Chapter 5

The Real Case

5.1 Extending the Field

Suppose $\mathcal{E}$ is an elliptic curve defined over k, given by the equation

$$F(X, Y, Z) = 0.$$

and suppose K is an extension field of k:

$$k \subset K.$$

Then the same equation defines an elliptic curve over K; and the group $\mathcal{E}(k)$ of points defined over k (if non-empty) is a subgroup of $\mathcal{E}(K)$:

$$\mathcal{E}(k) \subset \mathcal{E}(K).$$

The study of $\mathcal{E}(K)$ often gives us valuable information on $\mathcal{E}(k)$.

We shall be particularly interested in curves over the rationals: $k = \mathbb{Q}$. In this case there are several candidates for K: the reals $\mathbb{R}$; the complex numbers $\mathbb{C}$; the p-adic numbers $\mathbb{Q}_p$ for each prime p (defined in Chapter 5); and algebraic number fields such as the Gaussian field $\mathbb{Q}(i)$.

5.2 $\mathcal{E}(K)$ as a Topological Group

Each of the fields $K = \mathbb{R},\ \mathbb{C},\ \mathbb{Q}_p$ carries a natural *topology*, defined by a *metric*. This defines a topology on the corresponding projective space $\mathbb{P}^2(K)$, which in turn induces a topology on the group $\mathcal{E}(K)$.

In each of these cases, the space $\mathbb{P}^2(K)$ is *compact*. To see that, note that $\mathbb{P}^2(K)$ can be considered as the quotient-set of the sphere $S^2(K)$ under the equivalence E which identifies antipodal points:

$$\mathbb{P}^2(K) \cong S^2(K)/E.$$

It follows that the curve $\mathcal{E}(K)$, as a closed subset of $\mathbb{P}^2(K)$, is also compact.

We see therefore that in each of these 3 cases, $\mathcal{E}(K)$ is a *compact abelian group*.

The structure of compact abelian groups is essentially known. Two theorems — each of remarkable generality and beauty — describe this structure.

Firstly, every locally compact group G (not necessarily abelian) carries an *invariant measure* μ, unique up to a scalar multiple, known as the Haar measure.

In the case of a compact group G we can normalise the Haar measure by specifying that the whole group is to have measure — or volume — 1. Each continuous function $f(g)$ on G then has a well-defined integral

$$\int_G f(g)\, d\mu.$$

The measure is *invariant* in the sense that the functions $f(x)$ and $f(gx)$ have the same integral over G:

$$\int_G f(gx)\, d\mu(x) = \int_G f(x)\, d\mu(x).$$

Secondly, Pontriagin's Duality Theory for locally compact abelian groups associates to each such group A a dual group A^*, whose elements are the *unitary characters* of A, ie the continuous homomorphisms

$$\chi : A \to \mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}.$$

The group A^* carries a natural topology, under which it is locally compact.

There is a natural homomorphism

$$A \to A^{**}.$$

One of the basic results of the theory is that this is an isomorphism.

$$\mathbb{A}^{**} = A.$$

Examples of Pontriagin duality are

$$\mathbb{R}^* = \mathbb{R}, \mathbb{C}^* = \mathbb{C},$$
$$\mathbb{Z}^* = \mathbb{T} = \mathbb{R}/\mathbb{Z}, \mathbb{T}^* = \mathbb{Z},$$
$$\mathbb{Q}_p^* = \mathbb{Q}_p, \mathbb{Z}_p^* = \mathbb{Q}_p/\mathbb{Z}_p.$$

Other results are:

$$A \text{ compact} \iff A^* \text{ discrete},$$
$$A \text{ connected} \iff A^* \text{ torsion-free},$$
$$A \text{ totally-disconnected} \iff A^* \text{ torsion-group}$$

Pontriagin's theory is in effect a generalisaion of Fourier analysis. Fourier integrals correspond to the group $\mathbb{R}$, and Fourier series to the group $\mathbb{T}$.

We shall not assume either of these results. Our case is so trivial — the group $\mathcal{E}(\mathbb{R})$ being 1-dimensional — that to appeal these general theorems would be like taking a sledgehammer to crack a nut. However, they may motivate our method.

5.3 In the Neighbourhood of Infinity

We have seen that the flexes on an elliptic curve are determined by a polynomial equation (in one variable) of degree 9. Since an equation of odd degree always has a real root, an elliptic curve $\mathcal{E}$ over $\mathbb{R}$ always has a flex defined over $\mathbb{R}$. Thus we can take $\mathcal{E}$ in strict standard form

$$\mathcal{E}(\mathbb{R}) : y^2 = x^3 + bx + c \quad (b, c \in \mathbb{R}),$$

with the flex $[0, 1, 0]$ as neutral element.

A topological group G is *homogeneous*, that is, it looks the same at all points. For if $g \in G$, and U is a neighbourhood of the neutral element e then gU is a neighbourhood of g, and the map $x \mapsto gx$ establishes a homeomorphism between U and gU.

For this reason, the structure of a topological group is largely determined by its structure in the neighbourhood of the neutral element — in our case, the point $O = [0, 1, 0]$.

In studying such a neighbourhood, it is convenient to use the coordinates

$$(X, Z) = [X, 1, Z].$$

These are defined on the 'affine patch'

$$A_Y = \{[X, Y, Z] : Y \neq 0\},$$

containing the point O. In effect, our curve is covered by 2 affine patches: the 'usual' one

$$A_Z = \{[X, Y, Z] : Z \neq 0\},$$

on which we can use the coordinates $x = X/Z,\ y = Y/Z$), and A_Y. (To cover $\mathbb{P}^2$ we need a third patch, say

$$A_X = \{[X, Y, Z] : X \neq 0\}.$$

But $\mathcal{E} \subset A_Y \cup A_Z$.)

In (X, Z)-coordinates the curve takes the form (on setting $Y = 1$ in the homogeneous equation)

$$Z = X^3 + bXZ^2 + cZ^3.$$

If X and Z are sufficiently small (in other words, the point (X, Z) is sufficiently close to O) then this equation allows us to express Z recursively as a power series in X, taking $X = Z^3$ as our first approximation, and successively substituting in our equation:

$$\begin{aligned}
Z &= X^3 + bX(X^3 + \cdots)^2 + c(X^3 + \cdots)^3 \\
&= X^3 + bX^7 + \cdots \\
&= X^3 + bX(X^3 + bX^7 + \cdots)^2 + c(X^3 + bX^7 + \cdots)^3 \\
&= X^3 + bX^7 + cX^9 + 2b^2X^{11} + \cdots \\
&= X^3 + bX^7 + cX^9 + 2b^2X^{11} + 5bcX^{13} + \cdots \\
&= \ldots
\end{aligned}$$

Rigorously, this follows from the Implicit Function Theorem, since

$$\frac{\partial F}{\partial X} \neq 0$$

at $(X, Z) = (0, 0)$ (and therefore in a neighbourhood of $(0, 0)$), where

$$F(X, Z) \equiv Z - X^3 - bXZ^2 - cZ^3.$$

The Theorem tells us that Z is expressible as a power-series in X in some region $|X|, |Z| < \delta$.

In particular, this implies that $\mathcal{E}$ is locally homeomorphic to the open interval $(-\delta, \delta)$ in a neighbourhood $U \ni O$. It follows, from the homogeneity of the group $\mathcal{E}$, that $\mathcal{E}$ is a 1-dimensional topological manifold, ie it is locally homeomorphic to an open interval at each point $P \in \mathcal{E}$.

5.4 The Invariant Differential

It is not difficult to see intuitively why there is an invariant measure on a topological group. If we choose a 'standard volume', say a small box B, at the neutral element e, then its transform gB can be taken as a standard volume at $g \in G$. Thus we have a uniform measure of volume throughout G.

In the case of a manifold of dimension n we can implement this idea by taking an infinitesimal volume $dx_1 \cdots dx_n$ as standard. If $g \in G$ lies within the $(x_1, \dots, x_n)$ coordinate-system, say $g = (X_1, \dots, X_n)$, then the transformation $x \mapsto gx$ will define a volume

$$\phi(X_1, \dots, X_n) dx_1 \cdots dx_n$$

at g. The Haar integral of a function f is then given by

$$\int f(x_1, \dots, x_n) d\mu = \int f(x_1, \dots, x_n) \phi(x_1, \dots, x_n) dx_1 \cdots dx_n.$$

That is a crude description of Haar measure, and is not intended to be rigorous. In particular, we cannot in general cover the whole of G with a single coordinate system $x_1, \dots, x_n$; we have to 'stick together' patches with different coordinate-systems. This is no great problem, since we know that a change of coordinates to say

$$X_1(x_1, \dots, x_n), \dots, X_n(x_1, \dots, x_n)$$

requires multiplication by the Jacobian:

$$dx_1 \dots dx_n = \frac{\partial(x_1, \dots, x_n)}{\partial(X_1, \dots, X_n)} dX_1 \dots dX_n.$$

Thus in the new coordinates our invariant measure becomes

$$\Phi(X_1, \dots, X_n) \frac{\partial(x_1, \dots, x_n)}{\partial(X_1, \dots, X_n)} dX_1 \dots dX_n,$$

where

$$\Phi\left(X_1(x_1,\ldots,x_n),\ldots,X_n(x_1,\ldots,x_n)\right) = \phi(x_1,\ldots,x_1).$$

It is not difficult to make this rigorous in our 1-dimensional case. We can take X as our single coordinate in the neighbourhood of O, with corresponding differental dX. The coordinate Z as we have seen is expressible as a power-series

$$Z = Z(X) = X^3 + bX^7 + \cdots$$

for X in some interval $I = [-C, C]$. Let

$$U = \{[X, Z] \in \mathcal{E} : X \in I\}$$

be the corresponding neighbourhood of O. Each point $P \in U$ is uniquely determined by its X-coordinate, so we may write $P = P(X)$.

If $X_1, X_2 \in I$ are sufficiently small then $P(X_1) + P(X_2) \in U$, say

$$P(X_1) + P(X_2) = P(S(X_1, X_2)).$$

In other words,

$$P(X_1) + P(X_2) = \left(S(X_1, X_2), T(X_1, X_2)\right),$$

We can compute $S(X_1, X_2)$ by our usual technique. Let the line joining $P(X_1)$ and $P(X_2)$ be

$$Z = MX + D.$$

Then

$$M = \frac{Z_1 - Z_2}{X_1 - X_2}, \quad D = \frac{X_1Z_2 - X_2Z_1}{X_1 - X_2},$$

where $Z_1 = Z(X_1), Z_2 = Z(X_2)$. Suppose this line meets the curve $\mathcal{E}$ again at $P_3 = (X_3, Z_3)$. Then X_1, X_2, X_3 are the roots of

$$MX + D = X^3 + bX(MX + D)^2 + c(MX + D)^3.$$

It follows that

$$\begin{aligned} X_1 + X_2 + X_3 &= -\frac{\text{coeff of } x^2}{\text{coeff of } x^3} \\ &= -\frac{2bMD + 3CM^2D}{1 + bM^2 + cM^3}. \end{aligned}$$

Since $-(X, Z) = (-X, -Z)$,

$$\begin{aligned} S(X_1, X_2) &= -X_3 \\ &= X_1 + X_2 + \frac{2b + 3CM}{1 + aM + bM^2 + cM^3}\, DM. \end{aligned}$$

Let us leave these formulae aside for the moment. According to our argument above, if we integrate the invariant differential $d\theta$ we obtain an invariant or *normal coordinate* θ on I with the property that addition of points is defined by addition of their θ-coordinates. In other words, the function $\theta(X)$ satisfies the condition

$$\theta(X_1) + \theta(X_2) = \theta(S(X_1, X_2)).$$

On differentiating this with respect to X_2,

$$\frac{d\theta}{dX}(X_2) = \frac{d\theta}{dX}(S(X_1, X_2))\frac{\partial S}{\partial X_2}(X_1, X_2).$$

In particular, at the point $(X_1, X_2) = (X, 0)$,

$$\frac{d\theta}{dX}(0) = \frac{d\theta}{dX}(S(X, 0))\frac{\partial S}{\partial X_2}(X, 0).$$

But $S(X, 0) = X$ since $P(X) + P(0) = P(X) + 0 = P(X)$. Thus

$$\frac{d\theta}{dX}(0) = \frac{d\theta}{dX}(X)\frac{\partial S}{\partial X_2}(X, 0).$$

We may assume that $d\theta/dX(0) = 1$, since $d\theta$ is only defined up to a scalar multiple. (In theory we could normalise $d\theta$ by specifying that its integral around the whole curve should be 1:

$$\int_{\mathcal{E}} d\theta = 1.$$

But in practice there is little merit in this.) Thus

$$\frac{d\theta}{dX} = \frac{1}{\partial S/\partial X_2(X, 0)}.$$

The problem is reduced to computation of this partial derivative.

We see from our formula for $S(X_1, X_2)$ that this involves M and D and possibly their derivatives. We have

$$M(X, 0) = \frac{Z - 0}{X - 0} = \frac{Z}{X}, \quad D(X, 0) = \frac{X \cdot 0 - 0 \cdot Z}{X - 0} = 0.$$

It follows from this last result that

$$\frac{\partial S}{\partial X_2}(X, 0) = 1 + \frac{2b + 3CM}{1 + aM + bM^2 + cM^3}\, M\frac{\partial D}{\partial X_2}(X, 0).$$

But

$$\frac{\partial D}{\partial X_2} = \frac{X_1 \partial Z_2/\partial X_2 - Z_1}{X_1 - X_2} + \frac{X_1 Z_2 - X_2 Z_1}{(X_1 - X_2)^2};$$

and so, since $dZ/dX(0) = 0$ (as $Z = X^3 + \cdots$),

$$\frac{\partial D}{\partial X_2}(X, 0) = -\frac{Z}{X}.$$

Thus

$$\begin{aligned}\frac{\partial S}{\partial X_2}(X,0) &= 1 - \frac{2b+3c(Z/X)}{1+b(Z/X)^2+c(Z/X)^3}\left(-\frac{Z^2}{X^2}\right)\\ &= 1 - \frac{2bX+3cZ}{X^3+bXZ^2+cZ^3}\,Z^2\\ &= 1-2bXZ-3cZ^2,\end{aligned}$$

since

$$X^3+bXZ^2+cZ^3 = Z.$$

If we set

$$F(X,Z) \equiv Z - X^3 - bXZ^2 - cZ^3,$$

so that the curve has equation $F(X,Z)=0$, then we can write this as

$$\frac{\partial S}{\partial X_2}(X,0) = \frac{\partial F}{\partial Z}.$$

We conclude that the invariant differential is

$$d\theta = \Phi(X)dX = \frac{dX}{\partial F/\partial Z}.$$

On the curve $\mathcal{E}$ we have $F(X,Z)=0$, and so

$$\frac{\partial F}{\partial X}\,dX + \frac{\partial F}{\partial Z}\,dZ = 0.$$

Thus we have an alternative form for the differential:

$$d\theta = \Phi(X)dX = \frac{dX}{\partial F/\partial Z} = -\frac{dZ}{\partial F/\partial X};$$

or if preferred,

$$\frac{d\theta}{dX} = \frac{1}{\partial F/\partial Z},\ \frac{d\theta}{dZ} = -\frac{1}{\partial F/\partial X}.$$

The differential $d\theta$ is defined on the whole group, and so must be expressible in terms of dx and dy on the 'finite' (x,y)-patch A_Z. Since

$$(x,y) = [x,y,1] = [x/y, 1/y, 1] = (X,Z),$$

the coordinate transformation between the patches is given by

$$X = \frac{x}{y},\ Z = \frac{1}{y},$$

with the inverse transformation

$$x = \frac{X}{Z},\ y = \frac{1}{Z}.$$

Thus

$$\begin{aligned}
dx &= \frac{dX}{Z} - \frac{XdZ}{Z^2} \\
&= \frac{1}{Z^2}(ZdX - XdZ) \\
&= \frac{1}{Z^2}\left(Z\frac{dX}{d\theta} - X\frac{dZ}{d\theta}\right)d\theta \\
&= \frac{1}{Z^2}\left(Z\frac{\partial F}{\partial Z} + X\frac{\partial F}{\partial X}\right)d\theta \\
&= \frac{1}{Z^2}\left(Z - 3X^3 - 3bXZ^2 - 3CZ^3\right)d\theta \\
&= -\frac{2}{Z}d\theta \\
&= -2y d\theta \\
&= -\frac{\partial f}{\partial y}d\theta,
\end{aligned}$$

ie

$$\frac{d\theta}{dx} = -\frac{1}{\partial f/\partial y},$$

where

$$f(x, y) \equiv y^2 - x^3 - bx - c = 0$$

is the equation of the curve in (x, y)-coordinates.

As before,

$$\frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy = 0,$$

Thus

$$d\theta = -\frac{dx}{\partial f/\partial y} = \frac{dy}{\partial f/\partial x}.$$

5.5 No Miracles in Maths

These formulae for the invariant differential $d\theta$ are remarkable both for their simplicity and for the similarity between the formulae on the 2 patches. The reason is as follows — where we emphasize that our argument is not intended to be rigorous (and is more appropriate to the complex case, in any case).

An elliptic curve is a curve of genus 1. The genus g of a non-singular curve can be defined in various ways; but one definition is that g is *the dimension of the space of holomorphic differentials on the curve*, that is, differentials which are everywhere expressible in terms of a local coordinate u in the form $\Phi(u)du$ where $\Phi(u)$ can be written as a power-series in u.

Accordingly, there is just one such differential $d\theta$ on an elliptic curve $\mathcal{E}$, up to a scalar multiple — which is, of course, the differential defining the Haar measure.

Suppose (u, v) are the coordinates in an affine patch. Let the equation of the curve in these coordinates by $f(u, v) = 0$. At any point P we must have either $\partial f/\partial u \neq 0$ or $\partial f/\partial v \neq 0$. (Otherwise P would be a singular point.) Suppose $\partial f/\partial v \neq 0$. Then by the Implicit Function Theorem, we can express v as a function $v(u)$ of u, and so we can take u as local coordinate. Thus the differential

$$\frac{du}{\partial f/\partial v}$$

is holomorphic in the neighbourhood of this point. Similarly, if $\partial f/\partial u \neq 0$ then the differential

$$-\frac{dv}{\partial f/\partial u}$$

is holomorphic near P. Since the two differentials are equal wherever they are both defined, together they define a differential which is holomorphic everywhere in the patch.

The simplest way to see that the differentials defined in this way on the 3 affine patches $A_X, A_Y, A_Z \subset \mathbb{P}^2$ 'fit together' is to pass to homogeneous coordinates, with the whole curve defined by

$$H(X, Y, Z) = 0.$$

Let

$$\pi : \mathbb{R}^3 \setminus \{0\} \to \mathbb{P}^2$$

be the natural surjection

$$(X, Y, Z) \mapsto [X, Y, Z].$$

Each function u on an open subset of $\mathbb{P}^2$ defines a function $\pi^* u$ on the corresponding open subset of $\mathbb{R}^3$. For example, the functions x and y on $A_Z \subset \mathbb{P}^2$ give the functions

$$\pi^* x = \frac{X}{Z},\ \pi^* y = \frac{Y}{Z}.$$

Similarly each differential ω defined on $\mathbb{P}^2$ or on an open subset of $\mathbb{P}^2$ gives a differential $\pi^*\omega$ on the corresponding open subset of $\mathbb{R}^3$. For example, the differentials dx and dy on $\mathbb{A}_Z$ give the differentials

$$\pi^*(dx) = \frac{Z\,dX - X\,dZ}{Z^2},\ \pi^*(dy) = \frac{Z\,dY - Y\,dZ}{Z^2}.$$

on the subset $z \neq 0$ of $\mathbb{R}^3$.

The differentials

$$u(X, Y, Z)dX + v(X, Y, Z)dY + w(X, Y, Z)dZ$$

induced in this way all satisfy

$$Xu + Yv + Zw = 0.$$

(This arises from the fact that the functions we get on $\mathbb{R}^3$ are all homogeneous of degree 0; and if $u(X, Y, Z)$ is homogeneous of degree d then

$$X\frac{\partial u}{\partial X} + Y\frac{\partial u}{\partial Y} + Z\frac{\partial u}{\partial Z} = du,$$

as we noted earlier.)

The subspace of differentials on $\mathbb{R}^3 \setminus \{0\}$ satisfying this condition is in one-one correspondence with the differentials on $\mathbb{P}^2$, allowing us to identify the two.

We are actually interested in differentials on $\mathcal{E}$, not on $\mathbb{P}^2$. Two differentials on $A_Z \subset \mathbb{P}^2$ define the same differential on $\mathcal{E}$ if they differ by a multiple of

$$\frac{\partial f}{\partial X}dX + \frac{\partial f}{\partial Y}dY.$$

It follows that 2 differentials on $\mathbb{R}^3 \setminus \{0\}$ define the same differential on $\mathcal{E}$ if they differ by a multiple of

$$\frac{\partial H}{\partial x}dx + \frac{\partial H}{\partial y}dy + \frac{\partial H}{\partial z}dz.$$

Now we can describe the invariant differential in global terms on $\mathbb{R}^3$. It is given by

$$d\Theta = \frac{X\ dY - Y\ dX}{\partial H/\partial Z} = \frac{Y\ dZ - Z\ dY}{\partial H/\partial X} = \frac{Z\ dX - X\ dZ}{\partial H/\partial Y}.$$

We see now why the formulae on the 3 patches A_X, A_Y, A_Z are so similar.

5.6 The Functional Equation for θ

It remains to verify that

$$\theta(P_1) + \theta(P_2) = \theta(P_1 + P_2)$$

for P_1, P_2 sufficiently close to O, ie that

$$\theta(X_1) + \theta(X_2) = \theta\left(S(X_1, X_2)\right)$$

for X_1, X_2 sufficiently small.

If we regard this as an equation in X_2 then it holds for $X_2 = 0$. Hence it is sufficient to show that the derivative of the equation with respect to X_2 holds, ie

$$\frac{d\theta}{dX}\left(S(X_1, X_2)\right)\frac{\partial S}{\partial X_2}(X_1, X_2) = \frac{d\theta}{dX}(X_2).$$

From the definition of θ,

$$\frac{d\theta}{dX}(X)\frac{\partial S}{\partial X_2}(X, 0) = \frac{d\theta}{dX}(0) = 1.$$

Thus we have to show that

$$\frac{\partial S/\partial X_2(X_1, X_2)}{\partial S/\partial X_2(S(X_1, X_2), 0)} = \frac{1}{\partial S/\partial X_2(X_2, 0)},$$

that is,

$$\frac{\partial S}{\partial X_2}(X_1, X_2)\frac{\partial S}{\partial X_2}(X_2, 0) = \frac{\partial S}{\partial X_2}\left(S(X_1, X_2), 0\right).$$

But by the associative law,

$$S\left(X_1, S(X_2, X_3)\right) = S\left(S(X_1, X_2)X_3\right).$$

Differentiating this with respect to X_3 and setting $X_3 = 0$,

$$\frac{\partial S}{\partial X_2}(X_1, X_2)\frac{\partial S}{\partial X_2}(X_2, 0) = \frac{\partial S}{\partial X_2}\left(S(X_1, X_2), 0\right),$$

which is just the result required. We conclude that

$$\theta(X_1) + \theta(X_2) = \theta\left(S(X_1, X_2)\right).$$

We note that this result must be an identity in X_1 and X_2, which will therefore hold in *every* field F, for example in the p-adic field $\mathbb{Q}_p$.

5.7 The Components of $\mathcal{E}(\mathbb{R})$

The connected component of the neutral element e in a topological group G is a closed subgroup of G, which is generally denoted by G_0.

Proposition 5.1 *The group $\mathcal{E}(\mathbb{R})$ has either 1 or 2 connected components.*

Proof ▶ Suppose C is a component other than $\mathcal{E}(\mathbb{R})_0$. Then C does not meet the line at infinity, since O is the only point of $\mathcal{E}$ on this line. Hence C is a compact and therefore bounded set in the affine (x, y)-patch A_Z.

Thus the function x must attain its upper and lower bounds on C. If the curve has equation

$$y^2 = x^3 + bx + c,$$

then

$$2y\frac{dy}{dx} = 3x^2 + b.$$

Thus $y = 0$ where x attains its bounds, and so

$$f(x) = 0$$

at these points. There are at most 3 such points. Since each component apart from $\mathcal{E}_0$ contributes 2 points, we conclude that there are at most 2 components. ◀

Corollary 1 *Either*

$$\mathcal{E}(\mathbb{R}) = \mathcal{E}(\mathbb{R})_0 \text{ or } \mathcal{E}(\mathbb{R}) = \mathcal{E}(\mathbb{R})_0 \oplus \mathbb{Z}/(2).$$

In the first case, $\mathcal{E}(\mathbb{R})$ has just one point of order 2. In the second case, $\mathcal{E}(\mathbb{R})$ has three points of order 2, which together with O form a subgroup $\cong \mathbb{Z}/(2) \oplus \mathbb{Z}/(2)$.

Proof ▶ Recall that the point $P = (x, y)$ on the elliptic curve

$$\mathcal{E}(\mathbb{R}) : y^2 = f(x)$$

is of order 2 if and only if $y = 0$.

Suppose $\mathcal{E}(R)$ has two components. Then the second component contains at least two points of order 2, as we saw in the proof above. Let $A = (\alpha, 0)$ be one of these points.

In general, if G_0 is the connected component of the neutral element in a topological group G then each coset $G_0 g$ is a connected component of G.

Thus in our case $\mathcal{E}(\mathbb{R})_0 + A$ must be the second component of $\mathcal{E}(\mathbb{R})$. It follows that

$$\begin{aligned}\mathcal{E}(\mathbb{R}) &= \mathcal{E}(\mathbb{R})_0 \cup (\mathcal{E}(\mathbb{R})_0 + A) \\ &= \mathcal{E}(\mathbb{R})_0 \oplus \{O, A\} \\ &\cong \mathcal{E}(\mathbb{R})_0 \oplus \mathbb{Z}/(2),\end{aligned}$$

since $\{O, A\}$ is a subgroup $\cong \mathbb{Z}/(2)$. ◀

Corollary 2 *If $\mathcal{E}(\mathbb{R})$ has one component then it has just one element of order 2. If it has two components then it has three elements of order 2, which together with O form a subgroup $\cong \mathbb{Z}/(2) \oplus \mathbb{Z}/(2)$.*

Proof ▶ Suppose $\mathcal{E}(\mathbb{R})$ has two components. Then $f(x)$ has at least 2 real roots, as we saw. But a polynomial of degree 3 over $\mathbb{R}$ has either 1 or 3 real roots. Hence $f(x)$ has 3 real roots $\alpha, \beta, gamma$, giving 3 points of order 2, namely

$$A = (\alpha, 0),\ B = (\beta, 0),\ C = (\gamma, 0).$$

Evidently $\{O, A, B, C\}$ is a subgroup, $\cong \mathbb{Z}/(2) \oplus \mathbb{Z}/(2)$.

Conversely, suppose $f(x)$ has 3 real roots $\alpha, \beta, gamma$, where $\alpha < \beta < \gamma$. Then

$$y^2 \geq 0 \Longrightarrow \alpha \leq x \leq \beta \text{ or } \gamma \leq x.$$

Let

$$M = \max_{\alpha \leq x \leq \beta} f(x).$$

Then

$$\alpha \leq x \leq \beta \Longrightarrow (x, y) \in [\alpha, \beta] \times [-M^{1/2}, M^{1/2}].$$

Evidently the points of $\mathcal{E}(\mathbb{R})$ in this rectangle form a component or components distinct from $\mathcal{E}(\mathbb{R})_0$; and in particular $\mathcal{E}(\mathbb{R})$ has more than one component. ◀

5.8 The connected component $\mathcal{E}(\mathbb{R})_0$

Proposition 5.2 *The connected component of $\mathcal{E}(\mathbb{R})$ is isomorphic to the torus:*

$$\mathcal{E}(\mathbb{R})_0 \cong \mathbb{T}.$$

Proof ► We have seen that θ defines a *local isomorphism* of $\mathbb{R}$ into $\mathcal{E}(\mathbb{R})$; that is, θ is defined on an open interval $I \subset \mathbb{R}$ containing 0, and there satisfies

$$\theta(x + y) = \theta(x) + \theta(y) \quad (x, y, x + y \in I).$$

Lemma 1 *A local homomorphism θ of $\mathbb{R}$ into a topological group G extends uniquely to a homomorphism*

$$\theta : \mathbb{R} \to G.$$

Proof of Lemma ▷ Suppose $x \in R$. Then $x/n \in I$ for some integer n. If θ can be extended to the whole of $\mathbb{R}$ then

$$\theta(x) = n\theta(x/n).$$

But is this unique? Suppose $x/m \in I$. Then

$$\frac{x}{m}, \frac{x}{n}, \frac{x}{mn} \in I.$$

Hence, since θ is a local homomorphism,

$$\theta(x/n) = m\theta(x/mn),\ \theta(x/m) = n\theta(x/mn).$$

It follows that

$$n\theta(x/n) = m\theta(x/m) = mn\theta(x/mn).$$

Thus $\theta(x)$ is well-defined by the relation

$$\theta(x) = n\theta(x/n);$$

the definition is independent of n.

It is straightforward to verify that θ is a homomorphism; and its continuity follows from its continuity at 0. ◁

Remark: The last Lemma is a particular case of the general result that a local homomorphism of a *simply-connected* topological group G and a topological group H always be extended to a true homomorphism. We are applying this with $\mathbb{R}, G$ in place of G, H.

Lemma 2 *The homomorphism*

$$\theta : \mathbb{R} \to \mathcal{E}(\mathbb{R})_0$$

is surjective.

Proof of Lemma ▷ We know that $\operatorname{im}\theta$ includes an open interval I around O. It follows that $\operatorname{im}\theta$ is open; for

$$P \in \operatorname{im}\theta \Longrightarrow P + U \in \operatorname{im}\theta.$$

Thus $\operatorname{im}\theta$ is an open subgroup, and is therefore also closed. (For each coset is open; hence the subgroup, as the complement of the union of all the cosets except itself, is closed.) Since $\mathcal{E}(\mathbb{R})_0$ is connected, it follows that

$$\operatorname{im}\theta = \mathcal{E}(\mathbb{R})_0.$$

◁

Lemma 3 *The subgroup* $\ker\theta$ *is discrete.*

Proof of Lemma ▷ Since θ is a local isomorphism, it is a homeomorphism on an open interval $I \ni 0$. It follows that

$$\ker\theta \cap I = \{0\}.$$

Hence $\ker\theta$ is discrete. ◁

Lemma 4 *A discrete subgroup $S \subset \mathbb{R}$ is necessarily of the form*

$$S = \mathbb{Z}x = \{nx : n \in \mathbb{Z}\}$$

for some $x \in \mathbb{R}$.

Proof of Lemma ▷ If $S \neq 0$ then there are points $s \in S$, $s > 0$. Let x be the lower bound of these numbers. Since S is discrete, $x > 0$; and since a discrete subgroup is necessarily closed, $x \in S$.

Now suppose $s \in S$. If $n = [s/x]$ then

$$s = nx + r$$

where $0 \leq r < x$. Since

$$r = s - nx \in S$$

it follows that $r = 0$; for otherwise the minimality of x would be contradicted. Thus

$$s = nx.$$

◁

We conclude that

$$\mathcal{E}(\mathbb{R})_0 \cong \mathbb{R}/\ker\theta \cong \mathbb{R}/\mathbb{Z} = \mathbb{T}.$$

◀

5.9 The Structure of $\mathcal{E}(\mathbb{R})$

Theorem 5.1 *For each elliptic curve, either*

$$\mathcal{E}(\mathbb{R}) = \mathbb{T} \text{ or } \mathbb{T} \oplus \mathbb{Z}/(2).$$

Proof ► This follows at once from the Corollary to Proposition 5.1 and Proposition 5.2. ◄

Proposition 5.3 *The elliptic curve*

$$\mathcal{E}(\mathbb{R}) : y^2 = x^3 + ax^2 + bx + c$$

has one component or two according as

$$D(f) < 0 \text{ or } D(f) > 0.$$

Proof ► By Corollary 2 to Proposition 5.2, $\mathcal{E}(\mathbb{R})$ has one or two components according as $f(x)$ has 1 or 3 real roots.

Recall that the discriminant of a polynomial $f(x)$ with roots α_i is defined to be

$$D(f) = \prod(\alpha_i - \alpha_j)^2.$$

Suppose $f(x)$ has 3 real roots α, β, γ. Then

$$D(f) = [(\alpha - \beta)(\beta - \gamma)(\gamma - \alpha)]^2 > 0.$$

On the other hand, suppose $f(x)$ has 1 real root α, and complex conjugate roots $\beta \pm i\gamma$. Then

$$\begin{aligned} D(f) &= [(\alpha - \beta + i\gamma)(\alpha - \beta - i\gamma)(2i\gamma)]^2 \\ &= -4\left[(\alpha - \beta)^2 + \gamma^2)\gamma\right]^2 \\ &< 0. \end{aligned}$$

◄

5.10 Postscript: an Elementary Approach

If we only want to determine the structure of $\mathcal{E}(\mathbb{R})$, and do not want an explicit formula for $\theta = \theta(X)$, we can argue as follows.

We know that each connected component of $\mathcal{E}(\mathbb{R})$ is closed, from which it follows, as we have seen, that there are at most 2 components. So it sufficient to show that the connected component $\mathcal{E}(\mathbb{R})_0$ of the zero element — which is a subgroup of $\mathcal{E}(\mathbb{R})$ — is isomorphic to $\mathbb{T}$.

We know, from its structure in the neighbourhood of O that $\mathcal{E}(\mathbb{R})$ is *a 1-dimensional topological manifold*, by which we simply mean that it is locally isomorphic to an interval.

Proposition 5.4 *A connected abelian topological group A which is a topological 1-manifold is necessarily isomorphic to $\mathbb{T}$.*

Proof ► By hypothesis there is an open neighbourhood I of O isomorphic to $(-1, 1)$ (with the number 0 corresponding to the zero element O). By continuity we can find an interval $J \subset I$ such that

$$P, Q \in J \Longrightarrow P + Q \in I, \; -P \in I.$$

There is a natural *order* on the interval I (this is a characteristic of dimension 1), which we may denote by $P \prec Q$. If $P \prec Q \prec R$ we may say that Q *lies between* P and R. This is reflected in the fact that there exists a 'one-one path' from P to R in I (ie an injective continuous map $\pi : [0, 1] \to I$ with $\pi(0) = P, \; \pi(1) = R$) passing through Q.

Lemma 1 *Suppose $P, Q, R \in J$. Then*

$$Q \prec R \Longrightarrow P + Q \prec P + R.$$

Proof of Lemma ▷ This holds for $P = O$. It follows by continuity that it holds for all $P \in J$. ◁

Lemma 2 *Suppose $P, Q \in J$. Then*

$$P \prec Q \Longrightarrow -Q \prec -P.$$

Proof of Lemma ▷ By the previous Lemma,

$$P \prec Q \Longleftrightarrow 0 \prec Q - P.$$

Thus it is sufficient to prove the result with $P = O$, ie to show that

$$O \prec Q \Longrightarrow -Q \prec O.$$

Suppose not; then we can find a point $P \in J$ such that

$$0 \prec -P \prec P.$$

But on adding P this implies that $P \prec O$. ◁

Lemma 3 *Suppose $P \in J$. Then there exists a unique point in J, which we may denote by $\frac{1}{2}P$, such that*

$$\frac{1}{2}P + \frac{1}{2}P = P.$$

Proof of Lemma ▷ Suppose $0 \prec P$. Then by the first Lemma, $P \prec P + P$. On the other hand, $O + O \prec P$. Thus there is a point $Q \in [0, P]$ such that $Q + Q = P$.

This point is unique. For suppose $Q_1, Q_2 \in J$ and

$$2Q_1 = 2Q_2.$$

We may suppose that $Q_1 \prec Q_2$. But then

$$O \prec Q_2 - Q_1 \Longrightarrow O \prec 2(Q_1 - Q_2) = O.$$

If $P \prec O$ then $O \prec -P$ and

$$\frac{1}{2}P = -\frac{1}{2}(-P).$$

◁

By repeating this construction, we can define points

$$\frac{1}{2^n}P$$

for each $n > 0$.

Lemma 4 *As $n \to \infty$,*

$$\frac{1}{2^n}P \to O.$$

Proof of Lemma ▷ If not then (assuming $O \prec P$) we can find a point Q such that $O \prec Q$ and

$$2^n Q \prec P$$

for all n. It follows that the sequence is convergent, say

$$2^n Q \to R.$$

But then

$$2R = R,$$

which is impossible, since $O \prec R$. ◁

Now we can define λP for

$$\lambda = \frac{m}{2^n} \quad (0 \le m \le 2^n);$$

and it is a straightforward matter to verify that if $O \prec P$ then

$$\lambda < \mu \Longrightarrow \lambda P \prec \mu P.$$

But now we can define λP for $\lambda \in [-1, 1]$, by continuity; and we have a *local isomorphism* $\mathbb{R} \to A$, ie a map

$$\theta : [-1, 1] \to A$$

such that if $\lambda, \mu \in [-1, 1]$ then

$$\theta(-\lambda) = -\theta(\lambda),\ \theta(\lambda + \mu) = -\theta(\lambda) + \theta(\mu).$$

Lemma 5 *For any topological group G, every local homomorphism $\theta : \mathbb{R} \to G$ extends to a true homomorphism*

$$\theta : \mathbb{R} \to G.$$

Proof of Lemma ▷ This is straightforward. Suppose the local homomorphism is defined on the interval I around O. Given any real number λ we can find an integer $n > 0$ such that

$$\frac{\lambda}{n} \in I.$$

We set

$$\theta(\lambda) = n\theta\left(\frac{\lambda}{n}\right).$$

It is a straightforward matter to verify that this definition is independent of the integer n chosen. ◁

We have shown therefore that we have a homomorphism

$$\theta : \mathbb{R} \to A.$$

Moreover since θ is a local isomorphism it follows that $\ker\theta$ is discrete. But it is easy to see that a discrete subgroup $S \subset \mathbb{R}$ is generated by the least positive number μ in S (unless $S = \{0\}$):

$$S = \{n\mu : n \in \mathbb{Z}\} \cong \mathbb{Z}.$$

Finally, the homomorphism θ must be surjective. For $\operatorname{im}\theta$ is an open subgroup of A, since it contains an open neighbourhood of O. It is therefore also closed (since all its cosets are open). Since A is by hypothesis connected, this implies that $\operatorname{im}\theta = A$.

We conclude that

$$A \cong \mathbb{R}/\operatorname{im}\theta \cong \mathbb{R}/\mathbb{Z} = \mathbb{T}.$$

◀

Corollary $\mathcal{E}(\mathbb{R}) = \mathbb{T}$ *or* $\mathbb{T} \oplus \mathbb{Z}/(2)$.