[1]

ELLIPTIC CURVES MT 2025/26: PRELIMINARY READING

Section 0. Background Material in Algebra, Number Theory and Geometry

The following gives a summary of the main ideas you need to know as prerequisites to the lecture course on Elliptic Curves. Most of you should have seen most of this material before in lecture courses from previous years, but it is just as well to read through it carefully, in order to fill in any gaps.

Groups

Definition 0.1. A *group* is a set G with a binary operation $*$ which satisfies the following properties.
Closure: If $f, g \in G$ then $f * g \in G$.
Associativity: For all $f, g, h \in G$, $(f * g) * h = f * (g * h)$.
Existence of identity: There exists $e \in G$ such that, for all $g \in G$, $e * g = g * e = g$.
Existence of inverses: For all $g \in G$, there exists $h \in G$ such that $g * h = h * g = e$.

Comment 0.2. The element h is the *inverse* of g, and is typically denoted g^{-1}, when referring to a general group $G, *$, and any specific group whose operation is some type of multiplication. On the other hand, the inverse of g will typically be denoted $-g$ when dealing with a specific group whose operation is some form of addition.

Definition 0.3. We say that a group G is a *commutative* (or *Abelian*) group if it also satisfies
Commutativity: For all $f, g \in G$, $f * g = g * f$.

Examples 0.4.
(a) $\mathbb{Z}, +$ is an Abelian group (identity 0).
(b) $\mathbb{Z}, \times$ has identity $= 1$ but, for example, 2 has no inverse, and so this is not a group.
(c) $\mathbb{R}^+, \times$ (the positive real numbers under multiplication) is an Abelian group with identity 1.
(d) $\mathbb{R} \times \mathbb{R}, +$ [which means all pairs (a, b), with operation $(a_1, b_1) + (a_2, b_2) = (a_1 + a_2, b_1 + b_2)$] is an Abelian group with identity $(0, 0)$.
(e) $\{2 \times 2$ matrices with nonzero determinant$\}$ under matrix multiplication is a group. Identity $= \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$.

[1] These notes were written by Victor Flynn, with minor edits by James Newton.

(f) $C_6, +$ [the cyclic group of order 6], denoting $\{0, 1, 2, 3, 4, 5\}$ under $+$ modulo 6 [e.g. $3 + 4 = 1$]. This is an Abelian group with identity 0.
(g) $C_2 \times C_3, +$, which is $\{(0, 0), (0, 1), (0, 2), (1, 0), (1, 1), (1, 2)\}$ under the operation: $(a_1, b_1) + (a_2, b_2) = (a_1 + a_2 \bmod 2, b_1 + b_2 \bmod 3)$. This is an Abelian group with identity $(0, 0)$.
(h) Let $S_3, \circ$ be the set of permutations of $\{1, 2, 3\}$, with: $f \circ g$ = 'g-followed-by-f' as our operation [we shall normally abbreviate $f \circ g$ as fg]. This is a group and the elements are: $\{e, (12), (13), (23), (123), (132)\}$ [where, for example, (132) represents the permutation: $1 \to 3, 3 \to 2, 2 \to 1$, and (23) represents $2 \to 3, 3 \to 2$ (with $1 \to 1$)]. This is not an Abelian group since, for example, $(132)(12) = (23)$, but $(12)(132) = (13)$.

Definition 0.5. Let $G_1, *_1$ and $G_2, *_2$ be groups, and let $\phi : G_1 \to G_2$ [a map from G_1 to G_2]. We say that ϕ is a *homomorphism* if, for all $g, h \in G$, $\phi(g *_1 h) = \phi(g) *_2 \phi(h)$.

An *endomorphism* on a group G is a homomorphism from G to itself.

Examples 0.6.
(a) $\log : \mathbb{R}^+, \times \to \mathbb{R}, +$ is a homomorphism since, for all $a, b \in \mathbb{R}^+$, $\log(a \times b) = \log(a) + \log(b)$ [that is, $\log(a *_1 b) = \log(a) *_2 \log(b)$].
(b) $\phi : \mathbb{R} \times \mathbb{R}, + \to \mathbb{R}, +$ defined by $\phi\big((a, b)\big) = a$ [can also express this as $\phi : (a, b) \mapsto a$] is a homomorphism.
Proof. $\phi\big((a, b) *_1 (c, d)\big) = \phi\big((a, b) + (c, d)\big) = \phi\big((a + c, b + d)\big) = a + c$.

Also, $\phi\big((a, b)\big) *_2 \phi\big((c, d)\big) = \phi\big((a, b)\big) + \phi\big((c, d)\big) = a + c$, and these are the same.
(c) $\phi : \mathbb{Z}, + \to \mathbb{Z}, +$, defined by $\phi(a) = 2a$ is a homomorphism.
(d) $\phi : \mathbb{Z}, + \to \mathbb{Z}, + : a \mapsto a^2$ is not a homomorphism since, for example, $\phi(2 + 3) = \phi(5) = 5^2 = 25$, but $\phi(2) + \phi(3) = 2^2 + 3^2 = 13$, and these are not equal.

Definition 0.7. Let $\phi : S \to T$, for any sets S, T. We say that ϕ is *injective* (or 1–1 or an *injection*) if, for all $f, g \in S$, $\phi(f) = \phi(g) \implies f = g$; that is, $f \neq g \implies \phi(f) \neq \phi(g)$ [i.e. when it never happens that two distinct f and g are mapped by ϕ to the same element]. We say that ϕ is *surjective* (or *onto* or a *surjection*) if, for all $w \in T$, there exists $g \in S$ such that $w = \phi(g)$ [i.e. when every member of T is mapped onto by at least one element of S]. We say that ϕ is *bijective* (or a *bijection*) if it is both injective and surjective.

Definition 0.8. Let $\phi : G_1, *_1 \to G_2, *_2$ be a homomorphism. The *kernel* of ϕ (denoted ker ϕ) is defined as the set of all members of G_1 which are mapped to the identity element e_2 in G_2. That is: $\ker \phi = \{g \in G_1 : \phi(g) = e_2\}$. The image of ϕ (denoted im ϕ) is the set of all members of G_2 which are mapped onto by some member of G_1. That is to say: $\operatorname{im} \phi = \{\phi(g) : g \in G_1\}$.

Comment 0.9. Clearly, a homomorphism $\phi : G_1, *_1 \to G_2, *_2$ is injective if and only if $\ker \phi = \{e_1\}$, where e_1 is the identity element in G_1. It is surjective if and only if $\mathrm{im}\ \phi = G_2$.

Examples 0.10.
(a) $\log : \mathbb{R}^+, \times \to \mathbb{R}, +$ is an injection since, for any $f, g \in \mathbb{R}^+$: $\phi(f) = \phi(g) \implies \log f = \log g \implies e^{\log f} = e^{\log g} \implies f = g$.

It is also a surjection since, if $w \in \mathbb{R}$, we can take $g = e^w \in \mathbb{R}^+$ and $\phi(g) = \log(e^w) = w$. Hence ϕ is a bijection, since it is both an injection and a surjection. The kernel is $\{1\}$ [that is, 1 is the unique member of $\mathbb{R}^+, \times$ mapped by log to the identity element 0 in $\mathbb{R}, +$]. The image is all of $\mathbb{R}$ [since the map is surjective].
(b) Let $\phi : \mathbb{R} \times \mathbb{R}, + \to \mathbb{R}, +$ be defined by $\phi\big((a, b)\big) = a$. This is not an injection since, for example, $\phi\big((2, 1)\big) = 2$ and $\phi\big((2, 3)\big) = 2$, but $(2, 1) \neq (2, 3)$. It is a surjection since, for any $r \in \mathbb{R}$, we can take $(r, 0) \in \mathbb{R} \times \mathbb{R}$ which satisfies $\phi\big((r, 0)\big) = r$ [of course, we could just as easily have used $(r, 1)$; we merely had to show that every $r \in \mathbb{R}$ is mapped onto by at least one member of of $\mathbb{R} \times \mathbb{R}$]. The kernel is $\{(0, b) : b \in \mathbb{R}\}$ and the image is all of $\mathbb{R}$ [since ϕ is surjective].
(c) $\phi : \mathbb{Z}, + \to \mathbb{Z}, +, a \mapsto 2a$. This is an injection since, for any $a, b \in \mathbb{Z}$: $\phi(a) = \phi(b) \implies 2a = 2b \implies a = b$. It is not a surjection since nothing maps to 3 (for example). The kernel is $\{0\}$ and the image is $\{\ldots, -4, -2, 0, 2, 4, \ldots\}$.

Definition 0.11. Let $G_1, *_1$ and $G_2, *_2$ be groups and let $\phi : G_1 \to G_2$. If ϕ is both a bijection and a homomorphism, then we say that ϕ is an *isomorphism.* If there exists an isomorphism $\phi : G_1 \to G_2$, we say that the two groups are *isomorphic* (same shape) and we write $G_1 \cong G_2$.

Comment 0.12. If G_1 and G_2 are isomorphic groups, then G_2 can be regarded as the same group as G_1, merely with the elements relabelled. G_1 and G_2 will have all of the same structural properties (for example, G_1 will be Abelian iff G_2 is Abelian, G_1 will have an element $g \neq e$ satisfying $g * g = e$ iff G_2 has such an element, etc).

Example 0.13. $\log : \mathbb{R}^+, \times \to \mathbb{R}, +$ is an isomorphism, since it is both a homomorphism and a bijection. The groups $\mathbb{R}^+, \times$ and $\mathbb{R}, +$ are isomorphic.

Comment 0.14. Two finite groups G_1, G_2 are isomorphic if the group table of G_1 can have its elements relabelled to give the group table of G_2.

Example 0.15. Let $G_1 = C_2 \times C_3$ and $G_2 = C_6$. Let $\phi : G_1 \to G_2$ be defined by:

$(0, 0) \mapsto 0,\ (1, 1) \mapsto 1,\ (0, 2) \mapsto 2,\ (1, 0) \mapsto 3,\ (0, 1) \mapsto 4,\ (1, 2) \mapsto 5.$

The group table of G_1 is as follows.

+	(0,0)	(0,1)	(0,2)	(1,0)	(1,1)	(1,2)
(0,0)	(0,0)	(0,1)	(0,2)	(1,0)	(1,1)	(1,2)
(0,1)	(0,1)	(0,2)	(0,0)	(1,1)	(1,2)	(1,0)
(0,2)	(0,2)	(0,0)	(0,1)	(1,2)	(1,0)	(1,1)
(1,0)	(1,0)	(1,1)	(1,2)	(0,0)	(0,1)	(0,2)
(1,1)	(1,1)	(1,2)	(1,0)	(0,1)	(0,2)	(0,0)
(1,2)	(1,2)	(1,1)	(1,1)	(0,2)	(0,0)	(0,1)

Replacing all entries using ϕ gives the following table.

+	0	4	2	3	1	5
0	0	4	2	3	1	5
4	4	2	0	1	5	3
2	2	0	4	5	3	1
3	3	1	5	0	4	2
1	1	5	3	4	2	0
5	5	3	1	2	0	4

This is just the group table for C_6, which proves that $C_2 \times C_3 \cong C_6$.

The last example is a special case of the following result.

Lemma 0.16. *When $m, n \in \mathbb{Z}$ and m, n have no common factors (apart from 1) then $C_m \times C_n \cong C_{mn}$.*

The following is also quite a useful property of finite Abelian groups.

Lemma 0.17. *Any finite Abelian group G is isomorphic to the product of cyclic groups: $G \cong C_{m_1} \times C_{m_2} \times \ldots \times C_{m_k}$, for some $C_{m_1}, \ldots, C_{m_k}$.*

For any group G, it is natural to consider groups which lie inside G (that is to say, which are subsets of G).

Definition 0.18. Let $G, *$ be a group and let $H \subset G$ [H is a subset of G]. We say that H is a *subgroup* of G (written: $H \leqslant G$) if H is nonempty, and forms a groups with respect to the same operation $*$ as G. This is equivalent to:

$e_G \in H$ (where e_G is the identity element in G),

If $f, g \in H$ then $f * g \in H$,

If $h \in H$ then $h^{-1} \in H$. Note that associativity automatically holds in H since it holds in the group G, of which H is a subset.

Examples 0.19.

(a) $H = \{\ldots, -4, -2, 0, 2, 4, \ldots\} \leqslant \mathbb{Z}, +$.

(b) $H = \{\ldots, -3, -1, 1, 3, \ldots\} \not\leqslant \mathbb{Z}, +$, since the identity element 0 is not in the set (we could alternatively have used the fact that it is not closed; for example, $1, 3 \in H$ but $1 + 3 \notin H$).

(c) $H = \{0, 1, 2, 3, \ldots\} \not\leqslant \mathbb{Z}, +$. It is fine for containing the identity element and closure, but H does not contain the inverse of every element in H (for example, $3 \in H$ but $-3 \notin H$).

Definition 0.20. Let $H \leqslant G$ and let $g \in G$. The set $gH = \{g * h : h \in H\}$ is called a *left coset* of H and the set $Hg = \{h * g : h \in H\}$ is called a *right coset* of H.

Comment 0.21. When the group operation is some form of multiplication, one typically writes the left (or right) cosets, as above, in the style gH (or Hg). When the group operation is some form of addition, then one typically writes $g + H = \{g + h : h \in H\}$ (similarly for $H + g$).

Example 0.22. Let $G = \mathbb{Z}, +$ and let $H = 3\mathbb{Z} = \{\ldots, -6, -3, 0, 3, 6, \ldots\} \leqslant G$. Then some examples of left cosets are:

$0 + H = \{\ldots, 0 + (-6), 0 + (-3), 0 + 0, 0 + 3, 0 + 6, \ldots\} = \{\ldots, -6, -3, 0, 3, 6, \ldots\}$,
$1 + H = \{\ldots, 1 + (-6), 1 + (-3), 1 + 0, 1 + 3, 1 + 6, \ldots\} = \{\ldots, -5, -2, 1, 4, 7, \ldots\}$,
$2 + H = \{\ldots, 2 + (-6), 2 + (-3), 2 + 0, 2 + 3, 2 + 6, \ldots\} = \{\ldots, -4, -1, 2, 5, 8, \ldots\}$,
$3 + H = \{\ldots, 3 + (-6), 3 + (-3), 3 + 0, 3 + 3, 3 + 6, \ldots\} = \{\ldots, -3, 0, 3, 6, 9, \ldots\}$.
$4 + H = \{\ldots, 4 + (-6), 4 + (-3), 4 + 0, 4 + 3, 4 + 6, \ldots\} = \{\ldots, -2, 1, 4, 7, 10, \ldots\}$.

Note that $0 + H = 3 + H$ and $1 + H = 4 + H$. Clearly

$\ldots -6 + H = -3 + H = 0 + H = 3 + H = 6 + H = \ldots$
$\ldots -5 + H = -2 + H = 1 + H = 4 + H = 7 + H = \ldots$
$\ldots -4 + H = -1 + H = 2 + H = 5 + H = 8 + H = \ldots$

so that there are only 3 distinct left cosets.

The left coset $eH = H$, where e is the identity element, so that H is one of the left cosets of itself (and similarly is one of the right cosets of itself). It can be shown two left cosets g_1H and g_2H are either equal or disjoint and that every element of G is a member of some coset. When G is a finite group, it can also be shown that any g_1H and g_2H have the same number of elements (and so every left coset of H has the same number of elements as H). It follows that the left cosets of H give a partition of G, that is, they give G as a union of disjoint subsets. Since each of these subsets has the same number of elements as H, we see that $|G| = |H| + \ldots + |H| = k|H|$, where k is the number of distinct left cosets of H [here, $|S|$ is the standard notation for the number of elements in S, for any set S]. The following immediately follows.

Theorem 0.23. *(Lagrange's Theorem) Let G be a finite group, and let $H \leqslant G$. Then $|H|$ is a factor of $|G|$ [this can also be expressed as $|H|$ divides $|G|$, or as $|H| \mid |G|$].*

There are many situations where we would like to consider the elements of a group G, but in a simplified context, where we 'mod out' (or 'quotient out') by a subgroup, and focus on the information that remains. For example, when $G = \mathbb{Z}, +$, we might want to collapse $H = 3\mathbb{Z} \leqslant G$ down to a single element, and consider the elements mod H (considering elements to be the same if they lie in the same coset). The natural way to do this is to create a new group G/H, whose elements are (say) the left cosets of H, in which case there are only 3 distinct elements:

$$\{\ldots, -6, -3, 0, 3, 6, \ldots\}, \{\ldots, -5, -2, 1, 4, 7, \ldots\}, \{\ldots, -4, -1, 2, 5, 8, \ldots\},$$

which give all the members of G/H. It is natural to ask whether the group law on G carries over to give group law on G/H. How might we add, for example, the second and third of these? That is, we want to perform the addition:

$$\{\ldots, -5, -2, 1, 4, 7, \ldots\} + \{\ldots, -4, -1, 2, 5, 8, \ldots\}.$$

A natural attempt is add any element in the first coset to any element in the second coset, and see what coset the sum lies in. For example, -5 is in the first coset, and 2 is the second coset, and $-5 + 2 = -3$, which lies in: $\{\ldots, -6, -3, 0, 3, 6, \ldots\}$, suggesting that, in G/H:

$$\{\ldots, -5, -2, 1, 4, 7, \ldots\} + \{\ldots, -4, -1, 2, 5, 8, \ldots\} = \{\ldots, -6, -3, 0, 3, 6, \ldots\}.$$

Furthermore, it doesn't matter what members you take: you can add any member of $\{\ldots, -5, -2, 1, 4, 7, \ldots\}$ to any member of $\{\ldots, -4, -1, 2, 5, 8, \ldots\}$ and you will get a member of $\{\ldots, -6, -3, 0, 3, 6, \ldots\}$, reinforcing our confidence in this definition of the sum. It is easy to see that this gives a way of turning the 3 members of G/H into a group. We can also express this group law on G/H as: $(g_1 + H) + (g_2 + H) = (g_1 + g_2) + H$, where the well-definedness of this rule is due to the fact that, at least for this choice of G, H, whenever $g_1 + H = g_1' + H$ and $g_2 + H = g_2' + H$ then $(g_1 + g_2) + H = (g_1' + g_2') + H$. Even though the members of G/H are sets, it is often convenient to denote them by selected representative elements; for example, we can use $0, 1, 2$ to denote the cosets containing $0, 1, 2$, respectively, in which case that above addition could be expressed as: $1 + 2 = 0$ in G/H. Of course, 91 lies in the same coset as 1, so that $1 = 91$ in G/H; we could just as easily represent our 3 members of G/H as $0, 91, 2$ and say that $91 + 2 = 0$ in G/H.

Similarly, let $G = \mathbb{C}^*, \times$, the group of nonzero complex numbers under multiplication, and let $H = \{z : |z| = 1\} \leqslant G$, the unit circle on an Argand diagram. Then an example of a left coset is $(3 + 4i)H = \{(3 + 4i)z : |z| = 1\}$, which is easily seen to be just the circle, centre 0, with radius 5 (the modulus of $3 + 4i$). Note that the group operation is multiplication here, so the cosets are written as $gH = \{g * h : h \in H\} = \{gh : h \in H\}$ [rather than $g + H = \{g * h : h \in H\} = \{g + h : h \in H\}$, as in the previous example]. Two complex numbers are in the same coset iff they have the same modulus. Clearly, the left cosets are just the circles with centre 0, and these are the elements of G/H. We have 'modded out' by H, removing the argument information, and retaining only the modulus information. We can turn G/H into a group under multiplication: for example, the set of complex numbers of modulus 5 multiplied by the set of complex numbers of modulus 2 gives the set of complex numbers modulus 10. This is well defined, since it does not matter which representative is taken: any member of the first coset (any complex number of modulus 5) times any member of the second coset (any complex number of modulus 2) will give a member of the third coset (a complex number of modulus 10).

By way of contrast, let G be as in Example 0.4(h), that is, $G = S_3, \circ$, the group of permutations of $\{1, 2, 3\}$ under the operation $f \circ g =$ 'g-followed-by-f' [where, as usual, we shall abbreviate $f \circ g$ as fg]. Consider $H = \{e, (12)\} \leqslant G$. There are only 3 distinct left cosets of H:

$eH = (12)H = \{e, (12)\}$,
$(123)H = (13)H = \{(123), (13)\}$,
$(132)H = (23)H = \{(132), (23)\}$.

How might we try to perform: $\{e, (12)\}\{(123), (13)\}$? We could attempt the same approach as before: take any element from each set, combine them according to the group law on G and see what coset the results lies in. For example, e is a member of $\{e, (12)\}$ and (123) is a member of $\{(123), (13)\}$ and $e(123) = (123) \in \{(123), (13)\}$. So we might be tempted to say that $\{e, (12)\}\{(123), (13)\} = \{(123), (13)\}$. On the other hand, $(12) \in \{e, (12)\}$ and $(13) \in \{(123), (13)\}$, and $(12)(13) = (132) \in \{(132), (23)\}$, so this suggests that $\{e, (12)\}\{(123), (13)\} = \{(132), (23)\}$. We see that there is no sensible unambiguous way of defining $\{e, (12)\}\{(123), (13)\}$. To put it another way, our attempt to use the natural rule $(g_1H)(g_2H) = (g_1g_2)H$ to give a group law on G/H, has foundered on the fact that there are instances where $g_1H = g_1'H$ and $g_2H = g_2'H$, but $(g_1g_2)H \neq (g_1'g_2')H$ [for example, when $g_1 = e, g_1' = (12), g_2 = (123), g_2' = (13)$]. Any attempt to turn the set of right cosets into a group would also suffer the same problem. Note that if we keep the group $G = S_3$, as before, but use instead $H = \{e, (123), (132)\} \leqslant G$, then it is easy to check that everything is fine, and we can turn G/H into a group.

The key property which allows G/H to be a group is the following.

Definition 0.24. Let $G, *$ be a group and let $H \leqslant G$. We say that H is a *normal* subgroup of G, denoted $H \lhd G$ if, for every $g \in G$, $gH = Hg$.

An equivalent definition is: $\forall g \in G,\ \forall h \in H,\ \ g^{-1}hg \in H$.

Comment 0.25. When $H \lhd G$, the left cosets of H are the same as the right cosets, and so we can just refer to them as *cosets*, without needing to specify left or right.

Definition 0.26. Let $G, *$ be a group and let $H \lhd G$. Then G/H (or 'G quotient H' or 'G mod H') is defined as $G/H = \{gH : g \in G\}$, under the group operation: $(g_1H)(g_2H) = (g_1g_2)H$ [here, we are writing g_1g_2, g_1H, g_2H as shorthand for $g_1 * g_2, g_1 * H, g_2 * H$].

Why is it that the condition $H \lhd G$ is sufficient for this group operation on G/H to be well defined? Recall, the guarantee we need for unambiguity is that, whenever $g_1H = g_1'H$ and $g_2H = g_2'H$, then $(g_1g_2)H = (g_1'g_2')H$. So, suppose that $H \lhd G$ and that $g_1H = g_1'H, g_2H = g_2'H$. Then:

$$\begin{aligned}(g_1g_2)H &= g_1(g_2H) = g_1(g_2'H) = g_1(Hg_2') = (g_1H)g_2' \\ &= (g_1'H)g_2' = (Hg_1')g_2' = H(g_1'g_2') = (g_1'g_2')H, \text{ as required.}\end{aligned}$$

Comment 0.27. If $G, *$ is Abelian then any subgroup H must be normal, guaranteeing that we can always form the quotient group G/H.

Definition 0.28. Let X be any set, and let $\sim$ be a binary relation on X. We say that $\sim$ is an *equivalence relation* if it satisfies:

(1) $a \sim a$ for all $a \in X$ [reflexivity].
(2) $a \sim b \implies b \sim a$ for all $a, b \in X$ [symmetry].
(3) $a \sim b$ and $b \sim c \implies a \sim c$ for all $a, b, c \in X$ [transitivity].

The *equivalence class* of an element $a \in X$, denoted $[a]$, is the set of all members of X which are equivalent to a. This is to say: $[a] = \{x \in X : x \sim a\}$.

Given any $g_1, g_2 \in G$, it is easy to check that $g_1H = g_2H$ exactly when $g_1 = g_2 * h$, for some $h \in H$; that is, when $g_1 * g_2^{-1} \in H$. Define the relation $g_1 \sim g_2$ by:

$$g_1 \sim g_2 \iff g_1 = g_2 * h, \text{ for some } h \in H,$$

which gives an equivalence relation on G. Another way to describe members of G/H is to say that they are equivalence classes under this relation (or, we can also say that they are the members of G *modulo* the equivalence relation).

Comment 0.29. It can sometimes seem cumbersome to deal directly with the above definition of G/H, since the group elements in G/H are cosets (so that G/H is a set of sets). Suppose nobody had ever mentioned cosets. There is a more intuitive approach to quotient groups (which is in fact the way they are mostly dealt with in practice) which requires no explicit mention of cosets. Namely, one writes the elements of G/H exactly as the elements of G, except that certain elements become equal in G/H which were distinct in G. Specifically, one imposes the rule:

$$g_1 = g_2 \text{ in } G/H \iff g_1 = g_2 * (\text{some member of } H).$$

Equivalently: $g_1 = g_2$ in $G/H \iff g_1 * g_2^{-1} \in H$. When the operation in G is addition, this means two elements are equal in G/H exactly when their difference is in H. When the operation in G is multiplication, two elements are equal in G/H exactly when their quotient is in H [of course, when the group operation is neither an addition nor a multiplication, then just use the general criterion $g_1 * g_2^{-1} \in H$]. The following examples are described in this spirit, with no explicit mention of cosets.

Examples 0.30.

(a) Let $G = \mathbb{Z}, +$ and $H = 3\mathbb{Z} = \{\ldots, -6, -3, 0, 3, 6, \ldots\} \leqslant G$. We see that, for example, $1 = 16*(-15)$ in G [since $*$ is $+$ here], so that $1 = 16*(\text{member of } H)$, and so $1 = 16$ in G/H. Equivalently, $1 * 16^{-1} = 1 + (-16) = -15 \in H \implies 1 = 16$ in G/H [note that 16^{-1} is the inverse of 16 in G, which is -16]. On the other hand, $1 \neq 20$ in G/H, since $1 = 20 * (-19)$ and $-19 \notin H$.

In the group $G/H = \mathbb{Z}/3\mathbb{Z}$:

$\ldots = -6 = -3 = 0 = 3 = 6 = \ldots$

$\ldots = -5 = -2 = 1 = 4 = 7 = \ldots$

$\ldots = -4 = -1 = 2 = 5 = 8 = \ldots$

and so $\mathbb{Z}/3\mathbb{Z}$ contains only 3 distinct elements. The usual convention is to pick out $0, 1, 2$ as listing the distinct members of $\mathbb{Z}/3\mathbb{Z}$. We can see that $\mathbb{Z}/3\mathbb{Z}, +$ is isomorphic to $C_3, +$.

(b) Let $G = \mathbb{Q}^*, \times$ = nonzero members of $\mathbb{Q}$ under multiplication. Let $H = (\mathbb{Q}^*)^2 = \{\text{squares of nonzero members of } \mathbb{Q}\}$. For example, $4/9 \in H$ but $2 \notin H$.

In $\mathbb{Q}^*$, $2/3 = 6 \times \frac{1}{9}$ and $\frac{1}{9} \in (\mathbb{Q}^*)^2$ so that $2/3 = 6$ in $\mathbb{Q}^*/(\mathbb{Q}^*)^2$. Similarly, $6 = \frac{24}{25}$ in $\mathbb{Q}^*/(\mathbb{Q}^*)^2$ since $6 = \frac{24}{25} \times \frac{25}{4}$ and $\frac{25}{4} \in (\mathbb{Q}^*)^2$. However, $2 \neq 3$ in $\mathbb{Q}^*/(\mathbb{Q}^*)^2$ since $2 = 3 \times \frac{2}{3}$ and $\frac{2}{3} \notin (\mathbb{Q}^*)^2$.

Note that any $\frac{a}{b} \in \mathbb{Q}^*/(\mathbb{Q}^*)^2$ [where $a, b \in \mathbb{Z}$] can be written as $\frac{a}{b} = \frac{a}{b}b^2 = ab \in \mathbb{Z}$. We can write any integer in the form rs^2 where $r, s \in \mathbb{Z}$ and r is square-free [where *square free* means not divisible by any integer square except 1; for example, 6 is square free, but 12 is not square free, since it is divisible by 4]. Write the integer ab in the form rs^2, so that $\frac{a}{b} = ab = rs^2 = r$ in $\mathbb{Q}^*/(\mathbb{Q}^*)^2$. The standard way of working in $\mathbb{Q}^*/(\mathbb{Q}^*)^2$ is to write each distinct element as a square free integer. For example:

$\frac{20}{13} = \left(\frac{20}{13}\right)13^2 = 20 \times 13 = 4 \times 5 \times 13 = 5 \times 13 = 65$ in $\mathbb{Q}^*/(\mathbb{Q}^*)^2$,

which is a square free integer.

(c) Let $G = \mathbb{C}^*, \times$ and $H = \{z : |z| = 1\}$. Then $z_1 = z_2$ in $G/H \iff z_1/z_2 \in H \iff |z_1/z_2| = 1 \iff |z_1| = |z_2|$. That is, $z_1 = z_2$ in G/H exactly when they have the same modulus. So, for example, $3 + 4i = 5i = 5$ in G/H. Clearly, every member of G is equal in G/H to precisely one nonzero real number (namely, its modulus). So, each element of G/H

can be represented by a nonzero real numbers, and it is easy to see that G/H is isomorphic $\mathbb{R}^*, \times$.

The idea of a normal subgroup is related to homomorphisms in the following way.

Lemma 0.31. *Let $\phi : G_1, *_1 \to G_2, *_2$ be a homomorphism. Then ker $\phi \lhd G_1$ and im $\phi \leqslant G_2$.*

Since the kernel of a homomorphism is a normal subgroup, we can form the quotient group $G_1/\text{ker } \phi$. The map $g *_1 \text{ker } \phi \mapsto g$ can be shown to be well defined and an isomorphism, giving the following result (often called the First Isomorphism Theorem).

Theorem 0.32. *Let $\phi : G_1, *_1 \to G_2, *_2$ be a homomorphism. Then $G_1/$ker $\phi \cong$ im ϕ.*

Comment 0.33. Note that, in the case when ϕ is surjective, we have im $\phi = G_2$ and so $G_1/\text{ker } \phi \cong G_2$.

Examples 0.34.
(a) Let $\phi : \mathbb{R} \times \mathbb{R} \times \mathbb{R}, + \to \mathbb{R} \times \mathbb{R}, +$ be defined by $\phi\big((x, y, z)\big) = (x, y)$ [the projection map to the (x, y)-plane]. Then ker ϕ is the z-axis $\{(0, 0, z) : z \in \mathbb{R}\}$, and im ϕ is all of $\mathbb{R} \times \mathbb{R}$ (the map is surjective). The isomorphism theorem tells us that $\mathbb{R} \times \mathbb{R} \times \mathbb{R}/\text{ker } \phi \cong \mathbb{R} \times \mathbb{R}$.
(b) Let $\phi : \mathbb{C}^*, \times \to \mathbb{R}^*, \times : z \mapsto |z|$. Then ker $\phi = \{z : |z| = 1\}$ and im ϕ is all of $\mathbb{R}$ (the map is surjective). The isomorphism theorem tells us that $\mathbb{C}^*/\text{ker } \phi \cong \mathbb{R}^*$.

Another important idea is that of the order of an element.

Definition 0.35. Let $G, *$ be a group and $g \in G$. If there exists $k > 0$ such that $g * g * \ldots * g$ [k times] $= e$ then we say that g has *finite order* (or is a *torsion* element), and the smallest such k is the *order* of g, denoted o(g). If no such k exists, we say that g has *infinite order*. For an Abelian group G, the set of all elements in G of finite order is a subgroup of G, the *torsion subgroup* of G, denoted G_{tors}.

Since $\{e, g, g^2, \ldots, g^{o(g)-1}\}$ is a subgroup of G [the *subgroup generated by* g] with o(g) elements, we obtain the following consequence of Lagrange's Theorem.

Corollary 0.36. *Let $G, *$ be a group and $g \in G$. The order of g is always a factor of $|G|$. As a consequence, $g^{|G|} = e$.*

Definition 0.37. We say that $G, *$ is *Boolean* if, for all $g \in G$, $g * g = e$ [and so every element apart from the identity will have order 2].

Comment 0.38. Any finite Boolean group G is isomorphic to the product of a finite number of copies of C_2; that is: $G \cong C_2 \times C_2 \times \ldots \times C_2$. It follows that the order of G [that is, the number of elements in G] is a power of 2.

Definition 0.39. Let $G, *$ be an Abelian group. The *m-torsion subgroup* of G, denote by $G[m]$, is defined as $\{g \in G : g * g * \ldots * g \text{ [}m\text{ times]} = e\}$. This is same as the set of members of G whose orders are factors of m.

Comment 0.40. When G is an Abelian group, let $2G$ denote the subgroup $\{g * g : g \in G\}$. Clearly $G/2G$ is always a Boolean group When G is a finite Abelian group, it can be shown that $G/2G \cong G[2]$.

Elementary Number Theory

We have already seen the idea of the 'integers modulo m' developed as a quotient group in Example 0.30(a). The next few definitions rephrase this idea in the language of congruences (which we have already used in Examples 0.4(f),(g), but which we now formalise). First a few preliminaries are necessary.

Definition 0.41. For any $a, b \in \mathbb{Z}$, we say that a *divides* b [or that a is a *factor* of b, or that a is a *divisor* of b], denoted $a|b$, if there exists $k \in \mathbb{Z}$ such that $b = ka$. When a does not divide b, this is denoted $a \not| \, b$ [for example, $5|20$, but $7 \not| \, 20$ and $20 \not| \, 5$].

Example 0.42. If $x \in \mathbb{Z}$ is a root of a polynomial $f(x) = f_n x^n + \ldots + f_0$ with integer coefficients, then $x|f_0$ [since, $f(x) = 0$ implies $x(-f_n x^{n-1} - \ldots - f_1) = f_0$]. So, for example, to test whether $x^3 + 11x - 6 = 0$ has any integer solutions, it is only necessary to check the possibilities $x = \pm 1, \pm 2, \pm 3, \pm 6$. Since none of these are solutions, it follows that the equation $x^3 + 11x - 6 = 0$ has no integer solutions.

Definition 0.43. Let $m \in \mathbb{Z}, m > 1$. We say that m is *prime* [or a *prime number*] if its only divisors are 1 and m itself; otherwise m is *composite* [by convention, 1 is neither prime nor composite].

Definition 0.44. For any $m, n \in \mathbb{Z}$, the *greatest common divisor* of m, n, denoted $\gcd(m, n)$, is the largest $d \geqslant 1$ such that $d|m$ and $d|n$ (sometime also called the *highest common factor* of m, n or $\text{hcf}(m, n)$). The *least common multiple* of m, n, denoted $\text{lcm}(m, n)$, is the smallest $D \geqslant 1$ such that $m|D$ and $n|D$. Sometimes $\gcd(m, n)$ is abbreviated as (a, b) and $\text{lcm}(m, n)$ as $[a, b]$. When $\gcd(m, n) = 1$ we say that m and n are *coprime*.

For example, the positive divisors of 12 are: $1, 2, 3, 4, 6, 12$ and the positive divisors of 18 are: $1, 2, 3, 6, 9, 18$. The common divisors are: $1, 2, 3, 6$, the greatest of which is 6, and so $\gcd(12, 18) = 6$.

Note that any common divisor of a and b is also a common divisor of $a + kb$ and b, and vice versa, giving the following property of gcd's.

Lemma 0.45. *For any $a, b, k \in \mathbb{Z}$, $gcd(a + kb, b) = gcd(a, b) = gcd(a, b + ka)$.*

A fundamental property of $\mathbb{N}$ is that, given any $a, b \in \mathbb{N}$, one can find the highest multiple of b [say qb] $\leqslant a$, and the remainder $a-qb$ will be less than b. This is to say, given any $a, b \in \mathbb{N}$, there exist $q, r \in \mathbb{N}$ such that $a = qb+r$ and $r < b$. This is known as the *Division Algorithm*, and the existence of such q, r [given any a, b] can be proved by induction. For example, given $a = 22$ and $b = 5$, we can say that 5 goes into 22 a total of $q = 4$ times with remainder $r = 2$, and write: $22 = 4 \cdot 5 + 2$, and indeed $0 \leqslant 2 < 5$. Repeated applications of the Division Algorithm give the following technique for finding the greatest common divisor of two numbers.

Definition 0.46. Given positive integers m, n, *Euclid's Algorithm* for finding $\gcd(m, n)$ is as follows.

First find q_1, r_2 such that $m = q_1 n + r_2 \quad (0 \leqslant r_2 < n)$,
Then find q_2, r_3 such that $n = q_2 r_2 + r_3 \quad (0 \leqslant r_3 < r_2)$,
Then find q_3, r_4 such that $r_2 = q_3 r_3 + r_4 \quad (0 \leqslant r_4 < r_3)$, and so on.
Since the remainders $r_i \geqslant 0$ are strictly decreasing, we will at some point get remainder 0. The last nonzero remainder r_k is $\gcd(m, n)$.

The proof that Euclid's Algorithm gives $\gcd(m, n)$ is a repeated application of Lemma 0.45.

Example 0.47. Consider $m = 9108, n = 1121$. The first step of Euclid's Algorithm is: $9108 = 8 \cdot 1121 + 140$. The second step is: $1121 = 8 \cdot 140 + 1$, and the final step is $140 = 140 \cdot 1 = 0$, giving remainder 0. The last nonzero remainder is 1, which must be $\gcd(9108, 1121)$.

Note that we can reverse the steps of Euclid's Algorithm to express $\gcd(m, n)$ as an integer linear combination of m, n. In this example, we write the equation from the last-nonzero-remainder step as: $1 = 1121 - 8 \cdot 140$. We then use the previous equation [expressed as $140 = 9108 - 8 \cdot 1121$] to obtain: $1 = 1121 - 8 \cdot (9108 - 8 \cdot 1121)$ and so $1 = -8 \cdot 9108 + 65 \cdot 1121$.

Another way of performing the same computation is by row operations on the matrix $\left(\begin{smallmatrix} 1 & 0 & | & m \\ 0 & 1 & | & n \end{smallmatrix}\right)$. In this case:

$$\left(\begin{smallmatrix} 1 & 0 & | & 9108 \\ 0 & 1 & | & 1121 \end{smallmatrix}\right) \rightarrow^{R_1 - 8R_2} \left(\begin{smallmatrix} 1 & -8 & | & 140 \\ 0 & 1 & | & 1121 \end{smallmatrix}\right) \rightarrow^{R_2 - 8R_1} \left(\begin{smallmatrix} 1 & -8 & | & 140 \\ -8 & 65 & | & 1 \end{smallmatrix}\right) \rightarrow^{R_1 - 140R_2} \left(\begin{smallmatrix} * & * & | & 0 \\ -8 & 65 & | & 1 \end{smallmatrix}\right),$$

where the $*$ entries need not be computed. This gives us, all in the same computation, that $\gcd(9108, 1121) = 1$, and the bottom row of the last matrix gives $\gcd(9108, 1121)$ as a linear combination of $9108, 1121$, namely: $1 = -8 \cdot 9108 + 65 \cdot 1121$, as before.

This process can be performed for any m, n, giving the following result.

Lemma 0.48. *For any $m, n \in \mathbb{N}$, there exist $\lambda, \mu \in \mathbb{Z}$ such that $\lambda m + \mu n = gcd(m, n)$.*

Definition 0.49. Let $a, b, m \in \mathbb{Z}$. We say that $a \equiv b \pmod{m}$ ['a is congruent to b modulo m'] when $m | (a - b)$.

For example, $2 \equiv 12 \pmod 5$, since $5|(2-12)$. It is straightforward to show that, if $a \equiv b \pmod m$ and $c \equiv d \pmod m$, then

$a + c \equiv b + d,\ a - c \equiv b - d,\ ac \equiv bd,\ a^n \equiv b^n,\ ka \equiv kb \pmod n$,

for any $k \in \mathbb{Z}$ and any $n \in \mathbb{Z}, n \geqslant 0$. So, congruences in most ways can be manipulated like standard equations. An exception is cancellation: $ka \equiv kb \pmod m$ does not always imply that $a \equiv b \pmod m$; for example $2 \cdot 4 \equiv 2 \cdot 1 \pmod 6$ even though $4 \not\equiv 1 \pmod 6$. However, the implication is always true when k and m are coprime.

Lemma 0.50. *If $gcd(m, n) = 1$ then there exists $\lambda \in \mathbb{Z}$ such that $\lambda m \equiv 1 \pmod n$. In particular, if p is prime and $p \not| \ m$ then there exists $\lambda \in \mathbb{Z}$ such that $\lambda m \equiv 1 \pmod p$.*

Proof We know from Lemma 0.48 that there exist λ, μ such that $\lambda m + \mu n = \gcd(m, n)$. Reducing modulo n immediately gives the required result. □

Corollary 0.51. *For any $m \in \mathbb{N}$, the set $G_m = \{x : 1 \leqslant x \leqslant m, gcd(x, m) = 1\}$ is a group under multiplication modulo m. In particular, for any prime p, the set $\{1, 2, \ldots, p-1\}$ is a group under multiplication modulo p.*

Letting $G = \{1, 2, \ldots, p-1\}$, we can apply Corollary 0.36 to obtain the following.

Theorem 0.52. *(Fermat's Little Theorem). Let p be prime. If $p \not| \ a$ then $a^{p-1} \equiv 1 \pmod p$.*

As a consequence, $a^p \equiv a \pmod p$ for all a, regardless of whether $p|a$ or $p \not| \ a$.

Another natural problem in Number Theory is that of trying to decide when one number is congruent to a square modulo a prime.

Definition 0.53. Let p be prime and $m \in \mathbb{Z}$. We say that m is a *quadratic residue* mod p if there exists $x \in \mathbb{Z}$ such that $m \equiv x^2 \pmod p$. Otherwise m is a *quadratic non-residue* mod p.

For example, consider what happens modulo $p = 5$. Every number is congruent to one of $0, 1, 2, 3$ or $4 \pmod 5$ [which are the same as $0, 1, 2, -2, -1 \pmod 5$]. Now: $0^2 \equiv 0, 1^2 \equiv 1, 2^2 \equiv 4, 3^2 = (-2)^2 \equiv 4, 4^2 = (-1)^2 \equiv 1 \pmod 5$. So, $0, 1, 4$ are quadratic residues mod 5, but $2, 3$ are not.

Lemma 0.54. *For any prime $p \neq 2$, $\psi : \mathbb{F}_p^* \to \mathbb{F}_p^* : x \mapsto x^2$ is a 2-to-1 map [2 elements map to 1 element], with $\psi(x) = \psi(p-x)$, or equivalently $\psi(x) = \psi(-x)$ [since $(p-x)^2 \equiv (-x)^2 \equiv x^2 \pmod p$]. So exactly half of $\{1, \ldots, p-1\}$ are quadratic residues mod p and half are quadratic non-residues mod p.*

Definition 0.55. For prime p and $p \not| \ m$, define the *Legendre symbol* by:

$$\left(\frac{m}{p}\right) = \begin{cases} 1 & \text{if } m \text{ is a quadratic residue mod } p, \\ -1 & \text{otherwise.} \end{cases}$$

When $p|m$, we normally define $\left(\frac{m}{p}\right) = 0$.

For example, we have already seen that $\left(\frac{2}{5}\right) = -1$. Also, $\left(\frac{7}{5}\right) = \left(\frac{2}{5}\right) = -1$, since 7 and 2 are congruent (mod 5). Similarly, $\left(\frac{11}{5}\right) = \left(\frac{1}{5}\right) = 1$ and $\left(\frac{10}{5}\right) = 0$.

Lemma 0.56. *Let p be an odd prime and let $p \not| \ m, n, m_1, m_2$.*
(a) *If $m_1 \equiv m_2$ (mod p) then $\left(\frac{m_1}{p}\right) = \left(\frac{m_2}{p}\right)$.*
(b) *$\left(\frac{mn}{p}\right) = \left(\frac{m}{p}\right)\left(\frac{n}{p}\right)$, which is the same as saying:*
mn is a quadratic residue mod $p \iff$ either (m and n are both quadratic residues mod p)
or (m and n are both quadratic non-residues mod p)
(c) *$\left(\frac{-1}{p}\right) = 1 \iff p \equiv 1$ (mod 4) or $p = 2$. $\left(\frac{-1}{p}\right) = -1 \iff p \equiv 3$ (mod 4).*
(d) *$\left(\frac{2}{p}\right) = 1 \iff p \equiv \pm 1$ (mod 8). $\left(\frac{2}{p}\right) = -1 \iff p \equiv \pm 3$ (mod 8).*

Theorem 0.57. *(Gauss' Law of Quadratic Reciprocity). Let $p \neq 2, q \neq 2$ be distinct primes.*
If either $p \equiv 1$ (mod 4) or $q \equiv 1$ (mod 4) then $\left(\frac{p}{q}\right) = \left(\frac{q}{p}\right)$.
If both $p \equiv 3$ (mod 4) and $q \equiv 3$ (mod 4) then $\left(\frac{p}{q}\right) = -\left(\frac{q}{p}\right)$.

Example 0.58. Let us decide whether 6 is a quadratic residue mod 1019 [which is prime], using applications of quadratic reciprocity.
$\left(\frac{6}{1019}\right) = \left(\frac{2}{1019}\right)\left(\frac{3}{1019}\right) = (-1)\left(\frac{3}{1019}\right)$ [by Lemma 0.56(d)]
$= (-1)(-1)\left(\frac{1019}{3}\right)$ [by quadratic reciprocity, since both 1019 and 3 are $\equiv 3$ (mod 4)]
$= (-1)(-1)\left(\frac{2}{3}\right) = (-1)(-1)(-1) = -1$,
establishing that 6 is a quadratic non-residue mod 1019 [and so there does not exist an integer x such that $6 \equiv x^2$ (mod 1019)], in a way much quicker than checking that none of $0^2, 1^2, \ldots, 1018^2$ are congruent to 6 (mod 1019).

Rings

There are many situations where we have two operations on the same set, for example $\mathbb{Z}$ with both addition and multiplication.

Definition 0.59. Let R have two binary operations $+, \times$. R is a *ring* (with 1) if:
R is a commutative group under $+$ with identity 0.
There exists an element 1 ($\neq 0$) such that, for all $r \in R$, $1 \times r = r \times 1 = r$.
For all $r, s, t \in R$, $(r \times s) \times t = r \times (s \times t)$ [associativity of multiplication].
For all $r, s, t \in R$, $r \times (s + t) = r \times s + r \times t$, $(s + t) \times r = s \times r + t \times r$ [left and right distributivity].

Note that, for any ring, addition is always commutative, but multiplication need not be commutative. When multiplication is commutative [that is, $r \times s = s \times r$ for all $r, s \in R$] we say that R is a *commutative ring.*

Examples 0.60.
(a) $\mathbb{Z}, +, \times$ is a commutative ring.
(b) For any ring R, define $R[x] = \{\text{polynomials in } x \text{ with coefficients in } R\}$, which is also a ring, with the usual addition and multiplication of polynomials. Also define the ring $R[[x]] = \{\text{power series in } x \text{ with coefficients in } R\}$. The same is true when there are several variables, for example: $R[x, y], R[[x, y]]$.
(c) Let $G, +$ be any commutative group. Let $\operatorname{End}(G) = \{\phi : \phi \text{ is an endomorphism on } G\}$. Then $\operatorname{End}(G)$ is a ring, with operations: $(\phi_1 + \phi_2)(g) = \phi_1(g) + \phi_2(g)$ [defining ring addition $\phi_1 + \phi_2$], and with ring multiplication given by $\phi_1 \circ \phi_2$ [composition]. This is the *endomorphism ring* of the group G.
(d) $M_2(\mathbb{Z}) = \{2 \times 2 \text{ matrices with integer entries}\}$ is a non-commutative ring, with '0' given by $\left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and '1' given by $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$.
(e) The set $\{0, \ldots, n-1\}$ under addition and multiplication modulo n is a commutative ring.

Definition 0.61. A commutative ring R is an *integral domain* if, for all $r, s \in R$,

$$rs = 0 \implies (r = 0 \text{ or } s = 0).$$

For example, $\mathbb{Z}$ and $\mathbb{Z}[[x]]$ are integral domains, but $M_2(\mathbb{Z})$ is not, since $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)\left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right)$.

Definition 0.62. Let R_1, R_2 be rings. A function $f : R_1 \to R_2$ is a *ring homomorphism* if, for all $r, s \in R_1$, $f(r_1 + r_2) = f(r_1) + f(r_2)$ and $f(r_1 \times r_2) = f(r_1) \times f(r_2)$. If f is also a bijection, then f is a *ring isomorphism.*

If there exists an isomorphism from R_1 to R_2, then R_1 and R_2 are *isomorphic*, denoted $R_1 \cong R_2$.

The equivalent idea for rings to that of normal subgroups is as follows.

Definition 0.63. An *ideal* of a ring R is a subset $I \subset R$ satisfying:

$I, +$ is a subgroup of $R, +$.

For all $x \in I, r \in R$ we have $x \times r \in I$ and $r \times x \in I$.

This last condition can be phrased as: 'the product of anything in the ring with anything in the ideal must be in the ideal'. Note that $1 \in I \iff I = R$. If $I \neq R$ then I is a *proper ideal.* If I is a proper ideal and is not contained in a larger proper ideal, then I is a *maximal ideal.*

Definition 0.64. Let I be an ideal of a ring R; define the quotient ring $R/I = \{r+I : r \in R\}$, under the operations $(r_1+I)+(r_2+I) = (r_1+r_2)+I$ and $(r_1+I)\times(r_2+I) = (r_1 \times r_2)+I$.

Note that I is an ideal if and only if it occurs as the kernel of a ring homomorphism from R to some ring.

For example, $x\mathbb{Z}[x]$ [the polynomials with 0 constant term] is an ideal of the ring $\mathbb{Z}[x]$. It is the kernel of the ring homomorphism from $\mathbb{Z}[x]$ to $\mathbb{Z}$, defined by $p(x) \mapsto p(0)$. Furthermore: $\mathbb{Z}[x]/x\mathbb{Z}[x] \cong \mathbb{Z}$. Similarly, the ring of Example 0.60(e) is just the quotient ring $\mathbb{Z}/n\mathbb{Z}$.

Definition 0.65. Let R be a ring. If there exists an integer $n \geqslant 1$ such that $1+1+\ldots+1[n \text{ times}] = 0$, then the smallest such n is the *characteristic* of R. If no such n exists, then R is said to have characteristic 0.

For example, $\mathbb{Z}/n\mathbb{Z}$ has characteristic n, whereas $\mathbb{Z}, \mathbb{Q}, \mathbb{C}$ all have characteristic 0.

Fields

Definition 0.66. Let K have two binary operations $+, \times$. K is a *field* if:

K is an Abelian group under $+$ with identity 0,

The nonzero elements of K is an Abelian group under $\times$ with identity 1,

For all $a, b \in K$, $a \times (b+c) = a \times b + a \times c$ [distributivity].

Equivalently, we could define a field to be a commutative ring for which every nonzero element has a multiplicative inverse.

Examples 0.67.

(a) $\mathbb{Q}, +, \times$ is a field.

(b) Let $\mathbb{F}_p, +, \times$ denote $\{0, 1, \ldots, p-1\}$ under addition and multiplication modulo p, where p is prime [this is the same as $\mathbb{Z}/p\mathbb{Z}, +, \times$]. This a field with p elements (a *finite field*, since it has only finitely many elements, as opposed to the infinite field $\mathbb{Q}$). The fact that it is a group under addition modulo p is straightforward. The fact that the nonzero elements form a group under multiplication modulo p was shown in Corollary 0.51

(c) $\mathbb{R}, \mathbb{C}, \mathbb{Q}(\sqrt{2}), \mathbb{Q}(i)$ are all fields. By $\mathbb{Q}(\sqrt{2})$ we mean $\{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$, and similarly for $\mathbb{Q}(i)$.

(d) $\mathbb{Z}, +, \times$ is not a field since the nonzero integers is not a group under multiplication (for example, 3 has no inverse under multiplication).

(e) $\mathbb{Z}/6\mathbb{Z} = \{0, 1, 2, 3, 4, 5\}$ under addition and multiplication modulo 6 is not a field for the same reason.

(f) Given any ring R and maximal ideal $\mathcal{M}$, the quotient $R/\mathcal{M}$ is always a field.

(g) Given any integral domain R, define $K = \{\frac{a}{b} : a, b \in R, b \neq 0\}$, where we regard $\frac{a}{b} = \frac{a'}{b'}$ when $ab' = a'b$. This is the *field of fractions* of R. Addition and multiplication are defined as you would expect: $\frac{a_1}{b_1} + \frac{a_2}{b_2} = \frac{a_1 \times b_2 + a_2 \times b_1}{b_1 b_2}$ and $\frac{a_1}{b_1} \times \frac{a_2}{b_2} = \frac{a_1 a_2}{b_1 b_2}$. More pedantically, you could define the field of fractions as $\{(a, b) : a, b \in R\}$ modulo the equivalence relation: $(a, b) = (a', b') \iff ab' = a'b$, with addition and multiplication defined by: $(a_1, b_1) + (a_2, b_2) = (a_1 \times b_2 + a_2 \times b_1, b_1 b_2)$ and $(a_1, b_1) \times (a_2, b_2) = (a_1 a_2, b_1 b_2)$. For example, $\mathbb{Q}$ is the field of fractions of $\mathbb{Z}$.

(h) For any integral domain R, the field of fractions of $R[x]$ is denoted $R(x)$; it is the field of *rational functions* in x over R, that is, $R(x) = \{\frac{p(x)}{q(x)} : p(x), q(x) \in \mathbb{Z}[x], q(x) \neq 0\}$. Note that, if K is the field of fractions of R, then $R(x) = K(x)$.

Since fields are special cases of rings, the definitions for field homomorphism, field isomorphism and characteristic are exactly as described for rings. An isomorphism from a field to itself is an *automorphism*.

Definition 0.68. Let $K, +, \times$ be a field. Then K^* always denotes the group of nonzero elements of K under $\times$ [for example, $\mathbb{Q}^*, \mathbb{R}^*, \mathbb{C}^*$ are all groups under $\times$].

Definition 0.69. Let K be a field. Any $p(x) \in K[x]$ is *irreducible* if it cannot be written as a product of two polynomials in $K[x]$ both of degree $\geqslant 1$. It is *monic* if the leading coefficient [that is, the coefficient of the highest power of x] is 1. Let α be the root of any $p(x) \in K[x]$ (not necessarily irreducible); then α is *algebraic* over K. For example, $\sqrt{2}$ is algebraic over $\mathbb{Q}$, since it is a root of $x^2 - 2$; on the other hand, $\sqrt{-\pi}$ is algebraic over $\mathbb{R}$, but can be shown not to be algebraic over $\mathbb{Q}$. Given any α, algebraic over K, there always exists $p_\alpha(x) \in K[x]$ of smallest degree m_α which has α as a root, and which has the property that it is a factor of any other member of $K[x]$ which has α as a root. We say that $p_\alpha(x)$ is the *minimal polynomial* of α and that α is *algebraic of degree* m_α over K. A field K is *algebraically closed* if every polynomial $p(x) \in K[x]$ contains a root in K.

For example, $\mathbb{C}$ is algebraically closed, but $\mathbb{Q}$ is not. For any field K (whether algebraically closed or not), there exists a field $\overline{K}$, the *algebraic closure* of K, which is the smallest algebraically closed field containing K. Given α, algebraic of degree m_α over K, we can form the field $K(\alpha)$, which is the smallest subfield of $\overline{K}$ containing K and α. We say that $K(\alpha)$ is the field obtained by *adjoining* α to K. A similar definition applied for any $K(\alpha_1, \dots, \alpha_n)$. A field L is an *algebraic extension* of K if $K \subset L$ and every $\ell \in L$ is algebraic over K, otherwise L is a *transcendental* extension of K.

Examples 0.70.

(a) $\mathbb{C}$ is the algebraic closure of $\mathbb{R}$.

(b) The minimal polynomial of i over $\mathbb{Q}$ is $x^2 + 1$, so that i is algebraic of degree 2 over $\mathbb{Q}$, and $\mathbb{Q}(i) = \{a + bi : a, b \in \mathbb{Q}\}$.

Definition 0.71. Let L be a field extension of K [that is, K, L are fields and $K \subset L$; this is sometimes denoted L/K]. If there exists a finite set $\ell_1, \ldots, \ell_n \in L$ such that every $\ell \in L$ can be written as $\ell = k_1\ell_1 + \ldots + k_n\ell_n$, for some $k_1, \ldots, k_n \in K$, then L is a *finite extension* of K. In such cases, it is then always possible to find such a set with the extra property that $k_1\ell_1+\ldots+k_n\ell_n \neq 0$ except when $k_1 = \ldots = k_n = 0$, in which case we say that $\ell_1, \ldots, \ell_n$ is a *basis* for the field extension. We then say that n is the *degree* of the extension $L : K$, or that $[L : K] = n$. Of course, if you wish, you can also phrase this in terms of vector spaces. Letting the set of vectors be L and the field of scalars be K, then L forms a vector space with respect to vector addition: $\ell_1 + \ell_2$, for any $\ell_1, \ell_2 \in L$, being simply the usual addition in the field L, and scalar multiplication $k\ell$, for any $k \in K, \ell \in L$, being simple the usual multiplication in L. Then the degree of the extension $L : K$ is just the dimension of this vector space.

A *number field* is a finite extension of $\mathbb{Q}$

Example 0.72. The field $\mathbb{Q}(\sqrt{2})$ is a degree 2 extension of $\mathbb{Q}$, with basis $1, \sqrt{2}$.

Comment 0.73. Let $K \subset L \subset M$ be fields. Then $[M : K] = [M : L][L : K]$.

Definition 0.74. Let L be a field extension of K. Define the set

$$\mathrm{Aut}(L : K) = \{\sigma : L \to L : \sigma \text{ is an automorphism and } \sigma(k) = k \text{ for all } k \in K\},$$

that is, the set of all automorphisms of L which fix K [recall that an automorphism of L is a field isomorphism from L to itself]. Then $\mathrm{Aut}(L : K)$ forms a group under the operation of function composition, the *automorphism group* of the extension $L : K$.

For any subgroup $H \leqslant \mathrm{Aut}(L : K)$, the *fixed field* of H is the field $\{\ell \in L : \sigma(\ell) = \ell \text{ for all } \sigma \in H\}$. If K is the fixed field of $\mathrm{Aut}(L : K)$, we say that $L : K$ is a *Galois extension* and we refer to $\mathrm{Aut}(L : K)$ as the *Galois group* of the extension, denoted $\mathrm{Gal}(L : K)$ or $\mathrm{Gal}(L/K)$ or $\mathrm{Gal}_{L/K}$.

Example 0.75. The group $\mathrm{Gal}\big(\mathbb{Q}(\sqrt{2}) : \mathbb{Q}\big)$ has two elements: $e : a + b\sqrt{2} \mapsto a + b\sqrt{2}$ and $\sigma : a + b\sqrt{2} \mapsto a - b\sqrt{2}$.

Fields with Valuations

Definition 0.76. Let K be a field. A *valuation* on K is a function $|\ | : K \to \mathbb{R}$ satisfying:

(1) $|x| \geqslant 0$ for all $x \in K$, with equality if and only if $x = 0$.

(2) $|xy| = |x| \cdot |y|$ for all $x, y \in K$.

(3) $|x + y| \leqslant |x| + |y|$ for all $x, y \in K$ [the *triangle inequality*].

If a valuation also satisfies $|x + y| \leqslant \max\,(|x|, |y|)$, then we say that it is a *non-Archimedean valuation*; otherwise we say that it is an *Archimedean valuation*.

Example 0.77. Any of $\mathbb{Q}, \mathbb{R}$, together with the usual notion of absolute value: $|x| = \max\,(x, -x)$ [for example, $|-5| = |5| = 5$], is a field with an Archimedean valuation. The same is true of $\mathbb{C}$, together with the usual definition of modulus: $|a + bi| = \sqrt{a^2 + b^2}$.

Fields with valuations are special cases of metric spaces. We first recall what these are.

Definition 0.78. A *metric space* is a set M together with a *metric* d, which is a function $d : M \times M \to \mathbb{R}$, satisfying:

(1) $d(x, y) \geqslant 0$, for all $x, y \in K$, with equality if and only if $x = y$.

(2) $d(x, y) = d(y, x)$ for all $x, y \in M$.

(3) $d(x, z) \leqslant d(x, y) + d(y, z)$ for all $x, y, z \in M$ [the *triangle inequality*].

Example 0.79. Let $K, |\ |$ be a field with valuation. Then K is a metric space with respect to the associated metric $d(x, y) = |x - y|$.

Comment 0.80. The standard absolute value on $\mathbb{Q}$ is sometimes denoted $|\ |_\infty$, in order to distinguish is from other valuations on $\mathbb{Q}$ (the p-adic valuations) that will be mentioned in the lecture course. The associated metric is often denoted d_∞. So, for example, $|-5|_\infty = |5|_\infty = 5$ and $d_\infty(5, -3) = |5 - (-3)|_\infty = 8$.

We shall give the following definitions in the context of a field with a valuation, but they could just as easily be given for a general metric space.

Definition 0.81. Let $K, |\ |$ be a field with valuation. For $a_n, \ell \in K$, we say that the sequence a_n *converges* to ℓ [denoted $a_n \to \ell$] in $K, |\ |$ when $|a_n - \ell| \to 0$ in $\mathbb{R}, |\ |_\infty$ as $n \to \infty$. That is: for any $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that, $|a_n - \ell| < \epsilon$ for all $n > N$. Given a sequence $a_n \in K$, if there exists $\ell \in K$ such that $a_n \to \ell$ in $K, |\ |$ then we say that a_n *converges* in $K, |\ |$, or that it is *convergent* in $K, |\ |$.

Lemma 0.82. *Let $K, |\ |$ be a field with valuation and suppose that $a_n \to \ell$ and $b_n \to m$ in K and let $k \in K$. Then the following standard limit properties are always satisfied.*

$a_n{+}b_n \to \ell{+}m,\ \ a_n{-}b_n \to \ell{-}m,\ \ ka_n \to k\ell,\ \ a_nb_n \to \ell m,\ \ \frac{a_n}{b_n} \to \frac{\ell}{m}$ [*when* $b_n \neq 0, m \neq 0$].

Definition 0.83. Let $K, |\ |$ be a field with valuation. A sequence $a_n \in K$ is *Cauchy* if $|a_m - a_n| \to 0$ in $\mathbb{R}, |\ |_\infty$ as $m, n \to \infty$. That is: for any $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that, $|a_m - a_n| < \epsilon$ for all $m, n > N$.

Using $|a_m - a_n| = |a_m - \ell + \ell - a_n| \leqslant |a_m - \ell| + |a_n - \ell|$, the following is immediate.

Lemma 0.84. *If a sequence is convergent in $K, |\ |$ then it is also Cauchy.*

Definition 0.85. A field with valuation is *complete* if every Cauchy sequence is convergent.

Examples 0.86.

(a) Define the sequence $a_n = \frac{n-1}{n}$ in $\mathbb{Q}, |\ |_\infty$. Then $|a_n - 1|_\infty = |-\frac{1}{n}|_\infty = \frac{1}{n} \to 0$ as $n \to \infty$, so that $a_n \to 1$ in $\mathbb{Q}, |\ |_\infty$. This proves that a_n is convergent in $\mathbb{Q}, |\ |_\infty$ (and therefore is also Cauchy).

(b) Define $a_1 = \frac{1}{10},\ a_2 = \frac{1}{10} + \frac{1}{10^2},\ a_3 = \frac{1}{10} + \frac{1}{10^2} + \frac{1}{10^3}, \ldots$ in $\mathbb{Q}, |\ |_\infty$. Then $9a_n + \frac{1}{10^n} = 1$ and so $|a_n - \frac{1}{9}|_\infty = |-\frac{1}{9 \cdot 10^n}|_\infty = \frac{1}{9 \cdot 10^n} \to 0$, as $n \to \infty$, so that $a_n \to \frac{1}{9}$ in $\mathbb{Q}, |\ |_\infty$. This proves that a_n is convergent in $\mathbb{Q}, |\ |_\infty$ (and therefore is also Cauchy).

(c) Let $f(x) = x^2 - 2$, and let $\big(x_n, f(x_n)\big)$ be a point on $y = f(x)$. The tangent line to $y = f(x)$ at this point has equation: $y - f(x_n) = f'(x_n)(x - x_n)$. This tangent line cuts the x-axis when $y = 0$ and so solving for x gives: $x = x_n - \frac{f(x_n)}{f'(x_n)}$. In summary, the tangent line cuts the x-axis at the point $\big(x_n - \frac{f(x_n)}{f'(x_n)}, 0\big)$. Define x_{n+1} to be the x-coordinate, that is:

$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$ [the Newton-Raphson formula].

So, now define a sequence a_n by $a_1 = 1$ and $a_{n+1} = a_n - \frac{f(a_n)}{f'(a_n)} = a_n - \frac{a_n^2 - 2}{2a_n}$, the first few of which are: $a_1 = 1, a_2 = \frac{3}{2}, a_3 = \frac{17}{12}, \ldots$

Does a_n converge in $\mathbb{Q}, |\ |_\infty$? Well, imagine that $a_n \to \ell$ in $\mathbb{Q}, |\ |_\infty$ for some $\ell \in \mathbb{Q}$. Taking the limit as $n \to \infty$ of $a_{n+1} = a_n - \frac{a_n^2 - 2}{2a_n}$ gives $\ell = \ell - \frac{\ell^2 - 2}{2\ell}$ and so $\ell^2 - 2 = 0$; this is impossible for $\ell \in \mathbb{Q}$, which is a contradiction. We deduce that a_n is not convergent in $\mathbb{Q}, |\ |_\infty$. However, it is easy to check that a_n is Cauchy, and so the converse of Lemma 0.84 does not always hold. This shows that $\mathbb{Q}, |\ |_\infty$ is not a complete field.

Suppose that we were living in a world where all we know is $\mathbb{Q}$. What would we make of this last example? It would appear to have convergence-like properties, and yet does not converge to anything within $\mathbb{Q}$. We would feel that the sequence is approaching some gap or 'incompleteness' in our set, and that it would be nice to extend $\mathbb{Q}$ to a larger set in which the sequence actually converges. One way of constructing such a set is simply to consider the set of all Cauchy sequences in $\mathbb{Q}, |\ |_\infty$. This certainly includes a copy of $\mathbb{Q}$ since, for any $q \in \mathbb{Q}$, the sequence $q, q, q, \ldots$ is Cauchy. However, there is the problem that we now have many different Cauchy sequences representing what we would prefer to be just one element. For example, the sequences $a_n = 0$, $b_n = 1/n$ and $c_n = 1/n^2$ are all Cauchy, and we would prefer to have all of them simply be represented by the 0 element. A natural solution is to say that two Cauchy sequences a_n, b_n are equivalent if $a_n - b_n \to 0$. The set of equivalence classes of Cauchy sequences then gives us the set we want. We can extend the field operations $+, \times$ by defining $[(a_n)] + [(b_n)]$ to be $[(a_n + b_n)]$ and similarly for multiplication. Any equivalence class of sequences that actually converge to a given $q \in \mathbb{Q}$ can be denoted $[q]$ (or simply q), and these give a copy of $\mathbb{Q}$ within our new set. The equivalence class α containing the sequence a_n of Example 0.86(c) is a member of our new set, and $\alpha^2 = [2]$, so that our new larger set includes a square root of 2. This construction is generalised as follows.

Definition 0.87. Let $K, |\ |$ be a field with valuation. Define an equivalence relation on Cauchy sequences by: $(a_n) \sim (b_n) \iff a_n - b_n \to 0$. The *completion* K' of K is the set of equivalence classes of Cauchy sequences. This is a field with respect to the operations: $[(a_n)] + [(b_n)] = [(a_n + b_n)]$ and $[(a_n)] \times [(b_n)] = [(a_n \times b_n)]$. For any $k \in K$, we can use $[k]$ to denote the equivalence class of all sequences converging to k. Then $\{[k] : k \in K\}$ gives a copy of K in K'. The field K' is complete and it is the smallest complete field containing K.

Comment 0.88. The completion of $\mathbb{Q}, |\ |_\infty$ is typically denoted $\mathbb{R}$, and this gives a way of constructing the real numbers, having starting only with $\mathbb{Q}$. You might legitimately complain that $\mathbb{R}$ was mentioned in Definition 0.76, the original definition of valuation (since a valuation is defined to be a map from K to $\mathbb{R}$). However, this is easily overcome, since for $\mathbb{Q}, |\ |_\infty$ we can amend the definition of valuation to be a map from $\mathbb{Q}$ to $\mathbb{Q}$, avoiding any mention of $\mathbb{R}$ until its construction. The status of the sequence a_n of Example 0.86(c) is that it is both Cauchy and convergent in $\mathbb{R}, |\ |_\infty$, but only Cauchy (and not convergent) in $\mathbb{Q}, |\ |_\infty$.

It might seem cumbersome to think of a real number as being an equivalence class of Cauchy sequences, but of course in practice it is easier to try a similar trick here as for quotient groups, where we choose a representative for each equivalence class. After all, this is what we are really doing with decimal expansions of real numbers. When we write: $\sqrt{2} = 1.414\ldots$ this is a shorthand notation for the sequence $a_1 = 1.4, a_2 = 1.41, a_3 = 1.414, \ldots$, where $a_n \in \mathbb{Q}$ is the largest number to n decimal places satisfying $a_n^2 < 2$. This sequence is Cauchy but not convergent in $\mathbb{Q}, |\ |_\infty$, and it can be taken as the representative of its equivalence class, and labelled $\sqrt{2}$ in $\mathbb{R}$.

If one were to replace all occurrences of $|x - y|$ by $d(x, y)$, and remove all mention of the field operations, then the above discussion also describes completion in the more general context of an arbitrary metric space. However, we shall not require that level of generality here, since all completions in the lecture course will be for fields with valuations.

Of course, if a different valuation were to be used on $\mathbb{Q}$ then we would expect different sequences to converge, and a different completion. In the lecture course, we shall see another example of a valuation on $\mathbb{Q}$, the p-adic valuation $|\ |_p$. The completion of $\mathbb{Q}$ with respect to this valuation is called the field of p-adic numbers, denoted $\mathbb{Q}_p$. We shall see in the lecture course that the field $\mathbb{Q}_p$ is helpful for tackling certain types of problems in Number Theory.

Geometry

Definition 0.89. Let K be a field. $\mathbf{A}^n = \{(x_1, \ldots, x_n) : x_1, \ldots x_n \in K\}$ is called *affine n-space*[1]. When $P \in \mathbf{A}^n(K)$, we say that P is *K-rational* or *defined over K*.

[1] Usually the field K will be implicit, but we could write $\mathbf{A}^n_K$ if we want to remember it in the notation.

Example 0.90. $(\frac{1}{2},\frac{3}{4}) \in \mathbf{A}^2(\mathbb{Q}) \subset \mathbf{A}^2(\mathbb{C})$. The point $(\frac{1}{2},\frac{3}{4})$ is $\mathbb{Q}$-rational (we can also say: it is a $\mathbb{Q}$-rational point, or that it is defined over $\mathbb{Q}$). Of course, it is also $\mathbb{R}$-rational and $\mathbb{C}$-rational. The point $(\frac{1}{2}, 2+i, \sqrt{2}) \in \mathbf{A}^3(\mathbb{C})$ but is not a member of $\mathbf{A}^3(\mathbb{Q})$. The point $(\frac{1}{2}, 2+i, \sqrt{2})$ is defined over $\mathbb{C}$, but not defined over $\mathbb{Q}$ (it is $\mathbb{C}$-rational, but not $\mathbb{Q}$-rational).

Definition 0.91. A *monomial* is a product of the form $kx_1^{m_1}\ldots x_\ell^{m_\ell}$, where $x_1,\ldots,x_\ell$ are variables, $k \in K$, $m_1,\ldots,m_\ell \geqslant 0$, which has *degree* $m_1+\ldots+m_\ell$ [so that a polynomial is a sum of monomials; we also call the monomials the *terms* of the polynomials]. A *rational function* is a quotient of two polynomials.

For example, $\frac{1+x^2}{4+x+x^3}$ is a rational function in the variable x, and $\frac{s+\sqrt{2}t^2}{1+s+st^2}$ is a rational function in the variables s,t. Note that $x^{1/2}$ is neither a polynomial nor a rational function (all exponents in a polynomial must be integers $\geqslant 0$).

Definition 0.92. An algebraic expression such as a curve, polynomial, rational function, is said to be *defined over* K (or K-rational) if it can be described by an equation with coefficients in K.

Examples 0.93.
(a) x^3+1 is a polynomial in x, defined over $\mathbb{Q}$.
(b) $\frac{s+\sqrt{2}t^2}{1+s+st^2}$ is a rational function in s,t, defined over $\mathbb{Q}(\sqrt{2})$. We could also say that it is a $\mathbb{Q}(\sqrt{2})$-rational rational function. Note the two different uses of the word rational here: in the phrase $\mathbb{Q}(\sqrt{2})$-rational, which refers to the fact that the coefficients are in $\mathbb{Q}(\sqrt{2})$, and in the phrase 'rational function', which refers to the fact that the expression is a quotient of two polynomials.

Definition 0.94. A (nonzero) polynomial in two variables $f(x,y)$, with coefficients in K, defines an (affine) *curve defined over* K. For any field L with $K \subset L$, the set of L-rational points on a curve $\mathcal{C}$ is denoted $\mathcal{C}(L)$. The field K is often called the *field of definition* (or the *ground field*).

Example 0.95. Let $\mathcal{C} : f(x,y) = x^2+y^2 = 0$. This defines an affine curve over $\mathbb{Q}$ [so, we can also say it is a curve defined over $\mathbb{Q}$]. Of course, this same curve $\mathcal{C}$ could be regarded having field of definition (ground field) $\mathbb{Q}, \mathbb{Q}(\sqrt{2}), \mathbb{R}, \mathbb{C}$ or indeed any field containing $\mathbb{Q}$. When the field of definition is not stated explicitly, it is taken to be the smallest possible field over which the curve is defined (in this case, $\mathbb{Q}$). The point $(0,0)$ is $\mathbb{Q}$-rational [since all the coordinates are in $\mathbb{Q}$] and it is the only $\mathbb{Q}$-rational point on $\mathcal{C}$, so that $\mathcal{C}(\mathbb{Q}) = \{(0,0)\}$. It has many $\mathbb{C}$-rational points, for example $(i,1) \in \mathcal{C}(\mathbb{C})$, since $i \in \mathbb{C}, 1 \in \mathbb{C}$.

Comment 0.96. Of course, it is also possible to embed curves in higher dimensional space, as long as the number of 'independent' polynomials is one less than the number of variables; for example the 2 equations: $y^2 + 4x^2 - 1 = 0, z^2 - x^2 - x = 0$ define a curve in the variables x, y, z. However, we shall not concern ourselves with that here, and we shall assume that all of our affine curves are defined by a single polynomial in two variables.

Definition 0.97. The *degree* of a polynomial is the degree of its highest degree monomial. A *homogeneous* polynomial is a polynomial whose terms all have the same degree.

Example 0.98. $f(x, y) = x+y-8 = 0$ defines a curve of degree 1 (a *linear* curve), $g(x, y) = xy+y^2-y+3 = 0$ defines a curve of degree 2 (a *quadratic* curve) and $h(x, y) = x^3+y^3+y-1$ defines a curve of degree 3 (a *cubic* curve). None of these polynomials are homogeneous.

If you try drawing an accurate sketch of, for example, the three curves $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$ defined by $x^2 + y^2 = 1, y^2 = x^3, y^2 = x(x - 2)^2$, respectively, you will notice distinguishing features. The first curve $\mathcal{C}_1$ appears smooth at all points, and it is easy to see that there is a unique tangent at each point. The curve $\mathcal{C}_2$ has a 'sharp corner' at $(0, 0)$, and the third curve $\mathcal{C}_3$ crosses itself at the point $(2, 0)$, when there is a plausible choice of two distinct tangents. These sharp corners and crossing points are typified by the fact that both partial derivatives of f vanish, when the curve is written as $f(x, y) = 0$.

Definition 0.99. Let $\mathcal{C} : f(x, y) = 0$ be an (affine) curve defined over a field K and let $P = (x_0, y_0)$ be a point in $\mathcal{C}(\overline{K})$, where $\overline{K}$ is an algebraic closure of K. We say that P is a *singular point* (or a *singularity*) on $\mathcal{C}$ if $\frac{\partial f}{\partial x}(P) = 0$ and $\frac{\partial f}{\partial y}(P) = 0$. Otherwise, P is a *smooth point* (or a *nonsingular point*) on $\mathcal{C}$. A curve $\mathcal{C}$ is called *smooth* (or *nonsingular*) if it does not contain any singular points (the curve is called *singular* if it contains at least one singular point).

Comment 0.100. There is a standard technique for computing all tangents to $\mathcal{C} : f(x, y) = 0$ at a point $P = (x_0, y_0)$, in which we first translate the curve by $(-x_0, -y_0)$ [so that (x_0, y_0) is taken to $(0, 0)$], then use the fact that the lowest degree terms dominate near $(0, 0)$ and determine the tangent behaviour at $(0, 0)$, and then finally translate the curve back to its original position. This gives three steps.
Step 1. Consider $f(x+x_0, y+y_0)$ [same as $f(x, y)$ translated by $(-x_0, -y_0)$] which contains the point $x = y = 0$ and so has no constant term. We can write:

$$f(x + x_0, y + y_0) = R_k(x, y) + R_{k+1}(x, y) + \ldots + R_n(x, y),$$

where $k \geqslant 1$ and where each $R_i(x, y)$ is homogeneous of degree i (for $k \leqslant i \leqslant n$) and $R_k(x, y) \neq 0$.

Step 2. Consider $R_k(x,y)$, which is the lowest degree portion of $f(x+x_0,y+y_0)$, and factorise $R_k(x,y)=L_1(x,y)L_2(x,y)\dots L_k(x,y)$ over the algebraic closure, where $L_1,\dots,L_k$ are linear.

Step 3. There are k tangents to $f(x+x_0,y+y_0)=0$ at $(0,0)$ namely: $L_1(x,y)=0,\dots,L_k(x,y)=0$. So, after reversing the translation of Step 1, there are k tangents to $\mathcal{C}:f(x,y)=0$ at $P=(x_0,y_0)$, namely:

$$L_1(x-x_0,y-y_0)=0,\dots,L_k(x-x_0,y-y_0)=0.$$

Note that the same tangent may be repeated more than once [e.g. $\mathcal{C}:f(x,y)=y^2-x^3=0$ has 2 tangents at $(0,0)$, namely: $y=0$ twice, in which case we can say that the tangent $y=0$ occurs with multiplicity 2].

Comment 0.101. $P=(x_0,y_0)$ is a smooth point on $\mathcal{C}$

$\iff k=1$ in Step 1

$\iff$ there is only one tangent to $\mathcal{C}$ at P.

When $k\geqslant 2$, the singularity at P is called a *double point* ($k=2$), *triple point* ($k=3$), and so on.

Example 0.102. Let $\mathcal{C}_1:x^2+y^2=1$ [circle of radius 1 and centre $(0,0)$]. Then we can write: $\mathcal{C}_1:f(x,y)=x^2+y^2-1=0$, and so $\frac{\partial f}{\partial x}=2x$, $\frac{\partial f}{\partial y}=2y$. A point (x,y) is a singular point on $\mathcal{C}_1$ exactly when: it lies on $\mathcal{C}_1$ and both partial derivatives are zero, that is, when:

$$(1)\ x^2+y^2-1=0,\quad (2)\ 2x=0,\quad (3)\ 2y=0.$$

Assuming our ground field does not have characteristic 2, equations (2),(3) force $x=y=0$, but this does not satisfy (1). We conclude that there are no singular points and that $\mathcal{C}_1$ is smooth.

Example 0.103. Let $\mathcal{C}_2:y^2=x^3$, that is: $\mathcal{C}_2:f(x,y)=y^2-x^3=0$. Then $\frac{\partial f}{\partial x}=-3x^2$, $\frac{\partial f}{\partial y}=2y$. We can see that the only singular point is $(0,0)$. For computing tangents at $(0,0)$, we first take $f(x+0,y+0)=y^2-x^3=R_2(x,y)+R_3(x,y)$, where $R_2(x,y)=y^2$ and $R_3(x,y)=-x^3$. Then $R_2(x,y)=y^2=L_1(x,y)L_2(x,y)=y\cdot y$, so there are two tangents to $\mathcal{C}_2$ at $(0,0)$, namely: $L_1(x-0,y-0)=0$ and $L_2(x-0,y-0)=0$, that is: $y=0$ and $y=0$ (i.e. $y=0$ with multiplicity 2). A double point singularity where the same tangent line has multiplicity 2 is called a *cusp* (or a *cuspidal singularity*).

Example 0.104. Let $\mathcal{C}_3:y^2=x(x-2)^2$, that is: $\mathcal{C}_3:f(x,y)=y^2-x(x-2)^2=0$. The point (x,y) on $\mathcal{C}_3$ is singular when:

$$(1)\ y^2-x(x-2)^2=0,\quad (2)\ \frac{\partial f}{\partial x}=-3x^2+8x-4=0,\quad (3)\ \frac{\partial f}{\partial y}=2y=0.$$

Assuming our ground field does not have characteristic 2, from (3) we see that $y = 0$, and substituting this into (1) gives: $x(x-2)^2 = 0$, so that $x = 0$ or 2. Now, $x = 2$ satisfies (2), but $x = 0$ does not, giving $x = 2$ as the only common solution. So, the only possible singular point is $(2, 0)$ [conversely, check that $x = 2, y = 0$ satisfies (1),(2),(3) so that $(2, 0)$ is a singular point]. We conclude that $(2, 0)$ is the only singularity on $\mathcal{C}_3$.

For the tangents at $(2, 0)$, first compute $f(x+2, y+0) = y^2 - (x+2)x^2 = y^2 - 2x^2 - x^3 = R_2(x, y) + R_3(x, y)$, where $R_2(x, y) = y^2 - 2x^2$ and $R_3(x, y) = -x^3$. Factorising $R_2(x, y)$ into linear factors gives: $R_2(x, y) = (y + \sqrt{2}x)(y - \sqrt{2}x) = L_1(x, y)L_2(x, y)$. The tangents to the curve $\mathcal{C}_3$ at $(2, 0)$ are then: $L_1(x-2, y-0) = 0$ and $L_2(x-2, y-0) = 0$, that is: $y = -\sqrt{2}(x-2)$ and $y = \sqrt{2}(x-2)$. The point $(2, 0)$ is a double point with two distinct tangents; such a point is called a *node* (or a *nodal singularity*).

Note that the system of equations satisfied by singular points is over-represented, since there are 3 equations and only 2 variables. If you choose a curve 'at random', you would expect the first two of these equations to have only finitely many solutions, and it is rather a fluke if one of these solutions also happens to satisfy the third equation. So, a 'typical' curve will be smooth.

A useful tool, for computing singularities and other purposes, is the idea of the resultant of two polynomials.

Definition 0.105. Let $f(x) = f_m x^m + \ldots + f_0$ and $g(x) = g_n x^n + \ldots + g_0$, where $f_m \neq 0$ and $g_n \neq 0$. The *resultant* of $f(x)$ and $g(x)$, denoted $\mathrm{Res}\big(f(x), g(x)\big)$ or just $\mathrm{Res}(f, g)$, the following determinant of an $(m+n) \times (m+n)$ matrix.

$$\left|\begin{array}{ccccccccc}
 & \langle n-1 & 0\text{'s}\rangle & & & f_m & \ldots & \ldots & f_0 \\
 & \langle n-2 & 0\text{'s}\rangle & & f_m & \ldots & \ldots & f_0 & 0 \\
 & & & & \vdots & & & & \\
 & & & & \vdots & & & & \\
f_m & \ldots & \ldots & f_0 & & & \langle n-1 & 0\text{'s}\rangle & \\
 & \langle m-1 & 0\text{'s}\rangle & & & g_n & \ldots & \ldots & g_0 \\
 & \langle m-2 & 0\text{'s}\rangle & & g_n & \ldots & \ldots & g_0 & 0 \\
 & & & & \vdots & & & & \\
 & & & & \vdots & & & & \\
g_n & \ldots & \ldots & g_0 & & & \langle m-1 & 0\text{'s}\rangle &
\end{array}\right|$$

The following are easy to show.

Lemma 0.106. *Let $f(x), g(x) \in R[x]$ be polynomials of degree m, n, respectively, defined over a commutative ring R.*

(a) *There exist polynomials $p(x) \in R[x]$, of degree at most $n-1$, and $q(x) \in R[x]$, of degree at most $m-1$, such that: $p(x)f(x) + q(x)g(x) = Res\big(f(x), g(x)\big)$.*

(b) *When R is a field, Res$\big(f(x), g(x)\big) = 0 \iff f(x)$ and $g(x)$ have a non-constant common factor.*

Definition 0.107. The *discriminant* of a degree n polynomial $f(x) = f_n x^n + \ldots f_0$ is given by: $\text{Disc}(f) = \text{Res}(f, f')/f_n$.

Comment 0.108. **(a)** Given a monic polynomial $f(x) \in R[x]$, there exist polynomials $p(x), q(x) \in R[x]$ such that $p(x)f(x) + q(x)f'(x) = \text{Disc}(f)$.
(b) $\text{Disc}(f) = 0 \iff f$ and f' have a common root $\iff f$ has a repeated root. For example, $\text{Disc}(x^3 - 2x^2 + x) = 0$, whereas $\text{Disc}(x^2 + 1) \neq 0$.

Example 0.109. Let $f(x) = ax^2 + bx + c$. Then $\text{Disc}(f) = \text{Res}(f, f')/a$

$$= \text{Res}(ax^2 + bx + c, 2ax + b)/a = \tfrac{1}{a}\begin{vmatrix} a & b & c \\ 0 & 2a & b \\ 2a & b & 0 \end{vmatrix} = b^2 - 4ac,$$

which is the discriminant you know from school, appearing under the square root sign in the quadratic formula.

Example 0.110. Let $f(x) = x^3 + Ax + B$. Then $\text{Disc}(f) = \text{Res}(f, f')$

$$= \text{Res}(x^3 + Ax + B, 3x^2 + A) = \begin{vmatrix} 0 & 1 & 0 & A & B \\ 1 & 0 & A & B & 0 \\ 0 & 0 & 3 & 0 & A \\ 0 & 3 & 0 & A & 0 \\ 3 & 0 & A & 0 & 0 \end{vmatrix} = 4A^3 + 27B^2.$$

Example 0.111. An application of resultants to singularities is as follows. Consider the curve $\mathcal{C} : y^2 = x^3 + Ax + B$ [that is: $g(x, y) = x^3 + Ax + B - y^2 = 0$], where $A, B \in K$, a field of characteristic not equal to 2. Suppose (x_0, y_0) is a singular point on $\mathcal{C}$, so that:

$$(1)\ g(x_0, y_0) = 0, \quad (2)\ \frac{\partial g}{\partial x}(x_0, y_0) = 0, \quad (3)\ \frac{\partial g}{\partial y}(x_0, y_0) = 0,$$

giving:

$$(1)\ y_0^2 = x_0^3 + Ax_0 + B, \quad (2)\ 3x_0^2 + A = 0, \quad (3)\ 2y_0 = 0.$$

Since the characteristic of K is not equal to 2, we know that $2 \neq 0$, and so (3) gives $y_0 = 0$. Substituting this into (1) tells us that x_0 is a root of $x^3 + Ax + B$, and (2) tells us that x_0 is a root of its derivative; this is possible exactly when $x^3 + Ax + B$ has a repeated root – in other words, when $\text{Disc}(x^3 + Ax + B) = 0$. We have already seen in Example 0.110 that $\text{Disc}(x^3 + Ax + B) = 4A^3 + 27B^2$.

In summary, the curve $\mathcal{C}$ is smooth if and only if $4A^3 + 27B^2 \neq 0$.

Another basic idea in geometry applies to situations where $f(x, y)$ itself has a proper factorisation, for example: $\mathcal{C} : f(x, y) = x^2 - y^2 = 0$. This is a quadratic curve, but it factors as $(x+y)(x-y)=0$, and so the graph of $\mathcal{C}$ is just the union of the graphs of the lines $x+y=0$ and $x-y=0$. This seems geometrically different from curve such as $x^2-y^2+1=0$, which has no such factorisation. This is formalised in the following definition.

Definition 0.112. Let $\mathcal{C} : f(x,y)=0$ be a curve defined over K, and let L be any field containing K. We say that $\mathcal{C}$ is *irreducible over* L if $f(x,y)$ cannot be expressed as a product of two polynomials, both of degree $\geqslant 1$ and both defined over L [by the word *irreducible* on its own, we mean irreducible over K]. For any $\mathcal{C} : f(x,y)=0$, we can write f uniquely (up to constants and reordering) as a product $f=f_1f_2\ldots f_n$, where $f_1,\ldots,f_n$ are irreducible over L. The curves $\mathcal{C}_1 : f_1(x,y)=0,\ldots,\mathcal{C}_n : f_n(x,y)=0$ are called the *irreducible components* of $\mathcal{C}$ over L.

Examples 0.113.
(a) $\mathcal{C} : f(x,y)=y^2-2x^2=0$, defined over $\mathbb{Q}$. This is irreducible [by which we mean irreducible over $\mathbb{Q}$], but it becomes reducible over $\mathbb{C}$, with irreducible components $\mathcal{C}_1 : y=\sqrt{2}x$ and $\mathcal{C}_2 : y=-\sqrt{2}x$.
(b) $\mathcal{C} : f(x,y)=y^4-x^4=0$ is reducible. Its irreducible components (over $\mathbb{Q}$) are: $y-x=0, y+x=0, y^2+x^2=0$. The last of these becomes reducible over $\mathbb{C}$, and the irreducible components over $\mathbb{C}$ are: $y-x=0, y+x=0, y+ix=0, y-ix=0$.

It is also helpful to formalise the relationship between curves such as $x^2+y^3-5=0$ and $(x+1)^2+y^3-5=0$, where there are maps from one to the other. In this case, one can map each curve to the other with a linear map, but more generally we consider maps between curves described by rational functions (quotients of polynomials).

Definition 0.114. Let $\mathcal{C} : f(x,y)=0$ and $\mathcal{C}' : g(x,y)=0$ be curves over K. A *rational map* $\underline{\phi}$ over L from $\mathcal{C}$ to $\mathcal{C}'$ is a map given by a pair ϕ_1,ϕ_2 of rational functions in x,y, defined over L [i.e. ϕ_1,ϕ_2 are both of the form $\dfrac{\text{polynomial in } x,y}{\text{polynomial in } x,y}$ and the coefficients of ϕ_1,ϕ_2 are in L], with the property that, given any point $P=(x_0,y_0)$ on $\mathcal{C}$, then the point $\big(\phi_1(x_0,y_0),\phi_2(x_0,y_0)\big)$ lies on $\mathcal{C}'$ [for all but finitely many points (x_0,y_0) at which the denominators of ϕ_1,ϕ_2 are 0]. If there also exists a rational map $\underline{\psi}=\big(\psi_1(x,y),\psi_2(x,y)\big)$ from $\mathcal{C}'$ to $\mathcal{C}$ such that $\underline{\psi}\ \underline{\phi}$ is the identity on $\mathcal{C}$ and $\underline{\phi}\ \underline{\psi}$ is the identity on $\mathcal{C}'$ then we say that $\underline{\phi}$ is a *birational transformation* over L from $\mathcal{C}$ to $\mathcal{C}'$ and that $\mathcal{C}$ and $\mathcal{C}'$ are *birationally equivalent* over L.

Examples 0.115.
(a) Let $\mathcal{C} : x^4+y^4=1$ [i.e. $f(x,y)=x^4+y^4-1=0$] and let $\mathcal{C}' : x^4+y^2=1$ [i.e. $g(x,y)=x^4+y^2-1=0$]. Define $\underline{\phi} : \mathcal{C}\to\mathcal{C}'$ by $\underline{\phi}(x,y)=(x,y^2)$ [in the notation of the Definition 0.114: $\phi_1(x,y)=x$ and $\phi_2(x,y)=y^2$]. This is a rational map from $\mathcal{C}$ to $\mathcal{C}'$ over $\mathbb{Q}$

since, if (x, y) satisfies $\mathcal{C} : x^4 + y^4 = 1$ then $x^4 + (y^2)^2 = 1$ and so (x, y^2) lies on $\mathcal{C}'$. This is a rational map from $\mathcal{C}$ to $\mathcal{C}'$, but it is not a birational transformation, since there is no inverse map ($\underline{\phi}$ is 2-to-1).

(b) Let $\mathcal{C} : x^2 + y^3 - 5 = 0$ and $\mathcal{C}' : (x + 1)^2 + y^3 - 5 = 0$. If (x, y) is on $\mathcal{C}$ then $x^2 + y^3 - 5 = 0$ and so $((x - 1) + 1)^2 + y^3 - 5 = 0$, giving that $(x - 1, y)$ lies on $\mathcal{C}'$. The map $\underline{\phi}(x, y) = \big(\phi_1(x, y), \phi_2(x, y)\big) = (x - 1, y)$ is then a rational map over $\mathbb{Q}$ from $\mathcal{C}$ to $\mathcal{C}'$, and the inverse map is clearly $\underline{\psi}(x, y) = (x + 1, y)$. The map $\underline{\phi}$ is a birational transformation from $\mathcal{C}$ to $\mathcal{C}'$ over $\mathbb{Q}$, and so $\mathcal{C}$ and $\mathcal{C}'$ are birationally equivalent over $\mathbb{Q}$.

Note that the rational map from $\mathcal{C}$ to $\mathcal{C}'$ is in the opposite direction to the variable replacement which transforms the equations. In the above example, $\underline{\phi}(x, y) = (x - 1, y)$ is the map from $\mathcal{C}$ to $\mathcal{C}'$ [in that it maps points on $\mathcal{C}$ to points on $\mathcal{C}'$; for example, the point $(2, 1)$ on $\mathcal{C}$ maps to $(1, 1)$ on $\mathcal{C}'$], but the variable replacement 'replace x by $x - 1$ and y by y' changes the equation for $\mathcal{C}'$ into the equation for $\mathcal{C}$.

(c) Let $\mathcal{C} : x^2 - y^2 = 0$ and $\mathcal{C}' : x^2 + y^2 = 0$. Clearly $\underline{\phi} : \mathcal{C} \to \mathcal{C}'$, defined by $\underline{\phi}(x, y) = (x, iy)$ is a rational map from $\mathcal{C}$ to $\mathcal{C}'$, with inverse $\underline{\psi}(x, y) = (x, -iy)$. This shows that $\mathcal{C}$ and $\mathcal{C}'$ are birationally equivalent over $\mathbb{C}$. However, $\mathcal{C}$ and $\mathcal{C}'$ are not birationally equivalent over $\mathbb{Q}$, since any such map would take the infinitely many members of $\mathcal{C}(\mathbb{Q})$ to infinitely many members of $\mathcal{C}'(\mathbb{Q})$, contradicting the fact that $\mathcal{C}'(\mathbb{Q}) = \{(0, 0)\}$.

(d) Let $\mathcal{C} : y^2 = x^4 + 3x^2 + 5$ and $\mathcal{C}' : y^2 = 5x^4 + 3x^2 + 1$. Define $\underline{\phi}(x, y) = (\frac{1}{x}, \frac{y}{x^2})$. If (x, y) is a point on $\mathcal{C}$ then $y^2 = x^4 + 3x^2 + 5$ and so $\frac{y^2}{x^4} = 1 + \frac{3}{x^2} + \frac{5}{x^4}$, giving: $\big(\frac{y}{x^2}\big)^2 = 1 + 3\big(\frac{1}{x}\big)^2 + 5\big(\frac{1}{x}\big)^4$, so that $\big(\frac{1}{x}, \frac{y}{x^2}\big)$ is a point on $\mathcal{C}'$. Our map $\underline{\phi}$ is then a rational map (over $\mathbb{Q}$) from $\mathcal{C}$ to $\mathcal{C}'$. The inverse map is $\underline{\psi}(x, y) = (\frac{1}{x}, \frac{y}{x^2})$ [check that $\underline{\psi}\big(\underline{\phi}(x, y)\big) = \underline{\psi}\big(\frac{1}{x}, \frac{y}{x^2}\big) = \big(\frac{1}{1/x}, \frac{y/x^2}{(1/x)^2}\big) = (x, y)$, so that $\underline{\psi}\ \underline{\phi}$ is the identity, as is $\underline{\phi}\ \underline{\psi}$]. Hence $\underline{\phi}$ is a birational transformation over $\mathbb{Q}$; the curves $\mathcal{C}$ and $\mathcal{C}'$ are birationally equivalent over $\mathbb{Q}$.

(e) Let $\mathcal{C} : x^2 + y^2 = 1$ and $\mathcal{C}' : y = 0$. It might at first seem surprising that a circle should be birationally equivalent to a line, but we can establish the map first by fixing a specific point on $\mathcal{C}$, say $P_0 = (-1, 0)$, and mapping a point on $\mathcal{C}$ to $s = \frac{y}{x+1}$, the slope of the line from P_0 to (x, y) [literally, we are mapping it to $(s, 0)$]. Define: $\underline{\phi}(x, y) = (\frac{y}{x+1}, 0)$ from $\mathcal{C}$ to $\mathcal{C}'$ [defined everywhere except at the point $(-1, 0)$, but this is permissible, since the definition of rational map allows us to have a finite number of points where the map is not defined]. For the inverse, note that if the slope is s, then the line through P_0 and (x, y) has equation: $y = s(x + 1)$; substituting this into $\mathcal{C}$ gives $x^2 + s^2(x + 1)^2 = 1$, and so: $(x + 1)(x - 1 + s^2(x + 1)) = 0$. When $x \neq -1$, this gives $x = \frac{1-s^2}{1+s^2}$ and $y = s(x + 1) = \frac{2s}{1+s^2}$. This suggests that, for the inverse map, we should take: $\underline{\psi}(x, y) = \big(\frac{1-x^2}{1+x^2}, \frac{2x}{1+x^2}\big)$. It is straightforward to check that this is indeed a map from $\mathcal{C}'$ to $\mathcal{C}$ [since $\left(\frac{1-x^2}{1+x^2}\right)^2 + \left(\frac{2x}{1+x^2}\right)^2 = 1$ for any x], that $\underline{\psi}\ \underline{\phi} =$ identity on $\mathcal{C}$ and that $\underline{\phi}\ \underline{\psi} =$ identity on $\mathcal{C}'$. Hence $\mathcal{C}$ and $\mathcal{C}'$ are birationally equivalent over $\mathbb{Q}$.

Definition 0.116. A *parametrisation* of a curve $\mathcal{C}$ is a birational equivalence between $\mathcal{C}$ and a line.

Comment 0.117. The birational transformation in Example 0.115(e) is a parametrisation of the circle $x^2 + y^2 = 1$. Note that a parametrisation is an unusual type of birational transformation, in that it gives a map to a single variable; in this case, $\left(\frac{1-s^2}{1+s^2}, \frac{2s}{1+s^2}\right)$ gives a description of the points on $\mathcal{C}$ in terms of the parameter s. Since the maps $\underline{\phi}$ and $\underline{\psi}$ are defined over $\mathbb{Q}$, this gives a way of describing all $\mathbb{Q}$-rational points on $\mathcal{C}$, namely: $(x, y) \in \mathcal{C}(\mathbb{Q}) \iff s \in \mathbb{Q}$. For example, $s = 2$ gives $\left(-\frac{3}{5}, \frac{4}{5}\right) \in \mathcal{C}(\mathbb{Q})$.

The curve $x^2 + y^2 = 1$ is a special case of the following class of curves.

Definition 0.118. A *conic* is a quadratic curve: $ax^2 + 2bxy + cy^2 + 2dx + 2fy + g = 0$, satisfying

$$\begin{vmatrix} a & b & d \\ b & c & f \\ d & f & g \end{vmatrix} \neq 0 \quad \text{(which guarantees that the curve is smooth).}$$

A conic is an ellipse, hyperbola or parabola; the name ‘conic’ refers to the fact that these are the curves which can be obtained by intersecting a plane and a double-cone [two cones with the same axis, placed apex to apex]. The parametrisation of the circle given in Example 0.115(e) is a special case of the following result.

Theorem 0.119. *Any conic $\mathcal{C}$ (over K) with a K-rational point is birationally equivalent to a line [i.e. it is parametrisable].*

Proof We are given that there exists a K-rational point (x_0, y_0) on the curve $\mathcal{C} : f(x, y) = 0$. Let $g(x, y) = f(x + x_0, y + y_0)$. This contains the point $(0, 0)$ so that we can write: $g(x, y) = g_1(x, y) + g_2(x, y)$, where g_1 is homogeneous & linear, and g_2 is homogeneous & quadratic. Hence $g(x, tx) = x\phi_1(t) + x^2\phi_2(t) = 0$. Apart from $x = 0$, we can take $x = -\phi_1(t)/\phi_2(t), y = -t\phi_1(t)/\phi_2(t)$ [with inverse $t = y/x$] as a parametrisation of $g(x, y) = 0$. The parametrisation of $\mathcal{C}$ is then: $x = x_0 - \phi_1(t)/\phi_2(t), y = y_0 - t\phi_1(t)/\phi_2(t)$ [with inverse $t = (y - y_0)/(x - x_0)$]. □

Definition 0.120. The curves $\mathcal{C} : f(x, y) = 0$ and $\mathcal{C}' : g(x, y) = 0$ *intersect* at $P = (x_0, y_0)$ if P lies on both of $\mathcal{C}$ and $\mathcal{C}'$ [that is, $f(x_0, y_0) = g(x_0, y_0) = 0$].

Definition 0.121. Suppose the curves $\mathcal{C} : f(x, y) = 0$ and $\mathcal{C}' : g(x, y) = 0$ intersect at $P = (x_0, y_0) \in \mathcal{C}(L)$ (with L a field containing the field of definition of the curve). The curves *intersect with multiplicity* $r > 0$ at P if the dimension of the quotient ring

$$\dim_L L[\![x,y]\!]/(f(x+x_0,y+y_0),g(x+x_0,y+y_0)) = r.$$

The intersection multiplicity is ∞ if and only if $\mathcal{C}$ and $\mathcal{C}'$ have a common irreducible component containing P. We refer to Fulton *Algebraic Curves* for details and proofs of the fundamental properties of the intersection multiplicity. You can also take a look at Part B Algebraic Curves for an approach via resultants.

Lemma 0.122. *Consider a curve $\mathcal{C} : f(x,y) = 0$ over K and a line $\mathcal{D}$ parameterised by $x = at+b$, $y = ct+d$, with $a,b,c,d \in K$ and a,c not both zero. Then $\mathcal{C}$ and $\mathcal{D}$ intersect at the points $P = (at_0+b, ct_0+d)$ with t_0 a root of the polynomial $F(t) = f(at+b,ct+d)$. If $F(t)$ is identically 0, then $\mathcal{C}$ contains the line $\mathcal{D}$.*

Suppose $t_0 \in \overline{K}$ is a root of $F(t)$ and let $P = (at_0+b, ct_0+d)$. Then $\mathcal{C}$ and $\mathcal{D}$ intersect at P with multiplicity equal to the multiplicity of t_0 as a root of $F(t)$.

Proof. The intersection property is clear, so we need to verify the assertion about multiplicities. We can apply an affine transformation and assume WLOG that $t_0 = 0$, so $P = (b,d)$. The line $\mathcal{D}$ has equation $g(x,y) = cx - ay + ad - bc = 0$. We have to compute the dimension of the $\overline{K}$-vector space

$$\overline{K}[\![x,y]\!]/(f(x+b,y+d),g(x+b,y+d)).$$

It is not hard to check that the map

$$\begin{aligned}\overline{K}[\![x,y]\!]/(g(x+b,y+d) &\to \overline{K}[\![t]\!]\\ x &\mapsto at\\ y &\mapsto ct\end{aligned}$$

is an isomorphism. So we need to compute the dimension of the $\overline{K}$-vector space

$$\overline{K}[\![t]\!]/(f(at+b,ct+d)) = \overline{K}[\![t]\!]/(F(t)).$$

We claim that this is equal to the multiplicity of 0 as a root of $F(t)$. Write $F(t) = t^r\tilde{F}(t)$, where $\tilde{F}(t)$ has non-zero constant term. It is a nice exercise to show that $\tilde{F}(t)$ has a multiplicative inverse in the formal power series ring $\overline{K}[\![t]\!]$. So the ideal generated by $F(t)$ is equal to (t^r). Finally, we see that $\overline{K}[\![t]\!]/(t^r)$ has dimension r, since it has $1,t,\ldots,t^{r-1}$ as a basis. □

Lemma 0.123. *Suppose $\mathcal{C}$ and $\mathcal{C}'$ are two curves intersecting at a point $P \in \mathcal{C}(K) \cap \mathcal{C}'(K)$. Suppose moreover that P is a nonsingular point on both curves. Then the intersection multiplicity at P is > 1 if and only if the tangent lines to $\mathcal{C}$ and $\mathcal{C}'$ at P coincide.*

Proof. Translating x and y, we may assume that $P = (0,0)$. If $f(x,y)$ is the equation for $\mathcal{C}$, suppose WLOG that $\lambda = \frac{\partial f}{\partial y}(P) \neq 0$ (otherwise we can swap the roles of x and y). Then the tangent line to $\mathcal{C}$ at P has equation $y = -\lambda^{-1}\frac{\partial f}{\partial x}(P)x$. In this situation, the natural map $K[\![x]\!] \to K[\![x,y]\!]/(f(x,y))$ is an isomorphism, with inverse given by mapping y to a power series $Y(x)$ of the form $Y(x) = -\lambda^{-1}\frac{\partial f}{\partial x}(P)x +$ higher order terms. (This can be proved by a version of Hensel's lemma, which we will see in the course, and is the implicit function theorem for formal power series.) So we have to compute the dimension of the quotient $K[\![x]\!]/(g(x,Y(x)))$. As in the proof of Lemma 0.122, this is given by the multiplicity of 0 as a root of $g(x,Y(x))$. It is > 1 if and only if the linear part $ax + by$ of $g(x,y)$ satisfies $a - b\lambda^{-1}\frac{\partial f}{\partial x}(P) = 0$. On the other hand, the tangent to $\mathcal{C}'$ at P has equation $ax + by = 0$. A short calculation shows that this tangent is the same as the tangent for $\mathcal{C}$ if and only if we do indeed have $a - b\lambda^{-1}\frac{\partial f}{\partial x}(P) = 0$ (necessarily with b nonzero, since P is a nonsingular point of $\mathcal{C}'$). □

In the situation of the above proof, if $\frac{\partial f}{\partial y}(P) \neq 0$ and $\frac{\partial g}{\partial y}(P) \neq 0$, we can moreover compute the intersection multiplicity using the power series $Y(x), \tilde{Y}(x)$ which respectively satisfy $f(x,Y(x)) = 0, g(x,\tilde{Y}(x)) = 0$. Indeed, the ideals generated by $f(x,y)$ and $g(x,y)$ in $K[\![x,y]\!]$ are equal to $(y - Y(x))$ and $(y - \tilde{Y}(x))$ respectively, so $K[\![x,y]\!]/(f(x,y),g(x,y)) = K[\![x,y]\!]/(y - Y(x), y - \tilde{Y}(x)) = K[\![x]\!]/(Y(x) - \tilde{Y}(x))$. So the multiplicity is the order of vanishing at $x = 0$ of $Y(x) - \tilde{Y}(x)$. When K has characteristic 0, we deduce that the curves intersect with multiplicity $\geq r$ if $\frac{\mathrm{d}^i y}{\mathrm{d}x^i}(P)$ on $\mathcal{C} = \frac{\mathrm{d}^i y}{\mathrm{d}x^i}(P)$ on $\mathcal{C}'$ for all $1 \leqslant i \leqslant r-1$ (via the Taylor expansion formula, these derivatives determine the power series $Y(x)$ and $\tilde{Y}(x)$ up to degree $r-1$).

Example 0.124. Let $\mathcal{C} : y^2 = x^3 + 2x + 1$ and $\mathcal{D} : y = x + 1$. On substituting $\mathcal{D}$ into $\mathcal{C}$ we see that the x-coordinate of any point of intersection must satisfy $(x+1)^2 = x^3 + 2x + 1$, and so $x^2(x-1) = 0$, giving only $x = 0, 1$ as possibilities. Substituting $x = 0$ in $\mathcal{D}$ gives $y = 1$; substituting $x = 1$ in $\mathcal{D}$ gives $y = 2$. So, the only possible points of intersection are $(0,1)$ and $(1,2)$ [and these do indeed lie on $\mathcal{C}$ and $\mathcal{D}$]. It also follows from Lemma 0.122 that the intersection multiplicities at these points are 2 and 1 respectively.

Comment 0.125. For more complicated examples, we cannot always find the points of intersection by a straightforward substitution of one equation into the other. Given two curves $\mathcal{C} : f(x,y) = 0$ of degree m and $\mathcal{D} : g(x,y) = 0$ of degree n, a systematic approach to finding the points of intersection is possible via resultants. One initially picks one of the variables, y say, and computes the resultant of $f(x,y)$ and $g(x,y)$, regarded as polynomials in y, by writing them as: $f(x,y) = f_m(x)y^m + \ldots + f_0(x)$, and similarly for $g(x,y)$. The matrix in Definition 0.105 will have entries that are polynomials in x, and consideration of the degrees of these polynomials shows that the resultant of $f(x, y)$ and $g(x, y)$ (regarded as polynomials in y) will be a polynomial in x of degree at most mn. Any point of intersection of $\mathcal{C}$ and $\mathcal{D}$ must have x-coordinate which is a root of the at-most-degree-mn polynomial. For each value of x, one can then substitute back into $\mathcal{C}$ and $\mathcal{D}$ to find the corresponding y-coordinates.

Projective Space

There are several respects in which affine space is unsatisfying. Consider, for example, the true statement in affine space: two distinct lines meet at exactly one point, except when parallel. It would be much nicer to have a cleaner statement, in which we remove 'except when parallel'. Intuitively, parallel lines intersect 'at infinity', given that the point of intersection shoots off to infinity as two lines become closer and closer to parallel. Similarly, consider the affine curves: $\mathcal{C} : y^2 = x^3 + 1$ and $D : y = x + 1$; these meet at the points $(-1, 0), (0, 1), (2, 3)$, each with multiplicity 1. On trying other lines in place of $\mathcal{D}$, one typically finds again that there are 3 points of intersection (when counted with multiplicity). An apparent exception is $\mathcal{D} : x = 0$, which intersects $\mathcal{C}$ only at $(0, 1)$ and $(0, -1)$, and this is true for any vertical line. We seem to have a rule: any line intersects $\mathcal{C}$ at exactly 3 points (counted with multiplicity) except when the line is vertical. Again, we would like a cleaner statement, in which we remove 'except when the line is vertical'. Again, the third point of intersection seems to be 'at infinity'.

Points at infinity are intuitively points (x, y) where there is a denominator of 0. We cannot express this idea using only pairs (x, y), where x, y lie in a field K. A natural approach is to write: $x = X/Z, y = Y/Z$ and identify the point (x, y) with the triple (X, Y, Z). As long as $Z \neq 0$, we can go in the other direction from the triple (X, Y, Z) to (x, y). Note that, for any $k \in K^*$, the triple (kX, kY, kZ) corresponds to $(kX/kZ, kY/kZ) = (X/Z, Y/Z) = (x, y)$, and so we impose a relation, that two triples are regarded as being the same if they are nonzero scalar multiples of each other. Subject to this relation, there is then a $1 - 1$ correspondence between (x, y) and triples (X, Y, Z) with $Z \neq 0$. On the other hand, the triples (X, Y, Z) with $Z = 0$ do not correspond to any $x, y \in K$, and such triples give us a way of describing formally these new points at infinity.

Definition 0.126. Let K be a field. $\mathbb{P}^n(K) = \{(x_0, \ldots, x_n) : x_0, \ldots, x_n \in K, \text{ not all } 0\}$, subject to the relation that $(x_0, \ldots x_n) = (y_0, \ldots, y_n)$ in $\mathbb{P}^n(K)$ if there exists $r \in K, r \neq 0$, such that $(y_0, \ldots, y_n) = (rx_0, \ldots rx_n)$. $\mathbb{P}^n(K)$ is called *projective* n-space over K.

Example 0.127. $(1, 2, 3) = (3, 6, 9)$ in $\mathbb{P}^2(\mathbb{Q})$. [N.B. $(0, 0, 0) \notin \mathbb{P}^2(\mathbb{Q})$.]

Definition 0.128. A *polynomial in n projective variables* is an $(n+1)$-variable homogeneous polynomial. A *projective curve* in $\mathbb{P}^2$ is defined by a homogeneous polynomial in 3 variables $F(X,Y,Z)=0$, for example, $X^3+Y^3-Z^3=0$.

Definition 0.129. Let $\mathcal{C} : f(x,y) = 0$ be an (affine) curve. The *homogenisation* of $\mathcal{C}$ is the projective curve $F(X,Y,Z)=0$ of the same degree as $f(x,y)$, with the property that $F(x,y,1)=f(x,y)$. A point (X_0,Y_0,Z_0) on $F(X,Y,Z)=0$ with $Z_0=0$ is called a *point at infinity* on $\mathcal{C}$. When $Z_0 \neq 0$, the point (X_0,Y_0,Z_0) corresponds to $(X_0/Z_0, Y_0/Z_0)$ on $f(x,y)=0$.

Example 0.130. Let $\mathcal{C} : y^2 = 4x^2+1$, so that $f(x,y)=y^2-4x^2-1=0$. The associated projective curve (the homogenisation) is: $Y^2 = 4X^2+Z^2$ [so that $F(X,Y,Z)=Y^2-4X^2-Z^2$]. The two points at infinity are: $(1,2,0)$ and $(1,-2,0)$.

Example 0.131. For the curve $\mathcal{C} : y^2 = x^3+1$, the associated projective curve is $ZY^2 = X^3+Z^3$. To find the points at infinity (the points where $Z=0$), substitute $Z=0$ into the equation, giving $X^3=0$ and so $X=0$. This forces $Y \neq 0$ [since $(0,0,0)$ is not allowed as a point in $\mathbb{P}^2$]. So, the points at infinity are of the form $(0,Y,0)$, where $Y \neq 0$. But these are all the same in $\mathbb{P}^2$, since they are scalar multiples of each other; therefore this is exactly one point at infinity, which we can represent by $(0,1,0)$, say.

Comment 0.132. Two distinct affine lines $a_1x+b_1y+c_1=0$ and $a_2x+b_2y+c_2=0$ meet at exactly one point, except when parallel. For example, $x+y+2=0$ and $x+y+3=0$ do not intersect. For projective lines, the rule is the same, but we can remove the phrase 'except when parallel'. For example, the projective lines $X+Y+2Z=0$ and $X+Y+3Z=0$ have $(1,-1,0)$ as the unique point of intersection.

Definition 0.133. A projective curve $F(X,Y,Z)=0$ has a *singularity* at (X_0,Y_0,Z_0) when:

$$F(X_0,Y_0,Z_0) = \tfrac{\partial F}{\partial X}(X_0,Y_0,Z_0) = \tfrac{\partial F}{\partial Y}(X_0,Y_0,Z_0) = \tfrac{\partial F}{\partial Z}(X_0,Y_0,Z_0) = 0.$$

Comment 0.134. Note that, by multiplying through by denominators, we can take rational maps and birational transformations between projective curves to be of the form:

$$\underline{\phi}(X,Y,Z) = \big(\phi_1(X,Y,Z), \phi_2(X,Y,Z), \phi_3(X,Y,Z)\big),$$

where ϕ_1, ϕ_2, ϕ_3 are homogeneous polynomials, rather than rational functions.

Comment 0.135. Suppose that two projective curves $F(X,Y,Z)=0$ and $G(X,Y,Z)=0$ have a point of intersection (X_0,Y_0,Z_0). The multiplicity of intersection can always be computed by using some associated affine curve. At least one of X_0, Y_0, Z_0 must be nonzero, since $(0,0,0)$ is not allowed in $\mathbb{P}^2$. If $Z_0 \neq 0$ then the multiplicity of intersection is the same as that of $(X_0/Z_0, Y_0/Z_0)$ on the affine curves $F(x,y,1)=0$ and $G(x,y,1)=0$ [here, $x =$

$X/Z, y = Y/Z]$. If $Y_0 \neq 0$ then one can use $F(x,1,z), G(x,1,z)$, where $x = X/Y, z = Z/Y$. If $X_0 \neq 0$ then one can use $F(1,y,z), G(1,y,z)$, where $y = Y/X, z = Z/X$.

Theorem 0.136. *(Bézout's Theorem). Two projective curves, with no common component [i.e. with no common non-constant factor] of degrees m, n intersect at precisely mn points, counted with multiplicity.*

Example 0.137. The projective curves $ZY^2 = X^3 + Z^3$ and $X = 0$ intersect at the points $(0,1,1), (0,-1,1), (0,1,0)$, each with multiplicity 1.

Elliptic Curves

The following overlaps with the material that will be presented during the first week of the Part C Elliptic Curves lecture course.

Curves can be classified according to a property called *genus*, which is invariant under birational equivalence. We shall not go into the technicalities of what precisely is meant by genus, and its properties, which would be an entire lecture course in its own right. The simplest type are curves of genus 0, which can be defined by quadratic and linear equations. Recall from Theorem 0.119 that any conic with a rational point can be parametrised.

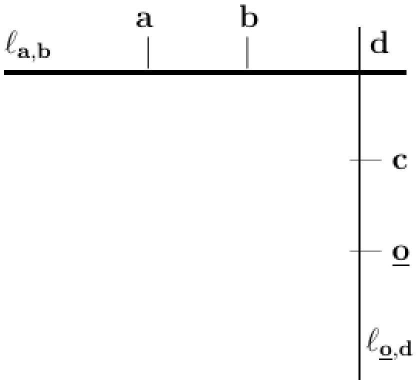
Curves of genus 1 are the next natural class of curves to consider; they are, in a sense, the next 'simplest' type of curve after conics. Please don't confuse 'elliptic curves' (which are of genus 1) with ellipses (which are of genus 0). The classical terminology comes from a relationship between cubic curves and elliptic integrals, which were much studied in the 19th century. It can be shown that a curve of genus 1 is not parametrisable. An *elliptic curve* over K is defined to be a nonsingular projective curve of genus 1, defined over K, together with a K-rational point on the curve. It can also be shown that any curve of genus 1 is birationally equivalent over K to a nonsingular projective cubic curve.

For the purposes of this lecture course, you can forget about the term 'genus' and simply take this as the definition of an elliptic curve, as follows.

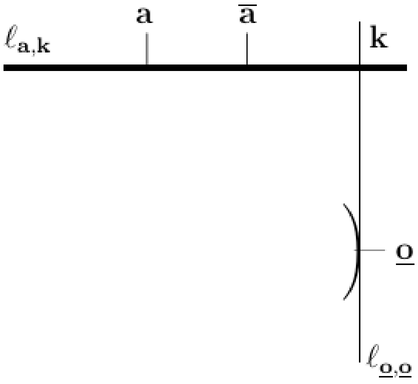
Definition 0.138. An elliptic curve over a field K is a nonsingular projective cubic curve, defined over K, with a specified K-rational point.

A property which makes elliptic curves of particular interest is the fact that there is a natural way to combine any two points $\mathbf{a}, \mathbf{b}$ on the curve to obtain a third point $\mathbf{a} + \mathbf{b}$. The following defines what we mean by $\mathbf{a} + \mathbf{b}$ and by $-\mathbf{a}$.

Definition 0.139. Let $\mathcal{C} : F(X,Y,Z) = 0$ be an elliptic curve $/K$ (the notation $/K$ means 'defined over K'; that is, all of the coefficients of $\mathcal{C}$ are in the field K). So, $\mathcal{C}$ is a nonsingular projective cubic curve, with a K-rational point, which we shall denote $\underline{\mathbf{o}}$. For any two points $\mathbf{a}, \mathbf{b}$ on $\mathcal{C}$ (defined over a common extension field L/K), let $\ell_{\mathbf{a},\mathbf{b}}$ denote the line which meets $\mathcal{C}$ at $\mathbf{a}, \mathbf{b}$ (if $\mathbf{a}, \mathbf{b}$ are distinct then $\ell_{\mathbf{a},\mathbf{b}}$ is the unique line through $\mathbf{a}, \mathbf{b}$; if $\mathbf{a} = \mathbf{b}$ then $\ell_{\mathbf{a},\mathbf{b}}$ is the line tangent to $\mathcal{C}$ at $\mathbf{a} = \mathbf{b}$).


$\mathbf{a}$
$\mathbf{b}$
$\ell_{\mathbf{a},\mathbf{b}}$
$\mathbf{d}$
$\mathbf{c}$
$\underline{\mathbf{o}}$
$\ell_{\underline{\mathbf{o}},\mathbf{d}}$

Let $\ell_{\mathbf{a},\mathbf{b}}$ denote the line which meets $\mathcal{C}$ at $\mathbf{a}, \mathbf{b}$.
Then $\ell_{\mathbf{a},\mathbf{b}}$ and $\mathcal{C}$ have 3 points of intersection (Bézout).
Let $\mathbf{d}$ be the third point of intersection between $\mathcal{C}$ and $\ell_{\mathbf{a},\mathbf{b}}$.
Now, let $\ell_{\underline{\mathbf{o}},\mathbf{d}}$ denote the line which meets $\mathcal{C}$ at $\underline{\mathbf{o}}$ and $\mathbf{d}$.
Let $\mathbf{c}$ be the third point of intersection between $\mathcal{C}$ and $\ell_{\underline{\mathbf{o}},\mathbf{d}}$.
Define $\mathbf{a} + \mathbf{b} = \mathbf{c}$.


$\mathbf{a}$
$\bar{\mathbf{a}}$
$\ell_{\mathbf{a},\mathbf{k}}$
$\mathbf{k}$
$\underline{\mathbf{o}}$
$\ell_{\underline{\mathbf{o}},\underline{\mathbf{o}}}$

Let $\ell_{\underline{\mathbf{o}},\underline{\mathbf{o}}}$ be the line tangent to $\mathcal{C}$ at $\underline{\mathbf{o}}$.
Let $\mathbf{k}$ be the third point of intersection between $\mathcal{C}$ and $\ell_{\underline{\mathbf{o}},\underline{\mathbf{o}}}$.
Now, let $\ell_{\mathbf{a},\mathbf{k}}$ be the line which meets $\mathcal{C}$ at $\mathbf{a}$ and $\mathbf{k}$.
Let $\bar{\mathbf{a}}$ be the third point of intersection between $\mathcal{C}$ and $\ell_{\mathbf{a},\mathbf{k}}$.
Define $-\mathbf{a}$ to be $\bar{\mathbf{a}}$.

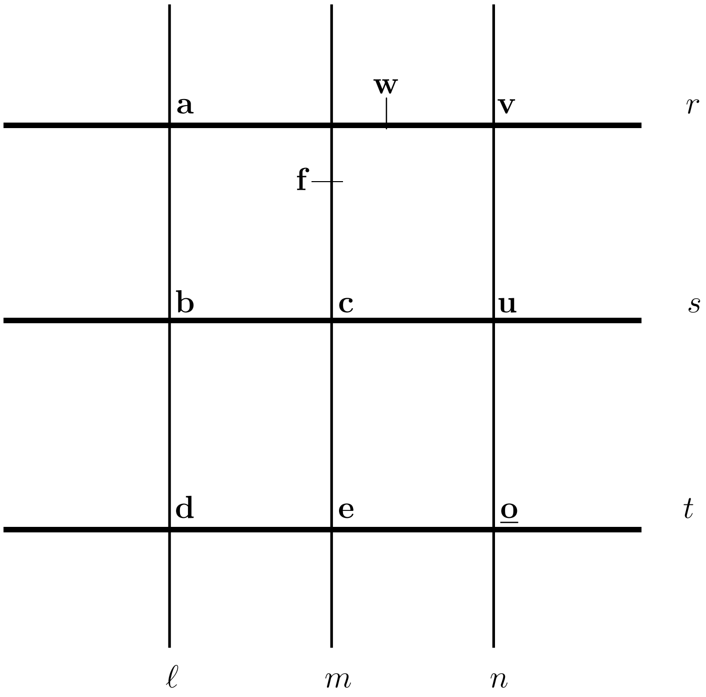
We shall soon show that $\mathbf{a} + \mathbf{b}$ is a commutative group law on the points on $\mathcal{C}$, with identity $\underline{\mathbf{o}}$ and the inverse of $\mathbf{a}$ given by $-\mathbf{a}$. First we need the following technical lemma.

Lemma 0.140. *Let $P_1, \ldots, P_8$ be such that no 4 points lie on a line and no 7 points lie on a conic. Then there exists a unique point P_9 which is a 9th point of intersection of any two cubics passing through $P_1, \ldots, P_8$.*

Proof (Sketch). Any projective cubic curve $\mathcal{C} : F(X, Y, Z) = f_1X^3 + f_2Y^3 + \ldots + f_{10}XYZ = 0$ has 10 coefficients $f_1, \ldots, f_{10}$, and the condition that $P_1, \ldots, P_8$ lie on $\mathcal{C}$ gives linear equations which must be satisfied by the coefficients [the condition that no 4 lie on a line and no 7 lie on a conic give that the equations are linearly independent]. So, there are two free parameters λ, μ and F can be written: $F = \lambda F_1 + \mu F_2$, where F_1, F_2 are fixed cubic through $P_1, \ldots, P_8$. By Bézout's Theorem, F_1, F_2 have a 9th point of intersection P_9, which must lie on any $\lambda F_1 + \mu F_2$. □

Theorem 0.141. *Let $\mathcal{C}$ be an elliptic curve $/K$, with K-rational point $\underline{\mathbf{o}}$. Then $\mathbf{a}+\mathbf{b}$, as in Definition 0.139, gives a commutative group law on the points on $\mathcal{C}$, with identity $\underline{\mathbf{o}}$. The inverse of $\mathbf{a}$ is given by the point $-\mathbf{a}$, constructed in in Definition 0.139. Further, the K-rational points $\mathcal{C}(K)$ form a subgroup, called the Mordell-Weil group.*

Proof It is easy to show commutativity, the fact that $\underline{\mathbf{o}}$ is the identity, and the fact that $-\mathbf{a}$ is the inverse of $\mathbf{a}$. The only difficult problem is associativity. In order to prove associativity, consider the following diagram.



Here, r, s, t, ℓ, m, n are lines. On each line, the labelled points are the points of intersection between $\mathcal{C}$ and that line. From the construction of Definition 0.139:

$$\mathbf{a}+\mathbf{b}=\mathbf{e},$$

and so:

$$(\mathbf{a}+\mathbf{b})+\mathbf{c}=\text{3rd point of intersection on } \ell_{\underline{\mathbf{o}},\mathbf{f}}.$$

Similarly:

$$\mathbf{b}+\mathbf{c}=\mathbf{v},$$

$$\mathbf{a}+(\mathbf{b}+\mathbf{c})=\text{3rd point of intersection on } \ell_{\underline{\mathbf{o}},\mathbf{w}}.$$

To show $(\mathbf{a}+\mathbf{b})+\mathbf{c}=\mathbf{a}+(\mathbf{b}+\mathbf{c})$, it is sufficient to show that $\mathbf{f}=\mathbf{w}$. Let $F_1=\ell mn$ and $F_2=rst$, both of which are cubic curves.

$\mathcal{C}$ and F_1 have 8 common points: $\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d},\mathbf{e},\mathbf{u},\mathbf{v},\underline{\mathbf{o}}$.

$\mathcal{C}$ and F_2 also have these 8 common points: $\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d},\mathbf{e},\mathbf{u},\mathbf{v},\underline{\mathbf{o}}$.

From Lemma 0.140, the 9th point of intersection of $\mathcal{C}$ and F_1 must be the same as the 9th point of intersection of $\mathcal{C}$ and F_2; that is, $\mathbf{f}=\mathbf{w}$, as required.

Hence, $+$ is a commutative group law.

It remains to show that $\mathcal{C}(K)$ is a subgroup. We are given that $\underline{\mathbf{o}} \in \mathcal{C}(K)$. Let $\mathbf{a}, \mathbf{b} \in \mathcal{C}(K)$. It is sufficient to show that $\mathbf{a}+\mathbf{b} \in \mathcal{C}(K)$ and that $-\mathbf{a} \in \mathcal{C}(K)$. In the following, we shall write points and equations of lines in affine form, as a shorthand notation for the corresponding projective points and lines (their homogenisations).

Let $\mathbf{a} = (x_1, y_1)$ and $\mathbf{b} = (x_2, y_2)$, where $x_1, y_1, x_2, y_2 \in K$. Then the line through $\mathbf{a}, \mathbf{b}$ is (in affine form) $\ell_{\mathbf{a},\mathbf{b}} : y = \ell x + m$, where $\ell = \frac{y_1 - y_2}{x_1 - x_2} \in K$ and $m = \frac{x_1 y_2 - x_2 y_1}{x_1 - x_2} \in K$. Substitute $y = \ell x + m$ into the cubic equation for $\mathcal{C}$ to get; $\phi(x) = x^3 + c_2 x^2 + c_1 x + c_0 = 0$, defined over K. Let $\phi(x) = (x - x_1)(x - x_2)(x - x_3)$ be the factorisation of $\phi(x)$. Then x_1, x_2, x_3 are the 3 roots of ϕ and so $x_1 + x_2 + x_3 = -c_2$, giving: $x_3 = -c_2 - x_1 - x_2 \in K$ and $y_3 = \ell x_3 + m \in K$. The line $\ell_{\mathbf{a},\mathbf{b}}$ then meets $\mathcal{C}$ at $\mathbf{a}, \mathbf{b}, \mathbf{d} = (x_3, y_3) \in \mathcal{C}(K)$. The same argument shows that the line $\ell_{\underline{\mathbf{o}},d}$ through $\underline{\mathbf{o}}, \mathbf{d}$ has 3rd point of intersection $\mathbf{c}$ which is also in $\mathcal{C}(K)$. But $\mathbf{c} = \mathbf{a} + \mathbf{b}$ and so we have shown that $\mathbf{a} + \mathbf{b} \in \mathcal{C}(K)$. A similar argument shows that if $\mathbf{a} \in \mathcal{C}(K)$ then $-\mathbf{a} \in \mathcal{C}(K)$. Hence $\mathcal{C}(K)$ is a subgroup, as required. □

It is apparent that, in the above proof, we have dealt with the 'typical' case, where none of our points are repeated (for the proof of associativity), and none are at infinity (for the proof that $\mathcal{C}(K)$ is a subgroup, since the points were written in affine form). It is straightforward to check these special cases; we shall not bother to do so here.

Comment 0.142. By an elliptic curve, we shall always mean a projective curve, but often write the equation in affine form. Note that, whichever way it is written, we are always referring to the projective curve. For example, if we say 'let $\mathcal{C} : y^2 = x^3 + 3$ be an elliptic curve', it should be understood that this is a shorthand notation for the corresponding projective curve $ZY^2 = X^3 + 3Z^3$.

It can be shown that any elliptic curve over K can be birationally transformed over K to *Weierstrass form*, which is quadratic in one of the variables (y, say) and cubic in the other variable (x, say):

$$\mathcal{E} : y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6.$$

Indeed, as long as we avoid fields with small characteristic, we can simplify the equation still further.

Theorem 0.143. *Let K be a field satisfying char$(K) \neq 2$ [recall – this means that $1+1 \neq 0$]. Then any elliptic curve over K is birationally equivalent over K to the form $y^2 =$ cubic in x. If char$(K) \neq 2, 3$ then the curve is birationally equivalent over K to a curve of the form $y^2 = x^3 + Ax + B$.*

Comment 0.144. When $char(K) \neq 2, 3$, we shall typically take our elliptic curves to have the form

$$\mathcal{E} : y^2 = x^3 + Ax + B, \text{ where } A, B \in K,$$

which should be regarded as shorthand for the projective curve $ZY^2 = X^3 + AXZ^2 + BZ^3$. Sometimes it will be convenient to include the x^2 term. Since $\mathcal{E}$ is nonsingular, we must have $\Delta = 4A^3 + 27B^2 \neq 0$, as was shown in Example 0.111 (note the assumption there that $char(K) \neq 2$). The notation $\Delta = 4A^3 + 27B^2$ is standard.

It is conventional to choose $\underline{\mathbf{o}} = (0, 1, 0)$, the point at infinity, as the identity [we shall always take $\underline{\mathbf{o}} = (0, 1, 0)$ unless otherwise stated]. Note that the line $Z = 0$ meets $\mathcal{E}$ at $\underline{\mathbf{o}}$ three times (such a point is called an *inflexion*). Given a point $\mathbf{a} = (X, Y, Z)$, if we take the line through $\mathbf{a}$ and $\underline{\mathbf{o}} = (0, 1, 0)$ then the third point of intersection is $(X, -Y, Z)$, which must then be $-\mathbf{a}$. In affine form:

$$-(x, y) = (x, -y).$$

This gives an easy rule for finding the inverse of a point, under the group law, namely: the inverse of $\mathbf{a}$ is its reflection in the x-axis.

So, for an elliptic curve $\mathcal{E}$ written in the form $y^2 =$ cubic in x, the points are $\underline{\mathbf{o}}$ (the point at infinity) and the affine points (x, y), and the group law has a simpler description:

Let $\mathbf{d} = (x_3, y_3)$ the 3rd point of intersection of $\mathcal{E}$ and $\ell_{\mathbf{a},\mathbf{b}}$.

Then $\mathbf{a} + \mathbf{b} = (x_3, -y_3)$, the reflection of $\mathbf{d}$ in the x-axis.

Comment 0.145. Note that any $y^2 =$ cubic in x, over $\mathbb{Q}$, can be birationally transformed to a curve of the form $y^2 = x^3+Ax+B$, where $A, B \in \mathbb{Z}$, using only linear changes in x, y. For example, starting with $y^2 = 5x^3+6x^2+4x+2$, the birational transformation $(x, y) \mapsto (5x, 5y)$ [with inverse $(x, y) \mapsto (\frac{x}{5}, \frac{y}{5})$] takes this curve to $\big(\frac{y}{5}\big)^2 = 5\big(\frac{x}{5}\big)^3 + 6\big(\frac{x}{5}\big)^2 + 4\big(\frac{x}{5}\big) + 2$, that is: $y^2 = x^3+6x^2+20x+50$. Then the birational transformation $(x, y) \mapsto (x+6/3, y) = (x+2, y)$ [with inverse $(x, y) \mapsto (x-2, y)$] takes this curve to: $y^2 = (x-2)^3+6(x-2)^2+20(x-2)+50 = x^3 + 8x + 26$.

We illustrate the group law with the following computation.

Example 0.146. Let $\mathcal{E} : y^2 = x^3 + 1$. Let us compute $\mathbf{a} + \mathbf{b}$, where $\mathbf{a} = (x_1, y_1) = (-1, 0)$ and $\mathbf{b} = (x_2, y_2) = (0, 1)$.

The line through $\mathbf{a}, \mathbf{b}$ is $\ell_{\mathbf{a},\mathbf{b}} : y = x + 1$. Substituting this into $\mathcal{E}$, we see that the x-coordinate of any point of intersection satisfies: $(x + 1)^2 = x^3 + 1$, and so:

$$x^3 - x^2 - 2x = 0. \qquad (*)$$

We are looking for (x_3, y_3), the 3rd point of intersection of $\mathcal{E}$ and $\ell_{\mathbf{a},\mathbf{b}}$. We first find x_3; note that x_1, x_2, x_3 must be the roots of $(*)$.

Method A (for finding x_3). Since the roots of $(*)$ are x_1, x_2, x_3, it follows that $x^3 - x^2 - 2x = (x - x_1)(x - x_2)(x - x_3)$; equating coefficients of x^2 gives that:

$$x_1 + x_2 + x_3 = -(\text{coefficient of } x^2 \text{ in } (*)) = -(-1) = 1,$$

so that $(-1) + 0 + x_3 = 1$, giving $x_3 = 2$.
Method B (for finding x_3). Factorise $(*)$ to give: $x(x+1)(x-2)$, whose roots are: $0, -1, 2$. Two of these are the already known $x_1 = -1, x_2 = 0$, and so x_3 must be the remaining root: $x_3 = 2$.

Having found x_3 (by either method), we use the equation of $\ell_{\mathbf{a},\mathbf{b}}$ to compute $y_3 = x_3 + 1 = 3$. In summary: $\mathcal{E}$ and $\ell_{\mathbf{a},\mathbf{b}}$ intersect at: $(-1, 0), (0, 1), (2, 3)$, and so $(-1, 0) + (0, 1) + (2, 3) = \underline{\mathbf{o}}$.

Finally, this gives: $(-1, 0) + (0, 1) = -(2, 3) = (2, -3)$, using the rule that negation is given by reflection in the x-axis.

One can also obtain an explicit general formula for the group law.

Lemma 0.147. *Let $\mathcal{E} : y^2 = x^3 + Ax + B$, where $A, B \in K$, with (as usual) $\underline{\mathbf{o}} =$ the point at infinity. Let $(x_3, y_3) = (x_1, y_1) + (x_2, y_2)$.*
Case 1. *When $x_1 \neq x_2$ then:*

$$x_3 = \frac{x_1 x_2^2 + x_1^2 x_2 + A(x_1 + x_2) + 2B - 2y_1 y_2}{(x_1 - x_2)^2}, \quad y_3 = -\ell x_3 - m,$$

$$\textit{where: } \ell = \frac{y_1 - y_2}{x_1 - x_2}, \quad m = \frac{x_1 y_2 - x_2 y_1}{x_1 - x_2}.$$

Case 2. *When $(x_1, y_1) = (x_2, y_2)$ then $(x_3, y_3) = (x_1, y_1) + (x_1, y_1)$ [which can be written as $2(x_1, y_1)$], and:*

$$x_3 = \frac{x_1^4 - 2Ax_1^2 - 8Bx_1 + A^2}{4y_1^2} = \frac{x_1^4 - 2Ax_1^2 - 8Bx_1 + A^2}{4(x_1^3 + Ax_1 + B)}, \quad y_3 = -\ell x_3 - m,$$

$$\textit{where: } \ell = \frac{3x_1^2 + A}{2y_1}, \quad m = \frac{-x_1^3 + Ax_1 + 2B}{2y_1}.$$

Proof:
Case 1. When $x_1 \neq x_2$, the line through $(x_1,y_1),(x_2,y_2)$ is $y = \ell x + m$, where $\ell = \frac{y_1-y_2}{x_1-x_2}$, $m = \frac{x_1y_2-x_2y_1}{x_1-x_2}$. Replacing y by $\ell x+m$ in $y^2 = x^3+Ax+B$, we see that x_1,x_2,x_3 satisfy $(\ell x+m)^2 = x^3+Ax+B$, and so:

$$x^3 - \ell^2x^2 + \text{terms of lower degree} = 0.$$

Hence $x_1+x_2+x_3 = -(\text{coefficient of } x^2) = \ell^2$, and so $x_3 = \ell^2 - x_1 - x_2 = \left(\frac{y_1-y_2}{x_1-x_2}\right)^2 - x_1 - x_2$. On expanding this expression and using the fact that each $y_i^2 = x_i^3+Ax_i+B$, for $i=1,2$ [since $(x_1,y_1),(x_2,y_2)$ lie on $\mathcal{E}$], we obtain the given formula for x_3. The 3 points of intersection of $\mathcal{E}$ and $y=\ell x+m$ are then $(x_1,y_1),(x_2,y_2),(x_3,\ell x_3+y_3)$, so that

$$(x_1,y_1)+(x_2,y_2)+(x_3,\ell x_3+y_3) = \underline{\mathbf{o}},$$

giving $(x_1,y_1)+(x_2,y_2) = -(x_3,\ell x_3+y_3) = (x_3,-\ell x_3-y_3)$, as required.
Case 2. When $(x_1,y_1) = (x_2,y_2)$, we should first compute the tangent to $\mathcal{E}$ at (x_1,y_1). Using $2yy' = 3x^2+A$, we see that the slope of the tangent at (x_1,y_1) is $\ell = \frac{3x_1^2+A}{2y}$. The equation of the tangent is then $y = \ell x+m$, where $m = y_1 - \ell x_1 = \frac{-x_1^3+Ax_1+2B}{2y_1}$. As in case 1, $x_3 = \ell^2 - x_1 - x_2$, which simplifies to the given formula, and $y_3 = -\ell x_3 - m$. □

The only case not considered above is when $(x_1,y_1) = (x_2,-y_2)$, but it this case the points are inverses and so $(x_1,y_1)+(x_2,y_2) = \underline{\mathbf{o}}$. The above formulas give an alternative method for computing the group law, although in practice it often turns out to be easier to compute the group law from first principles, as in Example 0.146.

This lecture course will concentrate on the number theoretic properties of elliptic curves, in particular results about $\mathcal{E}(\mathbb{Q})$, when $\mathcal{E}$ is an elliptic curve defined over $\mathbb{Q}$. The number theoretic properties of elliptic curves is a substantial area of research. Recently, elliptic curves were use in the proof of Fermat's conjecture, that there are no solutions in positive integers to the equation $a^n + b^n = c^n$, for any $n \geqslant 3$. It is sufficient to prove the result for $n = p$ prime. If there were a solution, then the elliptic curve $y^2 = x(x-a^p)(x+b^p)$ [often called the 'Frey curve'] cannot be modular, and so would disobey the Taniyama-Shimura conjecture [that all elliptic curves are modular], a conjecture that has recently been proved (alas, the proof is far beyond the scope of any first lecture course on elliptic curves).

The main goal of the lecture course will be to gain some understanding of the properties of the Mordell-Weil group $\mathcal{E}(\mathbb{Q})$. We shall begin by giving basic definitions and properties that apply to an elliptic curve over any field K [some of the first week will overlap with the above introduction to elliptic curves]. We shall then explore elliptic curves over finite fields, and over the field of p-adic numbers $\mathbb{Q}_p$. This will include the development of the *formal group* of an elliptic curve, which describes how to expand the group law as a power series in a neighbourhood of the identity. We shall see that this exploration over $\mathbb{Q}_p$ gives us results that help us to understand $\mathcal{E}(\mathbb{Q})$. We shall then discuss the subgroup of $\mathcal{E}(\mathbb{Q})$ consisting of points of finite order (the *torsion subgroup*). For elliptic curves with a rational point of order 2, there is a map (a 2-*isogeny*) to an associated elliptic curve, which is a homomorphism and has kernel of order 2. We shall describe properties of this isogeny and use it to prove that such curves satisfy the *Weak Mordell-Weil Theorem*, that $\mathcal{E}(\mathbb{Q})/2\mathcal{E}(\mathbb{Q})$ is always finite. We shall then describe the theory of heights on elliptic curves [essentially, a measure of the 'size' of a point on an elliptic curve], and use this to prove the climactic result of the lecture course: the *Mordell-Weil Theorem*, that $\mathcal{E}(\mathbb{Q})$ is always finitely generated. As an added extra at the end will be a short section on the relevance of elliptic curves to cryptography.

The sections will be as follows.

Section 1. The Group Law on an Elliptic Curve.

The group law, associativity, examples, torsion, the number of points on an elliptic curve over a finite field.

Section 2. The p-adic Numbers $\mathbb{Q}_p$.

The p-adic valuation on $\mathbb{Q}$, the definition of $\mathbb{Q}_p$ as the completion of $\mathbb{Q}$ with respect to this valuation, basic properties of $\mathbb{Q}_p$, Hensel's Lemma for determining when an approximate solution to a univariate polynomial equation lifts to a actual solution.

Section 3. The Reduction Map on an Elliptic Curve.

The reduction map on an elliptic curve, the use of Hensel's Lemma to decide when a given point over $\mathbb{F}_p$ has a preimage (a 'lift') under the reduction map.

Section 4. Formal Groups.

General 1-parameter formal groups and their properties, the invariant differential and formal logarithm, the formal group of an elliptic curve.

Section 5. Global Torsion.

Injectivity of the reduction map on torsion, the Nagell-Lutz Theorem [which gives an effective procedure for computing the torsion group of $\mathcal{E}(\mathbb{Q})$].

Section 6. A 2-isogeny on an Elliptic Curve.

Elliptic curves with a point of order 2, a 2-isogeny from such an elliptic curve to an associated elliptic curve and its properties. The Weak Mordell-Weil Theorem that $\mathcal{E}(\mathbb{Q})/2\mathcal{E}(\mathbb{Q})$ is finite, for the case when an elliptic curve has a point of order 2.

Section 7. The Mordell-Weil Theorem.

Height functions on Abelian groups, height functions on elliptic curves, the Mordell-Weil Theorem, that $\mathcal{E}(\mathbb{Q})$ is finitely generated.

Section 8. Cryptography. Public keys, cryptography, Pollard's $p-1$ method and the elliptic curve method for factorising large integers.

The following are the two main references for the lecture course.

J.W.S. Cassels. *Lectures on Elliptic Curves.* LMS–ST **24**. Cambridge University Press, Cambridge, 1991.

J.H. Silverman. *The Arithmetic of Elliptic Curves.* GTM **106**. Springer-Verlag, 1986.

During the lecture course, I shall refer to these simply as 'Cassels' and 'Silverman', respectively.

The following gives some possible pre-course reading options if you find that you have gaps in your knowledge of any of the pre-requisite material described in Section 0.

W. Keith Nicholson. *Introduction to Abstract Algebra.* (Second Edition, John Wiley, 1999).

Peter J. Cameron. *Introduction to Algebra.* OUP 1998.

Alan Baker. *A Concise Introduction to the Theory of Numbers.* CUP, 1985.

I.M. Niven, H.S. Zuckerman and H.L. Montgomery. *An Introduction to the Theory of Numbers.* Wiley, 1991.

W.A. Sutherland. *Introduction to Metric and Topological Spaces.* OUP, 1975.

Miles Reid. *Undergraduate Algebraic Geometry.* CUP, 1988.
