TRANSLATION INVARIANT QUADRATIC FORMS IN DENSE SETS

EUGEN KEIL

ABSTRACT. We generalize Roth's theorem on three term arithmetic progressions to translation invariant quadratic forms in at least 17 variables. We use Fourier-analysis, restriction theory, uniformity norms and Roth's density increment method to show quantitative estimates for subsets of the integers without any non-trivial solutions.

1. INTRODUCTION

In 1953 Roth [18] proved his theorem on 3-term arithmetic progressions in dense sets. It states that a subset $\mathcal{A} \subset \{1, 2, \dots, N\}$ with no arithmetic progressions of the form $x, x+h, x+2h$ with $h \geq 1$ cannot be too large. His theorem gives the bound $|\mathcal{A}| \leq CN(\log\log N)^{-1}$ for some constant $C \geq 1$. In other words, it is not possible to avoid 3-term arithmetic progressions as long as the density of the set $\mathcal{A}$ is big enough.

Arithmetic progressions can also be described as solutions to translation invariant equations (see explanation at the end of the introduction). In the case of 3-term progressions we have the equation $x_1 - 2x_2 + x_3 = 0$. Roth [19] went on to prove a version of his theorem for solutions to translation invariant linear systems in k equations with at least $2k+1$ variables. By recent work of Gowers [8] we can now solve translation invariant systems with as few as $k+2$ variables in sets $\mathcal{A}$ of cardinality at least $C_k N(\log\log N)^{-c_k}$ for some $C_k, c_k > 0$.

The aim of this work is to combine the ideas of Gowers [8], Green [9], Liu [16], Roth [18] and the previous work of the author [15] to give a version of Roth's theorem for quadratic forms.

Theorem 1.1. *Let $\mathbf{x}^T Q \mathbf{x} = 0$ be a translation invariant quadratic equation in $s \geq 17$ variables. Assume that it has a non-singular real solution, but only trivial solutions when the variables are restricted to $\mathcal{A} \subset \{1, 2, \dots, N\}$. Then $|\mathcal{A}| \leq C_Q N(\log\log N)^{-c}$ for some $c, C_Q > 0$.*

Theorem 1.1 will follow from the more precise Theorem 2.2. In most cases we only need $s \geq 10$ variables as in the work of Liu [16]. The bound $s \geq 17$ is a worst case scenario and can certainly be improved by a more complicated analysis.

Date: August 29, 2013.

2010 *Mathematics Subject Classification.* Primary 11B30; Secondary 11P55, 11D09.

Key words and phrases. dense set and circle method and quadratic form and translation invariant.

The next observation is that the exponent c in $C_Q N(\log\log N)^{-c}$ is independent of the quadratic form Q and the number of variables s (we use the letter Q interchangeably for the quadratic form and the underlying matrix). If one would allow such a dependence, it is possible to derive the above theorem (even for all $s \geq 5$) by the methods of Gowers [8] as follows. Take any integer $\mathbf{y}$ with $Q(\mathbf{y}) = 0$ (see Lemma 2.3) and consider the patterns $(a + qy_1, \ldots, a + qy_s)$ for $a \in \mathbb{Z}$ and $q \in \mathbb{N}$. Then the approach in [8] shows that a set that doesn't contain any of those patterns, must have a density bounded by $C_Q N(\log\log N)^{-c_s}$.

We want to point out that it is very likely that the methods of this work can be adapted to give an asymptotic count of the number of solutions to $Q(\mathbf{x}) = 0$ for $x_i \in \mathcal{B}$ for some relatively structured set $\mathcal{B}$ such as the primes, for example. This is not possible by relying on the work of [8] or other purely additive combinatorial results from the linear theory. This explains some of the motivation behind Theorem 1.1.

Recent years have seen huge advances in our understanding of linear equations in primes. Work of Green [9] and Green and Tao [11] introduced the concept of a 'pseudorandom measure', which led to amazing new developments in the linear theory [12]. These results can be used to find prime solutions for general diophantine equations, such as in recent work of Brüdern, Dietmann, Liu and Wooley [2] on the Birch-Goldbach problem.

If one is interested in asymptotics, on the other hand, one has to deal with the non-linear theory directly. Recent work on prime solutions for quadratic forms by Liu [16] uses a variant of the circle method to deal with a large class of quadratic forms in at least ten variables and provides one of the main ideas for this work.

Previous work on diagonal translation invariant forms was carried out by Smith [21], who considers the system

$$\begin{aligned} \lambda_1 x_1^2 + \lambda_2 x_2^2 + \ldots + \lambda_s x_s^2 &= 0, \\ \lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_s x_s &= 0 \end{aligned}$$

with $\lambda_1 + \ldots + \lambda_s = 0$ in $s \geq 9$ variables. The author simplified Smith's approach in [15] and reduced the number of variables down to $s \geq 7$. The methods of [15] play a significant role in the development of this work and we cite several results from [15] to simplify the exposition here. Readers interested in the restriction theory part of the argument are adviced to have a look at [15] for more explanations.

Before proceeding to the main part of the paper, we want to give the reader the opportunity to gain some intuition about the consequences of assuming 'translation invariance' in the context of quadratic forms. For linear equations, this geometric condition translates into the arithmetic statement that the sum of the coefficients in each equation is zero. This is also true for the diagonal quadratic system considered above. For a quadratic form $Q(\mathbf{x}) := \mathbf{x}^T Q \mathbf{x} = 0$ with a symmetric matrix $Q \in \mathbb{Z}^{s\times s}$ *translation invariance* means that $Q(\mathbf{x} + h\mathbf{1}) = Q(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{Z}^s$ and $h \in \mathbb{Z}$, where $\mathbf{1} = (1, \ldots, 1)^T \in \mathbb{Z}^s$.

This implies that $Q(h\mathbf{1}) = Q(\mathbf{0}) = 0$. We call the multiples of $\mathbf{1}$ the *trivial solutions* to our quadric. If we expand $Q(\mathbf{x} + h\mathbf{1})$, we obtain

$$Q(\mathbf{x}) = Q(\mathbf{x} + h\mathbf{1}) = Q(\mathbf{x}) + 2h\mathbf{x}^T Q\mathbf{1} + h^2 Q(\mathbf{1}).$$

It follows that $Q\mathbf{1} = \mathbf{0}$ and it is easy to check that it is a sufficient condition as well.

Another way of looking at this issue is to set $h = -x_s$. Then we get $Q(\mathbf{x} - x_s\mathbf{1}) = Q(\mathbf{x})$ and, therefore, any translation invariant quadratic form can be written in the form $Q'(x_1 - x_s, \dots, x_{s-1} - x_s)$ for some arbitrary quadratic form Q' in $s - 1$ variables.

To prove Theorem 1.1, it is not always necessary to assume translation invariance, as can be seen from considering only the first equation from the diagonal quadratic system above, but the condition $Q(\mathbf{1}) = 0$ is clearly necessary. Otherwise, we can choose $\mathcal{A}$ as the set of numbers congruent to one modulo n, where n is some large number (dependent only on the coefficients of Q) and obtain a contradiction.

Acknowledgements:
This work is part of the author's Ph.D. thesis and he would like to thank his supervisor Trevor Wooley for proposing the topic and constant encouragement. The doctoral studies of the author were partially supported by the EPSRC.

2. Notation and General Discussion

First we remind the reader about some standard notation. We write $e(x) = \exp(2\pi i x)$ and use $f = O(g)$ to express that $|f| \leq Cg$ for some constant $C > 0$ and similarly Vinogradov's notation $f \ll g$. We indicate dependencies on parameters by subscripts as in $O_p(N^s)$ or $\ll_{P,\epsilon}$, for example. The parameter $N \in \mathbb{N}$, governing the size of the variables x_i should be thought of as large and we write $[1, N]$ as abbreviation for the interval $\{1, 2, \dots, N\}$. The set $\mathcal{A}$ is always a subset of $[1, N]$ with density $\delta = |\mathcal{A}|/N$ and indicator function $1_{\mathcal{A}}$. The balanced function $f(x) = 1_{\mathcal{A}}(x) - \delta$ plays an important role at various places in this paper.

Bold face letters such as $\mathbf{x}$ denote vectors with components x_i and inequalities such as $\mathbf{x} \leq N$ or $\mathbf{x} \leq \mathbf{y}$ should be understood componentwise. A sum over natural numbers starts at one, if not otherwise indicated. The symbol $\mathbb{T}$ is used to refer to the 'circle' $\mathbb{R}/\mathbb{Z}$ with the circle norm $\|\alpha\| := \min\{|\alpha - z| : z \in \mathbb{Z}\}$, the distance of $\alpha \in \mathbb{R}$ to the nearest integer.

We don't want to distinguish between quadratic forms that are related by a simple renaming of variables. Given two matrices A and B in $\mathbb{R}^{s \times s}$ we say they are *permutation-equivalent* if

$$A = P^T B P$$

for an invertable matrix $P \in \{0, 1\}^{s \times s}$.

To explain the general structure of the work and the main theorem, we consider the following property of quadratic forms.

Definition 2.1 (Off-diagonal rank)**.** For a symmetric matrix $Q \in \mathbb{R}^{s\times s}$ we consider matrices M that are permutation equivalent to Q and write them in the form

$$M = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$$

for some matrices A, B and C. Then the *off-rank* r of Q is defined as

$$r = \max\{\operatorname{rank}(B) : M \text{ is permutation equivalent to } Q\}.$$

In other words, this is the maximal rank of a submatrix in Q, that doesn't contain any diagonal elements.

The off-rank r of a matrix determines the treatment of the corresponding quadratic equation. While for $r \geq 5$ we can apply the bilinear sum method inspired by the work of Liu [16] on prime solutions for quadratic forms, we need a more complicated approach for $r \leq 4$, based on 'partial diagonalisation'. This leads to the following main result of this work.

Theorem 2.2. *Let $Q \in \mathbb{Z}^{s\times s}$ be symmetric with $Q \cdot \mathbf{1} = \mathbf{0}$ and off-rank r. Assume that $Q(\mathbf{x}) = 0$ has a non-singular real solution and assume that $s \geq 5+3r$ for $1 \leq r \leq 4$ and $s \geq 10$ for $r \geq 5$. If there are only trivial solutions, when the variables are restricted to $\mathcal{A} \subset [1, N]$, then $|\mathcal{A}| \ll_Q N(\log\log N)^{-c}$ for some $c > 0$ independent of Q.*

Remark. Almost all quadratic forms in at least 10 variables have off-rank at least five. The exceptional cases lie in a lower-dimensional submanifold, meaning that we need only $s \geq 10$ for almost all quadratic forms.

Apart from the off-rank problem, there is the general issue of positive definiteness that needs to be addressed in the case of quadratic forms. We saw in the introduction that any translation invariant quadratic form can be written in the form $Q(\mathbf{x}) = Q(\mathbf{x} - x_s\mathbf{1}) = Q'(x_1 - x_s, \ldots, x_{s-1} - x_s)$, where the matrix for the quadratic form Q' is given by the upper left submatrix of size $(s-1) \times (s-1)$ in Q. Now we can diagonalize $Q'(z_1, \ldots, z_{s-1}) = 0$ over $\mathbb{Z}$, where $z_i = x_i - x_s$. This can be done by completing the square successively and then multiplying by a suitable integer to ensure that the rational coefficients that appear during this process become integers again. We get for some $\lambda_j \in \mathbb{Z}$ an equation of the form

$$\lambda_1 y_1^2 + \ldots + \lambda_{s-1} y_{s-1}^2 = 0, \tag{2.1}$$

where the y_i are independent linear forms in the variables z_i or equivalently, translation invariant linear forms in the variables $x_1, \ldots, x_s$. If we have $\lambda_i \geq 0$ for all $1 \leq i \leq s$, then all real solutions to this quadric are singular and this case is excluded by the assumptions in Theorem 2.2. This is also true for the case when $\lambda_i \leq 0$ for all $1 \leq i \leq s$.

In the remaining cases we have at least one negative and at least one positive coefficient λ_i. The existence of coefficients of both signs is equivalent to the existence of a non-singular real solution. If the number t of non-zero coefficients λ_i is less than five, we can get problems with p-adic solubility as in the example

$$y_1^2 + y_2^2 - 3(y_3^2 + y_4^2) = 0,$$

where we have only the zero solutions modulo eight. In this case we consider the linear system $y_k = 0$ for $1 \leq k \leq t \leq 4$ instead. This case is covered in Section 8.

For most quadratic forms we have $t \geq 5$ and have the following Lemma.

Lemma 2.3. *If $s \geq 5$, $d_i \neq 0$ for $1 \leq i \leq s$ and not all coefficients d_i in*

$$d_1 x_1^2 + \ldots + d_s x_s^2 = 0, \tag{2.2}$$

have the same sign, then (2.2) *has $C(\mathbf{d})N^{s-2} + o(N^{s-2})$ solutions with $x_i \in [1, N]$ for some constant $C(\mathbf{d}) > 0$ dependent on the coefficients d_i.*

Proof. The proof is essentially given in Chapter 8 of [5]. Chapter 2 of [23] also contains all the necessary estimates to deduce the result. Another approach can be found in the recent paper [14]. □

The last ingredient for the proof of Theorem 2.2 are L^p-properties for certain exponential sums. For a function $g : \mathbb{N} \to \mathbb{C}$ and $\mathcal{A} \subset [1, N]$ we define

$$L_g(\alpha) = \sum_{z \leq N} g(z) e(\alpha z) \quad \text{and} \quad V_g(\alpha, \beta) = \sum_{x \leq N} g(x) e(\alpha x^2 + \beta x) \tag{2.3}$$

and write $L_{\mathcal{A}}(\alpha)$ or $V_{\mathcal{A}}(\alpha, \beta)$ in the case $g = 1_{\mathcal{A}}$ and $L(\alpha), V(\alpha, \beta)$ for the sums without any weight g. We use Appendix C to derive two useful L^p estimates along the lines of [15].

The general structure of the paper is as follows. In Section 3 we treat the case $r \geq 5$ with a refinement of Liu's method [16]. Appendix A provides the new necessary ingredient, a sharp 'Vinogradov lemma'. In Sections 4 to 6 we consider the non-degenerate part of the case $r \leq 4$ and use 'convexity' methods to simplify our mean-value integrals to deduce a correlation estimate for the exponential sums in (2.3) with $g = f$. In Section 7, we use the correlation estimates to prove Theorem 2.2. Section 8 finally deals with the degenerate cases, where we can extract a linear subsystem from Q, which can be treated by Gowers' theory [8]. Appendix B provides a short proof for the uniformity norm estimate for completeness.

3. The Bilinear Sum Method

The main goal of this section is to deduce correlation estimate (3.7) in the case $r \geq 5$. We follow Liu [16] and simplify his approach by removing the 'geometry of numbers' argument. For a quadratic form $Q(\mathbf{x}) = \mathbf{x}^T Q \mathbf{x}$ define the exponential sum

$$S_g(\alpha) = \sum_{\mathbf{x} \leq N} g(\mathbf{x}) e(\alpha Q(\mathbf{x})),$$

where $g(\mathbf{x}) = \prod_{i=1}^{s} g_i(x_i)$ and $|g_i| \leq 1$. The main technical result in this section is an L^p-estimate for this quadratic exponential sum.

Theorem 3.1. *Let Q have off-rank $r \geq 1$. Then for $p > 4/r$ we have*

$$\int_0^1 |S_g(\alpha)|^p \, d\alpha \ll_p N^{ps-2}.$$

Assume the L^1-bound $\sum_{x \leq N} |g_i(x)| \leq 2\delta N$ and $r \geq 5$. Then $p > 4/5$ implies

$$\int_0^1 |S_g(\alpha)|^p \, d\alpha \ll_{p,s} \delta^{(s-10)p} N^{ps-2}.$$

Proof. The matrix Q is permutation equivalent to a matrix of the form

$$\begin{pmatrix} A & R \\ R^T & B \end{pmatrix}$$

with $\operatorname{rank}(R) = r$. To simplify notation, we can also assume that the first r rows of B are linearly independent. Decompose the variables $\mathbf{x} = (\mathbf{x}_a, \mathbf{x}_b)$ accordingly into two blocks of sizes a and b with at least r variables each. Then the quadratic form $Q(\mathbf{x})$ has the representation

$$Q(\mathbf{x}) = \mathbf{x}_a^T A \mathbf{x}_a + 2\mathbf{x}_a^T R \mathbf{x}_b + \mathbf{x}_b^T B \mathbf{x}_b,$$

Write $g(\mathbf{x}_a) = \prod_{i=1}^a g_i(x_i)$ and $g(\mathbf{x}_b) = \prod_{i=a+1}^s g_i(x_i)$. Then we have the estimate

$$\begin{aligned} |S_g(\alpha)| =& \Big| \sum_{\mathbf{x}_a \leq N} \sum_{\mathbf{x}_b \leq N} g(\mathbf{x}_a) g(\mathbf{x}_b) e(\alpha(\mathbf{x}_a^T A \mathbf{x}_a + 2\mathbf{x}_a^T R \mathbf{x}_b + \mathbf{x}_b B \mathbf{x}_b)) \Big| \\ \leq& \sum_{\mathbf{x}_a \leq N} \Big| \sum_{\mathbf{x}_b \leq N} g(\mathbf{x}_b) e(\alpha(2\mathbf{x}_a^T R \mathbf{x}_b + \mathbf{x}_b B \mathbf{x}_b)) \Big| \\ \leq& N^{a/2} \Big(\sum_{\mathbf{x}_a \leq N} \Big| \sum_{\mathbf{x}_b \leq N} g(\mathbf{x}_b) e(\alpha(2\mathbf{x}_a^T R \mathbf{x}_b + \mathbf{x}_b B \mathbf{x}_b)) \Big|^2 \Big)^{1/2} \end{aligned}$$

by the inequality of Cauchy-Schwarz. The expression in the parentheses on the right hand side is

$$\begin{aligned} & \sum_{\mathbf{x}_b \leq N} \sum_{\mathbf{x}'_b \leq N} g(\mathbf{x}_b) \overline{g(\mathbf{x}'_b)} \sum_{\mathbf{x}_a \leq N} e(\alpha(2\mathbf{x}_a^T R(\mathbf{x}_b - \mathbf{x}'_b) + \mathbf{x}_b B \mathbf{x}_b - \mathbf{x}'_b B \mathbf{x}'_b)) \\ \leq & \sum_{\mathbf{x}_b \leq N} \sum_{\mathbf{x}'_b \leq N} \Big| \sum_{\mathbf{x}_a \leq N} e(2\alpha \mathbf{x}_a^T R(\mathbf{x}_b - \mathbf{x}'_b)) \Big| \\ \leq & \sum_{\mathbf{x}_b \leq N} \sum_{\mathbf{x}'_b \leq N} \prod_{i=1}^a \min\left(N, \|2\alpha R_i(\mathbf{x}_b - \mathbf{x}'_b)\|^{-1}\right) \end{aligned}$$

where R_i is the ith row of R.

Consider the equations $y_i = 2R_i(\mathbf{x}_b - \mathbf{x}'_b)$. The variables y_i can vary over an interval $[-PN, PN]$ for some constant P depending on the size of the coefficients of R. Since R has rank r, the system of equations $y_i = 2R_i(\mathbf{x}_b - \mathbf{x}'_b)$ has $O(N^{2b-r})$ solutions for given $|y_1|, \ldots, |y_r| \leq PN$. We bound the other $a - r$ factors $(i > r)$ trivially by N and obtain the bound

$$|S_g(\alpha)|^2 \ll N^{2b+2a-2r} \sum_{|\mathbf{y}| \leq M} \prod_{i=1}^r \min\left(N, \|\alpha y_i\|^{-1}\right).$$

Since $a+b=s$ and the inner expression factors into r independent sums, we can apply Lemma A.1. Write $\alpha = a/q+\beta$ with $|\beta| \le 1/(qN)$ for some $q \le N$ by Dirichlet's approximation theorem and define

$$K(\alpha) = \Big(N \log q + \min\Big\{\frac{N^2}{q}, \frac{|\log(|\beta|N^2)|+1}{|\beta| q}\Big\}\Big)^{1/2}. \tag{3.1}$$

If the representation of α turns out to be non-unique, we take the minimal value attained by the various functions on the right hand side of (3.1). With this definition, we get from Lemma A.1 the estimate

$$|S_g(\alpha)| \ll N^{s-r} K(\alpha)^r. \tag{3.2}$$

To obtain the first L^p-bound in Theorem 3.1 we prove the following useful lemma.

Lemma 3.2. *For $p>4$ we have $\int_0^1 |K(\alpha)|^p \, d\alpha \ll_p N^{p-2}$.*

Proof. We decompose $[0,1]$ according to the Dirichlet approximations, which give us

$$\int_0^1 |K(\alpha)|^p \, d\alpha \ll \sum_{q\le N} \sum_{a=1}^{q} \int_{|\beta| \le 1/(qN)} |K(a/q+\beta)|^p \, d\beta,$$

where the summation in a is only over the elements with $(a;q)=1$. We decompose the integration over β further into sets with $|\beta| \le N^{-2}$ and a dyadic decomposition $|\beta| \in (2^i N^{-2}, 2^{i+1}N^{-2}]$ for $0 \le i \le \log_2(Nq^{-1})$. For $|\beta| \le N^{-2}$ we get by (3.1) the contribution

$$\sum_{q\le N} \sum_{a=1}^{q} N^{-2}\Big(N\log q + N^2 q^{-1}\Big)^{p/2} \ll N^{p-2}.$$

On each dyadic part $|\beta| \in (2^i N^{-2}, 2^{i+1}N^{-2}]$, we obtain the contribution

$$2^i N^{-2}\Big(N \log q + N^2(i+1)2^{-i}q^{-1}\Big)^{p/2}.$$

Apply the bound $|x+y|^{p/2} \ll |x|^{p/2} + |y|^{p/2}$ and sum over i to arrive at

$$\sum_{q\le N} \sum_{a=1}^{q} \Big((Nq)^{-1}(N\log q)^{p/2} + N^{-2}(N^2/q)^{p/2}\Big) \ll N^{p-2}.$$

□

Now the first part of Theorem 3.1 follows from (3.2). For the second part we have $r \ge 5$. By permutation equivalence we may assume, that

$$Q = \begin{pmatrix} A & R & X \\ R^T & B & Y \\ X^T & Y^T & C \end{pmatrix}, \tag{3.3}$$

where $R \in \mathbb{R}^{5\times 5}$ is a full rank matrix. We can divide the variable vector $\mathbf{x}$ into $(\mathbf{x}_a, \mathbf{x}_b, \mathbf{x}_c)$, where $\mathbf{x}_a$ and $\mathbf{x}_b$ contain five variables and $\mathbf{x}_c$ the remaining

$s-10$ variables. Then we have the bound

$$|S_g(\alpha)| \leq \sum_{\mathbf{x}_c \leq N} |g_c(\mathbf{x}_c)| \Big| \sum_{\mathbf{x}_a \leq N} \sum_{\mathbf{x}_b \leq N} g_a(\mathbf{x}_a) g_b(\mathbf{x}_b) e(\alpha Q(\mathbf{x})) \Big|, \tag{3.4}$$

where the functions g_a, g_b and g_c are defined in a similarlar way as above. If we expand the quadratic form $Q(\mathbf{x})$ by use of (3.3) and the decomposition $(\mathbf{x}_a, \mathbf{x}_b, \mathbf{x}_c)$ we get

$$\mathbf{x}^T Q \mathbf{x} = (\mathbf{x}_a^T, \mathbf{x}_b^T) \begin{pmatrix} A & R \\ R^T & B \end{pmatrix} \begin{pmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{pmatrix} + 2\mathbf{x}_a^T X \mathbf{x}_c + 2\mathbf{x}_b^T Y \mathbf{x}_c + \mathbf{x}_c^T C \mathbf{x}_c. \tag{3.5}$$

The last summand depends only on $\mathbf{x}_c$ and will disappear due to the absolute value signs in (3.4). The other two parts which contain $\mathbf{x}_c$ can be seen as a sum of linear forms $L_{i,\mathbf{x}_c}(x_i)$ for $i \leq 10$. Therefore, the absolute value of the inner sum in (3.4) can be seen as $|S_h(\alpha)|$ for a new function

$$h_{\mathbf{x}_c}(\mathbf{x}_a, \mathbf{x}_b) = \prod_{i=1}^{10} g_i(x_i) e(\alpha L_{i,\mathbf{x}_c}(x_i))$$

and the quadratic form corresponding to the 10×10 matrix in (3.5).

For $p > 4/5$ we use (3.2) and obtain the bound

$$\int_0^1 |S_g(\alpha)|^p \, d\alpha \leq \int_0^1 \Big| \sum_{\mathbf{x}_c \leq N} |g_c(\mathbf{x}_c)| |S_h(\alpha)| \Big|^p \, d\alpha$$

$$\ll \int_0^1 \Big| \sum_{\mathbf{x}_c \leq N} |g_c(\mathbf{x}_c)| N^5 K(\alpha)^5 \Big|^p \, d\alpha \ll \Big(\sum_{\mathbf{x}_c \leq N} |g_c(\mathbf{x}_c)| \Big)^p N^{10p-2}.$$

The result now follows from the assumption $\sum_{x \leq N} |g_i(x)| \leq 2\delta N$. □

Having proven Theorem 3.1, we can deduce a correlation estimate for the exponential sum $S_g(\alpha)$ in the cases $r \geq 5$. From the discussion of Section 2, the assumption in Theorem 2.2, we have by Lemma 2.3 the lower bound

$$\int_0^1 S(\alpha) \, d\alpha \gg N^{s-2},$$

where $S(\alpha)$ is the exponential sum with $g = 1$. On the other hand, if we write $S_{\mathcal{A}}(\alpha)$ for the exponential sum with the indicator function $g = 1_{\mathcal{A}}$ of the set $\mathcal{A}$, we obtain

$$\int_0^1 S_{\mathcal{A}}(\alpha) \, d\alpha = \delta N$$

since there are only trivial solutions in the set $\mathcal{A}$. By comparing those two quantities, we derive

$$\int_0^1 |S_g(\alpha)| \, d\alpha \gg \delta^s N^{s-2}$$

for $g(\mathbf{x}) = 1_{\mathcal{A}^s}(\mathbf{x}) - \delta^s$ with $\delta = |\mathcal{A}|/N$ as long as $N \gg_Q \delta^{-2}$, say. While this function g doesn't satisfy the conditions of Theorem 3.1, we can write it as a finite sum $g(\mathbf{x}) = \sum_{i=1}^{s} f_i(\mathbf{x})$ of functions

$$f_i(\mathbf{x}) = (1_{\mathcal{A}}(x_i) - \delta)\delta^{i-1} \prod_{j>i} 1_{\mathcal{A}}(x_j), \tag{3.6}$$

that factor into a product of factors h_j. Each of those satisfies the L^1-condition $\sum_{x\leq N} |h_j| \leq 2\delta N$. Therefore, by part two of Theorem 3.1 we have for some $i \leq s$ the estimate

$$\begin{aligned} \delta^s N^{s-2} &\ll \sup_{\alpha} |S_{f_i}(\alpha)|^{1-p} \int_0^1 |S_{f_i}(\alpha)|^p \, d\alpha \\ &\ll \sup_{\alpha} |S_{f_i}(\alpha)|^{1-p} (\delta N)^{(s-10)p} N^{10p-2}. \end{aligned}$$

Now set $p = 8/9 > 4/5$, for example, and deduce the correlation estimate

$$\sup_{\alpha} |S_{f_i}(\alpha)| \gg \delta^{s+80} N^s. \tag{3.7}$$

This correlation estimate is used in Section 7 to run the usual density increment method of Roth [18].

4. Partial diagonalisation

The correlation estimate for the case $r \leq 4$ requires more work and the problem splits into several subcases. In this section we take the first step and prove a structure result for quadratic forms of low off-rank to extract a partial diagonal structure.

By permutation equivalence, we may assume that

$$Q = \begin{pmatrix} A & R & M \\ R^T & B & N \\ M^T & N^T & C \end{pmatrix}, \tag{4.1}$$

where $R \in \mathbb{R}^{r\times r}$ is a full rank matrix. We want to show that we can 'diagonalize' C by adding at most r linear equations to $Q(\mathbf{x}) = 0$. Since the size of the matrix C in (4.1) is $s-2r$, we can hope that methods for diagonal quadrics can provide solutions, as long as s is not too small compared to r. We begin by stating a simple lemma.

Lemma 4.1. *Let $R \in \mathbb{R}^{r\times r}$ be a full rank matrix and $\mathbf{v}, \mathbf{w} \in \mathbb{R}^r$, then there is exactly one $c \in \mathbb{R}$ such that the matrix*

$$\begin{pmatrix} R & \mathbf{v} \\ \mathbf{w}^T & c \end{pmatrix}$$

has rank r, namely $c = \mathbf{w}^T R^{-1} \mathbf{v}$.

Proof. Multiplication by an invertable matrix on the left leads to the relation

$$\begin{pmatrix} R^{-1} & 0 \\ \mathbf{w}^T R^{-1} & -1 \end{pmatrix} \begin{pmatrix} R & \mathbf{v} \\ \mathbf{w}^T & c \end{pmatrix} = \begin{pmatrix} E_r & R^{-1}\mathbf{v} \\ 0 & \mathbf{w}^T R^{-1} \mathbf{v} - c \end{pmatrix},$$

where $E_r \in \mathbb{Z}^{r\times r}$ is the identity matrix. This implies the result. □

Now we apply Lemma 4.1 to 'diagonalize' C.

Lemma 4.2. *Let M, N, C, R be the matrices from* (4.1)*. Then*

$$C = N^T R^{-1} M + D,$$

where D is a diagonal matrix.

Proof. For a fixed element c_{ij} from C with $i \neq j$ consider the matrix

$$\begin{pmatrix} R & \mathbf{m}_j \\ \mathbf{n}_i^T & c_{ij} \end{pmatrix},$$

where $\mathbf{m}_j, \mathbf{n}_i^T$ are the jth column and ith row from the matrices M and N^T respectively. This is a submatrix of Q that lies completely off-diagonal and, therefore, cannot have rank more than r. By Lemma 4.1, we get $c_{ij} = \mathbf{n}_i^T R^{-1} \mathbf{m}_j$, which gives the desired claim. □

Lemma 4.2 says that there is a diagonal matrix D such that the rows of $C - D$ are linear combinations of rows of M. The next step is to find a common basis of linear forms, that span the rows of M, N and $C - D$.

Lemma 4.3. *Let M, N be as in* (4.1)*. There is a matrix $H \in \mathbb{Z}^{t \times (s-2r)}$ with $t \leq r$ and linearly independent rows such that $M = \hat{M}H$ and $N = \hat{N}H$ for some matrices $\hat{M}, \hat{N} \in \mathbb{Q}^{r \times t}$.*

Proof. The submatrix consisting of M and N, as in (4.1), namely

$$\begin{pmatrix} M \\ N \end{pmatrix},$$

has rank t with $t \leq r$, since Q has off-rank r. One can find t rows which span the rowspace of this matrix. Arrange these into a single matrix H. Then we can write

$$\begin{pmatrix} M \\ N \end{pmatrix} = \begin{pmatrix} \hat{M} \\ \hat{N} \end{pmatrix} \cdot H,$$

for some matrices $\hat{M}, \hat{N} \in \mathbb{Q}^{r \times t}$. □

Combining Lemma 4.2 and Lemma 4.3, we obtain

$$C = H^T \hat{N}^T R^{-1} \hat{M} H + D. \tag{4.2}$$

This insight is sufficient to diagonalize C. But there is another technical rank condition on H that we need for later parts of this chapter, which requires another step of linear algebra. Consider the following property for matrices.

AP: *If we remove any column from $H \in \mathbb{Z}^{t \times g}$, then the remaining matrix contains two disjoint $t \times t$ non-singular submatrices.*

This is the key property for the application of the classical circle method. We can always ensure that it is satisfied by passing to a submatrix and the use of the following lemma.

Lemma 4.4. *Let A be a $t \times m$ matrix over a field K and q be a positive integer. Then either A includes q disjoint $t \times t$ non-singular submatrices or all but at most $q(t-d)-1$ columns are contained in a d-dimensional subspace for some $0 \leq d \leq t-1$.*

Proof. This is a special case of Proposition 6.45 in [1] and a proof may also be found in [17]. □

Now, either H satisfies the property [AP] or we can remove the 'bad' column and apply Lemma 4.4 with $q = 2$. This gives us that at most $2(r-d)-1$ columns are not contained in a d-dimensional subspace for some $0 \leq d \leq r-1$. Remove these other exceptional columns as well and we obtain a matrix $\overline{H}$ with rank at most d and $w \geq s-2r-2(r-d)$ columns. Rename variables if necessary and write $H = \begin{pmatrix} H' & \overline{H} \end{pmatrix}$. By splitting the other matrices accordingly, we arrive at the form

$$Q = \begin{pmatrix} A & R & M' & \overline{M} \\ R^T & B & N' & \overline{N} \\ M'^T & N'^T & C_{11} & C_{12} \\ \overline{M}^T & \overline{N}^T & C_{12}^T & C_{22} \end{pmatrix},$$

where $\overline{M} = \hat{M}\overline{H}$, $\overline{N} = \hat{N}\overline{H}$, $M' = \hat{M}H'$, $N' = \hat{N}H'$. The matrix C splits according to formula (4.2) into parts $C_{11}, C_{12}, C_{12}^T, C_{22}$ with $C_{12} = H'^T\hat{N}^T R^{-1}\hat{M}\overline{H}$, for example.

If we choose d minimal in the above procedure, we end up with a matrix $\overline{H}$ that contains d linearly independent rows, which (as a matrix) satisfy property [AP]. Otherwise, we could apply Lemma 4.4 again and obtain the same result for a smaller value of d.

Now we decompose our variables suitable for this decomposition. Since $\overline{H}$ has w columns, we have $C_{22} \in \mathbb{Z}^{w\times w}$ with $w \geq s-4r+2d > 0$ if $s \geq 5+3r$. Call the first $s-w$ variables $\mathbf{y} = (y_1, \dots, y_{s-w})$ and the remaining w variables $\mathbf{x} = (x_1, \dots, x_w)$. The original equation $\mathbf{x}^T Q\mathbf{x} = 0$ decomposes into

$$\mathbf{y}^T \begin{pmatrix} A & R & M' \\ R^T & B & N' \\ M'^T & N'^T & C_{11} \end{pmatrix} \mathbf{y} + 2\mathbf{y}^T \begin{pmatrix} \overline{M} \\ \overline{N} \\ C_{12} \end{pmatrix} \mathbf{x} + \mathbf{x}^T C_{22}\mathbf{x} = 0. \tag{4.3}$$

For $1 \leq i \leq d$ we add linear equations

$$\mu_{i1}x_1 + \ldots + \mu_{it}x_w = h_i,$$

where the coefficients $\mu_{ij} \in \mathbb{Z}$ correspond to d linearly independent rows of $\overline{H}$ that satisfy [AP]. The variables h_i may have any integer value, but due to the restrictions on the x_j they will range over a bounded interval of size $O_Q(N)$ as well. Any occurrence of the term $\overline{H}\mathbf{x}$ can now be replaced by a suitable linear combination of the variables $\mathbf{h}$. Write Z_1 for the first matrix in (4.3) and $Z_2 = \begin{pmatrix} \hat{M} & \hat{N} & H'^T\hat{N}^T R^{-1}\hat{M} \end{pmatrix}^T$. From equation (4.2) we obtain

$$C_{22} = \overline{H}^T \hat{N}^T R^{-1}\hat{M}\overline{H} + D_1$$

for a diagonal matrix D_1 and can replace $\mathbf{x}^T C_{22}\mathbf{x}$ by $\mathbf{h}^T Z_3\mathbf{h} + \mathbf{x}^T D_1\mathbf{x}$ with $Z_3 = \hat{N}^T R^{-1}\hat{M}$. Then equation (4.3) changes into

$$\mathbf{y}^T Z_1\mathbf{y} + 2\mathbf{y}^T Z_2\mathbf{h} + \mathbf{h}^T Z_3\mathbf{h} + \mathbf{x}^T D_1\mathbf{x} = 0.$$

Combine the parts, which only contain $\mathbf{y}$ and $\mathbf{h}$, into a single quadratic form

$$P(\mathbf{y},\mathbf{h}) = \begin{pmatrix}\mathbf{y}^T & \mathbf{h}^T\end{pmatrix}\begin{pmatrix}Z_1 & Z_2\\ Z_2^T & Z_3\end{pmatrix}\begin{pmatrix}\mathbf{y}\\ \mathbf{h}\end{pmatrix}.$$

If we write λ_i for the diagonal entries in D_1, we obtain the system

$$\begin{aligned}\lambda_1 x_1^2 + \ldots + \lambda_w x_w^2 &= P(\mathbf{y},\mathbf{h}),\\ \mu_{11}x_1 + \ldots + \mu_{1w}x_w &= h_1,\\ \vdots \qquad\qquad \vdots \qquad &\quad \vdots\\ \mu_{d1}x_1 + \ldots + \mu_{dw}x_w &= h_d,\end{aligned} \tag{4.4}$$

where the matrix $(\mu_{ij}) \in \mathbb{Z}^{d\times w}$ has property [AP] and the matrix of the quadratic form P has off-rank at least $r \geq d$ uniformly in $\mathbf{h}$. The number of diagonal variables is at least $w \geq s-4r+2d$, which is at least $5+2d-r > 0$ for $s \geq 5+3r$ and $r \leq 4$.

If necessary, we can multiply the first equation (4.4) by a suitable non-zero integer to ensure that all the entries in the matrix P are integers, thereby avoiding any complications later.

5. Preparation for Section 6

In Section 4 we found $d \leq 4$ linear equations such that by adding them to our original quadric $Q(\mathbf{x}) = 0$ we end up with the partially diagonal form (4.4). Some of the coefficients λ_i can be zero and we split the vector $\mathbf{x} = (x_1,\ldots,x_w)$ according to this condition. Denote by z_i the variables with vanishing coefficients in the quadratic equation. We obtain after renaming

$$\begin{aligned}\lambda_1 x_1^2 + \ldots + \lambda_u x_u^2 &= P(\mathbf{y},\mathbf{h}),\\ \nu_{11}z_1 + \ldots + \nu_{1v}z_v + \mu_{11}x_1 + \ldots + \mu_{1u}x_u &= h_1,\\ \vdots \qquad\qquad \vdots \qquad \vdots \qquad\qquad \vdots \qquad &\quad \vdots\\ \nu_{d1}z_1 + \ldots + \nu_{dv}z_v + \mu_{d1}x_1 + \ldots + \mu_{du}x_u &= h_d,\end{aligned} \tag{5.1}$$

where $u+v=w$. We recall that the integer variables h_i can be restricted to a bounded range $[-CN, CN]$ for some $C \geq 1$.

There are now two cases to consider, which result in completely different treatments of the system of equations. In the first case (the one we consider in the next section) the columns $(\nu_{ij})_{1\leq i\leq d}$ in the linear part of the system are linearly independent. In particular, this implies that $v \leq d$. If they are not linearly independent, we deal with this system in Section 8 and use methods of Gowers [8] on linear equations in dense sets.

Recall the definition of $V_g(\alpha,\beta)$ and $L_g(\alpha)$ from (2.3) and define a new exponential sum corresponding to the right hand side of (5.1) by

$$T_g(\alpha,\boldsymbol{\beta}) = \sum_{\substack{\mathbf{y}\leq N\\ |\mathbf{h}|\leq CN}} g(\mathbf{y})e(\alpha P(\mathbf{y},\mathbf{h}) + \boldsymbol{\beta}\cdot\mathbf{h}).$$

In the case $g(\mathbf{y}) = 1_{\mathcal{A}^{s-w}}(\mathbf{y}) = 1_{\mathcal{A}}(y_1)\cdots 1_{\mathcal{A}}(y_{s-w})$ we write $T_{\mathcal{A}}(\alpha,\boldsymbol{\beta})$ instead. We can use the trivial estimate

$$|T_g(\alpha,\boldsymbol{\beta})| \le \sum_{|\mathbf{h}|\le CN} \Big|\sum_{\mathbf{y}\le N} g(\mathbf{y})e(\alpha P(\mathbf{y},\mathbf{h}))\Big|$$

to remove the linear term. Since the quadratic form $P(\mathbf{y},\mathbf{h})$ has off-rank at least r for any $\mathbf{h}$, we get (as in the proof of the second part of Theorem 3.1)

$$|T_g(\alpha,\boldsymbol{\beta})| \ll N^{s-w+d-r}K(\alpha)^r, \tag{5.2}$$

for $K(\alpha)$ as in (3.1) and any bounded $g(\mathbf{y}) = g_1(y_1)\cdots g_{s-w}(y_{s-w})$.

To simplify the estimation of (6.3) we introduce abbreviations for two often occuring 'actions'.

Rep: (Replace) *Suppose that one is given a set of linearly independent linear forms Ω and another linear form l. Choose one linear form l' from the set Ω and exchange it with l in such a way that $\{l\}\cup\Omega\backslash\{l'\}$ is still linearly independent. We use this action to replace one exponential sum by another in such a way that the corresponding linear forms obey the replacement property.*

EaS: (Estimate and Simplify) *If we estimate an exponential sum integral by a sum of several such integrals, we need to deal only with the maximal contribution. This happens, for example, when we apply the inequality $x_1^{p_1}\cdots x_n^{p_n} \le x_1^p + \ldots + x_n^p \ll x_1^p$ with $p = p_1+\ldots+p_n$ and assume without loss of generality that x_1^p is the largest term. If different terms require different treatments, we make an additional case analysis.*

6. Convexity and Correlation Estimates

Now that we set up the necessary notation and conventions, we can write the number of solutions to (5.1) as the Fourier-integral

$$\int_{\mathbb{T}^{d+1}} T_{\mathcal{A}}(\alpha,\boldsymbol{\beta}) \prod_{i=1}^{v} L_{\mathcal{A}}(\boldsymbol{\nu}_i\cdot\boldsymbol{\beta}) \prod_{j=1}^{u} V_{\mathcal{A}}(\lambda_j\alpha, \boldsymbol{\mu}_j\cdot\boldsymbol{\beta})\, d\alpha d\boldsymbol{\beta}. \tag{6.1}$$

For $\mathcal{A}\subset[1,N]$ and $\delta = N^{-1}|\mathcal{A}|$ we define the balanced function f by

$$f(n) = 1_{\mathcal{A}}(n) - \delta. \tag{6.2}$$

We replace each occurrence of $1_{\mathcal{A}}$ in the integral above by $f + \delta 1_{[1,N]}$ and expand (6.1) into 2^s integrals of the form

$$E = \int_{\mathbb{T}^{d+1}} T_g(\alpha,\boldsymbol{\beta}) \prod_{i=1}^{v} L_{f_i}(\boldsymbol{\nu}_i\cdot\boldsymbol{\beta}) \prod_{j=1}^{u} V_{g_j}(\lambda_j\alpha, \boldsymbol{\mu}_j\cdot\boldsymbol{\beta})\, d\alpha d\boldsymbol{\beta}, \tag{6.3}$$

where $f_i, g_j \in \{f, \delta 1_{[1,N]}\}$ and g is a product of $s-w$ such functions. We consider the $2^s - 1$ integrals that contain the function f as 'error terms' and give upper bounds on their size to deduce correlation estimates for the exponential sums later.

We start with the case, where $f_1 = f$ and show later how to modify the argument to obtain the estimate in the other cases. By pulling out half of L_f, we obtain from (6.3) the estimate

$$E \ll \sup_{\beta} |L_f(\beta)|^{1/2} \int_{\mathbb{T}^{d+1}} |T_g||L_{f_1}|^{1/2} \prod_{i=2}^{v} |L_{f_i}| \prod_{j=1}^{u} |V_{g_j}| \, d\alpha d\boldsymbol{\beta},$$

where we left out the dependences on the variables to save space. By property [AP] from Section 4, the linear forms $\boldsymbol{\nu}_i \cdot \boldsymbol{\beta}$ and $\boldsymbol{\mu}_j \cdot \boldsymbol{\beta}$ in the exponential sums L_{f_i} and V_{g_j} contain two basis sets after removing $\boldsymbol{\nu}_1 \cdot \boldsymbol{\beta}$. We can group these exponential sums into two products W_1 and W_2 corresponding to the two bases. Since the linear forms $\boldsymbol{\nu}_i \cdot \boldsymbol{\beta}$ are linearly independent by assumption, we can make sure that all L_{f_i} for $2 \leq i \leq u$ are contained in W_1. We obtain (after renaming) an estimate of the form

$$E \ll \sup_{\beta} |L_f(\beta)|^{1/2} \int_{\mathbb{T}^{d+1}} |T_g||L_{f_1}|^{1/2} |W_1||W_2| \prod_{j=1}^{w-2d-1} |V_{g_j}| \, d\alpha d\boldsymbol{\beta}. \tag{6.4}$$

Since the half-power of a linear exponential sum would cause problems later, we find by [Rep] an exponential sum V inside W_2 that can be replaced by L_{f_1}. Then we apply [EaS] with the estimate $|L|^{1/2}|V| \leq |V|^{1/2}|L| + |V|^{3/2}$ to replace $|L|^{1/2}$ by $|V|^{1/2}$. In the case of the first summand, we have substituted $|V|$ by $|L|$ inside W_2. Assume without loss of generality that V is $V_{g_{w-2d}}$. Our estimate is now

$$E \ll \sup_{\beta} |L_f(\beta)|^{1/2} \int_{\mathbb{T}^{d+1}} |T_g||V_{g_{w-2d}}|^{1/2} |W_1||W_2| \prod_{j=1}^{w-2d-1} |V_{g_j}| \, d\alpha d\boldsymbol{\beta}.$$

At this point we are in the position to apply a cascade of estimates. We replace $|T_g|$ by the right hand side of (5.2), use [EaS] to replace the product of the V_{g_j} (including $|V_{g_{w-2d}}|^{1/2}$) by $|V_{g_1}|^{w-2d-1/2}$ and again [EaS] with the estimate $|W_1 W_2| \leq |W_1|^2 + |W_2|^2$. We obtain a much simpler upper bound of the form

$$E \ll N^{s-w-r+d} \sup_{\beta} |L_f(\beta)|^{1/2} \int_{\mathbb{T}^{d+1}} K^r |V_{g_1}|^{w-2d-1/2} |W_1|^2 \, d\alpha d\boldsymbol{\beta}.$$

Before we proceed further, we reduce the power of $|V_{g_1}|$ in order to obtain a more uniform estimate in δ later on. We can replace $w - 2d \geq s - 4r \geq 5 - r$ by the minimal value $5 - r$ for which the following argument works and pull out any additional powers of $|V_{g_1}|$. With the trivial estimate $|V_{g_1}| \leq 2\delta N$ we obtain

$$E \ll \delta^{w-2d+r-5} N^{s-d-5} \sup_{\beta} |L_f(\beta)|^{1/2} \int_{\mathbb{T}^{d+1}} K^r |V_{g_1}|^{9/2-r} |W_1|^2 \, d\alpha d\boldsymbol{\beta}.$$

Now we continue with the main argument. The estimate

$$K^r |V_{g_1}|^{9/2-r} \leq K^{9/2} + |V_{g_1}|^{9/2} \tag{6.5}$$

results in two cases. If the first expression on the right hand side gives the dominating contribution, then

$$E \ll \delta^{w-2d+r-5} N^{s-d-5} \sup_{\beta} |L_f(\beta)|^{1/2} \int_{\mathbb{T}^{d+1}} K^{9/2} |W_1|^2 \, d\alpha d\boldsymbol{\beta}.$$

Here we can use the fact that K only depends on α with Parseval's identity (or equivalently 'interpretation of the integral as a diophantine equation')

$$\int_{\mathbb{T}^d} |W_1(\alpha, \boldsymbol{\beta})|^2 \, d\boldsymbol{\beta} = N^d,$$

to get the bound

$$E \ll \delta^{w-2d+r-5} N^{s-5} \sup_{\beta} |L_f(\beta)|^{1/2} \int_{\mathbb{T}} K(\alpha)^{9/2} \, d\alpha.$$

We can apply Lemma 3.2 to get the final upper bound

$$E \ll \delta^{w-2d+r-5} N^{s-5/2} \sup_{\beta} |L_f(\beta)|^{1/2}. \tag{6.6}$$

If the second term in (6.5) gives the dominating contribution, then

$$E \ll \delta^{w-2d+r-5} N^{s-d-5} \sup_{\beta} |L_f(\beta)|^{1/2} \int_{\mathbb{T}^{d+1}} |V_{g_1}|^{9/2} |W_1|^2 \, d\alpha d\boldsymbol{\beta}, \tag{6.7}$$

and things are slightly more complicated. First we find by [Rep] the linear form in W_1 that corresponds to $\boldsymbol{\mu}_1 \cdot \boldsymbol{\beta}$. This linear form can either belong to another function V_{g_j} or $\boldsymbol{\mu}_1 \cdot \boldsymbol{\beta}$ must be in the span of the linear forms of the linear exponential sums L_{f_i} that appear in W_1. We have to deal with these subcases differently.

In the first subcase, we can use [EaS] with the estimate

$$|V_{g_1}|^{9/2} |V_{g_j}|^2 \ll |V_{g_1}|^{13/2}$$

and make the change of variables $\gamma_i = \boldsymbol{\nu}_i \cdot \boldsymbol{\beta}$ and $\gamma_j = \boldsymbol{\mu}_j \cdot \boldsymbol{\beta}$ to bound the integral in (6.7) by a multiple of

$$\int_{\mathbb{T}^{d+1}} |V_{g_1}(\lambda_1 \alpha, \gamma_1)|^{13/2} \prod_{i=2}^{h} |L_{f_i}(\gamma_i)|^2 \prod_{j=h+1}^{d} |V_{g_j}(\lambda_j \alpha, \gamma_j)|^2 \, d\alpha d\boldsymbol{\gamma},$$

for some $h \leq d$ dependent on the previous steps. Since the variables γ_j appear now separately, we can use Parseval's identity $d-1$ times to integrate out γ_2 to γ_d and obtain

$$E \ll \delta^{w-2d+r-5} N^{s-6} \sup_{\beta} |L_f(\beta)|^{1/2} \int_{\mathbb{T}^2} |V_{g_1}(\lambda_1 \alpha, \gamma_1)|^{13/2} \, d\alpha d\gamma_1.$$

A change of variables (to remove λ_1) and Theorem C.1 imply the final bound

$$E \ll \delta^{w-2d+r-5} N^{s-5/2} \sup_{\beta} |L_f(\beta)|^{1/2}.$$

In the second subcase of the analysis of (6.7), we can bound the integral in (6.7) (after a change of variables) by a multiple of

$$\int_{\mathbb{T}^{d+1}} |V_{g_1}(\lambda_1\alpha, \boldsymbol{\nu}\cdot\boldsymbol{\gamma})|^{9/2} \prod_{i=1}^{h} |L_{f_i}(\gamma_i)|^2 \prod_{j=h+1}^{d} |V_{g_j}(\lambda_j\alpha, \gamma_j)|^2 \, d\alpha d\boldsymbol{\gamma},$$

for some $\boldsymbol{\nu} \in \mathbb{Q}^d$ with the property that the linear form $\boldsymbol{\nu}\cdot\boldsymbol{\gamma}$ only contains the variables $\gamma_1, \ldots, \gamma_h$. As before, we integrate out the functions $|V_{g_j}|^2$ by Parseval's identity and end up with

$$N^{d-h} \int_{\mathbb{T}^{h+1}} |V_{g_1}(\lambda_1\alpha, \boldsymbol{\nu}\cdot\boldsymbol{\gamma})|^{9/2} \prod_{i=1}^{h} |L_{f_i}(\gamma_i)|^2 \, d\alpha d\gamma_1 \cdots d\gamma_h.$$

Now we are again in a situation similar to the case with $K(\alpha)$. Observe that the functions L_{f_i} are independent of α. We can use Lemma C.4 and Parseval's identity for the integrals over $\gamma_1, \ldots, \gamma_h$ to bound this from above by

$$\begin{aligned} N^{d-h} \sup_{\beta} \int_{\mathbb{T}} |V_{g_1}(\lambda_1\alpha, \beta)|^{9/2} \, d\alpha \int_{\mathbb{T}^h} \prod_{i=1}^{h} |L_{f_i}(\gamma_i)|^2 \, d\gamma_1 \cdots d\gamma_h \\ \ll N^{d-h} N^{5/2} N^h = N^{d+5/2}. \end{aligned}$$

Combined with (6.7), we obtain estimate (6.6) again.

This is the end of the estimates under the assumption that $f_1 = f$. We now address the other cases. Obviously, the same procedure works, when $f_i = f$ for some other $i \le v$. If $g_j = f$ for some $j \le u$, the argument is even simpler. We obtain instead of (6.4) the inequality

$$E \ll \sup_{\alpha,\beta} |V_f(\alpha,\beta)|^{1/2} \int_{\mathbb{T}^{d+1}} |T_g||V_{g_{w-2d}}|^{1/2}|W_1||W_2| \prod_{j=1}^{w-2d-1} |V_{g_j}| \, d\alpha d\boldsymbol{\beta},$$

which can be treated in the same manner as before and gives the final bound

$$E \ll \delta^{w-2d+r-5} N^{s-5/2} \sup_{\alpha,\beta} |V_f(\alpha,\beta)|^{1/2}.$$

To describe the last remaining case write $g(\mathbf{y}) = \prod_{k=1}^{s-w} h_k(y_k)$ for the function g appearing in T_g with $h_k = f$ for some $1 \le k \le s-w$. Instead of pulling out half of T_g, we take the $1/(2r)$-th power. This gives us

$$E \ll \sup_{\alpha,\boldsymbol{\beta}} |T_g(\alpha,\boldsymbol{\beta})|^{1/2r} \int_{\mathbb{T}^{d+1}} |T_g|^{1-1/2r}|W_1||W_2| \prod_{j=1}^{w-2d} |V_{g_j}| \, d\alpha d\boldsymbol{\beta}$$

instead of (6.4). The same estimates as before apply, where $K(\alpha)^{1/2}$ is replaced by $|V_{g_1}|^{1/2}$. This does not matter since we use [EaS] with (6.5) to separate the cases with K and V_{g_1} anyway. The final estimate has the slightly different form

$$E \ll \delta^{w-2d+r-5} N^{s-2-(s-w+d)/2r} \sup_{\alpha,\boldsymbol{\beta}} |T_g(\alpha,\boldsymbol{\beta})|^{1/2r}.$$

This concludes the last case of the error term estimates and we proceed to deduce the corelation estimates.

One of the integrals appearing in the expansion of (6.1) is equal to

$$\delta^s \int_{\mathbb{T}^{d+1}} T(\alpha,\boldsymbol{\beta}) \prod_{i=1}^{v} L(\boldsymbol{\nu}_i \cdot \boldsymbol{\beta}) \prod_{j=1}^{u} V(\lambda_j \alpha, \boldsymbol{\mu}_j \cdot \boldsymbol{\beta})\, d\alpha d\boldsymbol{\beta} \tag{6.8}$$

and gives the expected number of solutions to (5.1) with $x_i, z_j \in [1,N]$. From the discussion in Section 2 and Lemma 2.3, we know that (6.8) is bounded from below by $\gg \delta^s N^{s-2}$. On the other hand, (6.1) is of size exactly δN since by assumption there are only trivial solutions to our system (5.1). Therefore, at least one of the $2^s - 1$ 'error terms' E has to be of size $\gg \delta^s N^{s-2}$ if $N \gg_Q \delta^{-2}$. The three previously obtained upper bounds transform with $w \geq s - 4r + 2d$ into one of the correlation estimates

$$\sup_{\alpha,\boldsymbol{\beta}} |T_g| \gg \delta^{2r(3r+5)} N^{s-w+d}, \quad \sup_{\alpha,\beta} |V_f| \gg \delta^{6r+10} N \quad \text{or} \quad \sup_{\alpha} |L_f| \gg \delta^{6r+10} N,$$

as long as $N \gg_Q \delta^{-2}$ and $r \leq 4$. In the next section, these lower bounds are used to obtain structural information about the set $\mathcal{A}$.

7. Density Increment Method

In equation (3.7) and Section 6 we have found the correlation estimates

$$\sup_{\alpha} |S_{f_i}(\alpha)| \gg \delta^{s+80} N^s, \qquad \sup_{\alpha,\boldsymbol{\beta}} |T_g(\alpha,\boldsymbol{\beta})| \gg \delta^{136} N^{s-w+d},$$
$$\sup_{\alpha} |L_f(\alpha)| \gg \delta^{34} N, \qquad \sup_{\alpha,\beta} |V_f(\alpha,\beta)| \gg \delta^{34} N,$$

that hold for suitable functions f_i, g as long as $N \gg_Q \delta^{-2}$. To reduce our work we transform the first three estimates to the fourth (with δ^{136} instead of δ^{34}) in the following way. We have from (3.6) the representation

$$\begin{aligned} S_{f_i}(\alpha) &= \sum_{\mathbf{x} \leq N} (1_{\mathcal{A}}(x_i) - \delta)\delta^{i-1} \prod_{j>i} 1_{\mathcal{A}}(x_j) e(\alpha Q(\mathbf{x})) \\ &= \sum_{\mathbf{x}' \leq N} \delta^{i-1} \prod_{j>i} 1_{\mathcal{A}}(x_j) \sum_{x_i \leq N} f(x_i) e(\alpha q_{\mathbf{x}'}(x_i)), \end{aligned}$$

where $\mathbf{x}'$ is the vector of variables without x_i and $q_{\mathbf{x}'}(x_i) = Q(\mathbf{x})$ is seen as a quadratic polynomial in x_i with linear and constant coefficients depending on $\mathbf{x}'$. Therefore, we deduce for some $d \in \mathbb{Z}$ and $\beta = \beta(\mathbf{x}', \alpha)$ the estimate

$$|S_{f_i}(\alpha)| \leq \sum_{\mathbf{x}' \leq N} \delta^{i-1} \prod_{j>i} 1_{\mathcal{A}}(x_j) |V_f(d\alpha, \beta)|.$$

Take the supremum over α and β of V outside the sum and conclude that

$$\begin{aligned} \delta^{s+80} N^s \ll \sup_{\alpha} |S_{f_i}(\alpha)| &\leq \sup_{\alpha,\beta} |V_f(\alpha,\beta)| \sum_{\mathbf{x}' \leq N} \delta^{i-1} \prod_{j>i} 1_{\mathcal{A}}(x_j) \\ &\leq \delta^{s-1} N^{s-1} \sup_{\alpha,\beta} |V_f(\alpha,\beta)|, \end{aligned}$$

which gives the desired implication.

In a similar way we may reduce the second estimate to the fourth. The function g in T_g is a product of functions f from (6.2) and $\delta 1_{[1,N]}$. We expand all but one of the balanced functions f in g into $1_{\mathcal{A}} - \delta$, giving us several exponential sums of the form

$$T_{g_i}(\alpha, \boldsymbol{\beta}) = \sum_{\mathbf{y} \leq N, |\mathbf{h}| \leq CN} f(y_i) \prod_{j \neq i} g_j(y_j) e(\alpha P(\mathbf{y}, \mathbf{h}) + \boldsymbol{\beta} \cdot \mathbf{h})$$

with $g_j \in \{\delta 1_{[1,N]}, 1_{\mathcal{A}}\}$. At least one of these has to be big and this implies a correlation estimate in exactly the same way as for S_{f_i} if we estimate the sums over $\mathbf{h}$ trivially. Finally, we observe the identity $L_f(\beta) = V_f(0, \beta)$ which reduces the third case to the fourth.

The fourth correlation can be used to obtain a density increment by the following lemma.

Lemma 7.1. *If $|V_f(\alpha, \beta)| \geq \eta N$ for some $(\alpha, \beta) \in \mathbb{T}^2$ and $\eta > 0$, then there is an arithmetic progression $P \subset [1, N]$ of length $|P| \gg \eta^2 N^{1/16}$ with*

$$|\mathcal{A} \cap P| \geq (\delta + \eta/4)|P|.$$

Proof. This is the Lemma B.1 in Appendix B of [15]. Similar statements for finite fields can be found in [8], for example. □

We use the large Fourier estimate $|V_f(\alpha, \beta)| \gg \delta^{136} N$ from above with Lemma 7.1 to find a progression P of length $\gg \delta^{272} N^{1/16}$, such that $\mathcal{A}$ has density $\geq \delta + \lambda \delta^{136}$ on P, where $\lambda > 0$ is an absolute constant. Due to the translation and dilation invariance of $Q(\mathbf{x}) = 0$, we end up with the same problem on a subprogression, but with a slightly higher density.

Since the density is bounded by one, this procedure cannot last more than $\lambda^{-1} \delta^{-136}$ steps before reaching a contradiction. This means that at some stage we have a non-trivial solution or the size of our progression is $\ll \delta^{-2}$. The first option is not available by assumption. Therefore, we have

$$\delta^{150} N^{(1/16)^{\lambda^{-1} \delta^{-136}}} \ll_Q \delta^{-2}.$$

Rearranging for δ we can deduce that $\delta \ll (\log \log N)^{-c}$ with $c = (137)^{-1}$, for example.

8. A Linear Subsystem

In this section we consider the last two cases of Theorem 2.2. First let us look at the case from Section 6, when the columns of the linear part in (5.1) turn out to be linearly dependent. This allows us to extract a linear subsystem as follows.

Since the quadric is translation invariant we can set $y_k = z_j = x_i = z_0 \in \mathcal{A}$ for any z_0, and get a trivial solution of (5.1) with uniquely defined $h_i(z_0) = -\nu_{i0} z_0$ for some $\nu_{i0} \in \mathbb{Z}$. Equipped with this information, we can set $y_k = x_i = z_0 \in \mathcal{A}$ and $h_i = -\nu_0 z_0$ for all k and i in (5.1) and obtain the translation invariant system

$$\begin{aligned} \nu_{11}z_1 + \ldots + \nu_{1v}z_v + \nu_{10}z_0 &= 0, \\ \vdots \qquad\qquad\qquad\qquad \vdots \quad & \vdots \\ \nu_{d1}z_1 + \ldots + \nu_{dv}z_v + \nu_{d0}z_0 &= 0, \end{aligned} \tag{8.1}$$

with coefficients $\nu_{i0} \in \mathbb{Z}$. By assumption, the rank of the first v columns is at most $v-1$. By translation invariance, the last column is a linear combination of the first v columns and this implies that the rank of the whole coefficient matrix is still at most $v-1$. This gives a linear system in $v+1$ variables with at most $v-1$ independent equations. A very similar system appears in Section 2 in the situation, where the number of non-zero coefficients in (2.1) is at most four. Both systems can be treated by the method of Gowers [8], as we show now.

First we remove equations from (8.1) until the remaining rows become linearly independent. To simplify notation, we can assume that we are left with w equations and $w+2$ variables. If more variables are left over, we can 'fuse' them by setting them equal to z_0.

We can also assume that the columns of E are in general position. To see this we take $\mathbb{Z}$-linear combinations of rows and rename variables to transform the equation $E\mathbf{z} = 0$ into the form

$$\begin{pmatrix} D & \mathbf{a} & \mathbf{b} \end{pmatrix} \cdot \mathbf{z} = 0, \tag{8.2}$$

where $D \in \mathbb{Z}^{w \times w}$ is a diagonal matrix of full rank with non-zero diagonal entries and $\mathbf{a}, \mathbf{b} \in \mathbb{Z}^w$ are integer vectors. The property that the columns of the matrix in (8.2) are in general position translates into the arithmetic conditions that $a_i \neq 0, b_i \neq 0$ for all i and $a_ib_j - a_jb_i \neq 0$ for $i \neq j$.

If this is not the case, we can combine two rows to deduce an equation of the form $mz_i = mz_j$ for some $1 \leq i < j \leq w+2$ and $m \neq 0$. (The form of this equation is forced by translation invariance.) It is then possible to reduce the system by one equation and one variable, leaving us with the same situation with parameter $w-1$ instead of w.

We should briefly discuss the case $w = 1$. In this case we have one equation in three variables of the form

$$d_1x_1 + d_2x_2 + d_3x_3 = 0,$$

with $d_1 + d_2 + d_3 = 0$. If $d_i \neq 0$ for all $1 \leq i \leq 3$, then we are in the situation of Roth's paper [18] and we obtain a bound of the form $O((\log\log N)^{-1})$ for the density of $\mathcal{A}$. If we have $d_i = 0$ for one of the coefficients, then we directly get a non-trivial solution in $\mathcal{A}$ as long as $|\mathcal{A}| \geq 2$.

Now consider the difference

$$\sum_{\mathbf{z}, E\mathbf{z}=0} 1_{\mathcal{A}}(z_0) \cdots 1_{\mathcal{A}}(z_{w+1}) - \sum_{\mathbf{z}, E\mathbf{z}=0} \delta 1_{[1,N]}(z_0) \cdots \delta 1_{[1,N]}(z_{w+1}).$$

Since we assume that our system has only trivial solutions, this difference is either of size around $\delta^{w+2}N^2$ or we have $N \ll \delta^{-w-1}$. In the second case, we are done. Otherwise, we can write $1_{\mathcal{A}} = f + \delta 1_{[1,N]}$ for the balanced function f from (6.2) and expand the first sum into 2^{w+2} terms, one of which is cancelled by the second term. Then we bound the remaining sums of the form

$$\sum_{\mathbf{z}, E\mathbf{z}=0} f_0(z_0) \cdots f_{w+1}(z_{w+1}). \tag{8.3}$$

with functions $f_i \in \{f, \delta 1_{[1,N]}\}$ and at least one of these f_i equal to f.

To simplify the exposition and be able to cite a result from [8], we convert this sum into one with variables $z_i \in \mathbb{Z}/M\mathbb{Z}$ for some prime M dependent on N and E. We choose the prime M in an interval $[C_E N, 2C_E N]$, which is possible by Bertrand's postulate. More precisely, we take C_E in such a way that the equation $E\mathbf{z} = 0$ in $\mathbb{Z}/M\mathbb{Z}$ with $z_i \in [1, N]$ implies that $E\mathbf{z} = 0$ in $\mathbb{Z}$. This is always possible, if C_E is big enough to avoid 'wrap-around issues'. We set the functions $f_i(z_i) = 0$ for $z_i \notin [1, N]$.

The 'error terms' in (8.3) can be bounded by more symmetric expressions, the so called *uniformity norms.* This is done in Appendix B, where we deduce Lemma B.1. We get the estimate (see Appendix B for a definition of Δ)

$$\begin{aligned} &M^{-2} \sum_{\mathbf{z}, E\mathbf{z}=0} f_0(z_0) \cdots f_{w+1}(z_{w+1}) \\ &\leq \Big(M^{-w-2} \sum_{h_1,\ldots,h_w} \Big| \sum_{z_{w+1}} \Delta(f_{w+1}; \mathbf{h})(z_{w+1}) \Big|^2 \Big)^{2^{-w-1}}. \end{aligned} \tag{8.4}$$

For this to be useful, we have to assume that $f_{w+1} = f$. If this was not the case, we can rename variables to obtain the desired outcome.

According to [8] we define a function $f : \mathbb{Z}/M\mathbb{Z} \to \mathbb{C}$ to be α-uniform of degree w if

$$\sum_{h_1,\ldots,h_w} \Big| \sum_z \Delta(f; \mathbf{h})(z) \Big|^2 \leq \alpha M^{w+2}.$$

A set $\mathcal{A}$ is α-uniform if its balanced function f given by (6.2) is α-uniform. By assumption, our error term is $\gg \delta^{w+2} N^2 \gg \delta^{w+2} M^2$, so in combination with the estimate (8.4), we derive that f is not $\mu(\delta^{w+2})^{2^{w+1}}$-uniform for some small $\mu > 0$. Combined with the following theorem of Gowers this allows us to deduce a density increment for $\mathcal{A}$ on a subprogression, similar to the one in Lemma 7.1.

Theorem 8.1. *Let $\alpha \leq 1/2$ and let $\mathcal{A} \subset \mathbb{Z}/M\mathbb{Z}$ be a set which fails to be α-unform of degree k. There exists a partition of $\mathbb{Z}/M\mathbb{Z}$ into arithmetic progressions $P_1, \ldots, P_K$ of average size $M^{\alpha^{2^{2^{k+10}}}}$ such that*

$$\sum_{j=1}^{K} \Big| \sum_{s \in P_j} f(s) \Big| \geq \alpha^{2^{2^{k+10}}} M.$$

Proof. This is Theorem 18.5 from [8]. □

We note that the term *arithmetic progression* in the theorem refers to 'proper' arithmetic progressions, which keep their structure if projected down to $[1, N]$. For details the reader is referred to [8].

To obtain the density increment we observe that

$$\sum_{j=1}^{K}\sum_{s\in P_j} f(s) = \sum_{s\leq N} f(s) = 0$$

by the definition of f given in (6.2). Therefore, by Theorem 8.1, we have

$$\sum_{s\in P_j} f(s) \geq \frac{1}{2}\alpha^{2^{2^{k+10}}}|P_j|$$

for some $j \leq K$, which can be translated into a density increment of size

$$\frac{1}{2}\left(\mu(\delta^{w+2})^{2^{w+1}}\right)^{2^{2^{w+10}}}$$

on P_j. If we perform the same iteration as in Section 7, we obtain the same result but with $c = 2^{-2^{15}}$. This is due to the fact that we need the argument only for $w \leq 4$, which makes c independent of the number of variables s.

Appendix A. A Version of Vinogradov's Lemma

Here we prove Lemma A.1, a version of Vinogradov's lemma, that we need for the estimation of bilinear exponential sums. While there are many versions in the literature, none of those seems to be good enough for our purpose. Since we need to reprove it with explicit dependance on β, we take the opportunity to state it with explicit constants as well, to provide a reference for possible numerical applications.

Lemma A.1. *Let $\alpha = a/q + \beta$ with $|\beta| \leq \frac{1}{qN}$ and $(a;q) = 1$, then we have*

$$\sum_{|h|\leq N} \min\{N, \|\alpha h\|^{-1}\} \leq 6N + 6\min\left\{\frac{N^2}{q}, \frac{1}{|\beta|q}\left(|\log(|\beta|N^2)| + 2\right)\right\} + (4N + 2q)\cdot(1 + \log q).$$

Proof. We insert the formula $\alpha = a/q + \beta$ and rearrange the expression according to the residue class of $h \mod q$, giving us

$$\sum_{|h|\leq N} \min\{N, \|ah/q + \beta h\|^{-1}\} = \sum_{c=0}^{q-1} \sum_{\substack{|h|\leq N \\ h\equiv c \mod q}} \min\{N, \|ac/q + \beta h\|^{-1}\}.$$

Since $|\beta| \leq \frac{1}{qN}$ and $h \leq N$, the term βh is bounded by $1/q$ and since $(a;q) = 1$, there are at most three values of c such that $\|ac/q + \beta h\| < 1/q$ is possible. For the other values of c, we can bound the expression $\|ac/q + \beta h\|$ from below by d/q for some $1 \leq d \leq q/2$ independent of $|h| \leq N$. Take the maximal d for which the inequality holds and arrange the values of c according to these d-values. There are at most two values of c for each d

and we end up with the bound

$$\sum_{|h|\le N} \min\{N, \|\alpha h\|^{-1}\} \le 3 \sup_{0\le c\le q-1} \sup_{x_0\in[0,1]} \sum_{\substack{|h|\le N\\ h\equiv c \mod q}} \min\{N, \|x_0+\beta h\|^{-1}\}$$
$$+ 2\cdot \frac{2N+q}{q} \sum_{1\le d\le q/2} \|d/q\|^{-1}.$$

We use the estimate $\sum_{1\le d\le q/2} d^{-1} \le 1+\log q$ and get for the second term

$$2\cdot \frac{2N+q}{q} \sum_{1\le d\le q/2} \|d/q\|^{-1} \le 2\cdot(2N+q)\cdot(1+\log q),$$

which accounts for the second line in the estimate of Lemma A.1. Write $h = c + lq$ such that the first expression changes into

$$3 \sup_{0\le c\le q-1} \sup_{x_0\in[0,1]} \sum_{|l+c/q|\le N/q} \min\{N, \|x_0+\beta q l\|^{-1}\}.$$

Since the function $\min\{N, \|\gamma\|^{-1}\}$ is monotone for $0 \le \gamma \le 1/2$ and symmetric around the origin, we can choose $x_0 = c = 0$ by taking care of a possible boundary term and obtain the upper bound

$$3N + 3\sum_{|l|\le N/q} \min\{N, \|\beta q l\|^{-1}\}. \tag{A.1}$$

There is a small problem with this argument for $q = 1$ due to 'wrap-around issues' in $\mathbb{R}/\mathbb{Z}$, but the bound in (A.1) is trivial in this case.

We assume without loss of generality that $\beta > 0$ and split the summation into $-1/(\beta q N) < l < 1/(\beta q N)$ and the positive and negative part of the sum over $1/(\beta q N) \le |l| \le N/q$. The first sum gives a contribution of at most $N\cdot(2/(\beta q N)+1)$ and the the two other sums give

$$\frac{2}{\beta q} \sum_{1/(\beta q N)\le l\le N/q} l^{-1} \le \frac{2}{\beta q}\left(|\log(N/q) - \log(1/(\beta q N))| + 1\right)$$
$$= \frac{2}{\beta q}\left(|\log(\beta N^2)| + 1\right).$$

There is also the trivial bound $N(2N/q+1)$ for the sum in (A.1), which is superior for $\beta \le 1/N^2$. We put everything together and replace β by $|\beta|$ to obtain the result. □

Appendix B. Uniformity Estimates

Here we want to give a proof for the uniformity estimate (8.4). This is a slightly different version of a result of Gowers [8] and can be deduced from the very general estimates in Appendix C of [12]. Since our result is a very special case with a simpler proof, we give it here for completeness.

Define $\Delta(f; h_1, \ldots, h_d)$ inductively by $\Delta(f;h)(x) = f(x)\overline{f(x-h)}$ and

$$\Delta(f; h_1, \ldots, h_{d+1}) = \Delta(\Delta(f; h_1, \ldots, h_d); h_{d+1}).$$

Consider

$$A(M,E,\mathbf{f}) := M^{-2} \sum_{\mathbf{x}, E\mathbf{x}=0} f_1(x_1)\cdots f_{w+2}(x_{w+2}), \tag{B.1}$$

where the variables are summed over $\mathbb{Z}/M\mathbb{Z}$ for some prime M.

Lemma B.1. *For a matrix $E \in \mathbb{Z}^{w\times(w+2)}$ with columns in general position over $\mathbb{Z}/M\mathbb{Z}$ and bounded functions $|f_i| \leq 1$, we have the bound*

$$|A(M,E,\mathbf{f})| \leq \Big(M^{-w-2} \sum_{h_1,\dots,h_w} \Big| \sum_x \Delta(f_{w+2};\mathbf{h})(x) \Big|^2 \Big)^{1/2^{w+1}}.$$

Remark B.2. There is nothing special about f_{w+2} here except for simplicity of notation.

To simplify the exposition of the proof, we don't write down all the normalisation constants M^{-t}, where t is the number of free variables in the summation. We write $\leq_M$ instead of $\leq$ to say that the estimate holds up to a power of M that has the right order of magnitude.

Proof. Since $|f_1| \leq 1$ we can estimate $A(M,E,\mathbf{f})$ by

$$\begin{aligned} A(M,E,\mathbf{f}) &\leq_M \sum_{x_1} \Big| \sum_{E\mathbf{x}=0}^{*} f_2(x_2)\cdots f_{w+2}(x_{w+2}) \Big| \\ &\leq_M \Big(\sum_{x_1} \Big| \sum_{E\mathbf{x}=0}^{*} f_2(x_2)\cdots f_{w+2}(x_{w+2}) \Big|^2 \Big)^{1/2}, \end{aligned}$$

where the star $*$ indicates that the inner sums run only over $x_2,\dots,x_s$. With $y_1 = x_1$ we can write the summation inside the square root as

$$\sum_{x_1} \sum_{E\mathbf{x}=0}^{*} \sum_{E\mathbf{y}=0}^{*} f_2(x_2)\cdots f_{w+2}(x_{w+2}) \overline{f_2(y_2)\cdots f_{w+2}(y_{w+2})}.$$

Define $h_1 := x_{w+2} - y_{w+2}$ and observe that from $E\mathbf{x} = 0$ and $E\mathbf{y} = 0$ we get $E(\mathbf{x}-\mathbf{y}) = 0$. By the definition of h_1 we obtain that

$$E \cdot (0, x_2 - y_2, \dots, x_{w+1} - y_{w+1}, h_1)^T = \mathbf{0}.$$

This equation uniquely determines the differences $x_i - y_i$ once h_1 is given and, therefore, y_i in terms of x_i and h_1. Denote this unique solution by y_{x_i,h_1} and set $F_i(h_1;x_i) = f_i(x_i)\overline{f_i(y_{x_i,h_1})}$. Then we can write the sum in the form

$$\sum_{h_1} \sum_{\mathbf{x}, E\mathbf{x}=0} F_2(h_1;x_2)\cdots F_{w+1}(h_1;x_{w+1}) \Delta(f_{w+2};h_1)(x_{w+2}).$$

Since this is just an average of another version of the original sum (B.1) with f_1 removed, we can use the estimate

$$\sum_{h_1} \Big(\sum \cdots \Big) \leq_M \Big(\sum_{h_1} \Big| \sum \cdots \Big|^2 \Big)^{1/2}$$

to apply the above procedure inductively to the inner sum. After w steps we have removed all functions but $\Delta(f_{w+2}; h_1, \dots, h_w)$ and some function $F_{\mathbf{h}}(x_{w+1})$. We are left with

$$A(M, E, \mathbf{f}) \leq_M \Big(\sum_{h_1,\dots,h_w} \sum_{\mathbf{x}, E\mathbf{x}=0} F_{\mathbf{h}}(x_{w+1}) \Delta(f_{w+2}; \mathbf{h})(x_{w+2}) \Big)^{1/2^w}.$$

Now we evaluate the sum over the variables $x_1, \dots, x_w$, since their values are given once we know the values of x_{w+1} and x_{w+2}. Having done this, we may rearrange the summation a last time and proceed by another application of the Cauchy-Schwarz-inequality and $|F_{\mathbf{h}}(x_{w+1})| \leq 1$ to obtain

$$\begin{aligned} A(M, E, \mathbf{f}) &\leq_M \left(\sum_{h_1,\dots,h_w} \Big(\sum_{x_{w+1}} 1 \Big) \Big| \sum_{x_{w+2}} \Delta(f_{w+2}; \mathbf{h})(x_{w+2}) \Big| \right)^{1/2^w} \\ &\leq_M \Big(\sum_{h_1,\dots,h_w} \Big| \sum_{x_{w+2}} \Delta(f_{w+2}; \mathbf{h})(x_{w+2}) \Big|^2 \Big)^{1/2^{w+1}}, \end{aligned}$$

which is the estimate in Lemma B.1 up to normalisation. □

Appendix C. L^p-Estimates for Quadratic Exponential Sums

In our recent work [15], we reproved a result of Bourgain [3], namely the following estimate for the two-dimensional quadratic exponential sum.

Theorem C.1. *Let $V_g(\alpha, \beta)$ be defined as in (2.3) for a function g with $|g(n)| \leq 1$. Then for $p > 6$, we have*

$$\int_{\mathbb{T}^2} |V_g(\alpha, \beta)|^p \, d\alpha d\beta \ll_p N^{p-3}.$$

Proof. This is Theorem 2.1 from [15]. □

Note that it is easy to get this result with N^{p-3} replaced by $N^{p-3} \log N$. In this appendix we use the same technique to prove a L^p-estimate for the corresponding one-dimensional exponential sum

$$U_g(\alpha) = \sum_{n \leq N} g(n) e(\alpha n^2),$$

where $g : \mathbb{N} \to \mathbb{C}$ is any function with $|g(n)| \leq 1$.

Theorem C.2. *For $p > 4$, we have*

$$\int_{\mathbb{T}} |U_g(\alpha)|^p \, d\alpha \ll_p N^{p-2}.$$

The proof is a simplified version of the proof in [15] and we use a general result (Theorem C.3 below) on L^p-estimates from our previous work. First we need some notation. Consider the square

$$U_g(\alpha)^2 = \sum_{n_1, n_2 \leq N} g(n_1) g(n_2) e(\alpha(n_1^2 + n_2^2)).$$

With $\omega(m) = \#\{n_1, n_2 \le N : m = n_1^2 + n_2^2\}$ we can write

$$f(m)\omega(m) = \sum_{\substack{n_1,n_2 \le N \\ n_1^2+n_2^2=m}} g(n_1)g(n_2)$$

for some function $|f| \le 1$. This leads to the definition of

$$W_f(\alpha) = U_g(\alpha)^2 = \sum_{m \le 2N^2} f(m)\omega(m)e(\alpha m).$$

For $f \equiv 1$ we get $W(\alpha)$. For an index set J decompose

$$W(\alpha) = \sum_{j\in J} W_j(\alpha) \quad \text{and} \quad \omega(m) = \sum_{j \in J} \omega_j(m),$$

where W_j is the exponential sum for ω_j. Define the L^p-norms as usual by

$$\|\omega\|_p := \Big(\sum_{m \le 2N^2} |\omega(m)|^p \Big)^{1/p} \text{ and } \|W\|_p := \Big(\int_{\mathbb{T}} |W(\alpha)|^p \, d\alpha \Big)^{1/p}_{.}$$

Now we can state the auxiliary result.

Theorem C.3. *For $p > 2$, $N \in \mathbb{N}$ and any $f : \mathbb{N} \to \mathbb{C}$ we have*

$$\|W_f\|_p \le \Big(\sum_{m \le 2N^2} |f(m)|^2 \omega(m) \Big)^{1/2} \Big(\sum_{j \in J} \|W_j\|_p^{(p-2)/p} \|\omega_j\|_{2p/(p-2)}^{2/p} \Big)^{1/2}_{.}$$

Proof. This is a special case of Theorem 4.1 from [15]. □

The first factor is just a weighted L^2-norm of f and is easily estimated in our context. The second factor needs much more attention and we will perform a variant of the major-minor-arc decomposition from the circle method. For small values of j we have the big major-arc contribitions in $\|W_j\|_p$ but the arithmetic counterparts ω_j are very regular 'almost periodic functions'. For larger values of j, the random fluctuations in ω_j contribute more and more to the sum, but are balanced by the savings on the side of the exponential sums W_j. The proposition below gives a quantitative version of these qualitative description.

First we need some notation. Define the local versions of $U(\alpha)$ to be

$$U(q, a) = \sum_{b=1}^{q} e(ab^2/q) \quad \text{and} \quad v(\alpha) = \int_1^N e(\alpha t^2)\, dt$$

and set the major arcs $\mathfrak{M}$ to be the union of

$$\mathfrak{M}(q, a) = \{\beta \in \mathbb{T} : \|\beta - a/q\| \le Q/N^2\}. \tag{C.1}$$

for $1 \le a \le q, (a; q) = 1$ and $q \le 4Q$. Note that they are disjoint for $4Q \le N^{2/3}$ and we set Q to be a small power of N with $16Q^6 \le N$ later. Define for $Y \le 2Q$ a dyadic part of the usual major arcs approximation for quadratic exponential sums (see [23, Theorem 7.2])

$$\mathfrak{U}_Y(\alpha) := \sum_{Y \le q < 2Y} q^{-2} \sum_{(a;q)=1} U(q, a)^2 v(\alpha - a/q)^2 \Big|_{\mathfrak{M}(q,a)}. \tag{C.2}$$

We take the Fourier transform and obtain the arithmetic functions

$$\begin{aligned}\omega_Y(m) &= \int_{\mathbb{T}} \mathfrak{U}_Y(\alpha)e(-\alpha m)\,d\alpha \\ &= \sum_{Y\le q<2Y} q^{-2} \sum_{(a;q)=1} U(q,a)^2 e(-am/q) \int_{|\beta|\le Q/N^2} v(\beta)^2 e(-\beta m)\,d\beta,\end{aligned}$$

which we restrict to $1 \le m \le 2N^2$. Set $\omega_Y(m) = 0$ for other values of m. Write $W(\alpha)$ for $U^2(\alpha)$ and define the corresponding exponential sums for $\omega_Y(m)$ as

$$W_Y(\alpha) = \sum_{m\le 2N^2} \omega_Y(m)e(\alpha m).$$

By inserting the definition of $\omega_Y(m)$ we see that $W_Y(\alpha)$ and $\mathfrak{U}_Y(\alpha)$ are related by the formula

$$W_Y(\alpha) = \int_{\mathbb{T}} \mathfrak{U}_Y(\beta)L_{2N^2}(\alpha-\beta)\,d\beta, \tag{C.3}$$

where $L_M(\alpha) = \sum_{n\le M} e(\alpha n)$ is the linear exponential sum.

Before we state the main proposition of this section, we set J to be the set $J = \{1, 2, 4, \ldots, 2^{D-1}, 2^D\}$ with $D \in \mathbb{N}$ between $\log_2 Q$ and $1 + \log_2 Q$. Decompose the function W and U^2 into

$$U^2(\alpha) = \sum_{Y\in J} \mathfrak{U}_Y(\alpha) + \mathfrak{U}'(\alpha) \quad \text{and} \quad W(\alpha) = \sum_{Y\in J} W_Y(\alpha) + W'(\alpha).$$

One can think of $\mathfrak{U}'$ as the minor-arc contribution which also contains the approximation error on the major arcs. Define ω' as the arithmetic function that belongs to W'.

Proposition. *For $Y \le 2Q$ we have the estimates*

$$\begin{aligned}&\int_{\mathbb{T}} |W_Y(\alpha)|^2\,d\alpha \ll N^2, &&\int_{\mathbb{T}} |W'(\alpha)|^2\,d\alpha \ll N^{2+\epsilon},\\ &\sup_\alpha |W_Y(\alpha)| \ll N^2Y^{-1}, &&\sup_\alpha |W'(\alpha)| \ll N^{2+\epsilon}Q^{-1}.\end{aligned}$$

For each $k \in \mathbb{N}$ with $Q^{8k} \le N$ we have

$$\sum_{m\le 2N^2} |\omega_Y(m)|^{2k} \ll_{\epsilon,k} Y^\epsilon N^2 \quad \textit{and} \quad \sum_{m\le 2N^2} |\omega'(m)|^{2k} \ll_{\epsilon,k} N^{2+\epsilon}.$$

Proof. We go through the estimates one by one. For the first one, we observe that by (C.3) the function W_Y is a projection of $\mathfrak{U}_Y$ and, therefore, by Bessel's inequality, we have the upper bound

$$\int_{\mathbb{T}} |W_Y(\alpha)|^2\,d\alpha \le \int_{\mathbb{T}} |\mathfrak{U}_Y(\beta)|^2\,d\beta,$$

which can also be established directly. Insert the definition of $\mathfrak{U}_Y$ from (C.2) and expand. We obtain due to the disjointness of the sets $\mathfrak{M}(q, a)$ the evaluation

$$\int_{\mathbb{T}} |\mathfrak{U}_Y(\beta)|^2\, d\beta = \sum_{Y\le q<2Y} q^{-4} \sum_{(a;q)=1} |U(q,a)|^4 \int_{\mathfrak{M}(q,a)} |v(\alpha - a/q)|^4\, d\alpha.$$

By the well known estimates $|U(q,a)| \ll q^{1/2}$ (see for example Lemma A.5 in [15]) and $|v(\beta)| \ll N(1+|\beta|N^2)^{-1/2}$ (see [23, Theorem 7.3]), we obtain the claim by a straightforward calculation.
The L^2-bound for W' follows from the previous bound since by Parseval's identity and $\omega(m) \ll_\epsilon m^\epsilon$ we get

$$\int_{\mathbb{T}} |W(\alpha)|^2\, d\alpha = \sum_{m\le 2N^2} \omega(m)^2 \ll N^{2+\epsilon}.$$

The L^∞-estimate for W_Y starts with (C.2) and (C.3) and gives us

$$\begin{aligned}|W_Y(\alpha)| &\le \sum_{Y\le q<2Y} q^{-2} \sum_{(a;q)=1} |U(q,a)|^2 \int_{\mathfrak{M}(q,a)} |v(\beta - a/q)|^2 |L_{2N^2}(\alpha-\beta)|\, d\beta \\ &\le Y^{-1} \sum_{Y\le q<2Y} \sum_{(a;q)=1} \int_{\mathfrak{M}(q,a)} |v(\beta - a/q)|^2 |L_{2N^2}(\alpha-\beta)|\, d\beta\end{aligned}$$

using again the estimate $|U(q,a)| \ll q^{1/2}$. For a given pair of q and a we can estimate the inner integral by Cauchy's inequality. The two resulting integrals can be dealt with the estimate $|v(\beta)| \ll N(1+|\beta|N^2)^{-1/2}$ and Parseval's identity to obtain the bound

$$\begin{aligned}&\int_{\mathfrak{M}(q,a)} |v(\beta - a/q)|^2 |L_{2N^2}(\alpha-\beta)|\, d\beta \\ \le &\Big(\int_{\mathfrak{M}(q,a)} |v(\beta - a/q)|^4\, d\beta\Big)^{1/2} \Big(\int_{\mathbb{T}} |L_{2N^2}(\alpha-\beta)|^2\, d\beta\Big)^{1/2} \ll N^2.\end{aligned}$$

This estimate is very wasteful if a/q is 'far' away from α and we use it only when $\|a/q-\alpha\| \le 2Q^4/N^2$. But there is only one such pair a and q since $\|a_i/q_i - \alpha\| \le 2Q^4/N^2$ for $i\in\{1,2\}$ implies that $1/q_1q_2 \le \|a_1/q_1 - a_2/q_2\| \le 4Q^4/N^2$, which isn't possible due to the restriction $16Q^6 \le N$.

For all the other values of a and q we can do better by using the bounds $|L(\alpha)| \le \|\alpha\|^{-1}$ and $|v(\beta)| \le N$. Since $\|a/q-\alpha\| > 2Q^4/N^2$ and $\|a/q-\beta\| \le Q/N^2$ we get $|L_{2N^2}(\alpha-\beta)| \le N^2/Q^4$. This implies

$$\int_{\mathfrak{M}(q,a)} |v(\beta - a/q)|^2 |L_{2N^2}(\alpha-\beta)|\, d\beta \ll \mu(\mathfrak{M}(q,a)) N^3/Q^4,$$

where μ is the Lebesgue measure. Summing over a and q and using the bound $\mu(\mathfrak{M}(q,a)) \le 2Q/N^2$, we obtain the L^∞-estimate for W_Y.

For the last part of the L^∞-estimates write

$$\mathfrak{U}^*(\alpha) = \sum_{Y\in J} \mathfrak{U}_Y(\alpha)$$

as an abbreviation. From $|L_{2N^2}(\alpha)| \leq \min\{2N^2, \|\alpha\|^{-1}\}$ we obtain

$$\int_{\mathbb{T}} |L_{2N^2}(\alpha)| \, d\alpha \ll \log N.$$

We use this observation together with (C.3) and the 'projection identity'

$$W(\alpha) = \int_{\mathbb{T}} U^2(\beta) L_{2N^2}(\alpha - \beta) \, d\beta$$

to reduce the L^∞-bound for W' to

$$|W'(\alpha)| \leq \int_{\mathbb{T}} |U^2(\beta) - \mathfrak{U}^*(\beta)||L_{2N^2}(\alpha - \beta)| \, d\beta \leq \log N \sup_{\beta \in \mathbb{T}} |U^2(\beta) - \mathfrak{U}^*(\beta)|.$$

Since $\mathfrak{U}^*$ is the major arc approximation for U^2 and zero outside of $\mathfrak{M}$, we can use Theorem 7.2 from [23] to estimate the approximation error $\Delta = U(\alpha) - q^{-1}U(q, a)v(\alpha - a/q)$ on the major arcs by $|\Delta| \ll q(1 + |\beta|N^2) \ll Q^2$. This is acceptable for our choice of Q. We estimate the size of U by Weyl's inequality (see [23, Lemma 2.4]) from above by

$$|U(\alpha)| \ll N^{1+\epsilon}(1/q + 1/N + q/N^2)^{1/2},$$

if $\alpha = a/q + \beta$ with $|\beta| \leq 1/q^2$. For $\alpha \notin \mathfrak{M}$ we have $q \geq Q$ and the estimate $|U(\alpha)| \ll N^{1+\epsilon} Q^{-1/2}$ as long as $Q \leq N$, which gives the desired estimate on the minor arcs.

The result for ω_Y is obtained from the decomposition

$$\omega_Y(m) = \sum_{Y \leq q < 2Y} q^{-2} \sum_{(a;q)=1} U(q, a)^2 e(-am/q) \int_{|\beta| \leq Q/N^2} v(\beta)^2 e(-\beta m) \, d\beta.$$

Since this factors into an analytic and an arithmetic part, we can use the Cauchy-Schwarz-inequality to estimate the $2k$-th moment over each part separately but with $4k$ instead of $2k$. The analytic part can be dealt with by the Hausdorff-Young inequality for $p = (1 - 1/(4k))^{-1}$ and contributes

$$\sum_{m \leq 2N^2} \Big| \int_{|\beta| \leq Q/N^2} v(\beta)^2 e(-\beta m) \, d\beta \Big|^{4k} \ll_k \Big(\int_{\mathbb{T}} |v(\beta)|^{2p} \, d\beta \Big)^{4k/p} \ll N^2.$$

On the other hand, the arithmetic moment can be bounded by [15, Lemma 5.3] and we obtain

$$\sum_{m \leq 2N^2} \Big| \sum_{Y \leq q < 2Y} q^{-2} \sum_{(a;q)=1} U(q, a)^2 e(-am/q) \Big|^{4k} \ll_{k,\epsilon} N^2 Y^\epsilon$$

as lond as $Y^{8k} \leq 2N^2$, which is satisfied, if Q is a sufficiently small power of N.

The result for $\omega' = \omega - \sum_{Y \in J} \omega_Y$ follows from by Hölder's inequality and the fact that the number of solutions to $m = x^2 + y^2$ is bounded by m^ϵ. □

Proof of Theorem C.2. We use the proposition above with Theorem C.3. We set $W_f = U_g^2$, $N_1 = 2N^2$, $J_0 = \{0\} \cup J$, where $W_0 = W'$, $\omega_0 = \omega'$ and

$r = p/2 > 2$. The first factor in

$$\|W_f\|_r \le \Big(\sum_{m\le 2N^2} |f(m)|^2\omega(m)\Big)^{1/2}\Big(\sum_{Y\in J_0} \|W_Y\|_r^{(r-2)/r}\|\omega_Y\|_{2r/(r-2)}^{2/r}\Big)^{1/2}$$

is $O(N)$ since $|f(m)| \le 1$ and $\omega(m) = \#\{n_1, n_2 \le N : m = n_1^2 + n_2^2\}$. The bound for the L^r-norm of W_Y for $Y \neq 0$ follows from

$$\int_0^1 |W_Y(\alpha)|^r \, d\alpha \le \sup_\alpha |W_Y(\alpha)|^{r-2}\int_0^1 |W_Y(\alpha)|^2 \, d\alpha \ll (N^2Y^{-1})^{r-2}N^2$$

by the first part of the proposition. For $Y = 0$ we obtain in the same way

$$\int_0^1 |W_0(\alpha)|^r \, d\alpha \ll (N^{2+\epsilon}Q^{-1})^{r-2}N^{2+\epsilon}.$$

The moment estimates for ω_Y give

$$\sum_{m\le 2N^2} |\omega_Y|^{2r/(r-2)} \ll N^2Y^\epsilon \quad \text{and} \quad \sum_{m\le 2N^2} |\omega_0|^{2r/(r-2)} \ll N^{2+\epsilon}.$$

If we parametrize J by $Y = 2^i$, we get for $D \approx \log_2 N$ the upper bound

$$\sum_{Y\in J_0} \|W_Y\|_r^{(r-2)/r}\|\omega_Y\|_{2r/(r-2)}^{2/r} \le \sum_{i\le D}\Big((N^2 2^{-i})^{r-2}N^2 \cdot 2^{\epsilon i}N^2\Big)^{(r-2)/r^2} + \\ + \Big((N^{2+\epsilon}Q^{-1})^{r-2}N^{2+\epsilon}\cdot N^{2+\epsilon}\Big)^{(r-2)/r^2}.$$

This is $O(N^{2(r-2)/r})$ if $r > 2$ and Q is a small power of N dependent on r. Combine the square-root of this with the first factor to get

$$\|W_f\|_r \ll N^{(2r-2)/r},$$

which gives the result when we take r-th powers and substitute $r = p/2$. □

We give a corollary of Theorem C.2 for the two-dimensional version.

Lemma C.4. *Let $|g| \le 1$, then for $p > 4$ we have*

$$\sup_\beta \int_{\mathbb{T}} |V_g(\alpha,\beta)|^p \, d\alpha \ll N^{p-2}.$$

Proof. Write

$$V_g(\alpha,\beta) = \sum_{n\le N} g(n)e(\alpha n^2 + \beta n) = \sum_{n\le N} g(n)e(\beta n)e(\alpha n^2) = U_h(\alpha)$$

for $h(n) = g(n)e(\beta n)$. The estimate now follows from Theorem C.2. □

References

[1] M. Aigner, *Combinatorial theory.* Springer-Verlag, Berlin-New York, 1979.

[2] J. Brüdern, R. Dietmann, J. Y. Liu and T. D. Wooley, *A Birch-Goldbach theorem.* Arch. Math. (Basel) 94 (2010), no. 1, 53–58

[3] J. Bourgain, *Fourier transform restriction phenomena for certain lattice subsets and applications to nonlinear evolution equations. I. Schrödinger equations.* Geom. Funct. Anal. 3 (1993), no. 2, 107–156.

[4] J. Bourgain, *Roth's theorem on progressions revisited.* J. Anal. Math. 104 (2008), 155–192.

[5] H. Davenport, *Analytic methods for Diophantine equations and Diophantine inequalities. Second edition.* Cambridge University Press, Cambridge, 2005.

[6] P. Erdős, P. Turan *On some sequences of integers.* J. Lond. Math. Soc. 11, 261–264 (1936).

[7] H. Furstenberg, Y. Katznelson and D. Ornstein, *The ergodic theoretical proof of Szemerédi's theorem.* Bull. Amer. Math. Soc. (N.S.) 7 (1982), no. 3, 527–552.

[8] T. Gowers, *A new proof of Szemerédi's theorem.* Geom. Funct. Anal. 11 (2001), no. 3, 465–588.

[9] B. Green, *Roth's theorem in the primes.* Ann. of Math. (2) 161 (2005), no. 3, 1609–1636.

[10] B. Green and T. Tao, *Restriction theory of the Selberg sieve, with applications.* J. Théor. Nombres Bordeaux 18 (2006), no. 1, 147–182.

[11] B. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions.* Ann. of Math. (2) 167 (2008), no. 2, 481–547.

[12] B. Green and T. Tao, *Linear equations in primes* Ann. of Math. (2) 171 (2010), no. 3, 1753–1850.

[13] D. R. Heath-Brown, *Integer sets containing no arithmetic progressions.* J. London Math. Soc. (2) 35 (1987), no. 3, 385–394.

[14] D. R. Heath-Brown, *A new form of the circle method, and its application to quadratic forms.* J. Reine Angew. Math. 481 (1996), 149–206.

[15] E. Keil, *On a diagonal quadric in dense variables.* Preprint.

[16] J. Liu, *Integral points on quadrics with prime coordinates.* Monatsh. Math. 164 (2011), no. 4, 439–465.

[17] L. Low, J. Pitman, A. Wolff. *Simultaneous diagonal congruences.* J. Number Theory 29 (1988), no. 1, 31–59.

[18] K. F. Roth, *On certain sets of integers.* J. London Math. Soc. 28, (1953), 104–109.

[19] K. F. Roth, *On certain sets of integers. II.* J. London Math. Soc. 29, (1954). 20–26.

[20] T. Sanders, *On Roth's theorem on progressions.* Ann. of Math. (2) 174 (2011), no. 1, 619–636.

[21] M. Smith, *On solution-free sets for simultaneous quadratic and linear equations.* J. Lond. Math. Soc. (2) 79 (2009), no. 2, 273–293.

[22] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression.* Acta Arith. 27 (1975), 199–245.

[23] R. C. Vaughan, *The Hardy-Littlewood method. Second edition.* Cambridge University Press, Cambridge, 1997.

Mathematical Institute, University of Oxford, Mathematical Institute, 24-29 St Giles', OX1 3LB Oxford, United Kingdom

E-mail address: `Eugen.Keil@maths.ox.ac.uk`