INSTITUTO NACIONAL DE MATEMÁTICA PURA
E APLICADA

DOCTORAL THESIS

On the Geometry of Semi-Arithmetic Riemann Surfaces

Gregory Cosac

impa

Rio de Janeiro
November, 2021

INSTITUTO NACIONAL DE MATEMÁTICA PURA
E APLICADA

On the Geometry of Semi-Arithmetic Riemann Surfaces

Author:
Gregory Cosac

Advisor:
Mikhail Belolipetskiy

A thesis submitted in fulfillment of the requirements for the degree of Doctor in Philosophy in Mathematics to the Posgraduate Program in Mathematics at Instituto Nacional de Matemática Pura e Aplicada.

Rio de Janeiro
November, 2021

To my mother, Laïs Cosac.

CONTENTS

ABSTRACT

In this thesis, we study geometric aspects of semi-arithmetic Riemann surfaces by means of number theory and hyperbolic geometry. First, we show the existence of infinitely many semi-arithmetic Riemann surfaces with arbitrarily small systole. Furthermore, this leads to a construction, for each genus $g \geq 2$, of infinite families of semi-arithmetic surfaces with pairwise distinct invariant trace fields, giving a negative answer to a conjecture of B. Jeon. Finally, we produce a bound on the number of automorphisms of non-semi-arithmetic surfaces and, in particular, of surfaces with non-integral traces.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Prof. Mikhail Belolipetsky for his unwavering support, invaluable advice and infinite patience. None of this would have been possible without him.

I would also like to thank Profs. Inkang Kim, Alexander Kolpakov, Hossein Movasati, Gisele Teixeira and Mikhail Verbitsky for being part of the Examination Committee, for commenting on my work, and for helping me improve this thesis.

I am grateful to IMPA for providing an exciting research environment that brought me joy and allowed me to prosper. The combination of its robust infrastructure, colossal library and idyllic landscape made IMPA a second home for me. I would also like to express my gratitude to the very friendly and helpful staff members: Andréia, Josenildo, Isabel, Kênia, André Valério, among others. I am especially grateful to the institute for being understanding when I found myself in need of a second chance. In that regard I would like to thank Prof. Roberto Imbuzeiro, then Head of Studies, for his patience and advice.

During the course of my studies at IMPA, I had the good fortune of making many friends, without whom I would not have been able to complete this thesis. This paragraph is dedicated to them. I sincerely thank my long-time friend, Jamerson Bezerra, whose friendship and support helped me through the most difficult parts of my postgraduate education. I was privileged to have enjoyed the company of the best classmates one could ever hope for, Alcides Carvalho, Sandoel Vieira and Thomas Spier, who were constantly encouraging me and helping me to grow. I am very grateful to my academic siblings Plinio Murillo and Gisele Teixeira for having paved the way. I am especially grateful to my academic brother, coauthor and good friend Cayo Dória, who taught me so much. My flatmate for nearly four years and friend for life, Ivan Passoni, I have to thank for all the healthy nonsensical debates and all the 3am burgers. I thank the guidance

of my godfather in IMPA, Fernando Lenarduzzi, and my good friends Felippe "Belezinha" Guimarães and Vanderson Lima. My friend Dmitrii Korshunov, I thank for the pleasant conversations in etymology and History. I thank José Eduardo Garcez for all we exchanged over the years about life, gym and procrastination. I am grateful to João Pedro "Ismailov" Ramos for his enduring friendship and hospitality. I am also grateful to my dear friends Mateus Sousa, Felipe Gonçalves and Hugo Araújo for their helpful advice and for setting a great example. I thank Mateus M. de Melo for the many conversations we shared since we first met in Topology of Manifolds. I also extend my gratitude to Pedro Campos, Rafael Ponte, Renan Santos, Daniel Yukimura, Luiz Paulo Freire, Daniel Ungaretti, Vitor Alves, Ermerson Araujo, Roberto Viveros, Shangjie Yang, Mauricio Poletti, Hudson Lima, Miguel Jimenez, Roberto Alvarenga, Marcelo Santos, Letícia Alves, Juan José Villarreal, José Ramon, Renan Finder, Dan Ketover, Renato Dias, Zoraida Rico, João Vicente "Velho" Lopes, Leonardo Ferrari, Gabriel Araújo and many others.

I would like to thank Prof. Vincent Emery for the opportunity of visiting him in Bern towards the end of my Ph.D. programme. And, of course, my deepest thanks and appreciation to Alejandro, Livio, Levi, Julia, Kevin and Ruben, who opened their homes and took me in while the world was falling apart.

I am grateful to my family for all they have done for me, especially my mother Laïs and my brother Álvaro. My most heartfelt thanks to my late grandmother, Mathilde, who backed me up in the very beginning of this long and winding road. Even though she is not here to see this work concluded, she is definitely a part of it.

Finally, I would like to extend my gratitude to all the people that, in one way or another, helped me in my journey. Special thanks to Maria do Carmo Lins, Damião e Cátia Balbino, Mauro Tavares, César Liberato, Júlio, Maria Raimunda, José Paulo Silva and João Paulo Silva.

CHAPTER 1

INTRODUCTION

1.1 The setting

Let Γ be a Kleinian group, i.e, a discrete subgroup of $\mathrm{PSL}(2, \mathbb{C})$. The group Γ acts on the hyperbolic 3-space $\mathbb{H}^3$ by isometries, and so the quotient $\Gamma\backslash\mathbb{H}^3$ is a hyperbolic 3-orbifold. In particular, there is a well-defined notion of (hyperbolic) volume in the quotient.

We associate to Γ a number theoretic object, $\mathbb{Q}(\operatorname{tr}\Gamma)$, called the *trace field* of Γ. It consists of the field of rational numbers $\mathbb{Q}$ with the traces of all elements of Γ adjoined. Already, an interplay between hyperbolic geometry and number theory can be observed: if $\Gamma\backslash\mathbb{H}^3$ has finite volume, then the trace field $\mathbb{Q}(\operatorname{tr}\Gamma)$ is a *number field* (i.e., a finite field-extension of $\mathbb{Q}$).

The proof of this fact is a consequence of Mostow's Rigidity Theorem, which makes it essentially a phenomenon in dimension at least 3. Indeed, the situation is drastically different in dimension 2.

Let Γ be now a Fuchsian group, i.e, a discrete subgroup of $\mathrm{PSL}(2, \mathbb{R})$, so that it acts on $\mathbb{H}^2$. Assume, for simplicity, that Γ is free of torsion and cocompact, so $\Gamma\backslash\mathbb{H}^2$ is a closed hyperbolic surface of genus $g \geq 2$. In other words, the group Γ determines a hyperbolic structure on the underlying topological surface S_g. The main difference in dimension 2 is that this structure can be continuously deformed. Indeed, if we perturb the inclusion homomorphism $\iota : \Gamma \hookrightarrow \mathrm{PSL}(2, \mathbb{R})$, we obtain non-equivalent hyperbolic structures on S_g. The space T_g of all (marked) hyperbolic structures on S_g, up to isometries (isotopic to the identity), is known as the *Teichmüller space* of S_g. Similarly, the space of all conjugacy classes of representations of Γ into $\mathrm{PSL}(2, \mathbb{R})$ is called the *Teichmüller space* of Γ, and is denoted by $\mathrm{Teich}(\Gamma)$. These two spaces are in one-to-one correspondence with each other and are often considered as two different points of view of the same space. They are, in fact, homeomorphic, when endowed with the appropriate topology.

The Teichmüller space $\mathrm{Teich}(\Gamma)$ is a manifold of real dimension $6g-6$ (homeomorphic to an open ball). One may associate to each point in $\mathrm{Teich}(\Gamma)$ a well-defined trace field since traces are invariant under conjugation. At most countably many points in $\mathrm{Teich}(\Gamma)$ can have a trace field that is a finite extension of $\mathbb{Q}$ (a number field), as may be seen after parametrising $\mathrm{Teich}(\Gamma)$ by finitely many trace functions. In particular, most points in $\mathrm{Teich}(\Gamma)$ have transcendental trace fields. All of these points, however, have the same finite coarea $4\pi(g-1)$, in contrast with the case of cofinite Kleinian groups mentioned above.

Among the most studied invariants used in order to understand the hyperbolic structure of a generic surface $X \in \mathrm{T}_g$, one finds the diameter $\mathrm{diam}(X)$, the spectral gap $\lambda_1(X)$ of the Laplace-Beltrami operator defined on X, the isometry group $\mathrm{Isom}(X)$, and the systole $\mathrm{sys}(X)$, defined as the minimal length of a closed geodesic of S.

Points of T_g that satisfy some arithmetic property tend to respond to a greater range of techniques due to their extra structure. This makes them somewhat easier to understand, as well as a plentiful source of examples. Besides, they often manifest extremal properties. To mention a few:

- The cofinite Fuchsian group of minimal coarea is the triangular group $(2,3,7)$ which is *arithmetic* by [51]. The closed Riemann surfaces of genus g whose automorphism group attains the maximal cardinality of $84(g-1)$ are, therefore, also arithmetic. Moreover, for *non-arithmetic* surfaces, the cardinality of the automorphism group is at most $\frac{156}{7}(g-1)$, as proved in [3].

- For any Riemann surface S, a simple geometric argument gives that:

$$\mathrm{sys}(S) \leq 2\log(g(S)) + A, \tag{1.1.1}$$

  where $g(S)$ denotes the genus of S and $A > 0$ is some absolute constant ([8, Lemma 5.2.1]). Buser-Sarnak [9] and Katz-Schaps-Vishne [29] proved that a sequence of *congruence coverings* of any closed arithmetic Riemann surface satisfy the following logarithmic systolic growth:

$$\mathrm{sys}(S_i) \gtrsim \frac{4}{3}\log(g(S_i)),$$

  In particular, the logarithmic upper bound (1.1.1) is optimal (up to a constant).

- While systoles of arithmetic surfaces can be very large, there exists an explicit lower bound in terms of the area. Indeed, Belolipetsky proved in [4] the following inequality:

$$\mathrm{sys}(S) \geq C_1 \left(\frac{\log\log\log \mathrm{area(S)}^{\mathrm{C_2}}}{\log\log \mathrm{area(S)}^{\mathrm{C_2}}} \right)^3 ,$$

  for any arithmetic Riemann surface S, where $C_1, C_2 > 0$ are universal constants. In fact, the *short geodesic conjecture* predicts the existence of a universal lower bound for the systole of any arithmetic Riemann surface.

- The *commensurability class* of an arithmetic Riemann surface is determined by its Laplace-Beltrami spectrum [46].

It becomes clear the desirability to research arithmetic properties on Fuchsian groups and to better understand their geometrical implications.

In addition to the trace field already introduced, another important arithmetic object often associated to a Fuchsian group Γ is the *quaternion algebra* $A_0\Gamma$ defined over $\mathbb{Q}(\mathrm{tr}\,\Gamma)$ as:

$$A_0\Gamma = \left\{ \sum_i a_i \gamma_i \mid a_i \in \mathbb{Q}(\mathrm{tr}\,\Gamma),\ \gamma_i \in \Gamma \right\},$$

where the sums are all finite. The algebra $A_0\Gamma$ is a central simple algebra of dimension 4.

Before we continue, let us make a technical adjustment. Two groups Γ_1 and Γ_2 are said to be *commensurable* if $\Gamma_1 \cap \Gamma_2$ has finite index both in Γ_1 and in Γ_2. The trace field and associated quaternion algebra are not invariant under commensurability, so we make a small adjustment: given a Fuchsian group Γ, let $\Gamma^{(2)}$ be the (finite index) subgroup generated by the square of every element in Γ. We then define the *invariant trace field* of Γ to be $k\Gamma := \mathbb{Q}(\mathrm{tr}\,\Gamma^{(2)})$, and the *invariant quaternion algebra* of Γ to be $A\Gamma := A_0\Gamma^{(2)}$. These are now, as their names suggest, invariants of the commensurability class of Γ.

We list the three most common arithmetic properties that one may require of a Fuchsian group Γ:

(i) The invariant trace field $k\Gamma$ is a *totally real* number field;

(ii) The invariant quaternion algebra $A\Gamma$ is *admissible*, meaning that $A\Gamma \otimes_{\mathbb{Q}} \mathbb{R} \cong M_2(\mathbb{R}) \times K$ where $M_2(\mathbb{R})$ is the algebra of 2×2 matrices with real coefficients and K is compact;

(iii) The traces of elements of Γ are algebraic integers.

Definition 1.1.1. We say a Fuchsian group Γ is:

- *arithmetic* if it satisfies (i)-(iii) (see Theorem 5.3.11);
- *quasi-arithmetic* if it satisfies (ii) (note that (ii) implies (i));
- *semi-arithmetic* if it satisfies (i) and (iii).

Note that each of these definitions are invariant under conjugacy and therefore it makes sense to say that a point in $\mathrm{Teich}(\Gamma)$ is (semi-, quasi-) arithmetic. Also, we say that a Riemann surface $X = \Gamma\backslash\mathbb{H}^2$ is (semi-, quasi-) arithmetic according as to Γ is (semi-, quasi-) arithmetic, respectively.

Arithmetic Fuchsian groups have been extensively studied since the 1970s and are known to enjoy several interesting properties, as indicated by the list above. However, they are somewhat rigid in the sense that there are at most finitely many arithmetic points in each $\mathrm{Teich}(\Gamma)$ ([7, Theorem 8.2]).

Quasi-arithmetic groups are more often studied in higher dimensions (for example, [5] and [16]), although recently some interesting examples of quasi-arithmetic groups of isometries of $\mathbb{H}^2$ where described in [15].

Finally, semi-arithmetic groups were formally introduced around the year 2000 by Schmutz Schaller and Wolfart ([49]). These Fuchsian groups satisfy weaker properties than arithmetic groups do but, on the other hand, they are much more abundant. Also, they can be embedded as infinite index subgroups of arithmetic lattices in semi-simple Lie groups of higher rank ($\mathrm{PSL}(2,\mathbb{R})^r$). Either as cofinite Fuchsian groups or as infinite index subgroups of arithmetic lattices, semi-arithmetic groups have demonstrated to be a pertinent topic of investigation. They are the main object of interest of the present thesis.

1.2 The results in this Thesis

1.2.1 Constructing semi-arithmetic Fuchsian groups

We generate semi-arithmetic Fuchsian groups from reflections across the sides of certain hyperbolic polygons, namely, the trirectangle and the right-angled hexagon. The main result of this thesis is the following:

Theorem A. *For any $g \geq 2$ there exists a length function $\ell_\alpha : \mathrm{T}_g \to \mathbb{R}$ such that*

$$\{\ell_\alpha(S) \mid S \in \mathrm{T}_g \text{ is semi-arithmetic}\}$$

is dense on the set of positive real numbers.

Here a *length function* is a function that associates to each hyperbolic structure $X \in \mathrm{T}_g$ the length of the unique geodesic contained in the free homotopy class of a fixed non-trivial closed curve on the underlying topological surface.

We recall that, for a hyperbolic element $\gamma \in \mathrm{PSL}(2, \mathbb{R})$, its translation length $\ell(\gamma)$ is related to its trace by the elementary formula:

$$|\mathrm{tr}\, \gamma| = 2 \cosh \frac{\ell(\gamma)}{2}.$$

With this in mind, we generate a reflection group from the right-angled hyperbolic hexagons with three non-adjacent sides of length $a > 0$. Its index 2 subgroup Γ_a of orientation-preserving isometries contains an element T_a with $|\mathrm{tr}\, T_a| = 2 \cosh a$. By picking the appropriate parameter a, we can guarantee that Γ_a is semi-arithmetic. Moreover, a can be chosen from a dense subset in $[0, +\infty)$.

Finally, we use group theoretic tools (the Reidemeister-Schreier rewriting process) to find, for each $g \geq 2$, a surface group of genus g contained in Γ_a that still contains the distinguished element T_a.

As a consequence of Theorem A, we see that there exist infinitely many semi-arithmetic surfaces in each genus $g \geq 2$, unlike the arithmetic case. Also, for any fixed genus $g \geq 2$, there are semi-arithmetic surfaces with arbitrarily small systole. In particular, the *short geodesic conjecture*, which is an open conjecture about arithmetic surfaces, could not possibly hold for semi-arithmetic surfaces. Moreover, as proved in [14], it also follows from Theorem A that the set $\{\mathrm{sys}(X) \mid X \text{ is a closed semi-arithmetic Riemann surface}\}$ is dense in $[0, +\infty)$.

1.2.2 Surfaces with integral traces

When $X = \Gamma \backslash \mathbb{H}^2$ and the traces of elements of Γ are algebraic integers, we say that X has integral traces. One application of Theorem A is as follows:

Theorem B. *Every totally real number field of prime degree at least 3 is realised as the invariant trace field of a genus g semi-arithmetic Riemann surface, for any $g \geq 2$.*

Theorem B settles the question as to whether only finitely many trace fields (and quaternion algebras) could be realised as the invariant trace field of a surface with integral traces of a fixed genus $g \geq 2$.

In particular, for each genus $g \geq 2$, there are semi-arithmetic surfaces with invariant trace fields of arbitrarily large degree.

1.2.3 Automorphism group of surfaces with non-integral traces

The classical Hurwitz bound states that, for a closed Riemann surface X_g of genus g, the order of its automorphism group $\mathrm{Aut}(X_g)$ is at most $84(g-1)$. In [3], Belolipetsky proves that this bound drops to $\frac{156}{7}(g-1)$ for non-arithmetic surfaces, and that this number is attained for infinitely many g. Finally, for non-semi-arithmetic Riemann surfaces and, in particular, for surfaces with non-integral traces, we have the following:

Theorem C. *The order of the automorphism group of a non-semi-arithmetic Riemann surface X_g of genus $g \geq 2$ satisfies the following bound:*

$$|\mathrm{Aut}(X_g)| \leq 12(g-1).$$

Moreover, this bound is attained in every genus $g \geq 2$.

1.3 Outline of the Thesis

This thesis is organised as follows:

Chapter 2 comprises the background material on Number Theory. We begin with a description of the ring of integers of a number field: unique factorisation into prime ideals, integral basis and the structure of its group of units (Dirichlet's Unit Theorem). Next, we introduce valuations and study valued fields and local and global fields. The topic of extending of valuations to field extensions is discussed in §§2.3.9 and is complemented by Appendix A, on Krull valuations.

In Chapter 3, we introduce quaternion algebras, building up on the material presented in the previous chapter. Quaternion algebras over local fields are fully described (Propositions 3.4.1 and 3.4.2 and Theorem 3.4.5) and the general theorem on classification of quaternion algebras (Theorem 3.4.11) is stated (without proof). Finally, we introduce orders in quaternion algebras. One could say that orders play the role of the ring of integers in number fields. They are fundamental in the definition of our main object of study.

Chapter 4 studies Fuchsian groups acting on the hyperbolic 2-space, where certain fundamental domains for these actions are described. We discuss Poincaré's Theorem and the presentation of finitely generated Fuchsian groups. Finally, the concept of the space of deformation of a Fuchsian group (its Teichmüller space) is introduced.

Chapter 5 is where the arithmetic of Chapters 2 and 3 is combined with the geometry of Chapter 4. The invariant trace field and quaternion algebra associated to a

Kleinian group are introduced. Arithmetic groups are defined and characterised by Takeuchi's theorem (Theorem 5.3.12). The invariant trace field and quaternion algebra are proved to be complete commensurability invariants for arithmetic groups (Theorem 5.3.13). At last, the central object of this thesis is defined: semi-arithmetic Fuchsian groups. Subsection §§5.4.3 comprises a list of properties of these groups that have been researched in recent years, including the contribution of the present thesis.

In Chapter 6, Theorems A and B are restated and proved.

Chapters 2, 3 and 4 were designed to be as self-contained as possible and might be somewhat too inclusive for our purposes. Nevertheless, they were kept unaltered in the spirit of exposition. None of the material therein is original (except the typos).

Theorems A and B appear in the following article:

> *Closed geodesics on semi-arithmetic Riemann surfaces* (with C. Dória), to appear in Mathematical Research Letters. arXiv:2004.13683 (2020). 32 pages.

CHAPTER 2

NUMBER THEORY

2.1 Integrality

Definition 2.1.1. Let $A \subset B$ be an extension of rings. An element b of B is said to be *integral* over A if it is the root of a *monic* polynomial with coefficients in A, i.e., if there exist elements $a_0, \dots, a_{n-1}$ in A such that:

$$b^n + a_{n-1}b^{n-1} + \cdots + a_1 b + a_0 = 0.$$

We say that B is integral over A in case every element of B is integral over A.

The integral closure of A in B, denoted $\overline{A}$, is the set of all elements in B that are integral over A. The fact that $\overline{A}$ is a ring is a consequence of the following characterisation of integrality, analogous to the field-theoretic notion of being algebraic over a field:

Proposition 2.1.2. *The elements $b_1, \dots, b_m \in B$ are integral over A if and only if the ring $A[b_1, \dots, b_m]$ viewed as an A-module is finitely generated.*

Proof. We first prove necessity: for $m = 1$, let $b \in B$ be integral over A, so that b satisfies:

$$b^n = -(a_{n-1}b^{n-1} + \cdots + a_1 b + a_0). \tag{2.1.1}$$

Multiplying by b on both sides and inserting the expression for b^n given by (2.1.1), one obtains an expression for b^{n+1} in terms of $\{1, b, \dots, b^{n-1}\}$. By repeating this argument, we see that any power of b may be written as a linear combination of $\{1, b, \dots, b^{n-1}\}$ with coefficients in A, and thus $A[b]$ is finitely generated. The argument is then concluded by induction on m.

Conversely, suppose $A[b_1,\dots,b_m]$ is finitely generated by the set $\{\alpha_1,\dots,\alpha_k\}$ and let b be an element in $A[b_1,\dots,b_m]$. For each $i=1,\dots,k$, there exist $a_{i1},\dots,a_{ik}\in A$ such that

$$b\alpha_i=\sum_{j=1}^{k}a_{ij}\alpha_j.$$

This system of equations can be written in the form of matrices as follows:

$$\begin{bmatrix} b & 0 & \cdots & 0\\ 0 & b & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & b\end{bmatrix}\begin{bmatrix}\alpha_1\\ \vdots\\ \\ \alpha_k\end{bmatrix}=\begin{bmatrix} a_{11} & \cdots & a_{1k}\\ \vdots & \ddots & \vdots\\ \\ a_{k1} & \cdots & a_{kk}\end{bmatrix}\begin{bmatrix}\alpha_1\\ \vdots\\ \\ \alpha_k\end{bmatrix}. \tag{2.1.2}$$

So, if we let Id_k denote de $k\times k$ identity matrix (with coefficients in A) and α the k-vector $(\alpha_1,\dots,\alpha_k)$, then (2.1.2) can be rewritten as:

$$(b\mathrm{Id}_k-(a_{ij}))\alpha=0.$$

Multiplying both sides by the classical adjoint of $(b\mathrm{Id}_k-(a_{ij}))$ yields

$$\det(b\mathrm{Id}_k-(a_{ij}))\alpha_i=0,\quad i=1,\dots,k.$$

The elements $\alpha_1,\dots,\alpha_r$ generate $A[b_1,\dots,b_m]$ and, in particular, 1. It follows that $\det(b\mathrm{Id}_k-(a_{ij}))=0$, which gives a monic polynomial of degree k and coefficients in A that has b as one of its roots. □

Corollary 2.1.3. *The integral closure of A in B, $\overline{A}=\{b\in B\mid b$ is integral over $A\}$, is a subring of B.*

Proof. Given $b_1,b_2\in\overline{A}$, it follows from the proof of Proposition 2.1.2 that any $b\in A[b_1,b_2]$ is integral over A and, in particular, so are b_1-b_2 and b_1b_2. Alternatively, one can conclude that $b\in A[b_1,b_2]$ is integral over A directly from the statement of Proposition 2.1.2, since $A[b_1,b_2,b]=A[b_1,b_2]$ is finitely generated over A. □

Corollary 2.1.4. *Let $A\subset B\subset C$ be ring extensions such that C is integral over B and B is integral over A. Then C is integral over A.*

Definition 2.1.5. As mentioned before, the set $\overline{A}$ defined above is called the *integral closure* of A in B. If $\overline{A}=A$ then A is said to be *integrally closed in B*.

If A is an integral domain, we say that A is *integrally closed* when it is integrally closed in its field of fractions.

For instance, it is an elementary fact that the ring $\mathbb{Z}$ is integrally closed. Indeed, its field of fractions is $\mathbb{Q}$ and if a/b is a rational number in its reduced form that is integral over $\mathbb{Z}$ then, after clearing denominators, a and b satisfy:

$$a^n + a_{n-1}ba^{n-1} + \cdots + a_1 b^{n-1}a + a_0 b^n = 0,$$

which implies that any divisor of b must also be a divisor of a. Since we assumed a/b to be reduced, it follows that $b = \pm 1$ and $a/b \in \mathbb{Z}$. Note that the same argument applies to any unique factorisation domain.

In what follows, we will be dealing mostly with the case where A is an integrally closed integral domain with field of fractions K, and $L \mid K$ is a finite field extension. Let B denote the integral closure of A in L. The following may be easily verified:

(i) Any element of L can be written in the form b/a, where $b \in B$ and $a \in A$. In particular, L can be recovered from B as its field of fractions.

(ii) An element of L is integral over A if and only if its minimal polynomial has coefficients in A.

We briefly recall that the *trace* and *norm* of an element $x \in L$, denoted respectively by $\operatorname{Tr}_{L|K}(x)$ and $\mathrm{N}_{L|K}(x)$, are defined to be the trace and determinant of the K-linear transformation $T_x : L \to L,\ \alpha \mapsto x\alpha$. These objects satisfy the following well-known properties:

Proposition 2.1.6 (Properties of the Trace and Norm)**.**

1. *The maps* $\operatorname{Tr}_{L|K} : L \to K$ *and* $\mathrm{N}_{L|K} : L^* \to K^*$ *are (group) homomorphisms.*

2. *For extensions* $K \subset L \subset M$ *one has that:*

$$\operatorname{Tr}_{M|K} = \operatorname{Tr}_{L|K} \circ \operatorname{Tr}_{M|L}, \qquad \mathrm{N}_{M|K} = \mathrm{N}_{L|K} \circ \mathrm{N}_{M|L}$$

3. *If* $L \mid K$ *is separable, let* $\sigma : L \to \overline{K}$ *run over all the* K*-embeddings of* L *into the algebraic closure of* K*. Then the following equalities hold:*

   (a) $f_x = \prod_\sigma (t - \sigma x)$ *where* f_x *is the characteristic polynomial of* T_x*;*

   (b) $\operatorname{Tr}_{L|K}(x) = \sum_\sigma \sigma x$*;*

   (c) $\mathrm{N}_{L|K}(x) = \prod_\sigma \sigma x$*.*

Proof. See, for example, [42]. □

It is worth noting that $\operatorname{Tr}_{L|K}(x)$ and $\mathrm{N}_{L|K}(x)$ appear as coefficients of the characteristic polynomial f_x of T_x, namely, if

$$f_x(t) = \det(t\mathrm{Id} - T_x) = t^n - a_{n-1}t^{n-1} + \cdots + (-1)^n a_0,$$

then $\operatorname{Tr}_{L|K}(x) = a_{n-1}$ and $\mathrm{N}_{L|K}(x) = a_0$. In particular, equalities (3b) and (3c) in the proposition above follow directly from (3a) in the light of this observation.

2.2 The ring of integers

2.2.1 The ring of integers of a number field

The framework of integral extension introduced in the previous section will now be applied to the case where the integral domain A is $\mathbb{Z}$. As pointed out before, the ring $\mathbb{Z}$ is integrally closed in its field of fractions $\mathbb{Q}$. If we consider a finite field extension of $\mathbb{Q}$, say K, then, by taking the integral closure of $\mathbb{Z}$ in K, we obtain a larger ring $\mathcal{O}_K$, which will be the object of study of this section. First, let us introduce some terminology.

Definition 2.2.1. We say that a complex number $\alpha \in \mathbb{C}$ is an *algebraic number* if it is a root of a polynomial with rational coefficients. We say it is an *algebraic integer*, if it is integral over $\mathbb{Z}$ according to Definition 2.1.1, i.e., if it is a root of a monic polynomial with integer coefficients. Note that, by eliminating denominators, every algebraic number is the root of a polynomial with integer coefficients, but this polynomial is not necessarily monic.

A *number field* K is a finite field extension of $\mathbb{Q}$. In particular, K is an algebraic extension of $\mathbb{Q}$. The *ring of integers* of K is the integral closure of $\mathbb{Z}$ in K, and is denoted by $\mathcal{O}_K$. In other words, $\mathcal{O}_K$ is the ring of all algebraic integers that lie in K.

The set of all algebraic integers is the integral closure of $\mathbb{Z}$ in $\mathbb{C}$ and, by Corollary 2.1.3, it is a subring of $\mathbb{C}$. The elements of $\mathbb{Z}$ are sometimes referred to as *rational integers*, in order to be distinguished from general algebraic integers.

The ring $\mathcal{O}_K$ enjoys some very interesting properties that are extensively studied in the field of Algebraic Number Theory. In this section, we will only explore some of its most basic features.

Firstly, note that $\mathcal{O}_K$ is clearly integrally closed, being the integral closure of $\mathbb{Z}$ in K. Indeed, it follows from Corollary 2.1.4 that the subring formed by the elements of K that are integral over $\mathcal{O}_K$, must also be integral over $\mathbb{Z}$, and are therefore contained in $\mathcal{O}_K$.

We know from Proposition 2.1.2 that the $\mathbb{Z}$-module $\mathbb{Z}[b_1, \ldots, b_k]$ is finitely generated for any finite collection of elements $b_1, \ldots, b_k \in \mathcal{O}_K$. In fact, as we shall see next, the whole ring $\mathcal{O}_K$ is finitely generated as a $\mathbb{Z}$-module and, moreover, $\mathcal{O}_K$ admits a $\mathbb{Z}$-basis.

Definition 2.2.2. An *integral basis* of $\mathcal{O}_K$ is a $\mathbb{Z}$-basis for the $\mathbb{Z}$-module (equivalently, abelian group) $\mathcal{O}_K$. It is sometimes referred to as an integral basis for K. Note that an integral basis is always finite since it is also a $\mathbb{Q}$-basis for K.

Proposition 2.2.3. *$\mathcal{O}_K$ has an integral basis of cardinality $[K : \mathbb{Q}]$.*

Proof. Let $\{\alpha_1, \ldots, \alpha_n\}$ be a $\mathbb{Q}$-basis for K. For any $c \in \mathscr{O}_K$, there exists a rational integer $N = N(c)$ such that $Nc \in \mathbb{Z}\alpha_1 + \cdots + \mathbb{Z}\alpha_n$. For example, one can take N to be the least common multiple of the denominators of the coefficients of c when expressed in the basis $\{\alpha_1, \ldots, \alpha_n\}$. The key observation here is that, if the α_i are algebraic integers (which we may assume without loss of generality, simply by eliminating denominators) we can choose a rational integer that works for every c. In other words, there exists an integer d such that $d \cdot \mathscr{O}_K \subset \mathbb{Z}\alpha_1 + \cdots + \mathbb{Z}\alpha_n$. This will be proved in Proposition 2.2.19. For now, let us assume it to be true. Then, as a subgroup of a finitely generated free abelian group, it follows that $d \cdot \mathscr{O}_K$ is a free abelian group of rank $\leq [K : \mathbb{Q}]$, and thus so is $\mathscr{O}_K$. On the other hand, as observed before, a $\mathbb{Z}$-basis for $\mathscr{O}_K$ generates K over $\mathbb{Q}$ and so $\operatorname{rank}(\mathscr{O}_K) \geq [K : \mathbb{Q}]$. □

Corollary 2.2.4. *Let $\mathfrak{a} \subset \mathscr{O}_K$ be a non-zero ideal. Then $\mathfrak{a}$ admits an integral basis of cardinality $n = [K : \mathbb{Q}]$.*

Proof. If $\{\omega_1, \ldots, \omega_n\}$ is an integral basis for $\mathscr{O}_K$, let $a \in \mathfrak{a},\ a \neq 0$. Then

$$\mathbb{Z}a\omega_1 + \cdots + \mathbb{Z}a\omega_n \subset \mathfrak{a} \subset \mathbb{Z}\omega_1 + \cdots + \omega_n.$$

Following the same reasoning as in the proof of the previous proposition, we conclude that $\mathfrak{a}$ is a free abelian group of rank n. □

Remark 2.2.5. More generally, any finitely generated $\mathscr{O}_K$-module $M \subset K$, is a free $\mathbb{Z}$-module of rank n. Indeed, take a generating set $\{\mu_1, \ldots, \mu_m\}$ of M. Since every element of K is of the form a/b where $a \in \mathscr{O}_K$ and $b \in \mathbb{Z}$, one can find a rational integer N such that $N\mu_i \in \mathscr{O}_K$ for every $i = 1, \ldots, m$ and then $N \cdot M \subset \mathscr{O}_K$. It follows that $N \cdot d \cdot M \subset d \cdot \mathscr{O}_K \subset \mathbb{Z}\alpha_1 + \cdots + \mathbb{Z}\alpha_n$, where $\alpha_1, \ldots, \alpha_n$ are as in the proof of Proposition 2.2.3. We conclude by using the same theorem for subgroups of finitely generated abelian groups as before.

Another important property of $\mathscr{O}_K$ is that, for any non-zero ideal $\mathfrak{a}$, the quotient $\mathscr{O}_K/\mathfrak{a}$ is finite. Indeed, let $a \in \mathfrak{a}$, $a \neq 0$, and $m = \mathrm{N}_{K|\mathbb{Q}}(a)$. Let $t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0 \in \mathbb{Z}[t]$ be the minimal polynomial of a. Note that $m = a_0 = -(a^n + a_{n-1}a^{n-1} + \cdots + a_1 a) \in \mathfrak{a}$ so that, in particular, $m\mathscr{O}_K \subset \mathfrak{a}$. Then $\mathscr{O}_K/m\mathscr{O}_K$ is a finitely generated abelian group and, moreover, since $m \in \mathbb{Z}$, every element of this quotient must be of finite order. It follows that $\mathscr{O}_K/m\mathscr{O}_K$ is finite and, consequently, so is $\mathscr{O}_K/\mathfrak{a}$.

Definition 2.2.6. The *(absolute) norm* of a non-zero ideal $\mathfrak{a} \subset \mathscr{O}_K$, $\mathrm{N}(\mathfrak{a})$, is defined to be the index of $\mathfrak{a}$ in $\mathscr{O}_K$, i.e.,

$$\mathrm{N}(\mathfrak{a}) = [\mathscr{O}_K : \mathfrak{a}],$$

By convention, the norm of the zero ideal is taken to be 0.

The properties enjoyed by $\mathcal{O}_K$ described so far, when put together, turn out to be quite fruitful. For this reason, a special denomination is given to integral domains bearing such qualities, according to the following definition:

Definition 2.2.7. An integral domain D is said to be a *Dedekind domain* if it satisfies the following properties:

(i) D is integrally closed;

(ii) D is Noetherian;

(iii) Every non-zero prime ideal of D is maximal.

Proposition 2.2.8. *$\mathcal{O}_K$ is a Dedekind domain.*

Proof. We have already showed that $\mathcal{O}_K$ is integrally closed. It follows from Corollary 2.2.4 that $\mathcal{O}_K$ is Noetherian. Finally, if $\mathfrak{p}$ is a non-zero prime ideal, the quotient $\mathcal{O}_K/\mathfrak{p}$ will not only be finite but also an integral domain, due to primality of $\mathfrak{p}$ and commutativity of $\mathcal{O}_K$. A finite integral domain is easily seen to be a field, whence it follows that $\mathfrak{p}$ is maximal. □

2.2.2 Factorisation into prime ideals

The ring of integers $\mathcal{O}_K$ of a number field K is not, in general, a unique factorisation domain. In other words, there may be more than one way to factor an element of $\mathcal{O}_K$ into irreducible elements, which makes the arithmetic in $\mathcal{O}_K$ fundamentally different from that of the rational integers. According to [42], Ernst Kummer's idea to overcome this failure in unique factorisation was to embed the ring $\mathcal{O}_K$ into a larger domain of "ideal numbers" where unique factorisation should hold. Something in the same spirit as the embedding of the real numbers in the larger field of complex numbers. Richard Dedekind then reinterpreted Kummer's ideas replacing the ideal numbers by ideals of $\mathcal{O}_K$, as we know them today. We will now briefly explore the factorisation of ideals of $\mathcal{O}_K$ into prime ideals. Let A be an integral domain with field of fractions K. In this section, D will always denote a Dedekind domain, unless otherwise stated.

Given ideals $\mathfrak{a}, \mathfrak{b} \subset A$, we say that $\mathfrak{a}$ divides $\mathfrak{b}$ when $\mathfrak{b} \subset \mathfrak{a}$. This definition is quite natural when one thinks of ideals of $\mathbb{Z}$.

Sum and product of ideals are defined in the customary way, namely:

$$\mathfrak{a} + \mathfrak{b} = \{a + b \mid a \in \mathfrak{a},\ b \in \mathfrak{b}\} \quad \text{and} \quad \mathfrak{a}\mathfrak{b} = \left\{ \sum_{i=1}^{k} a_i b_i \mid a_i \in \mathfrak{a},\ b_i \in \mathfrak{b}, k > 0 \right\}$$

Note that $\mathfrak{a}+\mathfrak{b}$ and $\mathfrak{a}\mathfrak{b}$ are also ideals. Furthermore, $\mathfrak{a}+\mathfrak{b}$ is the smallest ideal containing both $\mathfrak{a}$ and $\mathfrak{b}$. In other words, any ideal that divides $\mathfrak{a}$ and $\mathfrak{b}$, must also divide $\mathfrak{a}+\mathfrak{b}$. Therefore, it is only natural to say that $\mathfrak{a}+\mathfrak{b}$ is the greatest common divisor $\mathfrak{a}$ and $\mathfrak{b}$: $\gcd(\mathfrak{a},\mathfrak{b})$. In the same way, $\mathfrak{a}\cap\mathfrak{b}=\operatorname{lcm}(\mathfrak{a},\mathfrak{b})$. Observe that $\mathfrak{a}\mathfrak{b}\subset\mathfrak{a}\cap\mathfrak{b}$. We are now ready to state the main result of this subsection:

Theorem 2.2.9 (Unique factorisation)**.** *Every non-zero proper ideal $\mathfrak{I}$ of $\mathcal{O}_K$ may be factored into prime ideals of $\mathcal{O}_K$,*

$$\mathfrak{I}=\mathfrak{p}_1\cdots\mathfrak{p}_r,$$

in a unique way up to the order of the factors.

The proof of this theorem involves an interesting arithmetic in the ring $\mathcal{O}_K$, using, in particular, fractional ideals, which we now introduce:

Definition 2.2.10. Let D be a Dedekind domain with field of fractions K. We say that a D-submodule $\mathfrak{I}$ of K is a *fractional ideal* of D if there exists $a\in D$ such that $a\mathfrak{I}\subset D$, i. e., such that $a\mathfrak{I}$ is an ideal of D. When there may be risk of confusion, we refer to ideals of D as ordinary ideals. A *principal fractional ideal* is a fractional ideal of the form Dx for some $x\in K$.

Obviously, ordinary ideals are fractional ideals. Note that, even though we refer to fractional ideals of D, they are not actually subsets of D, unless they are ordinary ideals. The multiplication of ideals defined above can be extended to fractional ideals in the obvious way. The resulting operation is commutative, associative and admits a neutral element, namely, D (indeed, $D\mathfrak{I}=\mathfrak{I}D=\mathfrak{I}$ for every fractional ideal $\mathfrak{I}\subset K$). A fractional ideal $\mathfrak{I}$ is said to be *invertible* when there exists another fractional ideal $\mathfrak{I}'$ such that $\mathfrak{I}\mathfrak{I}'=\mathfrak{I}'\mathfrak{I}=D$. It is paramount to observe that every non-zero prime ideal is invertible:

Proposition 2.2.11 (Inverse of prime ideals)**.** *Let $\mathfrak{p}$ be a non-zero prime ideal of D. Define $\mathfrak{p}^{-1}$ to be*

$$\mathfrak{p}^{-1}=\{a\in K\mid a\mathfrak{p}\subset D\}.$$

Then $\mathfrak{p}^{-1}$ is a fractional ideal and $D\subsetneq\mathfrak{p}^{-1}$. Moreover, $\mathfrak{p}\mathfrak{p}^{-1}=D$.

Proof. It is clear from the definitions that $\mathfrak{p}^{-1}$ is a fractional ideal and that $D\subset\mathfrak{p}^{-1}$. We will show that this inclusion is proper. Take $a\in\mathfrak{p}$, $a\neq 0$. Then Da contains a product of non-zero prime ideals. Indeed, if not, let S be the set of all ideals of D failing to contain a product of non-zero prime ideals and let $\mathfrak{c}$ be its maximal element (in a Noetherian ring, a non-empty collection of ideals, partially ordered by inclusion, always has a maximal element, since ascending chains of ideals are stable). The ideal $\mathfrak{c}$ cannot be prime, so there exist $u,v\in D$ such that $uv\in\mathfrak{c}$ but

$u \notin \mathfrak{c}$ and $v \notin \mathfrak{c}$. Note that $\mathfrak{c} \subsetneq \mathfrak{c} + Du$ and $\mathfrak{c} \subsetneq \mathfrak{c} + Dv$ so, by maximality of $\mathfrak{c}$, each of these ideals must contain some product of prime ideals. However, since $(\mathfrak{c} + Du)(\mathfrak{c} + Dv) \subset \mathfrak{c}$, $\mathfrak{c}$ must also contain a product of prime ideals, contrary to our hypothesis. This proves there exist prime ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ such that Da contains $\mathfrak{p}_1 \cdots \mathfrak{p}_r$. Assume, furthermore, that r is minimal with this property. In particular, $\mathfrak{p}_1 \cdots \mathfrak{p}_r \subset \mathfrak{p}$ and so $\mathfrak{p}_i \subset \mathfrak{p}$ for some i which we will assume to be 1, without loss of generality. Then $\mathfrak{p}_1 = \mathfrak{p}$, since $\mathfrak{p}_1$ is maximal and $\mathfrak{p}$ is a proper ideal. For any $b \in \mathfrak{p}_2 \cdots \mathfrak{p}_r$ it follows that $b\mathfrak{p} \subset \mathfrak{p}\mathfrak{p}_2 \cdots \mathfrak{p}_r = \mathfrak{p}_1 \cdots \mathfrak{p}_r \subset Da$, and therefore that $(b/a)\mathfrak{p} \subset D$ so, by definition, $b/a \in \mathfrak{p}^{-1}$. Now, by minimality of r, we have that $\mathfrak{p}_2 \cdots \mathfrak{p}_r \not\subset Da$ so there exists some $b \in \mathfrak{p}_2 \cdots \mathfrak{p}_r$ such that $b \notin Da$, which means that $(b/a) \notin D$. This proves $D \subsetneq \mathfrak{p}^{-1}$.

Now, since $\mathfrak{p}\mathfrak{p}^{-1}$ is an ordinary ideal of D (indeed, $\mathfrak{p}\mathfrak{p}^{-1} \subset D$) and $\mathfrak{p}\mathfrak{p}^{-1}$ contains the prime ideal $\mathfrak{p}$, if we show that $\mathfrak{p}\mathfrak{p}^{-1} \neq \mathfrak{p}$, then it will follow that $\mathfrak{p}\mathfrak{p}^{-1} = D$, by maximality of $\mathfrak{p}$. We already know that $D \not\subset \mathfrak{p}^{-1}$. Note, however, that this does not immediately imply that $\mathfrak{p} \not\subset \mathfrak{p}\mathfrak{p}^{-1}$. In order to prove this we first need to observe that, for any fractional ideal $\mathfrak{J}$ of D, if $x \in K$ is such that $x\mathfrak{J} \subset \mathfrak{J}$, then $x \in D$. This is the content of Lemma 2.2.12, which we state and prove below. Having observed this, assume $\mathfrak{p} = \mathfrak{p}\mathfrak{p}^{-1}$. So, in particular, for any $x \in \mathfrak{p}^{-1}$ we have that $x\mathfrak{p} \subset \mathfrak{p}\mathfrak{p}^{-1} = \mathfrak{p}$ and then $x \in D$. This implies that $\mathfrak{p}^{-1} \subset D$ and, consequently, that $\mathfrak{p}^{-1} = D$, contradicting what we have established above. Therefore, $\mathfrak{p} \neq \mathfrak{p}\mathfrak{p}^{-1}$ and the proof is complete. □

Lemma 2.2.12. *Let $\mathfrak{J}$ be a fractional ideal of D. If $x \in K$ is such that $x\mathfrak{J} \subset \mathfrak{J}$, then $x \in D$.*

Proof. Let $\{\alpha_1, \ldots, \alpha_n\}$ be a generating set for $\mathfrak{J}$ (recall that D is Noetherian). Since $x\mathfrak{J} \subset \mathfrak{J}$, we have

$$x\alpha_j = \sum_{i=1}^{n} a_{ij}\alpha_i,$$

for $j = 1, \ldots, n$. This means that $(\mathrm{Id}_n x - A)\alpha = 0$, where Id_n is the $n \times n$ identity matrix (with coefficients in D), $A = (a_{ij})$ and $\alpha = (\alpha_1, \ldots, \alpha_n)$. Multiplying both sides by the classical adjoint of $(\mathrm{Id}_n x - A)$ gives that $\det(\mathrm{Id}_n x - A) = 0$. So x is a root of a monic polynomial with coefficients in D and thus integral over D. Since D is integrally closed, we conclude that $x \in D$. □

We are finally ready to prove the unique factorisation into prime ideals:

Proof of unique factorization. We prove the theorem for a general Dedekind domain D, since the reasoning is the same.

Let M be the set of proper ideals of D that do not admit factorization into prime ideals and suppose $M \neq \emptyset$. Since D is Noetherian, M has a maximal element.

Take $\mathfrak{a}$ to be a maximal element in M. As any ideal is contained in a maximal ideal, $\mathfrak{a}$ is contained in a maximal ideal $\mathfrak{m}$, which is, in particular, a prime ideal. It follows that $\mathfrak{a}\mathfrak{m}^{-1} \subset \mathfrak{m}\mathfrak{m}^{-1} = D$, meaning $\mathfrak{a}\mathfrak{m}^{-1}$ is an ordinary ideal of D. Just as in Proposition 2.2.11, $D \subsetneq \mathfrak{m}^{-1}$ implies that $\mathfrak{a} \subsetneq \mathfrak{a}\mathfrak{m}^{-1}$. Indeed, if $\mathfrak{a} = \mathfrak{a}\mathfrak{m}^{-1}$, then, by applying Lemma 2.2.12, we conclude that $\mathfrak{m}^{-1} = D$, contradicting Proposition 2.2.11. Now, $\mathfrak{a} \subsetneq \mathfrak{a}\mathfrak{m}^{-1}$ implies that $\mathfrak{a}\mathfrak{m}^{-1} \notin M$. Hence, there exist non-zero prime ideals $\mathfrak{p}_1, \dots, \mathfrak{p}_r$ such that $\mathfrak{a}\mathfrak{m}^{-1} = \mathfrak{p}_1 \cdots \mathfrak{p}_r$ and, finally, $\mathfrak{a} = \mathfrak{a}\mathfrak{m}^{-1}\mathfrak{m} = \mathfrak{p}_1 \cdots \mathfrak{p}_r\mathfrak{m}$.

For uniqueness, the reasoning follows the exact same lines as the proof of unique factorization for integers: suppose $\mathfrak{p}_1 \cdots \mathfrak{p}_r = \mathfrak{q}_1 \cdots \mathfrak{q}_s$ where $\mathfrak{p}_1, \dots, \mathfrak{q}_s$ are non-zero prime ideals. Then $\mathfrak{p}_1$ divides $\mathfrak{q}_1 \cdots \mathfrak{q}_s$ (meaning $\mathfrak{p}_1 \supset \mathfrak{q}_1 \cdots \mathfrak{q}_s$) and thus $\mathfrak{p}_1$ divides (contains) one of the ideals $\mathfrak{q}_1, \dots, \mathfrak{q}_s$ which we will assume, without loss of generality, to be $\mathfrak{q}_1$. Since non-zero prime ideals are maximal, we have that $\mathfrak{p}_1 = \mathfrak{q}_1$ and then $\mathfrak{p}_2 \cdots \mathfrak{p}_r = \mathfrak{q}_2 \cdots \mathfrak{q}_s$. Continuing inductively, we conclude that $r = s$ and $\mathfrak{p}_i = \mathfrak{q}_i$ for $i = 1, \dots, r$. □

Corollary 2.2.13 (Unique factorization for fractional ideals)**.** *Let $\mathfrak{J}$ be a fractional ideal of a Dedekind domain D. Then $\mathfrak{J}$ is expressible as*

$$\mathfrak{J} = \prod_{i=1}^{r} \mathfrak{p}_i^{\nu_i}, \tag{2.2.1}$$

in a unique way up to the order of the factors, where $\mathfrak{p}_i$ are pairwise distinct prime ideals of D and $\nu_i \in \mathbb{Z}$, for $i = 1, \dots, r$.

Proof. By definition, there is some $a \in D$ such that $a\mathfrak{J} \subset D$. In particular, $\mathfrak{a} = a\mathfrak{J}$ is an ordinary ideal of D and $\mathfrak{a} = (Da)\mathfrak{J}$. It follows from Theorem 2.2.9 that $\mathfrak{a}$ and Da can be uniquely factored into prime ideals and so, by inverting the prime factors of Da (Proposition 2.2.11), we obtain an expression for $\mathfrak{J}$ of the form (2.2.1). Uniqueness follows naturally from the uniqueness statement in the theorem. □

Corollary 2.2.14 (Group structure in the set of non-zero ideals.)**.** *Let $\mathscr{F}$ be the set of non-zero fractional ideals of a Dedekind domain D. Then $\mathscr{F}$, endowed with multiplication of ideals, is an abelian group.*

Proof. It has already been observed that multiplication of (fractional) ideals is commutative, associative and has D as a neutral element. Moreover, Proposition 2.2.11 tells us how to invert a prime element. Given any non-zero fractional ideal $\mathfrak{J}$, we know it to be of the form $\mathfrak{J} = \prod \mathfrak{p}_i^{\nu_i}$. Then, clearly, $\mathfrak{J}$ is invertible with inverse $\mathfrak{J}^{-1} = \prod \mathfrak{p}_i^{-\nu_i}$. □

Corollary 2.2.15 (Equivalent notions of divisibility for ideals)**.** *Let $\mathfrak{a}, \mathfrak{b}$ be two ordinary ideals of D. Then $\mathfrak{a}$ divides $\mathfrak{b}$ if and only if there exists an ordinary ideal $\mathfrak{c}$ such that $\mathfrak{b} = \mathfrak{a}\mathfrak{c}$.*

Proof. Sufficiency is straightforward. For necessity, we begin by factoring $\mathfrak{a}$ and $\mathfrak{b}$ into prime ideals. We have that $\mathfrak{q}_1 \cdots \mathfrak{q}_s = \mathfrak{b} \subset \mathfrak{a} = \mathfrak{p}_1 \cdots \mathfrak{p}_r$. In particular, $\mathfrak{q}_1 \cdots \mathfrak{q}_s \subset \mathfrak{p}_1$ and it follows from primality that $\mathfrak{q}_i \subset \mathfrak{p}_1$. We may assume without losses that $i = 1$ and, by maximality of prime ideals, we have that $\mathfrak{q}_1 = \mathfrak{p}_1$, so $\mathfrak{q}_2 \cdots \mathfrak{q}_s = \mathfrak{p}_2 \cdots \mathfrak{p}_r$. We continue inductively to find that $\mathfrak{q}_j = \mathfrak{p}_j$, $j = 2, \ldots, r$ and $r \leq s$ (otherwise we would have $D \subset \mathfrak{p}_{s+1} \cdots \mathfrak{p}_r \subset \mathfrak{p}_r$, a contradiction). Take $\mathfrak{c} = \mathfrak{q}_{r+1} \cdots \mathfrak{q}_s$. □

As another application of unique factorisation, we prove that the ideal norm (Definition 2.2.6) is multiplicative.

Proposition 2.2.16. *For ordinary ideals* $\mathfrak{a}, \mathfrak{b}$ *of* D*, one has that:*

1. $\mathrm{N}(\mathfrak{a}\mathfrak{b}) = \mathrm{N}(\mathfrak{a})\mathrm{N}(\mathfrak{b})$;
2. $\mathrm{N}(\mathfrak{a}) \in \mathfrak{a}$;
3. *if* K *is a number field, for any non-zero* $a \in \mathscr{O}_K$, $\mathrm{N}(\mathscr{O}_K a) = |N_{K|\mathbb{Q}}(a)|$.

Proof. (1) Note that it suffices to prove the case where $\mathfrak{b}$ is a prime ideal $\mathfrak{p}$. Since $\mathfrak{a}\mathfrak{p} \subset \mathfrak{a} \subset D$ we have that $|D/\mathfrak{a}\mathfrak{p}| = |D/\mathfrak{a}| \cdot |\mathfrak{a}/\mathfrak{a}\mathfrak{p}|$. We show that $D/\mathfrak{p}$ and $\mathfrak{a}/\mathfrak{a}\mathfrak{p}$ are isomorphic as abelian groups which concludes the argument. Take $a \in \mathfrak{a} \backslash \mathfrak{a}\mathfrak{p}$ which exists since $\mathfrak{a} \neq \mathfrak{a}\mathfrak{p}$ and consider the ideal $Da + \mathfrak{a}\mathfrak{p}$. Note that $\mathfrak{a}\mathfrak{p} \subset Da + \mathfrak{a}\mathfrak{p} \subset \mathfrak{a}$, so by unique factorisation, we have that either $Da + \mathfrak{a}\mathfrak{p} = \mathfrak{a}$ or $Da + \mathfrak{a}\mathfrak{p} = \mathfrak{a}\mathfrak{p}$, but the latter would imply that $a \in \mathfrak{a}\mathfrak{p}$, contrary to our assumptions. Consider the homomorphism $\phi : D/\mathfrak{p} \to \mathfrak{a}/\mathfrak{a}\mathfrak{p}$ given by $x + \mathfrak{p} \mapsto ax + \mathfrak{a}\mathfrak{p}$. It is well-defined since, for $x' \in D$ such that $x' - x \in \mathfrak{p}$, we note that $ax' - ax \in \mathfrak{a}\mathfrak{p}$. Now, given an element $u + \mathfrak{a}\mathfrak{p} \in \mathfrak{a}/\mathfrak{a}\mathfrak{p}$, because $u \in \mathfrak{a} = Da + \mathfrak{a}\mathfrak{p}$, one may write $u + \mathfrak{a}\mathfrak{p}$ as $ad + \mathfrak{a}\mathfrak{p} = \phi(d + \mathfrak{p})$, for some $d \in D$. This proves that ϕ is surjective. It is also injective since $\phi(x + \mathfrak{p}) = \mathfrak{a}\mathfrak{p}$ implies that $ax \in \mathfrak{a}\mathfrak{p}$ or, in other terms, that $\mathfrak{a}\mathfrak{p} \mid (Da)(Dx)$. Since a was chosen in $\mathfrak{a} \backslash \mathfrak{a}\mathfrak{p}$, we observe that the power of $\mathfrak{p}$ that divides Da is the same that divides $\mathfrak{a}$ and thus $\mathfrak{p} \mid Dx$, which is the same as saying that $x \in \mathfrak{p}$ proving that ϕ is injective.

(2) By definition, $\mathrm{N}(\mathfrak{a})$ is the order of the abelian group $D/\mathfrak{a}$ and hence a multiple of the order of any element of $D/\mathfrak{a}$. In particular, $\mathrm{N}(\mathfrak{a})d \in \mathfrak{a}$ for any $d \in D$. Take $d = 1$.

(3) This is Corollary 2.2.24. □

Proposition 2.2.16 (1) allows us to extend the notion of ideal norm to fractional ideals of $\mathscr{O}_K$ in a natural way, namely, if the non-zero fractional ideal $\mathfrak{J}$ is given by $\mathfrak{J} = \prod_{i=1}^{r} \mathfrak{p}_i^{\nu_i}$, where $\nu_i \in \mathbb{Z}$, we define $\mathrm{N}(\mathfrak{J}) = \prod_{i=1}^{r} \mathrm{N}(\mathfrak{p}_i)^{\nu_i}$. Note that the norm thus defined is still multiplicative. As a matter fact, if $\mathscr{F}$ denotes the group of all non-zero (fractional) ideals of $\mathcal{O}_K$, then one has a homomorphism $\mathrm{N} : \mathscr{F} \to \mathbb{R}^*_{>0}$.

Following P. Ribenboim, we conclude this subsection by recording a theorem that collects results from different mathematicians over the decades. It is a characterisation of Dedekind domains and most of its content has already been exposed here.

Theorem 2.2.17 (see [47, p.104]). *Let A be an integral domain. The following are equivalent:*

1. *The domain A is a Dedekind domain.*
2. *Every non-zero proper ideal of A can be factored into prime ideals in a unique way (up to the order of the factors).*
3. *Every non-zero proper ideal of A can be factored into prime ideals.*
4. *The set of non-zero fractional ideals of A form a multiplicative group.*

2.2.3 Discriminant

We have yet to prove the assertion, used in Proposition 2.2.3, that, for any number field K with a $\mathbb{Q}$-basis of algebraic integers $\{\alpha_1, \ldots, \alpha_n\}$, there exists a rational integer d such that $d\mathcal{O}_K \subset \mathbb{Z}\alpha_1 + \cdots + \mathbb{Z}\alpha_n$. This is the content of Proposition 2.2.19, below. But first, let us take this opportunity to introduce an important isomorphism invariant for number fields.

Definition 2.2.18. Let $\{\tau_1, \ldots, \tau_n\}$ be a K-basis of the separable extension $L \mid K$. We define the *discriminant* of this basis to be

$$\Delta(\tau_1, \ldots, \tau_n) = \det(\mathrm{Tr}_{L|K}(\tau_i\tau_j)), \tag{2.2.2}$$

which lies in K, being the determinant of a matrix with coefficients in K.

Let $\sigma_1, \ldots, \sigma_n : L \to \overline{K}$ denote the K-embeddings (i.e., embeddings that leave K fixed) of L into the algebraic closure of K. The following is a useful identity:

$$\Delta(\tau_1, \ldots, \tau_n) = (\det(\sigma_i\tau_j))^2. \tag{2.2.3}$$

Indeed, if we define $\delta(\tau_1, \ldots, \tau_n) = (\sigma_i\tau_j)$ then we get the matrix equation

$$(\mathrm{Tr}_{L|K}(\tau_i\tau_j)) = \delta(\tau_1, \ldots, \tau_n)\, \delta(\tau_1, \ldots, \tau_n)^T,$$

where $(\cdot)^T$ indicates the transpose matrix. Identity (2.2.3) then follows by taking determinants on both sides.

Now, let $\{\omega_1, \dots, \omega_n\}$ be another K-basis for L and let $C = (c_{ij})$ be defined by

$$\omega_j = \sum_{i=1}^{n} c_{ij}\tau_i\,, \quad j = 1, \dots, n,$$

i.e., C is the change of basis from $\{\omega_1, \dots, \omega_n\}$ to $\{\tau_1, \dots, \tau_n\}$. Since each c_{ij} belongs to K, it follows that

$$\sigma_k\omega_j = \sum_{i=1}^{n} c_{ij}(\sigma_k\tau_i) \quad k, j = 1, \dots, n, \tag{2.2.4}$$

which, in matrix form is but

$$\delta(\omega_1, \dots, \omega_n) = \delta(\tau_1, \dots, \tau_n)\, C,$$

whence we conclude that

$$\Delta(\omega_1, \dots, \omega_n) = (\det C)^2\, \Delta(\tau_1, \dots, \tau_n). \tag{2.2.5}$$

We observe that $\Delta(\omega_1, \dots, \omega_n) \neq 0$ for any basis $\{\omega_1, \dots, \omega_n\}$. Indeed, since $\det C \neq 0$, it follows from (2.2.5) that it is sufficient to check this assertion for a specific basis. The extension $L \mid K$ is assumed to be separable so, by the Primitive Element Theorem, there exists some $\theta \in L$ such that $L = K(\theta)$. We take the basis $\{1, \theta, \dots, \theta^{n-1}\}$. Note that $\delta(1, \theta, \dots, \theta^{n-1})$ is a Vandermonde matrix, and thus:

$$\Delta(1, \theta, \dots, \theta^{n-1}) = (\det \delta(1, \theta, \dots, \theta^{n-1}))^2 = \prod_{1 \leq i < j \leq n} (\sigma_j\theta - \sigma_i\theta)^2 \neq 0, \tag{2.2.6}$$

where the last passage is due to the fact that the K-embeddings $\sigma_1, \dots, \sigma_n$ are pairwise distinct and hence so are $\sigma_i\theta$ and $\sigma_j\theta$ for $i \neq j$.

Now we are ready to prove what was promised. We will do so in a slightly more general setting, since the proof remains the same. Let K be the field of fractions of an integrally closed integral domain A and let B denote the integral closure of A in the finite separable extension $L \mid K$. Keep in mind that the case of interest for us will be the one where $A = \mathbb{Z}$, $K = \mathbb{Q}$ and L is any number field (which, beware, is often denoted by K in this context).

Proposition 2.2.19. *Let $\{\alpha_1, \dots, \alpha_n\}$ be a K-basis for L consisting of elements of B. Let d be the discriminant of $\{\alpha_1, \dots, \alpha_n\}$, then*

$$dB \subset A\alpha_1 + \cdots + A\alpha_n$$

Proof. Let $b \in B$, then we may write $b = b_1\alpha_1 + \cdots + b_n\alpha_n$ where $b_1, \ldots, b_n \in K$. Multiplying both sides by α_i and taking traces, for $i = 1, \ldots, n$, we obtain the following linear system:

$$\begin{aligned}
\operatorname{Tr}_{L|K}(b\alpha_1) &= \operatorname{Tr}_{L|K}(\alpha_1\alpha_1)b_1 + \cdots + \operatorname{Tr}_{L|K}(\alpha_1\alpha_n)b_n \\
&\vdots \\
\operatorname{Tr}_{L|K}(b\alpha_n) &= \operatorname{Tr}_{L|K}(\alpha_n\alpha_1)b_1 + \cdots + \operatorname{Tr}_{L|K}(\alpha_n\alpha_n)b_n,
\end{aligned}$$

which we want to solve for the unknowns $b_1, \ldots, b_n$. Note that the traces appearing in this system all lie in A since they are traces of elements of B. Moreover, the determinant of the matrix of coefficients is precisely $\Delta(\alpha_1, \ldots, \alpha_n)$ which, according to (2.2.6) and the discussion preceding it, is non-zero. We may therefore apply Cramer's Rule and obtain that each b_j, for $j = 1, \ldots, n$, satisfies $db_j \in A$ where $d = \Delta(\alpha_1, \ldots, \alpha_n)$. □

This concludes, at last, the proof of Proposition 2.2.3. Before moving on, we point out that the concept of integral basis is perfectly suitable for being defined in greater generality. Indeed, let A, K, B and L be as above.

Definition 2.2.20. An *integral basis* of B over A is an A-basis for B, i.e., a set $\{\omega_1, \ldots, \omega_n\} \subset B$ such that any element of B can be uniquely written as a linear combination of $\omega_1, \ldots, \omega_n$ with coefficients in A. When there is no risk of confusion, this is simply referred to as an integral basis of $L \mid K$ or even an integral basis of L.

If such an integral basis exists, B is then a free A-module of finite rank. This means that B is an A-module that admits a finite basis. Moreover, it is immediate that this basis must also be a K-basis for L, whence $n = [L : K]$.

In view of this definition, Proposition 2.2.3 and Corollary 2.2.4 may be generalised as follows:

Proposition 2.2.21. *When A is a principal ideal domain, B admits an integral basis over A of cardinality $[L : K]$.*

Moreover, any finitely generated B-module $M \subset L$ is a free A-module of rank $[L : K]$.

Proof. The proof follows the same lines as the proof of Proposition 2.2.3, with one minor difference: in order to obtain that B is a free A-module of rank at most $[L : K]$, one needs to make use now of the fundamental theorem for free modules over principal ideal domains (hence the additional hypothesis that A is principal): any submodule of the free A-module A^k is free of rank $\leq k$. Note that, when

$A = \mathbb{Z}$, then this is just the result for free abelian groups that was evoked in the proof of Proposition 2.2.3.

The second part is a direct generalisation or Remark 2.2.5, obtained by performing the same adaptation as above. □

Let $\{\omega_1, \ldots, \omega_n\}$ and $\{\tau_1, \ldots, \tau_n\}$ be two integral basis of $\mathcal{O}_K$ and let C be the change of basis matrix. Note that C is invertible. Moreover, C has coefficients in $\mathbb{Z}$. This means that $\det C = \pm 1$ and, by equation (2.2.3), we conclude that $\Delta(\omega_1, \ldots, \omega_n) = \Delta(\tau_1, \ldots, \tau_n)$. The following definition is therefore justified:

Definition 2.2.22. Let K be a number field. We define the *discriminant* of K, denoted Δ_K, to be the discriminant of any integral basis of K.

Note that the discriminant of any basis consisting of algebraic integers is a rational integer. In particular, the discriminant of a number field is a rational integer. Furthermore, being defined in terms of Galois embeddings, the discriminant of a number field is an invariant of its isomorphism class. There exist, however, examples of non-isomorphic number fields with the same discriminant (see, for instance, [36, Example 0.2.11 (4)]).

Proposition 2.2.23. *Let $\mathfrak{a}$ be a non-zero ideal of $\mathcal{O}_K$ with $\mathbb{Z}$-basis $\{\alpha_1, \ldots, \alpha_n\}$. Then*

$$\Delta(\alpha_1, \ldots, \alpha_n) = m^2 \Delta_K,$$

where m is the index of $\mathfrak{a}$ in $\mathcal{O}_K$.

Proof. Let $\{\omega_1, \ldots, \omega_n\}$ be an integral basis of $\mathcal{O}_K$ in K, and let $a \in \mathfrak{a} \cap \mathbb{Z}$, $a \neq 0$ (take, for example, the norm of any non-zero element of $\mathfrak{a}$). We have that

$$\mathbb{Z}a\omega_1 + \cdots + \mathbb{Z}a\omega_n \subset \mathfrak{a} \subset \mathbb{Z}\omega_1 + \cdots + \mathbb{Z}\omega_n.$$

Define, for each $j = 1, \ldots, n$, the set $B_j = \{a_j\omega_j + \cdots + a_n\omega_n \in \mathfrak{a} \mid a_i, \ldots, a_n \in \mathbb{Z}\}$. Each B_j is non-empty since $a\omega_j \in B_j$. Pick τ_j in B_j such that the coefficient of ω_j is positive and minimal. Let us write $\tau_j = a_{jj}\omega_j + a_{j+1,j}\omega_{j+1} + \cdots + a_{nj}\omega_n$, for $j = 1, \ldots, n$. We claim that $\{\tau_1, \ldots, \tau_n\}$ generates $\mathfrak{a}$ over $\mathbb{Z}$: let $x \in \mathfrak{a}$ be written as $x = x_1\omega_1 + \cdots + x_n\omega_n$, where $x_1, \ldots, x_n \in \mathbb{Z}$. By the division algorithm, there exists q_1 and $0 \leq r_1 < a_{11}$ such that $x_1 = a_{11}q_1 + r_1$. Note that $x - q_1\tau_1 \in \mathfrak{a} = B_1$. Minimality of a_{11} implies that r_1 must be equal to zero and thus that $x - q_1\tau_1 \in B_2$. By the same argument, we find an integer q_2 such that $x - q_1\tau_1 - q_2\tau_2 \in B_3$. Proceeding inductively, we obtain integers $q_1, \ldots, q_n$ such that $x - q_1\tau_1 - \cdots - q_n\tau_n = 0$, as we wanted. Note also that $\tau_1, \ldots, \tau_n$ are linearly independent over $\mathbb{Z}$. This can be derived from the linear independency of $\omega_1, \ldots, \omega_n$ and from the change of basis matrix. This shows, in a more constructive manner, that $\mathfrak{a}$ admits an integral basis, which had already been proved in Corollary 2.2.4.

Moreover, if we repeat the algorithm described above, starting with an element $x = x_1\omega_1+\cdots+x_n\omega_n$ of $\mathcal{O}_K$, we find that x may be written as $x = x'+r_1\omega_1+\cdots+r_n\omega_n$, where $x' \in \mathfrak{a}$ and $0 \leq r_i < a_{11}$. Therefore, we have that $x \in \mathfrak{a}$ if and only if $r_1 = \cdots = r_n = 0$, i.e., if and only if a_{ii} divides x_i for $i = 1 \dots n$. This proves that the set $\{r_1\omega_1 + \cdots + r_n\omega_n \mid 0 \leq r_i < a_{ii}\}$ is a complete residue system modulo $\mathfrak{a}$. Its cardinality, which is easily seen to be $a_{11} \cdots a_{nn}$, is therefore equal to the index of $\mathfrak{a}$ in $\mathcal{O}_K$, denoted here by m.

Let $A = (a_{ij})$ denote the change of basis matrix from $\{\omega_1, \dots, \omega_n\}$ to $\{\tau_1, \dots, \tau_n\}$. Note that A is upper triangular and that $\det A = a_{11} \cdots a_{nn} = m$. It then follows from (2.2.5) that

$$\Delta(\alpha_1, \dots, \alpha_n) = \Delta(\tau_1, \dots, \tau_n) = (\det A)^2 \Delta(\omega_1, \dots, \omega_n) = m^2 \Delta_K.$$

□

Corollary 2.2.24. *For any non-zero $a \in \mathcal{O}_K$,*

$$\mathrm{N}(\mathcal{O}_K a) = |\mathrm{N}_{K|\mathbb{Q}}(a)|. \tag{2.2.7}$$

Proof. Let $\{\omega_1, \dots, \omega_n\}$ be an integral basis for $\mathcal{O}_K$. Then $\{a\omega_1, \dots, a\omega_n\}$ is an integral basis for $\mathcal{O}_K a$ and

$$\Delta(a\omega_1, \dots, a\omega_n) = [\det(\sigma_i(a\omega_j))]^2 = [\det(\sigma_i(a)\delta_{ij})]^2[\det \sigma_i\omega_j]^2 = \mathrm{N}_{K|\mathbb{Q}}(a)^2\Delta_K,$$

where the σ_i's denote, as usual, the embeddings of K into its algebraic closure, and (δ_{ij}) denotes the identity matrix (Kronecker delta).

This observation together with the proposition yields equation (2.2.7). □

We point out that, for a $\mathbb{Q}$-basis of K, $\{\alpha_1, \dots, \alpha_n\}$, consisting of algebraic integers (not necessarily an integral basis), it holds that

$$\Delta(\alpha_1, \dots, \alpha_n) = k^2 \Delta_K, \tag{2.2.8}$$

for some $k \in \mathbb{Z}$. Indeed, every α_i, when expressed as a linear combination of some integral basis, has rational integer coefficients. The change of basis matrix, from this integral basis to $\{\alpha_1, \dots, \alpha_n\}$, will then have rational integer coefficients and thus (2.2.8) follows from (2.2.5). As a consequence, we get the following simple criteria:

Proposition 2.2.25. *Let $\{\alpha_1, \ldots, \alpha_n\}$ be a $\mathbb{Q}$-basis for K consisting of algebraic integers. If $\Delta(\alpha_1, \ldots, \alpha_n)$ is square-free, then $\Delta(\alpha_1, \ldots, \alpha_n) = \Delta_K$ and $\{\alpha_1, \ldots, \alpha_n\}$ is an integral basis.*

Proof. From the previous paragraph it is clear that $\Delta(\alpha_1, \ldots, \alpha_n) = \Delta_K$. It remains to observe that $\{\alpha_1, \ldots, \alpha_n\}$ is an integral basis. Choose an integral basis $\{\omega_1, \ldots, \omega_n\}$ and let C be the change of basis matrix from $\{\alpha_1, \ldots, \alpha_n\}$ to this basis. Clearly, C has coefficients in $\mathbb{Z}$. From (2.2.5) and the fact that the discriminant of $\{\alpha_1, \ldots, \alpha_n\}$ equals the field discriminant, we obtain that $\det C = \pm 1$, meaning that C has an inverse with coefficients in $\mathbb{Z}$. Therefore, $\{\alpha_1, \ldots, \alpha_n\}$ is also a $\mathbb{Z}$-basis of $\mathcal{O}_K$. □

2.2.4 Extensions of Dedekind domains

In this subsection, we show that the integral closure of a Dedekind domain in a finite separable extension is also a Dedekind domain (Theorem 2.2.30). This fact will be used in Section 2.3 to prove, in some particular cases, the existence of extensions of valuations.

Before we prove Theorem 2.2.30, we need to address the problem of finding finitely generated submodules in a given finitely generated module. Unlike the case of vector spaces, it is not true in general that a submodule of a finitely generated module will also admit a finite generating set. In fact, this property deserves a special distinction and inspires the following definition:

Definition 2.2.26. A module for which every submodule is finitely generated is called a *Noetherian module*.

Since the ideals of a ring A are precisely its A-submodule, we point out that every Noetherian ring is in particular a Noetherian module. Next, we list some of the basic properties of Noetherian modules. For more details, see [47, Chapter 6].

Proposition 2.2.27.

1. *A module M is Noetherian if and only if every ascending chain of submodules eventually stabilises.*

2. *Every submodule or quotient of a Noetherian module is again Noetherian.*

3. *Let N be a Noetherian submodule of M such that M/N is also Noetherian. Then M is Noetherian.*

4. *If $M_1, \ldots, M_r$ are Noetherian then so is $M_1 \times \cdots \times M_r$.*

Sketch of proof. (1) is proved in the exact same way as in the case of Noetherian rings.

(2) follows easily from the characterisation given in (1).

In order to prove (3), let M' be a submodule of M and let $x_1, \dots, x_r \in M'$ be such that $\phi(x_1), \dots, \phi(x_r)$ span $\phi(M') \subset M/N$, where ϕ is the projection of M onto M/N. Pick a set $\{y_1, \dots, y_s\}$ generating the submodule $N \cap M'$ of N. It is easy to see that $\{x_1, \dots, x_r, y_1, \dots, y_s\}$ generates M'.

For $r = 2$ in (4), note that $(M_1 \times M_2)/M_2 \cong M_1$ is Noetherian and use (3). The general case follows by induction. □

These properties put together show us that modules defined over Noetherian rings behave "nicely", which is the content of the next theorem.

Theorem 2.2.28. *A finitely generated module over a Noetherian ring is Noetherian.*

Proof. Let M be a finitely generated A-module where A is a Noetherian ring. If $\{x_1, \dots, x_n\}$ generates M, there is an epimorphism $\phi : A^n \to M$ taking $(a_1, ..., a_n)$ to $\sum_{i=1}^n a_i x_i$, which implies that $M \cong A^n/\ker\phi$. The result now follows from Proposition 2.2.27 (4) and (2). □

Remark 2.2.29. We have encountered once before another aspect in which modules can be different from vector spaces, namely, that submodules of free modules of finite rank need not be free. In the occasion, we remedied this adversity by restricting to modules defined over principal ideal domains (see the proof of Proposition 2.2.21). If, however, we are only interested in the quality of being finitely generated, then it suffices to restrict to modules over Noetherian ring, as Theorem 2.2.28 demonstrates.

Theorem 2.2.30. *Let D be a Dedekind domain with field of fractions K. If $L \mid K$ is a finite separable extension, let E be the integral closure of D in L. Then E is also a Dedekind domain.*

Moreover, for any prime ideal $\mathfrak{p}$ of D, there exists a prime ideal $\mathfrak{P}$ of E such that $\mathfrak{P} \cap D = \mathfrak{p}$. We say that $\mathfrak{P}$ lies over *$\mathfrak{p}$.*

Proof. E is integrally closed by construction.

By Theorem 2.2.19, E is a submodule of a free D-module M of rank $n = [L : K]$. Every ideal $\mathfrak{I}$ of E, being a D-submodule of E, is also a submodule of M and, since D is Noetherian, Theorem 2.2.28 implies that $\mathfrak{I}$ is a finitely generated D-module, hence *a fortiori* a finitely generated E-module. This means that the ring E is Noetherian. Alternatively, one can use Proposition 2.2.27 (4) and (2) to prove that E is a Noetherian D-module, whence every ideal in E is a finitely generated D-module and consequently a finitely generated E-module.

Finally, if $\mathfrak{P}$ is a prime ideal of E, note first that $\mathfrak{p} = \mathfrak{P} \cap D$ is non-trivial. Indeed, take $x \in \mathfrak{P}$. Since x is integral over D, there are $d_0, \ldots, d_{n-1} \in D$, $d_0 \neq 0$, such that $x^n + d_{n-1}x^{n-1} + \cdots + d_0 = 0$, so in particular $d_0 \in \mathfrak{p}$. Moreover, it is immediate that $\mathfrak{p}$ is a prime ideal of D, and therefore maximal. So $D/\mathfrak{p}$ is a field, which is naturally embedded in the domain $E/\mathfrak{P}$. Let e be a non-zero element of the domain $E/\mathfrak{P}$. By the same reasoning as above, the ideal generated by e in $E/\mathfrak{P}$ intersects $D/\mathfrak{p}$ non-trivially (note that $E/\mathfrak{P}$ is integral over $D/\mathfrak{p}$) which means there exists a non-zero d in $D/\mathfrak{p}$ such that $d = ye$ for some $y \in E/\mathfrak{P}$. Let d' be the inverse of d in $D/\mathfrak{p}$, then $d'y$ is the inverse of e in $E/\mathfrak{P}$ which proves that $E/\mathfrak{P}$ is a field and, therefore, that $\mathfrak{P}$ is maximal. This concludes the proof that E is a Dedekind domain.

Now, let $\mathfrak{p}$ be a prime ideal of D. We denote by $E\mathfrak{p}$ the ideal generated by $\mathfrak{p}$ in the ring E, that is, all finite sums

$$\sum_i a_i x_i, \qquad \text{for } a_i \in E \text{ and } x_i \in \mathfrak{p}. \tag{2.2.9}$$

This is in accordance with the notation for product of ideals introduced earlier. Note that $E\mathfrak{p} \subsetneq E$. Indeed, suppose $E\mathfrak{p} = E$. Then, in particular, $1 \in E$ could be written in the form (2.2.9) and, for any $x \in \mathfrak{p}^{-1}$, one would have that $x = 1 \cdot x \in E$. This would imply that $\mathfrak{p}^{-1} \subset E \cap K = D$, contradicting Proposition 2.2.11. So $E\mathfrak{p}$ is a proper ideal and can thus be factored into prime ideals of E. Let $\mathfrak{P}$ be one of its prime factors. Then clearly $\mathfrak{p} \subset \mathfrak{P} \cap D$. In the previous paragraph we saw that $\mathfrak{P} \cap D$ is a non-empty prime ideal of D and hence proper. It follows from the maximality of $\mathfrak{p}$ that $\mathfrak{p} = \mathfrak{P} \cap D$. □

2.2.5 Dirichlet's Unit Theorem

In this subsection, we describe the structure of the multiplicative group of units of the ring of integers in a number field.

Let K be a number field of degree n. Since any finite extension of $\mathbb{Q}$ is separable, there exist n Galois embeddings of K into $\mathbb{C}$. We say a Galois embedding is *real* when its image lies in $\mathbb{R}$. Otherwise we say the embedding is *complex*. Note that if $\sigma : K \hookrightarrow \mathbb{C}$ is a complex embedding, its complex-conjugate $\overline{\sigma}$ is also an embedding (different from σ). So the complex embeddings of K come in pairs. Let r_1 be the number of real embeddings and r_2 the number of pairs of complex embeddings, so that $n = r_1 + 2r_2$. We denote these embeddings by $\sigma_1, \ldots, \sigma_{r_1}, \sigma_{r_1+1}, \overline{\sigma}_{r_1+1}, \ldots, \sigma_{r_1+r_2}, \overline{\sigma}_{r_1+r_2}$. In this notation, we state the following:

Theorem 2.2.31 (Dirichlet's Unit Theorem). *The group of units $\mathcal{O}_K^\times$ of $\mathcal{O}_K$ is a finitely generated group of rank $r_1 + r_2 - 1$. The torsion part is the (finite) cyclic subgroup formed by the roots of unity contained in K.*

In other words, there are elements $a_1, \dots, a_N, \zeta_t$ in $\mathcal{O}_K^\times$, where $N = r_1 + r_2 - 1$ and ζ_t is a tth root of unity, such that, any $a \in \mathcal{O}_K^\times$ can be uniquely written as

$$a = \zeta_t^{e_0} a_1^{e_1} \cdots a_N^{e_N}, \quad e_0 \in \mathbb{Z}/t\mathbb{Z}, e_1, \dots, e_N \in \mathbb{Z}.$$

The rest of this subsection is dedicated to proving this theorem.

By a *lattice* Λ in $\mathbb{R}^n$ we mean a subset of the form

$$\mathbb{Z}\omega_1 + \cdots + \mathbb{Z}\omega_m = \{a_1\omega_1 + \cdots a_m\omega_m \mid a_i \in \mathbb{Z}\}, \tag{2.2.10}$$

where $\{\omega_1, \dots, \omega_m\}$ are linearly independent vectors in $\mathbb{R}^n$ (cf. Definition 3.5.1). The set $\{\omega_1, \dots, \omega_m\}$ is called a basis of Λ and the subset

$$P = \{x_1\omega_1 + \cdots + x_m\omega_m \mid 0 \leq x_i \leq 1, i = 1, \dots, m\}$$

is called the *fundamental parallelepiped* of Λ with respect to this basis. A lattice is said to be *complete* when $m = n$.

Lattices and complete lattices can be characterised topologically, as the next two propositions show.

Proposition 2.2.32. *A subgroup $\Lambda \subset \mathbb{R}^n$ is a lattice if and only if Λ is discrete.*

Proof. If Λ is given by 2.2.10, then discreteness follows. Conversely, suppose Λ is discrete. Let V be the subspace of $\mathbb{R}^n$ spanned by Λ and let $m \leq n$ be the dimension of V. Choose a basis $\{v_1, \dots, v_m\}$ for V such that each v_i is in Λ and consider the group $\Lambda_0 = \mathbb{Z}v_1 + \cdots \mathbb{Z}v_m$. We claim that Λ_0 has finite index in Λ. Indeed, let $P_0 = \{x_1v_1 + \cdots x_mv_m \mid 0 \leq x_i \leq 1, i = 1, \dots, m\}$ be the fundamental parallelepiped of Λ_0 in V with respect to the basis $\{v_1, \dots, v_m\}$. Let $\{\lambda_j\}_{j\in J}$ be a set of representatives of the cosets of Λ_0 in Λ. Each λ_j, as a point in V, can be written as $a_j + \lambda'_j$ where $a_j \in P_0$ and $\lambda'_j \in \Lambda_0$ (note that $a_{j_1} \neq a_{j_2}$ if $j_1 \neq j_2$). In particular, for every, $j \in J$, $a_j = \lambda_j - \lambda'_j \in \Lambda \cap P_0$. Here is where the hypothesis of discreteness comes into play. Since Λ is a discrete subgroup of $\mathbb{R}^n$, it is closed (assume, by way of contradiction, that $\{\alpha_k\}_k$ is a Cauchy sequence of pairwise distinct elements of Λ, then $\{\alpha_k - \alpha_{k-1}\}$ is a sequence in Λ converging to 0, which violates the discreteness of Λ). As P_0 is compact, the intersection $\Lambda \cap P_0$ is a compact and discrete subset, thus finite. It follows that J is finite, of cardinality, say, N. This is precisely the index of Λ_0 in Λ, proving our claim. In particular, $N\Lambda \subset \Lambda_0$ so Λ is a subgroup of a free abelian group of rank m. By the fundamental theorem for finitely generated abelian groups, Λ is free abelian of rank at most m. On the other hand, by the same theorem, Λ has rank at least m since $\Lambda_0 \subset \Lambda$. The proposition follows. □

Proposition 2.2.33. *A lattice Λ is complete if and only if it there exists a bounded set $B \subset \mathbb{R}^n$ such that $\{\lambda + B\}_{\lambda\in\Lambda}$ covers $\mathbb{R}^n$.*

Proof. If Λ is complete, the fundamental parallelepiped P satisfies all the properties required.

Conversely, let $\Lambda = \mathbb{Z}v_1 + \cdots \mathbb{Z}v_m$ where $m \leq n$. Let V be the subspace of $\mathbb{R}^n$ spanned by Λ, i.e, $V = \mathbb{R}v_1 + \cdots + \mathbb{R}v_m$. In particular, $\dim V \leq m \leq n$. If we are able to prove that $V = \mathbb{R}^n$ then it would follow that $m = n$ and that $\{v_1, \ldots, v_n\}$ is linearly independent, thus a basis for $\mathbb{R}^n$. So let $x \in \mathbb{R}^n$. For each positive integer k, there exists some $a_k \in B$ and $\lambda_k \in \Lambda$ such that $kx = a_k + \lambda_k$. Since B is bounded, $\frac{a_k}{k} \to 0$ as $k \to \infty$. It follows that $\frac{\lambda_k}{k} \to x$ as $k \to \infty$ so, as each $\frac{\lambda_k}{k}$ is in V and V is closed in $\mathbb{R}^n$, we find that $x \in V$. □

The *volume*[1]of Λ is, by definition, the volume of P, that is:

$$\mathrm{vol}(\Lambda) = |\det(\omega_1, \ldots, \omega_n)|.$$

Note that this definition does not depend on the choice of basis for Λ since a basis change, in this case, must have determinant ±1.

Minkowski's famous theorem on the *geometry of numbers* asserts that a symmetric convex subset $V \in \mathbb{R}^n$ of volume sufficiently large must contain a point of $\Lambda \setminus \{0\}$.

Theorem 2.2.34 (Minkowski). *Let $\Lambda \subset \mathbb{R}^n$ be a complete lattice and $V \subset \mathbb{R}^n$ a measurable subset. Suppose also that V is symmetric (i.e., that $V = -V$) and convex. If $\mathrm{vol}(V) > 2^n\mathrm{vol}(\Lambda)$ then V contains a point of $\Lambda \setminus \{0\}$.*

Proof. Suppose all translates $\{\lambda + \frac{1}{2}V\}_{\lambda\in\Lambda}$ were pairwise disjoint. Then, in particular, their intersection with the fundamental parallelepiped P of Λ would also be pairwise disjoint. Each $(\lambda + \frac{1}{2}V) \cap P$, when translated by $-\lambda$, results in the set $\frac{1}{2}V \cap (P - \lambda)$, of same volume. Since the translates $\{(P - \lambda)\}_{\lambda\in\Lambda}$ cover $\mathbb{R}^n$, we have the following sequence of inequalities:

$$\begin{aligned}\mathrm{vol}(\Lambda) = \mathrm{vol}(P) &\geq \sum_{\lambda\in\Lambda} \mathrm{vol}\left(\left(\lambda + \frac{1}{2}V\right) \cap P\right) = \sum_{\lambda\in\Lambda} \mathrm{vol}\left(\frac{1}{2}V \cap (P - \lambda)\right)\\ &\geq \mathrm{vol}\left(\frac{1}{2}V\right) = \frac{1}{2^n}\mathrm{vol}(V) > \mathrm{vol}(\Lambda),\end{aligned}$$

[1] Regarding Λ as a group of isometries acting on $\mathbb{R}^n$, this number should actually be called the *covolume* of Λ. However, we shall maintain this more classical terminology for the present section.

a contradiction. Therefore, the translates $\{\lambda + \frac{1}{2}V\}_{\lambda\in\Lambda}$ cannot be pairwise disjoint, meaning there exist $v, w \in V$, $v \neq w$, and $\lambda_1, \lambda_2 \in \Lambda$ such that $(v-w)/2 = \lambda_2 - \lambda_1 \in \Lambda$. Since V is symmetric and convex, we have that $v, -w \in V$ and thus $(v-w)/2 \in V$. □

Now, let us define the map $\psi : K \to \mathbb{R}^n$ as:

$$x \mapsto (\sigma_1 x, \dots, \sigma_{r_1} x, \mathrm{Re}(\sigma_{r_1+1}x), \mathrm{Im}(\sigma_{r_1+1}x), \dots, \mathrm{Re}(\sigma_{r_1+r_2}x), \mathrm{Im}(\sigma_{r_1+r_2}x)).$$

Note that ψ is an injective homomorphism of groups. Moreover, if $\omega_1, \dots, \omega_n$ is an integral basis of $\mathscr{O}_K$ (Proposition 2.2.3), then $\psi(\omega_1), \dots, \psi(\omega_n)$ are linearly independent vectors in $\mathbb{R}^n$, proving that $\psi(\mathscr{O}_K)$ is a lattice in $\mathbb{R}^n$. Indeed, by performing elementary transformations on the rows of $(\psi(\omega_1), \dots, \psi(\omega_n))$, one can easily see that $\det(\psi(\omega_1), \dots, \psi(\omega_n))$ equals the determinant of the $n \times n$ matrix whose jth column is given by $(\sigma_1\omega_j, \dots \sigma_{r_1}\omega_j, \sigma_{r_1+1}\omega_j, -\frac{1}{2}\overline{\sigma_{r_1+1}\omega_j}, \dots, \sigma_{r_1+r_2}\omega_j, -\frac{1}{2}\overline{\sigma_{r_1+r_2}\omega_j})$. Therefore:

$$\det(\psi(\omega_1), \dots, \psi(\omega_n)) = \left(-\frac{1}{2}\right)^s \det(\sigma_i\omega_j),$$

where the right-hand side is known to be non-zero (see the discussion after Definition 2.2.18). It also follows that:

$$\mathrm{vol}(\psi(\mathscr{O}_K)) = 2^{-s}\sqrt{|\Delta(\omega_1, \dots, \omega_n)|}. \tag{2.2.11}$$

In order to study the group $\mathscr{O}_K^\times$, we will need a homomorphism that maps product to sum. Let $\mu : K^\times \to \mathbb{R}^{r+s}$ be defined as:

$$x \mapsto (\log|\sigma_1 x|, \dots, \log|\sigma_{r_1}x|, 2\log|\sigma_{r_1+1}x|, \dots, 2\log|\sigma_{r_1+r_2}x|).$$

Note first that, for any $a \in \mathscr{O}_K^\times$, $\log|\sigma_1 a| + \cdots + \log|\sigma_{r_1}x| + 2\log|\sigma_{r_1+1}x| + \cdots + 2\log|\sigma_{r_1+r_2}a| = \log|N_{K|\mathbb{Q}}(a)| = 0$, since $N_{K|\mathbb{Q}}(a) = \pm 1$, being a a unit of $\mathscr{O}_K$. It follows that $\mu(\mathscr{O}_K^\times)$ is a subgroup of the hyperplane $H = \{(x_1, \dots, x_{r_1+r_2}) \in \mathbb{R}^{r_1+r_2} \mid x_1 + \cdots + x_{r_1+r_2} = 0\}$. We aim to show that $\mu(\mathscr{O}_K^\times)$ is a complete lattice in H.

To prove completeness, we must find a bounded set $B \subset H$ such that H is covered by $\{\mu(a) + B\}_{a\in\mathscr{O}_K^\times}$. For that purpose, let us bring back the map $\psi : K \to \mathbb{R}^n$, except this time we make the identification $\mathbb{R}^n = \mathbb{R}^{r_1+2r_2} \cong \mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$. In other words, ψ maps x to $(\sigma_1 x, \dots, \sigma_{r_1}x, \sigma_{r_1+1}x, \dots, \sigma_{r_1+r_2}x)$. Note that $\mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$ is a ring with coordinate-wise multiplication. Define the *norm* of an element $y = (y_1, \dots, y_{r_1+r_2}) \in \mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$ to be $\mathscr{N}(y) = |y_1| \cdots |y_{r_1}||y_{r_1+1}|^2 \cdots |y_{r_1+r_2}|^2$ motivated by the fact that, in this way, $\mathscr{N}(\psi(x))$ is precisely $N_{K|\mathbb{Q}}(x)$. In particular, for $a \in \mathscr{O}_K^\times$, $\mathscr{N}(\psi(a)) = 1$. Also, it follows immediately from the definition that $\mathscr{N}$ is multiplicative and that, if $\mathscr{N}(y) \neq 0$, then y is invertible in $\mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$.

Consider the hypersurface $S = \{y \in \mathbb{R}^{r_1} \times \mathbb{C}^{r_2} \mid \mathcal{N}(y) = 1\}$ and observe that the restriction $\mu : \mathcal{O}_K^\times \to H$ factors through ψ in the following manner:

$$\mathcal{O}_K^\times \xrightarrow{\psi} S \xrightarrow{\lambda} H \subset \mathbb{R}^{r_1+r_2},$$

where λ here represents the "logarithmic" map, i.e., the map taking $(y_1, \dots, y_{r_1+r_2})$ to $(\log|y_1|, \dots, \log|y_{r_1}|, 2\log|y_{r_1+1}|, \dots, 2\log|y_{r_1+r_2}|)$. The map λ is surjective. Then it suffices to find $B' \subset S$, such that S is covered by the translates $\{\psi(a)B'\}_{a\in\mathcal{O}^\times}$ and the coordinates of B' are bounded away from 0 and from ∞. Indeed, we just set $B = \lambda(B')$. In the rest of the argument we will need the following:

Lemma 2.2.35. *For $M > 0$, there are only finitely many ideals $\mathfrak{a} \subset \mathcal{O}_K$ with norm* $\mathrm{N}(\mathfrak{a}) \leq M$.

Proof. The norm of $\mathfrak{a}$ is the index of $\mathfrak{a}$ as a subgroup of the additive group $\mathcal{O}_K$. Since $\mathcal{O}_K$ is finitely generated, there are only finitely many subgroups of index m for $m = 1, 2, \dots, M$.

Alternatively, let m be the norm of $\mathfrak{a}$, then $m \in \mathfrak{a}$ (Proposition 2.2.16 (2)), whence $\mathfrak{a} \mid m\mathcal{O}_K$. Since, for each m, $1 \leq m \leq M$, there are only finitely many prime divisors of $m\mathcal{O}_K$, it follows from unique factorisation that $m\mathcal{O}_K$ has finitely many divisors. Therefore, there are only finitely many possibilities for $\mathfrak{a}$. □

For $y \in S$, observe that $y\psi(\mathcal{O}_K)$ is still a lattice, where the product here denotes multiplication in the ring $\mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$ (which, we recall, is just coordinate-wise multiplication). The same computations used to obtain (2.2.11), can be used to show that

$$\begin{aligned}\mathrm{vol}(y\psi(\mathcal{O}_K)) &= 2^{-s}\sqrt{|\Delta_K|} \cdot \mathcal{N}(y)\\ &= 2^{-s}\sqrt{|\Delta_K|}.\end{aligned}$$

If we choose a bounded subset $V \subset \mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$, convex and symmetric, with $\mathrm{vol}(V)$ sufficiently large, it follows from Minkowski's Theorem 2.2.34 that, for any $y \in S$, there exists a in $\mathcal{O}_K$, $a \neq 0$, for which $y\psi(a) \in V$. The coordinates of $y\psi(a)$ are bounded, whence $\mathcal{N}(y\psi(a)) \leq C$, for some constant $C > 0$ depending only on V. But $\mathcal{N}(y\psi(a)) = \mathcal{N}(y)\mathcal{N}(\psi(a)) = \mathrm{N}_{K|\mathbb{Q}}(a)$, so the norm of the ideal $a\mathcal{O}_K$ must be bounded by C. Lemma 2.2.35 implies that there are only finitely many principal ideals with norm bounded by C. That is to say, there are $a_1, \dots, a_k \in \mathcal{O}_K$ such that any principal ideal of $\mathcal{O}_K$ with norm at most C must be equal to some $a_i\mathcal{O}_K$. In particular, $a\mathcal{O}_K = a_i\mathcal{O}_K$, for some i. This means there is a unit $u_i \in \mathcal{O}_K^\times$ for which $a = a_iu_i$. Note that $\mathcal{N}(\psi(a)) = \mathrm{N}_{K|\mathbb{Q}}(a) \neq 0$ so $\psi(a)$ is invertible in $\mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$. It follows that $y \in \psi(u_i^{-1})\psi(a_i)^{-1}V$. By defining $B' = S \cap (\psi(a_1)^{-1}V \cup \cdots \cup \psi(a_k)^{-1}V)$, we obtain that $y \in \psi(u_i^{-1})B'$. Moreover, every x in $\psi(a_1)^{-1}V \cup \cdots \cup \psi(a_k)^{-1}V$ has bounded coordinates in $\mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$. So, if x is in B', no coordinate of x can approach 0 nor ∞, since $\mathscr{N}(x) = 1$. This concludes the proof that $\mu(\mathscr{O}_K^\times)$ is a complete lattice in H and, therefore, has rank $r_1 + r_2 - 1$.

Now we must describe the kernel of $\mu : \mathscr{O}_K^\times \to H$. Let a in $\mathscr{O}_K$ be such that $\mu(a) = 0$. This means that $|\sigma_1 a| = \cdots = |\sigma_{r_1+r_2} a| = 1$. As we shall see, this implies that a is a root of unity. Indeed, recall the following:

Lemma 2.2.36. *There are only finitely many algebraic integers a of degree at most d and such that a and all of its conjugates are bounded.*

Proof. Let a be one such algebraic integer and let $P(X) = X^n + c_{n-1}X^{n-1} + \cdots + c_0$ be its minimal polynomial, so $n \leq d$. As a is an algebraic integer, we know $P \in \mathbb{Z}[X]$. On the other hand, P can be factored over $\mathbb{C}$ as $\prod_{i=1}^n (X - \sigma_i a)$. Since every $|\sigma_i a|$ is bounded by, say, M, every coefficient of P must be bounded by a constant M' depending only on M. Indeed, the coefficients of P are symmetric functions on the $\sigma_i a$. There are only finitely many integral monic polynomials of degree at most d and coefficients bounded by M', and therefore finitely many roots of such polynomials. The result follows. □

This lemma implies, in particular, that the kernel of μ is finite. Furthermore, if a is in the kernel of μ then $1, a, a^2, \ldots$ are also in the kernel of μ, whence $a^l = a^m$ for some $l > m$ and thus a is a root of unity, as claimed. The kernel of μ is thus a finite subgroup of the circle $\mathbb{S}^1 \subset \mathbb{C}$ and, as such, must be cyclic (this can be deduced, for example, from the fact that subgroups of the real line are either infinite cyclic or dense).

This concludes the proof of Theorem 2.2.31.

2.3 Valuations

In this section we introduce valuations, along with global and local fields. These objects are fundamental for the study of quaternion algebras in the next chapters. The material here can be complemented by the Appendix A, which comprises a brief introduction to Krull valuations. For more information on the vast theory of valuations and its applications, the reader may refer to [11], [10], [17], [48] or even to the fine set of notes [12].

2.3.1 Definition and basic properties

Definition 2.3.1. A *valuation* v on a field K is a non-negative function $v : K \to \mathbb{R}_{\geq 0}$ that satisfies

(i) $v(x) = 0$ if and only if $x = 0$;

(ii) (Multiplicativity) $v(xy) = v(x)v(y)$ for all $x, y \in K$;

(iii) There is a constant $C > 0$ such that $v(1 + x) \leq C$ whenever $v(x) \leq 1$.

One may also refer to the pair (K, v) as a *valued field*.

Note that v^a, $a > 0$, is a valuation whenever v is. Two valuations v_1, v_2 on K are said to be equivalent if $v_1 = v_2^a$ for some positive number $a \in \mathbb{R}$. A *place* of K is an equivalence class of valuations on K. It is immediate that any valuation on K is equivalent to a valuation satisfying condition (iii) with $C = 2$. Moreover, a valuation satisfies (iii) with $C = 2$ if and only if it is sub-additive, i.e., if and only if it satisfies the triangle inequality. Sufficiency is straightforward (note that $v(\pm 1) = 1$, by (ii)). Necessity, however, demands some computation: let x, y be in K, non-zero, and assume $v(x) \geq v(y)$. Then $v(x + y) = v(x)v(1 + y/x) \leq 2v(x) = 2\max\{v(x), v(y)\}$. By induction, we have that $v(\sum_{i=1}^{2^k} x_i) \leq 2^k \max\{v(x_i)\}$ so, for any integer $n > 0$, choosing k such that $2^{k-1} \leq n \leq 2^k$ yields $v(\sum_{i=1}^{n} x_i) \leq 2^k \max\{v(x_i)\} \leq 2n \max\{v(x_i)\}$. In particular, $v(n) = v(1 + \cdots + 1) \leq 2n$. It follows that

$$\begin{aligned} v(x + y)^n &= v\left(\sum_{i=0}^{n} \binom{n}{k} x^i y^{n-i}\right) \\ &\leq 2(n+1) \max_i \left\{ v\left(\binom{n}{i}\right) v(x)^i v(y)^{n-i} \right\} \\ &\leq 4(n+1) \max_i \left\{ \binom{n}{i} v(x)^i v(y)^{n-i} \right\} \\ &\leq 4(n+1)(v(x) + v(y))^n. \end{aligned}$$

Taking the nth root on both sides and letting n go to infinity gives $v(x + y) \leq v(x) + v(y)$.

Definition 2.3.2. A valuation is said to be *non-Archimedean* if it satisfies (iii) with $C = 1$. If that is the case, note that any valuation equivalent to it is also non-Archimedean. A valuation is *Archimedean* when it is not non-Archimedean.

Equivalently, a valuation v is non-Archimedean if it satisfies the *ultrametric inequality*

$$v(x + y) \leq \max\{v(x), v(y)\},$$

for every $x, y \in K$.

A non-Archimedean valuation may be characterised as follows:

Lemma 2.3.3. *A valuation v on K is non-Archimedean if and only if v is bounded on the subgroup generated by 1 of the additive group $(K, +)$.*

Proof. Necessity is clear. For sufficiency, suppose there is some $M > 0$ such that $v(1 + \cdots + 1) \leq M$ for any number of summands 1. Let $x \in K$ be such that $v(x) \leq 1$. As pointed out earlier, v is equivalent to a valuation satisfying the triangle inequality, and it is non-Archimedean if and only if the latter is. Therefore, we may assume without loss of generality that v satisfies the triangle inequality. Then

$$\begin{aligned} v(1+x)^n &\leq \sum_{i=0}^{n} v\left(\binom{n}{i}\right) v(x)^i \\ &\leq (n+1)M. \end{aligned}$$

Taking the nth roots on both sides and letting n go to infinity concludes the argument. □

Corollary 2.3.4. *If K has positive characteristic then every valuation on K is non-Archimedean.*

Corollary 2.3.5. *If $L \mid K$ is a field extension then a valuation on L is non-Archimedean if and only if its restriction to K is non-Archimedean.*

The following is a simple but quite useful observation about non-Archimedean valuations that follows directly from the definitions. It is commonly expressed by saying that, in a non-Archimedean *metric*, every triangle is isosceles.

Proposition 2.3.6. *If v is a non-Archimedean valuation and $v(a) < v(b)$, then $v(a+b) = v(b)$.*

Example 2.3.7. Let us see the model examples of valuations:

(i) Let $\sigma : K \to \mathbb{C}$ be a Galois embedding and define $v_\sigma(x) = |\sigma(x)|, \ \forall x \in K$, where $|\cdot|$ is the usual norm in $\mathbb{C}$. Then v_σ is an Archimedean valuation and, as we shall see, two such valuations, v_σ and $v_{\sigma'}$, are equivalent if and only if σ' is the complex conjugate of σ. An *infinite place* of K is the equivalence class of some Archimedean valuation on K. Every such v_σ thus defines an infinite place of K.

(ii) Let $\mathfrak{p}$ be any prime ideal in $\mathcal{O}_K$ and let c be a real number larger than 1. For any non-zero $x \in \mathcal{O}_K$, define $v_\mathfrak{p}(x) = c^{-\operatorname{ord}_\mathfrak{p}(x)}$, where $\operatorname{ord}_\mathfrak{p}(x)$ is the largest integer k such that $\mathfrak{p}^k$ divides the ideal $\mathcal{O}_K x$. It is natural to extend these functions to 0 as $\operatorname{ord}_\mathfrak{p}(0) = +\infty$ and $v_\mathfrak{p}(0) = 0$. Finally, since K is the field of fractions of $\mathcal{O}_K$, and $v_\mathfrak{p}$ is multiplicative, we may extend it to K as $v_\mathfrak{p}(x/y) := v_\mathfrak{p}(x)/v_\mathfrak{p}(y)$.

It is easy to check that $v_{\mathfrak{p}}$ defines a non-Archimedean valuation, known as the $\mathfrak{p}$-adic valuation, on K. A *finite place* of K is the equivalence class of some non-Archimedean valuation on K. Each $\mathfrak{p}$-adic valuation thus defines a finite place on K.

(iii) A particular case of example (ii) above is the p-adic valuations v_p on $\mathbb{Q}$, where $p > 0$ is a prime integer.

(iv) Let K be the field of fractions of $\mathbb{F}_q[X]$, i.e., $K = \mathbb{F}_q(X)$, where $\mathbb{F}_q$ is a finite field. For a polynomial $P(X) \in \mathbb{F}_q[X]$, define $|P|_\infty = q^{\deg P}$ and, for $P(X)/Q(X) \in \mathbb{F}_q(X)$, define $|P/Q|_\infty = q^{\deg P - \deg Q}$. Then $|\cdot|_\infty$ defines a valuation on $\mathbb{F}_q(X)$. Its equivalence class is usually referred to as the *infinite place* of $\mathbb{F}_q(X)$. Note that, unlike the the case of number fields, $|\cdot|_\infty$ receives the name "infinite place" even though it is non-Archimedean (by Corollary 2.3.4). The *finite places* of $\mathbb{F}_q(X)$ are determined in an analogous fashion as those of a number field, namely, given an irreducible polynomial $P(X) \in \mathbb{F}_q[X]$, any $f(X) \in F_q[X]$ may be (uniquely) factored as $f = P^n Q$ where Q is prime to P. Set $v_P(f) = c^n$ for some $0 < c < 1$, and extend v_P to $\mathbb{F}_q(X)$ in the usual manner.

It is an important result that (i) and (ii) comprise all valuations on a number field K, up to equivalence. In other words, any Archimedean valuation on K is equivalent to v_σ for some Galois embedding $\sigma : K \to \mathbb{C}$ (Corollary 2.3.53), and any non-Archimedean valuation on K is equivalent to $v_{\mathfrak{p}}$ for some prime ideal $\mathfrak{p} \subset \mathcal{O}_K$ (Proposition 2.3.54).

Remark 2.3.8. The function $v(x) = 1$ for all non-zero $x \in K$ is clearly a (not very interesting) valuation on K, known as the *trivial valuation*. Henceforth we will tacitly assume that valuations are non-trivial whenever needed.

In the non-Archimedean case, it is often convenient to consider the logarithmic counterpart of a valuation, what we call an *additive valuation*. Bear in mind that any statement in terms of valuations has an analogue for additive valuations which will not always be made explicit here.

Definition 2.3.9. A *(rank one) additive valuation* is a function $u : K \to \mathbb{R} \cup \{\infty\}$ satisfying

(i) $u(x) = \infty$ if and only if $x = 0$;

(ii) $u(xy) = u(x) + u(y)$;

(iii) $u(x + y) \geq \min\{u(x), u(y)\}$;

where we make the customary conventions that $x \leq \infty$ and $x + \infty = \infty$ for all $x \in \mathbb{R}$. Two additive valuations u_1 and u_2 are said to be *equivalent* if there exists a real number $a > 0$ such that $u_1 = au_2$. The additive valuation defined by $u(x) = 0$, for all non-zero $x \in K$, is called *trivial*.

Note that v is a valuation if and only if $-\log(v)$ is an additive valuation (equivalently, u is an additive valuation if and only if $\exp(-u)$ is a valuation).

Remark 2.3.10. Some authors prefer to call the object defined in Definition 2.3.1 an *absolute value*, reserving the term valuation for the functions defined in Definition 2.3.9.

2.3.2 Topology

Just as in the case of norms on a vector space, a valuation v satisfying the triangle inequality provides a metric on K by means of a distance function defined as $d(x, y) = v(x - y)$, for every $x, y \in K$. More generally, let v be an arbitrary valuation on K and consider the collection of sets $B_\epsilon(p) := \{x \in K \mid v(x - p) < \epsilon\}$ for every $p \in K$ and every $\epsilon > 0$. This collection is clearly a basis for a topology. We will see next that two equivalent valuations induce the same topology, and since any valuation is equivalent to one satisfying the triangle inequality, the topology just described is always a metric topology.

Proposition 2.3.11. *Two valuations K are equivalent if and only if they induce the same topology.*

Proof. Let v_1 and v_2 be two valuations on K and let τ_1 and τ_2 be the corresponding induced topologies. Necessity is clear, so suppose $\tau_1 = \tau_2$. Let $x \in K$ be any element. We claim that $v_1(x) < 1$ if and only if $v_2(x) < 1$. Since the hypothesis is of a topological nature, in order to prove the claim we translate the statement $v_i(x) < 1$ into topological terms, namely: $v_i(x) < 1$ if and only if $x^n \to 0$ in τ_i as $n \to \infty$, $i = 1, 2$. Now, the sequence $(x_n)_{n \geq 1}$ converges to 0 in τ_1 if and only if it converges to 0 in τ_2. The claim follows.

By taking inverses, we see that $v_1(x) < 1, = 1$ or > 1 if and only if $v_2(x) < 1, = 1$ or > 1, respectively.

Take any non-zero elements $x, y \in K$. Then $v_1(x^n y^m) \gtreqless 1$ if and only if $v_2(x^n y^m) \gtreqless 1$, for any integers n, m. Using multiplicativity of valuations and taking logarithms, we obtain that

$$q\,\frac{\log v_1(x)}{\log v_1(y)} + 1 \gtreqless 0 \quad \text{if and only if} \quad q\,\frac{\log v_2(x)}{\log v_2(y)} + 1 \gtreqless 0,$$

for every rational number q. This is only possible if

$$\frac{\log v_1(x)}{\log v_2(x)} = \frac{\log v_1(y)}{\log v_2(y)}.$$

Since x, y were taken arbitrarily, we conclude that, for every non-zero $x \in K$,

$$\frac{\log v_1(x)}{\log v_2(x)} = a > 0,$$

whence $v_1 = v_2^a$. □

In the course of the proof of Proposition 2.3.11 a characterisation for equivalent valuations was derived, which is of independent interest. We state this characterisation separately, with a small adjustment.

Proposition 2.3.12. *Let v_1 and v_2 be valuations on K such that*

$$v_1(x) < 1 \implies v_2(x) < 1, \quad \textit{for all } x \in K. \tag{2.3.1}$$

Then v_1 and v_2 are equivalent.

Proof. Taking reciprocals, we also have that $v_1(x) > 1 \implies v_2(x) > 1$ for all $x \in K$.

Suppose there exists $a \in K$ such that $v_1(a) = 1$ and $v_2(a) \neq 1$, say, $v_2(a) > 1$. Take $b \in K$ non-zero such that $v_1(b) < 1$ (which is possible once we assume v_1 is non-trivial. See Remark 2.3.8). Then $v_1(ba^n) < 1$ while, for sufficiently large n, $v_2(ba^n) > 1$, contradicting (2.3.1).

Thus $v_2(x) < 1$ if and only if $v_1(x) < 1$ and the rest follows as in the proof of Proposition 2.3.11. □

Finally, we observe that the operations of addition, multiplication and inversion are continuous in the topology induced by the valuation; i.e., K, endowed with this topology, is a topological field.

2.3.3 Completions

Since a valuation v on K induces a metric topology, we may complete K in the metric sense and obtain the metric space K_v which is easily seen to be a field with a valuation $\hat{v}$ that extends v.

More precisely, we may assume, without loss of generality, that v satisfies the triangle inequality and thus induces a metric d on K. Consider the set of Cauchy sequences in (K, d) equipped with pointwise addition and multiplication. Modding out by the subset (ideal) $\mathfrak{n}$ of Cauchy sequences converging to 0, we obtain the completion $(K_v, \hat{d})$, where the metric $\hat{d}$ is defined in the usual way: for

$x = (x_n)_{n\geq 1} + \mathfrak{n}$ and $y = (y_n)_{n\geq 1} + \mathfrak{n}$ in K_v, define $\hat{d}(x,y) = \lim_{n\to\infty} d(x_n, y_n)$. At this stage, $(K_v, \hat{d})$ is a complete metric space with K embedded in it as a dense subset. Indeed, the elements of K are naturally identified with the constant Cauchy sequences. Since K_v is a quotient ring, addition and multiplication on K_v are already defined. If $x = (x_n)_{n\geq 1} + \mathfrak{n}$ is non-zero, then x_n is bounded away from 0, for every sufficiently large index n, and we define x^{-1} to be $(x_n^{-1})_{n\geq 1} + \mathfrak{n}$. It is straightforward to check that inversion is well-defined and that all the field axioms hold. This makes K_v into a field. Now, for an arbitrary $x = (x_n)_{n\geq 1} + \mathfrak{n}$ in K_v note that $(v(x_n))_{n\geq 1}$ is a Cauchy sequence in the real line and therefore converges. We define the valuation $\hat{v}$ as $\hat{v}(x) = \lim_{n\to\infty}\{v(x_n)\}$. Note that $\hat{v}$, defined in this way, is indeed a valuation that extends v. Moreover, since $\hat{d}(x,y) = \lim_{n\to\infty} d(x_n, y_n) = \lim_{n\to\infty} v(x_n - y_n) = \hat{v}(x-y)$, it follows that $\hat{v}$ induces the complete metric $\hat{d}$ that extends d. It is a customary abuse of notation to continue to write v instead of $\hat{v}$.

The completion described is unique up to isomorphism: suppose the field L is complete with respect to a valuation v' and $\sigma : K \to L$ is a field embedding preserving the valuations, i.e., such that $v' \circ \sigma = v$. Then σ can naturally be extended to K_v and $\sigma(K_v)$ is clearly going to be the closure of $\sigma(K)$ in L, $\overline{\sigma(K_v)}$. In particular, if K is densely embedded in L, than $\sigma(K_v) = L$.

As a consequence of Lemma 2.3.3, we see that $\hat{v}$ is non-Archimedean if and only if v is non-Archimedean.

Finally, if v is non-Archimedean, the image of the function v on K coincides with that of v on K_v; i.e., $\{v(x) \mid x \in K\} = \{v(x) \mid x \in K_v\}$. Indeed, let $x \in K_v$. By definition, there exists $x' \in K$ such that $v(x' - x) < v(x)$. It follows from Proposition 2.3.6 that $v(x') = v(x)$.

2.3.4 Chinese Remainder Theorem and Weak Approximation

In this subsection, we prove an approximation result (Theorem 2.3.14) due to E. Artin and G. Whaples. This theorem might be seen as a metric version of the Chinese Remainder Theorem for rings. It is sometimes referred to as *Weak Approximation Theorem*, on account of deeper results existing in this direction. But first, let us recall the Chinese Remainder Theorem.

Theorem 2.3.13 (Chinese Remainder Theorem for Rings). *Let A be a ring and let $J_1, \dots, J_r$ be (two-sided) ideals in A that are pairwise* coprime, *i.e., such that $J_i + J_j = A$ for $i \neq j$. Let $J = \bigcap_{i=1}^r J_i$. Then there exists an isomorphism*

$$A/J \cong \bigoplus_{i=1}^{r} A/J_i,$$

where $x + J \mapsto (x + J_1, \dots, x + J_r)$. In other words, given any elements $a_1, \dots, a_r \in A$, there exists some $x \in A$ such that $x - a_i \in J_i$ for $i = 1, \dots, r$ and, moreover, if y is any other element of A satisfying this then $x - y \in J$.

Proof. First, we note that pairwise coprimality is equivalent to the condition that

$$J_i + \bigcap_{j \neq i} J_j = A \quad \text{for every} \quad i = 1, \dots, r. \tag{2.3.2}$$

Indeed, (2.3.2) clearly implies that the ideals are pairwise coprime. Conversely, for each $j \neq i$, since $J_i + J_j = A$, we may write $1 = a_j + b_j$ where $a_j \in J_i$ and $b_j \in J_j$. Then $1 - a_j \in J_j$ for every $j \neq i$. So, if we set $a = (1 - a_1) \cdots (1 - a_{j-1})(1 - a_{j+1}) \cdots (1 - a_r)$, then $a \in \bigcap_{j \neq i} J_j$. Furthermore, by the distributive property, one can easily see that $a = 1 + a'$ where $a' \in J_i$, so $1 \in J_i + \bigcap_{j \neq i} J_j$ and (2.3.2) follows. This is where we used the assumption that the ideals involved are two-sided. This theorem is often stated with (2.3.2) in lieu of pairwise coprimality.

Let us assume (2.3.2). Consider the ring homomorphism ϕ from A to $\bigoplus_{i=1}^{r} A/J_i$ given by $x \mapsto (x + J_1, \dots, x + J_r)$. The kernel of ϕ is clearly J, so all we have to show is that ϕ is surjective. Note that $\bigoplus_{i=1}^{r} A/J_i$ is generated by elements of the form $(a_1 + J_1, \dots, a_r + J_r)$ such that $a_j \in J_j$ for all j but one, say, $j = i$. So if we prove that every such element is in the image of ϕ, being that ϕ is a homomorphism, surjectivity will follow at once. Assume, for simplicity, that $i = 1$, that is, assume that $a_j \in J_j$ for $j = 2, \dots, r$. Write $a_1 = x + y$ where $x \in J_1$ and $y \in \bigcap_{j \neq 1} J_j$, which is possible, according to (2.3.2). Then $a_1 - x$ is clearly mapped to $(a_1 + J_1, \dots, a_r + J_r)$. □

As a particular case of the Chinese Remainder Theorem we see that, given r (pairwise distinct) rational primes $p_1, \dots, p_r$, along with positive integers $n_1, \dots, n_r$, and integers $a_1, \dots, a_r$, there exists some $x \in \mathbb{Z}$ that simultaneously satisfies $x \equiv a_j \pmod{p_j^{n_j}}$, for $j = 1, \dots, r$. Since, in the p-adic metric, two integers are close to each other precisely when their difference is divisible by a high power of p, we may reinterpret this conclusion as saying that one may find $x \in \mathbb{Z}$ arbitrarily close to a_j with respect to the p_j-adic metric for every $j = 1, \dots, r$ simultaneously.

Theorem 2.3.14 (Weak Approximation). *Let K be a field and let $v_1, \dots, v_m$ be non-equivalent valuations on K. Denote by K_{v_j} the completion of K with respect to the valuation v_j. The diagonal embedding*

$$K \hookrightarrow \prod_{j=1}^{m} K_{v_j}$$

has a dense image. Alternatively, for any choice of points $x_j \in K_j$, $j = 1, \dots, m$, and for any $\epsilon > 0$ there exists $x \in K$ such that $v_j(x - x_j) < \epsilon$, for all $j = 1, \dots, m$ (i.e., x is simultaneously ϵ-close to x_j, with respect to v_j, for $j = 1, \dots, m$).

Proof. We may assume, without loss of generality, that each x_j is in K. Indeed, by the definition of completion, we can find, for each $j = 1, \dots, m$, an element $x'_j \in K$ such that $v_j(x_j - x'_j) < \epsilon/2$.

The crucial step in this proof is to note that, for any i between 1 and m, there exists some $\alpha_i \in K$ such that $v_i(\alpha_i) > 1$ while $v_j(\alpha_i) < 1$ for every $j \neq i$. This is proved by induction on m. For simplicity, we assume $i = 1$.

For $m = 2$, since the valuations are non-equivalent, there exists, by Proposition 2.3.12, $\phi \in K$ such that $v_1(\phi) > 1$ and $v_2(\phi) \leq 1$. Likewise, there exists $\psi \in K$ such that $v_1(\psi) \leq 1$ and $v_2(\psi) > 1$. Take $\alpha_1 = \phi\psi^{-1}$.

Assume this is true for $m - 1$. Choose $\phi \in K$ such that $v_1(\phi) > 1$ and $v_j(\phi) < 1$ for $2 \leq j \leq m - 1$. By what we have already proved, there exists $\psi \in K$ such that $v_1(\psi) > 1$ and $v_m(\psi) < 1$. We define α_1 according to whether $v_m(\phi)$ is $<, =$ or > 1. Namely, if $v_m(\phi) < 1$, pick $\alpha_1 = \phi$ and we are done. If $v_m(\phi) = 1$, pick $\alpha_1 = \phi^r\psi$ for sufficiently large r. Finally, for $v_m(\alpha_1) > 1$, define $\alpha_1 = \frac{\phi^r}{1+\phi^r}\psi$, for sufficiently large r. It is easy to see that, in every situation, α_1 has the required properties.

Now, let $\alpha_i \in K$ be such that $v_i(\alpha_i) > 1$ and $v_j(\alpha_i) < 1$ for every $j \neq i$. Then pick

$$x = \sum_{j=1}^{m} \frac{\alpha_j^r}{1 + \alpha_j^r} x_j,$$

for sufficiently large r. □

2.3.5 Ostrowski's Theorems

We have seen in Example 2.3.7 that the usual absolute value on $\mathbb{Q}$, denoted here by $|\cdot|_\infty$, is an Archimedean valuation and that the p-adic valuations v_p on $\mathbb{Q}$ are non-Archimedean. A. Ostrowski proved in 1916 that, up to equivalence, these are the only valuations on $\mathbb{Q}$.

Theorem 2.3.15 (Ostrowski). *Any non-trivial valuation v on $\mathbb{Q}$ is equivalent either to the usual absolute value on $\mathbb{Q}$ or to some p-adic valuation, according to whether v is Archimedean or non-Archimedean.*

Proof. Let $a, b \in \mathbb{Z}$ be such that $a > 1$ and $b > 0$. Write b in base a expansion; i.e., write

$$b = b_m a^m + b_{m-1} a^{m-1} + \cdots + b_0, \tag{2.3.3}$$

where $b_j \in \mathbb{Z}$, $0 \leq b_j < a$ for $j = 1, \dots, m$, and $b_m > 0$. Note that, since $a^m \leq b$, we have that $m \leq \log b / \log a$.

Let $d = \max\{v(2), \ldots, v(a-1)\}$. Then, it follows from (2.3.3) and the triangle inequality (which we may assume v satisfies, without loss of generality) that

$$\begin{aligned} v(b) &\leq d \sum_{j=0}^{m} v(a)^j & (2.3.4) \\ &\leq d(m+1)\max\{1, v(a)^m\} & (2.3.5) \\ &\leq d\left(\frac{\log b}{\log a} + 1\right)\max\{1, v(a)^{\log b/\log a}\}, & (2.3.6) \end{aligned}$$

where the penultimate inequality is just the observation that if $v(a) < 1$ then $v(a)^j < 1 = \max\{1, v(a)^m\}$, and if $v(a) \geq 1$ then $v(a)^j \leq v(a)^m = \max\{1, v(a)^m\}$. Either way, $v(a)^j \leq \max\{1, v(a)^m\}$, for $j = 0, \ldots, m$.

Choose $b = c^n$ in (2.3.4), for any $c > 0$ and $n = 0, 1, \ldots$. Taking the nth root and letting $n \to \infty$, one obtains

$$v(c) \leq \max\{1, v(a)^{\log c/\log a}\}. \tag{2.3.7}$$

If v is Archimedean, then there exists some integer $c > 0$ such that $v(c) > 1$. For any integer $a > 1$, since $v(c) > 1$, (2.3.7) implies that $v(a)^{\log c/\log a} > 1$ and consequently that $v(a) > 1$. We may thus exchange a and c in (2.3.7) to find that

$$v(c)^{1/\log c} = v(a)^{1/\log a},$$

for every integer $a > 1$. In particular, if we let $\lambda = \log\left(v(c)^{1/\log c}\right) > 0$, then $v(a)^{1/\log a} = e^{\lambda}$ and, finally

$$v(a) = a^{\lambda} = |a|^{\lambda}, \quad \text{for every integer } a > 1. \tag{2.3.8}$$

One can easily check that (2.3.8) extends to every $a \in \mathbb{Q}$, proving that v is equivalent to $|\cdot|$.

If v is non-Archimedean, then $v(c) \leq 1$ for every integer c. Then there exists some integer $c > 1$ such that $v(c) < 1$, otherwise v would be trivial. Let $p > 1$ be the smallest integer such that $v(p) < 1$. Note that p is prime, by minimality. Any integer a such that $p \nmid a$ may be written as $a = pq + r$, where $0 < r < p$. By minimality of p, we must have $v(r) = 1$. Since $v(pq) = v(p)v(q) < 1$, it follows from Proposition 2.3.6 that $v(a) = v(pq + r) = 1$.

Let $\lambda = -\log v(p)/\log p > 0$. For any integer b, if we define $\mathrm{ord}_p(b)$ to be the largest integer k such that p^k divides b, then we may write $b = p^{\mathrm{ord}_p(b)}a$ where $p \nmid a$. Then

$$v(b) = v(p)^{\mathrm{ord}_p(b)} = (p^{-\mathrm{ord}_p(b)})^{\lambda} = v_p(b)^{\lambda}.$$

Just as before, one can easily extend the previous equality to all $b \in \mathbb{Q}$, showing that v is equivalent to v_p. □

Using weak approximation (Theorem 2.3.14) we derive, as a corollary, that the usual absolute value on $\mathbb{C}$ is, up to equivalence, the only Archimedean valuation on $\mathbb{C}$. This is a very particular case of another celebrated theorem due to Ostrowski (Theorem 2.3.17).

Corollary 2.3.16. *Any Archimedean valuation on $\mathbb{C}$ must be equivalent to the usual absolute value.*

Proof. Let v be an Archimedean valuation on $\mathbb{C}$ and let $|\cdot|$ denote the usual absolute value. We may assume without loss of generality that v satisfies the triangle inequality.

By Theorem 2.3.15, the restriction of v to the field $\mathbb{Q}$ must be equivalent to the absolute value on $\mathbb{Q}$ and, since $\mathbb{R}$ is the completion of $\mathbb{Q}$ with respect to $|\cdot|$, the same holds for the restrictions of v and $|\cdot|$ to $\mathbb{R}$; i.e., there exists $\lambda > 0$ such that $v = |\cdot|^\lambda$ on $\mathbb{R}$.

Now, let $z = x + iy \in \mathbb{C}$, where $x, y \in \mathbb{R}$. Clearly $|x| \leq |z|$ and $|y| \leq |z|$. Since i is a root of unity, we have that $v(i) = 1$, from where it follows that

$$v(z) \leq v(x) + v(y) = |x|^\lambda + |y|^\lambda \leq 2|z|^\lambda. \tag{2.3.9}$$

In other word, we cannot have $v(z)$ very large and $|z|$ very small. Thus, if v is not equivalent to $|\cdot|$ then we can use weak approximation and derive a contradiction. Indeed, let $a \in \mathbb{C}$ be such that $v(a) > 100$ and let $\epsilon > 0$. Weak approximation then provides us with a $b \in \mathbb{C}$ such that $|b| = |b - 0| < \epsilon$ while $v(b - a) < \epsilon$. Together with (2.3.9), these imply that $100 - \epsilon < v(b) \leq 2|b|^\lambda < 2\epsilon^\lambda$, a contradiction since $\epsilon > 0$ was arbitrary. □

Theorem 2.3.17 (Ostrowski). *If K is a complete field with respect to some Archimedean valuation then K is isomorphic either to $\mathbb{R}$ or to $\mathbb{C}$ and this valuation is equivalent to the usual absolute value.*

Although we will not prove Ostrowski's Theorem 2.3.17 here, we point out that it can be derived as a particular case of a more general result concerning real Banach algebras (i.e., unital algebras over $\mathbb{R}$ endowed with a complete norm that is submultiplicative and that takes the value 1 on the multiplicative unit).

Theorem 2.3.18 (Gelfand). *If a real Banach algebra $(A, |\cdot|)$ is a field then $(A, |\cdot|)$ is isomorphic either to $(\mathbb{R}, |\cdot|_\infty)$ or to $(\mathbb{C}, |\cdot|_\infty)$.*

Proof. See, for example, [12, Theorem 1.48]. □

For the sake of completeness, we mention that a result analogous to Ostrowski's Theorem 2.3.15 also holds for the field $\mathbb{F}_q(X)$, namely, that the valuations described in Example 2.3.7(iv) are the only valuations on $\mathbb{F}_q(X)$, up to equivalence.

Proposition 2.3.19. *Using the notation of Example 2.3.7(iv), any valuation v on the field $\mathbb{F}_q(X)$ is equivalent either to $|\cdot|_\infty$ or to v_P for some irreducible polynomial $P \in \mathbb{F}_q[X]$.*

Proof. See [44, Theorem 4-30]. □

2.3.6 Valuation ring, uniformiser and residue field

Throughout this subsection, v will denote a non-Archimedean valuation while $u : K \to \mathbb{R} \cup \{\infty\}$ will denote an additive valuation. A non-Archimedean valuation v on a field K naturally induces the following objects:

Definition 2.3.20.

(i) The set $\mathfrak{o} = \{x \in K \mid v(x) \leq 1\}$ is a ring called the ring of *(valuation) integers*. It is a valuation ring in the sense that, for any x in its field of fractions K, we have that $x \in \mathfrak{o}$ or $x^{-1} \in \mathfrak{o}$. For this reason, $\mathfrak{o}$ is sometimes referred to as the *valuation ring* of K with respect to v.

(ii) $\mathfrak{p} = \{x \in K \mid v(x) < 1\}$ is the unique maximal ideal in $\mathfrak{o}$.

(iii) The set $\mathfrak{o} \setminus \mathfrak{p} = \{x \in K \mid v(x) = 1\}$ is the group of units of $\mathfrak{o}$.

(iv) The quotient $\mathfrak{o}/\mathfrak{p}$ is a field called the *residue field*. It is, in fact, a field since the ideal $\mathfrak{p}$ is maximal in $\mathfrak{o}$.

For the additive valuation u, the valuation ring $\mathfrak{o}$ and its unique maximal ideal $\mathfrak{p}$ are given, respectively, by $\mathfrak{o} = \{x \in K \mid u(x) \geq 0\}$ and $\mathfrak{p} = \{x \in K \mid u(x) > 0\}$.

Note, first, that the ring of integers determines the valuation up to equivalence. In other words, two valuations are equivalent if and only if they have the same valuation ring. Indeed, necessity is obvious and sufficiency follows from Proposition 2.3.12.

For the completion K_v of K with respect to v, we will denote the corresponding ring of integers and maximal ideal respectively by $\overline{\mathfrak{o}}$ and $\overline{\mathfrak{p}}$ when the valuation in question is clear and there is no risk of confusion. Note that $\mathfrak{o} = \overline{\mathfrak{o}} \cap K$ and that $\mathfrak{p} = \overline{\mathfrak{p}} \cap K$. Therefore, there is a well-defined field embedding $\phi : \mathfrak{o}/\mathfrak{p} \to \overline{\mathfrak{o}}/\overline{\mathfrak{p}}$ taking $x + \mathfrak{p}$ to $x + \overline{\mathfrak{p}}$. The monomorphism ϕ is actually a field isomorphism:

$$\mathfrak{o}/\mathfrak{p} \cong \overline{\mathfrak{o}}/\overline{\mathfrak{p}}$$

Indeed, surjectivity of ϕ follows from the fact that, for any $x \in K_v$, there exists $x' \in K$ such that $v(x-x') < 1$, i.e., such that $x-x' \in \overline{\mathfrak{p}}$. Then $\phi(x'+\mathfrak{p}) = x'+\overline{\mathfrak{p}} = x+\overline{\mathfrak{p}}$.

At the end of §§2.3.3, we noted that v takes on the the same values whether defined over K or over its completion K_v. Let us look into this set with more attention. Denote by $K^\times$ the group of non-zero elements of K under multiplication. Then $\{v(x) \mid x \in K^\times\}$ is a subgroup of the multiplicative group of positive reals, called the *value group* of v. For the p-valuation on $\mathbb{Q}$, for example, the value group will be $\{c^n \mid n \in \mathbb{Z}\}$, for some $c > 0$. Note that this is a discrete subset of $\mathbb{R}_{>0}$. In fact, this is the case for most of the valuations we will be concerned with, and it has fundamental consequences, which motivate the next definition.

Definition 2.3.21. We say that a valuation v is *discrete* when its value group is discrete. Equivalently, v is discrete when there exists $\delta > 0$ such that

$$1 - \delta < v(x) < 1 + \delta \implies v(x) = 1, \tag{2.3.10}$$

for every $x \in K$.

Analogously, in the case of the additive valuation u, the value group $u(K^\times)$ is a subgroup of the additive group of real numbers $(\mathbb{R}, +)$. Then u is discrete precisely when its value group is a discrete subgroup of $(\mathbb{R}, +)$. In this case, $u(K^\times)$ is an infinite cyclic group of the form $\alpha\mathbb{Z}$, for some real number $\alpha > 0$. It follows that u is equivalent to an additive valuation with value group $\mathbb{Z}$ (such discrete valuations are said to be *normalised*).

Example 2.3.22. In this example we observe the existence of several non-discrete valuations. Recall that a division group is an abelian group $(G, +)$ satisfying the property that for every $g \in G$ and every positive integer n, g may be "divided by n", i.e, there exists $h \in G$ such that $g = nh$. Let $u : K \to \mathbb{R} \cup \{\infty\}$ be an additive valuation. Note that if K is algebraically closed, then the value group $u(K^\times)$ is a division group. Indeed, given g such that $g = u(x)$ and a positive integer n, let $y \in K$ be a root of the polynomial $X^n - x \in K[X]$ and set $h = u(y)$. Clearly $g = nh$. Now, it is easy to see that a discrete subgroup of $(\mathbb{R}, +)$ cannot be a division group. We conclude from this that a non-trivial valuation (additive or otherwise) on an algebraically closed field is never discrete.

In particular, we obtain that $\mathbb{Q}_p$ is not algebraically closed.

Proposition 2.3.23. *The valuation v is discrete if and only if $\mathfrak{p}$ is principal.*

Proof. Suppose $\mathfrak{p} = (\pi)$. If $v(x) < 1$ then $x \in \mathfrak{p}$ which mean that $x = a\pi$ for some $a \in \mathfrak{o}$. So $v(x) \leq v(\pi)$. On the other hand, if $v(x) > 1$ than $v(x^{-1}) < 1$ and, by what we have just established, $v(x) \geq v(\pi)^{-1}$. Condition (2.3.10) follows.

Conversely, suppose v is discrete. In particular, the set $\{v(x) \mid v(x) < 1\}$ attains its maximum. Let $\pi \in \mathfrak{p}$ be such that $v(\pi) = \max_{v(x)<1} v(x)$. If $x \in \mathfrak{p}$, then $v(x) < 1$ and $v(x) \leq v(\pi)$. So $x\pi^{-1} = a \in \mathfrak{o}$ whence $x = a\pi \in (\pi)$. □

Definition 2.3.24. An element π such that $\mathfrak{p} = (\pi)$ is called a *uniformiser* for v. Note that a uniformiser is not necessarily unique. If, however, π' is another uniformiser, then $v(\pi) = v(\pi')$.

There are, in fact, a few more properties that characterise a discrete valuation. We expand Proposition 2.3.23 and collect these properties below.

Proposition 2.3.25. *In the above notation, the following are equivalent:*

1. *v is discrete;*
2. *$\mathfrak{p}$ is principal;*
3. *$\mathfrak{o}$ is a principal ideal domain (PID);*
4. *$\mathfrak{o}$ is Noetherian.*

Proof. (1) $\iff$ (2) is precisely the content of Proposition 2.3.23.

(1) $\implies$ (3): Any proper ideal $\mathfrak{I}$ of $\mathfrak{o}$ is contained in $\mathfrak{p} = \{x \in \mathfrak{o} \mid v(x) < 1\}$. Discreteness implies that v attains its maximum in $\mathfrak{I}$, and the proof follows just as in Proposition 2.3.23.

(3) $\implies$ (4) is immediate.

(4) $\implies$ (2): Let $a_1, \dots, a_n$ be such that $\mathfrak{p} = \langle a_1, \dots, a_n \rangle$ and $0 < v(a_1) \leq v(a_2) \leq \cdots \leq v(a_n)$. For $i = 1, \dots, n-1$, $v(a_i a_n^{-1}) \leq 1$ so that $a_i a_n^{-1} = a \in \mathfrak{o}$ and $a_i \in \langle a_n \rangle$. It follows that $\mathfrak{p} = \langle a_n \rangle$. □

When the valuation group is discrete, the underlying set must be of the form $\{c^n \mid n \in \mathbb{Z}\}$ for some $0 < c < 1$, since it is isomorphic to a discrete subgroup of the additive group of real numbers (via the logarithmic function). Since the uniformiser is characterised by maximising $v(x)$ for $x \in \mathfrak{p}$, we see that $v(\pi) = c$ and thus, for every non-zero $x \in K$, there exists an integer n such that $v(x) = v(\pi)^n$. As in the example of $\mathfrak{p}$-adic valuations, this integer n is called the order of x and is denoted $\operatorname{ord}_v(x)$. It is clearly independent of the choice of uniformiser π since it only depends on $v(\pi)$. In particular, every non-zero x is of the form $x = \epsilon\pi^n$ for some ϵ such that $v(\epsilon) = 1$. One can be even more precise:

Proposition 2.3.26. *Suppose v is a discrete valuation on K, let π be a uniformiser and let A be a set of representatives for the cosets of $\mathfrak{o}/\mathfrak{p}$. Any non-zero element $a \in K_v$ can be uniquely written as*

$$a = \pi^m \sum_{n=0}^{+\infty} a_n \pi^n, \tag{2.3.11}$$

where $a_n \in A$, and $m = \operatorname{ord}_v(a) \in \mathbb{Z}$.

Proof. If $m = \mathrm{ord}_v(a)$ then $\pi^{-m}a \in \bar{\mathfrak{o}}$. Since $\bar{\mathfrak{o}}/\bar{\mathfrak{p}} \cong \mathfrak{o}/\mathfrak{p}$, A is also a set of representatives for the cosets in $\bar{\mathfrak{o}}/\bar{\mathfrak{p}}$. Let a_0 be the unique element in A such that $\pi^{-m}a \in a_0 + \bar{\mathfrak{p}}$. Then $\pi^{-m}a - a_0 \in \bar{\mathfrak{p}}$ and so it must be of the form $b_1\pi$ for some $b_1 \in \bar{\mathfrak{o}}$. Repeating this argument with b_1 instead of $\pi^{-m}a$, we find $a_1 \in A$ (also uniquely determined) such that $\pi^{-m}a = a_0 + a_1\pi + b_2\pi^2$ for some $b_2 \in \bar{\mathfrak{o}}$. We define a_n inductively for every $n = 0, 1, 2, \dots$ and the series $\sum_{n=0}^{+\infty} a_n\pi^n$ converges to $\pi^{-m}a$ since $v\left(\pi^{-m}a - \sum_{n=0}^{N} a_n\pi^n\right) = v(b_{N+1}\pi^{N+1}) \to 0$ as $N \to \infty$. □

Remark 2.3.27. If we assume $0 \in A$ in the proposition above, then $m = \mathrm{ord}_v(a)$ in (2.3.11) if and only if $a_0 \neq 0$. If we do not assume $0 \in A$ and only require that $a_0 \neq 0$, then uniqueness does not hold in general.

Until now, we have not made any assumption on the residue field $\mathfrak{o}/\mathfrak{p}$. We will see next that finiteness of $\mathfrak{o}/\mathfrak{p}$ has fruitful topological implications that enable us to pick a Haar measure on K.

Proposition 2.3.28. *Let v be a non-Archimedean valuation of K. Then K is locally compact if and only if all of the following hold:*

1. *v is discrete;*
2. *K is complete with respect to v;*
3. *The residue field is finite.*

Proof. Assume K satisfy conditions (1), (2) and (3). The ring of integers $\mathfrak{o}$ is a neighbourhood of 1. We will show that, any open cover $\{U_\lambda\}$ of $\mathfrak{o}$ admits a finite subcover. Suppose not. Let π be a uniformiser element, which exists since v is discrete, and let A be a set of representatives of the cosets in $\mathfrak{o}/\mathfrak{p}$. Since $\mathfrak{o}$ cannot be covered by finitely many sets in $\{U_\lambda\}$ and, on the other hand, $\mathfrak{o}$ is the disjoint union of finitely many cosets, there exists a_0 in A such that $a_0 + \mathfrak{o}\pi$ is not finitely covered by the sets in $\{U_\lambda\}$. Note that $\mathfrak{o}\pi$ is but the maximal ideal $\mathfrak{p}$. For the same reason. There exists $a_1 \in A$ for which $a_0 + a_1\pi + \mathfrak{o}\pi^2$ is not finitely covered by the sets in $\{U_\lambda\}$. If we proceed inductively, we define a sequence of elements $a_n \in A$ with the property that $a_0 + \cdots + a_m\pi^m + \mathfrak{o}\pi^{m+1}$ is not finitely covered by the sets in $\{U_\lambda\}$, for every $m = 0, 1, 2, \dots$. Note that $v(a_n\pi^n) \leq v(\pi)^n \to 0$ as $n \to +\infty$, so the series $\sum_{n=0}^{+\infty} a_n\pi^n$ converges to some element $a \in K$ since K is complete with respect to v. Moreover, $v(a) \leq \max_n v(a_n)v(\pi)^n \leq 1$, which means that a belongs to $\mathfrak{o}$ and therefore to some U_{λ_0}. But U_{λ_0} is open and $\{\mathfrak{o}\pi^n\}_{n\geq 0}$ is a neighbourhood basis for 1, so, for some $N \geq 0$, we have that $a + \mathfrak{o}\pi^N \subset U_{\lambda_0}$. It follows that $a_0 + \cdots + a_{N-1}\pi^{N-1} + \mathfrak{o}\pi^N = a + \mathfrak{o}\pi^N \subset U_{\lambda_0}$, a contradiction.

Conversely, suppose K is locally compact.

There exists a compact neighbourhood C of 1. For a sufficiently large n, $\mathfrak{o}\pi^n \subset C$. Since $\mathfrak{o}$ is closed, $\mathfrak{o}\pi^n$ is a closed subset of a compact set and is therefore compact, which implies that $\mathfrak{o}$ is compact as well. Since the topology of $\mathfrak{o}$ is metrizable, $\mathfrak{o}$ is sequentially compact. Now, given a Cauchy sequence $(x_n)_{n\geq 1}$, there exists n_0 such that $x_n \in x_{n_0} + \mathfrak{o}$ for every $n \geq n_0$. It follows that $(x_n)_{n\geq 1}$ has a converging subsequence and therefore converges, proving that K is complete.

Let A be a set of representatives for the cosets in $\mathfrak{o}/\mathfrak{p}$. Then $\{a+\mathfrak{p}\}_{a\in A}$ is a collection of open sets covering $\mathfrak{o}$ such that no subcollection still covers $\mathfrak{o}$. By compactness of $\mathfrak{o}$, this collection must then be finite, which means that A is finite and thus so is the residue field.

Finally, note that $\mathfrak{p}$ is compact being a closed subset of the compact set $\mathfrak{o}$ (indeed, since $\mathfrak{p}$ is open, its complement is also open being the union of translates of $\mathfrak{p}$, which makes $\mathfrak{p}$ closed). The collection of open sets $B_n = \{x \in \mathfrak{o} \mid v(x) < 1 - 1/n\}$, for $n = 1, 2, 3, \ldots$, cover $\mathfrak{p}$. By compactness, there must be some n_0 for which $\mathfrak{p} = B_{n_0}$. This means that $v(x) < 1 \implies x \in B_{n_0} \implies v(x) < 1 - 1/n_0$ and, by taking inverses, that $v(x) > 1 \implies v(x) > 1 + 1/(1 + n_0)$. Consequently, $1 - 1/n_0 \leq v(x) \leq 1 + 1/(1 + n_0) \implies v(x) = 1$, proving that v is discrete. □

2.3.7 Haar measure on locally compact fields

Let us recall that a Borel measure μ on a topological space T is a measure defined on the σ-algebra generated by the open subsets of T (the Borel σ-algebra). We say that μ is *regular* when

1. $\mu(E) = \inf\{\mu(U) \mid U \text{ is open and } E \subset U\}$ for every measurable set $E \subset T$;

2. $\mu(V) = \sup\{\mu(K) \mid K \text{ is compact and } K \subset V\}$ for every open set $V \subset T$.

If a regular Borel measure μ also satisfies

$$\mu(K) < \infty, \quad \text{for ever measurable compact set } K \subset T,$$

then μ is called a *Radon measure*.

Suppose G is a topological group. A *left Haar measure* on G is a Radon measure μ that is also left-invariant, meaning that, for every $g \in G$ and for every measurable subset $A \subset G$, one has that

$$\mu(gA) = \mu(A).$$

Alternatively, let L_g denote the left translation by g, which is the mapping defined by $h \mapsto gh$. Then μ is left-invariant if $(L_g)_*\mu = \mu$ for every $g \in G$, where $(L_g)_*$

denotes the push-forward by L_g, i.e, the measure defined by $A \mapsto \mu(g^{-1}A)$, for every measurable set $A \subset G$.

These measures are named after Alfréd Haar who introduced them in 1933 and proved his well known theorem:

Theorem 2.3.29 (Haar's Theorem)**.** *Let G be a locally compact Hausdorff topological group. There exists a left Haar measure μ on G that is unique up to scaling.*

A *right Haar measure* is defined analogously and the same result holds for right Haar measures. Indeed, let i denote inversion in G and consider the push-forward $i_*\mu$. Then μ is left-invariant if and only if $i_*\mu$ is right-invariant.

Now, let μ be a left Haar measure on G. For each $g \in G$, denote by R_g the right translations by g, i.e, the mapping defined by $h \mapsto hg$. Then $(R_g)_*\mu$ is also a left Haar measure since right translations and and left translations commute. By the uniqueness up to scaling, there exists a real number $\Delta(g) > 0$ such that $(R_g)_*\mu = \Delta(g)\mu$. In this way, we may define a function $\Delta : G \to \mathbb{R}_+$, called the *modular function* of G. Note that Δ does not depend on the choice of the left invariant Haar measure μ. It easy to see from the definition that Δ is a homomorphism from G to the multiplicative group of positive reals.

When Δ is constant equal to 1, the group G is said to be *unimodular*. In this case, every left-invariant measure is also right-invariant.

Remark 2.3.30. In the above discussion we have used the notation of a multiplicative group whereas, in the sequence, Haar's Theorem will be often applied to the additive group of a topological field. In any case, the group structure in consideration will always be made explicit and shall not lead to any confusion.

Suppose now that v is a discrete non-Archimedean valuation on a (non-discrete) field K, complete with respect to v, and such that the residue field $\mathfrak{o}/\mathfrak{p}$ is finite. Note that K is discrete if and only if v is trivial, which we tacitly assume not to be the case. By Proposition 2.3.28, we know that, in this case, K is a locally compact field.

Let K^+ denote the the topological group obtained from K with addition. It follows form Haar's Theorem that there exists a Haar measure μ in K^+. The group K^+ is unimodular since it is abelian, thus μ is bi-invariant. Let π be a uniformiser for v and let P be the cardinality of the residue field $\mathfrak{o}/\mathfrak{p}$. For any $\alpha \in K^+$ and any integer n, one has that

$$\alpha + \pi^n \mathfrak{o} = \coprod_{j=1}^{P} \alpha + c_j \pi^n + \pi^{n+1}\mathfrak{o},$$

where $c_1, \ldots, c_j$ are a set of representatives of the elements of the quotient $\mathfrak{o}/\mathfrak{p}$, and $\coprod$ indicates disjoint union. Since μ is translation invariant, it follows in particular that

$$\begin{aligned}\mu(\pi^n \mathfrak{o}) &= \sum_{j=1}^{P} \mu(c_j \pi^n + \pi^{n+1}\mathfrak{o}) \\ &= P\mu(\pi^{n+1}\mathfrak{o}),\end{aligned}$$

whence $\mu(\pi^n\mathfrak{o}) = P\mu(\pi^{n+1}\mathfrak{o})$. Using this relation inductively, one obtains that $\mu(\pi^n\mathfrak{o}) = P^{-n}\mu(\mathfrak{o})$.

We normalise μ such that $\mu(\mathfrak{o}) = 1$ and therefore

$$\mu(\pi^n\mathfrak{o}) = P^{-n}, \quad \text{for all } n \in \mathbb{Z}.$$

More generally, if $\beta \in K$, $\beta \neq 0$, we know that $\beta = \pi^n\epsilon$, where $n = \mathrm{ord}_v(\beta)$ and ϵ is such that $v(\epsilon) = 1$ (in particular, $\epsilon \in \mathfrak{o}$). Then, $\beta\mathfrak{o} = \pi^n\mathfrak{o}$ and $\mu(\beta\mathfrak{o}) = P^{-n}$.

Definition 2.3.31. We say that a discrete non-Archimedean valuation v on K, with residue field of finite cardinality P, is *normalised* if $v(\pi) = P^{-1}$. Note that in the case of $\mathbb{Q}_p$, this convention implies that $v_p(p) = p^{-1}$.

We have just proved the following:

Proposition 2.3.32. *Let K be a complete field with respect to a normalised valuation v, and let μ be the normalised Haar measure on K^+ (i.e., such that $\mu(\mathfrak{o}) = 1$). Then, for every $\beta \in K$, we have that*

$$\mu(\beta\mathfrak{o}) = v(\beta).$$

Alternatively, if we define the measure μ_β as $A \mapsto \mu(\beta A)$, then μ_β is still a Haar measure on K^+ and, by uniqueness it should be a rescaling of μ. The proposition says that

$$\mu_\beta = v(\beta)\mu, \tag{2.3.12}$$

for every $\beta \in K$.

When K is complete with respect to the Archimedean valuation v, it follows from Theorem 2.3.17 that $K = \mathbb{R}$ or $\mathbb{C}$ and that v is equivalent to the usual absolute value. In either case, the usual Lebesgue measure may be taken to be the Haar measure on K^+. In order for equation (2.3.12) to hold in this setting we make the following convention:

Definition 2.3.33. Let v be an Archimedean valuation on $K = \mathbb{R}$ or $\mathbb{C}$. We know that v is equivalent to the usual absolute value on K, that is, $v = |\cdot|^\lambda$ for some $\lambda > 0$, where $|\cdot|$ denote the usual absolute value. Then v is said to be *normalised* if

- $\lambda = 1$, i.e., $v(x) = |x|$ for $K = \mathbb{R}$, or
- $\lambda = 2$, i.e., $v(x) = |x|^2$ for $K = \mathbb{C}$.

2.3.8 Local and global fields

Definition 2.3.34. A *global field* is either one of the following:

(i) a number field, i.e., a finite (algebraic) extension of $\mathbb{Q}$;

(ii) a function field in one variable over a finite field, i.e., $\mathbb{F}_q(X)$ where q is a prime power[2].

The motivation for the above definition is the parallel treatment of number fields and function fields, which has been a beneficial analogy in the development of Algebraic Number Theory.

At this point, we have a good understanding of all the valuations that can be defined on a global field. This knowledge will allow us to fully characterise the completion of global fields with respect to such valuations, leading to the so called *local fields*. By the end of this subsection, we will have given three equivalent characterisations of local fields, any of which can be chosen as a definition. We start with ours:

Definition 2.3.35. A *local field* is one of the following:

(i) $\mathbb{R}$ or $\mathbb{C}$;

(ii) any finite (algebraic) extension of $\mathbb{Q}_p$;

(iii) the field of Laurent series $\mathbb{F}_q((X))$ over a finite field.

Note that every field in Definition 2.3.35 is locally compact. Indeed, $\mathbb{R}$ and $\mathbb{C}$ are clearly locally compact, while the fields in (ii) and (iii) are locally compact by Proposition 2.3.28 (recall that $\mathbb{F}_q((X))$ is the completion of $F_q(X)$ with respect to its infinite place). This topological property, in fact, turns out to be an alternative characterisation of local fields.

Theorem 2.3.36. *A (non-discrete) locally compact topological field is a local field.*

[2] It is worth noting that one may always assume that a finite extension $K \mid \mathbb{F}_q(X)$ is separable. The reason for this is that, since $\mathbb{F}_q$ is a perfect field, there exists an element $Y \in K$ (a separating transcendence basis), such that $K \mid \mathbb{F}_q(Y)$ is a separable extension (cf. [39, Theorem 26.3]).

Sketch of proof. There exists a Haar measure μ on the locally compact field K. For every automorphism ϕ of the additive group K^+, note that $\phi_*\mu$ is also left-invariant and therefore must be a multiple of μ. The positive real number $\mathrm{mod}_K(\phi)$ defined by

$$\phi_*\mu = \mathrm{mod}_K(\phi)\mu,$$

is called the module of ϕ. In particular, for a non-zero element $a \in K$, let $\mathrm{mod}_K(a)$ be the module of the automorphism induced by multiplication by a. This gives a homomorphism $\mathrm{mod}_K : K \to \mathbb{R}_{>0}$ (extended to 0 as 0). One easily checks that mod_K is a valuation on K.

It is important to note that the function $\mathrm{mod}_K : K \to \mathbb{R}_{\geq 0}$ is continuous (see, for example, [44, Proposition 4-1]) so that the topology on K induced by mod_K is also locally compact (it actually coincides with the original topology on K).

Local compactness of K implies completeness of the metric defined by the valuation mod_K, so K must contain the completion (with respect to mod_K) of its prime field.

If $\mathrm{char}K = 0$, the prime field of K is $\mathbb{Q}$ and, according to Ostrowski's Theorem (Theorem 2.3.15), the restriction of mod_K to $\mathbb{Q}$ must be either the usual absolute value on $\mathbb{Q}$ or a p-adic absolute value, whence K must contain either a copy of $\mathbb{R}$ or of $\mathbb{Q}_p$, respectively. Finally, local compactness also implies that K must be finite dimensional over this complete subfield.

If $\mathrm{char}K > 0$, we first observe that mod_K is non-Archimedean (Corollary 2.3.4) and non-trivial (otherwise K would be discrete). By Proposition 2.3.28, mod_K is discrete and $\mathfrak{o}/\mathfrak{p}$ is a finite field with, say, q elements, so $\mathfrak{o}/\mathfrak{p} \cong \mathbb{F}_q$. It follows from Proposition 2.3.26 that K is isomorphic to $\mathbb{F}_q((X))$, the field of formal Laurent series with coefficients in $\mathbb{F}_q$, just by mapping the element π to X.

□

Remark 2.3.37. The valuation mod_K for K equal to $\mathbb{R}$, $\mathbb{C}$ or $\mathbb{Q}_p$ coincides with our convention for a normalised valuation adopted earlier in Definitions 2.3.31 and 2.3.33.

As mentioned before, the completion of a global field with respect to some valuation is a local field. Indeed, the case of Archimedean valuations on number field clearly leads either to $\mathbb{R}$ or $\mathbb{C}$. In the remaining cases, one has a discrete non-Archimedean valuation v on a global field K. Both of these properties are transmitted to (the unique extension of) v on the completion K_v. By Proposition 2.3.28, K_v is locally compact, whence Theorem 2.3.36 implies that K_v must be one of the fields listed in Definition 2.3.35. Conversely, every local field arises as the completion of a global field, which brings us to the third characterisation of local fields. The proof of said equivalence is a neat application of Krasner's Lemma. We will sketch the proof of the equivalence without formally stating the lemma. For a more detailed exposition, we refer the reader to [35, Chapter 25].

Proposition 2.3.38. *Local fields are precisely the completion of global fields.*

Sketch of Proof. We have already established that the completion of global fields are local fields and shall now sketch the converse implication. The cases of $\mathbb{R}$, $\mathbb{C}$ or $\mathbb{F}_q((X))$ are clear and the only case that requires some justification is that of a finite extension of $\mathbb{Q}_p$. Let K be one such extension. We shall see in §§2.3.9 that there exists a unique extension of the p-adic valuation on $\mathbb{Q}_p$ to K, let us denote this extension by v_p. By the Primitive Element Theorem, K may be written as $\mathbb{Q}_p(\alpha)$ for some α. Let f be the minimal polynomial of α over $\mathbb{Q}_p$. By Krasner's Lemma, any $g \in \mathbb{Q}_p[X]$ with coefficients sufficiently close to the coefficients of f (in the p-adic metric) is separable, irreducible and of the same degree as f. Moreover, g has a root β such that $\mathbb{Q}_p(\beta) = \mathbb{Q}_p(\alpha) = K$. Since $\mathbb{Q}$ is dense in $\mathbb{Q}_p$, we may choose g in $\mathbb{Q}[X]$, so that β is algebraic over $\mathbb{Q}$. It follows that $\mathbb{Q}(\beta)$ is a number field which is clearly dense in $\mathbb{Q}_p(\beta) = K$ with respect to v_p, whence its completion yields K. □

2.3.9 Extension of valuations

In this subsection, we study the situation of a field extension $L \mid K$ when a valuation v is defined in K. In what ways, if any, can we extend v to L? We begin with the case of a finite separable extension, which cover most of the cases we are interested in and can be seen as a model for more general contexts.

Theorem 2.3.39. *Let v be a discrete non-Archimedean valuation on a field K and let $L \mid K$ be a finite separable extension. There exists a valuation w on L extending v, i.e., such that $w|_K = v$.*

Proof. As usual, let us denote by $\mathfrak{o}$ the ring of integers of v. Then K is the fraction field of $\mathfrak{o}$. Note that $\mathfrak{o}$ is a Dedekind domain. Indeed, given that v is discrete and non-Archimedean, it follows that $\mathfrak{o}$ is a principal ideal domain. It is a general fact that principal ideal domains are Dedekind domains and we will omit the proof of this statement (in fact, a ring is a principal ideal domain if and only if it is a Dedekind domain and a unique factorisation domain). It follows from Theorem 2.2.30 that the integral closure $\mathfrak{O}$ of $\mathfrak{o}$ in L is also a Dedekind domain and that there exists a non-zero prime ideal $\mathfrak{P}$ in $\mathfrak{O}$ (lying over $\mathfrak{p}$). Consider the valuation w' on L given by $w'(x) = c^{\mathrm{ord}_{\mathfrak{P}}(x)}$ where $0 < c < 1$ and $\mathrm{ord}_{\mathfrak{P}}(x)$ is the order of x with respect to $\mathfrak{P}$, i.e, $\mathrm{ord}_{\mathfrak{P}}(u)$ is the largest integer k for which $u \in \mathfrak{P}^k$ when u is an element of the ring $\mathfrak{O}$ and $\operatorname{ord}_{\mathfrak{P}}(u/v) = \operatorname{ord}_{\mathfrak{P}}(u) - \operatorname{ord}_{\mathfrak{P}}(v)$. The restriction $w'|_K$ is a valuation on K that satisfies

$$w'|_K(x) \leq 1 \iff x \in \mathfrak{O} \cap K = \mathfrak{o}.$$

It follows that $w'|_K$ and v have the same ring of integers and must then be equivalent (see the paragraph after Definition 2.3.20) so there exists $\lambda > 0$ such that $(w'|_K)^\lambda = v$. Define the valuation $w(x) = (c^\lambda)^{\operatorname{ord}_{\mathfrak{P}}(x)}$ for all $x \in L$. Then w extends v, as required. □

The previous result can be strengthened to include all valuations and all algebraic extensions, finite or otherwise. We state this generalisation next. A proof is given in the end of Appendix A.

Theorem 2.3.40. *Let $v : K \to \mathbb{R}_{\geq 0}$ be any (rank one) valuation on K. If $L \mid K$ is an algebraic field extension, then there exists an extension of v to L.*

In this setting, if the field K is complete with respect to v, the extension of v to L is unique. In order to prove this, we make use of a basic lemma in Real Analysis which we present here, adapted to our framework:

Lemma 2.3.41. *Let K be complete with respect to the valuation v and let V be a finite dimensional vector space over K. Any two norms on V are equivalent.*

Before we prove the lemma, let us recall, for the sake of clarity, some terminology.

Definition 2.3.42. Let V be be a vector space over the field K and let v be a valuation on K. A *norm* on V is a nonnegative real function $\|\cdot\|$ on V satisfying, for all $\xi, \eta \in V$ and all $a \in K$

(i) $\|\xi\| = 0$ if and only if $\xi = 0$;

(ii) $\|\xi + \eta\| \leq \|\xi\| + \|\eta\|$;

(iii) $\|a\xi\| = v(a)\,\|\xi\|$.

Any two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on V are said to be equivalent if there are constants $c, C > 0$ such that

$$c\,\|\xi\|_2 \leq \|\xi\|_1 \leq C\,\|\xi\|_2, \qquad \text{for all } \xi \in V.$$

In particular, $\|\cdot\|_1$ and $\|\cdot\|_2$ induce the same topology on V.

Proof of Lemma 2.3.41. Let $e_1, \cdots, e_n$ be a basis for V and denote by $|\cdot|$ the maximum norm, i.e., the norm given by $|\sum_i \xi_i e_i| = \max v(\xi_i)$. It is clearly sufficient to prove that any norm $\|\cdot\|$ is equivalent to $|\cdot|$. Put $C = \sum_i \|e_i\|$, then for any $\xi = \sum_i \xi_i e_i \in V$ we have that

$$\left\| \sum_{i=1}^{n} \xi_i e_i \right\| \leq \sum_{i=1}^{n} v(\xi_i) \|e_i\| \leq C|\xi|.$$

Now, suppose there is no $c > 0$ such that $c|\cdot| \leq \|\cdot\|$, then, for every $k = 1, 2, \ldots$ there exists $\xi_k = \sum_i \xi_{k,i} e_i \in V$ such that

$$0 < \|\xi_k\| \leq \frac{1}{k}|\xi_k|. \tag{2.3.13}$$

Without loss of generality, we may assume that $|\xi_k| = v(\xi_{k,n})$ for every k. Indeed, by restricting ourselves to a subsequence, if necessary, there exists an index $1 \leq i_0 \leq n$ such that $|\xi_k| = v(\xi_{k,i_0})$ for all $k = 1, 2, \ldots$, and then we relabel the basis so that $i_0 = n$. For each k, define $\eta_k = (\xi_{k,n})^{-1}\xi_k$ so that it may be written as $\eta_k = \sum_{i=1}^{n-1} \eta_{k,i} e_i + e_n$. It follows from (2.3.13) that $\|\eta_k\| < 1/k$ for every $k = 0, 1, 2, \ldots$ so that

$$\left\| \sum_{i=1}^{n-1} (\eta_{k,i} - \eta_{l,i}) e_i \right\| = \|\eta_k - \eta_l\| \to 0 \quad \text{as} \quad k, l \to \infty. \tag{2.3.14}$$

We then finish by induction on n. The lemma is trivial for $n = 1$. Suppose this is true for $n - 1$. Then (2.3.14) implies that

$$\left| \sum_{i=1}^{n-1} (\eta_{k,i} - \eta_{l,i}) e_i \right| \to 0 \quad \text{as} \quad k, l \to \infty,$$

and, consequently, that $(\eta_{k,i})_{k>1}$ is a Cauchy sequence in K for $i = 1, \ldots, n-1$. Since we are assuming that K is complete with respect to v, there exists $\eta_i^* \in K$ such that $\eta_{k,i} \to \eta_i^*$ for each $i = 1, \ldots, n-1$. But then

$$0 < \left\| \sum_{i=1}^{n-1} \eta_i^* e_i + e_n \right\| \leq \|\eta_k\| + \left\| \sum_{i=1}^{n-1} (\eta_i^* - \eta_{k,i}) e_i \right\| \to 0 \quad \text{as} \quad k \to \infty,$$

a contradiction. □

Theorem 2.3.43. *Let K be a complete field with respect to the valuation v and let $L \mid K$ be an algebraic extension. There exists a unique way of extending v to a valuation w on L.*

Moreover, if $L \mid K$ is a finite extension of degree $[L : K] = n$, then w is given by the formula

$$w(\alpha) = v\left(\mathrm{N}_{L|K}(\alpha)\right)^{1/n}, \tag{2.3.15}$$

for all $\alpha \in L$.

Remark 2.3.44. Although the proof of uniqueness given below works for both Archimedean and non-Archimedean v, we remark that existence and uniqueness in the Archimedean case are immediate consequences of Ostrowski's Theorem 2.3.17. Indeed, the theorem implies that either $L = \mathbb{C}$ and $K = \mathbb{R}$, in case L is strictly larger than K, or $L = K = \mathbb{R}$ or $\mathbb{C}$ otherwise, with v being the usual valuation (up to equivalence). The result follows.

Proof of Theorem 2.3.43.

Existence. Follows from Theorem 2.3.40.

Uniqueness. Suppose first that $L \mid K$ is finite. Let w_1 and w_2 be two extensions of v to L. Pick λ_1 and λ_2 such that $w_i^{\lambda_i}$ satisfies the triangle inequality, for $i = 1, 2$. Regarding L as a finite dimensional vector space over K, note that each $w_i^{\lambda_i}$ defines a norm on L, according to definition 2.3.42. It follows from Lemma 2.3.41 that $w_1^{\lambda_1}$ and $w_2^{\lambda_2}$ are equivalent as norms, and thus induce the same topology on L. Clearly, the topology induced by a valuation is the same topology induced by said valuation seen as a norm. It follows that the valuations $w_1^{\lambda_1}$ and $w_2^{\lambda_2}$ induce the same topology on L and, by Proposition 2.3.11, they are equivalent as valuations, so there exists $a > 0$ such that $w_1^{\lambda_1} = w_2^{a\lambda_2}$. Since w_1 and w_2 coincide when restricted to K, we find that $\lambda_1 = a\lambda_2$, thus $w_1 = w_2$.

For the general case, if w and w' are distinct extensions of v to L, there must be some $x \in L$ for which $w(x) \neq w'(x)$. Set $K' = K(x)$, then $w \mid_{K'}$ and $w' \mid_{K'}$ are distinct extensions of v to the finite-dimensional field extension $K' \mid K$, contradicting our previous argument.

Formula. Assuming existence and uniqueness, we derive formula (2.3.15). Note that even without the proof of existence, the reasoning that follows provides an educated guess for what the extended valuation should look like, if it exists. And then one can proceed to verify that (2.3.15) indeed defines a valuation, as done by Cassels in [10, Chapter 7].

Suppose first that $L \mid K$ is normal. Let w be an extension of v to L. For an automorphism σ of L fixing K, note that $w_\sigma(\alpha) = w(\sigma\alpha)$ defines an extension of v and thus, by uniqueness, it must be equal to w, i.e., $w(\sigma\alpha) = w(\alpha)$ for all $\alpha \in L$ and all $\sigma \in \operatorname{Aut}(L \mid K)$. Since $\mathrm{N}_{L|K}(\alpha) = \prod_{\sigma\in\operatorname{Aut}(L|K)} \sigma\alpha$ and since w extends v, we have that

$$v\left(\mathrm{N}_{L|K}(\alpha)\right) = w(\alpha)^n,$$

from where the formula follows. For a general finite extension $L \mid K$, let $L' \mid K$ be the smallest normal extension containing $L \mid K$ as a subextension. Then (2.3.15) provides an extension of v to L':

$$w'(x) = v\left(N_{L'|K}(x)\right)^{1/[L':K]} \quad \text{for all } x \in L',$$

and its restriction to L provides an extension of v to L. Note, however, that for $\alpha \in L$, one has that

$$\mathrm{N}_{L'|K}(\alpha) = \left(\mathrm{N}_{K(\alpha)|K}(\alpha)\right)^{[L':K(\alpha)]} = \left(\mathrm{N}_{L|K}(\alpha)\right)^{\frac{[L':K(\alpha)]}{[L:K(\alpha)]}} = \left(\mathrm{N}_{L|K}(\alpha)\right)^{[L':L]},$$

and so

$$w'(\alpha) = v\left(\mathrm{N}_{L|K}(\alpha)\right)^{\frac{[L':L]}{[L':K]}} = v\left(\mathrm{N}_{L|K}(\alpha)\right)^{\frac{1}{[L:K]}}.$$

□

Corollary 2.3.45. *Let L, K, v, w be as in the theorem, with $L \mid K$ finite. Suppose further that K is locally compact, so it makes sense to consider a normalised valuation on K (see Definition 2.3.31). If v is assumed to be normalised, then the normalised valuation w' in the equivalence class of w is given by*

$$w'(\alpha) = v(\mathrm{N}_{L|K}(\alpha)) \tag{2.3.16}$$

Proof. By (2.3.15), we know that $w'(\alpha) = v(\mathrm{N}_{L|K}(\alpha))^{\lambda}$ for some $\lambda > 0$. Let μ be a Haar measure on K^+. The product measure $\mu \otimes \cdots \otimes \mu$ is clearly a Haar measure on $\bigoplus_{i=1}^{n} K^+ \cong L$. Let $A \subset K$ be a measurable subset with non-zero measure and let $b \in K$. It follows that

$$\mu \otimes \cdots \otimes \mu \left(b \bigoplus_{i=1}^{n} A\right) = \prod_{i=1}^{n} \mu(bA) = v(b)^n \mu \otimes \cdots \otimes \mu \left(\bigoplus_{i=1}^{n} A\right).$$

On the other hand, $\mu \otimes \cdots \otimes \mu \left(b \bigoplus_{i=1}^{n} A\right) = w'(b) \mu \otimes \cdots \otimes \mu \left(\bigoplus_{i=1}^{n} A\right)$, and thus $w'(b) = v(b)^n = v(\mathrm{N}_{L|K}(b))$, so that $\lambda = 1$. □

Let (K, v) be a non-Archimedean valued field. There is a standard way of extending v to the field $K(X)$ of rational functions in one variable X and coefficients in K. Given a non-zero polynomial in $P \in K[X]$ such that $P(X) = \sum_{i=0}^{n} a_i X^i$, $a_1, \ldots, a_n \in K$, define $w(P) := \max\limits_{0 \le i \le n} v(a_i)$, known as the *Gauß norm*. It is easy to check that w satisfies the properties of a valuation on the ring $K[X] \setminus \{0\}$. For instance, let us check multiplicativity of w: Let $P(X) = \sum_{i=0}^{n} a_i X^i$, $Q(X) = \sum_{i=0}^{n} b_i X^i$, let a_{n_0}, b_{m_0} be such that $w(P) = v(a_{n_0})$ and $w(Q) = v(b_{m_0})$, where n_0 and m_0 are minimal with this property. We then, look at the coefficient of the $X^{n_0+m_0}$ term of PQ, namely, $\sum_{i=0}^{n_0+m_0} a_i b_{n_0+m_0-i}$. For every $i \neq n_0$, note that $v(a_i b_{n_0+m_0-i}) < v(a_{n_0} b_{m_0})$, whence $v(\sum_{i=0}^{n_0+m_0} a_i b_{n_0+m_0-i}) = v(a_{n_0} b_{m_0}) = w(P)w(Q)$ and, consequently, $w(P)w(Q) \le w(PQ)$. On the other hand, it follows from the triangle inequality (of v) that $w(PQ) \le w(P)w(Q)$. Now, since $K(X)$ is the field of fractions of $K[X]$, we may extend w to $K(X)$ by defining $w(P/Q) = w(P)/w(Q)$. Then w is a valuation on $K(X)$ extending v. One can develop this even further in the following manner: suppose $L \mid K$ is a purely transcendental extension.

Consider the collection of all valued fields (E, v') such that $E \mid K$ is a subextension and v' extends v, partially ordered by extension ($(E, v') \preceq (F, w')$ if and only if $F \mid E \mid K$ and w' extends v'). This collection is non-empty since $(K(X), w)$ is in it, and it is easily seen to satisfy the chain condition. By Zorn's Lemma, there exist a maximal element (M, u). We claim that $M = L$. If not, we could use the Gauß norm construction, as described above, to extend u to $M(T)$ (for some indeterminate $T \in L$ not in M), which violates maximality of (M, u). We have just proved the following:

Proposition 2.3.46. *Let (K, v) be a non-Archimedean valued field. If $L \mid K$ is a purely transcendental extension, then the valuation v can be extended to L.*

Combining Theorem 2.3.40 and Proposition 2.3.46 together, we obtain the following extension theorem:

Theorem 2.3.47 (Extension of non-Archimedean valuations)**.** *Let (K, v) be a non-Archimedean valued field. For* any *field extension $L \mid K$, there exits a valuation w on L extending v.*

Proof. Field theory says there exists a transcendence basis S of $L \mid K$, so that $K(S) \mid K$ is purely transcendental while $L \mid K(S)$ is algebraic. It follows from Theorem 2.3.40 and Proposition 2.3.46 that v can be extended first to $K(S)$ and then to L. □

Example 2.3.48. In the case where $L \mid K$ is not algebraic and the base field K is endowed with an Archimedean valuation v, the existence of extension is not guaranteed:

(i) Take, for instance, $\mathbb{R} \mid \mathbb{Q}$. The usual absolute value $|\cdot|_\infty$ on $\mathbb{Q}$ is an Archimedean valuation that can obviously be extended to $\mathbb{R}$ (as the usual absolute value on $\mathbb{R}$). In particular, it can also be extended to any purely transcendental field extension of $\mathbb{Q}$ that is intermediate to $\mathbb{R} \mid \mathbb{Q}$.

(ii) Let $L = \mathbb{R}(X)$ and consider the purely transcendental field extension $L \mid \mathbb{R}$. Take v to be the absolute value on $\mathbb{R}$. We claim that v cannot be extended to a valuation on L. Indeed, suppose w on L extends v. Then w can be naturally extended to a valuation (which we will keep denoting by w) on the completion $\overline{L}$ of L with respect to w. So $(\overline{L}, w)$ is a complete Archimedean valued field and, by Ostrowski's Theorem 2.3.17, must be isomorphic to $\mathbb{R}$ or $\mathbb{C}$ with its usual absolute value. In particular, $\overline{L} \mid \mathbb{R}$ must be algebraic, contrary to our assumptions.

When K is not complete with respect to v, there might exist multiple ways of extending v. We analyse the case of a finite *separable* extension $L \mid K$. Regarding L as a vector space over the field K, we may extend scalars to the completion K_v and simultaneously produce all (the finitely many) completions of L with respect to the different possible extensions of v.

Theorem 2.3.49. *Let $L \mid K$ be a finite separable extension of degree $[L : K] = n$ and let v be a valuation on K. There exist N extensions $v_1, \ldots, v_N$ of v to L, for nome $1 \leq N \leq n$. Moreover,*

$$K_v \otimes_K L \cong \bigoplus_{j=1}^{N} L_j, \tag{2.3.17}$$

where each L_j is the completion of L with respect to v_j, and both sides are regarded as topological K_v-vector spaces.

In particular, $\sum_{j=1}^{N} [L_j : K_v] = [L : K]$.

Proof. Since $L \mid K$ is assumed to be finite and separable, it follows from the Primitive Element Theorem that $L = K(\alpha)$ for some $\alpha \in L$. Let $f(X) \in K[X]$ be the minimal polynomial of α over K. Over K_v, f factors as

$$f(X) = \phi_1(X) \cdots \phi_N(X),$$

where $\phi_1, \ldots, \phi_N$ are irreducible in $K_v[X]$ and pairwise distinct (since L is separable over K). For each $j = 1, \ldots, N$, let α_j be a root of ϕ_j (in the algebraic closure of K_v) and put $L_j = K_v(\alpha_j)$. We have, from the Chinese Remainder Theorem, that

$$K_v[X]/(f(X)) \cong \bigoplus_{j=1}^{N} K_v[X]/(\phi_j(X)),$$

as rings. Now, since ϕ_j is the minimal polynomial of α_j, each $K_v(X)/(\phi_j(X))$ is field-isomorphic to $K_v(\alpha_j) = L_j$. On the other hand, $K_v \otimes_K K[X]/(f(X))$ is easily seen to be ring-isomorphic to $K_v[X]/(f(X))$, and the former is but $K_v \otimes_K L$, which gives us the desired isomorphism of the algebraic structures in (2.3.17). Note that, at this point, each L_j is merely $K_v(\alpha_j)$, as defined above. We have yet to prove that these are completions of K with respect to different valuations extending v.

Note that L is canonically embedded into each L_j. Indeed, L is naturally identified with a subset of $K_v \otimes_K L$ and we define a ring homomorphism $\mu_j : K_v \otimes_K L \to L_j$ by mapping any polynomial expression $g(\alpha) \in K_v \otimes_K L$ to $g(\alpha_j)$. Let us denote by λ_j the homomorphism thus defined:

$$\lambda_j : L \hookrightarrow K_v \otimes_K L \xrightarrow{\mu_j} L_j. \tag{2.3.18}$$

Since every L_j is a finite (separable) extension of the complete field K_v, it follows from Theorem 2.3.43 that there exists a unique extension of v (defined on K_v) to the valuation v_j^* on L_j, and then (2.3.18) induces a valuation v_j on L extending K. Moreover, L_j is the completion of L with respect to v_j. Indeed, $L \cong K \otimes_K L$ is dense in $K_v \otimes_K L$, and μ_j, being a surjective K_v-linear map between finite-dimensional normed K_v-vector spaces, is a surjective continuous map. It follows that $L \cong \lambda_j(L)$ is a dense subset of L_j (with the topology induced from v_j^*).

We must show next that the valuations v_j defined above are all the valuations on L extending v and that they are pairwise distinct, producing, in this way, N different extensions for v on L. Let w be any valuation on L extending v. Note that w is defined on the dense subset $K \otimes_K L \subset K_v \otimes_K L$ and thus we may extend w continuously to a nonnegative function w on $K_v \otimes_K L$ satisfying

(i) $w(xy) = w(x)w(y)$;

(ii) $w(x + y) \leq C \max\{w(x), w(y)\}$;

for every $x, y \in K_v \otimes_K L$ where C is a positive constant. Suppose there exists some $x \in L_j$ such that $w(x) \neq 0$. Then $w(y) \neq 0$ for all $y \in L_j$, $y \neq 0$, since $w(x) = w(y)w(y^{-1}x)$. It follows that the restriction of w to K_j, $w|_{L_j}$, is either identically 0 or it defines a valuation on L_j extending v, in which case it must be equal to v_j^*, by uniqueness, and therefore it restricts to L as v_j. Furthermore, w cannot induce a valuation on L_i and L_j ($i \neq j$) simultaneously since, for $x \in L_i$ and $y \in L_j$, one has that $xy = (0, \dots, 0, x, 0, \dots, 0) \cdot (0, \dots, 0, y, 0, \dots, 0) = (0, \dots, 0)$, where x occupies the i-th position and y occupies the j-position in their respective vector representation, and thus $w(x)w(y) = 0$. This implies that, for $i \neq j$, v_i and v_j are different (and therefore non-equivalent, since they both extend v). □

Corollary 2.3.50. *With the same notation as in the theorem, let $x \in L$. Then*

$$\operatorname{Tr}_{L|K}(x) = \sum_{j=1}^{N} \operatorname{Tr}_{L_j|K_v}(x), \qquad \textit{and} \qquad \mathrm{N}_{L|K}(x) = \prod_{j=1}^{N} \mathrm{N}_{L_j|K_v}(x). \tag{2.3.19}$$

Proof. Indeed, $\operatorname{Tr}_{L|K}$ and $\mathrm{N}_{L|K}$ are respectively the trace and determinant of the K-linear transformation on L (seen as a K-vector space) induced by multiplication by x. Note that these are the same trace and determinant of the K_v-linear transformation on $K_v \otimes_K L$ corresponding to multiplications by x. Given the isomorphism (2.3.17), the result follows. □

Corollary 2.3.51. *With the same notation as in the theorem, let $x \in L$. Then*

$$v(\mathrm{N}_{L|K}(x)) = \prod_{j=1}^{N} v_j(x)^{[L_j:K_v]}.$$

Proof. This is a direct consequence of relation (2.3.19) and formula (2.3.15). □

Corollary 2.3.52. *With the same notation as in the theorem, suppose v is normalised and let v'_j denote the normalised valuation extending v to L_j. Then, for every $x \in L$,*

$$v(\mathrm{N}_{L|K}(x)) = \prod_{j=1}^{N} v'_j(x).$$

Proof. It follows straight from (2.3.19) and formula (2.3.16). □

In the particular case of a finite extension of $\mathbb{Q}$, i.e., of a number field L, Theorem 2.3.49 allows us to characterise all possible Archimedean valuations on L. The number of such valuations is related to the nature of the embeddings of L into $\mathbb{C}$. Let r_1 denote the number of those embeddings whose images are contained in $\mathbb{R}$, i.e., real embeddings. All the other embeddings of L into $\mathbb{C}$ are complex embeddings. Recall that, for a complex embedding $\sigma : L \to \mathbb{C}$, one can define another embedding η as $\eta(x) = \overline{\sigma(x)}$ for every $x \in L$. Then σ and η are said to be complex-conjugate embeddings. The complex embeddings of L come in pairs of complex-conjugate embeddings. Let r_2 denote the number of these pairs and note that $[L : \mathbb{Q}] = r_1 + 2r_2$. With this notation, we state the following:

Corollary 2.3.53 (Characterisation of Archimedean valutions on number fields)**.** *If $L = \mathbb{Q}(\alpha)$ is a number field of degree $[L : \mathbb{Q}] = r_1 + 2r_2$, then every Archimedean valuation on L is (up to equivalence) of the form $v_\sigma(x) = |\sigma x|_\infty$, for $x \in L$, where $|\cdot|_\infty$ denotes the ordinary absolute value on $\mathbb{C}$ and σ is an embedding of L into $\mathbb{C}$.*

Furthermore, for $\sigma \neq \eta$, v_σ is equivalent to v_η if and only if σ and η are complex-conjugates. In particular, the number of equivalence classes of Archimedean valuations on L is $r_1 + r_2$.

Proof. We apply Theorem 2.3.49 (more precisely, its proof) with $K = \mathbb{Q}$ and $v = |\cdot|_\infty$, in which case $K_v = \mathbb{R}$. Let f be the minimal polynomial of α over $\mathbb{Q}$. We see that there are N extensions of $|\cdot|_\infty$ to L, where N is the number of irreducible factors of f over $\mathbb{R}$. Let v_j be one of these extensions. We know that the completion of L with respect to v_j, denoted in the proof of the theorem by L_j, must be $\mathbb{R}(\alpha_j)$ where α_j is some root of the factor ϕ_j of f. So L_j is either $\mathbb{R}$ or $\mathbb{C}$. In either case, since L is naturally embedded in L_j, it is then naturally embedded in $\mathbb{C}$ by some embedding, say, σ. Moreover, v_j was given in the theorem by the restriction to L of the valuation on L_j extending the one on $\mathbb{R}$ $(= K_v)$. Note that, in this case, said valuation on L_j must be the usual absolute value of $\mathbb{C}$ (or its restriction to $\mathbb{R}$). It follows that v_j is precisely v_σ.

Now, it follows from basic field theory that r_1 is the number of real roots of f while r_2 is the number of pairs of complex-conjugate roots of f. Note also that the number of factors of f is $N = r_1 + r_2$. For each one of the N factors of f, we obtained an extension v_σ of $|\cdot|_\infty$, and the theorem says that these N extensions are pairwise non-equivalent. So if v_σ is equivalent to v_η, for $\sigma \neq \eta$, then σ and η must be two embeddings of L into $\mathbb{C}$ arising from the same (quadratic) factor of f, in the process described above. It can only be the case that σ and η are complex-conjugates. Conversely, if they are complex-conjugates, then it follows straight from the definition that $v_\sigma = v_\eta$. □

We finish by characterising all non-Archimedean valuations on number fields. Just as every non-Archimedean valuation on $\mathbb{Q}$ is given by a p-adic valuation (Ostrowski's Theorem 2.3.15), the same holds for general number fields. We have now developed more than enough tools to verify this.

Proposition 2.3.54 (Characterisation of non-Archimedean valuations on number fields)**.** *On a number field L, every non-Archimedean valuation v is equivalent to a $\mathfrak{p}$-adic valuation $v_\mathfrak{p}$ for some prime ideal $\mathfrak{p}$ of the ring of integers $\mathscr{O}_L$.*

Proof. Let $\mathfrak{o}$ be the valuation ring of v and let $\mathfrak{m}$ denote its maximal ideal. Since $1 \in \mathfrak{o}$ and $\mathfrak{o}$ is integrally closed (Proposition A.0.2 1, in Appendix A), it follows that $\mathscr{O}_L \subset \mathfrak{o}$. The intersection $\mathfrak{p} = \mathfrak{m} \cap \mathscr{O}_L$ is a prime ideal of $\mathscr{O}_L$. Consider the $\mathfrak{p}$-adic valuation $v_\mathfrak{p}$, as described in Example 2.3.6(ii). Its valuation ring is clearly the localisation of $\mathscr{O}_L$ at $\mathfrak{p}$, i.e., the ring $\{\frac{a}{b} \in L \mid a, b \in \mathscr{O}_L,\ b \notin \mathfrak{p}\}$. Given $\frac{a}{b}$ in this ring, since $b \in \mathscr{O}_L \setminus \mathfrak{p}$ and $\mathfrak{p} = \mathfrak{m} \cap \mathscr{O}_L$, we see that b is a unit in $\mathfrak{o}$ and therefore $\frac{a}{b} \in \mathfrak{o}$. It follows that the valuation ring of $v_\mathfrak{p}$ is contained in $\mathfrak{o}$ or, in other words, that $v_\mathfrak{p}(x) \leq 1 \implies v(x) \leq 1$ for $x \in L$. This shows that v is equivalent to $v_\mathfrak{p}$ (cf. Proposition 2.3.12). □

The main extension theorems studied in this subsection are summarised in Table 2.1: the cases of Archimedean and non-Archimedean valuations are listed in the rows, while columns discriminate the types of fields extension.

	Algebraic	Purely Transcendental
Archimedean	🟩 Thm. 2.3.40 (Ostrowski's Thm. 2.3.17)	🟥 Example 2.3.48
non-Archimedean	🟩 Thm. 2.3.40	🟩 Prop. 2.3.46

Table 2.1: The green colour indicates that the valuation can be extended. The red indicates cases in which an extension does not always exist.

CHAPTER 3
QUATERNION ALGEBRAS

3.1 Central simple algebras

Definition 3.1.1. Let R be a commutative ring. An *R-algebra A* is a ring together with a ring homomorphism $f : R \to Z(A)$, where, as usual, $Z(A)$ denotes the centre of A. Note that, for $r \in R$ and $a \in A$, one may naturally define the scalar product $ra = f(r)a$, giving A the structure of a module over R. Furthermore, the R-module structure is compatible with the ring structure of A, namely

$$(ra)b = r(ab) = a(rb), \qquad \text{for all } a, b \in A \text{ and } r \in R.$$

When R is a field K, the algebra is, in particular, a vector space over K. This is the case with which we will be mostly concerned and, henceforth, every algebra will be assumed to be defined over a field. The field of definition will often be omitted when there is no risk of confusion.

Example 3.1.2. Some examples of algebras are:

(i) Every ring (with unity) A is a $\mathbb{Z}$-algebra. The homomorphism f is, in this case, the unique homomorphism mapping 1 to the unity in A.

(ii) The set of functions from a set X taking values in a field K is a ring where addition and multiplication are defined pointwise. This ring can be made into a K-algebra by considering the homomorphism that maps $a \in K$ to the constant function $g_a(x) = a$ for all $x \in X$.

(iii) The ring $M_n(K)$ of $n \times n$ matrices with coefficients in a field K is a K-algebra, with the homomorphism $f : K \to M_n(K)$ given by $k \to k\mathrm{Id}_n$, where Id_n denotes the identity matrix.

(iv) Let A be a ring and let M be an A-module. Consider the set $\mathrm{End}_A(M)$ of all endomorphisms of M, i.e., all A-homomorphisms from M to M. This set is clearly a ring, the *ring of endomorphisms of M*, where addition is defined pointwise and multiplication is given by composition. If A is itself an algebra over the field K, then $\mathrm{End}_A(M)$ is also an algebra over K. Moreover, if A is a division algebra (see Definition 3.1.3 below) and M is finitely generated over A then, by choosing a basis, we see that $\mathrm{End}_A(M) \cong M_n(A)$.

Definition 3.1.3. A *division ring* is a ring in which every non-zero element has a multiplicative inverse, i.e, where division can be carried out. An algebra D is said to be a *division algebra* when the underlying ring structure is a division ring. For finite-dimensional algebras, this is easily seen to be equivalent to not having zero divisors (that is, $ab = 0$ implies that $a = 0$ or $b = 0$).

Definition 3.1.4. Let A be an algebra defined over the field K. Then

(i) A is *central* if $Z(A) = K$;

(ii) A is *simple* if it has no proper nontrivial two-sided ideals.

Remark 3.1.5. In the theory of modules it is also common to define a *simple R-module* as one that has no nontrivial proper submodule. The reader is cautioned that, for an R-algebra A, these two notions are *not* the same in general. Indeed, if an R-algebra A is regarded as a left (or right) A-module, then a submodule of A is just a left (or right) ideal, and not necessarily a two-sided ideal. For this reason, an algebra A may be simple as an algebra but not simple as an A-module.

Example 3.1.6. (i) Every division algebra is simple;

(ii) Let D be a division algebra over K. Its centre, $Z(D)$, is a field. Indeed, it is commutative and closed under addition, multiplication and taking inverses. Therefore, if we regard D as an algebra over $Z(D)$, it becomes a central simple algebra.

(iii) If D is a division algebra (over K) then $M_n(D)$ is a simple algebra (over K). For $1 \leq, i, j \leq n$, let E_{ij} denote the elementary $n \times n$ matrix with coefficient 1 in the entry (i, j) and coefficient zero elsewhere. For a non-zero matrix M, denote by $\langle M \rangle$ the two-sided ideal of $M_n(D)$ generated by M. There must be some entry (i, j) of M, say m, which is non-zero. Then $E_{ij} = m^{-1} E_{ii} M E_{jj} \in \langle M \rangle$. Since $E_{kl} = E_{ki} E_{ij} E_{jl} \in \langle M \rangle$, for $1 \leq k, l \leq n$, and since any matrix in $M_n(D)$ is a D-linear combination of elementary matrices, it follows that $\langle M \rangle = M_n(D)$.

The reciprocal of Example 3.1.6 (iii) also holds, and it is one of the most important results in the theory of central simple algebras. The reader may find a thorough exposition of this theorem and its proof either in [23, Section 2.1] or in [35, Chapter 29]. We merely state it here:

Theorem 3.1.7 (Wedderburn's Structure Theorem)**.** *For a finite-dimensional simple K-algebra A, there exists a division algebra D over K and an integer n such that $A \cong M_n(D)$. Moreover, n is uniquely determined and so is D (up to isomorphism).*

Corollary 3.1.8. *If K is algebraically closed then a central simple algebra A over K is isomorphic to $M_n(K)$.*

Proof. Let D be a division algebra over K. In particular, $K \subset D$. Suppose this inclusion is strict and let $d \in D \setminus K$. Since D is finite-dimensional over K, the set $\{1, d, d^2, d^3, \dots\}$ cannot be linearly independent and thus there exists a polynomial $F \in K[X]$ such that $F(d) = 0$. Now, D is a division algebra, so we may take an irreducible factor $f \in K[X]$ of F such that $f(d) = 0$. It follows from irreducibility that the ideal generated by f in $K[X]$ is maximal, so that $K[X]/(f)$ is a field. Moreover, the K-algebra homomorphism $K[X] \to D$ taking X to d induces an embedding of $K[X]/(f)$ into D whose image contains d, proving that d is algebraic over K. Being K algebraically complete, we have that $d \in K$, contradicting our assumption.

We conclude from the argument above that the only division algebra over an algebraically closed field K is K itself. The corollary then follows from the Wedderburn Structure Theorem. □

Another fundamental result in the theory of central simple algebras is the Skolem-Noether Theorem, which characterises automorphisms of central simple algebras. A proof of this theorem may be found in [36, Theorem 2.9.8].

Theorem 3.1.9 (Skolem-Noether Theorem)**.** *Let A and B be finite-dimensional simple algebras over K. Furthermore, suppose B is also central. If $f, g : A \to B$ are algebra homomorphisms then there exists an invertible element $b \in B$ such that*

$$f(a) = bg(a)b^{-1}, \qquad \text{for all } a \in A.$$

Corollary 3.1.10. *If A is a finite-dimensional central simple algebra over K, then every endomorphism of A is inner. This is to say that, for every endomorphism $f : A \to A$, there exists an invertible element $c \in A$ such that*

$$f(a) = cac^{-1}, \qquad \text{for all } a \in A.$$

3.2 A brief review of quadratic spaces

Throughout this section we assume that the characteristic of K is $\neq 2$.

Definition 3.2.1. Let V be a finite-dimensional K-vector space over a field K and $B : V \times V \to K$ a symmetric bilinear map. The pair (V, B) is said to be a *quadratic space*. When the bilinear map is clear from the context, we abuse notation and refer to the quadratic space V.

The map B determines a *quadratic map* $q : V \to K$ given by $q(v) = B(v, v)$, which immediately implies that $q(av) = a^2 q(v)$ for all $a \in K$ and $v \in V$.

Note that B and q are related by the *polarisation identity*:

$$2B(v, w) = q(v + w) - q(v) - q(w),$$

so one can also determine a quadratic space by the pair (V, q). Again, when it is clear from the context, we may omit the quadratic map and simply refer to the quadratic space by V.

Choosing a basis $\{v_1, \ldots, v_n\}$ of V, we obtain a *quadratic form* on n variables, which we also denote by q, given by

$$q(x_1, \ldots, x_n) = \sum_{i,j} B(v_i, v_j) x_i x_j,$$

with associated symmetric matrix $M = (B(v_i, v_j))_{ij}$. A change of basis gives rise to a congruent symmetric matrix.

Two quadratic forms over K with associated matrices M and M' are *equivalent* if they are congruent, i.e, if there exists a non-singular matrix $X \in \mathrm{GL}(n, K)$ such that

$$M' = X^t M X. \tag{3.2.1}$$

Motivated by the classical Euclidean space, we say that a K-isomorphism $\tau : (V, B) \to (V', B')$ between quadratic spaces is an *isometry* if it preserves the bilinear map, i.e., if $B'(\tau(v), \tau(w)) = B(v, w)$ for all $v, w \in V$. Note that equivalence of quadratic forms, as described in (3.2.1), is equivalent to the isometry of the corresponding quadratic spaces.

Two vectors $v_1, v_2 \in V$ are *orthogonal* if $B(v_1, v_2) = 0$. Similarly, two subspaces W_1, W_2 of V are orthogonal when every vector in W_1 is orthogonal to every vector in W_2. If, furthermore, $V = W_1 \oplus W_2$, then we say that (V, B) can be decomposed into the *orthogonal summands* W_1 and W_2, and write $V = W_1 \widehat{\oplus} W_2$.

For a subspace $W \subset V$, we denote by $W^{\perp}$ the subspace of V consisting of all vectors orthogonal to W. The subspace $V^{\perp}$ itself is called the *radical* of V, and is denoted by $\mathrm{rad}(V)$. When $\mathrm{rad}(V) = \{0\}$, then (V, B) is said to be *regular* and the map B, as well as the quadratic map q, are said to be *non-degenerate*. Equivalently, (V, B) is regular if the dual map $v \mapsto B(\,\cdot\,, v)$, from V to V^{*}, is an isomorphism. This corresponds to a quadratic form whose matrix M is non-singular. Note that, when V is not regular, then $V = \mathrm{rad}(V) \mathbin{\widehat{\oplus}} W$, where $\mathrm{rad}(V)$ is the kernel of the dual map and W (with the quadratic map restricted to it) is a regular subspace.

A vector $v \neq 0$ is called *isotropic* if $q(v) = 0$, and *anisotropic* otherwise. Likewise, a subspace is called isotropic if it contains an isotropic vector and is called anisotropic otherwise. More generally, when there exists a non-zero $v \in V$ such that $q(v) = a$, we say that (V, B) *represents* $a \in K$.

Note that, when a regular quadratic space V represents $a \neq 0$ then it decomposes as $V = \langle v \rangle \mathbin{\widehat{\oplus}} \langle v \rangle^{\perp}$, where v is such that $q(v) = a$. By repeating this we obtain the following lemma:

Lemma 3.2.2. *If (V, B) is a quadratic space over K, then V has an orthogonal basis $\{v_1, \ldots, v_n\}$ with respect to which the associated matrix M is diagonal, i.e., every quadratic form is equivalent to a diagonal form $d_1 x_1^2 + \cdots + d_n x_n^2$.*

Proof. First decompose V as $\mathrm{rad}(V) \mathbin{\widehat{\oplus}} W$, where W is regular. Then apply the recurrence described above. □

The set of all isometries from V to itself form a group called the *orthogonal group* of (V, B), denoted by $O(V, B)$, or simply by $O(V)$, when there is no risk of confusion. Given a vector $v \in V$ such that $q(v) \neq 0$, the reflection across the subspace perpendicular to v is the isometry τ_v defined by:

$$\tau_v(x) = x - \frac{2B(x, v)}{q(v)} v. \tag{3.2.2}$$

The famous Cartan-Dieudonné Theorem states that, for a regular quadratic space (V, B) of dimension n over a field of characteristic $\neq 2$, the group $O(V, B)$ is generated by reflections. Moreover, every isometry is generated by at most n reflections.

Being V a vector space over K, given any field extension $L \mid K$, one may extend scalars and obtain the L-vector space $V \otimes L$. Naturally, the same B may now be regarded as a symmetric bilinear form on $V \otimes L$. In particular, when K is a number field, we can take L to be a completion K_w of K, with respect to a place w of K, in which case we denote the resulting quadratic space by (V_w, B). The quadratic spaces over all local fields, together, give information on the quadratic space over the global field. This is (an instance of) the *local-global principle* and we will content ourselves with only stating the following central result:

Theorem 3.2.3 (Haße-Minkowski Theorem)**.** *Let (V, B) be a vector space over the number field K. Then V is isotropic over K if and only if V_w is isotropic over* every *completion K_w of K.*

The Haße-Minkowski Theorem says that a quadratic form represents 0 in K, i.e, the equation $q(x_1, ..., x_n) = 0$ has a non-trivial solution over K, if and only if it has a non-trivial solution over K_w for every place w of K. The same result also holds when representing any other element of K:

Corollary 3.2.4. *Let (V, q) be a quadratic space over the number field K and let $a \in K$. Then V represents a if and only if V_w represents a for every place w of K.*

Proof. If $a = 0$, this is but the Haße-Minkowski Theorem. Suppose $a \neq 0$. Necessity is immediate, so we need only prove sufficiency.

First, we reduce to the case where q is non-degenerate. Recall that V decomposes as $V = \mathrm{rad}(V) \mathbin{\widehat{\oplus}} W$ and that q restricted to W is non-degenerate. The theorem holds for V if and only if it holds for $(W, q|_W)$. Therefore, we may assume (V, q) to be a regular space.

Consider the vector space $V \oplus K$ and define on it the quadratic map $q'(v, c) = q(v) - ac^2$. To say that a completion V_w represents a means that q satisfies $q(v) = a$, for some non-zero vector $v \in V_w$, which, in turn, means that the vector $(v, 1) \in V_w \oplus K_w$ is isotropic for q'. If this is the case for every place w of K, then, by the Haße-Minkowski Theorem, there exists a non-zero vector $(u, c) \in V \oplus K$ that is isotropic for the map q'. If $c \neq 0$, then $q(u/c) = a$ and we are done. If $c = 0$, this means u is an isotropic vector in V and we have to tweak it a little.

Since (V, q) is assumed to be non-degenerate, there exists $z \in V$ such that $B(u, z) = 1$, where B represents the symmetric bilinear map associated to q. Set $\tilde{z} = z - \frac{q(z)}{2}u$ and note that $\tilde{z}$ is a non-zero isotropic vector for which $B(u, \tilde{z}) = 1$. Now $q\left(u + \frac{a}{2}\tilde{z}\right) = a$. □

In the last part of the argument given above we could have represented any scalar. We point this out in the next corollary.

Corollary 3.2.5. *Let (V, q) be a regular quadratic space. If V contains an isotropic vector then $q(V) = K$.*

In the spirit of the local-global principle, one can use the Haße-Minkowski Theorem (or, in fact, Corollary 3.2.4) to prove that two quadratic spaces over a number field K are isometric if and only they are locally isometric for every place w of K. More precisely:

Theorem 3.2.6 (Haße-Minkowski for quadratic spaces). *Let U and V be two quadratic spaces over the number field K. Then U is isometric to V if and only if U_w is isometric to V_w for every place w of K.*

Proof. Necessity is clear. We prove sufficiency by induction on the dimension of U. If U is one-dimensional, let $u \in U$ be such that $q(u) = a \neq 0$. Every U_w then represents a and, since each U_w is isometric to V_w, every V_w also represents a. By Corollary 3.2.4, there exists some $v \in V$ that represents a, so an isometry between U and V can be defined simply by mapping $u \mapsto v$.

For the general case, let q denote the quadratic map of U and Q the quadratic map of V. By the same construction as before there are vectors u, v such that $q(u) = Q(v) = a \neq 0$. For each place w of K, there is an isometry $f : U_w \to V_w$ and, in particular, $Q(f(u)) = q(u) = Q(v) = a$.

Claim: There exists an isometry $\rho \in O(V_w)$ mapping $f(u)$ to v.

Proof of claim: writing the polarisation identity for the pairs $(f(u), v)$ and $(f(u), -v)$, and adding them up leads to $Q(f(u) + v) + Q(f(u) - v) = 4a \neq 0$ (here we use the assumption that characteristic of $K \neq 2$). This means that at least one of the vectors $f(u) \pm v$ is anisotropic. Suppose $f(u) - v$ is anisotropic. Then let ρ be the reflection across the plane perpendicular to $f(u) - v$, given by formula (3.2.2), and note that $\rho(f(u)) = v$. This proves the claim.

Now, the isometry $\rho \circ f : U_w \to V_w$ maps u to v and thus it restricts to an isometry between the orthogonal complement of $\langle u \rangle$ in U_w and the orthogonal complement of $\langle v \rangle$ in V_w. This is clearly equivalent to $(\langle u \rangle^\perp)_w$ being isometric to $(\langle v \rangle^\perp)_w$. Since this is true for every place w of K, the induction hypothesis implies that $\langle u \rangle^\perp$ must be isometric to $\langle v \rangle^\perp$ and this isometry can then be extended to an isometry between U and V in the obvious way. □

3.3 Quaternion algebras

We continue to assume that the characteristic of K is $\neq 2$.

3.3.1 Definition

Definition 3.3.1. A *quaternion algebra A* over K is a 4-dimensional K-vector space with basis $\{1, i, j, k\}$ and a multiplication operation defined by:

$$i^2 = a, \quad j^2 = b, \quad ij = -ji = k,$$

for some $a, b \in K^*$, and linearly extended to the whole space.

The set $\{1, i, j, k\}$ is called a *standard basis*. This quaternion algebra is denoted by the *Hilbert symbol* $\left(\frac{a,b}{K}\right)$.

Note that a quaternion algebra A does not have a unique standard basis and, for that reason, the Hilbert symbol is not uniquely determined.

The first examples of quaternion algebras are:

Example 3.3.2. (i) The Hamilton's quaternions: $\mathscr{H} = \left(\frac{-1,-1}{\mathbb{R}}\right)$;

(ii) $M_2(K) \cong \left(\frac{1,1}{K}\right)$ with generators $i = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right), \quad j = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$;

(iii) More generally, $M_2(K) \cong \left(\frac{1,t}{K}\right)$ for any non-zero $t \in K$ and the isomorphism is given by the map

$$\psi(x + yi + zj + wij) = \begin{pmatrix} x+y & z+w \\ t(z-w) & x-y \end{pmatrix},$$

with inverse

$$\psi^{-1}\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = \frac{1}{2}[a + d + (a-d)i + (b + t^{-1}c)j + (b - t^{-1}c)ij];$$

(iv) It is easy to see that $\left(\frac{a,b}{K}\right) \cong \left(\frac{b,a}{K}\right)$.

If $L \mid K$ is a field extension, then we may extend the scalars of the K-vector space $\left(\frac{a,b}{K}\right)$ to L. Multiplication of elements in this new L-vector space is defined in the natural way, turning it into a quaternion algebra over L. One clearly has that:

$$\left(\frac{a,b}{K}\right) \otimes_K L \cong \left(\frac{a,b}{L}\right).$$

More generally, if $\sigma : K \to L$ is a field embedding, we may regard L with the structure of a K-algebra induced by σ (see Definition 3.1.1) and then consider the tensor product of the K-algebras $\left(\frac{a,b}{K}\right)$ and L. We denote this product by $\left(\frac{a,b}{K}\right) \otimes_\sigma L$, in order to indicate the dependence of this construction on the embedding σ. Note that one has the following isomorphism

$$\left(\frac{a,b}{K}\right) \otimes_\sigma L \cong \left(\frac{\sigma(a), \sigma(b)}{L}\right), \tag{3.3.1}$$

given by

$$(x + yi + zj + wij) \otimes_\sigma \alpha \mapsto \alpha(\sigma(x) + \sigma(y)i' + \sigma(z)j' + \sigma(w)i'j'),$$

where $\{1, i', j', i'j'\}$ is a standard basis of $\left(\frac{\sigma(a),\sigma(b)}{L}\right)$.

Proposition 3.3.3 (Basic Properties)**.**

1. $\left(\frac{a,b}{K}\right) \cong \left(\frac{ax^2,by^2}{K}\right)$, *for any* $a, b, x, y \in K^*$;

2. *Every quaternion algebra is a central simple algebra.*

Proof. 1. Let $\{1, i, j, ij\}$ and $\{1', i', j', i'j'\}$ be standard basis for $\left(\frac{a,b}{K}\right)$ and $\left(\frac{ax^2,by^2}{K}\right)$ respectively. The linear map taking $1 \mapsto 1$, $i' \mapsto xi$, $j' \mapsto yj$ and $i'j' \mapsto xyij$, is a vector space isomorphism that respects multiplication and, therefore, an algebra homomorphism.

2. Consider the quaternion algebra $A = \left(\frac{a,b}{K}\right)$ and let $\overline{K}$ be the algebraic closure of K. Choose x and y in $\overline{K}$ such that $x^2 = a^{-1}$ and $y^2 = b^{-1}$, then it follows from part 1 that $\left(\frac{a,b}{K}\right) \cong \left(\frac{a,b}{\overline{K}}\right) \cong \left(\frac{1,1}{\overline{K}}\right) \cong M_2(\overline{K})$. The centre of $\left(\frac{a,b}{\overline{K}}\right)$ is, on the one hand, the centre of $M_2(\overline{K})$, which is $\overline{K}$, and, on the other hand, it is $Z(A) \otimes \overline{K}$. Counting dimensions over $\overline{K}$ shows that $Z(A) = K$.

We prove that A is simple in a similar way. Let I be a nontrivial two-sided ideal of A. Then $I \otimes \overline{K}$ is a non-trivial two-sided ideal of the simple algebra $M_2(\overline{K})$ and, as such, it must be all of $M_2(\overline{K})$. This can only be the case when $I = A$. □

The classical example of an algebra which is *not* a division algebra is the algebra of $n \times n$ matrices over a field. As a consequence of the Wedderburn Structure Theorem, it turns out that this is the only way in which a 4-dimensional simple algebra can fail to be a division algebra:

Proposition 3.3.4. *If A is a 4-dimensional simple algebra over K, then either A is a division algebra or A is isomorphic to $M_2(K)$. In particular, quaternion algebras are either a division algebra or they are isomorphic to $M_2(K)$.*

Proof. According to Wedderburn's Structure Theorem 3.1.7, A must be isomorphic to a matrix algebra $M_n(D)$, where D is a division algebra. The K-dimension of $M_n(D)$ is then $n^2 \dim_K D$. Since A is 4-dimensional, there are only two possibilities: $\dim_K D = 4,\ n = 1$ or $\dim_K D = 1,\ n = 2$. Thus A is isomorphic either to D or to $M_2(K)$. □

It is a remarkable fact that the properties of being central and simple characterise quaternion algebras among all 4-dimensional algebras.

Theorem 3.3.5. *Let A be a 4-dimensional central simple algebra over the field K (of characteristic $\neq 2$). Then A is a quaternion algebra.*

Proof. According to Proposition 3.3.4 A is either a division algebra or it is isomorphic to $M_2(K)$. We already know that $M_2(K)$ is a quaternion algebra, so assume A is a division algebra.

Let $x \in A \setminus K$ and consider the the subalgebra $L = K(x)$. Note that L is commutative and, since it is finite-dimensional, over K, x has a multiplicative inverse in L (consider the linear transformation given by multiplication by x), so L is actually a field extension of K. Moreover, we shall prove next that $L \mid K$ is a quadratic field extension. First, L, as an algebra over K, is not central ($Z(L) = L \neq K$) so the fact that A is assumed to be central implies that $A \neq L$. Let $y \in A \setminus L$ and note that $\{1, x, y, xy\}$ must be linearly independent. Indeed, $\{1, x, y\}$ is linearly independent by construction. If xy were a linear combination of the other three, than solving for y would imply that $y \in L$, a contradiction. It follows that $\{1, x, y, xy\}$ is a basis for A. In particular, there are $c_0, c_1, c_2, c_3 \in K$ such that

$$x^2 = c_0 + c_1 x + c_2 y + c_4 xy.$$

As before, if $c_2 + c_4 x \neq 0$, then $y \in L$, contrary to our assumptions. Therefore we actually have $x^2 = c_0 + c_1 x$, which proves that L is a quadratic extension of K. Let $i \in L$ be such that $L = K(i)$ and $i^2 = a \in K$ (choose, for instance, $i = x - c_1/2$ which is possible since the characteristic of K is $\neq 2$).

The construction so far (as well as the notation chosen) strongly suggests that 1 and i should be part of a standard basis. Now we want to find $j \in A \setminus K$ such that $j^2 \in K$ and $ij = -ji$. This can be achieved by means of the Skolem-Noether Theorem 3.1.9. Let σ denote the nontrivial field automorphism of $L \mid K$ mapping $i \mapsto -i$. By composing σ with the inclusion of L in A we obtain a K-algebra homomorphism $\sigma : L \to A$ (which we are also denoting by σ). The Skolem-Noether Theorem then gives an invertible element $j \in A$ such that $\sigma(z) = j^{-1}zj$ for every $z \in L$. In particular, $j^{-1}ij = \sigma(i) = -i$, which means $ij = -ji$. Note that $j \notin L$ since it does not commute with i. Also if ij were a linear combination of $\{1, i, j\}$ then, as we have seen before, this would imply that $j \in L$, contrary to what we have just established. We conclude that $\{1, i, j, ij\}$ is a basis for A. Finally, we show that $j^2 = b \in K$. Since A is central, it suffices to prove that j^2 commutes with i, for this will show that j^2 is in the centre of A. The relation $ij = -ji$ immediately implies that $ij^2 = -jij = j^2 i$. In other words, we have proved that $A \cong \left(\frac{a,b}{K}\right)$. □

Embedded in the proof of Theorem 3.3.5, there is a construction of standard basis for quaternion algebras which we single out in the following corollary.

Corollary 3.3.6. *Let A be a quaternion division algebra over K. For any $i \in A \setminus K$, $K(i) \mid K$ is a quadratic extension. Furthermore, if $i^2 \in K$, then there exists $j \in A$ such that $\{1, i, j, ij\}$ is a standard basis for A.*

Corollary 3.3.7. *Let A be a quaternion algebra over K and let $L \mid K$ be a quadratic extension that embeds into A. Then $A \otimes_K L \cong M_2(L)$.*

Proof. We may assume without loss of generality that $L \subset A$, so that $L = K(x)$ for some $x \in A \setminus K$. Pick $i = x + t$ for an appropriate $t \in K$ that makes $i^2 = a \in K$. It follows from Corollary 3.3.6 that there exists a standard basis $\{1, i, j, ij\}$ for A. With this basis in hand, we can find zero divisors in the algebra $A \otimes_K L$: take, for instance, $x = i \otimes (ia^{-1})$ and note that $x^2 = 1$, where the 1 here really denotes the unity in the algebra $A \otimes_K L$. Since x is clearly different from ± 1, we find that $x + 1$ and $x - 1$ are zero divisors. It thus follows from Proposition 3.3.4 that $A \otimes_K L$ must be isomorphic to $M_2(L)$. □

3.3.2 Trace and norm

Let A be a quaternion algebra over some field K, with standard basis $\{1, i, j, ij\}$. For an element $x = x_0 + x_1 i + x_2 j + x_3 ij$ in A, we make the following definitions:

Definition 3.3.8.

(i) The *conjugate* $\overline{x}$ of x is given by

$$\overline{x} = x_0 - x_1 i - x_2 j - x_3 ij.$$

Conjugation defines an *anti-involution* of the algebra since $\overline{(x+y)} = \overline{x} + \overline{y}$, $\overline{xy} = \overline{y}\,\overline{x}$, $\overline{\overline{x}} = x$ and $\overline{rx} = r\overline{x}$ for $r \in K$.

(ii) The *(reduced) trace* and *(reduced) norm* of an element $x \in A$ are defined, respectively, by $\operatorname{tr}(x) = x + \overline{x}$ and $\operatorname{n}(x) = x\overline{x}$. Explicitly,

$$\begin{aligned} \operatorname{tr}(x) &= 2x_0; \\ \operatorname{n}(x) &= x_0^2 - ax_1^2 - bx_2^2 + abx_3^2. \end{aligned} \tag{3.3.2}$$

Note that they both lie in K. Furthermore, the invertible elements of A are precisely those with non-zero norm. In this case, the inverse of x is $\overline{x}/\operatorname{n}(x)$.

(iii) The elements with trace 0 are called the *pure quaternions* of A. They form the subspace $A_0 = \operatorname{span}\{i, j, ij\}$ of A.

It is important to note that none of the definitions above depend on the chosen standard basis. Indeed, conjugation can be characterised in a purely algebraic manner, as the next proposition shows.

Proposition 3.3.9. *A non-zero element $x \in A$ is a pure quaternion if and only if $x \notin Z(A)$ and $x^2 \in Z(A)$.*

Proof. For $x = x_0 + x_1 i + x_2 j + x_3 ij$, we calculate that

$$x^2 = (x_0^2 + ax_1^2 + bx_2^2 - abx_3^2) + 2x_0(x_1 i + x_2 j + x_3 ij) \tag{3.3.3}$$

Since $Z(A) = K$, the result follows. □

So any $x \in A$ can be uniquely written in the form $x = a + \alpha$ where $a \in K$ and α is a pure quaternion.

From (3.3.3), we can also deduce that any element $x \in A$ satisfies a quadratic equation, namely:

$$x^2 - \mathrm{tr}(x)x + \mathrm{n}(x) = 0. \tag{3.3.4}$$

In particular, when A is a division algebra, it follows that, for any $x \in A \setminus K$, $K(x) \mid K$ is a quadratic extension, confirming what we established in Corollary 3.3.6.

Note that, in a 2×2 matrix algebra, the norm and trace are, respectively, the determinant and trace of a matrix. Indeed, considering the standard basis $\{1, i = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right), j = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right), ij\}$ given in Example 3.3.2 (ii) any matrix $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ in $M_2(K)$ can be written as

$$M = \frac{a+d}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{a-d}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + \frac{b+c}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \frac{b-c}{2}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Then the norm and trace of M, according to the equations in (3.3.2), are:

$$\mathrm{n}(M) = \left(\frac{a+d}{2}\right)^2 - \left(\frac{a-d}{2}\right)^2 - \left(\frac{b+c}{2}\right)^2 + \left(\frac{b-c}{2}\right)^2 = ad - bc = \det(M),$$

and

$$\mathrm{tr}(M) = 2\left(\frac{a+d}{2}\right) = a + d = \mathrm{tr}\,(M).$$

It is also worth noting the following addendum to Corollary 3.3.6:

Proposition 3.3.10. *Let A be a quaternion division algebra over K. For any $w \in A \setminus K$, $K(w) \mid K$ is a quadratic extension, and the reduced norm of A, when restricted to $K(w)$, coincides with the extension norm $\mathrm{N}_{K(w)|K}$.*

Proof. The part concerning the norms is the only one requiring explanation. Let $x \in K(w)$ be written as $x = x_0 + x_1 w$. Note that $x_0 \in K$ while $x_1 w$ is a pure quaternion, so the observation following Proposition 3.3.9 shows that the conjugate of x in the quaternion algebra is $\overline{x} = x_0 - x_1 w$, which coincides with the conjugate of x in the field extension $K(w) \mid K$. The result follows. □

The norm map $\mathrm{n} : A \to K$ is multiplicative: $\mathrm{n}(xy) = (xy)(\overline{xy}) = xy\overline{y}\,\overline{x} = x\mathrm{n}(y)\overline{x} = x\overline{x}\mathrm{n}(y) = \mathrm{n}(x)\mathrm{n}(y)$. Moreover, the following expression defines on A a symmetric bilinear form B whose associated quadratic map is n:

$$B(x,y) := \frac{1}{2}[n(x+y) - n(x) - n(y)] = \frac{x\overline{y} + y\overline{x}}{2}. \tag{3.3.5}$$

Then (A, n) becomes a quadratic space. The norm function n restricted to the subspace A_0 is non-degenerate. Indeed, if $x = x_1 i + x_2 j + x_3 ij \in \mathrm{rad}(A_0)$ one has, in particular, that $0 = B(x,i) = -ax_1$, $0 = B(x,j) = -bx_2$ and $0 = B(x,ij) = abx_3$, so $x = 0$.

As we shall see later, it is often important to be able to determine whether a quaternion algebra is isomorphic to a matrix algebra. We present a few criteria in the language of quadratic spaces.

Theorem 3.3.11. *Let $A = \left(\frac{a,b}{K}\right)$. The following are equivalent:*

1. *$A \cong M_2(K)$;*
2. *A is not a division algebra;*
3. *The quadratic space (A, n) is isotropic;*
4. *The quadratic space (A_0, n) is isotropic;*
5. *The quadratic form $ax^2 + by^2 = 1$ has a solution in K;*
6. *The element a is a norm from the field extension $K(\sqrt{b}) \mid K$.*

Proof. (1) $\iff$ (2) is just Proposition 3.3.4.

(2) $\implies$ (3): If A contains a non-zero element $x \in A$ that is a zero divisor, x cannot be invertible so then $\mathrm{n}(x) = 0$, which mean x is an isotropic vector.

(3) $\implies$ (4): Let $x = x_0 + x_1 i + x_2 j + x_3 ij$ be an isotropic vector in (A, n). If $x_0 = 0$, there is nothing to prove, so assume $x_0 \neq 0$. Then at least one of the x_1, x_2, x_3 is non-zero, which we will assume, without loss in generality, to be x_1. Suppose A_0 is anisotropic and we will arrive at a contradiction. Let $y = y_0 + y_1 i + y_2 j + y_3 ij$ be an element in A whose coordinates are yet to be determined. If we find y_0, y_1, y_2, y_3 such that xy has vanishing first coordinate, then $xy = 0$. Indeed, a vanishing first coordinate means $xy \in A_0$, but $\mathrm{n}(xy) = \mathrm{n}(x)\mathrm{n}(y) = 0$ and since we are assuming A_0 to be anisotropic, it must be the case that $xy = 0$. Of course, if we choose $y = \pm\overline{x}$, the equation $xy = 0$ will not tell us anything that we do not already know, so we try to stir away from this solution. Equating the expression for the first coordinate of xy to zero gives $x_0y_0 + ax_1y_1 + bx_2y_2 - abx_3y_3 = 0$, and one can guess a few solutions

by inspection. For instance, taking $y_0 = -bx_2/2$, $y_1 = bx_3/2$, $y_2 = x_0/2$, $y_3 = x_1/2$ leads to the pure quaternion

$$xy = b(x_0x_3 - x_1x_2)i + a(x_1^2 - bx_3^2)j + (x_0x_1 - bx_2x_3)ij.$$

Now, $xy = 0$ so that, in particular, $a(x_1^2 - bx_3^2) = 0$. But notice that $a(x_1^2 - bx_3^2) = -\mathrm{n}(x_1i + x_3ij)$ whence, using once again the assumption that A_0 is anisotropic, it follows that $x_1i + x_3ij = 0$ and, consequently, that $x_1 = 0$, a contradiction.

(4) $\Longrightarrow$ (5): Let $x_1i + x_2j + x_3ij$ be an isotropic vector. If $x_3 \neq 0$ then one finds the solution $x = x_2/ax_3$ and $y = x_1/bx_3$. If, on the other hand, $x_3 = 0$, then $ax_1^2 + bx_3^2 = 0$ and $x_1, x_2 \neq 0$. Set $f(r, q, r', q') = a(rx_1 + q)^2 + b(r'x_2 + q')^2$. Using that $f(0, 0, 0, 0) = 0$, it is (tedious but) not hard to vary the parameters r, q, r', q' in K until we eventually find $f(r, q, r', q') = 1$. For example, one can take $q = 1/2a$, $q' = -x_2/2ax_1$, $r = r' = 1/2x_1$, leading to the solution $x = (a + 1)/2a$ and $y = x_2(a - 1)/2ax_1$.

(5) $\Longrightarrow$ (6): Let $x, y \in K$ be such that $ax^2 + by^2 = 1$. If $x = 0$ then $\sqrt{b} \in K$ and $K(\sqrt{b}) = K$ so the result is trivial. Now, if $x \neq 0$, dividing both sides by x^2 gives that

$$a = \frac{1}{x^2} - b\frac{y^2}{x^2} = \left(\frac{1}{x} + \frac{y}{x}\sqrt{b}\right)\left(\frac{1}{x} - \frac{y}{x}\sqrt{b}\right) = N_{K(\sqrt{b})|K}\left(\frac{1}{x} + \frac{y}{x}\sqrt{b}\right).$$

(6) $\Longrightarrow$ (2): Again, we have to consider two cases: the extension $K(\sqrt{b}) \mid K$ has either degree 1 or degree 2. In the former case, it means that $\sqrt{b} \in K$ so, in particular, $\sqrt{b} \pm j \neq 0$. But note that $(\sqrt{b} + j)(\sqrt{b} - j) = b - j^2 = 0$, so A is not a division algebra. Now, suppose a is a norm of the quadratic extension $K(\sqrt{b}) \mid K$, so there are $x, y \in K$ for which $a = x^2 - by^2$. It follows that $\mathrm{n}(x + i + yj) = x^2 - a - by^2 = 0$. □

Corollary 3.3.12. *The quaternion algebras* $\left(\frac{1,a}{K}\right)$, $\left(\frac{a,-a}{K}\right)$ *and* $\left(\frac{a,1-a}{K}\right)$ *are isomorphic to* $M_2(K)$.

Proof. The cases $\left(\frac{1,a}{K}\right)$ and $\left(\frac{a,1-a}{K}\right)$ follow from criterion (5) in the above theorem. In $\left(\frac{a,-a}{K}\right)$, we observe that the vector $i + j$ is isotropic.

The fact that $\left(\frac{1,a}{K}\right) \cong M_2(K)$ had already been proved in Example 3.3.2 (iii), where an explicit isomorphism was given. □

The next theorem shows that the regular quadratic space (A_0, n) determines the isomorphism class of the quaternion algebra A.

Theorem 3.3.13. *Let A and A' be two quaternion algebras with norm, respectively, n and n'. Then $A \cong A'$ if and only if the quadratic spaces (A_0, n) and (A_0', n') are isometric.*

Proof. Let $f : A \to A'$ be an isomorphism. Since, by Proposition 3.3.9, pure quaternions are algebraically characterised, the isomorphism maps A_0 onto A'_0. Equation (3.3.4) shows that the norm of a pure quaternion x is given by $-x^2$. Therefore $n'(f(x)) = -f(x)^2 = f(n(x)) = n(x)$, which shows that $f : A_0 \to A'_0$ is an isometry.

Conversely, if $\phi : A_0 \to A'_0$ is an isometry, let $A = \left(\frac{a,b}{K}\right)$ with standard basis $\{1, i, j, ij\}$. We aim to show that $\{1, \phi(i), \phi(j), \phi(i)\phi(j)\}$ is a standard basis for A'. First note that $\phi(i)^2 = -n'(\phi(i)) = -n(i) = a$ and, similarly, that $\phi(j)^2 = b$. Then we need to check that $\phi(i)\phi(j)$ is actually a pure quaternion of A_0. Note that the relation $ij = -ji$ implies that i and j are orthogonal, which means that $\phi(i)$ and $\phi(j)$ are orthogonal. It is easy to see that, for any $x \in A_0$, $\phi(\overline{x}) = \overline{\phi(x)}$, and so (3.3.5) gives the relation $0 = \phi(i)\phi(j) + \phi(j)\phi(i)$. In particular, $\phi(i)[\phi(i)\phi(j)] = -\phi(i)[\phi(j)\phi(i)] = -[\phi(i)\phi(j)]\phi(i)]$, which shows that $\phi(i)\phi(j)$ is not in the centre of A'. On the other hand, $[\phi(i)\phi(j)]^2 = -ab \in Z(A')$ and so, by Proposition 3.3.9, $\phi(i)\phi(j) \in A'_0$. Finally, $\{\phi(i), \phi(j), \phi(i)\phi(j)\}$ is linearly independent since, if $c_1, c_2, c_3 \in K$ are such that $c_1\phi(i) + c_2\phi(j) + c_3\phi(i)\phi(j) = 0$, multiplying both sides of the equation on the left by $\phi(i)$ gives that $ac_1 = -c_2\phi(i)\phi(j) - ac_3\phi(j) \in A'_0$. But this can ony be the case if $c_1 = 0$. Similarly, we see that $c_2 = c_3 = 0$. Therefore $\{1, \phi(i), \phi(j), \phi(i)\phi(j)\}$ is a standard basis for A' which shows that $A' \cong \left(\frac{a,b}{K}\right) = A$. □

Theorem 3.3.13 allows us to employ tools we have developed for quadratic spaces over number fields in order to study the isomorphism class of a quaternion algebra defined over a number field, as the following corollary exemplifies. Theorem 3.4.10 is another example of of this interplay.

Corollary 3.3.14. *Two quaternion algebras $A = \left(\frac{a,b}{K}\right)$ and $A' = \left(\frac{a',b'}{K}\right)$, defined over the number field K, are isomorphic if and only if the quadratic forms $ax^2 + by^2 - abz^2$ and $a'x^2 + b'y^2 + a'b'z^2$ are equivalent over K.*

Proof. These two quadratic forms being equivalent means that the quadratic spaces A_0 and A'_0 are isometric. □

3.4 Classification of quaternion algebras

3.4.1 Quaternion algebras over local fields

In this subsection we give a complete description of quaternion algebras over local fields. This includes $\mathbb{C}$, $\mathbb{R}$ and $\mathfrak{p}$-adic fields.

Proposition 3.4.1 (Quaternion algebras over $\mathbb{C}$)**.** *The only quaternion algebra over $\mathbb{C}$ is the algebra $M_2(\mathbb{C})$.*

Proof. This follows straight from Proposition 3.3.3 (1), or from the more general Corollary 3.1.8, together with the fact that $\left(\frac{1,1}{K}\right) \cong M_2(K)$ (Example 3.3.2 (ii)). $\square$

Proposition 3.4.2 (Quaternion algebras over $\mathbb{R}$)**.** *Any quaternion algebra over $\mathbb{R}$ is isomorphic either to $M_2(\mathbb{R})$ or, if it is a division algebra, to the Hamilton's quaternions $\mathscr{H}$.*

Proof. Proposition 3.3.3 (1) shows that any quaternion algebra $\left(\frac{a,b}{\mathbb{R}}\right)$ is isomorphic to $\left(\frac{\pm 1,\pm 1}{\mathbb{R}}\right)$, depending on the signs of a and b. It follows from Example 3.3.2 (iii) and (iv), that $\left(\frac{1,\pm 1}{\mathbb{R}}\right) \cong M_2(\mathbb{R}) \cong \left(\frac{-1,1}{\mathbb{R}}\right)$. By definition, $\left(\frac{-1,-1}{\mathbb{R}}\right) \cong \mathscr{H}$, which is a division algebra. $\square$

Similarly, over a non-Archimedean local field, a quaternion division algebra is uniquely determined. The proof, however, is slightly more delicate and depends on a number-theoretic machinery that has not been presented here. We will sketch an argument following the proof in [32, Chapter VI. Proposition 2.10]. First, we introduce some new objects.

Let (K, v) be a complete non-Archimedean valued field of characteristic $\neq 2$, where $v : K \to \mathbb{Z} \cup \{\infty\}$ is an additive (non-Archimedean) valuation (see Definition 2.3.9) with valuation ring $\mathfrak{o}$. Denote the unique maximal ideal of $\mathfrak{o}$ by $\mathfrak{p}$, with uniformiser π, so that the residue field is given by $k = \mathfrak{o}/\mathfrak{p}$. For a quaternion *division* algebra A we define the analogue of a discrete additive valuation on A: let $w : A \to \mathbb{Z} \cup \{\infty\}$ be defined as $w(x) = v(\mathrm{n}(x))$, for $x \in A$. Recall that $v(0) = \infty$. In this way, w satisfies the following conditions, as an additive valuation should.

Lemma 3.4.3. *For any $x, y \in A$, the following is true:*

1. $w(xy) = w(x) + w(y)$*;*

2. $w(x + y) \geq \min\{w(x), w(y)\}$*, with equality if and only if* $w(x) = w(y)$*.*

Proof. Both statements are trivial if either x or y is 0, so let us assume otherwise. Property (1) follows from the multiplicativity of the norm together with the analogous property for v. Now, take any $x \in A \setminus K$. By Proposition 3.3.10, $K(x) \mid K$ is a quadratic extension and n restricts to the extension norm on $K(x)$. It follows from Theorem 2.3.43 that $v \circ \mathrm{N}_{K(x)|K}$ is a valuation on $K(x)$ and so property (2) is satisfied on $K(x)$. For $x, y \in A^*$, it follows from (1) that $w(x + y) - w(y) = w(xy^{-1} + 1)$. So, using property (1) for $K(xy^{-1})$, we get

$$w(x + y) \geq w(y) + \min\{w(xy^{-1}), 0\},$$

with equality if and only if $w(xy^{-1}) = w(1)$. The result now follows from (1). $\square$

Note that $w(A^*)$ is a subgroup of $\mathbb{Z}$ and, as such, it must be of the form $d\mathbb{Z}$ for some uniquely determined positive integer d. Since $w(\pi) = v(\pi^2) = 2$, it follows that $d = 1$ or 2. We normalise w by taking w/d instead. This does not alter anything that was done for w so far, so we might as well relabel w/d as w. Remember that now w is surjective and $w(x) = v(\mathrm{n}(x))/d$ where $d = 1$ or 2.

From the properties derived in Lemma 3.4.3, we can easily see, just as in the case of valued fields, that $\mathcal{O} = \{x \in A \mid w(x) \geq 0\}$ is a subring of A with the (unique) maximal two-sided ideal $\mathfrak{P} = \{x \in A \mid w(x) > 0\}$.

Let $L \mid K$ be a Galois extension of the non-Archimedean local fields K, with residue fields respectively l and k. There is a natural embedding of k into l and $[l : k]$ is called the *residual degree*. Note that, in this context, the residual degree is finite (both residue fields are finite since L and K are locally compact). In fact we have the inequality $[l : k] \leq [L : K]$ (see, for example, [35, Chapter 24. F1]). The extension $L \mid K$ is said to be *unramified* when $[l : k] = [L : K]$.

It is known in Algebraic Number Theory that unramified extensions of $\mathfrak{p}$-adic fields of a specified degree exist and are uniquely determined ([10, Chapter 8. §2]). In particular, there exists a unique quadratic extension of a $\mathfrak{p}$-adic field. This is actually true for any local field, not necessarily $\mathfrak{p}$-adic ([32, Chapter VI. Proposition 2.8]). Furthermore, we will also need the following result, which we state here for future reference:

Theorem 3.4.4 ([32, Chapter VI. Proposition 2.9])**.** *Let L be the unique unramified quadratic extension of the local field K. Then $L = K(\sqrt{u})$ for some unit $u \in \mathfrak{o}^\times$ whose square class in $K^*/(K^*)^2$ is uniquely determined. Moreover, every unit $u' \in \mathfrak{o}^\times$ is a norm of $L \mid K$.*

We are now ready to prove the following theorem:

Theorem 3.4.5 (Quaternion division algebras over non-Archimedean local fields)**.** *Over the non-Archimedean local field K there is a unique (up to isomorphism) quaternion division algebra. Moreover, it is isomorphic to $\left(\frac{u,\pi}{K}\right)$ where π is a uniformiser of K and u is such that $K(\sqrt{u})$ is the unique unramified quadratic extension of K.*

Proof. We first observe that $\left(\frac{u,\pi}{K}\right)$ is indeed a division algebra. There is a unique (up to equivalence) valuation on $K(\sqrt{u})$ extending v, let us call it v'. If σ denotes the nontrivial automorphism of $K(\sqrt{u}) \mid K$, then $v' \circ \sigma$ is another such valuation and must differ from v' by a scalar multiple. Since they coincide on K we conclude that they are actually equal. Given any $x \in K(\sqrt{u})$, one has that $v(\mathrm{N}_{K(\sqrt{u})|K}(x)) = v'(x) + v'(\sigma x) = 2v'(x)$. On the other hand, $v(\pi) = 1$, which proves that $\pi \notin \mathrm{N}_{K(\sqrt{u})|K}(K)$ and so, according to Theorem 3.3.11 (6), $\left(\frac{u,\pi}{K}\right)$ is a division algebra.

Let A be any division quaternion algebra over K. We must prove that it is isomorphic to $\left(\frac{u,\pi}{K}\right)$ and we will do so in several steps.

Step 1: $\mathcal{O}$ is a free $\mathfrak{o}$-module of rank 4.

Let $\{x_1, x_2, x_3, x_4\}$ be a basis of A. For m sufficiently large, each $\pi^m x_i \in \mathcal{O}$. Indeed, $\mathrm{n}(\pi^m x_i) = \pi^{2m}\mathrm{n}(x_i)$ and so $w(\pi^m x_i) \geq 0$ if m is large enough. We may therefore assume, without loss of generality, that each $x_i \in \mathcal{O}$. Whence

$$\mathfrak{o}[x_1, x_2, x_3, x_4] \subset \mathcal{O}.$$

On the other hand, let B be (twice) the symmetric bilinear map associated to the quadratic map n, i.e., $B(x, y) = \mathrm{n}(x + y) - \mathrm{n}(x) - \mathrm{n}(y)$ for $x, y \in A$. Then B is non-degenerate (since A is a division algebra), which means that the correspondence $v \mapsto B(\,\cdot\,, v)$ is an isomorphism. We may select $y_1, y_2, y_3, y_4 \in A$ such that $B(x_i, y_j) = 1$ if $i = j$, or 0 if $i \neq j$. Let x be any element in $\mathcal{O}$ and write it as $x = a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4$. Note that $\mathrm{n}(\mathcal{O}) \subset \mathfrak{o}$ and thus B maps $\mathcal{O} \times \mathcal{O}$ into $\mathfrak{o}$. In particular, each $a_i = B(x, y_i)$ is in $\mathfrak{o}$. This proves that

$$\mathcal{O} \subset \mathfrak{o}[x_1, x_2, x_3, x_4].$$

Since $\mathfrak{o}$ is a principal ideal domain (Proposition 2.3.25 (3)), it follows from the fundamental theorem for free modules over principal ideal domains that $\mathcal{O}$ is a free $\mathcal{O}$-module of rank 4 (cf. proof of Proposition 2.2.21).

Step 2: $\dim_k(\mathcal{O}/\mathfrak{P}) > 1$.

The $\mathfrak{o}$-module structure of $\mathcal{O}$ naturally induces an $\mathfrak{o}$-module structure on the quotient $\mathcal{O}/\mathfrak{p}\mathcal{O}$. Since the ideal $\mathfrak{p}$ of $\mathfrak{o}$ *annihilates* $\mathcal{O}/\mathfrak{p}\mathcal{O}$ (the terminology is self-explanatory), it descends to an $\mathfrak{o}/\mathfrak{p}$-module structure on $\mathcal{O}/\mathfrak{p}\mathcal{O}$. In other words, $\mathcal{O}/\mathfrak{p}\mathcal{O}$ is a k-vector space. It is easy to see that its dimension over k is at most 4. On the other hand, $\mathfrak{o}$ is a local ring with (unique) maximal ideal $\mathfrak{p}$ and so, from [1, Proposition 2.8], if $\{x_1 + \mathfrak{p}\mathcal{O}, \ldots, x_n + \mathfrak{p}\mathcal{O}\}$ is a basis for $\mathcal{O}/\mathfrak{p}\mathcal{O}$ over k, then $\{x_1, \ldots, x_n\}$ generates $\mathcal{O}$, which means n is at least 4. We conclude that $\dim_k(\mathcal{O}/\mathfrak{p}\mathcal{O}) = 4$.

In the same way as before, the $\mathfrak{o}$-module structure on $\mathcal{O}/\mathfrak{P}$ descends to a k-vector space structure on $\mathcal{O}/\mathfrak{P}$. In order to compare the dimension of these two k-vector spaces, we first need to observe that $\mathfrak{p}\mathcal{O} = \pi\mathcal{O} = \mathfrak{P}^{2/d}$, where the second equality follows from the definitions. Indeed, note that $\pi\mathcal{O} = \{x \in A \mid w(x) \geq 2/d\}$. If $d = 2$, the right-hand side is precisely $\mathfrak{P}$. For $d = 1$, any $x \in \mathfrak{P}^2$ satisfies $w(x) \geq 2$ and so $\mathfrak{P}^2 \subset \pi\mathcal{O}$. Conversely, if $y \in \pi\mathcal{O}$ then $w(y\pi^{-1}) = w(y) - w(\pi) \geq 2 - 1 = 1$ and $y\pi^{-1} \in \mathfrak{P}$, which implies that $y = (y\pi^{-1})\pi \in \mathfrak{P}^2$. Moreover, in this case, $\dim_k(\mathcal{O}/\mathfrak{P}^2) = 2\dim_k(\mathcal{O}/\mathfrak{P})$, as we prove next. Since we are assuming w to be surjective, there exists some $s \in \mathfrak{P}$ for which $w(s) = 1$. Consider the k-linear transformation from $\mathscr{O}/\mathfrak{P}$ to $\mathfrak{P}/\mathfrak{P}^2$ defined by $a + \mathfrak{P} \mapsto as + \mathfrak{P}^2$ and notice that it is injective and surjective. Finally, apply the rank-nullity theorem to the projection $\mathscr{O}/\mathfrak{P}^2 \to \mathscr{O}/\mathfrak{P}$, whose kernel is $\mathfrak{P}/\mathfrak{P}^2$. Therefore, in either case we have:

$$4 = \dim_k(\mathscr{O}/\pi\mathscr{O}) = \dim_k(\mathscr{O}/\mathfrak{P}^{2/d}) = \frac{2}{d}\dim_k(\mathscr{O}/\mathfrak{P}),$$

In particular, $\dim_k(\mathscr{O}/\mathfrak{P}) > 1$.

Now, $\mathscr{O}/\mathfrak{P}$ is a finite dimensional vector space over the finite field k and thus it must be finite. Furthermore, $\mathscr{O}/\mathfrak{P}$ is a division ring: a non-zero element $x + \mathfrak{P}$ is represented by some $x \in \mathscr{O} \setminus \mathfrak{P}$, which means that $w(x) = 0$ and thus $w(x^{-1}) = 0$, so $x^{-1} + \mathfrak{P}$ is the inverse of $x + \mathfrak{P}$. It follows from the well known Wedderburn's Little Theorem that a finite division ring must be commutative and, therefore, a field. In particular, the extension, being an extension of finite fields, is simple.

Choose $\alpha \in \mathscr{O}$ such that $\alpha + \mathfrak{P}$ generates $\mathscr{O}/\mathfrak{P}$ over k and let $L = K(\alpha)$. We know that $L \mid K$ is quadratic, and that the minimal polynomial of α has coefficients (norm and trace of α) in $\mathfrak{o}$. Reducing modulo $\mathfrak{p}$, we find that $\alpha + \mathfrak{P}$ is the root of a polynomial of degree at most 2 and coefficients in k, which means that $\dim_k(\mathscr{O}/\mathfrak{P}) \leq 2$. Together with the information we had before, this implies that the dimension is precisely 2 (in particular, $d = 1$). The field L is a local field with valuation ring $L \cap \mathscr{O}$ and residue field $l = L \cap \mathscr{O}/(L \cap \mathfrak{P})$ which is a field extension of $\mathfrak{o}/\mathfrak{p}$. Since the former contains the residue class of α, which generates $\mathscr{O}/\mathfrak{P}$ over k, it follows that $2 \leq [l : k] \leq [L : K] = 2$ and thus $L \mid K$ is unramified. By uniqueness of a quadratic unramified extension over the $\mathfrak{p}$-adic field K, we have that $\alpha^2 = u \in \mathfrak{o}$ (Theorem 3.4.4).

By Corollary 3.3.6, there exists $\beta \in A$ such that $\{1, \alpha, \beta, \alpha\beta\}$ is a standard basis for A, and we may rescale it in such a way that $\beta^2 \in \mathfrak{o}$. We know that $\beta^2 = \pi^m u'$ where u' is a unit in $\mathfrak{o}$ (see the discussion preceding Proposition 2.3.11). Eliminating squares, we have that $A \cong \left(\frac{u,\pi^m u'}{K}\right) \cong \left(\frac{u,\pi^\epsilon u'}{K}\right)$ where $\epsilon = 0$ or 1.

According to Theorem 3.4.4, u' in a norm of $K(\sqrt{u})$, so it follows from Theorem 3.3.11 (6) that $\left(\frac{u,u'}{K}\right)$ splits, which means that ϵ must be 1 and then $A = \left(\frac{u,\pi u'}{K}\right)$. We also get from Theorem 3.3.11 (5) that there are $a, b \in K$, $b \neq 0$, for which $ua^2 + u'b^2 = 1$. Using this we can easily solve the equation $\mathrm{diag}[u, \pi u', -u\pi u'] = M^T \mathrm{diag}[u, \pi, -u\pi] M$ for the 3×3 matrix M. Choose, for instance,

$$M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/b & ua/b \\ 0 & a/b & 1/b \end{pmatrix}.$$

And this implies that $A \cong \left(\frac{u,\pi}{K}\right)$. □

We can now state the analogue dichotomy of Theorem 3.4.2 for quaternion algebras over non-Archimedean local fields.

Corollary 3.4.6 (Quaternion algebras over non-Archimedean local fields)**.** *In the notation of Theorem 3.4.5, any quaternion algebra over K is isomorphic either to $M_2(K)$ or, if it is a division algebra, to $\left(\frac{u,\pi}{K}\right)$.*

Remark 3.4.7. Although we will only need Theorem 3.4.5 for $\mathfrak{p}$-adic fields, the argument given above is valid for any non-Archimedean local field of characteristic $\neq 2$. We have not defined quaternion algebras over fields of characteristic 2.

3.4.2 Quaternion algebras over number fields

We already know how to extend the scalars of a quaternion algebra A defined over K to a field $L \mid K$. When K is a number field, we may take L to be a completion of K, since the quaternion algebras over local fields have been successfully described in §3.4.1.

Propositions 2.3.53 and 2.3.54 characterised all valuations on K, which, let us recall, are either

- Archimedean: in this case the valuation is given by $v_\sigma(x) = |\sigma x|$ where σ is an embedding of K into $\mathbb{C}$, and $|\cdot| = |\cdot|_\infty$ here denotes the usual absolute value in $\mathbb{C}$;
- or non-Archimedean: in which case, the valuation is the $\mathfrak{p}$-adic valuation $v_\mathfrak{p}$ where $\mathfrak{p}$ is a prime ideal of the ring of integers of K.

If we denote by K_v the completion of K with respect to the valuation v and by σ the embedding of K into K_v, then, as in (3.3.1), there is an isomorphism

$$\left(\frac{a,b}{K}\right) \otimes_\sigma K_v \cong \left(\frac{\sigma(a),\sigma(b)}{K_v}\right).$$

Note that the embedding σ is canonically determined by the (equivalence class of the) valuation v (see §§2.3.3) and so, once the valuation is indicated in the notation, we may denote the tensor product just described simply by $\left(\frac{a,b}{K}\right) \otimes_K K_v$ with no risk of confusion. Both notations mean the same thing.

In particular, for the Archimedean case, we know that K_v is either $\mathbb{C}$ or $\mathbb{R}$. For a complex embedding $\sigma : K \to \mathbb{C}$, Proposition 3.4.1 gives that $\left(\frac{\sigma(a),\sigma(b)}{\mathbb{C}}\right) \cong M_2(\mathbb{C})$. Similarly, for a real embedding, Proposition 3.4.2 implies that $\left(\frac{\sigma(a),\sigma(b)}{\mathbb{R}}\right)$ is isomorphic either to $M_2(\mathbb{R})$ or to the division algebra $\mathscr{H}$. For the non-Archimedean case, there is also a dichotomy between the matrix algebra and a uniquely determined division algebra (Corollary 3.4.6). In order to distinguish between these two situations, we make the following:

Definition 3.4.8. Let A be a quaternion algebra over a number field K. We say that A *splits* over K if $A \cong M_2(K)$. More generally, let v be a place of K (i.e., a representative of an equivalence class of valuations). If $A \otimes_\sigma K_v$ is a division algebra (which exists and is uniquely determined unless v is a complex place) we say that A is *ramified* at v. Otherwise, $A \otimes_\sigma K_v \cong M_2(K_v)$ and we say that A is *unramified* (or *split*) at v.

Denote by $\mathrm{Ram}(A)$ the set of places of K at which A is ramified. The subset of Archimedean places at which A is ramified is denoted by $\mathrm{Ram}_\infty(A)$ while the subset of non-Archimedean places are denoted by $\mathrm{Ram}_f(A)$. With this notation, $\mathrm{Ram}(A) = \mathrm{Ram}_\infty(A) \cup \mathrm{Ram}_f(A)$.

Clearly, $M_2(K)$ splits over every place of K. Conversely, it is a consequence of the Haße-Minkowski Theorem that if a quaternion algebra defined over K splits at every place of K, then it must be the matrix algebra $M_2(K)$.

Theorem 3.4.9 (Albert-Brauer-Haße-Noether Theorem). *Let A be a quaternion algebra over the number field K. Then A splits over K if and only if $A \otimes_\sigma K_v$ splits over K_v, for every place v of K.*

Proof. Let $A = \left(\frac{a,b}{k}\right)$. By Theorem 3.3.11 (5), A splits over K if and only if the quadratic equation $ax^2 + by^2 = 1$ has a solution in K. By the Haße-Minkowski Theorem (more precisely, by Corollary 3.2.4), this is true if and only if $ax^2 + by^2 = 1$ has a solution in K_v for every place v, which is equivalent to the quaternion algebra $A \otimes_\sigma K_v$ splitting over K_v, again by Theorem 3.3.11 (5). □

Furthermore, the information encoded in $\mathrm{Ram}(A)$ uniquely determines the quaternion algebra A up to isomorphism, as the next theorem shows.

Theorem 3.4.10. *Let A and A' be two quaternion algebras defined over the number field K. Then $A \cong A'$ if and only if $\mathrm{Ram}(A) = \mathrm{Ram}(A')$.*

Proof. We have the following sequence of equivalences, where the first follows from Theorem 3.3.13, the second from the Haße - Minkowski Theorem for quadratic spaces, the third from the simple observation that $(A_0)_v = (A \otimes_\sigma K_v)_0$ and, finally, the fourth is just Theorem 3.3.13 yet again:

$$\begin{aligned} A \cong A' &\iff \text{the quadratic spaces } A_0 \text{ and } A_0' \text{ are isometric} \\ &\iff (A_0)_v \text{ and } (A_0')_v \text{ are isometric for every place } v \text{ of } K \\ &\iff (A \otimes_\sigma K_v)_0 \text{ and } (A' \otimes_\sigma K_v)_0 \text{ are isometric for every place } v \text{ of } K \\ &\iff A \otimes_\sigma K_v \cong A' \otimes_\sigma K_v \text{ for every place } v \text{ of } K \end{aligned}$$

Now, for each complex Archimedean place v, $A \otimes_\sigma K_v \cong A' \otimes_\sigma K_v$. For every other place, there are only two possibilities, depending on weather or not the algebra is ramified at v (Theorems 3.3.3 and 3.4.5). Therefore, $\mathrm{Ram}(A) = \mathrm{Ram}(A')$ if and only if $A \otimes_\sigma K_v \cong A' \otimes_\sigma K_v$ for all v. □

Theorem 3.4.11 (Classification of Quaternion Algebras)**.** *Let K be a number field. If A is a quaternion algebra over K then* $\mathrm{Ram}(A)$ *is a finite set of even cardinality. Conversely, for any finite set of (non-complex) places S of K, if S is of even cardinality, then there exists a unique (up to isomorphism) quaternion algebra A over K such that* $\mathrm{Ram}(A) = S$.

Proof. See [36, Theorem 7.3.6]. □

3.5 Orders in quaternion algebras

Quaternion algebras are, in many aspects, similar to fields. In this section, we briefly introduce orders, which are the quaternion algebra analogues of rings of integers. These object will be fundamental for the definition of arithmetic and semi-arithmetic groups later on.

Definition 3.5.1. In a vector space V over a number field K, an *$\mathcal{O}_K$-lattice L* is a finitely generated $\mathcal{O}_K$-module contained in V. The lattice L is said to be *complete* if $L \otimes_{\mathcal{O}_K} K \cong V$.

In order to extend the concept of an algebraic integer to elements of a quaternion algebra A, we use the characterisation of algebraic integers given by Proposition 2.1.2:

Definition 3.5.2. An element α of a quaternion algebra A over K is an *integer* if $\mathcal{O}_K[\alpha]$ is an $\mathcal{O}_K$-lattice, i.e., if it is a finitely generated $\mathcal{O}_K$-module.

Just as in the case of field extensions, it follows that:

Proposition 3.5.3. *An element $\alpha \in A$ is an integer if and only if* $\mathrm{tr}(\alpha)$ *and* $\mathrm{n}(\alpha)$ *are in* $\mathcal{O}_K$.

Proof. Sufficiency follows from the fact that α satisfies the polynomial $X^2 - \mathrm{tr}(\alpha)X + \mathrm{n}(\alpha)$.

Conversely, let α be an integer in A. If α lies in the centre K of A, then $\alpha \in \mathcal{O}_K$ and the result follows.

Suppose $\alpha \in A \setminus K$. We analyse two cases.

If $K(\alpha)$, the smallest subalgebra of A containing K, is not an integral domain, then A is not a division algebra and $A \cong M_2(K)$. Note that if $X^2 - \mathrm{tr}(\alpha)X + \mathrm{n}(\alpha)$ were irreducible over K, then $K(\alpha)$ would be an integral domain, so we conclude that the matrix $\alpha \in M_2(K)$ must have eigenvalues in K. Therefore, α is of the form $\left(\begin{smallmatrix} a & b \\ 0 & c \end{smallmatrix}\right)$, for $a, b, c \in K$ (up to conjugation in $M_2(K)$). Since $\alpha^n = \left(\begin{smallmatrix} a^n & * \\ 0 & c^n \end{smallmatrix}\right)$ and $\mathcal{O}_K[\alpha]$ is an $\mathcal{O}_K$-lattice, it follows that $\mathcal{O}_K[a]$ and $\mathcal{O}_K[c]$ are finitely generated $\mathcal{O}_K$-submodules of K, which means that $a, c \in \mathcal{O}_K$ and hence that $\mathrm{tr}(\alpha), \mathrm{n}(\alpha) \in \mathcal{O}_K$.

If $K(\alpha)$ is an integral domain, then $L = K(\alpha)$ is a quadratic field extension of K, and the conjugate $\overline{\alpha}$ of α in A coincides with the field conjugate of α (see proof of Proposition 3.3.10). In particular, the reduced trace and reduced norm of α coincide with its trace and norm with respect to $L \mid K$, respectively. Since $\mathcal{O}_K[\alpha]$ and $\mathcal{O}_K[\overline{\alpha}]$ are finitely generated $\mathcal{O}_K$-modules, it follows that α and $\overline{\alpha}$ belong to the integral closure of $\mathcal{O}_K$ in L, i.e., to $\mathcal{O}_L$. So $\mathrm{tr}(\alpha)$ and $\mathrm{n}(\alpha)$ are both in $\mathcal{O}_L \cap K = \mathcal{O}_K$. □

Unlike the case of algebraic integers in number fields, the sum and product of two integers in a quaternion algebra may not be an integer. For this reason, we have the following:

Definition 3.5.4. An *order* $\mathcal{O}$ in a quaternion algebra A over K is a complete $\mathcal{O}_K$-lattice which is also a ring with unity. The order is said to be *maximal* when it is maximal with respect to inclusion.

We then have the following characterisation.

Proposition 3.5.5.

1. *$\mathcal{O}$ is an order in A if and only if $\mathcal{O}$ is a ring of integers in A that contains $\mathcal{O}_K$ and such that $\mathcal{O} \otimes_{\mathcal{O}_K} K = A$;*
2. *Every order is contained in a maximal order.*

Proof. Necessity in (1) is clear. Sufficiency follows from the same “determinant trick" that was applied in the proof of Theorem 2.2.19. Let $\{x_1, x_2, x_3, x_4\}$ be a basis of A such that each $x_i \in \mathcal{O}$. The reduced trace defines a non-singular symmetric bilinear form on A: $(x, y) \mapsto \mathrm{tr}(xy)$. So $d = \det(\mathrm{tr}(x_i x_j)) \neq 0$. Let $L = \{\sum a_i x_i \mid a_i \in \mathcal{O}_K\} \subset \mathcal{O}$, which is an $\mathcal{O}_K$-lattice. Each $x \in \mathcal{O}$ can be written as $\sum b_i x_i$, where $b_i \in K$. Multiplying both sides by x_j and taking traces, we obtain the system of equations: $\mathrm{tr}(x x_j) = \sum b_i \mathrm{tr}(x_i x_j)$, $j = 1, \ldots, 4$. Since $\mathrm{tr}(x x_j), \mathrm{tr}(x_i x_j) \in \mathcal{O}_K$, it follows from Cramer's Rule that $\mathcal{O} \subset \frac{1}{d}L$, so $\mathcal{O}$ is an $\mathcal{O}_K$-lattice (see Theorem 2.2.28). The other assumptions on $\mathcal{O}$ imply that it is indeed an order.

The second assertion follows form the characterisation in (1) and a Zorn's Lemma argument. □

CHAPTER 4

FUCHSIAN GROUPS

4.1 Möbius transformations and automorphisms of the Riemann sphere

4.1.1 Möbius transformations

The *Riemann sphere*, Σ, is the 2-dimensional sphere $\mathbb{S}^2 \subset \mathbb{R}^3$ endowed with the usual complex structure induced by stereographic projection, i.e., $\Sigma = \mathbb{C} \cup \{\infty\}$. An automorphism of Σ is a bijective meromorphic function $T : \Sigma \to \Sigma$ (or, from the Riemann surface point of view, a biholomorphic function from Σ to itself). It is well known from Complex Analysis that any such T must be a degree 1 rational function:

$$T(z) = \frac{az+b}{cz+d}, \quad a, b, c, d \in \mathbb{C} \text{ and } ad - bc \neq 0. \tag{4.1.1}$$

These functions are called *Möbius transformations* or *linear fractional transformations*.

Let $\mathrm{Aut}(\Sigma)$ denote the group of automorphisms of Σ, under composition. Note that there exists a homomorphism ϕ from the general linear group $\mathrm{GL}(2, \mathbb{C})$ to $\mathrm{Aut}(\Sigma)$ given by means of Möbius transformations:

$$\phi : A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \phi_A(z) = \frac{az+b}{cz+d}.$$

Given the considerations above, this map is surjective. The kernel of ϕ is precisely the scalar matrices $\lambda\mathrm{Id}$ where $\lambda \in \mathbb{C}^*$ and Id denotes the two-by-two identity matrix. This induces the isomorphism $\mathrm{PGL}(2, \mathbb{C}) \cong \mathrm{Aut}(\Sigma)$.

Let $\mathrm{SL}(2,\mathbb{C})$ be the subgroup of $\mathrm{GL}(2,\mathbb{C})$ of matrices with determinant 1. Accordingly, $\mathrm{PSL}(2,\mathbb{C})$ is the group $\mathrm{SL}(2,\mathbb{C})$ modulo the subgroup of scalar matrices with determinant 1, i.e., $\{\pm\mathrm{Id}\}$. For an automorphism of the form (4.1.1) we can always factor out $\sqrt{ad-bc}$ and write it as $z \mapsto \frac{a'z+b'}{c'z+d'}$ where $a'd'-b'c'=1$. It follows that ϕ restricts to a surjective homomorphism from $\mathrm{SL}(2,\mathbb{C})$ onto $\mathrm{Aut}(\Sigma)$, with kernel $\{\pm I\}$. In this way, we obtain the following isomorphisms:

Theorem 4.1.1. $\mathrm{PGL}(2,\mathbb{C}) \cong \mathrm{Aut}(\Sigma) \cong \mathrm{PSL}(2,\mathbb{C})$.

Remark 4.1.2. Note that we were able to obtain an isomorphism $\mathrm{PSL}(2,\mathbb{C}) \cong \mathrm{PGL}(2,\mathbb{C})$ because any non-zero complex number has a square root in $\mathbb{C}$. The same is not true over $\mathbb{R}$ and, in fact, $\mathrm{PSL}(2,\mathbb{R})$ is not isomorphic to $\mathrm{PGL}(2,\mathbb{R})$.

4.1.2 Conformal maps

A *conformal* map is a map that preserves angles. More precisely, given two Riemannian manifolds (M,g) and (N,h), a map $f : M \to N$ is conformal if the pull-back metric tensor f^*h is a multiple of g, i.e, if there exists a positive function λ on M such that $f^*h = \lambda g$. Alternatively, one can think of a conformal map as a map sending any pair of intersecting curves to another pair of curves intersecting with the same angle. For example, let U be a domain (i.e. connected open subset) of $\mathbb{C}$. Then a holomorphic map $f : U \to f(U) \subset \mathbb{C}$ with non-zero derivative at every point is conformal (here the metrics involved are understood to be the Euclidean metric). Indeed, let γ_1 and γ_2 be two curves intersecting at $p = \gamma_1(0) = \gamma_2(0)$ where the angle of intersection is the argument of $\gamma_1'(0)/\gamma_2'(0)$. Then $f \circ \gamma_1$ and $f \circ \gamma_2$ intersect at $f(p)$ with the same angle, since $(f\circ\gamma_1)'(0)/(f\circ\gamma_2)'(0) = \gamma_1'(0)/\gamma_2'(0)$. The converse statement is also true (see Proposition 4.1.3 (2) below), which explains why complex structures and conformal structures on (orientable) surfaces are often treated as the same.

Note that in the example given above, not only the measure of the angles were preserved but also their orientation. This is not always the case. Complex conjugation, for instance, preserves angles but reverses orientation. We must therefore distinguish between *orientation-preserving* conformal maps and *orientation-reversing* conformal maps. Let us define an anti-automorphism of Σ to be the composition of an automorphism of Σ with complex conjugation. In other words, an anti-automorphism is a map of the form $z \to T(\overline{z})$, where T is a Möbius transformation. Denote the set of all anti-automorphisms by $\overline{\mathrm{PGL}}(2,\mathbb{C})$ and observe that $\mathrm{PGL}(2,\mathbb{C}) \cup \overline{\mathrm{PGL}}(2,\mathbb{C})$ form a group, of which $\mathrm{PGL}(2,\mathbb{C})$ is an index 2 subgroup.

We list a few basic facts concerning conformal maps which can be found in most textbooks and therefore shall not be proved here:

Proposition 4.1.3. *1. The stereographic projection, as well as its inverse, are conformal maps.*

2. *Let $U \subset \mathbb{C}$ be a domain. A function $f : U \to f(U) \subset \mathbb{C}$ is an orientation-preserving conformal map if and only if f is holomorphic with non-zero derivative everywhere.*

3. *Every automorphism (resp. anti-automorphism) of Σ is an orientation-preserving (resp. orientation-reversing) conformal homeomorphism of Σ.*

4. *Every orientation-preserving (resp. orientation-reversing) conformal map from Σ to Σ is an automorphism (resp. anti-automorphism) of Σ (see [26, Theorem 2.11.4]).*

Note that the set of all conformal homeomorphisms of Σ form a group under composition, which we denote by $\mathrm{Conf}(\Sigma)$. Proposition 4.1.3 (3) shows that $\mathrm{PGL}(2, \mathbb{C}) \cup \overline{\mathrm{PGL}}(2, \mathbb{C}) < \mathrm{Conf}(\Sigma)$. Conversely, (4) shows the reverse inclusion, and together they describe the group of conformal homeomorphisms of Σ. We highlight this fact below:

Theorem 4.1.4. $\mathrm{Conf}(\Sigma) = \mathrm{PGL}(2, \mathbb{C}) \cup \overline{\mathrm{PGL}}(2, \mathbb{C})$.

4.1.3 The hyperbolic plane

Let $\mathbb{H}$ denote the hyperbolic plane, i.e., the upper half-plane

$$\{z \in \mathbb{C} \mid \Im z > 0\},$$

together with the Riemannian metric

$$g_z(u, v) = \frac{\langle u, v\rangle}{y^2}, \tag{4.1.2}$$

where $u, v \in \mathbb{C} \cong \mathrm{T}_x\mathbb{H}$, $z = x + iy \in \mathbb{H}$ and $\langle \cdot, \cdot \rangle$ denotes the Euclidean metric on $\mathbb{C}$. We recall that, with respect to the hyperbolic metric, the geodesics are the vertical lines and semi-circles orthogonal to the boundary of $\mathbb{H}$. Let $\mathrm{Isom}(\mathbb{H})$ denote the group of isometries of $\mathbb{H}$ whereas $\mathrm{Isom}^+(\mathbb{H})$ denotes the index 2 subgroup of isometries that preserve orientation.

Being an open subset of the Riemann surface Σ, $\mathbb{H}$ naturally inherits a complex structure, as does the unit disc $\mathbb{D} = \{z \in \mathbb{C} \mid |z| < 1\}$. As a matter of fact, they are biholomorphic: the *Cayley transform* $W : \mathbb{H} \to \mathbb{D}$ given by $W(z) = \frac{z-i}{z+i}$ is a biholomorphism, with inverse $W^{-1}(z) = \frac{z+1}{i(z-1)}$. In particular, an automorphism S of $\mathbb{H}$, fixing the element i, induces an automorphism $W \circ S \circ W^{-1}$ of $\mathbb{D}$ fixing the origin 0. The latter is completely described by the Schwarz Lemma, namely, it is a rotation $z \mapsto e^{i\theta}z$ for some $\theta \in \mathbb{R}$, whence it follows that S must be a Möbius transformation. If we consider now any automorphism T of $\mathbb{H}$, let $a, b \in \mathbb{R}$, $a > 0$, be such that $T(i) = ai + b$ and define U to be $U(z) = az + b$. The transformation U is clearly an automorphism of $\mathbb{H}$ mapping i to $T(i)$, so that $U^{-1} \circ T$ is an automorphism of $\mathbb{H}$ fixing i and thus must be a Möbius transformation. We have showed that $\mathrm{Aut}(\mathbb{H}) < \mathrm{PGL}(2, \mathbb{C})$. In other words, the automorphisms of $\mathbb{H}$ are precisely those Möbius transformations that stabilise the subset $\mathbb{H} \subset \Sigma$. It is well known that the subgroup of $\mathrm{PGL}(2, \mathbb{C})$ stabilising $\mathbb{H}$ is the group $\mathrm{PSL}(2, \mathbb{R})$ (indeed, any automorphism of Σ stabilising $\mathbb{H}$ must also stabilise the extended real line $\mathbb{R} \cup \{\infty\}$. The stabiliser of $\mathbb{R} \cup \{\infty\}$ is easily seen to be the subgroup of "matrices with real coefficients", i.e., $\mathrm{PGL}(2, \mathbb{R})$. Among these, the ones that stabilise $\mathbb{H}$ are precisely those with positive determinant.). We have just established that:

Theorem 4.1.5. *The group of automorphisms of the hyperbolic plane is* $\mathrm{Aut}(\mathbb{H}) = \mathrm{PSL}(2, \mathbb{R})$.

It follows from Proposition 4.1.3 (2) that the orientation-preserving conformal homeomorphisms of $\mathbb{H}$ are precisely the automorphisms of $\mathbb{H}$ (i.e., biholomorphisms from $\mathbb{H}$ to itself). If we denote by $\mathrm{Conf}(\mathbb{H})$ the group of conformal homeomorphisms of $\mathbb{H}$, and by $\mathrm{Conf}^+(\mathbb{H})$ the index 2 subgroup of orientation-preserving homeomorphisms, we have that $\mathrm{Conf}^+(\mathbb{H}) = \mathrm{PSL}(2, \mathbb{R})$. Just as we described the whole group of conformal homeomorphisms of Σ in Theorem 4.1.1, we may describe the group $\mathrm{Conf}(\mathbb{H})$ in a similar fashion. Now, conjugation is not a transformation from $\mathbb{H}$ to itself, but $z \mapsto -\overline{z}$ is, and it is seen to be orientation reversing. Given any orientation-reversing conformal homeomorphism S of $\mathbb{H}$, then $z \mapsto -\overline{S(z)}$ is an orientation-preserving conformal homeomorphism and so it must equal some $T = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{PSL}(2, \mathbb{R})$. This means that $S(z) = \frac{-a\overline{z}-b}{c\overline{z}+d}$ where $\left(\begin{smallmatrix} -a & -b \\ c & d \end{smallmatrix}\right) \in \mathrm{PGL}(2, \mathbb{R})$. Conversely, it is immediate that any $S = \left(\begin{smallmatrix} a' & b' \\ c' & d' \end{smallmatrix}\right) \in \mathrm{PGL}(2, \mathbb{R})$, for which $a'd' - b'c' = -1$, acts on $\mathbb{H}$ as an orientation-reversing conformal map in the following way: $z \mapsto \frac{a'\overline{z}+b'}{c'\overline{z}+d'}$. We thus obtain:

Theorem 4.1.6. $\mathrm{Conf}(\mathbb{H}) \cong \mathrm{PGL}(2, \mathbb{R})$.

Moreover, $\mathrm{PGL}(2, \mathbb{R})$ *acts on* $\mathbb{H}$ *as follows:*

> *Each* $T \in \mathrm{PSL}(2, \mathbb{R}) < \mathrm{PGL}(2, \mathbb{R})$ *acts on* $\mathbb{H}$ *as a Möbius transformation;*
>
> *Each* $S \in \mathrm{PGL}(2, \mathbb{R})$ *for which* $\det S < 0$ *acts on* $\mathbb{H}$ *as* $z \mapsto S(\overline{z})$.

Finally, we point out that the action of $\mathrm{PGL}(2, \mathbb{R})$ just described preserves the metric tensor introduced in (4.1.2), meaning that every element of $\mathrm{PGL}(2, \mathbb{R})$ acts on $\mathbb{H}$ as an isometry. Conversely, every isometry is *a fortiori* a conformal map, whence we conclude that:

Theorem 4.1.7. $\mathrm{Isom}(\mathbb{H}) \cong \mathrm{PGL}(2, \mathbb{R})$ *and* $\mathrm{Isom}^+(\mathbb{H}) \cong \mathrm{PSL}(2, \mathbb{R})$.

4.1.4 The hyperbolic 3-space

Although we will mostly work with 2-dimensional hyperbolic surfaces, for the sake of completeness let us briefly recall the definition of the hyperbolic n-space $\mathbb{H}^n$ and, in particular, the action of $\mathrm{PSL}(2,\mathbb{C})$ on $\mathbb{H}^3$. The following is a short account where formal statements, as well as their proofs, have been omitted. For a good exposition about n-dimensional hyperbolic models, the reader is referred to [37, Chapter 2].

We define the hyperbolic n-space[1] $\mathbb{H}^n$ to be the set

$$\{(x_1,\dots,x_n)\in\mathbb{R}^n \mid x_n>0\},$$

together with the Riemannian metric

$$g_x^{\mathbb{H}^n}(u,v)=\frac{\langle u,v\rangle}{x_n^2},$$

where $u,v\in\mathbb{R}^n\cong \mathrm{T}_x\mathbb{H}^n$, $x=(x_1,\dots,x_n)$ and $\langle\cdot,\cdot\rangle$ denotes the Euclidean metric on $\mathbb{R}^n$ (compare with (4.1.2)). According to this notation, $\mathbb{H}^2$ is just the hyperbolic plane $\mathbb{H}$ as defined in the previous subsection.

As in the 2-dimensional case, the *boundary at infinity* of $\mathbb{H}^n$ is the set $\partial\mathbb{H}^n=\{x\in\mathbb{R}^n \mid x_n=0\}\cup\{\infty\}$ and the compactification of $\mathbb{H}^n$ is the space $\overline{\mathbb{H}}^n=\mathbb{H}^n\cup\partial\mathbb{H}^n$. Note that $\overline{\mathbb{H}}^n$ is topologically an n-dimensional (closed) ball. The geodesics of $\mathbb{H}^n$ are the vertical lines and the semi-circles orthogonal to $\{x_n=0\}$. More generally, the *k-subspaces* of $\mathbb{H}^n$ are the k-spheres and k-planes orthogonal to the boundary $\partial\mathbb{H}^n$. We remark that every k-subspace is isometric to $\mathbb{H}^k$.

Let $\mathbb{D}^n=\{x\in\mathbb{R}^n \mid \langle x,x\rangle<1\}$ be the open unit ball of $\mathbb{R}^n$. Equip $\mathbb{D}^n$ with the following metric tensor:

$$g_x^{\mathbb{D}^n}(\cdot,\cdot)=\left(\frac{2}{1-|x|^2}\right)^2\langle\cdot,\cdot\rangle,$$

where $x\in\mathbb{D}^n$, $\langle\cdot,\cdot\rangle$ denotes the Euclidean metric. Let $|\cdot|$ denote the Euclidean norm. The space $(\mathbb{D}^n,g^{\mathbb{D}^n})$ is called the *disc model* or the *Poincaré model* and is isometric to the half-space $(\mathbb{H}^n,g^{\mathbb{H}^n})$. Indeed, the map taking $\mathbb{D}^n$ to $\mathbb{H}^n$ is the *inversion* of $\mathbb{R}^n$ with centre $(0,\dots,0,-1)$ and radius $\sqrt{2}$:

$$\psi(x_1,\dots,x_n)=\frac{(2x_1,\dots,2x_{n-1},1-|x|)}{|x|^2+2x_n+1}. \tag{4.1.3}$$

Direct calculations show that ψ preserves the metric tensor and is therefore an isometry. The half-space model $\mathbb{H}^n$ and the disc model $\mathbb{D}^n$ are called *conformal*

[1] The reader should not confuse the hyperbolic n-space $\mathbb{H}^n$ with the cartesian product of n copies of the hyperbolic plane $\mathbb{H}$. The latter will always be denoted by $(\mathbb{H})^n$.

models of the hyperbolic space, since their metric tensors are conformal to the Euclidean metric tensor.

In the disc model, the boundary at infinity of $\mathbb{D}^n$ is the set $\partial\mathbb{D}^n = \mathbb{S}^{n-1} = \{x \in \mathbb{R}^n \mid |x| = 1\}$ and $\overline{\mathbb{D}}^n = \mathbb{D}^n \cup \partial\mathbb{D}^n$. Note that ψ takes $\partial\mathbb{H}^n$ to $\partial\mathbb{D}^n$. Furthermore, the geodesics in $\mathbb{D}^n$ are the diameters together with the semi-circles orthogonal to the boundary of $\mathbb{D}^n$.

The map ψ described in (4.1.3) is a particular case of a type of transformation called *inversion*: let S be a sphere in $\mathbb{R}^n$ centered in x_0 with radius r. The inversion along the sphere S is defined to be the map ϕ from $\mathbb{R}^n \cup \{\infty\}$ to $\mathbb{R}^n \cup \{\infty\}$ given by

$$\phi(x) = x_0 + r^2 \frac{x - x_0}{|x - x_0|^2},$$
$$\phi(\infty) = x_0, \ \ \phi(x_0) = \infty,$$

where $|\cdot|$ denotes the Euclidean norm. Inversions are orientation-reversing conformal maps that send k-spheres and k-planes to k-spheres and k-planes. For this reason, the k-subspaces in the disc model (i.e., the image of the k-subspaces through the map ψ^{-1}) are the k-spheres and k-planes orthogonal to the boundary $\partial\mathbb{D}^n$.

Furthermore, every inversion along spheres orthogonal to the boundary are isometries (in both models). We point out that inversions along spheres orthogonal to the boundary together with reflections across Euclidean planes orthogonal to the boundary generate the group of isometries, in the models $\mathbb{H}^n$ and $\mathbb{D}^n$ ([37, Proposition 2.1.28]).

Every isometry f from $\mathbb{H}^n$ to itself can be naturally extended to a homeomorphism f of $\overline{\mathbb{H}}^n$. Moreover, f is uniquely determined by its trace $f|_{\partial\mathbb{H}^n}$. An analogous statement holds for $\mathbb{D}^n$.

In the particular case of $n = 3$, the group of isometries of $\mathbb{D}^3$ is generated by inversions along spheres orthogonal to the boundary $\partial\mathbb{D}^3 = \mathbb{S}^2$ as well as by reflections across planes orthogonal to $\mathbb{S}^2$. The trace of these transformations on the boundary are inversions along circles in S^2. Conversely, for an inversion T along a circle C on the boundary $\mathbb{S}^2$, let S be a 2-space that intersects $\mathbb{S}^2$ on C, and choose f to be the inversion along S. The trace of f on the boundary is T. The group of isometries of $\mathbb{D}^3$ (or $\mathbb{H}^3$) is thus naturally isomorphic to the group of automorphisms of $\mathbb{S}^2$ generated by inversions along circles. Regarding $\mathbb{S}^2$ as the Riemann Sphere Σ, every inversion is an element of $\mathrm{Conf}(\Sigma)$. So it happens that the group generated by such inversions is the whole group of conformal homeomorphisms. In particular, the subgroup of orientation-preserving isometries of $\mathbb{H}^3$ is identified with the subgroup of orientation-preserving conformal homeomorphisms of Σ, that is $\mathrm{PSL}(2, \mathbb{C})$:

Proposition 4.1.8. $\mathrm{Isom}(\mathbb{H}^3) \cong \mathrm{Conf}(\Sigma)$ *and* $\mathrm{Isom}^+(\mathbb{H}^3) \cong \mathrm{PSL}(2, \mathbb{C})$.

One way to picture the action of $\mathrm{PSL}(2, \mathbb{C})$ on $\mathbb{H}^3$ is as follows: any point $p \in \mathbb{H}^3$ is the intersection of two geodesics lines, say, γ_1 and γ_2. Each γ_i has endpoints α_i and β_i on the boundary $\partial\mathbb{H}^3 \cong \Sigma$. A transformation $T \in \mathrm{PSL}(2, \mathbb{C})$ then acts on α_i and β_i. Consider the uniquely determined geodesics connecting $T(\alpha_1)$ to $T(\beta_1)$ and $T(\alpha_2)$ to $T(\beta_2)$. They must intersect (in a single point). Their intersection determines $T(p)$.

4.2 Fuchsian and Kleinian groups

In this section we study in greater detail the action of $\mathrm{PSL}(2, \mathbb{R})$ on $\mathbb{H}$ as its group of orientation-preserving isometries. Henceforth we shall abuse terminology and refer to elements of $\mathrm{PSL}(2, \mathbb{R})$ as matrices, even though they are really only matrices up to a sign. Also, we often identify an element of $\mathrm{PSL}(2, \mathbb{R})$ with its correspondent Möbius transformation. As a consequence, we remark that any transformation T of the form $z \mapsto \frac{az+b}{cz+d}$, where $a, b, c, d \in \mathbb{R}$ satisfy $ad - bc > 0$, may be regarded as an element of $\mathrm{PSL}(2, \mathbb{R})$. Indeed, T is unaltered when all of its coefficients are divided by $\sqrt{ad - bc}$, and the resulting transformation is the one induced by a matrix in $\mathrm{PSL}(2, R)$.

The first important thing to notice is the transitivity of the aforementioned action. We refer to $\mathbb{R} \cup \{\infty\}$ as the *boundary at infinity* and denote it by $\partial\mathbb{H}$.

Proposition 4.2.1. *The action of* $\mathrm{PSL}(2, \mathbb{R})$ *is transitive on* $\mathbb{H}$ *and doubly transitive on* $\partial\mathbb{H}$.

Proof. For any $a + bi$ with $a, b \in \mathbb{R}$ and $b > 0$, the transformation $z \mapsto bz + a$ maps i to $a + bi$.

Now, let $a, b \in \mathbb{R}$, then $z \mapsto \frac{bz-a}{z-1}$ maps the pair $(0, \infty)$ to (a, b). Similarly, $z \mapsto z + b$ maps the pair $(0, \infty)$ to (b, ∞) and $z \mapsto -\frac{1}{z}$ takes $(0, \infty)$ to $(\infty, 0)$. □

4.2.1 Types of transformations

Since $\mathbb{H} \cup \partial\mathbb{H}$ is topologically a disc, we know that any automorphism of $\mathbb{H}$ must have at least one fixed point, which could be in $\mathbb{H}$ or on the boundary at infinity. More precisely, z is a fixed point of the nonidentity transformation $T = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{PSL}(2, \mathbb{R})$ if and only if

$$z = \frac{az + b}{cz + d},$$

which leads to the equation

$$cz^2 + (d-a)z - b = 0. \tag{4.2.1}$$

Suppose first that $c = 0$, then $ad = 1$. If $a = d = 1$, then $T(z) = z + b$ and its only fixed point is ∞. If, $a \neq d$ then the fixed points of T are $z = b/(d-a)$ and $z = \infty$.

When $c \neq 0$ then (4.2.1) is a quadratic equation with discriminant

$$\Delta = (d-a)^2 + 4bc = (a+d)^2 - 4, \tag{4.2.2}$$

where we have used that $ad - bc = 1$. If $\Delta < 0$ then (4.2.1) has two conjugate complex roots, which means that T has one fixed point in $\mathbb{H}$. If $\Delta > 0$, T has two distinct fixed points in $\mathbb{R} \subset \partial\mathbb{H}$. Finally, when $\Delta = 0$, T has only one fixed point in $\mathbb{R} \subset \partial\mathbb{H}$. According to these three cases, we will classify the nonidentity elements of $\mathrm{PSL}(2,\mathbb{R})$.

Let tr denote the trace of a matrix. The maps $\gamma \mapsto |\mathrm{tr}\,\gamma|$ and $\gamma \mapsto \mathrm{tr}^2\,\gamma$ on $\mathrm{SL}(2,\mathbb{R})$ are constant on each equivalence class (i.e. on each coset of $\{\pm\mathrm{Id}\}$) and thus they both descend to well-defined functions on $\mathrm{PSL}(2,\mathbb{R})$.

Remark 4.2.2. Although the trace function is not defined on $\mathrm{PSL}(2,\mathbb{R})$, we will often abuse terminology and refer to the trace of an element of $\mathrm{PSL}(2,\mathbb{R})$ meaning the trace of any of its lifts to $\mathrm{SL}(2,\mathbb{R})$. This is a number defined only up to a sign, so we shall employ this practice in situations where the sign is irrelevant. The same is true, of course, for $\mathrm{PSL}(2,\mathbb{C})$.

The expression (4.2.2) for Δ shows that each class of elements of $\mathrm{PSL}(2,\mathbb{R})$ can be characterised in terms of the trace of T, as we next explain in more detail.

Elliptic elements. An element T of $\mathrm{PSL}(2,\mathbb{R})$ is called *elliptic* when $|\mathrm{tr}\,T| < 2$. In this case, T has only one fixed point in $\mathbb{H}$ (and none in $\partial\mathbb{H}$). Let $z_0 \in \mathbb{H}$ be this fixed point. Let $S \in \mathrm{PSL}(2,\mathbb{R})$ be such that $S(z_0) = i$. Then $W = STS^{-1} = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is an elliptic element with fixed point i. It follows from direct calculation that $a = d$ and $b = -c$. Since $1 = ad - bc = a^2 + c^2$, we may write $a = \cos\theta$ and $c = \sin\theta$. Thus, T is conjugate in $\mathrm{PSL}(2,\mathbb{R})$ to the rotation matrix $R_\theta = \left(\begin{smallmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{smallmatrix}\right)$. Furthermore, note that $R_{\theta+\pi} = R_\theta$ in $\mathrm{PSL}(2,\mathbb{R})$, so we need only to consider $\theta \in (0,\pi)$. It can be checked that if $\theta, \theta' \in (0,\pi)$ are two distinct angles then $\mathbb{R}_\theta$ is not conjugate to $\mathbb{R}_{\theta'}$ in $\mathrm{PSL}(2,\mathbb{R})$.

Hyperbolic Elements. An element T of $\mathrm{PSL}(2,\mathbb{R})$ is said to be a *hyperbolic* element when $|\mathrm{tr}\,T| > 2$. In this case, T has two fixed points $\alpha, \beta \in \partial\mathbb{H}$. According to

Proposition 4.2.1, there exists $S \in \mathrm{PSL}(2, \mathbb{R})$ such that S maps α to 0 and β to ∞. It is then easy to see that:

$$STS^{-1}(z) = \lambda z, \quad \text{for some } \lambda > 0,\ \lambda \neq 1.$$

Let $U_\lambda(z) = \lambda^2 z$ for $\lambda > 0, \lambda \neq 1$. Note that if $B(z) = -1/z$, then $BU_\lambda B^{-1} = U_{\lambda^{-1}}$, so that U_λ is conjugate to $U_{\lambda^{-1}}$ in $\mathrm{PSL}(2, \mathbb{R})$. Conversely, if U_λ is conjugate to U_κ then $\mathrm{tr}^2\, U_\lambda = \mathrm{tr}^2\, U_\kappa$ and a simple calculation shows that $\kappa = \lambda$ or $\kappa = \lambda^{-1}$.

We conclude that every hyperbolic element of $\mathrm{PSL}(2, \mathbb{R})$ is conjugate to a unique transformation of the form U_λ, $\lambda > 1$, i.e., to some

$$U_\lambda = \begin{pmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{pmatrix}, \quad \lambda > 1.$$

Remark 4.2.3. More generally, when $T \in \mathrm{PSL}(2, \mathbb{C})$ and $|\mathrm{tr}\, T| > 2$, we call T *loxodromic*. If, moreover, $\mathrm{tr}\, T$ is real, then T is said to be hyperbolic. A hyperbolic element is then a particular kind of loxodromic element with real trace.

Parabolic Elements. An element T of $\mathrm{PSL}(2, \mathbb{R})$ is said to be *parabolic* when $|\mathrm{tr}\, T| = 2$. In this case, T has a single fixed point $\alpha \in \partial\mathbb{H}$. Let $S \in \mathrm{PSL}(2, \mathbb{R})$ be such that $S(\alpha) = \infty$. Then STS^{-1} is parabolic with fixed point ∞ and it is easy to see that $STS^{-1}(z) = z + b$. If $V(z) = z/|b|$ then

$$(SV)T(SV)^{-1} = z \pm 1.$$

Trying to solve $A \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) A^{-1} = \left(\begin{smallmatrix} 1 & -1 \\ 0 & 1 \end{smallmatrix}\right)$ for $A \in \mathrm{PSL}(2, \mathbb{R})$, shows that $z \mapsto z + 1$ cannot be conjugate to $z \mapsto z - 1$ in $\mathrm{PSL}(2, \mathbb{R})$.

Summarising, a non-identity transformation $T \in \mathrm{PSL}(2, \mathbb{R})$ is conjugate to exactly one of the following:

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \ \theta \in (0, \pi); \qquad \begin{pmatrix} \lambda & 0 \\ 0 & \frac{1}{\lambda} \end{pmatrix}, \ \lambda > 1; \qquad \begin{pmatrix} 1 & \pm 1 \\ 0 & 1 \end{pmatrix}$$

if it is, respectively, elliptic, hyperbolic or parabolic.

4.2.2 Continuous group actions and homogeneous spaces

$\mathrm{PSL}(2, \mathbb{R})$ **and** $\mathrm{PSL}(2, \mathbb{C})$ **as Topological Groups.** Consider the closed subspace X of $\mathbb{R}^4$ defined by:

$$X = \{(a, b, c, d) \in \mathbb{R} \mid ad - bc = 1\}.$$

Once we identify $\mathrm{SL}(2, \mathbb{R})$ with the subspace X we may endow it with the subspace topology, inherited from $\mathbb{R}^4$. In this way, it is easy to check that $\mathrm{SL}(2, \mathbb{R})$ becomes a topological group. Furthermore, the quotient $\mathrm{PSL}(2, \mathbb{R})$ also becomes a topological group with the quotient topology (a Lie Group, in fact). Note that we can also induce from $\mathbb{R}^4$ a metric on $\mathrm{PSL}(2, \mathbb{R})$. The function $|\mathrm{tr}|$ on $\mathrm{PSL}(2, \mathbb{R})$ is continuous with respect to this topology.

It follows from Theorem 4.1.7 that the group of all isometries of $\mathbb{H}$, $\mathrm{Isom}(\mathbb{H})$, may be topologised similarly.

In an entirely analogous manner, we identify $\mathrm{PSL}(2, \mathbb{C})$ with a closed subspace of $\mathbb{C}^4$, from where it inherits the subspace topology. The group $\mathrm{PSL}(2, \mathbb{C})$ becomes a topological group endowed with a metric.

Definition 4.2.4. A *Fuchsian group* is a discrete subgroup of $\mathrm{PSL}(2, \mathbb{R})$. A *Kleinian group* is a discrete subgroup of $\mathrm{PSL}(2.\mathbb{C})$.

Note that if $\widetilde{\Gamma}$ is a discrete subgroup of $\mathrm{Isom}(\mathbb{H}) \cong \mathrm{PGL}(2, \mathbb{R})$, then the subgroup $\Gamma = \widetilde{\Gamma} \cap \mathrm{SL}(2, \mathbb{R})$ is a Fuchsian group.

We know that $\mathrm{PSL}(2, \mathbb{R})$ acts on $\mathbb{H}$ by isometries (in particular, also by homeomorphisms), and therefore so does any Fuchsian group. We will next recall some definitions in the general setting of a topological group G acting on a topological space X by homeomorphism, and analyse which properties are then reflected on the quotient space (the space of orbits).

A *left action* of a group G on a set X is a map from $\varphi : G \times X$ to X, $(g, x) \mapsto \varphi(g, x)$, satisfying:

(i) $\varphi(e, x) = x$, for every $x \in X$, where e is the identity element of G;

(ii) $\varphi(g_2, \varphi(g_1, x)) = \varphi(g_2 g_1, x)$, for every $x \in X$ and every $g_1, g_2 \in G$.

It is customary to denote $\varphi(g, x)$ simply by $g \cdot x$ or even gx. Note that for each $g \in G$, the map $x \mapsto gx$ is a bijection. Equivalently, one can define an action to be a homomorphism $\rho : G \to \mathrm{Bij}(X)$, where $\mathrm{Bij}(X)$ denotes the group of bijections from X to itself.

A *right action* is defined analogously. When it is not relevant for the discussion or when it is clear from the context, we will refer to an action without specifying if it is on the left or on the right.

The action ρ is *faithful* if its kernel is trivial. A stronger condition would be that every element of G different from the identity acts on X without fixed points, i.e.: if $g \in G$ and there exists $x \in X$ such that $\rho(g)(x) = x$ then $g = e$. If this is the case, we say that the action ρ is *free*.

If $x, y \in X$ are such that there exists $g \in G$ satisfying $gx = y$, we say that x and y are in the same orbit. The *G-orbit* (or just orbit, when it is clear which group is acting) of an element x is the set $Gx = \{gx \mid g \in G\}$. Note that the set of all G-orbits form a partition of the space X (in other words, being in the same orbit is an equivalence relation), and we denote this set by $G\backslash X$. There exists a canonical projection

$$\begin{aligned} \pi : X &\to G\backslash X \quad . \\ x &\mapsto Gx \end{aligned} \tag{4.2.3}$$

When X is a topological space, the projection π becomes a quotient map once we equip $G\backslash X$ with the *quotient topology* as follows: $U \subset G\backslash X$ is open if and only if $\pi^{-1}(U) \subset X$ is open. Note that π is an open map. Indeed, let $U \subset X$ be open, then $\pi^{-1}(\pi(U)) = \bigcup_{g \in G} gU$ is open and so, by definition, $\pi(U)$ is open.

The action is said to be *transitive* if the whole space X is a G-orbit, i.e., if, for every $x, y \in X$, there exists $g \in G$ such that $gx = y$.

The *stabiliser* (or *isotropy group*) of a point $x \in X$ is the subgroup $G_x = \{g \in G \mid gx = x\}$ of G.

An *action by homeomorphisms* is an action of G on the topological space X such that, for every $g \in G$, the map $x \mapsto gx$ is continuous (and hence a homeomorphism of X into itself). Just like before, this action can be regarded as a homomorphism ρ from G to $\mathrm{Bij(X)}$, except now ρ takes values in a smaller subgroup of $\mathrm{Bij(X)}$, namely, the group $\mathrm{Homeo}(X)$ of homeomorphisms of X, i.e., $\rho : G \to \mathrm{Homeo}(X)$. Moreover, we say that G acts *continuously* if the map from $G \times X \to X$, $(g, x) \mapsto gx$, is continuous. This is equivalent to ρ being continuous when $\mathrm{Homeo}(X)$ is endowed with the compact-open topology.

Henceforth we assume that X is a topological space and that G is a topological group acting *continuously* on X.

When X is assumed to be Hausdorff, the stabiliser G_x of any point x is a closed subgroup G. Give G/G_x the quotient topology. Then there is a continuous bijection between G/G_x and the G-orbit of x, mapping a coset gG_x to the point gx. If G and X are sufficiently well-behaved spaces, this bijection turns out to be a homeomorphism:

Proposition 4.2.5. *Let G be a second-countable locally compact topological group and let X be a locally compact Hausdorff space. If G acts continuously and transitively on X, then the map $h : G/G_x \to X$ taking gG_x to gx is a homeomorphism, where x is any point in X.*

Proof. The projection $G \to G/G_x$ is continuous, so it suffices to show that the *orbit map* $G \to X$, $g \mapsto gx$ is open. Let U be an open subset of G and let g be any element in U. We want to check that gx is an interior point of Ux. Choose a compact neighbourhood V of the identity element $e \in G$ such that $V = V^{-1}$ and $gV^2 \subset U$ (such a neighbourhood can easily be constructed in a topological group). Since G is second-countable, there are countably many elements $g_1, g_2, \dots$ of G such that $G = \bigcup_n g_nV$. Then $X = \bigcup_n g_nVx$, where each g_nVx is compact and hence closed in the Hausdorff space X. Moreover, X being a locally compact Hausdorff space, it follows that some g_nVx, and thus Vx, must have nonempty interior (Baire Category Theorem). Take $v \in V$ for which vx is an interior point of Vx. Now, $gx = gv^{-1}vx$ so that gx is an interior point of $gv^{-1}Vx \subset Ux$, given our choice of V. □

The case we are most interested in is when $G = \mathrm{PSL}(2, \mathbb{R})$ or $\mathrm{SL}(2, \mathbb{R})$ and $X = \mathbb{H}$. These objects satisfy all the hypothesis in Proposition 4.2.5 and the action via Möbius transformations is clearly continuos. When $\mathrm{SL}(2, \mathbb{R})$ acts on $\mathbb{H}$, one can check that the stabilizer of the element $i \in \mathbb{H}$ is isomorphic to the special orthogonal group

$$\mathrm{SO}(2, \mathbb{R}) = \{A \in SL(2, \mathbb{R}) \mid AA^T = \mathrm{Id}\}.$$

Proposition 4.2.5 then gives us that $\mathbb{H}$ is homeomorphic to $\mathrm{SL}(2, \mathbb{R})/\mathrm{SO}(2, \mathbb{R})$. Note also that $\mathrm{SO}(2, \mathbb{R})$ is compact. Indeed, it is closed and bounded (as a subset of $\mathbb{R}^4$).

Let H be a subgroup of G. The quotient topology on G/H is Hausdorff if and only if H is a closed subgroup. Furthermore, the natural projection $p : G \to G/H$ is an open map, which can be proved with the same argument used for the projection π in (4.2.3) (as a matter of fact, H acts continuously on G via multiplication on the right, so G/H may be regarded as the orbit space of this right action). If K is a compact subgroup of the Hausdorff topological group G (so K is, in particular, closed), then the projection $p : G \to G/K$ is also closed: for a closed subset $F \subset G$, $p^{-1}(p(F)) = FK$ and the latter is closed, being the product of a closed subset and a compact subset of a topological group; it follows that $p(F)$ is closed.

Moreover, when G is a locally compact Hausdorff topological group and K is a compact subgroup, we claim that the projection map p is also *proper*. Indeed, let C be a compact subset of G/K. Cover the whole space G with pre-compact open neighbourhoods $\{V_i\}$, then $\{p(V_i)\}$ is a collection of open neighbourhoods covering G/K and so C is covered by finitely many such neighbourhoods; in particular, $p^{-1}(C)$ is a closed set covered by the finite union $\bigcup p^{-1}(p(V_i)) = \bigcup KV_i \subset \bigcup K\overline{V_i}$ the latter being clearly a compact set.

Note that the group G itself acts continuously on the quotient G/K and, more generally, so does any subgroup $H < G$. The next proposition characterises the discrete subgroups in this setting.

Proposition 4.2.6. *Let G, K and p be as above. Let Γ be a subgroup of G. Then Γ is discrete if and only if it satisfies the following property:*

$$\text{For any two compact sets } C_1, C_2 \subset G/K \text{ the set } \{g \in \Gamma \mid gC_1 \cap C_2 \neq \emptyset\} \text{ is finite.} \tag{4.2.4}$$

Proof. Suppose Γ is discrete. If $g \in G$ is such that $gC_1 \cap C_2 \neq \emptyset$ then $gp^{-1}(C_1) \cap p^{-1}(C_2) \neq \emptyset$. Let $D_i = p^{-1}(C_i)$. According to the discussion above, D_1 and D_2 are compact and therefore so is $D_2(D_1)^{-1}$. Then $g \in \Gamma \cap D_2(D_1)^{-1}$ and this intersection is finite.

Conversely, let V be a compact neighbourhood of the identity $e \in G$. If any $g \in \Gamma$ is in V then, in particular, one would have $gK \cap VK \neq \emptyset$. So condition (4.2.4), when C_1 is the point $p(e) = K$ and $C_2 = p(V)$, implies that there are only finitely many elements of Γ in V, and the result follows. □

Corollary 4.2.7. *If G and K are as above, let Γ be a discrete subgroup of G. Then every $x \in G/K$ has a neighbourhood V such that $\{g \in \Gamma \mid gV \cap V \neq \emptyset\} = \{g \in \Gamma \mid gx = x\} = \Gamma_x$. Moreover, this set is finite.*

Proof. Start with a compact neighbourhood U of x. By Proposition 4.2.6, there are only finitely many $g \in \Gamma$ for which $gU \cap U \neq \emptyset$. Then choose a neighbourhood V that separates x from each of these (finitely many) gx that are actually distinct from x. □

Corollary 4.2.8. *If G, K and p are as above, let Γ be a discrete subgroup of G. If x and y are two points in G/K that are not on the the same Γ-orbit, then there are neighbourhoods U and V of x and y, respectively, such that $gU \cap V = \emptyset$ for every $g \in G$. In particular, the quotient space $\Gamma \backslash G/K$ is Hausdorff.*

Proof. This is simply an application of Proposition 4.2.6 with a similar adjustment as the one made in the proof above: take compact neighbourhoods U' of x and V' of y; there are only finitely many translates gU' that intersect V'; separate each of these gx from y by using the Hausdorff property of the space G/K, then reduce U' and V' accordingly.

Let Γx and Γy be two different points of $\Gamma \backslash G/K$. Then $x, y \in G/K$ are not on the same Γ-orbit and we can construct neighbourhoods U and V as above. If we let $p : G/K \to \Gamma \backslash G/K$ denote the canonical projection, then $p(U)$ and $p(V)$ are neighbourhoods of Γx and Γy that do not intersect. □

Property (4.2.4) is quite significant and, for this reason, an action satisfying this property receives the special name of a *properly discontinuous* actions:

Definition 4.2.9. The action of Γ on X is *properly discontinuous* when, for every compact set $K \subset X$, the number of elements $\gamma \in \Gamma$ for which $\gamma K \cap K \neq \emptyset$ is finite.

When the action is properly discontinuous, assuming X is a locally compact Hausdorff space, we conclude, just as we did in Corollary 4.2.8, that the quotient space $\Gamma \backslash X$ is Hausdorff.

Another common property that is often required from a group action is given by the next definition.

Definition 4.2.10. The action of a group G on a topological space X is said to be *wandering* (following Thurston in [52, Definition 8.2.4]) if it satisfies the following:

$$\text{Each point } x \in X \text{ has a neighbourhood } V \text{ such that } gV \cap V \neq \emptyset \text{ for only finitely many } g \in G. \tag{4.2.5}$$

Corollary 4.2.7 implies, in particular, that the action of a discrete subgroup Γ of G on the space G/K is wandering.

When the action of G on a Hausdorff space X is free and wandering, the projection map $\pi : X \to G\backslash X$ is a covering map. The quotient $G\backslash X$, however, is not necessarily Hausdorff (see Remark 4.2.11).

Together with the observation following Definition 4.2.9 we conclude that, if Γ acts freely and properly discontinuously on a locally compact Hausdorff space X, then $\Gamma\backslash X$ is Hausdorff and the natural projection $\pi : X \to \Gamma\backslash X$ is a covering map.

Remark 4.2.11. It is worth remarking that Proposition 4.2.6 and its corollaries are concerned with a particular type of spaces G/K, known as a *homogenous spaces*. In general, however, discreteness is not equivalent to proper discontinuity and the quotient need not be Hausdorff. A well-known counterexample is the action of $\mathbb{Z}$ on $X = \mathbb{R}^2 \setminus \{0\}$ given by $n \cdot (x, y) = (2^n x, 2^{-n} y)$. This action is easily seen to be free and wandering. Nevertheless, we observe that if U is any neighbourhood of $(0, 1)$ and V is any neighbourhood of $(1, 0)$, then $nU \cap V \neq \emptyset$ when n is sufficiently large. This implies that, in $\mathbb{Z}\backslash X$, the points $\pi((1, 0))$ and $\pi((0, 1))$ (which are clearly distinct) cannot be separated, i.e., they do not admit disjoint neighbourhoods. In particular, the quotient space $\mathbb{Z}\backslash X$ is not Hausdorff. Note that this action is not properly discontinuous (take, for example, $C =$ the line segment connecting $(0, 1)$ to $(1, 0)$).

Finally, in the case of a Fuchsian group Γ acting on $\mathbb{H}$, a Riemann surface structure passes on to the quotient, regardless of whether Γ acts freely or not. Note, however, that if Γ has torsion (which, in this case, is equivalent to not acting freely), then the projection map is not a covering map, but only a *branched* covering map. More precisely, we have the following:

Theorem 4.2.12. *Let Γ be a Fuchsian group. The quotient space $\Gamma\backslash\mathbb{H}$ admits a Riemann surface structure and, with this structure, the canonical projection $\pi : \mathbb{H} \to \Gamma\backslash\mathbb{H}$ is a holomorphic map.*

Proof. [26, Theorem 5.9.1]. □

4.3 The geometry of Fuchsian groups

4.3.1 Fundamental domains

In this section we describe fundamental domains for the action of a Fuchsian group on $\mathbb{H}$. As we shall see, a lot of geometrical information on these fundamental domains can be determined by the Fuchsian group, and vice versa.

Definition 4.3.1 (Fundamental domain)**.** Let Γ be a Fuchsian group acting on $\mathbb{H}$. A subset $F \subset \mathbb{H}$ is said to be a *fundamental domain* for Γ if

(i) The union of all translates fo F by elements of Γ cover the entire plane $\mathbb{H}$:
$$\bigcup_{\gamma\in\Gamma} \gamma F = \mathbb{H};$$

(ii) Any two translates of F by elements of Γ have disjoint interiors.

A fundamental domain always exists and, as we shall see below, it is clearly not unique.

Definition 4.3.2 (Dirichlet domain)**.** Let Γ be a Fuchsian group and let $p \in \mathbb{H}$ be a point that is not fixed by any element in Γ other than the identity. We define the *Dirichlet domain for* Γ *centered at* p to be the set
$$D_p(\Gamma) = \{z \in \mathbb{H} \mid d(z,p) \leq d(z,\gamma p), \text{ for all } \gamma \in \Gamma\}.$$

From the description of $D_p(\Gamma)$ we see that it is the intersection of certain hyperbolic half-planes (i.e. sets of all the points in $\mathbb{H}$ to one side of a geodesic line in $\mathbb{H}$) containing p so, in particular, $D_p(\Gamma)$ is connected and convex. Moreover, $D_p(\Gamma)$ is a fundamental domain for Γ.

Let $\mathrm{area}(\cdot)$ denote the hyperbolic area on $\mathbb{H}$. If D_1 and D_2 are two fundamental domains for Γ with topological boundary of hyperbolic area zero, then $\mathrm{area}(D_1) = \mathrm{area}(D_2)$. Define the *coarea* of Γ to be the area of a fundamental domain for Γ with boundary of area zero, which always exists (take a Dirichlet domain for instance). We observe here that the metric structure of $\mathbb{H}$ descends to the Riemann surface $\Gamma\backslash\mathbb{H}$ (see Theorem 4.2.12) making it a hyperbolic orbifold, whose area form we continue to denote by area. It follows that $\mathrm{area}(\Gamma\backslash\mathbb{H})$ equals the coarea of Γ.

Definition 4.3.3. We say Γ is *cofinite* or has *finite coarea* if $\mathrm{area}(\Gamma\backslash\mathbb{H}) < \infty$.

Now, $D_p(\Gamma)$ is bounded by geodesic segments called *sides* and, possibly, segments of the boundary at infinity, called *free sides*. When two sides intersect at a point in $\mathbb{H}$, this point is called an *ordinary vertex* of $D_p(\Gamma)$ or, sometimes, simply a *vertex*. Ordinary vertices are isolated as a consequence of discreteness of Γp and, in particular, $D_p(\Gamma)$ has at most countably many sides. The union of these sides and ordinary vertices constitutes the boundary of $D_p(\Gamma)$ in $\mathbb{H}$.

There is also the boundary of D_p at infinity, that is: take the topological boundary of $D_p(\Gamma)$ as a subset of the Riemann sphere Σ and then intersect it with $\partial\mathbb{H}$. A point w of the boundary of $D_p(\Gamma)$ at infinity falls into one of the following three types:

(i) Two sides of $D_p(\Gamma)$ meet at w, in which case w is called a *vertex at infinity* of $D_p(\Gamma)$;

(ii) Only one side of $D_p(\Gamma)$ meets w. This could happen for two reasons: either w is the endpoint of a free side, in which case it is called a *real vertex* of $D_p(\Gamma)$, or w is the accumulation point of an infinite sequence of sides of $D_p(\Gamma)$.

(iii) No side of $D_p(\Gamma)$ meets w. This, also, could happen for two reasons. The point w might be the interior point of a free side. Or, w could be the accumulation point of two sequences of sides of $D_p(\Gamma)$.

It is worth noting that, when a Fuchsian group Γ has finite coarea, boundary points of type (ii) and (iii) do not occur. Moreover, in this case, a Dirichlet domain for Γ has finitely many vertices and sides, i.e., it is a hyperbolic polygon.

Theorem 4.3.4. *If Γ is a Fuchsian group of finite coarea, then $D_p(\Gamma)$ has finitely many sides (and no free sides).*

Proof. See [34, Theorem I.5C]. □

Henceforth, we shall only be concerned with Fuchsian groups of finite coarea. Nevertheless, we point out that in the general case, although there may be infinitely many sides, the geometry of $D_p(\Gamma)$ is always well behaved at least locally:

Theorem 4.3.5 (Dirichlet domains are locally finite)**.** *A Dirichlet domain $D = D_p(\Gamma)$ for Γ is always* locally finite, *meaning that every $z \in D$ has a neighbourhood V such that $V \cap \gamma D \neq \emptyset$ for only finitely many elements $\gamma \in \Gamma$.*

Proof. See [26, Theorem 5.8.5]. □

In the classical example of $\mathbb{Z}^2$ acting on $\mathbb{R}^2$, it is intuitive that, instead of "folding up" the entire plane, one only needs to fold up the fundamental domain $[0,1] \times [0,1]$. In other words, it is sufficient to quotient the square $[0,1] \times [0,1]$. Similarly, in the case of a Fuchsian group Γ, we can also restrict to the quotient of a fundamental domain D, which we assume to be locally finite (for instance, a Dirichlet domain). Indeed, let π and Π be the projections respectively from D onto $\Gamma \backslash D$ and from $\mathbb{H}$ onto $\Gamma \backslash \mathbb{H}$. If $i : D \to \mathbb{H}$ denotes inclusion, we define $f : \Gamma \backslash D \to \Gamma \backslash \mathbb{H}$ as the map that makes the following diagram commute:

$$\begin{tikzcd}
D \arrow[r, "i"] \arrow[d, "\pi"'] & \mathbb{H} \arrow[d, "\Pi"] \\
\Gamma \backslash D \arrow[r, dashed, "f"'] & \Gamma \backslash \mathbb{H}
\end{tikzcd}$$

It is straightforward to check that f is bijective. Continuity of f follows from the fact that $\Pi \circ i$ is continuous and π is open. Moreover, we show that f is an open map. Let $\pi(z) \in U \subset \Gamma \backslash D$ where U is open. Since D is assumed to be locally finite, the preimage of $\pi(z)$ consists of finitely many points $z, \gamma_1 z, \ldots \gamma_k z$ and we can find a small neighbourhood V of z in $\mathbb{H}$ so that whenever a translate γV intersects D then γ must be one of the elements $\gamma_0, \gamma_1, \ldots, \gamma_k$, where γ_0 is the identity. In this way we see that the image $\Pi(V)$ is contained in $f(U)$: for $w \in V$, there exists $\gamma \in \Gamma$ such that $\gamma w \in D$, which means $\gamma V \cap D \neq \emptyset$ and γ is one of the γ_i. If the neighbourhood V is chosen sufficiently small, one can argue that $\gamma_i V \cap D \subset \pi^{-1}(U)$ for any $i = 0, \ldots, k$ and thus $\Pi(w) = \Pi(\gamma_i w) = f(\pi(\gamma_i w)) \in f(U)$, proving that $\Pi(V) \subset f(U)$. The fact that Π is an open map shows that $\Pi(V)$ is an open neighbourhood of $f(\pi(z))$. We have just proved that:

Theorem 4.3.6. *Let D be any locally finite fundamental domain for Γ (in particular, D could be a Dirichlet domain). Then $\Gamma \backslash D$ is homeomorphic to $\Gamma \backslash \mathbb{H}$.*

Corollary 4.3.7. *Let D be a locally finite fundamental domain for Γ. Then $\Gamma \backslash \mathbb{H}$ is compact if and only if D is compact. Moreover, if this is the case, then Γ contains no parabolic elements.*

Proof. Compactness of D immediately implies the compactness of the quotient. Conversely, using the fact that D is locally finite, (sequentially) compactness of D easily follows.

For the second part, let

$$\eta(z) = \inf\{d(z, \gamma z) \mid \gamma \in \Gamma, \gamma \text{ not elliptic}\}.$$

Then η is a continuous function of z and, since D is compact, η attains a positive minimum on D:

$$\eta(z) \geq r > 0, \quad \text{for all } z \in D.$$

As D is a fundamental domain for Γ, we claim that r is a lower bound for $\eta(z)$ when z varies in $\mathbb{H}$. Indeed, let z be any point in $\mathbb{H}$ and let $\tau \in \Gamma$ be such that $\tau z \in D$. For any $\gamma \in \Gamma \backslash \{\mathrm{Id}\}$ non-elliptic:

$$d(z, \gamma z) = d(\tau z, \tau\gamma z) = d(\tau z, \tau\gamma\tau^{-1}(\tau z)) \geq r > 0,$$

which then implies that $\eta(z) \geq r > 0$.

Now, if $\rho \in \Gamma$ is parabolic, then $d(z, \rho z)$ approaches 0 as z approaches the fixed point of ρ at infinity. Indeed, let ζ be a Möbius transformation such that $\zeta\rho\zeta^{-1}$: $w \mapsto w \pm 1$. Then $d(z, \rho z) = d(w, \zeta\rho\zeta^{-1}(w))$ where $w = \zeta(z)$ and the right-hand side is bounded above by $\frac{1}{\operatorname{Im} w}$, which tends to 0 as $w \to \infty$. □

Remark 4.3.8. The converse of the second part is also true, i.e., if Γ is non-cocompact then Γ contains parabolic elements (see, for instance, [28, Theorem 4.2.5]).

Let s be a side of a Dirichlet region D for Γ. If $\gamma \in \Gamma$ is such that $\gamma(s)$ is a side of D, then s and $\gamma(s)$ are said to be *conjugate* sides of D. Note that if $\gamma'(s)$ is also conjugate to s, then $\gamma' = \gamma$. If a side is conjugate to itself, then the two halves of this side are interchanged and, in this case, we consider the mid point to be a vertex with internal angle π. Thus, the sides of the Dirichlet domain fall into conjugate pairs and we remark that Γ is generated by these side-pairing transformations:

Proposition 4.3.9. *Let D be a Dirichlet domain for Γ and let $\{\gamma_n\}$ be the subset of Γ consisting of all transformations that pair two sides of D. Then $\{\gamma_n\}$ generates Γ.*

Proof. Let $\Lambda < \Gamma$ denote the subset generated by $\{\gamma_n\}$. We have that

$$\mathbb{H} = \bigcup_{\lambda \in \Lambda} \lambda D \;\cup \bigcup_{\gamma \in \Gamma\backslash\Lambda} \gamma D.$$

It follows from local finiteness that any union of translates of D is closed. Clearly, $\bigcup_{\lambda\in\Lambda} \lambda D$ is non-empty, so if we can prove that the two sets on the right-hand side of 4.3.1 are disjoint, it will follow from connectedness of $\mathbb{H}$ that $\Gamma\backslash\Lambda = \emptyset$.

Suppose γD intersects some λD for $\lambda \in \Lambda$. If their interiors intersect then $\gamma = \lambda$, so assume they only intersect on their boundary. They can either share a common side or a common vertex. Suppose first that they share a common side γs, where s is a side of D. Let s' be the side of D that is paired with s by the transformation γ_i. Then $\gamma\gamma_i\gamma^{-1}$ pairs $\gamma s'$ with γs. It follows that $\gamma\gamma_i\gamma^{-1}(\gamma D)$ and λD have intersecting interiors so, in particular, $\gamma\gamma_i = \lambda$ and $\gamma \in \Lambda$. Now, let it be the case that γD share a vertex γv with λD, where v is a vertex of D. We claim that there are finitely many side-pairing transformations $\gamma_{i_1}, \dots, \gamma_{i_N}$ such that $\zeta = \gamma_{i_N} \cdots \gamma_{i_1}$ fixes v (this will become clearer in the next paragraph. See also §§4.3.3). Due to discreteness of Γ, the transformation ζ must have finite order, so, for some k, the interiors of $\gamma\zeta^k\gamma^{-1}(\gamma D)$ and λD intersect. It then follows that $\gamma = \lambda\zeta^{-k} \in \Lambda$. □

The vertices of a Dirichlet domain can also fall into the same Γ-orbit. One subset of vertices belonging to the same Γ-orbit is called a *cycle*. Note that each cycle must be finite, since D is locally finite. The stabilisers of the vertices in a cycle are all conjugate to each other and, in particular, have the same order (recall that the stabilisers are finite cyclic groups due to discreteness of Γ). If one of these vertices has non-trivial stabiliser, i.e., if one of these vertices is the fixed point of an elliptic element of Γ, then its cycle is called an *elliptic cycle*. Note that the stabiliser of one such vertex is a maximal finite cyclic group, since any element in its centraliser must also fix that vertex. Conversely, a finite cyclic subgroup S of Γ must fix some point in $\mathbb{H}$ (for instance, the centre of mass of a finite S-orbit), and thus it corresponds to an elliptic vertex of D. The conjugacy class of S in Γ is then associated to an elliptic cycle of D. We have established the following:

Proposition 4.3.10. *The elliptic cycles of D are in one-to-one correspondence with conjugacy classes of non-trivial maximal finite cyclic subgroups of* Γ.

Since the translates of a Dirichlet domain form a tesselation of $\mathbb{H}$, the angles of all the translates meeting at the same vertex must sum 2π. The next proposition is then easily verified:

Proposition 4.3.11. *Let $\{v_1, \dots, v_r\}$ be an elliptic cycle of D such that the vertex v_i has internal angle θ_i. Let m be the order of the stabiliser of each v_i. Then $\theta_1 + \cdots + \theta_r = 2\pi/m$.*

Likewise, any vertex at infinity of D is a fixed point of a parabolic element in Γ (see [28, Theorem 4.2.5 (i)]) and we can partition these vertices into congruence cycles called *parabolic cycles*. Each of these cycles contain finitely many vertices (Theorem 4.3.4). There is also a correspondence between parabolic cycles and conjugacy classes of parabolic cyclic subgroups (cyclic subgroups constituted of parabolic elements):

Proposition 4.3.12. *The parabolic cycles of D are in one-to-one correspondence with conjugacy classes of maximal parabolic cyclic subgroups of* Γ.

Let Γ be cofinite and let D be a Dirichlet domain for Γ. Theorem 4.3.4 says that D is a hyperbolic polygon, so there are finitely many vertices (including the ones at infinity). It follows from Propositions 4.3.10 and 4.3.12 that the number of conjugacy classes of maximal finite cyclic subgroups and the number of conjugacy classes of maximal parabolic cyclic subgroups are both finite. The following definition conveys all this information into a string of integers.

Definition 4.3.13. Suppose there are r conjugacy classes of maximal finite cyclic subgroups of Γ. Let $m_1, \dots, m_r$ be the order of the subgroups in each of these conjugacy classes. Let s be the number of conjugacy classes of maximal parabolic cyclic subgroups of Γ. If g is the genus of the surface $\Gamma\backslash\mathbb{H}$, we say that Γ has *signature* $(g; m_1, \dots, m_r; s)$.

4.3.2 The hyperbolic area of a fundamental domain

Recall that the total area of $\Gamma\backslash\mathbb{H}$ equals the hyperbolic area of a Dirichlet domain D for Γ, or of any other fundamental domain (as long as it has boundary of hyperbolic area zero). In this context, Γ could be a Fuchsian group as well as a discrete subgroup of $\mathrm{Isom}(\mathbb{H})$.

Proposition 4.3.14. *Let $\Gamma < \mathrm{Isom}(\mathbb{H})$ be a discrete group of isometries and let $\Lambda < \Gamma$ be a finite index subgroup. Suppose*

$$\Gamma = \Lambda\gamma_1 \cup \cdots \cup \Lambda\gamma_n$$

is a decomposition of Γ into Λ right cosets. If D is a fundamental domain for Γ (with boundary of hyperbolic area zero), then

$$D' = \gamma_1 D \cup \cdots \cup \gamma_n D$$

is a fundamental domain for Λ.

Proof. It is easy to see that $\Lambda D' = \mathbb{H}$. Now, following [26], suppose z and $\gamma(z)$ are two points in the interior of D'. Let $\epsilon > 0$ be such that a hyperbolic open ball of radius ϵ around z, as well as the one around $\gamma(z)$, are both contained in D'. If we denote the former by B, then the latter is γB. Let $1 \leq i_1 < \cdots < i_k \leq n$ be the indices of all translates $\gamma_i \mathrm{int}(D)$ intersecting B, where $\mathrm{int}(D)$ denotes the interior of the set D. The ball γB must intersect some $\gamma_j \mathrm{int}(D)$, which means B intersects $\gamma^{-1}\gamma_j \mathrm{int}(D)$ and so $\gamma_j = \gamma\gamma_{i_l}$ for some i_l. It follows that $\Lambda\gamma_j = \Lambda\gamma\gamma_{i_l} = \Lambda\gamma_{i_l}$, thus $\gamma_j = \gamma_{i_l}$ and $\gamma = \mathrm{Id}$, proving that D' is indeed a fundamental domain for Λ. □

Corollary 4.3.15 (Riemann-Hurwitz)**.** *Let Γ be a discrete subgroup of $\mathrm{Isom}(\mathbb{H})$. If $\Lambda < \Gamma$ is of finite index, then*

$$\mathrm{area}(\Lambda\backslash\mathbb{H}) = [\Gamma : \Lambda] \cdot \mathrm{area}(\Gamma\backslash\mathbb{H}).$$

Remark 4.3.16. The Riemann-Hurwitz Theorem is a more general statement accounting for branching points in the covering map. This version, however, will be sufficient for our use.

The information contained in the signature of a Fuchsian group Γ is sufficient for us to calculate the area of $\Gamma\backslash\mathbb{H}$. Let D be a Dirichlet domain for Γ of signature $(g; m_1, \dots, m_r; s)$. The analysis prior to Definition 4.3.13 shows that D has r elliptic cycles, s parabolic cycles and, possibly, r' cycles made of vertices with trivial stabiliser. Suppose D has n pairs of sides. Remember that the action of Γ pairs these sides together. Being a hyperbolic polygon with $2n$ sides, the area of D is given by the Gauß-Bonnet Theorem as:

$$\text{area}(D) = (2n-2)\pi - \sum \alpha_i,$$

where $\sum \alpha_i$ is the sum of all the internal angles of D. This sum can be calculated using Proposition 4.3.11. Indeed, the sum of all the angles at elliptic vertices amounts to $2\pi \sum_{i=1}^{r}(1/m_i)$. Similarly, the sum of other vertices with trivial stabiliser gives $2\pi r'$. Finally, the vertices at infinity have internal angles 0. By putting all this together we obtain that:

$$\text{area}(D) = (2n-2)\pi - 2\pi\left(\sum_{i=1}^{r}\frac{1}{m_i} + r'\right), \tag{4.3.1}$$

Now, we want to express n in terms of the information given in the signature of Γ, which includes the genus g of $\Gamma\backslash\mathbb{H}$. This indicates that we should consider the situation from a topological perspective. Recall that $\Gamma\backslash\mathbb{H}$ is homeomorphic to $\Gamma\backslash D$. The n pairs of sides of D project to $\Gamma\backslash D$ as n edges. The elliptic cycles and the cycles of vertices with trivial stabiliser project to $r + r'$ vertices of $\Gamma\backslash D$. The vertices at infinity "project" to s punctures. The sides of D going to infinity project to edges with one end at these punctures. So we add s points to $\Gamma\backslash D$ in order to obtain a CW structure with $r + r' + s$ vertices, n edges and 1 face on a compact orientable surface of genus g. The Euler formula then gives us:

$$2 - 2g = r + r' + s - n + 1.$$

Substituting in (4.3.1) leads to:

$$\begin{aligned}\text{area}(D) &= (4g - 4 + 2r + 2r' + 2s)\pi - 2\pi\left(\sum_{i=1}^{r}\frac{1}{m_i} + r'\right)\\ &= 2\pi\left[2g - 2 + \sum_{i=1}^{r}\left(1 - \frac{1}{m_i}\right) + s\right].\end{aligned}$$

We have just proved the following:

Theorem 4.3.17. *Let Γ be a Fuchsian group of signature $(g; m_1, \dots, m_r; s)$. Then*

$$\operatorname{area}(\Gamma\backslash\mathbb{H}) = 2\pi \left[2g - 2 + \sum_{i=1}^{r} \left(1 - \frac{1}{m_i}\right) + s\right]. \tag{4.3.2}$$

4.3.3 Poincaré's Theorem and presentation of Fuchsian groups

We have stated several properties satisfied by a Dirichlet domain of a Fuchsian group Γ. Poincaré's Theorem goes in the opposite direction: starting from a polygon P, Poincaré established sufficient conditions for the group generated by the side-pairing transformations of P to be discrete.

Let us restrict ourselves to finite sided polygons, although the theorem is true for more general polygons.

Suppose P satisfies the following conditions:

1. An *identification* on P is given, i.e., a function associating to each side s of P another side s' and an isometry $\phi(s, s')$ of the hyperbolic plane so that:

   (a) $\phi(s, s')$ maps s onto s' and takes the exterior of the circle containing s (the half plane containing the interior of P) to the interior of the circle containing s';

   (b) $(s')' = s$ and $\phi(s', s) = \phi(s, s')^{-1}$;

   (c) if $s = s'$ then $\phi(s, s')$ is the reflection across the line containing s. In particular, $\phi(s, s')$ then satisfies a *reflection relation*:

   $$\phi(s, s')^2 = \mathrm{Id}; \tag{4.3.3}$$

2. For each *cycle of vertices* $\{v_1, \dots, v_r\}$ in $\mathbb{H}$ there exists an integer m such that the angles subtended at these vertices add up to $2\pi/m$;

3. For each cycle of vertices at infinity, the *cycle transformation* is parabolic.

Some of the terminology used requires further explanation. Start from a vertex v_1 with s_1 being one of the sides of P that contain v_1. The side s_1 is mapped onto s_1' by the isometry $A_1 = \phi(s_1, s_1')$, taking the vertex v_1 to $v_2 = A_1(v_1)$. Let s_2 be the other side of P that meets s_1' at v_2, then set $v_3 = A_2(v_2)$, where $A_2 = \phi(s_2, s_2')$. Continuing with this process leads to a sequence of vertices $\{v_1, v_2, \dots\}$, a sequence of isometries $\{A_1, A_2, \dots\}$ and, finally, a sequence of pairs of sides $\{(s_1, s_1'), (s_2, s_2'), \dots\}$. Since we are assuming there are only finitely many sides (hence finitely many vertices), these sequences are all periodic. Let r be the least integer such that these three sequences are periodic with period r. Then

$\{v_1, \ldots, v_r\}$ is a *cycle of vertices* and $A_r \cdots A_1$ is the *cycle transformation*. Notice from this construction that the cycle transformation fixes v_1. It thus follows from condition (2) that $A_n \cdots A_1$ is an elliptic transformation of order m (or the identity in case $m = 1$), so it satisfies the following *cycle relation*:

$$(A_n \cdots A_1)^m = \mathrm{Id}. \tag{4.3.4}$$

For the case of vertices at infinity, the same construction may be carried out. In this way we obtain a cycle of vertices at infinity and a cycle transformation that condition (3) requires to be parabolic.

In the setting of Proposition 4.3.9, the side-pairing transformations of a fundamental polygon for Γ constituted a generating set. Here, we have a polygon P whose side-pairing transformations will generate a group of isometries that has P for a fundamental domain. The point of Poincaré's Theorem is precisely ensuring that the group generated is discontinuous (and hence discrete):

Theorem 4.3.18 (Poincaré's Theorem)**.** *Let P be a polygon satisfying conditions* (1)-(3). *Then the group of isometries Γ generated by the side-pairing transformations is discontinuous and P is a fundamental domain for Γ. Moreover, the reflection relations* (4.3.3) *and the cycle relations* (4.3.4) *form a complete set of relations of Γ.*

Proof. See [38]. □

A direct application of Poincaré's Theorem guarantees the existence of a Fuchsian group of signature $(g; m_1, \ldots, m_r; s)$ as long as the right-hand side of (4.3.2) is positive. Sometimes, this result is also referred to as Poincaré's Theorem:

Theorem 4.3.19. *Let $g, r, s \geq 0$ and $m_1, \ldots, m_r \geq 1$ be integers such that*

$$2\pi \left[2g - 2 + \sum_{i=1}^{r} \left(1 - \frac{1}{m_i}\right) + s \right] > 0, \tag{4.3.5}$$

then there exists a Fuchsian group of signature $(g; m_1, \ldots, m_r; s)$.

Sketch of proof. For the sake of this proof consider the disc model $\mathbb{D}$. Divide $\mathbb{D}$ into $N = 4g + r + s$ circular sections of equal angles.

For $t > 0$, mark, on each radius, the point $v_i(t)$ whose hyperbolic distance to the origin is t. Connect each $v_i(t)$ to $v_{i+1}(t)$ (where, of course, $v_N(t) = v_1$) with a geodesic arc and form, in this way, a geodesic polygon with $N = 4g + r + s$ sides. Now, for each $j = 1, \ldots, r$, pick the unique point $w_j(t)$ on the exterior of this polygon such that $v_j(t)$, $v_{j+1}(t)$ and $w_j(t)$ form an isosceles triangle with base on the side of the polygon and with angle $2\pi/m_j$ at $w_j(t)$. Label the two congruent sides of each triangle d_j and d'_j. For $j = r + 1, \ldots, s$, do the same thing by picking $w_j(t)$ on the border at infinity, an endpoint of the geodesic line bisecting the corresponding side of the original polygon, so that, in this case, the angle subtended by the sides d_j and d'_j is zero. Label the remaining $4g$ sides of the original polygon $a_1, b_1, a'_1, b'_1, \ldots, a_g, b_g, a'_g, b'_g$. The sides $d_1, d'_1, \ldots, d_{r+s}, d'_{r+s}$ together with the a_i, b_i, a'_i, b'_i form a geodesic polygon we shall call $M(t)$, as in Figure 4.1.

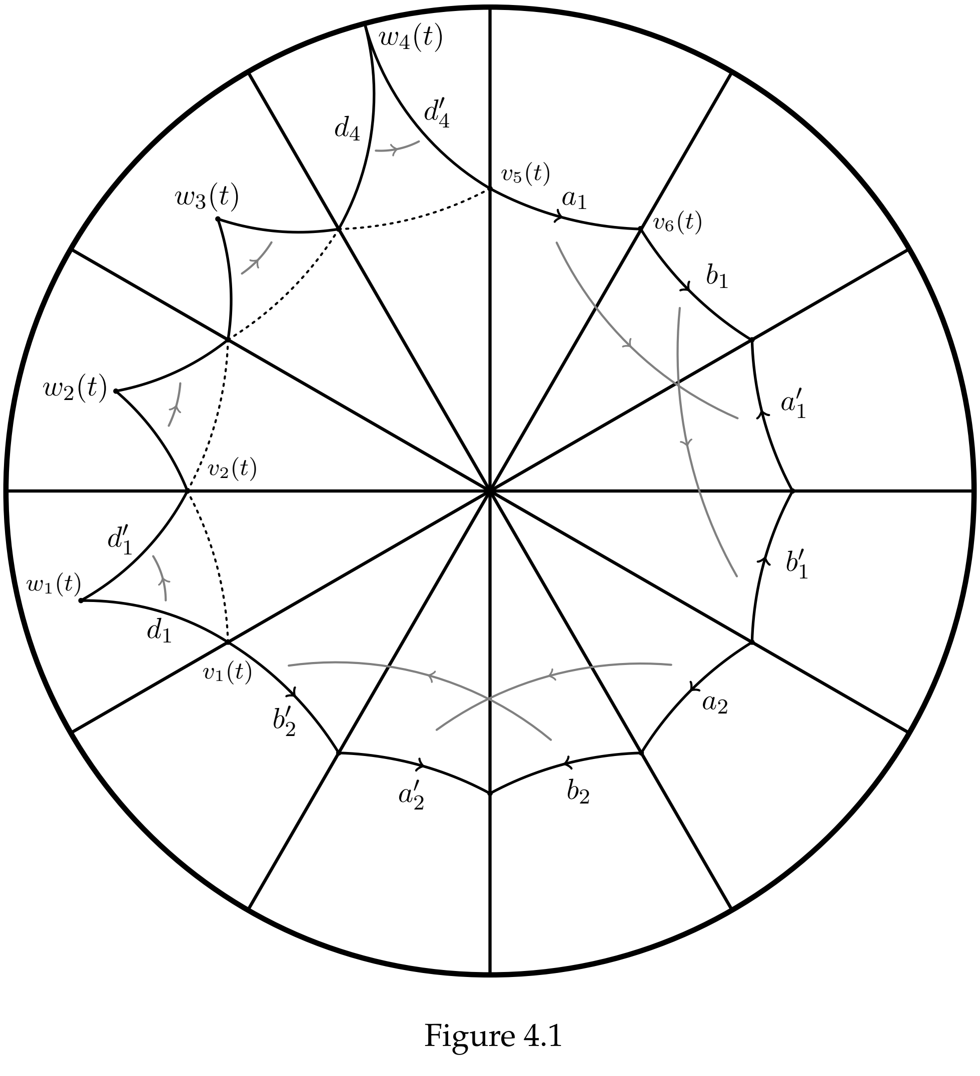

Figure 4.1

We pair the sides a_i with a'_i, b_i with b'_i and d_j with d'_j through the unique isometries of $\mathbb{D}$ taking each of these sides to its correspondent while mapping the exterior of the circle supporting one side to the interior of the circle supporting the corresponding side. In this way, almost all the conditions in Poincaré's Theorem are satisfied, except possibly one: the vertices $v_1(t), v_2(t), \ldots, v_N(t)$ form a cycle and we need the sum of the internal angles of $M(t)$ at these vertices to be a submultiple of 2π. In fact, in order to obtain the desired signature, we need this sum to be precisely 2π. In other words, we want this cycle to be accidental. This can be achieved by varying the parameter t continuously. Indeed, let $\mu(t)$ denote the hyperbolic area of $M(t)$. When $t \to 0$, $\mu(t)$ also approaches zero. On the other hand, when $t \to \infty$, by the Gauß-Bonnet formula, we see that $\mu(t)$ approaches

$$(4g+2r+2s) - \sum_{i=1}^{r} \frac{2\pi}{m_i} = 2\pi \left[2g - 1 + \sum_{i=1}^{r} \left(1 - \frac{1}{m_i} \right) + s \right]. \tag{4.3.6}$$

Since $\mu(t)$ varies continuously with t, there exists $t_0 > 0$ for which $\mu(t_0)$ is precisely the area that we expect a group of such signature to have:

$$\mu(t_0) = 2\pi \left[2g - 2 + \sum_{i=1}^{r} \left(1 - \frac{1}{m_i} \right) + s \right].$$

Once again, by the Gauß-Bonnet formula, the area $\mu(t_0)$ must be equal to the right-hand side of (4.3.6) minus the sum of the angles at the vertices of the cycle $v_1(t_0), v_2(t_0), \ldots, v_N(t_0)$. This immediately implies that, for $t = t_0$, this sum is 2π and the cycle in question is accidental as we wanted.

Now we can apply Poincaré's Theorem 4.3.18 and conclude that the group Γ generated by the identifications of the sides of $M(t_0)$ has signature $(h; m_1, \ldots, m_r; s)$, where h is the genus of the underlying topological surfaces of the quotient $\Gamma \backslash \mathbb{D}$. Just as in the discussion preceding Theorem 4.3.17, we see that $\Gamma \backslash \mathbb{D}$ admits a CW structure with $r+s+1$ vertices, $2g+r+s$ edges and 1 face, so the formula for its Euler characteristic yields

$$2 - 2h = (r+s+1) - (2g+r+s) + 1 = 2 - 2g.$$

We conclude that $h = g$ and thus Γ has the desired signature. □

Note that one can also describe a complete set of relations for the group Γ constructed in the proof above. Using the same notation, let A_i and B_i, $i = 1, \ldots, g$, be the hyperbolic transformations pairing a_i to a_i' and b_i to b_i', respectively. For $j = 1, \ldots, r$, let C_j be the elliptic transformation with fixed point $w_j(t_0)$, pairing the sides d_j and d_j'. Finally, for $k = 1, \ldots, s$, let P_k be the parabolic transformation with fixed point $w_{r+k}(t_0)$, pairing the sides d_{r+k} and d_{r+k}'. By going around each elliptic vertex one obtains the relations

$$C_j^{m_j} = \mathrm{Id}, \quad j = 1, \ldots, r.$$

By going around the unique accidental cycle gives that

$$\prod_{i=1}^{g} [A_i, B_i] \cdot C_1 \cdots C_r P_1 \cdots P_s = \mathrm{Id}.$$

It follows from Poincaré's Theorem 4.3.18 that these are all the relations we need, and thus Γ has presentation

$$\Big\langle A_1, B_1, \dots, A_g, B_g, C_1, \dots, C_r, P_1, \dots, P_s \;\Big|\; C_1^{m_1} = \cdots = C_r^{m_r} = \prod_{i=1}^{g} [A_i, B_i]\, C_1 \cdots C_r\, P_1 \cdots P_s = 1 \Big\rangle. \quad (4.3.7)$$

We were only able to get this presentation by reading off the relations from the very specific fundamental polygon that we constructed. We give this kind of polygon a special name:

Definition 4.3.20. Let Γ be a Fuchsian group of signature $(g; m_1, \dots, m_r; s)$. A fundamental polygon P for the action of Γ is said to be a *canonical polygon* if it is a polygon with $4g + 2(r + s)$ analytic sides in the following order:

$$a_1, b_1, a_1', b_1', \dots, a_g, b_g, a_g', b_g', d_1, d_1', \dots d_{r+s}, d_{r+s}',$$

where each side is paired to the corresponding primed side.

The presentation (4.3.7) is called the *standard presentation* of Γ.

Note that, in this definition, the canonical polygon is *not* required to be geodesic, i.e., to have geodesic arcs for sides, but merely to be bound by analytic arcs.

It is not hard to prove that, for a generic choice of p, the Dirichlet domain $D_p(\Gamma)$ only has elliptic and parabolic cycles of length 1 (see [2, Theorem 9.4.5]). Nonetheless, it is still necessary to have the vertices and sides ordered as in the canonical polygon in order to obtain a standard presentation.

Suppose P is a fundamental polygon for Γ whose sides can be traveled in the the sequence $WaXYa'Z$. Here, W, X, Y and Z represent blocks of letters (sides) and (a, a') is a pair of conjugate sides paired by the transformation T. If h is an analytic arc, connecting two vertices of P and splitting it into two regions P_1 and P_2 such that a is in P_1 and a' is in P_2 (see Figure 4.2), then $P_2 \cup T(P_1)$ is still a fundamental domain for Γ. Also, we note that this process preserves the total sum of internal angles as well as each angle subtended by two conjugate sides meeting at a fixed points. This cutting and pasting process is called an *admissible modification*. It is proved in [33, Chapter VII §4] that, after a finite number of admissible modifications, one obtains a fundamental domain for Γ with the desired order of sides of a canonical polygon, thus *every cofinite Fuchsian group admits a canonical polygon*. Moreover, it is known that every cofinite Fuchsian group admits a convex geodesic canonical polygon. For this result, we reference the reader to the foundational treatise by Fricke and Klein (pp. 240-260 of the English translation [20]). Also, cf. [30].

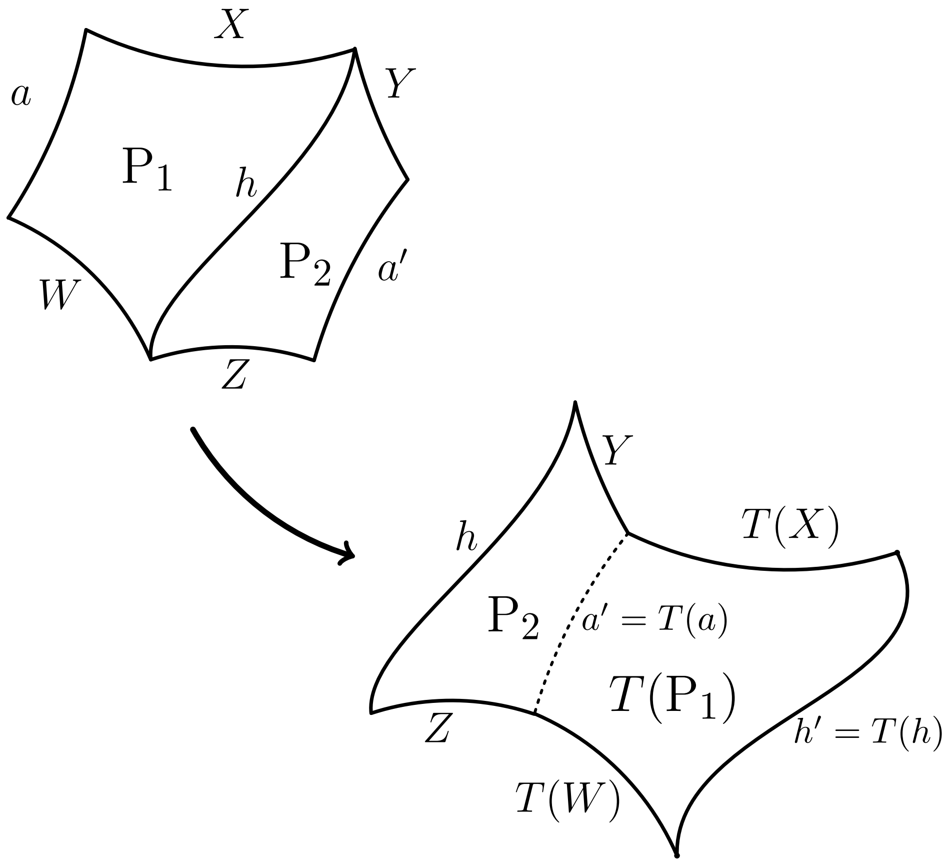

Figure 4.2: Example of an admissible modification.

In particular, it follows that every cofinite Fuchsian group admits a standard presentation as in (4.3.7). More precisely,

Theorem 4.3.21. *Suppose the integers $g, r, m_1, \ldots, m_r, s$ satisfy* (4.3.5) *and let G be the abstract group*

$$\left\langle \begin{array}{c} a_1, b_1, \ldots, a_g, b_g, c_1, \ldots, c_r, \\ p_1, \ldots, p_s \end{array} \middle| \; c_1^{m_1} = \cdots = c_r^{m_r} = \prod_{i=1}^{g} [a_i, b_i] \cdot c_1 \cdots c_r \, p_1 \cdots p_s = 1 \right\rangle .$$

Then any Fuchsian group Γ of signature $(g; m_1, \ldots, m_r; s)$ contains hyperbolic transformations $A_1, B_1, \ldots, A_g, B_g$, elliptic transformations $C_1, \ldots, C_r$ of order, respectively, $m_1, \ldots, m_r$ and parabolic transformations $P_1, \ldots, P_s$ such that the homomorphism ϕ : $G \to \Gamma$ mapping $a_i \mapsto A_i$, $b_i \mapsto B_i$, $c_i \mapsto C_i$ and $p_i \mapsto P_i$ is an isomorphism.

Example 4.3.22 (Reflection groups)**.** Let P be a hyperbolic polygon with n sides whose internal angles are submultiples of π. Let us label the sides of P cyclically as $s_1, \ldots, s_n$ and denote by σ_i the hyperbolic reflection across s_i. It follows from Poincaré's Theorem that the group $\widetilde{\Gamma}$ of hyperbolic isometries of $\mathbb{H}$, generated by $\sigma_1, \ldots, \sigma_n$, is a discrete subgroup of $\mathrm{Isom}(\mathbb{H})$ and P is a fundamental domain for the action of $\widetilde{\Gamma}$. Groups obtained in this way are called *reflection groups*.

Note however that $\widetilde{\Gamma}$ is not a subgroup of $\mathrm{PSL}(2,\mathbb{R})$ as it contains among its elements orientation-reversing isometries (for instance, the hyperbolic reflections generating it). We may, however, consider the subgroup Γ of orientation-preserving isometries in $\widetilde{\Gamma}$:

$$\Gamma = \widetilde{\Gamma} \cap \mathrm{PSL}(2,\mathbb{R})$$

This group is called a *Fuchsian group generated by reflections*. We will sometimes refer to it as the Fuchsian group *generated by P*. Despite this terminology, we emphasise that Γ does not contain any reflection, by definition.

It is immediate that Γ consists of all elements of $\widetilde{\Gamma}$ that can be expressed as the product of an even number of generators $\sigma_1, \ldots, \sigma_n$. One may write $\widetilde{\Gamma} = \Gamma \sqcup \Gamma\sigma_1$, where $\sqcup$ indicates disjoint union. In particular, $[\widetilde{\Gamma} : \Gamma] = 2$. Moreover, it follows from Proposition 4.3.14 that $P \cup \sigma_1 P$ is a fundamental domain for Γ.

While Poincaré's Theorem gives a presentation for the group generated by the side-pairing isometries of $P \cup \sigma_1 P$, Theorem 4.3.9 tells us that this group is precisely Γ. We are therefore able to produce a presentation for Γ as follows. The side-pairing isometries of $P \cup \sigma_1 P$ are $\sigma_1\sigma_2$ taking s_2 to s_2', $\sigma_1\sigma_3$ taking s_3 to s_3', and so on and so forth, until $\sigma_1\sigma_n$, taking s_n to s_n' (see Figure 4.3). Let the internal angle of P at the vertex v_i be π/m_i. We see that the vertex v_1 of $P \cup \sigma_1 P$ is the fixed point of an elliptic transformation rotating $2\pi/m_1$ and so $\sigma_1\sigma_2$ satisfies the relation $(\sigma_1\sigma_2)^{m_1} = \mathrm{Id}$. Likewise, $(\sigma_1\sigma_n)^{m_n} = \mathrm{Id}$. The other vertices are part of an elliptic cycle of length 2, so each of these cycles contributes with a cycle relation

$$[(\sigma_1\sigma_i)^{-1}(\sigma_1\sigma_{i+1})]^{m_i} = \mathrm{Id}, \quad i = 2, \ldots, n-1, \tag{4.3.8}$$

which comes from: starting at v_i with incident side s_{i+1}; then s_{i+1} is mapped to s_{i+1}' by $\sigma_1\sigma_{i+1}$, taking v_i to v_i'; the other side incident to v_i' is s_i' and this side is mapped to s_i by $(\sigma_1\sigma_i)^{-1}$, which takes v_i' back to v_i, thus completing the cycle.

This is sufficient to give a presentation of Γ. We will, however, rewrite this presentation in a more familiar form. Observe that

$$\sigma_1\sigma_{i+1} = (\sigma_1\sigma_i)(\sigma_i\sigma_{i+1}) \text{ and so } \sigma_i\sigma_{i+1} = (\sigma_1\sigma_i)^{-1}(\sigma_1\sigma_{i+1}),$$

If we let $C_i = \sigma_i\sigma_{i+1}$ for each $i = 1, \ldots, n-1$, each relation (4.3.8) can be expressed as $C_i^{m_i} = \mathrm{Id}$. Indeed, C_i is the elliptic element fixing the vertex v_i and rotating $2\pi/m_i$ (from s_{i+1} towards s_i). Also, for $i = 1$, $C_1^{m_1} = \mathrm{Id}$. It remains to rewrite the relation $(\sigma_1\sigma_n)^{m_n} = \mathrm{Id}$. Notice that

$$\sigma_1\sigma_n = (\sigma_1\sigma_2)(\sigma_2\sigma_3)\cdots(\sigma_{n-1}\sigma_n) = C_1C_2\cdots C_{n-1},$$

giving that $(C_1C_2\cdots C_{n-1})^{m_n} = \mathrm{Id}$. Alternatively, add the generator $C_n = \sigma_n\sigma_1$ that clearly satisfies $C_1C_2\cdots C_n = \mathrm{Id}$.

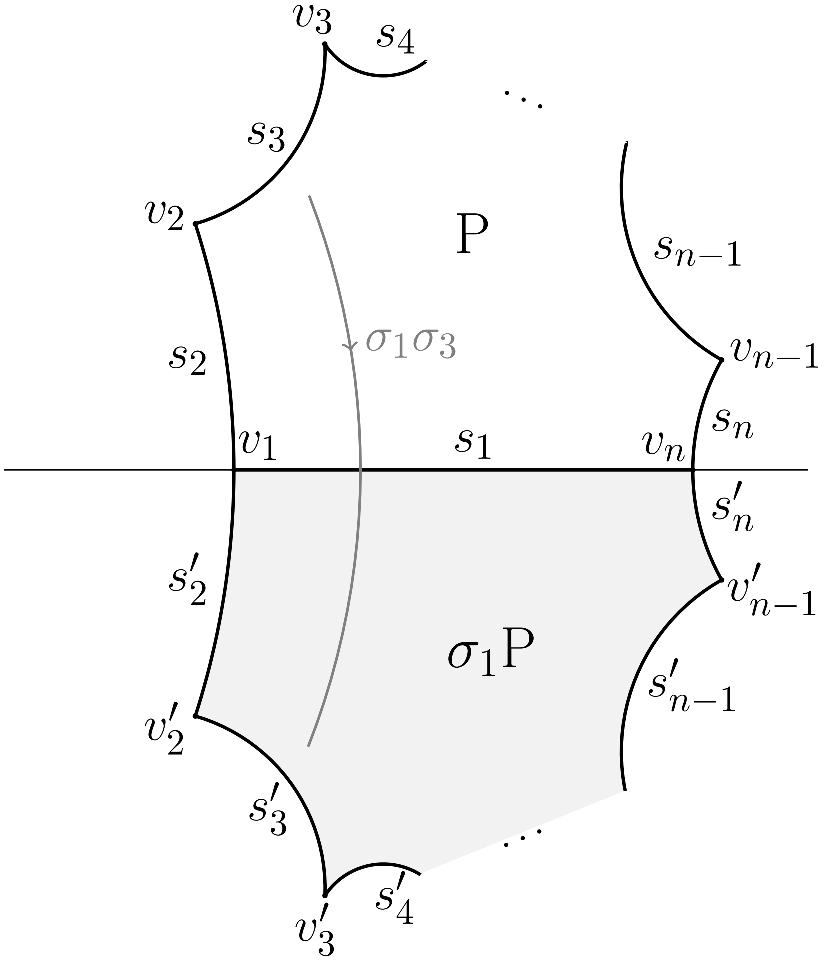

Figure 4.3

We therefore obtain the following presentation for the Fuchsian group Γ generated by P:

$$\left\langle C_1, \dots, C_n \mid C_1^{m_1} = \dots = C_n^{m_n} = C_1 \cdots C_r = 1 \right\rangle .$$

4.3.4 The Teichmüller space of a Fuchsian group

In this subsection we assume Γ to be cocompact of signature $(g; m_1, \dots, m_r)$.

By a *representation* of Γ into $\mathrm{PSL}(2,\mathbb{R})$ we mean a group homomorphisms $\phi : \Gamma \to \mathrm{PSL}(2,\mathbb{R})$. Note that ϕ is determined by the image of any set generating Γ. Moreover, the representations of Γ into $\mathrm{PSL}(2,\mathbb{R})$ are in one-to-one correspondence with the elements $(A_1, B_1, \dots, A_g, B_g, C_1, \dots, C_r)$ of $\mathrm{PSL}(2,\mathbb{R})^{2g+r}$ whose coordinates satisfy the relations

$$C_1^{m_1} = \dots = C_r^{m_r} = \prod_{i=1}^{g} [A_i, B_i]\, C_1 \cdots C_r = \mathrm{Id}.$$

Let $\mathfrak{R}'(\Gamma)$ denote the set of all representations of Γ into $\mathrm{PSL}(2,\mathbb{R})$. Given the above correspondence, we will identify $\mathfrak{R}'(\Gamma)$ with a closed subset of $\mathrm{PSL}(2,\mathbb{R})^{2g+r}$. This identification induces a natural topology on $\mathfrak{R}'(\Gamma)$. Let $\mathfrak{R}(\Gamma)$ be the subset of $\mathfrak{R}'(\Gamma)$ consisting of injective representations ϕ such that $\phi(\Gamma)$ is a discrete cocompact subgroup of $\mathrm{PSL}(2,\mathbb{R})$.

A. Weil proved that $\mathfrak{R}(\Gamma)$ is open in $\mathfrak{R}'(\Gamma)$. In fact, Weil's result holds, more generally, for cocompact lattices in connected Lie groups. This result was later extended to the non-cocompact case by Garland and Raghunathan. For our purposes, it is enough to know that:

Theorem 4.3.23 (Weil, [53]). *Let $\phi_0 : \Gamma \to \mathrm{PSL}(2,\mathbb{R})$ be an injective representation such that $\phi_0(\Gamma)$ is discrete and cocompact. Then any representation ϕ sufficiently close to ϕ_0 is also injective with image $\phi(\Gamma)$ discrete and cocompact.*

Note that $\mathrm{PGL}(2,\mathbb{R})$, the group of all conformal and anti-conformal homeomorphisms of the upper half-plane $\mathbb{H}$, acts on $\mathfrak{R}(\Gamma)$ by conjugation. We make the following definition:

Definition 4.3.24. The *Teichmüller space* of Γ, $\mathrm{Teich}(\Gamma)$, is defined to be the quotient of $\mathfrak{R}(\Gamma)$ by the action of $\mathrm{PGL}(2,\mathbb{R})$ by conjugation.

Remark 4.3.25. Let S_g be a closed orientable surface of genus g. The Teichmüller space $\mathrm{T}(S_g)$ of S_g may be defined as the set of all hyperbolic metrics on S_g up to isometries isotopic to the identity. Alternatively, one may consider all pairs formed by a genus g Riemann surface together with a *marking*, i.e, a choice of homotopy classes for the canonical generators of its fundamental group. Two such pairs are said to be equivalent when there exists a biholomorphism between the surfaces that respects the corresponding markings. The space of equivalence classes is called the *Teichmüller space of genus g*, and is denoted by T_g. We observe that the spaces $\mathrm{T}(S_g)$ and T_g may be identified. Moreover, when $S_g = \Gamma\backslash\mathbb{H}$, there exits a natural correspondence between $\mathrm{Teich}(\Gamma)$ and $\mathrm{T}(S_g) \cong \mathrm{T}_g$. For further discussion on this topic, one may refer to well known textbooks on the subject, such as [18] and [24].

Finally, it is known that for a Fuchsian group of signature $(g; m_1, \ldots, m_r)$, the Teichmüller space $\mathrm{Teich}(\Gamma)$ is a manifold of real dimension $6g - 6 + 2r$ homeomorphic to a ball.

CHAPTER 5

THE ARITHMETIC OF KLEINIAN AND FUCHSIAN GROUPS

In this chapter, the arithmetic introduced in Chapters 2 and 3 is combined with the geometry of Chapter 4. We will attach algebraic objects to Kleinian and Fuchsian groups that turn out to be invariants of their commensurability classes. These algebraic objects, the invariant trace field and quaternion algebra, often carry geometric information about the orbifold uniformised by these groups. Arithmetic Kleinian and Fuchsian groups will be defined, and also the object of interest of the present thesis: semi-arithmetic Fuchsian groups.

5.1 Trace field and associated quaternion algebra

Let $\Gamma < \mathrm{PSL}(2, \mathbb{C})$ be a Kleinian group. Denote by $\widetilde{\Gamma}$ the preimage of Γ in $\mathrm{SL}(2, \mathbb{C})$. We define the following algebraic object associated to Γ.

Definition 5.1.1. The *trace field* of Γ, denoted $\mathbb{Q}(\mathrm{tr}\,\Gamma)$, is defined as $\mathbb{Q}(\mathrm{tr}\,\tilde{\gamma} \mid \tilde{\gamma} \in \widetilde{\Gamma})$.

The first interplay between geometry and algebra that we can observe in this setting is the following:

Theorem 5.1.2. *Let Γ be a Kleinian group. If Γ is cofinite then $\mathbb{Q}(\mathrm{tr}\,\Gamma)$ is a number field.*

Sketch of proof. Let us treat Γ as a subgroup of $\mathrm{SL}(2, \mathbb{C})$ (look at its preimage in $\mathrm{SL}(2, \mathbb{C})$). If Γ is cofinite then it is finitely generated (finitely presented even). Let Γ_1 be a torsion free subgroup of finite index, which exists, by Selberg's Lemma. Note that Γ_1 is also cofinite and thus also finitely generated by, say, $\gamma_1, \ldots, \gamma_n$ in

$\mathrm{SL}(2, \mathbb{C})$. Normalise Γ such that γ_1 fixes 0 and ∞, and γ_2 fixes 1. In other words, if $\gamma_i = \left(\begin{smallmatrix} x_i & y_i \\ z_i & w_i \end{smallmatrix}\right)$, this normalisation means that $y_1 = z_1 = 0$ and $x_2 + y_2 = z_2 + w_2$.

Consider the algebraic set of all representations of Γ_1 into $\mathrm{SL}(2, \mathbb{C})$ (c.f. §4.3.4), which is identified with a subset of $\mathrm{SL}(2, \mathbb{C})^n$ (for an explicit description of this subset, see [36, §1.6]). Consider then the subset of those representations satisfying the normalisation imposed above and let $V(\Gamma_1)$ be an irreducible component of this algebraic set, containing the inclusion $i : \Gamma_1 \hookrightarrow \mathrm{SL}(2, \mathbb{C})$. It follows from Mostow's Rigidity Theorem that the variety $V(\Gamma_1)$ is zero dimensional. Indeed, if it had positive dimension, by perturbing the inclusion i in $V(\Gamma_1)$, one would obtain a continuous family of (distinct) subgroups of $\mathrm{SL}(2, \mathbb{C})$, all of which are isomorphic to Γ_1 (see Theorem 4.3.23 and the discussion preceding it). Mostow's Rigidity Theorem implies that each of these groups are conjugate to Γ_1 in $\mathrm{Isom}(\mathbb{H}^3)$. However if $g\Gamma g^{-1}$ is to satisfy the normalisation imposed, then there are only finitely many possibilities for g, a contradiction.

From the facts that $V(\Gamma)$ is defined over $\mathbb{Q}$ and is zero dimensional, one derives that $V(\Gamma)$ must be a single point with algebraic coordinates ([36, Lemma 3.1.5]). So, all the matrices in Γ_1 have algebraic entries. We claim that the same is true for Γ. Indeed, if all the entries of the matrices in Γ_1 are algebraic, then all the traces of Γ_1 are algebraic. This property is easily seen to be carried over to Γ, since the trace of a power of an element of $\mathrm{SL}(2, \mathbb{C})$ is an integral (monic) polynomial of the trace of that element. So $\mathbb{Q}(\mathrm{tr}\, \Gamma)$ is an algebraic field. We will see later in Corollary 5.1.7 that, up to conjugation, Γ is a subgroup of $\mathrm{SL}(2, \mathbb{Q}(\mathrm{tr}\, \Gamma)(x_1))$ (recall that x_1 is the first coordinate of γ_1) and since x_1 is algebraic, it follows that (up to conjugation) the elements of Γ have algebraic entries. Since Γ is finitely generated, all entries of the matrices in Γ lie in a finite extension of $\mathbb{Q}$. In particular, so do all the traces and we conclude that $\mathbb{Q}(\mathrm{tr}\, \Gamma)$ is a number field. □

For the remainder of this section, let us assume that Γ is *non-elementary*, i.e., that Γ acting on $\overline{\mathbb{H}^3} = \mathbb{H}^3 \cup \partial\mathbb{H}^3$ does not have any finite orbits. One important property of non-elementary groups is that they contain infinitely many loxodromic elements (see Remark 4.2.3) no two of which have a common fixed point ([2, Theorem 5.1.3]). When all elements of a subgroup of $\mathrm{PSL}(2, \mathbb{C})$ have a common fixed point, we say this group is *reducible*. Otherwise, the group is said to be *irreducible*. In particular, a non-elementary group Γ contains two elements g, h such that $\langle g, h\rangle$ is irreducible. Now, a simple criterion for reducibility is that $\langle g, h\rangle$ is reducible if and only if $\mathrm{tr}\,[g, h] = 2$ ([2, Theorem 4.3.5]). Given two elements $x, y \in \mathrm{PSL}(2, \mathbb{C})$, let X, Y denote representatives in $\mathrm{SL}(2, \mathbb{C})$ and let $m(x, y)$ denote the determinant of the 4×4 matrix with columns Id, X, Y and XY. Note that $m(x, y)$ does not depend on the choice of lifts X and Y and thus is well-defined. Direct computation shows that

$$m(x, y) = 2 - \mathrm{tr}\,[x, y].$$

Putting all this together we obtain the following:

Lemma 5.1.3. *Let $x, y \in \mathrm{PSL}(2, \mathbb{C})$. The group $\langle x, y\rangle$ is irreducible if and only if the vectors* Id, X, Y, XY *are linearly independent in $M_2(\mathbb{C})$.*

In particular, for a non-elementary group Γ, there exist $x, y \in \Gamma$ such that Id, X, Y, XY *are linearly independent in $M_2(\mathbb{C})$.*

Remark 5.1.4. In what follows, we will sometimes consider Γ to be a discrete subgroup of $\mathrm{PSL}(2, F)$ and sometimes a discrete subgroup of $\mathrm{SL}(2, F)$, where $F = \mathbb{R}, \mathbb{C}$. In most cases, one can either project or lift, if necessary, without affecting the reasoning. Because of that, we will often abuse notation an treat elements of $\mathrm{PSL}(2, F)$ simply as matrices without explicit warning.

Another algebraic object we associate to Γ is the algebra $A_0\Gamma$ defined over its trace field. Denote by $\widetilde{\Gamma}$ the preimage of Γ in $\mathrm{SL}(2, \mathbb{C})$.

Definition 5.1.5. The *associated quaternion algebra* $A_0\Gamma$ is defined over $\mathbb{Q}(\mathrm{tr}\,\Gamma)$ as:

$$A_0\Gamma = \left\{ \sum_i a_i\gamma_i \mid a_i \in \mathbb{Q}(\mathrm{tr}\,\Gamma),\ \gamma_i \in \widetilde{\Gamma} \right\}, \tag{5.1.1}$$

where the sums are all finite.

This terminology is justified by the next proposition, which proves that $A_0\Gamma$ is indeed a quaternion algebra over $\mathbb{Q}(\mathrm{tr}\,\Gamma)$.

Proposition 5.1.6. *$A_0\Gamma$ is a quaternion algebra over $\mathbb{Q}(\mathrm{tr}\,\Gamma)$, where Γ is assumed to be non-elementary.*

Proof. According to Theorem 3.3.5, we must show that $A_0\Gamma$ is 4-dimensional, central and simple over $\mathbb{Q}(\mathrm{tr}\,\Gamma)$.

Let $g, h \in \Gamma$ be such that $\langle g, h\rangle$ is irreducible (Lemma 5.1.3). Consider the trace form T on $M_2(\mathbb{C})$ defined by $T(X, Y) = \mathrm{tr}\,(XY)$ and note that T is a non-degenerate symmetric bilinear form. Let $\{\mathrm{Id}^*, g^*, h^*, (gh)^*\}$ be the dual basis with respect to T. Any $\gamma \in \Gamma$ can thus be written as a K-linear combination of this dual basis whose coefficients will be of the form

$$T(\gamma, \gamma_i) = \mathrm{tr}\,(\gamma\gamma_i),$$

where γ_i is one of the elements $\{\mathrm{Id}, g, h, gh\}$. But $\mathrm{tr}\,(\gamma\gamma_i) \in \mathbb{Q}(\mathrm{tr}\,\Gamma)$, whence

$$\mathbb{Q}(\mathrm{tr}\,\Gamma)[\mathrm{Id}, g, h, gh] \subset A_0\Gamma \subset \mathbb{Q}(\mathrm{tr}\,\Gamma)[\mathrm{Id}^*, g^*, h^*, (gh)^*].$$

So, $A_0\Gamma$ is 4-dimensional over $\mathbb{Q}(\operatorname{tr}\Gamma)$.

If c is in the centre of $A_0\Gamma$ then it is in the centre of $A_0\Gamma \otimes_{\mathbb{Q}(\operatorname{tr}\Gamma)} \mathbb{C} \cong M_2(\mathbb{C})$ and thus c must be a multiple of the identity. Similarly, if I is a two-sided ideal of $A_0\Gamma$, then $I \otimes_{\mathbb{Q}(\operatorname{tr}\Gamma)} \mathbb{C}$ is a two-sided ideal of $M_2(\mathbb{C})$. Since $M_2(\mathbb{C})$ is simple, $I \otimes_{\mathbb{Q}(\operatorname{tr}\Gamma)} \mathbb{C} = M_2(\mathbb{C})$, which implies that I has dimension 4 over $\mathbb{Q}(\operatorname{tr}\Gamma)$. □

As a consequence of the proof above, we obtain that

Corollary 5.1.7. *If g and h are two elements of the subgroup Γ of $\mathrm{SL}(2,\mathbb{R})$ such that $\langle g, h\rangle$ is irreducible, then $A_0\Gamma$ is a quaternion algebra over $\mathbb{Q}(\operatorname{tr}\Gamma)$ and*

$$A_0\Gamma = \mathbb{Q}(\operatorname{tr}\Gamma)[\mathrm{Id}, g, h, gh].$$

Corollary 5.1.8. *Let Γ be a non-elementary Kleinian group with trace field $K = \mathbb{Q}(\operatorname{tr}\Gamma)$ and let g be a loxodromic element of Γ with eigenvalue λ. Then Γ is conjugated to a subgroup of $\mathrm{SL}(2, K(\lambda))$.*

Proof. Let $g, h \in \mathrm{PSL}(2,\mathbb{C})$ be such that $\langle g, h\rangle$ is irreducible and g is loxodromic. After conjugation, we can assume the fixed points of g are 0 and ∞, that is

$$g = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}, \qquad h = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Note that b and c are both non-zero. We can normalise so that $b = 1$ (conjugate by a diagonal matrix of determinant 1).

Now, λ satisfies a quadratic equation over K, so that $K(\lambda)$ has degree at most 2 over K. Moreover, since $a + d = \operatorname{tr} h$ and $\lambda a + \lambda^{-1} d = \operatorname{tr} gh$ are both in K, it then follows that a, d and $c = ad - 1$ also lie in $K(\lambda)$. By Corollary 5.1.7, $A_0\Gamma \subset M_2(K(\lambda))$. The result follows. □

Corollary 5.1.9. *If Γ is a non-elementary subgroup of $\mathrm{SL}(2,\mathbb{C})$ such that $K = \mathbb{Q}(\operatorname{tr}\Gamma)$ is real, then Γ is conjugated to a subgroup of $\mathrm{SL}(2, K(\lambda))$ where $K(\lambda) \subset \mathbb{R}$.*

Proof. This is an immediate consequence of Corollary 5.1.8. Indeed, let g be as before. If $\operatorname{tr} g$ is real and $|\operatorname{tr} g| > 2$, then λ is a real root of the polynomial $X^2 - (\operatorname{tr} g)X + 1$. □

When the traces of Γ are all algebraic integers (a situation that will arise frequently in the future), an order of $A_0\Gamma$ can be described in a way similar to (5.1.1):

Proposition 5.1.10. *Let Γ be a non-elementary subgroup of $\mathrm{SL}(2,\mathbb{C})$ whose elements all have algebraic integral traces. Then*

$$\mathcal{O}\Gamma = \left\{ \sum_i a_i\gamma_i \mid a_i \in \mathcal{O}_{\mathbb{Q}(\operatorname{tr}\Gamma)},\ \gamma_i \in \Gamma \right\},$$

is an order in $A_0\Gamma$.

Proof. Let K denote the field of traces $\mathbb{Q}(\operatorname{tr}\Gamma)$. It is immediate that $\mathcal{O}\Gamma$ is an $\mathcal{O}_K$-module and a ring with unity. Moreover, it follows from Corollary 5.1.7 that $\mathcal{O}\Gamma$ contains a K-basis of $A_0\Gamma$. So we only need to check that $\mathcal{O}\Gamma$ is finitely generated as a module over $\mathcal{O}_K$.

Let $g, h \in \Gamma$ be such that $\langle g, h\rangle$ is irreducible. As in the proof of Theorem 5.1.6, consider the trace form T and let $\{\mathrm{Id}^*, g^*, h^*, (gh)^*\}$ be the dual basis with respect to T. Any $\gamma \in \Gamma$ can thus be written as a K-linear combination of this dual basis whose coefficients will be of the form

$$T(\gamma, \gamma_i) = \operatorname{tr}(\gamma\gamma_i),$$

where γ_i is one of the elements $\{\mathrm{Id}, g, h, gh\}$. But $\operatorname{tr}(\gamma\gamma_i)$ is, by hypothesis, an algebraic integer, which means that $\mathcal{O}\Gamma \subset \mathcal{O}_K[\mathrm{Id}^*, g^*, h^*, (gh)^*]$. Let M denote the module $\mathcal{O}_K[\mathrm{Id}^*, g^*, h^*, (gh)^*]$.

Now, every element in the basis $\{\mathrm{Id}^*, g^*, h^*, (gh)^*\}$ is a K-linear combination of $\{\mathrm{Id}, g, h, gh\}$ and so it follows that, for some appropriate integer a, $aM \subset \mathcal{O}\Gamma$. The quotient M/aM is finite (since $\mathcal{O}_K/a\mathcal{O}_K$ is finite). Therefore, we can pick a finite set of representatives for the lateral classes in M/aM (that intersect $\mathcal{O}\Gamma$), such that, together with $\{\mathrm{Id}^*, g^*, h^*, (gh)^*\}$, they constitute a finite generating set for $\mathcal{O}\Gamma$. (Alternatively, see Theorem 2.2.28). □

5.2 Invariant trace field and quaternion algebra

Definition 5.2.1. Two groups are said to be *commensurable* when their intersection has finite index in each of them. Two subgroups Γ_1 and Γ_2 of (P)SL$(2,\mathbb{C})$ are said to be *commensurable in the wide sense* if Γ_1 is commensurable to a conjugate of Γ_2.

Commensurability is clearly an equivalence relation. Geometrically, it means that two surfaces have a common finite sheeted cover (possibly of different degrees). More precisely, let Γ_1 and Γ_2 be two commensurable subgroups of $\mathrm{PSL}(2,\mathbb{R})$. Assume, for simplicity, that they are torsion-free. Then $(\Gamma_1 \cap \Gamma_2)\backslash\mathbb{H}$ is a cover of $\Gamma_i\backslash\mathbb{H}$ of degree $[\Gamma_i : \Gamma_1 \cap \Gamma_2] < \infty$, $i = 1, 2$.

The trace field defined in the previous section is *not* a commensurability invariant. Indeed, we have the following counter-example given in [45]:

Example 5.2.2. Let Γ be the Kleinian group generated by the image of the elements

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ -\omega & 1 \end{pmatrix},$$

in $\mathrm{PSL}(2,\mathbb{C})$, where $\omega = (-1+\sqrt{-3})/2$. Note that Γ is a subgroup of $\mathrm{PSL}(2,\mathbb{Z}[\omega])$ and the latter is discrete, since $\mathbb{Z}[\omega]$ is a discrete subring of $\mathbb{C}$. Moreover, $\mathbb{Q}(\mathrm{tr}\,\Gamma)$ = $\mathbb{Q}(\sqrt{-3})$. Now, let $\widetilde{X} = \left(\begin{smallmatrix} i & 0 \\ 0 & -i \end{smallmatrix}\right)$ and let X be its image in $\mathrm{PSL}(2,\mathbb{C})$. Then Γ is a subgroup of index 2 in $\Lambda = \langle \Gamma, X\rangle$. However, Λ contains the element $XBA = \left(\begin{smallmatrix} i & i \\ i\omega & -i+i\omega \end{smallmatrix}\right)$ so that, in particular, $i \in \mathbb{Q}(\mathrm{tr}\,\Lambda)$.

In order to remedy this situation, we focus on a finite index subgroup of Γ whose trace field is an invariant of the commensurability class of Γ.

For the remainder of this section, assume that Γ is a finitely generated non-elementary subgroup of $(\mathrm{P})\mathrm{SL}(2,\mathbb{C})$.

Definition 5.2.3. Let $\Gamma^{(2)} = \langle \gamma^2 \mid \gamma \in \Gamma\rangle$.

Proposition 5.2.4. $\Gamma^{(2)}$ *is a finite index normal subgroup of* Γ *and* $\Gamma/\Gamma^{(2)}$ *is a finite abelian torsion group.*

Proof. Let α be any element in Γ. Note that α normalises the set $\{\gamma^2 \mid \gamma \in \Gamma\}$ and therefore $\Gamma^{(2)}$. This proves that $\Gamma^{(2)}$ is normal in Γ. Now, given that every element of the quotient has order two, it follows that $\Gamma/\Gamma^{(2)}$ is abelian. Moreover, since Γ is finitely generated, $\Gamma/\Gamma^{(2)}$ is finite. □

Theorem 5.2.5 (Invariance of $\mathbb{Q}(\mathrm{tr}\,\Gamma^{(2)})$)**.** *The field* $\mathbb{Q}(\mathrm{tr}\,\Gamma^{(2)})$ *is an invariant of the commensurability class of* Γ*, for* Γ *a finitely generated non-elementary subgroup of* $\mathrm{SL}(2,\mathbb{C})$.

Proof. We begin by asserting the following claim, that says that $\Gamma^{(2)}$ has the minimal trace field among all finite index subgroups of Γ. More precisely:

Claim: If Γ_1 is a finite index subgroup of Γ then $\mathbb{Q}(\mathrm{tr}\,\Gamma^{(2)}) \subset \mathbb{Q}(\mathrm{tr}\,\Gamma_1)$.

Proof of claim: We may assume, without loss in generality, that Γ_1 is a normal subgroup of Γ. Indeed, let $C = \bigcap_{\gamma\in\Gamma} \gamma\Gamma_1\gamma^{-1}$, which is clearly a normal subgroup of Γ. Since Γ_1 has finite index in Γ, there exist $\gamma_1, \ldots, \gamma_n$ such that any $\gamma \in \Gamma$ may be written as $\gamma = \gamma_i\gamma'$ for some $\gamma' \in \Gamma_1$ and some $1 \leq i \leq n$. In particular, $C = \bigcap_{i=1}^{n} \gamma_i\Gamma_1\gamma_i^{-1}$ has finite index in Γ.

Let γ be any element in Γ. We want to prove that $\gamma^2 \in A_0\Gamma_1$. Since Γ_1 is normal, conjugation by γ is an automorphism of Γ_1, which, in turn, induces an automorphism of the quaternion algebra $A_0\Gamma_1$. By the Skolem-Noether Theorem (see Corollary 3.1.10), there exists an invertible element $a \in A_0\Gamma_1$ such that $\gamma x\gamma^{-1} = axa^{-1}$ for every $x \in A_0\Gamma_1$. It follows that $\gamma^{-1}a$ commutes with every element of $A_0\Gamma_1$ and, consequently, with every element of $A_0\Gamma_1 \otimes \mathbb{C} \cong M_2(\mathbb{C})$. It follows that $\gamma^{-1}a = c\mathrm{Id}$ for some complex number $c \in \mathbb{C}$. We observe next that $c^2 \in \mathbb{Q}(\mathrm{tr}\,\Gamma_1)$:

$$c^2 = \det(c\mathrm{Id}) = \det(\gamma^{-1})\det(a) = \det(a),$$

and since a satisfies its own characteristic equation, $\det(a)\mathrm{Id} = \mathrm{tr}\,(a)a - a^2 \in A_0\Gamma_1$ and so $\det(a) \in \mathbb{Q}(\mathrm{tr}\,\Gamma_1)$. Then, $\gamma^2 = c^2a^{-2}$ is in $A_0\Gamma_1$, as we wanted to show.

Finally, if $\{\gamma^2 \mid \gamma \in \Gamma\}$ is a subset of $A_0\Gamma_1$, then $\Gamma^{(2)} \subset A_0\Gamma_1$, whence $\mathrm{tr}\,\Gamma^{(2)} \subset \mathbb{Q}(\Gamma_1)$, concluding the proof of the claim.

The theorem follows at once: let Γ and Λ be commensurable groups. By transitivity, $\Gamma^{(2)}$ and $\Lambda^{(2)}$ are commensurable, which implies that $\Gamma^{(2)} \cap \Lambda^{(2)}$ also has finite index both in Γ and in Λ. From the claim we obtain

$$\mathbb{Q}(\mathrm{tr}\,\Gamma^{(2)}) \subset \mathbb{Q}(\mathrm{tr}\,(\Gamma^{(2)} \cap \Lambda^{(2)})),$$

where the latter is clearly a subfield of $\mathbb{Q}(\mathrm{tr}\,\Lambda^{(2)})$. Similarly, $\mathbb{Q}(\mathrm{tr}\,\Lambda^{(2)}) \subset \mathbb{Q}(\mathrm{tr}\,\Gamma^{(2)})$. □

By Corollary 5.1.7, the quaternion algebra $A_0\Gamma^{(2)}$ is generated over $\mathbb{Q}(\mathrm{tr}\,\Gamma^{(2)})$ with basis $\{\mathrm{Id}, g, h, gh\}$, for g, h in $\Gamma^{(2)}$ satisfying certain properties. If Λ is commensurable to Γ, by choosing g and h in $\Gamma^{(2)} \cap \Lambda^{(2)}$, we establish the following:

Corollary 5.2.6. *For a finitely generated non-elementary group $\Gamma < \mathrm{SL}(2,\mathbb{C})$, the quaternion algebra $A_0\Gamma^{(2)}$ is an invariant of the commensurability class of Γ.*

These two algebraic invariants associated to a Kleinian group are central to this study. They receive special names and notation according to the following definition:

Definition 5.2.7. The *invariant trace field* of Γ, denoted $k\Gamma$, is the trace field of $\Gamma^{(2)}$, i.e., $\mathbb{Q}(\mathrm{tr}\,\Gamma^{(2)})$. The *invariant quaternion algebra* of Γ, denoted $A\Gamma$, is the quaternion algebra $A_0\Gamma^{(2)}$.

With what we already know about quaternion algebras, we can readily establish the following:

Theorem 5.2.8. *If Γ is a non-elementary subgroup of $\mathrm{SL}(2,\mathbb{C})$ that contains a parabolic element, then $A_0\Gamma = M_2(\mathbb{Q}(\mathrm{tr}\,\Gamma))$.*

Proof. Let $\gamma \in \Gamma$ be parabolic. Its characteristic equation is $\gamma^2 \pm 2\gamma + \mathrm{Id} = 0$, which means that $\gamma \pm \mathrm{Id}$ is a zero divisor. The theorem then follows from Proposition 3.3.4. □

We shall see now some ways in which we can calculate the trace field and associated quaternion algebra of a given group Γ. It turns out that one does not need all the traces in order to determine the trace field. In fact, when Γ is finitely generated, only finitely many traces are necessary. This is a consequence of the trace relations between elements of $\mathrm{SL}(2,\mathbb{C})$. A detailed exposition can be found in [36, §3.4]. Here we merely state the main results to be used later on.

Recall that, for any two $n \times n$ matrices X and Y, we have that $\operatorname{tr} XY = \operatorname{tr} YX$. More generally, the trace is invariant under any cyclic permutation of the factors of a product of matrices and, in particular, it is invariant under conjugation.

For $X, Y \in \mathrm{SL}(2,\mathbb{C})$, one obtains, by direct calculation, the following identity:

$$\operatorname{tr}(XY) = (\operatorname{tr} X)(\operatorname{tr} Y) - \operatorname{tr}(XY^{-1}).$$

Note that making $X = Y$ yields:

$$\operatorname{tr} X^2 = (\operatorname{tr} X)^2 - 2,$$

which may also be deduced from the characteristic equation of $X \in \mathrm{SL}(2,\mathbb{C})$, namely:

$$X^2 - (\operatorname{tr} X)X + \mathrm{Id} = 0.$$

It is also worth noting that, once we are working with matrices in $\mathrm{SL}(2,\mathbb{C})$, we have that $\operatorname{tr} X = \operatorname{tr} X^{-1}$.

Assume Γ is finitely generated, say, by $\{\gamma_1, \dots, \gamma_n\}$. Let Q and R denote the following collections of elements of Γ:

$$Q = \{\gamma_{i_1} \cdots \gamma_{i_k} \mid 1 \le k \le n,\ 1 \le i_1 < \cdots < i_k \le n\}$$

and

$$R = \{\gamma_i, \gamma_{j_1}\gamma_{j_2}, \gamma_{k_1}\gamma_{k_2}\gamma_{k_3} \mid 1 \le i \le n,\ 1 \le j_1 < j_2 \le n,\ 1 \le k_1 < k_2 < k_3 \le n\}.$$

Proposition 5.2.9 ([36, Lemma 3.5.2]). *For $\gamma \in \Gamma$, its trace $\operatorname{tr}\gamma$ is an integer polynomial in $\{\operatorname{tr}\delta \mid \delta \in Q\}$. In particular, it follows that*

$$\mathbb{Q}(\operatorname{tr}\Gamma) = \mathbb{Q}(\operatorname{tr}\delta \mid \delta \in Q) \quad \textit{and} \quad \mathbb{Z}[\operatorname{tr}\Gamma] = \mathbb{Z}[\operatorname{tr}\delta \mid \delta \in Q]$$

Proposition 5.2.10 ([36, Lemma 3.5.3]). *For $\gamma \in \Gamma$, its trace $\operatorname{tr}\gamma$ is a rational polynomial in $\{\operatorname{tr}\delta \mid \delta \in R\}$. In particular, $\mathbb{Q}(\operatorname{tr}\Gamma) = \mathbb{Q}(\operatorname{tr}\delta \mid \delta \in R)$.*

Given a finitely generated $\Gamma < \mathrm{SL}(2,\mathbb{C})$, we may, in principle, calculate the invariant trace field $k\Gamma$, applying the results stated above to the (also finitely generated) group $\Gamma^{(2)}$. This may, however, present some technical difficulties as, for instance, finding a generating set for $\Gamma^{(2)}$. Lemma 5.2.12 and Theorem 5.2.13 below are quite useful in these situations.

Definition 5.2.11. For $\Gamma < \mathrm{SL}(2,\mathbb{C})$, with generators $\{\gamma_1, \ldots, \gamma_n\}$, define the subgroup Γ^{SQ} to be:

$$\Gamma^{SQ} = \langle \gamma_1^2, \ldots, \gamma_n^2 \rangle.$$

Lemma 5.2.12 ([36, Lemma 3.5.5])**.** *For a non-elementary group Γ generated by $\{\gamma_1, \ldots, \gamma_n\}$, such that $\mathrm{tr}\,\gamma_i \neq 0$, for $i = 1, \ldots, n$, we have that $k\Gamma = \mathbb{Q}(\mathrm{tr}\,\Gamma^{SQ})$.*

Theorem 5.2.13 ([36, Lemma 3.5.6])**.** *For a finitely generated non-elementary group $\Gamma < \mathrm{SL}(2,\mathbb{C})$, we have that*

$$k\Gamma = \mathbb{Q}(\mathrm{tr}\,(\gamma^2) \mid \gamma \in \Gamma) = \mathbb{Q}(\mathrm{tr}^{\,2}\gamma \mid \gamma \in \Gamma).$$

We saw that $A\Gamma$ is a quaternion algebra over $k\Gamma$, when $\Gamma^{(2)}$ is a non-elementary subgroup of $\mathrm{SL}(2,\mathbb{C})$. For the sake of completeness, we describe next the Hilbert symbol of $A\Gamma$ in terms of certain elements of $\Gamma^{(2)}$:

Proposition 5.2.14 ([36, Theorem 3.6.1])**.** *Let g and h be elements of $\Gamma^{(2)}$ such that g is not parabolic and the group $\langle g, h\rangle$ is irreducible, then*

$$A\Gamma = \left(\frac{\mathrm{tr}^2 g - 4, \mathrm{tr}\,[g,h] - 2}{k\Gamma}\right).$$

Finally, we apply the previous result in order to obtain a description of $A\Gamma$ in terms of elements of Γ instead of $\Gamma^{(2)}$:

Proposition 5.2.15 ([36, Theorem 3.6.2])**.** *If g and h are elements of Γ such that $\langle g, h\rangle$ is irreducible, g and h do not have order 2 in $\mathrm{PSL}(2,\mathbb{C})$ and g is not parabolic, then*

$$A\Gamma = \left(\frac{\mathrm{tr}^2 g\,(\mathrm{tr}^2 g - 4), \mathrm{tr}^2 g\,\mathrm{tr}^2 h\,(\mathrm{tr}\,[g,h] - 2)}{k\Gamma}\right).$$

5.3 Arithmetic groups

5.3.1 Definition of arithmetic Kleinian and Fuchsian groups

Let K be a number field such that $n = [K : \mathbb{Q}] = r_1 + 2r_2$ where r_1 is the number of real places and r_2 is the number of complex places of K. Denote by $\sigma_1, \ldots, \sigma_n$ the n embeddings of K into $\mathbb{C}$, and by v_{σ_i} the valuation in K arising from the embedding σ_i. For simplicity, we denote the completion of K with respect to the valuation v_{σ_i} as K_{σ_i}.

We saw in Theorem 2.3.49 (see also Corollary 2.3.53) that, possibly after re-indexing the embeddings:

$$K \otimes_{\mathbb{Q}} \mathbb{R} \cong \bigoplus_{i=1}^{r_1+r_2} K_{\sigma_i},$$

where $K_{\sigma_i} \cong \mathbb{R}$ (resp. $\mathbb{C}$) if σ_i is a real (resp. complex) embedding. An analogous isomorphism can be found in the non-commutative setting of algebraic groups, through a construction known as *restriction of scalars* (see [36, §10.3] and references therein).

Heuristically, if we think of $A = \left(\frac{a,b}{K}\right)$ as an algebraic group defined over K (since $a, b \in K$), the $\mathbb{R}$-rational points of the group obtained from A by restriction of scalars should be isomorphic to the direct sum of $\left(\frac{\sigma_i(a),\sigma_i(b)}{K_{\sigma_i}}\right)$, $i = 1, \ldots, r_1 + r_2$. Recall that a quaternion algebra over $\mathbb{C}$ is necessarily split and a quaternion algebra over $\mathbb{R}$ is either isomorphic to $M_2(\mathbb{R})$ or to the Hamilton quaternions $\mathscr{H}$. So, in this case, it is reasonable to assume that $A \otimes_{\mathbb{Q}} \mathbb{R}$ should be isomorphic to a direct sum of factors $\mathscr{H}$ (as many as the real places of K over which A is ramified), as well as factors $M_2(\mathbb{R})$ and $M_2(\mathbb{C})$. This is indeed the case, as we see next. In the reference given below, the proof is straightforward and makes no use of restriction of scalars.

Theorem 5.3.1. *If A is a quaternion algebra over the number field K, that is ramified over s_1 real places, then*

$$A \otimes_{\mathbb{Q}} \mathbb{R} \cong \mathscr{H}^{\oplus s_1} \oplus M_2(\mathbb{R})^{\oplus(r_1 - s_1)} \oplus M_2(\mathbb{C})^{\oplus r_2}, \tag{5.3.1}$$

where $E^{\oplus m}$ denotes the direct sum of m copies of E.

Proof. [36, Theorem 8.1.1]. □

Assume that A is unramified at at least one place in V_∞. By considering the projection of the right side of (5.3.1) onto its non-compact factors, we obtain an embedding

$$\psi : A \to \bigoplus_{v \in V_\infty \setminus \mathrm{Ram}_\infty(A)} M_2(K_v). \tag{5.3.2}$$

The composition of the isomorphism (5.3.1) with the projection onto the ith factor gives an embedding ρ_i of A into $A_i = \mathscr{H}, M_2(\mathbb{R})$ or $M_2(\mathbb{C})$, extending the embedding σ_i, that preserves the trace and the norm of A. This means that, for $\alpha \in A$, $\mathrm{n}_i(\rho_i(\alpha)) = \rho_i(\mathrm{n}(\alpha)) = \sigma_i(\mathrm{n}(\alpha))$ and, similarly, $\mathrm{tr}_i(\rho_i(\alpha)) = \sigma_i(\mathrm{tr}(\alpha))$. In particular, the elements of $A^1 = \{\alpha \in A \mid n(\alpha) = 1\}$ are mapped by ψ into the elements whose coordinates are matrices in $M_2(K_v)$ with determinant 1:

$$\psi : A^1 \to \bigoplus_{v \in V_\infty \setminus \mathrm{Ram}_\infty(A)} \mathrm{SL}(2, K_v)$$

The following is a theorem of vital importance for the definition of arithmetic groups. Its proof relies on deeper tools that are beyond the scope of this thesis. Items (1) and (2) can be seen as a version of the famous Borel-Harish-Chandra Theorem in this specific setting.

Theorem 5.3.2. *Let A be a quaternion algebra which is unramified at at least one infinite place. Let $\mathscr{O}$ be an order in A and denote by $\mathscr{O}^1$ the elements of norm 1 in $\mathscr{O}$. Then the embedding ψ given in* (5.3.2) *is such that:*

1. *$\psi(\mathscr{O}^1)$ is discrete and of finite covolume in $\bigoplus_{v \in V_\infty \setminus \mathrm{Ram}_\infty(A)} \mathrm{SL}(2, K_v)$;*

2. *$\psi(\mathscr{O}^1)$ is cocompact if A is a division algebra;*

3. *If $T \neq \emptyset$ is a proper subset of $V_\infty \setminus \mathrm{Ram}_\infty(A)$, then $\psi(\mathscr{O}^1)$ projects to a dense set in $\bigoplus_{v \in T} \mathrm{SL}(2, K_v)$.*

Proof. [36, Theorem 8.1.2]. □

Remark 5.3.3. Note that the group $\psi(\mathscr{O}^1)$ acts on a product of copies of $\mathbb{H}^2$ and $\mathbb{H}^3$ which have a natural volume form arising from the area/volume forms of its factors. Therefore, for $\psi(\mathscr{O}^1)$ to be of finite covolume simply means that it admits a fundamental domain of finite volume.

Now we are finally ready to define arithmetic Kleinian and Fuchsian groups:

Definition 5.3.4 (Arithmetic Kleinian group)**.** Let K be a number field with one complex place and let A be a quaternion algebra over K which is ramified at all real places of K. Let ρ be an embedding of A into $M_2(\mathbb{C})$ and let $\mathscr{O}$ be an order of A. A subgroup Γ of $\mathrm{SL}(2, \mathbb{C})$ is an arithmetic Kleinian group if there exist ρ and $\mathscr{O}$ as above such that Γ is commensurable to $\rho(\mathscr{O}^1)$. Similarly, a subgroup of $\mathrm{PSL}(2, \mathbb{C})$ shall be called arithmetic when it is commensurable to $\mathrm{P}\rho(\mathscr{O}^1)$.

The quotient $\Gamma \backslash \mathbb{H}^3$ is said to be arithmetic when Γ is an arithmetic Kleinian group.

If $\rho' : A \to M_2(\mathbb{C})$ is any other embedding, it follows from the Skolem-Noether Theorem that ρ and ρ' differ by a conjugation by an element of $\mathrm{GL}(2, \mathbb{C})$. For this reason, if Γ is arithmetic, we may very well consider ρ to be the embedding ψ given in (5.3.2).

Note also that being arithmetic is independent of the order $\mathscr{O}$. Indeed, if $\mathscr{D}$ is any other order, then $\mathscr{O} \cap \mathscr{D}$ is also an order. By Theorem 5.3.2 (1), both $\rho((\mathscr{O} \cap \mathscr{D})^1)$ and $\rho(\mathscr{D}^1)$ are cofinite, so $\rho((\mathscr{O} \cap \mathscr{D})^1)$ has finite index in $\rho(\mathscr{D}^1)$. Likewise, $\rho((\mathscr{O} \cap \mathscr{D})^1)$ has finite index in $\rho(\mathscr{O}^1)$, which means that $\rho(\mathscr{O}^1)$ and $\rho(\mathscr{D}^1)$ are commensurable.

Analogously, there is a definition for the case of Fuchsian groups. Recall that a *totally real* number field K, is one for which every Galois embedding $\sigma : K \to \mathbb{C}$ has its image $\sigma(K)$ lying in $\mathbb{R}$.

Definition 5.3.5 (Arithmetic Fuchsian group)**.** Let K be a totally real number field and let A be a quaternion algebra over K which is ramified at all real places of

K but one. Let ρ be an embedding of A into $M_2(\mathbb{R})$ and let $\mathcal{O}$ be an order of A. A subgroup Γ of $\mathrm{SL}(2,\mathbb{R})$ is an arithmetic Fuchsian group if there exist ρ and $\mathcal{O}$ as above such that Γ is commensurable to $\rho(\mathcal{O}^1)$. Similarly, a subgroup of $\mathrm{PSL}(2,\mathbb{R})$ shall be called arithmetic when it is commensurable to $\mathrm{P}\rho(\mathcal{O}^1)$.

A hyperbolic 2-orbifold (in particular, a hyperbolic surface) $\Gamma\backslash\mathbb{H}^2$ is said to be arithmetic when Γ is an arithmetic Fuchsian group.

Again this is independent of the embedding ρ and of the choice of the order $\mathcal{O}$. Note that we can always assume that ρ is unramified at the identity $i : K \to \mathbb{C}$.

Remark 5.3.6. More generally, for a connected semi-simple algebraic group G defined over $\mathbb{Q}$, a subgroup Γ of the $\mathbb{Q}$-rational points $G(\mathbb{Q})$ is said to be arithmetic if there exists an embedding $\rho : G \to \mathrm{GL}_n$, defined over $\mathbb{Q}$, such that $\rho(\Gamma)$ is commensurable to the integral points $\rho(G)(\mathbb{Z})$.

The arithmetic subgroups of PGL_2, according to this definition, are precisely the Kleinian and Fuchsian groups given by definitions 5.3.4 and 5.3.5. For more details, see [36, §10.3] and the references therein.

5.3.2 Takeuchi's characterisation

Theorem 5.3.7. *Let A be a quaternion algebra over the number field k and ρ a k-embedding of A into $M_2(\mathbb{C})$. If Γ is an arithmetic Kleinian (or Fuchsian) group commensurable with $\rho(\mathcal{O}^1)$, where $\mathcal{O}$ is an order in A, then $k\Gamma = k$ and $A\Gamma = \rho(A)$.*

Proof. Commensurability implies that $k\Gamma = k\rho(\mathcal{O}^1)$. Now, for every element x in A, one has that $\operatorname{tr}\rho(x) = \operatorname{tr}(x) \in k$, so then $k\Gamma \subset k$. We claim that equality holds. Let us assume that for the moment and finish the argument.

Pick $g, h \in \Gamma^{(2)} \cap \rho(\mathcal{O}^1)$ such that $\langle g, h\rangle$ is irreducible. From the characterisation of the associated quaternion algebra given in Corollary 5.1.7, we see that $A_0(\Gamma^{(2)}) \subset A_0(\rho(\mathcal{O}^1))$. Also, as the trace set of $\rho(\mathcal{O}^1)$ is contained in k, we have that $A_0(\rho(\mathcal{O}^1)) \subset \rho(A)$. These inclusions put together show that $A\Gamma \subset \rho(A)$ are two quaternion algebras over k, whence they must be equal.

Going back to the the equality $k\Gamma = k$, we divide the proof in two cases.

Case 1: When Γ is an arithmetic Kleinian group, the field k has only one complex place. Then every proper subfield of k must be totally real. Since $k\Gamma$ cannot be real (otherwise Γ would be conjugate to a subgroup of $\mathrm{PSL}(2,\mathbb{R})$ and could not possibly have finite covolume acting on $\mathbb{H}^3$), it follows that $k\Gamma = k$.

Case 2: When Γ is an arithmetic Fuchsian group, k is a totally real field and A is ramified at every real place of k different from the identity. This means that

$\sigma_1 = \mathrm{Id} : k \to \mathbb{R}$ extends to a k-embedding $\rho : A \to M_2(\mathbb{R})$, while every other $\sigma_i : k \to \mathbb{R}$, $i = 2, \ldots, n$, extends to an embedding $\rho_i : A \to \mathscr{H}$.

Suppose $k\rho(\mathcal{O}^1)$ is a proper subfield of k. Then there must be a non-trivial embedding σ_i that restricts to the identity on $k\rho(\mathcal{O}^1)$. Now, $\sigma_i(\mathrm{tr}(\mathcal{O}^1)) = \mathrm{tr}(\rho_i(\mathcal{O}^1)) \subset \mathrm{tr}(\mathscr{H}^1) \subset [-2, 2]$. Since $\mathrm{tr}(\mathcal{O}^1)^{(2)} = \mathrm{tr}(\rho(\mathcal{O}^1)^{(2)})$, applying σ_i on both sides yields $\mathrm{tr}(\rho(\mathcal{O}^1)^{(2)}) \subset [-2, 2]$. But this means that no element of $\rho(\mathcal{O}^1)^{(2)}$ is hyperbolic, which cannot be the case for a non-elementary Fuchsian group. This contradiction proves that $k\rho(\mathcal{O}^1) = k$, and therefore that $k\Gamma = k$. □

Theorem 5.3.8 (Characterisation of arithmetic Kleinian groups)**.** *A cofinite Kleinian group Γ is arithmetic if and only if it satisfies the following three conditions:*

1. *$k\Gamma$ is a number field with exactly one complex place;*

2. *For every $\gamma \in \Gamma$, $\mathrm{tr}\,\gamma$ is an algebraic integer;*

3. *$A\Gamma$ is ramified at all real places of $k\Gamma$.*

Proof. If Γ is arithmetic, then it is commensurable to some $\rho(\mathcal{O}^1)$ where $\mathcal{O}$ is an order in a quaternion algebra A over k, both satisfying the conditions in (1) and (3), and $\rho : A \to M_2(\mathbb{C})$ is a k-embedding. From Theorem 5.3.7, it follows that $k\Gamma$ and $A\Gamma$ also satisfy these two conditions. Moreover, every $x \in \mathcal{O}$ is an integer (of A) and therefore $\mathrm{tr}\,x$ is an integer of k (see Proposition 2.2.28). Since the trace of $\rho(x)$ equals the reduced trace of x, every element of $\rho(\mathcal{O}^1)$ has integral trace. For any $\gamma \in \Gamma$, there exists an integer m such that $\gamma^m \in \rho(\mathcal{O}^1)$ and so $\mathrm{tr}\,(\gamma^m)$ is an algebraic integer. Note that $\mathrm{tr}\,(\gamma^m) = p(\mathrm{tr}\,\gamma)$, where p is a monic polynomial with integral coefficients. It then follows that $\mathrm{tr}\,\gamma$ is also an algebraic integer.

Conversely, assume that Γ is a cofinite Kleinian group satisfying conditions (1)-(3). In this case, (5.3.1) gives an isomorphism between $A\Gamma \otimes_{\mathbb{Q}} \mathbb{R}$ and $M_2(\mathbb{C}) \oplus \mathscr{H} \oplus \cdots \oplus \mathscr{H}$, which, in turn, induces a $k\Gamma$-homomorphism $\rho : A\Gamma \to M_2(\mathbb{C})$. Note that $A\Gamma \subset M_2(\mathbb{C})$, so that, by the Skolem-Noether Theorem, there exists $g \in \mathrm{GL}(2, \mathbb{R})$ for which $\rho(x) = gxg^{-1}$ for all $x \in A\Gamma$. In particular, for any $\gamma \in \Gamma^{(2)}$, we have that $\gamma = g^{-1}\rho(\gamma)g$. If n denotes the reduced norm in $A\Gamma$, then it follows that

$$1 = \det(\gamma) = \det(\rho(\gamma)) = \mathrm{n}(\gamma). \tag{5.3.3}$$

Recall that $\mathcal{O}\Gamma = \left\{\sum_i a_i\gamma_i \mid a_i \in \mathcal{O}_{k\Gamma},\ \gamma_i \in \Gamma^{(2)}\right\}$ is an order of $A\Gamma$, according to Proposition 5.1.10. Moreover, it follows from (5.3.3) that $\Gamma^{(2)} \subset (\mathcal{O}\Gamma)^1$. So, in particular, $\Gamma^{(2)} \subset g^{-1}\rho((\mathcal{O}\Gamma)^1)g$. Now, by Theorem 5.3.2, we know that $\rho((\mathcal{O}\Gamma)^1)$ has finite covolume. The group $\Gamma^{(2)}$ also has finite covolume since it is a finite index subgroup of the cofinite group Γ (note that for $\Gamma^{(2)}$ to have finite index in Γ it is necessary to assume that Γ is finitely generated, which is the case since it is cofinite). Denote by $\tilde{\rho}$ the $k\Gamma$-homomorphism of $A\Gamma$ into $M_2(\mathbb{C})$ obtained by composing ρ with conjugation by g^{-1}. Finite covolume then implies that $\Gamma^{(2)}$ is a finite index subgroup of $\tilde{\rho}((\mathcal{O}\Gamma)^1)$. The latter is therefore commensurable with Γ, which is hence, by definition, arithmetic. □

Definition 5.3.9. A Kleinian (resp. Fuchsian) group Γ is said to be *derived from a quaternion algebra* if it is a finite index subgroup of $(\mathrm{P})\rho(\mathcal{O}^1)$ for some order $\mathcal{O}$ in a quaternion algebra A over a number field k such that k has precisely one complex place and A is ramified over all of its real places (resp. k is totally real and A ramifies at every real place except one).

It follows from the last paragraph in the proof of Theorem 5.3.8 that every arithmetic Kleinian group is virtually derived from a quaternion algebra. More precisely:

Corollary 5.3.10. *A cofinite Kleinian group Γ is arithmetic if and only if $\Gamma^{(2)}$ is derived from a quaternion algebra.*

A completely analogous characterisation of arithmetic Fuchsian groups is available with the same proof, *mutatis mutandis*.

Theorem 5.3.11 (Characterisation of arithmetic Fuchsian groups)**.** *A cofinite Fuchsian group Γ is arithmetic if and only if it satisfies the following three conditions:*

1. *$k\Gamma$ is a totally real number field;*
2. *For every $\gamma \in \Gamma$, $\operatorname{tr}\gamma$ is an algebraic integer;*
3. *$A\Gamma$ is ramified at all real places of $k\Gamma$ except one.*

Similarly, it follows that a cofinite arithmetic Fuchsian group Γ is arithmetic if and only if $\Gamma^{(2)}$ is derived from a quaternion algebra.

A characterisation of arithmetic Fuchsian groups was originally provided by K. Takeuchi in 1975, solely in terms of the invariant trace field of the group. We state a version of his theorem below, which will be the starting point for the definition of *semi-arithmeticity* (see Definition 5.4.3). Some bits of the argument have already appeared in this subsection, so we sketch the rest of the proof relying on Theorem 5.3.11. It is worth noting that an analogous characterisation also holds for arithmetic Kleinian groups.

Theorem 5.3.12 (Takeuchi, [50])**.** *A Fuchsian group Γ is arithmetic if and only it satisfies the following two conditions:*

1. *$k\Gamma$ is an algebraic number field and $\operatorname{tr}\Gamma^{(2)} \subset \mathcal{O}_{k\Gamma}$;*

2. *If $\phi : k\Gamma \to \mathbb{C}$ is any Galois embedding different from the identity then $\phi(\operatorname{tr}\Gamma^{(2)})$ is bounded in $\mathbb{C}$.*

Sketch of Proof. If Γ is arithmetic, then (1) and (2) follow from Theorem 5.3.11.

To prove the sufficiency of (1) and (2), we start by pointing out again that the invariant trace field of $\Gamma^{(2)}$ coincides with its trace field. Since $\Gamma^{(2)}$ arithmetic implies Γ arithmetic, we may assume without loss in generality that Γ is a Fuchsian group for which $\mathbb{Q}(\operatorname{tr}\Gamma) = k\Gamma =: k$. Let us do so.

We note that conditions (1) and (2) together imply that $k\Gamma$ is totally real. Indeed, let $\gamma \in \Gamma$ have eigenvalues λ and $1/\lambda$. Suppose $\sigma : k \to \mathbb{C}$ is a Galois embedding different from the identity and extend it to an embedding $\tilde{\sigma} : k(\lambda) \to \mathbb{C}$. Then $|\sigma(\operatorname{tr}(\gamma^m))| = |\tilde{\sigma}(\lambda)^m + 1/\tilde{\sigma}(\lambda)^m|$. If $|\tilde{\sigma}(\lambda)| \neq 1$, then clearly $|\sigma(\operatorname{tr}(\gamma^m))|$ goes to infinity as m goes to infinity, contradicting condition (2). It then follows that

$$\sigma(\operatorname{tr}(\gamma)) = \tilde{\sigma}(\lambda) + \frac{1}{\tilde{\sigma}(\lambda)} = \tilde{\sigma}(\lambda) + \overline{\tilde{\sigma}(\lambda)} = 2\mathrm{Re}(\tilde{\sigma}(\lambda)) \in [-2, 2],$$

which, in particular, implies that $\sigma(\operatorname{tr}\gamma)$ is real for every Galois embedding σ and every $\gamma \in \Gamma$.

Next, we will argue that $A\Gamma$ is ramified at all infinite places of k different from the identity, and then the theorem will follow from Theorem 5.3.11.

As in the proof of Corollary 5.1.8, let g and h be two elements of Γ generating an irreducible subgroup, and conjugate so that

$$g = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}, \qquad h = \begin{pmatrix} a & 1 \\ c & d \end{pmatrix},$$

where $\lambda^2 \neq 1$ and $c \neq 0$.

Let $K = k(\lambda)$. Since $a + d$ and $\lambda a + \lambda^{-1} d$ both lie in $k = \mathbb{Q}(\operatorname{tr}\Gamma)$, a quick calculation shows that $d = \eta(a)$, where $\eta : K \to K$ is the nontrivial automorphism of $K \mid k$, sending λ to λ^{-1}. For this, one needs to argue that K is a proper extension of k. This is clear when k is a proper extension of $\mathbb{Q}$ (see the comments after (5.3.5) below) and, for $k = \mathbb{Q}$, notice that λ is a root of $X^2 - \operatorname{tr} g + 1$, which is irreducible over $\mathbb{Q}$ since $\operatorname{tr} g$ is, in this case, a rational integer greater than 2. Having said that, we recall that $A\Gamma$ can be expressed as $k[\mathrm{Id}, g, h, gh]$, whence

$$A\Gamma = \left\{ \begin{pmatrix} A & B \\ c\eta(B) & \eta(A) \end{pmatrix} \mid A, B \in K \right\}. \tag{5.3.4}$$

Let $\sigma : k \to \mathbb{C}$ be a Galois embedding different from the identity and extend it to an embedding $\tilde{\sigma}$ from $K = k(\lambda)$ into $\mathbb{C}$. We note that $\tilde{\sigma}(c) < 0$:

$$\begin{aligned} \sigma(\mathrm{tr}\,(g^m h)) = \tilde{\sigma}(\lambda^m a) + \tilde{\sigma}(\lambda^{-m} d) = \tilde{\sigma}(\lambda^m a) + \tilde{\sigma}\eta(\lambda^m a) = \\ = \tilde{\sigma}(\lambda^m a) + \overline{\tilde{\sigma}(\lambda^m a)} = 2\mathrm{Re}(\tilde{\sigma}(\lambda)^m \tilde{\sigma}(a)). \end{aligned} \tag{5.3.5}$$

where the second to last equality follows from the fact that $\tilde{\sigma} \circ \eta$ is the same as $\tilde{\sigma}$ composed with complex conjugation (indeed, by assuming that σ is nontrivial, we are tacitly assuming that k is a proper extension of $\mathbb{Q}$, in which case $|\sigma(\mathrm{tr}\, g)| < 2$, so that $\tilde{\sigma}(\lambda)$ and $\tilde{\sigma}\eta(\lambda)$ are complex roots of the real polynomial $X^2 - \sigma(\mathrm{tr}\, g)X + 1$, and thus complex-conjugates). Now, $\tilde{\sigma}(\lambda)$ is not a root of unity, so $\{\tilde{\sigma}(\lambda)^m \mid m \in \mathbb{Z}\}$ is dense in the unit circle, and then (5.3.5) together with continuity and our hypothesis on σ implies that $2\mathrm{Re}(z\tilde{\sigma}(a)) \leq 2$ for any $z \in \mathbb{C}$ of norm 1. In particular, $|\tilde{\sigma}(a)| \leq 1$. Finally, since $c = ad - 1$ and $d = \eta(a)$, applying $\tilde{\sigma}$ to both sides yields $\tilde{\sigma}(c) = |\tilde{\sigma}(a)|^2 - 1 \leq 0$. As $c \neq 0$, we get $\tilde{\sigma}(c) < 0$.

Let $\sigma_i : k \to \mathbb{C}$, $i = 2, \dots, n$ be any of the nontrivial Galois embeddings of k, and extend each of these to a k-homomorphism $\rho_i : A\Gamma \to M_2(K)$ obtained by applying $\tilde{\sigma_i}$ to the coordinates of the elements of $A\Gamma$, as in (5.3.4). We then obtain that

$$\rho_i(A\Gamma) \otimes_{\sigma_i(k)} \mathbb{R} = \left\{ \begin{pmatrix} A & B \\ \tilde{\sigma_i}(c)\overline{B} & \overline{A} \end{pmatrix} \mid A, B \in \mathbb{C} \right\}.$$

The quaternion algebra on the right admits the following standard basis:

$$1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad I = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad J = \begin{pmatrix} 0 & 1 \\ \tilde{\sigma_i}(c) & 0 \end{pmatrix}, \quad IJ = \begin{pmatrix} 0 & i \\ -\tilde{\sigma_i}(c)i & 0 \end{pmatrix},$$

so its Hilbert symbol can be written as $\left(\frac{-1, \tilde{\sigma_i}(c)}{\mathbb{R}}\right)$, which is isomorphic to $\mathscr{H}$ since $\tilde{\sigma_i}(c) < 0$, as we had established before. □

We conclude this section by showing that the invariant trace field and invariant quaternion algebras are *complete* commensurability invariants of a cofinite arithmetic Kleinian or Fuchsian group.

Theorem 5.3.13 (Complete Commensurability Invariants). *Let Γ_1 and Γ_2 be cofinite arithmetic Kleinian (resp. Fuchsian) subgroups of* $\mathrm{PSL}(2,\mathbb{C})$ *(resp.* $\mathrm{PSL}(2,\mathbb{R})$*). Then Γ_1 and Γ_2 are commensurable in the wide sense in* $\mathrm{PSL}(2,\mathbb{C})$ *(resp.* $\mathrm{PSL}(2,\mathbb{R})$*) if and only if $k\Gamma_1 = k\Gamma_2 = k$ and there exists a k-algebra isomorphism between $A\Gamma_1$ and $A\Gamma_2$.*

Proof. Let Γ_1 and Γ_2 be commensurable in the wide sense, so there exists $g \in \mathrm{SL}(2,\mathbb{C})$ such that $g\Gamma_1 g^{-1}$ and Γ_2 are commensurable. This implies, as we know, that $A(g\Gamma_1 g^{-1})$ and $A\Gamma_2$ are the same. Now the map taking $\sum_i a_i\gamma_i$ to $\sum_i a_i(g\gamma g^{-1})$, where $a_i \in k\Gamma_1 = k\Gamma_2 = k$ and $\gamma_i \in \Gamma_1^{(2)}$, is clearly a k-isomorphism between $A\Gamma_1$ and $A(g\Gamma_1 g^{-1}) = A\Gamma_2$.

Conversely, let $\phi : A\Gamma_1 \to A\Gamma_2$ be a k-isomorphism of algebras, where $k = k\Gamma_1 = k\Gamma_2$. We want to prove commensurability in the wide sense of Γ_1 and Γ_2. By transitivity, this is equivalent to commensurability in the wide sense of $\Gamma_1^{(2)}$ and $\Gamma_2^{(2)}$. The invariant trace field and quaternion algebra of these subgroups are the same as before. There is, though, the additional benefit that the trace field of $\Gamma_i^{(2)}$ coincides with its invariant trace field (this follows from the invariance under commensurability). Therefore we may assume from the beginning that the trace field of Γ_i, $i = 1, 2$, coincides with its invariant trace field, i.e., that $k = \mathbb{Q}(\operatorname{tr}\Gamma_1) = \mathbb{Q}(\operatorname{tr}\Gamma_2)$.

From Corollaries 5.1.8 and 5.1.9, there exists a finite field extension $K \mid k$ (a finite real extension, in the case of Fuchsian groups) such that, up to conjugation, $\Gamma_2 < \mathrm{PSL}(2, K)$. Assume, then, without loss in generality, that Γ_2 is in fact a subgroup of $\mathrm{SL}(2, K)$, in which case, the invariant quaternion algebra $A\Gamma_2$ is a subalgebra of $M_2(K)$. Considering then ϕ as a k-algebra homomorphism from $A\Gamma_1$ into $M_2(K)$, we can apply the Skolem-Noether Theorem and obtain an invertible element $g \in \mathrm{SL}(2, K)$ (after normalising it), such that $\phi(x) = gxg^{-1}$ for any $x \in A\Gamma_1$. Now, as in the proof of Theorem 5.3.8, each $\mathcal{O}\Gamma_i^1$ is commensurable with Γ_i, $i = 1, 2$. Since ϕ preserves norm, it follows, in particular, that $\phi(\mathcal{O}\Gamma_1)^1 = \phi(\mathcal{O}\Gamma_1^1)$ is commensurable with $\phi(\Gamma_1)$. Note also that the unit subgroups $\mathcal{O}\Gamma_2^1$ and $\phi(\mathcal{O}\Gamma_1)^1$ are commensurable (see Theorem 5.3.2). Putting all these together, we find that Γ_2 is commensurable with $\phi(\Gamma_1)$, and the latter is but $g\Gamma_1 g^{-1}$.

□

5.4 Semi-arithmetic Fuchsian groups

5.4.1 Fuchsian groups with prescribed algebraic data

In this subsection we address the inverse problem of, given the quaternion algebra A over a number field k, finding a Fuchsian group with invariant trace field k and invariant quaternion algebra A.

Using their work on the Ehrenpreis Conjecture, J. Kahn and V. Markovic proved the following theorem:

Theorem 5.4.1 (Kahn-Markovic, [27]). *Let K be a real number field and let A be a quaternion algebra over K that splits over the identity and such that $A \not\cong M_2(\mathbb{R})$. Then there exists a cocompact Fuchsian group Γ such that $k\Gamma = K$ and $A\Gamma = A$.*

Moreover, in the construction of Theorem 5.4.1, the group Γ realising A and K has integral traces, i.e., $\operatorname{tr}\Gamma$ is composed by algebraic integers. There is no control, however, over the genus of Γ. If we fix a genus $g \geq 2$, it is still possible to realise quaternion algebras as before, though perhaps not with integral traces. This was proved by B. Jeon in [25]:

Theorem 5.4.2. *Let K be a real number field and let A be a quaternion algebra over K that splits over the identity and such that $A \not\cong M_2(\mathbb{R})$. Then, for any $g \geq 2$, there exists a cocompact surface group Γ of genus g such that $k\Gamma = K$ and $A\Gamma = A$.*

5.4.2 Definition of semi-arithmetic groups

Given the characterisation in Theorem 5.3.12, one possible way to define a class of Fuchsian groups larger than that of arithmetic groups is to simply waive the hypothesis on the boundedness of $\phi(\operatorname{tr}\Gamma^{(2)})$. Note, however, that conditions (1) and (2) together imply that $k\Gamma$ is totally real. This property is retained in the following definition:

Definition 5.4.3 ([49])**.** A cofinite Fuchsian group Γ is said to be *semi-arithmetic* when $k\Gamma = \mathbb{Q}(\operatorname{tr}\Gamma^{(2)})$ is a totally real number field and $\operatorname{tr}\Gamma^{(2)} \subset \mathcal{O}_{k\Gamma}$.

Equivalently, it follows from elementary trace relations that Γ is semi-arithmetic if $k\Gamma$ is a totally real number field and, for every $\gamma \in \Gamma$, $\operatorname{tr}\gamma$ is an algebraic integer.

As per usual, we say that a hyperbolic surface (or hyperbolic 2-orbifold) S is semi-arithmetic when $S = \Gamma\backslash\mathbb{H}$ where $\Gamma < \mathrm{PSL}(2,\mathbb{R})$ is semi-arithmetic.

We gather below the first examples of semi-arithmetic Fuchsian groups. Example 5.4.6 is going to be of particular importance in the proof of the main results.

Example 5.4.4 (Arithmetic groups)**.** Every arithmetic Fuchsian group is, of course, a semi-arithmetic group.

Example 5.4.5 (Triangle groups)**.** Let Γ be a triangle group (see Example 4.3.22), i.e., a Fuchsian group generated by three elements T_1, T_2, T_3 satisfying:

$$T_1^m = T_2^n = T_3^l = T_1T_2T_3 = \mathrm{Id},$$

where the integers m, n, l are such that $2 \leq m, n, l \leq \infty$ and $1/m + 1/m + 1/l < 1$ (if an integer equals ∞ it means that the corresponding T_i is parabolic).

Using Proposition 5.2.9, an explicit description of the (invariant) trace field and quaternion algebra of Γ can be given. In [51], Takeuchi shows that:

$$\mathbb{Q}(\operatorname{tr}\Gamma) = \mathbb{Q}\left(\cos\frac{\pi}{m}, \cos\frac{\pi}{n}, \cos\frac{\pi}{l}\right),$$

and

$$\operatorname{tr}\Gamma = \mathbb{Z}\left[2\cos\frac{\pi}{m}, 2\cos\frac{\pi}{n}, 2\cos\frac{\pi}{l}\right].$$

In particular, $k\Gamma$ is totally real, as a subfield of $\mathbb{Q}(\operatorname{tr}\Gamma)$, and the trace of any element of Γ is an algebraic integer. Therefore, *every triangle groups is semi-arithmetic*.

Furthermore, Takeuchi's characterisation of arithmetic Fuchsian groups (see Theorem 5.3.12) gives a criterion to determine whether triangle groups are arithmetic. It turns out that only 85 triangle groups are arithmetic, as listed in [51]. Since there are infinitely many triangle groups, one obtains in this way an infinite family of *strictly semi-arithmetic* groups, i.e., non-arithmetic semi-arithmetic groups.

Example 5.4.6 (Trirectangle)**.** The *trirectangle* is a geodesic quadrilateral with three right angles and one acute angle φ. Let $\mathcal{Q}$ be a trirectangle with acute angle $\varphi = \pi/3$ and vertices labeled $F_1, \ldots, F_4$,, as shown in Figure 5.1.

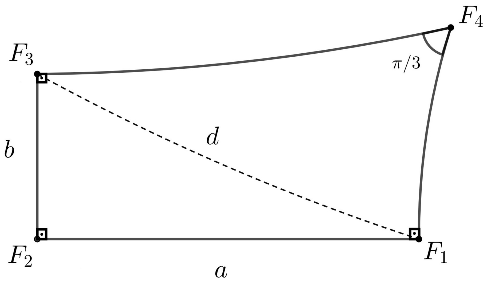

Figure 5.1: Trirectangle

Consider the abstract group

$$\Lambda = \langle s_1, \ldots, s_4 \mid s_1^2 = s_2^2 = s_3^2 = s_4^3 = s_1 \cdots s_4 = 1 \rangle. \tag{5.4.1}$$

Let $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ denote the reflections across the geodesic lines supporting, respectively, the sides $F_1F_2,\ F_2F_3,\ F_3F_4,\ F_4F_1$, and let $\mathrm{R}(\mathcal{Q})$ be the group of isometries of $\mathbb{H}$ generated by these reflections. According to Example 4.3.22, the index 2 subgroup of orientation-preserving isometries is a Fuchsian group with presentation given by (5.4.1). Moreover, the product of any two reflections across adjacent sides of $\mathcal{Q}$ gives an elliptic element in $\mathrm{PSL}(2,\mathbb{R})$ fixing the intersection point between the respective sides. In particular, the isometries defined by $S_j = \sigma_j\sigma_{j+1}$ for $j = 1, 2, 3, 4$ (where indices are taken modulo 4) are elliptic isometries fixing the vertex F_j, and the map $\rho_{\mathcal{Q}} : \Lambda \to \mathrm{PSL}(2,\mathbb{R})$ given by $\rho_{\mathcal{Q}}(s_i) = S_i$, $i = 1, 2, 3, 4$, induces an injective homomorphism onto the Fuchsian group in question.

The length a can be arbitrarily chosen while keeping the angle $\pi/3$ fixed. Furthermore, this length determines the other three sides of the trirectangle (see [8, Theorem 2.3.1]). Thus, we associate to each positive a a trirectangle $\mathscr{Q}_a$, uniquely determined up to isometry. The key fact is that the trace of the elements S_jS_{j+1} (indices mod 4) are given in terms of the lengths of the sides of $\mathscr{Q}_a$ which, in turn, are a function of a. So the idea is to select this length in such a way that the resulting group $\rho_{\mathscr{Q}_a}(\Lambda) < \mathrm{PSL}(2,\mathbb{R})$ will be semi-arithmetic. As we shall see later in more detail (Theorem 6.1.2), this is not only possible, but there is actually a dense set of parameters in $[0,+\infty)$ from where we can pick a and obtain a semi-arithmetic group $\rho_{\mathscr{Q}_a}(\Lambda)$.

Arithmetic Fuchsian groups were defined from quaternion algebras (Definition 5.3.9) and then characterised in terms of their invariant trace field (Theorem 5.3.12). Semi-arithmetic Fuchsian groups, on the other hand, were defined based on properties of their invariant trace field, and now we characterise them in the language of quaternion algebras.

In the proof of Theorem 5.3.12 we observed that the real places of $k\Gamma$ at which $A\Gamma$ is ramified are precisely those that map $\mathrm{tr}\,(\Gamma^{(2)})$ to a bounded set. Which indicates that, in order to define semi-arithmetic Fuchsian groups from quaternion algebras, we must allow algebras that are split at multiple infinite real places.

Let A be a quaternion algebra defined over the totally real number field k of degree n. Denote by $\sigma_1 = \mathrm{id}, \sigma_2, \dots, \sigma_r$ the embeddings of k into $\mathbb{R}$ at which A is split, and extend them to embeddings $\rho_1, \dots, \rho_r$ of A into $M_2(\mathbb{R})$. Let $\mathscr{O}$ be an order of A. From Theorem 5.3.2, we know that the k-homomorphism $\rho : A \to M_2(\mathbb{R}) \times \cdots \times M_2(\mathbb{R})$ defined by $x \mapsto (\rho_1(x), \dots \rho_r(x))$ maps $\mathscr{O}^1$ onto a cofinite discrete subgroup $\rho(\mathscr{O}^1)$ of $\mathrm{SL}(2,\mathbb{R}) \times \cdots \times \mathrm{SL}(2,\mathbb{R})$. In particular, it follows that $\rho(\mathscr{O}^1)$ acts on $(\mathbb{H})^r$ by componentwise Möbius transformations.

Definition 5.4.7. We say a subgroup Δ of (P)SL$(2,\mathbb{R})$ is an *arithmetic group acting on* $(\mathbb{H})^r$ if it is commensurable to some $(\mathrm{P})\rho_1(\mathscr{O}^1)$ as above. A finite index subgroup of $(\mathrm{P})\rho_1(\mathscr{O}^1)$ is said to be *derived from a quaternion algebra* (cf. Definition 5.3.9).

A subgroup G of a group derived from a quaternion algebra, then acts on $(\mathbb{H})^r$ as follows: for $g = \rho_1(x) \in G$,

$$g \cdot (z_1, \dots, z_r) = (\rho_1(x)z_1, \dots, \rho_r(x)z_r).$$

Note that, when $r = 1$, Δ is just an arithmetic Fuchsian group. If $r > 1$, on the other hand, then it follows from Theorem 5.3.2 that Δ is dense in $\mathrm{SL}(2,\mathbb{R})$.

Now, the traces of elements of $\rho_1(\mathscr{O}^1)$, being the reduced traces of elements of $\mathscr{O}^1$, are, in particular, algebraic integers of k. Let S be any subgroup of an arithmetic group Δ acting on $(\mathbb{H})^r$. Since the invariant trace field and the property of having integral traces are both commensurability invariants, it follows that kS is totally real and that every trace in $\operatorname{tr} S$ is an algebraic integer. In other words, if S is a Fuchsian group, then S is semi-arithmetic. This is a characterisation of semi-arithmetic Fuchsian groups, as the next theorem shows:

Theorem 5.4.8 ([49]). *A cofinite Fuchsian group Γ is semi-arithmetic if and only if Γ is commensurable to a subgroup S of an arithmetic group Δ acting on $(\mathbb{H}^2)^r$.*

Proof. Sufficiency was just established above. Now suppose Γ is a semi-arithmetic Fuchsian group. We know that $A\Gamma$ is a quaternion algebra over $k\Gamma$ and, since $\operatorname{tr} \Gamma^{(2)}$ is contained in the ring of algebraic integers, it follows from Proposition 5.1.10 that $\mathcal{O}\Gamma = \left\{\sum_i a_i\gamma_i \mid a_i \in \mathcal{O}_{k\Gamma}, \gamma \in \Gamma^{(2)}\right\}$ is an order in $A\Gamma$. Furthermore, the reduced norm of an element γ of $\Gamma^{(2)}$ equals to $\det \gamma = 1$, which means that $\Gamma^{(2)}$ is a subgroup of $\mathcal{O}\Gamma^1$. Since the inclusion of $A\Gamma$ in $M_2(\mathbb{R})$ is an embedding extending the identity embedding of k into $\mathbb{R}$, we find that $\Gamma^{(2)}$ is the subgroup of the arithmetic group $\mathcal{O}\Gamma^1$ acting on $(\mathbb{H}^2)^r$, and Γ is commensurable to $\Gamma^{(2)}$. □

Remark 5.4.9. When arithmetic subgroups of $\mathrm{PSL}(2,\mathbb{R})$ and $\mathrm{PSL}(2,\mathbb{C})$ were defined, being discrete and cofinite came as a consequence of Theorem 5.3.2. In the discussion above, we can see that discreteness and finite covolume does not always hold for subgroups of an arithmetic group acting on $(\mathbb{H}^2)^r$. Therefore, when defining semi-arithmetic groups, we must explicitly require them to be Fuchsian groups.

5.4.3 Properties of semi-arithmetic groups

In this subsection, we present some properties of semi-arithmetic Fuchsian groups that were studied in recent research ([13], [14], [21], [31] and [49]). In the course of this exposition, we also introduce the special subclass of semi-arithmetic groups admitting modular embedding. The lettered theorems indicate the original contributions of the present thesis. Chapter 6 will be devoted to proving Theorems A and B.

Congruence subgroups and systolic growth

Consider a quaternion algebra A over the totally real number field k, splitting over the identity embedding $k \hookrightarrow \mathbb{C}$, which gives the embedding of k-algebras $\rho_1 : A \to M_2(\mathbb{R})$. Let $\mathcal{O}$ be a maximal order of A. In particular, $\rho(\mathcal{O}^1) \subset \mathrm{SL}(2,\mathbb{R})$.

For any ideal $\mathfrak{a} \subset \mathscr{O}_K$, the set $\mathfrak{a} \cdot \mathscr{O} = \{\sum_j a_j\omega_j \mid a_j \in \mathfrak{a}, \omega_j \in \mathscr{O}\}$ is an ideal of the ring $\mathscr{O}$. The *principal congruence subgroup* of $\mathscr{O}^1$ of level $\mathfrak{a}$ is defined by

$$\mathscr{O}^1(\mathfrak{a}) = \{x \in \mathscr{O}^1 \mid x - 1 \in \mathfrak{a}\mathscr{O}\}.$$

We refer to [31] for a proof that $\mathscr{O}^1(\mathfrak{a})$ is a normal subgroup of $\mathscr{O}^1$ of finite index.

Now let $\Gamma < \mathrm{PSL}(2, \mathbb{R})$ be a semi-arithmetic Fuchsian group and denote by $\widetilde{\Gamma}$ its preimage in $\mathrm{SL}(2, \mathbb{R})$. According to Theorem 5.4.8, there exist A, k, ρ_1 and $\mathscr{O}$ as above such that $\mathrm{P}\rho_1(\mathscr{O}^1)$ contains $\Gamma^{(2)}$. For an ideal $\mathfrak{a}$ of $\mathscr{O}_k$ we define:

Definition 5.4.10. The *principal congruence subgroup* $\Gamma(\mathfrak{a})$ of level $\mathfrak{a}$ of Γ is the projection on $\mathrm{PSL}(2, \mathbb{R})$ of the intersection $\widetilde{\Gamma} \cap \rho(\mathscr{O}^1(\mathfrak{a}))$. A *congruence subgroup* of Γ is any subgroup containing a principal congruence subgroup.

Moreover, $\Gamma(\mathfrak{a})$ is a finite index subgroup of Γ. Indeed, the natural map

$$\Gamma^{(2)}/\Gamma^{(2)}(\mathfrak{a}) \to \mathrm{P}\rho(\mathscr{O}^1)/\mathrm{P}\rho(\mathscr{O}^1(\mathfrak{a})),$$

is well-defined and injective. The quotient on the right-hand side is finite, so $\Gamma^{(2)}(\mathfrak{a})$ has finite index in $\Gamma^{(2)}$ and hence in Γ. Since, $\Gamma^{(2)}(\mathfrak{a}) < \Gamma(\mathfrak{a}) < \Gamma$ it follows that $[\Gamma : \Gamma(\mathfrak{a})] < \infty$.

Note that if Γ is taken to be arithmetic, we can define congruence subgroups in the same way. As a matter of fact, if the algebra A in the above discussion is ramified over every non-trivial place of k, then any Fuchsian group that contains $\rho_1(\mathscr{O}^1(\mathfrak{a}))$ as a finite index subgroup is automatically (discrete and) arithmetic.

Let $S = \Gamma\backslash\mathbb{H}^2$, where Γ is a torsion-free arithmetic group with trace field $\mathbb{Q}$. P. Buser and P. Sarnak showed in [9] that the *principal congruence coverings* $S_m = \Gamma(m)\backslash\mathbb{H}$ of S satisfy the following logarithmic systolic growth:

$$\mathrm{sys}(S_m) \geq \frac{4}{3}\log(g(S_m)) - c,$$

where $g(S_m)$ denotes the genus of S_m, $\Gamma(m)$ denotes the principal congruence subgroup of level $(m) \subset \mathbb{Z}$ and c is a constant independent of m.

In [29], M. Katz, M. Schaps and U. Vishne extended this result to congruence coverings of *any* closed arithmetic Riemann surface.

Remark 5.4.11. For any closed Riemann surface S, a simple geometric argument gives that:

$$\mathrm{sys}(S) \leq 2\log(g(S)) + A,$$

where A is independent of S. In particular, it follows from the results discussed above that this logarithmic upper bound is optimal (up to the constants for arithmetic Riemann surfaces).

In [14], C. Dória shows that, for semi-arithmetic orbifolds, there also exists a sequence of congruence coverings with logarithmic systolic growth:

Theorem 5.4.12 ([14, Theorem 4.2])**.** *Let $\Gamma < \mathrm{PSL}(2,\mathbb{R})$ be a cocompact semi-arithmetic Fuchsian group and let k be its invariant trace field. Then, for infinitely many prime ideals $\mathfrak{p} \subset \mathcal{O}_k$, the corresponding principal congruence subgroups $\Gamma(\mathfrak{p}) < \Gamma$ are torsion-free and the closed Riemann surfaces $S_{\mathfrak{p}} = \Gamma(\mathfrak{p})\backslash\mathbb{H}$ satisfy*

$$\mathrm{sys}(S_{\mathfrak{p}}) \geq C \log(g(S_{\mathfrak{p}})) - c,$$

where $C > 0, c \in \mathbb{R}$ are constants that do not depend on $\mathfrak{p}$ and $g(S_{\mathfrak{p}})$ denotes the genus of $S_{\mathfrak{p}}$.

The multiplicative constant can be made explicit if one requires Γ to be in a more restrictive class of semi-arithmetic groups, which we introduce next.

Groups addmitting modular embedding

Definition 5.4.13 (Groups admitting modular embedding)**.** Let $\Gamma < \mathrm{PSL}(2,\mathbb{R})$ be a cofinite Fuchsian group such that Γ is contained in an arithmetic group Δ acting on $(\mathbb{H})^r$ (in fact, assume for simplicity that Δ is derived from a quaternion algebra as in Definition 5.4.7) and there exists an equivariant holomorphic embedding $F : \mathbb{H} \to (\mathbb{H})^r$, i.e., a holomorphic embedding F satisfying

$$F(\gamma z) = \gamma \cdot F(z),$$

for all $z \in \mathbb{H}$ and all $\gamma \in \Gamma$. In this case, we say that Γ *admits a modular embedding*.

Note that a group admitting modular embedding is, in particular, semi-arithmetic.

In [13], P. Cohen and J. Wolfart proved that all triangle groups admit modular embedding. In [49], Schaller and Wolfart describe infinite families of semi-arithmetic groups not admitting modular embedding by generating Fuchsian reflection groups from certain trirectangles and hyperbolic pentagons. In fact, they raise the question (see [49, Problem 1]) as to whether the only groups admitting modular embedding are arithmetic groups and finite index subgroups of triangle groups. R. Kucharczyk answered this question negatively in [31] by pointing out that some Veech groups are neither triangular nor arithmetic even though they do admit modular embedding. These examples are non-cocompact and it appears that the question remains open for cocompact semi-arithmetic groups admitting modular embedding.

Furthermore, Dória proved the following version of Theorem 5.4.12 for the case of surface-groups admitting modular embedding:

Theorem 5.4.14 ([14, Theorem 1.5]). *If $S = \Gamma\backslash\mathbb{H}$ is a semi-arithmetic Riemann surface such that Γ admits modular embedding (acting on $(\mathbb{H})^r$), then S admits a sequence of congruence coverings $S_i \to S$ of degree arbitrarily large satisfying*

$$\operatorname{sys}(S_i) \geq \frac{4}{3r}\log(g(S_i)) - c,$$

where the constant c does not depend on the i.

Note that, for arithmetic surfaces, $r = 1$ and the multiplicative constant given above reduces to the one previously known from [9] and [29].

Rigidity of semi-arithmetic groups

The famous Rigidity Theorem due to Mostow (cocompact case) and Prasad (non-cocompact case) states that an abstract isomorphism between two cofinite subgroups of $\mathrm{PSL}(2,\mathbb{C})$ must extend to a conjugation of $\mathrm{PSL}(2,\mathbb{C})$. It also holds in the isometry groups of higher dimensional hyperbolic spaces. In dimension 2, however, the absence of such rigidity is exactly what enables the rich theory of Teichmüller spaces.

Surprisingly, for semi-arithmetic Fuchsian groups admitting modular embeddings, Kucharczyk proves that there exists some rigidity, as long as the abstract isomorphism satisfies certain conditions:

Theorem 5.4.15 ([31, Theorem A]). *Let $\Gamma_1, \Gamma_2 < \mathrm{PSL}(2,\mathbb{R})$ be two semi-arithmetic Fuchsian groups which* virtually *admit modular embeddings (i.e., they each contain a finite index subgroup admitting modular embedding) and let $f : \Gamma_1 \to \Gamma_2$ be an isomorphism of abstract groups such that, for every subgroup $\Lambda < \Gamma_1$ of finite index, Λ is a congruence subgroup if and only if $f(\Lambda)$ is a congruence subgroup.*

Then f is a conjugation by some element $a \in \mathrm{PGL}(2,\mathbb{R})$. In particular, $\Gamma_2 = a\Gamma_1 a^{-1}$.

Thinness

In [21], S. Geninska describes the limit set of semi-arithmetic groups when acting on $(\mathbb{H})^r$. More generally, in [22], Geninska further investigates the limit set of infinite covolume subgroups of irreducible arithmetic lattices of $\mathrm{PSL}(2,\mathbb{C})^q \times \mathrm{PSL}(2,\mathbb{R})^r$. The results therein are beyond the scope of this thesis. We briefly point out here that strictly semi-arithmetic groups are *thin* when embedded in the appropriate ambient group ([31, Corollary 7.2]). Indeed, let Γ be a strictly semi-arithmetic group with invariant trace field $k = k\Gamma$ and invariant quaternion algebra $A = A\Gamma$. In the notation of the proof of Theorem 5.4.8, the group $\mathcal{O}\Gamma^1$ is derived from a quaternion algebra. The embedding $\rho : \mathcal{O}\Gamma^1 \to \mathrm{SL}(2,\mathbb{R})^r$, $r \geq 2$, descends to an embedding of $\mathrm{P}\mathscr{O}\Gamma^1$ into $\mathrm{PSL}(2,\mathbb{R})^r$ (which we shall also denote by ρ). Note that $\Gamma^{(2)}$ is a non-elementary subgroup of $\mathrm{P}\mathscr{O}\Gamma^1$. Moreover, $k = \mathbb{Q}(\mathrm{tr}\,\Gamma^{(2)})$ coincides with the invariant trace field of $\mathrm{P}\mathscr{O}\Gamma^1$, by construction. It then follows from [21, Corollary 2.2] that the group $\rho(\Gamma^{(2)})$ is Zariski-dense in $\mathrm{PSL}(2,\mathbb{R})^r$. On the other hand, we know $\Gamma^{(2)}$ to be of infinite covolume in $\mathrm{PSL}(2,\mathbb{R})^r$ since it is non-arithmetic ($r \geq 2$), proving that $\rho(\Gamma^{(2)})$ is thin.

Remark 5.4.16. Strictly speaking, we have not proved that Γ is a thin subgroup of $\mathrm{PSL}(2,\mathbb{R})^r$ (or even a subgroup, for that matter). We leave this discussion slightly informal and simply mention that ρ may be extended to an embedding of Γ into $\mathrm{PSL}(2,\mathbb{R})^r$, although we shall not pursue this any further in here.

Semi-arithmetic points of Teichmüller spaces

It has been observed before that a finite index subgroup $\Gamma' < \Gamma$ has integral traces if and only if Γ does. Since the invariant trace field is invariant under commensurability, it follows from the definitions that being (semi-)arithmetic is invariant under commensurability. It is also clearly invariant under conjugation, and hence it is invariant under commensurability in the wide sense.

In particular, if $\phi \in \mathfrak{R}(\Gamma)$ is a representation (see §§4.3.4) such that $\phi(\Gamma)$ is an arithmetic (resp. a semi-arithmetic) Fuchsian group, then this is also true for every representation of Γ that is equivalent to ϕ under the action of $\mathrm{PGL}(2,\mathbb{R})$ on $\mathfrak{R}(\Gamma)$ by conjugation. It thus makes sense to say that the point $[\phi]$ in the Teichmüller space $\mathrm{Teich}(\Gamma)$ is arithmetic (resp. semi-arithmetic).

In [7, Theorem 8.2], A. Borel proves that, for each $C > 0$, there are at most finitely many groups $\Gamma_1, \ldots, \Gamma_{n(C)}$ in $\mathrm{PSL}(2,\mathbb{R})$ such that any arithmetic group in $\mathrm{PSL}(2,\mathbb{R})$ with coarea $\leq C$ is conjugate to one of the Γ_i, $1 \leq i \leq n(C)$. In particular, there are at most finitely many arithmetic points in each Teichmüller space. The situation is drastically different for semi-arithmetic points, as will become clear from Theorem A below.

Given a closed topological surface S_g of genus $g \geq 2$ and a homotopically nontrival closed curve $\alpha \subset S_g$, we define the corresponding *length function* $\ell_\alpha : \mathrm{T}_g \to \mathbb{R}$, that associates to each Riemann surface X in T_g the length $\ell_\alpha(X)$ of the unique closed geodesic on X freely homotopic to α. Recall that T_g is the space of all equivalence classes of marked Riemann surfaces of genus g or, equivalently, the space hyperbolic metrics on S_g up to isometries isotopic to the identity (see Remark 4.3.25). We then have the following result:

Theorem A. *For any $g \geq 2$ there exists a length function $\ell_\alpha : \mathrm{T}_g \to \mathbb{R}$ such that*

$$\{\ell_\alpha(S) \mid S \in \mathrm{T}_g \text{ is semi-arithmetic}\}$$

is dense on the set of positive real numbers.

Recall that the *systole* of a closed hyperbolic surface S, denoted $\mathrm{sys}(S)$, is defined to be the minimum length of a closed geodesic on S. Note that $\mathrm{sys}(S) > 0$.

It follows immediately from Theorem A that, for any genus $g \geq 2$, one can find a sequence of closed semi-arithmetic surfaces of genus g with systole approaching 0. In particular, there are infinitely many semi-arithmetic surfaces in every Teichmüller space T_g.

In fact, if we let the genus vary, the set of systoles of semi-arithmetic surfaces is dense in the real line. Indeed, in [14, Theorem 1.2], the following result is proved:

Theorem 5.4.17. *The set* $\{\mathrm{sys}(S) \mid S$ *is a closed semi-arithmetic Riemann surface*$\}$ *is dense in the positive real numbers.*

Remark 5.4.18. The Teichmüller space $\mathrm{Teich}(\Gamma)$ is known to be parametrised by finitely many trace functions (see, for example, [40], [41], [43]). More precisely, let $\gamma \in \Gamma$ and define the *trace function* $\mathrm{tr}_{\gamma} : \mathrm{Teich}(\Gamma) \to \mathbb{R}$ as $\mathrm{tr}_{\gamma}([\phi]) = |\mathrm{tr}\,(\phi(\gamma))|$. Then there exist $N > 0$ and $\gamma_1, \ldots, \gamma_N \in \Gamma$ such that $(\mathrm{tr}_{\gamma_1}, \ldots, \mathrm{tr}_{\gamma_N}) : \mathrm{Teich}(\Gamma) \to \mathbb{R}^N$ is an (real-analytic) embedding. It then follows from Definition 5.4.3 that the semi-arithmetic points in $\mathrm{Teich}(\Gamma)$ are in one-to-one correspondence with a subset of the N-tuples with algebraic integer coordinates. Therefore, we conclude that there are at most countably many semi-arithmetic Fuchsian groups.

Another consequence of Theorem A shows that, for any fixed genus $g \geq 2$, infinitely many number fields are realised as the invariant trace field of a semi-arithmetic Fuchsian group of genus g:

Theorem B. *Every totally real number field of prime degree at least 3 is realised as the invariant trace field of a genus g semi-arithmetic Riemann surface, for any $g \geq 2$.*

Remark 5.4.19. In particular, there are semi-arithmetic Riemann surfaces of genus g with invariant trace fields of arbitrarily large degree.

This gives a negative answer to a conjecture made by B. Jeon (see [25, Conjectrue 2]):

Conjecture (Jeon). *For each $g \geq 2$ there exists only a finite number of real number fields and quaternion algebras that are realised as the invariant trace field and invariant quaternion algebra of a hyperbolic structure on S_g with integral traces.*

Automorphisms of surfaces with non-integral traces

Let Γ be a irreducible Fuchsian group and let $\mathrm{N}(\Gamma) = \{g \in \mathrm{PSL}(2,\mathbb{R}) \mid g\Gamma g^{-1} = \Gamma\}$ be its normaliser. If $\mathrm{N}(\Gamma)$ is not discrete, we can find a sequence of elements $\{g_n\}$ in $\mathrm{N}(\Gamma)$ approaching the identity Id. For any two elements γ_1 and γ_2 of Γ, we have that $g_n\gamma_i g_n^{-1}$ approach γ_i, $i = 1, 2$. Discreteness of Γ implies that both sequences must be constant eventually, which means that, for n sufficiently large, g_n commutes with γ_1 and with γ_2. In particular, γ_2 fixes the same points as g_n, which are, in turn, the same points fixed by γ_1. Since γ_1, γ_2 were arbitrary, we conclude that every element in Γ fixes the same point, contradicting the assumption that Γ is irreducible. We thus conclude that $\mathrm{N}(\Gamma)$ must also be a Fuchsian group.

Theorem 4.3.17 implies that the smallest coarea of a Fuchsian group is $\pi/21$, realised by the triangle group of signature $(0; 2, 3, 7)$. As a consequence, we obtain the famous Hurwitz bound on the cardinality of the automorphism group of a compact Riemann surface X_g of genus g. Indeed, let $X_g = \Gamma\backslash\mathbb{H}$. It is not hard to see that the automorphism group $\mathrm{Aut}(X_g)$ is isomorphic to $\mathrm{N}(\Gamma)/\Gamma$. The Riemann-Hurwitz Theorem (see Corollary 4.3.15) together with the above observation yields:

$$|\mathrm{N}(\Gamma)/\Gamma| = [\mathrm{N}(\Gamma) : \Gamma] = \frac{\mathrm{area}(\Gamma\backslash\mathbb{H})}{\mathrm{area}(\mathrm{N}(\Gamma)\backslash\mathbb{H})} \leq \frac{4\pi(g-1)}{\pi/21} = 84(g-1). \tag{5.4.2}$$

There are surfaces realising this bound with arbitrarily large genera.

In [3], M. Belolipetsky showed that, for non-arithmetic surfaces, the bound on their automorphism groups drops to $\frac{156}{7}(g-1)$. This bound is obtained by founding the minimal coarea of a non-arithmetic Fuchsian group, which is $7\pi/39$, the coarea of the triangle group $(0; 2, 3, 13)$. Moreover, this bound is attained in infinitely many genera.

Following the same idea, we can ask what would be the maximal cardinality of the automorphism group of a non-semi-arithmetic surface $X_g = \Gamma\backslash\mathbb{H}$ of genus g. To answer this, we must investigate the minimal coarea of a non-semi-arithmetic Fuchsian group. We know that every triangle group is semi-arithmetic. Excluding those, the signature providing the smallest coarea would be the one of the trirectangle group $(0; 2, 2, 2, 3)$, studied in Example 5.4.6. The Teichmüller space corresponding to this signature has real dimension 2 so, by a cardinality argument, there must be non-semi-arithmetic groups with this signature (see Remark 5.4.18). Alternatively, in the construction described in Example 5.4.6, one can pick a such that $\cosh(a)$ is transcendental. Now, the coarea of one such group is $\pi/3$ and inserting this into (5.4.2) leads to the following bound:

$$|\mathrm{N}(\Gamma)/\Gamma| \leq \frac{4\pi(g-1)}{\pi/3} = 12(g-1).$$

Furthermore, this bound is attained for every genus $g \geq 2$. Indeed, fix a non-semi-arithmetic trirectangular group $\Lambda = \Lambda_a$ by picking, for example, a such that $\cosh(a)$ is transcendental, as suggested above. From the construction in the proof of Theorem A, which will be presented in Chapter 6, it follows that, for every $g \geq 2$, there exists a genus g surface group Γ_g that is a normal subgroup of Λ. It follows that $\Lambda < \mathrm{N}(\Gamma_g)$ so that $\mathrm{area}(\mathrm{N}(\Gamma_g)\backslash\mathbb{H}) \leq \mathrm{area}(\Lambda\backslash\mathbb{H}) = \pi/3$. On the other hand, since $\pi/3$ is the minimal coarea for a non-semi-arithmetic group, we conclude that $\mathrm{area}(\mathrm{N}(\Gamma_g)\backslash\mathbb{H}) = \pi/3$, $\mathrm{N}(\Gamma_g) = \Lambda$ and $|\mathrm{N}(\Gamma_g)/\Gamma_g| = 12(g-1)$. We have, thus, proved that:

Theorem C. *The order of the automorphism group of a non-semi-arithmetic Riemann surface X_g of genus $g \geq 2$ satisfies the following bound:*

$$|\mathrm{Aut}(X_g)| \leq 12(g-1).$$

Moreover, this bound is attained in every genus $g \geq 2$.

CHAPTER 6

PROOFS OF THEOREMS A AND B

This chapter is devoted to the proof of the main theorems of this thesis. We will show that the set of numbers realised as the length of a closed geodesic in semi-arithmetic surfaces of some fixed genus is dense in the positive real numbers. More precisely, we describe a dense set $\mathscr{L} \subset [0, +\infty)$ such that, for any integer $g \geq 2$ and $l \in \mathscr{L}$, there exists a semi-arithmetic surface S of genus g and a closed geodesic γ in S with length $\ell(\gamma) = l$. This leads to Theorem A. Theorem B is derived as an application of Theorem A. Finally, in Section 6.2, we describe the Reidemeister-Schreier rewriting process used in the proof of Theorem A.

Let us first recall some tools:

Lemma 6.0.1. *Let G be a 1-dimensional Lie group with finitely many connected components. If $H < G$ has rank ≥ 2 then H is dense in G*

Proof. Assume, without loss of generality, that G is connected. If G is non-compact then it is Lie group-isomorphic to $(\mathbb{R}, +)$ by an isomorphism, say, f. It is known that additive subgroups of $\mathbb{R}$ are either cyclic infinite or dense. Since $f(H)$ has rank ≥ 2, it must be dense in $\mathbb{R}$ in which case H is dense in G. If G is compact, it will be isomorphic to $\mathbb{S}^1$, where a similar dichotomy holds. □

Lemma 6.0.1 together with the Dirichlet's Unit Theorem (Theorem 2.2.31) give us the following corollary.

Corollary 6.0.2. *Let K be a totally real number field such that $[K : \mathbb{Q}] \geq 3$. Then $\mathcal{O}_K^\times$ is dense in $\mathbb{R}$.*

Next, we list some formulae in classical hyperbolic geometry that will be used later on. For a systematic treatment of the subject, see, for example, [2], [8] or [19].

Lemma 6.0.3 ([2, Theorems 7.11.1 and 7.17.1]). *Consider a geodesic quadrilateral, known as* trirectangle, *with three right angles and one acute angle φ, with side lengths indicated as in Figure 6.1. Then the following hold:*

$$\cos\varphi = \sinh a \, \sinh b\,; \tag{6.0.1}$$

$$\cosh d = \cosh a \, \cosh b. \tag{6.0.2}$$

Note that equation (6.0.2) *is but the hyperbolic Pythagoras' Theorem.*

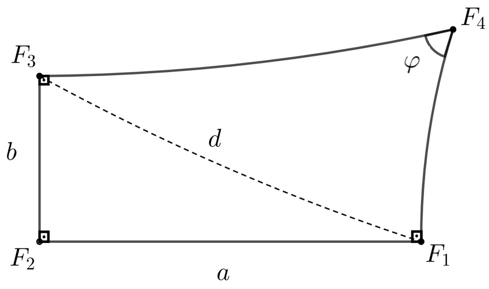

Figure 6.1: Trirectangle

Lemma 6.0.4 ([2, Theorem 7.19.2]). *For any three positive real numbers a_1, a_2, a_3, there exists a convex right-angled hexagon with three non-adjacent sides of length a_1, a_2, a_3.*

Moreover, this hexagon is unique (up to isometry). Indeed, if the side of length a_1 is opposed by a side of length, say, b_1, then the following equation holds:

$$\cosh b_1 \, \sinh a_2 \, \sinh a_3 = \cosh a_1 + \cosh a_2 \, \cosh a_3.$$

In other words, the lengths of all sides of the hexagon are determined by the lengths of three non-adjacent sides.

Finally, we recall the following (vital) relation between the displacement of a hyperbolic element and its trace.

Proposition 6.0.5. *Let $\gamma \in \mathrm{PSL}(2,\mathbb{R})$ be a hyperbolic element with translation length $\ell(\gamma)$. Then the following relation holds:*

$$|\operatorname{tr}\gamma| = 2\cosh\frac{\ell(\gamma)}{2}.$$

Proof. Both sides are invariant under conjugation, so we can assume γ to be $\gamma(z) = \lambda z$, $\lambda > 1$, translating along the imaginary axis. Then $\ell(\gamma) = d_{\mathbb{H}}(i, \lambda i) = \ln\lambda$, which gives $|\operatorname{tr}\gamma| = \sqrt{\lambda} + 1/\sqrt{\lambda} = e^{\ell(\gamma)/2} + e^{-\ell(\gamma)/2}$ and the result follows. □

Corollary 6.0.6 (cf. [2, Theorem 7.38.2]). *If $\gamma_1, \gamma_2 \in \mathrm{PSL}(2,\mathbb{R})$ are half-turns around p_1 and p_2, then*

$$\operatorname{tr} \gamma_2\gamma_1 = 2\cosh d_{\mathbb{H}}(p_1, p_2),$$

where $d_{\mathbb{H}}$ is the hyperbolic distance.

Proof. Let L be the geodesic line connecting p_1 to p_2 and let L_i be the geodesic line orthogonal to L at p_i, $i = 1, 2$. If σ, σ_i denote hyperbolic reflection across L, L_i, respectively, then $\gamma_1 = \sigma\sigma_1$ and $\gamma_2 = \sigma_2\sigma$, whence $\gamma_2\gamma_1 = \sigma_2\sigma_1$, which is easily seen to be a hyperbolic translation along L (in the direction from p_1 to p_2) with translation length $2d_{\mathbb{H}}(p_1, p_2)$. The result then follows from Proposition 6.0.5. □

6.1 Construction

Consider the abstract group

$$\Lambda = \langle s_1, \ldots, s_4 \mid s_1^2 = s_2^2 = s_3^2 = s_4^3 = s_1 \cdots s_4 = 1\rangle. \tag{6.1.1}$$

Just as we did in Example 5.4.6, we will realise Λ as the Fuchsian group generated by a trirectangle. We repeat the process here, for the convenience of the reader. Let $\mathscr{Q}$ be a trirectangle with acute angle measuring $\varphi = \pi/3$ and with vertices labeled $F_1, \ldots, F_4$, as shown in Figure 6.1. Let $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ denote the reflections across the geodesic lines supporting, respectively, the sides $F_1F_2,\ F_2F_3,\ F_3F_4,\ F_4F_1$, and let $\mathrm{R}(\mathscr{Q})$ be the group of isometries of $\mathbb{H}$ generated by these reflections. The index 2 subgroup of orientation-preserving isometries is a Fuchsian group with presentation given by (6.1.1) (cf. Example 4.3.22). Moreover, the product of any two reflections across adjacent sides of $\mathscr{Q}$ gives an elliptic element in $\mathrm{PSL}(2,\mathbb{R})$ fixing the intersection point between the respective sides. In particular, the isometries defined by $S_j = \sigma_j\sigma_{j+1}$ for $j = 1, 2, 3, 4$ (where indices are taken modulo 4) are elliptic isometries fixing the vertex F_j, and the map $\rho_{\mathscr{Q}} : \Lambda \to \mathrm{PSL}(2,\mathbb{R})$ given by $\rho_{\mathscr{Q}}(s_i) = S_i$, $i = 1, 2, 3, 4$, induces an injective homomorphism onto the Fuchsian group in question.

Consider the right-angled hexagon $\mathscr{H}$ that is tiled by six copies of one such trirectangle, as in Figure 6.2. One may verify that the isometries

$$\sigma_1\,,\ \sigma_2\,,\ \sigma_3\sigma_1\sigma_3\,,\ (\sigma_4\sigma_3\sigma_4)\sigma_2(\sigma_4\sigma_3\sigma_4)\,,\ \sigma_4(\sigma_3\sigma_1\sigma_3)\sigma_4\,,\ \sigma_4\sigma_2\sigma_4$$

are reflections across the geodesic lines supporting the respective sides of the hexagon. It follows that the group $\mathrm{R}(\mathscr{H})$, generated by reflections across the sides of $\mathscr{H}$, is a subgroup of $\mathrm{R}(\mathscr{Q})$. Let C_j denote the half-turn around the vertex E_j of

$\mathscr{H}$, i.e., the product of the reflections across the sides of $\mathscr{H}$ intersecting at E_j. Let Γ be the abstract group with presentation

$$\Gamma = \langle c_1, \dots, c_6 \mid c_1^2 = \cdots = c_6^2 = c_1 \cdots c_6 = 1 \rangle. \tag{6.1.2}$$

In the same spirit as before, we define a map $\rho_{\mathscr{H}}$ taking c_j to C_j and obtain an injective representation of Γ into $\mathrm{PSL}(2, \mathbb{R})$ such that

$$\rho_{\mathscr{H}}(\Gamma) < \rho_{\mathscr{Q}}(\Lambda). \tag{6.1.3}$$

Furthermore, we observe that $\rho_{\mathscr{H}}(\Gamma)$ has index 6 in $\rho_{\mathscr{Q}}(\Lambda)$.

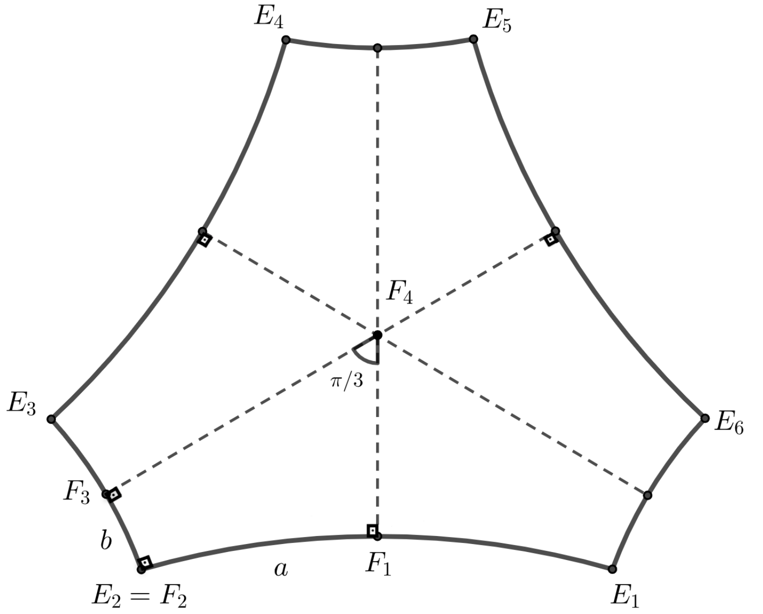

Figure 6.2: Right-angled hexagon

Definition 6.1.1. We say that a real number $t \geq 0$ is *realised* by the Fuchsian group Γ if there exists some element $\gamma \in \Gamma$ such that $|\mathrm{tr}\,\gamma| = t$. Similarly, we say that $l > 0$ is realised by the hyperbolic surface S if there exists some closed geodesic in S of length l. Note that $S = \mathbb{H}/\Gamma$ realises $l > 0$ if and only if Γ realises $2\cosh(l/2) > 2$. Indeed, recall that a closed geodesic in S is the projection of the axis of a hyperbolic element of Γ with translation length equal to the length this geodesic, then use Proposition 6.0.5.

Theorem 6.1.2. *The set $\mathscr{T}$, of all real numbers that are realised by a semi-arithmetic Fuchsian group of signature $(0; 2, 2, 2, 2, 2, 2)$, is dense in the interval $[2, +\infty)$.*

Proof. Consider a trirectangle $\mathscr{Q}_a$ with angle $\pi/3$ at the vertex F_4, and side F_1F_2 of length a, as shown in Figure 6.1. Let $\mathscr{H}_{2a}$ denote the corresponding right-angled hexagon tiled by $\mathscr{Q}_a$, as constructed above, and note that the side E_1E_2 has length $2a$.

By Theorem 2.3.1 in [8], the length a determines the other three sides of the trirectangle. Note that, by an argument of continuity, a can be arbitrarily chosen while keeping the acute angle with a fixed measure of $\pi/3$. We will select this length in such a way that the resulting group $\rho_{\mathscr{Q}_a}(\Lambda) < \mathrm{PSL}(2,\mathbb{R})$ will be semi-arithmetic and, as a consequence, so will be the finite index subgroup $\rho_{\mathscr{H}_{2a}}(\Gamma)$.

Let K be any totally real number field with $[K:\mathbb{Q}] \geq 3$ and let $V = \{v \in \mathbb{R}_{>0} \mid \sinh v \in \mathscr{O}_K^\times\}$. So, in particular, for every $v \in V$, $\sinh v$ is a totally real unit in the ring of integers of K. It follows from Corollary 6.0.2 that V is dense in $[0,\infty)$. Choose $a \in V$.

We first observe that $\cosh^2 a = \sinh^2 a + 1$ is a totally positive algebraic integer (i.e., an algebraic integer such that all its Galois conjugates are positive real numbers). It then follows that $\cosh a$ is a totally real algebraic integer.

Equation (6.0.1) gives that $\sinh b = \frac{(\sinh a)^{-1}}{2}$ and then

$$\cosh^2 b = \sinh^2 b + 1 = \frac{(\sinh a)^{-2} + 4}{4}.$$

By the same reasoning as before, we obtain $2\cosh b = \sqrt{(\sinh a)^{-2} + 4}$ is a totally real algebraic integer. Now, by the Pythagoras' Theorem, we have that

$$2\cosh d = \cosh a(2\cosh b)$$

is also a totally real algebraic integer.

Thus, by Corollary 6.0.6, the traces $|\mathrm{tr}\,(S_1S_2)|$, $|\mathrm{tr}\,(S_2S_3)|$, and $|\mathrm{tr}\,(S_1S_3)|$ are all totally real algebraic integers. Note also that the order 2 elements S_1, S_2, S_3 have trace 0, and that the order three element S_4 has trace 1 so, in view of Proposition 5.2.9 and (6.1.3):

$$\mathrm{tr}\,(\rho_{\mathscr{H}_{2a}}(\Gamma)) \subset \mathrm{tr}\,(\rho_{\mathscr{Q}_a}(\Lambda)) \subset \mathbb{Z}[\mathrm{tr}\,(S_1S_2), \mathrm{tr}\,(S_2S_3), \mathrm{tr}\,(S_1S_3)].$$

In particular, the invariant trace field of $\rho_{\mathscr{H}_{2a}}(\Gamma)$ is a subfield of

$$\mathbb{Q}(\mathrm{tr}\,(S_1S_2), \mathrm{tr}\,(S_2S_3), \mathrm{tr}\,(S_1S_3)),$$

which is, by construction, totally real. We therefore conclude that the Fuchsian group $\rho_{\mathscr{H}_{2a}}(\Gamma)$ is semi-arithmetic. Since a was arbitrarily chosen among a dense subset V of $[0,+\infty)$, the group $\rho_{\mathscr{H}_{2a}}(\Gamma)$ realises $t = 2\cosh 2a = 2 + 4\sinh^2 a$ as $\mathrm{tr}\,\rho_{\mathscr{H}_{2a}}(c_1c_2)$ for any $t = t(a)$ in $\mathscr{T} := 2\cosh 2V$, the latter being dense in $[2,+\infty)$. □

A *surface-kernel epimorphism* is an epimorphism $\theta\colon \Gamma \to G$ with torsion-free kernel. This is the case, for example, if θ preserves the order of the torsion elements of Γ. Note that it is sufficient to check if the orders of the generators of Γ are preserved (where Γ is still the group defined in (6.1.2)).

Theorem 6.1.3. *Let $\mathscr{L}_g$ be the set of real numbers that are realised by a semi-arithmetic surface of genus g. Then $\bigcap\limits_{g\geq 2} \mathscr{L}_g$ is dense in $[0, +\infty)$.*

Proof. Let $\mathscr{T}$ be the set obtained in Theorem 6.1.2. For each $t \in \mathscr{T}$ define $\Gamma_t = \rho_t(\Gamma) := \rho_{\mathscr{H}_{2a}}(\Gamma)$ where a is such that $2\cosh 2a = t$, as in the construction of the previous theorem. From what was proved, it follows that the Fuchsian group Γ_t is a semi-arithmetic group of signature $(0; 2, 2, 2, 2, 2, 2)$ with $\operatorname{tr}\rho_t(c_1c_2) = t$.

Case 1: $g = 2$

Let $\theta : \Gamma \to \mathbb{Z}/2\mathbb{Z}$ be a homomorphism defined by $c_i \mapsto \overline{1}$, $i = 1, \dots, 6$. It is immediate that θ is a surface-kernel epimorphism, since it preserves the order of the generators of Γ. It follows that $\mathscr{K} := \ker\theta$ is a torsion-free index 2 subgroup of Γ and, as such, it must be a surface group of genus g. From Theorem 4.3.17 and the Riemann-Hurwitz formula (Corollary 4.3.15), we compute that

$$2\pi(2g-2) = \operatorname{area}(\rho_t(\mathscr{K})\backslash\mathbb{H}) = [\Gamma_t : \rho_t(\mathscr{K})] \cdot \operatorname{area}(\Gamma_t\backslash\mathbb{H}) = 2 \cdot 2\pi,$$

and so $g = 2$. Moreover, $\theta(c_1c_2) = \overline{1} + \overline{1} = \overline{0}$ and thus $c_1c_2 \in \mathscr{K}$.

The quotient, $\rho_t(\mathscr{K})\backslash\mathbb{H}$, is a closed hyperbolic surface of genus 2. The axis of the hyperbolic element $\rho_t(c_1c_2)$ projects onto a closed geodesic γ_t in $\rho_t(\mathscr{K})\backslash\mathbb{H}$ that satisfies $2\cosh(\ell(\gamma_t)/2) = t$. If we let t vary in the set $\mathscr{T}$, then $\ell(\gamma_t) = 2\cosh^{-1}(t/2)$ covers a dense subset of $[0, +\infty)$.

Case 2: $g > 2$

For larger genera we proceed similarly. Using the Reidemeister-Schreier rewriting process (see §6.2 for the details) we can find an isomorphism Φ between $\langle x, y, x', y' \mid [x, y][x', y']\rangle$ and $\mathscr{K}$ such that $\Phi(y) = c_1c_2$. We will mildly abuse notation and say that

$$\mathscr{K} = \langle x, y, x', y' \mid [x, y][x', y']\rangle \text{ and } y = c_1c_2.$$

For each $n > 1$ we define the homomorphism $\eta_n\colon \mathscr{K} \to \mathbb{Z}/n\mathbb{Z}$ given by

$$\begin{aligned} \eta_n(x) &= \eta_n(x') = \eta_n(y') = \overline{1}, \\ \eta_n(y) &= \overline{0}. \end{aligned}$$

Therefore, for all $n > 1$ we have: η_n is an epimorphism, $y \in \mathscr{K}_n := \ker \eta_n$ and $\mathscr{K}_n$ is a surface group of genus $n+1$, since $[\mathscr{K} : \mathscr{K}_n] = n$. Moreover, the geodesic γ_t obtained in Case 1, lifts as a closed geodesic of the same length for all coverings $\rho_t(\mathscr{K}_n)\backslash\mathbb{H}$ of $\rho_t(\mathscr{K})\backslash\mathbb{H}$. Thus, the same conclusion as in the genus 2 case holds for any genus $g > 2$. □

Theorem A. *For any $g \geq 2$ there exists a length function $\ell_\alpha : \mathrm{T}_g \to \mathbb{R}$ such that*

$$\{\ell_\alpha(S) \mid S \in \mathrm{T}_g \textit{ is semi-arithmetic}\}$$

is dense on the set of positive real numbers.

Proof. This follows straight from (the proof of) Theorem 6.1.3 (see Remark 4.3.25). Indeed, the axis of the hyperbolic element $\rho_t(c_1c_2)$ projects onto a closed geodesic in $\rho_t(\mathscr{K}_n)\backslash\mathbb{H}$, whose free homotopy class induces the length function ℓ with the desired properties. □

Next, we show how to realise infinitely many number fields as the invariant trace field of a semi-arithmetic Fuchsian group with a fixed genus $g \geq 2$. The idea is essentially contained in Theorem 6.1.2 and Theorem 6.1.3.

Theorem B. *Every totally real number field of prime degree at least 3 is realised as the invariant trace field of a genus g semi-arithmetic Riemann surface, for any $g \geq 2$.*

Proof. Let K be a totally real number field of prime degree $p \geq 3$ and a a positive real number such that $\sinh a \in \mathscr{O}_K^\times$. For each $g \geq 2$, we can find a semi-arithmetic genus g Riemann surface whose uniformising Fuchsian group Δ realises $2\cosh 2a = 2 + 4\sinh^2 a$, i.e., there exists some $\gamma \in \Delta$ with $\operatorname{tr}\gamma = 2 + 4\sinh^2 a$. Note that a may be chosen in such a way that $\operatorname{tr}^2\gamma$ is not a rational number so that $k\Delta$ is strictly larger than $\mathbb{Q}$. Since $[k\Delta : \mathbb{Q}]$ divides the prime number $[K : \mathbb{Q}]$, we conclude that $k\Delta = K$. □

6.2 Reidemeister-Schreier rewriting process

In this section we use the Reidemeister-Schreier rewriting process in order to give a standard presentation for the group $\mathscr{K} \trianglelefteq \Gamma$ in terms of the generators of Γ, as defined in Section 6.1. For more information on the Reidemeister-Schreier rewriting process we refer the reader to [6, §2.9]. Recall that

$$\Gamma = \langle c_1, \ldots, c_6 \mid c_1^2 = \cdots = c_6^2 = c_1 \cdots c_6 = 1 \rangle.$$

Let $\theta \colon \Gamma \to \mathbb{Z}/2\mathbb{Z}$ be the epimorphism defined by $c_i \mapsto 1$, $i = 1, \ldots, 6$ and define $\mathscr{K} := \ker\theta \trianglelefteq \Gamma$.

Let $\phi\colon F_6 = F(x_1,\dots,x_6) \to \Gamma$ be an epimorphism from the free group of rank 6 with generators $\{x_1,\dots,x_6\}$ onto Γ given by $\phi(x_j) = c_j$, $j = 1,\dots,6$. Finally, let $\widetilde{\mathscr{K}}$ be the pre-image of $\mathscr{K}$ with respect to ϕ. Since $[\Gamma : \mathscr{K}] = 2$ and ϕ is surjective, it follows that $\widetilde{\mathscr{K}}$ also has index 2 in F_6. We pick the set $\mathfrak{T} = \{1, x_1\}$ as a Schreier transversal for $\widetilde{\mathscr{K}}$ in F_6. As usual, for any $g \in F_6$ we denote by $\overline{g}$ the (unique) element in $\mathfrak{T}$ with the property that $\widetilde{\mathscr{K}}g = \widetilde{\mathscr{K}}\overline{g}$. Then, by the Reidemeister-Schreier rewriting process process, the following elements generate the free group $\widetilde{\mathscr{K}}$:

$$\begin{aligned} 1 \cdot x_i \cdot (\overline{1 \cdot x_i})^{-1} &= x_i x_1^{-1}, \quad i = 2,\dots,6;\\ x_1 x_j \cdot (\overline{x_1 x_j})^{-1} &= x_1 x_j, \quad j = 1,\dots,6. \end{aligned}$$

Let us rename this generators as:

$$y_j = x_1 x_j,\ j = 1,\dots,6 \quad \text{and} \quad y_{5+i} = x_i x_1^{-1},\ i = 2\dots,6. \tag{6.2.1}$$

In order to find the defining relations we rewrite the words trt^{-1}, where $t \in \{1, x_1\}$ and $r \in \{x_1^2,\dots,x_6^2,\ x_1\cdots x_6\}$, in terms of $\{y_1,\dots,y_{11}\}$:

$$x_j^2 = (x_j x_1^{-1})(x_1 x_j) = \begin{cases} y_1, & \text{if } j = 1;\\ y_{5+j}\, y_j, & \text{if } j = 2,\dots,6; \end{cases}$$

$$x_1\cdots x_6 = (x_1x_2)(x_3x_1^{-1})(x_1x_4)(x_5x_1^{-1})(x_1x_6) = y_2\, y_8\, y_4\, y_{10}\, y_6\,;$$

$$x_1 x_j^2 x_1^{-1} = (x_1 x_j)(x_j x_1^{-1}) = \begin{cases} y_1, & \text{if } j = 1;\\ y_j\, y_{5+j}, & \text{if } j = 2,\dots,6; \end{cases}$$

$$x_1(x_1\cdots x_6)x_1^{-1} = x_1^2(x_2x_1^{-1})(x_1x_3)(x_4x_1^{-1})(x_1x_5)(x_6x_1^{-1}) = y_1\, y_7\, y_3\, y_9\, y_5\, y_{11}.$$

Note that we may eliminate the generators $y_1,\ y_7,\ y_8,\ y_9,\ y_{10},\ y_{11}$ and obtain the following presentation:

$$\mathscr{K} = \langle y_2, y_3, y_4, y_5, y_6 \mid y_2\, y_3^{-1}\, y_4\, y_5^{-1}\, y_6\ ,\ y_2^{-1}\, y_3\, y_4^{-1}\, y_5\, y_6^{-1} \rangle. \tag{6.2.2}$$

In order to make it less cumbersome, let us once again rename the generators, now as

$$y_2 =: a,\ y_3^{-1} =: b,\ y_4 =: c,\ y_5^{-1} =: d.$$

The two relations in (6.2.2) imply that

$$\begin{aligned}&y_6 = y_2^{-1}\, y_3\, y_4^{-1}\, y_5 = (y_5^{-1}\, y_4\, y_3^{-1}\, y_2)^{-1};\\&y_2\, y_3^{-1}\, y_4\, y_5^{-1}\, y_6 = 1.\end{aligned}$$

which, in terms of the new names introduced above, yield the following presentation

$$\mathscr{K} = \langle a, b, c, d \mid abcd(dcba)^{-1}\rangle.$$

Finally, we achieve the standard presentation by making one last change in the generators of $\mathscr{K}$:

$$x := ab^{-2},\ y := b,\ x' := bac,\ y' := dc.$$

Note that $\{x, y, x', y'\}$ generates $\mathscr{K}$ indeed, since

$$x\, y^2 = a,\ y = b,\ y^{-2}\, x^{-1}\, y^{-1}\, x' = c,\ y'\, (x')^{-1}\, y\, x\, y^2 = d.$$

With this new set of generators, the relation $abcd(dcba)^{-1} = 1$ becomes $[x, y][x', y'] = 1$, i.e.

$$\mathscr{K} = \langle x, y, x', y' \mid [x, y][x', y']\rangle.$$

By the definition of the homomorphism ϕ and (6.2.1), we conclude this section by pointing out that:

$$c_1c_2 = \phi(x_1x_2) = \phi(y_2) \in \mathscr{K}$$

Appendices

APPENDIX A

A BRIEF INTRODUCTION TO KRULL VALUATIONS

In this appendix we introduce Krull valuations. By the end, we provide a proof of Theorem 2.3.40.

Let us explore now another point of view when dealing with valuations: the valuation ring. Recall that an integral domain B with field of fractions K is said to be a valuation ring (of K) if, for every non-zero $x \in K$, it holds that $x \in B$ or $x^{-1} \in B$ (see Definition 2.3.20).

We saw that a non-Archimedean valuation v on K induces a valuation ring, namely, the ring $\mathfrak{o}_v = \{x \in K \mid v(x) \leq 1\}$. It is then natural to ask whether every valuation ring in K comes from some (non-Archimedean) valuation. The answer is no, in general, unless we allow for a more comprehensive notion of valuation, Krull valuations, to be introduced presently. First, let us state some of the basic properties of valuation rings.

Definition A.0.1. A ring A is said to be a *local ring* if it has a unique maximal ideal. It is immediate to check that A is a local ring with maximal ideal $\mathfrak{m}$ if and only if $A \setminus \mathfrak{m}$ is comprised of units of A.

Proposition A.0.2. *Let B be a valuation ring of K. Then*

1. *B is a local ring;*
2. *B is integrally closed.*

Proof. The proof of this proposition is routine. See [1, Propositon 5.12]. □

Definition A.0.3. An *ordered abelian group* is an (additive) abelian group Γ, together with a linear order, i.e., a binary relation $\leq$ satisfying:

(i) (Reflexivity) $x \leq x$, for all $x \in \Gamma$;

(ii) (Antisymmetry) $x \leq y,\ y \leq x \implies x = y$, for all $x, y \in \Gamma$;

(iii) (Transitivity) $x \leq y,\ y \leq z \implies x \leq z$, for all $x, y, z \in \Gamma$;

(iv) (Linearity) $x \leq y$ or $y \leq x$, for all $x, y \in \Gamma$;

which also satisfies:

5. (Monotonicity) $x \leq y \implies x + z \leq y + z$, for all $x, y, z \in \Gamma$.

A Krull valuation is much like an additive valuation, as defined in Definition 2.3.9, except it takes values in an ordered abelian group. More precisely,

Definition A.0.4. Let Γ be an ordered abelian group. A *Krull valuation* or Γ*-valuation* is a surjective function $v : K \twoheadrightarrow \Gamma \cup \{\infty\}$ satisfying:

(i) $v(x) = \infty$ if and only if $x = 0$;

(ii) $v(xy) = v(x) + v(y)$;

(iii) $v(x + y) \geq \min\{v(x), v(y)\}$.

As usual, $x \leq \infty$ and $x + \infty = \infty$ for all $x \in \Gamma$. Note that a Krull valuation as defined above is a direct generalisation of an additive valuation (see Definition 2.3.9). In an entirely analogous way, one can define an ordered *multiplicative* abelian group $(\Lambda, \cdot)$ and consider functions $v' : K \to \Lambda$ satisfying:

(i) $v'(x) = 0$ if and only if $x = 0$;

(ii) $v'(xy) = v'(x) \cdot v'(y)$;

(iii) $v'(x + y) \leq \max\{v'(x), v'(y)\}$,

as to obtain a direct generalisation of valuations as defined in Definition 2.3.1. Summarising, there are two equivalent notions: that of a valuation (Definition 2.3.1) and that of an additive valuation (Definition 2.3.9). Additive valuations generalise to Krull valuations (Definition A.0.4) while valuations generalise to the multiplicative version of Krull valuations explained above. These two more general notions are also equivalent. Even though we have worked mostly with valuations in this text (instead of additive valuations), we chose the additive version of Krull valuations as a preferred generalisation. We stress this is purely a matter of taste.

In the same way that (additive) valuations induced valuation rings, a Krull valuation $v : K \twoheadrightarrow \Gamma \cup \{\infty\}$ induces a valuation ring given by $\mathfrak{o}_v = \{x \in K \mid v(x) \geq 0\}$. Conversely, a valuation ring of K comes from a Krull valuation, as the next theorem shows.

Proposition A.0.5. *For every valuation ring $\mathfrak{o}$ of K there exists a valuation $v : K \twoheadrightarrow \Gamma \cup \{\infty\}$ such that $\mathfrak{o} = \mathfrak{o}_v$.*

Proof. Take Γ to be the abelian group $K^\times/\mathfrak{o}^\times$ with group law written additively, i.e., $x\mathfrak{o}^\times + y\mathfrak{o}^\times$ is defined to be $xy\mathfrak{o}^\times$.

Define an order $\leq$ on Γ as: $x\mathfrak{o}^\times \leq y\mathfrak{o}^\times$ if and only if $y/x \in \mathfrak{o}^\times$. The fact that $\mathfrak{o}$ is a valuation ring implies that the order thus defined is linear. Moreover, $(\Gamma, \leq)$ is an ordered abelian group.

The function $v : K \twoheadrightarrow \Gamma \cup \{\infty\}$ defined by $v(x) = x\mathfrak{o}^\times$ for $x \in K \setminus \{0\}$ and $v(0) = \infty$, is easily seen to be a valuation. Finally, note that the neutral element $0 \in \Gamma$ is given by $1\mathfrak{o}^\times$, and therefore $x\mathfrak{o}^\times = v(x) \geq 0 = 1\mathfrak{o}$ if and only if $x = x/1 \in \mathfrak{o}$. It follows that $\mathfrak{o}_v = \mathfrak{o}$. □

Definition A.0.6. Two valuations $v_1 : K \twoheadrightarrow \Gamma_1 \cup \{\infty\}$ and $v_2 : K \twoheadrightarrow \Gamma_2 \cup \{\infty\}$ are said to be equivalent when there exists an order-isomorphism $f : \Gamma_1 \to \Gamma_2$, i.e., a group-isomorphism that preserves the order, and f is such that $f \circ v_1 = v_2$.

The correspondence between valuation rings and Krull valuations, described by Proposition A.0.5 and the discussion preceding it, respects the equivalence relation just defined. This is precisely the content of the next proposition.

Proposition A.0.7. *Two valuations are equivalent if and only if they have the same associated valuation rings.*

Proof. Let $v_1 : K \twoheadrightarrow \Gamma_1 \cup \{\infty\}$ and $v_2 : K \twoheadrightarrow \Gamma_2 \cup \{\infty\}$ be two valuations. If v_1 and v_2 are equivalent then $v_1(x) \geq 0$ if and only if $v_2(x) \geq 0$ and thus it $\mathfrak{o}_{v_1} = \mathfrak{o}_{v_2}$.

Conversely, suppose $\mathfrak{o}_{v_1} = \mathfrak{o}_{v_2}$, which implies, in particular, that $\mathfrak{o}_{v_1}^\times = \mathfrak{o}_{v_2}^\times$. Now, for $i = 1, 2$, since $v_i : K^\times \twoheadrightarrow \Gamma_i$ is a group-homomorphism with kernel $\mathfrak{o}_{v_i}^\times$, it induces an isomorphism $V_i : K^\times/\mathfrak{o}_{v_i}^\times \to \Gamma_i$. Given that $\mathfrak{o}_{v_1}^\times = \mathfrak{o}_{v_2}^\times$, we may set $f = V_2 \circ V_1^{-1} : \Gamma_1 \to \Gamma_2$ and it is clear that f satisfies $f \circ v_1 = v_2$. □

A subgroup Λ of an ordered abelian group Γ is said to be *convex* if it satisfies the following condition: for all $\gamma \in \Gamma$, if there exists $\lambda \in \Lambda$ such that $0 \leq \gamma \leq \lambda$ then $\gamma \in \Lambda$. In this context, the *rank* of Γ is the order type of the collection of proper convex subgroups of Γ, which is linearly ordered, as one can easily check. Note that, for a non-trivial $\Gamma < (\mathbb{R}, +)$ (with the usual order), the only convex subgroup of Γ is the trivial subgroup $\{0\}$ and, therefore, we say that Γ has rank one. As it turns out, being of rank one is a defining characteristic of subgroups of $(\mathbb{R}, +)$, as the next proposition shows.

Proposition A.0.8. *An ordered abelian group Γ is of rank one if and only if it is order-isomorphic to a non-trivial subgroup of $(\mathbb{R}, +)$ with the usual order.*

Proof. See [17, Proposition 2.1.1]. □

Proposition A.0.8 justifies the terminology introduced in Definition 2.3.9, in the sense that the (Krull) valuations with a rank one value group are precisely the additive valuations defined in Definition 2.3.9, up to equivalence (of Krull valuations).

Moreover, if a valuation has value group of rank larger than one, than it cannot be one of the valuations defined earlier in Definition 2.3.9. It is easy to find ordered abelian groups with rank > 1, for example, the group $\mathbb{Z}\times\mathbb{Z}$ with the lexicographical order has rank 2 (or, more generally, $\mathbb{Z}^n$ with lexicographical order has rank n), which begs the question: is this ordered abelian group the value group of some valuation on some field? The answer is affirmative, every ordered abelian group Γ is the value group of some valuation v ([12, Exercise 1.64]): take any field K and consider the ring $K[\Gamma]$. An element of $K[\Gamma]$ is of the form $x = \sum_{\gamma\in\Gamma} x_\gamma \cdot \gamma$ where all but finitely many x_γ are 0, for which we define $v(x) = \min\{\gamma \mid x_\gamma \neq 0\}$, assuming x is non-zero. Note that $K[\Gamma]$ is an integral domain and that v thus defined can be extended to its field of fractions. The result is a valuation with value group Γ. This observation shows that one can produce many Krull valuations that are not rank one valuations, which means that the class of Krull valuations is strictly larger. For the field of rational numbers $\mathbb{Q}$, however, we do not gain anything new by considering Krull valuations, as the next proposition shows.

Proposition A.0.9. *Every non-trivial (Krull) valuation on $\mathbb{Q}$ is equivalent to a p-adic valuation.*

Proof. Let v be a non-trivial Krull valuation on $\mathbb{Q}$ with valuation ring $\mathfrak{o}$. Note that $\mathbb{Z} \subset \mathfrak{o}$ since $1 \in \mathfrak{o}$. If every prime number was a unit in $\mathfrak{o}$ then $\mathfrak{o}$ would contain every number of the form $1/n$ for $n \in \mathbb{Z}$ and, consequently, every rational number. But $\mathfrak{o} \subsetneq \mathbb{Q}$ (v is non-trivial) and hence there must be some rational prime p that is not a unit in $\mathfrak{o}$. In other words, p is in the maximal ideal $\mathfrak{m} \subset \mathfrak{o}$. For every other prime $q \neq p$ there exist integers a, b such that $ap + bq = \gcd(p, q) = 1$, whence $q \notin \mathfrak{m}$. It follows that $\mathfrak{o}$ is the ring $\mathbb{Z}$ localised at the ideal (p), i.e., the ring

$$\{m/n \in \mathbb{Q} \mid \gcd(m, n) = 1 \text{ and } p \nmid n\},$$

which is precisely the valuation ring of the p-adic valuation on $\mathbb{Q}$. The result then follows from Proposition A.0.7. □

Next is a theorem credited to Chevalley. We follow the proof in [17, Theorem 3.1.1].

Theorem A.0.10 (Chevalley). *Let A be a subring of a field K and let $\mathfrak{p}$ be a prime ideal of A. There exists a valuation ring $B \subset K$, with maximal ideal $\mathfrak{m}$, such that:*

$$A \subset B \quad \textit{and} \quad \mathfrak{m} \cap A = \mathfrak{p}. \tag{A.0.1}$$

Proof. For B and $\mathfrak{m}$ as in the statement of the theorem, condition (A.0.1) implies that: if $x \in A \setminus \mathfrak{p}$, then $x \notin \mathfrak{m}$, which means that $x^{-1} \in B$. In particular, one must have that $A_\mathfrak{p} \subset B$, where $A_\mathfrak{p}$ denotes the localisation of A at $\mathfrak{p}$. Moreover, since $\mathfrak{p} = \mathfrak{m} \cap A$, one must have $\mathfrak{p}A_\mathfrak{p} \subset \mathfrak{m}$.

It is, therefore, reasonable to start looking for B and $\mathfrak{m}$ among the subrings $R \subset K$ that contain $A_\mathfrak{p}$ and a proper ideal $\mathfrak{I}$ which contains $\mathfrak{p}A_\mathfrak{p}$. More precisely, consider the collection $\mathscr{C}$ defined by

$$\mathscr{C} = \{(R, \mathfrak{I}) \mid R \text{ is a ring, } \mathfrak{I} \text{ is a proper ideal of } R,\ A_\mathfrak{p} \subset R \subset K,\ \mathfrak{p}A_\mathfrak{p} \subset \mathfrak{I} \subset R\}.$$

The collection $\mathscr{C}$ is non-empty since $(A_\mathfrak{p}, \mathfrak{p}A_\mathfrak{p}) \in \mathscr{C}$, and is partially ordered by (componentwise) inclusion, i.e., $(R_1, \mathfrak{I}_1) \preceq (R_2, \mathfrak{I}_2)$ if and only if $R_1 \subset R_2$ and $\mathfrak{I}_1 \subset \mathfrak{I}_2$. This partial order is easily seen to satisfy the condition that every chain has an upper bound and so, by Zorn's Lemma, there exists a maximal element $(B, \mathfrak{m})$ in $\mathscr{C}$. It turns out that this maximal element is precisely what we were looking for.

Firstly, since $\mathfrak{m} \cap A_\mathfrak{p}$ is a proper ideal of $A_\mathfrak{p}$ and $\mathfrak{p}A_\mathfrak{p}$ is maximal, it follows that $\mathfrak{m} \cap A_\mathfrak{p} = \mathfrak{p}A_\mathfrak{p}$. Intersecting both sides with A yields $\mathfrak{m} \cap A = \mathfrak{p}$.

To conclude the proof, we only need to show that B is a valuation ring. It follows from the maximality of $(B, \mathfrak{m})$ that $\mathfrak{m}$ is maximal in B. Moreover, since $(B, \mathfrak{m}) \preceq (B_\mathfrak{m}, \mathfrak{m}B_\mathfrak{m})$, maximality also implies that $B_\mathfrak{m} = B$, and consequently that $B \setminus \mathfrak{m} = B^\times$, meaning that $\mathfrak{m}$ is the only maximal ideal of B. In other words, we have shown at this stage that B is a local ring.

Suppose B is not a valuation ring, so that there exists $x \in K$ such that $x \notin B$ and $x^{-1} \notin B$. In this case, $B \subsetneqq B[x]$ and $B \subsetneqq B[x^{-1}]$. Maximality of $(B, \mathfrak{m})$ implies that the ideal $\mathfrak{m}[x]$ cannot be proper, so $\mathfrak{m}[x] = B[x]$. Similarly, $\mathfrak{m}[x^{-1}] = B[x^{-1}]$. In particular, there are $x_0, \ldots, x_n, y_1, \ldots, y_m \in \mathfrak{m}$ such that

$$x_0 + x_1 x + \cdots + x_n x^n = 1 = y_0 + y_1 x^{-1} + \cdots + y_m x^{-m}, \tag{A.0.2}$$

and assume, without any loss in generality, that $m \leq n$ are minimal with this property. Since $1 - y_0 \notin \mathfrak{m}$ and B is a local ring, we have that $1 - y_0 \in B \setminus \mathfrak{m} = B^\times$ and thus that

$$1 = \frac{y_1}{1-y_0}x^{-1} + \cdots + \frac{y_m}{1-y_0}x^{-m}.$$

Multiplying both sides by x^n allows us to express x^n in terms of lower powers of x, namely

$$x^n = \frac{y_1}{1-y_0}x^{n-1} + \cdots + \frac{y_m}{1-y_0}x^{n-m}. \tag{A.0.3}$$

Substituting (A.0.3) into the left-hand side of (A.0.2) yields

$$z_0 + z_1 x + \cdots + z_{n-1}x^{n-1} = 1,$$

for some $z_0, \ldots, z_{n-1} \in \mathfrak{m}$, contradicting the minimality of n. □

Corollary A.0.11 (Characterisation of integral closure)**.** *Let A be a subring of the field K. Then the integral closure of A in K, $\overline{A}$, is given by:*

$$\overline{A} = \bigcap_{\substack{A \subset R \subset K \\ R \text{ valaution ring}}} R. \tag{A.0.4}$$

Moreover, let $\mathscr{B}$ denote the smaller collection consisting of those valuation rings B of K, with maximal ideal $\mathfrak{m}$, such that $A \subset B$ and $\mathfrak{m} \cap A$ is a maximal ideal of A. Then it still holds that

$$\overline{A} = \bigcap_{B \in \mathscr{B}} B.$$

Proof. Since every valuation ring is integrally closed (Proposition A.0.2 (2)), it is immediate that

$$\overline{A} \subset \bigcap_{\substack{A \subset R \subset K \\ R \text{ valaution ring}}} R \subset \bigcap_{B \in \mathscr{B}} B.$$

It is therefore sufficient to check that $\bigcap_{B \in \mathscr{B}} B \subset \overline{A}$. We prove the contrapostive: suppose $x \notin \overline{A}$. Then $x \notin A[x^{-1}]$, otherwise it would be integral over A. It follows that x^{-1} in not a unit of $A[x^{-1}]$ and, therefore, there exists a (maximal) prime ideal $\mathfrak{p} \subset A[x^{-1}]$ that contains x^{-1} (it is a consequence of Zorn's Lemma that every proper ideal is contained in a maximal ideal and, in particular, so is $x^{-1} \in A[x^{-1}]$). We apply Theorem A.0.10 to the pair $\mathfrak{p} \subset A[x^{-1}]$ and obtain a valuation ring $B \supset A[x^{-1}]$ with maximal ideal $\mathfrak{m}$ such that $\mathfrak{m} \cap A[x^{-1}] = \mathfrak{p}$, so $x^{-1} \in \mathfrak{m}$ which means that x^{-1} is a non-unit of B. In other words, $x \notin B$.

If we can show that $B \in \mathscr{B}$, the proof is finished. So far, we have that B is a valuation ring with maximal ideal $\mathfrak{m}$ such that $A \subset B$. The only thing left to verify is that $\mathfrak{m} \cap A$ is a maximal ideal of A. Observe that the canonical projection $g : A \to A[x^{-1}]/\mathfrak{p}$ is a surjective homomorphism since $x^{-1} \in \mathfrak{p}$. This implies that $A/\ker g$ is isomorphic to $A[x^{-1}]/\mathfrak{p}$, which is a field since $\mathfrak{p}$ is maximal in $A[x^{-1}]$. So $A/\ker g$ is also a field and, consequently, $\ker g$ is a maximal ideal of A. Since $\mathfrak{m} \cap A[x^{-1}] = \mathfrak{p}$, one sees that $\ker g = A \cap \mathfrak{p} = A \cap \mathfrak{m}$. □

In the case of number fields, Corollary A.0.11 gives us the following:

Corollary A.0.12 (Characterisation of the ring of integers of a number field)**.** *Let K be a number field and $\mathscr{O}_K$ its ring of integers. If V_f denotes the set of all non-Archimedean valuations on K, then*

$$\mathscr{O}_K = \bigcap_{v \in V_f} \{x \in K \mid v(x) \leq 1\}.$$

Proof. By definition, $\mathscr{O}_K$ is the integral closure of $\mathbb{Z}$ in K and thus, according to Corollary A.0.11, it is the intersection of all valuation rings of K that contain $\mathbb{Z}$.

Let B be one of these valuation rings of K. Then B is associated to a (Krull) valuation w on K. Moreover, $B \cap \mathbb{Q}$ is a valuation ring associated to a Krull valuation on $\mathbb{Q}$, which, by Proposition A.0.9, must be a p-adic valuation. Now, since $K \mid \mathbb{Q}$ is algebraic, it follows from Proposition A.0.17 (1) below and from Proposition A.0.8 that w is equivalent to a rank one additive valuation and thus $v = e^{-w}$ is a non-Archimedean valuation on K with ring of integers B. Conversely, all non-Archimedean valuations on K contain $\mathbb{Z}$ in its ring of integers. □

Definition A.0.13. Let $L \mid K$ be a field extension, let $\mathfrak{o}_1 \subset K$ and $\mathfrak{o}_2 \subset L$ be valuation rings. We say $\mathfrak{o}_2$ *lies over* $\mathfrak{o}_1$ (or that $\mathfrak{o}_2$ is an *extension* of $\mathfrak{o}_1$) if $\mathfrak{o}_2 \cap K = \mathfrak{o}_1$. If this is the case, we write $(K_1, \mathfrak{o}_1) \subset (L, \mathfrak{o}_2)$.

If $(K_1, \mathfrak{o}_1) \subset (L, \mathfrak{o}_2)$, note that $x \in \mathfrak{o}_2 \cap \mathfrak{o}_1$ is a non-unit in $\mathfrak{o}_2$ if and only if it is a non-unit in $\mathfrak{o}_1$. In particular, if we let $\mathfrak{m}_i$ denote the maximal ideal of $\mathfrak{o}_i$, $i = 1, 2$, it follows that

$$\begin{aligned} \mathfrak{m}_2 \cap K_1 &= \mathfrak{m}_2 \cap \mathfrak{o}_1 = \mathfrak{m}_1, \\ \mathfrak{o}_2^\times \cap K_1 &= \mathfrak{o}_2^\times \cap \mathfrak{o}_1 = \mathfrak{o}_1^\times. \end{aligned}$$

Corollary A.0.14. *If $\mathfrak{o}_1 \subset K$ is a valuation ring and $L \mid K$, then there exists a valuation ring $\mathfrak{o}_2 \subset L$ that lies over $\mathfrak{o}_1$.*

Proof. Regarding $\mathfrak{o}_1$ as a subring of L, with maximal ideal $\mathfrak{m}_1$, it follows from Theorem A.0.10 that there exists a valuation ring $\mathfrak{o}_2$ of L, with maximal ideal $\mathfrak{m}_2$, such that $\mathfrak{o}_1 \subset \mathfrak{o}_2$ and $\mathfrak{m}_2 \cap \mathfrak{o}_1 = \mathfrak{m}_1$.

All that is left to show now is that $\mathfrak{o}_2 \cap K = \mathfrak{o}_1$. The inclusion $\supset$ is immediate, so let $x \in \mathfrak{o}_2 \cap K$ and suppose $x \notin \mathfrak{o}_1$. Then $x^{-1} \in \mathfrak{m}_1$. It follows that $x^{-1} \in \mathfrak{m}_2$, whence $x \notin \mathfrak{o}_2$, a contradiction.

□

With this terminology, we may apply Corollary A.0.11 to a valuation ring of K in order to characterise its integral closure in a field extension $L \mid K$.

Corollary A.0.15. *Let $L \mid K$ and let $\mathfrak{o}$ be a valuation ring of K. The integral closure of $\mathfrak{o}$ in L, denoted by $\overline{\mathfrak{o}}^L$, is the intersection of all valuation rings of L lying over $\mathfrak{o}$.*

Proof. We claim that the collection of all valuation rings of L lying over $\mathfrak{o}$ is precisely $\mathscr{B}$ (in the notation of Corollary A.0.11). Indeed, let $\mathfrak{m}$ denote the maximal ideal of $\mathfrak{o}$. If B is a valuation ring of L, with maximal ideal $\mathfrak{m}_B$, lying over $\mathfrak{o}$, then, in particular, $\mathfrak{o} \subset B$ and $\mathfrak{m}_B \cap \mathfrak{o} = \mathfrak{m}$ is maximal in $\mathfrak{o}$, so $B \in \mathscr{B}$. Conversely, let $B \in \mathscr{B}$, then $\mathfrak{o} \subset B$ and $\mathfrak{m}_B \cap \mathfrak{o} = \mathfrak{m}$. Reasoning just as in the proof of Corollary A.0.14, we obtain that B lies over $\mathfrak{o}$.

Now the result follows from Corollary A.0.11. □

Let $(K_1, \mathfrak{o}_1) \subset (K_2, \mathfrak{o}_2)$ with corresponding valuations

$$v_1 : K_1 \twoheadrightarrow \Gamma_1 \cup \{\infty\} \quad \text{and} \quad v_2 : K_2 \twoheadrightarrow \Gamma_2 \cup \{\infty\}.$$

The homomorphism $v_i : K_1^\times \twoheadrightarrow \Gamma_1$ has kernel $\mathfrak{o}_i^\times$, so that $K_i^\times / \mathfrak{o}_i^\times \cong \Gamma_i$, $i = 1, 2$. The composition of homomorphisms $K_1^\times \hookrightarrow K_2^\times \twoheadrightarrow K_2^\times / \mathfrak{o}_2^\times \cong \Gamma_2$ has kernel $K_1^\times \cap \mathfrak{o}_2^\times = \mathfrak{o}_1^\times$, whence $\Gamma_1 \cong K_1^\times / \mathfrak{o}_1^\times \hookrightarrow K_2^\times / \mathfrak{o}_2^\times \cong \Gamma_2$ and we may regard Γ_1 as a subgroup of Γ_2.

Lemma A.0.16. *Let $(K_1, \mathfrak{o}_1) \subset (K_2, \mathfrak{o}_2)$ with corresponding valuations*

$$v_1 : K_1 \twoheadrightarrow \Gamma_1 \cup \{\infty\} \quad \textit{and} \quad v_2 : K_2 \twoheadrightarrow \Gamma_2 \cup \{\infty\}.$$

If $x_1, \dots, x_k \in K_2^\times$ are such that $v_2(x_1), \dots, v_2(x_k)$ represent different cosets in Γ_2/Γ_1, then $\{x_1, \dots, x_k\}$ are linearly independent over K_1.

Proof. For $\lambda_1, \dots, \lambda_k \in K_1$ not all zero, let $1 \le j \le k$ be such that

$$v_2(\lambda_j x_j) = \min\{v_2(\lambda_1 x_1), \dots, v_2(\lambda_k x_k)\} < \infty.$$

If there exists some $i \neq j$ such that $v_2(\lambda_j x_j) = v_2(\lambda_i x_i)$ (note that, in this case, $\lambda_i, \lambda_j \neq 0$), then $v_2(x_j) - v_2(x_i) = v_2(\lambda_i) - v_2(\lambda_j) \in \Gamma_1$, contrary to our hypothesis. So $v_2(\lambda_j x_j) < v_2(\lambda_i x_i)$ for all $i \neq j$. It follows that $v_2\left(\sum_{i=1}^k \lambda_i x_i\right) = v_2(\lambda_j x_j) < \infty$. In particular, $\sum_{i=1}^k \lambda_i x_i \neq 0$. □

Proposition A.0.17. *Suppose $K_2 \mid K_1$ is algebraic and $(K_1, \mathfrak{o}_1) \subset (K_2, \mathfrak{o}_2)$. Let v_i be the corresponding valuations with value group Γ_i, $i = 1, 2$. The following hold:*

1. Γ_2/Γ_1 *is a torsion group;*

2. Γ_1 *and* Γ_2 *have the same rank.*

Proof. For $\gamma \in \Gamma_2$, take $x \in K_2$ such that $v_2(x) = \gamma$. Let Γ denote the subgroup $v_2(K_1(x)) < \Gamma_2$. It follows from Lemma A.0.16 that $[\Gamma : \Gamma_1] \leq [K_1(x) : K_1] < +\infty$, so that the group Γ/Γ_1 is finite. In particular, $[\Gamma : \Gamma_1]\gamma \in \Gamma_1$ and (1) follows.

Now, consider the correspondence that associates a convex subgroup Λ of Γ_2 to the subgroup $\Lambda \cap \Gamma_1$ of Γ_1. First, note that $\Lambda \cap \Gamma_1$ is a proper convex subgroup of Γ_1, so what we really have is a correspondence, say Φ, between proper convex subgroups of Γ_2 and proper convex subgroups of Γ_1. We use the knowledge that Γ_2/Γ_1 is a torsion group to show that Φ is an order-preserving bijection between these two sets, which means that they have the same order type. Thus, by definition, Γ_2 and Γ_1 have the same rank.

Injectivity of Φ: let Λ and Λ' be convex subgroups of Γ_2 such that $\Lambda \cap \Gamma_1 = \Lambda' \cap \Gamma_1$. If $\lambda \in \Lambda$, suppose $\lambda \geq 0$. For some N, $N\lambda \in \Lambda \cap \Gamma_1$, so $N\lambda$ is also in $\Lambda' \cap \Gamma_1$ and, since $0 \leq \lambda \leq N\lambda$, convexity implies that $\lambda \in \Lambda'$. When $\lambda \leq 0$, we apply the same argument to $-\lambda$. This shows that $\Lambda \subset \Lambda'$ and, by symmetry, we obtain that $\Lambda = \Lambda'$.

Surjectivity of Φ: let Λ be a proper convex subgroup of Γ_1 and define the "radical" of Λ to be $\sqrt{\Lambda} := \{\gamma \in \Gamma_2 \mid N\gamma \in \Lambda \text{ for some } N\}$. It is straightforward to check that $\sqrt{\Lambda}$ is a convex subgroup of Γ_2 and that $\sqrt{\Lambda} \cap \Gamma_1 = \Lambda$. The last equality also shows that $\sqrt{\Lambda}$ is proper.

Finally, note that Φ preserves inclusion. □

We are finally ready to prove that, for an algebraic field extension $L \mid K$, a valuation on K can always be extended to L, as promised in §§2.3.9. We restate the theorem for the convenience of the reader.

Theorem 2.3.40. *Let $L \mid K$ be an algebraic extension and let $v : K \to R_{\geq 0}$ be a (rank one) valuation on K.*

There exists an extension of v to L.

Proof. We will treat the Archimedean and non-Archimedean cases separately. The non-Archimedean case will be proved with the Krull valuation machinery we have developed in this appendix, whereas the Archimedean case will follow from Ostrowski's Theorem 2.3.17.

Case 1: v is non-Archimedean.

In this case, we prove the (obviously equivalent) extension for additive valuations. Abusing notation slightly, we still denote by v the additive valuation $v : K \to \mathbb{R} \cup \{\infty\}$ (as in Definition 2.3.9). It induces a valuation ring $\mathfrak{o}_v = \{x \in K \mid v(x) \geq 0\}$ which, by Corollary A.0.14, extends to a valuation ring $\mathfrak{o} \subset L$, i.e., there exists a valuation ring $\mathfrak{o}$ of L such that $\mathfrak{o} \cap K = \mathfrak{o}_v$. By Proposition A.0.5, the ring $\mathfrak{o}$ is the valuation ring of some Krull valuation $w' : L \twoheadrightarrow \Gamma \cup \{\infty\}$.

Since $L \mid K$ is algebraic, it follows from Proposition A.0.17 that Γ has rank one and thus, Proposition A.0.8 implies the existence of an order-isomorphism f between Γ and some subgroup G of $(\mathbb{R}, +)$, whence $w'' = f \circ w' : L \to \mathbb{R} \cup \{\infty\}$ is an additive valuation on L. Moreover, note that

$$\{x \in K \mid w''(x) \geq 0\} = \{x \in L \mid w(x) \geq 0\} \cap K = \mathfrak{o} \cap K = \mathfrak{o}_v.$$

So $w''|_K$ and v are two additive valuations on K with the same valuation ring and therefore must be equivalent, again in the sense of Definition 2.3.9 (cf. Proposition 2.3.12). This means there exists $a > 0$ such that $v = aw''|_K$. Then the valuation $w = aw''$ on L is the valuation we want.

Case 2: v is Archimedean.

This case is a consequence of Ostrowski's Theorem 2.3.17. Indeed, suppose for a moment that K is complete with respect to v. Then, by Ostrowski's Theorem, we either have that $K = \mathbb{R}$ and $L = \mathbb{C}$ or that $K = L = \mathbb{R}$ or $\mathbb{C}$. In either case, v is equivalent to the usual absolute value and can be extended to L.

Now, back to the general case, in which K is not necessarily complete with respect to v, we may take K_v to be the metric completion of K with respect to v and $\overline{K_v}$ its algebraic closure. We know that v extends uniquely to a valuation (which we keep denoting by v) on K_v, thus we are in a position to apply the result of the previous paragraph to the algebraic extension $\overline{K_v} \mid K_v$ and extend v to a valuation w' on $\overline{K_v}$. Note that, at this point, the extension w' is still uniquely determined, up to equivalence. Since the extension $L|K$ is assumed to be algebraic, the embedding of K into the algebraically closed field $\overline{K_v}$ can be extended to an embedding $\phi : L \hookrightarrow \overline{K_v}$ whence we may define the valuation $w(x) = w'(\phi(x))$ for $x \in L$, which extends v.

The only choice made in this process was the embedding ϕ. As a matter of fact, (all) different extensions of v arise when we vary the embedding of L into $\overline{K_v}$. A sample of this fact was seen in Corollary 2.3.53. □

BIBLIOGRAPHY

[1] Michael F. Atiyah and Ian G. Macdonald. *Introduction to commutative algebra*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1969.

[2] Alan F. Beardon. *The geometry of discrete groups*, volume 91. Springer Science & Business Media, 2012.

[3] Mikhail Belolipetsky. On the number of automorphisms of a nonarithmetic Riemann surface. *Siberian Math. J*, 38:860867, 1997.

[4] Mikhail Belolipetsky. Geodesics, volumes and Lehmer's conjecture. *arXiv preprint arXiv:1106.1834*, 2011.

[5] Nikolay Bogachev and Alexander Kolpakov. On faces of quasi-arithmetic Coxeter polytopes. *International Mathematics Research Notices*, 2021(4):3078–3096, 2021.

[6] Oleg Bogopolski. *Introduction to group theory*. EMS Textbooks in Mathematics. European Mathematical Society (EMS), Zürich, 2008. Translated, revised and expanded from the 2002 Russian original.

[7] Armand Borel. Commensurability classes and volumes of hyperbolic 3-manifolds. *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze*, 8(1):1–33, 1981.

[8] Peter Buser. *Geometry and spectra of compact Riemann surfaces*, volume 106 of *Progress in Mathematics*. Birkhäuser Boston, Inc., Boston, MA, 1992.

[9] Peter Buser and Peter Sarnak. On the period matrix of a Riemann surface of large genus (with an appendix by JH Conway and NJA Sloane). *Inventiones mathematicae*, 117(1):27–56, 1994.

[10] J. W. S. Cassels. *Local fields*, volume 3 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1986.

[11] John William Scott Cassels. *Global fields, Algebraic Number Theory (edited by J. Cassels, A. Fröhlich)*. Academic Press, 1967.

[12] Peter C. Clark. *Algebraic Number Theory II: Valuations, Local Fields and Adèles*. `http://math.uga.edu/~pete/8410FULL.pdf`. Accessed: 21/11/2020.

[13] Paula Cohen and Jürgen Wolfart. Modular embeddings for some non-arithmetic Fuchsian groups. *Acta Arithmetica*, 56(2):93–110, 1990.

[14] Gregory Cosac and Cayo Dória. Closed geodesics on semi-arithmetic Riemann surfaces. *to appear in Mathematical Research Letters*.

[15] Edoardo Dotti and Alexander Kolpakov. Infinitely many quasi-arithmetic maximal reflection groups. *arXiv preprint arXiv:2109.03316*, 2021.

[16] Vincent Emery. On volumes of quasi-arithmetic hyperbolic lattices. *Selecta mathematica*, 23(4):2849–2862, 2017.

[17] Antonio J Engler and Alexander Prestel. *Valued fields*. Springer Science & Business Media, 2005.

[18] Benson Farb and Dan Margalit. *A primer on mapping class groups*, volume 49 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 2012.

[19] Werner Fenchel. *Elementary geometry in hyperbolic space*, volume 11 of *De Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 1989. With an editorial by Heinz Bauer.

[20] Robert Fricke and Felix Klein. *Lectures on the theory of automorphic functions. Vol. 1*, volume 3 of *CTM. Classical Topics in Mathematics*. Higher Education Press, Beijing, 2017. Translated from the German original by Arthur M. DuPre.

[21] Slavyana Geninska. Examples of infinite covolume subgroups of $\mathrm{PSL}(2,\mathbb{R})^r$ with big limit sets. *Mathematische Zeitschrift*, 272(1-2):389–404, 2012.

[22] Slavyana Geninska. The limit set of subgroups of arithmetic groups in $\mathrm{PSL}(2,\mathbb{C})^q \times \mathrm{PSL}(2,\mathbb{R})^r$. *Groups, Geometry, and Dynamics*, 8(4):1047–1099, 2014.

[23] Philippe Gille and Tamás Szamuely. *Central simple algebras and Galois cohomology*, volume 165. Cambridge University Press, 2017.

[24] Yoichi Imayoshi and Masahiko Taniguchi. *An introduction to Teichmüller spaces*. Springer Science & Business Media, 2012.

[25] BoGwang Jeon. Realizing algebraic invariants of hyperbolic surfaces. *Trans. Amer. Math. Soc.*, 371(1):147–172, 2019.

[26] Gareth A Jones and David Singerman. *Complex functions: an algebraic and geometric viewpoint*. Cambridge university press, 1987.

[27] Jeremy Kahn. *Finding cocompact Fuchsian groups of given trace field and quaternion algebra*, 2015. `https://www.youtube.com/watch?v=-HxjWz_igbs&t=109s&ab_channel=InstituteforAdvancedStudy`. Accessed: 15/8/2021.

[28] Svetlana Katok. *Fuchsian groups*. University of Chicago press, 1992.

[29] Mikhail G. Katz, Mary Schaps, and Uzi Vishne. Logarithmic growth of systole of arithmetic Riemann surfaces along congruence subgroups. *Journal of Differential Geometry*, 76(3):399–422, 2007.

[30] Linda Keen. Canonical polygons for finitely generated Fuchsian groups. *Acta Mathematica*, 115(1):1–16, 1966.

[31] Robert A. Kucharczyk. Modular embeddings and rigidity for Fuchsian groups. *Acta Arith.*, 169(1):77–100, 2015.

[32] Tsit-Yuen Lam. *Introduction to quadratic forms over fields*, volume 67. American Mathematical Soc., 2005.

[33] Joseph Lehner. *Discontinuous groups and automorphic functions*. Number 8. American Mathematical Soc., 1964.

[34] Joseph Lehner. *A short course in automorphic functions*. Courier Corporation, 2014.

[35] Falko Lorenz. *Algebra: Volume II: Fields with structure, algebras and advanced topics*. Springer Science & Business Media, 2007.

[36] Colin Maclachlan and Alan W. Reid. *The arithmetic of hyperbolic 3-manifolds*, volume 219 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2003.

[37] Bruno Martelli. *An introduction to geometric topology*. CreateSpace Independent Publishing Platform, 2016.

[38] Bernard Maskit. On Poincaré's theorem for fundamental polygons. *Advances in Mathematics*, 7(3):219–230, 1971.

[39] Hideyuki Matsumura. *Commutative ring theory*, volume 8. Cambridge university press, 1989.

[40] Gou Nakamura and Toshihiro Nakanishi. Parametrizations of Teichmüller spaces by trace functions and action of mapping class groups. *Conformal Geometry and Dynamics of the American Mathematical Society*, 20(2):25–42, 2016.

[41] Toshihiro Nakanishi and Marjatta Näätänen. Parametrization of Teichmüller space by length parameters. In *Analysis And Topology: A Volume Dedicated to the Memory of S. Stoilow*, pages 541–560. World Scientific, 1998.

[42] Jürgen Neukirch. *Algebraic number theory*, volume 322 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999. Translated from the 1992 German original and with a note by Norbert Schappacher, With a foreword by G. Harder.

[43] Yoshihide Okumura. On the global real analytic coordinates for Teichmüller spaces. *Journal of the Mathematical Society of Japan*, 42(1):91–101, 1990.

[44] Dinakar Ramakrishnan and Robert J Valenza. *Fourier analysis on number fields*, volume 186. Springer Science & Business Media, 2013.

[45] Alan W Reid. A note on trace-fields of Kleinian groups. *Bulletin of the London Mathematical Society*, 22(4):349–352, 1990.

[46] Alan W. Reid. Isospectrality and commensurability of arithmetic hyperbolic 2- and 3-manifolds. *Duke Math. J.*, 65(2):215–228, 1992.

[47] Paulo Ribenboim. *Algebraic numbers*, volume 27. John Wiley & Sons, 1972.

[48] Paulo Ribenboim. *The theory of classical valuations*. Springer Science & Business Media, 2012.

[49] Paul Schmutz Schaller and Jürgen Wolfart. Semi-arithmetic Fuchsian groups and modular embeddings. *Journal of the London Mathematical Society*, 61(1):13–24, 2000.

[50] Kisao Takeuchi. A characterization of arithmetic Fuchsian groups. *Journal of the Mathematical Society of Japan*, 27(4):600–612, 1975.

[51] Kisao Takeuchi. Arithmetic triangle groups. *Journal of the Mathematical Society of Japan*, 29(1):91–106, 1977.

[52] William P. Thurston. Three-dimensional geometry and topology. Vol. 2, 2002.

[53] André Weil. On discrete subgroups of Lie groups. *Annals of Mathematics*, pages 369–384, 1960.

INDEX