SOME REFINEMENTS FOR TRANSLATION INVARIANT QUADRATIC FORMS IN DENSE SETS

EUGEN KEIL

ABSTRACT. We improve the result of our previous paper on translation invariant quadratic forms in two special cases. We reduce the density bound $|\mathcal{A}|/N = O((\log \log N)^{-c})$ to $|\mathcal{A}|/N = O((\log N)^{-c})$ for most quadratic forms and handle almost diagonal equations in $s \geq 6$ variables instead of $s \geq 8$.

1. INTRODUCTION

This paper complements the author's work [3] on translation invariant quadratic forms and should be read subsequent to it. In [3] we have proven the following result.

Theorem 1.1. *Let $Q \in \mathbb{Z}^{s \times s}$ be symmetric with $Q \cdot \mathbf{1} = \mathbf{0}$ and off-rank r. Assume that $\mathbf{x}^T Q \mathbf{x} = 0$ 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 $\mathcal{A} \subset \{1, 2, \ldots, N\}$, then $|\mathcal{A}| \leq C_Q N (\log \log N)^{-c}$ for some $c, C_Q > 0$ and c independent of Q.*

The condition $Q \cdot \mathbf{1} = \mathbf{0}$ is equivalent to *translation invariance* of the equation $\mathbf{x}^T Q \mathbf{x} = 0$ as explained in [3]. The *'off-rank'* of a matrix is defined in Section 2.

The aim of this note is to give improvements on this theorem in two natural special cases. On the one hand, we can improve the bound $|\mathcal{A}| \leq C_Q N (\log \log N)^{-c}$ to $|\mathcal{A}| \leq C_Q N (\log N)^{-c}$ in the case $r \geq 5$ (which covers almost all quadratic forms) and on the other hand, solve the problem with as few as $s \geq 6$ variables in the almost diagonal situation $r = 1$. In the end, we want to discuss how the two approaches might be useful to improve on Theorem 1.1 for all quadratic forms.

The next section is devoted to explain the results, which are given in Theorem 2.2 and Theorem 2.4.

We introduce some standard notation. We use the Vinogradov notation $\ll$ and O-notation throughout the paper and indicate dependencies on parameters by subscripts, like in $O_{P,\epsilon}(N)$. The asymptotic parameter $N \in \mathbb{N}$ should be thought of as large and we work most of the time on the interval $[1, N] = \{1, 2, \ldots, N\}$. We write $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ for the 'circle' and identify it with $[0, 1]$ whenever convenient. As usual we abbreviate $e(x) = \exp(2\pi i x)$.

For a set $\mathcal{A} \subset [1, N]$ with the indicator function $1_{\mathcal{A}}$ and density $\delta = |\mathcal{A}|/N$ we define the 'balanced function' by

$$(1.1) \qquad\qquad f(n) = 1_{\mathcal{A}}(n) - \delta.$$

We are using bold face to denote vectors and the notation $\mathbf{x} \leq N$ means that each coordinate is bounded by N.

2. Statement of Results

First we summarize some of the definitions from [3] that are needed for the proofs below.

A quadratic form $Q(\mathbf{x}) = \mathbf{x}^T Q \mathbf{x}$ is translation invariant if $Q(\mathbf{x} + \mathbf{1}) = Q(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{Z}^s$. For a corresponding symmetric matrix $Q \in \mathbb{Z}^{s \times s}$, this translates into the condition $Q \cdot \mathbf{1} = \mathbf{0}$.

Definition 2.1 (Off-diagonal rank). *For a symmetric matrix $Q \in \mathbb{R}^{s \times s}$ we consider the set of matrices M such that $M = P^T Q P$ for a permutation matrix P. Such a matrix can be (non-uniquely) written as*

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

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

$$r = \max rank(B),$$

where the maximum is taken over all choices of P and decompositions of M. In other words, r is the maximal rank of a submatrix in Q, that doesn't contain any diagonal elements.

Now we can state the first theorem.

Theorem 2.2. *Let $Q \in \mathbb{Z}^{s \times s}$ be symmetric with $Q \cdot \mathbf{1} = \mathbf{0}$ and off-rank $r \geq 5$ (this implies that $s \geq 10$). Assume that it has a non-singular real solution to $\mathbf{x}^T Q \mathbf{x} = 0$, but only trivial solutions when the variables are restricted to $\mathcal{A} \subset [1, N]$. Then $|\mathcal{A}| \leq C_Q N (\log N)^{-c}$ for some $c, C_Q > 0$.*

The main idea for the proof can be summarized as follows. The usual density increment procedure depends on the fact that $e(\alpha p(x))$ for a polynomial p is constant on long arithmetic progressions. If α is close to a rational number with small denominator (on the 'major arcs'), the length of those progressions is comparable to N and we get a much better bound. It turns out that we can control the contribution of the 'minor arcs' (complement of the major arcs) in the proof of Theorem 2.2 in a uniform way independent of the structure of $\mathcal{A}$. This is done in Section 3. In Section 4 we perform

the density increment with α in the major arcs.

To motivate the second part of the paper, consider quadratic forms defined by the following matrices.

Example 2.3. *Consider the matrices*

$$\begin{pmatrix} 3 & 1 & 0 & 0 & 0 \\ 1 & 7 & -2 & -3 & 2 \\ 0 & -2 & 5 & 0 & 0 \\ 0 & -3 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 & -4 \end{pmatrix} \quad and \quad \begin{pmatrix} 0 & 2 & 2 & 2 & 1 \\ 2 & -1 & 4 & 4 & 2 \\ 2 & 4 & 6 & 4 & 2 \\ 2 & 4 & 4 & -5 & 2 \\ 1 & 2 & 2 & 2 & 3 \end{pmatrix}.$$

The first example is 'almost diagonal' in the sense that off-diagonal terms are all concentrated in the second column/row. In what sense is the second example almost diagonal? We can write the second matrix as a sum of a diagonal matrix D and a rank one perturbation R, a matrix of the form $R = \mathbf{v} \cdot \mathbf{v}^T$ for the vector $\mathbf{v} = (1, 2, 2, 2, 1)^T$. While the first matrix is an example of a small 'local' perturbation, the second is an instance of a diagonal matrix with a small 'global' perturbation.

One can easily check that the two matrices have off-rank one. As we will see in Section 5, it turns out that those two cases exhaust the possible ways that a symmetric matrix can have off-rank one. Armed with this classification, we can show that Theorem 1.1 holds for $s \geq 6$ (instead of $s \geq 5 + 3r = 8$) if the matrix underlying our quadratic form has off-rank one. The result is given in the following theorem.

Theorem 2.4. *Let $Q \in \mathbb{Z}^{s \times s}$ be symmetric with $Q \cdot \mathbf{1} = \mathbf{0}$ and off-rank $r = 1$. Assume that $s \geq 6$, and that there is a non-singular real solution to $\mathbf{x}^T Q \mathbf{x} = 0$. If there are only trivial solutions when the variables are restricted to $\mathcal{A} \subset [1, N]$, then $|\mathcal{A}| \ll_Q N (\log \log N)^{-1/15}$.*

Let us analyse the two cases in more detail. What does it mean that in the first example all off-diagonal terms are concentrated in the row/column of one variable? Let us assume for simplicity that the variable is x_s. We will see that by completing squares, we can remove all off-diagonal terms and replace each variable x_i by $x_i - x_s$ (translation invariance!), reducing the problem to a diagonal form in five variables, which is well within reach of the classical approach. We deal with this case in Section 6.

In Section 7 we reduce the second case to the following diagonal system considered by Smith [5] and the author [2].

$$\begin{aligned} d_1 x_1^2 + d_2 x_2^2 + \ldots + d_s x_s^2 &= 0, \\ d_1 x_1 + d_2 x_2 + \ldots + d_s x_s &= 0. \end{aligned} \tag{2.1}$$

The system (2.1) can be handled under the following conditions.

Conditions for system (2.1).
(*i*) $d_1 + d_2 + \ldots + d_s = 0$,
(*ii*) $s \geq 7$ and $d_i \neq 0$ for all $1 \leq i \leq s$,
(*iii*) there are at least two positive and at least two negative coefficients d_i.

Condition (i) encodes the translation invariance of the system. The bound $s \geq 7$ in condition (ii) will correspond to the condition $s \geq 6$ in Theorem 2.4. We assume that $d_i \neq 0$ since we can easily locate non-trivial solutions otherwise. Condition (iii) is needed for the existence of a non-singular real solution (see [2] for details).

In Section 7 we use the following theorem from [2] to deduce Theorem 2.4 in the second case.

Theorem 2.5. *Assume that the conditions above hold and system* (2.1) *has only trivial solutions for* $x_i \in \mathcal{A} \subset [1, N]$. *Then* $|\mathcal{A}| \leq CN (\log \log N)^{-1/15}$ *for some constant* C, *which depends only on the coefficients* d_i *of the system.*

3. CONTROLLING THE MINOR ARCS

As in [3], the main ingredient in the proof is a bilinear sum estimate for our exponential sum

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

Write $\alpha = a/q + \beta$ for a diophantine approximation with $q \leq N, |\beta| \leq (qN)^{-1}$ and

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

Then we can bound $S_g(\alpha)$ as in [3] for $|g| \leq 1$ by

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

As in [3, Theorem 3], we can refine the estimate if we are given L^1-bounds $\sum_{n \leq N} |g(n)| = O(\delta N)$. As in [3], we get the bound

$$(3.2) \qquad\qquad |S_g(\alpha)| \ll \delta^{s-10} N^{s-5} K(\alpha)^5.$$

This pointwise bound allows us to deduce a sharp L^p-estimate for $S_g(\alpha)$, which is necessary for the density increment strategy to work.

If we look at the proof of this L^p bound in [3], we see that the condition we need for the exponent of $K(\alpha)$ is that it is bigger than four. This means that we can pull out a small power of $K(\alpha)$, and improve the estimate on the minor arcs, where $K(\alpha)$ is small.

For an absolute constant $D > 1$ to be chosen later, define the major arcs for $q \leq D^4 \delta^{-40}$ and $(a; q) = 1$ by

$$(3.3) \qquad \mathfrak{M}(q, a) = \{ \alpha \in \mathbb{T} : \|\alpha - a/q\| \leq D^8 \delta^{-80} N^{-2} \}$$

(disjoint for $N \geq D^8 \delta^{-80}$) and set $\mathfrak{M}$ to be the union of all those sets. The minor arcs $\mathfrak{m} = \mathbb{T} \backslash \mathfrak{M}$ is the complement. The key point to notice is that the constants in the definition of $\mathfrak{M}$ only depend on δ and an absolute constant D. The precise numbers are less important.

By Dirichlet's approximation theorem, we have that every $\alpha \in \mathbb{T}$ is contained in at least one ball defined by $\|\alpha - a/q\| \leq (qN)^{-1}$ for some

$q \leq N$. This means that we have two types of minor arcs. Those Dirichlet neighbourhoods with $q > D^4\delta^{-40}$ and those with $q \leq D^4\delta^{-40}$ but $D^8\delta^{-80}N^{-2} < \|\alpha - a/q\| \leq (qN)^{-1}$.

If $q > D^4\delta^{-40}$, we have

$$K(\alpha) \ll (N \log q)^{1/2} + \frac{N}{q^{1/2}} \ll D^{-2}\delta^{20}N.$$

for $N \gg D^8\delta^{-80}$.

In the second case $q \leq D^4\delta^{-40}$, we split the two intervals $D^8(N^2\delta^{80})^{-1} < |\beta| \leq (qN)^{-1}$ into dyadic bits $|\beta| \in (2^i N^{-2}, 2^{i+1}N^{-2}]$ for $\log_2(\delta^{-80}D^8) \leq i \leq \log_2(Nq^{-1})$. For fixed i we have

$$K(\alpha) \ll (N \log q)^{1/2} + \left(N^2 \frac{i+1}{q2^i} \right)^{1/2}$$
$$\ll (N \log q)^{1/2} + N2^{-i/4} \ll D^{-2}\delta^{20}N$$

for $N \gg D^8\delta^{-80}$.

Now that we have good pointwise bounds, we cite [3, Lemma 2], which gives us a control of $K(\alpha)$ on average. For $p > 4$ we have

$$\int_0^1 |K(\alpha)|^p \, d\alpha \ll N^{p-2}.$$

We apply the two bounds with (3.2) and get

$$\int_{\mathfrak{m}} |S_g(\alpha)| \, d\alpha \ll (\delta^{s-10}N^{s-5}) \sup_{\alpha \in \mathfrak{m}} |K(\alpha)|^{1/2} \int_{\mathfrak{m}} |K(\alpha)|^{9/2} \, d\alpha$$
$$\ll (\delta^{s-10}N^{s-5})(D^{-1}\delta^{10}N^{1/2})N^{5/2} \ll D^{-1}\delta^s N^{s-2}.$$

We can use this bound in the argument in [3, Section 3]. We know that

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

for $g(\mathbf{x}) = 1_{\mathcal{A}^s}(\mathbf{x}) - \delta^s$ with $\delta = |\mathcal{A}|/N$ as long as $N \gg_Q \delta^{-2}$. By choosing D sufficiently small, this implies that

$$\int_{\mathfrak{M}} |S_g(\alpha)| \, d\alpha \gg \delta^s N^{s-2}.$$

By the same argument as in [3, Section 3] we get the bound

(3.4) $$\sup_{\alpha \in \mathfrak{M}} |S_{f_i}(\alpha)| \gg \delta^{s+80} N^s$$

for a function f_i that is a (tensor) product of functions f, δ and $1_{\mathcal{A}}$ (see (1.1)).

4. DENSITY INCREMENT ON MAJOR ARCS

Consider the density increment argument from [3, Section 7]. Instead of a general lower bound for the exponential sum, we have (3.4), where α lies in the major arcs $\mathfrak{M}$ defined in (3.3). This implies that there is a $q \leq D^4\delta^{-40}$

and $1 \leq a \leq q$ such that $\|\alpha - a/q\| \leq D^8\delta^{-80}N^{-2}$. For the following argument, D is a fixed constant and will be absorbed in the Vinogradov notation.

We decompose our interval $[1, N]$ into progressions modulo q. On each progression we have $e(\alpha Q(\mathbf{x})) = e(\beta Q(\mathbf{x}))e(aQ(\mathbf{x})/q)$, where $\alpha = a/q + \beta$. Since β is very small, we can estimate

$$|e(\beta Q(\mathbf{x})) - e(\beta Q(\mathbf{y}))| \leq 2\pi|\beta||Q(\mathbf{x}) - Q(\mathbf{y})| \ll (N^2\delta^{80})^{-1}|Q(\mathbf{x}) - Q(\mathbf{y})|.$$

The second factor is constant on progressions with difference q. If we write $\mathbf{x} = q\mathbf{m} + \mathbf{h}$ and $\mathbf{y} = q\mathbf{n} + \mathbf{h}$, we obtain

$$|e(\alpha Q(\mathbf{x})) - e(\alpha Q(\mathbf{y}))| = |e(\beta Q(\mathbf{x}))e(aQ(\mathbf{h})/q) - e(\beta Q(\mathbf{y}))e(aQ(\mathbf{h})/q)|$$

$$= |e(\beta Q(\mathbf{x})) - e(\beta Q(\mathbf{y}))| \ll (N^2\delta^{80})^{-1}|Q(\mathbf{x}) - Q(\mathbf{y})|$$

$$\ll (N^2\delta^{80})^{-1}q^2 N\|\mathbf{m} - \mathbf{n}\|_\infty \ll N^{-1}\delta^{-160}\|\mathbf{m} - \mathbf{n}\|_\infty.$$

If we cut our q-progressions $P_\mathbf{h}$ into subprogressions $P_{\mathbf{h},\mathbf{j}}$ of side length approximately $P = \eta\delta^{240}N$ for some small fixed η, our function $e(\alpha Q(\mathbf{x}))$ is constant up to a small error term of size $O(\delta^{80})$. This implies

$$\sum_{\mathbf{x} \leq N} f_i(\mathbf{x})e(\alpha Q(\mathbf{x})) = \sum_{\mathbf{h} \leq q} \sum_{\mathbf{j} \leq (q\eta\delta^{240})^{-1}} \sum_{\mathbf{x} \in P_{\mathbf{h},\mathbf{j}}} f_i(\mathbf{x})e(\alpha Q(\mathbf{x}))$$

$$= \sum_{\mathbf{h} \leq q} \sum_{\mathbf{j} \leq (q\eta\delta^{240})^{-1}} \sum_{\mathbf{x} \in P_{\mathbf{h},\mathbf{j}}} f_i(\mathbf{x})(c(\alpha, \mathbf{h}, \mathbf{j}) + \epsilon(\alpha, \mathbf{x})\delta^{80})$$

for some functions $|c(\alpha, \mathbf{h}, \mathbf{j})| \leq 1$ and $|\epsilon(\alpha, \mathbf{x})| < 1/2$. Taking absolute values and using 3.4 with $\sum_\mathbf{x}|f_i(\mathbf{x})| \ll \delta^s N^s$ gives

$$\sum_{\mathbf{h} \leq q} \sum_{\mathbf{j} \leq (q\eta\delta^{240})^{-1}} \left|\sum_{\mathbf{x} \in P_{\mathbf{h},\mathbf{j}}} f_i(\mathbf{x})\right| \gg \delta^{s+80}N^s.$$

For at least one value of $\mathbf{h}$ and $\mathbf{j}$, we get

$$\left|\sum_{\mathbf{x} \in P_{\mathbf{h},\mathbf{j}}} f_i(\mathbf{x})\right| \gg \delta^{240s+s+80}N^s.$$

Since f_i is a product of $f, 1_\mathcal{A}$ and δ the sum on the left hand side splits into s independent sums. We estimate all factors trivially apart from one of those that contain the balanced function f (which exists by construction of f_i). Write $P_{\mathbf{h},\mathbf{j}} = \mathbf{h} + qP_\mathbf{j}$ with $P_\mathbf{j} = P_{j_1} \times \ldots \times P_{j_s}$, then we have for some $k \leq s$ the inequality

$$(\delta P)^{s-1}\left|\sum_{x_k \in h_k + qP_{j_k}} f(x_k)\right| \gg \delta^{240s+s+80}N^s \gg \delta^{s+80}|P_{j_k}|^s.$$

This implies a density increment of size $\delta \to \delta + \theta\delta^{81}$ for a small $\theta > 0$ and a loss of progression length of $N \to \eta\delta^{240}N$. Now we perform a density increment argument, the details of which can be found in [3, Section 7].

The number of steps of this iteration is bounded by $\theta^{-1}\delta^{-81}$ and we end up with the condition $(\eta\delta^{240})^{\theta^{-1}\delta^{-81}}N \ll C_Q D^8\delta^{-80}$ for some constant C_Q, which depends on the smallest non-trivial solution of Q in $\mathbb{Z}$ (for a discussion

of this, see [3, Section 2]). Taking logarithms and rearranging for δ gives us $\delta \ll (\log N)^{-1/82}$, for example.

5. STRUCTURE OF QUADRATIC FORMS WITH OFF-RANK ONE

Now we turn our attention towards the second topic of this paper, the off-rank one situation. Given a symmetric matrix with off-rank one, what can we say about its structure? The following lemma provides a complete answer to this question. Write $\mathbf{v} \otimes \mathbf{w} := \mathbf{v} \cdot \mathbf{w}^T$ for $\mathbf{v}, \mathbf{w} \in \mathbb{R}^d$.

Lemma 5.1. *Let $Q \in \mathbb{Z}^{s \times s}$ be a symmetric matrix with off-rank $r = 1$. Then either*

(i) $Q = D + \mathbf{v} \otimes \mathbf{e}_j + \mathbf{e}_j \otimes \mathbf{v}$, *where $\mathbf{v} \in \mathbb{Z}^s$, $\mathbf{e}_j$ is the standard basis vector for some $1 \leq j \leq s$ and $D \in \mathbb{Z}^{s \times s}$ is diagonal, or*

(ii) $Q = m^{-1}(D + \mathbf{v} \otimes \mathbf{v})$ *with a diagonal quadric $D \in \mathbb{Z}^{s \times s}$, some vector $\mathbf{v} \in \mathbb{Z}^s$ and $m \in \mathbb{Z} \backslash \{0\}$.*

Proof. By permutating variables, if necessary, we can assume that Q has the form

$$Q = \begin{pmatrix} g & u & \mathbf{a}^T \\ u & h & \mathbf{b}^T \\ \mathbf{a} & \mathbf{b} & C \end{pmatrix}$$

for some matrix $C \in \mathbb{Z}^{(s-2) \times (s-2)}$ and $u \neq 0$. Now consider for some $i \neq j$ the off-diagonal submatrix

$$\begin{pmatrix} u & a_j \\ b_i & c_{ij} \end{pmatrix}.$$

For this to have rank at most one, we must have $uc_{ij} = a_j b_i$. If $\mathbf{a} = 0$ or $\mathbf{b} = 0$ this implies that $c_{ij} = 0$ for all $i \neq j$, which corresponds to case (i). Otherwise, we can assume that $\mathbf{a} \neq 0$ and $\mathbf{b} \neq 0$. Again by the off-rank property we deduce that the vectors $\mathbf{a}$ and $\mathbf{b}$ are linearly dependent and we have $\mathbf{b} = \lambda \mathbf{a}$ for some $\lambda \in \mathbb{Q} \backslash \{0\}$. By using $uc_{ij} = a_j b_i$ we deduce that $C = u^{-1} \mathbf{b} \otimes \mathbf{a} + E$ for some diagonal matrix E. It remains to 'lift' this information to the matrix Q.

If we consider the structure of Q modulo diagonal matrices, we have to show that one can choose diagonal entries x and y in such a way that the matrix

$$P = \begin{pmatrix} x & u & \mathbf{a}^T \\ u & y & \lambda \mathbf{a}^T \\ \mathbf{a} & \lambda \mathbf{a} & u^{-1} \lambda \mathbf{a} \otimes \mathbf{a} \end{pmatrix},$$

has the form $m^{-1} \mathbf{v} \otimes \mathbf{v}$ for some $m \in \mathbb{Z} \backslash \{0\}$ and $\mathbf{v} \in \mathbb{Z}^s$. There isn't any choice but to complete P to

$$\begin{pmatrix} u\lambda^{-1} & u & \mathbf{a}^T \\ u & \lambda u & \lambda \mathbf{a}^T \\ \mathbf{a} & \lambda \mathbf{a} & u^{-1} \lambda \mathbf{a} \otimes \mathbf{a} \end{pmatrix}$$

and check that this matrix has indeed rank one. To be able to write this as a tensor product of a vector $\mathbf{v}$ with itself, we multiply with the common

factor $m = k^2 \lambda u$. A suitable $k \in \mathbb{N}$ makes all entries to integers and allows us to choose $\mathbf{v}^T = k\,(u, \lambda u, \lambda \mathbf{a}^T)$. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

6. The Parameter Method

In this section we are going to analyse case (i) in Lemma 5.1. By changing notation, we can assume that the 'non-diagonal' variable is x_s and by completing the square we can write the quadratic equation $Q(\mathbf{x}) = 0$ in the form

$$a_1(b_1 x_1 - c_1 x_s)^2 + \ldots + a_{s-1}(b_{s-1} x_{s-1} - c_{s-1} x_s)^2 + a_s x_s^2 = 0$$

for some $a_i, b_i, c_i \in \mathbb{Z}$. We can assume that $b_i \neq 0$ for all $1 \leq i \leq s - 1$.

Now we can simplify this equation by incorporating the restriction of translation invariance. If we replace each x_i by $x_i + 1$ we obtain the same quadratic form with an additional linear term, which is two times

$$a_1(b_1 x_1 - c_1 x_s)(b_1 - c_1) + \ldots + a_{s-1}(b_{s-1} x_{s-1} - c_{s-1} x_s)(b_{s-1} - c_{s-1}) + a_s x_s,$$

as well as the constant term

$$a_1(b_1 - c_1)^2 + \ldots + a_{s-1}(b_s - c_s)^2 + a_s.$$

The linear term must be zero for all $\mathbf{x}$, which implies that $b_i = c_i$ for all i with $a_i \neq 0$. Where $a_i = 0$ we can simply assume that $b_i = c_i$ without changing the quadratic form. Furthermore this implies that $a_s = 0$. These conditions are sufficient for the vanishing of the constant term as well.

If we set $d_i = a_i b_i^2$ we end up with the almost diagonal form

(6.1) $$d_1(x_1 - x_s)^2 + \ldots + d_{s-1}(x_{s-1} - x_s)^2 = 0.$$

If $d_i = 0$ for some i, we can easily find a non-trivial solution in $\mathcal{A}$. Therefore, we can assume that all $d_i \neq 0$.

We view x_s as a free 'parameter' ranging over the set $\mathcal{A}$ and focus on the remaining $s - 1$ variables. We define the shifted exponential sum by

(6.2) $$U_{\mathcal{A}-y}(\alpha) = \sum_{x \in \mathcal{A}} e(\alpha(x - y)^2).$$

The number of solutions to (6.1) can be written as the Fourier integral

$$\sum_{x_s \in \mathcal{A}} \int_0^1 \prod_{i=1}^{s-1} U_{\mathcal{A}-x_s}(d_i \alpha)\, d\alpha.$$

We compare this to the situation, where $x_1, \ldots, x_{s-1} \in [1, N]$. The number of solutions in this case is given by

$$\sum_{x_s \in \mathcal{A}} \int_0^1 \prod_{i=1}^{s-1} U_{[1,N]-x_s}(d_i \alpha)\, d\alpha.$$

The inner integral is bounded from below by

(6.3) $$\int_0^1 \prod_{i=1}^{s-1} U_{[1,N/2]}(d_i \alpha)\, d\alpha$$

since the set $[1, N] - x_s$ covers either $[1, N/2]$ or $[-N/2, -1]$ completely and equation (6.1) has even degree.

As long as $s - 1 \geq 5$ and not all d_i have the same sign, the classical circle method gives us many integer solutions for a diagonal quadratic equation. (See the book of Davenport [1], for example.) The case, where all d_i have the same sign is excluded by the assumption that the quadric has a non-singular real solution. This implies a lower bound of size N^{s-3} for (6.3) and we obtain

$$\sum_{x_s \in \mathcal{A}} \int_0^1 \prod_{i=1}^{s-1} U_{[1,N]-x_s}(d_i \alpha) \, d\alpha \gg \delta N \cdot N^{s-3}.$$

Since there are only δN trivial solutions of (6.1) in $\mathcal{A}$ by assumption, we get

$$(6.4) \quad \sum_{x_s \in \mathcal{A}} \int_0^1 \left| \prod_{i=1}^{s-1} U_{\mathcal{A}-x_s}(d_i \alpha) - \prod_{i=1}^{s-1} \delta U_{[1,N]-x_s}(d_i \alpha) \right| d\alpha \gg \delta^s N^{s-2} - \delta N.$$

To deduce a correlation estimate, we replace the indicator function $1_{\mathcal{A}}$ in (6.2) by $1_{\mathcal{A}} = \delta 1_{[1,N]} + f$ with the balanced function f from (1.1). An expansion of the first product in (6.4) creates 2^{s-1} terms, where the first one is cancelled by the second product. The remaining contributions can be bounded from above by a finite sum of terms of the form

$$\sum_{x_s \in \mathcal{A}} \int_0^1 \prod_{i=1}^{s-1} |U_{T(x_s)g_i}(d_i \alpha)| \, d\alpha$$

with $(T_y g)(x) = g(x - y)$ and at least one of the g_i equal to f. By Hölder's inequality, we can bound this expression by

$$\sup_{x_s, \alpha} |U_{T_{x_s}f}(\alpha)|^{1/2} \sum_{x_s \in \mathcal{A}} \int_0^1 |U_g(\alpha)|^{s-3/2} \, d\alpha$$

for some function $|g| \leq 1$ defined on $[-N, N]$ with $\sum_{|n| \leq N} |g(n)| \leq 2\delta N$. This implies $|U_g(\alpha)| \leq 2\delta N$ and the upper bound

$$(2\delta N)^{s-6} \sup_{x_s, \alpha} |U_{T_{x_s}f}(\alpha)|^{1/2} \sum_{x_s \in \mathcal{A}} \int_0^1 |U_g(\alpha)|^{9/2} \, d\alpha.$$

We get the bound $O(N^{5/2})$ for the integral from [3, Theorem 6]. This leads to the estimate

$$(\delta N)^{s-5} \sup_{x_s, \alpha} |U_{T_{x_s}f}(\alpha)|^{1/2} N^{5/2} \gg \delta^s N^{s-2} - \delta N.$$

As long as $N \gg_Q \delta^{-2}$ we obtain the correlation estimate

$$\left| \sum_x f(x) e(\alpha(x - x_s)^2) \right| \gg \delta^{10} N$$

for some $x_s \in [1, N]$ and $\alpha \in \mathbb{T}$. By expanding the square, one can see that the left hand side is just a quadratic exponential sum. The density increment procedure in [3, Section 7] gives us the bound $\delta \ll (\log \log N)^{-1/11}$.

7. The Diagonal System

We are given a translation invariant quadratic form $\mathbf{x}^T Q \mathbf{x} = 0$ in $s \geq 6$ variables with off-rank one, which satisfies case (ii) in Lemma 5.1.

We can write the equation $Q(\mathbf{x}) = 0$ in the form

$$(7.1) \qquad d_1 x_1^2 + \ldots + d_s x_s^2 + (v_1 x_1 + \ldots + v_s x_s)^2 = 0,$$

where the coefficients d_i are equal to the diagonal elements in the matrix D from Lemma 5.1. Consider $Q(\mathbf{x} + \mathbf{1}) = Q(\mathbf{x}) + 2L(\mathbf{x}) + Q(\mathbf{1})$, where $L(\mathbf{x})$ is given by

$$d_1 x_1 + \ldots + d_s x_s + (v_1 x_1 + \ldots + v_s x_s)(v_1 + \ldots + v_s)$$

and the constant term $Q(\mathbf{1})$ is

$$(d_1 + \ldots + d_s) + (v_1 + \ldots + v_s)^2.$$

By translation invariance, the linear and constant terms have to disappear for all $\mathbf{x} \in \mathbb{Z}$. This can only happen if for $n = v_1 + \ldots + v_s$ we have $d_i = -n v_i$ and $d_1 + \ldots + d_s = -n^2$. In the special case $n = 0$ we end up with the linear equation

$$v_1 x_1 + \ldots + v_s x_s = 0,$$

which is covered by the method of Roth [4] and gives a better bound than needed for our theorem here. We assume from now on that $n \neq 0$ and write equation (7.1) in the form of a system. For a new variable $h = v_1 x_1 + \ldots + v_s x_s$ with $h \in \mathbb{Z}$ we obtain (after multiplication of the linear equation by n)

$$(7.2) \qquad \begin{aligned} d_1 x_1^2 + \ldots + d_s x_s^2 + h^2 &= 0, \\ d_1 x_1 + \ldots + d_s x_s + nh &= 0, \end{aligned}$$

with the condition $d_1 + \ldots + d_s = -n^2$. Now we (arbitrarily) restrict h to $n\mathbb{Z}$, write $h = n x_0$ and $d_0 = n^2$. Then system (7.2) reduces further into

$$\begin{aligned} d_0 x_0^2 + d_1 x_1^2 + \ldots + d_s x_s^2 &= 0, \\ d_0 x_0 + d_1 x_1 + \ldots + d_s x_s &= 0, \end{aligned}$$

with $d_0 + d_1 + \ldots + d_s = 0$.

Now Theorem 2.5 implies the result for the weaker system (7.2). One should note at this point that conditions (ii) and (iii) for Theorem 2.5 are implied by the assumptions in Theorem 2.4 or can be assumed to be true without loss of generality.

8. Remarks and Open Problems

Looking at the cases $r = 1$ and $r = 5$, where we can deal with equations in $5 + r$ variables, one might have a vague hope to be able to extend this result to off-ranks $2 \leq r \leq 4$. If this were the case, Theorem 1.1 would require only $s \geq 10$ variables instead of $s \geq 17$. (Recent work of Zhao [6] could probably reduce the bound even further to $s \geq 9$.)

To prove such a result, we need a structure theorem along the lines of Lemma 5.1 for quadratic forms with off-rank $2 \leq r \leq 4$. Let us look at the next simplest case $r = 2$.

When $r = 1$ we have two cases. Slightly simplified, they correspond to the decompositions $M = D + \mathbf{v} \otimes \mathbf{v}$ and $M = D + \mathbf{e}_s \otimes \mathbf{v} + \mathbf{v} \otimes \mathbf{e}_s$ for a diagonal matrix D. Combining the two ideas, we end up with three different structures for $r = 2$:

(i) $Q = D + \mathbf{v} \otimes \mathbf{v} + \mathbf{w} \otimes \mathbf{w}$,

(ii) $Q = D + \mathbf{v} \otimes \mathbf{v} + \mathbf{e}_s \otimes \mathbf{w} + \mathbf{w} \otimes \mathbf{e}_s$,

(iii) $Q = D + \mathbf{e}_{s-1} \otimes \mathbf{v} + \mathbf{v} \otimes \mathbf{e}_{s-1} + \mathbf{e}_s \otimes \mathbf{w} + \mathbf{w} \otimes \mathbf{e}_s$.

It is easy to check, that in all three cases we have indeed a matrix with off-rank at most two. Sadly this naïve idea doesn't work and they don't cover all possible cases of matrices with off-rank $r = 2$. Consider the matrix

$$\begin{pmatrix} * & 1 & 1 & 0 & 1 \\ 1 & * & 0 & 1 & 1 \\ 1 & 0 & * & 1 & 0 \\ 0 & 1 & 1 & * & -1 \\ 1 & 1 & 0 & -1 & * \end{pmatrix},$$

where stars mark arbitrary entries.

For it to be of type (i), we would need to find diagonal entries such that the resulting matrix has rank two. By choosing suitable 3×3 matrices with only one entry missing, we can fill in the diagonal entries easily and check that this cannot be done consistently.

It is also easy to see that option (iii) is not correct since there are non-zero entries in more than two rows. For option (ii), we have to show that any 4×4 submatrix, which results by deleting a row and corresponding column cannot be completed to have rank one. This follows directly from the existence of off-diagonal 2×2 matrices of full rank in each of those cases.

It is an interesting problem, whether this a is pathological counterexample that can be understood by adding a case (iv) to the above list or whether symmetric off-rank two matrices don't have a simple classification.

Even though a complete classification seems a non-trivial task, it is likely, that the ideas of this paper can be used to improve slightly on the variable bounds for $r \in \{2, 3, 4\}$ and, therefore, potentially reduce the overall bound of Theorem 1.1 from $s \geq 17$ to $s \geq 16$, for example.

Concerning the density bounds, it isn't hard to see that the proof given for Theorem 2.2 does generalize to any situation in [3], where we use the function $K(\alpha)$ from (3.1) to bound the L^p-norm of our exponential sum. This would take care of almost all quadratic forms with off-rank $1 \leq r \leq 4$ as well. The only cases, where this is not possible correspond to Section 8 in [3], where we reduce the problem to a linear system in four equations. Further advances in the linear theory could provide bounds of the form $(\log N)^{-c}$ for all quadratic forms in sufficiently many variables.

References

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

[2] E. Keil, *On a diagonal quadric in dense variables,* arXiv:1306.4524.

[3] E. Keil, *Translation invariant quadratic forms in dense sets,* arXiv:1308.6680.

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

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

[6] L. Zhao, *The quadratic form in 9 prime variables,* arXiv:1402.3697.

Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, United Kingdom

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