The arithmetic of quaternion algebras

John Voight
jvoight@gmail.com

June 18, 2014

Preface

Goal

In the response to receiving the 1996 Steele Prize for Lifetime Achievement [Ste96], Shimura describes a lecture given by Eichler:

> [T]he fact that Eichler started with quaternion algebras determined his course thereafter, which was vastly successful. In a lecture he gave in Tokyo he drew a hexagon on the blackboard and called its vertices clockwise as follows: automorphic forms, modular forms, quadratic forms, quaternion algebras, Riemann surfaces, and algebraic functions.

This book is an attempt to fill in the hexagon sketched by Eichler and to augment it further with the vertices and edges that represent the work of many algebraists, geometers, and number theorists in the last 50 years.

Quaternion algebras sit prominently at the intersection of many mathematical subjects. They capture essential features of noncommutative ring theory, number theory, K-theory, group theory, geometric topology, Lie theory, functions of a complex variable, spectral theory of Riemannian manifolds, arithmetic geometry, representation theory, the Langlands program—and the list goes on. Quaternion algebras are especially fruitful to study because they often reflect some of the general aspects of these subjects, while at the same time they remain amenable to concrete argumentation.

With this in mind, we have two goals in writing this text. First, we hope to introduce a large subset of the above topics to graduate students interested in algebra, geometry, and number theory. We assume that students have been exposed to some algebraic number theory (e.g., quadratic fields), commutative algebra (e.g., module theory, localization, and tensor products), as well as the basics of linear algebra, topology, and complex analysis. For certain sections, further experience with objects in arithmetic geometry, such as elliptic curves, is useful; however, we have endeavored to present the material in a way that is motivated and full of rich interconnections and examples, so that the reader will be encouraged to review any prerequisites with these examples in mind and solidify their understanding in this way. At the moment, one can find introductions for aspects of quaternion algebras taken individually, but there is no text that brings them together in one place and that draws the connections between them; we have tried to fill this gap. Second, we have written this text for researchers in these areas: we have collected results otherwise scattered in the literature, provide some clarifications and corrections and complete proofs in the hopes that this text will provide a convenient reference in the future.

In order to combine these features, we have opted for an organizational pattern that is "horizontal" rather than "vertical": the text has many chapters, each representing a different slice of the theory. Each chapter could be used in a (long) seminar afternoon or could fill a few hours of a semester course. To the extent possible, we have tried to make the chapters stand on their own (with explicit references to results used from previous chapters) so that they can be read based on the reader's interests—hopefully the interdependence of the material will draw the reader in more deeply! The introductory section of each chapter contains motivation and a summary of the results contained therein, and we often restrict the level of generality and make simplifying hypotheses so that the main ideas are made plain. Hopefully the reader who is new to the subject will find these helpful as way to dive in.

This book has three other features. First, as is becoming more common these days, paragraphs are numbered when they contain results that are referenced later on; we have opted not to put these always in a labelled environment (definition, theorem, proof, etc.) to facilitate the expositional flow of ideas, while at the same time we wished to remain precise about where and how results are used. Second, we have included in each chapter a section on "extensions and further reading", where we have indicated some of the ways in which the author's (personal) choice of presentation of the material naturally connects with the rest of the mathematical landscape. Our general rule (except the historical expository in Chapter 1) has been to cite specific results and proofs in the text where they occur, but to otherwise exercise restraint until this final section where we give tangential remarks, more general results, additional references, etc. Finally, in many chapters we have also included a section on algorithmic aspects, for those who want to pursue the computational side of the theory. And as usual, each section also contains a number of exercises at the end, ranging from checking basic facts used in a proof to more difficult problems that stretch the reader. For many of these exercises, there are hints at the end of the book; for any result that is used later, a complete argument is given.

Acknowledgements

This book began as notes from a course offered at McGill University in the Winter 2010 semester, entitled *Computational Aspects of Quaternion Algebras and Shimura Curves*. I would like to thank the members of my Math 727 class for their invaluable discussions and corrections: Dylan Attwell-Duval, Xander Faber, Luis Finotti, Andrew Fiori, Cameron Franc, Adam Logan, Marc Masdeu, Jungbae Nam, Aurel Page, Jim Parks, Victoria de Quehen, Rishikesh, Shahab Shahabi, and Luiz Takei. This course was part of the special thematic semester *Number Theory as Applied and Experimental Science* organized by Henri Darmon, Eyal Goren, Andrew Granville, and Mike Rubinstein at the Centre de Recherche Mathématiques (CRM) in Montréal, Québec, and the extended visit was made possible by the generosity of Dominico Grasso, dean of the College of Engineering and Mathematical Sciences, and Jim Burgmeier, chair of the Department of Mathematics and Statistics, at the University of Vermont. With gratitude, I acknowledge their support.

The writing continued while the author was on sabbatical at the University of California, Berkeley, sponsored by Ken Ribet. Several students attended these lectures and gave helpful feedback: Watson Ladd, Andrew Niles, Shelly Manber, Eugenia Rosu, Emmanuel Tsukerman, Victoria Wood, and Alex Youcis. My sabbatical from Dartmouth College for the Fall 2013 and Winter 2014 quarters was made possible by the efforts of Associate Dean David Kotz, and I thank him for his support. Further progress on the text was made in preparation for a minicourse on Brandt modules as part of *Minicourses on Algebraic and Explicit Methods in Number Theory*, organized by Cécile Armana and Christophe Delaunay at the Laboratoire de Mathématiques de Besançon in Salins-les-Bains, France.

It is somehow fitting that I would find myself composing this text while a faculty member at Dartmouth College, as the story of the quaternions has even woven its way into the history of mathematics at Dartmouth. The only mathematical output by a Dartmouth professor in the 19th century was by Arthur Sherburne Hardy, Ph.D., the author of an 1881 text on quaternions entitled *Elements of quaternions* [Har1881]. Brown describes it as “an adequate, if not inspiring text. It was something for Dartmouth to offer a course in such an abstruse field, and the course was actually given a few times when a student and an instructor could be found simultaneously” [Bro61, p. 2]. On that note, many thanks go the participants in my Math 125 *Quaternion algebras* class at Dartmouth in Spring 2014: Daryl Deford, Tim Dwyer, Zeb Engberg, Michael Firrisa, Jeff Hein, Nathan McNew, Jacob Richey, Tom Shemanske, Scott Smedinghoff, and David Webb. (I can only hope that this book will receive better reviews!)

Many thanks go to the others who offered helpful comments and corrections: France Dacar, Ariyan Javanpeykar, BoGwang Jeon, Chan-Ho Kim, Chao Li, Ben-

jamin Linowitz, Nicole Sutherland, Joe Quinn, and Jiangwei Xue. The errors and omissions that remain are, of course, my own.

I am profoundly grateful to those who offered their encouragement at various times during the writing of this book: Srinath Baba, Chantal David, Matthew Greenberg and Kristina Loeschner, David Michaels, and my mother Connie Voight. Thank you all! Finally, I would like to offer my deepest gratitude to my partner Brian Kennedy—this book would not have been possible without his patience and enduring support.

To do

[[Sections that are incomplete, or comments that need to be followed up on, are in red.]]

Contents

Part I

Algebra

Chapter 1

Introduction

In this chapter, we give an overview of the topics contained in this book. We follow the historical arc of quaternion algebras and see in broad stroke how they have impacted the development of many areas of mathematics. This account is selective and is mostly culled from existing historical surveys; two very nice surveys of quaternion algebras and their impact on the development of algebra are those by Lam [Lam03] and Lewis [Lew06].

1.1 Hamilton's quaternions

In perhaps the most famous act of mathematical vandalism, on October 16, 1843, Sir William Rowan Hamilton carved the following equations into the Brougham Bridge (now Broomebridge) in Dublin:

$$i^2 = j^2 = k^2 = ijk = -1. \tag{1.1.1}$$

His discovery of these multiplication laws was a defining moment in the history of algebra.

For at least ten years, Hamilton had been attempting to model three-dimensional space with a structure like the complex numbers, whose addition and multiplication model two-dimensional space. Just like the complex numbers had a "real" and "imaginary" part, so too did Hamilton hope to find an algebraic system whose elements had a "real" and two-dimensional "imaginary" part. His son William Edward Hamilton, while still very young, would pester his father [Ham67, p. xv]: "Well, papa, can you multiply triplets?" To which Hamilton would reply, with a sad shake of the head, "No, I can only add and subtract them." (For a history of the "multiplying triplets" problem—the nonexistence of division algebra over the reals of dimension 3—see May [May66, p. 290].)

Figure 1.1: Sir William Rowan Hamilton (1805–1865)

Then, on this dramatic day in 1843, Hamilton's had a flash of insight [Ham67, p. xx–xxvi]:

> On the 16th day of [October]—which happened to be a Monday, and a Council day of the Royal Irish Academy—I was walking in to attend and preside, and your mother was walking with me, along the Royal Canal, to which she had perhaps driven; and although she talked with me now and then, yet an under-current of thought was going on in my mind, which gave at last a result, whereof it is not too much to say that I felt at once the importance. An electric circuit seemed to close; and a spark flashed forth, the herald (as I foresaw, immediately) of many long years to come of definitely directed thought and work, by myself if spared, and at all events on the part of others, if I should even be allowed to live long enough distinctly to communicate the discovery. Nor could I resist the impulse—unphilosophical as it may have been—to cut with

> a knife on a stone of Brougham Bridge, as we passed it, the fundamental formula with the symbols, i, j, k; namely,
>
> $$i^2 = j^2 = k^2 = ijk = -1$$
>
> which contains the Solution of the Problem.

In this moment, Hamilton realized that he needed a fourth dimension, and so he coined the term *quaternions* for the real space spanned by the elements $1, i, j, k$, subject to his multiplication laws 1.1.1. He presented this theory to the Royal Irish Academy in a paper entitled "On a new Species of Imaginary Quantities connected with a theory of Quaternions" [Ham1843]. For more, there are several extensive, detailed accounts of this history of quaternions [Dic19, vdW76]. Although his carvings have long since worn away, a plaque on the bridge now commemorates this historically significant event. This magnificent story remains in the popular consciousness, and to commemorate Hamilton's discovery of the quaternions, there is an annual "Hamilton walk" in Dublin [ÓCa10].

Although Hamilton was undoubtedly responsible for advancing the theory of quaternion algebras, there are several precursors to his discovery that bear mentioning. First, the quaternion multiplication laws are already implicitly present in the four-square identity of Leonhard Euler:

$$\begin{aligned}(a_1^2 + a_2^2 + a_3^2 + a_4^2)(b_1^2 + b_2^2 + b_3^2 + b_4^2) &= c_1^2 + c_2^2 + c_3^2 + c_4^2 = \\ &(a_1b_1 - a_2b_2 - a_3b_3 - a_4b_4)^2 + (a_1b_2 + a_2b_1 + a_3b_4 - a_4b_3)^2 \\ &+ (a_1b_3 - a_2b_4 + a_3b_1 + a_4b_2)^2 + (a_1b_4 + a_2b_3 - a_3b_2 + a_4b_1)^2.\end{aligned} \tag{1.1.2}$$

Indeed, the multiplication law in the quaternion reads precisely

$$(a_1 + a_2i + a_3j + a_4k)(b_1 + b_2i + b_3j + b_4k) = c_1 + c_2i + c_3j + c_4k.$$

It was perhaps Carl Friedrich Gauss who first observed this connection. In a note dated around 1819 [Gau00], he interpreted the formula (1.1.2) as a way of composing real quadruples: to the quadruples (a_1, a_2, a_3, a_4) and (b_1, b_2, b_3, b_4) in $\mathbb{R}^4$, he defined the composite tuple (c_1, c_2, c_3, c_4) and noted the noncommutativity of this operation. Gauss elected not to publish these findings, as he was afraid of the unwelcome reception that the idea might receive. (In letters to De Morgan [Grav1885, Grav1889, p. 330, p. 490], Hamilton attacks the allegation that Gauss had discovered quaternions first.) Finally, Olinde Rodrigues (1795–1851) (of the *Rodrigues formula* for Legendre polynomials) gave a formula for the angle and axis of a rotation in $\mathbb{R}^3$ obtained from two successive rotations—essentially giving a different parametrization of the quaternions—but had left mathematics for banking long before the publication of his paper [Rod1840]. The story of Rodrigues and the quaternions is given by

Altmann [Alt89] and Pujol [Puj12] and the fuller story of his life by Altmann–Ortiz [AO05].

In any case, the quaternions consumed the rest of Hamilton's academic life and resulted in the publication of two treatises [Ham1853, Ham1866] (see also the review [Ham1899]). Hamilton's writing over these years became increasingly obscure, and many found his books to be impenetrable. Nevertheless, many physicists used quaternions extensively and for a long time in the mid-19th century, quaternions were an essential notion in physics. Hamilton endeavored to set quaternions as the standard notion for vector operations in physics as an alternative to the more general dot product and cross product introduced in 1881 by Willard Gibbs (1839–1903), building on remarkable but largely ignored work of Hermann Grassmann (1809–1877) [Gras1862]. The two are related by the beautiful equality

$$vw = v \cdot w + v \times w \tag{1.1.3}$$

for $v, w \in \mathbb{R}i + \mathbb{R}j + \mathbb{R}k$, relating quaternionic multiplication to dot and cross products. This rivalry between physical notation flared into a war in the latter part of the 19th century between the 'quaternionists' and the 'vectorists', and for some the preference of one system versus the other became an almost partisan split. On the side of quaternions, James Clerk Maxwell (1831–1879) (responsible for the equations which describe electromagnetic fields) wrote [Max1869, p. 226]:

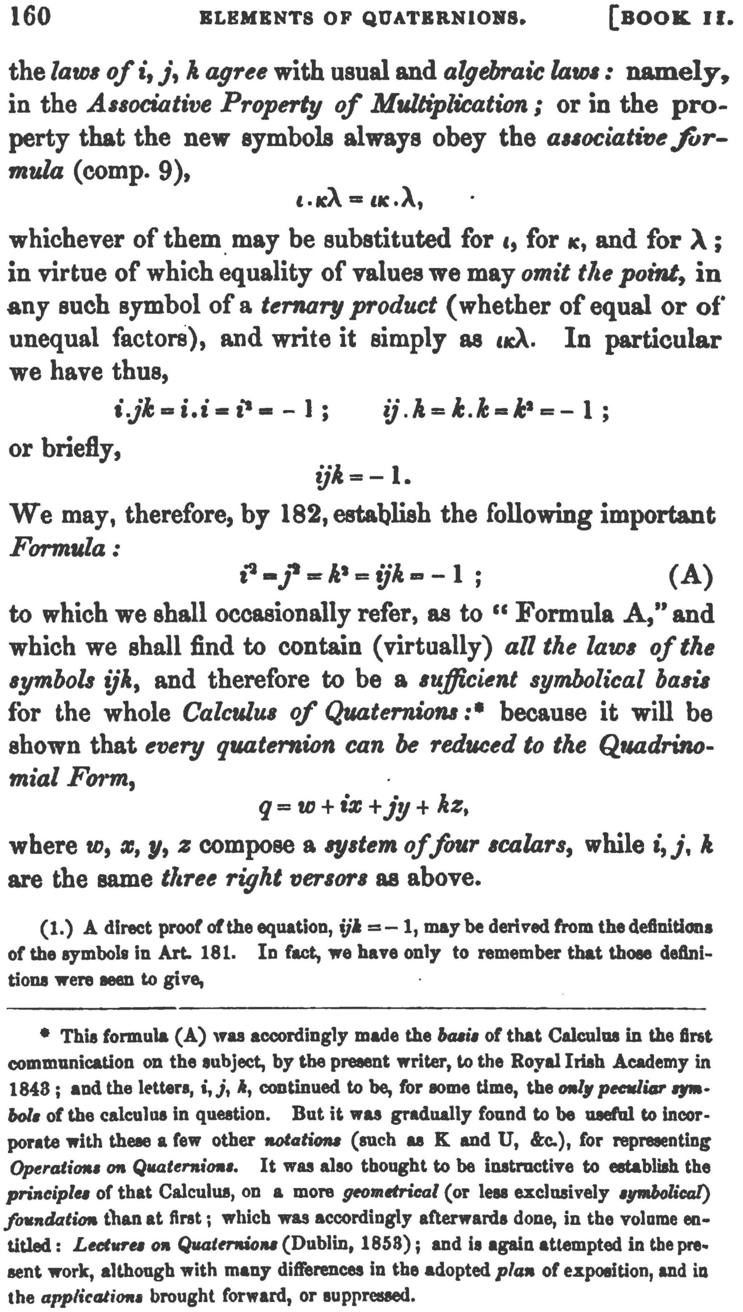
> The invention of the calculus of quaternions is a step towards the knowledge of quantities related to space which can only be compared, for its importance, with the invention of triple coordinates by Descartes. The ideas of this calculus, as distinguished from its operations and symbols, are fitted to be of the greatest use in all parts of science.

And Peter Tait (1831–1901), one of Hamilton's students, wrote in 1890 [Tai1890]:

> Even Prof. Willard Gibbs must be ranked as one the retarders of quaternions progress, in virtue of his pamphlet on *Vector Analysis*, a sort of hermaphrodite monster, compounded of the notation of Hamilton and Grassman.

On the vectorist side, Lord Kelvin (a.k.a. William Thomson, who formulated of the laws of thermodynamics), said in an 1892 letter to R. B. Hayward about his textbook in algebra (quoted in Thompson [Tho10, p. 1070]):

> Quaternions came from Hamilton after his really good work had been done; and, though beautifully ingenious, have been an unmixed evil to those who have touched them in any way, including Clerk Maxwell.



160 ELEMENTS OF QUATERNIONS. [BOOK II.

the *laws of* **i, j, k** *agree* with usual and *algebraic laws:* namely, in the *Associative Property of Multiplication;* or in the property that the new symbols always obey the *associative formula* (comp. 9),

$$\iota . \kappa\lambda = \iota\kappa . \lambda,$$

whichever of them may be substituted for ι, for κ, and for λ; in virtue of which equality of values we may *omit the point*, in any such symbol of a *ternary product* (whether of equal or of unequal factors), and write it simply as $\iota\kappa\lambda$. In particular we have thus,

$$i.jk = i.i = i^2 = -1; \qquad ij.k = k.k = k^2 = -1;$$

or briefly,

$$ijk = -1.$$

We may, therefore, by 182, establish the following important *Formula:*

$$i^2 = j^2 = k^2 = ijk = -1; \qquad \text{(A)}$$

to which we shall occasionally refer, as to "Formula A," and which we shall find to contain (virtually) *all the laws of the symbols ijk*, and therefore to be a *sufficient symbolical basis* for the whole *Calculus of Quaternions:*[*] because it will be shown that *every quaternion can be reduced to the Quadrinomial Form*,

$$q = w + ix + jy + kz,$$

where w, x, y, z compose a *system of four scalars*, while i, j, k are the same *three right versors* as above.

(1.) A direct proof of the equation, $ijk = -1$, may be derived from the definitions of the symbols in Art. 181. In fact, we have only to remember that those definitions were seen to give,

[*] This formula (A) was accordingly made the *basis* of that Calculus in the first communication on the subject, by the present writer, to the Royal Irish Academy in 1843; and the letters, *i, j, k*, continued to be, for some time, the *only peculiar symbols* of the calculus in question. But it was gradually found to be useful to incorporate with these a few other *notations* (such as K and U, &c.), for representing *Operations on Quaternions*. It was also thought to be instructive to establish the *principles* of that Calculus, on a more *geometrical* (or less exclusively *symbolical*) *foundation* than at first; which was accordingly afterwards done, in the volume entitled: *Lectures on Quaternions* (Dublin, 1853); and is again attempted in the present work, although with many differences in the adopted *plan* of exposition, and in the *applications* brought forward, or suppressed.

Figure 1.2: A page from Hamilton's *Elements of quaternions*

Ultimately, the superiority of vector notation carried the day, and only certain useful fragments of Hamilton's quaternionic notation remain in modern usage. For more on the history of quaternionic and vector calculus, see Crowe [Cro64] and Simons [Sim10].

The debut of the quaternions by Hamilton was met with some resistance in the mathematical world: it proposed a system of "numbers" that did not satisfy the usual commutative rule of multiplication. Quaternions predated the notion of matrices, introduced in 1855 by Arthur Cayley (1821-1895). Hamilton's bold proposal of a noncommutative multiplication law was the harbinger of an array of algebraic structures. In the words of J.J. Sylvester [Syl1883, pp. 271–272]:

> In Quaternions (which, as will presently be seen, are but the simplest order of matrices viewed under a particular aspect) the example had been given of Algebra released from the yoke of the commutative principle of multiplication—an emancipation somewhat akin to Lobachevsky's of Geometry from Euclid's noted empirical axiom; and later on, the Peirces, father and son (but subsequently to 1858) had prefigured the universalization of Hamilton's theory, and had emitted an opinion to the effect that probably all systems of algebraical symbols subject to the associative law of multiplication would be eventually found to be identical with linear transformations of schemata susceptible of matriculate representation.

Indeed, with the introduction of the quaternions the floodgates of algebraic possibilities had been opened. See Happel [Hap80] for the early development of algebra following Hamilton's quaternions.

1.2 Algebra after the quaternions

Soon after he discovered his quaternions, Hamilton sent a letter [Ham1844] describing them to his friend John T. Graves (1806-1870). Graves replied on October 26, 1843, with his complements, but added:

> There is still something in the system which gravels me. I have not yet any clear views as to the extent to which we are at liberty arbitrarily to create imaginaries, and to endow them with supernatural properties. ... If with your alchemy you can make three pounds of gold, why should you stop there?

Following this line of inquiry, on December 26, 1843, Graves wrote to Hamilton that he had successfully generalized quaternions to the "octaves", now called *octonions*

$\mathbb{O}$, an algebra in eight dimensions, with which he was able to prove that the product of two sums of eight perfect squares is another sum of eight perfect squares, a formula generalizing (1.1.2). In fact, Hamilton first invented the term *associative* in 1844, around the time of his correspondence with Graves. Unfortunately for Graves, the octonions were discovered independently and published already in 1845 by Cayley [Cay1845], who often is credited for their discovery. (Even worse, the eight squares identity was also previously discovered by C. F. Degen.) For a more complete account of this story and the relationships between quaternions and octonions, see the survey article by Baez [Bae02], the article by van der Blij[vdB60], and the book by Conway–Smith [CS03].

In this way, Cayley was able to reinterpret the quaternions as arising from a *doubling process*, also called the *Cayley–Dickson construction*, which starting from $\mathbb{R}$ produces $\mathbb{C}$ then $\mathbb{H}$ then $\mathbb{O}$, taking the ordered, commutative, associative algebra $\mathbb{R}$ and progressively deleting one adjective at a time. So algebras were first studied over the real and complex numbers and were accordingly called *hypercomplex numbers* in the late 19th and early 20th century. And this theory flourished. In 1878, Ferdinand Frobenius (1849–1917) proved that the only finite-dimensional division associative algebras over $\mathbb{R}$ are $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{H}$ [Fro1878]. (This result was also proven independently by C.S. Peirce, the son of Benjamin Peirce, below.) (Much later, work by topologists culminated in the theorem of Bott–Milnor [BM58] and Kervaire [Ker58]: the only finite-dimensional division (not-necessarily- associative) algebras have dimensions $1, 2, 4, 8$. As a consequence, the sphere $\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : \|x\|^2 = 1\}$ has a trivial tangent bundle only when $n = 1, 2, 4, 8$.)

In another attempt to seek a generalization of the quaternions to higher dimension, William Clifford (1845–1879) developed a way to build algebras from quadratic forms in 1876 [Cli1878]. Clifford constructed what we now call a *Clifford algebra* associated to $V = \mathbb{R}^n$; it is an algebra of dimension 2^n containing V with multiplication induced from the relation $x^2 = -\|x\|^2$ for all $x \in V$. We have $C(\mathbb{R}^1) = \mathbb{C}$ and $C(\mathbb{R}^2) = \mathbb{H}$, so the Hamilton quaternions arise as a Clifford algebra, but $C(\mathbb{R}^3)$ is not the octonions. Nevertheless, the theory of Clifford algebras is tightly connected to the theory of normed division algebras. For more on the history of Clifford algebras, see Diek–Kantowski [DK95].

The study of division algebras gradually evolved, including work by Benjamin Peirce [Pei1882] originating from 1870 on *linear associative algebra*; therein, he provides a decomposition of an algebra relative to an idempotent. The notion of a *simple* algebra had been found and developed around this time by Élie Cartan (1869–1951). But it was Joseph Henry Maclagan Wedderburn (1882–1948) who was the first to find meaning in the structure of simple algebras over an arbitrary field, in many ways leading the way forward. The jewel of his 1908 paper [Wed08] is still foundational in the structure theory of algebras: a simple algebra (finite-dimensional over a field) is isomorphic to the matrix ring over a division ring. Wedderburn also proved that a finite division ring is a field, a result that like his structure theorem has inspired much mathematics. For more on the legacy of Wedderburn, see Artin [Art50].

Around this time, other types of algebras over the real numbers were also being investigated, the most significant of which were Lie algebras. In the seminal work of Sophus Lie (1842–1899), group actions on manifolds were understood by looking at this action infinitessimally; one thereby obtains a *Lie algebra* of vector fields that determines the local group action. The simplest nontrivial example of a Lie algebra is the cross product of two vectors, related to quaternion multiplication in (1.1.3): it defines, in fact, give a binary operation on $\mathbb{R}^3$, but now

$$i \times i = j \times j = k \times k = 0.$$

The Lie algebra "linearizes" the group action and is therefore more accessible. Wilhelm Killing (1847–1923) initiated the study of the classification of Lie algebras in a series of papers [Kil1888], and this work was completed by Cartan. For more on this story, see Hawkins [Haw00].

The first definition of an algebra over an arbitrary field seems to have been given by Leonard E. Dickson (1874–1954) [Dic03] (even though at first he still called the resulting object a *system of complex numbers* and later adopting the name *(linear) algebra*). In the early 1900s, Dickson developed this theory further and in particular was the first to consider quaternion algebras over a general field. First, he considered algebras in which every element satisfies a quadratic equation [Dic12], leading to multiplication laws for what he later called a *generalized quaternion algebra* [Dic14, Dic23]. Today, we no longer employ the adjective "generalized", and we can reinterpret this vein of Dickson's work as showing that every 4-dimensional central simple algebra is a quaternion algebra (over a field F with $\operatorname{char} F \neq 2$).

At this time, Dickson [Dic19] (giving also a complete history) wrote on earlier work of Hurwitz from 1888 [Hur1888], who asked for generalizations of the composition laws arising from sum of squares laws like that of Euler (1.1.2) for four squares and Cayley for eight squares: for which n does there exist an identity

$$(a_1^2 + \cdots + a_n^2)(b_1^2 + \cdots + b_n^2) = c_1^2 + \cdots + c_n^2$$

with c_i bilinear in x and y? He showed they only exist for $n = 1, 2, 4, 8$ variables (so in particular, there is no formula expressing the product of two sums of 16 squares as the sum of 16 squares), the result being tied back to his theory of algebras.

Biquaternion (Albert) algebras. A. Adrian Albert.

Class field theory Hasse principle (1920s), class field theory, Noether, arithmetic of hypercomplex number systems. Cyclic algebras, cyclic cross product.

As "twisted forms" of 2 × 2-matrices, quaternion algebras in many ways are like "noncommutative quadratic field extensions", and just as the quadratic fields $\mathbb{Q}(\sqrt{d})$ are wonderously rich, so too are their noncommutative analogues. In this way, quaternion algebras provide a natural place to do noncommutative algebraic number theory. A more general study would look at central simple algebras (see Reiner).

Fuchsian groups The quotient gives rise to a *Riemann surface*. Riemann.

Hypergeometric functions also give examples. Fuchs and his differential equations.

After all, how do you get discrete groups? Start with real matrices, go to rational matrices, then to integral matrices, then make a group. Allow yourself entries in a number field, consider the algebra generated, take integral elements, make the group. When is this discrete? Something like 4-dimensional object gives you quaternion algebras.

Modular forms. The basic example being the group $\mathrm{SL}_2(\mathbb{Z})$. Quaternion algebras give rise therefore to objects of interest in geometry and low-dimensional topology. Classical modular forms.

discovered by Deuring.

Especially Jacobi and the sums of 4 squares, something that also can be seen using quaternion algebras. How often is an integer a sum of squares, or more generally, represented by a quadratic form in 4 variables? The generating function is a modular form.

Automorphic forms. Then discovery by Poincaré "when he was walking on a cliff," apparently in 1886, as he reminisced in his *Science et Méthode*. Holomorphic (complex analytic) functions that are invariant with respect to these groups are very interesting to study ("automorphic functions"). Set of matrices that preserve a quadratic form.

Soon after, this was followed by Fricke and Klein, who were interested in subgroups of $\mathrm{PGL}_2(\mathbb{R})$ that act discretely on the upper half-plane, such as the group generated by the matrices

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \text{ and } \begin{pmatrix} \sqrt{2} & 1+\sqrt{3} \\ 1-\sqrt{3} & \sqrt{2} \end{pmatrix}.$$

They still used the language of quadratic forms? In this way, quaternion algebras are useful in group theory.

Then other groups, Hilbert modular forms.

1.3 Modern theory

Composition laws Picking up again, work of Brandt.

Hecke operators, the basis problem, and the trace formula

Then Eichler: theory of Hecke operators in the 1950s. Selberg and trace formula. Basis problem.

Modularity and elliptic curves

Work of Shimura: find examples of zeta functions that could be given. Theory of complex multiplication and modularity of elliptic curves. Galois representations

Abelian varieties. Quaternion algebras arise also as the endomorphism rings of elliptic curves, and indeed they are the only noncommutative endomorphism algebras of simple abelian varieties over fields by Albert's classification. So that justifies there study already. The Rosati involution figures prominently in this classification.

Algebras with involution

Composition algebras. Algebras with involutions: Knus, etc. Connects back to Lie theory.

Riemannian manifolds

Back to Riemann surfaces. Vignéras.

Three-dimensional groups, arithmetic, some results.

Algorithmic aspects. Computations and algorithms can be done; this gives modular symbols, Brandt matrices, and their generalizations.

Today, quaternions have seen a revival in computer modeling and animation as well as in attitude control of aircraft and spacecraft [Han06]. A rotation in $\mathbb{R}^3$ about an axis through the origin can be represented by a 3×3 orthogonal matrix with determinant 1. However, the matrix representation is redundant, as there are only four degrees of freedom in such a rotation (three for the axis and one for the angle). Moreover, to compose two rotations requires the product of the two corresponding matrices, which requires 27 multiplications and 18 additions in $\mathbb{R}$. Quaternions, on the other hand, represent this rotation with a 4-tuple, and multiplication of two quaternions takes only 16 multiplications and 12 additions in $\mathbb{R}$. (What about Euler angles?)

In physics, quaternions yield elegant expression for the Lorentz transformations, the basis of the modern theory of relativity. Originally Hamilton's motivation, so we have come full circle (with many loops in between). There has been renewed interest by topologists in understanding quaternionic manifolds and by physicists who seek a quaternionic quantum physics, and some physicists still hope they will obtain a deeper understanding of physical principles in terms of quaternions.

And so although much of Hamilton's quaternionic physics fell out of favor long ago, we have somehow come full circle in our elongated historical arc. The enduring role of quaternion algebras as a progenitor of a vast range of mathematics promises a rewarding ride for years to come.

1.4 Extensions and further reading

1.4.1. There are three main biographies written about the life of William Rowan Hamilton, a man sometimes referred to as "Ireland's greatest mathematician", by Graves [Grav1882, Grav1885, Grav1889] in three volumes, Hankins [Han80], and O'Donnell [O'Do83]. Numerous other shorter biographies have been written [Lan67, ÓCa00].

1.4.2. If B is an $\mathbb{R}$-algebra of dimension 3, then either B is commutative or B is isomorphic to the subring of upper triangular matrices in $\mathrm{M}_2(\mathbb{R})$ (and consequently has a standard involution; see Chapter 3). A similar statement holds for free R-algebras of rank 3 over a (commutative) domain R; see Levin [Lev13].

1.4.3. [[Ways to visualize the spin group [HFK94].]]

Exercises

1.1. Hamilton originally sought an associative multiplication law on $B = \mathbb{R} + \mathbb{R}i + \mathbb{R}j \simeq \mathbb{R}^3$ where $i^2 = -1$, so in particular $\mathbb{C} \subset B$. Show that this is impossible.

1.2. Hamilton sought a multiplication $* : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ that preserves length:

$$\|v\|^2\|w\|^2 = \|v * w\|^2$$

for $v, w \in \mathbb{R}^3$. Expanding out in terms of coordinates, such a multiplication would imply that the product of the sum of three squares over $\mathbb{R}$ is again the sum of three squares in $\mathbb{R}$. (Such a law holds for the sum of two squares, corresponding to the multiplication law in $\mathbb{R}^2 \simeq \mathbb{C}$: we have

$$(x^2 + y^2)(u^2 + v^2) = (xu - yv)^2 + (xv + yu)^2.)$$

However, show that such a formula for three squares is impossible, as it would imply an identity in the polynomial ring in 6 variables over $\mathbb{Z}$. *[Hint: Find a natural number that is the product of two sums of three squares which is not itself the sum of three squares.]*

1.3. Show that there is no way to give $\mathbb{R}^3$ the structure of a ring (with 1) in which multiplication distributes over scalar multiplication by $\mathbb{R}$ and every nonzero element has a (two-sided) inverse, as follows.

a) Suppose otherwise, and $\mathbb{R}^3 = D$ is equipped with a multiplication law. Show that every $x \in D$ satisfies a polynomial of degree at most 3 with coefficients in $\mathbb{R}$.

b) By consideration of irreducible factors, show that every $x \in D$ satisfies a (minimal) polynomial of degree 1.

c) Derive a contradiction from the fact that every nonzero element has a (two-sided) inverse.

Chapter 2

Beginnings

In this chapter, we define quaternion algebras over fields by giving a multiplication table, following Hamilton; we then consider the classical application of understanding rotations in $\mathbb{R}^3$.

2.1 Conventions

Throughout this chapter, let F be a field with $\operatorname{char} F \neq 2$; the case $\operatorname{char} F = 2$ is considered in detail in Chapter 5.

We assume throughout the text (unless otherwise stated) that all rings are associative, not necessarily commutative, with 1, and that ring homomorphisms preserve 1. An *algebra* over the field F is a ring B equipped with a homomorphism $F \to B$ such that the image of F lies in the center of B. If B is not the zero ring, then this map is necessarily injective and we identify F with its image. Equivalently, an F-algebra is an F-vector space that is also compatibly a ring.

A *homomorphism* of F-algebras is a ring homomorphism which restricts to the identity on F. An F-algebra homomorphism is necessarily F-linear. The *dimension* $\dim_F B$ of an F-algebra B is its dimension as an F-vector space. If B is an F-algebra then we denote by $\operatorname{End}_F(B)$ the endomorphism ring of all F-linear homomorphisms $B \to B$ (where ring multiplication is given by functional composition) and by $\operatorname{Aut}_F(B)$ the automorphism group of all F-algebra isomorphisms $B \xrightarrow{\sim} B$.

2.2 Quaternion algebras

In this section, we define quaternion algebras by giving a set of generators and relations.

Definition 2.2.1. An algebra B over F (with $\operatorname{char} F \neq 2$) is a *quaternion algebra* if there is an F-basis $1, i, j, k$ for B such that

$$i^2 = a, \ j^2 = b, \ \text{ and } \ k = ij = -ji \tag{2.2.2}$$

for some $a, b \in F^\times$.

The multiplication table for a quaternion algebra B is determined by the multiplication rules (2.2.2): for example, we have that

$$k^2 = (ij)^2 = (ij)(ij) = i(ji)j = i(-ij)j = -ab$$

and $j(ij) = (-ij)j = -bi$. This multiplication table is associative (Exercise 2.1), and we have $\dim_F B = 4$.

It will be useful to have a symbol for quaternion algebras. For $a, b \in F^\times$, we define $\left(\dfrac{a,b}{F}\right)$ to be the algebra over F with basis $1, i, j, k$ subject to the multiplication rules 2.2.2. Thus an F-algebra B is a quaternion algebra over F if and only if B is isomorphic (as an F-algebra) to $\left(\dfrac{a,b}{F}\right)$ for some a, b.

The map which interchanges i and j gives an isomorphism $\left(\dfrac{a,b}{F}\right) \cong \left(\dfrac{b,a}{F}\right)$, so Definition 2.2.1 is symmetric in a, b. (See also Exercise 2.4.)

If $K \supseteq F$ is a field extension of F, then we have a canonical isomorphism

$$\left(\frac{a,b}{F}\right) \otimes_F K \cong \left(\frac{a,b}{K}\right)$$

so Definition 2.2.1 is functorial in F (with respect to inclusion of fields).

Example 2.2.3. The $\mathbb{R}$-algebra $\mathbb{H} = \left(\dfrac{-1,-1}{\mathbb{R}}\right)$ is the ring of quaternions over the real numbers, discovered by Hamilton; we call $\mathbb{H}$ the ring of *Hamiltonians*.

Example 2.2.4. The ring $\mathrm{M}_2(F)$ of 2×2-matrices with coefficients in F is a quaternion algebra over F: indeed, there is an isomorphism $\left(\dfrac{1,1}{F}\right) \xrightarrow{\sim} \mathrm{M}_2(F)$ of F-algebras induced by

$$i \mapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \ j \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Indeed, if $F = \overline{F}$ is algebraically closed and B is a quaternion algebra over F, then necessarily $B \cong \mathrm{M}_2(F)$ (Exercise 2.4).

A quaternion algebra B is generated by the elements i, j by definition (2.2.2). However, exhibiting an algebra by generators and relations can be subtle, as the dimension of such an algebra is not a priori clear. But working with presentations is quite useful, so we consider the following lemma.

Lemma 2.2.5. *An F-algebra B is a quaternion algebra if and only if there exist generators $i, j \in B$ (as an F-algebra) satisfying*

$$i^2 = a, \ \ j^2 = b, \ \ \text{and} \ \ ij = -ji \tag{2.2.6}$$

with $a, b \in F^{\times}$.

In other words, once the relations (2.2.6) are satisfied for generators i, j, then automatically B has dimension 4 as an F-vector space.

Proof. It is necessary and sufficient to prove that the elements $1, i, j, ij$ are linearly independent. So suppose that $t + xi + yj + zij = 0$ with $t, x, y, z \in F$ not all zero. The map $i \mapsto -i$ (and $j \mapsto j$) is an automorphism of B as an F-algebra, since it preserves the relations 2.2.6. Applying this automorphism, we obtain $t - xi + yj - zij = 0$, and adding we get $2(t + yj) = 0$. Since $\operatorname{char} F \neq 2$, this gives $t + yj = 0$. If $y \neq 0$, then $j \in F$ so lies in the center, contradicting $ij = -ji$; so $y = 0$. In a similar way, the automorphism $j \mapsto -j$ yields $t + xi = 0$ so $x = 0$, and their composition gives $t + zij = 0$ so $z = 0$. Thus $t = 0$ as well. □

Accordingly, we will call elements $i, j \in B$ satisfying (2.2.6) *standard generators* for a quaternion algebra B.

Remark 2.2.7. Invertibility of both a and b in F is needed for Lemma 2.2.5: the commutative algebra $B = F[i, j]/(i, j)^2$ is generated by the elements i, j that satisfy $i^2 = j^2 = ij = -ji = 0$ but B is *not* a quaternion algebra.

2.2.8. Every quaternion algebra $B = \left(\dfrac{a, b}{F}\right)$ can be viewed as a subalgebra of 2×2-matrices, as follows.

Let

$$K = F[i] = F \oplus Fi \cong F[x]/(x^2 - a)$$

be the (commutative) F-algebra generated by i. Suppose that K is a field: then $K \cong F(\sqrt{a})$ is a quadratic field extension of F. (We relax this assumption below; alternatively, replace "vector space" with "free module" in the argument that follows.)

The algebra B has the structure of a left K-vector space of dimension 2, with basis $1, j$: explicitly, we have

$$\alpha = t + xi + yj + zij = (t + xi) + (y + zi)j \in K \oplus Kj$$

for all $\alpha \in B$. We then define the *right regular representation* over K

$$\begin{aligned} \rho : B &\to \operatorname{End}_K(B) \\ \alpha &\mapsto \rho_\alpha \end{aligned}$$

by mapping an element $\alpha \in B$ to the map ρ_α given by right multiplication by α. Each map ρ_α is indeed a K-linear endomorphism in B (considered as a left K-vector space) by associativity in B: we have

$$(w\beta)\rho_\alpha = (w\beta)\alpha = w(\beta\alpha) = w(\beta)\rho_\alpha$$

for all $\alpha, \beta \in B$ and $w \in K$. Note here that we adopt the convention that endomorphisms act on the *right*, so $\rho_\alpha \cdot \rho_\beta$ means first ρ_α then ρ_β. A similar argument shows that ρ is further an F-algebra homomorphism.

In the basis $1, j$ we have $\operatorname{End}_K(B) \cong \mathrm{M}_2(K)$, and ρ is given by

$$i \mapsto \rho_i = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad j \mapsto \rho_j = \begin{pmatrix} 0 & 1 \\ b & 0 \end{pmatrix}. \tag{2.2.9}$$

By our convention, the matrices above act on row vectors on the right.

The map ρ is injective (ρ is a *faithful* representation) since $\rho_\alpha = 0$ implies $\rho_\alpha(1) = \alpha = 0$.

In particular, we have that

$$B \cong \left\{ \begin{pmatrix} t + xi & y + zi \\ b(y - zi) & t - xi \end{pmatrix} : t, x, y, z \in F \right\} \subset \mathrm{M}_2(K). \tag{2.2.10}$$

Now—even if K is not a field—we verify by definition that the F-subalgebra generated by the matrices ρ_i, ρ_j in (2.2.9) is a quaternion algebra using Lemma 2.2.5, so that the isomorphism (2.2.10) holds in all cases.

Here, B acts on rows on the right; if instead, one wishes to have B act on the left on columns, give B the structure of a right K-vector space and use the left regular representation instead.

This is not the only way to embed B as a subalgebra of 2×2-matrices; indeed, the "splitting" of quaternion algebras in this way is a theme that will reappear throughout this text.

2.3 Rotations

From Paragraph 2.2.8, we see that the Hamiltonians $\mathbb{H}$ have the structure of a left $\mathbb{C}$-vector space with basis $1, j$; the right regular representation (2.2.10) then yields an

$\mathbb{R}$-algebra embedding

$$\begin{aligned} \rho : \mathbb{H} &\hookrightarrow \operatorname{End}_{\mathbb{C}}(\mathbb{H}) \cong \mathrm{M}_2(\mathbb{C}) \\ t + xi + yj + zij &\mapsto \begin{pmatrix} t+xi & y+zi \\ -(y-zi) & t-xi \end{pmatrix} = \begin{pmatrix} u & v \\ -\overline{v} & \overline{u} \end{pmatrix} \end{aligned} \tag{2.3.1}$$

where $u = t + xi$ and $v = y + zi$ and $\overline{}$ denotes complex conjugation. Note that

$$\det\begin{pmatrix} u & v \\ -\overline{v} & \overline{u} \end{pmatrix} = |u|^2 + |v|^2 = t^2 + x^2 + y^2 + z^2.$$

It follows that $\mathbb{H}^\times = \mathbb{H} \setminus \{0\}$ and the subgroup of *unit Hamiltonians*

$$\mathbb{H}_1^\times = \{t + xi + yj + zk \in \mathbb{H} : t^2 + x^2 + y^2 + z^2 = 1\},$$

which as a set is identified with the 3-sphere in $\mathbb{R}^4$, is isomorphic as a group to

$$\begin{aligned} \mathbb{H}_1^\times \cong \mathrm{SU}(2) &= \left\{ \begin{pmatrix} u & v \\ -\overline{v} & \overline{u} \end{pmatrix} \in \mathrm{M}_2(\mathbb{C}) : |u|^2 + |v|^2 = 1 \right\} \\ &= \{A \in \mathrm{SL}_2(\mathbb{C}) : A^* = A^{-1}\} \\ &= \{A \in \mathrm{SL}_2(\mathbb{C}) : JA = \overline{A}J\} \end{aligned}$$

where $A^* = \overline{A}^t$ is the complex conjugate transpose of A and $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

To conclude this chapter, we return to Hamilton's original design: quaternions model rotations in 3-dimensional space. This development is important not only historically important but it also previews many aspects of the general theory of quaternion algebras over fields.

Definition 2.3.2. $\alpha \in \mathbb{H}$ is *real* if $\alpha \in \mathbb{R}$ and *pure* (or *imaginary*) if $\alpha \in \mathbb{R}i + \mathbb{R}j + \mathbb{R}ij$.

Therefore, just like over the complex numbers, every element of $\mathbb{H}$ is the sum of its real part and its pure (imaginary) part. And just like complex conjugation, we define a *conjugation* map

$$\begin{aligned} \overline{} : \mathbb{H} &\to \mathbb{H} \\ \alpha = t + (xi + yj + zk) &\mapsto \overline{\alpha} = t - (xi + yj + zk) \end{aligned} \tag{2.3.3}$$

by negating the imaginary part. We compute that

$$\alpha + \overline{\alpha} = 2t \quad \text{and} \quad \alpha\overline{\alpha} = \|\alpha\|^2 = t^2 + x^2 + y^2 + z^2.$$

The conjugate transpose map on $\mathrm{M}_2(\mathbb{C})$ restricts to conjugation on the image of $\mathbb{H}$ in 2.3.1. Thus the elements of $\mathbb{H}$ which are Hermitian matrices are the scalar matrices and those that are skew-Hermitian are exactly the pure quaternions. The conjugation map is the subject of the next chapter (Chapter 3), and we discuss it more generally there.

Let

$$\mathbb{H}^0 = \{v = xi + yj + zk \in \mathbb{H} : x, y, z \in \mathbb{R}\} \in \mathbb{R}^3$$

be the set of pure Hamiltonians, the three-dimensional real space on which the (unit) Hamiltonians will act by rotations. For $v \in \mathbb{H}^0 \cong \mathbb{R}^3$, we have

$$\|v\|^2 = x^2 + y^2 + z^2 = \det(\rho(v)), \tag{2.3.4}$$

and from (2.3.1), we have

$$\mathbb{H}^0 = \{v \in \mathbb{H} : \mathrm{Tr}(\rho(v)) = v + \bar{v} = 0\}.$$

Consequently for $v \in \mathbb{H}^0$ we again see that $\bar{v} = -v$.

The set $\mathbb{H}^0$ is not closed under multiplication: if $v, w \in \mathbb{H}^0$ we have

$$vw = -v \cdot w + v \times w \tag{2.3.5}$$

where $v \cdot w$ is the dot product on $\mathbb{R}^3$ and $v \times w$ is the *cross product* on $\mathbb{R}^3$, defined as the determinant

$$v \times w = \begin{vmatrix} i & j & k \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$$

where $v = v_1 i + v_2 j + v_3 k$ and $w = w_1 i + w_2 j + w_3 k$, so

$$v \cdot w = v_1 w_1 + v_2 w_2 + v_3 w_3$$

and

$$v \times w = (v_2 w_3 - v_3 w_2)i + (v_3 w_1 - v_1 w_3)j + (v_1 w_2 - v_2 w_1)k.$$

The formula (2.3.5) is striking: it contains three different kinds of multiplications!

2.3.6. The following statements follow directly from (2.3.5).

(a) If $v, w \in \mathbb{H}^0$, then $vw \in \mathbb{H}^0$ if and only if v, w are orthogonal.

(b) $v^2 = -\|v\|^2 \in \mathbb{R}$ for all $v \in \mathbb{H}^0$.

(c) $wv = -vw$ if $v, w \in \mathbb{H}^0$ are orthogonal.

The group $\mathbb{H}_1^\times$ acts on our three-dimensional space $\mathbb{H}^0$ (on the left) by conjugation:

$$\begin{aligned} \mathbb{H}_1^\times \circlearrowright \mathbb{H}^0 &\to \mathbb{H}^0 \cong \mathbb{R}^3 \\ v &\mapsto \alpha v \alpha^{-1}; \end{aligned} \tag{2.3.7}$$

indeed, $\operatorname{Tr}(\rho(\alpha v \alpha^{-1})) = \operatorname{Tr}(\rho(v)) = 0$ by properties of the trace, so $\alpha v \alpha^{-1} \in \mathbb{H}^0$. Or, we have

$$\mathbb{H}^0 = \{v \in \mathbb{H} : v^2 \leq 0\}$$

and this latter set is visibly stable under conjugation. The representation (2.3.7) is called the *adjoint representation*.

Let $\alpha \in \mathbb{H}_1^\times \setminus \{\pm 1\}$. Then there exists $\theta \in \mathbb{R}$ such that

$$\alpha = t + xi + yj + zk = \cos\theta + (\sin\theta) I(\alpha) \tag{2.3.8}$$

and $\|I(\alpha)\| = 1$: we take $\theta = \cos^{-1} t$ and

$$I(\alpha) = \frac{xi + yj + zk}{|\sin\theta|},$$

since otherwise $\alpha = \pm 1$. We call $I(\alpha)$ as in (2.3.8) the *axis* of α.

Remark 2.3.9. In analogy with Euler's formula, we can write (2.3.8) as

$$\alpha = \exp(I(\alpha)\theta).$$

Proposition 2.3.10. *$\mathbb{H}_1^\times$ acts by rotation on $\mathbb{R}^3$ via conjugation: specifically, α acts by rotation through the angle 2θ about the axis $I(\alpha)$.*

Proof. Let $\alpha \in \mathbb{H}_1^\times \setminus \{\pm 1\}$. Then for all $v \in \mathbb{H}^0$, we have

$$\|\alpha v \alpha^{-1}\|^2 = \|v\|^2$$

by (2.3.4), so at least α acts by an orthogonal matrix

$$\mathrm{O}(3) = \{A \in \mathrm{M}_3(\mathbb{R}) : AA^t = 1\}.$$

Let $0 \neq j' \in \mathbb{H}^0$ be a unit vector orthogonal to $i' = I(\alpha)$. Then $(i')^2 = (j')^2 = -1$ by 2.3.6(b) and $j'i' = -i'j'$ by 2.3.6(c), so (applying an automorphism of $\mathbb{H}$) without loss of generality we may assume that $I(\alpha) = i$ and $j' = j$.

Thus $\alpha = t + xi$ with $t^2 + x^2 = \cos^2\theta + \sin^2\theta = 1$, and $\alpha^{-1} = t - xi$. We have

$$\alpha i \alpha^{-1} = i$$

(computing in $\mathbb{C}$), and

$$\begin{aligned}\alpha j\alpha^{-1} &= (t+xi)j(t-xi) = (t+xi)(t+xi)j \\ &= ((t^2-x^2)+2txi)j = (\cos 2\theta)j + (\sin 2\theta)k\end{aligned}$$

by the double angle formula. Consequently,

$$\alpha k\alpha^{-1} = i(\alpha j\alpha^{-1}) = (-\sin 2\theta)j + (\cos 2\theta)k$$

so the matrix of α in the basis $1, i, j$ is

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos 2\theta & \sin 2\theta \\ 0 & -\sin 2\theta & \cos 2\theta \end{pmatrix},$$

a (counterclockwise) rotation (determinant 1) through the angle 2θ about i as desired. □

Corollary 2.3.11. *We have an exact sequence*

$$1 \to \{\pm 1\} \to \mathbb{H}_1^\times \to \mathrm{SO}(3) \to 1$$

where

$$\mathrm{SO}(3) = \{A \in \mathrm{M}_3(\mathbb{R}) : AA^t = A^tA = 1 \textit{ and } \det(A) = 1\}$$

is the group of rotations of $\mathbb{R}^3$.

In particular, since $\mathbb{S}^3 \cong \mathrm{SU}(2) \cong \mathbb{H}_1^\times$ we have $\mathrm{SO}(3) \cong \mathrm{SU}(2)/\{\pm 1\} \cong \mathbb{RP}^3$ is topologically real projective space.

Proof. The map $\mathbb{H}_1^\times \to \mathrm{SO}(3)$ is surjective, since every element of SO(3) is rotation about some axis. If α belongs to the kernel, then $\alpha = \cos\theta + (\sin\theta)I(\alpha)$ must have $\sin\theta = 0$ so $\alpha = \pm 1$. □

2.3.12. We conclude with one final observation, returning to the formula (2.3.5). There is another way to mix the dot product and cross product in $\mathbb{H}$: we define the *scalar triple product*

$$\begin{aligned}\mathbb{H}\times\mathbb{H}\times\mathbb{H} &\to \mathbb{R} \\ (u,v,w) &\mapsto u\cdot(v\times w).\end{aligned} \tag{2.3.13}$$

Amusingly, this gives a way to "multiply" triples of triples: in fact, the map (2.3.13) is an alternating, trilinear form (Exercise 2.14). If $u, v, w \in \mathbb{H}^0$, then the scalar triple product is a determinant

$$u\cdot(v\times w) = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$$

so $|u \cdot (v \times w)|$ is the volume of a parallelepiped in $\mathbb{R}^3$ whose sides are given by u, v, w.

2.4 Extensions and further reading

2.4.1. The main reference for quaternion algebras, which can serve as companions to the material in this book is the book by Vignéras [Vig80]. Maclachlan–Reid [MR03] also gives an overview of the subject with an eye toward the manifestation of quaternion algebras in the theory of Fuchsian and Kleinian groups.

An overview of the subject of associative algebras is given by Pierce [Pie82].

2.4.2. The matrix representation of $\mathbb{H} \otimes_{\mathbb{R}} \mathbb{C}$ in section 2.3, and its connections to unitary matrices, is still used in phisics. In the embedding with

$$i \mapsto \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, j \mapsto \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, k \mapsto \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

whose images are unitary matrices, we multiply by i to obtain Hermitian matrices

$$\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

which are the famous *Pauli spin matrices*. [[cites]].

Exercises

Let F be a field with char $F \neq 2$.

2.1. Show that a (not necessarily associative) F-algebra is associative if and only if the associative law holds on a basis, and thereby check that the multiplication table implied by (2.2.2) is associative.

2.2. Show that if B is an F-algebra generated by $i, j \in B$ and $1, i, j$ are linearly dependent, then B is commutative.

2.3. Verify directly that the map $\left(\dfrac{1,1}{F}\right) \xrightarrow{\sim} \mathrm{M}_2(F)$ in Example 2.2.4 is an isomorphism of F-algebras.

2.4. Let $a, b \in F^\times$.

a) Show that $\left(\dfrac{a,b}{F}\right) \cong \left(\dfrac{a,-ab}{F}\right) \cong \left(\dfrac{b,-ab}{F}\right)$.

b) Show that if $c,d \in F^\times$ then $\left(\dfrac{a,b}{F}\right) \cong \left(\dfrac{ac^2,bd^2}{F}\right)$. Conclude that if $F^\times/F^{\times 2}$ is finite, then there are only finitely many isomorphism classes of quaternion algebras over F, and in particular that if $F^{\times 2} = F^\times$ then there is only one isomorphism class $\left(\dfrac{1,1}{F}\right) \cong \mathrm{M}_2(F)$.

c) Let B be a quaternion algebra over F. Show that $B \otimes_F \overline{F} \cong \mathrm{M}_2(\overline{F})$, where $\overline{F}$ is an algebraic closure of F.

d) Refine part (c) as follows. A field $K \supseteq F$ is a *splitting field* for B if $B \otimes_F K \cong \mathrm{M}_2(K)$. Show that B has a splitting field K with $[K : F] \leq 2$.

2.5. Recall that a *division ring* is a ring R in which every nonzero element has a (two-sided) inverse, i.e., $R \setminus \{0\}$ is a group under multiplication.

Show that if B is a division quaternion algebra over $\mathbb{R}$ then $B \cong \mathbb{H}$.

2.6. Use the quaternion algebra $B = \left(\dfrac{-1,-1}{F}\right)$, multiplicativity of the determinant, and the left regular representation (2.2.10) to show that if two elements of F can be written as the sum of four squares, then so too can their product (a discovery of Euler in 1748).

2.7. Let B be an F-algebra. The *center* of B is

$$Z(B) = \{\alpha \in B : \alpha\beta = \beta\alpha \text{ for all } \beta \in B\}.$$

We say B is *central* if $Z(B) = F$. Show that if B is a quaternion algebra over F, then B is central.

2.8. Prove the following partial generalization of Exercise 2.4(c). Let B be a finite-dimensional algebra over F.

a) Show that every element $\alpha \in B$ satisfies a unique monic polynomial of smallest degree with coefficients in F.

b) Suppose that $B = D$ is a division algebra (cf. Exercise exc:BdivHH). Show that the minimal polynomial of $\alpha \in D$ is irreducible over F. Conclude that if $F = \overline{F}$ is algebraically closed, then $D = F$.

2.9. Show explicitly that every quaternion algebra B over F is isomorphic to a subalgebra of $\mathrm{M}_4(F)$ via the right regular representation over F. With respect to a suitable such embedding for $B = \mathbb{H}$, show that the quaternionic conjugation map $\alpha \mapsto \overline{\alpha}$ is the matrix transpose, and the matrix determinant is the *square* of the norm $\|\alpha\|^2 = \alpha\overline{\alpha}$.

2.10. In Corollary 2.3.11, we showed that $\mathrm{SU}(2) \cong \mathbb{H}_1^\times$ has a 2-to-1 map to $\mathrm{SO}(3)$, where $\mathbb{H}_1^\times$ acts on $\mathbb{H}^0 \cong \mathbb{R}^3$ by conjugation: quaternions model rotations in three-dimensional space, with a little bit of spin. They also do so in four-dimensional space, as follows.

a) Show that the map

$$\begin{aligned} (\mathbb{H}_1^\times \times \mathbb{H}_1^\times) \circlearrowright \mathbb{H} &\to \mathbb{H} \cong \mathbb{R}^4 \\ x &\mapsto \alpha x \beta^{-1} \end{aligned} \tag{2.4.3}$$

defines a (left) action of $\mathbb{H}_1^\times \times \mathbb{H}_1^\times$ on $\mathbb{H} \cong \mathbb{R}^4$, giving a group homomorphism

$$\phi : \mathbb{H}_1^\times \times \mathbb{H}_1^\times \to \mathrm{O}(4).$$

b) Show that ϕ surjects onto $\mathrm{SO}(4) < \mathrm{O}(4)$. *[Hint: If $A \in \mathrm{SO}(4)$ fixes $1 \in \mathbb{H}$, then A restricted to $\mathbb{H}^0$ is a rotation and so is given by conjugation. More generally, if $A \cdot 1 = \alpha$, consider $x \mapsto \alpha^{-1}Ax$.]*

c) Show that the kernel of ϕ is ± 1, so we have an exact sequence

$$1 \to \{\pm 1\} \to \mathrm{SU}(2) \times \mathrm{SU}(2) \to \mathrm{SO}(4) \to 1.$$

2.11. a) Show that the rotation $\rho(u,\theta) : \mathbb{R}^3 \to \mathbb{R}^3$ counterclockwise by the angle 2θ about the axis $u \in \mathbb{R}^3 \cong \mathbb{H}^0$ is given by conjugation by the quaternion $\alpha = \cos(\theta/2) + (\sin(\theta/2))u$.

b) Prove *Rodrigues's rotation formula*:

$$\rho(u,\theta; v) = (\cos\theta)v + (\sin\theta)(u \times v) + (1 - \cos\theta)(u \cdot v)u.$$

2.12. [[Do Hamilton's rotation bit with $B = \mathrm{M}_2(\mathbb{R})$ instead of $B = \mathbb{H}$.]]

2.13. Let B be a quaternion algebra and let $\mathrm{M}_2(B)$ be the ring of 2×2-matrices over B. (Be careful in the definition of matrix multiplication: B is noncommutative!)

a) By an explicit formula, show that $\mathrm{M}_2(B)$ has a *determinant* map $\det : \mathrm{M}_2(B) \to F$ that is multiplicative and left-B-multilinear in the rows of B.

b) Find a matrix $A \in \mathrm{M}_2(\mathbb{H})$ that is invertible (i.e., having a two-sided inverse) but has $\det(A) = 0$. Then find such an A with the further property that its transpose is not invertible but has nonzero determinant.

Moral: be careful with matrix rings over noncommutative rings! [[Check, compare with Dieudonné determinant, and give formula.]]

2.14. Verify that the map (2.3.13) is a trilinear, alternating form on $\mathbb{H}$.

Chapter 3

Involutions

In this chapter, we define the standard involution (also called conjugation) on a quaternion algebra. In this way, we characterize division quaternion algebras as noncommutative division rings equipped with a standard involution.

3.1 Conjugation

The conjugation map (2.3.3) defined on the Hamiltonians $\mathbb{H}$ arises naturally from the notion of real and imaginary parts, which Hamilton argued have a physical interpretation. This involution played an essential role, and it has a natural generalization to a quaternion algebra $B = \left(\dfrac{a,b}{F}\right)$ over a field F with $\operatorname{char} F \neq 2$: we define

$$\begin{aligned} \overline{} : B &\to B \\ \alpha = t + xi + yj + zij &\mapsto \overline{\alpha} = t - (xi + yj + zij) \end{aligned}$$

Multiplying out, we can then verify directly in analogy that

$$\|\alpha\|^2 = \alpha\overline{\alpha} = t^2 - ax^2 - by^2 + abz^2 \in F.$$

The way in which the cross terms cancel, because the basis elements i, j, ij skew commute, is an enchanting calculation to perform every time!

But this definition seems to depend on a basis: it is not intrinsically defined. What properties characterize it? Is it unique? We are looking for a good definition of conjugation $\overline{} : B \to B$ on an F-algebra B: we will call such a map a *standard involution.*

The involutions we consider should have the basic linearity properties: they are F-linear (with $\overline{1} = 1$) and have order 2 as an F-linear map. An involution should also respect the multiplication structure on B, but we should not require that it be an F-algebra isomorphism: instead, like the inverse map reverses order of multiplication, we ask that $\overline{\alpha\beta} = \overline{\beta}\,\overline{\alpha}$ for all $\alpha \in B$. Finally, we want the standard involution to give rise to a trace and norm (a measure of size), which is to say, we want $\alpha + \overline{\alpha} \in F$ and $\alpha\overline{\alpha} = \overline{\alpha}\alpha \in F$ for all $\alpha \in B$. The precise (minimal) definition is given in Definition 3.2.1. These properties are rigid: if an algebra B has a standard involution, then it is necessarily unique (Corollary 3.4.3).

The existence of a standard involution on B implies that every element of B satisfies a quadratic equation: by direct substitution, we see that $\alpha \in B$ is a root of the polynomial $x^2 - tx + n \in F[x]$ where $t = \alpha + \overline{\alpha}$ and $n = \alpha\overline{\alpha}$, since then

$$\alpha^2 - (\alpha + \overline{\alpha})\alpha + \alpha\overline{\alpha} = 0$$

identically. This is already a strong condition on B: we say that B has *degree* 2 if every element $\alpha \in B$ satisfies a (monic) polynomial in $F[x]$ of degree 2 and, to avoid trivialities, that $B \neq F$.

The main result of this section is that a division F-algebra of degree 2 over a field F with $\operatorname{char} F \neq 2$ is either a quadratic field extension of F or a division quaternion algebra over F. As a consequence, a noncommutative division algebra with a standard involution is a quaternion algebra (and conversely). This gives our first intrinsic characterization of (division) quaternion algebras, when $\operatorname{char} F \neq 2$.

3.2 Involutions

Throughout this chapter, let B be an F-algebra. For the moment, we allow F to be of arbitrary characteristic.

We begin by defining involutions on B.

Definition 3.2.1. An *involution* $\overline{} : B \to B$ is an F-linear map which satisfies:

(i) $\overline{1} = 1$;

(ii) $\overline{\overline{\alpha}} = \alpha$ for all $\alpha \in B$; and

(iii) $\overline{\alpha\beta} = \overline{\beta}\,\overline{\alpha}$ for all $\alpha, \beta \in B$ (the map $\overline{}$ is an *anti-automorphism*).

3.2.2. If B^{op} denotes the *opposite algebra* of B, so that $B^{\mathrm{op}} = B$ as abelian groups but with multiplication $\alpha \cdot_{\mathrm{op}} \beta = \beta \cdot \alpha$, then one can equivalently define an involution to be an F-algebra isomorphism $B \to B^{\mathrm{op}}$ whose underlying F-linear map has order at most 2.

Remark 3.2.3. What we have defined to be an involution is known in other contexts as an *involution of the first kind.* An *involution of the second kind* is a map which acts nontrivially when restricted to F, and hence is not F-linear; although these involutions are interesting in other contexts, they will not figure in our discussion, as one can always consider such an algebra over the fixed field of the involution.

Definition 3.2.4. An involution $\overline{}$ is *standard* if $\alpha\overline{\alpha} \in F$ for all $\alpha \in B$.

3.2.5. If $\overline{}$ is a standard involution, so that $\alpha\overline{\alpha} \in F$ for all $\alpha \in B$, then

$$(\alpha + 1)(\overline{\alpha + 1}) = (\alpha + 1)(\overline{\alpha} + 1) = \alpha\overline{\alpha} + \alpha + \overline{\alpha} + 1 \in F$$

and hence $\alpha + \overline{\alpha} \in F$ for all $\alpha \in B$ as well. It then also follows that $\alpha\overline{\alpha} = \overline{\alpha}\alpha$, since

$$(\alpha + \overline{\alpha})\alpha = \alpha(\alpha + \overline{\alpha}).$$

Example 3.2.6. The $\mathbb{R}$-algebra $\mathbb{C}$ has a standard involution, namely, complex conjugation.

Example 3.2.7. The adjoint map

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto A^{\dagger} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

is a standard involution on $\mathrm{M}_2(F)$ since $AA^{\dagger} = ad - bc = \det A \in F$.

Matrix transpose is an involution on $\mathrm{M}_n(F)$ but is a standard involution only if $n = 1$.

3.2.8. Suppose $\operatorname{char} F \neq 2$ and let $B = \left(\dfrac{a, b}{F}\right)$ be a quaternion algebra. Then the map

$$\begin{aligned} \overline{} : B &\to B \\ \alpha = t + xi + yj + zij &\mapsto \overline{\alpha} = 2t - \alpha = t - xi - yj - zij \end{aligned}$$

defines a standard involution on B. The map is F-linear with $\overline{1} = 1$ and $\overline{\overline{\alpha}} = \alpha$, so properties (i) and (ii) hold. By F-linearity, it is enough to check property (iii) on a basis (Exercise 3.1), and we verify e.g. that

$$\overline{ij} = -ij = ji = (-j)(-i) = \overline{j}\,\overline{i}$$

(Exercise 3.2). Finally, the involution is standard because

$$(t + xi + yj + zij)(t - xi - yj - zij) = t^2 - ax^2 - by^2 + abz^2 \in F. \tag{3.2.9}$$

3.3 Reduced trace and reduced norm

Let $\overline{} : B \to B$ be a standard involution on B. We define the *reduced trace* on B by

$$\begin{aligned} \operatorname{trd} : B &\to F \\ \alpha &\mapsto \operatorname{trd}(\alpha) = \alpha + \overline{\alpha} \end{aligned}$$

and similarly the *reduced norm*

$$\begin{aligned} \operatorname{nrd} : B &\to F \\ \alpha &\mapsto \operatorname{nrd}(\alpha) = \alpha\overline{\alpha}. \end{aligned}$$

Example 3.3.1. For $B = \mathrm{M}_2(F)$, equipped with the adjoint map as a standard involution as in Example 3.2.7, the reduced trace is the usual matrix trace and the reduced norm is the determinant.

3.3.2. The reduced trace trd is an F-linear map, since this is true for the standard involution:

$$\operatorname{trd}(\alpha+\beta) = (\alpha+\beta) + \overline{(\alpha+\beta)} = (\alpha+\overline{\alpha}) + (\beta+\overline{\beta}) = \operatorname{trd}(\alpha) + \operatorname{trd}(\beta)$$

for $\alpha, \beta \in B$. The reduced norm nrd is multiplicative, since

$$\operatorname{nrd}(\alpha\beta) = (\alpha\beta)\overline{(\alpha\beta)} = \alpha\beta\overline{\beta}\,\overline{\alpha} = \alpha \operatorname{nrd}(\beta)\overline{\alpha} = \operatorname{nrd}(\alpha)\operatorname{nrd}(\beta)$$

for all $\alpha, \beta \in B$.

Lemma 3.3.3. *If B is not the zero ring, then $\alpha \in B$ is a unit (has a two-sided inverse) if and only if* $\operatorname{nrd}(\alpha) \neq 0$.

Proof. Exercise 3.4. □

Lemma 3.3.4. *For all $\alpha, \beta \in B$, we have* $\operatorname{trd}(\beta\alpha) = \operatorname{trd}(\alpha\beta)$.

Proof. We have

$$\operatorname{trd}(\alpha\overline{\beta}) = \operatorname{trd}(\alpha(\operatorname{trd}(\beta) - \beta)) = \operatorname{trd}(\alpha)\operatorname{trd}(\beta) - \operatorname{trd}(\alpha\beta)$$

and yet

$$\operatorname{trd}(\alpha\overline{\beta}) = \operatorname{trd}(\overline{\alpha\overline{\beta}}) = \operatorname{trd}(\beta\overline{\alpha}) = \operatorname{trd}(\alpha)\operatorname{trd}(\beta) - \operatorname{trd}(\beta\alpha)$$

so $\operatorname{trd}(\alpha\beta) = \operatorname{trd}(\beta\alpha)$. □

Remark 3.3.5. The maps trd and nrd are called *reduced* for the following reason.

Let A be a finite-dimensional F-algebra, and consider the right regular representation $\rho : A \hookrightarrow \operatorname{End}_F(A)$ given by right-multiplication in A (cf. Paragraph 2.2.8). We then have a trace map $\operatorname{Tr} : A \to F$ and norm map $\operatorname{N} : A \to F$ given by mapping $\alpha \in B$ to the trace and determinant of the endomorphism $\rho(\alpha)$.

Now when $A = \mathrm{M}_2(F)$, a direct calculation (Exercise 3.10) reveals that

$$\operatorname{Tr}(\alpha) = 2\operatorname{trd}(\alpha) \text{ and } \operatorname{N}(\alpha) = \operatorname{nrd}(\alpha)^2 \text{ for all } \alpha \in A$$

whence the name *reduced.*

3.3.6. Since

$$\alpha^2 - (\alpha + \overline{\alpha})\alpha + \alpha\overline{\alpha} = 0 \tag{3.3.7}$$

identically we see that $\alpha \in B$ is a root of the polynomial

$$x^2 - \operatorname{trd}(\alpha)x + \operatorname{nrd}(\alpha) \in F[x] \tag{3.3.8}$$

which we call the *reduced characteristic polynomial* of α. (We might call this statement the *reduced Cayley-Hamilton theorem* for an algebra with standard involution.)

3.4 Uniqueness and degree

An F-algebra K with $\dim_F K = 2$ is called a *quadratic* algebra.

Lemma 3.4.1. *Let K be a quadratic F-algebra. Then K is commutative and has a unique standard involution.*

Proof. Let $\alpha \in K \setminus F$. Then $K = F \oplus F\alpha = F[\alpha]$, so in particular K is commutative. Then $\alpha^2 = t\alpha - n$ for unique $t, n \in F$, since $1, \alpha$ is a basis for K, and consequently $K \cong F[x]/(x^2 - tx + n)$.

Now if $\overline{} : K \to K$ is any standard involution, then from (3.3.7) and uniqueness we have $t = \alpha + \overline{\alpha}$ (and $n = \alpha\overline{\alpha}$), and so any involution must have $\overline{\alpha} = t - \alpha$. And indeed, the map $\alpha \mapsto t - \alpha$ extends to a unique involution of B because $t - \alpha$ is also a root of $x^2 - tx + n$ (and so (i)–(iii) hold in Definition 3.2.1), and it is standard because $\alpha(t - \alpha) = n \in F$. □

Remark 3.4.2. If K is a separable field extension of F, then the standard involution is just the nontrivial element of $\operatorname{Gal}(K/F)$.

Corollary 3.4.3. *If B has a standard involution, then this involution is unique.*

Proof. For any $\alpha \in B \setminus F$, we have from (3.3.7) that $\dim_F F[\alpha] = 2$, so the restriction of the standard involution to $F[\alpha]$ is unique. Therefore the standard involution on B is itself unique. □

We have seen that the equation (3.3.7), implying that if B has a standard involution then every $\alpha \in B$ satisfies a quadratic equation, has figured prominently in the above proofs. To further clarify the relationship between these two notions, we make the following definition.

Definition 3.4.4. The *degree* of B is the smallest integer $m \in \mathbb{Z}_{\geq 1}$ such that every element $\alpha \in B$ satisfies a monic polynomial $f(x) \in F[x]$ of degree m, if such an integer exists.

If B has finite dimension $n = \dim_F B < \infty$, then every element of F satisfies a polynomial of degree at most n: if $\alpha \in B$ then the elements $1, \alpha, \dots, \alpha^n$ are linearly dependent over F. Consequently, every finite-dimensional F-algebra has a well-defined degree.

Example 3.4.5. If B has degree 1, then $B = F$. If B has a standard involution, then either $B = F$ or B has degree 2 by (3.3.8).

3.5 Quaternion algebras

We are now ready to characterize algebras of degree 2.

Theorem 3.5.1. *Suppose* $\operatorname{char} F \neq 2$ *and let B be a division F-algebra. Then B has degree at most* 2 *if and only if one of the following hold:*

(i) *$B = F$;*

(ii) *$B = K$ is a quadratic field extension of F; or*

(iii) *B is a division quaternion algebra over F.*

Proof. From Example 3.4.5, we may assume $B \neq F$ and B has degree 2.

Let $i \in B \setminus F$. Then $F[i] = K$ is a (commutative) quadratic F-subalgebra of the division ring B, so $K = F(i)$ is a field. If $K = B$, we are done. Completing the square (since $\operatorname{char} F \neq 2$), we may suppose that $i^2 = a \in F^\times$.

Let $\phi : B \to B$ be the map given by conjugation by i, i.e., $\phi(x) = i^{-1}xi = a^{-1}ixi$. Then ϕ is a K-linear endomorphism of B, thought of as a (left) K-vector space, and

ϕ^2 is the identity on B. Therefore ϕ is diagonalizable, so we may decompose $B = B^+ \oplus B^-$ into eigenspaces for ϕ: explicitly, we can always write

$$\alpha = \frac{\alpha + \phi(\alpha)}{2} + \frac{\alpha - \phi(\alpha)}{2} \in B^+ \oplus B^-.$$

We now prove $\dim_K B^+ = 1$. Let $\alpha \in B^+$. Then $L = F(\alpha, i)$ is a field. Since $\operatorname{char} F \neq 2$, and L is a compositum of quadratic extensions of F, the primitive element theorem implies that $L = F(\beta)$ for some $\beta \in L$. But by hypothesis β satisfies a quadratic equation so $\dim_F L = 2$ and hence $L = K$. (For an alternative direct proof of this claim, see Exercise 3.8.)

Next, we prove that $\dim_K B^- = 1$. If $0 \neq j \in B^-$ then $i^{-1}ji = -j$, so $i = -j^{-1}ij$ and hence all elements of B^- conjugate i to $-i$. Thus if $0 \neq j_1, j_2 \in B^-$ then $j_1 j_2$ centralizes i so $j_1 j_2 \in B^+ = K$. Thus any two nonzero elements of B^- are indeed K-multiples of each other.

Finally, let $j \in B^- \setminus \{0\}$, so $ji = -ij$. Both j and $i^{-1}ji = -j$ satisfy the same minimal polynomial of degree 2 and belong to $F(j)$, so we must have $j^2 = b \in F^\times$ and B is a quaternion algebra. □

Remark 3.5.2. We need not assume in Theorem 3.5.1 that B is finite-dimensional; somehow, it is a consequence, and every division algebra over F (with $\operatorname{char} F \neq 2$) of degree ≤ 2 is finite-dimensional. By contrast, any boolean ring (see Exercise 3.9), has degree 2 as an $\mathbb{F}_2$-algebra, and there are such rings of infinite dimension over $\mathbb{F}_2$—such algebras are quite far from being division rings, of course.

Corollary 3.5.3. *Let B be a division F-algebra with* $\operatorname{char} F \neq 2$. *Then B has degree at most* 2 *if and only if B has a standard involution.*

Proof. In each of the cases (i)–(iii), B has a standard involution. □

Recall that an F-algebra B is *central* if the center $Z(B) = \{\alpha \in B : \alpha\beta = \beta\alpha \text{ for all } \beta \in B\}$ is $Z(B) = F$.

Corollary 3.5.4. *Let B be a division F-algebra with* $\operatorname{char} F \neq 2$. *Then B is a quaternion algebra if and only if B is noncommutative and has degree* 2 *if and only if B is central and has degree* 2.

An F-algebra B is *algebraic* if every $\alpha \in B$ is algebraic over F (satisfies a polynomial with coefficients in F). If $\dim_F B = n < \infty$, then B is algebraic since for every $\alpha \in B$ the elements $1, \alpha, \dots, \alpha^n$ are linearly dependent over F.

Corollary 3.5.5 (Frobenius)**.** *Let B be an algebraic division algebra over $\mathbb{R}$. Then either $B = \mathbb{R}$, $B \cong \mathbb{C}$, or $B \cong \mathbb{H}$.*

Proof. If $\alpha \in B \setminus \mathbb{R}$ then $\mathbb{R}(\alpha) \cong \mathbb{C}$, so α satisfies a polynomial of degree 2. Thus if $B \neq \mathbb{R}$ then B has degree 2 and so either $B \cong \mathbb{C}$ or B is a division quaternion algebra over $\mathbb{R}$, and hence $B \cong \mathbb{H}$ by Exercise 2.5. □

Example 3.5.6. Division algebras over $\mathbb{R}$ of infinite dimension abound. Transcendental field extensions of $\mathbb{R}$, for example $\mathbb{R}(x)$ or $\mathbb{R}((x))$, are examples of infinite-dimensional division algebras over $\mathbb{R}$. The free algebra in two (noncommuting) variables is a subring of a division ring B with center $\mathbb{R}$.

3.6 Extensions and further reading

3.6.1. The consideration of algebras by degree was initiated by Dickson [Dic12] in the early 1900s. Dickson considered algebras in which every element satisfies a quadratic equation over a field F with $\operatorname{char} F \neq 2$, exhibited a diagonalized basis for such an algebra, and considered when such an algebra can be a division algebra.

3.6.2. Algebras with involutions come from quadratic forms. More precisely, there is a bijection between the set of isomorphism classes of finite-dimensional simple F-algebras equipped with a F-linear involution and the set of similarity classes of nonsingular quadratic forms on finite-dimensional F-vector spaces. More generally, for involutions that act nontrivially on the base field, one looks at Hermitian forms. Consequently, we have three broad types of involutions on central simple algebras, depending on the associated quadratic or Hermitian form: orthogonal, symplectic and unitary. Consequently, algebras with involutions can be classified by the invariants of the associated form.

Associated with every quadratic form there is a central simple algebra with involution. In this way the theory of quadratic forms belongs to the theory of algebras with involution, which in turn is a part of the theory of linear algebraic groups.

3.6.3. Lam

3.6.4. The standard involution is also called *conjugation* on B, but in some circumstances this can be confused with conjugation by an element in $B^\times$. The terminology *standard* is employed because conjugation on a quaternion algebra is the "standard" example of such an involution.

Because of Corollary 3.4.3, a standard involution is often also called the *canonical involution*; however, there are other circumstances where involutions can be defined canonically that are not standard (like the map induced by $g \mapsto g^{-1}$ on the group ring $F[G]$), so we resist this terminology.

3.6.5. The proof of Theorem 3.5.1 has a bit of history, discussed by van Praag [vPr02] (along with several proofs). Moore [1, Theorem 14.4] in 1915 studied algebra of matrices over skew fields and in particular the role of involutions, and gives an elementary proof of this theorem (with the assumption char $F \neq 2$). Dieudonné [Die48, Die53] gave another proof that relies on structure theory for finite-dimensional division algebras.

Albert proved that a central simple algebra A over F admits an (F-linear) involution if and only if A is isomorphic to its opposite algebra A^{op}. This is further equivalent to A having order at most 2 in the Brauer group $\text{Br}(F)$. Albert proved that a central simple algebra of dimension 16 with an involution is a biquaternion algebra.

3.6.6. Algebras with involution play an important role in analysis, for example,

3.6.7. The statement of Corollary 3.5.3 holds more generally (even if B is not necessarily a division ring). Let B be an F-algebra with char $F \neq 2$. Then B has a standard involution if and only if B has degree at most 2 [Voi11b]. However, this is no longer true in characteristic 2 (Exercise 3.9).

3.7 Algorithmic aspects

In this section, we exhibit an algorithm to determine if an algebra has a standard involution (and, if so, to give it explicitly as a linear map).

First, some definitions.

Definition 3.7.1. A field F is *computable* if F comes equipped with a way of encoding elements of F in bits (i.e. the elements of F are recursively enumerable, allowing repetitions) along with deterministic algorithms to perform field operations in R (addition, subtraction, multiplication, and division by a nonzero element) and to test if $x = 0 \in F$; a field is *polynomial-time computable* if these algorithms run in polynomial time (in the bit size of the input).

For precise definitions and a thorough survey of the subject of computable rings we refer to Stoltenberg-Hansen and Tucker [SHT99] and the references contained therein.

Example 3.7.2. A field that is finitely generated over its prime ring is computable by the theory of Gröbner bases [vzGG03].

Let B be a F-algebra with $\dim_F B = n$ and basis $e_1, e_2, \dots, e_n$ as an F-vector space and suppose $e_1 = 1$. A *multiplication table* for B is a system of n^3 elements

$(c_{ijk})_{i,j,k=1,\dots,n}$ of F, called *structure constants*, such that multiplication in B is given by

$$e_ie_j = \sum_{k=1}^{n} c_{ijk}e_k$$

for $i, j \in \{1, \dots, n\}$.

An F-algebra B is represented in bits by a multiplication table and elements of B are represented in the basis e_i. Note that basis elements in B can be multiplied directly by the multiplication table but multiplication of arbitrary elements in B requires $O(n^3)$ arithmetic operations (additions and multiplications) in F; in either case, note the output is of polynomial size in the input for fixed B.

We now exhibit an algorithm to test if an F-algebra B (of dimension n) has a standard involution [Voi13, §2].

First, we note that if B has a standard involution $\bar{\ } : B \to B$, then this involution and hence also the reduced trace and norm can be computed efficiently. Indeed, let $\{e_i\}_i$ be a basis for B; then $\operatorname{trd}(e_i) \in F$ is simply the coefficient of e_i in e_i^2, and so $\overline{e_i} = \operatorname{trd}(e_i) - e_i$ for each i can be precomputed for B; one recovers the involution on B for an arbitrary element of B by F-linearity. Therefore the involution and the reduced trace can be computed using $O(n)$ arithmetic operations in F and the reduced norm using $O(n^2)$ operations in F.

Algorithm 3.7.3. Let B be an F-algebra given by a multiplication table in the basis $e_1, \dots, e_n$ with $e_1 = 1$. This algorithm returns true if and only if B has a standard involution.

1. For $i = 2, \dots, n$, let $t_i \in F$ be the coefficient of e_i in e_i^2, and let $n_i = e_i^2 - t_ie_i$. If some $n_i \notin F$, return false.

2. For $i = 2, \dots, n$ and $j = i + 1, \dots, n$, let $n_{ij} = (e_i + e_j)^2 - (t_i + t_j)(e_i + e_j)$. If some $n_{ij} \notin F$, return false. Otherwise, return true.

Proof of correctness. Let $F[x] = F[x_1, \dots, x_n]$ be the polynomial ring over F in n variables, and let $B_{F[x]} = B \otimes_F F[x]$. Let $\xi = x_1 + x_2e_2 + \dots + x_ne_n \in B_{F[x]}$, and define

$$t_\xi = \sum_{i=1}^{n} t_ix_i$$

and

$$n_\xi = \sum_{i=1}^{n} n_ix_i^2 + \sum_{1 \le i < j \le n} (n_{ij} - n_i - n_j)x_ix_j.$$

Let

$$\xi^2 - t_\xi \xi + n_\xi = \sum_{i=1}^n c_i(x_1, \dots, x_n) e_i$$

with $c_i(x) \in F[x]$. Each $c_i(x)$ is a homogeneous polynomial of degree 2. The algorithm then verifies that $c_i(x) = 0$ for $x \in \{e_i\}_i \cup \{e_i + e_j\}_{i,j}$, and this implies that each $c_i(x)$ vanishes identically. Therefore, the specialization of the map $\xi \mapsto \overline{\xi} = t_\xi - \xi$ is the unique standard involution on B. □

3.7.4. Algorithm 3.7.3 requires $O(n)$ arithmetic operations in F, since e_i^2 can be computed directly from the multiplication table and hence $(e_i + e_j)^2 = e_i^2 + e_i e_j + e_j e_i + e_j^2$ can be computed using $O(4n) = O(n)$ operations.

Exercises

3.1. Let B be a F-algebra and let $\overline{} : B \to B$ be an F-linear map with $\overline{1} = 1$. Show that $\overline{}$ is an involution if and only if (ii)–(iii) in Definition 3.2.1 hold for a basis of B (as an F-vector space).

3.2. Verify that the map $\overline{}$ in Example 3.2.8 is a standard involution.

3.3. Determine the standard involution on $K = F \times F$.

3.4. Let B be an F-algebra with a standard involution. Show that $0 \neq x \in B$ is a left zerodivisor if and only if x is a right zerodivisor if and only if $\operatorname{nrd}(x) = 0$. In particular, if B is not the zero ring, then $\alpha \in B$ is (left and right) invertible if and only if $\operatorname{nrd}(\alpha) \neq 0$.

3.5. Let G be a finite group. Show that the F-linear map induced by $g \mapsto g^{-1}$ for $g \in G$ is an involution on the group ring $F[G] = \bigoplus_{g \in G} Fg$. Under what conditions is this map a standard involution?

3.6. Show that $B = M_n(F)$ has a standard involution if and only if $n \leq 2$.

3.7. In this exercise, we examine when the identity map yields a standard involution on an F-algebra B.

a) Show that if $\operatorname{char} F \neq 2$, then $x \in B$ satisfies $\overline{x} = x$ if and only $x \in F$.

b) Suppose that $\dim_F B < \infty$. Show that the identity map is a standard involution on B if and only if (i) $B = F$ or (ii) $\operatorname{char} F = 2$ and B is a quotient of the commutative ring $F[x_1, \dots, x_n]/(x_1^2 - a_1, \dots, x_n^2 - a_n)$ with $a_i \in F$.

3.8. Suppose $\operatorname{char} F \neq 2$. Let K be a field of degree 2 over F, i.e., suppose that every element of $K \setminus F$ satisfies a quadratic polynomial. Show directly that K is a quadratic field extension of F.

3.9. In this exercise, we explore further the relationship between algebras of degree 2 and those with standard involutions (Paragraph 3.6.7).

a) Suppose $\operatorname{char} F \neq 2$ and let B be a finite-dimensional F-algebra. Show that B has a standard involution if and only if $\deg_F B \leq 2$.

b) Let $F = \mathbb{F}_2$ and let B be a Boolean ring, a ring such that $x^2 = x$ for all $x \in B$. (Verify that $2 = 0$ in B, so B is indeed an $\mathbb{F}_2$-algebra.) Prove that B does not have a standard involution unless $B = \mathbb{F}_2$ or $B = \mathbb{F}_2 \times \mathbb{F}_2$, but nevertheless any Boolean ring has degree at most 2.

3.10. Let $B = \mathrm{M}_n(F)$, and consider the map $\rho : B \hookrightarrow \operatorname{End}_F(B)$ given by right-multiplication in B. Show that for all $A \in \mathrm{M}_n(F)$, the characteristic polynomial of $\rho(A)$ is the nth power of the characteristic polynomial of A. Conclude when $n = 2$ that $\operatorname{Tr}(A) = 2\operatorname{trd}(A)$ and $\mathrm{N}(A) = \operatorname{nrd}(M)^2$.

3.11. Let V be an F-vector space and let $t : V \to F$ be an F-linear map. Let $B = F \oplus V$ and define the binary operation $x \cdot y = t(x)y$ for $x, y \in V$. Show that $\cdot$ induces a multiplication on B, and that the map $x \mapsto \overline{x} = t(x) - x$ for $x \in V$ induces a standard involution on B. Such an algebra is called an *exceptional algebra* [GL09, Voi11b]. Conclude that there exists a central F-algebra B with a standard involution in any dimension $n = \dim_F B$.

3.12. In this exercise, we mimic the proof of Theorem 3.5.1 to prove that a quaternion algebra over a finite field is not a division ring, a special case of Wedderburn's theorem.

Let B be a division quaternion algebra over $F = \mathbb{F}_q$ with q odd. Show that for any $i \in B \setminus F$ that the centralizer $C(i) = \{x \in B^\times : ix = xi\}$ is given by $C(i) = F(i)^\times$. Conclude that any noncentral conjugacy class in $B^\times$ has order q^2+1. Derive a contradiction from the class equation $q^4 - 1 = q - 1 + k(q^2+1)$ (where $k \in \mathbb{Z}$).

This argument can be generalized in a natural way to prove Wedderburn's theorem in full: see Schue [Schu88], for example.

3.13. Derive Euler's identity (1.1.2) that the product of the sum of four squares is again the sum of four squares as follows. Let $F = \mathbb{Q}(x_1, \dots, x_4, y_1, \dots, y_4)$ be a function field over $\mathbb{Q}$ in 8 variables and consider the quaternion algebra

$\left(\dfrac{-1,-1}{F}\right)$. Show (by "universal formula") that if R is any commutative ring and $x, y \in R$ are the sum of four squares in R, then xy is the sum of four squares in R.

3.14. Suppose char $F \neq 2$. For an F-algebra B, let

$$V(B) = \{\alpha \in B \setminus F : \alpha^2 \in F\} \cup \{0\}.$$

Let B be a division ring. Show that $V(B)$ is a vector space (closed under addition) if and only if $B = F$ or $B = K$ is a quadratic field extension of F or B is a quaternion algebra over F. (Conclude that $V(B)$ is a vector space if and only if B has a standard involution.)

Chapter 4

Quadratic forms

Quaternion algebras, as algebras equipped with a standard involution, are intrinsically related to quadratic forms. We explore this connection in this section.

4.1 Norm form

Let F be a field with $\operatorname{char} F \neq 2$ and let $B = \left(\dfrac{a,b}{F}\right)$ be a quaternion algebra over F. We have seen (3.2.8) that B has a unique standard involution and consequently a reduced norm map, with

$$\operatorname{nrd}(t + xi + yj + zk) = t^2 - ax^2 - by^2 + abz^2 \tag{4.1.1}$$

for $t, x, y, z \in F$. The reduced norm therefore defines a *quadratic form*, a homogeneous polynomial of degree 2 in $F[t, x, y, z]$, with respect to the basis $1, i, j, k$. It should come as no surprise, then, that the structure of the quaternion algebra B is related to properties of the quadratic form nrd.

A quadratic form (since $\operatorname{char} F \neq 2$) can be diagonalized by a change of basis, and such a form is *nonsingular* (or *nondegenerate*) if the product of these diagonal entries is nonzero. The reduced norm quadratic form (4.1.1) is already diagonal in the basis $1, i, j, k$, and it is nonsingular because $a, b \neq 0$.

More generally, we have seen that any algebra with a standard involution has a quadratic form nrd. We will show (Theorem 4.3.1) that this form is nonsingular if and only if B is one of the following: F, a reduced quadratic algebra, or a quaternion algebra. This gives another way of characterizing quaternion algebras more general than Theorem 3.5.1: they are noncommutative algebras with a nonsingular standard involution.

Implicitly, we are trying to compare the categories of quaternion algebras and quadratic forms. From that point of view, the quadratic form (4.1.1) is a bit too big:

after all, we know what it does when restricted to F. The form carries the same data as when it is restricted to the orthogonal complement of F: this space is

$$B^0 = \{\alpha \in B : \operatorname{trd}(\alpha) = 0\}$$

and is spanned by i, j, ij. The quadratic form then becomes

$$\operatorname{nrd}(xi + yj + zk) = -ax^2 - by^2 + abz^2$$

for $x, y, z \in F$. The *discriminant* of such a diagonal form is just the square class of the product of the diagonal entries $(-a)(-b)(ab) = 1 \in F^\times/F^{\times 2}$.

We might now try to classify quaternion algebras over F up to isomorphism in terms of this quadratic form. As it turns out, one needs to consider maps on quadratic forms coming from not just an invertible change of basis (an *isometry*) but allowing also scaling of the quadratic form by a nonzero element of F. Then (Theorem 4.4.5) the map $B \mapsto \operatorname{nrd}|_{B^0}$ gives a bijection

$$\left\{ \begin{array}{c} \text{Quaternion algebras over } F \\ \text{up to isomorphism} \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} \text{Nonsingular ternary} \\ \text{quadratic spaces over } F \\ \text{up to similarity} \end{array} \right\}.$$

The similarity class of a nonsingular ternary quadratic form also cuts out a unique plane conic $C \subseteq \mathbb{P}^2$, so one also has a bijection between isomorphism classes of quaternion algebras over F and isomorphism classes of conics over F.

The problem of classifying quaternion algebras over F is now rephrased in terms of quadratic forms, and consequently the answer depends in an intrinsic way of the field F. In this vein, the most basic question we can ask about a quaternion algebra B is if it is isomorphic to the matrix ring $B \cong \mathrm{M}_2(F)$: if so, we say that B is *split* over F. Every quaternion algebra over $\mathbb{C}$ (or an algebraically closed field) is split, and a quaternion algebra $\left(\dfrac{a,b}{\mathbb{R}}\right)$ is split if and only if $a > 0$ or $b > 0$. Ultimately, we will find six equivalent ways (Theorem 4.5.5) to check if a quaternion algebra B is split; in the language of quadratic forms, B is split if and only if the quadratic form $\operatorname{nrd}|_{B^0}$ is *isotropic*, meaning it represents 0 nontrivially (there exists $\alpha \in B^0 \setminus \{0\}$ such that $\operatorname{nrd}(\alpha) = 0$).

In later chapters, we will consider this problem as we gradually increase the "arithmetic complexity" of the field F. We have our answer already for $F = \mathbb{R}, \mathbb{C}$; the answer becomes more complex as we move to a finite field $F = \mathbb{F}_q$, then to nonarchimedean local fields (such as p-adic fields), and finally to global fields.

Before concluding this introductory section, we make one final connection to the theory of quadratic forms. In Section 2.3, we say that the unit Hamiltonians $\mathbb{H}_1^\times$ act on the pure Hamiltonians $\mathbb{H}^0$ (Section 2.3) by *rotations*: the Euclidean quadratic form is preserved by conjugation. This generalizes in a natural way to an arbitrary field, and so we can understand the group that preserves a ternary (or quaternary) form in terms of the unit group of a quaternion algebra (Proposition 4.8.6).

4.2 Definitions

In this section, we give a brief summary of some basic definitions and notation in the theory of quadratic forms over fields.

Definition 4.2.1. A *quadratic form* Q is a map $Q : V \to F$ on an F-vector space V satisfying:

(i) $Q(ax) = a^2 Q(x)$ for all $a \in F$ and $x \in V$; and

(ii) The map $T : V \times V \to F$ defined by

$$T(x,y) = Q(x+y) - Q(x) - Q(y)$$

is F-bilinear.

We call the pair (V, Q) a *quadratic space* and T the *associated bilinear form.*

We will often abbreviate (V, Q) by simply V.

Definition 4.2.2. A *similarity* of quadratic forms $Q : V \to F$ and $Q' : V' \to F$ is an F-linear isomorphism $f : V \to V'$ and $u \in F^\times$ such that $Q(x) = uQ'(f(x))$ for all $x \in V$, i.e., such that the diagram

$$\begin{tikzcd}
V \arrow[r, "Q"] \arrow[d, "f", "\wr"'] & F \arrow[d, "u", "\wr"'] \\
V' \arrow[r, "Q'"] & F
\end{tikzcd}$$

commutes.

An *isometry* of quadratic forms (or *isomorphism* of quadratic spaces) is a similarity with $u = 1$; we write in this case $Q \cong Q'$.

Remark 4.2.3. A similarity allows isomorphisms of the target F (as a one-dimensional F-vector space). The notion of "isometry" comes from the connection with measuring lengths, when working with the usual Euclidean norm form on a vector space over $\mathbb{R}$: similarity allows these lengths to scale uniformly.

If Q is a quadratic form then the associated bilinear form T satisfies $T(x, y) = T(y, x)$ for all $x, y \in V$, i.e., T is *symmetric*. Conversely, given a symmetric bilinear form T, if $\operatorname{char} F \neq 2$ then the map $Q : V \to F$ defined by $Q(x) = T(x, x)/2$ is a quadratic form whose associated bilinear form is T, and so there is an equivalence of categories between quadratic forms and symmetric bilinear forms over F.

Definition 4.2.4. A symmetric bilinear form $T : V \times V \to F$ is *nondegenerate* if for all $0 \neq x \in V$, the linear functional $T_x : V \to F$ by $T_x(y) = T(x, y)$ is nonzero.

A quadratic form Q (or quadratic space V) is *nondegenerate* (or *nonsingular* or *regular*) if the associated bilinear form T is nondegenerate.

A quadratic space is nondegenerate if and only if the map $V \to \operatorname{Hom}(V, F)$ given by $x \mapsto (y \mapsto T(x, y))$ is injective (and hence an isomorphism if $\dim_F V < \infty$).

4.2.5. Let B be an algebra over F with a standard involution. Then $\operatorname{nrd} : B \to F$ is a quadratic form on B. We have $\operatorname{nrd}(ax) = a^2\alpha$ for all $\alpha \in B$, and the map T given by

$$T(\alpha,\beta) = (\alpha+\beta)\overline{(\alpha+\beta)} - \alpha\overline{\alpha} - \beta\overline{\beta} = \alpha\overline{\beta} + \beta\overline{\alpha} = \alpha\overline{\beta} + \overline{\alpha\overline{\beta}} = \operatorname{trd}(\alpha\overline{\beta}) \qquad (4.2.6)$$

for $\alpha, \beta \in B$ is indeed bilinear.

In particular, we have $T(1, \alpha) = \operatorname{trd}(\alpha)$ and

$$\alpha\beta + \beta\alpha = \operatorname{trd}(\beta)\alpha + \operatorname{trd}(\alpha)\beta - T(\alpha, \beta). \qquad (4.2.7)$$

If nrd is nonsingular, then we say the standard involution on B is *nonsingular*.

From now on, let $Q : V \to F$ be a quadratic form with $n = \dim_F V < \infty$.

If $Q : V \to F$ is a quadratic form, then a basis $e_1, \dots, e_n$ for V gives an isomorphism $V \cong F^n$ in which Q can be written

$$Q(x) = Q(x_1e_1 + \dots + x_ne_n) = \sum_i Q(e_i)x_i^2 + \sum_{i<j} T(e_i, e_j)x_ix_j.$$

Let $e_1, \dots, e_n \in V$. We define

$$d(e_1, \dots, e_n) = \det(T(e_i, e_j))_{i,j}$$

where $(T(e_i, e_j))_{i,j} \in \mathrm{M}_n(F)$ is the (symmetric) matrix whose (i, j)th entry is equal to $T(e_i, e_j)$, called the *Gram matrix* of the quadratic form. If $A \in \mathrm{M}_n(F)$ is a matrix such that $Ae_i = e_i'$, then

$$\begin{aligned} d(e_1', \dots, e_n') &= \det(T(Ae_i, Ae_j))_{i,j} = \det(A)^2 \det(T(e_i, e_j))_{i,j} \\ &= \det(A)^2 d(e_1, \dots, e_n). \end{aligned} \qquad (4.2.8)$$

If n is odd, then working with a generic quadratic form $t_{11}x_1^2 + \cdots + t_{nn}x_n^2$ over the field $\mathbb{Q}(t_{11}, \dots, t_{nn})$ with t_{ij} transcendental, we find that $d(e_1, \dots, e_n)$ is divisible by 2 as an element of $\mathbb{Z}[t_{11}, \dots, t_{nn}]$; let $d/2$ denote this polynomial divided by 2.

Definition 4.2.9. The *discriminant* $\operatorname{disc}(Q) \in F/F^{\times 2}$ of Q is

$$\operatorname{disc}(Q) = \begin{cases} d(e_1, \dots, e_n) \in F/F^{\times 2}, & \text{if } n \text{ is even;} \\ (d/2)(e_1, \dots, e_n) \in F/F^{\times 2}, & \text{if } n \text{ is odd;} \end{cases}$$

where $e_1, \dots, e_n$ is any basis for V.

In odd dimension, the discriminant is sometimes called the *half-discriminant* .

The discriminant is well defined by (4.2.8). In particular, note that Q is nonsingular if and only if $\operatorname{disc}(Q) \in F^\times/F^{\times 2}$ (Exercise 4.1). When it will cause no confusion, we will represent the class of the discriminant in $F/F^{\times 2}$ simply by a representative element in F.

Let $T : V \times V \to F$ be the symmetric bilinear form associated to Q.

Definition 4.2.10. We say that $x \in V$ is *orthogonal* to $y \in V$ (with respect to Q) if $T(x, y) = 0$.

Since T is symmetric, x is orthogonal to y if and only if y is orthogonal to x for $x, y \in V$, and so we simply say x, y are orthogonal. If $S \subseteq V$ is a subset, we write

$$S^\perp = \{x \in V : T(v, x) = 0 \text{ for all } v \in S\}$$

for the subspace of V which is orthogonal to (the span of) S.

4.2.11. If B is an F-algebra with a standard involution (cf. Paragraph 4.2.5), then α, β are orthogonal with respect to nrd if and only if

$$\operatorname{trd}(\alpha\overline{\beta}) = \alpha\overline{\beta} + \beta\overline{\alpha} = 0.$$

Let $Q : V \to F$ and $Q' : V' \to F$ be quadratic forms. We define the *orthogonal sum* $Q \perp Q'$ on $V \oplus V'$, with associated bilinear pairing $T \perp T'$, by the conditions

$$(T \perp T')(x + x', y + y') = T(x, y) + T(x', y')$$
$$(Q \perp Q')(x + x') = Q(x) + Q(x')$$

for all $x, y \in V$ and $x', y' \in V'$; the latter follows from the former when $\operatorname{char} F \neq 2$. By definition, $V' \subseteq V^\perp$ (and $V \subseteq (V')^\perp$). We similarly define the *tensor product* $Q \otimes Q'$ on $V \otimes V'$ by

$$(T \otimes T')(x \otimes x', y \otimes y') = T(x, y)T'(x', y')$$
$$(Q \otimes Q')(x \otimes x') = Q(x)Q(x').$$

4.2.12. To some extent, one can often restrict to the case where a quadratic form Q is nondegenerate by *splitting off the radical*, as follows. We define the *radical* to be

$$\operatorname{rad}(Q) = V^\perp \cap \{v \in V : Q(v) = 0\}.$$

If $\operatorname{char} F \neq 2$, then any $v \in V^\perp$ has $Q(v) = T(v,v)/2 = 0$, so the second term is unnecessary. The map Q on $V^\perp$ is an additive map ($Q(v + w) = Q(v) + Q(w)$), so we could also write the radical as $\ker Q|_{V^\perp}$ (the kernel taken as an abelian group). Consequently, we have $\operatorname{rad}(Q) \subset V$ is an F-subspace, so completing a basis we can write $V = \operatorname{rad}(Q) \perp W$—the direct sum is an orthogonal direct sum by definition of the radical. Now $Q|_{\operatorname{rad}(Q)}$ is identically zero and $Q|_W$ is nondegenerate.

We now define several quadratic forms on F^n. For $a \in F$, the quadratic form $Q(x) = ax^2$ on F is denoted $\langle a\rangle$; for $a_1, \dots, a_n \in F$, we abbreviate

$$\langle a_1\rangle \perp \cdots \perp \langle a_n\rangle = \langle a_1, \dots, a_n\rangle$$

for the quadratic form on F^n.

The following result is a standard application of orthogonalization (Exercise 4.2) and can be proven by induction. Here is the first time where we assume that $\operatorname{char} F \neq 2$.

Lemma 4.2.13. *Suppose* $\operatorname{char} F \neq 2$*, and let* $Q : V \to F$ *be a quadratic form with* $\dim_F V < \infty$*. Then there exists a basis of* V *such that* $Q \cong \langle a_1, \dots, a_n\rangle$ *with* $a_i \in F$*.*

A form presented with a basis as in Lemma 4.2.13 is called *normalized* (or *diagonal*). For the modifications required when $\operatorname{char} F = 2$, we refer to Chapter 5.

Example 4.2.14. A normalized quadratic form $\langle a_1, \dots, a_n\rangle$ has discriminant $d = a_1 \cdots a_n \in F/F^{\times 2}$ and hence is nonsingular if and only if $a_1 \cdots a_n \neq 0$ (Exercise 4.1).

Example 4.2.15. Let $B = \left(\dfrac{a,b}{F}\right)$ be a quaternion algebra. Then by Paragraph 3.2.8, the quadratic form $\operatorname{nrd} : B \to F$ is normalized with respect to the basis $1, i, j, ij$. Indeed, we have

$$\operatorname{nrd} \cong \langle 1, -a, -b, ab\rangle \cong \langle 1, -a\rangle \otimes \langle 1, -b\rangle.$$

To conclude, we state and prove an important foundational result, due to Witt.

Proposition 4.2.16 (Witt cancellation)**.** *Let* $Q \cong Q'$ *be nondegenerate quadratic forms such that* $Q \cong Q_1 \perp Q_2$ *and* $Q' \cong Q_1 \perp Q_2'$*. Then* $Q_2 \cong Q_2'$*.*

Proof. Induction on the number of variables. □

4.3 Nonsingular standard involutions

In this section, we follow Theorem 3.5.1 with a characterization of quaternion algebras beyond division algebras by the nonsingularity of the standard involution (see Paragraph 4.2.5).

Theorem 4.3.1. *Suppose* char $F \neq 2$ *and let B be an F-algebra. Then B has a nonsingular standard involution if and only if one of the following holds:*

(i) *$B = F$;*

(ii) *$B = K$ is a quadratic F-algebra with either K a field or $K \cong F \times F$; or*

(iii) *B is a quaternion algebra over F.*

By Exercise 3.11, there exist F-algebras with standard involution having arbitrary dimension, so it is remarkable that the additional requirement that the standard involution be nonsingular gives such a tidy result. Case (ii) in Theorem 4.3.1 is equivalent to requiring that K be a quadratic F-algebra that is *reduced* (has no nonzero nilpotent elements).

Proof. If $B = F$, then the standard involution is the identity and nrd is nonsingular. If $\dim_F K = 2$, then after completing the square we have $K \cong F[x]/(x^2 - a)$ and in the basis $1, x$ we have nrd $\cong \langle 1, a\rangle$. By Example 4.2.14, nrd is nonsingular if and only if $a \in F^\times$ if and only if K is a quadratic field extension of F or $K \cong F \times F$.

So suppose that $\dim_F B > 2$. Let $1, i, j$ be a part of a normalized basis for B with respect to the quadratic form nrd. Then we have $T(1, i) = \operatorname{trd}(i) = 0$, so $i^2 = a \in F^\times$, since nrd is nonsingular. Note in particular that $\bar{i} = -i$. Similarly we have $j^2 = b \in F^\times$, and by (4.2.7) we have $\operatorname{trd}(ij) = ij + ji = 0$. We have $T(1, ij) = \operatorname{trd}(ij) = 0$, and $T(ij, i) = \operatorname{trd}(\bar{i}(ij)) = a\operatorname{trd}(j) = 0$ and similarly $T(ij, j) = 0$, so $ij \in \{1, i, j\}^\perp$. If $ij = 0$ then $i(ij) = aj = 0$ so $j = 0$, a contradiction. Since nrd is nonsingular, it follows then that the set $1, i, j, ij$ is linearly independent.

Therefore, the subalgebra A of B generated by i, j satisfies $A \cong \left(\dfrac{a, b}{F}\right)$, so if $\dim_F B = 4$ we are done. So let $k \in A^\perp$, so in particular $\operatorname{trd}(k) = 0$ and $k^2 = c \in F^\times$. Thus $k \in B^\times$, with $k^{-1} = c^{-1}k$. By Paragraph 4.2.11 we have $k\alpha = \overline{\alpha}k$ for any $\alpha \in A$ since $\bar{k} = -k$. But then

$$k(ij) = (\overline{ij})k = \bar{j}\,\bar{i}k = \bar{j}ki = k(ji). \tag{4.3.2}$$

But $k \in B^\times$ so $ij = ji = -ij$, and this is a contradiction. □

Corollary 4.3.3. *Let B be an F-algebra with* char $F \neq 2$. *Then B is a quaternion algebra if and only if B is noncommutative and has a nonsingular standard involution.*

4.4 Isomorphism classes of quaternion algebras

In Section 2.3, we found that the unit Hamiltonians act by conjugation on the pure quaternions $\mathbb{H}^0 \cong \mathbb{R}^3$ as rotations, preserving the standard inner product. In this section, we return to the theme suggested by this line of inquiry for a general quaternion algebra, and we characterize isomorphism classes of quaternion algebras as isometry classes of ternary quadratic forms.

Let B be a quaternion algebra over F.

Definition 4.4.1. $\alpha \in B$ is *scalar* if $\alpha \in F$ and *pure* if $\operatorname{trd}(\alpha) = 0$.

4.4.2. Let

$$B^0 = \{\alpha \in B : \operatorname{trd}(\alpha) = 0\} = \{1\}^{\perp}$$

be the F-vector space of pure elements of B. The standard involution restricted to B^0 is given by $\alpha \mapsto -\alpha$ for $\alpha \in B^0$. For $B = \left(\dfrac{a,b}{F}\right)$, we have $B^0 = Fi \oplus Fj \oplus Fij$ and in this basis

$$\operatorname{nrd}|_{B^0} \cong \langle -a, -b, ab\rangle \tag{4.4.3}$$

so that $\operatorname{disc}(\operatorname{nrd}|_{B^0}) = (ab)^2 = 1 \in F^\times/F^{\times 2}$.

Proposition 4.4.4. *Let A and B be quaternion algebras over F. Then the following are equivalent.*

(i) *$A \cong B$ as F-algebras;*

(ii) *$A \cong B$ as quadratic spaces; and*

(iii) *$A^0 \cong B^0$ as quadratic spaces.*

Proof. The implication (i) ⇒ (ii) follows from the fact that the standard involution on an algebra is unique and the reduced norm is determined by this standard involution, so the reduced norm on A corresponds to the reduced norm on B. The implication (ii) ⇒ (iii) follows since A^0 and B^0 are defined as the spaces orthogonal to 1 and so are preserved by an isometry of quadratic spaces.

So finally we prove (iii) ⇒ (i). Let $f : A^0 \to B^0$ be an isomorphism of quadratic spaces. Then f extends in the natural way to an F-linear map $f : A \to B$ by mapping $1 \mapsto 1$. We need to show that f is a homomorphism (hence isomorphism) of F-algebras. Suppose $A \cong \left(\dfrac{a,b}{F}\right)$. Then we have $\operatorname{nrd}(f(i)) = \operatorname{nrd}(i) = -a$ and

$$\operatorname{nrd}(f(i)) = f(i)\overline{f(i)} = -f(i)^2$$

so $f(i)^2 = a$. Similarly we have $f(j)^2 = b$. Finally, we have $ji = -ij$ since i, j are orthogonal (as in the proof of Theorem 4.3.1), but then $f(i), f(j)$ are orthogonal as well and so $f(j)f(i) = -f(i)f(j)$. [[Redefine $f(ij) = f(i)f(j)$ to obtain a ring homomorphism.]] Together, these imply that $B \cong \left(\dfrac{a,b}{F}\right) \cong A$. □

Theorem 4.4.5. *Let F be a field with* char $F \neq 2$. *Then the map $B \mapsto \operatorname{nrd}|_{B^0}$ yields a bijection*

$$\left\{\begin{array}{c}\textit{Quaternion algebras over } F\\ \textit{up to isomorphism}\end{array}\right\} \longleftrightarrow \left\{\begin{array}{c}\textit{Nonsingular ternary quadratic spaces}\\ \textit{over } F \textit{ with discriminant } 1 \in F^\times/F^{\times 2}\\ \textit{up to isometry}\end{array}\right\}$$
$$\longleftrightarrow \left\{\begin{array}{c}\textit{Nonsingular ternary}\\ \textit{quadratic spaces over } F\\ \textit{up to similarity}\end{array}\right\}$$

that is functorial with respect to F.

By the expression *functorial with respect to F*, we simply mean that this bijection respects (is compatible with) field extensions. Explicitly, if $F \hookrightarrow F'$ is an inclusion of fields, and B is a quaternion algebra associated to the ternary quadratic space $Q : B^0 \to F$, then $B' = B \otimes_F F'$ is associated to the extension $B^0 \otimes_F F' \to F'$ to F'; this boils down to the fact that $(B')^0 = (B \otimes_F F')^0 = B^0 \otimes_F F'$.

Proof. The association $B \mapsto \operatorname{nrd}|_{B^0}$ gives a well-defined functor from quaternion algebras to nonsingular ternary quadratic spaces with discriminant 1, by Paragraph 4.4.2; the map sends isomorphisms to isometries and vice versa by Proposition 4.4.4. To conclude, show that the functor is essentially surjective. Let V be a nonsingular ternary quadratic space with discriminant $1 \in F^{\times 2}$. Choose a normalized basis for V, so that $Q \cong \langle -a, -b, c\rangle$ with $a, b, c \in F^\times$. We have $\operatorname{disc}(Q) = abc \in F^{\times 2}$, so rescaling the third basis vector we may assume $c = ab$. We then associate to V the isomorphism class of the quaternion algebra $\left(\dfrac{a,b}{F}\right)$. The result follows.

For the second equivalence, we examine the natural map from isometry classes to similarity classes. Every ternary quadratic form (indeed, any quadratic form in odd dimension) is similar to a unique isometry class of quadratic forms with trivial discriminant: if $Q = \langle a, b, c\rangle$ with $a, b, c \in F^\times$, then

$$Q = \langle a, b, c\rangle \sim (abc)\langle a, b, c\rangle = \langle a^2bc, ab^2c, abc^2\rangle \cong \langle bc, ac, ab\rangle$$

and $\operatorname{disc}(\langle bc, ac, ab\rangle) = (abc)^2 = 1 \in F^\times/F^{\times 2}$. Therefore the map is essentially surjective, and a similarity between quadratic spaces yields an isometry of the associated spaces with trivial discriminant. □

Remark 4.4.6. In fact, Proposition 4.4.4 and Theorem 4.4.5 can be understood more conceptually via the *Clifford algebra* of a quadratic form (Exercise 4.5); the inverse functor is called the *even Clifford functor*. See Paragraph 4.10.2.

Corollary 4.4.7. *There is a functorial bijection between the set of isomorphism classes of quaternion algebras over F and the set of nonsingular ternary quadratic spaces over F.*

Therefore, the problem of classifying quaternion algebras depends on the theory of quadratic forms over that field. As fields increase in arithmetic complexity, the latter becomes increasingly difficult.

4.5 Splitting

With Proposition 4.4.4 in hand, we are now ready to characterize the matrix ring and division rings among quaternion algebras.

Let $Q : V \to F$ be a quadratic form. We say that Q *represents* an element $a \in F$ if there exists $x \in V$ such that $Q(x) = a$.

Definition 4.5.1. Q (or V) is *isotropic* if Q represents 0 nontrivially (there exists $0 \neq x \in V$ such that $Q(x) = 0$) and otherwise Q is *anisotropic*.

Definition 4.5.2. A quadratic form Q is a *hyperbolic plane* if $Q \cong H$ where $H : F^2 \to F$ is defined by $H(x, y) = xy$.

When $\operatorname{char} F \neq 2$, a hyperbolic plane is isometric to the form $\langle 1, -1 \rangle$. Clearly, a hyperbolic plane H represents every element of F; we say H is *universal*.

Lemma 4.5.3. *Suppose Q is nonsingular. Then Q is isotropic if and only if $Q \cong H \perp Q'$ with H a hyperbolic plane.*

Proof. Exercise 4.4. □

Lemma 4.5.4. *Suppose Q is nonsingular and let $a \in F^\times$. Then the following are equivalent.*

(i) *Q represents a;*

(ii) *$Q \cong \langle a \rangle \perp Q'$;*

(iii) *$\langle -a \rangle Q$ is isotropic.*

Proof. For (i) $\Rightarrow$ (ii), we take $Q' = Q|_W$ and $W = \{v\}^\perp \subset V$. For (ii) $\Rightarrow$ (iii), we note that $\langle -a\rangle \perp Q \cong \langle a, -a\rangle \perp Q'$ is isotropic. For (iii) $\Rightarrow$ (i), suppose $Q(v) = ax^2$ with $v \in V$ and $x \in F$. If $x = 0$, then Q is isotropic, so by Lemma 4.5.3 represents a; if $x \neq 0$, then by homogeneity $Q(v/x) = a$ and again Q represents a. □

From now on, we assume that $\operatorname{char} F \neq 2$.

Theorem 4.5.5. *Let* $B = \left(\dfrac{a,b}{F}\right)$ *be a quaternion algebra over* F *(with* $\operatorname{char} F \neq 2$*). Then the following are equivalent:*

(i) $B \cong \left(\dfrac{1,1}{F}\right) \cong \mathrm{M}_2(F)$*;*

(ii) *B is not a division ring;*

(iii) *The quadratic form* nrd *is isotropic;*

(iv) *The quadratic form* $\operatorname{nrd}|_{B^0}$ *is isotropic;*

(v) *The binary form* $\langle a, b\rangle$ *represents* 1*; and*

(vi) $b \in \mathrm{N}_{K/F}(K^\times)$ *where* $K = F[i]$.

In (vi), if K is not a field then $K \cong F \times F$ and $\mathrm{N}_{K/F}(K^\times) = \operatorname{nrd}(K^\times) = F^\times$.

Proof. The isomorphism $\left(\dfrac{1,1}{F}\right) \cong \mathrm{M}_2(F)$ in (i) follows from Example 2.2.4. The implication (i) $\Rightarrow$ (ii) is clear. The equivalence (ii) $\Leftrightarrow$ (iii) follows from the fact that $\alpha \in B^\times$ if and only if $\operatorname{nrd}(\alpha) \in F^\times$ (Exercise 3.4).

We now prove (iii) $\Rightarrow$ (iv). Let $0 \neq \alpha \in B$ be such that $\operatorname{nrd}(\alpha) = 0$. If $\operatorname{trd}(\alpha) = 0$, then we are done. Otherwise, we have $\operatorname{trd}(\alpha) \neq 0$. Let β be orthogonal to $1, \alpha$, so that $\operatorname{trd}(\alpha\beta) = 0$. We cannot have both $\alpha\beta = 0$ and $\overline{\alpha}\beta = (\operatorname{trd}(\alpha) - \alpha)\beta = 0$, so we may assume $\alpha\beta \neq 0$. But then $\operatorname{nrd}(\alpha\beta) = \operatorname{nrd}(\alpha)\operatorname{nrd}(\beta) = 0$ as desired.

To complete the equivalence of the first four we prove (iv) $\Rightarrow$ (i). Let $\beta \in B^0$ satisfy $\operatorname{nrd}(\beta) = 0$. Since $\operatorname{nrd}|_{B^0}$ is nonsingular, there exists $0 \neq \alpha \in B^0$ such that $\operatorname{trd}(\alpha\overline{\beta}) \neq 0$. Therefore, the restriction of nrd to $F\alpha \oplus F\beta$ is nonsingular and isotropic. By Lemma 4.5.3, we conclude there exists a basis for B^0 such that $\operatorname{nrd} \cong \langle 1, -1\rangle \perp \langle c\rangle = \langle 1, -1, c\rangle$; but $\operatorname{disc}(\operatorname{nrd}|_{B^0}) = -c = 1 \in F^\times/F^{\times 2}$ by Paragraph 4.4.2 so rescaling we may assume $c = -1$. But then by Proposition 4.4.4 we have $B \cong \left(\dfrac{1,1}{F}\right)$.

Now we show (iv) $\Rightarrow$ (v). For $\alpha \in B^0$, we have

$$\operatorname{nrd}(\alpha) = \operatorname{nrd}(xi + yj + zij) = -ax^2 - by^2 + abz^2$$

as in Paragraph 4.4.2. Suppose $\operatorname{nrd}(\alpha) = 0$. If $z = 0$, then the binary form $\langle a, b\rangle$ is isotropic so is a hyperbolic plane by Lemma 4.5.3 and thus represents 1. If $z \neq 0$ then

$$a\left(\frac{x}{az}\right)^2 + b\left(\frac{y}{bz}\right)^2 = 1.$$

Next we prove (v) ⇒ (vi). If $a \in F^{\times 2}$ then $K \cong F \times F$ and $\mathrm{N}_{K/F}(K^\times) = F^\times \ni b$. If $a \notin F^{\times 2}$, then given $ax^2 + by^2 = 1$ we must have $y \neq 0$ so

$$\left(\frac{1}{y}\right)^2 - a\left(\frac{x}{y}\right)^2 = \mathrm{N}_{K/F}\left(\frac{1 - x\sqrt{a}}{y}\right) = b.$$

To conclude, we prove (vi) ⇒ (iii). If $b = \mathrm{N}_{K/F}(K^\times) = x^2 - ay^2$, then $\alpha = x + yi + j \neq 0$ has $\operatorname{nrd}(\alpha) = x^2 - ay^2 - b = 0$. □

We give a name to the equivalent conditions in Theorem 4.5.5.

Definition 4.5.6. A quaternion algebra B over F is *split* if $B \cong \mathrm{M}_2(F)$. A field K containing F is a *splitting field* for B if $B \otimes_F K$ is split.

Lemma 4.5.7. *Let $K \supset F$ be a quadratic extension of fields with* $\operatorname{char} F \neq 2$. *Then K is a splitting field for B if and only if there is an injective F-algebra homomorphism $K \hookrightarrow B$.*

Proof. First, suppose $\iota : K \hookrightarrow B$. We may assume that $K = F(\sqrt{d})$ with $d \in F^\times$. Let $\mu = \iota(\sqrt{d})$. Then $1 \otimes \sqrt{d} - \mu \otimes 1$ is a zerodivisor in $B \otimes_F K$:

$$(1 \otimes \sqrt{d} - \mu \otimes 1)(1 \otimes \sqrt{d} + \mu \otimes 1) = 1 \otimes \sqrt{d}^2 - \mu^2 \otimes 1 = 1 \otimes d - d \otimes 1 = 0.$$

By Theorem 4.5.5, we have $B \otimes_F K \cong \mathrm{M}_2(K)$.

Conversely, suppose $B \otimes_F K \cong \mathrm{M}_2(K)$. Consider the standard involution on K, which we denote for emphasis by σ. Then σ acts as an F-linear involution on $B \otimes_F K$ by $\sigma(\alpha \otimes a) = \alpha \otimes \sigma(a)$, and so $\sigma(\alpha \otimes a) = \alpha \otimes a$ if and only if $a \in F$.

If $B \cong \mathrm{M}_2(F)$ already, then any quadratic field K embeds in B (take a matrix in rational normal form) and indeed $B \otimes_F K \cong \mathrm{M}_2(K)$ for any K. So by Theorem 4.5.5, we may suppose B is a division ring. Let $K = F(\sqrt{d})$. We have $B \otimes_F K \cong \mathrm{M}_2(K)$ if and only if $\langle -a, -b, ab\rangle$ is isotropic over K, which is to say there exist $x, y, z, u, v, w \in F$ such that

$$-a(x + u\sqrt{d})^2 - b(y + v\sqrt{d})^2 + ab(z + w\sqrt{d})^2 = 0. \qquad (4.5.8)$$

Let $\alpha = xi + yj + zij$ and $\beta = ui + vj + wij$. Then $\operatorname{trd}(\alpha) = \operatorname{trd}(\beta) = 0$. Expansion of (4.5.8) (Exercise 4.15) shows that α is orthogonal to β, so $\operatorname{trd}(\alpha\beta) = 0$, and that

$\operatorname{nrd}(\alpha) + d\operatorname{nrd}(\beta) = 0$. Since B is a division ring, if $\operatorname{nrd}(\beta) = c = 0$ then $\beta = 0$ so $\operatorname{nrd}(\alpha) = 0$ as well and $\alpha = 0$, a contradiction. So $\operatorname{nrd}(\beta) \neq 0$, and the element $\gamma = \alpha\beta^{-1} = c^{-1}\alpha\beta \in B$ has $\operatorname{nrd}(\gamma) = -d$ and $\operatorname{trd}(\gamma) = c^{-1}\operatorname{trd}(\alpha\beta) = 0$ so $\gamma^2 = d$ as desired. □

4.6 Conics

Following Theorem 4.4.5, we are led to consider the zero locus of the quadratic form $\operatorname{nrd}|_{B^0}$ up to scaling; this gives a geometric way to view the precedings results.

Definition 4.6.1. A *conic* $C \subset \mathbb{P}^2$ over F is a nonsingular projective plane curve of degree 2. Two conics C, C' are *isomorphic* over F if there exists a linear change of variables $f \in \mathrm{PGL}_3(F)$ such that $f(C) = C'$.

If we identify

$$\mathbb{P}(B^0) = (B^0 \setminus \{(0,0,0)\}/F^\times$$

with the points of the projective plane $\mathbb{P}^2(F)$ over F, then the vanishing locus $C = V(\operatorname{nrd}|_{B^0})$ of $\operatorname{nrd}|_{B^0}$ defines a conic over F: here, nonsingularity of the quadratic form implies nonsingularity of the associated plane curve (Exercise 4.12). As above, if we take the basis i, j, ij for B^0, then the conic C is defined by the vanishing of the equation $\operatorname{nrd}(xi + yj + zij) = -ax^2 - by^2 + abz^2 = 0$.

The following corollary is then simply a rephrasing of Theorem 4.4.5.

Corollary 4.6.2. *The map $B \mapsto C = V(\operatorname{nrd}|_{B^0})$ yields a bijection*

$$\left\{\begin{matrix} \textit{Quaternion algebras over } F \\ \textit{up to isomorphism} \end{matrix}\right\} \longleftrightarrow \left\{\begin{matrix} \textit{Conics over } F \\ \textit{up to isomorphism} \end{matrix}\right\}$$

that is functorial with respect to F.

Theorem 4.5.5 also extends to this context.

Theorem 4.6.3. *The following are equivalent:*

(i) $B \cong \mathrm{M}_2(F)$*;*

(vii) *The conic C associated to B has an F-rational point.*

4.7 Hilbert symbol

Theorem 4.5.5(v) is called *Hilbert's criterion* for the splitting of a quaternion algebra: the quaternion algebra $\left(\dfrac{a,b}{F}\right)$ is split if and only if the *Hilbert equation* $ax^2+by^2=1$ has a solution with $x,y\in F$.

Definition 4.7.1. We define the *Hilbert symbol*

$$\begin{aligned} F^\times \times F^\times &\to \{\pm 1\} \\ (a,b) &\mapsto (a,b)_F \end{aligned}$$

by the condition that $(a,b)_F = 1$ if and only if the quaternion algebra $\left(\dfrac{a,b}{F}\right)$ is split.

The similarity between the symbols $(a,b)_F$ and $\left(\dfrac{a,b}{F}\right)$ is intentional. The Hilbert symbol is in fact well defined as a map $F^\times/F^{\times 2}\times F^\times/F^{\times 2}\to\{\pm 1\}$ by Exercise 2.4.

Lemma 4.7.2. *Let $a\in F^\times$. Then the following statements hold:*

(a) $(1,a)_F = (a,-a)_F = 1$.

(b) $(a,1-a)_F = 1$ *if* $a\neq 1$.

Proof. For (a), the Hilbert equation $x^2+ay^2=1$ has the obvious solution $(x,y)=(1,0)$. And $\langle a,-a\rangle$ is isotropic (taking $(x,y)=(1,1)$) so is a hyperbolic plane so represents 1, as in the proof of Theorem 4.5.5, or we argue

$$(a,-a)_F = (a,a^2)_F = (a,1)_F = (1,a)_F = 1$$

by Exercise 2.4. Finally, part (c): we have $(a,1-a)_F=1$ since $ax^2+(1-a)y^2=1$ has the solution $(x,y)=(1,1)$. □

The study of symbols like the Hilbert symbol leads naturally to the definition of K_2 of a field (see Paragraph 4.10.5).

4.8 Orthogonal groups

In this section, we revisit the original motivation of Hamilton (Section 2.3) in a more general context, relating quaternions to the orthogonal group of a quadratic form.

Let Q be a nonsingular quadratic form on the finite-dimensional F-vector space V.

Definition 4.8.1. The *orthogonal group* of Q is the group of isometries of Q, i.e.,

$$\mathrm{O}(Q) = \{f \in \mathrm{Aut}_F(V) : Q(f(x)) = Q(x) \text{ for all } x \in V\}.$$

One can understand this orthogonal group quite concretely as follows. Choose a basis $e_1, \dots, e_n$ for $V \cong F^n$ and let $A = (T(e_i, e_j))_{i,j=1,\dots,n}$ be the Gram matrix of Q with respect to this basis, so that

$$Q(x) = x^t A x \quad \text{and} \quad T(x, y) = x^t A y.$$

Then $\mathrm{Aut}_F(V) \cong \mathrm{GL}_n(F)$, and so $M \in \mathrm{GL}_n(F)$ belongs to $\mathrm{O}(Q)$ if and only if

$$Q(Mx) = (Mx)^t A(Mx) = x^t M^t A M x = x^t A x = Q(x)$$

for all $x \in V \cong F^n$, so

$$\mathrm{O}(Q) = \{M \in \mathrm{GL}_n(F) : M^t A M = A\}. \tag{4.8.2}$$

There is a natural map

$$\det : \mathrm{O}(Q) \to F^\times;$$

the image of det is $\operatorname{img}\det = \{\pm 1\}$ by (4.8.2).

Definition 4.8.3. An isometry $f \in \mathrm{O}(Q)$ is *special* if it has $\det f = 1$. The *special orthogonal group* of Q is the group of special isometries of Q:

$$\mathrm{SO}(Q) = \ker\det = \{f \in \mathrm{O}(Q) : \det(f) = 1\}.$$

The subgroup $\mathrm{SO}(Q) \leq \mathrm{O}(Q)$ is a (normal) subgroup of index 2.

Example 4.8.4. If $V = \mathbb{R}^n$ and Q is the usual norm on V, then

$$\mathrm{O}(Q) = \mathrm{O}(n) = \{A \in \mathrm{GL}_n(\mathbb{R}) : AA^t = 1\}$$

and $\mathrm{SO}(Q)$ is the usual group of rotations of V. In particular, $\mathrm{SO}(2) \cong \mathbb{S}^1$ is the circle group.

In general, the structure of the orthogonal group of a quadratic form can be quite interesting and complicated. However, a classical theorem of Cartan and Dieudonné at least tells us that the orthogonal group is generated by reflections, as follows.

Theorem 4.8.5 (Cartan–Dieudonné). *Let (V, Q) be a nonsingular quadratic space with $\dim_F V = n$. Then every isometry $f \in \mathrm{O}(Q)$ is a product of at most n reflections.*

Let $B^0 = \{x \in B : \operatorname{trd}(B) = 0\}$. Then we have a (left) action $B^\times \circlearrowright B^0$ defined by $x \mapsto \alpha x \alpha^{-1}$, since $\operatorname{trd}(\alpha x \alpha^{-1}) = \operatorname{trd}(x) = 0$. Moreover, let $Q^0 = \operatorname{nrd}|_{B^0}$ is the restriction of the reduced norm to B^0 then $B^\times$ acts by isometries, since by multiplicativity

$$\operatorname{nrd}(\alpha x \alpha^{-1}) = \operatorname{nrd}(x)$$

for all $\alpha \in B$ and $x \in B^0$. The kernel of the action is given by those $\alpha \in B^\times$ with $\alpha x \alpha^{-1} = x$ for all $x \in B^0$, and this implies that $\alpha \in F^\times$.

Proposition 4.8.6. *The sequences*

$$1 \to F^\times \to B^\times \to \mathrm{SO}(Q^0) \to 1$$

is exact. If $\operatorname{nrd}(B^\times) \subseteq F^{\times 2}$*, then the sequence*

$$1 \to \{\pm 1\} \to B_1^\times \to \mathrm{SO}(Q^0) \to 1$$

is exact.

Example 4.8.7. If $B \cong \mathrm{M}_2(F)$, then $\operatorname{nrd} = \det$, so $\det^0 \cong \langle 1, -1, -1 \rangle$ and $\mathrm{PGL}_2(F) \cong \mathrm{SO}(\langle 1, -1, -1 \rangle)$.

If $F = \mathbb{R}$ and $B = \mathbb{H}$, then $\det(\mathbb{H}) = \mathbb{R}_{>0} = \mathbb{R}^{\times 2}$, and the second exact sequence is Hamilton's (Section 2.3).

4.9 Equivalence of categories

Definition 4.9.1. A quadratic space is *oriented* with a choice of isomorphism

$$i : \det(V) = \textstyle\bigwedge^n(V) \to F,$$

and a similarity or isometry is *oriented* if [[finish]].

One can beef up the set-theoretic bijection to a full equivalence of categories by putting in an orientation.

4.10 Extensions and further reading

4.10.1. The terminology *isotropic* is as least as old as Eichler [Eic53, p. 3], and goes perhaps back to Witt. The word can be used to mean "having properties that are identical in all directions", and so the motivation for this language probably comes from physics: the second fundamental form associated to a parametrized surface $z = f(x, y)$ in $\mathbb{R}^3$ is a quadratic form, and (roughly speaking) this quadratic form defines the curvature at a given point. In this sense, if the quadratic form vanishes, then the curvature is zero, and indeed things look the same in all directions.

4.10.2. Proposition 4.4.4 is just about the even Clifford algebra.

4.10.3. Lemma 4.5.7 requires a calculation; but for a general central simple algebra, there is no short proof, since one only knows in the hard direction that B is Brauer equivalent to an algebra where K embeds.

4.10.4. Brauer-Severi varieties.

4.10.5. *Algebraic K-theory* (K for the German "Klasse", following Grothendieck [Kar10]), in its various formulations, seeks to understand certain functors from rings to abelian groups. For a field F, the lower K-groups are trivial: $K_0 = \{1\}$ and $K_1(F) = F^\times$; however, by a theorem of Matsumoto [Mat69], the group $K_2(F)$ is the universal domain for symbols over F:

$$K_2(F) = F^\times \otimes_{\mathbb{Z}} F^\times / \langle a \otimes (1-a) : a \neq 0, 1 \rangle.$$

The map $a \otimes b \mapsto (a,b)_F$ extends to a map $K_2(F) \to \{\pm 1\}$, a *Steinberg symbol*, a homomorphism from $K_2(F)$ to a multiplicative abelian group. The higher K-groups are related to deeper arithmetic of commutative rings. For an introduction, see Weibel [Wei13].

4.11 Algorithmic aspects

Diagonalization of quadratic forms, making the equivalences algorithmic.

Recognizing quaternion algebras.

Exercises

4.1. Show that a quadratic form Q on a finite-dimensional space is nonsingular if and only if it has nonzero discriminant $d(Q) \in F^\times/F^{\times 2}$.

4.2. Prove that every (finite-dimensional) quadratic space has a normalized basis (Lemma 4.2.13).

4.3. Let B be a quaternion algebra over F. Let $N : B \to F$ and $\Delta : B \to F$ be defined by $N(x) = \operatorname{trd}(x^2)$ and $\Delta(x) = \operatorname{trd}(x)^2 - 4\operatorname{nrd}(x)$. Show that N, Δ are quadratic forms on B, describe their associated bilinear forms, and compute a normalized form (and basis) for each.

4.4. Prove Lemma 4.5.3: if V is a nonsingular quadratic space that is isotropic, then V contains a hyperbolic plane.

4.5. Let $Q : V \to F$ be a nonsingular ternary quadratic form. In this exercise, we construct directly the even Clifford algebra of Q and show that it is a quaternion algebra. This gives another proof of the results of section 4.4.

Let M be the F-vector space $F \oplus (V \otimes V)$ and let C be the quotient of M by the subspace of elements $x \otimes x - Q(x)$ for $x \in V$.

a) Show that the relation $x \otimes y + y \otimes x = T(x, y)$ holds in B for all $x, y \in V$. Conclude that $\dim_F C = 4$.

b) Show that $\otimes$ yields a multiplication law on C. *[Hint: Choose a normalized basis for V.]* We call C the *even Clifford algebra* of Q.

c) Show that the map $x \otimes y \mapsto T(x, y) - x \otimes y$ defines a nonsingular standard involution on C. (In particular, note that if $T(x, y) = 0$ then the involution is $x \otimes y \mapsto y \otimes x$.) Conclude that C is a quaternion algebra over F.

4.6. Let $Q : V \to F$ be quadratic form with $\dim_F V = n < \infty$. A subspace $W \subseteq V$ is *totally isotropic* if $Q|_W = 0$ is identically zero. The *Witt index* of a quadratic form $\nu(V) = \nu(Q)$ is the maximal dimension of a totally isotropic subspace.

a) Show that if $\nu(V) = m$ then $2m \leq n$.

b) [[Isotropy of Pfister forms]].

c) Suppose that Q is an isotropic Pfister form. Let $W \subset V$ be a subspace of dimension $n - 1$. Show that $Q|_W$ is isotropic, giving another proof of Theorem 4.5.5 (iii) $\Rightarrow$ (iv).

4.7. Recall Remark 3.3.5. Let B be a finite-dimensional F-algebra (not necessarily a quaternion algebra), and let $\mathrm{Tr} : B \to F$ be the (right) algebra trace.

a) Show that the map $B \to F$ defined by $x \mapsto \mathrm{Tr}(x^2)$ is a quadratic form on B; this form is called the *(right) trace form* on B.

b) Compute the trace form of $A \times B$ and $A \otimes_F B$ in terms of the trace form of A and B.

c) Show that if K/F is a finite inseparable field extension (with $[K : F] < \infty$) then the trace form on K (as an F-algebra) is identically zero. On the other hand, show that if K/F is a finite separable field extension (with $\operatorname{char} F \neq 2$) then the trace form is nonsingular.

d) Compute the trace form on $\mathbb{Q}(\sqrt{5})$ and $\mathbb{Q}(\alpha)$ where $\alpha = 2\cos(2\pi/7)$, so that $\alpha^3 + \alpha^2 - 2\alpha - 1 = 0$.

4.8. Show that the reduced norm is the unique nonzero quadratic form Q on B that is *multiplicative* , i.e., $Q(\alpha\beta) = Q(\alpha)Q(\beta)$ for all $\alpha,\beta \in B$.

4.9. Use Theorem 4.5.5(vi) to give another proof that there is no division quaternion algebra B over a finite field $F = \mathbb{F}_q$ (with q odd).

4.10. Let $a, b, c \in F^\times$. Show that

$$\left(\frac{a,b}{F}\right) \otimes_F \left(\frac{a,c}{F}\right) \cong D \otimes_F \left(\frac{c,-c}{F}\right) \cong \mathrm{M}_2(D)$$

where $D = \left(\dfrac{a,bc}{F}\right)$.

4.11. Show that $(-1,10)_{\mathbb{Q}} = 1$.

4.12. Show that if Q is a nonsingular quadratic form on a vector space V with $\dim_F V = n$, then the equation $Q(x) = 0$ defines a nonsingular projective variety in $\mathbb{P}^n$ of degree 2, called a *quadric*.

4.13. Show that

$$\left(\frac{-2,-3}{\mathbb{Q}}\right) \cong \left(\frac{-1,-1}{\mathbb{Q}}\right) \text{ but that } \left(\frac{-2,-5}{\mathbb{Q}}\right) \not\cong \left(\frac{-1,-1}{\mathbb{Q}}\right)$$

(cf. Lam [Lam05, Examples III.2.12–13]). *[Hint: In the first case, show that the quadratic forms are isometric directly; in the second case, note that the quadratic form* $\langle 2,5,10\rangle$ *represents* 7 *but* $\langle 1,1,1\rangle$ *does not.]*

4.14. Let p be prime. Show that $\left(\dfrac{-1,p}{\mathbb{Q}}\right) \cong \mathrm{M}_2(\mathbb{Q})$ if and only if $p = 2$ or $p \equiv 1 \pmod 4$.

4.15. Expand (4.5.8) and prove that if $\alpha = xi + yj + zij$ and $\beta = ui + vj + wij$, then α is orthogonal to β, so $\operatorname{trd}(\alpha\beta) = 0$, and that $\operatorname{nrd}(\alpha) + d\operatorname{nrd}(\beta) = 0$.

4.16. Let $a \in \mathbb{Q}^\times \setminus \mathbb{Q}^{\times 2}$.

a) Show that there are infinitely many distinct isomorphism classes of conics $x^2 - ay^2 = bz^2$ for $b \in \mathbb{Q}^\times$.

b) Show that

$$\lim_{X\to\infty} \frac{\#\{p < X : p \text{ prime}, x^2 - ay^2 = pz^2 \text{ has no } \mathbb{Q}\text{-rational points}\}}{\#\{p < X : p \text{ prime}\}} = \frac{1}{2}.$$

c) Generalize (b) to all integers.

d) What is the probability that a random conic in $\mathbb{P}^2$, ordered by height, has a rational point? What is the probability that a random quaternion algebra $\left(\dfrac{a,b}{\mathbb{Q}}\right)$ with $a, b \in \mathbb{Q}^\times$ is split?

Chapter 5

Quaternion algebras in characteristic 2

In this chapter, we extend the results from the previous three chapters to the neglected case where the base field has characteristic 2. Throughout this chapter, let F be a field with algebraic closure $\overline{F}$.

5.1 Separability and another symbol

To get warmed up, we give a different notation (symbol) for quaternion algebras that holds in any characteristic and which is convenient for many purposes.

Let B be a commutative, finite-dimensional algebra over F. We say B is *separable* if

$$B \otimes_F \overline{F} \cong \overline{F} \times \cdots \times \overline{F}.$$

If $B \cong F[x]/(f(x))$ with $f(x) \in F[x]$, then B is separable if and only if f has distinct roots in $\overline{F}$.

5.1.1. If $\operatorname{char} F \neq 2$, a quadratic F-algebra K is separable if and only if $K \cong F[x]/(x^2 - a)$ with $a \neq 0$ if and only if K is a field or $K \cong F \times F$.

5.1.2. If $\operatorname{char} F = 2$, then a quadratic F-algebra K is separable if and only if $K \cong F[x]/(x^2 + x + a)$ for some $a \in F$, and any quadratic algebra of the form $K = F[x]/(x^2 + a)$ with $a \in F$ is not separable.

Now the more general more general notation.

5.1.3. Let K be a separable quadratic F-algebra, and let $b \in F^\times$. We denote by

$$\left(\frac{K,b}{F}\right) = K + Kj$$

the F-algebra with basis $1, j$ as a left K-vector space and with the multiplication rules $j^2 = b$ and $j\alpha = \overline{\alpha} j$ for $\alpha \in K$, where $\overline{}$ is the standard involution on K. (Since K is separable over F, the standard involution is the nontrivial element of $\operatorname{Gal}(K/F)$.)

From Paragraph 5.1.1, we see that $\left(\dfrac{K,b}{F}\right)$ is a quaternion algebra over F if $\operatorname{char} F \neq 2$; we will show this also holds in characteristic 2 and so gives a characteristic-free way to define quaternion algebras.

In using this symbol, we are breaking the symmetry between the standard generators i, j, but otherwise have not changed anything about the definition.

5.2 Characteristic 2

From the previous section, we now see how to define quaternion algebras in characteristic 2. Throughout this section, suppose that $\operatorname{char} F = 2$.

Definition 5.2.1. An algebra B over F (with $\operatorname{char} F = 2$) is a *quaternion algebra* if there exists an F-basis $1, i, j, k$ for B such that

$$i^2 + i = a, \;\; j^2 = b, \;\; \text{and} \;\; k = ij = j(i+1) \tag{5.2.2}$$

with $a \in F$ and $b \in F^\times$.

Just as when $\operatorname{char} F \neq 2$, we find that the multiplication table for a quaternion algebra B is determined by the rules (5.2.2). We denote by $\left[\dfrac{a,b}{F}\right)$ the F-algebra with basis $1, i, j, ij$ subject to the multiplication rules (5.2.2). The algebra $\left[\dfrac{a,b}{F}\right)$ is not symmetric in a, b (explaining the choice of notation), but it is still functorial in the field F.

If we let $K = F[i] \cong F[x]/(x^2 + x + a)$, then

$$\left[\frac{a,b}{F}\right) \cong \left(\frac{K,b}{F}\right)$$

and our notation extends that of Section 5.1.

Example 5.2.3. The ring $\mathrm{M}_2(F)$ of 2×2-matrices with coefficients in F is again a quaternion algebra over F: indeed, we have an isomorphism $\left[\dfrac{1,1}{F}\right) \xrightarrow{\sim} \mathrm{M}_2(F)$ with $i \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$, $j \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

The following result can be proven in a similar way as Lemma 2.2.5.

Lemma 5.2.4. *An F-algebra B is a quaternion algebra if and only if there exist generators $i, j \in B$ satisfying*

$$i^2 + i = a, \ \ j^2 = b, \ \ \textit{and} \ \ ij = j(i+1). \tag{5.2.5}$$

5.2.6. Let $B = \left[\dfrac{a, b}{F}\right)$ be a quaternion algebra over F. Then B has a standard involution given by

$$\alpha = t + xi + yj + zij \mapsto \overline{\alpha} = x + \alpha = (t+x) + xi + yj + zij$$

since

$$(t + xi + yj + zij)((t+x) + xi + yj + zij) = t^2 + tx + ax^2 + by^2 + byz + abz^2 \in F.$$

Consequently, one has a reduced trace and reduced norm on B as in Chapter 3.

We now state a version of Theorem 3.5.1 in characteristic 2; the proof is similar and is left as an exercise.

Theorem 5.2.7. *Let B be a division F-algebra with a standard involution which is not the identity. Then either B is a separable quadratic field extension of F or B is a quaternion algebra over F.*

Proof. Exercise 5.5. □

5.3 Quadratic forms and characteristic 2

We now turn to the theory of quadratic forms in characteristic 2. Here, we have $T(x,x) = 2Q(x) = 0$, so there is no longer an equivalence between quadratic forms and symmetric bilinear forms—the former is the more "fundamental" object. To define normalized forms, we will also have need of the following form: for $a, b \in F$, we define $Q(x,y) = ax^2 + axy + by^2$ on F^2 and abbreviate $Q = [a, b]$.

Lemma 5.3.1. *Let $Q : V \to F$ be a quadratic form with $\dim_F V < \infty$. Then there exists a basis of V such that*

$$Q \cong [a_1, b_1] \perp \cdots \perp [a_m, b_m] \perp \langle c_1, \dots, c_n \rangle$$

with $a_i, b_i, c_j \in F$; morevoer, Q is nonsingular if and only if $a_1 \cdots a_m \neq 0$ and

$$(\dim_F V \textit{ is even and } n = 0) \textit{ or } (\dim_F V \textit{ is odd and } n = 1).$$

Proof. Exercise 5.6. □

Example 5.3.2. Let $B = \left[\dfrac{a,b}{F}\right)$ be a quaternion algebra. Then $1, i, j, ij$ is a normalized basis for B, indeed we have

$$\mathrm{nrd} \cong [1,a] \perp [b,ab] \cong [1,a] \otimes \langle 1, b\rangle,$$

and consequently nrd is nonsingular.

We now consider the characterization of quaternion algebras as those equipped with a nonsingular standard involution (Theorem 4.3.1).

Proposition 5.3.3. *Let B be an F-algebra (with* $\operatorname{char} F = 2$*). Then B has a nonsingular standard involution if and only if B is a separable quadratic F-algebra or B is a quaternion algebra over F.*

Proof. If $B = F$, then the standard involution is the identity but it is not nonsingular, as $\mathrm{trd}(1) = 2 = 0$.

If $\dim_F B = 2$, then $B = K$ has a unique standard involution (Lemma 3.4.1). By Paragraph 5.1.2, we see that the involution is nonsingular if and only if K is separable.

So suppose $\dim_F B > 2$. Since B has a nonsingular standard involution, there exists an element $i \in B$ such that $T(i,1) = \mathrm{trd}(i) \neq 0$. We have $i \notin F$ since $\mathrm{trd}(F) = \{0\}$. Rescaling we may assume $\mathrm{trd}(i) = 1$, whence $i^2 = i + a$ for some $a \in F$, and $\mathrm{nrd}\,|_{F+Fi} = [1,a]$. We have begun the proof of Lemma 5.3.1, and $1, i$ is part of a normalized basis, in this special case.

By nondegeneracy, there exists $j \in \{1,i\}^{\perp}$ such that $\mathrm{nrd}(j) = b \neq 0$. Thus $\mathrm{trd}(j) = 0$ so $\overline{j} = j$ and $j^2 = b \in F^{\times}$. Furthermore, we have

$$0 = \mathrm{trd}(ij) = ij + j\overline{i} = ij + j(i+1)$$

so $ij = j(i+1)$. Therefore i, j generate an F-subalgebra $A \cong \left[\dfrac{a,b}{F}\right)$.

The conclusion of the proof follows exactly as in (4.3.2): if $k \in \{1, i, j, ij\}^{\perp}$ then $k(ij) = k(ji)$, a contradiction. □

Next, we characterize isomorphism classes of quaternion algebras in characteristic 2 in the language of quadratic forms.

Let B be a quaternion algebra over F. We again define

$$B^0 = \{\alpha \in B : \mathrm{trd}(\alpha) = 0\} = \{1\}^{\perp}.$$

But now $B^0 = F \oplus Fj \oplus Fij$ and in this basis we have

$$\mathrm{nrd}\,|_{B^0} \cong \langle 1\rangle \perp [b, ab].$$

with discriminant

$$\operatorname{disc}(\operatorname{nrd}|_{B^0}) = b^2 = 1 \in F^\times/F^{\times 2}, \tag{5.3.4}$$

recalling that in odd dimension, the discriminant is halved (Definition 4.2.9).

Theorem 5.3.5. *The maps $B \mapsto \operatorname{nrd}|_{B^0} \mapsto V(\operatorname{nrd}|_{B^0})$ yield equivalences of categories*

$$\begin{aligned}\left\{\begin{matrix}\textit{Quaternion algebras over } F\\ \textit{up to isomorphism}\end{matrix}\right\} &\longleftrightarrow \left\{\begin{matrix}\textit{Nonsingular ternary}\\ \textit{quadratic forms over } F\\ \textit{with discriminant } 1 \in F^\times/F^{\times 2}\\ \textit{up to isometry}\end{matrix}\right\}\\ &\longleftrightarrow \left\{\begin{matrix}\textit{Nonsingular ternary}\\ \textit{quadratic forms over } F\\ \textit{up to similarity}\end{matrix}\right\}\\ &\longleftrightarrow \left\{\begin{matrix}\textit{Conics over } F\\ \textit{up to isomorphism}\end{matrix}\right\}\end{aligned}$$

Proof. We extend the proof of Proposition 4.4.4 to char $F = 2$. Let $f : A^0 \to B^0$ be an isomorphism of quadratic spaces. Let $A \cong \left[\dfrac{a,b}{F}\right)$. Extend f to an F-linear map $A \to B$ by mapping $i \mapsto b^{-1}f(ij)f(j)$. The map f preserves 1: it maps F to F by Exercise 5.8, since $F = (B^0)^\perp = (A^0)^\perp$, and $1 = \operatorname{nrd}(1) = \operatorname{nrd}(f(1)) = f(1)^2$ so $f(1) = 1$. We have $f(j)^2 = \operatorname{nrd}(f(j)) = \operatorname{nrd}(j) = b$ as before, since $j, ij \in A^0$. Thus

$$\begin{aligned}1 = \operatorname{trd}(i) = b^{-1}\operatorname{trd}((ij)j) = b^{-1}T(ij, j) =\\ = b^{-1}T(f(ij), f(j)) = \operatorname{trd}(b^{-1}f(ij)f(j)) = \operatorname{trd}(f(i))\end{aligned}$$

and similarly $\operatorname{nrd}(f(i)) = \operatorname{nrd}(i) = a$, so we have $f(i)^2 + f(i) + a = 0$. Finally, $f(i)f(j) = f(j)(f(i) + 1)$ by Exercise 5.9. So the map f is an isomorphism of quaternion algebras. This map is surjective since

To show the map is surjective, let V be a nonsingular ternary quadratic space with discriminant $1 \in F^{\times 2}$. Then $Q \cong \langle u\rangle \perp [b, c]$ for some $u, b \in F^\times$ and $c \in F$. Scaling by u shows that $Q \sim \langle 1\rangle \perp [bu^{-1}, cu^{-1}]$, and this arises from the quaternion algebra $\left[\dfrac{a, bu^{-1}}{F}\right)$ with $a = ucb^{-1}$.

The rest of the equivalence follows as in the proof of Theorem 4.4.5. □

We now turn to identifying the matrix ring in characteristic 2.

Definition 5.3.6. A quadratic form $H : V \to F$ is a *hyperbolic plane* if $H \cong [1, 0]$, i.e., $H(x, y) = x^2 + xy = x(x + y)$.

As in Lemma 4.5.3, we have the following result.

Lemma 5.3.7. *If Q is nonsingular and isotropic then $Q \cong H \perp Q'$ with H a hyperbolic plane.*

We may again characterize division quaternion algebras by examination of the reduced norm as a quadratic form as in Theorems 4.5.5 and 4.6.3.

Theorem 5.3.8. *Let $B = \left[\dfrac{a,b}{F}\right)$ (with $\operatorname{char} F = 2$). Then the following are equivalent:*

(i) $B \cong \left[\dfrac{1,1}{F}\right) \cong \mathrm{M}_2(F)$;

(ii) *B is not a division ring;*

(iii) *The quadratic form* nrd *is isotropic;*

(iv) *The quadratic form* $\operatorname{nrd}|_{B^0}$ *is isotropic;*

(v) *The binary form $[1,a]$ represents b;*

(vi) *$b \in \mathrm{N}_{K/F}(K^\times)$ where $K = F[i]$; and*

(vii) *The conic $C = V(\operatorname{nrd}|_{B^0}) \subset \mathbb{P}^2$ has an F-rational point.*

Proof. Only condition (v) requires significant modification in the case $\operatorname{char} F = 2$; see Exercise 5.7. □

Analogous to section 4.7, one can define a symbol $[a,b)_F$ for the splitting of quaternion algebras in characteristic 2. This symbol is no longer called the Hilbert symbol, but there is still an analogue of the *Hilbert equation*, namely $\left[\dfrac{a,b}{F}\right)$ is split if and only if $bx^2 + bxy + aby^2 = 1$ has a solution with $x, y \in F$.

Lemma 5.3.9. *Let $K \supset F$ be a quadratic extension of fields. Then K is a splitting field for B if and only if there is an injective F-algebra homomorphism $K \hookrightarrow B$.*

Proof. If $\iota : K \hookrightarrow B$ and $K = F(\alpha)$, then $1 \otimes \alpha - \iota(\alpha) \otimes 1$ is a zerodivisor in $B \otimes_F K$, with

$$(1 \otimes \alpha - \iota(\alpha) \otimes 1)(1 \otimes \alpha - \iota(\overline{\alpha}) \otimes 1) = 0 \tag{5.3.10}$$

by Exercise 5.10.

Conversely, let $K = F(\alpha)$ and suppose $B \otimes_F K \cong \mathrm{M}_2(K)$. Without loss of generality, we may assume that $B = \left[\dfrac{a,b}{F}\right)$ is a division ring. By Theorem 5.3.8(v), tthere exist $x, y, z, u, v, w \in F$ not all zero such that

$$(x + u\alpha)^2 + b(y + v\alpha)^2 + b(y + v\alpha)(z + w\alpha) + ab(z + w\alpha)^2 = 0 \tag{5.3.11}$$

so

$$(u^2 + bv^2 + bvw + abw^2)\alpha^2 + (vz + wy)b\alpha + (x^2 + by^2 + byz + abz^2) = 0. \quad (5.3.12)$$

Let $\beta = x+yj+zij$ and $\gamma = u+vj+wij$. Then $\gamma \in B^\times$, since $\gamma = 0$ implies $\operatorname{nrd}(\beta) = 0$ and yet B is a division ring. But then the element

$$\mu = \beta\gamma^{-1} = (u^2 + uv + av^2 + bw^2)^{-1}\beta(v + \gamma)$$

satisfies the same equation as (5.3.12), so the embedding $\alpha \mapsto \mu$ gives an embedding $K \hookrightarrow B$. □

5.4 Extensions and further reading

Exercises

5.1. Give a version of the primitive element theorem as follows. Let B be a commutative F-algebra with $\dim_F B > \operatorname{char} F$. Show that $B \cong F[x]/(f(x))$ for some $f(x) \in F[x]$.

5.2. Let B be a quaternion algebra over F with $\operatorname{char} F \neq 2$, and let $K \subseteq B$ be a separable quadratic F-algebra. Show that there exists $b \in F^\times$ such that $B \cong \left(\dfrac{K, b}{F}\right)$ (as in Paragraph 5.1.3).

5.3. Let K be a separable quadratic F-algebra and let $u, binF^\times$. Show that $\left(\dfrac{K, b}{F}\right) \cong \left(\dfrac{K, ub}{F}\right)$ if and only if $u \in \operatorname{nrd}(K^\times) = \mathrm{N}_{K/F}(K^\times)$.

5.4. Let $\operatorname{char} F = 2$ and let $a \in F$ and $b \in F^\times$.

a) Show that $\left[\dfrac{a, b}{F}\right) \cong \left[\dfrac{a, ab}{F}\right)$ if $a \neq 0$.

b) Show that if $t \in F$ a,nd $u \in F^\times$, then $\left[\dfrac{a, b}{F}\right) \cong \left[\dfrac{a + (t + t^2), bu^2}{F}\right)$.

5.5. Let $\operatorname{char} F = 2$ and let B be a division F-algebra with a standard involution. Prove that either the standard involution is the identity (and so B is classified by Exercise 3.7), or that the conclusion of Theorem 3.5.1 holds for B: namely, that either $B = K$ is a separable quadratic field extension of F or that B is a quaternion algebra over F. *[Hint: Replace conjugation by i by the map $\phi(x) = ix + xi$, and show that $\phi^2 = \phi$. Then diagonalize and proceed as in the case* $\operatorname{char} F \neq 2$.*]*

5.6. Prove Lemma 5.3.1, that every quadratic form over F with $\operatorname{char} F = 2$ has a normalized basis.

5.7. Prove Theorem 5.3.8.

5.8. Let B be a quaternion algebra over F (with F of arbitrary characteristic). Show that $F = (B^0)^{\perp}$.

5.9. Let A, B be quaternion algebras over F with $\operatorname{char} F = 2$, let $f : A^0 \to B^0$ be an isometry, and define $f(i) = f(ij)f(j)\operatorname{nrd}(j)^{-1}$. Show that $f(i)f(j) = f(j)(f(i) + 1)$, as in the proof of Theorem 5.3.5.

5.10. Verify (5.3.10).

5.11. Prove Wedderburn's theorem the following special case: a finite quaternion ring of even cardinality is not a division ring. *[Hint: See Exercise* 3.12.*]*

Chapter 6

Simple algebras

6.1 The "simplest" algebras

In this chapter, we return to the characterization of quaternion algebras. We initially defined quaternion algebras in terms of generators and relations in Chapter 2; in the chapters that followed, we showed that quaternion algebras are equivalently noncommutative algebras with a nonsingular standard involution. Here, we pursue another approach, and characterize quaternion algebras in a different way.

Consider now the "simplest" sorts of algebras. Field extensions are perhaps the simplest kinds of algebras, studied in a first course in abstract algebra. Division rings are closest to fields—they just lack commutativity—but miss the matrix rings. So we look for a concept that includes them both.

Like the primes among the integers or the finite simple groups among finite groups, it is natural to seek algebras that cannot be "broken down" any further. Accordingly, we say that a ring B is *simple* if it has nontrivial two-sided ideals, i.e., the only two-sided ideals are $\{0\}$ and B. To show the power of this notion, consider this: if $\phi : B \to A$ is a ring homomorphism and B is simple, then ϕ is either injective or the zero map (since $\ker \phi \subseteq B$ is a two-sided ideal).

A division ring is simple, since every element is a unit so every ideal (left, right, or two-sided) is trivial. In particular, a field is a simple ring, and in fact, a *commutative* ring is simple if and only if it is a field. The matrix ring $\mathrm{M}_n(F)$ over a field F is also simple, something that can be checked directly by multiplying by matrix units (Exercise 6.4).

Moreover, quaternion algebras are simple. The shortest proof of this statement, given what we have done so far, is to employ Theorems 4.5.5 and 5.3.8: a quaternion algebra B over F is either isomorphic to $\mathrm{M}_2(F)$ or is a division ring, and in either case is simple. One can also prove this directly (Exercise 6.1).

Although the primes are quite mysterious and the classification of finite simple groups is a monumental achievement in group theory, the situation for algebras is quite simple, indeed!

Theorem 6.1.1 (Wedderburn–Artin)**.** *Let F be a field and B be a finite-dimensional F-algebra. Then B is simple if and only if $B \cong \mathrm{M}_n(D)$ where $n \geq 1$ and D is a finite-dimensional division F-algebra.*

A corollary of this theorem is another characterization of quaternion algebras. Recall that an F-algebra B is *central* if the center of B is F. Quaternion algebras are central (Exercise 2.7).

Corollary 6.1.2. *Let B be an F-algebra. Then the following are equivalent:*

(i) *B is a quaternion algebra;*

(ii) *$B \otimes_F \overline{F} \cong \mathrm{M}_2(\overline{F})$, where $\overline{F}$ is an algebraic closure of F; and*

(iii) *B is a central simple algebra of dimension $\dim_F B = 4$.*

Moreover, a central simple algebra B of dimension $\dim_F B = 4$ is either a division algebra or has $B \cong \mathrm{M}_2(F)$.

This corollary has the neat consequence that a division algebra B over F is a quaternion algebra over F if and only if it is central of dimension $\dim_F B = 4$.

For the reader in a hurry, we now give a proof of this corollary without invoking the Wedderburn–Artin theorem, previewing some of the ideas that go into it.

Proof of Corollary. The statement (i) $\Rightarrow$ (ii) was proven in Exercise 2.4(c).

To prove (ii) $\Rightarrow$ (iii), suppose B is an algebra with $\overline{B} = B \otimes_F \overline{F} \cong \mathrm{M}_2(\overline{F})$. The $\overline{F}$-algebra $\overline{B}$ is central simple, from above. Thus $Z(B) = Z(\overline{B}) \cap F = F$. And if I is a two-sided ideal of B then $\overline{I} = I \otimes_F \overline{F}$ is a two-sided ideal of $\overline{B}$, so $\overline{I} = \{0\}$ or $\overline{I} = \overline{B}$ is trivial, whence $I = \overline{I} \cap F$ is trivial.

Finally, we prove (iii) $\Rightarrow$ (i). Let B a central simple F-algebra of dimension 4. If B is a division algebra we are done; so suppose not. Then B has a nontrivial left ideal, generated by a nonunit, and let $\{0\} \subsetneq I \subsetneq B$ be a nontrivial left ideal with $0 < m = \dim_F I$ minimal. Then we have a nonzero homomorphism $B \to \mathrm{End}_F(I) \cong \mathrm{M}_m(F)$ which is injective, since B is simple. By dimensions, we cannot have $m = 1$ and if $m = 2$, then $B \cong \mathrm{M}_2(F)$. So suppose $m = 3$. Then by minimality, every nontrivial left ideal of B has dimension 3. But for any $\alpha \in B$, we have that $I\alpha$ is a left ideal, so the left ideal $I \cap I\alpha$ is either $\{0\}$ or I; in either case, $I\alpha \subseteq I$, so I is a right ideal as well. But this contradicts the fact that B is simple. $\square$

The Wedderburn–Artin theorem is an important structural result used throughout mathematics, so we give in this chapter a reasonably self-contained account of its proof. More generally, it will be convenient to work with *semisimple* algebras, which for the purposes of this introduction can be thought of as finite direct products of simple algebras. Indeed, when treating ideals of an algebra we would be remiss if we did not discuss more generally modules over the algebra, and the notions of simple and semisimple module are natural concepts in linear algebra and representation theory: a semisimple module is one that is a direct sum of simple modules ("completely reducible"), analogous to a semisimple operator where every invariant subspace has an invariant complement (e.g., a diagonalizable matrix).

The second important result in this chapter is a theorem that concerns the simple subalgebras of a simple algebra.

Theorem 6.1.3 (Skolem–Noether)**.** *Let A, B be simple F-algebras and suppose that B is central. Suppose that $f, g : A \to B$ are homomorphisms. Then there exists $\beta \in B$ such that $f(\alpha) = \beta^{-1} g(\alpha)\beta$ for all $\alpha \in A$.*

Corollary 6.1.4. *Every F-algebra automorphism of a simple F-algebra B is inner, i.e.,* $\operatorname{Aut}_F(B) \cong B^\times/F^\times$.

Just as above, for our quaternionic purposes, we can give a direct proof.

Corollary 6.1.5. *Let B be a quaternion algebra over F and let $K_1, K_2 \subset B$ be quadratic subfields. Suppose that $\phi : K_1 \xrightarrow{\sim} K_2$ is an isomorphism of F-algebras. Then ϕ lifts to an inner automorphism of B, i.e., there exists $\beta \in B$ such that $K_2 = \beta^{-1} K_1 \beta$.*

Proof. We have $K_1 = F[\alpha_1] \cong F[x]/(x^2 - tx + n)$, where $\alpha_1 \in B$ and $t, n \in F$. Let $\alpha_2 = \phi(\alpha_1) \in K_2 \subseteq B$. We want to find $\beta \in B^\times$ such that $\beta\alpha_1 = \alpha_2\beta$.

In the special case $B \cong \mathrm{M}_2(F)$, then $\alpha_1, \alpha_2 \in \mathrm{M}_2(F)$ satisfy the same irreducible characteristic polynomial, so by the theory of rational canonical forms, we have $\alpha_2 = \beta^{-1}\alpha\beta$ where $\beta \in B^\times \cong \mathrm{GL}_2(F)$.

In general, the set

$$\{\beta \in B : \beta\alpha_1 = \alpha_2\beta\}$$

is an F-vector subspace of B. We have $B \otimes_F \overline{F} \cong \mathrm{M}_2(\overline{F})$, so there exists a nonzero $\beta \in B \otimes_F \overline{F}$ that will do; but by linear algebra, this means that there exists a nonzero $\beta \in B$ with the desired property. If $B \not\cong \mathrm{M}_2(F)$ then B is a division ring, so $\beta \in B^\times$, and we are done. □

6.2 Simple modules

Throughout this chapter, let B be a finite-dimensional F-algebra.

To understand the algebra B, we look at its representations. A *representation* of B (over F) is a vector space V over F together with an F-algebra homomorphism $B \to \mathrm{End}_F(V)$. Equivalently, a representation is given by a left (or right) B-module V: this is almost a tautology. Although one can define infinite-dimensional representations, they will not interest us here, and we assume throughout that $\dim_F V < \infty$, or equivalently that V is a finitely generated (left or right) B-module. If we choose a basis for V, we obtain an isomorphism $\mathrm{End}_F(V) \cong \mathrm{M}_n(F)$ where $n = \dim_F V$, so a representation is just a homomorphic way of thinking of the algebra B as an algebra of matrices.

Example 6.2.1. The space of column vectors F^n is a left $\mathrm{M}_n(F)$-module; the space of row vectors is a right $\mathrm{M}_n(F)$-module.

Example 6.2.2. B is itself a left B-module, giving rise to the *left regular representation* $B \to \mathrm{End}_F(B)$ (cf. Paragraph 2.2.8 and Remark 3.3.5).

Example 6.2.3. Let G be a finite group. Then a representation of $F[G]$ (is the same as an $F[G]$-module which) is the same as a homomorphism $G \to \mathrm{GL}(V)$, where V is an F-vector space (Exercise 3.5).

Definition 6.2.4. Let V be a left B-module. Then V is *simple* (or *irreducible*) if $V \neq \{0\}$ and the only B-submodules of V are $\{0\}$ and V.

We say V is *indecomposable* if V cannot be written as $V = V_1 \oplus V_2$ with $V_i \neq \{0\}$ for $i = 1, 2$.

A simple module is indecomposable, but the converse need not hold, and this is a central point of difficulty in understanding representations.

Example 6.2.5. If $B = \left\{\begin{pmatrix} a & b \\ 0 & c \end{pmatrix} : a, b, c \in F\right\} \subseteq \mathrm{M}_2(F)$, then the space $V = F^2$ of column vectors is not simple, since the subspace spanned by $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is a B-submodule; nevertheless, V is indecomposable (Exercise 6.3).

The importance of simple modules is analogous to that of simple groups. Indeed, arguing by induction on the dimension of V, we have the following lemma analogous to the Jordan–Hölder theorem on composition series.

Lemma 6.2.6. *A (finite-dimensional) left B-module V admits a filtration*

$$V = V_0 \supsetneq V_1 \supsetneq V_2 \supsetneq \cdots \supsetneq V_r = \{0\}$$

such that V_i/V_{i+1} is simple for each i.

This filtration is not unique, but up to isomorphism and permutation, the quotients V_i/V_{i+1} are unique.

Having defined the notion of simplicity for modules, we now consider simplicity of the algebra B.

Definition 6.2.7. An F-algebra B is *simple* if the only two-sided ideals of B are $\{0\}$ and B.

Equivalently, B is simple if and only if any F-algebra (or even ring) homomorphism $B \to A$ is either injective or the zero map.

Remark 6.2.8. The two notions of simplicity are related as follows: B is simple as an algebra if and only if it is simple as a left and right B-module.

Example 6.2.9. A division F-algebra D is simple. In fact, the F-algebra $\mathrm{M}_n(D)$ is simple for any division F-algebra D (Exercise 6.4), so in particular $\mathrm{M}_n(F)$ is simple.

Example 6.2.10. Let $\overline{F}$ be an algebraic closure of F. If $B \otimes_F \overline{F}$ is simple, then B is simple. Indeed, the association $I \mapsto I \otimes_F \overline{F}$ is an injective map from the set of two-sided ideals of B to the set of two-sided ideals of $B \otimes_F \overline{F}$.

6.2.11. If B is a quaternion algebra over F, then B is simple. By Exercise 2.7, we have $B \otimes_F \overline{F} \cong \mathrm{M}_2(\overline{F})$, which is simple by Example 6.2.9, so B is simple by Example 6.2.10.

Example 6.2.9 shows that algebras of the form $\mathrm{M}_n(D)$ with D a division F-algebra yield a large class of simple F-algebras. In fact, these are all such algebras, a fact we will now prove. First, a few preliminary results.

Lemma 6.2.12 (Schur)**.** *Let B be an F-algebra. Let V_1, V_2 be simple B-modules. Then any homomorphism $\phi : V_1 \to V_2$ of B-modules is either zero or an isomorphism.*

Proof. We have that $\ker\phi$ and $\operatorname{img}\phi$ are B-submodules of V_1 and V_2, respectively, so either $\phi = 0$ or $\ker\phi = 0$ and $\operatorname{img}\phi = V_2$, hence $V_1 \cong V_2$. □

Corollary 6.2.13. *If V is a simple B-module, then* $\mathrm{End}_B(V)$ *is a division ring.*

6.2.14. Let B be an F-algebra and consider B as a left B-module. Then we have a map

$$\begin{aligned} \rho : B^{\mathrm{op}} &\xrightarrow{\sim} \mathrm{End}_B(B) \\ \alpha &\mapsto (\rho_\alpha : \beta \mapsto \beta\alpha), \end{aligned}$$

where B^{op} is the opposite algebra of B (Paragraph 3.2.2).

The map ρ is injective since $\rho_\alpha = 0$ implies $\rho_\alpha(1) = \alpha = 0$; it is also surjective, since if $\phi \in \operatorname{End}_B(B)$ then letting $\alpha = \phi(1)$ we have $\phi(\beta) = \beta\phi(1) = \beta\alpha$ for all $\beta \in B$. Finally, it is an F-algebra homomorphism, since

$$\rho_{\alpha\beta}(\mu) = \mu(\alpha\beta) = (\mu\alpha)\beta = (\rho_\beta \circ \rho_\alpha)(\mu),$$

and therefore ρ is an isomorphism of F-algebras.

One lesson here is that a left module has endomorphisms that act naturally on the right; but the more common convention is that endomorphisms also act on the left, so in order to make this compatible, the opposite algebra intervenes.

6.2.15. Many theorems of linear algebra hold equally well over division rings as they do over fields, as long as one is careful about the direction of scalar multiplication. For example, let D be a division F-algebra and let V be a left D-module. Then $V \cong D^n$ is free, and choice of basis for V gives an isomorphism $\operatorname{End}_D(V) \cong \mathrm{M}_n(D^{\mathrm{op}})$. When $n = 1$, this becomes $\operatorname{End}_D(D) \cong D^{\mathrm{op}}$, as in Paragraph 6.2.14.

Lemma 6.2.16. *Let B be a (finite-dimensional) simple F-algebra. Then there exists a simple left B-module which is unique up to isomorphism.*

Proof. Since B is finite-dimensional over F, there is a nonzero left ideal I of B of minimal dimension, and such an ideal I is necessarily simple. Moreover, if $v \in I$ is nonzero then $Bv = I$, since $Bv \subseteq I$ is nonzero and I is simple. Let $I = Bv$ with $v \in I$.

Now let V be any simple B-module; we will show $I \cong V$ as B-modules. Since B is simple, the natural map $B \to \operatorname{End}_F(V)$ is injective (since it is nonzero). Therefore, there exists $x \in V$ such that $vx \neq 0$, so $Ix \neq \{0\}$. Thus, the map $I \to V$ by $\beta \mapsto \beta x$ is a nonzero B-module homomorphism, so it is an isomorphism by Schur's lemma. □

Example 6.2.17. The unique simple left $\mathrm{M}_n(F)$-module (up to isomorphism) is the space F^n of column vectors (Example 6.2.1).

6.3 Semisimple modules and the Wedderburn–Artin theorem

We continue our assumptions that B is a finite-dimensional F-algebra and a B-module V is finite-dimensional.

Definition 6.3.1. A B-module V is *semisimple* (or *completely reducible*) if $V = \bigoplus_i V_i$ is a (finite) direct sum of simple B-modules V_i.

B is a *semisimple* F-algebra if B is semisimple as a left B-module.

Remark 6.3.2. More precisely, we have defined the notion of *left semisimple* and could equally well define *right semisimple*; below we will see that these two notions are the same.

Example 6.3.3. If $B = F$, then simple F-modules are one-dimensional vector spaces, and as F is simple these are the only ones. Every F-vector space has a basis and so is the direct sum of one-dimensional subspaces, so every F-module is semisimple.

Example 6.3.4. A finite-dimensional commutative F-algebra B is semisimple if and only if B is the product of field extensions of F, i.e., $B \cong K_1 \times \cdots \times K_r$ with $K_i \supseteq F$ a finite extension of fields.

Lemma 6.3.5. *The following statements hold.*

(a) *A B-module V is semisimple if and only if it is the sum of simple B-modules.*

(b) *A submodule or a quotient module of a semisimple B-module is semisimple.*

(c) *If B is a semisimple F-algebra, then every B-module is semisimple.*

Proof. For (a), let $V = \sum_i V_i$ be the sum of simple B-modules. Since V is finite-dimensional, we can rewrite it as an irredundant finite sum; and then since each V_i is simple, the intersection of any two distinct summands is $\{0\}$, so the sum is direct.

For (b), let $W \subseteq V$ be a submodule of the semisimple B-module V. Every $0 \neq x \in W$ is contained in a simple B-submodule of W by minimality, so $W = \sum_i W_i$ is a sum of simple B-modules. The result now follows from (a) for submodules. For quotient modules, suppose $\phi : V \to Z$ is a surjective B-module homomorphism; then $\phi^{-1}(Z) \subseteq V$ is a B-submodule, so $\phi^{-1}(Z) = \sum_i W_i$ is a sum of simple B-modules, and hence by Schur's lemma $Z = \sum_i \phi(W_i)$ is semisimple.

For (c), let V be a B-module. Since V is finitely generated as a B-module, there is a surjective B-module homomorphism $B^r \to V$ for some $r \geq 1$. Since B^r is semisimple, so too is V by (b). □

Lemma 6.3.6. *If B is a simple F-algebra, then B is a semisimple F-algebra.*

Proof. Let $I \subseteq B$ be a minimal nonzero left ideal, the unique simple left B-module up to isomorphism as in Lemma 6.2.16. For all $\alpha \in B$, the left ideal $I\alpha$ is a homomorphic image of I, so by Schur's lemma, either $I\alpha = \{0\}$ or $I\alpha$ is simple. Let $A = \sum_\alpha I\alpha$. Then A is a nonzero two-sided ideal of B, so since B is simple, we have $A = B$. Thus B is the sum of simple B-modules, so the result follows from Lemma 6.3.5(a). □

Corollary 6.3.7. *A (finite) direct product of simple F-algebras is a semisimple F-algebra.*

Proof. If $B \cong B_1 \times \cdots \times B_r$ with each B_i simple, then by Lemma 6.3.6, each B_i is is semisimple so $B_i = \bigoplus_j I_{ij}$ is the direct sum of simple B_i-modules I_{ij}. Each I_{ij} has the natural structure of a B-module (extending by zero), and with this structure it is simple, so $B = \bigoplus_{i,j} I_{ij}$ is semisimple. □

The converse of Corollary 6.3.7 is true and is proven as Corollary 6.3.12, a consequence of the Wedderburn–Artin theorem.

In analogy to Paragraph 6.2.15, we have the following corollary.

Corollary 6.3.8. *Let B be a simple F-algebra and let V be a left B-module. Then $V \cong I^{\oplus n}$ for some $n \geq 1$, where I is a simple left B-module. In particular, two left B-modules V_1, V_2 are isomorphic if and only if* $\dim_F V_1 = \dim_F V_2$.

Proof. Since B is simple, B is semisimple by Lemma 6.3.6, so V is semisimple by Lemma 6.3.5. But by Lemma 6.2.16, there is a unique simple left B-module I, and the result follows. □

In other words, this corollary says that if B is simple then every left B-module V has a left basis over B; if we define the *rank* of a left B-module V to be cardinality of this basis (the integer n such that $V \cong I^{\oplus n}$ as in Corollary 6.3.8), then two such modules are isomorphic if and only if they have the same rank.

We now come to the main result of this chapter.

Theorem 6.3.9 (Wedderburn–Artin)**.** *Let B be a finite-dimensional F-algebra. Then B is semisimple if and only if there exist integers $n_1, \dots, n_r$ and division algebras $D_1, \dots, D_r$ such that*

$$B \cong \mathrm{M}_{n_1}(D_1) \times \cdots \times \mathrm{M}_{n_r}(D_r).$$

In such a decomposition, the integers $n_1, \dots, n_r$ are unique up to permutation and once these integers are fixed, the division rings $D_1, \dots, D_r$ are unique up to isomorphism.

Proof. If $B \cong \prod_i M_{n_i}(D_i)$, then each factor $M_{n_i}(D_i)$ is a simple F-algebra by Example 6.2.9, so by Corollary 6.3.7, B is semisimple.

So suppose B is semisimple. Then we can write B as a left B-module as the direct sum $B \cong I_1^{\oplus n_1} \oplus \cdots \oplus I_r^{\oplus n_r}$ of simple B-modules $I_1, \dots, I_r$, grouped up to isomorphism. We have $\mathrm{End}_B(B) \cong B^{\mathrm{op}}$ by Paragraph 6.2.14. By Schur's lemma, we have

$$\mathrm{End}_B(B) \cong \bigoplus_i \mathrm{End}_B\left(I_i^{\oplus n_i}\right);$$

by Paragraph 6.2.15, we have

$$\mathrm{End}_B\left(I_i^{\oplus n_i}\right) \cong \mathrm{M}_{n_i}(D_i)$$

where $D_i = \operatorname{End}_B(I_i)$ is a division ring. So

$$B \cong \operatorname{End}_B(B)^{\mathrm{op}} \cong \mathrm{M}_{n_1}(D_1^{\mathrm{op}}) \times \cdots \times \mathrm{M}_{n_r}(D_r^{\mathrm{op}}).$$

The statements about uniqueness are then clear. □

Corollary 6.3.10. *Let B be a simple F-algebra. Then $B \cong \mathrm{M}_n(D)$ for a uniquely determined integer $n \in \mathbb{Z}_{\geq 1}$ and division algebra D, unique up to isomorphism.*

Example 6.3.11. Let B be a division F-algebra. Then $V = B$ is a simple B-module, and in Corollary 6.3.10 we have $D = \operatorname{End}_B(B) = B^{\mathrm{op}}$, and the Wedderburn–Artin isomorphism is just $B \cong \mathrm{M}_1((B^{\mathrm{op}})^{\mathrm{op}})$.

Corollary 6.3.12. *An F-algebra B is semisimple if and only if B is the direct product of simple F-algebras.*

Proof. Immediate from the Wedderburn–Artin theorem, as each factor $\mathrm{M}_{n_i}(D_i)$ is simple. □

6.4 Central simple algebras

Definition 6.4.1. An F-algebra B is *central* if the center of B is equal to F, i.e., $Z(B) = \{\alpha \in B : \alpha\beta = \beta\alpha \text{ for all } \alpha \in B\} = F$.

Remark 6.4.2. If the F-algebra B has center $Z(B) = K$, then B is a K-algebra—this is true even if K is not a field, but we have considered so far only algebras over fields. Aside from this caveat, every algebra is a central algebra over its center.

Example 6.4.3. By Corollary 6.3.10, the center $Z(B)$ of a simple F-algebra B is a field, since $Z(\mathrm{M}_n(D)) = Z(D)$ (Exercise 6.4).

The category of central simple algebras is closed under tensor product, as follows.

Proposition 6.4.4. *Let A, B be F-algebras and suppose that B is central.*

(a) *The center of $A \otimes_F B$ is $Z(A) \subseteq A \otimes_F B$.*

(b) *Suppose that A, B are simple. Then $A \otimes_F B$ is simple.*

Proof. First, centrality in part (a). Suppose that $\gamma = \sum_i \alpha_i \otimes \beta_i \in Z(A \otimes B)$. Without loss of generality, we may assume that α_i are linearly independent over F. Then by properties of tensor products, the elements $\beta_i \in B$ are unique. But then for all $\beta \in B$ we have

$$\sum_i(\alpha_i \otimes \beta\beta_i) = (1 \otimes \beta)\left(\sum_i \alpha_i \otimes \beta_i\right) = \left(\sum_i \alpha_i \otimes \beta_i\right)(1 \otimes \beta) = \sum_i(\alpha_i \otimes \beta_i\beta)$$

so $\beta\beta_i = \beta_i\beta$ for each i; thus $\beta_i = b_i \in Z(B) = F$. Hence

$$\gamma = \sum_i \alpha_i \otimes b_i = \sum_i \alpha_i b_i \otimes 1 = \left(\sum_i \alpha_i b_i\right) \otimes 1;$$

since $\alpha \otimes 1$ also commutes with γ for all $\alpha \in B$, we have $\sum_i \alpha_i b_i \in Z(A)$. Thus $\gamma \in Z(A) \otimes F = Z(A)$.

Next, simplicity in part (b). Let I be a nontrivial two-sided ideal in $A \otimes B$, and let $\gamma = \sum_{i=1}^m \alpha_i \otimes \beta_i \in I \setminus \{0\}$. Without loss of generality, we may assume $\beta_1 \neq 0$. Then since B is simple, we have $B\beta_1 B = B$, so multiplying on the left and right by elements of $B \subseteq A \otimes B$, we may assume further that $\beta_1 = 1$. Let $\gamma \in I \setminus \{0\}$ be such an element that is minimal with respect to m; then in particular the elements β_i are linearly independent over F. Now for each $\beta \in B$, we have

$$(1 \otimes \beta)\gamma - \gamma(1 \otimes \beta) = \sum_{i=2}^m (\alpha_i \otimes (\beta\beta_i - \beta_i\beta)) \in I;$$

but by minimality of m, the right-hand side is zero, so $\beta\beta_i = \beta_i\beta$ for all i. Hence $\beta_i \in Z(B) = F$ for all i and as above $\gamma = \alpha \otimes 1$ for some $0 \neq \alpha \in A$. But then

$$I \supseteq (A \otimes 1)(\alpha \otimes 1)(A \otimes 1) = (A\alpha A) \otimes 1 = A \otimes 1$$

since A is simple, so $I \supseteq (A \otimes 1)(1 \otimes B) = A \otimes B$, and thus $I = A \otimes B$ and $A \otimes B$ is simple. □

Lemma 6.4.5. *If B is a central simple algebra, then so too is B^{op}, and $B \otimes_F B^{\mathrm{op}} \cong \mathrm{End}_F(B)$.*

Proof. Define $\phi : B \otimes_F B^{\mathrm{op}} \to \mathrm{End}_F(B)$ by $\phi(\mu) = \alpha\mu\beta$ for $\alpha, \beta, \mu \in B$. Then just as in Paragraph 6.2.14, ϕ is a nonzero F-algebra homomorphism. By Proposition 6.4.4, $B \otimes_F B^{\mathrm{op}}$ is simple, so ϕ is injective. Since $\dim_F(B \otimes_F B^{\mathrm{op}}) = \dim_F \mathrm{End}_F(B) = (\dim_F B)^2$, ϕ is an isomorphism. □

6.5 Quaternion algebras

Having the set the stage, we are now ready to prove the following final characterizations of quaternion algebras.

Proposition 6.5.1. *Let B be an F-algebra. Then the following are equivalent.*

(i) *B is a quaternion algebra;*

(ii) *B is a central simple F-algebra with $\dim_F B = 4$;*

(iii) *B is a central semisimple F-algebra with $\dim_F B = 4$; and*

(iv) *$B \otimes_F \overline{F} \cong \mathrm{M}_2(\overline{F})$, where $\overline{F}$ is an algebraic closure of F.*

Proof. First, (i) ⇒ (ii): if B is a quaternion algebra, then B is central simple (Paragraph 6.2.11).

The equivalence (ii) ⇔ (iii) follows from the Wedderburn–Artin theorem: we have

$$1 = \dim Z(B) = \sum_{i=1}^{r} \dim_F Z(D_i) \geq r$$

so $r = 1$.

Next we prove (ii) ⇒ (iv). If B is central simple, then $B \otimes_F \overline{F}$ is a central simple $\overline{F}$-algebra by Proposition 6.4.4. But by Exercise 2.8, the only division $\overline{F}$-algebra is $\overline{F}$, so by the Wedderburn–Artin theorem we have $B \otimes_F \overline{F} \cong \mathrm{M}_n(\overline{F})$; by dimensions, we have $n = 2$.

It remains to prove (iv) ⇒ (i). So suppose $B \otimes_F \overline{F} \cong \mathrm{M}_2(\overline{F})$. Then B is simple by Example 6.2.10 and $\dim_F B = 4$. By the Wedderburn–Artin theorem (Corollary 6.3.10), we have $B \cong \mathrm{M}_n(D)$ with $n \in \mathbb{Z}_{\geq 1}$ and D a division ring. So $4 = \dim_F B = n^2 \dim_F D$ so either $n = 2$ so $B \cong \mathrm{M}_2(F)$ and we are done, or $n = 1$ and B is a division ring.

In this latter case, the result will follow from Theorem 3.5.1 (and Theorem 5.2.7 for the case char $F = 2$) if we show that B has degree 2. But for any $\alpha \in B$ we have that $\alpha \in B \otimes_F \overline{F} \cong \mathrm{M}_2(\overline{F})$ satisfies its characteristic polynomial of degree 2, so that $1, \alpha, \alpha^2$ are linearly dependent over $\overline{F}$ and hence linearly dependent over F, by linear algebra. □

Inspired by the proof of this result, we reconsider and reprove the splitting criterion considered in the previous section.

Proposition 6.5.2. *Let B be a quaternion algebra over F. Then the following are equivalent:*

(i) *$B \cong \mathrm{M}_2(F)$;*

(ii) *B is not a division ring;*

(iii) *There exists $0 \neq \epsilon \in B$ such that $\epsilon^2 = 0$;*

(iv) *B has a nontrivial left ideal $I \subsetneq B$;*

Proof. The equivalence (i) $\Leftrightarrow$ (ii) follows from the Wedderburn–Artin theorem (also proved in Theorems 4.5.5 and 5.3.8). The implications (i) $\Rightarrow$ (iii) $\Rightarrow$ (ii) and (i) $\Rightarrow$ (iv) $\Rightarrow$ (ii) are clear. □

6.5.3. We showed in Lemma 6.2.16 that a simple algebra B has a unique simple left B-module I up to isomorphism, obtained as a minimal nonzero left ideal. If B is a quaternion algebra, this simple module I can be readily identified using the above proposition. If B is a division ring, then necessarily $I = B$. Otherwise, we have $B \cong \mathrm{M}_2(F)$, and then $\dim_F I = 2$; indeed, the map $B \to \mathrm{End}_F(I)$ given by left multiplication is then an isomorphism.

6.6 Skolem–Noether

We conclude this chapter with a fundamental result that characterizes the automorphisms of a simple algebra—and much more.

Theorem 6.6.1 (Skolem–Noether)**.** *Let A, B be simple F-algebras and suppose that B is central. Suppose that $f, g : A \to B$ are homomorphisms. Then there exists $\beta \in B$ such that $f(\alpha) = \beta^{-1} g(\alpha) \beta$ for all $\alpha \in A$.*

Proof. By Corollary 6.3.10, we have $B \cong \mathrm{End}_D(V) \cong \mathrm{M}_n(D^{\mathrm{op}})$ where V is a simple B-module and $D = \mathrm{End}_B(V)$ is a central F-algebra. Now the maps f, g give V the structure of an A-module in two ways. The A-module structure commutes with the D-module structure since $B \cong \mathrm{End}_D(V)$. So V has two $A \otimes_F D$-module structures via f and g.

By Proposition 6.4.4, since D is central over F, we have that $A \otimes_F D$ is a simple F-algebra. By Corollary 6.3.8 and a dimension count, the two $A \otimes_F D$-module structures on V are isomorphic. Thus, there exists an isomorphism $\beta : V \to V$ of $A \otimes_F D$-modules; i.e. we have $\beta(f(\alpha)x) = g(\alpha)\beta(x)$ for all $\alpha \in A$ and $x \in V$, and $\beta(\delta x) = \delta\beta(x)$ for all $\delta \in D$ and $x \in V$. We have $\beta \in \mathrm{End}_D(V) \cong B$ and so we can write $\beta f(\alpha) \beta^{-1} = g(\alpha)$ for all $\alpha \in A$, as claimed. □

Corollary 6.6.2. *If A_1, A_2 are simple F-subalgebras of a central simple F-algebra B and $\phi : A_1 \xrightarrow{\sim} A_2$ is an isomorphism of F-algebras, then ϕ extends to an inner automorphism of B.*

Corollary 6.6.3. *The group of F-algebra automorphisms of a central simple algebra B is* $\operatorname{Aut}_F(B) \cong B^\times/F^\times$.

As a consequence, for example, we get that $\operatorname{Aut}_F(\mathrm{M}_n(F)) = \mathrm{PGL}_n(F)$.

Definition 6.6.4. Let A be an F-subalgebra of B. Let

$$C(A) = C_B(A) = \{\beta \in B : \alpha\beta = \beta\alpha \text{ for all } \alpha \in A\}$$

be the *centralizer* of A (in B).

Note that $C(A)$ is an F-subalgebra of B.

Lemma 6.6.5. *Let B be a central simple F-algebra and let $A \subseteq B$ a simple F-subalgebra. Then the following statements hold:*

(a) *$C_B(A)$ is a simple F-algebra.*

(b) $\dim_F B = \dim_F A \cdot \dim_F C_B(A)$.

(c) $C_B(C_B(A)) = A$.

Part (c) of this lemma is called the *double centralizer property*.

Proof. First, part (a). We interpret the centralizer as arising from certain kinds of endomorphisms. We have that B is a left $A\otimes B^{\mathrm{op}}$ module by the action $(\alpha\otimes\beta)\cdot\mu = \alpha\mu\beta$ for $\alpha \otimes \beta \in A \otimes B^{\mathrm{op}}$ and $\mu \in B$. We claim that

$$C_B(A) = \operatorname{End}_{A\otimes B^{\mathrm{op}}}(B). \tag{6.6.6}$$

Indeed, any $\phi \in \operatorname{End}_{A\otimes B^{\mathrm{op}}}(B)$ is left multiplication by an element of B: if $\gamma = \phi(1)$, then $\phi(\mu) = \phi(1)\mu = \gamma\mu$ by $1 \otimes B^{\mathrm{op}}$-linearity. Now the equality

$$\gamma\alpha = \phi(\alpha) = \alpha\phi(1) = \alpha\gamma$$

shows that multiplication by γ is $A \otimes 1$-linear if and only if $\gamma \in C_B(A)$, proving (6.6.6).

By Proposition 6.4.4, the algebra $A \otimes B^{\mathrm{op}}$ is simple. By the Wedderburn–Artin theorem, we have $A \otimes B^{\mathrm{op}} \cong \mathrm{M}_n(D)$ for some $n \geq 1$ and division F-algebra D. Since $\mathrm{M}_n(D)$ is simple, its unique left D-module is $V = D^n$, and $\operatorname{End}_{\mathrm{M}_n(D)}(V) \cong D^{\mathrm{op}}$. In particular, we have $B \cong V^r$ for some $r \geq 1$ as an $A \otimes B^{\mathrm{op}}$-module. So

$$C_B(A) = \operatorname{End}_{A\otimes B^{\mathrm{op}}}(B) \cong \operatorname{End}_{\mathrm{M}_n(D)}(V^r) \cong \mathrm{M}_r(\operatorname{End}_{\mathrm{M}_n(D)}(V)) \cong \mathrm{M}_r(D^{\mathrm{op}}).$$

Thus $C_B(A)$ is simple.

For part (b), we have

$$\dim_F C_B(A) = \dim_F \mathrm{M}_r(D^{\mathrm{op}}) = r^2 \dim_F D$$

and

$$\dim_F(A \otimes B^{\mathrm{op}}) = \dim_F A \cdot \dim_F B = n^2 \dim_F D$$

and

$$\dim_F B = \dim_F V^r = r \dim_F D^n = rn \dim_F D$$

so $\dim_F A \cdot \dim_F C_B(A) = rn \dim_F D = \dim_F B$.

Finally, part (c) follows from the fact (a) and (b), giving

$$\dim_F B = \dim_F C_B(A) \cdot \dim_F C_B(C_B(A)) = \dim_F A \cdot \dim_F C_B(A)$$

so $\dim_F A = \dim_F C_B(C_B(A))$ and $A \subseteq C_B(C_B(A))$ so equality holds. □

Example 6.6.7. We always have the two extremes $A = F$ and $A = B$, with $C_B(F) = B$ and $C_B(B) = F$, accordingly.

Corollary 6.6.8. *Let B be a central simple F-algebra and let K be a maximal subfield. Then $[B : F] = [K : F]^2$.*

Proof. If K is maximal, then $C_B(K) = K$, so $[B : F] = [K : F]^2$. □

6.7 Reduced trace and norm

In this last section, we consider notions of reduced trace and reduced norm in the context of semisimple algebras.

6.7.1. Let B be a simple algebra over F, and let F^{sep} denote a separable closure of F. By Exercise 6.12, we have an F-algebra homomorphism

$$\iota : B \hookrightarrow B \otimes_F F^{\mathrm{sep}} \cong \mathrm{M}_n(F^{\mathrm{sep}})$$

for some $n \geq 1$. By the Skolem-Noether theorem (Theorem 6.6.1), if ι' is another such homomorphism, then there exists $M \in \mathrm{GL}_n(F^{\mathrm{sep}})$ such that $\iota'(\alpha) = M\iota(\alpha)M^{-1}$, so the characteristic polynomial of $\iota(\alpha)$ is independent of the choice of ι. We define the *reduced characteristic polynomial* of $\alpha \in B$ to be the characteristic polynomial of $\iota(\alpha)$ as an element of $F^{\mathrm{sep}}[T]$ and similarly the *reduced trace* and *reduced norm* of α to be the trace and determinant of $\iota(\alpha)$ as elements of F^{sep}.

In fact, the reduced characteristic polynomial descends to F, as follows. The absolute Galois group $G_F = \mathrm{Gal}(F^{\mathrm{sep}}/F)$ acts on $B \otimes_F F^{\mathrm{sep}} \cong \mathrm{M}_n(F^{\mathrm{sep}})$ by

$$\sigma(\alpha \otimes a) = \alpha \otimes \sigma(a)$$

for $\sigma \in G_F$, $\alpha \in B$, and $a \in F^{\mathrm{sep}}$. For $\sigma \in G_F$, we define ${}^{\sigma}\iota : B \hookrightarrow B \otimes_F F^{\mathrm{sep}}$ by $({}^{\sigma}\iota)(\alpha) = \sigma(\iota(\alpha))$. Just as ι is an F-algebra homomorphism, so too is ${}^{\sigma}\iota$, exactly because $\sigma \in G_F$ fixes F. By the preceding paragraph, therefore, the characteristic polynomial of $\iota(\alpha)$ and $({}^{\sigma}\iota)(\alpha) = \sigma(\iota(\alpha))$ are the same. And if

$$f(\alpha; T) = \det(T - \iota(\alpha)) = T^n + a_{n-1}T^{n-1} + \cdots + a_0$$

is the reduced characteristic polynomial of $\iota(\alpha)$, then the reduced characteristic polynomial of $({}^{\sigma}\iota)(\alpha)$ is

$${}^{\sigma}f(\alpha; T) = \det(T - \sigma(\iota(\alpha))) = T^n + \sigma(a_{n-1})T^{n-1} + \cdots + \sigma(a_0).$$

And then since $f(\alpha; T) = {}^{\sigma}f(\alpha; T)$ for all $\sigma \in G_F$, by Galois theory, we have $f(\alpha; T) \in F[T]$. Therefore, the reduced norm and reduced trace also belong to F.

6.7.2. We extend this definition to semisimple algebras B by writing $B \cong B_1 \times \cdots \times B_r$ where each B_i is simple, and defining the reduced characteristic polynomial to be the product of the reduced characteristic polynomials on each simple direct factor B_i. This is well-defined by the uniqueness statement in the Wedderburn–Artin theorem (Theorem 6.3.9).

More conceptually, we can reinterpret this as follows. Let V_i be the unique simple left $B_i \otimes_F F^{\mathrm{sep}}$-module and let $V = \bigoplus_i V_i$; then V is the unique minimal faithful (semisimple) left $B \otimes_F F^{\mathrm{sep}}$-module, up to isomorphism. We have a map $B \hookrightarrow \mathrm{End}_{F^{\mathrm{sep}}}(V)$, so we may accordingly define the reduced characteristic polynomial in this way, and argue as in Paragraph 6.7.1.

6.8 Separable algebras

For a (finite-dimensional) F-algebra, the notions of simple and semisimple are sensitive to the base field F in the sense that these properties need not hold after extending the base field. Indeed, let $K \supseteq F$ be a finite extension of fields, so K is a simple F-algebra. Then $K \otimes_F \overline{F}$ is simple only when $K = F$ and is semisimple if and only if $K \otimes_F \overline{F} \cong \overline{F} \times \cdots \times \overline{F}$, i.e., K is separable over F.

It is important to have a notion which is stable under base change, as follows.

Definition 6.8.1. Let B be a finite-dimensional F-algebra. We say that B is a *separable* F-algebra if B is semisimple and $Z(B)$ is the product of separable field extensions of F.

By the Wedderburn–Artin theorem, for a semisimple algebra B we have

$$B \cong \mathrm{M}_{n_1}(D_1) \times \cdots \times \mathrm{M}_{n_r}(D_r),$$

so by Example 6.4.3 we have

$$Z(B) \cong Z(D_1) \times \cdots \times Z(D_r),$$

and B is separable if and only if $Z(D_i)$ is separable for each $i = 1, \dots, r$. Like being central, the notion of separability depends on the base field F.

Lemma 6.8.2. *A finite-dimensional simple F-algebra is a separable algebra over its center K.*

Proof. The center of B is a field $K = Z(B)$ and as a K-algebra, the center $Z(B) = K$ is certainly separable over K. (Or use Proposition 6.4.4 and (iii) below.) □

The notion of separability in this context is incredibly robust and useful.

Theorem 6.8.3. *Let B be a finite-dimensional F-algebra. Then the following are equivalent:*

(i) *B is separable;*

(ii) *There exists a finite separable field extension K of F such that $B \otimes_F K \cong M_{n_1}(K) \times \cdots \times M_{n_r}(K)$ for integers $n_1, \dots, n_r \geq 1$;*

(iii) *For every extension $K \supseteq F$ of fields, the K-algebra $B \otimes_F K$ is semisimple; and*

(iv) *The reduced trace gives rise to a nondegenerate pairing*

$$\begin{aligned} B \times B &\to F \\ (\alpha, \beta) &\mapsto \operatorname{trd}(\alpha\beta). \end{aligned}$$

In particular by (iii), B is a separable F-algebra if and only if $B \otimes_F K$ is a separable K-algebra for all field extensions $K \supseteq F$.

For this reason, a separable F-algebra is sometimes called *absolutely semisimple*. For more, see Reiner [Rei03, Section 7c] or Pierce [Pie82, Chapter 10].

6.9 Extensions and further reading

6.9.1. Basic references for this section include Lam [Lam01, §2–3] and Farb–Dennis [FD93, Part I]. An elementary approach to the Weddernburn–Artin theorem is given by Brešar [Bre10].

6.9.2. Theorem 6.3.9 as it is stated was originally proven by Wedderburn [Wed08], and so is sometimes called *Wedderburn's theorem*. However, this term may also apply to the theorem of Wedderburn that a finite division ring is a field; and Artin generalized Theorem 6.3.9 to rings where the ascending and descending chain condition holds for left ideals [Art26], so we follow the common convention by referring to Theorem 6.3.9 as the Wedderburn–Artin theorem.

6.9.3. Doing linear algebra with semisimple modules mirrors very closely linear algebra over a field. We have already seen that every submodule and quotient module of a semisimple module is again semisimple. Moreover, every module homomorphism $V \to W$ with V semisimple splits, and every submodule of a semisimple module is a direct summand. In particular, a semisimple module is *projective*, and every module over a semisimple algebra is projective.

We have barely scratched the surface of this theory and we refer to [[cites]] for further reference.

6.9.4. Among central simple algebras over a field, quaternion algebras have an especially nice presentation because of the quadratic norm form can be diagonalized (normalized, in characteristic 2). More generally, one may look at algebras with a similarly nice presentation, as follows. [[Cyclic algebras]]

Exercises

6.1. Prove that a quaternion algebra $B = \left(\dfrac{a,b}{F}\right)$ with $\operatorname{char} F \neq 2$ is simple, as follows.

a) Let I be a nontrivial two-sided ideal. Show that if $\epsilon \in I$, then $\epsilon^2 = 0$.

b) Show that there exists $\alpha \in B$ such that $\operatorname{trd}(\alpha) = 0$ and $\operatorname{trd}(\epsilon\alpha) \neq 0$.

c) Considering $\epsilon\alpha + \alpha\epsilon$, derive a contradiction (cf. (4.2.7)).

Modify this argument to show that an algebra $B = \left[\dfrac{a,b}{F}\right)$ is simple when $\operatorname{char} F = 2$.

6.2. Let B be a quaternion algebra. Exhibit an explicit isomorphism $B \otimes_F B \xrightarrow{\sim} \mathrm{M}_4(F)$ (see Exercise 2.9).

6.3. Let $B = \left\{\begin{pmatrix} a & b \\ 0 & c \end{pmatrix} : a, b, c \in F\right\} \subseteq \mathrm{M}_2(F)$, and $V = F^2$ be the left B-module of column vectors. Show that B is indecomposable, but not simple (cf. Example 6.2.5).

6.4. a) Let D be a division F-algebra. Prove that $\mathrm{M}_n(D)$ is a simple F-algebra with center $Z(D)$ for all $n \geq 1$. *[Hint: Let E_{ij} be the matrix with* 1 *in the* ij*th entry and zeros in all other entries. Show that* $E_{ki} M E_{j\ell} = m_{ij} E_{k\ell}$ *where* m_{ij} *is the* ij*th entry of* M*.]*

b) More generally, let R be a ring (associative with 1, but potentially noncommutative). Show that $Z(\mathrm{M}_n(R)) = Z(R)$ and that any two-sided ideal of $\mathrm{M}_n(R)$ is of the form $\mathrm{M}_n(I) \subseteq \mathrm{M}_n(R)$ where I is a two-sided ideal of R.

6.5. Generalize the statement of Proposition 6.4.4(a) as follows. Let A, B be F-algebras, and let $A' \subseteq A$ and $B' \subseteq B$ be F-subalgebras. Prove that $C_{A\otimes B}(A' \otimes B') = C_A(A') \otimes C_B(B')$.

6.6. Let $G \neq \{1\}$ be a finite group. Show that the *augmentation ideal*, the two-sided ideal generated by $g - 1$ for $g \in G$, is a nontrivial ideal, and hence $F[G]$ is not simple as an F-algebra. (However, $F[G]$ is semisimple if and only if char F is coprime to $\#G$: this is Maschke's theorem [Lam01, Theorem 6.1].)

6.7. Let B be an F-algebra, and let $\overline{F}$ be an algebraic closure of F. Show that B is simple if and only if $B \otimes_F \overline{F}$ is simple.

6.8. Let D be a (finite-dimensional) division algebra over $\overline{F}$. Show that $D = \overline{F}$. Conclude that if B is a simple algebra over $\overline{F}$, then $B \cong \mathrm{M}_n(\overline{F})$ for some $n \geq 1$ and hence is central.

6.9. Show that if B is a semisimple F-algebra, then so is $\mathrm{M}_n(B)$ for any $n \in \mathbb{Z}_{\geq 1}$.

6.10. A *nil ideal* of a ring B is a (two-sided) ideal $I \subseteq B$ such that $I^n = (0)$ for some $n \in \mathbb{Z}_{>0}$.

Let B be a central (finite-dimensional) F-algebra with standard involution with $B \neq F$. Show that B is a quaternion algebra if and only if the largest nil ideal of B is (0). *[Hint: Suppose* $0 \neq e \in B$ *satisfies* $e^2 = 0$*. Show that* e *generates a nil ideal if and only if* $\operatorname{trd}(ex) = 0$ *for all* $x \in B$*.]*

[[Relate nil ideal to semisimple.]]

6.11. Give an example of (finite-dimensional) simple algebras A, B over a field F such that $A \otimes_F B$ is not simple. Same with not semisimple.

6.12. In Exercise 6.8, we saw that if D is a (finite-dimensional) division algebra over F then $D \otimes_F \overline{F} \cong \mathrm{M}_n(\overline{F})$ for some $n \geq 1$. In this exercise, we show the same is true if we consider the separable closure.

Let D be a finite-dimensional central division algebra over a *separably closed* field F, i.e. F contains a root of all separable polynomials with coefficients in F. Suppose $\operatorname{char} F = p$.

a) Prove that $\dim_F D$ is divisible by p.

b) Show that the minimal polynomial of each nonzero $d \in D$ has the form $x^{p^e} - a$ for some $a \in F$.

c) Choose an isomorphism $\phi : D \otimes_F \overline{F} \to \mathrm{M}_n(\overline{F})$. Show that the trace of $\phi(x \otimes 1) = 0$ for all $x \in D$.

d) Prove that D does not exist.

6.13. Use the Skolem–Noether theorem to give another proof that if $K \subset B$ is a separable quadratic F-algebra then $B \cong \left(\dfrac{K,b}{F}\right)$ for some $b \in F^\times$.

6.14. Give an alternate direct proof of Corollary 6.6.3. *[Hint: Use the fact that there is a unique simple left B-module.]*

6.15. Let $B = \left(\dfrac{a,b}{F}\right)$ be a quaternion algebra, and let $K = F[i]$. Show that the subgroup of $\operatorname{Aut}(B)$ that fixes $K \subseteq B$ is isomorphic to the group

$$K^\times/F^\times \cup j(K^\times/F^\times).$$

Show that the subgroup of $\operatorname{Aut}(B)$ that restricts to the identity on K is isomorphic to $K^\times/F^\times$.

6.16. Use the Skolem–Noether theorem and the fact that a finite group cannot be written as the union of the conjugates of a proper subgroup to prove Wedderburn's theorem: a finite division ring is a field.

6.17. Show that every *ring* automorphism of $\mathbb{H}$ is inner. (Compare this with automorphisms of $\mathbb{C}$!)

Chapter 7

Simple algebras and involutions

7.1 The Brauer group and involutions

Cyclic algebras. But are all central division algebras cyclic? No, and to get an example, it is enough to take a tensor product of quaternion algebras: these are biquaternion algebras, and the theory of Albert.

More generally, the set of isomorphism classes of central simple algebras is closed under tensor product; if we think that the matrix ring is something that is no more complicated than the ring that it is over, it is natural to introduce an equivalence relation on central simple algebras that identifies a division ring with the matrix ring (of any rank) over this division ring. This set of such equivalence classes then becomes a group under tensor product, known as the Brauer group. Large subject; over fields, the tensor products of quaternion algebras are elements of order at most 2 in the Brauer group.

An important application of the theory of semisimple algebras is in the structure of endomorphism algebras of abelian varieties. For example, supersingular elliptic curves. Instead of a standard involution, we consider positive involutions; the Rosati involution is a positive involution on the endomorphism algebra of an abelian variety, and it is a consequence that this algebra (over $\mathbb{Q}$) is semisimple! Weil classified simple algebras with positive involution, and these come very close to the theory of quaternion algebras.

We've seen quaternion algebras characterized in several different ways: possessing a standard involution is a strong condition indeed. In many contexts, we have a simple algebra with an involution of another kind; we see the consequences of that in this section.

First, we consider tensor products of quaternion algebras: the standard involution on each factor gives an involution, and the decomposition of this according to the

Wedderburn–Artin theorem says something about quadratic forms.

Once we see that tensor products are interesting to study, we make a group out of (classes of) central simple algebras: this is the Brauer group, and the quaternion algebras generate the 2-torsion.

Finally, we consider positive involutions, which arise naturally in the study of endomorphism algebras of abelian varieties. It turns out that quaternion algebras again figure prominently.

These sections are not essential for the rest of the text.

7.2 Biquaternion algebras

Something about the Brauer group here, and quaternion algebras are elements of order 2.

Work of Albert [Alb72] and the Albert form.

Let A, B be quaternion algebras over F. The tensor product $A \otimes B$ is a central simple algebra over F of dimension 16 called a *biquaternion algebra*. By the Wedderburn–Artin theorem, we have one of the following possibilities for this algebra:

(i) $A \otimes B$ is a division algebra;

(ii) $A \otimes B \cong \mathrm{M}_4(F)$;

(iii) $A \cong \mathrm{M}_2(D)$ where D is a quaternion division algebra over F.

We may prefer to combine the last two and just say that $A \otimes B \cong \mathrm{M}_2(C)$ where C is a quaternion algebra over F, since $\mathrm{M}_2(\mathrm{M}_2(F)) \cong \mathrm{M}_4(F)$.

There is an algebraic way to see what happens.

Proposition 7.2.1. *The following are equivalent:*

(i) *$A \otimes_F B$ is not a division algebra;*

(ii) *A, B have a common quadratic splitting field;*

(iii) *There exists a quadratic field extension K/F that can be F-embedded in both A and B.*

Proof. Lam gives the proof in his textbook. Characteristic 2 argument by Lam [Lam02], so you can add "separable". □

There is also a quadratic forms approach. We define the *Albert form* of the biquaternion algebra $A \otimes B$ is $Q\,\mathrm{nrd}\,|_{A^0} \perp \mathrm{nrd}\,|_{B^0}$. If $\mathrm{char}\, F \neq 2$ and $A \cong \left(\dfrac{a,b}{F}\right)$ and $B \cong \left(\dfrac{c,d}{F}\right)$ then the Albert form is $Q = \langle -a, -b, ab, -c, -d, cd\rangle$.

Proposition 7.2.2. *Let $A \otimes B$ be a biquaternion algebra with Albert form Q. Then*

$$A \otimes B \text{ is not a division algebra} \quad \Leftrightarrow \quad Q \text{ is isotropic}$$

and

$$A \otimes B \cong \mathrm{M}_4(F) \quad \Leftrightarrow \quad Q \cong H \perp H \perp H$$

where H is a hyperbolic plane.

[[Is this still OK in characteristic 2?]]

For the fields of interest in this book (finite fields, local fields, global fields), a biquaternion algebra is never a division algebra. But biquaternion division algebras exist: let $F = \mathbb{C}(x, y, z, w)$ be a rational function field in 4 variables and take $A = \left(\dfrac{x,y}{F}\right)$ and $B = \left(\dfrac{z,w}{F}\right)$. (Exercise.)

Implications for testing if two quaternion algebras are isomorphic.

7.3 Brauer group

The set of division algebras over F is indeed interesting to study, encoding deep information about the field F. This set can be given the structure of a group as follows: if D, D' are division F-algebras, then $D \otimes_F D' \cong \mathrm{M}_n(D'')$ for a unique division F-algebra D'' up to F-algebra isomorphism, so we could try to define $D \oplus D' = D''$, by taking the "division ring" part.

To make this work, recall that every simple F-algebra is the matrix ring over a division ring, by the Wedderburn–Artin theorem (Theorem 6.3.9). So let $\mathrm{CSA}(F)$ be the set of isomorphism classes of central simple F-algebras. The binary operation of tensor product on $\mathrm{CSA}(F)$ is commutative and F is the identity, but for dimension reasons, only F has an inverse. So we define an equivalence relation $\sim$ on this set:

$$A \sim A' \text{ if } \mathrm{M}_{n'}(A) \cong \mathrm{M}_n(A') \text{ for some } n, n' \geq 1. \tag{7.3.1}$$

In particular, $A \sim \mathrm{M}_n(A)$ for all $A \in \mathrm{CSA}(F)$ as needed above.

Lemma 7.3.2. *The set of equivalence classes of central simple F-algebras under the equivalence relation $\sim$ has the structure of an abelian group.*

Proof. Tensor product is compatible with this definition (Exercise 7.2). If $\dim_F A = n$ and A^{op} is the opposite algebra of A (Paragraph 3.2.2) then the map

$$\begin{aligned} A \otimes_F A^{\mathrm{op}} &\to \operatorname{End}_F(A) \cong \mathrm{M}_n(F) \\ \alpha \otimes \beta &\mapsto (\mu \mapsto \alpha\mu\beta) \end{aligned}$$

is a nonzero homomorphism of F-algebras, so since $A \otimes_F A^{\mathrm{op}}$ is simple it is injective, and since $\dim_F A \otimes_F A^{\mathrm{op}} = n^2 = \dim_F \mathrm{M}_n(F)$ it is an isomorphism, and so $[A]^{-1} = [A^{\mathrm{op}}]$ provides an inverse to $[A]$. □

So we make the following definition.

Definition 7.3.3. The *Brauer group* of F is the set $\mathrm{Br}(F)$ of equivalence classes of central simple F-algebras under the equivalence relation $\sim$ defined in (7.3.1).

Let B be a division quaternion algebra over F. Then the standard involution gives an isomorphism $B \xrightarrow{\sim} B^{\mathrm{op}}$, and hence in $\mathrm{Br}(F)$ we have $[B]^{-1} = [B]$ and so $[B]$ is an element of order at most 2. It follows that biquaternion algebras, or more generally tensor products of quaternion algebras, are also elements of order at most 2 in $\mathrm{Br}(F)$.

7.4 Positive involutions

Our interest in involutions in Chapter 3 began with an observation of Hamilton: the product of a nonzero element with its involute in $\mathbb{H}$ is a positive real number (its norm, or square length). We then proved that such the existence of a such an involution characterizes quaternion algebras in an essential way. However, one may want to relax this setup and instead consider when the product of an nonzero element with its involute merely has positve *trace*.

Throughout this section, let B be a finite-dimensional $\mathbb{R}$-algebra. Recall that there is a trace map $\mathrm{Tr} : B \to \mathbb{R}$ given by the trace of right multiplication.

Definition 7.4.1. An involution $^* : B \to B$ is *positive* if $\mathrm{Tr}(\alpha\alpha^*) > 0$ for all $\alpha \in B \setminus \{0\}$.

Example 7.4.2. The standard involution on $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{H}$ are positive involutions, but the standard involution on $\mathrm{M}_2(\mathbb{R})$ is not, since if $\alpha \in \mathrm{M}_2(\mathbb{R})$ then $\mathrm{Tr}(\alpha\overline{\alpha}) = 4\det(\alpha)$ takes on all values in $\mathbb{R}$.

7.4.3. Let D be one of $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$. Let $B = \mathrm{M}_n(D)$. The standard involution $^-$ on D extends to an involution on B, acting on coordinates.

Then the conjugate transpose map $\alpha \mapsto \alpha^* = \overline{\alpha}^t$ for $\alpha \in B$ is a positive involution on B. Indeed, if $\alpha = (a_{ij})_{i,j=1,\dots,n}$ then

$$\mathrm{Tr}(\alpha\alpha^*) = n\,\mathrm{Tr}(\alpha\alpha^*) = n\|\alpha\|^2 = n\sum_{i,j=1}^{n} \mathrm{nrd}(a_{ij})$$

is the *Frobenius norm* on B. , e.g., if $D = \mathbb{R}$ then $\mathrm{Tr}(\alpha\alpha^*) = \sum_{i,j} a_{ij}^2 > 0$ for $\alpha \neq 0$.

We will soon see that essentially all positive involutions are given in Paragraph 7.4.3. First, we reduce to the case where B is a semisimple algebra.

Lemma 7.4.4. *Suppose that B admits a positive involution. Then B is semisimple.*

Proof. Let $\alpha \in \operatorname{rad} B$. Since $J = \operatorname{rad} B$ is nilpotent, there exists n such that $J^n \neq \{0\}$ but $J^{n+1} = \{0\}$. Let $\alpha \in J$ be such that $\alpha^n \neq 0$ but $\alpha^{n+1} = 0$. We have $J^* = J$, so $\alpha^n\alpha^* = 0$ so $\mathrm{Tr}(\alpha^n(\alpha^n)^*) = \mathrm{Tr}(0) = 0$, contradicting that * is positive. □

An

Theorem 7.4.5 (Weil)**.** *Let B be a simple $\mathbb{R}$-algebra with a positive involution † : $B \to B$. Then there exists an element $\mu \in B^\times$ with $\mu^2 \in \mathbb{R}_{>0}$, such that*

$$\alpha^\dagger = \mu^{-1}\alpha^*\mu$$

and * is the conjugate transpose involution.

The element μ in Theorem 7.4.5 is unique up to multiplication by F^*, by the Skolem–Noether theorem.

7.5 Endomorphism algebras of abelian varieties

We conclude this section with an application that may be skipped on a first reading: we characterize endomorphism algebras of (simple) abelian varieties in terms of algebras with involutions.

Let k be a field with algebraic closure $\overline{k}$. A *variety* over k is a geometrically integral separated scheme of finite type over k. An *abelian variety* is a proper group variety, i.e., a group in the category of varieties over k. An abelian variety is projective and commutative, and any abelian variety over $\mathbb{C}$ is isomorphic to $\mathbb{C}^g/\Lambda$ for some $g \geq 0$ and discrete subgroup $\Lambda \subset \mathbb{C}^g$ satisfying $\Lambda \cong \mathbb{Z}^{2g}$. An abelian variety A is *simple* if A has no abelian subvariety other than $\{0\}$ and A.

An *isogeny* is a surjective homomorphism $\alpha : A \to A'$ of abelian varieties such that $\dim A = \dim A'$ with finite kernel $\ker(\alpha) \subseteq A$.

Let A be an abelian variety over an algebraically closed field $k = \overline{k}$.

Lemma 7.5.1. *A is isogenous to the product*

$$A_1^{n_1} \times \cdots \times A_r^{n_r}$$

where $A_1, \ldots, A_r$ are simple, pairwise nonisogenous, abelian subvarieties of A and $n_1, \ldots, n_r \in \mathbb{Z}_{>0}$.

Let $\operatorname{End}(A)$ be the ring of endomorphisms of A. It follows from Lemma 7.5.1 that

$$\operatorname{End}(A) \otimes_{\mathbb{Z}} \mathbb{Q} \cong \prod_{i=1}^{r} \mathrm{M}_{n_i}(D_i)$$

where $D_i = \operatorname{End}(A_i) \otimes_{\mathbb{Z}} \mathbb{Q}$. So from now on, we may suppose that A is simple; then $D = \operatorname{End}(A) \otimes \mathbb{Q}$ is a division algebra, hence simple.

We will need just three properties of the endomorphism algebra D. The first two properties are basic.

Proposition 7.5.2. *$D = \operatorname{End}(A) \otimes \mathbb{Q}$ is a finite-dimensional division algebra over $\mathbb{Q}$ that admits an involution $^\dagger : D \to D$.*

The same method of proof as Theorem 3.5.1 applies in this case.

Lemma 7.5.3. *Let D be a division algebra over $\mathbb{Q}$ that admits an involution †. Suppose that the fixed subspace Then there is a unique subfield $F \subseteq Z(D)$ such that one of the three posibilities holds:*

(i) *$D = F$ is a field and † is the identity;*

(ii) *$D = K$ is a quadratic field extension of F and † is the standard involution; or*

(iii) *D is a division quaternion algebra over F.*

Proof. Let

$$D^{\langle\dagger\rangle} = \{\alpha \in D : \alpha^\dagger = \alpha\}$$

be the subspace of D where † acts by the identity, and let $F = D^{\langle\dagger\rangle} \cap Z(D)$. Then D is an F-algebra and † is an involution of D as an F-algebra.

Since † is an involution, we can diagonalize and write $D = D^+ \oplus D^-$ where † acts as the identity on D^+ and acts as -1 on D^-.

Diagonalizing, we can write $D = D^+ \oplus D^-$ where. Then $D^{\langle\dagger\rangle}$ is a subfield of D, since fixed subfield of †. Then $(\alpha\alpha^\dagger)^\dagger = (\alpha^\dagger)^\dagger\alpha^\dagger = \alpha\alpha^\dagger$ by the properties of an involution, so $\alpha\alpha^\dagger \in F$. □

Definition 7.5.4. An involution $^*: B \to B$ of a finite-dimensional $\mathbb{Q}$-algebra B is *positive* if the map

$$\begin{aligned} B &\to \mathbb{Q} \\ \alpha &\mapsto \operatorname{Tr}(\alpha\alpha^*) \end{aligned}$$

is a positive definite quadratic form on B, i.e., $\operatorname{Tr}(\alpha\alpha^*) > 0$ for all $\alpha \in B \setminus \{0\}$.

The involution $^\dagger : D \to D$ is called the *Rosati involution* , and it arises from the existence of the dual isogeny. (It depends on a choice of *polarization* , an isogeny $\phi : A \to A^\vee$.)

We now use the results of this chapter to explicitly characterize the algebras D that can occur as endomorphism algebras of simple abelian varieties over an algebraically closed field, using Proposition 7.5.2.

First, we relate the existence of a positive involution to a standard involution.

Proposition 7.5.5. *There exists $\mu \in D$*

Remark 7.5.6. The Rosati involution depends on a choice of polarization, and the implicit claim in Proposition 7.5.2 is that A (up to isogeny) has a polarization such that the Rosati involution is the standard involution.

Proposition 7.5.7. *We have one of four possibilities for D.*

Type I. $D = F$ is a totally real field, and the Rosati involution

Type II. D is a division quaternion algebra over a totally real field F that is totally indefinite

7.6 Extensions and further reading

7.6.1. Merkurjev in 1981 proved that any division algebra with an involution is Brauer equivalent to a tensor product of quaternion algebras, i.e., if D is a division F-algebra with (not necessarily standard) involution, then there exists $n \in \mathbb{Z}_{\geq 1}$ such that $\mathrm{M}_n(D)$ is isomorphic to a tensor product of quaternion algebras. His theorem, more properly, says that the natural map $K_2(F) \to \operatorname{Br}(F)[2]$ is an isomorphism.

7.6.2. The notion of positive involution was introduced by Weil.

Exercises

7.1. Let G be a finite group. Show that the map induced by $g \mapsto g^{-1}$ for $g \in G$ defines an positive involution on $\mathbb{R}[G]$. Similarly, show that this map composed with coordinatewise complex conjugation defines a positive involution on $\mathbb{C}[G]$ (as an $\mathbb{R}$-algebra).

7.2. Show that if $\sim$ is the equivalence relation (7.3.1) on $\mathrm{CSA}(F)$, then $\sim$ is compatible with tensor product, i.e., if $A, A' \in \mathrm{CSA}(F)$ and $A' \sim A'' \in \mathrm{CSA}(F)$ then $A \otimes A' \sim A \otimes A''$.

7.3. Let K/F be a finite extension of global fields. Show that the set of isomorphism classes of plane conics over F such that $X \times_F K \cong \mathbb{P}^1_K$ is infinite, and the same for $X \times_F K \ncong \mathbb{P}^1_K$. (Compare Exercise 4.16.)

[[Examples of domains arising as coordinate rings of varieties.]]

Chapter 8

Orders

8.1 Integral structures

Inside the rational numbers $\mathbb{Q}$ are the integers $\mathbb{Z}$; inside a number field is its ring of integers. What happens if we concern ourselves with a notion of integrality for possibly noncommutative algebras? In this chapter, we consider some basic questions of this nature that work without hypothesis on the field.

First we have to understand the linear algebra aspects: these are modules inside a vector space. Then the algebra structure is a multiplication law on this lattice, and is called an *order* because something.

Some properties of orders can be deduced from the commutative case: orders still consist of *integral* elements, satisfying a monic polynomial with coefficients in $\mathbb{Z}$.

The matrix ring over a field are endomorphisms of a vector space; the orders in a matrix ring should look like endomorphisms of a lattice (perhaps with extra structure).

Do some examples over $\mathbb{Z}$.

8.2 Lattices

Throughout this chapter, let R be a noetherian domain with field of fractions F. To avoid trivialities, we assume $R \neq F$.

Definition 8.2.1. Let V be a finite-dimensional F-vector space. An *R-lattice* of V is a finitely generated R-submodule $M \subseteq V$ with $MF = V$.

Remark 8.2.2. Other authors omit the second condition in the definition of an R-lattice and say that I is *full* if $MF = V$. We will not encounter R-lattices that are not full (and when we do, we call them finitely generated R-submodules), so we avoid this added nomenclature.

By definition, an R-lattice contains a basis of V, and it can be thought of an R-submodule that "allows bounded denominators", as follows.

Lemma 8.2.3. *Let M be an R-lattice. Then for any $y \in V$, there exists $0 \neq r \in R$ such that $ry \in M$. Moreover, if J is a finitely generated R-submodule of V, then there exists $0 \neq r \in R$ such that $rJ \subseteq M$, and J is an R-lattice if and only if there exists $0 \neq r \in R$ such that $rM \subseteq J \subseteq r^{-1}M$.*

Proof. Since $FM = V$, the R-lattice M contains an F-basis $x_1, \dots, x_n$ for V, so in particular $M \supset Rx_1 \oplus \cdots \oplus Rx_n$. Writing $y \in V$ in the basis $x_1, \dots, x_n$, clearing denominators we see that there exists $0 \neq r \in R$ such that $rx \in M$.

For the second statement, let y_i be a set of R-module generators for J; then there exist $r_i \in R$ such that $r_i y_i \in M$ hence $0 \neq r = \prod_i r_i$ satisfies $rJ \subseteq M$, so $J \subseteq r^{-1}M$. Repeating this argument with M interchanged with J and taking the product of the two, we have the result. □

8.3 Orders

Let B be an F-algebra.

Definition 8.3.1. An R-*order* $\mathcal{O} \subseteq B$ is an R-lattice that is also a subring of B.

In particular, if $\mathcal{O}$ is an R-order then we insist that $1 \in \mathcal{O}$.

8.3.2. An R-*algebra* is a ring $\mathcal{O}$ equipped with an embedding $R \hookrightarrow \mathcal{O}$ whose image lies in the center of $\mathcal{O}$. An R-order $\mathcal{O}$ is an R-algebra, and if $\mathcal{O}$ is an R-algebra that is finitely generated as an R-module, then $\mathcal{O}$ is an R-order of $B = \mathcal{O} \otimes_R F$.

Example 8.3.3. The matrix algebra $\mathrm{M}_n(F)$ has the R-order $\mathrm{M}_n(R)$. The subring $R[G] = \bigoplus_g Rg$ is an R-order in the group ring $F[G]$.

Example 8.3.4. Let $a, b \in R \setminus \{0\}$ and consider the quaternion algebra $B = \left(\dfrac{a,b}{F}\right)$. Then $\mathcal{O} = R \oplus Ri \oplus Rj \oplus Rij$ is an R-order.

Let $I \subseteq B$ be an R-lattice in the F-algebra B.

8.3.5. An important construction of orders comes as follows. Define the set

$$\mathcal{O}_{\mathrm{L}}(I) = \{\alpha \in B : \alpha I \subseteq I\}.$$

Then $\mathcal{O}_{\mathrm{L}}(I)$ is an R-submodule of B which is a ring. We show it is also an R-lattice. For any $\alpha \in B$, by Lemma 8.2.3 there exists $0 \neq r \in R$ such that $r(\alpha I) \subseteq I$, hence $\mathcal{O}_{\mathrm{L}}(I)F = B$. Also by this lemma, there exists $0 \neq s \in R$ such that $s = s \cdot 1 \in I$; thus $\mathcal{O}_{\mathrm{L}}(I)s \subseteq I$ so $\mathcal{O}_{\mathrm{L}}(I) \subseteq s^{-1}I$. Since R is noetherian and $s^{-1}I$ is an R-lattice so finitely generated, we conclude that $\mathcal{O}_{\mathrm{L}}(I)$ is finitely generated and is thus an R-lattice.

It follows that every F-algebra B has an R-order, since if $B = \bigoplus_i F\alpha_i$ then $I = \bigoplus_i R\alpha_i$ is an R-lattice.

Definition 8.3.6. The order

$$\mathcal{O}_{\mathrm{L}}(I) = \{\alpha \in B : \alpha I \subseteq I\}$$

is called the *left order* of I. We similarly define the *right order* of I by

$$\mathcal{O}_{\mathrm{R}}(I) = \{\alpha \in B : I\alpha \subseteq I\}.$$

Orders are composed of integral elements, defined as follows. If $\alpha \in B$, we denote by $R[\alpha] = \sum_d R\alpha^d$ the (commutative) R-subalgebra of B generated by α.

Definition 8.3.7. An element $\alpha \in B$ is *integral* over R if α satisfies a monic polynomial with coefficients in R.

Lemma 8.3.8. *For $\alpha \in B$, the following are equivalent:*

(i) *α is integral over R;*

(ii) *$R[\alpha]$ is a finitely generated R-module;*

(iii) *α is contained in a subring A which is a finitely generated R-module.*

Proof. If $\alpha \in B$ is integral and is a root of $f(x) = x^n + a_{n-1}x^{n-1} + \dots + a_0 \in R[x]$, then obviously $R[\alpha] = R + R\alpha + \dots + R\alpha^{n-1}$. Conversely, if $R[\alpha]$ is finitely generated, then α satisfies the characteristic polynomial of left multiplication by α on a basis for B consisting of elements of $\mathcal{O}$. This proves (i) $\Leftrightarrow$ (ii).

For the final equivalence, we see that (ii) $\Rightarrow$ (iii) is immediate, and for the converse, if $\mathcal{O} \subseteq B$ is an R-order, then every $\alpha \in \mathcal{O}$ is integral over R, since $R[\alpha]$ is a submodule of $\mathcal{O}$ so (since R is noetherian) $R[\alpha]$ is finitely generated. □

Corollary 8.3.9. *If $\mathcal{O}$ is an R-order, then every $\alpha \in \mathcal{O}$ is integral over R.*

We say R is *integrally closed* (in F) if any $\alpha \in F$ integral over R has $\alpha \in R$.

Inside the field F, the set of elements integral over R (the *integral closure* of R in F) forms a ring: if α, β are integral over R then $\alpha + \beta$ and $\alpha\beta$ are integral since they lie in $R[\alpha, \beta]$ which is a finitely generated submodule of F. This ring is itself integrally closed.

Lemma 8.3.10. *Suppose that R is integrally closed. Then $\alpha \in B$ is integral over R if and only if the minimal polynomial of α over F has coefficients in R.*

Proof. Let $f(x) \in R[x]$ be a monic polynomial that α satisfies, and let $g(x) \in F[x]$ be the minimal polynomial of α. Let K be a splitting field for $g(x)$, and let $\alpha_1, \dots, \alpha_n$ be the roots of $g(x)$ in K. Since $g(x) \mid f(x)$, each such α_i is integral over R, and the set of elements in K integral over R forms a ring, so each coefficient of g is integral over R and belongs to F; but since R is integrally closed, these coefficients must belong to R, so $g(x) \in R[x]$. □

Corollary 8.3.11. *If B is an F-algebra with a standard involution, and R is integrally closed, then $\alpha \in B$ is integral over R if and only if* $\operatorname{trd}(\alpha), \operatorname{nrd}(\alpha) \in R$.

The integral closure of R in F is the largest ring containing integral elements. Accordingly, we make the following more general definition.

Definition 8.3.12. An R-order is *maximal* if it is not properly contained in another R-order.

If B is a commutative F-algebra and R is integrally closed in F, then the integral closure S of R in K is integrally closed and therefore S is a maximal R-order in K. However, if B is noncommutative, then the set of elements in B integral over R is no longer necessarily itself a ring, and so the theory of maximal orders is more complicated. (This may seem counterintuitive at first, but certain aspects of the noncommutative situation are indeed quite different!)

Example 8.3.13. Let $B = \mathrm{M}_2(\mathbb{Q})$ and let $\alpha = \begin{pmatrix} 0 & 1/2 \\ 0 & 0 \end{pmatrix}$ and $\beta = \begin{pmatrix} 0 & 0 \\ 1/2 & 0 \end{pmatrix}$. Then $\alpha^2 = \beta^2 = 0$, so α, β are integral over $R = \mathbb{Z}$, but $\alpha+\beta$ is not integral since $\operatorname{nrd}(\alpha+\beta) = -1/4$ (Corollary 8.3.11). Such a counterexample does not require the existence of zerodivisors: see Exercise 8.9.

The problem in the noncommutative setting is that although $R[\alpha]$ and $R[\beta]$ may be finitely generated as R-modules, this need not be the case for the R-algebra generated by α and β: indeed, in the example above, it is not!

The structure of (maximal) orders in a quaternion algebra over the domains of arithmetic interest is the subject of the second Part of this text. To conclude this chapter, we discuss some special cases over the next few sections.

8.4 Orders in separable algebras

We have also the following characterization of orders in separable algebras.

Lemma 8.4.1. *Let $\mathcal{O} \subseteq B$ be a subring of a separable F-algebra B such that $\mathcal{O}F = B$. Then $\mathcal{O}$ is an R-order if and only if every $\alpha \in \mathcal{O}$ is integral.*

Proof. Let $\mathcal{O} \subseteq B$ be a subring of an F-algebra B such that $\mathcal{O}F = B$. Recall that a separable F-algebra is a semisimple F-algebra such that the symmetric bilinear pairing $(\alpha,\beta) \mapsto \operatorname{trd}(\alpha\beta)$ is nondegenerate.

We need to show that $\mathcal{O}$ is finitely generated. Let $\alpha_1, \dots, \alpha_n$ be an F-basis for B contained in $\mathcal{O}$. If $\beta \in \mathcal{O}$ then $\beta = \sum_i a_i\alpha_i$ with $a_i \in F$. We have $\beta\alpha_i \in \mathcal{O}$ since $\mathcal{O}$ is a ring, so $\operatorname{trd}(\beta\alpha_i) = \sum_j a_j \operatorname{trd}(\alpha_j\alpha_i)$ with $\operatorname{trd}(\alpha_j\alpha_i) \in R$. Now since B is separable, the matrix $(\operatorname{trd}(\alpha_i\alpha_j))_{i,j=1,\dots,n}$ is invertible, say $r = \det(\operatorname{trd}(\alpha_i\alpha_j))$, so we can solve these equations for a_j using Cramer's rule and we find that $a_j \in r^{-1}R$. Consequently $\mathcal{O} \subseteq r^{-1}(R\alpha_1 \oplus \cdots \oplus R\alpha_n)$ is a submodule of a finitely generated module so (since R is noetherian) we have that $\mathcal{O}$ is finitely generated. □

Remark 8.4.2. Assuming the axiom of choice, it follows from Lemma 8.4.1 that a separable F-algebra B has a maximal order. By Paragraph 8.3.5, B has an R-order $\mathcal{O}$ (since it has a lattice, taking the R-span of any F-basis), so the collection of R-orders containing $\mathcal{O}$ is nonempty. Given any chain of R-orders containing $\mathcal{O}$, by Lemma 8.4.1 the union of these orders is again an R-order. Thus, by Zorn's lemma, there exists a maximal element in this collection and B has a maximal order.

8.5 Orders in a matrix ring

Next, we study orders in a matrix ring. The matrix ring over F is just the endomorphism ring of a finite-dimension vector space over F, and we seek a similar description for orders as endomorphism rings of lattices, following Paragraph 8.3.5.

Let V be an F-vector space with $\dim_F V = n$ and let $B = \operatorname{End}_F(V)$. Choosing a basis of V gives an identification $B = \operatorname{End}_F(V) \cong \mathrm{M}_n(F)$. Given an R-lattice $I \subseteq V$, we define

$$\operatorname{End}_R(I) = \{f \in \operatorname{End}_F(V) : f(I) \subseteq I\} \subset B.$$

Note that the definition of $\operatorname{End}(I)$ differs from that of the left order (8.3.5): we do not take $B = V$, but rather, consider endomorphisms of lattices of smaller rank.

Example 8.5.1. If $V = Fx_1 \oplus \cdots \oplus Fx_n$ and $I = Rx_1 \oplus \cdots \oplus Rx_n$, then $\operatorname{End}_R(I) \cong \mathrm{M}_n(R)$.

More generally, if I is *completely decomposable*, i.e. $I = \mathfrak{a}_1x_1 \oplus \cdots \oplus \mathfrak{a}_nx_n$ with $\mathfrak{a}_i$ projective R-submodules of F, then $\operatorname{End}_R(J) \subseteq \mathrm{M}_n(F)$ consists of those matrices whose ijth entry lies in $\operatorname{Hom}_R(\mathfrak{a}_i, \mathfrak{a}_j) \subseteq \operatorname{Hom}_F(F, F) = F$. For example, if $n = 2$ then

$$\operatorname{End}_R(I) \cong \begin{pmatrix} R & \operatorname{Hom}_R(\mathfrak{a}_2, \mathfrak{a}_1) \\ \operatorname{Hom}_R(\mathfrak{a}_1, \mathfrak{a}_2) & R \end{pmatrix} \subset \mathrm{M}_2(F).$$

Lemma 8.5.2. *Let I be an R-lattice of V. Then* $\operatorname{End}_R(I)$ *is an R-order in* $B = \operatorname{End}_F(V)$.

Proof. As in Paragraph 8.3.5, we have $\operatorname{End}_R(I)F = B$. Let $\alpha_1, \alpha_2, \dots, \alpha_n$ be an F-basis for V and let $J = R\alpha_1 \oplus \dots \oplus R\alpha_n$. Then by Lemma 8.2.3 there exists $0 \neq r \in R$ such that $rJ \subseteq I \subseteq r^{-1}J$. Therefore $\operatorname{End}_R(rJ) = r^n \operatorname{End}_R(J) \subseteq \operatorname{End}_R(I) \subseteq r^{-n} \operatorname{End}_R(J)$, and so $\operatorname{End}_R(I)$ is an R-order in B. □

Lemma 8.5.3. *Let $\mathcal{O} \subseteq B = \operatorname{End}_F(V)$ be an R-order. Then $\mathcal{O} \subseteq \operatorname{End}_R(I)$ for some R-lattice $I \subseteq V$.*

Proof. Let J be any R-lattice in V, and let $I = \{\alpha \in J : \mathcal{O}\alpha \subseteq J\}$. Then I is an R-submodule of J with $FI = V$ (as in Paragraph 8.3.5), so I is an R-lattice in V and $\mathcal{O} \subseteq \operatorname{End}_R(I)$. □

Corollary 8.5.4. *If R is a PID, then every maximal R-order $\mathcal{O} \subseteq B \cong \mathrm{M}_n(F)$ is conjugate in B to $\mathrm{M}_n(R)$.*

Proof. The isomorphism $B \cong \mathrm{M}_n(F)$ arises from a basis $x_1, \dots, x_n$; letting $J = \bigoplus_i Rx_i$ we have $\operatorname{End}_R(J) \cong \mathrm{M}_n(R)$. Now the R-order $\mathrm{M}_n(R)$ is maximal by Exercise 8.6, since a PID is integrally closed. By the lemma, we have $\mathcal{O} \subseteq \operatorname{End}_R(I)$ for some R-lattice $I \subseteq V$, so if $\mathcal{O}$ is maximal then $\mathcal{O} = \operatorname{End}_R(I)$. If R is a PID then $I = Ry_1 \oplus \dots \oplus Ry_n$, and the change of basis matrix from x_i to y_i realizes $\operatorname{End}_R(I)$ as a conjugate of $\operatorname{End}_R(J) \cong \mathrm{M}_n(R)$. □

An order $\mathcal{O} \subseteq \operatorname{End}_R(I)$ can be thought of as a subring of endomorphisms of a lattice preserving some extra structure. We consider this matter in detail in the quaternionic context of 2×2-matrices in Chapter 16.

8.6 Quadratic forms

In setting up an integral theory, we will also have need of an extension of the theory of quadratic forms over a PID; these notions generalize those over fields (Section 4.2) in a straightforward way.

Let R be a PID.

Definition 8.6.1. A *quadratic form* over R is a map $Q : M \to R$ where M is a (free) R-module satisfying:

(i) $Q(rx) = r^2 Q(x)$ for all $r \in R$ and $x \in R^n$; and

(ii) The map $T : R^n \times R^n \to R$ defined by

$$T(x,y) = Q(x+y) - Q(x) - Q(y)$$

is R-bilinear.

T is called the *associated bilinear map* .

8.6.2. Let $Q : V \to F$ be a quadratic form with F the field of fractions of R. Let $M \subseteq V$ be a finitely generated R-lattice such that $Q(M) \subseteq R$. Then the restriction $Q|_M : M \to R$ is a quadratic form. Conversely, if $Q : M \to R$ is a quadratic form over R, then the extension $Q : M \otimes_R F \to F$ is a quadratic form over F.

Definition 8.6.3. A *similarity* between quadratic forms $Q : M \to R$ and $Q' : M' \to R$ is an isomorphism $f : M \xrightarrow{\sim} M'$ and $u \in R^\times$ such that $Q(f(x)) = uQ'(x)$ for all $x \in M$. An *isometry* between quadratic forms is a similarity with $u = 1$.

Let $Q : M \to R$ be a quadratic form over R. Then Q is *nondegenerate* if the extension $Q : M \otimes_R F \to F$ is nondegenerate. [[Nonsingular?]] From now on, suppose that $M \cong R^n$ is free of finite rank n in the basis $e_1, \dots, e_n$. We then define the *discriminant* $\operatorname{disc}(Q)$ as the (half-)determinant of the *Gram matrix* $(T(e_i, e_j))_{i,j}$, as in Definition 4.2.9. [[Nonsingular? and differences between them?]]

8.7 Extensions and further reading

8.7.1. The hypothesis that R is noetherian is used in Paragraph 8.3.5; it seems possible that the left order may not be finitely generated. Perhaps noetherian induction will work? [[Used in other places?]].

Exercises

Let R be a noetherian domain with field of fractions F.

8.1. Let L, M be R-lattices in a vector space V with $\dim_F V < \infty$. Show that $L + M$ and $L \cap M$ are R-lattices.

8.2. Let B be an F-algebra and let $I \subset B$ be an R-lattice. Show that there exists a nonzero $r \in R \cap I$.

8.3. Let $\mathfrak{c} \subseteq R$ be a nonzero ideal. Show that

$$\begin{pmatrix} R & R \\ \mathfrak{c} & R \end{pmatrix} = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{M}_2(R) : c \in \mathfrak{c} \right\} \subseteq \mathrm{M}_2(R)$$

is an R-order in $\mathrm{M}_2(F)$.

8.4. Let $\mathcal{O}, \mathcal{O}' \subseteq B$ be R-orders. Show that $\mathcal{O} \cap \mathcal{O}'$ is an R-order.

8.5. Let $A_1, \dots, A_r$ be F-algebras and let $B = A_1 \times \cdots \times A_r$. Show that $\mathcal{O} \subseteq B$ is an R-order if and only if $\mathcal{O} \cap A_i$ is an R-order for each i.

8.6. Let R be integrally closed. Show that $\mathrm{M}_n(R)$ is a maximal R-order in $\mathrm{M}_n(F)$.

8.7. Let $B = \left(\dfrac{K, b}{F}\right)$ with $b \in R$ and let S be an R-order in K. Let $\mathcal{O} = S + Sj$. Show that $\mathcal{O}$ is an R-order in B.

8.8. Let B be an F-algebra with a standard involution and let $\alpha \in B$. Show that if α is integral over R then $\operatorname{trd}(\alpha^n) \in R$ for all $n \in \mathbb{Z}_{\geq 0}$. Is the converse true?

8.9. Generalize Example 8.3.13 as follows.

a) Find an algebra B over a field F and elements $\alpha, \beta \in B$ such that α, β are integral over $R \subseteq F$ but $\alpha\beta$ is not.

b) Find a division ring D over a field F and elements $\alpha, \beta \in D$ such that α, β are integral over $R \subseteq F$ but $\alpha + \beta$ is not.

8.10. Give an example of a non-noetherian ring R and modules $J \subset I$ such that I is finitely generated but J is not finitely generated. Does this yield an example where $\mathcal{O}_{\mathrm{L}}(I)$ is not an R-lattice (cf. Paragraph 8.3.5)?

8.11. Let $\alpha \in \mathrm{M}_n(F)$ have characteristic polynomial with coefficients in R. Show that α is conjugate by an element $\beta \in \mathrm{GL}_n(F)$ to an element of $\mathrm{M}_n(R)$. Explicitly, how do you find such a matrix β?

8.12. Let $\mathcal{O} \subseteq B$ be an R-order.

a) Show that $\mathcal{O}_{\mathrm{L}}(\mathcal{O}) = \mathcal{O}_{\mathrm{R}}(\mathcal{O}) = \mathcal{O}$.

b) Let $\alpha \in B^\times$, and let $\alpha\mathcal{O} = \{\alpha\beta : \beta \in \mathcal{O}\}$. Show that $\alpha\mathcal{O}$ is an R-lattice and that $\mathcal{O}_{\mathrm{L}}(\alpha\mathcal{O}) = \alpha\mathcal{O}\alpha^{-1}$.

8.13. Let $\mathcal{O} \subseteq B$ be an R-order and let $\gamma \in \mathcal{O}$ and let $N : B^\times \to F^\times$ be any multiplicative map. Show that $\gamma \in \mathcal{O}^\times$ if and only if $N(\gamma) \in R^\times$, and in particular, if B has a standard involution, then $\gamma \in \mathcal{O}^\times$ if and only if $\operatorname{nrd}(\gamma) \in R^\times$.

[[More explicit examples.]]

Part II

Algebraic number theory

Chapter 9

The Hurwitz order

Before we embark on a general treatment of quaternion algebras over number fields and the arithmetic of their orders, we consider the special case of the Hurwitz order. This is not only instructive for what follows, but this order possesses certain exceptional symmetries that make it worthy of specific investigation.

9.1 The Hurwitz order

We consider in this chapter the restriction of the Hamiltonians from $\mathbb{R}$ to $\mathbb{Q}$, namely, the quaternion algebra $B = \left(\dfrac{-1,-1}{\mathbb{Q}}\right)$. We consider first the natural further restriction to those elements with integer coordinates

$$\mathbb{Z}\langle i, j, k\rangle = \mathbb{Z} + \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}k;$$

by Example 8.3.4, this is a $\mathbb{Z}$-order in B, called the *Lipschitz order*. In the rest of this chapter, we will work over $\mathbb{Z}$ and so we will simply refer to $\mathbb{Z}$-lattices and $\mathbb{Z}$-orders as *lattices* and *orders*.

The Lipschitz order is not a maximal order, and as we will see later on, this makes it less suitable for the development of an algebraic theory; this is analogous to the fact that the ring $\mathbb{Z}[\sqrt{-3}]$ is a order in $\mathbb{Q}(\sqrt{-3})$ but is not maximal (not integrally closed) as it is properly contained in the maximal order $\mathbb{Z}[(-1+\sqrt{-3})/2]$ of Eisenstein integers.

Lemma 9.1.1. *The lattice*

$$\mathcal{O} = \mathbb{Z} + \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}\left(\frac{-1+i+j+k}{2}\right) \tag{9.1.2}$$

in B is a maximal order that properly contains $\mathbb{Z}\langle i, j\rangle$ (with index $[\mathcal{O} : \mathcal{O}'] = 2$).

The order $\mathcal{O}$ (9.1.2) is called the *Hurwitz order*.

Proof. By Exercise 9.1, the lattice $\mathcal{O}$ is an order. Suppose that $\mathcal{O}' \supseteq \mathcal{O}$ and let $\alpha = t + xi + yj + zk \in \mathcal{O}'$ with $t, x, y, z \in \mathbb{Q}$. Then $\operatorname{trd}(\alpha) = 2t \in \mathbb{Z}$, so by Corollary 8.3.11 we have $t \in \frac{1}{2}\mathbb{Z}$. Similarly, we have $\alpha i \in \mathcal{O}'$ so $\operatorname{trd}(\alpha i) = -2x \in \mathbb{Z}$, hence $x \in \frac{1}{2}\mathbb{Z}$, and in the same way $y, z \in \frac{1}{2}\mathbb{Z}$. Finally, $\operatorname{nrd}(\alpha) = t^2 + x^2 + y^2 + z^2 \in \mathbb{Z}$, and considerations modulo 4 imply that t, x, y, z either all belong to $\mathbb{Z}$ or to $\frac{1}{2} + \mathbb{Z}$; thus $\alpha \in \mathcal{O}$ and $\mathcal{O}' = \mathcal{O}$. □

Note that the element $\omega = (-1 + i + j + k)/2$ satisfies $\omega^2 + \omega + 1 = 0$, so the comparison with the Eisenstein integers is more than incidental: we have $\mathbb{Z}[\sqrt{-3}] \cong \mathbb{Z}[i + j + k] \subseteq \mathbb{Z}\langle i, j, k\rangle$, and both extend to a maximal order in a parallel way.

9.2 Hurwitz units and finite subgroups of the Hamiltonians

We now consider unit groups. An element $\gamma = t+xi+yj+zij \in \mathbb{Z}\langle i, j\rangle$ is a unit if and only if $\operatorname{nrd}(\gamma) = t^2 + x^2 + y^2 + z^2 = 1 \in \mathbb{Z}^\times$, and since $t, x, y, z \in \mathbb{Z}$ we immediately have

$$\mathbb{Z}\langle i, j, k\rangle^\times = \{\pm 1, \pm i, \pm j, \pm ij\} \cong Q_8$$

is the *quaternion group* of order 8. In a similar way, allowing $t, x, y, z \in \frac{1}{2}\mathbb{Z}$ we find that

$$\mathcal{O}^\times = \{(\pm 1 \pm i \pm j \pm k)/2\} \cup \{\pm 1, \pm i, \pm j, \pm k\}$$

is a group of order 24.

We have $\mathcal{O}^\times \not\cong S_4$ because there is already no embedding $Q_8 \hookrightarrow S_4$. (The permutation representation $Q_8 \to S_4$ obtained by the action on the cosets of the unique subgroup $\langle -1\rangle$ of index 4 factors through the quotient $Q_8 \to Q_8/\pm 1 \cong V_4 \hookrightarrow S_4$, where V_4 is the Klein 4-group.) There are 15 groups of order 24 up to isomorphism; we identify the right one as follows.

Lemma 9.2.1. *We have* $\mathcal{O}^\times \cong \mathrm{SL}_2(\mathbb{F}_3)$.

Proof. We obtain this isomorphism by reduction modulo 3. We have a ring homomorphism

$$\mathcal{O} \to \mathcal{O}/3\mathcal{O} \cong \mathbb{F}_3\langle i, j, k\rangle \cong \left(\frac{-1, -1}{\mathbb{F}_3}\right).$$

But any quaternion algebra over a finite field is isomorphic to the matrix ring by Wedderburn's theorem (Exercises 3.12, 5.11, and 6.16). Specifically, the element

$\epsilon = i + j + k$ has $\epsilon^2 = 0 \in \mathcal{O}/3\mathcal{O}$, and so the left ideal generated by ϵ has basis ϵ and $i\epsilon i = -1 - j + k$ and this yields an isomorphism (Proposition 6.5.2)

$$\begin{aligned} \mathcal{O}/3\mathcal{O} &\to \mathrm{M}_2(\mathbb{F}_3) \\ i, j &\mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \end{aligned}$$

(Exercise 9.2). This map gives an injective group homomorphism $\mathcal{O}^\times \hookrightarrow \mathrm{SL}_2(\mathbb{F}_3)$, since the reduced norm corresponds to the determinant. By a comparison of orders, we see that this is an isomorphism. □

We have a permutation representation $\mathrm{SL}_2(\mathbb{F}_3) \to S_4$ obtained from the natural action of $\mathrm{SL}_2(\mathbb{F}_3)$ on the set $\mathbb{P}^1(\mathbb{F}_3) = \mathbb{F}_3 \cup \{\infty\}$ by left multiplication; the kernel of this map is the subgroup generated by the scalar matrix -1 and so the image is $\mathrm{PSL}_2(\mathbb{F}_3) = \mathrm{SL}_2(\mathbb{F}_3)/\{\pm 1\} \cong A_4$, and in particular we have an exact sequence

$$1 \to \{\pm 1\} \to \mathcal{O}^\times \to A_4 \to 1. \tag{9.2.2}$$

We can also visualize this group (and the exact sequence (9.2.2)), thinking of the Hamiltonians as acting by rotations (Section 2.3). Recall we have an exact sequence

$$1 \to \{\pm 1\} \to \mathbb{H}_1^\times \to \mathrm{SO}(3) \to 1$$

obtained by the left action $\alpha \mapsto \alpha v \alpha^{-1}$ for $\alpha \in \mathbb{H}_1^\times$ and $v \in \mathbb{H}^0 \cong \mathbb{R}^3$; specifically, by Proposition 2.3.10, a quaternion $\alpha = \cos\theta + I(\alpha)\sin\theta$ acts by rotation through the angle 2θ about the axis $I(\alpha)$.

Then we can think of the group $\mathcal{O}^\times/\{\pm 1\} \cong A_4$ as the group of symmetries (rigid motions) of a tetrahedron (or rather, a tetrahedron and its dual), as follows. Inside the cube in $\mathbb{R}^3$ with vertices $(\pm 1, \pm 1, \pm 1)$, we can find four inscribed tetrahedra, for example, the tetrahedron T with vertices $(1, 1, 1), (1, -1, -1), (-1, 1, -1), (-1, -1, 1)$. Then the elements $\pm i, \pm j, \pm k$ act by rotation about the x, y, z axes by an angle $\pi = 180°$ (so interchanging points with the same x, y, z coordinate). The element $\pm\omega = \pm(-1 + i + j + k)/2$ rotates by the angle $2\pi/3 = 120°$ fixing the point $(1, 1, 1)$ and cyclically permuting the other three points, and by symmetry we understand the action of the other elements of $\mathcal{O}^\times$. We therefore call $\mathcal{O}^\times$ the *binary tetrahedral group*; the notations $2T \cong \widehat{A_4}$ are also used for this group.

The subgroup $Q_8 \trianglelefteq 2T$ is normal (as it is characteristic, consisting of all elements of $\mathcal{O}$ of order dividing 4), and so we can write $2T = Q_8 \rtimes \mathbb{Z}/3\mathbb{Z}$ where $\mathbb{Z}/3\mathbb{Z}\langle\omega\rangle$ acts on Q_8 by conjugation, cyclically rotating the elements i, j, k. Finally, the group $2T$ has a presentation

$$2T \cong \langle r, s, t \mid r^2 = s^3 = t^3 = rst = -1\rangle \tag{9.2.3}$$

via $r = i$, $s = \omega = (-1+i+j+k)/2$, and $t = (-1+i+j-k)/2$ (Exercise 9.5); we will see later (Section 24.1) that this realizes it as a spherical triangle group.

We conclude by noting that the difference between the Lipschitz and Hurwitz orders is "covered" by the extra units.

Lemma 9.2.4. *For any $\beta \in \mathcal{O}$, there exists $\gamma \in \mathcal{O}^\times$ such that $\beta\gamma \in \mathbb{Z}\langle i, j, k\rangle$.*

Proof. If $\beta \in \mathbb{Z}\langle i, j, k\rangle$ already, then we are done. Otherwise, $2\beta = t + xi + yj + zk$ with all $t, x, y, z \in \mathbb{Z}$ odd. Choosing matching signs, there exists $\gamma \in \mathcal{O}^\times$ such that $2\beta \equiv 2\gamma \pmod{4\mathcal{O}}$. Thus

$$(2\beta)\gamma^{-1} \equiv 2 \pmod{4\mathcal{O}}$$

so $\beta\gamma^{-1} \in \mathbb{Z} + 2\mathcal{O} = \mathbb{Z}\langle i, j, k\rangle$, as claimed. □

9.3 Euclidean algorithm, sums of four squares

The Eisenstein order $\mathbb{Z}[(-1+\sqrt{-3})/2]$ has several nice properties. Perhaps nicest of all is that it is a Euclidean domain, so in particular it is a PID and UFD. (Alas, the ring $\mathbb{Z}[\sqrt{-3}]$ just fails to be Euclidean.)

The Hurwitz order also has a left (or right) Euclidean algorithm generalizing the commutative case, as follows. We have an embedding $B \hookrightarrow B \otimes_{\mathbb{Q}} \mathbb{R} \cong \mathbb{H}$, and inside $\mathbb{H} \cong \mathbb{R}^4$ the Hurwitz order sits as a ($\mathbb{Z}$-)lattice equipped with the Euclidean inner product, so we can think of the reduced norm by instead thinking of distance. In the Lipschitz order, we see by rounding coordinates that for any $\gamma \in B$ there exists $\mu \in \mathbb{Z}\langle i, j, k\rangle$ such that $\operatorname{nrd}(\gamma - \mu) \leq 4 \cdot (1/2)^2 = 1$—a farthest point occurs at the center $(1/2, 1/2, 1/2, 1/2)$ of a unit cube. But this is precisely the point where the Hurwitz quaternions occur, and it follows that for any $\gamma \in B$, there exists $\mu \in \mathcal{O}$ such that $\operatorname{nrd}(\gamma - \mu) < 1$. (In fact, we can take $\operatorname{nrd}(\gamma - \mu) \leq 1/2$; see Exercise 9.6.)

This becomes a euclidean algorithm in the usual way.

Lemma 9.3.1 (Hurwitz order is Euclidean)**.** *For all $\alpha, \beta \in \mathcal{O}$ with $\beta \neq 0$, there exists $\mu, \rho \in \mathcal{O}$ such that*

$$\alpha = \mu\beta + \rho$$

and $\operatorname{nrd}(\rho) < \operatorname{nrd}(\beta)$.

A similar statement holds for division on the right, i.e., we may also take $\alpha = \beta\mu + \rho$.

Proof. If $\operatorname{nrd}(\alpha) < \operatorname{nrd}(\beta)$, we may take $\mu = 1$ and $\rho = 0$, so suppose $\operatorname{nrd}(\alpha) \geq \operatorname{nrd}(\beta) > 0$. Let $\gamma = \alpha\beta^{-1} \in B$. Then there exists $\mu \in \mathcal{O}$ such that $\operatorname{nrd}(\gamma - \mu) =$

$\mathrm{nrd}(\alpha\beta^{-1} - \mu) < 1$, so by multiplicativity, $\mathrm{nrd}(\alpha - \mu\beta) < \mathrm{nrd}(\beta)$. So we may take $\rho = \alpha - \mu\beta$. □

Lemma 9.3.2. *Every left (or right) ideal $I \subseteq \mathcal{O}$ is left principal, i.e., there exists $\beta \in I$ such that $I = \beta\mathcal{O}$.*

Proof. Let $I \subseteq \mathcal{O}$ be a left ideal. If $I = \{0\}$, we are done. Otherwise, there exists an element $0 \neq \beta \in I$ with minimal reduced norm $\mathrm{nrd}(\beta) \in \mathbb{Z}_{>0}$. We claim that $I = \mathcal{O}\beta$. Indeed, for any $\alpha \in I$, by the left Euclidean algorithm in Lemma 9.3.1, there exists $\mu \in \mathcal{O}$ such that $\alpha = \mu\beta + \rho$ with $\mathrm{nrd}(\rho) < \mathrm{nrd}(\beta)$; but $\rho = \alpha - \mu\beta \in I$, so by minimality, we have $\mathrm{nrd}(\rho) = 0$ so $\rho = 0$, hence $\alpha \in \mathcal{O}\beta$ as claimed. □

Corollary 9.3.3 (Bezout's theorem)**.** *For all $\alpha, \beta \in \mathcal{O}$ with $\alpha\mathcal{O} + \beta\mathcal{O} = \gamma\mathcal{O}$, there exists $\mu, \nu \in \mathcal{O}$ such that $\alpha\mu + \beta\nu = \gamma$.*

It does not follow that there is unique factorization in $\mathcal{O}$ in the traditional sense, as the order of multiplication matters. Nevertheless, there is a theory of prime factorization in $\mathcal{O}$ as follows.

Definition 9.3.4. An element $\pi \in \mathcal{O}$ is *prime* if whenever $\alpha\beta = \pi$ with $\alpha, \beta \in \mathcal{O}$ then $\alpha \in \mathcal{O}^\times$ or $\beta \in \mathcal{O}^\times$.

Lemma 9.3.5. *Let $\pi \in \mathcal{O}$. Then the following are equivalent:*

(i) *$I = \mathcal{O}\pi$ is a prime right ideal, i.e., if $\alpha\beta \in I$ then $\alpha \in I$ or $\beta \in I$;*

(ii) *$\pi\mathcal{O}$ is a prime left ideal; and*

(iii) *$\mathrm{nrd}(\pi) = p \in \mathbb{Z}$ is a prime.*

If any one of the conditions of Lemma 9.3.5 is satisfied, we say that $\pi \in \mathcal{O}$ is *prime*.

Proof. We begin by proving the last statement. □

Lemma 9.3.6. *For all primes p, there exists $\pi \in \mathcal{O}$ such that $\pi\overline{\pi} = \mathrm{nrd}(\pi) = p$.*

Proof. We have $2 = 1^2 + 1^2 + 0^2 + 0^2$, so we may assume $p \geq 3$ is odd. Then we have $\mathcal{O}/p\mathcal{O} \cong \left(\dfrac{-1,-1}{\mathbb{F}_p}\right)$ and by Wedderburn's theorem, we have

$$\left(\frac{-1,-1}{\mathbb{F}_p}\right) \cong \mathrm{M}_2(\mathbb{F}_p)$$

so there exists a left ideal $I \bmod p \subset \mathcal{O}/p\mathcal{O}$ with $\dim_{\mathbb{F}_p}(I \bmod p) = 2$. Let

$$I = \{\alpha \in \mathcal{O} : \alpha \bmod p \in I \bmod p\};$$

in particular we have $p\mathcal{O} \subsetneq I \subsetneq \mathcal{O}$. Then $I \subset \mathcal{O}$ is a left ideal, and $I \neq \mathcal{O}$. But $I = \beta\mathcal{O}$ is left principal by Lemma 9.3.2.

We claim that $\operatorname{nrd}(\beta) = p$. Since $p \in I$, we have $p = \beta\mu$ for some $\mu \in \mathcal{O}$, whence $\operatorname{nrd}(p) = p^2 = \operatorname{nrd}(\beta)\operatorname{nrd}(\mu)$ so $\operatorname{nrd}(\beta) \mid p^2$. We cannot have $\operatorname{nrd}(\beta) = 1$ or $\operatorname{nrd}(\beta) = p^2$, as these would imply $I = \mathcal{O}$ or $I = p\mathcal{O}$, impossible. So $\operatorname{nrd}(\beta) = p$. □

Definition 9.3.7. An element $\alpha \in \mathcal{O}$ is *primitive* if $\alpha \notin n\mathcal{O}$ for all $n \in \mathbb{Z}$.

[[Define left divisibility, etc.?]]

Theorem 9.3.8 (Conway–Smith)**.** *Let $\alpha \in \mathcal{O}$ be primitive and let $a = \operatorname{nrd}(\alpha)$. Factor $a = p_1 p_2 \cdots p_r$ into a product of primes. Then there exists $\pi_1, \pi_2, \dots, \pi_r \in \mathcal{O}$ such that*

$$\alpha = \pi_1\pi_2\cdots\pi_r, \quad \textit{and } \operatorname{nrd}(\pi_i) = p_i \textit{ for all } i.$$

Moreover, any other such factorization is of the form

$$\alpha = (\pi_1\gamma_1)(\gamma_1^{-1}\pi_2\gamma_2^{-1})\cdots(\gamma_{r-1}^{-1}\pi_r) \tag{9.3.9}$$

where $\gamma_1, \dots, \gamma_r \in \mathcal{O}^\times$.

The factorization (9.3.9) is said to be obtained from $\alpha = \pi_1 \cdots \pi_r$ by *unit migration*.

Theorem 9.3.10. *Every integer $n \geq 0$ is the sum of four squares, i.e., there exists $t, x, y, z \in \mathbb{Z}$ such that $n = t^2 + x^2 + y^2 + z^2$.*

Proof. We seek an element $\beta \in \mathbb{Z}\langle i, j, k\rangle$ such that $\operatorname{nrd}(\beta) = n$. By multiplicativity, it is sufficient to treat the case where $n = p$ is prime. We obtain $\pi \in \mathcal{O}$ such that $\operatorname{nrd}(\pi) = p$ by Lemma 9.3.6. But now the result follows from Lemma 9.2.4, as there exists $\gamma \in \mathcal{O}^\times$ such that $\pi\gamma \in \mathbb{Z}\langle i, j, k\rangle$. □

9.4 Sums of three squares

9.5 Extensions and further reading

Exercises

9.1. Show that

$$\mathcal{O} = \mathbb{Z} + \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}\left(\frac{1+i+j+k}{2}\right)$$

is an order in $B = \left(\dfrac{-1,-1}{\mathbb{Q}}\right)$.

9.2. Check that the map

$$\begin{aligned} \mathcal{O}/3\mathcal{O} &\to \mathrm{M}_2(\mathbb{F}_3) \\ i,j &\mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \end{aligned}$$

from Lemma 9.2.1 is the isomorphism obtained from the representation on the left ideal generated by $\epsilon = i + j + k$.

9.3. Draw the subgroup lattice for $\mathrm{SL}_2(\mathbb{F}_3)$, indicating normal subgroups (and their quotients).

9.4. Let $\mathcal{O} \subset B = \left(\dfrac{-1,-1}{\mathbb{Q}}\right)$ be the Hurwitz order.

a) Consider the natural ring homomorphism $\mathcal{O} \to \mathcal{O} \otimes_{\mathbb{Z}} \mathbb{F}_2 = \mathcal{O}_{\mathbb{F}_2}$ giving the reduction of the algebra $\mathcal{O}$ modulo 2. Show that $\#\mathcal{O}_{\mathbb{F}_2} = 16$, that $\mathcal{O}_{\mathbb{F}_2}^\times \cong A_4$ the alternating group on 4 elements. Conclude that $\mathcal{O}_{\mathbb{F}_2} \not\cong \mathrm{M}_2(\mathbb{F}_2)$ and hence that $\mathcal{O}_{\mathbb{F}_2}$ is not a quaternion algebra over $\mathbb{F}_2$.

b) Show that $\mathrm{Aut}_{\mathbb{F}_2}(\mathcal{O}_{\mathbb{F}_2}) \cong S_4$ (automorphisms as an $\mathbb{F}_2$-algebra). More generally, if F is a field of characteristic 2, show that $\mathcal{O}_F = \mathcal{O} \otimes_{\mathbb{Z}} F$ has $\mathrm{Aut}_F\, \mathcal{O}_F$ is an extension of $\mathrm{SL}_2(F)$ by the additive group F^2.

This kind of construction, considered instead over the octonions, arises when constructing the exceptional group G_2 in characteristic 2 [Wil09, §4.4.1].

9.5. Show that

$$2T \cong \langle r, s, t \mid r^2 = s^3 = t^3 = rst = -1 \rangle$$

(cf. (9.5)).

9.6. Let $\Lambda = \mathbb{Z}^4 + ((\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}) + \mathbb{Z}^4)$ be the image of the Hurwitz order $\mathcal{O}$ under the natural embedding $\mathcal{O} \hookrightarrow \mathbb{H} \cong \mathbb{R}^4$. Show that for every $x \in \mathbb{R}^4$, there exists $\lambda\Lambda$ such that $\|\lambda\|^2 \leq 1/2$. *[Hint: without loss of generality we may take* $0 \leq x_i \leq 1/2$ *for all i; then show we may take* $x_1 + x_2 + x_3 + x_4 \leq 1$*; conclude that the maximum value of* $\|x\|^2$ *with these conditions occurs at the point* $(\frac{1}{2}, \frac{1}{2}, 0, 0)$*.]*

Chapter 10

Quaternion algebras over local fields

In this chapter, we classify quaternion algebras over local fields; this generalizes the classification of quaternion algebras over $\mathbb{R}$.

10.1 Local quaternion algebras

Having spent the first part of this book exploring the properties of quaternion algebras, we now seek to classify them over a nice class of fields. Over any field F we have the matrix ring $\mathrm{M}_2(F)$, and if F is a finite field or an algebraically closed field F, then any quaternion algebra over F is isomorphic to the matrix ring. The 'first' quaternion algebra, of course, was the division ring $\mathbb{H}$ of Hamiltonians, and this ring is the only division quaternion ring over $\mathbb{R}$ up to isomorphism.

In this section, we will classify quaternion algebras over a field F that is in some sense similar to $\mathbb{R}$. We will insist that the field F is equipped with a topology compatible with the field operations in which F is Hausdorff and locally compact (every element of F has a compact neighborhood). To avoid trivialities, we will insist that this topology is not the discrete topology (where every subset of F is open): such a topological field is called a *local field*.

For purposes of illustration, we consider local fields F that contain the rational numbers $\mathbb{Q}$ as a dense subfield. Such a field F is the completion of $\mathbb{Q}$ with respect to an absolute value $|\,|$, so is obtained as the set of equivalence classes of Cauchy sequences, and has a topology induced by the metric $d(x, y) = |x - y|$. By a theorem of Ostrowski, such an absolute value is equivalent to either the usual archimedean absolute value or a *p-adic absolute value*, defined by $|0|_p = 0$ and

$$|c|_p = p^{-\operatorname{ord}_p(c)} \quad \text{for } c \in \mathbb{Q}^\times,$$

where $\operatorname{ord}_p(c)$ is the power of p occurring in c in its unique factorization (taken to be negative if p divides the denominator of c written in lowest terms).

Just as elements of $\mathbb{R}$ can be thought of infinite decimals, an element of $\mathbb{Q}_p$ can be thought of in its p-adic expansion

$$a = (\dots a_3 a_2 a_1 a_0 . a_{-1} a_{-2} \cdots a_{-k})_p = \sum_{n=-k}^{\infty} a_n p^n$$

where each $a_i \in \{0, \dots, p-1\}$ are the *digits* of a. We continue "to the left" because a decimal expansion is a series in the base $1/10 < 1$ and instead we have a base $p > 1$. Inside $\mathbb{Q}_p$ is the ring $\mathbb{Z}_p$ of *p-adic integers*, the completion of $\mathbb{Z}$ with respect to $|\ |_p$: the ring $\mathbb{Z}_p$ consists of those elements of $\mathbb{Q}_p$ with $a_n = 0$ for $n < 0$. The ring $\mathbb{Z}_p$ might be thought of intuitively as $\mathbb{Z}/p^\infty\mathbb{Z}$, if this made sense: they were first defined in this context by Hensel, who wanted a uniform language for when a Diophantine equation has a solution modulo p^n for all n.

By construction, the ring $\mathbb{Z}_p$ and the field $\mathbb{Q}_p$ come equipped with a topology arising from its metric $d_p(x, y) = |x - y|_p$. With respect to this topology, in fact $\mathbb{Z}_p$ is compact and $\mathbb{Q}_p$ is locally compact. It is easiest to see this by viewing $\mathbb{Z}_p$ as a projective limit with respect to the natural maps $\mathbb{Z}/p^{n+1}\mathbb{Z} \to \mathbb{Z}/p^n\mathbb{Z}$:

$$\begin{aligned} \mathbb{Z}_p &= \varprojlim_n \mathbb{Z}/p^n\mathbb{Z} \\ &= \left\{ x = (x_n)_n \in \prod_{n=0}^{\infty} \mathbb{Z}/p^n\mathbb{Z} : x_{n+1} \equiv x_n \pmod{p^n} \text{ for all } n \geq 0 \right\}. \end{aligned} \tag{10.1.1}$$

In other words, each element of $\mathbb{Z}_p$ is a compatible sequence of elements in $\mathbb{Z}/p^n\mathbb{Z}$ for each n. The equality (10.1.1) is just a reformulation of the notion of Cauchy sequence for $\mathbb{Z}$, and so for the purposes of this introduction it can equally well be taken as a definition. As for the topology in (10.1.1), each factor $\mathbb{Z}/p^n\mathbb{Z}$ is given the discrete topology, the product $\prod_{n=0}^{\infty} \mathbb{Z}/p^n\mathbb{Z}$ is given the product topology, and $\mathbb{Z}_p$ is given the restriction topology. Since each $\mathbb{Z}/p^n\mathbb{Z}$ is compact (it is a finite set!), by Tychonoff's theorem the product $\prod_{n=0}^{\infty} \mathbb{Z}/p^n\mathbb{Z}$ is compact; and $\mathbb{Z}_p$ is closed inside this product (a convergent limit of Cauchy sequences is a Cauchy sequence), so $\mathbb{Z}_p$ is compact. The set $\mathbb{Z}_p$ is a neighborhood of 0, indeed, it is the closed ball of radius 1 around 0:

$$\mathbb{Z}_p = \{x \in \mathbb{Q}_p : |a|_p \leq 1\}.$$

In a similar way, the disc of radius 1 around any $a \in \mathbb{Q}_p$ is a compact neighborhood of a homeomorphic to $\mathbb{Z}_p$, so $\mathbb{Q}_p$ is locally compact.

As is evident from this argument, although $\mathbb{Q}_p$ is Hausdorff and locally compact, it has a rather strange topology, akin to a Cantor set: $\mathbb{Q}_p$ is *totally disconnected* (the largest connected subsets consist of single points). Nevertheless, being able to make topological arguments like the one above is the whole point of looking at local fields like $\mathbb{Q}_p$: our understanding of algebraic objects is informed by the topology.

In particular, we have a result for quaternion algebras over $\mathbb{Q}_p$ that is quite analogous to that over $\mathbb{R}$.

Theorem 10.1.2. *There is a unique division quaternion algebra B over $\mathbb{Q}_p$, up to isomorphism. In fact, if $p \neq 2$, then B is given by*

$$B \cong \left(\frac{e, p}{\mathbb{Q}_p}\right)$$

where $e \in \mathbb{Z}$ is a quadratic nonresidue modulo p.

We approach this theorem in two ways in this section. The first way is using the language of quadratic forms, and for that we use the classification of isomorphism classes of quaternion algebras in terms of similarity classes of ternary quadratic forms. The following proposition then implies Theorem 10.1.2.

Proposition 10.1.3. *There is a unique ternary anisotropic quadratic form Q over $\mathbb{Q}_p$, up to similarity. If $p \neq 2$, then $Q \sim \langle 1, -e, -p \rangle$ where e is a quadratic nonresidue modulo p.*

This proposition can be proved using some rather direct manipulations with quadratic forms. On the other hand, it has the defect that quadratic forms behave differently in characteristic 2, and so one may ask for a more uniform proof. This is the second way that we approach the proof of Theorem 10.1.2: we extend the absolute value on $\mathbb{Q}_p$ to one on a division quaternion algebra B, and use this extension to show that B is unique by direct examination of its valuation ring and two-sided maximal ideal. While it requires a bit more theory, this method of proof also can be used to classify central division algebras over $\mathbb{Q}_p$ in much the same manner.

10.2 Local fields

Definition 10.2.1. A *topological ring* is a ring A equipped with a topology such that the ring operations (addition, negation, and multiplication) are continuous. A *homomorphism* of topological rings is a ring homomorphism that is continuous. A *topological field* is a field that is also a topological ring in such a way that division by a nonzero element is continuous.

A very natural way to equip a ring with a topology that occurs throughout mathematics is by way of an absolute value; to get started, we consider such notions first for fields. Throughout, let F be a field.

Definition 10.2.2. An *absolute value* on F is a map

$$|\,| : F \to \mathbb{R}_{\geq 0}$$

such that:

(i) $|x| = 0$ if and only if $x = 0$;

(ii) $|xy| = |x|y$ for all $x, y \in F$; and

(iii) $|x + y| \leq |x| + |y|$ for all $x, y \in F$ (*triangle inequality*).

An absolute value $|\,|$ on F gives F the structure of a topological field by the metric $d(x, y) = |x - y|$. Two absolute values $|\,|, \|\,\|$ on F are *equivalent* if there exists $c > 0$ such that $|x| = \|x\|^c$ for all $x \in F$; equivalent absolute values induces the same topology on F.

Definition 10.2.3. An absolute value is *nonarchimedean* if $|x + y| \leq \sup\{|x|, |y|\}$ for all $x, y \in F$ (*ultra metric inequality*), and *archimedean* otherwise.

Example 10.2.4. The fields $\mathbb{R}$ and $\mathbb{C}$ are topological fields with respect to the usual archimedean absolute value.

Remark 10.2.5. A field is archimedean if and only if it satisfies the *archimedean property*: for all $x \in F^\times$, there exists $n \in \mathbb{Z}$ such that $|nx| > 1$. In particular, a field F equipped with an archimedean absolute value has $\operatorname{char} F = 0$.

Example 10.2.6. Every field has the *trivial* (nonarchimedean) absolute value, defined by $|0| = 0$ and $|x| = 1$ for all $x \in F^\times$; the trivial absolute value induces the discrete topology on F.

A nonarchimedean absolute value on a field F arises naturally by way of a valuation, as follows.

Definition 10.2.7. A *valuation* of a field F is a map $v : F \to \mathbb{R} \cup \{\infty\}$ such that:

(i) $v(x) = \infty$ if and only if $x = 0$;

(ii) $v(xy) = v(x) + v(y)$ for all $x, y \in F$; and

(iii) $v(x + y) \geq \min(v(x), v(y))$ for all $x, y \in F$.

A valuation is *discrete* if the *value group* $v(F^\times)$ is discrete in $\mathbb{R}$ (has no accumulation points).

Here, we set the convention that $x + \infty = \infty + x = \infty$ for all $x \in \mathbb{R} \cup \{\infty\}$. By (ii), the value group $v(F^\times)$ is indeed a subgroup of the additive group $\mathbb{R}$, and so although an absolute value is multiplicative, a valuation is additive.

Example 10.2.8. For $p \in \mathbb{Z}$ prime, the the map $v(x) = \operatorname{ord}_p(x)$ is a valuation on $\mathbb{Q}$.

Example 10.2.9. Let k be a field and $F = k(t)$ the field of rational functions over k. For $f(t) = g(t)/h(t) \in k(t) \setminus \{0\}$, define $v(f(t)) = \deg g(t) - \deg h(t)$ and $v(0) = \infty$. Then v is a discrete valuation on F.

Given the parallels between them, it should come as no surprise that a valuation gives rise to an absolute value on F by defining $|x| = c^{-v(x)}$ for any $c > 1$; the induced topology on F is independent of the choice of c. By condition (iii), the absolute value associated to a valuation is nonarchimedean.

Example 10.2.10. The *trivial* valuation is the valuation v satisfying $v(0) = \infty$ and $v(x) = 0$ for all $x \in F^\times$. The trivial valuation gives the trivial absolute value on F.

Two valuations v, w are *equivalent* if there exists $a \in \mathbb{R}_{>0}$ such that $v(x) = aw(x)$ for all $x \in F$; equivalent valuations give the same topology on a field. A nontrivial discrete valuation is equivalent after rescaling (by the minimal positive element in the value group) to one with value group $\mathbb{Z}$, since a nontrivial discrete subgroup of $\mathbb{R}$ is cyclic; we call such a discrete valuation *normalized*.

Given a field F with a nontrivial discrete valuation v, we have the *valuation ring* $R = \{x \in F : v(x) \geq 0\}$. We have $R^\times = \{x \in F : v(x) = 0\}$ since $v(x) + v(x^{-1}) = v(1) = 0$ for all $x \in F^\times$. The valuation ring is a local ring with unique maximal ideal

$$\mathfrak{p} = \{x \in F : v(x) > 0\} = R \setminus R^\times.$$

An element $\pi \in \mathfrak{p}$ with smallest valuation is called a *uniformizer*, and comparing valuations we see that $\pi R = (\pi) = \mathfrak{p}$. Since $\mathfrak{p} \subsetneq R$ is maximal, the quotient $k = R/\mathfrak{p}$ is a field, called the *residue field* of R (or of F).

Recall that a topological space is *locally compact* if each point has a compact neighborhood.

Definition 10.2.11. A *local field* is a Hausdorff, locally compact topological field with a nondiscrete topology.

In a local field, we can hope to understand its structure by local considerations in a compact neighborhood, hence the name.

Local fields have a very simple classification as follows.

Theorem 10.2.12. *Every local field F is isomorphic as a topological field to one of the following:*

- *F is archimedean, and $F \cong \mathbb{R}$ or $F \cong \mathbb{C}$;*
- *F is nonarchimedean with $\operatorname{char} F = 0$, and F is a finite extension of $\mathbb{Q}_p$ for some prime p; or*
- *F is nonarchimedean with $\operatorname{char} F = p$, and F is a finite extension of $\mathbb{F}_p((t))$ for some prime p; in this case, there is a (non-canonical) isomorphism $F \cong \mathbb{F}_q((t))$ where q is a power of p.*

We have the following equivalent characterization of nonarchimedean local fields.

Lemma 10.2.13. *A field is a nonarchimedean local field if and only if it is complete with respect to a nontrivial discrete valuation $v : F \to \mathbb{R} \cup \{\infty\}$ with finite residue field.*

Although a local field is only locally compact, the valuation ring is itself compact, as follows.

Lemma 10.2.14. *Suppose F is nonarchimedean. Then F is totally disconnected and the valuation ring $R \subset F$ is a compact, totally disconnected topological ring.*

Proof. To see that F is totally disconnected (so too R is totally disconnected), by translation it suffices to show that the only connected set containing 0 is $\{0\}$. Let $x \in F^\times$ with $|x| = \delta > 0$. The image $|F^\times| \subseteq \mathbb{R}_{>0}$ is discrete, so there exists $0 < \epsilon < \delta$ so that $|y| < \delta$ implies $|y| \leq \delta - \epsilon$ for all $y \in F$. Thus an open ball is a closed ball

$$D(0,\delta) = \{y \in F : |y| < \delta\} = \{y \in F : |y| \leq \delta - \epsilon\} = D[0, \delta - \epsilon];$$

since $x \in F^\times$ and $\delta > 0$ were arbitrary, the only connected subset containing 0 is $\{0\}$.

Next, we show R is compact. We have a natural continuous ring homomorphism

$$\phi : R \to \prod_{n=1}^{\infty} R/\mathfrak{p}^n$$

where each factor $R/\mathfrak{p}^n$ is equipped with the discrete topology and the product is given the product topology. The map ϕ is injective, since $\bigcap_{n=1}^{\infty} \mathfrak{p}^n = \{0\}$ (every nonzero element has finite valuation). The image of ϕ is obviously closed. Therefore R is homeomorphic onto its closed image. But by Tychonoff's theorem, the product $\prod_{n=1}^{\infty} R/\mathfrak{p}^n$ of compact sets is compact, and a closed subset of a compact set is compact, so R is compact. □

One key property of local fields we will use is Hensel's lemma: it is the nonarchimedean analogue of Newton's method.

Lemma 10.2.15 (Hensel's lemma)**.** *Let F be a nonarchimedean local field with valuation v and valuation ring R, and let $f(x_1, \dots, x_n) \in R[x_1, \dots, x_n]$ with $n \geq 1$. Suppose that $a = (a_i)_i \in R^n$ satisfies*

$$k = v(f(a_1, \dots, a_n)) > 2v(f'(a_1, \dots, a_n)) \geq 0.$$

Then there exists $\widetilde{a} = (\widetilde{a}_i)_i \in R^n$ such that $f(\widetilde{a}) = 0$ and

$$\widetilde{a}_i \equiv a_i \pmod{\mathfrak{p}^k}$$

for all $i = 1, \dots, n$.

10.3 Unique division ring, first proof

We now seek to classify quaternion algebras over local fields. First, suppose F is archimedean. When $F = \mathbb{C}$, the only quaternion algebra over $\mathbb{C}$ up to isomorphism is $B \cong \mathrm{M}_2(\mathbb{C})$. When $F = \mathbb{R}$, by the theorem of Frobenius (Corollary 3.5.5), there is a unique quaternion division algebra over $\mathbb{R}$. The classification of quaternion algebras over nonarchimedean local fields is quite analogous to the classification over $\mathbb{R}$; indeed, we have the following.

Theorem 10.3.1. *Let $F \neq \mathbb{C}$ be a local field. Then there is a unique division quaternion algebra B over F up to F-algebra isomorphism.*

To prove this theorem, from the first paragraph of this section we may assume F is a nonarchimedean local field with discrete valuation v.

We approach the proof of Theorem 10.3.1 from two vantage points. In this section, we give a proof using quadratic forms (which excludes the case where $\operatorname{char} F = 2$); in the next section, we give another proof by extending the valuation (including all characteristics).

By Theorems 4.4.5 and 4.5.5, to prove Theorem 10.3.1 it is equivalent to prove the following proposition.

Proposition 10.3.2. *Let $F \neq \mathbb{C}$ be a local field. Then there is a unique nonsingular anisotropic ternary quadratic form over F up to similarity.*

So our task becomes a hands-on investigation of ternary quadratic forms over F. The theory of quadratic forms over F is linked to that over its residue field k, so we first need to examine isotropy of quadratic forms over a finite field.

Lemma 10.3.3. *A quadratic space V over a finite field with $\dim_F V \geq 3$ is isotropic.*

This statement is elementary (Exercise 10.1).

Lemma 10.3.4. *Suppose* $\operatorname{char} k \neq 2$. *Let $Q : M \to R$ be a nonsingular quadratic form over R. Then the reduction $Q \bmod \mathfrak{p} : M \otimes_R k \to k$ of Q modulo $\mathfrak{p}$ is nonsingular over k; moreover, Q is isotropic over R if and only if $Q \bmod \mathfrak{p}$ is isotropic.*

Lemma 10.3.4 is a consequence of Hensel's lemma (Lemma 10.2.15). Combining these two lemmas, we obtain the following.

Proposition 10.3.5. *Suppose* $\operatorname{char} k \neq 2$. *Let $Q : M \to R$ be a nonsingular quadratic form over R with M of rank ≥ 3. Then Q is isotropic.*

Considering valuations, we also deduce the following from Lemma 10.3.4.

Lemma 10.3.6. *Suppose* $\operatorname{char} k \neq 2$. *Then $\#F^\times/F^{\times 2} \cong (\mathbb{Z}/2\mathbb{Z})^2$ and is represented by the classes of $1, e, \pi, e\pi$ where $e \in R^\times$ is any element which reduces modulo $\mathfrak{p}$ to a nonsquare in k.*

We first consider the case $\operatorname{char} k \neq 2$.

Proof of Proposition 10.3.2 ($\operatorname{char} k \neq 2$*).* Let $Q \cong \langle a, -b, -c \rangle$ be a nonsingular, anisotropic ternary quadratic form over F. Rescaling the basis elements by a power of the uniformizer, we may assume that $v(a), v(b), v(c) \in \{0, 1\}$. Then, by the pigeonhole principle on this set of valuations, we may rescale the form and permute the basis to assume that $a = 1$ and $0 = v(b) \leq v(c)$. If $v(b) = v(c) = 0$ then the quadratic form modulo $\mathfrak{p}$ is nonsingular, so by Lemma 10.3.3 it is isotropic and by Lemma 10.3.4 we conclude Q is isotropic, a contradiction.

We are left with the case $v(b) = 0$ and $v(c) = 1$. By Lemma 10.3.6, we may assume $b = 1$ or $b = e$ where e is a nonsquare in k. If $b = 1$, then the form is obviously isotropic, so we have $b = e$. Similarly, we have $c = \pi$ or $c = e\pi$. In fact, the latter case is similar to the former: scaling by e we have

$$\langle 1, -e, -e\pi \rangle \sim \langle -1, e, -\pi \rangle$$

and since $\langle -1, e \rangle \cong \langle 1, -e \rangle$ (Exercise 10.2), we have $Q \sim \langle 1, -e, -\pi \rangle$.

To conclude, we show that the form $\langle 1, -e, -\pi \rangle$ is anisotropic. Suppose that $x^2 - ey^2 = \pi z^2$ with $x, y, z \in F^3$ not all zero. By homogeneity, we may assume $x, y, z \in R$ and at least one of $x, y, z \in R^\times$. Reducing modulo $\mathfrak{p}$ we have $x^2 \equiv ey^2 \pmod{\mathfrak{p}}$ so since e is a nonsquare we have $v(x), v(y) \geq 1$. But this implies that $v(z) = 0$ and so $v(\pi z^2) = 1 = v(x^2 - ey^2) \geq 2$, a contradiction. □

Now suppose that $\operatorname{char} k = 2$. Recall the issues with inseparability in characteristic 2 (Paragraph 5.1.2). Let $\wp(k) = \{z + z^2 : z \in k\}$ be the *Artin-Schreier group* of k. The polynomial $x^2 + x + t$ is reducible if and only if $t \in \wp(k)$, and since k is finite, we have $k/\wp(k) \cong \mathbb{Z}/2\mathbb{Z}$ (Exercise 10.3). Let $t \in R$ represent the nontrivial class in $k \setminus \wp(k)$.

Proof of Proposition 10.3.2 ($\operatorname{char} k = 2$*).* By nonsingularity and scaling, we may assume that $Q \sim [1, b] \perp \langle c \rangle$ with $b, c \in R$. If $v(b) > 0$, then $[1, b]$ is isotropic modulo $\mathfrak{p}$ and hence Q is isotropic, a contradiction. So $v(b) = 0$, and for the same reason b in the same class as $t \in k \setminus \wp(k)$. Scaling, we may assume $v(c) = 0, 1$. If $v(c) = 0$, then either c or $t + c$ belongs to $\wp(k)$ and so again we have a contradiction. Thus $v(c) = 1$ and $c = u\pi$ for some $u \in R^\times$; but then $[u, tu] \cong [1, t]$ so $Q \sim [1, t] \perp \langle \pi \rangle$. To conclude, we verify that this form is indeed anisotropic, applying the same argument as in the last paragraph in the proof when $\operatorname{char} k \neq 2$ to the quadratic form $x^2 + xy + ty^2 = \pi z^2$. □

In mixed characteristic where $\operatorname{char} F = 0$ and $\operatorname{char} F = 2$, the extension $K = F[x]/(x^2 + x + t)$ for t nontrivial in $k/\wp(k)$ is the unramified quadratic extension of F, and we can complete the square to obtain $K = F(\sqrt{e})$ with $e \in F^\times \setminus F^{\times 2}$—it is just no longer the case that e is nontrivial in $k^\times/k^{\times 2}$. Putting these cases together, we have the following corollary.

Corollary 10.3.7. *Let $F \neq \mathbb{C}$ be a local field and B be a quaternion algebra over F. Then B is a division quaternion algebra if and only if*

$$B \cong \left(\frac{K, \pi}{F}\right)$$

where K is the unramified quadratic extension of F. In particular, if $\operatorname{char} k \neq 2$*, then B is a division algebra if and only if*

$$B \cong \left(\frac{e, \pi}{F}\right), \text{ where } e \text{ is nontrivial in } k^\times/k^{\times 2}$$

and if $\operatorname{char} F = \operatorname{char} k = 2$*, then B is a division algebra if and only if*

$$B \cong \left[\frac{t, \pi}{F}\right), \text{ where } t \text{ is nontrivial in } k/\wp(k).$$

10.4 Local Hilbert symbol

Recall the definition of the Hilbert symbol (Section 4.7). In this section, we compute the Hilbert symbol over a local field F with $\operatorname{char} k \neq 2$. Let $a, b \in F^\times$.

We begin with the case where F is archimedean. If $F = \mathbb{C}$, then the Hilbert symbol is identically 1. If $F = \mathbb{R}$, then

$$(a,b)_{\mathbb{R}} = \begin{cases} 1, & \text{if } a > 0 \text{ or } b > 0; \\ -1, & \text{if } a < 0 \text{ and } b < 0. \end{cases}$$

Lemma 10.4.1. *The Hilbert symbol over a nonarchimedean local field F is bimultiplicative, i.e.*

$$(a,bc)_F = (a,b)_F(a,c)_F \quad \textit{and} \quad (ab,c)_F = (a,c)_F = (b,c)_F$$

for all $a,b,c \in F^\times$.

Remark 10.4.2. The bimultiplicativity property of the local Hilbert symbol is a special property and does not extend to a general field!

Proof. This will follow from the direct computation below (10.4.3), but it is helpful to know this fact independently.

We appeal to Theorem 4.5.5(vi): we have $(a,b)_F = 1$ if and only if $b \in \mathrm{N}_{K/F}(K^\times)$ where $K = F[x]/(x^2 - a)$. If K is not a field, then $(a,b)_F = 1$ identically, so it is certainly multiplicative. Otherwise, $F^\times / \mathrm{N}_{K/F}(K^\times) \cong \mathbb{Z}/2\mathbb{Z}$: when $\operatorname{char} k \neq 2$, this follows from Lemma 10.3.6, but it is true in general. Multiplicativity is then immediate. □

Since the Hilbert symbol is well-defined up to squares, the symbol $(a,b)_F$ is determined by the values with $a,b \in \{1, e, \pi, e\pi\}$ where e is a nonsquare in $k^\times$. Let $s = (-1)^{(\#k-1)/2}$, so that $s = 1, -1$ according as -1 is a square in k. Then we have:

$(a,b)_F$	1	e	π	$e\pi$
1	1	1	1	1
e	1	1	-1	-1
π	1	-1	s	$-s$
$e\pi$	1	-1	$-s$	s

(10.4.3)

The computation of this table is Exercise 10.7.

10.4.4. The following criteria follow from 10.4.3:

(a) If $v(ab) = 0$, then $(a,b)_F = 1$.

(b) If $v(a) = 0$ and $v(b) = v(\pi)$, then

$$(a,b)_F = \left(\frac{a}{\pi}\right) = \begin{cases} 1 & \text{if } a \in k^{\times 2}; \\ -1 & \text{if } a \in k^\times \setminus k^{\times 2}. \end{cases}$$

10.4.5. The computation of the Hilbert symbol for local fields with $\operatorname{char} F \neq 2$ but $\operatorname{char} k = 2$ is significantly more involved. We provide in Exercise 10.11 a way to understand this symbol for a general F. In this paragraph, we compute the Hilbert symbol for $F = \mathbb{Q}_2$.

To begin, the group $\mathbb{Q}_2^\times/\mathbb{Q}_2^{\times 2}$ is generated by $-1, -3, 2$, so representatives are $\{\pm 1, \pm 3, \pm 2, \pm 6\}$. The extension $\mathbb{Q}_2(\sqrt{-3}) \supset \mathbb{Q}_2$ is the unique unramified extension.

We recall Hilbert's criterion: $(a, b)_F = 1$ if and only if $ax^2 + by^2 = 1$ has a solution with $x, y \in F$.

If $a, b \in \mathbb{Z}$ are odd, then

$$\begin{aligned} ax^2 + by^2 = z^2 &\text{ has a nontrivial solution in } \mathbb{Q}_2 \\ \iff \quad & a \equiv 1 \pmod 4 \text{ or } b \equiv 1 \pmod 4; \end{aligned}$$

by homogeneity and Hensel's lemma, it is enough to check for a solution modulo 4. This deals with all of the symbols with a, b odd.

By the determination above, we see that $(-3, b) = -1$ for $b = \pm 2, \pm 6$ and $(2, 2)_{\mathbb{Q}_2} = (-1, 2)_{\mathbb{Q}_2} = 1$ the latter by Hilbert's criterion, as $-1 + 2 = 1$; knowing multiplicativity (Lemma 10.4.1), we have uniquely determined all Hilbert symbols. It is still useful to compute several of these symbols individually, in the same manner as (10.4.5) (working modulo 8): see Exercise 10.10. We summarize the results here:

$(a,b)_{\mathbb{Q}_2}$	1	-3	-1	3	2	-6	-2	6
1	1	1	1	1	1	1	1	1
-3	1	1	1	1	-1	-1	-1	-1
-1	1	1	-1	-1	1	1	-1	-1
3	1	1	-1	-1	-1	-1	1	1
2	1	-1	1	-1	1	-1	1	-1
-6	1	-1	1	-1	-1	1	-1	1
-2	1	-1	-1	1	1	-1	-1	1
6	1	-1	-1	1	-1	1	1	-1

(10.4.6)

[[Unramified square symbol.]]

10.5 Unique division ring, second proof

We now proceed to give a second proof of Theorem 10.3.1; we will extend the valuation v to one uniquely on a division quaternion algebra. For this, we will need to rely a bit more heavily on the theory of local fields. We retain our assumption that F is a nonarchimedean local field with valuation ring R, residue field k, and maximal ideal $\mathfrak{p}$ generated by a uniformizer π.

Let $K \supseteq F$ be a finite extension of fields. Then there exists a unique valuation w on K such that $w|_F = v$, and we say that w *extends* v: this valuation is defined by

$$w(x) = \frac{v(\mathrm{N}_{K/F}(x))}{[K:F]}; \tag{10.5.1}$$

in particular, K is also a nonarchimedean local field. (The only nontrivial thing to check is condition (iii), and this can be derived from the fact that

$$v(\mathrm{N}_{K/F}(x)) \geq 0 \Rightarrow v(\mathrm{N}_{K/F}(x+1)) \geq 0$$

for $x \in K$ and this follows by a direct examination of the minimal polynomial of x.)

Lemma 10.5.2. *The integral closure of R in K is the valuation ring $S = \{x \in K : w(x) \geq 0\}$, and S is an R-order in K.*

A finite extension K/F is *unramified* if a uniformizer π for F is also a uniformizer for K. There is a unique unramified extension of F of any degree $f \in \mathbb{Z}_{\geq 1}$ and such a field corresponds to the unique extension of the residue field k of degree f. In an unramified extension K/F of degree $[K:F] = f$, we have $N_{K/F}(K^\times) = R^\times \pi^{f\mathbb{Z}}$, so $b \in \mathrm{N}_{K/F}(K^\times)$ if and only if $f \mid v(b)$.

10.5.3. If $\operatorname{char} k \neq 2$, then by Hensel's lemma, the unramified extension of degree 2 is given by adjoining a square root of the unique nontrivial class in $k^\times/k^{\times 2}$; if $\operatorname{char} k = 2$, then the unramified extension of degree 2 is given by adjoining a root of the polynomial $x^2 + x + t$ where t is a nontrivial class in the Artin-Schreier group $k/\wp(k)$.

Let $K \supseteq F$ be a finite separable extension of fields. We say K/F with $e = [K:F]$ is *totally ramified* if a uniformizer π_K has the property that π_K^e is a uniformizer for F. For any finite separable extension K/F of degree n, there is a (unique) maximal unramified subextension K_{un}/F, and the extension K/K_{un} is totally ramified.

$$\begin{tikzcd} & & K \\ & K_{\mathrm{un}} \arrow[ur, dash, "e"] & \\ F \arrow[ur, dash, "f"] & & \end{tikzcd}$$

We say that $e = [K : K_{\mathrm{un}}]$ is the *ramification degree* and $f = [K_{\mathrm{un}} : F]$ the *inertial degree*, and we have the fundamental equality

$$n = [K:F] = ef. \tag{10.5.4}$$

We now seek to generalize these theorems to the noncommutative case. Let D be a central simple *division* algebra over F with $\dim_F D = [D : F] = n^2$. We extend the valuation v to a map

$$\begin{aligned} w : D &\to \mathbb{R} \cup \{\infty\} \\ \alpha &\mapsto \frac{v(\mathrm{N}_{D/F}(\alpha))}{[D : F]} = \frac{v(\mathrm{nrd}(\alpha))}{n}, \end{aligned}$$

where the equality follows from the fact that $\mathrm{N}_{D/F}(\alpha) = \mathrm{nrd}(\alpha)^n$ (see Section 6.7).

Lemma 10.5.5. *The map w defines a valuation on D, i.e., the following hold:*

(i) $w(\alpha) = \infty$ *if and only if* $\alpha = 0$.

(ii) $w(\alpha\beta) = w(\alpha) + w(\beta) = w(\beta\alpha)$ *for all* $\alpha, \beta \in D$.

(iii) $w(\alpha + \beta) \geq \min(w(\alpha), w(\beta))$ *for all* $\alpha, \beta \in D$.

(iv) $w(D^\times)$ *is discrete in* $\mathbb{R}$.

Proof. Statement (i) is clear (note it already uses that D is a division ring). Statement (ii) follows from the multiplicativity of nrd and v. To prove (iii), we may assume $\beta \neq 0$ and so $\beta \in D^\times$. We have $w(\alpha+\beta) = w(\beta) + w(\alpha\beta^{-1} + 1)$. But the restriction of w to $F(\alpha\beta^{-1})$ is a discrete valuation, so $w(\alpha\beta^{-1} + 1) \geq \min(w(\alpha\beta^{-1}), w(1))$ so by (ii) $w(\alpha+\beta) \geq \min(w(\alpha), w(\beta))$, as desired. Finally, (iv) holds since $w(D^\times) \subseteq v(F^\times)/n$ and the latter is discrete. □

From Lemma 10.5.5, we say that w is a *discrete valuation* on D since it satisfies the same axioms as for a field. It follows from Lemma 10.5.5 that the set

$$\mathcal{O} = \{\alpha \in D : w(\alpha) \geq 0\}$$

is a ring, called the *valuation ring* of D.

Proposition 10.5.6. *$\mathcal{O}$ is the unique maximal R-order in D, consisting of all elements of D that are integral over R.*

Proof. First, we prove that

$$\mathcal{O} = \{\alpha \in D : \alpha \text{ is integral over } R\}.$$

In one direction, suppose $\alpha \in D$ is integral over R. Since R is integrally closed, by Lemma 8.3.10 the coefficients of the minimal polynomial $f(x) \in F[x]$ of α belong to R. Since D is a division ring, $f(x)$ is irreducible and hence the reduced characteristic polynomial $g(x)$ is a power of $f(x)$ and thus has coefficients in R. The reduced norm is the constant coefficient of $g(x)$, so $\alpha \in \mathcal{O}$.

Now suppose $\alpha \in \mathcal{O}$, so that $w(\alpha) \geq 0$, and let $K = F(\alpha)$. Let $f(x) \in F[x]$ be the minimal polynomial of α. We want to conclude that $f(x) \in R[x]$ knowing that $w(\alpha) \geq 0$. But the restriction of w to K is the unique extension of v to K, and so this is a statement about the extension K/F of local fields and therefore follows from the theory in the commutative case. For completeness, we give the proof. Let L be a splitting field of $f(x)$ containing K. Then v extends to a unique valuation w_L on L. At the same time, the norm w on D restricts to a discrete valuation on K and hence by equivalence of valuations, we have $w_L(\alpha) \geq 0$. But now if $f(x) = \prod_{i=1}^{n}(x - \alpha_i) = x^n + \dots + a_0 \in F[x]$ with $\alpha_i \in L$, then $w_L(\alpha_i) = a_0 = w(\alpha) \geq 0$. Thus the coefficients of f (symmetric functions in the α_i) belong to R, and so α is integral over R.

We can now prove that $\mathcal{O}$ is an R-order. Scaling any element $\alpha \in D^\times$ by an appropriate power of π gives it positive valuation, so $\mathcal{O}F = D$. So to conclude we must show that $\mathcal{O}$ is finitely generated as an R-module. For this purpose, we may assume that D is central over F, since the center $K = Z(D)$ is a field extension of F of finite degree and the integral closure of R in K is finitely generated as an R-module (Lemma 10.5.2). A central division algebra is separable, so we may apply Lemma 8.4.1: every $\alpha \in \mathcal{O}$ is integral over R and $\mathcal{O}$ is a ring, so the lemma implies that $\mathcal{O}$ is an R-order.

Finally, it follows immediately that $\mathcal{O}$ is a maximal R-order: by Corollary 8.3.9, every element of an R-order is integral over R, and $\mathcal{O}$ contains all such elements. □

Remark 10.5.7. For a quaternion division algebra D, we can argue more directly in the proof of Proposition 10.5.6 using the reduced norm: see Exercise 10.14.

It follows from Proposition 10.5.6 that $\mathcal{O}$ is a finitely generated R-submodule of D. But R is a PID (every ideal is a power of the maximal ideal $\mathfrak{p}$) so in fact $\mathcal{O}$ is free of rank $[D : F]$ over R. We have

$$\mathcal{O}^\times = \{\alpha \in D : w(\alpha) = 0\} \tag{10.5.8}$$

since $w(\alpha^{-1}) = -w(\alpha)$ and $\alpha \in \mathcal{O}^\times$ if and only if $\operatorname{nrd}(\alpha) \in R^\times$. Consequently,

$$P = \{\alpha \in D : w(\alpha) > 0\} = \mathcal{O} \setminus \mathcal{O}^\times$$

is the unique maximal two-sided ideal of $\mathcal{O}$. Therefore $\mathcal{O}$ is a *noncommutative local ring*, a ring with a unique maximal left (or right) ideal.

We are now prepared to give the second proof of the main result in this chapter (Theorem 10.3.1). By way of analogy, we consider the commutative case: for an extension L of F of degree $[L : F] = n$, we have a ramification degree e and an inertial degree f with $ef = n$ (10.5.4). The same will be true when B is a division quaternion algebra: we will show that $\mathcal{O}/P$ is a quadratic field extension of k and hence that B contains an unramified separable quadratic extension K of F (extending the analogy, that $f = 2$); and then computing with valuations we will conclude that $P^2 = \pi\mathcal{O}$ (and $e = 2$) from which the result follows.

Proof of Theorem 10.3.1. Suppose that $v : F \to \mathbb{Z}_{\geq 0} \cup \{\infty\}$ is normalized and let $j \in P$ have minimal (positive) valuation $w(j) \geq 1/2$. Then for any $0 \neq \alpha \in P$ we have $w(\alpha j^{-1}) = w(\alpha) - w(j) \geq 0$ so $\alpha j^{-1} \in \mathcal{O}$ so $\alpha \in \mathcal{O}j$. Thus $P = \mathcal{O}j = j\mathcal{O} = \mathcal{O}j\mathcal{O}$, since P is a two-sided ideal. Arguing in the same way, since

$$w(j) \leq w(\pi) = v(\pi^2) = 1 \leq w(j^2),$$

we conclude that $P \supseteq \pi\mathcal{O} \supseteq P^2 = j^2\mathcal{O}$. The map $\alpha \mapsto \alpha j$ yields an isomorphism $\mathcal{O}/P \xrightarrow{\sim} P/P^2$ of k-vector spaces, so

$$4 = \dim_k \mathcal{O}/\pi\mathcal{O} \leq \dim_k \mathcal{O}/P + \dim_k P/P^2 = 2\dim_k \mathcal{O}/P \tag{10.5.9}$$

and thus $\dim_k \mathcal{O}/P \geq 2$; in particular, $\mathcal{O}/P \neq k$.

Since $\mathcal{O} \setminus P = \{x \in \mathcal{O} : w(x) = 0\} = \mathcal{O}^\times$, the ring $\mathcal{O}/P$ is a division algebra over k and hence a finite division ring. But then by Wedderburn's theorem (Exercises 3.12 and 5.11, or Exercise 6.16), the ring $\mathcal{O}/P$ is a field. Thus, there exists $i \in \mathcal{O}$ such that $\mathcal{O}/P = k(i)$, since k is a finite field. Then $K = F(i)$ is an unramified separable quadratic extension of F and consequently $\dim_k \mathcal{O}/P = f = 2$. Therefore equality holds in (10.5.9), and so $P^2 = \pi\mathcal{O}$.

By Exercise 6.13, there exists $b \in F^\times$ such that $B \cong \left(\dfrac{K, b}{F}\right)$. But B is a division ring if and only if $b \in F^\times \setminus N_{K/F}(K^\times)$ by Theorems 4.5.5 and 5.3.8. Since K/F is unramified we have $N_{K/F}(K^\times) = R^\times\pi^{2\mathbb{Z}}$, so we may take $b = \pi$ (Exercise 5.3) and $B \cong \left(\dfrac{K, \pi}{F}\right)$ is the unique division quaternion algebra over F. □

Corollary 10.5.10. *If F is nonarchimedean and $B = \left(\dfrac{K, b}{F}\right)$, then $B \cong \mathrm{M}_2(F)$ if K is unramified over F (so v is split or inert in K) and $v(b) = 0$.*

Remark 10.5.11. Let $B \cong \left(\dfrac{K, \pi}{F}\right)$ be a division quaternion algebra over F, so that K is a separable quadratic subfield and $j^2 = \pi$. As above, in analogy with the case of field extensions, we define the *ramification degree* of B over F as $e(B/F) = 2$ and the *inertial degree* of B over F as $f(B/F) = 2$ and note the equality $e(B/F)f(B/F) = 4 = [B : F]$, as in the commutative case. This equality carries over more generally to division algebras; see Exercise 10.17.

10.6 Topology

In this section, we conclude with some discussion about the topology of algebras over local fields.

Let F be a local field. Then F is locally compact (by definition) but is not itself compact. The subgroup $F^\times = F \setminus \{0\}$ is equipped the subspace topology; it is open in F so $F^\times$ is locally compact—this is quite visible when $F = \mathbb{R}, \mathbb{C}$ is archimedean. If F is nonarchimedean, with valuation ring R and valuation v, then $F^\times$ is totally disconnected and further $R^\times = \{x \in R : v(x) = 0\} \subset R$ is closed so is a topological abelian group that is compact (and totally disconnected).

Now let B be a finite-dimensional F-algebra. Then as a vector space over F, it has a unique topology compatible with the topology on F, as any two norms on a topological vector F-space (extending the norm on F) are equivalent (the sup norm is equivalent to the sum of squares norm, etc.). Two elements are close in the topology on B if and only if their coefficients are close with respect to a (fixed) basis: for example, two matrices in $\mathrm{M}_n(F)$ are close if and only if all of their coordinate entries are close. Consequently, B is locally compact topological ring (take a compact neighborhood in each coordinate). It is also true that $B^\times$ is a locally compact topological group: the norm $\mathrm{N}_{B/F} : B^\times \to F^\times$ is a continuous map, so $B^\times$ is open in B, and an open subset of a Hausdorff, locally compact space is locally compact in the subspace topology.

Example 10.6.1. If $B = \mathrm{M}_n(F)$, then $B^\times = \mathrm{GL}_n(F)$ is locally compact: any closed, bounded neighborhood that avoids the locus of matrices with determinant 0 is a compact neighborhood. When F is archimedean, this is quite visual: any matrix of nonzero determinant is at some finite distance away from the determinant zero locus! Note however that $\mathrm{GL}_n(F)$ is not itself compact since $F^\times = \mathrm{GL}_1(F)$ is not compact.

Now suppose F is nonarchimedean with valuation v and valuation ring R. Then R is the maximal compact subring of F. Indeed, $x \in F$ lies in a compact subring if and only if $v(x) \geq 0$ if and only if x is integral over R. The only new implication here is the statement that if $v(x) < 0$ then x does not lie in a compact subring, and that is because the sequence $x_n = x^n$ does not have a convergent subsequence as $|x_n| \to \infty$.

Next, let $\mathcal{O}$ be an R-order in B. Then $\mathcal{O} \cong R^n$ is a free R-module of finite rank. Choosing a basis, the above argument shows that $\mathcal{O}$ is compact as the Cartesian power of a compact set. The group $\mathcal{O}^\times$ is therefore also compact because it is closed: for $\gamma \in \mathcal{O}$, we have $\gamma \in \mathcal{O}^\times$ if and only if $\mathrm{N}_{B/F}(\gamma) \in R^\times \subset R$ and $R^\times = \{x \in R : v(x) = 0\} \subseteq R$ is closed.

Example 10.6.2. For $R = \mathbb{Z}_p \subseteq F = \mathbb{Q}_p$ and $B = \mathrm{M}_n(\mathbb{Q}_p)$, the order $\mathcal{O} = \mathrm{M}_n(\mathbb{Z}_p)$ is compact (neighborhoods can be taken coordinatewise) and the subgroup $\mathcal{O}^\times = \mathrm{GL}_n(\mathbb{Z}_p)$ is compact: there is no way to run off to infinity, either in a single coordinate or via the determinant.

10.6.3. Suppose $B = D$ is a division ring. Then the valuation ring $\mathcal{O}$ is the maximal compact subring of B, for the same reason as in the commutative case. In this situation, the unit group $\mathcal{O}^\times$ is a pro-solvable group! We have a filtration $\mathcal{O} \supset P \supset P^2 \supset \dots$ giving rise to a filtration

$$\mathcal{O}^\times \supset 1 + P \supset 1 + P^2 \supset \dots .$$

As in the second proof of Theorem 10.3.1, the quotient $\mathcal{O}/P$ is a finite extension of the finite residue field k, so $(\mathcal{O}/P)^\times$ is a cyclic abelian group. The maximal two-sided ideal P is principal, generated by an element j of minimal valuation, and multiplication by j^n gives an isomorphism $\mathcal{O}/P \xrightarrow{\sim} P^n/P^{n+1}$ of k-vector spaces (or abelian groups) for all $n \geq 1$.

Furthermore, for each $n \geq 1$, we have an isomorphism of groups

$$\begin{aligned} P^n/P^{n+1} &\xrightarrow{\sim} (1 + P^n)/(1 + P^{n+1}) \\ \alpha &\mapsto 1 + \alpha. \end{aligned} \tag{10.6.4}$$

Therefore, $\mathcal{O}^\times = \varprojlim_n (\mathcal{O}/P^n)^\times$ is an inverse limit of solvable groups.

10.6.5. We will also want to consider norm 1 groups; for this, we assume that B is a semisimple algebra. Let

$$B_1^\times = \{\alpha \in B : \mathrm{nrd}(\alpha) = 1\}.$$

Then $B_1^\times$ is a closed subgroup of $B^\times$, since the reduced norm is a continuous function. If B is a divison ring and F is nonarchimedean, then B has a valuation ring $\mathcal{O}$, and $B_1^\times = \mathcal{O}_1^\times$ is compact. If B is a division ring and F is archimedean, then $B \cong \mathbb{H}$ and $B_1^\times \cong \mathbb{H}_1^\times \cong \mathrm{SU}(2)$ is compact (it is identified with the 3-sphere in $\mathbb{R}^4$). Finally, if B is not a division ring, then either B is the product of two algebras or B is a matrix ring over a division ring, and correspondingly B is not compact by considering the subgroup $(\pi, 1/\pi)$ or a unipotent subgroup.

10.7 Splitting fields

Proposition 10.7.1. *Let B be a division quaternion ring over F, and let L be a separable field extension of F of finite degree. Then L is a splitting field for B if and only if $[L : F]$ is even.*

Proof. We have $B \cong \left(\dfrac{K,\pi}{F}\right)$ where K is the unramified quadratic extension of F. Let e, f be the ramification index and inertial degree of L. Then $[L:F] = n = ef$, so n is even if and only if e is even or f is even.

But f is even if and only if L contains an unramified quadratic subextension, necessarily isomorphic to K; but then K splits B so L splits B.

Otherwise, L is linearly disjoint from K so $K \otimes_F L = KL$ is the unramified quadratic extension of L. Therefore $B \otimes_F L \cong \left(\dfrac{KL,\pi}{L}\right)$. Let R_L be the valuation ring of L and let π_L be a uniformizer for L. Then $N_{KL/L}(KL^\times) = R_L^\times \pi_L^{2\mathbb{Z}}$. If L/F has ramification index e, then $\pi = u\pi_L^e$ for some $u \in R_L^\times$. Putting these together, we see that $B \otimes_F L$ is a division ring if and only if π is a norm from KL if and only if e is even. $\square$

Corollary 10.7.2. *If K/F is a separable quadratic field extension, then $K \hookrightarrow B$.*

In other words, B contains every separable quadratic extension of F!

10.8 Extensions and further reading

10.8.1. The p-adic numbers were developed by Hensel. In the early 1920s, Hasse used them in the study of quadratic forms and algebras over number fields. At the time, what is now called the "local–global principle" then was called the the p-adic transfer from the "small" to the "large".

10.8.2. Theory of local fields: [Neu99, Corollary V.1.2]. serre.

10.8.3. Weil started this game in his basic number theory.

10.8.4. Theory of local division rings more generally and noncommutative local rings.

10.9 Algorithmic aspects

Computing the Hilbert symbol

Exercises

10.1. Let k be a finite field and let $Q : V \to k$ be a ternary quadratic form. Show that q is isotropic. *[Hint: Reduce to the case of finding a solution to $y^2 = f(x)$ where f is a polynomial of degree 2. Then only the case #k odd remains; show* *that f takes on at least $(q+1)/2$ values in k but there are at most $(q-1)/2$ nonsquares in k.]*

Conclude again that there is no division quaternion ring over a finite field k.

10.2. Let k be a finite field with $\operatorname{char} k \neq 2$ and let $e \in k^\times$. Show that there is an isometry $\langle -1, e\rangle \cong \langle 1, -e\rangle$.

10.3. Let k be a finite field with even cardinality. Show that $\#k/\wp(k) = 2$, where $\wp(k)$ is the Artin-Schreier group.

10.4. Let $F \neq \mathbb{C}$ be a local field and let Q be a nonsingular ternary quadratic form over F. Show that Q is isotropic over any quadratic field extension of F.

10.5. Let B be a division quaternion algebra over a nonarchimedean local field F. Give another proof that the unramified quadratic extension K of F embeds in B as follows.

Suppose it does not: then for all $x \in \mathcal{O}$, the extension $F(x)/F$ is ramified, so there exists $a \in R$ such that $x - a \in P \cap K(x)$; then write $x = x_0 = a + jx_1$, where $P = j\mathcal{O}$, and iterate to conclude that $x = \sum_{n=0}^{\infty} a_n j^n$ with $a_n \in R$. But $F(j)$ is complete so $\mathcal{O} \subseteq F(j)$, a contradiction.

10.6. Let B be a division quaternion algebra over the nonarchimedean local field F.

a) Show that B is a complete, locally compact topological ring and that $\mathcal{O}$ is the maximal compact subring of B.

b) Show that $\mathcal{O}^\times$ and $B^\times/F^\times$ are compact topological groups.

c) Conclude that the smooth, irreducible complex representations of $B^\times$ are finite dimensional, and compare this with the alternative $B \cong \mathrm{M}_2(F)$.

10.7. Show that the table of Hilbert symbols (10.4.3) is correct.

10.8. Let F be a local field and K the unramified quadratic extension of F. Compute the K-left regular representation of a division quaternion algebra B over F (2.2.8) and identify the maximal order R and the maximal ideal P.

10.9. Prove a descent for the Hilbert symbol, as follows. Let K be a finite extension of the local field F with $\operatorname{char} F \neq 2$ and let $a, b \in F^\times$. Show that $(a,b)_K = (a, N_{K/F}(b))_F$.

10.10. Show that the table of Hilbert symbols (10.4.6) is correct by considering the equation $ax^2 + by^2 \equiv 1 \pmod{8}$.

10.11. [[Alternative method for computing the Hilbert symbol.]]

10.12. One can package Paragraph 10.4.4 together with multiplying by squares to prove the following more general criterion. For $a, b \in F^\times$, we have

$$(a,b)_F = (-1)^{v(a)v(b)(q-1)/2} \left(\frac{\overline{a}}{\pi}\right)^{v(b)} \left(\frac{\overline{b}}{\pi}\right)^{v(a)}$$

where $a = a_0\pi^{v(a)}$ and $b = b_0\pi^{v(b)}$ (and $v(\pi) = 1$).

10.13. Consider $B = \left(\dfrac{-1,-1}{\mathbb{Q}_2}\right)$ and let $\mathcal{O} = \mathbb{Z}_2 \oplus \mathbb{Z}_2 i \oplus \mathbb{Z}_2 j \oplus \mathbb{Z}_2(1+i+j+ij)/2$. Show that B is a division ring. Give an explicit formula for the discrete valuation w on B (extending the valuation v on $\mathbb{Q}_2$) and prove that $\mathcal{O}$ is its valuation ring.

10.14. Let B be a division quaternion algebra over F. Show that $\alpha \in B$ is integral over R if and only if $\operatorname{nrd}(x), \operatorname{nrd}(x+1) \in R$ if and only if $w(x), w(x+1) \geq 0$, where w is the valuation on B.

10.15. Let B be a division quaternion algebra over a nonarchimedean local field F, and let $\mathcal{O}$ be the valuation ring. Show that every one-sided (left or right) ideal of $\mathcal{O}$ is a power of the maximal ideal J and hence is two-sided.

10.16. Let F be a nonarchimedean local field, let $B = \mathrm{M}_2(F)$ and $\mathcal{O} = \mathrm{M}_2(R)$. Show that there are $q+1$ right $\mathcal{O}$-ideals of norm $\mathfrak{p}$ corresponding to the elements of $\mathbb{P}^1(k)$ or to the lines in k^2.

10.17. Let D be a finite-dimensional division algebra over a nonarchimedean local field F of degree $[D : F] = n^2$ with valuation ring $\mathcal{O}$ and two-sided ideal P. Show that $\mathcal{O}/P$ is finite extension of k of degree n and $J^n = \mathcal{O}\pi\mathcal{O}$.

10.18. Show that (10.6.4) is an isomorphism of (abelian) groups.

Chapter 11

Quaternion algebras over global fields

In this chapter, we discuss quaternion algebras over global fields and characterize them up to isomorphism.

11.1 Ramification

To motivate the classification of quaternion algebras over $\mathbb{Q}$, we consider by analogy a classification of quadratic fields. For this purpose, we restrict to the following class.

Definition 11.1.1. A quadratic field $F = \mathbb{Q}(\sqrt{d})$ of discriminant $d \in \mathbb{Z}$ is *mildly ramified* if $8 \nmid d$.

A quadratic field F is mildly ramified if and only if $F = \mathbb{Q}(\sqrt{m})$ where $m \neq 1$ is odd and squarefree; then $d = m$ or $d = 4m$ according as $m \equiv 1, 3 \pmod 4$.

Let $F = \mathbb{Q}(\sqrt{d})$ be a mildly ramified quadratic field of discriminant $d \in \mathbb{Z}$ and let R be its ring of integers. The primes p that ramify in F, so $pR = \mathfrak{p}^2$ for a prime ideal $\mathfrak{p} \subset R$, are precisely those with $p \mid d$.

But a discriminant d can be either positive or negative; to put this bit of data on the same footing, we define the set of *places* of $\mathbb{Q}$ to be the primes together with the symbol ∞, and we make the convention that ∞ ramifies in F if $d < 0$ and is unramified if $d > 0$. This convention is sensible, because when $d < 0$ we have only one way to embed $\mathbb{Q}(\sqrt{d}) \hookrightarrow \mathbb{C}$ up to complex conjugation—only one place above ∞ in F, so ramified—whereas there are two essentially distinct ways to embed $\mathbb{Q}(\sqrt{d}) \hookrightarrow \mathbb{R}$ when $d > 0$ (two places above ∞ in F).

Let $F = \mathbb{Q}(\sqrt{d})$ be a mildly ramified quadratic field, and let $\operatorname{Ram}(F)$ be the set of places that ramify in F. The set $\operatorname{Ram}(F)$ determines F up to isomorphism, since the discriminant of F is the product of the odd primes in $\operatorname{Ram}(F)$, multiplied by 4 if $2 \in \operatorname{Ram}(F)$ and by -1 if $\infty \in \operatorname{Ram}(F)$. (For bookkeeping reasons, in this context it would probably therefore be better to consider 4 and -1 as primes, but we will resist the inducement here.) However, not every finite set of places Σ occurs: the product d corresponding to Σ is a discriminant if and only if $d \equiv 0, 1 \pmod 4$. We call this a *parity condition* on the set of ramifying places of a mildly ramified quadratic field:

$$2 \in \Sigma \iff \text{there are an } \textit{odd} \text{ number of primes } p \in \Sigma \text{ with } p \equiv -1 \pmod 4$$

(with the convention that ∞ is congruent to $-1 \pmod 4$).

Note that if Σ is a finite subset of places of $\mathbb{Q}$ and $2 \notin \Sigma$, then precisely one of either Σ or $\Sigma \cup \{\infty\}$ satisfies the parity condition; accordingly, if we define $m(\Sigma)$ to be the product of all odd primes in Σ multiplied by -1 if $\infty \in \Sigma$, then we can recover Σ from $m(\Sigma)$.

We have proven the following result.

Lemma 11.1.2. *The maps $F \mapsto \operatorname{Ram}(F)$ and $\Sigma \mapsto m(\Sigma)$ furnishes a bijection*

$$\left\{ \begin{array}{c} \textit{Mildly ramified quadratic fields} \\ \mathbb{Q}(\sqrt{d}) \textit{ up to isomorphism} \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} \textit{Finite subsets of places of } \mathbb{Q} \\ \textit{satisfying the parity condition} \end{array} \right\}$$
$$\longleftrightarrow \{ \textit{Squarefree odd integers } m \neq 1 \}.$$

This classification procedure using sets of ramifying primes and discriminants works as well for quaternion algebras over $\mathbb{Q}$. Let B be a quaternion algebra over $\mathbb{Q}$. When is a prime p ramified in B? In Chapter 10, we saw that the completion $B_p = B \otimes_{\mathbb{Q}} \mathbb{Q}_p$ is either a division ring or the matrix ring $\mathrm{M}_2(\mathbb{Q}_p)$. Further, when B_p is a division ring, the valuation ring $\mathcal{O}_p \subset B_p$ is the unique maximal order, and the unique maximal ideal $P_p \subset \mathcal{O}_p$ satisfies $p\mathcal{O}_p = P_p^2$. So by analogy with the quadratic case, we say that a place v is *ramified* in B if the completion B_v is a division ring, and otherwise v is *unramified* (or *split*).

There are only finitely many places where $B = \left(\dfrac{a, b}{\mathbb{Q}}\right)$ is ramified: by the calculation of the Hilbert symbol (Paragraph 10.4.4), if $p \nmid 2ab$ is prime, then $(a, b)_{\mathbb{Q}_p} = 1$ and p is split in B. Therefore $\#\operatorname{Ram}(B) < \infty$.

Let $\operatorname{Ram}(B)$ be the set of ramified places of B. Not every finite subset Σ of places can occur as $\operatorname{Ram}(B)$ for a quaternion algebra B. It turns out that the *parity condition* here is that we must have $\#\Sigma$ even. So again, if Σ is a finite set of primes, then precisely one of either Σ or $\Sigma \cup \{\infty\}$ can occur as $\operatorname{Ram}(B)$. We define the *discriminant* of B to be the product $\operatorname{disc}(B)$ of primes that ramify in B, so $\operatorname{disc}(B)$ a squarefree positive integer.

The main result of this chapter, specialized to the case $F = \mathbb{Q}$, is then the following.

Theorem 11.1.3. *The maps $B \mapsto \operatorname{Ram}(B)$ and $\Sigma \mapsto \prod_{p \in \Sigma} p$ furnish bijections*

$$\left\{ \begin{matrix} \text{Quaternion algebras over } \mathbb{Q} \\ \text{up to isomorphism} \end{matrix} \right\} \longleftrightarrow \left\{ \begin{matrix} \text{Finite subsets of places of } \mathbb{Q} \\ \text{of even cardinality} \end{matrix} \right\}$$
$$\longleftrightarrow \{ d \in \mathbb{Z}_{>0} \text{ squarefree} \}.$$

The composition of these maps is $B \mapsto \prod_{p \in \operatorname{Ram}(B)} p = \operatorname{disc}(B)$.

Having stated Theorem 11.1.3, we will spend the next two sections giving a self-contained proof, assuming two statements from basic number theory (quadratic reciprocity, and primes in arithmetic progression). Although the proofs presented do not generalize to an arbitrary global field, the argument is simple enough and its structure is good motivation for the more involved treatment in the chapter ahead.

11.2 Hilbert reciprocity over the rationals

To begin, we look into the parity condition: it has a simple reformulation in terms of the Hilbert symbol (Section 4.7). For a place v of $\mathbb{Q}$, let $\mathbb{Q}_v$ denote the completion of $\mathbb{Q}$ at the absolute value associated to v: if $v = p$ is prime, then $\mathbb{Q}_v = \mathbb{Q}_p$ is the field of p-adic numbers; if $v = \infty$ is the real place, then $\mathbb{Q}_v = \mathbb{R}$. For $a, b \in \mathbb{Q}^\times$, we abbreviate $(a,b)_{\mathbb{Q}_v} = (a,b)_v$.

Proposition 11.2.1 (Hilbert reciprocity)**.** *For all $a, b \in \mathbb{Q}^\times$, we have*

$$\prod_v (a,b)_v = 1, \tag{11.2.2}$$

the product taken over all places v of $\mathbb{Q}$.

The product (11.2.2) is well-defined, because we just saw that for all primes $p \nmid 2ab$, we have $(a,b)_p = 1$. The following corollary is then immediate.

Corollary 11.2.3. *Let B be a quaternion algebra over $\mathbb{Q}$. Then the set $\operatorname{Ram}(B)$ is finite of even cardinality.*

The law of Hilbert reciprocity, as it turns out, is a core premise in number theory: it is *equivalent* to the law of *quadratic reciprocity*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}} \tag{11.2.4}$$

for odd primes p, q together with the *supplement*

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}} \quad \text{and} \quad \left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}} \tag{11.2.5}$$

for odd primes p.

We now give a proof of Hilbert reciprocity (Proposition 11.2.1), assuming the law of quadratic reciprocity and its supplement.

Proof of Proposition 11.2.1. Since the Hilbert symbol is bilinear, it suffices to prove the statement when $a, b \in \mathbb{Z}$ are equal to either -1 or a prime number. The Hilbert symbol is also symmetric, so we may interchange a, b.

If $a = b = -1$, then $B = \left(\dfrac{a, b}{\mathbb{Q}}\right) = \left(\dfrac{-1, -1}{\mathbb{Q}}\right)$ is the rational Hamiltonians, and we have $(-1, -1)_\infty = (-1, -1)_2 = -1$ and $(-1, -1)_v = 1$ if $v \neq 2, \infty$, by the computation of the even Hilbert symbol (Paragraph 10.4.5). Similarly, the cases with $a = -1, 2$ follow from the supplement (11.2.5), and are requested in Exercise 11.1.

So we may suppose $a = p$ and $b = q$ are primes. If $p = q$ then $\left(\dfrac{p, p}{\mathbb{Q}}\right) \cong \left(\dfrac{-1, p}{\mathbb{Q}}\right)$ and we reduce to the previous case, so we may suppose $p \neq q$. Since $p, q > 0$, we have $(p, q)_\infty = 1$. We have $(p, q)_\ell = 1$ for all primes $\ell \nmid 2pq$. We have

$$(p, q)_p = (q, p)_p = \left(\frac{q}{p}\right) \quad \text{and} \quad (p, q)_q = \left(\frac{p}{q}\right)$$

by Paragraph 10.4.4. Finally, we have

$$(p, q)_2 = -1 \text{ if and only if } p, q \equiv 3 \pmod 4$$

i.e., $(p, q)_2 = (-1)^{(p-1)(q-1)/4}$, again by the computation of the even Hilbert symbol (10.4.5). Thus the product becomes

$$\prod_v (p, q)_v = (-1)^{(p-1)(q-1)/4} \left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = 1$$

by quadratic reciprocity. □

Hilbert reciprocity has several aesthetic advantages over the law of quadratic reciprocity. For one, it is simpler to write down! Also, Hilbert believed that his reciprocity law is a kind of analogue of Cauchy's integral theorem, expressing an integral as a sum of residues (11.9.6). The fact that a normalized product over all places is trivial also arises quite naturally: if we define for $x \in \mathbb{Q}^\times$ and a prime p the normalized absolute value

$$|x|_p = p^{-\operatorname{ord}_p(x)},$$

and $|x|_\infty$ the usual archimedean absolute value, then

$$\prod_v |x|_v = 1$$

by unique factorization in $\mathbb{Z}$; this is called the *product formula* for $\mathbb{Q}$, for obvious reasons.

From the tight relationship between quaternion algebras and ternary quadratic forms, we obtain the following corollary.

Corollary 11.2.6. *Let Q be a nonsingular ternary quadratic form over $\mathbb{Q}$. Then the set of places v such that Q_v is anisotropic is finite and of even cardinality.*

In particular, by Corollary 11.2.6 we have that if Q_v is isotropic for all but one place v of $\mathbb{Q}$, then Q_v is in fact isotropic for all places v.

Proof. In the bijection of Theorem 4.4.5, the quadratic form Q corresponds to a quaternion algebra $B = \left(\dfrac{a,b}{\mathbb{Q}}\right)$, and by Theorem 4.5.5, Q is isotropic if and only if B is split if and only if $(a,b)_{\mathbb{Q}} = 1$. By functoriality, the same is true over each completion $\mathbb{Q}_v$ for v a place of $\mathbb{Q}$, and therefore the set of places v where Q_v is isotropic is precisely the set of ramified places in B. The result then follows by Hilbert reciprocity. □

To conclude this section, we show that every allowable product of Hilbert symbols is obtained.

Proposition 11.2.7. *Let Σ be a finite set of places of $\mathbb{Q}$ of even cardinality. Then there exists a quaternion algebra B over $\mathbb{Q}$ with* $\operatorname{Ram}(B) = \Sigma$.

Just as with Hilbert reciprocity, Proposition 11.2.7 touches on a deep statement in number theory concerning primes.

Theorem 11.2.8 (Primes in arithmetic progression)**.** *Given $a, n \in \mathbb{Z}$ coprime, there are infinitely many primes $p \equiv a \pmod{n}$.*

We now prove Proposition 11.2.7 assuming Theorem 11.2.8.

Proof. Let $d = \prod_{p \in \Sigma} p$ be the product of the primes in Σ, and let $u = -1$ if $\infty \in \Sigma$ and $u = 1$ otherwise. Let $d^* = ud$. We consider quaternion algebras of the form

$$B = \left(\frac{d^*, q^*}{\mathbb{Q}}\right)$$

with $q^* = uq$ (and q prime) satisfying certain congruence conditions to ensure $\operatorname{Ram}(B) = \Sigma$. To this end, we seek a prime q such that

$$q^* \text{ is a quadratic nonresidue modulo } p \text{ for all odd } p \mid d \tag{11.2.9}$$

and

$$q^* \equiv \begin{cases} 1 \pmod 8, & \text{if } 2 \nmid d; \\ 5 \pmod 8, & \text{if } 2 \mid d. \end{cases} \tag{11.2.10}$$

There exists a prime satisfying the conditions (11.2.9)–(11.2.10) by the theorem on primes in arithmetic progression (Theorem 11.2.8), since the condition to be a quadratic nonresidue is a congruence condition on q^* and hence on q modulo p.

We now verify that B has $\operatorname{Ram}(B) = \Sigma$. We have $(d^*, q^*)_\infty = u$ by choice of signs and $(d^*, q^*)_p = 1$ for all $p \nmid 2dq$. We compute that

$$(d^*, q^*)_p = \left(\frac{q^*}{p}\right) = -1 \quad \text{for all odd } p \mid d$$

by (11.2.9). For $p = 2$, we find that $(d^*, q^*)_2 = -1$ or $(d^*, q^*)_2 = 1$ according as $2 \mid d$ or not by the computation of the even Hilbert symbol (10.4.5). This shows that

$$\Sigma \subseteq \operatorname{Ram}(B) \subseteq \Sigma \cup \{q\}.$$

The final symbol $(d^*, q^*)_q$ is determined by Hilbert reciprocity (Proposition 11.2.1): since $\#\Sigma$ is already even, we must have $(d^*, q^*)_q = 1$ and $\Sigma = \operatorname{Ram}(B)$. □

11.3 Hasse–Minkowski theorem over the rationals

To complete the proof of Theorem 11.1.3, we need to show that the association $B \mapsto \operatorname{Ram}(B)$ is injective (on isomorphism classes).

Proposition 11.3.1. *Let A, B be quaternion algebras over $\mathbb{Q}$. Then $A \cong B$ if and only if $\operatorname{Ram}(A) = \operatorname{Ram}(B)$ if and only if $A_v \cong B_v$ for all (but one) places v.*

The statement of Proposition 11.3.1 is a local–global principle: the global isomorphism class is determined by the local behavior at primes. For example, we have the following consequence.

Corollary 11.3.2. *Let B be a quaternion algebra over $\mathbb{Q}$. Then $B \cong \mathrm{M}_2(\mathbb{Q})$ if and only if $B_p \cong \mathrm{M}_2(\mathbb{Q}_p)$ for all primes p.*

By the equivalence between quaternion algebras and quadratic forms (Chapter 4 and specifically Section 4.4), the statement of Proposition 11.3.1 is equivalent to the statement that a ternary quadratic form over $\mathbb{Q}$ is isotropic if and only if it is isotropic over all (but one) completions. In fact, the more general statement is true—and again we come in contact with a deep result in number theory.

Theorem 11.3.3 (Hasse–Minkowski)**.** *Let Q be a quadratic form over $\mathbb{Q}$. Then Q is isotropic if and only if Q_v is isotropic for all places v of $\mathbb{Q}$.*

This theorem of Hasse–Minkowski is more generally called the *Hasse principle* (see Paragraph 11.9.4).

We will prove the Hasse–Minkowski theorem by induction on the number of variables. Of particular interest is the case of (nondegenerate) ternary quadratic forms, for which we have the following theorem of Legendre.

Theorem 11.3.4 (Legendre)**.** *Let $a, b, c \in \mathbb{Z}$ be nonzero, squarefree integers that are relatively prime in pairs. Then the quadratic form*

$$ax^2 + by^2 + cz^2 = 0$$

has a nontrivial solution if and only if a, b, c do not all have the same sign and

$$-ab,\ -bc,\ -ac \text{ are quadratic residues modulo } |c|,\ |a|,\ |b|, \text{ respectively.}$$

Proof. First, the conditions for solvability are indeed necessary. The condition on signs is necessary for a solution in $\mathbb{R}$. If $ax^2 + by^2 + cz^2 = 0$ with $x, y, z \in \mathbb{Q}$ not all zero, then scaling we may assume $x, y, z \in \mathbb{Z}$ satisfy $\gcd(x, y, z) = 1$; if $p \mid c$ then $p \nmid y$ (else $p \mid x$ so $p \mid z$, contradiction), so $(x/y)^2 \equiv (-b/a) \pmod{|c|}$ so $-ba$ is a quadratic residue modulo $|c|$; the other conditions hold by symmetry.

So suppose the conditions hold. Multiplying through and rescaling by squares, we may assume a, b are squarefree (but not necessarily coprime) and $c = -1$, and we seek a nontrivial solution to $ax^2 + by^2 = z^2$. If $a \in \mathbb{Q}^{\times 2}$, then we are done. Otherwise, we need to solve

$$\frac{z^2 - ax^2}{y^2} = b = \mathrm{N}_{\mathbb{Q}(\sqrt{a})/\mathbb{Q}}\left(\frac{z + x\sqrt{a}}{y}\right)$$

for $x, y, z \in \mathbb{Q}$ and $y \neq 0$, i.e., we need to show that b is a norm from $F = \mathbb{Q}(\sqrt{a})$. By hypothesis, a, b are not both negative and

$$b \text{ is a square modulo } |a| \text{ and } a \text{ is a square modulo } |b|. \tag{11.3.5}$$

We may also assume $|a| \leq |b|$.

We use complete induction on $m = |a| + |b|$. If $m = 2$, then we have the equation $\pm x^2 \pm y^2 = z^2$ with the case both negative signs excluded, each of which has solutions. Now suppose that $m > 2$ so $|b| \geq 2$, and let $p \mid b$ be prime divisor. By hypothesis, there exist integers t, b' such that $t^2 = a + bb'$; taking a small residue, we may assume $|t| < |b|/2$. Thus

$$bb' = t^2 - a = N_{F/\mathbb{Q}}(t + \sqrt{a})$$

so bb' is a norm from F. Thus b is a norm if and only if b' is a norm. But

$$\left|b'\right| = \left|\frac{t^2 - a}{b}\right| \leq \frac{|b|}{4} + 1 < |b|$$

because $|b| \geq 2$. Now write $b' = b''u^2$ with $b'', u \in \mathbb{Z}$ and b'' squarefree. Then $|b''| \leq |b'| < |b|$ and b'' is a norm if and only if b' is a norm. With these manipulations, we propagate the hypothesis that $|a|$ is a square modulo $|b''|$ and $|b''|$ is a square modulo $|a|$. Therefore, the induction hypothesis applies to the equation $ax^2 + b''y^2 = z^2$, and we are done. □

Corollary 11.3.6. *Let Q be a nondegenerate ternary quadratic form over $\mathbb{Q}$. Then Q is isotropic if and only if Q_v is isotropic for all (but one) places v of $\mathbb{Q}$.*

Proof. If Q is isotropic, then Q_v is isotropic for all v. For the converse, suppose that Q_v is isotropic for all places v of $\mathbb{Q}$. As in the proof of Legendre's Theorem 11.3.4, we may assume $Q(x, y, z) = ax^2 + by^2 - z^2$. The fact that Q is isotropic over $\mathbb{R}$ implies that a, b are not both negative. Now let $p \mid a$ be odd. The condition that Q_p is isotropic is equivalent to $(a, b)_p = (b/p) = 1$; putting these together, we conclude that b is a quadratic residue modulo $|a|$. The same holds for a, b interchanged, so (11.3.5) holds and the result follows. □

We are now in a position to complete the proof of the Hasse–Minkowski theorem.

Proof of Theorem 11.3.3. We may assume that Q is nondegenerate in $n \geq 1$ variables. If $n = 1$, the statement is vacuous. If $n = 2$, the after scaling we may assume $Q(x, y) = x^2 - ay^2$ with $a \in \mathbb{Q}^\times$; since Q_p is isotropic for all primes p, we have $a \in \mathbb{Q}_p^{\times 2}$ so in particular $\operatorname{ord}_p(a)$ is even for all primes p; since Q is isotropic at ∞, we have $a > 0$; thus by unique factorization, we have $a \in \mathbb{Q}^{\times 2}$ and the result follows. If $n = 3$, the statement is proven in Corollary 11.3.6.

Now suppose $n \geq 4$. Write $Q = \langle a, b\rangle \perp -Q'$ where $Q' = \langle c_1, \dots, c_{n-2}\rangle$ and $a, b, c_i \in \mathbb{Z}$. Let $d = 2ab(c_1 \cdots c_{n-2}) \neq 0$. For each prime $p \mid d$, since Q is isotropic, there exists $t_p \in \mathbb{Q}_p^\times$ represented by both $\langle a, b\rangle$ and Q' in $\mathbb{Q}_p$. (This requires a small argument, which is left as Exercise 11.4.) Similarly, there exists $t_\infty \in \mathbb{R}^\times$ represented by these forms in $\mathbb{R}$.

By another application of primes in arithmetic progression (Exercise 11.5), there exists $t \in \mathbb{Q}^\times$ such that:

(i) $t \in t_p\mathbb{Q}_p^{\times 2}$ for all primes $p \mid d$,

(ii) t and t_∞ have the same sign, and

(iii) $p \nmid t$ for all primes $p \nmid d$ except possibly for one prime $q \nmid d$.

Now the quadratic form $\langle a, b, -t\rangle$ is isotropic for all $p \mid d$ and at ∞ by construction and at all primes $p \nmid d$ except $p = q$ since $p \nmid abt$. Therefore, by case $n = 3$ (using the "all but one" in Corollary 11.3.6), the form $\langle a, b, -t\rangle$ is isotropic.

On the other side, if $n = 4$, then the form $\langle t\rangle \perp Q'$ is isotropic by the same argument. If $n \geq 5$, then we apply the induction hypothesis to Q': the hypothesis holds, since we have Q' is isotropic at ∞ and all $p \mid d$ by construction, and for all $p \nmid d$ the completion Q'_p is a nondegenerate form in ≥ 3 variables over $\mathbb{Z}_p$ so is isotropic by the results of Section 10.3, using Hensel's lemma to lift a solution modulo the odd prime p.

Finally, putting these two together, we find that Q is isotropic over $\mathbb{Q}$. □

We conclude with the following consequence, which immediately implies Proposition 11.3.1.

Corollary 11.3.7. *Let Q, Q' be quadratic forms over $\mathbb{Q}$ in the same number of variables. Then $Q \cong Q'$ are isometric if and only if $Q_v \cong Q'_v$ for all places v.*

Proof. The implication ($\Rightarrow$) is immediate. We prove ($\Leftarrow$) by induction on the number of variables, the case of $n = 0$ variables being clear. By splitting the radical (Paragraph 4.2.12), we may assume that Q, Q' are nondegenerate. Let $a \in \mathbb{Q}^\times$ be represented by Q. Since $Q_v \cong Q'_v$ the quadratic form $\langle a\rangle \perp Q'$ is isotropic at v for all v, so Q' represents a (Lemma 4.5.4). So in both cases, we can write $Q \cong \langle a\rangle \perp Q_1$ and $Q' \cong \langle a\rangle \perp Q'_1$ for quadratic forms Q_1, Q'_1 in one fewer number of variables. Finally, by Witt cancellation (Proposition 4.2.16), from $Q_v \cong Q'_v$ we have $(Q_1)_v \cong (Q'_1)_v$ for all v, so by induction, we have $Q_1 \cong Q'_1$, and thus $Q \cong Q'$. □

In summary, the classification of quaternion algebras over $\mathbb{Q}$ embodies some deep statements in number theory: quadratic reciprocity (and its reformulation in Hilbert reciprocity), the Hasse-Minkowski theorem (the local–global principle for quadratic forms), and the proofs use the theorem of primes in arithmetic progression! It is a small blessing that we can make these essentially elementary arguments over $\mathbb{Q}$. In the more general case, we must dig more deeply.

11.4 Global fields

In this chapter and in many that remain, we focus on a certain class of fields of arithmetic interest: a *global field* is a finite extension of $\mathbb{Q}$ (a *number field*) or $\mathbb{F}_p(t)$ (a *function field*) for a prime p. Global fields are strongly governed by their completions with respect to nontrivial absolute values, which are local fields. Throughout this text, we will return to this theme that global behavior is governed by local behavior.

For the rest of this chapter, let F be a global field. The set of *places* of F is the set of equivalence classes of embeddings $\iota_v : F \to F_v$ where F_v is a local field and $\iota_v(F)$ is dense in F_v; two embeddings $\iota_v : F \to F_v$ and $\iota_v' : F \to F_v'$ are equivalent if there is an isomorphism of topological fields $\phi : F_v \to F_v'$ such that $\iota_v' = \phi \circ \iota_v$.

Every valuation $v : F \to \mathbb{R} \cup \{\infty\}$, up to scaling, defines a place $\iota_v : F \to F_v$ where v is the completion of F with respect to the absolute value induced by v; we call such a place *nonarchimedean*, and using this identification we will write v for both the place of F and the corresponding valuation. For a nonarchimedean place v corresponding to a local field F_v, we denote by R_v its valuation ring, $\mathfrak{p}_v$ its maximal ideal, and k_v its residue field. If F is a function field, then all places of F are nonarchimedean. If F is a number field, a place $F \hookrightarrow \mathbb{R}$ is called a *real place* and a place $F \hookrightarrow \mathbb{C}$ (equivalent to its complex conjugate) is called a *complex place*. A real or complex place is *archimedean*.

A global field F has a set of *preferred* embeddings $\iota_v : F \hookrightarrow F_v$ corresponding to each place v (equivalently, a preferred choice of absolute values $|\ |_v$ for each place v) such that the *product formula* holds: for all $x \in F^\times$, we have

$$\prod_v |x|_v = 1. \tag{11.4.1}$$

The preferred absolute values are defined as follows.

11.4.2. The set of places of $\mathbb{Q}$ consists of the archimedean real place, induced by the embedding $\mathbb{Q} \hookrightarrow \mathbb{R}$ and the usual absolute value $|x|_\infty$, and the set of nonarchimedean places indexed by the primes p given by the embeddings $\mathbb{Q} \hookrightarrow \mathbb{Q}_p$, with the preferred absolute value

$$|x|_p = p^{-\operatorname{ord}_p(x)}.$$

The statement of the product formula for $x \in \mathbb{Q}$ is

$$\prod_p p^{-\operatorname{ord}_p(x)}|x|_\infty = 1$$

and this follows from unique factorization in $\mathbb{Z}$.

11.4.3. The set of places of $\mathbb{F}_p(t)$ is indexed by monic irreducible polynomials $f(t) \in \mathbb{F}_p[t]$ with preferred absolute value

$$|x(t)|_f = p^{-(\deg f)\operatorname{ord}_f(x)}$$

and $1/t$, the place at infinity, with preferred absolute value

$$|x(t)|_{1/T} = p^{\deg x}.$$

Then the statement of the product formula for $x(t) \in \mathbb{F}_p(t)$ is

$$\prod_f p^{(\deg f)\operatorname{ord}_f(x)} = p^{\deg x}$$

which follows from unique factorization in $\mathbb{F}_p[t]$.

11.4.4. More generally, let K/F be a finite extension of global fields. Let v be a place of F with a preferred absolute value and let w is a place of K above v. Then the preferred absolute value for w is

$$|x|_w = |\mathrm{N}_{K/F}(x)|_v^{1/[K:F]}$$

for $x \in K$; note that if $x \in F$ then

$$|x|_w = |\mathrm{N}_{K/F}(x)|_v^{1/[K:F]} = |x|_v,$$

so the absolute value $|\ |_w$ extends $|\ |_v$, and thus this definition is compatible with further field extensions. [[In a potentially different way than valuations extend.]]

If F satisfies the product formula (11.4.1) with respect to preferred absolute values, then so does K, since

$$\prod_w |x|_w = \prod_v \left(\prod_{w|v} |x|_w \right) = \prod_v |\mathrm{N}_{K/F}(x)|_v = 1.$$

Therefore every global field satisfies the product formula with respect to preferred absolute values.

11.4.5. Let F be a global field. Let S be a nonempty finite set of places of F containing all archimedean places of F (if F is a number field).

The *ring of S-integers* in F is the set

$$R_S = \{x \in F : v(x) \geq 0 \text{ for all } v \notin S\}.$$

This definition makes sense as if $v \notin S$ then by hypothesis v is nonarchimedean. We will often abbreviate $R = R_S$.

Example 11.4.6. If F is a number field and S consists only of the archimedean places in F then R_S is the *ring of integers* in F, the integral closure of $\mathbb{Z}$ in F. If F is a function field, corresponding to a curve X, then R_S is the ring of all rational functions with no poles outside S.

11.5 Ramification and discriminant

Let $R = R_S$ be an S-integer ring of F. Let B be a quaternion algebra over F.

Definition 11.5.1. Let v be a place of F. We say that B is *ramified* at v if $B_v = B \otimes_F F_v$ is a division ring; otherwise we say that B is *split* (or *unramified*) at v. Let $\operatorname{Ram}(B)$ denote the set of ramified places of B.

We use the term *ramified* for the following reason: if $B_{\mathfrak{p}}$ is a division ring with valuation ring $\mathcal{O}_{\mathfrak{p}}$, then $\mathfrak{p}\mathcal{O}_{\mathfrak{p}} = P^2$ for a two-sided maximal ideal P, by the results of Section 10.5. In a similar way, we will see that if $\mathfrak{p}$ is a prime of R unramified in B, then there exist R-lattices of reduced norm $\mathfrak{p}$, so the term *split* is justified.

Lemma 11.5.2. *The set* $\operatorname{Ram}(B)$ *of ramified places of B is finite.*

Proof. Let $B = \left(\dfrac{K,b}{F}\right)$. Since F has only finitely many archimedean places, we may suppose v is nonarchimedean. The extension K/F is ramified at only finitely many places, so we may assume that K/F is unramified at v (the corresponding prime $\mathfrak{p}$ is split or inert). Finally, we have $v(b) = 0$ for all but finitely many v, so we may assume $v(b) = 0$. But then under these hypotheses, we have $B_v = \left(\dfrac{K_v,b}{F_v}\right)$ is split, by Corollary 10.5.10. □

Motivated by the fact that the discriminant of a quadratic field extension is divisible by ramifying primes, we make the following definition.

Definition 11.5.3. The *discriminant* of B (relative to S) is the R-ideal

$$\operatorname{disc}_S(B) = \prod_{\substack{\mathfrak{p} \text{ ramified} \\ \mathfrak{p} \notin S}} \mathfrak{p} \subseteq R$$

obtained as the product of all primes $\mathfrak{p}$ of R ramified in B.

11.5.4. When F is a number field and S consists of archimedean places only, so R is the ring of integers of F, then we sometimes abbreviate $\operatorname{disc}_S(B) = \operatorname{disc}(B)$. Note that the discriminant $\operatorname{disc}_S(B)$ discards information about primes in S, so one should refer to the set of ramified places for something independent of S.

The discriminant captures maximal R-orders in B, as follows: in particular, we will see that maximal R-orders are characterized by their discriminants (and hence by the ramified places of B not in S).

Proposition 11.5.5. *Let $\mathcal{O}$ be an R-order in B. Then $\mathcal{O}$ is maximal if and only if* $\operatorname{discrd}(\mathcal{O}) = \operatorname{disc}_S(B)$.

Proof. The order $\mathcal{O}$ is maximal if and only if $\mathcal{O}_{\mathfrak{p}}$ is maximal for all primes $\mathfrak{p}$ of R by Lemma 12.4.2, so it suffices to prove this statement locally. If B is split at $\mathfrak{p}$ then $B_{\mathfrak{p}} \cong \mathrm{M}_2(F_{\mathfrak{p}})$, and an order is maximal if and only if it is isomorphic to $\mathrm{M}_2(R_{\mathfrak{p}})$. But $\mathrm{M}_2(R_{\mathfrak{p}})$ has discriminant $R_{\mathfrak{p}}$ by Exercise 13.3, and by Lemma 13.2.16, an order is maximal if and only if it has discriminant $R_{\mathfrak{p}}$.

In a similar way, if B is ramified at $\mathfrak{p}$ then $B_{\mathfrak{p}}$ has a unique maximal order of discriminant $\mathfrak{p}R_{\mathfrak{p}}$, and the same argument applies. □

Remark 11.5.6. The preceding results have an analogue in the case of a quadratic field extension. Let K be a quadratic extension of F and suppose that $1/2 \in R$. Let S be the integral closure of R in K. Then an R-order in K is maximal if and only if its discriminant is equal to

$$\operatorname{disc}(S) = \prod_{\substack{\mathfrak{p} \text{ ramified in } K \\ \mathfrak{p} \notin S}} \mathfrak{p} \subseteq R.$$

11.5.7. It follows from Proposition 11.5.5 that

$$\operatorname{disc}_S(B) \mid \operatorname{discrd}(\mathcal{O}).$$

So we define the *level* of an order $\mathcal{O}$ is the R-ideal

$$\mathfrak{N} = \mathfrak{N}(\mathcal{O}) = \operatorname{discrd}(\mathcal{O}) \operatorname{disc}_S(B)^{-1} \subseteq R.$$

11.6 Quaternion algebras over global fields

In the final section of this chapter, we deduce results characterizing isomorphism classes of quaternion algebras assuming two results from number theory. In later chapters, we give two self-contained proofs: an analytic approach using L-functions (Chapter [[??]]) and a cohomological approach using ideles (Chapter [[??]]). These arguments are more involved, and so for now we merely provide an exposition of the statements.

The main result is as follows.

Theorem 11.6.1. *The map* $B \mapsto \operatorname{Ram}(B)$ *gives a bijection*

$$\left\{ \begin{array}{c} \text{Quaternion algebras over } F \\ \text{up to isomorphism} \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} \text{Finite subsets of noncomplex places} \\ \text{of } F \text{ of even cardinality} \end{array} \right\}.$$

Theorem 11.6.1 is comparable in depth and difficulty to the main theorems of class field theory; no proof that avoids these techniques is known. Several corollaries are important to note.

Corollary 11.6.2. *Let B be a quaternion algebra over a global field. Then the set of places of F where B is ramified is finite and of even cardinality.*

Recall the definition of the Hilbert symbol (Section 4.7). For a place v of F, we abbreviate $(a,b)_{F_v} = (a,b)_v$; this symbol was computed for v an odd nonarchimedean place (10.4.3).

Corollary 11.6.3 (Hilbert reciprocity)**.** *Let F be a global field with* $\operatorname{char} F \neq 2$ *and let $a, b \in F^\times$. Then*

$$\prod_v (a,b)_v = 1.$$

The statement of Hilbert reciprocity (Corollary 11.6.2) can be rightly seen as a law of quadratic reciprocity for number fields; indeed, it is equivalent to the law of quadratic reciprocity over $\mathbb{Q}$, as we saw in Section 11.2.

Corollary 11.6.4. *Let A, B be quaternion algebras over F. Then $A \cong B$ if and only if* $\operatorname{Ram}(A) = \operatorname{Ram}(B)$ *if and only if $A_v \cong B_v$ for all (but one) places v of F.*

In particular, $B \cong$ $\mathrm{M}_2(F)$ *if and only if* $\operatorname{Ram}(B) = \emptyset$.

The statement of Corollary 11.6.4 is a local–global principle: the global isomorphism class is determined by the local behavior at primes.

Corollary 11.6.5. *A finite extension K/F splits B if and only if K_w/F_v splits B_v for all places w (over v) of K.*

Proof. K splits B if and only if $B \otimes_F K \cong \mathrm{M}_2(K)$ if and only if $B \otimes_F K_w = B_v \otimes_{F_v} K_w \cong \mathrm{M}_2(K_w)$ for all places w over v of K. □

To prove Theorem 11.6.1, we will use the following two fact, whose proof we delay until Chapter [[??]]: it is the case of what is known as the *Hasse norm theorem* for quadratic extensions.

Claim 11.6.6. Let K/F be a separable quadratic F-algebra, and let $b \in F^\times$. Then $b \in \mathrm{N}_{K/F}(K^\times)$ if and only if $b \in \mathrm{N}_{K_v/F_v}(K_w^\times)$ for all places v of F.

This statement is summarized as *everywhere local norms are global norms*. Since every quaternion algebra is of the form $B = \left(\dfrac{K,b}{F}\right)$ with K separable over F, and $B \cong \mathrm{M}_2(F)$ if and only if $b \in \mathrm{N}_{K/F}(K^\times)$, the claim is equivalent to the statement that $B \cong \mathrm{M}_2(F)$ if and only if $\operatorname{Ram}(B) = \emptyset$.

11.7 Hasse–Minkowski theorem

Claim 11.7.1. Let K/F be a separable quadratic F-algebra. Then $K \cong F \times F$ if and only if $K_v \cong F_v \times F_v$ for all places v of F.

If $\operatorname{char} F \neq 2$, and we can complete the square to obtain $K = F(\sqrt{a})$ with $a \in F^{\times}$, then this claim becomes $a \in F^{\times 2}$ if and only if $a \in F_v^{\times 2}$ for all places v of F: *everywhere local squares are global squares*. This claim is the simpler, commutative analogue of the final statement in Corollary 11.6.4.

We now use these two claims to prove the Hasse–Minkowski theorem.

Theorem 11.7.2 (Hasse-Minkowski)**.** *Let F be a field with* $\operatorname{char} F \neq 2$ *and let Q be a quadratic form over F. Then Q is isotropic if and only if Q_v is isotropic over F_v for all places v of F.*

Proof. It suffices to prove sufficiency. We may assume, without loss of generality, that Q is nondegenerate. If $n = \dim Q = 1$, the theorem is vacuous. If $n = 2$, then after scaling we may assume $Q = \langle 1, -a\rangle$; the equivalence follows immediately from Claim 11.7.1. If $n = 3$, then we may take $Q = \langle 1, -a, -b\rangle$, and the equivalence follows from Claim 11.6.6. □

11.8 Representative quaternion algebras

Write down explicit representation of quaternion algebras and maximal orders based on their discriminant.

Over $\mathbb{Q}$, can do this with formulas.

More generally, use primes in arithmetic progression. See also [Lem11, Proposition 2.10, Proposition 6.9].

11.9 Extensions and further reading

11.9.1. Conway would tell us to take -1 as a prime.

11.9.2. Proof of quadratic reciprocity.

11.9.3. Primes in arithmetic progression.

11.9.4. Called the Hasse principle because of Hasse's contribution over number fields. See Fenster and SchwŁrmer [FS07]. [FS07].

11.9.5. The definitions for the preferred absolute values may seem boring, but we will see later that they are natural from the perspective of Haar measure.

11.9.6. Hilbert saw his reciprocity law for the product of the Hilbert symbols as an analogue of Cauchy's integral theorem; more precisely (according to Shafarevich), it is the analogue of the corollary to Cauchy's integral theorem that say that the sum of the residues of a holomorphic differential at all points of a Riemann surfaces is zero. In this analogy, the symbol $(a,b)_p$ is the analogue of the differential $a\,db$ at the point p.

11.10 Algorithmic aspects

Effective solution of conics, geometry of numbers.

Computing the set of ramified places.

Computing a representative algebra.

Exercises

11.1. Complete the proof of Hilbert reciprocity (Proposition 11.2.1) in the remaining cases $(a,b) = (-1,2), (2,2), (-1,p), (2,p)$. In particular, show that

$$\left(\frac{-1,2}{\mathbb{Q}}\right) \cong \left(\frac{2,2}{\mathbb{Q}}\right) \cong \mathrm{M}_2(\mathbb{Q})$$

and

$$(a,p)_2 = (a,p)_p = \left(\frac{a}{p}\right)$$

for $a = -1, 2$ (cf. Paragraph para:Hilbm1m1).

11.2. Show that the law of Hilbert reciprocity (Proposition 11.2.1) implies the law of quadratic reciprocity; with the argument given in section 11.1, this completes the equivalence of these two laws.

11.3. Show that Legendre's theorem can be deduced from the statement where $a, b > 0$ and $c = -1$.

11.4. Let $Q = Q' \perp Q''$ be an orthogonal sum of two nondegenerate quadratic forms over a field F. Show that Q is isotropic if and only if there exists $c \in F$ that is represented by both Q' and Q''.

11.5. Let S be a finite set of places of $\mathbb{Q}$ containing ∞. For each $v \in S$, let $t_v \in \mathbb{Q}_v^\times$ be given. Show that there exists $t \in \mathbb{Q}^\times$ such that $t \in t_v\mathbb{Q}_v^{\times 2}$ for all $v \in S$ and $\operatorname{ord}_p(t) = 0$ for all $p \notin S \setminus \{\infty\}$ except (possibly) for one prime $p = q$.

By another application of primes in arithmetic progression (Exercise 11.5), there exists $t \in \mathbb{Q}^\times$ such that $t \in t_p\mathbb{Q}_p^{\times 2}$ for all primes $p \mid d$, t and t_∞ have the same sign, and $p \nmid t$ for all primes $p \nmid d$ except for one prime $q \nmid d$.

11.6. Let $F = \mathbb{Q}(\sqrt{d})$ be a real quadratic field. Find $a, b \in F^\times$ such that $\left(\dfrac{a, b}{F}\right)$ is a division ring unramified at all finite places.

11.7. Let $K \supseteq F$ be finite separable extension of global fields. Let B be a quaternion algebra over K. We say that B *descends* to F if there exists a quaternion algebra A over F such that $A \otimes_F K \cong B$. Show that B descends to F if and only if $\operatorname{Ram}(B)$ is invariant under $\operatorname{Gal}(K/F)$.

11.8. Show without using primes in arithmetic progression that over $\mathbb{Q}$ all possible (even cardinality) ramification sets can occur. Do this by using the Brauer relation and linkage.

Chapter 12

Lattices and localization

Properties of a domain are governed in an important way by its localizations, and consequently the structure of lattices, orders, and algebras can often be understood by looking at their localizations and completions. This chapter develops these themes, a local-global principle that we will carry throughout the text.

12.1 Localization of integral lattices

A lattice over the integers $\mathbb{Z}$ is a

12.2 Localizations

Throughout this chapter, let R be a noetherian domain with field of fractions F.

For a prime ideal $\mathfrak{p} \subseteq R$, we denote by

$$R_{(\mathfrak{p})} = \{r/s : s \notin \mathfrak{p}\} \subseteq F$$

the localization of R at $\mathfrak{p}$. (We reserve the simpler subscript notation for the completion, which will feature more prominently.) Since R is a domain, the map $R \hookrightarrow R_{(\mathfrak{p})}$ is an embedding and we can recover R by

$$R = \bigcap_{\mathfrak{p}} R_{(\mathfrak{p})} = \bigcap_{\mathfrak{m}} R_{(\mathfrak{m})} \subseteq F \qquad (12.2.1)$$

where the intersections are over all prime ideals of R and all maximal ideals of R, respectively.

We now prove a version of the equality (12.2.1) for R-lattices.

Let V be a finite-dimensional F-vector space and let M be an R-lattice in V. For a prime $\mathfrak{p}$ of R, let $M_{(\mathfrak{p})} = M \otimes_R R_{(\mathfrak{p})} = R_{(\mathfrak{p})}M$. Then $M_{(\mathfrak{p})}$ is an $R_{(\mathfrak{p})}$-lattice in V. In this way, M determines a collection $(M_{(\mathfrak{p})})_{\mathfrak{p}}$ indexed over the primes $\mathfrak{p}$ of R.

Lemma 12.2.2. *Let M be an R-lattice in V. Then*

$$M = \bigcap_{\mathfrak{p}} M_{(\mathfrak{p})} = \bigcap_{\mathfrak{m}} M_{(\mathfrak{m})} \subseteq V$$

where the intersection is over all prime (maximal) ideals $\mathfrak{p}$.

Proof. It suffices to prove the statement for maximal ideals since $M_{(\mathfrak{m})} \subseteq M_{(\mathfrak{p})}$ whenever $\mathfrak{m} \supset \mathfrak{p}$.

The inclusion $M \subseteq \bigcap_{\mathfrak{m}} M_{(\mathfrak{m})}$ is clear. So let $x \in V$ satisfy $x \in \bigcap_{\mathfrak{m}} M_{(\mathfrak{m})}$. Let

$$(M : x) = \{r \in R : rx \in M\}.$$

Then $(M : x)$ is an ideal of R. For any maximal ideal $\mathfrak{m}$ of R, since $x \in M_{(\mathfrak{m})}$ there exists $0 \neq r_{\mathfrak{m}} \in R \setminus \mathfrak{m}$ such that $r_{\mathfrak{m}} x \in M$. Thus $r_{\mathfrak{m}} \in (M : x)$ and so $(M : x)$ is not contained in any maximal ideal of R. Therefore $(M : x) = R$ and hence $x \in M$. □

Corollary 12.2.3. *Let M, N be R-lattices in V. Then $M \subseteq N$ if and only if $M_{(\mathfrak{p})} \subseteq N_{(\mathfrak{p})}$ for all prime ideals $\mathfrak{p}$ of R if and only if $M_{(\mathfrak{m})} \subseteq N_{(\mathfrak{m})}$ for all maximal ideals $\mathfrak{m}$ of R.*

Proof. We have $M = \bigcap_{\mathfrak{p}} M_{(\mathfrak{p})} \subseteq \bigcap_{\mathfrak{p}} N_{(\mathfrak{p})} = N$ and similarly for $\mathfrak{m}$; the other inclusion is clear. □

Lemma 12.2.4. *Let B be a finite-dimensional F-algebra. An R-lattice $I \subseteq B$ is an R-order if and only if $I_{(\mathfrak{p})}$ is an $R_{(\mathfrak{p})}$-order for all primes $\mathfrak{p}$ of R if and only if $I_{\mathfrak{m}}$ is an $R_{\mathfrak{m}}$-order for all maximal ideals $\mathfrak{m}$ of R.*

Proof. If I is an R-order then $I_{(\mathfrak{p})}$ is an $R_{(\mathfrak{p})}$-order for all primes $\mathfrak{p}$, hence for all maximal ideals $\mathfrak{m}$.

Conversely, suppose that $I_{\mathfrak{m}}$ is an $R_{\mathfrak{m}}$-order for all maximal ideals $\mathfrak{m}$. Then $1 \in \bigcap_{\mathfrak{m}} I_{(\mathfrak{m})} = I$ and for any $\alpha, \beta \in M$ we have $\alpha\beta \in \bigcap_{\mathfrak{m}} I_{(\mathfrak{m})} = I$, so I is a subring of B and hence an order. The statement for prime ideals follows *a fortiori*. □

A property like that of being an R-order is called a *local property*, and a lemma like Lemma 12.2.4 is thereby called a *local-global principle*.

We now pass to completions. Let $R_{\mathfrak{p}}$ denote the completion of R at $\mathfrak{p}$, and let $F_{\mathfrak{p}} = F \otimes_R R_{\mathfrak{p}}$ be the completion of F at $\mathfrak{p}$ and $V_{\mathfrak{p}} = V \otimes_F F_{\mathfrak{p}}$.

Lemma 12.2.5. *The map*

$$M_{(\mathfrak{p})} \mapsto M_{\mathfrak{p}} = M_{(\mathfrak{p})} \otimes_{R_{(\mathfrak{p})}} R_{\mathfrak{p}}$$

yields a bijection between $R_{(\mathfrak{p})}$-lattices in $V_{(\mathfrak{p})}$ and $R_{\mathfrak{p}}$-lattices in $V_{\mathfrak{p}}$, with inverse

$$M_{\mathfrak{p}} \mapsto M_{\mathfrak{p}} \cap V_{(\mathfrak{p})}.$$

Proof. This lemma follows as above once we show that if $M_{(\mathfrak{p})}$ is an $R_{(\mathfrak{p})}$-lattice, then $M_{\mathfrak{p}} \cap V_{(\mathfrak{p})} = M_{(\mathfrak{p})}$: for the details, see Exercise 12.2. □

12.3 Bits of commutative algebra and Dedekind domains

So far, we have worked with finitely generated modules. We pause to consider some relevant bits of commutative algebra in our context.

12.3.1. A free R-module is projective; the converse is true when R is a local ring or PID. In particular, a finitely generated R-module is projective if and only if it is locally free (since R is noetherian). The ability to argue locally and then with free objects is very useful in our investigations (as well as many others), and so very often we will restrict our attention to projective R-modules.

As a basic counterexample to keep in mind, let k be a field and $R = k[x, y]$. Then the R-module (x, y) is *not* projective (Exercise 12.1). Similarly, if $R = \mathbb{Z}/n^2\mathbb{Z}$ for $n \in \mathbb{Z}_{>1}$, then $\mathbb{Z}/n\mathbb{Z}$ is not projective as an R-module.

There is a class of rings where finitely generated modules are projective, namely when R is a *Dedekind domain*: a (noetherian) integrally closed domain such that every nonzero prime ideal is maximal.

Example 12.3.2. Trivially, any field is a Dedekind domain.

The rings $\mathbb{Z}$ and $\mathbb{F}_p[t]$ are Dedekind domains. If K is a finite extension of $\mathbb{Q}$ or $\mathbb{F}_p(t)$, then the integral closure of $\mathbb{Z}$ or $\mathbb{F}_p[t]$ in K is a Dedekind domain.

The localization or completion of a Dedekind domain R at a prime $\mathfrak{p}$ is again a Dedekind domain.

Suppose R is a Dedekind domain. Then every nonzero ideal $\mathfrak{a}$ of R can be written uniquely as the product of prime ideals (up to reordering). A *fractional ideal* of R is a nonzero projective R-submodule $\mathfrak{a} \subseteq F$; a subset $\mathfrak{a} \subseteq F$ is a fractional ideal if and only if there exists $d \in R \setminus \{0\}$ such that $d\mathfrak{a} \subseteq R$ is an ideal in the usual sense. If $\mathfrak{a}, \mathfrak{b}$ are fractional R-ideals, then $\mathfrak{a} \subseteq \mathfrak{b}$ if and only if $\mathfrak{a} \otimes_R R_{(\mathfrak{p})} = \mathfrak{a}_{(\mathfrak{p})} \subseteq \mathfrak{b}_{(\mathfrak{p})}$ for all primes $\mathfrak{p}$, and hence equality holds if and only if it holds locally. Indeed, we have for any fractional ideal $\mathfrak{a}$ that $\mathfrak{a} = \bigcap_{\mathfrak{p}} \mathfrak{a}_{(\mathfrak{p})}$. For every fractional ideal $\mathfrak{a}$ of R, the set $\mathfrak{a}^{-1} = \{a \in F : a\mathfrak{a} \subseteq R\}$ is a fractional ideal with $\mathfrak{a}\mathfrak{a}^{-1} = R$. Therefore the set of fractional ideals of R forms a group under multiplication.

The localization of a Dedekind domain is a discrete valuation ring (DVR), hence a PID. Consequently, every fractional ideal of R is *locally principal*, i.e., if $\mathfrak{a} \subseteq R$ is a fractional ideal, then for all primes $\mathfrak{p}$ of R we have $\mathfrak{a}_{(\mathfrak{p})} = \mathfrak{a} \otimes_R R_{(\mathfrak{p})} = a_{\mathfrak{p}} R_{(\mathfrak{p})}$ for some $a_{\mathfrak{p}} \in R_{(\mathfrak{p})}$.

Let $M \subseteq V$ be an R-lattice. Then M is a projective R-module, and it follows the structure theorem of projective modules over Dedekind domains that M is *completely decomposable*, i.e. there exist fractional ideals $\mathfrak{a}_1, \dots, \mathfrak{a}_n$ of R and elements $x_1, \dots, x_n \in V$ such that

$$M = \mathfrak{a}_1 x_1 \oplus \cdots \oplus \mathfrak{a}_n x_n. \tag{12.3.3}$$

We call the elements $x_1, \dots, x_n$ a *pseudobasis* for M with respect to the *coefficient ideals* $\mathfrak{a}_1, \dots, \mathfrak{a}_n$.

12.4 Lattices and localization

We now characterize in a simple way the conditions under which a collection $(M_{(\mathfrak{p})})_{\mathfrak{p}}$ of $R_{(\mathfrak{p})}$-lattices arise from a global R-lattice, as in the first section. We will see that just as a nonzero ideal of R can be factored uniquely into a product of prime ideals, and hence by the data of these primes and their exponents, so too can a lattice be understood by a finite number of localized lattices, once a "reference" lattice has been chosen (to specify the local behavior of the lattice at the other places).

We retain the hypothesis that R is a Dedekind domain.

Proposition 12.4.1. *Let $M \subseteq V$ be an R-lattice. Then the map $N \mapsto (N_{(\mathfrak{p})})_{\mathfrak{p}}$ establishes a bijection between R-lattices N and collections $(N_{(\mathfrak{p})})_{\mathfrak{p}}$ where $M_{(\mathfrak{p})} = N_{(\mathfrak{p})}$ for all but finitely many primes $\mathfrak{p}$.*

This proposition gives an extension of the local-global principle for lattices: not only can a lattice be recovered by its localizations, but *any* lattice is obtained from a fixed one by making (arbitrary) choices at finitely many localizations.

Proof. Let $N \subseteq V$ be an R-lattice. Then there exists $0 \neq r \in R$ such that $rM \subseteq N \subseteq r^{-1}M$. But r is contained in only finitely many prime (maximal) ideals of R, so for all but finitely many primes $\mathfrak{p}$ we have that r is a unit in $R_{(\mathfrak{p})}$ and thus $M_{(\mathfrak{p})} = N_{(\mathfrak{p})}$.

So consider the set of collections $(N_{(\mathfrak{p})})_{\mathfrak{p}}$ where $N_{(\mathfrak{p})}$ is an $R_{(\mathfrak{p})}$-lattice for each prime $\mathfrak{p}$ with the property that $M_{(\mathfrak{p})} = N_{(\mathfrak{p})}$ for all but finitely many primes $\mathfrak{p}$ of R. Given such a collection, we define $N = \bigcap_{(\mathfrak{p})} N_{(\mathfrak{p})} \subseteq V$. Then N is an R-submodule of F. We show it is an R-lattice in V. For each $\mathfrak{p}$ such that $M_{(\mathfrak{p})} \neq N_{(\mathfrak{p})}$, there exists $r_{\mathfrak{p}} \in R$ such that $r_{\mathfrak{p}} M_{(\mathfrak{p})} \subseteq N_{(\mathfrak{p})} \subseteq r_{\mathfrak{p}}^{-1} M_{(\mathfrak{p})}$. Therefore, if $r = \prod_{\mathfrak{p}} r_{\mathfrak{p}}$ is the product of these elements, then $rM_{(\mathfrak{p})} \subseteq N \subseteq r^{-1}M_{(\mathfrak{p})}$ for all primes $\mathfrak{p}$ with $M_{(\mathfrak{p})} \neq N_{(\mathfrak{p})}$. On the other hand, if $M_{(\mathfrak{p})} = N_{(\mathfrak{p})}$ then already $rM_{(\mathfrak{p})} \subseteq M_{(\mathfrak{p})} = N_{(\mathfrak{p})} \subseteq r^{-1}N_{(\mathfrak{p})} = r^{-1}M_{(\mathfrak{p})}$. Therefore by Corollary 12.2.3, we have $rM \subseteq N \subseteq r^{-1}M$, and so N is an R-lattice.

By Lemma 12.2.2, the association $(N_{(\mathfrak{p})})_{\mathfrak{p}} \mapsto \bigcap_{\mathfrak{p}} N_{(\mathfrak{p})}$ is an inverse to $N \mapsto (N_{(\mathfrak{p})})_{\mathfrak{p}}$. Conversely, given a collection $(N_{(\mathfrak{p})})_{\mathfrak{p}}$, for a nonzero prime $\mathfrak{p}$, we have $\left(\bigcap_{\mathfrak{q}} N_{\mathfrak{q}}\right)_{(\mathfrak{p})} = N_{(\mathfrak{p})}$ since $(R_{\mathfrak{q}})_{(\mathfrak{p})} = F$ so $(N_{\mathfrak{q}})_{(\mathfrak{p})} = V$ whenever $\mathfrak{q} \neq \mathfrak{p}$. □

In this way, we can show that the property of being a maximal order is a local property.

Lemma 12.4.2. *Let B be a finite-dimensional F-algebra. An R-order $\mathcal{O} \subseteq B$ is maximal if and only if $\mathcal{O}_{(\mathfrak{p})}$ is a maximal $R_{(\mathfrak{p})}$-order for all primes $\mathfrak{p}$ of R.*

Proof. If $\mathcal{O}_{(\mathfrak{p})}$ is maximal for each prime $\mathfrak{p}$ then by Corollary 12.2.3 we see that $\mathcal{O}$ is maximal. Conversely, suppose $\mathcal{O}$ is maximal and suppose that $\mathcal{O}_{(\mathfrak{p})} \subsetneq \mathcal{O}'_{(\mathfrak{p})}$ is a proper containment of orders for some nonzero prime $\mathfrak{p}$. Then the set $\mathcal{O}' = (\bigcap_{\mathfrak{q}\neq\mathfrak{p}} \mathcal{O}_{\mathfrak{q}}) \cap \mathcal{O}'_{(\mathfrak{p})}$ is an R-order properly containing $\mathcal{O}$ by Lemma 12.2.4 and Proposition 12.4.1. □

12.5 Adelic completions

For a lattice M, all localizations $M_{(\mathfrak{p})}$ are submodules of the vector space $M_{(0)} = V$ over F, so for example, the notion of intersection makes sense. It is often helpful to be able to do this with the completions as well; this gives us a preview for the addles, which will feature prominently later on.

Define the restricted direct product. The module embeds diagonally and is dense. Prove at least one nice thing.

12.6 Extensions and further reading

12.6.1. Noetherianness is only used when...

12.7 Exercises

12.1. Let k be a field and $R = k[x, y]$. Show that the R-module (x, y) is not projective.

12.2. Let V be a finite-dimensional vector space over F and $I \subseteq V$ an R-lattice. Let $\mathfrak{p}$ be a prime of R, let $R_{(\mathfrak{p})}$ be the localization of R at $\mathfrak{p}$ and let $R_{\mathfrak{p}}$ be the completion of R at $\mathfrak{p}$. Show that if $I_{(\mathfrak{p})} \subseteq V_{(\mathfrak{p})}$ is an $R_{\mathfrak{p}}$-lattice then $I_{\mathfrak{p}} \cap V_{(\mathfrak{p})} = I_{(\mathfrak{p})}$. Conclude that Lemma 12.2.5 holds.

12.3. Prove Lemma 13.2.15: if R is a Dedekind domain and $J \subseteq I \subseteq V$ are R-lattices in a finite-dimensional vector space V over F, then $[I : J]$ is the product of the invariant factors (or elementary divisors) of the torsion R-module I/J.

12.4. Find R-lattices $I, J \subseteq V$ such that $[I : J] = R$ but $I \neq J$ (cf. Proposition 13.2.14).

12.5. Consider the following ‘counterexamples’ to Proposition 12.4.1 for more general integral domains as follows. Let $R = \mathbb{Q}[x, y]$ be the polynomial ring in two variables over $\mathbb{Q}$, so that $F = \mathbb{Q}(x, y)$. Let $V = F$ and $I = R$.

a) Show that yR has the property that $yR \neq R$ for infinitely many prime ideals $\mathfrak{p}$ of R.

b) Consider the collection of lattices given by $J_{\mathfrak{p}} = f(x)R_{\mathfrak{p}}$ if $\mathfrak{p} = (y, f(x))$ where $f(x) \in \mathbb{Q}[x]$ is irreducible and $J_{\mathfrak{p}} = R_{\mathfrak{p}}$ otherwise. Show that $\bigcap_{\mathfrak{p}} J_{\mathfrak{p}} = (0)$.

Instead, to conclude that a collection $(J_{\mathfrak{p}})_{\mathfrak{p}}$ of $R_{\mathfrak{p}}$-lattices arises from a global R-lattice J, one needs that the collection forms a *sheaf* [[cite]].

Chapter 13

Discriminants

13.1 Discriminantal notions

Let $x_1, \dots, x_n \in \mathbb{R}^n$, and let A be the matrix with columns x_i. Then the volume of the box with edges v_i (originating at the origin) has ordinary volume $|\det(A)|$. In this way, we can measure the volume of a number field F by taking the volume of a fundamental parallelepiped for its ring of integers R: if $x_1, \dots, x_n$ is a $\mathbb{Z}$-basis for R and $\iota : F \hookrightarrow F \otimes_{\mathbb{Q}} \mathbb{R} \cong \mathbb{R}^n$, then the volume of R in this embedding is $|\det(A)|$ where A is the matrix with columns $\iota(x_i)$.

We could compute this volume in a different way: we have

$$|\det(A)| = \sqrt{|\det(A^t A)|} = \sqrt{\det(M)}$$

where M has ijth entry equal to the ordinary dot product $v_i \cdot v_j$. The square volume has some advantages: for example, the square volume of the ring of integers computed above is the discriminant of R, an integer: indeed, the associated dot product is the trace form $(x, y) \mapsto \mathrm{Tr}_{F/\mathbb{Q}}(xy)$ (with values in $\mathbb{Z}$ on R).

More generally, whenever we have a symmetric bilinear form $T : F \times F \to F$ (associated to a quadratic form), we have a square volume given by the determinant $\det(T(x_i, x_j))_{i,j}$: this is the (twice-)discriminant of the quadratic form Q associated to T (when $\operatorname{char} F \neq 2$).

Now if B is a finite-dimensional algebra over F, we have a bilinear form $(\alpha, \beta) \mapsto \mathrm{Tr}_{B/F}(\alpha\beta)$, and so in this manner we obtain a discriminant ("square volume") measuring in some way the complexity of B.

In this section, we establish basic facts about discriminants, including how they behave under inclusion (measuring index) and localization. For quaternion orders, it turns out that the discriminant so defined is always a square, and there is an intrinsic way to define the square root of this discriminant, called the *reduced discriminant*.

13.2 Discriminant

As in the commutative case, ramification and maximality of orders is intrinsically related to discriminants. In this section, we defined the discriminant and show that it behaves well with respect to localization.

Let R be a noetherian domain and let $F = \operatorname{Frac}(R)$. Let B be a semisimple algebra over F with $\dim_F B = n$. For elements $\alpha_1, \dots, \alpha_n \in B$, we define

$$d(\alpha_1, \dots, \alpha_n) = \det(\operatorname{trd}(\alpha_i \alpha_j))_{i,j=1,\dots,n}.$$

Let $\mathcal{O} \subseteq B$ be an R-order.

Definition 13.2.1. The *discriminant* of $\mathcal{O}$ is the ideal $\operatorname{disc}(\mathcal{O})$ of R generated by the set

$$\{d(\alpha_1, \dots, \alpha_n) : \alpha_1, \dots, \alpha_n \in \mathcal{O}\}.$$

Although Definition 13.2.1 may look unwieldly, it works as well in the commutative case as in the noncommutative case. We can at least see immediately that if $\mathcal{O} \subseteq \mathcal{O}'$ are R-orders, then $\operatorname{disc}(\mathcal{O}') \mid \operatorname{disc}(\mathcal{O})$. And the function d itself transforms in a nice way as the systems α_i vary (such as in under a change of basis), as follows.

Lemma 13.2.2. *Let $\alpha_1, \dots, \alpha_n \in B$ and suppose $\beta_1, \dots, \beta_n \in B$ have the form $\beta_i = \sum_{j=1}^n m_{ij}\alpha_j$ with $m_{ij} \in F$. Let $M = (m_{ij})_{i,j=1,\dots,n}$. Then*

$$d(\beta_1, \dots, \beta_n) = \det(M)^2 d(\alpha_1, \dots, \alpha_n). \tag{13.2.3}$$

Proof. By properties of determinants, if $\alpha_i = \alpha_j$ for some $i \neq j$, then $d(\alpha_1, \dots, \alpha_n) = 0$. Consequently, if $\alpha_1, \dots, \alpha_n$ are linearly dependent (over F), then $d(\alpha_1, \dots, \alpha_n) = 0$. So we may assume that $\alpha_1, \dots, \alpha_n$ are linearly independent and that $\beta_1, \dots, \beta_n$ are linearly independent as well, so the matrix M, a change of basis matrix, is invertible. But by Gaussian reduction, any invertible matrix is the product of elementary matrices, so it is enough to check that the equality holds when M is an elementary matrix. But in each of these cases (a diagonal matrix, a permutation matrix, or a transvection matrix), the equality is immediate. □

Corollary 13.2.4. *If $\alpha_1, \dots, \alpha_n$ is an R-basis for $\mathcal{O}$, then*

$$\operatorname{disc}(\mathcal{O}) = d(\alpha_1, \dots, \alpha_n)R.$$

Example 13.2.5. Suppose $\operatorname{char} F \neq 2$. Let $B = \left(\dfrac{a,b}{F}\right)$ with $a, b \in R$. Let $\mathcal{O} = R \oplus Ri \oplus Rj \oplus Rij$.

Then $\operatorname{disc}(\mathcal{O})$ is the principal R-ideal generated by

$$\operatorname{disc}(1, i, j, ij) = \det\begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2a & 0 & 0 \\ 0 & 0 & 2b & 0 \\ 0 & 0 & 0 & -2ab \end{pmatrix} = -(4ab)^2.$$

The calculation when $\operatorname{char} F = 2$ is Exercise 13.1.

13.2.6. More generally, if $\mathcal{O}$ is completely decomposable with

$$\mathcal{O} = \mathfrak{a}_1\alpha_1 \oplus \cdots \oplus \mathfrak{a}_n\alpha_n$$

such as in (12.3.3), then from (13.2.3) we have

$$\operatorname{disc}(\mathcal{O}) = (\mathfrak{a}_1 \cdots \mathfrak{a}_n)^2 \operatorname{disc}(\alpha_1, \dots, \alpha_n).$$

Example 13.2.7. Let $B = \mathrm{M}_n(F)$ and $\mathcal{O} = \mathrm{M}_n(R)$. Then $\operatorname{disc}(\mathcal{O}) = R$ (Exercise 13.3).

13.2.8. It follows from Lemma 13.2.2 that $\operatorname{disc}(\mathcal{O})$ is finitely generated as an R-module (apply d to all subsets of a set of generators for $\mathcal{O}$ as an R-module).

13.2.9. Equation (13.2.3) and the fact that $\mathcal{O}_{(\mathfrak{p})} = \mathcal{O} \otimes_R R_{(\mathfrak{p})}$ implies that

$$\operatorname{disc}(\mathcal{O}_{(\mathfrak{p})}) = \operatorname{disc}(\mathcal{O})_{(\mathfrak{p})}.$$

We also have an equality for the completions

$$\operatorname{disc}(\mathcal{O}_\mathfrak{p}) = \operatorname{disc}(\mathcal{O})_\mathfrak{p}$$

because R is dense in $R_\mathfrak{p}$ for the $\mathfrak{p}$-adic topology. In this way, the discriminant respects localization and completion. Therefore, from the local-global principle (Lemma 12.2.2), we have

$$\operatorname{disc}(\mathcal{O}) = \bigcap_{\mathfrak{p}} \operatorname{disc}(\mathcal{O})_{(\mathfrak{p})}.$$

13.2.10. Let $B = \left(\dfrac{K, b}{F}\right)$ with $b \in R$ and let S be an R-order in K. Let $\mathcal{O} = S + Sj$; then $\mathcal{O}$ is an R-order in B by Exercise 8.7. Let $\mathfrak{d} = \operatorname{disc}(S)$. Then $\operatorname{disc}(\mathcal{O}) = (b\mathfrak{d})^2$, by Exercise 13.4.

Lemma 13.2.11. *Suppose that B is separable. Then the discriminant* $\operatorname{disc}(\mathcal{O})$ *of $\mathcal{O} \subseteq B$ is a nonzero ideal of R. If further $\mathcal{O}$ is projective as an R-module, then* $\operatorname{disc}(\mathcal{O})$ *is also a projective R-module.*

Proof. Since $\alpha_i\alpha_j \in \mathcal{O}$, we have $\operatorname{trd}(\alpha_i\alpha_j) \in R$ so $d(\alpha_1, \dots, \alpha_n) \in R$. Since $\mathcal{O}$ is an R-lattice, there exist elements $\alpha_1, \dots, \alpha_n$ which are linearly independent over F, and since B is separable, trd is a nondegenerate bilinear pairing on F, hence $\operatorname{disc}(\mathcal{O})$ is a nonzero ideal of R.

To show that $\operatorname{disc}(\mathcal{O})$ is projective, by Paragraph 12.3.1, we show that $\operatorname{disc}(\mathcal{O})$ is locally principal. Let $\mathfrak{p}$ be a prime ideal of R. Since $\mathcal{O}$ is a projective R-module, its localization $\mathcal{O}_{(\mathfrak{p})}$ is free; thus from Corollary 13.2.4, we conclude that $\operatorname{disc}(\mathcal{O})_{(\mathfrak{p})} = \operatorname{disc}(\mathcal{O}_{(\mathfrak{p})})$ is principal over $R_{(\mathfrak{p})}$, generated by $\operatorname{disc}(\alpha_1, \dots, \alpha_n)$ for any $R_{(\mathfrak{p})}$-basis $\alpha_1, \dots, \alpha_n$ of $\mathcal{O}$. □

We can compare orders by their index and discriminant; we set these up in some useful generality. Let $M, N \subseteq V$ be R-lattices in a finite-dimensional vector space V.

Definition 13.2.12. The *R-index* of N in M is the R-submodule $[M : N]$ of F generated by the set

$$\{\det(\delta) : \delta \in \operatorname{End}_F(V) \text{ and } \delta(N) \subseteq M\}.$$

The index $[M : N]$ is a nonzero R-module, arguing as in Paragraph 8.3.5 (Exercise 13.7).

13.2.13. If M, N are free, then $[M : N]$ is a free R-module generated by the determinant of any $\delta \in \operatorname{End}_F(V)$ giving a change of basis from N to M. In particular, if $N = rM$ with $r \in R$, then $[M : N] = r^n R$ where $n = \dim_F V$.

Proposition 13.2.14. *Suppose that M, N are projective R-modules. Then $[M : N]$ is a projective R-module. Moreover, if $N \subseteq M$ then $[M : N] = R$ if and only if $M = N$.*

Proof. By Paragraph 12.3.1, we can show that $[M : N]$ is locally principal. So let $\mathfrak{p}$ be a prime of R and consider the localization $[M : N]_{(\mathfrak{p})}$ at $\mathfrak{p}$. Since M, N are projective R-modules, they are locally free. Then by Paragraph 13.2.13, if $\delta_{\mathfrak{p}} \in F$ gives a change of basis from $N_{(\mathfrak{p})}$ to $M_{(\mathfrak{p})}$, then $[M : N]_{(\mathfrak{p})}$ is generated by $\det(\delta_{\mathfrak{p}})$.

The second statement follows in a similar way. We may assume that R is local and thus $N \subseteq M$ are free, in which case $M = N$ if and only if a change of basis matrix from N to M has determinant in $R^\times$. □

For Dedekind domains, the R-index measure the discrepancy between M and N, as follows.

Lemma 13.2.15. *If R is a Dedekind domain and $N \subseteq M$, then $[M : N]$ is the product of the invariant factors (or elementary divisors) of the torsion R-module M/N.*

Proof. Exercise 12.3. □

We can now compare discriminants and indices, as in the following lemma.

Lemma 13.2.16. *Let $\mathcal{O}, \mathcal{O}'$ be projective R-orders. Then*

$$\operatorname{disc}(\mathcal{O}) = [\mathcal{O}' : \mathcal{O}]^2 \operatorname{disc}(\mathcal{O}');$$

if $\mathcal{O} \subseteq \mathcal{O}'$, then equality holds if and only if $\mathcal{O} = \mathcal{O}'$.

Proof. For the first statement, we argue locally, and combine (13.2.3) and Paragraph 13.2.13. For the second statement, clearly $\operatorname{disc}(\mathcal{O}') \subseteq \operatorname{disc}(\mathcal{O})$, and if $\mathcal{O} = \mathcal{O}'$ then equality holds; and conversely, if $\operatorname{disc}(\mathcal{O}) = [\mathcal{O}' : \mathcal{O}]^2 \operatorname{disc}(\mathcal{O}') = \operatorname{disc}(\mathcal{O}')$ then $[\mathcal{O}' : \mathcal{O}] = R$, hence $\mathcal{O}' = \mathcal{O}$. □

Using the discriminant as a measure of index, we can ensure the existence of maximal orders in this general context as follows.

Proposition 13.2.17. *There exists a maximal order $\mathcal{O} \subseteq B$.*

Proof. This proof does not assume Zorn's lemma, but compare the proof with Remark 8.4.2).

B has an order $\mathcal{O}$ by Paragraph 8.3.5. If $\mathcal{O}$ is not maximal, there exists an order $\mathcal{O}' \supsetneq \mathcal{O}$ with $\operatorname{disc}(\mathcal{O}') \mid \operatorname{disc}(\mathcal{O})$. If $\mathcal{O}'$ is maximal, we are done; otherwise, we can continue in this way to obtain orders $\mathcal{O} = \mathcal{O}_1 \subsetneq \mathcal{O}_2 \subsetneq \dots$ and an ascending chain of ideals $\operatorname{disc}(\mathcal{O}_1) \subsetneq \operatorname{disc}(\mathcal{O}_2) \subsetneq \dots$ of R; but since R is noetherian, the latter terminates, a contradiction. □

13.3 Reduced discriminant

Now suppose that B is a quaternion algebra over F. We have already seen that the index of one order in another yields a difference in the index by a square. In fact, the discriminant itself is always a square, and so by defining a reduced discriminant as a square root of this ideal, we avoid this unnecessary exponent.

In fact, there is a way to define this square root directly.

13.3.1. For $\alpha_1, \alpha_2, \alpha_3 \in B$, we define

$$\{\alpha_1, \alpha_2, \alpha_3\} = \operatorname{trd}((\alpha_1\alpha_2 - \alpha_2\alpha_1)\overline{\alpha_3}) = (\alpha_1\alpha_2 - \alpha_2\alpha_1)\overline{\alpha_3} - \alpha_3(\overline{\alpha_2}\,\overline{\alpha_1} - \overline{\alpha_1}\,\overline{\alpha_2}).$$

[[Brzezinski: if i, j, k have trace zero, then $\{i, j, k\} = \operatorname{trd}(ijk)$. In any case, give some motivation! It comes from the fact that the lie algebra of the quaternion algebra has a symmetric nondegenerate invariant bilinear form.]]

Lemma 13.3.2. *The form* $\{\ \} : B \times B \times B \to F$ *is an alternating trilinear form which descends to such a pairing on* B/F.

Proof. The form is alternating because clearly $\{\alpha_1, \alpha_1, \alpha_2\} = 0$ and

$$\{\alpha_1, \alpha_2, \alpha_1\} = \operatorname{trd}((\alpha_1\alpha_2 - \alpha_2\alpha_1)\overline{\alpha_1}) = \operatorname{trd}(\operatorname{nrd}(\alpha_1)\alpha_2) - \operatorname{trd}(\alpha_2 \operatorname{nrd}(\alpha_1)) = 0$$

and similarly $\{\alpha_1, \alpha_2, \alpha_2\} = 0$ for all $\alpha_1, \alpha_2 \in B$. The trilinearity follows from the linearity of the reduced trace. Finally, from these two properties, the descent to B/F follows from the computation $\{1, \alpha_1, \alpha_2\} = 0$ for all $\alpha_1, \alpha_2 \in B$.

(Alternatively, one can check that the pairing descends to B/F first, so that the involution becomes $\overline{\alpha + F} = -\alpha + F$, and then the alternating condition is immediate.) □

13.3.3. If $\beta_i = M\alpha_i$ for some $M \in \mathrm{M}_3(F)$ and $\alpha_i, \beta_i \in B$, then

$$\{\beta_1, \beta_2, \beta_3\} = \det(M)\{\alpha_1, \alpha_2, \alpha_3\} \tag{13.3.4}$$

by Exercise 13.6. It follows that if $\mathcal{O} \subseteq \mathcal{O}'$ are projective R-orders in B, then

$$\operatorname{discrd}(\mathcal{O}) = [\mathcal{O}' : \mathcal{O}] \operatorname{discrd}(\mathcal{O}').$$

Definition 13.3.5. The *reduced discriminant* of an R-order $\mathcal{O} \subseteq B$ is the R-submodule $\operatorname{discrd}(\mathcal{O})$ of F generated by

$$\{\{\alpha_1, \alpha_2, \alpha_3\} : \alpha_1, \alpha_2, \alpha_3 \in \mathcal{O}\}.$$

Lemma 13.3.6. *If* $\mathcal{O}$ *is a projective* R*-order in* B*, then* $\operatorname{disc}(\mathcal{O}) = \operatorname{discrd}(\mathcal{O})^2$.

Proof. First, we claim that

$$\{i, j, ij\}^2 = -d(1, i, j, ij).$$

If $\operatorname{char} F \neq 2$, then we have $\operatorname{disc}(1, i, j, ij) = -(4ab)^2$ by Example 13.2.5 and

$$\{i, j, ij\} = \operatorname{trd}((ij - ji)\overline{ij}) = \operatorname{trd}(2ij(\overline{ij})) = 4ab,$$

as claimed. See Exercise 13.1 for the case $\operatorname{char} F = 2$. (This computation shows verifies the result for the order $\mathcal{O} = R \oplus Ri \oplus Rj \oplus Rij$.) The lemma now follows from (13.2.3) and (13.3.4), for it shows that

$$\{\alpha_1, \alpha_2, \alpha_3\}^2 = -d(1, \alpha_1, \alpha_2, \alpha_3)$$

for all $\alpha_1, \alpha_2, \alpha_3 \in B$, and Exercise 13.5. □

Remark 13.3.7. [[The definition of the reduced discriminant for a general semisimple algebra is the reduced norm of the different?]]

13.4 Extensions and further reading

13.4.1. We have defined the discriminant only for a separable algebra because then we can ensure that the discriminant is nonzero (see Lemma 13.2.11 below). More generally, can take algebra trace instead of reduced trace.

13.4.2. Fitting ideal, in place of $[J : I]$.

13.4.3. An order in a separable algebra over a Dedekind domain has a reduced discriminant.

Exercises

Let R be a noetherian domain with field of fractions F.

13.1. Let $\operatorname{char} F = 2$ and let $\left[\dfrac{a,b}{F}\right)$ be a quaternion algebra over F with $a, b \in R$ and $b \neq 0$. Show that $\mathcal{O} = R + Ri + Rj + Rij$ is an R-order in B and compute the (reduced) discriminant of $\mathcal{O}$.

13.2. Let B be a division quaternion algebra over a nonarchimedean local field F with uniformizer π, and let $\mathcal{O}$ be the valuation ring of B. Show that $\operatorname{discrd}(\mathcal{O}) = (\pi)$.

13.3. Let $B = \mathrm{M}_n(F)$ and $\mathcal{O} = \mathrm{M}_n(R)$ with $n \geq 1$. Show that $\operatorname{disc}(\mathcal{O}) = R$. *[Hint: Compute directly on a basis of matrix units.]*

13.4. Let $B = \left(\dfrac{K,b}{F}\right)$ with $b \in R$ and let S be an R-order in K. Let $\mathcal{O} = S + Sj$; then $\mathcal{O}$ is an R-order in B by Exercise 8.7. Let $\mathfrak{d} = \operatorname{disc}(S)$. Show that $\operatorname{disc}(\mathcal{O}) = (b\mathfrak{d})^2$.

13.5. Let $\mathcal{O}$ be an R-order. Show that $\operatorname{disc}(\mathcal{O})$ is generated by

$$\{d(1, \alpha_1, \dots, \alpha_{n-1}) : \alpha_1, \dots, \alpha_{n-1} \in \mathcal{O}\}.$$

13.6. Let B be a quaternion algebra over F. Define $\{\,\} : B \times B \times B \to F$ by $\{\alpha_1, \alpha_2, \alpha_3\} = \operatorname{trd}([\alpha_1, \alpha_2]\overline{\alpha_3})$ for $\alpha_i \in B$. If $\beta_i = M\alpha_i$ for some $M \in \mathrm{M}_3(F)$, show that

$$\{\beta_1, \beta_2, \beta_3\} = \det(M)\{\alpha_1, \alpha_2, \alpha_3\}.$$

13.7. Let I, J be R-lattices in an F-vector space V. Show that the index $[I : J]$ is a nonzero R-module.

13.8. Show that if I is an R-lattice in B then the dual $I^\sharp$ is an R-lattice in B.

Chapter 14

Quaternion ideals over Dedekind domains

14.1 Ideals and modules

Much like a space can be understood by studying functions on that space (the subject of functional analysis), often the first task to understand a ring A is to understand the ideals of A and modules over A (the subject of commutative algebra). The simplest ideals of a ring are the principal ideals—but not all ideals are principal, and various algebraic structures are built to understand the difference between these two. In the next section, we consider such structures; we begin here by considering just the ideals themselves. To get warmed up for the noncommutative situation, we consider first the simple case of quadratic rings.

Let $D \in \mathbb{Z}$ be a nonsquare *discriminant*, so $D \equiv 0, 1 \pmod 4$. Let S be the quadratic order of nonsquare discriminant $D \in \mathbb{Z}$, namely,

$$S = S(D) = \mathbb{Z} \oplus \mathbb{Z}[(D + \sqrt{D})/2] \subset K = \mathbb{Q}(\sqrt{D}).$$

The set of ideals of S has a natural commutative multiplication structure with identity element S (it has the structure of a monoid), but this set lacks inverses and we would surely feel more comfortable with a group structure. So more generally we consider S-lattices $\mathfrak{a} \subset K$, and call them *fractional ideals* of S. To get some kind of group structure, we must restrict our attention to the *invertible* fractional ideals $\mathfrak{a} \subset K$, i.e., those such that there exists a fractional ideal $\mathfrak{b}$ with $\mathfrak{a}\mathfrak{b} = S$. If $\mathfrak{a}$ has an inverse then this inverse is unique, given by $\mathfrak{a}^{-1} = \{x \in F : x\mathfrak{a} \subseteq S\}$. If S is the ring of integers of K (the maximal order), then all nonzero fractional ideals of S are invertible—in fact, this property characterizes Dedekind domains, in that a noetherian commutative ring is a Dedekind domain if and only if every nonzero (prime) ideal is invertible.

For some purposes, one might as well assume S is a Dedekind domain and leave the subtler issues aside.

In the quaternionic generalization, noncommutativity poses some thorny issues. Let B be a quaternion algebra over $\mathbb{Q}$ and let $\mathcal{O} \subset B$ be an order. Right away, to study ideals of $\mathcal{O}$ we must distinguish between left or right ideals, and the product of two (say) right $\mathcal{O}$-ideals need not be again a right $\mathcal{O}$-ideal! To address these issues, for $\mathbb{Z}$-lattices $I, J \subset B$, we say that I is *compatible* with J if the right order of I is equal to the left order of J, so that what comes between I and J in the product $I \cdot J$ "matches up".

A $\mathbb{Z}$-lattice $I \subset B$ is *invertible* if there exists a $\mathbb{Z}$-lattice I' such that $II' = \mathcal{O}_{\mathrm{L}}(I) = \mathcal{O}_{\mathrm{R}}(I')$ and $I'I = \mathcal{O}_{\mathrm{R}}(I) = \mathcal{O}_{\mathrm{L}}(I')$, so in particular both of these products are compatible. The simplest kind of invertible lattices are those the principal lattices $I = \mathcal{O}_{\mathrm{L}}(I)\alpha = \alpha\mathcal{O}_{\mathrm{R}}(I)$. More generally, all lattices I with a maximal left or right order are invertible. Therefore, if one is not interested in the more subtle algebraic issues of invertibility, one can restrict to working over a maximal order $\mathcal{O}$.

The set of invertible $\mathbb{Z}$-lattices in B under compatible product has only a partial product defined: and so Brandt coined the term *groupoid* (*gruppoid*) for this kind of object, a nonempty set with an inverse function and a partial product that satisfies the associativity, inverse, and identity properties whenever they are defined. (Despite this humble beginning, groupoids now figure prominently in category theory, as well as many other contexts. For a category theorist, a groupoid is a small category such that every morphism is an isomorphism.)

In addition to studying compatible products, the major subject of this chapter will be pinning down the notion of invertibility to make it seem both natural and concrete. Recall that a fractional ideal $\mathfrak{a}$ of S is invertible if and only if $\mathfrak{a}$ is *locally principal*, i.e., $\mathfrak{a} \otimes_{\mathbb{Z}} \mathbb{Z}_{(p)} = \mathfrak{a}_{(p)}$ is a principal fractional ideal of the localization $S_{(p)}$ for all primes p. Every locally principal ideal is invertible, and the extent to which the converse is something that arises in an important way more generally in algebraic geometry (comparing Weil and Cartier divisors on a scheme). In the language of commutative algebra, a locally principal S-module is equivalently a projective S-module of rank 1. Each of these characterizations has its uses.

If $S(D)$ is not maximal, so that $D = df^2$ with $d \in \mathbb{Z}$ a fundamental discriminant and $f \in \mathbb{Z}_{>1}$ the *conductor* of S, then there is always an ideal of S that is not invertible. Specifically, consider the ideal $\mathfrak{f} = (f, \sqrt{D})$. Then $\mathfrak{f}$ is a free $\mathbb{Z}$-module of rank 2, with

$$\mathfrak{f} = f\mathbb{Z} + \sqrt{D}\mathbb{Z} = f(\mathbb{Z} + \sqrt{D}\mathbb{Z}) + \sqrt{D}(\mathbb{Z} + \sqrt{D}\mathbb{Z}).$$

Then

$$\mathfrak{f}^2 = (f\mathbb{Z} + \mathbb{Z}\sqrt{D})^2 = f^2\mathbb{Z} + f\sqrt{D}\mathbb{Z} = f\mathfrak{f}$$

so if $\mathfrak{f}$ were invertible with inverse $\mathfrak{f}^{-1}$, then

$$\mathfrak{f}^2\mathfrak{f}^{-1} = \mathfrak{f} = (f) \subseteq S(D) \tag{14.1.1}$$

but this is a contradiction, since

$$(f) = f\mathbb{Z} + f\sqrt{D}\mathbb{Z} \neq \mathfrak{f} = f\mathbb{Z} + \sqrt{D}\mathbb{Z}$$

whenever $f > 1$.

This example also suggests the real issue with noninvertible modules for quadratic orders. As an abelian group, we have

$$\mathfrak{f} = f\mathbb{Z} + f\sqrt{d}\mathbb{Z} = f \cdot S(d),$$

so $\mathfrak{f}$ is principal and hence certainly invertible as an ideal of $S(d)$—but not as an ideal of the smaller order $S(D)$. More generally, if $\mathfrak{a} \subset K = \mathbb{Q}(\sqrt{D})$ is a $\mathbb{Z}$-lattice in K (free $\mathbb{Z}$-module of rank 2), we define its *multiplicator ring* as

$$S(\mathfrak{a}) = \{a \in K : a \cdot \mathfrak{a} \subseteq \mathfrak{a}\};$$

the ring $S(\mathfrak{a})$ is an order of K and so is also called the *order* of $\mathfrak{a}$. For example, in the example above we have $S(\mathfrak{a}) = S(d)$. (Sometimes, an ideal $\mathfrak{a} \subseteq S$ is called *proper* if $S = S(\mathfrak{a})$; the term *proper* is quite overloaded in mathematics, so we will mostly resist this notion.) It turns out that every $\mathbb{Z}$-lattice in K is projective over its multiplication ring, and this statement plays an important role in the theory of complex multiplication.

The major task of this chapter will be to investigate this notion of invertibility in a general quaternionic context. We will prove the following theorem.

Theorem 14.1.2. *Let $I \subset B$ be a $\mathbb{Z}$-lattice. Then the following are equivalent:*

(i) *I is invertible (as a $\mathbb{Z}$-lattice);*

(ii) *I is locally principal (as a $\mathbb{Z}$-lattice), i.e., $I_{(p)}$ is principal for all primes p; and*

(iii) *I is projective as a left $\mathcal{O}_{\mathrm{L}}(I)$-module and as a right $\mathcal{O}_{\mathrm{R}}(I)$-module.*

In this theorem, we consider only lattices as (proper) modules over their full left and right orders. Unlike the quadratic case, not every lattice $I \subset B$ is projective as a left module over its left order (or with right). In Chapter 16, we classify orders $\mathcal{O}$ with the property that every lattice I with $\mathcal{O}_{\mathrm{L}}(I) = \mathcal{O}$ is projective as an $\mathcal{O}$-module.

14.2 Locally principal lattices

The simplest lattices to understand are those that are principal; but as we saw in Chapter 12, lattices are inherently local in nature. So instead we are led to consider the more general class of locally principal lattices. We work first work with lattices yet unattached to an order, and later we will sort them by their left and right orders.

To begin, we can work in quite some generality. Let R be a noetherian domain with field of fractions F, let B be a finite-dimensional F-algebra, and let I be an R-lattice in B.

Definition 14.2.1. An R-lattice I is *principal* if there exists $\alpha \in B$ such that $I = \mathcal{O}_{\mathrm{L}}(I)\alpha = \alpha\mathcal{O}_{\mathrm{R}}(I)$; we say that I is *generated* by α.

If I is generated by $\alpha \in B$, then necessarily $\alpha \in B^\times$ since otherwise $IF = B\alpha \subsetneq B$, contradicting the fact that I is a lattice.

14.2.2. If $I = \mathcal{O}_{\mathrm{L}}(I)\alpha$, then $\mathcal{O}_{\mathrm{R}}(I) = \alpha^{-1}\mathcal{O}_{\mathrm{L}}(I)\alpha$, by Exercise 14.3, so automatically

$$I = \alpha(\alpha^{-1}\mathcal{O}_{\mathrm{L}}(I)\alpha) = \alpha\mathcal{O}_{\mathrm{R}}(I).$$

Therefore it is sufficient to check for a one-sided generator (and if we defined the obvious notions of *left principal* or *right principal*, these would be equivalent to principal).

Definition 14.2.3. An R-lattice I is *integral* if $I \subseteq \mathcal{O}_{\mathrm{L}}(I) \cap \mathcal{O}_{\mathrm{R}}(I)$.

Remark 14.2.4. If $I \subseteq \mathcal{O}_{\mathrm{L}}(I)$ then already $II \subseteq I$ so $I \subseteq \mathcal{O}_{\mathrm{R}}(I)$ as well. Hence I is integral if and only if $I \subseteq \mathcal{O}_{\mathrm{L}}(I)$ if and only if $I \subseteq \mathcal{O}_{\mathrm{R}}(I)$. (And hence we do not define notions of left and right integral.)

If I is integral, then every element of I is integral over R (Lemma 8.3.8).

14.2.5. An R-lattice I is integral if and only if I is a right ideal of $\mathcal{O}_{\mathrm{R}}(I)$ and a left ideal of $\mathcal{O}_{\mathrm{L}}(I)$ in the usual sense.

For any R-lattice I, there exists $d \in R \setminus \{0\}$ such that dI is integral, so any R-lattice $I = (dI)/d$ is *fractional* in the sense that it is obtained from an integral lattice with denominator.

Definition 14.2.6. Let $\mathcal{O} \subset B$ be an order. A *left* (resp. *right*) *fractional $\mathcal{O}$-ideal* is a lattice $I \subset B$ such that $\mathcal{O} \subseteq \mathcal{O}_{\mathrm{L}}(I)$ (resp. $\mathcal{O} \subseteq \mathcal{O}_{\mathrm{R}}(I)$).

Next, we extend the reduced norm to lattices.

Definition 14.2.7. The *reduced norm* $\operatorname{nrd}(I)$ of I is the R-submodule of F generated by the set $\{\operatorname{nrd}(\alpha) : \alpha \in I\}$.

Lemma 14.2.8. *The R-module* $\operatorname{nrd}(I)$ *is finitely generated.*

Proof. Since I is an R-lattice, we have $I = \sum_i R\alpha_i$ generated by a finite set $\{\alpha_i\}_i$ as an R-module. The R-module $\operatorname{nrd}(I)$ is then generated by the values $a_{ii} = \operatorname{nrd}(\alpha_i)$ and $a_{ij} = \operatorname{nrd}(\alpha_i + \alpha_j)$, since

$$\operatorname{nrd}\left(\sum_i a_i\alpha_i\right) = \sum_{i \leq j} a_{ij}\alpha_i\alpha_j$$

for $a_i \in R$ as nrd is a quadratic form. □

By the local-global property of lattices (Lemma 12.2.2), we have

$$\operatorname{nrd}(I) = \bigcap_{\mathfrak{p}} \operatorname{nrd}(I_{\mathfrak{p}}). \tag{14.2.9}$$

14.2.10. If I is a principal R-lattice generated by $\alpha \in I$ then $\operatorname{nrd}(I) = \operatorname{nrd}(\alpha)R$; more generally, if I is an R-lattice and $\alpha \in B^\times$ then $\operatorname{nrd}(\alpha I) = \operatorname{nrd}(\alpha)\operatorname{nrd}(I)$ (Exercise 14.9).

The notion of principality naturally extends locally.

Definition 14.2.11. An R-lattice I is *locally principal* if $I_{(\mathfrak{p})} = I \otimes_R R_{(\mathfrak{p})}$ is a principal $R_{(\mathfrak{p})}$-lattice for all primes $\mathfrak{p}$ of R.

We will show in the following subsections that the properties of being locally principal, projective, and (in the presence of a standard involution) invertible are all equivalent for a lattice I.

14.3 Compatible lattices

Now let I, J be R-lattices in B. We define the product IJ to be the R-submodule of B generated by the set

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

The product IJ is an R-lattice: it is finitely generated as this is true of I, J individually, and there exists a nonzero $r \in I$ (Exercise 8.2) so $rJ \subset IJ$ and thus $B = FJ = F(rJ) \subseteq F(IJ)$.

If I, J are R-lattices then we have $\operatorname{nrd}(IJ) \supseteq \operatorname{nrd}(I)\operatorname{nrd}(J)$. However, we need not have equality, as the following example indicates.

Example 14.3.1. It is not always true that $\operatorname{nrd}(IJ) = \operatorname{nrd}(I)\operatorname{nrd}(J)$. For example, if $a \in R$ is neither zero nor a unit, then $I = \begin{pmatrix} (a) & R \\ (a) & R \end{pmatrix}$ and $J = \begin{pmatrix} (a) & (a) \\ R & R \end{pmatrix}$ are R-lattices in $\mathrm{M}_2(F)$ with $\operatorname{nrd}(I) = \operatorname{nrd}(J) = (a)$ but $IJ = \mathrm{M}_2(R)$ and so $\operatorname{nrd}(IJ) = R$. However, $\operatorname{nrd}(JI) = (a)^2 = \operatorname{nrd}(J)\operatorname{nrd}(I)$.

The issue present in Example 14.3.1 is that the product is not as well-behaved for noncommutative rings as for commutative rings; we need the elements coming between I and J to "match up".

Definition 14.3.2. We say that I is *compatible* with J if $\mathcal{O}_{\mathrm{R}}(I) = \mathcal{O}_{\mathrm{L}}(J)$.

We will also sometimes just say that the product IJ is *compatible* to mean that I is compatible with J.

The relation "compatible with" is in general *not* a symmetric relation.

Example 14.3.3. Continuing with Example 14.3.1, we find that $\mathcal{O}_{\mathrm{R}}(J) = \mathrm{M}_2(R) = \mathcal{O}_{\mathrm{L}}(I)$, so J is compatible with I; but

$$\mathcal{O}_{\mathrm{R}}(I) = \begin{pmatrix} R & (a^{-1}) \\ (a) & R \end{pmatrix} \quad \text{and} \quad \mathcal{O}_{\mathrm{L}}(J) = \begin{pmatrix} R & (a) \\ (a^{-1}) & R \end{pmatrix},$$

so I is not compatible with J.

Lemma 14.3.4. *Suppose that I is compatible with J and that either I or J is locally principal. Then* $\operatorname{nrd}(IJ) = \operatorname{nrd}(I)\operatorname{nrd}(J)$.

Proof. By the local–global property for norms (14.2.9) and since localization commutes with multiplication, i.e.,

$$(\mathfrak{ab})_{(\mathfrak{p})} = \mathfrak{a}_{(\mathfrak{p})}\mathfrak{b}_{(\mathfrak{p})} \text{ for all (finitely generated) } R\text{-modules } \mathfrak{a}, \mathfrak{b} \subseteq F,$$

we may assume that either I or J is principal. Suppose I is (right) principal. Then $I = \alpha\mathcal{O}$ for some $\alpha \in B$ where $\mathcal{O} = \mathcal{O}_{\mathrm{R}}(I) = \mathcal{O}_{\mathrm{L}}(J)$. Then

$$IJ = (\alpha\mathcal{O})J = \alpha(\mathcal{O}J) = \alpha J$$

and so $\operatorname{nrd}(IJ) = \operatorname{nrd}(\alpha)\operatorname{nrd}(J) = \operatorname{nrd}(I)\operatorname{nrd}(J)$ by Paragraph 14.2.10. The case where J is principal follows in the same way. □

14.4 Invertible lattices

We are now in a position to investigate the class of invertible lattices. Let $I \subseteq B$ be an R-lattice.

Definition 14.4.1. An R-lattice I is *invertible* if there exists an R-lattice $I' \subseteq B$ that is a *(two-sided) inverse* to I, i.e.

$$II' = \mathcal{O}_{\mathrm{L}}(I) = \mathcal{O}_{\mathrm{R}}(I') \text{ and } I'I = \mathcal{O}_{\mathrm{R}}(I) = \mathcal{O}_{\mathrm{L}}(I'). \tag{14.4.2}$$

In particular, both of the products in (14.4.2) are compatible.

14.4.3. If I, J are invertible lattices and I is compatible with J, then IJ is an invertible lattice (Exercise 14.6).

14.4.4. If I is a principal lattice, then I is invertible: if $I = \mathcal{O}\alpha$ with $\alpha \in B^\times$ and $\mathcal{O} = \mathcal{O}_{\mathrm{L}}(I)$, then $I' = \alpha^{-1}\mathcal{O}$ has

$$II' = (\mathcal{O}\alpha)(\alpha^{-1}\mathcal{O}) = \mathcal{O}(\alpha\alpha^{-1})\mathcal{O} = \mathcal{O}\mathcal{O} = \mathcal{O}$$

so I' is a right inverse, and

$$I'I = (\alpha^{-1}\mathcal{O})(\mathcal{O}\alpha) = \alpha^{-1}\mathcal{O}\alpha = \mathcal{O}_{\mathrm{R}}(I)$$

so I' is also a left inverse.

A candidate for the inverse presents itself quite naturally. If $II' = \mathcal{O}_{\mathrm{L}}(I)$ and $I'I = \mathcal{O}_{\mathrm{R}}(I)$, then $II'I = I$. So we define

$$I^{-1} = \{\alpha \in B : I\alpha I \subseteq I\}. \tag{14.4.5}$$

The same proof as in Paragraph 8.3.5 implies that I^{-1} is an R-lattice.

Proposition 14.4.6. *The following are equivalent:*

(i) *I^{-1} is a (two-sided) inverse for I;*

(ii) *I is invertible; and*

(iii) *We have a compatible product $II^{-1}I = I$ and both $1 \in II^{-1}$ and $1 \in I^{-1}I$.*

Proof. The implication (i) ⇒ (ii) is clear. For the statement (ii) ⇒ (i), suppose that I' is an inverse to I. Then $I = II'I \subseteq II^{-1}I \subseteq I$, so equality holds throughout. Multiplying by I' on the left and right gives $(I'I)I^{-1}(II') = I'II'$ whence $I^{-1} = I'$.

Again the implication (i) ⇒ (iii) is immediate. To prove (iii) ⇒ (i), we need to show that $II^{-1} = \mathcal{O}_{\mathrm{L}}(I)$ and $I^{-1}I = \mathcal{O}_{\mathrm{R}}(I)$; we show the former. By compatibility, we have $\mathcal{O}_{\mathrm{R}}(I^{-1}) = \mathcal{O}_{\mathrm{L}}(I) = \mathcal{O}$ so if $II^{-1} = J$ then $J = II^{-1} = \mathcal{O}(II^{-1})\mathcal{O} = \mathcal{O}J\mathcal{O}$, so $J \subseteq \mathcal{O}$ is a two-sided ideal of $\mathcal{O}$ containing 1 hence $J = \mathcal{O}$. □

Not every lattice is invertible, and it is helpful to have a counterexample at hand.

Example 14.4.7. Let $\mathfrak{a} \subsetneq R$ be a nonzero, proper ideal, and for simplicity let $\mathcal{O} = \mathrm{M}_2(R) \subset B = \mathrm{M}_2(F)$. Let $I = \mathfrak{a} + \mathcal{O}$. Then $\mathcal{O}_\mathrm{L}(I) = \mathcal{O}_\mathrm{R}(I) = R + \mathfrak{a}\mathcal{O}$, and $I^{-1} = \mathcal{O}_\mathrm{L}(I) = \mathcal{O}_\mathrm{R}(I)$ and yet $II^{-1} = I \neq \mathcal{O}_\mathrm{L}(I)$ (and similarly on the right), so I is not invertible. Indeed, we have a compatible product $II^{-1}I = I$, but in Proposition 14.4.6(iii) we have $1 \notin II^{-1} = I$. Note interestingly that the compatible product

$$I^2 = II = \mathfrak{a}^2 + \mathfrak{a}\mathcal{O} + \mathcal{O} = \mathcal{O} \tag{14.4.8}$$

has $\mathcal{O}_\mathrm{L}(I^2) = \mathcal{O}_\mathrm{R}(I^2) = \mathcal{O} \supsetneq \mathcal{O}_\mathrm{L}(I) = \mathcal{O}_\mathrm{R}(I)$.

A compatible product with an invertible lattice (compare (14.4.8)) respects taking left (and right) orders, as follows.

Lemma 14.4.9. *If I is compatible with J and J is invertible, then $\mathcal{O}_\mathrm{L}(IJ) = \mathcal{O}_\mathrm{L}(I)$.*

Proof. We prove the first statement. Without the hypothesis of compatibility, we have $\mathcal{O}_\mathrm{L}(I) \subseteq \mathcal{O}_\mathrm{L}(IJ)$. To show the other containment, suppose that $\alpha \in \mathcal{O}_\mathrm{L}(IJ)$, so that $\alpha IJ \subseteq IJ$. Multiplying by J^{-1}, we conclude $\alpha I \subseteq I$, so $\alpha \in \mathcal{O}_\mathrm{L}(I)$. □

Invertibility is a local property, as one would expect.

Lemma 14.4.10. *I is invertible if and only $I_{(\mathfrak{p})}$ is invertible for all primes $\mathfrak{p}$.*

Proof. The implication (⇒) follows from the fact that taking products commutes with localization. For the other implication, we show right invertibility.

Let $\mathcal{O} = \mathcal{O}_\mathrm{L}(I)$ and $J = (\mathcal{O} : I)_\mathrm{R}$; then J is an R-lattice (14.7.4) and $IJ \subseteq \mathcal{O}$. Since $I_{(\mathfrak{p})}$ is right invertible, there exists an $R_{(\mathfrak{p})}$-lattice $I'_{(\mathfrak{p})}$ such that $I_{(\mathfrak{p})}I'_{(\mathfrak{p})} = \mathcal{O}_{(\mathfrak{p})}$. But then

$$\mathcal{O}_{(\mathfrak{p})} = I_{(\mathfrak{p})}I'_{(\mathfrak{p})} \subseteq I_{(\mathfrak{p})}J_{(\mathfrak{p})} \subseteq \mathcal{O}_{(\mathfrak{p})}$$

so equality holds, and thus $I_{(\mathfrak{p})}J_{(\mathfrak{p})} = \mathcal{O}_{(\mathfrak{p})}$ for all $\mathfrak{p}$. Intersecting, we obtain $IJ = \mathcal{O}$, so I is right invertible. □

Corollary 14.4.11. *If I is locally principal, then I is invertible.*

To conclude this section, we note the multiplicative structure on the set of invertible lattices.

Definition 14.4.12. A *groupoid* G is a set with a unary operation $^{-1}$ and a *partial function* $* : G \to G$ such that $*$ and $^{-1}$ satisfy the associativity, inverse, and identity properties (as in a group) whenever they are defined.

1. [Associativity] For all $a, b, c \in G$, such that $a * b$ is defined and $(a * b) * c$ is defined, we have that $b * c$ and $a * (b * c)$ is defined and the equality
$$(a * b) * c = a * (b * c)$$
holds.

2. [Inverses] For all $a \in G$, there exists $a^{-1} \in G$ such that $a * a^{-1}$ and $a^{-1} * a$ are defined.

3. [Identity] For all $a, b \in G$ such that $a * b$ is defined, we have
$$(a * b) * b^{-1} = a \quad \text{and} \quad a^{-1} * (a * b) = b. \tag{14.4.13}$$

A *homomorphism* $\phi : G \to G'$ of groupoids is a map satisfying $\phi(a * b) = \phi(a) * \phi(b)$ for all $a, b \in G$.

The products in the identity law (14.4.13) are defined by the associative and inverse laws, and it follows that $e = a * a^{-1}$ and $f = a^{-1} * a$ satisfy $e * a = a = a * f$. (We may have that $e \neq f$, i.e., the left and right identities for a disagree.)

Remark 14.4.14. Equivalently, a groupoid is a small category (the class of objects in the category is a set) such that every morphism is an isomorphism.

Example 14.4.15. The set of homotopy classes of paths in a topological space X forms a groupoid under composition: the paths $\gamma_1, \gamma_2 : [0, 1] \to X$ can be composed to a path $\gamma_2 \circ \gamma_1 : [0, 1] \to X$ if and only if $\gamma_2(0) = \gamma_1(1)$.

Proposition 14.4.16. *The set of invertible lattices in B form a groupoid under compatible product.*

Proof. For the associative law, suppose I, J, K are invertible R-lattices with IJ and $(IJ)K$ compatible products. Then $\mathcal{O}_{\mathrm{R}}(I) = \mathcal{O}_{\mathrm{L}}(J) = \mathcal{O}_{\mathrm{L}}(JK)$ and $\mathcal{O}_{\mathrm{R}}(IJ) = \mathcal{O}_{\mathrm{R}}(J) = \mathcal{O}_{\mathrm{L}}(K)$ by Lemma 14.4.9, so the products JK and $I(JK)$ are compatible. Multiplication is associative in B, and it follows that $I(JK) = (IJ)K$. Inverses exist exactly because we restrict to the invertible lattices. The law of identity holds as follows: if I, J are invertible R-lattices such that IJ is a compatible product, then $(IJ)J^{-1}$ is a compatible product since $\mathcal{O}_{\mathrm{R}}(IJ) = \mathcal{O}_{\mathrm{R}}(J) = \mathcal{O}_{\mathrm{L}}(J^{-1})$, and by associativity we have
$$(IJ)J^{-1} = I(JJ^{-1}) = I\mathcal{O}_{\mathrm{L}}(J) = I\mathcal{O}_{\mathrm{R}}(I) = I,$$
with a similar argument on the left. □

14.5 Invertibility with a standard involution

The main result of this section is as follows.

Theorem 14.5.1. *Let R be a Dedekind domain and suppose that B has a standard involution. Then an R-lattice I is invertible if and only if I is locally principal.*

We have already seen (Corollary 14.4.11) the implication ($\Rightarrow$) in Theorem 14.5.1; the remaining implication is the topic of this section.

One reason to suppose that R is a Dedekind domain is the following: if $\mathfrak{a} \subset R$ is not invertible as an R-module, and $\mathcal{O} \subset B$ is any R-order, then $\mathfrak{a}\mathcal{O}$ is not invertible as an R-lattice. To make the simplest kind of arguments here, we would like for all (nonzero) ideals $\mathfrak{a} \subseteq R$ to be invertible, and this is equivalent to the requirement that R is a Dedekind domain.

So suppose B has a standard involution $\overline{}$ and that R is a Dedekind domain. Let $I \subset B$ be an R-lattice. The following concept will be useful in this section.

Definition 14.5.2. We say I is a *semi-order* if $1 \in I$ and $\operatorname{nrd}(I) \subseteq R$.

Lemma 14.5.3. *An R-lattice I is a semi-order if and only if $1 \in I$ and every $\alpha \in I$ is integral over R.*

Proof. We have that $\alpha \in I$ is integral over R if and only if $\operatorname{trd}(\alpha) \in R$ and $\operatorname{nrd}(\alpha) \in R$ (by Corollary 8.3.11, since R is integrally closed) if and only if $\operatorname{nrd}(\alpha) \in R$ and $\operatorname{nrd}(\alpha + 1) = \operatorname{nrd}(\alpha) + \operatorname{trd}(\alpha) + 1 \in R$. □

In particular, Lemma 14.5.3 implies that an order is a semi-order (by Corollary 8.3.9); we will see that semi-orders behave enough like orders that we can deduce local principality from their structure.

14.5.4. Let $\overline{I} = \{\overline{\alpha} : \alpha \in I\}$. Then $\overline{I}$ is an R-lattice in B. If I, J are R-lattices then $\overline{IJ} = \overline{J}\,\overline{I}$ (even if this product is not compatible).

If I is a semi-order, then $\overline{I} = I$ (Exercise 14.7). In particular, if $\mathcal{O}$ is an R-order then $\overline{\mathcal{O}} = \mathcal{O}$.

Lemma 14.5.5. *We have $\mathcal{O}_{\mathrm{L}}(I) = \mathcal{O}_{\mathrm{R}}(\overline{I})$ and $\mathcal{O}_{\mathrm{R}}(I) = \mathcal{O}_{\mathrm{L}}(\overline{I})$.*

Proof. We have $\alpha \in \mathcal{O}_{\mathrm{L}}(I)$ if and only if $\alpha I \subseteq I$ if and only if $\overline{\alpha I} = \overline{I}\,\overline{\alpha} \subseteq \overline{I}$ if and only if $\overline{\alpha} \in \mathcal{O}_{\mathrm{R}}(\overline{I})$ if and only if $\alpha \in \overline{\mathcal{O}_{\mathrm{R}}(\overline{I})} = \mathcal{O}_{\mathrm{R}}(\overline{I})$. □

Corollary 14.5.6. *If I is a semi-order, then $\mathcal{O}_{\mathrm{L}}(I) = \mathcal{O}_{\mathrm{R}}(I)$.*

By Lemma 14.5.5 the standard involution gives a bijection between the set of lattices I with $\mathcal{O}_{\mathrm{L}}(I) = \mathcal{O}$ and those with $\mathcal{O}_{\mathrm{R}}(I)$.

14.5.7. Suppose that R is a DVR (e.g., a localization of R at a prime ideal $\mathfrak{p}$). We will show how to reduce the proof of Theorem 14.5.1 to that of a semi-order.

When R is a DVR, the fractional R-ideal $\operatorname{nrd}(I) \subseteq R$ is principal, generated by an element with minimal valuation, so let $\alpha \in I$ achieves this minimum. Then the R-lattice $J = \alpha^{-1}I$ now satisfies $1 \in J$ and $\operatorname{nrd}(J) = R$. Thus J is a semi-order, and J is (locally) principal if and only if I is (locally) principal.

Proof of Theorem 14.5.1. Suppose I is invertible; we wish to show that I is locally principal. The conclusion is local, so localizing we may assume R is a DVR.

We may reduce to the case where I is a semi-order by Paragraph 14.5.7. Then $1 \in I$. Let $\alpha_1, \dots, \alpha_n$ be an R-basis for I.

We claim that $I^{n+1} = I^n$. Since $1 \in I$, we have $I^n \subseteq I^{n+1}$. It suffices then to prove that a product of $n + 1$ basis elements of I lies in I^n. By the pigeonhole principle, there must be a repeated term α_i among them. But we have the formula (4.2.7)

$$\alpha\beta + \beta\alpha = \operatorname{trd}(\beta)\alpha + \operatorname{trd}(\alpha)\beta - \operatorname{trd}(\alpha\overline{\beta}) \tag{14.5.8}$$

for all $\alpha, \beta \in B$. We can use this relation to "push" the second instance of the repeated element until it meets with its mate, at the expense of terms lying in I^n. Specifically, in the R-module I^2/I, (14.5.8) implies

$$\alpha_i\alpha_j \equiv -\alpha_j\alpha_i \pmod{I}$$

for all i, j; it follows that in I^{n+1}/I^n we have

$$\mu(\alpha_i\alpha_j)\nu \equiv -\mu(\alpha_j\alpha_i)\nu \pmod{I^n}$$

for any μ, ν appropriate products of basis elements. Therefore we may assume that the repetition α_i^2 is adjacent; but then α_i satisfies a quadratic equation, so $\alpha_i^2 = \operatorname{trd}(\alpha_i)\alpha_i - \operatorname{nrd}(\alpha_i) \in I$, so in fact the product belongs to I^n, and the claim follows.

To conclude, from the equality $I^{n+1} = I^n$, we multiply both sides of this equation by $(I^n)^{-1}$ and obtain $I = \mathcal{O} = \mathcal{O}_{\mathrm{L}}(I) = \mathcal{O}_{\mathrm{R}}(I)$. In particular, I is principal, generated by 1. □

We have the following immediate corollary of the above proof.

Corollary 14.5.9. *An R-lattice I is an R-order if and only if* $1 \in I$*, every element of I is integral, and I is invertible.*

14.5.10. We conclude with two important consequences.

First, let I, J be invertible R-lattices such that I is compatible with J. Then we have $\operatorname{nrd}(IJ) = \operatorname{nrd}(I)\operatorname{nrd}(J)$, since it is enough to check this locally, and locally both I and J are principal and we have proved the statement in this case (Lemma 14.3.4).

Dressed up a little bit, this implies that the reduced norm is a homomorphism from the groupoid of invertible lattices to the group(oid) of fractional R-ideals in F.

Second, in the presence of a standard involution, we can write the inverse in another way: we have

$$\overline{I}I = \operatorname{nrd}(I)\mathcal{O}_{\mathrm{R}}(I) \text{ and } I\overline{I} = \operatorname{nrd}(I)\mathcal{O}_{\mathrm{L}}(I)$$

by checking these statements locally (where they follow immediately by computing the norm on a local generator). Since $\operatorname{nrd}(I)$ is a fractional R-ideal and thus invertible, since R is a Dedekind domain, it follows that

$$I^{-1} = \overline{I}\operatorname{nrd}(I)^{-1}.$$

14.6 Projective and proper modules

In this section, we move from lattices to modules, relating invertibility to projectivity. This section is a bit technical and can be skipped for the reader who does not need this reinterpretation. We restore the generality that R is a noetherian domain.

Definition 14.6.1. Let $\mathcal{O}$ be an order and let P be a left $\mathcal{O}$-module. Then P is *projective* (as a left $\mathcal{O}$-module) if it is a direct summand of a free left $\mathcal{O}$-module.

We see that a left $\mathcal{O}$-module P is projective if and only if there exists a left $\mathcal{O}$-module Q such that $P \oplus Q$ is free. This definition may seem opaque on first reading; it turns out that projective modules are fundamental in many areas of algebra, as the following proposition indicates.

Proposition 14.6.2. *Let P be a left $\mathcal{O}$-module. Then the following are equivalent:*

(i) *P is projective;*

(ii) *Every surjective homomorphism $f : M \to P$ (of left $\mathcal{O}$-modules) has a splitting $g : P \to M$ (i.e., $f \circ g = \mathrm{id}_P$);*

(iii) *Every diagram*

$$\begin{tikzcd}
 & P \arrow[dl, dashed, "q"'] \arrow[d, "p"] & \\
N \arrow[r, "f"] & M \arrow[r] & 0
\end{tikzcd}$$

with exact bottom row can be extended (i.e., $p = f \circ q$); and

(iv) $\operatorname{Hom}_{\mathcal{O}}(P, -)$ *is an exact functor.*

Let $I \subseteq B$ be an R-lattice. Then I has the structure of a finitely generated left $\mathcal{O}_{\mathrm{L}}(I)$-module and a finitely generated right $\mathcal{O}_{\mathrm{R}}(I)$-module.

[[Relationship between compatibility, multiplication, and tensor product]]

In the commutative case, an R-lattice $\mathfrak{a} \subseteq F$ is invertible if and only if $\mathfrak{a}$ is projective as a (left and right) R-module. The same is true in this context, as follows.

Theorem 14.6.3. *An R-lattice I is invertible if and only if I is projective as a left $\mathcal{O}_{\mathrm{L}}(I)$-module and as a right $\mathcal{O}_{\mathrm{R}}(I)$-module.*

Proof. To prove the implication ($\Rightarrow$), suppose that I is invertible. Then $I^{-1}I = \mathcal{O}_{\mathrm{R}}(I)$, so there exist $\alpha_i \in I$ and $\alpha_i^* \in I^{-1}$ such that $\sum_i \alpha_i^* \alpha_i = 1$. We may extend the set α_i to generate I as a left $\mathcal{O}_{\mathrm{L}}(I)$-module by taking $\alpha_i^* = 0$ if necessary. We define the surjective map

$$\begin{aligned} f : M = \bigoplus_i \mathcal{O}_{\mathrm{L}}(I) e_i &\to I \\ e_i &\mapsto \alpha_i. \end{aligned}$$

Consider the map

$$\begin{aligned} g : I &\to M \\ \beta &\mapsto (\beta \alpha_i^*)_i; \end{aligned}$$

the map g is defined because we have $\beta\alpha_i^* \in II^{-1} \subseteq \mathcal{O}_{\mathrm{L}}(I)$ for all $\beta \in I$. The map g is a splitting of f since

$$(f \circ g)(\beta) = \sum_i \beta \alpha_i^* \alpha_i = \beta \sum_i \alpha_i^* \alpha_i = \beta.$$

Therefore I is a direct summand of M, so I is projective as a left $\mathcal{O}_{\mathrm{L}}(I)$-module. A similar argument works on the right.

Next we prove ($\Leftarrow$). There exists a nonzero $r \in I \cap R$ (Exercise 8.2), so to show that I is invertible, we may replace I with $r^{-1}I$ and therefore assume that $1 \in I$.

Following in similar lines as in the previous paragraph, let α_i generate I as a left $\mathcal{O}_{\mathrm{L}}(I)$-module, and consider the surjective map $f : M = \bigoplus_i \mathcal{O}_{\mathrm{L}}(I)e_i \to I$ by $e_i \mapsto \alpha_i$. Then since I is projective as a left $\mathcal{O}_{\mathrm{L}}(I)$-module, this map splits by a map $g : I \to M$; suppose that $g(1) = (\alpha_i^*)_i$ with $\alpha_i^* \in \mathcal{O}_{\mathrm{L}}(I)$; then

$$(f \circ g)(1) = 1 = \sum_i \alpha_i^* \alpha_i \tag{14.6.4}$$

For any $\beta \in I$, we have $g(\beta) = (\beta\alpha_i^*)_i \in M$, so $\beta\alpha_i^* \in \mathcal{O}_{\mathrm{L}}(I)$ for all i; therefore for all $\alpha, \beta \in I$ we have $\beta\alpha_i^*\alpha \in \mathcal{O}_{\mathrm{L}}(I)I \subseteq I$, whence $\alpha_i^* \in I^{-1}$ by definition.

We always have $II^{-1}I \subseteq I$; to show the reverse inclusion, note that if $\beta \in I$ then $\beta = (f \circ g)(\beta) = \sum_i \beta\alpha_i^*\alpha_i \in II^{-1}I$. Therefore we have equality $II^{-1}I = I$ and $1 \in I^{-1}I$. Repeating the argument on the right, we have also $1 \in II^{-1}$.

By Proposition 14.4.6, it remains only to show that the product $II^{-1}I$ is compatible; we will show $\mathcal{O}_{\mathrm{R}}(I) = \mathcal{O}_{\mathrm{L}}(I^{-1})$, the other following similarly. By definition, we have $\mathcal{O}_{\mathrm{R}}(I) \subseteq \mathcal{O}_{\mathrm{L}}(I^{-1})$, so we prove the other implication. Let $\mu \in \mathcal{O}_{\mathrm{L}}(I^{-1})$, so $\mu I^{-1} \subseteq I^{-1}$, i.e., the implication

$$(I\alpha^* I \subseteq I) \Rightarrow (I\mu\alpha^* I \subseteq I) \tag{14.6.5}$$

holds for all $\alpha^* \in I^{-1}$. We need to show $I\mu \subseteq I$. We know that each α_i^* satisfies $I\alpha_i^*I \subseteq II^{-1}I = I$, so by the implication (14.6.5) we have that $I\mu\alpha_i^*I \subseteq I$ for all i. Thus for all $\beta \in I$, we have $\beta\mu\alpha_i^*\alpha_i \in I$ hence by (14.6.4) we have $\beta\mu = \sum_i \beta\mu\alpha_i^*\alpha_i \in I$, so $\mu \in \mathcal{O}_{\mathrm{R}}(I)$ as desired. □

14.6.6. In Theorem 14.6.3, we only considered an R-lattice I as a module over its left and right orders. Of course, I has the structure of a left $\mathcal{O}$-module for any $\mathcal{O} \subseteq \mathcal{O}_{\mathrm{L}}(I)$, and similarly on the right.

We defined invertibility for the lattice I in terms of its left and right orders. But only $\mathcal{O}_{\mathrm{L}}(I)$ works in the definition: if I' is an R-lattice and $II' = \mathcal{O}$ for some $\mathcal{O} \subseteq \mathcal{O}_{\mathrm{L}}(I)$, then multiplying on both sides on the left by $\mathcal{O}_{\mathrm{L}}(I)$ gives

$$\mathcal{O}_{\mathrm{L}}(I)II' = II' = \mathcal{O}_{\mathrm{L}}(I)\mathcal{O} = \mathcal{O}_{\mathrm{L}}(I),$$

with a similar statement on the right.

Lemma 14.6.7. . *If $I \subset B$ is projective as an $\mathcal{O}, \mathcal{O}'$-bimodule, then $\mathcal{O} = \mathcal{O}_{\mathrm{L}}(I)$ and $\mathcal{O}' = \mathcal{O}_{\mathrm{R}}(I)$.*

In other words, even when interpreting an R-lattice I as a left or right module, to get good behavior (like invertibility) we will want to take this structure over the full left or right endomorphism ring.

Proof. Combine Theorem 14.6.3 and Paragraph 14.6.6. □

Definition 14.6.8. We say $I \subset B$ is *proper* as a $\mathcal{O}, \mathcal{O}'$-bimodule if $\mathcal{O} = \mathcal{O}_{\mathrm{L}}(I)$ and $\mathcal{O}' = \mathcal{O}_{\mathrm{R}}(I)$.

The term *proper* is badly overloaded in mathematics, so we will not make extensive use of this term.

Example 14.6.9. The converse of Lemma 14.6.7. Consider again Example 14.4.7. The lattice I has the structure of a proper $\mathcal{O}', \mathcal{O}'$-module, where $\mathcal{O}' = R + \mathfrak{a}\mathcal{O}$ module. However, since I is not invertible, we conclude from Theorem 14.6.3 that I is *not* projective as a left or right $\mathcal{O}'$-module. (In [[???]], we characterize orders $\mathcal{O}$ for which every lattice which is proper as a left $\mathcal{O}$-module is projective: they are the Gorenstein orders $\mathcal{O}$.

14.7 One-sided invertibility

We conclude this chapter with some comments on one-sided notions of invertibility.

Definition 14.7.1. An R-lattice I is *right invertible* if there exists an R-lattice $I' \subseteq B$, a *right inverse*, such that the product II' is compatible and $II' = \mathcal{O}_{\mathrm{L}}(I)$ In a similar way, we define *left invertible* and *left inverse*.

Applying the same reasoning as in Lemma 14.4.10, we see that left (or right) invertibility is a local property.

Remark 14.7.2. The compatibility condition in invertibility is important. Consider Example 14.3.1: we have $IJ = \mathrm{M}_2(R) = \mathcal{O}_{\mathrm{L}}(I)$, and indeed if we let $J = \begin{pmatrix} \mathfrak{b} & \mathfrak{b} \\ R & R \end{pmatrix}$ for *any* nonzero ideal $\mathfrak{b} \subseteq R$, the equality $IJ = \mathrm{M}_2(R)$ remains true.

Again in this context a natural potential left (and right) inverse presents itself: if $II' = \mathcal{O}_{\mathrm{L}}(I)$, then I' maps I into $\mathcal{O}_{\mathrm{L}}(I)$ on the right. Accordingly, we make the following definition.

Definition 14.7.3. The *left colon lattice* of I with respect to J is the set

$$(I : J)_{\mathrm{L}} = \{\alpha \in B : \alpha J \subseteq I\}$$

and similarly the *right colon lattice* is

$$(I : J)_{\mathrm{R}} = \{\alpha \in B : J\alpha \subseteq I\}.$$

14.7.4. Note that $(I : I)_{\mathrm{L}} = \mathcal{O}_{\mathrm{L}}(I)$ is the left order of I (and similarly on the right). The same proof as in Paragraph 8.3.5 shows that $(I : J)_{\mathrm{L}}$ and $(I : J)_{\mathrm{R}}$ are indeed R-lattices.

Left invertibility does not imply right invertibility, and so the sided notions can be a bit slippery: see Exercise 14.12.

Remark 14.7.5. For rings, the (left or) right inverse of an element need not be unique even though a two-sided inverse is necessarily unique. Once can say at least there if I has a right inverse I' then I has a unique *maximal* right inverse (under inclusion); one may hope that this maximal right inverse is$I' = (\mathcal{O}_{\mathrm{L}}(I) : I)_{\mathrm{R}}$, but the compatibility is not clear.

In the presence of a standard involution over a Dedekind domain R—the case of interest in this book—the sided notions of invertibility are equivalent to the two-sided notion.

Lemma 14.7.6. *Suppose R is a Dedekind domain and that B has a standard involution. Then an R-lattice I is left invertible if and only if I is right invertible if and only if I is invertible.*

Proof. We will show that if I is right invertible then I is left invertible. By localizing, we reduce to the case where R is a DVR. By the results of Paragraph 14.5.7, we may assume that I is a semi-order, so that $\mathcal{O}_{\mathrm{L}}(I) = \mathcal{O}_{\mathrm{R}}(I) = \mathcal{O}$ and $I = \overline{I}$. Suppose $II' = \mathcal{O}$. $\overline{I'} = \overline{I'}I = \overline{\mathcal{O}} = \mathcal{O}$, and $\overline{I'}$ is compatible with I since

$$\mathcal{O} = \mathcal{O}_{\mathrm{R}}(I) = \mathcal{O}_{\mathrm{L}}(I') = \mathcal{O}_{\mathrm{R}}(\overline{I'})$$

as desired. □

Corollary 14.7.7. *An R-lattice I is left invertible with $I'I = \mathcal{O}_{\mathrm{R}}(I)$ if and only if $I' = (\mathcal{O}_{\mathrm{R}}(I) : I)_{\mathrm{L}} = I^{-1}$.*

A similar statement holds for the right inverse.

Proof. Let $\mathcal{O} = \mathcal{O}_{\mathrm{R}}(I)$. Then we have

$$\mathcal{O} = I'I \subseteq (\mathcal{O} : I)_{\mathrm{L}} I \subseteq \mathcal{O}$$

so equality must hold, and $I'I = (\mathcal{O} : I)_{\mathrm{L}} I$. By 14.7.6, I is invertible, and multiplying both sides by I^{-1} gives $I' = (\mathcal{O} : I)_{\mathrm{L}}$ as desired. □

14.8 Extensions and further reading

14.8.1. Cox [Cox89, §7] discusses orders in quadratic fields and the connections to quadratic forms and class numbers.

14.8.2. Due to Kaplansky [Kap69]; he calls our compatible product instead *concordant*.

14.8.3. To relax the condition that R is noetherian, one could work instead with *Prüfer domains* (generalizing Dedekind domains to the non-noetherian context).

14.8.4. This is really all about Morita equivalence.

Proposition 14.6.3 also goes by the name "dual basis" lemma in other places.

14.8.5. Groupoids are fun, and show up in stacks. There is a slight difference with the notion of a Brandt groupoid and a groupoid, but we are careful about the distinction.

14.8.6. [[Link the inverse to the dual of a lattice.]]

Exercises

Unless otherwise specified, throughout these exercises let R be a domain with field of fractions F, let V be an F-vector space, and let B be an F-algebra.

14.1. Let $D \in \mathbb{Z} \setminus \{0,1\}$ be a discriminant, and let $S(D) = \mathbb{Z}[(D+\sqrt{D})/2]$ be the quadratic ring of discriminant D. Suppose that $d = df^2$ with $f > 1$. Show that the ideal $(f, \sqrt{D})$ of $S(D)$ where is not invertible. (In particular, there are number rings with class number 1 that are not PIDs!)

14.2. Let $I \subseteq B$ be a principal R-lattice. Show that if α generates I then $\alpha \in B^\times$.

14.3. Show that if $I = \mathcal{O}_{\mathrm{L}}(I)\alpha \subseteq V$ then $\mathcal{O}_{\mathrm{R}}(I) = \alpha^{-1}\mathcal{O}_{\mathrm{L}}(I)\alpha$.

14.4. Let $I \subseteq B$ be an R-lattice. Show that if $\mathcal{O}_{\mathrm{L}}(I)$ is maximal, then $\mathcal{O}_{\mathrm{R}}(I)$ is maximal. Show that all maximal orders are connected.

14.5. Let $I \subset \mathrm{M}_2(F)$ be a lattice with $\mathcal{O}_{\mathrm{R}}(I) = \mathrm{M}_2(R)$. By considering $I \otimes_R F$ show that

$$I \subseteq \begin{pmatrix} F & F \\ 0 & 0 \end{pmatrix} \mathrm{M}_2(R) \oplus \begin{pmatrix} 0 & 0 \\ F & F \end{pmatrix} \mathrm{M}_2(R).$$

Now suppose that R is a PID. Conclude that I is principal. Conclude (again) that all maximal orders in $\mathrm{M}_2(F)$ are conjugate.

14.6. Show that if I, J are invertible and I is compatible with J, then IJ is invertible and $(IJ)^{-1} = J^{-1}I^{-1}$.

14.7. Let B be an F-algebra with a standard involution $\overline{}$. Show that if I is a semi-order then $\overline{I} = I$.

14.8. Let I be an R-lattice, and let $\alpha_1, \dots, \alpha_n$ generate I as an R-module. Give an explicit example where $\operatorname{nrd}(I)$ is not generated by $\operatorname{nrd}(\alpha_i)$ (cf. Lemma 14.2.8). Moreover, show that for any R-lattice I, there exists a set of R-module generators α_i such that $\operatorname{nrd}(I)$ is in fact generated by $\operatorname{nrd}(\alpha_i)$.

14.9. Show that if I is an R-lattice and $\alpha \in B$ then $\operatorname{nrd}(\alpha I) = \operatorname{nrd}(\alpha)\operatorname{nrd}(I)$. Conclude that if I is a principal R-lattice, generated by $\alpha \in I$, then $\operatorname{nrd}(I) = \operatorname{nrd}(\alpha)R$.

14.10. Let R be a Dedekind domain with field of fractions F, let $K \supset F$ be a quadratic field extension and let S be an R-order in K. Let S_K be the integral closure of R in K.

a) Show that there exists a (unique) ideal $\mathfrak{f} = \mathfrak{f}(S) \subset S_K$ (called the *conductor*) such that $S = R + \mathfrak{f}S_K$.

b) Now let $\mathfrak{b} \subset K$ be a fractional S-ideal. Show that the following are equivalent:

(i) $\mathfrak{b}$ is a locally principal S-ideal;

(ii) $\mathfrak{b}$ is invertible as a fractional S-ideal, i.e., there exists a fractional ideal $\mathfrak{b}^{-1}$ such that $\mathfrak{b}\mathfrak{b}^{-1} = S$ (necessarily $\mathfrak{b}^{-1} = (S : \mathfrak{b})$);

(iii) There exists $d \in K^\times$ such that $d\mathfrak{b} + \mathfrak{f} \cap S = S$; and

(iv) $\mathfrak{b}$ is proper, i.e., $S = \mathcal{O}(\mathfrak{b}) = \{x \in K : x\mathfrak{b} \subseteq \mathfrak{b}\}$.

14.11. Let $\mathcal{O} \subseteq B$ be an R-order.

a) Let $\alpha \in B^\times$. Show that $I = \mathcal{O}\alpha$ is a lattice with $\mathcal{O}_{\mathrm{L}}(I) = \mathcal{O}_{\mathrm{R}}(I) = \mathcal{O}$ if and only if $\alpha \in B^\times$ and $\mathcal{O}\alpha = \alpha\mathcal{O}$. Conclude that the set of invertible two-sided principal lattices I with $\mathcal{O}_{\mathrm{L}}(I) = \mathcal{O}_{\mathrm{R}}(I) = \mathcal{O}$ forms a group.

b) Show that the normalizer of $\mathcal{O}$,

$$N(\mathcal{O}) = \{\alpha \in B^\times : \alpha\mathcal{O}\alpha^{-1} = \mathcal{O}\}$$

is the group generated by $\alpha \in B^\times$ such that $\mathcal{O}\alpha$ is a two-sided $\mathcal{O}$-ideal.

14.12. Let R be a DVR with field of fractions F, and let $a \in R$ be neither zero nor a unit. Consider the R-lattice

$$I = \begin{pmatrix} (a) & (a) & R \\ (a^2) & (a^2) & R \\ R & R & R \end{pmatrix} \subset B = \mathrm{M}_3(F)$$

Show that I is left invertible but is not right invertible.

14.13. Let I be an R-lattice in B over F, let K be a finite extension field of F, and let S be a domain containing R with field of fractions K. Show that

$$\operatorname{disc}(I \otimes_R S) = \operatorname{disc}(I) \otimes_R S = \operatorname{disc}(I)S.$$

Chapter 15

Classes of quaternion ideals

15.1 Composition laws and ideal multiplication

Following the previous chapter, we now study *classes* of quaternion ideals. To guide these investigations, we first appeal to the quadratic case: it is quite instructive to see how the theory is built in the simpler but still incredibly rich commutative case.

Let $D \in \mathbb{Z}$ be a discriminant. A subject of classical interest was the set of *integral primitive binary quadratic forms* of discriminant D, namely

$$\mathcal{Q}(D) = \{ax^2+bxy+cy^2 : a, b, c \in \mathbb{Z},\, b^2 - 4ac = D, \text{ and } \gcd(a, b, c) = 1\} \subset \mathbb{Z}[x, y].$$

Of particular interest to early number theorists (Fermat, Legendre, Lagrange, and Gauss) was the set of primes represented by a quadratic form $Q \in \mathcal{Q}(D)$—inquiries of this nature proved to be foundational, giving rise to the law of quadratic reciprocity and the beginnings of the theory of complex multiplication and class field theory.

An invertible (oriented) change of variables on a quadratic form $Q \in \mathcal{Q}(D)$ does not alter the set of primes represented, so one is naturally led to study the *classes* of quadratic forms under the action of the group $\mathrm{SL}_2(\mathbb{Z})$ given by

$$(g^{-1} \cdot Q)(x, y) = Q(px + qy, rx + sy) \quad \text{for} \quad g = \begin{pmatrix} p & q \\ r & s \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}).$$

The set $\mathrm{Cl}(D)$ of $\mathrm{SL}_2(\mathbb{Z})$-classes of forms in $\mathcal{Q}(D)$ is finite, by reduction theory: every form in $\mathcal{Q}(D)$ is equivalent under the action of $\mathrm{SL}_2(\mathbb{Z})$ to a *reduced* form with $|b| \leq a \leq c$ (and further $b \geq 0$ in the boundary cases $a = |b|$ or $a = c$). To study this finite set, Gauss defined a *composition law* on $\mathrm{Cl}(D)$, giving $\mathrm{Cl}(D)$ the structure of an abelian group by an explicit formula.

Today, we see this composition law as a consequence of a natural identification of $\mathrm{Cl}(D)$ with a set with an obvious group structure. There is a bijection between $\mathrm{Cl}(D)$ and the *narrow class group* $\mathrm{Cl}^+(S) = \mathrm{Cl}^+(S(D))$, the group of invertible fractional ideals of the quadratic order S under multiplication

modulo the subgroup of

nonzero principal fractional ideals generated by a totally positive element

(i.e., one that is positive in every embedding into $\mathbb{R}$—if $D < 0$ then this is no condition). Specifically, to the class of the quadratic form $Q = ax^2 + bxy + cy^2 \in \mathcal{Q}(D)$, we associate the class of the ideal

$$\mathfrak{a} = a\mathbb{Z} + \left(\frac{-b + \sqrt{D}}{2}\right)\mathbb{Z} \subset S(D).$$

Conversely, the quadratic form is recovered as the norm form on $K = \mathbb{Q}(\sqrt{D})$ restricted to $\mathfrak{a}$:

$$\mathrm{N}\left(ax + \frac{-b + \sqrt{D}}{2}y\right) = ax^2 + bxy + cy^2, \quad \text{where } c = \frac{b^2 - D}{4a} \in \mathbb{Z}.$$

Much of the same structure can be found in the quaternionic case, with several interesting twists. Historically, it was Brandt who first asked if there was a composition law for (integral, primitive) quaternary quadratic forms: it would arise naturally from some kind of multiplication of ideals in a quaternion order, with the analogous bijection furnished by the reduced norm form, and we group together lattices based on orders that are connected to one another, as follows.

Let B be a quaternion algebra over $\mathbb{Q}$. Recall we have defined the notion of a compatible product on the set of lattices $I \subset B$. In the consideration of classes of such lattices, we make a choice and consider lattices as right modules—considerations on the left are analogous. We say that lattices I and J are *in the same right class*, and write $I \sim J$, if there exists $\alpha \in B^\times$ such that $\alpha I = J$. The relation $\sim$ is an equivalence relation, and the class of a lattice I is denoted $[I]$. If I is invertible, then every lattice in the class $[I]$ is invertible and we simply call the class invertible. We say that $[I]$ is *compatible* with $[J]$ if there exists $I' \in [I]$ and $J' \in [J]$ such that $\mathcal{O}_\mathrm{R}(I') = \mathcal{O}_\mathrm{L}(J)$.

Now let $\mathcal{O} \subset B$ be an order. We say that $\mathcal{O}'$ is *connected* to $\mathcal{O}$ if there exists an invertible lattice I with $\mathcal{O}_\mathrm{L}(I) = \mathcal{O}$ and $\mathcal{O}_\mathrm{R}(I) = \mathcal{O}'$. If $I' \in [I]$, then $\mathcal{O}_\mathrm{R}(I') = \mathcal{O}_\mathrm{R}(I)$ and we write simply $\mathcal{O}_\mathrm{R}([I])$.

Theorem 15.1.1. *The set*

$$\mathcal{B}(\mathcal{O}) = \{[I] : \mathcal{O}_\mathrm{R}([I]) \text{ connected to } \mathcal{O}\} \tag{15.1.2}$$

has the structure of a finite groupoid under compatible product.

We call the set $\mathcal{B}(\mathcal{O})$ in (15.1.2) the *Brandt groupoid* of $\mathcal{O}$. The Brandt groupoid is *strongly connected*: for all $[I],[K] \in \mathcal{B}(\mathcal{O})$, there exists $[J] \in \mathcal{O}$ such that $[I]$ is compatible with $[J]$ and $[I][J] = [K]$.

Theorem 15.1.1 follows from general principles in the geometry of numbers. The methods of the geometry of numbers do not provide a sharp bound on $\#\mathcal{B}(\mathcal{O})$; in Chapter 18, we consider a more refined approach that gives a weighted formula for the number of classes.

In this way, one recovers a partial composition law on certain classes of quaternary quadratic forms by restricting the reduced norm nrd to a representative ideal in each class.

Example 15.1.3. Consider the quaternion algebra $B = \left(\dfrac{-1,-11}{\mathbb{Q}}\right)$ and the maximal order $\mathcal{O}$ generated by i and $(j+1)/2$,

$$\mathcal{O} = \mathbb{Z} + i\mathbb{Z} + \mathbb{Z}\frac{j+1}{2}\mathbb{Z} + i\frac{j+1}{2}\mathbb{Z}.$$

The set of orders connected to $\mathcal{O}$ has exactly two isomorphism classes, represented by $\mathcal{O}_1 = \mathcal{O}$ and

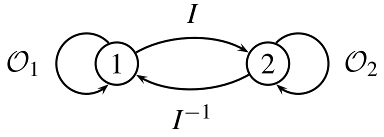
$$\mathcal{O}_2 = \mathbb{Z} + 2i\mathbb{Z} + \frac{1-2i+j}{2}\mathbb{Z} + \frac{2-i-ij}{2}\mathbb{Z}.$$

The Brandt groupoid associated to (the class of orders containing) $\mathcal{O}$ has four elements, represented by $\mathcal{O}_1$, $\mathcal{O}_2$, the right ideal

$$I = 2\mathcal{O} + \frac{1-2i-j}{2}\mathcal{O}$$

with right order $\mathcal{O}_1$ and left order $\mathcal{O}_2$, and its inverse I^{-1} with these reversed, so the products $I^{-1}I = \mathcal{O}_1$ and $II^{-1} = \mathcal{O}_2$ are compatible.

We can visualize this groupoid as a graph as follows, with directed edges for multiplication:



(One could consider the dual graph, interchanging edges for vertices, if preferred.) Restricting the reduced norm to these lattices, we then have a description of the partial composition law on classes of quaternary quadratic forms of discriminant 11^2:

$$\begin{aligned}
\operatorname{nrd}|_{\mathcal{O}_1} &= x^2 + xw + y^2 + yz + 3z^2 + 3w^2 \\
\operatorname{nrd}|_{\mathcal{O}_2} &= x^2 + xy + xz + y^2 + yz + yw + 4z^2 + 4zw + 4w^2 \\
\operatorname{nrd}|_I \cong \operatorname{nrd}|_{I^{-1}} &= 2x^2 + 2xy + xz + 2y^2 + yz + yw + 2z^2 + 2zw + 2w^2
\end{aligned}$$

In the final part of this book, we will see that by counting the representation of primes represented by quaternary quadratic forms we uncover deep arithmetic structure: the Brandt groupoid encodes a space of modular forms.

15.2 Two-sided ideals

To get warmed up for the one-sided notion of classes, we begin by considering first the two-sided notions, where we will be able to define a group structure.

Let R be a noetherian domain with field of fractions F, let B be an F-algebra, and let $\mathcal{O}$ be an R-order in B.

15.2.1. Let I, J be invertible two-sided fractional $\mathcal{O}$-ideals (cf. Definition 14.2.6). Then IJ is also an invertible two-sided fractional $\mathcal{O}$-ideal by Lemma 14.4.9, as $\mathcal{O}_{\mathrm{L}}(IJ) = \mathcal{O}_{\mathrm{L}}(I) = \mathcal{O}$ and $\mathcal{O}_{\mathrm{R}}(IJ) = \mathcal{O}_{\mathrm{R}}(J) = \mathcal{O}$.

Let $\mathcal{I}(\mathcal{O})$ be the set of invertible two-sided fractional $\mathcal{O}$-ideals. Then multiplication defines an associative bilinear operation on $\mathcal{I}(\mathcal{O})$ with identity element $\mathcal{O}$, so $\mathcal{I}(\mathcal{O})$ has the structure of a group.

Lemma 15.2.2. *The group $\mathcal{I}(\mathcal{O})$ is abelian.*

Proof. Let $I, J \in \mathcal{I}(\mathcal{O})$. Consider $IJI^{-1} \in \mathcal{I}(\mathcal{O})$. There exists a nonzero $r \in R \cap I$ (Exercise 8.2), so $J = rJr^{-1} \subseteq IJI^{-1}$. Similarly, we have $J \subseteq I^{-1}JI$ so $IJI^{-1} \subseteq J$, so equality holds and $\mathcal{I}(\mathcal{O})$ is abelian. □

Definition 15.2.3. The *Picard group* of the R-order $\mathcal{O}$, denoted $\operatorname{Pic}_R(\mathcal{O})$, is the quotient of $\mathcal{I}(\mathcal{O})$ by the subgroup of principal two-sided fractional $\mathcal{O}$-ideals.

A principal fractional $\mathcal{O}$-ideal is invertible by Paragraph 14.4.4. Note that if I, J are two-sided fractional $\mathcal{O}$-ideals then they are in the same two-sided ideal class in $\operatorname{Pic}_R(\mathcal{O})$ if and only if IJ^{-1} is a principal two-sided fractional $\mathcal{O}$-ideal.

At this point, we cannot say much more about $\operatorname{Pic}_R(\mathcal{O})$, and we introduce it first just for comfort; we will return later to its study.

15.3 One-sided ideals

We now study one-sided notions. Let $I, J \subset B$ be invertible right fractional $\mathcal{O}$-ideals.

Definition 15.3.1. We say I, J are *in the same (right) ideal class*, and we write $I \sim J$, if there exists $\alpha \in B^\times$ such that $\alpha I = J$.

Clearly $\sim$ defines an equivalence relation on the set of invertible right fractional $\mathcal{O}$-ideals, and we let $\operatorname{Cl}_{\mathrm{R}}\mathcal{O}$ denote the set of equivalence classes under this equivalence.

Lemma 15.3.2. *I is isomorphic to J as a right $\mathcal{O}$-module if and only if I and J are in the same ideal class if and only if $(J : I)_{\mathrm{L}} = JI^{-1}$ is principal.*

Proof. Suppose that $\phi : I \xrightarrow{\sim} J$ is an isomorphism of right $\mathcal{O}$-modules. Then $\phi_F : I \otimes_R F = B \xrightarrow{\sim} J \otimes_R F = B$ is an automorphism of B as a right B-module. Then as in Example 6.2.14, such an isomorphism is obtained by left multiplication by $\alpha \in B^\times$, so by restriction we have ϕ is given by this map as well. Conversely, if $\alpha I = J$ then left multiplication by α gives an isomorphism $I \xrightarrow{\sim} J$ by associativity in B

For the second equivalence, suppose $\alpha I = J$ with $\alpha \in B^\times$. Then $\alpha(II^{-1}) = \alpha\mathcal{O}_{\mathrm{L}}(I) = JI^{-1}$, and $\mathcal{O}_{\mathrm{R}}(JI^{-1}) = \mathcal{O}_{\mathrm{R}}(I^{-1}) = \mathcal{O}_{\mathrm{L}}(I)$, so JI^{-1} is indeed principal. The converse follows similarly. □

We will interchangeably use the language of ideal classes and isomorphisms.

It makes sense to identify isomorphic orders, as the isomorphism will identify these ideal classes.

Definition 15.3.3. Two orders $\mathcal{O}, \mathcal{O}'$ are *of the same type* if there exist $x \in B^\times$ such that $\mathcal{O}' = x^{-1}\mathcal{O}x$.

Lemma 15.3.4. *Let B be a central simple F-algebra. Then two orders $\mathcal{O}, \mathcal{O}' \subseteq B$ are isomorphic as R-algebras if and only if they are of the same type.*

Proof. If $\phi : \mathcal{O} \xrightarrow{\sim} \mathcal{O}'$ is an isomorphism of F-algebras, then extending scalars to F we obtain an F-algebra automorphism of B which is given by conjugation by the theorem of Skolem–Noether. □

Definition 15.3.5. Let $\mathcal{O}, \mathcal{O}' \subseteq B$ be R-orders. We say that $\mathcal{O}, \mathcal{O}'$ are *connected* if there exists an invertible fractional $\mathcal{O}, \mathcal{O}'$-ideal in B, called a *connecting ideal*.

Lemma 15.3.6. *Suppose that B has a standard involution. Then the orders $\mathcal{O}, \mathcal{O}'$ are connected if and only if $\mathcal{O}_\mathfrak{p}$ and $\mathcal{O}'_\mathfrak{p}$ are of the same type for all primes $\mathfrak{p}$.*

Proof. Let I be an invertible fractional $\mathcal{O}, \mathcal{O}'$-ideal I. Then I is locally principal, so $I_\mathfrak{p} = \mathcal{O}_\mathfrak{p}x_\mathfrak{p}$ and consequently $\mathcal{O}'_\mathfrak{p} = \mathcal{O}_{\mathrm{R}}(I_\mathfrak{p}) = x_\mathfrak{p}^{-1}\mathcal{O}_\mathfrak{p}x_\mathfrak{p}$. Conversely, since $\mathcal{O}_\mathfrak{p} = \mathcal{O}'_\mathfrak{p}$ for all but finitely many primes $\mathfrak{p}$, if we have $\mathcal{O}'_\mathfrak{p} = x_\mathfrak{p}^{-1}\mathcal{O}_\mathfrak{p}x_\mathfrak{p}$ then the R-lattice I with $I_\mathfrak{p} = \mathcal{O}_\mathfrak{p}x_\mathfrak{p}$ is an invertible fractional $\mathcal{O}, \mathcal{O}'$-ideal. □

Remark 15.3.7. Note that if two R-orders $\mathcal{O}, \mathcal{O}'$ are isomorphic then they are connected, but the converse is not in general true. We will return to this question in earnest in section [[??]].

Lemma 15.3.8. *Let $\mathcal{O}, \mathcal{O}'$ be connected R-orders. Then there exists a bijection $\operatorname{Cl}(\mathcal{O}) \xrightarrow{\sim} \operatorname{Cl}(\mathcal{O}')$ of sets and an isomorphism $\operatorname{Pic}_R(\mathcal{O}) \cong \operatorname{Pic}_R(\mathcal{O}')$ of groups.*

Proof. Let J be an invertible fractional $\mathcal{O}, \mathcal{O}'$-ideal, so that $\mathcal{O}_{\mathrm{L}}(I) = \mathcal{O}$ and $\mathcal{O}_{\mathrm{R}}(I) = \mathcal{O}'$. The map $I \mapsto IJ$ induces a bijection between the set of right $\mathcal{O}$-ideals and the set of right $\mathcal{O}'$-ideals, with inverse given by $I' \mapsto I'J^{-1}$, since each of the products are compatible. This induces a bijection $\operatorname{Cl}(\mathcal{O}) \xrightarrow{\sim} \operatorname{Cl}(\mathcal{O}')$, since is compatible with left multiplication in B, i.e., $(\alpha I)J = \alpha(IJ)$ for all $\alpha \in B^\times$.

In a similar way, the map $I \mapsto J^{-1}IJ$ yields an isomorphism from the group of invertible two-sided fractional ideals of $\mathcal{O}$ to those of $\mathcal{O}'$. [[This respects classes.]] □

Remark 15.3.9. There is no reason to expect a bijection of pointed sets. [[Draw a picture of a graph.]]

Proposition 15.3.10. *The map*

$$\begin{aligned} \operatorname{Cl}(\mathcal{O}) &\to T(\mathcal{O}) \\ [I] &\mapsto [\mathcal{O}_L(I)] \end{aligned}$$

has fiber canonically identified with $\operatorname{Pic}_R(\mathcal{O}')$ for $[\mathcal{O}'] \in T(\mathcal{O})$. In particular,

$$\# \operatorname{Cl}(\mathcal{O}) = \sum_{[\mathcal{O}'] \in T(\mathcal{O})} \# \operatorname{Pic}(\mathcal{O}').$$

Proof. Let I be an invertible right fractional $\mathcal{O}$-ideal. Then $\mathcal{O}_{\mathrm{L}}(I) \cong \mathcal{O}_i$ for a uniquely determined i, so $\mathcal{O}_{\mathrm{L}}(I) = x^{-1}\mathcal{O}_i x$ for some $x \in B^\times$. But then $I_i = xKI$ where $K = x^{-1}I_iI^{-1}$ is a two-sided invertible fractional $\mathcal{O}_{\mathrm{L}}(I)$-ideal, and so $I \sim KI_i \sim J_{i,j}I_i$ for some j, again uniquely determined. □

Corollary 15.3.11. *Let $\mathcal{O}$ be an R-order. Let $\mathcal{O}_i$ be representatives of the orders in B connected to $\mathcal{O}$, up to isomorphism. For each i, let I_i be a connecting ideal for $\mathcal{O}_i, \mathcal{O}$, and let $J_{i,j}$ be representatives of the two-sided invertible fractional $\mathcal{O}_i$-ideal classes.*

Then the set $\{J_{i,j}I_i\}_{i,j}$ is a complete set of representatives of $\operatorname{Cl}\mathcal{O}$.

15.4 Minkowski theory

[[Also known as Jordan-Zassenhaus]]

Now suppose that F is a number field and R is the ring of integers in F. Let B be a quaternion algebra over F and let $\mathcal{O}$ be an R-order in B. We will show that the set $\operatorname{Cl}\mathcal{O}$ of invertible right (fractional) $\mathcal{O}$-ideals is finite. This proof will be drastically improved upon in section [[??]] by considering the zeta function of $\mathcal{O}$; the proof in this section, using the geometry of numbers, is nevertheless worth giving.

We argue roughly as follows: if J is an invertible right $\mathcal{O}$-ideal, then there exists $x \in J^{-1}$ with the property that $xJ = I$ has $N_{F/\mathbb{Q}}(\mathrm{nrd}(I)) \leq C$ where $C \in \mathbb{R}_{>0}$ depends only on $\mathcal{O}$. The result will then follow from the fact that there are only finitely many right $\mathcal{O}$-ideals of bounded norm.

A subset $D \subseteq \mathbb{R}^n$ is *convex* if $tx + (1-t)y \in D$ whenever $x, y \in D$ and $t \in [0,1]$ and is *symmetric* if $-x \in D$ whenever $x \in D$. A *lattice* L in $\mathbb{R}^n$ is a discrete additive subgroup; the rank of L is the dimension of $L \otimes_{\mathbb{Z}} \mathbb{R}$ as an R-vector space. The *covolume* $\mathrm{covol}(L)$ of a lattice is equal to the volume of the quotient $\mathbb{R}^n/L$ or equivalently if $a_i = (a_{i1}, \dots, a_{in})$ is a basis for L then $\mathrm{covol}(L) = |\det(a_{ij})|$.

Theorem 15.4.1 (Minkowski)**.** *Let $D \subseteq \mathbb{R}^n$ be a closed, convex, symmetric subset of $\mathbb{R}^n$, and let L be a lattice in $\mathbb{R}^n$. If $\mu(D) \geq 2^n$ $\mathrm{covol}(L)$, then there exists $0 \neq x \in L \cap D$.*

Let $B = \left(\dfrac{a,b}{F}\right)$. For an infinite place v of F and $u = x + yi + zj + wij$, define

$$Q_v(u) = |v(x)|^2 + |v(a)||v(y)|^2 + |v(b)||v(z)|^2 + |v(ab)||v(w)|^2. \tag{15.4.2}$$

We then define the *absolute reduced norm* by

$$\begin{aligned} Q : B &\to \mathbb{R} \\ u &\mapsto \textstyle\sum_v Q_v(u); \end{aligned}$$

by construction, the form Q is positive definite and gives an R-lattice the structure of a lattice of full rank in $\mathbb{R}^{4n}$, where $n = [F : \mathbb{Q}]$. Note that $|v(\mathrm{nrd}(u))| \leq Q_v(u) \leq Q(u)$ for all infinite places v so $|\mathrm{N}_{F/\mathbb{Q}}(\mathrm{nrd}(u))| \leq Q(u)^n$ for all $u \in B$.

Lemma 15.4.3. *For all invertible fractional $\mathcal{O}$-ideals I in B we have* $\mathrm{covol}(I) = \mathrm{N}_{F/\mathbb{Q}}(\mathrm{nrd}(I))^2 \,\mathrm{covol}(\mathcal{O})$.

Proof. First, by definition we have

$$\mathrm{covol}(I) = [\mathcal{O} : I]_{\mathbb{Z}} \,\mathrm{covol}(\mathcal{O})$$

where $[\mathcal{O} : I]_{\mathbb{Z}}$ denotes the index as lattices. But we have $[\mathcal{O} : I]_{\mathbb{Z}} = \mathrm{N}_{F/\mathbb{Q}}([\mathcal{O} : I])$ where now the index is taken as R-modules, and $[\mathcal{O} : I] = \mathrm{nrd}(I)^2$ since this can be checked locally and then if $I_{\mathfrak{p}} = x_{\mathfrak{p}}\mathcal{O}_{\mathfrak{p}}$ then the F-endomorphism of B given by left multiplication by $x_{\mathfrak{p}}$ has determinant $\mathrm{nrd}(x_{\mathfrak{p}})^2$. Putting these together, we have $\mathrm{covol}(I) = \mathrm{N}_{F/Q}(\mathrm{nrd}(I))^2 \,\mathrm{covol}(\mathcal{O})$. □

Proposition 15.4.4. *There exists $C \in \mathbb{R}_{>0}$ such that every right ideal class of* $\mathrm{Cl}\,\mathcal{O}$ *is represented by an integral right $\mathcal{O}$-ideal with* $\mathrm{N}_{F/\mathbb{Q}}\,\mathrm{nrd}(I) \leq C$.

Proof. Let J be an invertible right fractional $\mathcal{O}$-ideal. Then

$$\operatorname{covol}(J^{-1}) = \mathrm{N}\operatorname{nrd}(J)^{-2}\operatorname{covol}(\mathcal{O}).$$

Let D be the convex body $\{x \in \mathbb{R}^{4n} : Q(x) \leq 1\}$, and let $a > 0$ be such that

$$a^{4n}\operatorname{vol}(D) = 2^{4n}\operatorname{covol}(J^{-1}) = 2^{4n}\,\mathrm{N}\operatorname{nrd}(J)^{-2}\operatorname{covol}(\mathcal{O}).$$

By Minkowski's theorem, there exists $0 \neq x \in J^{-1} \cap aD$. Therefore

$$|\mathrm{N}_{F/\mathbb{Q}}(\operatorname{nrd}(x))|^{1/n} \leq Q(x) \leq a^2.$$

Consequently

$$\begin{aligned}\mathrm{N}\operatorname{nrd}(xJ)^2 &= |\mathrm{N}\operatorname{nrd}(x)|^2\,\mathrm{N}\operatorname{nrd}(J)^2 \leq a^{4n}\,\mathrm{N}\operatorname{nrd}(J)^2\\ &= 2^{4n}\operatorname{covol}(\mathcal{O})\operatorname{vol}(D)^{-1} = C^2.\end{aligned}$$

Since $x \in J^{-1}$, the ideal $xJ = I$ is an $\mathcal{O}$-ideal in the same ideal class as J, which completes the proof. □

Lemma 15.4.5. *For any fractional ideal $\mathfrak{a}$ of R, there are only finitely many right $\mathcal{O}$-ideals with* $\operatorname{nrd}(I) = \mathfrak{a}$.

Proof. We may assume $\mathfrak{a} \subseteq R$. Since $\operatorname{nrd}(I) = \mathfrak{a}$ we see that $\mathfrak{a}\mathcal{O} \subseteq I \subseteq \mathcal{O}$. But $\mathcal{O}/\mathfrak{a}\mathcal{O}$ is a finite set, so there only finitely many possibilities for I. □

[[Cite Markus]]

[[This is like an "almost Euclidean algorithm", analogous to the class group of a number field. The pigeonhole principle should work, too.]]

15.5 Extensions and further reading

15.5.1. Cox [Cox89, §7] discusses orders in quadratic fields and the connections to quadratic forms and class numbers.

15.5.2. The composition law on binary quadratic forms can be understood quite concretely using 2 × 2 × 2 Rubik's cubes, by a beautiful result of Bhargava [[cite]].

[[Manjul and Melanie take a different approach, and consider *oriented* quadratic rings and ideal classes. Is this really any different?]]

Remark 15.5.3. The notation $\operatorname{Picent}(\mathcal{O})$ is also used for $\operatorname{Pic}_{Z(\mathcal{O})}(\mathcal{O})$ where $Z(\mathcal{O})$ is the center of $\mathcal{O}$ and we consider $\mathcal{O}$ as a $Z(\mathcal{O})$-algebra [[reference a remark on orders over rings that are not domains]].

Exercises

Unless otherwise specified, throughout these exercises let R be a domain with field of fractions F, let V be an F-vector space, and let B be an F-algebra.

[[Some explicit computations.]]

15.1.

Chapter 16

Quaternion orders over Dedekind domains

In this section, we begin to classify orders over a Dedekind domain.

16.1 Classifying orders

Let B be a quaternion algebra over $\mathbb{Q}$. A maximal order in B is analogous to the ring of integers in a number field, but because of noncommutativity, maximal orders in B are not unique (and in general, not unique even up to conjugation in B). Restricting our investigations to maximal orders would come at a cost, as many natural orders are not maximal: even the Lipschitz order quaternions $\mathbb{Z}+\mathbb{Z}i+\mathbb{Z}j+\mathbb{Z}k$, which arises when considering if a positive integer is the sum of four squares, is not maximal, contained inside the Hurwitz order $\mathbb{Z}+\mathbb{Z}i+\mathbb{Z}j+\mathbb{Z}(1+i+j+k)/2$.

How do we classify quaternion orders? In Chapter 4 (see Theorem 4.4.5), we saw that quaternion algebras over a field F (with $\operatorname{char} F \neq 2$) are classified by ternary quadratic forms over F. It should come as no surprise, then, that quaternion orders over a Dedekind domain R are classified by ternary quadratic modules over R (Theorem 16.3.1). The "module" part of this statement account for the possibility that R may not be a PID, and so some aspects of the class group of R creep in. Over $\mathbb{Z}$, we have an especially simple statement:

Theorem 16.1.1. *There is a discriminant-preserving bijection*

$$\left\{ \begin{array}{c} \textit{Quaternion orders over } \mathbb{Z} \\ \textit{up to isomorphism} \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} \textit{Nonsingular ternary quadratic forms over } \mathbb{Z} \\ \textit{up to sign under the action of } \mathrm{GL}_3(\mathbb{Z}) \end{array} \right\}$$

where $\mathrm{GL}_3(\mathbb{Z})$ *acts on quadratic forms by the natural change of variable.*

The bijection is defined by restricting the reduced norm, as follows. Let $\mathcal{O} \subset B$ be a quaternion order over $\mathbb{Z}$ with reduced discriminant $D = D(\mathcal{O}) \in \mathbb{Z}$, a quantity well defined up to sign. Let $1, i, j, k$ be a $\mathbb{Z}$-basis for $\mathcal{O}$: we do not assume that these are standard generators! Let $\frac{1}{2}, i^\sharp, j^\sharp, k^\sharp$ be the dual basis with respect to the standard bilinear form $\alpha, \beta \mapsto \operatorname{trd}(\alpha\overline{\beta})$, arising from the reduced norm. So,

$$\operatorname{trd}(i^\sharp) = 0, \quad \operatorname{trd}(ii^\sharp) = 1, \quad \operatorname{trd}(ij^\sharp) = \operatorname{trd}(ik^\sharp) = 0, \quad \text{etc.}$$

We then associate to $\mathcal{O}$ the quadratic form

$$Q(x, y, z) = D\operatorname{nrd}(xi^\sharp + yj^\sharp + zk^\sharp),$$

defined up to sign. This quadratic form has discriminant D.

Conversely, to the quadratic form Q

$$Q(x, y, z) = ax^2 + by^2 + cz^2 + uyz + vxz + wxy \in \mathbb{Z}[x, y, z]$$

up to sign, we associate the quaternion order $\mathcal{O}$ with basis $1, i, j, k$ where

$$\begin{aligned} i^2 &= ui - bc & jk &= a\overline{i} = a(u - i) \\ j^2 &= vj - ac & ki &= b\overline{j} = b(v - j) \\ k^2 &= wk - ab & ij &= c\overline{k} = c(w - k); \end{aligned} \tag{16.1.2}$$

this quaternion order has reduced discriminant

$$D = \operatorname{disc}(Q) = 4abc - uvw - au^2 - bv^2 - cw^2 \neq 0,$$

defined up to sign. (Note that if we scale the quadratic form by -1, the multiplication laws (16.1.2) do not change!) Completing the square, since Q is nonsingular, we see that at least one of the discriminants

$$u^2 - 4bc,\ v^2 - 4ac,\ w^2 - 4ab$$

is nonzero; assuming $w^2 - 4ab \neq 0$, we have

$$\mathcal{O} \subset B = \left(\frac{w^2 - 4ab, aD}{\mathbb{Q}}\right)$$

as well as the others by symmetry (when they apply).

The formulas may seem a little involved, so it is clarifying to work out the diagonal case. Let $B = \left(\frac{b, a}{\mathbb{Q}}\right)$ (the interchange is deliberate!), and let

$$\mathcal{O} = \mathbb{Z} + \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}k \subset B$$

be the tame order generated by the standard generators with

$$i^2 = a, \quad j^2 = b, \quad k = ij = -ji, \quad k^2 = -ab.$$

Then $\mathcal{O}$ has reduced discriminant $D = -4ab$. The dual basis is

$$i^\sharp = -\frac{i}{2b}, \quad j^\sharp = -\frac{j}{2a}, \quad k^\sharp = \frac{k}{2ab}$$

and the reduced norm on this basis is

$$D\,\mathrm{nrd}(xi^\sharp + yj^\sharp + zk^\sharp) = -4ab\left(\frac{-b}{4b^2}x^2 + \frac{-a}{4a^2}y^2 + \frac{ab}{4(ab)^2}z^2\right) = ax^2 + by^2 - z^2.$$

Conversely, to the quadratic form $Q(x, y, z) = ax^2 + by^2 - z^2$, we recover the order $\mathbb{Z} + \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}k$ with multiplication (16.1.2) by taking $u = v = w = 0$ and $c = -1$.

Why not just take the reduced norm restricted to the trace zero sublattice $\mathcal{O}^0$ of $\mathcal{O}$? In general, taking the trace zero subspace is not a good thing to do with respect to the prime 2. For example, we cannot recover $\mathcal{O}$ from its trace zero subspace: the Lipschitz and Hurwitz orders both have trace zero subspace spanned by i, j, k. The second more serious problem is that taking the trace zero quadratic form does not preserve discriminants: for the tame order, we have $Q(x, y, z) = \mathrm{nrd}(xi + yj + zk) = -bx^2 - ay^2 + abz^2$ of discriminant $4(ab)^2$, not $4ab$. We could recover the above form by taking instead

$$-(ab)^{-1}Q(ax, by, z) = \frac{a^2bx^2 + ab^2y^2 - abz^2}{ab} = ax^2 + by^2 - z^2,$$

but this change of variables is achieved by taking the dual basis above, in a more natural way!

Just as in the case of fields, the translation from quaternion orders to ternary quadratic forms makes the classification problem potentially easier: we replace the potentially complicated notion of finding a lattice closed under multiplication in a quaternion algebra with the simpler notion of choosing coefficients of a quadratic form. Using the local-global correspondence, in the next chapter, we tackle the classification problem for ternary quadratic forms over a local ring, where we can be quite explicit.

However, before we do so, we take a tour of the zoo of orders and identify those with good properties. To start, recall that a Dedekind domain R is a (commutative) domain that is *hereditary*: every submodule of a projective module is again projective. (Hence the name: projectivity is inherited by a submodule.) A domain is hereditary if and only if every ideal of R is projective, or equivalently, that any submodule of a free R-module is a direct sum of ideals of R. This property is used in the proof of unique factorization of ideals and makes the structure theory of modules over a Dedekind quite nice. (Note, however, that any order which is not maximal in a number ring is not hereditary.)

This definition carries over to the quaternionic context: an order $\mathcal{O}$ is *right hereditary* if every right ideal of $\mathcal{O}$ is projective as a right $\mathcal{O}$-module. The standard involution turns a right ideal into a left ideal, and so right hereditary is equivalent to the obvious notion of left hereditary, and so we simply refer to an order as *hereditary*.

There is a simple numerical criterion to test if an order is hereditary.

Theorem 16.1.3. *A $\mathbb{Z}$-order $\mathcal{O}$ is hereditary if and only if* $\operatorname{discrd}(\mathcal{O})$ *is squarefree; in particular, a maximal order $\mathcal{O}$ is hereditary.*

Orders $\mathcal{O}$ with squarefree discriminant are easily described, via their completions: they are orders of elements that are upper triangular modulo p for a finite set of primes $p \nmid \operatorname{disc}(B)$. More precisely, let B be a quaternion algebra of discriminant D and let $\mathcal{O}(1)$ be a maximal order in B, so $\operatorname{discrd}(\mathcal{O}(1)) = D$. Let N be a squarefree integer coprime to $\operatorname{disc}(B)$, and let $\mathcal{O} \subseteq \mathcal{O}(1)$ be an order with squarefree discriminant $\operatorname{discrd}(\mathcal{O}) = DN$. Then at all primes $p \nmid N$, we have $\mathcal{O}_p = \mathcal{O}(1)_p$ is maximal. For the primes $p \mid N$, there exists an isomorphism $\mathcal{O}(1)_p \cong \mathrm{M}_2(\mathbb{Z}_p)$ such that

$$\mathcal{O}_p = \left\{ \begin{pmatrix} a & b \\ pc & d \end{pmatrix} : a,b,c,d \in \mathbb{Z}_p \right\} \subset \mathrm{M}_2(\mathbb{Z}_p).$$

We can combine these into one, and write simply

$$\mathcal{O}_N \cong \left\{ \begin{pmatrix} a & b \\ Nc & d \end{pmatrix} : a,b,c,d \in \mathbb{Z}_N \right\} \subset \mathrm{M}_2(\mathbb{Z}_N)$$

where $\mathbb{Z}_N \cong \prod_{p|N} \mathbb{Z}_p$ is the completion at N.

More generally, Eichler considered those orders for which

$$\mathcal{O}_N \cong \left\{ \begin{pmatrix} a & b \\ Nc & d \end{pmatrix} : a,b,c,d \in \mathbb{Z}_N \right\}$$

for an integer N coprime to D but not necessarily squarefree: these orders are called *Eichler orders*. Equivalently, an Eichler order is the intersection of two maximal orders. From the perspective of matrices, these are those which are not endomorphisms of a full rank 2 module but preserve an incidence modulo N.

There is one final way of classifying rings that extends nicely to the noncommutative context. By way of analogy, note that orders in quadratic fields are characterized simply by their conductor. Let $K = \mathbb{Q}(\sqrt{D})$, where $D \in \mathbb{Z}$ is a fundamental discriminant. Then the maximal order in K is $\mathbb{Z} + \mathbb{Z}w$ where $w = (D + \sqrt{D})/2$. Any order S

in K is of the form $\mathbb{Z} + fS = \mathbb{Z} + \mathbb{Z}fw$, where $f \in \mathbb{Z}_{\geq 1}$ is the *conductor* of S, and the discriminant of S is $d = f^2 D$. Even in classical considerations, these orders arise naturally when considering binary quadratic forms of nonfundamental discriminant. The condition that $f = 1$, which ensures maximality, also ensures that the dual is projective, in other words, that S is *Gorenstein*. Gorenstein rings are natural from a linear algebra point of view.

An order $\mathcal{O}$ is *Gorenstein* if its dual $\mathcal{O}^\sharp$ is projective as a right $\mathcal{O}$-module. In a similar way, to a $\mathbb{Z}$-order $\mathcal{O} \subset B$, there is a unique maximal integer $f = f(\mathcal{O}) \in \mathbb{Z}_{\geq 1}$ such that $\mathcal{O} = \mathbb{Z} + f\mathcal{O}'$ where $\mathcal{O}' \subset B$ is a superorder of $\mathcal{O}$. We call f the *conductor* of the order; it is also sometimes called the *Brandt invariant* after Brandt. An order is Gorenstein if and only if $f = 1$. An Eichler order is Gorenstein, but there are Gorenstein orders that are not Eichler. If $\mathcal{O} = \mathbb{Z} + f\mathcal{O}'$ where $f = f(\mathcal{O})$ is the conductor of $\mathcal{O}$, then we call $\mathcal{O}' \supseteq \mathcal{O}$ the *Gorenstein closure* of $\mathcal{O}$. To understand orders, therefore, it is enough to understand Gorenstein orders. In the language of quadratic forms, an order is Gorenstein if and only if its associated quadratic form is *primitive*, meaning the greatest common divisor of its coefficients is 1. Because of the importance of the Gorenstein property, we say that an order $\mathcal{O}$ is *Bass* if every order that contains $\mathcal{O}$ is Gorenstein.

To summarize, we have

$$\text{maximal} \implies \text{hereditary} \implies \text{Eichler} \implies \text{Bass} \implies \text{Gorenstein}$$

and each of these implications is strict.

Finally, for a number ring R there is an ideal $\mathfrak{d}$ of R called the *different* whose norm is the discriminant; the different has better behavior under base extension. (There is probably some intrinsic geometric thing that the different measures.) Give the basic review; over quadratic fields, the different is just the difference between the squareroots.

In a same way, we define the *different* of a quaternion order $\mathcal{D}(\mathcal{O})$, a two-sided ideal. [[It is also just the commutator of the order, at least for a nice enough class of orders?]]

16.2 Quadratic modules over rings

To begin, we consider the theory of quadratic modules; this generalizes the theory of quadratic forms in Section 8.6 by keeping track of the codomain (target) of the quadratic map.

Let R be a (commutative) noetherian domain with field of fractions F.

Definition 16.2.1. A *quadratic map* is a map $Q : M \to N$ between R-modules, satisfying:

(i) $Q(rx) = r^2Q(x)$ for all $r \in R$ and $x \in M$; and

(ii) The map $T : M \times M \to N$ defined by
$$T(x, y) = Q(x + y) - Q(x) - Q(y)$$
is R-bilinear.

The map T in (ii) is called the *associated bilinear map*.

A *quadratic module* over R is a quadratic map $Q : M \to L$ where M is a projective R-module of finite rank and L is an invertible R-module. A *quadratic form* over R is a quadratic module with $L = R$.

Example 16.2.2. Let $Q : V \to F$ be a quadratic form. Let $M \subseteq V$ be an R-lattice such that $Q(M) \subseteq L$ where L is an invertible R-module. Then the restriction $Q|_M : M \to L$ is a quadratic module over R.

Conversely, if $Q : M \to L$ is a quadratic module over R, then the extension $Q : M \otimes_R F \to L \otimes_R F \cong F$ is a quadratic form over F.

Example 16.2.3. If $Q : M \to L$ is a quadratic module and $\mathfrak{c} \subseteq R$ is a projective R-ideal, then Q extends naturally by property (i) to a quadratic module $\mathfrak{a}M \to \mathfrak{a}^2L$.

Definition 16.2.4. A *similarity* between quadratic modules (M, L, Q) and (M', L', Q') is a pair of R-module isomorphisms $f : M \xrightarrow{\sim} M'$ and $g : L \xrightarrow{\sim} L'$ such that $Q'(f(x)) = g(Q(x))$ for all $x \in M$, i.e., such that the diagram

$$\begin{tikzcd}
M \arrow[r, "Q"] \arrow[d, "f", "\wr"'] & L \arrow[d, "g", "\wr"'] \\
M' \arrow[r, "Q'"] & L'
\end{tikzcd}$$

commutes. An *isometry* between quadratic modules is a similarity with $L = L'$ and g the identity map.

Remark 16.2.5. Similarity and isometry are different notions of "isomorphism" for quadratic modules where either you are allowed to move the codomain or not.

A similarity is a uniform These are just different notions of isomorphism, depending on if you can move the target or not. The notion of "isometry" comes from the connection with measuring lengths.

Think of a similarity like in the quaternion algebra case, where we can post multiply by any nonzero element of the field and we get an isomorphic conic. This may change the values of the quadratic form: $q(x)$ and $uq(x)$ are similar but not necessarily isometric.

If free, then the same theory works fine.

Discriminant, odd and even dimensions.

16.3 Connection with ternary quadratic forms

In this section, we classify quaternion orders over R in terms of ternary quadratic modules, due to Brzezinski. There is a further connection with quaternary forms, but this is for ideals; see the next section.

There are actually two different ways of doing it. I would attribute them to the authors as follows.
The one of Brandt/Eichler/Peters/Brzsinski which rescales the lattice by nr(L)^-1 in the definition of O(L). This way O(L)=O(aL) for each fractional ideal a. This works for any Dedekind ring and induces bijections between similarity classes of forms and isomorphism classes of Gorenstein orders.

The other one used by Pall/Nipp/Lemurell. Their definition of O(L) does not use the factor nr(L)^-1.
Over a PID this induces a bijection between similarity classes of forms and isomorphism classes of arbitrary orders.

There is a very detailed description by Shimura that works out the image of this second map over non PIDs.

We return to the notation we have used throughout this section (except the previous one, where we allowed more generality): let R be a Dedekind domain with field of fractions F and let $\mathcal{O}$ be an R-order in a quaternion algebra B over F.

Theorem 16.3.1. *There is a discriminant-preserving bijection between quaternion orders $\mathcal{O}$ over R and twisted isometry classes of ternary quadratic modules over R with nonzero discriminant.*

We can be more specific; it's also functorial.

Let $\mathcal{O}^\sharp \subset B$ be the dual of $\mathcal{O}$. Then

$$(\mathcal{O}^\sharp)^0 = \mathcal{O}^\sharp \cap B^0 = \{\alpha \in \mathcal{O}^\sharp : \operatorname{trd}(\alpha) = 0\}$$

is projective (locally free) of rank 3 so projective. We associate to $\mathcal{O}$ the quadratic map $Q(\mathcal{O}) = \operatorname{nrd} : \mathcal{O}^\sharp \cap B^0 \to \mathfrak{d}^{-2}$. [[This should be the same thing as taking the inverse different!]]

Remark 16.3.2. Over a field, this is the same. See also Lucianovic's thesis. Over odd characteristic fields this makes no difference.

Lemma 16.3.3. *$Q(\mathcal{O})$ is a quadratic module.*

Proof. □

Compute an explicit representation when free. Check morphisms (similarity).

Example 16.3.4. The Clifford functor respects morphisms of quadratic spaces but the inverse does not. In particular, maximal quadratic spaces do not correspond to maximal orders: the correspondence is only in one direction. There's a specific example where bad things happen at 2, but the problem is more general than that.

To show we get a bijection, we work in reverse. Let $M = \mathfrak{a}_1 e_1 \oplus \mathfrak{a}_2 e_2 \oplus \mathfrak{a}_3 e_3$ and $I = R$ and the quadratic form $Q : M \to R$ by

$$Q(xe_1 + ye_2 + ze_3) = Q(x, y, z) = ax^2 + by^2 + cz^2 + uyz + vxz + wxy, \tag{16.3.5}$$

with $a, b, c, u, v, w \in R$. We just write down an algebra:

$$B = R \oplus Ri \oplus Rj \oplus Rk$$

with multiplication laws

$$\begin{aligned}
i^2 &= ui - bc & jk &= a\bar{i} = a(u - i) & kj &= \overline{\bar{j}\,\bar{k}} = -vw + ai + wj + vk \\
j^2 &= vj - ac & ki &= b\bar{j} = b(v - j) & ik &= \overline{\bar{k}\,\bar{i}} = -uw + wi + bj + uk \\
k^2 &= wk - ab & ij &= c\bar{k} = c(w - k) & ji &= \overline{\bar{i}\,\bar{j}} = -uv + vi + uj + ck
\end{aligned} \tag{Q}$$

Check various things (exercise). This construction has been attributed to Eichler and appears in Brzezinski [Brz83] in the case $R = \mathbb{Z}$. In this association, the reversal map corresponds to the standard involution $\bar{\ }$ on B.

Remark 16.3.6. In fact, this is the *Clifford algebra* of M.

Proof. Finish proof. □

OK, now to read off invariants.

Proposition 16.3.7. *If $Q' = bQ$ then $\mathfrak{d}(\mathcal{O}') = b\mathfrak{d}(\mathcal{O})$. In particular, if Q is primitive then $\mathfrak{d}(\mathcal{O}') = (b)$ and $G(\mathcal{O}') = \mathcal{O}$.*

Remark 16.3.8. Relationship to the *discriminant* quadratic form.

Musing: is there any relationship to representation numbers for this form? E.g. splitting?

16.4 Hereditary orders

Definition 16.4.1. An order $\mathcal{O}$ is *hereditary* if every right fractional $\mathcal{O}$-ideal is invertible.

[[Nope! Hereditary is for one-sided ideals. Unless these are related?]]

16.4.2. Let $\mathcal{O} \subseteq B$ be a maximal order. Then $I \subseteq \mathcal{O}$ is a (maximal) invertible two-sided ideal if and only if $I_{\mathfrak{p}}$ is so for all primes $\mathfrak{p}$.

We have two cases: either $B_{\mathfrak{p}} \cong \mathrm{M}_2(F_{\mathfrak{p}})$ or $B_{\mathfrak{p}}$ is a division ring. In the former case, we have that $\mathcal{O}_{\mathfrak{p}} \cong \mathrm{M}_2(R_{\mathfrak{p}})$ and $I_{\mathfrak{p}} = \mathrm{M}_2(\mathfrak{a}_{\mathfrak{p}}) = \mathfrak{a}_{\mathfrak{p}} \mathrm{M}_2(R)$ for some $\mathfrak{a}_{\mathfrak{p}} \subseteq R_{\mathfrak{p}}$, and such an ideal is principal since $\mathfrak{a}_{\mathfrak{p}}$ is. If $B_{\mathfrak{p}}$ is a division ring, then all (two-sided) ideals of the maximal order $\mathcal{O}_{\mathfrak{p}}$ are generated by a power of the unique maximal ideal $P_{\mathfrak{p}}$ which is also principal.

Thus, all two-sided fractional $\mathcal{O}$-ideals are locally principal, so $\mathcal{O}$ is hereditary, and the group of invertible two-sided fractional $\mathcal{O}$-ideals is generated by $\{P_{\mathfrak{p}} \cap \mathcal{O} : \mathfrak{p} \text{ ramified in } B\} \cup \{\mathfrak{p}\mathcal{O} : \mathfrak{p} \subseteq R \text{ prime}\}$.

[[Same with ideal classes?]]

Definition 16.4.3. Let P be a two-sided integral $\mathcal{O}$-ideal and suppose $P \neq \mathcal{O}$. We say that P is *prime* if $IJ \subseteq P$ implies $I \subseteq P$ or $J \subseteq P$, where I and J are two-sided integral $\mathcal{O}$-ideals.

Theorem 16.4.4. *Suppose that $\mathcal{O}$ is hereditary. Then the set of two-sided (invertible) fractional $\mathcal{O}$-ideals of B forms an abelian group under multiplication, generated by the prime ideals.*

Proof. Let I be a two-sided integral $\mathcal{O}$-ideal. Then since $\mathcal{O}$ is a finitely generated R-module and R is noetherian, we conclude that I is contained in a proper maximal (integral) $\mathcal{O}$-ideal M. From $I \subseteq M$ we conclude that $IM^{-1} \subseteq \mathcal{O}$, so IM^{-1} is integral. But $\operatorname{nrd}(IM^{-1}) = \operatorname{nrd}(I)/\operatorname{nrd}(M) \mid \operatorname{nrd}(I)$. It follows that I can be written as the product of maximal ideals M by induction (on the reduced norm).

We will now show that in fact a maximal $\mathcal{O}$-ideal is prime. For suppose that $IJ \subseteq M$ and that $I \not\subseteq M$. Then $I + M$ is a two-sided $\mathcal{O}$-ideal strictly containing M so $I + M = \mathcal{O}$. But then $J = IJ + MJ \subseteq M$. Conversely, if P is prime and $P \subseteq I$ where I is a proper two-sided integral $\mathcal{O}$-ideal, then $P = I(I^{-1}P)$, but $I^{-1}P \subseteq P$ implies $I^{-1} \subseteq \mathcal{O}$ which is impossible so $I \subseteq P$ hence $P = I$.

To conclude, we show that this group is abelian. Let P, Q be prime ideals. Then $PQ \subseteq P$, so as above PQP^{-1} is integral, say $PQP^{-1} = Q'$. If $Q' = \mathcal{O}$ then $Q = \mathcal{O}$, a contradiction. But then choosing $0 \neq p \in P \cap R$ then $Q = pQp^{-1} \subseteq Q'$, but Q is maximal so $Q = Q'$. Thus $PQ = Q'P$, so the group is abelian. □

16.5 Eichler orders

An *Eichler order* is the intersection of two maximal orders, and it is this class of orders which we will study throughout. The *level* of an Eichler order $\mathcal{O}$ is the ideal $\mathfrak{N} \subseteq R$ satisfying $\mathfrak{d} = \mathfrak{D}\mathfrak{N}$; the level $\mathfrak{N}$ is coprime to the discriminant $\mathfrak{D}$ of B. Alternatively, given a maximal order $\mathcal{O} \subseteq B$, an ideal $\mathfrak{N}$ coprime to $\mathfrak{D}$ and an embedding $\iota_{\mathfrak{N}} : \mathcal{O} \hookrightarrow \mathrm{M}_2(\mathbb{Z}_{F,\mathfrak{N}})$ where $\mathbb{Z}_{F,\mathfrak{N}}$ denotes the completion of $\mathbb{Z}_F$ at $\mathfrak{N}$, an Eichler order of level $\mathfrak{N}$ is given by

$$\mathcal{O}_0(\mathfrak{N}) = \{\gamma \in \mathcal{O} : \iota_{\mathfrak{N}}(\gamma) \text{ is upper triangular modulo } \mathfrak{N}\}\,, \tag{16.5.1}$$

and all Eichler orders arise in this way up to conjugation. In particular [Rei03, Theorem 39.14], an order $\mathcal{O}$ is hereditary (all one-sided ideals of $\mathcal{O}$ are projective) if and only if $\mathcal{O}$ is an Eichler order with squarefree level.

Being Eichler is a local condition.

Two orders $\mathcal{O}, \mathcal{O}'$ are *conjugate* (also *isomorphic* or *of the same type*) if there exists $\nu \in B^*$ such that $\mathcal{O}' = \nu^{-1}\mathcal{O}\nu$, and we write $\mathcal{O} \cong \mathcal{O}'$.

Proposition 16.5.2 ([Vig80, Corollaire III.5.5])**.** *The number of isomorphism classes of Eichler orders $\mathcal{O} \subseteq B$ of level $\mathfrak{N}$ is finite.*

16.6 Gorenstein orders

Definition 16.6.1. $\mathcal{O}$ is *Gorenstein* if $\mathcal{O}^\sharp = \mathrm{Hom}(\mathcal{O}, R) = \{\alpha \in B : \mathrm{trd}(\alpha\mathcal{O}) \subseteq R\}$ [[why is this equivalent?]] is projective as a (left or) right $\mathcal{O}$-module.

Lemma 16.6.2. *$\mathcal{O}$ is Gorenstein if and only if $\mathcal{O}_{\mathfrak{p}}$ is Gorenstein for all primes $\mathfrak{p}$.*

Let $\mathcal{O}$ be an order.

Proposition 16.6.3. *There exists a unique Gorenstein order $G(\mathcal{O}) \subset \mathcal{O}$ and a unique ideal $\mathfrak{b}(\mathcal{O}) \subset R$ such that*

$$\mathcal{O} = R + \mathfrak{b}(\mathcal{O})G(\mathcal{O}).$$

Proof. □

Lemma 16.6.4. *We have $\mathcal{O} \cong \mathcal{O}'$ if and only if $\mathfrak{b}(\mathcal{O}) = \mathfrak{b}(\mathcal{O}')$ and $G(\mathcal{O}) \cong G(\mathcal{O}')$.*

Proof. □

Explicit example (Santi Molina).

16.7 Different

As in the commutative case, the discriminant can be realized as the norm of the different, which here is two-sided and invertible.

Let B be a separable F-algebra and let I be an R-lattice in B.

Definition 16.7.1. The *dual* of I is

$$I^\sharp = \{\alpha \in B : \operatorname{trd}(\alpha I) = \operatorname{trd}(I\alpha) \subseteq R\}.$$

In particular, we have $\operatorname{trd}(I^\sharp I) \subseteq R$.

16.7.2. $I^\sharp$ is an R-lattice in B (Exercise 13.8). Clearly, if $I \subseteq J$ then $I^\sharp \supset J^\sharp$. For all $\beta \in B^\times$ we have $(\beta I)^\sharp = I^\sharp \beta^{-1}$.

16.7.3. We have $I \subseteq (I^\sharp)^\sharp$, since if $\alpha \in I$ and $\beta \in I^\sharp$ then $\operatorname{trd}(\alpha\beta) \subseteq R$ so $\alpha \in (I^\sharp)^\sharp$.

Lemma 16.7.4. *We have $\mathcal{O}_{\mathrm{R}}(I) \subseteq \mathcal{O}_{\mathrm{L}}(I^\sharp) \subseteq \mathcal{O}_{\mathrm{R}}((I^\sharp)^\sharp)$, so equality holds if $I = (I^\sharp)^\sharp$.*

Of course, a similar statement holds in Lemma 16.7.4, interchanging left and right.

Proof. Let $\alpha \in \mathcal{O}_{\mathrm{R}}(I)$; then $I\alpha \subseteq I$, so $I^\sharp I\alpha \subseteq I^\sharp I$ so

$$\operatorname{trd}(\alpha I^\sharp I) = \operatorname{trd}(I^\sharp I\alpha) \subseteq \operatorname{trd}(I^\sharp I) \subseteq R$$

hence $\alpha I^\sharp \subseteq I^\sharp$ and $\alpha \in \mathcal{O}_{\mathrm{L}}(I^\sharp)$. Conversely, if $\alpha \in \mathcal{O}_{\mathrm{L}}(I^\sharp)$ so $\alpha I^\sharp \subseteq I^\sharp$, then $(\alpha I^\sharp)^\sharp = (I^\sharp)^\sharp \alpha^{-1} \supseteq (I^\sharp)^\sharp$ so $(I^\sharp)^\sharp \supseteq (I^\sharp)^\sharp \alpha$ and $\alpha \in \mathcal{O}_{\mathrm{R}}((I^\sharp)^\sharp)$. □

Now let $\mathcal{O}$ be an R-order.

Definition 16.7.5. We define the *different* of $\mathcal{O}$ to be $\mathcal{O}^* = (\mathcal{O} : \mathcal{O}^\sharp)_{\mathrm{L}} = (\mathcal{O} : \mathcal{O}^\sharp)_{\mathrm{R}}$.

[[Would Lenstra use a † here? What is $((\mathcal{O})^\sharp)^\sharp$?]]

Lemma 16.7.6. *We have $\mathcal{O}^*$ is a two-sided integral ideal of $\mathcal{O}$ with* $\operatorname{nrd}(\mathcal{O}^*)^2 = \operatorname{disc}(\mathcal{O})$.

In particular, we may define the *reduced discriminant* of $\mathcal{O}$ to be $\operatorname{discrd}(\mathcal{O}) = \operatorname{nrd}(\mathcal{O}^*)$ and this agrees with the definition in the quaternion case.

16.8 Other orders

Definition 16.8.1. An order is *Bass* if every order $\mathcal{O}' \supseteq \mathcal{O}$ is Gorenstein.

Gorenstein is weaker than Bass is weaker than Eichler is weaker than hereditary is weaker than maximal.

Lemma 16.8.2. *$\mathcal{O}$ is Bass if and only if $\mathcal{O}_{\mathfrak{p}}$ is Bass for all primes $\mathfrak{p}$.*

Finally, Eichler introduced a class of orders. Eichler

Definition 16.8.3. An order $\mathcal{O}$ is *primitive* if it contains a maximal R-order of a quadratic subfield of B.

Lemma 16.8.4. *A primitive order is a Bass order.*

Proof. □

Is every Bass order primitive? True in the local case and for rational orders

Lemma 16.8.5. *$\mathcal{O}$ is primitive if and only if $\mathcal{O}_{\mathfrak{p}}$ is primitive for all primes $\mathfrak{p}$ of R.*

Proof. Local global. □

16.9 Extensions and further reading

Condition (ii) can be given purely in terms of Q

$$Q(x+y+z) = Q(x+y) + Q(x+z) + Q(y+z) - Q(x) - Q(y) - Q(z)$$

for all $x, y, z \in M$.

16.9.1. In some lattice contexts, with R a Dedekind domain, a quadratic form with values in a fractional ideal $\mathfrak{a}$ is called an $\mathfrak{a}$-*modular quadratic form*. Given the overloaded meanings of the word *modular*, we do not employ this terminology. In the geometric context, a quadratic module is called a *line-bundle valued quadratic form*. Sometimes called $\mathfrak{a}$-modular quadratic form or something. Sometimes you just want to keep track of where the image lies.

First construction is due to Brandt clarified by Peters and generalizing a result of Latimer General form due to Brandt Lemurell (also writing under the previous family name Johansson), discusses the relationship between quaternion algebras and orders and ternary quadratic forms [Lem11]. Also consult history in Gross and Lucianovic.

For more on quadratic forms over rings, see O'Meara, Knus, Scharlau, ...

Other treatment by Gross and Lucianovic. Most general theorem is work of the author [Voi11a].

Exercises

16.1. Finite intersection of maximal orders?

16.2. Let E be a field and let A be a F a finite extension of E [[Lassina's exercise: there exists a Galois stable maximal order.]]

16.3. Prove the statement in Example 16.2.3. Conclude that up to similarity, the target of a quadratic module only depends on $\operatorname{Pic}(R)/2\operatorname{Pic}(R)$. Twisted discriminants.

Chapter 17

Quaternion orders over local PIDs

Carry out explicit local descriptions of all orders. Throughout, let F be a nonarchimedean local field, R its ring of integers, $\mathfrak{p} = (\pi)$ its maximal ideal, and $k = R/\mathfrak{p}$ its (finite) residue field.

17.1 Eichler symbol

Let $\mathcal{O}$ be an R-order. We want to know if the order gives rise to a split, ramified, or inert extension of the residue field.

Let $J(\mathcal{O})$ be the Jacobson radical of $\mathcal{O}$.

Lemma 17.1.1. *$J(\mathcal{O}) = \mathcal{O}$ if and only if $\mathcal{O} \cong \mathrm{M}_2(R)$.*

Proof. Exercise, using Wedderburn. □

Suppose now that $J(\mathcal{O}) \neq \mathcal{O}$; then $J(\mathcal{O}) \supseteq \mathfrak{p}\mathcal{O}$ and $\mathcal{O}/J(\mathcal{O})$ is a semisimple k-algebra, and hence a product of fields.

Exercise: only three possibilities.

Accordingly, we make the following definition.

Definition 17.1.2. Define the *Eichler symbol*

$$e(\mathcal{O}) = \begin{cases} 1, & \text{if } J(\mathcal{O}) = R \text{ or } \mathcal{O}/J(\mathcal{O}) \cong k \times k; \\ 0, & \text{if } \mathcal{O}/J(\mathcal{O}) \cong k; \\ -1, & \text{if } \mathcal{O}/J(\mathcal{O}) \text{ is a quadratic field extension of } k. \end{cases} .$$

Remark 17.1.3. Recall the definition of the discriminant quadratic form (Remark 16.3.8). Then: $e(\mathcal{O}) = 0$ if and only if Δ is identically zero modulo $\mathfrak{p}$; and if Δ is not identically zero, then $e(\mathcal{O}) = 1$ or -1 according as if $\left(\frac{\Delta(\alpha)}{\mathfrak{p}}\right) = -1$ for some $\alpha \in \mathcal{O}$.

This is like wanting the separable quadratic field extension inside the quaternion algebra, which was so important in Chapter ...

In the language of quadratic forms, the Eichler symbol becomes:

$$e_{\mathfrak{p}}(\mathcal{O}) = \begin{cases} -1, & \text{if } q \text{ modulo } \mathfrak{p} \text{ is irreducible} \\ 0, & \text{if } q \text{ modulo } \mathfrak{p} \text{ is the square of a linear factor} \\ 1, & \text{if } q \text{ modulo } \mathfrak{p} \text{ is the product of two different linear factors.} \end{cases} \tag{17.1.4}$$

This is just recording the reduction type of the associated conic in the plane, and hence also the definition (inert, ramified, split).

17.2 Odd characteristic

By Lemma 16.6.4, it is enough to characterize local Gorenstein orders, and by ... these correspond to similarity classes of primitive ternary quadratic forms over R. We begin with the case where k has odd characteristic.

Let ϵ be a quadratic nonresidue modulo $\mathfrak{p}$, so that $R^\times/R^{\times 2}$ is represented by the classes of $1, \epsilon$.

Proposition 17.2.1. *Let Q be a primitive nondegenerate ternary quadratic form. Then Q is similar to a unique quadratic form*

$$Q \sim \langle 1, u\pi^e, v\pi^f \rangle$$

where $u, v \in \{1, \epsilon\}$ and $0 \leq e \leq f$ satisfying

$$???$$

Proof. Diagonalize the form. □

We call a form as in Proposition 17.2.1 is in *standard form*.

Proof. Immediate from (17.1.4). □

Lemma 17.2.2. *The associated quaternion order $\mathcal{O} = C^0(Q)$ is maximal if and only if $r = s = 0$ or ($r = s = 1$ and $v = \epsilon$) and Bass if and only if $r \leq 1$.*

Proof. Computation. □

Tree of orders.

17.3 Even characteristic

Atomic. Do explicitly for some cases like unramified extension of $\mathbb{Z}_2$ and totally ramified extension $\mathbb{Z}_2[\sqrt{2}]$?

17.4 Extensions and further reading

Discussed by Lemurell [Lem11, §5].

Exercises

17.1. Let $\mathcal{O}$ be a local quaternion order. Show that if $\mathcal{O}$ is not a local Bass order, then $e(\mathcal{O}) = 0$.

17.2. Prove Lemma 17.2.2.

17.3. Let Q be a ternary quadratic form in standard form and let $\mathcal{O} = C^0(Q)$ be the associated quaternion order. Then

$$e(\mathcal{O}) = \pm 1 \text{ if and only if } e = 0 \text{ and } f \geq 1 \text{ and } (-u/\mathfrak{p}) = \pm 1.$$

Chapter 18

Zeta functions and the mass formula

In this chapter, we introduce zeta functions of a quaternion order and use them to investigate the class number of a totally definite quaternion order.

18.1 Zeta functions of quadratic fields

Gauss, in his investigation of binary quadratic forms was led to conjecture that there were finitely many imaginary quadratic orders of class number 1. There are many approaches to this problem, involving some beautiful and deep mathematics. Given that we want to prove some kind of lower bound for the class number in terms of the discriminant, it is natural to seek an analytic expression for this class number: this is provided by the *analytic class number formula* of Dirichlet, and it turns the class number problem of Gauss into a (still hard, but tractable) problem of estimation. In a similar way, we may ask: what are the definite quaternion orders of class number 1? The method to prove Dirichlet's formula generalizes to quaternion orders as well, as pursued by Eichler in his *mass formula*.

In this chapter, we treat these topics in detail. To introduce the circle of ideas, let $K = \mathbb{Q}(\sqrt{d})$ be a quadratic field of discriminant $d \in \mathbb{Z}$ and let R be its ring of integers. We encode information about the field K by its *zeta function*. Over $\mathbb{Q}$, we would consider the *Riemann zeta function*

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \tag{18.1.1}$$

as the prototypical such function; this series converges for $\operatorname{Re} s > 1$, by the comparison test. By unique factorization, we have an *Euler product*

$$\zeta(s) = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}$$

where the product is over all primes p. The function $\zeta(s)$ can be meromorphically continued to the right half-plane $\operatorname{Re} s > 0$ using the fact that the sum

$$\zeta_2(s) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n^s}$$

converges for $\operatorname{Re} s > 0$ and

$$\zeta(s) + \zeta_2(s) = 2^{1-s}\zeta(s)$$

so that

$$\zeta(s) = \frac{1}{2^{1-s} - 1}\zeta_2(s)$$

and the right-hand side makes sense for any $\operatorname{Re} s > 0$ except for possible poles where $2^{1-s} = 1$. For real values of $s > 1$, we have

$$\frac{1}{s-1} = \int_1^{\infty} \frac{dx}{x^s} \leq \zeta(s) \leq 1 + \int_1^{\infty} \frac{dx}{x^s} = \frac{s}{s-1}$$

so

$$1 \leq (s-1)\zeta(s) \leq s;$$

therefore, as $s \to^+ 1$, we have $(s-1)\zeta(s) \to 1$, so $\zeta(s)$ has a simple pole at $s = 1$ with residue $\operatorname{res}_{s=1} \zeta(s) = 1$.

For the field K, modeled after (18.1.1) we define the *Dedekind zeta function* by

$$\zeta_K(s) = \sum_{\mathfrak{a} \subseteq R} \frac{1}{\mathrm{N}(\mathfrak{a})^s} \tag{18.1.2}$$

where the sum is over all *nonzero* ideals of R and the series is defined for $\operatorname{Re} s > 1$. We can also write this as a *Dirichlet series*

$$\zeta_K(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}$$

where a_n is the number of ideals of norm n in R. By unique factorization of ideals, we again have an Euler product expansion

$$\zeta_K(s) = \prod_{\mathfrak{p}} \left(1 - \frac{1}{\mathrm{N}\,\mathfrak{p}^s}\right)^{-1}, \tag{18.1.3}$$

the product over all nonzero prime ideals $\mathfrak{p} \subset R$.

18.2 Analytic class number formula for imaginary quadratic fields

In order to introduce a formula that involves the class number, we group the ideals in (18.1.2) by their ideal class: for $[\mathfrak{b}] \in \mathrm{Cl}(K)$, we define

$$\zeta_{K,[\mathfrak{b}]}(s) = \sum_{\substack{\mathfrak{a} \subseteq R \\ [\mathfrak{a}]=[\mathfrak{b}]}} \frac{1}{\mathrm{N}\,\mathfrak{a}^s}$$

so that

$$\zeta_K(s) = \sum_{[\mathfrak{b}] \in \mathrm{Cl}(K)} \zeta_{K,[\mathfrak{b}]}(s). \tag{18.2.1}$$

In general, for $[\mathfrak{b}] \in \mathrm{Cl}(K)$, we have $[\mathfrak{a}] = [\mathfrak{b}]$ if and only if there exists $a \in K^\times$ such that $\mathfrak{a} = a\mathfrak{b}$, but since $\mathfrak{a} \subseteq R$, in fact we have

$$a \in \mathfrak{b}^{-1} = \{a \in R : a\mathfrak{b}^{-1} \subseteq R\};$$

this gives a bijection

$$\{\mathfrak{a} \subseteq R : [\mathfrak{a}] = [\mathfrak{b}]\} \leftrightarrow \mathfrak{b}^{-1}/R^\times,$$

(since the generator of an ideal is unique up to units). Thus

$$\zeta_{K,[\mathfrak{b}]}(s) = \frac{1}{\mathrm{N}\,\mathfrak{b}^s} \sum_{0 \neq a \in \mathfrak{b}^{-1}/R^\times} \frac{1}{(\mathrm{N}\,a)^s}. \tag{18.2.2}$$

for each class $[\mathfrak{b}] \in \mathrm{Cl}(K)$.

Everything we have done so far works equally as well for real as for imaginary quadratic fields. But here, to make sense of $\mathfrak{b}^{-1}/R^\times$ in the simplest case, we want $R^\times$ to be a finite group, which means exactly that K is imaginary quadratic. So from now on this section, we assume $d < 0$. Then $w = \#R^\times = 2$, except when $d = -3, -4$ where $w = 6, 4$, respectively.

Under this hypothesis, the sum (18.2.2) can be transformed into sum over lattice points with the fixed factor w of overcounting. Before estimating the sum over reciprocal norms, we first estimate the count. Let $\Lambda \subset \mathbb{C}$ be a lattice. We can estimate the number of lattice points $\lambda \in \Lambda$ with $|\lambda| \leq x$ by the ratio $\pi x^2/A$, where A is the area of a fundamental parallelogram P for Λ: roughly speaking, this says that we can tile a circle of radius x with approximately $\pi x^2/A$ parallelograms P.

More precisely, we have the following lemma.

Lemma 18.2.3. *Let $\Lambda \subset \mathbb{C}$ be a lattice with coarea* $\operatorname{area}(\mathbb{C}/\Lambda) = A$*. Then there is a constant C such that for all $x > 1$, we have*

$$\left| \#\{\lambda \in \Lambda : |\lambda| \leq x\} - \frac{\pi x^2}{A} \right| \leq Cx.$$

We leave this lemma as an exercise (Exercise 18.1) in tiling a circle with radius x with fundamental parallelograms for the lattice Λ.

Now we apply this lemma to a lattice $\mathfrak{b}^{-1} \subset \mathbb{C}$. We write

$$\zeta_{K,[\mathfrak{b}]}(s) = \frac{1}{w(\mathrm{N}\,\mathfrak{b}^s)} \sum_{b=1}^{\infty} \frac{b_n}{n^s}$$

where

$$b_n = \#\{a \in \mathfrak{b}^{-1} : \mathrm{N}\,a = n\}.$$

Since $\mathrm{N}\,a = |a|^2$, for all $x > 1$ we have

$$\sum_{n \leq x} b_n = \#\{a \in \mathfrak{b}^{-1} : 0 < |a| \leq \sqrt{x}\};$$

from Lemma 18.2.3, we conclude

$$\left| \sum_{n \leq x} b_n - \frac{\pi x}{A} \right| \leq C\sqrt{x}$$

where A is the coarea of $\mathfrak{b}^{-1}$ and C is a constant that does not depend on x. We compute that

$$A = N(\mathfrak{b}^{-1}) \frac{\sqrt{|d|}}{2}.$$

Now consider the Dirichlet series

$$f(s) = \frac{1}{w(N\mathfrak{b})^s} \sum_{n=1}^{\infty} \left(b_n - \frac{\pi}{A} \right) \frac{1}{n^s}.$$

Then the estimate

$$\left| \sum_{n \leq x} \left(b_n - \frac{\pi}{A} \right) \right| = \left| \sum_{n \leq x} b_n - \frac{\pi x}{A} \right| \leq C\sqrt{x}$$

by the comparison test implies that $f(s)$ converges for all $\operatorname{Re} s > 1/2$ and in particular $f(s)$ converges at $s = 1$. For $s > 1$, we have

$$f(s) = \zeta_{K,[\mathfrak{b}]}(s) - \frac{\pi}{Aw(N\mathfrak{b})^s} \zeta(s)$$

so

$$\zeta_{K,[\mathfrak{b}]}(s) = f(s) + \frac{2\pi}{w\sqrt{|d|}}(N\mathfrak{b})^{1-s}\zeta(s).$$

hence

$$\begin{aligned}\operatorname{res}_{s=1}\zeta_K(s) &= \lim_{s\to^+1}(s-1)\zeta_{K,[\mathfrak{b}]}(s)\\ &= \lim_{s\to^+1}(s-1)f(s) + \frac{2\pi}{w\sqrt{|d|}}\lim_{s\to^+1}(s-1)(N\mathfrak{b})^{1-s}\zeta(s)\\ &= 0 + \frac{2\pi}{w\sqrt{|d|}}\cdot 1 = \frac{2\pi}{w\sqrt{|d|}}.\end{aligned}$$

In particular, $\zeta_{K,[\mathfrak{b}]}(s)$ has a simple pole at $s = 1$ with residue independent of $[\mathfrak{b}]$. Summing the residues over $[\mathfrak{b}] \in \mathrm{Cl}(K)$, from (18.2.1) we have the following result.

Theorem 18.2.4 (Analytic class number formula for imaginary quadratic field)**.** *Let $K = \mathbb{Q}(\sqrt{d})$ be an imaginary quadratic field with discriminant $d < 0$. Let h be the class number of K and w the number of roots of unity in K. Then*

$$\operatorname{res}_{s=1}\zeta_K(s) = \frac{2\pi h}{w\sqrt{|d|}}.$$

This formula simplifies slightly if we cancel the pole at $s = 1$ with $\zeta(s)$, as follows. Like in the Dirichlet series, we can combine terms in (18.1.3) to get

$$\zeta_K(s) = \prod_p \prod_{\mathfrak{p}|p}\left(1 - \frac{1}{\mathrm{N}\,\mathfrak{p}^s}\right)^{-1}$$

and

$$L_p(s) = \prod_{\mathfrak{p}|p}\left(1 - \frac{1}{\mathrm{N}\,\mathfrak{p}^s}\right)^{-1} = \begin{cases}\left(1-p^{-s}\right)^{-2}, & \text{if } (p) = \mathfrak{p}\mathfrak{p}' \text{ splits in } K;\\ \left(1-p^{-s}\right)^{-1}, & \text{if } (p) = \mathfrak{p}^2 \text{ ramifies in } K;\\ \left(1-p^{-2s}\right)^{-1}, & \text{if } (p) \text{ is inert in } K.\end{cases}$$

We condition of being split, ramified, or inert in K is recorded in a character:

$$\chi(p) = \chi_d(p) = \begin{cases}1, & \text{if } p \text{ splits in } K;\\ 0, & \text{if } p \text{ ramifies in } K;\\ -1, & \text{if } p \text{ is inert in } K\end{cases}$$

for prime p and extended to all positive integers by multiplicativity. If $p \nmid d$ is an odd prime, then

$$\chi(p) = \left(\frac{d}{p}\right)$$

is the usual Legendre symbol, equal to 1 or -1 according as if d is a quadratic residue or not modulo p. Then

$$L_p(s) = (1 - p^{-s})(1 - \chi(p)p^{-s}).$$

Expanding the Euler product term-by-term and taking a limit, we have

$$\zeta_K(s) = \zeta(s)L(s,\chi) \tag{18.2.5}$$

where

$$L(s,\chi) = \prod_p \left(1 - \frac{\chi(p)}{p^s}\right)^{-1} = \sum_n \frac{\chi(n)}{n^s}.$$

The function $L(s,\chi)$ is in fact holomorphic for all $\operatorname{Re} s > 0$; this follows from the fact that the partial sums $\sum_{n \leq x} \chi(n)$ are bounded and the mean value theorem. So in particular the series

$$L(1,\chi) = 1 + \frac{\chi(2)}{2} + \frac{\chi(3)}{3} + \frac{\chi(4)}{4} + \dots$$

converges (slowly). Combining (18.2.5) with the analytic class number formula yields:

$$L(1,\chi) = \frac{2\pi h}{w\sqrt{|d|}}.$$

For example, taking $d = -4$, so $\chi(2) = 0$ and $\chi(p) = (-1/p) = (-1)^{(p-1)/2}$, we have

$$L(1,\chi) = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \dots = \prod_{p \geq 3} \left(1 - \frac{(-1)^{(p-1)/2}}{p}\right)^{-1} = \frac{\pi}{4} = 0.7853\dots.$$

Remark 18.2.6. The fact that $L(1,\chi) \neq 0$, and its generalization to complex characters χ, is the key ingredient to prove Dirichlet's theorem on primes in arithmetic progression (Theorem 11.2.8), used in the classification of quaternion algebras over $\mathbb{Q}$. The arguments to complete the proof are requested in Exercise 18.8.4.

To approach the class number problem of Gauss, we would then seek lower bounds on $L(1,\chi)$ in terms of the discriminant $|d|$. Rather than go into these estimates here, we refer to the additional reading in Paragraph 18.8.1.

18.3 Eichler mass formula over the rationals

We are now prepared to consider the analogue of the above for quaternion orders. Let B be a quaternion algebra over $\mathbb{Q}$ and let $\mathcal{O}$ be a maximal order in B. We define the *zeta function* of $\mathcal{O}$ to be

$$\zeta_{\mathcal{O}}(s) = \sum_{I \subseteq \mathcal{O}} \frac{1}{\mathrm{N}(I)^s}, \tag{18.3.1}$$

the sum over all nonzero right ideals in $\mathcal{O}$. A maximal order is hereditary, so every right ideal is projective hence locally principal (and invertible), and here we need not concern ourselves with the subtler aspects of ideal theory as in Chapter 14.

Let a_n be the number of right ideals in $\mathcal{O}$ of reduced norm (n) for $n > 0$. Then $\mathrm{N}(I) = \mathrm{nrd}(I)^2$, so

$$\zeta_{\mathcal{O}}(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^{2s}}.$$

To establish an Euler product for $\zeta_{\mathcal{O}}(s)$, we will give a kind of factorization formula for right ideals of $\mathcal{O}$, but by necessity, writing an ideal as a compatible product will involve the entire set of orders connected to $\mathcal{O}$; in any case, we will show that $a_{mn} = a_m a_n$ whenever m, n are coprime. Next, we will count the ideals of a given reduced norm $q = p^e$ a power of a prime: the answer will depend on whether p is ramified or not in B. As a result, we will find

$$\zeta_{\mathcal{O}}(s) = \prod_p \zeta_{\mathcal{O},p}(s)$$

where

$$\zeta_{\mathcal{O},p}(s) = \begin{cases} \left(1 - p^{-2s}\right)^{-1}, & \text{if } p \text{ is ramified;} \\ \left(1 - p^{-2s}\right)^{-1}\left(1 - p^{1-2s}\right)^{-1}, & \text{if } p \text{ is unramified.} \end{cases} \tag{18.3.2}$$

In particular, this formula shows that $\zeta_{\mathcal{O}}(s)$ does not depend on the choice of maximal order $\mathcal{O}$, so we may also write $\zeta_B(s) = \zeta_{\mathcal{O}}(s)$. From (18.3.2), we have

$$\zeta_{\mathcal{O}}(s) = \zeta(2s)\zeta(2s-1)\prod_{p \mid D}\left(1 - \frac{1}{p^{2s-1}}\right)$$

where $D > 0$ is the discriminant of B. In particular, we have

$$\operatorname{res}_{s=1} \zeta_{\mathcal{O}}(s) = \lim_{s \to^+ 1} (s-1)\zeta_{\mathcal{O}}(s) = \frac{\zeta(2)}{2}\prod_{p \mid D}\left(1 - \frac{1}{p}\right) = \frac{\pi^2}{12}\prod_{p \mid D}\left(1 - \frac{1}{p}\right). \tag{18.3.3}$$

(We could also look to cancel the poles of $\zeta_{\mathcal{O}}(s)$ in a similar way to define an L-function, holomorphic for $\operatorname{Re} s > 0$.)

Now we break up the sum (18.3.1) according to right ideal class:

$$\zeta_{\mathcal{O}}(s) = \sum_{[J]\in \operatorname{Cl}(\mathcal{O})} \zeta_{\mathcal{O},[J]}(s)$$

where

$$\zeta_{\mathcal{O},[J]}(s) = \sum_{\substack{I\subseteq \mathcal{O}\\ [I]=[J]}} \frac{1}{\operatorname{N}(I)^s}.$$

Since $[I] = [J]$ if and only if $I = \alpha J$ for some invertible $\alpha \in J^{-1}$, and $\mu J = J$ if and only if $\mu \in \mathcal{O}_{\mathrm{L}}(J)^\times$, we have

$$\zeta_{\mathcal{O},[J]}(s) = \frac{1}{N(J)^s} \sum_{0\neq\alpha\in J^{-1}/\mathcal{O}_{\mathrm{L}}(J)^\times} \frac{1}{N(\alpha)^s}.$$

In order to proceed, we assume that B is definite (ramified at ∞) or equivalently that $\#\mathcal{O}_{\mathrm{L}}(J)^\times < \infty$. Let $w_J = \#\mathcal{O}_{\mathrm{L}}(J)/\{\pm 1\}$. We again argue by counting lattice points to prove the following proposition.

Proposition 18.3.4. *The function $\zeta_{\mathcal{O},[J]}(s)$ has a simple pole at $s = 1$ with residue*

$$\operatorname{res}_{s=1} \zeta_{\mathcal{O},[J]}(s) = \frac{\pi^2}{w_J D}.$$

Rather than give a proof of this proposition (which will be proven in this chapter in greater generality), we only give a sketch of the idea. Using a *Tauberian theorem*, we will show that

$$\operatorname{res}_{s=1} \zeta_{\mathcal{O},[J]}(s) = \frac{1}{2w_J \operatorname{N}(J)} \frac{\operatorname{vol}(\mathbb{R}^4_{\leq 1})}{\operatorname{covol}(J)}$$

where

$$\operatorname{vol}(\mathbb{R}^4_{\leq 1}) = \operatorname{vol}(\{x \in \mathbb{R}^4 : |x| \leq 1\}) = 2\pi^2$$

and $J \hookrightarrow J \otimes_{\mathbb{R}} \cong \mathbb{H} \cong \mathbb{R}^4$ has the structure of a lattice in $\mathbb{R}^4$ with $\operatorname{covol}(J) = \operatorname{covol}(\mathcal{O})/\operatorname{N}(J) = D/\operatorname{N}(J)$. Putting all of these facts together, we have

$$\operatorname{res}_{s=1} \zeta_{\mathcal{O},[J]}(s) = \frac{2\pi^2}{2w_J \operatorname{N}(J)} \frac{\operatorname{N}(J)}{D} = \frac{\pi^2}{w_J D}.$$

Combining Proposition 18.3.4 with (18.3.3), we have

$$\operatorname{res}_{s=1} \zeta_{\mathcal{O}}(s) = \frac{\pi^2}{D} \sum_{[J]\in\operatorname{Cl}(\mathcal{O})} \frac{1}{w_J} = \frac{\pi^2}{12} \prod_{p\mid D} \left(1 - \frac{1}{p}\right)$$

and we conclude the following theorem.

Theorem 18.3.5 (Eichler mass formula)**.** *Let $\mathcal{O}$ be a maximal order in a definite quaternion algebra over $\mathbb{Q}$ of discriminant D. Then*

$$\sum_{[J]\in \mathrm{Cl}(\mathcal{O})} \frac{1}{w_J} = \frac{\phi(D)}{12}$$

where $w_J = \#\mathcal{O}_{\mathrm{L}}(J)/\{\pm 1\}$ and ϕ is the Euler ϕ-function.

The Eichler mass formula does not quite give us a formula for the class number; rather, it gives us a formula for a "weighted" class number (we can think of w_J as the order of the nonscalar automorphism group of J), and this is still useful for applications.

Corollary 18.3.6. *We have $\#\,\mathrm{Cl}(\mathcal{O}) = 1$ if and only if $D = 2, 3, 5, 7, 13$.*

Proof. We have $\#\,\mathrm{Cl}(\mathcal{O}) = 1$ if and only if

$$\frac{1}{w} = \frac{\phi(D)}{12}$$

where $w = \#\mathcal{O}/\{\pm 1\}$. For the cases $D = 2, 3$, we have seen already that the maximal order $\mathcal{O}$ is fact Euclidean and $w = 12, 6$, respectively; this independently verifies the Eichler mass formula in these cases. We will see in Section 21.1 that if $D > 3$ then $w \leq 3$. By elementary arguments with the Euler ϕ function (Exercise 18.2), we have $\phi(D)/12 = 1/w$ with $w \leq 3$ only if $D = 5, 7, 13, 42$, and we can check each discriminant in turn. For $D = 5$, we may take $B = \left(\dfrac{-5, -3}{\mathbb{Q}}\right)$ and a maximal order $\mathcal{O}$ containing $\mathbb{Z}[(-1 + j)/2]$, so $w = 3$ and $\#\,\mathrm{Cl}(\mathcal{O}) = 1$. The other cases follow similarly; the details are requested in Exercise 18.3. □

In Section 18.6, we go farther and find *all* definite quaternion orders of class number 1, following Brzezinski.

18.4 Analytic class number formula

Let F be a number field with r real places and c complex places, so that $[F : \mathbb{Q}] = n = r + 2c$. Let R be the ring of integers in F. Define the Dedekind zeta function for $s \in \mathbb{C}$ with $\Re(s) > 1$ by

$$\zeta_F(s) = \sum_{\mathfrak{a}\subseteq R} \frac{1}{\mathrm{N}\mathfrak{a}^s} = \prod_{\mathfrak{p}} \left(1 - \frac{1}{\mathrm{N}\mathfrak{p}^s}\right)^{-1}$$

where the sum is over all nonzero ideals of R, the product is over all primes of R, and $\mathrm{N}\mathfrak{a}$ is the absolute norm. Then ζ_F admits an analytic continuation to all of $\mathbb{C}$ with a simple pole at $s = 1$ with residue

$$\zeta_F^*(1) = \lim_{s\to 1}(s-1)\zeta_F(s) = \frac{2^{r_1}(2\pi)^{r_2}}{w\sqrt{d_F}} h_F \mathrm{Reg}_F \tag{18.4.1}$$

where w_F is the number of roots of unity in F, $h_F = \#\mathrm{Cl}\,R$ is the class number of F, Reg_F is the regulator of F, and d_F is the absolute discriminant of F

More generally, for $a \in \mathbb{C}$ we write $\zeta_F^*(a)$ for the leading coefficient in the Laurent series expansion for ζ_F at $s = a$.

The formula (18.4.1) is known as Dirichlet's *analytic class number formula* (even though Dirichlet's theorem concerned quadratic forms rather than classes of ideals). Using the functional equation for ζ_F, we can also write this more simply as

$$\zeta_F^*(0) = \lim_{s\to 0} s^{-(r_1+r_2-1)}\zeta_F(0) = \frac{h_F \mathrm{Reg}_F}{w_F};$$

here, ζ_F has a zero at $s = 0$ of order $r_1 + r_2 - 1$.

In the situation where F is an imaginary quadratic field ($r_1 = 0$ and $r_2 = 1$), we have $\mathrm{Reg}_F = 1$, so we find that

$$h_F = w_F \zeta_F(0)$$

and in particular if $d_F > 4$ then $h_F = 2\zeta_F(0)$.

In brief, this formula is proved as follows. We write the Dedekind zeta function as a sum over ideals in a given ideal class: we define the partial zeta function

$$\zeta_{F,[\mathfrak{b}]}(s) = \sum_{\substack{\mathfrak{a}\subseteq R \\ [\mathfrak{a}]=[\mathfrak{b}]}} \frac{1}{\mathrm{N}\mathfrak{a}^s}$$

so that

$$\zeta_F(s) = \sum_{[\mathfrak{b}]\in \mathrm{Cl}\,R} \zeta_{F,[\mathfrak{b}]}(s).$$

Now note that $[\mathfrak{a}] = [\mathfrak{b}]$ if and only if $\mathfrak{a} = a\mathfrak{b}$ for some nonzero

$$a \in \mathfrak{b}^{-1} = \{x \in F : x\mathfrak{b} \subseteq R\},$$

so we have a bijection between nonzero ideals $\mathfrak{a} \subseteq R$ such that $[\mathfrak{a}] = [\mathfrak{b}]$ and the set of nonzero elements in $\mathfrak{b}^{-1}/R^\times$. So

$$\zeta_{F,[\mathfrak{b}]}(s) = \frac{1}{\mathrm{N}\mathfrak{b}^s} \sum_{0\neq a\in \mathfrak{b}^{-1}/R^\times} \frac{1}{|\mathrm{N}a|^s}.$$

One now reduces to a problem concerning lattice points after choosing a fundamental domain for the action of $R^\times$, which reduces to a volume computation (an exercise in multivariable integration with polar coordinates): one finds that

$$\zeta^*_{F,[\mathfrak{b}]}(s) = \frac{2^{r_1}(2\pi)^{r_2}}{w\sqrt{d_F}}\operatorname{Reg}_F$$

(independent of $[\mathfrak{b}]$!) and Dirichlet's formula follows.

18.5 Zeta functions of quaternion algebras

We now mimic the above proof in our quaternionic setting.

Recall from strong approximation that if F is a function field, or if F is a number field and B satisfies the Eichler condition—i.e., that there is an unramified archimedean place for B—then one may identify the class number of any order $\mathcal{O} \subseteq B$ with a suitable class group.

So only one case remains to consider. Let F be a totally real number field and let B be a quaternion algebra which is ramified at all real places of F; we say that B is a *(totally) definite* (and otherwise we say B is *indefinite*).

Remark 18.5.1. To extend the analogy, we also note that this situation is analogous to one without a regulator term (as would be necessary, as otherwise the unit groups would be infinite and noncommutative!).

Let $\mathcal{O}$ be an order in B.

Let I be an integral right $\mathcal{O}$-ideal, so that $I \subseteq \mathcal{O}$. We define $N(I) = \#(\mathcal{O}/I)$. We have $N(I) = N_{F/\mathbb{Q}}\operatorname{nrd}(I)^2$. For example, if $\mathfrak{a} \subseteq R$ is a nonzero ideal then $N(\mathfrak{a}\mathcal{O}) = N_{F/\mathbb{Q}}\mathfrak{a}^4$. Note this agrees with the covolume computation in Lemma 15.4.3.

We then define the *zeta function* of $\mathcal{O}$ to be

$$\zeta_{\mathcal{O}}(s) = \sum_{I \subseteq \mathcal{O}} \frac{1}{N(I)^s} = \sum_{\mathfrak{n}} \frac{a_{\mathfrak{n}}}{N(\mathfrak{n})^{2s}} \tag{18.5.2}$$

where the first sum is over all (nonzero) integral right $\mathcal{O}$-ideals I and in the second sum $a_{\mathfrak{n}}$ is the number of right $\mathcal{O}$-ideals I with $\operatorname{nrd}(I) = \mathfrak{n}$.

Our first order of business is to establish an Euler product for $\zeta_{\mathcal{O}}(s)$.

Lemma 18.5.3. *Let I be an integral $\mathcal{O}$-ideal with* $\operatorname{nrd}(I) = \mathfrak{a}\mathfrak{b}$ *and* $\mathfrak{a}$ *and* $\mathfrak{b}$ *coprime. Then I can be uniquely written as a compatible product $I = I(\mathfrak{a})I(\mathfrak{b})$ with* $\operatorname{nrd}(I(\mathfrak{a})) = \mathfrak{a}$ *and* $\operatorname{nrd}(I(\mathfrak{b})) = \mathfrak{b}$.

Proof. Let $\mathcal{O}' = \mathcal{O}_{\mathrm{L}}(I)$. Define $I(\mathfrak{a}) = I + \mathfrak{a}\mathcal{O}'$ and $I(\mathfrak{b}) = I + \mathfrak{b}\mathcal{O}$. The statement holds if and only if it holds locally, so let $\mathfrak{p}$ be a prime of R. If $\mathfrak{p} \nmid \mathfrak{a}\mathfrak{b}$ then $I_{\mathfrak{p}} = \mathcal{O}_{\mathfrak{p}} =$

$I(\mathfrak{a})_\mathfrak{p} = I(\mathfrak{b})_\mathfrak{p}$. If $\mathfrak{p} \mid \mathfrak{a}$, then $I(\mathfrak{a}) = I_\mathfrak{p}$ and $I(\mathfrak{b})_\mathfrak{p} = \mathcal{O}_\mathfrak{p}$ and so $\operatorname{nrd}(I(\mathfrak{a})_\mathfrak{p}) = \operatorname{nrd}(I_\mathfrak{p})$ and $(I(\mathfrak{a})I(\mathfrak{b}))_\mathfrak{p} = I_\mathfrak{p}\mathcal{O}_\mathfrak{p} = I_\mathfrak{p}$ and the product is compatible. We have a similar statement for $\mathfrak{p} \mid \mathfrak{b}$. The result follows. □

If we factor an $\mathcal{O}$-ideal into a product of ideals whose norms are powers of a prime, then these ideals will have orders which will be in general different than $\mathcal{O}$. Of course, if $\mathcal{O}$ and $\mathcal{O}'$ are connected (locally isomorphic) orders, then since ideals can be recovered from its localizations there is a bijection between the right ideals of $\mathcal{O}$ and of $\mathcal{O}'$ of any reduced norm $\mathfrak{a}$.

Suppose that $\mathcal{O}$ is hereditary. Then when we factor I into a product of ideals of prime power norm, each left and right order that appears in the product is connected to $\mathcal{O}$, since each ideal is locally principal. It follows that if $a_\mathfrak{n}$ is the number of right $\mathcal{O}$-ideals of norm $a_\mathfrak{n}$, then $a_{\mathfrak{mn}} = a_\mathfrak{m} a_\mathfrak{n}$ whenever $\mathfrak{m}$ and $\mathfrak{n}$ are relatively prime. Consequently

$$\zeta_\mathcal{O}(s) = \prod_\mathfrak{p} \zeta_{\mathcal{O}_\mathfrak{p}}(s)$$

where

$$\zeta_{\mathcal{O}_\mathfrak{p}}(s) = \sum_{I_\mathfrak{p} \subseteq \mathcal{O}_\mathfrak{p}} \frac{1}{N(I_\mathfrak{p})^s} = \sum_{e=0}^{\infty} \frac{a_{\mathfrak{p}^e}}{N\mathfrak{p}^{2s}}.$$

We are now reduced to counting ideals locally.

Lemma 18.5.4. *Let $\mathcal{O}_\mathfrak{p}$ be a maximal order in $B_\mathfrak{p}$ and let $\mathfrak{q} = \mathfrak{p}^e$ be a prime power of R.*

If $\mathfrak{p}$ is ramified in B, then there is a unique (left or) right integral $\mathcal{O}$-ideal of norm $\mathfrak{q}$.

If $\mathfrak{p}$ is unramified in B, then the number of (left or) right integral $\mathcal{O}$-ideals of norm $\mathfrak{p}$ is equal to $1 + N\mathfrak{p} + \cdots + N\mathfrak{p}^e$.

Proof. If $\mathfrak{p}$ is ramified then the unique maximal order $\mathcal{O}_\mathfrak{p}$ has a unique (two-sided) maximal ideal P with $\operatorname{nrd}(P) = \mathfrak{p}$ and all ideals of $\mathcal{O}_\mathfrak{p}$ are powers of P.

So suppose that $\mathfrak{p}$ is split in B. Let $I_\mathfrak{p} = x_\mathfrak{p}\mathcal{O}_\mathfrak{p}$ be a right integral $\mathcal{O}_\mathfrak{p}$-ideal of norm $\mathfrak{p}^e$ and let π be a uniformizer for $\mathfrak{p}$. Then we can write

$$x_\mathfrak{p} = \begin{pmatrix} \pi^{e-f} & 0 \\ r & \pi^f \end{pmatrix}$$

for unique $f \in \mathbb{Z}_{\geq 0}$ and $r \in R/\mathfrak{p}^f$ (Exercise 19.5. Since $\#R/\mathfrak{p}^f = N\mathfrak{p}^f$, it follows that the number of such ideals is equal to $1 + N\mathfrak{p} + \cdots + N\mathfrak{p}^e$. □

From this lemma we have

$$\zeta_{\mathcal{O}_\mathfrak{p}}(s) = \left(1 - \frac{1}{N\mathfrak{p}^{2s}}\right)^{-1}$$

if $\mathfrak{p}$ is ramified, and

$$\zeta_{\mathcal{O}_\mathfrak{p}}(s) = \sum_{e=0}^{\infty} \frac{1 + N\mathfrak{p} + \cdots + N\mathfrak{p}^e}{N\mathfrak{p}^{2es}} = \left(1 - \frac{1}{N\mathfrak{p}^{2s}}\right)^{-1}\left(1 - \frac{1}{N\mathfrak{p}^{2s-1}}\right)^{-1}$$

if $\mathfrak{p}$ is split.

Recall that

$$\zeta_F(s) = \prod_{\mathfrak{p}} \left(1 - \frac{1}{N\mathfrak{p}^s}\right)^{-1}.$$

Thus

$$\zeta_{\mathcal{O}}(s) = \prod_{\mathfrak{p}} \zeta_{\mathcal{O}_\mathfrak{p}}(s) = \zeta_F(2s)\zeta_F(2s-1) \prod_{\mathfrak{p} \mid \mathfrak{D}} \left(1 - \frac{1}{N\mathfrak{p}^{2s-1}}\right)$$

where $\mathfrak{D}$ is the discriminant of B. (Here one needs to break up the sum into ideals of norm $\leq X$ and those $> X$ and show that the product converges, etc.)

Since the ζ_F has only a simple pole at $s = 1$, there is a single simple pole of $\zeta_{\mathcal{O}}$ at $s = 1$, and we conclude that

$$\zeta_{\mathcal{O}}^*(1) = \lim_{s\to 1}(s-1)\zeta_{\mathcal{O}}(s) = \frac{\zeta_F(2)}{2}\zeta_F^*(1)\prod_{\mathfrak{p}\mid\mathfrak{D}}(1 - N\mathfrak{p}^{-1}). \tag{18.5.5}$$

We now write $\zeta_{\mathcal{O}}(s)$ as a sum over right ideal classes and analyze the residue at $s = 1$ by a volume a computation as with the analytic class number formula. Combining this with (18.5.5), we will obtain a *mass formula* for the order $\mathcal{O}$.

For an integral right $\mathcal{O}$-ideal J, let

$$\zeta_{\mathcal{O},[J]}(s) = \sum_{\substack{I\subseteq\mathcal{O}\\ [I]=[J]}} \frac{1}{N(I)^s}.$$

Then obviously

$$\zeta_{\mathcal{O}}(s) = \sum_{[J]\in\operatorname{Cl}\mathcal{O}} \zeta_{\mathcal{O},[J]}(s).$$

We have $[I] = [J]$ if and only if $I \cong J$ if and only if $I = xJ$ for nonzero $x \in J^{-1}$ (recall that B is a division ring since it is ramified at its real places). Since $uJ = J$ if and only if $u \in \mathcal{O}_{\mathrm{L}}(J)^\times$ (Exercise 19.7), it follows that

$$\zeta_{\mathcal{O},[J]}(s) = \frac{1}{N(J)^s} \sum_{0\neq x\in J^{-1}/\mathcal{O}_{\mathrm{L}}(J)^\times} \frac{1}{N(x)^s}.$$

But now $\mathcal{O}_{\mathrm{L}}(J)^\times/R^\times$ is finite since B is totally definite. Let $w(J) = \#\mathcal{O}_{\mathrm{L}}(J)^\times/R^\times$. Then

$$\zeta_{\mathcal{O},[J]}(s) = \frac{1}{w(J)N(J)^s} \sum_{0\neq x\in J^{-1}/R^\times} \frac{1}{N(x)^s}.$$

Proposition 18.5.6. *$\zeta_{\mathcal{O},[J]}(s)$ has a simple pole at $s = 1$ with residue*

$$\zeta^*_{\mathcal{O},[J]}(1) = \frac{2^n(2\pi)^{2n}\mathrm{Reg}_F}{8w(J)d_F^2 N\mathfrak{D}}.$$

With this lemma in hand, we find that

$$\zeta^*_{\mathcal{O}}(1) = \frac{2^n(2\pi)^{2n}\mathrm{Reg}_F}{8d_F^2 N\mathfrak{D}} \sum_{[J]\in\mathrm{Cl}\,\mathcal{O}} \frac{1}{w(J)} = \frac{\zeta_F(2)}{2}\left(\frac{2^n}{2\sqrt{d_F}}h_F\mathrm{Reg}_F\right)\prod_{\mathfrak{p}|\mathfrak{D}}\left(1-\frac{1}{N\mathfrak{p}}\right)$$

since $w_F = 2$ (F is totally real) so since $\mathfrak{D}$ is squarefree we have

$$\sum_{[J]\in\mathrm{Cl}\,\mathcal{O}} \frac{1}{w(J)} = \zeta_F(2)\frac{2}{(2\pi)^{2n}}h_F d_F^{3/2}\prod_{\mathfrak{p}|\mathfrak{D}}(N\mathfrak{p}-1).$$

By the functional equation for ζ_F, we have

$$\zeta_F(-1) = \left(\frac{-1}{2\pi^2}\right)^n d_F^{3/2}\zeta_F(2)$$

so we can also write

$$\sum_{[J]\in\mathrm{Cl}\,\mathcal{O}} \frac{1}{w(J)} = 2^{1-n}|\zeta_F(-1)|h_F\Phi(\mathfrak{D})$$

where we define

$$\Phi(\mathfrak{D}) = \prod_{\mathfrak{p}|\mathfrak{D}}(N\mathfrak{p}-1),$$

the generalization of Euler's Φ-function. We have proven the following theorem.

Theorem 18.5.7 (Eichler's mass formula)**.** *We have*

$$\sum_{[J]\in\mathrm{Cl}\,\mathcal{O}} \frac{1}{[\mathcal{O}_{\mathrm{L}}(J)^\times : R^\times]} = 2^{1-n}|\zeta_F(-1)|h_F\Phi(\mathfrak{D}).$$

Corollary 18.5.8. *If $\mathcal{O}$ is the maximal order in a definite quaternion algebra of discriminant D over $\mathbb{Q}$, then*

$$\sum_{[J]\in\mathrm{Cl}\,\mathcal{O}} \frac{1}{\#\mathcal{O}_{\mathrm{L}}(J)^\times} = \frac{1}{24}\phi(D).$$

In the (unlikely) situation where $\mathcal{O}_{\mathrm{L}}(J)^\times = u$ is independent of J, we would conclude that

$$\#\operatorname{Cl}\mathcal{O} = H = u2^{1-n}|\zeta_F(-1)|h_F\Phi(\mathfrak{D}).$$

[[As an approximation, H is roughly $d_F^{3/2}N\mathfrak{D}/2^n$.]]

Proof of Proposition 18.5.6. We first prove this statement in the case $F = \mathbb{Q}$, since then we can see the main ideas minimizing technicalities. We will come back and treat the general case. [[Sucker!]]

In this situation, our formula greatly simplifies and we need to show that $\zeta_{\mathcal{O},[J]}(s)$ has a simple pole at $s = 1$ with residue $\pi^2/(w(J)D)$ where D is the discriminant of B. We have written

$$\zeta_{\mathcal{O},[J]}(s) = \frac{1}{w(J)N(J)^s} \sum_{0 \neq x \in J^{-1}/\{\pm 1\}} \frac{1}{N(x)^s}.$$

The key technical fact is the following. Let $X \subseteq \mathbb{R}^n$ be a cone and suppose $0 \notin N(X)$. Suppose that $X_{\leq 1} = X \cap \{x \in \mathbb{R}^n : N(x) \leq 1\}$ is bounded and has volume $\operatorname{vol}(X_{\leq 1})$. Let $L \subseteq \mathbb{R}^n$ be a lattice of full rank in $\mathbb{R}^n$ and has covolume $\operatorname{covol}(L)$. Let

$$\zeta_{L,X}(s) = \sum_{x \in X \cap L} \frac{1}{N(x)^s}.$$

Then $\zeta_{L,X}(s)$ converges for $\operatorname{Re} s > 1$ and has a simple pole at $s = 1$ with residue

$$\zeta_{L,X}^*(1) = \lim_{s \to 1^+} (s-1)\zeta_{L,X}(s) = \frac{\operatorname{vol}(X_{\leq 1})}{\operatorname{covol}(L)}.$$

Whew! In our case, we consider the lattice

$$J^{-1} \hookrightarrow B \otimes_{\mathbb{Q}} \mathbb{R} \cong \mathbb{H} \cong \mathbb{R}^4.$$

The choice of sign allows us to choose the first coordinate to be positive, so we take $X = \mathbb{R}_{\geq 0} \times \mathbb{R}^3$. We have $\operatorname{vol}(X) = 1/2(2\pi^2) = \pi^2$. From the previous section on Minkowski theory, we have $\operatorname{covol}(J^{-1}) = \operatorname{covol}(\mathcal{O})/N(J)$. And almost by definition we have $\operatorname{covol}(\mathcal{O}) = \sqrt{\operatorname{disc}\mathcal{O}} = D$. Putting this together with the technical fact, we have

$$\zeta_{\mathcal{O},[J]}^*(1) = frac1w(J)N(J)\frac{\pi^2 N(J)}{D} = \frac{\pi^2}{w(J)D}$$

as claimed. □

18.6 Class number 1 problem

The list of all definite quaternion orders of class number 1 over the integers was determined by Brzezinski [Brz95] in the language of ternary quadratic forms.

(

Theorem 18.6.1 (Brzezinski)**.** *There are* 25 *isomorphism classes of definite quaternion orders over* $\mathbb{Z}$ *with class number* 1.

Proof. Mass formula for maximal orders. □

Also work with Markus.

```
It turns out
that there are no non-principal Euclidean ideal classes for maximal
Z-orders in definite rational quaternion algebras.

But the way things worked out was somewhat interesting: the class
number 2 (reduced) discriminants are 11, 17, 19, 30, 42, 70, 78.

For 11, 17, 19, it turns out that the type number is also 2.  This
means that the Brandt groupoid has the following structure:
there are the two nonsiomorphic maximal orders, say O1 and O2.  The
right Picard set [I think you might call this the "left Picard set in
your notes", and you might well be right; anyway you get the point] of
O1 consists of O1 together with an ideal I with left order O1 and
right order O2.  Similarly, the right Picard set of O2 consists of O2
together with an ideal I' with left order O2 and right order O1.  Thus
in the Brandt groupoid I' is the inverse of I, whereas O1 and O2 are
self-inverse.

[So the picture is a graph with two vertices and four oriented edges,
two of which are loops and a pair of mutually inverse edges running
between the two vertices.]

These four "arrows" in the groupoid represent three different
quadratic forms: an ideal and its inverse give rise to the same
quadratic form.  In fact these three quadratic forms comprise a full
genus (I checked that independently on the quadratic lattice side).
Before I was looking only at one of the maximal orders and was very
confused as to why MAGMA was telling me that the class number of the
associated quadratic form was equal to 3.  I think I get it now!

Note that this means that the two-sided Picard group of each of O1 and
O2 is trivial (in particular they are isomorphic!),
```

and this means that if there is any Euclidean ideal class, then there would be a Euclidean order, forcing the class number to be 1 rather than 2. (Of course I checked this by computing the Euclidean minima of the 3 forms in question: they are all bigger than 1, although not tremendously so.)

For 30,42,70,78, it turns out that the type number is 1. So in this case the Brandt groupoid is simply equal to the two-sided Picard group, which has cardinality 2. In other words there is a unique nonprincipal two-sided ideal, which is a candidate for being a Euclidean ideal class...but it turns out not to be. (On the other hand, I think I could invert a suitable small set of primes in Z in such a way as to keep the class numbers, type numbers, etc. the same but make the nonprincipal ideal class be Euclidean. I should probably do that: it would be nice to have an actual example.)

You might want to use this example in your book. (Or you might not: just a late night idea.)

Let me say that although obviously the first three discriminants have one prime factor and the last four have three prime factors, in and of itself this doesn't explain to me why the two Picard groups have different structures. I happen to know that the two-sided Picard group is an elementary abelian two-group whose F_2-dimension is *at most* the number of ramified primes, but that doesn't explain everything here: certainly there are examples of quaternion algebras of prime discriminant with nontrivial two-sided Picard group, e.g. D = 73, 89. (If there is some clear answer as to when the Picard group is trivial and when it has order 2 in the prime discriminant case, I am not aware of it.)

Best regards,

Pete

18.7 Zeta functions over function fields

18.8 Extensions and further reading

18.8.1. Further reading for Gauss class number 1 problem.

18.8.2. Eichler mass formula and supersingular elliptic curves.

18.8.3. The primes $p = 2, 3, 5, 7, 13$ in Corollary 18.3.6 are also the primes p such that the modular curve $X_0(p)$ has genus 0. This is not a coincidence, and reflects a deep correspondence between classical and quaternionic modular forms.

18.8.4. Prove the theorem on primes in arithmetic progression (Theorem 11.2.8) using the nonvanishing $L(1,\chi) \neq 0$ for characters χ, as follows.

18.8.5. More generally, we can replace the orthogonal group with another compact, semisimple Lie group G_∞, an integral model for it $G_{\mathbb{Z}}$, and ask for the weighted sum of the class set of objects with automorphism group $G_{\mathbb{Z}}$. (For example, unitary group and Hermitian lattices, symplectic group, G_2, etc.) Weil conjectured a formula for this in terms of a product, and Langlands, Lai, and Kottwitz proved general statements for number fields. The right way to formulate this is in terms of double cosets, where the final answer comes down to a Tamagawa number calculation.

The function field equivalent is a problem in topology.

Exercises

18.1. Prove Lemma 18.2.3 as follows.

a) Let P be a fundamental parallelogram for Λ, and for $\lambda \in \Lambda$ let $P_\lambda = P+\lambda$. For $x > 1$, let $D(x) = \{z \in \mathbb{C} : |z| \leq x\}$, and

$$\begin{aligned} N(x) &= \#\{\lambda \in \Lambda : \lambda \in D(x)\} \\ N_P(x) &= \#\{\lambda \in \Lambda : P_\lambda \subseteq D(x)\} \\ N_P^+(x) &= \#\{\lambda \in \Lambda : P_\lambda \cap D(x) \neq \emptyset\}. \end{aligned}$$

Show that
$$N(x) \leq N_P(x) \leq N_P^+(x).$$

b) Show that $N_P(x) \leq \pi x^2/A \leq N_P^+(x)$.

c) Let d be the length of a long diagonal in P. Show that for any $\lambda \in \Lambda \cap D(x)$, we have $P_\lambda \subseteq D(x+d)$, so

$$N(x) \leq N_P(x+\delta) \leq \frac{\pi(x+d)^2}{A}.$$

Similarly, show that if $P_\lambda \cap D(x-d)$ then $P_\lambda \subseteq D(x)$ and $\lambda \in D(x)$, so

$$\frac{\pi(x-d)^2}{A} \leq N_P^+(x-d) \leq N(x).$$

d) Conclude that Lemma 18.2.3 holds with $C = \pi/A(2x + x^2)$. See [Wes, Lemma 1.19].

18.2. Show that if D is a squarefree integer with an odd number of prime factors with $\phi(D)/12 \in \{1, 1/2, 1/3\}$, then $D \in \{5, 7, 13, 42\}$.

18.3. Using the explicit description of maximal orders, show that maximal orders of discriminant $D = 5, 7, 13$ have class number 1 and a maximal order of discriminant 42 has class number > 1.

Chapter 19

Adelic framework

19.1 Adeles and ideles

We have already seen that the local-global dictionary is a powerful tool in understanding the arithmetic of quaternion algebras. In this section, we formalize this connection by consideration of ideles.

Throughout this section, let F be a global field. For a place v of F, we denote by F_v the completion of F at the place v, with preferred (normalized) associated absolute value $|\,|_v$ so that the product formula holds in F. If v is nonarchimedean, we let $R_v = \{x \in F_v : |x|_v \leq 1\}$ be the valuation ring of F_v, where by abuse of notation we write v also for the discrete valuation associated to the place v. If v is archimedean, we write $v \mid \infty$.

19.2 Adeles

The *adele ring* of F is

$$\mathbf{A}_F = \prod_v{}' F_v = \left\{(x_v)_v \in \textstyle\prod_v F_v : |x_v|_v \leq 1 \text{ for all but finitely many } v\right\}.$$

The product $\mathbf{A}_F$ is given a topology as follows: $U \subseteq \mathbf{A}_F$ is open if and only if for all $a \in \mathbf{A}_F$, one has that the set

$$(a + U) \cap \left(\textstyle\prod_{v\mid\infty} F_v \times \prod_{v\nmid\infty} R_v\right)$$

is open in the product topology.

We embed $F \subseteq \mathbf{A}_F$ by $x \mapsto (x)_v$; this map is well-defined because $|x|_v > 1$ for only finitely many places v of F. The image of F in $\mathbf{A}_F$ has the discrete topology and hence it is closed in $\mathbf{A}_F$; the quotient $\mathbf{A}_F/F$ is a connected, compact Hausdorff topological group; we say that F is *cocompact* in $\mathbf{A}_F$. (In some sense, this is like how $\mathbb{Z}$ sits in $\mathbb{R}$: the quotient is the compact circle group $\mathbb{R}/\mathbb{Z}$.)

Let S be a finite set of places of F containing the infinite places. We will often write

$$\mathbf{A}_F = \widehat{F}_S \times \prod_{v \in S} F_v,$$

where $\widehat{F} = \prod\nolimits'_{v \notin S} F_v$; we call $\widehat{F}$ the *S-finite adele ring* of F.

19.2.1. We define

$$\begin{aligned}\widehat{\mathbb{Z}} &= \varprojlim_n \mathbb{Z}/n\mathbb{Z} \\ &= \left\{(a_n)_{n=1}^{\infty} \in \textstyle\prod_{n=1}^{\infty}(\mathbb{Z}/n\mathbb{Z}) : \text{for all } n \mid m, a_m \equiv a_n \pmod{n}\right\}.\end{aligned}$$

We give each $\mathbb{Z}/n\mathbb{Z}$ the discrete topology, and $\prod_n(\mathbb{Z}/n\mathbb{Z})$ the product topology. This product is compact, as a result of the theorem of Tychonoff (the product of compact topological spaces is itself compact); the restriction $\widehat{\mathbb{Z}}$ is therefore itself compact, as $\widehat{\mathbb{Z}}$ is closed in $\prod_n(\mathbb{Z}/n\mathbb{Z})$. The ring homomorphism $\mathbb{Z} \to \prod_n(\mathbb{Z}/n\mathbb{Z})$ which takes every element to its reduction modulo n realizes $\widehat{\mathbb{Z}}$ as the closure of $\mathbb{Z}$ in the product $\prod_n(\mathbb{Z}/n\mathbb{Z})$.

By unique factorization, we obtain a ring isomorphism $\widehat{\mathbb{Z}} \xrightarrow{\sim} \prod_p \mathbb{Z}_p$. This notation is justified, since we have $\widehat{\mathbb{Q}} = \prod\nolimits'_p \mathbb{Q}_p \cong \mathbb{Q} \otimes_{\mathbb{Z}} \widehat{\mathbb{Z}}$. More generally, we have

$$\widehat{F} \cong F \otimes_{\mathbb{Q}} \widehat{\mathbb{Q}} = F \otimes_{\mathbb{Z}} \widehat{\mathbb{Z}}.$$

19.3 Ideles

We pass now to the multiplicative situation. The *idele group* of F is

$$\mathbf{J}_F = \left\{(x_\mathfrak{p})_\mathfrak{p} \in \textstyle\prod_\mathfrak{p} F_v^\times : |x_v|_v = 1 \text{ for all but finitely many } v\right\}.$$

We have $\mathbf{J}_F = \widehat{F}^\times \times \prod_{v \mid \infty} F_v^\times$.

In the relative topology of $\mathbf{J}_F$ in $\mathbf{A}_F$, inversion is not a continuous operation! To get the correct topology, we declare instead that $U \subseteq \mathbf{J}_F$ is open if and only if for all $a \in \mathbf{J}_F$, the set

$$aU \cap \left(\textstyle\prod_{v \mid \infty} F_v^\times \times \prod_{v \nmid \infty} R_v^\times\right)$$

is open in the product topology.

Remark 19.3.1. In general, if A is a topological ring, $A^\times$ becomes a topological group when $A^\times$ is given the relative topology from

$$\begin{aligned} A^\times &\subseteq A \times A \\ x &\mapsto (x, x^{-1}). \end{aligned}$$

(See Exercise 19.15.)

Just as $F \subseteq \mathbf{A}_F$ is discrete, $F^\times \subseteq \mathbf{J}_F$ is also discrete, but this time it is not quite cocompact: the group $C_F = \mathbf{J}_F/F^\times$, called the *idele class group*, warrants further study.

As above, if S is a finite set of places containing the archimedean places, we define $\widehat{F}_S^\times = \prod'_{v \notin S} F_v^\times$ so that $\mathbf{J}_F = \widehat{F}_S^\times \times \prod_{v \in S} F_v^\times$.

Example 19.3.2. Take the example $K = \mathbb{Q}$. We have a canonical isomorphism

$$\mathbb{Q}_p^\times \cong \langle p \rangle \times \mathbb{Z}_p^\times$$

defined by taking the p-adic valuation. If we identify $\mathbb{Z} \cong \langle p \rangle$, then

$$\mathbf{J}_\mathbb{Q} = \mathbb{R}^\times \times \prod_p{}' \mathbb{Q}_p^\times \cong \{\pm 1\} \times \mathbb{R}_{>0} \times \prod_p \mathbb{Z}_p^\times \times \bigoplus_p \mathbb{Z}.$$

(A direct sum appears because an element of the restricted direct product is a p-adic unit for all but finitely many p.)

We project $\mathbf{J}_\mathbb{Q}$ onto the product of the first and last factor:

$$\mathbf{J}_\mathbb{Q} \to \{\pm 1\} \times \bigoplus_p \mathbb{Z} \to 0.$$

Looking at $\mathbb{Q}^\times \subseteq \mathbf{J}_\mathbb{Q}$, if we write $r = \epsilon \prod_p p^{n(p)}$, where $\epsilon \in \{\pm 1\}$ and $n(p) = \operatorname{ord}_p(r)$, then $r \mapsto (\epsilon, (n(p))_p)$ in the projection. Therefore $\mathbb{Q}^\times$ is canonically identified with $\{\pm 1\} \times \bigoplus_p \mathbb{Z}$ in $\mathbf{J}_\mathbb{Q}$. Putting these together, we see that

$$\mathbf{J}_\mathbb{Q} \cong \mathbb{Q}^\times \times \mathbb{R}_{>0} \times \prod_p \mathbb{Z}_p^\times.$$

By the logarithm map, we have an isomorphism $\mathbb{R}_{>0} \cong \mathbb{R}$ of topological groups, therefore

$$\mathbf{J}_\mathbb{Q} \cong \mathbb{Q}^\times \times \mathbb{R} \times \widehat{\mathbb{Z}}^\times$$

and $\mathbf{J}_\mathbb{Q}/\mathbb{Q}^\times \cong \mathbb{R} \times \widehat{\mathbb{Z}}^\times$.

In a similar way, we see that $\widehat{\mathbb{Q}}^\times/\mathbb{Q}_+^\times \cong \widehat{\mathbb{Z}}^\times$, where $\mathbb{Q}_+^\times = \{x \in \mathbb{Q} : x > 0\}$.

(We have used that $\mathbb{Z}$ has unique factorization and $\mathbb{Z}^\times = \{\pm 1\}$; for a general number field, we face problems associated with units and the class group of the field, and the exact sequence will not split!)

19.3.3. If F is a number field, we have seen that even for $F = \mathbb{Q}$ the idele class group is neither profinite or compact; but the noncompactness is only because of the presence of the term $\mathbb{R}$ (a volume term). We therefore map

$$\begin{aligned} \mathbf{J}_F &\to \mathbb{R}_{>0} \\ (x_v)_v &\mapsto \prod_v |x_v|_v. \end{aligned}$$

This map is clearly surjective, so we obtain an exact sequence

$$1 \to \mathbf{J}_F^0 \to \mathbf{J}_F \to \mathbb{R}_{>0} \to 1.$$

In fact, the kernel $\mathbf{J}_F^0$ is compact, a statement equivalent to the Dirichlet unit theorem and the finiteness of the class group. This exact sequence splits, and so $\mathbf{J}_F \cong \mathbb{R} \times \mathbf{J}_F^0$ as topological groups. Since $F^\times \subseteq \mathbf{J}_F^0$ (by the product formula), we have $C_F \cong \mathbb{R} \times C_F^0$.

19.3.4. [[Function field case]]

19.4 Idelic class field theory

Let $\overline{F}$ be a separable closure of F. The main theorem of idèlic class field theory is as follows. We have a bijection

$$\begin{aligned} \{K \subseteq \overline{F} : K \supset F \text{ finite abelian}\} &\leftrightarrow \{H \subseteq C_F : H \text{ open subgroup}\} \\ K &\mapsto N_{K/F} C_K. \end{aligned}$$

What are these open subgroups? Let $\mathfrak{f} = \prod_{\mathfrak{p}} \mathfrak{p}^{n(\mathfrak{p})}$ be a *cycle*, a formal product where $n(\mathfrak{p}) \geq 0$ for all $\mathfrak{p}$, $n(\mathfrak{p}) = 0$ for almost all $\mathfrak{p}$, and

$$n(\mathfrak{p}) = \begin{cases} 0 \text{ or } 1, & \mathfrak{p} \text{ real}, \\ 0, & \mathfrak{p} \text{ complex}. \end{cases}$$

Given such a cycle $\mathfrak{f}$, we have an open subgroup $W_{\mathfrak{f}} \subseteq \mathbf{J}_F$, where

$$W_{\mathfrak{f}} = \prod_{n(\mathfrak{p})=0}{}' F_{\mathfrak{p}}^\times \times \prod_{\substack{\mathfrak{p} \text{ real} \\ n(\mathfrak{p}) \neq 0}} F_{\mathfrak{p},>0}^\times \times \prod_{\substack{\mathfrak{p} < \infty \\ n(\mathfrak{p}) > 0}} (1 + \mathfrak{p}^{n(\mathfrak{p})}).$$

A subgroup of $\mathbf{J}_F$ is open if and only if it contains an $W_{\mathfrak{f}}$: one can read this off almost immediately from the definition of the topology. Note that for such an open subgroup, at every complex place one has the entire component and at every real place one has the component up to finite index. We may take the image $\overline{W}_{\mathfrak{f}} =$

$(W_{\mathfrak{f}}F^\times)/F^\times \subseteq C_F$, and we see that a subgroup of C_F is open if and only if it contains $\overline{W}_{\mathfrak{f}}$ for some $\mathfrak{f}$. Then under the above correspondence, we have the isomorphisms $\mathrm{Gal}(K/F) \cong C_F/H$ and $C_F/\overline{W}_{\mathfrak{f}} \cong \mathrm{Cl}_{\mathfrak{f}}$, where $\mathrm{Cl}_{\mathfrak{f}}$ is the ray class group of conductor $\mathfrak{f}$.

Combining the surjections $C_F \to \mathrm{Gal}(K/F)$, we obtain a surjective map

$$C_F \to \varprojlim_{L} \mathrm{Gal}(K/F) = G_F^{\mathrm{ab}}$$

where G_F^{ab} is the Galois group of the maximal abelian extension of F in $\overline{F}$. Let D_F be the connected component of 1 in C_F. Then D_F is a closed subgroup, and it is exactly the kernel of the above map, so $C_F/D_F \cong G_F^{\mathrm{ab}}$. In fact, the topological group D_F is isomorphic to

$$D_F \cong \mathbb{R} \times (\mathbb{R}/\mathbb{Z})^{r_2} \times \mathbb{S}^{r_1+r_2-1}$$

where r_1 is the number of real primes, r_2 the number of complex primes, and $\mathbb{S}$ is the solenoid defined in the exercises.

19.5 Adelic dictionary

Let S be a finite, nonempty set of places of F containing the archimedean places, and let $R = R_S$ denote the ring of S-integers. Then R is a Dedekind domain with field of fractions F.

To an invertible fractional ideal $\mathfrak{a} \subseteq F$ of R_S, we consider the product of its images in the completions

$$(\mathfrak{a}_\mathfrak{p})_{\mathfrak{p}\notin S} \subseteq \prod_{\mathfrak{p}\notin S}{}' F_\mathfrak{p} = \widehat{F}_S.$$

(The image indeed lies in the restricted direct product since $\mathfrak{a}_\mathfrak{p} = R_\mathfrak{p}$ for all but finitely many primes $\mathfrak{p}$.) But recall that a fractional ideal of R is invertible if and only if it is locally principal. Therefore we can write $(\mathfrak{a}_\mathfrak{p})_\mathfrak{p} = (x_\mathfrak{p}R_\mathfrak{p})_\mathfrak{p}$ for some $\widehat{x} = (x_\mathfrak{p})_\mathfrak{p} \in \widehat{F}^\times$, and $\widehat{\mathfrak{a}} = \mathfrak{a}\widehat{R} = \widehat{xR} \subseteq \widehat{F}$. Recall that we can recover $\mathfrak{a}$ from $\widehat{\mathfrak{a}}$ as $\mathfrak{a} = \widehat{\mathfrak{a}} \cap F$. We have shown therefore that the group of invertible fractional ideals is isomorphic to $\widehat{F}_S^\times/\widehat{R}_S^\times$.

The principal (invertible) fractional ideals correspond to the image of $F^\times$ in $\widehat{F}^\times$. Therefore we have a canonical isomorphism

$$\mathrm{Cl}\, R_S \xrightarrow{\sim} \widehat{F}_S^\times/\widehat{R}_S^\times F^\times.$$

More generally, one may restrict to a subgroup of principal fractional ideals. For example, if we restrict to the subgroup of principal fractional ideals which have a totally positive generator, then we obtain instead the *narrow* (or *strict*) *class group*

$$\operatorname{Cl} R_S \xrightarrow{\sim} \widehat{F}_S^\times / \widehat{R}_S^\times F_+^\times .$$

(This is the quotient of the usual strict class group by the nonarchimedean primes in S.) Indeed, the open subgroups of $\widehat{F}_S^\times / F_+^\times$ correspond to the open subgroups of $C_F/D_F = \mathbf{J}_F/F^\times D_F$ in which one takes the full component $F_\mathfrak{p}^\times$ for all nonarchimedean primes $\mathfrak{p} \in S$. Therefore, for every open subgroup $H \subseteq \widehat{F}_S^\times / F_+^\times$, there exists an abelian extension K of F with the property that

$$\operatorname{Gal}(K/F) \xrightarrow{\sim} \operatorname{Cl}_H R_S = \widehat{F}_S^\times / H F_+^\times .$$

Remark 19.5.1. [[Remark on the idele class group à la Schoof.]]

Now let B be a quaternion algebra over F. In this section, we extend the above notions to B. Let $\mathcal{O}$ be an R-order in B.

The *adele ring* of B is $\mathbf{A}_B = B \otimes_F \mathbf{A}_F = B \otimes_\mathbb{Q} \mathbf{A}_\mathbb{Q}$ and the *idele ring* of B is $\mathbf{J}_B = \mathbf{A}_B^\times$ with the topology as in Remark 19.3.1. [[The topology is induced from a choice of order...]]

Lemma 19.5.2. *The set of invertible right fractional $\mathcal{O}$-ideals are in bijection with $\widehat{B}^\times / \widehat{\mathcal{O}}^\times$.*

The set $\operatorname{Cl} \mathcal{O}$ *of isomorphism classes of right invertible fractional $\mathcal{O}$-ideals is in bijection with $B^\times \backslash \widehat{B}^\times / \widehat{\mathcal{O}}^\times$.*

Proof. Let I be an invertible right fractional $\mathcal{O}$-ideal. Then $I_\mathfrak{p} = x_\mathfrak{p} \mathcal{O}_\mathfrak{p}$ is principal for all primes $\mathfrak{p}$ of R, so to I we associate $(x_\mathfrak{p}\mathcal{O}_\mathfrak{p})_\mathfrak{p} = \widehat{x}\widehat{\mathcal{O}} \subseteq \widehat{B}$. The generator $\widehat{x}$ is well-defined up to $\mathcal{O}_\mathfrak{p}^\times$, so to I we obtain an element of $\widehat{B}^\times / \widehat{\mathcal{O}}^\times$. Conversely, given $\widehat{x} \in \widehat{B}^\times / \widehat{\mathcal{O}}^\times$ we recover I as $I = \widehat{x}\widehat{\mathcal{O}} \cap B$.

The principal (invertible) right fractional $\mathcal{O}$-ideals are the image of $B^\times$, and so the second statement follows. □

Remark 19.5.3. Of course on the left we have instead $\widehat{\mathcal{O}}^\times \backslash \widehat{B}^\times$, and the map $\widehat{x} \mapsto \overline{\widehat{x}}$ yields a bijection between these two sets.

19.5.4. In a similar way, we see that the group of invertible two-sided $\mathcal{O}$-ideals is in bijection with

$$\widehat{\mathcal{O}}^\times \backslash N(\widehat{\mathcal{O}}) / \widehat{\mathcal{O}}^\times = N(\widehat{\mathcal{O}}) / \widehat{\mathcal{O}}^\times = \widehat{\mathcal{O}}^\times \backslash N(\widehat{\mathcal{O}})$$

where

$$N(\widehat{\mathcal{O}}) = \{x \in \widehat{B}^\times : x\widehat{\mathcal{O}} = \widehat{\mathcal{O}}x\}$$

is the normalizer of $\widehat{\mathcal{O}}$.

The group of isomorphism classes of invertible two-sided $\mathcal{O}$-ideals is therefore in bijection with $N(\mathcal{O}) \backslash N(\widehat{\mathcal{O}}) / \widehat{\mathcal{O}}^\times$ where $N(\mathcal{O}) = \{x \in B^\times : x\mathcal{O} = \mathcal{O}x\}$.

[[$B^\times\backslash\widehat{B}^\times$ is compact and $\widehat{\mathcal{O}}^\times$ is open so the double coset is finite.]]

19.6 Norms

Let B be a quaternion algebra over F. Let $\mathcal{O}$ be an R-order.

In order to understand the set $\mathrm{Cl}\,\mathcal{O}$, we examine the effect of the reduced norm on the double coset $B^\times\backslash\widehat{B}^\times/\widehat{\mathcal{O}}^\times$. For this, we will need to determine the image on each component.

Note that if v is a place of F then $\mathrm{nrd}(B^\times) \subseteq \mathrm{nrd}(B_v^\times) \subseteq F_v^\times$. So we first characterize the groups $\mathrm{nrd}(B_v^\times) \subseteq F_v^\times$ for v place of F.

Lemma 19.6.1. *Let v be a place of F. If v is a ramified real place, then* $\mathrm{nrd}(B_v^\times) = \mathbb{R}_{>0}^\times$. *Otherwise,* $\mathrm{nrd}(B_v^\times) = F_v^\times$*; moreover, if $\mathcal{O}_v \subseteq B_v$ is a maximal order, then* $\mathrm{nrd}(\mathcal{O}_v^\times) = R_v^\times$.

Proof. If v is split, then $B_v \cong \mathrm{M}_2(F_v)$ and clearly $\mathrm{nrd}(B_v^\times) = \det(\mathrm{GL}_2(F_v)) = F_v^\times$. So suppose v is ramified. If v is real then $B_v \cong \mathbb{H}$ and $\mathrm{nrd}(B_v^\times) = \mathbb{R}_{>0}^\times$. If v is nonarchimedean, then $B_v \cong \left(\dfrac{K_v, \pi_v}{F_v}\right)$ where K_v is the unramified quadratic extension of F_v and π_v is a uniformizer. But $F_v^\times = R_v^\times \times \langle \pi_v \rangle$, and we have $\mathrm{nrd}(K_v^\times) = \mathrm{N}_{K_v/F_v}(K_v^\times) = R_v^\times \times \langle \pi_v^2 \rangle$ and $\mathrm{nrd}(j) = \pi_v$, so the result follows by multiplicativity of the norm. The second clause follows similarly. □

It follows from this lemma that $\mathrm{nrd}(\widehat{B}^\times) = \widehat{F}^\times$. Now we turn to the image of $\mathrm{nrd}(B^\times) \subseteq F^\times$. Let $F_{(+)}^\times$ denote the set of $x \in F$ such that $v(x) > 0$ for all ramified (real) archimedean places $v \mid \infty$. We have just seen that $\mathrm{nrd}(B^\times) \subseteq F_{(+)}^\times$. The following converse is due to Eichler.

Theorem 19.6.2 (Theorem on norms)**.** *We have* $\mathrm{nrd}(B^\times) = F_{(+)}^\times$.

We will use the following lemmas.

Lemma 19.6.3. *Let v be a noncomplex place of F. Let $n \in F_v^\times$, and if v is real suppose $n > 0$. Then there exists $t \in F_v$ such that $T^2 - tT + n$ is separable and irreducible over F_v.*

Proof. We must show that there exists $t \in F_v$ such that $t^2 - 4n \notin F_v^{\times 2}$. We suppose that $\mathrm{char}\, F_v \neq 2$ and leave the other case as an exercise (Exercise 19.18. If v is real, or if v is nonarchimedean and $v(n)$ is odd, we may take $t = 0$. So suppose v is nonarchimedean and without loss of generality that $\mathrm{ord}_v(n) = 0$. Let $e \in R_v^\times \setminus R_v^{\times 2}$. Then there exists a solution to the quadratic form $x^2 - 4ny^2 = ez^2$ with $t, u, z \in R_v$ by previous results; note $y \in R_v^\times$. The element $t = x/y$ is then the desired element. □

Lemma 19.6.4 (Weak approximation for global fields)**.** *Let $v_1, \dots, v_r$ be places of F, let $a_i \in F_{v_i}$, and let $\epsilon > 0$. Then there exists $a \in F$ such that $|a - a_i|_{v_i} < \epsilon$ for all i.*

Proof. This is just the Chinese remainder theorem in drag. □

Proof of Theorem 19.6.2. Let $n \in F_{(+)}^{\times}$ we will construct a separable quadratic extension K/F with $K \hookrightarrow B$ such that $n \in \mathrm{N}_{K/F}(K^{\times})$; it suffices to find K/F with the property that K_v splits B for all ramified places v of F. It is enough to find K/F such that K_v is a field for each ramified v.

By the first lemma above, for each ramified place v of B, there exists $t_v \in F_v$ such that $T^2 - t_vT + n$ is separable and irreducible over F_v. By the second lemma, there exists $t \in F$ which is arbitrarily close to t_v for each v. It follows that such a t exists so that $T^2 - tT + n$ is irreducible over each F_v. Let K be the extension of F obtained by adjoining a root of this polynomial. Then K_v is a field for each ramified v, as desired. □

19.6.5. The surjectivity of $\mathrm{nrd}_v : B_v^{\times} \to F_v^{\times}$ follows, so the reduced norm yields a surjective map

$$\mathrm{nrd} : B^{\times}\backslash\widehat{B}^{\times}/\widehat{\mathcal{O}}^{\times} \to F_{(+)}^{\times}\backslash\widehat{F}^{\times}/\,\mathrm{nrd}(\widehat{\mathcal{O}}^{\times}).$$

If we let $H = \mathrm{nrd}(\widehat{\mathcal{O}}^{\times})F_{(+)}^{\times}$, then H is an open subgroup of $\widehat{F}^{\times}/F^{\times}$, and so there exists a class field K for H, i.e. there exists a surjective map

$$\mathrm{nrd} : \mathrm{Cl}\,\mathcal{O} \to \mathrm{Gal}(K/F).$$

Example 19.6.6. Suppose F is a number field, and S consists of the archimedean places of F, so that R is the full ring of integers in F. Suppose that $\mathcal{O}$ is maximal. Then we have $H = F_{(+)}^{\times}\widehat{R}^{\times}$, and we obtain a surjective map

$$\mathrm{nrd} : \mathrm{Cl}\,\mathcal{O} \to \mathrm{Cl}_{(+)}\,R$$

where $\mathrm{Cl}_{(+)}$ is the strict class group of R corresponding to the cycle given by the product of the real places of F which ramify in B. In particular, if B is unramified at all real places, then we have a surjection $\mathrm{nrd} : \mathrm{Cl}\,\mathcal{O} \to \mathrm{Cl}\,R$.

It is a fundamental result of Eichler that if there exists a place $v \in S$ such that B is unramified at v then this map is injective, and the reduced norm is in fact a bijection; we pursue this in the next chapter.

19.7 Extensions and further reading

Cassels and Frohlich, Chapter 2. Several exercises above were taken from Lorentz workshop. Lenstra and Stevenhagen.

Exercises

19.1. Show that $B^\times \widehat{\mathcal{O}}^\times \cap \widehat{B}_1^\times = B_1^\times \widehat{\mathcal{O}}_1^\times$ if and only if $\operatorname{nrd}(\mathcal{O}^\times) = F_{(+)}^\times \cap \operatorname{nrd}(\widehat{\mathcal{O}}^\times)$ (Remark 20.1.6).

19.2. Let B a quaternion algebra over a global field and suppose that S satisfies the Eichler condition for B. Let $\mathcal{O}$ be a norm-maximal order. Give a direct proof using strong approximation that if $\mathcal{O} \subseteq B$ is an R-order and I is an invertible right fractional $\mathcal{O}$-ideal, then I is principal if and only if $[\operatorname{nrd}(I)]$ is trivial in $\operatorname{Cl}_{(+)} R$. *[Hint: If $x \in B^\times$ satisfies $\operatorname{nrd}(x)R = \operatorname{nrd}(I)$, consider $x^{-1}I$.]*

19.3. Give another proof of Lemma 19.6.1 using quadratic forms (as in Section 10.3).

19.4. Prove an integral version of Eichler's theorem of norms as follows. Let B a quaternion algebra over a global field and suppose that S satisfies the Eichler condition for B. Let $\mathcal{O}$ be a norm-maximal order. Show that $\operatorname{nrd}(\mathcal{O}) = R \cap (F_{(+)} \cup \{0\})$.

19.5. Let I be a (invertible) integral right $M_2(R)$-ideal where R is a DVR with uniformizer π. Show that I is generated by

$$x = \begin{pmatrix} \pi^{e-f} & 0 \\ r & \pi^f \end{pmatrix}$$

where $e, f \in \mathbb{Z}_{\geq 0}$ and $r \in R/\pi^f$ are unique.

19.6. Show that for $|q| > 1$ and $s \in \mathbb{C}$ with $\operatorname{Re} s > 1$ we have

$$\sum_{e=0}^{\infty} \frac{1 + q + \cdots + q^e}{q^{2es}} = \left(1 - \frac{1}{q^{2s}}\right)^{-1} \left(1 - \frac{1}{q^{2s-1}}\right)^{-1}.$$

19.7. Show that if J is an R-lattice in B then $uJ = J$ if and only if $u \in \mathcal{O}_{\mathrm{L}}(J)^\times$.

19.8. Let F be a global field and let K be a finite separable extension of F.

a) Show that $\mathbf{A}_K \cong \mathbf{A}_F \otimes_F K$. *[Hint: Use the fact that $F_v \otimes_F K \cong \prod_w K_w$ where w denotes the places above v.]*

b) Show that we have

$$\mathbf{A}_K = K \otimes_F \mathbf{A}_F = \left\{(x_w)_w \in \textstyle\prod_w K_w : |\mathrm{N}_{K_w/F_v}\, x_v|_v \leq 1 \text{ for almost all } v\right\}$$

so $\mathbf{A}_K$, but that the corresponding statement is *not* true for a quaternion algebra $\widehat{B}$: i.e., that the inclusion

$$\mathbf{A}_B = B \otimes_F \mathbf{A}_F \subset \{(x_v)_v \in \textstyle\prod_v B_v : |\mathrm{nrd}(x_v)|_v \leq 1 \text{ for almost all } v\}$$

is strict.

19.9. For a prime p, let $\widehat{p} = (1, \dots, 1, p, 1, \dots, p) \in \mathbf{A}_{\mathbb{Q}}$ be the adele which is equal to p in the pth and ∞th component and 1 elsewhere. Show that the sequence $\widehat{p}$, ranging over primes p, does not converge in $\mathbf{J}_{\mathbb{Q}}$ (so in particular $\mathbf{J}_{\mathbb{Q}}$ is not compact). However, show that this sequence does have a convergent subsequence (converging to 1) in $\mathbf{J}_{\mathbb{Q}}/\mathbb{Q}^\times$.

19.10. Recall that we have $\widehat{\mathbb{Z}} = \varprojlim_n \mathbb{Z}/n\mathbb{Z} \cong \prod_p \mathbb{Z}_p$.

a) Prove that each $\widehat{x} \in \widehat{\mathbb{Z}}$ has a unique representation as $\widehat{x} = \sum_{n=1}^{\infty} c_n n!$ where $c_n \in \mathbb{Z}$ and $0 \leq c_n \leq n$.

b) Prove that $\widehat{\mathbb{Z}}^\times \cong \widehat{\mathbb{Z}} \times \prod_{n=1}^{\infty} \mathbb{Z}/n\mathbb{Z}$ as profinite groups. *[Hint: Consider the product of the p-adic logarithm maps and use the fact that for every prime power p^e there are infinitely many primes q such that $p^e \parallel (q-1)$.]*

c) Prove that for every positive integer n the natural map $\mathbb{Z}/n\mathbb{Z} \to \widehat{\mathbb{Z}}/n\widehat{\mathbb{Z}}$ is an isomorphism.

d) Prove that there is a bijection from the set of positive integers to the set of open subgroups of $\widehat{\mathbb{Z}}$ mapping $n \mapsto n\widehat{\mathbb{Z}}$.

e) Can you classify all closed subgroups of $\widehat{\mathbb{Z}}$?

19.11. View $\mathbb{Z}$ as a subgroup of $\mathbb{R} \times \widehat{\mathbb{Z}}$ by identifying $n \in \mathbb{Z}$ with (n, n). Give $R \times \widehat{\mathbb{Z}}$ the product topology, and give $\mathbb{S} = (\mathbb{R} \times \widehat{\mathbb{Z}})/\mathbb{Z}$ the quotient topology. The topological group $\mathbb{S}$ is called the *solenoid*.

a) Prove that $\mathbb{S}$ is compact, Hausdorff, and connected.

b) Prove that $\mathbb{S}$ has the structure of a vector space over $\mathbb{Q}$. *[Hint: Show it is torsion-free and divisible.]*

c) Prove that $\widehat{\mathbb{Z}} \cong \mathrm{End}(\mathbb{Q}/\mathbb{Z})$ (as rings).

d) Prove that $\mathbb{S} \cong \mathrm{Hom}(\mathbb{Q}, \mathbb{R}/Z)$ (as groups).

e) Prove that $\mathbb{S} \cong \mathbf{A}_{\mathbb{Q}}/\mathbb{Q}$ (as topological groups).

19.12. Show that if $\mathcal{O}$ is an R-order in B that

$$\mathbf{A}_B = \{(x_v)_v \in \textstyle\prod_v B_v : x_v \in \mathcal{O}_v \text{ for all but finitely many } v\}$$

and

$$\mathbf{J}_B = \left\{(x_v)_v \in \textstyle\prod_v B_v : x_v \in \mathcal{O}_v^\times \text{ for all but finitely many } v\right\}$$

and therefore that this definition is independent of the choice of $\mathcal{O}$.

19.13. Show that B is discrete in $\mathbf{A}_B$ and $B^\times$ is discrete in $\mathbf{J}_B$. *[Hint: F is discrete in $\mathbf{A}_F$ by the product formula.]*

[[Generalization/consequence: if V is a vector space over F then F is discrete in $\mathbf{A}_V = V \otimes_F \mathbf{A}_F$ and $\mathbf{A}_V/V$ is compact. Reduce to the case $V = F = \mathbb{Q}$.]]]

[[Let F be a topological field. Show that the coarsest topology in which multiplication on $\mathrm{M}_2(F)$ is continuous is given by the induced (coordinate) topology.]]

19.14. Give a fundamental system of neighborhoods of 0 in $\widehat{B}$ and of 1 in $\widehat{B}^\times$.

19.15. Let A be a topological ring. Suppose that $A^\times \subseteq A$ has the induced topology. Show that the map $x \mapsto x^{-1}$ on $A^\times$ is not necessarily continuous.

Embed $A^\times \hookrightarrow A \times A$ by $x \mapsto (x, x^{-1})$ and give $A^\times$ the subspace topology. Show that $A^\times$ in this topology is a topological group.

19.16. Show that the topology on $\mathbf{J}_F$ agrees with the subspace topology induced on $\mathbf{J}_F \hookrightarrow \mathbf{A}_F \times \mathbf{A}_F$ by $x \mapsto (x, x^{-1})$.

19.17. Let $\mathcal{O}$ be an R-order in B with $R = R_S$ the ring of S-integers in F. Show that the set of R-orders which are connected to $\mathcal{O}$ is in bijection with $\widehat{B}^\times/N(\widehat{\mathcal{O}})$, where $N(\widehat{\mathcal{O}})$ is the normalizer of $\widehat{\mathcal{O}}$ in $\widehat{B}$.

19.18. Let F_v be a local field with $\operatorname{char} F_v = 2$. Let $n \in F_v$. Show that there exists $t \in F_v$ such that $T^2 - tT + n$ is separable and irreducible.

Chapter 20

Strong approximation

20.1 Strong approximation

In this section, we prove the following important result characterizing the class group in many cases.

Let F be a global field and let S be a finite set of places of F containing all archimedean places. Let $R = R_S$ be the ring of S-integers in B. Let B be a quaternion algebra over F, and let $\mathcal{O} \subseteq B$ be an R-order. Then by Paragraph 19.6.5, the reduced norm map

$$\operatorname{nrd} : \operatorname{Cl} \mathcal{O} = B^\times \backslash \widehat{B}^\times / \widehat{\mathcal{O}}^\times \to F_{(+)}^\times \backslash \widehat{F}^\times / \operatorname{nrd}(\widehat{\mathcal{O}}^\times) \tag{20.1.1}$$

is surjective, where we recall that $\widehat{F}^\times$ is an abbreviation for $\widehat{F}_S^\times = \prod'_{v \notin S} F_v^\times$, etc.

Quite surprisingly, as mentioned at the end of the previous section, in many situations this map is in fact bijective!

Let us investigate the injectivity of the reduced norm map above; it is only a map of sets, after all, but we will show it suffices to look at an appropriate kernel.

Remark 20.1.2. For any $\widehat{y} \in \widehat{B}^\times$, the map $\widehat{x}\mathcal{O} \mapsto \widehat{x}\mathcal{O}\widehat{y}^{-1}$ gives a bijection

$$\operatorname{Cl} \mathcal{O} = B^\times \backslash \widehat{B}^\times / \widehat{\mathcal{O}}^\times \longleftrightarrow B^\times \backslash \widehat{B}^\times / \widehat{\mathcal{O}}'^\times = \operatorname{Cl} \mathcal{O}'$$

where $\mathcal{O}' = B \cap \widehat{y}\widehat{\mathcal{O}}\widehat{y}^{-1}$ is connected (locally isomorphic) to $\mathcal{O}$. So it is sensible to consider the maps (20.1.1) for all orders $\mathcal{O}'$ connected to $\mathcal{O}$.

Let

$$\widehat{B}_1^\times = \{\widehat{x} \in \widehat{B}^\times : \operatorname{nrd}(\widehat{x}) = 1\}$$

be the kernel of $\operatorname{nrd} : \widehat{B}^\times \to \widehat{F}^\times$ (and define similarly $B_1^\times$, etc.).

Lemma 20.1.3. *Let $\Lambda \subseteq B$ be an R-order. Then the map* (20.1.1) *is injective for all orders $\mathcal{O}$ which are locally isomorphic to Λ if and only if $\widehat{B}_1^\times \subseteq B^\times \widehat{\mathcal{O}}^\times$ for all such orders $\mathcal{O}$.*

Proof. One direction is easy: if (20.1.1) is injective then given $\widehat{x} \in \widehat{B}_1^\times$ we have $\operatorname{nrd}(\widehat{x}\widehat{\mathcal{O}}^\times) = \operatorname{nrd}(\widehat{\mathcal{O}}^\times)$ so $\widehat{x}\widehat{\mathcal{O}}^\times = z\widehat{\mathcal{O}}^\times$ for some $z \in B^\times$ so $\widehat{x} \in z\widehat{\mathcal{O}}^\times \subseteq B^\times\widehat{\mathcal{O}}^\times$.

For the converse, since $\operatorname{nrd} : B^\times \to F_{(+)}^\times$ and $\operatorname{nrd} : \widehat{\mathcal{O}}^\times \to \operatorname{nrd}(\widehat{\mathcal{O}}^\times)$ are both surjective, to show nrd is injective for $\mathcal{O}$ we may show that if $\operatorname{nrd}(\widehat{x}) = \operatorname{nrd}(\widehat{y}) \in \widehat{F}^\times$ then $\widehat{x}\widehat{\mathcal{O}}^\times = z\widehat{y}\widehat{\mathcal{O}}^\times$ for some $z \in B^\times$. We consider $(\widehat{xy}^{-1})(\widehat{y}\widehat{\mathcal{O}}\widehat{y}^{-1}) = (\widehat{xy}^{-1})\widehat{\mathcal{O}}'$ where as above $\mathcal{O}' = B \cap \widehat{y}\widehat{\mathcal{O}}\widehat{y}^{-1}$ is locally isomorphic to $\mathcal{O}$ and hence also to Λ. Since $\widehat{xy}^{-1} \in \widehat{B}_1^\times$, by hypothesis we have $\widehat{xy}^{-1} = z\widehat{u}' = z(\widehat{yuy}^{-1})$ where $z \in B^\times$ and $u \in \widehat{\mathcal{O}}^\times$, and consequently $\widehat{x}\widehat{\mathcal{O}} = z\widehat{yu}\widehat{\mathcal{O}} = z\widehat{y}\widehat{\mathcal{O}}$, and hence the map is injective. □

Remark 20.1.4. Above is just the idelic proof of the following statement: if $\widehat{B}_1^\times \subseteq B^\times\widehat{\mathcal{O}}^\times$, then a right invertible fractional $\mathcal{O}$-ideal I is principal if and only if the class of $\operatorname{nrd}(I)$ in $F_{(+)}\backslash\widehat{F}/\operatorname{nrd}(\widehat{\mathcal{O}})$ is trivial.

In order to compare two such ideals I, J with the same norm and show they are isomorphic, we need to show that the colon ideal $(I : J)_{\mathrm{L}} = IJ^{-1}$ is principal; but this colon ideal has right order equal to $\mathcal{O}' = \mathcal{O}_{\mathrm{L}}(J)$ (which is by definition connected to $\mathcal{O}$), and so in order to apply the previous statement we need to know $\widehat{B}_1^\times \subseteq B^\times\widehat{\mathcal{O}}'^\times$.

Question 20.1.5. Can one prove directly that $\widehat{B}_1^\times \subseteq B^\times\mathcal{O}^\times$ for one order $\mathcal{O}$ implies the same statement for all connected orders?

Remark 20.1.6. We have $B^\times\widehat{\mathcal{O}} \cap \widehat{B}_1^\times = B_1^\times\widehat{\mathcal{O}}_1^\times$ if and only if $\operatorname{nrd}(\mathcal{O}^\times) = F_{(+)}^\times \cap \operatorname{nrd}(\widehat{\mathcal{O}}^\times)$ (Exercise 19.1.

If $B = \left(\dfrac{a,b}{F}\right)$, then $B_1^\times = \{(x,y,z,w) \in F^4 : x^2 - ay^2 - bz^2 + abw^2 = 1\}$.

20.1.7. We could hope that $B_1^\times$ is already dense in $\widehat{B}_1^\times$. This would imply in fact that $\widehat{B}_1^\times \subseteq B_1^\times\widehat{\mathcal{O}}_1^\times$, since if $\widehat{x} = (x_v)_v \in \widehat{B}_1^\times$ then $x_v \in \widehat{\mathcal{O}}_1^\times$ for all but finitely many places v, and thus by density there exists $y \in B_1^\times$ which is arbitrarily close to x for these finitely many v and such that $y \in (\widehat{\mathcal{O}}_v)_1^\times$ for all other v, hence for y sufficiently close to x we have $x_v y_v^{-1} \in (\mathcal{O}_v)_1^\times$ for all v (since $\mathcal{O}_v^\times$ contains some neighborhood of 1).

Example 20.1.8. Let $F = \mathbb{Q}$ and take $S = \{\infty\}$ so that $R = \mathbb{Z}$.

If $B = \mathrm{M}_2(\mathbb{Q})$ and $\mathcal{O} = \mathrm{M}_2(\mathbb{Z})$, then one can show using Hensel's lemma (and the Chinese remainder theorem) that indeed $B_1^\times = \mathrm{SL}_2(\mathbb{Q})$ is dense in $\widehat{B}_1^\times = \mathrm{SL}_2(\widehat{\mathbb{Q}})$ [[Exercise]]. We recover the fact that every right ideal in $\mathrm{M}_2(\mathbb{Z})$ is principal.

Example 20.1.9. Now let p, q be odd primes with $q > p$. Consider the quaternion algebra $B = \left(\dfrac{-p,-q}{\mathbb{Q}}\right)$ and let $\mathcal{O} = \mathbb{Z} \oplus \mathbb{Z}i \oplus \mathbb{Z}j \oplus \mathbb{Z}ij$. Then we claim $\widehat{B}_1^\times \not\subseteq B^\times \widehat{\mathcal{O}}^\times$. Indeed, let ℓ be a prime such that $(-p/\ell) = (-q/\ell) = 1$ and $\ell^2 < p$, and suppose

$$a^2 + pb^2 = c^2 + qd^2 = t\ell$$

with $a, b, c, d, t \in \mathbb{Z}$ and $\ell \nmid t$. Now let $\widehat{x} = (x_v)_v \in \widehat{B}^\times$ be such that

$$x_\ell = (a+i)(b+j)^{-1} = \frac{(a+i)(b-j)}{t\ell}$$

and $x_v = 1$ if $v \neq \ell$. Then $\operatorname{nrd}(\widehat{x}) = 1$ so $\widehat{x} \in \widehat{B}_1^\times$. We claim that $\widehat{x} \notin B^\times \widehat{\mathcal{O}}^\times$. Indeed, suppose that $\widehat{x} = y\widehat{u}$ with $y \in B^\times$ and $\widehat{u} \in \widehat{\mathcal{O}}^\times$. Since $\operatorname{nrd}(\mathcal{O}^\times) \cap \mathbb{Q}^\times = \{\pm 1\}$, we may assume $y \in B_1^\times$. Then $\ell\widehat{x}\widehat{u}^{-1} = \ell y = z \in B \cap \widehat{\mathcal{O}} = \mathcal{O}$; thus $\operatorname{nrd}(z) = \ell^2$. But $\operatorname{nrd}|_{\mathcal{O}} = \langle 1, -p, -q, pq\rangle$ only represents ℓ^2 by $\pm\ell$, a contradiction.

We now state the main result.

Definition 20.1.10. Let F be a global field and let S be a finite set of places of F containing all archimedean places. Let B be a quaternion algebra over F. Then we say S *satisfies the Eichler condition* for B if S contains a place which is unramified in B. If F is a number field, we say B *satisfies the Eichler condition* if the set of archimedean places satisfies the Eichler condition for B.

Theorem 20.1.11 (Strong approximation)**.** *Let B be a quaternion algebra over a global field and let S satisfy the Eichler condition for B. Then $B_1^\times$ is dense in $\widehat{B}_1^\times$.*

From the discussion above (Paragraph 20.1.7), we have the following important resulting proposition.

Proposition 20.1.12. *If S satisfies the Eichler condition for B, then the map* (20.1.1) *is a bijection for all R-orders $\mathcal{O} \subseteq B$.*

We say that an R-order $\mathcal{O} \subseteq B$ is *norm-maximal* if $\operatorname{nrd}(\widehat{\mathcal{O}}^\times) = \widehat{R}^\times$.

Corollary 20.1.13. *Let F be a number field and let $S = \{v : v \mid \infty\}$ consist of the archimedean places of F. Let S satisfy the Eichler condition for B and let $\mathcal{O} \subseteq B$ be a norm-maximal R-order. Then* $\operatorname{nrd} : \operatorname{Cl}\mathcal{O} \to \operatorname{Cl}_{(+)} R$ *is a bijection.*

Corollary 20.1.14. *Let F be a global field. Let S be a set of places of F and let $\mathcal{O} \subseteq B$ be a norm-maximal R-order. Let T be a set of primes which generate* $\operatorname{Cl}_{(+)} R$ *and suppose $S \cup T$ satisfies the Eichler condition for B. Then every ideal class in* $\operatorname{Cl}\mathcal{O}$ *contains an integral $\mathcal{O}$-ideal whose reduced norm is supported in T.*

Proof. Let R_T denote the localization of R at the primes in T, that is to say, R_T is the ring of $S \cup T$-integers in F. We apply the above corollary to the order $\mathcal{O}_T = \mathcal{O} \otimes_R R_T$: then we have a bijection $\operatorname{Cl} \mathcal{O}_T \to \operatorname{Cl} R_T$. But $\operatorname{Cl} R_T$ is the quotient of $\operatorname{Cl} R$ by the primes in T and so is trivial. Therefore if I is a right $\mathcal{O}$-ideal, then $I_T = I \otimes_R R_T$ satisfies $I_T = x\mathcal{O}_T$ for some $x \in B^\times$. But now $J = x^{-1}I$ is now a right fractional $\mathcal{O}$-ideal in the same class as I and $(x^{-1}I)_\mathfrak{p} = \mathcal{O}_\mathfrak{p}$ for any prime $\mathfrak{p} \notin T$ and so J has reduced norm supported in T. Replacing J by aJ with $a \in R$ supported in T, we may suppose further that J is integral. □

Example 20.1.15. Let B be a definite quaternion algebra over a totally real field F, let $S = \{v : v \mid \infty\}$ so that R is the ring of integers of F. Let $\mathcal{O}$ be a norm-maximal R-order in B. Suppose that $\operatorname{Cl}_{(+)} R = \{1\}$ and let $\mathfrak{p}$ be a prime of R which is unramified in B. Then every ideal class in $\operatorname{Cl} \mathcal{O}$ contains an integral $\mathcal{O}$-ideal whose reduced norm is a power of $\mathfrak{p}$.

20.1.16. One can think of strong approximation from the following informal perpsective: if we are allowed to "forget" at least one unramified place of B, or equivalently a place where $(B_v)_1^\times$ is not compact, then there is enough room for $B_1^\times$ to "spread out" so that $B_1^\times$ is dense in the S-finite part $\widehat{B}_1^\times$. In other words, we need $(\widehat{B}^S)_1^\times = \prod_{v \in S} (B_v)_1^\times$ to be noncompact. (Recall that

$$\mathbf{J}_B = \widehat{B}^S \times \widehat{B}_S = \prod_{v \in S} B_v^\times \times \prod_{v \notin S}{}' B_v^\times;$$

we consider each factor embedded in $\mathbf{J}_B$ by extension by 1.) This condition is indeed necessary: If $B_1^\times$ was dense in $(\widehat{B}_S)_1^\times$, then the closure of $B_1^\times (\widehat{B}^S)_1^\times$ would be all of $\mathbf{J}_B$; but if $(\widehat{B}^S)_1^\times$ is compact, then $B_1^\times (\widehat{B}^S)_1^\times$ would be a closed subgroup of $(\mathbf{J}_B)_1$ not equal to $(\mathbf{J}_B)_1$, since $B_1^\times$ is discrete in $(\mathbf{J}_B)_1$, a contradiction.

Now we proceed with the proof; we follow roughly the same lines as in the proof of Eichler's theorem on norms, but here instead we will be concerned with traces.

Proof of Theorem 20.1.11 (Strong approximation). Let $\mathcal{O}$ be an R-order in B. We need to show that for any open set $U \subseteq \widehat{B}_1^\times$ that $B_1^\times \cap U \neq \emptyset$.

For this, it suffices to consider open neighborhoods. Let $\widehat{x} = (x_v)_v \in \widehat{B}_1^\times$ and let U be an open neighborhood of $\widehat{x}$, which we can take to be of the form

$$U = \prod_{v \in S_U} x_v U_v \times \prod_{v \notin S_U}{}' (\mathcal{O}_v)_1^\times$$

where S_U is a finite set of places disjoint from S and U_v is an open neighborhood of $1 \in B_v^\times$. Therefore, we may assume $x_v = 1$ for $v \notin S_U$: in other words, every neighborhood of an element $\widehat{x}$ is also a neighborhood of an element which is 1 at all but finitely many places). We can shrink U so that $U_v^2 \subseteq U_v$, therefore we may further assume that $x_v \notin F_v^\times$ for $v \in S_U$, multiplying x_v by any element of $U_v \setminus F_v^\times$.

Let $\widehat{t} = \operatorname{trd}(\widehat{x}) = (t_v)_v$, so that $t_v = 2$ for $v \notin S_U$. For each $v \in S_U$ such that B is ramified, the minimal polynomial $T^2 - t_vT + 1$ of x_v is irreducible, since B_v is a division ring and $x_v \notin F_v^\times$. For each $v \in S$ in which v is ramified, choose any $t_v \in F_v$ such that $T^2 - t_vT + 1$ is irreducible. By hypothesis, there exists a place $w \in S$ which is unramified in B. Since F is dense in $\widehat{F}_{\{w\}} = \prod'_{v \neq w} F_v$, there exists $u \in F$ such that u_v is arbitrarily close to t_v for all places v such that $v \notin S$ or v ramified in B. (We do not require anything at places $v \in S$ where $v \neq w$ and v is unramified!)

It follows that the polynomial $f(t) = T^2 - uT + 1$ is irreducible at all places v where B is ramified, so defines a quadratic extension that embeds in B as in the proof of Eichler's theorem of norms. Let $y \in B_1^\times$ have minimal polynomial $f(t)$. Then since $\operatorname{trd}(y_v)$ has been made arbitrarily close to $t_v = \operatorname{trd}(x_v)$ for $v \notin S$, and since trd is an open map, there exists $\widehat{x'} = (x'_v)_v \in U$ such that $\operatorname{trd}(y_v) = \operatorname{trd}(x'_v)$ for all $v \notin S$. It follows that y_v and x'_v have the same minimal polynomial for all $v \notin S$ and so y and $\widehat{x'}$ are conjugate in $\widehat{B}^\times$, say $\widehat{c} \in \widehat{B}^\times$ satisfies $y = \widehat{c}^{-1}\widehat{x'}\widehat{c}$.

We have therefore shown that for every open neighborhood U of $\widehat{x}$, there exists $\widehat{c} \in \widehat{B}^\times$ such that $B_1^\times \cap \widehat{c}^{-1}U\widehat{c} \neq \emptyset$.

Now let U_n be open neighborhoods of $\widehat{x}$ such that $\bigcap_n U_n = \{\widehat{x}\}$, and let

$$B_1^\times \ni y_n = \widehat{c_n}^{-1}\widehat{x'_n}\widehat{c_n} \in \widehat{c_n}^{-1}U_n\widehat{c_n}.$$

Then $\widehat{x'_n} \to \widehat{x}$.

By the geometry of numbers, the set $\operatorname{Cl}\mathcal{O} = \widehat{\mathcal{O}}^\times\backslash\widehat{B}^\times/B^\times$ is finite, so $\widehat{B}^\times/B^\times$ is compact. Restricting to a subsequence, we may write $\widehat{c_n} = \widehat{d_n}z_n$ with $z_n \in B^\times$ and $\widehat{d_n} \to \widehat{d} = (d_v)_v \in \widehat{B}^\times$.

Now $B^\times$ is dense in $\widehat{B}^\times_{S_U}$ by weak approximation, so there exists a sequence $d_n \in B^\times$ such that $(d_n)_v \to d_v$ for all $v \in S_U$. Therefore

$$B_1^\times \ni (d_n^{-1}z_n)y_n(z_n^{-1}d_n) = d_n^{-1}\widehat{d_n}\widehat{x'_n}\widehat{d_n}^{-1}d_n \to \widehat{x};$$

for $v \in S_U$ this follows since $(d_n)_v \to d_v$, and for $v \notin S_U$ we have $(x'_n)_v \to x_v = 1$ so $(d_n^{-1})_v d_v (x'_n)_v d_v^{-1}(d_n)_v \to 1$ as well.

So we have shown that $\widehat{x} \in \overline{B_1^\times}$ and hence $B_1^\times \cap U \neq \emptyset$. □

20.2 Maps between class sets

When $\mathcal{O}' \subseteq \mathcal{O}$, we have a maps comparing $\operatorname{Cl}\mathcal{O}$ to $\operatorname{Cl}\mathcal{O}'$ by restriction and extension. Strong approximation tells you what these look like.

20.3 Extensions and further reading

Exercises

Chapter 21

Unit groups

21.1 Quaternion unit groups

Having dealt with the question of class numbers, we now turn to another important object associated to a quaternion order: the unit group. By way of analogy, we consider what happens for quadratic orders. In this case, just as with class groups, the behavior of unit groups is quite different depending on if the asociated quadratic field K is real or imaginary.

In the imaginary case, the unit group is finite, as the norm equation $\mathrm{N}_{K/\mathbb{Q}}(\alpha) = 1$ has only finitely many solutions for integral α: these are elements of a 2-dimensional lattice in $\mathbb{C}$ with bounded size. And an element of finite order in $K^\times \hookrightarrow \mathbb{C}^\times$ is a root of unity that satisfies a quadratic equation over $\mathbb{Q}$, and the only such roots of unity have orders $1, 2, 4, 6$. Therefore, only two imaginary quadratic orders have units other than ± 1: the Gaussian order $\mathbb{Z}[\sqrt{-1}]$ of discriminant -4 and the Eisenstein order $\mathbb{Z}[(-1 + \sqrt{-3})/2]$ of discriminant -3.

[[pictures of lattices]]

Orders $\mathcal{O}$ in a *definite* quaternion algebra B over $\mathbb{Q}$ behave like orders in an imaginary quadratic field. The unit group of such an order is finite, as the solutions to $\mathrm{nrd}(\alpha) = 1$ with $\alpha \in \mathcal{O}$ are elements of a 4-dimensional lattice in $\mathbb{R}^4$ again with bounded size.

As with quadratic orders, we can say more: now this unit group embeds as a finite subgroup of $\mathbb{H}_1^\times = \mathrm{SU}(2)$ and so $\mathcal{O}_1^\times/\{\pm 1\} \hookrightarrow \mathrm{SO}(3)$ is a finite group of rotations in $\mathbb{R}^3$. But these groups have been classified by Dickson: they are either cyclic, dihedral, or *exceptional*, a subgroup of A_4, S_4, or A_5 corresponding to the symmetry groups of the tetrahedron, octahedron, or icosahedron. In this chapter, we take up the task of describing explicitly the noncyclic subgroups, working of course in the context of a general definite quaternion order.

However, here we can say what happens over $\mathbb{Q}$ quite neatly. Among the cyclic groups, only subgroups of order $2,4,6$ are possible over $\mathbb{Q}$ for the same reason as the commutative case: an element of $B^\times \setminus \mathbb{Q}$ also satisfies a quadratic equation with rational coefficients (it generates an imaginary quadratic field!). The question of whether or not there is a unit of specified order is the question of whether or not the Gaussian order or the Eisenstein (quadratic) order embeds in the quaternion order $\mathcal{O}$, and these embedding questions are the subject of Chapter 25.

Suppose that $\mathcal{O}^\times$ is dihedral, and let $j \in \mathcal{O}^\times \setminus \{\pm 1\}$ act by inversion (equivalently, conjugation) on a cyclic group (of order 4 or 6). Let K be the (imaginary quadratic) field generated by this group. We have $j^2 \in \mathbb{Q}$ so $j^2 = -1$, and so $j\alpha = \overline{\alpha} j$ for all $\alpha \in K$. Thus we have $B \cong \left(\dfrac{K,-1}{\mathbb{Q}}\right)$, and this leaves only two possibilities.

(i) $B \cong \left(\dfrac{-1,-1}{\mathbb{Q}}\right)$ with discriminant 2, and $\mathcal{O}$ contains the order generated by i, j. If equality holds, and $\mathcal{O}$ has reduced discriminant 4, then $\mathcal{O}^\times$ is the quaternion group Q_8 of order 8. Otherwise, up to isomorphism the order $\mathcal{O}$ is the Hurwitz order (Chapter 9) and $\mathcal{O}^\times$ is an exceptional group of order 24 containing A_4 with index 2.

(ii) $B \cong \left(\dfrac{-3,-1}{\mathbb{Q}}\right)$ with discriminant 3, and $\mathcal{O}$ contains the order generated by $(1+i)/2$ and j which is maximal in B. In this case, $\mathcal{O}^\times \cong D_{12}$ is a dihedral group of order 12.

Therefore, we conclude that if $\mathcal{O}$ is a definite quaternion order over $\mathbb{Z}$ with reduced discriminant greater than 4, then $\mathcal{O}^\times$ is cyclic.

Now we turn to the indefinite case, which like the commutative case is quite different. For the real quadratic order $\mathbb{Z}[\sqrt{d}]$ with $d > 0$, the units are again given by solutions to the Pell equation $x^2 - dy^2 = \pm 1$ with $x, y \in \mathbb{Z}$. All solutions up to sign are given by powers of a *fundamental solution* which can be computed explicitly using continued fractions; consequently, $\mathbb{Z}[\sqrt{d}]^\times = \langle -1, u\rangle \cong \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}$ where $u = x + y\sqrt{d}$ is the *fundamental unit* . However, the fundamental unit is often (but not always) very large, being of exponential size in the discriminant, by theorems of Schur and Siegel.

In a similar way, we consider units in an order $\mathcal{O} = \mathbb{Z} \oplus \mathbb{Z}i \oplus \mathbb{Z}j \oplus \mathbb{Z}ij$ in an indefinite quaternion algebra $B = \left(\dfrac{a,b}{\mathbb{Q}}\right)$ with $a, b > 0$. The norm equation then reads

$$t^2 - ax^2 - by^2 + abz^2 = \pm 1$$

with $t, x, y, z \in \mathbb{Z}$. Amusingly, this "quaternion Pell equation" combines Pell equations for $\mathbb{Z}[\sqrt{a}]$ and $\mathbb{Z}[\sqrt{b}]$ by restriction, and in fact by considering embeddings of quadratic orders (the subject of Chapter 25), we see that this equation combines *all* Pell equations satisfying certain congruence conditions. Combining these Pell equations, we see that the group of solutions is an infinite, noncommutative group.

We seek understand the group $\mathcal{O}^\times$ by its action on a suitable space, and in this way we are led to consider groups acting discretely on symmetric spaces; we will discover that the group $\mathcal{O}^\times$ is finitely presented. This investigation is detailed but fruitful, involving the theory of Fuchsian and Kleinian groups, and it the focus of our investigation in Part III of this text. In this chapter, we discuss a few issues concerning the structure of these groups.

21.2 Structure of units

Throughout this chapter, let F be a global field, let S be a nonempty set of places of F containing the archimedean places, and let R be the ring of S-integers in F.

21.2.1. From Dirichlet's unit theorem (and its generalization to the function field case), the group $R^\times$ of units of R is a finitely generated abelian group of rank $\#S - 1$, so that $R^\times \cong \mathbb{Z}/w\mathbb{Z} \oplus \mathbb{Z}^{\#S-1}$ where w is the number of roots of unity in F.

Let B be a quaternion algebra over R, and let $\mathcal{O}$ be an R-order in B. We are interested in the structure of the group $\mathcal{O}^\times$: as this group is (in general) noncommutative, it is much more difficult to describe than for $R^\times$! Its description will depend on the set S, and in particular in the number field case on the number of real and complex places. To begin, we consider some basic structure of this group.

Since the center of $B^\times$ is $F^\times$, the center of $\mathcal{O}^\times$ is $R^\times$. We understand the structure of $R^\times$ by Dirchlet's unit theorem (Paragraph 21.2.1), so we turn first to understand the group $\mathcal{O}^\times/R^\times$.

Example 21.2.2. If $B = \mathrm{M}_2(F)$ and $\mathcal{O} = \mathrm{M}_2(R)$, then $\mathcal{O}^\times = \mathrm{GL}_2(R)$ and $\mathcal{O}^\times/R^\times = \mathrm{PGL}_2(R)$.

The reduced norm gives a map $\mathrm{nrd} : \mathcal{O}^\times \to R^\times$. The image of the reduced norm is determined by Eichler's theorem of norms (the integral version is discussed in Exercise 19.4): we have an exact sequence

$$1 \to \mathcal{O}_1^\times \to \mathcal{O}^\times \xrightarrow{\mathrm{nrd}} R_{(+)}^\times$$

where $\mathcal{O}_1^\times = \{x \in \mathcal{O}^\times : \mathrm{nrd}(x) = 1\}$ and

$$R_{(+)}^\times = \{x \in R^\times : v(x) > 0 \text{ for all real } v \in \mathrm{Ram}(B)\}.$$

Now we have $R^\times \subseteq \mathcal{O}^\times$, and $\operatorname{nrd}(R^\times) = R^{\times 2}$ so $\mathcal{O}_1^\times \cap R^\times = \{\pm 1\}$, so this yields an exact sequence

$$1 \to \frac{\mathcal{O}_1^\times}{\{\pm 1\}} \to \frac{\mathcal{O}^\times}{R^\times} \xrightarrow{\text{nrd}} \frac{R_{(+)}^\times}{R^{\times 2}} \tag{21.2.3}$$

Example 21.2.4. If $B = \mathrm{M}_2(F)$ and $\mathcal{O} = \mathrm{M}_2(R)$, then the exact sequence (21.2.3) becomes

$$1 \to \mathrm{PSL}_2(R) \to \mathrm{PGL}_2(R) \xrightarrow{\det} R^\times/R^{\times 2} \to 1.$$

The group $R^\times$ is finitely generated, so the group $R_{(+)}^\times/R^{\times 2}$ is a finite, elementary abelian 2-group, isomorphic to a Cartesian power of $\mathbb{Z}/2\mathbb{Z}$. If F is a function field, then $R_{(+)} = R^\times$.

Remark 21.2.5. If F is a number field, then the group $R_{(+)}^\times/R^{\times 2}$ is a class group; specifically, if $\mathrm{Cl}^{(+)}\,\mathbb{Z}_F$ denotes the ray class group of F with modulus equal to the real places in F ramified in B, then $R_{(+)}^\times/R^{\times 2}$ is isomorphic to the quotient of $\mathrm{Cl}^{(+)}\,\mathbb{Z}_F/\,\mathrm{Cl}\,\mathbb{Z}_F$ by the finite primes in S.

In general, the exact sequence (21.2.3) does not split, so in general the group $\mathcal{O}^\times/R^\times$ will be an extension of $\mathcal{O}_1^\times/\{\pm 1\}$ by an elementary abelian 2-group: see already Example (21.2.4).

21.3 Units in definite quaternion orders

We now begin our investigation of the unit group.

First, we recall the proof of Dirichlet's unit theorem: we embed $R^\times$ modulo torsion as a discrete subgroup of a nice topological group. To build intuition, consider the case where F is a number field with r real places and c complex places, and S is the set of infinite places so that R is the ring of integers of F. Then we embed

$$F \hookrightarrow F_\infty = \mathbb{R}^r \times \mathbb{C}^c \tag{21.3.1}$$

and R sits discretely inside F_∞ as a $\mathbb{Z}$-lattice; this is the embedding of R into its completions at all places in S. Consequently, we have a map $R^\times \to \mathbb{R}^{r+c}$ given by $x \mapsto (\log |x|_v)_v$; the kernel of this map is the group of roots of unity and the image is still discrete. What is more, the image of this map is cocompact inside the hyperplane $\sum_v x_v = 0$, and consequently it is isomorphic to $\mathbb{Z}^{r+c-1}$. From this, we see that the most basic question about $R^\times$, whether it is finite or infinite, is determined by the set S: we have $\#R^\times < \infty$ if and only if $r + c - 1 = 0$, which leaves only the possibilities $(r, c) = (1, 0)$ (and $F = \mathbb{Q}$) or $(r, c) = (0, 1)$ (and F is an imaginary quadratic field).

One might say that $R^\times$ is finite only when the completions provide no room for the unit group to become infinite. With this in mind, we make the following definition.

Definition 21.3.2. A quaternion algebra B over a global field F is *S-definite* if every place in S is ramified in B.

Since a complex place is necessarily unramified, we see that if B is S-definite over F then in particular F is a totally real number field.

Proposition 21.3.3. *The group $\mathcal{O}^\times/R^\times$ is finite if and only if B is S-definite.*

Proof of Proposition 21.3.3. By the exact sequence (21.2.3), the group $\mathcal{O}^\times/R^\times$ is finite if and only if the group $\mathcal{O}_1^\times$ is finite.

First, suppose B is not S-definite. Then there is an place $v \in S$ that is unramified. Therefore there exists a separable quadratic field extension K of F that embeds in B such that v splits in K: if v is infinite, then either v is complex or v is real and there are two real places above v in K. Let S be the integral closure of R in K (note that the ring S is not the set S). Then by the Dirichlet S-unit theorem (Paragraph 21.2.1), the rank of $S^\times/R^\times$ is at least 1. The order $S \cap \mathcal{O}$ has finite index in S, so $S^\times/(S \cap \mathcal{O})^\times$ is a finite group. Therefore, if a sufficiently high power of any $S^\times \setminus R^\times$ will lie in $S^\times \cap \mathcal{O}^\times = (S \cap \mathcal{O})^\times$, so $\mathcal{O}^\times$ is infinite.

Now suppose that B is S-definite. Then by definition, for each $v \in S$, the completion $B_v = D_v$ is a division algebra over F_v. Consider the setup in analogy with Dirichlet's unit theorem. We consider the embedding of B into the completions at all places in S:

$$B \hookrightarrow B_S = \prod_{v \in S} D_v.$$

Since R is discrete in $F_S = \prod_{v\in S} F_v$ (in the number field case, S contains all archimedean places), the ring $\mathcal{O}$ is discrete in B_S (Exercise 23.1). Consequently, $\mathcal{O}^\times$ is discrete in $B_S^\times$ and $\mathcal{O}_1^\times$ sits discretely in

$$\prod_{v \in S} (D_v)_1^\times.$$

But each $(D_v)_1^\times$ is compact, from the discussion in Section 10.6. Therefore $\mathcal{O}_1^\times$ is a discrete subgroup of a compact group and hence finite. □

Example 21.3.4. Let $B = \left(\dfrac{-1,-1}{\mathbb{Q}}\right)$ and let $\mathcal{O}$ be the $\mathbb{Z}$-order generated by i, j, so that $S = \{\infty\}$. Then $\mathcal{O}^\times = \langle i, j\rangle \cong Q_8$ is the quaternion group of order 8.

Now invert 5, and consider $S = \{5, \infty\}$ and the order $\mathcal{O}[1/5]$ over $R = \mathbb{Z}[1/5]$. Then $\mathcal{O}$ contains the element $1 + i$ of norm $5 \in R^\times$, and this element has infinite order.

On the other hand, now invert 2. Then $\mathcal{O}[1/2]^\times \cong \langle 2, i, j, 1 + i\rangle$ so the group $\mathcal{O}[1/2]^\times/\langle -1, 2\rangle$ is an extension of Q_8 by $\mathbb{Z}/2\mathbb{Z}$ (Exercise 21.1).

21.4 Explicit definite unit groups

If B is definite, then $\mathcal{O}_1^\times$ is a finite subgroup of $\mathbb{H}_1^\times$. The exact sequence

$$1 \to \{\pm 1\} \to \mathbb{H}_1^\times \to \mathrm{SO}_3(\mathbb{R}) \to 1$$

implies that $\mathcal{O}_1^\times/\{\pm 1\}$ is a subgroup of a Coxeter group, and can be realized inside the automorphism group of a regular polyhedron (Platonic solid) and so is a subgroup of one of the following groups: cyclic, dihedral, or one of the three exceptional groups A_4, S_4, or A_5.

Cyclic, dihedral, tetrahedral, octahedral, icosahedral and their extensions; these have presentations as triangle groups.

Can detect cyclic groups because we must have $\mathbb{Q}(\zeta_{2m})^+ \subseteq F$ and

$$B \cong \left(\frac{-1, \lambda_{2m}^2 - 4}{F}\right).$$

Dihedral can happen. So for $F = \mathbb{Q}$ and "most" fields, just cyclic groups of order $4, 6$.

For A_4, S_4, A_5, we must have $B = \left(\dfrac{-1, -1}{F}\right)$ and $F \supseteq \mathbb{Q}, \mathbb{Q}(\sqrt{2}), \mathbb{Q}(\sqrt{5})$, respectively.

Consider the extensions.

21.5 Extensions and further reading

21.5.1. Quaternion Pell equation investigated by [Jah10].

Exercises

21.1. Let $B = \left(\dfrac{-1, -1}{\mathbb{Q}}\right)$ and let $\mathcal{O}$ be the $\mathbb{Z}$-order generated by i, j. Prove that $\mathcal{O}[1/2]^\times \cong \langle 2, i, j, 1 + i\rangle$ and describe $\mathcal{O}[1/2]^\times/\mathbb{Z}[1/2]^\times$ as an extension of Q_8 by $\mathbb{Z}/2\mathbb{Z}$.

Chapter 22

Picard groups

22.1 Locally free class groups

Now we really want a group, not a set, so we consider the Grothendieck group. Let B be a quaternion algebra and $\mathcal{O} \subset B$ be an order.

Definition 22.1.1. A right $\mathcal{O}$-module M is *locally free* if $\widehat{M} = M \otimes_{\mathcal{O}} \widehat{\mathcal{O}} \cong \widehat{\mathcal{O}}^r$ (as right $\mathcal{O}$-modules) for some $r \in \mathbb{Z}_{\geq 1}$, called the *rank* $r = \operatorname{rk} M$.

What about rank 0? Compare with other definitions.

Definition 22.1.2. Two projective right $\mathcal{O}$-modules M and M' of finite rank are *stably isomorphic* if there exists an isomorphism

$$M \oplus \mathcal{O}^s \cong \mathcal{M}' \oplus \mathcal{O}^s$$

for some $s \in \mathbb{Z}_{\geq 0}$.

M is *stably free* if M is stably isomorphic to $\mathcal{O}^r$ (for $r = \operatorname{rk} M$).

What about s, s'? This should just follow from the Grothendieck formalism.

Let $\operatorname{Pic} \mathcal{O}$ be the set of stable isomorphism classes of right $\mathcal{O}$-modules.

Theorem 22.1.3. *Let M be a locally free right $\mathcal{O}$-module. Then there exists a locally free (fractional) $\mathcal{O}$-ideal I such that*

$$M \cong \mathcal{O}^{r-1} \oplus I.$$

Proof. □

It follows that if I_1, I_2 are right $\mathcal{O}$-ideals, then there exists a locally free right $\mathcal{O}$-ideal I_3 such that

$$I_1 \oplus I_2 \cong \mathcal{O} \oplus J. \tag{22.1.4}$$

Probably you can be much more explicit in this quaternion case.

This gives a group law and a map to $\mathbb{Z}$ on the Grothendieck group. Maybe it's better to call it K_0?

Proposition 22.1.5. $\mathrm{Pic}(\mathcal{O})$ *is a finite group.*

Proof. Geometry of numbers, or the Jordan-Zassenhaus theorem. □

Lemma 22.1.6. *If* $\mathcal{O}, \mathcal{O}'$ *are of the same type, then* $\mathrm{Pic}(\mathcal{O}) \cong \mathrm{Pic}(\mathcal{O}')$.

If $\mathcal{O}' \subseteq \mathcal{O}$ *then the extension map gives a surjection* $\mathrm{Pic}(\mathcal{O}') \to \mathrm{Pic}(\mathcal{O})$.

Reduced norm map, comparison between Pic^1 and Cl given by (22.1.4).

22.2 Cancellation

Now the version of the class number 1 problem.

Definition 22.2.1. $\mathcal{O}$ has the *cancellation property* if for all right $\mathcal{O}$-modules M_1, M_2, N of finite rank, we have $M_1 \oplus N \cong M_2 \oplus N$ implies $M_1 \cong M_2$.

Lemma 22.2.2. *The following are equivalent:*

(i) *$\mathcal{O}$ has the cancellation property;*

(ii) *M is stably free if and only if M is free; and*

(iii) $\mathrm{Cl}(\mathcal{O}) = \mathrm{Pic}^1(\mathcal{O})$.

It follows from Lemmas 22.1.6 and 22.2.2(c) that if $\mathcal{O}' \subseteq \mathcal{O}$ and $\mathcal{O}'$ has the cancellation property, then so does $\mathcal{O}$.

Theorem 22.2.3 (Eichler, Jacobinski)**.** *If B satisfies the Eichler condition, then $\mathcal{O}$ has the cancellation property.*

Proof. (Don't we need class number 1 for this?) □

In 1962, Swan gave an example. (See work of Smertnig.)

Theorem 22.2.4 (Vignéras, Hallouin–Maire, Smertnig)**.** *There are finitely many isomorphism classes of definite quaternion orders for which the cancellation property holds. There are exactly 128? definite Eichler orders with the cancellation property.*

Proof. □

22.3 Extensions and further reading

Exercises

Part III

Arithmetic geometry

Chapter 23

Geometry

23.1 Arithmetic groups

[[Adelic measure is the statement]]

$$\frac{\pi^2}{6} = \prod_p \frac{1}{1 - 1/p^2}.$$

So we turn to the groups $\mathcal{O}_1^\times$ when B is indefinite. From the above, $\mathcal{O}_1^\times$ is a discrete subgroup of

$$(B_{\mathbb{R}})_1^\times \cong (\mathbb{H}_1^\times)^d \times \mathrm{SL}_2(\mathbb{R})^{r-d} \times \mathrm{SL}_2(\mathbb{C})^c;$$

However, since $\mathbb{H}_1^\times$ is compact, it follows that $\mathcal{O}_1^\times$ is a discrete subgroup of the product $\mathrm{SL}_2(\mathbb{R})^{r-d} \times \mathrm{SL}_2(\mathbb{C})^c$ (Exercise 23.6). It follows that

$$\Gamma(\mathcal{O}) = \mathcal{O}_1^\times/\{\pm 1\} \hookrightarrow \mathrm{PSL}_2(\mathbb{R})^{r-d} \times \mathrm{PSL}_2(\mathbb{C})^c.$$

It follows that any projection down to a smaller group is *not* discrete (Exercise 23.3).

Definition 23.1.1. A *Fuchsian group* is a discrete subgroup of $\mathrm{PSL}_2(\mathbb{R})$.

A *Kleinian group* is a discrete subgroup of $\mathrm{PSL}_2(\mathbb{C})$.

We see that when $d = r - 1$ and $c = 0$, i.e. F is totally real and B is ramified at all but one real place, the group $\Gamma(\mathcal{O})$ is a Fuchsian group; and when $d = r$ and $c = 1$, i.e. F has exactly one complex place and B is ramified at all real places of F, the group $\Gamma(\mathcal{O})$ is a Kleinian group. In fact, these statements are equivalences. [[Losing a factor loses discreteness.]]

The groups $\mathrm{PSL}_2(\mathbb{R})$ and $\mathrm{PSL}_2(\mathbb{C})$ arise in geometry as groups of orientation-preserving isometries of hyperbolic two- and three-space.

A small class of groups, as there are only finitely many conjugacy classes of arithmetic Fuchsian groups of any signature or with bounded coarea.

For nonarithmetic groups, the commensurator is finite (theorem of Margulis), so there the correspondences/Hecke operators are finite.

Detecting arithmetic groups: theorem of Takeuchi. More generally, can always define a trace field, S-integral group.

23.2 Fuchsian groups

We consider first the simplest case of Fuchsian groups.

First, a brief review of hyperbolic geometry. We define the *upper half-plane*

$$\mathcal{H} = \{z = x + iy \in \mathbb{C} : \Re(z) = y > 0\}.$$

This space is equipped with the *hyperbolic metric* given by

$$ds^2 = \frac{dx^2 + dy^2}{y^2} = \frac{|dz|^2}{(\mathrm{Im}\, z)^2}.$$

Thus, the hyperbolic area of a region $D \subseteq \mathcal{H}$ is given by

$$\mathrm{area}(D) = \int\!\!\int_D \frac{dx\,dy}{y^2}.$$

[[As in the Iwasawa-Tate style treatment of zeta functions of number fields and residue of first pole as volume of idele class group (with the non-compact "ray" removed), this volume is essentially the residue of the leading pole of the zeta function of the quaternion algebra, and this zeta function factors as zeta and a shift of zeta of the groundfield, up to finitely-many factors depending on ramification of the quaternion algebra. This kind of computation is treated in Weil's "Basic Number Theory", and also in his "Adeles and Algebraic Groups".]]

The group $\mathrm{PSL}_2(\mathbb{R})$ acts on $\mathcal{H}$ as orientation-preserving isometries by linear fractional transformation

$$z \mapsto \frac{az + b}{cz + d}$$

with $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{R})$. A discrete subgroup $\Gamma \subseteq \mathrm{PSL}_2(\mathbb{R})$ acts *discontinuously* on $\mathcal{H}$, i.e. for each compact subset $K \subseteq \mathcal{H}$, the set $K \cap gK = \emptyset$ for all but finitely many $g \in \Gamma$. Thus, the stabilizer of a point is finite.

The topology on $\mathrm{SL}_2(\mathbb{R})$ and $\mathrm{PSL}_2(\mathbb{R})$ is determine by matrix entries.

Definition 23.2.1. A *Fuchsian group* is a discrete subgroup of $\mathrm{PSL}_2(\mathbb{R})$.

Remark 23.2.2. First and second kind.

Example 23.2.3. [[Exercise!]] Compute the area of the usual fundamental domain for the action of $\mathrm{SL}_2(\mathbb{Z})$:

$$D = \{z \in \mathcal{H} : |\Re z| \leq 1 \text{ and } |z| \geq 1.$$

Now let F be a totally real number field and let B be a quaternion algebra which is ramified at all but one real place. Let $\mathcal{O}$ be a maximal order in B. Then by the above, we have an embedding

$$\Gamma(\mathcal{O}) = \mathcal{O}_1^\times/\{\pm 1\} \hookrightarrow \mathrm{PSL}_2(\mathbb{R})$$

realizing $\Gamma(\mathcal{O})$ as discrete subgroup.

It is another exercise in multivariable integration to prove the following formula using the zeta function.

Proposition 23.2.4 (Volume formula)**.** *The group $\Gamma(\mathcal{O})$ has finite coarea in $\mathcal{H}$, and*

$$\operatorname{area}(\Gamma(\mathcal{O})\backslash\mathcal{H}) = \frac{8\pi}{(4\pi^2)^n} d_F^{3/2} \zeta_F(2)\Phi(\mathfrak{D})$$

where $\Phi(\mathfrak{D}) = \prod_{\mathfrak{p}|\mathfrak{D}}(N\mathfrak{p} - 1)$.

The quotient $\Gamma(\mathcal{O}) \setminus \mathcal{H}$ can be given the structure of a Riemann surface, called a *Shimura curve*.

Theorem 23.2.5. *$\Gamma(\mathcal{O}) \setminus \mathcal{H}$ is compact if and only if B is a division ring.*

Definition 23.2.6. Two subgroups are *commensurable* if their intersection is of finite index in both, and *commmensurable in the wide sense* if conjugates of the groups are commensurable.

Commensurable groups are related by correspondences on the Riemann surfaces, and the commensurability class is the quaternion algebra.

Example of $\mathrm{PSL}_2(\mathbb{Z})$.

23.3 Kleinian groups

In a similar way, we can identify the group $\mathrm{PSL}_2(\mathbb{C})$ as a group of isometries. Here, we define *hyperbolic three-space* by

$$\mathcal{H}^3 = \{(z,t) \in \mathbb{C} \times \mathbb{R} : t > 0\}.$$

The hyperbolic metric on $\mathcal{H}^3$ is induced from the line element ds by

$$ds^2 = \frac{dx^2 + dy^2 + dt^2}{t^2}.$$

The space $\mathcal{H}^3$ is the unique three-dimensional connected and simply connected Riemannian manifold with constant sectional curvature -1. The volume element is accordingly $(dx\,dy\,dt)/t^3$.

The group $\mathrm{PSL}_2(\mathbb{C})$ acts as orientation-preserving isometries of $\mathcal{H}^3$ by linear fractional transformations: letting $w = z + tj$, we have

$$w \mapsto (aw + b)(cw + d)^{-1}$$

for $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{C})$ (normalized so that $ad - bc = 1$?).

Now let F be a number field with one complex place and let B be a quaternion algebra over F which is ramified at all real places of F. Let $\mathcal{O}$ be a maximal order in B. Then by the above, we have an embedding

$$\Gamma(\mathcal{O}) = \mathcal{O}_1^\times/\{\pm 1\} \hookrightarrow \mathrm{PSL}_2(\mathbb{C})$$

realizing $\Gamma(\mathcal{O})$ as discrete subgroup.

It is another exercise in multivariable integration to prove the following formula using the zeta function.

Proposition 23.3.1 (Volume formula)**.** *The group $\Gamma(\mathcal{O})$ has finite covolume in $\mathcal{H}^3$ and*

$$\mathrm{vol}(\Gamma(\mathcal{O})\backslash\mathcal{H}^3) = \frac{1}{(4\pi^2)^{n-1}} d_F^{3/2} \zeta_F(2)\phi(\mathfrak{D}).$$

[[Example of $\mathrm{PSL}_2(\mathbb{Z}[i])$.]]

23.4 Arithmetic groups, revisited

There is a more general notion of arithmetic group, defined for a reductive group; in this section, we show that this definition is equivalent to ours.

23.5 Orthogonal group

Rephrase in terms of orthogonal group, with an eye toward generalizations.

[[I'm pretty sure that this implies that an automorphism of a quaternion algebra has to be inner, giving a quadratic forms proof of Skolem–Noether.]]

23.6 Extensions and further reading

Exercises

In these exercises, we maintain the notation in this section: let F be a number field with ring of integers R, and let $\mathcal{O}$ be an R-order in a quaternion algebra B over F.

23.1. Show that R is discrete in $F_{\mathbb{R}} = F \otimes_{\mathbb{Q}} \mathbb{R}$. *[Hint: it is enough to show this for a neighborhood of* 0*, and then use the fact that the norm must be an integer.]* Use this to show that $\mathcal{O}$ is discrete in $B \otimes_{\mathbb{Q}} \mathbb{R}$.

23.2. Let $d \in \mathbb{R} \setminus \mathbb{Q}$. Show that $\mathbb{Z}[\sqrt{2}]$ is not discrete in $\mathbb{R}$. (This gives a reason to worry about discreteness of number fields when we project.)

23.3. Show that the image of $\mathcal{O}_1^\times$ in a projection to any proper factor of $\mathrm{SL}_2(\mathbb{R})^{r-d} \times \mathrm{SL}_2(\mathbb{C})^c$ is not discrete.

23.4. Let $B = \left(\dfrac{a,b}{F}\right)$ be a quaternion algebra over a number field $F \hookrightarrow \mathbb{R}$. Let $\mathcal{O} \subset B$ be an order. Show that $\mathcal{O}_1^\times$ is discrete in $\mathrm{SL}_2(\mathbb{R})$ if and only if k is totally real and for all nonidentity real places v, we have $v(a), v(b) < 0$.

23.5. In this exercise, we give a direct argument for the discreteness of an arithmetic Fuchsian group. Let $B = \left(\dfrac{a,b}{F}\right)$ be a quaternion algebra over a totally real number field $F \hookrightarrow \mathbb{R}$. Let $\mathcal{O} \subset B$ be an order. Suppose that B is ramified at all nonidentity real places. We will show that $\mathcal{O}_1^\times$ is discrete in $\mathrm{SL}_2(\mathbb{R})$.

a) Suppose not: then there exists a sequence $\alpha_n = t_n + x_n i + y_n j + z_n ij \to 1$ with $t_n, x_n, y_n, z_n \in F$ with bounded denominators. Multiplying through, assume that all coordinates are integral. Thus for n sufficiently large, all of the coordinates are integral and bounded.

b) Show that for all nonidentity v, the coordinates of $v(\alpha_n)$ are also bounded using compactness.

c) Show that there are only a finite number of elements in R that are bounded in each coordinate (all conjugates are bounded); look at the coefficients of a minimal polynomial. Derive a contradiction.

23.6. If $H \hookrightarrow G_1, G_2$ is a subgroup of topological groups G_1 and G_2, $H \hookrightarrow G_1 \times G_2$ is discrete, and G_1 is compact, then $H \hookrightarrow G_2$ is discrete.

Chapter 24

Fuchsian and Kleinian groups: examples

24.1 Triangle groups

Chapter 25

Embedding numbers

25.1 Representation numbers

Of binary quadratic forms by indefinite quaternary forms or integers by indefinite ternary quadratic forms.

25.2 Selectivity

25.3 Isospectral, nonisometric orbifolds

Sunada and Vignéras.

25.4 Extensions and further reading

Exercises

25.1.

Chapter 26

Formalism of Shimura varieties

In this section, we introduce quaternionic Shimura varieties which give a geometric way of viewing quaternion algebras over number fields. Roughly speaking, the unit group of an order in a quaternion algebra acts on a hyperbolic space and the quotient is an arithmetic manifold.

26.1 Modular curves

As motivation, we consider the case of classical modular curves. This case corresponds to the simplest situation, that where $F = \mathbb{Q}$ and $B = \mathrm{M}_2(\mathbb{Q})$. The order $\mathcal{O} = \mathrm{M}_2(\mathbb{Z})$ is maximal in B and any maximal order is conjugate in B to $\mathcal{O}$.

We have $\mathbf{A}_B \cong \mathrm{M}_2(\mathbf{A}_\mathbb{Q}) \cong \mathrm{M}_2(\widehat{\mathbb{Q}}) \times \mathrm{M}_2(\mathbb{R})$. We have seen that $B = \mathrm{M}_2(\mathbb{Q})$ sits discretely in $\mathbf{A}_B \cong \mathrm{M}_2(\mathbf{A}_\mathbb{Q})$ and the quotient $\mathrm{M}_2(\mathbf{A}_\mathbb{Q})/\mathrm{M}_2(\mathbb{Q}) \cong \mathrm{M}_2(\mathbf{A}_\mathbb{Q}/\mathbb{Q})$ is compact—and like the adeles themselves, not very interesting (from this perspective).

We turn then to $\mathbf{J}_B = \mathrm{GL}_2(\mathbf{A}_\mathbb{Q}) \cong \mathrm{GL}_2(\widehat{\mathbb{Q}}) \times \mathrm{GL}_2(\mathbb{R})$. In the previous section, we understood the double quotient space

$$B^\times \backslash \widehat{B}^\times / \mathcal{O}^\times = \mathrm{GL}_2(\mathbb{Q}) \backslash \mathrm{GL}_2(\widehat{\mathbb{Q}}) / \mathrm{GL}_2(\widehat{\mathbb{Z}}) = \mathrm{Cl}\,\mathrm{M}_2(\mathbb{Z}) = \{1\}$$

as identifying the set of right invertible fractional $\mathcal{O}$-ideals up to isomorphism, i.e. the set of $\mathbb{Z}$-lattices I in B with $\mathrm{End}(I) \cong \mathrm{M}_2(\mathbb{Z})$ (acting on the right), up to isomorphism.

But this description leaves out the real archimedean place! Indeed, we have $B_\mathbb{R}^\times = (B \otimes_\mathbb{Q} \mathbb{R})^\times \cong \mathrm{GL}_2(\mathbb{R})$, and $B^\times = \mathrm{GL}_2(\mathbb{Q}) \hookrightarrow \mathrm{GL}_2(\mathbb{R})$. From this description, we see that a $\mathbb{Z}$-lattice $I \subseteq B$ embeds as $I \hookrightarrow I \otimes_\mathbb{Z} \mathbb{R} \cong \mathrm{M}_2(\mathbb{R})$.

So what does the set

$$B^\times \backslash (B_\mathbb{R}^\times \times \widehat{B}^\times / \mathcal{O}^\times) = \mathrm{GL}_2(\mathbb{Q}) \backslash (\mathrm{GL}_2(\mathbb{R}) \times \mathrm{GL}_2(\widehat{\mathbb{Q}}) / \mathrm{GL}_2(\widehat{\mathbb{Z}}))$$

represent? Here, note that $\mathcal{O}^\times$ acts by right multiplication on $\widehat{B}^\times$ but $B^\times$ acts by left multiplication (embedded diagonally) on both factors $B_\mathbb{R}^\times \times \widehat{B}^\times$.

As it stands, this is not yet interesting: by the above

Other orders include those defined by congruence conditions: for example, for each $N \in \mathbb{Z}_{>0}$, we have

$$\mathcal{O}_0(N) = \left\{\begin{pmatrix} a & b \\ c & d \end{pmatrix} : N \mid c\right\}$$

and

$$\mathcal{O}(N) = \left\{\begin{pmatrix} a & b \\ c & d \end{pmatrix} : N \mid c, N \mid b, a \equiv d \pmod N\right\}.$$

To introduce the adelic description, we first give its derivation for the space $Y(1)_\mathbb{C}$. We have seen that $Y(1)_\mathbb{C}$ is in bijection with the set of lattices in $\mathbb{C}$ up to isomorphism, which we denote $\cong \backslash \mathrm{Lat}(\mathbb{C})$. But in fact a lattice in $\mathbb{C}$ is really a lattice in $\mathbb{R}^2$ together with a *complex structure* $\psi : \mathbb{C} \to \mathrm{End}_\mathbb{R}(\mathbb{R}^2)$, and these are in bijection with $\mathbb{C} \setminus \mathbb{R} = \mathfrak{H}^\pm$ as follows: choose $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{M}_2(\mathbb{R})$ such that $M\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}M^{-1} = \psi(i)$; then we take $\tau = \dfrac{ai+b}{ci+d} \in \mathfrak{H}^\pm$. Therefore $Y(1)_\mathbb{C}$ is in bijection with

$$\cong \backslash(\mathfrak{H}^\pm \times \mathrm{Lat}(\mathbb{R}^2)) = GL_2(\mathbb{R})\backslash(\mathfrak{H}^\pm \times \mathrm{Lat}(\mathbb{R}^2)).$$

And for every lattice $\Lambda \in \mathrm{Lat}(\mathbb{R}^2)$, we can find an $M \in GL_2(\mathbb{R})$ such that $M\Lambda \subseteq \mathbb{Q}^2 \subseteq \mathbb{R}^2$, so this set is also in bijection with $GL_2(\mathbb{Q})\backslash(\mathfrak{H}^\pm \times \mathrm{Lat}(\mathbb{Q}^2))$.

Next, we introduce the adeles. For a $\mathbb{Z}$-module S, we define $\widehat{S} = S \otimes_\mathbb{Z} \widehat{\mathbb{Z}}$, where $\widehat{\mathbb{Z}} = \varprojlim \mathbb{Z}/n\mathbb{Z}$. We then see that the map

$$\begin{aligned} \mathrm{Lat}(\mathbb{Q}^2) &\to \mathrm{Lat}(\widehat{\mathbb{Q}}^2) \\ \Lambda &\mapsto \widehat{\Lambda} \end{aligned}$$

is a bijection, with inverse $\widehat{\Lambda} \mapsto \widehat{\Lambda} \cap \mathbb{Q}^2$. Now since $GL_2(\widehat{\mathbb{Q}})$ acts transitively on $\mathrm{Lat}(\widehat{\mathbb{Q}}^2)$, with stabilizer of a lattice $\widehat{\Lambda}$ given by $GL_2(\widehat{\mathbb{Z}})$, we have in sum a bijection

$$Y(1)(\mathbb{C}) \leftrightarrow GL_2(\mathbb{Q})\backslash(\mathfrak{H}^\pm \times GL_2(\widehat{\mathbb{Q}})/GL_2(\widehat{\mathbb{Z}})).$$

This gives a description for the quaternion algebra $B = M_2(\mathbb{Q})$. For any indefinite quaternion algebra B over $\mathbb{Q}$ with maximal order $\mathcal{O}$, we have in a similar way the set

$$B^*\backslash(\mathfrak{H}^\pm \times \widehat{B}^*/\widehat{\mathcal{O}}^*).$$

This description can be seen exactly as the set of (real) lattices inside $B \otimes_{\mathbb{Q}} \mathbb{R}$ together with a complex structure, up to isomorphism.

One then proves that this set has the structure of a complex manifold, and we have:

Theorem 26.1.1 (Deligne)**.** *There exists a curve $X^B_{\mathbb{Q}}$ defined over $\mathbb{Q}$, and an analytic isomorphism*

$$X^B_{\mathbb{Q}}(\mathbb{C}) \xrightarrow{\sim} B^* \backslash (\mathfrak{H}^{\pm} \times \widehat{B}^* / \widehat{\mathcal{O}}^*).$$

Remark 26.1.2. It is more natural to define a Shimura curve by giving an *incoherent* quaternion algebra.

26.2 Modular forms

26.3 Global embeddings

26.4 Extensions and further reading

Exercises

Chapter 27

Definite quaternion algebras

27.1 Class numbers

Class number 1

27.2 Two-sided ideals

Exercise from Kimberly.

27.3 Theta functions

Application to sums of squares, the level is the minimal integer N such that NQ^{-1} is integral. For individual quadratic forms, get an explicit answer. For general quadratic forms, estimate the coefficients.

27.4 Brandt matrices

27.5 Jacquet-Langlands correspondence

[[Trace formula for $GL_2(F)$ in Ling's book; Zagier]]. [[Hijikata, "Explicit formula"]].

27.6 Relationship to elliptic curves

In this section, we discuss the relationship between endomorphism rings of supersingular elliptic curves and quaternion algebras, with applications to modular forms.

Let E be an elliptic curve over field k. For each (nonzero) *isogeny* (finite surjective homomorphism) $f : E \to E'$, there exists a *dual isogeny* $f^\vee : E' \to E$ such that $f^\vee \circ f$ is equal to multiplication by the degree $\deg f \in \mathbb{Z}_{>0}$. In particular, the dual $^\vee$ on $\operatorname{End}(E) = \operatorname{End}_{\overline{k}}(E)$ yields a nonsingular standard involution on $\operatorname{End}(E)$.

The $\mathbb{Q}$-algebra $\operatorname{End}(E)_{\mathbb{Q}} = \operatorname{End}(E) \otimes_{\mathbb{Z}} \mathbb{Q}$ is a division ring, and therefore from previous work, it follows that $\operatorname{End}(E)_{\mathbb{Q}}$ is either $\mathbb{Q}$, an imaginary quadratic field K, or a quaternion algebra over $\mathbb{Q}$.

In fact, the latter possibility can only occur when $\operatorname{char} k = p$, and we say E is *supersingular*: equivalently,

(i) $\operatorname{End}(E)_{\mathbb{Q}}$ has rank 4 as a $\mathbb{Q}$-algebra;

(ii) $E[p](\overline{k}) = 0$;

(iii) $\operatorname{Tr}(\phi) \equiv 0 \pmod{p}$ where ϕ is the Frobenius endomorphism.

If $\#k = p \geq 5$ then E is supersingular if and only if $\#E(k) = p + 1$.

Let E be a supersingular elliptic curve; then E may be defined over a finite field k of characteristic p. We will start from (i) and show in fact that $\operatorname{End}(E)$ is a maximal order in the quaternion algebra ramified at p and ∞; this result is due to Deuring, but we follow a proof given by Lenstra.

Let $\mathcal{O} = \operatorname{End} E$ and $B = \mathcal{O} \otimes_{\mathbb{Z}} \mathbb{Q}$. Whenever $n \in \mathbb{Z}_{>0}$ is prime to p, there is an isomorphism

$$E[n] = E[n](\overline{k}) \cong \mathbb{Z}/n\mathbb{Z} \oplus \mathbb{Z}/n\mathbb{Z}$$

as abelian groups, so the endomorphism ring of this abelian group is $\operatorname{End} E[n] \cong \mathrm{M}_2(\mathbb{Z}/n\mathbb{Z})$. By the existence of the dual isogeny, $E[n]$ is a faithful module over $\mathcal{O}/n\mathcal{O}$, i.e. the map $\mathcal{O}/n\mathcal{O} \to \operatorname{End} E[n]$ is injective. Since $\#\mathcal{O}/n\mathcal{O} = \# \operatorname{End} E[n]$, this map is an isomorphism.

It follows that for every prime $\ell \neq p$, the map

$$\mathcal{O} \hookrightarrow \mathcal{O}_\ell = \mathcal{O} \otimes_{\mathbb{Z}} \mathbb{Z}_\ell \xrightarrow{\sim} \operatorname{End} E[\ell^\infty] = \lim_n E[\ell^n](\overline{k}) = T_\ell E \cong \mathrm{M}_2(\mathbb{Z}_\ell)$$

is an isomorphism, and in particular $\mathcal{O}_\ell$ is maximal and B is split at ℓ.

Since B is a division ring, it follows that B is ramified at p at infinity.

For $f \in \mathcal{O}$, let $\deg_i f$ be the inseparable degree of f, which is a power of p. We put $\deg_i 0 = \infty$. Then we have $\deg_i(fg) = \deg_i(f)\deg_i(g)$ and $\deg_i(f + g) \geq \min\{\deg_i f, \deg_i g\}$—this follows from the fact that $\deg_i f$ is divisible by a given power q of p if and only if f factors via the qth power Frobenius morphism $E \to E^{(q)}$. It follows that $\deg_i : \operatorname{End}(E)_{\mathbb{Q}} \to \mathbb{Z} \cup \{\infty\}$ is a valuation on $\operatorname{End}(E)_{\mathbb{Q}}$. Extended to $\operatorname{End}(E)_{\mathbb{Q}_p}$, this valuation "agrees" with the usual valuation on $\mathbb{Q}_p$, since $\deg_i p = p^2$; factoring an isogeny into its separable and inseparable parts (the latter a power of

Frobenius) shows that $\deg_i f = \operatorname{ord}_p(f^\vee \circ f)$. Hence $\mathcal{O}_p$ is the valuation ring of B_p and hence maximal. This shows that $\mathcal{O}$ itself is maximal in B.

Consider the category SSEll of supersingular elliptic curves over $\overline{k}$, an algebraically closed field of characteristic p. The objects of SSEll are supersingular elliptic curves over $\overline{k}$ and the morphisms are isogenies. Choose a base object $E \in$ SSEll, and let $\mathcal{O} = \operatorname{End}(E)$ and $B = \mathcal{O} \otimes_{\mathbb{Z}} \mathbb{Q}$ as above. For any other $E' \in$ SSEll, the $\operatorname{Hom}_{\overline{k}}(E, E') = \operatorname{Hom}(E, E')$ has the structure of a right $\mathcal{O}$-module (precomposing) and the structure of a left $\mathcal{O}' = \operatorname{End}(E')$-module (postcomposing).

Let J be a (nonzero) integral *left* $\mathcal{O}$-ideal. Define

$$E[J] = E[J](\overline{k}) = \{P \in E(\overline{k}) : f(P) = O \text{ for all } f \in I\}.$$

Then $E[J]$ is a finite subgroup of E and every finite subgroup of E is of the form $E[J]$ for some left $\mathcal{O}$-ideal J. (In fact, $\#E[J] = \operatorname{nrd}(J)$.)

Let $E' = E/E[J]$ and consider the quotient map $\phi : E \to E'$ and its dual $\phi^\vee : E' \to E$. Then post-composing with $\phi^\vee$ gives an injective map $\operatorname{Hom}(E, E') \to \operatorname{End}(E) = \mathcal{O}$.

Let $\mathcal{O}' = \operatorname{End}(E')$. Then we have an embedding $\mathcal{O}' \hookrightarrow \mathcal{O}$ of $\mathbb{Z}$-modules by $g \mapsto \overset{\vee}{\phi} g\phi$, and an element $f \in \mathcal{O}$ gives rise to such an endomorphism if and only if $f(E[J]) \subset E[J]$ if and only if $(Jf)(E[J]) = \{O\}$ if and only if $Jf \subset J$ if and only if $f \in \mathcal{O}_{\mathrm{R}}(J)$. In this way, we identify $\operatorname{End}(E') \cong \mathcal{O}_{\mathrm{R}}(J)$.

If $E' \in$ SSEll then $\operatorname{End}(E')$ is isomorphic to a maximal order in B so $\operatorname{End}(E') \cong \mathcal{O}_{\mathrm{R}}(J)$ for some left $\mathcal{O}$-ideal J; considering the dual isogeny, we see that $\operatorname{End}(E') \cong \mathcal{O}_{\mathrm{L}}(I)$ for the right $\mathcal{O}$-ideal $I = J^{-1}$.

In this correspondence, we identify $\operatorname{Hom}(E', E) = J$ since $s \in \mathcal{O}$ factors via E' if and only if $s \in J$. Dualizing, we have $\operatorname{Hom}(E, E') = I$. In particular, this implies that $\operatorname{Hom}(E_1, E_2) \cong I_2 I_1^{-1} \cong (I_2 : I_1)_{\mathrm{L}}$ and so

$$\begin{aligned} \operatorname{Hom}(E_1, E_2) &\to \operatorname{Hom}(I_1, I_2) \\ f &\mapsto (\phi \mapsto f \circ \phi) \end{aligned}$$

is bijective.

Let $\mathsf{Mod}_{\mathcal{O}}$ be the category of right $\mathcal{O}$-modules with morphisms given by nonzero (right $\mathcal{O}$-module) homomorphisms. We have proven the following proposition.

Proposition 27.6.1 (Kohel [Koh96, Theorem 45]). *The association*

$$\operatorname{Hom}(E, -) : \mathsf{SSEll} \to \mathsf{Mod}_{\mathcal{O}}$$

is a functor and defines an equivalence of categories.

Proof. We have shown that $\operatorname{Hom}(E, -)$ is full, faithful, and essentially surjective. □

Remark 27.6.2. By using right $\mathcal{O}$-modules, the functor $\operatorname{Hom}(E,-)$ is covariant.

Note in particular that the set of isomorphism classes of supersingular elliptic curves corresponds the set of right $\mathcal{O}$-ideal classes. From the Eichler mass formula, we conclude that

$$\sum_{[E]} \frac{1}{\#\operatorname{Aut} E} = \sum_{[I]\in \operatorname{Cl}\mathcal{O}} \frac{1}{\#\mathcal{O}_{\mathrm{L}}(I)} = \frac{p-1}{24}.$$

This is an equivalence between right ideal classes and *not* the left orders.

Lemma 27.6.3. *Let $\mathcal{O}$ be a maximal order. Then there exist one or two supersingular elliptic curves E up to isomorphism over $\overline{k}$ (equivalently, j-invariants) such that* $\operatorname{End}(E) \cong \mathcal{O}$.

Moreover, there exist two such elliptic curves if and only if $j(E) \in \mathbb{F}_{p^2} \setminus \mathbb{F}_p$ if and only if the unique two-sided ideal of $\mathcal{O}'$ of reduced norm p is not principal.

Recall that two-sided ideals can be recovered locally and thus for maximal orders generate a group isomorphic to $\mathbb{Z}^d$ where d is the number of finite ramified primes in B.

Proof. There exists at least one by the above. Without loss of generality, suppose that $\operatorname{End}(E') \cong \mathcal{O}$. Suppose that E' corresponds to the right $\mathcal{O}$-ideal I, namely $\operatorname{End}(E') = \mathcal{O}_{\mathrm{L}}(I)$. Then there exists $x \in B^\times$ such that $\operatorname{End}(E') \cong \mathcal{O}_{\mathrm{L}}(xI) = x\mathcal{O}_{\mathrm{L}}(I)x^{-1} = \mathcal{O}$ by hypothesis of the above isomorphism. Thus xI is a two-sided (fractional) $\mathcal{O}$-ideal which is principal if and only if $E \cong E'$.

Therefore there is a bijection between isomorphism classes of supersingular elliptic curves with $\operatorname{End}(E) \cong \mathcal{O}$ and two-sided ideal classes in $\mathcal{O}$. This proves the result. □

We can generalize this setup slightly as follows. Let $N \in \mathbb{Z}_{>0}$ be coprime to p, and let $C \subset E(\overline{k})$ be a cyclic subgroup of order N. Then $\operatorname{End}_{\overline{k}}(E,C) = \mathcal{O}_0(N)$ is an Eichler order of level N in B. Then $\operatorname{Hom}((E,C),-)$ defines an equivalence of categories between the category of supersingular elliptic curves equipped with a cyclic N-isogeny (with morphisms given by isogenies which identify the cyclic subgroups), to the category of right invertible $\mathcal{O}_0(N)$-modules (with morphisms given by homomorphisms). The mass formula now reads

$$\sum_{[(E,C)]} \frac{1}{\#\operatorname{Aut}(E,C)} = \sum_{[I]\in \operatorname{Cl}\mathcal{O}_0(N)} \frac{1}{\#\mathcal{O}_{\mathrm{L}}(I)} = \frac{p-1}{24}\psi(N).$$

One can also consider instead the category of cyclic N-isogenies $f : E \to E'$ and note that this is equivalent to the category of *cyclic* homomorphisms $f : I \to J$ where $J/\phi(I) \cong (\mathbb{Z}/N\mathbb{Z})^2$, so that $J/\phi(I)$ is a principal right $\mathcal{O}$-module.

Example 27.6.4. Consider $p = 11$. The algebra $B = \left(\frac{-1,-11}{\mathbb{Q}}\right)$ has discriminant 11 and the maximal order $\mathcal{O}$ generated by i and $(1+j)/2$. We have $\#\operatorname{Cl}\mathcal{O} = 2$, with the nontrivial class represented by the ideal I generated by 2 and $1 + i(1+j)/2$.

We have $\mathcal{O}^\times = \langle i \rangle$ of order 4 and $\mathcal{O}_{\mathrm{L}}(I) = \langle 1/2 - i(1+j)/4 \rangle$ of order 6, and indeed $1/4 + 1/6 = 10/24 = 5/12$. The two supersingular curves modulo 11 are the ones with j-invariants 0 and $1728 \equiv 1 \pmod{11}$, and $\operatorname{End}(E) \cong \mathcal{O}$ if $j(E) = 1728$ whereas for $\operatorname{End}(E') \cong \mathcal{O}'$ we have $\operatorname{Hom}(E, E') \cong I$, in other words, $E' \cong E/E[I]$.

Let $I_1, \dots, I_h$ be representatives for the set $\operatorname{Cl}\mathcal{O}$. Let $\mathcal{O}_i = \mathcal{O}_{\mathrm{L}}(I_i)$ and let $w_i = \#\mathcal{O}_i^\times$. The products $M_{ij} = I_j I_i^{-1}$ are fractional $\mathcal{O}_i, \mathcal{O}_j$-ideals. If E_i is an elliptic curve with $\operatorname{End}(E_i) = \mathcal{O}_i$, then we have $I_i = \operatorname{Hom}(E, E_i)$ hence $M_{ij} = I_j I_i^{-1} = \operatorname{Hom}(E_i, E_j)$.

Define the *theta series* $\theta_{ij}(q)$ by

$$\theta_{ij}(q) = \frac{1}{w_j} \sum_{x \in M_{ij}} \exp(2\pi i \operatorname{nrd}(x)/\operatorname{nrd}(M_{ij})\tau) = \sum_{m \geq 0} B_{ij}(m) q^m$$

where $q = \exp(2\pi i \tau)$.

The functions θ_{ij} as functions on the upper half-plane (for $\tau \in \mathcal{H}$) are modular forms of weight 2 for the group $\Gamma_0(p)$. Their Fourier coefficients $B_{ij}(m)$ give the entries of the Brandt matrix $B(m) = (B_{ij}(m))_{i,j=1,\dots,h}$. We have $B(1)$ is the identity matrix.

Example 27.6.5. We return to the example $p = 11$.

Let $I_1 = \mathcal{O}$ and $I_2 = I$. We have $M_{11} = \mathcal{O}$, and in the basis $1, i, (1+j)/2, i(1+j)/2$ we have $\operatorname{nrd}(x, y, z, w) = x^2 + xz + y^2 + yw + 3w^2 + 3z^2$, so

$$\theta_{11}(q) = \frac{1}{4}(1 + 4q + 4q^2 + 8q^3 + 20q^4 + 16q^5 + 32q^6 + \dots).$$

In a similar way, we have the basis $2, 2i, 1 - 3/2i - 1/2ij, 3/2 - i - 1/2j$ for I and

$$\theta_{12}(q) = \frac{1}{6}(1 + 12q^2 + 12q^3 + 12q^4 + 12q^5 + \dots)$$

and $\theta_{21}(q) = \frac{3}{2}\theta_{12}(q)$ and

$$\theta_{22}(q) = \frac{1}{6}(1 + 6q + 6q^3 + 24q^4 + 18q^5 + 32q^6 + \dots).$$

Indeed, $B(1)$ is the identity matrix.

Example 27.6.6. In the case $p = 11$, we have $B(2) = \begin{pmatrix} 1 & 2 \\ 3 & 0 \end{pmatrix}$ with characteristic polynomial $(T - 3)(T + 2)$. The eigenvalue $3 = 2 + 1$ corresponds to an Eisenstein series, and the eigenvalue -2 corresponds to the (isogeny class of the) elliptic curve $X_0(11)$ given by the equation $y^2 + y = x^3 - x^2 - 10x - 20$: indeed, $\#X_0(11)(\mathbb{F}_2) = 2 + 1 - (-2) = 5$.

Lemma 27.6.7. *The entry $B_{ij}(m)$ is equal to the number of subgroups C of order m in E_i such that $E_i/C = E_j$.*

Return to example with maximal unit group (icosahedron group) and mass formula.

27.7 Ramanujan graphs

Let G be a k-regular connected graph with n vertices and with adjacency matrix T and *combinatorial Laplacian* $k - T$ whose eigenvalues are $0 < \mu - 1 \leq \mu_2 \leq \cdots \leq \mu_{n-1} \leq 2k$. (Adjusted average a function on the neigbors of v.) The *expansion coefficient* of G is

$$h(G) = \min_{\#S \leq n/2} \frac{\#\partial S}{\#S}.$$

One is interested in getting a large coefficient.

Tanner, Alon-Milman:

$$\frac{2\mu_1}{k + 2\mu_1} \leq h(G) \leq \sqrt{2k\mu_1}$$

Alon-Boppana:

$$\liminf \lambda_2(G) \geq 2\sqrt{k-1}.$$

Define Ramanujan graph, random graph is Ramanujan.

A k-regular infinite tree is the ideal expander, with expansion coefficient $k - 1$. Find subgroups of its automorphism group that does not identify vertices that are too close to each other. Bruhat-Tits tree, identify units of norm 1.

Bound on the eigenvalues of the adjacency matrix is given by the Ramanujan-Petersson bound on coefficients.

27.8 Extensions and further reading

Exercises

27.1. Consider the analogous isogeny tree of CM elliptic curves.

Chapter 28

Drinfeld modules and function fields

In this chapter, we discuss the function field side of the global field picture.

28.1 Extensions and further reading

Exercises

Part IV

Concluding material

Chapter 29

Other topics

29.1 Quaternionic polynomial rings

29.2 Matrix rings over quaternion rings

29.3 Unitary groups and Hermitian forms

29.4 Unit groups of integral group rings

In $\mathbb{Z}[G]$ for G a finite group, get interesting unit groups.

29.5 Representation theory of quaternion algebras

29.6 Quaternion rings and Azumaya algebras

Neukirch, get references from my paper, including GL and Lucianovic.

29.7 Octonions and composition algebras

29.8 Lie theory

Appendix A

Hints and solutions to selected exercises

1.1 If B contains $\mathbb{C}$, then B is a $\mathbb{C}$-vector space, so B has even dimension as an $\mathbb{R}$-vector space. Or see May [May66, p. 290]: if $ij = a + bi + cj$ with $a, b, c \in \mathbb{R}$, multiply on the left by i, and derive a contradiction.

2.3 For such a map, we must have $ij \mapsto \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Check that the four matrices are linearly independent, so the map is an F-linear isomorphism. Then, using the universal property of algebras given by generators and relations, show that the given matrices satisfy the relations in B, so the map is an F-algebra homomorphism.

2.7 Use Exercise 2.4(c) and show that the center over $\overline{F}$ has dimension 1 or compute directly with $xi - ix = xj - jx = 0$ for $x = u + vi + wj + zk \in B$.

2.13 Use the left regular representation either to F or a subfield K, and use the block matrix determinant. See also Aslaksen [Asl96].

3.3 It is $(x, y) \mapsto (y, x)$. Note that F embeds diagonally in $F \times F$.

3.5 $g \mapsto g^{-1}$ is a standard involution if and only if G has exponent 2 and char $F = 2$ (so the standard involution is the identity and $F[G]$ is commutative).

3.8 Let $i, j \in K \setminus F$. Then $i + j$ satisfies a quadratic polynomial, but $ji = ij$, so we have $(i+j)^2 = i^2 + 2ij + j^2 \in F(i+j) + F$ hence $2ij = c(i+j) + d$ with $c, d \in F$: but then since $2 \neq 0$, we have $2i \neq c \in F$ so $j = (ci + d)/(2i - c) \in K$.

3.9 For part (a), Suppose B has degree 2. Choose a basis $1, x_2, \dots, x_m$. For each i, the quadratic F-algebras $F[x_i]$ have a standard involution, and so extending by F-linearity we obtain a map $\overline{} : B \to B$. For $x \in B$, let $t(x) = x + \overline{x}$ and $n(x) = x\overline{x}$.

By induction and F-linearity, we may suppose $1, x, y$ are F-linearly independent. Suppose $(x + y)^2 - s(x + y) + m = 0$ with $s, m \in F$. We show that $s = t(x) + t(y)$. We have

$$\begin{aligned}(x-y)^2 = x^2 - (xy+yx) + y^2 &= 2(x^2+y^2) - s(x+y) + m \\ &= (2t(x) - s)x + (2t(y) - s)y + (m - 2n(x) - 2n(y))\end{aligned}$$

But $(x - y)^2 \in F(x - y) + F$ so $2t(x) - s = s - 2t(y)$, i.e. $2s = 2t(x) + 2t(y)$. Since $\operatorname{char} F \neq 2$, we have $s = t(x) + t(y)$ as desired.

To conclude, we show $\overline{xy} = \overline{y}\,\overline{x}$. We may suppose $xy \notin F$. We verify that both $(xy)^2 - (xy + \overline{y}\,\overline{x})xy + (\overline{y}\,\overline{x})(xy) = 0$ and $(xy)^2 - (xy + \overline{xy})xy + \overline{xy}(xy) = 0$, so the result follows by uniqueness of the minimal polynomial.

For part (b), by the uniqueness of the standard involution, we have $\overline{x} = x + 1$ if $x \notin F$. But then if $1, x, y$ are F-linearly independent we have $x + y + 1 = \overline{x+y} = \overline{x} + \overline{y} = (x + 1) + (y + 1) = x + y$, a contradiction. So $\dim_{\mathbb{F}_2} B \leq 2$. Since a Boolean ring consists of idempotents, we have $B = \mathbb{F}_2$ or $B \cong \mathbb{F}_2^2$.

3.10 Under right multiplication by $B = \mathrm{M}_n(F)$, a matrix is nothing other than the direct sum of its rows, so in particular, the characteristic polynomial of right multiplication by $A \in \mathrm{M}_n(F)$ acting on $\mathrm{M}_n(F)$ will be the nth power of the usual characteristic polynomial of A acting on row vectors $V \cong F^n$. (In the language of Chapter 6, $B = \mathrm{M}_n(F)$ as a right B-module is $B \cong V^n$ where $V \cong F^n$ is the unique simple right B-module.)

3.11 By F-linearity, it suffices to verify these statements on a basis for B.

3.14 See van Praag [vPr68] (or its summary [vPr02, Remark 4]).

4.5 For part (a), to simplify the proof of the second statement, choose a normalized basis for V.

4.7 For part (c), by the transitivity of trace, we may assume K/F is purely inseparable and $[K : F]$ is a multiple of p. But then all roots of the minimal polynomial of $x \in K$ over F are equal, so the characteristic polynomial of multiplication by $x \in K$ has all roots equal and there are a multiple of p of them and thus the trace is zero.

For part (d), $\operatorname{Tr}((a+b\sqrt{5})^2) = 2(a^2+5b^2)$ and

$$\operatorname{Tr}((a+b\alpha+c\alpha^2)^2) = 2(a^2-2ab+10ac+5b^2-8bc+13c^2).$$

4.9 If $B = \left(\dfrac{a,b}{\mathbb{F}_q}\right)$ then $K = \mathbb{F}_q(i) \cong \mathbb{F}_{q^2}$ and $\mathrm{N} : \mathbb{F}_{q^2} \to \mathbb{F}_q$ is surjective so $b \in \mathrm{N}_{K/F}(K^\times)$.

4.10 If i, j and i', j' are generators, respectively, then consider the subalgebras generated by $i \otimes 1$ and $j \otimes j'$, and $i \otimes i'$ and $1 \otimes j'$.

4.13 For the first, exhibit an explicit isometry $\langle 1,1,1\rangle \cong \langle 2,3,6\rangle$. For the second, note that $\langle 2,5,10\rangle$ represents 7 but $\langle 1,1,1\rangle$ does not (by showing $x^2+y^2+z^2+w^2 \not\equiv 0 \pmod 8$ for $x,y,z,w \in \mathbb{Z}$ with $\gcd(x,y,z,w)=1$); or note that $\langle 1,1,1\rangle$ represents 1 but $\langle 2,5,10\rangle$ (looking modulo 5, and arguing similarly).

5.2 Choose $0 \neq y \in K^\perp$.

14.10 See Shimura [Shi71, Proposition 4.11, (5.4.2)].

6.2 The map $a \otimes b \mapsto (x \mapsto ax\overline{b})$ gives an F-algebra homomorphism $B \otimes_F B \to \operatorname{End}_F(B) \cong \mathrm{M}_4(F)$, which is injective since $B \otimes_F B$ is simple and therefore an isomorphism by a dimension count.

6.6 The augmentation ideal is the kernel of the surjective map $\sum_g a_g g \mapsto \sum_g a_g$, so is nontrivial.

6.12 This exercise was given in a course by Bjorn Poonen in Spring 2000 at the University of California, Berkeley.

First, parts (a) and (b). Choose $x \in D \setminus F$. Then $K = F(x)$ is a purely inseparable extension of F so the minimal polynomial of x in D (or in $\overline{F}$) is of the form $T^{p^n} - a$. In particular, $p \mid [K:F]$, but D is a left K-vector space and $[D:F] = [D:K][K:F]$ so $p \mid [D:F]$.

For part (c), all roots of the minimal polynomial of x are equal, hence all eigenvalues of $x \otimes 1 \in \mathrm{M}_n(F)$ are equal, and the number of them is divisible by p by (a), so the trace is zero. For part (d), by (c), all elements of $\mathrm{M}_n(\overline{F})$ have trace zero, which is a contradiction.

6.13 Let $j \in B^\times$ satisfy $jxj^{-1} = \overline{x}$. Then $B = K \oplus Kj$, but $j^2xj^{-2} = x$ so $j^2 \in Z(B)$ so $j^2 = b \in F^\times$.

6.16 By Corollary 6.6.8, every maximal subfield K of B has the same dimension, so since F is a finite field they are isomorphic (as abstract fields). But then by the Skolem–Noether theorem, since every element lies in a maximal subfield, we have $B^\times = \bigcup_{x \in B^\times} x^{-1}K^\times x$, which is a contradiction.

One can also proceed without using the maximal subfield dimension theorem. Suppose B is a minimal counterexample (by cardinality); then B is a division ring, but every subalgebra of B is a field. Let $F = Z(B)$. Let $i \in B \setminus F$; then by minimality, the centralizer of i is a maximal subfield K. We may assume $K = F(i)$. If $B = K$, we are done. Otherwise, let i have multiplicative order m. Consider $L : B \to B$ by $L(x) = ixi^{-1}$. Then L is a K-linear map with L^m equal to the identity. We may therefore decompose B into eigenspaces for L. Arguing as in the case of quaternion division rings, we show that each such nonzero eigenspace has dimension 1 as K-vector space. Now consider the normalizer $N = N_B(K)$. Then there is a bijection between the set of cosets of $N/K^\times$ and the eigenspaces of L. But N acts on K as F-linear automorphisms with kernel $K^\times$, so $N/K^\times$ is a subgroup of the Galois group $\operatorname{Gal}(K/F)$. It must be the full Galois group, otherwise $N/K^\times$ fixes some subfield and its centralizer is a noncommutative F-subalgebra, contradicting minimality. Therefore $\dim_K B = \dim_F K$. We now proceed as above.

8.6 Using the matrix units, show that if $M = (m_{ij})_{i,j} \in \mathcal{O}$ then $m_{ij} \cdot 1 \in \mathcal{O}$, but then m_{ij} is integral over R so in fact $m_{ij} \in R$ and hence $M \in \mathrm{M}_n(R)$.

8.8 The converse is true if $\operatorname{char} F \neq 2$ and R is integrally closed. It is immediate if $1/2 \in R$ since $\operatorname{trd}(x^2) = \operatorname{trd}(x)^2 - 2\operatorname{nrd}(x)$, so $2\operatorname{nrd}(x) \in R$. But for the same reason more generally we have $2\operatorname{nrd}(x^n) = 2\operatorname{nrd}(x)^n \in R$ so $R[\operatorname{nrd}(x)] \subseteq (1/2)R$; so if R is integrally closed we have in fact $\operatorname{nrd}(x) \in R$.

The statement is false if $\operatorname{char} F = 2$: take $B = F \times F$ (with $\operatorname{char} F = 2$) and $x = (a, a)$ with $a \in F$ not integral over R. Then $\operatorname{trd}(x^n) = 2a^n = 0$ for all n but x is not integral.

10.2 The quadratic form $\langle -1, e, -1 \rangle$ is isotropic by a previous exercise, so diagonalizing we have $\langle -1, e \rangle \cong \langle 1, s \rangle$ for some $s \in k^\times$. But $\operatorname{disc}(\langle -1, e \rangle) = -e = s = \operatorname{disc}(\langle 1, s \rangle) \in k^\times/k^{\times 2}$, so $\langle 1, s \rangle \cong 1, -e\rangle$. More generally, this argument shows that two nonsingular binary quadratic forms over a finite field are isometric if and only if they have the same discriminant.

10.6 The proof that addition and multiplication are continuous with respect to the absolute value $|\,|$ induced by w is identical to the commutative case. We have a filtration $\mathcal{O} \supset P \supset P^2 \supset \dots$ where P is generated by j and thus to show that B is complete it suffices to note that the limit of the partial sums $x_0 + x_1 j + x_2 j^2 + \cdots = (x_0 + x_2\pi + \dots) + (x_1 + x_3\pi + \dots)j \in K + Kj$ exists since K is complete. The set $\mathcal{O}$ is compact since it is complete and totally bounded. By translating, since $\mathcal{O}$ is open we have that B is locally compact. Finally, if $x \notin \mathcal{O}$ then $w(x) < 0$ so the ring generated by $\mathcal{O}$ and x is equal to B; but B is not compact, since the open cover $\bigcup_i \pi^{-i}\mathcal{O}$ has no subcover.

10.13 Write B in the form $B = \left(\dfrac{K, 2}{\mathbb{Q}_2}\right)$ with $K/\mathbb{Q}_2$ the unique unramified extension of $\mathbb{Q}_2$.

12.2 The lattices are free, so by induction we reduce to the one-dimensional case, which is simply the statement that $\widehat{R}_{\mathfrak{p}} \cap F = R_{\mathfrak{p}} \subseteq \widehat{F}_{\mathfrak{p}}$ and follows since $R_{\mathfrak{p}} = \{x \in F : v(x) \geq 0\}$.

11.5 Take $t = \pm q \prod_{p \in \Sigma \setminus \{\infty\}} p^{\operatorname{ord}_p(t_p)}$. Select the prime q to satisfy congruences to ensure that the conditions hold. See [[Cassels, Corollary to Theorem 6.5.1]]

14.2 There exists nonzero $r \in I = \mathcal{O}x$ so $1 = (y/r)x$ for some $y \in \mathcal{O}$ and hence $x \in B^\times$.

14.4 Reduce to the local case; the result follows from Paragraph 14.5.7. Consider the connecting ideal $I = \mathcal{O}\mathcal{O}'$: clearly $\mathcal{O} \subseteq \mathcal{O}_{\mathrm{L}}(I)$ so equality holds since $\mathcal{O}$ is maximal. [[Or use hereditary?]]

14.12 This exercise is due to Kaplansky [Kap69]. We compute that

$$\mathcal{O}_{\mathrm{L}}(I) = \begin{pmatrix} R & R & (a) \\ (a) & R & (a^2) \\ R & R & R \end{pmatrix} \quad \text{and} \quad \mathcal{O}_{\mathrm{R}}(I) = \begin{pmatrix} R & R & R \\ R & R & R \\ (a^2) & (a^2) & R \end{pmatrix}$$

and

$$I^{-1} = \begin{pmatrix} R & R & R \\ R & R & R \\ (a) & R & (a^2) \end{pmatrix}$$

has $I^{-1}I = \mathcal{O}_{\mathrm{R}}(I)$ but

$$II^{-1} = \begin{pmatrix} (a) & R & (a) \\ (a) & R & (a^2) \\ R & R & R \end{pmatrix} \neq \mathcal{O}_{\mathrm{L}}(I).$$

21.1 First compute all elements in $\mathcal{O}$ of norm 2, then show the product of any two of these elements belongs to $2\mathcal{O}$.

Appendix B

Orders of small class number

Etc.

Bibliography

[Alb72] A. A. Albert, *Tensor products of quaternion algebras*, Proc. Amer. Math. Soc. **35** (1972), no. 1, 65–66.

[Alt89] Simon L. Altmann, *Hamilton, Rodrigues, and the quaternion scandal*, Math. Magazine **62** (1989), no. 5, 291–308.

[AO05] Simon L. Altmann and Eduardo L. Ortiz, eds., *Mathematics and social utopias in France: Olinde Rodrigues and his times*, Amer. Math. Society, Providence, RI, 2005.

[Art26] Emil Artin, *Zur Theorie der hyperkomplexen Zahlen*, Abh. Math. Sem. Hamburgischen Univ. **5** (1926), 251-260.

[Art50] Emil Artin, *The influence of J. H. M. Wedderburn on the development of modern algebra*, Bull. Amer. Math. Soc. **56** (1950), no. 1, 65–72.

[Asl96] Helmer Aslaksen, *Quaternionic determinants*, Math. Intelligencer **18** (1996), no. 3, Summer 1996, 57–65.

[Bae02] John C. Baez, *The octonions*, Bull. Amer. Math. Soc. (N.S.) **39** (2002), 145–205; *Errata for "The octonions"*, Bull. Amer. Math. Soc. (N.S.) **42** (2005), 213.

[BM58] R. Bott and J. Milnor, *On the parallelizability of the spheres*, Bull. Amer. Math. Soc. **64** (1958), 87–89.

[Bre10] Matej Brešar, *An elementary approach to Wedderburn's structure theory*, Exposition. Math. **28** (2010), 79–83.

[Bro61] Bancroft H. Brown, *Mathematics at Dartmouth: 1769–1961*, Dedicatory Conference, Albert Bradley Center for Mathematics, Dartmouth College, Hanover, November 3, 1961.

[Brz95] Juliusz Brzezinski, *Definite quaternion orders of class number one*, J. Théorie Nombres Bordeaux **7** (1995), 93–96.

[Brz83] Julius Brzezinski, *On orders in quaternion algebras*, Comm. Algebra **11** (1983), no. 5, 501-522.

[Cay1845] Arthur Cayley, *On Jacobi's elliptic functions, in reply to the Rev. B. Bronwin; and on quaternions*, Phil. Mag. **26** (1845), 208–211.

[Cli1878] William K. Clifford, *Applications of Grassmann's extensive algebra*, Amer. Jour. Math. **1** (1878), 350–358.

[CS03] John H. Conway and Derek A. Smith, *On quaternions and octonions: their geometry, arithmetic, and symmetry*, A. K. Peters, Ltd., Natick, MA, 2003.

[Cox89] David A. Cox, *Primes of the form $x^2 + ny^2$: Fermat, class field theory, and complex multiplication*, John Wiley & Sons, Inc., New York, 1989.

[Cro64] Michael J. Crowe, *A history of vector analysis*, Mineola, NY, Dover Press, 1964.

[Dic03] Leonard Eugene Dickson, *Definitions of a linear associative algebra by independent postulates*, Trans. Amer. Math. Soc. **4** (1903), no. 1, 21–26.

[Dic12] Leonard Eugene Dickson, *Linear algebras*, Trans. Amer. Math. Soc. **13** (1912), no. 1, 59–73.

[Dic14] Leonard Eugene Dickson, *Linear associative algebras and abelian equations*, Trans. Amer. Math. Soc. **15** (1914), 31–46.

[Dic19] L. E. Dickson, *On quaternions and their generalization and the history of the eight square theorem*, Ann. Math. (2) **20** (1919), no. 3, 155–171.

[Dic23] L. E. Dickson, *Algebras and their arithmetics*, Univ. of Chicago Press, Chicago, 1923.

[Die48] Jean Dieudonné, *Sur les groupes classiques*, Actualitiés scientifique et industrielles, no. 1040, Paris, Hermann, 1948.

[Die53] Jean Dieudonné, *On the structure of unitary groups (II)*, Amer. J. Math. **75** (1953), 665–678.

[DK95] Adel Diek and R. Kantowski, *Some Clifford algebra history*, *Clifford Algebras and Spinor Structures*, Math. Appl. **321** (1995), 3–12.

[Eic53] Martin Eichler, *Quadratische Formen und orthogonale Gruppen*, Grundlehren Math. Wiss., vol. 63, Springer, New York, 1974.

[FD93] Benson Farb and R. Keith Dennis, *Noncommutative algebra*, Grad. Texts in Math., vol. 144, Springer-Verlag, New York, 1993.

[FS07] Della D. Fenster and Joachim Schwermer, *Beyond class field theory: Helmut Hasse's arithmetic in the theory of algebras in 1931*, Arch. Hist. Exact Sci. **61** (2007), 425–456.

[Fro1878] F.G. Frobenius, *Über lineare Substitutionen und bilineare Formen*, J. Reine Angew. Math. **84** (1878), 1–63.

[vzGG03] Joachim von zur Gathen and Jürgen Gerhard, *Modern computer algebra*, 2nd edition, Cambridge University Press, Cambridge, 2003.

[Gau00] Carl Friedrich Gauss, *Mutation des Raumes*, in Carl Friedrich Gauss Werke, Band 8, König. Gesell. Wissen., Göttingen, 1900, 357–361.

[Gras1862] Hermann Grassmann, *Die Ausdehnungslehre. Vollstaändig und in strenger Form begründet* [*Extension theory*], Berlin, Wiegand, 1862.

[Grav1882] Robert Perceval Graves, *Life of Sir William Rowan Hamilton*, Volume I, Dublin University Press, 1882.

[Grav1885] Robert Perceval Graves, *Life of Sir William Rowan Hamilton*, Volume II, Dublin University Press, 1885.

[Grav1889] Robert Perceval Graves, *Life of Sir William Rowan Hamilton*, Volume III, Dublin University Press, 1889.

[GL09] Benedict H. Gross and Mark W. Lucianovic, *On cubic rings and quaternion rings*, J. Number Theory **129** (2009), no. 6, 1468-1478.

[Ham1843] William R. Hamilton, *On a new species of imaginary quantities connected with a theory of quaternions*, Proc. Royal Irish Acad. **2** (1843), 424–434.

[Ham1844] William R. Hamilton, *On quaternions, or on a new system of imaginaries in algebra: copy of a letter from Sir William R. Hamilton to John T. Graves, Esq. on quaternions*, Philosophical Magazine **25** (1844), 489–495.

[Ham1853] W. R. Hamilton, *Lectures on quaternions*, Cambridge, Cambridge University Press, 1853.

[Ham1866] W. R. Hamilton, *Elements of quaternions*, Cambridge, Cambridge University Press, 1866.

[Ham1899] *Hamilton's quaternions*, review of *Elements of quaternions*, Nature, August 24, 1899, 387.

[Ham67] W. R. Hamilton, *The mathematical papers of Sir William Rowan Hamilton, Vol. III: Algebra*, eds. H. Halberstam and R. E. Ingram, Cambridge University Press, 1967.

[Han80] Thomas L. Hankins, *William Rowan Hamilton*, Johns Hopkins University Press, 1980.

[Han06] Andrew J. Hanson, *Visualizing quaternions*, Morgan Kaufmann, San Francisco, 2006.

[Hap80] Dieter Happel, *Klassifikationstheorie endlich-dimensionaler Algebren in der Zeit von 1880 bis 1920*, Enseign. Math. (2) **26** (1980), no. 1–2, 91-102.

[Har1881] A. S. Hardy, *Elements of quaternions*, Ginn, Heath, and Company, Boston, 1881.

[HFK94] John C. Hart, George K. Francis, Louis H. Kauffman, *Visualizing quaternion rotation*, ACM Trans. Graphics **13** (1994), no. 3, 256–276.

[Haw00] Thomas Hawkins, *The emergence of the theory of Lie groups: an essay in the history of mathematics*, 1869–1926, Springer, New York, 2000.

[Hur1888] Adolf Hurwitz, *Über die Komposition der quadratischen Formen von beliebig vielen Variablen*, Nach. der köbig. Gesell. Wissen. Göttingen, Math.-Physik. Klasse, 1898, 309–316.

[Jah10] Majid Jahangiri, *Generators of arithmetic quaternion groups and a Diophantine problem*, Int. J. Number Theory **6** (2010), no. 6, 1311–1328.

[Kap69] Irving Kaplansky, *Submodules of quaternion algebras*, Proc. London Math. Soc. (3) **19** (1969), 219–232

[Kar10] Max Karoubi, *K-theory, an elementary introduction*, Cohomology of groups and algebraic K-theory, Adv. Lect. Math. (ALM), vol. 12, Int. Press, Somerville, MA, 2010, 197-215.

[Ker58] M. Kervaire, *Non-parallelizability of the n-sphere for $n > 7$*, Proc. Nat. Acad. Sci. **44** (1958) 280-283.

[Kil1888] Wilhelm Killing, *Die Zusammensetzung der stetigen endlichen Transformations-gruppen*, Math. Ann. **31** (1888), 252–290; Math. Ann. **33** (1888), 1–48; Math. Ann. **34** (1889), 57–122; Math. Ann. **36** (1890), no. 2, 161–189.

[Koh96] David Kohel, *Endomorphism rings of elliptic curves over finite fields*, Ph.D. thesis, University of California, Berkeley, 1996.

[Lam01] Tsit-Yuen Lam, *A first course in noncommutative rings*, 2nd. ed., Graduate texts in math., vol. 131, Springer-Verlag, New York, 2001.

[Lam02] Tsit-Yuen Lam, *On the linkage of quaternion algebras*, Bull. Belg. Math. Soc. **9** (2002), 415–418.

[Lam03] Tsit-Yuen Lam, *Hamilton's quaternions*, Handbook of algebra, vol. 3, North Holland, Amsterdam, 2003, 429–454.

[Lam05] Tsit-Yuen Lam, *Introduction to quadratic forms over fields*, Graduate studies in math., vol. 67, American Mathematical Society, Providence, 2005.

[Lan67] C. Lanczos, William Rowan Hamilton—An appreciation, *Amer. Scientist* **55** (1967), 129–143.

[Lem11] Stefan Lemurell, *Quaternion orders and ternary quadratic forms*, 2011, `arXiv:1103.4922`.

[Lev13] Alex Levin, *On the classification of algebras*, M.Sc. thesis, University of Vermont, 2013.

[Lew06] David W. Lewis, *Quaternion algebras and the algebraic legacy of Hamilton's quaternions*, Irish Math. Soc. Bull. **57** (2006), 41–64.

[MR03] Colin Maclachlan and Alan W. Reid, *The arithmetic of hyperbolic* 3-*manifolds*, Grad. Texts in Math., vol. 219, Springer-Verlag, New York, 2003.

[Mat69] Hideya Matsumoto, *Sur les sous-groupes arithmétiques des groupes semi-simples déployés*, Ann. Sci. École Norm. Sup. (4) **2** (1969), 1-62.

[Max1869] J. C. Maxwell, *Remarks on the mathematical classification of physical quantities*, Proc. London Math. Soc. **3** (1869), 224–232.

[May66] Kenneth O. May, *The impossibility of a division algebra of vectors in three dimensional space*, Amer. Math. Monthly **73** (1966), no. 3, 289–291.

[Mil] J.S. Milne, *Class field theory*, `http://www.jmilne.org/math/CourseNotes/cft.html`.

[1] Eliakim Hastings Moore, *General analysis*, Part I, Memoirs Amer. Phil. Soc., vol. 1, Amer. Phil. Soc., Philadelphia, 1935.

[Neu99] Jürgen Neukirch, *Algebraic number theory*, Grundlehren Math. Wiss., vol. 322, Springer-Verlag, Berlin, 1999.

[ÓCa00] Fiacre Ó Cairbre, *William Rowan Hamilton (1805–1865), Ireland's greatest mathematician*, Ríocht na Midhe (Meath Archaeological and Historical Society) **11** (2000), 124–150.

[ÓCa10] Fiacre Ó Cairbre, *Twenty years of the Hamilton walk*, Irish Math. Soc. Bulletin **65** (201), 33–49.

[O'Do83] Sean O'Donnell, *William Rowan Hamilton*, Boole Press Limited, 1983.

[Pei1882] Benjamin Peirce, *Linear associative algebra*, Amer. J. Math. **4** (1881), no. 1–4, 97-229.

[Pie82] Richard S. Pierce, *Associative algebras*, Springer–Verlag, New York, 1982.

[Puj12] Jose Pujol, *Hamilton, Rodrigues, Gauss, quaternions, and rotations: a historical reassessment*, Comm. Math. Anal. **13** (2012), no. 2, 1–14.

[Rei03] Irving Reiner, *Maximal orders*, London Math. Soc. Monogr. (N.S.), vol. 28, Clarendon Press, Oxford University Press, Oxford, 2003.

[Rod1840] Olinde Rodrigues, *Des lois geometriques qui regissent les déplacements d'un système solide dans l'espace, et la variation des coordonnées provenant de ses déplacements considérés indépendamment* *des causes qui peuvent les produire*, J. de Mathematiques Pures et Appliqués **5** (1840), 380–440.

[Schu88] John Schue, *The Wedderburn theorem of finite division rings*, American Math. Monthly **95** (1988), no. 5, 436–437.

[Shi71] Goro Shimura, *Introduction to the arithmetic theory of automorphic functions*, Kanô Memorial Lectures, no. 1, Publications of the Mathematical Society of Japan, no. 11, Iwanami Shoten, Tokyo, Princeton University Press, Princeton, 1971.

[Ste96] *1996 Steele Prizes*, Notices Amer. Math. Soc. **43** (1996), no. 11, 1340–1347.

[Sim10] Peter Simons, *Vectors and beyond: geometric algebra and its philosophical significance*, dialectica **63** (2010), no. 4, 381–395.

[SHT99] Viggo Stoltenberg-Hansen and John V. Tucker, Computable rings and fields, *Handbook of computability theory*, ed. Edward R. Griffor, North-Holland, Amsterdam, 1999, 336–447.

[Syl1883] J. J. Sylvester, *Lectures on the principles of universal algebra*, American J. Math. **6** (1883–1884), no. 1, 270–286.

[Tai1890] P. G. Tait, *An elementary treatise on quaternions*, 3rd ed., Cambridge University Press, Cambridge, 1890.

[Tho10] Silvanus P. Thompson, *The Life of Lord Kelvin, Baron Kelvin of Largs*, Vol. II, Macmillan, London, 1910,

[vdB60] F. van der Blij, *History of the octaves*, Simon Stevin **34** (1960/1961), 106-125.

[vdW76] B. L. van der Waerden, *Hamilton's discovery of quaternions*, Math. Magazine **49** (1976), 227–234.

[vPr68] Paul van Praag, *Une caractérisation des corps de quaternions*, Bull. Soc. Math. Belgique **10** (1968), 283–285.

[vPr02] Paul van Praag, *Quaternions as reflexive skew fields*, Adv. Appl. Clifford Algebr. **12** (2002), no. 2, 235–249.

[Vig80] Marie-France Vignéras, *Arithmétique des algèbres de quaternions*, Lecture Notes in Math., vol. 800, Springer, Berlin, 1980.

[Voi11a] John Voight, *Characterizing quaternion rings over an arbitrary base*, J. Reine Angew. Math. **657** (2011), 113–134.

[Voi11b] John Voight, *Rings of low rank with a standard involution*, Illinois J. Math. **55** (2011), no. 3, 1135–1154.

[Voi13] John Voight, *Identifying the matrix ring: algorithms for quaternion algebras and quadratic forms*, Quadratic and higher degree forms, Developments in Math., vol. 31, Springer, New York, 2013, 255–298.

[Wed08] J.H. Maclagan Wedderburn, *On hypercomplex numbers*, Proc. London Math. Soc. **2** (1908), vol. 6, 77-118.

[Wei13] Charles A. Weibel, *The K-book: An introduction to algebraic K-theory*, Grad. Studies in Math., vol. 145, Amer. Math. Soc., Providence, 2013.

[Wes] Tom Weston, *Lectures on the Dirichlet class number formula for imaginary quadratic fields*.

[Wil09] Robert A. Wilson, *The finite simple groups*, Grad. Texts in Math., Springer-Verlag, London, 2009.