CHAPTER 1

Sums of squares

One of the great theorems in arithmetic in the 19th century is Lagrange's *four square theorem*:

THEOREM (Lagrange, 1770). *A non-zero integer n is representable as a sum of four squares, that is, there exists $(a, b, c, d) \in \mathbb{Z}^4$ such that*

$$n = a^2 + b^2 + c^2 + d^2. \tag{0.1}$$

if and only if $n > 0$.

A quadruple $(a, b, c, d) \in \mathbb{Z}^4$ satisfying the above equation is called *an integral representation* of n as a sum of four squares. Alternatively, let q_4 denote the Euclidean quadratic form on $\mathbb{Q}^4$

$$q_4(a, b, c, d) = a^2 + b^2 + c^2 + d^2,$$

the quadruple (a, b, c, d) is called an (integral) *representation* of n by q_4.

Let $\mathrm{R}_4(n)$ be the set of all such representations:

$$\mathrm{R}_4(n) = \{(a, b, c, d) \in \mathbb{Z}^4,\ a^2 + b^2 + c^2 + d^2 = n\};$$

for $n \neq 0$, that set is non-empty if an only if n is positive.

1. A proof of Lagrange's theorem after A. Venkov

We shall now give a proof of Lagrange's theorem following ideas of A: Venkov [**Ven22**, **Ven29**]: this involve the introduction into the picture of another object: the algebra of *Hamilton*[1] *quaternions*.

1.1. The Hamilton quaternion. This is the 4-dimensional associative, non commutative, $\mathbb{Q}$-algebra

$$\mathrm{B}(\mathbb{Q}) = \mathbb{Q} + \mathbb{Q}.i + \mathbb{Q}.j + \mathbb{Q}.k$$

with i, j, k satisfying the relations

$$i^2 = j^2 = k^2 = -1,\ ij = -ji = k.$$

The center of B is the algebra of scalars

$$Z_{\mathrm{B}}(\mathbb{Q}) = \mathbb{Q} = \mathbb{Q}.1.$$

[1] Discovered by Hamilton in 1843

The algebra B may be explicitely realized as a $\mathbb{Q}$-algebra of 2×2 (complex) matrices via

$$1 \mapsto \mathrm{Id},\ i \mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},\ j \mapsto \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix},\ k \mapsto \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}. \tag{1.1}$$

Alternatively, letting $\mathrm{B}(\mathbb{Q})$ act on itself by left multiplications (viewing $\mathrm{B}(\mathbb{Q}) \simeq \mathbb{Q}^4$ as a $\mathbb{Q}$-vector space), one also may realize B as an algebra of 4×4 matrices: the map defined by

$$a + bi + cj + dk \mapsto \begin{pmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{pmatrix} \tag{1.2}$$

is an isomorphism of $\mathbb{Q}$-algebras.

1.2. Reduced Norm, Reduced trace. The algebra B is endowed with an involutive $\mathbb{Q}$-linear anti-automorphism: the *canonical involution*

$$z = a + bi + cj + dk \mapsto \overline{z} = a - bi - cj - dk$$

(ie. satisfying

$$\overline{z_1 z_2} = \overline{z_2}.\overline{z_1}.)$$

Remark 1.1. When B is identified with the algebra of 2×2 complex matrices generated by the four matrices given in (1.1), the canonical involution is the "conjugate-transpose" map. When B is realized as an algebra of 4×4 rational matrices, the canonical involution is just the "transpose" map

Out of the canonical involution, one defines the $\mathbb{Q}$-valued *reduced trace* and *reduced norm* maps

$$\mathrm{tr}(z) = z + \overline{z} = 2a, \tag{1.3}$$

$$\mathrm{Nr}(z) = z\overline{z} = a^2 + b^2 + c^2 + d^2. \tag{1.4}$$

The trace is $\mathbb{Q}$-linear and satisfies

$$\mathrm{tr}(z_1 z_2) = \mathrm{tr}(z_2 z_1).$$

From its concrete expression (1.4), the norm defines a quadratic form, who polarization is the bilinear, symetric function of two variables given by

$$(z_1, z_2) \to \mathrm{Nr}(z_1 + z_2) - \mathrm{Nr}(z_1) - \mathrm{Nr}(z_2) = z_1\overline{z_2} + \overline{z_1}z_2 = \mathrm{tr}(z_1\overline{z}_2).$$

In particular the linear map

$$\begin{array}{ccc} \mathrm{B}(\mathbb{Q}) & \mapsto & \mathbb{Q}^4 \\ a + bi + cj + dk & \mapsto & (a, b, c, d) \end{array}, \tag{1.5}$$

is an isometry between the quadratic spaces $(\mathrm{B}(\mathbb{Q}), \mathrm{Nr})$ and $(\mathbb{Q}^4, q_4)$ sending the basis $\{1, i, j, k\}$ to the canonical basis $\{e_1, e_2, e_3, e_4\}$.

Another quite remarkable property of the norm is that it is *multiplicative*:

$$\mathrm{Nr}(z_1 z_2) = \mathrm{Nr}(z_1)\,\mathrm{Nr}(z_2);$$

indeed

$$\begin{aligned}\mathrm{Nr}(z_1z_2) = z_1z_2\overline{z_1z_2} = z_1z_2\overline{z_2}\,\overline{z_1} = z_1\,\mathrm{Nr}(z_2)\overline{z_1}\\ = \mathrm{Nr}(z_2)z_1\overline{z_1} = \mathrm{Nr}(z_1)\,\mathrm{Nr}(z_2).\end{aligned}$$

(in that way one retrieves the so-called *Lagrange identity* in dimension 4)

The quadratic form Nr is obviously *anisotropic*: given $z \in \mathrm{B}(\mathbb{Q})$, one has $\mathrm{Nr}(z) = 0$ if and only if $z = 0$.

A consequence is that a quaternion z is invertible in $\mathrm{B}(\mathbb{Q})$ if and only if it is non zero, its inverse being:

$$z^{-1} = \overline{z}/\,\mathrm{Nr}(z).$$

In other terms $\mathrm{B}(\mathbb{Q})$ is a *division algebra* (or a skew field): the multiplicative group of invertible elements of $\mathrm{B}(\mathbb{Q})$ equals

$$\mathrm{B}^\times(\mathbb{Q}) = \mathrm{B}(\mathbb{Q}) - \{0\}.$$

Remark. The adjective *reduced*, for the trace or the norm come from the fact that, if we view $\mathrm{B}(\mathbb{Q})$ as a subalgebra of the algebra, $\mathrm{End}_{\mathbb{Q}}(\mathrm{B}(\mathbb{Q}))$, of linear endomorphism of the $\mathbb{Q}$-vector space (cf. the matrix realization 1.2), the trace and the determinant of a quaternion z are equal respectively, to *twice* the *reduced trace* and to the *square* of the reduced norm of z.

1.2.1. *The characteristic polynomial.* Any $z \in \mathrm{B}(\mathbb{Q})$ is annihilated by the quadratic polynomial with rational coefficients:

$$P_z(X) = X^2 - \mathrm{tr}(z)X + \mathrm{Nr}(z) = (X - z)(X - \overline{z}); \tag{1.6}$$

We have evidently

$$|\,\mathrm{tr}(z)|^2 \leq 4\,\mathrm{Nr}(z)$$

and moreover this inequality is strict if and only if z is not a scalar ($z \notin \mathbb{Q}$); in such a case, $P_z(X)$ is an irreducible polynomial (even over $\mathbb{R}$) and the commutative $\mathbb{Q}$-algebra generated by z, $\mathbb{Q}[z] = \mathbb{Q}(z)$ is a (imaginary) quadratic field embedded into $\mathrm{B}(\mathbb{Q})$ and one has the equalities between traces and norms

$$\mathrm{tr}(z) = \mathrm{tr}_{\mathbb{Q}(z)/\mathbb{Q}}(z),\ \mathrm{Nr}(z) = N_{\mathbb{Q}(z)/\mathbb{Q}}(z).$$

1.2.2. *Trace* 0*-quaternions.* The kernel of the trace map is a linear subspace of $\mathrm{B}(\mathbb{Q})$, called the space of *trace zero* or *pure* quaternions

$$\mathrm{B}^0(\mathbb{Q}) = \{z \in \mathrm{B}(\mathbb{Q}),\ \mathrm{tr}(z) = 0\} = \{bi + cj + dk,\ b, c, d \in \mathbb{Q}\}.$$

1.2.3. *Norm*-1 *quaternions.* Similarly we denote by $\mathrm{B}^{(1)}(\mathbb{Q})$ the kernel of the norm homomorphism on $\mathrm{B}^\times(\mathbb{Q})$:

$$\mathrm{B}^{(1)}(\mathbb{Q}) = \{z \in \mathrm{B}(\mathbb{Q}),\ \mathrm{Nr}(z) = 1\} = \{a+bi+cj+dk,\ a^2+b^2+c^2+d^2 = 1\}.$$

1.3. The integral quaternions and the Hurwitz quaternions. From the shape of the reduced norm (1.4), we see that the question of determining the representations of n as a sum of four squares is equivalent to determining the quaternions of norm n with having integral entries: in other terms, let

$$\mathrm{B}(\mathbb{Z}) = \mathbb{Z} + \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}k$$

be the set of integral quaternion and let

$$\mathrm{B}^{(n)}(\mathbb{Z}) = \{z = a + bi + cj + dk,\ a, b, c, d \in \mathbb{Z},\ \mathrm{Nr}(z) = n\}$$

be the subset of integral quaternion of norm n; the map

$$(a, b, c, d) \mapsto a + bj + cj + dk \tag{1.7}$$

yields a bijection between $\mathrm{R}_4(n)$ and $\mathrm{B}^{(n)}(\mathbb{Z})$.

The $\mathbb{Z}$-module $\mathrm{B}(\mathbb{Z})$ is a subring of $\mathrm{B}(\mathbb{Q})$ (ie. it contain the identity and is stable under multiplication) and is free rank 4: in other terms $\mathrm{B}(\mathbb{Z})$ is an *order* in $\mathrm{B}(\mathbb{Q})$. As it turns out, $\mathrm{B}(\mathbb{Z})$ is contained is a slightly bigger order (with index 2): the ring of *Hurwitz quaternions*

$$\mathscr{O}_{\mathrm{B}} = \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}k + \mathbb{Z}\frac{1 + i + j + k}{2} = \mathbb{Z}[i, j, k, \frac{1 + i + j + k}{2}].$$

The proof of Lagrange's theorem rests of the following:

THEOREM 1.1. *The ring of Hurwitz quaternions satisfies the following properties*

1. *The elements of $\mathscr{O}_{\mathrm{B}}$ have integral reduced norm and trace.*
2. *The group of units of $\mathscr{O}_{\mathrm{B}}$ and $\mathrm{B}(\mathbb{Z})$ are*

$$\mathrm{B}(\mathbb{Z})^{\times} = \{\pm 1, \pm i, \pm j, \pm k\},$$

$$\mathscr{O}_{\mathrm{B}}^{\times} = \{\pm 1, \pm i, \pm j, \pm k, \frac{\pm 1 \pm i \pm j \pm k}{2}\}.$$

3. *For any $z \in \mathscr{O}_{\mathrm{B}}$ there is a unit $w \in \mathscr{O}_{\mathrm{B}}^{\times}$ such that $wz \in \mathrm{B}(\mathbb{Z})$.*
4. *The ring $\mathscr{O}_{\mathrm{B}}$ is principal: any left (resp. right) $\mathscr{O}_{\mathrm{B}}$-ideal $I \subset \mathrm{B}(\mathbb{Q})$ is of the form $\mathscr{O}_{\mathrm{B}}.z$ (resp. $z.\mathscr{O}_{\mathrm{B}}$) for some $z \in \mathrm{B}^{\times}$*

PROOF. (1) is a general property of elements of finitely generated (as $\mathbb{Z}$-module) subrings of $\mathrm{B}(\mathbb{Q})$: if z is contained into some finitely generated subring, the commutative ring $\mathbb{Z}[z]$ is finitely generated and hence the monic polynomial $P_z(X)$ (cf. (1.6)) which annihilates z has integral coefficients.

Property (2) follows immediately from the fact, a consequence of (1), that $z \in \mathscr{O}_{\mathrm{B}}$ is a unit if and only if $\mathrm{Nr}(z) = 1$.

(3) is a simple computation.

As for (4): recall that a left (resp. right) $\mathscr{O}_{\mathrm{B}}$-ideal $I \subset \mathrm{B}$ is a finitely generated $\mathbb{Z}$-module of maximal rank (4) such that $\mathscr{O}_{\mathrm{B}}.I \subset I$ (resp. $I.\mathscr{O}_{\mathrm{B}} \subset I$). The principality is a consequence of the following

CLAIM. *the order* $\mathscr{O}_{\mathrm{B}}$ *is* euclidean *w.r.t the norm* Nr*: for any* $q, z \in \mathscr{O}_{\mathrm{B}}$, $q \neq 0$, *there is* $h, r \in \mathscr{O}_{\mathrm{B}}$ *such that*

$$z = hq + r,\ \mathrm{Nr}(r) < \mathrm{Nr}(q).$$

PROOF. (of the Claim) Indeed, considering $zq^{-1} \in \mathrm{B}$ there is $l \in \mathrm{B}(\mathbb{Z})$ such that $z.q^{-1} - l = a + bi + cj + dk$ has its coordinates contained in the cube of radius $1/2$ $[-1/2, 1/2]^4 = \{|a|, |b|, |c|, |d| \leq 1/2\}$ so that $\|(a, b, c, d)\| = (a^2 + b^2 + c^2 + d^2)^{1/2} = \mathrm{Nr}(z.q^{-1} - l)^{1/2} \leq 1$; moreover equality hold if and only if $|a| = |b| = |c| = |d| = 1/2$ but in that case replacing l by some $h = l + \frac{\pm 1 \pm i \pm j \pm k}{2} \in \mathscr{O}_{\mathrm{B}}$ we can always insure that $\|(a, b, c, d)\| = \mathrm{Nr}(z.q^{-1} - h)^{1/2} < 1$ hence $\mathrm{Nr}(z - hq) < \mathrm{Nr}(q)$. □

Now given I a $\mathscr{O}_{\mathrm{B}}$-left ideal; up to multiplying I by a scalar $a \in \mathbb{Q}^{\times}$ we may assume that $I \subset \mathscr{O}_{\mathrm{B}}$. Let $q \in I - \{0\}$ of minimal norm (it exists because I is discrete in $B(\mathbb{R})$); then for any $z \in I$, there is $h \in \mathscr{O}_{\mathrm{B}}$ such that $r = z - hq \in I$ has norm $< \mathrm{Nr}(q)$, but the minimality of $\mathrm{Nr}(q)$ guaranties that $r = 0$. □

We have the following consequence of (4)

COROLLARY 1.1. *The ring of Hurwitz quaternions is a maximal order of* $\mathrm{B}(\mathbb{Q})$. *Any order, and more generally, any finitely generated subring* $\mathscr{O} \subset \mathrm{B}(\mathbb{Q})$ *is conjugate to a subring of* $\mathscr{O}_{\mathrm{B}}$*: there is* $q \in \mathrm{B}^{\times}(\mathbb{Q})$ *such that*

$$q^{-1}\mathscr{O}q \subset \mathscr{O}_{\mathrm{B}}.$$

PROOF. The maximality property follows from the second part of the corollary. Let $\mathscr{O} \subset \mathrm{B}(\mathbb{Q})$ be any finitely generated subring, then $\mathscr{O}.\mathscr{O}_{\mathrm{B}}$ is finitely generated hence is a right $\mathscr{O}_{\mathrm{B}}$-ideal; it i therefore principal:

$$\mathscr{O}.\mathscr{O}_{\mathrm{B}} = q\mathscr{O}_{\mathrm{B}}$$

for some $q \in \mathrm{B}^{\times}(\mathbb{Q})$ and then

$$q^{-1}\mathscr{O}q \subset q^{-1}\mathscr{O}q\mathscr{O}_{\mathrm{B}} = q^{-1}\mathscr{O}.\mathscr{O}.\mathscr{O}_{\mathrm{B}} \subset q^{-1}\mathscr{O}.\mathscr{O}_{\mathrm{B}} = \mathscr{O}_{\mathrm{B}}.$$

□

REMARK. The ring $\mathrm{B}(\mathbb{Z})$ is not principal.

1.4. Proof of Lagrange theorem. Lagrange's theorem is equivalent to showing for any $n \geq 1$ the existence of an integral quaternion $z \in \mathrm{B}(\mathbb{Z})$ of norm n.

1. If $n = 1$, take for z any element of $\mathrm{B}(\mathbb{Z})^{\times}$, so we may assume that $n \geq 2$.

2. Since $\mathrm{Nr}(z_1 z_2) = \mathrm{Nr}(z_1)\,\mathrm{Nr}(z_2)$, we see, by decomposing n into a product of primes, that it is sufficient to show the existence of such a quaternion when $n = p$ a prime.

3. If $p = 2$, $z = 1 + i$ has norm 2 and clearly the set of integral quaternion of norm 2 is precisely $\mathrm{B}(\mathbb{Z})^{\times}.(1 + i)$ (so has order 8).

4. Let now p be an odd prime, we claim the following

LEMMA. *The equation*

$$a^2 + b^2 + c^2 = 0$$

has a solution in ${\mathbb{F}_p}^3 - \{(0,0,0)\}$.

PROOF. There are precisely $\frac{p+1}{2}$ squares in $\mathbb{F}_p$ $(({\mathbb{F}_p}^\times)^2 \cup \{0\})$, hence the sets $\{1 + b^2,\ b \in \mathbb{F}_p\}$ and $\{-c^2,\ c \in \mathbb{F}_p\}$ have cardinality $\frac{p+1}{2}$; since $\frac{p+1}{2} + \frac{p+1}{2} > p$ these sets have non empty intersection □

From the above lemma, there exists $(a,b,c) \in \mathbb{Z}^3$ such that $a \equiv 1 (\mathrm{mod}\ p)$ and

$$a^2 + b^2 + c^2 \equiv 0 (\mathrm{mod}\ p).$$

Set $w = ai + bj + ck \in \mathscr{O}_{\mathrm{B}}$ and let I be the left ideal $I = \mathscr{O}_{\mathrm{B}} w + \mathscr{O}_{\mathrm{B}} p$; one has

$$\mathscr{O}_{\mathrm{B}}.p \subsetneq I \subsetneq \mathscr{O}_{\mathrm{B}}.$$

Indeed the first inclusion is strict since $w \notin \mathscr{O}_{\mathrm{B}} p$ and the second because all the elements of I have norm divisible by p. By the above lemma I is principal; let z be a generatorof I, then $z \notin \mathscr{O}_{\mathrm{B}}^\times$ there is $z' \in \mathscr{O}_{\mathrm{B}}$ such that $p = z'z$. We have $\mathrm{Nr}(p) = p^2 = \mathrm{Nr}(z)\,\mathrm{Nr}(z')$ and since z' is not a unit (since I contains $\mathscr{O}_{\mathrm{B}}.p$ properly) we must have $\mathrm{Nr}(z) = \mathrm{Nr}(z') = p$. A priori z belong to $\mathscr{O}_{\mathrm{B}}$ but from (3) of the previous proposition we may multiply it by a unit $w \in \mathscr{O}_{\mathrm{B}}^\times$ such that $wz \in \mathrm{B}(\mathbb{Z})$ and $\mathrm{Nr}(wz) = \mathrm{Nr}(z) = p$. □

1.5. Quaternions and quadratic spaces. As we have seen, one of the main virtue of introducing the Hamilton quaternions has been to endow the quadratic space $(\mathbb{Q}^4, a^2 + b^2 + c^2 + d^4)$, with an extra multiplicative structure, via the linear isomorphism,

$$\begin{array}{ccc} \mathrm{B}(\mathbb{Q}) & \mapsto & \mathbb{Q}^4 \\ a + bi + cj + dk & \mapsto & (a,b,c,d) \end{array}, \tag{1.8}$$

so that the four squares quadratic form becomes multiplicative with respect to that structure[2].

In that section, we discuss in greater details the relationships between the quadratic and the quaternionic structures. The norm form is a quadratic form on $\mathrm{B}(\mathbb{Q})$ whose polarization is the inner product

$$\langle z, z' \rangle = \mathrm{Nr}(z + z') - \mathrm{Nr}(z) - \mathrm{Nr}(z') = \mathrm{tr}(z\overline{z}') \tag{1.9}$$

and the map (1.8) is an isometry of quadratic spaces $(\mathbb{Q}^4, q_4) \simeq (\mathrm{B}(\mathbb{Q}), \mathrm{Nr})$, hence

$$\mathrm{SO}_4 \simeq \mathrm{SO}_{\mathrm{B}}\,.$$

We now express SO_{B} in terms of quaternions; we refer to Chapter 10 for the details and for more general statements. Observe that given $z, z' \in \mathrm{B}^\times(\mathbb{Q})$, and $w \in \mathrm{B}(\mathbb{Q})$ one has

$$\mathrm{Nr}(zw{z'}^{-1}) = \mathrm{Nr}(z/z')\,\mathrm{Nr}(w),$$

[2] For general quadratic forms a related construction exists: it yields the *Clifford algebra* (or *algebra of spinors*).

therefore the linear map

$$\rho_{z,z'} : \begin{array}{ccc} \mathrm{B}(\mathbb{Q}) & \mapsto & \mathrm{B}(\mathbb{Q}) \\ w & \mapsto & zwz'^{-1} \end{array}$$

is an orthogonal similitude of the quadratic space $(\mathrm{B}(\mathbb{Q}), \mathrm{Nr})$ with similitude factor

$$\mu(\rho_{z,z'}) = \mathrm{Nr}(z/z').$$

Moreover a computation shows that

$$\det(\rho_{z,z'}) = \mu^2(\rho_{z,z'}),$$

so the above maps defines a group homomorphism

$$\rho : \begin{array}{ccc} \mathrm{B}^\times(\mathbb{Q}) \times \mathrm{B}^\times(\mathbb{Q}) & \mapsto & \mathrm{GSO}_\mathrm{B}(\mathbb{Q}) \simeq \mathrm{GSO}_4(\mathbb{Q}) \\ (z, z') & \mapsto & \rho_{z,z'} \end{array} .$$

where GSO_B denote the connected component of GO_B the group of orthogonal similitudes of $(\mathrm{B}, \mathrm{Nr})$ (the kernel of the character

$$\rho \in \mathrm{GO}_\mathrm{B} \mapsto \det(\rho)/\mu(\rho)^2).$$

This map is surjective and its kernel is the diagonal scalar subgroup

$$\Delta Z_B^\times(\mathbb{Q}) = \{(\lambda, \lambda),\ \lambda \in \mathbb{Q}^\times\}.$$

In fact, we obtain an isomorphism of $\mathbb{Q}$-algebraic groups

$$\mathrm{B}^\times \times \mathrm{B}^\times / \Delta Z_\mathrm{B}^\times \simeq \mathrm{GSO}_\mathrm{B}\,. \tag{1.10}$$

The kernel of the similitude factor μ is the group of special orthgonal isometries SO_B and restricting to it we obtain a isomorphism of $\mathbb{Q}$-algebraic groups

$$(\mathrm{B}^\times \times \mathrm{B}^\times)^{\mathrm{Nr}} / \Delta Z_\mathrm{B}^\times \simeq \mathrm{SO}_\mathrm{B} \tag{1.11}$$

where

$$(\mathrm{B}^\times \times \mathrm{B}^\times)^{\mathrm{Nr}} = \{(z, z') \in \mathrm{B}^\times \times \mathrm{B}^\times,\ \mathrm{Nr}(z) = \mathrm{Nr}(z')\}.$$

1.5.1. *The trace zero quaternions as quadratic space.* We consider now the three dimensional space of pure quaternion

$$\mathrm{B}^0(\mathbb{Q}) = \{z \in \mathrm{B}(\mathbb{Q}),\ \mathrm{tr}(z) = 0\}$$

with its quadratic space structure given by the norm form. From (1.9) we see that $\mathrm{B}^0(\mathbb{Q})$ is precisely the subspace of vectors orthogonal to the quaternion 1,

$$\mathrm{B}^0(\mathbb{Q}) = \mathbb{Q}.1^\perp = \{z \in \mathrm{B}(\mathbb{Q}),\ \langle z, 1\rangle = \mathrm{tr}(z.1) = \mathrm{tr}(z) = 0\}.$$

It follows that its special orthogonal group $\mathrm{SO}_{\mathrm{B}^0}$ is the stabilizer of the vector 1 in GSO_B. In term of quaternions, this subgroup is given by the isometries of the shape $\rho_{z,z}$ for $z \in \mathrm{B}^\times$; in other terms

$$\mathrm{SO}_{\mathrm{B}^0} = \mathrm{Stab}_{\mathrm{GSO}_\mathrm{B}}(1) \simeq \{(z, z),\ z \in \mathrm{B}^\times\}/\Delta Z_\mathrm{B}^\times \simeq \mathrm{B}^\times / Z_\mathrm{B}^\times = \mathrm{PB}^\times$$

the *projective group* of invertible quaternions.

The map (1.8) induces an isomorphism between the quadratic spaces

$$(\mathbb{Q}^3, b^2 + c^2 + d^2) \simeq (\mathrm{B}^0(\mathbb{Q}), \mathrm{Nr})$$

and therefore we have an isomorphism of $\mathbb{Q}$-algebraic groups

$$\rho : \mathrm{PB}^\times \simeq \mathrm{SO}_{\mathrm{B}^0} \simeq \mathrm{SO}_3\,. \tag{1.12}$$

1.5.2. *The group of quaternions of norm one.* Let

$$\mathrm{B}^{(1)} = \{z \in \mathrm{B},\ \mathrm{Nr}(z) = 1\} \subset \mathrm{B}^\times$$

denote the variety of quaternions of norm 1; being the kernel of the norm homomorphism, this is a normal algebraic subgroup of $\mathrm{B}^\times$. By the inclusion $\mathrm{B}^{(1)} \hookrightarrow \mathrm{B}^\times$, $\mathrm{B}^{(1)}$ acts on B^0 and one has a short exact sequence of algebraic groups (over $\mathbb{C}$)

$$1 \to \{\pm 1\} \to \mathrm{B}^{(1)} \to \mathrm{PB}^\times \to 1. \tag{1.13}$$

In fact relative to the identification $\mathrm{PB}^\times \simeq \mathrm{SO}_{\mathrm{B}^0} \simeq \mathrm{SO}_3$, $\mathrm{B}^{(1)}$ gets naturally identified with the spin group (the simply connected 2-covering group of the special orthogonal group)

$$\mathrm{B}^{(1)} \simeq \mathrm{Spin}_{\mathrm{B}^0} \simeq \mathrm{Spin}_3.$$

1.5.3. *The Hopf fibration.* The previous section is especially useful to discuss the *Hopf filtration*: over the reals the sequence (1.13) remains exact:

$$1 \to \{\pm 1\} \to \mathrm{B}^{(1)}(\mathbb{R}) \to \mathrm{PB}^\times(\mathbb{R}) \to 1.$$

By Witt's theorem $\mathrm{B}^{(1)}(\mathbb{R})$ acts transitively on $\mathrm{B}^{0,(1)}(\mathbb{R})$ the set of vectors of norm 1 (by conjugation) making it a principal homogeneous space: the map

$$\begin{array}{ccc} \mathrm{B}^{(1)}(\mathbb{R}) & \mapsto & \mathrm{B}^{0,(1)}(\mathbb{R}) \\ z & \mapsto & \rho_{z,z}(k) = zkz^{-1} \end{array} \tag{1.14}$$

is surjective, its fiber above any $w \in \mathrm{B}^{0,(1)}(\mathbb{R})$ is the stabilizer of w in $\mathrm{B}^{(1)}(\mathbb{R})$,

$$\mathrm{B}^{(1)}(\mathbb{R})_w = \mathrm{Stab}_{\mathrm{B}^{(1)}(\mathbb{R})}(w) = \{a + d.w,\ a, d \in \mathbb{R},\ a^2 + d^2 = 1\} \simeq S^1$$

and yields to the homeomorphism

$$\mathrm{B}^{0,(1)}(\mathbb{R}) \simeq \mathrm{B}^{(1)}(\mathbb{R})/\mathrm{B}^{(1)}(\mathbb{R})_k;$$

(indeed the set of quaternions z commuting with w (such that $zw = wz$) is the set of quaternions of the shape $a + d.w,\ a, d \in \mathbb{R}$.)

Now, via (1.8), $\mathrm{B}^{(1)}(\mathbb{R})$ is identified with the 3-sphere

$$S^3 = \{(a, b, c, d) \in \mathbb{R}^4,\ a^2 + b^2 + c^2 + d^2 = 1\},$$

$\mathrm{B}^{0,(1)}(\mathbb{R})$ with the 2-sphere

$$S^2 = \{(0, b, c, d) \in \mathbb{R}^4,\ b^2 + c^2 + d^2 = 1\},$$

and from the above discusion we obtain a realization of S^3 as a (topological) S^1-bundle over the 2-sphere

$$S^1 \hookrightarrow S^3 \twoheadrightarrow S^2. \tag{1.15}$$

This is the so-called *Hopf fibration*.

2. The number of sums of four squares representations

After having established the existence of representations of and integer as a sum of four squares, the next step is: *how many representations are there ?*

Thirty years after Lagrange, Jacobi, using theta series, gave a closed formula of the number of such representations:

THEOREM (Jacobi, 1828). *For $n \geq 1$,*

$$|\mathrm{R}_4(n)| = |\mathrm{B}^{(n)}(\mathbb{Z})| = 8 \sum_{\substack{d|n \\ 4\nmid d}} d = 8(1+2\delta_{2|n}) \prod_{p^\alpha \| n, p\neq 2} \frac{p^{\alpha+1}-1}{p-1}.$$

The factor 8 is the order of the group of units of the ring $\mathrm{B}(\mathbb{Z})$

$$\mathrm{B}^\times(\mathbb{Z}) = \{\pm 1, \pm i, \pm j, \pm k\}$$

which acts faithfully on $\mathrm{B}^{(n)}(\mathbb{Z})$ by left multiplication. So Jacobi's formula could be rewritten

$$|\mathrm{B}^\times(\mathbb{Z})\backslash \mathrm{B}^{(n)}(\mathbb{Z})| = \sum_{\substack{d|n \\ 4\nmid d}} d = (1+2\delta_{2|n}) \prod_{p^\alpha || n, p\neq 2} \frac{p^{\alpha+1}-1}{p-1}. \tag{2.1}$$

It follows from these formulas and from the estimate for the number of divisors

$$d(n) = \sum_{d|n} 1 = n^{o(1)}$$

that, as $n \to +\infty$, $4 \nmid n$

$$|\mathrm{R}_4(n)| = n^{1+o(1)}. \tag{2.2}$$

We will not discuss the proof here; for now let us merely say that the main ingredient of Jacobi's proof is the introduction of an analytic object: Jacobi's *theta* function, defined for $z \in \mathbb{C}$, $\Im(z) > 0$ by

$$\theta_4(z) = \sum_{(a,b,c,d)\in\mathbb{Z}^4} \exp(2\pi i(a^2+b^2+c^2+d^2)z) = \sum_{n\geq 0} |\mathrm{R}_4(n)| \exp(2\pi i n z).$$

Using the Poisson summation formulation, one shows that $\theta_4(z)$ is an holomorphic modular form of weight 2 relative to the subgroup

$$\Gamma_0(4) = \{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}),\ c \equiv 0(4)\};$$

that is for $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(4)$

$$\theta_4(\frac{az+b}{cz+d}) = (cz+d)^2 \theta_4(z),$$

and in fact is an *Eisenstein series*. Using this information one can then obtain Jacobi's formula (cf. [**Iwa97**, **Sar91**]).

3. The distribution of four squares representations

As we have seen, as $n \to +\infty$, there are "more and more" representation of n as a sum of four squares. Our next question then is

How are these representations distributed as $n \to \infty$?

One of the simplest way to investigate this question is the following: any representation (a, b, c, d) is a vector in $\mathbb{R}^4$ of euclidean norm $n^{1/2}$. Projecting such points radially to the 3-sphere we obtain a sequence of finite (essentially growing) sets of points

$$n^{-1/2}.\mathrm{R}_4(n) \subset S^3.$$

3.1. Equidistribution of points on the 3-sphere. As it turn out, these points become distributed in the most natural possible way: let μ_{S^3} denote the "Lebesgue", rotationally invariant, probability measure on the 3-sphere

THEOREM (Malyshev). *As $n \to +\infty$ $n \not\equiv 0(4)$, the set $n^{-1/2}.\mathrm{R}_4(n)$ becomes equidistributed on S^3 with respect to μ_{S_3}: for any $\varphi \in \mathcal{C}(S^3)$,*

$$\frac{1}{|\mathrm{R}_4(n)|} \sum_{\mathbf{x} \in \mathrm{R}_4(n)} \varphi(\frac{\mathbf{x}}{n^{1/2}}) \to \int_{S^3} \varphi(x) d\mu_{S^3}(x)$$

Below we give several equivalent formulations of Malyshev's equidistribution theorem.

3.2. Equidistribution of quaternions. As we have seen $\mathbb{R}^4$ is isometric to $\mathrm{B}(\mathbb{R})$, and under this isometry, $R_4(n)$ is identified with the of integral quaternions of norm n, $B^{(n)}(\mathbb{Z})$, and the 3-sphere S^3 with the group real quaternion of norm one $\mathrm{B}^{(1)}(\mathbb{R})$. The later a compact group and under this identification, the Lebesgue measure on S^3 correspond the Haar probability measure $\mu_{\mathrm{B}^{(1)}(\mathbb{R})}$. Therefore an equivalent formulation of Malyshev's theorem is: *for any $\varphi \in \mathcal{C}(\mathrm{B}^1(\mathbb{R}))$,*

$$\frac{1}{|\mathrm{B}^{(n)}(\mathbb{Z})|} \sum_{z \in \mathrm{B}^{(n)}(\mathbb{Z})} \varphi(\frac{z}{n^{1/2}}) \to \int_{\mathrm{B}^1(\mathbb{R})} \varphi(z) d\mu_{\mathrm{B}^1}(z), \ n \to +\infty. \tag{3.1}$$

3.3. Equidistribution of rotations. The Hopf map discussed in §1.5.3 makes it possible to represent graphically this equidistribution property: to each non-zero quaternion z is associated a rotation $r_z \in \mathrm{SO}_3(\mathbb{R})$, namely the rotation corresponding to the rotation $\rho_{z,z} \in \mathrm{SO}_{\mathrm{B}^0}(\mathbb{R})$ via the isometry $\mathrm{B}^0(\mathbb{R}) \simeq \mathbb{R}e_2 + \mathbb{R}e_3 + \mathbb{R}e_4 \simeq \mathbb{R}^3$. This yields an homeomorphism

$$\mathrm{B}^{(1)}(\mathbb{R})/\{\pm 1\} \simeq \mathrm{SO}_3(\mathbb{R})$$

and the Haar probability measure on $\mathrm{SO}_3(\mathbb{R})$ correspond to the Haar probability measure on $\mathrm{B}^{(1)}(\mathbb{R})$ (restricted to the continuous functions invariant under multiplication by ± 1). Therefore we obtain

THEOREM 3.1. *As $n \to +\infty$ $n \not\equiv 0(4)$, the set of rotations*

$$\{r_{z/n^{1/2}},\ z \in \mathrm{B}^{(n)}(\mathbb{Z})\} \subset \mathrm{SO}_3(\mathbb{R})$$

becomes equidistributed on $\mathrm{SO}_3(\mathbb{R})$ with respect to the Haar probability measure: for any $f \in \mathcal{C}(\mathrm{SO}_3(\mathbb{R}))$,

$$\frac{1}{|\mathrm{R}_4(n)|} \sum_{z \in \mathrm{R}_4(n)} f(r_{z/n^{1/2}}) \to \int_{\mathrm{SO}_3(\mathbb{R})} f(\rho) d\mu_{\mathrm{SO}_3}(\rho).$$

3.3.1. *Equidistribution on the 2-sphere.* By Witt's theorem the action of $\mathrm{SO}_3(\mathbb{R})$ on S^2 is transitive; so any $v_0 \in S^2$, the map

$$r \in \mathrm{SO}_3(\mathbb{R}) \mapsto r(v_0)$$

induces and homeomorphism

$$\mathrm{SO}_3(\mathbb{R})/\,\mathrm{SO}_3(\mathbb{R})_{v_0} \simeq S^2.$$

Under this map, the rotation invariant Lebesgue probability measure μ_{S^2} on the 2-sphere corresponds to the Haar probability measure on $\mathrm{SO}_3(\mathbb{R})$ (restricted to the subspace of continuous function on $\mathrm{SO}_3(\mathbb{R})$ which are $\mathrm{SO}_3(\mathbb{R})_{v_0}$. Therefore we deduce

COROLLARY 3.1. *Given any $v_0 \in S^2$, as $n \to +\infty$ $n \not\equiv 0(4)$, the multiset $\{r_{z/n^{1/2}}(v_0),\ z \in \mathrm{B}^{(n)}(\mathbb{Z})\}$ becomes equidistributed on S^2 w.r.t. μ_{S^2}: for any $\varphi \in \mathcal{C}(S^2)$,*

$$\frac{1}{|\mathrm{B}^{(n)}|} \sum_{z \in \mathrm{B}^{(n)}} \varphi(\rho_{z/n^{1/2}}(v_0)) \to \int_{S^2} \varphi(v) d\mu_{S^2}(v).$$

In fact this Corollary (the equidistribution of the translates of v_0 by the ρ_z *for any* v_0 is equivalent to Theorem 3.1.

3.4. Sketch of a proof. Malyshev's theorem is to be proven in the course of this book in a much more general form but let us sketch the principle of the proof as it is exposed in [**Iwa97**, Chap. 11].

It is known ([**Far08**]) that any continuous function on S^3 may be approximated uniformly on S^3 by a linear combination of *harmonic homogeneous polynomials*: let us recall that an *harmonic homogeneous polynomial*, is the restriction to S^3 of a polynomial ϕ_ν ($\nu \in \mathbb{N}$ an integer) on $\mathbb{R}^4$ which is homogeneous of degree ν, that is such that for $\lambda \in \mathbb{R}$

$$\phi_\nu(\lambda \mathbf{x}) = \lambda^\nu \phi_\nu(\mathbf{x}), \text{ and such that } \Delta_{\mathbb{R}^4}\phi_\nu = 0$$

were

$$\Delta_{\mathbb{R}^4} = \frac{\partial^2}{\partial^2 x} + \frac{\partial^2}{\partial^2 y} + \frac{\partial^2}{\partial^2 z} + \frac{\partial^2}{\partial^2 t}$$

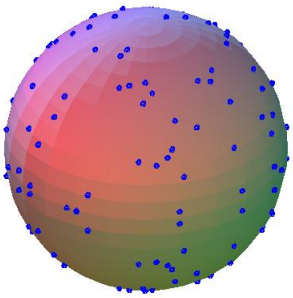

FIGURE 1. The distribution of $\{\rho_z(v_0),\ z \in \mathrm{R}_4(5^3)\}$, $v_0 = (1,0,0)$.

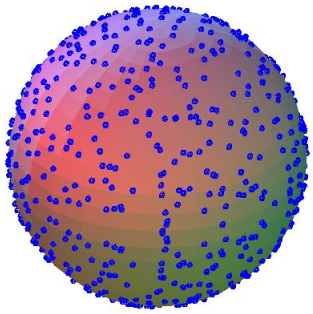

FIGURE 2. The distribution of $\{\rho_z(v_0),\ z \in \mathrm{R}_4(5^4)\}$, $v_0 = (1,0,0)$.

is the Laplace operators on $\mathbb{R}^4$. This is a consequence of the *Peter-Weyl* theorem[3] for the compact group $\mathrm{SO}_4(\mathbb{R})$. Explicitely, one can show that the space of harmonic homogeneous polynomial of degree ν is generated by

[3]A continuous function on a compact group may be approximated uniformly by linear combination of the matrix coefficients of irreductible continuous representations the group

polynomials of the shape

$$\phi_\nu(x,y,z,t) = ax + by + ct + dt$$

for $(a,b,c,d) \in \mathbb{C}^4$, varying over the set of q_4 *isotropic* vectors: ie. such that

$$q_4(a,b,c,d) = a^2 + b^2 + c^2 + d^2 = 0.$$

In view of this, it is sufficient (by Weyl's equidistribution criterion) to show that for any such polynomial

$$W(\phi_\nu; n) = \frac{1}{|\mathrm{R}_4(n)|} \sum_{\mathbf{x} \in \mathrm{R}_4(n)} \phi_\nu(\frac{\mathbf{x}}{|n|^{1/2}}) \to \int_{S^3} \phi_\nu(x) d\mu_{S^3}(x) \tag{3.2}$$

The quantities $W(\phi_\nu; n)$ are called the *Weyl sum* associated with this equidistribution problem.

– If $\nu = 0$, ϕ_ν is a constant polynomial and

$$W(\phi_\nu; n) = \phi_\nu = \int_{S^3} \phi_\nu(x) d\mu_{S^3}(x).$$

– If $\nu > 0$, then $\int_{S^3} \phi_\nu(x) d\mu_{S^3}(x) = 0$ (ϕ_ν is orthogonal to the constants) and we have to show that $W(\phi_\nu; n) \to 0$.

– If ν is odd, then since $\mathrm{R}_4(n) = -\mathrm{R}_4(n)$,

$$W(\phi_\nu; n) = (-1)^\nu W(\phi_\nu; n) = 0$$

and we are done.

– Let us now suppose that ν is even. We consider, for

$$z \in \mathbb{H} = \{z \in \mathbb{C},\ \Im(z) > 0\}$$

(the upper half-plane), the theta function

$$\begin{aligned}\theta(\phi_\nu, z) &= \sum_{(a,b,c,d) \in \mathbb{Z}^4} \phi_\nu(a,b,c,d) \exp(2\pi i(a^2+b^2+c^2+d^2)z) \\ &= \sum_{n \geq 1} \Big(\sum_{\mathbf{x} \in \mathrm{R}_4(n)} \phi_\nu(\mathbf{x}) \Big) \exp(2\pi i n z) \\ &= \sum_{n \geq 1} |R_4(n)| W(\phi_\nu; n) n^{\nu/2} \exp(2\pi i n z).\end{aligned}$$

Using the Poisson summation formula, one can show that $\theta(\phi_\nu, z)$ is a holomorphic modular cusp form of weight $k = 2 + \nu$ with respect to the congruence subgroup

$$\Gamma_0(4) = \{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}),\ c \equiv 0(4) \}.$$

In other terms, $\theta(\phi_\nu, z)$ this is an holomorphic function on $\mathbb{H}$, such that for any $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(4)$,

$$\theta(\phi_\nu, \gamma z) = \theta(\phi_\nu, \frac{az+b}{cz+d}) = (cz+d)^{2+\nu} \theta(\phi_\nu, z)$$

such that the $\Gamma_0(4)$ invariant function $z \mapsto (\Im z)^{k/2}|\theta(\phi_\nu, z)|$ is bounded on $\mathbb{H}$.

The sums

$$\sum_{\mathbf{x}\in \mathrm{R}_4(n)} \phi_\nu(\mathbf{x}) = \int_0^1 \theta(\phi_\nu, x+iy)e(-2\pi i n(x+iy))dx$$

are then the *Fourier coefficients* of this holomorphic cusp form; by the Petersson formula, the square of their modulus are bounded by

$$n^{-(k-1)}|\sum_{\mathbf{x}\in \mathrm{R}_4(n)} \phi_\nu(\mathbf{x})|^2 \ll_{\phi_\nu} 1 + \sum_{c\equiv 0(4)} \frac{|S(n,n;c)|}{c}|J_{k-1}(\frac{4\pi n}{c})|$$

where $J_{k-1}(x)$ is the Bessel function of order $k-1$ and

$$S(n,n;c) = \sum_{\substack{x(\mathrm{mod}\ c)\\(x,c)=1}} \exp(2\pi i \frac{nx+nx^{-1}}{c})$$

is a *Kloosterman* sum. Such sums were bounded non-trivially for the first time by Kloosterman (for c prime which is the fundamental case) and Salié (for general c): there is an absolute constant $\delta > 0$ such that

$$|S(n,n;c)| \leq c^{1-\delta+o(1)}(n,c)^\delta. \tag{3.3}$$

It follows from such a bound that (see [**Iwa97**, Chap. 11])

$$\sum_{\mathbf{x}\in \mathrm{R}_4(n)} \phi_\nu(\mathbf{x}) \ll_{\phi_\nu} n^{\frac{\nu}{2}+1-\frac{\delta}{2}},$$

hence (using the estimate $|\mathrm{R}_4(n)| = n^{1+o(1)}$) we obtain

$$W(\phi_\nu; n) \ll_{\phi_\nu} n^{-\delta/2+o(1)} \to 0,\ n \to +\infty$$

which conclude the proof of equidistribution. □

Remark 3.1. Kloosterman proved the bound (3.3) for $\delta = 1/3$ and Weil using his proof of the Riemann hypothesis for curves over finite fields improved the bound to $\delta = 1/2$; 30 years later, Deligne [`?DelBour`, `?WeilI`] proved the Weil conjectures and as a consequence of it the Ramanujan-Petterson conjecture on the size of Fourier coefficients of holomorphic modular cuspforms of integral weight. This implies the following bound

$$W(\phi_\nu; n) \ll_{\phi_\nu} n^{-1/2+o(1)}.$$

This bound is up to the term $n^{o(1)}$ optimal; moreover since $|R_4(n)| = n^{1+o(1)}$ the Weyl sum is essentially bounded by the inverse of squareroot of the number of terms occuring the sum; this is a manifestation of the *square-root cancellation* phenomenon.

4. Sums of three squares

We consider now the question of the representability of a given integer n as a sum of *three squares*: the existence of solutions $(a, b, c) \in \mathbb{Z}^3$ to the equation

$$n = a^2 + b^2 + c^2. \tag{4.1}$$

This question is significantly harder than for than for the sums of four squares. Its solution came 30 years after the proof Lagrange's theorem, when Legendre gave a classification integers which are sums of three squares; Legendre's argument however was not complete[4] and a different unconditional proof was given by Gauss

THEOREM (Legendre, 1798; Gauss 1801). *Every non-negative integer n not of the form $4^k(8l-1)$, $k, l \in \mathbb{N}$ may be represented as a sum of three squares*

$$n = a^2 + b^2 + c^2,\ a, b, c \in \mathbb{Z}.$$

4.1. An idea of the proof of the Gauss-Legendre Theorem. We will not give the complete argument here but at least we discuss the arithmetical part of the proof using the Hamilton quaternions.

Let us make first a trivial reduction: let $a^2+b^2+c^2 = n$ be a representation of n as a sum of three squares; if $4|n$ then $a^2+b^2+c^2 \equiv 0 (\mathrm{mod}\ 4)$ and by inspecting the sums of three squares in $\mathbb{Z}/4\mathbb{Z}$, we see that 2 must divide a, b and c so $(a/2, b/2, c/2)$ is a representation of $n/4$; in other terms, if $n = 4^k n_0$ with $4 \nmid n_0$ then the representations of n are of the form $2^k(a_0, b_0, c_0)$ for (a_0, b_0, c_0) ranging over the representations of n_0. Considering these, we see that since 7 is not a sum of three squares modulo 8, n_0 is not representable as a sum of three squares unless $n \not\equiv 7 (\mathrm{mod}\ 8)$. This proves the necessity of the condition $n \neq 4^k(8l-1)$.

The hard part of the Gauss-Legendre Theorem is that this condition is sufficient. The starting and key point is the fact that, under that condition, n admit at least a *rational* representation: there is $(a, b, c) \in \mathbb{Q}^3$ such that $n = a^2 + b^2 + c^2$. This is special case of the *Hasse-Minkowski principle* (see Chap. 3 below for a discussion and [**Ser73**][Thm. 8, p. 41] for a proof.)

From this, one can easily conclude the proof: we denote by q_3 the "three square" quadratic form (the euclidean quadratic form on $\mathbb{Q}^3$)

$$q_3(a, b, c) = a^2 + b^2 + c^2.$$

As noted before, the euclidian 3 space is isometric to the space of pure Hamilton quaternions

$$(\mathbb{Q}^3, q_3) \simeq (\mathrm{B}^0(\mathbb{Q}), \mathrm{Nr}).$$

[4]Legendre's proof assumed the inifiniteness of the set primes in a given arithmetic progression which was only proven about 50 years later by Dirichlet

The problem of representing n as a sum of three squares is equivalent to the problem of finding a pure integral quaternion of that norm. Consider the pure (rational) quaternion

$$z = ai + bj + ck \in \mathrm{B}^{(0)}(\mathbb{Q}).$$

It has norm n, and so, by (1.6), satisfies the quadratic equation

$$z^2 + n = 0.$$

The commutative ring $\mathbb{Z}[z] \subset \mathscr{O}_{\mathrm{B}}$ is isomorphic to the quadratic ring $\mathbb{Z}[\sqrt{-n}]$ hence is finitely generated and by the Corollary 1.1 there is $q \in \mathrm{B}^{\times}(\mathbb{Q})$ such that $z' = q^{-1}zq = \rho_q(z) \in \mathscr{O}_{\mathrm{B}}$. Since $\mathrm{Nr}(z') = \mathrm{Nr}(z) = n$, $\mathrm{tr}(z') = \mathrm{tr}(z) = 0$ and $\mathscr{O}_{\mathrm{B}} \cap \mathrm{B}^{(0)}(\mathbb{Q}) = \mathrm{B}^{(0)}(\mathbb{Z})$, $z' = a'i + b'j + c'k$ has integral entries and (a', b', c') is a representation on n. □

5. The number of three squares representations

Let

$$\mathrm{R}_3(n) = \{(a, b, c) \in \mathbb{Z}^4,\ a^2 + b^2 + c^2 = n\}$$

be the set of integral representations of n by q_3. We know precisely when $\mathrm{R}_3(n)$ is non empty and, we will discuss now its finer structure and properties as n varies. The next question coming is the size of that set. On that matter we have the following

THEOREM 5.1. *For $n \not\equiv 0, 4, 7 (\mathrm{mod}\ 8)$ one has, as $n \to +\infty$*

$$|\mathrm{R}_3(n)| = n^{1/2+o(1)}.$$

REMARK. From the previous discussion, we may and will assume that $4 \nmid n$.

The proof of this theorem will be discussed in the next sections. Again its proof is significantly harder than the corresponding question for $\mathrm{R}_4(n)$ and builds on three principal ingredients:

The first ingredient is due to Gauss: the set of representations $\mathrm{R}_3(n)$ is closely related to another arithmetic set which we will discuss again in the next chapter: set $d = -n$ if $n \equiv 3 \,(\mathrm{mod}\, 8)$ or $-4d$ if $n \equiv 1, 2, 5, 6 \,(\mathrm{mod}\, 8)$, and let

$$R_{\mathrm{disc}}(d) = \{(a', b', c') \in \mathbb{Z}^3,\ (b')^2 - 4a'c' = d\},$$

be the set of integral representations of d by the *discriminant* quadratic form

$$\mathrm{disc}(a', b', c') = (b')^2 - 4a'c'.$$

To describe the relationship more precisely we need the notion of primitive representation.

5.0.1. *Relation with the class number.* Let (a_0, b_0, c_0) be an integral representation of some integer n_0 by q_3, then for any integer $f \neq 0$, $f.(a_0, b_0, c_0) = (fa_0, fb_0, fc_0)$ is a representation of $n = n_0 f^2$. A representation (a, b, c) of n is *primitive* if it does not come from a representation of some proper divisor $n_0 = n/f^2$, or in other terms if and only if a, b, c are *coprime*[5]. The set of primitive representations is noted

$$\mathrm{R}_3^*(n) = \{(a, b, c) \in \mathbb{Z}^3,\ a^2 + b^2 + c^2 = n,\ a, b, c \text{ coprime}\}.$$

One has

$$\mathrm{R}_3(n) = \bigsqcup_{f^2|n} f.\mathrm{R}_3^*(n/f^2). \tag{5.1}$$

Similarly, one consider the set of primitive representations of d by the discriminant

$$\mathrm{R}_{\mathrm{disc}}^*(d) = \{(a', b', c') \in \mathbb{Z}^3,\ (b')^2 - 4a'c' = d,\ a', b', c' \text{ coprimes}\}.$$

It is easy to see that the set $\mathrm{R}_{\mathrm{disc}}^*(d)$ is non-empty if and only $d \equiv 0, 1 \pmod 4$ (search a representation with $a' = 1$ or $b' = 1$) and in this case $\mathrm{R}_{\mathrm{disc}}^*(d)$. The reason is that $\mathrm{R}_{\mathrm{disc}}^*(d)$ is acted by the infinite group $\mathrm{SL}_2(\mathbb{Z})$ and the stabilizers of the points $\mathrm{R}_{\mathrm{disc}}^*(d)$ in are finite (since $d < 0$). This action is defined by noting that $\mathrm{R}_{\mathrm{disc}}^*(d)$ correspond bijectively with the set of *integral, primitive, binary* quadratic forms of discriminant d via the map

$$(a', b', c') \to q_{a',b',c'}(X, Y) = a'X^2 + b'XY + c'Y^2,$$

and $\mathrm{SL}_2(\mathbb{Z})$ acts on the set of such quadratic forms via the linear transformation in $\mathbb{Z}^2$: for $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$

$$(X, Y).\begin{pmatrix} a & b \\ c & d \end{pmatrix} = (aX + cY, bX + dY).$$

A fundamental result of Gauss is

THEOREM 5.2 (Gauss). *The set of orbits $\mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{R}_{\mathrm{disc}}^*(d)$ is finite. Its cardinality is called the class number of the discriminant d and is noted*

$$h(d) = |\mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{R}_{\mathrm{disc}}^*(d)|.$$

The set $\mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{R}_{\mathrm{disc}}^*(d)$ has the structure of an abelian group which nowaday is expressed as *the ideal class group of the imaginary quadratic order of discriminant d* and Gauss proved that this group acts on transitively the quotient $\mathrm{SO}_3(\mathbb{Z})\backslash\mathrm{R}^*(n)$ and computed the size of the stabilizers for this action. From this, it follows that

THEOREM (Gauss). *For $n \geq 1$, $n \not\equiv 0, 4, 7 \pmod 8$. One has*

$$|\mathrm{R}_3^*(n)| = (48/w)h(d), \quad \text{with } d = -n \quad \text{if } n \equiv 3(8),$$
$$|\mathrm{R}_3^*(n)| = (24/w)h(d), \quad \text{with } d = -4n \quad \text{if } n \equiv 1, 2, 5, 6(4);$$

here $w = 6$ for $d = -3$, $w = 4$ for $d = -4$, $w = 2$ for $d < -4$.

[5] one also says that (a, b, c) is a primitive vector of $\mathbb{Z}^3$

REMARK. If $n = 1, 2$ or 3 , the above theorem remains valid with, the 12 replaced by 6, the 12 remains unchanged and the 24 replaced by 8 respectively.

5.0.2. *The class number formula.* The second ingredient is the *Dirichlet's class number formula* which express the order of the ideal class group of imaginary quadratic orders in terms of L-functions: for d as above, let

$$\chi_d(.) = (\frac{d}{.}) : (\mathbb{Z}/d\mathbb{Z})^\times \mapsto \{\pm 1\}$$

be the Kronecker symbol of d. This is a (not-necessarily primitive) real character of modulus d and conductor $d_K = d/f^2$ the unique fundamental discriminant in the square class of d (ie. the discriminant of the quadratic field $K = \mathbb{Q}(\sqrt{d})$). Let

$$L(\chi_d, s) = \sum_{n\geq 1} (\frac{d}{n})\frac{1}{n^s} = \prod_p (1 - \frac{(d/p)}{p^s})^{-1}$$

be the Dirichlet L-function associated with χ_d. One has

THEOREM (Dirichlet). *The class number $h(d)$ has the following expression*

$$h(d) = \frac{w}{2}\frac{L(\chi_d, 1)}{\pi}|d|^{1/2}.$$

Combining Gauss and Dirichlet formulae we obtain

$$|R_3^*(n)| = 12\frac{L(\chi_d, 1)}{\pi}n^{1/2}.$$

REMARK. We are grateful to Gergely Harcos for pointing out that the formula above is valid for any n not divisible by 4 and that it is not restricted to squarefree n.

5.0.3. *Siegel's theorem.* By Dirichlet's formula (and the fact that a group has order at least 1) $L(\chi_d, 1) \gg 1$. Landau and subsequently Siegel showed that this lower bound is far from optimal: (see [**Dav00**, Chap. 21]):

THEOREM (Siegel). *As $d \to \infty$*

$$L(\chi_d, 1) = |d|^{o(1)}.$$

Combining these three results we obtain

$$|R_3^*(n)| = n^{1/2+o(1)}$$

and using (5.1) we obtain Theorem (5.1).

6. Sums of three squares and quadratic fields

In this section, we discuss the arguments leading to the proof of Gauss Theorem above, in the language of quaternion; this lead us naturally to consider ideal classes of imaginary quadratic order instead classes of binary quadratic forms.

6.1. Embedding of quadratic field. As pointed out before, the map

$$(a,b,c)\mapsto z=ai+bj+ck$$

identifies the quadratic space $(\mathbb{Q}^3,q_3)$ with $(\mathrm{B}^0(\mathbb{Q}),\mathrm{Nr})$; under that map $\mathrm{R}_3(n)$ is identified with the pure integral quaternions of reduced norm n, $\mathrm{B}^{0,(n)}(\mathbb{Z})=\mathscr{O}_{\mathrm{B}}^{0,n}$, $\mathrm{R}_3^*(n)$ with the *primitive* ones noted $\mathrm{B}^{0,(n)}(\mathbb{Z})^*=\mathscr{O}_{\mathrm{B}}^{0,(n)*}$. In the sequel we will use the notations $\mathrm{R}_3(n)$, $\mathrm{R}_3^*(n)$, $\mathrm{B}^{0,(n)}(\mathbb{Z})$, $\mathscr{O}_{\mathrm{B}}^{0,n}$ $\mathrm{B}^{0,(n)}(\mathbb{Z})$, $\mathscr{O}_{\mathrm{B}}^{0,(n)*}$ etc... interchangeably.

Given $z=ai+bj+ck\in\mathrm{B}^{(0)}(\mathbb{Q})$ a pure quaternion of norm n, z satisfies the equation

$$z^2+n=0$$

The polynomial X^2+n is irreducible and therefore z yield an embedding ι_z of the imaginary quadratic field $K=\mathbb{Q}(\sqrt{-n})$ into $\mathrm{B}(\mathbb{Q})$

$$\iota_z:u+v\sqrt{-n}\mapsto u+vz.$$

To such an embedding, one associate an order in K in the following way

Definition 6.1. *Let $z=ai+bj+ck\in\mathrm{B}^{(0)}(\mathbb{Q})$ be a non-zero pure quaternion. The order associated with z is the order*

$$\mathscr{O}_z=\mathbb{Q}[z]\cap\mathscr{O}_{\mathrm{B}}.$$

We would like now to determine the *preimage* of $\mathscr{O}_z$ in K

$$\iota_z^{-1}(\mathscr{O}_z)\subset K.$$

Obviously the order $\mathscr{O}_z$ depends on z only up to multiplication by scalars: for any $\lambda\in\mathbb{Q}^\times$

$$\mathscr{O}_z=\mathscr{O}_{\lambda z},$$

we may assume that z is an integral primitive pure quaternions (that is z has integral coprime coordinates). If is remarkable that after such a reduction, the preimage in K depends only on the norm of z, n:

Proposition 6.1. *Let $z=ai+bj+ck$ be a pure integral primitive quaternion of norm n. We have*

$$\iota_z^{-1}(\mathscr{O}_z)=\begin{cases}\mathbb{Z}[\sqrt{-n}] & \text{if } -n\equiv 2,3(4)\\ \mathbb{Z}[\frac{1+\sqrt{-n}}{2}] & \text{if } -n\equiv 1(4)\end{cases}.$$

In other terms,

$$\iota_z^{-1}(\mathscr{O}_z)=\mathscr{O}_d=\mathbb{Z}[\frac{d+\sqrt{d}}{2}]$$

is the quadratic order of discriminant $d=d(n)$ where $d=-n$ if $-n\equiv 1(\mathrm{mod}\ 4)$ and $d=-4n$ if $-n\equiv 2,3(\mathrm{mod}\ 4)$.

Proof. Let $u+\sqrt{-n}v\in\iota_z^{-1}(\mathscr{O}_z)$, $u,v\in\mathbb{Q}$, $\iota_z(u+\sqrt{-n}v)=u+vai+vbj+vck\in\mathscr{O}_{\mathrm{B}}$ thus either $(u,va,vb,vc)\in\mathbb{Z}^4$ which impose $u,v\in\mathbb{Z}$ since a,b,c are coprime or $(u,va,vb,vc)\in(\frac{1}{2}+\mathbb{Z})^4$ which can occur if and only

if $a, b, c \equiv 1(2)$ which is equivalent to $a^2 + b^2 + c^2 \equiv 3(4)$ and which impose that $u, v \in \frac{1}{2} + \mathbb{Z}$. Evidently if $a, b, c \equiv 1(2)$, then $\frac{1+z}{2} \in \mathscr{O}_{\mathrm{B}}$. □

Definition 6.2. *Given K a quadratic field, $\mathscr{O} \subset K$ an order and $\iota : K \mapsto \mathrm{B}$ an embedding of $\mathbb{Q}$-algebras. ι is an* optimal *embedding of $\mathscr{O}$ into $\mathscr{O}_{\mathrm{B}}$ if $\iota(\mathscr{O}) \subset \mathscr{O}_{\mathrm{B}}$ and $\iota^{-1}(\mathscr{O}_{\mathrm{B}}) = \mathscr{O}$ is maximal for this property.*

In view of this definition, ι_z defines an *optimal* embedding of $\mathscr{O}_d$ into $\mathscr{O}_{\mathrm{B}}$.

6.2. The ideal class group of a quadratic order. Let $\mathscr{O} \subset K$ be an order in a quadratic field, an *ideal* $I \subset K$ is an $\mathscr{O}$-module which is of rank 2 as a $\mathbb{Z}$-module. An $\mathscr{O}$-ideal is *proper* if its stabilizer in K is precisely $\mathscr{O}$ (and not a bigger order)

$$\mathscr{O}_I = \{z \in K,\ zI \subset I\} = \mathscr{O}.$$

Let $\mathrm{Prop}(\mathscr{O})$ be the set of propers $\mathscr{O}$-ideals: the group $K^\times$ acts on the set of propers $\mathscr{O}$-ideals by multiplication

$$(\lambda, I) \in K^\times \times \mathrm{Prop}(\mathscr{O}) \to \lambda.I \in \mathrm{Prop}(\mathscr{O})$$

and the set of proper ideal classes is the set of orbits under this action, it is noted

$$\mathrm{Pic}(\mathscr{O}) = \mathrm{Prop}(\mathscr{O})/K^\times$$

and is finite.

Moreover $\mathrm{Prop}(\mathscr{O})$ has the structure of commutative group for the multiplication of ideals

$$I, J \to I.J = \mathbb{Z}\text{-module generated by the products } xy,\ x \in I, y \in J;$$

this gives $\mathrm{Pic}(\mathscr{O})$ the structure of a finite commutative group: this is the *Picard group* of $\mathscr{O}$.

7. The action of the class group on sums of three squares

Let $z \in \mathrm{R}_3^*(n)$ be a primitive pure quaternion of norm n and let $\mathscr{O}_z$ be its corresponding order and $\mathscr{O} = \iota_z^{-1}(\mathscr{O}_z) \subset K$ by the corresponding "abstract order". The Theorem 5.0.1 of Gauss is a consequence of the existence of a transitive action of the Picard group $\mathrm{Pic}(\mathscr{O})$ on a quotient of $\mathrm{R}_3^*(n)$. In this section we will describe this action.

We have seen that $\mathrm{B}^\times(\mathbb{Q})$ act isometrically on $\mathrm{B}^{(0)}(\mathbb{Q})$ by conjugation and that the group of projective units $\mathrm{PB}^\times(\mathbb{Q}) = \mathrm{B}^\times(\mathbb{Q})/\mathbb{Q}^\times$ is isomorphic to the orthogonal group $\mathrm{SO}_{\mathrm{B}^{(0)}}(\mathbb{Q})$; one can also show thst the stabilizer of the lattice $\mathrm{B}^{(0)}(\mathbb{Z}) \simeq \mathbb{Z}^3$, $\mathrm{SO}_{\mathrm{B}^{(0)}}(\mathbb{Z}) \simeq \mathrm{PB}^\times(\mathbb{Z})$ say, is of order 24 and equals to the image of the group generated by $\mathscr{O}_{\mathrm{B}}^\times$ and the image of any integral quaternion of norm 2. Clearly $\mathrm{PB}^\times(\mathbb{Z})$ acts on $\mathrm{R}_3^*(n)$. Denote the quotient by

$$\widetilde{\mathrm{R}_3}(n) := \mathrm{PB}^\times(\mathbb{Z})\backslash\mathrm{R}_3^*(n).$$

The generalization of Theorem 5.0.1 is

THEOREM 7.1. *The quotient* $\widetilde{\mathrm{R}_3}(n)$ *is endowed with a transitive action of the Picard group* $\mathrm{Pic}(\mathscr{O})$ *and the stabilizer of an element under this action has order at most* 2. *Consequently*

$$\frac{1}{2}|\mathrm{Pic}(\mathscr{O})| \leq |\mathrm{R}_3^*(n)| \leq 24|\mathrm{Pic}(\mathscr{O})|$$

PROOF. We will first show how, from an integral representation z of n and a proper[6] $\mathscr{O}$-ideal $I \subset K$, one can obtain a new representation of n. The $\mathbb{Z}$-module $\iota_z(I).\mathscr{O}_{\mathrm{B}}$ is a left $\mathscr{O}_{\mathrm{B}}$-ideal; since $\mathscr{O}_{\mathrm{B}}$ is principal, one has $\iota_z(I)\mathscr{O}_{\mathrm{B}} = q\mathscr{O}_{\mathrm{B}}$ for some $q \in \mathrm{B}^\times(\mathbb{Q})$. Then we claim that $z' = q^{-1}zq \in \mathrm{R}_3^*(n)$. Indeed $\mathrm{Nr}(z') = n$;

$$z' = q^{-1}zq \in q^{-1}zq\mathscr{O}_{\mathrm{B}} = q^{-1}z\iota_z(I).\mathscr{O}_{\mathrm{B}} \subset q^{-1}\iota_z(I).\mathscr{O}_{\mathrm{B}} = \mathscr{O}_{\mathrm{B}},$$

since $z\iota_z(I) = \iota_z(\sqrt{-n}I) \subset \iota_z(I)$; it is easy to check that z' is primitive.

Note that the quaternion q is defined only up to multiplication on the right by an element of $\mathscr{O}_{\mathrm{B}}^\times$; if we replace q by qu for $u \in \mathscr{O}_{\mathrm{B}}^\times$ we see that z' is replaced by the conjugate of z' $\rho_{u^{-1}}(z') \in \mathrm{R}_3^*(n)$. Thus the ideal I defines a map between orbits

$$[I] : \begin{array}{ccc} \widetilde{\mathrm{R}_3}(n) & \mapsto & \widetilde{\mathrm{R}_3}(n) \\ \tilde{z} & \mapsto & [I].\tilde{z} := \tilde{z'} \end{array} .$$

If we replace I by an homothetic lattice $\lambda.I$ $\lambda \in K^\times$, then q is replaced by $q' = \iota_z(\lambda)q$ and $q'^{-1}zq' = q^{-1}\iota_z(\lambda^{-1}\sqrt{-n}\lambda)q = q^{-1}zq$ so that $[\lambda.I] = [I]$; finally it is not difficult to check that $[I.J] = [I] \circ [J]$ and that $[\mathscr{O}] = \mathrm{Id}$ so that we obtain a well defined action of $\mathrm{Pic}(\mathscr{O})$ on $\widetilde{\mathrm{R}_3}(n)$. We postpone the proof of the transitivity of this action and of the size of the stabilizer to Chapter 6.1 were these results are obtained in greater generality.

To prove (**??**), we need first to to evaluate the size of $\widetilde{\mathrm{R}_3}(n)$: for any $z \in \mathrm{R}_3^*(n)$, the stabilizer of z for the action of $\mathscr{O}_{\mathrm{B}}^\times$ is the set of $u \in \mathscr{O}_{\mathrm{B}}^\times$ commuting with z. It follows from Chap **??** that $u \in \iota_z(K) \cap \mathscr{O}_{\mathrm{B}}^\times = \iota_z(\mathscr{O}^\times)$. The later has order $4, 6$ or 2 for $n = 1, 3,\ \neq 1, 3$ respectively. Hence

$$|\mathrm{R}_3^*(n)| \leq$$

□

8. The distribution of the representations

As for the case of four squares, since $|\mathrm{R}_3(n)|$ (when non-zero) grows relatively rapidly with n, we may wish toconsider the distribution of the directions in $\mathbb{R}^3$ defined by the points of $\mathrm{R}_3(n) \subset \mathbb{R}^3$ or equivalently we may consider the distribution of their projection on the sphere S^2. Again, these points turn out to be well distributed

[6]Since K is a quadratic field, the notions of proper and locally principal ideal coincide

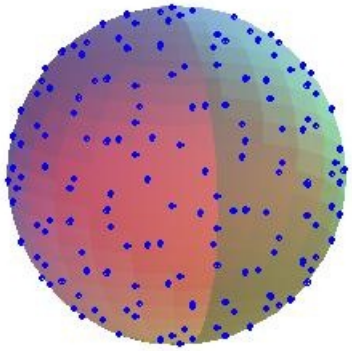

FIGURE 3. The distribution of $\mathrm{R}_3(101)$.

THEOREM 8.1 (Duke). *As $n \to +\infty$, $n \not\equiv 0, 4, 7(8)$, the set*

$$n^{-1/2}.\mathrm{R}_3(n) \subset S^2,$$

become equidistributed on S^2 w.r.t. the Lebesgue measure: for any $f \in \mathcal{C}(S^2)$,

$$\frac{1}{|\mathrm{R}_3(n)|} \sum_{\mathbf{x} \in \mathrm{R}_3(n)} \varphi(\frac{\mathbf{x}}{n^{1/2}}) \to \int_{S^2} \varphi(v) d\mu_{S^2}(v).$$

As we will see below, this beautiful theorem is significantly deeper than Malyshev's Theorem. Until the work of Duke [**Duk88**] the best result was the work of Linnik [**Lin68**] who, by his "ergodic method" proved this equidistribution of $n^{-1/2}R_3(n)$ for n satisfying an extra congruence condition (at some fixed odd prime). One can find in [**EMV10**] an exposition of Linnik's proof in a modern language.

8.1. Principle of the proof. Duke's original approach follow closely the principles of the proof of the equidistribution of representations as sums of four squares sketched previously. In the course of this book we will discuss an alternative approach but for now, and for completeness, let us present Duke's original approach as in [**Duk88**].

As for the 3-sphere, continuous function on S^2 may be uniformly approximated by a linear combinations of *harmonic homogeneous polynomials*:

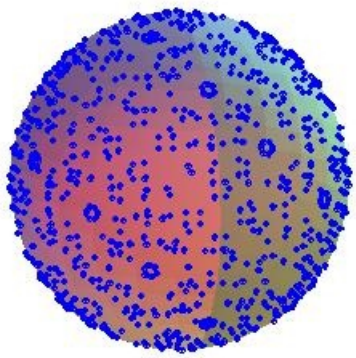

FIGURE 4. The distribution of $\mathrm{R}_3(78540)$.

the restriction to S^2 of a polynomial ϕ_ν, of degree ν, on $\mathbb{R}^3$ such that for $\lambda \in \mathbb{R}$

$$\phi_\nu(\lambda \mathbf{x}) = \lambda^\nu \phi_\nu(\mathbf{x}), \text{ and } \Delta_{\mathbb{R}^3}\phi_\nu = 0$$

were

$$\Delta_{\mathbb{R}^3} = \frac{\partial^2}{\partial^2 y} + \frac{\partial^2}{\partial^2 z} + \frac{\partial^2}{\partial^2 t}$$

is the Laplace operator. Such polynomial may be taken of the form

$$\phi_\nu(y,z,t) = (by + cz + dt)^\nu. \tag{8.1}$$

Where $(b,c,d) \in \mathbb{C}^3$, is an *isotropic* complex vector: ie. such that $q_3(b,c,d) = b^2 + c^2 + d^2 = 0$.

By Weyl's equidistribution criterion is it sufficient to show that for such non-constant polynomials, the corresponding Weyl sums decay:

$$W(\phi_\nu; n) = \frac{1}{|\mathrm{R}_3(n)|} \sum_{z \in \mathrm{R}_3(n)} \phi_\nu(\frac{z}{|n|^{1/2}}) \to \int_{S^2} \phi_\nu(x) d\mu_{S^2}(x) = 0. \tag{8.2}$$

Again, we may restrict to even degree ν (otherwise the Weyl sum is zero). In view of Theorem 5.1 and by homogeneity, it will be sufficient to show that

$$\sum_{\mathbf{x} \in \mathrm{R}_3(n)} \phi_\nu(\mathbf{x}) \ll_\phi n^{\frac{1}{2} + \frac{\nu}{2} - \delta} \tag{8.3}$$

for some absolute constant $\delta > 0$.

As before, the theta function

$$\theta(\phi_\nu, \mathbf{x}) = \sum_{(a,b,c)\in\mathbb{Z}^3} \phi_\nu(a,b,c)\exp(2\pi i(a^2+b^2+c^2)z)$$
$$= \sum_{n\geq 1}\Big(\sum_{\mathbf{x}\in \mathrm{R}_3(n)} \phi_\nu(\mathbf{x})\Big)\exp(2\pi i n z).$$

is a holomorphic modular cusp form of weight $k = 3/2+\nu$ with respect to the congruence subgroup

$$\Gamma_0(4) = \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}),\ c \equiv 0(4)\}.$$

That means that $\theta(\phi_\nu, z)$ is bounded on the upper-half plane $\mathbb{H}$ and that for any $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$,

$$\theta(\phi_\nu, \gamma z) = \theta(\phi_\nu, \frac{az+b}{cz+d}) = j(\gamma, z)^{3+2\nu}\theta(\phi_\nu, z),$$

with

$$j(\gamma, z) = (\frac{c}{d})\varepsilon_d^{-1}(cz+d)^{1/2}$$

with $(\frac{c}{d})$ the extended Legendre symbol, $\varepsilon_d = 1, \iota$ for $d \equiv 1, 3(4)$ and $(.)^{1/2}$ is the usual branch of the square root which is positive on $\mathbb{R}_{>0}$.

The sums $\sum_{\mathbf{x}\in \mathrm{R}_3(n)} \phi_\nu(\mathbf{x})$ which we seek to bound are the *Fourier coefficients* of $\theta(\phi_\nu, z)$ and by the Petersson formula, their square may be bounded by sums of *Salié* sums. Salié sums may be evaluated by elementary means and bounded individually in an optimal way; from this one obtains the following bound

$$\sum_{\mathbf{x}\in \mathrm{R}_3(n)} \phi_\nu(x) \ll_{\phi_\nu} n^{\frac{k-1}{2}+\frac{1}{4}+o(1)} = n^{\frac{1}{2}+\frac{\nu}{2}+o(1)}$$

which is *just* not sufficient. In the remarkable paper, Iwaniec [**Iwa87**] was able to take into account the summation over the Salié sums and to improve the constant $\frac{1}{4}$ above to $\frac{1}{4} - \frac{1}{22}$ replacing the final exponent by $\frac{1}{2} + \frac{\nu}{2} - \frac{1}{22}$ at least for weights $k \geq 7/2$. Duke generalized Iwaniec's method to bound Fourier coefficients to forms of weight $\geq 3/2$ as well as to more general –not necessarily holomorphic– forms of 1/2-integral weight and used this bound to prove the equidistribution of integral three squares representations as well as for other (possibly indefinite) quadratic forms. We discuss these in the next chapter.