University of Nevada, Reno

Lattice Point Problems about the Paraboloid

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Mathematics

by

Zachary Miller

Dr. Jing-Jing Huang/Thesis Advisor

August, 2021

THE GRADUATE SCHOOL

We recommend that the thesis
prepared under our supervision by

ZACHARY MILLER

entitled

Lattice Points Problems about the Paraboloid

be accepted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

Jing-Jing Huang, PhD.
Advisor

Valentin Deaconu
Committee Member

Matthew Aguirre
Graduate School Representative

David W. Zeh, Ph.D., Dean
Graduate School

August, 2021

Abstract

The Gauss circle problem, which asks for the best possible error term when approximating the number of lattice points inside a dilating circle centered at the origin by its area, is a longstanding open question in number theory. One may as well ask similar questions for regions bounded by other conics such as hyperbola and parabola, or their higher dimensional generalizations. Building off of the techniques of Huang and Li, we establish in this thesis asymptotic formulae for the number of lattice points under and near the standard paraboloid of dimension two and higher. The upper bound estimates we obtain on the error terms nearly meet those in the omega result of Chamizo and Pastor, and therefore are essentially best possible.

Acknowledgements

I am indebted to and owe great thanks to my advisor Jing-Jing Huang. His expert advice, patience and encouragement were invaluable to me during the work and completion of this thesis.

I am also grateful to my committee, family, and friends.

Contents

List of Symbols

$\|\cdot\|$		$\|x\| = \min_{k\in\mathbb{Z}} \|x-k\|$
$(\cdot,\cdot)$	The greatest common divisor	
$O(\cdot)$	Landau's notation	$f(x) = O(g(x))$ if $\limsup_{x\to\infty} \frac{\|f(x)\|}{g(x)} < \infty$
$\ll$	Vindogradov's notation	$f(x) \ll g(x) \iff f(x) = O(g(x))$
$\Omega(\cdot)$	Hardy-Littlewood's notation	$f(x) = \Omega(g(x))$ if $\limsup_{x\to\infty} \left\|\frac{f(x)}{g(x)}\right\| > 0$
$\exp(\cdot)$	The exponential function	$\exp(x) := e^x$
$e(\cdot)$		$e(\theta) := \exp(2\pi i\theta)$
$\lfloor\cdot\rfloor$	The floor function	Round down to the closest integer
$\{\cdot\}$	The fractional part	$\{x\} := x - \lfloor x\rfloor$

1 Introduction

Let $\mathcal{K}$ be a bounded region in $\mathbb{R}^d$. It is a classical question in number theory to study the number of lattice points (i.e. elements of $\mathbb{Z}^d$) within a dilation of this region $q\mathcal{K}$ ($q \geq 1$). Hence we are naturally led to study the counting function

$$N_{\mathcal{K}}(q) \coloneqq \#q\mathcal{K} \cap \mathbb{Z}^d.$$

If we consider each lattice point as the center of a cube with sides of length 1 parallel to the axis, then finding $N_{\mathcal{K}}(q)$ is equivalent to finding the total area of the squares corresponding to those lattice points inside $q\mathcal{K}$, which for large enough q is approximated very well by the volume of $q\mathcal{K}$. In other words we expect that as $q \to \infty$

$$N_{\mathcal{K}}(q) \sim |\mathcal{K}|q^d,$$

where $|\mathcal{K}|$ stands for the volume of $\mathcal{K}$. Therefore, it is very natural to ask how good this approximation is. To that end, we may consider the error term

$$E_{\mathcal{K}}(q) = N_{\mathcal{K}}(q) - |\mathcal{K}|q^d.$$

The smaller the error term, the better our approximation.

Friedrich Gauss initially studied this lattice point problem when $\mathcal{K} = \mathcal{C}$, where $\mathcal{C}$ is the unit disk centered at the origin $\mathcal{C} = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$. In this case, Gauss' circle problem asks to count the number of lattice points (a_1, a_2) that are contained in the homothetically dilated disk $q\mathcal{C}$, that is,

$$N_{\mathcal{C}}(q) = \#\{(a, b) \in \mathbb{Z}^2 \mid a^2 + b^2 \leq q^2\},$$

and the corresponding error term is $E_{\mathcal{C}}(q) = N_{\mathcal{C}}(q) - \pi q^2$. In his efforts, Gauss proved that

$$|E_{\mathcal{C}}(q)| \leq 2\sqrt{2}\pi q.$$

Gauss's argument is based on the following simple geometric observation. Let $S(q)$ be the union of all unit squares centered at lattice points inside the disk $q\mathcal{C}$, then

$$\left(q - \frac{\sqrt{2}}{2}\right)\mathcal{C} \subseteq S(q) \subseteq \left(q + \frac{\sqrt{2}}{2}\right)\mathcal{C}.$$

Let us expound the reasoning behind this observation. Note that the distance between any point in $\mathbb{R}^2$ to the nearest lattice point is at most $\frac{\sqrt{2}}{2}$. For any $\boldsymbol{x} \in \left(q - \frac{\sqrt{2}}{2}\right)\mathcal{C}$, let $\boldsymbol{a}$ be its nearest lattice point, then by the triangle inequality, we see that

$$|\boldsymbol{a}| \leq |\boldsymbol{a} - \boldsymbol{x}| + |\boldsymbol{x}| \leq \frac{\sqrt{2}}{2} + \left(q - \frac{\sqrt{2}}{2}\right) = q,$$

or equivalently $\boldsymbol{a} \in q\mathcal{C}$. Since $\boldsymbol{x}$ lies in the unit square centered at $\boldsymbol{a}$, this means that $\boldsymbol{x} \in S(q)$, which explains the first inclusion above. Similarly, let $\boldsymbol{x} \in S(q)$, then by definition we can find a lattice point $\boldsymbol{a} \in q\mathcal{C}$ such that $|\boldsymbol{x} - \boldsymbol{a}| \leq \frac{\sqrt{2}}{2}$. Consequently by the triangle inequality again we have

$$|\boldsymbol{x}| \leq |\boldsymbol{x} - \boldsymbol{a}| + |\boldsymbol{a}| \leq \frac{\sqrt{2}}{2} + q,$$

which gives the second inclusion above.

Now equipped with this inclusion relation, we immediately obtain

$$\left|\left(q - \frac{\sqrt{2}}{2}\right)\mathcal{C}\right| \leq |S(q)| \leq \left|\left(q + \frac{\sqrt{2}}{2}\right)\mathcal{C}\right|,$$

which reduces to

$$\pi\left(q-\frac{\sqrt{2}}{2}\right)^2 \leq N_{\mathcal{C}}(q) \leq \pi\left(q+\frac{\sqrt{2}}{2}\right)^2.$$

After expanding the squares and subtracting the main term πq^2, we arrive at

$$|E_{\mathcal{C}}(q)| \leq \pi\left(\sqrt{2}q+\frac{1}{2}\right) < 2\sqrt{2}\pi q$$

when $q \geq 1$.

We may rephrase Gauss' bound as $E_{\mathcal{C}}(q) = O(q)$ to suppress the coefficient since it plays a lesser role in this estimate. As one can easily perceive in the above argument, this estimate essentially relies on the fact that there are $O(q)$ many unit squares centered at lattice points which intersect with $q\mathcal{C}$. However, while some of these squares are included in $S(q)$, some are not! Therefore, if this rounding error problem exhibits random behavior one should expect the error term be much smaller than $O(q)$; indeed, it has been famously conjectured that for any $\varepsilon > 0$

$$E_{\mathcal{C}}(q) = O(q^{\frac{1}{2}+\varepsilon}).$$

Throughout the thesis, we use ε to denote a fixed positive real number, which value may not necessarily be the same at each occurrence.

Further improvement on Gauss's bound took over a century when finally Voronoi (1903) showed $E_{\mathcal{C}}(q) = O(q^{2/3})$. Over the next century the progress slowed down and the state-of-the-art bound is $O(q^{517/824})$ thanks to Bourgain and Watt [2]. It would not be an understatement to say this is an extremely difficult problem!

As work on this problem continued, generalizations were beginning form. Nowadays, the lattice point problem can be studied in any bounded region $\mathcal{K} \subset \mathbb{R}^d$, but

the cases that $\mathcal{K}$ is a region bounded by conics or their higher dimensional analogues continue to serve as pivotal models and therefore are of fundamental importance. For example, Dirichlet has studied a companion of Gauss' circle problem, known as the divisor problem, where $\mathcal{K}$ is a hyperbolic region in the plane bounded by $y = \frac{1}{x}$ and the coordinate axes (admittedly $\mathcal{K}$ is not bounded in this case and $|\mathcal{K}| = \infty$, but we may cut off the two thin cusps of $q\mathcal{K}$ which do not contain any lattice point and still make sense of the question). While an optimal estimate remains far out of reach for the above two problems, it has been achieved in the case of the parabola.

Let $\mathcal{P}_1 = \{(x, y) \in \mathbb{R}^2 \mid 0 < x,\ 0 < y \leq x^2\}$, and consider its truncated homothetic dilation

$$q\mathcal{P}_1(t) = \{(x, y) \in \mathbb{R}^2 \mid 0 < x \leq t, 0 < y \leq x^2/q\},$$

where t is a positive integer. The reason for truncation is obvious, as otherwise the region is unbounded. Thus we are led to the counting function

$$N_{\mathcal{P}_1}(q, t) = \#\{(a, b) \in \mathbb{Z}^2 \mid (a, b) \in q\mathcal{P}_1(t)\}.$$

Huang and Li obtained in [9, Theorem 1] that for any $\varepsilon > 0$

$$E_{\mathcal{P}_1}(q, t) = O\left(\sqrt{q}\log q + tq^{-\frac{1}{2}}\exp\left(\frac{(2+\varepsilon)\sqrt{\log q}}{\log\log q}\right)\right),$$

where

$$E_{\mathcal{P}_1}(q, t) = N_{\mathcal{P}_1}(q, t) - \sum_{a=1}^{t}\left(\frac{a^2}{q} - \frac{1}{2}\right)$$

and $q \geq 3$. Note that the main term in the sum $\sum_{a=1}^{t}\left(\frac{a^2}{q} - \frac{1}{2}\right)$ is $\frac{t^3}{3q}$, which is exactly the area of $q\mathcal{P}_1(t)$. However, there are some secondary main terms in the sum as well,

which are caused by the two flat edges of $q\mathcal{P}_1(t)$. In general, flat points (i.e. points with zero curvature) on the boundary of the region tend to have significant contribution to the counting function and sometimes cause larger than expected fluctuations of the error term. Fortunately, in the case under consideration, this is not much of a problem as their contribution can be computed exactly, so the authors still managed to obtain the much smaller conjectural error term.

In particular, if we let $t = q$, the above mentioned estimate becomes

$$E_{\mathcal{P}_1}(q,q) = O\left(\sqrt{q}\exp\left(\frac{(2+\varepsilon)\sqrt{\log q}}{\log\log q}\right)\right).$$

On the other hand, Chamizo and Pastor showed [1] in [3, Theorem 5.2] that

$$E_{\mathcal{P}_1}(q,q) = \Omega\left(\sqrt{q}\exp\left(\frac{(\sqrt{2}-\varepsilon)\sqrt{\log q}}{\log\log q}\right)\right).$$

It is readily seen that the above upper and lower bounds are incredibly close to each other and therefore are both optimal except for some logarithmic factors.

The main purpose of this thesis is to generalize the above result of Huang and Li from the standard parabola to the standard paraboloid. To that end, let

$$\mathcal{P}_n = \left\{(x_1,\cdots,x_n,y)\in\mathbb{R}^{n+1} \mid 0<x_1,\cdots,x_n,\ 0<y\le x_1^2+\cdots+x_n^2\right\},$$

and similarly we consider the truncated dilation

$$q\mathcal{P}_n(t) = \left\{(x_1,\cdots,x_n,y)\in\mathbb{R}^{n+1} \mid 0<x_1,\cdots,x_n\le t,\ 0<y\le \frac{x_1^2+\cdots+x_n^2}{q}\right\}$$

[1] Their formulation of the problem is a bit different from ours. Here we have taken the liberty of translating their results using our notation. An interested reader can take the conversion process as a fun exercise.

and the corresponding counting function

$$N_{\mathcal{P}_n}(q,t) = \#q\mathcal{P}_n(t) \cap \mathbb{Z}^{n+1}.$$

Our first main result is the following estimate of $N_{\mathcal{P}_n}(q,t)$.

Theorem 1. *For any positive integers q, t, and n with $q \geq 3$ and $n \geq 2$, we have*

$$E_{\mathcal{P}_n}(q,t) = O\left(t^n q^{-1}\xi(q) + q^{\frac{n}{2}}\log q\right),$$

where

$$E_{\mathcal{P}_n}(q,t) = N_{\mathcal{P}_n}(q,t) - \sum_{a_1,\cdots,a_n=1}^{t}\left(\frac{a_1^2+\cdots+a_n^2}{q} - \frac{1}{2}\right),$$

and

$$\xi(q) = \begin{cases} \exp\left((\log 2+\varepsilon)\dfrac{\log q}{\log\log q}\right) & \text{when } n=2, \\ \exp\left((2+\varepsilon)\dfrac{\sqrt{\log q}}{\log\log q}\right) & \text{when } n=3, \\ \log q\log\log q & \text{when } n=4, \\ \log q & \text{when } n\geq 5. \end{cases}$$

Again, when $t = q$, our Theorem 1 reduces to

$$E_{\mathcal{P}_n}(q,q) = O\left(q^{n-1}\xi(q)\right),$$

noting that $n-1 \geq \frac{n}{2}$ when $n \geq 2$. In view of the fact that $\xi(q) = O(q^{\varepsilon})$, this improves upon an earlier result of Chamizo and Pastor [3, Theorem 4.1] where they have obtained the upper bound $O(q^{n-1+\varepsilon})$.

In the same paper [3, Theorem 5.3], Chamizo and Pastor also have an asymptotic

lower bound

$$E_{\mathcal{P}_n}(q,q) = \Omega\left(q^{n-1}\eta(q)\right)$$

where

$$\eta(q) = \begin{cases} \exp\left((\log 2 - \varepsilon)\dfrac{\log q}{\log\log q}\right) & when\ n = 2, \\ \log\log q & when\ n = 3, \\ \sqrt{\log\log q} & when\ n = 4, \\ 1 & when\ n \geq 5. \end{cases}$$

This shows that our upper bound is extremely close to the true order of the error term.

As mentioned earlier, $\mathcal{K}$ can be any bounded region. So in a similar spirit to Theorem 1, let

$$A_n(q,t,\delta) = \sum_{\substack{1\leq a_1,\cdots,a_n\leq t \\ \left\|\frac{a_1^2+\cdots+a_n^2}{q}\right\|\leq\delta}} 1,$$

where $\|\cdot\|$ is the distance to the closest integer. Intuitively the function $A_n(q,t,\delta)$ counts the number of lattice points that lie within δ of the dilated paraboloid $y = \frac{x_1^2+\cdots+x_n^2}{q}$ with $0 < x_1, \cdots, x_n \leq t$.

We remark in passing that when $n = 1$, this problem has been well studied for the standard parabola [7, 9] and for general parabolas [8]. In fact, this kind of counting problem is not only interesting in its own right, but also closely related to metric diophantine approximation on manifolds, a very hot topic of late. We refer the interested readers to [1, 4, 5, 6, 10] and the references therein for an overview and some recent advances in this fast-growing field.

Just as in Theorem 1, we achieve an essentially optimal estimate for $A_n(q,t,\delta)$.

Theorem 2. *Let $\delta \in (0, 1/2)$. For any positive integers q, t, and n with $q \geq 3$ and $n \geq 2$, we have*

$$A_n(q, t, \delta) = 2\delta t^n + O\left(t^n q^{-1} \xi(q) + q^{\frac{n}{2}} \log q\right),$$

where $\xi(q)$ is defined the same with Theorem 1.

Before embarking on the detailed proofs, here we briefly outline our strategy. Our approach starts with an elementary counting argument based on the orthogonality of additive characters, which then naturally leads us to treat some incomplete quadratic Gauss sums. It is at this stage that the major cancellation occurs, which eventually results in the essentially sharp error terms in our theorems. A seasoned worker can perceive the spirit of Fourier-analytic methods here, which perhaps is not surprising at all in a lattice point problem. It is, however, worth noting that our approach is completely elementary, and does not utilize any advanced analytic tools, which means that this thesis is accessible to motivated undergraduate students with an introductory course in number theory.

This thesis is structured as follows. In Section 2, we present some necessary lemmata that are needed in the proofs of our main theorems in Section 3 and 4. In Section 5, we will discuss some possible future projects based off the methods developed in this thesis.

2 Preliminary Lemmata

Lemma 1 (Orthogonality of additive characters)**.** *For a positive integer q, we have*

$$\frac{1}{q}\sum_{h=1}^{q} e\left(\frac{hd}{q}\right) = \begin{cases} 1, & \text{if } q \mid d, \\ 0, & \text{otherwise.} \end{cases}$$

Proof. Let $x = e(d/q)$. Notice that $x = 1$ if and only if $q \mid d$. Also $x^q = e(d) = 1$. Now we discuss two cases:

(i) If $q \mid d$, then $x = 1$ and

$$\sum_{h=1}^{q} e\left(\frac{hd}{q}\right) = \sum_{h=1}^{q} x^h = q.$$

(ii) If $q \nmid d$, then $x \neq 1$ and by the geometric summation formula we have

$$\sum_{h=1}^{q} x^h = x \cdot \frac{x^q - 1}{x - 1} = 0.$$

□

Lemma 2. *For a positive integer q and an integer h, we have*

$$\sum_{j=0}^{q-1} je\left(-\frac{hj}{q}\right) = \begin{cases} \frac{-q}{1-e\left(-\frac{h}{q}\right)}, & \text{if } q \nmid h, \\ \frac{(q-1)q}{2}, & \text{if } q \mid h. \end{cases}$$

Proof. In the case that $q \mid h$, we have

$$\sum_{j=0}^{q-1} je\left(-\frac{hj}{q}\right) = \sum_{j=0}^{q-1} j = \frac{(q-1)q}{2}.$$

Now suppose that $q \nmid h$. Observe that

$$\begin{aligned}
&\left(\sum_{j=0}^{q-1} je\left(-\frac{hj}{q}\right)\right)\left(1-e\left(\frac{-h}{q}\right)\right)\\
&= \left(0+e\left(\frac{-h}{q}\right)+2e\left(\frac{-2h}{q}\right)+\cdots+(q-1)e\left(\frac{-(q-1)h}{q}\right)\right)\left(1-e\left(\frac{-h}{q}\right)\right)\\
&= e\left(\frac{-h}{q}\right)+2e\left(\frac{-2h}{q}\right)+\cdots+(q-1)e\left(\frac{-(q-1)h}{q}\right)\\
&\qquad -e\left(\frac{-2h}{q}\right)-\cdots-(q-2)e\left(\frac{-(q-1)h}{q}\right)-(q-1)e\left(\frac{-qh}{q}\right)\\
&= e\left(\frac{-h}{q}\right)+e\left(\frac{-2h}{q}\right)+\cdots+e\left(\frac{-(q-1)h}{q}\right)+1-q\\
&= \sum_{j=1}^{q} e\left(\frac{-jh}{q}\right)-q
\end{aligned}$$

Then the desired conclusion immediately follows in view of Lemma 1 and the assumption that $q \nmid h$.

□

Lemma 3. *For a positive integer q and an integer h such that $q \nmid h$, we have the following inequality*

$$\left|\frac{1}{1-e\left(-\frac{h}{q}\right)}\right| \leq \frac{1}{4\left\|\frac{h}{q}\right\|}.$$

Proof. Without loss of generality we may assume $0 < h < q$ as the general case follows by periodicity. Notice that

$$\left|\frac{1}{1-e\left(-\frac{h}{q}\right)}\right| = \left|\frac{e\left(\frac{h}{2q}\right)}{e\left(\frac{h}{2q}\right)-e\left(-\frac{h}{2q}\right)}\right| = \frac{1}{2\sin\left(\frac{h}{q}\pi\right)},$$

where in the second equation above we use Euler's identity

$$e\left(\pm\frac{h}{2q}\right) = \cos\left(\frac{h}{q}\pi\right) \pm i\sin\left(\frac{h}{q}\pi\right).$$

Next we see that the inequality $\sin(\theta\pi) \geq 2\theta > 0$ holds for all $\theta \in \left(0, \frac{1}{2}\right]$ by the concavity of the function $\sin(\theta\pi)$. Moreover when $\theta \in [\frac{1}{2}, 1)$, by the reflection formula we have

$$\sin(\theta\pi) = \sin(\pi - \theta\pi) \geq 2(1-\theta).$$

In any case, this gives

$$\sin(\theta\pi) \geq 2\|\theta\| > 0, \quad \text{when } 0 < \theta < 1.$$

Therefore

$$\left|\frac{1}{1-e\left(-\frac{h}{q}\right)}\right| = \frac{1}{2\sin\left(\frac{h}{q}\pi\right)} \leq \frac{1}{4\left\|\frac{h}{q}\right\|}.$$

□

Lemma 4 ([11, Corollary, Page 53])**.** *Let q, t and h be integers such that $1 \leq t \leq q$ and $(q,h) = 1$. Then*

$$\left|\sum_{a=1}^{t} e\left(\frac{ha^2}{q}\right)\right| < 3.9071\sqrt{q}.$$

Lemma 5 ([12, §I.5.5 Theorem 5])**.** *For a positive integer q and a real number s, we define the sum of divisor function as follows*

$$\sigma_s(q) = \sum_{d|q} d^s.$$

Then we have the following bounds

$$\sigma_0(q) \ll \exp\left((\log 2 + o(1))\frac{\log q}{\log\log q}\right),$$
$$\sigma_{\frac{1}{2}}(q) \ll \sqrt{q}\exp\left((2 + o(1))\frac{\sqrt{\log q}}{\log\log q}\right),$$
$$\sigma_1(q) \ll q\log\log q,$$

and for $s > 1$

$$\sigma_s(q) \ll q^s.$$

3 The proof of Theorem 1

First, it is easily seen that

$$N_{\mathcal{P}_n}(q,t) = \sum_{a_1,\cdots,a_n=1}^{t} \left\lfloor \frac{a_1^2 + \cdots + a_n^2}{q} \right\rfloor.$$

Next, in view of $\lfloor y \rfloor = y - \{y\}$, it suffices to investigate the above summation with the floor function replaced by the fractional part. Moreover, if $y = a/q$ for some positive integers a, q, then there exists a unique integer j with $a \equiv j \bmod q$ and $0 \le j \le q-1$ so that

$$\left\{\frac{a}{q}\right\} = \frac{j}{q}.$$

From this observation we see that

$$\sum_{a_1,\cdots,a_n=1}^{t} \left\{\frac{a_1^2 + a_2^2 + \cdots + a_n^2}{q}\right\} = \sum_{j=0}^{q-1} \frac{j}{q} \sum_{\substack{1\le a_1,\cdots,a_n\le t \\ a_1^2+\cdots+a_n^2 \equiv j \bmod q}} 1. \tag{1}$$

We then use Lemma 1 to pick up the congruence condition $a_1^2 + \cdots + a_n^2 \equiv j \bmod q$, so after rearranging the order of summation the inner sum on the right hand side becomes

$$\sum_{a_1,\cdots,a_n=1}^{t} \frac{1}{q} \sum_{h=1}^{q} e\left(h\frac{a_1^2 + \cdots + a_n^2 - j}{q}\right) = \frac{1}{q}\sum_{h=1}^{q} e\left(-\frac{hj}{q}\right) S(h,q,t)^n,$$

where

$$S(h,q,t) = \sum_{a=1}^{t} e\left(h\frac{a^2}{q}\right)$$

is an incomplete Gauss sum. Therefore the right hand side of (1) is

$$\frac{1}{q^2}\sum_{h=1}^{q}\sum_{j=0}^{q-1} je\left(-\frac{hj}{q}\right) S(h,q,t)^n,$$

which after applying Lemma 2 to the summation over j, yields

$$\sum_{a_1,\cdots,a_n=1}^{t} \left\{\frac{a_1^2 + a_2^2 + \cdots + a_n^2}{q}\right\} = \frac{q-1}{2q}t^n + \frac{1}{q}\sum_{h=1}^{q-1} \frac{-1}{1 - e\left(-\frac{h}{a}\right)} S(h,q,t)^n, \qquad (2)$$

where the first term on the right hand side corresponds to $h = q$.

We would like to estimate the Gauss sum $S(h,q,t)$ via lemma 4. However, the fact that q and h may not be coprime creates an extra nuisance, which can be taken care of by working with $h' := h/d$ and $q' := q/d$ instead, where $d = (h,q)$ is the greatest common divisor of h and q. Another issue is that the lemma only allows sums of length at most q'. This can also be resolved by dissecting the range $[1,t]$ into at most $\lfloor t/q' \rfloor + 1$ blocks of length at most q' and utilizing periodicity of the Gauss sum. By

Lemma 4, each block contributes $O(\sqrt{q'})$. So we deduce that

$$S(h,q,t) = S(h',q',t) \ll \left(\frac{t}{q'}+1\right)\sqrt{q'} \ll \frac{t}{\sqrt{q'}} + \sqrt{q'},$$

which together with Lemma 3 gives

$$\begin{aligned}\frac{1}{q}\sum_{h=1}^{q-1}\frac{-1}{1-e\left(-\frac{h}{q}\right)}S(h,q,t)^n &\ll \sum_{h=1}^{q-1}\frac{1}{q\|\frac{h}{q}\|}\left(\frac{t}{\sqrt{q'}}+\sqrt{q'}\right)^n \\ &\ll \sum_{h=1}^{q-1}\frac{1}{q\|\frac{h}{q}\|}\left(\frac{t^n}{\sqrt{q'^n}}+\sqrt{q'}^n\right).\end{aligned}$$

To estimate the resulting sum, we may split it into two sums $\sum_{1\le h<q/2}$ and $\sum_{q/2\le h<q}$. We will only treat the former case as the latter is completely analogous. From here, we write $h = dk$. Then we must have $(k,q') = 1$ since $(h,q) = d$. Hence the former sum is

$$\sum_{d|q}\sum_{\substack{k<q'/2\\(k,q')=1}}\frac{1}{kd}\left(\frac{t^n}{\sqrt{q^n}}\sqrt{d^n}+\sqrt{\frac{q^n}{d^n}}\right) = \sum_{d|q}\left(\frac{t^n}{\sqrt{q^n}}d^{\frac{n}{2}-1}+\sqrt{\frac{q^n}{d^{n+2}}}\right)\sum_{\substack{k<q'/2\\(k,q')=1}}\frac{1}{k}.$$

We shall note that

$$\sum_{d|q} d^{-\frac{n+2}{2}} \le \sum_{d=1}^{\infty}\frac{1}{d^2} = \frac{\pi^2}{6}$$

and that

$$\sum_{\substack{k<q'/2\\(k,q')=1}}\frac{1}{k} \le \sum_{k<q}\frac{1}{k} \ll \log q.$$

Thus

$$\begin{aligned}\sum_{h<\frac{q}{2}} \frac{1}{q\|\frac{h}{q}\|}\left(\frac{t^n}{\sqrt{q'}^n}+\sqrt{q'}^n\right) \ll & \log q\left(\frac{t^n}{\sqrt{q^n}}\sum_{d|q} d^{\frac{n}{2}-1}+\sqrt{q^n}\right)\\ \ll & t^n q^{-\frac{n}{2}}(\log q)\sigma_{\frac{n}{2}-1}(q)+q^{\frac{n}{2}}\log q\\ \ll & t^n q^{-1}\xi(q)+q^{\frac{n}{2}}\log q\end{aligned}$$

where in the second last line we bound $\sigma_s(q)$ using Lemma 5.

Therefore, we obtain from (2) that when $n \geq 2$

$$\sum_{a_1,\cdots,a_n=1}^{t}\left(\frac{1}{2}-\left\{\frac{a_1^2+a_2^2+\cdots+a_n^2}{q}\right\}\right)=O\left(t^n q^{-1}\xi(q)+q^{\frac{n}{2}}\log q\right),$$

from which Theorem 1 follows immediately.

4 The proof of Theorem 2

Let $J=\lfloor \delta q\rfloor$. Our goal is to count the number of lattice points $(a_1,\cdots,a_n)\in\{1,2,\cdots,t\}^n$ such that $\left\|\frac{a_1^2+\cdots+a_n^2}{q}\right\|\leq\delta$. Note that $\left\|\frac{a_1^2+\cdots+a_n^2}{q}\right\|\leq\delta$ if and only if there exists an integer k, such that

$$k-\delta\leq\frac{a_1^2+\cdots+a_n^2}{q}\leq k+\delta,$$

i.e.

$$kq-\delta q\leq a_1^2+\cdots+a_n^2\leq kq+\delta q,$$

which happens if and only if $a_1^2 + \cdots + a_n^2 \equiv j \pmod{q}$ for some $|j| \le J$. It therefore follows that

$$A_n(q,t,\delta) = \sum_{\substack{1 \le a_1,\cdots,a_n \le t \\ \left\| \frac{a_1^2+\cdots+a_n^2}{q} \right\| \le \delta}} 1 = \sum_{|j| \le J} \sum_{\substack{1 \le a_1,\cdots,a_n \le t \\ a_1^2+\cdots a_n^2 \equiv j \bmod q}} 1.$$

Using Lemma 1 and rearranging the order of summation yield

$$\begin{aligned} A_n(q,t,\delta) =& \sum_{|j| \le J} \sum_{a_1,\cdots,a_n=1}^{t} \frac{1}{q} \sum_{h=1}^{q} e\left(h\frac{a_1^2+\cdots+a_n^2-j}{q}\right) \\ =& \frac{1}{q} \sum_{h=1}^{q} \sum_{|j| \le J} e\left(-\frac{hj}{q}\right) S(h,q,t)^n. \end{aligned} \tag{3}$$

The terms with $h = q$ contribute

$$(2J+1)\frac{t^n}{q} = 2\delta t^n + O\left(\frac{t^n}{q}\right). \tag{4}$$

When $h \neq q$, a successive application of geometric summation and Lemma 3 gives

$$\left| \sum_{|j| \le J} e\left(-\frac{hj}{q}\right) \right| = \left| e\left(\frac{hJ}{q}\right) \frac{1 - e\left(-\frac{h(2J+1)}{q}\right)}{1 - e\left(-\frac{h}{q}\right)} \right| \le \left| \frac{2}{1 - e\left(-\frac{h}{q}\right)} \right| \le \left(2 \left\| \frac{h}{q} \right\| \right)^{-1}.$$

So the total contribution of the terms with $1 \le h < q$ in (3) is

$$\ll \frac{1}{q} \sum_{h=1}^{q-1} \left\| \frac{h}{q} \right\|^{-1} |S(h,q,t)|^n,$$

which, as demonstrated in the proof of Theorem 1, is

$$\ll t^n q^{-1} \xi(q) + q^{\frac{n}{2}} \log q. \tag{5}$$

Finally Theorem 2 follows by combining (3), (4), and (5).

5 Future Work

Since we have successfully treated the standard paraboloid

$$y = x_1^2 + x_2^2 + \cdots + x_n^2,$$

it is very likely that our method can be employed as well to treat rational diagonal paraboloids

$$y = c_1x_1^2 + c_2x_2^2 + \cdots + c_nx_n^2, \quad c_1, \cdots, c_n \in \mathbb{Q}.$$

Clearly our method does not work for the irrational case. Nevertheless, it is conceivable that a proper adaptation of the method of Huang and Li in [8] will enable us to treat the general diagonal case, likely at the expense of slightly weaker error terms than those obtained here.

Another immediate project is to shave off the extra logarithmic factor in $\xi(q)$ when $n \geq 4$. This can be done via a more careful analysis of the divisor sum function. If this is done, then $\xi(q)$ exactly matches $\eta(q)$ when $n \geq 5$, which would render absolutely sharp upper and lower bounds for the corresponding error term $E_{\mathcal{P}_n}(q, q)$.

A more ambitious goal would be to treat general quadratic hypersurfaces of the form

$$y = Q(x_1, x_2, \cdots, x_n),$$

where Q is a quadratic form in n variables with real coefficients. Chamizo and Pastor [3, Theorem 1.1] have obtained a result in this regard with some restrictions on the coefficients of Q.

References

[1] Victor Beresnevich. Rational points near manifolds and metric Diophantine approximation. *Ann. of Math. (2)*, 175(1):187–235, 2012.

[2] Jean Bourgain and Nigel Watt. Decoupling for perturbed cones and the mean square of $|\zeta(\frac{1}{2}+it)|$. *Int. Math. Res. Not. IMRN*, 2018(17):5219–5296, 2018.

[3] Fernando Chamizo and Carlos Pastor. Lattice points in elliptic paraboloids. *Publ. Mat.*, 63(1):343–360, 2019.

[4] Jing-Jing Huang. Rational points near planar curves and Diophantine approximation. *Adv. Math.*, 274:490–515, 2015.

[5] Jing-Jing Huang. Integral points close to a space curve. *Math. Ann.*, 374(3-4):1987–2003, 2019.

[6] Jing-Jing Huang. The density of rational points near hypersurfaces. *Duke Math. J.*, 169(11):2045–2077, 2020.

[7] Jing-Jing Huang. Diophantine approximation on the parabola with non-monotonic approximation functions. *Math. Proc. Cambridge Philos. Soc.*, 168(3):535–542, 2020.

[8] Jing-Jing Huang and Huixi Li. On a generalization of a theorem of Popov. *Houston J. Math.*, 46(1):27–38, 2020.

[9] Jing-Jing Huang and Huixi Li. On two lattice points problems about the parabola. *Int. J. Number Theory*, 16(4):719–729, 2020.

[10] Jing-Jing Huang and Jason J. Liu. Simultaneous approximation on affine subspaces. *Int. Math. Res. Not. IMRN*, in press. arXiv:1811.06531, November 2018.

[11] M. A. Korolev. On incomplete Gaussian sums. *Proc. Steklov Inst. Math.*, 290(1):52–62, 2015.

[12] Gérald Tenenbaum. *Introduction to analytic and probabilistic number theory*, volume 163 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, third edition, 2015.