Algebra & Number Theory

Volume 14

2020

No. 6

Gowers norms control diophantine inequalities

Aled Walker

msp

ALGEBRA AND NUMBER THEORY 14:6 (2020)
https://doi.org/10.2140/ant.2020.14.1457

Gowers norms control diophantine inequalities

Aled Walker

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 $\|f\|_{U^{s+1}[N]}$ 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. Furthermore, we provide a necessary and sufficient condition for a system of real linear forms to be governed by Gowers norms in this way. We present applications to cancellation of the Möbius function over certain sequences.

The machinery developed in this paper can be adapted to 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. Substantial extra difficulties occur in this setting, however, and we have prepared a separate paper on these issues.

1.	Introduction	1458
2.	The main theorem	1467
3.	Upper bounds	1480
4.	Normal form	1485
5.	Dimension reduction	1489
6.	Transfer from $\mathbb{Z}$ to $\mathbb{R}$	1504
7.	Degeneracy relations	1508
8.	A generalised von Neumann theorem	1509
9.	Constructions	1517
Appendix A.	Gowers norms	1523
Appendix B.	Lipschitz functions	1525
Appendix C.	Rank matrix and normal form: proofs	1527
Appendix D.	Additional linear algebra	1530
Appendix E.	The approximation function in the algebraic case	1532
Acknowledgements		1535
References		1535

MSC2010: primary 11D75; secondary 11B30, 11J25.
Keywords: Gowers norms, diophantine inequalities, Möbius orthogonality, generalised von Neumann theorem.

1. Introduction

The field of *diophantine inequalities* is a large, and somewhat loosely defined, collection of problems which lie at the intersection of traditional number theory and diophantine approximation. As far as this paper is concerned, we will restrict our attention to the following class of questions. Let A be a set of integers, let ε be a positive parameter, and let L be an m-by-d real matrix. One might then ask whether there are infinitely many solutions to

$$\|L\boldsymbol{a}\|_\infty \leqslant \varepsilon \tag{1-1}$$

in which all of the coordinates of $\boldsymbol{a}$ lie in A. Further, letting N be an integer parameter tending to infinity, one might seek an asymptotic formula for the number of such solutions $\boldsymbol{a}$ which lie in the box given by the condition $\|\boldsymbol{a}\|_\infty \leqslant N$. One might even try to count solutions in certain cases in which L depends on N.

Much is known about these problems for certain special sets A (see [Baker 1967; 1986; Davenport and Heilbronn 1946; Margulis 1989; Müller 2005; Parsell 2002a; 2002b]), in particular for the image sets of polynomials. This work is discussed in Section 1A below. However, as far we are aware, the inequality (1-1) has not been considered before in such generality. Naturally there are some advantages and some disadvantages in pursuing such a general formulation, the main disadvantage being that the statements of our main results must perforce include some complicated technical hypotheses on the matrix L.

It will take us the rest of Sections 1 and 2 to properly motivate these hypotheses, culminating in the statement of Theorem 2.12 (our main theorem). Section 1 will focus on qualitative results and applications, whereas Section 2 goes on to explore the issues of diophantine approximation and nondegeneracy which are required for a quantitative treatment when L depends on N. At the end of Section 2 we will give a detailed sketch of our entire proof strategy. For now, we present the reader with a certain corollary of our main theorem, which we hope will encourage further reading through this long introduction.

Corollary 1.1 (example of Möbius orthogonality). *Let $\theta_1, \dots, \theta_s \in \mathbb{R}$ be distinct irrational numbers, let N be an integer parameter, and let $f_1, f_2, \dots, f_{s+1} : \{1, \dots, N\} \to [-1, 1]$ be arbitrary 1-bounded functions. Then*

$$\frac{1}{N^2} \sum_{\substack{x,d\in\mathbb{Z} \\ 1\leqslant x\leqslant N}} \mu(x) f_1(x+d)\Bigl(\prod_{i=2}^{s+1} f_i([x+\theta_{i-1}d])\Bigr) = o(1) \tag{1-2}$$

as $N \to \infty$, where μ denotes the Möbius function and $[x] := \lfloor x + \frac{1}{2} \rfloor$ is the nearest integer to x. The $o(1)$ error term may depend on the numbers $\theta_1, \dots, \theta_s$ but is independent of the choice of functions $f_1, \dots, f_{s+1}$.

1A. ***Classical results.*** As we said above, much is known about the inequality (1-1) for certain special sets A, particularly when $m = 1$. For example, if A is the set of squares, it was shown by Davenport and Heilbronn [1946] that there are infinitely many solutions to (1-1) for $m = 1$ and $d = 5$, i.e., infinitely many solutions to

$$|\lambda_1 n_1^2 + \lambda_2 n_2^2 + \lambda_3 n_3^2 + \lambda_4 n_4^2 + \lambda_5 n_5^2| \leqslant \varepsilon, \tag{1-3}$$

provided the coefficients λ_i are nonzero, 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 $2^k + 1$. Some 55 years after Davenport and Heilbronn, Freeman [2002] refined the analysis from [Davenport and Heilbronn 1946] to obtain asymptotic formulas for the number of solutions to (1-3) in which $n_i \leqslant N$ for every i, and he also reduced the number of variables required in the case of k-th powers, to $k(\log k + \log\log k + O(1))$. Wooley [2003] further reduced this number, particularly for small k.

The Davenport–Heilbronn method is Fourier-analytic. One begins by replacing the interval $[-\varepsilon, \varepsilon]$ with a Lipschitz cutoff function, and then one expresses the solution count via the Fourier inversion formula (see [Davenport 1963, Chapter 20] or [Vaughan 1981, Chapter 11]). The device of replacing $[-\varepsilon, \varepsilon]$ with a friendlier cutoff plays an important role in our argument too, and we discuss it at length in Section 2E.

There are also some results on the inequality (1-1) when $m \geqslant 2$, although this setting has been studied less intensively. For example, Parsell [2002b] considered the setting of k-th powers, with Müller [2005] developing a refined result in the case of inequalities for general real quadratics. Parsell's result is rather technical to state, and we defer the interested reader to the original paper. Later on in Section 2, however, we will state Müller's result precisely, as one of his hypotheses is closely related to a hypothesis in our main theorem.

One of our main goals, for this paper and for our follow-up [Walker 2019], is to find a method of proving asymptotic formulae for the number of solutions to diophantine inequalities which goes beyond what can be done using the Davenport–Heilbronn method. Of particular interest to us is the case of inequality (1-1) when A is the set of prime numbers. A result first claimed by A. Baker [1967][1] states that for any fixed positive ε there exist infinitely many triples of primes (p_1, p_2, p_3) satisfying

$$|\lambda_1 p_1 + \lambda_2 p_2 + \lambda_3 p_3| \leqslant \varepsilon, \tag{1-4}$$

assuming again that the coefficients λ_i are nonzero, not all of the same sign, and not all in pairwise rational ratio. Parsell [2002a] then used a similar refinement to that of Freeman to prove a lower bound on the number of solutions to (1-4) in the box $p_1, p_2, p_3 \leqslant N$. For m simultaneous inequalities, and for a generic matrix L, Parsell's method is powerful enough[2] to prove an asymptotic formula for the number of solutions to (1-1) in prime variables $p_1, p_2, \dots, p_d \leqslant N$, provided that $d \geqslant 2m + 1$. In [Walker 2019], building on the work of the present paper, we manage to reach the same conclusion under the weaker hypothesis that $d \geqslant m + 2$, provided that L has algebraic coefficients.

[1] In fact Baker [1967] proved a slightly different result, writing that the result we quote here followed easily from the then-existing methods.

[2] This does not seem to be present in the literature except in an appendix of our paper [Walker 2019].

A discussion of the literature on diophantine inequalities would not be complete without at least making reference to Margulis's famous resolution [1989] of the Oppenheim conjecture. With this work Margulis reduced the number of variables required to show the existence of infinitely many solutions to the inequality (1-3) from 5 to 3. Margulis's approach used dynamical methods, and is rather different in flavour to anything in this paper. In particular this method does not provide an asymptotic formula for the number of solutions in which the variables are bounded in a box.

1B. ***Notation.*** Before continuing with the rest of our introduction, we feel that, given the technical nature of some of the statements to follow, it is prudent to fix all our notation at the outset.

We will use standard asymptotic notation O, o, and Ω. 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 $|f(N)| \leqslant Cg(N)$ *for N sufficiently large*. Rather we require the inequality to hold for all N in some prespecified range. If N is a natural number, the range is always assumed to be $\mathbb{N}$ unless otherwise specified. For us, $0 \notin \mathbb{N}$.

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 κ notation from [Green and Tao 2010a]: $\kappa(x)$ 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 4.8, by n). If an implied constant depends on other parameters, we will denote these by subscripts, e.g., $O_{c,C,\varepsilon}(1)$, or $f \asymp_\varepsilon g$. By notation such as $o_\rho(1)$ we mean a term which tends to zero as the asymptotic parameter tends to infinity with ρ fixed.

If N is a natural number, we use $[N]$ to denote $\{n \in \mathbb{N} : n \leqslant N\}$, 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 $\boldsymbol{x} \in \mathbb{R}^m$, we let $\|\boldsymbol{x}\|_{\mathbb{R}^m/\mathbb{Z}^m}$ denote $\sup_i |x_i - [x_i]|$.

If $X, Y \subset \mathbb{R}^d$ for some d, we define

$$\operatorname{dist}(X, Y) := \inf_{x \in X, y \in Y} \|x - y\|_\infty.$$

If X is the singleton $\{x\}$, we write $\operatorname{dist}(x, Y)$ for $\operatorname{dist}(\{x\}, Y)$. 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 $\operatorname{dist}(X, Y)$ 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 use the notation $L^* : (\mathbb{R}^m)^* \to (\mathbb{R}^d)^*$ for the dual linear map between the dual spaces. For a set $U \subset \mathbb{R}^d$ we use U^0 to denote the annihilator of U, i.e., the set of all f in the dual space $(\mathbb{R}^d)^*$ for which $f|_U \equiv 0$.

We let $\partial(X)$ denote the topological boundary of a set $X \subset \mathbb{R}^d$. Given $S \subset \mathbb{R}$ and $\lambda \in \mathbb{R}$, we let $\lambda S := \{x \in \mathbb{R} : \exists s \in S \text{ for which } \lambda s = x\}$. 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 $e(\theta)$ to mean $e^{2\pi i\theta}$.

In Section 8, if $\boldsymbol{x} \in \mathbb{R}^d$ and if a and b are two subscripts with $1 \leqslant a \leqslant b \leqslant d$, we use the notation $\boldsymbol{x}_{\boldsymbol{a}}^{\boldsymbol{b}}$ to denote the vector $(x_a, x_{a+1}, \dots, x_b)^T \in \mathbb{R}^{b-a+1}$.

1C. ***The main corollary.*** We will now begin the process of developing our first main result, namely Corollary 1.4. This result is the first to link diophantine inequalities, such as (1-1), to *Gowers norms*.

Given natural numbers N and d, and a function $f : [N] \to \mathbb{C}$, the *Gowers U^d norm* $\|f\|_{U^d[N]}$ was introduced into the literature around twenty years ago, as part of Gowers' [2001] proof of Szemerédi's theorem.[3] The U^d norms are genuine norms for $d \geqslant 2$, with $\|f\|_{U^d[N]}$ measuring the density of certain linear patterns weighted by f. Their presence in analytic number theory is by now well established (see for instance [Green and Tao 2008a; 2010a; Tao and Teräväinen 2018; 2019]), but, to help any readers who are unfamiliar with these norms, in Appendix A we have given a summary of the necessary definitions and basic notions.

Our present endeavour is motivated by one particular application of Gowers norms, namely the so-called "generalised von Neumann theorem" developed by Green and Tao [2008a; 2010a] to study linear equations with rational coefficients.

Theorem 1.2 (generalised von Neumann theorem for rational forms (nonquantitative)). *Let m, d be natural numbers, satisfying $d \geqslant m + 2$. Let L be an m-by-d real matrix with integer coefficients, with rank m. Suppose that there does not exist any nonzero row-vector in the row-space of L that has two or fewer nonzero coordinates. Then there is some natural number s at most $d - 2$ that satisfies the following. Let N be an integer parameter, let $f_1, \dots, f_d : [N] \to [-1, 1]$ be arbitrary functions, and suppose that*

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

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

$$\frac{1}{N^{d-m}} \sum_{\substack{\boldsymbol{n} \in [N]^d \\ L\boldsymbol{n} = \boldsymbol{0}}} \prod_{j=1}^{d} f_j(n_j) \ll_L \rho^{\Omega(1)} + o_\rho(1)$$

as $N \to \infty$.

[3] Gowers worked over the cyclic group $\mathbb{Z}/N\mathbb{Z}$ rather than $[N]$, but this is a very minor difference.

Results similar to Theorem 1.2 are central to Green and Tao's approach [2010a] to counting solutions to linear equations in primes. It seems reasonable to hope that, if one could combine Gowers norms and diophantine inequalities in a suitable way, then one might be able to develop a strong understanding of linear inequalities in primes. As we have already intimated in Section 1A, when describing our improvements to Parsell's results, this can be done. However, many additional technical difficulties occur for the primes, as the von Mangoldt function is unbounded; we have chosen to present a separate work on these issues [Walker 2019].

We should briefly discuss the nondegeneracy condition on L in the statement of Theorem 1.2, namely that "there does not exist any nonzero row-vector in the row-space of L that has two or fewer nonzero coordinates", as it may seem a little unnatural at first sight.[4] Suppose that such a row-vector existed. Suppose also that it is the coordinates at index i and index j which are nonzero. Then, by a short linear algebra argument (see Proposition 4.5), for any linear parametrisation $(\psi_1, \dots, \psi_d) = \Psi : \mathbb{R}^{d-m} \to \ker L$, ψ_i is a multiple of ψ_j. Such a coupling between the coordinates has dire consequences for any averaging approach built upon the independence of the different coordinates, such as the averaging in Gowers norms, and so this coupling must be precluded. We will present a rigorous version of this principle in the context of linear inequalities, in Theorem 2.14 below.

Regarding the condition $d \geqslant m+2$, if L has rank m and $d \leqslant m+1$ then in fact, as follows from Gaussian elimination, there must always exist a nonzero vector in the row space of L with two or fewer nonzero coordinates. Thus, the condition $d \geqslant m+2$ is a necessary one if the coordinates of $\ker L$ are to be suitably independent.

Remark 1.3. Theorem 1.2 is implicit in [Green and Tao 2010a], but there is no explicit such statement presented there, as those authors were focussed on results over the primes. We will describe how to extract Theorem 1.2 from the arguments of [loc. cit.], but we postpone this task until Section 5, as at that point we will also introduce a quantitative version (this will be Theorem 5.2).

Our first main result is a version of Theorem 1.2 for diophantine inequalities.

Corollary 1.4 (Gowers norms control diophantine inequalities (nonquantitative)). *Let m, d be natural numbers, satisfying $d \geqslant m+2$, and let ε 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 nonzero row-vector in the row-space of L that has two or fewer nonzero coordinates. Then there is some natural number s at most $d-2$, independent of ε, such that the following is true. Let N be an integer parameter and let $f_1, \dots, f_d : [N] \to [-1, 1]$ be arbitrary functions, and suppose that*

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

[4] For readers who are already familiar with the notion of Cauchy–Schwarz complexity, imposing this nondegeneracy condition on L is equivalent to insisting that $\ker L$ may be parametrised by a system of linear forms with finite Cauchy–Schwarz complexity.

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

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

as $N \to \infty$.

We can provide detailed information about how the implied constant and the $o_{\rho,L}(1)$ term depend on L, but we leave that to the next section and to Theorem 2.12.

Before giving some examples, let us make a few remarks about Corollary 1.4. Firstly, note that if L has integer coefficients then, by picking ε small enough, Corollary 1.4 immediately implies Theorem 1.2, since the inequality $\|L\boldsymbol{n}\|_\infty \leqslant \varepsilon$ is only satisfied if $L\boldsymbol{n} = \boldsymbol{0}$.

Next, due to the nested property of Gowers norms (see Appendix A) one sees that Corollary 1.4 may be fruitfully applied under the hypothesis $\min_j \|f_j\|_{U^{d-1}[N]} \leqslant \rho$.

Finally, we note that it might be tempting to think that Corollary 1.4 would follow easily from taking rational approximations of the coefficients of L and then using Theorem 1.2 as a black box. Though of course we cannot completely rule out an alternative approach to that of this paper, when one investigates the quantitative details it seems that such an argument will only quickly succeed if the coefficients of L are all extremely well-approximable by rationals, else the height of the rational approximations becomes too great to apply results like Theorem 1.2. We will need to employ a different strategy, and we discuss this at length in Section 2E.

1D. ***Fourier uniform sets and other examples.*** Let us illustrate the applications of Corollary 1.4 with certain explicit examples.

Example 1.5 (three-term irrational AP). The first example could have been analysed by Davenport and Heilbronn using the methods they developed [1946], but we include it here to demonstrate the simplest case in which Corollary 1.4 applies.

Let

$$L := (1 \quad -\sqrt{2} \quad -1+\sqrt{2}).$$

Then $m = 1$ and $d = 3$, and manifestly there does not exist any nonzero row-vector in the row-space of L that has two or fewer nonzero coordinates. Therefore Corollary 1.4 applies, and so, if for each N we have three functions $f_1, f_2, f_3 : [N] \to [-1, 1]$ satisfying $\min_j \|f_j\|_{U^2[N]} \leqslant \rho$ for some ρ in the range $0 < \rho \leqslant 1$, then we have

$$\left| \frac{1}{N^2} \sum_{\substack{n_1,n_2,n_3 \leqslant N \\ |n_1 - \sqrt{2}n_2 + (-1+\sqrt{2})n_3| \leqslant \varepsilon}} f_1(n_1) f_2(n_2) f_3(n_3) \right| \ll_\varepsilon \rho^{\Omega(1)} + o_\rho(1) \tag{1-5}$$

as $N \to \infty$.

The statement (1-5) 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 $\lfloor \theta + \frac{1}{2} \rfloor$, i.e., the nearest integer to θ. Then for any three functions $f_1, f_2, f_3 : [N] \to [-1, 1]$, we make the definition

$$T(f_1, f_2, f_3) := \frac{1}{N^2} \sum_{x,d \in \mathbb{Z}} f_3(x) f_2(x+d) f_1([x + \sqrt{2}d]). \tag{1-6}$$

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 performing the 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 that

$$T(f_1, f_2, f_3) = \frac{1}{N^2} \sum_{\substack{n_1, n_2, n_3 \leqslant N \\ |n_1 - \sqrt{2}n_2 + (-1+\sqrt{2})n_3| \leqslant \frac{1}{2}}} f_1(n_1) f_2(n_2) f_3(n_3). \tag{1-7}$$

By (1-5), this means that if $\min_j \|f_j\|_{U^2[N]} \leqslant \rho$ then

$$|T(f_1, f_2, f_3)| \ll \rho^{\Omega(1)} + o_\rho(1) \tag{1-8}$$

as $N \to \infty$.

Now suppose that A_N is a subset of $[N]$ with $|A_N| = \alpha_N N$. Let

$$f_{A_N} := 1_{A_N} - \alpha_N 1_{[N]}$$

be its so-called "balanced function". By the usual telescoping trick, $T(1_{A_N}, 1_{A_N}, 1_{A_N})$ is equal to

$$T(\alpha_N 1_{[N]}, \alpha_N 1_{[N]}, \alpha_N 1_{[N]}) + T(f_{A_N}, \alpha_N 1_{[N]}, \alpha_N 1_{[N]}) + T(1_{A_N}, f_{A_N}, \alpha_N 1_{[N]}) + T(1_{A_N}, 1_{A_N}, f_{A_N}).$$

One may then bound the final three terms using $\|f_{A_N}\|_{U^2[N]}$ and, from the relation (1-7), one has then established that, provided $\|f_{A_N}\|_{U^2[N]} \leqslant \rho$,

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

is equal to

$$\frac{\alpha_N^3}{N^2} \sum_{x,d \in \mathbb{Z}} 1_{[N]}(x) 1_{[N]}(x+d) 1_{[N]}([x + \sqrt{2}d]) + O(\rho^{\Omega(1)}) + o_\rho(1) \tag{1-9}$$

as $N \to \infty$. If $\|f_{A_N}\|_{U^2[N]} = o(1)$ as $N \to \infty$ then, by picking $\rho = \rho(N)$ to be a quantity that tends to zero suitably slowly, one can ensure that the error term in (1-9) is $o(1)$ as $N \to \infty$.

As is familiar from [Gowers 2001], for bounded functions the U^2-norm is closely related to the Fourier transform. Indeed, we say that the family of sets A_N is Fourier-uniform if the balanced functions f_{A_N} satisfy

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

as $N \to \infty$, and it is a standard result (see [Tao 2012, Exercise 1.3.18]) that A_N is Fourier uniform if and only if $\|f_{A_N}\|_{U^2[N]} = o(1)$ as $N \to \infty$. Therefore expression (1-9), and the remarks following it, imply the following corollary.

Corollary 1.6 (Fourier-uniform sets). *Let $\beta \in \mathbb{R} \setminus \mathbb{Q}$, and for each natural number N let A_N be a subset of $[N]$ with $|A_N| = \alpha_N N$. Suppose that A_N is a Fourier-uniform family of sets. Then*

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

is equal to

$$\frac{\alpha_N^3}{N^2} \sum_{x,d \in \mathbb{Z}} 1_{[N]}(x) 1_{[N]}(x+d) 1_{[N]}([x+\beta d]) + o_\beta(1)$$

as $N \to \infty$, where the $o_\beta(1)$ term also depends on the $o(1)$ term in the Fourier-uniformity expression for the family A_N.

Example 1.7. Let

$$L := \begin{pmatrix} 1 & 0 & -\sqrt{2} & -1+\sqrt{2} \\ 0 & 1 & -\sqrt{3} & -1+\sqrt{3} \end{pmatrix}. \tag{1-10}$$

Since $\sqrt{2}$ and $\sqrt{3}$ are distinct irrationals it is not hard to see that all elements of the row-space of L must have three or four nonzero coordinates, and so Corollary 1.4 applies. Letting $f_1, f_2, f_3, f_4 : [N] \to [-1, 1]$ be arbitrary functions, the reparametrisation $n_1 = [x + \sqrt{2}d]$, $n_2 = [x + \sqrt{3}d]$, $n_3 = x + d$, $n_4 = x$, shows that

$$\frac{1}{N^2} \sum_{\substack{\boldsymbol{n} \in [N]^4 \\ \|L\boldsymbol{n}\|_\infty \leqslant \frac{1}{2}}} \left(\prod_{j=1}^{4} f_j(n_j) \right) = \frac{1}{N^2} \sum_{x,d \in \mathbb{Z}} f_4(x) f_3(x+d) f_1([x+\sqrt{2}d]) f_2([x+\sqrt{3}d]).$$

Corollary 1.4 controls the left-hand side of this expression in terms of the Gowers norms of the functions f_j, and so the right-hand side is controlled as well.

We can generalise the previous two examples as follows.

Corollary 1.8. *Let $\theta_1, \dots, \theta_s \in \mathbb{R}$ be distinct irrational numbers. For each natural number N let A_N be a subset of $[N]$, with $|A_N| = \alpha_N N$ and with balanced function f_{A_N}. Suppose that $\|f_{A_N}\|_{U^{s+1}[N]} = o(1)$ as $N \to \infty$. Then*

$$\frac{1}{N^2} \sum_{x,d \in \mathbb{Z}} 1_{A_N}(x) 1_{A_N}(x+d) \left(\prod_{i=1}^{s} 1_{A_N}([x+\theta_i d]) \right) \tag{1-11}$$

is equal to

$$\frac{\alpha_N^{s+2}}{N^2} \sum_{x,d \in \mathbb{Z}} 1_{[N]}(x) 1_{[N]}(x+d) \left(\prod_{i=1}^{s} 1_{[N]}([x+\theta_i d]) \right) + o(1)$$

as $N \to \infty$, where the $o(1)$ error term may depend on $\theta_1, \dots, \theta_s$ and on the rate of decay of $\|f_{A_N}\|_{U^{s+1}[N]}$.

Proof. Apply Corollary 1.4 to the s-by-$s+2$ matrix

$$L = (I \quad -\boldsymbol{\theta} \quad -1+\boldsymbol{\theta}), \tag{1-12}$$

where I denotes the identity matrix and $\boldsymbol{\theta}$ denotes the vector $(\theta_1, \dots, \theta_s)^T \in \mathbb{R}^s$. □

The question remains as to whether one can use Corollary 1.8, perhaps in conjunction with a density increment argument such as is used in [Green and Tao 2010b], to deduce that there are infinitely many $s+2$-tuples of the form $(x, x+d, [x+\theta_1 d], \dots, [x+\theta_s d])$ inside any set of natural numbers with positive upper Banach density. More generally, one might wish to find tuples in which all the coordinates are of the form $x + p(d)$ where p is a generalised polynomial of degree 1 without a constant term. This result is already known, in fact, but as it stands the only proof uses ergodic theory methods (see [McCutcheon 2005, Theorem B]). We view Corollary 1.8 as a promising first step towards a purely combinatorial proof of this result, with a chance to prove explicit bounds.

Corollary 1.4 has immediate consequences for counting solutions to diophantine inequalities weighted by explicit bounded pseudorandom functions. In particular there is the following natural analogue of [Green and Tao 2010a, Proposition 9.1] concerning the cancellation of the Möbius function, which we mentioned earlier in Corollary 1.1.

Corollary 1.9 (Möbius orthogonality). *Let m, d be natural numbers satisfying $d \geqslant m+2$, and let ε 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 nonzero row-vector in the row-space of L that has two or fewer nonzero coordinates. Let μ denote the Möbius function and let N be an integer parameter. Then, for any bounded functions $f_2, \dots, f_d : [N] \to [-1, 1]$,*

$$\sum_{\substack{\boldsymbol{n} \in [N]^d \\ \|L\boldsymbol{n}\|_\infty \leqslant \varepsilon}} \mu(n_1) \biggl(\prod_{j=2}^{d} f_j(n_j) \biggr) = o_{L,\varepsilon}(N^{d-m})$$

as $N \to \infty$. The same is true with μ replaced by the Liouville function λ.

Proof. This follows immediately from Corollary 1.4 and the deep facts (stated in [Green and Tao 2010a], proved in [Green and Tao 2012] and [Green et al. 2012]) that $\|\mu\|_{U^{s+1}[N]} = o_s(1)$ and $\|\lambda\|_{U^{s+1}[N]} = o_s(1)$ as $N \to \infty$. □

Corollary 1.9, when applied to the matrix (1-12), yields Corollary 1.1 from earlier in this introduction. It also yields cancellation in expressions such as

$$\sum_{\substack{\boldsymbol{n} \in [N]^4 \\ n_1 - n_2 = n_2 - n_3 \\ |(n_2 - n_3) - \sqrt{2}(n_3 - n_4)| \leqslant \frac{1}{2}}} \mu(n_1)\mu(n_2)\mu(n_3)\mu(n_4) = o(N^2) \tag{1-13}$$

as $N \to \infty$. There are of course many such examples; we chose (1-13) to emphasise that one can choose configurations that combine rational and irrational relations.

2. The main theorem

In our results from the previous section, the quantitative nature of the dependence of the error terms on the matrix L was hidden. Our main theorem (Theorem 2.12 below) addresses this point, which turns out to be surprisingly subtle.

To start off, 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 $L : \mathbb{R}^d \to \mathbb{R}^m$ be a linear map. Let $F : \mathbb{R}^d \to [0, 1]$ and $G : \mathbb{R}^m \to [0, 1]$ be two functions, with F supported on $[-N, N]^d$ and G compactly supported. Let $f_1, \ldots, f_d : [N] \to [-1, 1]$ be arbitrary functions. We define

$$T^L_{F,G,N}(f_1, \ldots, f_d) := \frac{1}{N^{d-m}} \sum_{\boldsymbol{n}\in\mathbb{Z}^d} \biggl(\prod_{j=1}^{d} f_j(n_j)\biggr) F(\boldsymbol{n})G(L\boldsymbol{n}). \tag{2-1}$$

The normalisation factor of N^{d-m} is appropriate; in Lemma 3.2 we show that $T^L_{F,G,N}(f_1, \ldots, f_d) \ll_G 1$.

In Theorem 2.12 we will bound $T^L_{F,G,N}(f_1, \ldots, f_d)$ above by Gowers norms. The error term will depend on three further notions: the rational relations present in L; the "approximation function" A_L, which will measure the approximate rational relations present in L; and the nondegeneracy of L, which is related to the nondegeneracy conditions in Theorem 1.2. These three notions will be introduced in the next three subsections, before we (finally!) state Theorem 2.12 in Section 2D.

2A. ***Rational relations.*** Let us consider some properties of the image $L(\mathbb{Z}^d)$. It is certainly true that if $L : \mathbb{R}^d \to \mathbb{R}^m$ is a surjective linear map then $\operatorname{span}(L(\mathbb{Z}^d)) = \mathbb{R}^m$. However, $L(\mathbb{Z}^d)$ need not be dense in $\mathbb{R}^m$, as the matrix L may satisfy some rational relations. These in turn restrict $L(\mathbb{Z}^d)$ to various affine subspaces of $\mathbb{R}^m$.

This observation motivates the following definitions:

Definition 2.2 (rational dimension, rational map, purely irrational). Let m and d be natural numbers, with $d \geqslant m+1$. 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 $\Theta : \mathbb{R}^m \to \mathbb{R}^u$ for which $\Theta L(\mathbb{Z}^d) \subseteq \mathbb{Z}^u$. We call u the *rational dimension* of L, and we call any map Θ 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

$$\Theta := (5 \;\; 1),$$

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 a subset of a 2-dimensional lattice, which it is not. So the rational dimension of L is equal to 1.

Ours is certainly not the first paper on the topic of diophantine inequalities to have considered issues such as this. For example, in the previous section we remarked that Müller [2005] came across a similar phenomenon. Given quadratic forms $Q_1, \dots, Q_r$ he found infinitely many solutions $\boldsymbol{x}$ to the inequalities

$$|Q_1(\boldsymbol{x})| < \varepsilon, \dots, |Q_r(\boldsymbol{x})| < \varepsilon,$$

under the hypotheses that every quadratic form in the set

$$\left\{ \sum_{i=1}^{r} \alpha_i Q_i : \alpha_1, \dots, \alpha_r \in \mathbb{R},\ \boldsymbol{\alpha} \neq \boldsymbol{0} \right\}$$

was irrational and had rank greater than $8r$. One can use the coefficients of the quadratic forms to translate this problem into one of trying to understand $T^L_{F,G,N}(f_1, \dots, f_d)$ for a certain linear map L and for functions $f_1, \dots, f_d$ supported on the image of quadratic monomials. Then, Müller's hypothesis that all the linear combinations of the Q_i are irrational is transformed into the hypothesis that L is purely irrational. In this paper we consider all L, not just those which are purely irrational, and this causes some added complications.

In our definition of rational dimension, there is some flexibility over the exact choice of map Θ. The next lemma identifies an invariant.

Lemma 2.3. *Let m and d be natural numbers, with $d \geqslant m+1$. 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 Φ be any surjective map from $\operatorname{im}((\Theta_1, \Theta_2))$ to $\mathbb{R}^{u+1}$ for which $\Phi(\mathbb{Z}^{2u} \cap \operatorname{im}((\Theta_1, \Theta_2))) \subseteq \mathbb{Z}^{u+1}$. 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. □

We will also need to identify the quantitative aspects of these rational relations, in order to properly state the main theorem.

Definition 2.4 (rational complexity). Let m and d be natural numbers, with $d \geqslant m+1$. 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 Θ that is a rational map for L and for which $\|\Theta\|_\infty \leqslant C$. If L is purely irrational, then L has rational complexity 0.

In this definition, 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 with respect to the standard bases.

We observe that a linear map with maximal rational dimension, i.e., with rational dimension m, is equivalent to a linear map with integer coefficients, in the following sense.

Lemma 2.5 (maximal rational dimension). *Let m and d be natural numbers, with $d \geqslant m+1$. 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 Θ and an m-by-d matrix S with integer coefficients such that, as matrices, $\Theta L = S$. Furthermore, $\|\Theta\|_\infty \leqslant C$.*

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

We will use this lemma in Section 5, to reduce the study of maps L with maximal rational dimension to the study of maps L with integer coefficients.

2B. ***Approximation function.*** We must also 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 ΘL 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. The critical place where it enters the argument is Lemma 3.4, whose content we briefly sketch here, so as to further motivate our introduction of the "approximation function" below.

This will be the first, of many, places in the paper in which we have to manipulate Lipschitz functions. For the reader's benefit, in Appendix B we have collected together the definitions and results that we will use.

Let L be an m-by-d matrix, which may depend on the asymptotic parameter N. Suppose that one is seeking an upper bound on $T^L_{F,G,N}(1, \dots, 1)$, where G is a Lipschitz function supported on $[-1, 1]^m$ and F is a function supported on $[-N, N]^d$. We note that this task is a special case of bounding $T^L_{F,G,N}(f_1, \dots, f_d)$ above by the Gowers norms of the functions f_i. In our main proof, bounds on $T^L_{F,G,N}(1, \dots, 1)$ will be useful when controlling some error terms which occur when the inequality is perturbed (see Section 2E for a full discussion of this point).

Also suppose, for simplicity, that the first m columns of L form the identity matrix, and let $\mathbf{v}_j \in \mathbb{R}^m$ denote the j-th column of L. Then, by summing over the variables $n_1, \dots, n_m \in \mathbb{Z}$, one quickly derives that

$$T^L_{F,G,N}(1, \dots, 1) \ll \frac{1}{N^{d-m}} \sum_{\substack{n_{m+1},\dots,n_d \in \mathbb{Z} \\ |n_{m+1}|,\dots,|n_d| \leqslant N}} \tilde{G}\Big(\sum_{j=m+1}^{d} \mathbf{v}_j n_j\Big),$$

where $\tilde{G}$ is a $\mathbb{Z}^m$-periodic Lipschitz function formed by taking translates of G.

We consider $\tilde{G}$ as a function on $\mathbb{R}^m/\mathbb{Z}^m$, and approximate it by a short exponential sum (as one may do for all such Lipschitz functions).[5] As is familiar in these kind of problems, one is left with having to bound the expression that arises from the nonzero Fourier modes. Here, one ends up with terms

$$\frac{1}{N^{d-m}} \sum_{\substack{n_{m+1},\dots,n_d \in \mathbb{Z} \\ |n_{m+1}|,\dots,|n_d| \leqslant N}} e\Big(\mathbf{k} \cdot \sum_{j=m+1}^{d} \mathbf{v}_j n_j\Big)$$

[5] See Lemma B.3.

with $\boldsymbol{k} \in \mathbb{Z}^m \setminus \{\mathbf{0}\}$, which one may sum as geometric progressions over n_{m+1} to n_d. This means we have to control

$$\max_{\substack{\boldsymbol{k} \in \mathbb{Z}^m \\ 0<\|\boldsymbol{k}\|_\infty \leqslant X}} \Bigl(\prod_{j=m+1}^{d} \min(1, N^{-1} \|\boldsymbol{k} \cdot \boldsymbol{v_j}\|_{\mathbb{R}/\mathbb{Z}}^{-1}) \Bigr),$$

where X is some threshold, and the above expression is certainly bounded by

$$N^{-1} \max_{\substack{\boldsymbol{k} \in \mathbb{Z}^m \\ 0<\|\boldsymbol{k}\|_\infty \leqslant X}} \|L^T \boldsymbol{k}\|_{\mathbb{R}^d/\mathbb{Z}^d}^{-1}, \tag{2-2}$$

as the first m columns of L have integer coordinates. One hopes to bound expression (2-2) by $o(1)$ as $N \to \infty$.

We observe two facts about (2-2). Firstly, if L is not purely irrational and if X is larger than the rational complexity of L, then the expression (2-2) is infinite! Secondly, even if L is purely irrational then it could still be the case that (2-2) tends to infinity with N, as L may depend on N. We conclude that, with the state of our current argument, the size of expression (2-2) — or an expression like it — must be included in our error terms.

Motivated by the above discussion, we introduce the "approximation function". The definition is phrased in terms of dual functions — this will make linear algebraic manipulations more straightforward later on — and for real vectors φ rather than for integer vectors $\boldsymbol{k}$, which reflects the general situation in which the first m columns of L are not the identity matrix. We also generalise to the case of arbitrary rational dimension u, rather than just $u = 0$.

Following this definition and some remarks, we will show how to calculate the approximation function in an explicit example. This should hopefully serve to clarify the properties of this somewhat technical object.

Definition 2.6 (approximation function). Let m and d be natural numbers, with $d \geqslant m+1$. 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 $u \leqslant m-1$. Then we define the *approximation function of* L, denoted $A_L : (0, 1] \times (0, 1] \to (0, \infty)$, by

$$A_L(\tau_1, \tau_2) := \inf_{\substack{\varphi \in (\mathbb{R}^m)^* \\ \operatorname{dist}(\varphi, \Theta^*((\mathbb{R}^u)^*)) \geqslant \tau_1 \\ \|\varphi\|_\infty \leqslant \tau_2^{-1}}} \operatorname{dist}(L^*\varphi, (\mathbb{Z}^d)^T),$$

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 τ_1.[6]

[6]When $u = m$ we've already seen (Lemma 2.5) that L may be transformed into a matrix S with integer coefficients, and thus L is somewhat degenerate from the point of view of diophantine approximation. We define $A_L(\tau_1, \tau_2)$ for such matrices only to avoid having to discuss this special case in the statement of Theorem 2.12 later on.

From our discussion of (2-2) above, one sees that upper bounds on $A_L(\tau_1, \tau_2)^{-1}$ will be the main focus. The threshold τ_2^{-1} plays the role of the threshold X in (2-2), and the condition $\operatorname{dist}(\varphi, \Theta^*((\mathbb{R}^u)^*)) \geqslant \tau_1$ corresponds to the condition $\|\boldsymbol{k}\|_\infty \geqslant 1$ which is implicit in (2-2).

There is some notation to unpack in Definition 2.6. Regarding the notion "dist", we remind the reader of some material from Section 1B, namely 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 ℓ^∞ 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

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

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

Note that Definition 2.6 is independent of the choice of Θ. Indeed, by basic linear algebra $\Theta^*((\mathbb{R}^u)^*) = (\ker \Theta)^0$, where $(\ker \Theta)^0$ is the annihilator of $\ker \Theta$ (see Section 1B). By Lemma 2.3, $\ker \Theta$ is independent of the choice of Θ, and therefore so is $(\ker \Theta)^0$.

Example 2.7. Suppose that, as a matrix,

$$L := (1 \quad -\sqrt{2} \quad -1 + \sqrt{2})$$

as in Example 1.5. Then L is purely irrational, i.e., $u = 0$, since its coefficients are not all in rational ratio. Therefore $A_L(\tau_1, \tau_2)$ is equal to

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

As we said above, we seek an upper bound on $A_L(\tau_1, \tau_2)^{-1}$. To this end, we claim that, for this particular L and for all $\tau_1, \tau_2 \in (0, 1]$, one has

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

Indeed, we know that, for all $q \in \mathbb{N}$,

$$\|q\sqrt{2}\|_{\mathbb{R}/\mathbb{Z}} \geqslant \frac{1}{10q}.$$

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

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

which is a contradiction.

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

$$\max(\|-k\sqrt{2}\|_{\mathbb{R}/\mathbb{Z}}, \|-k + k\sqrt{2}\|_{\mathbb{R}/\mathbb{Z}}) \geqslant \|[k]\sqrt{2}\|_{\mathbb{R}/\mathbb{Z}} - \frac{\tau_2}{50} \geqslant \frac{1}{10[k]} - \frac{\tau_2}{50} \geqslant \frac{1}{10\tau_2^{-1} + 10} - \frac{\tau_2}{50} \gg \tau_2.$$

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.

It is not too difficult to show that if L is an m-by-d matrix with rank m and with algebraic coefficients, then

$$A_L(\tau_1, \tau_2) \gg_L \min(\tau_1, \tau_2^{O_L(1)}), \tag{2-3}$$

where the $O_L(1)$ term in the exponent depends on the algebraic degree of the coefficients of L.[7] We shall give a proof of this statement in Appendix E. In general, however, $A_L(\tau_1, \tau_2)$ could tend to zero arbitrarily quickly as τ_2 tends to zero, for example in the case when $L = (1 \;\; -\lambda \;\; -1+\lambda)$ and λ 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 2.8. For all permissible choices of L, τ_1 and τ_2 in Definition 2.6, $A_L(\tau_1, \tau_2)$ is positive.

Proof. Let u be the rational dimension of L. Without loss of generality we may assume that $u \leqslant m-1$. Then, for all $\varphi \in (\mathbb{R}^m)^* \setminus \Theta^*((\mathbb{R}^u)^*)$ we have that $\operatorname{dist}(L^*\varphi, (\mathbb{Z}^d)^T) > 0$. (If this were not the case then the map $(\Theta, \varphi) : \mathbb{R}^m \to \mathbb{R}^{u+1}$ 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. □

The expression $A_L(\tau_1, \tau_2)^{-1}$ will appear in the error term of our main theorem; Claim 2.8 shows that such an error term still has content.

2C. *Nondegeneracy.*

In the statement of Theorem 1.2, which we remind the reader was the result of Green and Tao that used Gowers norms to control the number of solutions to linear equations with integer coefficients, one recalls that there were certain linear-algebraic notions of nondegeneracy for the matrix L. These concerned the rank of L and the properties of its row space. In the setting of diophantine inequalities it will transpire that the same notions of nondegeneracy are important — this much was obvious from the statement of Corollary 1.4 — except that, in order to control the error terms when L depends on N, one must assume that L is not even "approximately" degenerate.

In order to make these notions precise, we will first give some names to the sets of degenerate maps that we wish to avoid.

Definition 2.9 (low-rank variety). Let m, d be natural numbers satisfying $d \geqslant m+1$. Let $V_{\mathrm{rank}}(m, d)$ denote the set of all linear maps $L : \mathbb{R}^d \to \mathbb{R}^m$ whose rank is less than m. We call $V_{\mathrm{rank}}(m, d)$ the *low-rank variety*.

[7] One 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.

Let $V_{\text{rank}}^{\text{unif}}(m,d)$ 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 $\boldsymbol{e_i}$, for which $L|_{\text{span}(\boldsymbol{e_j}:j\neq i)}$ has rank less than m. We call $V_{\text{rank}}^{\text{unif}}(m,d)$ the *uniform low-rank variety*.

We make the trivial observation that $V_{\text{rank}}^{\text{unif}}(m,d)$ contains $V_{\text{rank}}(m,d)$. For certain technical reasons it will be much more convenient to work with matrices $L\notin V_{\text{rank}}^{\text{unif}}(m,d)$, as opposed to merely working with matrices $L\notin V_{\text{rank}}(m,d)$, as we will be able to fix an arbitrary coordinate and still be left with a full rank linear map.

Definition 2.10 (dual degeneracy variety). Let m,d be natural numbers satisfying $d\geqslant m+2$. Let $\boldsymbol{e_1},\dots,\boldsymbol{e_d}$ denote the standard basis vectors of $\mathbb{R}^d$, and let $\boldsymbol{e_1^*},\dots,\boldsymbol{e_d^*}$ 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 $i,j\leqslant d$, and some real number λ, such that $\boldsymbol{e_i^*}-\lambda\boldsymbol{e_j^*}$ is nonzero and $(\boldsymbol{e_i^*}-\lambda\boldsymbol{e_j^*})\in L^*((\mathbb{R}^m)^*)$. We call $V_{\text{degen}}^*(m,d)$ the *dual degeneracy variety*.

It may be easily verified that this definition does nothing more than rephrase the condition that appeared in the statements of Corollary 1.4 and Theorem 1.2 concerning the row-space of a degenerate map L, namely that there exists a nonzero row-vector in the row-space of L that has two or fewer nonzero coordinates. The formulation in terms of dual spaces will be particularly convenient, however, for some of the algebraic manipulations in Section 5. This is the reason why we use the term “dual” in the name of $V_{\text{degen}}^*(m,d)$.[8]

Having introduced $V_{\text{rank}}(m,d)$, $V_{\text{rank}}^{\text{unif}}(m,d)$ and $V_{\text{degen}}^*(m,d)$, we can articulate the relationship between the nondegeneracy conditions in Theorem 1.2 (for linear equations given by L) and our nondegeneracy conditions in Theorem 2.12 below (for linear inequalities given by L). Indeed, for equations, L is nondegenerate if

$$L\notin V_{\text{rank}}(m,d)\quad\text{or}\quad L\notin V_{\text{degen}}^*(m,d). \tag{2-4}$$

For inequalities, L is nondegenerate if

$$\operatorname{dist}(L,V_{\text{rank}}^{\text{unif}}(m,d))\geqslant c\quad\text{or}\quad \operatorname{dist}(L,V_{\text{degen}}^*(m,d))\geqslant c', \tag{2-5}$$

for some fixed parameters c and c'. One can see immediately how the conditions for inequalities are “approximate” versions of the conditions for equations.

Example 2.11. It may be instructive to consider a matrix such as

$$L=\begin{pmatrix}1+N^{-1} & \sqrt{3}+N^{-\frac12} & \pi & -\pi+\sqrt{2}\\ 2 & 2\sqrt{3}+N^{-\frac12} & -\sqrt{5} & e\end{pmatrix}.$$

We observe that L has rank 2 and $L\notin V_{\text{degen}}^*(2,4)$. If one knew Theorem 1.2 and the conditions (2-4), then one might perhaps have hoped to apply the theory of Gowers norms to bound the number of solutions

[8] Later on (in Definition 4.4) we will have a set of degenerate maps $V_{\text{degen}}(d-m,d)$ which will parametrise the kernel of maps in $V_{\text{degen}}^*(m,d)$. Since these maps feel somewhat dual to those maps in $V_{\text{degen}}^*(m,d)$, we will come to call $V_{\text{degen}}(d-m,d)$ the “degeneracy variety”.

to inequalities given by L. However, by considering perturbations of the first two columns, we see that $\operatorname{dist}(L, V^*_{\mathrm{degen}}(2,4)) = o(1)$ as $N \to \infty$. 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. So, despite the fact that L is nondegenerate from the point of view of equations, L *is* degenerate from the point of view of inequalities and the conditions (2-5). Thus, our main theorem on inequalities will not apply to this L.

Furthermore, we have another result (Theorem 2.14 below) which shows that one cannot possibly use Gowers norms to control inequalities given by such an L. Therefore, whatever methods we use to prove Theorem 2.12, these methods must necessarily break down when applied to this example.

2D. ***The main theorem and a partial converse.*** Having laid the groundwork, we may now state the main theorem of this paper.

Theorem 2.12 (Main Theorem). *Let m, d be natural numbers, satisfying $d \geqslant m+2$, and let ε, c, C, C' be positive reals. Let N be an integer parameter and let $L = L(N) : \mathbb{R}^d \to \mathbb{R}^m$ be a surjective linear map that satisfies $\|L\|_\infty \leqslant C$. Let $A_L : (0,1] \times (0,1] \to (0,\infty)$ be the approximation function of L. Suppose further that $\operatorname{dist}(L, V^{\mathrm{unif}}_{\mathrm{rank}}(m,d)) \geqslant c$, that $\operatorname{dist}(L, V^*_{\mathrm{degen}}(m,d)) \geqslant c'$, and that L has rational complexity at most C'. Then there exists a natural number s at most $d-2$, independent of ε, such that the following is true. Let $F : \mathbb{R}^d \to [0,1]$ be the indicator function of $[1,N]^d$, and let $G : \mathbb{R}^m \to [0,1]$ be the indicator function of a convex domain contained in $[-\varepsilon,\varepsilon]^m$. Let $f_1, \dots, f_d : [N] \to [-1,1]$ be arbitrary functions, and suppose that*

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

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

$$T^L_{F,G,N}(f_1, \dots, f_d) \ll_{c,c',C,C',\varepsilon} \rho^{\Omega(1)} + o_{\rho,A_L,c,c',C,C'}(1) \tag{2-6}$$

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

$$N^{-\Omega(1)} \rho^{-O(1)} A_L(\Omega_{c,c',C,C'}(1), \rho)^{-1}.$$

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, and c and c' by a single constant c, without weakening the conclusion. We proceed with this assumption.

Let us note some consequences of this theorem. Firstly, since $A_L(\Omega_{c,C}(1), \rho)^{-1}$ is finite (by Claim 2.8), Theorem 2.12 immediately implies Corollary 1.4 (this was the qualitative statement around which we structured Section 1). Hence Theorem 2.12 also implies all the other corollaries from Section 1. Secondly, from (2-3), or rather from our full quantitative version in Lemma E.1, we have another corollary for matrices L with algebraic coefficients.

Corollary 2.13 (inequalities with algebraic coefficients). *Assume the same hypotheses as Theorem 2.12, 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*

$$T^L_{F,G,N}(f_1,\dots,f_d) \ll_{c,C,\varepsilon,H} \rho^{\Omega(1)} + N^{-\Omega(1)}\rho^{-O_k(1)}.$$

The reader may wonder how the implied constant in these statements depends on ε. Ultimately the implied constant in (2-6) tends to infinity as ε tends to zero, as our approximation argument in Section 6 will not be efficient in powers of ε. Yet, to prevent our notation becoming too unreadable, we choose not to keep track of the precise behaviour of implied constants involving ε.

As we remarked in Section 2C and Example 2.11, we can also prove a partial converse to Theorem 2.12. This result demonstrates that the nondegeneracy condition $\operatorname{dist}(L, V^*_{\text{degen}}(m,d)) \geqslant c$ is necessary in order to use Gowers norms to control inequalities given by L.

Theorem 2.14. *Let* m, d *be natural numbers, satisfying* $d \geqslant m+2$*, and let* ε, c, C *be positive constants. For each natural number* N*, let* $L = L(N) : \mathbb{R}^d \to \mathbb{R}^m$ *be a linear map satisfying* $\|L\|_\infty \leqslant C$*. Let* $F : \mathbb{R}^d \to [0,1]$ *denote the indicator function of* $[1,N]^d$ *and* $G : \mathbb{R}^m \to [0,1]$ *denote the indicator function of* $[-\varepsilon,\varepsilon]^m$*. Assume further that* $\operatorname{dist}(L, V_{\text{rank}}(m,d)) \geqslant c$ *and that* $T^L_{F,G,N}(1,\dots,1) \gg_{c,C,\varepsilon} 1$ *for large enough* N*.*

Suppose that

$$\liminf_{N\to\infty} \operatorname{dist}(L, V^*_{\text{degen}}(m,d)) = 0.$$

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

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

and

$$|T^L_{F,G,N}(f_1,\dots,f_d)| > H(\rho) + E_\rho(N). \tag{2-7}$$

In other words, the conclusion of Theorem 2.12 cannot possibly hold if $\operatorname{dist}(L, V^*_{\text{degen}}(m,d))$ is arbitrarily close to 0, even if one replaces the $\rho^{\Omega(1)}$ dependence in (2-6) with a function $H(\rho)$ that could potentially decay to zero arbitrarily slowly as ρ tends to zero.

The proof of Theorem 2.14 is contained in Section 9, which can be read independently of the rest of the paper.

2E. *The proof strategy.* All the corollaries from Sections 1 and 2 are implied by Theorem 2.12, so our remaining task is to prove this theorem. Speaking somewhat informally, we wish to bound $T^L_{F,G,N}(f_1,\dots,f_d)$ in terms of some Gowers norms $\|f_j\|_{U^{s+1}[N]}$ when the functions F and G are the indicator functions of certain convex domains. Now, one might expect the proof to be easier if, instead, F and G were functions with nicer analytic properties — Lipschitz functions, for example. This is indeed the case, and thus our proof splits naturally into two parts. The first part, contained in Sections 3 and 5, reduces Theorem 2.12

to a similar statement in which the functions F and G are Lipschitz — this will be Theorem 5.6. The second part of the paper is devoted to proving Theorem 5.6. For the rest of this subsection we will try to articulate the strategies for each part, and to elucidate the main technical difficulties.

In [Green and Tao 2010a], replacing convex cutoffs with Lipschitz cutoffs was an easy operation, accomplished in a couple of pages in Appendices A and C of that paper. Somewhat surprisingly, this part turns out to be the trickiest element in the setting of inequalities, at least when L is not purely irrational.

Replacing F with a Lipschitz cutoff is no issue, but the difficulty comes from replacing G. Consider the example

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

from Section 2A, in which we established that L has rational dimension 1 and that

$$L(\mathbb{Z}^4) \subset \left\{ \boldsymbol{x} \in \mathbb{R}^2 : \boldsymbol{x} \cdot \begin{pmatrix} 5 \\ 1 \end{pmatrix} \in \mathbb{Z} \right\}.$$

Take G to be the indicator of the compact convex domain

$$\left\{ \boldsymbol{x} \in \mathbb{R}^2 : \|\boldsymbol{x}\|_\infty \leqslant 10,\ -1 \leqslant \boldsymbol{x} \cdot \begin{pmatrix} 5 \\ 1 \end{pmatrix} \leqslant 1 \right\}.$$

Then

$$T^L_{F,G,N}(f_1, \dots, f_4) = \frac{1}{N^2} \sum_{\substack{\boldsymbol{n} \in \mathbb{Z}^4 \\ \|L\boldsymbol{n}\|_\infty \leqslant 10 \\ (5\ 1)L\boldsymbol{n} = -1,0,1}} \left(\prod_{j=1}^{4} f_j(n_j) \right) F(\boldsymbol{n}). \tag{2-8}$$

To replace the convex cutoff G by a Lipschitz cutoff, a natural approach is to take a Lipschitz function $\tilde{G}$ that is a minorant for G,[9] with $\tilde{G}$ supported on the set

$$\left\{ \boldsymbol{x} \in \mathbb{R}^2 : \|\boldsymbol{x}\|_\infty \leqslant 10 - \delta, \quad -1 + \delta \leqslant \boldsymbol{x} \cdot \begin{pmatrix} 5 \\ 1 \end{pmatrix} \leqslant 1 - \delta \right\}$$

for some small positive parameter δ, and $\tilde{G}$ identically equal to 1 on the set

$$\left\{ \boldsymbol{x} \in \mathbb{R}^2 : \|\boldsymbol{x}\|_\infty \leqslant 10 - 2\delta, \quad -1 + 2\delta \leqslant \boldsymbol{x} \cdot \begin{pmatrix} 5 \\ 1 \end{pmatrix} \leqslant 1 - 2\delta \right\}.$$

One has $\|G - \tilde{G}\|_1 = \kappa(\delta)$, so one might hope that for any functions $f_1, \dots, f_4$ one would have

$$|T^L_{F,G,N}(f_1, \dots, f_4) - T^L_{F,\tilde{G},N}(f_1, \dots, f_4)| = \kappa(\delta). \tag{2-9}$$

However, no matter how small we choose δ,

$$T^L_{F,\tilde{G},N}(f_1, \dots, f_4) \approx \frac{1}{N^2} \sum_{\substack{\boldsymbol{n} \in \mathbb{Z}^4 \\ \|L\boldsymbol{n}\|_\infty \leqslant 10-\delta \\ (5\ 1)L\boldsymbol{n} = 0}} \left(\prod_{j=1}^{4} f_j(n_j) \right) F(\boldsymbol{n}). \tag{2-10}$$

[9] One also finds a majorant Lipschitz function, but that won't feature in this example.

In moving from G to $\tilde{G}$ the range of summation for $\boldsymbol{n}$ between expressions (2-8) and (2-10) has been cut by a factor of two-thirds! Thus we have no reason to expect that (2-9) should hold for all functions $f_1, \ldots, f_4$.

We circumvent these difficulties by employing the following idea. Rather than replacing G with a Lipschitz cutoff straight away, when faced with an expression such as (2-8) we can perform some initial reparametrisation, observing that there is a linear map $\Xi : \mathbb{R}^3 \to \mathbb{R}^4$ with integer coefficients which gives a lattice parametrisation of those $\boldsymbol{n} \in \mathbb{Z}^4$ for which $(5\ \ 1)L\boldsymbol{n} = 0$, namely

$$\Xi \begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix} = \begin{pmatrix} m_1 \\ -5m_1 - 5m_2 \\ m_3 \\ m_2 \end{pmatrix}.$$

Moreover, $\boldsymbol{n} \in \mathbb{Z}^4$ with $(5\ \ 1)L\boldsymbol{n} = \pm 1$ if and only if there are integers m_1, m_2, m_3 for which

$$\boldsymbol{n} = \Xi \begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix} + \begin{pmatrix} 0 \\ \pm 1 \\ 0 \\ 0 \end{pmatrix}.$$

This enables us to decompose $T^L_{F,G,N}(f_1, \ldots, f_4)$ into three separate expressions, each of the form

$$\frac{1}{N^2} \sum_{\boldsymbol{m} \in \mathbb{Z}^3} \Bigl(\prod_{j=1}^{4} f_j(\Xi(\boldsymbol{m})_j + \tilde{\boldsymbol{r}}_j) \Bigr) F(\Xi(\boldsymbol{m}) + \tilde{\boldsymbol{r}}) 1_{[-10,10]^2}(L(\Xi\boldsymbol{m} + \tilde{\boldsymbol{r}})) \tag{2-11}$$

for some different vector $\tilde{\boldsymbol{r}} \in \mathbb{Z}^4$, where $\Xi(\boldsymbol{m})_j$ denotes the j-th coordinate of $\Xi(\boldsymbol{m})$. Now, replace the convex cutoff function $1_{[-10,10]^2}$ with some Lipschitz minorant $\tilde{G}$ which is supported on $[-10+\delta, 10-\delta]^2$ and equal to 1 on $[-10+2\delta, 10-2\delta]^2$, in each of the three expressions (2-11) separately. Then the size of these expressions *will* stay roughly constant.

To quantify this step, the approximation function A_L enters the picture. Indeed, if $\tilde{\boldsymbol{r}} = \boldsymbol{0}$ the error term introduced by applying such an approximation to (2-11) is bounded above by

$$T^{L\Xi}_{F,G^*,N}(1, \ldots, 1),$$

where G^* is some other Lipschitz function supported on

$$\{\boldsymbol{x} \in \mathbb{R}^2 : 10 - 2\delta \leqslant \|\boldsymbol{x}\|_\infty \leqslant 10 + 2\delta\}.$$

Finding an upper bound on expressions such as $T^{L\Xi}_{F,G^*,N}(1, \ldots, 1)$ is exactly the endeavour we discussed in Section 2B, when motivating the introduction of the approximation function A_L. The only difference is that now we are dealing with the function $A_{L\Xi}$, rather than A_L.

It turns out that the map $L\Xi$ is most naturally viewed as a map from $\mathbb{R}^3$ into a one dimensional space, i.e.,

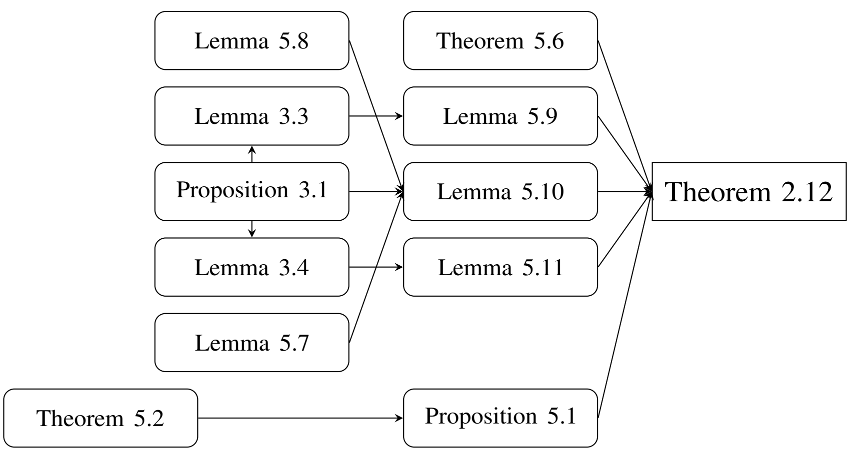
$$L\Xi : \mathbb{R}^3 \to \left\{ \boldsymbol{x} \in \mathbb{R}^2 : \boldsymbol{x} \cdot \binom{5}{1} = 0 \right\},$$

whereas L maps from $\mathbb{R}^4$ to $\mathbb{R}^2$. This is the "dimension reduction" which gives Section 5 its name.

The reader will have noticed that, after replacing $1_{[-10,10]^2}$ with the Lipschitz function $\tilde{G}$, expression (2-11) is not equal to an expression of the form $T^{L}_{F,\tilde{G},N}(f_1, \dots, f_4)$, since the map Ξ and the shift $\tilde{\boldsymbol{r}}$ are now both on the scene. This complicates matters in the second half of the proof, and thus Theorem 5.6 will not be exactly the same statement as Theorem 2.12 apart from the Lipschitz cutoffs. Rather, Theorem 5.6 will bound an object that we will come to denote by $T^{L,\Xi,\tilde{\boldsymbol{r}}}_{F,G,N}(f_1, \dots, f_d)$, which will be a general version of expression (2-11). The reader may consult Definition 5.3 for the full definition.

In order to make this argument rigorous we will have to verify that in replacing the map L with the map $L\Xi$ we haven't introduced any extra rational relations;[10] to work out how to relate A_L and $A_{L\Xi}$; and to work out how to identify a suitable Ξ in the general case. Furthermore, we will have to carry the nondegeneracy relations (such as $\operatorname{dist}(L, V^*_{\mathrm{degen}}(m, d)) \geqslant c$) through this reparametrisation by Ξ, and then establish what the new nondegeneracy notions should be for the pairs $(\Xi, L\Xi)$. This is all done in the (somewhat alarming) Lemma 5.10, which has 9 parts. The upper bounds on expressions like $T^{L\Xi}_{F,G^*,N}(1, \dots, 1)$ are established earlier, in Lemma 3.4, with everything combined at the end of Section 5.

The diagram of the dependency of the various lemmas — excluding those which are found in Appendices A, B and D, which are somewhat standard — is as follows:



It remains to resolve Theorem 5.6, and it turns out that this second part of the proof is significantly more straightforward than the first. In particular neither the statement of Theorem 5.6 nor its proof make any reference to the rational dimension of L nor to the approximation function A_L.

[10] This is essentially the statement that $L\Xi$ should be purely irrational.

The idea is as follows. For a function $f : [N] \to [-1, 1]$ and a small parameter η, let $\tilde{f} : \mathbb{R} \to [-1, 1]$ denote the function

$$\tilde{f}(x) = \begin{cases} f(n) & \text{if } |n-x| \leqslant \eta, \\ 0 & \text{otherwise,} \end{cases}$$

i.e., $\tilde{f}$ is a "fattened" version of f. Then, for Lipschitz functions F and G, let

$$\tilde{T}^L_{F,G,N}(\tilde{f}_1, \dots, \tilde{f}_d) := \frac{1}{N^{d-m}} \int_{\boldsymbol{x}\in\mathbb{R}^d} \biggl(\prod_{j=1}^d \tilde{f}_j(x_j)\biggr) F(\boldsymbol{x}) G(L\boldsymbol{x})\, d\boldsymbol{x}$$

represent the "real solution density" for the inequality weighted by the functions $\tilde{f}_j$. The expression $\tilde{T}^L_{F,G,N}(\tilde{f}_1, \dots, \tilde{f}_d)$ is more convenient to work with than $T^L_{F,G,N}(f_1, \dots, f_d)$, as we are now working in a setting in which the coefficients of L are invertible.[11]

The expression $\tilde{T}^L_{F,G,N}(\tilde{f}_1, \dots, \tilde{f}_d)$ enjoys the following two properties. Firstly, it is closely related to $T^L_{F,G,N}(f_1, \dots, f_d)$. Indeed, just by expanding out the definition of $\tilde{f}_j$, we see that

$$\begin{aligned} \tilde{T}^L_{F,G,N}(\tilde{f}_1, \dots, \tilde{f}_d) &= \frac{1}{N^{d-m}} \sum_{\boldsymbol{n}\in\mathbb{Z}^d} \biggl(\prod_{j=1}^d f_j(n_j)\biggr) \int_{\boldsymbol{y}\in\mathbb{R}^d} F(\boldsymbol{y}) G(L\boldsymbol{y}) 1_{[-\eta,\eta]^d}(\boldsymbol{y}-\boldsymbol{n})\, d\boldsymbol{y} \\ &\approx \frac{1}{N^{d-m}} \sum_{\boldsymbol{n}\in\mathbb{Z}^d} \biggl(\prod_{j=1}^d f_j(n_j)\biggr) F(\boldsymbol{n}) G(L\boldsymbol{n}) \int_{\boldsymbol{y}\in\mathbb{R}^d} 1_{[-\eta,\eta]^d}(\boldsymbol{y}-\boldsymbol{n})\, d\boldsymbol{y} \\ &\approx (2\eta)^d T^L_{F,G,N}(f_1, \dots, f_d), \end{aligned} \tag{2-12}$$

by using the Lipschitz properties of F and G to replace $F(\boldsymbol{y})$ and $G(L\boldsymbol{y})$ by $F(\boldsymbol{n})$ and $G(L\boldsymbol{n})$ respectively. This analysis is performed rigorously in Section 6, and is the only place in the proof where the Lipschitz property of G is used.

Secondly, $\tilde{T}^L_{F,G,N}(\tilde{f}_1, \dots, \tilde{f}_d)$ may be bounded above by expressions involving the Gowers norms (over the reals) of the functions $\tilde{f}_j$. Indeed, after some small manipulations using the compact support of G, one ends up with the bound

$$|\tilde{T}^L_{F,G,N}(\tilde{f}_1, \dots, \tilde{f}_d)| \ll_G \frac{1}{N^{d-m}} \int_{\substack{\boldsymbol{x}\in\mathbb{R}^d \\ L\boldsymbol{x}=\boldsymbol{0}}} \biggl(\prod_{j=1}^d \tilde{f}_j(x_j)\biggr) F(\boldsymbol{x})\, d\mu(\boldsymbol{x}), \tag{2-13}$$

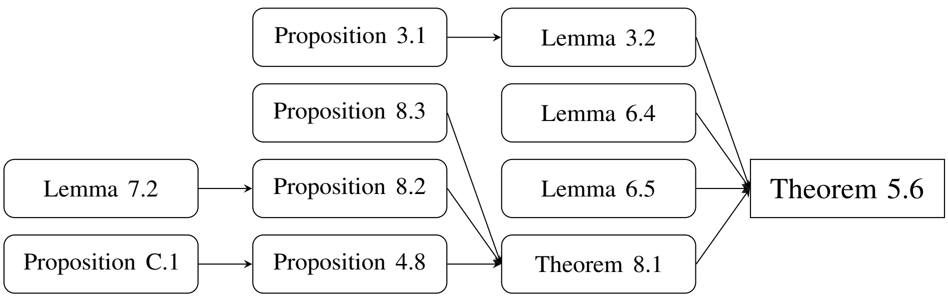
where $\mu(\boldsymbol{x})$ is a suitable measure supported on $\ker L$. The reader will then notice that the right-hand side of (2-13) bears a structural similarity to the expression considered in Theorem 1.2 above, i.e., in the generalised von Neumann theorem for equations with integer coefficients. One may then rejig Green and Tao's proof of Theorem 1.2 to apply in this setting, and thereby bound $\tilde{T}^L_{F,G,N}(\tilde{f}_1, \dots, \tilde{f}_d)$ by the Gowers norms of the functions $\tilde{f}_j$. This is done in Section 8. Finally, there is an elementary argument (Lemma 6.5) that relates the Gowers norms of the functions $\tilde{f}_j$ to the Gowers norms of the original functions f_j, thus completing the proof of our result.

[11] This manoeuvre is somewhat analogous to the device used by Green and Tao [2010a] of passing from $[N]$ to some cyclic group $\mathbb{Z}/N'\mathbb{Z}$, where N' is a prime number larger than N.

As will be familiar to readers of [Green and Tao 2010a], the key manoeuvre in analysing (2-13) is parametrising $\ker L$ in a certain special way (in *normal form*, see Section 4), in order to facilitate repeated applications of the Cauchy–Schwarz inequality. When working over the reals, maintaining quantitative control over the size of the coefficients after this reparametrisation is no longer trivial, and requires the assumption that $\mathrm{dist}(L, V^*_{\mathrm{degen}}(m, d)) \geqslant c$. The details of this piece of quantitative linear algebra are given in Proposition 4.8 and Appendix C. It is this part of the argument which would break down were one to attempt to use Gowers norms to bound inequalities such as the one given by the matrix L in Example 2.11.

We have already remarked that Theorem 5.6 does not just concern the objects $T^L_{F,G,N}(f_1, \dots, f_d)$ but actually concerns the more general objects $T^{L,\Xi,\tilde{\boldsymbol{r}}}_{F,G,N}(f_1, \dots, f_d)$, which are similar to (2-11). This adds an extra veneer of complication, centred largely around the notion of degeneracy for the pair of maps $(\Xi, L\Xi)$. Matters are resolved by a linear algebra argument in Section 7, relating different notions of degeneracy.

The diagram of the dependency of the lemmas used in the proof of Theorem 5.6 is as follows:



The appendices contain some extra material which we felt to be best kept apart from the main narrative. In the case of the first two appendices, they comprise standard results from the literature on Gowers norms and Lipschitz functions, which we include to assist any readers who are unfamiliar with these topics. In the case of Appendices C and D, we present a handful of arguments of a linear algebraic nature which, though perhaps not already present in the literature in the exact form we require, are nonetheless easy to establish. Finally, Appendix E concerns the analysis of the approximation function A_L when L has algebraic coefficients. This argument has a similar flavour to Example 2.7, and is included for the sake of completeness.

3. Upper bounds

This section is devoted to proving three upper bounds on the expression $T^L_{F,G,N}(1, \dots, 1)$. 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 $d \geqslant m+1$. Let c, C be positive constants. Then there are positive constants $D_{c,C}$, $D'_{c,C}$ for which the following holds. Let $L : \mathbb{R}^d \to \mathbb{R}^m$ be a surjective* *linear map, denoted by matrix* $(\lambda_{ij})_{i \leqslant m, j \leqslant d}$, *and assume that* $\|L\|_\infty \leqslant C$ *and* $\mathrm{dist}(L, V_{\mathrm{rank}}(m,d)) \geqslant c$. *Then there exists a matrix* M *that is an* m*-by-*m *submatrix of* L *and enjoys the following properties*:

(1) $|\det M| \geqslant D_{c,C}$.

(2) $\|M^{-1}\|_\infty \leqslant D'_{c,C}$.

We call such a matrix M *a* rank matrix *of* L. *Furthermore*:

(3) *Let* $\boldsymbol{v} \in \mathbb{R}^d$ *be a vector such that* $\boldsymbol{v}^T$ *is in the row-space of* L, *and suppose that* $\|\boldsymbol{v}\|_\infty \leqslant C_1$ *for some positive constant* C_1. *Then for* i *in the range* $1 \leqslant i \leqslant m$ *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* $1 \leqslant j \leqslant d$.

Finally:

(4) *If* L *satisfies the stronger hypothesis* $\mathrm{dist}(L, V_{\mathrm{rank}}^{\mathrm{unif}}(m,d)) \geqslant c$, *then*, *for each* j, *there exists a rank matrix of* L *that does not 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 6.

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

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

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 $m+1 \leqslant j \leqslant d$, let the vector $\boldsymbol{v_j} \in \mathbb{R}^m$ be the j-th column of the matrix $M^{-1}L$. Then $N^{d-m} T^L_{F,G,N}(1,\ldots,1) \leqslant \|G\|_\infty \cdot Z$, where Z is the number of solutions to

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

in which $n_1, \ldots, n_d$ are integers that satisfy $|n_1|, \ldots, |n_d| \leqslant N$. 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* $d \geqslant m+1$, *and let* c, C, ε *be positive constants. Let* $L : \mathbb{R}^d \to \mathbb{R}^m$ *be a surjective linear map*, *and suppose that* $\|L\|_\infty \leqslant C$ *and* $\mathrm{dist}(L, V_{\mathrm{rank}}^{\mathrm{unif}}(m,d)) \geqslant c$. *Let* σ *be a real number in the range* $0 < \sigma < \frac{1}{2}$. *Let* $F : \mathbb{R}^d \to [0,1]$ *and* $G : \mathbb{R}^m \to [0,1]$ *be two functions*, *with* F *supported on*

$$\{\boldsymbol{x} \in \mathbb{R}^d : \mathrm{dist}(\boldsymbol{x}, \partial([1,N]^d)) \leqslant \sigma N\}$$

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

$$T^L_{F,G,N}(1, \dots, 1) \ll_{c,C,\varepsilon} \sigma \|G\|_\infty.$$

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

$$\{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_\infty \leqslant 2N,\ |x_d - 1| \leqslant \sigma N\} \quad \text{or} \quad \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_\infty \leqslant 2N,\ |x_d - N| \leqslant \sigma N\}.$$

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 $m + 1 \leqslant j \leqslant d$, let the vector $\boldsymbol{v_j} \in \mathbb{R}^m$ be the j-th column of the matrix $M^{-1}L$. Then the expression $N^{d-m}T^L_{F,G,N}(1, \dots, 1)$ may be bounded above by $\|G\|_\infty$ times the number of solutions to

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

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

In the second case, the relevant equation is

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

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

Our third estimate is more refined, and will be needed in Section 5 when we replace the sharp cutoff $1_{[-\varepsilon,\varepsilon]^m}$ with a Lipschitz cutoff. For the definition of the approximation function A_L, we refer the reader to Definition 2.6.

Lemma 3.4. *Let N, m, d be natural numbers, with $d \geqslant m + 1$. Let c, C, ε be positive constants, and let σ_G be a parameter in the range $0 < \sigma_G < \frac{1}{2}$. Suppose that $L : \mathbb{R}^d \to \mathbb{R}^m$ is a purely irrational surjective linear map, satisfying $\|L\|_\infty \leqslant C$ and $\operatorname{dist}(L, V_{\mathrm{rank}}(m, d)) \geqslant c$. Let A_L denote the approximation function of L. Let $F : \mathbb{R}^d \to [0, 1]$ be supported on $[-N, N]^d$, and let $G : \mathbb{R}^m \to [0, 1]$ be a Lipschitz function, with Lipschitz constant $O(1/\sigma_G)$, supported on $[-\varepsilon, \varepsilon]^m$. Assume further that $\int_{\boldsymbol{x}} G(\boldsymbol{x})\, d\boldsymbol{x} = O_\varepsilon(\sigma_G)$. Then for all τ_2 in the range $0 < \tau_2 \leqslant 1$,*

$$T^L_{F,G,N}(1, \dots, 1) \ll_{c,C,\varepsilon} \sigma_G + \frac{\tau_2^{1/2}}{\sigma_G} + \frac{\tau_2^{-O(1)} A_L(\Omega_{c,C}(1), \tau_2)^{-1}}{N}.$$

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

$$T^L_{F,G,N}(1,\dots,1) \ll_{c,C,\varepsilon} \frac{1}{N^{d-m}} \sum_{\substack{n_{m+1},\dots,n_d\in\mathbb{Z}\\ |n_{m+1}|,\dots,|n_d|\leqslant N}} \tilde{G}\Big(\sum_{j=m+1}^{d} \boldsymbol{v}_j n_j\Big), \tag{3-1}$$

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

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

It remains to estimate the right-hand side of (3-1).

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

$$\int_{\boldsymbol{x}\in[0,1)^m} \tilde{G}(\boldsymbol{x})\,d\boldsymbol{x} = \int_{\boldsymbol{x}\in\mathbb{R}^m} (G\circ M)(\boldsymbol{x})\,d\boldsymbol{x} = O_{c,C,\varepsilon}(\sigma_G).$$

By [Green and Tao 2008b, Lemma A.9], which we recall in Lemma B.3, for any X at least 2 we may write

$$\tilde{G}(\boldsymbol{x}) = \sum_{\substack{\boldsymbol{k}\in\mathbb{Z}^m\\ \|\boldsymbol{k}\|_\infty\leqslant X}} b_X(\boldsymbol{k})e(\boldsymbol{k}\cdot\boldsymbol{x}) + O_{c,C,\varepsilon}\Big(\frac{\log X}{\sigma_G X}\Big), \tag{3-2}$$

where $b_X(\boldsymbol{k})\in\mathbb{C}$ and satisfies $|b_X(\boldsymbol{k})| = O(1)$. Moreover $b_X(\boldsymbol{0}) = \int_{\boldsymbol{x}\in[0,1)^m} \tilde{G}(\boldsymbol{x})\,d\boldsymbol{x}$.[12]

Returning to (3-1), we see that for any X at least 2 we may write

$$T^L_{F,G,N}(1,\dots,1) \ll_{c,C,\varepsilon} \sigma_G + \frac{\log X}{\sigma_G X} + X^{O(1)} \max_{\substack{\boldsymbol{k}\in\mathbb{Z}^m\\ 0<\|\boldsymbol{k}\|_\infty\leqslant X}} \Big(\prod_{j=m+1}^{d} \min(1, N^{-1}\|\boldsymbol{k}\cdot\boldsymbol{v}_j\|^{-1}_{\mathbb{R}/\mathbb{Z}})\Big), \tag{3-3}$$

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 (3-3) to the approximation function A_L. Since L is purely irrational,

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

[12]This final fact is not given explicitly in the statement of [Green and Tao 2008b, Lemma A.9], although it is given in the proof. In any case, it may be immediately deduced from (3-2), by letting X tend to infinity and integrating (3-2) over all $\boldsymbol{x}\in\mathbb{R}^m/\mathbb{Z}^m$.

Let τ_2 be in the range $0 < \tau_2 \leqslant 1$. Then there exist positive parameters D and D', depending only on c and C, for which

$$\begin{aligned}\min_{\substack{\boldsymbol{k}\in\mathbb{Z}^m\\ 0<\|\boldsymbol{k}\|_\infty\leqslant D\tau_2^{-1}}} \max(\{\|\boldsymbol{k}\cdot\boldsymbol{v_j}\|_{\mathbb{R}/\mathbb{Z}} : m+1\leqslant j\leqslant d\}) &= \min_{\substack{\boldsymbol{k}\in\mathbb{Z}^m\\ 0<\|\boldsymbol{k}\|_\infty\leqslant D\tau_2^{-1}}} \operatorname{dist}(\boldsymbol{k}^T M^{-1}L, (\mathbb{Z}^d)^T)\\ &\geqslant \inf_{\substack{\boldsymbol{k}\in\mathbb{R}^m\\ 1\leqslant\|\boldsymbol{k}\|_\infty\leqslant D\tau_2^{-1}}} \operatorname{dist}(\boldsymbol{k}^T M^{-1}L, (\mathbb{Z}^d)^T)\\ &\geqslant \inf_{\substack{\boldsymbol{k}\in\mathbb{R}^m\\ D'\leqslant\|\boldsymbol{k}\|_\infty\leqslant \tau_2^{-1}}} \operatorname{dist}(\boldsymbol{k}^T L, (\mathbb{Z}^d)^T)\\ &= A_L(D',\tau_2). \end{aligned} \tag{3-4}$$

Letting $X = D\tau_2^{-1}$, and substituting the bound (3-4) into (3-3), one derives

$$T^L_{F,G,N}(1,\dots,1) \ll_{c,C,\varepsilon} \sigma_G + \frac{\tau_2^{1/2}}{\sigma_G} + \frac{\tau_2^{-O(1)} A_L(\Omega_{c,C}(1),\tau_2)^{-1}}{N}$$

as required. □

The relations (3-4) formalise the estimate (2-2), which we first discussed when introducing the approximation function A_L in Definition 2.6. With the details all here, one can now see that it would have been enough to define the approximation function, at least if L is purely irrational, to be the function

$$\tau_2 \mapsto \min_{\substack{\boldsymbol{k}\in\mathbb{Z}^m\\ 0<\|\boldsymbol{k}\|_\infty\leqslant \tau_2^{-1}}} \operatorname{dist}(\boldsymbol{k}^T M^{-1}L, (\mathbb{Z}^d)^T). \tag{3-5}$$

One might now be concerned that, in defining A_L using real vectors φ rather than integer vectors $\boldsymbol{k}$, we might have constructed a much weaker object than (3-5), making (3-4) a wasteful step in our estimation. This is not the case, because if $\varphi \in (\mathbb{R}^m)^*$ and

$$\operatorname{dist}(L^*\varphi, (\mathbb{Z}^d)^T) \leqslant \delta$$

then $\operatorname{dist}(L^*(M^*)^{-1}M^*\varphi, (\mathbb{Z}^d)^T) \leqslant \delta$, and so in particular $\operatorname{dist}(M^*\varphi, (\mathbb{Z}^m)^T) \leqslant \delta$ (as M is a rank matrix). Letting $\boldsymbol{k}^T \in (\mathbb{Z}^m)^T$ be the nearest integer vector to $M^*\varphi$, we have that

$$\operatorname{dist}(\boldsymbol{k}^T M^{-1}L, (\mathbb{Z}^d)^T) \ll_{c,C} \delta.$$

So, up to some constants depending on c and C, there is essentially no difference between working with Definition 2.6 or with (3-5).

Restricting to integer vectors $\boldsymbol{k}$ may seem more natural from the point of view of diophantine approximation, but on the other hand the expression (3-5) depends on the choice of the particular rank matrix M, which is not canonical. It was more to our taste to present a definition of A_L which was intrinsic to L. Lemma 8.1 of our follow-up paper [Walker 2019] is also a setting in which having real vectors in the definition of A_L seems to be more natural.

It is also worth highlighting the exact moment in the proof of Lemma 3.4 in which it was vital that L was purely irrational. Considering expression (3-3), if L was not purely irrational and X was bigger than the rational complexity of L then the final error term is just $X^{O(1)}$, which is not $o(1)$ as $N \to \infty$.

4. Normal form

In this section we recall a technical notion from [Green and Tao 2010a] 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 $(\psi_1, \dots, \psi_m) := \Psi : \mathbb{R}^n \to \mathbb{R}^m$ as a system of homogeneous linear forms. The crux of the theory from [Green and Tao 2010a] is that, provided Ψ is of so-called "finite Cauchy–Schwarz complexity", Ψ 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 [Green and Tao 2010a, Section 1] and [Gowers and Wolf 2010].

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

$$\begin{aligned} \psi_1(t,u,v) &= u+v \\ \psi_2(t,u,v) &= v+t \\ \psi_3(t,u,v) &= u+t \\ \psi_4(t,u,v) &= u+v+t \end{aligned} \tag{4-1}$$

are in normal form with respect to ψ_4, since ψ_4 is the only form to utilise all three of the variables. However, this system is not in normal form with respect to ψ_3, say. However, the system

$$\begin{aligned} \psi_1(t,u,v,w) &= u+v+2w \\ \psi_2(t,u,v,w) &= v+t-w \\ \psi_3(t,u,v,w) &= u+t-w \\ \psi_4(t,u,v,w) &= u+v+t, \end{aligned} \tag{4-2}$$

which parametrises the same subspace of $\mathbb{R}^4$, *is* in normal form for all i.

We repeat the precise definition from [Green and Tao 2010a].

Definition 4.1. Let m, n be natural numbers, and let $(\psi_1, \dots, \psi_m) = \Psi : \mathbb{R}^n \to \mathbb{R}^m$ be a system of homogeneous linear forms. Let $i \in [m]$. We say that Ψ is in normal form with respect to ψ_i if there exists a nonnegative integer s and a collection $J_i \subseteq \{\boldsymbol{e_1}, \dots, \boldsymbol{e_n}\}$ of the standard basis vectors, satisfying $|J_i| = s+1$, such that

$$\prod_{\boldsymbol{e} \in J_i} \psi_{i'}(\boldsymbol{e})$$

is nonzero when $i' = i$ and vanishes otherwise. We say that Ψ is in normal form if it is in normal form with respect to ψ_i for every i.

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

Definition 4.2. For a system of homogeneous linear forms $(\psi_1, \ldots, \psi_m) = \Psi : \mathbb{R}^n \to \mathbb{R}^m$, an extension $(\psi'_1, \ldots, \psi'_m) = \Psi' : \mathbb{R}^{n'} \to \mathbb{R}^m$ is a system of homogeneous linear forms on $\mathbb{R}^{n'}$, for some n' with $n' \geqslant n$, such that

(1) $\Psi'(\mathbb{R}^{n'}) = \Psi(\mathbb{R}^n)$;

(2) if we identify $\mathbb{R}^n$ with the subset $\mathbb{R}^n \times \{0\}^{n'-n}$ in the obvious manner, then Ψ is the restriction of Ψ' to this subset.

The paper [Green and Tao 2010a] 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 (4-1) and (4-2) 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 notion of finite Cauchy–Schwarz complexity introduced in [Green and Tao 2010a].[13] 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 4.3 (suitable partitions). Let m, n be natural numbers, with $m \geqslant 2$, and let $(\psi_1, \ldots, \psi_m) = \Psi : \mathbb{R}^n \to \mathbb{R}^m$ be a system of homogeneous linear forms. Fix $i \in [m]$. Let $\mathcal{P}_i$ be a partition of $[m] \setminus \{i\}$, i.e.,

$$[m] \setminus \{i\} = \bigcup_{k=1}^{s+1} \mathcal{C}_k$$

for some s satisfying $0 \leqslant s \leqslant m-2$ and some disjoint sets $\mathcal{C}_k$. We say that $\mathcal{P}_i$ is *suitable* for Ψ if

$$\psi_i \notin \operatorname{span}_{\mathbb{R}}(\psi_j : j \in \mathcal{C}_k)$$

for any k.

Definition 4.4 (degeneracy varieties). Let m, n be natural numbers, with $m \geqslant 2$. 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 Ψ. Finally, the *degeneracy variety $V_{\text{degen}}(n, m)$* is given by

$$V_{\text{degen}}(n, m) := \bigcup_{i=1}^{m} \bigcap_{\mathcal{P}_i} V_{\mathcal{P}_i},$$

where the inner intersection is over all possible partitions $\mathcal{P}_i$.

[13] In [Green and Tao 2010a] this is just called "complexity".

It is easy to observe that $\Psi \in V_{\mathrm{degen}}(n,m)$ if and only if, for some $i \neq j$, ψ_i is a real multiple of ψ_j. This also yields the following:

Proposition 4.5 (relating $V^*_{\mathrm{degen}}(m,d)$ and $V_{\mathrm{degen}}(d-m,d)$). *Let m, d be natural numbers with $d \geqslant m+2$, and let $L : \mathbb{R}^d \to \mathbb{R}^m$ be a surjective linear map. Let $\Psi : \mathbb{R}^{d-m} \to \mathbb{R}^d$ be any system of homogeneous linear forms whose image is* $\ker L$. *Then $L \in V^*_{\mathrm{degen}}(m,d)$ if and only if $\Psi \in V_{\mathrm{degen}}(d-m,d)$.*

Proof. We know that $L \in V^*_{\mathrm{degen}}(m,d)$ if and only if there exist some nonzero vector $\boldsymbol{e}^*_i - \lambda \boldsymbol{e}^*_j \in L^*((\mathbb{R}^m)^*)$. But $L^*((\mathbb{R}^m)^*) = (\ker L)^0 = (\Psi(\mathbb{R}^{d-m}))^0$, so this occurs if and only if $\psi_i = \lambda \psi_j$ for some i and j. □

We will prove a more general version of this statement in Lemma 7.1.

Green and Tao [2010a, Definition 1.5] refer to those $\Psi \in V_{\mathrm{degen}}(n,m)$ as having infinite Cauchy–Schwarz complexity, and develop their theory for $\Psi \notin V_{\mathrm{degen}}(n,m)$. As we did for describing degeneracy properties of L, we need to quantify such a notion.

Definition 4.6 (c_1-Cauchy–Schwarz complexity). Let m, n be natural numbers, with $m \geqslant 3$, and let c_1 be a positive constant. Let $(\psi_1, \dots, \psi_m) = \Psi : \mathbb{R}^n \to \mathbb{R}^m$ 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 $\mathrm{dist}(\Psi, V_{\mathcal{P}_i}) \geqslant c_1$, 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 Ψ.

We remark, for readers familiar with [Green and Tao 2010a], 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 $s+1 \geqslant 2$. This is why we need the condition $m \geqslant 3$ in the above definition. We also take this opportunity to note that if s satisfies the above definition, and $s \neq \infty$, then $2 \leqslant s+1 \leqslant m-1$.

We note an easy consequence of these definitions.

Lemma 4.7. *Let m and n be natural numbers, with $m \geqslant 3$, and let c_1 be a positive constant. Let $(\psi_1, \dots, \psi_m) = \Psi : \mathbb{R}^n \to \mathbb{R}^m$ be a system of homogeneous linear forms. Furthermore, suppose that* $\mathrm{dist}(\Psi, V_{\mathrm{degen}}(n,m)) \geqslant c_1$. *Then Ψ has finite c_1-Cauchy–Schwarz complexity.*

Proof. We have already observed that $\Psi \in V_{\mathrm{degen}}(n,m)$ if and only if, for some $i \neq j$, ψ_i is a real multiple of ψ_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 ψ_k is in its own part. Our initial observation then implies that $\Psi \in V_{\mathrm{degen}}(n,m)$ if and only if $\Psi \in V_{\mathcal{P}_i}$ for some i. So $\mathrm{dist}(\Psi, V_{\mathrm{degen}}(n,m)) \geqslant c_1$ implies that $\mathrm{dist}(\Psi, V_{\mathcal{P}_i}) \geqslant c_1$ for all i. Therefore, by using these partitions $\mathcal{P}_i$ in Definition 4.6, we conclude that Ψ has finite c_1-Cauchy–Schwarz complexity. □

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

(1) $n' = n + s + 1 \leqslant n + m - 1$.

(2) Ψ' *is of the form*

$$\Psi'(\boldsymbol{u}, w_1, \dots, w_{s+1}) := \Psi(\boldsymbol{u} + w_1 \boldsymbol{f_1} + \cdots + w_{s+1} \boldsymbol{f_{s+1}})$$

for some vectors $\boldsymbol{f_k} \in \mathbb{R}^n$, *such that* $\|\boldsymbol{f_k}\|_\infty = O_{c_1,C_1}(1)$ *for every* k.

(3) Ψ' *is in normal form with respect to* ψ'_i.

(4) $\psi'_i(\boldsymbol{0}, \boldsymbol{w}) = w_1 + \cdots + w_{s+1}$.

The proof is deferred to Appendix C, as it is very similar to the proof from [Green and Tao 2010a] (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

$$\begin{aligned} \psi_1(u_1, u_2, u_3) &= (1 + \iota(N)^{-1})u_1 + u_2 \\ \psi_2(u_1, u_2, u_3) &= u_1 + u_2 \\ \psi_3(u_1, u_2, u_3) &= u_3. \end{aligned}$$

and let $\Psi := (\psi_1, \psi_2, \psi_3)$. Notice that $\operatorname{dist}(\Psi, V_{\text{degen}}(3, 3)) \to 0$ as $N \to \infty$. Therefore, for any $c_1 > 0$, if N is large enough then Ψ does not have finite c_1-Cauchy–Schwarz complexity. One may nonetheless construct a normal form reparametrisation

$$\begin{aligned} \psi'_1(u_1, u_2, u_3, w_1, w_2) &= (1 + \iota(N)^{-1})u_1 + u_2 + w_1 \\ \psi'_2(u_1, u_2, u_3, w_1, w_2) &= u_1 + u_2 + w_2 \\ \psi'_3(u_1, u_2, u_3, w_1, w_2) &= u_3. \end{aligned}$$

However, since

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

Ψ' is not obtained by bounded shifts of the u_i variables, and so (if N is large enough) it fails to satisfy part (2) of the conclusion of the above proposition. Such an extension Ψ' would not be suitable for our requirements in Section 8.

Remark 4.9. In [Green and Tao 2010a], 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 [loc. cit.] demonstrates that it is sufficient for Ψ merely to admit, for each i separately, an extension that is in normal form with respect to ψ_i, but this is of little consequence in [loc. cit.]. 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 4.8 for normal forms only with respect to a single i, in order to avoid this technical annoyance.

5. Dimension reduction

As we described in our proof strategy (Section 2E), in this section we reduce Theorem 2.12 to a different result, namely Theorem 5.6. This second theorem will be simpler in one key respect: the replacement of sharp cutoffs by Lipschitz cutoffs. It is the proof of Theorem 5.6 in which the Lipschitz property is actually used, and this will begin in Section 6. Any reader only wishing to consider the case of diophantine inequalities with Lipschitz cutoffs may eschew Section 5 of this paper entirely.

We begin by dismissing the case of maximal rational dimension.

Proposition 5.1. *Theorem 2.12 holds under the additional assumption that L has rational dimension m.*

To prove this, we will appeal to a quantitative version of Theorem 1.2.

Theorem 5.2 (generalised von Neumann theorem for rational forms (quantitative version)). *Let N, m, d be natural numbers, satisfying $d \geqslant m+2$, and let C_1, C_2 be positive constants. Let $S = S(N)$ be an m-by-d matrix with integer coefficients, satisfying $\|S\|_\infty \leqslant C_1$, and let $\boldsymbol{r} \in \mathbb{Z}^m$ be some vector with $\|\boldsymbol{r}\|_\infty \leqslant C_2 N$. Suppose S has rank m, and $S \notin V^*_{\text{degen}}(m,d)$. Let $K \subseteq [-N,N]^d$ be convex. Then there exists some natural number s at most $d-2$ that satisfies the following. Let $f_1, \dots, f_d : [N] \to \mathbb{C}$ be arbitrary functions with $\|f_j\|_\infty \leqslant 1$ for all j, and assume that*

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

for some ρ in the range $0 < \rho \leqslant 1$. Then

$$\frac{1}{N^{d-m}} \sum_{\substack{\boldsymbol{n} \in \mathbb{Z}^d \cap K \\ S\boldsymbol{n} = \boldsymbol{r}}} \prod_{j=1}^{d} f_j(n_j) \ll_{C_1,C_2} \rho^{\Omega(1)} + o_\rho(1)$$

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

Let us sketch a proof of this result, assuming a certain familiarity with the methods and terminology of [Green and Tao 2010a].

Proof sketch of Theorem 5.2. One follows the proof of Theorem 1.8 of [Green and Tao 2010a]. Firstly, recall that in our language, the nondegeneracy condition in the statement of Theorem 1.8 of [loc. cit.] is exactly the condition that $S \notin V^*_{\text{degen}}(m,d)$. One then follows the same linear algebraic reductions as those used in Section 4 of [loc. cit.] 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 [Tao 2012, 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 5.2, where s is the Cauchy–Schwarz complexity of some system of forms $(\psi_1, \dots, \psi_d)$ 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 5.2 is proved. □

Now let us use Theorem 5.2 to resolve Proposition 5.1.

Proof of Proposition 5.1. Let L be as in Theorem 2.12, and assume that L has rational dimension m and rational complexity at most C. Let $\Theta : \mathbb{R}^m \to \mathbb{R}^m$ be some linear isomorphism satisfying $\Theta L(\mathbb{Z}^d) \subseteq \mathbb{Z}^m$ and $\|\Theta\|_\infty \leqslant C$. 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 $((\Theta M) \circ (M^{-1}L))(\mathbb{Z}^d) = \Theta L(\mathbb{Z}^d) \subseteq \mathbb{Z}^m$. The matrix $M^{-1}L$ also has rational complexity $O_{c,C}(1)$. Therefore, replacing L with $M^{-1}L$, we may assume that the first m columns of L form the identity matrix.

As in Lemma 2.5, we write $\Theta L = S$, where S has integer coefficients and $\|\Theta\|_\infty \ll_{c,C} 1$. Hence $\|S\|_\infty \ll_{c,C} 1$. But Θ must also have integer coefficients, as the first m columns of L form the identity matrix, and hence $\|\Theta^{-1}\|_\infty \ll_{c,C} 1$ as well. Note finally that $S \notin V^*_{\mathrm{degen}}(m,d)$, since $L \notin V^*_{\mathrm{degen}}(m,d)$.

Now, suppose that $G : \mathbb{R}^m \to [0,1]$ is the indicator function of some convex domain D, with $D \subseteq [-\varepsilon, \varepsilon]^m$. Then there are at most $O_{c,C,\varepsilon}(1)$ possible vectors $\boldsymbol{r} \in \mathbb{Z}^m$ such that $\boldsymbol{r} \in S(\mathbb{Z}^d) \cap \Theta(D)$. Let R be the set of all such vectors. Therefore, with F being the indicator function of the set $[1,N]^d$, we have

$$T^L_{F,G,N}(f_1, \dots, f_d) = \sum_{\boldsymbol{r} \in R} \sum_{\substack{\boldsymbol{n} \in [N]^d \\ S\boldsymbol{n} = \boldsymbol{r}}} \prod_{j=1}^d f_j(n_j) \ll_{c,C,\varepsilon} \rho^{\Omega(1)} + o_\rho(1) \tag{5-1}$$

as $N \to \infty$, by Theorem 5.2. The $o_\rho(1)$ term may be bounded above by $\rho^{-O(1)}N^{-\Omega(1)}$. This is the desired conclusion of Theorem 2.12 in the case when L has rational dimension m. □

Having dismissed this case, we prepare to state Theorem 5.6. We begin with a definition that generalises Definition 2.1.

Definition 5.3. Let N, m, d, h be natural numbers, with $d \geqslant h \geqslant m+2$. Let ε be positive. Let $\Xi = (\xi_1, \dots, \xi_d) : \mathbb{R}^h \to \mathbb{R}^d$ 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 compactly supported. Let $\tilde{\boldsymbol{r}} \in \mathbb{Z}^d$ be some vector, and let $f_1, \dots, f_d : \mathbb{R} \to [-1,1]$ be arbitrary functions. We then define

$$T^{L,\Xi,\tilde{\boldsymbol{r}}}_{F,G,N}(f_1, \dots, f_d) := \frac{1}{N^{h-m}} \sum_{\boldsymbol{n} \in \mathbb{Z}^h} \biggl(\prod_{j=1}^d f_j(\xi_j(\boldsymbol{n}) + \tilde{\boldsymbol{r}}_j) \biggr) F(\boldsymbol{n}) G(L\boldsymbol{n}). \tag{5-2}$$

In the paper so far we have introduced many degeneracy relations (Definitions 2.9, 2.10, 4.4). In order to state Theorem 5.6, we must introduce another.

Definition 5.4 (dual pair degeneracy variety). Let m, d, h be natural numbers satisfying $d \geqslant h \geqslant m+2$. Let $\boldsymbol{e_1}, \dots, \boldsymbol{e_d}$ denote the standard basis vectors of $\mathbb{R}^d$, and let $\boldsymbol{e_1^*}, \dots, \boldsymbol{e_d^*}$ denote the dual basis of $(\mathbb{R}^d)^*$. Then let $V^*_{\text{degen},2}(m, d, h)$ denote the set of all pairs of linear maps $\Xi : \mathbb{R}^h \to \mathbb{R}^d$ and $L : \mathbb{R}^h \to \mathbb{R}^m$ for which there exist two indices $i, j \leqslant d$, and some real number λ, such that $(\boldsymbol{e_i^*} - \lambda \boldsymbol{e_j^*})$ is nonzero and $\Xi^*(\boldsymbol{e_i^*} - \lambda \boldsymbol{e_j^*}) \in L^*((\mathbb{R}^m)^*)$. We call $V^*_{\text{degen},2}(m, d, h)$ the *dual pair degeneracy variety*.

One can motivate this definition as follows. We noted in Proposition 4.5 that, if $L : \mathbb{R}^d \to \mathbb{R}^m$ is a surjective linear map, then saying that $L \notin V^*_{\text{degen}}(m, d)$ is equivalent to saying that any parametrisation $\Psi = (\psi_1, \dots, \psi_d) : \mathbb{R}^{d-m} \to \mathbb{R}^d$ of $\ker L$ has finite Cauchy–Schwarz complexity. In this paper, following our sketched idea in expression (2-11), we will end up needing to replace the map L with two maps, an injective map $\Xi : \mathbb{R}^{d-u} \to \mathbb{R}^d$ and a purely irrational surjective map $L' : \mathbb{R}^{d-u} \to \mathbb{R}^{m-u}$ (here u will be the rational dimension of L). It will turn out that after this manipulation the system of forms that we will require to have finite Cauchy–Schwarz complexity (in order to bring in Gowers norms) will be $\Xi\Psi' : \mathbb{R}^{d-m} \to \mathbb{R}^d$, where $\Psi' : \mathbb{R}^{d-m} \to \mathbb{R}^{d-u}$ is a parametrisation of $\ker L'$. One can easily show (and we do, in Lemma 7.1), that $(\Xi, L') \notin V^*_{\text{degen}}(m-u, d, d-u)$ is the exactly the right condition to ensure that $\Xi\Psi' : \mathbb{R}^{d-m} \to \mathbb{R}^d$ has finite Cauchy–Schwarz complexity.

As ever, we need a quantitative version of nondegeneracy.

Definition 5.5 (distance metric for pairs of matrices). Let m, d, h be natural numbers, with $d \geqslant h \geqslant m+2$, and let $V^*_{\text{degen},2}(m, d, h)$ be the dual pair degeneracy variety. Let $\Xi : \mathbb{R}^h \to \mathbb{R}^d$ and $L : \mathbb{R}^h \to \mathbb{R}^m$ be linear maps. We say that $\operatorname{dist}((\Xi, L), V^*_{\text{degen},2}(m, d, h)) \geqslant c$ if $(\Xi + Q, L) \notin V^*_{\text{degen},2}(m, d, h)$ for all $Q : \mathbb{R}^h \to \mathbb{R}^d$ with $\|Q\|_\infty < c$.

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 our theorem on linear inequalities with Lipschitz cutoffs.

Theorem 5.6 (Lipschitz case). *Let N, m, d, h be natural numbers, with $d \geqslant h \geqslant m+2$, and let c, C, ε be positive constants. Let $\Xi = \Xi(N) : \mathbb{R}^h \to \mathbb{R}^d$ be an injective linear map with integer coefficients, and assume that $\Xi(\mathbb{Z}^h) = \mathbb{Z}^d \cap \operatorname{im} \Xi$. Let $L = L(N) : \mathbb{R}^h \to \mathbb{R}^m$ be a surjective linear map. Assume that $\|\Xi\|_\infty \leqslant C$, $\|L\|_\infty \leqslant C$, $\operatorname{dist}(L, V_{\text{rank}}(m, d)) \geqslant c$ and $\operatorname{dist}((\Xi, L), V^*_{\text{degen},2}(m, d, h)) \geqslant c$. Then there exists a natural number s at most $d-2$, independent of ε, such that the following holds. Let σ_F, σ_G be any two parameters in the range $0 < \sigma_F, \sigma_G < \frac{1}{2}$. 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 a Lipschitz function supported on $[-\varepsilon, \varepsilon]^m$ with Lipschitz constant $O(1/\sigma_G)$. Let $\tilde{\boldsymbol{r}}$ be a fixed vector in $\mathbb{Z}^d$, satisfying $\|\tilde{\boldsymbol{r}}\|_\infty = O_{c,C,\varepsilon}(1)$. Suppose that $f_1, \dots, f_d : [N] \to [-1, 1]$ are arbitrary bounded functions satisfying*

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

for some ρ in the range $0 < \rho \leqslant 1$. Then

$$T^{L,\Xi,\tilde{\boldsymbol{r}}}_{F,G,N}(f_1, \dots, f_d) \ll_{c,C,\varepsilon} \rho^{\Omega(1)}(\sigma_F^{-O(1)} + \sigma_G^{-O(1)}) + \sigma_F^{-O(1)} N^{-\Omega(1)}. \tag{5-3}$$

Although the above theorem contains more technical conditions than even Theorem 2.12 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 (5-3).

As we described in Section 2E, the presence of Lipschitz cutoffs rather than convex cutoffs will be especially convenient when approximating the discrete solution count by a continuous solution count. This will be done in Section 6.

The remainder of this section will be devoted to proving the main theorem (Theorem 2.12), assuming the truth of Theorem 5.6.

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 5.10.

Lemma 5.7 (parametrising the image lattice). *Let u, d be integers with $d \geqslant u+1$. Let $S : \mathbb{R}^d \to \mathbb{R}^u$ be a surjective linear map with $S(\mathbb{Z}^d) \subseteq \mathbb{Z}^u$, and suppose that $\|S\|_\infty \leqslant C$. Then there exists a set $\{\boldsymbol{a_1}, \dots, \boldsymbol{a_u}\} \subset \mathbb{Z}^u$ that is a basis for the lattice $S(\mathbb{Z}^d)$ and for which $\|\boldsymbol{a_i}\|_\infty = O_C(1)$ for every i. Furthermore there exist $\boldsymbol{x_1}, \dots, \boldsymbol{x_u} \in \mathbb{Z}^d$ such that, for every i, $S(\boldsymbol{x_i}) = \boldsymbol{a_i}$ and $\|\boldsymbol{x_i}\|_\infty = O_C(1)$.*

Proof. The lattice $S(\mathbb{Z}^d)$ is u dimensional, as S is surjective. If $\{\boldsymbol{e_j} : j \leqslant d\}$ denotes the standard basis of $\mathbb{R}^d$ then integer combinations of elements from the set $\{S(\boldsymbol{e_j}) : j \leqslant d\}$ span $S(\mathbb{Z}^d)$. Since $\|S\|_\infty \leqslant C$, these vectors also satisfy $\|S(\boldsymbol{e_j})\|_\infty = O_C(1)$. Therefore the u successive minima of the lattice $S(\mathbb{Z}^d)$ are all $O_C(1)$, and so, by Mahler's theorem [Tao and Vu 2006, Theorem 3.34] the lattice $S(\mathbb{Z}^d)$ has a basis $\{\boldsymbol{a_1}, \dots, \boldsymbol{a_u}\}$ of the required form.

Note that S has integer coefficients. The construction of suitable $\boldsymbol{x_1}, \dots, \boldsymbol{x_u}$ 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 $\boldsymbol{v_1}, \dots, \boldsymbol{v_{d-u}} \in \mathbb{Z}^d$ whose integer span is a full-dimensional sublattice of the $d-u$ dimensional lattice $\mathbb{Z}^d \cap \ker S$, and that satisfy $\|\boldsymbol{v_i}\|_\infty = O_C(1)$ for all i. Now given some $\boldsymbol{a_i}$, by its construction there must be some $\boldsymbol{x_i} \in \mathbb{Z}^d$ that satisfies $S(\boldsymbol{x_i}) = \boldsymbol{a_i}$. Write $\boldsymbol{x_i} = \boldsymbol{x_i}|_{\ker S} + \boldsymbol{x_i}|_{(\ker S)^\perp}$ as the sum of the obvious projections. By adding a suitable integer combination of the vectors $\boldsymbol{v_1}, \dots, \boldsymbol{v_{d-u}}$ to $\boldsymbol{x_i}$ one may find such an $\boldsymbol{x_i}$ that satisfies $\|\boldsymbol{x_i}|_{\ker S}\|_\infty = O_C(1)$. Furthermore, $\operatorname{dist}(S, V_{\mathrm{rank}}(m, d)) = \Omega_C(1)$, since S has integer coordinates, and so (by Lemma D.1) $\|\boldsymbol{x_i}|_{(\ker S)^\perp}\|_\infty = O_C(1)$. Hence $\|\boldsymbol{x_i}\|_\infty = O_C(1)$, as desired. □

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

Lemma 5.8 (dual space decomposition). *Let u, d, be integers with $d \geqslant u+1$, and let C, η be constants. Let $S : \mathbb{R}^d \to \mathbb{R}^u$ be a surjective linear map with $S(\mathbb{Z}^d) \subseteq \mathbb{Z}^u$, and suppose that $\|S\|_\infty \leqslant C$. Let $\{\boldsymbol{a_1}, \ldots, \boldsymbol{a_u}\}$ be a basis for the lattice $S(\mathbb{Z}^d)$ that satisfies $\|\boldsymbol{a_i}\|_\infty = O_C(1)$ for every i. Let $\boldsymbol{x_1}, \ldots, \boldsymbol{x_u} \in \mathbb{Z}^d$ be vectors such that, for every i, $S(\boldsymbol{x_i}) = \boldsymbol{a_i}$ and $\|\boldsymbol{x_i}\|_\infty = O_C(1)$. Suppose that $\Xi : \mathbb{R}^{d-u} \to \mathbb{R}^d$ is an injective linear map such that* $\operatorname{im} \Xi = \ker S$ *and such that* $\Xi(\mathbb{Z}^{d-u}) = \mathbb{Z}^d \cap \operatorname{im} \Xi$. *Suppose further that $\|\Xi\|_\infty \leqslant C$.*

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

(1) *The set $\mathcal{B} := \{\boldsymbol{x_i} : i \leqslant u\} \cup \{\Xi(\boldsymbol{w_j}) : j \leqslant d-u\}$ is a basis for $\mathbb{R}^d$, and a lattice basis for $\mathbb{Z}^d$.*

(2) *Writing $\mathcal{B}^* := \{\boldsymbol{x_i^*} : i \leqslant u\} \cup \{\Xi(\boldsymbol{w_j})^* : j \leqslant d-u\}$ for the dual basis, both the change of basis matrix between the standard dual basis and $\mathcal{B}^*$ 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 $V :=$ $\operatorname{span}(\boldsymbol{x_i^*} : i \leqslant u)$ *and $W :=$* $\operatorname{span}(\Xi(\boldsymbol{w_j})^* : j \leqslant d-u)$. *Then*:

(3) $V = S^*((\mathbb{R}^u)^*)$.

(4) *Suppose that $\varphi \in (\mathbb{R}^d)^*$ satisfies $\|\Xi^*(\varphi)\|_\infty \leqslant \eta$. 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 $\mathcal{B}$ is a basis for $\mathbb{R}^d$ is just a manifestation of the familiar principle $\mathbb{R}^d \cong \ker S \oplus \operatorname{im} S$. To show that $\mathcal{B}$ is a lattice basis for $\mathbb{Z}^d$, let $\boldsymbol{n} \in \mathbb{Z}^d$ and write

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

for some $\lambda_i, \mu_j \in \mathbb{R}$. Applying S, we see $S(\boldsymbol{n}) = \sum_i^u \lambda_i \boldsymbol{a_i}$, and hence $\lambda_i \in \mathbb{Z}$ for all i, as $\{\boldsymbol{a_1}, \ldots, \boldsymbol{a_u}\}$ is a basis for the lattice $S(\mathbb{Z}^d)$. But this implies $\sum_{j=1}^{d-u} \mu_j \Xi(\boldsymbol{w_j}) \in \mathbb{Z}^d \cap \operatorname{im}(\Xi)$. 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 $\|\Xi^*(\varphi)\|_\infty \leqslant \eta$ means that $|\Xi^*(\varphi)(\boldsymbol{w_j})| \leqslant \eta$. Hence $|\varphi(\Xi(\boldsymbol{w_j}))| \leqslant \eta$. But, writing $\varphi_W = \sum_{j=1}^{d-u} \mu_j \Xi(\boldsymbol{w_j})^*$, this implies that $|\mu_j| \leqslant \eta$. Since the coefficients of the change of basis matrix between $\mathcal{B}^*$ and the standard dual basis are bounded in absolute value by $O_C(1)$, this implies that $\|\varphi_W\|_\infty \leqslant O_C(\eta)$. □

We now begin the attack on Theorem 2.12 in earnest. Assume the hypotheses of Theorem 2.12. As a reminder, we have natural numbers m, d satisfying $d \geqslant m+2$, and positive reals ε, c, C. For a natural number N, we have $L = L(N) : \mathbb{R}^d \to \mathbb{R}^m$ being a surjective linear map with approximation function A_L, with $\operatorname{dist}(L, V_{\text{rank}}^{\text{unif}}(m,d)) \geqslant c$, $\operatorname{dist}(L, V_{\text{degen}}^*(m,d)) \geqslant c$, and with rational complexity at most C. We have $F : \mathbb{R}^d \to [0,1]$ being the indicator function of $[1,N]^d$ and $G : \mathbb{R}^m \to [0,1]$ being the indicator function of a convex domain contained in $[-\varepsilon, \varepsilon]^m$. For some $s \leqslant d-2$, to be determined, we also have functions $f_1, \ldots, f_d : [N] \to [-1,1]$ that satisfy $\min_j \|f_j\|_{U^{s+1}[N]} \leqslant \rho$ for some ρ in the range $0 < \rho \leqslant 1$.

The proof has four parts:

- Lemma 5.9, in which we replace the indicator function of $[1, N]^d$ with a Lipschitz cutoff.
- Lemma 5.10, in which we replace L by a pair of maps (Ξ, L') where L' is purely irrational.
- Lemma 5.11, in which we replace the function G by a Lipschitz cutoff (using Lemma 3.4).
- Finally, the application of Theorem 5.6 to the pair (Ξ, L').

The second of these steps is by far the most technically intricate, and, as we mentioned when discussing our proof strategy in Section 2E, Lemma 5.10 will have 9 subparts. One might well ask why it is necessary to expend so much effort creating a purely irrational map L', given that Theorem 5.6 does not include this condition in its hypotheses. The point is that in order to replace G with a Lipschitz cutoff (and thus in order to be able to apply Theorem 5.6 at all) it is vital that L' is purely irrational. If $L' : \mathbb{R}^{d-u} \to \mathbb{R}^{m-u}$ failed to be purely irrational then $L'\mathbb{Z}^{d-u}$ would not equidistribute in $\mathbb{R}^{m-u}$; it would instead be restricted to certain proper affine subspaces. This would affect our ability to perturb the function G without drastically altering the number of solutions to the inequality. For more on this issue, the reader may consult Section 2E.

One does note from the above discussion, however, that in order to deduce Theorem 2.12 it would be enough to prove Theorem 5.6 under the additional assumption that L is purely irrational. Yet it turns out that the general version of Theorem 5.6 that we have stated is no harder to prove than the restricted version.

We begin with the first of our four parts.

Lemma 5.9 (replacing variable cutoff). *Assume the hypotheses of Theorem 2.12 (in particular let F be the indicator function $1_{[1,N]^d}$), and let σ_F be any parameter in the range $0 < \sigma_F < \frac{1}{2}$. 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*

$$|T^L_{F,G,N}(f_1, \dots, f_d)| \ll |T^L_{F_{1,\sigma_F},G,N}(f_1, \dots, f_d)| + O_{c,C}(\sigma_F).$$

Proof. By Lemma B.2, for any parameter σ_F in the range $0 < \sigma_F < \frac{1}{2}$ 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 $\int_{\boldsymbol{x}} F_{2,\sigma_F}(\boldsymbol{x})\, d\boldsymbol{x} = O(\sigma_F N^d)$. Moreover, F_{2,σ_F} is supported on

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

Therefore

$$T^L_{F,G,N}(f_1, \dots, f_d) \ll |T^L_{F_{1,\sigma_F},G,N}(f_1, \dots, f_d)| + |T^L_{F_{2,\sigma_F},G,N}(1, \dots, 1)|.$$

Therefore, since $\operatorname{dist}(L, V^{\text{unif}}_{\text{rank}}(m, d)) \geqslant c$, by Lemma 3.3 we have

$$|T^L_{F_{2,\sigma_F},G,N}(f_1, \dots, f_d)| = O_{c,C}(\sigma_F).$$

This gives the lemma. □

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 2.6.

Lemma 5.10 (generating a purely irrational map). *Let σ_F be a parameter in the range $0 < \sigma_F < \frac{1}{2}$. Assume the hypotheses of Theorem 2.12, 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 $u \leqslant m - 1$. 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 $\tilde{R} \subset \mathbb{Z}^d$, and, for each $\tilde{\boldsymbol{r}} \in \tilde{R}$, functions $F_{\tilde{\boldsymbol{r}}} : \mathbb{R}^{d-u} \to [0, 1]$ and $G_{\tilde{\boldsymbol{r}}} : \mathbb{R}^{m-u} \to [0, 1]$, that together satisfy the following properties*:

(1) *Ξ has integer coefficients, $\|\Xi\|_\infty = O_{c,C}(1)$, and $\Xi(\mathbb{Z}^{d-u}) = \mathbb{Z}^d \cap \operatorname{im} \Xi$.*

(2) *$|\tilde{R}| = O_{c,C}(1)$, and $\|\tilde{\boldsymbol{r}}\|_\infty = O_{c,C}(1)$ for all $\tilde{\boldsymbol{r}} \in \tilde{R}$.*

(3) *$F_{\tilde{\boldsymbol{r}}}$ is supported on $[-O_{c,C}(N), O_{c,C}(N)]^{d-u}$, with Lipschitz constant $O_{c,C}(1/\sigma_F N)$, and $G_{\tilde{\boldsymbol{r}}}$ is the indicator function of a convex domain contained in $[-O_{c,C,\varepsilon}(1), O_{c,C,\varepsilon}(1)]^{m-u}$.*

(4) *$T^L_{F,G,N}(f_1, \dots, f_d) = \sum_{\tilde{\boldsymbol{r}} \in \tilde{R}} T^{L',\Xi,\tilde{\boldsymbol{r}}}_{F_{\tilde{\boldsymbol{r}}},G_{\tilde{\boldsymbol{r}}},N}(f_1, \dots, f_d)$.*

(5) *L' is purely irrational.*

(6) *$\|L'\|_\infty = O_{c,C}(1)$ and $\operatorname{dist}(L', V_{\mathrm{rank}}(m-u, d-u)) = \Omega_{c,C}(1)$.*

(7) *$\operatorname{dist}((\Xi, L'), V^*_{\mathrm{degen},2}(m-u, d, d-u)) = \Omega_{c,C}(1)$.*

(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 5.6 are satisfied. Part (9) is included for completeness, and to assist the calculations in Appendix E.

Before giving the full details of the proof, we sketch the idea. Let $\Theta : \mathbb{R}^m \to \mathbb{R}^u$ 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 $\Xi : \mathbb{R}^{d-u} \to \ker(\Theta L)$. Without too much difficultly, Ξ can be chosen to satisfy $\Xi(\mathbb{Z}^{d-u}) = \mathbb{Z}^d \cap \operatorname{im} \Xi$. Then

$$L\Xi : \mathbb{R}^{d-u} \to \ker \Theta,$$

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 5.10.

Of course this is not quite possible, as we only defined the notion of purely irrational maps between vector spaces of the form $\mathbb{R}^a$. 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 $\Xi : \mathbb{R}^d \to \mathbb{R}^d$ to be the identity map, $\tilde{\boldsymbol{r}}$ 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 5.1, in which we may assume that the first m columns of L form the identity matrix.

Indeed, let $\Theta : \mathbb{R}^m \to \mathbb{R}^u$ be a rational map for L with $\|\Theta\|_\infty \leqslant C$. 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

$$T^{L}_{F,G,N}(f_1, \dots, f_d) = T^{\tilde{L}}_{F,\tilde{G},N}(f_1, \dots, f_d),$$

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 $\operatorname{dist}(\tilde{L}, V^*_{\text{degen}}(m,d)) = \Omega_{c,C}(1)$. Finally, for all $\tau_1, \tau_2 \in (0,1]$, we have that $A_{\tilde{L}}(\tau_1,\tau_2) \asymp_{c,C} A_L(\Omega_{c,C}(\tau_1), \Omega_{c,C}(\tau_2))$.

Therefore, by replacing L with $\tilde{L}$ and G with $\tilde{G}$, we may assume throughout the proof of Lemma 5.10 that the first m columns of L form the identity matrix. This is at the cost of replacing ε by $O_{c,C}(\varepsilon)$, C by $O_{c,C}(1)$, and c by $\Omega_{c,C}(1)$.

Now let $\Theta : \mathbb{R}^m \to \mathbb{R}^u$ be a rational map for L with $\|\Theta\|_\infty = O_{c,C}(1)$. Since the first m columns of L form the identity matrix, Θ 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 ΘL 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 $\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(d-u)} \in \mathbb{Z}^d$ that satisfies $\|\boldsymbol{v}^{(i)}\|_\infty = O_{c,C}(1)$ for every i. Define $\Xi : \mathbb{R}^{d-u} \to \mathbb{R}^d$ by

$$\Xi(\boldsymbol{w}) := \sum_{i=1}^{d-u} w_i \boldsymbol{v}^{(i)}.$$

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

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

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

Before proceeding to prove part (3) of the lemma, we pause to apply Lemmas 5.7 and 5.8. Indeed, applying these lemmas to the map $S := \Theta L$, there exists a set $\{\boldsymbol{a_1}, \dots, \boldsymbol{a_u}\} \subset \mathbb{Z}^u$ that is a basis for the lattice $\Theta L(\mathbb{Z}^d)$ and for which $\|\boldsymbol{a_i}\|_\infty = O_{c,C}(1)$ for each i. Also, there exists a set of vectors $\{\boldsymbol{x_1}, \dots, \boldsymbol{x_u}\} \subset \mathbb{Z}^d$ such that $\Theta L(\boldsymbol{x_i}) = \boldsymbol{a_i}$ for each i, and $\|\boldsymbol{x_i}\|_\infty = O_{c,C}(1)$. By Lemma 5.8,

$$\mathcal{B} := \{\boldsymbol{x_i} : i \leqslant u\} \cup \{\Xi(\boldsymbol{w_j}) : j \leqslant d-u\} \tag{5-4}$$

is a basis for $\mathbb{R}^d$ and a lattice basis for $\mathbb{Z}^d$, where $\boldsymbol{w_1}, \dots, \boldsymbol{w_{d-u}}$ 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

$$T^L_{F,G,N}(f_1,\ldots,f_d) = \sum_{\tilde{\boldsymbol{r}}\in\tilde{R}} \frac{1}{N^{d-m}} \sum_{\boldsymbol{n}\in\mathbb{Z}^{d-u}} \Bigl(\prod_{j=1}^{d} f_j(\xi_j(\boldsymbol{n})+\tilde{\boldsymbol{r}}_j)\Bigr) F(\Xi(\boldsymbol{n})+\tilde{\boldsymbol{r}}) G(L\Xi(\boldsymbol{n})+L\tilde{\boldsymbol{r}}), \tag{5-5}$$

where $\tilde{\boldsymbol{r}}_j$ denotes the j-th coordinates of $\tilde{\boldsymbol{r}}$. Now by an easy linear algebraic argument (recorded in Lemma D.4),

$$\mathbb{R}^m = \operatorname{span}(L\boldsymbol{x_i} : i \leqslant u) \oplus \ker\Theta \tag{5-6}$$

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

$$P((\operatorname{span}(L\boldsymbol{x_i} : i \leqslant u))) = \mathbb{R}^u \times \{0\}^{m-u}, \tag{5-7}$$

$$P(\ker\Theta) = \{0\}^u \times \mathbb{R}^{m-u}, \tag{5-8}$$

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

We have

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

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

$$G_{\tilde{\boldsymbol{r}}}(\boldsymbol{x}) := (G\circ P^{-1})(\boldsymbol{x_0} + PL\tilde{\boldsymbol{r}}),$$

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

Define

$$F_{\tilde{\boldsymbol{r}}}(\boldsymbol{n}) := F(\Xi(\boldsymbol{n}) + \tilde{\boldsymbol{r}}).$$

Then $F_{\tilde{\boldsymbol{r}}}$ has Lipschitz constant $O_{c,C}(1/\sigma_F N)$ and $F_{\tilde{\boldsymbol{r}}}$ 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 Ξ). So $F_{\tilde{\boldsymbol{r}}}$ and $G_{\tilde{\boldsymbol{r}}}$ satisfy part (3).

Part (4): Writing $\pi_{m-u} : \mathbb{R}^m \to \mathbb{R}^{m-u}$ for the projection onto the final $m-u$ coordinates, expression (5-5) is equal to

$$\sum_{\tilde{\boldsymbol{r}}\in\tilde{R}} \frac{1}{N^{d-m}} \sum_{\boldsymbol{n}\in\mathbb{Z}^{d-u}} \Bigl(\prod_{j=1}^{d} f_j(\xi_j(\boldsymbol{n})+\tilde{\boldsymbol{r}}_j)\Bigr) F_{\tilde{\boldsymbol{r}}}(\boldsymbol{n}) G_{\tilde{\boldsymbol{r}}}(\pi_{m-u}PL\Xi(\boldsymbol{n})). \tag{5-9}$$

Let

$$L' := \pi_{m-u}PL\Xi. \tag{5-10}$$

Then $L' : \mathbb{R}^{d-u} \to \mathbb{R}^{m-u}$ is surjective, and

$$T^L_{F,G,N}(f_1,\ldots,f_d) = \sum_{\tilde{\boldsymbol{r}}\in\tilde{R}} T^{L',\Xi,\tilde{\boldsymbol{r}}}_{F_{\tilde{\boldsymbol{r}}},G_{\tilde{\boldsymbol{r}}},N}(f_1,\ldots,f_d).$$

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 $\varphi : \mathbb{R}^{m-u} \to \mathbb{R}$ with $\varphi L'(\mathbb{Z}^{d-u}) \subseteq \mathbb{Z}$, i.e., with $\varphi \pi_{m-u} P L \Xi(\mathbb{Z}^{d-u}) \subseteq \mathbb{Z}$. Then define the map $\Theta' : \mathbb{R}^m \to \mathbb{R}^{u+1}$ by

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

Then Θ' 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 $\mathcal{B}$ from (5-4)). 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 Ξ, L, P, and π_{m-u} separately.

We wish to prove that

$$\operatorname{dist}(L', V_{\mathrm{rank}}(m-u, d-u)) \gg_{c,C} 1,$$

i.e., that

$$\operatorname{dist}(\pi_{m-u} P L \Xi, V_{\mathrm{rank}}(m-u, d-u)) \gg_{c,C} 1.$$

Suppose for contradiction that, for a small parameter η, there exists a linear map $Q : \mathbb{R}^{d-u} \to \mathbb{R}^{m-u}$ such that $\|Q\|_\infty < \eta$ and $\pi_{m-u} P L \Xi + Q$ has rank less than $m-u$. Recall that $P L \Xi(\mathbb{R}^{d-u}) = \{0\}^u \times \mathbb{R}^{m-u}$. So, extending Q by zeros to a map $Q : \mathbb{R}^{d-u} \to \{0\}^u \times \mathbb{R}^{m-u}$, and applying P^{-1}, there is a map $Q' : \mathbb{R}^{d-u} \to \mathbb{R}^m$ 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

$$\mathcal{B} := \{\boldsymbol{x}_i : i \leqslant u\} \cup \{\Xi(\boldsymbol{w}_j) : j \leqslant d-u\}$$

be the basis of $\mathbb{R}^d$ from (5-4), i.e., the basis formed by applying Lemma 5.8 to the map $S := \Theta L$. Define the linear map H by $H(\Xi(\boldsymbol{w}_j)) := Q'(\boldsymbol{w}_j)$ for each j and $H(\boldsymbol{x}_i) := \mathbf{0}$ for each i. Since the change of basis matrix between $\mathcal{B}$ 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 $\Xi : \mathbb{R}^{d-u} \to \mathbb{R}^d$ is injective, so this implies that the rank of $L+H$ is less than m. Hence $\operatorname{dist}(L, V_{\mathrm{rank}}(m,d)) = O_{c,C}(\eta)$, which contradicts the assumptions of the lemma (if η is small enough). So $\operatorname{dist}(L', V_{\mathrm{rank}}(m-u, d-u)) \gg_{c,C} 1$ as required.

Part (7): We wish to show that $\operatorname{dist}((\Xi, L'), V^*_{\mathrm{degen},2}(m-u, d, d-u)) = \Omega_{c,C}(1)$. Suppose for contradiction that, for a small parameter η, there exists a linear map $Q : \mathbb{R}^{d-u} \to \mathbb{R}^d$ such that $\|Q\|_\infty \leqslant \eta$ and $\operatorname{dist}((\Xi + Q, L'), V^*_{\mathrm{degen},2}(m-u, d, d-u)) \leqslant \eta$. In other words, we suppose there exist two indices $i, j \leqslant d$, and a real number λ, such that $\boldsymbol{e}_i^* - \lambda \boldsymbol{e}_j^*$ is nonzero and

$$(\Xi + Q)^*(\boldsymbol{e}_i^* - \lambda \boldsymbol{e}_j^*) \in (L')^*((\mathbb{R}^{m-u})^*),$$

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

$$\Xi^*(\boldsymbol{e}_i^* - \lambda \boldsymbol{e}_j^* - L^*(P^*\pi_{m-u}^*(\varphi))) = -Q^*(\boldsymbol{e}_i^* - \lambda \boldsymbol{e}_j^*).$$

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

$$\|\Xi^*(\boldsymbol{e}_i^* - \lambda \boldsymbol{e}_j^* - L^*(P^*\pi_{m-u}^*(\varphi)))\|_\infty = O(\eta \|\boldsymbol{e}_i^* - \lambda \boldsymbol{e}_j^*\|_\infty). \tag{5-11}$$

Let

$$\mathcal{B}^* := \{\boldsymbol{x}_i^* : i \leqslant u\} \cup \{\Xi(\boldsymbol{w}_j)^* : j \leqslant d-u\} \tag{5-12}$$

denote the basis of $(\mathbb{R}^d)^*$ that is dual to the basis $\mathcal{B}$ from (5-4). It follows from part (4) of Lemma 5.8 and (5-11) that

$$\boldsymbol{e}_i^* - \lambda \boldsymbol{e}_j^* - L^*(P^*\pi_{m-u}^*(\varphi)) = \omega_V + \omega_W,$$

where $\omega_V \in L^*\Theta^*((\mathbb{R}^u)^*)$, $\omega_W \in \operatorname{span}(\Xi(\boldsymbol{w}_j)^* : j \leqslant d-u)$, and $\|\omega_W\|_\infty = O_{c,C}(\eta \|\boldsymbol{e}_i^* - \lambda \boldsymbol{e}_j^*\|_\infty)$. So therefore

$$\boldsymbol{e}_i^* - \lambda \boldsymbol{e}_j^* = L^*(\alpha) + \omega_W,$$

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

Therefore $\|\boldsymbol{e}_i^* - \lambda \boldsymbol{e}_j^* - \omega_W\|_\infty \geqslant \frac{1}{2}\|\boldsymbol{e}_i^* - \lambda \boldsymbol{e}_j^*\|_\infty$, provided η is small enough. Since $\|L^*\|_\infty = O_{c,C}(1)$, we conclude that $\|\alpha\|_\infty = \Omega_{c,C}(\|\boldsymbol{e}_i^* - \lambda \boldsymbol{e}_j^*\|_\infty)$.

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

$$\boldsymbol{e}_i^* - \lambda \boldsymbol{e}_j^* \in (L+E)^*((\mathbb{R}^m)^*),$$

and hence $\operatorname{dist}(L, V_{\text{degen}}^*(m,d)) = O_{c,C}(\eta)$. This is a contradiction to the hypotheses of Theorem 2.12, provided η is small enough, and hence $\operatorname{dist}((\Xi, L'), V_{\text{degen},2}^*(m-u, d, d-u)) = \Omega_{c,C}(1)$.

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

$$A_{L'}(\tau_1, \tau_2) \gg_{c,C} A_L(\Omega_{c,C}(\tau_1), \Omega_{c,C}(\tau_2)), \tag{5-13}$$

where L' is as in (5-10).

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 η, there exists some $\varphi \in (\mathbb{R}^{m-u})^*$ for which $\tau_1 \leqslant \|\varphi\|_\infty \leqslant \tau_2^{-1}$ and for which

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

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

$$\operatorname{dist}(L^*(P^*\pi^*_{m-u}(\varphi)), (\mathbb{Z}^d)^T) \ll_{c,C} \eta; \tag{5-14}$$

$$\|P^*\pi^*_{m-u}(\varphi)\|_\infty \ll_{c,C} \tau_2^{-1}; \tag{5-15}$$

$$\operatorname{dist}(P^*\pi^*_{m-u}(\varphi), \Theta^*((\mathbb{R}^u)^*)) \gg_{c,C} \tau_1, \tag{5-16}$$

from which (5-13) immediately follows.

Let us prove (5-14). Indeed, we already know that $\operatorname{dist}(\Xi^* L^* P^* \pi^*_{m-u}(\varphi), (\mathbb{Z}^{d-u})^T) < \eta$, i.e., that

$$\|\Xi^* L^* P^* \pi^*_{m-u}(\varphi) - \alpha\|_\infty < \eta, \tag{5-17}$$

for some $\alpha \in (\mathbb{Z}^{d-u})^T$. Let us write $\alpha = \sum_{j=1}^{d-u} \lambda_j \boldsymbol{w}_j^*$ for some $\lambda_j \in \mathbb{Z}$, where $\boldsymbol{w}_1, \dots, \boldsymbol{w}_{d-u}$ denotes the standard basis for $\mathbb{R}^{d-u}$ and $\boldsymbol{w}_1^*, \dots, \boldsymbol{w}_{d-u}^*$ denotes the dual basis. Let $\mathcal{B}^*$ be as in (5-12). Then $\boldsymbol{w}_j^* = \Xi^*((\Xi(\boldsymbol{w}_j)^*)$, and so

$$\alpha = \Xi^*\biggl(\sum_{j=1}^{d-u} \lambda_j \Xi(\boldsymbol{w}_j)^*\biggr).$$

So from (5-17) and the final part of Lemma 5.8,

$$L^* P^* \pi^*_{m-u}(\varphi) - \sum_{j=1}^{d-u} \lambda_j \Xi(\boldsymbol{w}_j)^* = \omega_V + \omega_W, \tag{5-18}$$

where $\omega_V \in \operatorname{span}(\boldsymbol{x}_i^* : i \leqslant u)$, $\omega_W \in \operatorname{span}(\Xi(\boldsymbol{w}_j)^* : j \leqslant d-u)$, and $\|\omega_W\|_\infty = O_{c,C}(\eta)$.

But $L^* P^* \pi^*_{m-u}(\varphi) \in \operatorname{span}(\Xi(\boldsymbol{w}_j)^* : j \leqslant d-u)$ too. Indeed, for every i at most $d-u$,

$$L^* P^* \pi^*_{m-u}(\varphi)(\boldsymbol{x}_i) = \varphi(\pi_{m-u} P L \boldsymbol{x}_i) = \varphi(\boldsymbol{0}) = 0,$$

by the properties of P (see (5-7)). Therefore $\omega_V = \boldsymbol{0}$, and so

$$\biggl\| L^* P^* \pi^*_{m-u}(\varphi) - \sum_{j=1}^{d-u} \lambda_j \Xi(\boldsymbol{w}_j)^* \biggr\|_\infty = O_{c,C}(\eta).$$

Since $\sum_{j=1}^{d-u} \lambda_j \Xi(\boldsymbol{w}_j)^* \in (\mathbb{Z}^d)^T$, this implies (5-14) as claimed.

The bound (5-15) is immediate from the bounds on the coefficients of P^* and π^*_{m-u}, so it remains to prove (5-16). Suppose for contradiction that, for some small parameter δ,

$$P^* \pi^*_{m-u}(\varphi) = \alpha_1 + \alpha_2,$$

where $\alpha_1 \in \Theta^*((\mathbb{R}^u)^*)$ and $\|\alpha_2\|_\infty \leqslant \delta\tau_1$. We know that $\|\varphi\|_\infty \geqslant \tau_1$, which means that there is some standard basis vector $\boldsymbol{f}_k \in \mathbb{R}^{m-u}$ for which $|\varphi(\boldsymbol{f}_k)| \geqslant \tau_1$. Let $\boldsymbol{b}_{k+u}$ be the standard basis vector of $\mathbb{R}^m$ for which $\pi_{m-u}(\boldsymbol{b}_{k+u}) = \boldsymbol{f}_k$. Recall the properties of P (given in (5-7) and (5-8)), in particular recall that $P : \ker\Theta \to \{0\}^u \times \mathbb{R}^{m-u}$ is an isomorphism. Then

$$|P^* \pi^*_{m-u}(\varphi)(P^{-1}(\boldsymbol{b}_{k+u}))| = |\pi^*_{m-u}(\varphi)(\boldsymbol{b}_{k+u})| = |\varphi(\boldsymbol{f}_k)| \geqslant \tau_1.$$

Note that $\Theta^*((\mathbb{R}^u)^*) = (\ker\Theta)^0$, and so

$$|P^*\pi^*_{m-u}(\varphi)(P^{-1}(\boldsymbol{b_{k+u}}))| = |(\alpha_1+\alpha_2)(P^{-1}(\boldsymbol{b_{k+u}}))| = |\alpha_2(P^{-1}(\boldsymbol{b_{k+u}}))| \ll_{c,C} \delta\tau_1,$$

as $P^{-1}(\boldsymbol{b_{k+u}}) \in \ker\Theta$ and satisfies $\|P^{-1}(\boldsymbol{b_{k+u}})\|_\infty = O_{c,C}(1)$. This is a contradiction if δ is small enough, and so (5-16) holds. This resolves part (8).

Part (9): Let $\tau_1, \tau_2 \in (0, 1]$. We desire to prove the relationship

$$A_{L'}(\tau_1, \tau_2) \ll_{c,C} A_L(\Omega_{c,C}(\tau_1), \Omega_{c,C}(\tau_2)), \tag{5-19}$$

where L' is as in (5-10). 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 η. Then there exists some $\varphi \in (\mathbb{R}^m)^*$ such that $\operatorname{dist}(\varphi, \Theta^*((\mathbb{R}^u)^*)) \geqslant \tau_1$, $\|\varphi\|_\infty \leqslant \tau_2^{-1}$, and $\operatorname{dist}(L^*\varphi, (\mathbb{Z}^d)^T) < \eta$. So there exists some $\omega \in (\mathbb{Z}^d)^T$ for which

$$\|L^*\varphi - \omega\|_\infty < \eta.$$

We expand both $L^*\varphi$ and ω with respect to the dual basis $\mathcal{B}^*$ from (5-12). So,

$$L^*\varphi = \sum_{i=1}^{u} \lambda_i \boldsymbol{x_i^*} + \sum_{j=1}^{d-u} \mu_j \Xi(\boldsymbol{w_j})^* \quad \text{and} \quad \omega = \sum_{i=1}^{u} \lambda_i' \boldsymbol{x_i^*} + \sum_{j=1}^{d-u} \mu_j' \Xi(\boldsymbol{w_j})^*.$$

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 and 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 5.8), one has $|\lambda_i - \lambda_i'| = O_{c,C}(\eta)$ and $|\mu_j - \mu_j'| = O_{c,C}(\eta)$ for each i and j.

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

$$\omega' := \sum_{j=1}^{d-u} \mu_j' \boldsymbol{w_j^*}.$$

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 (5-19) and part (9).

Indeed, recall the decomposition $\mathbb{R}^m = (\operatorname{span}(L\boldsymbol{x_i} : i \leqslant u)) \oplus \ker\Theta$ as an algebraic direct sum from (5-6). Let $\varphi = \varphi_1 + \varphi_2$, where $\varphi_1 \in (\operatorname{span}(L\boldsymbol{x_i} : i \leqslant u))^0$ and $\varphi_2 \in (\ker\Theta)^0$. Since $\operatorname{dist}(\varphi, (\ker\Theta)^0) \geqslant \tau_1$, we have $\|\varphi_1\|_\infty \geqslant \tau_1$. By the properties of the matrix P ((5-7) and (5-8)) there exists some $\varphi' \in (\mathbb{R}^{m-u})^*$ such that

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

Furthermore, by evaluating φ' 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 φ'.

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

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

Hence

$$\Xi^* L^* P^* \pi_{m-u}^* \varphi' = \sum_{j=1}^{d-u} \mu_j \boldsymbol{w}_{\boldsymbol{j}}^*,$$

in other words $(L')^*\varphi' = \sum_{j=1}^{d-u} \mu_j \boldsymbol{w}_{\boldsymbol{j}}^*$. But since $|\mu_j - \mu_j'| = O_{c,C}(\eta)$ for each j, one has $\|(L')^*\varphi' - \omega'\|_\infty = O_{c,C}(\eta)$ as required. This settles part (9).

The entire lemma is settled. □

The final lemma we need in order to deduce Theorem 2.12 involves removing the sharp cutoff G.

Lemma 5.11 (removing image cutoff). *Let m, d, h be natural numbers, satisfying $d \geqslant h \geqslant m+1$. Let c, C, ε be positive, and let σ_G be any parameter in the range $0 < \sigma_G < \frac{1}{2}$. Let $L' : \mathbb{R}^h \to \mathbb{R}^m$ be a purely irrational surjective map, and let $\Xi : \mathbb{R}^h \to \mathbb{R}^d$ be an injective map. Suppose that $\|L'\|_\infty \leqslant C$ and that $\operatorname{dist}(L', V_{\operatorname{rank}}(m, h)) \geqslant c$. Let $F_{\tilde{\boldsymbol{r}}} : \mathbb{R}^h \to [0, 1]$ be any function supported on $[-N, N]^h$, and let $G_{\tilde{\boldsymbol{r}}} : \mathbb{R}^m \to [0, 1]$ be the indicator function of a convex set contained within $[-\varepsilon, \varepsilon]^m$. Then there exists a Lipschitz function $G_{\tilde{\boldsymbol{r}},\sigma_G,1}$ 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 τ_2 in the range $0 < \tau_2 \leqslant 1$ and for any functions $f_1, \ldots, f_d : [N] \to [-1, 1]$,*

$$|T^{L',\Xi,\tilde{\boldsymbol{r}}}_{F_{\tilde{\boldsymbol{r}}},G_{\tilde{\boldsymbol{r}}},N}(f_1, \ldots, f_d)| \ll_{c,C,\varepsilon} |T^{L',\Xi,\tilde{\boldsymbol{r}}}_{F_{\tilde{\boldsymbol{r}}},G_{\tilde{\boldsymbol{r}},\sigma_G,1},N}(f_1, \ldots, f_d)| + \sigma_G + \frac{\tau_2^{1/2}}{\sigma_G} + \frac{\tau_2^{-O(1)} A_L(\Omega_{c,C}(1), \tau_2)^{-1}}{N}.$$

Proof. Applying Lemma B.2 to the function $G_{\tilde{\boldsymbol{r}}}$, we have

$$G_{\tilde{\boldsymbol{r}}} = G_{\tilde{\boldsymbol{r}},\sigma_G,1} + O(G_{\tilde{\boldsymbol{r}},\sigma_G,2}),$$

where $G_{\tilde{\boldsymbol{r}},\sigma_G,1}, G_{\tilde{\boldsymbol{r}},\sigma_G,2} : \mathbb{R}^m \to [0, 1]$ are Lipschitz functions with Lipschitz constant $O_{c,C,\varepsilon}(1/\sigma_G)$, both supported on $[-O_{c,C,\varepsilon}(1), O_{c,C,\varepsilon}(1)]^m$, and with $\int_{\boldsymbol{x}} G_{\tilde{\boldsymbol{r}},\sigma_G,2}(\boldsymbol{x})\, d\boldsymbol{x} = O_{c,C,\varepsilon}(\sigma_G)$.

By the triangle inequality,

$$|T^{L',\Xi,\tilde{\boldsymbol{r}}}_{F_{\tilde{\boldsymbol{r}}},G_{\tilde{\boldsymbol{r}},\sigma_G,2},N}(1, \ldots, 1)| \leqslant T^{L'}_{F_{\tilde{\boldsymbol{r}}},G_{\tilde{\boldsymbol{r}},\sigma_G,2},N}(1, \ldots, 1).$$

We now apply Lemma 3.4, with linear map L' and Lipschitz function $G_{\tilde{\boldsymbol{r}},\sigma_G,2}$. Inserting the bound from Lemma 3.4, the present lemma follows. □

We conclude this section by combining the three previous lemmas, along with Theorem 5.6, to deduce our main result.

Proof of Theorem 2.12 assuming Theorem 5.6. Assume the hypotheses of Theorem 2.12. Let σ_F and σ_G be any parameters satisfying $0 < \sigma_F, \sigma_G < \frac{1}{2}$, and let τ_2 be any parameter satisfying $0 < \tau_2 \leqslant 1$.

By Lemma 5.9,

$$|T^L_{F,G,N}(f_1, \dots, f_d)| \leqslant |T^L_{F_{1,\sigma_F},G,N}(f_1, \dots, f_d)| + O_{c,C}(\sigma_F),$$

for some function $F_{1,\sigma_F} : \mathbb{R}^d \to [0, 1]$ supported on $[-2N, 2N]^d$ and with Lipschitz constant $O(1/\sigma_F N)$. By part (4) of Lemma 5.10, writing F_{1,σ_F} for F, we have

$$|T^L_{F_{1,\sigma_F},G,N}(f_1, \dots, f_d)| \leqslant \sum_{\tilde{\boldsymbol{r}}\in\tilde{R}} |T^{L',\Xi,\tilde{\boldsymbol{r}}}_{F_{\tilde{\boldsymbol{r}}},G_{\tilde{\boldsymbol{r}}},N}(f_1, \dots, f_d)|,$$

where the objects $F_{\tilde{\boldsymbol{r}}}$, $G_{\tilde{\boldsymbol{r}}}$, L', Ξ and $\tilde{R}$ satisfy all the conclusions of that lemma.

Parts (1), (5) and (6) of Lemma 5.10 show that Ξ and L' satisfy the hypotheses of Lemma 5.11, where in the notation of Lemma 5.11 we take $h := d - u$ and rewrite m for $m - u$. So, applying Lemma 5.11, there are some Lipschitz functions $G_{\tilde{\boldsymbol{r}},\sigma_G,1} : \mathbb{R}^{m-u} \to [0, 1]$ supported on $[-O_{c,C,\varepsilon}(1), O_{c,C,\varepsilon}(1)]^{m-u}$ and with Lipschitz constant $O_{c,C,\varepsilon}(1/\sigma_G)$ such that

$$|T^L_{F,G,N}(f_1, \dots, f_d)| \ll_{c,C,\varepsilon} \sum_{\tilde{\boldsymbol{r}}\in\tilde{R}} |T^{L',\Xi,\tilde{\boldsymbol{r}}}_{F_{\tilde{\boldsymbol{r}}},G_{\tilde{\boldsymbol{r}},\sigma_G,1},N}(f_1, \dots, f_d)| + \sigma_G + \frac{\tau_2^{1/2}}{\sigma_G} + \frac{\tau_2^{-O(1)} A_{L'}(\Omega_{c,C}(1), \tau_2)^{-1}}{N} + \sigma_F. \quad (5\text{-}20)$$

(Recall that $|\tilde{R}| = O_{c,C,\varepsilon}(1)$, by part (2) of Lemma 5.10.)

By conclusion (8) of Lemma 5.10, 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 $F_{\tilde{\boldsymbol{r}}}$, L', Ξ, and $\tilde{R}$ together satisfy conclusions (1), (2), (3), (6), and (7) of Lemma 5.10, the hypotheses are satisfied so that we may apply Theorem 5.6 to the expression $T^{L',\Xi,\tilde{\boldsymbol{r}}}_{F_{\tilde{\boldsymbol{r}}},G_{\tilde{\boldsymbol{r}},\sigma_G,1},N}(f_1, \dots, f_d)$. (We take $h = d - u$ and rewrite m for $m - u$, as above). Therefore there exists an s at most $d - 2$, independent of $F_{\tilde{\boldsymbol{r}}}$, $G_{\tilde{\boldsymbol{r}}}$ and $\tilde{\boldsymbol{r}}$, such that, if

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

for some ρ in the range $0 < \rho \leqslant 1$ then $|T^L_{F,G,N}(f_1, \dots, f_d)|$ is

$$\ll_{c,C,\varepsilon} \rho^{\Omega(1)}(\sigma_F^{-O(1)} + \sigma_G^{-O(1)}) + \sigma_F^{-O(1)} N^{-\Omega(1)} + \sigma_G + \frac{\tau_2^{1/2}}{\sigma_G} + \frac{\tau_2^{-O(1)} A_L(\Omega_{c,C}(1), \Omega_{c,C}(\tau_2))^{-1}}{N} + \sigma_F. \quad (5\text{-}21)$$

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

$$|T^L_{F,G,N}(f_1, \dots, f_d)| \ll_{c,C,\varepsilon} \rho^{\Omega(1)} + o_{\rho,A_L,c,C}(1)$$

as $N \to \infty$, where, after the combining the various error terms from (5-21), 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 τ_2 decreases. This is the desired conclusion of Theorem 2.12. □

6. Transfer from $\mathbb{Z}$ to $\mathbb{R}$

Our remaining task is to prove Theorem 5.6. We devote this section to the formulation and proof of a certain "transfer" argument, whereby we replace the discrete summation in the definition of $T^{L,\Xi,\tilde{\boldsymbol{r}}}_{F,G,N}(f_1,\dots,f_d)$ with an integral $\tilde{T}^{L,\Xi,\tilde{\boldsymbol{r}}}_{F,G,N}(g_1,\dots,g_d)$. This manoeuvre will be extremely useful in the sequel, as it gives us access to the standard techniques of manipulating real integrals (in particular reparametrisation of variables). These reparametrisations may be attempted directly in the context of the discrete summation $T^{L,\Xi,\tilde{\boldsymbol{r}}}_{F,G,N}(f_1,\dots,f_d)$, but the results will be messy, and one will need to control the error term each time such a reparametrisation is undertaken. It is easier in our view to do a single approximation at the beginning, so that we may subsequently reparametrise at will. As we remarked in Section 2E, there is a somewhat analogous device in [Green and Tao 2010a], in which the authors transfer their combinatorial expressions into summations over a field (a finite field $\mathbb{Z}/N'\mathbb{Z}$ for some prime N', in their case), in order that their algebraic manipulations may be simplified. The natural field to use in our setting is $\mathbb{R}$.

Let us introduce some notation for the integral in question.

Definition 6.1. Let N,m,d,h be natural numbers, with $d \geqslant h \geqslant m+2$. Let ε be positive. Let $\Xi = (\xi_1,\dots,\xi_d) : \mathbb{R}^h \to \mathbb{R}^d$ 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 $g_1,\dots,g_d : \mathbb{R} \to [-1,1]$ be arbitrary functions. We define

$$\tilde{T}^{L,\Xi,\tilde{\boldsymbol{r}}}_{F,G,N}(g_1,\dots,g_d) := \frac{1}{N^{h-m}} \int_{\boldsymbol{x}\in\mathbb{R}^h} \Bigl(\prod_{j=1}^{d} g_j(\xi_j(\boldsymbol{x})+\tilde{\boldsymbol{r}}_j)\Bigr) F(\boldsymbol{x})G(L\boldsymbol{x})\,d\boldsymbol{x}. \tag{6-1}$$

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

Definition 6.2 (η-supported). Let $\chi : \mathbb{R} \to [0,1]$ be a measurable function, and let η be a positive parameter. We say that χ is η-*supported* if χ is supported on $[-\eta,\eta]$ and $\chi(x) \equiv 1$ for all $x \in [-\eta/2,\eta/2]$.

Definition 6.3 (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 χ is η-supported, for small enough η, then there is only one possible integer n that makes a nonzero contribution to above summation.

We now state the key lemma.

Lemma 6.4. *Let N, m, d, h be natural numbers, with $d \geqslant h \geqslant m+2$, and let c, C, ε, η be positive constants. Let $\Xi : \mathbb{R}^h \to \mathbb{R}^d$ be an injective linear map with integer coefficients, and assume that $\Xi(\mathbb{Z}^h) = \mathbb{Z}^d \cap \operatorname{im} \Xi$. Let $L : \mathbb{R}^h \to \mathbb{R}^m$ be a surjective linear map. Assume that $\|\Xi\|_\infty \leqslant C$, $\|L\|_\infty \leqslant C$, and $\operatorname{dist}(L, V_{\mathrm{rank}}(m,h)) \geqslant c$. 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 $\tilde{\boldsymbol{r}}$ be a fixed vector in $\mathbb{Z}^d$, satisfying $\|\tilde{\boldsymbol{r}}\|_\infty = O_{c,C,\varepsilon}(1)$. Let $\chi : \mathbb{R} \to [0,1]$ be an η-supported measurable function. Then, if η is small enough (in terms of the dimensions m, d, h, C, and ε) there exists some positive real number $C_{\Xi,\chi}$ such that, if $f_1, \dots, f_d : [N] \to [-1,1]$ are arbitrary functions,*

$$T^{\Xi,L,\tilde{\boldsymbol{r}}}_{F,G,N}(f_1,\dots,f_d) = \frac{1}{C_{\Xi,\chi}\eta^h}\tilde{T}^{\Xi,L,\tilde{\boldsymbol{r}}}_{F,G,N}(f_1*\chi,\dots,f_d*\chi) + O_{C,c,\varepsilon}(\eta/\sigma_G) + O_{C,c,\varepsilon}(\eta/\sigma_F N). \quad (6\text{-}2)$$

Moreover, $C_{\Xi,\chi} \asymp_C 1$.

This lemma is a rigorous formulation of (2-12) from the proof strategy in Section 2E. It is in fact the only part of the proof of Theorem 5.6 in which we use the fact that G is Lipschitz.

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

$$C_{\Xi,\chi} := \frac{1}{\eta^h}\int_{\boldsymbol{x}\in\mathbb{R}^h} \boldsymbol{\chi}(\Xi(\boldsymbol{x}))\,d\boldsymbol{x}.$$

Since χ is η-supported, $C_{\Xi,\chi} \asymp_C 1$.

Then, expanding the definition of the convolution,

$$\frac{1}{C_{\Xi,\chi}\eta^h}\tilde{T}^{L,\Xi,\tilde{\boldsymbol{r}}}_{F,G,N}(f_1*\chi,\dots,f_d*\chi)$$

equals

$$\frac{1}{N^{h-m}}\sum_{\boldsymbol{n}\in\mathbb{Z}^d}\biggl(\prod_{j=1}^d f_j(n_j)\biggr)\frac{1}{C_{\Xi,\chi}\eta^h}\int_{\boldsymbol{y}\in\mathbb{R}^h} F(\boldsymbol{y})G(L\boldsymbol{y})\boldsymbol{\chi}(\Xi(\boldsymbol{y})+\tilde{\boldsymbol{r}}-\boldsymbol{n})\,d\boldsymbol{y}. \quad (6\text{-}3)$$

Note that any vector $\boldsymbol{n} \in \mathbb{Z}^d$ that gives a nonzero contribution to expression (6-3) satisfies

$$\|\boldsymbol{n} - \Xi(\boldsymbol{y}) - \tilde{\boldsymbol{r}}\|_\infty \ll \eta,$$

for some $\boldsymbol{y} \in \mathbb{R}^h$. Therefore, $\boldsymbol{n}$ must be of the form $\Xi(\boldsymbol{n}') + \tilde{\boldsymbol{r}}$ for some unique $\boldsymbol{n}' \in \mathbb{Z}^h$. (This is proved in full in Lemma D.2). Therefore, writing $\Xi = (\xi_1,\dots,\xi_d)$, we may reformulate (6-3) as

$$\frac{1}{N^{h-m}}\sum_{\boldsymbol{n}\in\mathbb{Z}^h}\biggl(\prod_{j=1}^d f_j(\xi_j(\boldsymbol{n})+\tilde{\boldsymbol{r}}_j)\biggr)\frac{1}{C_{\Xi,\chi}\eta^h}\int_{\boldsymbol{y}\in\mathbb{R}^h} F(\boldsymbol{y})G(L\boldsymbol{y})\boldsymbol{\chi}(\Xi(\boldsymbol{y}-\boldsymbol{n}))\,d\boldsymbol{y},$$

which is equal to

$$\frac{1}{N^{h-m}}\sum_{\boldsymbol{n}\in\mathbb{Z}^h}\Bigl(\prod_{j=1}^{d} f_j(\xi_j(\boldsymbol{n})+\tilde{\boldsymbol{r}}_j)\Bigr)\frac{1}{C_{\Xi,\chi}\eta^h}\int_{\boldsymbol{y}\in\mathbb{R}^h}(F(\boldsymbol{n})+O_C(\eta/\sigma_F N))G(L\boldsymbol{y})\boldsymbol{\chi}(\Xi(\boldsymbol{y}-\boldsymbol{n}))\,d\boldsymbol{y}. \quad (6\text{-}4)$$

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

Continuing, expression (6-4) is equal to

$$\frac{1}{N^{h-m}}\sum_{\boldsymbol{n}\in\mathbb{Z}^h}\Bigl(\prod_{j=1}^{d} f_j(\xi_j(\boldsymbol{n})+\tilde{\boldsymbol{r}}_j)\Bigr)F(\boldsymbol{n})H(L\boldsymbol{n})+E \quad (6\text{-}5)$$

where

$$H(\boldsymbol{x})=\frac{1}{C_{\Xi,\chi}\eta^h}\int_{\boldsymbol{y}\in\mathbb{R}^h}\boldsymbol{\chi}(\Xi(\boldsymbol{y}))G(\boldsymbol{x}+L\boldsymbol{y})\ d\boldsymbol{y}$$

and E is a certain error, which may be bounded above by

$$\ll_C \frac{\eta}{\sigma_F N}\frac{1}{N^{h-m}}\sum_{\boldsymbol{n}\in[-O(N),O(N)]^h} H(L\boldsymbol{n}). \quad (6\text{-}6)$$

Let us deal with the first term of (6-5), in which we wish to replace H with G. We therefore consider

$$\Bigl|\frac{1}{N^{h-m}}\sum_{\boldsymbol{n}\in\mathbb{Z}^h}\Bigl(\prod_{j=1}^{d} f_j(\xi_j(\boldsymbol{n})+\tilde{\boldsymbol{r}}_j)\Bigr)F(\boldsymbol{n})(G(L\boldsymbol{n})-H(L\boldsymbol{n}))\Bigr|,$$

which is

$$\leqslant \frac{1}{N^{h-m}}\sum_{\boldsymbol{n}\in\mathbb{Z}^h}F(\boldsymbol{n})|G-H|(L\boldsymbol{n}). \quad (6\text{-}7)$$

Using Lemma D.3 again, the function H is supported on $[-\varepsilon-O_C(\eta),\varepsilon+O_C(\eta)]^m$. Thus, if η is small enough in terms of ε, 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,

$$\begin{aligned}G(\boldsymbol{x})-\frac{1}{C_{\Xi,\chi}\eta^h}\int_{\boldsymbol{y}\in\mathbb{R}^h}G(\boldsymbol{x}+L\boldsymbol{y})\boldsymbol{\chi}(\Xi(\boldsymbol{y}))\,d\boldsymbol{y} &= G(\boldsymbol{x})-\frac{1}{C_{\Xi,\chi}\eta^h}\int_{\boldsymbol{y}\in\mathbb{R}^h}(G(\boldsymbol{x})+O_C(\eta/\sigma_G))\boldsymbol{\chi}(\Xi(\boldsymbol{y}))\,d\boldsymbol{y}\\ &= O_C(\eta/\sigma_G),\end{aligned}$$

by the definition of $C_{\Xi,\chi}$ and using the Lipschitz property of G. So, by the crude bound given in Lemma 3.2, (6-7) may be bounded above by $O_{c,C,\varepsilon}(\eta/\sigma_G)$.

Turning to the error E from (6-5), we've already remarked that it may be bounded above by expression (6-6). Applying Lemma 3.2 again, expression (6-6) may be bounded above by $O_{c,C,\varepsilon}(\eta/\sigma_F N)$.

Lemma 6.4 follows immediately upon substituting the estimates on (6-6) and (6-7) into (6-5). □

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

Lemma 6.5 (relating different Gowers norms). *Let s be a natural number, and assume that η is a positive parameter that is small enough in terms of s. Let $\chi:\mathbb{R}\to[0,1]$ be an η-supported measurable function. Let N be a natural number, and let $f:[N]\to\mathbb{R}$ be an arbitrary function. View $f*\chi$ as a function supported on $[-2N,2N]$. Then we have*

$$\|f*\chi\|_{U^{s+1}(\mathbb{R},2N)} \ll \eta^{(s+2)/2^{s+1}} \|f\|_{U^{s+1}[N]}. \tag{6-8}$$

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

Proof. From expression (A-5), we have

$$\|f*\chi\|_{U^{s+1}(\mathbb{R},2N)}^{2^{s+1}} \ll \frac{1}{N^{s+2}} \int_{(x,\boldsymbol{h})\in\mathbb{R}^{s+2}} \prod_{\boldsymbol{\omega}\in\{0,1\}^{s+1}} (f*\chi)(x+\boldsymbol{h}\cdot\boldsymbol{\omega})\,dx\,d\boldsymbol{h}.$$

Substituting in the definition of $f*\chi$, this is equal to

$$\frac{1}{N^{s+2}} \sum_{(n_{\boldsymbol{\omega}})_{\boldsymbol{\omega}\in\{0,1\}^{s+1}}\in\mathbb{Z}^{\{0,1\}^{s+1}}} \Bigg(\prod_{\boldsymbol{\omega}\in\{0,1\}^{s+1}} f(n_{\boldsymbol{\omega}})\Bigg) \int_{(x,\boldsymbol{h})\in\mathbb{R}^{s+2}} \boldsymbol{\chi}(\Psi(x,\boldsymbol{h})-\boldsymbol{n})\,dx\,d\boldsymbol{h}, \tag{6-9}$$

where $\Psi:\mathbb{R}^{s+2}\to\mathbb{R}^{2^{s+1}}$ has coordinate functions $\psi_{\boldsymbol{\omega}}$, indexed by $\boldsymbol{\omega}\in\{0,1\}^{s+1}$, where $\psi_{\boldsymbol{\omega}}(x,\boldsymbol{h}):=x+\boldsymbol{h}\cdot\boldsymbol{\omega}$. In similar notation to that used in the previous proof, for $\boldsymbol{x}\in\mathbb{R}^{2^{s+1}}$, we let $\boldsymbol{\chi}(\boldsymbol{x}):=\prod_{i=1}^{2^{s+1}}\chi(x_i)$. Note that Ψ is injective, $\Psi(\mathbb{Z}^{s+2})=\mathbb{Z}^{2^{s+1}}\cap\operatorname{im}\Psi$, and $\|\Psi\|_\infty=O_s(1)$.

The contribution to the inner integral of (6-9) from a particular $\boldsymbol{n}$ is zero unless $\|\boldsymbol{n}-\Psi(x,\boldsymbol{h})\|_\infty\ll\eta$, for some $(x,\boldsymbol{h})\in\mathbb{R}^{s+2}$. Therefore, if η is small enough we can conclude that $\boldsymbol{n}$ must be of the form $\Psi(p,\boldsymbol{k})$, for some unique $(p,\boldsymbol{k})\in\mathbb{Z}^{s+2}$. (To spell it out, apply Lemma D.2 with the map Ψ in place of the map Ξ). So (6-9) is equal to

$$\frac{1}{N^{s+2}} \sum_{(p,\boldsymbol{k})\in\mathbb{Z}^{s+2}} \Bigg(\prod_{\boldsymbol{\omega}\in\{0,1\}^{s+1}} f(p+\boldsymbol{k}\cdot\boldsymbol{\omega})\Bigg) \int_{(x,\boldsymbol{h})\in\mathbb{R}^{s+2}} \boldsymbol{\chi}(\Psi(x-p,\boldsymbol{h}-\boldsymbol{k}))\,dx\,d\boldsymbol{h}, \tag{6-10}$$

which, after a change of variables, is equal to

$$\frac{C}{N^{s+2}} \sum_{(p,\boldsymbol{k})\in\mathbb{Z}^{s+2}} \prod_{\boldsymbol{\omega}\in\{0,1\}^{s+1}} f(p+\boldsymbol{k}\cdot\boldsymbol{\omega}), \tag{6-11}$$

where

$$C := \int_{(x,\boldsymbol{h})\in\mathbb{R}^{s+2}} \boldsymbol{\chi}(\Psi(x,\boldsymbol{h}))\,dx\,d\boldsymbol{h}.$$

Since $\boldsymbol{\chi}$ has support contained within $[-\eta,\eta]^{2^{s+1}}$, a vector $(x,\boldsymbol{h})$ only makes a nonzero contribution to the above integral if $\|\Psi(x,\boldsymbol{h})\|_\infty\ll\eta$. This implies that $\|(x,\boldsymbol{h})\|_\infty\ll\eta$. (To prove this is full, apply Lemma D.3 to the linear map Ψ). Since $\|\boldsymbol{\chi}\|_\infty=O(1)$, this means $C=O(\eta^{s+2})$. The lemma then follows from (6-11). □

7. Degeneracy relations

Our aim for this short section is to establish a quantitative relationship between the dual pair degeneracy variety $V^*_{\mathrm{degen},2}(m, d, h)$ and the degeneracy variety $V_{\mathrm{degen}}(h-m, d)$ (see Definitions 5.4 and 4.4 respectively), which will be needed in the next section. It is here that we show that $V^*_{\mathrm{degen},2}(m, d, h)$ was indeed the appropriate notion for guaranteeing finite Cauchy–Schwarz complexity of the relevant system of homogeneous linear forms. We direct the reader to Proposition 4.5 and the discussion after Definition 5.4 for more on this issue.

To introduce the ideas, we first prove a nonquantitative proposition (which is a generalisation of Proposition 4.5).

Lemma 7.1. *Let m, d, h be natural numbers, with $d \geqslant h \geqslant m+2$. 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 $(\Xi, L) \notin V^*_{\mathrm{degen},2}(m, d, h)$. 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 homogeneous linear forms, is not in $V_{\mathrm{degen}}(h-m, d)$.*

Proof. Let $\boldsymbol{e_1}, \dots, \boldsymbol{e_d}$ denote the standard basis vectors in $\mathbb{R}^d$, and let $\boldsymbol{e}^*_{\boldsymbol{1}}, \dots, \boldsymbol{e}^*_{\boldsymbol{d}}$ denote the dual basis of $(\mathbb{R}^d)^*$. Suppose for contradiction that $\Xi\Phi \in V_{\mathrm{degen}}(h-m, d)$. Then by Proposition 4.5 there exist two indices $i, j \leqslant d$, and a real number λ, such that $\boldsymbol{e}^*_{\boldsymbol{i}} - \lambda \boldsymbol{e}^*_{\boldsymbol{j}}$ is nonzero and $\Xi\Phi(\mathbb{R}^{h-m}) \subset \ker(\boldsymbol{e}^*_{\boldsymbol{i}} - \lambda \boldsymbol{e}^*_{\boldsymbol{j}})$.

But then $\Phi(\mathbb{R}^{h-m}) \subset \ker(\Xi^*(\boldsymbol{e}^*_{\boldsymbol{i}} - \lambda \boldsymbol{e}^*_{\boldsymbol{j}}))$, i.e., $\Xi^*(\boldsymbol{e}^*_{\boldsymbol{i}} - \lambda \boldsymbol{e}^*_{\boldsymbol{j}}) \in (\ker L)^0$. But $(\ker L)^0 = L^*((\mathbb{R}^m)^*)$, and so $\Xi^*(\boldsymbol{e}^*_{\boldsymbol{i}} - \lambda \boldsymbol{e}^*_{\boldsymbol{j}}) \in L^*((\mathbb{R}^m)^*)$.

Then, by the definition of $V^*_{\mathrm{degen},2}(m, d, h)$, we have $(\Xi, L) \in V^*_{\mathrm{degen},2}(m, d, h)$, 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 $d \geqslant h \geqslant m+2$, 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 $\|\Xi\|_\infty \leqslant C$, and* $\operatorname{dist}((\Xi, L), V^*_{\mathrm{degen},2}(m, d, h)) \geqslant c$. *Let K denote* $\ker L$, *choose any orthonormal basis $\{\boldsymbol{v}^{(\boldsymbol{1})}, \dots, \boldsymbol{v}^{(\boldsymbol{h-m})}\}$ for K, and let $\Phi : \mathbb{R}^{h-m} \to K$ denote the associated parametrisation, i.e., $\Phi(\boldsymbol{x}) := \sum_{i=1}^{h-m} x_i \boldsymbol{v}^{(\boldsymbol{i})}$. Then $\|\Xi\Phi\|_\infty = O(C)$ and* $\operatorname{dist}(\Xi\Phi, V_{\mathrm{degen}}(h-m, d)) = \Omega(c)$.

For the definition of $\operatorname{dist}((\Xi, L), V^*_{\mathrm{degen},2}(m, d, h))$, consult Definition 5.5.

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

Let $\boldsymbol{e_1}, \dots, \boldsymbol{e_d}$ denote the standard basis vectors in $\mathbb{R}^d$, and let $\boldsymbol{e}^*_{\boldsymbol{1}}, \dots, \boldsymbol{e}^*_{\boldsymbol{d}}$ denote the dual basis of $(\mathbb{R}^d)^*$. Suppose for contradiction that $\operatorname{dist}(\Xi\Phi, V_{\mathrm{degen}}(h-m, d)) \leqslant \eta$ for some small parameter η. In other words, assume that there exists a linear map $P : \mathbb{R}^{h-m} \to \mathbb{R}^d$ with $\|P\|_\infty \leqslant \eta$ such that $\Xi\Phi + P \in V_{\mathrm{degen}}(h-m, d)$. By definition, this means that

$$(\Xi\Phi + P)(\mathbb{R}^{h-m}) \subset \ker(\boldsymbol{e}^*_{\boldsymbol{i}} - \lambda \boldsymbol{e}^*_{\boldsymbol{j}}),$$

for some two indices $i, j \leqslant d$, and some real number λ, such that $\boldsymbol{e}^*_{\boldsymbol{i}} - \lambda \boldsymbol{e}^*_{\boldsymbol{j}}$ is nonzero.

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 $\boldsymbol{f_1},\dots,\boldsymbol{f_{h-m}}$ denote the standard basis vectors in $\mathbb{R}^{h-m}$, and for all k at most $h-m$ define

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

(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 $\boldsymbol{v}^{(\boldsymbol{k})}$ seems to be natural). Complete $\{\boldsymbol{v}^{(\mathbf{1})},\dots,\boldsymbol{v}^{(\boldsymbol{h-m})}\}$ to an orthonormal basis $\{\boldsymbol{v}^{(\mathbf{1})},\dots,\boldsymbol{v}^{(\boldsymbol{h})}\}$ for $\mathbb{R}^h$ and, for k in the range $h-m+1\leqslant k\leqslant h-m$, define $Q(\boldsymbol{v}^{(\boldsymbol{k})}) := \mathbf{0}$. Then $P=Q\Phi$, and $\|Q\|_\infty = O(\eta)$, since $\{\boldsymbol{v}^{(\mathbf{1})},\dots,\boldsymbol{v}^{(\boldsymbol{h})}\}$ is an orthonormal basis.

Thus,

$$(\Xi\Phi + Q\Phi)(\mathbb{R}^{h-m}) \subset \ker(\boldsymbol{e}_{\boldsymbol{i}}^* - \lambda \boldsymbol{e}_{\boldsymbol{j}}^*).$$

So

$$\Phi(\mathbb{R}^{h-m}) \subset \ker((\Xi+Q)^*(\boldsymbol{e}_{\boldsymbol{i}}^* - \lambda \boldsymbol{e}_{\boldsymbol{j}}^*)).$$

Like the previous proof, we conclude that

$$(\Xi+Q)^*(\boldsymbol{e}_{\boldsymbol{i}}^* - \lambda \boldsymbol{e}_{\boldsymbol{j}}^*) \in L^*((\mathbb{R}^m)^*).$$

Hence $((\Xi+Q), L) \in V^*_{\text{degen},2}(m,d,h)$, which, if η is small enough, contradicts the assumption that $\operatorname{dist}((\Xi,L), V^*_{\text{degen},2}(m,d,h)) \geqslant c$. □

8. A generalised von Neumann theorem

In this section we complete the proof of Theorem 5.6, and therefore complete the proof of Theorem 2.12. It will be enough to prove the following statement.

Theorem 8.1. *Let N,m,d,h be natural numbers, with $d\geqslant h\geqslant m+2$, and let c,C,ε 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 $\|L\|_\infty\leqslant C$, $\|\Xi\|_\infty\leqslant C$, $\operatorname{dist}(L, V_{\text{rank}}(m,d))\geqslant c$ and $\operatorname{dist}((\Xi,L), V^*_{\text{degen},2}(m,d,h))\geqslant c$. Then there is some natural number s at most $d-2$, independent of ε, such that the following holds. Let $\tilde{\boldsymbol{r}}\in\mathbb{Z}^d$ be some vector with $\|\tilde{\boldsymbol{r}}\|_\infty = O_{c,C,\varepsilon}(1)$, and let σ_F be a parameter in the range $O<\sigma_F<\frac{1}{2}$. 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 $g_1,\dots,g_d:[-2N,2N]^d\to[-1,1]$ be arbitrary measurable functions. Suppose*

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

for some ρ at most 1. *Then*

$$|\tilde{T}^{L,\Xi,\tilde{\boldsymbol{r}}}_{F,G,N}(g_1,\dots,g_d)| \ll_{c,C,\varepsilon} \rho^{\Omega(1)}\sigma_F^{-1}. \tag{8-1}$$

Proof that Theorem 8.1 implies Theorem 5.6. Assume the hypotheses of Theorem 5.6. This gives natural numbers N,m,d,h, linear maps $L:\mathbb{R}^h\to\mathbb{R}^m$ and $\Xi:\mathbb{R}^h\to\mathbb{R}^d$, 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

$$\delta := T^{L,\Xi,\tilde{r}}_{F,G,N}(f_1, \ldots, f_d).$$

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

Let $\eta := c_1\delta\sigma_G$, where c_1 is small enough depending on m, d, h, c, C, and ε, and let $\chi : \mathbb{R} \to [0, 1]$ be an η-supported measurable function (see Definition 6.2). For all j at most d, let $g_j := f_j * \chi$. Finally, suppose $\min_j \|f_j\|_{U^{s+1}[N]} \leqslant \rho$, for some parameter ρ in the range $0 < \rho \leqslant 1$.

We proceed by bounding $\tilde{T}^{L,\Xi,\tilde{r}}_{F,G,N}(g_1, \ldots, g_d)$. Indeed, by Lemma 6.5, if c_1 is small enough

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

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

$$\tilde{T}^{L,\Xi,\tilde{r}}_{F,G,N}(g_1, \ldots, g_d) \ll_{c,C,\varepsilon} \rho^{\Omega(1)}\sigma_F^{-1}. \tag{8-2}$$

Now we use this to bound δ by Gowers norms. Indeed, by Lemma 6.4, we have

$$\delta \ll_{c,C,\varepsilon} \frac{1}{(c_1\delta\sigma_G)^h}\tilde{T}^{L,\Xi,\tilde{r}}_{F,G,N}(g_1, \ldots, g_d) + c_1\delta + c_1\delta\sigma_G\sigma_F^{-1}N^{-1}.$$

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

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

and so

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

This yields the desired conclusion of Theorem 5.6. □

So it remains to prove Theorem 8.1, for which the bulk of the work will be done in the following two propositions. In Proposition 8.2, we will reduce the integral in $\tilde{T}^{L,\Xi,\tilde{r}}_{F,G,N}(g_1, \ldots, g_d)$ to an integral over the kernel of L. This kernel will be parametrised by a map Ψ, which will have finite c_1-Cauchy–Schwarz complexity for some suitable c_1. In Proposition 8.3 we will then work out the details of applying the Cauchy–Schwarz inequality to such a map, thereby producing Gowers norms.

Proposition 8.2 (separating out the kernel). *Let N, m, d, h be natural numbers, with $d \geqslant h \geqslant m + 2$, and let c, C, ε be positive constants. Let σ_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 $\|L\|_\infty \leqslant C$, $\|\Xi\|_\infty \leqslant C$, $\mathrm{dist}(L, V_{\mathrm{rank}}(m, h)) \geqslant c$ and $\mathrm{dist}((\Xi, L), V^*_{\mathrm{degen},2}(m, d, h)) \geqslant c$. 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 measurable function supported on $[-\varepsilon, \varepsilon]^m$. Let $\tilde{\boldsymbol{r}}$ be a fixed vector in $\mathbb{Z}^d$, satisfying $\|\tilde{\boldsymbol{r}}\|_\infty = O_C(1)$. Then there exists a system of linear forms $(\psi_1, \ldots, \psi_d) = \Psi : \mathbb{R}^{h-m} \to \mathbb{R}^d$ satisfying $\|\Psi\|_\infty = O_C(1)$, 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* $g_1, \dots, g_d : [-2N, 2N] \to [-1, 1]$ *are arbitrary functions,*

$$|\tilde{T}^{L,\Xi,\tilde{r}}_{F,G,N}(g_1, \dots, g_d)| \ll_{c,C,\varepsilon} \left| \frac{1}{N^{h-m}} \int_{\boldsymbol{x}} \prod_{j=1}^{d} g_j(\psi_j(\boldsymbol{x}) + a_j) F_1(\boldsymbol{x})\, d\boldsymbol{x} \right|, \tag{8-3}$$

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* Ψ *has* $\Omega_{c,C}(1)$*-Cauchy–Schwarz complexity at most* s, *in the sense of Definition 4.6.*

Proof of Proposition 8.2. For ease of notation, let

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

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

$$\Phi(\boldsymbol{x}) := \sum_{i=1}^{h-m} x_i \boldsymbol{v}^{(i)}, \tag{8-4}$$

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

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

$$\frac{1}{N^{h-m}} \int_{\boldsymbol{x} \in \mathbb{R}^h} F\Big(\sum_{i=1}^{h} x_i \boldsymbol{v}^{(i)}\Big) G\Big(L\Big(\sum_{i=1}^{h} x_i \boldsymbol{v}^{(i)}\Big)\Big) \Big(\prod_{j=1}^{d} g_j\Big(\xi_j\Big(\Phi(\boldsymbol{x}_{\boldsymbol{1}}^{\boldsymbol{h-m}}) + \sum_{i=h-m+1}^{h} x_i \boldsymbol{v}^{(i)}\Big) + \tilde{\boldsymbol{r}}_j\Big)\Big)\, d\boldsymbol{x}, \tag{8-5}$$

using $\boldsymbol{x}_{\boldsymbol{1}}^{\boldsymbol{h-m}}$ to refer to the vector in $\mathbb{R}^{h-m}$ given by the first the first $h-m$ coordinates of $\boldsymbol{x}$.

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

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

The vector $\sum_{i=h-m+1}^{h} x_i \boldsymbol{v}^{(i)}$ is in $(\ker L)^{\perp}$. Hence, due to the limited support of G, there is a domain D, contained in $[-O_{\varepsilon,c,C}(1), O_{\varepsilon,c,C}(1)]^m$, such that $G\big(L\big(\sum_{i=h-m+1}^{h} x_i \boldsymbol{v}^{(i)}\big)\big)$ is equal to zero unless $(x_{h-m+1}, \dots, x_h)^T \in D$. (This is proved in full in Lemma D.1).

We can use this observation to bound the right-hand side of (8-5). Indeed, we have

$$\beta \ll \operatorname{vol} D \times \sup_{\boldsymbol{x}_{\boldsymbol{h-m+1}}^{\boldsymbol{h}} \in D} \frac{1}{N^{h-m}} \bigg| \int_{\boldsymbol{x}_{\boldsymbol{1}}^{\boldsymbol{h-m}} \in \mathbb{R}^{h-m}} F\Big(\sum_{i=1}^{h} x_i \boldsymbol{v}^{(i)}\Big) G\Big(L\Big(\sum_{i=h-m+1}^{h} x_i \boldsymbol{v}^{(i)}\Big)\Big) \times \Big(\prod_{j=1}^{d} g_j\Big(\xi_j\Big(\Phi(\boldsymbol{x}_{\boldsymbol{1}}^{\boldsymbol{h-m}}) + \sum_{i=h-m+1}^{h} x_i \boldsymbol{v}^{(i)}\Big) + \tilde{\boldsymbol{r}}_j\Big)\Big)\, d\boldsymbol{x}_{\boldsymbol{1}}^{\boldsymbol{h-m}} \bigg|. \tag{8-6}$$

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

$$\beta \ll_{c,C,\varepsilon} \frac{1}{N^{h-m}} \left| \int_{\boldsymbol{x_1^{h-m}} \in \mathbb{R}^{h-m}} F\Big(\sum_{i=1}^{h} x_i \boldsymbol{v}^{(\boldsymbol{i})}\Big) G\Big(L\Big(\sum_{i=h-m+1}^{h} x_i \boldsymbol{v}^{(\boldsymbol{i})}\Big)\Big) \times \Big(\prod_{j=1}^{d} g_j\Big(\xi_j\Big(\Phi(\boldsymbol{x_1^{h-m}}) + \sum_{i=h-m+1}^{h} x_i \boldsymbol{v}^{(\boldsymbol{i})}\Big) + \tilde{\boldsymbol{r}}_j\Big)\Big) \, d\boldsymbol{x_1^{h-m}} \right|. \tag{8-7}$$

Define the function $F_1 : \mathbb{R}^{h-m} \to [0, 1]$ by

$$F_1(\boldsymbol{x_1^{h-m}}) := F\Big(\Phi(\boldsymbol{x_1^{h-m}}) + \sum_{i=h-m+1}^{h} x_i \boldsymbol{v}^{(\boldsymbol{i})}\Big)$$

and for each j at most d, a shift

$$a_j := \xi_j\Big(\sum_{i=h-m+1}^{h} x_i \boldsymbol{v}^{(\boldsymbol{i})}\Big) + \tilde{\boldsymbol{r}}_j.$$

Then

$$\beta \ll_{c,C,\varepsilon} \left| \frac{1}{N^{h-m}} \int_{\boldsymbol{x} \in \mathbb{R}^{h-m}} F_1(\boldsymbol{x}) \prod_{j=1}^{d} g_j(\xi_j(\Phi(\boldsymbol{x})) + a_j) \, d\boldsymbol{x} \right|, \tag{8-8}$$

and F_1 and a_j satisfy the conclusions of the proposition.

Finally, since $\operatorname{dist}((\Xi, L), V^*_{\text{degen},2}(m, d, h)) \geqslant c$ and $\|\Xi\|_\infty, \|L\|_\infty \leqslant C$, Lemma 7.2 tells us that $\Xi\Phi : \mathbb{R}^{h-m} \to \mathbb{R}^d$ satisfies $\operatorname{dist}(\Xi\Phi, V_{\text{degen}}(h-m, d)) \gg_{c,C} 1$. (One may consult Definitions 4.4 and 5.4 for the definitions of $V_{\text{degen}}(h-m, d)$ and $V^*_{\text{degen},2}(m, d, h)$). Thus, by Lemma 4.7, there exists some s at most $d-2$ for which $\Xi\Phi$ has $\Omega_{c,C}(1)$-Cauchy–Schwarz complexity at most s.

Writing Ψ for $\Xi\Phi$, the proposition is proved. □

Proposition 8.3 (Cauchy–Schwarz argument). *Let s, d be natural numbers, with $d \geqslant 3$, and let C be a positive constant. Let σ_F be a parameter in the range $0 < \sigma_F < \frac{1}{2}$. Let $(\psi_1, \dots, \psi_d) = \Psi : \mathbb{R}^{s+1} \to \mathbb{R}^d$ be a linear map, and suppose that $\psi_1(\boldsymbol{e_k}) = 1$, for all the standard basis vectors $\boldsymbol{e_k} \in \mathbb{R}^{s+1}$. Suppose that, for all j in the range $2 \leqslant j \leqslant s+1$, there exists some k such that $\psi_j(\boldsymbol{e_k}) = 0$. Let $N \geqslant 1$ be real, and let $g_1, \dots, g_d : [-N, N] \to [-1, 1]$ be arbitrary measurable functions, and, for each j at most d, let a_j be some real number with $|a_j| \leqslant CN$. Let $F : \mathbb{R}^{s+1} \to [0, 1]$ be any Lipschitz function, supported on $[-CN, CN]^{s+1}$ with Lipschitz constant $O(1/\sigma_F N)$. Suppose that $\|g_1\|_{U^{s+1}(\mathbb{R},N)} \leqslant \rho$, for some parameter ρ in the range $0 < \rho \leqslant 1$. Then*

$$\left| \frac{1}{N^{s+1}} \int_{\boldsymbol{w} \in \mathbb{R}^{s+1}} \prod_{j=1}^{d} g_j(\psi_j(\boldsymbol{w}) + a_j) F(\boldsymbol{w}) \, d\boldsymbol{w} \right| \ll_C \rho^{-\Omega(1)} \sigma_F^{-1}. \tag{8-9}$$

We stress again that implied constants may depend on the implicit dimensions (so the $\Omega(1)$ term in (8-9) may depend on s).

Proof. This theorem is very similar to the usual generalised von Neumann theorem (see [Tao 2012, 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 from line to line. Next, since $\psi_1(\boldsymbol{w}) = w_1 + w_2 + \cdots + w_{s+1}$, 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 $\boldsymbol{w}$ to $[-CN, CN]^{s+1}$. By Lemma B.4, for any $Y > 2$ there is a function $\boldsymbol{c}_Y : \mathbb{R}^{s+1} \to \mathbb{C}$ satisfying $\|\boldsymbol{c}\|_\infty \ll 1$ such that we may replace $F(\boldsymbol{w})$ by

$$\int_{\substack{\boldsymbol{\theta}\in\mathbb{R}^{s+1}\\ \|\boldsymbol{\theta}\|_\infty \leqslant Y}} c_Y(\boldsymbol{\theta}) e\Big(\frac{\boldsymbol{\theta}\cdot\boldsymbol{w}}{N}\Big)\, d\boldsymbol{\theta} + O_C\Big(\frac{\log Y}{\sigma_F Y}\Big).$$

We will determine a particularly suitable Y later (which will depend on ρ).

This means that

$$\Big|\frac{1}{N^{s+1}}\int_{\boldsymbol{w}\in\mathbb{R}^{s+1}} \prod_{j=1}^{d} g_j(\psi_j(\boldsymbol{w}) + a_j) F(\boldsymbol{w})\, d\boldsymbol{w}\Big|$$
$$\ll \int_{\substack{\boldsymbol{\theta}\in\mathbb{R}^{s+1}\\ \|\boldsymbol{\theta}\|_\infty \leqslant Y}} \Big|\frac{1}{N^{s+1}}\int^{*}_{\boldsymbol{w}\in\mathbb{R}^{s+1}} e\Big(\frac{\boldsymbol{\theta}}{N}\cdot\boldsymbol{w}\Big)\Big(\prod_{j=1}^{d} g_j(\psi_j(\boldsymbol{w}) + a_j)\Big)\, d\boldsymbol{w}\Big|\, d\boldsymbol{\theta} + O_C\Big(\frac{\log Y}{\sigma_F Y}\Big), \quad (8\text{-}10)$$

where $\int^*$ indicates the limits $\boldsymbol{w} \in [-CN, CN]^{s+1}$. Fix $\boldsymbol{\theta}$. The inner integral of (8-10) will be our primary focus.

Firstly, we wish to "absorb" the exponential phases $e\big(\frac{\boldsymbol{\theta}}{N}\cdot\boldsymbol{w}\big)$. To do this, we write $e\big(\frac{\boldsymbol{\theta}}{N}\cdot\boldsymbol{w}\big)$ as a product of functions $\prod_{k=1}^{s+1} b_k(\boldsymbol{w})$, where, for each k, the function $b_k : \mathbb{R}^{s+1} \to \mathbb{C}$ is bounded in absolute value by 1 and does not depend on the variable w_k. Since $s + 1 \geqslant 2$, this is possible. Now write

$$\prod_{j=2}^{d} g_j(\psi_j(\boldsymbol{w}) + a_j) = \prod_{k=1}^{s+1} b'_k(\boldsymbol{w}),$$

where each $b'_k : \mathbb{R}^{s+1} \to \mathbb{C}$ is bounded in absolute value by 1 and does not depend on the variable w_k. This is possible since ψ_1 is the only function ψ_j that includes all the variables $w_1, \ldots, w_{s+1}$.

Therefore we may rewrite the inner integral of (8-10) as

$$\frac{1}{N^{s+1}}\int^{*}_{\boldsymbol{w}\in\mathbb{R}^{s+1}} g_1(\psi_1(\boldsymbol{w})) \prod_{k=1}^{s+1} b'_k(\boldsymbol{w}) b_k(\boldsymbol{w})\, d\boldsymbol{w}. \quad (8\text{-}11)$$

A brief aside: readers familiar with the arguments of [Green and Tao 2010a, 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 [loc. cit.]). The key point there is that, if the linear form $\boldsymbol{w} \mapsto \boldsymbol{\theta} \cdot \boldsymbol{w}$ fails to be in the set $\operatorname{span}(\psi_j : 1 \leqslant j \leqslant d)$, 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 is not 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 $\psi_1(\boldsymbol{w}) = w_1 + w_2 + \cdots + w_{s+1}$. 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

$$\ll_C \left(\frac{1}{N^{2s+2}} \int^*_{\boldsymbol{w} \in \mathbb{R}^{s+1}} \int^*_{\boldsymbol{z} \in \mathbb{R}^{s+1}} \prod_{\boldsymbol{\alpha} \in \{0,1\}^{s+1}} g_1\Biggl(\sum_{\substack{k \leqslant s+1 \\ \boldsymbol{\alpha}_k = 0}} w_k + \sum_{\substack{k \leqslant s+1 \\ \boldsymbol{\alpha}_k = 1}} z_k \Biggr) \, d\boldsymbol{w} \, d\boldsymbol{z} \right)^{1/2^{s+1}}. \tag{8-12}$$

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

$$\ll \left(\frac{1}{N^{2s+2}} \int_{(u, \boldsymbol{m}, \boldsymbol{z}_2^{s-1}) \in D} \prod_{\boldsymbol{\alpha} \in \{0,1\}^{s+1}} g_1(u + \boldsymbol{\alpha} \cdot \boldsymbol{m}) \, du \, d\boldsymbol{m} \, d\boldsymbol{z}_{\mathbf{2}}^{\mathbf{s+1}} \right)^{1/2^{s+1}}, \tag{8-13}$$

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

By Lemma B.2 we may write

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

for any σ in the range $0 < \sigma < \frac{1}{2}$, 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

$$\int_{\boldsymbol{x}} G_\sigma(\boldsymbol{x}) \, d\boldsymbol{x} = O_C(\sigma N^{2s+2}).$$

Then, since $\|g_1\|_\infty \leqslant 1$, we may bound (8-13) above by

$$\left(\frac{1}{N^{2s+2}} \int^*_{u, \boldsymbol{m}, \boldsymbol{z}_2^{s-1}} F_\sigma(u, \boldsymbol{m}, \boldsymbol{z}_{\mathbf{2}}^{\mathbf{s-1}}) \prod_{\boldsymbol{\alpha} \in \{0,1\}^{s+1}} g_1(u + \boldsymbol{\alpha} \cdot \boldsymbol{m}) du \, d\boldsymbol{m} \, d\boldsymbol{z}_{\mathbf{2}}^{\mathbf{s-1}} + O_C(\sigma) \right)^{1/2^{s+1}}, \tag{8-14}$$

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

By applying Lemma B.4 to F_σ, for any $X > 2$ the absolute value of expression (8-14) is

$$\ll_C \Biggl(\Biggl(\frac{1}{N^{2s+2}} \int_{\substack{\boldsymbol{\xi} \in \mathbb{R}^{2s+2} \\ \|\boldsymbol{\xi}\|_\infty \leqslant X}} \left| \int^*_{u, \boldsymbol{m}, \boldsymbol{z}_2^{s-1}} e\Bigl(\frac{\boldsymbol{\xi}}{N} \cdot (u, \boldsymbol{m}, \boldsymbol{z}_{\mathbf{2}}^{\mathbf{s-1}}) \Bigr) \prod_{\boldsymbol{\alpha} \in \{0,1\}^{s+1}} g_1(u + \boldsymbol{\alpha} \cdot \boldsymbol{m}) du \, d\boldsymbol{m} \, d\boldsymbol{z}_{\mathbf{2}}^{\mathbf{s-1}} \right| d\boldsymbol{\xi} \Biggr) + O(\sigma) + O\Bigl(\frac{\log X}{\sigma X} \Bigr) \Biggr)^{1/2^{s+1}}. \tag{8-15}$$

Integrating over the variables $z_2, \dots, z_{s+1}$, and splitting the exponential phase amongst the different functions, expression (8-15) is

$$\ll_C \Biggl(\biggl(\frac{1}{N^{s+2}} \int_{\substack{\boldsymbol{\xi}\in\mathbb{R}^{2s+2}\\ \|\boldsymbol{\xi}\|_\infty\leqslant X}} \biggl| \int_{(u,\boldsymbol{m})\in[-CN,CN]^{s+2}} \prod_{\boldsymbol{\alpha}\in\{0,1\}^{s+1}} g_{\boldsymbol{\alpha}}(u+\boldsymbol{\alpha}\cdot\boldsymbol{m})\,du\,d\boldsymbol{m}\biggr|\,d\boldsymbol{\xi}\biggr) + O_C(\sigma) + O_C\biggl(\frac{\log X}{\sigma X}\biggr)\Biggr)^{1/2^{s+1}}, \quad (8\text{-}16)$$

where each function $g_{\boldsymbol{\alpha}}$ is of the form

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

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

Recall that g_1 is supported on $[-2N, 2N]$. Therefore, if $\prod_{\boldsymbol{\alpha}\in\{0,1\}^{s+1}} g_{\boldsymbol{\alpha}}(u+\boldsymbol{\alpha}\cdot\boldsymbol{m}) \neq 0$ then $(u,\boldsymbol{m}) \in [-O(N), O(N)]^{s+2}$. So, if C is large enough in terms of s, we may replace the restriction $(u,\boldsymbol{m}) \in [-CN, CN]^{s+2}$ in (8-16) with the condition $(u,\boldsymbol{m}) \in \mathbb{R}^{s+2}$, without changing the value of (8-16).

Then, by the Gowers–Cauchy–Schwarz inequality (Proposition A.4) and the triangle inequality, (8-16) is

$$\begin{aligned} &\ll_C \biggl(X^{O(1)}\|g_1\|_{U^{s+1}(\mathbb{R})}^{2^{s+1}} + \sigma + \frac{\log X}{\sigma X}\biggr)^{1/2^{s+1}} \\ &\ll_C \biggl(X^{O(1)}\rho^{2^{s+1}} + \sigma + \frac{\log X}{\sigma X}\biggr)^{1/2^{s+1}}. \end{aligned} \quad (8\text{-}17)$$

Choosing $X = \rho^{-c_1}$, with c_1 suitably small in terms of s, and $\sigma = \rho^{c_1/2}$, expression (8-17) is $O_C(\rho^{\Omega(1)})$.

Putting this estimate into (8-10), we get a bound on (8-10) of

$$\ll_C Y^{O(1)}\rho^{\Omega(1)} + O\biggl(\frac{\log Y}{\sigma_F Y}\biggr). \quad (8\text{-}18)$$

Picking $Y = \rho^{-c_1}$, with c_1 suitably small in terms of s, we may ensure that (8-18) is $O_C(\rho^{\Omega(1)}\sigma_F^{-1})$, 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 $\tilde{T}^{L,\Xi,\tilde{\boldsymbol{r}}}_{F,G,N}(g_1, \dots, g_d)$. Thus

$$|\tilde{T}^{L,\Xi,\tilde{\boldsymbol{r}}}_{F,G,N}(g_1, \dots, g_d)| \ll_{c,C,\varepsilon} \biggl| \frac{1}{N^{h-m}} \int_{\boldsymbol{x}\in\mathbb{R}^{h-m}} F_1(\boldsymbol{x}) \prod_{j=1}^{d} g_j(\psi_j(\boldsymbol{x}) + a_j)\,d\boldsymbol{x}\biggr|, \quad (8\text{-}19)$$

where $\Psi : \mathbb{R}^{h-m} \to \mathbb{R}^d$ has $\Omega_{c,C}(1)$-Cauchy–Schwarz complexity at most s, for some s at most $d-2$, $F_1 : \mathbb{R}^{h-m} \to [0,1]$ 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 $a_j = O_{c,C,\varepsilon}(1)$. Furthermore $\|\Psi\|_\infty = O_C(1)$.

We apply Proposition 4.8 to Ψ. Therefore, for *any* real numbers $w_1, \ldots, w_{s+1}$,

$$|\tilde{T}^{L,\Xi,\tilde{\boldsymbol{r}}}_{F,G,N}(g_1,\ldots,g_d)| \ll \left|\frac{1}{N^{h-m}}\int_{\boldsymbol{x}\in\mathbb{R}^{h-m}} F_1\Bigl(\boldsymbol{x}+\sum_{k=1}^{s+1} w_k \boldsymbol{f_k}\Bigr)\prod_{j=1}^{d} g_j(\psi'_j(\boldsymbol{x},\boldsymbol{w})+a_j)\,d\boldsymbol{x}\right|, \tag{8-20}$$

where

- for each j at most d, $\psi'_j : \mathbb{R}^{h-m}\times\mathbb{R}^{s+1}\to\mathbb{R}$ is a linear form;
- $\psi'_1(\boldsymbol{0},\boldsymbol{w}) = w_1+\cdots+w_{s+1}$;
- $\boldsymbol{f_1},\ldots,\boldsymbol{f_{s+1}}\in\mathbb{R}^{h-m}$ are some vectors that satisfy $\|\boldsymbol{f_k}\|_\infty = O_{c,C}(1)$ for each k at most $s+1$;
- the system of forms $(\psi'_1,\ldots,\psi'_d)$ is in normal form with respect to ψ'_1.

We remark that the right-hand side of expression (8-20) is independent of $\boldsymbol{w}$, as it was obtained by applying the change of variables $\boldsymbol{x}\mapsto\boldsymbol{x}+\sum_{k=1}^{s+1} w_k\boldsymbol{f_k}$ 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 $P(\boldsymbol{x})\equiv 1$ if $\|\boldsymbol{x}\|_\infty \leqslant N/2$. Integrating over $\boldsymbol{w}$, we have that $|\tilde{T}^{L,\Xi,\tilde{\boldsymbol{r}}}_{F,G,N}(g_1,\ldots,g_d)|$ is

$$\ll_{c,C,\varepsilon} \frac{1}{N^{h-m+s+1}}\int_{\boldsymbol{w}\in\mathbb{R}^{s+1}} P(\boldsymbol{w})\left|\int_{\boldsymbol{x}\in\mathbb{R}^{h-m}} F_1\Bigl(\boldsymbol{x}+\sum_{k=1}^{s+1} w_k\boldsymbol{f_k}\Bigr)\prod_{j=1}^{d} g_j(\psi'_j(\boldsymbol{x},\boldsymbol{w})+a_j)\,d\boldsymbol{x}\right| d\boldsymbol{w}$$

$$\ll_{c,C,\varepsilon} \left|\frac{1}{N^{h-m+s+1}}\int_{\substack{\boldsymbol{x}\in\mathbb{R}^{h-m}\\ \boldsymbol{w}\in\mathbb{R}^{s+1}}} H(\boldsymbol{x},\boldsymbol{w})\prod_{j=1}^{d} g_j(\psi'_j(\boldsymbol{x},\boldsymbol{w})+a_j)\,d\boldsymbol{x}\,d\boldsymbol{w}\right|, \tag{8-21}$$

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

$$H(\boldsymbol{x},\boldsymbol{w}) := F_1\Bigl(\boldsymbol{x}+\sum_{k=1}^{s+1} w_k\boldsymbol{f_k}\Bigr)P(\boldsymbol{w}).$$

Since the vectors $\boldsymbol{f_k}$ satisfy

$$\|\boldsymbol{f_k}\|_\infty = O_{c,C}(1),$$

H is a Lipschitz function supported on $[-O_{c,C,\varepsilon}(N), O_{c,C,\varepsilon}(N)]^{h-m+s+1}$, with Lipschitz constant $O_{c,C}(1/\sigma_F N)$. Notice in (8-21) that we were able to move the absolute value signs outside the integral, as P is positive and the integral over $\boldsymbol{x}$ is independent of $\boldsymbol{w}$ (so in particular has constant sign).

Fix $\boldsymbol{x}$. Then the integral over $\boldsymbol{w}$ in (8-21) satisfies the hypotheses of Proposition 8.3. Applying Proposition 8.3 to this integral, and then integrating over $\boldsymbol{x}$, one derives

$$|\tilde{T}^{L,\Xi,\tilde{\boldsymbol{r}}}_{F,G,N}(g_1,\ldots,g_d)| \ll_{c,C,\varepsilon} \rho^{\Omega(1)}\sigma_F^{-1}.$$

Theorem 8.1 is proved. □

By our long series of reductions, this means that both Theorem 5.6 and Theorem 2.12 are proved. □

9. Constructions

In this section we prove Theorem 2.14, which, we remind the reader, is the partial converse of Theorem 2.12. In other words, we show that L being bounded away from $V^*_{\text{degen}}(m, d)$ is a necessary hypothesis for Theorem 2.12 to be true.

Proof of Theorem 2.14. Recall the hypotheses of Theorem 2.14. In particular, we suppose that

$$\liminf_{N\to\infty} \operatorname{dist}(L, V^*_{\text{degen}}(m, d)) = 0,$$

i.e., we assume that $\operatorname{dist}(L, V^*_{\text{degen}}(m, d)) = \omega(N)^{-1}$, for some function $\omega(N)$ such that

$$\limsup_{N\to\infty} \omega(N) = \infty.$$

Let η 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 $\omega(N) \geqslant \eta^{-1}$ and $\eta N \geqslant \max(1, \varepsilon)$. All implied constants to follow will be independent of η.

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

$$T^L_{F,G,N}(f_1, \ldots, f_d) = \frac{1}{N^{d-m}} \sum_{\substack{\boldsymbol{n}\in[N]^d \\ \|L\boldsymbol{n}\|_\infty \leqslant \varepsilon}} \prod_{j=1}^{d} f_j(n_j).$$

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

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

for some ρ at most 1 and that

$$T^L_{F,G,N}(f_1, \ldots, f_d) > H(\rho) + E_\rho(N). \tag{9-1}$$

We begin by observing that the condition $\|L\boldsymbol{n}\|_\infty \leqslant \varepsilon$ implies certain constraints on two of the variables n_i. Indeed, let $L' \in V^*_{\text{degen}}(m, d)$ be such that $\|L - L'\|_\infty = \operatorname{dist}(L, V^*_{\text{degen}}(m, d))$. Write λ_{ij} for the coefficients of L and λ'_{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 such that for all j in the range $3 \leqslant j \leqslant d$ we have

$$\sum_{i=1}^{m} a_i \lambda'_{ij} = 0, \tag{9-2}$$

and further we may assume that for all i we have $\lambda'_{i1} = \lambda_{i1}$ and $\lambda'_{i2} = \lambda_{i2}$ (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 $\boldsymbol{n} \in [N]^d$ be some solution to $\|L\boldsymbol{n}\|_\infty \leqslant \varepsilon$. The critical observation is that (9-2), combined with the assumptions on the a_i, implies that

$$|b_1 n_1 + b_2 n_2| \ll \eta N. \tag{9-3}$$

Indeed, for j in the range $3 \leqslant j \leqslant d$ 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 $\|L\boldsymbol{n}\|_\infty \leqslant \varepsilon$, 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 $\eta N \geqslant \varepsilon$.

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 η. All constants C_1 and C_2 to follow will be assumed to satisfy $C_1, C_2 = O_{c,C}(1)$.

Case 1 $|b_1|, |b_2| \leqslant c_1$.

Under the assumptions of Theorem 2.14, 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

$$\lambda''_{11} = \lambda'_{11} - \frac{b_1}{a_1} \quad \text{and} \quad \lambda''_{12} = \lambda'_{12} - \frac{b_2}{a_1}.$$

Then

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

for all j in the range $1 \leqslant j \leqslant d$. In other words we have shown that $\|L - L''\|_\infty \leqslant \eta + c_1$ for some matrix L'' with rank less than m. Since $\eta + c_1 < c$ (if c_1 is small enough), this implies that $\operatorname{dist}(L, V_{\mathrm{rank}}(m, d)) < c$, which contradicts the assumptions of Theorem 2.14. Therefore this case is indeed precluded.

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

In this case, (9-3) implies that $n_1 \leqslant C_1 \eta N$ for some constant C_1.[14] Now, define $f_1 : [N] \to [-1, 1]$ to be the indicator function of the interval $[\lceil C_1 \eta N \rceil, N] \cap \mathbb{N}$. We then have

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

[14] The same conclusion is true for n_2, but this will not be needed.

for some constant C_2. However, observe that

$$|T^L_{F,G,N}(f_1-1,1,\dots,1)| = |T^L_{F,G,N}(f_1,1,\dots,1) - T^L_{F,G,N}(1,1,\dots,1)|$$
$$= |0 - T^L_{F,G,N}(1,1,\dots,1)|$$
$$\gg_{c,C,\varepsilon} 1$$

by the hypotheses of Theorem 2.14. If $T^L_{F,G,N}(f_1-1,1,\dots,1)$ did not satisfy (9-1), then

$$1 \ll_{c,C,\varepsilon} H(\rho) + E_\rho(1),$$

where $\rho := C_2(C_1\eta)^{(s+2)/2^{s+1}}$. Picking η small enough, then N large enough, this inequality cannot possibly hold, and we have a contradiction. So $T^L_{F,G,N}(f_1-1,1,\dots,1)$ satisfies (9-1).

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

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 coupled in such a way that condition (9-3) is very likely to hold, we can guarantee that $T^L_{F,G,N}(f_1-p, f_2-p, 1,\dots,1)$ 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 conclude that $\|f_1-p\|_{U^{s+1}[N]} = o(1)$ as $N\to\infty$, and the same for f_2.

We now fill in the technical details. Relation (9-3) implies that

$$|b_1n_1+b_2n_2| \leqslant C_1\eta N, \tag{9-4}$$

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 η, but in magnitude we may pick $C_2 \asymp 1$. We consider the real interval $[0,N]$ modulo N, and for $x\in[0,N]$ and i in the range $0\leqslant i\leqslant (C_1C_2\eta)^{-1}-1$ we define the half-open interval modulo N

$$I_{x,i} := [x + iC_1C_2\eta N, x+(i+1)C_1C_2\eta N).$$

This choice guarantees that

$$[0,N] = \bigcup_{i=0}^{(C_1C_2\eta)^{-1}-1} I_{x,i}, \tag{9-5}$$

and the union is disjoint. Now, for δ a small constant to be chosen later,[15] we define

$$I^\delta_{x,i} := \big[x + \big(i+\tfrac12-\delta\big)C_1C_2\eta N, x+\big(i+\tfrac12+\delta\big)C_1C_2\eta N\big).$$

We will use the partition (9-5) to construct a function f_1, using an averaging argument to choose an x so that the $I^\delta_{x,i}$ intervals capture a positive proportion of the solution density of the linear inequality

[15] This δ is unrelated to the notation $\delta = T^L_{F,G,N}(f_1,\dots,f_d)$ used in previous sections.

system. Indeed, for $n_1 \in [N]$ let the weight $u(n_1)$ denote the number of $d-1$-tuples $n_2, \dots, n_d \leqslant N$ that together with n_1 satisfy the inequality $\|L\boldsymbol{n}\|_\infty < \varepsilon$. The weight $u(n_1)$ could be zero, of course. Let

$$E_{x,\delta} := \bigcup_i I_{x,i}^\delta.$$

Then

$$\begin{aligned}\frac{1}{N}\int_0^N \sum_{n \in [N]} u(n) 1_{E_{x,\delta}}(n)\,dx &= \frac{1}{N}\sum_{n\in[N]} u(n) \int_0^N 1_{E_{x,\delta}}(n)\,dx \\ &= \sum_{n \in [N]} u(n) 2\delta \\ &= 2\delta N^{d-m} T^L_{F,G,N}(1,\dots,1)\end{aligned}$$

Therefore, by the assumptions of Theorem 2.14, we may fix an x such that

$$\sum_{n\in[N]} u(n) 1_{E_{x,\delta}}(n) \gg_{c,C} \delta N^{d-m} T^L_{F,G,N}(1,\dots,1). \tag{9-6}$$

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 $I_{x,i} \cap \mathbb{N}$ to be members of A, with probability p, or none of $I_{x,i} \cap \mathbb{N}$ to be members of A, with probability $1-p$. We then make this same choice for each i in the range $0 \leqslant i \leqslant (C_1C_2\eta)^{-1} - 1$, independently. Observe immediately that 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

$$J_{x,i} = \frac{b_1}{|b_2|} I_{x,i} \cap (0, N],$$

where the dilation of the interval $I_{x,i}$ is not considered modulo N but rather just as an operator on subsets of $\mathbb{R}$ (see Section 1B for this notation). Since $b_1 \geqslant |b_2|$ we have that these $J_{x,i}$ also form a disjoint partition of $[0, N]$. (NB: If $b_1 > |b_2|$ it may be that certain $J_{x,i}$ are empty, since the dilate of the corresponding $I_{x,i}$ may land entirely outside $[0, N]$). Then let B be the subset of $[N]$ defined so that for each i with $J_{x,i}$ nonempty we have $J_{x,i} \cap \mathbb{N} \subseteq B$ if and only if $I_{x,i} \cap \mathbb{N} \subseteq A$. 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 δ,

$$|\mathbb{E} T^L_{F,G,N}(f_1, f_2, 1, \dots, 1) - T^L_{F,G,N}(p, p, 1, \dots, 1)| \gg_{c,C,\varepsilon} \delta^2. \tag{9-7}$$

Indeed, suppose that $I_{x,i}$ is included in the set A, and suppose that $n_1 \in I_{x,i}^\delta$. If $n_2 \in [N]$ satisfies $\left|\frac{b_1}{|b_2|} n_1 - n_2\right| \leqslant \frac{1}{b_2} C_1 \eta N$ and if δ is small enough in terms of b_1 and b_2, then $n_2 \in J_{x,i}$.[16] Thus, by the

[16]This fact is the reason why we introduced the parameter δ.

observation (9-4), $n_2 \in B$, for every integer n_2 that is the second coordinate of a solution vector $\boldsymbol{n}$ for which the first coordinate is n_1.[17] Therefore

$$\begin{aligned}
\mathbb{E}T^L_{F,G,N}(f_1, f_2, 1, \dots, 1) &= \frac{1}{N^{d-m}} \sum_{\substack{\boldsymbol{n}\in[N]^d \\ \|L\boldsymbol{n}\|_\infty \leqslant \varepsilon}} \mathbb{P}(n_1 \in A \wedge n_2 \in B) \\
&\geqslant \frac{1}{N^{d-m}} \sum_{\substack{\boldsymbol{n}\in[N]^d \\ \|L\boldsymbol{n}\|_\infty \leqslant \varepsilon}} \mathbb{P}(n_1 \in A \wedge n_1 \in I^\delta_{x,i} \text{ for some } i \wedge n_2 \in B) \\
&\geqslant \frac{1}{N^{d-m}} \sum_{\substack{\boldsymbol{n}\in[N]^d \\ \|L\boldsymbol{n}\|_\infty \leqslant \varepsilon}} \mathbb{P}(n_1 \in A \wedge n_1 \in I^\delta_{x,i} \text{ for some } i) \\
&= \frac{1}{N^{d-m}} \sum_{n_1\in[N]} u(n_1) p 1_{E_{x,\delta}}(n_1) \\
&\geqslant 2\delta p T^L_{F,G,N}(1, \dots, 1),
\end{aligned}$$

where the final line follows from (9-6). On the other hand $T^L_{F,G,N}(p, p, 1, \dots, 1) = p^2 T^L_{F,G,N}(1, \dots, 1)$, and hence

$$\mathbb{E}T^L_{F,G,N}(f_1, f_2, 1, \dots, 1) - T^L_{F,G,N}(p, p, 1, \dots, 1) \geqslant (2\delta p - p^2) T^L_{F,G,N}(1, \dots, 1). \tag{9-8}$$

Picking p small enough in terms of δ, and using the assumption that $T^L_{F,G,N}(1, \dots, 1) = \Omega_{c,C,\varepsilon}(1)$, this proves the relation (9-7).

Our second claim is that

$$\mathbb{E}\|f_1 - p\|_{U^{s+1}[N]}, \mathbb{E}\|f_2 - p\|_{U^{s+1}[N]} \ll \eta^{1/2^{s+1}}. \tag{9-9}$$

We first consider f_1. Then

$$\mathbb{E}\|f_1 - p\|^{2^{s+1}}_{U^{s+1}[N]} \ll \frac{1}{N^{s+2}} \sum_{(x,\boldsymbol{h})\in\mathbb{Z}^{s+2}} \mathbb{E}\Bigg(\prod_{\boldsymbol{\omega}\in\{0,1\}^{s+1}} (f_1 - p1_{[N]})(x + \boldsymbol{h}\cdot\boldsymbol{\omega})\Bigg).$$

Observe that for fixed $(x, \boldsymbol{h})$ the random variables $(f_1 - p1_{[N]})(x + \boldsymbol{h}\cdot\boldsymbol{\omega})$ each have mean zero and, unless some two of the expressions $x + \boldsymbol{h}\cdot\boldsymbol{\omega}$ 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

$$\mathbb{E}\Bigg(\prod_{\boldsymbol{\omega}\in\{0,1\}^{s+1}} (f_1 - p1_{[N]})(x + \boldsymbol{h}\cdot\boldsymbol{\omega})\Bigg) = \prod_{\boldsymbol{\omega}\in\{0,1\}^{s+1}} \mathbb{E}((f_1 - p1_{[N]})(x + \boldsymbol{h}\cdot\boldsymbol{\omega})) = 0.$$

Therefore,

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

[17] i.e., a vector $\boldsymbol{n}$ such that $\|L\boldsymbol{n}\|_\infty \leqslant \varepsilon$.

where

$$R=\{\boldsymbol{h}:|\boldsymbol{h}\cdot(\boldsymbol{\omega_1}-\boldsymbol{\omega_2})|\leqslant C_1C_2\eta N \text{ for some } \boldsymbol{\omega_1},\boldsymbol{\omega_2}\in\{0,1\}^{s+1},\ \boldsymbol{\omega_1}\neq\boldsymbol{\omega_2}\}.$$

Thus by Jensen's inequality we have

$$\mathbb{E}\|f_1-p\|_{U^{s+1}[N]}\ll\eta^{1/2^{s+1}}, \tag{9-10}$$

as claimed in (9-9).

The calculation for f_2 is essentially identical, noting that the length of the blocks $J_{x,i}$ 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 $f_1,\dots,f_d$ that satisfied (9-1). Then, by (9-7), if we pick p to be small enough in terms of δ we have

$$\begin{aligned}\delta^2&\ll_{c,C,\varepsilon}|\mathbb{E}T^L_{F,G,N}(f_1,f_2,1,\dots,1)-T^L_{F,G,N}(p,p,1,\dots,1)|\\&\ll|\mathbb{E}T^L_{F,G,N}(f_1-p,f_2,1,\dots,1)|+|\mathbb{E}T^L_{F,G,N}(p,f_2-p,1,\dots,1)|\\&\ll\mathbb{E}(H(\rho_1)+E_{\rho_1}(N))+\mathbb{E}(H(\rho_2)+E_{\rho_2}(N)),\end{aligned} \tag{9-11}$$

where ρ_1 (resp. ρ_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 ρ_i may be random variables themselves.

We claim that the random variables ρ_1 and ρ_2 may be chosen so that the right-hand side of (9-11) is $\kappa(\eta)+o_\eta(1)$ as $N\to\infty$. To prove this, we make two observations. Note first that by Markov's inequality

$$\mathbb{P}(\|f_1-p\|_{U^{s+1}[N]}\geqslant\eta^{1/2^{s+2}})\ll\eta^{1/2^{s+2}}$$

We choose the (random) upper-bound ρ_1 satisfying

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

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,

$$\mathbb{E}(H(\rho_1)+E_{\rho_1}(N))\ll H(\mathbb{E}\rho_1)+\mathbb{E}(E_{\rho_1}(1))\ll\kappa(\eta^{1/2^{s+2}})+o_\eta(1)\ll\kappa(\eta)+o_\eta(1). \tag{9-12}$$

We do the same manipulation for f_2. Combining (9-12) with (9-11) we conclude that

$$\delta^2\ll_{c,C,\varepsilon}\kappa(\eta)+o_\eta(1). \tag{9-13}$$

The only condition on δ occurred in the proof of (9-7), in which we assumed that δ was small enough in terms of b_1 and b_2. Therefore there exists a suitable δ that satisfies $\delta=\Omega_{c,C}(1)$. Picking such a δ, and then picking η small enough and N large enough, (9-13) is a contradiction. So there must be some functions $f_1,\dots,f_d$ that satisfy (9-1).

Case 4 *Exactly one of* b_1, b_2 *satisfies* $b_i \geqslant c_1$.

Without loss of generality we may assume that $b_1 \geqslant c_1$. But then, as in Case 2, (9-3) implies that $n_1 \leqslant C_1 \eta N$ 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 2.14. □

Appendix A: Gowers norms

There are several existing accounts of the basic theory of Gowers norms — for example in [Green 2007] and [Tao 2012] — 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 [Green and Tao 2010a]. 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 are 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 nonnegative solution to

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

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

For example,

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

and

$$\|f\|_{U^2(N)} = \left(\frac{1}{N^3} \sum_{x, h_1, h_2} f(x) \overline{f(x + h_1) f(x + h_2)} f(x + h_1 + h_2) \right)^{1/4}.$$

It is not immediately obvious that the right-hand side of (A-1) is always a nonnegative real, nor why the U^d norms are genuine norms if $d \geqslant 2$: proofs of both these facts may be found in [Tao and Vu 2006]. An immediate Cauchy–Schwarz argument, which may also be found in [loc. cit.], gives the so-called "nesting property" of Gowers norms, namely the fact that

$$\|f\|_{U^2(N)} \leqslant \|f\|_{U^3(N)} \leqslant \|f\|_{U^4(N)} \leqslant \cdots.$$

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 $N' \geqslant N$. 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 nonnegative real solution to the equation

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

where $f1_{[N]}$ is the extension by zero of f to $\mathbb{Z}/N'\mathbb{Z}$, the summation is over $x, h_1, \dots, h_d \in \mathbb{Z}/N'\mathbb{Z}$, and R is the set

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

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 $\|f\|_{U^d[N]} \asymp \|f1_{[N]}\|_{U^d(N')}$. (See [Green and Tao 2010a, Lemma B.5] for more detail on this).

We observe that there is only a contribution to the summand in (A-2) when $x \in [N]$ and for every i we have $h_i \in \{-N, -N+1, \dots, N-1, N\}$ modulo N'. Further, it may be easily seen that $|R| \asymp N^{d+1}$. Therefore, choosing N'/N sufficiently large, we conclude that

$$\|f\|_{U^d[N]} \asymp \biggl(\frac{1}{N^{d+1}} \sum_{x,h_1,\dots,h_d \in \mathbb{Z}} \prod_{\boldsymbol{\omega} \in \{0,1\}^d} \mathcal{C}^{|\boldsymbol{\omega}|} f(x + \boldsymbol{h} \cdot \boldsymbol{\omega}) \biggr)^{1/2^d}. \tag{A-3}$$

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 [Cook et al. 2017, Lemma 4.2; Durcik et al. 2018, 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 [Tao 2015]).

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 nonnegative real satisfying

$$\|f\|_{U^d(\mathbb{R})}^{2^d} = \int_{(x,\boldsymbol{h}) \in \mathbb{R}^{d+1}} \prod_{\boldsymbol{\omega} \in \{0,1\}^d} \mathcal{C}^{|\boldsymbol{\omega}|} f\Bigl(x + \sum_{i=1}^{d} h_i \omega_i\Bigr) \, dx \, dh_1 \cdots dh_d \tag{A-4}$$

where $|\boldsymbol{\omega}| = \sum_i \omega_i$, and $\mathcal{C}$ 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

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

Explicitly, a change of variables shows that

$$\|g\|_{U^d(\mathbb{R},N)}^{2^d} \asymp \frac{1}{N^{d+1}} \int_{(x,\boldsymbol{h})\in\mathbb{R}^{d+1}} \prod_{\boldsymbol{\omega}\in\{0,1\}^d} \mathcal{C}^{|\boldsymbol{\omega}|} g\Big(x+\sum_{i=1}^{d} h_i\omega_i\Big)\, dx\, dh_1 \cdots dh_d. \tag{A-5}$$

We require one further fact about Gowers norms.

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

$$\langle (f_{\boldsymbol{\omega}})_{\boldsymbol{\omega}\in\{0,1\}^d}\rangle := \int_{(x,\boldsymbol{h})\in\mathbb{R}^{d+1}} \prod_{\boldsymbol{\omega}\in\{0,1\}^d} \mathcal{C}^{|\boldsymbol{\omega}|} f_{\boldsymbol{\omega}}\Big(x+\sum_{i=1}^{d} h_i\omega_i\Big)\, dx\, dh_1 \cdots dh_d.$$

Then

$$|\langle (f_{\boldsymbol{\omega}})_{\boldsymbol{\omega}\in\{0,1\}^d}\rangle| \leqslant \prod_{\boldsymbol{\omega}\in\{0,1\}^d} \|f_{\boldsymbol{\omega}}\|_{U^d(\mathbb{R})}.$$

Proof. See [Tao and Vu 2006, Chapter 11] for the proof in the finite group setting. The modification to the setting of the reals is trivial. □

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 $F : \mathbb{R}^m \to \mathbb{C}$ is Lipschitz, with Lipschitz constant at most M, if

$$M \geqslant \sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^m \\ \boldsymbol{x}\neq\boldsymbol{y}}} \frac{|F(\boldsymbol{x}) - F(\boldsymbol{y})|}{\|\boldsymbol{x}-\boldsymbol{y}\|_\infty}.$$

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

$$M \geqslant \sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^m/\mathbb{Z}^m \\ \boldsymbol{x}\neq\boldsymbol{y}}} \frac{|G(\boldsymbol{x}) - G(\boldsymbol{y})|}{\|\boldsymbol{x}-\boldsymbol{y}\|_{\mathbb{R}^m/\mathbb{Z}^m}}.$$

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 σ be some parameter in the range $0 < \sigma < \frac{1}{2}$. Then there exist Lipschitz functions $F_\sigma, G_\sigma : \mathbb{R}^m \to [0, 1]$ supported on $[-2N, 2N]^m$, both with Lipschitz constant at most $O\big(\frac{1}{\sigma N}\big)$, such that*

$$1_K = F_\sigma + O(G_\sigma)$$

and $\int_{\boldsymbol{x}} G_\sigma(\boldsymbol{x})\, d\boldsymbol{x} = O(\sigma N^m)$. Furthermore, $F_\sigma(\boldsymbol{x}) \geqslant 1_K(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^m$, and G is supported on $\{\boldsymbol{x} \in \mathbb{R}^m : \operatorname{dist}(\boldsymbol{x}, \partial(K)) \leqslant \sigma N\}$.

This is [Green and Tao 2010a, Corollary A.3]. It was be used in Lemmas 5.9 and 5.11 to replace sums with sharp cutoffs by sums with Lipschitz cutoffs.

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

$$F(\boldsymbol{x}) = \sum_{\substack{\boldsymbol{k}\in\mathbb{Z}^m \\ \|\boldsymbol{k}\|_\infty \leqslant X}} c_X(\boldsymbol{k}) e(\boldsymbol{k}\cdot\boldsymbol{x}) + O\biggl(M\frac{\log X}{X}\biggr) \tag{B-1}$$

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

This is [Green and Tao 2008b, Lemma A.9], and was used in Lemma 3.4 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 $F : \mathbb{R}^m \to \mathbb{C}$ be a Lipschitz function, supported on $[-CN, CN]^m$, such that $\|F\|_\infty \leqslant 1$ and the Lipschitz constant of F is at most M. Then*

$$F(\boldsymbol{x}) = \int_{\substack{\boldsymbol{\xi}\in\mathbb{R}^m \\ \|\boldsymbol{\xi}\|_\infty \leqslant X}} c_X(\boldsymbol{\xi}) e\biggl(\frac{\boldsymbol{\xi}\cdot\boldsymbol{x}}{N}\biggr) d\boldsymbol{\xi} + O_C\biggl(MN\frac{\log X}{X}\biggr) \tag{B-2}$$

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

Lemma B.4 is very similar to Lemma B.3, and may be easily proved by adapting that standard harmonic analysis argument found in [Green and Tao 2008b, Lemma A.9] from $\mathbb{R}^m/\mathbb{Z}^m$ to $\mathbb{R}^m$. For completeness, we sketch the proof.

Sketch of proof. By rescaling the variable $\boldsymbol{x}$ 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

$$K_X(\boldsymbol{x}) := \prod_{i=1}^{m} \frac{1}{X}\biggl(\frac{\sin(\pi X x_i)}{\pi x_i}\biggr)^2.$$

Then

$$\widehat{K_X}(\boldsymbol{\xi}) = \prod_{i=1}^{m} \max\biggl(1 - \frac{|\xi_i|}{X}, 0\biggr).$$

We have

$$(F * K_X)(\boldsymbol{x}) = \int_{\substack{\boldsymbol{\xi}\in\mathbb{R}^m \\ \|\boldsymbol{\xi}\|_\infty \leqslant X}} \widehat{F}(\boldsymbol{\xi})\widehat{K_X}(\boldsymbol{\xi}) e(\boldsymbol{\xi}\cdot\boldsymbol{x})\, d\boldsymbol{x},$$

and, since $|\widehat{F}(\boldsymbol{\xi})| \leqslant \|F\|_1 \ll_C 1$, letting $c_X(\boldsymbol{\xi}) := \widehat{F}(\boldsymbol{\xi})\widehat{K_X}(\boldsymbol{\xi})$ gives a main term of the desired form.

It remains to show that

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

By writing

$$|F(\boldsymbol{x}) - (F * K_X)(\boldsymbol{x})| = \biggl|\int_{\boldsymbol{y}\in\mathbb{R}^m} (F(\boldsymbol{x}) - F(\boldsymbol{y})) K_X(\boldsymbol{x}-\boldsymbol{y})\, d\boldsymbol{y}\biggr|,$$

one sees that it suffices to show that

$$\int_{\|z\|_\infty \leqslant 2C} \|z\|_\infty K_X(z)\, dz \ll_C \frac{\log X}{X}.$$

But this bound follows immediately from a dyadic decomposition. □

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 4.8.

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*

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

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

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 $\{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_\infty \leqslant C,\ \operatorname{dist}(\boldsymbol{x}, V(S)) \geqslant c\}$. □

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 $\binom{d}{m}$, 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 $\|L\|_\infty \leqslant C$ and $\operatorname{dist}(L, V_{\text{rank}}(m, d)) \geqslant c$ 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 immediate 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 $1 \leqslant j \leqslant d$. (Recall that $(\lambda_{ij})_{i \leqslant m, j \leqslant d}$ denotes the coefficients of L). Restricting to j in the range $1 \leqslant j \leqslant m$, 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 $a_i = O_{c,C,C_1}(1)$ for all i.

The final part of the proposition is to show that if $\operatorname{dist}(L, V_{\mathrm{rank}}^{\mathrm{unif}}(m, d)) \geqslant c$ then, for each j, there exists a rank matrix of L that does not include the j-th column. But this statement follows immediately from the above, after having deleted the j-th column. □

We now prove Proposition 4.8 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 $V_{\mathcal{P}_i}$, which consists of all systems of linear forms for which the partition $\mathcal{P}_i$ is not "suitable", the reader may consult Definition 4.4. The reader may also find the example that follows the proof to be informative.

Proof of Proposition 4.8. Fix i, and let $\mathcal{P}_i$ be a partition of $[m] \setminus \{i\}$ such that $\operatorname{dist}(\Psi, V_{\mathcal{P}_i}) \geqslant c_1$ (such a $\mathcal{P}_i$ exists by the definition of c_1-Cauchy–Schwarz complexity, i.e., by Definition 4.6). The partition $\mathcal{P}_i$ has $s_i + 1$ parts, for some s_i at most s.

It is clear from Definition 4.6 that we may, without loss of generality, further subdivide the partition and assume that the partition $\mathcal{P}_i$ has exactly $s+1$ parts. Call the parts $\mathcal{C}_1$ through $\mathcal{C}_{s+1}$.

Following Section 4 of [Green and Tao 2010a], for each $k \in [s+1]$ there exists a vector $\boldsymbol{f_k} \in \mathbb{R}^n$ that witnesses the fact that $\operatorname{dist}(\Psi, V_{\mathcal{P}_i}) > 0$, i.e., for which $\psi_i(\boldsymbol{f_k}) = 1$ but $\psi_j(\boldsymbol{f_k}) = 0$ for all $j \in \mathcal{C}_k$. Such a vector can be found using Gaussian elimination, say. Consider the extension

$$\Psi'(\boldsymbol{u}, w_1, \ldots, w_{s+1}) := \Psi(\boldsymbol{u} + w_1 \boldsymbol{f_1} + \cdots + w_{s+1} \boldsymbol{f_{s+1}}).$$

Then, if $\Psi' = (\psi'_1, \ldots, \psi'_m)$, the form $\psi'_i(\boldsymbol{u}, w_1, \ldots, w_{s+1})$ is the only one that uses all of the w_k variables. Furthermore, $\psi'_i(\boldsymbol{0}, \boldsymbol{w}) = w_1 + \cdots + w_{s+1}$. Also, $n' = n+s+1$, which is at most $n+m-1$. So Proposition 4.8 is proved if for each k we can find such a vector $\boldsymbol{f_k}$ that additionally satisfies $\|\boldsymbol{f_k}\|_\infty = O_{c_1,C_1}(1)$.

Consider a fixed k, and let Γ 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 $\boldsymbol{f_k}$. If in the course of implementing these algorithms we are given a free choice for a coordinate of $\boldsymbol{f_k}$, 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 $\boldsymbol{f_k}$. 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

$$Q_\gamma := \prod_{j \leqslant n} q_{\gamma,j}.$$

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 $\boldsymbol{f_k}$, and this in turn implies that $\mathcal{P}_i$ is not suitable for Ψ and hence that $\Psi \in V_{\mathcal{P}_i}$.

Since $V(I) \subseteq V_{\mathcal{P}_i}$, the assumptions of Proposition 4.8 imply that $\operatorname{dist}(\Psi, V(I)) \geqslant c_1$. Applying Proposition C.1 to the polynomials $\{Q_\gamma : \gamma \in \Gamma\}$, we conclude that there is some $\gamma \in \Gamma$ such that $|Q_\gamma(\Psi)| = \Omega_{c_1,C_1}(1)$. In particular, we conclude that the solution vector $\boldsymbol{f_k}$ obtained by the implementation γ has coefficients that are $O_{c_1,C_1}(1)$. This concludes the proof of Proposition 4.8. □

Let us illustrate the above proof with an example which we hope will be instructive. 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 $\boldsymbol{f_1}$ simply by applying Gaussian elimination. Implementing the algorithm a certain way we have

$$\boldsymbol{f_1} = \begin{pmatrix} a_{22}/(a_{11}a_{22} - a_{12}a_{21}) \\ -a_{21}/(a_{11}a_{22} - a_{12}a_{21}) \\ 0 \end{pmatrix}$$

as a solution, in the case where $a_{11}a_{22} - a_{12}a_{21}$ is nonzero. Of course if $a_{11}a_{23} - a_{13}a_{21}$ is nonzero too, we have another solution

$$\boldsymbol{f_1} = \begin{pmatrix} a_{23}/(a_{11}a_{23} - a_{13}a_{21}) \\ 0 \\ -a_{21}/(a_{11}a_{23} - a_{13}a_{21}) \end{pmatrix}.$$

So, if one applied Gaussian elimination idly, one might end up with either of these two solutions. Unfortunately it could be the case that $\operatorname{dist}(\Psi, V_{\mathcal{P}_i}) \geqslant c_1$ whilst one of these determinants, $a_{11}a_{22} - a_{12}a_{21}$ say, was nonzero yet $o(1)$ (as the unseen variable N, on which Ψ will ultimately depend, tends to infinity). In this instance, applying the first implementation of the algorithm would not give a desirable solution vector $\boldsymbol{f_1}$. For this reason we had to apply somewhat indirect arguments in order to find the appropriate vector $\boldsymbol{f_1}$.

It is worth including a brief discussion on why these quantitative subtleties do not arise in the setting of [Green and Tao 2010a]. Indeed, assume that Ψ has rational coefficients of naive height at most C_1 and that $\Psi \notin V_{\mathcal{P}_i}$. Since there are only $O_{C_1}(1)$ many possible choices of Ψ we immediately conclude that $\operatorname{dist}(\Psi, V_{\mathcal{P}_i}) \gg_{C_1} 1$, without needing to assume this as an extra hypothesis. Then *any* implementation of Gaussian elimination succeeds in finding a suitably bounded $\boldsymbol{f_k}$, 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 5.10.

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

Lemma D.1. *Let m, d be natural numbers, with $d\geqslant m+1$, and let c, C, l be positive constants. Let $L:\mathbb{R}^d\to\mathbb{R}^m$ be a surjective linear map, and suppose $\|L\|_\infty\leqslant C$ and $\operatorname{dist}(L, V_{\mathrm{rank}}(m,d))\geqslant c$. Let K denote* $\ker L$. *Let R be a convex set contained in $[-l,l]^m$. Then, if $\boldsymbol{v}\in K^{\perp}$, $L\boldsymbol{v}\in R$ only when $\boldsymbol{v}\in R'$, 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 $\boldsymbol{\lambda_i}\in\mathbb{R}^d$ denote the column vector such that $\boldsymbol{\lambda_i}^T$ is the i-th row of L. Since $\operatorname{dist}(L, V_{\mathrm{rank}}(m,d))\geqslant c$, the vectors $\boldsymbol{\lambda_i}$ are linearly independent. Moreover, we may extend the set $\{\boldsymbol{\lambda_i}: i\leqslant m\}$ by orthogonal vectors of unit length to form a basis $\{\boldsymbol{\lambda_i}: i\leqslant d\}$ for $\mathbb{R}^d$.

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

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

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

Now notice that $\operatorname{span}(\boldsymbol{\lambda_i}: m+1\leqslant i\leqslant d)$ is exactly equal to K. Let $\boldsymbol{v}\in K^{\perp}$, and suppose $L\boldsymbol{v}\in R$. Letting L' be the d-by-d matrix whose rows are $\boldsymbol{\lambda_i}^T$, we have that $L'\boldsymbol{v}=\boldsymbol{w}$ for some vector $\boldsymbol{w}$ satisfying $\|\boldsymbol{w}\|_\infty\ll l$. Premultiplying by the matrix $A=(a_{ki})$, we immediately get $\boldsymbol{v}=A\boldsymbol{w}$, and hence $\|\boldsymbol{v}\|_\infty=O_{c,C}(l)$. The region $R':=(L^{-1}R)\cap K^{\perp}$ is therefore bounded. R' is clearly convex, and so the lemma is proved. □

The second lemma concerns vectors, with integer coordinates, that lie near to a subspace.

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

$$\operatorname{dist}(\boldsymbol{n}, \Xi(\mathbb{R}^h)+\tilde{\boldsymbol{r}})\leqslant\eta. \tag{D-1}$$

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

Proof. By replacing $\boldsymbol{n}$ with $\boldsymbol{n}-\tilde{\boldsymbol{r}}$, we can assume without loss of generality that $\tilde{\boldsymbol{r}}=\boldsymbol{0}$. It will also be enough to show that $\boldsymbol{n}\in\Xi(\mathbb{R}^h)$, as the injectivity of Ξ 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 $\boldsymbol{m}$.

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

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

$$\operatorname{dist}((\widetilde{\Xi})^{-1}(\boldsymbol{n}),(\widetilde{\Xi})^{-1}(\Xi(\mathbb{R}^h)))\gg C^{-O(1)}.$$

Applying $\widetilde{\Xi}$, we conclude that

$$\operatorname{dist}(\boldsymbol{n},\Xi(\mathbb{R}^h))\gg C^{-O(1)},$$

which is a contradiction to (D-1) if η is small enough. □

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

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

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

Finally, we give the linear algebra argument used to construct the matrix P during the proof of Lemma 5.10.

Lemma D.4. *Let m,d be natural numbers, with $d\geqslant m+1$. 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 $\|L\|_\infty\leqslant C$ and $\|\Theta\|_\infty\leqslant C$. Equating L with its matrix, suppose that the first m columns of L form the identity matrix. Let $\{\boldsymbol{a_1},\dots,\boldsymbol{a_u}\}$ be a basis for the lattice $\Theta L(\mathbb{Z}^d)$ that satisfies $\|\boldsymbol{a_i}\|_\infty=O_C(1)$ for every i. Let $\boldsymbol{x_1},\dots,\boldsymbol{x_u}\in\mathbb{Z}^d$ be vectors such that, for every i, $\Theta L(\boldsymbol{x_i})=\boldsymbol{a_i}$ and $\|\boldsymbol{x_i}\|_\infty=O_C(1)$. Then*

$$\mathbb{R}^m=\operatorname{span}(L\boldsymbol{x_i}:i\leqslant u)\oplus\ker\Theta \tag{D-2}$$

and there is an invertible linear map $P:\mathbb{R}^m\to\mathbb{R}^m$ such that

$$P((\operatorname{span}(L\boldsymbol{x_i}:i\leqslant u)))=\mathbb{R}^u\times\{0\}^{m-u},\quad P(\ker\Theta)=\{0\}^u\times\mathbb{R}^{m-u},$$

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

Note that both $\{\boldsymbol{a_1}, \ldots, \boldsymbol{a_u}\}$ and $\boldsymbol{x_1}, \ldots, \boldsymbol{x_u} \in \mathbb{Z}^d$ exist by applying Lemma 5.7 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 Θ has integer coefficients.

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

$$\begin{aligned} P(L\boldsymbol{x_i}) &:= \boldsymbol{b_i}, & 1 \leqslant i \leqslant u, \\ P(\boldsymbol{y_j}) &:= \boldsymbol{b_{j+u}}, & 1 \leqslant j \leqslant m-u, \end{aligned} \tag{D-3}$$

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

To this end, for a standard basis vector $\boldsymbol{b_k}$ write

$$\boldsymbol{b_k} = \sum_{i=1}^{u} \lambda_i L\boldsymbol{x_i} + \sum_{j=1}^{d-u} \mu_j \boldsymbol{y_j}.$$

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, $\boldsymbol{b_k} \in L(\mathbb{Z}^d)$. Also $\Theta(\boldsymbol{b_k}) = \sum_{i=1}^{u} \lambda_i \boldsymbol{a_i}$. So $\boldsymbol{a} := \sum_{i=1}^{u} \lambda_i \boldsymbol{a_i}$ is an element of $\Theta L(\mathbb{Z}^d)$ that satisfies $\|\boldsymbol{a}\|_\infty = O_C(1)$. Since $\|\boldsymbol{a_i}\|_\infty = O_C(1)$ for every i, and $\{\boldsymbol{a_1}, \ldots, \boldsymbol{a_u}\}$ is a basis for the lattice $\Theta L(\mathbb{Z}^d)$, this implies that $|\lambda_i| = O_C(1)$ for every i.

So then $\sum_{j=1}^{d-u} \mu_j \boldsymbol{y_j}$ is a vector in $\ker\Theta$ satisfying $\bigl\|\sum_{j=1}^{d-u} \mu_j \boldsymbol{y_j}\bigr\|_\infty = O_C(1)$. Since $\{\boldsymbol{y_1}, \ldots, \boldsymbol{y_{m-u}}\}$ 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 $d \geqslant m + 1$, and let c, C be positive constants. Let $L : \mathbb{R}^d \to \mathbb{R}^m$ 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.5 for definitions). Suppose that $\|L\|_\infty \leqslant C$, that $\operatorname{dist}(L, V_{\mathrm{rank}}(m, d)) \geqslant c$, and that L has rational complexity at most C. Let τ_1, τ_2 be two parameters in the range $0 < \tau_1, \tau_2 \leqslant 1$. 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 5.10 and replace L by the map L' (expression (5-10)). By part (9) of Lemma 5.10, $A_{L'}(\tau_1, \tau_2) \ll_{c,C} A_L(\Omega_{c,C}(\tau_1), \Omega_{c,C}(\tau_2))$. 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 τ_1 and τ_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 $\tau_1 \leqslant \|\alpha\|_\infty \leqslant \tau_2^{-1}$ and

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

Fix α and φ 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 α 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 3.1), 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

$$\operatorname{dist}(\beta, (\mathbb{Z}^m)^T) < c_1 \min(\tau_1, \tau_2^{C_2}). \tag{E-3}$$

We know that $\|\beta\|_\infty = \Omega_{c,C}(\tau_1)$. Also, considering (E-3), by perturbing β 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

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

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

$$\|L^*(M^{-1})^*\rho - \varphi\|_\infty < c_1 C_3 \tau_2^{C_2} \tag{E-4}$$

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 α 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 could not 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 φ 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 nonzero), we may pick some coordinate i at most d for which $(L^*(M^{-1})^*\rho)_i - \varphi_i \neq 0$. So there are algebraic numbers $\lambda_1, \ldots, \lambda_m$ with algebraic degree $O_k(1)$ and algebraic height $O_{c,C,k,H}(1)$ for which

$$0 < \left| \sum_{j=1}^{m} \lambda_j \rho_j - \varphi_i \right| < c_1 C_3 \tau_2^{C_2}, \tag{E-5}$$

where $(\rho_1 \ \cdots \ \rho_m)$ is the representation of ρ 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 $Q\big(\sum_{j\leqslant m} \lambda_j \rho_j\big) = 0$, and then to apply Liouville's original argument.

Assume without loss of generality that each $\lambda_j \rho_j$ is nonzero. 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 λ_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 [Cohen 1993, Section 4.2.1]), this implies that there is a polynomial $Q \in \mathbb{Z}[X]$ with degree $O_k(1)$ such that $Q\big(\sum_{j\leqslant m} \lambda_j \rho_j\big) = 0$ and $H(Q) = O_{c,C,k,H}(\tau_2^{-O_k(1)})$.

Now, it could be that φ_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, $H(Q \cdot (X - \varphi_i)^{-1}) \ll_{c,C,k,H} (\varphi_i + 1)^{O_k(1)} \tau_2^{-O_k(1)}$. By repeating this process as necessary, since $|\varphi_i| = O_{c,C}(\tau_2^{-1})$ we may assume therefore that φ_i is not a root of Q.

This immediately implies a bound on the derivative of Q, namely that, for any θ,

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

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

$$1 \leqslant |Q(\varphi_i)| = \left| Q\left(\sum_{j=1}^{m} \lambda_j \rho_j \right) - Q(\varphi_i) \right| \leqslant |Q'(\theta)| \left| \sum_{j=1}^{m} \lambda_j \rho_j - \varphi_i \right| \ll_{c,C,k,H} c_1 C_3 \tau_2^{-O_k(1)} \tau_2^{C_2}.$$

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. □

Acknowledgements

We would like to thank several anonymous referees, for their very careful reading of earlier versions of this work, and Ben Green, for his advice and comments. We also benefited from conversations with Sam Chow, Trevor Wooley, and Yufei Zhao. Some of the paper was 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 project the author was supported by EPSRC grant no. EP/M50659X/1.

References

[Baker 1967] A. Baker, "On some Diophantine inequalities involving primes", *J. Reine Angew. Math.* **228** (1967), 166–181. MR Zbl

[Baker 1986] R. C. Baker, *Diophantine inequalities*, Lond. Math. Soc. Monogr. (N.S) **1**, Oxford Univ. Press, 1986. MR Zbl

[Cohen 1993] H. Cohen, *A course in computational algebraic number theory*, Grad. Texts in Math. **138**, Springer, 1993. MR Zbl

[Cook et al. 2017] B. Cook, A. Magyar, and M. Pramanik, "A Roth-type theorem for dense subsets of $\mathbb{R}^d$", *Bull. Lond. Math. Soc.* **49**:4 (2017), 676–689. MR Zbl

[Davenport 1963] H. Davenport, *Analytic methods for Diophantine equations and Diophantine inequalities*, Ann Arbor Publ., 1963. MR Zbl

[Davenport and Heilbronn 1946] H. Davenport and H. Heilbronn, "On indefinite quadratic forms in five variables", *J. Lond. Math. Soc.* **21** (1946), 185–193. MR Zbl

[Durcik et al. 2018] P. Durcik, V. Kovač, and L. Rimanić, "On side lengths of corners in positive density subsets of the Euclidean space", *Int. Math. Res. Not.* **2018**:22 (2018), 6844–6869. MR Zbl

[Freeman 2002] D. E. Freeman, "Asymptotic lower bounds and formulas for Diophantine inequalities", pp. 57–74 in *Number theory for the millennium, II* (Urbana, IL, 2000), edited by M. A. Bennett et al., Peters, Natick, MA, 2002. MR Zbl

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

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

[Green 2007] B. Green, "Montréal notes on quadratic Fourier analysis", pp. 69–102 in *Additive combinatorics* (Montréal, 2006), edited by A. Granville et al., CRM Proc. Lecture Notes **43**, Amer. Math. Soc., Providence, RI, 2007. MR Zbl

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

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

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

[Green and Tao 2010b] B. Green and T. Tao, "Yet another proof of Szemerédi's theorem", pp. 335–342 in *An irregular mind*, edited by I. Bárány and J. Solymosi, Bolyai Soc. Math. Stud. **21**, Bolyai Math. Soc., Budapest, 2010. MR Zbl

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

[Green et al. 2012] B. Green, T. Tao, and T. Ziegler, "An inverse theorem for the Gowers $U^{s+1}[N]$-norm", *Ann. of Math.* (2) **176**:2 (2012), 1231–1372. MR Zbl

[Margulis 1989] G. A. Margulis, "Discrete subgroups and ergodic theory", pp. 377–398 in *Number theory, trace formulas and discrete groups* (Oslo, 1987), edited by K. E. Aubert et al., Academic Press, Boston, 1989. MR Zbl

[McCutcheon 2005] R. McCutcheon, "FVIP systems and multiple recurrence", *Israel J. Math.* **146** (2005), 157–188. MR Zbl

[Müller 2005] W. Müller, "Systems of quadratic Diophantine inequalities", *J. Théor. Nombres Bordeaux* **17**:1 (2005), 217–236. MR Zbl

[Parsell 2002a] S. T. Parsell, "Irrational linear forms in prime variables", *J. Number Theory* **97**:1 (2002), 144–156. MR Zbl

[Parsell 2002b] S. T. Parsell, "On simultaneous diagonal inequalities, III", *Q. J. Math.* **53**:3 (2002), 347–363. MR Zbl

[Tao 2012] T. Tao, *Higher order Fourier analysis*, Grad. Stud. Math. **142**, Amer. Math. Soc., Providence, RI, 2012. MR Zbl

[Tao 2015] T. Tao, "An inverse theorem for the continuous Gowers uniformity norm", blog post, 2015, Available at https://tinyurl.com/taogowers.

[Tao and Teräväinen 2018] T. Tao and J. Teräväinen, "Odd order cases of the logarithmically averaged Chowla conjecture", *J. Théor. Nombres Bordeaux* **30**:3 (2018), 997–1015. MR Zbl

[Tao and Teräväinen 2019] T. Tao and J. Teräväinen, "The structure of logarithmically averaged correlations of multiplicative functions, with applications to the Chowla and Elliott conjectures", *Duke Math. J.* **168**:11 (2019), 1977–2027. MR Zbl

[Tao and Vu 2006] T. Tao and V. Vu, *Additive combinatorics*, Cambridge Stud. Adv. Math. **105**, Cambridge Univ. Press, 2006. MR Zbl

[Vaughan 1981] R. C. Vaughan, *The Hardy–Littlewood method*, Cambridge Tracts in Math. **80**, Cambridge Univ. Press, 1981. MR Zbl

[Walker 2019] A. Walker, "Linear inequalities in primes", 2019. To appear in *J. Anal. Math.* arXiv

[Wooley 2003] T. D. Wooley, "On Diophantine inequalities: Freeman's asymptotic formulae", pp. art. id. 30 in *Proc. Session in Analytic Number Theory and Diophantine Equations* (Bonn, Germany, 2002), edited by D. R. Heath-Brown and B. Z. Moroz, Bonner Math. Schriften **360**, Univ. Bonn, 2003. MR Zbl

Communicated by Roger Heath-Brown
Received 2019-03-06 Revised 2019-10-30 Accepted 2020-02-06

aledwalker@gmail.com *Centre de Recherches Mathématiques, Montréal QC, Canada*

Algebra & Number Theory

msp.org/ant

See inside back cover or msp.org/ant for submission instructions.

The subscription price for 2020 is US $415/year for the electronic version, and $620/year (+$60, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP.

Algebra & Number Theory (ISSN 1944-7833 electronic, 1937-0652 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

ANT peer review and production are managed by EditFLOW® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

Algebra & Number Theory

Volume 14 No. 6 2020

Unobstructedness of Galois deformation rings associated to regular algebraic conjugate self-dual cuspidal automorphic representations 1331
DAVID-ALEXANDRE GUIRAUD

The Hilbert scheme of hyperelliptic Jacobians and moduli of Picard sheaves 1381
ANDREA T. RICOLFI

Endomorphism algebras of geometrically split abelian surfaces over $\mathbb{Q}$ 1399
FRANCESC FITÉ and XAVIER GUITART

Uniform Yomdin–Gromov parametrizations and points of bounded height in valued fields 1423
RAF CLUCKERS, ARTHUR FOREY and FRANÇOIS LOESER

Gowers norms control diophantine inequalities 1457
ALED WALKER

Modular invariants for real quadratic fields and Kloosterman sums 1537
NICKOLAS ANDERSEN and WILLIAM D. DUKE

Generically free representations, I: Large representations 1577
SKIP GARIBALDI and ROBERT GURALNICK

Classification of some vertex operator algebras of rank 3 1613
CAMERON FRANC and GEOFFREY MASON

1937-0652(2020)14:6;1-W