7 Quadratic Integers

In the last chapter, we used unique factorization in $\mathbb{Z}[i]$ to study sums of squares. Here we give a study what happens in the rings $\mathbb{Z}[\sqrt{-2}]$ and $\mathbb{Z}[\sqrt{-3}]$. We will go through these cases fairly quickly, and not give all of the details, as we will soon look at the general case. The sections in the notes no longer match up with the sections in the text for this chapter.

First, for the record we define for $n \in \mathbb{N}$, the *imaginary quadratic rings*

$$\mathbb{Z}[\sqrt{-n}] = \left\{a + b\sqrt{-n} | a, b \in \mathbb{Z}\right\}$$

These are all rings, in the sense that they contain $\mathbb{Z}$ and are closed under addition and multiplication.

Fix n. For $\alpha = a + b\sqrt{-n}$, the **conjugate** is $\overline{\alpha} = a - b\sqrt{-n}$ and the **norm** is $N(\alpha) = \alpha\overline{\alpha} = a^2 + nb^2$. Note that $\alpha = a + ib\sqrt{n}$ and $\overline{\alpha} = a - ib\sqrt{n}$, so conjugation in $\mathbb{Z}[\sqrt{-n}]$ coincides with complex conjugation. This also means $N(\alpha) = |\alpha|^2$.

Note the norm is never negative. We say $u \in \mathbb{Z}[\sqrt{-n}]$ is a **unit** of $\mathbb{Z}[\sqrt{-n}]$ if u is invertible in $\mathbb{Z}[\sqrt{-n}]$, which is equivalent to $N(u) = 1$. As before we have the properties $\overline{\alpha\beta} = \overline{\alpha}\overline{\beta}$ (ordinary complex conjugation) and $N(\alpha\beta) = N(\alpha)N(\beta)$.

For $\alpha, \beta \in \mathbb{Z}[\sqrt{-n}]$, we say β is a **divisor** of α, and write $\beta|\alpha$, if $\alpha = \beta\gamma$ for some $\gamma \in \mathbb{Z}[\sqrt{-n}]$. In rings without unique factorization, we do not usually use the term *prime* as we have been (or it is defined differently). Instead, we talk about *irreducibles* (which include the units). Let's start using this terminology now. We say α is **irreducible** if the only divisors of α are u and $u\alpha$ where u is a unit, i.e., if α has no non-trivial factorizations.

Example. *The irreducible elements of $\mathbb{Z}$ are ± 1 and $\pm p$ for p prime.*

Example. *The irreducible elements of $\mathbb{Z}[i]$ are the units and the primes of $\mathbb{Z}[i]$.*

Proposition 7.1. *Any non-zero $\alpha \in \mathbb{Z}[\sqrt{-n}]$ factors into a product of irreducibles.*

The proof is identical to the proof of existence of prime factorzation in $\mathbb{Z}$ and $\mathbb{Z}[i]$.

7.1 Factorization in $\mathbb{Z}[\sqrt{-2}]$ and $\mathbb{Z}[\sqrt{-3}]$

Definition 7.2. *We say $\mathbb{Z}[\sqrt{-n}]$ has* **unique factorization** *if for any non-zero, non-unit $\alpha \in \mathbb{Z}[\sqrt{-n}]$ and any two irreducible factorizations*

$$\alpha = \prod \pi_i = \prod \pi'_j$$

(where π_i, π'_j are irreducible non-units) we have, up to relabeling, $\pi_i = u_i\pi'_i$ for some unit u_i.

This coincides with our notions of unique factorization in the cases $\mathbb{Z}$ and $\mathbb{Z}[i]$. Let us recall our approach for proving unique factorization in these cases.

In $\mathbb{Z}$ (or rather $\mathbb{N}$), we first observed that $\gcd(a, b) = \gcd(b, a - b)$. Then we used the Euclidean algorithm to show that $\gcd(a, b) = ma + nb$ for some $m, n \in \mathbb{Z}$. From this, we got the prime divisor property ($p|ab \implies p|a$ or $p|b$), and unique factorization follows formally.

In $\mathbb{Z}[i]$, we first demonstrated the **division property**: for any $\alpha, \beta \in \mathbb{Z}[i]$ and non-zero, there exist $\mu, \rho \in \mathbb{Z}[i]$ such that $\alpha = \mu\beta + \rho$ where $N(\rho) < N(\beta)$. This gave us a Euclidean algorithm

for computing some (non-unique) $\gcd(\alpha, \beta)$ in $\mathbb{Z}[i]$. This implies any $\gcd(\alpha, \beta) = \mu\alpha + \nu\beta$ for some $\mu, \nu \in \mathbb{Z}[i]$. Then we concluded the prime divisor property and unique factorization as before.

The key step (meaning the only one that can fail) is the division property. The idea of the proof of the division property was that $L = \{\mu\beta : \mu \in \mathbb{Z}[i]\}$ formed a square lattice in the complex plane, so the furthest α could be from a corner of L was $\frac{|\beta|}{\sqrt{2}}$.

Proposition 7.3. *Let $\alpha, \beta \in \mathbb{Z}[\sqrt{-2}]$ be non-zero. Then there exist $\mu, \rho \in \mathbb{Z}[\sqrt{-2}]$ such that*

$$\alpha = \mu\beta + \rho, \quad |\rho| < |\beta|$$

or equivalently,

$$\alpha = \mu\beta + \rho, \quad N(\rho) < N(\beta).$$

Proof. The multiples of β are

$$L = \left\{ m\beta + n\sqrt{-2}\beta = m\beta + ni\sqrt{2}\beta \right\},$$

in other words, the lattice in $\mathbb{C}$ generated by β and $i\sqrt{2}\beta$. Recall multiplication by i rotates 90 degrees, so $i\sqrt{2}\beta$ is the vector obtained by rotating β 90 degrees and scaling by $\sqrt{2}$. Hence L is a rectangular lattice with side lengths $|\beta|$ and $\sqrt{2}|\beta|$. The distance from the midpoint of one of these rectangles to the corner is

$$\sqrt{\left(\frac{|\beta|}{2}\right)^2 + \left(\frac{\sqrt{2}|\beta|}{2}\right)^2} = \sqrt{\frac{|\beta|^2 + 2|\beta|^2}{4}} = \frac{\sqrt{3}}{2}|\beta|$$

Thus the closest lattice point (or corner) $\mu\beta$ to α is at most distance $\frac{\sqrt{3}}{2}|\beta|$ away. Letting $\rho = \alpha - \mu\beta$, we see $|\rho| = \frac{\sqrt{3}}{2}|\beta| < |\beta|$. □

Corollary 7.4. *$\mathbb{Z}[\sqrt{-2}]$ has unique factorization.*

This follows in the same way. We omit the details.

If we try to do the same thing in $\mathbb{Z}[\sqrt{-3}]$, we have a rectangular lattice and the rectangles have dimension $|\beta| \times \sqrt{3}|\beta|$. But then the midpoint of one such rectangle has distance

$$\sqrt{\left(\frac{|\beta|}{2}\right)^2 + \left(\frac{\sqrt{3}|\beta|}{2}\right)^2} = \sqrt{\frac{|\beta|^2 + 3|\beta|^2}{2}} = |\beta|$$

from the corners, so we don't have a similar division property in $\mathbb{Z}[\sqrt{-3}]$. This suggests in fact

Proposition 7.5. *$\mathbb{Z}[\sqrt{-3}]$ does not have unique factorization.*

Proof.

$$4 = 2 \cdot 2 = (1 + \sqrt{-3})(1 - \sqrt{-3}).$$

Further 2, $1 + \sqrt{-3}$ and $1 - \sqrt{-3}$ are all irreducible in $\mathbb{Z}[\sqrt{-3}]$ because they all have norm 4, and $\mathbb{Z}[\sqrt{-3}]$ has no elements of norm 2 (since $2 \neq x^2 + 3y^2$). □

However, an amazing thing happens. Let

$$\zeta_3 = e^{2\pi i/3} = \cos(2\pi/3) + i\sin(2\pi/3) = -\frac{1}{2} + i\frac{\sqrt{3}}{2} = \frac{-1+\sqrt{-3}}{2}.$$

If we extend the ring $\mathbb{Z}[\sqrt{-3}]$ to include ζ_3, namely set

$$\mathbb{Z}[\zeta_3] = \{a + b\zeta_3 | a, b \in \mathbb{Z}\}.$$

Just like $\mathbb{Z}[\sqrt{-n}]$ this is a ring. The two things to check are that it's closed under addition, which is obvious, and closed under multiplication.

Exercise 7.1. *Show $\zeta_3^2 + \zeta_3 + 1 = 0$. Using this, deduce that $\mathbb{Z}[\zeta_3]$ is closed under multiplication. (Note the set of integer linear combinations of 1 and α, $\{a + b \cdot \alpha | a, b \in \mathbb{Z}\}$, is not always closed under multiplication. For example when $\alpha = \sqrt[3]{2}$ or $\alpha = e^{2\pi i/5}$, α^2 is not a linear combination of 1 and α.)*

Then we get the following.

Proposition 7.6. *Let $\alpha, \beta \in \mathbb{Z}[\zeta_3]$ be non-zero. Then there exist $\mu, \rho \in \mathbb{Z}[\zeta_3]$ such that*

$$\alpha = \mu\beta + \rho, \ \ |\rho| < |\beta|,$$

or equivalently,

$$\alpha = \mu\beta + \rho, \ \ N(\rho) < N(\beta).$$

Proof. Note that multiplication by $\zeta_3 = e^{2\pi i/3}$ in $\mathbb{C}$ is simply rotation by 120 degrees. Thus the multiples $\mu\beta$ form the lattice

$$L = \{m\beta + n\zeta_3\beta\}.$$

This divides the complex plane not into rectangles, but parallograms. In fact, into equilateral triangles, each of side length $|\beta|$. Now α lies in (or on the border of) one of these triangles, and the distance to the closest corner $\mu\beta$ must be less than $|\beta|$. □

Exercise 7.2. *For $\alpha \in \mathbb{Z}[\zeta_3]$, let $\overline{\alpha}$ be the complex conjugate of α, and define the norm by $N(\alpha) = |\alpha|^2 = \alpha\overline{\alpha}$. (The norm of $a + b\zeta_3$ is* not *$(a + b\zeta_3)(a - b\zeta_3)$.) Show $\overline{\alpha} \in \mathbb{Z}[\zeta_3]$ for any $\alpha \in \mathbb{Z}[\zeta_3]$. Compute $\overline{\zeta}_3, N(\zeta_3)$ and $N(1 + \zeta_3)$. Write down a formula for $N(a + b\zeta_3)$ where $a, b \in \mathbb{Z}$.*

Exercise 7.3. *Determine the units of $\mathbb{Z}[\zeta_3]$ (the elements of norm 1).*

One also defines units and irreducibles for $\mathbb{Z}[\zeta_3]$ in the same way. As before, the division property yields

Corollary 7.7. *$\mathbb{Z}[\zeta_3]$ has unique factorization.*

In the general case, $\mathbb{Z}[\sqrt{-n}]$ typically does not have unique factorization (it happens only finitely many times). Nevertheless, one can always enlarge the ring $\mathbb{Z}[\sqrt{-n}]$ to some slightly bigger ring R such that any element of $\mathbb{Z}[\sqrt{-n}]$ factors uniquely into irreducibles in R; however, R itself may not have unique factorization. This route essentially was what Kummer pursued (and had great success); however Dedekind's ideal theory provides a much simpler approach. This is what we will aim for by the end of the course.

Exercise 7.4. *Exercise 7.4.1. This resolves the non-unique factorization of $4 = 2 \cdot 2 = (1+\sqrt{-3})(1-\sqrt{-3})$ in $\mathbb{Z}[\sqrt{-3}]$ by going to $\mathbb{Z}[\zeta_3]$.*

7.2 Applications

Here we give a couple examples of the power of unique factorizations. There are infinitely many solutions to $y^3 = x^2+2$ in $\mathbb{Q}$, which was known to Diophantus. However Fermat claimed, and Euler mostly proved,

Proposition 7.8. *The only solution to $y^3 = x^2 + 2$ in $\mathbb{N}$ is $(x, y) = (5, 3)$.*

Proof. (Sketch) Suppose

$$y^3 = x^2 + 2 = (x + \sqrt{-2})(x - \sqrt{-2}).$$

It is not too hard to show that $x + \sqrt{-2}$ and $x - \sqrt{-2}$ are relatively prime in $\mathbb{Z}[\sqrt{-2}]$. Since their product is a cube, then both $x + \sqrt{-2}$ and $x - \sqrt{-2}$ are cubes. Write

$$x + \sqrt{-2} = (a + b\sqrt{-2})^3 = a^2 - 6ab^2 + (3a^2b - 2b^3)\sqrt{-2}.$$

Hence

$$x = a^3 - 6ab^2, \;\; 1 = b(3a^2 - 2b^2).$$

From the second equation, we have $b = \pm 1$ and $3a^2 - 2b^2 = 3a^2 - 2 = 1$, so $a = \pm 1$ and $x = \pm 5$. □

Theorem 7.9. *$x^3 + y^3 = z^3$ has no solutions in $\mathbb{N}$.*

Proof. (Sketch) Suppose we have a solution $z^3 = x^3 + y^3$. Dividing by any common factors, we may assume x, y, z are (pairwise) relatively prime. Note

$$z^3 = x^3 + y^3 = (x + y)(x^2 - xy + y^2) = (x + y)(x + \zeta_3 y)(x + \zeta_3^2 y).$$

One can show that $x+y$, $x+\zeta_3 y$ and $x+\zeta_3^2 y$ are relatively prime in $\mathbb{Z}[\zeta_3]$. Then unique factorization implies that, up to units, they are all cubes in $\mathbb{Z}[\zeta_3]$. In fact, one may assume that $x+\zeta_3 y$ is a cube (multiplying by units just reorders these three factors and interchanges x and y), i.e.,

$$x + \zeta_3 y = (a + b\zeta_3)^3 = a^3 + b^3 - 3ab^2 + 3ab(a - b)\zeta_3.$$

Hence $y = 3cd(c - d) \equiv 0 \bmod 3$. Thus we have

$$z^3 - x^3 = (z - x)(z - x\zeta_3)(z - x\zeta_3^2)$$

is divisible by 3. As before, $z - x, z - x\zeta_3, z - x\zeta_3^2$ are relatively prime in $\mathbb{Z}[\zeta_3]$.

The element $\pi = \sqrt{-3}$ is a prime in $\mathbb{Z}[\zeta_3]$ (it's norm is 3, which is prime in $\mathbb{Z}$). We say $\alpha \equiv \beta \bmod \pi$ if $\pi|(\alpha - \beta)$. This is well defined.

By unique factorization, π must divide one of $z - x, z - x\zeta_3, z - x\zeta_3^2$. But in fact one can show

$$z - x \equiv z - x\zeta_3 \equiv z - x\zeta_3^2 \bmod \pi,$$

hence π divides all three factors, contradicting relative primeness. □

Exercise 7.5. *What are the possible remainders mod $\pi = \sqrt{-3}$ in $\mathbb{Z}[\zeta_3]$? (Hint: they will be the elements whose norm is less than that of π.) Show that for any $z, x \in \mathbb{Z}$ (or even $\mathbb{Z}[\zeta_3]$) $z - x \equiv z - x\zeta_3 \equiv z - x\zeta_3^2$, which we need in the proof of Fermat's Last Theorem for $n = 3$. (If you need at hint, look at p. 131.)*

Lamé thought he could generalize this kind of proof and show Fermat's Last Theorem in general, but his proof failed due to failure of unique factorization in the *cyclotomic integers* $\mathbb{Z}[\zeta_n]$ for most n. Nevertheless, Kummer developed a theory to try to recover unique factorization by going to larger rings, as mentioned above, and used this to prove Fermat's Last Theorem in a large number of cases. If people show interest, I will try to say more about this later.

Exercise 7.6. *Fermat's Last Theorem says $x^n + y^n = z^n$ has no solutions in $\mathbb{N}$ if $n > 2$. Show that if $d|n$ then a solution to $x^n + y^n = z^n$ give a solution to $x^d + y^d = z^d$. Deduce that Fermat's Last Theorem is true for $n \equiv 0 \bmod 3$. Also deduce that to prove Fermat's Last Theorem, it suffices to prove it for $n = 4$ and $n = p$ where p is any odd prime.*

7.3 Discussion

For more details, see the text. The discussion section is worth reading. Also Section 7.5, which I didn't touch on, is interesting.

8 The four square theorem

8.1 Some theorems

After having answered a question, which numbers are sums of two squares, it is natural to ask which numbers are sums of three squares or four squares.

Surprisingly, case of 3 squares turns out to be harder than 4 squares! Using the theory of quadratic forms (largely developed by him), the great genius Gauss showed

Theorem 8.1. *(Legendre–Gauss, 1801) $n \in \mathbb{N}$ is a sum of three* integral *squares if and only if n is not of the form $4^a(8m+7)$.*

He also determined the number of ways n is a sum of three cubes. We will not go into that here, but may revisit it next semester when we cover binary quadratic forms if there is interest. In any case, this was considerably more different than the following result, which seems to have been observed without proof by Diophantus.

Theorem 8.2. *(Diophantus–Lagrange, 1770) Every $n \in \mathbb{N}$ is a sum of four* integral *squares.*

Note that Lagrange's theorem does not say every n is a sum of exactly 4 positive squares. For example, $1 = 1^2 + 0^2 + 0^2 + 0^2$. Therefore, a common way to state Lagrange's theorem is that every number is a sum of at most 4 squares. Before we talk about how to prove Lagrange's theorem, let me mention a nice generalization observed by Fermat.

I can think of n^2 as the number of points square lattice, with side length n. We can also make an equilateral triangular lattice, of side length n. In the top row there is 1 point, in the second 2 and so on. Hence there are $1 + 2 + \cdots + n = \frac{n(n+1)}{2}$ total points. Thus we call the numbers $\frac{n(n+1}{)}2$ triangular numbers.

Theorem 8.3. *(Fermat–Gauss, 1796) Every $n \in \mathbb{N}$ is a sum of at most 3 triangular numbers.*

More generally, one can define k-gonal numbers as the number of points one can get from regular k-gonal lattices.

Theorem 8.4. *(Fermat–Cauchy, 1813) Every $n \in \mathbb{N}$ is a sum of at most k k-gonal numbers.*

Legendre showed that in fact 4 or 5 k-gonal numbers suffice depending on whether n is odd or even. (Cf. L.E. Dickson's, *History of the theory of numbers.*) In the case of three triangular numbers or four squares, one knows formulas for the number of ways n can be written in these forms, but—if I remember correctly—formulas for general k are not known.

Another related question, still unsolved, is Waring's problem. In 1909, Hilbert show for every $k \in \mathbb{N}$, there is a smallest number $g(k)$ such that every $n \in \mathbb{N}$ is a sum of at most $g(k)$ k-th powers. Lagrange's theorem says $g(2) = 4$. Further $g(3) = 9$ and and $g(4) = 19$. What is $g(k)$? (Rather the more interesting question is, what is the smallest number $G(k)$ of k-th powers needed to represent every sufficiently large integer. For example $G(4) = 16$. I.e., *almost* every integer is a sum of 16 4-th powers, but there are a finite number of special ones which require up to 19.)

8.2 Three proofs

There are several different proofs of Lagrange's four square theorem. One can deduce it from either Gauss's theorem on the sum of 3 squares or Gauss's theorem on the sum of 3 triangular numbers

(cf. Grosswald's *Representations of Integers as Sums of Squares*), but both of Gauss's theorems are harder to prove.

Below we briefly discuss three relatively simple ways one can prove Lagrange's four square theorem. In all cases, we will just sketch the proofs, leaving out the more tedious details. The first two proofs both use the following two lemmas.

Lemma 8.5. *Suppose n and m are sums of four squares. Then so is mn.*

Hence it suffice to prove every odd prime p is a sum of four squares ($p = 2$ clearly is $1^2 + 1^2 + 0^2 + 0^2$).

Exercise 8.1. *Let p be an odd prime. We say a is a* square *or* quadratic residue *mod p if $a \equiv x^2 \bmod p$ for some x. Prove there are $\frac{p+1}{2}$ distinct squares mod p.*

Exercise 8.2. *Let p be an odd prime. Show $x^2 + y^2 \equiv -1 \bmod p$ for some $x, y \in \mathbb{Z}$. (Hint: use the previous exercise and the pigeonhole principle.)*

Note this exercise looks similar to Lagrange's lemma from Chapter 6. In fact it was also proved by Lagrange, and we restate it in the following form.

Lemma 8.6. *(Lagrange) Let p be an odd prime. Then p divides $x^2 + y^2 + 1$ for some $x, y \in \mathbb{Z}$.*

8.2.1 The algebraic proof

One idea, which is the proof given in the text, is to generalize the proof of Fermat's two square theorem using $\mathbb{Z}[i]$. Hence we want a ring R and a norm map $N : R \to \mathbb{Z}$ which looks like $a^2 + b^2 + c^2 + d^2$. Answer: the quaternions.

Let

$$\mathbb{H} = \mathbb{R}[i, j, k] = \{a + bi + cj + dk | a, b, c, d \in \mathbb{R}\},$$

where

$$i^2 = j^2 = k^2 = ijk = -1.$$

This is a *non-commutative* division algebra, known as Hamilton's quaternions. It is some sort of generalization of the complex numbers. Just as $\mathbb{C}$ is 2-dimensional over $\mathbb{R}$, this is 2-dimensional over $\mathbb{C}$, or 4-dimensional over $\mathbb{R}$ (as a vector space). $\mathbb{H}$ has many applications, and is closely related to the algebra of 2×2 real matrices. Note that going from $\mathbb{R}$ to $\mathbb{C}$, one loses the natural well-ordering one had on $\mathbb{R}$, and going from $\mathbb{C}$ to $\mathbb{H}$ one loses commutativity. Incidentally one can extend the quaterions to the octonions $\mathbb{O}$, but then one loses associativity. Nevertheless, the octonions also have various applications.

The norm on $\mathbb{H}$ is given by

$$N(a + bi + cj + dk) = a^2 + b^2 + c^2 + d^2.$$

The integers of $\mathbb{H}$, also called the Hurwitz integers are

$$R = \mathbb{Z}[\frac{1 + i + j + k}{2}, i, j, k] = \left\{ \frac{a + bi + cj + dk}{2} | a, b, c, d \in \mathbb{Z}, a \equiv b \equiv c \equiv d \bmod 2 \right\}.$$

Then

$$N : R \to \mathbb{Z}.$$

Theorem 8.7. *R has unique factorization.*

Lemma 8.8. *n is a sum of four squares if and only if $n = N(\alpha)$ for some $\alpha \in R$.*

Note that Lemma 8.5 is a simple consequence of this lemma and the fact that the norm is multiplicative.

Corollary 8.9. *p is a sum of four squares if and only if p is not prime (irreducible) in R.*

So we want to show no prime p of $\mathbb{N}$ is prime in R. Let p be an odd prime. Then by Lagrange's Lemma 8.6 we have, for some $x, y \in \mathbb{Z}$,

$$p|1 + x^2 + y^2 = (1 + xi + yj)(1 - xi - xj).$$

Since

$$\frac{1 \pm xi \pm yj}{p} \notin R$$

p does not divide either factor on the right. Hence by unique factorization/prime divisor property, p is not prime in R.

8.2.2 The geometric proof

This proof is taken from *Algebraic Number Theory and Fermat's Last Theorem* by Stewart and Tall.

Suppose you have a sublattice L of $\mathbb{Z}^2$ generated by two elements $(a, 0)$ and $(0, b)$, i.e., $L = \{am + bn | m, n \in \mathbb{Z}\}$. This provides a tiling of $\mathbb{R}^2$ by rectangles, each of area ab. A theorem of Minkowski says that if X is a bounded symmetric ($X = -X$) domain of $\mathbb{R}^2$ and has volume greater than $4ab$, it contains a non-zero point of L.

Minkowski's theorem applies in any dimension, and we consider it in dimension 4. Specifically, let p be an odd prime, and let $x, y \in \mathbb{Z}$ such that $p|x^2 + y^2 + 1$. Consider the lattice $L \subset \mathbb{Z}^4$ given by

$$L = \{(a, b, c, d) | a, b, c, d \in \mathbb{Z}, \;\; c \equiv xa + yb \bmod p, \;\; d \equiv xb - ya \bmod p\}.$$

The point is then that each lattice point satisfies

$$a^2 + b^2 + c^2 + d^2 \equiv a^2 + b^2 + x^2a^2 + y^2b^2 + x^2b^2 + y^2a^2 \equiv a^2(1 + x^2 + y^2) + b^2(1 + x^2 + y^2) \equiv 0 \bmod p.$$

The lattice L tiles $\mathbb{R}^4$ into 4-dimensional rectangles, each of volume p^2. Let X be a 4-dimenional sphere of radius r, which has volume $\pi^2 r^4/2$. If we choose r such that $r^2 = 1.9p$, then the volume of X is greater than $16p^2$, which by Minkowski's theorem say X contains a non-zero lattice point (a, b, c, d). This means

$$a^2 + b^2 + c^2 + d^2 \leq r^2 = 1.9p < 2p.$$

On the other hand $a^2 + b^2 + c^2 + d^2 \equiv 0 \bmod p$ means $a^2 + b^2 + c^2 + d^2 = p$.

One can do a similar proof to show any prime of the form $4n + 1$ is a sum of two squares.

8.2.3 The analytic proof

This proof is basically due to Jacobi, but recast in the language of *modular forms*. Let $r_4(n)$ denote the number of solutions to $n = a^2 + b^2 + c^2 + d^2$ in $\mathbb{Z}$. Consider the function

$$\theta(z) = \sum_{n=-\infty}^{\infty} z^{n^2}.$$

Then

$$\Theta(z) = \theta(z)^4 = \sum a_n z^n = 1 + \sum_{n=1}^{\infty} r_4(n) z^n = 1 + 8z + 12z^2 + \cdots$$

To see this, first we know that $\Theta(z)$ has a power series expansion because $\theta(z)$ does. Now what is the coefficient a_n of the n-th term? We just need to count the number of ways z^n is a product $z^{a^2} z^{b^2} z^{c^2} z^{d^2}$, i.e, it is $r_4(n)$ for $n > 0$.

One shows that $\Theta(z)$ satisfies certain transformation properties, which means that it is what is called a *modular form* of weight 2 on $\Gamma_0(4)$. However this is a two dimensional space of functions generated by

$$F(z) = \frac{1}{24} + \sum_{n \text{ odd}} \sigma(n) z^n + \sum_{2n} (\sigma(2n) - 2\sigma(n)) z^{2n} = \frac{1}{24} + z + z^2 + 4z^3 + z^4 + \cdots$$

and

$$G(z) = \frac{1}{24} + \sum_{n \text{ odd}} \sigma(2n) z^{2n} + \sum_{2n} (\sigma(4n) - 2\sigma(2n)) z^{4n} = \frac{1}{24} + z + z^4 + 4z^6 + \cdots$$

where $\sigma(n) = \sum_{d|n} d$. Hence for some a, b we have

$$\Theta(z) = aF(z) + bG(z).$$

By comparing the first two coefficients, we have in fact

$$\Theta(z) = \frac{F(z) + 2G(z)}{3}.$$

Solving for the coefficients gives

$$r_4(n) = \begin{cases} 8\sigma(n) & n \not\equiv 0 \bmod 4 \\ 8(\sigma(n) - 4\sigma(n/4)) & n \equiv 0 \bmod 4. \end{cases}$$

Since we see $r_4(n) > 0$ this gives Lagrange's theorem, together with a formula for number of ways n is a sum of 4 squares.

Jacobi also used this idea to prove a formula for $r_2(n)$. I'll go through this in detail in the modular forms course next year.