



# FREE GROUPS IN $SL(n, \mathbb{Q})$

Rupert McCallum

Supervisor: Professor Michael Cowling

School of Mathematics,  
The University of New South Wales.

June 2003

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF THE DEGREE OF  
BACHELOR OF SCIENCE WITH HONOURS

---

## Acknowledgements

---

I would like to thank my parents, my supervisor Professor Michael Cowling, Dr Daniel Chan who commented on an early version of my Honours talk, and my Honours Co-ordinator Dr Ian Doust, for all the support and help they provided me with during the writing of this thesis.

---

## Contents

---

|           |                                 |    |
|-----------|---------------------------------|----|
| Chapter 1 | Dedekind domains and valuations | 1  |
| Chapter 2 | $SL(2, \mathbb{Q})$             | 23 |
| Chapter 3 | $SL(3, \mathbb{Q})$             | 37 |
|           | References                      | 44 |

---

# CHAPTER 1

## Dedekind domains and valuations

---

In this chapter, we develop some of the properties of Dedekind domains and prove that the ring of algebraic integers in an algebraic number field - by which we mean a subfield of  $\mathbb{C}$  which is a finite extension of  $\mathbb{Q}$  - is always a Dedekind domain. We also bring out the relationship between the prime ideals in a Dedekind domain and the nonarchimedean valuations on its field of fractions.

We call a  $\mathfrak{o}$ -module  $M$  a Noetherian  $\mathfrak{o}$ -module if all its  $\mathfrak{o}$ -submodules are finitely generated over  $\mathfrak{o}$ . This is equivalent to the condition that every ascending chain of  $\mathfrak{o}$ -submodules of  $M$  is eventually constant, and also to the condition that every nonempty family of  $\mathfrak{o}$ -submodules of  $M$  contains a maximal element. We call a ring  $\mathfrak{o}$  a Noetherian ring if  $\mathfrak{o}$  is a Noetherian  $\mathfrak{o}$ -module.

A ring  $R$  is said to be integrally closed in a larger ring  $S$  if every element of  $S$  which is integral over  $R$  - i.e. a root of a monic polynomial with coefficients in  $R$  - is in  $R$  itself.

**Definition 1.1** An integral domain  $\mathfrak{o}$  is said to be a Dedekind domain if

- (i)  $\mathfrak{o}$  is a Noetherian ring;
- (ii)  $\mathfrak{o}$  is integrally closed in its field of fractions  $K$ ;
- (iii) all nonzero prime ideals of  $\mathfrak{o}$  are maximal ideals.

We introduce the notion of a *fractional ideal*. Given an integral domain  $\mathfrak{o}$  with field of fractions  $K$ , a fractional ideal of  $\mathfrak{o}$  is an  $\mathfrak{o}$ -submodule  $\mathfrak{b}$  of  $K$  of the form  $\mathfrak{b} = c\mathfrak{a} = \{x \in K \mid x = ca \text{ for some } a \in \mathfrak{a}\}$  where  $c \in K^*$  and  $\mathfrak{a}$  is a non-zero  $\mathfrak{o}$ -ideal. If  $\mathfrak{o}$  is a Noetherian ring, then we easily see that a fractional  $\mathfrak{o}$ -ideal is a finitely generated  $\mathfrak{o}$ -submodule of  $K$ ; conversely we easily see that a finitely generated  $\mathfrak{o}$ -submodule of  $K$  is a fractional  $\mathfrak{o}$ -ideal. We define the product  $\mathfrak{b}_1 \cdot \mathfrak{b}_2$  of two fractional  $\mathfrak{o}$ -ideals to be the set of all finite sums  $\sum x_i y_i$  with  $x_i \in \mathfrak{b}_1, y_i \in \mathfrak{b}_2$ . This is clearly again a fractional  $\mathfrak{o}$ -ideal. For  $a \in K^*$  we define  $(a) = a\mathfrak{o}$ , then clearly  $(a)$  is always a fractional  $\mathfrak{o}$ -ideal and we have  $(a)(b) = (ab)$ .

We now work to prove the fundamental theorem of factorization of ideals in Dedekind domains, namely that every non-zero ideal  $\mathfrak{a}$  of a Dedekind domain  $\mathfrak{o}$  can be written as a product  $\mathfrak{a} = \mathfrak{p}_1 \dots \mathfrak{p}_n$  of prime ideals  $\mathfrak{p}_i$  of  $\mathfrak{o}$ ; moreover this representation is unique up to the order of the factors. We follow [3], chapter II, section 1.

For any fractional  $\mathfrak{o}$ -ideal  $\mathfrak{a}$  we write  $R_{\mathfrak{a}} = \{x \in K \mid x\mathfrak{a} \subset \mathfrak{a}\}$ .

**Proposition 1.1** If  $\mathfrak{o}$  is a Noetherian ring which is integrally closed, then  $R_{\mathfrak{a}} = \mathfrak{o}$ .

*Proof.* We clearly have  $\mathfrak{o} \subset R_{\mathfrak{a}}$ . Let  $a_1, \dots, a_n$  be a generating set of  $\mathfrak{a}$  over  $\mathfrak{o}$ , and let  $b \in R_{\mathfrak{a}}$ . Then we have  $ba_i = \sum_j c_{ij} a_j$  with  $c_{ij} \in \mathfrak{o}$ . It follows that  $b$  is a root of the monic polynomial  $\det(X \cdot 1_n - (c_{ij}))$ , which has coefficients in  $\mathfrak{o}$ , therefore  $b \in \mathfrak{o}$ . Thus  $R_{\mathfrak{a}} \subset \mathfrak{o}$  and the theorem is proved.  $\square$

**Proposition 1.2** If all the prime ideals of an integral domain  $\mathfrak{o}$  are maximal, then an inclusion  $\mathfrak{p} \supseteq \mathfrak{p}_1 \dots \mathfrak{p}_r$  where  $\mathfrak{p}$  and all the  $\mathfrak{p}_j$  are non-zero prime ideals, implies that  $\mathfrak{p} = \mathfrak{p}_i$  for some  $i$ .

*Proof.* Suppose  $r=1$ . Then  $\mathfrak{p} \supseteq \mathfrak{p}_1$  and since  $\mathfrak{p}_1$  is maximal we have  $\mathfrak{p} = \mathfrak{p}_1$ . Now suppose  $r > 1$  and the result is true for  $r-1$ . If  $\mathfrak{p} \neq \mathfrak{p}_r$ , then we have  $\mathfrak{p} \not\supseteq \mathfrak{p}_r$ . Let  $c \in \mathfrak{p}_r, c \notin \mathfrak{p}$ , and

let  $b \in \mathfrak{p}_1 \dots \mathfrak{p}_{r-1}$ . Then  $bc \in \mathfrak{p}_1 \dots \mathfrak{p}_r \subseteq \mathfrak{p}$ , and since  $\mathfrak{p}$  is prime and  $c \notin \mathfrak{p}$  we have  $b \in \mathfrak{p}$ . Thus  $\mathfrak{p} \supseteq \mathfrak{p}_1 \dots \mathfrak{p}_{r-1}$ .  $\square$

Given a fractional ideal  $\mathfrak{a}$  of an integral domain  $\mathfrak{o}$  with field of fractions  $K$ , we define  $\mathfrak{a}^{-1} = \{x \in K \mid x\mathfrak{a} \subset \mathfrak{o}\}$ . Note that if  $a \in \mathfrak{a}$ , we have  $\mathfrak{a}^{-1} \subset (a)^{-1}$ , and therefore if  $\mathfrak{o}$  is Noetherian then  $\mathfrak{a}^{-1}$  is a finitely generated  $\mathfrak{o}$ -module and so a fractional  $\mathfrak{o}$ -ideal.

Call a nonzero ideal  $\mathfrak{a}$  of an integral domain  $\mathfrak{o}$  weakly invertible if there is a  $c \in \mathfrak{a}^{-1}$  with  $c \notin \mathfrak{o}$ . Call a fractional ideal  $\mathfrak{a}$  of an integral domain  $\mathfrak{o}$  invertible if there is a fractional  $\mathfrak{o}$ -ideal  $\mathfrak{b}$  such that  $\mathfrak{a}\mathfrak{b} = \mathfrak{o}$ .

**Proposition 1.3** Let  $\mathfrak{o}$  be a Dedekind domain which is not a field and let  $\mathfrak{m}$  be a nonzero  $\mathfrak{o}$ -ideal which is maximal in the class of weakly invertible  $\mathfrak{o}$ -ideals. Then  $\mathfrak{m}$  is an invertible prime ideal of  $\mathfrak{o}$ .

*Proof.* Let  $a \in \mathfrak{o} \setminus \mathfrak{m}$  and suppose that  $b \in \mathfrak{o}, ab \in \mathfrak{m}$ . Since  $\mathfrak{m}$  is weakly invertible, there exists a  $c \in \mathfrak{m}^{-1} \setminus \mathfrak{o}$ , and furthermore since  $\mathfrak{m}$  is maximal in the weakly invertible ideals we have  $(\mathfrak{m} + a)c \not\subset \mathfrak{o}$ ; hence  $ac \notin \mathfrak{o}$ . Since  $ab \in \mathfrak{m}$ , we have  $b(ac) = (ab)c \in \mathfrak{m}\mathfrak{m}^{-1} \subset \mathfrak{o}$ ; thus  $ac \in (b)^{-1}$ . Furthermore since  $a \in \mathfrak{o}$  we have  $ac \in \mathfrak{m}^{-1}$ . It follows that  $ac \in (\mathfrak{m} + b\mathfrak{o})^{-1}$ . Thus  $\mathfrak{m} + b\mathfrak{o}$  is weakly invertible, and by the maximality of  $\mathfrak{m}$ , we have  $b \in \mathfrak{m}$ . We have derived that  $b \in \mathfrak{m}$  from the supposition that  $a \in \mathfrak{o} \setminus \mathfrak{m}, b \in \mathfrak{o}, ab \in \mathfrak{m}$ . This shows that  $\mathfrak{m}$  is a prime ideal.

By Proposition 1.1 we have  $\mathfrak{m}^{-1} \not\subset R_{\mathfrak{m}}$ . It follows that  $\mathfrak{m}\mathfrak{m}^{-1} \not\subset \mathfrak{m}$ . On the other hand we have  $\mathfrak{m} \subset \mathfrak{m}\mathfrak{m}^{-1} \subset \mathfrak{o}$ , and since  $\mathfrak{m}$  is maximal this means  $\mathfrak{m}\mathfrak{m}^{-1} = \mathfrak{o}$ .  $\square$

**Proposition 1.4** Let  $\mathfrak{o}$  be a Dedekind domain. A nonzero ideal  $\mathfrak{a}$  of  $\mathfrak{o}$  is invertible if and only if we have  $\mathfrak{a} = \mathfrak{m}_1 \dots \mathfrak{m}_r$  where the  $\mathfrak{m}_j$  are invertible prime ideals of  $\mathfrak{o}$ .

Note that the empty product is to be interpreted as  $\mathfrak{o}$ .

*Proof.* First note that, in the event  $\mathfrak{a}$  is invertible, we have  $\mathfrak{a}\mathfrak{b} = \mathfrak{o}$  for some fractional ideal  $\mathfrak{b}$ , so that  $\mathfrak{b} \subset \mathfrak{a}^{-1}$ , hence  $\mathfrak{o} = \mathfrak{a}\mathfrak{b} \subset \mathfrak{a}\mathfrak{a}^{-1} \subset \mathfrak{o}$ , so that  $\mathfrak{a}\mathfrak{a}^{-1} = \mathfrak{o}$  and  $\mathfrak{b} = \mathfrak{b}\mathfrak{a}\mathfrak{a}^{-1} = \mathfrak{a}^{-1}$ .

Now, if  $\mathfrak{a}$  is a proper invertible ideal, then  $\mathfrak{a}^{-1} \not\subseteq \mathfrak{o}$ , so that  $\mathfrak{a}$  is weakly invertible and so, since  $\mathfrak{o}$  is Noetherian, it is contained in a maximal weakly invertible ideal  $\mathfrak{m}_1$ . By Proposition 1.3 this will be invertible. Therefore  $\mathfrak{o} = \mathfrak{m}_1\mathfrak{m}_1^{-1} \supset \mathfrak{a}\mathfrak{m}_1^{-1} \supset \mathfrak{a}$ , and by Proposition 1.1  $\mathfrak{a}\mathfrak{m}_1^{-1} \neq \mathfrak{a}$ . Furthermore  $\mathfrak{a}\mathfrak{m}_1^{-1}$  is clearly also invertible. If  $\mathfrak{a}\mathfrak{m}_1^{-1} = \mathfrak{o}$ , then we have  $\mathfrak{a} = \mathfrak{m}_1$ . Otherwise we can repeat the process, and since  $\mathfrak{o}$  is Noetherian it must halt after finitely many steps so that we get an equation  $\mathfrak{o} = \mathfrak{a}\mathfrak{m}_1^{-1}\dots\mathfrak{m}_r^{-1}$ ; whence  $\mathfrak{a} = \mathfrak{m}_1\dots\mathfrak{m}_r$ .  $\square$

**Proposition 1.5** Every prime ideal  $\mathfrak{p}$  of a Dedekind domain is invertible.

*Proof.* Let  $a \in \mathfrak{p} \setminus \{0\}$ . Then  $(a)$  is invertible and we have  $\mathfrak{p} \supset (a) = \mathfrak{m}_1\dots\mathfrak{m}_r$  where the  $\mathfrak{m}_j$  are invertible prime ideals. Then by Proposition 1.2, we conclude that  $\mathfrak{p} = \mathfrak{m}_j$  for some  $j$ .  $\square$

**Proposition 1.6** Every nonzero ideal  $\mathfrak{a}$  of a Dedekind domain  $\mathfrak{o}$  can be written as a product  $\mathfrak{a} = \mathfrak{p}_1\dots\mathfrak{p}_r$  with the  $\mathfrak{p}_i$  prime ideals of  $\mathfrak{o}$ .

*Proof.* If  $\mathfrak{a} \neq \mathfrak{o}$ , then  $\mathfrak{a} \subset \mathfrak{p}$  for some prime ideal  $\mathfrak{p}$ . By Proposition 1.5,  $\mathfrak{p}$  will be invertible, and then  $\mathfrak{o} \supset \mathfrak{p}^{-1}\mathfrak{a} \supset \mathfrak{a}$ , with it following from Proposition 1.1 that  $\mathfrak{p}^{-1}\mathfrak{a} \neq \mathfrak{a}$ . As in the proof of Proposition 1.4, we iterate this process and it must terminate after finitely many steps, at the end of which we get an equality  $\mathfrak{a} = \mathfrak{p}_1\dots\mathfrak{p}_r$ .  $\square$

**Proposition 1.7** The representation of an ideal  $\mathfrak{a}$  as a product of prime ideals  $\mathfrak{p}_1\dots\mathfrak{p}_r$  is unique up to the order of factors.

*Proof.* Suppose  $\mathfrak{p}_1\dots\mathfrak{p}_s = \mathfrak{q}_1\dots\mathfrak{q}_r$  for some nonzero prime ideals  $\mathfrak{p}_i, \mathfrak{q}_j$ . Then we have  $\mathfrak{p}_s \supset \mathfrak{q}_1\dots\mathfrak{q}_r$ , whence by Proposition 1.2, renumbering the  $\mathfrak{q}_i$  if necessary, we get  $\mathfrak{p}_s = \mathfrak{q}_r$ .

We then multiply the equation by  $\mathfrak{p}_s^{-1} = \mathfrak{q}_r^{-1}$ ; this gives  $\mathfrak{p}_1 \dots \mathfrak{p}_{s-1} = \mathfrak{q}_1 \dots \mathfrak{q}_{r-1}$  and the result follows by induction on the number of factors.  $\square$

We now wish to prove that the ring of algebraic integers in an algebraic number field is always a Dedekind domain. We shall need various propositions for this, again we follow [3], Chapters 1 and 2.

**Proposition 1.8** Let  $E$  be a finite separable extension of a field  $F$  of degree  $n$ . Let  $L$  be an algebraically closed field containing  $F$  and let  $\sigma_i$ , for  $i = 1, \dots, n$ , be the  $n$  embeddings  $E \rightarrow L$  which extend the inclusion map  $F \hookrightarrow L$ . Given a basis  $u_1, \dots, u_n$  of  $E$  over  $F$ , let  $V^*(u_1, \dots, u_n) = \det(u_j^{\sigma_i})_{j,i}$ . Then  $V^*(u_1, \dots, u_n) \neq 0$ .

*Proof.* Given any basis  $w_1, \dots, w_n$  of  $E$  over  $F$ , we can write  $w_k = \sum_{j=1}^n c_{k,j} u_j$ , for  $k=1, \dots, n$ , where  $c_{k,j} \in F$  and  $\det(c_{k,j}) \neq 0$ . More generally we have  $w_k^{\sigma_j} = \sum_{i=1}^n c_{k,i} u_i^{\sigma_j}$ . Therefore  $V^*(w_1, \dots, w_n) = V^*(u_1, \dots, u_n) \det(c_{k,j})$ . Hence to prove the result for all bases of  $E$  over  $F$ , it will suffice to prove it for one such basis. Now it is a well-known theorem that a finite separable extension is simple, (see, for example, [4], Chapter 13), so take a generating element  $\alpha$ . Then  $1, \alpha, \dots, \alpha^{n-1}$  is a basis of  $E$  over  $F$ . Let us define

$$V = V(T_1, \dots, T_n) = \det \begin{pmatrix} 1 & 1 & \dots & 1 \\ T_1 & T_2 & \dots & T_n \\ \vdots & \vdots & & \vdots \\ T_1^{n-1} & T_2^{n-1} & \dots & T_n^{n-1} \end{pmatrix}.$$

Now  $V$  is a polynomial of total degree at most  $n(n-1)/2$  in the indeterminates. However, when we subtract the  $j$ th column from each of the other columns in turn, we do not affect the determinant and so we see that it is divisible by  $\prod_{i=1, i \neq j}^n (T_i - T_j)$ . Now the  $T_i - T_j$  are non-associate irreducible elements of the unique factorization domain  $L[T_1, \dots, T_n]$ . Therefore we conclude  $\prod_{1 \leq i < j \leq n} (T_i - T_j)$  divides  $V$ . Comparing degrees we see that in fact  $V$  is a constant multiple of this expression. This is enough to show that when  $\alpha_1, \dots, \alpha_n$  are the algebraic

conjugates of the generator  $\alpha$  of a separable extension, we have  $V(\alpha_1, \dots, \alpha_n) \neq 0$ . Hence  $V^*(1, \alpha, \dots, \alpha^{n-1}) \neq 0$ . We have now shown that  $V^*(w_1, \dots, w_n) \neq 0$  for one basis  $w_1, \dots, w_n$  and this is enough to prove our result.  $\square$

Now, if we define  $t_{E/F}(\alpha) = \sum_i \alpha^{\sigma_i}$ , then, multiplying the matrix  $(u_j^{\sigma_i})_{j,i}$  by its transpose, we see that  $\det(t_{E/F}(u_i u_j)) = V^*(u_1, \dots, u_n)^2 \neq 0$ . This shows that the  $F$ -bilinear mapping  $(x, y) \mapsto t_{E/F}(xy)$  is non-singular.

We now require some properties of Noetherian rings and modules. We take as given the well-known fact that a module over a ring is Noetherian if and only if every submodule is finitely generated.

**Proposition 1.9** Let  $0 \rightarrow M \rightarrow N \rightarrow P \rightarrow 0$  be an exact sequence of  $\mathfrak{o}$ -modules. Then  $N$  is a Noetherian  $\mathfrak{o}$ -module if and only if  $M$  and  $P$  are both Noetherian  $\mathfrak{o}$ -modules. Thus if  $N$  is Noetherian, both  $M$  and  $P$  are.

*Proof.* Let  $\rho$  be the map  $M \rightarrow N$  and let  $\pi$  be the map  $N \rightarrow P$ . Suppose that  $N$  is Noetherian. Every submodule of  $M$  is then isomorphic to a submodule of  $N$  via  $\rho$  and is therefore finitely generated over  $\mathfrak{o}$ . If  $Q$  is a submodule of  $P$ , then  $\pi^{-1}(Q)$  is a finitely generated  $\mathfrak{o}$ -submodule of  $N$  since  $N$  is Noetherian, and therefore  $Q$  itself must be finitely generated over  $\mathfrak{o}$ .

Suppose  $P$  and  $M$  are both Noetherian  $\mathfrak{o}$ -modules. Let  $\{N_i | i = 1, 2, 3, \dots\}$ , denote an ascending chain of  $N$ -submodules. We shall prove this chain eventually stabilizes, thereby proving  $N$  is Noetherian. We have that the chains  $\{\pi(N_i)\}$ ,  $\{\rho(M) \cap N_i\}$ , stabilize after their  $n$ th term for some  $n$ . We claim that  $\{N_i\}$  must stabilize after its  $n$ th term. Let  $m \leq n$ , we shall prove  $N_m \subset N_n$ . Let  $a \in N_m$ . We can find  $b \in N_n$  such that  $\pi(a) = \pi(b)$ . Therefore, by exactness of the sequence, we have  $a - b = \rho(c)$  for some  $c \in M$ . Hence  $a - b \in \rho(M) \cap N_m = \rho(M) \cap N_n$  so that  $a \in N_n$  as required. This completes the proof.  $\square$

An easily corollary of this result is that a finitely generated module over a Noetherian ring is Noetherian.

**Proposition 1.10** Let  $\mathfrak{o}$  be a Noetherian ring. Then the polynomial ring  $\mathfrak{o}[X]$  is a Noetherian ring.

*Proof.* Let  $\mathfrak{a}$  be an  $\mathfrak{o}[X]$  ideal. Let  $\mathfrak{b}$  be the  $\mathfrak{o}$ -ideal of leading coefficients of elements in  $\mathfrak{a}$ . By the Noetherianness of  $\mathfrak{o}$ , let  $\{a_i | i = 1, \dots, n\}$ , be a finite set of generators for  $\mathfrak{b}$  over  $\mathfrak{o}$  and let  $a_i$  be the leading coefficient of  $f_i \in \mathfrak{a}$ ,  $d_i = \deg(f_i)$ ,  $d = \max_i(d_i)$ . Let  $\mathfrak{c}$  be the  $\mathfrak{o}$ -module of elements in  $\mathfrak{a}$  with degree less than or equal to  $d$ , together with 0. This is a Noetherian  $\mathfrak{o}$ -module; let  $\{g_j | j = 1, \dots, m\}$ , be a set of generators for  $\mathfrak{c}$  over  $\mathfrak{o}$ . Let  $\mathfrak{d}$  be the  $\mathfrak{o}[X]$  ideal generated by all the  $f_i$  and  $g_j$ . Evidently  $\mathfrak{d} \subset \mathfrak{a}$ . On the other hand let  $h \in \mathfrak{a}$ . We shall prove by induction on the degree of  $h$  that  $h \in \mathfrak{d}$ . Let  $m$  be the degree of  $h$ . First suppose  $m \leq d$ . Then it is clear that  $h \in \mathfrak{d}$ . Suppose  $m > d$ , and let  $a$  be the leading coefficient of  $h$ . Let  $a = \sum_{i=1}^n a_i \lambda_i$  with  $\lambda_i \in \mathfrak{o}$ . Then  $h - \sum_i \lambda_i X^{m-d_i} f_i$  lies in  $\mathfrak{a}$  and has degree less than  $m$ , so the result follows by induction.  $\square$

A corollary of this result is that a finitely generated ring extension of a Noetherian ring is a Noetherian ring.

We now prove that the set of elements integral over  $\mathfrak{o}$  in a given extension ring of  $\mathfrak{o}$  form a ring.

**Proposition 1.11** An element  $a$  in an extension ring of  $\mathfrak{o}$  is integral over  $\mathfrak{o}$  if and only if the ring  $\mathfrak{o}[a]$  is finitely generated as a  $\mathfrak{o}$ -module.

*Proof.* Suppose that  $a$  is integral over  $\mathfrak{o}$ , so that we have  $f(a)=0$ , for some monic polynomial  $f = X^n + a_1 X^{n-1} + \dots + a_n$ . We prove by induction that for  $m \geq n - 1$ , we have  $(1, a, \dots, a^m)\mathfrak{o} = (1, a, \dots, a^{n-1})\mathfrak{o}$ . The case  $m = n - 1$  is trivial. Suppose the result holds for  $m - 1 \geq n$ . We have  $a^{m-n} f(a) = 0$  and so  $a^m = -a_1 a^{m-1} + \dots - a_n a^{m-n}$  and therefore

$a^m \in (1, a, \dots, a^{m-1})\mathfrak{o} = (1, a, \dots, a^{n-1})\mathfrak{o}$ . On the other hand, suppose that  $\mathfrak{o}[a]$  is generated over  $\mathfrak{o}$  by  $f_1(a), \dots, f_r(a)$  with  $f_i(x) \in \mathfrak{o}[X]$ . Put  $n = \max_i(\deg(f_i))$ . Then by our supposition we can write  $a^{n+1} = \sum a_i f_i(a)$  with  $a_i \in \mathfrak{o}$ , and then  $a$  is a root of the monic polynomial  $X^{n+1} - \sum a_i f_i(X)$ .  $\square$

**Proposition 1.12** An element  $a$  of an extension ring of  $\mathfrak{o}$  is integral over  $\mathfrak{o}$  if and only if  $a$  lies in some extension ring  $\mathfrak{R}$  of  $\mathfrak{o}$  which is finitely generated as an  $\mathfrak{o}$ -module.

*Proof.* The “only if” part follows easily from the previous proposition. On the other hand, suppose that  $\mathfrak{R} = (r_1, \dots, r_n)\mathfrak{o}$ . Then we can write  $r_i a = \sum_{j=1}^n r_j a_{ji}$  with the  $a_{ij} \in \mathfrak{o}$ . Let  $A$  be the matrix  $a1_n - (a_{ji})$ , and let  $AA^* = d.1_n$ . Then, letting  $\mathbf{r}$  be the vector whose  $i$ th entry is  $r_i$ , we have  $\mathbf{r}AA^* = 0$ . Therefore  $r_i d = 0$  for all  $i$ . Since  $1 \in \mathfrak{R} = (r_1, \dots, r_n)\mathfrak{o}$ , this implies  $d = 0$ . Therefore  $a$  is a root of the monic polynomial  $\det(X.1_n - (a_{ji}))$ .  $\square$

**Proposition 1.13** Let  $a_1, \dots, a_n$  all be integral over  $\mathfrak{o}$ ; then  $\mathfrak{o}[a_1, \dots, a_n]$  is a finitely generated  $\mathfrak{o}$ -module.

*Proof.* We proceed by induction on  $n$ . The case where  $n = 1$  follows from Proposition 1.11. Suppose that  $\mathfrak{o}[a_1, \dots, a_{n-1}]$  is a finitely generated module over  $\mathfrak{o}$ . By Proposition 1.11  $\mathfrak{o}[a_n]$  is a finitely generated  $\mathfrak{o}$ -module, and it follows that  $\mathfrak{o}[a_1, \dots, a_n]$  is a finitely generated  $\mathfrak{o}[a_1, \dots, a_{n-1}]$ -module, and so since  $\mathfrak{o}[a_1, \dots, a_{n-1}]$  is a finitely generated  $\mathfrak{o}$ -module it now follows that  $\mathfrak{o}[a_1, \dots, a_n]$  is a finitely generated  $\mathfrak{o}$ -module.  $\square$

As a corollary of the last three propositions we have that the set of elements integral over  $\mathfrak{o}$  forms a ring. For if  $a, b$  are integral over  $\mathfrak{o}$  then their sum, difference, and product are both contained in  $\mathfrak{o}[a, b]$ , which is a finitely generated  $\mathfrak{o}$ -module by Proposition 1.13; it then follows that they are all integral over  $\mathfrak{o}$  by Proposition 1.12.

We can also deduce that the integral closure of  $\mathfrak{o}$  in an extension ring is integrally closed. For suppose  $\mathfrak{R} \supset \mathfrak{D} \supset \mathfrak{o}$  is a tower of ring extensions, and  $\mathfrak{D}$  is a finitely generated  $\mathfrak{o}$ -module,

and  $r \in \mathfrak{R}$  is integral over  $\mathfrak{D}$ . Then  $\mathfrak{D}[r]$  is a finitely generated  $\mathfrak{D}$ -module by Proposition 1.11, and  $\mathfrak{D}$  is a finitely generated  $\mathfrak{o}$ -module by hypothesis. Therefore  $\mathfrak{D}[r]$  is a finitely generated  $\mathfrak{o}$ -module, so that by Proposition 1.12,  $r$  is integral over  $\mathfrak{o}$ .

**Proposition 1.14** Let  $F$  be a field. An indecomposable finite-dimensional commutative  $F$ -algebra  $A$  with  $\text{Rad}(A)=\{0\}$  is a field.

*Proof.* Let  $a \in A, a \neq 0$ . Then  $a^n A$  is an ideal of  $A$ , and hence an  $F$ -subspace, and  $a^{n+1}A \subset a^n A$ . We have  $\dim_F(A)$  is finite and  $\dim_F(a^{n+1}A) \leq \dim_F(a^n A)$ , so that we must have  $a^{n+1}A = a^n A$  for some  $n$ . Then  $a^{2n}A = a^n A$ , so that  $a^n = a^{2n}b$  for some  $b \in A$ . Hence  $a^n b = (a^n b)^2$ , so that  $a^n b$  is an idempotent. Since  $A$  is indecomposable,  $a^n b$  must be a primitive idempotent, that is, it must be impossible to write it as a sum of nonzero orthogonal idempotents, orthogonal idempotents being ones which multiply to zero. Likewise  $1$  must be a primitive idempotent of  $A$ , and we now prove that  $a^n b$  must be either  $1$  or  $0$  by showing that distinct primitive idempotents multiply to zero. Let  $e', e''$  be two distinct primitive idempotents of  $A$ . If  $e' = e' e''$  and  $e'' = e' e''$  then clearly  $e' = e''$ . So, interchanging  $e'$  and  $e''$  if necessary, we may assume  $e' \neq e' e''$ , and then  $e' = e' e'' + e'(1 - e'')$ . Since the two right-hand terms are orthogonal idempotents and  $e'$  is primitive we deduce  $e' e'' = 0$ . We have now shown that  $a^n b$  equals  $0$  or  $1$ . If  $a^n b = 0$ , then  $a^n = a^{2n}b = a^n(a^n b) = 0$ , and so  $a = 0$  since  $\text{Rad}(A)=\{0\}$ . However, if  $a^n b = 1$ , then  $a$  has a multiplicative inverse. This shows that  $A$  is a field.  $\square$

As a corollary to this we have that if  $A$  is a finite-dimensional commutative  $F$ -algebra with  $\text{Rad}(A)=\{0\}$  then  $A \simeq \prod A_i$  with each  $A_i$  a field.

We now prove

**Proposition 1.15** Let  $\mathfrak{o}$  be a Dedekind domain with field of fractions  $K$  and let  $L/K$  be a finite separable extension, with  $\mathfrak{D}$  the integral closure of  $\mathfrak{o}$  in  $L$ . Then  $\mathfrak{D}$  is a finitely generated  $\mathfrak{o}$ -module which spans  $L$  over  $K$ , and is a Dedekind domain.

*Proof.* Let  $\alpha \in \mathfrak{D}$ . Then for some  $a_i, b_i \in \mathfrak{o}$  with  $b_i \neq 0$  for all  $i$ ,  $\alpha^m + \frac{a_{m-1}}{b_{m-1}}\alpha^{m-1} + \dots + \frac{a_1}{b_1}\alpha + \frac{a_0}{b_0} = 0$ . Then writing  $b = \prod_{i=0}^{m-1} b_i$ , we get  $(\alpha b)^m + (a_{m-1} \frac{b}{b_{m-1}})(\alpha b)^{m-1} + \dots + (a_1 \frac{b^{m-1}}{b_1})\alpha b + a_0 \frac{b^m}{b_0} = 0$ . Thus  $\alpha b$  is a root of a monic polynomial in  $\mathfrak{o}[x]$ . Thus  $\alpha b$  lies in  $\mathfrak{D}$ , so that  $\mathfrak{D}$  spans  $L$  over  $K$ .

For any  $\mathfrak{o}$ -submodule  $X$  of  $L$  which spans  $L$  over  $K$ , let  $X^D = \{x \in L \mid t_{L/K}(xy) \in \mathfrak{o} \forall y \in X\}$ . If  $X$  is a free  $\mathfrak{o}$ -module on a basis  $\{x_i \mid 1 \leq i \leq (L : K)\}$ , then we have, by the non-singularity of  $(x, y) \mapsto t_{L/K}(xy)$ , proved after Proposition 1.8, that there exist unique elements  $y_i \in L$  with  $t_{L/K}(x_i y_j) = \delta_{ij}$ . Now, clearly the  $y_i$  are all linearly independent and all lie in  $X^D$ . Suppose  $t_{L/K}(x_i y) = c_i \in \mathfrak{o}$  for each  $i$ . Then  $t_{L/K}(x_i(y - \sum_j c_j y_j)) = 0$  for each  $i$ . By non-singularity of the trace form,  $y = \sum c_j y_j$ , thus  $X^D$  is spanned by the  $y_j$ .

Let  $N$  be a Galois extension of  $K$  containing a copy of  $L$ , and let  $\{\sigma\}$  denote the set of embeddings  $L \rightarrow N$  over  $K$ . If  $x \in \mathfrak{D}$ , then all the  $x^\sigma$  are integral over  $\mathfrak{o}$ , hence so is  $\sum_\sigma x^\sigma$ , which equals  $t_{L/K}(x)$  by definition. Thus we have shown that  $t_{L/K}(\mathfrak{D}) \subset \mathfrak{o}$ . We can infer that  $\mathfrak{D} \subset \mathfrak{D}^D$ . Now since  $\mathfrak{D}$  spans  $L$  over  $K$ , it contains a basis  $\{x_i\}$  of  $L$  over  $K$ . Then the  $\mathfrak{o}$ -module  $X$  generated by the  $x_i$  is a free  $\mathfrak{o}$ -module, and  $X \subset \mathfrak{D}$ , whence  $\mathfrak{D} \subset \mathfrak{D}^D \subset X^D$ . Now  $X^D$  is a free  $\mathfrak{o}$ -module on  $(L:K)$  generators, and  $\mathfrak{o}$  is a Noetherian ring, so by the corollary to Proposition 1.9  $X^D$  is a Noetherian  $\mathfrak{o}$ -module, hence  $\mathfrak{D}$  is finitely generated over  $\mathfrak{o}$ . Hence, by the corollary to Proposition 1.10,  $\mathfrak{D}$  is a Noetherian ring. By the second corollary to Proposition 1.13, we see that the integral closure of  $\mathfrak{o}$  in  $L$  is integrally closed in  $L$ ; furthermore  $\mathfrak{D} \cap K = \mathfrak{o}$  since  $\mathfrak{o}$  is integrally closed in  $K$ .

Now let  $\mathfrak{B}$  be a prime ideal of  $\mathfrak{D}$ , and let  $a \in \mathfrak{B} \setminus \{0\}$ . Since  $a$  is integral over  $\mathfrak{o}$ , we have  $a^n + b_{n-1}a^{n-1} + \dots + b_1 a + b_0 = 0$  for some  $n$  and some  $b_{n-1}, \dots, b_0 \in \mathfrak{o}$  with  $b_0 \neq 0$ . But then  $b_0 \in \mathfrak{B} \cap \mathfrak{o}$ , and so  $\mathfrak{p} = \mathfrak{B} \cap \mathfrak{o}$  is non-zero and is a prime ideal of  $\mathfrak{o}$ . Clearly  $\mathfrak{B} \supset \mathfrak{p}\mathfrak{D}$ .

Now consider any prime ideal  $\mathfrak{p}$  of  $\mathfrak{o}$ . Since  $\mathfrak{D} \cap K = \mathfrak{o}$ , we have  $\mathfrak{p}^{-1} \not\subset \mathfrak{D}$  and so  $\mathfrak{D} \not\supseteq \mathfrak{p}\mathfrak{D}$ . Since  $\mathfrak{D}$  is a finitely generated  $\mathfrak{o}$ -module, we have that  $A = \mathfrak{D}/\mathfrak{p}\mathfrak{D}$  is a finite-dimensional

commutative  $\mathfrak{o}/\mathfrak{p}$ -algebra. By the corollary to Proposition 1.14,  $A$  has finitely many maximal ideals  $J_i$ , and the inverse image  $\mathfrak{B}_i$  of  $J_i$  under the quotient map  $\mathfrak{D} \rightarrow A$  is a maximal ideal of  $\mathfrak{D}$ . Indeed, for each  $i$ ,  $\mathfrak{D}/\mathfrak{B}_i \simeq A/J_i$  is a field with finite degree over  $\mathfrak{o}/\mathfrak{p}$ . Moreover, if  $\mathfrak{B}$  is a prime ideal of  $\mathfrak{D}$  which contains  $\mathfrak{p}\mathfrak{D}$ , then  $\mathfrak{D}/\mathfrak{B}$  is an integral domain and is a finite-dimensional algebra over the field  $\mathfrak{o}/\mathfrak{p}$ ; hence by Proposition 1.14 is a field;  $\mathfrak{B}$  is therefore maximal and so coincides with one of the above  $\mathfrak{B}_i$ .  $\square$

As a corollary we have that the ring of algebraic integers in an algebraic number field is a Dedekind domain.

We must now define the notion of a valuation.

**Definition 1.1** A *valuation* on a field  $K$  is a mapping  $v : K \rightarrow \mathbb{R}^+$  satisfying the following three properties:

- (i)  $v(x)=0$  if and only if  $x=0$
- (ii)  $v(xy)=v(x)v(y)$
- (iii)  $v(x + y) \leq v(x) + v(y)$

If we even have  $v(x + y) \leq \max(v(x),v(y))$ , then the valuation is said to be **non-archimedean**, otherwise **archimedean**.

**Proposition 1.16** A valuation  $v$  on a field  $K$  induces a topology on that field via the metric  $d(x, y) = v(x - y)$ .

*Proof.* We follow [2], chapter 1, section 1, p. 2. We have  $(v(-1))^2 = 1$ , whence  $v(-1) = 1$ , whence  $v(x - y) = v(y - x)$ . It follows that  $d(x, y) = d(y, x)$ , and that it satisfies the other requirements for a metric is easily checked.

We have  $v((x+y)-(x_0+y_0)) \leq v(x-x_0)+v(y-y_0)$ , and we have  $v((-y)-(-y_0)) = v(y-y_0)$ , and the continuity of addition and negation follow.

We have

$$v(xy - x_0y_0) \leq v(x - x_0)v(y - y_0) + v(x - x_0)v(y_0) + v(x_0)v(y - y_0),$$

whence the continuity of multiplication. And if  $v(x - x_0) < \frac{1}{2}\min(v(x_0), \epsilon(v(x_0))^2)$ , then

$$v(x^{-1} - x_0^{-1}) = \frac{v(x_0 - x)}{v(x)v(x_0)} < \frac{(\epsilon/2)(v(x_0))^2}{(v(x_0) - v(x - x_0))(v(x_0))} < \frac{(\epsilon/2)(v(x_0))^2}{(1/2)(v(x_0))^2} = \epsilon,$$

whence the continuity of multiplicative inversion. It follows that the metric makes  $K$  into a topological field.  $\square$

Given a field  $K$  and a valuation  $v$ , we may consider  $K$  as endowed with the metric induced by  $v$ . The field  $K$  is said to be complete with respect to the valuation  $v$  if and only if it is complete as a metric space with respect to the metric induced by  $v$ . If it is not complete, we may complete it as a metric space in the usual manner, by passing to the set of equivalence classes of convergent Cauchy sequences. It may also be seen without much difficulty how to make this completion into a valued field  $K_v$  which extends the original field  $K$  and is complete with respect to its valuation.

Two valuations  $\phi$  and  $\psi$  are said to be equivalent if  $\phi = \psi^\rho$  for some positive real number  $\rho$ . Every field  $K$  admits the trivial valuation  $v$  such that  $v(0)=0$  and  $v(x)=1$  whenever  $x \neq 0$ .

We now describe the valuations on  $\mathbb{Q}$ . For any prime  $p$ , we obtain a valuation  $v$  on  $\mathbb{Q}$  by setting  $v(\frac{m}{n}p^r) = p^{-r}$  when  $m, n$  are coprime to  $p$ ; this is called the  $p$ -adic valuation on  $\mathbb{Q}$ .

We now have

**Proposition 1.17** Every archimedean valuation  $v$  on  $\mathbb{Q}$  is equivalent to the usual absolute value. Every nontrivial nonarchimedean valuation  $v$  on  $\mathbb{Q}$  is equivalent to the  $p$ -adic valuation for some  $p$ .

*Proof.* We follow [5], chapter II, section 1. Let  $v$  be any nontrivial valuation on  $\mathbb{Q}$ . Let  $m$  and  $n$  be integers  $> 1$ . We may write  $m = a_0 + a_1n + \dots + a_rn^r$  with  $a_i$  an integer,  $0 \leq a_i < n$

and  $n^r \leq m$ . Let  $N = \max(1, v(n))$ . We get  $v(m) \leq \sum v(a_i)v(n)^i \leq \sum v(a_i)N^r$ . The number  $r$  satisfies  $r \leq \log m / \log n$  and the numbers  $a_i$  are less than  $n$ , so  $v(a_i) < n$ . Substituting these results into the previous inequality, we get  $v(m) \leq (1 + \log m / \log n)nN^{\log m / \log n}$ . Replace  $m$  by  $m^s$  and raise both sides to the power  $1/s$ , with  $s$  an integer. We get  $v(m) \leq (1 + s \log m / \log n)^{1/s} n^{1/s} N^{\log m / \log n}$ . Let  $s$  increase without bound. The terms in the inequality involving  $s$  converge to 1. We get  $m \leq N^{\log m / \log n}$ . We now have two cases.

*Case 1.* If for some  $n > 1$  we have  $v(n) \leq 1$ , then  $N=1$  and so  $v(m) \leq 1$  for all integers  $m$ , whence the valuation is nonarchimedean. Then let  $R = \{x \in \mathbb{Q} | v(x) \leq 1\}$  be the valuation ring. The units in this ring are precisely those  $x$  so that  $v(x)=1$ , so that the unique maximal ideal is  $\{x \in \mathbb{Q} | v(x) < 1\}$ , let this ideal be  $\mathfrak{B}$ . Then we have  $\mathbb{Z} \subset R$  and  $\mathfrak{B} \neq \{0\}$  because the valuation is not trivial. Hence  $\mathfrak{B} \cap \mathbb{Z} = (p)$  is a prime ideal. If  $m$  is in  $\mathbb{Z}$  but not in  $(p)$  then  $m$  is a unit in  $R$  and  $v(m)=1$ . Thus  $v(mp^r) = v(p)^r$  and this valuation on  $\mathbb{Q}$  is equivalent to the  $p$ -adic valuation.

*Case 2.* We have  $n > 1$  implies  $v(n) > 1$ . Then we always have  $N = v(n)$  and so the inequality becomes  $v(m)^{1/\log m} \leq v(n)^{1/\log n}$ . We may reverse the roles of  $m$  and  $n$  to obtain the inequality in the reverse direction. Thus  $c = v(m)^{1/\log m} = v(n)^{1/\log n}$  for some constant  $c$ , for all integers  $m$  and  $n$ , and it follows that  $v(x) = c^{\log x}$  for all positive rationals  $x$ . Therefore clearly  $v(x) = c^{\log |x|}$  for general rationals  $x$ . Thus the valuation is equivalent to the ordinary absolute value.  $\square$

We now classify the archimedean and nonarchimedean valuations on an algebraic number field.

**Proposition 1.18** Let  $K$  be a field which is complete with respect to an archimedean valuation. Then  $K$  is isomorphic either to the real or the complex field and the valuation is equivalent to the usual absolute value.

*Proof.* We follow [5], chapter II, section 4. First we note that a valuation  $v$  is non-archimedean if and only if it is bounded on  $\{n1\}$  as  $n$  runs through  $\mathbb{Z}$ . For if  $v$  satisfies  $v(x+y) \leq \max(v(x), v(y))$  then we have  $v(n1) = v(1+1+\dots+1) \leq v(1)$ , so that the values on  $\{n1\}$  are bounded. On the other hand, suppose  $v(n1) \leq N$  for all integers  $n$ . For any  $x, y \in K$  we have for any positive integer  $n$ ,  $v(x+y)^n = v(\sum_r \binom{n}{r} x^r y^{n-r}) \leq \sum_r v(\binom{n}{r} v(x)^r v(y)^{n-r})$ . Now if  $v(x) \geq v(y)$  then  $v(x)^r v(y)^{n-r} \leq v(x)^n$ . Since  $\binom{n}{r}$  is an integer, we get

$$v(x+y)^n \leq N(n+1)\max(v(x), v(y))^n;$$

whence

$$v(x+y) \leq N^{1/n}(n+1)^{1/n}\max(v(x), v(y)).$$

Since this is true for all positive  $n$  we conclude that  $v(x+y) \leq \max(v(x), v(y))$ .

To return to our proof of Proposition 1.18, suppose  $K$  is a field complete with respect to an archimedean valuation  $v$ . Since the valuation is archimedean, the values  $v(n)$  for  $n \in \mathbb{Z}$  are unbounded. Therefore  $K$  has characteristic zero and the restriction of  $v$  to  $\mathbb{Q}$  must be an archimedean valuation on  $\mathbb{Q}$  and therefore equivalent to the ordinary absolute value. Replacing the valuation by an equivalent one, we may assume the valuation is the ordinary absolute value on  $\mathbb{Q}$ .

Since the field  $K$  is complete, it must contain the completion of  $\mathbb{Q}$ . This will be a copy of  $\mathbb{R}$  with the usual absolute value; let us identify it with  $\mathbb{R}$ . If  $K$  does not contain a root of  $X^2 + 1$ , we adjoin one to  $K$ , and extend the valuation to the enlarged field by the formula  $v(a+ib) = (v(a)^2 + v(b)^2)^{1/2}$ . It can easily be checked that this gives a valuation on  $K(i)$  and that  $K(i)$  is complete under this valuation. Thus it suffices to prove the theorem on the assumption that  $K$  contains a copy of  $\mathbb{C}$ . First of all, we have proved that the valuation agrees with the usual absolute value on  $\mathbb{R}$ ; we show that it agrees with the usual absolute value on  $\mathbb{C}$  also. We have  $v(i) = 1$  since  $i^4 = 1$ . Thus for  $\alpha = a+ib \in \mathbb{C}$  we have  $v(\alpha) = v(a+ib) \leq v(a) + v(b) \leq \sqrt{2}(a^2 + b^2)^{1/2}$ . The function  $f(a+ib) = v(a+ib)/(a^2 + b^2)^{1/2}$

for  $a + ib \neq 0$  is thus bounded by  $\sqrt{2}$ . Since  $f(\alpha^n) = f(\alpha)^n$  it follows that  $f(\alpha) \leq 1$ . Furthermore we have  $f(\alpha^{-1}) = f(\alpha)^{-1}$ , so that we get  $f(\alpha) \geq 1$  for all  $\alpha \neq 0$  too. Thus  $f(\alpha) = 1$  as required.

Now, we have  $\mathbb{C} \subseteq K$  and it is necessary to show equality. Suppose  $z \in K$  but  $z \notin \mathbb{C}$ . Let  $m = \inf_{\alpha \in \mathbb{C}} v(z - \alpha)$ . For any positive number  $\epsilon$ , the set of complex numbers  $\alpha$  for which  $v(z - \alpha) \leq m + \epsilon$  is contained in the set  $\{\beta \in \mathbb{C} | v(\beta) \leq m + \epsilon + v(z)\}$ . This is a disc and the function  $f(\beta) = v(z - \beta)$  is a continuous function from the disc into  $\mathbb{R}$ . Therefore the minimum of this function is attained at some  $\alpha_0 \in \mathbb{C}$ . Since  $z \notin \mathbb{C}$ , we have  $z - \alpha_0 \notin \mathbb{C}$ . Replacing our original  $z$  with  $z - \alpha_0$ , we get (a)  $z \notin \mathbb{C}$  and (b)  $m = v(z) \leq v(z - \alpha), \alpha \in \mathbb{C}$ . We have  $m = v(z) \neq 0$  because  $z \notin \mathbb{C}$ .

Let  $n$  denote any positive integer and  $\omega$  a primitive  $n$ th root of unity in  $\mathbb{C}$ . We have  $v(z - \alpha)v(z - \omega\alpha)\dots v(z - \omega^{n-1}\alpha) = v(z^n - \alpha^n) \leq v(z^n) + v(\alpha^n)$ . Each term  $v(z - \omega^i\alpha) \geq m$  so  $v(z - \alpha)m^{n-1} \leq v(z)^n(1 + v(\alpha)^n/v(z)^n) = m^n(1 + v(\alpha)^n/m^n)$ . Therefore  $v(z - \alpha) \leq m(1 + v(\alpha)^n/m^n)$ . This holds for all  $n$ , so for  $v(\alpha) < m$ , let  $n$  increase without bound. Then we get  $v(z - \alpha) \leq m$ , and by minimality of  $m$  we have  $v(z - \alpha) = m$ . Thus when we replace  $z$  by  $z - \alpha$ , the conditions (a) and (b) above are satisfied. We may then repeat the above procedure and get  $v(z - \alpha - \beta) = m$  whenever  $\beta \in \mathbb{C}$  and  $v(\beta) < m$ . Then by induction we get for any positive integer  $n$ ,  $v(z - n\alpha) = m, \alpha \in \mathbb{C}, v(\alpha) < m$ . It follows that  $v(z - \alpha) = m$  for all  $\alpha \in \mathbb{C}$  since  $m \neq 0$ . Then, for any  $\alpha, \beta \in \mathbb{C}$ , we get  $|\alpha - \beta| \leq v(z - \alpha) + v(z - \beta) = 2m$ , which is clearly false in general. We have derived a contradiction from supposing that  $K \neq \mathbb{C}$ , so we conclude  $K = \mathbb{C}$ .  $\square$

A corollary of this result is that all the archimedean valuations on an algebraic number field are obtained from the ordinary absolute value via embeddings into the real or complex field.

Before classifying the nonarchimedean valuations on an algebraic number field, we need a few results about localizations of rings. We follow [5], chapter I. Given an integral domain  $R$  and a multiplicative subset  $S$  of  $R$  - i.e. a subset of  $R$  which does not contain zero and is closed under multiplication - the *localization* of  $R$  at  $S$ ,  $R_S$ , is defined to be the following ring. For the underlying set we take the equivalence classes of ordered pairs  $(r, s) \in R \times S$  modulo the equivalence relation whereby  $(r, s)$  and  $(q, t)$  are equivalent if and only if  $qs = rt$ . The equivalence class of  $(r, s)$  is written  $r/s$ . Addition and multiplication are defined by  $r/s + r'/s' = (rs' + r's)/ss'$ ,  $(r/s) \cdot (r'/s') = rr'/ss'$ . It is trivial to verify that these operations are well-defined and make  $R_S$  an integral domain.

**Proposition 1.19** Let  $R$  be an integral domain and  $S$  a multiplicative subset of  $R$ . There is a one-to-one correspondence between the prime ideals of  $R_S$  and the prime ideals of  $R$  that have empty intersection with  $S$ . Under this correspondence, a prime  $\mathfrak{B}$  of  $R$  is associated with the ideal  $\mathfrak{B}R_S$  in  $R_S$ .

*Proof.* Let  $\mathfrak{Q}$  be a prime ideal in  $R_S$ . We immediately see that  $\mathfrak{B} = \mathfrak{Q} \cap R$  is a prime ideal of  $R$ . Then  $\mathfrak{B}R_S$  is an ideal of  $R_S$  contained in  $\mathfrak{Q}$ . We must show that these are equal. Let  $q/s$  be any element in  $\mathfrak{Q}$  with  $q \in R$  and  $s \in S$ . Then  $q = (q/s)s \in R \cap \mathfrak{Q} = \mathfrak{B}$ . Thus  $q/s \in \mathfrak{B}R_S$  since  $q(1/s) = q/s$ ,  $q \in \mathfrak{B}$ ,  $1/s \in R_S$ . Thus we have proved that every prime ideal in  $R_S$  has the form  $\mathfrak{Q} = \mathfrak{B}R_S$  with  $\mathfrak{B} = \mathfrak{Q} \cap R$  uniquely determined by  $\mathfrak{Q}$ . Since every element in  $S$  has an inverse in  $R_S$  we know  $\mathfrak{Q} \cap S$  is empty. Thus  $\mathfrak{B} \cap S$  is empty.

Now let us start with a prime ideal  $\mathfrak{B}$  of  $R$  which has no elements in  $S$ . Let  $\mathfrak{Q} = \mathfrak{B}R_S$ . This is an ideal of  $R_S$ ; we shall show it is prime. Suppose  $a, b$  are elements of  $R_S$  with  $ab \in \mathfrak{Q}$ , then  $ab = x/s$  for some  $x \in \mathfrak{B}$  and  $s \in S$ . Suppose  $a = r_1/s_1, b = r_2/s_2$  with  $r_1, r_2 \in R$  and  $s_1, s_2 \in S$ . We have that  $r_1r_2s = xs_1s_2 \in \mathfrak{B}$ . Thus either  $r_1, r_2$ , or  $s \in \mathfrak{B}$  because  $\mathfrak{B}$  is prime. Also  $s \notin \mathfrak{B}$  by choice of  $\mathfrak{B}$ . Thus either  $r_1$  or  $r_2$  belongs to  $\mathfrak{B}$  and so either  $a$  or  $b$  is in  $\mathfrak{Q}$ . Therefore  $\mathfrak{Q}$  is prime. Now we prove  $\mathfrak{Q} \cap R = \mathfrak{B}$ . If  $u \in \mathfrak{Q} \cap R$  then  $u = x/s$  with  $x \in \mathfrak{B}$

because  $\mathfrak{Q} = \mathfrak{B}R_s$ . But we also have  $u \in R$  and so  $x = us$  implies that  $u$  or  $s$  is in  $\mathfrak{B}$ . Since  $s$  is not, we have  $u \in \mathfrak{B}$ . Thus the correspondences  $\mathfrak{B} \rightarrow \mathfrak{B}R_s$  and  $\mathfrak{Q} \rightarrow \mathfrak{Q} \cap R$  are inverse to one another and the proposition is proved.  $\square$

**Proposition 1.20** Suppose  $R'$  is an integral domain extending  $R$ ,  $b \in R'$ , and there exists in  $R'$  an  $R[b]$ -module  $M$  such that  $M$  is finitely generated over  $R$  and the only element  $y \in R[b]$  for which  $yM = 0$  is  $y = 0$ . Then  $b$  is integral over  $R$ .

*Proof.* Let  $m_1, \dots, m_n$  be a set of generators for  $M$  over  $R$ . Let  $r_{ij}$  be elements of  $R$  such that  $bm_i = \sum_j r_{ij}m_j$ . Then we get  $0 = \sum_j (r_{ij} - b\delta_{ij})m_j$ , and if we let  $A$  be the matrix  $(r_{ij} - \delta_{ij})$  and  $d = \det(A)$ , then by a similar argument to the proof of Proposition 1.12 we get  $dM = 0$  and so  $d = 0$ . Considering the polynomial  $\det(XI - r_{ij}) = f(X)$  we see that this is a monic polynomial with coefficients in  $R$  of which  $b$  is a root, so that  $b$  is integral over  $R$ .  $\square$

A *local ring* is a ring with precisely one maximal ideal. A *discrete valuation ring*, or DVR, is a principal ideal domain with precisely one maximal ideal. We now prove that for every nonzero prime ideal  $\mathfrak{B}$  of a Dedekind domain  $R$ ,  $R_{\mathfrak{B}}$  is a DVR.

**Proposition 1.21** Given a Dedekind domain  $R$  with a nonzero prime ideal  $\mathfrak{B}$ ,  $R_{\mathfrak{B}}$  is a DVR.

*Proof.* Let  $\mathfrak{B}$  be a maximal ideal of  $R$ . Then  $R_{\mathfrak{B}}$  is a Noetherian local ring with  $\mathfrak{B}R_{\mathfrak{B}}$  the only nonzero prime ideal. We have that  $R$  is integrally closed and it is straightforward to deduce that  $R_{\mathfrak{B}}$  is integrally closed. We must check that these conditions imply that  $R_{\mathfrak{B}}$  is a PID.

Selected any  $a \neq 0$  in  $\mathfrak{B}$  and let  $M = R/Ra$ . For each  $m \in M$  let  $\text{ann}(m) = \{r \in R \mid rm = 0\}$ . Since  $R$  is Noetherian there is a maximal element in the collection  $\{\text{ann}(m) \mid m \neq 0, m \in M\}$ . Let  $b$  be an element of  $R$  such that  $\mathfrak{Q} = \text{ann}(b + Ra)$  is such a maximal element.  $\mathfrak{Q}$

is nonzero because  $a \neq 0$  and  $a \in \mathfrak{Q}$ . If  $\mathfrak{Q}$  is not prime, let  $x, y \notin \mathfrak{Q}$  with  $xy \in \mathfrak{Q}$ . Then we have  $y(b + Ra) \neq 0 + Ra$  because  $y \notin \mathfrak{Q}$ . Then  $\text{ann}(yb + Ra)$  contains both  $\mathfrak{Q}$  and  $x$ , contradicting the maximal choice of  $\mathfrak{Q}$ . Therefore  $\mathfrak{Q}$  is prime, but  $R$  has only one nonzero prime ideal, so  $\mathfrak{B}$  must be the set of all elements which multiply  $b$  into  $Ra$ . Thus  $\mathfrak{B}b \subseteq Ra$  but  $b \notin Ra$ . Clearly  $b/a \notin R$ , since  $b + Ra \neq 0 + Ra$ . Now since  $\mathfrak{B}b \subseteq Ra$ ,  $\mathfrak{B}b/a$  is an ideal in  $R$ . If  $\mathfrak{B}b/a \subset \mathfrak{B}$ , then by Proposition 1.20  $b/a$  is integral over  $R$ , whence  $b/a \in R$  contradicting what we observed earlier. Thus  $\mathfrak{B}b/a = R$  and  $\mathfrak{B} = Ra/b$ . We now know the maximal ideal is principal, let us write  $\mathfrak{B} = R\pi$ . Now let  $\mathfrak{U}$  be any nonzero ideal, and consider the chain  $\mathfrak{U} \subset \mathfrak{U}\pi^{-1} \subset \mathfrak{U}\pi^{-2} \subset \dots$ . If  $\mathfrak{U}\pi^{-k} = \mathfrak{U}\pi^{-k-1}$  then  $\pi^{-1}$  sends  $\mathfrak{U}\pi^{-k}$  to itself so  $\pi^{-1}$  is integral over  $R$ , but  $\pi^{-1} \notin R$ . Thus the chain is strictly ascending, and since  $R$  is Noetherian the part that lies in  $R$  is finite. Let  $\mathfrak{U}\pi^{-n} \subseteq R, \mathfrak{U}\pi^{-n-1} \not\subseteq R$ . If  $\mathfrak{U}\pi^{-n} \subseteq \mathfrak{B} = R\pi$  then  $\mathfrak{U}\pi^{-n-1} \subseteq R$  so we must have  $\mathfrak{U}\pi^n = R$ . Thus  $\mathfrak{U} = R\pi^n$ .  $\square$

Let  $K$  be a field with a nonarchimedean valuation  $v_p$ . Let  $R$  be the valuation and assume  $R$  is a DVR with maximal ideal  $\mathfrak{p} = \pi R$ . Let  $L$  be a finite separable extension of  $K$  and  $R'$  the integral closure of  $R$  in  $L$ . We consider the problem of finding all valuations on  $L$  which restrict on  $K$  to a valuation equivalent to  $v_p$ .

Now  $R'$  is a Dedekind domain and we have the factorization  $\mathfrak{p}R' = \pi R' = \mathfrak{B}_1^{e_1} \dots \mathfrak{B}_g^{e_g}$  with the  $\mathfrak{B}_i$  distinct primes in  $R'$ . To every prime  $\mathfrak{B}_i$  in  $R'$  we have the  $\mathfrak{B}_i$ -adic valuation  $v_{\mathfrak{B}_i}$  on  $L$  as follows.  $R'_{\mathfrak{B}_i}$  is a DVR with maximal ideal  $\mathfrak{B}_i R'_{\mathfrak{B}_i}$ , and we can select a generator  $\tau$  for the maximal ideal. Then we can set  $v_{\mathfrak{B}_i}(u) = 1$  for units  $u \in R'_{\mathfrak{B}_i}$ , and we can set  $v_{\mathfrak{B}_i}(\tau)$  to some real number  $r$  between 0 and 1. Note that the valuation can also be described as follows. Given a Dedekind domain  $R$  with field of fractions  $K$ , and a fractional ideal  $\mathfrak{M}$  in  $K$  with generators  $m_1, \dots, m_k$  we can find a common denominator for them  $s$  such that we always have  $m_i s \in R$ . Then we have the factorizations  $Rs = \prod \mathfrak{Q}_j^{b_j}, Ms = \prod \mathfrak{B}_i^{a_i}$ , where the  $\mathfrak{B}_i$  and the  $\mathfrak{Q}_j$  are prime ideals of  $R$ . Then we have  $\mathfrak{M}\mathfrak{Q}_1^{b_1} \dots \mathfrak{Q}_t^{b_t} = \mathfrak{B}_1^{a_1} \dots \mathfrak{B}_k^{a_k}$ . Then we get

$\mathfrak{M} = \prod \mathfrak{B}_i^{a_i} \cdot \prod \mathfrak{Q}_j^{b_j}$ . This shows that every fractional ideal is factorizable as a product of prime ideals with integral exponents, moreover it is easy to see that this factorization must be unique. Now we can describe  $v_{\mathfrak{B}_i}$  by saying that for  $x \in L$ ,  $v_{\mathfrak{B}_i}(x)$  is  $r$  raised to the power of the exponent with which  $B_i$  occurs in the factorization of the fractional ideal  $xR'$ . It is also easy to see that the restriction of  $v_{\mathfrak{B}_i}$  to  $K$  is equivalent to  $v_{\mathfrak{p}}$ . On the other hand, suppose  $v$  is a valuation on  $L$  whose restriction to  $K$  is equivalent to  $v_{\mathfrak{p}}$ . Let  $R_0 = \{x \in L \mid v(x) \leq 1\}$ . Then  $R_0$  is a local ring with maximal ideal  $\mathfrak{M}$  and  $\mathfrak{M} \cap R = \mathfrak{p}$ . We shall first show  $R' \subseteq R_0$ . Suppose there is an  $x \in R'$  with  $x \notin R_0$ . Then  $v(x) > 1$ . Hence  $v(x^{-1}) < 1$  so  $x^{-1} \in \mathfrak{M}$ . The element  $x$  is integral over  $R$  so there is a relation  $x^n + a_1x^{n-1} + \dots + a_n = 0$  with  $a_i \in R$ . By dividing both sides of this by  $x^n$  we obtain  $1 \in \mathfrak{M}$ , but this is impossible because  $v(1) = 1$ . Thus  $R' \subseteq R_0$ . Let  $\mathfrak{B} = R' \cap \mathfrak{M}$ .  $\mathfrak{B}$  is a prime ideal of  $R'$  which contains  $\mathfrak{p}$ . The localization  $R'_{\mathfrak{B}}$  is also in  $R_0$  because all elements in  $R'$  outside  $\mathfrak{M}$  are units in  $R_0$ . Let the maximal ideal of  $R'_{\mathfrak{B}}$  be generated by  $\tau$ . Every element in  $L$  has the form  $u\tau^n$  for some unit  $u \in R'_{\mathfrak{B}}$ . This element  $u$  is also a unit in  $R_0$  so  $v(u) = 1$ . Thus  $v(u\tau^n) = v(\tau)^n$ . It follows that  $v$  is the  $\mathfrak{B}$ -adic valuation on  $L$ .

It follows from this result that a nontrivial nonarchimedean valuation on an algebraic number field is equivalent to the  $\mathfrak{B}$ -adic valuation for some prime ideal  $\mathfrak{B}$  in the ring of algebraic integers in that algebraic number field.

One corollary of this result is that if  $x$  and  $y$  are in an algebraic number field  $K$ , there fails to exist a nonarchimedean valuation giving  $x$  and  $y$  different values if and only if  $xy^{-1}$  is a unit in the ring of algebraic integers in  $K$ . Under what circumstances does there fail to exist an archimedean valuation giving  $x$  and  $y$  different values? The answer is, when  $xy^{-1}$  is a root of unity, as we shall now prove.

**Proposition 1.22** Given two elements  $x$  and  $y$  in a field  $K$ , there exists a valuation separating  $x$  and  $y$  if and only if  $xy^{-1}$  is a root of unity.

*Proof.* We fix an algebraic number field  $K$  and set  $W = \mathbb{R}^{(s+t)}$ , where  $s$ , and  $2t$ , are the number of real, and imaginary, embeddings respectively of  $K$  into  $\mathbb{C}$ . We let  $U_K$  be the group of units in the ring of the algebraic integers of  $K$  and define a group homomorphism  $\psi : U_K \rightarrow W$  by  $\psi(u) = \sum_{h \leq s} d_h \log |u^{\sigma_h}| + \sum_{s < h \leq s+t} d_h 2 \log |u^{\sigma_h}|$  where  $\{d_h\}$  is the canonical basis of  $W$ . Here the  $\sigma_h$  are ordered so that the first  $s$  are real, and the  $\sigma_h$ ,  $s < h \leq s+t$  are a set of representatives of the imaginary embeddings under the action of complex conjugation. We prove that for any positive real number  $\beta$ , the set  $\mathcal{S} = \{a \in \mathfrak{o}_K \mid |a^{\sigma_h}| \leq \beta \forall h\}$  is finite, where  $\mathfrak{o}_K$  denotes the ring of algebraic integers in  $K$ . Multiplication by  $a$  acts as a vector space automorphism of  $K$  considered as a vector space over  $K$  and this automorphism has a characteristic polynomial which we denote by  $c_{K/\mathbb{Q},a}$ . Now, if we let  $u_1, \dots, u_n$  denote a  $\mathbb{Q}$ -basis of  $K$ , then  $u_j a = \sum_{k=1}^n c_{jk} u_k$  ( $j=1, \dots, n$ ), for some matrix with rational entries  $(c_{jk})$ . Considering the image of this equation under various embeddings  $\sigma_i$  of  $K$  in  $\mathbb{C}$  we get the matrix equation  $(u_j^{\sigma_i})(\text{diag}_i(a^{\sigma_i})) = (c_{jk})(u_k^{\sigma_i})$ . By Proposition 1.8,  $\det(u_j^{\sigma_i}) \neq 0$ . Therefore  $c_{K/\mathbb{Q},a}$  equals the characteristic polynomial of  $\text{diag}(a^{\sigma_i})$ . Thus the coefficients of  $c_{K/\mathbb{Q},a}$  are all symmetric functions in the  $a^{\sigma_h}$ , so that if  $a \in \mathcal{S}$ , then  $c_{K/\mathbb{Q},a}$  has bounded coefficients in  $\mathbb{Z}$  and there are only a finite number of possibilities for the characteristic polynomials. We now prove that  $\ker \psi = \mu_K$ ,  $\mu_K$  being defined to be the group of roots of unity in  $K$ , which will be enough to prove our promised proposition that every archimedean valuation fails to separate  $x$  and  $y$  if and only if  $xy^{-1}$  is a root of unity. It is clear that  $\mu_K \subset \ker \psi$ , and if  $u \in \ker \psi$ , then the  $u^{\sigma_i}$  all have absolute value 1, and so by the above result  $\ker \psi$  is a finite subgroup of  $K^*$ , and must therefore consist entirely of roots of unity. This completes the proof.  $\square$

**Proposition 1.23** Let  $R$  be a Dedekind domain with field of fractions  $K$ ,  $L$  a finite separable extension of  $K$  and  $R'$  the integral closure of  $R$  in  $L$ . Let  $\mathfrak{U}$  be an ideal of  $R'$  such that  $\mathfrak{U} \cap R = \mathfrak{p}$  is a nonzero prime ideal. Then  $(R'/\mathfrak{U} : R/\mathfrak{p}) \leq (L : K)$ .

*Proof.* Let  $S$  be the complement of  $\mathfrak{p}$  in  $R$  so that  $R_S$  is a DVR. Then  $\mathfrak{U}R'_S \cap R_S = \mathfrak{p}R_S$  and  $R'_S/\mathfrak{U}R'_S \simeq R'/\mathfrak{U}$ . Hence it suffices to prove the lemma with  $R_S, R'_S$ , etc., in place of  $R, R'$ , etc. Thus we may suppose that  $\mathfrak{p} = R\pi$  is principal.

Let  $\{x_i\}$  be a finite set of elements of  $R'$  whose cosets  $x_i + \mathfrak{U}$  are linearly independent over  $R/\mathfrak{p}$ . Suppose there is a relation  $\sum a_i x_i = 0$  with certain elements  $a_i$  in  $K$ . Multiplying the  $a_i$  by a suitable common denominator we obtain such a relation with the  $a_i \in R$ . Suppose not all the  $a_i$  are zero. Then there is a highest power of  $\pi$  which divides all the  $a_i$ . Cancelling this highest power we obtain a relation in which not all the  $a_i$  are in  $R\pi$ . Then we get a relation of dependence  $\sum \overline{a_i x_i} = \overline{0}$  in  $R'/\mathfrak{U}$  contrary to the assumed linear independence of the cosets  $\overline{x_i} = x_i + \mathfrak{U}$ . Consequently the  $x_i$  are linearly independent over  $K$  and the lemma follows.  $\square$

As a corollary to the above proposition we have that when  $R$  is the ring of algebraic integers in an algebraic number field and  $\mathfrak{p}$  is a prime ideal in  $R$ , then  $R/\mathfrak{p}$  is a finite field. Finally we shall need

**Proposition 1.24** The completion of an algebraic number field with respect to a nonarchimedean valuation is locally compact.

*Proof.* Let  $v$  be a nonarchimedean valuation on an algebraic number field  $K$ ; by abuse of notation we shall also use  $v$  to denote the extension of  $v$  to the completion  $K_v$ . As a matter of fact the valuation ring  $R = \{x \in K_v | v(x) \leq 1\}$  is compact. It is not difficult to see that any  $x \in R$  may be written as an infinite sum  $a_0 + a_1\pi + a_2\pi^2 + \dots$ , where  $v(\pi) = 1$  and the  $a_i$  all come from a set of representatives for the field  $R/\mathfrak{p}$ , where  $\mathfrak{p}$  is the unique maximal ideal of  $R$ . We easily see from the foregoing proposition and its corollary that  $R/\mathfrak{p}$  is a finite field. Thus  $R$  is the inverse limit of a directed system of finite topological rings  $R/\mathfrak{p}, R/\mathfrak{p}^2, \dots$ . So topologically it is homeomorphic to a closed subset of the topological product of all these topological rings, which is compact by Tychonoff's theorem. Therefore  $R$  is compact.  $\square$ .

As a corollary to the above result we see that the ideals  $\mathfrak{p}, \mathfrak{p}^2, \dots$  are also compact.

---

## CHAPTER 2

### SL(2, $\mathbb{Q}$ )

---

An important theorem in group theory, Tits' theorem, states that given a field  $K$  of characteristic zero, every subgroup of  $GL(n, K)$  either contains a free subgroup of rank two or is almost solvable (that is, it contains a solvable subgroup of finite index). In this chapter we prove a special case of this theorem, namely that every subgroup of  $SL(2, \mathbb{R})$  is either solvable or contains a free subgroup of rank two. We also show how to get further information, namely that given a  $g \in SL(2, \mathbb{Q})$  of infinite order, there exists a Zariski dense set of  $h \in SL(2, \mathbb{Q})$  such that  $g$  and a sufficiently high power of  $h$  generate a free group of rank two. The term "Zariski dense" will be defined later.

Recall that the projective line  $P^1(\mathbb{R})$  is defined to be the set of equivalence classes of nonzero elements of  $\mathbb{R}^2$  by the equivalent relation whereby two elements are equivalent if and only if one is a nonzero real multiple of the other.  $GL(2, \mathbb{R})$  acts on  $P^1(\mathbb{R})$  in a natural way, namely the matrix  $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$  sends the class containing the vector  $\begin{pmatrix} x \\ y \end{pmatrix}$  to the class containing the vector  $\begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}$ . It is easy to see that this is a well-defined map.

It is possible to view the projective real line  $P^1(\mathbb{R})$  as the ordinary real line  $\mathbb{R}$  with  $\infty$  adjoined. Namely, the class containing  $\begin{pmatrix} x \\ 1 \end{pmatrix}$  can be identified with the real number  $x$ , and

the class containing  $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$  can be identified with  $\infty$ . Under these identifications we see that a matrix  $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$  will map  $x$  to  $\frac{ax+b}{cx+d}$ , with the convention that when  $ax+b \neq 0$  and  $cx+d = 0$  this is  $\infty$ , and will map  $\infty$  to  $\frac{a}{c}$  when  $c \neq 0$ , and to  $\infty$  when  $c = 0$ . The group of such transformations of the projective real line is isomorphic to  $PSL(2, \mathbb{R})$ . We prove the theorem for this group, which evidently suffices.

By the Jordan Normal Form theorem a matrix in  $SL(2, \mathbb{R})$  will be conjugate in  $GL(2, \mathbb{C})$  to a matrix of the form  $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$  or  $\begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$  or  $\begin{pmatrix} \lambda & 0 \\ 0 & \frac{1}{\lambda} \end{pmatrix}$ . In the first two cases and in the latter cases when  $\lambda$  is real, it will even be conjugate in  $GL(2, \mathbb{R})$  to a matrix of this form. Furthermore in the latter case when  $\lambda$  is not real, the eigenvalues will occur in conjugate pairs since the characteristic polynomial of the matrix has real coefficients, so that we will have  $\lambda\bar{\lambda} = 1$ .

If the matrix in question is conjugate in  $GL(2, \mathbb{R})$  to  $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$  or  $\begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$ , then the map is conjugate to a map of the form  $x \mapsto x + 1$  or  $x \mapsto x - 1$ , which have exactly one fixed point, namely  $\infty$ , and so the original map itself has exactly one fixed point. In this case the element of  $SL(2, \mathbb{R})$  is called **parabolic**. In the case where it is conjugate in  $GL(2, \mathbb{R})$  to  $\begin{pmatrix} \lambda & 0 \\ 0 & \frac{1}{\lambda} \end{pmatrix}$  with  $\lambda$  real and  $\neq 1$ , there are two distinct fixed points and in this case the element is said to be **hyperbolic**. Note that in this case the map is conjugate to a map of the form  $x \mapsto \lambda^2 x$ . In the case where the matrix is conjugate in  $GL(2, \mathbb{C})$  to  $\begin{pmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{pmatrix}$  with  $\lambda$  not real, then there are no fixed points and in this case the element is said to be **elliptic**.

Let us now prove

**Theorem 2.1.** Every subgroup of  $SL(2, \mathbb{R})$  is either solvable or contains a free subgroup of rank two.

*Proof.* We follow [1], section 3. Let  $\Gamma$  be a subgroup of  $SL(2, \mathbb{R})$ . Suppose that  $\Gamma$  contains at least one parabolic element  $g$  and let it have fixed point  $P$ . If all other elements in the group have fixed point  $P$  then we can assume by conjugating the group if necessary  $P = \infty$ , and then all the elements of the group have maps of the form  $x \mapsto ax + b$ , and so generate a solvable group. If not then we can select another element  $h$  such that  $h(P) = Q \neq P$ , and then  $k = hgh^{-1}$  has fixed point  $Q$ . Select disjoint compact neighbourhoods of  $P$  and  $Q$  in  $P^1(\mathbb{R})$ . Since a parabolic element  $g$  of  $SL(2, \mathbb{R})$  is conjugate in  $GL(2, \mathbb{R})$  to a map of the form  $x \mapsto x + 1$  or  $x \mapsto x - 1$ , then it is easy to see that the fixed point is an attracting fixed point, that for any  $x$  the limit of  $g^n(x)$  as  $n$  approaches  $\infty$  or  $-\infty$  is the fixed point of  $g$ . Given two compact neighbourhoods of  $P$  and  $Q$ , some iterate of  $k$  and likewise some iterate of  $k^{-1}$  will map both the neighbourhood of  $P$  and the neighbourhood of  $Q$  entirely into the neighbourhood of  $Q$ , and some iterate of  $g$  and likewise some iterate of  $g^{-1}$  will map both the neighbourhood of  $P$  and the neighbourhood of  $Q$  entirely into the neighbourhood of  $P$ . Let  $n$  be such that  $k^n, k^{-n}$ , map the neighbourhoods of  $P$  and  $Q$  entirely into the neighbourhood of  $Q$ , and  $g^n, g^{-n}$  map the neighbourhoods of  $P$  and  $Q$  entirely into the neighbourhood of  $P$ . Then no nontrivial word in  $g^n, k^n$  can be the identity, for it either maps the neighbourhood of  $P$  into the neighbourhood of  $Q$  or vice versa. Consequently  $g^n, k^n$  generate a free group of rank two.

Now suppose  $\Gamma$  contains at least one hyperbolic element  $h$ , and no parabolic elements. If there are two hyperbolic elements without a common fixed point, then we may select a compact neighbourhood of the fixed points of one of the hyperbolic elements, and a disjoint compact neighbourhood of the fixed points of the other, and proceed as in the previous case, because in light of the fact that a hyperbolic element is conjugate to a map of the

form  $x \mapsto ax$  with  $a$  real and positive, it may be seen that one of the fixed points of the hyperbolic element is attracting for positive iterates of the hyperbolic element, and the other for negative iterates. Consequently, if we suppose that  $\Gamma$  has no free subgroup of rank two then any two hyperbolic elements of  $\Gamma$  have a common fixed point. We might think it possible for two hyperbolic elements to have fixed point sets  $\{P, Q\}$  and  $\{P, R\}$  respectively, with  $Q \neq R$ , and then for a third hyperbolic element to have fixed point set  $\{Q, R\}$ . But in this case conjugating the hyperbolic element with fixed point set  $\{P, Q\}$  by the hyperbolic element with fixed point set  $\{Q, R\}$  would yield a hyperbolic element with fixed point set  $\{Q, S\}$ , where  $S$  is different from  $P, Q$ , and  $R$ , and this together with the hyperbolic element with fixed point set  $\{P, R\}$  would yield two hyperbolic elements without a common fixed point, contrary to our assumption that  $\Gamma$  has no free subgroup of rank two. So all hyperbolic elements of  $\Gamma$  have a common fixed point.

Now, if all elements of  $\Gamma$  have a common fixed point, then, by conjugating  $\Gamma$  if necessary, we can make this fixed point  $\infty$  and then all the elements will be maps of the form  $x \mapsto ax + b$  and the group will be solvable as before. If not, then there is an element  $g$  which has no fixed points in common with  $h$ . We may distinguish two cases, the case where  $g$  swaps the fixed points of  $h$  and the case where it doesn't. If  $g$  swaps the fixed points of  $h$  and there's another hyperbolic element with only one fixed point in common with  $h$ , then  $g$  doesn't swap the fixed points of that one. So our two cases become Case 1: all hyperbolic elements in  $\Gamma$  have the same two fixed points and every elliptic element  $g$  swaps them, and Case 2: there's at least one hyperbolic element  $h$  and at least one elliptic element  $g$  such that  $g$  doesn't swap the fixed points of  $h$ .

*Case 1.* All hyperbolic elements in  $\Gamma$  have the same two fixed points and every elliptic element  $g$  swaps the fixed points of every hyperbolic element  $h$ .

Pick a hyperbolic element  $h$ . Since  $h$  is conjugate to a map of the form  $x \mapsto ax$ , with  $a > 0$ , then, by conjugating  $\Gamma$  if necessary, we may suppose that  $h$  is of this form and that the fixed points of  $h$  are 0 and  $\infty$ . Then every hyperbolic element, since it has the fixed points in question, will be of this form. If there are no elliptic elements, then the hyperbolic elements will generate an abelian group. Suppose there is at least one elliptic element  $g$ . In that case we can show that  $g$  must be of the form  $x \mapsto \frac{1}{\lambda^2}x$ . If there are two different elliptic elements  $g_1, g_2$ , then one is the composite of the other with some hyperbolic element. So the hyperbolic elements together with our single chosen elliptic element  $g$  generate the entire group. Furthermore  $ghg^{-1} = h^{-1}$  for all hyperbolic elements  $h$ . Thus the group is solvable.

*Case 2.* There is at least one hyperbolic element  $h$  and at least one elliptic element  $g$  such that  $g$  does not swap the fixed points of  $h$ .

Let the fixed points of  $h$  be  $P, Q$ , and suppose that  $g(P) = R$  is neither  $P$  nor  $Q$ . If  $g(P) \neq Q$ , then  $h$  and  $ghg^{-1}$  are two hyperbolic elements without common fixed points and we get a free subgroup of  $\Gamma$  of rank two. So we may assume  $g(P) = Q$ . If  $g(R) \neq P$  then  $h$  and  $g^2hg^{-2}$  are two hyperbolic elements without common fixed points and again we get a free subgroup of rank two. So suppose  $g(R) = P$ . Consider  $h' = g^{-1}hg$ , a hyperbolic element with fixed points  $R$  and  $P$ , as well as  $h'' = ghg^{-1}hgh^{-1}g^{-1}$ , a hyperbolic element with fixed points  $Q = ghg^{-1}(Q)$  and  $S = ghg^{-1}(P)$ . We have  $h(R) \neq Q$  and thus  $S = gh(R) \neq g(Q) = R$ ; we also have  $h(R) \neq R$  and  $S \neq g(R) = P$ . Consequently  $h'$  and  $h''$  are two hyperbolic elements without a common fixed point and again we get a free subgroup of rank two.

Finally there is the case where the group consists entirely of elliptic elements. Before dealing with this case we must pause to investigate the manner in which  $GL(2, \mathbb{C})$  acts on  $P^1(\mathbb{C})$ .

Clearly,  $GL(2, \mathbb{C})$  may be viewed as acting on  $P^1(\mathbb{C})$ . We may identify the class in  $P^1(\mathbb{C})$  containing the vector  $\begin{pmatrix} z \\ 1 \end{pmatrix}$  with the complex number  $z$ , and the class in  $P^1(\mathbb{C})$  containing the vector  $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$  with  $\infty$ , and in this way  $P^1(\mathbb{C})$  is identified with the Riemann sphere  $\mathbb{C} \cup \{\infty\}$ . Under this identification, an element  $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$  will map a complex number  $z$  to  $\frac{az+b}{cz+d}$ , it being understood that when  $az+b \neq 0$  and  $cz+d = 0$  this is  $\infty$ , and will map  $\infty$  to  $\frac{a}{c}$ . We call a transformation of the Riemann sphere of this form a **linear fractional transformation**.

We view an extended line of the Riemann sphere as an ordinary straight line in  $\mathbb{C}$  together with  $\infty$ . We may reflect the Riemann sphere in an extended line by fixing each point on the extended line, and reflecting all other points in the straight line from which the extended line came in the ordinary way. We may reflect the Riemann sphere in a circle by mapping  $\infty$  to the centre of the circle and vice versa, and mapping each other point  $p$  to a point on the ray from the centre of the circle to  $p$ , where the distance of the new point from the centre of the circle is  $\frac{r^2}{x}$ , where  $r$  is the radius of the circle and  $x$  is the distance of  $p$  from the centre of the circle. We define a Moebius transformation of the Riemann sphere to be one which can be obtained as a composition of a finite series of reflections in extended lines and circles. Our next job is to show that every linear fractional transformation is a Moebius transformation.

If  $c = 0$ , then the map  $z \mapsto \frac{az+b}{cz+d}$  is the composition of a translation, rotation, and possibly a dilation. A translation and rotation can be obtained as the composition of a series of reflections in extended lines and a dilation can be obtained as the composition of two reflections in circles. Thus a linear fractional transformation of this form is a Moebius transformation. If  $c \neq 0$ , then

$$\frac{az+b}{cz+d} = \frac{a}{c} + \frac{bc-ad}{c^2z+cd}.$$

This can again be obtained by composing translations, rotations, dilations, and reflections. Thus all linear fractional transformations are Moebius transformations.

We must now show

**Proposition 2.2** A Moebius transformation maps extended lines and circles to extended lines or circles.

*Proof.* We follow [6], chapter 4, section 3. We switch from working in  $\mathbb{C} \cup \{\infty\}$  to working in  $\mathbb{R}^2 \cup \{\infty\}$ , identifying the complex number  $x+iy$  with the vector  $\begin{pmatrix} x \\ y \end{pmatrix}$ . Now the circle with centre  $a$  and radius  $r$  is precisely the set of points  $x$  in  $\mathbb{R}^2$  satisfying  $|x|^2 - 2a \cdot x + |a|^2 - r^2 = 0$ . Further, the set of finite points on an extended line is of the form  $-2a \cdot x + 2t = 0$ , for some unit vector  $a$  and  $t$ , in which case let us call the line  $L(a, t)$ . These can be written in the common form  $a_0 x^2 - 2a \cdot x + a_{n+1} = 0$  with  $|a|^2 > a_0 a_{n+1}$ . Conversely, given any vector  $(a_0, \dots, a_{n+1})$  in  $\mathbb{R}^{n+2}$  such that  $|a|^2 > a_0 a_{n+1}$ , if we set  $a = (a_1, \dots, a_n)$ , we get a circle or line  $\Sigma$  consisting of the set of points satisfying the equation  $a_0 |x|^2 - 2a \cdot x + a_{n+1} = 0$ . Namely, if  $a_0 \neq 0$ , then  $\Sigma$  is the circle with centre  $\frac{a}{a_0}$  and radius  $\frac{(|a|^2 - a_0 a_{n+1})^{\frac{1}{2}}}{|a_0|}$ , whereas if  $a_0 = 0$ , then  $\Sigma$  is the line  $L(\frac{a}{|a|}, \frac{a_{n+1}}{2|a|})$ . The vector  $(a_0, \dots, a_{n+1})$  is called the **coefficient vector** for  $\Sigma$ , and is uniquely determined by  $\Sigma$  up to multiplication by a nonzero scalar.

Now we wish to prove that a Moebius transformation maps an extended line or circle to an extended line or circle. It suffices to prove that a reflection in an extended line or circle does so and it is easy to see that a reflection in an extended line does so. A reflection in a circle is conjugate by a similarity transformation to the map  $x \mapsto \frac{x}{|x|^2}$ , and similarity transformations obviously map extended lines and circles to extended lines or circles, so it suffices to prove the result for the map  $x \mapsto \frac{x}{|x|^2}$ . Let  $(a_0, \dots, a_{n+1})$  be a coefficient vector for

the extended line or circle  $\Sigma$ . Then the finite points on  $\Sigma$  are precisely those satisfying the equation  $a_0|x|^2 - 2a.x + a_{n+1} = 0$ . Suppose  $x \neq 0$  and let  $y = \frac{x}{|x|^2}$ . Then  $y$  satisfies the equation  $a_0 - 2a.y + a_{n+1}|y|^2 = 0$ . If  $x = 0$  and satisfies the equation  $a_0|x|^2 - 2a.x + a_{n+1} = 0$ , then we have  $a_{n+1} = 0$ , and  $y = \infty$  and lies on the extended line whose finite part has the equation  $a_0 - 2a.y + a_{n+1}|y|^2 = 0$ . In either case, the extended line or circle  $\Sigma$  is mapped into the the extended line or circle  $\Sigma'$  whose finite points satisfy the equation  $a_0 - 2a.y + a_{n+1}|y|^2 = 0$ , and a similar argument shows that  $\Sigma'$  is mapped into  $\Sigma$ , so that  $\Sigma$  is mapped onto  $\Sigma'$ , as desired. This completes the proof.  $\square$ .

Let us now return to the discussion, in the proof of Theorem 1.1, of the case in which the group consists entirely of elliptic elements. We know by the theory of the Jordan normal form that the matrix of an elliptic element is conjugate in  $GL(2, \mathbb{C})$  to a matrix of the form  $\begin{pmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{pmatrix}$ , where  $|\lambda| = 1$ . Thus the elliptic element is conjugate by a linear fractional transformation to a map of the form  $z \mapsto \lambda^2 z$ , where  $|\lambda| = 1$  and furthermore  $\lambda \neq 1$ . This linear fractional transformation must map the extended real line to a circle fixed setwise by such a map, *i.e.*, a circle with centre the origin. And in fact dilations by a real number commute with the map  $z \mapsto \lambda^2 z$ , so we may assume that it is the unit circle. Thus, the elliptic element is conjugate, by a linear fractional transformation which maps the extended real line to the unit circle, to the map  $z \mapsto \alpha z$  where  $|\alpha| = 1$ . A different elliptic element is conjugate by the same linear fractional transformation to an automorphism (itself given by a linear fractional transformation) of the unit circle, and it may be seen that the new elliptic element commutes with the original elliptic element if and only if the automorphism in question is of the form  $z \mapsto \beta z$  with  $|\beta| = 1$ .

To return to the consideration of the case where the group has only elliptic elements. Pick one fixed elliptic element  $g$ , and let it be conjugate via some fixed linear fractional transformation to the rotation of the unit circle by an angle  $\alpha$ , and from now on view all

elliptic elements as conjugated via this fixed linear fractional transformation and acting on the unit circle. If all the other elements commute with it, the group is abelian. Otherwise, let  $h$  be an elliptic element such that  $k = hgh^{-1} \neq g$ . Now  $k$ , conjugated by the previously fixed linear fractional transformation, may be viewed as an automorphism of the unit circle. Let  $\tilde{k} : \mathbb{R} \mapsto \mathbb{R}$  be the lifting of this automorphism to  $\mathbb{R}$ , considered as the universal covering of the unit circle, with  $0 \leq \tilde{k}(0) < 2\pi$ . Then we easily see that  $\lim_{n \rightarrow \infty} \frac{1}{n}(\tilde{k}^n(x) - x)$  exists for all  $x \in \mathbb{R}$  and that this limit is  $\alpha$ , and furthermore  $\min_{x \in \mathbb{R}}(\tilde{k}(x) - x) \leq \alpha \leq \max_{x \in \mathbb{R}}(\tilde{k}(x) - x)$ . It follows that there exists  $P \in \mathbf{S}^1$  with  $k(P) = g(P)$ , so that  $g^{-1}k$  has a fixed point, contradicting the hypothesis that the group consists entirely of elliptic elements.

The proof of Theorem 1.1 is now complete.  $\square$

Now we prove our next result. To state it, we must define the Zariski topology on  $SL(2, \mathbb{Q})$ . In general, if  $K$  is a field then a subset of the affine space  $K^n$  is said to be a variety if it is the set of common zeroes of a finite family of polynomials in  $K[x_1, \dots, x_n]$ . By Proposition 1.10, every ideal of  $K[x_1, \dots, x_n]$  is finitely generated, so the qualifier "finite" is redundant here. A variety which is a subset of another variety is said to be a subvariety of that variety. Thus  $SL(2, \mathbb{Q})$  is a subvariety of  $\mathbb{Q}^4$ . The subvarieties of a given variety are clearly closed under intersection, and they are also closed under finite union, for if  $S_1$  is a set of polynomials defining the variety  $V_1$  and  $S_2$  is a set of polynomials defining the variety  $V_2$ , then  $\{fg | f \in S_1, g \in S_2\}$  defines the variety  $V_1 \cup V_2$ . Thus there is a topology on  $SL(2, \mathbb{Q})$  whose closed sets are precisely the subvarieties, and this topology we call the Zariski topology. Our next result states:

**Theorem 2.3.** For each  $g \in SL(2, \mathbb{Q})$  of infinite order, there exists a Zariski dense set of  $h \in SL(2, \mathbb{Q})$  such that  $g$  and a sufficiently high power of  $h$  generate a free group of rank two.

*Proof.* We begin with a result which states roughly that the eigenvectors of  $h$  depend

continuously upon  $h$ . Suppose we consider the eigenvectors as elements of  $\mathbb{C}^3$ , and perturb the coefficients of the characteristic polynomial by a small amount in the ordinary topology. Let  $H$  be the matrix of  $h$  and let  $\lambda$  be one of its eigenvalues. By perturbing the entries of  $H$  by a small amount we will perturb  $\lambda$  and therefore  $H - \lambda I$  by a small amount. Now the perturbed  $H - \lambda I$  is, like the original  $H - \lambda I$ , a projection of  $\mathbb{C}^3$  onto a 2-dimensional subspace. If we consider a vector  $v \in \mathbb{C}^3$ , we can decompose  $v = v_1 + v_2$  where  $v_2$  is in the kernel of the original  $H - \lambda I$  and  $v_1$  is in the corresponding 2-dimensional subspace. Now the image of  $v$  under the perturbed  $H - \lambda I$  is close to the image of  $v_1$  under the original  $H - \lambda I$ , and also therefore to the image of  $v_1$  under the perturbed  $H - \lambda I$ , which means that something close to  $v_2$  must be in the kernel of the perturbed  $H - \lambda I$ . Therefore the eigenvector of the perturbed  $h$  corresponding to the perturbed  $\lambda$  is close to the eigenvector of the original  $h$  corresponding to the original  $\lambda$ . A similar argument works if the perturbation takes place in the topology induced by a nonarchimedean valuation.

We now return to the main proof. We begin with the case where  $g$  is hyperbolic. Now, it follows from our proof of Theorem 1.1 that the set of  $h$  such that  $g$  and a sufficiently high power of  $h$  generate a free group of rank two include all the hyperbolic elements  $h$  which have no fixed point in common with  $g$ , such that no power of  $g$  maps one fixed point of  $h$  to another. For, in the proof of Theorem 1.1 we showed that if  $g$  is hyperbolic and  $h$  is hyperbolic having no fixed points in common with  $g$ , then there exist distinct attracting fixed points  $a, b, c, d$  for  $g, g^{-1}, h, h^{-1}$  respectively. Now, select compact neighbourhoods of  $c$  and  $d$  whose images under arbitrarily high powers of  $g$  are contained in a compact neighbourhood of  $a$ , and whose images under arbitrarily high powers of  $g^{-1}$  are contained in a compact neighbourhood of  $b$ , each compact neighbourhood disjoint from the original pair of compact neighbourhoods. This is possible because, for example, we may select a compact neighbourhood of  $a$ , disjoint from the compact neighbourhood of  $c$ , such that sufficiently high powers of  $g$  will take the compact neighbourhood of  $c$  into this neighbourhood. We

must then select compact neighbourhoods of images of  $c$  under finitely many powers of  $g$ , and we can arrange it so that these avoid the compact neighbourhood of  $c$  by making the latter sufficiently small. In this way we obtain a compact neighbourhood of  $a$  disjoint from the compact neighbourhood of  $c$  such that arbitrarily high powers of  $g$  take the compact neighbourhood of  $c$  into the compact neighbourhood of  $a$ , and a similar argument will work in the other cases. Further we can arrange it that the compact neighbourhood of  $a$  avoids the compact neighbourhood of  $d$ , and the compact neighbourhood of  $b$  avoids the compact neighbourhood of  $c$ , in virtue of the stipulation that no power of  $g$  map one fixed point of  $h$  to the other. Then a sufficiently high power of  $h$  will take both the compact neighbourhoods of  $a$  and  $b$  inside the neighbourhood of  $c$ . and a sufficiently high power of  $h^{-1}$  will take both compact neighbourhoods inside the neighbourhood of  $d$ . It is easily seen that this is a sufficient condition for  $g$  and a sufficiently high power of  $h$  to generate a free group of rank two. So, in the case where  $g$  is hyperbolic the set of all  $h$  such that  $g$  and a sufficiently high power of  $h$  generate a free group of rank two includes all the hyperbolic elements with no fixed points in common with  $g$ , which set is clearly nonempty. We shall prove that this set is the complement of a Zariski closed set in a set which is open in the ordinary topology, and therefore is a nonempty set which is open in the ordinary topology, and we shall also prove that any nonempty set which is open in the ordinary topology, or even contains an open ball, is Zariski dense, and this will complete the proof of the hyperbolic case.

Now it is easily seen that the hyperbolic elements are precisely those whose trace has absolute value greater than 2. Hence the set of hyperbolic elements is open in the usual topology. Now, in light of the fact that the eigenvectors of  $h$  depend continuously upon  $h$ , the set of  $h$  with the requisite properties will be open in the usual topology, and it is not difficult to show that it is nonempty.

We have now proved that the set of hyperbolic elements  $h$  with the requisite properties is a nonempty open set, in particular it is a set which contains an open ball. We must now prove that every set containing an open ball is Zariski dense. To show this we must show that every Zariski closed set containing an open ball of  $SL(2, \mathbb{Q})$  must contain the whole of  $SL(2, \mathbb{Q})$ . It will suffice to prove this for those Zariski closed sets defined by a single polynomial. Suppose that a Zariski closed subset of  $SL(2, \mathbb{Q})$  defined by a single polynomial  $p$  contains an open ball of  $SL(2, \mathbb{Q})$ . Then by continuity the Zariski closed subset of  $SL(2, \mathbb{R})$  defined by the same polynomial contains an open ball of  $SL(2, \mathbb{R})$ . Now  $SL(2, \mathbb{R})$  is a real-analytic 3-manifold. This means there exists an atlas of open balls in  $\mathbb{R}^3$  together with analytic embeddings into  $SL(2, \mathbb{R})$ , such that the images of the embeddings cover all of  $SL(2, \mathbb{R})$ , and such that when two of the images of the embeddings have nonempty intersection, the composition with the analytic map onto one of the images with the preimage of that onto the other is analytic. A sufficiently small open ball on which our polynomial  $p$  vanishes will be contained in the image of one of the analytic embeddings in this atlas. Then the composition of  $p$  with our analytic embedding will yield an analytic map from an open ball of  $\mathbb{R}^3$  into  $SL(2, \mathbb{R})$  which vanishes on an open ball and therefore everywhere. Furthermore, given an analytic embedding  $g$  in the atlas whose image has nonempty intersection with the image of the analytic embedding  $f$  under consideration,  $g^{-1} \circ f$  will be analytic and  $p \circ g$  will be an analytic map from an open ball of  $\mathbb{R}^3$  to  $SL(2, \mathbb{R})$  which vanishes on an open ball and therefore everywhere. Thus every open ball in the atlas which is connected to our original open ball by the transitive closure of the relation of the images of the analytic embeddings having nonempty intersection will be such that  $p$  vanishes on the image of the analytic embedding corresponding to the open ball in question. Since  $SL(2, \mathbb{R})$  is connected and is a countable union of compact sets, this means that  $p$  will vanish everywhere and so that the Zariski closed set defined by  $p$  is actually all of  $SL(2, \mathbb{R})$ . Thus the Zariski closed set in  $SL(2, \mathbb{Q})$  originally considered must contain all of  $SL(2, \mathbb{Q})$ . This proves our result.

This completes the proof that to every hyperbolic element  $g$  of  $SL(2, \mathbb{Q})$  there is a Zariski dense set of  $h$  in  $SL(2, \mathbb{Q})$  such that  $g$  and a sufficiently high power of  $h$  generate a free group of rank two. We now consider the elliptic case.

Let  $g$  be an elliptic element in  $SL(2, \mathbb{Q})$  and let  $K$  be the splitting field for the characteristic polynomial. By Proposition 1.22, if  $g$  is of infinite order then there will exist a nonarchimedean valuation  $v$  on  $K$  separating the eigenvalues of  $g$ . Then there will be two fixed points of  $g$  in  $P^1(K_v)$ , one attracted to by iteration of  $g$ , the other attracted to by iteration of  $g^{-1}$ . If we consider a second element  $h$  with  $v(\text{trace } h) > 1$ , then if we consider the smallest field  $K'$  over which the characteristic polynomials of  $g$  and  $h$  split, any valuation on  $K'$  extending  $v$  will separate the roots of  $h$ . By abuse of notation let us let  $v$  denote such a valuation and consider the completion  $K'_v$ . Then provided the 1-dimensional subspaces of  $\mathbb{Q}^2$  spanned by the eigenvectors of  $h$  are different to those of  $g$ ,  $h$  will have two fixed points in  $P^1(K'_v)$ , one attracted to by iteration of  $h$ , and the other attracted to by iteration of  $h^{-1}$ . Let us stipulate that these fixed points be different from those of  $g$  and that no power of  $g$  map one to the other. Since  $K'_v$  is locally compact and indeed in it points have arbitrarily small neighbourhoods with compact closure, this is enough for a version of the earlier argument to go through to show that  $g$  and a sufficiently high power of  $h$  generate a free group of rank two. The set of  $h$  with the requisite properties is clearly nonempty and, given a point it contains, it contains the intersection of an open neighbourhood of that point in the ordinary topology and an open neighbourhood of that point in the topology induced by some  $p$ -adic valuation. Thus it contains a set whose closure is an open neighbourhood of the point in the ordinary topology. By our earlier discussion, it is Zariski dense. This completes the proof of the elliptic case.

Finally there remains the parabolic case. A parabolic element  $g$  has one attracting fixed point in  $P^1(\mathbb{Q})$  and by similar arguments to those foregoing any hyperbolic element  $h$  with

no fixed points in common will be such that  $g$  and a sufficiently high power of  $h$  generate a free group of rank two. The set of such hyperbolic elements  $h$  is evidently Zariski dense so this completes the proof of the parabolic case. We have now proved the proposition.  $\square$

In the next chapter we discuss the extent to which this proposition generalizes to  $SL(3, \mathbb{Q})$ .

---

## CHAPTER 3

### $SL(3, \mathbb{Q})$

---

In the last chapter we saw that, given an element  $g$  of  $SL(2, \mathbb{Q})$  of infinite order, there exists a Zariski dense set of elements  $h$  in  $SL(2, \mathbb{Q})$  such that  $g$  and a sufficiently high power of  $h$  generate a free group of rank two. We seek to explore the extent to which this generalizes to  $SL(3, \mathbb{Q})$ .

First we prove

**Theorem 3.1** Suppose  $g \in SL(3, \mathbb{Q})$  and suppose there exists a valuation  $v$  on the splitting field  $K$  for the characteristic polynomial of  $g$ , such that, for some appropriate labelling  $\lambda_1, \lambda_2, \lambda_3$  of the eigenvalues of  $g$ , we have  $v(\lambda_1) > v(\lambda_2) > v(\lambda_3)$ . Then there exists a Zariski dense set of  $h \in SL(3, \mathbb{Q})$  such that  $g$  and a sufficiently high power of  $h$  generate a free group of rank two.

*Proof.* Suppose that  $h \in SL(3, \mathbb{Q})$  and, for some appropriate labelling  $\mu_1, \mu_2, \mu_3$  of the eigenvalues of  $h$ , for some extension  $v'$  of  $v$  to the smallest field  $K'$  over which the characteristic polynomials of  $g$  and  $h$  both split, we have  $v'(\mu_1) > v'(\mu_2) > v'(\mu_3)$ . Let  $g^+$  and  $g^-$  be eigenvectors of  $g$  corresponding to the eigenvalues  $\lambda_1$  and  $\lambda_3$  respectively, and let  $h^+$  and  $h^-$  be eigenvectors of  $h$  corresponding to the eigenvalues  $\mu_1$  and  $\mu_3$  respectively. Suppose that the components of  $h^+$  and  $h^-$  respectively in  $g^+$  and  $g^-$  are nonzero, and the components of  $g^+$  and  $g^-$  respectively in  $h^+$  and  $h^-$  are nonzero. Suppose further that the components of

$g^k(h^+)$  and  $g^k(h^-)$  respectively, for arbitrary nonzero integer values of  $k$ , in  $h^+$  and  $h^-$ , are nonzero.

Now, the images of  $g^+$  and  $g^-$  in  $P^2(K'_\nu)$  are fixed points of  $g$ . In what follows, we will identify vectors with their images in  $P^2(K'_\nu)$ . Furthermore,  $g^+$  is attracting under iteration of  $g$  for any seed value whose component in  $g^+$  is not zero, and  $g^-$  is attracting under iteration of  $g^{-1}$  for any seed value whose component in  $g^-$  is not zero. Similar remarks can be made about  $h^+$  and  $h^-$ . Now, one may select a compact neighbourhood of the image of  $g^+$  in  $P^2(K'_\nu)$  such that everything in the neighbourhood avoids the “bad set”, the set of images of vectors which have zero component in  $h^+$  or  $h^-$ . This neighbourhood will contain the images of  $g^k(h^+)$  and  $g^k(h^-)$  for sufficiently large positive  $k$ . Indeed, we may take a compact neighbourhood  $H^+$  of  $h^+$  and a compact neighbourhood  $H^-$  of  $h^-$  such that the neighbourhood contains the images of  $g^k(H^+)$  and  $g^k(H^-)$  for sufficiently large positive  $k$ . For the finitely many remaining values of  $k$ , one may likewise take compact neighbourhoods of the images of  $g^k(h^+)$  and  $g^k(h^-)$  which avoid the “bad set”. Then, by making these neighbourhoods small if necessary, one may take a compact neighbourhood of  $h^+$ , for example, disjoint from all these neighbourhoods, such that everything in it has nonzero component in  $g^+$  and  $g^-$ , and mapped into union of the compact neighbourhoods by the action of  $g$  and arbitrarily high powers thereof. Then the union of the neighbourhoods will be mapped back into the original compact neighbourhood of  $h$  by a sufficiently high power of  $h$ . In like manner one may choose compact neighbourhoods of  $g^-$  and  $h^-$ , such that either the compact neighbourhood of  $h^-$  or  $h^+$  gets mapped into the compact neighbourhood of  $g^+$  or  $g^-$ , respectively, by a single application of  $g$  or  $g^{-1}$  respectively, and arbitrarily high powers thereof, and the compact neighbourhoods of  $g^+$  or  $g^-$  get mapped into the compact neighbourhood of  $h^+$  or  $h^-$ , respectively, by a sufficiently high power of  $h$  or  $h^{-1}$  respectively. This is a sufficient condition for  $g$  and a sufficiently high power of  $h$  to generate a free group of rank two. We must now show that the set of  $h$  with the requisite properties is Zariski dense.

We prove this in two stages, firstly showing given any point it contains, it contains a set whose closure is an open neighbourhood of that point, and secondly showing it is nonempty. Since  $SL(3, \mathbb{R})$  is connected a similar argument to the previous chapter works to show that this is sufficient for it to be Zariski dense. To show the first property, we must show roughly that the eigenvalues of  $h$  depend continuously on  $h$ , in the topology induced by  $v'$ .

Evidently the coefficients of the characteristic polynomial of  $h$  depend continuously on  $h$  in the topology induced by any valuation. Any cubic may be replaced by a cubic  $x^3 + px + q$  with the same roots where  $p$  and  $q$  depend continuously on the coefficients of the original cubic in the topology induced by any valuation. Now Cardano's formula tells us we may obtain the roots of the cubic as follows: let  $\beta$  and  $\gamma$  be cube roots of  $-\frac{27}{2}q \pm \frac{3}{2}\sqrt{-3\Delta}$ , where  $\Delta = -4p^3 - 27q^2$ , the cube roots being chosen so that  $\beta\gamma = -3p$ . Then the roots are  $\frac{1}{3}(\beta + \gamma)$ ,  $\frac{1}{3}(\omega^2\beta + \omega\gamma)$ ,  $\frac{1}{3}(\omega\beta + \omega^2\gamma)$ , where  $\omega$  is a primitive cube root of unity. Hence evidently the eigenvalues of  $h$  depend continuously on  $h$ , when considered as complex numbers, with respect to the the ordinary topologies on  $\mathbb{C}$  and  $SL(3, \mathbb{Q})$ .

But now consider the nonarchimedean case. Let  $K$  be an extension by radicals of  $\mathbb{Q}$ , in which the roots exist, and let  $K_v$  be a completion of  $K$  with respect to an extension of the valuation under consideration, assumed nonarchimedean. By consideration of the power series for  $(1+x)^{\frac{1}{2}}$  and  $(1+x)^{\frac{1}{3}}$  it is not too difficult to see that the square root and cube root operators will be continuous functions with respect to the topology induced by the valuation in question. Hence if we perturb the coefficients of the original characteristic polynomial by a small amount in the appropriate topology, the roots will exist in the same complete field  $K_v$  and will be perturbed by only a small amount in the topology induced by  $v$ . These considerations show that the set of  $h$  with its eigenvalues separated by  $v$  is open in the appropriate topology.

Bearing in mind what the requisite properties of  $h$  were for  $g$  and a sufficiently high power of  $h$  to generate a free group of rank two, we now easily see that the set of  $h$  with the requisite properties is such that, given any point it contains, it contains a set whose closure is an open neighbourhood of that point. We must now show that it is nonempty. We do this by keeping the eigenvalues of  $h$  rational and selecting rational eigenvectors of  $h$ . This will lead to  $h$  being in  $SL(3, \mathbb{Q})$ , so this is acceptable. We can then easily select the eigenvalues in such a way that the component of  $h^+$  and  $h^-$  in each of  $g^+$ ,  $g^-$  is nonzero, and vice versa. Moreover we can preserve this property by perturbing the eigenvectors by a small amount. We choose a finite positive integer  $k_0$  which remains an upper bound, even when  $h^+$  and  $h^-$  are perturbed slightly, on the absolute value of the  $k$  for which we have to verify that the components of  $g^k(h^+)$  and  $g^k(h^-)$  in  $h^+$  and  $h^-$  are nonzero. We can then achieve this for each such  $k$  by perturbing the eigenvector of  $h$  other than  $h^+$  and  $h^-$  by a small amount. In this way we show the set of  $h$  with the requisite properties to be nonempty. This completes the proof of Theorem 3.1.  $\square$

We now prove

**Theorem 3.2** Suppose  $g$  is an element of  $SL(3, \mathbb{Q})$  of infinite order, conjugate to a Jordan matrix with a nontrivial Jordan block. Then there is a Zariski dense set of  $h \in SL(3, \mathbb{Q})$  such that  $g$  and a sufficiently high power of  $h$  generate a free group of rank two.

*Proof.* The case where  $g$  is conjugate to a Jordan matrix with a nontrivial Jordan block is not difficult. If  $g$  is conjugate to  $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ , then with respect to the ordinary topology  $g$  has one attracting fixed point with respect to the ordinary topology on  $\mathbb{C}$ , attracted to by either the action of  $g$  or  $g^{-1}$ , and the conditions on  $h$  for  $g$  and a sufficiently high power of  $h$  to generate a free group of rank two are similar to previously and a similar argument

goes through. If  $g$  is conjugate to  $\begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \mu \end{pmatrix}$ , with  $\lambda^2\mu = 1$ , then if  $|\lambda| \neq |\mu|$  then  $g$  has two attracting fixed points with respect to the ordinary topology on  $\mathbb{C}$ , one attracted to by  $g$  and the other by  $g^{-1}$ , and if  $|\lambda| = |\mu|$  then  $g$  has just one attracting fixed point, for both  $g$  and  $g^{-1}$ , and in either case a version of the argument will go through.  $\square$

There remains the case where the matrix of  $g$  is diagonalizable. If no two of the eigenvalues can be separated by a valuation, then  $g$  has finite order. There remains the case where one eigenvalue can be separated from the other two, but it is not possible to separate all three. Suppose we consider the case where one eigenvalue can be separated from the others by an archimedean valuation. Dealing with this case would be sufficient because if no two of the eigenvalues can be separated by an archimedean valuation, then the eigenvalues must all be complex numbers of absolute value 1, so one of them must be  $\pm 1$ , and the remaining two, since  $g$  is of infinite order, would have to be separable by a nonarchimedean valuation, and then all three eigenvalues would be separable by that nonarchimedean valuation. In this case, letting  $\lambda_1, \lambda_2, \lambda_3$  be the three eigenvalues, we have, relabelling if necessary,  $|\lambda_1| \neq |\lambda_2| = |\lambda_3|$ . Assume for convenience of discussion that  $|\lambda_1| > |\lambda_2|$ , as we can by replacing  $g$  with  $g^{-1}$ . Then  $h^+$ , say, assuming its components the vector corresponding to  $\lambda_1$  and in the vector subspace corresponding to  $\lambda_2, \lambda_3$  are both nonzero, will be attracted to a single fixed point under iteration of  $g$ , but will be attracted towards a complex projective line, a copy of the Riemann sphere, under iteration of  $g^{-1}$ . However, the attracting set will be a proper subset of this complex projective line. The projection of  $h^+$  onto the subspace in question will correspond to a certain point in  $P^1(\mathbb{C})$ , and the set of points attracting  $h^+$  will consist of this point multiplied by various values of  $e^{i\theta}$  as  $\theta$  ranges over a subset of the real numbers. The difficulty with getting the argument to go through lies in ensuring that this entire attracting circle misses the “bad set” consisting of those points whose components in  $h^+$  or  $h^-$  are zero.

One might be tempted to suppose that we could get the eigenvector of  $h$  which is not  $h^+$  or  $h^-$ , the second eigenvector of  $h$ , to be a point such that the intersections of the lines joining respectively  $h^+$  and the second eigenvector of  $h$ , and  $h^-$  and the second eigenvector of  $h$ , with the copy of  $P^1(\mathbb{C})$ , are off the attracting circle. However, this proof-idea fails because in order for the eigenvalues of  $h$  to be separated by some extension of the valuation that separates  $g$ , the eigenvalues and so the eigenvectors would have to be real, and it can be shown that it is not possible to choose such a real eigenvector. Indeed, the eigenvectors of  $g$  will consist of a real eigenvector and two conjugate eigenvectors, and it can be shown that the inverse of the matrix with the eigenvectors of  $g$  as columns will have a real row and two conjugate rows, and this means that the intersection of say, the line joining  $h^+$  and the second eigenvector of  $h$  with the copy of  $P^1(\mathbb{C})$  will have components in the second two eigenvectors of  $g$  which are conjugate, and therefore it will lie on the circle centre the origin and radius 1, which will be the attracting circle in question. So, alas, the proof fails, although a version of it can be made to work in the case where the ratio of the two conjugate eigenvalues of  $g$  is a root of unity.

Nevertheless we still can find a rich supply of  $h$  such that  $g$  and a sufficiently high power of  $h$  generate a free group of rank two in this case, even if this set is not Zariski dense. Any  $g$  which is diagonalizable and has a real eigenvalue and two complex conjugate eigenvalues will

be conjugate in  $SL(3, \mathbb{R})$  to a matrix of the form 
$$\begin{pmatrix} \lambda^{-2} & 0 & 0 \\ 0 & \lambda a & \lambda b \\ 0 & \lambda c & \lambda d \end{pmatrix},$$
 with  $a, b, c, d$  algebraic

and  $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{R})$ . Since  $a, b, c, d$  are algebraic, a similar argument to that of the previous chapter will show that there is a Zariski dense set of  $a', b', c', d' \in \mathbb{Q}$  such that  $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$  and a sufficiently high power of  $\begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}$  generate a free group of rank two.

If we select such  $a', b', c', d'$ , then our  $\begin{pmatrix} \lambda^{-2} & 0 & 0 \\ 0 & \lambda a & \lambda b \\ 0 & \lambda c & \lambda d \end{pmatrix}$  and a sufficiently high power of

anything of the form  $\begin{pmatrix} \nu^{-2} & k & l \\ 0 & \nu d' & \nu b' \\ 0 & \nu c' & \nu d' \end{pmatrix}$  will generate a free group of rank two, because the

map  $\begin{pmatrix} a & b & c \\ 0 & d & e \\ 0 & f & g \end{pmatrix} \mapsto \begin{pmatrix} d & e \\ f & g \end{pmatrix}$

is a homomorphism. Further, there may be other subgroups containing  $g$  which can be homomorphically mapped onto  $SL(2, \mathbb{R})$  or  $SL(2, \mathbb{C})$  for which a similar argument will work. Thus there is a rich supply of  $h$  for which  $g$  and a sufficiently high power of  $h$  will generate a free group of rank two. These considerations make it probable that the conjecture is true in the case of  $SL(3, \mathbb{Q})$ , but we do not see how to prove this at present.

---

## References

---

- [1] Pierre de la Harpe. Free groups in linear groups. *L'Enseignement Mathématique*, (29):129–144, 1983.
- [2] Otto Endler. *Valuation Theory*. Springer-Verlag, 1972.
- [3] A. Frohlich and M. J. Taylor. *Algebraic number theory*. Cambridge University Press, 1991.
- [4] D. J. H. Garling. *A Course in Galois Theory*. Cambridge University Press, 1986.
- [5] Gerald J. Janusz. *Algebraic Number Fields*. Academic Press, 1973.
- [6] John G. Ratcliffe. *Foundations of Hyperbolic Manifolds*. Springer-Verlag, 1994.