10
Sums of Squares

Our main aim in this chapter is to determine which integers can be expressed as the sum of a given number of squares, that is, which have the form $x_1^2+\cdots+x_k^2$, where each $x_i \in \mathbb{Z}$, for a given k. We shall concentrate mainly on the two most important cases, characterising the sums of two squares, and showing that every non-negative integer is a sum of four squares. We shall adopt two completely different approaches to this problem: the first is mainly algebraic, making use of two number systems, the Gaussian integers and the quaternions; the second approach is geometric, based on the fact that the expression $x_1^2 + \cdots + x_k^2$ represents the square of the length of the vector $(x_1, \ldots, x_k)$ in $\mathbb{R}^k$. We shall therefore give two different proofs for several of the main theorems in this chapter. In mathematics, it is often useful to have more than one proof of a result, not because this adds anything to its validity (a single correct proof is enough for this), but rather because the extra proofs may add to our understanding of the result, and may enable us to extend it in different directions.

10.1 Sums of two squares

Definition

For each integer $k \geq 1$, let $S_k = \{n \mid n = x_1^2 + \cdots + x_k^2$ for some $x_1, \ldots, x_k \in \mathbb{Z}\}$, the set of all sums of k squares.

Example 10.1

$S_1 = \{0, 1, 4, 9, \ldots\}$ is the set of all squares. By inspection, S_2, the set of sums of two squares, contains $0, 1, 2, 4, 5$ and 8, but not $3, 6$ or 7.

Lemma 10.1

The set S_2, consisting of the sums of two squares, is closed under multiplication, that is, if $s, t \in S_2$ then $st \in S_2$.

Proof

Let $s = a_1^2 + b_1^2$ and $t = a_2^2 + b_2^2$ be elements of S_2, where $a_1, b_1, a_2, b_2 \in \mathbb{Z}$. Then the identity

$$(a_1^2 + b_1^2)(a_2^2 + b_2^2) = (a_1a_2 - b_1b_2)^2 + (a_1b_2 + b_1a_2)^2 \qquad (10.1)$$

shows that $st \in S_2$, since $a_1a_2 - b_1b_2, a_1b_2 + b_1a_2 \in \mathbb{Z}$. □

Example 10.2

We have $8 = 2^2+2^2$ and $10 = 3^2+1^2$, so $80 = 8.10 = (2.3-2.1)^2+(2.1+2.3)^2 = 4^2 + 8^2$.

Comments

1 It follows immediately that the product of *any* finite set of elements of S_2 is also in S_2.

2 The identity (10.1) can be verified directly by expanding each side. However, it is more useful to define a pair of complex numbers $z_i = a_i + \mathrm{i}b_i$ for $i = 1, 2$ (where $\mathrm{i} = \sqrt{-1}$), so that $a_i^2 + b_i^2 = z_i\overline{z_i} = |z_i|^2$. Now the rules for multiplying complex numbers (with $\mathrm{i}^2 = -1$) give $z_1z_2 = (a_1a_2-b_1b_2)+\mathrm{i}(a_1b_2+b_1a_2)$, so that $|z_1z_2|^2 = (a_1a_2-b_1b_2)^2+(a_1b_2+b_1a_2)^2$. One can therefore prove the identity by arguing that

$$|z_1z_2|^2 = (z_1z_2).\overline{(z_1z_2)} = z_1\overline{z_1}.z_2\overline{z_2} = |z_1^2|.|z_2^2|$$

for all $z_1, z_2 \in \mathbb{C}$. We will see later that a similar identity holds for S_4, but not for S_3, so that sums of four squares are easier to deal with than sums of three squares.

3 By replacing b_1 with $-b_1$ in (10.1) we obtain the equivalent identity

$$(a_1^2 + b_1^2)(a_2^2 + b_2^2) = (a_1a_2 + b_1b_2)^2 + (a_1b_2 - b_1a_2)^2, \qquad (10.2)$$

which we will need later.

Lemma 10.1 suggests that, in determining the elements of S_2, we should first consider prime numbers: each integer $n \geq 2$ is a product of primes, and if these prime factors are all in S_2 then so is n. However, not all primes are sums of two squares, the prime 3 being the first counterexample.

Exercise 10.1

Which of the primes $p < 100$ are elements of S_2? Do you notice a pattern emerging?

The following *Two Squares Theorem* was stated by Fermat in 1640, and proved by Euler in 1754.

Theorem 10.2

Each prime $p \equiv 1 \bmod (4)$ is a sum of two squares.

Proof

Since $p \equiv 1 \bmod (4)$, Corollary 7.7 implies that $-1 \in Q_p$; thus $-1 \equiv u^2 \bmod (p)$ for some u, so $u^2 + 1 = rp$ for some integer r. We can choose u so that $0 \leq u \leq p - 1$, giving $0 < r < p$. Now $rp = u^2 + 1^2 \in S_2$, so it follows that there is a smallest integer m such that $mp \in S_2$ and $0 < m < p$. If $m = 1$ then $p \in S_2$ and we are done, so assume that $m > 1$.

Since $mp \in S_2$, we have $mp = a_1^2 + b_1^2$ for some integers a_1 and b_1. Let a_2 and b_2 be the least absolute residues of a_1 and $b_1 \bmod (m)$, so that $a_2 \equiv a_1$ and $b_2 \equiv b_1 \bmod (m)$ and $|a_2|, |b_2| \leq m/2$. Then $a_2^2 + b_2^2 \equiv a_1^2 + b_1^2 \equiv 0 \bmod (m)$, so $a_2^2 + b_2^2 = sm$ for some integer s; since $|a_2|, |b_2| \leq m/2$ we have $a_2^2 + b_2^2 \leq 2(m/2)^2 = m^2/2$, so $s \leq m/2$ and hence $s < m$.

We also have $s > 0$: if $s = 0$ then $a_2^2 + b_2^2 = 0$, so $a_2 = b_2 = 0$, giving $a_1 \equiv b_1 \equiv 0 \bmod (m)$; then m divides a_1 and b_1, so m^2 divides $a_1^2 + b_1^2 = mp$ and hence m divides p, which is impossible since p is prime and $1 < m < p$. Thus $0 < s < m$.

Now $(a_1^2 + b_1^2)(a_2^2 + b_2^2) = mp.sm = m^2sp$, and identity (10.2) following Lemma 10.1 shows that

$$(a_1^2 + b_1^2)(a_2^2 + b_2^2) = (a_1a_2 + b_1b_2)^2 + (a_1b_2 - b_1a_2)^2,$$

so that

$$(a_1a_2 + b_1b_2)^2 + (a_1b_2 - b_1a_2)^2 = m^2sp.$$

Since $a_1a_2 + b_1b_2 \equiv a_2^2 + b_2^2 \equiv 0 \bmod (m)$ and $a_1b_2 - b_1a_2 \equiv a_2b_2 - b_2a_2 \equiv 0 \bmod (m)$, we can divide this equation through by m^2 to give

$$\left(\frac{a_1a_2 + b_1b_2}{m}\right)^2 + \left(\frac{a_1b_2 - b_1a_2}{m}\right)^2 = sp$$

where both $(a_1a_2 + b_1b_2)/m$ and $(a_1b_2 - b_1a_2)/m$ are integers. Thus $sp \in S_2$ with $0 < s < m$, contradicting the minimality of m. Hence $m = 1$ and the proof is complete. □

We will give an alternative geometric proof of Theorem 10.2 in Section 10.6. We can now give a complete description of the elements of S_2 in terms of their prime-power factorisations.

Theorem 10.3

A positive integer n is a sum of two squares if and only if every prime $q \equiv 3 \bmod (4)$ divides n to an even power (which may, of course, be 0 if $q \nmid n$).

Proof

($\Leftarrow$) By assumption, we can write

$$n = 2^e p_1^{e_1} \dots p_k^{e_k} q_1^{2f_1} \dots q_l^{2f_l} = 2^e p_1^{e_1} \dots p_k^{e_k} (q_1^2)^{f_1} \dots (q_l^2)^{f_l},$$

for some set of primes $p_i \equiv 1 \bmod (4)$ and $q_j \equiv 3 \bmod (4)$, where the exponents are integers $e \geq 0, e_i > 0$ and $f_j > 0$. Now $2 = 1^2 + 1^2 \in S_2$, Theorem 10.2 shows that each $p_i \in S_2$, and also each $q_j^2 = q_j^2 + 0^2 \in S_2$. Thus n is a product of elements of S_2, so Lemma 10.1 implies that $n \in S_2$.

($\Rightarrow$) Let $n \in S_2$, say $n = x^2 + y^2$. Let q be any prime such that $q \equiv 3 \bmod (4)$, let $q^f \parallel n$, and suppose (for a contradiction) that f is odd. If d denotes the greatest common divisor of x and y, then $x = ad$ and $y = bd$ where $\gcd(a, b) = 1$, so $n = (a^2 + b^2)d^2$ and hence $nd^{-2} = a^2 + b^2$. If $q^e \parallel d$ then $q^{f-2e} | nd^{-2}$; now $f - 2e$ is odd and hence non-zero, so $q | nd^{-2} = a^2 + b^2$ and hence $a^2 \equiv -b^2 \bmod (q)$. Now b cannot be divisible by q (for then q would divide both a and b, whereas they are coprime), so b is a unit mod (q). If c is the inverse of b in U_q, then multiplying through by c^2 we have $(ac)^2 \equiv -1 \bmod (q)$, so that $-1 \in Q_q$. This is impossible for a prime $q \equiv 3 \bmod (4)$ by Corollary 7.7, so f must be even. □

Example 10.3

The integer 60 ($= 2^2.3.5$) is not a sum of two squares, since the exponent of 3 dividing it is odd. However, 180 ($= 2^2.3^2.5$) is a sum of two squares. To find them, first write 5 as a sum of two squares: $5 = 2^2 + 1^2$. Now multiplying through by $2^2.3^2$ we get $180 = 2^2.3^2.5 = (2.3.2)^2 + (2.3.1)^2 = 12^2 + 6^2$.

Example 10.4

The integer 221 ($= 13.17$) is a sum of two squares, since $13 \equiv 17 \equiv 1 \bmod (4)$. To find these squares, imitate the proof of Lemma 10.1, with $13 = 3^2 + 2^2$ and $17 = 4^2 + 1^2$ corresponding to the equations $s = a_1^2 + b_1^2$ and $t = a_2^2 + b_2^2$. Then

$$221 = st = (a_1a_2 - b_1b_2)^2 + (a_1b_2 + b_1a_2)^2 = (3.4 - 2.1)^2 + (3.1 + 2.4)^2 = 10^2 + 11^2.$$

(Note that equation (10.2) sometimes gives a different expression, e.g. $221 = 14^2 + 5^2$ in this case.) Similarly, one can express 6409 ($= 221.29$) as a sum of two squares by repeating this process: $221 = 10^2 + 11^2$ and $29 = 5^2 + 2^2$, so $6409 = (10.5 - 11.2)^2 + (10.2 + 11.5)^2 = 28^2 + 75^2$.

Exercise 10.2

Write each of the following integers as a sum of two squares, or show that this is impossible: 130, 260, 847, 980, 1073.

Exercise 10.3

Find all the pairs $(x, y) \in \mathbb{Z}^2$ satisfying $x^2 + y^2 = 50$.

As a special case of Theorem 10.3, we have the following stronger form of Theorem 10.2:

Corollary 10.4

A prime p is a sum of two squares if and only if $p = 2$ or $p \equiv 1 \bmod (4)$.

10.2 The Gaussian integers

A representation of an integer n as a *difference* of two squares, say $n = x^2 - y^2$, gives rise to a factorisation of n as a product of two integers,

$$n = x^2 - y^2 = (x + y)(x - y),$$

where $x + y \equiv x - y \bmod (2)$. Conversely, if $n = rs$ where $r \equiv s \bmod (2)$, then by writing $x = (r + s)/2$ and $y = (r - s)/2$ we get $n = x^2 - y^2$ with $x, y \in \mathbb{Z}$. This link with factorisations can be extended to *sums* of two squares if we write

$$n = x^2 + y^2 = (x + y\mathrm{i})(x - y\mathrm{i}),$$

where $\mathrm{i} = \sqrt{-1} \in \mathbb{C}$: given any factorisation $n = rs$ of this form, we now write $x = (r+s)/2$ and $y = (r-s)/2\mathrm{i}$, provided these are integers. This suggests that we should study the complex numbers of the form $x + y\mathrm{i}$ ($x, y \in \mathbb{Z}$), known as the *Gaussian integers*, and in particular their factorisations.

The set

$$\mathbb{Z}[\mathrm{i}] = \{x + y\mathrm{i} \mid x, y \in \mathbb{Z}\}$$

of all Gaussian integers is closed under addition, subtraction and multiplication: for instance, $(a + b\mathrm{i})(c + d\mathrm{i}) = (ac - bd) + (ad + bc)\mathrm{i}$, and if a, b, c, d are integers then so are $ac - bd$ and $ad + bc$. The usual axioms (associativity, distributivity, etc.) are satisfied, so $\mathbb{Z}[\mathrm{i}]$ is a ring; however, it is not a field, since not every non-zero element of $\mathbb{Z}[\mathrm{i}]$ is a unit (see Exercise 10.4). Thus $\mathbb{Z}[\mathrm{i}]$ shares many of the basic properties of $\mathbb{Z}$, so it is not surprising that many of our earlier results about divisibility and factorisation of integers extend in a natural way to Gaussian integers.

There are two other important properties of $\mathbb{Z}$ shared by $\mathbb{Z}[\mathrm{i}]$. The first is that if $r, s \neq 0$ then $rs \neq 0$, or equivalently, if $rs = 0$ then $r = 0$ or $s = 0$ (this is true in $\mathbb{Z}[\mathrm{i}]$ since it is true in $\mathbb{C}$, which contains $\mathbb{Z}[\mathrm{i}]$). A ring with this property is called an *integral domain*. This property is useful since it allows one to cancel non-zero factors: if $r's = r''s$ with $s \neq 0$, then $(r' - r'')s = 0$ so that $r' - r'' = 0$ and hence $r' = r''$.

The second important property of $\mathbb{Z}$ is the Division Algorithm (Corollary 1.2), which allows one to divide an integer a by a non-zero integer b, with a remainder which is small compared with b. An integral domain R is a *Euclidean domain* if, for each $a \in R \setminus \{0\}$ there is an integer $d(a) \geq 0$ such that

(1) $d(ab) \geq d(b)$ for all $a, b \neq 0$, with equality if and only if a is a unit;

(2) for all $a, b \in R$ with $b \neq 0$, there exist $q, r \in R$ such that $a = qb + r$, with $r = 0$ or $d(r) < d(b)$.

The function d assigns a measure of size to each non-zero element of R. Condition (1) simply states that this function behaves reasonably with respect to products, while condition (2) is the analogue of Corollary 1.2, giving a remainder r which is small in comparison with b.

Example 10.5

The ring $\mathbb{Z}$ is a Euclidean domain, with $d(r) = |r|$: condition (1) is clear, using the fact that the units of $\mathbb{Z}$ are ± 1, while condition (2) is simply Corollary 1.2.

Example 10.6

If F is any field, then the ring $F[x]$ of polynomials in one variable x, with coefficients in F, is a Euclidean domain. For each non-zero $f = f(x) \in F[x]$ we define $d(f) = \deg(f)$, the degree of the polynomial $f(x)$. Then condition (1) follows from the facts that $\deg(fg) = \deg(f) + \deg(g)$, and that f is a unit if and only if it is a non-zero constant polynomial, that is, $\deg(f) = 0$; condition (2) follows from polynomial division.

Example 10.7

The ring $\mathbb{Z}[i]$ of Gaussian integers is a Euclidean domain, with $d(z) = z\overline{z} = |z|^2 = x^2 + y^2$ for each $z = x + yi \in \mathbb{Z}[i]$. (This is sometimes called the *norm* of z, written $N(z)$.) Condition (1) is straightforward, using the facts that $d(zw) = d(z)d(w)$ for all z, w (see Comment 2 of Section 1), and that the units in $\mathbb{Z}[i]$ are ± 1 and $\pm i$ (see Exercises 10.4 and 10.5).

Exercise 10.4

Show that if $u \in \mathbb{Z}[i]$ then the following are equivalent:

(a) u is a unit in $\mathbb{Z}[i]$;

(b) $d(u) = 1$;

(c) $u = \pm 1$ or $\pm i$.

Exercise 10.5

Verify condition (1) for $\mathbb{Z}[i]$, that $d(ab) \geq d(b)$ for all non-zero $a, b \in \mathbb{Z}[i]$, with equality if and only if a is a unit.

We will give a geometric proof of condition (2) for $\mathbb{Z}[\mathrm{i}]$, though an algebraic proof is also possible. The Gaussian integers $q = x+y\mathrm{i} \in \mathbb{Z}[\mathrm{i}]$ can be regarded as the points $(x, y) \in \mathbb{R}^2$ with integer coordinates; as such they are the vertices of a tessellation (tiling) of the plane by squares with side-length 1. If b is a non-zero element of $\mathbb{Z}[\mathrm{i}]$, then the multiples qb of b, with $q \in \mathbb{Z}[\mathrm{i}]$, are also the vertices of a tessellation by squares, obtained by multiplying the original tessellation by b; equivalently, we rotate the original tessellation about the origin by an angle $\arg(b)$, and expand it by a factor $|b|$, so these squares have side-length $|b|$. Every complex number (and hence every Gaussian integer) a is in at least one of these squares, and its distance from one of the vertices qb is at most $|b|/\sqrt{2}$ (attained if a is the centre of a square). If we define $r = a - qb$, then $r \in \mathbb{Z}[\mathrm{i}]$, $a = qb + r$, and $|r| \leq |b|/\sqrt{2}$, so $d(r) = |r|^2 \leq |b|^2/2 < |b|^2 = d(b)$, as required.

The main results from Chapters 1 and 2 concerning divisors and factorisations all extend to any Euclidean domain R, with only minor changes of terminology. It is a useful exercise to adapt their proofs to establish the following more general results, and to see what they mean for $\mathbb{Z}[\mathrm{i}]$. First we need some terminology.

Two elements $a, b \in R$ are *associates* if $a = ub$ for some unit u of R; this is an equivalence relation, since the units form a group under multiplication. An element $a \in R$, which is not 0 or a unit, is *irreducible* if its only divisors are units or its associates; otherwise, a is *reducible*. In $\mathbb{Z}$, for instance, the associates of a are $\pm a$, and the irreducible elements are those of the form $\pm p$ where p is prime.

The main result we need here states that each element of a Euclidean domain R, other than 0 or a unit, is a product of powers of irreducible elements; moreover, this representation is unique, apart from permuting factors and replacing them with their associates. In the case $R = \mathbb{Z}$ this is the Fundamental Theorem of Arithmetic (Theorem 2.3). The proof for general Euclidean domains is very similar, and we will omit the details, since they can be found in many algebra textbooks.

In order to apply this to the Gaussian integers, we need to determine the irreducible elements of $\mathbb{Z}[\mathrm{i}]$.

Each prime $q \equiv 3 \bmod (4)$ in $\mathbb{Z}$ is irreducible in $\mathbb{Z}[\mathrm{i}]$. For if $q = zw$ with $z, w \in \mathbb{Z}[\mathrm{i}]$, then $d(z)d(w) = d(q) = q^2$ in $\mathbb{Z}$, so either $d(z) = d(w) = q$ or $\{d(z), d(w)\} = \{1, q^2\}$. In the first case, putting $z = x+y\mathrm{i}$ we get $q = x^2+y^2 \in S_2$, which is impossible by Corollary 10.4; hence either $d(z) = 1$ or $d(w) = 1$, so z or w is a unit, and q is irreducible. For instance, the primes $q = 3, 7, 11, 19, \ldots$ are all irreducible as Gaussian integers. Each prime $q \equiv 3 \bmod (4)$ gives rise to four associates $\pm q$ and $\pm q\mathrm{i}$, all irreducible in $\mathbb{Z}[\mathrm{i}]$; by the uniqueness of factorisation in $\mathbb{Z}[\mathrm{i}]$, these are the only irreducible elements dividing q.

On the other hand, any prime $p = 2$ or $p \equiv 1 \bmod (4)$ is in S_2, so $p = x^2+y^2$

for some integers x and y, giving a factorisation $p = (x + yi)(x - yi)$ of p in $\mathbb{Z}[i]$. These factors $\pi = x + yi$ and $\overline{\pi} = x - yi$ must be irreducible: if $x \pm yi = st$ in $\mathbb{Z}[i]$ then $d(s)d(t) = d(x \pm yi) = p$ in $\mathbb{Z}$, so $d(s) = 1$ or $d(t) = 1$ and hence s or t must be a unit. For instance, $2 = (1 + i)(1 - i)$, so $1 + i$ and $1 - i$ are irreducible, and multiplying by units we obtain four associate irreducible elements: $1+i$, $i(1+i) = -1+i$, $-(1+i) = -1-i$ and $-i(1+i) = 1-i$. However, if $p \equiv 1 \bmod (4)$ we obtain eight irreducible elements $\pm x \pm yi$ and $\pm y \pm xi$, consisting of four associates of each of π and $\overline{\pi}$; for instance $5 = 1^2 + 2^2 = (1 + 2i)(1 - 2i)$, giving the irreducible elements $\pm 1 \pm 2i$ and $\pm 2 \pm i$. In either case, the uniqueness of factorisation implies that these are the only irreducible divisors of p in $\mathbb{Z}[i]$.

These irreducible elements $\pi, \overline{\pi}$ and q, together with their associates, are in fact the only irreducible elements of $\mathbb{Z}[i]$. For suppose that z is irreducible in $\mathbb{Z}[i]$. Now z divides the positive integer $z\overline{z} = d(z)$, so there is a least positive integer n such that z divides n in $\mathbb{Z}[i]$; since z is not a unit, $n > 1$. Now n must be prime, for if $n = ab$ in $\mathbb{Z}$ with $a, b > 0$, then z divides a or b in $\mathbb{Z}[i]$ (by the analogue of Lemma 2.1(b) for $\mathbb{Z}[i]$), so $a = n$ or $b = n$ by the minimality of n. We have already determined which irreducible Gaussian integers divide the various primes $n = 2$, $n = p \equiv 1$ and $n = q \equiv 3 \bmod (4)$, so z must be an associate of some $\pi, \overline{\pi}$ or q, as required.

The uniqueness of factorisation in $\mathbb{Z}[i]$ implies that the representation of a prime $p \in S_2$ as $x^2 + y^2$ is essentially unique, apart from the obvious changes of transposing x and y, and multiplying either or both of them by -1. More precisely, if $r(n)$ denotes the number of pairs $(x, y) \in \mathbb{Z}^2$ such that $x^2 + y^2 = n$, then $r(2) = 4$, from the representations $2 = (\pm 1)^2 + (\pm 1)^2$, and $r(p) = 8$ for primes $p \equiv 1 \bmod (4)$, from $p = (\pm x)^2 + (\pm y)^2 = (\pm y)^2 + (\pm x)^2$; of course, Corollary 10.4 gives $r(q) = 0$ for primes $q \equiv 3 \bmod (4)$.

Using our knowledge of the irreducible elements of $\mathbb{Z}[i]$, we can in fact evaluate $r(n)$ for all n. Suppose that n factorises in $\mathbb{Z}$ as

$$n = 2^e p_1^{e_1} \dots p_k^{e_k} q_1^{f_1} \dots q_l^{f_l}$$

for primes $p_i \equiv 1$ and $q_j \equiv 3 \bmod (4)$, and integers $e \geq 0$ and $e_i, f_j > 0$. By factorising each prime $2, p_i$ and q_j in $\mathbb{Z}[i]$, we find that n has factorisation

$$\begin{aligned} n &= (1+i)^e (1-i)^e \prod_i \pi_i^{e_i} (\overline{\pi_i})^{e_i} \prod_j q_j^{f_j} \\ &= i^e (1-i)^{2e} \prod_i \pi_i^{e_i} (\overline{\pi_i})^{e_i} \prod_j q_j^{f_j} \end{aligned}$$

in $\mathbb{Z}[i]$, where $1-i$, π_i, $\overline{\pi_i}$ and q_j are all irreducible, and $\pi_i \overline{\pi_i} = p_i$ for $i = 1, \dots, k$. Now $n = x^2 + y^2$ if and only if $n = (x+yi)(x-yi)$ in $\mathbb{Z}[i]$, so $r(n)$ is the number

of distinct factors z of n in $\mathbb{Z}[\mathrm{i}]$ such that $n = z\overline{z}$. Any factor z of n must be a product of a unit u and various irreducible factors of n, that is,

$$z = u(1-\mathrm{i})^a \prod_i \pi_i^{a_i} (\overline{\pi_i})^{b_i} \prod_j q_j^{c_j}$$

where $0 \le a \le 2e$, $0 \le a_i, b_i \le e_i$, and $0 \le c_j \le f_j$. Then

$$\begin{aligned} \overline{z} &= \overline{u}(1+\mathrm{i})^a \prod_i (\overline{\pi_i})^{a_i} \pi_i^{b_i} \prod_j q_j^{c_j} \\ &= \overline{u}\mathrm{i}^a (1-\mathrm{i})^a \prod_i \pi_i^{b_i} (\overline{\pi_i})^{a_i} \prod_j q_j^{c_j}, \end{aligned}$$

and by using $u\overline{u} = 1$ we can combine these factorisations of z and $\overline{z}$ to obtain the factorisation

$$z\overline{z} = \mathrm{i}^a (1-\mathrm{i})^{2a} \prod_i \pi_i^{a_i+b_i} (\overline{\pi_i})^{a_i+b_i} \prod_j q_j^{2c_j}.$$

By comparing this with the factorisation of n, and using the uniqueness of factorisation of Gaussian integers, we see that $z\overline{z} = n$ if and only if $a = e$, $a_i + b_i = e_i$ and $2c_j = f_j$ for all i and j. Now $r(n)$ is the number of such factors z of n, so $r(n) = 0$ unless each f_j is even, confirming Theorem 10.3; when this happens, the values of a $(= e)$ and c_j $(= f_j/2)$ are uniquely determined by n, while there are four choices for the unit u, and $e_i + 1$ choices for each pair a_i, b_i (since $a_i = e_i - b_i = 0, 1, \ldots, e_i$). Thus

$$r(n) = \begin{cases} 4\prod_{i=1}^k (e_i+1) & \text{if } f_1, \ldots, f_l \text{ are all even,} \\ 0 & \text{otherwise.} \end{cases}$$

Note that Theorem 8.3 implies that $\prod_{i=1}^k (e_i + 1) = \tau(n_1)$, the number of divisors of $n_1 = \prod_{i=1}^k p_i^{e_i}$.

Example 10.8

Let $n = 30420 = 2^2.5.13^2.3^2$. The only prime $q_j \equiv 3 \bmod (4)$ dividing n is $q_1 = 3$, so $l = 1$ with $f_1 = 2$ (the exponent of 3), which is even. The primes $p_i \equiv 1 \bmod (4)$ dividing n are $p_1 = 5$ and $p_2 = 13$, so $k = 2$ with $e_1 = 1$ and $e_2 = 2$ (the exponents of 5 and 13). Thus $r(30420) = 4(1+1)(2+1) = 24$. Since $5 = (2+\mathrm{i})(2-\mathrm{i})$ and $13 = (3+2\mathrm{i})(3-2\mathrm{i})$, and since $e = 2$, the factors z such that $z\overline{z} = n$ have the form

$$z = u(1-\mathrm{i})^2.(2+\mathrm{i})^{a_1}(2-\mathrm{i})^{1-a_1}(3+2\mathrm{i})^{a_2}(3-2\mathrm{i})^{2-a_2}.3$$

where u is a unit, $0 \le a_1 \le 1$ and $0 \le a_2 \le 2$. Now $u(1-\mathrm{i})^2.3 = 6v$, where $v = -\mathrm{i}u$ is a unit, so if we take $a_1 = 0, 1$ and $a_2 = 0, 1, 2$ in turn then we get 24 factors

$$z = x + y\mathrm{i} = 6v(-2 \pm 29\mathrm{i}), \quad 6v(26 \pm 13\mathrm{i}), \quad 6v(22 \pm 19\mathrm{i}).$$

These give 24 representations of $n = 30420$ as x^2+y^2, each of which is obtained from one of

$$12^2 + 174^2, \ 156^2 + 78^2, \ 132^2 + 114^4$$

by transposing x and y, or multiplying them by ± 1, or both.

Exercise 10.6

Calculate $r(221)$, and find all representations of 221 as a sum of two squares. (See Example 10.4.)

Exercise 10.7

Calculate $r(16660)$, and find all representations of 16660 as a sum of two squares.

Exercise 10.8

Show that $r(n) = 4\big(\tau_1(n) - \tau_3(n)\big)$ for all n, where $\tau_1(n)$ and $\tau_3(n)$ denote the number of divisors d of n such that $d \equiv 1$ or $3 \bmod (4)$ respectively (see Chapter 8, Exercise 8.21).

10.3 Sums of three squares

Gauss proved that $n \in S_3$ if and only if $n \neq 4^e(8k+7)$; thus $7, 15, 23, 28, \ldots$ are not sums of three squares. It is a simple exercise (see below) to prove that no integer $n = 4^e(8k+7)$ can be a sum of three squares. The converse, which we will omit for lack of space, is rather harder, mainly because the set S_3 is not closed under multiplication: for instance 3 and 5 are sums of three squares, but 15 is not.

Exercise 10.9

Show that if $n \in S_3$ then $n \not\equiv 7 \bmod (8)$.

Exercise 10.10

Show that if $n \in S_3$ and n is divisible by 4, then $n/4 \in S_3$.

Exercise 10.11

Deduce that if $n = 4^e(8k+7)$ then $n \notin S_3$.

Exercise 10.12

In how many ways can 14 and 11 be written as sums of three squares?

10.4 Sums of four squares

Perhaps surprisingly, it is easier to deal with sums of four squares: first we need the following result.

Lemma 10.5

The set S_4, consisting of the sums of four squares, is closed under multiplication, that is, if $s, t \in S_4$ then $st \in S_4$.

Proof

This follows immediately from the identity

$$\begin{aligned}(a_1^2+b_1^2+c_1^2+d_1^2)(a_2^2+b_2^2+c_2^2+d_2^2) \;=\; & (a_1a_2-b_1b_2-c_1c_2-d_1d_2)^2 \\ & +(a_1b_2+b_1a_2+c_1d_2-d_1c_2)^2 \\ & +(a_1c_2-b_1d_2+c_1a_2+d_1b_2)^2 \\ & +(a_1d_2+b_1c_2-c_1b_2+d_1a_2)^2, \end{aligned} \tag{10.3}$$

which can be proved directly (at some length) by expanding each side. (We will shortly give an alternative explanation of this identity in terms of quaternions.)

□

Comment

By replacing b_1, c_1 and d_1 with $-b_1, -c_1$ and $-d_1$, we obtain the identity

$$\begin{aligned}(a_1^2+b_1^2+c_1^2+d_1^2)(a_2^2+b_2^2+c_2^2+d_2^2) &= (a_1a_2+b_1b_2+c_1c_2+d_1d_2)^2 \\ &\quad +(a_1b_2-b_1a_2-c_1d_2+d_1c_2)^2 \\ &\quad +(a_1c_2+b_1d_2-c_1a_2-d_1b_2)^2 \\ &\quad +(a_1d_2-b_1c_2+c_1b_2-d_1a_2)^2,\end{aligned} \tag{10.4}$$

which we will need later.

The following *Four Squares Theorem* was proved by Lagrange in 1770.

Theorem 10.6

Every non-negative integer is a sum of four squares.

Proof

Clearly $0, 1, 2 \in S_4$, so by Lemma 10.5 it is sufficient to prove that every odd prime p is in S_4. We do this by following the method of proof of Theorem 10.2 as far as possible. First we show that some positive multiple of p is in S_4.

Of the elements of $\mathbb{Z}_p$, exactly $(p+1)/2$ are squares, namely 0 and the $(p-1)/2$ quadratic residues in Q_p. Thus the set $K = \{z \in \mathbb{Z}_p \mid z = k^2,\ k \in \mathbb{Z}_p\}$ contains $(p+1)/2$ elements, and a similar argument shows that the set $L = \{z \in \mathbb{Z}_p \mid z = -1 - l^2,\ l \in \mathbb{Z}_p\}$ also has $(p+1)/2$ elements. Each of these two subsets thus accounts for more than half of the elements of $\mathbb{Z}_p$, so their intersection $K \cap L$ is non-empty. This means that there exist integers $u, v \in \mathbb{Z}$ such that $u^2 \equiv -1 - v^2 \bmod (p)$, that is, $u^2 + v^2 + 1 = rp$ for some integer $r > 0$. (As an example, if $p = 7$ then $K = \{0, 1, 2, 4\}$ and $L = \{2, 4, 5, 6\}$ so we can take $u^2 \equiv -1 - v^2 \equiv 2$ (or 4) mod (7), say $u = 3$ and $v = 2$ with $u^2 + v^2 + 1 = 14$.) Since $u^2 + v^2 + 1 = u^2 + v^2 + 1^2 + 0^2$, we have shown that some positive multiple rp of p is in S_4. By replacing u and v with their least absolute residues mod (p) we may assume that $|u|, |v| \leq p/2$, so that $r < p$. It follows that there exists a least positive integer $m < p$ such that $mp \in S_4$, say

$$mp = a_1^2 + b_1^2 + c_1^2 + d_1^2. \tag{10.5}$$

If $m = 1$ then $p \in S_4$ and we are home, so assume that $m > 1$.

Imitating the proof of Theorem 10.2, we take the least absolute residues a_2, b_2, c_2, d_2 of a_1, b_1, c_1, d_1 mod (m), so $a_2 \equiv a_1 \bmod (m)$, etc., and $|a_2|$, $|b_2|$,

$|c_2|, |d_2| \leq m/2$. We have $a_2^2+b_2^2+c_2^2+d_2^2 \equiv a_1^2+b_1^2+c_1^2+d_1^2 \equiv 0 \bmod (m)$, so

$$a_2^2+b_2^2+c_2^2+d_2^2 = sm \tag{10.6}$$

for some integer s. Now we would like to be able to assert that $s < m$, so that the proof could be completed as in Theorem 10.2; unfortunately, our bound on a_2, b_2, c_2 and d_2 merely implies that $sm \leq 4.(m/2)^2 = m^2$, so that $s \leq m$, which is not strong enough. However, if m is odd then since least absolute residues are integers we actually have $|a_2|, |b_2|, |c_2|, |d_2| < m/2$, so $s < m$ as required. We therefore need to eliminate the possibility that m is even. If m is even, then equation (10.5) implies that all, or two, or none of a_1, b_1, c_1 and d_1 are odd; by renaming these variables we may assume that a_1 and b_1 have the same parity, as do c_1 and d_1, that is, $a_1 \pm b_1$ and $c_1 \pm d_1$ are all even. But then

$$\left(\frac{a_1+b_1}{2}\right)^2+\left(\frac{a_1-b_1}{2}\right)^2+\left(\frac{c_1+d_1}{2}\right)^2+\left(\frac{c_1-d_1}{2}\right)^2 = \frac{a_1^2+b_1^2+c_1^2+d_1^2}{2} = \frac{mp}{2}$$

is an element of S_4 which is a positive multiple of p, contradicting the minimality of m. Thus m is odd, so $s < m$ as shown above.

We now show that $s > 0$. If $s = 0$, then equation (10.6) implies that $a_2 = b_2 = c_2 = d_2 = 0$, so a_1, b_1, c_1 and d_1 are all divisible by m, and equation (10.5) implies that p is divisible by m. This is impossible, since p is prime and $1 < m < p$, so we must have $s > 0$.

Equations (10.5) and (10.6) show that

$$(a_1^2+b_1^2+c_1^2+d_1^2)(a_2^2+b_2^2+c_2^2+d_2^2) = mp.sm = m^2sp,$$

and we can use identity (10.4) to write this as

$$\begin{aligned} m^2sp &= (a_1a_2+b_1b_2+c_1c_2+d_1d_2)^2+(a_1b_2-b_1a_2-c_1d_2+d_1c_2)^2 \\ &\quad +(a_1c_2+b_1d_2-c_1a_2-d_1b_2)^2+(a_1d_2-b_1c_2+c_1b_2-d_1a_2)^2. \end{aligned} \tag{10.7}$$

The congruences $a_1 \equiv a_2 \bmod (m)$, etc., together with (10.6), show that

$$a_1a_2+b_1b_2+c_1c_2+d_1d_2 \equiv a_2^2+b_2^2+c_2^2+d_2^2 \equiv 0 \bmod (m),$$

$$a_1b_2-b_1a_2-c_1d_2+d_1c_2 \equiv a_2b_2-b_2a_2-c_2d_2+d_2c_2 \equiv 0 \bmod (m),$$

etc., so each bracketed term on the right-hand side of (10.7) is divisible by m. We can therefore rewrite (10.7) as

$$\begin{aligned} sp &= \left(\frac{a_1a_2+b_1b_2+c_1c_2+d_1d_2}{m}\right)^2+\left(\frac{a_1b_2-b_1a_2-c_1d_2+d_1c_2}{m}\right)^2 \\ &\quad +\left(\frac{a_1c_2+b_1d_2-c_1a_2-d_1b_2}{m}\right)^2+\left(\frac{a_1d_2-b_1c_2+c_1b_2-d_1a_2}{m}\right)^2, \end{aligned}$$

so that sp is a positive multiple of p contained in S_4. Since $s < m$ this contradicts the minimality of m, so $m = 1$ and the result is proved. (We will give an alternative geometric proof later in this chapter.) □

Exercise 10.13

Express the following integers as sums of four squares: 247, 308, 465.

Exercise 10.14

In how many ways can 28 be written as a sum of four squares?

10.5 Digression on quaternions

Just as the two-squares identity (10.1) in Lemma 10.1 can be explained in terms of complex numbers, the four-squares identity (10.3) in the proof of Lemma 10.5 can be derived from a generalisation of complex numbers known as the quaternions. In the first half of the 19th century, Hamilton tried to find a 3-dimensional number system which would model the real world $\mathbb{R}^3$ in the same way as the 2-dimensional system $\mathbb{C}$ of complex numbers is an algebraic model of the plane $\mathbb{R}^2$. He wanted this system to retain as many as possible of the basic properties of $\mathbb{C}$, and in particular he wanted the length of a product to be equal to the product of the lengths of its factors (see Comment 2 following Lemma 10.1). After several years without success, he eventually realised in 1843 that this property would require a 4-dimensional system, rather than one based on $\mathbb{R}^3$. Its elements, which Hamilton called *quaternions*, are the points $q = (a, b, c, d) \in \mathbb{R}^4$, and addition and subtraction are performed by the usual method for vectors. To define multiplication, it is useful to write each quaternion in the form

$$q = a + b\mathrm{i} + c\mathrm{j} + d\mathrm{k} = a1 + b\mathrm{i} + c\mathrm{j} + d\mathrm{k},$$

where $a, b, c, d \in \mathbb{R}$ and $1, \mathrm{i}, \mathrm{j}, \mathrm{k}$ denote the standard basis vectors of $\mathbb{R}^4$. (This is analogous to identifying each point $(a, b) \in \mathbb{R}^2$ with the complex number $a + b\mathrm{i} = a1 + b\mathrm{i}$.) Hamilton defined the products of the basis vectors by

$$\mathrm{i}^2 = \mathrm{j}^2 = \mathrm{k}^2 = -1, \quad \mathrm{ij} = \mathrm{k} = -\mathrm{ji}, \quad \mathrm{jk} = \mathrm{i} = -\mathrm{kj}, \quad \mathrm{ki} = \mathrm{j} = -\mathrm{ik},$$

together with the rule that $1^2 = 1, 1\mathrm{i} = \mathrm{i} = \mathrm{i}1$ and so on. (Notice that multiplication is *not* commutative, since $\mathrm{ij} \neq \mathrm{ji}$ for example.) By assuming distributivity (that is, $q(q' + q'') = qq' + qq''$ and $(q' + q'')q = q'q + q''q$ for all q, q', q''), we find that the product of any pair of quaternions

$$q_1 = a_1 + b_1\mathrm{i} + c_1\mathrm{j} + d_1\mathrm{k} \quad \text{and} \quad q_2 = a_2 + b_2\mathrm{i} + c_2\mathrm{j} + d_2\mathrm{k}$$

is given by

$$\begin{aligned} q_1q_2 &= (a_1a_2 - b_1b_2 - c_1c_2 - d_1d_2) + (a_1b_2 + b_1a_2 + c_1d_2 - d_1c_2)\mathrm{i} \\ &\quad +(a_1c_2 - b_1d_2 + c_1a_2 + d_1b_2)\mathrm{j} + (a_1d_2 + b_1c_2 - c_1b_2 + d_1a_2)\mathrm{k}. \end{aligned} \tag{10.8}$$

The *conjugate* of a quaternion $q = a + b\mathrm{i} + c\mathrm{j} + d\mathrm{k}$ is the quaternion $\bar{q} = a - b\mathrm{i} - c\mathrm{j} - d\mathrm{k}$, and the *length* $|q|$ of q is given by

$$|q| = \sqrt{a^2 + b^2 + c^2 + d^2}.$$

Exercise 10.15

Verify that $|q|^2 = q\bar{q}$ for all quaternions q, and that $\overline{q_1q_2} = \overline{q_2}.\overline{q_1}$ for all quaternions q_1, q_2.

Exercise 10.16

Use Exercise 10.15 to prove that $|q_1q_2|^2 = |q_1|^2.|q_2|^2$, and deduce identity (10.3).

The quaternion number system is usually denoted by $\mathbb{H}$, in honour of Hamilton. He wrote two books and numerous papers on quaternions, exploiting their 4-dimensional nature to study space and time simultaneously. Soon after Hamilton's discovery of the quaternions, Cayley and Graves independently discovered a non-associative 8-dimensional system $\mathbb{O}$, the *octonions*, which leads to an eight-squares identity, analogous to those we have seen for two and four squares. Hamilton's failure to find a 3-dimensional number system was not due to any lack of effort or ability on his part: in 1878 Frobenius proved that $\mathbb{R}, \mathbb{C}$ and $\mathbb{H}$ are the only systems with the required properties (to be precise, these are the only finite-dimensional associative division algebras over $\mathbb{R}$), and in 1898 Hurwitz showed there is a k-squares identity of the required form only for $k = 1, 2, 4$ and 8. These facts help to explain why sums of three squares are harder to study than sums of two or four squares. For more on quaternions and related number systems, see Ebbinghaus *et al.* (1991).

10.6 Minkowski's Theorem

We will now reconsider some of the preceding results from a geometric point of view. The proof of Theorem 7.11 (quadratic reciprocity) involves the counting

of lattice-points in a subset of Euclidean space. This idea is useful elsewhere, for instance in studying the function $r(n)$ considered in Section 2 (see Exercise 10.26). Before applying lattices to sums of squares, we first need to study their properties a little more formally.

Definition

A *lattice* in $\mathbb{R}^n$ is a set of the form

$$\Lambda = \{\alpha_1 v_1 + \cdots + \alpha_n v_n \mid \alpha_i \in \mathbb{Z}\}$$

where $v_1, \ldots, v_n$ form a basis for the vector space $\mathbb{R}^n$. We then call $v_1, \ldots, v_n$ a *basis* for Λ.

Example 10.9

If $n = 1$ then $\mathbb{R}^n = \mathbb{R}$ and $\Lambda = \{\alpha_1 v_1 \mid \alpha_1 \in \mathbb{Z}\}$ for some non-zero $v_1 \in \mathbb{R}$, so Λ is the subgroup of $\mathbb{R}$ generated by v_1. If $v_1 = 1$ or -1, for instance, we get $\Lambda = \mathbb{Z}$.

Example 10.10

If $n = 2$ and we choose v_1 and v_2 to be the standard basis vectors $(1,0)$ and $(0,1)$ of $\mathbb{R}^2$, then $\Lambda = \{(\alpha_1, \alpha_2) \mid \alpha_1, \alpha_2 \in \mathbb{Z}\}$ is the *square lattice*, or *integer lattice* $\mathbb{Z}^2 \subset \mathbb{R}^2$.

Example 10.11

Similarly, if we choose v_1, v_2 and v_3 to be the standard basis vectors for $\mathbb{R}^3$ then Λ is the *simple cubic lattice* $\mathbb{Z}^3 \subset \mathbb{R}^3$, which plays a major role in crystallography.

Lemma 10.7

If Λ is a lattice in $\mathbb{R}^n$, then Λ is a subgroup of $\mathbb{R}^n$ under addition.

Proof

Let $v_1, \ldots, v_n$ be a basis for Λ. Clearly the zero vector $0 = \sum 0.v_i$ is in Λ. If $v = \sum \alpha_i v_i$ and $w = \sum \beta_i v_i$ are in Λ then $\alpha_i, \beta_i \in \mathbb{Z}$ for all i, so $\alpha_i - \beta_i \in \mathbb{Z}$ and hence $v - w = \sum(\alpha_i - \beta_i)v_i \in \Lambda$. □

Definition

If Λ is a lattice in $\mathbb{R}^n$, then vectors $v, w \in \mathbb{R}^n$ are *equivalent* (modulo Λ), written $v \sim w$, if $v - w \in \Lambda$. It follows from Lemma 10.7 that $\sim$ is an equivalence relation; the equivalence classes are simply the cosets $\Lambda + v = v + \Lambda$ of the subgroup Λ in the group $\mathbb{R}^n$. If $v_1, \ldots, v_n$ is a basis for Λ, we call the set

$$F = \{\alpha_1 v_1 + \cdots + \alpha_n v_n \mid 0 \leq \alpha_i < 1\}$$

a *fundamental region* for Λ; the sets $F + l$ ($l \in \Lambda$) tessellate $\mathbb{R}^n$, that is, they cover $\mathbb{R}^n$ without overlapping. This is equivalent to the following property:

Lemma 10.8

For each $v \in \mathbb{R}^n$ there is a unique $w \in F$ with $v \sim w$.

Proof

Let $v = \sum \alpha_i v_i \in \mathbb{R}^n$, so each $\alpha_i \in \mathbb{R}$. If we define $\beta_i = \alpha_i - \lfloor \alpha_i \rfloor$, the fractional part of α_i, and put $w = \sum \beta_i v_i$, then $w \in F$ since $0 \leq \beta_i < 1$ for all i, and $w = \sum \alpha_i v_i - \sum \lfloor \alpha_i \rfloor v_i = v - l$ with $l = \sum \lfloor \alpha_i \rfloor v_i \in \Lambda$, so $v \sim w$. For the uniqueness of w, suppose that we also have $v \sim w' \in F$. We have $w' = \sum \beta_i' v_i$ where $0 \leq \beta_i' < 1$ for each i, so $|\beta_i - \beta_i'| < 1$. Since v is equivalent to both w and w', we have $w \sim w'$, so $w - w' \in \Lambda$ and hence $\beta_i - \beta_i' \in \mathbb{Z}$. It follows that $\beta_i = \beta_i'$ for all i, so $w = w'$ as required. □

Comment

An alternative interpretation of this result is that each $v \in \mathbb{R}^n$ lies in a set $F + l = \{f + l \mid f \in F\}$ for a unique $l \in \Lambda$ (namely, $l = v - w$, so that $v = w + l \in F + l$). These sets $F + l$ are called *translates* of F, since they are obtained from F by translating F by l. Lemma 10.8 then asserts that these translates tessellate $\mathbb{R}^n$, that is, they cover $\mathbb{R}^n$ without overlapping.

We can use Lemma 10.8 to define a function $\phi : \mathbb{R}^n \to F$ by $\phi(v) = w$, where $v \sim w \in F$; thus w is the unique coset representative for $v + \Lambda$ contained in F. We can apply ϕ to a subset $X \subset \mathbb{R}^n$ by dividing X into regions $X \cap (F + l)$, one in each translate of F, and then translating each region by $-l$ to $(X - l) \cap F \subseteq F$. Note that the images $(X - l) \cap F$ of different regions may overlap if X is sufficiently large (see Lemma 10.9).

To make this last remark more precise we define the n-dimensional volume

of a set $X \subset \mathbb{R}^n$ to be

$$\mathrm{vol}(X) = \iint \dots \int 1 \, dx_1 \, dx_2 \dots dx_n ,$$

provided this exists and is finite, where the integration is over all $(x_1, x_2, \dots, x_n) \in X$. When $n = 1, 2$ or 3 this represents the length, area or volume of X.

Example 10.12

If X is the n-dimensional unit cube, defined by $0 \le x_i \le 1$ for $i = 1, \dots, n$, then $\mathrm{vol}(X) = 1$. Similarly the subset $C \subset X$ defined by $0 \le x_i < 1$, which is a fundamental region for the integer lattice $\Lambda = \mathbb{Z}^n$, also has $\mathrm{vol}(C) = 1$.

An important example we will need is the n-dimensional open ball $B_n(r)$ of radius r, the subset of $\mathbb{R}^n$ defined by $x_1^2 + \dots + x_n^2 < r^2$.

Exercise 10.17

Let $V_n = \mathrm{vol}(B_n(1))$, the volume of the n-dimensional unit ball. By considering cross-sections $x_n = x$ for $-1 \le x \le 1$, show that

$$V_n = \int_{-1}^{1} V_{n-1}(1 - x^2)^{(n-1)/2} dx = 2V_{n-1}I_n$$

for all $n \ge 2$, where $I_n = \int_0^{\pi/2} \sin^n \theta \, d\theta$ satisfies the reduction formula $I_n = (n-1)I_{n-2}/n$. By evaluating V_1, I_0 and I_1 directly, show that

$$V_n = \begin{cases} (2\pi)^m / n(n-2) \dots 4.2 & \text{if } n = 2m \text{ is even,} \\ 2^{m+1}\pi^m / n(n-2) \dots 3.1 & \text{if } n = 2m+1 \text{ is odd.} \end{cases}$$

(For those familiar with the gamma function, one can write this more concisely as $V_n = \pi^{n/2} / \Gamma(\frac{n}{2} + 1)$.)

Exercise 10.18

Deduce that the n-dimensional open ball $B_n(r)$ of radius r has volume $2r, \pi r^2, 4\pi r^3/3$ or $\pi^2 r^4/2$ for $n = 1, 2, 3$ or 4.

Exercise 10.19

By inscribing an octagon in a disc, show that $\pi > 2\sqrt{2}$. (We will need this inequality later.)

Exercise 10.20

Prove that the ellipse $x^2/a^2 + y^2/b^2 = 1$ encloses a set $X \subset \mathbb{R}^2$ (defined by $x^2/a^2 + y^2/b^2 < 1$) satisfying $\mathrm{vol}(X) = \pi ab$.

Lemma 10.9

If $\mathrm{vol}(X) > \mathrm{vol}(F)$ then the restriction $\phi|_X$ of ϕ to X is not one-to-one.

(In other words, if X is sufficiently large then there are at least two distinct equivalent points in X.)

Proof

Because the translates $F+l$ $(l \in \Lambda)$ tessellate $\mathbb{R}^n$, it follows that X is the disjoint union of the subsets $X_l = X \cap (F + l)$ $(l \in \Lambda)$. If $v \in X_l$ then $\phi(v) = v - l$, so ϕ translates X_l to a congruent subset $\phi(X_l) = X_l - l$ of F. Since translations preserve volumes, we have $\mathrm{vol}(\phi(X_l)) = \mathrm{vol}(X_l)$. Now

$$\mathrm{vol}(X) = \sum_{l \in \Lambda} \mathrm{vol}(X_l) = \sum_{l \in \Lambda} \mathrm{vol}(\phi(X_l)) .$$

If $\phi|_X$ is one-to-one then the translates $\phi(X_l)$ cannot overlap, so

$$\sum_{l \in \Lambda} \mathrm{vol}(\phi(X_l)) \leq \mathrm{vol}(F)$$

and hence $\mathrm{vol}(X) \leq \mathrm{vol}(F)$, against our assumption. □

Definition

A subset X of $\mathbb{R}^n$ is *centrally symmetric* if, whenever $v \in X$, we also have $-v \in X$; it is *convex* if, whenever $v, w \in X$, the line-segment vw also lies in X, that is, $tv + (1-t)w \in X$ for all t such that $0 \leq t \leq 1$.

Example 10.13

$B_n(r)$ is centrally symmetric, since if $x_1^2 + \cdots + x_n^2 < r^2$ then $(-x_1)^2 + \cdots + (-x_n)^2 < r^2$. Similarly, the region $x^2/a^2 + y^2/b^2 < 1$ bounded by an ellipse is centrally symmetric.

Exercise 10.21

Show that $B_n(r)$ is convex.

The following theorem, proved by Minkowski around 1890, has some far-reaching consequences.

Theorem 10.10

Let Λ be a lattice in $\mathbb{R}^n$ with fundamental region F, and let X be a centrally symmetric convex set in $\mathbb{R}^n$ with $\mathrm{vol}(X) > 2^n \mathrm{vol}(F)$. Then X contains a non-zero lattice-point of Λ.

Proof

The lattice $2\Lambda = \{2v \mid v \in \Lambda\}$ has a fundamental region $2F = \{2v \mid v \in F\}$ with $\mathrm{vol}(2F) = 2^n \mathrm{vol}(F)$. Thus $\mathrm{vol}(X) > \mathrm{vol}(2F)$, so by applying Lemma 10.9 to X and 2Λ we see that there exist $v \neq w$ in X with $v - w \in 2\Lambda$. Since $w \in X$ and X is centrally symmetric, we have $-w \in X$. Since X is convex and $v, -w \in X$, the midpoint $\frac{1}{2}(v - w)$ of the line-segment from v to $-w$ is also in X. Now $v - w \in 2\Lambda$, so $\frac{1}{2}(v - w) \in \Lambda$, giving the required non-zero lattice-point in X. □

Example 10.14

The lattice $\Lambda = \mathbb{Z}^n$ has a fundamental region F (such as the unit cube in $\mathbb{R}^n$) with $\mathrm{vol}(F) = 1$. The set $X = \{\sum \alpha_i v_i \mid |\alpha_i| < 1\}$ is centrally symmetric and convex, with $\mathrm{vol}(X) = 2^n = 2^n \mathrm{vol}(F)$, but X contains no non-zero lattice-points. This shows that Minkowski's Theorem fails if we relax the lower bound on $\mathrm{vol}(X)$.

Exercise 10.22

By finding suitable counterexamples, show that Minkowski's Theorem fails if either of the conditions 'centrally symmetric' or 'convex' is omitted.

In order to apply Minkowski's Theorem one needs to be able to calculate volumes of fundamental regions. This is easily done using determinants. Suppose that Λ has a basis $\{v_1, \ldots, v_n\}$, where each $v_i = (\alpha_{i1}, \ldots, \alpha_{in}) \in \mathbb{R}^n$. If A is the $n \times n$ matrix (α_{ij}) formed from these vectors v_i (as row or column vectors), then

$$\mathrm{vol}(F) = |\det(A)|.$$

This is because the linear transformation $\mathbb{R}^n \to \mathbb{R}^n$ induced by A sends the standard basis vectors e_i of $\mathbb{R}^n$ to the basis vectors v_i of Λ, and hence sends

the set $C = \{\alpha_1 e_1 + \cdots + \alpha_n e_n \mid 0 \leq \alpha_i < 1\}$ to the fundamental region $F = \{\alpha_1 v_1 + \cdots + \alpha_n v_n \mid 0 \leq \alpha_i < 1\}$ for Λ. Since $\mathrm{vol}(C) = 1$, and since any linear transformation A multiplies volumes by $|\det(A)|$, it follows that $\mathrm{vol}(F) = |\det(A)|$.

We now apply these ideas to give another proof of Theorem 10.2, the *Two Squares Theorem*, which we state again:

Theorem 10.2

Each prime $p \equiv 1 \bmod (4)$ is a sum of two squares.

Proof

Since $p \equiv 1 \bmod (4)$, Corollary 7.7 implies that $u^2 \equiv -1 \bmod (p)$ for some integer u. Now suppose that the following condition is true:

$$\text{there exist} \quad x, y \in \mathbb{Z} \quad \text{with} \quad y \equiv ux \bmod (p) \quad \text{and} \quad 0 < x^2 + y^2 < 2p. \tag{10.9}$$

Then $x^2 + y^2 \equiv x^2 + u^2x^2 \equiv x^2 - x^2 \equiv 0 \bmod (p)$, so $x^2 + y^2 = kp$ for some integer k. The inequalities in (10.9) become $0 < kp < 2p$, so $k = 1$ and $x^2 + y^2 = p$, as required. It is therefore sufficient to prove (10.9). We can do this using Minkowski's Theorem, since the first condition $y \equiv ux \bmod (p)$ in (10.9) defines a lattice Λ in $\mathbb{R}^2$ (as we shall prove below), the condition $x^2 + y^2 < 2p$ defines a centrally symmetric convex set X, namely the disc $B_2(\sqrt{2p})$, and the condition $0 < x^2 + y^2$ specifies a non-zero point (x, y); Minkowski's Theorem then guarantees the existence of a point (x, y) satisfying all three of these conditions, so (10.9) is proved.

To justify this, we must verify all the hypotheses in Minkowski's Theorem. First let

$$\Lambda = \{(x, y) \in \mathbb{Z}^2 \mid y \equiv ux \bmod (p)\}.$$

It is easily checked that this is a subgroup of $\mathbb{R}^2$ containing the linearly independent vectors $v_1 = (1, u)$ and $v_2 = (0, p)$. If $(x, y) \in \Lambda$ then let $\alpha_1 = x$ and $\alpha_2 = (y - ux)/p$; these are integers, with $\alpha_1 v_1 + \alpha_2 v_2 = (x, y)$, so Λ is generated by v_1 and v_2. Thus Λ is a lattice with v_1 and v_2 forming a basis, so it has a fundamental region F with 2-dimensional volume (or area)

$$\mathrm{vol}(F) = \left|\det \begin{pmatrix} 1 & u \\ 0 & p \end{pmatrix}\right| = p.$$

Now let $X = B_2(\sqrt{2p}) = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 2p\}$, an open disc of radius $r = \sqrt{2p}$ centred at the origin. This is centrally symmetric and convex, and Exercise 10.18 gives $\mathrm{vol}(X) = \pi r^2 = 2\pi p$. Now $\pi > 2\sqrt{2} > 2$ by Exercise

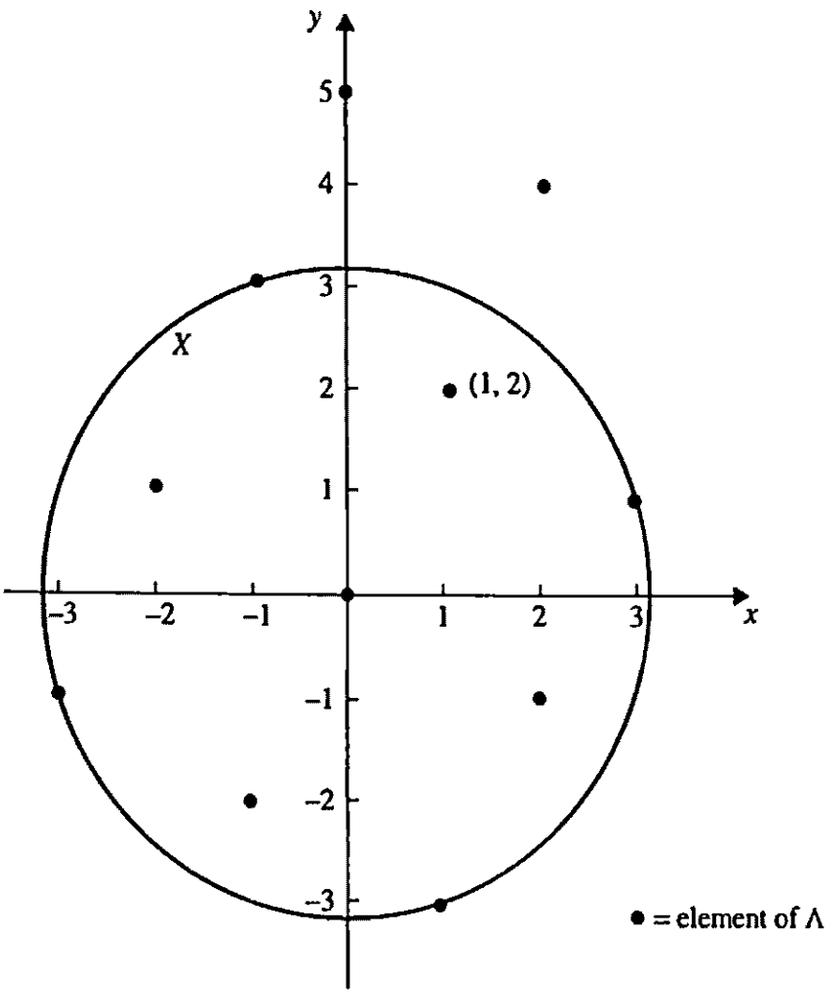

Figure 10.1. The proof of Theorem 10.2 for $p = 5$, with $u = 2$.

10.19, so $\text{vol}(X) > 2^2\text{vol}(F)$, and hence Minkowski's Theorem gives a non-zero lattice point $(x, y) \in X \cap \Lambda$. (Figure 10.1 illustrates this in the case $p = 5$, with $u = 2$ and $(x, y) = (1, 2)$.) We now have a pair of integers x and y satisfying (10.9), so the proof is complete. □

We can also use this method to prove Theorem 10.6, the Four Squares Theorem.

Theorem 10.6

Every non-negative integer is a sum of four squares.

Proof

As in our earlier proof of Theorem 10.6, in Section 4, it is sufficient to prove that every odd prime p is a sum of four squares. First we show (as before) that there exist integers u, v satisfying $u^2 + v^2 \equiv -1 \mod (p)$. For such a pair u, v

let

$$\Lambda = \{(x,y,z,t) \in \mathbb{Z}^4 \mid z \equiv ux + vy \text{ and } t \equiv vx - uy \mod (p)\}.$$

Exercise 10.23

Show that Λ is a lattice in $\mathbb{R}^4$ with basis

$$v_1 = (1,0,u,v), \quad v_2 = (0,1,v,-u), \quad v_3 = (0,0,p,0), \quad v_4 = (0,0,0,p).$$

Continuing the proof, we deduce from Exercise 10.23 that a fundamental region F for Λ has volume

$$\text{vol}(F) = \left|\det \begin{pmatrix} 1 & 0 & u & v \\ 0 & 1 & v & -u \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}\right| = p^2.$$

Now let $X = B_4(\sqrt{2p}) = \{(x,y,z,t) \in \mathbb{R}^4 \mid x^2+y^2+z^2+t^2 < 2p\}$, an open ball of radius $r = \sqrt{2p}$. This is centrally symmetric and convex, with 4-dimensional volume

$$\text{vol}(X) = \frac{\pi^2 r^4}{2} = 2\pi^2 p^2$$

by Exercise 10.18. Now Exercise 10.19 gives $\pi^2 > 8$, so $\text{vol}(X) > 2^4 \text{vol}(F)$ and hence Minkowski's Theorem implies that X contains a non-zero lattice point. Thus there exist integers x,y,z,t such that $0 < x^2+y^2+z^2+t^2 < 2p$ and $z \equiv ux+vy$, $t \equiv vx-uy \mod (p)$. Then

$$\begin{aligned} x^2+y^2+z^2+t^2 &\equiv x^2+y^2+u^2x^2+2uvxy+v^2y^2+v^2x^2-2uvxy+u^2y^2 \\ &\equiv (1+u^2+v^2)(x^2+y^2) \\ &\equiv 0 \mod (p) \end{aligned}$$

since $u^2+v^2 \equiv -1 \mod (p)$, so $x^2+y^2+z^2+t^2 = p$ and the proof is complete. □

10.7 Supplementary exercises

Exercise 10.24

Show that an odd prime p can be written in the form $2x^2+y^2$ if and only if $-2 \in Q_p$, or equivalently $p \equiv 1$ or $3 \mod (8)$. (Hint: apply Minkowski's Theorem, with X the interior of an ellipse.)

Exercise 10.25

Show that a prime p can be written in the form $x^2 + xy + y^2$ if and only if $p = 3$ or $p \equiv 1 \bmod (3)$.

Exercise 10.26

Show that the number $r(n)$ of representions of n as a sum of two squares has average value π, that is,

$$\frac{1}{n}\sum_{m=1}^{n} r(m) \to \pi \quad \text{as} \quad n \to \infty .$$

(Hint: representations $m = x^2 + y^2$ of integers $m \leq n$ correspond to integer lattice points (x, y) within distance $\sqrt{n}$ of the origin.) What can be said about the average number of representations of n as a sum of k squares?

Exercise 10.27

Show that $\sum_{n=1}^{\infty} r(n)/n^s = 4L(s)\zeta(s)$ for all $s > 1$, where $L(s) = 1^{-s} - 3^{-s} + 5^{-s} - \cdots$.