Quadratic phenomena in additive combinatorics and number theory

Sofia Lindqvist
Keble College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Trinity 2019

Acknowledgements

I am very thankful to my supervisor Prof. Ben Green for his support, for many valuable discussions and suggestions, for his guidance and for his willingness to invest so much time on his students.

I wish to thank my supervisor Prof. Roger Heath-Brown for all the things he taught me during the first two years of my DPhil, and for all his valuable feedback during this period.

I would also like to thank Dr. Sam Chow and Dr. Sean Prendiville for an excellent collaboration.

I am thankful to the anonymous referees of the paper [48] for their very helpful remarks and suggestions, which greatly improved the structure of that paper, included here as Chapter 3.

Finally, I am thankful to James Aaronson, Dr. Aled Walker, Dr. Jakub Konieczny, Dr. Rudi Mrazović, Dr. Sean Eberhard and Dr. Freddie Manners for many valuable discussions.

During my DPhil I have been supported by Ben Greens Simons Investigator Grant number 376201, by Ben Green's ERC Starting Grant 279438, *Approximate Algebraic Structure and Applications*, and by an EPSRC award. I am very thankful for these sources of funding.

In February 2017 I took part in the workshop in analytic number theory at the Mathematical Sciences Research Institute in Berkeley. I am grateful for the use of their facilities during this period.

Abstract

This thesis deals with problems related to the structure of the solutions to some specific polynomial equations. A brief introduction to the type of problems we are interested in is given in Chapter 1.

In Chapter 2 we recall some standard results in number theory and additive combinatorics.

In Chapter 3 we look at partition regularity of equations of the form $x^a + y^b = z^c$ over $\mathbb{Z}/p\mathbb{Z}$. In particular we look at the equation $x + y = z^2$.

In Chapter 4 we prove that any 2-colouring of $\mathbb{N}$ has infinitely many monochromatic solutions to the equation $x + y = z^2$. This work is joint with Ben Green.

In Chapter 5 we use the same methods as in Chapter 4 to prove partition regularity of the equation $x - y = z^2$.

In Chapter 6 we show that a linear combination of kth powers is partition regular if and only if the corresponding linear equation is partition regular, provided the number of variables is large enough. This is based on joint work with Sam Chow and Sean Prendiville.

In Chapter 7 we look at Heath-Brown's method of counting the zeros of a quadratic form in four variables, and in particular how the error term in this count is affected by the weight function used.

In Chapter 8 we try to count the number of zeros of a quadratic form in four variables that lie in a fixed congruence class.

Contents

Notation

The notation presented here is used throughout this thesis.

We let $\mathbb{Z}$, $\mathbb{N}$, $\mathbb{R}$ and $\mathbb{C}$ be the set of integers, natural numbers (0 is not a natural number), reals and complex numbers, respectively. The notation $\mathbb{Z}_{\geqslant 0}$ denotes the set $\{x \in \mathbb{Z} : x \geqslant 0\}$, with a similar definition for $\mathbb{R}_{\geqslant 0}$.

For q a prime power we write $\mathbb{F}_q$ for the finite field of size q. If $q = p$ is a prime we identify $\mathbb{F}_p$ with $\mathbb{Z}/p\mathbb{Z}$.

For a set A we let 1_A be the characteristic function of A, that is

$$1_A(x) \coloneqq \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A. \end{cases}$$

For sets A and B we define $A + B \coloneqq \{a + b : a \in A, b \in B\}$ and $A - B \coloneqq \{a - b : a \in A, b \in B\}$.

We write $\mathbb{E}_{x \in S} \coloneqq \frac{1}{|S|} \sum_{x \in S}$ for the expectation over some set S, as is standard in additive combinatorics. When the set S is clear from context we will just write $\mathbb{E}_x$.

We write $e_q(x) \coloneqq e^{2\pi i x/q}$ and $e(x) \coloneqq e^{2\pi i x}$.

Let G be an additive group and let $f : G \to \mathbb{C}$ be a function. We define the multiplicative differencing operator by

$$\Delta_h f(x) \coloneqq \overline{f(x+h)} f(x).$$

We use standard Landau and Vinogradov asymptotic notation: for functions f and positive valued functions g we write $f = O(g)$ if $|f| \leqslant Cg$ for some absolute constant C. This is also denoted by $f \ll g$ or $g \gg f$. The notation $f \asymp g$ is equivalent to $f \ll g \ll f$. We write $f = o(g)$ as $x \to \infty$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$, and $f \sim g$ as $x \to \infty$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$. We write $f = O_\alpha(g)$ if the implied constant in the O-notation depends on a parameter α, with similar notation for $\ll, \gg, \asymp, \sim$ and o.

If G is a compact Abelian group we denote probability Haar measure on G by μ_G. We always use counting measure on $\mathbb{N}$ and $\mathbb{Z}$, denoted by either $|\cdot|$ or $\#$. We say that a subset A of $\mathbb{N}$ is *dense* if $\limsup_{N\to\infty} \frac{|A \cap [N]|}{N} > 0$.

We write $\lfloor x \rfloor$ for the greatest integer less than or equal to x and $\lceil x \rceil$ for the smallest integer greater than or equal to x.

For $Y \geqslant 1$, let $[Y] = \{1, 2, \dots, \lfloor Y \rfloor\}$.

We write $\mathbb{T}^d$ for the d-dimensional torus $\mathbb{R}^d/\mathbb{Z}^d$, and we endow it with the metric $(\alpha, \beta) \mapsto \|\alpha - \beta\|$, where

$$\|\alpha\| := \sum_{i=1}^{d} \min_{n \in \mathbb{Z}} |\alpha_i - n|.$$

We warn the reader that in Chapter 3 we use a slightly different norm on $\mathbb{T}^d$, but the notation used there should not cause any confusion.

For a function $f : \mathbb{Z} \to \mathbb{C}$ or $f : \mathbb{N} \to \mathbb{C}$ we define

$$\|f\|_p := \begin{cases} \left(\sum_x |f(x)|^p\right)^{1/p} & \text{if } p < \infty \\ \max_x |f(x)| & \text{if } p = \infty \end{cases}$$

where the sums are over $\mathbb{Z}$ or $\mathbb{N}$ respectively.

For a function $f : \mathbb{Z}/q\mathbb{Z} \to \mathbb{C}$ we define

$$\|f\|_p := \begin{cases} \left(\mathbb{E}_x |f(x)|^p\right)^{1/p} & \text{if } p < \infty \\ \max_x |f(x)| & \text{if } p = \infty. \end{cases}$$

If G is a ring we write G^* for the multiplicative group of invertible elements in G.

We write

$$\sum_{x \,(\mathrm{mod}\ q)}{}^{*} := \sum_{\substack{x \,(\mathrm{mod}\ q) \\ (x,q)=1}} \quad \text{or} \quad \mathbb{E}^*_{x \,(\mathrm{mod}\ q)} := \frac{1}{\phi(q)} \sum_{\substack{x \,(\mathrm{mod}\ q) \\ (x,q)=1}}$$

for sums over $(\mathbb{Z}/q\mathbb{Z})^*$.

Chapter 1

Introduction

Our main interest lies in problems concerning the solutions of various polynomial equations. More precisely, there are two main flavours of problems we consider.

The first property we investigate is that of partition regularity. We say that an equation is *partition regular* in, for example, the integers, if any finite colouring of the integers admits a non-trivial solution to the equation where all variables have the same colour. In particular, partition regularity implies solvability. An even stronger property than partition regularity is that any dense set has a solution to the equations in question. This property is seen in for example Roth's theorem, which states that any dense set of integers contains a 3-term arithmetic progression. As we shall see, partition regularity is a natural property to ask for in the cases where such a density result does not hold.

The second flavour of problems we consider is about counting the number of solutions to an equation. In particular, we are interested in quadratic forms in few variables. This can be done for example by looking at the solutions in a box with side lengths that tend to infinity, or, as we do in chapters 7 and 8, by counting the solutions weighted by some weight function. In chapter 8 we restrict this further by only counting the solutions in some fixed congruence class.

A common theme in all the equations we consider is that they are quadratic in nature. This is not strictly true, as we also prove results about more general equations, but in all cases these more general results are motivated by their quadratic special cases.

We now give a brief introduction to these central topics, before giving an outline of the remainder of this thesis.

1.1 Partition regularity

As indicated above, we define partition regularity as follows.

Definition (Partition regularity)**.** Given a system of equations and a set S, we say that the system is *partition regular over* S if for any finite colouring of S one can find a non-trivial solution to the equations with all variables of the same colour.

The notion of non-trivial in the above definition may vary from case to case. This restriction is included to avoid degenerate cases like a solution where all variables are equal, as such a solution trivially has all variables the same colour.

We now mention some important results about partition regularity. One of the first such results is Schur's theorem [59], which states that the equation

$$x + y = z$$

is partition regular over the natural numbers. Later van der Waerden [68] showed that any finite colouring of the natural numbers contains arbitrarily long monochromatic arithmetic progressions. That is, for any $k \in \mathbb{N}$ the system

$$\{x, x + y, \ldots, x + (k-1)y\}$$

is partition regular over $\mathbb{N}$.

Both of these results fall under the framework of linear systems of equations. In [56] Rado fully classified all linear systems of equations that are partition regular over $\mathbb{N}$. Indeed, a system of linear equations is partition regular if and only if the coefficients satisfy an algebraic condition known as *the columns condition* or *Rado's criterion*. For the purposes of this text we will only be concerned with the case of a single equation, and so we note that in this case Rado's theorem states that the equation

$$a_1 x_1 + \cdots + a_s x_s = 0,$$

where $a_1, \cdots, a_s \in \mathbb{Z} \setminus \{0\}$ are integer coefficients and $x_1, \cdots, x_s$ are the variables, is partition regular over $\mathbb{N}$ if and only if some nonempty subset of $\{a_1, \ldots, a_s\}$ sums to zero. As a special case this of course covers both Schur's theorem and Van der Waerden's theorem.

For linear equations where the coefficients sum to zero one can say even more. A famous theorem of Roth [57] states that any dense subset of $\mathbb{N}$ contains a 3-AP, that is, a non-trivial solution to the equation

$$x + y = 2z.$$

Note that such a statement is not true for an equation where the coefficients do not sum to zero. For example, the set of odd numbers is certainly dense, but contains no solution to Schur's equation $x + y = z$.

Roth's theorem was later generalised to produce Szemerédi's theorem [60], which states that any dense subset of $\mathbb{N}$ contains arbitrarily long arithmetic progressions. Thus Szemerédi's theorem is a much stronger version of van der Waerden's theorem, and in fact any such density result trivially implies the Ramsey counterpart. Indeed, given a finite colouring, at least one colour class must be dense, and then one can apply the density result to a dense colour class to get a monochromatic solution.

So far we have only mentioned linear equations and configurations. The Bergelson–Leibman theorem [5] is a vast generalisation of Szemerédi's theorem. Instead of looking at AP's it looks at configurations of the form

$$\{x + p_1(y), \dots, x + p_s(y)\}$$

where $p_1, \dots, p_s \in \mathbb{Z}[x]$ are polynomials satisfying $p_1(0) = \cdots = p_s(0)$. Just as in Szemerédi's theorem, the Bergelson–Leibman theorem gives that any dense set will contain such a configuration.

The Bergelson–Leibman theorem is one of many results in this area proved using ergodic methods. In addition to many density results, the ergodic methods have also provided several genuine colouring results, in the sense that a density counterpart does not hold. A survey of some such results can be found in [2]. As an example we mention that the equation

$$x - y = z^2$$

is partition regular over $\mathbb{N}$.[1]

One can of course also consider partition regularity over other sets than $\mathbb{N}$. As we shall see $\mathbb{F}_p$ is in many ways a much easier setting to work in than $\mathbb{N}$, so in addition to being of interest in its own right, it may be of interest to establish partition regularity results over $\mathbb{F}_p$ before attempting to prove their counterparts over $\mathbb{N}$. A famous open problem of Hindman [40] asks whether the configuration

$$\{x, y, xy, x + y\}$$

is partition regular over $\mathbb{N}$. In [32] Green–Sanders proved that this configuration is in fact partition regular over $\mathbb{F}_p$, provided p is large enough. For comparison, Moreira

[1] Note that taking the set of odd numbers shows that for this equation it is not true that any dense set contains a solution.

[52] proved that the configuration

$$\{x, xy, x+y\}$$

is partition regular over $\mathbb{N}$, but the full conjecture of Hindman is still open.

In the setting of squares, a famous open problem of Erdős and Graham [28] asks whether or not the Pythagorean equation

$$x^2 + y^2 = z^2$$

is partition regular over $\mathbb{N}$. This is of course equivalent to asking whether or not Schur's equation is partition regular over the set of squares.

1.2 Counting zeros of quadratic forms

A famous problem of Waring asks if, for each $k \geqslant 2$, there is an s such that every sufficiently large[2] natural number can be expressed as a sum of exactly s kth powers. In 1909 Hilbert proved that this is indeed the case. It is still of great interest to find the best possible bounds on the size of s needed in Waring's problem. Let $G(k)$ be the smallest integer such that any sufficiently large integer can be expressed as the sum of $G(k)$ kth powers.

Hardy and Littlewood showed that $G(k) \leqslant k^2 + 1$. Their method, known as the Hardy–Littlewood circle method, gives more than just an upper bound on $G(k)$. They get that for any large $N \in \mathbb{N}$ the equation

$$x_1^k + \ldots + x_s^k = N \tag{1.1}$$

has $\mathfrak{S}(N)N^{\frac{s}{k}-1}(1+o(1))$ solutions, provided $s \geqslant k^2 + 1$. Here the *singular series* $\mathfrak{S}(N)$ satisfies the bounds $1 \ll \mathfrak{S}(N) \ll 1$.

Subsequently there were many improvements to the bounds on $G(k)$ but we will not go into the full history here. For a detailed account of this see [67].

We will be particularly interested in the case $k = 2$. Jacobi gave an explicit formula for the number of representations of an integer as the sum of four squares. This is a case which is not covered by the classical Hardy–Littlewood method, but in [43] Kloosterman refined the methods of Hardy and Littlewood to obtain an asymptotic count for the number of solutions to (1.1) in the case $k = 2$ and $s = 4$.

[2]Strictly speaking Waring asked for all natural numbers, but in the modern version of Waring's problem one usually only cares about sufficiently large integers

The Hardy–Littlewood circle method can also be applied to other polynomial equations. In particular, we are interested in counting the solutions to

$$c_1 x_1^2 + \cdots + c_s x_s^2 = 0 \tag{1.2}$$

where $x_i \leqslant X$ for $i = 1, \ldots, s$. In this case the Hardy–Littlewood circle method gives that there are asymptotically CX^{s-2} solutions as $X \to \infty$, provided $s \geqslant 5$. When trying to reduce the number of variables below 5 the regular circle method is not enough.

In [37] Heath-Brown used what is known as the Heath-Brown circle method or the δ-method to find an asymptotic count for the number of solutions to (1.2) in the cases $s = 4$ and $s = 3$. His methods do not require the quadratic forms to be diagonal, and so they give asymptotic results for any non-singular quadratic forms in 3 or 4 variables. There are examples of such forms with no zeros[3], but we will not consider such forms here. For the remaining cases, perhaps somewhat surprisingly, the asymptotic count for $s = 4$ variables is either of order X^2 or of order $X^2 \log X$, depending on whether or not the discriminant of the quadratic form is a square or not. For $s = 3$ the asymptotic count is always of order $X \log X$.

1.3 Outline

In the next chapter we recall some standard results from the literature for later use.

In Chapter 3 we prove partition regularity of the generalised Fermat equation $x^a + y^b = z^c$ in $\mathbb{F}_p$, for $a, b, c \in \mathbb{N}$. This chapter is essentially the paper [48]. The methods used are based very heavily on the paper [32].

In Chapter 4 we prove that any 2-colouring of $\mathbb{N}$ has a monochromatic solution to the equation $x + y = z^2$. This chapter is essentially the joint paper [31]. In Chapter 5 we show how the methods in Chapter 4 can be used to prove partition regularity of the equation $x - y = z^2$.

Chapter 6 is based on the joint paper [14]. We do not reproduce the full contents of [14] here. Instead, we aim to prove a slightly weaker version of the main result of [14], namely that the equation $c_1 x_1^k + \cdots + c_s x_s^k = 0$ is partition regular over $\mathbb{N}$ if and only if some nonempty subset of the coefficients $c_1, \ldots, c_s$ sum to zero. We prove this result provided $s \geqslant k^2 + 1$, while in [14] we give a proof of this result where one may take $s \geqslant 8$ for $k = 3$ and $s \geqslant k(\log k + O(\log\log k))$ for $k \geqslant 4$. The most interesting application of our result is probably the fact that the equation $x_1^2 + x_2^2 + x_3^2 + x_4^2 = x_5^2$ is

[3] other than the trivial $x_1 = x_2 = \ldots = 0$

partition regular, which does not require the improvements on the number of variables for $k \geqslant 3$ presented in [14].

Chapters 7 and 8 are about counting zeroes of quadratic forms in four variables. In Chapter 8 we consider the problem of counting such zeros restricted to some congruence class. This is done using the Heath-Brown circle method. In Chapter 7 we study the error term for the unrestricted count, and in particular how this depends on the choice of weight function used in the Heath-Brown circle method.

This thesis is structured in such a way that chapters 3 and onwards, except for Chapter 5, can be read independently from one another. Chapter 5 depends heavily on Chapter 4, and so these two chapters should be read together. All of these later chapters reference results stated in Chapter 2.

Finally, in appendices A and B we include two results which are tangentially related to Chapters 3 and 4, respectively. In Appendix C we establish some facts about smooth cutoff functions used in Chapter 4 and Chapter 5. In Appendix D we establish several technical results used in Chapter 6.

Chapter 2

Results from the literature

In this chapter we collect some standard results from number theory and additive combinatorics for later reference.

2.1 Exponential sums

We begin by recording Euler–Maclaurin summation as presented in [66, Lemma 4.2], and summation by parts as presented in [1, Theorem 4.2].

Lemma 2.1.1 (Euler–Maclaurin summation)**.** *Suppose that $X < Y$, F' exists and is continuous on $[X, Y]$ and F' is monotonic on $[X, Y]$. Let H_1, H_2 denote integers such that $H_1 \leqslant F'(\alpha) \leqslant H_2$ for every $\alpha \in [X, Y]$. Then*

$$\sum_{X<x\leqslant Y} e(F(x)) = \sum_{h=H_1}^{H_2} \int_X^Y e(F(\alpha) - \alpha h) d\alpha + O(\log(2 + \max\{|H_1|, |H_2|\})).$$

Lemma 2.1.2 (Summation by parts)**.** *Let $(a_n)_{n=1}^{\infty}$ be a sequence of real or complex numbers. Write $A(t) = \sum_{n\leqslant t} a_n$. Suppose that $X < Y$ and that F is continuously differentiable on $[X, Y]$. Then*

$$\sum_{X<n\leqslant Y} a_n F(n) = A(Y)F(Y) - A(X)F(X) - \int_X^Y A(u)F'(u)du.$$

Next we record the standard Gauss sum bound, found in e.g. [18, Chapter 9].

Lemma 2.1.3 (Gauss sum bound)**.** *For $q, a \in \mathbb{N}$ it holds that*

$$\Big| \sum_{x \pmod q} e_q(ax^2) \Big| \leqslant q^{1/2}(q, a)^{1/2}.$$

More generally we have the following.

Lemma 2.1.4. *For p a prime and g a polynomial of degree at most k with integral coefficients, which isn't identically zero, it holds that*

$$\frac{1}{p}\Big|\sum_{x(\mathrm{mod}\ p)} e_p(g(x))\Big| \ll p^{-1/2^{k-1}}.$$

Dirichlet's approximation lemma [66, Lemma 2.1] states the following.

Lemma 2.1.5 (Dirichlet's approximation lemma)**.** *Let $\alpha \in \mathbb{R}$. Then for each real $X \geqslant 1$ there exists a rational number a/q with $(a,q) = 1$, $1 \leqslant q \leqslant X$ and*

$$|\alpha - a/q| \leqslant \frac{1}{qX}.$$

We will need two versions of Weyl's inequality. The standard version is as follows, found in [66, Lemma 2.4].

Lemma 2.1.6 (Weyl's inequality)**.** *Suppose that $(a,q) = 1$, $|\alpha - a/q| \leqslant q^{-2}$ and $g(x) = \alpha x^k + \alpha_{k-1}x^{k-1} + \cdots \alpha_1 + x$. Then*

$$\sum_{n \leqslant N} e(g(n)) \ll_\epsilon N^{1+\epsilon}(q^{-1} + N^{-1} + qN^{-k})^{1/2^{k-1}}.$$

We will also need a version of Weyl's inequality without the factor N^ϵ. Such "ϵ-free" results are well-known to experts, but it is hard to locate a convenient reference. Wooley [70] discusses the pure power case (that is, sums of the form $\sum_{n \leq N} e(\alpha n^k)$), and it is likely that the same methods apply in greater generality, though the verification of this would involve a foray into the inner workings of [66, Chapter 4].

A self-contained source for the purposes of this chapter is [35, Lemma 4.4] (described in that paper as a "reformulation" of Weyl's inequality, a slightly inaccurate statement). Here is the statement.

Proposition 2.1.7 (Weyl's inequality, ϵ-free version)**.** *Let $k \in \mathbb{N}$. Then there is a constant C_k such that the following is true. Let $0 < \delta < 1/2$. Let $g : \mathbb{Z} \to \mathbb{R}$ be a polynomial of degree k with leading coefficient α_k. Suppose that $|\mathbb{E}_{n \in I} e(g(n))| \geq \delta$, where $I \subset \mathbb{Z}$ is a discrete interval. Then there is some $q \in \mathbb{N}$, $q \leq \delta^{-C_k}$, such that $\|q\alpha_k\|_{\mathbb{R}/\mathbb{Z}} \leq \delta^{-C_k}|I|^{-k}$.*

In Chapter 4 we will need this result in the cases $k = 2$ and $k = 4$. The proof in the latter case is essentially as hard as that of the general case. We remark that in Lemma [35, Lemma 4.4] the result is stated with $I = [N]$, but the general case follows trivially from this by translation (which does not affect the leading coefficient α_k).

2.2 Fourier analysis

We will make frequent use of Fourier analysis. We first look at the Fourier transform over a finite additive group G (see also [62, Chapter 4]). As we will only need the case $G = \mathbb{Z}/q\mathbb{Z}$ we specialise to $G = \mathbb{Z}/q\mathbb{Z}$ for now.

Let $f : G \to \mathbb{C}$. The Fourier transform of f is then defined by

$$\widehat{f}(\xi) = \mathbb{E}_{x \in G} f(x) e_q(-\xi x),$$

where $\xi \in G$. Note that here we have identified the additive character $x \mapsto e_q(\xi x)$ with the element $\xi \in G$.

We now recall the standard properties of the Fourier transform that we will need. The inversion formula tells us that for all $x \in G$ it holds that

$$f(x) = \sum_{\xi \in G} \widehat{f}(\xi) e_q(\xi x).$$

Parseval's identity states that

$$\|f\|_2^2 = \mathbb{E}_x |f(x)|^2 = \sum_{\xi \in G} |\widehat{f}(\xi)|^2.$$

The convolution between two functions $f, g : G \to \mathbb{C}$ is defined by

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

We then have that convolution is transformed to multiplication under the Fourier transform, that is,

$$\widehat{f * g}(\xi) = \widehat{f}(\xi)\widehat{g}(\xi).$$

We will also need the Fourier transform for functions on $\mathbb{Z}^d$. Let $g : \mathbb{Z}^d \to \mathbb{C}$ have compact support. The Fourier transform of g is then defined by

$$\widehat{g}(\alpha) = \sum_{x \in \mathbb{Z}^d} g(x) e(-\alpha \cdot x)$$

where $\alpha \in \mathbb{T}^d$ and $\alpha \cdot x = \alpha_1 x_1 + \cdots + \alpha_d x_d$. Then the inversion formula, Parseval's identity and the convolution result still hold, but with the appropriate normalisations and averages. In other words,

$$g(x) = \int_{\mathbb{T}^d} \widehat{g}(\alpha) e(\alpha \cdot x) d\alpha,$$

$$\|g\|_2^2 = \sum_{x \in \mathbb{Z}^d} |g(x)|^2 = \int_{\mathbb{T}^d} |\widehat{g}(\alpha)|^2 d\alpha$$

and

$$\widehat{g_1 * g_2}(\alpha) = \widehat{g_1}(\alpha)\widehat{g_2}(\alpha),$$

where

$$g_1 * g_2(x) = \sum_{y \in \mathbb{Z}^d} g_1(x-y)g_2(y).$$

For a function $\phi : \mathbb{T}^d \to \mathbb{C}$ we can of course then define

$$\widehat{\phi}(x) = \int_{\mathbb{T}^d} \phi(\alpha)e(-\alpha \cdot x)d\alpha,$$

and the usual results still hold.

Finally we mention that for $g : \mathbb{R} \to \mathbb{R}$ the Fourier transform is defined by

$$\widehat{g}(\xi) = \int_{\mathbb{R}} g(x)e(-\xi x)dx.$$

We will only ever need this for continuous functions g of compact support, and so we do not need to worry about convergence here.

For functions $g : \mathbb{R} \to \mathbb{R}$ one has the *Poisson summation formula*, which states that

$$\sum_{n \in \mathbb{Z}} g(n) = \sum_{k \in \mathbb{Z}} \widehat{g}(k).$$

2.3 Controlling linear patterns

Let $f : G \to \mathbb{C}$ for some Abelian group G. The Gowers U^k-norm is then defined by

$$\|f\|_{U^k(G)}^{2^k} = \mathbb{E}_{x,h_1,\dots,h_k \in G} \prod_{\omega \in \{0,1\}^k} \mathcal{C}^{|\omega|} f(x + h \cdot \omega),$$

where $\mathcal{C}$ denotes complex conjugation. It is a well-known fact that for $k \geqslant 2$ the Gowers U^k-norm truly is a norm, thus justifying the name. When the group G is clear from context we will just write $\|\cdot\|_{U^k}$ instead of $\|\cdot\|_{U^k(G)}$.

We also define the Gowers norm for functions supported on the interval $[N]$ by

$$\|f\|_{U^k([N])} = \|f\|_{U^k(\mathbb{Z}/M\mathbb{Z})},$$

where we have defined $M = 2^k N$ and extended f to a function on $\mathbb{Z}/M\mathbb{Z}$ in the obvious way.

We note that for $k = 2$ the U^2-norm reduces to

$$\|f\|_{U^2(G)}^4 = \mathbb{E}_{x,h_1,h_2} f(x)\overline{f(x+h_1)f(x+h_2)}f(x+h_1+h_2) = \sum_{r \in G} |\widehat{f}(r)|^4, \qquad (2.1)$$

and so the U^2-norm is very closely related to the Fourier transform of f.

The importance of Gowers norms in additive combinatorics is highlighted by the following result, which follows by work of Green and Tao (for an explicit statement see [61, Exercise 1.3.23.]).

Theorem 2.3.1 (Generalised von Neumann inequality)**.** *Let $(\psi_1, \cdots, \psi_t)$ be a collection of linear forms, $\psi_i : \mathbb{Z}^d \to \mathbb{Z}$, of Cauchy–Schwarz complexity s. Let $f_1, \cdots, f_t$ be functions $f_i : [N] \to \mathbb{C}$ with $\|f_i\|_\infty \leqslant 1$. Then*

$$|\mathbb{E}_{x_1,\ldots,x_d \in [N]} f_1(\psi_1(x_1, \cdots, x_d)) \cdots f_t(\psi_t(x_1, \cdots, x_d))| \leqslant \min_{1 \leqslant i \leqslant t} \|f_i\|_{U^{s+1}([N])}.$$

This theorem tells us is that linear configurations in a set are controlled by Gowers norms. For example, if a set A contains no kAP's then $\|1_A\|_{U^{k-1}}$ must be "small". The same result also holds with a cyclic group in place of $[N]$. We will not define Cauchy–Schwarz complexity here, but for the full definition see [61, Definition 1.3.2].

We also mention a class of (semi)norms closely related to the Gowers norms. Let $\mathrm{Poly}_{\leqslant d}$ be the set of polynomials $\mathbb{F}_p^n \to \mathbb{F}_p$ of degree at most d. Then for $f : \mathbb{F}_p^n \to \mathbb{C}$ we define

$$\|f\|_{u^k} = \sup\{|\mathbb{E}_x f(x) e_p(\pi(x))| : \pi \in \mathrm{Poly}_{\leqslant k-1}\}. \tag{2.2}$$

An idea that is used several times in this thesis is that a bounded function can be decomposed into a structured part and an error, where the precise form of the structured part is determined by which norm one wants to use to control the error. Such results are often referred to as regularity lemmas. In the case of the Gowers U^k norm one has the famous arithmetic regularity lemma of Ben Green, developed fully in [33]. Roughly speaking this states that for a bounded function $f : [N] \to [0, 1]$ one can decompose

$$f = f_{\mathrm{str}} + f_{\mathrm{unf}} + f_{\mathrm{sml}}$$

where f_{str} is a nilsequence of degree $\leqslant k$, $\|f_{\mathrm{unf}}\|_{U^{k+1}}$ is small and $\|f_{\mathrm{sml}}\|_2$ is small.

We also note that the precise statement of the regularity lemma for $k = 2$ is much simpler than the general case, as it does not involve nilsequences (see [22]).

2.4 The Heath-Brown circle method

Let $F(\mathbf{x})$ be a quadratic form in n variables with integer coefficients, and consider the problem of counting the number of ways in which F represents 0. More precisely, consider the quantity

$$N_F(P) \coloneqq \#\big\{\mathbf{x} \in \mathbb{Z}^n : F(\mathbf{x}) = 0, \|\mathbf{x}\|_\infty \leqslant P\big\}, \tag{2.3}$$

which counts the number of solutions to $F(\mathbf{x}) = 0$ in a box with side lengths $2P$. The Hardy–Littlewood circle method gives an asymptotic for $N_F(P)$ when $n \geqslant 5$. Given a quadratic form F the *singular series* is defined by

$$\sigma(F) = \prod_p \sigma_p$$

and

$$\sigma_p = \lim_{\nu \to \infty} \frac{M(p^\nu)}{p^{\nu(n-1)}} \tag{2.4}$$

with $M(q)$ the number of solutions to $F(\mathbf{x}) \equiv 0 (\text{mod } q)$ in $(\mathbb{Z}/q\mathbb{Z})^n$. We will not give the precise definition of the *singular integral* $\sigma_\infty(F)$ here, but we mention that it can be shown to be positive provided $F(\mathbf{x}) = 0$ has the "expected" number of solutions over $\mathbb{R}$.

Theorem 2.4.1 (Hardy–Littlewood)**.** *Let $n \geqslant 5$ and let F be a diagonal quadratic form in n variables. Then*

$$N_F(P) \sim \sigma_\infty(F)\sigma(F)P^{n-2}$$

as $P \to \infty$.

To avoid special cases we will only be interested in non-singular forms F. That is, if M is the underlying matrix of F such that $F(\mathbf{x}) = \mathbf{x}^T M \mathbf{x}$, we require that $\det M \neq 0$.

The Hardy–Littlewood method does not give an asymptotic for the cases $n = 3$ and $n = 4$, but such an asymptotic was found in [37]. Specifically, let

$$N_{F,w}(P) = \sum_{F(\mathbf{x})=0} w(P^{-1}\mathbf{x}), \tag{2.5}$$

where $w : \mathbb{R}^n \to [0, \infty)$ is some sufficiently smooth weight function and we are summing over all $\mathbf{x} \in \mathbb{Z}^n$ such that $F(\mathbf{x}) = 0$. A typical choice of w would be to choose a smooth approximation of $1_{[-1,1]^n}$ with compact support. Heath-Brown then proves the following theorem.

Theorem 2.4.2 ([37, Theorems 6, 7, 8])**.** *Let F be a non-singular quadratic form in $n = 4$ variables and let $w : \mathbb{R}^n \to [0, \infty)$ be an infinitely differentiable function with compact support.*

If $\det M$ *is not a square then, for all $\epsilon > 0$, it holds that*

$$N_{F,w}(P) = \sigma_\infty(F, w)L(1, \chi)\sigma^*(F)P^2 + O_{F,w,\epsilon}(P^{3/2+\epsilon}),$$

where

$$\sigma_\infty(w) = \lim_{\epsilon\to 0}(2\epsilon)^{-1}\int_{|F(\mathbf{x})|\leqslant\epsilon} w(\mathbf{x})\,d\mathbf{x}, \tag{2.6}$$

$$\sigma^*(F) = \prod_p \big(1-\chi(p)p^{-1}\big)\sigma_p,$$

with $\chi(p) = \left(\frac{\det(M)}{p}\right)$ *and* σ_p *is as in* (2.4).

If $\det M$ *is a square then for all* $\epsilon > 0$ *it holds that*

$$N_{F,w}(P) = \sigma_\infty(F,w)\sigma^*_{sq}(F)P^2\log P + \sigma_1(F,w)P^2 + O_{F,w,\epsilon}(P^{3/2+\epsilon}), \tag{2.7}$$

where $\sigma_1(F,w)$ *is some constant depending only on* w *and* F, $\sigma_\infty(F,w)$ *is defined as in* (2.6) *and* $\sigma^*_{sq}(F)$ *is defined by*

$$\sigma^*_{sq}(F) = \prod_p \big(1-p^{-1}\big)\sigma_p. \tag{2.8}$$

If instead $n = 3$ *then*

$$N_{F,w}(P) = \frac{1}{2}\sigma_\infty(w)\sigma^*_{sq}(F)P\log P + \sigma_1(F,w)P + O_{F,w,\epsilon}(P^{5/6+\epsilon}),$$

where $\sigma_1(F,w)$ *is some constant depending only on* w *and* F, *and* $\sigma_\infty(w)$ *and* $\sigma^*_{sq}(F)$ *are defined as in* (2.6) *and* (2.8).

Heath-Brown proved the above results by using a δ-method of Duke, Friedlander and Iwaniec [21], which we will refer to as the Heath-Brown circle method. The starting point of this method is the following identity.

Theorem 2.4.3 ([21])**.** *For any integer* n *let*

$$\delta_n = \begin{cases} 1, & n = 0, \\ 0, & n \neq 0. \end{cases}$$

Then for any $Q > 1$ *there is a positive constant* c_Q *and an infinitely differentiable function* $h : (0,\infty)\times\mathbb{R}\to\mathbb{R}$, *such that*

$$\delta_n = c_Q Q^{-2}\sum_{q=1}^{\infty}\sideset{}{^*}\sum_{a(\mathrm{mod}\ q)} e_q(an)h(q/Q, n/Q^2).$$

The constant c_Q *satisfies* $c_Q = 1 + O_N(Q^{-N})$ *for any* $N > 0$.

Much more can be said about the function h above, but we will not state any further properties until they are needed later.

By combining Theorem 2.4.3 with (2.5), together with an application of Poisson summation, Heath-Brown shows that in fact the following holds.

Theorem 2.4.4 ([37, Theorem 2]). *Let F be a polynomial in four variables. For any $P > 1$ there is a positive constant c_P and an infinitely differentiable function $h(x, y)$ defined on the set $(0, \infty) \times \mathbb{R}$, such that*

$$\sum_{\mathbf{x} \in \mathbb{Z}^4 : F(\mathbf{x}) = 0} w(P^{-1}\mathbf{x}) = c_P P^{-2} \sum_{\mathbf{c} \in \mathbb{Z}^4} \sum_{q=1}^{\infty} q^{-4} S_q(\mathbf{c}) I_q(\mathbf{c}),$$

where

$$S_q(\mathbf{c}) = \sideset{}{^*}\sum_{a \pmod q} \sum_{\mathbf{b} \pmod q} e_q(aF(\mathbf{b}) + \mathbf{c} \cdot \mathbf{b}) \tag{2.9}$$

and

$$I_q(\mathbf{c}) = \int_{\mathbb{R}^4} w(P^{-1}\mathbf{x}) h\left(\frac{q}{P}, \frac{F(\mathbf{x})}{P^2}\right) e_q(-\mathbf{c} \cdot \mathbf{x}) d\mathbf{x}. \tag{2.10}$$

The constant c_P satisfies $c_P = 1 + O_N(P^{-N})$ for any $N > 0$.

Heath-Brown proves a sequence of results concerning $I_q(\mathbf{c})$ and $S_q(\mathbf{c})$ above, which we will state when needed.

Chapter 3

Partition regularity of generalised Fermat equations

3.1 Introduction

Let p be a prime, and consider $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$. We study partition regularity in $\mathbb{F}_p$ of the equation

$$x + y = z^2 \tag{3.1}$$

and of the more general equation

$$x^a + y^b = z^c \tag{3.2}$$

where $a, b, c \in \mathbb{N}$. We are interested in the following two questions, which were asked in [15].

Question 1. Given an r-colouring of $\mathbb{F}_p$, will there always be a non-trivial solution to (3.1) with x, y and z the same colour?

Question 2. Given an r-colouring of $\mathbb{F}_p$, will there always be a non-trivial solution to (3.2) with x, y and z the same colour?

Here any solution with $x = y = z$ is counted as trivial. It turns out that the answer to both questions is indeed positive, provided p is sufficiently large, which is the content of our two main theorems.

Theorem 3.1.1. *Suppose that $\mathbb{F}_p$ is r-coloured. Then there are at least $c_r p^2$ monochromatic triples (x, y, z) that satisfy (3.1), where $c_r > 0$ depends on r but not on p.*

Note that an immediate corollary of Theorem 3.1.1 is that provided $p > \sqrt{2}c_r^{-1/2}$ then there is a monochromatic non-trivial solution to (3.1), as this ensures more solutions than the trivial ones $x = y = z = 0$ and $x = y = z = 2$.

Theorem 3.1.2. *Suppose that $\mathbb{F}_p$ is r-coloured. Then there are at least $c_{r,a,b,c}p^2$ monochromatic triples (x, y, z) that satisfy (3.2), where $c_{r,a,b,c} > 0$ depends on r, a, b and c but not on p.*

As noted in [15] there is no general density result for (3.1). By this it is meant that given a set $A \subset \mathbb{F}_p$ with $|A| \geqslant \alpha p$ for some α which is independent of p, there is not necessarily a solution to (3.1) with $x, y, z \in A$. Indeed, the set

$$A = \left\{x \in \mathbb{F}_p : 0 \leqslant x < p/3, 2p/3 \leqslant x^2 < p\right\}$$

has size $|A| = \frac{1}{9}p + o(p)$, but clearly no solutions to (3.1). This can be proven using Fourier analysis on $\mathbb{F}_p$, which we leave as an exercise for the reader.

It is also worth noting that (3.1) is not partition regular over $\mathbb{Z}/q\mathbb{Z}$, where $q = p^n$ and p is a fixed odd prime greater than 3. Indeed, by modifying a counterexample given in [15] which shows that (3.1) is not partition regular over $\mathbb{N}$ by using 16 colours, one can obtain a counterexample over $\mathbb{Z}/p^n\mathbb{Z}$ where the number of colours needed depends on p but not n. Such a colouring is given in Appendix A. This indicates that the primality of p must play an important role in the proof of Theorem 3.1.1, and thus also in Theorem 3.1.2.

For Question 2, the special case of proving partition regularity of the Fermat equation

$$x^n + y^n = z^n$$

was done in [15]. Their proof uses Schur's lemma [59] on partition regularity of the equation $x+y = z$. This result strengthened a previous result of Dickson [20] proving existence of non-trivial solutions to this equation over $\mathbb{F}_p$.

The proofs of theorems 3.1.1 and 3.1.2 use the methods developed by Green and Sanders in [32], where they establish partition regularity for quadruples $(x, y, xy, x+y)$ over $\mathbb{F}_p$. In several instances the results we need are simply weaker versions of results proved in [32], and in these cases our proofs will closely resemble theirs. In some places we will cite the corresponding results in [32] so that the reader may compare proofs.

Even though the statement of Theorem 3.1.2 is more general than that of Theorem 3.1.1, which is a corollary of the former, the proofs use essentially the same ingredients. For this reason we will introduce all the main tools needed in the context of Theorem 3.1.1, as this avoids the additional clutter caused by the parameters a, b and c. The proof consists of three main ingredients: a *regularity lemma*, a *counting lemma* and a *Ramsey lemma*. Theorem 3.1.1 is proved using these in Section 3.4, and then the

proofs of the lemmas are given in Sections 3.5, 3.6 and 3.7 respectively. Finally, in Section 3.8 we mention the necessary modifications of the proof that are needed to establish the more general Theorem 3.1.2.

3.2 Counting solutions

For functions $f_1, f_2, f_3 : \mathbb{F}_p \to \mathbb{C}$, define

$$T(f_1, f_2, f_3) = \frac{1}{p^2} \sum_{\substack{x,y,z \in \mathbb{F}_p \\ x+y=z^2}} f_1(x) f_2(y) f_3(z). \tag{3.3}$$

Clearly $p^2 T(1_A, 1_A, 1_A)$ counts the number of solutions to (3.1) with $x, y, z \in A$, and so we would like to control this quantity. The main content of this section is to show that $|T(f_1, f_2, f_3)|$ can be controlled by some norm of the functions f_1, f_2, f_3.

We remind the reader of the definitions of the u^2 and u^3 norms given in (2.2).

Definition 3.2.1. Let $f : \mathbb{F}_p \to \mathbb{C}$. The u^2 norm of f is defined by

$$\|f\|_{u^2} = \sup_{\xi} |\widehat{f}(\xi)|.$$

Definition 3.2.2. Let $f : \mathbb{F}_p \to \mathbb{C}$. The u^3 norm of f is defined by

$$\|f\|_{u^3} = \sup \left\{ \left| \mathbb{E}_{x \in \mathbb{F}_p} f(x) e_p(ax^2 + bx) \right| : a, b \in \mathbb{F}_p \right\}.$$

Note that by the definition of $\widehat{f}$ it immediately follows that $\|f\|_{u^2} \leqslant \|f\|_{u^3}$.

We now turn to the main result of this section.

Proposition 3.2.3. *If* $\|f_1\|_2, \|f_2\|_2, \|f_3\|_2 \leqslant 1$ *then*

$$|T(f_1, f_2, f_3)| \leqslant \sqrt{2} \min_i \|f_i\|_{u^3}.$$

Proof. Define $g(w) = \sum_{z^2 = w} f_3(z)$, where $g(w)$ is understood to be zero if there is no z such that $z^2 = w$. Then by the properties of convolution and the inverse formula for the Fourier transform one has

$$\begin{aligned} T(f_1, f_2, f_3) &= \mathbb{E}_x \mathbb{E}_y f_1(x) f_2(y) g(x+y) \\ &= \sum_{\xi \in \mathbb{F}_p} \widehat{f_1}(-\xi) \widehat{f_2}(-\xi) \widehat{g}(\xi). \end{aligned} \tag{3.4}$$

By the Cauchy–Schwarz inequality and Parseval one gets

$$|T(f_1,f_2,f_3)| \leqslant \|f_1\|_{u^2}\Big(\sum_\xi |\widehat{f_2}(-\xi)|^2\Big)^{1/2}\Big(\sum_\xi |\widehat{g}(\xi)|^2\Big)^{1/2}$$
$$= \|f_1\|_{u^2}\|f_2\|_2\Big(\sum_\xi |\widehat{g}(\xi)|^2\Big)^{1/2}.$$

Furthermore,

$$\sum_\xi |\widehat{g}(\xi)|^2 = \mathbb{E}_x|g(x)|^2 = \mathbb{E}_x \sum_{z^2=x} f_3(z) \sum_{w^2=x} \overline{f_3(w)}$$
$$= \mathbb{E}_z|f_3(z)|^2 + \mathbb{E}_z f_3(z)\overline{f_3(-z)} - \frac{|f_3(0)|^2}{p},$$

where we use Parseval and the definition of g. The triangle inequality and Cauchy–Schwarz applied to the second sum above gives that

$$\Big(\sum_\xi |\widehat{g}(\xi)|^2\Big)^{1/2} \leqslant \sqrt{2}\|f_3\|_2,$$

and so

$$|T(f_1,f_2,f_3)| \leqslant \sqrt{2}\|f_1\|_{u^2}\|f_2\|_2\|f_3\|_2 \leqslant \sqrt{2}\|f_1\|_{u^2},$$

by the assumptions on $\|f_2\|_2$ and $\|f_3\|_2$. The exact same argument with the roles of f_1 and f_2 interchanged gives $|T(f_1,f_2,f_3)| \leqslant \sqrt{2}\|f_2\|_{u^2}$.

To get the bound in terms of $\|f_3\|_{u^3}$, note that from (3.4), Cauchy–Schwarz and Parseval one also gets

$$|T(f_1,f_2,f_3)| \leqslant \|g\|_{u^2}\Big(\sum_\xi |f_1|^2\Big)^{1/2}\Big(\sum_\xi |f_2|^2\Big)^{1/2} = \|g\|_{u^2}\|f_1\|_2\|f_2\|_2,$$

which is less than $\|g\|_{u^2}$ by the assumptions on $\|f_1\|_2, \|f_2\|_2$. Furthermore,

$$\|g\|_{u^2} = \sup_\xi \Big|\mathbb{E}_x g(x)e_p(-x\xi)\Big| = \sup_\xi \Big|\mathbb{E}_z f_3(z)e_p(-z^2\xi)\Big| \leqslant \|f_3\|_{u^3},$$

which completes the proof. □

In addition to Proposition 3.2.3 we record the following rather trivial bound.

Lemma 3.2.4. *Let $f_1, f_2, f_3 : \mathbb{F}_p \to \mathbb{C}$. Then*

$$|T(f_1,f_2,f_3)| \leqslant \|f_1\|_2\|f_2\|_2\|f_3\|_2. \tag{3.5}$$

Proof. By the triangle inequality and Cauchy–Schwarz we have

$$\begin{aligned}|T(f_1,f_2,f_3)| &\leqslant \mathbb{E}_{x,z}|f_1(x)f_2(z^2-x)f_3(z)| \\ &\leqslant \mathbb{E}_z|f_3(z)|(\mathbb{E}_x|f_1(x)|^2)^{1/2}(\mathbb{E}_x|f_2(z^2-x)|^2)^{1/2} \\ &= \|f_3\|_1\|f_1\|_2\|f_2\|_2 \leqslant \|f_1\|_2\|f_2\|_2\|f_3\|_2,\end{aligned}$$

where we use that as x runs through $\mathbb{F}_p$, so does z^2-x. □

Proposition 3.2.3 and Lemma 3.2.4 will be used in the proof of Theorem 3.1.1 at the end of Section 3.4. It is the u^3-norm in Proposition 3.2.3 that decides the correct form of the Regularity lemma, and thus also the other key lemmas in Section 3.4.

3.3 Quadratic systems and trigonometric polynomials

Following the approach in [32], we make the following definitions. To begin with, write $\mathbb{G} = (\mathbb{R}/\mathbb{Z}) \times (\mathbb{R}/\mathbb{Z})$ and $\mathbb{G}^d = (\mathbb{R}/\mathbb{Z})^d \times (\mathbb{R}/\mathbb{Z})^d$. The group operation in $\mathbb{G}$ will be denoted by $+$.

Definition 3.3.1. A quadratic system of dimension d is a map $\Psi : \mathbb{F}_p \to \mathbb{G}^d$ of the form

$$\Psi(x) = (a_i x^2/p, a_i x/p)_{i=1}^d,$$

where $(a_i)_{i=1}^d \subset \mathbb{F}_p^d$.

If Ψ is a quadratic system we write

$$\Lambda_\Psi := \{\xi \in \mathbb{Z}^d : \xi_1 a_1 + \cdots + \xi_d a_d \equiv 0 (\mathrm{mod}\ p)\}, \tag{3.6}$$

where (a_i) are the coefficients appearing in the definition of Ψ. In other words, Λ_Ψ is a lattice encoding the linear relations between the a_i modulo p. Note that Λ_Ψ is a lattice of full rank, as $p\mathbb{Z}^d \subset \Lambda_\Psi$. Further, define the closed subgroup

$$G_\Psi := \{g \in (\mathbb{R}/\mathbb{Z})^d : \xi \cdot g = 0 \text{ for all } \xi \in \Lambda_\Psi\} \tag{3.7}$$

of $(\mathbb{R}/\mathbb{Z})^d$. We also write $H_\Psi = G_\Psi \times G_\Psi$. Note that we now can think of Ψ as a map from $\mathbb{F}_p \to H_\Psi$.

Lemma 3.3.2. *Let Λ_Ψ and G_Ψ be defined as in (3.6) and (3.7) respectively, and suppose that $\lambda \in \mathbb{Z}^d$ satisfies $\lambda \cdot g = 0$ for all $g \in G_\Psi$. Then $\lambda \in \Lambda_\Psi$.*

Proof. Note that $\xi_1 a_1 + \cdots + \xi_d a_d \equiv 0 (\mathrm{mod}\ p)$ is the same as saying $\xi \cdot a/p = 0$ in $\mathbb{R}/\mathbb{Z}$. It is thus clear that $a/p \in G_\Psi$, and so by assumption $\lambda \cdot a/p = 0$. This is again equivalent to saying $\lambda \cdot a \equiv 0 (\mathrm{mod}\ p)$, and so we have $\lambda \in \Lambda_\Psi$. □

The above allows us to prove a useful orthogonality relation.

Lemma 3.3.3. *It holds that*

$$\int e(\xi \cdot t)\, d\mu_{G_\Psi}(t) = 1_{\Lambda_\Psi}(\xi). \tag{3.8}$$

Proof. If $\xi \in \Lambda_\Psi$ it is clear that the above integral is 1. Suppose instead that the integral is nonzero. Take $g \in G_\Psi$ and make the substitution $t \mapsto t+g$, which preserves Haar measure, to get

$$\int e(\xi \cdot t)\, d\mu_{G_\Psi}(t) = e(\xi \cdot g) \int e(\xi \cdot t)\, d\mu_{G_\Psi}(t).$$

This gives $\xi \cdot g = 0$ in $\mathbb{R}/\mathbb{Z}$, and as this holds for any $g \in G_\Psi$ it follows from Lemma 3.3.2 that $\xi \in \Lambda_\Psi$. □

In addition to the notion of quadratic systems we will need the notion of trigonometric polynomials, and the trig-norm, as defined next.

Definition 3.3.4. Let $F : \mathbb{G}^d \to \mathbb{C}$ be a function. We write $\|F\|_{\mathrm{trig}}$ for the smallest M such that F has a Fourier expansion

$$F(\theta_1, \theta_2) = \sum_{\|\xi_1\|_1, \|\xi_2\|_1 \leqslant M} \widehat{F}(\xi_1, \xi_2) e(\xi_1 \cdot \theta_1 + \xi_2 \cdot \theta_2) \tag{3.9}$$

and $\sum_{\xi_1,\xi_2} |\widehat{F}(\xi_1, \xi_2)| \leqslant M$. If $\|F\|_{\mathrm{trig}} < \infty$ we say that F is a *trigonometric polynomial*.

The Fourier coefficients appearing in the above definition are given by

$$\widehat{F}(\xi_1, \xi_2) = \int F(\theta_1, \theta_2) e(-\xi_1 \cdot \theta_1 - \xi_2 \cdot \theta_2)\, d\theta_1 d\theta_2. \tag{3.10}$$

We also note that the trigonometric polynomials are dense in $C(\mathbb{G}^d)$, the space of continuous, complex valued functions on $\mathbb{G}^d$.

The following proposition can be thought of as the most basic form of the counting lemma presented in Section 3.6, and will be used to prove Lemma 3.3.7 on approximating points in H_Ψ by points in the image of Ψ.

Lemma 3.3.5. *Let Ψ be a quadratic system of dimension d, and let $F : \mathbb{G}^d \to \mathbb{C}$ be a trigonometric polynomial. Then*

$$\mathbb{E}_x F \circ \Psi(x) = \int F \, d\mu_{H_\Psi} + O(\|F\|_{\mathrm{trig}} p^{-1/2}).$$

We give the proof of Lemma 3.3.5 in Section 3.6, as the proof is just a much easier version of the proof of the counting lemma.

At this point we need to establish a notion of absolute value of points in $\mathbb{G}^d$. For a point $(\theta, \phi) \in \mathbb{G}$ we write $|(\theta, \phi)| := \max\{\|\theta\|_{\mathbb{R}/\mathbb{Z}}, \|\phi\|_{\mathbb{R}/\mathbb{Z}}\}$, where $\|x\|_{\mathbb{R}/\mathbb{Z}}$ is the distance from x to the nearest integer. We then let $|\eta| = \max_{i=1}^d \{|\eta_i|\}$ for $\eta \in \mathbb{G}^d$.

Before proving our final result of the section we will need the following result from [62].

Lemma 3.3.6. *For $\epsilon \geqslant 0$ it holds that*

$$\mu_{H_\Psi}\big(\{x \in H_\Psi : |x| \leqslant \epsilon\}\big) \geqslant \epsilon^{2d}.$$

Proof. This follows directly from [62, Lemma 4.20]. Indeed, in the notation of the lemma we take the ambient group Z to be H_Ψ and set $S = \{\xi_1, \dots, \xi_d, \zeta_1, \dots, \zeta_d\}$ where $\xi_i \cdot (\theta, \phi) = \theta_i$ and $\zeta_i \cdot (\theta, \phi) = \phi_i$. □

The next lemma tells us that any point in H_Ψ can be well approximated by $\Psi(x)$ for some value of x, provided p is large, and so in some sense we are saying that H_Ψ is well approximated by the image of Ψ. This result will be needed in the proof of Proposition 3.4.4.

Lemma 3.3.7 (Cf. [32, Corollary 3.4])**.** *There is a function $p_1 : \mathbb{Z}_{\geqslant 0} \times (0, 1] \to \mathbb{R}_{\geqslant 0}$ such that the following holds. For any d-dimensional quadratic system Ψ, $h \in H_\Psi$ and $\epsilon \in (0, 1]$ we have*

$$\#\{x \in \mathbb{F}_p : |\Psi(x) - h| \leqslant \epsilon\} \geqslant \frac{1}{8}\left(\frac{\epsilon}{2}\right)^{2d} p$$

provided $p \geqslant p_1(d, \epsilon)$.

Proof. Fix $h \in H_\Psi$. We can find a trigonometric polynomial F satisfying $-\delta \leqslant F \leqslant 2$ on $\mathbb{G}^d$, $F(\theta) \leqslant 0$ for $|\theta - h| > \epsilon$ and $F(\theta) \geqslant 1$ for $|\theta - h| \leqslant \epsilon/2$. Furthermore $\|F\|_{\mathrm{trig}}$ is bounded in terms of ϵ, δ and d, uniformly in the choice of h. Applying Lemma 3.3.6 and using translation invariance of Haar measure we have

$$\int F \, d\mu_{H_\Psi} = \int F(\theta - h) \, d\mu_{H_\Psi}(\theta) \geqslant \left(\frac{\epsilon}{2}\right)^{2d} - \delta.$$

We set $\delta = \frac{1}{2}\left(\frac{\epsilon}{2}\right)^{2d}$ and note that $\mathbb{E}_x F \circ \Psi(x) \leqslant 2\mu_{\mathbb{F}_p}(\{x : |\Psi(x) - h| \leqslant \epsilon\})$, as $F \leqslant 2$ everywhere. By invoking Lemma 3.3.5 and choosing p larger than some function $p_1(d, \epsilon)$ we then have

$$\mu_{\mathbb{F}_p}(\{x : |\Psi(x) - h| \leqslant \epsilon\}) \geqslant \frac{1}{2}\left(\left(\frac{\epsilon}{2}\right)^{2d} - \frac{1}{2}\left(\frac{\epsilon}{2}\right)^{2d}\right) - \frac{1}{8}\left(\frac{\epsilon}{2}\right)^{2d} = \frac{1}{8}\left(\frac{\epsilon}{2}\right)^{2d}.$$

□

3.4 Proof of main theorem for the case $x + y = z^2$

In order to prove Theorem 3.1.1 we will make use of the following three key lemmas.

Lemma 3.4.1 (Regularity lemma)**.** *There are functions $p_3, M : \mathbb{Z}_{\geqslant 0} \times (0, 1] \to \mathbb{R}_{\geqslant 0}$ such that the following holds. Let $c : \mathbb{F}_p \to [r]$ be an r-colouring of $\mathbb{F}_p$ and let $\epsilon > 0$. Then there is a quadratic system Ψ of dimension d, functions $F_1, \dots, F_r : \mathbb{G}^d \to \mathbb{R}_{\geqslant 0}$, and functions $g_1, \dots, g_r : \mathbb{F}_p \to [-1, 1]$, such that the following holds provided $p \geqslant p_3(r, \epsilon)$.*

(i) $\|F_i\|_{\text{trig}} \leqslant M(r, \epsilon)$ *for all* i*;*

(ii) $d \leqslant 8r\epsilon^{-2} + 1$*;*

(iii) $\|F_i \circ \Psi - g_i\|_2 \leqslant \epsilon$ *for all* i*;*

(iv) $\|1_{c^{-1}(i)} - g_i\|_{u^3} \leqslant \epsilon$ *for all* i*; and*

(v) $\sum_i F_i \circ \Psi \geqslant 1$ *pointwise.*

Lemma 3.4.2 (Counting lemma)**.** *Let Ψ be a d-dimensional quadratic system and $F : \mathbb{G}^d \to \mathbb{C}$ be a trigonometric polynomial. Then*

$$T(F \circ \Psi, F \circ \Psi, F \circ \Psi) = \int F(t, u)F(t', u')F(u + u', u'')\, d\mu_{G_\Psi}^{\otimes 5}(u, u', u'', t, t') + O\left(p^{-1/2}M^3\right), \quad (3.11)$$

where $M = \|F\|_{\text{trig}}$.

Lemma 3.4.3 (Ramsey lemma)**.** *There is a positive function $\rho : \mathbb{N} \to (0,1]$ with the following property. Suppose that G is a compact Abelian group with Haar probability measure μ. Assume also that $F_1, \dots, F_r : G \times G \to \mathbb{R}_{\geqslant 0}$ are continuous functions that satisfy $\sum_{i=1}^r F_i(g_1, g_2) \geqslant 1$ for all $(g_1, g_2) \in G \times G$. Then there is an $i \in [r]$ such that*

$$\int F_i(t,u)F_i(t',u')F_i(u+u',u'')\, d\mu^{\otimes 5}(u,u',u'',t,t') \geqslant \rho(r). \tag{3.12}$$

The Regularity lemma allows us to replace the characteristic functions of the colour classes by structured functions of the form $F \circ \Psi$, up to a small error. For these special functions the counting lemma tells us that counting configurations of the form $x+y=z^2$ in $\mathbb{F}_p$ is the same as counting some linear configuration in H_Ψ. Finally, the Ramsey lemma establishes that there are in fact sufficiently many monochromatic linear configurations in H_Ψ.

Lemmas 3.4.2 and 3.4.3 will be used together, so we therefore state the combination as a separate proposition.

Proposition 3.4.4 (Cf. [32, Proposition 4.5])**.** *There is a function $p_2 : \mathbb{R}_{\geqslant 0} \times \mathbb{Z}_{\geqslant 0} \times \mathbb{Z}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$ such that the following holds. Let Ψ be a quadratic system of dimension d. Suppose that $F_1, \dots, F_r : \mathbb{G}^d \to \mathbb{R}_{\geqslant 0}$ satisfy $\|F_i\|_{\mathrm{trig}} \leqslant M$ and $\sum_i F_i \circ \Psi \geqslant 1$ pointwise on $\mathbb{F}_p$. Then there is some i such that*

$$T(F_i \circ \Psi, F_i \circ \Psi, F_i \circ \Psi) \geqslant 2^{-4}\rho(r), \tag{3.13}$$

with ρ as given by Lemma 3.4.3, provided $p \geqslant p_2(M,r,d)$.

Proof. Apply Lemma 3.4.2 to each F_i composed with Ψ to get that

$$T(F_i \circ \Psi, F_i \circ \Psi, F_i \circ \Psi) \geqslant \int F_i(t,u)F_i(t',u')F_i(u+u',u'')\, d\mu_{G_\Psi}^{\otimes 5}(t,t',u,u',u'') - \frac{1}{16}\rho(r),$$

provided p is large enough, say $p \geqslant p_2'(M,r)$.

Let $h \in H_\Psi$. Note that $|F_i(w) - F_i(v)| \leqslant 4\pi d M^2 |w-v|$ for all $v, w \in \mathbb{G}^d$, and so

$$\Big| \sum_{i=1}^r F_i(h) - \sum_{i=1}^r F_i(\Psi(z)) \Big| \leqslant 4\pi d r M^2 |\Psi(z) - h|$$

for any $z \in \mathbb{F}_p$. By Lemma 3.3.7 we can find a z such that $|\Psi(z) - h| \leqslant \frac{1}{8\pi d r M^2}$ provided $p \geqslant p_1(d, \frac{1}{8\pi d r M^2})$. For this value of z we then have

$$\sum_{i=1}^r F_i(h) \geqslant \sum_{i=1}^r F_i \circ \Psi(z) - \frac{1}{2} \geqslant \frac{1}{2}. \tag{3.14}$$

Applying Lemma 3.4.3 to the functions $(2F_i)_{i=1}^r$ then gives that there is some i such that

$$T(F_i \circ \Psi, F_i \circ \Psi, F_i \circ \Psi) \geqslant 2^{-3}\rho(r) - 2^{-4}\rho(r) = 2^{-4}\rho(r), \tag{3.15}$$

provided $p \geqslant p_2(M, r, d)$, where we define $p_2(M, r, d)$ to be the largest of $p_2'(M, r)$ and $p_1(d, \frac{1}{8\pi d r M^2})$. □

We are now ready to prove the main theorem. Note that in contrast to [32, proof of Proposition 4.1 p. 22] we will not need to induct on the number of colours.

Proof of Theorem 3.1.1. Given the number of colours r, fix $\epsilon = 2^{-7}3^{-1}\rho(r)$ and

$$p_0(r) = \max\Big\{p_3(r, \epsilon), \sup_{d \leqslant 4r\epsilon^{-2}} p_2\big(M(r, \epsilon), r, d\big)\Big\}, \tag{3.16}$$

where ρ and p_2 are the functions appearing in Proposition 3.4.4 and p_3 and M are the functions appearing in Lemma 3.4.1. Note that the above supremum over d runs over positive integers, and so it is finite. Assume first that $p \geqslant p_0$. Apply Lemma 3.4.1 with the above ϵ and our colouring c to find F_i, g_i, Ψ and d that satisfy the properties listed in the Lemma, provided $p \geqslant p_3(r, \epsilon)$, which is satisfied by the choice of p_0. By Proposition 3.4.4 there is some $i \in [r]$ such that

$$T(F_i \circ \Psi, F_i \circ \Psi, F_i \circ \Psi) \geqslant 2^{-4}\rho(r)$$

provided $p \geqslant p_2(M(r, \epsilon), r, d)$. This holds by the choice of p_0 and property (ii) from Lemma 3.4.1. Note also that $\|g_i\|_2 \leqslant 1$ as $g_i : \mathbb{F}_p \to [-1, 1]$, $\|F_i \circ \Psi - g_i\|_2 \leqslant \epsilon \leqslant 1$ and so $\|F_i \circ \Psi\|_2 \leqslant 2$. Lemma 3.2.4 then gives

$$\begin{aligned}|T(g_i, g_i, g_i) - T(F_i \circ \Psi, F_i \circ \Psi, F_i \circ \Psi)| \leqslant \\ |T(g_i - F_i \circ \Psi, g_i, g_i)| + |T(F_i \circ \Psi, g_i - F_i \circ \Psi, g_i)| + |T(F_i \circ \Psi, F_i \circ \Psi, g_i - F_i \circ \Psi)| \\ \leqslant 7\epsilon,\end{aligned}$$

as $\|F_i \circ \Psi - g_i\|_2 \leqslant \epsilon$, so that

$$|T(g_i, g_i, g_i)| \geqslant 2^{-4}\rho(r) - 7\epsilon. \tag{3.17}$$

Let

$$A_i = c^{-1}(i), \quad i = 1, \dots, r.$$

By Proposition 3.2.3 we get that

$$\begin{aligned}|T(g_i, g_i, g_i) - T(1_{A_i}, 1_{A_i}, 1_{A_i})| \leqslant \\ |T(g_i - 1_{A_i}, g_i, g_i)| + |T(1_{A_i}, g_i - 1_{A_i}, g_i)| + |T(1_{A_i}, 1_{A_i}, g_i - 1_{A_i})| \\ \leqslant 3\sqrt{2}\|g_i - 1_{A_i}\|_{u^3} \leqslant 3\sqrt{2}\epsilon,\end{aligned}$$

where the final inequality follows by property (iv) of Lemma 3.4.1. Now this combined with (3.17) gives

$$T(1_{A_i}, 1_{A_i}, 1_{A_i}) \geqslant 2^{-4}\rho(r) - 7\epsilon - 3\sqrt{2}\epsilon > 2^{-5}\rho(r)$$

by our choice of ϵ.

If instead $p \leqslant p_0$ we have the two trivial solutions $x = y = z = 0$ and $x = y = z = 2$, giving at least $\frac{2}{p_0(r)^2}p^2$ monochromatic solutions to (3.1). Finally we set

$$c_r = \min\left\{\frac{2}{p_0(r)^2}, 2^{-5}\rho(r)\right\},$$

which finishes the proof.

□

3.5 Regularity lemma

The conclusion of Lemma 3.4.1 is implied by the regularity lemma given in [32]. The proof given here is therefore also just the necessary parts of the proof given by Green–Sanders in their paper. Nevertheless we believe it is instructive to see the full proof written out in this simpler setting, as many of the unpleasant technicalities of [32] go away in this case.

We begin by defining intervals in $\mathbb{G}^d$. Let $I(x, R) = \big[(2x-1)/2R, (2x+1)/2R\big)$ be an interval in $\mathbb{R}/\mathbb{Z}$ and define

$$I_{R;t,u} = \big\{(\theta, \phi) \subset \mathbb{G}^d : \theta_j \in I(t_j, R), \phi_j \in I(u_j, R) \text{ for } j = 1, \dots, d\big\}.$$

Next, given a quadratic system Ψ, consider the σ-algebra generated by $\Psi^{-1}(I_{R;t,u})$ for $t, u \in \{0, \dots, R-1\}^d$, and let Π_R^Ψ be the projection operator onto this σ-algebra. Explicitly, for any $f : \mathbb{F}_p \to \mathbb{C}$, we have that

$$\Pi_R^\Psi f(x) = \frac{1}{|A(x)|}\sum_{x' \in A(x)} f(x'),$$

where $A(x) = \Psi^{-1}(I_{R;t,u})$ with t, u such that $\Psi(x) \in I_{R;t,u}$.

Lemma 3.5.1. *Suppose that $\|f\|_\infty \leqslant 1$, $\|f\|_{u^3} \geqslant \delta$ and $R > 8\pi\delta^{-1}$. Then there is a function $g \in L^\infty(\mathbb{F}_p)$ with $\|g\|_\infty \leqslant 1$ and a quadratic system Φ of dimension 2, such that $|\langle f, \Pi_R^\Phi g\rangle| \geqslant \frac{1}{2}\delta$.*

Proof. By the definition of the u^3-norm there are a_1, a_2 such that

$$\left|\mathbb{E}_x f(x)\overline{e_p(a_1x^2 + a_2x)}\right| \geqslant \delta. \tag{3.18}$$

Let $\Phi(x) = \left(\frac{a_i x^2}{p}, \frac{a_i x}{p}\right)_{i=1,2}$ and $F(\theta_1, \phi_1, \theta_2, \phi_2) = e(\theta_1 + \phi_2)$, and set $g = F \circ \Phi$. Now (3.18) states precisely that $|\langle f, g\rangle| \geqslant \delta$, and $\|g\|_\infty \leqslant 1$ follows from $\|F\|_\infty \leqslant 1$. Furthermore,

$$\|\Pi_R^\Phi g - g\|_\infty \leqslant \sup_x \sup_{x' \in A(x)} |F(\Phi(x')) - F(\Phi(x))| \leqslant 4\pi R^{-1},$$

as $|F(x) - F(y)| \leqslant 4\pi|x - y|$. The assumption on R gives that this is less than $\frac{1}{2}\delta$. This together with $\|f\|_\infty \leqslant 1$ gives

$$|\langle f, \Pi_R^\Phi g\rangle| \geqslant |\langle f, g\rangle| - |\langle f, g - \Pi_R^\Phi g\rangle| \geqslant \frac{1}{2}\delta,$$

as required. □

By repeated application of the above lemma and an energy increment argument we are able to establish the following Koopman von Neumann type Lemma.

Lemma 3.5.2. *Suppose that $f_1, \dots, f_r : \mathbb{F}_p \to \mathbb{C}$ are such that $\|f_i\|_\infty \leqslant 1$ for all $i \in \{1, \dots, r\}$ and that $R > 16\pi\delta^{-1}$. Then there is a quadratic system Ψ with $\dim \Psi \leqslant 8r\delta^{-2} + 1$ such that $\|f_i - \Pi_R^\Psi f_i\|_{u^3} \leqslant \delta$ for all $i \in \{1, \dots, r\}$.*

Proof. We construct Ψ in an iterative manner. To begin with, pick any 1-dimensional quadratic system and let this be Ψ_0. At stage j, define $f_{i,j} = f_i - \Pi_R^{\Psi_j} f_i$. If $\|f_{i,j}\|_{u^3} \leqslant \delta$ for all i we are done and take $\Psi = \Psi_j$ of dimension $2j + 1$. If not, there is some i such that $\|f_{i,j}\|_{u^3} > \delta$, in which case we can apply Lemma 3.5.1 to $\frac{1}{2}f_{i,j}$. The factor $\frac{1}{2}$ is added as $\|f_{i,j}\|_\infty \leqslant 2\|f_i\|_\infty \leqslant 2$. This gives a quadratic system Φ of dimension 2 and a function g with $\|g\|_\infty \leqslant 1$, such that

$$|\langle f_{i,j}, \Pi_R^\Phi g\rangle| \geqslant \frac{1}{2}\delta.$$

Assume that Ψ_j is defined by the coefficients $(a_i)_{i=1}^d$ and that Φ is defined by the coefficients $(b_i)_{i=1}^2$. We then define Ψ_{j+1} as the quadratic system of dimension $d + 2$ with coefficients $(a_1, \dots, a_d, b_1, b_2)$. Noting that Π_R^Φ is idempotent and self adjoint, and that $\Pi_R^\Phi \Pi_R^{\Psi_{j+1}} = \Pi_R^\Phi$ we get that

$$\langle f_{i,j+1}, \Pi_R^\Phi g\rangle = \langle \Pi_R^\Phi f_{i,j+1}, \Pi_R^\Phi g\rangle = \langle \Pi_R^\Phi f_i - \Pi_R^\Phi \Pi_R^{\Psi_{j+1}} f_i, \Pi_R^\Phi g\rangle = 0.$$

This in turn gives

$$|\langle \Pi_R^{\Psi_{j+1}} f_i - \Pi_R^{\Psi_j} f_i, \Pi_R^{\Phi} g\rangle| = |\langle f_{i,j}, \Pi_R^{\Phi} g\rangle - \langle f_{i,j+1}, \Pi_R^{\Phi} g\rangle| \geqslant \frac{1}{2}\delta,$$

and by the Cauchy–Schwarz inequality and the fact that $\|g\|_2 \leqslant \|g\|_\infty \leqslant 1$ we get that

$$\|\Pi_R^{\Psi_{j+1}} f_i - \Pi_R^{\Psi_j} f_i\|_2 \geqslant \frac{1}{2}\delta.$$

Using the fact that $\Pi_R^{\Psi_j}$ is idempotent and self-adjoint, and that $\Pi_R^{\Psi_j}\Pi_R^{\Psi_{j+1}} = \Pi_R^{\Psi_j}$, we get $\langle \Pi_R^{\Psi_j} f_i, \Pi_R^{\Psi_{j+1}} f_i\rangle = \|\Pi_R^{\Psi_j} f_i\|_2^2$, which then gives

$$\|\Pi_R^{\Psi_{j+1}} f_i\|_2^2 - \|\Pi_R^{\Psi_j} f_i\|_2^2 = \|\Pi_R^{\Psi_{j+1}} f_i - \Pi_R^{\Psi_j} f_i\|_2^2 \geqslant \frac{1}{4}\delta^2.$$

This is what is needed to complete an energy increment argument. Indeed, defining

$$E_j = \sum_{i=1}^{r} \|\Pi_R^{\Psi_j} f_i\|_2^2,$$

we get that

$$E_{j+1} - E_j \geqslant \frac{1}{4}\delta^2.$$

The trivial bound $E_j \leqslant r$ gives that we can continue the iteration process at most $4r\delta^{-2}$ times, and we are done. □

Note that by definition of Π_R^{Ψ} we can write $\Pi_R^{\Psi} f = F_0 \circ \Psi$ for some F_0. The only thing that remains to settle before embarking on the proof of Lemma 3.4.1 is a way of controlling the trig norm of F_0, which is what the following lemma does.

Lemma 3.5.3 (Cf. [32, Lemma 5.3])**.** *There are functions $M_0, p_4 : (0,1] \times \mathbb{N} \times \mathbb{N} \to \mathbb{R}_{\geqslant 0}$ such that the following holds. Fix $R \in \mathbb{N}$, a quadratic system Ψ of dimension $d \in \mathbb{N}$ and $\epsilon \in (0,1]$. Then for all functions $f : \mathbb{F}_p \to [0,1]$, provided $p \geqslant p_4(\epsilon, d, R)$, there is a function $F : \mathbb{G}^d \to \mathbb{R}_{\geqslant 0}$ such that*

(i) $F \circ \Psi \geqslant \Pi_R^{\Psi} f$ *pointwise;*

(ii) $\|F\|_{\mathrm{trig}} \leqslant M_0(\epsilon, d, R)$*;*

(iii) $\|F \circ \Psi - \Pi_R^{\Psi} f\|_2 \leqslant \epsilon$.

Proof. We start by defining $A_{R;t,u} = \Psi^{-1}(I_{R;t,u})$, and note that $\Pi_R^\Psi f$ is constant on each of $A_{R;t,u}$. Denote this constant value by $c_{R;t,u}$, and note that then $0 \leqslant c_{R;t,u} \leqslant 1$ by assumption. We will construct F by approximating $1_{I_{R;t,u}}$ for each $t, u \in \{0, \dots, R-1\}$ and then adding these with weights $c_{R;t,u}$.

Let $\eta > 0$ and define

$$I^{\pm}_{R;t,u} = \Big\{(\theta, \phi) \in \mathbb{G}^d : \Big\|\theta_j - \frac{t_j}{R}\Big\|_{\mathbb{R}/\mathbb{Z}} < \frac{1}{2R} \pm \eta, \Big\|\phi_j - \frac{u_j}{R}\Big\|_{\mathbb{R}/\mathbb{Z}} < \frac{1}{2R} \pm \eta\Big\}.$$

As the trigonometric polynomials are dense in $C(\mathbb{G}^d)$, given $\delta > 0$ which we will specify later, we can find $F_{R;t,u} : \mathbb{G}^d \to \mathbb{R}$ which satisfy the following properties:

1. $0 \leqslant F_{R;t,u} \leqslant 1 + \delta$;

2. $F_{R;t,u}(z) \geqslant 1$ for all $z \in I_{R;t,u}$;

3. $F_{R;t,u}(z) \leqslant \delta$ for all $z \notin I^{+}_{R;t,u}$;

4. $\|F_{R;t,u}\|_{\text{trig}} \leqslant M_1(\delta, \eta, d, R)$

for some function $M_1 : (0,1) \times (0,1) \times \mathbb{N} \times \mathbb{N} \to \mathbb{R}_{\geqslant 0}$. With this in hand we set

$$F = \sum_{t,u \in \{0,\dots,R-1\}^d} c_{R;t,u} F_{R;t,u},$$

and note that $\Pi_R^\Psi f = F_0 \circ \Psi$, where

$$F_0 = \sum_{t,u \in \{0,\dots,R-1\}^d} c_{R;t,u} 1_{I_{R;t,u}}.$$

It remains to verify that F indeed satisfies each of the properties stated in the lemma, with appropriate choices of $\eta = \eta(\epsilon, d, R)$ and $\delta = \delta(\epsilon, d, R)$.

(i) If $z \in A_{R;t,u}$ then $\Pi_R^\Psi f(z) = c_{R;t,u}$, and by property (1) and (2) above we get that $F(z) \geqslant c_{R;t,u} F_{R;t,u}(z) \geqslant c_{R;t,u}$, and so (i) holds.

(ii) Set $M_0(\epsilon, d, R) = M_1(\delta(\epsilon, d, R), \eta(\epsilon, d, R), d, R)$ and then the result follows by property (4) above.

(iii) Begin by noting that $I^-_{R;t,u} \cap I^+_{R;t',u'} = \emptyset$ unless $(t,u) = (t',u')$. Define

$$E = \bigcup_{t,u} I^+_{R;t,u} \setminus I^-_{R;t,u}$$

and let z be such that $z \notin E$. As $\bigcup_{t,u} I^+_{R;t,u}$ covers $\mathbb{G}^d$ we must then have that $z \in I^-_{R;t,u}$ for some t, u. Then

$$|F(z) - F_0(z)| \leqslant \sum_{(t',u') \neq (t,u)} c_{R;t',u'} F_{R;t',u'}(z) + c_{R;t,u}\delta \leqslant R^{2d}\delta \tag{3.19}$$

by properties (1), (2) and (3) above.

Next, we will need to bound the size of $\Psi^{-1}(E)$. If J is an interval in $\mathbb{R}/\mathbb{Z}$ not containing 0 we have that

$$\#\{x \in \mathbb{F}_p : ax/p \in J\} \leqslant p|J| + C, \tag{3.20}$$

for any $a \in \mathbb{F}_p$ and an absolute constant C. This is because ax takes on each nonzero value in $\mathbb{F}_p$ at most once (exactly once if $a \neq 0$). We now note that

$$E \subset \bigcup_{u_1 \in \{0,\dots,R-1\}} \left\{(\theta,\phi) \in \mathbb{G}^d : \left\|\phi_1 - \frac{2u_1+1}{2R}\right\|_{\mathbb{R}/\mathbb{Z}} < \eta\right\}$$

and by (3.20) we then get $|\Psi^{-1}(E)| \leqslant 2\eta p + C < 3\eta p$, provided $p \geqslant C\eta^{-1}$. To ensure this, define $p_4(\epsilon, d, R) = C\eta^{-1}$. Using property (1) and $|f(x)| \leqslant 1$ then gives

$$\sum_{x \in \Psi^{-1}(E)} \left|F \circ \Psi(x) - F_0 \circ \Psi(x)\right|^2 < (R^{2d}(1+\delta)+1)^2 3\eta p. \tag{3.21}$$

Combining (3.19) and (3.21) now gives

$$\begin{aligned} \|F \circ \Psi - \Pi^\Psi_R f\|_2^2 = \frac{1}{p}\Big(\sum_{x \notin \Psi^{-1}(E)} + \sum_{x \in \Psi^{-1}(E)}\Big)(F \circ \Psi(x) - F_0 \circ \Psi(x))^2 \\ < R^{4d}\delta^2 + 27R^{4d}\eta = \epsilon^2, \end{aligned} \tag{3.22}$$

where we have defined

$$\eta = \frac{\epsilon^2}{28R^{4d}}, \qquad \delta = \frac{\epsilon}{\sqrt{28}R^{2d}}.$$

□

At this point we can complete the proof of Lemma 3.4.1.

Proof of Lemma 3.4.1. Assume that a colouring c and $\epsilon > 0$ is given. Write $f_i = 1_{c^{-1}(i)}$ for $i = 1, \dots, r$ and apply Lemma 3.5.2 to the f_i with $\delta = \epsilon$ and $R = \lceil 16\pi\epsilon^{-1} \rceil + 1$ to get a quadratic system Ψ of dimension $d \leqslant 8r\epsilon^{-2}+1$. Now we apply Lemma 3.5.3 to each of the f_i with this choice of Ψ and R as before to get functions F_i satisfying the conclusion of the lemma. Define $g_i = \Pi_R^\Psi f_i$ for $i = 1, \dots, r$. We will now show that the functions F_i, g_i satisfy the desired properties.

(*i*) As $d \leqslant 8r\epsilon^{-2}+1$ we set $M(r,\epsilon) = \sup_{d \leqslant 8r\epsilon^{-2}+1} M_0(\epsilon, d, R)$, where M_0 is as given in Lemma 3.5.3. Provided that p satisfies the conditions of Lemma 3.5.3, this follows immediately from property (*ii*) of Lemma 3.5.3. We therefore also set $p_3(r,\epsilon) = \sup_{d \leqslant 8r\epsilon^{-2}+1} p_4(\epsilon, d, R)$, with p_4 as given in Lemma 3.5.3. Note that both of these supremums are taken over integer values of d, and are therefore finite.

(*ii*) This follows immediately by the choice of Ψ from Lemma 3.5.2.

(*iii*) This follows immediately by the choice of F_i from Lemma 3.5.3.

(*iv*) By the choice of Ψ, Lemma 3.5.2 and the definition of g_i we get

$$\|1_{c^{-1}(i)} - g_i\|_{u^3} = \|f_i - \Pi_R^\Psi f_i\|_{u^3} \leqslant \epsilon.$$

(*v*) By property (*i*) of Lemma 3.5.3 we get that

$$\sum_i F_i \circ \Psi \geqslant \sum_i \Pi_R^\Psi f_i = \Pi_R^\Psi \sum_i f_i = \Pi_R^\Psi 1 = 1.$$

□

3.6 Counting lemma

Before proving the main counting lemma, we give the proof of Lemma 3.3.5.

Proof of Lemma 3.3.5. Write F in terms of its Fourier coefficients to get

$$\mathbb{E}_x F \circ \Psi(x) = \sum_{\xi_1, \xi_2} \widehat{F}(\xi_1, \xi_2) \mathbb{E}_x e_p(\xi_1 \cdot a\, x^2 + \xi_2 \cdot a\, x)$$

We now use that

$$\mathbb{E}_x e_p(ax^2 + bx) \ll p^{-1/2} \qquad \text{if } (a,b) \neq (0,0), \tag{3.23}$$

in order to discard all terms for which $\xi_1 \cdot a \neq 0$ or $\xi_2 \cdot a \neq 0$. This contributes an error $O(Mp^{-1/2})$, where $M = \|F\|_{\text{trig}}$, and so we have

$$\mathbb{E}_x F \circ \Psi(x) = \sum_{\xi_1,\xi_2 \in \Lambda_\Psi} \widehat{F}(\xi_1,\xi_2) + O(Mp^{-1/2}).$$

At the same time we have

$$\int F \, d\mu_{H_\Psi} = \sum_{\xi_1,\xi_2} \widehat{F}(\xi_1,\xi_2) \int e(\xi_1 \cdot t) e(\xi_2 \cdot u) \, d\mu_{H_\Psi}(t,u),$$

which by Lemma 3.3.3 is equal to the main term above. □

Proof of Lemma 3.4.2. Write $a = (a_i)_{i=1}^d$ for the coefficients of the quadratic system Ψ. Using the Fourier expansion of F and multilinearity of T one gets

$$T(F \circ \Psi, F \circ \Psi, F \circ \Psi) = \sum_{\xi_1,\dots,\xi_6 \in \mathbb{Z}^d} \widehat{F}(\xi_1,\xi_2)\widehat{F}(\xi_3,\xi_4)\widehat{F}(\xi_5,\xi_6) T(f_1,f_2,f_3)$$

where $f_1(x) = e_p(\xi_1 \cdot a\, x^2 + \xi_2 \cdot a\, x)$, $f_2(y) = e_p(\xi_3 \cdot a\, y^2 + \xi_4 \cdot a\, y)$ and $f_3(z) = e_p(\xi_5 \cdot a\, z^2 + \xi_6 \cdot a\, z)$. We proceed as in Proposition 3.2.3 to get that

$$T(f_1,f_2,f_3) = \sum_{\xi \in \mathbb{F}_p} \widehat{f_1}(-\xi)\widehat{f_2}(-\xi)\widehat{g}(\xi).$$

Now $\widehat{f_1}(-\xi) \ll p^{-1/2}$ unless $\xi_1 \cdot a = 0$ and $\xi_2 \cdot a + \xi = 0$, by (3.23). Similarly we have $\widehat{f_2}(-\xi) \ll p^{-1/2}$ unless $\xi_3 \cdot a = 0$ and $\xi_4 \cdot a + \xi = 0$, and from considering f_3 we have that $\widehat{g}(\xi) \ll p^{-1/2}$ unless $\xi_5 \cdot a - \xi = 0$ and $\xi_6 \cdot a = 0$. If all of these fail at once we say that $(\xi_1,\dots,\xi_6,\xi)$ is *exceptional*. For the sum over non-exceptional values we apply Proposition 3.2.3 to get that $T(f_1,f_2,f_3) \ll p^{-1/2}$, and then we bound the sum over $\xi_1,\dots,\xi_6$ by M^3.

For the exceptional values we have that $\xi = -\xi_2 \cdot a = -\xi_4 \cdot a = \xi_5 \cdot a$ and $\xi_1 \cdot a = \xi_3 \cdot a = \xi_6 \cdot a = 0$. Recalling (3.6) we may write this as

$$T(F \circ \Psi, F \circ \Psi, F \circ \Psi) = \sum_{\substack{\xi_1,\xi_3,\xi_6 \in \Lambda_\Psi \\ \xi_2+\xi_5,\xi_4+\xi_5 \in \Lambda_\Psi}} \widehat{F}(\xi_1,\xi_2)\widehat{F}(\xi_3,\xi_4)\widehat{F}(\xi_5,\xi_6) + O(p^{-1/2}M^3).$$

By inserting the Fourier expansion of F into the integral in (3.11) and using the orthogonality relation in Lemma 3.3.3 we are done, in exactly the same way as in the proof of Lemma 3.3.5. □

3.7 Ramsey lemma

The linear Ramsey problem addressed by the Ramsey lemma is to count triples (t, u), (t', u'), (t'', u'') such that $u + u' = t''$. In [32] Green–Sanders's linear Ramsey problem appears quite similar to ours, namely finding (t, u), (t', u') and (t'', u'') such that $u'' = t' - t$ and $u = u'$. Although their problem does not directly imply ours, it turns out that their proof can be easily adapted to solve our problem, which is the approach we take here. We note that the basic scheme of their proof is inspired by [16].

Lemma 3.7.1. *Let (X, ν_X) and (Y, ν_Y) be probability spaces, $A \subset X \times Y$, $(\nu_X \times \nu_Y)(A) = \alpha$ and $\eta \in (0, 1]$ be a parameter. Then there is a measurable set $Y' \subset Y$, with $\nu_Y(Y') \geqslant \frac{1}{2}\alpha$, such that the set*

$$E = \left\{ y \in Y : \nu_X(x \in X : (x, y) \in A) \leqslant \frac{1}{2}\eta\alpha \right\}$$

satisfies $\nu_Y(E \cap Y') \leqslant \eta \nu_Y(Y')$.

Proof. Define

$$N_X(y) = \{x \in X : (x, y) \in A\},$$
$$N_Y(x) = \{y \in Y : (x, y) \in A\}.$$

Then

$$\alpha = (\nu_X \times \nu_Y)(A) = \int \nu_X(N_X(y))\, d\nu_Y(y)$$

by Fubini's theorem. By the definition of E we have

$$\int 1_E(y) \nu_X(N_X(y))\, d\nu_Y(y) \leqslant \frac{1}{2}\eta\alpha,$$

so that

$$\int \left(1 - \frac{1}{\eta} 1_E(y)\right) \nu_X(N_X(y))\, d\nu_Y(y) \geqslant \frac{1}{2}\alpha.$$

Another application of Fubini's theorem then gives

$$\iint \left(1 - \frac{1}{\eta} 1_E(y)\right) 1_{N_Y(x)}(y)\, d\nu_Y(y) d\nu_X(x) \geqslant \frac{1}{2}\alpha,$$

and in particular there is some $x \in X$ such that

$$\int \left(1 - \frac{1}{\eta} 1_E(y)\right) 1_{N_Y(x)}(y) d\nu_Y(y) \geqslant \frac{1}{2}\alpha. \tag{3.24}$$

For this x, set $Y' = N_Y(x)$. From (3.24) one gets that $\nu_Y(Y') \geqslant \frac{1}{2}\alpha$ and that $\nu_Y(E \cap Y') \leqslant \eta \nu_Y(Y')$, as desired. □

Throughout this section, let G be a compact Abelian group with probability Haar measure μ. For any measurable subset $T \subset G$ we write $\mu_T = \frac{1}{\mu(T)}\,\mu|_T$. Define

$$\delta_T(A) = \int 1_A(t_1 + t_2, t)\, d\mu_T^{\otimes 3}(t_1, t_2, t) \tag{3.25}$$

and

$$\Lambda_T(A) = \int 1_A(t_1 + t_6, t_2) 1_A(t_3 + t_7, t_4) 1_A(t_4 + t_2, t_5) d\mu_T^{\otimes 7}(t_1, \ldots, t_7). \tag{3.26}$$

The main Ramsey lemma will follow from the following slightly stronger proposition. We need to allow for partial colourings and make $\rho(r)$ explicit in order to perform induction on r.

Proposition 3.7.2. *Set $\epsilon_r = 2^{-7r}(r!)^{-3}$. Suppose that G is a compact Abelian group, $T \subset G$ is a measurable set, $E \subset (T+T) \times T$ is measurable and $c : (T+T) \times T \to [r]$ is a measurable partial colouring defined outside of E, where $\delta_T(E) \leqslant \epsilon_r$. Then there is an $i \in [r]$ which satisfies $\Lambda_T(c^{-1}(i)) \geqslant \epsilon_r^3$.*

Proof. We proceed by induction on r. For $r = 1$ we have that

$$\Lambda_T(c^{-1}(1)) = \Lambda_T(((T+T) \times T) \setminus E) \geqslant 1 - 3\delta_T(E) \geqslant 1 - 3\epsilon_1 > \epsilon_1^3,$$

and so the proposition holds in this case.

Assume that the statement holds for $r-1$ colours. Now $\epsilon_r < \frac{1}{2}$ so by the pigeon hole principle there is some i with $\delta_T(A_i) \geqslant \frac{1}{2r}$, where we have defined

$$A_i = c^{-1}(i), \quad i = 1, \ldots, r.$$

Apply Lemma 3.7.1 with $X = T + T$, $Y = T$, $A = A_i$, $\eta = \frac{1}{6}\epsilon_{r-1}$, $\nu_Y = \mu_T$ and ν_X defined by

$$\nu_X(C) = \int 1_C(t_1 + t_2)\, d\mu_T^{\otimes 2}(t_1, t_2).$$

Note that $\nu_X \times \nu_Y = \delta_T$. The lemma gives us some $T' \subset T$ with $\mu_T(T') \geqslant \frac{1}{4r}$ and some $Z \subset T$ with a proportion $1 - \eta$ of all of the elements in T', which satisfies

$$\int 1_{A_i}(t_1 + t_2, t)\, d\mu_T^{\otimes 2}(t_1, t_2) \geqslant \frac{1}{4r}\eta$$

for all t in Z.

We now consider two cases. In the first case, assume that $\delta_{T'}(A_i) \geqslant \frac{1}{2}\epsilon_{r-1}$. Then restricting the integrals over t_2, t_4 and t_5 in (3.26) to T' gives

$$\begin{aligned}\Lambda_T(A_i) &\geqslant \mu_T(T')^3 \int 1_{A_i}(t_2+t_4,t_5)\left(\int 1_{A_i}(t_1+t_6,t_2)\,d\mu_T^{\otimes 2}(t_1,t_6)\right)\\ &\qquad\cdot\left(\int 1_{A_i}(t_3+t_7,t_4)\,d\mu_T^{\otimes 2}(t_3,t_7)\right)\,d\mu_{T'}^{\otimes 3}(t_2,t_4,t_5)\\ &\geqslant \mu_T(T')^3\frac{\eta^2}{2^4r^2}\int 1_{A_i}(t_2+t_4,t_5)1_Z(t_2)1_Z(t_4)\,d\mu_{T'}(t_2,t_4,t_5)\\ &\geqslant \mu_T(T')^3\frac{\eta^2}{2^4r^2}\left(\int 1_{A_i}(t_2+t_4,t_5)\,d\mu_{T'}(t_2,t_4,t_5)-\frac{1}{3}\epsilon_{r-1}\right)\\ &\geqslant \mu_T(T')^3\frac{\eta^2}{2^4r^2}\frac{1}{6}\epsilon_{r-1} \geqslant 2^{-13}3^{-3}r^{-5}\epsilon_{r-1}^3 = 2^8 3^{-3}r^4\epsilon_r^3 > \epsilon_r^3,\end{aligned}$$

and we are done in this case. If instead $\delta_{T'}(A_i) < \frac{1}{2}\epsilon_{r-1}$ we will use that

$$\delta_{T'}(E) \leqslant \mu_T(T')^{-3}\delta_T(E) \leqslant 2^6 r^3\epsilon_r = \frac{1}{2}\epsilon_{r-1}.$$

Defining $E' = (E\cup A_i)\cap((T'+T')\times T')$ then gives $\delta_{T'}(E') \leqslant \epsilon_{r-1}$. We now have a partial colouring of $(T'+T')\times T'$ outside E' using $r-1$ colours, and the inductive hypothesis then gives some colour class A_j with $\Lambda_{T'}(A_j) \geqslant \epsilon_{r-1}^3$. Considering the corresponding colour class A_j on $(T+T)\times T$ gives

$$\Lambda_T(A_j) \geqslant \mu_T(T')^7\Lambda_{T'}(A_j) \geqslant (4r)^{-7}\epsilon_{r-1}^3 = 2^7r^2\epsilon_r^3 > \epsilon_r^3,$$

and so we are done. □

Proof of Lemma 3.4.3. Put $f_i = \min(F_i, 1)$, such that $0 \leqslant f_i \leqslant 1$ and $\sum_{i=1}^r f_i(g_1,g_2) \geqslant 1$ for all $(g_1,g_2)\in G\times G$. Define

$$A_i = \Big\{(t,u)\in G\times G : f_i(t,u) \geqslant \frac{1}{r}, (t,u)\notin \bigcup_{j=1}^{i-1} A_j\Big\}. \tag{3.27}$$

If $(t,u)\notin\bigcup_i A_i$ then $\sum_i f_i(t,u) < 1$, and so we must have $\bigcup_i A_i = G\times G$. Restricting to A_i we get

$$\begin{aligned}&\int f_i(t,u)f_i(t',u')f_i(u+u',u'')\,d\mu^{\otimes 5}(t,t',u,u',u'')\\ &\qquad\geqslant \frac{1}{r^3}\int 1_{A_i}(t,u)1_{A_i}(t',u')1_{A_i}(u+u',u'')\,d\mu^{\otimes 5}(t,t',u,u',u'')\end{aligned}$$

Introducing two dummy integrations and renaming the variables then gives

$$\frac{1}{r^3}\int 1_{A_i}(t_1+t_6,t_2)1_{A_i}(t_3+t_7,t_4)1_{A_i}(t_2+t_4,t_5)\,d\mu^{\otimes 7}(t_1,\dots,t_7).$$

Setting $T=G$ in Proposition 3.7.2 we finally get that for some i this is $\geqslant r^{-3}\epsilon_r^3$, which is the desired function $\rho(r)$. □

3.8 Proof of main theorem in the general case

Before sketching the proof of Theorem 3.1.2, we note that one needs three different versions of the Ramsey lemma, depending on the exponents a, b, c in (3.2). The first version is needed when $a = b = c$, the second version is needed when exactly two of a, b and c are equal, and the third version is used when all of a, b and c are different. The necessity for different Ramsey lemmas arises because the conclusion of the counting lemma depends on the number of unique exponents, and this is of course what dictates the correct shape of the Ramsey lemma.

As mentioned in the introduction, the case where $a = b = c$ was already solved in [15] by using Schur's result on partition regularity of the equation $x + y = z$. Rather satisfyingly this is the same equation we end up having to address in the Ramsey lemma when using our methods.

At a first glance it may seem strange that the proof of Theorem 3.1.1 carries over to Theorem 3.1.2 without any major modifications. In particular we note that the proof does not rely in any significant way on the fact that (3.1) has a linear variable.

We begin by noting the changes that need to be made to the basic definitions in Section 3.2 and Section 3.3. This will essentially consist of redefining $T(f_1, f_2, f_3)$, $\Psi(x)$ and defining a new norm.

We consider the quantity

$$T(f_1, f_2, f_3) = \frac{1}{p} \sum_{\substack{x,y,z \in \mathbb{F}_p \\ x^a + y^b = z^c}} f_1(x) f_2(y) f_3(z),$$

and define a norm

$$\|f\|_{a,b,c} := \sup \left\{ |\mathbb{E}_x f(x) e_p(rx^a + sx^b + tx^c)| : r, s, t \in \mathbb{F}_p \right\}.$$

In this setting the analogue of Proposition 3.2.3 becomes the following.

Proposition 3.8.1. *If* $\|f_1\|_2, \|f_2\|_2, \|f_3\|_2 \leqslant 1$ *then*

$$|T(f_1, f_2, f_3)| \leqslant \sqrt{abc} \min_i (\|f_i\|_{a,b,c}).$$

Sketch of proof. Define $g_1(w) = \sum_{x^a = w} f_1(x)$, $g_2(w) = \sum_{u^b = w} f_2(u)$ and $g_3(w) = \sum_{z^c = w} f_3(z)$, and perform the same manipulations as in the proof of Proposition 3.2.3. □

We also have the analogue of Lemma 3.2.4, at the cost of a constant factor.

Lemma 3.8.2. *Let $f_1, f_2, f_3 : \mathbb{F}_p \to \mathbb{C}$. Then*

$$|T(f_1, f_2, f_3)| \leqslant \sqrt{k}\|f_1\|_2\|f_2\|_2\|f_3\|_2,$$

where $k = \min\{a, b, c\}$.

Sketch of proof. Assume that $k = a$ and define $g(w) = \sum_{x^a=w} f_1(x)$. Doing as in the proof of Lemma 3.2.4 and using the estimate $\|g\|_2 \leqslant \sqrt{a}\|f_1\|_2$ we are then done. The cases where $a > k$ are treated similarly. □

In addition we will need a notion of a *polynomial* system instead of just a quadratic system. In this section we let $\mathbb{G} = (\mathbb{R}/\mathbb{Z}) \times (\mathbb{R}/\mathbb{Z}) \times (\mathbb{R}/\mathbb{Z})$.

Definition 3.8.3. A polynomial system of dimension d is a map $\Psi : \mathbb{F}_p \to \mathbb{G}^d$ of the form

$$\Psi(x) = (a_i x^a/p, a_i x^b/p, a_i x^c/p)_{i=1}^d,$$

where $(a_i)_{i=1}^d \subset \mathbb{F}_p^d$.

Furthermore we will need to define trigonometric polynomials on $\mathbb{G}^d$ instead of on $(\mathbb{R}/\mathbb{Z})^{2d}$ in the obvious way.

We now state the three key lemmas in this setting, and briefly sketch the proof of each.

Lemma 3.8.4 (Regularity lemma)**.** *There are functions $p_3, M, D : \mathbb{Z}_{\geqslant 0} \times (0, 1] \times \mathbb{N} \to \mathbb{R}_{\geqslant 0}$ such that the following holds. Let $c : \mathbb{F}_p \to [r]$ be an r-colouring of $\mathbb{F}_p$ and let $\epsilon > 0$. Then there is a polynomial system Ψ of dimension d, functions $F_1, \dots, F_r : \mathbb{G}^d \to \mathbb{R}_{\geqslant 0}$, and functions $g_1, \dots, g_r : \mathbb{F}_p \to [-1, 1]$, such that the following holds provided $p \geqslant p_3(r, \epsilon, K)$.*

(i) $\|F_i\|_{\text{trig}} \leqslant M(r, \epsilon, K)$ *for all i;*

(ii) $d \leqslant D(r, \epsilon, K)$*;*

(iii) $\|F_i \circ \Psi - g_i\|_2 \leqslant \epsilon$ *for all i;*

(iv) $\|1_{c^{-1}(i)} - g_i\| \leqslant \epsilon$ *for all i;*

(v) $\sum_i F_i \circ \Psi \geqslant 1$ *pointwise.*

Here $K = \max\{a, b, c\}$.

Sketch of proof. Lemma 3.4.1 is proved via the three intermediate lemmas 3.5.1, 3.5.2 and 3.5.3. The analogue of each of these in the current setting goes through with nearly identical proofs, by modifying a few constants. In particular we note that in the proof of Lemma 3.5.3 we will need the following fact instead of (3.20). Let J be an interval in $\mathbb{R}/\mathbb{Z}$ not containing 0, and let $n \in \mathbb{N}$. Then

$$\#\{x \in \mathbb{F}_p : ax^n/p \in J\} \leqslant np|J| + C$$

for any $a \in \mathbb{F}_p$ and an absolute constant C. This holds because ax^n takes each value in $\mathbb{F}_p$ at most n times.

Combining the three is also done in exactly the same way as before. □

The conclusion of the counting lemma looks slightly different depending on how many of a, b and c are equal. In general we may write the integrand in the counting lemma as

$$F(t_a, t_b, t_c)F(u_a, u_b, u_c)F(v_a, v_b, v_c)1_{t_a+u_b=v_c}.$$

If say $a = b = c$ this should be interpreted as

$$\widetilde{F}(t)\widetilde{F}(u)\widetilde{F}(t+u),$$

where we have defined $\widetilde{F}(w) = F(w, w, w)$. Similarly, if $a = b \neq c$ we recover the integrand in Lemma 3.4.2 with $\overline{F}$ in place of F, where we define $\overline{F}(u, v) = F(v, v, u)$. Below we state and sketch the proof of the counting lemma for the case where none of a, b, c are equal, and leave it to the reader to examine the other cases.

Lemma 3.8.5 (Counting lemma)**.** *Let* Ψ *be a* d*-dimensional polynomial system with distinct exponents* a, b, c*, and let* $F : \mathbb{G}^d \to \mathbb{C}$ *be a trigonometric polynomial. Then*

$$\begin{aligned} T(F \circ \Psi, F \circ \Psi, F \circ \Psi) \\ = \int F(t_1, u_1, v_1)F(t_2, u_2, v_2)F(t_3, u_3, t_1 + u_2)\, d\mu_{G_\Psi}^{\otimes 8}(t_1, \dots, u_3) \\ + O_K(p^{-1/2^{K-1}}M^3), \end{aligned}$$

where $K = \max\{a, b, c\}$ *and* $M = \|F\|_{\mathrm{trig}}$.

Sketch of proof. We follow the steps of the proof of Lemma 3.4.2, with the standard bound in Lemma 2.1.4 in place of (3.23). This will give

$$T(F\circ\Psi, F\circ\Psi, F\circ\Psi) \\ = C_{a,b,c}\sum_{\substack{\xi_2,\xi_3,\xi_4,\xi_6,\xi_7,\xi_8\in\Lambda_\Psi\\ \xi_1+\xi_9,\xi_5+\xi_9\in\Lambda_\Psi}} \widehat{F}(\xi_1,\xi_2,\xi_3)\widehat{F}(\xi_4,\xi_5,\xi_6)\widehat{F}(\xi_7,\xi_8,\xi_9) \\ + O_K(p^{-1/2^{K-1}}M^3),$$

where

$$C_{a,b,c} = \frac{1}{p^2}\sum_{x^a+y^b=z^c} 1 = \mathbb{E}_{u,v}f_a(u)f_b(v)f_c(u+v) = \sum_\xi \widehat{f_a}(\xi)\widehat{f_b}(\xi)\widehat{f_c}(-\xi) \\ = \widehat{f_a}(0)\widehat{f_b}(0)\widehat{f_c}(0) + O_K(p^{-1/2^{K-1}}),$$

where $f_n(u) = \sum_{x^n=u} 1$ for $n\in\mathbb{N}$, and we have bounded the terms with $\xi\neq 0$ as usual by noting $|\widehat{f_a}(\xi)| \ll p^{-1/2^{K-1}}$ and $\|f_b\|_2 \leqslant \sqrt{b}, \|f_c\|_2 \leqslant \sqrt{c}$. Now

$$\widehat{f_n}(0) = \mathbb{E}_u f_n(u) = \mathbb{E}_x 1 = 1$$

for $n\in\mathbb{N}$, and so $C_{a,b,c} = 1 + O_K(p^{-1/2^{K-1}})$, which completes the proof. □

Note that the constant $C_{1,1,2}$ was implicitly evaluated to 1 in the proof of Lemma 3.4.2, which is why it was not necessary to introduce this quantity there.

We also note that if some of a,b,c are equal, the values of $\xi_1,\dots,\xi_9$ which contribute to the main term will be different. This is because for a sum of the form $\mathbb{E}_x e_p(rx^a+sx^b+tx^c)$ to be small it is not enough that $(r,s,t)\neq(0,0,0)$ in this case.

Finally, because of the different forms of the counting lemma, the Ramsey lemma will have to look slightly different depending on which case we are in. In the second case the needed Ramsey lemma is in fact Lemma 3.4.3. We here therefore prove the first and the third case.

Lemma 3.8.6 (Ramsey lemma, case 1)**.** *There is a positive function $\rho:\mathbb{Z}_{\geqslant 0}\to(0,1]$ with the following property. Assume that $F_1,\dots,F_r : G_\Psi\to\mathbb{R}_{\geqslant 0}$ are continuous functions that satisfy $\sum_{i=1}^r F_i(x)\geqslant 1$ for all $x\in G_\Psi$. Then there is an $i\in[r]$ such that*

$$\int F_i(t_1)F_i(t_2)F_i(t_1+t_2)\,d\mu_{G_\Psi}^{\otimes 2}(t_1,t_2) \geqslant \rho(r).$$

Proof. We could prove this by using the same techniques as in the proof of Lemma 3.4.3, but in this case our result actually just follows from a quantitative version of Schur's theorem. Note that if $a_i \equiv 0 (\mathrm{mod}\ p)$ for all i then $G_\Psi = \{0\}$, and if $a_i \not\equiv 0 (\mathrm{mod}\ p)$ for some i we have $G_\Psi = \{ax/p : x \in \mathbb{F}_p\} \simeq \mathbb{F}_p$, by the definition of G_Ψ. If $G_\Psi = \{0\}$ the statement we are trying to prove is trivial, as $F(0) \geqslant 1/r$ for some i, and for this i the above integral is $\geqslant 1/r^3$.

In the non-trivial case, define sets A_i as in the proof of Lemma 3.4.3 to get an r colouring of G_Ψ. We then invoke [24, Theorem 1], which states that given an r-colouring of $\mathbb{N}$ there are $\geqslant \nu(r)N^2$ monochromatic solutions to $x + y = z$ with $x, y, z \leqslant N$. Set $p = N$, then certainly we also have $\geqslant \nu(r)p^2$ monochromatic solutions to $x + y = z$ with $x, y, z \in \mathbb{F}_p$, which completes the proof. □

Lemma 3.8.7 (Ramsey lemma, case 3)**.** *There is a positive function $\rho : \mathbb{Z}_{\geqslant 0} \to (0, 1]$ with the following property. Suppose that G is a compact Abelian group with Haar probability measure μ. Assume also that $F_1, \dots, F_r : G \times G \times G \to \mathbb{R}_{\geqslant 0}$ are continuous functions that satisfy $\sum_{i=1}^r F_i(g_1, g_2, g_3) \geqslant 1$ for all $(g_1, g_2, g_3) \in G \times G \times G$. Then there is an $i \in [r]$ such that*

$$\int F_i(t_1, u_1, v_1) F_i(t_2, u_2, v_2) F_i(t_3, u_3, t_1 + u_2)\, d\mu^{\otimes 8}(t_1, \dots, u_3) \geqslant \rho(r).$$

Sketch of proof. The linear Ramsey problem we are addressing is slightly different from the one in Lemma 3.4.3, but essentially the same proof goes through. Introducing two dummy integrations we can view this as a problem on $T \times T \times (T + T)$. Invoking Lemma 3.7.1 with $X = T \times (T + T)$ and $Y = T$ to get a set $T' \subset T$, and then restricting t_1 and u_2 to this set, the proof goes through as in the proof of Lemma 3.4.3.

□

Finally the three lemmas are combined in almost exactly the same way as in Section 3.4.

Chapter 4

Monochromatic solutions to $x + y = z^2$

4.1 Introduction

In this chapter we will be concerned with the Ramsey theory of the equation $x+y = z^2$. It was shown by Csikvári, Gyarmati and Sárközy in [15] that this equation is *not* partition regular. Indeed, a 16-colouring of $\mathbb{N}$ is exhibited with no monochromatic solutions to $x + y = z^2$ other than the trivial one $x = y = z = 2$. There remains the question of whether the 16 here is optimal. The main theorem completely answers this question.

Theorem 4.1.1. *There is a 3-colouring of $\mathbb{N}$ with no monochromatic solution to $x + y = z^2$ other than the trivial one. On the other hand, every 2-colouring of $\mathbb{N}$ has infinitely many monochromatic solutions to $x + y = z^2$.*

The proof of the first statement is rather simple. It is given in Section 4.2. By contrast, the proof that every 2-colouring has infinitely many monochromatic solutions to $x + y = z^2$ is complicated and involves a surprisingly large number of tools from additive combinatorics and number theory.

In [53] Pach presents a much shorter combinatorial proof of Theorem 4.1.1. Furthermore in [49] this result was improved further to show that in fact any 2-colouring of $[N]$ has at least $N^{1/4-o(1)}$ monochromatic solutions to $x + y = z^2$. Despite Pach's much shorter proof of Theorem 4.1.1 we believe that our methods can be applied to a much wider range of problems. In particular, in the next chapter we show how to adapt the methods presented here to prove partition regularity of the superficially similar looking equation $x - y = z^2$.

We now outline the proof of Theorem 4.1.1.

If $\mathbb{N} = V \cup W$ then let us assume that there are infinitely many N such that $|V \cap [N, 2N)| \geq N/2$. If this is not the case then a corresponding statement holds for W and we may switch the roles of V and W in what follows. Suppose that there are no solutions to $x + y = z^2$ in either V or W. By a fairly elaborate sequence of arguments involving the arithmetic regularity lemma as well as certain Fourier-analytic and Diophantine arguments, as well as a deep result of Lagarias, Odlyzko and Shearer, we use this to show that for some $q \in N$ and $c > 0$ the set W contains the progression $\mathbb{P}([1, 1 + c]; M, q) := \{n \in \mathbb{Z} : M \leq n \leq (1 + c)M, n \equiv 0 (\text{mod } q)\}$ for infinitely many integers M. The details of these arguments may be found in Sections 4.4 and 4.5, certain preliminary results having been assembled in Section 4.3. The proof is concluded in Section 4.7 by performing an iterative argument to get a collection of further progressions inside W, eventually showing that all sufficiently large multiples of q lie in W. An important ingredient here is a result concerning gaps between sums of two squares with certain constraints, proven in Section 4.6.

The fact that all sufficiently large multiples of q lie in W leads immediately to a contradiction, since W then obviously contains infinitely many solutions to $x+y = z^2$.

We make heavy use of smooth cutoff functions in the latter half of the chapter. The properties and constructions of these are recalled in Appendix C.

Let us remark on the nice work of Khalfalah and Szemerédi [42] which, despite its rather similar title, concerns a somewhat different problem. They show that any finite colouring of $\mathbb{N}$ contains a solution to $x + y = z^2$ with x and y having the same colour (but not necessarily z). They show that if one passes to an appropriate subprogression then in fact a density result holds. This allows them to apply Fourier analysis. In Appendix B we adapt their proof to $\mathbb{Z}/q\mathbb{Z}$.

We also remark that for the modular version of the problem the answer is very different. Indeed, in [48] (included as Chapter 3) we show that if $p > p_0(k)$ is a prime and if $\mathbb{Z}/p\mathbb{Z}$ is k-coloured, then there are $\gg_k p^2$ monochromatic solutions to $x + y = z^2$.

Notation. We collect here some notation used in this chapter. The notation $\widehat{f}$ always denotes Fourier transform. At various points in this chapter f may be a function on $\mathbb{Z}$, $\mathbb{R}$ or $\mathbb{T}^d$. The definitions we are using are recalled in the text when there is any danger of confusion.

It is convenient to introduce a piece of notation which is less standard, but very useful. If $\Lambda \subset \mathbb{N}$ is a set of integers then we write $\sqrt{\Lambda} := \{n \in \mathbb{N} : n^2 \in \Lambda\}$ (this is *not* the same as $\{\sqrt{n} : n \in \Lambda\}$). If $A \subset \mathbb{N}$ is a set, we write $2A = A + A := \{a + a' : a, a' \in A\}$. We will sometimes use notation such as $2\sqrt{2A}$, which means $\sqrt{A + A} + \sqrt{A + A}$.

Finally, as hinted above, when $I \subset \mathbb{R}$ is a closed interval we write $\mathbb{P}(I; N, q) := \{n \in \mathbb{Z} : \frac{n}{N} \in I, q|n\}$.

4.2 A 3-colouring

In this short section we establish the easy part of Theorem 4.1.1. That is, we exhibit a 3-colouring of $\mathbb{N}$ for which the only monochromatic solution to $x + y = z^2$ is the trivial solution $x = y = z = 2$. We colour all the points in each dyadic block

$$A_i = \{n \in \mathbb{N} : 2^i \leq n < 2^{i+1}\},$$

$i = 0, 1, 2, \dots$, in one colour c_i. We assign c_0, c_1, c_2 to be distinct, and then assign the colours c_i, $i \geq 3$, inductively in such a way that $c_i \notin \{c_{\lfloor i/2 \rfloor}, c_{\lfloor i/2 \rfloor + 1}\}$. Note that this is possible since $\lfloor i/2 \rfloor + 1 < i$ for $i \geq 3$.

Assume now that $x, y, z \in \mathbb{N}$ have the same colour and that $x + y = z^2$. Without loss of generality we may assume that $x \leq y$. Let $i \in \{0, 1, 2, \dots\}$ be such that $y \in A_i$. Then $2^i < x + y < 2^{i+2}$, and hence $2^{i/2} < z < 2^{(i+2)/2}$. Since $i/2 \geq \lfloor i/2 \rfloor$ and $(i + 2)/2 \leq \lfloor i/2 \rfloor + 2$, it follows that $z \in A_{[i/2]} \cup A_{[i/2]+1}$. By construction, the only way that such a z can have the same colour as y is if $i \in \{0, 1, 2\}$, in which case $x \leq y < 8$, and so $z = 2$ or 3. An easy case check confirms that $x = y = z = 2$.

4.3 Results from the literature

The rest of this chapter is devoted to the harder part of Theorem 4.1.1. In this section we assemble some basic ingredients from the literature.

The following definition is relevant to much of this chapter. Recall that for $x \in \mathbb{R}$ we write $\|x\|_{\mathbb{T}} = \min_{n \in \mathbb{Z}} |x - n|$.

Definition 4.3.1. Suppose that $\theta \in \mathbb{R}^d$. Let $N \geq 1$ be an integer and let $A > 0$ be some real parameter. We say that θ is (A, N)-irrational if whenever $\mathbf{r} \in \mathbb{Z}^d \setminus \{0\}$ and $\|\mathbf{r}\|_1 \leq A$ we have $\|\mathbf{r} \cdot \theta\|_{\mathbb{T}} \geq A/N$.

We record a corollary of Proposition 2.1.7, phrased in the language of this definition. This corollary is the variant of Weyl's inequality that we have found to be most useful in this chapter.

Corollary 4.3.2. *Let $k, N \in \mathbb{N}$. Suppose that $I \subset \mathbb{Z}$ is a (discrete) interval of length $\leq N^{1/k}$. Suppose that $\theta \in \mathbb{R}^d$ is (A, N)-irrational, and suppose that $\mathbf{r} \in \mathbb{Z}^d \setminus \{0\}$. Then*

$$|\sum_{n \in I} e(\mathbf{r} \cdot \theta n^k + \ldots)| \leq N^{1/k} \|\mathbf{r}\|_1 A^{-1/C_k}.$$

Here, $\ldots$ denotes polynomial terms in n of degree $k-1$ or lower, and the estimate is uniform in the choice of these terms.

Proof. Suppose that the sum is $\geq \delta |I|$. Then, by Proposition 2.1.7 there is some $q \in \mathbb{N}$, $q \leq \delta^{-C_k}$, such that $\|q\mathbf{r} \cdot \theta\|_{\mathbb{R}/\mathbb{Z}} \leq \delta^{-C_k} |I|^{-k}$. Since θ is (A, N)-irrational, we have either (1) $q\|\mathbf{r}\|_1 \geq A$ or (2) $\delta^{-C_k} |I|^{-k} \geq A/N$. In case (1), the bound on q implies that $\delta^{-C_k} \|\mathbf{r}\|_1 \geq A$. In case (2), we have $\delta^{-C_k} \geq A$. Hence in either case we have $\delta^{-C_k} \|\mathbf{r}\|_1 \geq A$, and hence $\delta \leq (\|\mathbf{r}\|_1 / A)^{1/C_k}$. The result follows (in fact with $\|\mathbf{r}\|_1$ replaced by the smaller quantity $\|\mathbf{r}\|_1^{1/C_k}$). □

Turning to a different type of ingredient, we require the following estimate.

Proposition 4.3.3. *Let $S \subset \{1, \ldots, N\}$ be any set of squares. For $t \in \mathbb{R}/\mathbb{Z}$, write $\widehat{1}_S(t) := \sum_{n \in S} e(tn)$. Then $\int_0^1 |\widehat{1}_S(t)|^6 dt \ll N^2$.*

Proof. It is easy to see that the integral is $\sum_{x \leq 3N} r_{3,S}(x)^2$, where $r_{3,S}(x)$ is the number of ways of writing x as $n_1 + n_2 + n_3$ with $n_1, n_2, n_3 \in S$. This quantity is obviously largest when S is the set of *all* squares $\leq N$. In this case, the stated bound is a well-known consequence of the Hardy-Littlewood method. □

Remark. Using more advanced methods of harmonic analysis (related to the Tomas-Stein restriction theorem) one can show a bound $\int_0^1 |\widehat{1}_S(t)|^q \ll_q N^{q/2-1}$ for any $q > 4$.

Finally, we will also use the following result of Lagarias, Odlyzko and Shearer [45].

Proposition 4.3.4. *Suppose that $S \subset \mathbb{Z}/q\mathbb{Z}$, where q is a positive integer, and that $|S| > \frac{11}{32} q$. Then $S + S$ contains a quadratic residue modulo q.*

Remarks. The $\frac{11}{32}$ in this theorem is sharp, as was shown by an example of Massias [50]. To see this simply take the set of $x \equiv 1 (\mathrm{mod}\ 4)$ together with $\{14, 26, 30\}$ as a subset of $\mathbb{Z}/32\mathbb{Z}$. For our purposes, $\frac{11}{32}$ could be replaced by any constant less than $\frac{1}{2}$. A simpler proof of such a statement could probably be extracted from [45] or the companion paper [44], but we do not know of any argument that could be described as in any way routine.

Instead of the result of Lagarias, Odlyzko and Shearer, it would suffice to have the following statement: there is some $\eta_k > 0$ such that if $(1-\eta_k)q$ of the elements of $\mathbb{Z}/q\mathbb{Z}$ are k-coloured then there are x, y of the same colour with $x+y$ a square. Such a statement can be established relatively painlessly using a simplified version of the arguments of Khalfalah and Szemerédi [42]. An account of this argument is included in Appendix B.

4.4 Capturing most of the squares in a Bohr set

This section contains the technical heart of this chapter. Our aim is to prove the following result. Here, and in what follows, $\mathfrak{S}(b,q)$ denotes the number of solutions to $x^2 \equiv b (\mathrm{mod}\ q)$ with $x \in \mathbb{Z}/q\mathbb{Z}$.

Proposition 4.4.1. *Let $\eta > 0$, and let $\Omega : \mathbb{N}^3 \to \mathbb{N}$ be a function (which may depend on η), nondecreasing in each variable. Suppose that $N > N_0(\Omega, \eta)$ is sufficiently large, and let $A \subset [N, 2N]$ be a set of size at least $N/2$. Then there are $q, d = O_{\eta,\Omega}(1)$, $\epsilon \gg_{\eta,\Omega} 1$, $b \in \mathbb{Z}/q\mathbb{Z}$, $x \in [2,4]$, and $\theta, z \in \mathbb{R}^d$ such that*

1. *b is a quadratic residue modulo q;*
2. *θ is $(\Omega(q,d,1/\epsilon), N)$-irrational;*
3. *$A + A$ contains all but at most $\eta\mathfrak{S}(b,q)(2\epsilon)^{d+1}q^{-1}N^{1/2}$ of the squares in the set $\{n \in \mathbb{N} : n \equiv b (\mathrm{mod}\ q), |\frac{n}{N} - x|, \|\theta n - z\|_{\mathbb{T}^d} \leq \epsilon\}$.*

Remarks. The assumption that $|A| \geq N/2$ could be weakened to $|A| \geq cN$ for any $c > 11/32$, using essentially the same proof. We do not record this explicitly as Proposition 4.4.1 seems unlikely to be of independent interest. In our applications, η will be an absolute constant which could be specified explicitly if desired ($\eta = 10^{-10}$ should certainly be admissible).

The key tool in the proof of Proposition 4.4.1 will be the *arithmetic regularity lemma*, introduced in [29]. The formulation we use here, in a more general guise, is the main result of [34]. That paper is long and quite difficult, but only Sections 1 and 2 of it are relevant to us. Furthermore, that paper establishes a regularity lemma for the Gowers U^{s+1}-norm for general s, whereas we only need the case $s = 1$. This means that the notion of a *nilsequence*, beyond the abelian case, is not relevant here. A complete, self-contained proof of the arithmetic regularity lemma in the form we need it here can be written up in less than 10 pages. Conveniently, such a writeup has been provided by Sean Eberhard [22].

Here is the arithmetic regularity lemma in the form we will need it.

Proposition 4.4.2. *Suppose we are given $\delta > 0$ and an increasing function $\mathcal{F} : \mathbb{N} \to \mathbb{R}_+$. Then there exists $M_{\max} \ll_{\delta,\mathcal{F}} 1$ such that for any function $f : [N, 2N) \to [0, 1]$ there is an $M \leq M_{\max}$ and a decomposition $f = f_{\text{tor}} + f_{\text{sml}} + f_{\text{unf}}$ into functions taking values in $[-1, 1]$, where $\sum_{N \leq n < 2N} |f_{\text{sml}}(n)| \leq \delta N$, $\|\widehat{f}_{\text{unf}}\|_\infty \leq N/\mathcal{F}(M)$ and $f_{\text{tor}}(n) = F(n(\text{mod } q), n/N, \theta n)$ for some $q, d \leq M$ and some function $F : \mathbb{Z}/q\mathbb{Z} \times [1, 2] \times \mathbb{T}^d \to [0, 1]$ with Lipschitz constant at most M. Furthermore θ may be taken to be $(\mathcal{F}(M), N)$-irrational.*

We remark that in the works previously cited the function f_{unf} was controlled in terms of the Gowers U^2-norm, rather than in terms of the supremum norm of the Fourier transform, defined by

$$\widehat{f}_{\text{unf}}(t) := \sum_{N \leq n < 2N} f_{\text{unf}}(n) e(-tn).$$

However it is well-known that for bounded functions these norms are essentially equivalent, as seen by (2.1).

Moreover f_{sml} is traditionally controlled in the ℓ^2-norm, rather than the ℓ^1-norm as we have here. However, since f_{sml} is bounded by 1, these two norms are equivalent too. Thus Proposition 4.4.2 is equivalent to the arithmetic regularity lemma as usually stated.

Let us now begin the proof of Proposition 4.4.1 in earnest. Apply Proposition 4.4.2 with $f = 1_A$, $\delta < \eta$ some small constant ($\delta = 10^{-100}$ would be permissible), and the function $\mathcal{F}$ to be specified later (it will depend on Ω and η). This gives integers $q, d \leq M$, $\theta \in \mathbb{R}^d$ and $F : \mathbb{Z}/q\mathbb{Z} \times [1, 2] \times \mathbb{T}^d \to [0, 1]$ and a decomposition

$$1_A = f_{\text{tor}} + f_{\text{sml}} + f_{\text{unf}} \tag{4.1}$$

with the properties described in the statement of Proposition 4.4.2 just given.

Lemma 4.4.3. *Suppose that δ is sufficiently small and that $\mathcal{F}$ grows sufficiently rapidly. Then $\int F d\mu > \frac{9}{20}$, where μ denotes the natural[1] measure on $\mathbb{Z}/q\mathbb{Z} \times \mathbb{R} \times \mathbb{T}^d$.*

Remark. Here, $\frac{9}{20}$ is simply a convenient fraction less than $\frac{1}{2}$. In fact, $\int F d\mu$ can be made as close to $\frac{1}{2}$ as one wishes by reducing δ and increasing $\mathcal{F}(M)$.

[1] The product of the uniform probability measure on $\mathbb{Z}/q\mathbb{Z}$, Lebesgue measure on $\mathbb{R}$ and normalised Lebesgue measure on $\mathbb{T}^d$.

Proof. We begin by noting that, by assumption,

$$\mathbb{E}_{N\leq n<2N}1_A(n) \geq \frac{1}{2}. \tag{4.2}$$

If $\delta < \frac{1}{100}$ then

$$|\mathbb{E}_{N\leq n<2N}f_{\mathrm{sml}}(n)| < \frac{1}{100}. \tag{4.3}$$

Also, introducing a smooth majorant ψ for $[N,2N)$ with $\psi(n)=1$ for $N\leq n<2N$ we have

$$\begin{aligned}|\mathbb{E}_{N\leq n<2N}f_{\mathrm{unf}}(n)| &= |\frac{1}{N}\sum_n \psi(n)f_{\mathrm{unf}}(n)|\\ &= |\frac{1}{N}\int_0^1 \widehat{\psi}(t)\widehat{f_{\mathrm{unf}}}(t)dt|\\ &\leq \frac{\|\widehat{\psi}\|_1}{\mathcal{F}(M)},\end{aligned}$$

where we have used Parseval's Theorem and simply extended f_{unf} from a function on $[N,2N)$ to a function on $\mathbb{Z}$ by setting $f_{\mathrm{unf}}(x)=0$ for $x\notin[N,2N)$. With an appropriate choice of ψ (see Lemma C.1 for details) we have $\|\widehat{\psi}\|_1 = O(1)$, and so if $\mathcal{F}(M)$ is sufficiently large it follows that

$$|\mathbb{E}_{N\leq n<2N}f_{\mathrm{unf}}(n)| < \frac{1}{100}. \tag{4.4}$$

We also have

$$\mathbb{E}_{N\leq n<2N}f_{\mathrm{tor}}(n) = \mathbb{E}_{N\leq n<2N}F(n(\mathrm{mod}\ q), \frac{n}{N}, \theta n).$$

However, it was proven[2] in [23, Lemma A.4] that, if $\mathcal{F}$ grows sufficiently rapidly in terms of δ and if N is big enough in terms of δ,

$$|\mathbb{E}_{N\leq n<2N}F(n(\mathrm{mod}\ q), \frac{n}{N}, \theta n) - \int F d\mu| < \frac{1}{100}. \tag{4.5}$$

Combining (4.2), (4.3), (4.4), (4.5) concludes the proof. □

Now let $U\subset\mathbb{Z}/q\mathbb{Z}$ be the set of all $u\in\mathbb{Z}/q\mathbb{Z}$ for which

$$\int_1^2\int_{\mathbb{T}^d} F(u,x,z)dzdx \geq \frac{1}{20} \tag{4.6}$$

and for which

$$\sum_{\substack{N\leq n<2N\\ n\equiv u(\mathrm{mod}\ q)}} |f_{\mathrm{sml}}(n)| \leq \frac{20\delta}{q}N. \tag{4.7}$$

[2]This is not an especially difficult argument: roughly, one approximates F by a function with finite Fourier support, then uses the irrationality of θ in estimating the resulting exponential sums.

One should think, informally, of these being the residue classes (mod q) on which A has "significant mass".

Lemma 4.4.4. *Suppose that δ is sufficiently small and that $\mathcal{F}$ grows sufficiently rapidly. There are elements $u, u' \in U$ such that $u + u'$ is a quadratic residue modulo q.*

Proof. Let $U_1 \subset \mathbb{Z}/q\mathbb{Z}$ be the set of all u for which (4.6) fails, and U_2 the set of all u for which (4.7) fails. Since $\sum_{N \le n < 2N} |f_{\mathrm{sml}}(n)| \le \delta N$, we have

$$|U_2| \le \frac{q}{20}.$$

Furthermore by Lemma 4.4.3 we have

$$\begin{aligned}\frac{9}{20} < \int F d\mu &= \frac{1}{q} \sum_{u \in \mathbb{Z}/q\mathbb{Z}} \int_1^2 \int_{\mathbb{T}^d} F(u, x, z) dz dx \\ &\le \frac{1}{20} + \frac{1}{q} |(\mathbb{Z}/q\mathbb{Z}) \setminus U_1|.\end{aligned}$$

It follows that

$$|U| \ge |(\mathbb{Z}/q\mathbb{Z}) \setminus U_1| - |U_2| \ge \left(\frac{9}{20} - \frac{1}{20} - \frac{1}{20}\right) q > \frac{11q}{32}.$$

The result now follows from Proposition 4.3.4. □

Henceforth, we will fix two residue classes $u, u' \in U$ for which $u + u'$ is a quadratic residue modulo q. Define parameters $\epsilon > \epsilon' > 0$ by

$$\epsilon := \frac{\delta}{M} \tag{4.8}$$

and

$$\epsilon' := \frac{\delta}{dq} (2\epsilon)^{d+1}. \tag{4.9}$$

Note that since $q, d \le M$ we have

$$\epsilon' \ge \frac{\delta}{M^2} \left(\frac{2\delta}{M}\right)^{M+1} \gg_{\delta, M} 1. \tag{4.10}$$

(The precise form of this bound is unimportant; what matters is that there is a lower bound depending only on δ and M.)

For $x, x' \in [1, 2]$ and $z, z' \in \mathbb{T}^d$, define

$$E_{x,z} := \sum_{\substack{N \le n < 2N \\ n \equiv u \,(\mathrm{mod}\ q) \\ |\frac{n}{N} - x| \le \epsilon \\ \|\theta n - z\|_{\mathbb{T}^d} \le \epsilon}} |f_{\mathrm{sml}}(n)| \qquad \text{and} \qquad E'_{x',z'} := \sum_{\substack{N \le n < 2N \\ n \equiv u' \,(\mathrm{mod}\ q) \\ |\frac{n}{N} - x'| \le \epsilon' \\ \|\theta n - z'\|_{\mathbb{T}^d} \le \epsilon'}} |f_{\mathrm{sml}}(n)|.$$

We have

$$\int_1^2 \int_{\mathbb{T}^d} E_{x,z} dz dx = \sum_{\substack{N \le n < 2N \\ n \equiv u (\mathrm{mod}\ q)}} |f_{\mathrm{sml}}(n)| \int_1^2 1_{|\frac{n}{N} - x| \le \epsilon} dx \int_{\mathbb{T}^d} 1_{\|\theta n - z\|_{\mathbb{T}^d} \le \epsilon} dz$$

$$\le (2\epsilon)^{d+1} \sum_{\substack{N \le n < 2N \\ n \equiv u (\mathrm{mod}\ q)}} |f_{\mathrm{sml}}(n)| \le (2\epsilon)^{d+1} \frac{20\delta}{q} N,$$

the last step being a consequence of (4.7). It follows from this and (4.6) that

$$\int_1^2 \int_{\mathbb{T}^d} \left(F(u,x,z) - \frac{q}{800 N \delta (2\epsilon)^{d+1}} E_{x,z} \right) dz dx \ge \frac{1}{40},$$

and so there are specific choices of x, z such that

$$F(u,x,z) - \frac{q}{800 N \delta (2\epsilon)^{d+1}} E_{x,z} \ge \frac{1}{40},$$

which implies that

$$F(u,x,z) \ge \frac{1}{40} \qquad \text{and} \qquad E_{x,z} \le \frac{800\delta N}{q} (2\epsilon)^{d+1}. \tag{4.11}$$

Similarly, there are x', z' such that

$$F(u',x',z') \ge \frac{1}{40} \qquad \text{and} \qquad E_{x',z'} \le \frac{800\delta N}{q} (2\epsilon')^{d+1}. \tag{4.12}$$

From now on, we fix these specific choices of x, z, x', z' and set

$$X := \{ n \in \mathbb{N} : n \equiv u (\mathrm{mod}\ q), |\frac{n}{N} - x|, \|\theta n - z\|_{\mathbb{T}^d} \le \epsilon \}, \tag{4.13}$$

$$X' := \{ n \in \mathbb{N} : n \equiv u' (\mathrm{mod}\ q), |\frac{n}{N} - x'|, \|\theta n - z'\|_{\mathbb{T}^d} \le \epsilon' \}, \tag{4.14}$$

and

$$Y := \{ n \in \mathbb{N} : n \equiv u + u' (\mathrm{mod}\ q), |\frac{n}{N} - (x + x')|, \|\theta n - (z + z')\|_{\mathbb{T}^d} \le \epsilon \}. \tag{4.15}$$

Note that with this notation (4.11), (4.12) imply

$$\sum_{n \in X} |f_{\mathrm{sml}}(n)| \ll \delta (2\epsilon)^{d+1} q^{-1} N, \qquad \sum_{n \in X'} |f_{\mathrm{sml}}(n)| \ll \delta (2\epsilon')^{d+1} q^{-1} N. \tag{4.16}$$

Lemma 4.4.5. *Suppose that $\mathcal{F}$ grows sufficiently rapidly, and that N is sufficiently large in terms of δ, M. Then the number of squares in Y is $\ll (2\epsilon)^{d+1} q^{-1} \mathfrak{S}(u + u', q) N^{1/2}$.*

Proof. Let $\mathcal{A}$ be the set of all $a \in \mathbb{Z}/q\mathbb{Z}$ for which $a^2 \equiv u + u' (\mathrm{mod}\ q)$. Thus $|\mathcal{A}| = \mathfrak{S}(u+u', q)$. An upper bound for the number of squares in Y is then

$$\sum_{a \in \mathcal{A}} \sum_{n \in I} 1_{n \equiv a (\mathrm{mod}\ q)} \psi_\epsilon^+(\theta n^2 - z - z'),$$

where $I = [(x + x' - \epsilon)^{1/2} N^{1/2}, (x + x' + \epsilon)^{1/2} N^{1/2}]$ and ψ_ϵ^+ is the majorant for the characteristic function of the ball $B_\epsilon(0)$ in $\mathbb{T}^d$ constructed in Lemma C.2. Fourier expanding

$$1_{n \equiv a (\mathrm{mod}\ q)} = \frac{1}{q} \sum_{r (\mathrm{mod}\ q)} e\left(-\frac{ra}{q}\right) e\left(\frac{rn}{q}\right)$$

and

$$\psi_\epsilon^+(t) = \sum_{\mathbf{r} \in \mathbb{Z}^d} \widehat{\psi_\epsilon^+}(\mathbf{r}) e(\mathbf{r} \cdot t),$$

this may be written as

$$\sum_{a \in \mathcal{A}} \frac{1}{q} \sum_{r (\mathrm{mod}\ q)} e\left(-\frac{ra}{q}\right) \sum_{\mathbf{r} \in \mathbb{Z}^d} \widehat{\psi_\epsilon^+}(\mathbf{r}) e(-\mathbf{r} \cdot (z + z')) \sum_{n \in I} e\left(\mathbf{r} \cdot \theta n^2 + \frac{rn}{q}\right). \tag{4.17}$$

The contribution from $\mathbf{r} = 0$ is

$$\frac{1}{q} \left(\int \psi_\epsilon^+\right) \sum_{a \in \mathcal{A}} \sum_{r (\mathrm{mod}\ q)} e\left(-\frac{ra}{q}\right) \sum_{n \in I} e\left(\frac{rn}{q}\right).$$

If $r \neq 0$, the inner sum over n is at most q in magnitude, since the sum of $e(rn/q)$ over any interval of length q is zero. The total contribution from these terms is thus bounded independently of N, and so may be ignored if N is large enough. The contribution from $r = 0$ is $\frac{1}{q} \mathfrak{S}(u+u', q) (\int \psi_\epsilon^+) |I|$, which is $\ll (2\epsilon)^{d+1} q^{-1} \mathfrak{S}(u + u', q) N^{1/2}$ by Lemma C.2 (1) and the bound $|I| \ll \epsilon N^{1/2}$. The contribution to (4.17) from $\mathbf{r} \neq 0$ is bounded above by

$$\mathfrak{S}(u+u', q) \sum_{\mathbf{r} \in \mathbb{Z}^d \setminus \{0\}} |\widehat{\psi_\epsilon^+}(\mathbf{r})| \sup_{r (\mathrm{mod}\ q)} \Big| \sum_{n \in I} e(\mathbf{r} \cdot \theta n^2 + \frac{r}{q} n) \Big|.$$

By Corollary 4.3.2 and Lemma C.2 (2), this is

$$\ll q N^{1/2} \mathcal{F}(M)^{-1/C_2} \sum_{\mathbf{r} \in \mathbb{Z}^d \setminus \{0\}} |\widehat{\psi_\epsilon^+}(\mathbf{r})| \|\mathbf{r}\|_1 \ll_{\delta, M} N^{1/2} \mathcal{F}(M)^{-1/C_2}.$$

(Lemma C.2 (2) gives an implied constant depending on d, ϵ, but we have $d \leq M$ and $\epsilon = \delta/M$.) Hence if $\mathcal{F}$ is chosen to be sufficiently rapidly-growing, this is smaller than $(\frac{2\delta}{M})^{M+1} M^{-1} N^{1/2}$, which is at most $N^{1/2} (2\epsilon)^{d+1} q^{-1} N^{1/2}$. □

We will also need the following fact, proven using very similar techniques.

Lemma 4.4.6. *Suppose that $\mathcal{F}$ grows sufficiently rapidly, and that N is sufficiently large in terms of δ, M. Suppose that $n \in X$. Then the number of $n' \in X'$ for which $n+n'$ is a square is $\ll (2\epsilon')^{d+1}q^{-1}\mathfrak{S}(u+u',q)N^{1/2}$, uniformly in n.*

Proof. Once again, write $\mathcal{A}$ for the set of square roots of $u+u'$ in $\mathbb{Z}/q\mathbb{Z}$. Writing $m^2 = n+n'$, an upper bound for the quantity in question is

$$\sum_{a\in\mathcal{A}}\sum_{m\in J} 1_{m\equiv a(\mathrm{mod}\ q)}\psi^+_{\epsilon'}(\theta m^2 - \theta n - z'),$$

where $J = [(n+(x'-\epsilon')N)^{1/2}, (n+(x'+\epsilon')N)^{1/2}]$ and $\psi^+_{\epsilon'}$ is the majorant constructed in Lemma C.2 (but now with the smaller parameter ϵ'). Expanding in Fourier series much as before, this may be written as

$$\sum_{a\in\mathcal{A}}\frac{1}{q}\sum_{r(\mathrm{mod}\ q)} e\left(-\frac{ra}{q}\right)\sum_{\mathbf{r}\in\mathbb{Z}^d}\widehat{\psi^+_{\epsilon'}}(\mathbf{r})e(-\mathbf{r}\cdot\theta(n+z'))\sum_{m\in J} e\left(\mathbf{r}\cdot\theta m^2 + \frac{rm}{q}\right).$$

Arguing in an essentially identical fashion to the proof of Lemma 4.4.5, we see that this is bounded by a main term of size $\ll (2\epsilon')^{d+1}q^{-1}\mathfrak{S}(u+u',q)N^{1/2}$ plus an error of size $\ll_{\delta,M} N^{1/2}\mathcal{F}(M)^{-1/C_2}$. Choosing $\mathcal{F}$ to be sufficiently rapidly-growing, and recalling from (4.10) that $\epsilon' \gg_{\delta,M} 1$, this can be made $\ll (2\epsilon')^{d+1}q^{-1}\mathfrak{S}(u+u',q)N^{1/2}$. □

Finally, we need yet another fact with a similar proof. Define the set $Y_- \subset Y$ to be

$$\{n\in\mathbb{N} : n \equiv u+u'(\mathrm{mod}\ q), |\frac{n}{N} - (x+x')|, \|\theta n - (z+z')\|_{\mathbb{T}^d} \le \epsilon - 2\epsilon'\}. \tag{4.18}$$

Lemma 4.4.7. *Suppose that $\mathcal{F}$ grows sufficiently rapidly. Then the number of squares in $Y \setminus Y_-$ is $\ll \delta(2\epsilon)^{d+1}q^{-1}N^{1/2}$.*

Proof. If $n \in Y \setminus Y_-$ then either

$$\epsilon - 2\epsilon' < |\frac{n}{N} - (x+x')| < \epsilon \tag{4.19}$$

or

$$\epsilon - 2\epsilon' < \|\theta_i n - (z_i + z_i')\|_{\mathbb{T}^d} < \epsilon \tag{4.20}$$

for some $i \in \{1,\dots,d\}$. The number of squares satisfying (4.19) is elementarily seen to be $O(\epsilon' N^{1/2})$, which[3] is bounded as desired because of the choice of ϵ' (cf. (4.9)).

[3]Obviously this bound is rather crude, as we have completely ignored the fact that additionally $n \equiv u+u'(\mathrm{mod}\ q)$ and $\|\theta n - (z+z')\|_{\mathbb{T}^d} \le \epsilon$, but this is of little consequence in the grand scheme of the argument.

We now obtain an upper bound for the number of squares satisfying (4.20). By translating the function ψ_ϵ^+ constructed in Lemma C.2 (with $d = 1$ in that lemma) we may obtain a smooth majorant ψ for the interval $\{t \in \mathbb{T} : \epsilon - 2\epsilon' < \|t - (z_i + z_i')\|_{\mathbb{T}} < \epsilon\}$ such that

$$\int \psi \ll \epsilon', \quad \sum_r |\widehat{\psi}(r)||r| \ll_{\epsilon'} 1. \tag{4.21}$$

Then the number of squares satisfying (4.20) is bounded above by

$$\sum_{n \le 2N^{1/2}} \psi(\theta_i n^2) = \sum_{r \in \mathbb{Z}} \widehat{\psi}(r) \sum_{n \le 2N^{1/2}} e(r\theta_i n^2).$$

The term with $r = 0$ is $2N^{1/2}(\int \psi) \ll \epsilon' N^{1/2}$. By Corollary 4.3.2 (applied with $d = 1$) the contribution from the terms with $r \neq 0$ is

$$\ll N^{1/2} \mathcal{F}(M)^{-1/C_2} \sum_{r \neq 0} |\widehat{\psi}(r)||r|.$$

By (4.21) this is $\ll_{\epsilon'} N^{1/2}\mathcal{F}(M)^{-1/C_2}$ which, in view of (4.10), is $O(\epsilon' N^{1/2})$ provided $\mathcal{F}(M)$ grows sufficiently rapidly. Thus the total number of n satisfying (4.20) for some $i \in \{1, \dots, d\}$ is $O(\epsilon' d N^{1/2})$, which is bounded as claimed by the choice of ϵ'. □

To complete the proof of Proposition 4.4.1 it suffices to show that $A + A$ contains all but $\ll \delta(2\epsilon)^{d+1} q^{-1} \mathfrak{S}(u + u', q) N^{1/2}$ of the squares in Y. Indeed if δ is chosen small enough then this will be $\le \eta(2\epsilon)^{d+1} q^{-1} \mathfrak{S}(u + u', q) N^{1/2}$, the bound claimed. Let $S \subset Y$ be the set of all squares in Y which are not in $A + A$; thus it suffices to establish the bound

$$|S| \ll \delta(2\epsilon)^{d+1} q^{-1} \mathfrak{S}(u + u', q) N^{1/2}. \tag{4.22}$$

Recall the definitions (4.13), (4.14) of X, X'. We will need to introduce smoothed approximants χ, χ' to the characteristic functions of X, X' respectively, with the following properties.

1. χ is a minorant for X, that is to say $0 \le \chi(n) \le 1_X(n)$ for all n;
2. χ' is a minorant for X', that is to say $0 \le \chi'(n) \le 1_X(n)$ for all n;
3. $\chi(n) = 1$ on the set $\{n \in \mathbb{N} : n \equiv u (\mathrm{mod}\ q), |\frac{n}{N} - x|, \|\theta n - z\|_{\mathbb{T}^d} \le \epsilon - \epsilon'\}$;
4. $\int_0^1 |\widehat{\chi}(t)| dt, \int_0^1 |\widehat{\chi'}(t)| dt = O_M(1)$;
5. $\sum_n \chi'(n) \gg (2\epsilon')^{d+1} q^{-1} N$.

Such a function is constructed in Lemma C.3 (which must be applied twice, once with parameter ϵ and once with parameter ϵ').

In particular it follows from (4.16) that

$$\sum_n |f_{\mathrm{sml}}\chi(n)| \ll \delta(2\epsilon)^{d+1}q^{-1}N, \qquad \sum_n |f_{\mathrm{sml}}\chi'(n)| \ll \delta(2\epsilon')^{d+1}q^{-1}N. \tag{4.23}$$

Our assumption that $A+A$ is disjoint from S implies that

$$\sum_{n\in S}(1_A\chi * 1_A\chi')(n) = 0. \tag{4.24}$$

To investigate this expression, we use the decomposition from the regularity lemma,

$$1_A = f_{\mathrm{tor}} + f_{\mathrm{sml}} + f_{\mathrm{unf}}.$$

The left-hand side of (4.24) may then be expanded as a sum of 9 terms

$$T_{\bullet,\bullet'} := \sum_{n\in S}(f_\bullet\chi * f_{\bullet'}\chi')(n),$$

where $\bullet, \bullet' \in \{\mathrm{tor}, \mathrm{sml}, \mathrm{unf}\}$. Thus

$$|T_{\mathrm{tor,tor}}| \leq \sum_{(\bullet,\bullet')\neq(\mathrm{tor,tor})} |T_{\bullet,\bullet'}|. \tag{4.25}$$

We analyse these 9 terms $T_{\bullet,\bullet'}$ separately, beginning with the "main term" $T_{\mathrm{tor,tor}}$.

Writing

$$f_{\mathrm{tor}}(n) = F(n(\mathrm{mod}\ q), \frac{n}{N}, \theta n),$$

we may expand $T_{\mathrm{tor,tor}}$ as

$$\sum_{n\in S}\sum_m F(m(\mathrm{mod}\ q), \frac{m}{N}, \theta m)\chi(m)$$
$$\times F(n-m(\mathrm{mod}\ q), \frac{n-m}{N}, \theta(n-m))\chi'(n-m).$$

Since $\chi(m)$ is supported where $m \equiv u(\mathrm{mod}\ q)$ and $|\frac{m}{N} - x|, \|\theta m - z\|_{\mathbb{T}^d} \leq \epsilon$, and since F is M-Lipschitz, we have using (4.11) that

$$F(m(\mathrm{mod}\ q), \frac{m}{N}, \theta m)\chi(m) = (F(u,x,z) + O(M\epsilon))\chi(m) \geq \frac{1}{80}\chi(m)$$

if δ is sufficiently small (note, recalling the definition (4.8) of ϵ, that $M\epsilon = \delta$). Similarly,

$$F(n-m(\mathrm{mod}\ q), \frac{n-m}{N}, \theta(n-m))\chi'(n-m) \geq \frac{1}{80}\chi'(n-m).$$

It follows that

$$T_{\mathrm{tor,tor}} \gg \sum_{n\in S}\sum_{m} \chi(m)\chi'(n-m)$$
$$= \sum_{n\in S}\sum_{m} \chi(n-m)\chi'(m). \tag{4.26}$$

Recall the definition (4.18) of $Y_- \subset Y$. If $n \in Y_-$ and $m \in \mathrm{Supp}(\chi') \subset X'$ then $n-m \equiv u (\mathrm{mod}\ q)$ and $|\frac{n-m}{N} - x|, \|\theta(n-m) - z\|_{\mathbb{T}^d} \leq \epsilon - \epsilon'$, and therefore by property (3) of χ we have $\chi(n-m) = 1$. It follows from these observations, (4.26) and point (5) of the properties of χ, χ' that

$$T_{\mathrm{tor,tor}} \gg \sum_{n\in S\cap Y_-}\sum_{m} \chi(n-m)\chi'(m)$$
$$\gg |S\cap Y_-| \sum_{m} \chi'(m)$$
$$\gg |S\cap Y_-|(2\epsilon')^{d+1}q^{-1}N. \tag{4.27}$$

We set this estimate aside for later use.

Next we look at the terms $T_{\bullet,\bullet'}$ in which $\bullet' = \mathrm{sml}$. Here we require the *a priori* bound

$$|S| \ll (2\epsilon)^{d+1}q^{-1}\mathfrak{S}(u+u',q)N^{1/2}. \tag{4.28}$$

This is, of course, weaker than the result we are trying to prove, but it follows immediately from Lemma 4.4.5. All of these terms $T_{\bullet,\mathrm{sml}}$ have the form

$$T_{\bullet,\mathrm{sml}} = \sum_{n\in S}(g * f_{\mathrm{sml}}\chi')(n) = \sum_{n\in S}\sum_{m} g(n-m)f_{\mathrm{sml}}\chi'(m),$$

where g is some function bounded pointwise by 1. Thus

$$|T_{\bullet,\mathrm{sml}}| \leq |S| \sum_{m} |f_{\mathrm{sml}}\chi'(m)|$$

and so, by (4.28) and (4.23),

$$T_{\bullet,\mathrm{sml}} \ll \delta(4\epsilon\epsilon')^{d+1}q^{-2}\mathfrak{S}(u+u',q)N^{3/2}. \tag{4.29}$$

Next we turn to the bounding of

$$T_{\mathrm{sml,tor}} = \sum_{n\in S}(f_{\mathrm{sml}}\chi * f_{\mathrm{tor}}\chi')(n).$$

This expands as

$$\sum_{n\in S}\sum_{m} f_{\mathrm{sml}}\chi(n-m)F(m(\mathrm{mod}\ q), \frac{m}{N}, \theta m)\chi'(m).$$

By the Lipschitz property of F and the fact that χ' is supported on X', this is

$$F(u',x',z')\sum_{\substack{m\\ n\in S}} f_{\mathrm{sml}}\chi(n-m)\chi'(m) + O(\epsilon' M)\sum_{\substack{m\\ n\in S}} |f_{\mathrm{sml}}\chi(n-m)|\chi'(m).$$

Since $\epsilon' < \epsilon < 1/M$, it follows that

$$T_{\mathrm{sml,tor}} \ll \sum_{n\in S}\sum_{m} |f_{\mathrm{sml}}\chi(n-m)|\chi'(m) = \sum_{n',m} |f_{\mathrm{sml}}\chi(n')|\chi'(m)1_S(n'+m).$$

By (4.23), this is

$$\ll \delta(2\epsilon)^{d+1}q^{-1}N \sup_{n'\in \mathrm{Supp}\,\chi}\sum_{m}\chi'(m)1_S(n'+m).$$

By Lemma 4.4.6 and the fact that $\mathrm{Supp}\,\chi \subset X$, $\mathrm{Supp}\,\chi' \subset X'$, we conclude that

$$T_{\mathrm{sml,tor}} \ll \delta(4\epsilon\epsilon')^{d+1}q^{-2}\mathfrak{S}(u+u',q)N^{3/2}. \tag{4.30}$$

In all of the remaining terms $T_{\bullet,\bullet'}$ that we have yet to bound, at least one of $\bullet, \bullet'$ is unf. If $\bullet = \mathrm{unf}$ then such a term has the form

$$T_{\mathrm{unf},\bullet'} = \sum_{n\in S}(f_{\mathrm{unf}}\chi * g)(n),$$

where g is some function bounded pointwise by 1. This may be written in Fourier space as

$$\int_0^1 \widehat{f_{\mathrm{unf}}\chi}(t)\widehat{g}(t)\widehat{1}_S(t)dt,$$

where g is a bounded function. By Hölder's inequality, the right-hand side here is bounded above by

$$\|\widehat{f_{\mathrm{unf}}\chi}\|_\infty^{1/3}\Big(\int_0^1 |\widehat{f_{\mathrm{unf}}\chi}|^2\Big)^{1/3}\Big(\int_0^1 |\widehat{g}|^2\Big)^{1/2}\Big(\int_0^1 |\widehat{1}_S|^6\Big)^{1/6}. \tag{4.31}$$

By Parseval's identity and the boundedness of $f_{\mathrm{unf}}, g, \chi$ we have

$$\int_0^1 |\widehat{f_{\mathrm{unf}}\chi}|^2, \int_0^1 |\widehat{g}|^2 \ll N, \tag{4.32}$$

and Proposition 4.3.3 tells us that

$$\int_0^1 |\widehat{1}_S(t)|^6 dt \ll N^2.$$

Finally, we note that

$$\widehat{f_{\mathrm{unf}}\chi}(t) = \int_0^1 \widehat{f_{\mathrm{unf}}}(t')\widehat{\chi}(t-t')dt',$$

and so by property (4) of χ we have

$$\|\widehat{f_{\mathrm{unf}}\chi}\|_\infty \leq \|\widehat{f_{\mathrm{unf}}}\|_\infty \|\widehat{\chi}\|_1 \ll_M N\mathcal{F}(M)^{-1}.$$

Combining all these estimates together gives

$$T_{\mathrm{unf},\bullet'} = \sum_n (f_{\mathrm{unf}}\chi * g)(n)1_S(n) \ll_M N^{3/2}\mathcal{F}(M)^{-1/3}.$$

If the growth of $\mathcal{F}$ is sufficiently rapid, we obtain in view of the fact that $d, q \leq M$, $\epsilon = \delta/M$ and (4.10) that

$$T_{\mathrm{unf},\bullet'} \ll \delta(4\epsilon\epsilon')^{d+1}q^{-2}N^{3/2}. \tag{4.33}$$

An almost identical argument (relying instead on the bound $\|\chi'\|_1 = O_M(1)$) yields

$$T_{\bullet,\mathrm{unf}} \ll \delta(4\epsilon\epsilon')^{d+1}q^{-2}N^{3/2}. \tag{4.34}$$

Combining (4.27), (4.29), (4.30), (4.33) and (4.34) with (4.25) we obtain

$$|S \cap Y_-|(2\epsilon')^{d+1}q^{-1}N \ll \delta(4\epsilon\epsilon')^{d+1}q^{-2}\mathfrak{S}(u+u',q)N^{3/2},$$

and therefore

$$|S \cap Y_-| \ll \delta(2\epsilon)^{d+1}q^{-1}\mathfrak{S}(u+u',q)N^{1/2}.$$

Lemma 4.4.7 provides the bound

$$|S \cap (Y \setminus Y_-)| \ll \delta(2\epsilon)^{d+1}q^{-1}\mathfrak{S}(u+u',q)N^{1/2}.$$

Combining this with the preceding yields

$$|S| \ll \delta(2\epsilon)^{d+1}q^{-1}\mathfrak{S}(u+u',q)N^{1/2},$$

which is exactly (4.22). This completes the proof of Proposition 4.4.1.

4.5 The square-root of a Bohr set

Suppose that $\mathbb{N}$ is partitioned into two colour classes V and W, neither of which has a monochromatic solution to $x+y=z^2$. The main result of the last section, Proposition 4.4.1, shows that if $V \cap [N, 2N)$ has size at least $N/2$ then $V+V$ contains almost all of the squares in a "Bohr set" $\Lambda := \{n \in \mathbb{N} : n \equiv b(\mathrm{mod}\ q), |\frac{n}{N} - x|, \|\theta n - z\|_{\mathbb{T}^d} \leq \epsilon\}$. This means that most of $\sqrt{\Lambda}$ must lie in W. In this section we examine the additive properties of such square roots $\sqrt{\Lambda}$. (Recall that $\sqrt{\Lambda}$ is by definition the set of *integers* n such that $n^2 \in \Lambda$.)

Here is the main result of the section.

Proposition 4.5.1. *Let $\eta > 0$. Then there is a function $\Omega : \mathbb{N}^3 \to \mathbb{R}_+$ with the following property. Suppose we have $q, d \in \mathbb{N}$, $\epsilon > 0$, $x \in [0,3]$, $\theta, z \in \mathbb{T}^d$ and $N \in \mathbb{N}$. Suppose that θ is $(\Omega(q,d,1/\epsilon), N)$-irrational. Suppose that b is a square modulo q and set*

$$Y := \{n \in \mathbb{N} : n \equiv b (\mathrm{mod}\ q), |\frac{n}{N} - x|, \|\theta n - z\|_{\mathbb{T}^d} \leq \epsilon\}.$$

Let $Y' \subset Y$ be a set containing all but at most $\eta(2\epsilon)^{d+1}q^{-1}\mathfrak{S}(b,q)N^{1/2}$ of the squares in Y. Then, for all but at most $O(\eta\epsilon q^{-1}N^{1/4})$ of the elements $t \in Q$, where

$$Q := \mathbb{P}(\big[(2x)^{1/4} - \frac{\epsilon}{100}, (2x)^{1/4} + \frac{\epsilon}{100}\big]; N^{1/4}, q), \tag{4.35}$$

we have $t^2 \in \sqrt{Y'} + \sqrt{Y'}$.

(Recall that $\mathbb{P}(I; N, q) := \{n \in \mathbb{Z} : n/N \in I, q|n\}$.)

The proof of this is a little complicated so we break it down into a few lemmas. We have $\sqrt{Y} = \bigcup_{a \in \mathcal{A}} Z_+^a \cup Z_-^a$, where

$$\begin{aligned} Z_\pm^a := \{n \in \mathbb{N} : n \equiv \pm a (\mathrm{mod}\ q), &(x-\epsilon)^{1/2}N^{1/2} \leq n \leq (x+\epsilon)^{1/2}N^{1/2}, \\ &\|\theta n^2 - z\|_{\mathbb{T}^d} \leq \epsilon\}, \end{aligned} \tag{4.36}$$

and $\mathcal{A}$ is the set of square roots of b in $\mathbb{Z}/q\mathbb{Z}$. Define

$$\widetilde{Z}_\pm^a := \sqrt{Y'} \cap Z_\pm^a;$$

then

$$\sum_{a \in \mathcal{A}} |Z_\pm^a \setminus \widetilde{Z}_\pm^a| \ll \eta(2\epsilon)^{d+1}q^{-1}\mathfrak{S}(b,q)N^{1/2},$$

by assumption. It follows that there is some $a \in \mathcal{A}$ such that

$$|Z_\pm^a \setminus \widetilde{Z}_\pm^a| \ll \eta(2\epsilon)^{d+1}q^{-1}N^{1/2}. \tag{4.37}$$

Henceforth, we fix this value of a and write $Z_\pm = Z_\pm^a$ for brevity. To orient ourselves we remark that, if Ω grows sufficiently rapidly then one could prove that

$$|Z_\pm| \asymp (2\epsilon)^{d+1}q^{-1}N^{1/2}.$$

We will not need to explicitly prove any statement of this kind separately.

Lemma 4.5.2. *Suppose that $n_+ \in Z_+$. Then*

$$\#\{n_- \in Z_- : n_- + n_+ = q^2m^2 \text{ for some } m \in \mathbb{Z}\} \ll (2\epsilon)^{d+1}q^{-1}N^{1/4},$$

the implied constant being uniform in n_+ and independent of a (recall that $Z_\pm$ depends on a). Similarly, if $n_- \in Z_-$ then

$$\#\{n_+ \in Z_+ : n_- + n_+ = q^2 m^2 \text{ for some } m \in \mathbb{Z}\} \ll (2\epsilon)^{d+1} q^{-1} N^{1/4},$$

the implied constant being uniform in n_- and in a.

Proof. The quantity we are interested in can be written as

$$\sum_{m \in I(n_+)} 1_{\|\theta(q^2m^2 - n_+)^2 - z\|_{\mathbb{T}^d} \leq \epsilon},$$

where $I(n_+)$ is the interval

$$\frac{1}{q}\big((x-\epsilon)^{1/2}N^{1/2} + n_+\big)^{1/2} \leq m \leq \frac{1}{q}\big((x+\epsilon)^{1/2}N^{1/2} + n_+\big)^{1/2},$$

the cardinality of which satisfies

$$|I(n_+)| \ll \epsilon q^{-1} N^{1/4} \tag{4.38}$$

uniformly in n_+. To bound this above, take a majorant ψ_ϵ^+ to the unit ball $B_\epsilon(0) \subset \mathbb{T}^d$, as in Lemma C.2. Then our quantity is at most

$$\sum_{m \in I(n_+)} \psi_\epsilon^+(\theta(q^2m^2 - n_+)^2 - z).$$

Fourier expanding ψ_ϵ^+, this is

$$\sum_{\mathbf{r} \in \mathbb{Z}^d} \widehat{\psi_\epsilon^+}(\mathbf{r}) \sum_{m \in I(n_+)} e(q^4 \mathbf{r} \cdot \theta m^4 + \ldots),$$

where the dots denote terms of degree at most 2 in m (which can depend on $\mathbf{r}, n_+, \theta, z, q$). The contribution from $\mathbf{r} = 0$ is $|I(n_+)|(\int \psi_\epsilon^+)$ which, by (4.38) and property (1) of Lemma C.2, is $\ll (2\epsilon)^{d+1} q^{-1} N^{1/4}$. By Corollary 4.3.2 (and since $|I(n_+)| \leq N^{1/4}$), we have

$$\Big|\sum_{m \in I(n_+)} e(q^4 \mathbf{r} \cdot \theta m^4 + \ldots)\Big| \leq N^{1/4} \left(\frac{q^4 \|\mathbf{r}\|_1}{\Omega(q, d, 1/\epsilon)}\right)^{1/C_4}.$$

By Lemma C.2 (2), the contribution from $\mathbf{r} \neq 0$ is therefore

$$\begin{aligned} &\ll N^{1/4} \left(\frac{q^4}{\Omega(q, d, 1/\epsilon)}\right)^{1/C_4} \sum_{\mathbf{r} \in \mathbb{Z}^d \setminus \{0\}} |\widehat{\psi_\epsilon^+}(\mathbf{r})| \|\mathbf{r}\|_1 \\ &\ll_{\epsilon, d} N^{1/4} \left(\frac{q^4}{\Omega(q, d, 1/\epsilon)}\right)^{1/C_4}, \end{aligned}$$

which is also $\ll (2\epsilon)^{d+1} q^{-1} N^{1/4}$ if Ω is chosen appropriately. □

Define progressions P_+, P_- by

$$P_\pm := \{n \in N : n \equiv \pm a (\mathrm{mod}\ q), (x-\epsilon)^{1/2}N^{1/2} \leq n \leq (x+\epsilon)^{1/2}N^{1/2}\}, \tag{4.39}$$

and recall from the statement of Proposition 4.5.1 the definition of Q, viz.

$$Q := \mathbb{P}([(2x)^{1/4} - \frac{\epsilon}{100}, (2x)^{1/4} + \frac{\epsilon}{100}]; N^{1/4}, q).$$

Observe that if $t \in Q$ then t^2 is a sum $p_+ + p_-$ in $\gg \epsilon q^{-1}N^{1/2}$ ways. Indeed

$$\Big((2x)^{1/2} - \frac{\epsilon}{10}\Big)N^{1/2} < t^2 < \Big((2x)^{1/2} + \frac{\epsilon}{10}\Big)N^{1/2}$$

and $t^2 \equiv 0 (\mathrm{mod}\ q)$, hence for any of the $\gg \epsilon q^{-1}N^{1/2}$ values of p_+ with $(x^{1/2} - \frac{\epsilon}{10})N^{1/2} < p_+ < (x^{1/2} + \frac{\epsilon}{10})N^{1/2}$ and $p_+ \equiv a (\mathrm{mod}\ q)$ we have $t^2 - p_+ \in P_-$.

Note that from (4.36) and (4.39) we have

$$Z_\pm = \{n \in P_\pm : \|\theta n^2 - z\|_{\mathbb{T}^d} \leq \epsilon\}. \tag{4.40}$$

This suggests the intuition behind the arguments that follow, which is that $Z_\pm$ behaves like a "pseudorandom" subset of $P_\pm$ of density $(2\epsilon)^d$. Thus it is reasonable to expect that a typical t^2, $t \in Q$, will have $\gg (2\epsilon)^{2d+1}q^{-1}N^{1/2}$ representations as $z_+ + z_-$ with $z_+ \in Z_+$, $z_- \in Z_-$.

Lemma 4.5.3. *Suppose that Ω grows sufficiently rapidly. Write $r(n)$ for the number of representations of n as $z_+ + z_-$ with $z_\pm \in Z_\pm$. Suppose that Ω grows fast enough. Then all but at most $\eta\epsilon q^{-1}N^{1/4}$ of elements $t \in Q$ have $r(t^2) \gg (2\epsilon)^{2d+1}q^{-1}N^{1/2}$.*

Proof. If the lemma is false then for any absolute constant c (which we may specify later) there is a set $T \subset Q$, $|T| \geq \eta\epsilon q^{-1}N^{1/4}$, such that

$$\sum_{t \in T} r(t^2) \leq c(2\epsilon)^{2d+1}q^{-1}|T|N^{1/2}. \tag{4.41}$$

We first introduce a smoothed variant of r, defined by

$$\widetilde{r}(n) = f_+ * f_-(n),$$

where

$$f_\pm(n) = 1_{P_\pm}(n)\psi_\epsilon^-(\theta n^2 - z),$$

where ψ_ϵ^- is a suitable minorant to $B_\epsilon(0)$, as constructed in Lemma C.2. From (4.40) we see that $1_{Z_\pm} \geq f_\pm$ pointwise, and so

$$r(n) \geq \widetilde{r}(n)$$

pointwise. Define

$$g_{\pm}(n) = 1_{P_{\pm}}(n)\left(\psi_{\epsilon}^{-}(\theta n^2 - z) - \int \psi_{\epsilon}^{-}\right).$$

Fourier expanding ψ_{ϵ}^{-}, we see that

$$\widehat{g_{\pm}}(t) = \sum_{\mathbf{r}\in\mathbb{Z}^d\setminus\{0\}} \widehat{\psi_{\epsilon}^{-}}(\mathbf{r}) \sum_{n\in P_+} e(\mathbf{r}\cdot\theta n^2 + nt - \mathbf{r}\cdot z).$$

Parametrising $n \in P_+$ as $n = qm + b$ for m in some interval I with $|I| = |P_+| < N^{1/2}$, it follows from Corollary 4.3.2 that the inner sum is $\ll N^{1/2}\Omega(q,d,1/\epsilon)^{-1/C_2}\|\mathbf{r}\|_1$. Therefore, by property (2) of Lemma C.2, we have

$$\begin{aligned}\|\widehat{g_{\pm}}\|_{\infty} &\ll N^{1/2}\Omega(q,d,1/\epsilon)^{-1/C_2}\sum_{\mathbf{r}\in\mathbb{Z}^d}|\widehat{\psi_{\epsilon}^{-}}(\mathbf{r})|\|\mathbf{r}\|_1 \\ &\ll_{\epsilon,d} N^{1/2}\Omega(q,d,1/\epsilon)^{-1/C_2}. \end{aligned} \tag{4.42}$$

Now, writing

$$f_{\pm} = 1_{P_{\pm}}\int \psi_{\epsilon}^{-} + g_{\pm},$$

we may expand $\sum_{t\in T}\widetilde{r}(t^2)$ as a sum of four terms. The "main term" is

$$E_{\text{main}} = (\int \psi_{\epsilon}^{-})^2 \sum_{t\in T} 1_{P_+} * 1_{P_-}(t^2).$$

The three error terms each have the shape

$$E_{\text{error}} = \sum_{t\in T} g_{\pm} * h_{\mp}(t^2),$$

where $h_{\mp}$ is bounded pointwise by 1 and supported on $P_{\mp}$.

We have already remarked that if $t \in Q$ then t^2 has $\gg \epsilon q^{-1}N^{1/2}$ representations as $p_+ + p_-$, and therefore

$$E_{\text{main}} \gg (2\epsilon)^{2d}\cdot|T|\cdot\epsilon q^{-1}N^{1/2} \gg \eta(2\epsilon)^{2d+2}q^{-2}N^{3/4}. \tag{4.43}$$

On the other hand

$$E_{\text{error}} = \int_0^1 \widehat{g_{\pm}}(\theta)\widehat{h_{\mp}}(\theta)\widehat{1_{T^2}}(\theta)d\theta,$$

where here $T^2 := \{t^2 : t \in T\}$. Using the same application of Hölder's inequality as in (4.31),

$$E_{\text{error}} \ll \|\widehat{g_{\pm}}\|_{\infty}^{1/3}\Big(\int_0^1 |\widehat{g_{\pm}}|^2\Big)^{1/3}\Big(\int_0^1 |\widehat{h_{\mp}}|^2\Big)^{1/2}\Big(\int_0^1 |\widehat{1}_{T^2}|^6\Big)^{1/6}.$$

By Parseval and the crude bound $|P_\pm| \ll N^{1/2}$ we have

$$\int_0^1 |\widehat{g_\pm}|^2, \int_0^1 |\widehat{h_\mp}|^2 \ll N^{1/2}.$$

Proposition 4.3.3 tells us that

$$\int_0^1 |\widehat{1}_{T^2}|^6 \ll N.$$

Putting this together with (4.42) gives

$$E_{\text{error}} \ll \Omega(q, d, 1/\epsilon)^{-1/3C_2} N^{3/4}.$$

Choosing Ω to grow sufficiently quickly, we see from (4.43) that this can be made less than $\frac{1}{10}$ of E_{main}. It follows from (4.43) that

$$\sum_{t \in T} \widetilde{r}(t^2) \geq E_{\text{main}} - 3E_{\text{error}} > \frac{1}{2} E_{\text{main}} \gg (2\epsilon)^{2d+1} q^{-1} |T| N^{1/2},$$

contrary to (4.41) if c was chosen small enough. □

Finally we put Lemmas 4.5.2 and 4.5.3 together to establish Proposition 4.5.1. It is certainly enough (in view of the definitions of $\widetilde{Z}_\pm$) to show that $\widetilde{Z}_+ + \widetilde{Z}_-$ contains t^2 for all but at most $O(\eta\epsilon q^{-1} N^{1/4})$ of the elements $t \in Q$. By Lemma 4.5.3, all but at most $\eta\epsilon q^{-1} N^{1/4}$ elements $t \in Q$ are such that t^2 is *well-represented* in $Z_+ + Z_-$, by which we mean that $r(t^2) \gg (2\epsilon)^{2d+1} q^{-1} N^{1/2}$, where $r(t^2)$ is the number of representations of t^2 as $z_+ + z_-$. Suppose now that we pass from $Z_\pm$ to $\widetilde{Z}_\pm$. The number of pairs (z_+, z_-) with $z_+ + z_-$ the square of an element in Q that are lost in this way is, by Lemma 4.5.2, bounded above by $\ll |Z_\pm \setminus \widetilde{Z}_\pm| (2\epsilon)^{d+1} q^{-1} N^{1/4}$. By (4.37), this is bounded by $\ll \eta (2\epsilon)^{2d+2} q^{-2} N^{3/4}$. The number of t for which t^2 is well-represented but does not lie in $\widetilde{Z}_+ + \widetilde{Z}_-$ is therefore bounded above by

$$\ll \frac{\eta (2\epsilon)^{2d+2} q^{-2} N^{3/4}}{(2\epsilon)^{2d+1} q^{-1} N^{1/2}} = O(\eta\epsilon q^{-1} N^{1/4}).$$

This completes the proof of Proposition 4.5.1.

4.6 Gaps between sums of two squares

In this section we prove a result, Proposition 4.6.1, that we will need in the next section. It seems possible that such a result appears in the literature already, but we do not know a reference. We prove a slightly more general result than we actually need since this is plausibly of independent interest.

Proposition 4.6.1. *Let $\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2$ be nonnegative reals with $\alpha_1 < \beta_1$, $\alpha_2 < \beta_2$, $\alpha_1^2 + \alpha_2^2 < \gamma_1 < \gamma_2 < \beta_1^2 + \beta_2^2$. Let $q \in \mathbb{N}$ and set $P_i := \mathbb{P}([\alpha_i, \beta_i]; N, q)$ for $i = 1, 2$. Suppose that $\gamma_1 \leq n/N^2 \leq \gamma_2$. Then there are $n_1 \in P_1$, $n_2 \in P_2$ such that*

$$|n_1^2 + n_2^2 - n| \ll \sqrt{N}.$$

The implied constant may depend on $\alpha_i, \beta_i, \gamma_i, q$ but is independent of n and N.

Remark. A well-studied case is that in which $P_1 = P_2 = \{1, \dots, N\}$. Then it is well-known that there is a sum of two squares $n_1^2 + n_2^2$ within $O(N^{1/2})$ of any $n \leq N^2$. One argument to prove this is very simple: take $n_1 = \lfloor \sqrt{n} \rfloor$, noting that $|n - n_1^2| \ll N$, and then set $n_2 := \lfloor \sqrt{n - n_1^2} \rfloor$. No bound of the form $o(N^{1/2})$ is known, a problem Montgomery [51, Problem 64, p. 208] attributes to Littlewood. The argument just sketched does not adapt to our case since the n_2 produced is necessarily very small. However, there is another type of argument giving a similar bound and allowing us to take $n_1 \approx n_2$. The idea here is to take $n_1(k) = \lfloor \sqrt{n/2} \rfloor + k$, $n_2(k) = \lfloor \sqrt{n/2} \rfloor - k$, where $k \in \mathbb{Z}$ is to be specified later. Observe that

$$n_1(k)^2 + n_2(k)^2 = 2\lfloor \sqrt{n/2} \rfloor^2 + 2k^2,$$

and so in particular

$$n_1(0)^2 + n_2(0)^2 \leq n,$$
$$n_1(k)^2 + n_2(k)^2 \geq n - 2\sqrt{n} + 2k^2 > n$$

for $k = \lceil \sqrt{n} \rceil$ and

$$(n_1(k+1)^2 + n_2(k+1)^2) - (n_1(k)^2 - n_2(k))^2 = 4k + 2 \ll \sqrt{n}$$

uniformly for $k \leq \lceil \sqrt{n} \rceil$. It follows from the "discrete intermediate value theorem" that there is some k for which $|n_1(k)^2 + n_2(k)^2 - n| \ll \sqrt{n}$.

It turns out that this argument *does* generalise to allow us to prove Proposition 4.6.1.

Proof. For the duration of this proof, the implied constant in the $O()$ and $\ll, \gg$ notations may depend on $\alpha_i, \beta_i, \gamma_i, q$. We may clearly assume that N is sufficiently large.

For each $\gamma \in [\gamma_1, \gamma_2]$, define I_γ to be the set of all $\lambda \in \mathbb{R}$ for which there exist $t_1, t_2 \in \mathbb{R}$ with $\alpha_1 \leq t_1 \leq \alpha_2$, $\beta_1 \leq t_2 \leq \beta_2$, $t_1/t_2 = \lambda$ and $t_1^2 + t_2^2 = \gamma$. Let $\widetilde{I}_\gamma$ be the middle half of I_γ. It is easy to see that I_γ is a closed interval whose length is positive and varies continuously as a function of γ, and is therefore bounded below uniformly in γ. The same is true for $\widetilde{I}_\gamma$. This implies that

1. There is an absolute $\epsilon \gg 1$ such that if $\lambda \in \widetilde{I}_\gamma$ then we may find t_1, t_2 with $t_1/t_2 = \lambda$ and
$$\alpha_i + \epsilon \leq t_i \leq \beta_i - \epsilon; \tag{4.44}$$

2. $\widetilde{I}_\gamma$ contains a rational $a(\gamma)/b(\gamma)$ with $a(\gamma), b(\gamma) = O(1)$ and neither $a(\gamma)$ nor $b(\gamma)$ zero.

Now suppose that n is given satisfying $\gamma_1 \leq n/N^2 \leq \gamma_2$. Set $\gamma := n/N^2$, and select rationals $a = a(\gamma)$, $b = b(\gamma)$, not both zero, as in (2) above. According to (1), there are t_1, t_2 with $t_1^2 + t_2^2 = \gamma$, $t_1/t_2 = a/b$ and such that (4.44) is satisfied.

Now set
$$n_1(k) := q\left\lfloor \frac{t_1 N}{q} \right\rfloor + qkb, n_2(k) := q\left\lfloor \frac{t_2 N}{q} \right\rfloor - qka.$$
Evidently $q | n_1(k), n_2(k)$. Moreover from (4.44) it follows that $\alpha_i \leq n_i(k)/N \leq \beta_i$ provided $|k| \leq cN$ for suitably small $c \gg 1$. Therefore for k in this range we have $n_i(k) \in P_i$. Observe that
$$n_1(0)^2 + n_2(0)^2 \leq (t_1^2 + t_2^2)N^2 = n.$$
Also
$$\begin{aligned} &n_1(k)^2 + n_2(k)^2 \\ &\quad = q^2\Big(\left\lfloor \frac{t_1 N}{q} \right\rfloor^2 + \left\lfloor \frac{t_2 N}{q} \right\rfloor^2 + 2k(a\left\{ \frac{t_2 N}{q} \right\} - b\left\{ \frac{t_1 N}{q} \right\}) + k^2(a^2 + b^2)\Big) \\ &\quad \geq n - O(N) - O(k) + q^2k^2(a^2 + b^2), \end{aligned} \tag{4.45}$$
and in particular
$$n_1(k)^2 + n_2(k)^2 > n$$
for some $k = O(\sqrt{N})$.

Moreover, from (4.45) again we have
$$\left|(n_1(k+1)^2 + n_2(k+1)^2) - (n_1(k)^2 - n_2(k))^2\right| = O(k).$$
It follows from these properties and a discrete intermediate value argument that there is some $k = O(\sqrt{N})$ for which $|n_1(k)^2 + n_2(k)^2 - n| \ll \sqrt{N}$. The result follows. □

4.7 Proof of the main theorem

In Proposition 4.7.2 below we will synthesize the main results of Sections 4.4 and 4.5, together with the following small (and well-known) lemma.

Lemma 4.7.1. *Let $Q \subset \mathbb{N}$ be a finite arithmetic progression of size at least 100, and suppose that $S \subset Q$ is a set of size at least $\frac{9}{10}|Q|$. Then $S+S$ contains a subprogression of $Q+Q$ of size at least $|Q|$ with the same common difference as Q.*

Proof. By translating we may assume that $Q = \{1, \dots, m\}$. Suppose that $x \leq m$. Then the pairs $\{j, x-j\}$, $1 \leq j < x/2$, are disjoint. If $S+S$ does not contain x, then S cannot contain both elements of any such pair, and hence $|Q \setminus S| \geqslant \lfloor x/2 \rfloor$. Therefore $\lfloor x/2 \rfloor \leq \frac{m}{10}$, and so $x \leq \frac{m}{5} + 2$. A similar argument holds for $x \geq m$, with the conclusion now being that $2m - x \leq \frac{m}{5} + 2$. Thus $S+S$ contains the progression $\frac{m}{5} + 2 < x < 2m - \frac{m}{5} - 2$. This is more than m elements if $m \geq 100$. □

Proposition 4.7.2. *Suppose that $A \subset [N, 2N)$ is a set of size at least $N/2$. Then $\sqrt{2\sqrt{2\sqrt{2A}}}$ contains a progression $\mathbb{P}(I; N^{1/8}, q)$ for some interval $I \subset [0.1, 10]$ with $|I| \gg 1$ and for some $q = O(1)$.*

Proof. Let $\eta > 0$ be a quantity to be specified later. Let $\Omega : \mathbb{N}^3 \to \mathbb{R}_+$ be the growth function appearing in the statement of Proposition 4.5.1. Apply Proposition 4.4.1 with this function. Let $q, d, \epsilon, \theta, z, b$ be as in the conclusion of that proposition. Taking Y as in the statement of Proposition 4.5.1, Proposition 4.4.1 then tells us that $Y' := (A+A) \cap Y = 2A \cap Y$ satisfies the hypotheses of Proposition 4.5.1. It follows that $2\sqrt{Y'}$, and hence $2\sqrt{2A}$, contains t^2 for all but at most $O(\eta \epsilon q^{-1} N^{1/4})$ values of $t \in Q = \mathbb{P}([(2x)^{1/4} - \frac{\epsilon}{100}, (2x)^{1/4} + \frac{\epsilon}{100}]; N^{1/4}, q)$. Therefore $\sqrt{2\sqrt{2A}}$ contains all but at most $O(\eta \epsilon q^{-1} N^{1/4})$, and therefore at least $(1 - C\eta)|Q|$, of the elements of Q. If η is chosen suitably, this is at least $\frac{9}{10}|Q|$ elements of Q, and so by Lemma 4.7.1 we see that $2\sqrt{2\sqrt{2A}}$ contains a subprogression $Q' \subset Q$ of the form $Q' = \mathbb{P}(I; N^{1/4}, q)$ with $|I| \gg \epsilon$. Finally, note that $\sqrt{Q'}$ contains a progression of the form $\mathbb{P}(I'; N^{1/8}, q)$ for some $I' \subset [0.1, 10]$ with $|I| \gg \epsilon$. □

We are finally ready to complete the proof of Theorem 4.1.1. Suppose we have a 2-colouring $V \cup W$ of all sufficiently large positive integers, with no monochromatic solution to $x + y = z^2$. Without loss of generality, there are infinitely many N such that $|V \cap [N, 2N)| \geq \frac{N}{2}$. Then we have the following chain of inclusions:

$$\sqrt{2V} \subset W,$$

$$\sqrt{2\sqrt{2V}} \subset \sqrt{2W} \subset V,$$
$$\sqrt{2\sqrt{2\sqrt{2V}}} \subset \sqrt{2V} \subset W.$$

It follows from Proposition 4.7.2 that W contains , for infinitely many N, a progression $\mathbb{P}(I_N; N^{1/8}, q_N)$, where $I_N \subset [0.1, 10]$, $|I_N| \gg 1$ and $q_N = O(1)$, both of these uniformly in N. By pigeonholing in the value of q_N, we may assume that $q_N = q$ does not depend on N. Moreover, taking $M = \lceil 10/\inf|I_N| \rceil$ we see that every I_N contains one of the finite collection of intervals $[\frac{i}{M}, \frac{i+1}{M}]$, $M/10 \leq i \leq 10M$. Therefore we may pigeonhole in the choice of interval as well and assume that $I_N = I$ does not depend on N. Thus W contains $\mathbb{P}(I; N^{1/8}, q)$ for some $I \subset [0, 1, 10]$ and for infinitely many N. Rescaling N, we see that W contains $\mathbb{P}([1, 1+c]; N, q)$ for infinitely many N and for some $c > 0$.

From now on, this is the only consequence of the elaborate techniques of the earlier parts of the chapter that we will require.

Using Proposition 4.6.1 as a tool, we find longer and longer progressions inside W. The following lemma formalises this process.

Lemma 4.7.3. *Let $P_1 = \mathbb{P}([\alpha_1, \beta_1]; N, q)$ and $P_2 = \mathbb{P}([\alpha_2, \beta_2], N, q)$. Suppose that $\gamma_1 > \sqrt{\alpha_1^2 + \alpha_2^2}$ and that $\gamma_2 < \sqrt{\beta_1^2 + \beta_2^2}$. Then if N is large enough (depending on $\alpha_i, \beta_i, \gamma_i, q$) we have*

$$\mathbb{P}([\gamma_1, \gamma_2]; N, q) \subset \sqrt{P_1^2 + P_2^2 - P_1 - P_2}.$$

Remark. Here and in what follows, A^2 means $\{a^2 : a \in A\}$ and *not* $\{a \cdot a' : a, a' \in A\}$ as one might find in other literature.

Proof. Fix $\widetilde{\gamma}_1, \widetilde{\gamma}_2$ with $\gamma_1 > \widetilde{\gamma}_1 > \sqrt{\alpha_1^2 + \alpha_2^2}$ and $\gamma_2 < \widetilde{\gamma}_2 < \sqrt{\beta_1^2 + \beta_2^2}$. By Proposition 4.6.1, $P_1^2 + P_2^2$ has a point within $O(\sqrt{N})$ of every point of $\mathbb{P}([\widetilde{\gamma}_1^2, \widetilde{\gamma}_2^2]; N^2, q)$. $P_1 + P_2$ is a progression of length $\gg N$ consisting of multiples of q, and so it is easy to see that $P_1^2 + P_2^2 - P_1 - P_2$ contains all of $\mathbb{P}([\widetilde{\gamma}_1^2, \widetilde{\gamma}_2^2]; N^2, q)$ with the possible exception of points within $O(N)$ of the endpoints, and hence it contains $\mathbb{P}([\gamma_1, \gamma_2]; N^2, q)$. □

Starting from the fact that

$$\mathbb{P}([1, 1+c]; N, q) \subset W \qquad \text{for infinitely many } N, \tag{4.46}$$

we apply Lemma 4.7.3 iteratively. Observe that if $n_1, n_2, n_3, n_4 \in W$ then $n_1^2 - n_3 \in V, n_2^2 - n_4 \in V$, and hence (if it is an integer)

$$\sqrt{n_1^2 + n_2^2 - n_3 - n_4} \in W.$$

Thus if $P_1, P_2 \subset W$ then $\sqrt{P_1^2 + P_2^2 - P_1 - P_2} \subset W$. Using this observation and repeated applications of Lemma 4.7.3, we see that for any finite k and any choice of closed intervals $I_i \subset (\sqrt{i}, (1+c)\sqrt{i})$ there is an infinite sequence of Ns such that $\mathbb{P}(I_i; N, q) \subset W$ for $i = 1, 2, \dots, k$.

We claim that there is some $k = k(c)$ and some choice of $I_1, \dots, I_k$ such that $\bigcup_{i=1}^k I_i$ contains an interval of the form $[x, 3x]$. First note that if $i > 1/2c$ then $(1+c)\sqrt{i} > \sqrt{i+1}$, and so the intervals $(\sqrt{i}, (1+c)\sqrt{i})$ and $(\sqrt{i+1}, (1+c)\sqrt{i+1})$ overlap. Thus if we set $i_0 := \lceil 1/2c \rceil$ and $i_1 := 9i_0$ then $\bigcup_{i_0 \le i \le i_1} (\sqrt{i}, (1+c)\sqrt{i})$ is an interval containing a subinterval of the form $[x, 3x]$.

Thus W contains $\mathbb{P}([x, 3x]; N, q)$ for infinitely many N, and hence (replacing N by $\lfloor 1.1xN \rfloor$) we see that we have bootstrapped (4.46) to the stronger statement that

$$\mathbb{P}([1, 2]; N, q) \subset W \qquad \text{for infinitely many } N.$$

Pick one such $N = N_0$, sufficiently large. Thus

$$\mathbb{P}([1, 2]; N_0, q) \subset W. \tag{4.47}$$

By Lemma 4.7.3 once more (and the inequalities $\sqrt{2} < \frac{3}{2} < \frac{5}{2} < \sqrt{8}$) we have

$$\mathbb{P}(\left[\frac{3}{2}, \frac{5}{2}\right]; N_0, q) \subset W.$$

Together with (4.47), this implies that

$$\mathbb{P}([1, 2]; N_0 + 1, q) \subset W.$$

Continuing inductively, we obtain

$$\bigcup_{N \ge N_0} \mathbb{P}([1, 2]; N, q) \subset W.$$

This implies that all sufficiently large multiples of q lie in W. But there are arbitrarily large multiples x, y, z of q satisfying $x + y = z^2$, and so at last we obtain a contradiction.

Chapter 5

Another application of the Green–Lindqvist approach

This chapter builds directly on Chapter 4, and so we adopt all the notation used there.

5.1 The equation $x - y = z^2$

We wish to prove that the equation

$$x - y = z^2 \tag{5.1}$$

is partition regular over $\mathbb{N}$. This is already known by ergodic methods, see [2], and can also be proven by Fourier analytic methods [54]. The aim of this section is to reprove this fact using the machinery of [31]. As these methods are finitary in nature one in fact gets that for all $k \in \mathbb{N}$ there is a $N_0(k)$ such that any k-colouring of $[N_0(k)]$ has a monochromatic solution to (5.1), and in principle one should even be able to specify what the quantity $N_0(k)$ is. Unfortunately, since we apply the arithmetic regularity lemma[1] this will have an astronomical dependence on k, and is therefore not of much practical use.

In very rough terms the strategy will be as follows. Assume that we have a k-colouring of $[N]$ with no solutions to (5.1). Let $A \subset [N]$ be the largest colour class. Then we know that $A \cap \sqrt{A - A} = \emptyset$, and so all of $\sqrt{A - A}$ must be $(k-1)$-coloured. We will then proceed iteratively in this manner, reducing the number of colours by one at each step by removing the largest colour class and passing to a new set, until eventually we reach a 0-colouring, which is clearly a contradiction.

[1] in fact we apply it roughly $\frac{1}{2}k^2$ times.

In practice we implement this by finding some structured sets which are almost fully contained in $\sqrt{A-A}$. This then gives a "99%"-colouring of the structured set, but using only $k-1$ colours, as otherwise one would get a solution to (5.1) in the colour class A.

At this point we have a problem. Indeed, 99% of a subset of $[M]$ could avoid all of $[0.01M]$, and so if M is large such a set cannot possibly contain a solution to (5.1). We can avoid this problem by keeping track of sets at several scales at once. If $\sqrt{A-A}$ contains 99% of some subset of $[M]$ *and* 99% of some subset of $[M^{1/2}]$, then it is possible that one can find a solution to (5.1) within these sets. In practice this forces us to work at $k+1$ different scales for a k-colouring. Each iteration will lower the number of colours and scales by one, eventually leaving us with a 0-colouring at a single scale.

The main theorem will follow from a slightly technical proposition, which is similar to Proposition 4.4.1.

Proposition 5.1.1. *Let $\eta, \alpha > 0$, $\Omega : \mathbb{N}^3 \to \mathbb{N}$ be a growth function, and let K be sufficiently large in terms of η, α and Ω. Then the following holds.*

Let

$$B_0 = \{n \in \mathbb{N} : n \leqslant \epsilon_0 N, n \equiv 0 (\text{mod } q_0), \|\theta_0 n^2\|_{\mathbb{T}^{d_0}} \leqslant \epsilon_0\},$$

where $\epsilon_0 \in (0,1)$, $q_0, d_0 \in \mathbb{N}$, $\theta_0 \in \mathbb{T}^{d_0}$ and θ_0 is (K, N^2)-irrational. Let $A \subset [N^{2^l}]$ for some $l \geqslant 1$ have $|A| \geqslant \alpha N^{2^l}$. Then there are $d, q = O_{\eta,\Omega,\epsilon_0,q_0,d_0,\alpha}(1)$, $c, \epsilon \gg_{\eta,\Omega,\epsilon_0,q_0,d_0,\alpha} 1$ and $\theta \in \mathbb{T}^d$ such that

$$B = \{n \in \mathbb{N} : n \leqslant \epsilon N, n \equiv 0 (\text{mod } q), \|\theta n^2\|_{\mathbb{T}^d} \leqslant \epsilon\}$$

is contained in B_0 and $\sqrt{A-A}$ contains all but $\eta \epsilon^{d+1} q^{-1} N$ of the elements in B. Furthermore, θ is $(\Omega(q, d, 1/\epsilon), N^2)$-irrational.

By the iterative scheme outlined above this will allow us to prove the main theorem.

Theorem 5.1.2. *For each $k \in \mathbb{N}$ there is some $N_0(k) \in \mathbb{N}$ such that if $[N_0(k)]$ is k-coloured then there is a monochromatic solution to $x - y = z^2$.*

In Section 5.2 we prove Proposition 5.1.1, and in Section 5.3 we show how to prove Theorem 5.1.2.

5.2 Finding squares in Bohr sets

In this section we prove Proposition 5.1.1. The proof closely follows the proof of Proposition 4.4.1. The two main modifications which are needed here are the following.

Firstly, the local problem modulo q becomes much easier to handle in this case. For the equation $x+y=z^2$ we needed to know that if $A \subset \mathbb{Z}/q\mathbb{Z}$ is large enough, then $2A$ contains a quadratic residue modulo q. In the case of $x-y=z^2$ this becomes much easier, as one always has $0^2 \in A-A$. This is why in this case the proof can go through with α much smaller than $\frac{1}{2}$.

The second modification needed is the fact that we require our resulting set B to lie in the specified set B_0. This turns out to be quite easy. Indeed, it suffices to make sure that $q_0|q$, $\epsilon \leqslant \epsilon_0$, $d \geqslant d_0$ and that θ_0 overlaps with d_0 of the coordinates of θ. This last requirement is the only one which could cause trouble, as we need to make sure that we have a sufficient level of irrationality. This is the reason for requiring K to be sufficiently large in terms of η, α and Ω.

A third small modification is that we are passing from the scale N^{2^l} to the scale N, as opposed to passing from the scale N to the scale $N^{1/2}$. This is in fact a rather superficial difference. Note that if $A \subset [N^{2^l}]$ has size $|A| \geqslant \alpha N^{2^l}$, then by the pigeonhole principle there must be some $x \in \{0, N^2, 2N^2, \dots, N^{2^l}-N^2\}$ such that $|A \cap (x+[N^2])| \geqslant \alpha N^2$. Without loss of generality we may therefore assume that $l=1$, as otherwise we may just pass to the set $\widetilde{A}=\{a \in [N^2] : x+a \in A\}$ and note that $\widetilde{A}-\widetilde{A} \subset A-A$. We warn the reader that this will mean that any instance of N in the argument in Section 4.4 is replacedby N^2 in this section.

For later use we record the size of the type of sets appearing in Proposition 5.1.1.

Lemma 5.2.1. *Let* $X=\{n \in \mathbb{N} : n \equiv 0 (\mathrm{mod}\ q), n \leqslant N, \|\theta n^2\|_{\mathbb{T}^d} \leqslant \epsilon\}$ *where* θ *is* (A,N)*-irrational. Then*

$$\frac{1}{4}\frac{N}{q}(2\epsilon)^d \leqslant |X| \leqslant 4\frac{N}{q}(2\epsilon)^d$$

provided N *is sufficiently large in terms of the other parameters.*

Proof. An upper bound is given by

$$|X| \leqslant \sum_{n \leqslant N/q} \psi_\epsilon^+(q^2\theta n^2),$$

where ψ_ϵ^+ is the majorant for $B_\epsilon(0) \subset \mathbb{T}^d$ constructed in Lemma C.2. Expressing ψ_ϵ^+ by its Fourier coefficients this gives

$$|X| \leqslant \sum_{\mathbf{r}} \widehat{\psi_\epsilon^+}(\mathbf{r}) \sum_{n \leqslant N/q} e(q^2 n^2 \mathbf{r} \cdot \theta) = \left(\frac{N}{q} + O(1)\right) \widehat{\psi_\epsilon^+}(\mathbf{0}) + \sum_{\mathbf{r} \neq \mathbf{0}} \widehat{\psi_\epsilon^+}(\mathbf{r}) \sum_{n \leqslant N/q} e(q^2 n^2 \mathbf{r} \cdot \theta).$$

By Corollary 4.3.2 the innermost sum is bounded by $\sqrt{N/q}\|\mathbf{r}\|_1 A^{-1/C_2}$, and by property (2) of Lemma C.2 we also have that

$$\sum_{\mathbf{r} \neq \mathbf{0}} |\widehat{\psi_\epsilon^+}| \|\mathbf{r}\|_1 = O_{\epsilon,d}(1).$$

Furthermore, by property (1) of Lemma C.2 we have that

$$\widehat{\psi_\epsilon^+}(\mathbf{0}) \leqslant 2(2\epsilon)^d.$$

Combining all of this and assuming that N is large enough we get the upper bound.

The lower bound is proved in a similar way, but with a minorant in place of a majorant. □

Assume now that we are given $A \subset [N^2]$ such that $|A| = \alpha N^2$ and B_0 is as in Proposition 5.1.1. We apply Proposition 4.4.2, with the interval $[N^2]$ in place of the interval $[N, 2N]$, to the function 1_A with some $\delta, \mathcal{F}$ which ultimately will depend on α, η and Ω. This gives a decomposition

$$1_A = f_{\mathrm{str}} + f_{\mathrm{unf}} + f_{\mathrm{sml}}$$

where $\sum_{n \leqslant N^2} |f_{\mathrm{sml}}(n)| \leqslant \delta N^2$, $\|\widehat{f_{\mathrm{unf}}}\|_\infty \leqslant \frac{N^2}{\mathcal{F}(M)}$ and $f_{\mathrm{tor}}(n) = F(n(\mathrm{mod}\ q), n/N^2, \theta n)$ with q, d, F, θ satisfying the properties listed in Proposition 4.4.2. We may assume that $q_0 | q$ by modifying M such that now $M \ll_{q_0,\delta,\mathcal{F}} 1$. We may also redefine d and θ such that $d \geqslant d_0$ and θ contains θ_0 as its first d_0 coordinates, at the cost of θ now being $(\min\{K, \mathcal{F}(M)\}, N^2)$-irrational instead of $(\mathcal{F}(M), N^2)$-irrational. But if K is large enough in terms of $\mathcal{F}(M)$, that is, in terms of α, η and Ω, then θ remains $(\mathcal{F}(M), N^2)$-irrational. This is assumed to be the case.

As mentioned, the proof of Proposition 5.1.1 is very similar to the proof of Proposition 4.4.1. Here we outline the places where the choice of parameters differs between the two, but we will be quite sparse on details.

Lemma 5.2.2. *Suppose that δ is sufficiently small and $\mathcal{F}$ grows sufficiently rapidly. Then $\int F d\mu > \frac{9}{10}\alpha$.*

Proof. This is Lemma 4.4.3 with $\frac{9}{10}\alpha$ in place of $\frac{9}{20} = \frac{9}{10} \cdot \frac{1}{2}$. □

Let $U \subset \mathbb{Z}/q\mathbb{Z}$ be the set of $u \in \mathbb{Z}/q\mathbb{Z}$ satisfying

$$\int_0^1 \int_{\mathbb{T}^d} F(u,x,z)dzdx \geqslant \frac{1}{10}\alpha, \tag{5.2}$$

and

$$\sum_{n\leqslant N^2, n\equiv u (\mathrm{mod}\ q)} |f_{\mathrm{sml}}(n)| \leqslant \frac{10\delta}{q\alpha}N^2. \tag{5.3}$$

If U_1 is the set where (5.2) fails then by Lemma 5.2.2 we have

$$\frac{9}{10}\alpha < \int F d\mu = \frac{1}{q}\sum_{u(\mathrm{mod}\ q)} \int_0^1 \int_{\mathbb{T}^d} F(u,x,z)dzdx \leqslant \frac{1}{10}\alpha + \frac{1}{q}|(\mathbb{Z}/q\mathbb{Z}) \setminus U_1|.$$

Let U_2 be the set where (5.3) fails. Since $\sum_{n\leqslant N^2} |f_{\mathrm{sml}}(n)| \leqslant \delta N^2$ we have that

$$|U_2| \leqslant \frac{q\alpha}{10}.$$

Combining these inequalities we get

$$|U| \geqslant |(\mathbb{Z}/q\mathbb{Z}) \setminus U_1| - |U_2| \geqslant (\frac{9}{10}\alpha - \frac{1}{10}\alpha - \frac{1}{10}\alpha)q > 0,$$

and so U is nonempty. Fix some $u \in U$ and define parameters $\rho > \rho' > 0$ by

$$\rho = \frac{\delta}{M}, \quad \rho' = \frac{\delta}{qd}(\epsilon)^{d+1}.$$

For $x \in [0,1]$and $z \in \mathbb{T}^d$, define

$$E_{x,z} = \sum_{\substack{n\leqslant N^2 \\ n\equiv u(\mathrm{mod}\ q) \\ |\frac{n}{N}-x|\leqslant\rho \\ \|\theta n - z\|_{\mathbb{T}^d}\leqslant\rho}} |f_{\mathrm{sml}}(n)|, \quad E'_{x,z} = \sum_{\substack{n\leqslant N^2 \\ n\equiv u(\mathrm{mod}\ q) \\ |\frac{n}{N}-x|\leqslant\rho' \\ \|\theta n - z\|_{\mathbb{T}^d}\leqslant\rho'}} |f_{\mathrm{sml}}(n)|.$$

Note that

$$\int_0^1 \int_{\mathbb{T}^d} E_{x,z}dzdx \leqslant (2\rho)^{d+1}\frac{10\delta}{q\alpha}N^2$$

and

$$\int_0^1 \int_{\mathbb{T}^d} E'_{x,z}dzdx \leqslant (2\rho')^{d+1}\frac{10\delta}{q\alpha}N^2.$$

Therefore,

$$\int_0^1 \int_{\mathbb{T}^d} \left(F(u,x,z) - \frac{q\alpha^2}{400N^2\delta(2\rho)^{d+1}}E_{x,z} - \frac{q\alpha^2}{400N^2\delta(2\rho')^{d+1}}E'_{x,z}\right) dzdx \geqslant \frac{1}{20}\alpha,$$

and so in particular there is some x, z such that

$$F(u,x,z) - \frac{q\alpha^2}{400N^2\delta(2\rho)^{d+1}}E_{x,z} - \frac{q\alpha^2}{400N^2\delta(2\rho')^{d+1}}E'_{x,z} \geqslant \frac{1}{20}\alpha.$$

Now for this value of x, z we get

$$F(u,x,z) \geqslant \frac{1}{20}\alpha, \quad E_{x,z} \leqslant \frac{400\delta(2\rho)^{d+1}}{q\alpha^2}N^2, \quad E'_{x,z} \leqslant \frac{400\delta(2\rho')^{d+1}}{q\alpha^2}N^2.$$

Fix these values of x and z and set

$$X = \{n \in \mathbb{N} : n \equiv u(\mathrm{mod}\ q), \left|\frac{n}{N^2} - x\right|, \|\theta n - z\|_{\mathbb{T}^d} \leqslant \rho\},$$
$$X' = \{n \in \mathbb{N} : n \equiv u(\mathrm{mod}\ q), \left|\frac{n}{N^2} - x\right|, \|\theta n - z\|_{\mathbb{T}^d} \leqslant \rho'\},$$
$$Y = \{n \in \mathbb{N} : n \equiv 0(\mathrm{mod}\ q), n \leqslant \rho N^2, \|\theta n\|_{\mathbb{T}^d} \leqslant \rho\},$$

and

$$Y_- = \{n \in \mathbb{N} : n \equiv 0(\mathrm{mod}\ q), n \leqslant (\rho - 2\rho')N^2, \|\theta n\|_{\mathbb{T}^d} \leqslant \rho - 2\rho'\} \subset Y.$$

At this stage we may assume that $\rho \leqslant \epsilon_0^2$, as if this does not hold we may take M to be larger, at the cost of letting M depend on ϵ_0. This implies that the set

$$\{n \in \mathbb{N} : n^2 \in Y\} \subset B_0,$$

and so we are done if we can show that $A - A$ contains all but $\ll_\alpha \delta(2\rho)^{d+1}q^{-1}N$ of the squares in Y. The remainder of the proof is virtually identical to the remainder of the proof of Proposition 4.4.1, with X, X', Y and Y_- in place of (4.13), (4.14), (4.15) and (4.18), respectively, and all implied constants are now allowed to depend on α. We do not repeat this argument here, see Section 4.4 for the full details.

5.3 Proof of the main theorem

Proof of Theorem 5.1.2. Fix N large enough, and assume we have a k-colouring of $[N^{2^k}]$ with no monochromatic solution to (5.1). We will repeatedly remove one colour by passing to subsets. At the stage where i colours remain we will be left with an i-colouring of all but a fraction η_i of each of the sets $B_0^{(i)}, \dots, B_i^{(i)}$, where for $i = 0, \cdots, k$ we have that

$$B_i^{(i)} \subset B_i^{(i+1)} \subset \cdots \subset B_i^{(k)} \subset [N^{2^i}].$$

This relationship is summarised in the following diagram.

#colours:	k		$k-1$		$\cdots$		1		0
	$[N^{2^k}] = B_k^{(k)}$								
	$[N^{2^{k-1}}] = B_{k-1}^{(k)}$	$\supset$	$B_{k-1}^{(k-1)}$						
	$\vdots$		$\vdots$		$\vdots$	$\ddots$			
	$[N^2] = B_1^{(k)}$	$\supset$	$B_1^{(k-1)}$	$\supset$	$\cdots$	$\supset$	$B_1^{(1)}$		
	$[N] = B_0^{(k)}$	$\supset$	$B_0^{(k-1)}$	$\supset$	$\cdots$	$\supset$	$B_0^{(0)}$	$\supset$	$B_0^{(0)}$

The sets take the form

$$B_j^{(i)} = \left\{ n \in \mathbb{N} : n \leqslant \epsilon_j^{(i)} N^{2^j}, n \equiv 0 (\text{mod } q_j^{(i)}), \|\theta_j^{(i)} n^2\|_{\mathbb{T}^{d_j^{(i)}}} \leqslant \epsilon_j^{(i)} \right\}, \tag{5.4}$$

where $d_j^{(i)}, q_j^{(i)} \ll_k 1$, $\epsilon_j^{(i)} \gg_k 1$, $\theta_j^{(i)} \in \mathbb{T}^{d_j^{(i)}}$ and $\theta_j^{(i)}$ is $\left(\Omega^{(i)}\left(q_j^{(i)}, d_j^{(i)}, 1/\epsilon_j^{(i)}\right), N^{2^{j+1}}\right)$-irrational. The growth functions $\Omega^{(i)} : \mathbb{N}^3 \to \mathbb{N}$ will depend ultimately only on k. Furthermore, all but a fraction η_i of the set $B_j^{(i)}$ is $i-1$ coloured for $j = 0, \dots, i$, where we simply set $\eta_i = \frac{1}{2i}$ for $i \geqslant 1$, and $\eta_0 = \frac{1}{2}$.

Note that we will choose the $\Omega^{(i)}$'s in the opposite order of the other parameters. That is, we first fix $\Omega^{(0)}$, then $\Omega^{(1)}$, and so on all the way up to $\Omega^{(k-1)}$.

The first step Passing from k to $k-1$ colours follows the same procedure as passing from i to $i-1$ colours, but since the initial sets $B_0^{(k)}, \cdots, B_k^{(k)}$ are full intervals, this case is a bit easier, which we highlight here.

To begin with, assume that we have a k-colouring of $[N^{2^k}]$ and set $B_j^{(k)} = [N^{2^j}]$ for $j = 0, \cdots, k$. Let A_k denote the largest colour class on $B_k^{(k)} = [N^{2^k}]$, and so in particular $|A_k| \geqslant \frac{1}{k} N^{2^k}$. We will now apply Proposition 5.1.1 k times to find sets $B_0^{(k-1)}, \cdots, B_{k-1}^{(k-1)}$.

To find the set $B_j^{(k-1)}$, apply Proposition 5.1.1 with $B_0 = B_j^{(k)}$ and N^j in place of N. That is, $l = k-j$, $q_0 = 1$, $d_0 = 0$, $A = A_k$, $\alpha = \frac{1}{k}$, $\eta = \eta_{k-1}$ and $\Omega = \Omega^{(k-1)}$, where we will specify $\Omega^{(k-1)}$ later. The set $B_j^{(k-1)}$ is then chosen to be precisely the resulting set B. Note that the requirement of K being sufficiently large holds vacuously in this case.

We have that $\sqrt{A_k - A_k}$ contains a fraction $(1-\eta_{k-1})$ of each of the sets

$$B_0^{(k-1)}, \dots, B_{k-1}^{(k-1)}.$$

Because of the assumption $\sqrt{A_k - A_k} \cap A_k = \emptyset$ we therefore get $(k-1)$-colourings of these sets.

The induction step Assume now that we are at stage i, that is, we have an i-colouring of a fraction $(1-\eta_i)$ of each of the sets $B_0^{(i)}, \dots, B_i^{(i)}$ which are defined as in (5.4).

Let $A_i \subset B_i^{(i)}$ be the largest colour class (of the remaining colours) on $B_i^{(i)}$, and so in particular $|A_i| \geqslant \left(\frac{1}{i} - \eta_i\right)|B_i^{(i)}| = \frac{1}{2i}|B_i^{(i)}|$. By Lemma 5.2.1 we have that $|B_i^{(i)}| \gg_k N^{2^i}$ provided N is large enough, and so we get that indeed $|A_i| \gg_k N^{2^i}$.

For $j = 0, \dots, i-1$, to find the set $B_j^{(i-1)}$ we apply Proposition 5.1.1 with $B_0 = B_j^{(i)}$, $A = A_i$, $\eta = \eta_{i-1}$ and $\Omega = \Omega^{(i-1)}$. Then $B_j^{(i)}$ is chosen to be the resulting set B.

Note that the requirement in Proposition 5.1.1 of K being large enough in terms of η, α and Ω translates into $\Omega^{(i)}$ growing sufficiently rapidly in terms of k and $\Omega^{(i-1)}$. Given $\Omega^{(i-1)}$ we choose $\Omega^{(i)}$ in such a way that this is the case.

The base step Finally, at stage 0 we have a 0-colouring of a fraction $(1-\eta_0) = \frac{1}{2}$ of some set

$$B_0^{(0)} = \left\{ n \in \mathbb{N} : n \leqslant \epsilon_0^{(0)} N, n \equiv 0 (\text{mod } q_0^{(0)}), \|\theta_0^{(0)} n^2\|_{\mathbb{T}^{d_0^{(0)}}} \leqslant \epsilon_0^{(0)} \right\}$$

where $q_0^{(0)}, d_0^{(0)} \ll_k 1$, $\epsilon_0^{(0)} \gg_k 1$. We now fix some choice of $\Omega^{(0)}$, which enables us to fix $\Omega^{(1)}, \dots, \Omega^{(k-1)}$. All that remains in order to get a contradiction is to show that $B_0^{(0)}$ is nonempty. By Lemma 5.2.1 this is indeed the case, provided N is large enough. □

Chapter 6

Rado's criterion for squares and higher powers

6.1 Introduction

Schur's theorem [59] is a foundational result in Ramsey theory, asserting that in any finite colouring of the positive integers there exists a monochromatic solution to the equation $x + y = z$ (a solution in which each variable receives the same colour). A notorious question of Erdős and Graham asks if the same is true for the Pythagorean equation $x^2 + y^2 = z^2$, offering \$250 for an answer [28, 27]. The computer-aided verification [38] of the two colour case of this problem is reported to be the largest mathematical proof in existence, consuming 200 terabytes [46]. In this chapter we provide an affirmative answer to the analogue of the Erdős–Graham question for the Pythagorean equation in five variables.

Theorem 6.1.1 (Schur-type theorem in the squares)**.** *In any finite colouring of the positive integers there exists a monochromatic solution to the equation*

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = x_5^2. \tag{6.1}$$

This is an example of the following property.

Definition 6.1.2 (Partition regular)**.** Given a polynomial $P \in \mathbb{Z}[x_1, \dots, x_s]$ and a set S call the equation $P(x) = 0$ *partition regular over* S if, in any finite colouring of S, there exists a solution $x \in S^s$ whose coordinates all receive the same colour. We say that the equation is *non-trivially partition regular* if every finite colouring of S has a monochromatic solution in which each variable is distinct.

Rado [56] established an elegant algebraic characterisation of partition regular homogeneous linear equations.

Rado's criterion for one equation. Let $c_1, \dots, c_s \in \mathbb{Z}\setminus\{0\}$, where $s \geqslant 3$. Then the equation $\sum_{i=1}^{s} c_i x_i = 0$ is (non-trivially) partition regular over the positive integers if and only if there exists a non-empty set $I \subset [s]$ such that $\sum_{i\in I} c_i = 0$.

A number of authors [2, 3, 27, 19] have sought algebraic characterisations of partition regularity within families of non-linear Diophantine equations. The example of the Fermat equation shows that one cannot hope for something as simple as Rado's criterion for diagonal forms. Nevertheless, provided that the number of variables s is sufficiently large in terms of the degree k, we establish that the same criterion characterises partition regularity for equations in kth powers.

Theorem 6.1.3 (Rado over kth powers)**.** *Let $s \geq k^2 + 1$ and $c_1, \dots, c_s \in \mathbb{Z} \setminus \{0\}$. Then the equation*

$$\sum_{i=1}^{s} c_i x_i^k = 0 \tag{6.2}$$

is (non-trivially) partition regular over the positive integers if and only if there exists a non-empty set $I \subset [s]$ such that $\sum_{i\in I} c_i = 0$.

In fact, by incorporating smooth numbers into the argument one may take $s \geqslant 8$ if $k = 3$ and

$$s \geqslant k\left(\log k + \log\log k + 2 + O(\log\log k/\log k)\right). \tag{6.3}$$

for $k \geqslant 4$. The argument using smooth numbers is rather technical, and so we do not include it here. The full proof can be found in [14].

Notice that Rado's criterion for a linear equation shows that the condition $\sum_{i\in I} c_i = 0$ is necessary for (6.2) to be partition regular. The content of Theorem 6.1.3 is that this condition is also sufficient.

For higher-degree equations one cannot avoid the assumption of some lower bound on the number of variables, as the example of the Fermat equation demonstrates. Given current knowledge on the solubility of diagonal Diophantine equations [69], the bound (6.3) is at the cutting edge of present technology. Indeed, it is unlikely that one could improve this condition without making an analogous breakthrough in Waring's problem, since partition regularity implies the existence of a non-trivial integer solution to the equation (6.2).

6.1.1 Non-triviality

It may be that (6.2) possesses a wealth of monochromatic solutions for 'trivial' reasons. For instance, if $c_1 + \cdots + c_s = 0$ then taking $x_1 = \cdots = x_s$ yields many uninteresting solutions. We have delineated between partition regularity and non-trivial partition regularity to ensure that Rado's criterion still has content in such a situation. However, since Rado's criterion is necessary for 'trivial' partition regularity, the two notions are in fact equivalent.

6.1.2 Previous work

To the knowledge of the authors, work on non-linear partition regularity begins with papers of Furstenberg and Sárközy [25, 58], independently resolving a conjecture of Lovász—a line of investigation which culminates in the polynomial Szemerédi theorem of Bergelson–Leibman [5], proved using ergodic methods. Such methods have also established colouring results for which no density analogue exists, such as partition regularity of the equation $x - y = z^2$ as shown in [2, p.53][1]. Interestingly, the story is more complicated for the superficially similar equation $x + y = z^2$ studied in [42, 15, 53] and Chapter 4.

A recent breakthrough of Moreira [52] resolves a longstanding conjecture of Hindman [39], proving partition regularity of the equation $x + y^2 = yz$. More intuitively: in any finite colouring of the positive integers there exists a monochromatic configuration of the form $\{a, a + b, ab\}$. This result is a consequence of a general theorem which also yields partition regularity of equations of the form $x_0 = c_1x_1^2 + \cdots + c_sx_s^2$, subject to the condition that $c_1 + \cdots + c_s = 0$.

Notice that all of the above results involve an equation with at least one linear term. There are fewer results in the literature concerning genuinely non-linear equations such as (6.2). Certain diagonal quadrics are dealt with in Lefmann [47, Fact 2.8], using Rado's theorem to locate a long monochromatic progression whose common difference possesses a (well-chosen) multiple of the same colour. This results in the following sufficient condition for partition regularity.

Theorem 6.1.4 (Lefmann). *Let $c_1, \ldots, c_s \in \mathbb{Z} \setminus \{0\}$, and suppose that $\sum_{i \in I} c_i = 0$*

[1] For a different proof see Chapter 5

with $I \neq \emptyset$. Moreover, suppose that the auxiliary system

$$\begin{aligned} \Big(\sum_{i\notin I} c_i\Big)x_0^2 + \sum_{i\in I} c_i x_i^2 = 0, \\ \sum_{i\in I} c_i x_i = 0 \end{aligned} \tag{6.4}$$

possesses a rational solution with $x_0 \neq 0$. Then the equation

$$c_1 x_1^2 + \cdots + c_s x_s^2 = 0 \tag{6.5}$$

is partition regular.

This result reduces the combinatorial problem of establishing partition regularity of (6.5) to a task in number theory: find a rational point of a certain form on a variety determined by a diagonal quadric and linear equation. In [14] we show that Lefmann's result together with the Hardy–Littlewood circle method yields the following.

Theorem 6.1.5 (Lefmann + Hardy–Littlewood circle method)**.** *Let $c_1, \ldots, c_s \in \mathbb{Z} \setminus \{0\}$, and suppose that $\sum_{i\in I} c_i = 0$ with $I \neq \emptyset$. Suppose in addition that $|I| \geq 6$ and at least two c_i are positive and at least two are negative. Then*

$$c_1 x_1^2 + \cdots + c_s x_s^2 = 0 \tag{6.6}$$

is partition regular.

This result does not encompass all equations amenable to Lefmann's criterion: fewer variables may suffice, for instance

$$x^2 + 9y^2 = 2z^2 + 8w^2 \qquad \text{or} \qquad 4x^2 + y^2 = 2z^2 + 2w^2.$$

We emphasise that Lefmann's criterion cannot hope to be a necessary condition for partition regularity, as there are partition regular equations for which the auxiliary Lefmann system (6.4) has no rational point of the required form. Such equations include the generalised Pythagorean equation (6.1), as well as the 'convex' equation

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = 4x_5^2 \tag{6.7}$$

addressed in [11].

6.2 Overview of the argument

In this section we prove Theorem 6.1.3 by piecing together various key theorems which we prove in the next sections. The theorem follows from the following finitary version.

Theorem 6.2.1 (Finitary Schur-type theorem)**.** *Let $s > 0$, $\lambda_1, \dots, \lambda_s, \mu_1, \dots, \mu_t \in \mathbb{Z} \setminus \{0\}$ with $\lambda_1 + \cdots + \lambda_s = 0$ and $s + t \geqslant k^2 + 1$. For any $r \in \mathbb{N}$ there exists $N_0 \in \mathbb{N}$ such that for any $N \geq N_0$ the following is true. Given an r-colouring of $[N]$ there exists a monochromatic solution to the equation*

$$\lambda_1 x_1^k + \cdots + \lambda_s x_s^k = \mu_1 y_1^k + \cdots + \mu_t y_t^k.$$

Inspired by work of Cwalina–Schoen [16] and Green–Sanders [32], we derive Theorem 6.2.1 in Section 6.3 by an induction on the number of colours, in combination with a density result concerning what we have termed *homogeneous sets*.

Definition 6.2.2 (Homogeneous set)**.** Call a set B of positive integers *M-homogeneous* if for any $q \in \mathbb{N}$ we have

$$B \cap q \cdot [M] \neq \emptyset. \tag{6.8}$$

Given a set $S \subset \mathbb{N}$, we say that B is *M-homogeneous in S* if (6.8) holds for all homogeneous progressions $q \cdot [M]$ contained in S. Notice that the latter does not require that $B \subset S$.

Chapman [12] has observed that this is a quantitative variant of what it means to be *multiplicatively syndetic* (see Bergelson–Glasscock [4]), and that such sets appear to have a number of interesting properties in regard to the partition regularity of homogeneous systems of polynomial equations.

We remark that if B is an M-homogeneous set then $|B \cap [N]| \gg_M N$ for N sufficiently large in terms of M, so homogeneous sets are dense (see Lemma 6.3.2). In fact they are dense on all sufficiently long homogeneous arithmetic progressions.

To prove partition regularity of the generalised Pythagorean equation we induct on the number of colours. We then divide into two cases based on the level of homogeneity of the colour classes. The inhomogeneous case is dealt with by passing to a subprogression and applying the induction hypothesis. In the remaining case we may assume that all colour classes are homogeneous. In this situation we are able to show that every colour class contains many solutions to our non-linear equation by employing the following density result.

Theorem 6.2.3 (Non-linear homogeneous Sárközy)**.** *Let $\lambda_1, \dots, \lambda_s, \mu_1, \dots, \mu_t \in \mathbb{Z} \setminus \{0\}$ with $s + t \geq k^2 + 1$ and $\lambda_1 + \cdots + \lambda_s = 0$. For any $\delta > 0$ and $M \in \mathbb{N}$ there exist $N_0 \in \mathbb{N}$ and $c_0 > 0$ such that for any $N \geq N_0$ the following holds. Let $A \subset [N]$ have density at least δ and let B be a M-homogeneous subset of the positive integers. Then*

$$\#\Big\{(x, y) \in A^s \times B^t : \sum_{i=1}^{s} \lambda_i x_i^k = \sum_{j=1}^{t} \mu_j y_j^k\Big\} \geq c_0 N^{s+t-k}.$$

Using Green's Fourier-analytic transference principle [30], as elucidated for squares in [11, 55], the deduction of Theorem 6.2.3 is reduced (in sections 6.4–6.5) to a linear analogue in which the kth powers have been removed from the dense variables. This can be thought of as a generalisation of the Furstenberg–Sárközy theorem [25, 58], extended to homogeneous sets.

Theorem 6.2.4 (Supersaturated homogeneous Sárközy)**.** *Let $\lambda_1, \dots, \lambda_s, \mu_1, \dots, \mu_t \in \mathbb{Z} \setminus \{0\}$ with $\lambda_1 + \cdots + \lambda_s = 0$. For any $\delta > 0$ and $M \in \mathbb{N}$ there exists $N_0 \in \mathbb{N}$ and $c_0 > 0$ such that for any $N \geq N_0$ the following holds. Let $A \subset [N]$ have density at least δ in $[N]$ and B be M-homogeneous in $[N^{1/k}]$. Then there are at least $c_0 N^{s+\frac{t}{k}-1}$ tuples $(x, y) \in A^s \times B^t$ satisfying the equation*

$$\lambda_1 x_1 + \cdots + \lambda_s x_s = \mu_1 y_1^k + \cdots + \mu_t y_t^k. \tag{6.9}$$

Our ability to remove the kth powers from the dense variables is intrinsically linked to the fact that the coefficients corresponding to these variables sum to zero. One consequence of this is that we may restrict all of the dense variables to lie in the same congruence class, without destroying solutions to the equation in the process.

Theorem 6.2.4 is ultimately derived, in Section 6.7, from a homogeneous version of the Bergelson–Leibman theorem given in Section 6.6.

6.3 Induction on colours

In this section we derive Theorem 6.2.1 from Theorem 6.2.3 by induction on the number of colours.

Definition 6.3.1 (T counting operator)**.** Given functions $f_1, \dots, f_s : \mathbb{Z} \to \mathbb{C}$ with finite support, define the counting operator

$$T(f_1, \dots, f_s) := \sum_{c_1 x_1^k + \cdots + c_s x_s^k = 0} f_1(x_1) f_2(x_2) \cdots f_s(x_s).$$

We write $T(f)$ for $T(f, f, \dots, f)$.

By Theorem D.3, there exist $N_1 \in \mathbb{N}$ and $c_1 > 0$ such that for $N \geq N_1$ we have

$$T(1_{[N]}) \geq c_1 N^{s-k}.$$

Since the latter quantity is positive, Theorem 6.2.1 follows for 1-colourings (the base case of our induction).

Let $[N] = C_1 \cup \cdots \cup C_r$ be an r-colouring. We split our proof into two cases depending on the homogeneity of the C_i.

Let $M := N_0(r-1)$ be the quantity whose existence is guaranteed by our inductive hypothesis. We first suppose that some C_i is not M-homogeneous in $[N]$ (see Definition 6.2.2). Consequently there exists $q \in \mathbb{N}$ such that

$$q \cdot [M] \subset [N] \quad \text{and} \quad C_i \cap q \cdot [M] = \emptyset. \tag{6.10}$$

For $j \neq i$ let us define

$$C_j' := \{x \in [M] : qx \in C_j\}.$$

Then it follows from (6.10) that $\bigcup_{j \neq i} C_j' = [M]$. By the induction hypothesis, there exist $y_k \in C_j'$ for some $j \neq i$ such that $c_1 y_1^k + \cdots + c_s y_s^k = 0$. Setting $x_k := q y_k$ we obtain elements of C_j which solve (6.2).

In the second case we assume that every colour class is M-homogeneous in $[N]$. We claim that Theorem 6.2.3 then implies that each C_i contains a solution to equation (6.2). First we observe that each colour class is dense.

Lemma 6.3.2 (Homogeneous sets are dense)**.** *If $B \subset [N]$ is M-homogeneous in $[N]$ then*

$$|B| \geq \frac{1}{M} \left\lfloor \frac{N}{M} \right\rfloor.$$

Proof. We proceed by a variant of Varnavides averaging [64]. For each $q \leqslant N/M$ the definition of homogeneity gives

$$B \cap q \cdot [M] \neq \emptyset.$$

Summing over q then yields

$$\sum_{q \leqslant N/M} |B \cap q \cdot [M]| \geq \lfloor N/M \rfloor.$$

Interchanging the order of summation, we see that

$$\sum_{x \in B} \#\{(q,m) \in [N/M] \times [M] : x = qm\} \geq \lfloor N/M \rfloor.$$

The result follows on noting that

$$\#\{(q,m) \in [N/M] \times [M] : x = qm\} \leqslant M.$$

□

Setting $A = B = C_i$ in Theorem 6.2.3 we deduce that if $N \geq N_0(M)$ then

$$T(1_{C_r}) \geq c_0(M)N^{s-k}.$$

Since the latter quantity is positive the induction step follows, completing the proof of Theorem 6.2.1. Note that a quantity dependent on $M = N_0(r-1)$ is ultimately dependent only on r.

6.4 A pseudorandom Furstenberg–Sárközy theorem

In Section 6.3 we reduced partition regularity of (6.2) to Theorem 6.2.3. In Section 6.5 we will deduce the latter result from Theorem 6.2.4. To prepare the ground for this deduction, we first modify Theorem 6.2.4 to accommodate sets which are relatively dense in a suitably pseudorandom set. The goal is to find the weakest possible pseudorandomness conditions required for such a result to hold. Our primary quantity of interest is the following.

Definition 6.4.1 (T_l counting operator)**.** Given functions $f_1, \dots, f_s : \mathbb{Z} \to \mathbb{C}$ with finite support, $l = 1$ or $l = k$ and $B \subset \mathbb{Z}$, define

$$T_l(f_1, \dots, f_s; B) := \sum_{\lambda_1 x_1^l + \dots + \lambda_s x_s^l = \mu_1 y_1^k + \dots + \mu_t y_t^k} f_1(x_1) \cdots f_s(x_s) 1_B(y_1) \cdots 1_B(y_t).$$

We write $T_l(f; B)$ for $T_l(f, \dots, f; B)$ and $T_l(A; B)$ for $T_l(1_A; B)$.

We begin by showing how Theorem 6.2.4 implies a result in which the indicator function 1_A can be replaced by a function $f : [N] \to [0,1]$ with sufficiently large average.

Lemma 6.4.2 (Functional Sárközy)**.** *For any $\delta > 0$ and $M \in \mathbb{N}$ there exists $N_0 \in \mathbb{N}$ and $c_0 > 0$ such that for any $N \geq N_0$ the following holds. Let $f : [N] \to [0,1]$ with $\|f\|_1 \geq \delta N$ and let B be M-homogeneous in $[N^{1/k}]$. Then*

$$T_1(f; B) \geq c_0 N^{s + \frac{t}{k} - 1}.$$

Proof. Let $A = \{x \in [N] : f(x) \geqslant \delta/2\}$. As $\|f\|_1 \geqslant \delta N$ and $f \leq 1$, we have $|A| \geqslant \delta N/2$. Since $f \geqslant \delta 1_A/2$, we deduce that

$$T_1(f;B) \geqslant (\delta/2)^s T_1(A;B),$$

and an application of Theorem 6.2.4 completes the proof. □

Our next step is to weaken the assumptions of Theorem 6.2.4 even further, replacing bounded functions with unbounded functions which are sufficiently pseudorandom. The pseudorandomness we enforce posits the existence of a 'random-like' majorising function ν, whose properties are given in the following two definitions.

Definition 6.4.3 (Fourier decay)**.** We say that $\nu : [N] \to [0,\infty)$ has *Fourier decay of level* θ (with respect to $1_{[N]}$) if

$$\left\| \frac{\widehat{\nu}}{\|\nu\|_1} - \frac{\widehat{1}_{[N]}}{\left\|1_{[N]}\right\|_1} \right\|_\infty \leqslant \theta.$$

Definition 6.4.4 (p-restriction)**.** We say that $\nu : [N] \to [0,\infty)$ satisfies a *p-restriction estimate with constant K* if

$$\sup_{|\phi| \leqslant \nu} \int_{\mathbb{T}} \left|\widehat{\phi}(\alpha)\right|^p \mathrm{d}\alpha \leqslant K \|\nu\|_1^p N^{-1}.$$

Theorem 6.4.5 (Pseudorandom Sárközy)**.** *Let $s + t \geqslant k^2 + 1$. For any $\delta > 0$ and $M \in \mathbb{N}$ there exist $N_0, c_0, \theta > 0$ such that for any $N \geq N_0$ the following holds. Let B be M-homogeneous in $[N^{1/k}]$. Let $\nu : [N] \to [0,\infty)$ satisfy a $(s + t - \frac{1}{2})$-restriction estimate and have Fourier decay of level θ. Then for any $f : [N] \to [0,\infty)$ with $f \leqslant \nu$ and $\|f\|_1 \geq \delta \|\nu\|_1$ we have*

$$T_1(f;B) \geq c_0 \|\nu\|_1^s N^{\frac{t}{k}-1}. \tag{6.11}$$

Proof. We may replace B by $B \cap [N^{1/k}]$, so we assume that $B \subset [N^{1/k}]$. Since ν has Fourier decay of level θ, we may apply the dense model lemma recorded in [55, Theorem 5.1], rescaling as appropriate, to conclude the existence of $g : \mathbb{Z} \to \mathbb{C}$ satisfying $0 \leqslant g \leqslant 1_{[N]}$ and

$$\left\| \frac{\widehat{f}}{\|\nu\|_1} - \frac{\widehat{g}}{N} \right\|_\infty \ll \log(\theta^{-1})^{-3/2}. \tag{6.12}$$

Provided that $\theta \leqslant \exp(-C\delta^{-1})$ with C a large positive constant, we can compare Fourier coefficients at 0 to deduce that $\|g\|_1 \gg \delta N$. Applying Lemma 6.4.2 then gives

$$T_1(g;B) \gg_{\delta,M} N^{s+\frac{t}{k}-1}. \tag{6.13}$$

Let h denote the indicator function of the set $\{x^k : x \in B \cap [N^{1/k}]\}$. Then for functions $h_1, \cdots, h_s : [N] \to \mathbb{C}$ we have

$$T_1(h_1, \cdots, h_s; B) = \sum_{\lambda \cdot x = \mu \cdot y} h_1(x_1) \cdots h_s(x_s) h(y_1) \cdots h(y_t). \tag{6.14}$$

The function h is majorised by the indicator function of the set

$$\{x^k : x \in [N^{1/k}]\}$$

which, by Lemma D.4, satisfies a $(s+t-\frac{1}{2})$-restriction estimate with constant $O(1)$.

The function g is majorised by $1_{[N]}$, which satisfies a $(s+t-\frac{1}{2})$-restriction estimate with constant $O(1)$. Employing the generalised von Neumann lemma (Lemma D.8), together with (6.12) and (6.14), we deduce that

$$\left| \|\nu\|_1^{-s} T_1(f; B) - N^{-s} T_1(g; B) \right| \ll N^{\frac{t}{k}-1} \log(\theta^{-1})^{-0.75}.$$

Combining this with (6.13) and choosing $\theta \leqslant \theta_0(\delta, M)$ completes the proof. □

6.5 The W-trick for kth powers

Our objective in this section is to use Theorem 6.4.5 to deduce Theorem 6.2.3. This deduction proceeds by using the W-trick for kth powers, analogous to that developed for prime powers in [13]. Let

$$W = k^{k-1} \prod_{p \leqslant w} p^k, \tag{6.15}$$

where $w = w(\delta, M)$ is a constant to be determined, and the product is over primes.

Lemma 6.5.1. *Let W be as in (6.15) and $\delta > 0$. For any set $A \subset [N]$ with $|A| \geqslant \delta N$ there exist $\zeta \ll_{\delta,w} 1$ and $\xi \in [W]$ with $(W, \xi) = 1$ such that*

$$\#\{x \in \mathbb{Z} : \zeta(\xi + Wx) \in A\} \geq \tfrac{1}{2}\delta \#\{x \in \mathbb{Z} : \zeta(\xi + Wx) \in [N]\}. \tag{6.16}$$

Proof. Observe that any $x \in \mathbb{N}$ can be represented as $x = \zeta(\xi + Wy)$ with ζ w-smooth (see Definition D.1), $\xi \in [W]$, $(\xi, W) = 1$ and $y \in \mathbb{Z}$ in exactly one way. Let $M = 4\delta^{-2}10^{2w}$. By Lemma D.2 there are at most $10^w N M^{-1/2} = \frac{1}{2}\delta N$ elements of $[N]$ divisible by a w-smooth number greater than M. This means that

$$\delta N \leqslant \sum_{\substack{\zeta \in [M] \\ \zeta \text{ } w\text{-smooth}}} \sum_{\substack{\xi \in [W] \\ (\xi, W) = 1}} \#\Big\{x \in A : x = \zeta(\xi + Wy) \text{ for some } y \in \mathbb{Z}\Big\} + \frac{1}{2}\delta N,$$

and so the result follows by the pigeonhole principle. □

Define

$$P := \frac{N}{\zeta}, \qquad X := \frac{P^k}{kW} \tag{6.17}$$

and set

$$A_1 := \left\{ \tfrac{(Wx+\xi)^k - \xi^k}{kW} : \zeta(Wx+\xi) \in A \text{ and } Wx+\xi \in [P] \right\} \setminus \{0\}. \tag{6.18}$$

Then $A_1 \subset [X]$. By Lemma 6.5.1 we have the lower bound

$$|A_1| \geq \tfrac{\delta}{2} \#\{x \in [P] : x \equiv \xi \bmod W\}. \tag{6.19}$$

Noting that $(kW)^{1/k}$ is a positive integer, let

$$B_1 := \left\{ y \in \mathbb{N} : \zeta (kW)^{1/k} y \in B \right\}, \tag{6.20}$$

such that B_1 is M-homogeneous in the positive integers. Recalling that $\sum_{i=1}^{s} \lambda_i = 0$, we have

$$T_k(A; B) \geq T_1(A_1; B_1). \tag{6.21}$$

Define $\nu : [X] \to [0, \infty)$ by

$$\nu(n) = \begin{cases} x^{k-1}, & \text{if } n = \frac{x^k - \xi^k}{kW} \text{ for some } x \in [P] \text{ with } x \equiv \xi \bmod W \\ 0, & \text{otherwise.} \end{cases} \tag{6.22}$$

Observe that

$$\sum_n \nu(n) = \sum_{y \leqslant \frac{P}{W}} (Wy)^{k-1} + O(P^{k-1}) = \frac{P^k}{kW} + O(P^{k-1}) = X + O(X^{1-\frac{1}{k}}), \tag{6.23}$$

so ν has average 1 on $[X]$.

Lemma 6.5.2 (Density transfer)**.** *For N large in terms of k, w and δ we have*

$$\sum_{n \in A_1} \nu(n) \gg_k \delta^k \sum_n \nu(n). \tag{6.24}$$

Proof. We employ (6.19) to conclude that

$$\begin{aligned} &\#\left\{ x \in [P] : x \equiv \xi \bmod W, \quad \frac{x^k - \xi^k}{kW} \in A_1, \quad x > Z \right\} \\ &\geqslant |A_1| - ZW^{-1} - 1 \\ &\geqslant \tfrac{\delta}{2} \#\{x \in [P] : x \equiv \xi \bmod W\} - ZW^{-1} - 1 \\ &\geqslant \tfrac{\delta P}{2W} - ZW^{-1} - 2. \end{aligned}$$

Choosing

$$Z = \frac{\delta P}{4}$$

and taking N (and thus P) sufficiently large then gives

$$\sum_{n\in A_1} \nu(n) \geqslant \tfrac{(\delta/4)^k P^k}{2W} = \tfrac{k}{2}(\delta/4)^k X,$$

where we have recalled (6.17). An application of (6.23) completes the proof. □

The following two ingredients are established in Appendix D.

Lemma 6.5.3 (Fourier decay)**.** *We have*

$$\left\| \frac{\widehat{\nu}}{\|\nu\|_1} - \frac{\widehat{1}_{[X]}}{X} \right\|_\infty \ll_\eta w^{-1/k}. \tag{6.25}$$

Lemma 6.5.4 (Restriction estimate)**.** *If $p > k^2$ then*

$$\sup_{|\phi|\leqslant\nu} \int_{\mathbb{T}} \left|\widehat{\phi}(\alpha)\right|^p \mathrm{d}\alpha \ll_k \|\nu\|_1^p X^{-1}.$$

Proof of Theorem 6.2.3. We employ Theorem 6.4.5 with majorant ν given by (6.22), homogeneous set B_1 given by (6.20), and function $f = \nu 1_{A_1}$ (recall (6.18)). It is first necessary to check that these choices satisfy the hypotheses of Theorem 6.4.5.

By Lemma 6.5.4, the function ν satisfies a $(s + t - \frac{1}{2})$-restriction estimate with constant $K = O_k(1)$. Let c_k denote the implied constant in (6.24) and set $\widetilde{\delta} := c_k\delta^k$. Theorem 6.4.5 guarantees the existence of a positive constant

$$\theta = \theta(\widetilde{\delta}, M) \tag{6.26}$$

such that provided ν has Fourier decay of level θ and $\|f\|_1 \geq \widetilde{\delta}\|\nu\|_1$ we may conclude that (6.11) holds. Taking

$$w = C\theta^k$$

guarantees sufficient Fourier decay, by Lemma 6.5.3. We note that this choice of w satisfies $w \ll_{\delta,M} 1$, as can be checked by unraveling the dependencies in (6.26). We obtain $\|f\|_1 \geq \widetilde{\delta}\|\nu\|_1$ via Lemma 6.5.2. This requires us to take N sufficiently large in terms of k, w and δ. By our choice of w, this is ensured if N is sufficiently large in terms of δ and M (as we may assume).

Applying Theorem 6.4.5 and (6.23) yields

$$T_1(\nu 1_{A_1}; B_1) \gg_{\delta,M} \|\nu\|_1^s X^{\frac{t}{k}-1} \gg_{\delta,M} X^{s+\frac{t}{k}-1}.$$

By (6.21) and the bound $\|\nu\|_\infty \ll_{\delta,M} N^{k-1}$, we finally have

$$T_k(A; B) \geqslant T_1(A_1; B_1) \geqslant \|\nu\|_\infty^{-s} T_1(\nu 1_{A_1}; B_1) \gg_{\delta,M} N^{s+t-k}.$$

□

6.6 The homogeneous Bergelson–Leibman theorem

In this section we prove a Theorem similar in flavour to the following special case of the multidimensional polynomial Szemerédi theorem of Bergelson–Leibman [5].

Theorem 6.6.1 (Bergelson–Leibman)**.** *Let $k \in \mathbb{N}$, $\delta > 0$ and let $F \subset \mathbb{Z}^d$ be a finite set. There exists $N_0 = N_0(k, \delta, F)$ such that for any $N \geq N_0$, if $A \subset [N]^d$ has size $|A| \geq \delta N^d$ then there exists $x \in \mathbb{Z}^d$ and $y \in \mathbb{N}$ such that*

$$x + y^k \cdot F \subset A.$$

We require a version of this result in which the kth power comes from a homogeneous set. Fortunately, this strengthening can be deduced from the original. It is convenient to set up the following notation.

Notation. Given $q, y, k \in \mathbb{N}^d$ define

$$q \otimes y := (q_1 y_1, \dots, q_d y_d), \qquad y^{\otimes k} := (y_1^{k_1}, \dots, y_d^{k_d}).$$

For $F \subset \mathbb{Z}^d$, write $q \otimes F$ for the set

$$\{q \otimes y : y \in F\}.$$

Here is our version of the Bergelson–Leibman theorem with common difference arising from a homogeneous set.

Corollary 6.6.2 (Homogeneous Bergelson–Leibman)**.** *Let $k \in \mathbb{N}^d$, $M \in \mathbb{N}$, $\delta > 0$ and let $F \subset \mathbb{Z}^d$ be a finite set. There exists N_0 such that for any $N \geq N_0$, if $A \subset [N]^d$ has size $|A| \geq \delta N^d$ and $B_1, \dots, B_d \subset \mathbb{N}$ are M-homogeneous, then there exists $x \in \mathbb{Z}^d$ and $y_1 \in B_1, \dots, y_d \in B_d$ such that*

$$x + y^{\otimes k} \otimes F \subset A. \tag{6.27}$$

Proof. Let $K := \prod_i k_i$ and consider the finite set

$$F' := [M^K]^d \otimes F.$$

By the Bergelson–Leibman theorem, provided that N is sufficiently large in terms of M, K, F and δ, there exist $x \in \mathbb{Z}^d$ and $t \in \mathbb{N}$ such that

$$x + t^K \cdot F' \subset A.$$

The result follows if the progression $t^K \cdot [M^K]$ contains an element of the form $y_i^{k_i}$ for some $y_i \in B_i$.

Let $z_i := t^{K/k_i}$. Then

$$\left\{z_i^{k_i}, (2z_i)^{k_i}, \dots, (Mz_i)^{k_i}\right\} = t^K \cdot \left\{1^{k_i}, 2^{k_i}, \dots, M^{k_i}\right\} \subset t^K \cdot [M^K].$$

Since each B_i is M-homogeneous, it intersects the set $z_i \cdot [M]$. □

Next we require a counting analogue of this result.

Theorem 6.6.3 (Varnavides averaging)**.** *Let $k_1, \dots, k_d, M \in \mathbb{N}$, $\delta \in (0,1]$, and let $F \subset \mathbb{Z}^d$ be a finite set. There exist $N_0 \in \mathbb{N}$ and $c_0 > 0$ such that for any $N \geq N_0$, if $A \subset [N]^d$ has $|A| \geq \delta N^d$ and $B \subset \mathbb{N}$ is M-homogeneous, then the number of tuples $(x, y) \in \mathbb{Z}^d \times B^d$ for which (6.27) holds is at least*

$$c_0 N^{d + \frac{1}{k_1} + \dots + \frac{1}{k_d}}.$$

Proof. Increasing the size of F if necessary, we may assume that F contains two elements which differ in the ith coordinate for each $i \in [d]$. Let N_0 be the quantity given by Corollary 6.6.2 with respect to the density $\delta/2^{d+1}$. Suppose that

$$N \geq N_0,$$

and let

$$N_i := \left\lfloor \sqrt[k_i]{N/N_0} \right\rfloor.$$

Interchanging the order of summation, we have

$$\sum_{z \in \mathbb{Z}^d} \sum_{q_1 \in [N_1]} \cdots \sum_{q_d \in [N_d]} |A \cap \left(z + q^{\otimes k} \otimes [N_0]^d\right)| \geq \delta N_1 \cdots N_d (NN_0)^d.$$

Notice that there are at most $(2N)^d$ choices for z for which there exists $q \in [N_1] \times \cdots \times [N_d]$ such that

$$|A \cap \left(z + q^{\otimes k} \otimes [N_0]^d\right)| \neq 0.$$

Hence there are at least $\frac{1}{2}\delta N^d N_1 \cdots N_d$ choices for $(z, q) \in \mathbb{Z}^d \times \prod_i [N_i]$ for which

$$|A \cap \left(z + q^{\otimes k} \otimes [N_0]^d\right)| \geq 2^{-d-1} \delta N_0^d. \tag{6.28}$$

Call such a choice of (z, q) a *good* tuple.

Claim 1. For each good tuple (z, q) the set $A \cap \left(z + q^{\otimes k} \otimes [N_0]^d\right)$ contains a configuration of the form $x + y^{\otimes k} \otimes F$ for some $x \in \mathbb{Z}^d$ and some $y \in B^d$.

To see this, define

$$A_{z,q} := \left\{x \in [N_0]^d : z + q^{\otimes k} \otimes x \in A\right\}.$$

Then $|A_{z,q}| \geq 2^{-d-1}\delta N_0^d$. Let

$$B_i = \{y_i \in [N_0] : q_i y_i \in B\} \cup (N_0, \infty).$$

Using the fact that B is M-homogeneous one gets that each B_i is M-homogeneous. Invoking Corollary 6.6.2, we see that there exist $x \in \mathbb{Z}^d$ and $y \in B_1 \times \cdots \times B_d$ such that

$$x + y^{\otimes k} \otimes F \subset A_{z,q}.$$

Translating and dilating, we deduce that $A \cap \left(z + q^{\otimes k} \otimes [N_0]^d\right)$ contains a configuration of the form $x' + (q \otimes y)^{\otimes k} \otimes F$. By definition of the B_i and the fact that F is non-constant in each coordinate, we see that $y \in [N_0]^d$ and thus each coordinate of $q \otimes y$ lies in B. This establishes Claim 1.

For fixed $(x, y) \in \mathbb{Z}^d \times \mathbb{N}^d$ let $G(x, y)$ denote the number of tuples $(z, q) \in \mathbb{Z}^d \times \mathbb{N}^d$ satisfying

$$x + y^{\otimes k} \otimes F \subset z + q^{\otimes k} \otimes [N_0]^d. \tag{6.29}$$

Define

$$\mathcal{A} := \left\{(x, y) \in \mathbb{Z}^d \times B^d : x + y^{\otimes k} \otimes F \subset A\right\}. \tag{6.30}$$

Then interchanging the order of summation shows that the sum $\sum_{(x,y)\in\mathcal{A}} G(x, y)$ is at least

$$\sum_{z\in\mathbb{Z}^d} \sum_{q_1\in[N_1]} \cdots \sum_{q_d\in[N_d]} \left|\left\{(x, y) \in \mathbb{Z}^d \times B^d : x + y^{\otimes k} \otimes F \subset A \cap \left(z + q^{\otimes k} \otimes [N_0]^d\right)\right\}\right|$$
$$\geq \left|\left\{(z, q) \in \mathbb{Z}^d \times \prod_i [N_i] : (z, q) \text{ is good}\right\}\right| \geq \tfrac{1}{2}\delta N^d N_1 \cdots N_d.$$

We therefore have that

$$\sum_{(x,y)\in\mathcal{A}} G(x, y) \gg_{k,\delta,N_0} N^{d+\frac{1}{k_1}+\cdots+\frac{1}{k_d}}.$$

Since the theorem asserts a lower bound on the size of $\mathcal{A}$, the result is proved provided we have the following upper bound on $G(x, y)$.

Claim 2. Suppose that F contains two elements which differ in the ith coordinate for each $i \in [d]$. Then $G(x, y) \leq N_0^{2d}$.

To see this, first note that if $x + y^{\otimes k} \otimes F \subset z + q^{\otimes k} \otimes [N_0]^d$ then, since F contains two elements differing in their ith coordinate, there exist integers $f_i < f_i'$ such that

$$x_i + y_i^{k_i} f_i,\ x_i + y_i^{k_i} f_i' \in z_i + q_i^{k_i} \cdot [N_0].$$

Subtracting these elements, we deduce that there exists $n_i \in [N_0]$ for which

$$q_i^{k_i} = \frac{y_i^{k_i}(f_i' - f_i)}{n_i}.$$

As there are at most N_0 choices for n_i, and y_i is fixed, there are at most N_0^d choices for q. Once one has fixed this choice of q, for any $f \in F$ we have

$$z \in x + y^{\otimes k} \otimes f - q^{\otimes k} \otimes [N_0]^d,$$

so there are at most N_0^d choices for z. In summary $G(x, y) \leq N_0^{2d}$, which establishes Claim 2. □

6.7 A supersaturated generalisation of both Roth and Sárközy's theorems

In this section we deduce Theorem 6.2.4 by projecting down the multidimensional Theorem 6.6.3. Notice that Theorem 6.2.4 is a common generalisation of both the Furstenberg–Sárközy theorem (take $s = 2$ and $t = 1$) and Roth's theorem (take $\lambda = (1, -2, 1)$ and $t = 0$).

Proof of Theorem 6.2.4. Given $A \subset [N]$ of density at least δ, let us define

$$\widetilde{A} := \Big\{ x \in [N]^{s+t-2} : \sum_i x_i \in A \Big\}.$$

A stars and bars argument shows that for $n \in [N]$ we have

$$\#\left\{(n_1, \dots, n_d) \in [N]^d : n = n_1 + \cdots + n_d\right\} = \binom{n-1}{d-1}.$$

Since there are at most $\frac{1}{2}|A|$ elements x of A satisfying the inequality $x \leq \frac{1}{2}|A|$, it follows that for $N \geq C_{s,t}\delta^{-1}$ we have

$$|\widetilde{A}| = \sum_{n \in A} \binom{n-1}{s+t-3} \gg_{s,t} \delta^{s+t-2} N^{s+t-2}. \tag{6.31}$$

In the statement of Theorem 6.2.4, at least one of the coefficients λ_i must be positive. Relabeling indices, we may assume that $\lambda_s > 0$. For a technical reason, it will be useful in a later part of the argument if we can ensure that

$$\widetilde{A} - \widetilde{A} \subset \left[-\tfrac{N}{\lambda_s}, \tfrac{N}{\lambda_s}\right]^{s+t-2}. \tag{6.32}$$

This follows on partitioning the hypercube $[N]^{s+t-2}$ into subhypercubes of sufficiently small side length and applying the pigeonhole principle to ensure that $\widetilde{A}$ has large density on one such part (worsening the density (6.31) by a factor of $O_{s,t,\lambda_s}(1)$ in the process).

Define $F \subset \mathbb{Z}^{s+t-2}$ to be the set consisting of the zero vector together with the rows of the following matrix

$$\begin{pmatrix} -\lambda_s & & & & & & \\ & -\lambda_s & & & & & \\ & & \ddots & & & & \\ & & & -\lambda_s & & & \\ \lambda_1 & \lambda_2 & \dots & \lambda_{s-2} & \lambda_s^{k-1}\mu_1 & \dots & \lambda_s^{k-1}\mu_t \end{pmatrix}. \tag{6.33}$$

Consider the set

$$\widetilde{B} := \{y \in \mathbb{N} : \lambda_s y \in B\} \cup (N^{1/k}\lambda_s^{-1}, \infty)$$

which is M-homogeneous. Applying Theorem 6.6.3, we find that there are at least $c_0 N^{s+t-2+s-2+\frac{t}{k}}$ tuples $(x, y, z) \in \mathbb{Z}^{s+t-2} \times \widetilde{B}^{s-2} \times \widetilde{B}^t$ such that $\widetilde{A}$ contains the configuration

$$x + (y_1, \dots, y_{s-2}, z_1^k, \dots, z_t^k) \otimes F.$$

By (6.32) and (6.33) we have $\lambda_s^k \mu_i z_i^k \in [-N, N]$, hence by definition of $\widetilde{B}$ we deduce that $\lambda_s z_i \in B$. Projecting down to one dimension and taking into account the multiplicities of representations, we obtain $\gg N^{s+\frac{t}{k}-1}$ tuples $(x, y, z) \in \mathbb{Z} \times \mathbb{N}^{s-2} \times \mathbb{N}^t$ with $\lambda_s z_i \in B$ and such that A contains the configuration

$$x, \quad x - \lambda_s y_1, \quad \dots, \quad x - \lambda_s y_{s-2}, \quad x + \sum_{i=1}^{s-2} \lambda_i y_i + \lambda_s^{k-1} \sum_{j=1}^{t} \mu_j z_j^k.$$

Let us set $x_i := x - \lambda_s y_i$ for $i = 1, \dots, s-2$, along with $x_{s-1} = x$ and

$$x_s := x + \sum_{i=1}^{s-2} \lambda_i y_i + \lambda_s^{k-1} \sum_{j=1}^{t} \mu_j z_j^k.$$

One can then check that the tuple $(x_1, \dots, x_s, \lambda_s z_1, \dots, \lambda_s z_t)$ is an element of $A^s \times B^t$ satisfying (6.9). By construction there are $\gg N^{s+\frac{t}{k}-1}$ such tuples. □

Chapter 7

Counting zeros of a quadratic form in four variables with a weight

7.1 Introduction

Let $F(\mathbf{x})$ be a quadratic form in n variables with integer coefficients, and consider the problem of counting the number of ways in which F represents 0. Recall the quantity

$$N(P) = \#\big\{\mathbf{x} \in \mathbb{Z}^n : F(\mathbf{x}) = 0, \|\mathbf{x}\|_\infty \leqslant P\big\}, \tag{7.1}$$

from Section 2.4. This counts the number of solutions to $F(\mathbf{x}) = 0$ in a box with side lengths $2P$. To avoid special cases we will only be interested in non-singular forms F. That is, if M is the underlying matrix of F such that $F(\mathbf{x}) = \mathbf{x}^T M\mathbf{x}$, we require that $\det M \neq 0$.

One way of tackling this problem is by the Hardy–Littlewood circle method, as done in e.g. [17, chapters 8,10]. The circle method in its classical form allows one to give an asymptotic for $N(P)$ in the cases $n \geqslant 5$. For these cases it turns out that a simple heuristic argument predicts the correct asymptotic order. The heuristic argument is as follows. There are roughly P^n different values of $\mathbf{x} \in \mathbb{Z}$ with $\|\mathbf{x}\|_\infty \leqslant P$, and $F([-P, P]^n \cap \mathbb{Z})$ should spread over roughly P^2 different values. Based on this we expect CP^{n-2} of these $\mathbf{x}$'s to satisfy $F(\mathbf{x}) = 0$, for some constant $C > 0$. The precise asymptotic is given by Theorem 2.4.1.

We also mention that one often is concerned with showing that $\sigma_\infty\sigma$ actually is positive. In this chapter we will assume that we are dealing with quadratic forms F which satisfy $\sigma_\infty > 0$ and $\sigma > 0$, as we are concerned with the asymptotic expression for quadratic forms that *do* represent zero, rather than the question of whether or not a specific form does represent zero.

When one instead looks at $n = 4$, it turns out that the correct asymptotic expression does not always agree with the heuristic argument. The correct asymptotic expression in fact depends on the determinant of the underlying matrix M.

Let

$$N_w(P) = \sum_{F(\mathbf{x})=0} w(P^{-1}\mathbf{x}), \tag{7.2}$$

where $w : \mathbb{R}^n \to [0, \infty)$ is some sufficiently smooth function and we are summing over all $\mathbf{x} \in \mathbb{Z}^n$ such that $F(\mathbf{x}) = 0$. The asymptotic count for $N_w(P)$ is then given in Theorem 2.4.2.

In this chapter our main concern is the size of the error term in the expression for $N_w(P)$ for $n = 4$ variables and $\det(M)$ a square. If one just considers $N(P)$ it is clear that the error term must at least be of the same order as $N(P) - N(P-1)$, as $\mathbf{x}$ only runs over integers. This difference counts the number of solutions to $F(\mathbf{x}) = 0$ where at least one of x_1, x_2, x_3, x_4 is equal to P. Heuristically speaking there are then P^3 choices for $\mathbf{x}$, and the values of F are spread over a range of P^2, such that the error term is expected to be of size $P^{3-2} = P$, at the very least.

There is no such natural limitation for $N_w(P)$, and we are therefore interested in whether one can find some class of weight functions for which the error term in (2.7) can be made an arbitrarily small power of P. The aim of this chapter is to answer this question. First of all, we will consider a rather natural Gaussian weight function and show that in this case it appears to be very difficult to obtain an error of the form P^α with $\alpha < 1$. From this consideration it becomes clear how we can construct some weight function that gives an error term $O_A(P^{-A})$ for any $A > 0$, which answers the original question. Finally, we elaborate on the proof of Theorem 2.4.2 given in [37] in order to get an error term of the form $O_\epsilon(P^{1+\epsilon})$ for the general class of weight functions considered in Theorem 2.4.2. The same error was obtained in [26, Theorem 3.3] by adelic methods.

7.1.1 Notation

In this chapter we will not explicitly include the dependence in F of various functions and error terms. Any constants appearing may therefore depend on F.

7.2 Gaussian weight function

In this section and the next we fix a diagonal quadratic form in four variables. Let

$$F(\mathbf{x}) = x_1^2 + x_2^2 - x_3^2 - x_4^2, \tag{7.3}$$

such that $\det M = 1$ is a square. Define $N_w(P)$ as in (7.2). In this section we consider the weight function given by

$$w(\mathbf{x}) = \exp\left(-\mathbf{x}^T\mathbf{x}\right). \tag{7.4}$$

Although this weight does not have compact support and thus does not satisfy the assumptions in Theorem 2.4.2, it still seems like a natural weight to consider.

The advantage of using this Gaussian weight is that it gives a simple expression for $N_w(P)$ as a sum over only one variable. To see this we write

$$\begin{aligned} N_w(P) &= \sum_{x_1^2+x_2^2=x_3^2+x_4^2} \exp(-P^{-2}\mathbf{x}^T\mathbf{x}) \\ &= \sum_{n=0}^{\infty} \sum_{x_1^2+x_2^2=x_3^2+x_4^2=n} e^{-2nP^{-2}} \\ &= 1 + \sum_{n=1}^{\infty} r^2(n) e^{-2nP^{-2}}, \end{aligned}$$

where

$$r(n) \coloneqq \#\left\{(x,y) \in \mathbb{Z}^2 : x^2+y^2 = n\right\}$$

is the number of ways of representing n as the sum of two squares. Note that we can write

$$r(n) = 4\sum_{d|n} \chi_4(d), \tag{7.5}$$

where χ_4 is the non-principal character modulo 4, as shown in e.g. [36].

We will need this in the context of the Dirichlet series $\sum_n r^2(n)n^{-s}$. Restricting to $\Re s > 1$, from (7.5) we get

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{r^2(n)}{n^s} &= 16\sum_{n=1}^{\infty} \Big(\sum_{d|n} \chi_4(d)\Big)^2 n^{-s} \\ &= 16\sum_{m=1}^{\infty}\sum_{k=1}^{\infty}\sum_{d=1}^{\infty}\sum_{e=1,(e,d)=1}^{\infty} \chi_4(kd)\chi_4(ke)(mkde)^{-s} \\ &= 16\zeta(s)\Big(\sum_{k=1}^{\infty} \chi_4(k)^2 k^{-s}\Big) \sum_{(d,e)=1} \chi_4(ed)(ed)^{-s} \\ &= 16\zeta(s)L(s,\chi_0)\sum_{n=1}^{\infty} 2^{\omega(n)}\chi_4(n)n^{-s}, \end{aligned}$$

where $\chi_0 = \chi_4^2$ is the principal character modulo 2 and $\omega(n)$ is the number of distinct prime factors of n. The last sum we express as an Euler product to get

$$\sum_{n=1}^{\infty} 2^{\omega(n)}\chi_4(n)n^{-s} = \prod_p (1+\chi_4(p)p^{-s})(1+\chi_4(p)p^{-s}+\chi_4(p)^2p^{-2s}+\cdots) = \frac{L(s,\chi_4)^2}{L(2s,\chi_4^2)}.$$

Combining all of the above we now have

$$\sum_{n=1}^{\infty} \frac{r^2(n)}{n^s} = 16\zeta(s)L(s,\chi_0)L(s,\chi_4)^2L(2s,\chi_0)^{-1}. \tag{7.6}$$

Returning to $N_w(P)$, we have by the Mellin inversion formula for $\Gamma(s)$ that

$$e^{-t} = \frac{1}{2\pi i}\int_{(c)} t^{-s}\Gamma(s)\,ds,$$

where (c) denotes the vertical line with real part c, and we choose $c = 2$, say. This gives

$$\begin{aligned} N_w(P) &= 1 + \sum_{n=1}^{\infty} \frac{1}{2\pi i}\int_{(c)} 2^{-s}P^{2s}\Gamma(s)r^2(n)n^{-s}\,ds \\ &= 1 + \frac{16}{2\pi i}\int_{(c)} 2^{-s}P^{2s}\zeta(s)L(s,\chi_0)L(s,\chi_4)^2L(2s,\chi_0)^{-1}\Gamma(s)\,ds, \end{aligned} \tag{7.7}$$

where we have used (7.6) and interchanged the order of integration and summation. This is justified as the integrand is $\ll_A P^4n^{\epsilon-2}t^{-A}$ for any $A > 0$.

At this point we evaluate the integral in (7.7) by moving the path of integration as far left as possible. In order to do this we need to know the residues of the integrand. From $\zeta(s)L(s,\chi_0)$ one gets a second order pole at $s = 1$. The factor $L(s,\chi_4)^2$ does not contribute any poles, as χ_4 is a non-principal character. We will only be interested in the region $\sigma \geqslant \frac{1}{2}+\epsilon$, so Γ does not contribute any poles. Finally, there will be a pole at any point where $L(2s,\chi_0) = 0$, and it is these poles that will limit how far to the left we can move the contour. Using the fact that $L(2s,\chi_0)$ is zero-free for $2\sigma > 1$, we will next prove the following result.

Theorem 7.2.1. *Let F and w be defined as in (7.3) and (7.4). Then*

$$N_w(P) = CP^2\log P + C'P^2 + O_\epsilon(P^{1+\epsilon})$$

for any $\epsilon > 0$ and some constants C, C'.

Remark. Note that strictly speaking one could probably obtain an error term of the same form as in the prime number theorem, by using that $L(2s,\chi_0)$ has no zeros for $2\sigma \geqslant 1 - \frac{c}{\log t}$, $t > 2$, but here we are mainly concerned with the power of P, and are thus satisfied with the error term stated above.

Proof of Theorem 7.2.1. Starting from (7.7), we would like to move the line of integration to $\sigma = \frac{1}{2} + \epsilon$. For this purpose we recall some growth properties of the functions involved. First of all,

$$\zeta(\sigma + it) \ll_\epsilon |t|^{1/2}, \quad \sigma \geqslant \epsilon, \tag{7.8}$$

as shown in e.g. [63, Equation (5.1.5)]. We will also use the standard convexity bound for a character χ modulo q, namely that

$$L(s, \chi) \ll_\epsilon (q(1 + |t|))^{\frac{1-\sigma}{2}+\epsilon}, \quad \epsilon \leqslant \sigma \leqslant 1, \tag{7.9}$$

and we will need the bound

$$L(s, \chi) \gg_\epsilon 1, \quad \sigma \geqslant 1 + \epsilon, \tag{7.10}$$

all valid for any $\epsilon > 0$. Finally, we have from Stirling's formula that

$$\Gamma(s) \ll_A |t|^{-A} \tag{7.11}$$

as $t \to \infty$, for any $A > 0$. Denote the integrand in (7.7) by $u(s)$. By choosing $A = 4$, say, we then have

$$\int_{c\pm iT}^{c\pm i\infty} u(s)\, ds \ll P^4 \int_T^\infty t^{-4}\, dt \ll P^4 T^{-3},$$
$$\int_{1/2+\epsilon-iT}^{1/2+\epsilon+iT} u(s)\, ds \ll P^{1+2\epsilon} \int_{-T}^{T} (1 + |t|)^{-2}\, dt \ll_\epsilon P^{1+2\epsilon}$$

and

$$\int_{1/2+\epsilon+iT}^{c+iT} u(s)\, ds \ll P^4 T^{-2}.$$

Letting $T \to \infty$ in these formulae then gives

$$N_w(P) = 16\mathrm{Res}_{s=1} + O_\epsilon(P^{1+2\epsilon}),$$

where $\mathrm{Res}_{s=1}$ denotes the residue of $u(s)$ at $s = 1$.

For the residue we see that $\zeta(s)$ and $L(s, \chi_0)$ both have first order poles at $s = 1$, and the rest of the integrand is nonzero here. Thus

$$\mathrm{Res}_{s=1} = \lim_{s\to 1} \frac{d}{ds}\big((s-1)^2 2^{-s} P^{2s} \Gamma(s)\zeta(s)L(s, \chi_0)L(s, \chi_4)^2 L(2s, \chi_0)^{-1}\big)$$
$$= CP^2 \log P + C'P^2$$

for some constants C, C' which can be expressed explicitly if desired. Upon redefining C, C' and ϵ appropriately this finishes the proof. □

The above proof indicates that obtaining an error term of lower order than $P^{1+\epsilon}$ for this particular weight should be very difficult, as it appears to be connected with a zero-free strip $\sigma > 1-\alpha$ for $L(s,\chi_0)$. Indeed, if one knew that $L(s,\chi_0)$ had no zeros on this strip, one could move the contour to $2\sigma = 1-\alpha+\epsilon$, and obtain a final error of $O_\epsilon(P^{1-\alpha+\epsilon})$. The Generalised Riemann hypothesis would therefore immediately give an error term of size $O_\epsilon(P^{1/2+\epsilon})$.

Of course we cannot exclude the possibility of obtaining an error of $O_\epsilon(P^{1-\alpha+\epsilon})$ without also proving that the strip $\sigma \geqslant 1-\alpha$ is zero-free, as different zeros conceivably could cancel each other out.

7.3 Weight function with low error term

In the previous section, if it were not for the fact that $L(2s,\chi_0)$ appeared in the denominator in (7.6), we could have obtained a much lower error term in Theorem 7.2.1. This issue can be avoided if we choose our weight function in such a way that the resulting Dirichlet series in the integral is as in Section 7.2, but multiplied by $L(2s,\chi_0)$. We see that

$$\sum_{n=1}^{\infty}\frac{r^2(n)}{n^s}L(2s,\chi_0) = \sum_{n=1}^{\infty}r^2(n)\sum_{m=1}^{\infty}\chi_0(m)(nm^2)^{-s},$$

and so we try to set

$$w(\mathbf{x}) = \sum_{m=1}^{\infty}\chi_0(m)\exp(-(\mathbf{x}^T\mathbf{x})m^2) = \sum_{m=1}^{\infty}\exp(-(\mathbf{x}^T\mathbf{x})(2m-1)^2). \tag{7.12}$$

Unfortunately $w(\mathbf{x})$ diverges as $\mathbf{x}^T\mathbf{x}\to 0$, which is an undesirable property of a weight function. To remedy this we try a weight function written as linear combinations of sums of the form

$$\sum_{m=1}^{\infty}f(m^2\mathbf{x}^T\mathbf{x}),$$

where f is some function with a nice Mellin transform. In fact, the choice

$$w(\mathbf{x}) = W(\|\mathbf{x}\|_2^2) \tag{7.13}$$

with

$$W(t) = \theta(t)+\theta(2t)-2\theta(4t)-2\theta(8t)$$

and

$$\theta(t) = \sum_{m=1}^{\infty}\exp(-tm^2),$$

will work. We begin by showing that w has some reasonable properties one would expect from a weight function.

Proposition 7.3.1. *Let W be defined as in (7.12). Then $\lim_{t\to 0} W(t) = 1$, $W \geqslant 0$ for $t \in (0,\infty)$ and $\lim_{t\to\infty} W(t) = 0$.*

Proof. Grouping together terms we have that

$$\theta(t) - 2\theta(4t) = \sum_{m=1}^{\infty} (-1)^{m+1} \exp(-tm^2),$$

and since $\exp(-tm^2) - \exp(-t(m+1)^2) > 0$ for $t > 0$ and all $m \geqslant 1$ we have that $\theta(t) - 2\theta(4t) > 0$ for $t \in (0,\infty)$. The same holds if we replace t by $2t$ and so $W(t) > 0$ for $t \in (0,\infty)$. For the case $t \to \infty$ we have that

$$|\theta(t) - 2\theta(4t)| \leqslant \sum_{m=1}^{\infty} \exp(-tm^2) \leqslant \int_0^{\infty} \exp(-tx^2)\, dx = \left(\frac{\pi}{t}\right)^{1/2} \to 0$$

as $t \to \infty$. Again the same holds when replacing t by $2t$, and thus also for $W(t)$.

To study the behaviour of W as t tends to 0, we make use of the closely related theta function

$$\vartheta(x) = \sum_{m\in\mathbb{Z}} \exp(-m^2\pi x),$$

which satisfies the functional equation

$$\vartheta(x^{-1}) = \sqrt{x}\vartheta(x) \quad \text{for } x > 0,$$

as shown in e.g. [18, Chapter 8]. Again we deal first with $\theta(t) - 2\theta(4t)$. Writing

$$\theta(t) - 2\theta(4t) = \frac{1}{2}\vartheta(t/\pi) - \vartheta(4t/\pi) + \frac{1}{2},$$

and making use of the functional equation we then get

$$\begin{aligned}
\theta(t) - 2\theta(4t) &= \frac{1}{2}\frac{\sqrt{\pi}}{t^{1/2}}\left(\vartheta(\pi/t) - \vartheta(\pi/4t)\right) + \frac{1}{2} \\
&= \frac{\sqrt{\pi}}{t^{1/2}} \sum_{m=1}^{\infty} \left(\exp(-m^2\pi^2/t) - \exp(-m^2\pi^2/4t)\right) + \frac{1}{2} \\
&= -\frac{\sqrt{\pi}}{t^{1/2}} \sum_{m \text{ odd}} \exp(-m^2\pi^2/4t) + \frac{1}{2}.
\end{aligned}$$

For this final sum we have

$$\begin{aligned}
t^{-1/2} \sum_{m \text{ odd}} \exp(-m^2\pi^2/4t) &\leqslant t^{-1/2} \exp(-\pi^2/4t) + t^{-1/2} \int_1^{\infty} \exp(-x^2\pi^2/4t)\, dx \\
&= t^{-1/2} \exp(-\pi^2/4t) + \frac{2}{\pi} \int_{\pi t^{-1/2}/2}^{\infty} \exp(-x^2)\, dx,
\end{aligned}$$

which goes to 0 as $t \to 0$. This in turn gives

$$\lim_{t\to 0} W(t) = \lim_{t\to 0} \big(\theta(t) - 2\theta(4t)\big) + \lim_{t\to 0} \big(\theta(2t) - 2\theta(8t)\big) = 1.$$

□

In order to find an asymptotic expression for $N_w(P)$, with F still given by (7.3), we do the same type of manipulations as in Section 7.2 to get

$$\begin{aligned} N_w(P) &= 1 + \sum_{n=1}^{\infty} r^2(n) \sum_{m=1}^{\infty} (-1)^{m+1} \big(\exp(-2m^2 n/P^2) + \exp(-4m^2 n/P^2) \big) \\ &= 1 + \frac{1}{2\pi i} \int_{(c)} P^{2s} \big(2^{-s} - 2^{1-3s}\big)\big(1 + 2^{-s}\big)\Gamma(s) \bigg(\sum_{n=1}^{\infty} \frac{r^2(n)}{n^s} \bigg) \bigg(\sum_{m=1}^{\infty} m^{-2s} \bigg)\, ds \\ &= 1 + \frac{16}{2\pi i} \int_{(c)} P^{2s} \big(2^{-s} - 2^{1-3s}\big)\Gamma(s)\zeta(s)^2 L(s,\chi_4)^2\, ds, \end{aligned} \tag{7.14}$$

where we use (7.6) and $L(s,\chi_0) = \zeta(s)(1-2^{-s})$ to remove all instances of $L(\cdot,\chi_0)$ in the last equality. Using this we will prove the following theorem.

Theorem 7.3.2. *Let w and F be defined as in (7.13) and (7.3) respectively. Then*

$$N_w(P) = CP^2 \log P + C'P^2 + C'' + O_A(P^{-A})$$

for any $A > 0$, where C, C', C'' are some constants depending only on F and w.

Proof. The proof is similar to the proof of Theorem 7.2.1, except that there no longer is a $L(2s,\chi_0)^{-1}$ term. This means that we now can move the line of integration to $\sigma = -A$ for any $A > 0$ without worrying about zeros. We also have that the poles of $\Gamma(s)$ at $-1, -2, \dots$ will be cancelled by the trivial zeros of $\zeta(s)$ and $L(s,\chi_4)$. At $s = 0$ the factor $\Gamma(s)$ contributes a simple pole.

In addition to the bounds given in (7.8)-(7.11) we will need that for $\sigma \geqslant -A$

$$\begin{aligned} \zeta(s) &\ll_A |t|^{\frac{1}{2}+A}, \\ L(s,\chi) &\ll_A |t|^{\frac{1}{2}+A}, \end{aligned}$$

for any $A > 0$, with $\chi = \chi_4$. Observe that since χ_4 is an odd character, $L(s,\chi_4)$ has trivial zeros at the even negative integers. Denote the integrand in (7.14) by $u(s)$. By choosing $4A + 4$ as the exponent in (7.11) and setting $c = 2$ we then have

$$\begin{aligned} \int_{c\pm iT}^{c\pm i\infty} u(s)\, ds &\ll P^4 \int_T^{\infty} t^{-2}\, dt \ll P^4 T^{-1}, \\ \int_{-A-iT}^{-A+iT} u(s)\, ds &\ll_A P^{-A} \int_{-T}^{T} (1+|t|)^{-2}\, dt \ll_A P^{-A}, \end{aligned}$$

and

$$\int_{-A\pm iT}^{c\pm iT} u(s)\,ds \ll_A P^4 T^{-2}.$$

By closing the contour in (7.14) around the box with edges $\sigma = -A, \sigma = c$ and $t = \pm T$ and choosing T appropriately, we get

$$N_w(P) = 1 + 16(\mathrm{Res}_{s=1} + \mathrm{Res}_{s=0}) + O_A(P^{-A}),$$

As before the residue at $s = 1$ can be written in the form $CP^2 \log P + C'P^2$. For the residue at $s = 0$ we similarly get something of the form C''. □

7.4 Improving the error term by the Heath-Brown circle method

We now turn to Theorem 2.4.2. The goal of this section is to refine the argument given in [37] in order to improve the error term to $O_\epsilon(P^{1+\epsilon})$. As mentioned in the introduction this was also done by Getz in [26]. Using adelic methods he proves a version of Theorem 2.4.2 with an error term of order $O_\epsilon(P^{1+\epsilon})$, valid over any number field. He also improves the existing best known error terms for $n > 4$ by the same methods.

We will only prove this for the case where $\det M$ is a square, although a very similar argument should give the same improvement in the case where $\det M$ is not a square.

Most of the argument will be the same as is given in [37], so we state a series of lemmas taken from there. In order to do this we first need to settle some notation.

Definition. Let $w : \mathbb{R}^n \to \mathbb{C}$ be an infinitely differentiable function with compact support. Let $\mathrm{Rad}(w)$ be the smallest R such that the support of w is contained in the hypercube $[-R, R]^n$, and for $j = 0, 1, \ldots$ let

$$\kappa_j(w) = \max\left\{\left|\frac{\partial^{j_1+\cdots+j_n} w}{\partial^{j_1} x_1 \cdots \partial^{j_n} x_n}\right| : \mathbf{x} \in \mathbb{R}^n, \sum_i j_i = j\right\}.$$

If each of $\mathrm{Rad}(w), \kappa_0(w), \kappa_1(w), \ldots$ can be bounded by quantities depending only on the parameters in some set S we say that $w \in \mathcal{C}(S)$.

We will prove our result for weight functions in $\mathcal{C}(S)$. For technical reasons we also impose a non-singularity condition, namely $\nabla F(\mathbf{x}) \neq \mathbf{0}$ in the closure of $\mathrm{supp}(w)$. In particular this assumption implies that $0 \notin \mathrm{supp}(w)$. The result we seek to prove can then be stated as follows.

Theorem 7.4.1. *Let $n = 4$, assume that $w \in \mathcal{C}(S)$ and that $\det M$ is a square. Then*

$$N_w(P) = \sigma_\infty(w)\sigma^* P^2 \log P + \sigma_1(w)P^2 + O_{S,\epsilon}(P^{1+\epsilon}) \tag{7.15}$$

for any $\epsilon > 0$. Here $\sigma_\infty(w), \sigma^*$ and $\sigma_1(w)$ are defined as in Theorem 2.4.2.

We will start from Theorem 2.4.4. The improvements we make on the error term come mainly from considerations of the sum $S_q(\mathbf{c})$, in combination with existing results for $I_q(\mathbf{c})$ extended to also hold for various partial derivatives. We start by citing the needed results concerning $I_q(\mathbf{c})$.

7.4.1 Bounds on $I_q(\mathbf{c})$

When dealing with the integral $I_q(\mathbf{c})$ it is technically easier to restrict oneself to a smaller class of weight functions than $\mathcal{C}(S)$, namely $\mathcal{C}_0(S)$.

Definition. We say that $w \in \mathcal{C}_0(S)$ if $w \in \mathcal{C}(S)$ and there exist an $R \ll_S 1$ with the following property. Whenever $\mathbf{x} = (x_0, \mathbf{y}) \in \mathrm{supp}(w)$, the function $F(x_0, \mathbf{y})$ satisfies

$$\frac{\partial F(x, \mathbf{y})}{\partial x} \gg_S 1$$

in the range $|x - x_0| \leqslant R$, and F has exactly one zero on this range.

It turns out that if $|\nabla F| \geqslant \lambda > 0$ holds on $\mathrm{supp}(w)$, then proving Theorem 7.4.1 only for weight functions in $\mathcal{C}_0(S)$ will imply the same result for any weight function in $\mathcal{C}(S, \lambda)$. The justification for this can be found in [37], and we will not repeat it here. From now on we therefore assume that the weight functions we work with are in $C_0(S)$. This technicality is mentioned to allow us to quote results concerning $I_q(\mathbf{c})$ without worrying about the fact that they strictly speaking only hold for $w \in \mathcal{C}_0(S)$, but not necessarily for $w \in \mathcal{C}(S)$.

In order to make the P-dependency of $I_q(\mathbf{c})$ more explicit, define

$$I_r^*(\mathbf{v}) = \int_{\mathbb{R}^n} w(\mathbf{x})h(r, F(\mathbf{x}))e_r(-\mathbf{v}\cdot\mathbf{x})\, d\mathbf{x},$$

such that $I_q(\mathbf{c}) = P^n I^*_{P^{-1}q}(\mathbf{c})$ by the substitution $\mathbf{x} \mapsto P\mathbf{x}$ in (2.10). We have the following bound from [37].

Lemma 7.4.2 ([37, Lemma 18])**.** *When $\mathbf{c} \neq \mathbf{0}$ we have*

$$I_r^*(\mathbf{c}) \ll_{S,N} r^{-1}|\mathbf{c}|^{-N},$$

for any $N > 0$.

The next two results for $I_r^*(\mathbf{c})$ that we take from [37] need to be extended slightly in order to deal with also the first and second partial derivatives with respect to r.

Lemma 7.4.3 ([37, Lemma 13]). *Suppose that $w \in \mathcal{C}_0(S)$. Then for any $N > 0$ we have*

$$I_r^*(\mathbf{0}) = \sigma_\infty(F, w) + O_{S,N}(r^N),$$
$$\frac{\partial^j}{\partial r^j} I_r^*(\mathbf{0}) \ll_{S,N} r^N, \quad j = 1, 2,$$

for $r \ll 1$.

Proof. The first result is just [37, Lemma 13]. The proof of this hinges on properties of the function $h(x, y)$ in Theorem 7.4.1, and it does not require any new ideas to prove similar results for $\frac{\partial^j}{\partial x^j} h(x, y)$, which is what is needed to establish the remaining part of the lemma. We do not show the full proof here. □

Lemma 7.4.4 ([37, Lemma 22]). *Suppose that $n \geqslant 3$. Then for any $\epsilon \in (0, 1/2)$ we have*

$$\frac{\partial^j}{\partial r^j} I_r^*(\mathbf{c}) \ll_{S,\epsilon} r^{-j} \Big(\frac{|\mathbf{c}|}{r^2}\Big)^{\epsilon} \Big(\frac{|\mathbf{c}|}{r}\Big)^{1-n/2}, \quad j = 0, 1, 2.$$

Proof. The cases $j = 0, 1$ are shown in [37]. The ideas used there can be modified in a natural and fairly straight forward way to arrive at the same type of bounds for the second derivative. For brevity we do not repeat this argument here. □

7.4.2 The sum $S_q(\mathbf{c})$

We now turn to $S_q(\mathbf{c})$. To begin with, we cite a series of lemmas proved in [37], which we will build on.

Lemma 7.4.5 ([37, Lemma 24]). *Let $t \geqslant 2$ and $s = [t/2]$. Then*

$$S_{p^t}(\mathbf{c}) = p^{s(n+1)} \sideset{}{^*}\sum_{d \,(\mathrm{mod}\ p^{t-s})} \sum_{\mathbf{x} \,(\mathrm{mod}\ p^{t-s})}^{(1)} e_{p^t}(dF(\mathbf{x}) + \mathbf{x} \cdot \mathbf{c}),$$

where $\sum^{(1)}$ indicates the sum over all $\mathbf{x}$ which satisfy the conditions $p^s | F(\mathbf{x})$ and $p^s | d\nabla F(\mathbf{x}) + \mathbf{c}$.

From now on, let $\Delta = 2 \det(M)$.

Lemma 7.4.6 ([37, Lemma 25]). *One has*

$$S_q(\mathbf{c}) \ll_\Delta q^{1+n/2}.$$

As M has integer coefficients and is assumed to be non-singular we have that M^{-1} has rational coefficients. We write $M^{-1}(\mathbf{c}) \coloneqq \mathbf{c}^T M^{-1}\mathbf{c}$. If $(q, \Delta) = 1$ one can interpret M^{-1} modulo q. In the next lemma $M^{-1}(\mathbf{c})$ should be interpreted modulo p.

Lemma 7.4.7 ([37, Lemma 26])**.** *Let p be a prime such that $p \nmid \Delta$. If n is even then*

$$S_p(\mathbf{c}) = \begin{cases} -\left(\frac{(-1)^{n/2}\det(M)}{p}\right) p^{n/2} & \text{if } p \nmid M^{-1}(\mathbf{c}) \\ (p-1)\left(\frac{(-1)^{n/2}\det(M)}{p}\right) p^{n/2} & \text{if } p | M^{-1}(\mathbf{c}). \end{cases}$$

With this we are ready to prove our results. In [37] Heath-Brown goes on to find an expression for $S_{p^2}(\mathbf{c})$ in various cases. We will use the same type of arguments to find expressions for $S_{p^t}(\mathbf{c})$ for all $t \geqslant 2$.

Lemma 7.4.8. *If $p \nmid \Delta$ and $p \nmid M^{-1}(\mathbf{c})$ then $S_{p^t}(\mathbf{c}) = 0$ for any $t \geqslant 2$.*

Proof. We start with the expression in Lemma 7.4.5. The condition $p^s | d\nabla F(\mathbf{x}) + \mathbf{c}$ in the sum over $\mathbf{x}$ gives

$$\mathbf{x} \equiv -\overline{2d} M^{-1}\mathbf{c} (\text{mod } p^s), \tag{7.16}$$

where an overline denotes the inverse modulo p^s. This in turn gives

$$F(\mathbf{x}) = \mathbf{x}^T M \mathbf{x} \equiv \overline{4d^2} M^{-1}(\mathbf{c}) (\text{mod } p^s), \tag{7.17}$$

but as $M^{-1}(\mathbf{c}) \not\equiv 0 (\text{mod } p^s)$ this is not compatible with the condition $p^s | F(\mathbf{x})$, and so the sum over $\mathbf{x}$ is empty. □

Lemma 7.4.9. *Assume that $p \nmid \Delta$ and $p^t | M^{-1}(\mathbf{c})$. If $t = 2s \geqslant 2$ then*

$$S_{p^t}(\mathbf{c}) = p^{\frac{t}{2}n + t - 1}(p-1).$$

If $t = 2s + 1 \geqslant 3$ then

$$S_{p^t}(\mathbf{c}) = p^{\frac{t-1}{2}n + t - 1} S_p(\mathbf{0}).$$

Proof. Again we start from Lemma 7.4.5, and from (7.16) we can write

$$\mathbf{x} = \mathbf{y}p^s - \overline{2d} M^{-1}\mathbf{c}.$$

By the assumption on $M^{-1}(\mathbf{c})$ and (7.17) we see that the condition $p^s | F(\mathbf{x})$ is automatically satisfied for any $\mathbf{y}$. Computing, we get

$$dF(\mathbf{x}) + \mathbf{c}\cdot\mathbf{x} = dp^{2s}F(\mathbf{y}) - p^s \mathbf{y}\cdot\mathbf{c} + \overline{4d} M^{-1}(\mathbf{c}) + p^s \mathbf{y}\cdot\mathbf{c} - \overline{2d} M^{-1}(\mathbf{c}) \equiv dp^{2s}F(\mathbf{y}) (\text{mod } p^t),$$

where in the last step we use that $M^{-1}(\mathbf{c}) \equiv 0 (\mathrm{mod}\ p^t)$. Inserting this into the expression for $S_{p^t}(\mathbf{c})$ gives

$$S_{p^t}(\mathbf{c}) = p^{s(n+1)} \sideset{}{^*}\sum_{d (\mathrm{mod}\ p^{t-s})} \sum_{\mathbf{y} (\mathrm{mod}\ p^{t-2s})} e_{p^{t-2s}}(dF(\mathbf{y})).$$

If $t = 2s$ is even this reduces to

$$S_{p^{2s}}(\mathbf{c}) = p^{s(n+1)} \sideset{}{^*}\sum_{d (\mathrm{mod}\ p^s)} 1 = p^{s(n+1)} \phi(p^s) = p^{sn+2s-1}(p-1),$$

as desired. If instead $t = 2s+1$ is odd we get

$$\begin{aligned} S_{p^{2s+1}}(\mathbf{c}) &= p^{s(n+1)} \sideset{}{^*}\sum_{d (\mathrm{mod}\ p^{s+1})} \sum_{\mathbf{y} (\mathrm{mod}\ p)} e_p(dF(\mathbf{y})) \\ &= p^{sn+2s} \sideset{}{^*}\sum_{d (\mathrm{mod}\ p)} \sum_{\mathbf{y} (\mathrm{mod}\ p)} e_p(dF(\mathbf{y})), \end{aligned}$$

but the remaining sums are just $S_p(\mathbf{0})$, which gives the desired result. □

The next step is to form the Dirichlet series

$$\zeta(s, \mathbf{c}) = \sum_{q=1}^{\infty} q^{-s} S_q(\mathbf{c}), \tag{7.18}$$

which in light of Lemma 7.4.6 converges absolutely for $\sigma > 2 + n/2$. A standard computation shows that

$$S_{uv}(\mathbf{c}) = S_u(\mathbf{c}) S_v(\mathbf{c})$$

when $(u, v) = 1$, and so we can express $\zeta(s, \mathbf{c})$ as an Euler product. This gives

$$\zeta(s, \mathbf{c}) = \prod_p \Big\{ \sum_{t \geqslant 0} p^{-ts} S_{p^t}(\mathbf{c}) \Big\},$$

where the inner sum over t converges absolutely for $\sigma > 1 + \frac{n}{2}$ by Lemma 7.4.6. We remark that

$$\sum_{t \geqslant 0} p^{-4t} S_{p^t}(\mathbf{0}) = \sigma_p,$$

as defined in (2.4).

Now assume that $n = 4$, $M^{-1}(\mathbf{c}) = 0$ and $\sigma > 3$. From Lemmas 7.4.7 and 7.4.9 one then has for p such that $p \nmid \Delta$ that

$$\begin{aligned} \sum_{t \geqslant 0} p^{-st} S_{p^t}(\mathbf{c}) &= 1 + (p-1)\chi_p p^{2-s} + (p-1)p^{5-2s} + \dots \\ &= \frac{1}{1 - p^{6-2s}} (1 + \chi(p) p^{2-s}(p-1) - p^{5-2s}), \end{aligned}$$

where $\chi(\cdot) = \left(\frac{\det(M)}{\cdot}\right)$. From this point onwards we will specialise to the case where $\det(M)$ is a square, such that this becomes

$$z\sum_{t\geqslant 0} p^{-st}S_{p^t}(\mathbf{c}) = \frac{1}{1-p^{6-2s}}(1+p^{3-s}-p^{2-s}-p^{5-2s}) = \frac{1-p^{2-s}}{1-p^{3-s}},$$

provided $\sigma > 3$ and $p \nmid \Delta$.

For $p|\Delta$ we simply have $\sum_{t\geqslant 0} p^{-st}S_{p^t}(\mathbf{c}) \ll_{\Delta,\delta} 1$ for $\sigma \geqslant 3+\delta$ by Lemma 7.4.6. Furthermore the factors $\frac{1-p^{3-s}}{1-p^{2-s}}$ for $p|\Delta$ are $O_\Delta(1)$ in the range $\sigma \geqslant 3+\delta$, and so combining these observations we arrive at the first part of the following lemma.

Lemma 7.4.10. *Let $n = 4$, $\det(M)$ be a square, and assume that $M^{-1}(\mathbf{c}) = 0$. Then*

$$\zeta(s,\mathbf{c}) = \frac{\zeta(s-3)}{\zeta(s-2)}\nu(s,\mathbf{c})$$

where

$$\nu(s,\mathbf{c}) \ll_{\delta,\Delta} 1$$

uniformly for $\sigma \geqslant 3+\delta$.

If instead $M^{-1}(\mathbf{c}) \neq 0$, then

$$\zeta(s,\mathbf{c}) = \frac{1}{\zeta(s-2)}\nu(s,\mathbf{c}),$$

where

$$\nu(s,\mathbf{c}) \ll_{\delta,\Delta,\epsilon} |\mathbf{c}|^{\epsilon}$$

uniformly for $\sigma \geqslant 3+\delta$.

Proof. For the second part, note that by Lemma 7.4.8 we have

$$\sum_{t\geqslant 0} p^{-st}S_{p^t}(\mathbf{c}) = 1-p^{2-s}$$

for primes which do not divide $M^{-1}(\mathbf{c})$ and Δ. For $p|M^{-1}(\mathbf{c})$ or $p|\Delta$ we use the bound

$$\sum_{t\geqslant 0} p^{-st}S_{p^t}(\mathbf{c}) = 1+O_\Delta(p^{-\delta}).$$

The number of such primes is $\ll_F \log|\mathbf{c}|$, which gives the result. □

7.4.3 Combining the bounds

Our goal is now to combine the different bounds for $S_q(\mathbf{c})$ and $I_q(\mathbf{c})$ in order to prove Theorem 7.4.1. In order to do this we will make repeated use of the Mellin inversion formula and moving contours. For $\mathbf{c} = \mathbf{0}$ we will also need to introduce a bump function $W \in C^\infty(0,\infty)$ which satisfies the following properties:

(*i*) $\operatorname{supp}(W) \subset [0,2]$,

(*ii*) $W(x) = 1$ for $x \in [0,1]$,

(*iii*) $W^{(n)} \ll_n 1$ for $n = 0, 1, \dots$.

We write

$$\phi(s) = \int_0^\infty x^{s-1} W(x)\, dx$$

for the Mellin transform of W, and note that

$$\phi(s) = \frac{1}{s(s+1)} \int_0^\infty x^{s+1} W''(x)\, dx$$

by integration by parts, for $\sigma > 0$. Writing $f(s)$ for the last integral we see that $f(s) \ll 1$ for $\sigma \geqslant -1$. Furthermore ϕ has a first order pole at $s = 0$, and no other poles in the range $\sigma \geqslant -1 + \epsilon$, where $\epsilon > 0$.

In addition to W, we will work with the Mellin transform of $I_r^*(\mathbf{c})$ with respect to r. Define

$$\phi(s; \mathbf{c}) = \int_0^\infty r^{s-1} I_r^*(\mathbf{c})\, dr$$

for $\mathbf{c} \neq \mathbf{0}$. For $\mathbf{c} = \mathbf{0}$ we will instead have to consider

$$\phi_1(s) = \int_0^\infty r^{s-1} W(r/C) J_r\, dr, \tag{7.19}$$

where

$$I_r^*(\mathbf{0}) = \sigma_\infty(w) + J_r \tag{7.20}$$

and C is a constant (depending on w and F) such that $\operatorname{supp}(I_r^*(\mathbf{0})) \subset [0, C]$. Observe that by Lemma 7.4.3 we then have good bounds for J_r.

Lemma 7.4.11. *Let $\mathbf{c} \neq \mathbf{0}$. For $\sigma \geqslant -1 + \epsilon$ it holds that*

$$\phi(s; \mathbf{c}) = \frac{1}{s(s+1)} f(s; \mathbf{c}),$$

where $f \ll_{w,F,\epsilon} 1$. Furthermore, $\phi(s; \mathbf{c}) \ll_{w,F,\epsilon} 1$ on this range, so in particular there are no poles.

For $\sigma \geqslant -1+\epsilon$ it also holds that

$$\phi_1(s) = \frac{1}{s(s+1)} f_1(s),$$

where $f_1 \ll_{w,F,\epsilon} 1$ and $\phi_1(s) \ll_{w,F,\epsilon} 1$ on this range.

Proof. We have that $I_r^*(\mathbf{c})$ and thus also $\frac{\partial}{\partial r} I_r^*(\mathbf{c})$ are supported on $[0, C]$. Using Lemma 7.4.4 with an appropriate choice of ϵ we have that

$$\lim_{r\to 0} r^s I_r^*(\mathbf{c}) = 0,$$
$$\lim_{r\to 0} r^{s+1} \frac{\partial I_r^*(\mathbf{c})}{\partial r} = 0$$

for the relevant values of s, $\mathbf{c} \neq \mathbf{0}$. This allows us to use integration by parts to get

$$\phi(s;\mathbf{c}) = -\frac{1}{s}\int_0^C r^s \frac{\partial I_r^*(\mathbf{c})}{\partial r}\, dr = \frac{1}{s(s+1)}\int_0^C r^{s+1} \frac{\partial^2 I_r^*(\mathbf{c})}{\partial^2 r}\, dr,$$

where we denote the last integral by $f(s;\mathbf{c})$. By Lemma 7.4.4 we get the desired bound on f. To see that $\phi \ll_{w,F,\epsilon} 1$, apply Lemma 7.4.4 directly in the definition of ϕ, and use that $I_r^*(\mathbf{c}) = 0$ for $r \geqslant C$.

For the case $\mathbf{c} = \mathbf{0}$ we have by Lemma 7.4.3 that $J_r, \frac{\partial}{\partial r} J_r, \frac{\partial^2}{\partial r^2} J_r \ll_N r^N$ for $r \ll 1$, and we have that $W(r) = 0$ for $r \geqslant 2$. Integration by parts then gives

$$\phi_1(s) = \frac{1}{s(s+1)}\int_0^\infty r^{s+1} \frac{\partial^2}{\partial r^2}(W(r/C)J_r)\, dr.$$

Lemma 7.4.3 together with $W^{(n)} \ll_n 1$ then gives the desired bound on $f_1(s)$, which we define to be the last integral. To show that $\phi_1 \ll_{w,F,\epsilon} 1$ we apply Lemma 7.4.3 directly to the integrand in (7.19). □

We are now in a position to combine the bounds for $S_q(\mathbf{c})$ and $I_q(\mathbf{c})$ in order to arrive at our final result.

Proof of Theorem 7.4.1. Begin by observing that

$$\sum_{|\mathbf{c}|>P^\epsilon} \sum_{q=1}^\infty q^{-n} S_q(\mathbf{c}) I_q(\mathbf{c}) \ll_\epsilon 1,$$

because of Lemmas 7.4.2 and 7.4.6.

For $\mathbf{c} \neq \mathbf{0}$ and $M^{-1}(\mathbf{c}) = 0$, we have from the Mellin inversion formula that

$$\sum_{q=1}^\infty q^{-4} S_q(\mathbf{c}) I_q(\mathbf{c}) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \zeta(s+4,\mathbf{c}) P^{4+s} \phi(s;\mathbf{c})\, ds,$$

with $c = 2$, say. We cut the integral at height T to get that this is

$$\frac{1}{2\pi i}\int_{c-iT}^{c+iT} \zeta(s+4,\mathbf{c})P^{s+4}\phi(s;\mathbf{c})\,ds + O(P^6T^{-1}),$$

by Lemmas 7.4.10 and 7.4.11. We now close the contour along the box with sides $\sigma = -1+\epsilon, \sigma = 2$ and $t = \pm T$. By Lemmas 7.4.10 and 7.4.11 we get that the integrand has a simple pole at $s = 0$, and no other poles in this range. The residue here is given by

$$\lim_{s\to 0} s\zeta(s+4,\mathbf{c})P^{s+4}\phi(s;\mathbf{c}) = \sigma(\mathbf{c})\sigma_\infty(w,\mathbf{c})P^4,$$

where we have written

$$\sigma_\infty(w,\mathbf{c}) \coloneqq \int_0^\infty r^{-1}I_r^*(\mathbf{c})\,dr,$$
$$\sigma(\mathbf{c}) \coloneqq \prod_p \Big((1-p^{-1})\sum_{t\geqslant 0} p^{-4t}S_{p^t}(\mathbf{c})\Big).$$

Note that $\sigma(\mathbf{c}) \ll_\Delta 1$ uniformly for all nonzero $\mathbf{c}$ with $M^{-1}(\mathbf{c}) = 0$ by Lemmas 7.4.7 and 7.4.9, and $\sigma_\infty(w,\mathbf{c}) \ll_{S,M} |\mathbf{c}|^{-M}$ for any $M > 0$. To see this last fact, apply Lemma 7.4.4 when $r \ll |\mathbf{c}|^{-N/2}$ and Lemma 7.4.2 for the remaining range of r in the integral defining $\sigma_\infty(w,\mathbf{c})$, and then choose N sufficiently large in terms of M.

For the other sides of the contour we have

$$\int_{-1+\epsilon-iT}^{-1+\epsilon+iT} \frac{\zeta(s+1)}{\zeta(s+2)}\nu(s,\mathbf{c})P^{s+4}\frac{f(s)}{s(s+1)}\,ds \ll_\epsilon P^{3+\epsilon}\int_{-T}^{T}|t|^{1/2}\frac{dt}{|t(t+1)|} \ll_\epsilon P^{3+\epsilon}T^\epsilon,$$

by using the bound (7.8), and

$$\int_{-1+\epsilon+iT}^{c+iT} \frac{\zeta(s+1)}{\zeta(s+2)}\nu(s,\mathbf{c})P^{s+4}\frac{f(s)}{s(1+s)}\,ds \ll_\epsilon P^6T^{-1}$$

by the same bound. The same applies to the edge $t = -T$. Combining this and taking $T = P^6$, say, gives

$$\sum_{q=1}^\infty q^{-4}S_q(\mathbf{c})I_q(\mathbf{c}) = \sigma(\mathbf{c})\sigma_\infty(w,\mathbf{c})P^4 + O_\epsilon(P^{3+7\epsilon}).$$

For $M^{-1}(\mathbf{c}) \neq 0$ we do the same thing, but in this case there is no pole at $s = 1$, which gives

$$\sum_{q=1}^\infty q^{-4}S_q(\mathbf{c})I_q(\mathbf{c}) \ll_\epsilon P^{3+\epsilon}$$

for these values of $\mathbf{c}$.

Combining all the bounds for $\mathbf{c} \neq \mathbf{0}$ we get

$$\sum_{\mathbf{c}\neq 0}\sum_{q=1}^{\infty} q^{-4}S_q(\mathbf{c})I_q(\mathbf{c}) = \sum_{\substack{0<|\mathbf{c}|<P^{\epsilon}\\ M^{-1}(\mathbf{c})=0}} \sigma(\mathbf{c})\sigma_{\infty}(w,\mathbf{c})P^4 + O_{\epsilon}(P^{3+11\epsilon})$$
$$= \sum_{\substack{|\mathbf{c}|>0\\ M^{-1}(\mathbf{c})=0}} \sigma(\mathbf{c})\sigma_{\infty}(w,\mathbf{c})P^4 + O_{\epsilon}(P^{3+11\epsilon}).$$

For $\mathbf{c} = 0$ we note that as $I_q(\mathbf{c}) = 0$ for $q \geqslant CP$, with C as in the proof of Lemma 7.4.11, we have

$$\sum_{q=1}^{\infty} q^{-4}S_q(\mathbf{0})I_q(\mathbf{0}) = \sum_{q\leqslant 2CP} q^{-4}S_q(\mathbf{0})I_q(\mathbf{0})W(q/CP),$$

where W is the bump function defined at the start of Section 7.4.3. Using the decomposition in (7.20) we have that this is

$$\sigma_{\infty}(w)P^4 \sum_{q\leqslant 2CP} q^{-4}S_q(\mathbf{0})W(q/CP) + P^4 \sum_{q\leqslant 2CP} q^{-4}S_q(\mathbf{0})J_{q/P}W(q/CP).$$

Both these sums are dealt with using Mellin inversion. Using the same bounds as above we arrive at

$$\sum_{q\leqslant 2CP} q^{-4}S_q(\mathbf{0})W(q/CP) = \mathrm{Res}_{s=0}\big(\zeta(s+4,\mathbf{0})\phi(s)P^sC^s\big) + O_{w,F,\epsilon}(P^{-1+\epsilon})$$
$$= \sigma^* \log P + \sigma' + O_{\epsilon}(P^{-1+\epsilon}),$$

where σ^* is defined as in Theorem 2.4.2 and we use that $\lim_{s\to 0} s\phi(s) = 1$. This last statement follows by integration by parts and the fact that $W'(x) = 0$ for $0 \leqslant x \leqslant 1$ and for $x \geqslant 1$, and $W(1) = 1, W(2) = 0$. Here we have written $\sigma' = \lim_{s\to 0}\frac{d}{ds}\big(s^2\zeta(s,\mathbf{0})\phi(s)\big)$.

Similarly, by using Lemma 7.4.11, we also get that

$$\sum_{q\leqslant 2CP} q^{-4}S_q(\mathbf{0})J_{q/P}W(q/CP) = \mathrm{Res}_{s=0}\big(\zeta(s+4,\mathbf{0})\phi_1(s)P^s\big) + O_{\epsilon}(P^{-1+\epsilon})$$
$$= \sigma^*\phi_1(0) + O_{\epsilon}(P^{-1+\epsilon}).$$

Finally we define

$$\sigma_1(w) = \sigma_{\infty}(w)\sigma' + \sigma^*\phi_1(0) + \sum_{\substack{|\mathbf{c}|>0\\ M^{-1}(\mathbf{c})=0}} \sigma(\mathbf{c})\sigma_{\infty}(w,\mathbf{c}),$$

redefine ϵ and plug everything into Theorem 2.4.4 to get

$$N_w(P) = \sigma^*\sigma_{\infty}(w)P^2\log P + \sigma_1(w)P^2 + O_{\epsilon}(P^{1+\epsilon}),$$

which completes the proof.

□

Chapter 8

Weak approximation of quadratic forms in four variables

8.1 Introduction

Let F be a non-singular quadratic form in 4 variables with integer coefficients, and let $m \in \mathbb{N}, \mathbf{k} \in \mathbb{Z}^4$ satisfy $F(\mathbf{k}) \equiv 0 (\text{mod } m)$. We are interested in counting integer solutions to

$$F(\mathbf{x}) = 0, \quad \mathbf{x} \equiv \mathbf{k} (\text{mod } m) \tag{8.1}$$

inside a box of width P, as $P \to \infty$. We also assume that $(m, \mathbf{k}) = 1$, as otherwise one could just cancel the common factor everywhere.

The same problem without the congruence condition $\mathbf{x} \equiv \mathbf{k} (\text{mod } m)$ was solved in [37], and we will adopt the same methods here. As in Chapter 7 we will therefore count solutions with a smooth weight $w : \mathbb{R}^4 \to [0, \infty)$ with compact support. Fix a quadratic form F and let

$$N_{F,w}(P, m, \mathbf{k}) = \sum_{\substack{\mathbf{x} \in \mathbb{Z}^4 \\ \mathbf{x} \equiv \mathbf{k} (\text{mod } m) \\ F(\mathbf{x}) = 0}} w(P^{-1}\mathbf{x}).$$

As in [37] we will require that $\nabla F \neq \mathbf{0}$ on the closure of $\text{supp}(w)$. In particular this implies that $\text{supp}(w)$ does not contain the origin. From now on we assume that this technical condition is satisfied without further comment.

For the corresponding quantity

$$N_{F,w}(P) = \sum_{\substack{\mathbf{x} \in \mathbb{Z}^4 \\ F(\mathbf{x}) = 0}} w(P^{-1}\mathbf{x}),$$

Heath-Brown obtains Theorem 2.4.2. We note that the quantity $\sigma_\infty(w)$ can be shown to be positive if $w(\mathbf{x}) > 0$ for some real solution $\mathbf{x}$ to $F(\mathbf{x}) = 0$. For quadratic forms

in fewer than five variables the usual Hardy–Littlewood singular series may converge only conditionally or not at all, which is why the main terms in Theorem 2.4.2 have a modified singular series $\sigma^*(F)$ or $\sigma^*_{\mathrm{sq}}(F)$. In exactly the same way as for the usual singular series one can show that these are positive provided $F(\mathbf{x}) = 0$ has a nonzero p-adic solution for every prime p.

Naively one would perhaps expect $N_{F,w}(P; m, \mathbf{k})$ to have a similar looking asymptotic expression as $N_{F,w}(P)$. Defining $F'(\mathbf{x}) = F(m\mathbf{x}+\mathbf{k})$ and $w'(\mathbf{x}) = w(m\mathbf{x}+\mathbf{k})$ it seems reasonable to expect $N_{F,w}(P; m, \mathbf{k})$ to have the main term predicted by Theorem 2.4.2 for $N_{F',w'}(P)$, keeping in mind that the theorem is not applicable in this case, as F' is not a quadratic form. Rather surprisingly this turns out to not always be the case.

As an example of a case where the above prediction does not give the correct answer, consider the following. Let p, q be odd primes satisfying $p \equiv q \equiv 1 (\mathrm{mod}\ 8)$ and set

$$F(\mathbf{x}) = x_1^2 - pqx_2^2 - x_3x_4.$$

Let k satisfy $\left(\frac{k}{p}\right) = 1$ and $\left(\frac{k}{q}\right) = -1$. We set $m = pq$ and $\mathbf{k} = (a, b, k, k)$, where a, b are any integers satisfying $F(\mathbf{k}) \equiv 0 (\mathrm{mod}\ pq)$, e.g. $a = k, b = 0$. We shall show in Section 8.5 that then in fact

$$N_{F,w}(P; m, \mathbf{k}) \sim CP^2\Big(1 - \frac{L(2, \chi_\Delta)}{L(2, \chi_0)}\Big),$$

where C is the leading coefficient predicted by Theorem 2.4.2, χ_0 is the principal Dirichlet character of conductor pq and $\chi_\Delta = \left(\frac{\cdot}{pq}\right)$.

To understand why the above example has the "wrong" number of solutions we need to look at primitive solutions. We say that a vector $\mathbf{x}$ is *primitive* if the coordinates of $\mathbf{x}$ have no common factor. Let

$$N^{\mathrm{prim}}_{F,w}(P, m, \mathbf{k}) = \sum_{\substack{\mathbf{x}\ \text{primitive} \\ \mathbf{x} \equiv \mathbf{k} (\mathrm{mod}\ m) \\ F(\mathbf{x}) = 0}} w(P^{-1}\mathbf{x}).$$

The quantities $N_{F,w}(P, m, \mathbf{k})$ and $N^{\mathrm{prim}}_{F,w}(P, m, \mathbf{k})$ are then related via the formulae

$$N_{F,w}(P, m, \mathbf{k}) = \sum_{\substack{d=1 \\ (d,m)=1}}^{\infty} N^{\mathrm{prim}}_{F,w}(P/d, m, \overline{d}^{(m)}\mathbf{k}) \tag{8.2}$$

and

$$N^{\mathrm{prim}}_{F,w}(P, m, \mathbf{k}) = \sum_{\substack{d=1 \\ (d,m)=1}}^{\infty} \mu(d) N_{F,w}(P/d, m, \overline{d}^{(m)}\mathbf{k}), \tag{8.3}$$

where $\overline{d}^{(m)}$ is the inverse of d modulo m.

Assume that $\mathbf{x}$ is a primitive solution to $F(\mathbf{x}) = 0$ for the above example, with $\mathbf{x} \equiv \mathbf{k} (\mathrm{mod}\ pq)$. Let $r|x_3$ be an odd prime. Then $x_1^2 - pqx_2^2 \equiv 0 (\mathrm{mod}\ r)$, and so either $\left(\frac{pq}{r}\right) = \left(\frac{r}{pq}\right) = 1$ or $r|x_1$ and $r|x_2$. In the second case, we cannot have $r = p$ or $r = q$, and since $\mathbf{x}$ is primitive we then get $r^2|x_3$. Repeating this argument we see that any odd prime factor of x_3 either appears as an even power or is a quadratic residue mod pq. Noting that $\left(\frac{2}{pq}\right) = 1$ and $\left(\frac{-1}{pq}\right) = 1$ this then gives that in fact $\left(\frac{x_3}{pq}\right) = 1$, a contradiction. This shows that the example has no primitive solutions. By (8.2) we then expect that also the number of unrestricted solutions is biased, which is exactly what we observe.

Another example of this phenomenon was given in [9] and studied further in [10]. Indeed, let

$$F(\mathbf{x}) = x_1^2 + 47x_2^2 - 103x_3^2 - 17 \cdot 47 \cdot 103x_4^2$$

and let $\mathbf{k}_0$ satisfy $F(\mathbf{k}_0) \equiv 0 (\mathrm{mod}\ 17)$. Taking a such that $(a, 17) = 1$ and setting $\mathbf{k} = a\mathbf{k}_0$ it is then shown in [9] that (8.1) has primitive integer solutions for at most half of the possible values of a. In Proposition 8.5.1 we improve this to show that (8.1) has primitive solutions for exactly half of the possible values of a. In [10] the authors show that this curious fact is due to a Brauer–Manin obstruction.

Unfortunately we are not able to fully understand when the above phenomenon occurs. Let Δ be the conductor of the Dirichlet character $\chi_\Delta = \left(\frac{\det F}{\cdot}\right)$. What we are able to say is that $N_{F,w}(P; m, \mathbf{k})$ takes the expected form if $\det M$ is a square or if Δ has some prime factor that does not divide m. On the other hand, if all prime factors of Δ divide m then there are two terms of order P^2 in the asymptotic expression for $N_{F,w}(P; m, \mathbf{k})$, namely the usual one and some sort of bias term.

The bias term appears in the analysis in a similar way as the term $\sigma_1(F, w)P^2$ in Theorem 2.4.2 for the square determinant case (see Chapter 7). The quantity $\sigma_1(F, w)$ is not well understood, to the extent that there does not even seem to be a good way of predicting if it is nonzero. It is therefore not too surprising that it is hard to say anything quantitative about the bias term in our case. On the other hand, we are able to extract some qualitative information, so that knowing $N_{F,w}(P; m, \mathbf{k})$ or $N_{F,w}^{\mathrm{prim}}(P; m, \mathbf{k})$ for some value of $\mathbf{k}$ is enough to compute $N_{F,w}(P; m, d\mathbf{k})$ and $N_{F,w}^{\mathrm{prim}}(P; m, d\mathbf{k})$ for any multiple $d\mathbf{k}$ of $\mathbf{k}$.

Theorem 8.1.1. *Let F be a non-singular quadratic form in 4 variables with underlying matrix M, let $m \in \mathbb{N}$ and $\mathbf{k} \in \mathbb{Z}^4$ be such that $F(\mathbf{k}) \equiv 0 (\mathrm{mod}\ m)$, and let w be*

a smooth weight function. Then, if $\det M$ *is a square it holds that*

$$\begin{aligned} N_{F,w}(P;m,\mathbf{k}) =& \sigma_\infty(F,w)\sigma^*_{sq}(F,m,\mathbf{k})m^{-4}P^2\log P \\ &+ \sigma_1(F,w,m,\mathbf{k})P^2 + O_{F,w,m,\epsilon}(P^{3/2+\epsilon}) \end{aligned}$$

for any $\epsilon > 0$*. Here* $\sigma_\infty(F,w)$ *is as in* (2.6) *and* $\sigma^*_{sq}(F,m,\mathbf{k})$ *is the singular series, the definition of which will be given in* (8.13)*. The coefficient* $\sigma_1(F,w,m,\mathbf{k})$ *is some constant not depending on* P*.*

Theorem 8.1.2. *Let* F *be a non-singular quadratic form in* 4 *variables with underlying matrix* M*, let* $m \in \mathbb{N}$ *and* $\mathbf{k} \in \mathbb{Z}^4$ *be such that* $F(\mathbf{k}) \equiv 0 (\mathrm{mod}\ m)$*, and let* w *be a smooth weight function. Then if* $\det M$ *is not a square it holds that, for any* $\epsilon > 0$*,*

$$\begin{aligned} N_{F,w}(P;m,\mathbf{k}) = \big(\sigma_\infty(F,w)L(1,\chi_\Delta)\sigma^*(F,m,\mathbf{k}) + \tau(F,w,m,\mathbf{k})\big)m^{-4}P^2 \\ + O_{F,w,\epsilon,m}(P^{5/3+\epsilon}), \end{aligned}$$

where we postpone the definition of the singular series $\sigma^*(F,m,\mathbf{k})$ *until* (8.14)*,* $\chi_\Delta = \left(\frac{\det M}{\cdot}\right)$*, and* $\tau(F,w,m,\mathbf{k})$ *is some quantity not depending on* P *that satisfies the following two properties.*

- $\tau(F,w,m,\mathbf{k}) = 0$ *unless every prime factor of* Δ *divides* m*;*
- *If* $(d,m) = 1$ *then* $\tau(F,w,m,d\mathbf{k}) = \chi_\Delta(d)\tau(F,w,m,\mathbf{k})$*.*

Unfortunately we are not able to say much else about the quantity $\tau(F,w,m,\mathbf{k})$ than the two properties listed in the theorem. These do however immediately imply the following corollary, which in particular tells us that in projective space everything works out as expected.

Corollary 8.1.3. *Let* F, w, m *and* $\mathbf{k}$ *be as in Theorem 8.1.2 and assume that every prime factor of* Δ *divides* m*. Then if* $\chi_\Delta(d) = 1$ *it holds that, for any* $\epsilon > 0$*,*

$$N_{F,w}(P;m,\mathbf{k}) - N_{F,w}(P;m,d\mathbf{k}) = O_{F,w,m,\epsilon}(P^{5/3+\epsilon}).$$

Furthermore, it holds that

$$\begin{aligned} \sideset{}{^*}\sum_{d (\mathrm{mod}\ m)} N_{F,w}(P;m,d\mathbf{k}) = \sigma_\infty(F,w)L(1,\chi_\Delta)\sigma^*(F,m,\mathbf{k})\phi(m)m^{-4}P^2 \\ + O_{F,w,\epsilon,m}(P^{5/3+\epsilon}) \end{aligned}$$

for any $\epsilon > 0$*.*

The error terms in Theorem 8.1.2 and Corollary 8.1.3 are both worse than the error terms in theorems 2.4.2 and 8.1.1. This is not because of some actual limitation of the method, indeed one could obtain the same error term in all results by some extra work. However, allowing for a loss of $P^{1/6}$ in the error greatly simplifies the amount of technical work needed, which is why we have chosen to go for a slightly weaker approach here.

We remark that the problem we are trying to solve was mentioned in [37, discussion surrounding Corollary 3]. There Heath-Brown points out that the counting function $N_{F,w}(P; m, \mathbf{k})$ may be handled with the same methods as those applied to $N_{F,w}(P)$, which is exactly what we do. On the other hand, the conclusion is perhaps not as straightforward as indicated. In particular, it is not true that excluding the obvious obstructions is enough to guarantee primitive solutions, as our initial example showed.

Notation

For a 4×4 matrix A and a vector $\mathbf{x} \in \mathbb{R}^4$ we'll write $A^{-1}(\mathbf{x}) = \mathbf{x}^T A^{-1} \mathbf{x}$. The matrix M will always be the symmetric matrix satisfying $F(\mathbf{x}) = \mathbf{x}^T M \mathbf{x}$ and Δ will always be the conductor of the Dirichlet character $\left(\frac{\det M}{\cdot}\right)$, which we denote by χ_Δ. We will write χ_0 for the principal Dirichlet character modulo m.

If $(a, b) = 1$ we will write $\overline{a}^{(b)}$ for the inverse of a modulo b, that is, $\overline{a}^{(b)} a \equiv 1 (\text{mod } b)$. The notation $a | b^\infty$ means that if $p | a$ then $p | b$.

For a vector $\mathbf{x}$ we write $|\mathbf{x}|$ for the supremum norm $\|\mathbf{x}\|_\infty$.

Outline

In section 8.2 we apply the Heath-Brown circle method as described in [37] to get an expression for $N_{F,w}(P; m, \mathbf{k})$. Most of the analysis is nearly identical as in [37], and the parts which require some extra care are dealt with in Section 8.3. In Section 8.4 we sketch how to prove Theorem 8.1.1, Theorem 8.1.2 and Corollary 8.1.3. Finally, in Section 8.5 we revisit the example mentioned in the introduction together with a second example, and compute $N_{F,w}(P; m, \mathbf{k})$ for these two cases.

8.2 The Heath-Brown circle method

We start form Theorem 2.4.4. Note that in the theorem's original form F is a form of degree d, but the homogeneity of F is not actually used in the proof. We will use

that there is a constant $C > 0$ such that $I_q(\mathbf{c}) = 0$ for all $q \geqslant CP$, by properties of h outlined in [37].

Applying Theorem 2.4.4 with $F(m\mathbf{x}+\mathbf{k})$ and $w(P^{-1}(m\mathbf{x}+\mathbf{k}))$ in place of $F(\mathbf{x})$ and $w(P^{-1}(\mathbf{x}))$ respectively, we get

$$N_{F,w}(P,m,\mathbf{k}) = c_P P^{-2} \sum_{\mathbf{c}} \sum_{q=1}^{\infty} q^{-4} S_q(\mathbf{c};m,\mathbf{k}) I_q(\mathbf{c};m,\mathbf{k}) \tag{8.4}$$

with

$$S_q(\mathbf{c};m,\mathbf{k}) = \sum_{a (\mathrm{mod}\ q)}^{*} \sum_{\mathbf{b} (\mathrm{mod}\ q)} e_q(aF(m\mathbf{b}+\mathbf{k}) + \mathbf{c}\cdot\mathbf{b}) \tag{8.5}$$

and

$$I_q(\mathbf{c};m,\mathbf{k}) = \int w(P^{-1}(m\mathbf{x}+\mathbf{k})) h(P^{-1}q, P^{-2}F(m\mathbf{x}+\mathbf{k})) e_q(-\mathbf{c}\cdot\mathbf{x}) d\mathbf{x}.$$

Substitute $\mathbf{y} = m\mathbf{x}+\mathbf{k}$ in the above integral to get

$$I_q(\mathbf{c};m,\mathbf{k}) = m^{-4} e_{mq}(\mathbf{c}\cdot\mathbf{k}) I_q(\mathbf{c}/m), \tag{8.6}$$

where $I_q(\mathbf{c})$ is defined as in (2.10).

The fact that $\mathbf{c} \in \mathbb{Z}^4$ rather than just $\mathbb{R}^4$ is not needed to prove any of the results from [37] concerning $I_q(\mathbf{c})$, so all of the following results apply with $\frac{1}{m}\mathbf{c}$ in place of $\mathbf{c}$. We will need these bounds when we combine everything in Section 8.4.

Lemma 8.2.1 ([37, Lemma 13]). *For all $N > 0$ and $q \ll P$ it holds that*

$$I_q(\mathbf{0}) = P^4\big(\sigma_\infty(F,w) + O_{F,w,N}((q/P)^N)\big),$$

where $\sigma_\infty(F,w)$ is as in (2.6).

Lemma 8.2.2 ([37, Lemma 16]). *It holds that*

$$\frac{\partial^j I_q(\mathbf{0})}{\partial q^j} \ll_{F,w} P^4 q^{-j}$$

for $j = 0, 1$.

Lemma 8.2.3 ([37, Lemma 19]). *For $\mathbf{c} \neq \mathbf{0}$ we have*

$$I_q(\mathbf{c}) \ll_{F,w,N} P^5 q^{-1} |\mathbf{c}|^{-N}$$

for any $N > 0$.

Lemma 8.2.4 ([37, Lemma 22]). *For $j = 0, 1$ it holds that*

$$\frac{\partial^j I_q(\mathbf{c})}{\partial q^j} \ll_{F,w,\epsilon} P^{3+\epsilon} q^{1-j} |\mathbf{c}|^{-1}.$$

The main work needed to establish a result is to deal with the sums $S_q(\mathbf{c};m,\mathbf{k})$, which is what we do in the next section.

8.3 The sum $S_q(\mathbf{c}; m, \mathbf{k})$

We seek to prove an asymptotic for the sum

$$S(X, \mathbf{c}, m, \mathbf{k}) = \sum_{q \leqslant X} S_q(\mathbf{c}; m, \mathbf{k}) e_{mq}(\mathbf{c} \cdot \mathbf{k}), \tag{8.7}$$

where the factor $e_{mq}(\mathbf{c} \cdot \mathbf{k})$ is included because of (8.6).

For a fixed value of q write $q = uv$, where $(m, u) = 1$ and $v | m^\infty$. Let $a = va_u + ua_v$ and $\mathbf{b} = u\overline{u}^{(v)}\mathbf{b}_v + v\overline{v}^{(u)}\mathbf{b}_u$ in (8.5) to get that

$$S_{uv}(\mathbf{c}; m, \mathbf{k}) = S_u(\overline{v}^{(u)}\mathbf{c}; m, \mathbf{k}) S_v(\overline{u}^{(v)}\mathbf{c}; m, \mathbf{k}). \tag{8.8}$$

Note that if $\mathbf{c} = \mathbf{0}$ this simplifies to $S_q(\mathbf{0}; m, \mathbf{k}) = S_u(\mathbf{0}; m, \mathbf{k}) S_v(\mathbf{0}; m, \mathbf{k})$. In other words, $S_q(\mathbf{0}; m, \mathbf{k})$ is multiplicative in q, which will make this case much easier than the general case.

The case $\mathbf{c} = \mathbf{0}$

As noted above $S_q(\mathbf{0}; m, \mathbf{k})$ is multiplicative as a function of q, and so we can use the the exact same methods as in [37] to understand the sum $\sum_{q \leqslant X} q^{-4} S_q(\mathbf{0}; m, \mathbf{k})$. The resulting main term will involve the quantity

$$\sigma_p(m, \mathbf{k}) = \sum_{s \geqslant 0} p^{-4s} S_{p^s}(\mathbf{0}; m, \mathbf{k}). \tag{8.9}$$

In much the same way as one can show that the usual p-adic density

$$\sigma_p = \sum_{s \geqslant 0} p^{-4s} S_{p^s}(\mathbf{0}) \tag{8.10}$$

satisfies

$$\sigma_p = \lim_{\nu \to \infty} p^{-3\nu} \#\{\mathbf{x} \in (\mathbb{Z}/p^\nu\mathbb{Z})^4 : F(\mathbf{x}) \equiv 0 (\mathrm{mod}\ p^\nu)\},$$

one can show that

$$\sigma_p(m, \mathbf{k}) = \lim_{\nu \to \infty} p^{-3\nu} \#\{\mathbf{x} \in (\mathbb{Z}/p^\nu\mathbb{Z})^4 : F(m\mathbf{x} + \mathbf{k}) \equiv 0 (\mathrm{mod}\ p^\nu)\}. \tag{8.11}$$

If $(p, m) = 1$ it is clear that $\sigma_p(m, \mathbf{k}) = \sigma_p$, for example one can substitute $m\mathbf{x} + \mathbf{k} \mapsto \mathbf{x}$ in the above.

We remark that for any d satisfying $(d, m) = 1$,

$$\sigma_p(m, d\mathbf{k}) = \sigma_p(m, \mathbf{k}), \tag{8.12}$$

and so in reality $\sigma_p(m, \mathbf{k})$ only depends on the projective vector $\mathbf{k}$.

Lemma 8.3.1. *If* $\det M$ *is a square then*

$$\sum_{q\leqslant X} q^{-4}S_q(\mathbf{0};m,\mathbf{k}) = \sigma^*_{sq}(F,m,\mathbf{k})\log X + \sigma_1'(F,m,\mathbf{k}) + O_{F,m,\epsilon}(X^{-1/2+\epsilon}),$$

where

$$\sigma^*_{sq}(F,m,\mathbf{k}) = \big(\prod_{p\nmid m}(1-p^{-1})\sigma_p\big)\big(\prod_{p\mid m}(1-p^{-1})\sigma_p(m,\mathbf{k})\big), \tag{8.13}$$

$\sigma_p(m,\mathbf{k})$ *and* σ_p *are as in* (8.9) *and* (8.10) *respectively and* $\sigma_1'(F,m,\mathbf{k})$ *is some constant depending on* F, m *and* $\mathbf{k}$ *but not on* X.

If $\det M$ *is not a square then*

$$\sum_{q\leqslant X} q^{-4}S_q(\mathbf{0};m,\mathbf{k}) = L(1,\chi_\Delta)\sigma^*(F,m,\mathbf{k}) + O_{F,m,\epsilon}(X^{-1/2+\epsilon}),$$

where

$$\sigma^*(F,m,\mathbf{k}) = \big(\prod_{p\nmid m}(1-\chi_\Delta(p)p^{-1})\sigma_p\big)\big(\prod_{p\mid m}(1-\chi_\Delta(p)p^{-1})\sigma_p(m,\mathbf{k})\big) \tag{8.14}$$

and $\sigma_p(m,\mathbf{k})$ *and* σ_p *are as above.*

Proof. See [37, Lemma 31], which gives a similar statement with $S_q(\mathbf{0})$ in place of $S_q(\mathbf{0};m,\mathbf{k})$. The details are otherwise almost exactly the same, so we do not repeat them here. □

The case $\mathbf{c}\neq\mathbf{0}$

If $(u,m)=1$ it holds that

$$S_u(\overline{v}^{(u)}\mathbf{c};m,\mathbf{k}) = e_u(-\overline{mv}^{(u)}\mathbf{c}\cdot\mathbf{k})S_u(\overline{mv}^{(u)}\mathbf{c}) = e_u(-\overline{mv}^{(u)}\mathbf{c}\cdot\mathbf{k})S_u(\mathbf{c}),$$

by the substitution $\mathbf{b}' = m\mathbf{b}+\mathbf{k}$ in (8.5) and the fact that $S_q(k\mathbf{c}) = S_q(\mathbf{c})$ for $(q,k)=1$. Using this and (8.8) gives

$$e_{uv}(\mathbf{c}\cdot\mathbf{k}/m)S_{uv}(\mathbf{c};m,\mathbf{k}) = e_{mv}(\overline{u}^{(mv)}\mathbf{c}\cdot\mathbf{k})S_u(\mathbf{c})S_v(\overline{u}^{(v)}\mathbf{c};m,\mathbf{k}),$$

where we also use that

$$\overline{a}^{(b)} + \overline{b}^{(a)} \equiv 1 (\mathrm{mod}\ ab)$$

to write $e_{muv}(\mathbf{c}\cdot\mathbf{k})e_u(-\overline{mv}^{(u)}\mathbf{c}\cdot\mathbf{k}) = e_{mv}(\overline{u}^{(mv)}\mathbf{c}\cdot\mathbf{k})$.

We will use this to write the sum over q in (8.7) as a sum over u and v, where $(u,m)=1$ and $v|m^\infty$. We then split the sum into two parts depending on the size of v, namely $S(X,\mathbf{c},m,\mathbf{k})=S_1+S_2$ with

$$S_1=\sum_{\substack{v|m^\infty\\ v\leqslant X^{1/3}}}\sum_{u\leqslant X/v}e_{mv}(\overline{u}^{(mv)}\mathbf{c}\cdot\mathbf{k})S_u(\mathbf{c})S_v(\overline{u}^{(v)}\mathbf{c};m,\mathbf{k})$$

and

$$S_2=\sum_{\substack{v|m^\infty\\ X^{1/3}<v\leqslant X}}\sum_{u\leqslant X/v}e_{mv}(\overline{u}^{(mv)}\mathbf{c}\cdot\mathbf{k})S_u(\mathbf{c})S_v(\overline{u}^{(v)}\mathbf{c};m,\mathbf{k}).$$

For S_2 we use the following basic bound.

Lemma 8.3.2. *It holds that*

$$S_q(\mathbf{c},m,\mathbf{k})\ll_{m,\Delta}q^3.$$

Proof. This is essentially [37, Lemma 25], which establishes the bound for $S_q(\mathbf{c})$, but pretty much the exact same proof goes through for $S_q(\mathbf{c},m,\mathbf{k})$. □

Taking absolute values and using the lemma gives

$$|S_2|\ll_{m,\Delta}\sum_{\substack{v|m^\infty\\ X^{1/3}<v\leqslant X}}\sum_{u\leqslant X/v}u^3v^3\leqslant\sum_{\substack{v|m^\infty\\ X^{1/3}<v\leqslant X}}\sum_{u\leqslant X/v}X^3\leqslant\sum_{\substack{v|m^\infty\\ v\leqslant X}}\sum_{u\leqslant X^{2/3}}X^3\ll_{\epsilon,m}X^{11/3+\epsilon}.$$

For a function f supported only on integers coprime to mv with period mv one can decompose

$$f(u)=\sum_{\chi(\mathrm{mod}\ mv)}a_\chi\chi(u),$$

where

$$a_\chi=\frac{1}{\phi(mv)}\sum_{x(\mathrm{mod}\ mv)}f(x)\overline{\chi(x)}.$$

Using this in S_1 we get

$$S_1=\sum_{v|m^\infty,v\leqslant X^{1/3}}\sum_{\chi(\mathrm{mod}\ mv)}a_\chi(v,\mathbf{c},m,\mathbf{k})\sum_{u\leqslant X/v}\chi(u)S_u(\mathbf{c})$$

with

$$a_\chi(v,\mathbf{c},m,\mathbf{k})=\frac{1}{\phi(mv)}\sum_{x(\mathrm{mod}\ mv)}\chi(x)e_{mv}(x\mathbf{c}\cdot\mathbf{k})S_v(x\mathbf{c};m,\mathbf{k}).\tag{8.15}$$

Dealing with the innermost sum over u will be done nearly identically as in [37], the only difference is the twist by $\chi(u)$.

Lemma 8.3.3. *Let χ be a Dirichlet character of conductor mv for some $v|m^\infty$. Then for any $\epsilon > 0$,*

$$\sum_{u \leqslant X} \chi(u) S_u(\mathbf{c}) = \eta(\mathbf{c})\xi(\chi)\sigma'(F,m)\frac{X^4}{4} + O_\epsilon(X^{7/2+\epsilon}|\mathbf{c}|^\epsilon),$$

where

$$\sigma'(F,m) = \big(\prod_{p \nmid m}(1-p^{-1})\sigma'_p\big)\big(\prod_{p|m}(1-p^{-1})\big), \tag{8.16}$$

$$\sigma'_p = \sum_{t \geqslant 0} p^{-4t} S_{p^t}(\mathbf{0})\chi_\Delta(p)^t, \tag{8.17}$$

$\eta(\mathbf{c}) = 1$ if $M^{-1}(\mathbf{c}) = 0$ and 0 otherwise, and $\xi(\chi) = 1$ if $\chi = \chi_0\chi_\Delta$ and 0 otherwise. Recall that χ_0 is the principal character modulo m.

Proof. If $M^{-1}(\mathbf{c}) \neq 0$ this is just [37, Lemma 28] with the condition $|\mathbf{c}| \leqslant P$ removed and the error P^ϵ replaced by $|\mathbf{c}|^\epsilon$.

If $M^{-1}(\mathbf{c}) = 0$ this is essentially the same as [37, Lemma 30]. The only real difference is that if $\chi = \chi_0\chi_\Delta$ then one gets a main term even when $\det M$ is not a square. Note that in the main term we have used the fact that $S_{p^t}(\mathbf{c}) = S_{p^t}(\mathbf{0})$ whenever $M^{-1}(\mathbf{c}) = 0$ and $p \nmid \Delta$. Indeed,

$$aF(\mathbf{b}) + \mathbf{c}\cdot\mathbf{b} = aF(\mathbf{b} + (2a)^{-1}M^{-1}\mathbf{c}) - (4a)^{-1}M^{-1}(\mathbf{c})$$

can be used in the definition of $S_{p^t}(\mathbf{c})$. □

Applying Lemma 8.3.3 we now have that S_1 is

$$\sum_{v|m^\infty, v\leqslant X^{1/3}} \sum_{\chi (\mathrm{mod}\ mv)} a_\chi(v,\mathbf{c},m,\mathbf{k})\Big(\eta(\mathbf{c})\xi(\chi)\sigma'(F,m)\frac{X^4}{4v^4} + O_\epsilon((X/v)^{7/2+\epsilon}|\mathbf{c}|^\epsilon)\Big).$$

For the error term we use that

$$|a_\chi(v,\mathbf{c},m,\mathbf{k})| \ll_{m,\Delta} v^3 \tag{8.18}$$

by (8.15) and the bound in Lemma 8.3.2. The error is then bounded above by

$$\sum_{v|m^\infty, v\leqslant X^{1/3}} O_{m,\Delta,\epsilon}(v^{1/2-\epsilon}X^{7/2+\epsilon}|\mathbf{c}|^\epsilon) \ll_{m,\Delta,\epsilon} X^{11/3+\epsilon}|\mathbf{c}|^\epsilon,$$

where we have used that the number of values of v we are summing over is of order $(\log X)^{\omega(m)} \ll_{m,\epsilon} X^\epsilon$, where $\omega(m)$ is the number of unique prime factors of m.

If $\eta(\mathbf{c}) = 1$ and $\Delta | m^\infty$ we might also get a main term. Assume therefore that $\Delta | m^\infty$, such that for at least some v the character $\chi = \chi_0\chi_\Delta$ appears in the sum. The main term is then

$$\eta(\mathbf{c})\sigma'(F,m)\frac{X^4}{4}\sum_{\substack{v|m^\infty, v\leqslant X^{1/3}\\ \Delta|mv}} a_{\chi_0\chi_\Delta}(v,\mathbf{c},m,\mathbf{k})v^{-4}.$$

Finally we extend the sum over v to also include values with $v > X^{1/3}$. By (8.18) this introduces an error $O_{m,\Delta,\epsilon}(X^{11/3+\epsilon})$. Combining all of this we thus have

$$S_1 = \frac{X^4}{4}\eta(\mathbf{c})\sigma'(F,m)A(\mathbf{c},m,\mathbf{k}) + O_{m,\Delta,\epsilon}(X^{11/3+\epsilon}),$$

where we have defined

$$A(\mathbf{c},m,\mathbf{k}) = \sum_{v|m^\infty,\Delta|mv} a_{\chi_0\chi_\Delta}(v,\mathbf{c},m,\mathbf{k})v^{-4}. \tag{8.19}$$

Together with the bound on S_2 we have thus proved the following.

Lemma 8.3.4. *For any $\epsilon > 0$ and $\mathbf{c} \neq \mathbf{0}$ it holds that*

$$S(X,\mathbf{c},m,\mathbf{k}) = \frac{X^4}{4}\eta(\mathbf{c})\sigma'(F,m)A(\mathbf{c},m,\mathbf{k}) + O_{m,\Delta,\epsilon}(X^{11/3+\epsilon}|\mathbf{c}|^\epsilon).$$

Remark. One can in fact evaluate $a_\chi(v,\mathbf{c},m,\mathbf{k})$ by factoring v and m into separate prime factors and using standard results on Gauss sums. However, the resulting expressions do not seem very enlightening, and the dependence on $\mathbf{c}$ is particularly nasty. In proving our main theorem we will eventually need to sum an expression involving $\eta(\mathbf{c})A(\mathbf{c},m,\mathbf{k})$ over $\mathbf{c} \neq \mathbf{0}$, and here we were not able to use the closed form of $a_\chi(v,\mathbf{c},m,\mathbf{k})$ and $A(\mathbf{c},m,\mathbf{k})$ in any meaningful sense. Because computing a closed form expression is a fairly long and technical computation we thus omit it here.

Finally, we record the following transformation property for $A(\mathbf{c},m,\mathbf{k})$.

Lemma 8.3.5. *Let $(d,m) = 1$. Then*

$$A(\mathbf{c},m,d\mathbf{k}) = \chi_\Delta(d)A(\mathbf{c},m,\mathbf{k}).$$

Proof. By (8.15) and (8.5) we have that $a_\chi(v,\mathbf{c},m,\mathbf{k})$ is

$$\sum_{x \pmod{mv}} \sideset{}{^*}\sum_{a \pmod{v}} \sum_{\mathbf{b} \pmod{v}} \chi(x)e_{mv}(x\mathbf{c}\cdot(m\mathbf{b}+\mathbf{k}))e_v(aF(m\mathbf{b}+\mathbf{k})).$$

If $(d,m) = 1$ we can make the substitutions $d\mathbf{b} \mapsto \mathbf{b}$, $x \mapsto dx$ and $a \mapsto d^2a$ to get $a_\chi(v,\mathbf{c},m,\mathbf{k}) = \chi(d)a_\chi(v,\mathbf{c},m,d\mathbf{k})$. Inserting this into (8.19) then gives the result. □

8.4 Proof of main results

We are now ready to prove our main results. Assume first that $\det M$ is a square. The proof of Theorem 8.1.1 is exactly the same as the proof of [37, Theorem 7], except that the main term comes from Lemma 8.3.1 instead of an analogous result for $\sum_{q\leqslant X} q^{-4}S_q(\mathbf{0})$, and similarly one needs to use Lemma 8.3.4 instead of an analogous result for $\sum_{q\leqslant X} S_q(\mathbf{c})$ to achieve the term of order P^2. Note that by (8.18), $A(\mathbf{c}, m, \mathbf{k})$ can be bounded above by $O_m(1)$ uniformly in $\mathbf{c}$, and so showing convergence of the sum over $\mathbf{c} \neq \mathbf{0}$ can be done exactly as before.

When $\det M$ is not a square the proof is essentially just a combination of the proofs of [37, theorems 6 & 7]. As usual $\mathbf{c} = \mathbf{0}$ in (8.4) gives rise to a term of order P^2, but unlike in Theorem 2.4.2 the terms with $\mathbf{c} \neq \mathbf{0}$ also contribute something of order P^2, in the same way that one gets a secondary term of order P^2 in the case where $\det M$ is a square. As this is the most interesting case we repeat a rough sketch of the proofs in [37].

Sketch of proof of Theorem 8.1.2. Using lemmas 8.2.2 and 8.3.1 and the identity (8.6) we have by partial summation (Lemma 2.1.2) that

$$\sum_{R<q\leqslant 2R} q^{-4}S_q(\mathbf{0}; m, \mathbf{k})I_q(\mathbf{0}; m, \mathbf{k}) \ll_{F,w,m,\epsilon} P^4R^{-1/2+\epsilon}.$$

Since $I_q(\mathbf{0}) = 0$ for $q \gg P$ this then gives that

$$\sum_{q>P^{1-\epsilon}} q^{-4}S_q(\mathbf{0}; m, \mathbf{k})I_q(\mathbf{0}; m, \mathbf{k}) \ll_{F,w,m,\epsilon} P^{7/2+2\epsilon}.$$

For $q \leqslant P^{1-\epsilon}$, lemmas 8.2.1 and 8.3.1 give that

$$\sum_{q\leqslant P^{1-\epsilon}} q^{-4}S_q(\mathbf{0}; m, \mathbf{k})I_q(\mathbf{0}; m, \mathbf{k}) = \sigma_\infty(F, w)L(1, \chi_\Delta)\sigma^*(F, m, \mathbf{k})m^{-4}P^4 + O_{F,w,m,\epsilon}(P^{7/2+\epsilon}).$$

Lemmas 8.2.3 and 8.3.2 together with the fact that $I_q(\mathbf{c})$ is supported only for $q \ll P$ gives

$$\sum_{|\mathbf{c}|\geqslant P^\epsilon} \sum_{q=1}^{\infty} q^{-4}S_q(\mathbf{c}; m, \mathbf{k})I_q(\mathbf{c}; m, \mathbf{k}) \ll_{F,w,m,\epsilon} 1.$$

For $|\mathbf{c}| \leqslant P^\epsilon$ and $\mathbf{c} \neq \mathbf{0}$ we use partial summation (Lemma 2.1.2) together with Lemma 8.2.4 and Lemma 8.3.4 to get

$$\sum_{R<q\leqslant 2R} q^{-4}e_{mq}(\mathbf{c}\cdot\mathbf{k})S_q(\mathbf{c}; m, \mathbf{k})I_q(\mathbf{c}/m) = \eta(\mathbf{c})\sigma'(F, m)A(\mathbf{c}, m, \mathbf{k})\int_R^{2R} t^{-1}I_t(\mathbf{c}/m)dt + O_{F,w,m,\epsilon}(P^{3+\epsilon}R^{2/3+\epsilon}).$$

Summing over all q and using that $I_q(\mathbf{c}) = 0$ for $q \gg P$ we then get

$$\sum_{q=1}^{\infty} q^{-4} e_{mq}(\mathbf{c}\cdot\mathbf{k}) S_q(\mathbf{c}; m, \mathbf{k}) I_q(\mathbf{c}/m)$$
$$= \eta(\mathbf{c})\sigma'(F,m) A(\mathbf{c}, m, \mathbf{k}) \int_0^{\infty} t^{-1} I_t(\mathbf{c}/m) dt + O_{F,w,m,\epsilon}(P^{11/3+2\epsilon})$$

for any $\mathbf{c} \neq \mathbf{0}$. Now we define

$$I_r^*(\mathbf{c}) = P^{-4} I_{rP}(\mathbf{c}),$$

which does not depend on P. The last integral above is then $P^4 \sigma_\infty(F, w, \mathbf{c}/m)$, where we have defined

$$\sigma_\infty(F, w, \mathbf{c}) = \int_0^{\infty} r^{-1} I_r^*(\mathbf{c}) dr.$$

The convergence of this is shown in [37], but we do not repeat the argument here. In fact it holds that

$$\sigma_\infty(F, w, \mathbf{c}) \ll_{F,w,N} |\mathbf{c}|^{-N}$$

for any $N > 0$. For the main term we may therefore extend the sum $\sum_{0<|\mathbf{c}|\leqslant P^\epsilon}$ to all $\mathbf{c} \neq \mathbf{0}$, which upon plugging everything into (8.4) finally gives us that

$$N_{F,w}(P; m, \mathbf{k}) = m^{-4} P^2 \Big(\sigma'(F,m) \sum_{\mathbf{c}\neq\mathbf{0}} \eta(\mathbf{c}) A(\mathbf{c}, m, \mathbf{k}) \sigma_\infty(F, w, \mathbf{c}/m)$$
$$+ \sigma_\infty(F, w) L(1, \chi_\Delta) \sigma^*(F, m, \mathbf{k}) \Big) + O_{\epsilon,F,w,m}(P^{5/3+\epsilon}),$$

where we have redefined ϵ. By defining

$$\tau(F, w, m, \mathbf{k}) = \sigma'(F,m) \sum_{\mathbf{c}\neq\mathbf{0}} \eta(\mathbf{c}) A(\mathbf{c}, m, \mathbf{k}) \sigma_\infty(F, w, \mathbf{c}/m)$$

we then have the required asymptotic expression.

Finally, $A(\mathbf{c}, m, \mathbf{k}) = 0$ whenever $\Delta \nmid m^\infty$, and so the same holds for $\tau(F, w, m, \mathbf{k})$. Lemma 8.3.5 gives that $\tau(F, w, m, d\mathbf{k}) = \chi_\Delta(d)\tau(F, w, m, \mathbf{k})$ for any d satisfying $(m, d) = 1$. □

Proof of Corollary 8.1.3. For the first part of the corollary we just need that for $(m, d) = 1$ it holds that $\sigma^*(F, m, d\mathbf{k}) = \sigma^*(F, m, \mathbf{k})$ by (8.12), together with the fact that $\tau(F, w, m, d\mathbf{k}) = \chi_\Delta(d)\tau(F, w, m, \mathbf{k})$.

For the second part of the corollary we simply use

$$\tau(F, w, m, d\mathbf{k}) = \chi_\Delta(d)\tau(F, w, m, \mathbf{k}),$$

together with the fact that $\chi_\Delta(d) = -1$ for exactly half of the values $d \in (\mathbb{Z}/m\mathbb{Z})^*$ when $\Delta | m^\infty$. Here we also use that $\Delta \neq 1, 2$, as for squarefree D the quadratic character $\left(\frac{D}{\cdot}\right)$ has conductor $|D|$ or $4|D|$, and so $\Delta = 2$ will never occur. □

Remark. Now that we have the precise definition of $\tau(F, w, m, \mathbf{k})$ it should be clear what causes the difficulty in analysing it. In particular we see that in addition to understanding $A(\mathbf{c}, m, \mathbf{k})$ one needs to say something more about $\sigma_\infty(F, w, \mathbf{c})$, a feat that also would tell us something about $\sigma_1(F, w)$ in Theorem 2.4.2.

8.5 Examples

We begin by noting that the second result of Corollary 8.1.3 implies a similar result for $N^{\text{prim}}_{F,w}(P; m, \mathbf{k})$. Indeed, by the corollary and (8.3) we get that

$$\begin{aligned}\sideset{}{^*}\sum_{d(\text{mod } m)} N^{\text{prim}}_{F,w}(P; m, d\mathbf{k}) &= \sum_{\substack{l=1\\(l,m)=1}}^{\infty} \mu(l) \sideset{}{^*}\sum_{d(\text{mod } m)} N_{F,w}(P/l; m, d\bar{l}^{(m)}\mathbf{k}) \\ &= CP^2 \sum_{(l,m)=1} \mu(l) l^{-2} + O_{F,w,m,\epsilon}\left(P^{5/3+\epsilon}\right) \\ &= \frac{C}{L(2,\chi_0)} P^2 + O_{F,w,m,\epsilon}(P^{5/3+\epsilon}),\end{aligned}$$

where $C = \sigma_\infty(F, w) L(1, \chi_\Delta) \sigma^*(F, m, \mathbf{k}) \phi(m) m^{-4}$. Note that we have used that $\sigma^*(F, m, \mathbf{k})$ is unchanged when $\mathbf{k}$ is multiplied by $l \in (\mathbb{Z}/m\mathbb{Z})^*$.

From the first part of Corollary 8.1.3 together with (8.3) we also get that

$$N^{\text{prim}}_{F,w}(P; m, \mathbf{k}) - N^{\text{prim}}_{F,w}(P; m, d\mathbf{k}) = O_{F,w,m,\epsilon}(P^{5/3+\epsilon})$$

whenever $\chi_\Delta(d) = 1$.

If we know that $N^{\text{prim}}_{F,w}(P; m, \mathbf{k}) = 0$ for some value of $\mathbf{k}$, as will be the case in both examples, we then get the following.

Proposition 8.5.1. *Assume that $\Delta | m^\infty$, that F has a non-square determinant and that $N^{prim}_{F,w}(P; m, \mathbf{k}) = 0$. Then for $(d, m) = 1$ and any $\epsilon > 0$ it holds that*

$$N^{prim}_{F,w}(P; m, d\mathbf{k}) = \begin{cases} 0, & \text{if } \chi_\Delta(d) = 1 \\ \frac{2}{L(2,\chi_0)} \sigma_\infty(F, w) L(1, \chi_\Delta) \sigma^*(F, m, \mathbf{k}) m^{-4} P^2, & \text{if } \chi_\Delta(d) = -1 \\ \quad + O_{F,w,m,\epsilon}(P^{5/3+\epsilon}) \end{cases}$$

and

$$\begin{aligned}N_{F,w}(P; m, d\mathbf{k}) =& \sigma_\infty(F, w) L(1, \chi_\Delta) \sigma^*(F, m, \mathbf{k}) m^{-4} P^2 \left(1 - \chi_\Delta(d) \frac{L(2, \chi_0\chi_\Delta)}{L(2, \chi_0)}\right) \\ &+ O_{F,w,m,\epsilon}(P^{5/3+\epsilon}).\end{aligned}$$

Proof. The first part follows by the above discussion and the fact that $\chi_\Delta(d) = 1$ for exactly half the values of $d \in (\mathbb{Z}/m\mathbb{Z})^*$.

For the second part we use (8.2) to get

$$\begin{aligned} N_{F,w}(P; m, d\mathbf{k}) &= \sum_{(l,m)=1} N^{\mathrm{prim}}_{F,w}(P/l, m, d\bar{l}^{(m)}\mathbf{k}) \\ &= C_1 P^2 \sum_{l:\chi_0\chi_\Delta(l)=1} \frac{1}{l^2} + C_2 P^2 \sum_{l:\chi_0\chi_\Delta(l)=-1} \frac{1}{l^2} + O_{F,w,m,\epsilon}(P^{5/3+\epsilon}), \end{aligned}$$

where if $\chi_\Delta(d) = 1$ we have $C_1 = 0$ and $C_2 = 2\sigma_\infty(F, w)L(1, \chi_\Delta)\sigma^*(F, m, \mathbf{k})m^{-4}$, and otherwise C_1 and C_2 swap roles. Finally,

$$\sum_{l:\chi_0\chi_\Delta(l)=1} l^{-2} = \frac{1}{2} \sum_{(l,m)=1} (\chi_\Delta(l) + 1)l^{-2} = \frac{L(2, \chi_0\chi_\Delta) + L(2, \chi_0)}{2},$$

with a similar expression for $\sum_{l:\chi_0\chi_\Delta=-1} l^{-2}$, finishing the proof. □

Example 1. As in the introduction, let

$$F(\mathbf{x}) = x_1^2 - pqx_2^2 - x_3x_4,$$

let $p \equiv q \equiv 1 (\mathrm{mod}\ 8)$ be odd primes and take k such that $\left(\frac{k}{p}\right) = 1$ and $\left(\frac{k}{q}\right) = -1$. Set $m = pq$ and $\mathbf{k} = (k, 0, k, k)$. As shown in the introduction we know that

$$N^{\mathrm{prim}}_{F,w}(P; m, \mathbf{k}) = 0,$$

and so Proposition 8.5.1 is applicable.

Example 2. As a second example, set $m = 4$, $\mathbf{k} = (0, 3, 3, 3)$ and

$$F(\mathbf{x}) = x_1^2 + x_2^2 - x_3x_4.$$

Furthermore, let w be a smooth weight function which satisfies $\mathrm{supp}(w) \subset (0, \infty)^4$. Assume that $\mathbf{x}$ is a primitive solution to $F(\mathbf{x}) = 0$ with $\mathbf{x} \equiv \mathbf{k} (\mathrm{mod}\ m)$ and let $p|x_3$ be an odd prime. Then $x_1^2 + x_2^2 \equiv 0 (\mathrm{mod}\ p)$, and as x_1, x_2 are positive this implies that $p \equiv 1 (\mathrm{mod}\ 4)$ or $p|x_1$ and $p|x_2$. In the second case, as $\mathbf{x}$ is primitive we get $p^2|x_3$. In this way one can show that any prime $p \equiv 3 (\mathrm{mod}\ 4)$ divides x_3 to an even power, contradicting the fact that $x_3 \equiv 3 (\mathrm{mod}\ 4)$, and so there are no primitive solutions. Proposition 8.5.1 is thus applicable also in this case.

Example 3 ([9, 10])**.** Let

$$F(\mathbf{x}) = x_1^2 + 47x_2^2 - 103x_3^2 - 17 \cdot 47 \cdot 103x_4^2.$$

As mentioned in the introduction, Bright shows that for any $\mathbf{k}_0$ satisfying $F(\mathbf{k}_0) \equiv 0 (\text{mod } 17)$, (8.1) fails to have primitive solutions with $\mathbf{k} = a\mathbf{k}_0$ for at least half of the values $a \in (\mathbb{Z}/17\mathbb{Z})^\times$. Combining this with Proposition 8.5.1 then shows that in fact (8.1) has primitive solutions for exactly half of the possible values of a, these being either all the quadratic residues or all the quadratic non-residues.

Appendix A

A colouring of $\mathbb{Z}/p^n\mathbb{Z}$

In [15] Csikvári, Gyarmati and Sárközy give a counterexample showing that the equation

$$x + y = z^2, \quad x, y, z \text{ not all equal} \tag{A.1}$$

is not partition regular over $\mathbb{N}$. In this chapter we show how this construction can be slightly modified to give a counterexample also over $\mathbb{Z}/q\mathbb{Z}$, where $q = p^n$ and p is a prime. The number of colours used will depend on p but not on n.

Theorem A.1. *Let $q = p^n$, with $p > 2$ a fixed prime and $n \in \mathbb{N}$. Then for any such n there is a colouring using $4(p-1)$ colours such that (A.1) has no non-trivial monochromatic solutions in $\mathbb{Z}/q\mathbb{Z}$.*

Proof. We note that the counterexample given in [15] is a 5-adic construction. The changes that need to be made in $\mathbb{Z}/q\mathbb{Z}$ are to replace 5 by p and to make sure that the "wrap-around" in $\mathbb{Z}/q\mathbb{Z}$ is taken care of.

We now define the colouring. For $x \in \mathbb{Z}/q\mathbb{Z}$, write $x = m(x)p^{w(x)} + a(x)$ where $a(x) \in \{0, 1, \ldots, p-1\}$, $p \nmid m(x)$ and $m(x)p^{w(x)} + a(x) \in \{0, 1, \ldots, p^n - 1\}$. The colourclasses are defined by

$$\begin{aligned}
\mathcal{A}_i &= \{x \in \mathbb{Z}/q\mathbb{Z} : a(x) = i\} \quad \text{for } i = 1, 3, 4, \ldots, p-1, \\
\mathcal{B}_i &= \{x \in \mathbb{Z}/q\mathbb{Z} : a(x) = 0, m(x) \equiv i (\text{mod } p), w(x) = 2^{2u}(2v+1), u, v \in \mathbb{Z}_{\geqslant 0}\} \\
&\quad \text{for } i = 1, 2, \ldots, p-1, \\
\mathcal{C}_i &= \{x \in \mathbb{Z}/q\mathbb{Z} : a(x) = 0, m(x) \equiv i (\text{mod } p), w(x) = 2^{2u+1}(2v+1), u, v \in \mathbb{Z}_{\geqslant 0}\} \\
&\quad \text{for } i = 1, 2, \ldots, p-1, \\
\mathcal{D}_i &= \{x \in \mathbb{Z}/q\mathbb{Z} : a(x) = 2, m(x) \equiv i (\text{mod } p), w(x) \geqslant 1\}, \quad \text{for } i = 1, 2, \ldots, p-1, \\
\mathcal{E} &= \{0, 2\},
\end{aligned}$$

which is a $4(p-1)$-colouring of $\mathbb{Z}/q\mathbb{Z}$. We now show that none of the colour classes contain a solution to (A.1).

Case $\mathcal{A}_i$ Let $x, y, z \in \mathcal{A}_i$. Note that $x + y \equiv 2i \pmod{p}$ and $z^2 \equiv i^2 \pmod{p}$, such that $x + y \not\equiv z^2 \pmod{p}$, and thus $x + y \not\equiv z^2 \pmod{q}$, provided $i \neq 0, 2$.

Case $\mathcal{B}_i$ Let $x, y, z \in \mathcal{A}_i$. Write $w(a) = 2^{2u(a)}(2v(a) + 1)$ for $a \in \mathcal{B}_i$. We have that

$$z^2 = m(z)^2 p^{2^{2u(z)+1}(2v(z)+1)}.$$

From here we consider two cases.

- If $w(x) = w(y)$ then

  $$x + y = (m(x) + m(y))p^{2^{2u(x)}(2v(x)+1)},$$

  and since p is odd we have $m(x)+m(y) \equiv 2i \not\equiv 0 \pmod{p}$, so $x+y \not\equiv 0 \pmod{p^n}$. Thus we can assume $2w(z) < n$, such that also $z^2 \not\equiv 0 \pmod{p^n}$. But now $x + y$ and z^2 are divisible by different powers of p, which shows $x + y \neq z^2$.

- If $w(x) \neq w(y)$ assume without loss of generality that $w(x) > w(y)$, such that

  $$x + y = (m(x)p^{w(x)-w(y)} + m(y))p^{w(y)},$$

  but again $m(x)p^{w(x)-w(y)} + m(y) \equiv i \not\equiv 0 \pmod{p}$, and so $x + y$ and z^2 are divisible by different powers of p.

Case $\mathcal{C}_i$ This case is treated almost exactly the same as case $\mathcal{B}_i$.

Case $\mathcal{D}_i$ Let $x, y, z \in \mathcal{D}_i$. We have $z^2 = (m(z)^2 p^{w(z)} + 4m(z))p^{w(z)} + 4$, $w(z) < n$ and $m(z)^2 p^{u(z)} + 4m(z) \equiv 4i \not\equiv 0 \pmod{p}$, so in particular $z^2 - 4 \not\equiv 0 \pmod{p^n}$. Again we consider two cases.

- If $w(x) = w(y)$ we have

  $$x + y = (m(x) + m(y))p^{w(x)} + 4.$$

  Comparing with z^2, subtracting 4 and cancelling as many powers of p as possible we see that $w(x) = w(z)$ and in particular $m(x) + m(y) \equiv 4m(z) \pmod{p}$, but since $2i \not\equiv 4i \pmod{p}$ this cannot hold.

- If $w(x) \neq w(y)$ assume without loss of generality that $w(x) > w(y)$, so that

$$x + y = (m(x)p^{w(x)+w(y)} + m(y))p^{w(y)} + 4.$$

  Doing the same as before we now get $w(y) \equiv 4w(z) (\text{mod } p)$, which is not possible.

Case $\mathcal{E}$ This colour class clearly only contains the trivial solutions $0 + 0 = 0^2$ and $2 + 2 = 2^2$, which are not counted. □

Appendix B

Monochromatic solutions to $x+y$ a square in $\mathbb{Z}/q\mathbb{Z}$

Introduction

The aim of this appendix is to prove the following theorem.

Theorem B.1. *Let $\mathbb{Z}/q\mathbb{Z}$ be k-coloured. Then there is a colour class with $\gg_k |S(q)|q$ tuples x, y such that the sum $x+y$ is a quadratic residue modulo q, where $S(q)$ is the set of quadratic residues modulo q.*

In [42] Khalfalah–Szemerédi prove a similar statement over $\mathbb{N}$, namely that any k-colouring of $\mathbb{N}$ will have monochromatic x, y such that $x+y$ is a perfect square. The proof of Theorem B.1 is just their proof adapted to the modular case.

The outline of the proof is roughly as follows. Let

$$M = \prod_{p \text{ prime} \leqslant P} p, \tag{B.1}$$

where P is some large prime that will be chosen in terms of k. We then restrict ourselves to solutions to

$$x + y \equiv z^2 (\text{mod } q)$$

of the form

$$\begin{aligned} x &= 16Mx' + 2 \\ y &= 16My' + 2 \\ z &= 4Mz' + 2, \end{aligned}$$

If $\{A_i\}_{i=1}^k$ is the k colouring of $\mathbb{Z}/q\mathbb{Z}$ this induces a colouring $\{B_i\}_{i=1}^k$ of $\mathbb{Z}/q'\mathbb{Z}$, where $q' = \frac{q}{(q,16M)}$ and

$$B_i = \{x' \in \mathbb{Z}/q'\mathbb{Z} : 16Mx' + 2 \in A_i\}. \tag{B.2}$$

Let B be the largest induced colour class. We are then aiming to prove that

$$\#\{(x,y) \in B^2 : x+y = Mz^2+z \text{ for some } z \in \mathbb{Z}/q'\mathbb{Z}\} \gg_k |S(q')|q'.$$

As $q' \gg_k q$ this last quantity is $\gg_k |S(q)|q$, and so we will have found the required number of monochromatic solutions to the original problem.

We note that the only property about B that we use is that $|B| \gg_k q'$. This means that we could just as well prove the following slightly stronger version of Theorem B.1.

Theorem B.2. *There is a constant $\eta_k > 0$ which satisfies the following.*

Let $A \subset \mathbb{Z}/q\mathbb{Z}$ satisfying $|A| \geqslant (1-\eta_k)q$ be k-coloured. Then there is a colour class with $\gg_k |S(q)|q$ tuples x, y such that the sum $x+y$ is a quadratic residue modulo q, where $S(q)$ is the set of quadratic residues modulo q.

To prove Theorem B.2, note that $q' \geqslant \frac{q}{16M}$, and so if $\eta_k \leqslant \frac{1}{32M}$ we get that the induced colouring (B.2) covers at least half of $\mathbb{Z}/q'\mathbb{Z}$, such that the largest induced colour class B satisfies $|B| \geqslant \frac{1}{2k}q'$. The rest of the proof is identical to the proof of Theorem B.1.

We will make use of the sets

$$S(q) \coloneqq \{z^2 (\text{mod } q) : z \in \mathbb{Z}/q\mathbb{Z}\}$$

and

$$S(q;M) \coloneqq \{Mz^2 + z (\text{mod } q) : z \in \mathbb{Z}/q\mathbb{Z}\}.$$

In the next section we will record some basic facts about these sets, before giving the proof of Theorem B.1.

Notation: If $X \subset \mathbb{Z}/q\mathbb{Z}$ and $a \in \mathbb{Z}/q\mathbb{Z}$ we will write $aX = \{ax : x \in X\}$ and $X + a = \{x + a : x \in X\}$. If $(a,q) = 1$ we write $\overline{a}$ for the multiplicative inverse of a in $(\mathbb{Z}/q\mathbb{Z})^*$. The modulus q will be clear from context.

Squares modulo p^α

Consider the sum

$$\sum_{x \in S(q;M)} e_q(ax)$$

for some $a \in \mathbb{Z}/q\mathbb{Z}$. For any $x \in S(q;M)$ we can write $x = Mz^2 + z$ for some z. Expressing $z = z_1 \prod_{i\neq 1} p_i^{\alpha_i} + z_2 \prod_{i\neq 2} p_i^{\alpha_i} + \cdots + z_d \prod_{i\neq d} p_i^{\alpha_i}$ where $z_i \in \mathbb{Z}/p_i^{\alpha_i}\mathbb{Z}$ we see that this sum factorizes as

$$\sum_{x\in S(q;M)} e_q(ax) = \prod_{i=1}^{d} \sum_{x\in S(p_i^{\alpha_i};M_i)} e_{p_i^{\alpha_i}}(ax),$$

where $M_i = M\prod_{j\neq i} p_j^{\alpha_j}$. As these are the type of sums we will be interested in, we specialise to prime powers from now on. The next two lemmas show how to sum over the set $S(p^\alpha;M)$ in the case $p|M$ and the case $(p,M)=1$.

Lemma B.3. *Let p be a prime and let $p|M$. Then $S(p^\alpha;M) = \mathbb{Z}/p^\alpha\mathbb{Z}$.*

Proof. If this were not the case there must be some $z, w \in \mathbb{Z}/p^\alpha\mathbb{Z}$, $z \neq w$, such that $Mz^2+z \equiv Mw^2+w (\text{mod } p^\alpha)$. This is the same as $(z-w)(1+M(z+w)) \equiv 0 (\text{mod } p^\alpha)$. But since $p|M$ the second factor is not divisible by p, and so $z - w \equiv 0 (\text{mod } p^\alpha)$, a contradiction. □

Lemma B.4. *Let $p > 2$ be a prime, and let $(p,M) = 1$. Then $S(p^\alpha;M) = MS(p^\alpha) - \overline{4M}$,*

$$\sum_{x\in S(p^\alpha;M)} e_{p^\alpha}(ax) = \frac{1}{2} e_{p^\alpha}(-\overline{4M}) \sum_{\beta=0}^{\lceil \alpha/2\rceil -1} \sideset{}{^*}\sum_{z (\text{mod } p^{\alpha-2\beta})} e_{p^{\alpha-2\beta}}(Maz^2) + 1,$$

and consequently

$$|S(p^\alpha;M)| = |S(p^\alpha)| = \frac{1}{2}\frac{p^\alpha - p^{\alpha - 2\lceil \alpha/2\rceil}}{1+p^{-1}} + 1.$$

Proof. Write $Mz^2 + z = M(z + \overline{2M})^2 - \overline{4M}$, where an overline indicates the inverse modulo p^α. This shows that if x runs over $S(p^\alpha)$, Mx will run over $S(p^\alpha;M) - \overline{4M}$, and so $S(p^\alpha;M) = MS(p^\alpha) - \overline{4M}$.

For a nonzero element $x \in \mathbb{Z}/q\mathbb{Z}$, write $x = ap^\beta$, where $(a,p) = 1$. Then x is a quadratic residue if and only if a is a quadratic residue mod p and $\beta = 2\beta'$ is even [36, Chapter 8.3]. If this is the case there are exactly two values $z \in (\mathbb{Z}/p^{\alpha-\beta}\mathbb{Z})^*$ such that $(zp^{\beta'})^2 \equiv ap^\beta (\text{mod } p^\alpha)$. This means that if we sum over such values of β' and z we will count all squares twice. Finally we need to include the square 0, which gives the expression in the lemma.

□

Proof of Theorem B.1

For notational simplicity we will use q instead of q'. The number of x, y in $\mathbb{Z}/q\mathbb{Z}$ such that $x + y \in S(q; M)$ is given by q^2 times

$$\mathbb{E}_{x,y\in\mathbb{Z}/q\mathbb{Z}} 1_{S(q;M)}(x+y)1_B(x)1_B(y).$$

This can be rewritten as

$$\sum_{\xi\in\mathbb{Z}/q\mathbb{Z}} \widehat{1_{S(q;M)}}(-\xi)\widehat{1_B}(\xi)^2.$$

The term $\xi = 0$ gives

$$\widehat{1_{S(q;M)}}(0)\widehat{1_B}(0)^2 = q^{-3}|S(q;M)||B|^2,$$

which will be the main term. For the other values of ξ we have

$$|\sum_{\xi\neq 0} \widehat{1_{S(q;M)}}(-\xi)\widehat{1_B}(\xi)^2| \leqslant \sup_{\xi\neq 0} |\widehat{1_{S(q;M)}}(\xi)| \sum_{\xi} |\widehat{1_B}(\xi)|^2 \leqslant \sup_{\xi\neq 0} |\widehat{1_{S(q;M)}}(\xi)|.$$

where the last inequality follows from Parseval's identity.

Referring to the discussion in the last section we see that if $q = \prod_{i=1}^d p_i^{\alpha_i}$ then

$$\widehat{1_{S(q;M)}}(\xi) = \prod_{i=1}^d \widehat{1_{S(p_i^{\alpha_i};M_i)}}(\xi), \tag{B.3}$$

where $M_i = M\prod_{j\neq i} p_j^{\alpha_j}$. Here we take $\widehat{1_{S(p^\alpha;M)}}(\xi)$ to be the Fourier transform over $\mathbb{Z}/p^\alpha\mathbb{Z}$ instead of over $\mathbb{Z}/q\mathbb{Z}$. We believe that this abuse of notation should not cause any confusion, as if A is defined as a subset of G then $\widehat{1_A}$ is always taken to be the Fourier transform over G.

Lemma B.5. *If $p|M$ and $\xi \not\equiv 0 (\mathrm{mod}\ p^\alpha)$ then*

$$\widehat{1_{S(p^\alpha;M)}}(\xi) = 0.$$

Proof. This is a direct consequence of Lemma B.3 and the fact that

$$\mathbb{E}_x e(-\xi x) = 0.$$

□

Lemma B.6. *Let $(p, M) = 1$ and $(\xi, p^\alpha) = p^\gamma$ with $\gamma < \alpha$. Then*

$$|\widehat{1_{S_{p^\alpha}(M)}}(\xi)| \leqslant CP^{-1/2}|S_{p^\alpha}(M)|p^{-\alpha},$$

where C is some absolute constant not depending on M and q, and P is the prime appearing in the definition (B.1).

Proof. Applying Lemma B.4 gives

$$\begin{aligned} p^{\alpha}\widehat{1_{S(p^{\alpha};M)}}(\xi) &= \frac{1}{2}\sum_{\beta=0}^{\lceil \alpha/2\rceil -1} \sideset{}{^*}\sum_{z (\mathrm{mod}\ p^{\alpha-2\beta})} e_{p^{\alpha-2\beta}}(-M\xi z^2) + 1 \\ &= \frac{1}{2}\sum_{\beta=0}^{\lceil \frac{\alpha-\gamma}{2}\rceil -1} p^{\gamma} \sideset{}{^*}\sum_{z (\mathrm{mod}\ p^{\alpha-2\beta-\gamma})} e_{p^{\alpha-2\beta-\gamma}}(-M\xi' z^2) \\ &\quad + \frac{1}{2}\frac{p^{\alpha-2\lceil \frac{\alpha-\gamma}{2}\rceil} - p^{\alpha-2\lceil \alpha/2\rceil}}{1+p^{-1}} + 1, \end{aligned}$$

where we write $\xi = \xi' p^{\gamma}$. For the remaining sum over z we have

$$\begin{aligned} &\Big| \sideset{}{^*}\sum_{z (\mathrm{mod}\ p^{\alpha-2\beta-\gamma})} e_{p^{\alpha-2\beta-\gamma}}(-M\xi' z^2)\Big| \\ &\quad \leqslant \Big| \sum_{z (\mathrm{mod}\ p^{\alpha-2\beta-\gamma})} e_{p^{\alpha-2\beta-\gamma}}(-M\xi' z^2)\Big| + \Big| \sum_{z (\mathrm{mod}\ p^{\alpha-2\beta-\gamma-1})} e_{p^{\alpha-2\beta-\gamma-1}}(-pM\xi' z^2)\Big| \\ &\qquad\qquad \leqslant 2p^{\frac{\alpha-2\beta-\gamma}{2}}, \end{aligned}$$

where we use the standard Gauss sum estimate 2.1.3.

This in turn gives

$$p^{\alpha}|\widehat{1_{S(p^{\alpha};M)}}(\xi)| \leqslant p^{\frac{\alpha+\gamma}{2}}\frac{1-p^{\lceil \frac{\alpha+\gamma}{2}\rceil}}{1-p^{-1}} + \frac{1}{2}\frac{p^{\alpha-2\lceil \frac{\alpha-\gamma}{2}\rceil} - p^{\alpha-\gamma-2\lceil \alpha/2\rceil}}{1+p^{-1}} + 1.$$

Comparing this with the size of $S(p^{\alpha};M)$ given in Lemma B.4, and using that $\alpha-\gamma \geqslant 1$, we get that this is $\leqslant Cp^{-1/2}$ for some absolute constant C. Finally we use that $p \nmid M$, i.e. $p \geqslant P$, to finish the proof. □

Proof of Theorem B.1. If $\xi \not\equiv 0 (\mathrm{mod}\ q)$ there is some i such that $p_i^{\alpha_i} \nmid \xi$. Write $(\xi, p_i^{\alpha_i}) = p_i^{\gamma}$. If $\gamma = 0$, Lemma B.5 gives that $\widehat{1_{S(p_i^{\alpha_i};M_i)}}(\xi) = 0$, and thus also $\widehat{1_{S(q;M)}}(\xi) = 0$ by (B.3). If $\gamma \geqslant 1$, we can apply Lemma B.6 together with the trivial bound $\widehat{1_{S(p_j^{\alpha_j};M_j)}}(\xi) \leqslant p^{-\alpha_j}|S(p_j^{\alpha_j};M_j)|$ for $j \neq i$, to get that

$$|\widehat{1_{S(q;M)}}(\xi)| \leqslant CP^{-1/2}q^{-1}|S(q;M)|.$$

Combining these bounds, the fact that $|B| \geqslant \frac{1}{k}q$ and choosing $P \geqslant \frac{C^2}{4k^4}$, we get that

$$\mathbb{E}_{x,y} 1_{S(q;M)}(x+y)1_B(x)1_B(y) \geqslant \frac{1}{2k^2}q^{-1}|S(q;M)| \geqslant \frac{1}{2k^2}q^{-1}|S(q)|,$$

which completes the proof.

□

Appendix C

Some smooth cutoff functions

In the main body of Chapter 4 we required various smooth cutoff functions to (characteristic functions of) discrete intervals, balls in the torus $\mathbb{T}^d$ and Bohr sets. In this appendix we prove the existence of functions with the required properties.

It is convenient to have a C^∞-function $f : \mathbb{R} \to [0,1]$ with $\mathrm{Supp}(f) \subset [-1,1]$ and $\int f(x)dx = 1$. Such a function can be constructed with a "trick", for example defining $f(x) = C\exp(\frac{1}{x^2-1})$ for an appropriate constant C (for a very elegant analysis of this, see [6, Lemma 9]), or by convolving an infinite sequence of normalised characteristic functions of intervals $[-\ell_j, \ell_j]$ with $\sum_j \ell_j \leq 1$.

Let $g : \mathbb{R} \to \mathbb{R}$ be any compactly supported C^∞ function (for example, f). Then, since the Mth derivative $g^{(M)}$ is continuous and supported on $[-1,1]$, we have the bound $\|g^{(M)}\|_\infty = O_M(1)$. By integration by parts this leads to the standard bound

$$|\widehat{g}(\xi)| \ll_M \min(1, |\xi|^{-M}) \tag{C.1}$$

for $\xi \in \mathbb{R}$, where here $\widehat{g}(\xi) = \int_{\mathbb{R}} g(x)e(-\xi x)dx$.

Lemma C.1. *Let $N \in \mathbb{N}$. There is a function $\psi = \psi_N : \mathbb{N} \to [0,\infty)$ with $\psi(n) = 1$ for $N \leq n < 2N$ and $\|\widehat{\psi}\|_1 = O(1)$ (uniformly in N), where the Fourier transform $\widehat{\psi}(t)$ is defined to be $\sum_n \psi(n)e(-tn)$ for $t \in \mathbb{T}$.*

Proof. (Sketch.) Define first a function $g : \mathbb{R} \to \mathbb{R}$ via $g = 1_{[0,3]} * f$. It is easy to check that g is C^∞, compactly supported, and that $g(x) = 1$ for $x \in [1,2]$. We may then define $\psi(n) := g(n/N)$. By the Poisson summation formula we have

$$\widehat{\psi}(\theta) = N\sum_{k\in\mathbb{Z}} \widehat{g}(N(k+\theta)),$$

and so

$$\|\widehat{\psi}\|_1 \leq N\int_\infty^\infty |\widehat{g}(Nu)|du = \|\widehat{g}\|_1,$$

where the ℓ^1 norm on the right is taken on $\mathbb{R}$. The bound $\|\widehat{g}\|_1 = O(1)$ follows quickly by taking $M = 2$ in (C.1).

Alternatively, one may take ψ to be a de la Vallée Poussin type kernel as in the below figure and proceed quite explicitly using the fact that this is a difference of two Fejér kernels. Details may be found in [41, Section 1.2]. □

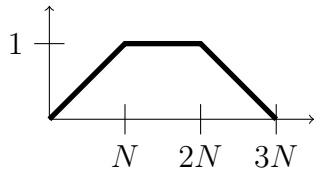

Figure C.1: de la Vallée Poussin kernel.

Suppose now that $\epsilon > 0$ and that $d \in \mathbb{N}$. Let us define $f_\epsilon : \mathbb{T}^d \to [0, \infty)$ by $f_\epsilon(x) = (2\epsilon)^{-d} \prod_{i=1}^d f(\widetilde{x}_i/\epsilon)$, where $\widetilde{x}$ is the unique element of $(-\frac{1}{2}, \frac{1}{2}]^d$ mapping to x under the natural projection. Note that $\int_{\mathbb{T}^d} f_\epsilon(x)dx = 1$.

Lemma C.2. *There is a majorant ψ_ϵ^+ and a minorant ψ_ϵ^- to the ball $B_\epsilon(0)$ in $\mathbb{T}^d$ satisfying*

1. $\frac{1}{2} \leq (2\epsilon)^{-d} \int_{\mathbb{T}^d} \psi_\epsilon^\pm(t)dt \leq 2$ *and*
2. $\sum_{\mathbf{r} \in \mathbb{Z}^d \setminus \{0\}} |\widehat{\psi_\epsilon^\pm}(\mathbf{r})| \|\mathbf{r}\|_1 = O_{\epsilon,d}(1)$.

Proof. We construct ψ_ϵ^+. The construction of ψ_ϵ^- is very similar and is left to the reader. Set $\epsilon' := \epsilon/10d$. For $x \in \mathbb{T}^d$ set

$$\psi_\epsilon^+(x) = 1_{B_{\epsilon+\epsilon'}(0)} * f_{\epsilon'}(x) = \int_{\mathbb{T}^d} f_{\epsilon'}(x-y) 1_{B_{\epsilon+\epsilon'}(0)}(y)dy.$$

Since $f_{\epsilon'}$ is supported on $B_{\epsilon'}(0)$, $\psi_\epsilon^+(x) = 1$ for $x \in B_\epsilon(0)$, and in particular ψ_ϵ^+ is a majorant to the ball $B_\epsilon(0)$.

Moreover ψ_ϵ is bounded pointwise by 1 and is supported on $B_{\epsilon+\epsilon'}(0)$, whence

$$\int_{\mathbb{T}^d} \psi_\epsilon^+(t)dt \leq \mu_{\mathbb{T}^d}(B_{\epsilon+\epsilon'}(0)) = (1 + \frac{\epsilon'}{\epsilon})^d (2\epsilon)^d \leq 2(2\epsilon)^d.$$

Thus (1) is satisfied.

Next we turn to point (2). Suppose that $\mathbf{r} \in \mathbb{Z}^d \setminus \{0\}$. Write $\mathbf{r} = (r_1, \ldots, r_d)$, and assume without loss of generality that $|r_1| = \|\mathbf{r}\|_\infty$. Performing M integration by parts in the integral

$$\widehat{\psi_\epsilon^+}(\mathbf{r}) = \int_{\mathbb{T}^d} \psi_\epsilon^+(x) e(-x \cdot \mathbf{r})dx$$

with respect to x_1, to get that

$$\widehat{\psi_\epsilon^+}(\mathbf{r}) = \frac{1}{(-2\pi i r_1)^M} \int \frac{\partial^M \psi_\epsilon^+(x)}{\partial x_1^M} e(-x \cdot \mathbf{r}) dx \ll_{\epsilon,d,M} \|\mathbf{r}\|_\infty^{-M}$$

for any $M \in \mathbb{N}$ (this is essentially the same bound as (C.1)). The ℓ^1 and ℓ^∞ norms of $\mathbf{r}$ are comparable up to factors of $O_d(1)$, and hence

$$\sum_{\mathbf{r} \in \mathbb{Z}^d \setminus \{0\}} |\widehat{\psi_\epsilon^+}(\mathbf{r})| \|\mathbf{r}\|_1 \ll_{\epsilon,d,M} \sum_{\mathbf{r} \in \mathbb{Z}^d \setminus \{0\}} \|\mathbf{r}\|_1^{1-M}.$$

Taking $M = d+2$, it is easy to see that the sum on the right converges and is bounded by $O_d(1)$. □

Finally we turn to the most complicated of our constructions, a smooth approximant for the Bohr-type set X considered in Section 4.5.

Lemma C.3. *Let $0 < \epsilon' < \epsilon < 1$, $d, q \in \mathbb{N}$, $x \in \mathbb{R}$ and $\theta, z \in \mathbb{T}^d$. Then there is an $A = A(\epsilon, \epsilon', d, q)$ with the following property. Suppose that N is sufficiently large in terms of $\epsilon, \epsilon', d, q, A$. Set*

$$X = \{n \in \mathbb{N} : n \equiv u (\mathrm{mod}\ q), |\frac{n}{N} - x|, \|\theta n - z\|_{\mathbb{T}^d} \leq \epsilon\}$$

and

$$X_- = \{n \in \mathbb{N} : n \equiv u (\mathrm{mod}\ q), |\frac{n}{N} - x|, \|\theta n - z\|_{\mathbb{T}^d} \leq \epsilon - \epsilon'\}.$$

Suppose that $\epsilon' < \epsilon/10d$ and θ is (A, N)-irrational. Then there exists a function χ satisfying

1. *$1_{X_-}(n) \leq \chi(n) \leq 1_X(n)$ for all n;*
2. *$\|\widehat{\chi}\|_1 = O_{\epsilon,\epsilon',q,d}(1)$ and*
3. *$\sum_n \chi(n) \geq \frac{1}{2}(2\epsilon)^{d+1} q^{-1} N$.*

Proof. Let $g : \mathbb{R} \to [0, \infty)$ be a C^∞ function with $g(t) = 1$ for $|t - x| \leq \epsilon - \epsilon'$ and $g(t) = 0$ for $|t - x| > \epsilon$. Such a function can be obtained by convolving the characteristic function of the interval $\{t : |t - x| \leq \epsilon - \frac{1}{2}\epsilon'\}$ with the function $\frac{2}{\epsilon'} f(\frac{2t}{\epsilon'})$.

Define a function $h : \mathbb{T}^d \to [0, \infty)$ by

$$h := f_{\epsilon'/2} * 1_{B_{\epsilon - \epsilon'/2}(z)}.$$

Now define

$$\chi(n) := g(\frac{n}{N}) h(\theta n) 1_{n \equiv u (\mathrm{mod}\ q)}.$$

The relevant support properties (1) may be easily checked. Turning to point (2), we begin by noting the expansion

$$1_{n\equiv u(\mathrm{mod}\ q)} = q^{-1}\sum_{s\in\mathbb{Z}/q\mathbb{Z}} e(\frac{(n-u)s}{q}).$$

This implies that

$$\widehat{\chi}(t) = q^{-1}\sum_{s\in\mathbb{Z}/q\mathbb{Z}} e(-\frac{us}{q})\widehat{g(\frac{\cdot}{N})h(\theta\cdot)}(t+\frac{s}{q}). \tag{C.2}$$

Therefore in order to establish (2) it suffices to prove that

$$\left\|\widehat{g(\frac{\cdot}{N})h(\theta\cdot)}\right\|_1 = O_{\epsilon,\epsilon',d}(1). \tag{C.3}$$

Fourier expanding h and applying Poisson summation, we have

$$\begin{aligned}\widehat{g(\frac{\cdot}{N})h(\theta\cdot)}(t) &= \sum_n g(\frac{n}{N})h(\theta n)e(-tn)\\ &= \sum_n g(\frac{n}{N})\sum_{\mathbf{r}}\widehat{h}(\mathbf{r})e\big((\mathbf{r}\cdot\theta - t)n\big)\\ &= N\sum_{\mathbf{r}}\widehat{h}(\mathbf{r})\sum_{k\in\mathbb{Z}}\widehat{g}\big(N(t+k-\mathbf{r}\cdot\theta)\big).\end{aligned} \tag{C.4}$$

Thus

$$\|\widehat{g(\frac{\cdot}{N})h(\theta\cdot)}\|_1 \le N\sum_{\mathbf{r}}|\widehat{h}(\mathbf{r})|\int_{-\infty}^{\infty}|\widehat{g}(Nu)|du = \|\widehat{g}\|_1\|\widehat{h}\|_1,$$

where here the ℓ^1 norms are on $\mathbb{Z}^d$ and $\mathbb{R}$ respectively.

That $\|\widehat{g}\|_1 \ll_{\epsilon,\epsilon'} 1$ follows immediately from (C.1) with $M = 2$.

By essentially the same reasoning used in the proof of Lemma C.2 we have

$$|\widehat{h}(\mathbf{r})| \ll_{\epsilon,\epsilon',d,M} \|\mathbf{r}\|_\infty^{-M}. \tag{C.5}$$

Taking $M = d+1$ we obtain

$$\|\widehat{h}\|_1 = O_{\epsilon,\epsilon',d}(1).$$

Putting these facts together completes the proof of (C.3) and hence of (2).

It remains to verify (3). Note that we have not yet used the irrationality of θ. From (C.2) we have

$$\sum_n \chi(n) = \widehat{\chi}(0) = q^{-1}\sum_{s\in\mathbb{Z}/q\mathbb{Z}} e(-\frac{us}{q})\widehat{g(\frac{\cdot}{N})h(\theta\cdot)}(\frac{s}{q}).$$

By (C.4), it follows that

$$\sum_n \chi(n) = Nq^{-1} \sum_{\mathbf{r}\in\mathbb{Z}^d} \sum_{s\in\mathbb{Z}/q\mathbb{Z}} \sum_{k\in\mathbb{Z}} e(-\frac{us}{q})\widehat{h}(\mathbf{r})\widehat{g}\big(N(\frac{s}{q}+k-\mathbf{r}\cdot\theta)\big). \tag{C.6}$$

The contribution from $\mathbf{r}=0$, $s=0$, $k=0$ is $Nq^{-1}(\int_{\mathbb{T}^d} h)(\int_{\mathbb{R}} g)$. Since $\epsilon' < \epsilon/10d$ we have $\int_{\mathbb{T}^d} h \geq \mu_{\mathbb{T}^d}(B_{\epsilon-\epsilon'}(0)) \geq 0.9(2\epsilon)^d$, and evidently $\int_{\mathbb{R}} g \geq 2(\epsilon-\epsilon') > 0.9(2\epsilon)$. Thus the contribution from this term is $\geq \frac{3}{4}(2\epsilon)^{d+1}q^{-1}N$. To complete the proof of (3) it suffices to show that the contribution of the other terms to (C.6) is at most $\frac{1}{4}(2\epsilon)^{d+1}q^{-1}N$, to which end it is enough to show that

$$\sum_{\mathbf{r}\in\mathbb{Z}^d} \sum_{s\in\mathbb{Z}/q\mathbb{Z}} \sum_{k\in\mathbb{Z}} |\widehat{h}(\mathbf{r})||\widehat{g}(N(\frac{s}{q}+k-\mathbf{r}\cdot\theta)| \leq \frac{1}{4}(2\epsilon)^{d+1}, \tag{C.7}$$

where the sum omits the term $\mathbf{r}=0$, $s=0$, $k=0$.

By (C.5) (with $M=d+1$) and (C.1) (with $M=2$), the left hand side is bounded by

$$O_{\epsilon,\epsilon',d}(1) \sum_{\mathbf{r}\in\mathbb{Z}^d} \sum_{s\in\mathbb{Z}/q\mathbb{Z}} \sum_{k\in\mathbb{Z}} \min(1, \|\mathbf{r}\|^{-d-1}) \min(1, N^{-2}|k+\frac{s}{q}-\mathbf{r}\cdot\theta|^{-2}). \tag{C.8}$$

If $0 < \|\mathbf{r}\|_1 \leq A/q$ then it follows from the fact that θ is (A,N)-irrational that $|k+\frac{s}{q}-\mathbf{r}\theta| \geq \frac{A}{qN}$ (no matter the value of s or k). The same is trivially true when $\mathbf{r}=0$, provided that not both of s,k are zero and that N is sufficiently large. In the inner sum over k in (C.8), the contribution from all but at most one term is $\ll N^{-2}\sum_{m\in\mathbb{Z}\setminus\{0\}}|m|^{-2} \ll N^{-2}$, and so when $\|\mathbf{r}\|_1 \leq A/q$ the inner sum over k is $\ll \frac{q^2}{A^2}+N^{-2}$, which is $\ll q^2/A^2$ if N is big enough. Therefore

$$\sum_{\substack{\mathbf{r}\in\mathbb{Z}^d \\ \|\mathbf{r}\|\leq A/q}} \sum_{s\in\mathbb{Z}/q\mathbb{Z}} \sum_{k\in\mathbb{Z}} \min(1, \|\mathbf{r}\|^{-d-1}) \min(1, N^{-2}|k+\frac{s}{q}-\mathbf{r}\cdot\theta|^{-2})$$

$$\ll \frac{q^3}{A^2} \sum_{\mathbf{r}} \|\mathbf{r}\|^{-d-1} \ll_{d,q} A^{-2}.$$

All other terms in (C.8) have $\|\mathbf{r}\| \geq \frac{A}{q}$. Using the trivial bound

$$\sum_{k\in\mathbb{Z}} \min(1, N^{-2}|k+\frac{s}{q}-\mathbf{r}\cdot\theta|^{-2}) \ll 1,$$

the contribution from these is bounded by

$$O_{d,\epsilon,\epsilon',q}(1) \sum_{\|\mathbf{r}\|\geq A/q} \|\mathbf{r}\|^{-d-1} \ll_{d,\epsilon,\epsilon',q} A^{-1}.$$

Putting all of this together shows that (C.8) is bounded by $O_{d,\epsilon,\epsilon,q}(A^{-1})$, and so (C.7) does indeed hold if A is large enough as a function of ϵ,ϵ',d,q. □

Appendix D

Restriction estimates and counting estimates for kth powers

Unrestricted counting and mean value estimates

Definition D.1 (R-smooth numbers)**.** We say that a positive integer is *R-smooth* if all of its prime divisors are at most R.

Lemma D.2. *There are at most $10^w N M^{-1/2}$ elements of $[N]$ divisible by a w-smooth number greater than M.*

Proof. It follows from Rankin's trick that the number of integers in $[N]$ divisible by a w-smooth number exceeding M is at most

$$\sum_{\substack{m>M \\ m \text{ is } w\text{-smooth}}} \frac{N}{m} \leqslant \sum_{m \text{ is } w\text{-smooth}} \frac{N}{m}\Big(\frac{m}{M}\Big)^{1/2} = NM^{-1/2} \prod_{p \leq w} \Big(1 + \frac{1}{p^{1/2}-1}\Big).$$

The result follows on noting that $1 + \frac{1}{p^{1/2}-1} \leq 10$. □

The following is a known consequence of Bourgain–Demeter–Guth [8].

Lemma D.3. *Let $c_1, \dots, c_s \in \mathbb{Z} \setminus \{0\}$ with $s \geqslant W(k)$, where $W(2) = 5$, $W(3) = 8$ and*

$$W(k) = k^2 - 1 \qquad (k \geqslant 4).$$

Then for large N we have

$$\#\Big\{x \in [N]^s : \sum_{i=1}^{s} c_i x_i^k = 0\Big\} \ll_{\mathbf{c}} N^{s-k}.$$

Moreover, if $\sum_{i \in I} c_i = 0$ for some $I \neq \emptyset$ then we have

$$\#\Big\{x \in [N]^s : \sum_{i=1}^{s} c_i x_i^k = 0\Big\} \gg_{\mathbf{c}} N^{s-k}.$$

One proves this using the circle method [66], imitating [71, Theorem 4.1] if $k \geqslant 4$; see also [13, pages 4–5]. The local solubility conditions are certainly met if $\sum_{i\in I} c_i = 0$ for some $I \neq \emptyset$, leading to the lower bound. For $k = 2$, the result was known to Hardy and Littlewood. In an influential paper, Kloosterman [43] opened with a discussion of this, then adapted the Hardy–Littlewood method to address the quaternary problem. For $k = 3$ there is Vaughan's breakthrough result [65].

We also need the following bounded restriction inequality.

Lemma D.4. *If $p > k^2$ is a real number and $f : [N] \to [0, 1]$ then*

$$\int_{\mathbb{T}} \Big|\sum_{n\leqslant N} f(n)e(\alpha n^k)\Big|^p \mathrm{d}\alpha \ll_p N^{p-k}.$$

Proof. When $k = 2$ this follows quickly from [7, Eq. (4.1)]. We now assume that $k \geqslant 3$, and write $2t$ for the greatest even integer less than or equal to k^2. by orthogonality and Lemma D.3, we have

$$\int_{\mathbb{T}} \Big|\sum_{n\leqslant N} f(n)e(\alpha n^k)\Big|^{2t} \mathrm{d}\alpha \leqslant \#\Big\{(x, y) \in [N]^t \times [N]^t : \sum_{i\leqslant t} x_i^k = \sum_{i\leqslant t} y_i^k\Big\} \ll N^{2t-k}.$$

The trivial estimate $\big|\sum_{n\leqslant N} f(n)e(\alpha n^k)\big| \leqslant N$ completes the proof. □

Finally we need an upper bound on the number of trivial solutions.

Lemma D.5. *Let $c_1, \dots, c_s \in \mathbb{Z} \setminus \{0\}$ with $s \geq k^2 + 1$. Then*

$$\#\Big\{x \in [N]^s : \sum_{i=1}^{s} c_i x_i^k = 0 \text{ and } x_i = x_j \text{ for some } i \neq j\Big\} \ll_{\mathbf{c},\epsilon} N^{s-k-\frac{1}{2}+\epsilon}.$$

Proof. By the union bound, it suffices to prove an estimate of this shape for the number of solutions with $x_{s-1} = x_s$. In this case we are estimating

$$\#\Big\{x \in [N]^{s-1} : \sum_{i=1}^{s-2} c_i x_i^k + (c_{s-1} + c_s)x_{s-1}^k = 0\Big\}.$$

It may be that $c_{s-1} + c_s = 0$, so we estimate the contribution from the x_{s-1} variables trivially. Using orthogonality and Hölder's inequality we are therefore reduced to showing that

$$\int_{\mathbb{T}} \Big|\sum_{n\leq N} e(\alpha n^k)\Big|^{s-2} \mathrm{d}\alpha \ll_\epsilon N^{s-1-k-\frac{1}{2}+\epsilon}.$$

When $s-2>k^2$ this follows from Lemma D.4. When $s-2\le k^2$ we apply Hölder's inequality to deduce that

$$\int_{\mathbb{T}}\Big|\sum_{n\le N}e(\alpha n^k)\Big|^{s-2}\mathrm{d}\alpha \ll \left(\int_{\mathbb{T}}\Big|\sum_{n\le N}e(\alpha n^k)\Big|^{k^2}\mathrm{d}\alpha\right)^{\frac{s-2}{k^2}} \ll_\epsilon N^{s-2-\frac{s-2}{k}+\epsilon}.$$

When $k\ge 3$ the latter inequality follows without the N^ϵ factor by employing Lemma D.3. When $k=2$ the estimate is classical, and recorded for instance in [7, Eq. (1.9)].

Finally we observe that

$$2+\tfrac{s-2}{k}\ge 2+\tfrac{k^2-1}{k}=k+2-\tfrac{1}{k}\ge k+\tfrac{3}{2}.$$

□

A generalised von Neumann lemma

Recall the notion of p-restriction introduced in Definition 6.4.4.

Lemma D.6. *Let $\nu_1,\nu_2:[N]\to[0,\infty)$. If both ν_1 and ν_2 satisfy a p-restriction estimate with constant K, then so does $\nu_1+\nu_2$.*

Proof. Let $|\phi|\le\nu_1+\nu_2$. Then $\phi=\psi\times\theta$, where $\psi:[N]\to[0,\infty)$ satisfies $\psi\le\nu_1+\nu_2$ and $\theta:[N]\to\mathbb{C}$ satisfies $|\theta|\le 1$. Put $\psi_1:=\min\{\psi,\nu_1\}$ and $\psi_2:=\psi-\psi_1$. On setting $\phi_i:=\psi_i\theta$, we have $\phi=\phi_1+\phi_2$ with $|\phi_i|\le\nu_i$. Applying the triangle inequality and restriction estimates for each ν_i gives

$$\begin{aligned}\big\|\widehat{\phi}\big\|_p &\le \big\|\widehat{\phi_1}\big\|_p+\big\|\widehat{\phi_2}\big\|_p\\ &\le (K/N)^{1/p}\left(\|\nu_1\|_1+\|\nu_2\|_1\right).\end{aligned}$$

Positivity gives that $\|\nu_1\|_1+\|\nu_2\|_1=\|\nu_1+\nu_2\|_1$, and the result then follows on taking pth powers. □

Lemma D.7. *Let $c_1,\dots,c_s\in\mathbb{Z}\setminus\{0\}$, $\delta\in(0,1)$ and suppose that $\nu_1,\dots,\nu_s:[N]\to[0,\infty)$ each satisfy a $(s-\delta)$-restriction estimate with constant K. Then for any $|f_i|\le\nu_i$ we have*

$$\left|\sum_{\mathbf{c}\cdot\mathbf{x}=0}\frac{f_1(x_1)}{\|\nu_1\|_1}\dots\frac{f_s(x_s)}{\|\nu_s\|_1}\right|\le\frac{K}{N}\min_i\left\|\frac{\widehat{f_i}}{\|\nu_i\|_1}\right\|_\infty^{\delta}. \tag{D.1}$$

Proof. We prove the upper bound with $i = 1$, the remaining cases following by relabeling indices. Let $p = s - \delta$. by orthogonality and Hölder's inequality, we have

$$\Big|\sum_{\mathbf{c}\cdot\mathbf{x}=0} f_1(x_1)\cdots f_s(x_s)\Big| = \Big|\int_{\mathbb{T}} \widehat{f}_1(c_1\alpha)\cdots\widehat{f}_s(c_s\alpha)\mathrm{d}\alpha\Big| \leqslant \int_{\mathbb{T}}\Big|\widehat{f}_1(c_1\alpha)\cdots\widehat{f}_s(c_s\alpha)\Big|\mathrm{d}\alpha$$
$$\leq \big\|\widehat{f}_1\big\|_\infty^{\delta}\big\|\widehat{f}_1\big\|_p^{1-\delta}\big\|\widehat{f}_2\big\|_p\cdots\big\|\widehat{f}_s\big\|_p.$$

Inequality (D.1) then follows from our p-restriction assumption. □

Lemma D.8 (Generalised von Neumann)**.** *Let $c_1,\dots,c_s \in \mathbb{Z}\setminus\{0\}$, $\delta\in(0,1)$ and suppose that $\nu_i,\mu_i : [N]\to[0,\infty)$ each satisfy a $(s-\delta)$-restriction estimate with constant K. Then for any $|f_i|\leq\nu_i$ and $|g_i|\leq\mu_i$ we have*

$$\Big|\sum_{\mathbf{c}\cdot\mathbf{x}=0}\Big(\frac{f_1(x_1)}{\|\nu_1\|_1}\cdots\frac{f_s(x_s)}{\|\nu_s\|_1} - \frac{g_1(x_1)}{\|\mu_1\|_1}\cdots\frac{g_s(x_s)}{\|\mu_s\|_1}\Big)\Big|$$
$$\leq \frac{sK}{N}\max_i\Big\|\frac{\widehat{f}_i}{\|\nu_i\|_1} - \frac{\widehat{g}_i}{\|\mu_i\|_1}\Big\|_\infty^{\delta}.$$

Proof. Let $p = s-\delta$. By Lemma D.6, the weight

$$\frac{\nu_i}{\|\nu_i\|_1} + \frac{\mu_i}{\|\mu_i\|_1}$$

satisfies a p-restriction estimate with constant K and majorises the difference

$$\frac{f_i}{\|\nu_i\|_1} - \frac{g_i}{\|\mu_i\|_1}.$$

Observing that this weight has L^1 norm equal to two, the lemma follows on applying the telescoping identity

$$a_1\cdots a_s - b_1\cdots b_s = \sum_{i=1}^{s}(a_i-b_i)\prod_{j<i}a_j\prod_{j>i}b_j,$$

together with Lemma D.7. □

Pointwise exponential sum estimates

The primary objective of this section is to establish Fourier decay (Lemma 6.5.3). Define

$$S(q,a) = \sum_{r \bmod q} e\Big(\frac{a}{q}\cdot\frac{(Wr+\xi)^k-\xi^k}{kW}\Big)$$

and

$$I(\beta) = \int_0^X e(\beta z)\mathrm{d}z,$$

where W is defined by (6.15) and ξ is as in (6.22).

Lemma D.9. *Suppose $q \in \mathbb{N}$, $a \in \mathbb{Z}$ and $\|q\alpha\| = |q\alpha - a|$. Then*

$$\widehat{\nu}(\alpha) = q^{-1}S(q,a)I(\alpha - a/q) + O((XW)^{\frac{k-1}{k}}(q + X\|q\alpha\|)).$$

Proof. Observe that

$$\frac{(Wy+\xi)^k - \xi^k}{kW}$$

is a polynomial in y with integer coefficients. With $\beta = \alpha - a/q$, we compute that

$$\begin{aligned}\widehat{\nu}(\alpha) &= \sum_{y \leqslant Y} (Wy+\xi)^{k-1} e\Big(\alpha \frac{(Wy+\xi)^k - \xi^k}{kW}\Big) \\ &= \sum_{r \leqslant q} e\Big(\frac{a}{q} \cdot \frac{(Wr+\xi)^k - \xi^k}{kW}\Big) \sum_{-\frac{r}{q} < x \leqslant \frac{Y-r}{q}} \phi(x),\end{aligned}$$

where

$$\phi(x) = (W(qx+r)+\xi)^{k-1} e\Big(\beta \frac{(W(qx+r)+\xi)^k - \xi^k}{kW}\Big).$$

As in the proof of [11, Lemma 5.1], we apply Euler–Maclaurin summation (Lemma 2.1.1). This gives

$$\begin{aligned}\sum_{-\frac{r}{q} < x \leqslant \frac{Y-r}{q}} \phi(x) &= \int_{-r/q}^{(Y-r)/q} \phi(t)\mathrm{d}t \\ &\qquad + O(\|\phi\|_{L^\infty([-r/q,(Y-r)/q])} + \frac{Y}{q}\|\phi'\|_{L^\infty([-r/q,(Y-r)/q])}) \\ &= \int_{-r/q}^{(Y-r)/q} \phi(t)\mathrm{d}t + O((XW)^{\frac{k-1}{k}}(1 + X|\beta|)).\end{aligned}$$

For the main term we make the change of variables

$$z = \frac{(W(qt+r)+\xi)^k - \xi^k}{kW},$$

giving

$$\int_{-r/q}^{(Y-r)/q} \phi(t)\mathrm{d}t = q^{-1}\int_0^{X+O(X/Y)} e(\beta z)\mathrm{d}z = q^{-1}(I(\beta) + O(X/Y)).$$

Therefore

$$\sum_{-\frac{r}{q} < x \leqslant \frac{Y-r}{q}} \phi(x) = q^{-1}I(\beta) + O((XW)^{\frac{k-1}{k}}(1 + X|\beta|)),$$

and so

$$\widehat{\nu}(\alpha) = q^{-1}S(q,a)I(\alpha - a/q) + O((XW)^{\frac{k-1}{k}}(q + X\|q\alpha\|)).$$

□

Next, we make a Hardy–Littlewood dissection. For $q \in \mathbb{N}$ and $a \in \mathbb{Z}$, let $\mathfrak{M}(q,a)$ be the set of $\alpha \in \mathbb{T}$ such that $|\alpha - a/q| \leqslant Q/X$, where

$$Q = Y^{\kappa}, \qquad \kappa = \frac{1}{100}.$$

Let $\mathfrak{M}(q)$ be the union of the sets $\mathfrak{M}(q,a)$ over integers a such that $(a,q) = 1$, and let $\mathfrak{M}$ be the union of the sets $\mathfrak{M}(q)$ over $q \leqslant Q$. Put $\mathfrak{m} = \mathbb{T} \setminus \mathfrak{M}$. By identifying $\mathbb{T}$ with a unit interval, we may write

$$\mathfrak{M}(q) = \bigcup_{\substack{a=0 \\ (a,q)=1}}^{q-1} \mathfrak{M}(q,a).$$

Lemma D.10. *If $\alpha \in \mathfrak{m}$ then $\widehat{\nu}(\alpha) \ll XY^{-\kappa 2^{-k}}$.*

Proof. Let $\alpha \in \mathfrak{m}$. by partial summation

$$\widehat{\nu}(\alpha) \ll (WY)^{k-1} \sup_{T \leqslant Y} |g(\alpha, T)|,$$

where

$$g(\alpha, T) = \sum_{y \leqslant T} e\Big(\alpha \frac{(Wy+\xi)^k - \xi^k}{kW}\Big).$$

Put $\gamma = W^{k-1}\alpha/k$ and, by Dirichlet's approximation theorem (Lemma 2.1.5), choose relatively prime integers r, a such that

$$1 \leqslant r \leqslant X/Q, \qquad |r\gamma - a| \leqslant Q/X.$$

Now Weyl's inequality (Lemma 2.1.6) yields

$$g(\alpha, T) \ll T^{1+\epsilon}(r^{-1} + T^{-1} + rT^{-k})^{2^{1-k}}.$$

With $q = rW^{k-1}/k$ we have $|\alpha - a/q| \leqslant |q\alpha - a| \leqslant Q/X$, so as $\alpha \notin \mathfrak{M}$ we must have $q > Q$, and so $r > kQW^{1-k}$. Now

$$\frac{1}{r} < \frac{W^{k-1}}{kQ} \leqslant \frac{X}{QT^k},$$

and so

$$\begin{aligned} g(\alpha, T) &\ll T^{1+\epsilon}\left(\frac{1}{T} + \frac{X}{QT^k}\right)^{2^{1-k}} \\ &\ll T^{1-2^{1-k}+\epsilon} + T^{1-k2^{1-k}+\epsilon}\left(\frac{W^{k-1}Y^k}{Q}\right)^{2^{1-k}}. \end{aligned}$$

As $T \leqslant Y$ and Y is large compared to $w = w(\delta, M)$, we obtain

$$g(\alpha, T) \ll Y^{1-\kappa 2^{-k}}.$$

Therefore

$$\widehat{\nu}(\alpha) \ll W^{k-1} Y^{k-\kappa 2^{-k}} \ll X Y^{-\kappa 2^{-k}}.$$

□

Lemma D.11. *If* $(q, a) = 1$ *then* $S(1, 0) = 1$,

$$S(q, a) = 0 \qquad (2 \leqslant q \leqslant w) \tag{D.2}$$

and

$$S(q, a) \ll q^{1-1/k}. \tag{D.3}$$

Proof. Plainly $S(1, 0) = 1$, so let $q \geqslant 2$, and let $a \in \mathbb{Z}$ with $(q, a) = 1$. The Binomial expansion gives

$$S(q, a) = \sum_{r \bmod q} e_q\left(a \sum_{\ell=1}^{k} \frac{\binom{k}{\ell} W^{\ell-1}}{k} \xi^{k-\ell} r^{\ell} \right),$$

and we note that $\frac{\binom{k}{\ell} W^{\ell-1}}{k} \in \mathbb{Z}$ $(1 \leqslant \ell \leqslant k)$. Write $q = uv$, where u is w-smooth and $(v, W) = 1$. Since $(u, v) = 1$, a standard calculation reveals that

$$S(q, a) = S(u, a_1) S(v, a_2), \tag{D.4}$$

where $a_1 = a v^{-1} \in (\mathbb{Z}/u\mathbb{Z})^{\times}$ and $a_2 = a u^{-1} \in (\mathbb{Z}/v\mathbb{Z})^{\times}$ (see [66, Lemma 2.10]).

Put $u = hu'$, where $h = (u, W/k)$. Letting $r = r_1 + u' r_2$, where $r_1 \bmod u'$ and $r_2 \bmod h$, gives

$$\begin{aligned} S(u, a_1) &= \sum_{\substack{r_1 \bmod u' \\ r_2 \bmod h}} e_{hu'}\left(a_1 \sum_{\ell=1}^{k} \binom{k}{\ell} \frac{\binom{k}{\ell} W^{\ell-1}}{k} \xi^{k-\ell} (r_1 + u' r_2)^{\ell} \right) \\ &= \sum_{r_1=0}^{u'-1} e_{hu'}\left(a_1 \sum_{\ell=1}^{k} \frac{\binom{k}{\ell} W^{\ell-1}}{k} \xi^{k-\ell} r_1^{\ell} \right) \\ &\quad \cdot \sum_{r_2=0}^{h-1} e_h\left(a_1 \sum_{\ell=1}^{k} \frac{\binom{k}{\ell} W^{\ell-1}}{k} \xi^{k-\ell} (u')^{\ell-1} r_2^{\ell} \right). \end{aligned}$$

The inner sum is

$$\sum_{r_2 \bmod h} e_h(a_1 \xi^{k-1} r_2),$$

which vanishes unless $h \mid a_1\xi^{k-1}$. As $(h, a_1) = (h, \xi) = 1$, and as

$$(u, W/k) = 1 \Leftrightarrow (u, W) = 1 \Leftrightarrow u = 1,$$

we conclude that

$$S(u, a_1) = \begin{cases} 0 & \text{if } u \neq 1 \\ 1 & \text{if } u = 1. \end{cases} \tag{D.5}$$

This completes the proof of (D.2), by (D.4).

Next we prove (D.3). By (D.4) and (D.5), we may assume $u = 1$. Consider

$$e_{kWv}(a_2\xi^k)S(v, a_2) = \sum_{r \bmod v} e_v\Big(a_2\frac{(Wr+\xi)^k}{kW}\Big).$$

As $(v, W) = 1$, we can change variables by $t = \xi W^{-1} + r \in \mathbb{Z}/v\mathbb{Z}$, which gives

$$e_{kWv}(a_2\xi^k)S(v, a_2) = \sum_{t \bmod v} e_v\Big(a_2\frac{W^{k-1}}{k}t^k\Big).$$

Since $\Big(v, a_2\frac{W^{k-1}}{k}\Big) = 1$, we may apply [66, Theorem 4.2], which gives

$$S(v, a_2) \ll v^{1-1/k} = q^{1-1/k}.$$

by (D.4) and (D.5), we now have $S(q, a) \ll q^{1-1/k}$. □

It is convenient to also record the following standard estimate, for later use.

Lemma D.12. *We have*

$$I(\beta) \ll \min\{X, \|\beta\|^{-1}\}.$$

We are ready to prove Lemma 6.5.3. We need to show that if $\alpha \in \mathbb{T}$ then

$$\widehat{\nu}(\alpha) - \widehat{1_{[X]}}(\alpha) \ll Xw^{-1/k}. \tag{D.6}$$

by a geometric series, we have

$$\widehat{1_{[X]}}(\alpha) = \sum_{x \leqslant X} e(\alpha x) \leqslant \|\alpha\|^{-1}. \tag{D.7}$$

First suppose $\alpha \in \mathfrak{m}$. By Lemma 2.1.5, we obtain relatively prime integers q and a such that $1 \leqslant q \leqslant Q$ and $|q\alpha - a| \leqslant Q^{-1}$. As $\alpha \notin \mathfrak{M}$, we must have $|q\alpha - a| > qQ/X$, so

$$\widehat{1_{[X]}}(\alpha) \ll \|\alpha\|^{-1} \ll \frac{q}{\|q\alpha\|} \ll X/Q.$$

by Lemma D.10, we now have (D.6).

Next we consider the case in which $q = 1$ and $\alpha \in \mathfrak{M}(q)$, in other words $\|\alpha\| \leqslant Q/X$. by Lemma D.9, we have

$$\widehat{\nu}(\alpha) - I(\alpha) \ll (XW)^{\frac{k-1}{k}}(1 + X\|\alpha\|) \ll (XW)^{\frac{k-1}{k}}Q. \tag{D.8}$$

by Euler–Maclaurin summation (Lemma 2.1.1), we have

$$\widehat{1_{[X]}}(\alpha) - I(\alpha) \ll 1 + X\|\alpha\| \ll Q. \tag{D.9}$$

Coupling (D.8) with (D.9) yields

$$\widehat{\nu}(\alpha) - \widehat{1_{[X]}}(\alpha) \ll (XW)^{\frac{k-1}{k}}Q \ll Xw^{-1/k}.$$

Finally, let $\alpha \in \mathfrak{M}(q, a)$ with $2 \leqslant q \leqslant Q$ and $(a, q) = 1$, and put

$$\beta = \alpha - a/q \in [-Q/X, Q/X].$$

Substituting

$$\|\alpha\| \geqslant q^{-1} - |\beta| \geqslant q^{-1} - Q/X \gg q^{-1}$$

into (D.7) gives

$$\widehat{1_{[X]}}(\alpha) \ll q \ll Q.$$

by Lemma D.9, we also have

$$\widehat{\nu}(\alpha) \ll Q^2(XW)^{\frac{k-1}{k}} + X|q^{-1}S(q, a)|,$$

and now Lemma D.11 yields (D.6).

The restriction estimate

In this section we prove the restriction estimate, Lemma 6.5.4. We follow Bourgain's two step procedure [7]. Let $2m$ be the greatest even integer strictly below p.

Lemma D.13. *If $\phi : \mathbb{Z} \to \mathbb{C}$ with $|\phi| \leqslant \nu$ then*

$$\int_{\mathbb{T}} |\widehat{\phi}(\alpha)|^{2m} \mathrm{d}\alpha \ll_\epsilon X^{2m-1+\epsilon}.$$

Proof. By orthogonality and the triangle inequality

$$\int_{\mathbb{T}} |\widehat{\phi}(\alpha)|^{2m} \mathrm{d}\alpha \ll (WY)^{2m(k-1)}\mathcal{N},$$

where $\mathcal{N}$ is the number of solutions $(\mathbf{x}, \mathbf{y}) \in [Y]^m \times [Y]^m$ to the Diophantine equation

$$\sum_{i \leqslant m} (Wx_i + \xi)^k = \sum_{i \leqslant m} (Wy_i + \xi)^k.$$

By Lemma D.3, we now have

$$\begin{aligned} \int_{\mathbb{T}} |\widehat{\phi}(\alpha)|^{2m} \mathrm{d}\alpha &\ll (WY)^{2m(k-1)} (WY)^{2m-k} = (WY)^{k(2m-1)} \\ &\ll (WX)^{2m-1} \ll X^{2m-1+\epsilon}. \end{aligned}$$

□

In this subsection only, we denote by δ an arbitrary parameter in the range

$$0 < \delta < 1,$$

for consistency with existing literature. Consider the large spectra

$$\mathcal{R}_\delta = \{\alpha \in \mathbb{T} : |\widehat{\phi}(\alpha)| > \delta X\},$$

and note from (6.23) that

$$\|\widehat{\phi}\|_\infty \leqslant \|\phi\|_1 \leqslant \|\nu\|_1 \ll X.$$

As in [11, §6], it suffices to prove that if $\epsilon_0 > 0$ then

$$\mathrm{meas}(\mathcal{R}_\delta) \ll_{\epsilon_0} \frac{1}{\delta^{2m+\epsilon_0} X} \tag{D.10}$$

and, further, we may assume that

$$X^{-\epsilon} < \delta < 1. \tag{D.11}$$

Let $\theta_1, \ldots, \theta_R$ be X^{-1}-spaced points in $\mathcal{R}_\delta$. As $m \geqslant k$, it remains to show that

$$R \ll_{\epsilon_0} \delta^{-2k-\epsilon_0}. \tag{D.12}$$

Put $\gamma = k + \epsilon_0/3$. by the calculation in [11, §6], we have

$$\delta^{2\gamma} X^\gamma R^2 \ll \sum_{1 \leqslant r, r' \leqslant R} |\widehat{\nu}(\theta_r - \theta'_r)|^\gamma. \tag{D.13}$$

Consider

$$\theta = \theta_r - \theta'_r$$

in the summand on the right hand side of (D.13). by Lemma D.10, the contribution from $\theta \in \mathfrak{m}$ to the right hand side of (D.13) is $O(R^2(XY^{-\kappa 2^{-k}})^\gamma)$. by (D.11), this is $o(\delta^{2\gamma} X^\gamma R^2)$. Hence

$$\delta^{2\gamma} X^\gamma R^2 \ll \sum_{\substack{1 \leqslant r, r' \leqslant R: \\ \theta = \theta_r - \theta'_r \in \mathfrak{M}}} |\widehat{\nu}(\theta)|^\gamma. \tag{D.14}$$

If $\theta \in \mathfrak{M}(q,a)$ with $(a,q) = 1$ and $q \leqslant Q$ then, by Lemmas D.9, D.11 and D.12 we have

$$\begin{aligned} \widehat{\nu}(\theta) &\ll (XW)^{\frac{k-1}{k}}(q + X\|q\alpha\|) + q^{-1/k} \min\{X, \|\alpha - a/q\|^{-1}\} \\ &\ll X^{1-\frac{1}{2k}} + q^{-1/k} \frac{X}{1 + X\|\alpha - a/q\|}. \end{aligned}$$

With C a large positive constant, the contribution to the right hand side of (D.14) from denominators $q > Q_1 := C + \delta^{-3k}$ is therefore bounded, up to a constant, by

$$R^2 X^\gamma (Q_1^{-\gamma/k} + X^{-\frac{\gamma}{2k}})$$

which, by (D.11), is negligible compared to the left hand side of (D.14). Therefore

$$\delta^{2\gamma} R^2 \ll \sum_{1 \leqslant r, r' \leqslant R} G(\theta_r - \theta'_r), \tag{D.15}$$

where

$$G(\alpha) = \sum_{q \leqslant Q_1} \sum_{a=0}^{q-1} \frac{q^{-\gamma/k}}{(1 + X|\sin(\alpha - a/q)|)^\gamma}.$$

The inequality (D.15) is very similar to [7, Eq. (4.16)], but with N^2 replaced by X, and with $Q_1 \sim \delta^{-3k}$ rather than $Q_1 \sim \delta^{-5}$. The exponents differ but, since $\gamma > k$, Bourgain's argument carries through, and we obtain (D.12). This completes the proof of Lemma 6.5.4.

Bibliography

[1] T. M. Apostol. *Introduction to analytic number theory*. Undergraduate Texts in Mathematics. Springer-Verlag, New York-Heidelberg, (1976).

[2] V. Bergelson. *Ergodic Ramsey theory—an update*. In: *Ergodic theory of* $\mathbf{Z}^d$ *actions (Warwick, 1993–1994)*. Vol. 228. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, (1996), pp. 1–61.

[3] V. Bergelson. *Mutually enriching connections between ergodic theory and combinatorics - lecture 7*. CIRM lecture available at https://bit.ly/2GNaL7d. (2016).

[4] V. Bergelson and D. Glasscock. *Interplay between notions of additive and multiplicative largeness*. arXiv:1610.09771. Preprint. (2016).

[5] V. Bergelson and A. Leibman. *Polynomial extensions of van der Waerden's and Szemerédi's theorems*. In: *J. Amer. Math. Soc.* 9.3 (1996), pp. 725–753.

[6] E. Bombieri, J.B. Friedlander, and H. Iwaniec. *Primes in arithmetic progressions to large moduli. II,* in: *Math. Ann.* 277.3 (1987), pp. 361–393.

[7] J. Bourgain. *On* $\Lambda(p)$*-subsets of squares*. In: *Israel J. Math.* 67.3 (1989), pp. 291–311.

[8] J. Bourgain, C. Demeter, and L. Guth. *Proof of the main conjecture in Vinogradov's mean value theorem for degrees higher than three*. In: *Ann. of Math. (2)* 184.2 (2016), pp. 633–682.

[9] M. Bright. *The Brauer-Manin obstruction on a general diagonal quartic surface*. In: *Acta Arith.* 147.3 (2011), pp. 291–302.

[10] M. Bright and I. Kok. *Failure of strong approximation on an affine cone*. arxiv:1707.04177. Preprint. (2017).

[11] T. D. Browning and S. M. Prendiville. *A transference approach to a Roth-type theorem in the squares*. In: *Int. Math. Res. Not. IMRN* 7 (2017), pp. 2219–2248.

[12] J. Chapman. *Partition regularity and multiplicatively syndetic sets*. arxiv:1902.01149. Preprint. (2019).

[13] S. Chow. *Roth-Waring-Goldbach*. In: *Int. Math. Res. Not. IMRN* 8 (2018), pp. 2341–2374.

[14] S. Chow, S. Lindqvist, and S. Prendiville. *Rado's criterion over squares and higher powers*. arXiv:1806.05002. Preprint. (2018).

[15] P. Csikvári, K. Gyarmati, and A. Sárközy. *Density and ramsey type results on algebraic equations with restricted solution sets*. In: *Combinatorica* 32 (2012), pp. 425–449.

[16] K. Cwalina and T. Schoen. *Tight bounds on additive Ramsey-type numbers*. In: *preprint* (2015).

[17] H. Davenport. *Analytic methods for diophantine equations and diophantine inequalities*. 2nd edition. Cambridge university press, (2005).

[18] H. Davenport. *Multiplicative Number Theory*. 3rd edition. Springer, (2000).

[19] M. Di Nasso and L. Luperi Baglini. *Ramsey properties of nonlinear Diophantine equations*. In: *Adv. Math.* 324 (2018), pp. 84–117.

[20] L. E. Dickson. *On the congrence* $x^n + y^n + z^n = 0 (\mathrm{mod}\ p)$. In: *Journal für die reine und angewandte Mathematk* 135 (1909), pp. 134–141.

[21] W. Duke, J. Friedlander, and H. Iwaniec. *Bounds for automorphic L-function*. In: *Invent. Math.* 112 (1993), pp. 1–8.

[22] S. Eberhard. *The abelian arithmetic regularity lemma*. arXiv:1606.09303. Preprint. (2016).

[23] S. Eberhard, B. Green, and F. Manners. *Sets of integers with no large sum-free subset*. In: *Ann. of Math.* 180 (2014), pp. 621–652.

[24] P. Frankl, R. L. Graham, and V. Rödl. *Quantitative theorems for regular systems of equations*. In: *J. Combin. Theory Ser. A* 47.2 (1988), pp. 246–261.

[25] H. Furstenberg. *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*. In: *J. Analyse Math.* 31 (1977), pp. 204–256.

[26] J. R. Getz. *Secondary terms in asymptotics for the number of zeros of quadratic forms over number fields*. In: *J. Lond. Math. Soc. (2)* 98.2 (2018), pp. 275–305.

[27] R. Graham. *Old and new problems and results in Ramsey theory*. In: *Horizons of combinatorics*. Vol. 17. Bolyai Soc. Math. Stud. Springer, Berlin, (2008), pp. 105–118.

[28] R. Graham. *Some of my favorite problems in Ramsey theory*. In: *Combinatorial number theory*. de Gruyter, Berlin, (2007), pp. 229–236.

[29] B. Green. *A Szemerédi-type regularity lemma in abelian groups, with applications*. In: *Geom. funct. anal.* 15 (2005), pp. 340–376.

[30] B. Green. *Roth's theorem in the primes*. In: *Ann. of Math. (2)* 161.3 (2005), pp. 1609–1636.

[31] B. Green and S. Lindqvist. *Monochromatic solutions to* $x + y = z^2$. In: *Can. J. Math.* (2017).

[32] B. Green and T. Sanders. *Monochromatic sums and products*. In: *Discrete Analysis* 5 (2016), pp. 1–43.

[33] B. Green and T. Tao. *An arithmetic regularity lemma, an associated counting lemma, and applications*. In: *An irregular mind*. Vol. 21. Bolyai Soc. Math. Stud. János Bolyai Math. Soc., Budapest, (2010), pp. 261–334.

[34] B. Green and T. Tao. *Linear equations in primes*. In: *Ann. of Math. (2)* 171.3 (2010), pp. 1753–1850.

[35] B. Green and T. Tao. *The quantitative behaviour of polynomial orbits on nilmanifolds*. In: *Ann. of Math.* 175.2 (2012), pp. 465–540.

[36] G. Hardy and E. Wright. *An Introduction to the Theory of Numbers*. 6th edition. Oxford University Press, (2008).

[37] R. Heath-Brown. *A New Form of the Circle Method, and its Application to Quadratic Forms*. In: *J. Reine Angew. Math.* 481 (1996).

[38] M. J. H. Heule, O. Kullmann, and V. W. Marek. *Solving and verifying the Boolean Pythagorean triples problem via cube-and-conquer*. In: *Theory and applications of satisfiability testing—SAT 2016*. Vol. 9710. Lecture Notes in Comput. Sci. Springer, [Cham], (2016), pp. 228–245.

[39] N. Hindman. *Partitions and sums of integers with repetition*. In: *J. Combin. Theory Ser. A* 27.1 (1979), pp. 19–32.

[40] N. Hindman, I. Leader, and D. Strauss. *Open problems in partition regularity*. In: *Combin. Probab. Comput.* 12.5-6 (2003). Special issue on Ramsey theory, pp. 571–583.

[41] Y. Katznelson. *An introduction to Harmonic analysis*. 2nd edition. Dover, (1976).

[42] A. Khalfalah and E. Szemerédi. *On the Number of Monochromatic Solutions of $x + y = z^2$*. In: *Combinatorics, Probability and Computing* 15 (2006), pp. 213–227.

[43] H. D. Kloosterman. *On the representation of numbers in the form $ax^2 + by^2 + cz^2 + dt^2$*. In: *Acta Math.* 49.3-4 (1927), pp. 407–464.

[44] J. C. Lagarias, A. M. Odlyzko, and J. B. Shearer. *On the density of sequences of integers the sum of no two of which is a square. I. Arithmetic progressions*. In: *J. Combin. Theory Ser. A* 33.2 (1982), pp. 167–185.

[45] J. C. Lagarias, A. M. Odlyzko, and J. B. Shearer. *On the density of sequences of integers the sum of no two of which is a square. II. General sequences*. In: *J. Combin. Theory Ser. A* 34.2 (1983), pp. 123–139.

[46] E. Lamb. *Maths proof smashes size record*. In: *Nature* 534 (2016), pp. 17–18.

[47] H. Lefmann. *On partition regular systems of equations*. In: *J. Combin. Theory Ser. A* 58.1 (1991), pp. 35–53.

[48] S. Lindqvist. *Partition regularity of generalised Fermat equations*. In: *Combinatorica* (2017).

[49] H. Liu, P. Pach, and C. Sándor. *Polynomial Schur's theorem*. arxiv:1811.05200. Preprint. (2018).

[50] J. P. Massias. *Sur les suites dont les sommes des termes 2 à 2 ne sont pas des carrés*. Published by département de mathématiques de Limoges. (1982).

[51] H. L. Montgomery. *Ten lectures on the interface between analytic number theory and harmonic analysis*. Vol. 84. CBMS Regional Conference Series in Mathematics. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, (1994), pp. xiv+220.

[52] J. Moreira. *Monochromatic sums and products in* $\mathbb{N}$. In: *Ann. of Math. (2)* 185.3 (2017), pp. 1069–1090.

[53] P. Pach. *Monochromatic solutions to* $x + y = z^2$ *in the interval* $[N, cN^4]$. arxiv:1805.06279. Preprint. (2018).

[54] S. Prendiville. Private communication. (2018).

[55] S. Prendiville. *Four variants of the Fourier-analytic transference principle*. In: *Online J. Anal. Comb.* 12 (2017), p. 25.

[56] R. Rado. *Studien zur Kombinatorik*. In: *Math. Z.* 36.1 (1933), pp. 424–470.

[57] K. F. Roth. *On certain sets of integers*. In: *J. London Math. Soc.* 28 (1953), pp. 104–109.

[58] A. Sárkőzy. *On difference sets of sequences of integers. I*. In: *Acta Math. Acad. Sci. Hungar.* 31.1–2 (1978), pp. 125–149.

[59] J. Schur. *Über die Kongruenz* $x^m + y^m \equiv z^m \pmod{p}$. In: *Jahresber. Deutschen Math. Verein.* 25 (1916), pp. 114–117.

[60] E. Szemerédi. *On sets of integers containing no k elements in arithmetic progression*. In: *Acta Arith.* 27 (1975). Collection of articles in memory of Juriĭ Vladimirovič Linnik, pp. 199–245.

[61] T. Tao. *Higher order Fourier analysis*. Vol. 142. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, (2012), pp. x+187.

[62] T. Tao and V. Vu. *Additive Combinatorics*. Cambridge University Press, (2010).

[63] E. Titshmarch. *The Theory of the Riemann Zeta-function*. 2nd edition. Oxford University Press, (1987).

[64] P. Varnavides. *On certain sets of positive density*. In: *J. London Math. Soc.* 34 (1959), pp. 358–360.

[65] R. C. Vaughan. *On Waring's problem for cubes*. In: *J. Reine Angew. Math.* 365 (1986), pp. 122–170.

[66] R. C. Vaughan. *The Hardy-Littlewood method*. Second. Vol. 125. Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge, (1997), pp. xiv+232.

[67] R. C. Vaughan and T. D. Wooley. *Waring's problem: a survey*. In: *Number theory for the millennium, III (Urbana, IL, 2000)*. A K Peters, Natick, MA, (2002), pp. 301–340.

[68] B. L. van der Waerden. *Beweis einer Baudetschen Vermutung*. In: *Nieuw Arch. Wisk.* 15 (1927), pp. 212–216.

[69] T. D. Wooley. *Large improvements in Waring's problem*. In: *Ann. of Math. (2)* 135.1 (1992), pp. 131–164.

[70] T. D. Wooley. *On Diophantine inequalities: Freeman's asymptotic formulae*. In: *Proceedings of the Session in Analytic Number Theory and Diophantine Equations*. Vol. 360. Bonner Math. Schriften. Univ. Bonn, Bonn, (2003), p. 32.

[71] T. D. Wooley. *The asymptotic formula in Waring's problem*. In: *Int. Math. Res. Not. IMRN* 7 (2012), pp. 1485–1504.