NEW BOUNDS FOR THE DISCRETE FOURIER RESTRICTION TO THE SPHERE IN FOUR AND FIVE DIMENSIONS

JEAN BOURGAIN AND CIPRIAN DEMETER

Abstract. We improve the range for the discrete Fourier restriction to the four and five dimensional spheres. We rely on two new ingredients, incidence theory and Siegel’s mass formula.

1. Introduction

Let $n \geq 2$ and $\lambda \geq 1$ be two integers. Define $N = \left\lfloor \lambda^{1/2} \right\rfloor + 1$ and

$$\mathcal{F}_{n,\lambda} = \{ \xi = (\xi_1, \ldots, \xi_n) \in \mathbb{Z}^n : |\xi_1|^2 + \cdots + |\xi_n|^2 = \lambda \}.$$ We will use the notation $e(z) = e^{iz}$. Recall the following conjecture from [3], about the eigenfunctions of the Laplacian on the torus.

Conjecture 1.1. For each $n \geq 3$, $a_\xi \in \mathbb{C}$, $\epsilon > 0$ and each $p \geq \frac{2n}{n-2}$ we have

$$\| \sum_{\xi \in \mathcal{F}_{n,\lambda}} a_\xi e(\xi \cdot x) \|_{L^p(T^n)} \lesssim \epsilon N^{\frac{2n-2}{p} + \frac{2n}{p} + \epsilon} \| a_\xi \|_{l^2(\mathcal{F}_{n,\lambda})}.$$ This can be thought of as a discrete version of the Thomas-Stein restriction theorem. We refer the reader to [5] and [7] for the necessary background.

Here we make progress when $n = 4$ and $n = 5$.

Theorem 1.2. (i) For each $a_\xi \in \mathbb{C}$, $\epsilon > 0$ and $p > \frac{44}{7}$ we have

$$\| \sum_{\xi \in \mathcal{F}_{4,\lambda}} a_\xi e(\xi \cdot x) \|_{L^p(T^4)} \lesssim \epsilon N^{\frac{4}{p} + \epsilon} \| a_\xi \|_{l^2(\mathcal{F}_{4,\lambda})}.$$ (ii) For each $a_\xi \in \mathbb{C}$, $\epsilon > 0$ and $p > \frac{14}{3}$ we have

$$\| \sum_{\xi \in \mathcal{F}_{5,\lambda}} a_\xi e(\xi \cdot x) \|_{L^p(T^5)} \lesssim \epsilon N^{\frac{5}{p} + \epsilon} \| a_\xi \|_{l^2(\mathcal{F}_{5,\lambda})}.$$ This improves the result in [7] where the conjecture was verified for $p > 8$ when $n = 4$ and for $p > 5$ when $n = 5$. The result from [7] relied essentially on two ingredients. One is the sharp supercritical estimate ($p > \frac{2(n+1)}{n-3}$) from [3] proved by combining the circle method with the Thomas-Stein argument. See Proposition 6.1 below. The second one is the sharp subcritical estimate ($p = \frac{2n}{n-1}$) from [6], whose proof did not rely at all on number theory.

Here we replace that subcritical estimate by a new $L^4$ estimate. While this $L^4$ estimate is not sharp, it is strong enough to improve the range in the conjecture. Note that the

---

The first author is supported by the NSF grant DMS 1301619. The second author is supported by a Sloan Research Fellowship and by the NSF Grant DMS-1161752.

AMS subject classification: Primary 11L03; Secondary 42A16, 42A25, 52C35.
index $p = 4$ is critical when $n = 4$ and supercritical when $n \geq 5$. Thus, when $n = 4$, $L^4$ is precisely the right space to consider; getting the sharp $L^4$ estimate would completely solve Conjecture 1.1. On the other hand, the $L^4$ approach we develop is only useful for $n = 4, 5$ since the result in [7] already proved the sharp $L^4$ bound in dimensions $n \geq 6$.

To derive the $L^4$ estimate we rely on two new methods: incidence theory and Siegel’s mass formula. Interestingly, the application of both methods is rather sharp, see Remarks 3.2, 3.5, 5.2 and 5.4. We mention that the use of incidence theory, while new in the context of Conjecture 1.1, has been in the last twenty years or so one of the important tools in various other problems with restriction theory flavor. It suffices to mention [16] and the more recent [13], [2].

We describe the incidence theory approach in sections 2 and 3 while the number theoretical approach appears in sections 4 and 5. These tools are then combined in section 6 to prove our main theorem. In the last section we speculate on possible ways to further improve our result.

The first author would like to thank Peter Sarnak for clarifying discussions around the Siegel mass formula.

2. SOME BACKGROUND FROM INCIDENCE THEORY

Let $\mathcal{P}$ be a collection of points in $\mathbb{R}^n$ and let $\mathcal{H}$ be a collection of sets in $\mathbb{R}^n$. We will not assume at this point that $\mathcal{H}$ consists of hyperplanes. Consider the standard incidence bipartite graph $G(\mathcal{P}, \mathcal{H})$ with vertex sets $\mathcal{P}$ and $\mathcal{H}$, where we have an edge between $P \in \mathcal{P}$ and $H \in \mathcal{H}$ whenever we have the incidence $P \in H$. So the number of edges $E$ in $G(\mathcal{P}, \mathcal{H})$ is the same as the number of incidences $I(\mathcal{P}, \mathcal{H})$ between $\mathcal{P}$ and $\mathcal{H}$.

Our approach in this section is an adaptation of Theorem 8 from [9] to our needs. The next two lemmas prove some weaker bounds that are then amplified to optimal bounds in Proposition 2.3. The first one is tailored for applications to four dimensions, while the second one for five dimensions.

**Lemma 2.1.** Fix $\gamma > 4$. Assume $|H \cap H' \cap \mathcal{P}| \leq \gamma$ for each $H \neq H' \in \mathcal{H}$. Then

$$I(\mathcal{P}, \mathcal{H}) \leq \gamma(|\mathcal{P}| + |\mathcal{H}|\sqrt{|\mathcal{P}|}).$$

**Proof** The argument is a standard double counting. Let $\mathcal{P}_H$ be the points in $\mathcal{P} \cap H$ and let $\mathcal{H}_P$ be the sets in $\mathcal{H}$ that contain the point $P$. We estimate $\sum_{H \in \mathcal{H}} I(\mathcal{P}_H, \mathcal{H})$ from above by

$$\leq \sum_{H \in \mathcal{H}} (\gamma|\mathcal{H}| + |\mathcal{P}_H|) = \gamma|\mathcal{H}|^2 + |E|$$

and also from below by

$$\sum_{P \in \mathcal{P}} |\mathcal{H}_P|^2 \geq \frac{1}{|\mathcal{P}|}(\sum_{P \in \mathcal{P}} |\mathcal{H}_P|)^2 = \frac{E^2}{|\mathcal{P}|}.$$

Thus

$$|\mathcal{P}| \gamma |\mathcal{H}|^2 \geq |E|^2(1 - \frac{|\mathcal{P}|}{E}).$$

Then either $E \leq 2|\mathcal{P}|$ or, if not, the above implies $E^2 \leq 2|\mathcal{P}| \gamma |\mathcal{H}|^2$. In either case we are fine.

\[ \blacksquare \]
Lemma 2.2. Fix $\gamma > 4$. Assume that for each $H \neq H' \in \mathcal{H}$
\[
|\{H'' \in \mathcal{H} : |H \cap H' \cap H'' \cap \mathcal{P}| \geq \gamma\}| \leq \gamma.
\]
Then
\[
I(\mathcal{P}, \mathcal{H}) \leq 4\gamma (|\mathcal{P}| + |\mathcal{H}| |\mathcal{P}|^{2/3})
\]
Proof. We will apply twice the double counting argument from the proof of Lemma 2.1. Let as before $\mathcal{P}_H$ be the points in $\mathcal{P} \cap H$ and let $\mathcal{H}_P$ be the sets in $\mathcal{H}$ that contain the point $P$. Set $E_1 = I(\mathcal{P}, \mathcal{H})$. Let us assume for a moment that we have the following inequality for each $H \in \mathcal{H}$
\[
I(\mathcal{P}_H, \mathcal{H}) \leq 2\gamma (|\mathcal{P}_H| + |\mathcal{H}| \sqrt{|\mathcal{P}_H|})
\]
(1)
We can then estimate $\sum_{H \in \mathcal{H}} I(\mathcal{P}_H, \mathcal{H})$ from above using Cauchy-Schwartz by
\[
\sum_{H \in \mathcal{H}} 2\gamma (|\mathcal{P}_H| + |\mathcal{H}| \sqrt{|\mathcal{P}_H|}) = 2\gamma (E_1 + |\mathcal{H}|^{3/2} E_1^{1/2})
\]
and also from below by
\[
= \sum_{P \in \mathcal{P}} |\mathcal{H}_P|^2 \geq \frac{1}{|\mathcal{P}|} (\sum_{P \in \mathcal{P}} |\mathcal{H}_P|)^2 = \frac{E_1^2}{|\mathcal{P}|}
\]
Thus
\[
2\gamma |\mathcal{P}| |\mathcal{H}|^{3/2} \geq E_1^{3/2} (1 - \frac{2\gamma |\mathcal{P}|}{E_1}).
\]
Then either $E_1 \leq 4\gamma |\mathcal{P}|$ or, if not, the above implies $E_1^{3/2} \leq 4\gamma |\mathcal{P}| |\mathcal{H}|^{3/2}$. In either case we are fine.

It remains to prove (1). Fix $H \in \mathcal{H}$ and define for simplicity $\mathcal{P}' = \mathcal{P}_H$, $H' = \mathcal{H} \setminus \{H\}$ and $E = I(\mathcal{P}', \mathcal{H}')$. Since $I(\mathcal{P}', \mathcal{H}) = I(\mathcal{P}', \mathcal{H}') + |\mathcal{P}'|$, we are left with estimating $I(\mathcal{P}', \mathcal{H}')$. We apply again the double counting argument. Let as before $\mathcal{P}'_H$ be the points in $\mathcal{P}' \cap H'$ and let $\mathcal{H}'_P$ be the sets in $\mathcal{H}'$ that contain the point $P$. For each $H' \in \mathcal{H}'$ define $\mathcal{H}'_1, \mathcal{H}'_2$ to consist of those $H'' \in \mathcal{H}'$ such that $|H' \cap H'' \cap \mathcal{P}'| \geq \gamma$. Note that our hypothesis implies that $|\mathcal{H}'_1| \leq \gamma$. Let $\mathcal{H}'_2 = \mathcal{H}' \setminus \mathcal{H}'_1$.

We estimate $\sum_{H' \in \mathcal{H}'} I(\mathcal{P}'_{H'}, \mathcal{H}')$ from above by
\[
\sum_{H' \in \mathcal{H}'} I(\mathcal{P}'_{H'}, \mathcal{H}'_1, \mathcal{H}') + \sum_{H' \in \mathcal{H}'} I(\mathcal{P}'_{H'}, \mathcal{H}'_2, \mathcal{H}') \leq \sum_{H' \in \mathcal{H}'} (|\mathcal{P}'_{H'}| + |\mathcal{H}'|) = \gamma (E + |\mathcal{H}'|^2)
\]
and also from below by
\[
= \sum_{P \in \mathcal{P}'} |\mathcal{H}'_P|^2 \geq \frac{1}{|\mathcal{P}'|} (\sum_{P \in \mathcal{P}'} |\mathcal{H}'_P|)^2 = \frac{E^2}{|\mathcal{P}'|}
\]
Thus
\[
|\mathcal{P}'| |\mathcal{H}'|^2 \geq |E|^2 (1 - \frac{\gamma |\mathcal{P}'|}{E}).
\]
Then either $E \leq 2|\mathcal{P}'|$ or, if not, the above implies $E^2 \leq 2|\mathcal{P}'| |\mathcal{H}'|^2$. In either case we are fine.

From this point on we assume $\mathcal{H}$ consists of hyperplanes.
Proposition 2.3. Let $\mathcal{H}$ of finite collection of hyperplanes in $\mathbb{R}^n$ and let $\mathcal{P}$ be a finite collection of points in $\mathbb{R}^n$. Assume the following hold for a given $\gamma \geq 1$

(a) $I(\mathcal{P}', \mathcal{H}') \leq \gamma (|\mathcal{P}'| + |\mathcal{H}'| |\mathcal{P}'|^{\frac{n-3}{n-2}})$ for each $\mathcal{P}' \subset \mathcal{P}$, $\mathcal{H}' \subset \mathcal{H}$

(b) Any $\gamma$ hyperplanes in $\mathcal{H}$ share fewer than $\gamma$ points in $\mathcal{P}$

Then the number of incidences satisfies for each $\epsilon > 0$

$$I(\mathcal{P}, \mathcal{H}) \leq C_\epsilon \gamma (|\mathcal{P}|^\alpha |\mathcal{H}|^\beta + |\mathcal{P}| + |\mathcal{H}|(1 + \log_2 |\mathcal{P}|)),$$  \hspace{1cm} (2)

where $\alpha = \frac{n(n-3)}{n^2 - 2n - 1}$, $\beta = \frac{(n-1)(n-2)}{n^2 - 2n - 1} + \epsilon$ and $C_\epsilon$ depends only on $\epsilon$ and $n$.

Recall the following Cutting Lemma (see Theorem 6.5.3 in [14]). This will enable a proof of the Proposition via induction.

Lemma 2.4. Given $s$ hyperplanes in $\mathbb{R}^n$ and a positive integer $r < s$, there exists a partition of $\mathbb{R}^n$ into fewer than $r^n$ parts, such that for each part there are at most $Bs/r$ hyperplanes which cut it (this means intersect it without containing it). $B$ will be a large number depending on $n$, but independent of $s, r$.

We now begin the proof of the Proposition 2.3, following [9]. By performing a translation, we can assume that neither of the hyperplanes in $\mathcal{H}$ contains the origin $0$, and also that $0 \notin \mathcal{P}$.

Choose $r$ large enough so that

$$\frac{B^\alpha}{r^{\alpha - n(1-\beta)}} + \frac{B}{r} < \frac{1}{2},$$  \hspace{1cm} (3)

Here $B$ is the constant from Lemma 2.4. Note that $\alpha > n(1 - \beta)$ and also that $r$ will only depend on $n, \epsilon$.

Let $s = |\mathcal{P}|$ and $t = |\mathcal{H}|$. We prove (2) via induction on $s$. The case $s = 1$ holds trivially true. Assume now it holds for $1, \ldots, s - 1$. We split the analysis in three cases.

Case 1: If $s \leq r$ then we trivially have $I(s, t) \leq st \leq sr$, so it suffices to choose $C_\epsilon > r$.

Case 2: If $r^{1-\alpha}s \geq t^{n-2}$ then hypothesis (a) implies that

$$I(s, t) \leq \gamma (s + ts^{\frac{n-4}{n-2}}) \leq \gamma s (1 + r^{\frac{n}{n-1}} t^{\frac{n}{n-2}}),$$

so it suffices to choose $C_\epsilon > 1 + r^{\frac{n}{1-\alpha(n-2)}}$.

Case 3: We now focus on the case when $r < s$ and $r^{\frac{n}{1-\alpha}}s < t^{n-2}$. By raising both terms in the second inequality to the power $1 - \alpha$ we get

$$s < r^{-\alpha} s^\alpha t^\beta.$$  \hspace{1cm} (4)

Next, we dualize. That is, we identify each point $P \in \mathcal{P}$ with the hyperplane

$$H_P := \{x \in \mathbb{R}^n : \langle x, P \rangle = 1\}$$

(call the resulting collection $\mathcal{H}'$) and each hyperplane $H \in \mathcal{H}$ with the point $P \in \mathbb{R}^n \setminus \{0\}$ such that

$$H_P = H$$

(call the resulting collection $\mathcal{P}'$). It is easy to see that incidences are preserved, that is $P \in H_P'$ if and only if $P' \in H_P$.

Apply the cutting lemma to the collections $\mathcal{H}'$, $\mathcal{P}'$ and $r$. Note that we operate under the assumption $r < s$, which makes the lemma applicable. Assign each point in $\mathcal{P}'$ to the part that contains it, and to each part we assign all hyperplanes in $\mathcal{H}'$ which cut it. So a
hyperplane can be assigned to more than one part, and there may be parts that are not assigned any hyperplanes. Call \( s_i \) and \( t_i \) the hyperplanes and points assigned to the \( i \)th of the \( M \) parts. We have

\[
M \leq r^n, \quad \sum_{i=1}^{M} t_i = t, \quad s_i \leq \frac{Bs}{r}.
\]

Each part contributes with two types of incidences. First, with the hyperplanes that cut it. Second, with those that contain it. The first contribution is bounded using the induction hypothesis (after undualizing) by

\[
I(t_i, s_i) \leq C\gamma (s^\alpha t^\beta_i + s_i + t_i(1 + \log_2 s_i)).
\]

The second contribution is bounded by \( \gamma(s + t) \). Indeed, if the part contains fewer than \( \gamma \) points then there are at most \( \gamma s \) incidences. If there are at least \( \gamma \) points in the part, there can be at most \( \gamma \) hyperplanes in \( \mathcal{H}' \) containing the part (undualize and use hypothesis (b)). Thus, there are fewer than \( \gamma t \) incidences. We conclude that

\[
I(s, t) \leq C\gamma \left( \sum_{i=1}^{M} (s^\alpha t^\beta_i + s_i + t_i(1 + \log_2 s_i)) + r^n \gamma(s + t) \right)
\]

\[
\leq C\gamma \left[ \left( \frac{Bs}{r} \right)^\alpha \sum_{i=1}^{M} t^\beta_i + B M s \frac{s}{r} + t(1 + \log_2 \frac{Bs}{r}) \right] + r^n \gamma(s + t).
\]

Using (3) we can further bound this by

\[
C\gamma \left[ \left( \frac{Bs}{r} \right)^\alpha M^{1-\beta} \left( \sum_{i=1}^{M} t_i \right)^\beta + B s r^{n-1} + s \frac{r^n}{C\epsilon} + t \left( \frac{r^n}{C\epsilon} + \log_2 s \right) \right].
\]

Since the second term does not fit well (\( B r^{n-1} \) is greater than 1), we need to replace it using (4). We further bound the above by

\[
C\gamma \left[ \left( \frac{B}{r} \right)^\alpha n(1-\beta) + B \frac{s}{r} t^\beta + s \frac{r^n}{C\epsilon} + t \left( \frac{r^n}{C\epsilon} + \log_2 s \right) \right].
\]

It now suffices to choose \( C\epsilon > r^n \). This ends the proof of the Proposition.

3. The Incidence Theory Approach

For \( \Lambda \subset \mathbb{R}^n \) define its additive energy

\[
\mathcal{E}(\Lambda) = \left| \mathcal{F}_{n,\epsilon} \right| = |\{(\xi_1, \xi_2, \xi_3, \xi_4) \in \Lambda^4 : \xi_1 + \xi_2 = \xi_3 + \xi_4\}|.
\]

We now show how to use the incidence theory developed so far to estimate the additive energy of subsets of the sphere. We will rely on the well known estimates, see [12]

\[
|\mathcal{F}_{n,\epsilon}| \lesssim N^{n-2+\epsilon}, \quad n = 2, 3, 4
\]

\[
|\mathcal{F}_{n,\epsilon}| \approx N^{n-2}, \quad n \geq 5
\]

For \( v \in \mathbb{Z}^n \) with \( |v| < 2\lambda^{1/2} \) let \( H_v \) be the unique hyperplane in \( \mathbb{R}^n \) containing the \( n-2 \) dimensional sphere

\[
S_v = \{ \xi \in \lambda^{1/2} S^{n-1} : \|\xi - v\| = \lambda^{1/2} \}.
\]
Theorem 3.1. Let $\Lambda$ be an arbitrary subset of $\mathcal{F}_{4,\Lambda}$. Then its energy satisfies for each $\epsilon > 0$

$$E(\Lambda) \lesssim_{\epsilon} N^\epsilon |\Lambda|^{7/3}. \quad (7)$$

Proof. Note that given $\xi, \eta \in \lambda^{1/2} S^3$, we have that $\xi + \eta = v$ if and only if $\xi, \eta \in S_v$ and $\xi, \eta$ are diametrically opposite on $S_v$. For $0 \leq k \leq \lfloor \log_2 |\Lambda| \rfloor$ let $M_k$ denote the number of hyperplanes $H_v$ containing between $2^k$ and $2^{k+1} - 1$ pairs $(\xi, \eta) \in \Lambda^2$ such that $\xi + \eta = v$. Since $E(\Lambda) \leq \sum_k 2^{2k+2} M_k$ and $|\Lambda| \lesssim N^4$, it suffices to prove that for each $k$

$$M_k 2^{2k} \lesssim_{\epsilon} N^\epsilon |\Lambda|^{7/3}. \quad (8)$$

We will find two upper bounds for $M_k$. First, note the trivial bound

$$M_k 2^k \lesssim |\Lambda|^2. \quad (9)$$

Next, note that $M_k$ is smaller that the number $N_k$ of hyperplanes $H_v$-call the collection $\mathcal{H}$- satisfying

$$2^k \leq |H_v \cap \Lambda|.$$

Recall that any circle on $\lambda^{1/2} S^3$ contains $O(N^\epsilon)$ points in $\mathbb{Z}^4$, see [1]. Thus $\mathcal{H}$ satisfies the requirement in Lemma 2.1, for $\gamma$ large enough but satisfying $\gamma \lesssim N^\epsilon$ for each $\epsilon > 0$. Note that there are at least $2^k N_k$ incidences between $\mathcal{H}$ and $\Lambda$. Apply now Proposition 2.3 with $P = \Lambda$ and $n = 4$ to get

$$N_k 2^k \lesssim_{\epsilon} N^\epsilon (|\Lambda|^4 N_k^6 + |\Lambda| + N_k (1 + \log_2 |\Lambda|)). \quad (10)$$

If $N_k 2^k \lesssim_{\epsilon} N^\epsilon |\Lambda|^4 N_k^6$, then since $N_k \lesssim N^4$ we get

$$N_k \lesssim_{\epsilon} N^\epsilon \frac{|\Lambda|^4}{2^{7k}}.$$

Combining this with (9) and $M_k \leq N_k$ gives (8).

If either $N_k 2^k \lesssim_{\epsilon} N^\epsilon |\Lambda|$ or $N_k 2^k \lesssim_{\epsilon} N^\epsilon N_k (1 + \log_2 |\Lambda|)$ then (8) follows immediately from (9).

Remark 3.2. Note that the expected result is

$$E(\Lambda) \lesssim_{\epsilon} N^\epsilon |\Lambda|^2.$$

To prove (7) we have relied on the incidence bound (10). This bound holds for any collection of hyperplanes in $\mathbb{R}^4$ subject to the only requirement that any two of them share at most $O(N^\epsilon)$ points in $\Lambda$. We now show that (10) can not in general be improved unless this requirement is strengthened in some way.

On the one hand, note that the argument in the proof of Theorem 3.1 can be applied with no essential modifications to the paraboloid

$$P_N^3 = \{\xi := (\xi_1, \xi_2, \xi_3, \xi_1^2 + \xi_2^2 + \xi_3^2) : |\xi_i| \leq N\}.$$

Indeed, if $\xi + \eta = v := (v_1, v_2, v_3, v_4)$, then

$$\sum_{i=1}^3 [\xi_i^2 + (\xi_i - v_i)^2] = v_4.$$
and thus $\xi, \eta$ belong to the hyperplane
\[ H_v := \{ \theta \in \mathbb{R}^4 : 2v_1\theta_1 + 2v_2\theta_2 + 2v_3\theta_3 - 2\theta_4 = v_1^2 + v_2^2 + v_3^2 - v_4 \} \].

Next note that for $v \neq v'$ the projection onto the first three components of $H_v \cap H_{v'} \cap P^3_N$ is a subset of $C \cap \mathbb{Z}^3$, where $C$ is a certain circle of radius $O(N)$. Thus
\[ |H_v \cap H_{v'} \cap P^3_N| \lesssim \epsilon N^\epsilon. \]

On the other hand, the estimate (7) is sharp for $\Lambda = P^3_N \cap \mathbb{Z}^4$. Indeed, note that $\mathbb{E}(P^3_N) = \|K\|_{L^4(\mathbb{T}^1)}^4$, where
\[ K(x) = \sum_{\xi \in P^3_N} e(\xi \cdot x). \]

Since $|\xi \cdot x| \ll 1$ for $|x_1|, |x_2|, |x_3| \ll \frac{1}{N}$ and $|x_4| \ll \frac{1}{N^2}$, it follows that $|K(x)| \gtrsim N^3$ for $x$ in a set of measure $\gtrsim \frac{1}{N^7}$. This shows $\|K\|_{L^4}^4 \gtrsim N^7$.

We will now obtain a similar result in five dimensions. The new observation that we need in this case is

**Lemma 3.3.** There are $O(N^\epsilon)$ hyperplanes $H_v$ in $\mathbb{R}^5$ containing a given three dimensional affine subspace $W$ of $\mathbb{R}^5$.

**Proof** This will follow from a few easy observations. Call $V$ the collection of all such $v$ and fix $\eta \in W$. First, it is easy to see that $v$ is orthogonal to $H_v$, in particular each $v \in V$ is orthogonal to $W$. Second, note that if $\xi \in S_v$ then also $v - \xi \in S_v$. Thus $\xi, v - \xi \in H_v$ which forces $v/2 \in H_v$. Combining this with the first observation further implies that $\langle v/2, \eta - v/2 \rangle = 0$. Thus $v/2$ belongs to the sphere centered at $\eta/2$ of radius $|\eta|/2$. Since $v/2$ is orthogonal to $W$, it is confined to a two dimensional subspace. As a result, all $v \in V$ will belong to a fixed circle of radius $O(\lambda^{1/2})$. It now suffices to invoke again the result in [1].

**Theorem 3.4.** Let $\Lambda$ be an arbitrary subset of $\mathcal{F}_{5,\lambda}$. Then its energy satisfies for each $\epsilon > 0$
\[ \mathbb{E}(\Lambda) \lesssim \epsilon N^\epsilon |\Lambda|^{5/2}. \]  

**Proof** The analysis is very similar with that in the proof of Theorem 3.1, we will use the notation $M_k, N_k, \mathcal{H}$ from there. The crucial difference is that the new collection $\mathcal{H}$ does not satisfy the requirement in Lemma 2.1, since two hyperplanes intersect $\lambda^{1/2}S^4$ along a three dimensional sphere that may contain as many as $N$ points in $\Lambda$. However, Lemma 3.3 shows that Lemma 2.2 is applicable in our situation. We will choose as before a $\gamma$ large enough but satisfying $\gamma \lesssim \epsilon N^\epsilon$ for each $\epsilon > 0$. Indeed, given distinct $H, H', H'' \in \mathcal{H}$ with $|H \cap H' \cap H'' \cap \Lambda| \geq \gamma$, it must be that $H \cap H' \cap H''$ is a three dimensional linear subspace $W$. This is because any lower dimensional subspace contains fewer that $\gamma$ points, if $\gamma$ is chosen large enough. But then $W = H \cap H'$ and thus $H''$ contains $W$. Lemma 3.3 produces the desired upper bound.

Apply Proposition 2.3 to $\mathcal{H}$, $\mathcal{P} = \Lambda$ and $n = 5$ to get
\[ N_k 2^k \lesssim \epsilon N^\epsilon |\Lambda|^{\frac{5}{2}}N_k^{\frac{6}{5}+\epsilon} + |\Lambda| + N_k (1 + \log_2 |\Lambda|). \]

The argument then follows closely the lines of that in the proof of Theorem 3.1.
Remark 3.5. The analogue of Remark 3.2 applies in this context, too.

4. COUNTING SOLUTIONS TO SYSTEMS OF QUADRATIC EQUATIONS

In this section we develop the necessary number theoretical machinery that will enable us to prove a different type of estimate for the energy of the lattice points on the sphere. The main theorem is as follows.

Theorem 4.1. (a) Let \( N_{a,b,\lambda} \) be the number of solutions \((x, y, z) \in (\mathbb{Z}^4)^3 \) of the system of equations

\[
\begin{bmatrix}
x_1 & x_2 & x_3 & x_4 \\
y_1 & y_2 & y_3 & y_4 \\
z_1 & z_2 & z_3 & z_4
\end{bmatrix}
\begin{bmatrix}
x_1 & y_1 & z_1 \\
x_2 & y_2 & z_2 \\
x_3 & y_3 & z_3 \\
x_4 & y_4 & z_4
\end{bmatrix}
= \begin{bmatrix}
\lambda & a & \lambda + a - b \\
a & \lambda & b \\
\lambda + a - b & b & \lambda
\end{bmatrix}
\]

Then

\[
\sum_{|a|,|b| \leq \lambda} N_{a,b,\lambda} \lesssim \epsilon \lambda^{2+\epsilon}
\]

for each \( \epsilon > 0 \).

(b) Let \( N_{a,b,c,d,\lambda} \) be the number of solutions \((u, v, x, y) \in (\mathbb{Z}^5)^4 \) with \( x \neq y \) of the system of equations

\[
\begin{bmatrix}
u_1 & u_2 & u_3 & u_4 & u_5 \\
v_1 & v_2 & v_3 & v_4 & v_5 \\
x_1 & x_2 & x_3 & x_4 & x_5 \\
y_1 & y_2 & y_3 & y_4 & y_5
\end{bmatrix}
\begin{bmatrix}
u_1 & v_1 & x_1 & y_1 \\
u_2 & v_2 & x_2 & y_2 \\
u_3 & v_3 & x_3 & y_3 \\
u_4 & v_4 & x_4 & y_4 \\
u_5 & v_5 & x_5 & y_5
\end{bmatrix}
= \begin{bmatrix}
a & c & a/2 & a/2 \\
c & b & b/2 & b/2 \\
a/2 & b/2 & \lambda & d \\
a/2 & b/2 & d & \lambda
\end{bmatrix}
\]

Then

\[
\sum_{|a|,|b|,|c|,|d| \leq \lambda} N_{a,b,c,d,\lambda} \lesssim \epsilon \lambda^{4+\epsilon}
\]

for each \( \epsilon > 0 \).

Note that for part (b) we have to exclude solutions with \( x = y \). The computations from Section 5 show that the sum over \(|a|, |b|, |c|, |d| \lesssim \lambda\) of this type of solutions is roughly

\[
|\{(u, v, x) \in \mathbb{Z}^5 \times \mathbb{Z}^5 \times F_{5,\lambda} : u - x, v - x \in F_{5,\lambda}\}| = |F_{5,\lambda}|^3 \sim \lambda^{9/2}.
\]

Our main tool will be Siegel’s mass formula which we recall below. In a nutshell, this formula relates the number of integral solutions to a system of quadratic equations with the number of solutions of the same system in \( \mathbb{Z}_{p^r} \), with \( p \) prime and \( r \to \infty \). The necessary background and the proof of Siegel’s mass formula are in [15]. More precisely, we will use the formula on page 10, case (i) from Lecture No. 2, which is proved in Lecture No. 6.

Let \( m \geq n + 1 \) and let \( \gamma \in M_{m,n}(\mathbb{Z}) \) and \( \Lambda \in M_{n,n}(\mathbb{Z}) \) be two positive definite matrices with integer entries. Denote by \( A(\gamma, \Lambda) \) the number of solutions \( \mathcal{L} \in M_{m,n}(\mathbb{Z}) \) for

\[
\mathcal{L}^* \gamma \mathcal{L} = \Lambda.
\]
Then Siegel’s mass formula asserts that

\[
\frac{h}{A(\gamma_i, \Lambda)} \sum_{i=1}^{h} \frac{1}{A(\gamma_i, \gamma_i)}^{-1} = C_{n,m,\gamma} A_0(\gamma, \Lambda) \prod_{p \text{ prime}} \nu_p(\gamma, \Lambda). \tag{13}
\]

Here \(h\) is the number of classes in the genus of \(\gamma\) and \(\gamma_i\) is a (any) representative for its class. On pages 9 and 10 of [15] it is stated that

\[
\nu_p(\gamma, \Lambda) = \lim_{r \to \infty} \frac{1}{p^r(mn - \frac{n(n+1)}{2})} |\{ \mathcal{L} \in M_{m,n}(\mathbb{Z}_{p^r}) : \mathcal{L}^* \mathcal{L} \equiv \Lambda \mod p^r\}|,
\]

while the computations on page 41 in [15] show that

\[
A_0(\gamma, \Lambda) = K_{n,m}(\det(\gamma))^{-n/2} (\det(\Lambda))^{m-n-1/2}. \tag{14}
\]

As \(A(\gamma, \Lambda) = A(\gamma_i, \Lambda)\) for some \(1 \leq i \leq h\), we immediately get

\[
A(\gamma, \Lambda) \lesssim_{n,m,\gamma} (\det(\Lambda))^{m-n-1/2} \prod_{p \text{ prime}} \nu_p(\gamma, \Lambda). \tag{15}
\]

In our forthcoming applications \(m = n + 1\), \(\gamma\) will always be the identity matrix \(I_{n+1}\) and \(m\) is the dimension of the ambient space where the lattice points live.

Fix \(\Lambda \in M_{n,n}(\mathbb{Z})\), a nonsingular positive definite matrix, in particular \(\det(\Lambda) \neq 0\). In evaluating \(\nu_p(I_{n+1}, \Lambda)\) we distinguish two separate cases: \(p \nmid \det(\Lambda)\) and \(p | \det(\Lambda)\). We start with the first case.

**Proposition 4.2.** Assume \(p\) is not a factor of \(\det(\Lambda)\). Then

\[
\nu_p(I_{n+1}, \Lambda) \leq 1 + \frac{C}{p^2},
\]

where \(C\) is independent of \(p, \Lambda\).

To prove the proposition we first analyze the case \(r = 1\). Using the same invariance considerations as in the evaluation of the term \(A_0\) in [15] Lecture 6, we get that

\[
|\{ \mathcal{L} \in M_{n+1,n}(\mathbb{Z}_p) : \mathcal{L}^* \mathcal{L} \equiv \Lambda \mod \mathcal{L} \}| \tag{16}
\]

only depends on the Legendre symbol \((\frac{\det(\Lambda)}{p})\). Thus we can replace \(\Lambda\) with the diagonal matrix

\[
\Lambda_\xi = \xi e_1 \otimes e_1 + \sum_{j=2}^{n} e_j \otimes e_j \tag{17}
\]

where \((\xi) = (\frac{\det(\Lambda)}{p})\).

We will rely on the following elementary fact, see Exercise 13 on page 31 in [8]

**Lemma 4.3.** Let \(g(x) = [x]^* C [x]\) be a quadratic form with \(C \in M_{l,l}(\mathbb{F}_p)\) symmetric and \(d := \det(C) \equiv 0 \mod p\). Denote for \(\xi \in \mathbb{F}_p\)

\[
N_\xi(d, l) = |\{x \in (\mathbb{F}_p)^l : g(x) = \xi\}|.
\]
Then
\[
N_0(d, l) = \begin{cases} 
 p^{l-1} & : \text{if } l = 2\nu + 1, \ \nu \geq 0 \\
 p^{l-1} + (p - 1)p^{\nu-1}\left(\frac{(-1)^\nu d}{p}\right) & : \text{if } l = 2\nu, \ \nu > 0
\end{cases}
\]

\[
N_1(d, l) = \begin{cases} 
 p^{l-1} + p^{\nu}\left(\frac{(-1)^\nu d}{p}\right) & : \text{if } l = 2\nu + 1, \ \nu \geq 0 \\
 p^{l-1} - p^{\nu-1}\left(\frac{(-1)^\nu d}{p}\right) & : \text{if } l = 2\nu, \ \nu > 0
\end{cases}
\]

where \((\frac{a}{p})\) is the Legendre symbol.

It is easy to see that if \(\eta \in \mathbb{F}_p^*\) then \(N_\xi(d, l)\) only depends on the class of \(\eta\) in the two element group \(\mathbb{F}_p^*/(\mathbb{F}_p^*)^2\). Thus
\[
N_0(d, l) + \frac{p-1}{2} N_1(d, l) + \frac{p-1}{2} N_\eta(d, l) = p^l
\]

for each \(\eta \in \mathbb{F}_p^* \setminus (\mathbb{F}_p^*)^2\), and we conclude that
\[
N_\eta(d, l) = \begin{cases} 
 p^{l-1} - p^{\nu}\left(\frac{(-1)^\nu d}{p}\right) & : \text{if } l = 2\nu + 1, \ \nu \geq 0 \\
 p^{l-1} - p^{\nu-1}\left(\frac{(-1)^\nu d}{p}\right) & : \text{if } l = 2\nu, \ \nu > 0
\end{cases}
\]

We now evaluate
\[
|\{\mathcal{L} \in M_{n+1,n}(\mathbb{Z}_p): \mathcal{L}^*\mathcal{L} \equiv \xi e_1 \otimes e_1 + \sum_{j=2}^n e_j \otimes e_j \mod p\}|
\]

for \(n \geq 2\). We need to count pairwise orthogonal vectors \(x^1, \ldots, x^{n-1} \in \mathbb{F}_p^{n+1}\) such that
\[
x^1 \cdot x^1 = \xi \text{ and } x^j \cdot x^j = 1 \text{ for } 2 \leq j \leq n.
\]

We have \(N_{\xi}(1, n + 1)\) choices for \(x^1\). Once we have chosen \(x^1\), there will be \(N_{1}(1, n)\) possibilities for \(x^2\), since \(x^2 \in (x^1)^\perp\). By repeating this reasoning and then using (19), (20), we bound the term (21) by
\[
N_{\xi}(1, n + 1) \prod_{k=2}^n N_{1}(1, k) \leq \]
\[
p^n(1 + \frac{1}{p}) \cdots p^3(1 + \frac{1}{p^2})(p - \frac{1}{p}) = p^{n(n+1)/2}(1 + O(\frac{1}{p^2})).
\]

To get the proof of Proposition 4.2 we need to recall Hensel’s lemma (see for example [11], Chapter 5 and [17])

**Lemma 4.4.** Let \(r \geq 1\). Let \(f_1, \ldots, f_k \in \mathbb{Z}[X_1, \ldots, X_k]\) for \(1 \leq j \leq k\) be a collection of polynomials, and set
\[
J(X) = \det(\frac{\partial f_j}{\partial X_i}(X))_{1 \leq i, j \leq k}.
\]

Let \(X \in \mathbb{Z}_p^k\) be a solution of
\[
f_j(X) \equiv 0 \mod p, \ 1 \leq j \leq k
\]
such that \( J(X) \not\equiv 0 \mod p \). Then there is a unique solution \( X' \in \mathbb{Z}_p^k \) of
\[
f_j(X') \equiv 0 \mod p^r, \quad 1 \leq j \leq k
\]
that satisfies \( X' \equiv X \mod p \).

Recall that we need to count the number of solutions for
\[
L \in M_{n+1,n}(\mathbb{Z}_p) : \quad L^*L \equiv \Lambda \mod p^r
\]
with \( \det(\Lambda) \not\equiv 0 \mod p \) it follows that any solution \( L = (x_0^1, \ldots, x_0^n) \) consists of linearly independent vectors \( x_0^i \) over \( \mathbb{Z}_p \). Equivalently, the rank \( \mod p \) of the matrix \( L \) is maximal (it equals \( n \)). Hensel’s lemma with \( k = \frac{n(n+1)}{2} \) shows that each solution for
\[
L \in M_{n+1,n}(\mathbb{Z}_p) : \quad L^*L \equiv \Lambda \mod p^r
\]
gives rise to exactly \( p^{\frac{n(n+1)}{2}r(1 + \frac{C}{p^r})} \) solutions for \( L' \in M_{n+1,n}(\mathbb{Z}_p^r) : \quad (L')^*L' \equiv \Lambda \mod p^r \) such that \( L' \equiv L \mod p \). Indeed, for each such \( L \), let \( X_1, \ldots, X_{\frac{n(n+1)}{2}} \) be the variables among \( x_1^1, \ldots, x_n^1 \) that correspond to \( \frac{n(n+1)}{2} \) independent columns of \( M(x_0^1, \ldots, x_0^n) \). Also let \( f_1, \ldots, f_{\frac{n(n+1)}{2}} \) be the functions \( x_i \cdot x_j \), considered as functions of only the variables \( X_1, \ldots, X_{\frac{n(n+1)}{2}} \). The remaining \( \frac{n(n+1)}{2} \) variables are fixed and note that there are \( p^{\frac{n(n+1)}{2}(r-1)} \) ways to fix them. For each such choice Hensel’s lemma provides exactly one way to complete the solution \( L' \). Combining this with estimate (22) produces the upper bound \( p^{\frac{n(n+1)}{2}r(1 + \frac{C}{p^r})} \) for the number of solutions of (23). This finishes the proof of Proposition 4.2.

Next we analyze the case of those primes \( p \) which divide \( \det(\Lambda) \). Since there are
\[
O\left(\frac{\log \det(\Lambda)}{\log \log \det(\Lambda)}\right)
\]
such primes, we will content ourselves with obtaining cruder bounds for the densities \( \nu_p \), which are only sharp up to a multiplicative constant. We will denote by \( o_p(T) \) the largest \( \alpha \) such that \( p^\alpha \mid T \).

One of our main tools here is the following result in [17]

**Lemma 4.5.** Let \( f_1, \ldots, f_d \in \mathbb{Z}[X_1, \ldots, X_d] \) be polynomials of degrees \( k_1, \ldots, k_d \) and set
\[
J(X) = \det(\frac{\partial f_j}{\partial X_i}(X))_{1 \leq i,j \leq d}.
\]
Then the number of solutions \( X \in (\mathbb{Z}_p)^d \) of
\[
f_j(X) \equiv 0 \mod p^r, \quad 1 \leq j \leq d
\]
for which \( J(X) \not\equiv 0 \mod p \) is at most \( k_1k_2 \ldots k_d \).
For an $n \times n$ matrix $\Lambda$ and for $A, B \subset \{1, \ldots, n\}$ with $|A| = |B|$ we define

$$
\mu_{A, B} = \det((\Lambda_{i,j})_{i \in A, j \in B}).
$$

**Proposition 4.6.** Let $\Lambda \in M_{n,n}(\mathbb{Z})$ be a positive definite matrix and let $p | \det(\Lambda)$. Then

$$
\nu_p(I_{n+1}, \Lambda) \lesssim \sum_{0 \leq \beta_1 + \cdots + \beta_n \leq \det(\Lambda)} p^{\beta_1(l_1, \ldots, l_n) + \cdots + \beta_n(l_1, \ldots, l_n)},
$$

where $\beta_i = \beta_i(l_1, \ldots, l_n)$ satisfies

$$
\beta_i = \min \{ (i-1)l_i, (i-2)l_i + \min_{|A| = 1} \mu_p(\mu_{1, A}) - l_1, (i-3)l_i + \min_{|A| = 2} \mu_p(\mu_{1,2, A}) - l_1 - l_2, \ldots, \min_{|A| = i-1} \mu_p(\mu_{1,2,\ldots,i-1, A}) - l_1 - l_2 - \cdots - l_{i-1} \}.
$$

**Proof.** We first show how to count the non-degenerate solutions for

$$
\mathcal{L} \in M_{n+1,n}(\mathbb{Z}_p^r): \mathcal{L}^* \mathcal{L} \equiv \Lambda \mod p^r,
$$

by which we mean the solutions $(x^1, \ldots, x^n)$ such that $x^j$ are linearly independent in the vector space $\mathbb{Z}_p^{n+1}$ over the field $\mathbb{Z}_p$. Recall that this implies that the $n(n+1)/2 \times n(n+1)$ matrix $M = M(x^1, \ldots, x^n)$ has rank $p$ equal to $n(n+1)/2$. Pick $n(n+1)/2$ independent columns of $M$. We fix $\mod p^r$ the values of the $n(n+1)/2$ variables corresponding to the remaining columns of $M$, and apply Lemma 4.5 with $d = n(n+1)/2$ to get at most $O(1)$ solutions. Thus the overall contribution of the non-degenerate solutions is $O(p^{n(n+1)/2\cdot r})$.

We next use a sequence of reductions that will allow us to relate the number of degenerate solutions to the number of non-degenerate ones. The analysis will be split into $n$ stages.

In the first stage, let us count the solutions for (25) satisfying $o_p(x^1) = l_1$ for some fixed $l_1 \geq 0$. By that we mean that $l_1$ is the largest integer such that $p^{l_1} | x^1_i$ for each $1 \leq i \leq n+1$. We can work with $r$ large enough so that $r \geq 2l_1 + 1$. Write $x^1 = p^{l_1} \tilde{x}^1$ where $\tilde{x}^1 \not\equiv 0 \mod p$. Note that the entry $\Lambda_{1,1}$ must be divisible by $p^{2l_1}$ since it equals $x^1 \cdot \tilde{x}^1 \mod p^r$. Similarly, $\Lambda_{1,j}$ must be divisible by $p^{l_1}$ for $j \geq 2$. Setting

$$
\Lambda_{1,1} = p^{2l_1} \tilde{\Lambda}_{1,1}, \quad \Lambda_{1,j} = p^{l_1} \tilde{\Lambda}_{1,j}, \quad j \geq 2
$$

we derive the new system of congruences with $(\tilde{x}^1, x^2, \ldots, x^n) \in \mathbb{Z}_p^{n+1} \times \mathbb{Z}_p^{n+1} \times \cdots \times \mathbb{Z}_p^{n+1}$

$$
(a) \quad \tilde{x}^1 \cdot \tilde{x}^1 \equiv \tilde{\Lambda}_{1,1} \mod p^{r-2l_1}
$$

$$
(b) \quad \tilde{x}^1 \cdot x^j \equiv \tilde{\Lambda}_{1,j} \mod p^{r-l_1}, \quad j \geq 2
$$

$$
(c) \quad x^i \cdot x^j \equiv \Lambda_{i,j} \mod p^r, \quad i, j \geq 2.
$$

Note that we also require $o_p(\tilde{x}^1) = 0$.

We argue that the number of solutions to the above system can be bounded by the maximum over all $0 \leq \tilde{\Lambda}_{1,1}, \tilde{\Lambda}_{1,2}, \ldots, \tilde{\Lambda}_{1,n} \leq p^r - 1$ satisfying

$$
\tilde{\Lambda}_{1,1} \equiv \tilde{\Lambda}_{1,1} \mod p^{r-2l_1}, \quad \tilde{\Lambda}_{1,j} \equiv \tilde{\Lambda}_{1,j} \mod p^{r-l_1}, \quad j \geq 2
$$

(27)
of the number of solutions of the system
\[
\mathcal{L}^* \mathcal{L} = \mathcal{L}' \mod p^r : \mathcal{L} = (x^1, \ldots, x^n) \in M_{n+1,n}(\mathbb{Z}_{p^r}), \quad o_p(x^1) = 0.
\] (28)

Here \(\mathcal{L}'\) is the symmetric matrix whose entries \(\Lambda'_{i,j} = \Lambda'_{j,i}\) have been defined in (27), while we set \(\Lambda'_{i,j} := \Lambda_{i,j}\) for the remaining pairs \((i,j)\). To see this we first note that the system (a) – (c) has \(p^{(n+1)/2}\) fewer solutions than the same system where \(x^1 \in \mathbb{Z}_{p^r}^{n+1}\) is replaced with \(x^1 \in \mathbb{Z}_{p^r}^{n+1}\) (we keep all the modular conditions unchanged). This follows since each \(\bar{x}_i \in \mathbb{Z}_{p^r-1}\) can be lifted in exactly \(p^1\) ways to some \(x_i^1 \in \mathbb{Z}_{p^r}\) with \(\bar{x}_i \equiv x_i^1 \mod p^r-1\).

Now the number of solutions to this new system is the sum over all \(\Lambda'_{1,1}, \ldots, \Lambda'_{1,n}\) as in (27) of the number of solutions of the system (28). It now suffices to note that there are exactly \(p^{(n+1)/2}\) choices for \(\Lambda'_{1,1}, \ldots, \Lambda'_{1,n}\), and to use the fact that the average is bounded by the maximum.

In the second stage of our reduction we fix \(0 \leq \Lambda'_{1,1}, \ldots, \Lambda'_{1,n} \leq p^r - 1\) as in (27) and count the number of solutions for (28). It suffices to focus attention on those particular solutions for which \(o_p(x^1 \wedge x^2) = l_2\) for fixed \(l_2 \geq 0\). By that we mean that \(l_2\) is the largest integer such that \(p^{l_2}\) divides the determinant of all the \(2 \times 2\) minors of the \((n + 1) \times 2\) matrix \([x^1, x^2]\). It follows that there must exist \(0 \leq t_{2,1} \leq p^{l_2} - 1\) and \(\bar{x}^2 \in \mathbb{Z}_{p^r}^{n+1}\) such that
\[
x^2 = t_{2,1}x^1 + p^{l_2}\bar{x}^2, \quad \text{with } o_p(x^1 \wedge \bar{x}^2) = 0.
\] (29)

Of course, to get this one relies crucially on the fact that \(o_p(x^1) = 0\). Also, we allow for \(l_2\) to be 0, in which case we can take \(t_{2,1} = 0, \bar{x}^2 = x^2\).

It suffices again to only consider \(r \geq 2l_2 + 1\).

Fix \(0 \leq t_{2,1} \leq p^{l_2} - 1\). We reformulate the system (28) using the variables \((x^1, x^2, \ldots, x^n) \in \mathbb{Z}_{p^r}^{n+1} \times \mathbb{Z}_{p^r-2}^{n+1} \times \mathbb{Z}_{p^r}^{n+1}\), with \(o_p(x^1) = o_p(x^1 \wedge x^2) = 0\)
\[
(a') \quad \bar{x} \cdot \bar{x}^2 \equiv \bar{\Lambda}_{2,2} \mod p^{r-2l_2}
\]
\[
(b') \quad x^1 \cdot x^2 \equiv \bar{\Lambda}_{i,2} \mod p^{r-l_2}, \quad i \neq 2
\]
\[
(c') \quad x^1 \cdot x^j \equiv \bar{\Lambda}_{i,j} \mod p^r, \quad i, j \neq 2
\]
where
\[
\begin{align*}
\bar{\Lambda}_{i,2} &= p^{-l_2}(\Lambda'_{i,2} - t_{2,1}\Lambda_{1,1}), \quad i \neq 2 \\
\bar{\Lambda}_{2,2} &= p^{-2l_2}(\Lambda'_{2,2} + t_{2,1}\Lambda_{1,1} - 2t_{2,1}\Lambda'_{1,2})
\end{align*}
\] (30)

Reasoning as we did in the previous stage, the number of solutions of the system (a') – (c') is bounded by the maximum over all \(0 \leq \Lambda''_{1,2}, \Lambda''_{2,2}, \ldots, \Lambda''_{n,2} \leq p^r - 1\) satisfying
\[
\Lambda''_{2,2} \equiv \bar{\Lambda}_{2,2} \mod p^{r-2l_2}, \quad \Lambda''_{i,2} \equiv \bar{\Lambda}_{i,2} \mod p^{r-l_2}, \quad \text{for } i \neq 2
\] (31)
of the number of solutions of the system
\[
\mathcal{L}^* \mathcal{L} = \mathcal{L}'' \mod p^r : \mathcal{L} = (x^1, \ldots, x^n) \in M_{n+1,n}(\mathbb{Z}_{p^r}), \quad o_p(x^1 \wedge x^2) = 0.
\] (32)

Here \(\mathcal{L}''\) is the symmetric matrix whose entries \(\Lambda''_{i,j} = \Lambda''_{j,i}\) have been defined in (31), while \(\Lambda''_{i,j} := \Lambda'_{i,j}\) for the remaining pairs \((i,j)\).

Before we go to the next stage, we bound the number of possible values for \(t_{2,1}\). First, we have the trivial bound \(p^{l_2}\). Also, since
\[
\Lambda'_{1,1} \equiv x^2 \cdot x^1 \mod p^r, \quad x^1 \cdot x^1 \equiv \Lambda'_{1,1} \mod p^r
\]
and since $p^j x^2$ is determined mod $p^r$, it follows that $t\Lambda'_{1,1}$ is determined mod $p^r$. But given the fact that $\Lambda'_{1,1}$ is determined mod $p^{r-2l_1}$, it follows that $t_{2,1}$ is determined modulo $p^{r-o_p(\Lambda_{1,1})-2l_1}$. Since also $0 \leq t_{2,1} \leq p^r - 1$, we get the upper bound $p^{r-o_p(\Lambda_{1,1})-2l_1}$ for the number of admissible values of $t_{2,1}$. A very similar reasoning will also produce the bound $p^{r-o_p(\Lambda_{1,j})-l_1}$ for $j \geq 2$. Combining the two bounds we get an upper bound $p^{\beta_2}$ for the number of admissible values of $t_{2,1}$, where

$$\beta_2 = \min\{l_2, \min_{1 \leq j \leq n} o_p(\Lambda_{1,j}) - l_1\}$$

We now begin the third stage of the reduction, which we hope will completely clarify the process. We will as before look for solutions for (32) which in addition satisfy $o_p(x^1 \land x^2 \land x^3) = l_3$ for fixed $l_3 \geq 0$. We can write

$$x^3 = t_{3,1} x^1 + t_{3,2} x^2 + p^j x^3 \quad (33)$$

where $o_p(x^1 \land x^2 \land x^3) = 0$ and $0 \leq t_{3,1}, t_{3,2} \leq p^j - 1$.

For such a solution to exist it must be that $\det(\Lambda'') \equiv 0 \mod p^j$. Using (26), (27), (30) and (31) we easily get that

$$l_1 + l_2 + l_3 \leq o_p(\det(\Lambda)). \quad (34)$$

Using (33) we get for each $1 \leq i \leq n$

$$\Lambda''_{i,3} = t_{3,1} \Lambda''_{i,1} + t_{3,2} \Lambda''_{i,2} \mod p^j \quad (35)$$

We now show how to bound the number of admissible pairs $(t_{3,1}, t_{3,2})$. First, there is the trivial bound $p^{2l_3}$.

Fix $1 \leq i \neq j \leq n$. We prove that the two equations (35) for $i$ and $j$ determine the pair $(t_{3,1}, t_{3,2}) \in \mathbb{Z}_{p^{l_3}} \times \mathbb{Z}_{p^{l_3}}$ up to at most $p^\alpha$ choices, where $p^\alpha$ is the largest power of $p$ that divides $\det[\Lambda''_{i,1} \quad \Lambda''_{i,2} \nabla \Lambda''_{j,1} \quad \Lambda''_{j,2}]$. We can assume this determinant to be nonzero, otherwise there is nothing to prove. Thus $(\Lambda''_{i,1}, \Lambda''_{i,2}) \neq (0, 0)$, and write $(\Lambda''_{i,1}, \Lambda''_{i,2}) = p^\alpha u$, where $u = (u_1, u_2)$ satisfies $o_p(u) = 0$ and $o \leq \alpha$. We can assume that $p^\alpha | \Lambda''_{j,1}$ and $p^\alpha | \Lambda''_{j,2}$; otherwise we do the argument for $(\Lambda''_{i,1}, \Lambda''_{i,2})$ instead. Note that in particular $o \leq \alpha/2$. We can now write as before

$$(\Lambda''_{j,1}, \Lambda''_{j,2}) = w u + p^{\alpha-o} v,$$

for some $w$ and $v = (v_1, v_2)$ with $\begin{bmatrix} u_1 & u_2 \\ v_1 & v_2 \end{bmatrix}$ nonsingular mod $p$. Note that $w$ must be divisible by $p^\alpha$. We thus get that

$$\begin{cases} u_1 t_{3,1} + u_2 t_{3,2} \equiv 0 \mod p^{j-\alpha} \\ v_1 t_{3,1} + v_2 t_{3,2} \equiv 0 \mod p^{j-\alpha+o} \end{cases}.$$

Since $\begin{bmatrix} u_1 & u_2 \\ v_1 & v_2 \end{bmatrix}$ is nonsingular mod $p$, the pair $(t_{3,1}, t_{3,2})$ will be uniquely determined in $\mathbb{Z}_{p^{j-\alpha}} \times \mathbb{Z}_{p^{j-\alpha+o}}$. Note that this can be lifted in exactly $p^\alpha$ ways to a $\mathbb{Z}_{p^{l_3}} \times \mathbb{Z}_{p^{l_3}}$ pair,
which proves the claim. It is easy to see as before that
\[ \alpha \leq o_p(\det \begin{bmatrix} \Lambda_{i,1} & \Lambda_{i,2} \\ \Lambda_{j,1} & \Lambda_{j,2} \end{bmatrix}) - l_1 - l_2. \]

Finally, by fixing \(0 \leq t_{3,2} \leq p^{l_3} - 1\), (35) for a fixed \(i\) will determine the value of \(t_{3,1}\) within \(p^{o_p(\Lambda_{i,1}) - l_1}\) possibilities.

Combining all three bounds derived above we get the upper bound \(p^{\beta_3}\) for the number of pairs \((t_{3,1}, t_{3,2})\), where
\[ \beta_3 = \min\{2l_3, \min_{1 \leq j \leq n} o_p(\Lambda_{1,j}) + l_3 - l_1, \min_{|A|=2} o_p(\mu_{\{1,2\}, A}) - l_2 - l_1\}. \]

It is now clear how to complete the remaining stages of the reduction. In the end we are left with counting non-degenerate solutions corresponding to fixed values of \(l_i, t_{i,j}\). As shown in the beginning of the proof, we have the bound \(O(p^{n(n+1)/2})\) for the number of these solutions. Also, the computations behind (34) easily extend to prove
\[ l_1 + \ldots + l_n \leq o_p(\det \Lambda). \]

The bound for the number of admissible tuples \((t_{i,1}, \ldots, t_{i,i-1})\) will follow as indicated before. This ends the proof of the proposition.

4.1. The four dimensional case. We start by proving part (a) of Theorem 4.1. Part (b) will be discussed in the next subsection.

Note that now \(m = 4, n = 3, \gamma = I_4\),
\[ \Lambda = \Lambda_{a,b} = \begin{bmatrix} \lambda & a & \lambda + a - b \\ a & \lambda & b \\ \lambda + a - b & b & \lambda \end{bmatrix} \]
and
\[ \nu_p = \nu_p(I_4, \Lambda) = \lim_{r \to \infty} \frac{1}{p^{6r}} |\{ \mathcal{L} \in M_{4,3}(\mathbb{Z}_{p^r}) : \mathcal{L}^* \mathcal{L} \equiv \Lambda \mod p^r \}|. \]

We will spend the rest of this subsection mainly evaluating \(\nu_p\). We will be interested only in values of \(a, b\) for which the equation \(\mathcal{L}^* \mathcal{L} = \Lambda_{a,b}\) has at least one solution \(\mathcal{L}\). In this case, it will be immediate that \(\Lambda_{a,b}\) is positive semi-definite, and in fact positive definite if its determinant \(2(b - \lambda)(a + \lambda)(a - b)\) is not zero. But then (15) will imply that
\[ N_{a,b,\lambda} \lesssim \prod_{p \text{ prime}} \lim_{r \to \infty} \frac{1}{p^{6r}} |\{ \mathcal{L} \in M_{4,3}(\mathbb{Z}_{p^r}) : \mathcal{L}^* \mathcal{L} \equiv \Lambda_{a,b} \mod p^r \}|. \]

We first note the easy estimate which takes care of the singular case
\[ \sum_{a=b \text{ or } b=\lambda \text{ or } a=-\lambda} |\mathcal{L} \in M_{4,3}(\mathbb{Z}_{p^r}) : \mathcal{L}^* \mathcal{L} \equiv \Lambda_{a,b} \mod p^r \}| \lesssim \lambda^{2+\epsilon}. \]

Let us see the \(a = b\) case, the other two cases are very similar. Note that if \(\mathcal{L} = (x, y, z)\) satisfies \(\mathcal{L}^* \mathcal{L} = \Lambda_{a,a}\) for some \(a\) then
\[ x \cdot (x - z) = y \cdot (x - z) = z \cdot (x - z) = 0, \]
which immediately implies that $x, y, z$ are linearly dependent. If $x$ and $y$ are fixed, then $z$ is hence constrained to a circle on $F_{4, \lambda}$ and can only take $O(\lambda')$ values. Note also that since $x, y \in F_{4, \lambda}$, there are $O(\lambda^{2+\epsilon})$ such pairs $(x, y)$.

We next focus on the nonsingular case. An immediate consequence of Proposition 4.6 is

**Proposition 4.7.** Assume $\lambda \not\in \{-a, b\}$ and $a \neq b$. If $p \mid \det(\Lambda_{a,b})$ then
\[
\nu_p \lesssim o_p((\det(\Lambda_{a,b}))^2 p^{o_p(\gcd(\lambda^2 - a^2, \lambda^2 - b^2))}).
\]

**Proof** First use the bounds $\beta_2 \leq l_2, \beta_3 \leq o_p(\lambda^2 - a^2) - l_1 - l_2$ and sum over $l_1 + l_2 + l_3 \leq o_p(\det(\Lambda_{a,b}))$ to get
\[
\nu_p \lesssim o_p((\det(\Lambda_{a,b}))^2 p^{o_p(\lambda^2 - a^2)}).
\]
Then repeat the argument with indices 1, 2 replaced by 2, 3.

We can now prove

**Corollary 4.8.** We have
\[
\sum_{|a|, |b| \leq \lambda, \ a \neq b, \lambda \not\in \{-a, b\}} N_{a,b,\lambda} \lesssim \lambda^{2+\epsilon}.
\]

**Proof** Using (36), (24), Proposition 4.2 and Proposition 4.7 we get that
\[
N_{a,b,\lambda} \lesssim \lambda' \gcd(\lambda^2 - a^2, \lambda^2 - b^2).
\]
Fix $1 \leq d \leq \lambda$. We rely on the fact that the number of divisors of an integer $l$ is $O(l')$. Then
\[
\sum_{s : s^2 \mid \lambda} \sum_{d_1, d_2 \mid \lambda, \gcd(d_1, d_2) = s} |\{k_1, k_2 \in \mathbb{Z}^2 : 1 \leq k_i \leq \frac{2\lambda}{d_i}, k_1d_1 + k_2d_2 = 2\lambda\}| =
\sum_{s : s^2 \mid \lambda} \sum_{d_1, d_2 \mid \lambda, \gcd(d_1, d_2) = s} \frac{2\lambda}{d_1d_2} \leq \sum_{s : s^2 \mid \lambda} \sum_{d_1 \mid \lambda} \frac{2s\lambda}{d} \lesssim \lambda' \sum_{s : s^2 \mid \lambda} \frac{2s\lambda}{d}.
\]
We now conclude that
\[
\sum_{|a|, |b| \leq \lambda} \gcd(\lambda^2 - a^2, \lambda^2 - b^2) \lesssim \sum_{1 \leq d \leq \lambda} d(2 + ||a| < \lambda : d | \lambda^2 - a^2||)^2 \lesssim
\sum_{1 \leq d \leq \lambda} d + \lambda' \sum_{s^2 \mid \lambda} s^2 \lambda^2 \frac{1}{d^2} \lesssim \lambda^2 + \lambda' \sum_{s \mid \lambda} \sum_{1 \leq d \leq \lambda} \frac{s^2 \lambda^2 d}{d} \lesssim \lambda^{2+\epsilon}.
\]
Note that Corollary 4.8 combined with (37) proves part (a) of Theorem 4.1.

4.2. The five dimensional case. We now discuss the proof of Theorem 4.1 (b). In this case $m = 5$, $n = 4$, $\gamma = I_5$,

$$\Lambda = \Lambda_{a,b,c,d} = \begin{bmatrix} a & c & a/2 & a/2 \\ c & b & b/2 & b/2 \\ a/2 & b/2 & \lambda & d \\ a/2 & b/2 & d & \lambda \end{bmatrix}.$$

and

$$\nu_p = \nu_p(I_5, \Lambda) = \lim_{r \to \infty} \frac{1}{p^{10r}} \left| \{ \mathcal{L} \in M_{5,4}(\mathbb{Z}_p^r) : \mathcal{L}^* \mathcal{L} \equiv \Lambda \mod p^r \} \right|.$$

We first analyze solutions corresponding to the degenerate cases, by which we mean $a \in \{0, 4\lambda\}$ or $b \in \{0, 4\lambda\}$ or $a = b$ or $\det(\Lambda_{a,b,c,d}) = 0$. Note first that each solution $(u, v, x, y)$ counting towards some $N_{a,b,c,d,\lambda}$ with $\det(\Lambda_{a,b,c,d}) = 0$ will necessarily satisfy $2 \leq \text{rank}[u, v, x, y] < 4$. The computations in section 5 combined with (5), (6) show that the sum over $a, b, c, d$ of the number of all solutions with $\text{rank}[u, v, x, y] = 2$ is in fact bounded by

$$\sum_{x,y \in F_{5,\lambda}} \sum_{x \neq y} \left| \{ (u, v) : u, v \in (x + F_{5,\lambda}) \cap (y + F_{5,\lambda}) \cap \langle x, y \rangle \} \right| \leq \sum_{x,y \in F_{5,\lambda}} \sum_{x \neq y} \left| \{ u : u \in (x + F_{5,\lambda}) \cap \langle x, y \rangle \} \right|^2 \lesssim \lambda^4 |F_{5,\lambda}|^2 \lesssim \lambda^{3+\epsilon}.$$

To count the solutions with $\text{rank}[u, v, x, y] = 3$, note first that since $x, y$ are linearly independent, we must have $\text{rank}[u, x, y] = 3$ or $\text{rank}[v, x, y] = 3$. By symmetry we focus on the first case. Reasoning as before, the sum over $a, b, c, d$ of the number of all such solutions is bounded by

$$\sum_{x,y \in F_{5,\lambda}} \sum_{x \neq y} |(x, y, u) \cap (x + F_{5,\lambda}) \cap (y + F_{5,\lambda})| \lesssim |F_{5,\lambda}|^2 \lambda^{1+\epsilon} \lesssim \lambda^{4+\epsilon}.$$

Next, we count the solutions corresponding to the remaining degenerate cases, under the additional assumption that now $\det(\Lambda_{a,b,c,d}) \neq 0$. If $a = 0$ then $u = 0$, so this corresponds to a zero determinant. If $a = 4\lambda$ then $|u| = 2\sqrt{\lambda}$. Note that in addition $u \in x + F_{5,\lambda} \subset F_{5,\lambda} + F_{5,\lambda}$ and these force $x = u$. Such a solution is again excluded, since it corresponds to a singular $\Lambda_{a,b,c,d}$.

To close the analysis of the degenerate cases, we count the contribution from the $a = b$ case. Note that we must have $|u| = |v|$ and $x \cdot (u - v) = y \cdot (u - v) = 0$. The corresponding contribution is bounded by

$$\lambda^{1/2+\epsilon} \sum_{x,y \in F_{5,\lambda}} |(x + F_{5,\lambda}) \cap (y + F_{5,\lambda})| \lesssim \lambda^{1/2+\epsilon} \mathbb{E}(|F_{5,\lambda}|) \lesssim \lambda^{4+\epsilon},$$
where for the last inequality we used (53).

We begin the analysis of the non-degenerate case by recording the following consequence of Proposition 4.2

**Proposition 4.9.** Assume $p$ is not a factor of $\det(\Lambda_{a,b,c,d}) \neq 0$. Then

$$\nu_p \leq 1 + \frac{C}{p^2},$$

where $C$ is independent of $p, a, b, c, d, \lambda$.

Note as before that the product of these $\nu_p$ is $O(1)$.

Also, by using the bounds

$$\beta_2 \leq l_2$$
$$\beta_3 \leq l_3 + \min_{|A|=1} \omega_p(\mu_{\{1\},A}) - l_1$$
$$\beta_4 \leq \min_{|A|=3} \omega_p(\mu_{\{1,2,3\},A}) - l_1 - l_2 - l_3$$

in Proposition 4.6, and by permuting indices we get

**Proposition 4.10.** Assume $\Lambda := \Lambda_{a,b,c,d}$ is nonsingular and $p \mid \det(\Lambda)$. Then

$$\nu_p \lesssim \omega_p(\det(\Lambda))^3 \omega_p(\gcd(\lambda, \lambda)) \omega_p(\min_{A,B \subset \{1,2,3,4\}} \omega_p(\mu_{A,B})).$$

Since two of the rows of $\Lambda$ contain (divisors of) both $a$ and $\lambda$, we conclude

$$\nu_p \lesssim \omega_p(\det(\Lambda))^3 p \omega_p(\gcd(a, \lambda)) \omega_p(\min_{A,B \subset \{1,2,3,4\}} \omega_p(\mu_{A,B})).$$

By using various choices for $A, B$, then invoking Propositions 4.9, 4.10 and equations (24), (15) we conclude that whenever $\det(\Lambda_{a,b,c,d}) \neq 0$

$$N_{a,b,c,d,\lambda} \lesssim \lambda^4 \gcd(a, \lambda) \cdot \gcd(\lambda(ab-c^2) + \frac{ab}{4}(2c-a-b), d(ab-c^2) + \frac{ab}{4}(2c-a-b), a(\lambda-d)(b-c), b(\lambda-d)(a-c)) \leq \lambda^4 \gcd(a, \lambda) \gcd((39), (40), (41), (42))$$

where

$$\lambda(ab-c^2) + \frac{ab}{4}(2c-a-b)$$  \hspace{1cm} (39)

$$\lambda - d)(ab - c^2)$$  \hspace{1cm} (40)

$$a(\lambda-d)(b-c)$$  \hspace{1cm} (41)

$$b(\lambda-d)(a-c).$$  \hspace{1cm} (42)
We use $\mathcal{A}$ to denote the non degenerate four-tuples $(a, b, c, d)$ with $|a|, |b|, |c|, |d| \lesssim \lambda$, $a, b \notin \{0, 4\lambda\}$, $a \neq b$ and $\det(\Lambda_{a,b,c,d}) \neq 0$. To finish the proof of part (b) of Theorem 4.1 we are left with evaluating

$$\sum_{(a,b,c,d)\in \mathcal{A}} \gcd(a, \lambda) \prod_{p|\lambda} p^{\nu_p(\gcd((39),(40),(41),(42)))} \prod_{p|\lambda} p^{\nu_p(\gcd((39),(40),(41),(42)))}.$$ 

This can be trivially bounded by

$$\sum_{k_0|\lambda} k_0 k_1 k_2 \sum_{(a,b,c,d)\in \mathcal{A}} 1.$$ 

Here and in the future we denote by $\tilde{k}$ the product of all primes dividing $k$. Note that if $(a, b, c, d) \in \mathcal{A}$ then the two terms (41)-(42) can not be zero simultaneously, which justifies the finiteness restriction $k_1, k_2 \lesssim \lambda^3$. Indeed, $d \neq \lambda$ since $\det(\Lambda_{a,b,c,d}) \neq 0$. Also $a, b \neq 0$ and moreover $b - c$ and $a - c$ can not be both zero since we have assumed $a \neq b$.

Note that the number of integers $k_0$ and $k_1$ in the sum is $O(\lambda^6)$ and that $(k_0 k_1, k_2) = 1$. Since $k_1(|\lambda - d| b(a - c)$, there exists a decomposition

$$k_1 = k'_1 k''_1 k'''_1$$

with

$$k'_1|\lambda - d, \ k''_1|b, \ k'''_1|a - c.$$ 

We further bound the sum by

$$\sum_{k_0|\lambda} k_0 k_1 \sum_{k_1 \leq \lambda^3} \sum_{a,b} k_0 a, k'_1 b \sum_{k_2 \leq \lambda^3} \sum_{c,d} k_2 c, d (a,b,c,d) \in \mathcal{A}.$$ 

Note that for each $k_2, a, b, c, d$ contributing to the sum there must exist a decomposition $k_2 = k'_2 k''_2$ such that

$$k'_2|\lambda - d$$

$$k''_2|ab - c^2, \ k''_2|a(b - c), \ k''_2|b(a - c).$$

This gives rise to the new bound

$$\sum_{k_0|\lambda} k_0 k_1 \sum_{k_1 \leq \lambda^3} \sum_{a,b} k_0 a, k'_1 b \sum_{k_2 \leq \lambda^3} \sum_{c,d} k_2 c, d (a,b,c,d) \in \mathcal{A}.$$ 

Equation (39) written as a quadratic polynomial in $c$ is

$$4\lambda c^2 - 2abc + (a^2 b + ab^2 - 4\lambda ab) \equiv 0 \mod k_2.$$ 

The discriminant is $D_{a,b} = 4ab(4\lambda - a)(4\lambda - b)$. For $a, b$ fixed we run a second decomposition for $k_2$ namely

$$k_2 = k_{2,1,a,b} k_{2,2,a,b}.$$
with
\[ (k_{2,1,a,b}, k_{2,2,a,b}) = 1, \quad (k_{2,1,a,b}, D_{a,b}) = 1, \quad k_{2,2,a,b}|D_{a,b}. \] (50)

Note that this decomposition is unique, since \( D_{a,b} \neq 0 \).

For each \( a, b, k'_2, k''_2 \) as above, define
\[ k'_{2,1,a,b} = \gcd(k'_2, k_{2,1,a,b}), \quad k'_{2,2,a,b} = \gcd(k'_2, k_{2,2,a,b}), \]
\[ k''_{2,1,a,b} = \gcd(k''_2, k_{2,1,a,b}), \quad k''_{2,2,a,b} = \gcd(k''_2, k_{2,2,a,b}), \]
and note that due to (43), (49), (50) we have
\[ \lambda \sum_{k_0 k_1 \mid \lambda} k_0 k_1 \sum_{k'_{2,1,a,b}} \sum_{k''_{2,1,a,b}} \sum_{k'_{2,2,a,b}} \sum_{k''_{2,2,a,b}} \gcd(a, b, k''_{2,2,a,b}) = \frac{k'_{2,1,a,b} k''_{2,1,a,b} k'_{2,2,a,b} k''_{2,2,a,b} k'_1 k''_1 k''_2}{\gcd(a,b,k''_{2,2,a,b})}, \]
where we have used that \( \gcd(ab - c^2, a(b - c), b(a - c))|ab(a - b) \) and the various decompositions for \( k_1, k_2 \).

Since \( \gcd(\lambda, k_2) = 1 \), we can choose a large enough integer \( M \) with \( \gcd(M, \lambda) = 1 \) such that \( \gcd(a, b, k''_{2,2,a,b}) \leq \gcd(a, b, M) \) for all admissible \( a, b, k''_{2,2,a,b} \). We can now bound the sum above by
\[ \lambda \sum_{k_0 k_1 \mid \lambda} k_0 \sum_{k'_1} \sum_{k''_1} \gcd(a, b, M) \sum_{k'_{2,1,a,b}} \sum_{k''_{2,1,a,b}} \gcd(k_{2,1,a,b}, k''_{2,1,a,b}) \] (51)
We next fix \(a, b\) and evaluate
\[
\sum_{k_2 \leq \lambda^3, \lambda} \sum_{k_2' \geq k_2, k_2'' \geq k_2, k_2'' \neq k_2} \frac{\gcd(k_{2,1}, a, b, k_2'')}{k_{2,1}}.
\]

By using the divisor function bound and the divisibility relations \(k_2'' \mid ab(a-b)\) and \(k_{2,1} \mid D_{a,b}\) we get that \(k_2''\) and \(k_{2,1} \mid D_{a,b}\) are both determined within \(O(\lambda')\) values. Call \(A''\) and \(A_{2,2}\) the corresponding set of admissible values. On the other hand, fixing \(k_{2,1} \mid D_{a,b}\) and \(1 \leq k_{2,1} \leq \lambda^3\) will certainly uniquely determine both \(k_2\) and \(k_2''\). We thus can write
\[
\sum_{k_2' \in A''} \sum_{k_2'' \in A_{2,2}} \sum_{1 \leq k_{2,1} \leq \lambda^3} \frac{\gcd(k_{2,1}, k_2'')}{k_{2,1}}.
\]

Note however that with \(k_2''\) fixed,
\[
\sum_{1 \leq k_{2,1} \leq \lambda^3} \frac{\gcd(k_{2,1}, k_2'')}{k_{2,1}} \leq \sum_d \sum_{1 \leq m \leq \lambda^3/d} \frac{1}{md} \lesssim \lambda'.
\]

Thus
\[
(51) \lesssim \lambda^{2+\varepsilon} \sum_{k_0 \leq \lambda} k_0 \sum_{k_1' \leq \lambda} k_1' \sum_{k_1'' \leq \lambda} \sum_{k_0 \leq \lambda} \gcd(a, b, M).\]

Observe now that \(k_0, k_1\) and thus also \(k_1', k_1'', k_1''\) can take \(O(\lambda')\) values. Note also that \(\gcd(a, b, M) \leq \gcd(\frac{a}{k_0}, \frac{b}{k_0})\) since \(\gcd(M, \lambda) = 1\). For fixed \(k_0, k_1''\) we estimate
\[
k_0 k_1'' \sum_{|a|, |b| \leq \lambda} \gcd(a, b, M) \leq k_0 k_1'' \sum_{|a|, |b| \leq \lambda} \gcd(\frac{a}{k_0}, \frac{b}{k_0}) \leq \lambda^{2+\varepsilon}.
\]

We can now finish the argument by observing that
\[
\lambda^{2+\varepsilon} \sum_{k_1 \leq \lambda^3} k_0 \sum_{k_1' \leq \lambda} k_1' \sum_{k_1'' \leq \lambda} \gcd(a, b, M) \lesssim \lambda^{4+\varepsilon}.
\]

5. Energy estimates using Siegel’s mass formula

In this section we show how Theorem 4.1 produces a different type of upper bound for the additive energy. When \(n = 4\) this method seems to only work for the whole \(\mathcal{F}_{4,\lambda}\).

**Theorem 5.1.** We have
\[
\mathbb{E}(\mathcal{F}_{4,\lambda}) \lesssim_{\varepsilon} N^{4+\varepsilon}
\]
Proof. Note that we need to count the number of quadruples \((x, y, z, w) \in (\mathbb{Z}^4)^4\) that satisfy
\[
\begin{align*}
x + y &= z + w \\
|x|^2 &= |y|^2 = |z|^2 = |w|^2 = \lambda.
\end{align*}
\]
Since
\[
\lambda = |x + y - z|^2 = 3\lambda + 2x \cdot y - 2x \cdot z - 2y \cdot z,
\]
it turns out that
\[
E(F_4, \lambda) = \left| \left\{ (x, y, z) \in (\mathbb{Z}^4)^3 : |x|^2 = |y|^2 = |z|^2 = \lambda, -x \cdot y + x \cdot z + y \cdot z = \lambda \right\} \right|
\]
\[
= \sum_{a, b \in \mathbb{Z} \atop |a|, |b| \leq \lambda} \left| \left\{ (x, y, z) \in (\mathbb{Z}^4)^3 : |x|^2 = |y|^2 = |z|^2 = \lambda, x \cdot y = a, y \cdot z = b, x \cdot z = \lambda + a - b \right\} \right|.
\]
Note that a triple \((x, y, z)\) as above satisfies
\[
\begin{pmatrix}
x_1 & x_2 & x_3 & x_4 \\
y_1 & y_2 & y_3 & y_4 \\
z_1 & z_2 & z_3 & z_4
\end{pmatrix}
\begin{pmatrix}
x_1 & y_1 & z_1 \\
x_2 & y_2 & z_2 \\
x_3 & y_3 & z_3 \\
x_4 & y_4 & z_4
\end{pmatrix}
= \begin{pmatrix}
\lambda & a & \lambda + a - b \\
& a & \lambda & b \\
& & \lambda + a - b & \lambda
\end{pmatrix}.
\]
Now Theorem 4.1 will provide the bound \(O(\lambda^{2+\epsilon})\) for the sum above.

Remark 5.2. It is easy to see that (apart from \(\epsilon\)) the bound (52) is sharp, if no further restriction is placed on \(\lambda\). Indeed,
\[
E(F_4, \lambda) = \| \sum_{\xi \in F_4, \lambda} e(x \cdot \xi) \|_4^4.
\]
Since \(|e(x \cdot \xi) - 1| < 1/2\) if \(|x| \ll \frac{1}{N}\) and \(\xi \in F_4, \lambda\), it follows that
\[
\left| \sum_{\xi \in F_4, \lambda} e(x \cdot \xi) \right| \gtrsim |F_4, \lambda|
\]
if \(|x| \ll \frac{1}{N}\). It now suffices to choose \(\lambda\) such that \(|F_4, \lambda| \gtrsim N^2\).

Let us now switch attention to five dimensions. Note that the proof of (52) combined with the case \(m = 5, n = 3\) in Siegel’s mass formula proves
\[
E(F_5, \lambda) \lesssim N^7.
\]
Alternatively, one could count the number of solutions of
\[
\{(x, y, z, w) \in F_5^4 : x + y = z + w\}
\]
by fixing \(x_5, y_5, z_5\) (there \(O(N^3)\) ways) and then applying the bound in (52).

While as observed below (53) is sharp, we can gain slightly more by applying Siegel’s mass formula with \(m = 5, n = 4\).

Theorem 5.3. For each \(\Lambda \subset F_5, \lambda\) we have
\[
E(\Lambda) \lesssim N^{4+\epsilon}|\Lambda|
\]
Proof. Note that
\[
\mathbb{E}(\Lambda) = \sum_{x, y \in \Lambda} |(\Lambda + x) \cap (\Lambda + y)| \leq \|\Lambda\|^2 + \Lambda \left[ \sum_{x \neq y \in \Lambda} |(\Lambda + x) \cap (\Lambda + y)|^2 \right]^{1/2} =
\]
\[
|\Lambda|^2 + |\Lambda| \|\{(u, v, x, y) \in Z^5 \times Z^5 \times F_{5, \lambda} \times F_{5, \lambda} : x \neq y, u - x, v - x, u - y, v - y \in F_{5, \lambda}\}|^{1/2}.
\]
Note that each \((u, v, x, y)\) as above we have
\[
|u|^2 = 2x \cdot u = 2y \cdot u \quad \text{and} \quad |v|^2 = 2x \cdot v = 2y \cdot v.
\]
It thus follows that
\[
|\{(u, v, x, y) \in Z^5 \times Z^5 \times F_{5, \lambda} \times F_{5, \lambda} : x \neq y, u - x, v - x, u - y, v - y \in F_{5, \lambda}\}| \leq \sum_{a, b, c, d \in Z} |\{a, b, c, d \leq \lambda : (u, v, x, y) \in (Z^5)^4 : x \neq y, |u|^2 = a, |v|^2 = b, |x|^2 = |y|^2 = \lambda, u \cdot v = c, x \cdot y = d, u \cdot x = u \cdot y = \frac{a}{2}, v \cdot x = v \cdot y = \frac{b}{2}\}|.
\]
The result follows by invoking part (b) of Theorem 4.1.

**Remark 5.4.** A computation similar to the one in Remark 5.2 combined with (6) shows that
\[
\mathbb{E}(F_{5, \lambda}) \gtrsim N^7,
\]
this time for each \(\lambda > 0\). This shows that Theorem 5.3 is sharp, in the sense that one cannot lower one of the exponents in either \(N^4\) or \(|\Lambda|^4\) from (54), without increasing the other one.

In the next section we will combine the estimates for the energy obtained here with the different type of estimates we have derived using incidence theory.

### 6. Proof of Theorem 1.2

Fix \(\|a_\xi\|_2 = 1\) and define
\[
F(x) = \sum_{\xi \in F_{n, \lambda}} a_\xi e(x \cdot \xi).
\]
We start by recalling the following estimate (24) from [7].

**Proposition 6.1.** For \(\alpha > N^{\frac{n-1}{n} + \epsilon}\) we have
\[
|\{|F| > \alpha\}| \lesssim \alpha^{-\frac{n-1}{n-3}} N^{\frac{2}{n-3}}. \tag{55}
\]

We now work the details for Theorem 1.2 in the case \(n = 4\) and then briefly explain how to modify the argument when \(n = 5\). First note that (7) and (52) imply that
\[
\mathbb{E}(\Lambda) \lesssim N^{\frac{4}{7} + \epsilon} |\Lambda|^2 \tag{56}
\]
for each \(\Lambda \subset F_{4, \lambda}\). This is equivalent with the fact that the linear operator
\[
T((a_\xi)_{\xi \in F_{4, \lambda}}) = \sum_{\xi \in F_{4, \lambda}} a_\xi e(x \cdot \xi)
\]
has a restricted weak type bound $O(\sqrt[4]{1 + \epsilon})$ when acting
\[ T : l^2(F_{4,\lambda}) \rightarrow L^4(T^4). \]
In other words, for each $\Lambda \subset F_{4,\lambda}$ and each $|a_\xi| \leq 1_\Lambda(\xi)$ we have
\[ \|T((a_\xi)_{\xi \in F_{4,\lambda}})\|_{L^4} \lesssim \epsilon \sqrt[4]{1 + \epsilon} |\Lambda|^{1/2}. \]
It is very easy to convert this into a strong bound. Note that, say
\[ \|T((a_\xi)_{\xi \in F_{4,\lambda}})\|_{L^5} \leq \|T((a_\xi)_{\xi \in F_{4,\lambda}})\|_{L^\infty} \leq N^2 \|a_\xi\|_{L^\infty} \leq N^2 \|a_\xi\|_{l^2}. \]
Restricted type interpolation now shows that for each $\epsilon$
\[ \|T((a_\xi)_{\xi \in F_{4,\lambda}})\|_{L^{4+\epsilon}} \lesssim \epsilon N^{4/3+\epsilon} \|a_\xi\|_{L^{2+\epsilon}}, \]
where $\epsilon', \epsilon'' \rightarrow 0$ as $\epsilon \rightarrow 0$. This trivially implies that for each $\epsilon > 0$
\[ \|T((a_\xi)_{\xi \in F_{4,\lambda}})\|_{L^4} \lesssim \epsilon N^{4/3+\epsilon} \|a_\xi\|_{L^2}. \]
We conclude that for $\alpha > 0$
\[ \{|\{F\} > \alpha\} \lesssim \epsilon \alpha^{-4} N^{4/3+\epsilon}. \]
Next note that (55) gives
\[ \{|\{F\} > \alpha\} \lesssim \alpha^{-6} N^2, \quad \alpha \gtrsim N^{3/4} \]
Combining (57) with (58), we get for $p > 6$
\[ \int_{T^4} |F|^p = \int_{N^{3/4} \leq |F| \leq N^{1+\epsilon}} |F|^p + \int_{|F| \leq N^{3/4}} |F|^p \lesssim \epsilon \]
\[ N^{p-4+\epsilon} + N^{4/3(p-4)} \int_{T^4} |F|^4 \lesssim \epsilon N^{p-4+\epsilon} + N^{4/3(p-4)+\epsilon}. \]
It suffices now to note that this is bounded by $N^{p-4+\epsilon}$ when $p \geq \frac{44}{7}$.
When $n = 5$ we will rely instead on the sharp $L^4$ estimate that follows from (55)
\[ \{|\{F\} > \alpha\} \lesssim \alpha^{-4} N, \quad \alpha \gtrsim N. \]
Also (11) and (54) give
\[ \mathbb{E}(\Lambda) \lesssim \epsilon N^{4/3+\epsilon}|\Lambda|^2 \]
for each $\Lambda \subset F_{5,\lambda}$. 

7. Possible further improvements

Our incidence theory approach relies on two ingredients. The main one is the hyperplane-point incidence theorem, which exploits the fact that quadruples contributing to the additive energy of the sphere concentrate on hyperplanes. This theorem relies crucially on the topology of $\mathbb{R}^n$, as can be seen in the proof of the Cutting Lemma 2.4. The second ingredient is the fact that circles contain a negligible number of lattice points, which seems to be a rather weak use of the fact that our points lie on the sphere. It may be possible that by using finer properties about the distribution of lattice points on spheres, one might gain additional information about the relevant hyperplanes, and possibly further improve the estimates on the energy.

In light of the sharp subcritical estimate from [6], one might wonder whether further progress is possible in the supercritical regime of Conjecture 1.1 by methods that completely avoid number theory. We believe the answer is yes. It seems natural to conjecture that for $p \geq \frac{2(n+1)}{n-1}$

$$\| \sum_k f_k \|_p \lesssim \delta^{-\frac{n+1}{2p} + \frac{2n+1}{np} - \epsilon} (\sum_k \|f_k\|_p^2)^{1/2},$$

for each partition of the unit sphere $S^{n-1}$ into $\delta^{1/2} - \text{caps } C_k$, and each $\hat{f}_k$ supported on a $\delta$ neighborhood of $C_k$. See (1.5) and (1.8) in [10] for some partial results in this direction, in the more general (and difficult) case of cones. If this conjecture is indeed true, its proof would naturally not involve any number theory. Moreover, using $\delta = N^{-2}$, it would imply via the dilation argument and the use of Dirac deltas as in [6] that

$$\| \sum_{\xi \in F_{n,\lambda}} a_{\xi} e(\xi \cdot x) \|_{L^p(T^n)} \lesssim \epsilon \| a_{\xi} \|_{l^2(F_{n,\lambda})},$$

for each $a_{\xi} \in \mathbb{C}$, $\epsilon > 0$ and each $p \geq \frac{2(n+1)}{n-1}$. On the other hand, if we assume (60) for the critical index $p = \frac{2(n+1)}{n-1}$, and if we combine this (55) as in section 6, we further improve the range in Conjecture 1.1 to $p \geq 6$ when $n = 4$ and $p \geq 4$ when $n = 5$.

We mention as a side remark that (59) is expected to be true in the case of the (truncated) paraboloid. If indeed true, this would in turn completely solve the discrete analog for the paraboloid considered in [4].

This discrepancy between the sphere and the paraboloid in the discrete world is due to the non uniform distribution of lattice points on the sphere. It is likely that to detect these irregularities and get the full range in Conjecture 1.1, some involved number theory will be needed. One step in this direction is made by our use of Siegel’s mass formula, which produces sharp results for the energy of the whole sphere. Another possible avenue is described in the last section of [7].

References

[1] Bombieri, E. and Pila, J. *The number of integral points on arcs and ovals*, Duke Math. J. 59 (1989), no. 2, 337-357
[2] Bombieri, E. and Bourgain, J. *A problem on sums of two squares*, to appear in IMRN
[3] Bourgain, J. *Eigenfunction bounds for the Laplacian on the n-torus*, Internat. Math. Res. Notices (1993), no. 3, 61-66.
[4] Bourgain, J. 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.

[5] Bourgain, J. Analysis results and problems related to lattice points on surfaces, Harmonic analysis and nonlinear differential equations (Riverside, CA, 1995), 85-109, Contemp. Math., 208, Amer. Math. Soc., Providence, RI, 1997.

[6] Bourgain, J. Moment inequalities for trigonometric polynomials with spectrum in curved hypersurfaces, Israel J. Math. 193 (2013), no. 1, 441-458.

[7] Bourgain, J. and Demeter, C. Improved estimates for the discrete Fourier restriction to the higher dimensional sphere, preprint (2012) available on arxiv, to appear in Illinois Journal of Mathematics.

[8] Cassels, J. W. S. Rational quadratic forms, London Mathematical Society Monographs, 13. Academic Press, Inc., London-New York, 1978.

[9] Elekes, G. and Szabó, E. How to find groups?(and how to use them in Erdos geometry), Combinatorica 32 (2012), no. 5, 537-571.

[10] Garrigós, G., Schlag, W. and Seeger, A. Improvements in Wolff’s inequality for decompositions of cone multipliers, preprint available online.

[11] Greenberg, M. J. Lectures on forms in many variables, W. A. Benjamin, Inc., New York-Amsterdam 1969 v+167 pp.

[12] Grosswald, E. Representations of integers as sums of squares, Springer-Verlag, 1985.

[13] Lewko, M. New restriction estimates for the 3-d paraboloid over finite fields, available on arxiv.

[14] Matousek, J. Lectures on Discrete Geometry, Springer-Verlag New York 2002.

[15] Siegel, L. C. Lectures on the analytical theory of quadratic forms: second term 1934-35, The Institute for Advanced Study and Princeton University, revised edition 1949, reprinted January 1955.

[16] Wolff, T. Local smoothing type estimates on $L^p$ for large $p$, Geom. Funct. Anal. 10 (2000), no. 5, 1237-1288.

[17] Wooley, T. D. A note on simultaneous congruences, J. Number Theory 58 (1996), no. 2, 288-297.

School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540
E-mail address: bourgain@math.ias.edu

Department of Mathematics, Indiana University, 831 East 3rd St., Bloomington IN 47405
E-mail address: demeterc@indiana.edu