The Journal of Geometric Analysis (2021) 31:9113–9130
https://doi.org/10.1007/s12220-020-00579-9

An Introduction to the Circle Method of Hardy, Littlewood, and Ramanujan

Stephen Wainger[1]

Accepted: 10 January 2020 / Published online: 25 January 2021

Abstract
We discuss the number of lattice points with integer coordinates on the sphere of radius λ and Vinogradov's Theorem on the representation of integers as a sum of three primes.

Keywords Circle method · Lattice points on spheres · Sums of three primes

Mathematics Subject Classification 11045

1 Introduction

In the last 15 years, the circle method has been applied to problems in Harmonic Analysis, Ergodic Theory, and Partial Differential Equations. See e.g., [1,2,4–6,11,14, 15,19,21,23,27–29,34]. Thus it seems worthwhile to have an introductory exposition to the topic. The present article tries to give such an exposition. As such, it is a slightly expanded version of a talk given at the IAS Park City Institute in July 2003. Over the last several years, I have been working with A. Magyar and E. M. Stein on several applications of the circle method. Conversations I have had with Magyar and Stein have greatly enriched my understanding of this method. I have also profited greatly from Magyar's paper [14] and unpublished lecture notes of Stein [25].

The circle method of Hardy, Littlewood, and Ramanujan is a method of studying asymptotically the number of solutions of diophantine equations. For example, Hardy and Littlewood [10] (with later improvements by Vinogradov [32]) studied the number of representations of an integer m as a sum of ℓ kth powers. That is they studied the number of solutions in positive integers $n_1, n_2, \ldots, n_\ell$ of the equation

✉ Stephen Wainger
wainger@math.wisc.edu

[1] Mathematics Department, University of Wisconsin - Madison, 480 Lincoln Drive, 213 Van Vleck Hall, Madison, WI 53706, USA

$$m = n_1^k + \cdots + n_\ell^k.$$

Another example is a theorem of Vinogradov [32] asserting that every sufficiently large odd integer can be written as a sum of three primes.

We will begin by considering $r_d(\lambda)$, the number of lattice points in $\mathbb{R}^d$ on the sphere centered at the origin of radius λ. A lattice point in $\mathbb{R}^d$ is a point $n = (n_1, n_2, \ldots, n_d)$ with $n_1, n_2, \ldots, n_d$ integers. Then we shall give a brief discussion of Vinogradov's Theorem asserting that every sufficiently large odd number can be written as a sum of 3 primes, and will try to explain why it seems so difficult to use the method to show that every sufficiently large even integer can be written as a sum of 2 primes. Finally, we will give some references for further reading.

2 The Number of Lattice Points on the Sphere of Radius λ

$r_d(\lambda)$, the number of lattice points in $\mathbb{R}^d$ on the sphere of radius λ centered at the origin, is the number of solutions in integers $(n_1, n_2, \ldots, n_d)$of the equation

$$\lambda^2 = n_1^2 + \cdots + n_d^2. \tag{1}$$

So $r_d(\lambda) = 0$ unless λ^2 is an integer, and we will always assume λ^2 is an integer. Then there is the following theorem.

Theorem 1 *For $d \geq 5$, there are positive constants $c_1(d)$ and $c_2(d)$ such that*

$$c_1(d)\lambda^{d-2} \leq r_d(\lambda) \leq c_2(d)\lambda^{d-2}.$$

See [8,13] if $d \geq 6$. The statement is false for $d \leq 4$. Note that the power of λ that occurs in Theorem 1 is λ^{d-2} while the area of the corresponding sphere in $\mathbb{R}^d$ is $C(d)\lambda^{d-1}$. We should expect the power λ^{d-2} to arise for the following reason: We expect the number of lattice points in the annulus $\Lambda \leq |x| \leq 2\Lambda$ to be about $c\Lambda^d$ for large Λ. On the other hand, the number of spheres having lattice points of radius λ with $\Lambda \leq \lambda \leq 2\Lambda$ is the number of λ with $\Lambda \leq \lambda \leq 2\Lambda$ such that λ^2 is an integer, and thus the number of integers in the interval $[\Lambda^2, 4\Lambda^2]$. Thus if $r_d(\lambda) \sim \lambda^a$, $\Lambda^a \cdot \Lambda^2 \sim \Lambda^d$ so $a = d - 2$.

We shall try to outline the proof of the upper bound in Theorem 1 by the circle method, and the lower bound for $d \geq 27$. We will then briefly indicate how to obtain the lower bound for $5 \leq d \leq 27$.

To prove Theorem 1, one shows

$$r_d(\lambda) = M_d(\lambda)\lambda^{d-2} + E_d(\lambda), \tag{2}$$

where

$$C_1(d) \leq M_d(\lambda) \leq C_2(d)$$

with $C_1(d)$ and $C_2(d)$ positive, and

$$|E_d(\lambda)| \leq C\lambda^{d/2}.$$

Because in Eq. (1), the power of the n_j is 2, other methods can be used to study $r_d(\lambda)$. In particular, Hardy [9] showed $E_d(\lambda) = 0$ if $5 \leq d \leq 8$. See [8,13]. Even when $d = 3$ or 4, it can be shown that $E_d(\lambda) = 0$. See [3,26]. A better understanding of this can be found in the work of Mordell [17]. See [20] for a more recent development using the modular group. In general, the estimate for $E_d(\lambda)$ can be improved. See [12].

$M_d(\lambda)$ itself is complicated to describe. In particular, it involves Gauss sums, $S(a, q)$. If $(a, q) = 1$, that is a and q are relatively prime, and $1 \leq a \leq q$,

$$S(a,q) = \sum_{n=0}^{q-1} e^{-2\pi i n^2 a/q}.$$

If $q = 1$, the only integer a with $(a, q) = 1$ is $a = 1$ and

$$S(1,1) = 1.$$

If $q = 2$, the only integer a with $(a, q) = 1$ is again $a = 1$ and

$$S(1,2) = \sum_{n=0}^{1} e^{-2\pi i n^2 1/2} = 1 + e^{-i\pi} = 1 - 1 = 0.$$

There is the following estimate for the size of $S(a, q)$.

Lemma 2

$$|S(a,q)| \leq \sqrt{2q}.$$

We defer the proof of Lemma 2 until later.

Now

$$M_d(\lambda) = C(d) \sum_{q=1}^{\infty} \sum_{\substack{a=1 \\ (a,q)=1}}^{q} e^{-2\pi i \lambda^2 \frac{a}{q}} \left(\frac{1}{q} S(a,q)\right)^d. \tag{3}$$

$M_d(\lambda)$ is generally referred to as the singular series. Note that Lemma 2 easily implies that $|M_d(\lambda)| \leq C(d)$ for $d \geq 5$. Also our remarks on $S(1, 1)$ and $S(1, 2)$ together with Lemma 2 imply

$$|M_d(\lambda)| \geq C(d) \left(1 - \sum_{q \geq 3}^{\infty} q \cdot \left(\sqrt{\frac{2}{q}}\right)^d\right) > \overline{C}(d)$$

if $d \geq 27$. The condition $d \geq 27$ could of course easily be improved.

In thinking about $r_d(\lambda)$, our first task is to change the combinatorial problem of studying the number of solutions of Eq. (1) to an analytic problem. A key observation is that

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i(\lambda^2 - n_1^2 - \cdots - n_d^2)\theta}\, \mathrm{d}\theta = \begin{cases} 1, & \text{if } n_1^2 + \cdots + n_d^2 = \lambda^2; \\ 0, & \text{otherwise.} \end{cases}$$

So

$$r_d(\lambda) = \sum_{n_1,\ldots,n_d=-\infty}^{\infty} \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i(\lambda^2 - n_1^2 - \cdots - n_d^2)\theta}\, \mathrm{d}\theta \tag{4}$$

or formally

$$r_d(\lambda) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i\lambda^2\theta} \sum_{n_1,\ldots,n_d=-\infty}^{\infty} \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-2\pi i(n_1^2+\cdots+n_d^2)\theta}\, \mathrm{d}\theta$$

or

$$r_d(\lambda) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i\lambda^2\theta} \left(\sum_{n=-\infty}^{\infty} e^{2\pi i n^2\theta} \right)^d \mathrm{d}\theta. \tag{5}$$

Of course, the infinite sun in (5) does not converge. To make (5) rigorous, we can either truncate the sum in (4) or introduce an ϵ. It turns out that in the case of squares, introducing an $\epsilon > 0$ is more convenient. Thus, we note that

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i\lambda^2\theta}\, e^{-2\pi i(n_1^2+\cdots+n_d^2)\theta}\, e^{-2\pi\epsilon(n_1^2+\cdots+n_d^2)}\, \mathrm{d}\theta$$
$$= \begin{cases} e^{-2\pi\epsilon(n_1^2+\cdots+n_d^2)} = e^{-2\pi\epsilon\lambda^2}, & \text{if } \ n_1^2+\cdots+n_d^2 = \lambda^2; \\ 0, & \text{otherwise.} \end{cases}$$

So

$$\begin{aligned} r_d(\lambda) &= e^{2\pi\epsilon\lambda^2} \sum_{n_1,\ldots,n_d=-\infty}^{\infty} \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-2\pi i(n_1^2+\cdots+n_d^2)\theta}\, e^{-2\pi\epsilon(n_1^2+\cdots+n_d^2)\theta}\, \mathrm{d}\theta \\ &= e^{2\pi\epsilon\lambda^2} \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i\lambda^2\theta} \left(\sum_{n=-\infty}^{\infty} e^{-2\pi(\epsilon+i\theta)(n_1^2+\cdots+n_d^2)} \right)^d \mathrm{d}\theta, \end{aligned}$$

or

$$r_d(\lambda) = e^{2\pi\epsilon\lambda^2} \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i\lambda^2\theta} \{F(\epsilon + i\theta)\}^d\, \mathrm{d}\theta. \tag{6}$$

where

$$F(\epsilon + i\theta) = \sum_{n=-\infty}^{\infty} e^{-2\pi(\epsilon+i\theta)n^2}. \tag{7}$$

We will always take $\epsilon = \dfrac{1}{\lambda^2}$ so that the factor $e^{2\pi\epsilon\lambda^2}$ will be a constant. To study the analytical problem posed by (6), we have to understand $F(\epsilon + i\theta)$, which is of course essentially a classical theta function. A convenient way to study $F(\epsilon + i\theta)$ is via the Poisson summation formula. The Poisson summation formula asserts that under suitable hypothesis on a function f,

$$\sum_n f(n) = \sum_n \hat{f}(n)$$

where

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} e^{-2\pi x\cdot\xi} f(x)\,\mathrm{d}x.$$

See [30].

We are going to apply the Poisson summation formula with

$$f(x) = e^{-2\pi(\epsilon+i\theta)x^2}.$$

Now $e^{-\pi x^2}$ is its own Fourier transform. See [30]. So by a change of variables

$$\hat{f}(\xi) = \frac{1}{\sqrt{2(\epsilon+i\theta)}} e^{-\pi\frac{\xi^2}{2(\epsilon+i\theta)}}.$$

Then the Poisson summation formula asserts

$$F(\epsilon+i\theta) = \sum_{n=-\infty}^{\infty} \frac{1}{\sqrt{2(\epsilon+i\theta)}} e^{-\pi\frac{n^2}{2(\epsilon+i\theta)}}. \tag{8}$$

If one is lucky in using the Poisson summation formula, the main term in $\sum \hat{f}(n)$ is $\hat{f}(0)$. Thus, we might hope

$$F(\epsilon+i\theta) = \frac{1}{\sqrt{2(\epsilon+i\theta)}} + \text{Error} \tag{9}$$

Just to see that we are on the right track, let us see what would happen if

$$F(\epsilon+i\theta) = \frac{1}{\sqrt{2(\epsilon+i\theta)}}.$$

Then we would have

$$r_d(\lambda) = C_d \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i\lambda^2\theta} \frac{1}{(\epsilon+i\theta)^{d/2}}\,\mathrm{d}\theta.$$

Now let us make a change of variable $\theta = x\epsilon$. Then recalling the fact that $\epsilon = \frac{1}{\lambda^2}$, we arrive at the equation

$$r_d(\lambda) = C_d\,\lambda^{d-2} \int_{-\frac{\lambda^2}{2}}^{\frac{\lambda^2}{2}} e^{2\pi ix} \frac{1}{(1+ix)^{d/2}}\,\mathrm{d}x.$$

Notice that the integrand is independent of λ, so that we would have

$$r_d(\lambda) = C_d\, \lambda^{d-2} \int_{-\infty}^{\infty} e^{2\pi i x} \frac{1}{(1+ix)^{d/2}}\, \mathrm{d}x + \mathcal{O}(1)$$

(which of course is too good an error to be true).

It is not hard to see that $\int_{-\infty}^{\infty} \frac{e^{2\pi i x}}{(1+ix)^{d/2}}\, \mathrm{d}x \neq 0$. If $d \geq 5$ is an even integer, one sees this by the residue theorem. If $d = 1$, this follows by distorting the contour to an integral over $[i, i\infty]$ and using the fact that $(1+ix)^{-1/2}$ is multiple valued. If d is a larger odd integer, we can reduce the matters to $d = 1$ by integration by parts.

Let us now return to Eq. (9), and consider the error.

The error is

$$\frac{1}{\sqrt{2(\epsilon + i\theta)}} \sum_{n \neq 0} e^{-\pi \frac{n^2}{\epsilon + i\theta}}.$$

The absolute value of the nth term in the above series is $\exp\left(-C\frac{n^2 \epsilon}{\epsilon^2+\theta^2}\right)$ for some positive C. Thus, we can expect to control the error only if $\theta^2 \leq \epsilon$, that is, if $|\theta| \leq \frac{1}{\lambda}$. If $|\theta| \leq \frac{1}{\lambda}$, then

$$|\text{Error}| \leq C_1 \frac{1}{(\epsilon^2+\theta^2)^{1/4}}\, e^{-C_2 \frac{\epsilon}{\epsilon^2+\theta^2}}.$$

So

$$\begin{aligned} F^d(\epsilon + i\theta) &= \left(\frac{1}{2(\epsilon+i\theta)}\right)^{d/2} + \mathcal{O}\left(\left(\frac{1}{\epsilon^2+\theta^2}\right)^{d/4}\right) e^{-C_3 \frac{\epsilon}{\epsilon^2+\theta^2}} \\ &= \left(\frac{1}{2(\epsilon+i\theta)}\right)^{d/2} + \mathcal{O}\left(\frac{1}{\epsilon^{d/4}}\left(\frac{\epsilon}{\epsilon^2+\theta^2}\right)^{d/4}\right) e^{-C_3 \frac{\epsilon}{\epsilon^2+\theta^2}} \\ &= \left(\frac{1}{2(\epsilon+i\theta)}\right)^{d/2} + \mathcal{O}(\lambda^{d/2}). \end{aligned}$$

This leads to the estimate

$$\int_{-\frac{1}{\lambda}}^{\frac{1}{\lambda}} e^{2\pi i \lambda^2 \theta} \{F(\epsilon + i\theta)\}^d\, \mathrm{d}\theta = C(d)\lambda^{d-2} + \mathcal{O}(\lambda^{\frac{d}{2}-1})$$

Now the thrust of the circle method is that the main contribution to the integral in (6) should come from small intervals around rationals a/q with $1 \leq a \leq q$, $(a, q) = 1$ and q not too large. In the present example, we define

$$I(a, q) = \left\{\theta : \left|\theta - \frac{a}{q}\right| \leq \frac{1}{q\lambda}\right\}.$$

The intervals $I(a,q)$ are disjoint for $q \le \frac{\lambda}{20}$ since if $I(a,q) \cap I(a_1,q_1) \ne \varnothing$

$$\left|\frac{a}{q} - \frac{a_1}{q_1}\right| \le 2\frac{1}{q^*\lambda}$$

where $q^* = \min(q, q_1)$. Since $(a,q) = (a_1,q_1)$=1,

$$\left|\frac{a}{q} - \frac{a_1}{q_1}\right| \ge \frac{1}{qq_1}.$$

Thus

$$\frac{1}{2}q^*\lambda \le qq_1,$$

and $\min(q,q_1) \ge \frac{\lambda}{2}$. Thus

$$r_d(\lambda) = \sum_{q=1}^{\frac{\lambda}{20}} \sum_{\substack{a=1\\(a,q)=1}}^{q} \int_{I(a,q)} e^{2\pi i\lambda^2\theta}\{F(\epsilon+i\theta)\}^d\,\mathrm{d}\theta + \int_{\mathcal{E}_\lambda} e^{2\pi i\lambda^2\theta}\{F(\epsilon+i\theta)\}^d\,\mathrm{d}\theta.$$

According to well-known principle of Dirichlet, for each $\theta \in [0,1]$ there is a q with $(a,q)=1$ such that $\left|\theta - \frac{a}{q}\right| \le \frac{1}{\lambda q}$, $q \le \lambda$. Thus

$$\mathcal{E}_\lambda \subset \bigcup_{\frac{\lambda}{2\theta} \le q \le \lambda} I(a,q).$$

Now we would like to find an approximation to $F(\epsilon+i\theta)$ for $\theta \in I(a,q)$ with $q \le \lambda$. To this end for $\theta \in I(a,q)$, we write n in the sum defining $F(\epsilon+i\theta)$ as

$$n = mq + \mu, \quad 0 \le \mu \le q-1.$$

Thus

$$F(\epsilon+i\theta) = \sum_{n=-\infty}^{\infty} e^{-2\pi n^2(\epsilon+i\theta)} = \sum_{\mu=0}^{q-1}\sum_{m=-\infty}^{\infty} e^{-2\pi(mq+\mu)^2(\epsilon+i(\theta-\frac{a}{q})+i\frac{a}{q})}.$$

Since

$$e^{-2\pi i(mq+\mu)^2\frac{a}{q}} = e^{-2\pi i\mu^2\frac{a}{q}},$$

$$F(\epsilon+i\theta) = \sum_{\mu=0}^{q-1} e^{-2\pi i\mu^2\frac{a}{q}}\,F_\mu\left(\epsilon + i\left(\theta - \frac{a}{q}\right)\right).$$

where

$$F_\mu\left(\epsilon+i\left(\theta-\frac{a}{q}\right)\right)=\sum_{m=-\infty}^{\infty}e^{-2\pi(mq+\mu)^2(\epsilon+i(\theta-\frac{a}{q}))}.$$

We study $F_\mu\left(\epsilon+i\left(\theta-\frac{a}{q}\right)\right)$ by the Poisson's summation formula with $f(x)=e^{-2\pi(xq+\mu)^2(\epsilon+i(\theta-\frac{a}{q}))}$. Then

$$\hat{f}(0)=C(d)\frac{1}{q(\epsilon+i(\theta-\frac{a}{q}))^{1/2}},$$

(which is independent of μ). Arguing as in the case that $\dfrac{a}{q}=0$, we find for $\theta\in I(a,q), q\leq\lambda$

$$(F(\epsilon+i\theta))^d=\left(\frac{S(a,q)}{q}\right)^d\frac{C(d)}{(\epsilon+i(\theta-\frac{a}{q}))^{d/2}}+O(\lambda^{d/2}).$$

To obtain this approximate expression for $F(\epsilon+i\theta)$, it is necessary to show if $(a,q)=1$

$$|S(a,q,m)|\leq Cq^{1/2}$$

where

$$S(a,q,m)=\sum_{n=1}^{q}e^{-2\pi in^2\frac{a}{q}}e^{2\pi im\frac{a}{q}}.$$

This estimate is proved in the same manner as Lemma 2 below.

So, by a change of variables

$$\begin{aligned}&\int_{I(a,q)}(F(\epsilon+i\theta))^d e^{2\pi i\lambda^2\theta}\,\mathrm{d}\theta\\&=C(d)\left(\frac{S(a,q)}{q}\right)^d e^{2\pi i\lambda^2\frac{a}{q}}\int_{|\beta|\leq\frac{1}{\lambda q}}\frac{e^{2\pi i\lambda^2\beta}}{(\epsilon+i\beta)^{d/2}}\,d\beta+O(\lambda^{d/2-1}).\end{aligned}$$

Notice the factors $\left(\frac{S(a,q)}{q}\right)^d e^{2\pi i\lambda^2\frac{a}{q}}$ are just those arising in the formula (3) for $M_d(\lambda)$. Next we replace the range of integration $|\beta|\leq\dfrac{1}{\lambda q}$ by the entire real axis, making another error of order $\lambda^{d/2-1}$. Thus we find

$$\begin{aligned}r_d(\lambda)=C(d)\lambda^{d-2}\sum_{q\leq\frac{\lambda}{20}}\sum_{\substack{a=1\\(a,q)=1}}^{q}\left(\frac{S(a,q)}{q}\right)^d e^{2\pi i\lambda^2\frac{a}{q}}\\+O\left(\int_{\mathcal{E}_\lambda}|F(\epsilon+i\theta)|^d\,\mathrm{d}\theta+O(\lambda^{d/2})\right).\end{aligned}$$

But for $\theta \in \mathcal{E}_\lambda$, $q \geq \frac{\lambda}{20}$, so $|F(\epsilon + i\theta)| \leq C \left(\frac{S(a,q)}{q}\right)^d \frac{1}{\epsilon^{d/2}} \leq C \left(\frac{S(a,q)}{q}\right)^d \lambda^{d/2}$, and finally, since $\left(\frac{S(a,q)}{q}\right)^d \leq C \left(\frac{1}{\lambda}\right)^{d/2}$ for $q \geq \frac{\lambda}{20}$, we arrive at the formula (2).

It remains to prove Lemma 2. We use what is commonly called *Weyl differencing*. See [16] or [33].

$$S(a, q) = \sum_{n=0}^{q-1} e^{-2\pi i n^2 \frac{a}{q}}.$$

Now

$$|S(a, q)|^2 = \sum_{n=0}^{q-1} \sum_{m=0}^{q-1} e^{2\pi i (m^2 - n^2) \frac{a}{q}} = \sum_{n=0}^{q-1} \sum_{m=n}^{n+q-1} e^{2\pi i (m^2 - n^2) \frac{a}{q}}$$
$$= \sum_{n=0}^{q-1} \sum_{k=0}^{q-1} e^{2\pi i k(k+2n) \frac{a}{q}},$$

so

$$|S(a, q)|^2 \leq \sum_{k=0}^{q-1} \left| \sum_{n=0}^{q-1} e^{4\pi i k n \frac{a}{q}} \right|.$$

Since $(a, q) = 1$, the inner sum is zero for all but at most two values of k. Thus

$$|S(a, q)|^2 \leq 2q.$$

To show $M_d(\lambda)$ is bounded below for $d \geq 5$, we must consider

$$A_\lambda(q) = \sum_{\substack{a=1 \\ (a,q)=1}}^{q} e^{2\pi i \lambda^2 \frac{a}{q}} \left(\frac{S(a, q)}{q}\right)^d.$$

It turns out that if

$$(q_1, q_2) = 1, \quad A_\lambda(q_1 q_2) = A_\lambda(q_1) A_\lambda(q_2).$$

This is done in [8, Chap. 12] and in the more general context of studying the number of representations of an integer as a sum of d k–th powers in [18,32]. Thus

$$M_d(\lambda) = C(d) \prod_{\substack{p \\ p \text{ prime}}} (1 + A_\lambda(p) + \cdots + A_\lambda(p^m) + \cdots).$$

Next one can see that the proof of Lemma 2 shows $|S(a, q)| \leq \sqrt{2q}$ if $q \equiv 0(4)$, $S(a, q) = 0$ if $q \equiv 2(4)$ and $|S(a, q)| \leq \sqrt{q}$, if q is odd. Using these estimates and

Eq. (3), it is straight forward to check that the infinite product is bounded below. I learned this argument from [25].

Another argument can be found in [8].

It is interesting to note that, for p prime, one may interpret $1+A_\lambda(p)+\cdots+A_\lambda(p^m)$ in terms of solutions to the congruence

$$n_1^2+\cdots+n_d^2 \equiv \lambda^2 (\text{mod } p^m). \qquad (*)$$

Note that the number of solutions of

$$n_1^2+\cdots+n_d^2 \equiv \lambda^2 (\text{mod } p^m)$$

with $1 \le n_j \le p^m$ is

$$\begin{aligned}
&\frac{1}{p^m}\sum_{n_1=1}^{p^m}\cdots\sum_{n_d=1}^{p^m}\sum_{a=1}^{p^m} e^{2\pi i(\lambda^2-n_1^2-\cdots-n_d^2)\frac{a}{p^m}}\\
&= \frac{1}{p^m}\sum_{j=0}^{m}\sum_{n_1=1}^{p^m}\cdots\sum_{n_d=1}^{p^m}\sum_{\substack{a=1\\(a,p^{m-j})=1}}^{p^{m-j}} e^{2\pi i(\lambda^2-n_1^2-\cdots-n_d^2)\frac{a}{p^{m-j}}}\\
&= \frac{1}{p^m}\sum_{j=0}^{m} p^{jd}\sum_{n_1=1}^{p^{m-j}}\cdots\sum_{n_d=1}^{p^{m-j}}\sum_{\substack{a=1\\(a,p^{m-j})=1}}^{p^{m-j}} e^{2\pi i(\lambda^2-n_1^2-\cdots-n_d^2)\frac{a}{p^{m-j}}}\\
&= \frac{1}{p^m}\sum_{j=0}^{m} p^{jd} e^{2\pi i\lambda^2\frac{a}{p^{m-j}}}[S(a,\, p^{m-j})]^d\\
&= p^{m(d-1)}\sum_{j=0}^{m}\frac{1}{p^{(m-j)d}}\sum_{\substack{a=1\\(a,p^{m-j})=1}}^{p^{m-j}} e^{2\pi i\lambda^2\frac{a}{p^{m-j}}}[S(a,\, p^{m-j})]^d.
\end{aligned}$$

So

$$\sum_{j=0}^{m}\sum_{\substack{a=1\\(a,p^{m-j})=1}}^{p^j}\left(\frac{S(a,\, p^j)}{p^j}\right)^d e^{2\pi i\lambda^2\frac{a}{p^{m-j}}} = \frac{1}{p^{md-1}}\cdot(\text{number of solutions of } (*)).$$

Thus

$$1+A_\lambda(p)+\cdots+A_\lambda(p^m) = p^{m(1-d)}N(\lambda,\, p^m) \qquad (**)$$

where

$$N(\lambda,\, p^m) = \text{number of solutions of the congruence } (*)$$

A generalization of $(**)$ becomes important in studying the number of ways of representing an integer m as a sum of ℓ kth powers. See [32, Chap. 2] or [18, Chap. 5].

3 The Number of Representations of an Integer as a Sum of Primes

In the discussion of $r_d(\lambda)$, we were able to accurately describe the generating function, $F(\epsilon + i\theta)$, for every θ. In many applications of the circle method, this is not possible, and the major difficulty arises in estimating the generating function on the set on which a really good approximation is unknown. A case in point is the problem of representing an integer N as a sum of two or three primes. We will give a short introduction to this topic. Details may be found in [7] or [22]. Thus, we let $\rho_2(N)$ denote the number of representations of an even integer as a sum of two primes and $\rho_3(N)$ the number of representations of an odd integer as a sum of three primes. We will first discuss what we might expect the size of $\rho_2(N)$ and $\rho_3(N)$ to be. Then we shall try to understand why one can successfully treat $\rho_3(N)$ but not $\rho_2(N)$. The substitute for $F(\epsilon + i\theta)$ will be

$$S_N(\theta) = \sum_{p \le N} e^{2\pi i p\theta}.$$

(In this section, p will always denote a prime.) We will indicate how $S_N(\theta)$ is described well on a small set called the major arcs, and finally we shall try to give some hint as to how $S_N(\theta)$ is estimated for θ not in the major arcs.

Let us first make some guess as to the size of $\rho_2(N)$ and $\rho_3(N)$. Consider first $\rho_2(N)$. The number of ways of writing

$$n = p_1 + p_2$$

with p_1 and p_2 is the number of primes in the sequence $n - p_1$, with p_1 prime. This latter sequence has about $\dfrac{n}{\log n}$ terms, so if the primes were uniformly distributed in this sequence we would expect

$$p_2(n) \sim \frac{\frac{n}{\log n}}{\log\left(\frac{n}{\log n}\right)} \sim \frac{n}{\log^2 n}.$$

We proceed to discuss $\rho_3(N)$. Again the sequence $n - p$, p prime has roughly $\dfrac{n}{\log n}$ elements. If for most of these p, $n - p = p_2 + p_3$ in about $\dfrac{n}{\log^2 n}$ ways, we would expect

$$\rho_3(N) \sim \frac{n^2}{\log^3 n}.$$

And in fact Vinogradov proved

Theorem 3 *There are positive constants N_0, C_1 and C_2 such that for $n \geq N_0$ and n odd,*

$$C_1 \frac{n^2}{\log^3 n} \leq \rho_3(N) \leq C_2 \frac{n^2}{\log^3 n}.$$

We give a brief introduction to the proof of Theorem 3 together with an explanation as to why the study of representing integers as sums of three primes is more tractable that handling the analogous problem for two primes. For more details, consult [22], which is the book I followed when I taught the material.

Let $S_N(\theta) = \sum\limits_{1 \leq p \leq N} e^{2\pi i p\theta}$. At the present time, it is possible to find a good approximation to S_N for only small set of θ's (as N gets large). Call this set U_N. Then write

$$\rho_3(N) = \int_0^1 e^{2\pi i N\theta} [S_N(\theta)]^3 \, \mathrm{d}\theta = \int_{U_N} + \int_{C(U_N)}.$$

The integral over U_N will give the main contribution $\sim \dfrac{N^2}{\log^3 N}$. Thus we have to prove $\displaystyle\int_{C(U_N)}$ is say $\mathcal{O}\dfrac{N^2}{\log^4 N}$. We can estimate $\displaystyle\int_{C(U_N)}$ by

$$\sup_{\theta \in C(U_N)} |S_N(\theta)| \int_0^1 |S_N(\theta)|^2 \, \mathrm{d}\theta \leq C \sup_{\theta \in C(U_N)} |S_N(\theta)| \frac{N}{\log N}$$

for some constant C by the Plancherel Theorem since the coefficients, a_n, of $S_N(\theta)$ are 1 if n is a prime and 0 otherwise, $\displaystyle\int_0^1 |S_N(\theta)|^2 \, \mathrm{d}\theta$ is just the number of primes $\leq N$.

On the other hand, one could not proceed this way in studying $\rho_2(N)$, for if one took a power of $S_N(\theta)$ out of the integral

$$\int_0^1 [S_N(\theta)]^2 \, e^{2\pi i N\theta} \, \mathrm{d}\theta,$$

one would no longer be in a position to use Plancherel's Theorem.

It turns out that the main contribution comes from small intervals around $\dfrac{a}{q}$ with $(a, q) = 1$, and $1 \leq q \leq \log^4 N$ with N large.

Let us see how one finds an approximation for $S_N(\theta)$. Note first that if $(a, q) = 1$,

$$S_N\left(\frac{a}{q}\right) = \sum_{r=1}^{q} e^{2\pi i r \frac{a}{q}} \, \pi(N, r, q)$$

where $\pi(N, r, q)$ is the number of primes $\leq N$ which are congruent to $r \mod q$. A theorem of Siegel asserts that if $(r, q) = 1$

$$\pi(N, r, q) = \frac{1}{\phi(q)} L(N) + \mathcal{O}_A\, N\, e^{-c\sqrt{\log N}}$$

uniformly for $q \leq (\log N)^A$ for any positive A. Here $\phi(q) =$ the number of integers q which are relatively prime to q and $L(N) = \int_2^N \frac{\mathrm{d}t}{\ln t}$. Thus for $q \leq (\log N)^A$

$$S_N\left(\frac{a}{q}\right) = \frac{1}{\phi(q)} L(N) \sum_{\substack{r=1\\(r,q)=1}}^{q} e^{2\pi i r \frac{a}{q}} + \mathcal{O}(N\, e^{-c\sqrt{\log N}}).$$

The r sum can be evaluated with the help of the Möbius inversion formula. The Möbius function $\mu(d)$ is defined as follows:

$$\mu(d) : \begin{cases} \mu(1) = 1; \\ \mu(d) = (-1)^r, \text{ if } d \text{ is the product of } r \text{ distinct primes;} \\ \mu(d) = 0, \qquad \text{if } d \text{ is divisible by a square.} \end{cases}$$

The Möbius inversion formula states that

$$\sum_{d|n} \mu(d) = \begin{cases} 1, \text{ if } n = 1; \\ 0, \text{ otherwise.} \end{cases}$$

The standard proof of the Möbius inversion formula given in elementary number theory always seemed mysterious to me. There is another proof using the Riemann zeta function that seems more natural to me.

For Re$s > 1$,

$$\zeta(s) = \sum_1^\infty \frac{1}{n^s} = \prod_{\substack{p\\p \text{ prime}}} \frac{1}{\left(1 - \frac{1}{p^s}\right)}.$$

So

$$\frac{1}{\zeta(s)} = \prod_p \left(1 - \frac{1}{p^s}\right) = \sum_1^\infty \frac{\mu(n)}{n^s}.$$

Now if $A(s) = \sum_{n=1}^\infty \frac{a(n)}{n^s}$ and $B(s) = \sum_{n=1}^\infty \frac{b(n)}{n^s}$ are two Dirichlet series then

$$A(s)B(s) = \sum_{n=1}^\infty \sum_{m=1}^\infty \frac{a(n)}{n^s} \frac{b(m)}{m^s} = \sum_{k=1}^\infty \frac{1}{k^s} \sum_{\substack{n=1\\k|n}}^\infty a(n)\, b\left(\frac{n}{k}\right).$$

$$\zeta(s)\frac{1}{\zeta(s)} = 1 = \frac{1}{1^s} + \frac{0}{2^s} + \frac{0}{3^s} + \cdots.$$

Take $a(n) = 1$ and $b(m) = \mu(m)$ and get

$$\sum_{k|n} \mu\left(\frac{n}{k}\right) = 0, \text{ if } k \neq 1,$$
$$\sum_{d|n} \mu(d) = 0, \text{ if } n \neq 1,$$
$$\mu(1) = \sum_{d|1} \mu(d) = 1.$$

The Möbius inversion formula is often used in summing over values of r where r is restricted to be relatively prime to another integer q. Thus

$$\sum_{\substack{r=1 \quad q \\ (r,q)=1}} e^{2\pi i r \frac{a}{q}} = \sum_{r=1}^{q} e^{2\pi i r \frac{a}{q}} \sum_{d|(r,q)} \mu(d) = \sum_{d|q} \mu(d) \sum_{\substack{d|r \\ 1\leq r\leq q}} e^{2\pi i r \frac{a}{q}}$$
$$= \sum_{d|q} \mu(d) \sum_{m=1}^{q/d} e^{2\pi i m d \frac{a}{q}} = \mu(q)$$

Since the inner sum is zero for $d \neq q$. Thus one finds

$$S_N\left(\frac{a}{q}\right) = \frac{\mu(q)}{\phi(q)} L(N) + \mathcal{O}(N\, e^{-c\sqrt{\log N}}).$$

Note that the main term does not depend on a as opposed to $S(a, q)$ arising in the study of $r_d(\lambda)$. This is a big advantage in some problems. Also $\phi(q) > c\,\dfrac{q}{\ln \ln q}$. Thus the factor $\dfrac{\mu(q)}{\phi(q)}$ is better than the corresponding factor $\dfrac{S(a,q)}{q}$ which arose before. Nest we note that we can find a good approximation to $S_N\left(\dfrac{a}{q} + \beta\right)$ if $|\beta| \leq \dfrac{|\log N|^A}{N}$, $q \leq \log^A N$. To see this, we write

$$S_N\left(\frac{a}{q} + \beta\right) = \sum_{p\leq N} e^{2\pi i p \frac{a}{q}} e^{2\pi i p \beta}.$$

Put $\Lambda(x) = \sum_{p\leq x} e^{2\pi i p \frac{a}{q}}$. Then

$$S_N\left(\frac{a}{q} + \beta\right) = \int_{3/2}^{N} e^{2\pi i t \beta} d\Lambda(t)\, \mathrm{d}t$$
$$= \Lambda(N) e^{2\pi i N \beta} - 2\pi i \beta \int_{3/2}^{N} \Lambda(t) e^{2\pi i t \beta}\, \mathrm{d}t$$

$$= \Lambda(N)e^{2\pi i N\beta} - 2\pi i\beta \int_{3/2}^{N} \frac{\mu(q)}{\phi(q)} L(t)e^{2\pi i t\beta}\,\mathrm{d}t + \mathcal{O}(N\,e^{-\sqrt{\log N}})$$

if $|\beta| \le C\dfrac{(\log N)^U}{N}$.

Thus another integration by parts shows

$$S_N\left(\frac{a}{q}+\beta\right) = \frac{\mu(q)}{\phi(q)} \int_{3/2}^{N} \frac{1}{\ln t}\, e^{2\pi i t\beta}\,\mathrm{d}t + \mathcal{O}(N\,e^{-c\sqrt{\log N}})$$

for $q \le (\log N)^A$ and $|\beta| \le \dfrac{(\log N)^A}{N}$. This is the set U_N. Note that $|U_N| \le \dfrac{(\log N)^{3A}}{N}$.

To estimate $S_N(\theta)$ in the complement of U_N, Vinogradov used Schwartz's inequality in a very clear way. Suppose

$$T = \sum_{n=1}^{N}\sum_{m=1}^{N} d_n b_m e^{2\pi i n m\theta}.$$

Then even if the d_n and b_m are very rough, there can be cancelation in the double sum for T. To fix matters consider

$$T = \sum_{n=1}^{q}\sum_{m=1}^{q} d_n b_m e^{2\pi i m n \frac{a}{q}}$$

with $(a,q)=1$. Let $D = \left(\sum_{n=1}^{q} d_n^2\right)^{1/2}$ and $B = \left(\sum_{m=1}^{q} b_m^2\right)^{1/2}$. Then the trivial estimate would be

$$T \le qDB.$$

In fact one has the estimate

$$T \le \sqrt{q}DB. \qquad (*)$$

If $q \ge (\log N)^A$, since we are talking about beating a trivial estimate by a small power of $\log N$, this makes a tremendous saving.

To see $(*)$ apply Schwartz's inequality to the outer sum to see

$$|T| = D\left\{\sum_{n=1}^{q}\left|\sum_{m=1}^{q} b_m e^{2\pi i m n\frac{a}{q}}\right|^2\right\}^{1/2} = D\left\{\sum_{m_1=1}^{q}\sum_{m_2=1}^{q}\sum_{n=1}^{q} b_{m_1}\overline{b_{m_2}} e^{2\pi i(m_1-m_2)n\frac{a}{q}}\right\}^{1/2}.$$

Now the sum on n is zero unless $m_1 = m_2$ in which case it is q. Thus $|T| \le \sqrt{q}DB$.

To see how double sums arise in studying $S_N(\theta)$ note that

$$S_N(\theta) = \sum_{\substack{\sqrt{N} \le n \le N \\ (n, \underline{P})=1}} e^{2\pi i n\theta} + \mathcal{O}(\sqrt{N})$$

where

$$\underline{P} = \prod_{p \le N} p.$$

Now we have to apply the Möbius inversion formula.

$$\sum_{d|n} \mu(d) = \begin{cases} 1, & \text{if } n = 1; \\ 0, & \text{otherwise.} \end{cases}$$

So

$$S_N(\theta) = \sum_{\sqrt{N} \le n \le N} \sum_{d|(n,\underline{P})} \mu(d) e^{2\pi i n\theta} + \mathcal{O}(\sqrt{N}) = \sum_{d|\underline{P}} \mu(d) \sum_{\substack{n \\ d|n \\ \sqrt{N} \le n \le N}} e^{2\pi i n\theta}$$

The d's we have to be careful about are those of the size roughly of N. We write $n = md$ and get a sum of the form

$$\sum_d \mu(d) \sum_m e^{2\pi i m\theta}.$$

If d is large, m must be small. Also all but a negligible number of large d have a large prime factor p, say $p > e^{\sqrt{\log N}}$. Now the idea roughly to write $d = pd_1$ where $p > e^{\sqrt{\log N}}$. Then the range of summation on d_1 is $d_1 \le \frac{N}{e^{\sqrt{\log N}}}$. Now the sum is something like

$$\sum_{d_1 \le \frac{N}{e^{\sqrt{\log N}}}} \mu(d_1) \sum_m e^{2\pi i d_1 m p\theta}$$

and one can control the size of mp.

So roughly the sum becomes

$$\sum_{d_1 \le \frac{N}{e^{\sqrt{\log N}}}} \mu(d_1) \sum_{\substack{\ell \\ \ell \text{ not too large}}} d(\ell) e^{2\pi i d_1 \ell\theta},$$

where $d(\ell)$ is dominated by the number of divisors of ℓ. This is now the type of double sum that can be controlled by an application of Schwartz's inequality as described above. For more details see Pracher [22].

4 Further Reading

I taught a one semester course in the Fall semester 2002 on the circle method. I covered two topics: (1) The number of solutions in integers of $m = n_1^k + \cdots + n_l^k$ and (2) Vinogradov's Theorem on the representation of an integer as a sum of three primes. The study of $r_d(\lambda)$ was a simplified version of [15] together with arguments for the singular series for the general problem of the number of solution of $m = n_1^k + \cdots + n_l^k$, I followed [15]. See also [31]. For Vinogradov's Theorem, I followed the treatment in Pracher [22], an algebraic approach to the study of $r_d(\lambda)$ can be found in [24].

Acknowledgements This research supported in part by NSF grant DMS-0098757 at the University of Wisconsin. I would like to thank Guido for over 60 years of friendship, guidance, and encouragement. I would also like to thank the referee for his helpful suggestions.

References

1. Arkipov, G.I., Oskolkov, K.I.: On a special trigonometric series and its applications. Mat. Sb. **134**(176), 147–158 (1987)
2. Arkipov, G.I., Oskolkov, K.I.: On a special trigonometric series and its applications. Sov. Math. **62**, 145–156 (1989)
3. Bateman, P.T.: On the representation of a number as the sum of three squares. Trans. Am. Soc. **71**, 70–101 (1951)
4. Bourgain, J.: Onthe maximal ergodic theorem for certain sequence of integers. Isr. J. Math. **61**, 39–72 (1988). 73–83
5. Bourgain, J.: Pointwise ergodic theorems for arithmetic sets with an appendix by the author, Furstenberg, Kutznelson, and Ornstein. Inst. Hautes Etudes Sci. Publ. Math. **69**, 5–45 (1989)
6. Bourgain, J.: Fourier transform restriction phenomena for certain lattice subsets and its applications to nonlinear evolution equations. Geom. Funct. Anal. **3**, 107–156, 157–178, 209–262 (1993)
7. Estermann, T.: Introduction to Prime Number Theory. Cambridge University Press, Cambridge (1952)
8. Grosswald, Emil: Representation of Integers as Sums of Squares. Springer, New York (1985)
9. Hardy, G.H.: On the representation of a number of a number as a sum of any number of squares, and in particular of five. Trans. Am. Math. Soc. **21**, 255–284 (1920)
10. Hardy, G.H., Littlewood, J.E.: A new solution of Waring Problem. Q. J. Math. **48**, 272–293 (1920)
11. Ionescu, A.: An endpoint estimate for the discrete spherical maximal function. Proc. Am. Soc. **132**(5), 1411–1417 (2004)
12. Kloosterman, H.D.: On the representation of numbers in the form $ax^2 + by^2 + cz^2 + dt^2$. Acta. Math. **49**, 407–464 (1927)
13. Knopp, M.I.: Modular Forms in Analytic Number Theory. Markham Publishing Company, Chicago (1970)
14. Magyar, A.: Diophantine equations and Ergdis theorems. Am. J. Math. **124**, 921–953 (2002)
15. Magyar, A., Stein, E.M., Wainger, S.: Discrete analogues in harmonic analysis: spherical averages. Ann. Math. **155**, 189–208 (2002)
16. Montgomery, H.: Ten Lectures on the Interface Between Analytic Number Theory and Harmonic Analysis. American Mathematical Society, Providence (1994)
17. Mordell, L.J.: On the representation of numbers as a sum of $2r$ squares. Q. J. Pure Appl. Math. **48**, 93–104 (1917)
18. Nathanson, M.: Additive Number Theory. Springer, New York (1996)
19. Oberlin, D.: Two discrete fractional integrals. Math. Res. Lett. **8**, 1–6 (2001)
20. Ono, K.: Representation of integers as a sum of squares. J. Number Theory **95**, 253–258 (2002)
21. Oskolkow, K.: Schrodinger equation and oscillatory Hilbert transform of second degree. J. Fourier Anal. Appl. **4**, 341–356 (1988)
22. Pracher, K.: Primzahlverteilung. Springer, New York (1957)

23. Schlag, W.: On minima of the absolute value of certain random exponential sums. Am. J. Math. **122**, 483–514 (2000)
24. Siegel, C.L.: Lectures on Analytic Theory of Quadratic Forms. Prince University Press, Princeton (1962)
25. Stein, E.M.: Discrete Analogues of Singular Integral Operators. Unpublished lecture notes
26. Stein, E.M., Shakarchi, R.: Princeton Lectures in Analysis II, Complex Analysis. Prince University Press, Princeton (2003)
27. Stein, E.M., Wainger, S.: Discrete analogues of singular Radon transforms. Bull. A.M.S. **23**, 537–544 (1990)
28. Stein, E.M., Wainger, S.: Discrete analogues in harmonis analysis, I: ℓ^2 estimates for singular Radon transforms. Am. J. Math. **121**, 1291–1336 (1999)
29. Stein, E.M., Wainger, S.: Two discrete fractional integral operators revisited. J. D'Anul. Math. **87**, 451–479 (2002)
30. Stein, E.M., Weiss, G.: Introduction to Fourier Analysis on Euclidean Spaces. Prince University Press, Princeton (1971)
31. Vaughn, R.C.: The Hardy–Littlewood Method. Cambridge University Press, Cambridge (1997)
32. Vinogradov, I.M.: The Method of Trigonometrical Sums in the Theory of Numbers. Interscience, New York (1954)
33. Weyl, H.: Uber die Gleichverteilung von Zahlen mod. Eins. Math. Ann. **77**, 313–336 (1916)
34. Wierdl, M.: Pointwise ergodic theorems along the prime numbers. Isr. J. Math. 64, 315–336 (1988)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.