Notes on Algebraic Numbers

Robin Chapman

January 20, 1995 (corrected November 3, 2002)

1 Introduction

This is a summary of my 1994–1995 course on Algebraic Numbers. (Revised and improved on 1993–1994!) The background assumed is standard elementary number theory—as found in my Level III course—and a little (Abelian) group theory. Corrections and suggestions for improvement are welcome, and will be credited in future editions!

I first learned algebraic number theory from Stewart & Tall's book ([3]) and this is an excellent account. However it's more abstract than the approach of this course and deals with general algebraic number theory while I deal mainly with the theory of quadratic fields. A book dealing mainly with quadratic fields is Cohn ([1]); I have incorporated many of the ideas in this book into this course, but this is a rather difficult book to read.

I am grateful to Jeremy Bygott for corrections to and suggestions on a previous version, and to Paul Epstein for a further correction.

2 Algebraic Numbers and Integers

An *algebraic number* is a root of a polynomial equation with rational coefficients, i.e., α is an algebraic number if and only if there exist $a_1, a_2, \ldots, a_n \in \mathbb{Q}$ with

$$\alpha^n + a_1\alpha^{n-1} + a_2\alpha^{n-2} + \cdots + a_{n-1}\alpha + a_n = 0.$$

Similarly an *algebraic integer* is a root of a monic polynomial equation with integer coefficients, i.e., β is an algebraic number if and only if there exist $b_1, b_2, \ldots, b_m \in \mathbb{Z}$ with

$$\beta^m + b_1\beta^{m-1} + b_2\beta^{m-2} + \cdots + b_{m-1}\beta + b_m = 0.$$

Clearly every algebraic integer is also an algebraic number. (As a convention I shall use Greek letters for algebraic numbers, and Roman letters for rational numbers.)

We think of algebraic integers and algebraic numbers as generalizations of integers and rationals. All integers are algebraic integers, and all rationals are algebraic numbers. Some examples of algebraic integers are i, $\sqrt{2}$, $\sqrt[3]{10}$, $(1+\sqrt{5})/2$ and $2\cos 2\pi/9$. The algebraic numbers include $(1+i)/2$, $\sqrt{1/5}$ and $\sin 2\pi/7$. The numbers e and π are not algebraic—they are *transcendental*. Transcendence proofs are very difficult, and will not be covered in this course. (See [2] for proofs that e and π are transcendental.)

A basic fact about algebraic integers is that a rational number which is also an algebraic integer is an (ordinary) integer:

Proposition 1 *If α is an algebraic integer and $\alpha \in \mathbb{Q}$, then $\alpha \in \mathbb{Z}$.*

Proof Write $\alpha = a/b$ in lowest terms. Suppose that α isn't an integer so $b \neq \pm 1$. As α is an algebraic integer, there exist $c_1, c_2, \ldots c_n \in \mathbb{Z}$ with

$$\alpha^n + c_1\alpha^{n-1} + \cdots + c_{n-1}\alpha + c_n = 0.$$

Writing $\alpha = a/b$ and clearing denominators gives

$$a^n + c_1 a^{n-1}b + \cdots + c_{n-1}ab^{n-1} + c_n b^n = 0,$$

which implies that $a^n \equiv 0 \pmod{b}$. As $b \neq \pm 1$ then b has a prime factor, p, say. Hence $p \mid a^n$ which implies that $p \mid a$. This means that p is a factor of both a and b contradicting the assumption that a/b is in its lowest terms. This contradiction establishes the result. $\square$

This result is useful as it shows that some algebraic numbers aren't algebraic integers.

The most important result in this section is that sums and product of algebraic numbers (resp. algebraic integers) are also algebraic numbers (resp. integers). Before we can prove this we need an alternative description of algebraic numbers and integers by means of matrix theory.

Lemma 1 *A number α is an algebraic number if and only if it is an eigenvalue of a (square) matrix with rational entries, and it is an algebraic integer if and only if it is an eigenvalue of a (square) matrix with integer entries.*

Proof Suppose first that A is a square matrix with integer entries. The characteristic polynomial of A is $\chi_A(\lambda) = \det(\lambda I - A)$. As the entries in $\lambda I - A$ involve only integer coefficients, it follows that its determinant $\chi_A(\lambda)$ is a monic polynomial with integer coefficients. Now if α is an eigenvalue of A, then $\chi_A(\alpha) = 0$ and so α is an algebraic integer. If A has rational entries then an identical argument shows that its eigenvalues are algebraic numbers.

For the converse suppose that α is an algebraic integer, so that there exist $b_1, b_2 \ldots, b_n \in \mathbb{Z}$ with

$$\alpha^n + b_1\alpha^{n-1} + \cdots + b_{n-1}\alpha + b_n = 0.$$

We can rewrite this as

$$\alpha^n = -b_1\alpha^{n-1} - \cdots - b_{n-1}\alpha - b_n. \qquad (*)$$

Let $\mathbf{v}$ be the (column) vector $(1 \;\; \alpha \;\; \alpha^2 \;\; \cdots \;\; \alpha^{n-1})^T$, so

$$\alpha\mathbf{v} = \begin{pmatrix} \alpha \\ \alpha^2 \\ \vdots \\ \alpha^{n-1} \\ \alpha^n \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -b_n & -b_{n-1} & -b_{n-2} & \cdots & -b_1 \end{pmatrix} \begin{pmatrix} 1 \\ \alpha \\ \vdots \\ \alpha^{n-2} \\ \alpha^{n-1} \end{pmatrix} = A\mathbf{v}$$

where A is a matrix with integer entries. (NB we have used $(*)$ here.) Hence α is an eigenvalue of A, with eigenvector $\mathbf{v}$. Similarly if α is an algebraic number then we can find a matrix A with rational entries having α as an eigenvalue. $\square$

A slightly more elaborate argument proves that sums and products of algebraic numbers or integers are also algebraic numbers or integers.

Theorem 1 *If β and γ are algebraic numbers (resp. algebraic integers), then $\beta + \gamma$ and $\beta\gamma$ are also algebraic numbers (resp. algebraic integers).*

Proof I'll just give the proof for algebraic integers, as that for algebraic numbers follows *mutatis mutandis* as in Lemma 1.

Suppose β and γ are algebraic integers. The idea of the proof is to find a non-zero vector $\mathbf{w}$ and two integer matrices B and C with $B\mathbf{w} = \beta\mathbf{w}$ and $C\mathbf{w} = \gamma\mathbf{w}$. This will imply that $\beta + \gamma$ and $\beta\gamma$ are eigenvalues of $B + C$ and BC respectively.

Suppose that

$$\beta^n + b_1\beta^{n-1} + \cdots + b_{n-1}\beta + b_n = 0$$

and

$$\gamma^m + c_1\gamma^{m-1} + \cdots + c_{m-1}\gamma + c_m = 0$$

where the b_js and c_ks are integers. Let $\mathbf{w}$ be the mn by 1 vector

$$(1 \;\; \beta \;\; \beta^2 \;\; \cdots \;\; \beta^{n-1} \;\; \gamma \;\; \beta\gamma \;\; \beta^2\gamma \;\; \cdots \;\; \beta^{n-1}\gamma \;\; \gamma^2 \;\; \beta\gamma^2 \;\; \cdots \;\; \beta^{n-1}\gamma^{m-1})^T.$$

Consider $\beta\mathbf{w}$. The entries in this vector all have the form $\beta^{j+1}\gamma^k$ where $0 \le j < n$ and $0 \le k < m$. If $j + 1 < n$ this is already an entry in $\mathbf{w}$; if $j + 1 = n$ it equals

$$\beta^n\gamma^k = -b_1\beta^{n-1}\gamma^k - \cdots - b_{n-1}\beta\gamma^k - b_n\gamma^k.$$

In all cases $\beta^{j+1}\gamma^k$ is a linear combination, with integer coefficients, of entries in $\mathbf{w}$. Hence there is an integer matrix B with $B\mathbf{w} = \beta\mathbf{w}$. Similarly there is an integer matrix C with $C\mathbf{w} = \gamma\mathbf{w}$. (Writing down general expressions for B and C is fairly horrible so I won't do it, but from the numerical examples I'll give in lectures I hope it's plain that this construction is a lot easier than it looks.)

Now

$$(B + C)\mathbf{w} = B\mathbf{w} + C\mathbf{w} = \beta\mathbf{w} + \gamma\mathbf{w} = (\beta + \gamma)\mathbf{w}$$

and

$$(BC)\mathbf{w} = B(C\mathbf{w}) = B(\gamma\mathbf{w}) = \gamma B\mathbf{w} = (\beta\gamma)\mathbf{w}$$

so $\beta + \gamma$ and $\beta\gamma$ are eigenvalues of $B + C$ and BC respectively. As $B + C$ and BC have integer coefficients, then by Lemma 1 $\beta + \gamma$ and $\beta\gamma$ are algebraic integers. □

3 Quadratic Fields

From now on we abandon the general study of algebraic numbers to concentrate on a specific class of such numbers, *quadratic numbers*, i.e., those which are solutions of quadratic equations with rational coefficients. By the solution formula for quadratics each such number has the form $a + b\sqrt{d}$ where a, b and d are rational numbers. We won't even look at all quadratic numbers at once; if we fix a non-square rational number d we shall consider the set of all numbers of the form $a + b\sqrt{d}$ with a and $b \in \mathbb{Q}$. This set is called a *quadratic field* and is denoted by $\mathbb{Q}(\sqrt{d})$. (NB if $d > 0$ then $\sqrt{d}$ is by convention the positive square root, and if $d = -D < 0$ then $\sqrt{d}$ will denote $i\sqrt{D}$.) If $d > 0$ then every element of $\mathbb{Q}(\sqrt{d})$ is real, and we say is a *real quadratic field*; if $d < 0$ then $\mathbb{Q}(\sqrt{d})$ contains non-real numbers—it is an *imaginary quadratic field*. The structure of real quadratic fields is radically different from that of imaginary quadratic fields, as we shall see on many subsequent occasions.

We should perhaps pause to justify our terminology; a field is by definition an algebraic structure which admits addition, subtraction, multiplication and division (except by 0) where these have all the familiar properties. This is indeed the case.

Lemma 2 *Let $K = \mathbb{Q}(\sqrt{d})$ be a quadratic field, where d is a non-square rational. Then if α, $\beta \in K$ then $\alpha + \beta$, $\alpha - \beta$, $\alpha\beta$ and, provided $\beta \neq 0$, $\alpha/\beta \in K$.*

Proof The proofs for addition, subtraction and multiplication are easy so I'll skip them. For division it's OK to treat the special case of $\alpha = 1$, for if $1/\beta \in K$ then by the multiplication rule $\alpha/\beta = \alpha(1/\beta) \in K$. Now if $\beta = a + b\sqrt{d}$ then we might try to write

$$\frac{1}{\beta} = \frac{1}{a + b\sqrt{d}} = \frac{a - b\sqrt{d}}{(a + b\sqrt{d})(a - b\sqrt{d})} = \frac{a - b\sqrt{d}}{a^2 - b^2 d}.$$

But remembering the commandment, "thou shalt not divide by zero", we must check that the denominator $a^2 - b^2 d$ is non-zero. But if $a^2 - b^2 d = 0$ then either $b \neq 0$ and then $d = (a/b)^2$ is a square contrary to hypothesis, or $b = 0$ and then $a = 0$ implying that $\beta = 0$, again contrary to hypothesis. Hence

$$\frac{1}{\beta} = \frac{a}{a^2 - b^2 d} + \frac{-b}{a^2 - b^2 d}\sqrt{d} \in K$$

as desired. □

In this proof we've employed the useful trick of rationalizing a denominator by multiplying by its *conjugate*. Formally speaking we say that the *conjugate* of $\beta = a + b\sqrt{d}$ is $\beta^* = a - b\sqrt{d}$. The operation of conjugation reflects the idea that both square roots of a number d are "equally good". (For a much deeper application of this principle see the course or books on Galois Theory, e.g. [2].) Conjugation respects the algebraic properties of $\mathbb{Q}(\sqrt{d})$.

Lemma 3 *Suppose that α, $\beta \in \mathbb{Q}(\sqrt{d})$. Then $(\alpha^*)^* = \alpha$, $(\alpha + \beta)^* = \alpha^* + \beta^*$, $(\alpha - \beta)^* = \alpha^* - \beta^*$, $(\alpha\beta)^* = \alpha^*\beta^*$ and, provided $\beta \neq 0$, $(\alpha/\beta)^* = \alpha^*/\beta^*$.*

Proof These are all straightforward apart, perhaps, for division, which can be deduced from multiplication by the following artifice. Put $\gamma = \alpha/\beta$ so $\alpha = \beta\gamma$. Then $\alpha^* = \beta^*\gamma^*$ and so $\alpha^*/\beta^* = \gamma^* = (\alpha/\beta)^*$. □

In showing that $1/\beta$ was in $\mathbb{Q}(\sqrt{d})$ I multiplied β by β^*, obtaining a non-zero rational. Hence we define the *norm* of β to be $N(\beta) = \beta\beta^*$. Similarly we define the *trace* of β to be $T(\beta) = \beta + \beta^*$. The norm and the trace of β are both rational numbers. Hence β and β^* are the roots of

$$(x - \beta)(x - \beta^*) = x^2 - T(\beta)x + N(\beta) = 0,$$

an equation with rational coefficients which is called the *characteristic equation* of β. The trace respects addition and subtraction, and the norm respects multiplication and division.

Lemma 4 *If α, $\beta \in \mathbb{Q}(\sqrt{d})$ then $T(\alpha + \beta) = T(\alpha) + T(\beta)$, $T(\alpha - \beta) = T(\alpha) - T(\beta)$, $N(\alpha\beta) = N(\alpha)N(\beta)$ and, provided $\beta \neq 0$, $N(\beta) \neq 0$ and $N(\alpha/\beta) = N(\alpha)/N(\beta)$.*

Proof First we note that $T(\alpha \pm \beta) = (\alpha \pm \beta) + (\alpha \pm \beta)^* = \alpha \pm \beta + \alpha^* \pm \beta^* = (\alpha + \alpha^*) \pm (\beta + \beta^*) = T(\alpha) \pm T(\beta)$. Similarly $N(\alpha\beta) = (\alpha\beta)(\alpha\beta)^* = \alpha\beta\alpha^*\beta^* = (\alpha\alpha^*)(\beta\beta^*) = N(\alpha)N(\beta)$. Finally if $\beta \neq 0$ then $N(\beta) \neq 0$ by the proof of Lemma 2, and so $N(\alpha/\beta) = (\alpha/\beta)(\alpha/\beta)^* = (\alpha/\beta)(\alpha^*/\beta^*) = (\alpha\alpha^*)/(\beta\beta^*) = N(\alpha)/N(\beta)$. □

In imaginary quadratic fields $\beta^* = \overline{\beta}$, the complex conjugate, and so $N(\beta) = \beta\overline{\beta} = |\beta|^2 \geq 0$; but in real quadratic fields $\overline{\beta} = \beta \neq \beta^*$ in general, and as $N(1) = 1$ and $N(\sqrt{d}) = -d$ the norm takes on both positive and negative values.

How many quadratic fields are there? If we have two such fields $\mathbb{Q}(\sqrt{d_1})$ and $\mathbb{Q}(\sqrt{d_2})$ how can we tell if they're the same? The following result tells us.

Proposition 2 *If $K_1 = \mathbb{Q}(\sqrt{d_1})$ and $K_2 = \mathbb{Q}(\sqrt{d_2})$, then $K_1 = K_2$ if and only if d_1/d_2 is a square of a rational.*

Proof If $d_1/d_2 = r^2$ where r is rational, then $\sqrt{d_1} = r\sqrt{d_2}$ and $\sqrt{d_2} = \frac{1}{r}\sqrt{d_1}$. Hence $a + b\sqrt{d_1} = a + br\sqrt{d_2}$ and $u + v\sqrt{d_2} = u + (v/r)\sqrt{d_1}$. It follows that $K_1 \subseteq K_2$ and $K_2 \subseteq K_1$, i.e., that $K_1 = K_2$.

The converse is slightly trickier. If $K_1 = K_2$ then $\sqrt{d_1} \in K_2$ and so we can write $\sqrt{d_1} = a + b\sqrt{d_2}$ with a and b rational. Squaring gives

$$d_1 = (a^2 + b^2 d_2) + 2ab\sqrt{d_2}.$$

This means that $d_1 = a^2 + b^2 d_2$ and $2ab = 0$. Hence either $a = 0$ or $b = 0$; $a = 0$ means that $d_1/d_2 = b^2$ and we're OK, while $b = 0$ means that $d_1 = a^2$ contrary to $\mathbb{Q}(\sqrt{d_1})$ being a quadratic field. □

We can now distinguish between quadratic fields. Can we now classify them? We say that an integer d is *squarefree* if it's not divisible by any square bigger than 1. This means that $d = \pm 1$ or $\pm p_1 p_2 \cdots p_k$ where the p_j are **distinct** primes.

Proposition 3 *If K is a quadratic field, then $K = \mathbb{Q}(\sqrt{d})$ where d is a uniquely determined squarefree integer and $d \neq 1$.*

Proof If $K = \mathbb{Q}(\sqrt{a/b})$ where a and b are integers then $K = \mathbb{Q}(\sqrt{b^2(a/b)}) = \mathbb{Q}(\sqrt{ab})$. Hence $K = \mathbb{Q}(\sqrt{r})$ where r is an integer. If r isn't squarefree then $c^2 \mid r$ where c is an integer bigger than 1. Then $K = \mathbb{Q}(\sqrt{r/c^2})$ and we can keep removing square factors until we reach $K = \mathbb{Q}(\sqrt{d})$ with d squarefree. Note that $d \neq 1$ as $\mathbb{Q}(\sqrt{1})$ isn't a quadratic field.

Now we address the question of uniqueness. If $\mathbb{Q}(\sqrt{d_1}) = \mathbb{Q}(\sqrt{d_2})$ with d_1 and d_2 squarefree then $d_1/d_2 = s^2$ where s is a **rational** number. It's now clear that d_1 and d_2 must have the same sign. Putting $s = u/v$ where u and v are integers we get that $d_1 v^2 = d_2 u^2$. If p is a prime dividing d_1 then p occurs to an odd power in the prime factorization of $d_1 v^2$ and so of $d_2 u^2$. But this means that $p \mid d_2$ also. Similarly every prime factor of d_2 is also a prime factor of d_1. As d_1 and d_2 are squarefree, have the same prime factors and the same sign, they are equal. □

Hence from now on when we write $\mathbb{Q}(\sqrt{d})$ we tacitly assume that d is a squarefree integer.

4 Rings of Integers

To do number theory we need the idea of integer. We have been considering quadratic fields $\mathbb{Q}(\sqrt{d})$ as generalizations of the rationals $\mathbb{Q}$. What should we consider instead of the integers $\mathbb{Z}$? Our first section gives us the answer. As the intersection of the set of algebraic integers and the set of rationals is the set of integers, we shall try to identify which elements of each quadratic field are algebraic integers. The norm and trace provide a useful characterization.

Proposition 4 *If $\alpha \in \mathbb{Q}(\sqrt{d})$ then α is an algebraic integer if and only if $N(\alpha)$ and $T(\alpha)$ are (ordinary) integers.*

Proof Suppose first that $N(\alpha)$ and $T(\alpha)$ are integers. Then

$$\alpha^2 - T(\alpha)\alpha + N(\alpha) = 0$$

and so α is an algebraic integer.

Now suppose that α is an algebraic integer. Then there exist integers b_j with

$$\alpha^n + b_1\alpha^{n-1} + \cdots + b_{n-1}\alpha + b_n = 0.$$

Taking conjugates gives

$$\alpha^{*n} + b_1\alpha^{*n-1} + \cdots + b_{n-1}\alpha^* + b_n = 0$$

which means that α^* is also an algebraic integer. As sums and products of algebraic integers are algebraic integers, then $T(\alpha) = \alpha + \alpha^*$ and $N(\alpha) = \alpha\alpha^*$ are algebraic integers. But $T(\alpha)$ and $N(\alpha)$ are rational numbers, and so $T(\alpha)$ and $N(\alpha)$ must be integers by Proposition 1. □

In a given quadratic field we can now identify the algebraic integers. If d is a squarefree integer other than 1 we denote by R_d the set of all algebraic integers in $\mathbb{Q}(\sqrt{d})$. We call R_d the *ring of integers* of $\mathbb{Q}(\sqrt{d})$. Incidentally note that the above Proposition shows that if $\alpha \in R_d$ then $\alpha^* \in R_d$.

Theorem 2 *If* $d \not\equiv 1 \pmod 4$ *then*

$$R_d = \{a + b\sqrt{d} : a, b \in \mathbb{Z}\}$$

while if $d \equiv 1 \pmod 4$ *then*

$$R_d = \left\{ \frac{r + s\sqrt{d}}{2} : r, s \in \mathbb{Z}, r \equiv s \pmod 2 \right\}.$$

Proof If $\alpha = a + b\sqrt{d}$ then $T(\alpha) = 2a \in \mathbb{Z}$ and $N(\alpha) = a^2 - b^2 d \in \mathbb{Z}$ and α is an algebraic integer. If $d \equiv 1 \pmod 4$ and $\alpha = \frac{1}{2}(r + s\sqrt{d})$, then $T(\alpha) = r \in \mathbb{Z}$ and $N(\alpha) = \frac{1}{4}(r^2 - s^2 d)$. If r and s are both even then $r^2 \equiv s^2 \equiv 0 \pmod 4$, while if r and s are both odd then $r^2 \equiv s^2 \equiv 1 \pmod 4$. In any case then $r^2 - s^2 d \equiv 0 \pmod 4$ and so $N(\alpha) \in \mathbb{Z}$ and α is an algebraic integer.

Conversely suppose that $\alpha \in R_d$. We can write $\alpha = a + b\sqrt{d}$ with a and b **rational**. Now $T(\alpha) = 2a$ is an integer, so we'll split the argument into two cases: $2a$ even or $2a$ odd.

If $2a$ is even, then a is an integer. As $N(\alpha) = a^2 - b^2 d$ is an integer it follows that so is $b^2 d$. Now b is rational and d is squarefree. I claim that this means that b must be an integer. Put $b = u/v$ in lowest terms with u and v integers. Then du^2/v^2 is an integer. But if p is a prime factor of v we must have $p^2 \mid du^2$, and $p \nmid u$. This means that $p^2 \mid d$ which is squarefree—a contradiction! So $v = \pm 1$ and b is an integer.

We now turn to the case where $2a$ is odd. Put $r = 2a$ and $s = 2b$. The number $r^2 - ds^2 = 4(a^2 - b^2 d)$ is an integer which is divisible by 4. Hence ds^2 is an integer, and repeating the above argument we see that s is an integer also. As $r^2 - ds^2$ is even then s can't be even. As r and s are both odd then $r^2 \equiv s^2 \equiv 1 \pmod 4$ and so $0 \equiv r^2 - ds^2 \equiv 1 - d \pmod 4$, i.e., $d \equiv 1 \pmod 4$. □

We can rephrase this result as follows. If we define $\tau_d = \sqrt{d}$ if $d \not\equiv 1 \pmod 4$ and $\tau_d = \frac{1}{2}(1 + \sqrt{d})$ if $d \equiv 1 \pmod 4$ then

$$R_d = \{a + b\tau_d : a, b \in \mathbb{Z}\}.$$

It may seem odd at first to be considering numbers like $\frac{1}{2}(1 + \sqrt{5})$ as elements of R_d as these appear to have "denominators", but to exclude them would make the theory a lot harder—the crucial Corollary 3 would be **false** if we excluded these numbers. (In future I'll usually write τ instead of τ_d when it's clear what d is.) One of the most important things about R_d is that it's a *ring*, i.e., it's closed under addition, subtraction and multiplication.

Proposition 5 *If* α, $\beta \in R_d$, *then* $\alpha + \beta$, $\alpha - \beta$ *and* $\alpha\beta \in R_d$.

Proof Recall that $\alpha \in R_d$ if and only if α is an algebraic integer and $\alpha \in \mathbb{Q}(\sqrt{d})$. Now the algebraic integers are closed under addition, subtraction and multiplication (Theorem 1), and $\mathbb{Q}(\sqrt{d})$ is also closed under these operations. The result is now immediate.

Alternatively you can use the description

$$R_d = \{a + b\tau : a, b \in \mathbb{Z}\}$$

to give a direct proof, putting $\alpha = a + b\tau$, $\beta = a' + b'\tau$ and computing $\alpha \pm \beta$ and $\alpha\beta$ explicitly. (Details left as an exercise!) □

5 Divisibility and Factorization

We now have the basic materials we need to start to do number theory in quadratic fields. Classical number theory deals with divisibility, congruences and primes, all in the integers. Can we define similar notions inside R_d?

Divisibility and congruences are easy: if $\alpha, \beta \in R_d$ with $\alpha \neq 0$ then we say that α *divides* β (and write $\alpha \mid \beta$) if $\beta/\alpha \in R_d$. Similarly we say that α *is congruent to* β modulo γ (and write $\alpha \equiv \beta \pmod{\gamma}$) if $\gamma \mid (\alpha - \beta)$. All the familiar properties of divisibility and congruences are true in R_d also, so I won't repeat them. Note though that if $\alpha \mid \beta$ then $N(\alpha) \mid N(\beta)$.

Some new things do happen in the various R_d. Consider $\varepsilon = 1 + \sqrt{2} \in R_2$. As $1/\varepsilon = -1 + \sqrt{2}$ it follows that $\alpha/\varepsilon \in R_2$ for all $\alpha \in R_2$, and so ε divides every element of R_2. An element of R_d with this property is called a *unit*; formally we define a *unit* to be an element $\varepsilon \in R_d$ whose inverse $1/\varepsilon$ is also an element of R_d. We denote the set of units of R_d by U_d. The existence of units complicates slightly the question of how to generalize the notion of prime. We digress slightly to study the theory of units. There's a simple characterization of units.

Proposition 6 *An element ε of R_d is a unit if and only if $N(\varepsilon) = \pm 1$.*

Proof Suppose first that $N(\varepsilon) = \pm 1$. This means that $\varepsilon\varepsilon^* = \pm 1$ and so $1/\varepsilon = \pm\varepsilon^* \in R_d$, and ε is a unit.

Conversely suppose that $\varepsilon \in U_d$. Then $1/\varepsilon \in R_d$ and so

$$1 = N(1) = N(\varepsilon(1/\varepsilon)) = N(\varepsilon)N(1/\varepsilon).$$

But as $N(\varepsilon)$ and $N(1/\varepsilon)$ are integers this implies that $N(\varepsilon) = \pm 1$. □

Another important fact is that U_d is a group under multiplication.

Lemma 5 *If ε and $\eta \in U_d$, then $\varepsilon\eta \in U_d$ and $1/\varepsilon \in U_d$. Also $1 \in U_d$.*

Proof An easy exercise. □

We now determine the size of U_d. This result shows a radical difference between the arithmetic of real and imaginary quadratic fields.

Theorem 3 *(a) If $d < 0$ then $U_d = \{\pm 1\}$ unless $d = -1$ when $U_{-1} = \{\pm 1, \pm i\}$, or $d = -3$ when $U_{-3} = \{\pm 1, \frac{1}{2}(\pm 1 \pm \sqrt{-3})\}$.*

(b) If $d > 0$ then U_d is an infinite group.

Proof (a) It's easy to check that the numbers claimed to be units really are units. Put $D = -d > 0$. Let $\varepsilon \in U_d$. We can certainly write $\varepsilon = \frac{1}{2}(r + s\sqrt{d})$ with r and s integers. As $N(\varepsilon) = 1$ (remember that as $d < 0$, $N(\varepsilon) \geq 0$) we get

$$4 = r^2 - s^2 d = r^2 + s^2 D.$$

As $s^2 D \geq 0$ we must have $|r| \leq 2$. If $r = \pm 2$ then $s = 0$ and $\varepsilon = \pm 1$. If $r = \pm 1$ then $s^2 D = 3$, and so $D = 3$ and $s = \pm 1$; hence $\varepsilon = \frac{1}{2}(\pm 1 \pm \sqrt{-3})$. If $r = 0$ then $s^2 D = 4$, and as D is squarefree $D = 1$ and $s = \pm 2$; hence $\varepsilon = \pm i$.

(b) By the theory of Pell's equation, there is a solution of the equation $x^2 - dy^2 = 1$ with x and y positive integers. If $\varepsilon = x + y\sqrt{d}$ then $\varepsilon \in U_d$ and $\varepsilon > 1$. Now $\varepsilon^2, \varepsilon^3, \ldots, \varepsilon^n, \ldots$ are all units and $\varepsilon < \varepsilon^2 < \varepsilon^3 < \cdots < \varepsilon^n < \cdots$ so they are all distinct. Hence U_d forms an infinite group. □

In the language of group theory we see easily that for negative d, U_d is a cyclic group (a generator is -1, i or $\frac{1}{2}(1 + \sqrt{-3})$ as appropriate). What is the structure of U_d when $d > 0$?

Theorem 4 *If $d > 0$ there exists a unique unit $\varepsilon \in U_d$ such that $\varepsilon > 1$ and every $\xi \in U_d$ has the form $\pm\varepsilon^n$ for some $n \in \mathbb{Z}$.*

Proof We try to take ε to be the smallest unit that's bigger than 1. But how do we know that this description makes sense? By Theorem 3 there is a unit bigger than 1, but maybe the set of all such units doesn't have a least element? (Cf. the set of all positive rationals.)

Let η be any unit bigger than 1, and let

$$U = \{\xi \in U_d : 1 < \xi \leq \eta\}.$$

If there is a least unit bigger than 1, it must be the least element of U. I claim that U is a finite non-empty set. It's non-empty as $\eta \in U$. If $\xi \in U$ then $\xi\xi^* = \pm 1$ and so $|\xi^*| = 1/\xi < 1$. As $T(\beta) = \xi + \xi^* = \xi \pm |\xi^*|$ it follows that $0 < T(\xi) < \eta + 1$. As $\xi^2 - T(\xi)\xi + N(\xi) = 0$ and $N(\xi) = \pm 1$ then

$$\xi = \frac{T(\xi) \pm \sqrt{T(\xi)^2 - 4N(\xi)}}{2} = \frac{t \pm \sqrt{t^2 \mp 4}}{2}$$

where t is an integer between 0 and $\eta + 1$. Hence ξ is confined to a finite set, and so the set U is finite. As U is finite and non-empty it has a least element ε; ε is a unit, $\varepsilon > 1$ and there are no other units between 1 and ε.

Now consider any $\xi \in U_d$. First assume that $\xi > 0$. Consider the real number $(\log \xi)/(\log \varepsilon)$. There is an integer n with

$$n \leq \frac{\log \xi}{\log \varepsilon} < n + 1,$$

and so

$$0 \leq \frac{\log \xi}{\log \varepsilon} - n = \frac{\log(\xi\varepsilon^{-n})}{\log \varepsilon} < 1$$

which implies that $1 \leq \xi\varepsilon^{-n} < \varepsilon$. Now $\xi\varepsilon^{-n}$ is a unit and so can't be bigger than 1 and less than ε. Hence $\xi\varepsilon^{-n} = 1$ and $\xi = \varepsilon^n$. If $\xi < 0$ then $-\xi$ is a positive unit and so $-\xi = \varepsilon^n$ for some integer n, and $\xi = -\varepsilon^n$.

I claim that ε is uniquely determined, for if ε' also satisfied the same conditions then $\varepsilon' = \varepsilon^n$ and $\varepsilon = \varepsilon'^m$ for some integers n and m and so $\varepsilon = \varepsilon^{mn}$ meaning that $mn = 1$. Hence $m = n = \pm 1$ and we must have $n = +1$ as $\varepsilon' = \varepsilon^n > 1$. $\square$

The unit ε in the above theorem is called the *fundamental unit* of R_d. In terms of group theory the theorem states that the group U_d is isomorphic to the direct product $\mathbb{Z} \times \mathbb{Z}_2$. If $d \not\equiv 1 \pmod 4$ the fundamental unit can be found by applying the continued fraction algorithm for Pell's equation. If $d \equiv 1 \pmod 4$ this algorithm may or may not find the fundamental unit, but a variant method where you start with $\frac{1}{2}(1 + \sqrt{d})$ instead of $\sqrt{d}$ always does work.

The above proof illustrates a useful technique. If $\alpha \in R_d$ and we know $N(\alpha)$ and have inequalities bounding the size of α, then we can bound the size of $T(\alpha)$ and can restrict α to lie in a explicitly specified finite set.

We now try to generalize the notion of prime. As every element of R_d is divisible by every unit, we make the following definition. We say that $\alpha \in R_d$ is *irreducible* if it's non-zero, not a unit, and whenever $\alpha = \beta\gamma$ with β, $\gamma \in R_d$ then either β or γ is a unit. (We don't call such elements primes, the reason being is that we reserve this term for elements satisfying a more stringent condition which we'll meet later.) It's easy to show (by induction on $|N(\alpha)|$) that every $\alpha \in R_d$ is either 0 or a unit or irreducible or a product of irreducibles. It's now natural to ask how unique this factorization into irreducibles is. For instance in R_{-1} we have $14 - 5i = (4 + i)(3 - 2i) = (1 - 4i)(2 + 3i)$, and all of these factors are irreducible. But these two factorizations are really the same in disguise. For i is a

unit and $4+i = i(1-4i)$ and $(3-2i) = -i(2+3i)$. We can go to and fro between these factorizations by juggling with units. We make the definition that non-zero elements α and $\beta \in R_d$ are *associates* if α/β is a unit. An easy exercise shows that being associates is an equivalence relation.

We now say that two factorizations

$$\alpha = \pi_1\pi_2\cdots\pi_m = \rho_1\rho_2\cdots\rho_n$$

of α into irreducibles are *equivalent* if $m = n$ and we can re-order the π_js so that π_j is an associate of ρ_j for each j. We might hope that factorization into irreducibles is unique up to equivalence. Alas this is not so. In R_{-5} consider

$$6 = 2\cdot 3 = (1+\sqrt{-5})(1-\sqrt{-5}).$$

These are two non-equivalent factorizations into irreducibles. We thus say that R_d has *unique factorization* if every factorization into irreducibles is unique up to equivalence. It follows that R_{-5} doesn't have unique factorization, but we shall see that there do exist some R_d which do have unique factorization. (NB to show that some R_d has unique factorization one needs a proof not just an example!)

If we go back to standard number theory and study the proof of unique factorization of natural numbers, we find that the proof ultimately depends on the Euclidean algorithm. If we could find a generalized version of this algorithm inside a particular R_d then perhaps we could show that this R_d has unique factorization. This is indeed so. We say that R_d is *Euclidean* if whenever α, $\beta \in R_d$ and $\beta \neq 0$, then there exists a $\lambda \in R_d$ with $|N(\alpha-\beta\lambda)| < |N(\beta)|$. Equivalently R_d is Euclidean if and only if for all $\xi \in \mathbb{Q}(\sqrt{d})$ there exists $\gamma \in R_d$ with $|N(\xi-\gamma)| < 1$. (To see this put $\xi = \alpha/\beta$.) Several examples of Euclidean R_d are known, including R_{-1}, R_{-2}, R_{-3}, R_{-7}, R_{-11}, R_2, R_3, R_5 and R_{13}. I shall prove some of these in lectures, and set others as exercises.

By means of the classical Euclidean algorithm one can find the greatest common divisor g of two numbers a and b, and write $g = ra + sb$. We can also do this in R_d provided that R_d is Euclidean. I'll give an "abstract" proof of this result, but it can also be proved algorithmically, and I'll illustrate this in lectures.

Proposition 7 *Let R_d be Euclidean. If α and β are non-zero elements of R_d then there exists $\gamma \in R_d$ with*
(i) $\gamma \mid \alpha$ and $\gamma \mid \beta$, and
(ii) if $\delta \mid \alpha$ and $\delta \mid \beta$ then $\delta \mid \gamma$.
Also there exist η, $\xi \in R_d$ with $\gamma = \eta\alpha + \xi\beta$.

Proof Let I be the set

$$\{\eta\alpha + \xi\beta : \eta, \xi \in R_d\}.$$

We look for a non-zero element γ in this set as "small" as possible in an appropriate sense. It's easy to check that I contains α and β, it's closed under addition and subtraction and if $\delta \in I$ and $\zeta \in R_d$ then $\zeta\delta \in I$. If γ is a non-zero element of I, then $|N(\gamma)|$ is a positive integer, so we can choose such a γ with $|N(\gamma)|$ as small as possible. Certainly there exist η, $\xi \in R_d$ with $\gamma = \eta\alpha + \xi\beta$. If $\delta \mid \alpha$ and $\delta \mid \beta$ then $\delta \mid \eta\alpha$ and $\delta \mid \xi\beta$ and so $\delta \mid \gamma$ and (ii) is verified.

It only remains to prove (i). By the Euclidean property there exists $\lambda \in R_d$ with $|N(\alpha-\lambda\gamma)| < |N(\gamma)|$. Now $\alpha \in I$ and $\lambda\gamma \in I$ so $\alpha - \lambda\gamma \in I$. But as γ was chosen to make $|N(\gamma)|$ as small as possible for non-zero $\gamma \in I$ then $\alpha - \lambda\gamma = 0$ and $\gamma \mid \alpha$. A similar argument shows $\gamma \mid \beta$. □

If γ has the properties described in the proof we call γ a *greatest common divisor* of α and β. Greatest common divisors are only unique up to multiplication by units. Given α and β we can find

γ by a generalization of the standard Euclidean algorithm, and also η and ξ by a generalization of the extended Euclidean algorithm.

To prove unique factorization for the ordinary integers we need the result that if p is prime and $p \mid ab$ then either $p \mid a$ or $p \mid b$. The corresponding result is true for irreducibles in Euclidean R_d.

Proposition 8 *Suppose that R_d is Euclidean and π is irreducible in R_d. If $\pi \mid \alpha\beta$ with α, $\beta \in R_d$ then either $\pi \mid \alpha$ or $\pi \mid \beta$.*

Proof Suppose $\pi \nmid \alpha$, and consider a greatest common divisor γ of α and π. As π is irreducible and $\gamma \mid \pi$ then either γ is a unit or an associate of π. But if γ were an associate of π we'd have $\pi \mid \gamma \mid \alpha$ and so $\pi \mid \alpha$. Hence γ is a unit and we may assume that $\gamma = 1$. There exist η and $\xi \in R_d$ with $1 = \eta\alpha + \xi\pi$ and so $1 \equiv \eta\alpha \pmod{\pi}$. But then $\beta \equiv \eta\alpha\beta \equiv 0 \pmod{\pi}$ and $\pi \mid \beta$ as required. □

Because of this result we define $\pi \in R_d$ to be *prime* if π isn't zero or a unit, and if $\pi \mid \alpha\beta$ implies that $\pi \mid \alpha$ or $\pi \mid \beta$. This Proposition now states that if R_d is Euclidean, then every irreducible is prime. This is **not** necessarily true if R_d isn't Euclidean. For instance 2 is irreducible but not prime in R_{-5}. (Note that $2 \mid (1+\sqrt{-5})(1-\sqrt{-5})$ but $2 \nmid (1 \pm \sqrt{-5})$.) On the other hand it's not hard to show that every prime is irreducible.

Lemma 6 *If π is prime in R_d then it's irreducible.*

Proof Suppose that $\pi = \alpha\beta$. Then either $\pi \mid \alpha$ or $\pi \mid \beta$. Let's suppose it's the former. Then $\pi \mid \alpha$ and $\alpha \mid \pi$ so α is an associate of β and β is a unit. This means that π is irreducible. □

An easy induction argument shows that if π is prime and $\pi \mid \alpha_1\alpha_2\cdots\alpha_n$ then $\pi \mid \alpha_j$ for some j. If in R_d all irreducibles are primes then we have unique factorization. The proof follows the lines of the unique factorization proof for the integers.

Theorem 5 *If every irreducible element of R_d is prime then R_d has unique factorization.*

Proof Suppose that the non-zero element α of R_d has two factorizations

$$\alpha = \pi_1\pi_2\cdots\pi_r = \rho_1\rho_2\cdots\rho_s$$

into irreducibles. If $r = 1$ then α is irreducible and we must have $s = 1$ and $\rho_1 = \alpha$. We may suppose that $r > 1$. As π_1 is prime then $\pi_1 \mid \rho_1\rho_2\cdots\rho_s$ and so $\pi_1 \mid \rho_j$ for some j. We can re-arrange the ρ_js so that $\pi_1 \mid \rho_1$. As ρ_1 is irreducible then π_1 must be an associate of ρ_1. Put $\rho_1 = \varepsilon\pi_1$ where ε is a unit. Dividing through by π_1 we get

$$\pi_2\pi_3\cdots\pi_r = \varepsilon\rho_2\rho_3\cdots\rho_s.$$

We can repeat the argument; π_2 must divide one of the terms in the second product, and as it cannot divide the unit ε, then we may assume it divides ρ_2 and so is an associate of ρ_2. We divide through and keep going. Eventually we re-order the ρ_js so that π_j is an associate of ρ_j for all j, and $r = s$. □

6 Ideal Theory

We can remedy the lack of unique factorization in the general R_d by considering factorizations of ideals rather than factorizations of elements. We've already seen an application of this idea in the proof of Proposition 7. The set I defined therein has the properties:

(i) $0 \in I$,

(ii) if α, $\beta \in I$ then $\alpha + \beta \in I$,
(iii) if $\alpha \in I$ and $\lambda \in R_d$ then $\lambda\alpha \in I$.
We define an *ideal* of R_d to be a subset I of R_d satisfying the above properties. If I is an ideal and $\alpha \in I$ then $-\alpha = (-1)\alpha \in I$ by (iii) and so $(I, +)$ is a subgroup of $(R_d, +)$. Obvious examples of ideals include both $\{0\}$ and R_d. Sometimes we want to exclude these from consideration and so we say that an ideal I is *proper* if $I \neq R_d$ and *non-zero* if $I \neq \{0\}$. Another basic class of ideals are the *principal* ideals; if $\alpha \in R_d$ the set

$$\langle\alpha\rangle = \{\alpha\beta : \beta \in R_d\} = \{\gamma : \alpha \mid \gamma\}$$

is an ideal, the *principal* ideal generated by α. It's easy to see that $\langle\alpha\rangle = \langle\beta\rangle$ if and only if α and β are associates; in particular $\langle\alpha\rangle = R_d$ if and only if α is a unit. If R_d is Euclidean then all its ideals are principal.

Theorem 6 *If R_d is Euclidean and I is an ideal of R_d then I is principal.*

Proof If $I = \{0\}$ then $I = \langle 0\rangle$. If $I \neq \{0\}$ then the proof follows the proof of Proposition 7 almost exactly. Choose a nonzero element γ of I making $|N(\gamma)|$ as small as possible. I claim that $I = \langle\gamma\rangle$. As $\beta\gamma \in I$ for all $\beta \in R_d$ then $\langle\gamma\rangle \subseteq I$. For the reverse inclusion assume that $\alpha \in I$. By the Euclidean property there exists $\lambda \in R_d$ with $|N(\alpha - \lambda\gamma)| < |N(\gamma)|$. As I is an ideal $\lambda\gamma \in I$ and so $\alpha - \lambda\gamma \in I$. But as γ was chosen to make $|N(\gamma)|$ as small as possible for non-zero $\gamma \in I$ then $\alpha - \lambda\gamma = 0$ and $\gamma \mid \alpha$. □

We say R_d is *principal* if all its ideals are principal. The above theorem shows that every Euclidean R_d is principal. (The converse isn't true; for instance R_{-19} is principal but not Euclidean.) If R_d is principal then it has unique factorization.

Theorem 7 *If all ideals of R_d are principal, then R_d has unique factorization.*

Proof By Theorem 5 it suffices to show that every irreducible in R_d is prime. Let π be irreducible, and suppose that $\pi \mid \alpha\beta$, but $\pi \nmid \alpha$. As in Proposition 7 the set

$$I = \{\eta\pi + \xi\alpha : \eta, \xi \in R_d\}$$

is an ideal, and so $I = \langle\gamma\rangle$ for some γ. As $\pi \in I$ then $\gamma \mid \pi$ and as π is irreducible then γ is either a unit or an associate of π. But if γ is an associate of π then $\pi \mid \gamma \mid \alpha$ as $\alpha \in I$. This is contrary to hypothesis, so γ is a unit and $I = \langle\gamma\rangle = R_d$. Hence $1 \in I$ and so $1 = \eta\pi + \xi\alpha$ with η, $\xi \in R_d$. It follows that $1 \equiv \xi\alpha \pmod{\pi}$ and so $\beta \equiv \xi\alpha\beta \equiv 0 \pmod{\pi}$. This means that $\pi \mid \beta$ and we've shown that π is prime. □
(NB this proof is a rehash of that of Proposition 8.)

The two basic operations on ideals are addition and multiplication. Addition is easy to define. If I and J are ideals of R then we define their sum to be

$$I + J = \{\alpha + \beta : \alpha \in I, \beta \in J\}$$

which it's easy to show is an ideal. The sum $I + J$ is the smallest ideal containing both I and J. The sum has some familiar properties: $I + J = J + I$, $I + (J + K) = (I + J) + K$ and $I + \langle 0\rangle = I$, and some unfamiliar ones: $I + I = I$ and $I + \langle 1\rangle = \langle 1\rangle$. All of these make instructive exercises.

The product is harder to define. We might try to define IJ to be

$$\{\alpha\beta : \alpha \in I, \beta \in J\}$$

but this isn't always an ideal. (There are examples where it's not closed under addition.) Instead we consider the additive subgroup of R_d generated by the above set which is

$$IJ = \{\alpha_1\beta_1 + \alpha_2\beta_2 + \cdots + \alpha_n\beta_n : \alpha_1, \ldots, \alpha_n \in I, \beta_1, \ldots, \beta_n \in J, n \in \mathbb{N}\}$$

as the product of I and J. This **is** an ideal. The product IJ is a subset of both I and J. The product also has some familiar properties: $IJ = JI$, $I(JK) = (IJ)K$, $I(J + K) = IJ + IK$, $I\langle 0\rangle = \langle 0\rangle$ and $I\langle 1\rangle = I$. Again these make good exercises!

There's also a conjugation operation on ideals. If I is an ideal then

$$I^* = \{\alpha^* : \alpha \in I\}$$

is easily seen to be an ideal. Conjugation respects addition and multiplication: $(I + J)^* = I^* + J^*$ and $(IJ)^* = I^*J^*$ (exercises!).

Addition and multiplication preserve inclusions: if $J \subseteq K$ then $I + J \subseteq I + K$ and $IJ \subseteq IK$.

We would like to be able to do arithmetic with ideals explicitly, as we do for elements of R_d. As it stands this is difficult as we don't even have a way of specifying ideals (apart from principal ones). One case is easy though; multiplication by a principal ideal. If I is a principal ideal then multiplication by I is easy.

Lemma 7 *Let $\alpha \in R_d$ and J be an ideal. Then*

$$\langle\alpha\rangle J = \alpha J = \{\alpha\beta : \beta \in J\}.$$

In particular if $\beta \in R_d$ then

$$\langle\alpha\rangle\langle\beta\rangle = \langle\alpha\beta\rangle.$$

Proof If $\beta \in J$ then as $\alpha \in \langle\alpha\rangle$ we have $\alpha\beta \in \langle\alpha\rangle J$ and so $\alpha J \subseteq \langle\alpha\rangle J$. Now the typical element of $\langle\alpha\rangle J$ has the form

$$\gamma_1\beta_1 + \gamma_2\beta_2 + \cdots + \gamma_n\beta_n$$

with each $\gamma_j \in \langle\alpha\rangle$. If we write $\gamma_j = \alpha\lambda_j$ we get

$$\gamma_1\beta_1 + \gamma_2\beta_2 + \cdots + \gamma_n\beta_n = \alpha(\lambda_1\beta_1 + \lambda_2\beta_2 + \cdots + \lambda_n\beta_n) \in \alpha J$$

as $\lambda_1\beta_1 + \cdots + \lambda_n\beta_n \in J$. Hence $\langle\alpha\rangle J \subseteq \alpha J$ and so these two ideals are equal.

If $J = \langle\beta\rangle$ it's easy to see that $\langle\alpha\rangle\langle\beta\rangle = \alpha\langle\beta\rangle = \langle\alpha\beta\rangle$. □

We generalize our notation for principal ideals as follows. We shall denote the ideal

$$\langle\alpha_1\rangle + \langle\alpha_2\rangle + \cdots + \langle\alpha_n\rangle$$

by $\langle\alpha_1, \alpha_2, \ldots, \alpha_n\rangle$. More explicitly

$$\langle\alpha_1, \alpha_2, \ldots, \alpha_n\rangle = \{\beta_1\alpha_1 + \beta_2\alpha_2 + \cdots + \beta_n\alpha_n : \beta_j \in R_d\}.$$

Using the associative and distributive laws we see that if $I = \langle\alpha_1, \ldots, \alpha_n\rangle$ and $J = \langle\beta_1, \ldots, \beta_m\rangle$ then

$$I + J = \langle\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_m\rangle$$

and

$$IJ = \langle\alpha_1\beta_1, \alpha_1\beta_2, \ldots, \alpha_1\beta_n, \alpha_2\beta_1, \ldots, \alpha_n\beta_m\rangle.$$

In general not all ideals are principal. For instance $\langle 2, 1+\sqrt{-5}\rangle$ is a non-principal ideal of R_{-5}. I'll prove this and give other examples in the lectures. But although we can't always write an ideal in the form $\langle\alpha\rangle$, can we always write ideals in the form $\langle\alpha_1, \alpha_2, \ldots, \alpha_n\rangle$? The answer to this is yes. In fact every ideal of R_d can actually be written in the form $\langle\alpha, \beta\rangle$. To demonstrate this I'll prove an even stronger result.

Every ideal is a subgroup of $(R_d, +)$ and it is convenient to first classify all possible subgroups of R_d and then to ask which of them are ideals. We introduce some notation. If $\alpha_1, \ldots, \alpha_n \in R_d$ then I denote by $[\,\alpha_1, \ldots, \alpha_n\,]$ the **subgroup** of R_d generated by the α_js, i.e.,

$$[\,\alpha_1, \ldots, \alpha_n\,] = \{b_1\alpha_1 + \cdots + b_n\alpha_n : b_1, \ldots, b_n \in \mathbb{Z}\}.$$

In fact we only need two generators.

Proposition 9 *Let G be a subgroup of R_d. There exist integers a, m and n such that m, $n \geq 0$ and $G = [\,a + m\tau, n\,]$.*

Proof A typical element of G has the form $r + s\tau$ where r and s are integers. We consider first the set H of all the s that occur, i.e.,

$$H = \{s : r + s\tau \in G\}.$$

It is easy to see that H is a subgroup of $\mathbb{Z}$. Now every subgroup of $\mathbb{Z}$ has the form $m\mathbb{Z}$ for some $m \geq 0$. As $m \in H$ there's an a with $a + m\tau \in G$. Also $G \cap \mathbb{Z}$ is a subgroup of $\mathbb{Z}$ and so $G \cap \mathbb{Z} = n\mathbb{Z}$ for some $n \geq 0$. Certainly $n \in G$.

I claim that $G = [\,a + m\tau, n\,]$. It's clear that $[\,a + m\tau, n\,] \subseteq G$. For the reverse inclusion take $r + s\tau \in G$. As $s \in H$ there exists $u \in \mathbb{Z}$ with $s = um$. Now consider $r - ua = r + s\tau - u(a + m\tau)$. As this is an element of $G \cap \mathbb{Z}$ then $r - ua = vn$ where $v \in \mathbb{Z}$. Hence

$$r + s\tau = r - ua + u(a + m\tau) = vn + u(a + m\tau) \in [\,a + m\tau, n\,].$$

It follows that $G = [\,a + m\tau, n\,]$ as required. □

Corollary 1 *If I is a non-zero ideal of R_d, then there exist integers a, m and n with $m > 0$ and $n > 0$ such that $I = [\,a + m\tau, n\,]$.*

Proof Let $G = I$ in the above proof. Choose α to be any non-zero element of I. Note that $N(\alpha) = \alpha\alpha^* \in I$ and so $N(\alpha)\tau \in I$. It follows that the sets H and $I \cap \mathbb{Z}$ of the above proof both contain $N(\alpha)$ and so are non-zero. This means we can take $m > 0$ and $n > 0$. □

Given a subgroup G we can express it in the above form by the following algorithm. Suppose that

$$G = [\,b_1 + c_1\tau, b_2 + c_2\tau, \ldots, b_k + c_k\tau\,].$$

Let $m = \gcd(c_1, c_2, \ldots, c_k)$. Then we can find integers $r_1, r_2, \ldots, r_k$ with $m = r_1c_1 + \cdots + r_kc_k$. Now

$$r_1(b_1 + c_1\tau) + \cdots + r_k(b_k + c_k\tau) = a + m\tau$$

say. Also each $c_j = s_jm$ where s_j is an integer. Put $t_j = (b_j + c_j\tau) - s_j(a + m\tau) = b_j - s_ja$. Then $n = \gcd(t_1, \ldots, t_k)$.

Representing G as $[\,a + m\tau, n\,]$ is very convenient for computation for we can easily test whether a particular element α is an element of G. For if $\alpha = b + c\tau$, then $\alpha \in G$ if and only if there are integers r and s with

$$b + c\tau = r(a + m\tau) + sn.$$

It follows that r must equal c/m and s must equal $(b-ra)/n$, i.e., $\alpha \in G$ if and only if these are both integers. Similarly we can test if two subgroups are equal for it is easy to show (exercise!) that provided m, m', n, $n' > 0$ then $[\, a+m\tau, n\,] = [\, a'+m'\tau, n'\,]$ if and only if $m'=m$, $n'=n$ and $a' \equiv a \pmod{n}$.

Another consequence is that every ideal can be generated by two elements.

Corollary 2 *If I is an ideal in R_d then $I = \langle \alpha, \beta \rangle$ for some α, β.*

Proof Certainly $I = [\, m, a+n\tau\,]$ where a, m and n are integers. Put $\alpha = a+m\tau$ and $\beta = n$. Then $\alpha,\ \beta \in I$ and so $\langle \alpha, \beta \rangle \subseteq I$. Conversely if $\gamma \in I$ then $\gamma = b\alpha + c\beta$ with b and c integers, and so $\gamma \in \langle \alpha, \beta \rangle$. Hence $I \subseteq \langle \alpha, \beta \rangle$ as required. $\square$

Given any ideal $I = \langle \alpha_1, \alpha_2, \ldots, \alpha_k \rangle$ then it's easy to show (exercise!) that

$$I = [\, \alpha_1, \tau\alpha_1, \alpha_2, \tau\alpha_2, \ldots, \alpha_k, \tau\alpha_k\,],$$

and by applying the above algorithm we can write I in the form $[\, a+m\tau, n\,]$.

As an example of this procedure let $I = \langle 2, 1+\sqrt{-5} \rangle$ be an ideal of R_{-5}. Then as $\tau = \sqrt{-5}$ we have

$$\begin{aligned} I &= [\, 2, 1+\sqrt{-5}, 2\sqrt{-5}, \sqrt{-5}(1+\sqrt{-5})\,] \\ &= [\, 2, 1+\sqrt{-5}, 2\sqrt{-5}, -5+\sqrt{-5}\,]. \end{aligned}$$

As $\gcd(0,1,2,1) = 1$ then $m=1$ and we can take $a+m\tau = 1+\sqrt{-5}$. Now $2-0(1+\sqrt{-5}) = 2$, $(1+\sqrt{-5})-1(1+\sqrt{-5}) = 0$, $2\sqrt{-5}-2(1+\sqrt{-5}) = -2$, and $(-5+\sqrt{-5})-1(1+\sqrt{-5}) = -6$, and so $n = \gcd(2,0,-2,-6) = 2$. Thus $I = [\, 2, 1+\sqrt{-5}\,]$. We can show that this is a proper ideal, as if $1 \in I$ then $1 = r2 + s(1+\sqrt{-5})$ ($r,\ s \in \mathbb{Z}$) and so $s=0$ and $r = 1/2$ which is impossible.

Our next goal is to vindicate the study of ideals by proving that every ideal can be uniquely expressed as the product of prime ideals (we'll find out what a prime ideal is later). We start by defining a notion of "size" of an ideal. When considering elements of R_d, $|N(\alpha)|$, the absolute value of the norm of the element α gives an idea of the "size" of α—if $|N(\alpha)| = 0$ then $\alpha = 0$, if $|N(\alpha)| = 1$ then α is a unit etc. As the norm is defined by $N(\alpha) = \alpha\alpha^*$ we might consider the product II^* for an ideal I. The following result is the cornerstone of our theory of ideals.

Theorem 8 (Hurwitz' Lemma) *Suppose that $\alpha,\ \beta \in R_d$ and $g \in \mathbb{N}$. If $N(\alpha)$, $N(\beta)$ and $T(\alpha\beta^*)$ are all divisible by g then $g \mid \alpha\beta^*$ and $g \mid \alpha^*\beta$ in R_d.*

Proof Let $\gamma = \alpha\beta^*/g$, so that $\gamma^* = \alpha^*\beta/g$. We want to show that $\gamma \in R_d$ for then also $\gamma^* \in R_d$. We use the criterion of Proposition 4, namely that $\gamma \in R_d$ if and only if both $T(\gamma) \in \mathbb{Z}$ and $N(\gamma) \in \mathbb{Z}$. Now

$$T(\gamma) = \gamma + \gamma^* = \frac{\alpha\beta^* + \alpha^*\beta}{g} = \frac{T(\alpha\beta^*)}{g} \in \mathbb{Z}$$

and

$$N(\gamma) = \gamma\gamma^* = \frac{\alpha\beta^*\alpha^*\beta}{g^2} = \frac{\alpha\alpha^*}{g}\frac{\beta\beta^*}{g} = \frac{N(\alpha)}{g}\frac{N(\beta)}{g} \in \mathbb{Z}.$$

Hence $\gamma \in R_d$ and also $\gamma^* \in R_d$. $\square$

Corollary 3 *If I is an ideal of R_d then $II^* = \langle N \rangle$ for some integer $N \geq 0$.*

Proof We know that $I = \langle \alpha, \beta \rangle$ for some α and β and so

$$II^* = \langle \alpha, \beta \rangle \langle \alpha^*, \beta^* \rangle = \langle \alpha\alpha^*, \alpha\beta^*, \beta\alpha^*, \beta\beta^* \rangle.$$

Hence the rational integers $\alpha\alpha^* = N(\alpha)$, $\beta\beta^* = N(\beta)$ and $\alpha\beta^* + \alpha^*\beta = T(\alpha\beta^*)$ all lie in II^*. Let N be the greatest common divisor of these three integers. Clearly $N \in II^*$ and so $\langle N \rangle \subseteq II^*$. As $N \mid N(\alpha)$, $N \mid N(\beta)$ and $N \mid T(\alpha\beta^*)$ then by Hurwitz' Lemma $N \mid \alpha\beta^*$ and $N \mid \beta\alpha^*$. Hence $II^* \subseteq \langle N \rangle$ as required. □

The integer N above is uniquely defined and is called the *norm* of the ideal I. We denote the norm of I by $\mathbf{N}(I)$. The above result gives us the formula

$$\mathbf{N}(\langle \alpha, \beta \rangle) = \gcd(N(\alpha), N(\beta), T(\alpha\beta^*)).$$

If $I = \langle \alpha \rangle$ is principal then $II^* = \langle \alpha\alpha^* \rangle$ and so $\mathbf{N}(\langle \alpha \rangle) = |N(\alpha)|$. More generally if $\alpha \in I$ then $\alpha^* \in I^*$ and $N(\alpha) = \alpha\alpha^* \in II^* = \langle \mathbf{N}(I) \rangle$, and so $\mathbf{N}(I) \mid N(\alpha)$. Also the ideal norm is multiplicative as

$$\langle \mathbf{N}(IJ) \rangle = (IJ)(IJ)^* = IJI^*J^* = (II^*)(JJ^*) = \langle \mathbf{N}(I) \rangle \langle \mathbf{N}(J) \rangle = \langle \mathbf{N}(I)\mathbf{N}(J) \rangle$$

and so $\mathbf{N}(IJ) = \mathbf{N}(I)\mathbf{N}(J)$. If $\mathbf{N}(I) = 0$ then $N(\alpha) = \alpha\alpha^* = 0$ for all $\alpha \in I$ and so $I = \{0\}$; if $\mathbf{N}(I) = 1$, then $II^* = \langle 1 \rangle = R_d$ and so $I \supseteq II^* = R_d$ which means that $I = R_d$. Hence non-zero ideals have non-zero norm, and non-zero proper ideals have norm bigger than 1. This observation will be very useful!

We can link the definition of norm with group theory by means of the following result.

Proposition 10 *Let $I = [\, n, a + m\tau \,]$ be a non-zero* **ideal** *of R_d, with m, $n \in \mathbb{N}$ and $a \in \mathbb{Z}$. Then $\mathbf{N}(I) = mn$.*

Proof For convenience let $\alpha = a + m\tau$. Now $I = [\, n, \alpha \,]$ and so $I^* = [\, n, \alpha^* \,]$. Let $N = \mathbf{N}(I)$ so that $II^* = \langle N \rangle$. As $n\alpha = an + mn\tau \in I^*I = \langle N \rangle$ it follows that $N \mid an$ and $N \mid mn$. To show that $N = mn$ it suffices to show that $mn \mid N$.

I claim that every element of II^* has the form $A + Bmn\tau$ where A and B are integers. As every element of II^* is a sum of elements of the form $\beta\gamma$ where $\beta \in I$ and $\gamma \in I^*$ it suffices to show that all such elements have this form. Now $\beta = rn + s\alpha$ and $\gamma = un + v\alpha^*$ so

$$\beta\gamma = run^2 + rvn\alpha^* + sun\alpha + sv\alpha\alpha^*.$$

Hence it suffices to show that $n\alpha$ and $n\alpha^*$ have the form $A + Bmn\tau$. But $n\alpha = na + mn\tau$ and $n\alpha^* = T(n\alpha) - n\alpha = T(n\alpha) - na - mn\tau$, and the claim is established. Now $N\tau \in II^*$ and so $mn \mid N$ as required, concluding the proof. □

Recall from group theory the definition of the *index* $[\, G : H \,]$ of a subgroup H of a group G as the number of left (or right) cosets of H in G. It is not hard to show that the index of $I = [\, a + m\tau, n \,]$ in R_d is mn (the cosets are $(r + s\tau) + I$ where $0 \leq r < n$ and $0 \leq s < m$). Hence we get the formula, valid for non-zero ideals I,

$$\mathbf{N}(I) = [\, R_d : I \,]$$

which is taken by many books as the **definition** of norm.

The fact that II^* is always principal allows us to "cancel" by a non-zero ideal.

Proposition 11 *Let I, J and K be ideals of R_d with I non-zero and $IJ = IK$. Then $J = K$.*

Proof Suppose first that $I = \langle \alpha \rangle$ is principal. Then $IJ = \alpha J$ and so $J = \alpha^{-1}(IJ) = \{\alpha^{-1}\beta : \beta \in IJ\}$. Similarly $K = \alpha^{-1}(IK) = \alpha^{-1}(IJ) = J$.

Now suppose that I is any non-zero ideal. From $IJ = IK$ it follows that

$$(II^*)J = (IJ)I^* = (IK)I^* = (II^*)K$$

and as II^* is principal and nonzero, then by the previous paragraph it follows that $J = K$. □

We next investigate divisibility of ideals. We say that I divides J (and write $I \mid J$) if there is an ideal K with $J = IK$. If $I \mid J$ then $I \supseteq J$ as we can write $J = IK \subseteq I$. A crucial result is that the converse holds.

Theorem 9 *If I and J are non-zero ideals of R_d then $I \mid J$ if and only if $I \supseteq J$.*

Proof We have already noted that if $I \mid J$ then $I \supseteq J$. Conversely suppose that $I \supseteq J$. Now $JI^* \subseteq II^* = \langle \mathbf{N}(I) \rangle$. Then

$$K = \frac{1}{\mathbf{N}(I)} JI^* = \{\mathbf{N}(I)^{-1}\alpha : \alpha \in JI^*\} \subseteq R_d$$

and it is easy to see that K is an ideal. Now

$$IK = \frac{1}{\mathbf{N}(I)} I(JI^*) = \frac{1}{\mathbf{N}(I)} J(II^*) = \frac{1}{\mathbf{N}(I)} J\langle \mathbf{N}(I) \rangle = J$$

and so $I \mid J$, as required. □

This theorem is summarized by the maxim *to contain is to divide*. Learn it by heart!

Naturally we say that an ideal I is *irreducible* if I is non-zero, $I \neq \langle 1 \rangle$ and if $J \mid I$ then $J = I$ or $J = \langle 1 \rangle$. If I is a non-zero proper ideal that isn't irreducible, then we can write I as a product of two ideals of smaller norm. By repeating the process if necessary we can write I as a product of irreducible ideals. (If this argument doesn't convince you, write out a formal proof.) Each irreducible ideal I is also *maximal*. This means that the only ideals containing I are I itself and R_d. This follows as to contain is to divide. More importantly irreducible ideals are also *prime* i.e., if I is irreducible and $I \mid JK$ then either $I \mid J$ or $I \mid K$.

Theorem 10 *Let I, J and K be ideals of R_d with I irreducible. If $I \mid JK$ then either $I \mid J$ or $I \mid K$.*

Proof Suppose that $I \nmid J$. As to contain is to divide then $I \not\supseteq J$. The ideal $I + J$ contains I, and so divides I but cannot equal I for then $I = I + J \supseteq J$. Hence as I is irreducible we have $I + J = R_d$, and so $1 = \alpha + \beta$ for some $\alpha \in I$ and $\beta \in J$. If $\gamma \in K$ then $\gamma = (\alpha + \beta)\gamma = \alpha\gamma + \beta\gamma$. Now $\alpha\gamma \in I$ as $\alpha \in I$; also $\beta\gamma \in JK$ but as $I \mid JK$ then $I \supseteq JK$ and $\beta\gamma \in I$. It follows that $\gamma \in I$ and $I \supseteq K$, or equivalently $I \mid K$. □

An obvious extension of this result is that if I is irreducible and $I \mid J_1 J_2 \cdots J_k$ then $I \mid J_j$ for some j. From now on we use the term *prime ideal* as a synonym for irreducible ideal. We can now prove the most important result of the theory.

Theorem 11 (Unique Factorization) *If I is a non-zero proper ideal of R_d and*

$$I = P_1 P_2 \cdots P_r = Q_1 Q_2 \cdots Q_s$$

are two factorizations of A into prime ideals, then $r = s$ and we can re-order the Q_js so that $P_j = Q_j$ for all j.

Proof Use induction on r. If $r = 1$, I is irreducible and so $s = 1$ and $Q_1 = I = P_1$. Assume that $r > 1$. As $P_1 \mid Q_1 Q_2 \cdots Q_s$ then $P_1 \mid Q_j$ for some j. Relabel the Q_js so that $j = 1$, so now $P_1 \mid Q_1$. As Q_1 is prime then $P_1 = Q_1$. We cancel P_1 to get

$$P_2 P_3 \cdots P_r = Q_2 Q_3 \cdots Q_s$$

and using the inductive hypothesis now gives the result. □

Note that this proof is similar to, but easier than, the unique factorization theorem for elements when R_d is principal, as we don't have to worry about units.

To understand factorization into prime ideals we need to find out what the prime ideals are. The next result is the first step in this direction. It shows a connection between prime ideals and (ordinary) prime numbers.

Proposition 12 *If P is a prime ideal of R_d then $P \mid \langle p \rangle$ for some prime number p.*

Proof Certainly $P \mid PP^* = \langle N \rangle$ where $N = \mathbf{N}(P) > 1$ is P as non-zero and proper. Write $N = p_1 p_2 \cdots p_k$ with each p_j an (ordinary) prime number. Then $P \mid \langle p_1 \rangle \langle p_2 \rangle \cdots \langle p_k \rangle$, but as P is prime then $P \mid \langle p_j \rangle$ for some j and this establishes the proposition. □

This proposition means that to find all prime ideals all we need is to factorize all ideals of the form $\langle p \rangle$, where p is a prime number, into prime ideals. Now the ideal $\langle p \rangle$ has norm p^2, and so any proper ideal dividing it has norm p or p^2. This means that either $\langle p \rangle$ is itself prime, or $\langle p \rangle$ is the product of two prime ideals each of norm p. If $\langle p \rangle$ is already prime we say that p is *inert* in R_d, if $\langle p \rangle$ is the product of two equal prime ideals of norm p we say that p is *ramified* in R_d, and if $\langle p \rangle$ is the product of two distinct prime ideals we say that p is *split* in R_d. All three possibilities do occur: in R_{-1}, 3 is inert, 2 is ramified ($\langle 2 \rangle = \langle 1 + i \rangle^2$) and 5 is split ($\langle 5 \rangle = \langle 2 + i \rangle \langle 2 - i \rangle$).

To find out whether a given prime p is inert, ramified or split it suffices to count the number of ideals of norm p. For if $\mathbf{N}(P) = p$ then $PP^* = \langle p \rangle$ and $P \mid \langle p \rangle$. If p is inert there are no ideals of norm p. If p ramifies there is exactly one ideal of norm p, and if p splits there are two ideals of norm p. To classify ideals of norm p we employ Proposition 10.

Proposition 13 *Let p be a prime number, and let $P = [\, a + m\tau, n \,]$ where a, m and n are integers with $m > 0$ and $n > 0$. Then P is an ideal of norm p if and only if $m = 1$, $n = p$ and $p \mid N(a + \tau)$.*

Proof Suppose that P is an ideal of norm p. By Proposition 10 $mn = p$ and so either $m = 1$ and $n = p$, or $m = p$ and $n = 1$. But in the latter case $1 \in P$ so $P = R_d$ which does not have norm p. Hence $m = 1$ and $n = p$. Let $\alpha = a + \tau \in P$. Now $N(\alpha) = \alpha\alpha^* \in PP^* = \langle p \rangle$ so $p \mid N(\alpha)$.

Now suppose that $P = [\, a + \tau, p \,]$ with $p \mid N(a + \tau)$. If P is an ideal it will have norm p, and so it suffices to show that P is an ideal. As P is a subgroup of R_d all we need to show is that $\beta\gamma \in P$ whenever $\beta \in P$ and $\gamma \in R_d$. Now $\beta = rp + s(a + \tau)$ and $\gamma = u + v\tau$ with r, s, u and v integers, and so

$$\beta\gamma = rup + rvp\tau + su(a + \tau) + sv\tau(a + \tau).$$

As $p \in P$ and $a + \tau \in P$ it suffices to show that $p\tau \in P$ and $\tau(a + \tau) \in P$. But

$$p\tau = p(a + \tau) - ap \in P$$

and

$$\tau(a + \tau) = (T(\tau) - \tau^*)(a + \tau) = (a + T(\tau) - a - \tau^*)(a + \tau) = (a + T(\tau))(a + \tau) - N(a + \tau) \in P$$

as $p \mid N(a + \tau)$. This completes the proof. □

We can now find all ideals of prime norm p. It's easy to see that $[a+\tau, p] = [b+\tau, p]$ if and only if $a \equiv b \pmod{p}$. Hence the number of ideals of norm p is the number of distinct solutions modulo p of the congruence $N(a+\tau) \equiv 0 \pmod{p}$. What does this congruence look like? If $d \not\equiv 1 \pmod 4$ then $\tau = \sqrt{d}$ and $N(a+\tau) = a^2 - d$; if $d \equiv 1 \pmod 4$ then $\tau = \frac{1}{2}(1+\sqrt{d})$ and $N(a+\tau) = a^2 + a - \frac{1}{4}(d-1)$. Hence this congruence is quadratic, and we can use the theory of quadratic residues to count the number of solutions.

Proposition 14 *(i) If p is an odd prime then p is inert, ramified or split in R_d according to whether the Legendre symbol $\left(\frac{d}{p}\right) = -1$, 0 or 1.*

(ii) If $p = 2$ then p is ramified in R_d if $d \not\equiv 1$ (mod 4), is inert in R_d if $d \equiv 5 \pmod 8$, and split in R_d if $d \equiv 1 \pmod 8$.

Proof (i) Suppose first that $d \not\equiv 1 \pmod 4$ so that $\tau = \sqrt{d}$. Then the congruence is

$$N(a+\tau) = a^2 - d \equiv 0 \pmod{p}$$

which has zero, one or two solutions according to whether $\left(\frac{d}{p}\right) = -1$, 0 or 1. On the other hand if $d \equiv 1 \pmod 4$ then $\tau = \frac{1}{2}(1+\sqrt{d})$ and the congruence is and $N(\tau) = (1-d)/4$.

$$N(a+\tau) = a^2 + a - \frac{d-1}{4} \equiv 0 \pmod{p},$$

or equivalently

$$4a^2 + 4a + 1 - d \equiv 0 \pmod{p}.$$

This reduces to $b^2 \equiv d \pmod{p}$ where $b = 2a+1$. We thus have $1 + \left(\frac{d}{p}\right)$ solutions for b modulo p giving rise to $1 + \left(\frac{d}{p}\right)$ solutions for a modulo p.

(ii) Let $p = 2$. If $d \not\equiv 1 \pmod 4$ then the congruence is

$$a^2 \equiv d \pmod 2$$

which has the unique solution $a \equiv 0 \pmod 2$ if d is even and the unique solution $a \equiv 1 \pmod 2$ if d is odd. If $d \equiv 1 \pmod 4$ then the congruence is

$$a^2 + a + \frac{1-d}{4} \equiv 0 \pmod 2.$$

If $d \equiv 5 \pmod 8$ then $(1-d)/4$ is odd and this congruence is insoluble. If $d \equiv 1 \pmod 8$ then $(1-d)/4$ is even and this congruence has the solutions $a \equiv 0$ and $a \equiv 1 \pmod 2$. □

We have the easy corollary.

Corollary 4 *In each R_d only a finite number of primes are ramified.*

Proof The only ramified primes are the odd prime factors of d, and possibly also 2. □

In more advanced algebraic number theory ramification theory becomes very important.

7 Ideal Classes and the Class Group

If every ideal of R_d is principal, then R_d has unique factorization. But more often than not R_d has non-principal ideals. It is easy to classify principal ideals, but can we do the same for non-principal ideals? The notion of *equivalence* of ideals gives us a handle on this problem.

We say that two non-zero ideals I and J of R_d are *equivalent* if $I = \lambda J$ for some $\lambda \in \mathbb{Q}(\sqrt{d})^*$. We write $I \sim J$ if I and J are equivalent. It is an easy exercise to show that $\sim$ is an equivalence relation: i.e., $I \sim I$ for all I, $I \sim J$ implies that $J \sim I$, and $I \sim J \sim K$ implies that $I \sim K$. We write the equivalence class of I as $[\,I\,]$, i.e.,

$$[\,I\,] = \{J : I \sim J\}$$

and we call $[\,I\,]$ an *ideal class* of R_d. The ideal class $[\,R_d\,]$ is the collection of all principal ideals. Hence R_d is principal if and only if R_d has exactly one ideal class.

Equivalence of ideals is respected by multiplication: if $I \sim I'$ and $J \sim J'$ then $IJ \sim I'J'$, as $I' = \alpha I$ and $J' = \beta J$, $(\alpha, \beta \in \mathbb{Q}(\sqrt{d})^*)$ and so $I'J' = (\alpha\beta)IJ$. This means that the ideal class $[\,IJ\,]$ depends only on the ideal classes $[\,I\,]$ and $[\,J\,]$. We can therefore define multiplication of ideal classes by $[\,I\,][\,J\,] = [\,IJ\,]$. This multiplication is obviously commutative and associative. As $[\,I\,][\,R_d\,] = [\,I\,]$ and $[\,I\,][\,I^*\,] = [\,\langle \mathbf{N}(I) \rangle\,] = [\,R_d\,]$ the set $\mathrm{Cl}(R_d)$ of ideal classes of R_d forms an Abelian group—the identity being the class of principal ideals, and inversion being the operation of conjugation. We call $\mathrm{Cl}(R_d)$ the *class group* of R_d.

The most important fact about $\mathrm{Cl}(R_d)$ is that it is a **finite** group. The size of this group shows how far away R_d is from having unique factorization. To prove that the class group is finite we need the following lemma, which shows that in a particular R_d the minimum size of an element of an ideal I isn't much bigger than $\mathbf{N}(I)$.

Lemma 8 *If d is a squarefree integer and $d \neq 1$ then there is a constant C_d, depending only on d, such that given any non-zero ideal I of R_d there is a non-zero element α of I with $|N(\alpha)| \leq C_d \mathbf{N}(I)$.*

Proof Let $N = \mathbf{N}(I)$ and choose K to be the biggest integer with $K \leq \sqrt{N}$. Then $K^2 \leq N < (K+1)^2$. Write $I = [\,a + m\tau, n\,]$ with a, m, $n \in \mathbb{Z}$ and $m > 0$, $n > 0$. Then by Proposition 10 $N = mn$. Consider the set

$$S = \{r + s\tau : r, s \in \mathbb{Z}, 0 \leq r \leq K, 0 \leq s \leq K\}.$$

This set has $(K+1)^2$ elements, and $(K+1)^2 > mn$. Define subsets S_0, $S_1, \ldots, S_{m-1}$ of S by

$$S_j = \{r + s\tau \in S : s \equiv j \pmod{m}\}.$$

As S is the union of the S_js at least one of these sets must have more than n elements. Suppose S_j is such a subset so

$$S_j = \{r_1 + s_1\tau, r_2 + s_2\tau, \ldots, r_k + s_k\tau\}$$

where $k > n$. Put $t_i = r_i - a(s_i - j)/m$. Note that t_j is an integer as $s_i \equiv j \pmod{m}$. As $k > n$ there must exist $i_1 < i_2$ with $t_{i_1} \equiv t_{i_2} \pmod{n}$. Let $\alpha = r_{i_1} + s_{i_1}\tau$ and $\beta = r_{i_2} + s_{i_2}\tau$. Then $\alpha \neq \beta$, α, $\beta \in S$ and

$$\begin{aligned}
\alpha - \beta &= r_{i_1} - r_{i_2} + (s_{i_1} - s_{i_2})\tau \\
&= r_{i_1} - r_{i_2} + (s_{i_1} - j - s_{i_2} + j)\tau \\
&= r_{i_1} - r_{i_2} + (s_{i_1} - j - s_{i_2} + j)\frac{a + m\tau}{m} - \frac{a}{m}((s_{i_1} - j) - (s_{i_2} - j)) \\
&= \frac{t_{i_1} - t_{i_2}}{n} n + \frac{s_{i_1} - s_{i_2}}{m}(a + m\tau) \in I.
\end{aligned}$$

We must estimate the size of $\gamma = \alpha - \beta$. Now $\gamma = u + v\tau$ where $|u|$, $|v| \leq K$, and so

$$\begin{aligned}
|N(u + v\tau)| &= |(u + v\tau)(u + v\tau^*)| \\
&= |u^2 + uv(\tau + \tau^*) + v^2\tau\tau^*| \\
&\leq u^2 + |uvT(\tau)| + v^2|N(\tau)| \\
&\leq K^2(1 + |T(\tau)| + |N(\tau)|) \leq C_d \mathbf{N}(I)
\end{aligned}$$

if we put $C_d = 1 + |T(\tau)| + |N(\tau)|$. $\square$

(One can accelerate the proof of the above Lemma by using the idea of the index of a subgroup—I may expand on this in lectures.)

We can now show the finiteness of $\mathrm{Cl}(R_d)$.

Proposition 15 *If C_d is as in the above Lemma, then every ideal class in $\mathrm{Cl}(R_d)$ contains an ideal of norm $\leq C_d$.*

Proof Let I be an ideal, and consider its class $[\,I\,]$. By the Lemma the ideal I^* contains a non-zero element α with $|N(\alpha)| \leq C_d \mathbf{N}(I^*)$. As $\langle\alpha\rangle \subseteq I^*$ then $I^* \mid \langle\alpha\rangle$ and we can write $\langle\alpha\rangle = I^*J$ for an ideal J. As $[\,I^*\,][\,J\,] = [\,\langle\alpha\rangle\,] = [\,R_d\,]$ it follows that the class $[\,J\,]$ is the inverse of $[\,I^*\,]$ which is the inverse of $[\,I\,]$ and so $[\,I\,] = [\,J\,]$. Now $|N(\alpha)| = \mathbf{N}\langle\alpha\rangle = \mathbf{N}(I^*)\mathbf{N}(J)$, and so $\mathbf{N}(I^*)\mathbf{N}(J) \leq C_d\mathbf{N}(I^*)$ and $\mathbf{N}(J) \leq C_d$ as required. $\square$

Corollary 5 *The group $\mathrm{Cl}(R_d)$ is finite.*

Proof By the Proposition every ideal class in R_d has the form $[\,I\,]$ where $\mathbf{N}(I) \leq C_d$. But for each particular norm N there are only a finite number of ideals of norm N, as such an ideal must be a product of prime ideals of norm p or p^2 where p runs through the prime factors of N. As there are only finitely many such prime ideals, there are only finitely many ideals of norm N. Hence there are only finitely many ideals of norm $\leq C_d$ and so there are only finitely many ideal classes. $\square$

We call the order of $\mathrm{Cl}(R_d)$ the *class number* of R_d and denote it as $h(d)$. To compute $\mathrm{Cl}(R_d)$ and so $h(d)$ one finds all prime ideals of norm $\leq C_d$, then uses these to find all ideals of norm $\leq C_d$, and then decides which of these are equivalent. Once we have done this we have found all ideal classes in R_d. In practice one finds that there are usually a lot of ideals of norm $\leq C_d$ and this procedure takes a long time. However the following theorem, due to Minkowski, gives a lower value for the constant C_d and so enables us to calculate the class group faster.

Theorem 12 (Minkowski) *Let d be a squarefree integer different from 1. Let $D = d$ if $d \equiv 1 \pmod 4$ and $D = 4d$ if $d \not\equiv 1 \pmod 4$. Define*

$$M_d = \begin{cases} \dfrac{\sqrt{D}}{2} & \text{if } d > 0, \\ \dfrac{2\sqrt{|D|}}{\pi} & \text{if } d < 0. \end{cases}$$

Then every non-zero ideal I of R_d contains a non-zero element α with $|N(\alpha)| \leq M_d\mathbf{N}(I)$.

Proof Omitted. See [1] or [3]. The proof isn't hard but it involves a technique—Geometry of Numbers—which I haven't developed. $\square$

References

[1] H. Cohn, *Advanced Number Theory* (a.k.a. *A Second Course in Number Theory*), Dover, 1962, 1980

[2] I.N. Stewart, *Galois Theory*, Chapman and Hall, 1973, 1989

[3] I.N. Stewart & D.O. Tall, *Algebraic Number Theory*, Chapman and Hall, 1979