)\) holds. This resolves part (8).
Part (9): Let \( \tau_1, \tau_2 \in (0, 1] \). We desire to prove the relationship

\[
A_{L'}(\tau_1, \tau_2) \ll c, c \ A_L(\Omega_{c, C}(\tau_1), \Omega_{c, C}(\tau_2)),
\]

where \( L' \) is as in (4.10). This inequality is the reverse inequality of part (8), and in fact it will not be required in the proof of any of our main theorems. However, it will be required in order to analyse \( A_L(\tau_1, \tau_2) \) when \( L \) has algebraic coefficients (in Appendix E), so we choose to state and prove it here, close to our argument for part (8).

Suppose that \( A_L(\tau_1, \tau_2) < \eta \), for some parameter \( \eta \). Then there exists some \( \varphi \in (\mathbb{R}^m)^* \) such that \( \text{dist}(\varphi, \Theta^*((\mathbb{R}^n)^*)) \geq \tau_1 \), \( \|\varphi\|_{\infty} \leq \tau_2^{-1} \), and \( \text{dist}(L^*\varphi, (\mathbb{Z}^d)^T) < \eta \). So there exists some \( \omega \in (\mathbb{Z}^d)^T \) for which

\[
\|L^*\varphi - \omega\|_{\infty} < \eta.
\]

We expand both \( L^*\varphi \) and \( \omega \) with respect to the dual basis \( \mathcal{B}^* \) from (4.12). So,

\[
L^*\varphi = \sum_{i=1}^{u} \lambda_i x_i^* + \sum_{j=1}^{d-u} \mu_j \Xi(w_j)^* \\
\omega = \sum_{i=1}^{u} \lambda_i x_i^* + \sum_{j=1}^{d-u} \mu_j \Xi(w_j)^*.
\]

Since \( \mathcal{B}^* \) is a lattice basis for \((\mathbb{Z}^d)^T\), we have \( \lambda_i \in \mathbb{Z} \) and \( \mu_j \in \mathbb{Z} \) for each \( i, j \). Since the change of basis matrix between \( \mathcal{B}^* \) and the standard dual basis has integer coefficients that are bounded in absolute value by \( O_{c, C}(1) \) (part (2) of Lemma 4.8), one has \( |\lambda_i - \lambda'_i| = O_{c, C}(\eta) \) and \( |\mu_j - \mu'_j| = O_{c, C}(\eta) \) for each \( i, j \).

Let \( w_1^*, \ldots, w_{d-u}^* \) denote the standard dual basis of \((\mathbb{R}^{d-u})^*\), and define

\[
\omega' := \sum_{j=1}^{d-u} \mu'_j w_j^*.
\]

Certainly \( \omega' \in (\mathbb{Z}^{d-u})^T \). We claim that there exists a map \( \varphi' \in (\mathbb{R}^{m-u})^* \) such that \( \tau_1 \ll c, c \ \|\varphi'\|_{\infty} \ll c, c \ \tau_2^{-1} \) and \( \|(L^*)^*\varphi' - \omega'\|_{\infty} \ll c, c \ \eta \), which will immediately resolve (4.19) and part (9).

Indeed, recall the decomposition \( \mathbb{R}^m = (\text{span}(Lx_i : i \leq u)) \oplus \ker \Theta \) as an algebraic direct sum from (4.13). Let \( \varphi = \varphi_1 + \varphi_2 \), where \( \varphi_1 \in (\text{span}(Lx_i : i \leq u))^0 \) and \( \varphi_2 \in (\ker \Theta)^0 \). Since \( \text{dist}(\varphi_1(\ker \Theta)^0) \geq \tau_1 \), we have \( \|\varphi_1\|_{\infty} \geq \tau_1 \). By the properties of the matrix \( P \) (4.17 and 4.18) there exists some \( \varphi' \in (\mathbb{R}^{m-u})^* \) such that

\[
\varphi_1 = P^* \pi^*_{m-u} \varphi'.
\]

Furthermore, by evaluating \( \varphi' \) at the standard basis vectors, one sees that

\[
\tau_1 \ll c, c \ \|\varphi'\|_{\infty} \ll c, c \ \tau_2^{-1}.
\]

We shall use this \( \varphi' \).

By evaluating \( L^*\varphi_1 \) at the elements of \( \mathcal{B} \) one immediately sees that

\[
L^*\varphi_1 = \sum_{j=1}^{d-u} \mu_j \Xi(w_j)^*.
\]

Hence

\[
\Xi^* L^* P^* \pi^*_{m-u} \varphi' = \sum_{j=1}^{d-u} \mu_j w_j^*.
\]
in other words \((L')^\ast \varphi' = \sum_{j=1}^{d-u} \mu_j \mathbf{W}_j^\ast\). But since \(|\mu_j - \mu_j'| = O_{c,C}(\eta)\) for each \(j\), one has \(\|(L')^\ast \varphi' - \omega'\|_\infty = O_{c,C}(\eta)\) as required. This settles part (9).

The entire lemma is settled. \qed

The final lemma we need in order to deduce Theorem \ref{thm:2.10} involves removing the sharp cut-off \(G\).

**Lemma 4.11** (Removing image cut-off). Let \(m, d, h\) be natural numbers, satisfying \(d \geq h \geq m + 1\). Let \(c, C, \varepsilon\) be positive, and let \(\sigma_G\) be any parameter in the range \(0 < \sigma_G < 1/2\). Let \(L' : \mathbb{R}^h \rightarrow \mathbb{R}^m\) be a purely irrational surjective map, and let \(\Xi : \mathbb{R}^h \rightarrow \mathbb{R}^d\) be an injective map. Suppose that \(\|L'\|_\infty \leq C\) and that \(\text{dist}(L', V_{\text{rank}}(m, h)) \geq c\).

Let \(F' : \mathbb{R}^h \rightarrow [0, 1]\) be any function supported on \([-N, N]^h\), and let \(G' : \mathbb{R}^m \rightarrow [0, 1]\) be the indicator function of a convex set contained within \([-\varepsilon, \varepsilon]^m\). Then there exists a Lipschitz function \(G'_{\bar{r}, \sigma_G, 1}\) supported on \([-O_{c,C,\varepsilon}(1), O_{c,C,\varepsilon}(1)]^m\), and with Lipschitz constant \(O_{c,C,\varepsilon}(1/\sigma_G)\), such that, for any parameter \(\tau_2\) in the range \(0 < \tau_2 \leq 1\) and for any functions \(f_1, \ldots, f_d : [N] \rightarrow [-1, 1]\),

\[
|T_{F', G'_{\bar{r}, \sigma_G, 1}, N}(f_1, \ldots, f_d)| + \sigma_G + \frac{\tau_2^{-o(1)}}{\sigma_G} + \frac{\tau_2^{-O(1)}}{N}.
\]

**Proof.** Applying Lemma \ref{lem:3.2} to the function \(G'\), we have

\[G'_{\bar{r}} = G'_{\bar{r}, \sigma_G, 1} + O(G'_{\bar{r}, \sigma_G, 2}),\]

where \(G'_{\bar{r}, \sigma_G, 1}, G'_{\bar{r}, \sigma_G, 2} : \mathbb{R}^m \rightarrow [0, 1]\) are Lipschitz functions with Lipschitz constant \(O_{c,C,\varepsilon}(1/2L)\), both supported on \([-O_{c,C,\varepsilon}(1), O_{c,C,\varepsilon}(1)]^m\), and with \(\int_x G'_{\bar{r}, \sigma_G, 2}(x) \, dx = O_{c,C,\varepsilon}(\sigma_G)\).

By the triangle inequality,

\[
|T_{F', G'_{\bar{r}, \sigma_G, 2}, N}(f_1, \ldots, 1)| \leq T_{F', G'_{\bar{r}, \sigma_G, 2}, N}(f_1, \ldots, 1).
\]

We now apply Lemma \ref{lem:3.3} with linear map \(L'\) and Lipschitz function \(G'_{\bar{r}, \sigma_G, 2}\). Inserting the bound from Lemma \ref{lem:3.3} the present lemma follows. \qed

We conclude this section by combining the three previous lemmas, along with Theorem \ref{thm:4.6} to deduce our main result.

**Proof of Theorem 2.10 assuming Theorem 4.6.** Assume the hypotheses of Theorem \ref{thm:2.10}; let \(\sigma_F\) and \(\sigma_G\) be any parameters satisfying \(0 < \sigma_F, \sigma_G < 1/2\), and let \(\tau_2\) be any parameter satisfying \(0 < \tau_2 \leq 1\).

By Lemma \ref{lem:4.9},

\[
|T_{F, G, N}(f_1, \ldots, f_d)| + \sigma_F + O_{c,C}(\sigma_F),
\]

for some function \(F_{1, \sigma_F} : \mathbb{R}^d \rightarrow [0, 1]\) supported on \([-2N, 2N]^d\) and with Lipschitz constant \(O(1/\sigma_F N)\). By part (4) of Lemma \ref{lem:4.10}, writing \(F_{1, \sigma_F}\) for \(F\), we have

\[
|T_{F_{1, \sigma_F}}(f_1, \ldots, f_d)| \leq \sum_{F' \in \mathcal{R}} |T_{F'_{\bar{r}, \sigma_G, N}}(f_1, \ldots, f_d)|,
\]

where the objects \(F', G_{\bar{r}}, L', \Xi\) and \(\tilde{R}\) satisfy all the conclusions of that lemma.

Parts (1), (5) and (6) of Lemma \ref{lem:4.10} show that \(\Xi\) and \(L'\) satisfy the hypotheses of Lemma \ref{lem:4.11} where in the notation of Lemma \ref{lem:4.11} we take \(h := d - u\) and rewrite
Lemma 4.10, the hypotheses are satisfied so that we may apply Theorem 4.6 to the expression \( R \) above). Therefore there exists an \( \rho \) for some \( A \) the term \(|\langle \rangle| \) (Recall that \(|\tilde{R}| = O_{c,c,\varepsilon}(1)\), by part (2) of Lemma 4.10).

By conclusion (8) of Lemma 4.10 we may replace the term \( A_L(\Omega_{c,C}(1), \tau_2)^{-1} \) with the term \( A_L(\Omega_{c,C}(1), \Omega_{c,C}(\tau_2))^{-1} \).

Since \( F, L, \Xi, \) and \( \tilde{R} \) together satisfy conclusions (1), (2), (3), (6), and (7) of Lemma 4.10, the hypotheses are satisfied so that we may apply Theorem 4.6 to the expression \( T_{F,\tilde{R},G,\Xi,\tau}^{L,N}(f_1, \cdots, f_d) \). (We take \( h = d - u \) and rewrite \( m \) for \( m - u \), as above). Therefore there exists an \( s \) at most \( d - 2 \), independent of \( F, G, \) and \( \tilde{R} \), such that, if

\[
\min_j \| f_j \|_{U^{s+1}[N]} \leq \rho,
\]

for some \( \rho \) in the range \( 0 < \rho \leq 1 \) then \(|T_{F,G,N}^L(f_1, \cdots, f_d)|\) is

\[
\ll_{c,c,\varepsilon} \rho^\Omega(1) (\sigma_F^{-O(1)} + \sigma_G^{-O(1)}) + \sigma_F^{-O(1)} N^{-\Omega(1)}
\]

\[
+ \sigma_G + \frac{\tau_2^{1-\sigma(1)}}{\sigma_G} + \frac{\tau_2^{-O(1)} A_L(\Omega_{c,C}(1), \Omega_{c,C}(\tau_2))^{-1}}{N} + \sigma_F.
\]  

(4.20)

It remains to pick appropriate parameters. Let \( C_1 \) be a constant that is suitably large in terms of \( c, C, \) and all \( O(1) \) constants, and let \( c_1 \) be a constant that is suitably small in terms of all \( O(1) \) constants. Pick \( \sigma_F := \sigma_G := \rho^{c_1} \) and \( \tau_2 := C_1 \rho \). Then

\[
|T_{F,G,N}^L(f_1, \cdots, f_d)| \ll_{c,c,\varepsilon} \rho^\Omega(1) + o_{\rho,A_L,c,C}(1),
\]

where, after the combining the various error terms from (4.21), the \( o_{\rho,A_L,c,C}(1) \) term may be bounded above by

\[
N^{-\Omega(1)} \rho^{-O(1)} A_L(\Omega_{c,C}(1), \rho)^{-1},
\]

as \( A_L(\tau_1, \tau_2) \) is monotonically decreasing as \( \tau_2 \) decreases. This is the desired conclusion of Theorem 2.10  

In order to resolve our main result, then, it suffices to prove\(^5\) Theorem 4.6.

5. Transfer from \( \mathbb{Z} \) to \( \mathbb{R} \)

As remarked above, our present task is to prove Theorem 4.6. Any reader only wishing to consider the case of diophantine inequalities with Lipschitz cut-offs may begin here, and eschew section 4.

We devote this section to the formulation and proof of a certain ‘transfer’ argument, whereby we replace the discrete summation in the definition of \( T_{F,G,N}^{L,\Xi,\delta}(f_1, \cdots, f_d) \) with

\(^5\)The reader may have noticed from the proof above that, in fact, it suffices to prove Theorem 4.6 in the case that \( L \) is purely irrational, but the general version is no harder to prove.
an integral.

Let us introduce some notation for the integral in question.

**Definition 5.1.** Let \( N, m, d, h \) be natural numbers, with \( d \geq h \geq m + 2 \). Let \( \varepsilon \) be positive. Let \( \Xi : \mathbb{R}^h \to \mathbb{R}^d \) and \( L : \mathbb{R}^h \to \mathbb{R}^m \) be linear maps. Let \( F : \mathbb{R}^h \to [0, 1] \) and \( G : \mathbb{R}^m \to [0, 1] \) be two functions, with \( F \) supported on \([-N, N]^h\) and \( G \) supported on \([-\varepsilon, \varepsilon]^m\). Let \( g_1, \ldots, g_d : \mathbb{R} \to [-1, 1] \) be arbitrary functions. We define

\[
\widehat{T}_{F,G,N}(g_1, \ldots, g_d) := \frac{1}{N^{h-m}} \int_{x \in \mathbb{R}^h} \left( \prod_{j=1}^{d} g_j(x) \right) F(x) G(Lx) \, dx. \tag{5.1}
\]

Next, we determine a particular class of measurable functions that will be useful to us.

**Definition 5.2** (\( \eta \)-supported). Let \( \chi : \mathbb{R} \to [0, 1] \) be a measurable 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 \( x \in [-\eta/2, \eta/2] \).

**Definition 5.3** (Convolution). If \( f : \mathbb{Z} \to \mathbb{R} \) has finite support, and \( \chi : \mathbb{R} \to [0, 1] \) is a measurable function, we may define the (rather singular) convolution \((f * \chi)(x) : \mathbb{R} \to \mathbb{R}\) by

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

We note that if \( \chi \) is \( \eta \)-supported, for small enough \( \eta \), then there is only one possible integer \( n \) that makes a non-zero contribution to above summation.

We now state the key lemma.

**Lemma 5.4.** Let \( N, m, d, h \) be natural numbers, with \( d \geq h \geq m + 2 \), and let \( c, C, \varepsilon, \eta \) be positive constants. Let \( \Xi : \mathbb{R}^h \to \mathbb{R}^d \) be an injective linear map with integer coefficients, and assume that \( \Xi(Z^h) = Z^d \cap \text{im} \Xi \). Let \( L : \mathbb{R}^h \to \mathbb{R}^m \) be a surjective linear map. Assume that \( \|\Xi\|_{\infty} \leq C, \|L\|_{\infty} \leq C, \) and \( \text{dist}(L, V_{\text{rank}}(m, h)) \geq c \). Let \( F : \mathbb{R}^h \to [0, 1] \) be a Lipschitz function supported on \([-N, N]^h\) with Lipschitz constant \( O(1/\sigma_F N) \), and let \( G : \mathbb{R}^m \to [0, 1] \) be a Lipschitz function supported on \([-\varepsilon, \varepsilon]^m\) with Lipschitz constant \( O(1/\sigma_G) \). Let \( \tilde{r} \) be a fixed vector in \( \mathbb{Z}^d \), satisfying \( \|\tilde{r}\|_{\infty} = O(c, c, \varepsilon)(1) \). Let \( \chi : \mathbb{R} \to [0, 1] \) be an \( \eta \)-supported measurable function. Then, if \( \eta \) is small enough (in terms of the dimensions \( m, d, h \), and \( \varepsilon \)) there exists some positive real number \( C_{\Xi, \chi} \) such that, if \( f_1, \ldots, f_d : [N] \to [-1, 1] \) are arbitrary functions,

\[
\widehat{T}_{F,G,N}(f_1, \ldots, f_d) = \frac{1}{C_{\Xi, \chi} \eta^h} \widehat{T}_{F,G,N}(f_1 \ast \chi, \ldots, f_d \ast \chi) + O_{C, c, \varepsilon}(\eta/\sigma_G) + O_{C, c, \varepsilon}(\eta/\sigma_F N). \tag{5.2}
\]

Moreover, \( C_{\Xi, \chi} \ll C \).

**Proof.** Let \( \chi : \mathbb{R}^d \to [0, 1] \) denote the function \( x \mapsto \prod_{i=1}^{d} \chi(x_i) \). We choose

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

Since \( \chi \) is \( \eta \)-supported, \( C_{\Xi, \chi} \ll C \).
Then, expanding the definition of the convolution,

\[ \frac{1}{C_{\Xi,\eta}^h} \tilde{T}_{f, \Xi, \eta}^{L, \Xi, \eta}(f_1 * \cdots * f_d * \chi) \]

equals

\[ \frac{1}{N^{h-m}} \sum_{n \in \mathbb{Z}^d} \left( \prod_{j=1}^d f_j(n_j) \right) \frac{1}{C_{\Xi,\eta}^h} \int_{y \in \mathbb{R}^h} F(y)G(Ly)\chi(\Xi(y) + \tilde{r} - n) \, dy. \tag{5.3} \]

Note that any vector \( n \in \mathbb{Z}^d \) that gives a non-zero contribution to expression (5.3) satisfies \( \|n - \Xi(y) - \tilde{r}\|_\infty \ll \eta \), for some \( y \in \mathbb{R}^h \). Therefore, \( n \) must be of the form \( \Xi(n') + \tilde{r} \) for some unique \( n' \in \mathbb{Z}^h \). (This is proved in full in Lemma D.2). Therefore, writing \( \Xi = (\xi_1, \cdots, \xi_d) \), we may reformulate (5.3) as

\[ \frac{1}{N^{h-m}} \sum_{n \in \mathbb{Z}^h} \left( \prod_{j=1}^d f_j(\xi_j(n) + \tilde{r}_j) \right) \frac{1}{C_{\Xi,\eta}^h} \int_{y \in \mathbb{R}^h} F(y)G(Ly)\chi(\Xi(y) - n) \, dy, \]

which is equal to

\[ \frac{1}{N^{h-m}} \sum_{n \in \mathbb{Z}^h} \left( \prod_{j=1}^d f_j(\xi_j(n) + \tilde{r}_j) \right) \frac{1}{C_{\Xi,\eta}^h} \int_{y \in \mathbb{R}^h} (F(n) + O_C(\eta/\sigma_{F,N}))G(Ly)\chi(\Xi(y) - n) \, dy. \tag{5.4} \]

Indeed, the inner integral is only non-zero when \( \|\Xi(y) - \Xi(n)\|_\infty \ll \eta \), and this implies that \( \|y - n\|_\infty \ll C^{O(1)} \eta \). (This is proved in full in Lemma D.3). Then recall that \( F \) has Lipschitz constant \( O(1/\sigma_{F,N}) \).

Continuing, expression (5.3) is equal to

\[ \frac{1}{N^{h-m}} \sum_{n \in \mathbb{Z}^h} \left( \prod_{j=1}^d f_j(\xi_j(n) + \tilde{r}_j) \right) \frac{1}{C_{\Xi,\eta}^h} \int_{y \in \mathbb{R}^h} (F(n) + O_C(\eta/\sigma_{F,N}))G(Ly)\chi(\Xi(y) - n) \, dy \]

(5.5)

where

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

and \( E \) is a certain error, that may be bounded above by

\[ \ll_C \frac{\eta}{\sigma_{F,N}} \frac{1}{N^{h-m}} \sum_{n \in [-O(N),O(N)]^h} H(Ln). \tag{5.6} \]

Let us deal with the first term of (5.5), in which we wish to replace \( H \) with \( G \). We therefore consider

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

which is

\[ \ll \frac{1}{N^{h-m}} \sum_{n \in \mathbb{Z}^h} F(n)|G - H|(Ln). \tag{5.7} \]
Using Lemma A.3 again, the function $H$ is supported on $[-\varepsilon - O_C(\eta), \varepsilon + O_C(\eta)]^m$. Thus, if $\eta$ is small enough in terms of $\varepsilon$, the function $|G - H| : \mathbb{R}^m \to \mathbb{R}$ is supported on $[-O_C(\varepsilon), O_C(\varepsilon)]^m$. Furthermore, $\|G - H\|_\infty = O_C(\eta/\sigma_G)$. Indeed,
\[
G(x) = -\frac{1}{C_{\Xi, \chi} \eta^h} \int_{y \in \mathbb{R}^h} G(x + Ly) \chi(\Xi(y)) \, dy \\
= G(x) - \frac{1}{C_{\Xi, \chi} \eta^h} \int_{y \in \mathbb{R}^h} (G(x) + O_C(\eta/\sigma_G)) \chi(\Xi(y)) \, dy \\
= O_C(\eta/\sigma_G),
\]
by the definition of $C_{\Xi, \chi}$. So, by the crude bound given in Lemma 3.2, (5.7) may be bounded above by $O_{c, C, \varepsilon}(\eta/\sigma_G)$.

Turning to the error $E$ from (5.5), we’ve already remarked that it may be bounded above by expression (5.6). Applying Lemma 3.2 again, expression (5.6) may be bounded above by $O_{c, C, \varepsilon}(\eta/\sigma_F N)$.

Lemma 5.4 follows immediately upon substituting the estimates on (5.6) and (5.7) into (5.8).

We finish this section by noting a simple relationship between the Gowers norms $\|f \ast \chi\|_{U^{s+1}(\mathbb{R}, 2N)}$ and the Gowers norms $\|f\|_{U^{s+1}[N]}$.

**Lemma 5.5** (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 $\chi : \mathbb{R} \to [0, 1]$ be an $\eta$-supported measurable function. Let $N$ be a natural number, and let $f : [N] \to \mathbb{R}$ be an arbitrary function. View $f \ast \chi$ as a function supported on $[-2N, 2N]$. Then we have
\[
\|f \ast \chi\|_{U^{s+1}(\mathbb{R}, 2N)} \ll \eta^{\frac{s+1}{2s+2}} \|f\|_{U^{s+1}[N]}.
\]

The definition of the real Gowers norm $\|f \ast \chi\|_{U^{s+1}(\mathbb{R}, 2N)}$ is recorded in Definition A.3.

**Proof.** From expression (A.5), we have
\[
\|f \ast \chi\|_{U^{s+1}(\mathbb{R}, 2N)} \ll \frac{1}{N^{s+2}} \int \prod_{x, h \in \mathbb{R}^{2s+2}} (f \ast \chi)(x + h \cdot \omega) \, dx \, dh.
\]
Substituting in the definition of $f \ast \chi$, this is equal to
\[
\frac{1}{N^{s+2}} \sum_{n = (n_\omega)_{\omega \in \{0, 1\}^{s+1} \in \mathbb{Z}^{\{0, 1\}^{s+1}}}} \left( \prod_{\omega \in \{0, 1\}^{s+1}} f(n_\omega) \right) \int_{(x, h) \in \mathbb{R}^{s+2}} \chi(\Psi(x, h) - n) \, dx \, dh,
\]
where $\Psi : \mathbb{R}^{s+2} \to \mathbb{R}^{2s+1}$ has coordinate functions $\psi_\omega$, indexed by $\omega \in \{0, 1\}^{s+1}$, where $\psi_\omega(x, h) := x + h \cdot \omega$. In similar notation to that used in the previous proof, for $x \in \mathbb{R}^{2s+1}$, we let $\chi(x) := \prod_{i=1}^{2s+1} \chi(x_i)$. Note that $\Psi$ is injective, $\Psi(\mathbb{Z}^{s+2}) = \mathbb{Z}^{2s+1} \cap \text{im} \, \Psi$, and $\|\Psi\|_\infty = O_{s}(1)$.

The contribution to the inner integral of (5.8) from a particular $n$ is zero unless $\|n - \Psi(x, h)\|_\infty \ll \eta$, for some $(x, h) \in \mathbb{R}^{s+2}$. Therefore, if $\eta$ is small enough we can conclude that $n$ must be of the form $\Psi(p, k)$, for some unique $(p, k) \in \mathbb{Z}^{s+2}$. (To spell
it out, apply Lemma D.2 with the map Ψ in place of the map Ξ). So (5.9) is equal to
\[
\frac{1}{N^{s+2}} \sum_{(p,k) \in \mathbb{Z}^{s+2}} \left( \prod_{\omega \in \{0,1\}^{s+1}} f(p + k \cdot \omega) \right) \int_{(x,h) \in \mathbb{R}^{s+2}} \chi(\Psi(x - p, h - k)) \, dx \, dh,
\]
(5.10)
which, after a change of variables, is equal to
\[
\frac{C}{N^{s+2}} \sum_{(p,k) \in \mathbb{Z}^{s+2}} \prod_{\omega \in \{0,1\}^{s+1}} f(p + k \cdot \omega),
\]
(5.11)
where
\[
C := \int_{(x,h) \in \mathbb{R}^{s+2}} \chi(\Psi(x,h)) \, dx \, dh.
\]
Since \( \chi \) has support contained within \([-\eta, \eta]^{2s+1} \), a vector \((x,h)\) only makes a non-zero contribution to the above integral if \( \|\Psi(x,h)\|_\infty \ll \eta \). This implies that \( \|(x,h)\|_\infty \ll \eta \).

To prove this is full, apply Lemma D.3 to the linear map Ψ. Since \( \|\chi\|_\infty = O(1) \), this means \( C = O(\eta^{s+2}) \). The lemma then follows from (5.11). □

6. Normal form

In this section we recall a technical notion from [15] that those authors refer to as normal form. In section 8 we will need to appeal to a quantitative refinement of this notion, which we also develop here.

Let \( \Psi : \mathbb{R}^n \to \mathbb{R}^m \) be a linear map. Putting the standard coordinates on \( \mathbb{R}^n \) and \( \mathbb{R}^m \), we may write \( (\psi_1, \ldots, \psi_m) := \Psi : \mathbb{R}^n \to \mathbb{R}^m \) as a system of homogeneous linear forms. The crux of the theory from [15] is that, provided \( \Psi \) is of so-called ‘finite Cauchy-Schwarz complexity’, \( \Psi \) may be reparametrised in such a way that it interacts particularly well with certain applications of the Cauchy-Schwarz inequality (see Proposition 8.3). Below we will give a brief overview of this terminology, before introducing our own quantitative versions; a much fuller discussion may be found in [15, Section 1] and [11].

In words, a reparametrisation into normal form is one in which each linear form is the only one that mentions all of its particular collection of variables. For example, the forms
\[
\begin{align*}
\psi_1(t,u,v) &= u + v \\
\psi_2(t,u,v) &= v + t \\
\psi_3(t,u,v) &= u + t \\
\psi_4(t,u,v) &= u + v + t
\end{align*}
\]
(6.1)
are in normal form with respect to \( \psi_1 \), since \( \psi_4 \) is the only form to utilise all three of the variables. However, this system is not in normal form with respect to \( \psi_3 \), say. However, the system
\[
\begin{align*}
\psi_1(t,u,v,w) &= u + v + 2w \\
\psi_2(t,u,v,w) &= v + t - w \\
\psi_3(t,u,v,w) &= u + t - w \\
\psi_4(t,u,v,w) &= u + v + t,
\end{align*}
\]
(6.2)
that parametrises the same subspace of \( \mathbb{R}^4 \), is in normal form for all \( i \).

We repeat the precise definition from [15].
Definition 6.1. Let $m,n$ be natural numbers, and let $(\psi_1, \cdots, \psi_m) = \Psi : \mathbb{R}^n \to \mathbb{R}^m$ be a system of homogeneous linear forms. Let $i \in [m]$. We say that $\Psi$ is in normal form with respect to $\psi_i$ if there exists a non-negative integer $s$ and a collection $J_i \subseteq \{e_1, \cdots, e_n\}$ of the standard basis vectors, satisfying $|J_i| = s + 1$, such that
\[ \prod_{e \in J_i} \psi_i(e) \text{ is non-zero when } i' = i \text{ and vanishes otherwise}. \]
We say that $\Psi$ is in normal form if it is in normal form with respect to $\psi_i$ for every $i$.

Let us also recall what it means for a certain system of forms $\Psi'$ to extend the system of forms $\Psi$.

Definition 6.2. For a system of homogeneous linear forms $(\psi_1, \cdots, \psi_m) = \Psi : \mathbb{R}^n \to \mathbb{R}^m$, an extension $(\psi'_1, \cdots, \psi'_m) = \Psi' : \mathbb{R}^{n'} \to \mathbb{R}^m$ is a system of homogeneous linear forms on $\mathbb{R}^{n'}$, for some $n'$ with $n' \geq n$, such that
\begin{enumerate}
  \item $\Psi'(\mathbb{R}^{n'}) = \Psi(\mathbb{R}^n)$;
  \item if we identify $\mathbb{R}^n$ with the subset $\mathbb{R}^n \times \{0\}^{n'-n}$ in the obvious manner, then $\Psi$ is the restriction of $\Psi'$ to this subset.
\end{enumerate}

The paper [15] includes a result (Lemma 4.4) on the existence of extensions in normal form, but we will need a quantitative refinement of this analysis.

The reader will note from examples (6.1) and (6.2) that the property of ‘being in normal form’ is a property of the parametrisation, and not of the underlying space that is being parametrised. It is natural to wonder whether there is some property of a space that can enable one to find a parametrisation in normal form, even if the original parametrisation is not. Fortunately there is such a notion, and it is the finite (Cauchy-Schwarz) complexity introduced in [15]. We introduce this notion in the following definitions, which we have phrased in such a way as to help us formulate a quantitative version.

Definition 6.3 (Suitable partitions). Let $m,n$ be natural numbers, with $m \geq 2$, and let $(\psi_1, \cdots, \psi_m) = \Psi : \mathbb{R}^n \to \mathbb{R}^m$ be a system of homogeneous linear forms. Fix $i \in [m]$. Let $\mathcal{P}_i$ be a partition of $[m] \setminus \{i\}$, i.e.
\[ [m] \setminus \{i\} = \bigcup_{k=1}^{s+1} C_k \]
for some $s$ satisfying $0 \leq s \leq m - 2$ and some disjoint sets $C_k$. We say that $\mathcal{P}_i$ is suitable for $\Psi$ if
\[ \psi_i \notin \text{span}_k(\psi_j : j \in C_k) \]
for any $k$.

Definition 6.4 (Degeneracy varieties). Let $m,n$ be natural numbers, with $m \geq 2$. Let $\mathcal{P}_i$ be a partition of $[m] \setminus \{i\}$. We define the $\mathcal{P}_i$-degeneracy variety $V_{\mathcal{P}_i}$ to be the set of all the systems of homogeneous linear forms $\Psi : \mathbb{R}^n \to \mathbb{R}^m$ for which $\mathcal{P}_i$ is not suitable for $\Psi$. Finally, the degeneracy variety $V_{\text{degen}}(n,m)$ is given by
\[ V_{\text{degen}}(n,m) := \bigcap_{i=1}^m V_{\mathcal{P}_i}, \]
where the inner intersection is over all possible partitions $\mathcal{P}_i$. 
It is easy to observe that $\Psi \in V_{\text{degen}}(n,m)$ if and only if, for some $i \neq j$, $\psi_i$ is a real multiple of $\psi_j$.

In [15, Definition 1.5], the authors refer to those $\Psi \in V_{\text{degen}}(n,m)$ as having infinite (Cauchy-Schwarz) complexity, and develop their theory for $\Psi \notin V_{\text{degen}}(n,m)$. As we did for describing degeneracy properties of $\Psi$, we remark for readers familiar with [15] that we preclude the ‘complexity 0’ case.

**Definition 6.6 (c₁-Cauchy-Schwarz complexity).** Let $m, n$ be natural numbers, with $m \geq 3$, and let $c_1$ be a positive constant. Let $(\psi_1, \ldots, \psi_m) = \Psi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a system of homogeneous linear forms. For $i \in [m]$, we define a quantity $s_i$ either by defining $s_i + 1$ to be the minimal number of parts in a partition $\mathcal{P}_i$ of $[m] \setminus \{i\}$ such that $\text{dist}(\Psi, V_{\mathcal{P}_i}) \geq c_1$, or by $s_i = \infty$ if no such partition exists. Then we define $s := \max(1, \max_i s_i)$. We say that $s$ is the $c_1$-Cauchy-Schwarz complexity of $\Psi$.

We remark, for readers familiar with [15], that we preclude the ‘complexity 0’ case. This is for a mundane technical reason, that occurs when absorbing the exponential phases in section 8, when it will be convenient that $s + 1 \geq 2$. This is why we need the condition $m \geq 3$ in the above definition. We also take this opportunity to note that if $s$ satisfies the above definition, and $s \neq \infty$, then $2 \leq s + 1 \leq m - 1$.

We note an easy consequence of these definitions.

**Lemma 6.6.** Let $m, n$ be natural numbers, with $m \geq 3$, and let $c_1$ be a positive constant. Let $(\psi_1, \ldots, \psi_m) = \Psi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a system of homogeneous linear forms. Suppose that $\text{dist}(\Psi, V_{\text{degen}}(n,m)) \geq c_1$. Then $\Psi$ has finite $c_1$-Cauchy-Schwarz complexity.

**Proof.** We have already observed that $\Psi \in V_{\text{degen}}(n,m)$ if and only if, for some $i \neq j$, $\psi_i$ is a real multiple of $\psi_j$. From now until the end of the proof, fix $\mathcal{P}_i$ to be the partition of $[m] \setminus \{i\}$ in which every form $\psi_k$ is in its own part. Our initial observation then implies that $\Psi \in V_{\text{degen}}(n,m)$ if and only if $\Psi \in V_{\mathcal{P}_i}$ for some $i$. So $\text{dist}(\Psi, V_{\text{degen}}(n,m)) \geq c_1$ implies that $\text{dist}(\Psi, V_{\mathcal{P}_i}) \geq c_1$ for all $i$. Therefore, by using these partitions $\mathcal{P}_i$ in Definition 6.5, we conclude that $\Psi$ has finite $c_1$-Cauchy-Schwarz complexity. \[\square\]

After having built up these definitions, we state the key proposition on the existence of normal form extensions to systems of real linear forms. We remind the reader that all implied constants may depend on the dimensions of the underlying spaces.

**Proposition 6.7 (Normal form algorithm).** Let $m, n$ be natural numbers, with $m \geq 3$, and let $c_1, C_1$ be positive constants. Let $(\psi_1, \ldots, \psi_m) = \Psi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a system of homogeneous linear forms, and suppose that the coefficients of $\Psi$ are bounded above in absolute value by $C_1$. Furthermore, suppose that $\Psi$ has $c_1$-Cauchy-Schwarz complexity $s$, for some finite $s$. Then, for each $i \in [m]$, there is an extension $\Psi' : \mathbb{R}^n \rightarrow \mathbb{R}^m$ such that:

1. $n' = n + s + 1 \leq n + m - 1$;
2. $\Psi'$ is of the form $\Psi'(u, w_1, \ldots, w_{s+1}) := \Psi(u + w_1 f_1 + \cdots + w_{s+1} f_{s+1})$ for some vectors $f_k \in \mathbb{R}^n$, such that $\|f_k\|_\infty = O_{c_1, C_1}(1)$ for every $k$;
3. $\Psi'$ is in normal form with respect to $\psi'_i$;
4. $\psi'_i(0, w) = w_1 + \cdots + w_{s+1}$.
The proof is deferred to Appendix C as it is very similar to the proof from [15] (although with one important extra subtlety, which we mention in the appendix).

We conclude this discussion of normal form by noting an example of a system of homogeneous linear forms that may be reparametrised in normal form, but without quantitative control over the resulting extension.

Indeed, take \( \iota(N) \) to be some function such that \( \iota(N) \to \infty \) as \( N \to \infty \). Consider the forms

\[
\psi_1(u_1, u_2, u_3) = (1 + \iota(N)^{-1})u_1 + u_2 \\
\psi_2(u_1, u_2, u_3) = u_1 + u_2 \\
\psi_3(u_1, u_2, u_3) = u_3.
\]

Notice that \( \text{dist}(\Psi, V_{\text{degen}}(3, 3)) \to 0 \) as \( N \to \infty \), so \( \Psi \) does not have finite \( c_1 \)-Cauchy-Schwarz complexity for any positive absolute constant \( c_1 \). One may nonetheless construct a normal form reparametrisation

\[
\psi'_1(u_1, u_2, u_3, w_1, w_2) = (1 + \iota(N)^{-1})u_1 + u_2 + w_1 \\
\psi'_2(u_1, u_2, u_3, w_1, w_2) = u_1 + u_2 + w_2 \\
\psi'_3(u_1, u_2, u_3, w_1, w_2) = u_3.
\]

The system \( \Psi \) does have all its non-zero coefficients bounded away from 0 and \( \infty \), but

\[
\Psi'(u_1, u_2, u_3, w_1, w_2) = \Psi(u_1 + \iota(N)w_1 - \iota(N)w_2, u_2 - \iota(N)w_1 + (\iota(N) + 1)w_2, u_3),
\]

so \( \Psi' \) is not obtained by bounded shifts of the \( u_i \) variables. Such an extension would not be suitable for our requirements in section 8.

One final remark: in [15], the simple algorithm that constructs normal form extensions with respect to a fixed \( i \) may easily be iterated, and so the authors work with systems that are in normal form with respect to every index \( i \). A careful analysis of the proof in Appendix C of [15] demonstrates that it is sufficient for \( \Psi \) merely to admit, for each \( i \) separately, an extension that is in normal form with respect to \( \psi_i \), but this is of little consequence in [15]. Yet certain quantitative aspects of the iteration of the normal form algorithm, critical to our application of these ideas, are not immediately clear to us. We have stated Proposition 6.7 for normal forms only with respect to a single \( i \), in order to avoid this technical annoyance.

7. Degeneracy relations

Our aim for this short section is to establish a quantitative relationship between the dual pair degeneracy variety \( V^*_{\text{degen}, 2}(m, d, h) \) and the dual degeneracy variety \( V_{\text{degen}}(h - m, d) \) (see Definitions 4.4 and 2.3 respectively), which will be needed in the next section. To introduce the ideas, we first prove a non-quantitative proposition.

**Lemma 7.1.** Let \( m, d, h \) be natural numbers, with \( d \geq h \geq m + 2 \). Let \( \Xi : \mathbb{R}^h \to \mathbb{R}^d \) be an injective linear map, let \( L : \mathbb{R}^h \to \mathbb{R}^m \) be a surjective linear map, and suppose that \( (\Xi, L) \notin V^*_{\text{degen}, 2}(m, d, h) \). Let \( \Phi : \mathbb{R}^{h-m} \to \ker L \) be any surjective linear map. Then the linear map \( \Xi \Phi : \mathbb{R}^{h-m} \to \mathbb{R}^d \), viewed as a system of homogenous linear forms, is not in \( V_{\text{degen}}(h - m, d) \).
Proof. Let $e_1, \ldots, e_d$ denote the standard basis vectors in $\mathbb{R}^d$, and let $e_1^*, \ldots, e_d^*$ denote the dual basis of $(\mathbb{R}^d)^*$. Suppose for contradiction that $\Xi \Phi \in V_{\text{degen}}(h-m, d)$. Then by definition there exist two indices $i, j \leq d$, and a real number $\lambda$, such that $e_i^* - \lambda e_j^*$ is non-zero and $\Xi \Phi \in (\mathbb{R}^h-m) \subset \ker(e_i^* - \lambda e_j^*)$.

But then $\Phi(\mathbb{R}^{h-m}) \subset \ker(\Xi^*(e_i^* - \lambda e_j^*))$, i.e. $\Xi^*(e_i^* - \lambda e_j^*) \in (\ker L)^\circ$. But $(\ker L)^\circ = L^*((\mathbb{R}^m)^*)$, and so $\Xi^*(e_i^* - \lambda e_j^*) \in L^*((\mathbb{R}^m)^*)$.

Then, by the definition of $V_{\text{degen,2}}(m, d, h)$, we have $(\Xi, L) \in V_{\text{degen,2}}(m, d, h)$, which is a contradiction.

□

The ideas having been introduced, we state the quantitative version we require.

**Lemma 7.2.** Let $m, d, h$ be natural numbers, with $d \geq h \geq m + 2$, and let $c, C$ be positive constants. Let $\Xi : \mathbb{R}^h \to \mathbb{R}^d$ be a linear map, and let $L : \mathbb{R}^h \to \mathbb{R}^m$ be a surjective linear map. Suppose that $||\Xi||_\infty \leq C$, and $\text{dist}((\Xi, L), V_{\text{degen,2}}^*(m, d, h)) \geq c$.

Let $K$ denote $\ker L$, choose any orthonormal basis $\{v(1), \ldots, v(h-m)\}$ for $K$, and let $\Phi : \mathbb{R}^{h-m} \to K$ denote the associated parametrisation, i.e. $\Phi(x) := \sum_{i=1}^{h-m} x_i v(i)$.

Then $||\Xi \Phi||_\infty = O(C)$ and $\text{dist}(\Xi \Phi, V_{\text{degen}}(h-m, d)) = \Omega(c)$.

For the definition of $\text{dist}((\Xi, L), V_{\text{degen,2}}^*(m, d, h))$, consult Definition 4.5.

Proof. Certainly $||\Phi||_\infty = O(1)$, as the chosen basis $\{v(1), \ldots, v(h-m)\}$ is orthonormal. Therefore $||\Xi \Phi||_\infty = O(C)$.

Let $e_1, \ldots, e_d$ denote the standard basis vectors in $\mathbb{R}^d$, and let $e_1^*, \ldots, e_d^*$ denote the dual basis of $(\mathbb{R}^d)^*$. Suppose for contradiction that $\text{dist}(\Xi \Phi, V_{\text{degen}}(h-m, d)) \leq \eta$ for some small parameter $\eta$. In other words, assume that there exists a linear map $P : \mathbb{R}^{h-m} \to \mathbb{R}^d$ with $||P||_\infty \leq \eta$ such that $\Xi \Phi + P \in V_{\text{degen}}(h-m, d)$. By definition, this means that

$$(\Xi \Phi + P)(\mathbb{R}^{h-m}) \subset \ker(e_i^* - \lambda e_j^*),$$

for some two indices $i, j \leq d$, and some real number $\lambda$, such that $e_i^* - \lambda e_j^*$ is non-zero.

We can factorise $P = Q \Phi$, for some linear map $Q : \mathbb{R}^h \to \mathbb{R}^d$ with $||Q||_\infty \ll \eta$. Indeed, let $f_1, \ldots, f_{h-m}$ denote the standard basis vectors in $\mathbb{R}^{h-m}$, and for all $k$ at most $h-m$ define

$$Q(v^{(k)}) := P(f_k).$$

(If the notation for the indices seems odd here, it is designed to match the notation in Proposition 8.2 in which having superscript on the vectors $v^{(k)}$ seems to be natural). Complete $\{v^{(1)}, \ldots, v^{(h-m)}\}$ to an orthonormal basis $\{v(1), \ldots, v(h)\}$ for $\mathbb{R}^h$ and, for $k$ in the range $h-m+1 \leq k \leq h-m$, define $Q(v^{(k)}) := 0$. Then $P = Q \Phi$, and $||Q||_\infty = O(\eta)$, since $\{v^{(1)}, \ldots, v^{(h)}\}$ is an orthonormal basis.

Thus,

$$(\Xi \Phi + Q \Phi)(\mathbb{R}^{h-m}) \subset \ker(e_i^* - \lambda e_j^*).$$

So

$$\Phi(\mathbb{R}^{h-m}) \subset \ker((\Xi + Q)^* (e_i^* - \lambda e_j^*)).$$

Like the previous proof, we conclude that

$$(\Xi + Q)^* (e_i^* - \lambda e_j^*) \in L^*((\mathbb{R}^m)^*).$$

Hence $(\Xi + Q, L) \in V_{\text{degen,2}}^*(m, d, h)$, which, if $\eta$ is small enough, contradicts the assumption that $\text{dist}((\Xi, L), V_{\text{degen,2}}^*(m, d, h)) \geq c$. □
8. A Generalised von Neumann Theorem

In this section we complete the proof of Theorem 8.1 and therefore complete the proof of our main result (Theorem 2.10). It will be enough to prove the following statement.

**Theorem 8.1.** Let $N, m, d, h$ be natural numbers, with $d \geq h \geq m + 2$, and let $c, C, \varepsilon$ be positive reals. Let $\Xi = \Xi(N) : \mathbb{R}^h \to \mathbb{R}^d$ be an injective linear map with integer coefficients, and let $L = L(N) : \mathbb{R}^d \to \mathbb{R}^m$ be a surjective linear map. Suppose further that $\|L\|_\infty \leq C$, $\|\Xi\|_\infty \leq C$, $\text{dist}(L, V_{\text{rank}}(m, d)) \geq c$ and $\text{dist}((\Xi, L), V_{\text{deg}, 2}(m, d, h)) \geq c$. Then there is some natural number $n$ at most $d - 2$, independent of $\varepsilon$, such that the following holds. Let $\tilde{r} \in \mathbb{Z}^d$ be some vector with $\|\tilde{r}\|_\infty = O_{c, C, \varepsilon}(1)$, and let $\sigma_F$ be a parameter in the range $0 < \sigma_F < 1/2$. Let $F : \mathbb{R}^h \to [0, 1]$ be a Lipschitz function supported on $[-N, N]^h$, with Lipschitz constant $O(1/\sigma_F N)$, and let $G : \mathbb{R}^m \to [0, 1]$ be any function supported on $[-\varepsilon, \varepsilon]^m$. Let $g_1, \ldots, g_d : [-2N, 2N]^d \to [-1, 1]$ be arbitrary measurable functions. Suppose

$$\min_{j \in \mathbb{Z}} \|g_j\|_{U^{s+1}(\mathbb{R}, 2N)} \leq \rho$$

for some $\rho$ at most $1$. Then

$$|T_{F, G, N}^L(\tilde{r}, g_1, \ldots, g_d)| \ll_{c, C, \varepsilon} \rho^{O(1)} \sigma_F^{-1}$$  \hspace{1cm} (8.1)

**Proof that 8.1 implies Theorem 4.6.** Assume the hypotheses of Theorem 4.6. This gives natural numbers $N, m, d, h$, linear maps $L : \mathbb{R}^h \to \mathbb{R}^m$ and $\Xi : \mathbb{R}^d \to \mathbb{R}^d$, and functions $F : \mathbb{R}^h \to [0, 1]$ and $G : \mathbb{R}^m \to [0, 1]$. Let $f_1, \ldots, f_d : [N] \to [-1, 1]$ be arbitrary functions, and for ease of notation let

$$\delta := T_{F, G, N}^L(f_1, \ldots, f_d).$$

From Lemma 3.2 and the triangle inequality, we have the crude bound $\delta = O_{c, C, \varepsilon}(1)$.

Let $\eta := c_1 \delta g$, where $c_1$ is small enough depending on $m, d, h, c, C,$ and $\varepsilon$, and let $\chi : \mathbb{R} \to [0, 1]$ be an $\eta$-supported measurable function (see Definition 5.2). For all $j$ at most $d$, let $g_j := f_j \ast \chi$. Finally, suppose $\min_j \|f_j\|_{U^{s+1}[N]} \leq \rho$, for some parameter $\rho$ in the range $0 < \rho < 1$.

We proceed by bounding $T_{F, G, N}^L(\tilde{r}, g_1, \ldots, g_d)$. Indeed, by Lemma 5.3, if $c_1$ is small enough

$$\min_j \|g_j\|_{U^{s+1}(\mathbb{R})} \ll \frac{\chi}{\eta^{\frac{d+2}{d}}} \min_j \|f_j\|_{U^{s+1}[N]} \ll_{c, C, \varepsilon} \rho.$$  \hspace{1cm} (8.2)

Applying Theorem 8.1 to these functions $g_1, \ldots, g_d$, the above implies

$$T_{F, G, N}^L(\tilde{r}, g_1, \ldots, g_d) \ll_{c, C, \varepsilon} \rho^{-O(1)} \sigma_F^{-1}$$  \hspace{1cm} (8.2)

Now we use this to bound $\delta$ by Gowers norms. Indeed, by Lemma 5.4, we have

$$\delta \ll_{c, C, \varepsilon} \frac{1}{(c_1 \delta g)\text{Ch}} T_{F, G, N}^L(\tilde{r}, g_1, \ldots, g_d) + c_1 \delta + c_1 \delta g \sigma_g^{-1} N^{-1}.$$  \hspace{1cm} (8.2)

Picking $c_1$ small enough, we may move the $c_1 \delta$ term to the left-hand side to get an $\Omega(\delta)$ term. The bound (8.2) then yields

$$\delta^{h+1} \ll_{c, C, \varepsilon} \rho^{O(1)} \sigma_F^{-1} \sigma_g^{-h} + \sigma_g^{-1} N^{-1},$$

and so

$$\delta \ll_{c, C, \varepsilon} \rho^{O(1)} (\sigma_F^{-O(1)} + \sigma_g^{-O(1)}) + \sigma_F^{-O(1)} N^{-O(1)}.$$  \hspace{1cm} (8.2)

This yields the desired conclusion of Theorem 4.6. □
So it remains to prove Theorem 8.1. The bulk of the work will be done in the following two propositions.

**Proposition 8.2** (Separating out the kernel). Let $N, m, d, h$ be natural numbers, with $d \geq h \geq m+2$, and let $c, C, \varepsilon$ be positive constants. Let $\sigma_F$ be a parameter in the range $0 < \sigma_F < 1/2$. Let $\Xi: \mathbb{R}^h \to \mathbb{R}^d$ be an injective linear map with integer coefficients, and let $L: \mathbb{R}^h \to \mathbb{R}^m$ be a surjective linear map. Assume further that $\|L\|_\infty \leq C$, $\|\Xi\|_\infty \leq C$, $\text{dist}(L, V_{\text{rank}}(m, h)) \geq c$ and $\text{dist}((\Xi, L), V_{\text{degen}}^\perp(m, d, h)) \geq c$. Let $F: \mathbb{R}^h \to [0, 1]$ be a Lipschitz function supported on $[-CN, CN]^h$, with Lipschitz constant $O_C(1/\sigma_F N)$, and let $G: \mathbb{R}^m \to [0, 1]$ be a Lipschitz function supported on $[-\varepsilon, \varepsilon]^m$. Let $\tilde{F}$ be a vector in $\mathbb{Z}^d$, satisfying $\|\tilde{F}\|_\infty = O_C(1)$. Then there exists a system of linear forms $(\psi_1, \ldots, \psi_d) = \Psi: \mathbb{R}^{h-m} \to \mathbb{R}^d$, and a Lipschitz function $F_1: \mathbb{R}^{h-m} \to [0, 1]$ supported on $[-O_{c,C,\varepsilon}(N), O_{c,C,\varepsilon}(N)]^{h-m}$ with Lipschitz constant $O(1/\sigma_F N)$, such that, if $g_1, \ldots, g_d: [-2N, 2N] \to [-1, 1]$ are arbitrary functions,

$$
|\hat{T}_{E,F,N}(g_1, \ldots, g_d)| \ll c, \varepsilon, \sigma
$$

where, for each $j$, $a_j$ is some real number that satisfies $a_j = O_{c,C,\varepsilon}(1)$.

Furthermore, there exists a natural number $s$ at most $d - 2$ such that the system $\Psi$ has $\Omega_{c,C}(1)$-Cauchy-Schwarz complexity at most $s$, in the sense of Definition 6.3.

**Proof of Proposition 8.2** For ease of notation, let

$$
\beta := \hat{T}_{E,F,N}(g_1, \ldots, g_d).
$$

Noting that $\ker L$ is a vector space of dimension $h - m$, define $\{v^{(1)}, \ldots, v^{(h-m)}\} \subset \mathbb{R}^h$ to be an orthonormal basis for $\ker L$. Then the map $\Phi: \mathbb{R}^{h-m} \to \mathbb{R}^h$, defined by

$$
\Phi(x) := \sum_{i=1}^{h-m} x_i v^{(i)},
$$

is an injective map that parametrises $\ker L$. (This is reminiscent of Lemma 7.2).

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

$$
\frac{1}{N^{h-m}} \int_{\mathbb{R}^h} F\left(\sum_{i=1}^{h} x_i v^{(i)}\right) G\left(L\left(\sum_{i=1}^{h} x_i v^{(i)}\right)\right) \prod_{j=1}^{d} g_j(\xi_j(\Phi(x_1^{h-m}) + \sum_{i=h-m+1}^{h} x_i v^{(i)}) + \tilde{F}_j) \, dx,
$$

using $x_1^{h-m}$ to refer to the vector in $\mathbb{R}^{h-m}$ given by the first the first $h - m$ coordinates of $x$.

We wish to remove the presence of the variables $x_{h-m+1}, \ldots, x_h$. To set this up, note that, by the choice of the vectors $v^{(i)}$,

$$
G(L\left(\sum_{i=1}^{h} x_i v^{(i)}\right)) = G(L\left(\sum_{i=h-m+1}^{h} x_i v^{(i)}\right)).
$$

The vector $\sum_{i=h-m+1}^{h} x_i v^{(i)}$ is in $(\ker L)^\perp$. Hence, due to the limited support of $G$, there is a domain $D$, contained in $[-O_{c,C}(1), O_{c,C}(1)]^m$, such that $G(L(\sum_{i=h-m+1}^{h} x_i v^{(i)}))$ is equal to zero unless $(x_{h-m+1}, \ldots, x_h)^T \in D$. (This is proved in full in Lemma 5.1).
We can use this observation to bound the right-hand side of \((8.5)\). Indeed, we have

\[
\beta \ll \mathrm{vol} \, D \times \sup_{x_{h-m+1} \in D} \frac{1}{N^{h-m}} \int_{x_{1}^{h-m} \in \mathbb{R}^{h-m}} F \left( \sum_{i=1}^{h} x_{i} v^{(i)} \right) G \left( L \left( \sum_{i=h-m+1}^{h} x_{i} v^{(i)} \right) \right) \left( \prod_{j=1}^{d} g_{j} \left( \Phi \left( x_{1}^{h-m} \right) \right) + \sum_{i=h-m+1}^{h} x_{i} v^{(i)} \right) d x_{1}^{h-m}. \tag{8.6}
\]

So there exists some fixed vector \((x_{h-m+1}, \ldots, x_{h})^{T}\) in \(D\) such that

\[
\beta \ll_{c,C,\varepsilon} \frac{1}{N^{h-m}} \int_{x_{1}^{h-m} \in \mathbb{R}^{h-m}} F \left( \sum_{i=1}^{h} x_{i} v^{(i)} \right) G \left( L \left( \sum_{i=h-m+1}^{h} x_{i} v^{(i)} \right) \right) \left( \prod_{j=1}^{d} g_{j} \left( \Phi \left( x_{1}^{h-m} \right) \right) + \sum_{i=h-m+1}^{h} x_{i} v^{(i)} \right) d x_{1}^{h-m}. \tag{8.7}
\]

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

\[
F_{1}(x_{1}^{h-m}) := F \left( \Phi \left( x_{1}^{h-m} \right) \right) + \sum_{i=h-m+1}^{h} x_{i} v^{(i)}
\]

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

\[
a_{j} := \xi_{j} \left( \sum_{i=h-m+1}^{h} x_{i} v^{(i)} \right) + \tilde{r}_{j}.
\]

Then

\[
\beta \ll_{c,C,\varepsilon} \left| \frac{1}{N^{h-m}} \int_{x \in \mathbb{R}^{h-m}} F_{1}(x) \prod_{j=1}^{d} g_{j} \left( \Phi(x) \right) + a_{j} \right| d x, \tag{8.8}
\]

and \(F_{1}\) and \(a_{j}\) satisfy the conclusions of the proposition.

Finally, since \(\mathrm{dist}(\Xi, L, V_{\text{deg}, 2}(m, d, h)) \geq c\) and \(\|\Xi\|_{\infty}, \|L\|_{\infty} \leq C\), Lemma 3.2 tells us that \(\Xi \Phi : \mathbb{R}^{h-m} \rightarrow \mathbb{R}^{d}\) satisfies \(\mathrm{dist}(\Xi \Phi, V_{\text{deg}, 2}(h-m, d)) \gg c, 1\). (One may consult Definitions 6.4 and Definition 14.1 for the definitions of \(V_{\text{deg}, 2}(h-m, d)\) and \(V_{\text{deg}, 2}(m, d, h)\). Thus, by Lemma 6.6 there exists some \(s\) at most \(d-2\) for which \(\Xi \Phi\) has \(\Omega_{c,C}(1)\)-Cauchy-Schwarz complexity at most \(s\).

Writing \(\Psi\) for \(\Xi \Phi\), the proposition is proved. \(\square\)

We now proceed to the second proposition, which is a standard Cauchy-Schwarz argument.

**Proposition 8.3** (Cauchy-Schwarz argument). Let \(s, d\) be natural numbers, with \(d \geq 3\), and let \(C\) be a positive constant. Let \(\sigma_{F}\) be a parameter in the range \(0 < \sigma_{F} < 1/2\). Let \((\psi_{1}, \ldots, \psi_{d}) = \Psi : \mathbb{R}^{s+1} \rightarrow \mathbb{R}^{d}\) be a linear map, and suppose that \(\psi_{1}(e_{k}) = 1\), for all the standard basis vectors \(e_{k} \in \mathbb{R}^{s+1}\). Suppose that, for all \(j\) in the range \(2 \leq j \leq s + 1\), there exists some \(k\) such that \(\psi_{j}(e_{k}) = 0\). Let \(N \geq 1\) be real, and let \(g_{1}, \ldots, g_{d} : [-N, N] \rightarrow [-1, 1]\) be arbitrary measurable functions, and, for each \(j\) at most \(d\), let \(a_{j}\) be some real number with \(|a_{j}| \leq CN\). Let \(F : \mathbb{R}^{s+1} \rightarrow [0, 1]\) be any
Lipschitz function, supported on $[-CN, CN]^{s+1}$ with Lipschitz constant $O(1/\sigma_FN)$. Suppose that $\|g_1\|_{U^{s+1}(\mathbb{R}, \mathbb{C})} \leq \rho$, for some parameter $\rho$ in the range $0 < \rho \leq 1$. Then
\[
\left| \frac{1}{N^{s+1}} \int_{\mathbb{R}^{s+1}} \prod_{j=1}^{d} g_j(\psi_1(w) + a_j) F(w) \, dw \right| \ll_C \rho^{-\Omega(1)}\sigma_F^{-1}.
\]
(8.9)

We stress again that implied constants may depend on the implicit dimensions (so the $\Omega(1)$ term in (8.9) may depend on $\sigma$).

**Proof.** This theorem is very similar to the usual Generalised von Neumann Theorem (see [24 Exercise 1.3.23]), and the proof is very similar too. A few extra technicalities arise from our dealing with the reals rather than with a finite group, but these are easily surmountable.

We begin with some simple reductions. First, we assume that $C$ is large enough in terms of all other $O(1)$ parameters. For notational convenience, we will also allow $C$ to vary form line to line. Next, since $\psi_1(w) = w_1 + w_2 + \cdots + w_{s+1}$, by shifting $w_1$ we can assume that $a_1 = 0$ in (8.9). Due to the restricted support of $F$, we may restrict the integral over $w$ to $[-CN, CN]^{s+1}$. By Lemma [34], for any $Y > 2$ there is a function $c_Y : \mathbb{R}^{s+1} \to \mathbb{C}$ satisfying $|c|_\infty \ll 1$ such that we may replace $F(w)$ by
\[
\int_{\theta \in \mathbb{R}^{s+1}, \|\theta\|_\infty \leq Y} c_Y(\theta) e(\theta \cdot w/N) \, d\theta + O_C \left( \frac{\log Y}{\sigma_F Y} \right).
\]

We will determine a particularly suitable $Y$ later (which will depend on $\rho$).

This means that
\[
\left| \frac{1}{N^{s+1}} \int_{\mathbb{R}^{s+1}} \prod_{j=1}^{d} g_j(\psi_1(w) + a_j) F(w) \, dw \right| \ll \int_{\theta \in \mathbb{R}^{s+1}, \|\theta\|_\infty \leq Y} \left| \frac{1}{N^{s+1}} \int_{\mathbb{R}^{s+1}} e(\theta \cdot w/N) \left( \prod_{j=1}^{d} g_j(\psi_1(w) + a_j) \right) \, dw \, d\theta \right| + O_C \left( \frac{\log Y}{\sigma_F Y} \right),
\]
(8.10)

where $\int^*$ indicates the limits $w \in [-CN, CN]^{s+1}$. Fix $\theta$. The inner integral of (8.10) will be our primary focus.

Firstly, we wish to ‘absorb’ the exponential phases $e(\theta \cdot w/N)$. To do this, we write $e(\theta \cdot w)$ as a product of functions $\prod_{k=1}^{s+1} b_k(w)$, where, for each $k$, the function $b_k : \mathbb{R}^{s+1} \to \mathbb{C}$ is bounded and does not depend on the variable $w_k$. Since $s + 1 \geq 2$, this is possible. Therefore we may rewrite the inner integral of (8.10) as
\[
\frac{1}{N^{s+1}} \int_{\mathbb{R}^{s+1}} g_1(\psi_1(w)) \prod_{k=1}^{s+1} b_k(w) \, dw,
\]
(8.11)

where the functions $b_k : \mathbb{R}^{s+1} \to \mathbb{C}$ are (possibly different) functions, satisfying $\|b_k\|_\infty \leq 1$ for all $k$, and such that $b_k$ does not depend on the variable $w_k$.

A brief aside: readers familiar with the arguments of [15, Appendix C] (which motivate the present proof) may note that a different device is used in that paper to absorb...
the exponential phases. Those authors work in the setting of the finite group \( \mathbb{Z}/N\mathbb{Z} \),
and there the exponential phases can be absorbed simply by twisting the functions
\( g_j : \mathbb{Z}/N\mathbb{Z} \to [-1, 1] \) by a suitable linear phase function (witness the discussion surrounding expression (C.7) from [15]).
The key point there is that, if the linear form \( w \mapsto \theta \cdot w \) fails to be in the set span(\( \psi_j : 1 \leq j \leq d \)),
then a Fourier expansion of \( g_j \) demonstrates that a certain expression, analogous to the inner integral of (8.10),
is equal to zero. This clean argument isn’t quite so easy to apply here, as the linear phases
are not integrable over all of \( \mathbb{R} \), which is why we choose a different approach.

Returning to (8.11), recall that \( \psi(w) = w_1 + w_2 + \cdots + w_{s+1} \). Therefore, applying
the Cauchy-Schwarz inequality in each of the variables \( w_1 \) through \( w_{s+1} \) in turn, one establishes that the absolute value of expression (8.11) is at most

\[
\ll C \left( \frac{1}{N^{2s+2}} \int_{w \in \mathbb{R}^{s+1}} \int_{z \in \mathbb{R}^{s+1}} \prod_{\alpha \in \{0,1\}^{s+1}} g_1 \left( \sum_{k \leq s+1} w_k + \sum_{k \leq s+1} z_k \right) dw \, dz \right)^{\frac{1}{2s+1}}. \tag{8.12}
\]

This expression may be immediately related to the real Gowers norm as given in Definition A.3,
by the change of variables \( m_k := z_k - w_k \), for all \( k \) at most \( s+1 \),
and \( u := w_1 + \cdots + w_{s+1} \). Performing this change of variables shows that (8.12) is

\[
\ll \left( \frac{1}{N^{2s+2}} \int_{(u,m,z^{s+1}) \in D} \prod_{\alpha \in \{0,1\}^{s+1}} g_1(u + \alpha \cdot m) \, du \, dm \, dz_2^{s+1} \right)^{\frac{1}{2s+1}}, \tag{8.13}
\]

where \( D \) is a convex domain contained within \([-CN,CN]^{2s+2} \). It remains to replace \( D \)
by a Cartesian box.

By Lemma [3.2] we may write

\[ 1_D = F_\sigma + O(G_\sigma), \]

for any \( \sigma \) in the range \( 0 < \sigma < 1/2 \), where \( F_\sigma, G_\sigma : \mathbb{R}^{2s+2} \to [0, 1] \) are Lipschitz
functions supported on \([-CN,CN]^{2s+2} \), with Lipschitz constant \( O_C(1/\sigma N) \), such that
\( \int_{\mathbb{R}^{2s+2}} G_\sigma(x) \, dx = O_C(\sigma N^{2s+2}) \). Then, since \( \|g_1\|_\infty \leq 1 \),
we may bound (8.13) above by

\[
\left( \frac{1}{N^{2s+2}} \int_{u,m,z^{s+1}} F_\sigma(u,m,z^{s+1}) \prod_{\alpha \in \{0,1\}^{s+1}} g_1(u + \alpha \cdot m) \, du \, dm \, dz_2^{s+1} + O_C(\sigma) \right)^{\frac{1}{2s+1}}, \tag{8.14}
\]

where \( \int^* \) now refers to the domain of integration \([-CN,CN]^{2s+2} \).

By applying Lemma [3.4] to \( F_\sigma \), for any \( X > 2 \) the absolute value of expression (8.14) is

\[
\ll C \left( \frac{1}{N^{2s+2}} \int_{\xi \in \mathbb{R}^{2s+2} \atop \|\xi\|_\infty \leq X} \int_{u,m,z^{s+1}} e\left( \frac{\xi \cdot (u,m,z^{s+1})}{N} \right) \prod_{\alpha \in \{0,1\}^{s+1}} g_1(u + \alpha \cdot m) \, du \, dm \, dz_2^{s+1} \right)^{\frac{1}{2s+1}} + O(\sigma) + O\left( \frac{\log X}{\sigma X} \right)^{\frac{1}{2s+1}}. \tag{8.15}
\]
Integrating over the variables $z_2, \ldots, z_{s+1}$, and splitting the exponential phase amongst the different functions, expression (8.15) is

$$\ll_C \left( \frac{1}{N^{s+2}} \int_{\xi \in \mathbb{R}^{2s+2}} \int_{\|\xi\|_{\infty} \leq X} \prod_{(u,m) \in [-CN,CN]^{s+2}} g_\alpha(u + \alpha \cdot m) \, du \, dm \right) \frac{1}{d\xi},$$

where each function $g_\alpha$ is of the form

$$g_\alpha(u) := g_1(u) e(k_\alpha u)$$

for some real $k_\alpha$. Note that $\|g_\alpha\|_{L^{s+1}([R,N])} = \|g_1\|_{L^{s+1}([R,N])}$.

Recall that $g_1$ is supported on $[-2N, 2N]$. Therefore, if $\prod_{\alpha \in \{0,1\}^{s+1}} g_\alpha(u + \alpha \cdot m) \neq 0$ then $(u, m) \in [-O(N), O(N)]^{s+2}$. So, if $C$ is large enough in terms of $s$, we may replace the restriction $(u, m) \in [-CN, CN]^{s+2}$ in (8.16) with the condition $(u, m) \in \mathbb{R}^{s+2}$, without changing the value of (8.16).

Then, by the Gowers-Cauchy-Schwarz inequality (Proposition A.4) and the triangle inequality, (8.10) is

$$\ll_C \left( X^{O(1)} \|g_1\|_{L^{s+1}([R,N])}^2 + \frac{\log X}{\sigma X} \right)^{\frac{1}{2s+1}},$$

Choosing $X = \rho^{-c_1}$, with $c_1$ suitably small in terms of $s$, and $\sigma = \rho^{c_1/2}$, expression (8.17) is $O_C(\rho^{O(1)})$.

Putting this estimate into (8.10), we get a bound on (8.10) of

$$\ll_C Y^{O(1)} \rho^{O(1)} + O\left( \frac{\log Y}{\sigma F Y} \right).$$

Picking $Y = \rho^{-c_1}$, with $c_1$ suitably small in terms of $s$, we may ensure that (8.18) is $O_C(\rho^{O(1)} \sigma F^{-1})$, thus proving the proposition.

With these propositions in hand, Theorem 8.1 follows quickly.

Proof of Theorem 8.1. Assuming all the hypotheses of Theorem 8.1, apply the result of Proposition 8.2 to $T_{F,G,N}(g_1, \ldots, g_d)$. Thus

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

where $\Psi : \mathbb{R}^{h-m} \rightarrow \mathbb{R}^d$ has $O_{c,C}(1)$-Cauchy-Schwarz complexity at most $s$, for some $s$ at most $d - 2$, $F_1 : \mathbb{R}^{h-m} \rightarrow [0,1]$ is a Lipschitz function supported on $[-O_{c,C,\varepsilon}(N), O_{c,C,\varepsilon}(N)]^{h-m}$ with Lipschitz constant $O(1/\sigma F N)$, and $a_j = O_{c,C,\varepsilon}(1)$.

We apply Proposition 6.7 to $\Psi$. Therefore, for any real numbers $w_1, \ldots, w_{s+1}$,

$$|\tilde{T}_{F,G,N}(g_1, \ldots, g_d)| \ll \left| \frac{1}{N^{h-m}} \int_{x \in \mathbb{R}^{h-m}} F_1(x + \sum_{k=1}^{s+1} w_k f_k) \prod_{j=1}^d g_j(\psi_j'(x, w) + a_j) \, dx \right|,$$

where

- for each $j$ at most $d$, $\psi_j' : \mathbb{R}^{h-m} \times \mathbb{R}^{s+1} \rightarrow \mathbb{R}$ is a linear form;
\( \psi'_i(0, w) = w_1 + \cdots + w_{s+1}; \)
\( f'_1, \cdots, f'_{s+1} \in \mathbb{R}^{h-m} \) are some vectors that satisfy \( \|f_k\|_\infty = O_{c,C}(1) \) for each \( k \) at most \( s+1; \)
\( \) the system of forms \((\psi'_1, \cdots, \psi'_d)\) is in normal form with respect to \( \psi'_i. \)

We remark that the right-hand side of expression (8.20) is independent of \( w, \) as it was obtained by applying the change of variables \( x \mapsto x + \sum_{k=1}^{s+1} w_k f_k \) to expression (8.19).

Now, let \( P : \mathbb{R}^{s+1} \to [0, 1] \) be some Lipschitz function, supported on \([-N, N]^{s+1}, \) with Lipschitz constant \( O(1/N). \) Also suppose that \( P(x) \equiv 1 \) if \( \|x\|_\infty \ll N/2. \) Integrating over \( w, \) we have that \( |\tilde{T}_{F,G,N}(g_1, \cdots, g_d)| \) is

\[
\ll_{c,C,\epsilon} \frac{1}{Nh-m+s+1} \left| \int_{x \in \mathbb{R}^{h-m}} \int_{w \in \mathbb{R}^{s+1}} P(w) \left| \int_{x \in \mathbb{R}^{h-m}} F_1(x + \sum_{k=1}^{s+1} w_k f_k) \prod_{j=1}^d g_j(\psi'_j(x, w) + a_j) \, dx \right| \, dw \right|,
\]

where the function \( H : \mathbb{R}^{h-m+s+1} \to [0, 1] \) is defined by

\[
H(x, w) := F_1(x + \sum_{k=1}^{s+1} w_k f_k) P(w).
\]

Since the vectors \( f_k \) satisfy \( \|f_k\|_\infty = O_{c,C}(1), \) \( H \) is a Lipschitz function supported on \([-O_{c,C,\epsilon}(N), O_{c,C,\epsilon}(N)]^{h-m+s+1}, \) with Lipschitz constant \( O_{c,C}(1/\sigma_F N). \) Notice 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 \( w \) (so in particular has constant sign).

Fix \( x. \) Then the integral over \( w \) in (8.21) satisfies the hypotheses of Proposition 8.3. Applying Proposition 8.3 to this integral, and then integrating over \( x, \) one derives

\[
|\tilde{T}_{F,G,N}(g_1, \cdots, g_d)| \ll_{c,C,\epsilon} \rho^{O(1)} \sigma_F^{-1}.
\]

Theorem 8.1 is proved. \( \square \)

By our long series of reductions, this means that both Theorem 4.6 and the main result (Theorem 2.10) are proved. \( \square \)

9. Constructions

In this section we prove Theorem 2.12, which, we remind the reader, is the partial converse of main result (Theorem 2.10). In other words, we show that \( L \) being bounded away from \( V_{d_{\text{degen}}}^*(m, d) \) is a necessary hypotheses for Theorem 2.10 to be true.

Proof of Theorem 2.12 Recall the hypotheses of Theorem 2.12. In particular, we suppose that

\[
\liminf_{N \to \infty} \text{dist}(L, V_{d_{\text{degen}}}^*(m, d)) = 0,
\]
i.e. we assume that \( \text{dist}(L, V_{d_{\text{degen}}}^*(m, d)) = \omega(N)^{-1}, \) for some function \( \omega(N) \) such that

\[
\limsup_{N \to \infty} \omega(N) = \infty.
\]
Let $\eta$ be a small positive quantity, picked small enough in terms of $c$ and $C$, and let $N$ be a natural number that is large enough so that $\omega(N) \geq \eta^{-1}$ and $\eta N \geq \max(1, \varepsilon)$. All implied constants to follow will be independent of $\eta$.

Since $F$ is the indicator function of $[1, N]^2$ and $G$ is the indicator function of $[-\varepsilon, \varepsilon]^m$, one has

$$T_{F,G,N}^\varepsilon(f_1, \ldots, f_d) = \frac{1}{N^{d-m}} \sum_{n \in [N]^d} \prod_{j=1}^d f_j(n_j).$$

Our aim is to construct functions $f_1, \ldots, f_d : [N] \to [-1, 1]$ such that

$$\min_j \|f_j\|_{U^{s+1}[N]} \leq \rho$$

for some $\rho$ at most 1 and that

$$T_{F,G,N}^\varepsilon(f_1, \ldots, f_d) > H(\rho) + E_\rho(N). \quad (9.1)$$

We begin by observing that the condition $\|L n\|_\infty \leq \varepsilon$ implies certain constraints on two of the variables $n_i$. Indeed, let $L' \in V_{\text{degen}}^*(m, d)$ be such that $\|L - L'\|_\infty = \text{dist}(L, V_{\text{degen}}^*(m, d))$. Write $\lambda'_{ij}$ for the coefficients of $L'$. By reordering columns, without loss of generality we may assume that there exist real numbers $\{a_i\}_{i=1}^m$ not all 0 s.t. for all $j$ in the range $3 \leq j \leq d$ we have

$$\sum_{i=1}^m a_i \lambda'_{ij} = 0, \quad (9.2)$$

and further we may assume that for all $i$ we have $\lambda'_{11} = \lambda_{i1}$ and $\lambda'_{22} = \lambda_{i2}$ (else $L' \in V_{\text{degen}}^*(m, d)$ is not one of the closest matrices to $L$). By reordering rows and rescaling, we may assume that $a_1$ has maximal absolute value amongst all the $a_i$, and that $|a_1| = 1$.

Define

$$b_1 := \sum_{i=1}^m a_i \lambda_{i1}, \quad b_2 := \sum_{i=1}^m a_i \lambda_{i2},$$

and let $n \in [N]^d$ be some solution to $\|L n\|_\infty \leq \varepsilon$. The critical observation is that (9.2), combined with the assumptions on the $a_i$, implies that

$$|b_1 n_1 + b_2 n_2| \ll \eta N. \quad (9.3)$$

Indeed, for $j$ in the range $3 \leq j \leq d$ we have

$$\left| \sum_{i=1}^m a_i \lambda_{ij} \right| = \left| \sum_{i=1}^m a_i (\lambda_{ij} - \lambda'_{ij}) \right| \ll \eta.$$

Since $\|L n\|_\infty \leq \varepsilon$, we certainly have that

$$\left| b_1 n_1 + b_2 n_2 + \sum_{j=3}^d n_j \sum_{i=1}^m a_i \lambda_{ij} \right| \ll \varepsilon,$$

and then (9.3) follows by the triangle inequality and the fact that $\eta N \geq \varepsilon$.

The constraint (9.3) will turn out to be enough for the proof. We consider various cases, constructing different counterexample functions $f_1$ and $f_2$ based on the size and sign of $b_1$ and $b_2$. To facilitate this, we let $c_1$ be a suitably small positive constant,
depending on $c$ and $C$, but independent of $\eta$. All constants $C_1$ and $C_2$ to follow will be assumed to satisfy $O_{c,C}(1)$.

**Case 1:** $|b_1|, |b_2| \leq c_1$.

Under the assumptions of Theorem 2.12 this case is actually precluded. Indeed, consider the matrix $L''$, defined by taking $\lambda''_{ij} = \lambda'_{ij}$ for all pairs $(i,j) \in [m] \times [d]$, except for $(1,1)$ and $(1,2)$. In these cases we let $\lambda''_{11} = \lambda'_{11} - \frac{b_1}{a_1}$ and $\lambda''_{12} = \lambda'_{12} - \frac{b_2}{a_1}$.

Then

$$\sum_{i=1}^{m} a_i \lambda''_{ij} = 0$$

for all $j$ in the range $1 \leq j \leq d$. In other words we have shown that $\|L - L''\|_\infty \leq \eta + c_1$ for some matrix $L''$ with rank less than $m$. Since $\eta + c_1 < c$ (if $c_1$ is small enough), this implies that $\text{dist}(L, V_{\text{rank}}(m, d)) < c$, which contradicts the assumptions of Theorem 2.12. Therefore this case is indeed precluded.

**Case 2:** $b_1, b_2$ both of the same sign, and $b_1, b_2 \geq c_1$.

In this case, \((9.3)\) implies\(^6\) that $n_1 \leq C_1 \eta N$ for some constant $C_1$. Now, define $f_1 : [N] \rightarrow [-1,1]$ to be the indicator function of the interval $[\lfloor C_1 \eta N \rfloor, N] \cap \mathbb{N}$. We then have

$$\|f_1 - 1\|_{U^{s+1}[N]} \ll \left( \frac{1}{N^{s+2}} \sum_{x,h_1,\ldots,h_{s+1} \leq C_1 \eta N} 1 \right) \frac{1}{2^{s+1}} \leq C_2 (C_1 \eta)^{\frac{s+2}{2^{s+1}}},$$

for some constant $C_2$. However, observe that

$$|T_{F,G,N}^L(f_1 - 1, 1, \cdots, 1)| = |T_{F,G,N}^L(f_1, 1, \cdots, 1) - T_{F,G,N}^L(1, 1, \cdots, 1)|$$

$$= |0 - T_{F,G,N}^L(1, 1, \cdots, 1)| \gg_c \varepsilon$$

by the hypotheses of Theorem 2.12. If $T_{F,G,N}^L(f_1 - 1, 1, \cdots, 1)$ did not satisfy \((9.1)\), then

$$1 \ll_c \varepsilon H(\rho) + E(1),$$

where $\rho := C_2 (C_1 \eta)^{\frac{s+2}{2^{s+1}}}$. Picking $\eta$ small enough, then $N$ large enough, this inequality cannot possibly hold, and we have a contradiction. So $T_{F,G,N}^L(f_1 - 1, 1, \cdots, 1)$ satisfies \((9.1)\).

**Case 3:** $b_1, b_2$ of opposite signs, and $b_1, b_2 \geq c_1$.

This is the most involved case, although the central idea is very simple. The condition \((9.3)\) confines $n_2$ to lie within a certain distance of a fixed multiple of $n_1$. By constructing functions $f_1$ and $f_2$ using random choices of blocks of this length, but

---

\(^6\) The same conclusion is true for $n_2$, but this will not be needed.
coupled in such a way that condition (9.3) is very likely to hold, we can guarantee that
\( T_{F,G,N}^L(f_1 - p, f_2 - p, 1, \cdots, 1) \) is bounded away from zero, where \( p \) is the probability
used to choose the random blocks. However, despite the block construction and the
coupling, the functions \( f_1 \) and \( f_2 \) still individually exhibit enough randomness to con-
clude that \( \| f_1 - p \|_{U_1[N]} = o(1) \), and the same for \( f_2 \).

We now fill in the technical details. Relation (9.3) implies that
\[
|b_1n_1 + b_2n_2| \leq C_1 \eta N,
\]
for some \( C_1 \) satisfying \( C_1 = O(1) \), and without loss of generality assume that \( b_1 \) is
positive, \( b_2 \) is negative, and \( |b_1| \) is at least \( |b_2| \). Let \( C_2 \) be some parameter, chosen so
that \( (C_1C_2\eta)^{-1} \) is an integer. Such a \( C_2 \) will of course depend on \( \eta \), but in magnitude
we may pick \( C_2 \simeq 1 \). We consider the real interval \([0, N]\) modulo \( N \), and for \( x \in [0, N] \)
and \( i \) in the range \( 0 \leq i \leq (C_1C_2\eta)^{-1} - 1 \) we define the half-open interval modulo \( N \)
\[
I_i := [x + iC_1C_2\eta N, x + (i + 1)C_1C_2\eta N).
\]
This choice guarantees that
\[
[0, N] = \bigcup_{i=0}^{(C_1C_2\eta)^{-1} - 1} I_i,
\]
and the union is disjoint. Now, for \( \delta \) a small constant to be chosen later, we define
\[
I_i^\delta := [x + (i + \frac{1}{2} - \delta)C_1C_2\eta N, x + (i + \frac{1}{2} + \delta)C_1C_2\eta N).
\]

We will use the partition (9.5) to construct a function \( f_1 \), using an averaging ar-
gument to choose an \( x \) so that the \( I_i^\delta \) intervals capture a positive proportion of the
solution density of the linear inequality system. Indeed, for \( n_1 \in [N] \) let the weight
\( u(n_1) \) denote the number of \( d - 1 \)-tuples \( n_2, \cdots, n_d \leq N \) that together with \( n_1 \) satisfy
the inequality \( \|Ln\|_\infty < \varepsilon \). The weight \( u(n_1) \) could be zero, of course. Let
\[
E_\delta := \cup_i I_i^\delta.
\]
Then
\[
\frac{1}{N} \int_0^N \sum_{n \in [N]} u(n)1_{E_\delta}(n) \, dx = \frac{1}{N} \sum_{n \in [N]} u(n) \int_0^N 1_{E_\delta}(n) \, dx
= \sum_{n \in [N]} u(n) 2\delta
= 2\delta N^{d-m}T_{F,G,N}^L(1, \cdots, 1)
\]
Therefore, by the assumptions of Theorem 2.12, we may fix an \( x \) such that
\[
\sum_{n \in [N]} u(n)1_{E_\delta}(n) \gg_{c, C} \delta N^{d-m}T_{F,G,N}^L(1, \cdots, 1).
\]

Let us finally define the function \( f_1 \). Let \( p \) be a small positive constant (to be
decided later). Fix a value of \( x \) such that (9.6) holds. Then we define a random subset
\( A \subseteq [N] \) by picking all of \( I_i \cap N \) to be members of \( A \), with probability \( p \), or none of
\( I_i \cap N \) to be members of \( A \), with probability \( 1 - p \). We then make this same choice for
each \( i \) in the range \( 0 \leq i \leq (C_1C_2\eta)^{-1} - 1 \), independently. Observe immediately that
\footnote{This \( \delta \) is unrelated to the notation \( \delta = T_{F,G,N}^L(f_1, \cdots, f_d) \) used in previous sections.}
for each $n \in [N]$ the probability that $n \in A$ is always $p$ (though these events are not always independent). We let $f_1(n)$ be the indicator function $1_A(n)$.

The function $f_2$ is defined in terms of $f_1$. Indeed, let

$$J_i = \frac{b_1}{|b_2|} I_i \cap (0, N],$$

where the dilation is not considered modulo $N$ but rather just as an operator on subsets of $\mathbb{R}$. Since $b_1 \geq |b_2|$ we have that these $J_i$ also form a disjoint partition of $[0, N]$.

[NB: If $b_1 > |b_2|$ it may be that certain $J_i$ are empty, since the dilate of the corresponding $I_i$ may land entirely outside $[0, N]$.] Then let $B$ be the subset of $[N]$ defined so that for each $i$ with $I_i$ non-empty we have $I_i \cap N \subseteq B$ if and only if $I_i \cap N \subseteq A$. Note again that for each individual $n \in [N]$ the probability that $n \in B$ is always $p$. We let $f_2(n)$ be the indicator function $1_B(n)$.

Our first claim is that, if $p$ is small enough in terms of $\delta$,

$$\mathbb{E}T^L_{F,G,N}(f_1, f_2, 1 \cdots, 1) - T^L_{F,G,N}(p, p, 1 \cdots, 1) \geq \epsilon \delta^2. \tag{9.7}$$

Indeed, suppose that $I_i$ is included in the set $A$, and suppose that $n_1 \in I_i^\delta$. If $n_2 \in [N]$ satisfies $\frac{b_1}{|b_2|} n_1 - n_2 \leq \frac{1}{b_2} C_1 \eta N$ and if $\delta$ is small enough in terms of $b_1$ and $b_2$, then $n_2 \in J_i$. Thus, by the observation $\{9.4\}$, $n_2 \in B$, for every integer $n_2$ that is the second coordinate of a solution vector\(^8\) for which the first coordinate is $n_1$. Therefore

$$\mathbb{E}T^L_{F,G,N}(f_1, f_2, 1 \cdots, 1) = \frac{1}{N^{d-m}} \sum_{\|L_n\|_{\infty} \leq \epsilon} \mathbb{P}(n_1 \in A \wedge n_2 \in B)$$

$$\geq \frac{1}{N^{d-m}} \sum_{\|L_n\|_{\infty} \leq \epsilon} \mathbb{P}(n_1 \in A \wedge n_1 \in I_i^\delta \text{ for some } i \wedge n_2 \in B)$$

$$\geq \frac{1}{N^{d-m}} \sum_{\|L_n\|_{\infty} \leq \epsilon} \mathbb{P}(n_1 \in A \wedge n_1 \in I_i^\delta \text{ for some } i)$$

$$= \frac{1}{N^{d-m}} \sum_{n_1 \in [N]} u(n_1) p 1_{E_3}(n_1)$$

$$\geq 2 \delta p T^L_{F,G,N}(1, \cdots, 1),$$

where the final line follows from $\{9.6\}$. On the other hand $T^L_{F,G,N}(p, p, 1 \cdots, 1) = p^2 T^L_{F,G,N}(1, \cdots, 1)$, and hence

$$\mathbb{E}T^L_{F,G,N}(f_1, f_2, 1 \cdots, 1) - T^L_{F,G,N}(p, p, 1 \cdots, 1) \geq (2 \delta p - p^2) T^L_{F,G,N}(1, \cdots, 1). \tag{9.8}$$

Picking $p$ small enough in terms of $\delta$, and using the assumption that $T^L_{F,G,N}(1, \cdots, 1) = \Omega_{\epsilon,C} \delta(1)$, this proves the relation $\{9.7\}$.

Our second claim is that

$$\mathbb{E}\|f_1 - p\|_{U^{s+1}[N]}, \mathbb{E}\|f_2 - p\|_{U^{s+1}[N]} \ll \eta^{\frac{1}{2s+1}}. \tag{9.9}$$

\(^8\)This fact is the reason why we introduced the parameter $\delta$.

\(^9\)i.e a vector $n$ such that $\|L n\|_{\infty} \leq \epsilon$. 
we may factor the expectation and conclude that these random variables are independent. Hence, apart from those exceptional cases, \( x \) mean zero and, unless some two of the expressions \( x + h \cdot \omega \) lie in the same block \( I_i \), these random variables are independent. Hence, apart from those exceptional cases, we may factor the expectation and conclude that

\[
\mathbb{E}\left( \prod_{\omega \in \{0,1\}^{s+1}} (f_1 - p_{1[N]})(x + h \cdot \omega) \right) = \prod_{\omega \in \{0,1\}^{s+1}} \mathbb{E}(f_1 - p_{1[N]})(x + h \cdot \omega) = 0.
\]

Therefore,

\[
\mathbb{E}\|f_1 - p\|^2_{U^{s+1}|N} \ll \frac{1}{N^{s+2}} \sum_{(x,h) \in \mathbb{Z}^{s+2}} 1_R(h) \ll \eta,
\]

where

\[
R = \{ h : |h \cdot (\omega_1 - \omega_2)| \leq C_1 C_2 \eta N \text{ for some } \omega_1, \omega_2 \in \{0,1\}^{s+1}, \omega_1 \neq \omega_2 \}.
\]

Thus by Jensen’s inequality we have

\[
\mathbb{E}\|f_1 - p\|_{U^{s+1}|N} \ll \eta^{\frac{1}{s+2}}, \tag{9.10}
\]

as claimed in (9.9).

The calculation for \( f_2 \) is essentially identical, noting that the length of the blocks \( J_i \) is also \( O(\eta N) \).

It is possible that one could finish the argument here by considering a second moment, and choosing some explicit \( f_1 \) and \( f_2 \). To avoid calculating a second moment, we argue as follows. Suppose for contradiction that there were no functions \( f_1, \ldots, f_d \) that satisfied (9.11). Then, by (9.7), if we pick \( p \) to be small enough in terms of \( \delta \) we have

\[
\delta^2 \ll_{C_1, C_2} |\mathbb{E}T^I_{F,G,N}(f_1, f_2, 1, \ldots, 1) - T^I_{F,G,N}(p, p, 1 \ldots, 1)| \ll |\mathbb{E}T^I_{F,G,N}(f_1 - p, f_2, 1, \ldots, 1) + |\mathbb{E}T^I_{F,G,N}(p, f_2 - p, 1, \ldots, 1)| \ll \mathbb{E}(H(\rho_1) + E_{\rho_1}(N)) + \mathbb{E}(H(\rho_2) + E_{\rho_2}(N)), \tag{9.11}
\]

where \( \rho_1 \) (resp. \( \rho_2 \)) is any chosen upper-bound on \( \|f_1 - p\|_{U^{s+1}|N} \) (resp. \( \|f_2 - p\|_{U^{s+1}|N} \)). Note that the values \( \rho_1 \) may be random variables themselves.

We claim that the random variables \( \rho_1 \) and \( \rho_2 \) may be chosen so that the right-hand side of (9.11) is \( \kappa(\eta) + o_1(1) \). To prove this, we make two observations. Note first that by Markov’s inequality

\[
\mathbb{P}(\|f_1 - p\|_{U^{s+1}|N} \geq \eta^{\frac{1}{s+2}}) \ll \eta^{\frac{1}{2s+2}}.
\]

We choose the (random) upper-bound \( \rho_1 \) satisfying

\[
\rho_1 = \begin{cases} 1 & \text{if } \|f_1 - p\|_{U^{s+1}|N} \geq \eta^{\frac{1}{2s+2}} \\ \eta^{\frac{1}{2s+2}} & \text{otherwise} \end{cases}.
\]

Secondly, we may upper-bound \( H \) by a concave envelope, so without loss of generality we may assume that \( H \) is concave.
Then by Jensen’s inequality,
\[
\mathbb{E}(H(\rho_1) + E_{\rho_1}(N)) \ll H(\mathbb{E}\rho_1) + \mathbb{E}(E_{\rho_1}(1))
\]
\[
\ll \kappa(\eta^{2^{\frac{1}{2}+\varepsilon} + 1}) + o(1)
\]
\[
\ll \kappa(\eta) + o(1).
\]  
\tag{9.12}

We do the same manipulation for \(f_2\). Combining (9.12) with (9.11) we conclude that
\[
\delta^2 \ll_{c,c,\varepsilon} \kappa(\eta) + o(1).
\]  
\tag{9.13}

The only condition on \(\delta\) occurred in the proof of (9.7), in which we assumed that \(\delta\) was small enough in terms of \(b_1\) and \(b_2\). Therefore there exists a suitable \(\delta\) that satisfies \(\delta = \Omega_{c,c}(1)\). Picking such a \(\delta\), and then picking \(\eta\) small enough and \(N\) large enough, (9.13) is a contradiction. So there must be some functions \(f_1, \ldots, f_d\) that satisfy (9.1).

**Case 4: Exactly one of \(b_1, b_2\) satisfies \(b_i \geq c_1\).**

Without loss of generality we may assume that \(b_1 \geq c_1\). But then, as in Case 2, (9.3) implies that \(n_1 \leq C_1 \eta N\) for some constant \(C_1\). The same construction as in Case 2 then applies.

We have covered all cases, and thus have concluded the proof of Theorem 2.12. \qed
Appendix A. Gowers norms

There are several existing accounts of the basic theory of Gowers norms – for example in [12] and [24] – and the reader looking for an introduction to the theory in its full generality should certainly consult these references, as well as Appendices B and C of [15]. However, in the interests of making this paper as self-contained as possible, we use this section to pick out the central definitions and notions that will be used in the main text.

Definition A.1. Let $N$ be a natural number. For a function $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$, and a natural number $d$, we define the Gowers $U^d$ norm $\|f\|_{U^d(N)}$ to be the unique non-negative solution to

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

where $|\omega| = \sum_i \omega_i$, $h = (h_1, \ldots, h_d)$, $\mathcal{C}$ is the complex-conjugation operator, and the summation is over $x, h_1, \ldots, h_d \in \mathbb{Z}/N\mathbb{Z}$.

For example,

$$\|f\|_{U^1(N)} = \left( \frac{1}{N} \sum_x |f(x)| \right)^\dagger,$$

and

$$\|f\|_{U^2(N)} = \left( \frac{1}{N^3} \sum_{x,h_1,h_2} f(x) \overline{f(x+h_1)} f(x+h_2) f(x+h_1+h_2) \right)^\dagger.$$

It is not immediately obvious that the right-hand side of (A.1) is always a non-negative real, nor why the $U^d$ norms are genuine norms if $d \geq 2$: proofs of both these facts may be found in [25]. An immediate Cauchy-Schwarz argument, which may also be found in [25], gives the so-called ‘nesting property’ of Gowers norms, namely the fact that

$$\|f\|_{U^d(N)} \leq \|f\|_{U^{d+1}(N)} \leq \|f\|_{U^{d+2}(N)} \leq \cdots.$$

The functions in the main text do not have a cyclic group as a domain but rather the interval $[N]$, but the theory may easily be adapted to this case.

Definition A.2. Let $N, N'$ be natural numbers, with $N' \geq N$. Identify $[N]$ with a subset of $\mathbb{Z}/N'\mathbb{Z}$ in the natural way, i.e. $[N] = \{1, \ldots, N\} \subseteq \{1, \ldots, N'\}$, which we then view as $\mathbb{Z}/N'\mathbb{Z}$. For a function $f : [N] \to \mathbb{C}$, and a natural number $d$, we define the Gowers norm $\|f\|_{U^d[N]}$ to be the unique non-negative real solution to the equation

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

where $f_{1[N]}$ is the extension by zero of $f$ to $\mathbb{Z}/N'\mathbb{Z}$, the summation is over $x, h_1, \ldots, h_d \in \mathbb{Z}/N'\mathbb{Z}$, and the set $R$ is the set

$$R := \{x, h_1, \ldots, h_d \in \mathbb{Z}/N'\mathbb{Z} : \text{for every } \omega \in \{0,1\}^d, x + h \cdot \omega \in [N]\}.$$

One can immediately see that this definition is equivalent to

$$\|f\|_{U^d[N]} = \|f_{1[N]}\|_{U^d(N')}/|1[N]|_{U^d(N')}$$

and is also independent of the choice of $N'$ as long as $N'/N$ is large enough (in terms of $d$). Taking $N' = O(N)$ we have $|1[N]|_{U^d(N')} \asymp 1$, and thus $\|f\|_{U^d[N]} \asymp \|f_{1[N]}\|_{U^d(N')}$. (See [15] Lemma B.5 for more detail on this).
We observe that there is only a contribution to the summand in equation (A.2) when \( x \in [N] \) and for every \( i \) we have \( h_i \in \{-N, -N + 1, \cdots, N - 1, N\} \) modulo \( N' \). Further, it may be easily seen that \( |R| \approx N^{d+1} \). Therefore, choosing \( N'/N \) sufficiently large, we conclude that

\[
\|f\|_{U^d([N])} \approx \left( \frac{1}{N^{d+1}} \sum_{x,h_1,\ldots,h_d \in \mathbb{Z}} \prod_{\omega \in \{0,1\}^d} C^{[\omega]} f(x + \mathbf{h} \cdot \mathbf{\omega}) \right)^{\frac{1}{2^d}}. \tag{A.3}
\]

The relation (A.3) is implicitly assumed throughout the main text.

In order to succinctly state Theorem 8.1, we had to refer to a Gowers norm \( U^d(\mathbb{R}) \), which has been used in some recent work on linear patterns in subsets of Euclidean space (see [5, Lemma 4.2], [8, Proposition 3.3]). This Gowers norm is a less well-studied object, as the theory was originally developed over finite groups. Nevertheless it may be perfectly well defined, and even deep aspects of its inverse theory may be deduced from the corresponding theory of the discrete Gowers norm (see [23]).

**Definition A.3.** Let \( f : [0,1] \to \mathbb{C} \) be a bounded measurable function, and let \( d \) be a natural number. Then we define the Gowers norm \( \|f\|_{U^d(\mathbb{R})} \) to be the unique non-negative real satisfying

\[
\|f\|_{U^d(\mathbb{R})}^2 = \int \prod_{(x,\mathbf{h}) \in \mathbb{R}^{d+1}} C^{[\omega]} f(x + \sum_{i=1}^d h_i \omega_i) \, dx \, dh_1 \cdots dh_d \tag{A.4}
\]

where \(|\omega| = \sum \omega_i\), and \( C \) is the complex-conjugation operator.

Let \( N \) be a positive real, and let \( g : [-N,N] \to \mathbb{C} \) be a measurable function. Define the function \( f : [0,1] \to \mathbb{C} \) by \( f(x) := g(2Nx - N) \), and then set

\[
\|g\|_{U^d([N],\mathbb{R})} := \|f\|_{U^d(\mathbb{R})}.
\]

Explicitly, a change of variables shows that

\[
\|g\|_{U^d([N],\mathbb{R})} \approx \frac{1}{N^{d+1}} \int \prod_{(x,\mathbf{h}) \in \mathbb{R}^{d+1}} C^{[\omega]} g(x + \sum_{i=1}^d h_i \omega_i) \, dx \, dh_1 \cdots dh_d. \tag{A.5}
\]

We require one further fact about Gowers norms.

**Proposition A.4** (Gowers-Cauchy-Schwarz inequality). Let \( d \) be a natural number, and, for each \( \omega \in \{0,1\}^d \), let \( f_\omega : [0,1] \to \mathbb{C} \) be a bounded measurable function. Define the Gowers inner-product

\[
\langle (f_\omega)_{\omega \in \{0,1\}^d}\rangle := \int \prod_{(x,\mathbf{h}) \in \mathbb{R}^{d+1}} C^{[\omega]} f_\omega(x + \sum_{i=1}^d h_i \omega_i) \, dx \, dh_1 \cdots dh_d.
\]

Then

\[
|\langle (f_\omega)_{\omega \in \{0,1\}^d}\rangle| \leq \prod_{\omega \in \{0,1\}^d} \|f_\omega\|_{U^d(\mathbb{R})}.
\]

**Proof.** See [24, Chapter 11] for the proof in the finite group setting. The modification to the setting of the reals is trivial. \( \square \)
Appendix B. Lipschitz functions

In the body of the paper we made extensive use of properties of Lipschitz functions.

Definition B.1 (Lipschitz functions). We say that a function \( F : \mathbb{R}^m \rightarrow \mathbb{C} \) is Lipschitz, with Lipschitz constant at most \( M \), if

\[
M \geq \sup_{x, y \in \mathbb{R}^m \atop x \neq y} \frac{|F(x) - F(y)|}{\|x - y\|_\infty}.
\]

We say that a function \( G : \mathbb{R}^m/\mathbb{Z}^m \rightarrow \mathbb{C} \) is Lipschitz, with Lipschitz constant at most \( M \), if

\[
M \geq \sup_{x, y \in \mathbb{R}^m/\mathbb{Z}^m \atop x \neq y} \frac{|G(x) - G(y)|}{\|x - y\|_{\mathbb{R}^m/\mathbb{Z}^m}}.
\]

We record the three properties of Lipschitz functions that we will require.

Lemma B.2. Let \( N \) be a positive real, let \( m \) be a natural number, let \( K \) be a convex subset of \([-N, N]^m\), and let \( \sigma \) be some parameter in the range \( 0 < \sigma < 1/2 \). Then there exist Lipschitz functions \( F_\sigma, G_\sigma : \mathbb{R}^m \rightarrow [0, 1] \) supported on \([-2N, 2N]^m\), both with Lipschitz constant at most \( O\left(\frac{1}{\sigma N}\right)\), such that

\[
1_K = F_\sigma + O(G_\sigma)
\]

and \( \int x G_\sigma(x) \, dx = O(\sigma N^m) \). Furthermore, \( F_\sigma(x) \geq 1_K(x) \) for all \( x \in \mathbb{R}^m \), and \( G \) is supported on \( \{ x \in \mathbb{R}^m : \text{dist}(x, \partial(K)) \leq \sigma N \} \).

This is [15] Corollary A.3. It was be used in Lemmas 4.9 and 4.11 to replace sums with sharp cut-offs with sums with Lipschitz cut-offs.

Lemma B.3. Let \( X \) be a positive real, with \( X > 2 \). Let \( F : \mathbb{R}^m/\mathbb{Z}^m \rightarrow \mathbb{C} \) be a Lipschitz function such that \( \|F\|_\infty \leq 1 \) and the Lipschitz constant of \( F \) is at most \( M \). Then

\[
F(x) = \sum_{k \in \mathbb{Z}^m} c_X(k)e(k \cdot x) + O\left(M \frac{\log X}{X}\right) \tag{B.1}
\]

for every \( x \in \mathbb{R}^m/\mathbb{Z}^m \), for some function \( c_X(k) \) satisfying \( \|c_X(k)\|_\infty \ll 1 \). (The implied constant in the error term above may depend on the underlying dimensions, as always in this paper).

This is [14] Lemma A.9, and was used in Lemma \( \mathbf{3.3} \) as a way of bounding the number of solutions to a certain inequality.

Lemma B.4. Let \( X, N, C \) be positive reals, with \( X > 2 \) and \( N > 1 \). Let \( F : \mathbb{R}^m \rightarrow \mathbb{C} \) be a Lipschitz function, supported on \([-CN, CN]^m\), such that \( \|F\|_\infty \leq 1 \) and the Lipschitz constant of \( F \) is at most \( M \). Then

\[
F(x) = \int_{\xi \in \mathbb{R}^m} c_X(\xi)e\left(\frac{\xi \cdot x}{N}\right) d\xi + O\left(MN \frac{\log X}{X}\right) \tag{B.2}
\]

for every \( x \in \mathbb{R}^m \), for some function \( c_X(\xi) \) satisfying \( \|c_X(\xi)\|_\infty \ll C \cdot 1 \).

Lemma \( \mathbf{B.3} \) is very similar to Lemma \( \mathbf{B.3} \) and may be easily proved by adapting that standard harmonic analysis argument found in [14] Lemma A.9 from \( \mathbb{R}^m/\mathbb{Z}^m \) to \( \mathbb{R}^m \). For completeness, we sketch the proof.
Sketch proof. By rescaling the variable $x$ by a factor of $N$, we reduce to the case where $F$ is supported on $[-C, C]^m$ and has Lipschitz constant at most $MN$.

Let 

$$K_X(x) := \prod_{i=1}^{m} \frac{1}{X} \left( \frac{\sin(\pi x \xi_i)}{\pi x_i} \right)^2.$$ 

Then 

$$\hat{K}_X(\xi) = \prod_{i=1}^{m} \max(1 - \frac{\xi_i}{X}, 0).$$

We have 

$$(F * K_X)(x) = \int_{\xi \in \mathbb{R}^m} \hat{F}(\xi) \hat{K}_X(\xi) e(\xi \cdot x) \, dx,$$ 

and, since $|\hat{F}(\xi)| \leq \|F\|_1 \ll 1$, letting $c_X(\xi) := \hat{F}(\xi) \hat{K}_X(\xi)$ gives a main term of the desired form.

It remains to show that 

$$\|F - F * K_X\|_{\infty} \ll MN \frac{\log X}{X}.$$ 

By writing 

$$|F(x) - (F * K_X)(x)| = \left| \int_{y \in \mathbb{R}^m} (F(x) - F(y)) K_X(x - y) \, dy \right|,$$ 

one sees that it suffices to show that 

$$\int_{\|z\|_{\infty} \leq 2C} \frac{\log X}{X} \, dz \ll C.$$

But this bound follows immediately from a dyadic decomposition. \qed 

We used Lemma B.4 extensively in the Generalised von Neumann Theorem argument in section 8.

**Appendix C. Rank matrix and normal form: proofs**

In this appendix we prove the two quantitative statements from earlier in the paper, namely Propositions 3.1 and 6.7.

**Proposition C.1.** Let $n$ be a natural number, and let $S = \{f_1, \ldots, f_k\}$ be a finite set of continuous functions $f_1, \ldots, f_k : \mathbb{R}^n \to \mathbb{R}$. Let 

$$V(S) = \{x \in \mathbb{R}^n : f_i(x) = 0 \text{ for all } i \leq k\}.$$ 

Suppose that $x \in \mathbb{R}^n$ is a point with $\|x\|_{\infty} \leq C$ and with $\text{dist}(x, V(S)) \geq c$, for some absolute positive constants $c$ and $C$. Then, there is some $f_j$ such that $|f_j(x)| = \Omega_{c, C, S}(1)$.

**Proof.** This is nothing more than the fact that every continuous function on a compact set is bounded, applied to the continuous function $\min(1/|f_1|, \ldots, 1/|f_k|)$ and the compact set $\{x \in \mathbb{R}^n : \|x\|_{\infty} \leq C, \text{dist}(x, V(S)) \geq c\}$. \qed 

From Proposition C.1 it is easy to deduce the existence of rank matrices, namely Proposition 3.1.
Proof of Proposition 3.1\ Let \( k \) be equal to \( (d \choose m) \), and identify \( \mathbb{R}^{md} \) with the space of \( m \)-by-\( d \) real matrices. Then let \( f_1, \ldots, f_k \) be the \( k \) polynomials on \( \mathbb{R}^{md} \) that are given by the \( k \) determinants of \( m \)-by-\( m \) submatrices of \( L \). One then sees that \( V_{\text{rank}}(m, d) \) is exactly the set of common zeros of the functions \( f_i \). This is since row rank equals column rank, and linear independence of columns in a square matrix can be detected by the determinant.

Since we assume that \( \|L\|_\infty \leq C \) and \( \text{dist}(L, V_{\text{rank}}(m, d)) \geq c \) we can fruitfully apply Proposition C.1 to deduce that there is some \( j \) for which \( |f_j(L)| = \Omega_{c,C}(1) \). The matrix \( M \) whose determinant corresponds to the polynomial \( f_j \) is exactly the claimed rank matrix.

This settles the first part of Proposition 3.1. The second part then follows immediately by the construction of \( M^{-1} \) as the adjugate matrix of \( M \) divided by \( \det M \).

The third part, namely the statement about linear combinations of rows, follows quickly from the others. Indeed, without loss of generality, assume that the rank matrix \( M \) is realised by columns 1 through \( M \). Then, the fact that the rows of \( L \) are linearly independent means that there are unique real numbers \( a_i \) such that \( \sum_{i=1}^{m} a_i \lambda_{ij} = v_j \) for all \( j \) in the range \( 1 \leq j \leq d \). (Recall that \( (\lambda_{ij})_{i \leq m, j \leq d} \) denotes the coefficients of \( L \).)

Restricting to \( j \) in the range \( 1 \leq j \leq m \), we observe that the \( a_i \) are forced to satisfy

\[
\begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} = (M^T)^{-1} \begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix}.
\]

Since \( \|(M^{-1})^T\|_\infty = \|M^{-1}\|_\infty = O_{c,C}(1) \), we conclude that \( a_i = O_{c,C;C}(1) \) for all \( i \).

The final part of the proposition is to show that if \( \text{dist}(L, V_{\text{rank}}(m, d)) \geq c \) then, for each \( j \), there exists a rank matrix of \( L \) that doesn’t include the \( j \)th column. But this statement follows immediately from the above, after having deleted the \( j \)th column. \( \square \)

We now prove Proposition 6.7 on the existence of quantitative normal form parametrisations. We remind the reader that, in the proof, the implied constants may depend on the dimensions of the underlying spaces, namely \( m \) and \( n \). For the definition of the variety \( V_{P_i} \), which consists of all systems of linear forms for which the partition \( P_i \) is not ‘suitable’, the reader may consult Definition 6.4. The reader may also find the example that follows the proof to be informative.

Proof of Proposition 6.7\ Fix \( i \), and let \( P_i \) be a partition of \( [m] \setminus \{i\} \) such that \( \text{dist}(\Psi, V_{P_i}) \geq c_1 \) (such a \( P_i \) exists by the definition of \( c_1 \)-Cauchy-Schwarz complexity, i.e. by Definition 6.5). The partition \( P_i \) has \( s_i + 1 \) parts, for some \( s_i \) at most \( s \).

It is clear from Definition 6.5 that we may, without loss of generality, further subdivide the partition and assume that the partition \( P_i \) has exactly \( s + 1 \) parts. Call the parts \( C_1 \) through \( C_{s+1} \).

Following section 4 of [15], for each \( k \in [s + 1] \) there exists a vector \( f_k \in \mathbb{R}^n \) that witnesses the fact that \( \text{dist}(\Psi, V_{P_i}) > 0 \), i.e. for which \( \psi_i(f_k) = 1 \) but \( \psi_j(f_k) = 0 \) for all \( j \in C_k \). Such a vector can be found using Gaussian elimination, say. Consider the extension

\[
\Psi'(u, w_1, \ldots, w_{s+1}) := \Psi(u + w_1 f_1 + \cdots + w_{s+1} f_{s+1}).
\]
Then, if \( \Psi' = (\psi'_1, \ldots, \psi'_s) \), the form \( \psi'_i(u, w_1, \ldots, w_{s+1}) \) is the only one that uses all of the \( w_k \) variables. Furthermore, \( \psi'_i(0, w) = w_1 + \cdots + w_{s+1} \). Also, \( s' = n + s + 1 \), which is at most \( n + m - 1 \). So Proposition [6.7] is proved if for each \( k \) we can find such a vector \( \mathbf{f}_k \) that additionally satisfies \( \|\mathbf{f}_k\|_\infty = O_{c_1, c_1}(1) \).

Consider a fixed \( k \), and let \( \Gamma \) be the set of possible implementations of Gaussian elimination on the set of forms \( \psi_i \cup \{ \psi_j : j \in \mathcal{C}_k \} \) to find a solution vector \( \mathbf{f}_k \). If in the course of implementing these algorithms we are given a free choice for a co-ordinate of \( \mathbf{f}_k \), we set it to be equal to zero. Note that \( |\Gamma| = O(1) \).

Now, for each \( \gamma \in \Gamma \), let the rational functions
\[
\frac{p_{\gamma,1}(\Psi)}{q_{\gamma,1}(\Psi)}, \ldots, \frac{p_{\gamma,n}(\Psi)}{q_{\gamma,n}(\Psi)}
\]
be the \( n \) rational functions defining the claimed coefficients of \( \mathbf{f}_k \). One may assume without loss of generality that, for all \( j \), we have \( p_{\gamma,j}, q_{\gamma,j} \in \mathbb{Z}[X_1, \ldots, X_n] \) with coefficients of size \( O(1) \). Now let
\[
Q_{\gamma} := \prod_{j \leq n} q_{\gamma,j}.
\]
We claim that \( V(I) \subseteq V_{\mathcal{P}_i} \), where \( I \) is the ideal generated by the set of polynomials \( \{ Q_{\gamma} : \gamma \in \Gamma \} \) and \( V(I) \) is the affine algebraic variety generated by \( I \). Indeed, if \( Q_{\gamma}(\Psi) = 0 \) for all \( \gamma \in \Gamma \) then there is no Gaussian elimination implementation that finds a solution \( \mathbf{f}_k \), and this in turn implies that \( \mathcal{P}_i \) is not suitable for \( \Psi \) and hence that \( \Psi \notin V_{\mathcal{P}_i} \).

Since \( V(I) \subseteq V_{\mathcal{P}_i} \), the assumptions of Proposition [6.7] imply that \( \text{dist}(\Psi, V(I)) \geq c_1 \).

Applying Proposition [C.1] to the polynomials \( \{ Q_{\gamma} : \gamma \in \Gamma \} \), we conclude that there is some \( \gamma \in \Gamma \) such that \( |Q_{\gamma}(\Psi)| = O_{c_1, c_1}(1) \). In particular, we conclude that the solution vector \( \mathbf{f}_k \) obtained by the implementation \( \gamma \) has coefficients that are \( O_{c_1, c_1}(1) \). This concludes the proof of Proposition [6.7]. \( \square \)

Let us illustrate the above proof with an instructive example. Consider \( n = 3, m = 2, i = 1 \), and denote
\[
\Psi = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}.
\]
Then the partition \( \mathcal{P}_i \) consists of the singleton \( \{2\} \), and suppose one wished to construct a suitable \( \mathbf{f}_1 \) simply by applying Gaussian elimination. Implementing the algorithm in a certain way we have
\[
\mathbf{f}_1 = \begin{pmatrix} a_{22}/(a_{11}a_{22} - a_{12}a_{21}) \\ -a_{21}/(a_{11}a_{22} - a_{12}a_{21}) \\ 0 \end{pmatrix}
\]
as a solution, in the case where \( a_{11}a_{22} - a_{12}a_{21} \) is non-zero. Of course if \( a_{11}a_{23} - a_{13}a_{21} \) is non-zero too, we have another solution
\[
\mathbf{f}_1 = \begin{pmatrix} a_{23}/(a_{11}a_{23} - a_{13}a_{21}) \\ -a_{21}/(a_{11}a_{23} - a_{13}a_{21}) \\ 0 \end{pmatrix}.
\]
So, if one applied Gaussian elimination idly, one might end up with either of these two solutions. Unfortunately it could be the case that \( \text{dist}(\Psi, V_{\mathcal{P}_i}) \geq c_1 \) whilst one of these determinants, \( a_{11}a_{22} - a_{12}a_{21} \) say, was non-zero yet \( o(1) \) (as the unseen variable \( N_i \) on which \( \Psi \) will ultimately depend, tends to infinity). In this instance, applying the first implementation of the algorithm would not give a desirable solution vector.
appropriate vector $f_1$. For this reason we had to apply somewhat indirect arguments in order to find the appropriate vector $f_1$.

It is worth including a brief discussion on why these quantitative subtleties do not arise in the setting of [15]. Indeed, assume that $\Psi$ has rational coefficients of naive height at most $C_1$ and that $\Psi \notin V_{R_1}$. Since there are only $O(1)$ many possible choices of $\Psi$ we immediately conclude that $\text{dist}(\Psi, V_{R_1}) \gg C_1$, without needing to assume this as an extra hypothesis. Then any implementation of Gaussian elimination succeeds in finding a suitably bounded $f_k$, since one is only ever working with rationals of bounded height.

**Appendix D. Additional linear algebra**

In this appendix, we collect together the assortment of standard linear algebra lemmas that we used at various points throughout the paper. We also give the linear algebra argument that we used to construct the matrix $P$ during the proof of Lemma 4.10.

This first lemma demonstrates the intuitive fact, that if $L : \mathbb{R}^d \rightarrow \mathbb{R}^m$ is a linear map then $L : (\ker L)^\perp \rightarrow \mathbb{R}^m$ has bounded inverse.

**Lemma D.1.** Let $m, d$ be natural numbers, with $d \geq m + 1$, and let $c, C, l$ be positive constants. Let $L : \mathbb{R}^d \rightarrow \mathbb{R}^m$ be a surjective linear map, and suppose $\|L\|_\infty \leq C$ and $\text{dist}(L, V_{\text{rank}(m, d)}) \geq c$. Let $K$ denote $\ker L$. Let $R$ be a convex set contained in $[-l, l]^m$. Then, if $v \in K^\perp$, $Lv \in R$ only when $v \in R'$, where $R'$ is some convex region that satisfies $R' \subseteq [-O_{c, C}(l), O_{c, C}(l)]^d$.

**Proof.** We choose to prove this statement using the concept of the 'rank matrix' introduced earlier. Writing $L$ as a $m$-by-$d$ matrix with respect to the standard bases, let $\lambda_i \in \mathbb{R}^d$ denote the column vector such that $\lambda_i^T$ is the $i$th row of $L$. Since $\text{dist}(L, V_{\text{rank}(m, d)}) \geq c$, the vectors $\lambda_i$ are linearly independent. Moreover, we may extend the set $\{\lambda_i : i \leq m\}$ by orthogonal vectors of unit length to form a basis $\{\lambda_i : i \leq d\}$ for $\mathbb{R}^d$.

We claim that for all $k \in [d]$ we have

$$\sum_{i=1}^{d} a_{ki} \lambda_i = e_k,$$

for some coefficients $a_{ki}$ satisfying $|a_{ki}| = O_{c, C}(1)$, where $e_k \in \mathbb{R}^d$ is the $k$th standard basis vector. Indeed, fix $k$, and note that $e_k = x_k + y_k$, where $x_k \in \text{span}(\lambda_i : i \leq m)$ and $y_k \in \text{span}(\lambda_i : m + 1 \leq i \leq d)$. The vectors $x_k$ and $y_k$ are orthogonal by construction, so in particular $\|x_k\|_2^2 + \|y_k\|_2^2 = 1$, and hence $\|x_k\|_\infty, \|y_k\|_\infty \leq 1$. By the third part of Proposition 3.1 applied to $x_k$ we get $|a_{ki}| = O_{c, C}(1)$ when $i \leq m$, and the orthogonality of $\{\lambda_i : m + 1 \leq i \leq d\}$ implies that $|a_{ki}| = O(1)$ when $i$ is in the range $m + 1 \leq i \leq d$.

Now notice that $\text{span}(\lambda_i : m + 1 \leq i \leq d)$ is exactly equal to $K$. Let $v \in K^\perp$, and suppose $Lv \in R$. Letting $L'$ be the $d$-by-$d$ matrix whose rows are $\lambda_i^T$, we have that $L'v = w$ for some vector $w$ satisfying $\|w\|_\infty \ll l$. Pre-multiplying by the matrix $A = (a_{ki})$, we immediately get $v = Aw$, and hence $\|v\|_\infty = O_{c, C}(l)$. The region $R' := (L^{-1}R) \cap K^\perp$ is therefore bounded. $R'$ is clearly convex, and so the lemma is proved.\qed
The second lemma concerns vectors, with integer coordinates, that lie near to a subspace.

Lemma D.2. Let \( h, d \) be natural numbers, with \( h \leq d \), and let \( C, \eta \) be positive reals.
Let \( \Xi : \mathbb{R}^h \to \mathbb{R}^d \) be an injective linear map, with \( \|\Xi\|_\infty \leq C \). Suppose further that \( \Xi(\mathbb{Z}^h) = \mathbb{Z}^d \cap \Xi(\mathbb{R}^h) \). Let \( n, r \in \mathbb{Z}^d \). Suppose that

\[
\text{dist}(n, \Xi(\mathbb{R}^h) + r) \leq \eta. \tag{D.1}
\]

Then, if \( \eta \) is small enough in terms of \( C, h \) and \( d \), \( n = \Xi(m) + r \), for some unique \( m \in \mathbb{Z}^h \).

Proof. By replacing \( n \) with \( n - r \), we can assume without loss of generality that \( r = 0 \). It will also be enough to show that \( n \in \Xi(\mathbb{R}^h) \), as the injectivity of \( \Xi \) and the assumption that \( \Xi(\mathbb{Z}^h) = \mathbb{Z}^d \cap \Xi(\mathbb{R}^h) \) immediately go on to imply the existence of a unique \( m \).

Suppose for contradiction then that \( n \notin \Xi(\mathbb{R}^h) \). In matrix form, \( \Xi \) is a \( d \)-by-\( h \) matrix with linearly independent columns, all of whose coefficients are integers with absolute value at most \( C \). We can extend this matrix to a \( d \)-by-\( d \) matrix \( \tilde{\Xi} \), with linearly independent columns, all of whose coefficients are integers with absolute value at most \( C \). Then \( (\tilde{\Xi})^{-1} \) is a \( d \)-by-\( d \) matrix with rational coefficients of naive height at most \( O(C^1) \), and \( (\tilde{\Xi})^{-1}(\Xi(\mathbb{R}^h)) = \mathbb{R}^h \times \{0\}^{d-h} \).

Since \( n \notin \Xi(\mathbb{R}^h) \), we have \( (\tilde{\Xi})^{-1}(n) \notin \mathbb{R}^h \times \{0\}^{d-h} \). But \( (\tilde{\Xi})^{-1}(n) \in \frac{1}{h}\mathbb{Z}^d \), for some natural number \( K \) satisfying \( K = O(C^1) \). Therefore

\[
\text{dist}( (\tilde{\Xi})^{-1}(n), (\tilde{\Xi})^{-1}(\Xi(\mathbb{R}^h))) \gg C^{-O(1)}.
\]

Applying \( \tilde{\Xi} \), we conclude that

\[
\text{dist}(n, \Xi(\mathbb{R}^h)) \gg C^{-O(1)},
\]

which is a contradiction to (D.1) if \( \eta \) is small enough. \( \square \)

The construction of the matrix \( \tilde{\Xi} \) in the above proof also has an even more basic consequence, namely that \( \Xi^{-1} : \im \Xi \to \mathbb{R}^h \) is bounded.

Lemma D.3. Let \( h, d \) be natural numbers, with \( h \leq d \), and let \( C, \eta \) be positive reals.
Suppose that \( \Xi : \mathbb{R}^h \to \mathbb{R}^d \) is an injective linear map, with \( \|\Xi\|_\infty \leq C \). Suppose further that \( \Xi(\mathbb{Z}^h) \subseteq \mathbb{Z}^d \cap \Xi(\mathbb{R}^h) \). Then if \( \|\Xi(y)\|_\infty \leq \eta \), we have \( \|y\|_\infty \ll C^{-O(1)} \eta \).

Proof. Construct the matrix \( \tilde{\Xi} \) as in the previous proof. Then \( \| (\tilde{\Xi})^{-1}(\Xi(y)) \|_\infty \ll C^{-O(1)} \eta \), by the bound on the size of the coefficients of \( \tilde{\Xi} \). But \( (\tilde{\Xi})^{-1}(\Xi(y)) \in \mathbb{R}^d \) is nothing more than the vector \( y \in \mathbb{R}^h \) extended by zeros. So \( \|y\|_\infty \ll C^{-O(1)} \eta \) as claimed. \( \square \)

Finally, we give the linear algebra argument used to construct the matrix \( P \) during the proof of Lemma 4.10.

Lemma D.4. Let \( m, d \) be natural numbers, with \( d \geq m + 1 \). Let \( L : \mathbb{R}^d \to \mathbb{R}^m \) be a surjective linear map with rational dimension \( u \), and let \( \Theta : \mathbb{R}^m \to \mathbb{R}^u \) be a rational map for \( L \). Suppose that \( \|L\|_\infty \leq C \) and \( \|\Theta\|_\infty \leq C \). Equating \( L \) with its matrix, suppose that the first \( m \) columns of \( L \) form the identity matrix. Let \( \{a_1, \ldots, a_u\} \) be a basis for the lattice \( \Theta L(\mathbb{Z}^d) \) that satisfies \( \|a_i\|_\infty = O_C(1) \) for every \( i \). Let \( x_1, \ldots, x_u \in \mathbb{Z}^d \) be vectors such that, for every \( i \), \( \Theta L(x_i) = a_i \) and \( \|x_i\|_\infty = O_C(1) \). Then

\[
\mathbb{R}^m = \text{span}(Lx_i : i \leq u) \oplus \ker \Theta \tag{D.2}
\]
and there is an invertible linear map \( P : \mathbb{R}^m \to \mathbb{R}^m \) such that

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

\[
P(\ker \Theta) = \{0\}^u \times \mathbb{R}^{m-u},
\]

and both \( \|P\|_\infty = O_C(1) \) and \( \|P^{-1}\|_\infty = O_C(1) \).

Note that both \( \{a_1, \ldots, a_u\} \) and \( x_1, \ldots, x_u \in \mathbb{Z}^d \) exist by applying Lemma 4.7 to the map \( S := \Theta L \).

**Proof.** The expression (D.2) is immediate from the definitions, so it remains to construct \( P \). We may assume, since the first \( m \) columns of \( L \) form the identity matrix, that \( \Theta \) has integer coefficients.

As \( \|\Theta\|_\infty = O_C(1) \), we may pick a basis \( \{y_1, \ldots, y_{m-u}\} \) for \( \ker \Theta \) in which \( y_j \in \mathbb{Z}^m \) and \( \|y_j\|_\infty = O_C(1) \) for all \( j \). Let \( b_1, \ldots, b_m \) denote the standard basis of \( \mathbb{R}^m \), and define \( P \) by letting

\[
P(Lx_i) := b_i, \quad 1 \leq i \leq u,
\]

\[
P(y_j) := b_{j-u}, \quad 1 \leq j \leq m-u,
\]

and then extending linearly to all of \( \mathbb{R}^m \). Clearly \( P((\text{span}(Lx_i : i \leq u))) = \mathbb{R}^u \times \{0\}^{m-u} \) and \( P(\ker \Theta) = \{0\}^u \times \mathbb{R}^{m-u} \). It is also immediate that \( \|P^{-1}\|_\infty = O_C(1) \), since \( \|Lx_i\|_\infty = O_C(1) \) and \( \|y_j\|_\infty = O_C(1) \) for all \( i \) and \( j \). It remains to bound \( \|P\|_\infty \). If \( Lx_i \) were all vectors with integer coordinates then this bound would be immediate as well, as then \( P^{-1} \) would have integer coordinates and hence \( |\det P^{-1}| \geq 1 \). As it is, we have to proceed more slowly.

To this end, for a standard basis vector \( b_k \) write

\[
b_k = \sum_{i=1}^u \lambda_i Lx_i + \sum_{j=1}^{d-u} \mu_j y_j.
\]

It will be enough to show that \( |\lambda_i|, |\mu_j| = O_C(1) \) for all \( i \) and \( j \). First note that, since the first \( m \) columns of \( L \) form the identity, \( b_k \in L(\mathbb{Z}^d) \). Also \( \Theta(b_k) = \sum_{i=1}^u \lambda_i a_i \). So \( a := \sum_{i=1}^u \lambda_i a_i \) is an element of \( \Theta L(\mathbb{Z}^d) \) that satisfies \( \|a\|_\infty = O_C(1) \). Since \( \|a_i\|_\infty = O_C(1) \) for every \( i \), and \( \{a_1, \ldots, a_u\} \) is a basis for the lattice \( \Theta L(\mathbb{Z}^d) \), this implies that \( |\lambda_i| = O_C(1) \) for every \( i \).

So then \( \sum_{j=1}^{d-u} \mu_j y_j \) is a vector in \( \ker \Theta \) satisfying \( \|\sum_{j=1}^{d-u} \mu_j y_j\|_\infty = O_C(1) \). Since \( \{y_1, \ldots, y_{m-u}\} \) is a set of linearly independent vectors, each of which has integer coordinates with absolute value \( O_C(1) \), this implies that \( |\mu_j| = O_C(1) \) for every \( j \).

Therefore \( P \) satisfies the conclusions of the lemma.

**Remark D.5.** We note the effects of the above construction in the case when \( L \) has algebraic coefficients. We use a rudimentary version of height: if \( Q \in \mathbb{Z}[X] \) we define

\[
H(Q) := \max(|q_i| : q_i \text{ a coefficient of } Q)
\]

to be the *height* of \( Q \), and we say that the height of an algebraic number is the height of its minimal polynomial. (So there are \( O_{k,H}(1) \) algebraic numbers of degree at most \( k \) and height at most \( H \).) Then, if in the statement of Lemma D.4 all the coefficients of \( L \) are algebraic numbers with degree at most \( k \) and height at most \( H \), all the coefficients of \( P \) are algebraic numbers of degree \( O_k(1) \) and height \( O_{C,k,H}(1) \).
Appendix E. The approximation function in the algebraic case

We use this final appendix to give the proof of relation \([2,3]\). The following lemma makes this relation quantitatively precise.

**Lemma E.1.** Let \(m, d\) be natural numbers, with \(d \geq m + 1\), and let \(c, C\) be positive constants. Let \(L : \mathbb{R}^d \rightarrow \mathbb{R}^m\) be a surjective linear map, and suppose that the matrix of \(L\) has algebraic coefficients of algebraic degree at most \(k\) and algebraic height at most \(H\) (see Remark \([D,3]\) for definitions). Suppose that \(\|L\|_\infty \leq C\), that \(\text{dist}(L, V_{\text{rank}}(m, d)) \geq c\), and that \(L\) has rational complexity at most \(C\). Let \(\tau_1, \tau_2\) be two parameters in the range \(0 < \tau_1, \tau_2 \leq 1\). Then

\[
A_L(\tau_1, \tau_2) \gg_{k,H,c,C} \min(\tau_1, \tau_2^{O_k(1)}).
\]

**Proof.** We begin by reducing to the case when \(L\) is purely irrational. Indeed, consider Lemma \([4,10]\) and replace \(L\) by the map \(L'\) (expression \([4,10]\)). By part (9) of Lemma \([4,10]\), \(A_L(\tau_1, \tau_2) \ll_{c,C} A_L(\Omega_{c,C}(\tau_1), \Omega_{c,C}(\tau_2))\). Also, using Remark \([D,5]\) it follows that \(L'\) has algebraic coefficients of algebraic degree at most \(O_k(1)\) and algebraic height at most \(O_{c,C,k,H}(1)\). So, replacing \(L\) with \(L'\), without loss of generality we may assume that \(L\) is purely irrational.

Suppose for contradiction that for all choices of constants \(c_1\) and \(C_2\), there exist parameters \(\tau_1\) and \(\tau_2\) such that \(A_L(\tau_1, \tau_2) < c_1 \min(\tau_1, \tau_2^{C_2})\), i.e. there exists a map \(\alpha \in (\mathbb{R}^m)^*\) and a map \(\varphi \in (\mathbb{Z}^d)^T\) such that \(\tau_1 \leq \|\alpha\|_\infty \leq \tau_2^{-1}\) and

\[
\|L^*\alpha - \varphi\|_\infty < c_1 \min(\tau_1, \tau_2^{C_2}). \tag{E.1}
\]

Fix \(\alpha\) and \(\varphi\) so that they satisfy \([E.1]\). We will obtain a contradiction if \(c_1\) is small enough in terms of \(c, C, k, H\), and if and \(C_2\) is large enough in terms of \(k\).

In the first part of the proof, we apply various reductions to enable us to replace \(\alpha\) with a map that has integer coordinates with respect to the standard dual basis of \((\mathbb{R}^m)^*\).

Let \(M\) be a rank matrix of \(L\) (Proposition \([3,3]\)), and assume without loss of generality that \(M\) consists of the first \(m\) columns of \(L\). Then there exists a map \(\beta \in (\mathbb{R}^m)^*\), namely \(\beta := M^*\alpha\), such that \(\tau_1 \ll_{c,C} \|\beta\|_\infty \ll_{c,C} \tau_2^{-1}\) and

\[
\|L^*(M^{-1})^*\beta - \varphi\|_\infty < c_1 \min(\tau_1, \tau_2^{C_2}). \tag{E.2}
\]

Since the first \(m\) columns of \(M^{-1}L\) form the identity matrix, \([E.2]\) implies that

\[
\text{dist}(\beta, (\mathbb{Z}^m)^T) < c_1 \min(\tau_1, \tau_2^{C_2}) \tag{E.3}
\]

We know that \(\|\beta\|_\infty = \Omega_{c,C}(\tau_1)\). Also, considering \([E.3]\), by perturbing \(\beta\) by a suitable element \(\gamma \in (\mathbb{R}^m)^*\) with \(\|\gamma\|_\infty < c_1 \min(\tau_1, \tau_2^{C_2})\) we may obtain a map \(\rho \in (\mathbb{Z}^m)^T\). Combining these facts, note how

\[
\|\rho\|_\infty \geq \|\beta\|_\infty - c_1 \min(\tau_1, \tau_2^{C_2}) \gg_{c,C} \tau_1
\]

if \(c_1\) is small enough, and so certainly \(\rho \neq 0\).

From \([E.2]\), we therefore conclude that there exists some \(\rho \in (\mathbb{Z}^m)^T \setminus \{0\}\), satisfying \(\|\rho\|_\infty = O_{c,C}(\tau_2^{-1})\), such that

\[
\|L^*(M^{-1})^*\rho - \varphi\|_\infty < c_1 C_2 \tau_2^{C_2} \tag{E.4}
\]
where $C_3$ is some constant that depends on $c$ and $C$. Referring back to (E.1), we see that we have achieved our goal of replacing $\alpha$ with a map that has integer coefficients.

Expression (E.4) leads to a contradiction. Morally this follows from Liouville’s theorem on the diophantine approximation of algebraic numbers, but we couldn’t find exactly the statement we needed in the literature, so we include a short argument here.

Indeed, let $\varphi = (\varphi_1 \cdots \varphi_d)$ be the representation of $\varphi$ with respect to the standard dual basis of $(\mathbb{R}^d)^*$ (with analogous notation for $L^* (M^{-1})^* \rho$). Since $L$ is assumed to be purely irrational, so is $M^{-1} L$. Therefore, since $\rho : \mathbb{R}^m \to \mathbb{R}$ is surjective (since it is non-zero), we may pick some co-ordinate $i$ at most $d$ for which $(L^* (M^{-1})^* \rho)_i - \varphi_i \neq 0$.

So there are algebraic numbers $\lambda_1, \cdots, \lambda_m$ with algebraic degree $O_k(1)$ and algebraic height $O_{c,C,k,H}(1)$ for which

$$0 < |\sum_{j=1}^m \lambda_j \rho_j - \varphi_i| < c_1 C_3 \tau_2^{C_2}, \quad \text{(E.5)}$$

where $(\rho_1 \cdots \rho_m)$ is the representation of $\rho$ with respect to the standard dual basis.

Note that if $c_1$ is small enough, by (E.5) and the fact that $\|\rho\|_\infty = O_{c,C}(\tau_2^{-1})$ one has $|\varphi_i| = O_{c,C}(\tau_2^{-1})$.

Our aim will be to find a suitable polynomial $Q$ for which $Q(\sum_{j=1}^m \lambda_j \alpha_j) = 0$, and then to apply Liouville’s original argument.

Assume without loss of generality that each $\lambda_j \rho_j$ is non-zero. For each $j$ at most $m$, let $Q_j \in \mathbb{Z}[X]$ denote the minimal polynomial of $\lambda_j \rho_j$. Note that the degree of $Q_j$ is $O_k(1)$ (since $\rho_j \in \mathbb{Z}$). By the bounds on the degree and height of $\lambda_j$, and since $\|\rho\|_\infty = O_{c,C}(\tau_2^{-1})$, we have $H(Q_j) = O_{c,C,k,H}(\tau_2^{-O_k(1)})$.

By using the standard construction based on resultants (see [4 section 4.2.1]), this implies that there is a polynomial $Q \in \mathbb{Z}[X]$ with degree $O_k(1)$ such that $Q(\sum_{j=1}^m \lambda_j \rho_j) = 0$ and $H(Q) = O_{c,C,k,H}(\tau_2^{-O_k(1)})$.

Now, it could be that $\varphi_i$ is a root of $Q$. If this is the case, we use the factor theorem and Gauss’ Lemma to replace $Q$ by the integer-coefficient polynomial $Q \cdot (X - \varphi_i)^{-1}$. In this case, $H(Q \cdot (X - \varphi_i)^{-1}) \ll c_{c,C,k,H} \tau_2^{-O_k(1)}$. By repeating this process as necessary, since $|\varphi_i| = O_{c,C}(\tau_2^{-1})$ we may assume therefore that $\varphi_i$ is not a root of $Q$ and that there exists a constant $C_L$ depending on $L$ such that $H(Q) = O_{c,C,k,H}(\tau_2^{-O_k(1)})$.

This immediately implies a bound on the derivative of $Q$, namely that, for any $\theta$,

$$|Q'(\theta)| \ll c_{c,C,k,H} \tau_2^{-O_k(1)} \sum_{0 \leq a \leq O_k(1)} \theta^a.$$ 

But then the mean value theorem implies that for some $\theta$ in the interval $[\sum_j \lambda_j \alpha_j, \varphi_i]$ one has

$$1 \leq |Q(\varphi_i)| = |Q(\sum_{j=1}^m \lambda_j \rho_j) - Q(\varphi_i)| \leq |Q'(\theta)||\sum_{j=1}^m \lambda_j \rho_j - \varphi_i| \ll c_{c,C,k,H} c_1 C_3 \tau_2^{-O_k(1)} \tau_2^{C_2}.$$
If $C_2$ is large enough in terms of $k$, this implies that $c_1 = \Omega_{c,C,k,H}(1)$, which is a contradiction if $c_1$ is small enough. Therefore the lemma holds. 

\[ \square \]

**References**

[1] A. Baker. On some diophantine inequalities involving primes. *J. Reine Angew. Math.*, 228:166–181, 1967.

[2] R. C. Baker. *Diophantine inequalities*, volume 1 of *London Mathematical Society Monographs. New Series*. The Clarendon Press, Oxford University Press, New York, 1986. Oxford Science Publications.

[3] J. Bourgain. A quantitative Oppenheim theorem for generic diagonal quadratic forms. *Israel J. Math.*, 215(1):503–512, 2016.

[4] H. Cohen. *A course in computational algebraic number theory*, volume 138 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, 1993.

[5] B. Cook, Á. Magyar, and M. Pramanik. A Roth-type theorem for dense subsets of $\mathbb{R}^d$. *Bulletin of the London Mathematical Society*, 49(4):676–689, 2017.

[6] H. Davenport. *Analytic methods for Diophantine equations and Diophantine inequalities*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, second edition, 2005. With a foreword by R. C. Vaughan, D. R. Heath-Brown and D. E. Freeman, Edited and prepared for publication by T. D. Browning.

[7] H. Davenport and H. Heilbronn. On indefinite quadratic forms in five variables. *J. London Math. Soc.*, 21:185–193, 1946.

[8] P. Durcik, V. Kovač, and L. Rimanić. On side lengths of corners in positive density subsets of the Euclidean space. *International Mathematics Research Notices*, page rnx093, 2017.

[9] D. E. Freeman. Asymptotic lower bounds and formulas for Diophantine inequalities. In *Number theory for the millennium, II (Urbana, IL, 2000)*, pages 57–74. A K Peters, Natick, MA, 2002.

[10] W. T. Gowers. A new proof of Szemerédi’s theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001.

[11] W. T. Gowers and J. Wolf. The true complexity of a system of linear equations. *Proc. Lond. Math. Soc. (3)*, 100(1):155–176, 2010.

[12] B. Green. Montréal notes on quadratic Fourier analysis. In *Additive combinatorics*, volume 43 of *CRM Proc. Lecture Notes*, pages 69–102. Amer. Math. Soc., Providence, RI, 2007.

[13] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. *Ann. of Math. (2)*, 167(2):481–547, 2008.

[14] B. Green and T. Tao. Quadratic uniformity of the Möbius function. *Ann. Inst. Fourier (Grenoble)*, 58(6):1863–1935, 2008.

[15] B. Green and T. Tao. Linear equations in primes. *Ann. of Math. (2)*, 171(3):1753–1850, 2010.

[16] B. Green and T. Tao. The Möbius function is strongly orthogonal to nilsequences. *Ann. of Math. (2)*, 175(2):541–566, 2012.

[17] B. Green, T. Tao, and T. Ziegler. An inverse theorem for the Gowers $U^{2+L}[N]$-norm. *Ann. of Math. (2)*, 176(2):1231–1372, 2012.

[18] G. A. Margulis. Discrete subgroups and ergodic theory. In *Number theory, trace formulas and discrete groups (Oslo, 1987)*, pages 377–398. Academic Press, Boston, MA, 1989.

[19] R. McCutcheon. FVIP systems and multiple recurrence. *Israel J. Math.*, 146:1231–1372, 2012.

[20] T. Tao. Higher order Fourier analysis, volume 142 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.

[21] T. Tao and V. Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.

[22] R. C. Vaughan. *The Hardy-Littlewood method*, volume 125 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, second edition, 1997.

[23] A. Walker. Linear inequalities in primes. In preparation.
[28] T. D. Wooley. On Diophantine inequalities: Freeman’s asymptotic formulae. In *Proceedings of the Session in Analytic Number Theory and Diophantine Equations*, volume 360 of *Bonner Math. Schriften*, page 32. Univ. Bonn, Bonn, 2003.

Aled Walker, Trinity College, Cambridge, United Kingdom, CB2 1TQ
aw530@cam.ac.uk