2. Examples of Conics

Our primary objects of study in this book are conics. As we will see, the set of rational points on conics with at least one such rational point carries a group structure. The two simplest examples are the parabola and the hyperbola, where the group structure turns out to be the additive group and the unit group of the underlying ring. The unit circle, perhaps the simplest conic from a geometric point of view, has a more interesting group structure.

Before we come to the geometric interpretations of the group laws we will be studying, let us briefly review the necessary abstract nonsense: the most basic notions of group theory.

2.1. Groups

Where we go abstract.

The readers already familiar with the notions of groups and homomorphisms may skip this section; the others are invited to do the same and come back here as soon as they stumble across something they do not know.

Recall that a **group** is a set G endowed with a composition $G \times G \longrightarrow G$ sending $a, b \in G$ to $a \circ b \in G$, such that the following properties are verified:

1. Existence of a unit: there is an element $e \in G$ such that $e \circ a = a \circ e = a$ for all $a \in G$.
2. Existence of an inverse: for every $a \in G$ there is a $b \in G$ such that $a \circ b = 1$.
3. Associativity: we have $a(bc) = (ab)c$ for all $a, b, c \in G$.

The most basic mathematical objects in everday life are the natural numbers 1, 2, 3, ... used for counting. The set $\mathbb{N} = \{1, 2, 3, \ldots\}$ of natural numbers has a composition called addition, but it does not form a group with respect to addition even if we include 0 (which mathematicians tend to do although historically the invention of 0 came long after the concept of numbers had been discovered). The problem is that the numbers 1, 2, 3, ... do not have additive inverses: there are no natural numbers x such that $x + 1 = 0$. The fact that such inverses are useful in everday life for describing objects that you do not possess (debts etc.) made Indian mathematician invent negative numbers. In mathematical terms, they have completed the semigroup $\mathbb{N}$ to the group $\mathbb{Z}$ of integers.

Most groups we shall deal with will be **commutative**: these are groups satisfying $a \circ b = b \circ a$ for all $a, b \in G$. Commutative groups are usually called **abelian**.

In many cases, the composition is either some form of addition or multiplication. If we write G additively ($a \circ b = a + b$), then we denote the neutral element e by 0, and the inverse of a by $-a$. If we write G multiplicatively, then the neutral element is denoted by 1 and the inverse of a by a^{-1}.

The number of elements of G is denoted by $|G|$ or $\#G$; groups with finitely many elements are called **finite groups**. The most important example of an infinite group is the group $\mathbb{Z}$ of integers (observe that $\mathbb{Z}$ is a group with respect to addition, not with respect to multiplication).

Examples of groups are the additive groups of various fields such as the rationals $\mathbb{Q}$, the reals $\mathbb{R}$ or the complex numbers $\mathbb{C}$. The nonzero elements $K^\times$ of a field K form a group with respect to multiplication. Finally, matrix groups such as $\mathrm{SL}_2(\mathbb{Z})$, the group of all 2×2 matrices with integral entries and determinant $+1$, play a fundamental role in number theory.

A group G is called **cyclic** if there is an element $g \in G$ such that every element of G can be written in the form ng (for additively written groups) or g^n (for multiplicatively written groups), where $n \in \mathbb{Z}$. Standard examples of cyclic groups are the additive group of integers $\mathbb{Z}$, and the residue class groups $\mathbb{Z}/n\mathbb{Z}$. In fact it is easy to show that these are the only examples (see Exercise 2.3).

A subset $H \subseteq G$ is called a **subgroup** of G if H is a group with respect to the "same" composition as G, that is, if

1. for all $h_1, h_2 \in H$, we have $h_1 \circ h_2 \in H$;
2. there is an element $1 \in H$ with $1h = h$ for all $h \in H$;
3. for every $h \in H$ there is an element $h' \in H$ such that $h \cdot h' = 1$.

For example, the even integers $2\mathbb{Z}$ form a subgroup in the group $\mathbb{Z}$ of integers.

A **homomorphism** between groups $(G, \circ)$ and $(H, *)$ is a map $f : G \longrightarrow H$ that respects the group laws in the sense that we have $f(g \circ g') = f(g) * f(g')$. An **isomorphism** is a bijective homomorphism. Here are some examples:

1. Multiplication by $a \in \mathbb{Z}$ induces a homomorphism $m_a : \mathbb{Z} \longrightarrow \mathbb{Z}$. In fact we have $m_a(x+y) = a(x+y) = ax + ay = m_a(x) + m_a(y)$.
   Addition of a nonzero integer m, however, does not define a homomorphism: setting $a_m(x) = x+m$ we find $a_m(x+y) = (x+y)+m = x+m+y+m-m = a_m(x)+a_m(y)-m$, and this shows that a_m is not a homomorphism unless $m = 0$, when a_m is the trivial homomorphism sending every element to itself.
2. The inclusion map $\iota : \mathbb{Q} \longrightarrow \mathbb{R}$ is a homomorphism because $\iota(r+s) = \iota(r)+\iota(s)$ (if you already know rings: ι is even a ring homomorphism, i.e., it also respects multiplication: $\iota(rs) = \iota(r)\iota(s)$). This means that it does not matter whether we add (resp. multiply) two rational numbers in $\mathbb{Q}$ or as elements of $\mathbb{R}$.
   Informally, this inclusion allows us to regard $\mathbb{Q}$ as a subset of $\mathbb{R}$ in such a way that the group law on $\mathbb{Q}$ is simply the restriction of the group law on $\mathbb{R}$ to $\mathbb{Q}$, in other words: $\mathbb{Q}$ is a subgroup (and even a subring and a subfield) of $\mathbb{R}$.
3. The exponential function is a homomorphism $\exp : (\mathbb{R}, +) \longrightarrow (\mathbb{R}_{>0}, \cdot)$ because
   $$\exp(a+b) = \exp(a)\exp(b).$$
4. The logarithm is a homomorphism $\log : (\mathbb{R}_{>0}, \cdot) \longrightarrow (\mathbb{R}, +)$ because
   $$\log ab = \log a + \log b.$$
   Note that exp and log are inverse maps of each other.
   Historically, the logarithm owes its existence to the fact that it is an isomorphism between the multiplicative group of positive reals and the additive group of all reals: logarithms were invented to reduce the problem of multiplying numbers to that of adding them. The slide rule is just a mechanical incarnation of the isomorphisms log and exp between $\mathbb{R}^\times$ and $\mathbb{R}_{>0}$.
5. The set C^∞ of all infinitely often differentiable functions $(0,1) \longrightarrow \mathbb{R}$ is an additive group, and $\frac{d}{dx} : C^\infty \longrightarrow C^\infty$ is a homomorphism (actually, C^∞ is an $\mathbb{R}$-vector space, and map is linear) because $(f+g)' = f' + g'$.

6. If $f : V \longrightarrow W$ is a linear map between K-vector spaces V and W, then f is also a homomorphism between the additive groups $(V, +)$ and $(W, +)$.

Observe that if $f : G \longrightarrow H$ is a homomorphism between additively written groups, then $f(0) = 0$ and $f(-g) = -f(g)$. This follows easily from the axioms.

A group homomorphism $f : A \longrightarrow A$ from a group A to itself is called an **endomorphism**, and an **automorphism** if f is bijective.

The **kernel** of a homomorphism $f : G \longrightarrow H$ is by definition

$$\ker f = \{g \in G : f(g) = 0\},$$

where 0 is the neutral element of H. The kernel of a homomorphism $G \longrightarrow H$ is a subgroup of G containing the neutral element 0 of G, and f is injective if and only if $\ker f = \{0\}$.

Fact 2.1. *If A is a finite abelian group with order n, and if we define an endomorphism $f : A \longrightarrow A$ by $f(a) = a^m$ for some integer m, then f is an automorphism whenever* $\gcd(m, n) = 1$.

In fact we have $\ker f = \{a \in A : a^m = 1\}$. Since $\gcd(m, n) = 1$, there exist, by Bezout, integers r, s with $mr + ns = 1$. Since $a^n = 1$ for all $a \in A$ we now have $1 = a^m = a^{rm} = a^{rm+sn} = a$ for all $a \in \ker f$, and this shows that $\ker f = 1$.

As an application, consider the unit group of the residue class group $G = (\mathbb{Z}/p\mathbb{Z})^\times$ for prime numbers p. This group has $p - 1$ elements, so if $p \equiv 2 \bmod 3$, its order is coprime to 3, and the map $f([a]) = [a]^3$ sending each residue class $[a]$ to its cube is an automorphism. In particular, if $p \equiv 2 \bmod 3$ is prime, each residue class modulo p is a cube.

Above we have attached a subgroup of G to any homomorphism $f : G \longrightarrow H$. Dually, we can define the **image** of a homomorphism $f : G \longrightarrow H$ as the subgroup (!)

$$\operatorname{im} f = \{f(g) \in H : g \in G\}$$

of H. The homomorphism $f : G \longrightarrow H$ is surjective if and only if $\operatorname{im} f = H$. Given a homomorphism $f : G \longrightarrow H$ and a subgroup $B \subseteq A$, the **preimage** of B under f is the subgroup $A \subseteq G$ defined by $A = f^{-1}(B) = \{g \in G : f(g) \in B\}$.

If $(G, \circ)$ and $(H, *)$ are groups, then the cartesian product $G \times H$ can be given a group structure by defining $(g, h)(g', h') = (g \circ g', h * h')$. Checking the axioms is straightforward. Moreover $G \times H$ is abelian if and only if G and H are.

In the next chapter we will study residue class rings $\mathbb{Z}/m\mathbb{Z}$; these are examples of factor groups, which can be defined in full generality as follows: let H be a subgroup of the additively written abelian group G. A coset $g + H$ consists of all elements $g + h$ with $h \in H$.

Lemma 2.2. *Let H be a subgroup of a group G. Then the cosets $g + H$, where $g \in G$, are either equal or disjoint.*

Proof. Assume that $g_1 + H$ and $g_2 + H$ have an element g in common. Then g can be written in the form $g = g_1 + h_1$ and $g = g_2 + h_2$ for elements $h_1, h_2 \in H$. Thus $g_1 - g_2 = h_2 - h_1 \in H$, say $g_1 - g_2 = h$. But then $g_1 + H = g_2 + h + H = g_2 + H$, and the cosets are equal. □

Thus G can be written as a disjoint union of cosets: $G = \bigcup (g_j + H)$. If there are only finitely many cosets $g_1 + H$, $g_2 + H$, …, $g_r + H$, then we call r the **index** of H in G and write $r = (G : H)$.

Proposition 2.3. *If H is a subgroup of a finite abelian group G, then $\#H \mid \#G$: the order of a subgroup divides the order of the group. More exactly we have $\#G = (G : H) \cdot \#H$.*

Proof. Observe that G can be written as the disjoint union of the cosets $g_1 + H$, $\ldots$, $g_r + H$, where $r = (G : H)$. Since each of these cosets has $\#H$ elements, we find $\#G = r \cdot \#H = (G : H) \cdot \#H$. □

The cosets $g_j + H$ can be made into a group by defining

$$(g_i + H) + (g_j + H) = g_k + H,$$

where $g_k + H$ is the coset containing the element $g_i + g_j$. The group of these cosets is called the **factor group**[1] of G by H and is denoted by G/H. The following result follows directly from the definition of the factor group:

Proposition 2.4. *If H is a subgroup of an abelian group G such that the index $(G : H)$ is finite, then the order of the factor group G/H is equal to $(G : H)$.*

If $f : G \longrightarrow G'$ is a homomorphism between abelian groups, then $H = \ker f$ is a subgroup of G and $\operatorname{im} f = \{f(g) : g \in G\}$ a subgroup of G'. One of the fundamental results in group theory is

Theorem 2.5. *If $f : G \longrightarrow G'$ is a group homomorphism, then*

$$G/\ker f \simeq \operatorname{im} f. \tag{2.1}$$

For proving this isomorphism we have to define a map $\phi : G/\ker f \longrightarrow \operatorname{im} f$ and show that it is an isomorphism. We set $\phi(g + \ker f) = f(g)$ and have to make sure that ϕ is well defined; this means that $\phi(g)$ must not depend on the choice of the element g representing the coset. In fact, if we choose $g + h$ for some $h \in \ker f$ as a representative, then $g + \ker f = (g + h) + \ker f$, and $\phi(g + h + \ker f) = f(g + h) = f(g) + f(h) = f(g) = \phi(g + \ker f)$ since f is a homomorphism and $f(h) = 0$. Next, ϕ is a homomorphism since $\phi(g + g' + \ker f) = f(g + g') = f(g) + f(g') = \phi(g + \ker f) + \phi(g' + \ker f)$.

Finally we have to show that ϕ is bijective. Clearly ϕ is surjective since every element of $\operatorname{im} f$ has the form $f(g)$ for some $g \in G$, and then $f(g) = \phi(g + \ker f)$ is in the image of ϕ. Next ϕ is injective: if $\phi(g + \ker f) = 0$, then $f(g) = 0$, hence $g \in \ker f$ and thus $g + \ker f = 0 + \ker f$. Thus $\ker \phi = \{0 + \ker f\}$ consists of the neutral element of $G/\ker f$, and this shows that ϕ is injective.

Equation (2.1) may easily be generalized from groups to rings:

Theorem 2.6. *Let $f : R \longrightarrow S$ be a ring homomorphism. Then we can give the quotient group $R/\ker f$ a ring structure such that $R/\ker f \simeq \operatorname{im} f$ becomes an isomorphism of rings.*

Proof. Recall that $\ker f = \{a \in R : f(a) = 0\}$. It is easy to see that $\ker f$ is a subring of R. We claim that we can give the set $R/\ker f = \{r + \ker f : r \in R\}$ a ring structure. Of course we define $(r + \ker f) \pm (s + \ker f) = (r \pm s) + \ker f$ and $(r + \ker f)(s + \ker f) = rs + \ker f$. For showing that multiplication is well defined, we have to show that $(r + \ker f)(s + \ker f) = (r + a + \ker f)(s + b + \ker f)$ for all $a, b \in \ker f$. Now

$$(r + a + \ker f)(s + b + \ker f) = rs + rb + as + ab + \ker f,$$

hence it is sufficient to show that $rb, as, ab \in \ker f$. But $f(rb) = f(r)f(b) = 0$ since $b \in \ker f$, and similarly $f(as) = f(ab) = 0$. Thus the claim follows.

It remains to construct an isomorphism $\lambda : R/\ker f \longrightarrow \operatorname{im} f$. To this end we set $\lambda(r + \ker f) = f(r)$. This is well defined since $f(\ker f) = 0$. Next, λ is a ring homomorphism since $\lambda((r + \ker f)(s + \ker f)) = \lambda(rs + \ker f) = f(rs) = f(r)f(s) = \lambda(r + \ker f) \cdot \lambda(s + \ker f)$.

[1] For nonabelian groups, there are problems ahead. See Exer. 2.7.

Moreover, λ is surjective by definition since if $t \in \operatorname{im} f$, say $t = f(r)$, then $t = \lambda(r + \ker f)$. Finally, $\ker \lambda = \{r + \ker f : f(r) = 0\} = 0 + \ker f$, so λ is injective. We remark in addition that if f is a homomorphism of unital rings, then so is λ. □

Groups (as well as rings) will be all over the place in this book. We will gradually become acquainted first with elements of a group, their orders, or the subgroups that they generate. Later, we will add a level of abstraction and try to understand interesting groups by constructing homomorphisms into groups that we understand well.

2.2. The Parabola

Where we give the points on a parabola a structure.

Consider the parabola $\mathcal{P} : y = x^2$. This is an object that most students have studied extensively in school. It is all the more surprising that the parabola has kept a couple of secrets. In fact, in particular the younger readers may have never heard of the following jewels in the classical literature: there is the fantastic computation of the areas of segments of parabolas by Archimedes, the books on just the geometric aspects of conic sections by Appolonius, or a famous theorem of Pascal on hexagons in conics.

One of the parabolic secrets we are about to discuss is the fact that the points on a parabola form a group. We will show below how to add two **rational points** (these are points with coordinates in the field $\mathbb{Q}$ of rational numbers), and then observe that the formulas we obtain in this way are valid also for adding **integral points** (points whose coordinates are integers) or points whose coordinates lie in some residue class ring, and even, in complete generality, for adding points whose coordinates lie in an arbitrary ring R.

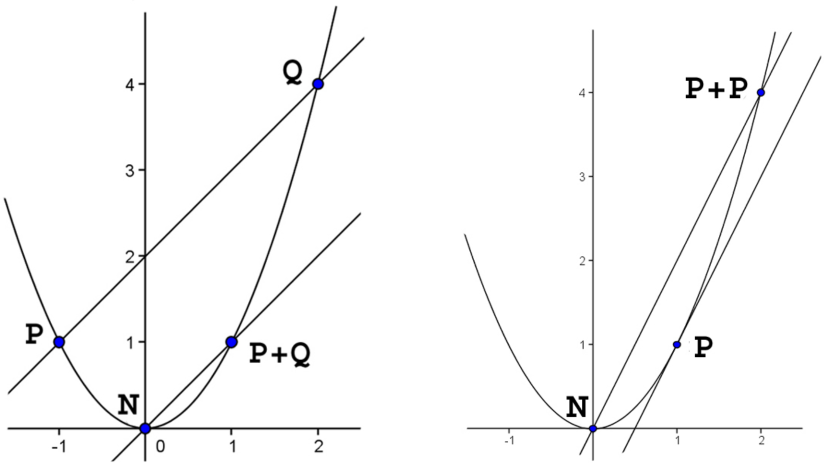


Fig. 2.1. Addition of points on the parabola $Y = X^2$: the examples $2 - 1 = 1$ and $1 + 1 = 2$.

In fact, fix the vertex $N = (0, 0)$ of the parabola. We define the sum $P \oplus Q$ of two points P and Q is the second point of intersection of the parabola and the parallel to $\overline{PQ}$ through N; if $P = Q$, replace the line $\overline{PQ}$ by the tangent to the parabola at P. As an example, the picture on the right displays the addition $1 + 1 = 2$.

Let us derive formulas for the addition of two points $P_1(x_1, x_1^2)$ and $P_2(x_2, x_2^2)$. If $P_1 \neq P_2$, the line through P_1 and P_2 has slope $m = \frac{x_2^2 - x_1^2}{x_2 - x_1} = x_2 + x_1$. The parallel to $\overline{P_1P_2}$ through N has equation $y = mx$, and for intersecting it with the parabola we have to solve $mx = x^2$. This gives us the two points of intersection $N = (0,0)$ and $R = (m, m^2)$.

By our definition of addition we have

$$(x_1, x_1^2) \oplus (x_2, x_2^2) = (x_1 + x_2, (x_1 + x_2)^2). \tag{2.2}$$

This formula remains valid even if $P_1 = P_2$. In fact, in this case the tangent to $P_1 = (x_1, x_1^2)$ has slope $2x_1$ by calculus, and as above we find $P_1 \oplus P_1 = (2x_1, (2x_1)^2)$.

You may have observed that we did not specify that our points P_1 and P_2 be rational. As a matter of fact, the addition law (2.2) defines a group law on the parabola for all points in

$$\mathcal{P}(R) = \{(x, y) \in R \times R : y = x^2\}$$

for arbitrary rings R. For example, the parabola over $\mathbb{Z}/4\mathbb{Z}$ has the following points:

$$\mathcal{P}(\mathbb{Z}/4\mathbb{Z}) = \{([0],[0]), ([1],[1]), ([2],[0]), ([3],[1])\}.$$

The sum of $P = ([1],[1])$ and $Q = ([2],[0])$ is given by $P \oplus Q = ([3],[1])$.

Proposition 2.7. *For any ring R, the addition formulas (2.2) define a group law on $\mathcal{P}(R)$. The map $\alpha : (x, x^2) \to x$ defines an isomorphism between $(\mathcal{P}(R), \oplus)$ and the ordinary additive group $(R, +)$ of the ring R.*

Proof. We verify the group axioms directly.

a) The element $N = (0,0)$ is the neutral element since clearly $N + P = P + N = P$ for all $P \in \mathcal{P}(R)$.
b) The inverse of $P = (x, x^2)$ is the point $-P = (-x, x^2)$; as it follows easily from (2.2) that $P \oplus -P = N$.
c) Given three points $P_j = (x_j, x_j^2) \in \mathcal{P}(R)$ $(j = 1,2,3)$ it is easily seen that $P_1 \oplus (P_2 \oplus P_3)$ and $(P_1 \oplus P_2) \oplus P_3$ both have the same coordinates $(x_1 + x_2 + x_3, (x_1 + x_2 + x_3)^2)$; here we have used associativity in $(R, +)$.

Checking that α is a homomorphism is also trivial:

$$\alpha(P_1 \oplus P_2) = \alpha(x_1 + x_2, (x_1 + x_2)^2) = x_1 + x_2 = \alpha(P_1) + \alpha(P_2)$$

for any two points $P_j = (x_j, x_j^2) \in \mathcal{P}(R)$. The map α is obviously onto: given $x \in R$ we have $x = \alpha(P)$ for $P = (x, x^2) \in \mathcal{P}(R)$. Moreover, α is injective because

$$\ker \alpha = \{P \in \mathcal{P}(R) : \alpha(P) = 0\} = \{(0,0)\} = \{N\}.$$

This proves all claims. □

The fact that $(\mathcal{P}(R), \oplus) \simeq (R, +)$ means that the group law on the points of a parabola defined over some ring R is not very exciting since it gives us only the "known" group $(R, +)$. The only advantage of interpreting this group structure geometrically is that it provides us with the simplest not completely trivial example of an algebraic group. In the next section we will see that the group of points on the hyperbola $\mathcal{H} : XY = 1$ also gives a group we already know, namely the unit group $R^\times$ of R. The advantage of describing this group law geometrically as a group law on conics will become clear in the chapters on Pell conics and on algorithmic number theory.

2.3. The Hyperbola

Where we show that the points on a hyperbola defined over some ring R form a group isomorphic to the unit group of R.

In this section we will study the hyperbola

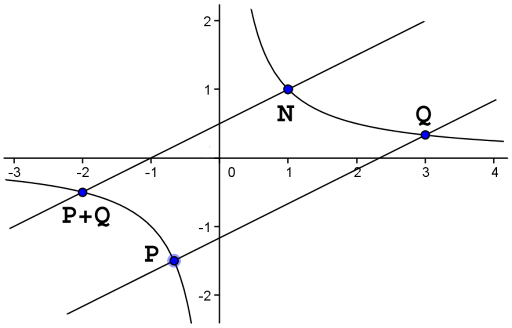
$$\mathcal{H} : XY = 1. \tag{2.3}$$

Over the integers, $\mathcal{H}$ only has two points, namely $(1,1)$ and $(-1,-1)$. Over a general ring R, we set

$$\mathcal{H}(R) = \{(x,y) \in R \times R : xy = 1\}.$$

Clearly both x and y must be units in R; conversely, if x is a unit in R, then $(x, \frac{1}{x}) \in \mathcal{H}(R)$. Thus studying points on the hyperbola $\mathcal{H}$ with coordinates in some ring R is the same as studying the unit group $R^\times$ of R.

This fact will become even more obvious after having given a group law on the hyperbola: fix a point $N = (1,1)$, and define the sum $P \oplus Q$ of two points P and Q to be the second point of intersection of the hyperbola $\mathcal{H}$ and the parallel to PQ through N; as before, if $P = Q$ then we replace the line through P and Q by the tangent to $\mathcal{H}$ at P.



For computing addition formulas, we write $P = (a, \frac{1}{a})$ and $Q = (b, \frac{1}{b})$; the slope of the line PQ is $m = \frac{\frac{1}{b} - \frac{1}{a}}{b-a} = -\frac{1}{ab}$, and the line through N parallel to PQ is given by the equation $y = -\frac{1}{ab}(x-1) + 1$. Plugging this into $xy = 1$ gives $-\frac{1}{ab}(x-1) \cdot x + x = 1$, or $x - 1 = \frac{1}{ab}(x-1) \cdot x$. The solution $x_1 = 1$ gives us the first point of intersection N; canceling $x - 1$ gives us the x-coordinate $x_2 = ab$ of the second point of intersection.

Thus the group law on the hyperbola with neutral element $N = (1,1)$ is given by

$$\Big(a, \frac{1}{a}\Big) \oplus \Big(b, \frac{1}{b}\Big) = \Big(ab, \frac{1}{ab}\Big). \tag{2.4}$$

It is easily checked that this formula also is valid in the case where $a = b$.

Although we have worked with rings R contained in $\mathbb{R}$ in the calculation above, Equation (2.4) allows us to define a group law on $\mathcal{H}(R)$ for an arbitrary ring R. The map μ sending $P = (x, \frac{1}{x}) \in \mathcal{H}(R)$ to its x-coordinate is a homomorphism $\mu : \mathcal{H}(R) \longrightarrow R^\times$, which is easily seen to be an isomorphism. Thus we have proved

Proposition 2.8. *For any ring R, the addition formulas (2.4) define a group law on $\mathcal{H}(R)$. The map $\mu : (x, \frac{1}{x}) \to x$ defines an isomorphism between $(\mathcal{H}(R), \oplus)$ and the unit group $(R^\times, \cdot)$ of the ring R.*

Chapter 3 will be devoted to a detailed investigation of the hyperbola over residue class rings: we will study the group $\mathcal{H}(\mathbb{Z}/m\mathbb{Z}) \simeq (\mathbb{Z}/m\mathbb{Z})^\times$ for integers $m \geq 2$, determine its order and its structure, and give a few applications. In Chapter 4 we will study general Pell conics over residue class rings.

Consequences of the Group Structure

After having defined group structures on the points of parabolas and hyperbolas let me remark that groups were not invented to give group theorists something to work on. The existence of a group structure on a set of objects usually may be exploited for computing and proving results.

For example, the fact that the points on the hyperbola $XY = 1$ form a group may be used for proving "Wilson's Theorem". Consider the hyperbola $\mathcal{H}$ over the field $R = \mathbb{Z}/p\mathbb{Z}$. The map $\iota : \mathcal{H} \longrightarrow \mathcal{H}$ defined by $\iota(x, y) = (y, x)$ is a bijection since $\iota \circ \iota$ is the identity map (maps with this property are called **involutions**). Thus points on $\mathcal{H}$ come in pairs $\{(x, y), (y, x)\}$. There are, however, also "single" points that are fixed by ι: if $\iota(P) = P$ for $P = (x, y)$, then $x = y$, hence $P = (1, 1)$ or $P = (-1, -1)$.

Now take the product of the x-coordinates of all points on $\mathcal{H}(\mathbb{Z}/p\mathbb{Z})$. Each pair $\{(x, y), (y, x)\}$ contributes a trivial factor 1 since $xy = 1$. The two single points contribute -1. Thus the product of the x-coordinates of all points is equal to -1, and since the x-coordinates are just the nonzero elements of $\mathbb{Z}/p\mathbb{Z}$, we have proved that $[1] \cdot [2] \cdots [p - 1] = [(p - 1)!] = [-1]$. This result is called Wilson's Theorem:

Proposition 2.9. *For every odd prime number p, we have $(p - 1)! \equiv -1 \bmod p$.*

We remark that the converse of Wilson's Theorem is also true: if $(p-1)! \equiv -1 \bmod p$, then none of the integers $1, 2, 3, \ldots, p-1$ is divisible by a prime factor of p, hence p must be irreducible and therefore prime.

In the next chapter we will discuss more consequences of the group law on the hyperbola, in particular Fermat's First Theorem and Euler's generalization.

The Hyperbola $X^2 - Y^2 = 1$

Consider the hyperbolas $\mathcal{H} : XY = 1$ and $\mathcal{H}' : T^2 - U^2 = 1$. Since we can write $\mathcal{H}'$ in the form $(T + U)(T - U) = 1$, the map $\lambda : (T, U) \to (T + U, T - U)$ maps $\mathcal{H}'$ to $\mathcal{H}$, and the point $N' = (1, 0)$ on $\mathcal{H}'$ gets mapped to $N = (1, 1)$ on $\mathcal{H}$. It is an instructive exercise to show that $\psi : (t, u) \to (t + u, t - u)$ is a group homomorphism $\mathcal{H}' \longrightarrow \mathcal{H}$, where the group structure on $\mathcal{H}'$ is defined geometrically exactly as that on $\mathcal{H}$: the sum P_3 of two points P_1 and P_2 is the second point of intersection of the line through N' parallel to P_1P_2.

For verifying the claim that λ is a homomorphism we first compute the algebraic group law. Given two points $P_1 = (t_1, u_1)$ and $P_2 = (t_2, u_2)$ on $\mathcal{H}'$, the point $P_3 = P_1 + P_2$ has coordinates

$$(t_3, u_3) = (t_1t_2 + u_1u_2, t_1u_2 + t_2u_1).$$

For verifying this addition formula we have to check that the lines P_1P_2 and $N'P_3$ are parallel, i.e., that

$$\frac{u_2 - u_1}{t_2 - t_1} = \frac{t_1u_2 + t_2u_1}{t_1t_2 + u_1u_2 - 1}.$$

Clearing denominators shows that this is equivalent to

$$(u_2 - u_1)(t_1t_2 + u_1u_2 - 1) = (t_1u_2 + t_2u_1)(t_2 - t_1),$$

which in turn can be written in the form

$$u_1 - u_2 = (t_2^2 - u_2^2)u_1 - (t_1^2 - u_1^2)u_2.$$

Since $t_1^2 - u_1^2 = t_2^2 - u_2^2 = 1$, this implies the claim.

The converse map $\mu : \mathcal{H} \longrightarrow \mathcal{H}'$ is given by

$$\mu(x,y) = \Big(\frac{x+y}{2}, \frac{x-y}{2}\Big), \tag{2.5}$$

and is defined only in domains in which 2 has an inverse, for example over the ring $R = \mathbb{Q}$, or over $R = \mathbb{Z}/m\mathbb{Z}$ for odd numbers m. In these cases, the composition of μ and λ is the identity map, which implies that $\mathcal{H}(R) \simeq \mathcal{H}'(R)$ for such domains.

2.4. The Unit Circle

Where we give various interpretations of the group law on the unit circle and use it for deriving the quadratic character of -1.

Our next example of a conic is the unit circle $\mathcal{C} : X^2 + Y^2 = 1$. We will explain how to make the set of rational points on $\mathcal{C}$ into a group in several different ways, all of which may be generalized to general Pell conics.

The geometric group law. Fix the point $N = (1,0)$ and define the sum of two points P and Q on the unit circle as the second point of intersection of the circle with the parallel to PQ through N. Writing $P = (r,s)$ and $Q = (t,u)$, the slope of the line PQ is $m = \frac{s-u}{r-t}$, and the line $y = m(x-1)$ through N parallel to PQ intersects the unit circle in $P+Q = (x,y)$ with

$$x = \frac{m^2-1}{m^2+1}, \qquad y = \frac{-2m}{m^2+1}.$$

It can be checked with some effort that these formulas can be simplified to

$$x = rt - su, \qquad y = ru + st.$$

In fact, all we have to do is verify that (x,y) is on the unit circle, which is easy:

$$x^2 + y^2 = (rt-su)^2 + (ru+st)^2 = (r^2+s^2)(t^2+u^2) = 1,$$

and that the slope of the line through (x,y) and N is $\frac{y}{x-1} = m$. This is more involved: the equation

$$\frac{y}{x-1} = \frac{ru+st}{rt-su-1} = \frac{s-u}{r-t}$$

is equivalent to

$$r^2u + rst - rtu - st^2 = rst - s^2u - s - rtu + su^2 + u,$$

that is, to

$$(r^2+s^2-1)u = s(t^2+u^2-1),$$

which clearly holds since $r^2 + s^2 = t^2 + u^2 = 1$. The calculations in the case $P = Q$ are left to the readers.

The algebraic group law. Identify the points $(x,y) \in \mathbb{R}^2$ with the complex number $x + iy \in \mathbb{C}$. Then two points (r,s) and (t,u) on the unit circle can be added by multiplying the corresponding complex numbers and pulling the result back to $\mathbb{R}^2$. Since

$$(r+si)(t+ui) = (rt-su) + (ru+st)i,$$

we find

$$(r,s) + (t,u) = (rt-su, ru+st).$$

The analytic group law. Parametrize the unit circle via $X = \cos\alpha$ and $Y = \sin\alpha$; we then get a group law by adding the angles corresponding to the points on the unit circle. In fact, if

$$(r, s) = (\cos\alpha, \sin\alpha) \quad \text{and} \quad (t, u) = (\cos\beta, \sin\beta),$$

then

$$\begin{aligned} \cos(\alpha+\beta) &= \cos\alpha\cos\beta - \sin\alpha\sin\beta = rt - su, \\ \sin(\alpha+\beta) &= \cos\alpha\sin\beta + \cos\beta\sin\alpha = ru + st, \end{aligned}$$

which gives us the same group law as above.

The diagrams in Fig. 2.2 illustrate the fact that addition of points on the unit circle amounts to adding the corresponding angles.

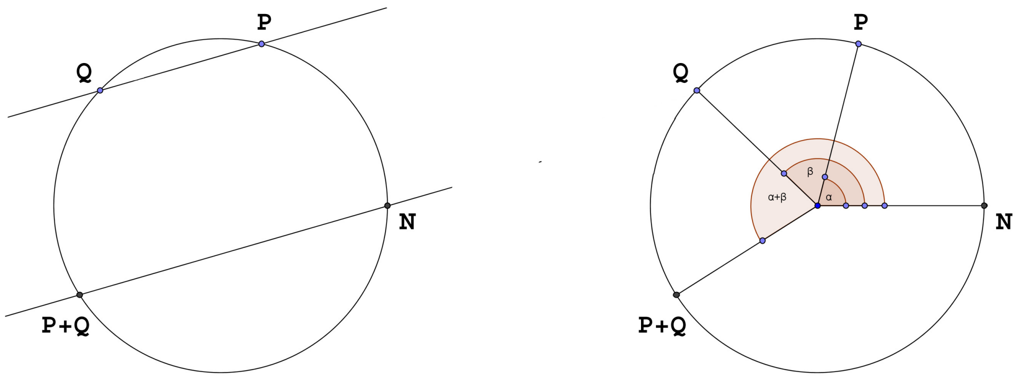


Fig. 2.2. Group Law on the Unit Circle

The complex analytic group law. This method explains the addition formulas for the trigonometric functions: points on the unit circle are parametrized by $\cos t + i\sin t = e^{it}$, and the group law on the real unit circle comes from the multiplication of complex numbers via $e^{ia}e^{ib} = e^{i(a+b)}$. In this description it is not at all obvious that the sum of two rational points is rational again. This is explained, however, by identifying $e^{it} = \cos t + i\sin t$ with the point $(\cos t, \sin t)$ on the real unit circle $X^2 + Y^2 = 1$ and using the addition formulas for the sine and the cosine

$$\begin{aligned} \cos(\alpha+\beta) &= \cos\alpha\cos\beta - \sin\alpha\sin\beta, \\ \sin(\alpha+\beta) &= \sin\alpha\cos\beta + \cos\alpha\sin\beta, \end{aligned}$$

which in turn follow from the interpretation of points $(\cos t, \sin t)$ on the real unit circle as complex numbers e^{it}: the addition formulas follow from comparing real and imaginary parts on both sides of the following equation:

$$\begin{aligned} \cos(\alpha+\beta) + i\sin(\alpha+\beta) = e^{i(\alpha+\beta)} &= e^{i\alpha}e^{i\beta} = (\cos\alpha + i\sin\alpha)(\cos\beta + i\sin\beta) \\ &= \cos\alpha\cos\beta - \sin\alpha\sin\beta + i(\sin\alpha\cos\beta + \cos\alpha\sin\beta) \end{aligned}$$

The fact that adding points on the unit circle corresponds to adding the corresponding angles generalizes to arbitrary Pell conics. We will come back to this observation in the last section of this chapter.

2.5. The Pythagorean Pell Conic

Where we find that certain Pell conics have been studied since antiquity.

As our last example of a Pell conic in this chapter we now look at the Pell conic

$$\mathcal{C} : X^2 - 2Y^2 = 1.$$

It is perhaps the most interesting example in that it has infinitely many *integral* points (points whose coordinates are integers), which are, however, not found as easily as in the (rather trivial) case of the parabola.

As before, we define a group law on the set $\mathcal{C}(R)$ of R-integral points on $\mathcal{C}$ by demanding that the sum of two points P_1, P_2 is the second point of intersection of the parallel to P_1P_2 through $N = (1,0)$. We claim that algebraically, this group law is given by the addition formula

$$(x_1,y_1) + (x_2,y_2) = (x_3,y_3), \quad \text{where} \quad x_3 = x_1x_2 + 2y_1y_2, \quad y_3 = x_1y_2 + x_2y_1. \tag{2.6}$$

Our proof will work over fields R, such as $R = \mathbb{Q}$, $R = \mathbb{R}$ or $R = \mathbb{Z}/p\mathbb{Z}$. In this case, the line through P_1P_2 has slope $m = \frac{y_2-y_1}{x_2-x_1}$, hence the parallel through $N = (1,0)$ is given by the equation $y = m(x-1)$. Intersecting this line with the conic gives $x^2-1-2m^2(x-1)^2 = 0$, which can be written in the form $(x-1)(x+1-2m^2(x-1)) = 0$. This equation has two solutions, namely $x_0 = 1$ (giving rise to the point of intersection N) and

$$x_3 = \frac{2m^2+1}{2m^2-1} = \frac{2(y_2-y_1)^2 + (x_2-x_1)^2}{2(y_2-y_1)^2 - (x_2-x_1)^2}.$$

We claim that

$$x_3 = x_1x_2 + 2y_1y_2 \quad \text{and} \quad y_3 = m(x_3-1) = x_1y_2 + x_2y_1.$$

These identities may be verified by a rather technical calculation (or by typing suitable commands into a computer algebra system, for example `pari`), and the reader is invited to do so. It turns out to be much simpler to verify the following claims, which also establish what we want:

1. If $P_1, P_2 \in \mathcal{C}(R)$, then $P_3 = P_1 \oplus P_2 = (x_3,y_3) \in \mathcal{C}(R)$.
   This follows easily from

$$\begin{aligned} x_3^2 - 2y_3^2 &= (x_1x_2 + 2y_1y_2)^2 - 2(x_1y_2 + x_2y_1)^2 \\ &= x_1^2x_2^2 + 4x_1x_2y_1y_2 + 4y_1^2y_2^2 - 2x_1^2y_2^2 - 4x_1x_2y_1y_2 - 2x_2^2y_1^2 \\ &= (x_1^2 - 2y_1^2)(x_2^2 - 2y_2^2) = 1. \end{aligned}$$

2. The lines P_1P_2 and NP_3 are parallel.
   In fact, the slopes m and n of the lines P_1P_2 and NP_3, respectively, are

$$m = \frac{y_2-y_1}{x_2-x_1} \quad \text{and} \quad n = \frac{y_3}{x_3-1} = \frac{x_1y_2 + x_2y_1}{x_1x_2 + 2y_1y_2 - 1}.$$

   Thus $m = n$ is equivalent to $(x_2-x_1)y_3 = (y_2-y_1)(x_3-1)$, that is, to

$$(x_2-x_1)(x_1y_2 + x_2y_1) = (y_2-y_1)(x_1x_2 + 2y_1y_2 - 1).$$

   Simplifying these expressions we end up with

$$y_1(x_2^2 - 2y_2)^2 - y_2(x_1^2 - 2y_1^2) = y_1 - y_2,$$

   which holds since $x_1^2 - 2y_1^2 = x_2^2 - 2y_2^2 = 1$.

Thus we have shown that over fields, the geometric group law gives the simple addition formulas (2.6). On the other hand, (2.6) allows us to define the sum of two points on $\mathcal{C}(R)$ for an arbitrary ring R, since verifying the group axioms is a trivial task: the neutral element is $N = (1, 0)$, the inverse of $P = (r, s)$ is $-P = (r, -s)$.

Proposition 2.10. *Let $\mathcal{C} : X^2 - 2Y^2 = 1$ denote the Pythagorean Pell conic. There are infinitely many integers $(x_n, y_n) \in \mathcal{C}(\mathbb{Z})$ given by the following recursion:*

$$(x_0, y_0) = (1, 0), \qquad x_{n+1} = 3x_n + 4y_n, \quad y_{n+1} = 2x_n + 3y_n.$$

In fact, every solution in positive integers is given by (x_n, y_n) for some $n \geq 1$.

Proof. Clearly (x_{n+1}, y_{n+1}) is an integral point on the Pythagorean Pell conic $X^2 - 2Y^2 = 1$ since

$$x_{n+1}^2 - 2y_{n+1}^2 = (3x_n + 4y_n)^2 - 2(2x_n + 3y_n)^2 = x_n^2 - 2y_n^2 = 1.$$

For showing that there are no other positive solutions, assume that (t, u) is an integral point on $\mathcal{C}$ with positive integral coordinates. We may and will assume that $t \geq 3$ and $u \geq 2$. Then there exists an integer n such that $x_n \leq t < x_{n+1}$. This implies $y_n \leq u < y_{n+1}$: for example, $2u^2 = t^2 - 1 \geq x_n^2 - 1 = 2y_n^2$ shows that $u \geq y_n$.

Now define a sequence of points (t_k, u_k) with $1 \leq k \leq n$ by setting $t_n = t$, $u_n = u$, as well as

$$t_{n-1} = 3t_n - 4u_n, \qquad u_{n-1} = 3u_n - t_n.$$

With some effort (and induction) it can be shown that the inequalities

$$x_n < t_n < x_{n+1} \quad \text{and} \quad y_n < u_n < y_{n+1}$$

imply

$$x_{n-1} < t_{n-1} < x_n \quad \text{and} \quad y_{n-1} < u_{n-1} < y_n.$$

In fact, the inequality

$$x_{n-1} = 3x_n - 4y_n < 3t_n - 4u_n = t_{n-1},$$

for example, is equivalent to

$$\frac{t_n - x_n}{u_n - y_n} < \frac{4}{3}.$$

This in turn is equivalent to the fact that the slope of the line through the points (t_n, u_n) and (x_n, y_n) is bounded above by $\frac{4}{3}$, the slope of the tangent at the point $(3, 2)$.

Now we can perform descent: if there is a point (t_n, u_n) not given by our formulas, there must be an integral point (t_0, u_0) satisfying

$$1 \leq t_0 < x_1 = 3, \qquad 0 \leq u_0 \leq y_1 = 2.$$

The only such point is $(t_0, u_0) = (x_0, y_0)$, and this implies that $t = x_n$ and $u = y_n$, which proves our claims. □

This result provides us with the following solutions:

n	$T^2 - 2U^2 = 1$	$T^2 - 2U^2 = -1$
0	$(1, 0)$	$(1, 1)$
1	$(3, 2)$	$(7, 5)$
2	$(17, 12)$	$(41, 29)$
3	$(99, 70)$	$(239, 169)$

The proof of Prop. 2.10 becomes a lot easier to follow when we embed $\mathcal{P}(\mathbb{Z})$ into the 2-dimensional Euclidean vector space $\mathbb{R} \times \mathbb{R}$. In fact, to each rational point $Q = (x, y) \in \mathcal{C}(\mathbb{Q})$ on the Pythagorean Pell conic $\mathcal{C} : X^2 - 2Y^2 = 1$ we associate the real number $\pi(Q) = x + y\sqrt{2} \in \mathbb{Q}(\sqrt{2}\,)$, where $\mathbb{Q}(\sqrt{2}\,)$ is the set of all elements $x + y\sqrt{2}$ with rational x, y. This set is a domain with respect to addition and multiplication defined by

$$\begin{aligned} (x_1 + y_1\sqrt{2}\,) + (x_2 + y_2\sqrt{2}\,) &= (x_1 + x_2) + (y_1 + y_2)\sqrt{2}, \\ (x_1 + y_1\sqrt{2}\,) \cdot (x_2 + y_2\sqrt{2}\,) &= (x_1x_2 + 2y_1y_2) + (x_1y_2 + x_2y_1)\sqrt{2}, \end{aligned}$$

and it is a field since

$$\frac{x_1 + y_1\sqrt{2}}{x_2 + y_2\sqrt{2}} = \frac{(x_1 + y_1\sqrt{2}\,)(x_2 - y_2\sqrt{2}\,)}{(x_2 + y_2\sqrt{2})(x_2 - y_2\sqrt{2}\,)} = \frac{x_1x_2 - 2y_1y_2}{x_2^2 - 2y_2^2} + \frac{x_2y_1 - x_1y_2}{x_2^2 - 2y_2^2}\sqrt{2}$$

for all nonzero elements $x_2 + y_2\sqrt{2}$.

The element corresponding to the "fundamental solution" $P = (3, 2)$ is $\pi(P) = 3 + 2\sqrt{2}$. The basic observation is that the sum $P_1 \oplus P_2$ on the Pell conic corresponds to multiplication $\pi(P_1) \cdot \pi(P_2)$ in $\mathbb{Q}(\sqrt{2}\,)^\times$, the nonzero elements of the field $\mathbb{Q}(\sqrt{2}\,)$. In other words:

Lemma 2.11. *The map $\pi : \mathcal{C}(\mathbb{Q}) \longrightarrow \mathbb{Q}(\sqrt{2}\,)^\times$ defined by $\pi((x, y)) = x + y\sqrt{2}$ is a homomorphism.*

In fact, if $P_1 \oplus P_2 = P_3$ for points $P_j = (x_j, y_j)$ $(j = 1, 2, 3)$ on $\mathcal{C}(\mathbb{Q})$, then

$$x_3 = x_1x_2 + 2y_1y_2, \quad y_3 = x_1y_2 + x_2y_1$$

according to (2.6); on the other hand,

$$\begin{aligned} \pi(P_1) \cdot \pi(P_2) &= (x_1 + y_1\sqrt{2}\,)(x_2 + y_2\sqrt{2}\,) \\ &= x_1x_2 + 2y_1y_2 + (x_1y_2 + x_2y_1)\sqrt{2} = \pi(P_3) = \pi(P_1 \oplus P_2). \end{aligned}$$

The homomorphism π has no chance of being surjective since $\mathbb{Q}(\sqrt{2}\,)$ contains many elements that do not come from rational points on $\mathcal{C}$. In fact, if $x + y\sqrt{2} = \pi((x, y))$, then clearly $x^2 - 2y^2 = 1$, and so the unit $1 + \sqrt{2}$, which satisfies $1^2 - 2 \cdot 1^2 = -1$, is an element of $\mathbb{Q}(\sqrt{2}\,)^\times$ that is not in the image of π. Let us call $N(x + y\sqrt{2}\,) = x^2 - 2y^2$ the **norm** of the element $x + y\sqrt{2}$; it is easily checked that the norm is a homomorphism $N : \mathbb{Q}(\sqrt{2}\,) \longrightarrow \mathbb{Q}^\times$:

$$\begin{aligned} N(x_1 + y_1\sqrt{2}\,) \cdot N(x_2 + y_2\sqrt{2}\,) &= (x_1^2 - 2y_1^2)(x_2^2 - 2y_2^2) \\ &= (x_1x_2 + 2y_1y_2)^2 - 2(x_1y_2 + x_2y_1)^2 \\ &= N((x_1 + y_1\sqrt{2}\,) \cdot (x_2 + y_2\sqrt{2}\,)) \end{aligned}$$

The elements with norm 1 are those in the kernel of the norm map, and in fact we have $\operatorname{im} \pi = \ker N$.

We now have constructed the following maps:

$$\begin{tikzcd} 0 \arrow[r] & \mathcal{C}(\mathbb{Z}) \arrow[r, "\pi"] \arrow[d] & \mathbb{Z}[\sqrt{2}\,]^\times \arrow[r, "N"] \arrow[d] & \mathbb{Z}^\times \arrow[r] \arrow[d] & 1 \\ 0 \arrow[r] & \mathcal{C}(\mathbb{Q}) \arrow[r, "\pi"] & \mathbb{Q}(\sqrt{2}\,)^\times \arrow[r, "N"] & \mathbb{Q}^\times & \end{tikzcd}$$

The 0's and 1's occurring here (the 0 stands for the trivial subgroup of an additively written group, the 1 for the trivial subgroup of a multiplicatively written group) have a meaning that is explained by the fact that the rows in this diagram are **exact sequences**. We say that a sequence $A \longrightarrow B \longrightarrow C$ of, say, abelian groups is exact at B if the kernel of the map $B \longrightarrow C$ is equal to the image of the preceding map $A \longrightarrow B$. In our example, the kernel of the norm map $N : \mathbb{Z}[\sqrt{2}\,]^\times \longrightarrow \mathbb{Z}^\times = \{\pm 1\}$ consists of units $x + y\sqrt{2}$ with $1 = N(x + y\sqrt{2}\,) = x^2 - 2y^2$, hence have the form $\pi((x, y))$ for $(x, y) \in \mathcal{C}(\mathbb{Z})$. You can verify similarly that the diagram is exact at all the other places where it makes sense. See Exer. 2.38 – 2.39 for more on exact sequences.

For now, diagrams such as the one above only help us keep track of the groups and maps involved in more complicated proofs, and so have the main purpose of adding clarity.

Let us now return to Prop. 2.10. Using the number field $\mathbb{Q}(\sqrt{2}\,)$, we can state its content as

Theorem 2.12. *The solutions of the equation*

$$T^2 - 2U^2 = +1$$

are given by

$$T + U\sqrt{2} = (-1)^a(3 + 2\sqrt{2}\,)^b, \tag{2.7}$$

where $a \in \{0, 1\}$ *and* $b \in \mathbb{Z}$.

Since the map sending $T + U\sqrt{2} \longmapsto (a, b) \in \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}$, with $T + U\sqrt{2}$ as in (2.7), is an isomorphism of groups, we can say that, as an abstract group, we have

$$\mathcal{C}(\mathbb{Z}) \simeq \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}.$$

This is less precise than explicitly giving the solutions, but in this form it can be generalized to all Pell conics.

For giving a (second) proof of Thm. 2.12, we have to verify the following assertions:

1. $(T, U) \in \mathcal{C}(\mathbb{Z})$, which is trivial;
2. if $(t, u) \in \mathcal{C}(\mathbb{Z})$, then there exist integers a, b such that $t + u\sqrt{2} = (-1)^a(3 + 2\sqrt{2}\,)^b$.

For proving the second claim it is sufficient to assume that $t, u > 0$; other signs come from replacing a by $a + 1$, or by replacing b with $-b$.

Assume therefore that $t^2 - 2u^2 = 1$ for positive integers t, u. Then there is a unique positive integer n such that

$$(3 + 2\sqrt{2}\,)^n \le t + u\sqrt{2} < (3 + 2\sqrt{2}\,)^{n+1}.$$

Multiplying everything through by $(3 - 2\sqrt{2}\,)^n = (3 + 2\sqrt{2}\,)^{-n}$ we find

$$1 \le t' + u'\sqrt{2} = (t + u\sqrt{2}\,)(3 - 2\sqrt{2}\,)^n < 3 + 2\sqrt{2}.$$

But now t' and u' are positive integers with $(t')^2 - 2(u')^2 = 1$: in fact, we have

$$1 \le t' + u'\sqrt{2} < 3 + 2\sqrt{2} \quad \text{and} \quad 3 - 2\sqrt{2} < t' - u'\sqrt{2} \le 1,$$

which easily implies $2 - \sqrt{2} < t' < 2 + \sqrt{2}$. The only solution of $T^2 - 2U^2 = 1$ in this interval is $(t', u') = (1, 0)$, and now everything follows.

The integral points with small coordinates on the two curves $\mathcal{C} : T^2 - 2U^2 = +1$ and $T^2 - 2U^2 = -1$ are displayed in Fig. 2.3, where the left and the right branches belong to the hyperbola $T^2 - 2U^2 = 1$, and the other two are part of $T^2 - 2U^2 = -1$.

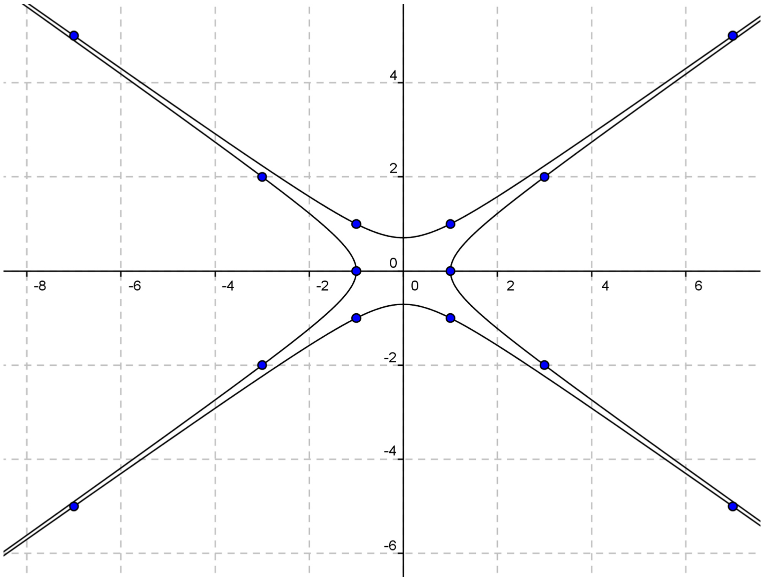


Fig. 2.3. Integral points on the curves $T^2 - 2U^2 = \pm 1$

2.6. Angles and Integrals

Where we show that adding points on Pell conics corresponds to adding suitable angles.

Angles are another object known from our school days; angles are so familiar to us that we tend to forget that they measure arc lengths, and that their proper definition involves integrals. In particular, the addition of angles is an addition of integrals, and this observation deserves to be studied in detail.

Group Structure on the Unit Circle via Integrals. Let us start by recalling that

$$s = \int_a^b \sqrt{1 + (f'(x))^2}\, dx$$

is the formula for the arc length of the graph of a sufficiently nice function $y = f(x)$ between $x = a$ and $x = b$. In fact, for small values of $\varDelta x = b - a$ we have $(\varDelta s)^2 \approx (\varDelta x)^2 + (\varDelta y)^2$, hence

$$\varDelta s \approx \sqrt{1 + \frac{(\varDelta y)^2}{(\varDelta x)^2}} \cdot \varDelta x,$$

and letting $\varDelta x \longrightarrow 0$ the claim follows in the usual way from standard mean value theorems.

For the unit circle defined by $x^2+y^2 = 1$ we get $2x+2yy' = 0$ by implicit differentiation, that is, $y' = -\frac{x}{y}$. In particular, the arc length of a circle between two points $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ is given by

$$s = \int_{x_1}^{x_2} \sqrt{1 + \frac{x^2}{y^2}}\, dx = \int_{x_1}^{x_2} \frac{\sqrt{x^2 + y^2}}{y}\, dx = \int_{x_1}^{x_2} \frac{dx}{y} = \int_{x_1}^{x_2} \frac{dx}{\sqrt{1 - x^2}}\,.$$

Thus the angle $\alpha = \sphericalangle NOP_1$ between $N(1, 0)$ and $P_1(x_1, y_1)$ is given by

$$\alpha = \int_0^{x_1} \frac{dx}{\sqrt{1-x^2}}\,.$$

The addition law $P_1 \oplus P_2 = P_3$ for $P_j = (x_j, y_j)$ given by $x_3 = x_1x_2 - y_1y_2$ thus would follow from the addition formula for integrals

$$\int_0^{x_1} \frac{dx}{\sqrt{1-x^2}} + \int_0^{x_2} \frac{dx}{\sqrt{1-x^2}} = \int_0^{x_3} \frac{dx}{\sqrt{1-x^2}},$$

where

$$x_3 = x_1x_2 - \sqrt{1-x_1^2} \cdot \sqrt{1-x_2^2}.$$

For proving this addition formula we have to define the cosine as the inverse function of α via

$$s = \int_0^{\cos s} \frac{dx}{\sqrt{1-x^2}}\,,$$

the sine $\sin s = -\frac{d}{ds}\cos s$ as its negative derivative, and then prove that

$$\cos(s_1+s_2) = \cos s_1 \cdot \cos s_2 - \sin s_1 \cdot \sin s_2.$$

This can indeed be done from first principles (by this we mean that we do not need to know beforehand any properties of $\sin\alpha$ or $\cos\alpha$, and this method of introducing the trigonometric functions is completely analogous to Abel's definition of elliptic functions as inverse functions of elliptic integrals such as $\int_0^s \frac{dx}{\sqrt{1-x^4}}$. The fact that the last integral cannot be transformed into the integral of a rational functions is related to the observation that the diophantine equation $X^4 - Y^4 = Z^2$ does not have a nonzero solution; see the corresponding remarks in the Notes of Chap. 1.

Hyperbolic Angles

For giving the addition of points another geometric interpretation we define the hyperbolic angle $\alpha = \sphericalangle NOP$ attached to a point $P = (a, \frac{1}{a})$ on the hyperbola $XY = 1$ as the area cut out by the lines ON and OP and the hyperbola (see Fig. 2.4).

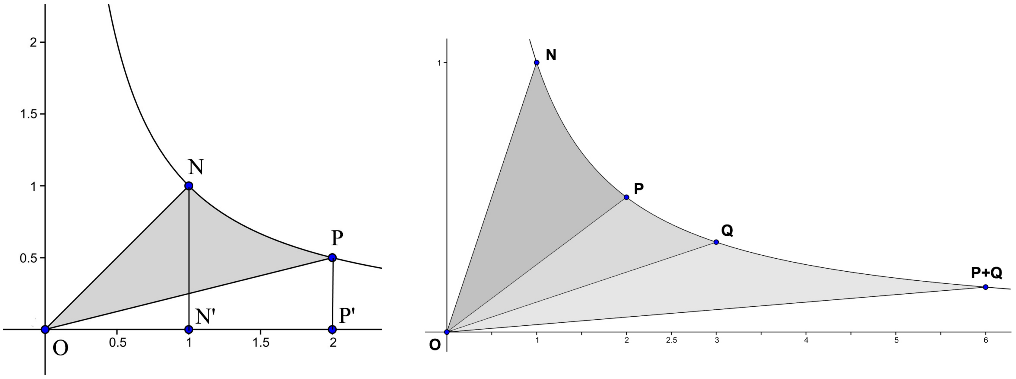


Fig. 2.4. Hyperbolic Angles

For computing α we clearly have to add

- the area of the triangle $ON'N$ with $N' = (1,0)$ and
- the area cut out by the hyperbola and the the lines $x = 1$, $x = a$, and $y = 0$,

and then subtract

- the area of the triangle $OP'P$ with $P' = (a, 0)$.

A simple calculation yields

$$\alpha = \text{area}(ON'N) + \text{area}(N'NPP') - \text{area}(OP'P) = \frac{1}{2} + \int_1^a \frac{dx}{x} - \frac{1}{2} = \log(a).$$

Since $(a, \frac{1}{a}) + (b, \frac{1}{b}) = (ab, \frac{1}{ab})$ and $\log(a) + \log(b) = \log(ab)$, the map sending $P = (a, \frac{1}{a})$ with $a > 0$ to $\log a$ is a group homomorphism from the group of real (or rational) points on $\mathcal{H}$ with positive coordinates to the reals.

Proposition 2.13. *The addition of points with positive coordinates on cH : $XY = 1$ corresponds to adding their hyperbolic angles.*

In particular, we have $\sphericalangle NOP + \sphericalangle POQ = \sphericalangle QOR$ for $R = P \oplus Q$ (see Fig. 2.4).

We can make this explicit by writing down the addition formula in the following form: given points $P(a, 1/a)$ and $Q(b, 1/b)$ with $P \oplus Q = R$, where $R = (ab, 1/ab)$, then

$$\int_1^a \frac{dx}{x} + \int_1^b \frac{dx}{x} = \int_1^{ab} \frac{dx}{x}.$$

Notes

It is not known exactly who discovered when that the square root of 2 is an irrational number, but since both Aristotle and Plato refer to this result it must have happened before 400 BC. The classical formulation of the irrationality of $\sqrt{2}$ involves the side and the diagonal of a square: these were said to be incommensurable, which meant that there is no line segment x such that both the side and the diagonal of a square are integral multiples of x.

Allusions to this result can be found in many places; Plato mentions that the diagonal of a square with sides of length 5 has length about 7; the ratio $\frac{7}{5} = 1.4$ is a rational approximation of $\sqrt{2} = 1.4142135\ldots$, and the fact that this is an approximation is reflected by the fact that $7^2 - 2 \cdot 5^2 = -1$ differs only slightly from 0. In fact, any integral point (x, y) on the two conics $X^2 - 2Y^2 = \pm 1$ gives rational approximations of $\sqrt{2}$. In his memoir on the measurement of the circle, Archimedes had to use approximations to $\sqrt{3}$, and the approximations he used come from integral points on the conics $X^2 - 3Y^2 = 1$ and $X^2 - 3Y^2 = -2$ (observe that $X^2 - 3Y^2 = -1$ does not have any rational point): the inequalities

$$\frac{1351}{780} > \sqrt{3} > \frac{265}{153}$$

come from $1351^2 - 3 \cdot 780^2 = 1$ and $265^2 - 3 \cdot 153^2 = -2$.

In fact we learn from the books on number theory by Nicomachus that the Pythagoreans must have studied numbers that happen to be coordinates of integral points on the conics $X^2 - 2Y^2 = 1$ and $X^2 - 2Y^2 = -1$.

The solutions of the Pell equations $T^2 - 2U^2 = \pm 1$ occur repeatedly in several classical Greek sources. Plato mentions in his “Republic” that the diagonal of a square on five feet is approximately seven feet. In his commentaries on this passage, Proclus quotes a result he credits to the Pythagoreans saying that

> when the diameter receives the side of which it is diameter, while the sides added to itself and receiving its diameter, becomes a diameter.

Thus if d is the diagonal of a square with side length s, then $2s+d$ is the diagonal of a square with side length $s+d$, which is easily confirmed. Starting with the approximation $s=2$ and $d=3$ (satisfying $3^2-2\cdot 2^2=1$), the new "square" has $s=5$ and $d=7$ satisfying $7^2-2\cdot 5^2=-1$.

In his comments on Plato, Theon of Smyrna even starts with $(s,d)=(1,1)$ and explicitly constructs the solutions $(3,2)$, $(7,5)$ and $(17,12)$ of $d^2-2s^2=\pm 1$ (see [Tho1978]).

Wilson's Theorem was first published (without proof) by E. Waring [War1770, p. 218] and credited to Wilson. Proofs were given by Euler (Opuscula Analytica I, p. 329), Lagrange (1771; Œuvres III, p. 425) and Gauss [Gau1801, art. 77]. Lagrange already observed that $(\frac{p-1}{2})!\pm 1$ is divisible by the prime number p according as $p=4n\pm 1$. A remarkable proof of Wilson's Theorem due to Stern [Ste18??, p. 391] is sketched in Exer. 2.27.

The group law on the rational points of the unit circle is implicit in Šimerka's article on rational triangles. On [Sim1870, p. 205], he writes

> If
>
> $$\tan\frac{1}{2}A=\frac{t}{u},\quad \text{i.e.,}\quad \sin A=\frac{2tu}{t^2+u^2},\quad \cos A=\frac{u^2-t^2}{t^2+u^2},$$
>
> and
>
> $$\tan\frac{1}{2}B=\frac{t'}{u'},\quad \text{i.e.,}\quad \sin A=\frac{2t'u'}{t'^2+u'^2},\quad \cos A=\frac{u'^2-t'^2}{t'^2+u'^2},$$
>
> then using $C=180^\circ-(A+B)$ we obtain
>
> $$\sin C=\sin(A+B)=\frac{2tu(u'^2-t'^2)}{(t^2+u^2)(t'^2+u'^2)}+\frac{2t'u'(u^2-t^2)}{(t^2+u^2)(t'^2+u'^2)},$$
>
> that is,
>
> $$\sin C=\frac{2(uu'-tt')(tu'+t'u)}{(t^2+u^2)(t'^2+u'^2)}.$$

Later Šimerka remarks (here $C=180^\circ-A-B$ is the third angle in his rational triangle):

> The memoir of Mr. Ligowski, vol. XLVI, p. 503 of this Archive coincides with these investigations if we set there
>
> $$x=\frac{u}{t}=\cot\frac{1}{2}A,\quad y=\frac{u'}{t'}=\cot\frac{1}{2}B,\quad z=\frac{u''}{t''}=\cot\frac{1}{2}C,\quad \rho=\frac{t''p''}{tt'},$$
>
> and each of the lines a, b, c, S, $S-a$ etc. is taken $\frac{k}{t^2t'^2}$ times. Then it follows from
>
> $$xyz=x+y+z$$
>
> that
>
> $$\frac{uu'u''}{tt't''}=\frac{u}{t}+\frac{u'}{t'}+\frac{u''}{t''},$$
>
> a result also flowing from our theory.

This is related to the addition formula for the tangent function

$$\tan(x+y)=\frac{\tan x+\tan y}{1-\tan x\tan y}.$$

The article [Nie1907] by Niewenglowski (mentioned by Dickson [Dic1920, vol II, p. 396]) also contained a hint at the geometric group law on conics. Niewenglowski considers the hyperbola $x^2-ay^2=1$ and writes

> Soient $A(1,0)$ le sommet, $A_1(x_1,y_1)$ le premier point entier à coordonnées positives; la parallèle menée par A à la tangente an A_1 donnera le point $A_2(x_2,y_2)$; la corde A_1A_3 sera parallèle à la tangente en A_2, etc., et l'on obtiendra ainsi tous les points à coordonnées entières et positives.[2]

A more recent reference to the group law on Pell conics appears in Primrose [Pri1959]. He parametrized the Pell conic $\mathcal{P} : X^2 - DY^2 = 1$ by hyperbolic functions $X = \cosh\theta$ and $Y\sqrt{D} = \sinh\theta$. If $P = (T,U)$ is an integral point on $\mathcal{P}$ and belongs to the parameter θ, then $P_r = (T_r,U_r)$ defined by $T_r + U_r\sqrt{D} = (T + U\sqrt{D}\,)^r$ belongs to $r\theta$. Moreover, if P_r and P_s are two points on $\mathcal{P}$, then the slopes of the lines P_rP_s only depend on the sum $r+s$. He then concludes that if P_1 and P_{n-1} are known, the point P_n is the second point of intersection of the conic and the line through $P_0 = (1,0)$ parallel to P_1P_{n-1}.

The field structure of the points on a parabola (see Exer. 2.11) was posed as a problem by Jacques Verdier [Ver1985] in a rather obscure French Journal called Le Petit Vert. Verdier gave a geometric proof for associativity, the geometric proof of distributivity was later provided by Joël Kieffer [Kie2003].

Our results on the structure of $\mathcal{C}(\mathbb{Z})$ may be summarized as follows:

conic	$\mathcal{C}$	$\mathcal{C}(\mathbb{Z})$
parabola	$Y = X^2$	$\mathbb{Z}$
hyperbola	$XY = 1$	$\mathbb{Z}/2\mathbb{Z}$
unit circle	$X^2 + Y^2 = 1$	$\mathbb{Z}/4\mathbb{Z}$
Eisenstein conic	$X^2 + XY + Y^2 = 1$	$\mathbb{Z}/6\mathbb{Z}$
Pythagorean Pell conic	$Y^2 - 2X^2 = 1$	$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}$

For the result concerning the Eisenstein conic, see Exer. 2.15.

Exercises

2.1 Let $f : G \longrightarrow H$ be an isomorphism of finite abelian groups, i.e., a bijective homomorphism. Show that the inverse map is also a homomorphism.

2.2 Let $f : G \longrightarrow H$ be an isomorphism of finite abelian groups.
1. Show that the order of $g \in G$ is the same as the order of $f(g) \in H$.
2. Show that if U is a subgroup of G with index r, then $f(U)$ is a subgroup of $f(G) = H$ with the same index r.
3. Show that if G is cyclic (i.e., generated by a single element), then so is H.

Informally speaking we may say that isomorphic groups have the same group theoretical properties.

2.3 Show that if G is a cyclic group, then $G \simeq \mathbb{Z}$ if G is infinite, and $G \simeq \mathbb{Z}/n\mathbb{Z}$ if G is finite with n elements.
In particular, cyclic groups are abelian, and subgroups of cyclic groups are cyclic.

2.4 Let $p \equiv 3 \bmod 4$ be a prime number. Show that every square modulo p is a fourth power.

2.5 Consider the additive group C^∞ of all infinitely often differentiable functions $(0,1) \longrightarrow \mathbb{R}$. Show that C^∞ is a ring with respect to multiplication of functions.
Verify that the derivative operator $\frac{d}{dx}$ induces a group homomorphism $C^\infty \longrightarrow C^\infty$, but that it is not a ring homomorphism.

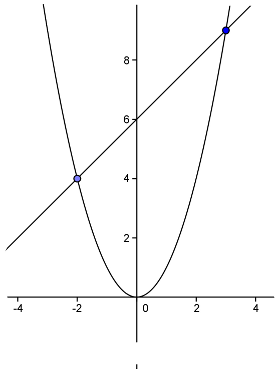
[2] Let $A(1,0)$ be the vertex, $A_1(x_1,y_1)$ the first integral point with positive coordinates; the parallel through A to the tangent at A_1 will give the point $A_2(x_2,y_2)$; the secant A_1A_3 will be parallel to the tangent at A_2, etc., and in this way we obtain all the integral points with positive coordinates.

Show that, for every $a \in (0,1)$, the map induced by $f_a \to f'(a)$ is a surjective group homomorphism $C^\infty \longrightarrow \mathbb{R}$.

2.6 Let G be a finite group with subgroup H. Show that $\#H$ divides $\#G$ (we have proved this for abelian groups; all you have to do is check carefully that everything works even for nonabelian groups).

2.7 Let G be a group with subgroup H. Show that the cosets g_iH ($g_i \in G$) can be made into a group via $g_iH \cdot g_jH = g_ig_jH$ if and only if H is a **normal subgroup**, that is, if $g^{-1}hg \in H$ for all $g \in G$ and all $h \in H$.

2.8 Given positive rational numbers a, b, consider the points $P = (a, a^2)$ and $Q = (b, b^2)$ on the parabola.



Define a map $\mathcal{P}(R) \times \mathcal{P}(R) \longrightarrow R$ by letting $P * Q$ denote the point you get by intersecting PQ with the y-axis and reflecting it at the origin. Show that $P * Q = (0, ab)$.
Show that the special case of composing $P * P$ is equivalent to the Greeks' construction of the tangent to a parabola. For drawing the tangent at a point $P = (a, a^2)$, they connected P with $(0, -a^2)$. It goes without saying that the Greeks formulated this property of parabolas without using coordinates, which were invented by Fermat and Descartes.

2.9 Show that the line connecting $P(a, \frac{1}{a})$ and $Q(2a, 0)$ is the tangent to the hyperbola $\mathcal{H} : XY = 1$ in P.

2.10 Show that the line connecting the points $P(a, b)$ on the unit circle $\mathcal{C} : X^2 + Y^2 = 1$ and $Q(\frac{1}{a}, 0)$ is the tangent to $\mathcal{C}$ at P.

2.11 Consider the parabola $\mathcal{P} : Y = X^2$ with $N = (0,0)$ and any fixed integral point $I \neq N$, say $I = (1,1)$. Define the addition of two points as in Sect. 2.2., and define a multiplication $P * Q$ for points $P, Q \in \mathcal{P}(\mathbb{Q}) \setminus \{N\}$ as follows: The line PQ intersects the y-axis in A, and the line IA intersects $\mathcal{P}$ in $B = P * Q$.
Show that $\mathcal{P}(\mathbb{Q}, +, *)$ is a field. Is it isomorphic to $\mathbb{Q}$?

2.12 Show that $X = e^t$, $Y = e^{-t}$ parametrizes the real hyperbola $\mathcal{H} : XY = 1$, and that the group law on $\mathcal{H}(\mathbb{R})$ is induced by adding exponents.

2.13 Show that the homomorphism $\mu : \mathcal{H} \longrightarrow \mathcal{H}'$ in (2.5) transforms the parametrization of $\mathcal{H}$ by the exponential function into the parametrization of $\mathcal{H}'$ by hyperbolic functions.

2.14 Show that the map $\lambda : (x, y) \to (t, u) = (x + y, x - y)$ defines a map from the unit circle $\mathcal{C} : X^2 + Y^2 = 1$ to the circle $\mathcal{C}' : T^2 + U^2 = 2$, and that it sends $N = (1,0)$ to $N' = (1,1)$. As in the case of the hyperbolas $\mathcal{H}$ and $\mathcal{H}'$, check that λ is a group homomorphism (with respect to the geometric group law on $\mathcal{C}'$ with neutral element N') and that it is an isomorphism over rings R in which 2 has an inverse, for example in the case $R = \mathbb{Q}$.

2.15 Define a group law on the Eisenstein conic $\mathcal{E} : X^2 + XY + Y^2 = 1$, determine all six integral points on $\mathcal{E}$, and show that $\mathcal{E}(\mathbb{Z}) \simeq \mathbb{Z}/6\mathbb{Z}$. Determine the primes p for which the congruence $x^2 \equiv -3 \bmod p$ is solvable.

2.16 Show that if $(\frac{x}{n}, \frac{y}{n})$ is a rational point on the unit circle with n positive and $\gcd(x, n) = 1$, then $\gcd(y, n) = 1$ as well, and $n \equiv 1 \bmod 4$. Show more exactly that every prime divisor $p > 0$ of n has the form $p \equiv 1 \bmod 4$.

2.17 Show that Pythagorean triples of the form $(m, m+1, n)$ are in bijection with integral points on the conic $\mathcal{C} : X^2 - 2Y^2 = -1$. Hint: multiply $m^2 + (m+1)^2 = n^2$ through by 2 and complete the square.

2.18 Let $N \equiv 1 \bmod 4$ be a positive integer. Show that, in Fermat's factorization method, it is sufficient to check whether any of the numbers $X^2 - N$ is a square with X odd.

2.19 Let $n \equiv 4 \bmod 9$ be an integer. Show that n cannot be written as a sum of four cubes.

2.20 Compute the addition and multiplication tables for the ring $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$, and compare the result to those for $\mathbb{Z}/4\mathbb{Z}$.

2.21 Do the same exercise for the rings $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/3\mathbb{Z}$ and $\mathbb{Z}/6\mathbb{Z}$.

2.22 Find all points on the unit circle $X^2 + Y^2 = 1$ over $\mathbb{F}_3$ and $\mathbb{F}_5$.

2.23 Let $P = (2, 2)$ be a point on the unit circle over $\mathbb{F}_7$. Compute $2P$, $3P$, $\ldots$; what is the order of P?

2.24 Consider the isomorphism $\exp : \mathbb{R} \longrightarrow \mathbb{R}_{>0}$ between the additive group of the reals and the multiplicative group of the positive reals. When you restrict the exponential function to the subgroup $\mathbb{Q}$ of $\mathbb{R}$, do you get an isomorphism between $\mathbb{Q}$ and the multiplicative subgroup of $\mathbb{R}$ consisting of positive rational numbers?

2.25 Show that subgroups of finite cyclic groups are cyclic.

2.26 The group of rational points on the unit circle is actually a topological group in that the addition of points is compatible with the Euclidean metric defined by

$$d(P, Q) = \sqrt{(x_P - x_Q)^2 + (y_P - y_Q)^2}.$$

In fact, show that the following statement is true:
Given a pair of rational points $P, Q \in \mathcal{C}(\mathbb{Q})$ of the unit circle $\mathcal{C} : X^2 + Y^2 = 1$ and some $\varepsilon > 0$, there exists a $\delta > 0$ such that $|d(P, Q) - d(P', Q')| < \varepsilon$ whenever $d(P, P') < \delta$ and $d(Q, Q') < \delta$.

2.27 (Stern [Ste18??, p. 391]) Stern's proof of Wilson's Theorem: Recall that

$$-\log(1 - x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \ldots$$

and deduce that

$$e^x e^{x^2/2} e^{x^3/3} \cdots = 1 + x + x^2 + x^3 + \ldots.$$

Develop the left hand side into a product of power series and compare the coefficient of x^p on both sides.
The coefficient of x^p on the left side is a sum of fractions, only two of which have a denominator divisible by p. Deduce that $\frac{1}{p!} + s + \frac{1}{p} = 1$, where s is a fraction whose denominator is not divisible by p. Finally show how to derive Wilson's Theorem from this observation.

2.28 Let m be a squarefree integer. Show that there are infinitely many prime numbers p with $(\frac{m}{p}) = +1$.
Hint. Let p_1, $\ldots$, p_n be primes (necessarily coprime to $2m$) with $(\frac{m}{p}) = -1$, and consider $N = (2^r p_1 \cdots p_n)^2 - m$, where r is chosen so large that $N > 1$ is not a power of 2. Show that $p_j \nmid N$, and that every prime divisor p of N satisfies $(\frac{m}{p}) = +1$.

2.29 Let m be a squarefree integer. Show that there are infinitely many prime numbers p congruent to a quadratic nonresidue modulo m.

2.30 Define the numbers p_n, d_n by $p_1 = d_1 = 1$ and $p_{k+1} = p_k + d_k$, $d_{k+1} = 2p_k + d_k$.
Show that $p_n^2 - 2d_n^2 = (-1)^n$, and that

$$\Big|\sqrt{2} - \frac{d_{k+1}}{p_{k+1}}\Big| < \frac{\sqrt{2} - 1}{\sqrt{2}} \cdot \Big|\sqrt{2} - \frac{d_k}{p_k}\Big|.$$

Also show that

$$\Big|\sqrt{2} - \frac{d_k}{p_k}\Big| < \frac{1}{2\sqrt{2}p_k^2}.$$

2.31 The unit circle $S^1 : x^2 + y^2 = 1$ and the hyperbola $H^1 : y^2 - x^2 = 1$ can be defined simultaneously by defining bilinear forms on $\mathbb{R}^2$ by

$$\langle p, q\rangle = p^t Dq,$$

where $p, q \in \mathbb{R}^2$, p^t is the transpose of p, and

$$D = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) \quad \text{for } S^1, \qquad \text{and} \qquad D = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right) \quad \text{for } H^1.$$

Show that S^1 and H^1 are the points described by the equation $p^T Dp = +1$.
Matrices $S \in \mathrm{GL}_2(\mathbb{R})$ act on $\mathbb{R}^2$; show that those leaving S^1 and H^1 invariant satisfy $A^T DA = D$.

2.32 Studnička [Stu18??] defined the hyperbolic analogue ϖ of π by $\sinh \frac{\varpi}{2} = 1$ and $\cosh \frac{\varpi}{2} = \sqrt{2}$.

1. Show that $\varpi = \log(1 + \sqrt{2}\,)$.
2. Show that

$$\frac{\pi}{2} = 1 + \frac{1}{3!} + \frac{3^2}{5!} + \frac{3^2 \cdot 5^2}{7!} + \ldots \quad \text{and} \quad \frac{\varpi}{2} = 1 - \frac{1}{3!} + \frac{3^2}{5!} - \frac{3^2 \cdot 5^2}{7!} \pm \ldots.$$

3. Show that

$$\exp\Big(\frac{n\varpi}{2}\Big) = (1 + \sqrt{2}\,)^n,$$

4. Deduce that the integers T_n, U_n defined by $\sinh\left(\frac{n\varpi}{2}\right) + \sqrt{2}\cosh\left(\frac{n\varpi}{2}\right) = T_n + U_n\sqrt{2}$ form the Pell sequence of solutions of $T^2 - 2U^2 = \pm 1$.

2.33 Show that the projection of the parabola $Y = X^2$ onto the X-axis is an isomorphism of abelian groups $\mathcal{P}(R) \longrightarrow (R, +)$ for arbitrary rings R.
Show that the projection onto the Y-axis is not a homomorphism.

2.34 Show that the projection of the real unit circle $X^2 + Y^2 = 1$ onto the tangent $X = 1$ at $(1, 0)$ (with a point ∞ at infinity added) defined by

$$(x, y) \longmapsto \begin{cases} (1, \frac{y}{x}) & \text{if } x \neq 0, \\ \infty & \text{if } x = 0, \end{cases}$$

is a group homomorphism.
Hint: Show that the addition formulas for the sine and cosine imply that

$$\tan\alpha + \tan\beta = \frac{\tan\alpha + \tan\beta}{1 - \tan\alpha \cdot \tan\beta}.$$

Deduce that the composition defined by

$$(r, s) \oplus (t, u) = \begin{cases} \frac{ru+st}{rt-su} & \text{if } rt \neq su, \\ \infty & \text{if } rt = su \end{cases}$$

and

$$(r, s) \oplus \infty = (-s, r)$$

induces a group structure on the line $\overline{L} = L \cup \{\infty\}$.

2.35 (See [Eri2011, p. 2; p. 151]) Consider the hyperbola $\mathcal{H}' : X^2 - Y^2 = 1$ and the lemniscate $\mathcal{L} : (X^2 + Y^2)^2 = X^2 - Y^2$. Show that the map

$$\lambda : (x, y) \longmapsto \Big(\frac{x}{x^2 + y^2}, \frac{y}{x^2 + y^2}\Big)$$

is a bijective map $\lambda : \mathcal{H}'(\mathbb{Q}) \longrightarrow \mathcal{L}(\mathbb{Q}) \setminus \{(0, 0)\}$.
Also show that

$$x = \frac{t^3 + t}{1 + t^4}, \quad y = \frac{t^3 - t}{1 + t^4}$$

is a rational parametrization of $\mathcal{L}$ induced by the map $\mu : \mathcal{H}(\mathbb{Q}) \longrightarrow \mathcal{L}(\mathbb{Q}); \left(t, \frac{1}{t}\right) \to (x, y)$.

Show that μ induces a group law on $\mathcal{L}(\mathbb{Q}) \setminus \{(0,0)\}$, and that

$$(x_1, y_1) \oplus (x_2, y_2) = (x_3, y_3), \quad \text{where} \quad x_3 = \frac{x_1^3 x_2^3 + x_1 x_2}{x_1^4 x_2^4 + 1}, \quad y_3 = \frac{x_1^3 x_2^3 - x_1 x_2}{x_1^4 x_2^4 + 1}.$$

Also show that $-(x, y) = (x, -y)$.
Since $\frac{x-y}{2} = \frac{t}{1+t^4}$ and $\frac{x+y}{x-y} = t^2$ we find

$$t = \frac{x-y}{2} \cdot \Big(1 + \Big(\frac{x+y}{x-y}\Big)^2\Big) = \frac{x-y}{2} + \frac{(x+y)^2}{2(x-y)}.$$

Show that

$$\#\mathcal{L}(\mathbb{Z}/p\mathbb{Z}) = \begin{cases} p & \text{if } p \equiv 3, 5, 7 \bmod 8, \\ p-4 & \text{if } p \equiv 1 \bmod 8. \end{cases}$$

The following tables give the points on $\mathcal{H}(\mathbb{Z}/p\mathbb{Z})$, $\mathcal{H}'(\mathbb{Z}/p\mathbb{Z})$ and $\mathcal{L}(\mathbb{Z}/p\mathbb{Z}) \setminus \{(0,0)\}$ for $p = 5$, $p = 7$ and $p = 17$.

$\mathcal{H}(\mathbb{Z}/5\mathbb{Z})$	$\mathcal{H}'(\mathbb{Z}/5\mathbb{Z})$	$\mathcal{L}(\mathbb{Z}/5\mathbb{Z})$
$(1,1)$	$(1,0)$	$(1,0)$
$(2,3)$	$(0,2)$	$(0,3)$
$(3,2)$	$(0,3)$	$(0,2)$
$(4,4)$	$(4,0)$	$(4,0)$

$\mathcal{H}(\mathbb{Z}/7\mathbb{Z})$	$\mathcal{H}'(\mathbb{Z}/7\mathbb{Z})$	$\mathcal{L}(\mathbb{Z}/7\mathbb{Z})$
$(1,1)$	$(1,0)$	$(4,0)$
$(2,4)$	$(3,6)$	$(1,2)$
$(3,5)$	$(4,6)$	$(6,2)$
$(4,2)$	$(3,1)$	$(1,5)$
$(5,3)$	$(4,1)$	$(6,5)$
$(6,6)$	$(6,0)$	$(6,0)$

$\mathcal{H}(\mathbb{Z}/17\mathbb{Z})$	$\mathcal{H}'(\mathbb{Z}/17\mathbb{Z})$	$\mathcal{L}(\mathbb{Z}/17\mathbb{Z})$
$(1,1)$	$(1,0)$	$(1,0)$
$(2,9)$	$(14,5)$	
$(3,6)$	$(13,7)$	$(7,9)$
$(4,13)$	$(0,4)$	$(0,13)$
$(5,7)$	$(6,16)$	$(2,11)$
$(6,3)$	$(13,10)$	$(7,8)$
$(7,5)$	$(6,1)$	$(2,6)$
$(8,15)$	$(3,5)$	
$(9,2)$	$(14,12)$	
$(10,12)$	$(11,16)$	$(15,11)$
$(11,14)$	$(4,7)$	$(10,9)$
$(12,10)$	$(11,1)$	$(15,6)$
$(13,4)$	$(0,13)$	$(0,4)$
$(14,11)$	$(4,10)$	$(10,8)$
$(15,8)$	$(3,12)$	
$(16,16)$	$(16,0)$	$(16,0)$

2.36 Triangular numbers are numbers of the form $T_n = \frac{n(n+1)}{2}$. Show that there are infinitely many triangular numbers that are squares. Hint: multiply $T_n = m^2$ through by 8 and complete the square. Then reduce the problem to solving the Pythagorean Pell equation.

2.37 (Burnside [Bur1891]) Let P_1, P_2, P_3, P_4 be points on the hyperbola $\mathcal{H} : XY = 1$, and assume that P_1P_2 is parallel to P_3P_4. Show that the midpoints M_1 of P_1P_2 and M_2 of P_3P_4 lie on a line through the origin.

2.38 Let A, B, C be abelian groups. Show that

1. $0 \longrightarrow B \longrightarrow C$ is exact (at B) if and only if the map $B \longrightarrow C$ is injective.
2. $A \longrightarrow B \longrightarrow 0$ is exact (at B) if and only if the map $A \longrightarrow B$ is surjective.
3. $0 \longrightarrow B \longrightarrow 0$ is exact (at B) if and only if $B = 0$.

2.39 Let A, B, C be abelian groups. Show that

$$\begin{tikzcd} 0 \arrow[r] & A \arrow[r, "\iota"] & B \arrow[r, "\pi"] & C \arrow[r] & 0 \end{tikzcd}$$

is exact if and only if $B/\iota(A) \simeq C$.

2.40 Find exact sequences

$$0 \longrightarrow \mathbb{Z} \longrightarrow \mathbb{Z} \oplus \mathbb{Z}/m\mathbb{Z} \longrightarrow \mathbb{Z}/m\mathbb{Z} \longrightarrow 0$$

and

$$0 \longrightarrow \mathbb{Z} \longrightarrow \mathbb{Z} \longrightarrow \mathbb{Z}/m\mathbb{Z} \longrightarrow 0.$$

2.41 Show that if

$$0 \longrightarrow A \xrightarrow{\iota} B \xrightarrow{\pi} C \longrightarrow 0$$

is an exact sequence of finite abelian groups, then $\#B = \#A \cdot \#C$.

2.42 Let $f : G \longrightarrow H$ be a homomorphism of abelian groups. Show that there is an exact and commutative diagram

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & & \\
0 \arrow[r] & \ker f \arrow[r] \arrow[d] & \ker f \arrow[r] \arrow[d] & 0 \arrow[d] & \\
0 \arrow[r] & \ker f \arrow[r] \arrow[d] & G \arrow[r] \arrow[d] & G/\ker f \arrow[r] \arrow[d] & 0 \\
 & 0 \arrow[r] & \operatorname{im} f \arrow[r] \arrow[d] & \operatorname{im} f \arrow[r] \arrow[d] & 0 \\
 & & 0 & 0 &
\end{tikzcd}$$

Also show that the sequence

$$0 \longrightarrow \operatorname{im} f \longrightarrow H \longrightarrow \operatorname{coker} f \longrightarrow 0$$

is exact, where the **cokernel** $\operatorname{coker} f$ is defined by $\operatorname{coker} f = H/\operatorname{im} f$.

2.43 The Six Blade Knife: Let $\alpha : A \longrightarrow B$ and $\beta : B \longrightarrow C$ be homomorphisms of abelian groups. Show that there is an exact sequence

$$0 \longrightarrow \ker \alpha \longrightarrow \ker \beta \circ \alpha \longrightarrow \ker \beta \longrightarrow \operatorname{coker} \alpha \longrightarrow \operatorname{coker} \beta \circ \alpha \longrightarrow \operatorname{coker} \beta \longrightarrow 0.$$

Bass has observed that this sequence nicely fits into the following diagram:

$$\begin{tikzcd}
 & & \ker \beta \arrow[rr] \arrow[dr, tail] & & \operatorname{coker} \alpha \arrow[ddrr] & & \\
 & & & B \arrow[ur, two heads] \arrow[dr] & & & \\
\ker \beta \circ \alpha \arrow[uurr] \arrow[rr, hook] & & A \arrow[ur] \arrow[rr] & & C \arrow[rr, two heads] \arrow[dr, two heads] & & \operatorname{coker} \beta \circ \alpha \arrow[dl] \\
 & \ker \alpha \arrow[ul] \arrow[ur, tail] & & 0 \arrow[ll] & & \operatorname{coker} \beta \arrow[ll] &
\end{tikzcd}$$