12 Endomorphism algebras

The key to improving the efficiency of elliptic curve primality proving (and many other algorithms) is the ability to directly construct an elliptic curve $E/\mathbb{F}_q$ with a specified number of rational points, rather than generating curves at random until a suitable curve is found. To do this we need to develop the theory of *complex multiplication.* As a first step in this direction we introduce the endomorphism algebra of an elliptic curve and classify the possible endomorphism algebras of an elliptic curve.

Recall from Lecture 6 that the endomorphism ring $\operatorname{End}(E)$ of an elliptic curve E/k consists of the isogenies from E to itself, together with the zero morphism; addition is defined point-wise and multiplication is composition. The ring $\operatorname{End}(E)$ is not necessarily commutative, but its center (elements that commute with every other element of the ring) always contains the multiplication-by-n maps $[n]$; there form a subring of $\operatorname{End}(E)$ isomorphic to $\mathbb{Z}$. We will identify this subring with $\mathbb{Z}$, and may write n rather than $[n]$ without risk of confusion: note that $n\phi = \phi + \cdots + \phi$ is the same as $[n]\circ\phi$. We thus have $\mathbb{Z} \subseteq \operatorname{End}(E)$, but this inclusion is not necessarily an equality. The following facts about $\operatorname{End}(E)$ were proved in Lecture 6:

- $\operatorname{End}(E)$ has no zero divisors;
- $\deg\colon \operatorname{End}(E) \to \mathbb{Z}_{\geq 0}$ defined by $\alpha \mapsto \deg\alpha$ is multiplicative (with $\deg 0 \coloneqq 0$);
- $\deg n = n^2$ for all $n \in \mathbb{Z} \subseteq \operatorname{End}(E)$;
- each $\alpha \in \operatorname{End}(E)$ has a *dual* $\hat\alpha \in \operatorname{End}(E)$ with $\alpha\hat\alpha = \hat\alpha\alpha = \deg\alpha = \deg\hat\alpha$, and $\hat{\hat\alpha} = \alpha$;
- $\hat n = n$ for all $n \in \mathbb{Z} \subseteq \operatorname{End}(E)$;
- $\widehat{\alpha+\beta} = \hat\alpha + \hat\beta$ and $\widehat{\alpha\beta} = \hat\beta\hat\alpha$ for all $\alpha,\beta \in \operatorname{End}(E)$;
- $\operatorname{tr}\alpha \coloneqq \alpha + \hat\alpha$ satisfies $\operatorname{tr}\alpha = \operatorname{tr}\hat\alpha$ and $\operatorname{tr}(\alpha+\beta) = \operatorname{tr}\alpha + \operatorname{tr}\beta$;
- $\operatorname{tr}\alpha = \deg\alpha + 1 - \deg(\alpha - 1) \in \mathbb{Z}$ for all $\alpha \in \operatorname{End}(E)$;
- α and $\hat\alpha$ are the roots of the characteristic equation $x^2 - (\operatorname{tr}\alpha)x + \deg\alpha \in \mathbb{Z}[x]$.

These facts imply that the map $\varphi \mapsto \hat\varphi$ is an *involution* of $\operatorname{End}(E)$.

Definition 12.1. An *anti-homomorphism* $\varphi\colon R \to S$ of rings is a homomorphism of their additive groups that satisfies $\varphi(1_R) = 1_S$ and $\varphi(\alpha\beta) = \varphi(\beta)\varphi(\alpha)$ for all $\alpha, \beta \in R$. An *involution* (or *anti-involution*) is an anti-homomorphism $\varphi\colon R \to R$ that is its own inverse: $\varphi \circ \varphi$ is the identity map.

A nontrivial involution of a commutative ring is an automorphism of order 2.

12.1 The endomorphism algebra of an elliptic curve

The additive group of $\operatorname{End}(E)$, like all abelian groups, is a $\mathbb{Z}$-module. Recall that if R is a commutative ring, an *R-module M* is an (additively written) abelian group that admits a scalar multiplication by R compatible with its structure as an abelian group. This means that for all $\alpha, \beta \in M$ and $r, s \in R$ we have

$$(r+s)\alpha = r\alpha + s\alpha, \qquad r\alpha + r\beta = r(\alpha+\beta), \qquad r(s\alpha) = (rs)\alpha, \qquad 1\alpha = \alpha$$

(one can check these conditions also imply $0\alpha = 0$ and $(-1)\alpha = -\alpha$).

Lecture by Andrew Sutherland

The ring $\operatorname{End}(E)$ is not only a $\mathbb{Z}$-module. Like all rings, it has a multiplication that is compatible with its structure as a $\mathbb{Z}$-module, making it a $\mathbb{Z}$-algebra. For any commutative ring R, an (associative unital) *R-algebra* A is a (not necessarily commutative) ring equipped with a ring homomorphism $R \to A$ that maps R into the center of A.[1] In our situation the map $\mathbb{Z} \to \operatorname{End}(E)$ sending n to $[n]$ is injective and we simply view $\mathbb{Z}$ as a subring of $\operatorname{End}(E)$ that necessarily lies in its center. When we have a ring A with an involution that is also an R-algebra, we typically require the involution to fix R, so that we may view it as an R-algebra involution; this holds for the involution $\alpha \mapsto \hat{\alpha}$ on our $\mathbb{Z}$-algebra $\operatorname{End}(E)$.

We now want to "upgrade" our $\mathbb{Z}$-algebra $\operatorname{End}(E)$ to a $\mathbb{Q}$-algebra (in other words, a $\mathbb{Q}$-vector space with a multiplication that is compatible with its structure as a vector space), To do this we take the tensor product of $\operatorname{End}(E)$ with $\mathbb{Q}$.

Definition 12.2. The *endomorphism algebra* of E is $\operatorname{End}^0(E) := \operatorname{End}(E) \otimes_{\mathbb{Z}} \mathbb{Q}$.

Recall that for a commutative ring R, the *tensor product* $A \otimes_R B$ of two R-modules A and B can be defined as the R-module generated by the formal symbols $\alpha \otimes \beta$ with $\alpha \in A$ and $\beta \in B$, subject to the relations

$$(\alpha_1+\alpha_2)\otimes\beta = \alpha_1\otimes\beta+\alpha_2\otimes\beta, \quad \alpha\otimes(\beta_1+\beta_2) = \alpha\otimes\beta_1+\alpha\otimes\beta_2, \quad r\alpha\otimes\beta = \alpha\otimes r\beta = r(\alpha\otimes\beta),$$

for $\alpha_1, \alpha_2 \in A$, $\beta_1, \beta_2 \in B$ and $r \in R$. The elements of $A \otimes_R B$ are finite sums of *pure tensors* $\alpha \otimes_R \beta$. We can use the relations above to simplify these sums. In general not every element of $A \otimes_R B$ can be reduced to a pure tensor, but in our situation this is in fact the case (see Lemma 12.5 below). The tensor product behaves quite differently than the direct product (for example, $A \times 0 = A$ but $A \otimes_R 0 = 0$), but we do have a canonical R-bilinear map $\varphi\colon A \times B \to A \otimes_R B$ defined by $(\alpha, \beta) \mapsto \alpha \otimes \beta$. This map is universal in the sense that every R-bilinear map of R-modules $\psi\colon A \times B \to C$ can be written uniquely as a composition

$$\begin{tikzcd}
A \times B \arrow[r, "\varphi"] \arrow[dr, "\psi"'] & A \otimes_R B \arrow[d, dashed, "\exists!"] \\
& C
\end{tikzcd}$$

This *universal property* can also be taken as a definition of the tensor product (without guaranteeing its existence).

When A and B are not only R-modules but R-algebras, we give the tensor product $A \otimes_R B$ the structure of an R-algebra by defining multiplication of purse tensors

$$(\alpha_1 \otimes \beta_1)(\alpha_2 \otimes \beta_2) = \alpha_1\alpha_2 \otimes \beta_1\beta_2$$

and extending linearly; this means we can compute $(\sum_i \alpha_i \otimes \beta_i)(\sum_j \alpha_j \otimes \beta_j)$ using the distributive law. The multiplicative identity is necessarily $1_A \otimes 1_B$. The R-algebras A and B can be canonically mapped to $A \otimes_R B$ via $\alpha \mapsto \alpha \otimes 1_B$ and $\beta \mapsto 1_A \otimes \beta$. These maps need not be injective; indeed, $A \otimes_R B$ may be the zero ring even when A and B are not.

Example 12.3. The tensor product $\mathbb{Z}/2\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/3\mathbb{Z}$ is the zero ring. To see why, note that for any pure tensor $\alpha \otimes \beta$ we have

$$\alpha \otimes \beta = a \otimes -2\beta = 2\alpha \otimes -\beta = 0 \otimes -\beta = 0 \otimes 0 = 0.$$

[1] Here we consider only associative unital algebras; one can define a more general notion of an R-algebra that is not necessarily a ring (Lie algebras, for example).

Example 12.4. If V is a k-vector space with basis $(v_1,\ldots,v_n)$ and L/k is any field extension, then $V\otimes_k L$ is an L-vector space with basis $(v_1\otimes 1,\ldots,v_n\otimes 1)$; multiplication by scalars in L takes place on the RHS of each pure tensor. This implies that if V is a k-algebra of k-kdimension n, then $V\otimes_k L$ is an L-algebra of L-dimension n.

Lemma 12.5. *Let R be an integral domain with fraction field B, and let A be an R-algebra. Every element of $A\otimes_R B$ can be written as a pure tensor $\alpha\otimes\beta$.*

Proof. It suffices to show that $\alpha_1\otimes\beta_1+\alpha_2\otimes\beta_2$ can be written as $\alpha_3\otimes\beta_3$. Let $\beta_1=r_1/s_1$ and $\beta_2=r_2/s_2$ with $r_1,r_2,s_1,s_2\in R$. Then

$$\begin{aligned}\alpha_1\otimes\beta_1+\alpha_2\otimes\beta_2&=\alpha_1\otimes\frac{r_1}{s_1}+\alpha_2\otimes\frac{r_2}{s_2}\\&=\alpha_1\otimes\frac{r_1s_2}{s_1s_2}+\alpha_2\otimes\frac{r_2s_1}{s_1s_2}\\&=(r_1s_2\alpha_1)\otimes\frac{1}{s_1s_2}+(r_2s_1\alpha_2)\otimes\frac{1}{s_1s_2}\\&=(r_1s_2\alpha_1+r_2s_1\alpha_2)\otimes\frac{1}{s_1s_2},\end{aligned}$$

so we may take $\alpha_3=r_1s_2\alpha_1+r_2s_1\alpha_2$ and $\beta_3=1/(s_1s_2)$. □

The lemma implies that every element of $\operatorname{End}^0(E)=\operatorname{End}(E)\otimes_{\mathbb Z}\mathbb Q$ can be written as $\phi\otimes r$ for some $\phi\in\operatorname{End}(E)$ and $r\in\mathbb Q$; to simplify notation we will simply use $r\phi$ to denote $\phi\otimes r$. Note that this representation is not unique (if $r'=r/n$ and $\phi'=n\phi$ then $r'\phi'=r\phi$). The only difference between $r\phi$, with $r\in\mathbb Q$, and $n\phi$, with $n\in\mathbb Z$, is that the former is not necessarily an endomorphism, but if we multiply $r\alpha$ by the denominator of r we will get an element of $\operatorname{End}^0(E)$ that corresponds to an endomorphism.

The canonical homomorphisms $\operatorname{End}(E)\to\operatorname{End}^0(E)$ and $\mathbb Q\to\operatorname{End}^0(E)$ are injective, because $\operatorname{End}(E)$ and $\mathbb Q$ are torsion-free $\mathbb Z$-algebras, so we may identify both $\operatorname{End}(E)$ and $\mathbb Q$ with corresponding subrings of $\operatorname{End}^0(E)$ that intersect in $\mathbb Z$. Every element of $\operatorname{End}^0(E)$ has an integer multiple that lies in the subring $\operatorname{End}(E)$, and the subring $\mathbb Q$ lies in the center of $\operatorname{End}^0(E)$, which makes $\operatorname{End}^0(E)$ a $\mathbb Q$-algebra. We also note that $\operatorname{End}^0(E)$ has no zero divisors: if $(r\phi)(r'\phi')=rr'\phi\phi'=0$ then either $rr'=0$ or $\phi\phi'=0$, so one of r,r',ϕ,ϕ' is zero (since $\mathbb Q$ and $\operatorname{End}(E)$ have no zero divisors); this implies that one of $r\phi$ or $r'\phi'$ is zero.

12.2 The Rosati involution and the reduced norm and trace

We now extend the involution $\alpha\mapsto\hat\alpha$ on $\operatorname{End}(E)$ to $\operatorname{End}^0(E)$ by defining $\widehat{r\alpha}=r\hat\alpha$ for all $r\in\mathbb Q$. This implies that $\hat r=r$ for all $r\in\mathbb Q$ (take $\alpha=1$), and therefore $\hat{\hat\alpha}=\alpha$ holds for all $\alpha\in\operatorname{End}^0(E)$. We also have $\widehat{\alpha\beta}=\hat\beta\hat\alpha$ and $\widehat{\alpha+\beta}=\hat\alpha+\hat\beta$ for all $\alpha,\beta\in\operatorname{End}^0(E)$, since these hold for elements of $\operatorname{End}(E)$ and scalars are fixed by $\alpha\mapsto\hat\alpha$ and commute. Thus the map $\alpha\mapsto\hat\alpha$ is an involution of the $\mathbb Q$-algebra $\operatorname{End}^0(E)$, and it is known as the *Rosati involution.*

The Rosati involution allows us to extend the notions of degree and trace on $\operatorname{End}(E)$ to a norm and a trace defined on all of $\operatorname{End}^0(E)$.

Definition 12.6. Let $\alpha\in\operatorname{End}^0(E)$. The (reduced) *norm* of α is $\mathrm N\alpha=\alpha\hat\alpha$ and the (reduced) *trace* of α is $\mathrm T\alpha=\alpha+\hat\alpha$.[2]

[2] $\mathrm N\alpha$ and $\mathrm T\alpha$ are often called the reduced norm and reduced trace and may be denoted $\operatorname{Nrd}\alpha$ and $\operatorname{Trd}\alpha$ to distinguish them from the more general notion of norm and trace in a $\mathbb Q$-algebra, which involve taking the determinant or trace of the $\mathbb Q$-linear transformation $\beta\mapsto\alpha\beta$ (this coincides with the reduced norm and trace when $\dim_{\mathbb Q}\operatorname{End}^0(E)=2$, but not otherwise). We shall only consider the reduced norm and trace.

We now show that $\mathrm{N}\alpha$ and $\mathrm{T}\alpha$ lie in $\mathbb{Q}$, and prove some other facts we will need.

Lemma 12.7. *For all $\alpha \in \operatorname{End}^0(E)$ we have $\mathrm{N}\alpha \in \mathbb{Q}_{\geq 0}$, with $\mathrm{N}\alpha = 0$ if and only if $\alpha = 0$. We also have $\mathrm{N}\hat{\alpha} = \mathrm{N}\alpha$ and $\mathrm{N}(\alpha\beta) = \mathrm{N}\alpha\mathrm{N}\beta$ for all $\alpha, \beta \in \operatorname{End}^0(E)$.*

Proof. Write $\alpha = r\phi$, with $r \in \mathbb{Q}$, $\phi \in \operatorname{End}(E)$. Then $\mathrm{N}\alpha = \alpha\hat{\alpha} = r^2 \deg \phi \geq 0$. If r or ϕ is zero then $\alpha = 0$ and $\mathrm{N}\alpha = 0$, and otherwise $\mathrm{N}\alpha > 0$. We have $\alpha\mathrm{N}\hat{\alpha} = \alpha\hat{\alpha}\alpha = (\mathrm{N}\alpha)\alpha = \alpha\mathrm{N}\alpha$, so $\mathrm{N}\hat{\alpha} = \mathrm{N}\alpha$ when $\alpha \neq 0$ (since $\operatorname{End}^0(E)$ has no zero divisors), and $\mathrm{N}\hat{\alpha} = \mathrm{N}\alpha = 0$ when $\alpha = 0$. Finally, for any $\alpha, \beta \in \operatorname{End}^0(E)$ we have

$$\mathrm{N}(\alpha\beta) = \alpha\beta\widehat{\alpha\beta} = \alpha\beta\hat{\beta}\hat{\alpha} = \alpha(\mathrm{N}\beta)\hat{\alpha} = \alpha\hat{\alpha}\mathrm{N}\beta = \mathrm{N}\alpha\mathrm{N}\beta. \qquad \square$$

Corollary 12.8. *Every nonzero $\alpha \in \operatorname{End}^0(E)$ has a multiplicative inverse α^{-1}.*

Proof. If we put $\beta = \hat{\alpha}/\mathrm{N}\alpha$, then $\alpha\beta = \mathrm{N}\alpha/\mathrm{N}\alpha = 1$ and $\beta\alpha = \mathrm{N}\hat{\alpha}/N\alpha = 1$, so $\beta = \alpha^{-1}$. □

The corollary implies that $\operatorname{End}^0(E)$ is a *division ring*; it satisfies all the field axioms except that multiplication need not be commutative. This means that $\operatorname{End}^0(E)$ is a field if and only if it is commutative.

Lemma 12.9. *For all $\alpha \in \operatorname{End}^0(E)$ we have $\mathrm{T}\hat{\alpha} = \mathrm{T}\alpha \in \mathbb{Q}$. For any $r \in \mathbb{Q}$, $\alpha, \beta \in \operatorname{End}^0(E)$ we have $\mathrm{T}(\alpha + \beta) = \mathrm{T}\alpha + \mathrm{T}\beta$, and $\mathrm{T}(r\alpha) = r\mathrm{T}\alpha$.*

Proof. We first note that $\mathrm{T}\hat{\alpha} = \hat{\alpha} + \hat{\hat{\alpha}} = \hat{\alpha} + \alpha = \alpha + \hat{\alpha} = \mathrm{T}\alpha$, and

$$\mathrm{T}\alpha = \alpha + \hat{\alpha} = 1 + \alpha\hat{\alpha} - (1 - \alpha)(1 - \hat{\alpha}) = 1 + \mathrm{N}\alpha - \mathrm{N}(1 - \alpha) \in \mathbb{Q}.$$

We also have

$$\mathrm{T}(\alpha + \beta) = \alpha + \beta + \widehat{\alpha + \beta} = \alpha + \beta + \hat{\alpha} + \hat{\beta} = \alpha + \hat{\alpha} + \beta + \hat{\beta} = \mathrm{T}\alpha + \mathrm{T}\beta.$$

and

$$\mathrm{T}(r\alpha) = r\alpha + \widehat{r\alpha} = r\alpha + \hat{\alpha}\hat{r} = r\alpha + \hat{\alpha}r = r\alpha + r\hat{\alpha} = r(\alpha + \hat{\alpha}) = r\mathrm{T}\alpha,$$

since $\mathbb{Q}$ lies in the center of $\operatorname{End}^0(E)$ and is fixed by the Rosati involution. □

Lemma 12.10. *Let $\alpha \in \operatorname{End}^0(E)$. Then α and $\hat{\alpha}$ are roots of the polynomial*

$$x^2 - (\mathrm{T}\alpha)x + \mathrm{N}\alpha \in \mathbb{Q}[x].$$

Proof. We have

$$0 = (\alpha - \alpha)(\alpha - \hat{\alpha}) = \alpha^2 - \alpha(\alpha + \hat{\alpha}) + \alpha\hat{\alpha} = \alpha^2 - (\mathrm{T}\alpha)\alpha + \mathrm{N}\alpha,$$

and similarly for $\hat{\alpha}$, since $\mathrm{T}\hat{\alpha} = \mathrm{T}\alpha$ and $\mathrm{N}\hat{\alpha} = \mathrm{N}\alpha$. □

Corollary 12.11. *For any nonzero $\alpha \in \operatorname{End}^0(E)$, if $\mathrm{T}\alpha = 0$ then $\alpha^2 = -\mathrm{N}\alpha < 0$. An element $\alpha \in \operatorname{End}^0(E)$ is fixed by the Rosati involution if and only if $\alpha \in \mathbb{Q}$.*

Proof. The first statement follows immediately from $\alpha^2 - (\mathrm{T}\alpha)\alpha + \mathrm{N}\alpha = 0$. For the second, we have $\hat{r} = r$ for $r \in \mathbb{Q}$, and if $\hat{\alpha} = \alpha$ then $\mathrm{T}\alpha = \alpha + \hat{\alpha} = 2\alpha$, so $\alpha = (\mathrm{T}\alpha)/2 \in \mathbb{Q}$. □

12.3 Quaternion algebras

Before we can give a complete classification of the possible endomorphism algebras $\operatorname{End}^0(E)$ that can arise, we need to introduce quaternion algebras.

Definition 12.12. A *quaternion algebra* over a field k is a k-algebra that has a k-basis of the form $\{1, \alpha, \beta, \alpha\beta\}$, with $\alpha^2, \beta^2 \in k^\times$ and $\alpha\beta = -\beta\alpha$.

Let H be a quaternion algebra over a field k. Then H is a 4-dimensional k-vector space with basis $\{1, \alpha, \beta, \alpha\beta\}$, and we may distinguish the subspace $k \subseteq H$ spanned by 1, which does not depend on the choice of α and β. The complementary subspace H_0 (spanned by $\alpha, \beta, \alpha\beta$) is the space of *pure quaternions*. Every $\gamma \in H$ has a unique decomposition of the form $a + \gamma_0$ with $a \in k$ and $\gamma_0 \in H_0$. The element $\hat\gamma := a - \gamma_0$ is the *conjugate* of γ. If γ is a pure quaternion then $\hat\gamma = -\gamma$, and for $\gamma \in k$ we have $\hat\gamma = \gamma$.

The map $\gamma \mapsto \hat\gamma$ is an involution of the k-algebra H, and we define the (reduced) trace $\mathrm{T}\gamma := \gamma + \hat\gamma$ and (reduced) norm $\mathrm{N}\gamma := \gamma\hat\gamma$, both of which lie in k. It is easy to check that $\mathrm{T}\gamma = \mathrm{T}\hat\gamma$ and $\mathrm{N}\gamma = \mathrm{N}\hat\gamma$, the trace is additive, the norm is multiplicative, and for $a \in k$ we have $\mathrm{T}a = 2a$ and $\mathrm{N}a = a^2$.

Lemma 12.13. *A quaternion algebra is a division ring if and only if* $\mathrm{N}\gamma = 0$ *implies* $\gamma = 0$.

Proof. Let γ be a nonzero element of a quaternion algebra H. Then $\hat\gamma \neq 0$ (since $\hat 0 = 0 \neq \gamma$) If H is a division ring, then x has an inverse γ^{-1} and $\gamma^{-1}\mathrm{N}\gamma = \gamma^{-1}\gamma\hat\gamma = \hat\gamma \neq 0$, so $\mathrm{N}\gamma \neq 0$. Conversely, if $\mathrm{N}\gamma \neq 0$ then $\gamma(\hat\gamma/\mathrm{N}\gamma) = 1$ and $(\hat\gamma/\mathrm{N}\gamma)\gamma = 1$, so γ has an inverse $\hat\gamma/\mathrm{N}\gamma$, which implies that H is a division ring. □

Example 12.14. The most well known example of a quaternion algebra is the ring of *Hamilton quaternions* (or *Hamiltonians*) $\mathbb{H}$: the $\mathbb{R}$-algebra with basis $\{1, i, j, ij\}$, where $i^2 = j^2 = -1$ and $ij = -ji$ (the product ij is often denoted k). This was the first example of a noncommutative division ring and has many applications in mathematics and physics.

Remark 12.15. The elements $i, j \in \mathbb{H}$ have the same characteristic polynomial $x^2 + 1$, but they are not conjugate; in fact, $x^2 + 1$ has infinitely many solutions in $\mathbb{H}$ (one can take any $x = bi + cj + dk$ with $b^2 + c^2 + d^2 = 1$).

Example 12.16. Let $H = \mathrm{M}_2(k)$ be the ring of 2×2 matrices over a field k with

$$\alpha := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \beta := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \alpha\beta = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad \beta\alpha = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

then $\alpha^2 = \beta^2 = 1 \in k^\times$ and $\alpha\beta = -\beta\alpha$, so H is a quaternion algebra, but it is not a division ring, by Lemma 12.13, since $\mathrm{N}(1+\alpha) = (1+\alpha)(1-\alpha) = 0$ but $1 + \alpha \neq 0$. Every quaternion algebra that is not a division ring arises in this way. Such quaternion algebras are said to be *split*, while those that are division rings are called *non-split*.

12.4 Classification theorem for endomorphism algebras

Theorem 12.17. *Let E/k be an elliptic curve. Then* $\operatorname{End}^0(E)$ *is isomorphic to one of:*

(i) *the field of rational numbers* $\mathbb{Q}$*;*

(ii) *an imaginary quadratic field* $\mathbb{Q}(\alpha)$ *with* $\alpha^2 < 0$*;*

(iii) *a quaternion algebra* $\mathbb{Q}(\alpha,\beta)$ *with* $\alpha^2,\beta^2<0$.

Proof. We always have $\mathbb{Q}\subseteq \operatorname{End}^0(E)$, and if $\mathbb{Q}=\operatorname{End}^0(E)$ we are in case (i).

Otherwise, let α be an element of $\operatorname{End}^0(E)$ not in $\mathbb{Q}$. By replacing α with $\alpha-\frac{1}{2}\mathrm{T}\alpha$, we may assume without loss of generality that $\mathrm{T}\alpha=0$, since

$$\mathrm{T}\left(\alpha-\frac{1}{2}\mathrm{T}\alpha\right)=\mathrm{T}\alpha-\frac{1}{2}\mathrm{T}\mathrm{T}\alpha=\mathrm{T}\alpha-\frac{1}{2}2\mathrm{T}\alpha=0,$$

where $\mathrm{T}\mathrm{T}\alpha=2\mathrm{T}\alpha$ because $\mathrm{T}\alpha\in\mathbb{Q}$. Now $\alpha^2<0$, by Corollary 12.11, and $\mathbb{Q}(\alpha)\subseteq\operatorname{End}^0(E)$ is an imaginary quadratic field. If $\mathbb{Q}(\alpha)=\operatorname{End}^0(E)$ then we are in case (ii).

Otherwise, let β be an element of $\operatorname{End}^0(E)$ not in $\mathbb{Q}(\alpha)$. As with α, we may assume without loss of generality that $\mathrm{T}\beta=0$ so that $\beta^2<0$. By replacing β with

$$\beta-\frac{\mathrm{T}(\alpha\beta)}{2\alpha^2}\alpha \tag{1}$$

we can also assume $\mathrm{T}(\alpha\beta)=0$ (to check, multiply (1) by α and compute the trace; replacing β with (1) does not change its trace because $\mathrm{T}\alpha=0$). Thus $\mathrm{T}\alpha=\mathrm{T}\beta=\mathrm{T}(\alpha\beta)=0$. This implies $\alpha=-\hat{\alpha}$, $\beta=-\hat{\beta}$, and $\alpha\beta=-\widehat{\alpha\beta}=-\hat{\beta}\hat{\alpha}$. Substituting the first two equalities into the third yields $\alpha\beta=-\beta\alpha$. Applying this together with the fact that $\alpha^2<0$ and $\beta^2<0$ lie in $\mathbb{Q}$, it is clear that $\{1,\alpha,\beta,\alpha\beta\}$ spans $\mathbb{Q}(\alpha,\beta)$ as a $\mathbb{Q}$-vector space.

To show that $\mathbb{Q}(\alpha,\beta)$ is a quaternion algebra, we need to show that 1, α, β, and $\alpha\beta$ are $\mathbb{Q}$-linearly independent. By construction, 1, α, β are linearly independent: note $\beta\notin\mathbb{Q}(\alpha)$ implies $\alpha\notin\mathbb{Q}(\beta)$, since $\mathbb{Q}(\beta)=\{r+s\beta : r,s\in\mathbb{Q}\}$ (because $\beta^2\in\mathbb{Q}$). Now suppose for the sake of contradiction that

$$\alpha\beta=a+b\alpha+c\beta,$$

for some $a,b,c\in\mathbb{Q}$. We must have $a,b,c\neq 0$, since $\beta,\alpha\beta\notin\mathbb{Q}(\alpha)$ and $\alpha\notin\mathbb{Q}(\beta)$. Squaring both sides yields

$$(\alpha\beta)^2=(a^2+b^2\alpha^2+c^2\beta^2)+2a(b\alpha+c\beta)+bc(\alpha\beta+\beta\alpha).$$

The LHS lies in $\mathbb{Q}$, since $\mathrm{T}(\alpha\beta)=0$, as does the first term on the RHS, since $\mathrm{T}\alpha=\mathrm{T}\beta=0$. The last term on the RHS is zero, since $\alpha\beta=-\beta\alpha$. Thus $d:=2a(b\alpha+c\beta)$ lies in $\mathbb{Q}$, but then $\beta=(d-2ab\alpha)/(2ac)$ lies $\in\mathbb{Q}(\alpha)$, a contradiction.

Thus $\mathbb{Q}(\alpha,\beta)\subseteq\operatorname{End}^0(E)$ is a quaternion algebra with $\alpha^2,\beta^2<0$. If $\mathbb{Q}(\alpha,\beta)=\operatorname{End}^0(E)$ then we are in case (iii).

Otherwise, let γ be an element of $\operatorname{End}^0(E)$ that does not lie in $\mathbb{Q}(\alpha,\beta)$. As with β, we may assume without loss of generality that $\mathrm{T}\gamma=0$ and $\mathrm{T}(\alpha\gamma)=0$, which implies $\alpha\gamma=-\gamma\alpha$. Then $\alpha\beta\gamma=-\beta\alpha\gamma=\beta\gamma\alpha$, so α commutes with $\beta\gamma$. By Lemma 12.18 below, $\beta\gamma\in\mathbb{Q}(\alpha)$. This implies $\gamma\in\mathbb{Q}(\alpha,\beta)$, contrary to our assumption that $\gamma\notin\mathbb{Q}(\alpha,\beta)$. □

Lemma 12.18. *If* $\alpha,\beta\in\operatorname{End}^0(E)$ *commute and* $\alpha\notin\mathbb{Q}$ *then* $\beta\in\mathbb{Q}(\alpha)$.

Proof. As in the proof of the Theorem 12.17, we can transform α and β so that $\mathrm{T}\alpha=\mathrm{T}\beta=\mathrm{T}(\alpha\beta)=0$, and therefore $\alpha\beta=-\beta\alpha$; this involves replacing α with $\alpha-r$ and then replacing β with $\beta-s-t\alpha$ for some $r,s,t\in\mathbb{Q}$; if α and β commute then so do all $\mathbb{Q}$-linear combinations, so the hypothesis still holds. We then have $\alpha\beta+\beta\alpha=2\alpha\beta=0$, which implies $\alpha=0$ or $\beta=0$, since $\operatorname{End}^0(E)$ has no zero divisors. We cannot have $\alpha=0$, since $\alpha\notin\mathbb{Q}$, so $\beta=0\in\mathbb{Q}(\alpha)$. □

Remark 12.19. In the proofs of Theorem 12.17 and Lemma 12.18 we never used the fact that $\operatorname{End}^0(E)$ is the endomorphism algebra of an elliptic curve. Indeed, one can replace $\operatorname{End}^0(E)$ with any $\mathbb{Q}$-algebra A possessing an involution $\alpha \mapsto \hat{\alpha}$ that fixes $\mathbb{Q}$ such that the associated norm $\mathrm{N}\alpha = \alpha\hat{\alpha}$ maps nonzero elements of A to positive elements of $\mathbb{Q}$; all other properties of $\operatorname{End}^0(E)$ that we used can be derived from these.

Having classified the possible endomorphism algebras $\operatorname{End}^0(E)$, our next task is to classify the possible endomorphism rings $\operatorname{End}(E)$. We begin with the following corollary to Theorem 12.17.

Corollary 12.20. *Let E/k be an elliptic curve. The endomorphism ring* $\operatorname{End}(E)$ *is a free $\mathbb{Z}$-module of rank r, where $r = 1, 2, 4$ is the dimension of* $\operatorname{End}^0(E)$ *as a $\mathbb{Q}$-vector space.*

Recall that a free $\mathbb{Z}$-module of rank r is an abelian group isomorphic to $\mathbb{Z}^r$.

Proof. Let us pick a basis $\{e_1, \ldots, e_r\}$ for $\operatorname{End}^0(E)$ as a $\mathbb{Q}$-basis with the property that $\mathrm{T}(e_i e_j) = 0$ unless $i = j$ (use the basis $\{1, \alpha\}$ when $\operatorname{End}^0(E) = \mathbb{Q}(\alpha)$ and $\{1, \alpha, \beta, \alpha\beta\}$ when $\operatorname{End}^0 = \mathbb{Q}(\alpha, \beta)$, where α and β are constructed as in the proof of Theorem 12.17). After multiplying by suitable integers if necessary, we can assume without loss of generality that $e_1, \ldots e_r \in \operatorname{End}(E)$ (this doesn't change $\mathrm{T}(e_i e_j) = 0$ for $i \neq j$).

For any $\mathbb{Z}$-module $A \subseteq \operatorname{End}^0(E)$ we have an associated *dual* $\mathbb{Z}$-module

$$A^* := \{\alpha \in \operatorname{End}^0(E) : \mathrm{T}(\alpha\phi) \in \mathbb{Z}\ \forall \phi \in A\}.$$

Note that A^* is closed under addition and multiplication by integers (if $\mathrm{T}(\alpha\phi), \mathrm{T}(\beta\phi) \in \mathbb{Z}$ then $\mathrm{T}(m\alpha\phi + n\beta\phi) \in \mathbb{Z}$ for all $m, n \in \mathbb{Z}$), so A^* is also a $\mathbb{Z}$-module. It is clear from the definition that if A and B are any $\mathbb{Z}$-modules in $\operatorname{End}(E)^0$, then $A \subseteq B$ implies $B^* \subseteq A^*$ (making A bigger imposes a stronger constraint on A^*).

Now let A be the $\mathbb{Z}$-module spanned by $e_1, \ldots, e_r \in \operatorname{End}(E)$. Then $A \subseteq \operatorname{End}(E)$, and therefore $\operatorname{End}(E)^* \subseteq A^*$. We also note that $\operatorname{End}(E) \subseteq \operatorname{End}(E)^*$, since $\mathrm{T}(\alpha\phi) \in \mathbb{Z}$ for all $\alpha, \phi \in \operatorname{End}(E)$. Thus

$$A \subseteq \operatorname{End}(E) \subseteq \operatorname{End}(E)^* \subseteq A^*.$$

We can write any $\alpha \in A^* \subseteq \operatorname{End}^0(E)$ as $a_1 e_1 + \cdots + a_r e_r$ for some $a_1, \ldots, a_r \in \mathbb{Q}$ (since $e_1, \ldots, e_r$ is a $\mathbb{Q}$-basis for $\operatorname{End}^0(E)$). For each e_i we then have

$$\mathrm{T}(\alpha e_i) = a_1 \mathrm{T}(e_1 e_i) + \cdots + a_r \mathrm{T}(e_r e_i) = a_i \mathrm{T}(e_i^2),$$

since $\mathrm{T}(e_i e_j) = 0$ for $i \neq j$, and $\mathrm{T}(\alpha e_i) = a_i \mathrm{T}(e_i^2) \in \mathbb{Z}$ since $\alpha \in A^*$ and $e_i \in A$. Thus a_i is an integer multiple of $1/\mathrm{T}(e_i^2)$, and it follows that $\{e_1/\mathrm{T}(e_1^2), \ldots, e_r/\mathrm{T}(e_r^2)\}$ is a basis for A^* as a $\mathbb{Z}$-module, which is therefore a free $\mathbb{Z}$-module of rank r, as is A (both are torsion free because $\operatorname{End}^0(E)$ is torsion free). It follows that $\operatorname{End}(E)$ and $\operatorname{End}(E)^*$ both free $\mathbb{Z}$-modules of rank r, since they are both contained in and contain a free $\mathbb{Z}$-module of rank r (every subgroup of $\mathbb{Z}^r$ is isomorphic to $\mathbb{Z}^s$ for some $0 \leq s \leq r$).[3] □

Definition 12.21. An elliptic curve E for which $\operatorname{End}(E) \not\simeq \mathbb{Z}$ is said to have *complex multiplication.*

[3] More generally, if R is a principal ideal domain (PID) then every submodule of a free R-module of rank r is free of rank $s \leq r$. This fails when R is not a PID (submodules of a free module need not be free)

It follows from Theorem 12.17 that if E has complex multiplication then $\operatorname{End}^0(E)$ is either an imaginary quadratic field or a quaternion algebra. Each element of $\operatorname{End}(E)$ that does not lie in $\mathbb{Z}$ is the root of quadratic polynomial in $\mathbb{Z}[x]$ that has no real roots, which we could view as a complex number (an algebraic integer, in fact). Elements ϕ of $\operatorname{End}(E)$ that lie in $\mathbb{Z}$ correspond to multiplication by some integer n, and we may view elements of $\operatorname{End}(E)$ that do not lie in $\mathbb{Z}$ as "multiplication" by some complex number that corresponds to an algebraic integer that is a root of the characteristic polynomial of ϕ.

12.5 Orders in $\mathbb{Q}$-algebras

Definition 12.22. Let K be a $\mathbb{Q}$-algebra of finite dimension r as a $\mathbb{Q}$-vector space. An *order* $\mathcal{O}$ in K is a subring of K that is a free $\mathbb{Z}$-module of rank r. Equivalently, $\mathcal{O}$ is a subring of K that is finitely generated as a $\mathbb{Z}$-module and satisfies $K = \mathcal{O} \otimes_{\mathbb{Z}} \mathbb{Q}$.

Note that an order is required to be both a lattice (a free $\mathbb{Z}$-module of maximal rank) *and* a ring; in particular it must contain 1.

Example 12.23. The integers $\mathbb{Z}$ are the unique example of an order in $\mathbb{Q}$. Non-examples include the even integers, which is a lattice but not a ring, and the set $\{a/2^n : a, n \in \mathbb{Z}\}$, which is a ring but not a lattice (because it is not finitely generated as a $\mathbb{Z}$-module).

It follows from Corollary 12.20 that the endomorphism ring $\operatorname{End}(E)$ is an order in the $\mathbb{Q}$-algebra $\operatorname{End}^0(E)$. Note that if $\operatorname{End}^0(E) = \mathbb{Q}$, then we must have $\operatorname{End}(E) = \mathbb{Z}$, but in general there are many infinitely many non-isomorphic possibilities for $\operatorname{End}(E)$.

Every order lies in some *maximal order* (an order that is not contained in any other); this follows from an application of Zorn's lemma, using the fact that elements of an order necessarily have monic minimal polynomials. In general, maximal orders need not be unique, but when the $\mathbb{Q}$-algebra K is a number field (a finite extension of $\mathbb{Q}$), this is the case. In view of Theorem 12.17, we are primarily interested in the case where K is an imaginary quadratic field, but it is just as easy to prove this for all number fields. We first need to recall a few standard results from algebraic number theory.

Definition 12.24. An *algebraic number* α is a complex number that is the root of a polynomial with coefficients in $\mathbb{Q}$. An *algebraic integer* is a complex number that is the root of a monic polynomial with coefficients in $\mathbb{Z}$.

Two fundamental results of algebraic number theory are (1) the set of algebraic integers in a number field form a ring, and (2) every number field has an *integral basis* (a basis whose elements are algebraic integers). The following theorem gives a more precise statement.

Theorem 12.25. *The set of algebraic integers $\mathcal{O}_K$ in a number field K form a ring that is a free $\mathbb{Z}$-module of rank r, where $r = [K : \mathbb{Q}]$ is the dimension of K as a $\mathbb{Q}$-vector space.*

Proof. See Theorem 2.1 and Corollary 2.30 in [1] (or Theorems 2.9 and 2.16 in [3]).[4] □

Theorem 12.26. *The ring of integers $\mathcal{O}_K$ of a number field K is its unique maximal order.*

[4] The proof of the second part of this theorem is essentially the same as the proof of Corollary 12.20; instead of the reduced trace in $\operatorname{End}^0(E)$, one uses the trace map from K to $\mathbb{Q}$, which has similar properties.

Proof. The previous theorem implies that $\mathcal{O}_K$ is an order. To show that it is the unique maximal order, we need to show that every order $\mathcal{O}$ in K is contained in $\mathcal{O}_K$. It suffices to show that every $\alpha \in \mathcal{O}$ is an algebraic integer. Viewing $\mathcal{O}$ as a $\mathbb{Z}$-lattice of rank $r = [K : \mathbb{Q}]$, consider the sublattice generated by all powers of α. Let $[\beta_1, \ldots, \beta_r]$ be a basis for this sublattice, where each β_i is a $\mathbb{Z}$-linear combination of powers of α. Let n be an integer larger than any of the exponents in any of the powers of α that appear in any β_i. Then $\alpha^n = c_1\beta_1 + \cdots + c_r\beta_r$, for some $c_1, \ldots, c_n \in \mathbb{Z}$, and this determines a monic polynomial of degree n with α as a root. Therefore α is an algebraic integer. □

Finally, we characterize the orders in imaginary quadratic fields, which are the number fields we are most interested in.

Theorem 12.27. *Let K be an imaginary quadratic field with ring of integers $\mathcal{O}_K$. The orders $\mathcal{O}$ in K are precisely the subrings $\mathbb{Z} + f\mathcal{O}_K$, where f is any positive integer.*

Proof. The maximal order $\mathcal{O}_K$ is a free $\mathbb{Z}$-module (a lattice) of rank 2 that contains 1, so it has a $\mathbb{Z}$-basis of the form $[1, \tau]$ for some $\tau \notin \mathbb{Z}$. Let $\mathcal{O} = \mathbb{Z} + f\mathcal{O}_K$. It is clear that $\mathcal{O}$ is a sub-lattice of $\mathcal{O}_K$ that properly contains $\mathbb{Z}$, hence it is of rank 2. The $\mathbb{Z}$-module $\mathcal{O}$ is a subset of the ring $\mathcal{O}_K$ and contains 1, so to show that $\mathcal{O}$ is a ring it suffices to show that it is closed under multiplication. So let $a + f\alpha$ and $b + f\beta$ be arbitrary elements of $\mathcal{O}$, with $a, b \in \mathbb{Z}$ and $\alpha, \beta \in \mathcal{O}_K$. Then

$$(a + f\alpha)(b + f\beta) = ab + af\beta + bf\alpha + f^2\alpha\beta = ab + f(a\beta + b\alpha + f\alpha\beta) \in \mathcal{O},$$

since $ab \in \mathbb{Z}$ and $(a\beta + b\alpha + f\alpha\beta) \in \mathcal{O}_K$. So $\mathcal{O}$ is a subring of K. To see that $\mathcal{O}$ is an order, note that $\mathcal{O} \otimes_{\mathbb{Z}} \mathbb{Q} = \mathcal{O}_K \otimes_{\mathbb{Z}} \mathbb{Q} = K$.

Now let $\mathcal{O}$ be any order in K. Then $\mathcal{O}$ is a rank-2 sub-lattice of $\mathcal{O}_K = [1, \tau]$ that contains 1, so $\mathcal{O}$ must contain an integer multiple of τ. Let f be the least positive integer for which $f\tau \in \mathcal{O}$. The lattice $[1, f\tau]$ lies in $\mathcal{O}$, and we claim that in fact $\mathcal{O} = [1, f\tau]$. Any element α of $\mathcal{O}$ must lie in $\mathcal{O}_K$ and is therefore of the form $\alpha = a + b\tau$ for some $a, b \in \mathbb{Z}$. The element $b\tau = \alpha - a$ then lies in $\mathcal{O}$, and the minimality of f implies that f divides b. Thus $\mathcal{O} = [1, f\tau] = \mathbb{Z} + f\mathcal{O}_K$. □

Remark 12.28. In the theorem above we never actually used the fact that the quadratic field K is imaginary; in fact, the theorem holds for real quadratic fields as well.

The integer f in Theorem 12.27 is called the *conductor* of the order $\mathcal{O} = \mathbb{Z} + f\mathcal{O}_K$. It is equal to the index $[\mathcal{O}_K : \mathcal{O}]$, which is necessarily finite.

References

[1] J. S. Milne, *Algebraic number theory*, course notes, version 3.06, 2014.

[2] Joseph H. Silverman, *The arithmetic of elliptic curves*, second edition, Springer 2009.

[3] Ian Stewart and David Tall, *Algebraic number theory and Fermat's last theorem*, third edition, A.K. Peters, 2002.