Introduction to Algebra

Shu Sasaki

9th April 2024

1 Introduction

"*Stand firm in your refusal to remain conscious during algebra. In real life, I assure you, there is no such thing as algebra.*"– Fran Lebowitz

What this course is about?

We axiomatise (=express a theory as a set of axioms) what we know very well (integers, GCD, LCM, Euclid's algorithm, modular arithmetic, complex numbers, polynomials etc), i.e., spot 'common denominators (structures)' of mathematical concepts we have learned (or will have learned), and build/extrapolate a theory out of them. For example, we will see that the set $\mathbb{Z}$ of integers and the set of polynomials in one variable with rational coefficients have the same 'algebraic' structure– they are examples of what we call rings. Similarly, the set of rational numbers (with addition, subtraction, multiplication and division) is 'similar (algebraically)' to the set of Laurent series with rational coefficients– they are examples of fields.

This means that if we can prove a statement about a ring (a set that satisfies a bunch of axioms as $\mathbb{Z}$ does for example), then the assertion would hold for any ring we find. Conversely, any abstract statement can always be specialised to examples. In my opinion, algebra is a concrete subject, contrary to the prevailing opinion amongst those who haven't seen the magic.

"*Algebra is the intellectual instrument which has been created for rendering clear the quantitative aspects of the world*"– Alfred North Whitehead

Related modules: NSF, Number Theory, Group Theory, Cryptography...

Books:
R. Allenby, *Rings, Fields and Groups: Introduction to Abstract Algebra*, Butterworth-Heinemann,
P. Cameron, *Introduction to Algebra*, Oxford University Press,
B. Hartley and T. O. Hawkes, *Rings, Modules and Linear Algebra*, Chapman and Hall,
M. A. Armstrong, *Groups and Symmetry*, Springer-Verlag,
D. J. H. Garling, *Galois theory*, Cambridge University Press,
M. Artin, *Algebra*, Prentice Hall,
NRICH, *Development of Algebra*, https://nrich.maths.org/6485, https://nrich.maths.org/6546

2 Revising bits and bobs from NSF

By $\mathbb{Z}$, we mean the set of integers. By a non-negative (resp. positive) integer, we mean an integer $\geqslant 0$ (resp. $\geqslant 1$). You will never hear 'natural numbers' from me.

2.1 Integer division

Proposition 1. Let a and b be integers. Assume that $b > 0$ (but a can be negative). Then there exist integers q and r such that

$$a = bq + r$$

with $0 \leq r < b$. Moreover, q and r are unique, i.e., if (q_1, r_1) and (q_2, r_2) are two pairs of integers satisfying $a = bq_1 + r_1$ with $0 \leq r_1 < b$ and $a = bq_2 + r_2$ with $0 \leq r_2 < b$ respectively, then $q_1 = q_2$ and $r_1 = r_2$.

Remark. The numbers q and r are referred to as the quotient and remainder when a is divided by b.

Examples.
$a = 100, b = 7$: $100 = 7 \times 14 + 2$.
$a = -100, b = 7$: $-100 = 7 \times (-15) + 5$. Note that q is forced to be a negative integer for the reminder to be in the range $[0, 7)$!
$a = 2, b = 3$: $2 = 3 \times 0 + 2$. It is possible for q to be zero!
$a = -2, b = 3$: $-2 = 3 \times (-1) + 1$.

Proof of the proposition. To see the existence of q and r, let S denote the set of integers of the form $a + sb \geqslant 0$ where s ranges over $\mathbb{Z}$.

We firstly show that S is non-empty. If $a \geqslant 0$, then $a = a + 0b$ defines an element of S; on the other hand, if $a < 0$, then $a + (-a)b = a(1 - b) = (-a)(b - 1) \geq 0$ and therefore defines an element of S (note that b assumed to be a positive integer, therefore $b - 1 \geqslant 0$). Since S is non-empty, it makes sense to take the smallest element, say r, of S. By definition, In this case, r is of the form $a + (-q)b \geq 0$ for some integer q.

We show that $r < b$. If $r \geqslant b$, then $0 \leqslant r - b = a - (q + 1)b < a - bq = r$ and $r - b$ would define an element of S that is strictly smaller than r. This contradicts the minimality of r. Therefore $r < b$.

The uniqueness is left as an exercise! □

Definition. Let a and b be integers. We say that a divides b if and only if there exists an integer c such that $b = ac$.

Remark. Note that a and b can be negative; in fact a can be zero! According to the definition, for the statement 'zero divides b' to hold, there has to be an integer c such that $b = 0 \times c$; but the RHS is nothing other than 0, forcing b to be zero! In other words, the only integer zero divides is zero itself! We are only considering 'zero divides zero' and are NOT trying to make sense of $\frac{b}{a} = \frac{0}{0}$.

Examples.
Every integer, including zero, divides 0. Indeed, $0 = a \times 0$.

If a and b are non-negative integers such that a divides b as well as b divides a, then $a = b$. This seems 'obvious' but proving this formally requires a bit of work: Firstly, suppose $a = 0$. It then follows, by the remark above and the assumption, a divides b, that $b = 0$. So $a = b$ holds. Swapping the roles, we can also prove, if $b = 0$, then $a = b(= 0)$. Having dealt with these two degenerate case, we may now assume that a and b are both positive integers. By assumption, there exist integers r and s such that $a = rb$ and $b = sa$– since a and b are positive integers, r and s are positive integers. We see that $a = rb = r(sa) = rsa$. Because a is non-zero, we may divide this through and get $rs = 1$. As we know r and s are positive integers, we deduce $r = 1$ and $s = 1$. It therefore follows that $a = rb = b$.

2.2 GCD and Euclid's algorithm

Definition. Let a and b be integers. A common divisor of a and b is a *non-negative* integer r with the property that r divides a and r divides b. We call a common divisor r of a and b the high common factor, or the greatest common divisor (GCD) in this course, if any other common divisor is smaller than r, i.e. if s is another common divisor of a and b, then $s < r$ holds.

Remark (important). The GCD of a and b, written often as $\gcd(a, b)$, is defined to be a non-negative integer, even if a and b are negative.

Example.
$a = 12, b = 18$: $\gcd(12, 18) = 6$.
$a = 12, b = -18$: $\gcd(12, -18) = 6$.
If a is a non-zero integer, then $\gcd(a, 0) = a$ (see below as to why).

Remark (not so important but useful to know) For $r = \gcd(a, b)$ to be 'the greatest', it is decreed, as part of the definition of GCD, that if s is a common divisor of a and b, then $s < r$ needs to hold. In fact, $s < r$ is equivalent to:

Proposition 2. s divides r.

Example. $a = 50, b = 100$. Then $r = \gcd(50, 100) = 50$. For example, $2, 5, 10, 25$ are all common divisors s of 50 and 100, and they all divide r.

Remark. If we know the Fundamental Theorem of Arithmetic– any positive integer can be written as a product of primes numbers, and this product expression is unique up to recording of the factors, then it is possible to completely unravel the common divisors, and this proposition follows immediately. A lowbrow approach requires the Bezout's identity– if p and q are integers, there exist integers x and y such that $px + qy = \gcd(p, q)$.

Proof. Firstly, if $s = 0$, then it forces both a and b to be 0, which in turn forces $r = 0$. But this contradicts the assumption that s is 'another' common divisor. So we may assume $s > 0$ (note that s is assumed to be non-negative). Since $s < r$, we see that r is also positive. It then follows that $r = sq + \gamma$ for some integers $q(> 1)$ and $0 \leq \gamma < s$. It suffices to see $\gamma = 0$.

Firstly, $a = rp$ and $b = rq$ for some integers. Since r is GCD, we see that p and q have no common divisor (i.e. $\gcd(p, q) = 1$); indeed, if it did have a common divisor, then multiplying r

by that common divisor (necessarily a positive integer) would yield a common divisor greater than r (which contradicts r being the 'greatest'). On the other hand, $a = (sq + \gamma)p$ and $b = (sq + \gamma)q$ and it follows (from the assumption that s is also a common divisor of a and b) that s divides both γp and γq. Since $\gcd(p, q) = 1$, it follows from Bezout's identity that there exist integers x, y such that $px + qy = 1$. Since $\gamma px + \gamma qy = \gamma$ and the LHS is divisible by s, the RHS, γ, is also divisible by s. However, since $0 \leq \gamma < s$, the only possibility left is that $\gamma = 0$. □

Definition. We say that a pair of integers a and b are coprime, if $\gcd(a, b) = 1$.

Example. 2 and 5 are coprime, but 2 and 4 are not comprime.

Definition. Let a and b be as above. A positive integer s is a common multiple of a and b if a divides s and b divides s. A common multiple of a and b is the least common multiple of a and b, $\mathrm{lcm}(a, b)$, if it is smaller than any other common multiple, in the sense that if s is another common multiple of a and b, then $r < s$.

Analogous to the case of GCD, we know:

Proposition 3. r divides s.

Proof. Observe, firstly, that there exist q and $0 \leq \gamma < r$ such that $s = rq + \gamma$. It suffices to show that $\gamma = 0$. By definition, a divides s, hence it also divides γ (as a divides r). Similarly, b divides s, therefore b divides γ. Combining, both a and b divide γ. This implies $\gamma = 0$ as if γ was non-zero, then it would mean that γ is a common multiple of a and b but is also smaller than r (by definition); and this contradicts the minimality of $r = \mathrm{lcm}(a, b)$. □

Example.
$a = 12, b = 18$: $\mathrm{lcm}(12, 18) = 36$.

Proposition 4. Let a be a non-negative integer. Then $\gcd(a, 0) = a$.

Proof. Let $\gamma = \gcd(a, 0)$. By definition, γ divides a; this means that there exists an integer λ such that $a = \gamma\lambda$. On the other hand, because a divides both a (itself) and 0 (as $0 = a0$), a is a common divisor of a and 0, and therefore a divides γ (Proposition 2) and we may write $\gamma = a\mu$ for some integer μ. Plugging this into $a = \gamma\lambda$, we obtain $a = a\lambda\mu$. From this equality, we deduce that if a is zero, then $\gamma = 0\mu = 0$; if, on the other hand, a is non-zero, then $\lambda\mu = 0$, i.e. $(\lambda, \mu) = (1, 1)$ or $(-1, -1)$. However, since both γ (by definition) and a (by assumption) are non-negative, $(\lambda, \mu) = (1, 1)$, i.e., $\gamma = a$. □

Euclid's algorithm is a very useful tool to compute gcd.

Example.

$\gcd(198, 78) = 6$.

$$\begin{aligned} 198 &= 78 \cdot 2 + 42 \\ 78 &= 42 \cdot 1 + 36 \\ 42 &= 36 \cdot 1 + 6 \\ 36 &= 6 \cdot 6 + 0 \end{aligned}$$

$\gcd(-78, 198) = 6$

$$\begin{aligned} -78 &= 198 \cdot (-1) + 120 \\ 198 &= 120 \cdot 1 + 78 \\ 120 &= 78 \cdot 1 + 42 \\ 78 &= 42 \cdot 1 + 36 \\ 42 &= 36 \cdot 1 + 6 \\ 36 &= 6 \cdot 6 + 0 \end{aligned}$$

In fact,

Lemma 5. If a and b are (positive) integers, then $\gcd(a, b) = \gcd(-a, b) = \gcd(a, -b) = \gcd(-a, -b)$.

Proof (non-examinable). Let us prove the first quality. If $\gcd(a, b) = 0$, then it forces $a = b = 0$. In this case, $\gcd(-a, b) = 0$. We may therefore assume that $\gcd(a, b) > 0$ (note that $\gcd(a, b)$ is defined to be non-negative). Since $\gcd(a, b)$ divides a and b, it also divides $-a$ and b. Therefore $\gcd(a, b)$ is a common divisor of $-a$ and b. By the Proposition ? above, $\gcd(a, b)$ divides $\gcd(-a, b)$. Swapping the role of $\gcd(a, b)$ and $\gcd(-a, b)$, we may also conclude that $\gcd(-a, b)$ divides $\gcd(a, b)$. Combining (together with the fact that they are both positive integer), $\gcd(a, b) = \gcd(-a, b)$. □

Euclid's algorithm is based on the following proposition– by the lemma above, we can always make 'b' positive when it comes to computing $\gcd(a, b)$.

Proposition 6. Let a and b be integers and suppose that $b > 0$ and $a = bq + r$ for some uniquely determined integers q and $0 \leqslant r < b$. Then $\gcd(a, b) = \gcd(b, r)$.

Proof. Let $\gamma = \gcd(a, b)$ and $\lambda = \gcd(b, r)$.

Firstly, suppose that $\gamma = 0$. Since γ divides a and b, the assumption forces both a and b to be zero. This in turn forces $r = 0$ and therefore $\lambda = 0$.

We may henceforth assume that γ is non-zero; since GCD is assumed to be non-negative integer, it is forced that $\gamma > 0$. It then follows that $\lambda > 0$. Indeed, if $\lambda = 0$, an argument similar to the one above would force $\gamma = 0$ which contradicts the running assumption $\gamma > 0$.

We claim that λ divides γ. Since γ divides a, and b, by definition, it follows that γ divides $a - bq = r$. Combined the fact that γ divides b by definition, γ makes a common divisor of b and r. By Proposition 2, we may therefore conclude that λ divides γ.

We also claim that γ divides λ. Since λ divides b, and r, by definition, it follows that λ divides $bq + r = a$. Combined with the assumption that λ divides b, λ defines a common divisor of a and b. By Proposition 2 (again!), we conclude that γ divides λ.

Since γ divides λ, as well as λ divides γ, it follows that $\gamma = \lambda$. □

2.3 Euclid's algorithm extended

Theorem 7 (Bezout's identity). Let a and b be integers. Then there exist integers r and s such that $ar + bs = \gcd(a, b)$. These integers r and s can be found from Euclid's algorithm.

Proof (non-examinable). Let S be the set of integers of the form $a\lambda + b\mu$, where λ and μ range over $\mathbb{Z}$. Since both a and b lie in S, the set is non-empty and let γ denote the smallest *non-negative* integer in S. We show that $\gamma = \gcd(a, b)$.

Since $\gcd(a, b)$ divides both a and b, it divides any integer linear combination of a and b (i.e. any element of S). In particular, $\gcd(a, b)$ divides γ and consequently $\gcd(a, b) \leqslant \gamma$ (since both $\gcd(a, b)$ and γ are non-negative integers).

Let $a = \gamma q + r$ for some integer $0 \leqslant r < \gamma$. Since γ is an element of S, so is r. If r is non-zero, it contradicts the minimality of γ. Therefore $r = 0$, in other words, γ divides a. We may similarly prove that γ divides b. Combining, γ divides both a and b, i.e. γ is a common divisor of a and b. Therefore $\gamma \leqslant \gcd(a, b)$.

As we have shown that $\gcd(a, b) \leqslant \gamma$, as well as $\gamma \leqslant \gcd(a, b)$, it follows that $\gamma = \gcd(a, b)$. □

Examples. Looking at the steps computing $\gcd(198, 78) = 6$ in the reverse order,

$$\begin{aligned} 6 &= 42 - 1 \cdot 36 \\ &= 42 - 1 \cdot (78 - 1 \cdot 42) \\ &= 2 \cdot 42 - 1 \cdot 78 \\ &= 2 \cdot (198 - 2 \cdot 78) - 1 \cdot 78 \\ &= 2 \cdot 198 - 5 \cdot 78 \end{aligned}$$

so $\gcd(198, 78) = 6 = 2 \cdot 198 + (-5) \cdot 78$.

Looking at the steps computing $\gcd(-78, 198) = 6$ in the reverse order,

$$\begin{aligned} 6 &= 42 - 1 \cdot 36 \\ &= 42 - 1 \cdot (78 - 1 \cdot 42) \\ &= 2 \cdot 42 - 1 \cdot 78 \\ &= 2 \cdot (120 - 1 \cdot 78) - 1 \cdot 78 \\ &= 2 \cdot 120 - 3 \cdot 78 \\ &= 2 \cdot 120 - 3 \cdot (198 - 1 \cdot 120) \\ &= 5 \cdot 120 - 3 \cdot 198 \\ &= 5 \cdot (1 \cdot (-78) + 1 \cdot 198) - 3 \cdot 198 \\ &= 2 \cdot 198 - 5 \cdot (-78) \end{aligned}$$

so $\gcd(198, -78) = 6 = 2 \cdot 198 + 5 \cdot (-78)$.

Subtexts (evidently non-examinable). In this section, we set off by leaning that '$\mathbb{Z}$ is a Euclid domain'. The set S in the proof of Bezout is, by definition, an 'ideal' of the ring (actually a domain) $R = \mathbb{Z}$; and 'γ divides $\gcd(a, b)$' sees the standard technique of proving that $\mathbb{Z}$ is a principal ideal domain. It is because of the comparative 'easier' direction– '$\gcd(a, b)$ divides γ' – that $\mathbb{Z}$ is a Bezout domain. The FTA below sees $\mathbb{Z}$ is a UFD; note however that Bezout is not necessarily a UFD.

2.4 Prime numbers

A prime number is (defined to be) a positive integer which can only be divided by 1 or itself (i.e. no *proper* factors). From the viewpoint of theory of algebra, the right definition of a prime number is that it is a positive integer p which satisfies the property that if p divides the product ab of integers a and b (not necessarily positive), then p divides either a or b. We will go with the standard definition though.

Remark (non-examinable, though it really touches upon the essence of this course). In $\mathbb{Z}$, the only elements that divide 1 are 1 and -1; and they are called *units* of $\mathbb{Z}$ (we will learn more about this towards the end of the course).

We say that an element π of $\mathbb{Z}$ (i.e. π is an integer) is *irreducible* if the following condition holds: if $\pi = ab$ (for a, b in $\mathbb{Z}$), then either a, or b, is a unit. For example, a prime number p in the standard sense (e.g. $2, 3, 5, \dots$), is irreducible in $\mathbb{Z}$. In fact, the negative integer $-p$ is also irreducible, by definition.

On the other hand, we say that an element π is *prime* if the following condition holds: if π divides ab, then either π divides a, or π divides b; this is the definition that pins down what it means for an integer to be prime.

If π is prime, then it is irreducible. Let π be a prime element of $\mathbb{Z}$, and suppose that $\pi = ab$. Since π divides π, it follows that π divides ab. Since π is assumed to be prime, π divides either a, or b. Without loss of generality, suppose that π divides a. Let $a = \pi c$ for some integer c. Feeding this back into $\pi = ab$, we deduce $\pi = \pi bc$, i.e. $\pi(bc - 1) = 0$. Since π is non-zero, and there is no 'zero-divisor', $bc - 1$, i.e. $bc = 1$. Hence b is a unit. Similarly, if π divides b, one can deduce a is a unit. In both cases, a or b is a unit, hence π is irreducible.

To prove the converse, we use Bezout's identity. Let π be an irreducible element of $\mathbb{Z}$. Suppose that π divides ab for a pair of integers a and b. Suppose furthermore that π does not divide a. The assertion– π is prime– follows if we establish that π divides b.

We claim that $\gcd(a, \pi) = 1$. To see this, let $\gamma = \gcd(a, \pi)$. Since γ divides π, it follows from the irreducibility of π (and the assumption that π does not divide a) that γ is a unit. Since γ is by definition positive, $\gamma = 1$. It then follows from Bezout that there exist integers r and s such that $ar + \pi s = 1$. Multiplying the equality through by b, we obtain $b = b(ar + \pi s)$. The both terms on the RHS are divisible by π (π divides ab by assumption!). It therefore follows that π divides b, as desired.

Theorem 8. (The Fundamental Theorem of Arithmetic) Every integer can be expressed as the product $(-1)^r \prod_p p^{r_p}$ of prime numbers (in the standard sense), where r is an element of $\{\pm 1\}$ and, for every prime number p, r_p is a non-negative integer. Furthermore, the expression is unique up to re-ordering of the prime factors.

Proof. See Example Sheet. □

Remarks.

The FTA is the reason why you often here 'the prime numbers are the building blocks of the (whole) numbers'.

You might find discussions of T. Gowers (a Fields medalist):

https://gowers.wordpress.com/2011/11/13/why-isnt-the-fundamental-theorem-of-arithmetic-obvious/

https://gowers.wordpress.com/2011/11/18/proving-the-fundamental-theorem-of-arithmetic/

enlightening.

What else do we know about prime numbers?

Proposition. There are infinitely many prime numbers.

Proposition. There are infinitely many prime numbers congruent to -1 mod 4.

Proof. Example Sheet. □

3 Modular arithmetic

3.1 Equivalence relations and partitions

Suppose that S is a set. In NSF, a relation $\mathcal{R}$ on S is defined to be a property which may, or may not, hold for each ordered pair of elements in S (i.e. an element of the set $S \times S$ of ordered pairs in S).

A relation $\mathcal{R}$ is said to be

- reflexive if $a\mathcal{R}a$ for every element a of S,
- symmetric if $a\mathcal{R}b$ implies $b\mathcal{R}a$ for all elements a, b of S,
- anti-symmetric if $a\mathcal{R}b$ and $b\mathcal{R}a$ implies $a = b$ for all elements a, b of S,
- transitive if $a\mathcal{R}b$ and $b\mathcal{R}c$ implies $a\mathcal{R}c$ for all elements a, b, c of S,

A reflexive, symmetric and transitive relation is said to be an equivalence relation.

Examples/Exercises. Which of the following are equivalence relations?

(1) $S = \mathbb{R}$ and $a\mathcal{R}b$ if and only if $a = b$ or $a = -b$. (2) $S = \mathbb{Z}$ and $a\mathcal{R}b$ if and only if $ab = 0$. (3) $S = \mathbb{R}$ and $a\mathcal{R}b$ if and only if $a^2 + a = b^2 + b$. (4) $S = \{\text{people in the world}\}$ and $a\mathcal{R}b$ if and only if a lives within 100km of b. (5) $S = \{\text{the points in the plane}\}$ and $a\mathcal{R}b$ if and only if a and b are of the same distance from the origin. (6) $S = \{\text{positive integers}\}$ and $a\mathcal{R}b$ if and only if ab is a square (of positive integers). (7) $S = \{1, 2, 3\}$ and $a\mathcal{R}b$ if and only if $a = 1$ or $b = 1$. (8) $S = \mathbb{R} \times \mathbb{R}$ and $p\mathcal{R}q$ (where $p = (x(p), y(q))$ and $q = (x(q), y(q))$) if and only if $x(p)^2 + y(p)^2 = x(q)^2 + y(q)^2$.

(1), (3), (5), (6) and (8) are equivalence relations.

Remark. The hardest to verify is the transitivity of $\mathcal{R}$ in (6): if a, b and c are positive integers and ab and bc are respectively squares of positive integers, then can ac be a square of positive integers? Yes! To see this, suppose that $ab = r^2$ and $bc = s^2$ for some positive integers r and s. Multiplying them together, we obtain $ab^2c = (rs)^2$. It suffices to establish that b divides rs, as if this is the case, then ac is a square of $(rs)/b$. How do we prove this? Recall from Proposition 8 that b is a product of prime factors of the form $\prod_p p^{r_p}$ where p ranges over the prime numbers and r_p is a non-negative integer for every p. If p^{r_p} and q^{r_q} are prime factors of b at distinct primes p and q, and if each of them divides rs, then the product $p^{r_p}q^{r_q}$ divides rs (this follows from the 'correct' definition of prime numbers). If we repeat the argument, then we may conclude that $\prod_p p^{r_p}$, i.e., b divides rs. To sum up, it boils down to showing that, for every prime number p that divides b (i.e. $r_p \geqslant 1$), the prime factor p^{r_p} of b divides rs. Since p^{r_p} divides b, it follows that p^{2r_p} divides b^2 and therefore that p^{2r_p} divides $(rs)^2$. If p^{s_p} is the prime factor of rs at p, then p^{2r_p} divides p^{2s_p}, i.e. $2r_p \leqslant 2s_p$, i.e. $r_p \leqslant s_p$. This manifests that p^{r_p} divides rs.

If $\mathcal{R}$ is a relation on S and a is an element of $\mathcal{R}$, we denote by $[a]_{\mathcal{R}}$, or simply $[a]$ if it is clear which relation we are considering from the context, the set

$$\{b \in S \mid a\mathcal{R}b\}$$

of all elements b in S which are 'in relation to' b with respect to $\mathcal{R}$. If $\mathcal{R}$ is an equivalence relation, we refer to $[a]$ an equivalence class (represented by a).

Examples/Exercises For those relations (1)-(8) above, describe the equivalence classes.

Remark. By definition, if $\mathcal{R}$ is an equivalence relation, then $a\mathcal{R}b$ if and only if $[a]_{\mathcal{R}} = [b]_{\mathcal{R}}$. To see 'only if', let c be an element of $[a]_{\mathcal{R}}$. By definition, this means that $a\mathcal{R}c$. Since $\mathcal{R}$ is reflexive, $c\mathcal{R}a$ holds. Since $a\mathcal{R}b$ by assumption, it follows from the transitivity of $\mathcal{R}$ that $c\mathcal{R}b$. By the reflexivity (again!), it then follows that $b\mathcal{R}c$, i.e. c is a element of $[b]_{\mathcal{R}}$. To sum up, we have established that $[a]_{\mathcal{R}} \subseteq [b]_{\mathcal{R}}$. Swapping the roles, it is also possible to prove $[b]_{\mathcal{R}} \subseteq [a]_{\mathcal{R}}$ (exercise!). Combining, we have $[a]_{\mathcal{R}} = [b]_{\mathcal{R}}$ as desired.

In preparation of a theorem to follow, we need:

Definition. Let S be a set. A partition of S is a set $\mathcal{P}$ of subsets of S, whose elements are called its parts, having the following properties:

- $\varnothing$ is not a part of $\mathcal{P}$.
- If A and B are distinct parts of $\mathcal{P}$, then $A \cap B = \varnothing$,
- The union of all parts of $\mathcal{P}$ is S.

Examples.

$S = \mathbb{Z}$, $\mathcal{P} = \{\{\text{even integers}\}, \{\text{odd integers}\}\}$.

$S = \{1, 2, 3, 4, 5\}$. $\{\{1, 2\}, \{3, 4\}, \{5\}\}$ and $\{\{1\}, \{2, 3, 4, 5\}\}$ are partitions but $\{\{1, 2\}, \{2, 3\}, \{4, 5\}\}$ is not.

Theorem 9 (Equivalence Relation Theorem).

- Let $\mathcal{R}$ be an equivalence relation on a set S. Then the set $[a]_{\mathcal{R}}$, as a ranges over S, form a partition of S.
- Conversely, given any partition $\mathcal{P}$ of S, there is a unique equivalence relation $\mathcal{R}$ on S such that the parts of $\mathcal{P}$ are the same as the sets $[a]_{\mathcal{R}}$ for a in S. This $\mathcal{R}$ is defined as: $a\mathcal{R}b$ if a and b lies in the same part defined by $\mathcal{P}$.

Proof. (a) We need to check the definitions one by one.

- No element of $\{[a]_{\mathcal{R}}\}$ is $\varnothing$. To see this, observe that, since $a\mathcal{R}a$ (since $\mathcal{R}$ is reflexive), a lies in $[a]_{\mathcal{R}}$; therefore $[a]_{\mathcal{R}}$ is non-empty.
- If $[a]_{\mathcal{R}}$ and $[b]_{\mathcal{R}}$ are distinct, then $[a]_{\mathcal{R}} \cap [b]_{\mathcal{R}} = \varnothing$; or equivalently, if $[a]_{\mathcal{R}} \cap [b]_{\mathcal{R}} \neq \varnothing$, then $[a]_{\mathcal{R}} = [b]_{\mathcal{R}}$. To prove the latter, let c be an non-trivial element of $[a]_{\mathcal{R}} \cap [b]_{\mathcal{R}}$ (made possible by assumption). By definition, this means that $a\mathcal{R}c$ and $b\mathcal{R}c$, or equivalently $c\mathcal{R}b$ (because $\mathcal{R}$ is symmetric). Because $\mathcal{R}$ is transitive, it follows from $a\mathcal{R}c$ and $c\mathcal{R}b$ that $a\mathcal{R}b$. From the remark above, it follows that $[a]_{\mathcal{R}} = [b]_{\mathcal{R}}$.
- The union T of $[a]_{\mathcal{R}}$, as a ranges over S, equals S. Since $[a]_{\mathcal{R}} \subseteq S$ as sets, $T \subseteq S$. Therefore it suffices to prove $S \subseteq T$. Let a be an element of S. Then a lies in $[a]_{\mathcal{R}}$ (see the proof for the first part). Since $[a]_{\mathcal{R}} \subseteq S$, it follows that a lies in S.

(b) We check the conditions of an equivalence relation one by one, following the definition of $\mathcal{R}$ given in the statement.

- reflexive. Since a and a (!) both lie in the same part, $a\mathcal{R}a$ holds.
- symmetric. If a and b lies in the same part, then so do b and a. So the reflexivity follows.
- transitive. Suppose that a and b lies in a part A of $\mathcal{P}$, i.e. a subset A of S. Similarly, suppose that b and c lie in a part B of $\mathcal{P}$. Since b lies in both A and B, it follows from the second condition of the definition of a partition that A and B are *not* distinct, i.e. $A = B$. Therefore a and c both lie in the same part $A = B$, i.e. $a\mathcal{R}c$.

By definition, $[a]_{\mathcal{R}}$ is the set of elements b in S which lie in the same part, say A, as a does. This set is nothing other than A! Hence $[a]_{\mathcal{R}} = A$. So the partition $\mathcal{P}$ of S is the subsets of the form $[a]_{\mathcal{R}}$.

To see the uniqueness ($\mathcal{R}$ is the only equivalence relation whose parts are the subsets $[a]_{\mathcal{R}}$), suppose that $\mathcal{R}$ and $\mathcal{R}'$ are equivalence relations giving rise to the partition $\mathcal{P}$. Since the parts $\{b \mid a\mathcal{R}b\} = [a]_{\mathcal{R}}$ and $\{b \mid a\mathcal{R}'b\} = [a]_{\mathcal{R}'}$ both contain a, they are the same subsets of S. □

Remark. The theorem asserts that every element a of S belongs to exactly one equivalence class $[a]$.

Example. Let $S = \{1, 2, 3\}$.

Partition	Relations	Equivalence classes
$\{1,2,3\}$	$a\mathcal{R}b$ for all $a,b \in \{1,2,3\}$	$[1]$
$\{1\},\{2,3\}$	$1\mathcal{R}1$, $a\mathcal{R}b$ for all $a,b \in \{2,3\}$	$[1],[2]$
$\{2\},\{1,3\}$	$2\mathcal{R}2$, $a\mathcal{R}b$ for all $a,b \in \{1,3\}$	$[2],[1]$
$\{3\},\{1,2\}$	$3\mathcal{R}3$, $a\mathcal{R}b$ for all $a,b \in \{1,2\}$	$[3],[1]$
$\{1\},\{2\},\{3\}$	$1\mathcal{R}1$, $2\mathcal{R}2$, $3\mathcal{R}3$	$[1],[2],[3]$

3.2 Congruence mod n

Let n be a positive integer.

Definition. We define a relation $\equiv$ on the set $\mathbb{Z}$ as follows:

if a and b are elements of $\mathbb{Z}$ (i.e. integers), then $a \equiv b$ if and only if $b - a$ is divisible by n.

Proposition 10. $\equiv$ on $\mathbb{Z}$ is an equivalence relation.

Proof. We need to check that it is reflexive, symmetric and transitive.

- $a \equiv a$.

  Since $a - a = 0$ and this is divisible by n (or any integer, for that matter), $a \equiv a$.

- If $a \equiv b$, then $b \equiv a$.

  Since $a \equiv b$, there exists $b - a$ is divisible by n, i.e., there exists an integer r such that $b - a = rn$. It then follows that $a - b = (-r)n$, i.e. $a - b$ is divisible by n, hence $b \equiv a$.

- If $a \equiv b$ and $b \equiv c$, then $a \equiv c$.

  By assumption, there exist integers r and s such that $b - a = rn$ and $c - b = sn$. It then follows that $c - a = (c - b) + (b - a) = rn + sn = (r + s)n$, hence $a \equiv c$. $\square$

This means that the set of integers is partitioned into equivalence classes by $\equiv$.

Definition. We write $\mathbb{Z}_n$ for the set of equivalence classes modulo n. Personally, I prefer to write $\mathbb{Z}/n\mathbb{Z}$. When n is a prime number p, we write $\mathbb{F}_p$ instead of '$\mathbb{Z}_p$'.

Examples

$$\mathbb{Z}_5 = \left\{ \begin{array}{ccccc} \vdots & \vdots & \vdots & \vdots & \vdots \\ [-5] & [-4] & [-3] & [-2] & [-1] \\ || & || & || & || & || \\ [0] & [1] & [2] & [3] & [4] \\ || & || & || & || & || \\ [5] & [6] & [7] & [8] & [9] \\ || & || & || & || & || \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{array} \right\}$$

In a standard clock, keeping track of hours $= \mathbb{Z}_{12}$ while minutes $= \mathbb{Z}_{60}$.

Proposition 11. The cardinality of $\mathbb{Z}_n$ is n, i.e. there are exactly n equivalence classes with respect to $\equiv$ modulo n, namely $[0], [1], \ldots, [n-1]$.

Proof. Firstly, we show that every integer s belongs to one of the congruence classes $[0], \ldots, [n-1]$. Indeed, there exist integers q and $0 \leqslant r \leqslant n-1$ such that $s = nq + r$, i.e. $s \equiv r \bmod n$. Therefore s lies in $[r]$.

Suppose r and s are integers satisfying $0 \leqslant r < s \leqslant n-1$. If $[r] = [s]$, then it would follow that $r - s$ is divisible by n. But this contradicts $0 < r - s < n - 1$. □

3.3 Arithmetic with congruence classes

We define addition, subtraction and multiplication on $\mathbb{Z}_n$ as follows:

$$[a] + [b] = [a + b]$$
$$[a] - [b] = [a - b]$$
$$[a][b] = [ab]$$

What about 'division'? Can we make sense of it? It is NOT true that when we divide $[a]$ by $[b]$, we get $\left[\frac{a}{b}\right]$. In the first place, $\frac{a}{b}$ might not even be an integer! Would it be surprising if I tell you, for example, that when $n = 11$, we can even divide $[1]$ by $[3]$ to get $[4]$! This is because $[3][4] = [12] = [1]$.

Examples

$\mathbb{Z}_3 = \mathbb{F}_3 = \{[0], [1], [2]\}$. Then $[1] + [2] = [1 + 2] = [3] = [0]$ while $[2][2] = [2 \times 2] = [4] = [1]$.

$\mathbb{Z}_6 = \{[0], [1], [2], [3], [4], [5]\}$. Then $[2] + [5] = [2 + 5] = [7] = [1]$ while $[2][3] = [2 \times 3] = [6] = [0]$. Since 2 divides 6, we know very well that $\frac{6}{2} = 3$ but $\frac{[6]}{[2]} = [3]$? In the first place, $[6] = [0]$, so this should mean the same thing as $\frac{[0]}{[2]} = [3]$ but if we allowed $\frac{[0]}{[2]} = [\frac{0}{2}] = [0]$, then we would get $[0] = [3]$ which is evidently false!

It is necessary to check that these definitions do not depend on our choice of representatives. For example, we've seen $[1] + [2] = [0]$ in $\mathbb{Z}_3$ but we could have had $[4]$ instead of $[1]$, as $[1] = [4]$. In this case, $[4] + [2] = [6] = [0]$, so it does not matter whether we choose 1 or 4 (or any integer congruent to 1 mod 3 for that matter) as a representative of the equivalence class $[1]$.

More rigorously, suppose that a, b and c are integers and that $a \equiv b \bmod n$. To show that the definition of 'addition' does not depend on choice of representatives, we need to show $[a] + [c] = [b] + [c]$. Since the LHS (resp. RHS) is defined to be $[a + c]$ (resp. $[b + c]$), this is equivalent to showing that $[a + c] = [b + c]$. However, it follows from $a \equiv b \bmod n$ that $(a + c) - (b + c)$ is divisible by n and therefore that $(a + c) \equiv (b + c) \bmod n$. It follows that $[a + c] = [b + c]$.

Similarly, it is necessary to check that $[a][c] = [b][c]$, i.e. $[ac] = [bc]$. Since n divides $a - b$, it also divides $c(a - b) = ac - bc$. Therefore $ac \equiv ab$, i.e. $[ac] = [ab]$.

3.4 Modular inverses

Let n be a fixed positive integer. Throughout this section, $\equiv$ denotes the 'congruence modulo n' and $[a]$ denote the congruence class of integers congruent to a modulo n.

Theorem 12. The elements $[a]$ of $\mathbb{Z}_n$ has a multiplicative inverse if and only if $\gcd(a, n) = 1$.

Proof. Suppose that $[a]$ has a multiplicative inverse, i.e. $[b]$ such that $[a][b] = [1]$, i.e. $[ab] = [1]$. This means that $ab - 1$ is divisible by n, hence there exists an integer c such that $ab + (-c)n = 1$. As $\gcd(a, n)$ divides the LHS, it does so the RHS, i.e. 1. The only non-negative integer diving 1 is 1, so $\gcd(a, n) = 1$.

Conversely, suppose $\gcd(a, n) = 1$. By Bezout, there exist integers b and c such that $ab+nc = 1$. Since $ar \equiv 1 \bmod n$, it follows that $[a][b] = [ab] = [1]$. The multiplicative inverse of $[a]$ is therefore $[b]$. □

Examples.

What is the multiplicative inverse of $[4]_{21}$? Since $\gcd(4, 21) = 1$, the theorem assures us of the multiplicative inverse. How do we compute it? The proof indeed explains how. Since $\gcd(4, 21) = 1$, Euclid's algorithm (backed up by Bezout) gives us a pair of integers r and s such that $4r + 21s = \gcd(4, 21) = 1$. Indeed, $(r, s) = (-5, 1)$ does the job. In particular, $4r \equiv 1 \bmod 21$ and it therefore follows that $[4][r] = [4r] = [1]$. So $[-5] = [16]$ is the multiplicative inverse of $[4]$.

What is the multiplicative inverse of $[23]_{2023}$? Firstly, we compute $\gcd(23, 2023)$ by Euclid's algorithm:

$$\begin{aligned} 2023 &= 23 \cdot 87 + 22 \\ 23 &= 22 \cdot 1 + 1. \end{aligned}$$

Hence $1 = 23 - 1 \cdot 22 = 23 - 1 \cdot (2023 - 23 \cdot 87) = 1 \cdot 2023 + (-86) \cdot 23$ and $[-86] = [1937]$ is the multiplicative inverse of $[23]$.

What is the multiplicative inverse of $[17]_{2023}$? Since $2023 = 119 \cdot 17$ and 17 is a prime number, $\gcd(2023, 17) = 17$. It follows from the theorem above that $[17]$ has no multiplicative inverse.

If p is a prime number, then $\mathbb{Z}_p = \{[0], [1], \ldots, [p-1]\}$ and, by the theorem, it follows that $\gcd(a, p-1) = 1$ if and only if a is prime to p. Therefore the congruence classes $[1], \ldots, [p-1]$ all have inverses.

Proposition 13. Suppose $n > 1$. The element $[a]$ of $\mathbb{Z}_n$ has no multiplicative inverse if and only if there exists an integer b, not congruent to 0 modulo n, such that $[a][b] = [0]$.

Proof. Suppose that $[a]$ has no multiplicative inverse. It then follow from the theorem above that $c = \gcd(a, n) > 1$. If we let $b = n/c$, then b is a positive integer not congruent to 0 mod n (if it were congruent to 0 mod n, then b would be n and force $c = 1$). By definition, $ab = an/c = (a/c)n$ is divisible by n, for a/c is an integer. It follows that $ab \equiv 0 \bmod n$, hence that $[a][b] = [ab] = [0]$.

To prove the converse, suppose that $[a]$ has a multiplicative inverse– we aim at establishing that no integer b, not congruent to n, satisfies $[a][b] = [0]$. By assumption, there exists an integer c such that $[a][c] = [1]$. Let b be an integer not congruent to 0 mod n. Multiplying the both sides of $[a][c] = [1]$ by $[b]$, we obtain $[b] = [b][a][c] = [c]([a][b])$. If $[a][b] = [0]$, then the RHS is $[0]$, hence the LHS $[b]$ is $[0]$, in other words, b is divisible by n. However this contradicts the assumption that b is not. □

Remark. Proposition 13 is paraphrasing $\gcd(a, n) > 1$.

Given a positive integer n, how many elements in $\mathbb{Z}_n$ has multiplicative inverses? In theory, we ask, for every $0 \leq a \leq n-1$, whethere $\gcd(a, n) = 1$ (or not) to compile a list. For example, if $n = 24$, $\{1, 5, 7, 11, 13, 17, 19, 23\}$ (incidentally they are all prime numbers!) is the set of integers $0 \leq a \leq n - 1 = 23$ such that $\gcd(a, 24) = 1$. Hence there are 8 elements in total.

What about $n = 108$? That seems to entail a lot of computations. There is a formula!– it goes by the name of Euler's totient function. Recall from the fundamental theorem of arithmetic that n may be written as the product $\prod_p p^{r_p}$ of prime factors. Then the number we are looking for is computed by

$$\phi(n) = \prod_p (p-1)p^{r_p - 1}.$$

For example, $24 = 2^3 \cdot 3$, so $\phi(24) = (2-1)2^2 \cdot (3-1) = 8$ which is consistent with the computation above. Similarly, $108 = 3^3 \cdot 2^2$, so $\phi(108) = (3-1) \cdot 3^2 \cdot (2-1) \cdot 2 = 36$. Is this consistent with your computation?

4 Algebraic structures

Which is more symmetrical, a scalene triangle or an equilateral triangle? The equilateral triangle has lots of symmetries (reflection and rotation), while the scalene triangle has no symmetry at all. We are capable of sensing, or even quantifying, 'symmetries' (we often find symmetry rather pleasing, evidenced in art, architecture etc.). But what exactly is symmetry?

By a 'symmetry' of an object, we mean an 'action' we perform on the object (e.g. rotating 'r', reflecting 's', etc.) while preserving its structure (e.g. vertices, edges, etc.). For the equilateral triangle, what are all the symmetries?

Let us call the vertices of an equilateral triangle a, b and c clockwise, with a sitting at '12', b at '4' and c at '8' if we fit the triangle in an old-fashioned clock. For ease of reference, let us write this configuration $[a, b, c]$. Rotating $\pi/3$ clockwise changes $[a, b, c]$ into $[c, a, b]$, while reflecting the triangle with respect to the line passing through a and the mid-point of the edge bc turns $[a, b, c]$ into $[a, c, b]$. We call the former 'action' r and the latter s. Playing around this a bit, we should see that $\{e, r, s, r^2, sr, sr^2\}$ (e denotes 'doing nothing'), subject to conditions such as $r^3 = e, s^2 = e$ and $srs = r^{-1}$, i.e. the composition of actions, s followed by r and then by s, equals $-\pi/3$ rotation represented by r^{-1}, are all possible actions, as there are in total $3!$ possible configurations$[*, *, *]$ in $\{a, b, c\}$. To sum up, we have completely described the symmetries of an equilateral triangle in terms of rotations and reflections. In fact, this 'labelling of actions' has given a 'structure' to the symmetries– for example, rs is manifestly different from sr!

A group is an axiomatisation/formalisation of 'actions' (we often forget about the 'object') as we have seen in this example.

Why groups? Groups pin down what we intuitively sense as a 'structure' (antithesis of which is 'chaos') and for that reason they are everywhere! Would it be surprising that group theory predicted the existence of many elementary particles before they were found experimentally? Groups theory is also a powerful tool in public-key cryptography and 'conceptually solving' Rubik's Cube?

Representation theory is a subject that aims at describing symmetry in terms of matrices (as we see in linear algebra). This is a subject area very much related to physics, for example.

4.1 Groups

Definition. A group is a set G with an operation $*$ on G satisfying the following axioms:

(G0) If a, b are elements of G, then $a * b$ is an element of G.

(G1) If a, b, c are elements of G, then $a * (b * c) = (a * b) * c$.

(G2) There is an element e in G (called the identity element) such that $a * e = e * a = a$ for every element of G.

(G3) For every element a of G, there exists b in G such that $a * b = b * a = e$. The element b is called the inverse of a.

(G4) If a, b are elements of G, then $a * b = b * a$.

When these five conditions hold, we say $(G, *)$ (or simply G if the operation $*$ is clear from the context) is a commutative/abelian group. By groups, I shall mean abelian groups unless otherwise specified.

Examples.

• $(G, *) = (\mathbb{Q}, +)$ is a group– this is an additive group (where the identity element e is '0' as we know well).

• $(\mathbb{Q} - \{0\}, \times)$ is a group– this is a multiplicative group (where the identity element is '1' as we know well).

• $(\mathbb{Z}, +)$ is an additive group.

• $(\mathbb{Z}, \times)$ is not a group. It seems 1 is a perfect candidate for the identity element (as it does the job in a bigger set $\mathbb{Q}$) but, for example, 2 does not have (multiplicative) inverse, i.e. there is no integer b such that $2 \times b = 1$!

• $(\{\text{The roots of } X^n - 1 \text{ in } \mathbb{C}\}, \times)$ is a multiplicative group (with identity element 1).
• $(\{\text{The roots of } X^n - 1 \text{ in } \mathbb{C}\}, +)$ is not a group.
• $(\{\text{The 2-by-2 invertible matrices with entries in } \mathbb{C}\}, \times)$ is a group but not an abelian group!

• $(\mathbb{Z}_n, +)$ is a group.
• $(\mathbb{Z}_n, \times)$ is not a group. The element $[1]$ satisfies (G2) but Theorem 12 proves that only those $[a]$ with $\gcd(a, n) = 1$ has (multiplicative) inverses, failing (G3).
• $(\mathbb{R}, *)$, where $*$ is defined as $a * b = a^2 b$, is not a group. (G2) fails for this example, i.e., there is no identity element. To see this, suppose that e is an element of $\mathbb{R}$ satisfying (G2). Firstly, $e * e = e$ yields $e^3 = e$. The only real numbers which satisfy this are $0, 1$ or -1. Secondly, for every element a in $\mathbb{R}$, the equality $a * e = e * a = a$ yields $a^2 e = e^2 a = a$. If $e = 0$, then $a = 0$ and this is false (as it says any element a of $\mathbb{R}$ is 0). If $e = 1$, then $a^2 = a$ forcing a to be either 0 or 1. If $e = -1$, then $-a^2 = a$ forcing a to be either 0 or -1.
• $(\mathbb{Z}_{\geqslant 1}, *)$, where $*$ is defined as $a * b = |a - b|$, is not a group. (G1) fails for this example. Indeed, $1 * (2 * 5) = 2$ but $(1 * 2) * 5 = 4$.

• Let S be a non-empty set. Let $\mathrm{Sym}(S)$ be the set of *bijective* functions $a : S \to S$ and $*$ be the composition $\circ$– if a and b are elements of G, then $a \circ b$ is the composite $S \xrightarrow{b} S \xrightarrow{a} S$ sending s to $a(b(s))$. Then $(\mathrm{Sym}(S), \circ)$ is a group.

(G0) If a and b are bijective, so is $a \circ b$. To see $a \circ b$ is injective, suppose $(a \circ b)(s) = (a \circ b)(s')$ for some elements s, s' of S (and aim at proving $s = s'$). By definition, $a(b(s)) = a(b(s'))$. Since a is injective, $b(s) = b(s')$. Since b is injective, $s = s'$ as desired. To prove $a \circ b$ is surjective, let s be an element of S (and aim at finding s'' such that $(a \circ b)(s'') = s$). Since a is surjective, there exists an element s' in S such that $a(s') = s$. Since b is surjective, there exists s'' in S such that $b(s'') = s'$. It then follows that $(a \circ b)(s'') = a(b(s'')) = a(s') = s$ by definition.

(G1) Let a, b, c be elements of $\mathrm{Sym}(S)$ (and aim at proving $a \circ (b \circ c) = (a \circ b) \circ c$). Indeed, $[a \circ (b \circ c)](s) = a((b \circ c)(s)) = a(b(c(s))) = (a \circ b)(c(s)) = [(a \circ b) \circ c](s)$.

(G2) The identity element is the identity map 'id' sending s to s. Then $(a \circ \mathrm{id})(s) = a(\mathrm{id}(s)) = a(s)$ and $(\mathrm{id} \circ a)(s) = \mathrm{id}(a(s)) = a(s)$.

(G3) If a is an element of $\mathrm{Sym}(S)$, then it follows from the surjectivity of a that, for any element s' of S, there exists s in S such that $a(s) = s'$. Note that this s is unique; indeed if r and s are elements of S satisfying $a(r) = s'$ and $a(s) = s'$, then $a(r) = a(s)$ holds. By injectivity of a, we have $r = s$. Granted, we define b to be the map that sends s' in S to the element s of S uniquely defined such that $a(s) = s'$– as is clear from the definition, this is well-defined only because a is bijective to start with. It remains for us to check that b fulfils the role of being the inverse of a. Since $(a \circ b)(s') = a(b(s')) = a(s) = s'$, we see that $a \circ b = \mathrm{id}$. Similarly, since $(b \circ a)(s) = b(a(s)) = b(s') = s$, we see that $b \circ a = \mathrm{id}$. This map b is often written as a^{-1}.

This last example formalises what we previously discussed as 'symmetries of an equilateral triangle' (where S is taken to be the vertices of the triangle and 'rotations' and 'reflections' are bijections). In fact, it underlies the historical development of the group theory:

Non-examinable Examples.
Symmetry groups of regular polygons (e.g. an equilateral triangle).
Symmetry groups of platonic solids.

Galois groups. E. Galois found a way to describe the solutions of a polynomial over $\mathbb{Q}$ in terms of a group. In fact, this was the motivation behind the development of group theory!

I find R. Borcherd's video https://www.youtube.com/watch?v=D908X1JAowY enlightening.

4.2 Elementary properties of groups

Proposition 14. Let $(G, *)$ be a group.

- The identity element of G is unique.
- Each element a of G has a unique inverse (written multiplicatively as a^{-1}).
- If $a * b = a * c$, then $b = c$. Similarly, if $b * a = c * a$, then $b = c$.
- For any a, b in G, then $(a * b)^{-1} = b^{-1} * a^{-1}$.

Proof. (1) If e and e' have properties that $e*a = a*e = a$ and $e'*a = a*e' = a$ for every element a of G, then $e * e' = e'$ (by letting $a = e'$ int the former) and $e * e' = e$ (by letting $a = e'$ in the latter). Combining, $e = e'$. (2) Let b and b' are elements of G has properties that $a * b = b * a = e$ and $a * b' = b' * a = e$. One observes that $b = b * e = b * (a * b') = (b * a) * b' = e * b' = b'$. (3) Let a^{-1} be the inverse of a. It then follows that $a^{-1} * (a * b) = a^{-1} * (a * c)$. The LHS equals $(a^{-1}*a)*b = e*b = b$ and similarly the RHS equals $(a^{-1}*a)*c = e*c = c$. Hence $b = c$, as desired. The second assertion can be proved analogously. (4) The inverse $(a * b)^{-1}$ is the unique element c of G that satisfies $c * (a * b) = (a * b) * c = e$. Firstly, $(b^{-1} * a^{-1}) * (a * b) = b^{-1} * (a^{-1} * a) * b = b^{-1}*e*b = b^{-1}*b = e$. Secondly, $(a*b)*(b^{-1}*a^{-1}) = a*(b*b^{-1})*a^{-1} = a*e*a^{-1} = a*a^{-1} = e$. Since $b^{-1} * a^{-1}$ has the properties that uniquely characterise the inverse of $a * b$, we see that $(a * b)^{-1} = b^{-1} * a^{-1}$. □

4.3 Rings

Definition. A ring is a set R which comes equipped with two operations, $+$ (addition) and $\times$ (multiplication), satisfying the following axioms:

(R+0) If a, b are elements of R, then $a + b$ is an element of R.

(R+1) If a, b, c are elements of R, then $a + (b + c) = (a + b) + c$.

(R+2) There is an element 0 in R such that $a+0 = 0+a = a$ for every element of R– the element is sometimes referred to as the additive identity element, or the identity element with respect to $+$/addition.

(R+3) For every element a of R, there exists b in G such that $a + b = b + a = 0$.

(R+4) If a, b are elements of R, then $a + b = b + a$.

(R×0) If a, b are elements of R, then $a \times b$ is an element of R.

(R×1) If a, b, c are elements of R, then $a \times (b \times c) = (a \times b) \times c$.

(R×+) If a, b, c are elements of R, then

$$a \times (b + c) = a \times b + a \times c.$$

(R+×) If a, b, c are elements of R, then

$$(b + c) \times a = b \times a + c \times a.$$

Remark. The first five axioms say that $(G, *) = (R, +)$ is an additive (abelian) group.

Remark. As seen in groups, the operations $+$ and $\times$ are just symbols/names given to operations that satisfy a bunch of conditions that pin down $+$ and $\times$ on $\mathbb{Z}$ (it is precisely for this reason that the symbols '$+$' and '$\times$' are used conventionally). See examples below.

Remark. We often write ab instead of $a \times b$.

Definition. A ring R is said to be a commutative ring if $a \times b = b \times a$ holds for all a, b in R.

Examples.

• $\{0\}$, where 0 is the additive identity element in $\mathbb{Z}$ with addition $0 + 0 = 0$ and $0 \times 0 = 0$, is a (commutative) ring– this is the smallest ring there is.

• Let $(G, *)$ be an abelian group (with identity element e). Define $+$ in terms of $*$; and define $\times$ by $a \times b = e$ for all elements a, b in G. Then $(G, +, \times)$ is a commutative ring.

• $(\mathbb{Z}, +, \times)$ is a commutative ring.

• $(\{\text{The non-negative integers}\}, +, \times)$ is not a ring, because they do not have inverses with respect to $+$.

• $(\{\text{The positive integers}\}, +, \times)$ is not a ring, because there is no additive identity element '0'.

• The set $\mathbb{R}[X]$ of polynomials in one variable X with coefficients in $\mathbb{R}$ is a commutative ring (this may be thought of as infinitely many copies of $\mathbb{R}$). We will revisit this example again, so we will be brief. An element of $\mathbb{R}[X]$ is of the form $f = c_n(f)X^n + c_{n-1}(f)X^{n-1} + \cdots + c_1(f)X + c_0(f)$ and it is said to be a polynomial of degree $n = \deg(f)$ with coefficients $c_n(f), c_{n-1}(f), \ldots c_0(f)$ in $\mathbb{R}$, when c_n is non-zero.

• The set $M_2(\mathbb{R})$ of n-by-n matrices with entries in $\mathbb{R}$ is a ring but not a commutative ring. More generally, if R is a ring, the set $M_n(R)$ of n-by-n matrices with entries in R is a (non-commutative) ring.

• $\mathbb{Z}[\sqrt{-1}] = \{a + b\sqrt{-1} \mid a, b \in \mathbb{Z}\}$ is a commutative ring with addition $(a + b\sqrt{-1}) + (c + d\sqrt{-1}) = (a + c) + (b + d)\sqrt{-1}$ and multiplication $(a + b\sqrt{-1})(c + d\sqrt{-1}) = (ac - bd) + (ad + bc)\sqrt{-1}$. This ring is often referred to as the Gaussian integers (named after F. Gauss). If you are suddenly gripped by the desired to know more, https://kconrad.math.uconn.edu/blurbs/ugrad-numthy/Zinotes.pdf might be enlightening.

• $\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$ is a commutative ring with addition $(a + b\sqrt{2}) + (c + d\sqrt{2}) = (a + c) + (b + d)\sqrt{2}$ and multiplication $(a + b\sqrt{2})(c + d\sqrt{2}) = (ac + 2bd) + (ad + bc)\sqrt{2}$.

• If R and S are rings, then the Cartesian product $R \times S = \{(r, s) \mid r \in R, s \in S\}$ is a ring by addition and multiplication on each coordiante.

• The set of all (resp. continuous, resp, differentiable etc.) functions on $\mathbb{R}$ to itself is a ring.

We shall see more examples and will study them in depth.

4.4 Elementary 'additive' properties of rings

Proposition 15. Let $(R, +, \times)$ be a ring.

- There is a unique zero element,
- Any element has a unique additive inverse.
- If $a + b = a + c$, then $b = c$.

Proof. This is proved in Proposition 14 (because $(R, +)$, amongst other thing, is an (additive) abelian group), but we spell out details, just in case.

Suppose that 0 and $0'$ are elements of R satisfying $a + 0 = 0 + a = a$ and $a + 0' = 0' + a = a$ for every element a of R. It suffices to show that $0 = 0'$. Letting $a = 0$ in the former, we obtain $(*)$ $0 + 0' = 0'$; while letting $a = 0$ in the latter, we obtain $(**)$ $0 + 0' = 0$. Combining $(*)$ and $(**)$, we have $0 = 0'$, as desired.

Suppose that b and b' are additive inverses of a, i.e. satisfying $a + b = b + a = 0$ and $a + b' = b' + a = 0$. To see $b = b'$, we observe $b = b + 0 = b + (a + b') = (b + a) + b' = 0 + b' = b'$.

To prove the last assertion, observe that $b = 0 + b = (-a + a) + b = -a + (a + b) = -a + (a + c) = (-a + a) + c = 0 + c = c$. □

Proposition 16. Let R be a ring. For every element a of R, we have $0a = a0 = 0$.

Proof. Since 0 is the additive identity, we have $0 + 0 = 0$ (by letting '$a = 0$' in the definition). Multiplying both sides by a, we get $a(0 + 0) = a0$. The LHS equals $a0 + a0$, while the RHS equals $a0 + 0$ (because 0 is the additive identity!). It therefore follows that $a0 + a0 = a0 + 0$. By Proposition 15, we then deduce that $a0 = 0$. A proof for $0a = 0$ is similar. □

4.5 Elementary 'multiplicative' properties of rings

Definition. Let R be a ring. If R has an element 1 (the multiplicative identity element) such that, for every a in R, we have $a \times 1 = 1 \times a = a$, then we say R is a ring with identity (commonly understood as *multiplicative* identity). The additive identity 0 and the multiplicative identity (if exists) do not have to be distinct.

Examples. Most of rings we have (and will have) seen have identity. To add a few,

- $\{0\}$ is a ring with identity– the additive and multiplicative identities are both 0.
- If R is a ring with identity, $\mathrm{M}_n(R)$ is a ring with identity $\begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix}$, where '1' is the identity element of R assured to exist by assumption.
- If R and S are rings with identity, so is $R \times S$ with identity $(1_R, 1_S)$.

Theorem 17. The set $\mathbb{Z}_n$, with addition and multiplication modulo n as defined before, is a commutative ring with identity [1].

Proof. See Chapter 3.

Examples (of rings without identity). It is not very easy to find rings without identity!

• The set of even integers is a ring (with respect to usual $+$ and $\times$) without identity– the set of odd integers is not even a ring!

• Let R be the set of continuous functions $f : \mathbb{R} \to \mathbb{R}$ such that $\int_0^\infty f < \infty$. This is a ring. However, the identity function 1 is not an element of R as $\int_0^\infty 1 = \infty$.

• $(G, *, \times)$ as seen above is not a ring with identity, unless $G = \{e\}$.

Definition. Let R be a ring with identity element 1. An element a in R is called a unit if there is an element b in R such that $ab = ba = 1$. The element b is called the inverse of b, and is written as a^{-1}.

Remark. If R is a ring with identity, an element a is a unit if and only if a has multiplicative inverse. To put it another way,

$$\{\text{units in } R\} = \{\text{elements in } R \text{ with multiplicative inverses}\}.$$

Definition. We will denote by $R^\times$ the units of R.

Examples.

• The units in $\mathbb{Z}$ are exactly $\{-1, 1\}$.

• The units in $\mathrm{M}_2(\mathbb{R})$ are exactly the group $\mathrm{GL}_2(\mathbb{R})$ of invertible (i.e. non-zero determinant) matrices. In fact, it can be n-by-n for any positive integer n, as well as $\mathbb{R}$ can be replaced by any field (to be defined shortly).

• $\mathbb{Z}[\sqrt{-1}]^\times = \{a + b\sqrt{-1} \mid a^2 + b^2 = 1\} = \{1, -1, \sqrt{-1}, -\sqrt{-1}\}$. To see this, observe that $a + b\sqrt{-1}$ is a unit if and only if there exist integers c, d such that $(a + b\sqrt{-1})(c + d\sqrt{-1}) = 1$. Taking the absolute values on both sides, we obtain $(a^2 + b^2)(c^2 + d^2) = 1$. Since $a^2 + b^2 \geqslant 0$, $a^2 + b^2 = 1$.

Proposition 18. The units of $\mathbb{Z}_n$ are the subset of equivalence classes $[a]$ in $\mathbb{Z}$ represented by integers a such that $\gcd(a, n) = 1$. Furthermore, $|\mathbb{Z}_n| = \phi(n)$.

Proof. See Theorem 12 in Chapter 3.

The following proposition puts together some of the key properties of the multiplicative identity 1.

Proposition 19. Let R be a ring with (multiplicative) identity 1.

- The identity element 1 is unique.
- If 1 is distinct from the additive identity 0, then 0 is NOT a unit.
- 1 is a unit and its inverse is 1 itself.

Proof. (1) This can be proved as in the proof of Proposition 14. Suppose that r and s are elements of R satisfying the properties $ra = ar = a$ and $sa = as = a$ for every element a of R. Letting

$a = s$ in the former (resp. $a = r$ in the latter), we obtain $rs = s$ (resp. $rs = r$). Combining, we deduce $r = rs = s$. (2) If 0 were a unit, there exists a say, such that $a0 = 1$. On the other hand, Proposition 16 asserts that $a0 = 0$. This contradicts the assumption that $0 \neq 1$. Hence 0 is not a unit. (3) $1 \times 1 = 1$, hence 1 is a unit and it is the inverse of itself. □

Proposition 20. Let R be a ring with (multiplicative) identity 1.

- If a is a unit, the inverse of a is unique.
- If a is a unit, then so is a^{-1}– the inverse of a^{-1} is indeed a.
- If a and b are units, then so is ab; and its inverse is $b^{-1}a^{-1}$.

Proof. The assertions are proved as in the proof of Proposition 14. (1) Suppose that b and b' are elements of R such that $ab = ba = 1$ and $ab' = b'a = 1$. It then follows that $b = b \times 1 = b(ab') = (ba)b' = 1 \times b' = b'$ (because $ab' = 1$ and $ba = 1$ by assumption). (2) Since $aa^{-1} = a^{-1}a = 1$, a is the inverse of a^{-1}. (3) Since a (resp. b) is a unit, a^{-1} (resp. b^{-1}) is the unique element of R such that $aa^{-1} = a^{-1}a = 1$ (resp. $bb^{-1} = b^{-1}b = 1$). Then $(b^{-1}a^{-1})(ab) = b^{-1}(a^{-1}a)b = b = b^{-1}b = 1$. Also $(ab)(b^{-1}a^{-1}) = a(bb^{-1})a^{-1} = aa^{-1} = 1$. Therefore, ab is a unit and its inverse is $b^{-1}a^{-1}$ (by the uniqueness established above). □

The frequency with which the proof of Proposition 14 was useful in proving statements in the propositions is suggestive of:

Theorem 21. If $(R, +, \times)$ is a ring with identity, $(R^\times, \times)$ is a group. If, furthermore, $(R, +, \times)$ is commutative, $(R^\times, \times)$ is abelian.

Proof. The last assertion of Proposition 20 shows (G0). (G1) follows by thinking of elements of $R^\times$ as elements of R (and make appeal to (R+1) for R). Since 1 is the (unique) element of R satisfying $1a = a1 = a$ for any element of R, it is certainly the case that $1a = a1 = a$ for any element of $R^\times$ (note that $R^\times$ is a subset of R), hence (G2) holds. The second assertion of Proposition 20 shows (G3). □

We end this section (about rings) by an example and an exercise that I find very instructive. I strongly recommend you study these carefully.

Example. Let $(\mathbb{Z}, +, \times)$ be the ring of integers with usual addition $+$ and multiplication $\times$. Define new addition $\boxplus$:

$$a \boxplus b = a + b + 1$$

and new multiplication

$$a \boxtimes b = a + b + ab$$

in terms of old $+$ and $\times$. Then this is a commutative ring with identity, where the zero identity (the identity element with respect to addition, as prescribed by (R+2)) is -1 and the multiplicative identity is 0!

- (R+0) Since $a + b + 1 \in \mathbb{Z}$, we have $a \boxplus b = a + b + 1 \in \mathbb{Z}$.

• (R+1) On one hand,

$$a \boxplus (b \boxplus c) = a \boxplus (b + c + 1) = a + (b + c + 1) + 1 = a + b + c + 1.$$

On the other hand,

$$(a \boxplus b) \boxplus c = (a + b + 1) \boxplus c = (a + b + 1) + c + 1 = a + b + c + 1.$$

Therefore

$$a \boxplus (b \boxplus c) = (a \boxplus b) \boxplus c.$$

• (R+2) (-1) is the identity element with respect to $\boxtimes$. Indeed,

$$a \boxplus (-1) = a + (-1) + 1 = a$$

and

$$(-1) \boxplus a = (-1) + a + 1 = a.$$

[To find the identity, we need to find b in $\mathbb{Z}$ such that $a \boxplus b = a$ holds for any a. By definition, this is equivalent to finding b satisfying $a + b + 1 = a$, i.e. $b + 1 = 0$. Therefore $b = -1$.]

• (R+3) The inverse of a with respect to $\boxplus$ is $-a - 2$. Indeed,

$$a \boxplus (-a - 2) = a + (-a - 2) + 1 = -1$$

and

$$(-a - 2) \boxplus a = (-a - 2) + a + 1 = -1.$$

[To find the inverse of a, we need to find b such that $a \boxplus b = -1$ (since -1 is the identity with respect to $\boxplus$!) for example. This is equivalent to $a + b + 1 = -1$, i.e., $b = -a - 2$.]

• (R+4)

$$a \boxplus b = a + b + 1 = b + a + 1 = b \boxplus a.$$

• (R×0) Since $a + b + ab \in \mathbb{Z}$, we have $a \boxtimes b = a + b + ab \in \mathbb{Z}$.

• (R× 1) On one hand,

$$a \boxtimes (b \boxtimes c) = a \boxtimes (b + c + bc) = a + (b + c + bc) + a(b + c + bc).$$

On the other hand,

$$(a \boxtimes b) \boxtimes c = (a + b + ab) \boxtimes c = (a + b + ab) + c + (a + b + ab)c.$$

It follows from (R+4), (R×1), (R×+) and (R+×) for $(\mathbb{Z}, +, \times)$ that

$$a \boxtimes (b \boxtimes c) = (a \boxtimes b) \boxtimes c.$$

• (R×+) On one hand,

$$a \boxtimes (b \boxplus c) = a \boxtimes (b + c + 1) = a + (b + c + 1) + a(b + c + 1).$$

On the other hand,

$$(a \boxtimes b) \boxplus (a \boxtimes c) = (a + b + ab) \boxplus (a + c + ac) = (a + b + ab) + (a + c + ac) + 1.$$

It then follows from (R+4), (R×+) and (R+×) for $(\mathbb{Z}, +, \times)$ that

$$a \boxtimes (b \boxplus c) = (a \boxtimes b) \boxplus (a \boxtimes c).$$

- (R+×) On one hand,

$$(b \boxplus c) \boxtimes a = (b + c + 1) \boxtimes a = (b + c + 1) + a + (b + c + 1)a.$$

On the other hand,

$$(b \boxtimes a) \boxplus (c \boxtimes a) = (b + a + ba) \boxplus (c + a + ca) = (b + a + ba) + (c + a + ca) + 1.$$

It then follows from (R+4), (R×+) and (R+×) for $(\mathbb{Z}, +, \times)$ that

$$(b \boxplus c) \boxtimes a = (b \boxtimes a) \boxplus (c \boxtimes a).$$

- $(\mathbb{Z}, \boxplus, \boxtimes)$ is commutative. Since $(\mathbb{Z}, +, \times)$ is a commutative ring,

$$a \boxtimes b = a + b + ab = b + a + ba = b \boxtimes a.$$

- The multiplicative identity with respect to $\boxtimes$ is 0. Indeed,

$$a \boxtimes 0 = a + 0 + a0 = a$$

and

$$0 \boxtimes a = 0 + a + 0a = a.$$

[To find this, we need to find b in $\mathbb{Z}$ such that $a \boxtimes b = a$ holds for every a. This is equivalent to finding b satisfying $a + b + ab = a$, i.e. $b(1 + a) = 0$, holds for every a. Therefore $b = 0$.]

The units of $(\mathbb{Z}, \boxplus, \boxtimes)$ are $\{0, -2\}$. To see this, we need to find integers a (and b) such that $a \boxtimes b = 0$, i.e. $a + b + ab = 0$. This is equivalent to $(a+1)(b+1) = -1$. Therefore, $(a+1, b+1)$ is either $(1, -1)$ or $(-1, 1)$. In other words, (a, b) is either $(0, -2)$ or $(-2, 0)$.

The following exercise taught me a lot about rings and abelian groups. I strongly recommend you have a go at it. This is another example of constructing a (commutative) ring out of (abelian) groups.

Exercise. Let $(G, *)$ is an abelian group with identity e. Given an element g in G and a positive integer n, we write ng to mean $g * \cdots * g$, where g is repeated n times, for brevity. Show that the set $R = \mathbb{Z} \times G$ of ordered pairs (n, g) of elements n in $\mathbb{Z}$ and g in G is a commutative ring with identity $(1, e)$ under the addition

$$(n, g) \boxplus (n', g') = (n + n', g * g')$$

and multiplication

$$(n, g) \boxtimes (n', g') = (nn', ng * n'g).$$

The units of $(R, \boxplus, \boxtimes)$ are $\{\pm 1\} \times G$.

4.6 Fields

Definition. A field is a *commutative* ring $(F, +, \times)$ satisfying the axioms

- $(F, +)$ is an additive group (with identity element 0)
- $(F - \{0\}, \times)$ is a multiplicative group (with identity element 1).
- The additive identity '0' (the identity element in the group $(F, +)$) is distinct from the multiplicative identity '1' (the identity element in the group $(F - \{0\}, \times)$).

Perhaps, it might be useful to spell out the field axioms: a field is a set F which comes equipped with addition $\times$ and multiplication+ which satisfy the following:

- It satisfies (R+0) through to (R+4) [which make $(F, +)$ an additive group with additive identity element 0), (R×0), (R×1), (R×+), (R+×) [which make $(F, +, \times)$ a ring]
- For all elements a and b, $a \times b = b \times a$ [which makes $(F, +, \times)$ a commutative ring]
- There exists an element, denoted 1 ,in F such that for every a in F, $1a = a1 = a$ holds– this is often referred to as the (multiplicative) identity element.
- For every a in $F - \{0\}$, there exists an element b in F such that $ab = ba = 1$– in which case, we write a^{-1} for b.
- $0 \neq 1$.

Remark. If $1 = 0$, then $a = 1 \times a = 0 \times a = 0$ (the last equality needs to be justified; see Proposition ?). So the condition $1 \neq 0$ denies any set with one element $\{1 = 0\}$ any chance of being a field.

Remark. By definition,

$$\text{Field} \Rightarrow \text{Ring} \Rightarrow \text{Group}$$

Remark. Groups encapsulate 'symmetry'. Why rings (and not fields)? In general, elements of a ring do not have (multiplicative) inverses and this is not a bad things and this actually makes rings interesting. For example, the division algorithm would be vacuous if everything in $\mathbb{Z}$ had an inverse (i.e. is divisible).

Examples.
• $\mathbb{Q}, \mathbb{R}$ are fields.
• $\mathbb{Z}$ is a ring but not a field. For example, 2 does not have a multiplicative inverse in $\mathbb{Z}$.

Theorem 22. If p is a prime number, then $\mathbb{F}_p = \mathbb{Z}_p$ is a field.

Proof. Firstly, [0] is distinct from [1]. If not, p would divide 1 and consequently force p to be 1.

Suppose that $[a]$ is not equal to [0]. This means that p does not divide a. It follows that $\gcd(a, p) = 1$ and therefore $[a]$ has multiplicative inverse by Theorem 12. □

The field of complex number is a field. It is worth studying it carefully:

Definition. The set $\mathbb{C}$ of complex numbers is the set of elements of the form $a + b\sqrt{-1}$ where a, b are real numbers.

We define addition and multiplication on $\mathbb{C}$ by

$$(a + b\sqrt{-1}) + (c + d\sqrt{-1}) = (a + c) + (b + d)\sqrt{-1}$$

$$(a + b\sqrt{-1}) \times (c + d\sqrt{-1}) = (ac - bd) + (ad + bc)\sqrt{-1}.$$

Theorem 23. The set $\mathbb{C}$ is a field.

Proof. The non-trivial part of this exercise is to see any non-zero element of $\mathbb{C}$ has a multiplicative inverse. Let $a + b\sqrt{-1}$ is a non-zero element of $\mathbb{C}$– in which case, either a or b is non-zero, and therefore $a^2 + b^2$ is non-zero. It then follows that

$$\frac{a}{a^2 + b^2} + \frac{(-b)}{a^2 + b^2}\sqrt{-1}$$

is a non-zero element of $\mathbb{C}$ and one can check easily

$$(a + b\sqrt{-1}) \left(\frac{a}{a^2 + b^2} + \frac{(-b)}{a^2 + b^2}\sqrt{-1} \right) = 1$$

and

$$\left(\frac{a}{a^2 + b^2} + \frac{(-b)}{a^2 + b^2}\sqrt{-1} \right) (a + b\sqrt{-1}) = 1.$$

□

Remark. We spot the inverse by calculating

$$\frac{1}{a + b\sqrt{-1}} = \frac{(a - b\sqrt{-1})}{(a + b\sqrt{-1})(a - b\sqrt{-1})} = \frac{a}{a^2 + b^2} + \frac{(-b)}{a^2 + b^2}\sqrt{-1}$$

but we should be mindful that the inverse, or rather the symbol, '$\frac{1}{a + b\sqrt{-1}}$' makes sense only if we know that $\mathbb{C}$ is a field– the symbol $\frac{1}{a + b\sqrt{-1}}$ is *defined to be* the (unique) element of $\mathbb{C}$ which yields 1 when multiplied by the non-zero element $a + b\sqrt{-1}$, so without knowing $a + b\sqrt{-1}$ is invertible (i.e. has a multiplicative inverse) in advance, how can we make sense of the element?! It would be a catch 22! if our proof mentions $\frac{1}{a + b\sqrt{-1}}$ at all. To check that $a + b\sqrt{-1}$ is invertible, all we need is to spot an element of $\mathbb{C}$ that does the job and no one asks how we find it. Which is why, in the proof, we are pretending that we magically pull the element $\frac{a}{a^2 + b^2} + \frac{(-b)}{a^2 + b^2}\sqrt{-1}$ out of our hat!

(Subtext) Similarly, it is possible to prove that the set $\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} \,|\, a, b \in \mathbb{Q}\}$ (where addition and multiplication are defined analogously with $\sqrt{2}$ in place of $\sqrt{-1}$) is a field– indeed,

the multiplicative inverse of a non-zero element $a+b\sqrt{2}$ (by assumption, a and b are both non-zero) is

$$\frac{a-b\sqrt{2}}{(a+b\sqrt{2})(a-b\sqrt{2})} = \frac{a}{a^2-2b^2} + \frac{(-b)}{a^2-2b^2}\sqrt{2} \in \mathbb{Q}(\sqrt{2}).$$

Note that a^2-2b^2 is never 0. To see this, suppose $a^2-2b^2=0$, i.e. $a^2=2b^2$ (and aim at finding contradiction). Ultimately, we may conclude from the fact that 2 is not a square of rational number (a/b in our case), but how do we formalise this argument?

4.7 Rings that are not fields

We have special names for rings which satisfy some, but not all, of the axioms a field needs to satisfy.

Definition. We say that a ring R with identity is called a division ring/skew field if it satisfies all the axioms except the commutativity of multiplication ($a \times b = b \times a$ for all a, b in R)– a field assumes the set of non-zero elements is an abelian group with respect to $\times$.

The name 'division ring' is justified by the following assertion:

Proposition 24. Let R be a division ring and a is non-zero element of R. If $ab = ac$, then $b = c$.

Proof. Since a is non-zero, it has an inverse a^{-1} in R. Multiplying $ab = ac$ by this, we get $b = c$. □

Example. Let $1, p, q, r$ be symbols subject to the 'multiplicative relations'

- $1p = p1 = p$, $1q = q1 = q$, $1r = r1 = r$
- $p^2 = -1, q^2 = -1, r^2 = -1$
- $pq = r, qp = -r$,
- $qr = p, rq = -p$,
- $rp = q, qr = -q$

The last three set of relations can be more succinctly described via

$$pqr = -1.$$

Indeed, combined with the first three sets of relations, it is possible to recover the last three (Exercise). Let $\mathbb{H}$ (often referred to as Hamilton's quaternions) be the set of elements of the form $c1 + c(p)p + c(q)q + c(r)r$ where $c, c(p), c(q), c(r)$ range over $\mathbb{R}$. In terms of natural addition and multiplication (prescribed by the conditions), $\mathbb{H}$ defines a division ring.

The table of (row)(column) is as follows:

	1	p	q	r
1	1	p	q	r
p	p	-1	r	$-q$
q	q	$-r$	-1	p
r	r	q	$-p$	-1

By assumption, $pq = -qp, qr = -rq, rp = -pr$ and therefore the ring is evidently non-commutative. The multiplicative inverse is 1 (the element of $\mathbb{H}$ given by $(c, c(p), c(q), c(r)) = (1, 0, 0, 0)$).

Let $a = c + c(p)p + c(q)q + c(r)r$ and $b = c - c(p)p - c(q)q - c(r)r$. It then follows that ab is the (non-negative) real number $\mathcal{R} = c^2 + c(p)^2 + c(q)^2 + c(r)^2$ (exercise!). Therefore, if a is non-zero, i.e. $c, c(p), c(q), c(r)$ are not simultaneously 0, then the inverse of a is

$$\frac{b}{\mathcal{R}} = \frac{1}{\mathcal{R}}\left(c - c(p)p - c(q)q - c(r)r\right) = \frac{1}{\mathcal{R}}c - \frac{1}{\mathcal{R}}c(p)p - \frac{1}{\mathcal{R}}c(q)q - \frac{1}{\mathcal{R}}c(r)r \in \mathbb{H}.$$

The element b plays the same role as the complex conjugation in $\mathbb{C}$!

Remark. If you are interested in how complex numbers and Hamilton's quaternions are related, you might find the following paper:

https://maa.org/sites/default/files/pdf/upload_library/22/Allendoerfer/0025570x.di021097.02p0154a.pdf

inspiring.

4.8 Groups, rings and fields from modular arithmetic

The set $\mathbb{Z}_n$ of equivalence classes with respect to 'congruence mod n' is a rich source of non-trivial examples of groups, rings and fields:

- $(\mathbb{Z}_n, +)$ is a group.
- $(\mathbb{Z}_n, +, \times)$ is a commutative ring with identity. There are $\phi(n)$ units in $\mathbb{Z}_n$. If n is not a prime number, this is neither a field nor a division ring.
- If n is a prime number p, then $\mathbb{Z}_p = \mathbb{F}_p$ is a field.

4.9 Properties of fields

If you are interested in these, take 'Further topics in algebra (Galois theory)'!.

5 Polynomials

5.1 Defining polynomials

Definition. Let R be a ring. A polynomial f in one variable X with coefficients in R is:

$$f = c_n X^n + c_{n-1}X^{n-1} + \cdots + c_1 X + c$$

where $c_n, c_{n-1}, \ldots, c_1, c$ are elements of R which are often referred to as the coefficients of f.

The set of all polynomials in one variable X with coefficients in R will be denoted by $R[X]$.

Definition. The degree, denoted $\deg(f)$, of a non-zero polynomial f (in one variable X) is the largest integer n for which its coefficient 'c_n' of X^n is non-zero.

Definition. A non-zero polynomial $f = c_n X^n + c_{n-1}X^{n-1} + \cdots + c_1 X + c$ of degree n is called monic if the leading coefficient $c_n = 1$. The zero polynomial is defined to be monic.

5.2 Polynomial rings

Theorem 25. If R is a ring, then so is $R[X]$ in terms of addition

$$(f+g)(X) = f(X) + g(X) = \sum_n (c_n(f) + c_n(g))X^n$$

and multiplication

$$(fg)(X) = f(X)g(X) = \sum_n \left(\sum_r c_r(f)c_{n-r}(g) \right) X^n.$$

If R is a ring with identity, then so is $R[X]$. If R is commutative, then so is $R[X]$.

Proof.

• (R+0) Since $c_n(f)$ and $c_n(g)$ are both elements of R, it follows from (R+0) for $(R, +, \times)$ that $c_n(f) + c_n(g)$ is an element of R. Therefore $\sum_n (c_n(f) + c_n(g))X^n \in R[X]$.

• (R+1) Since $c_n(f) + (c_n(g) + c_n(\gamma)) = (c_n(f) + c_n(g)) + c_n(\gamma)$ by (R+1) for $(R, +, \times)$,

$$\sum_n c_n(f)X^n + \sum_n (c_n(g) + c_n(\gamma))X^n = \sum_n (c_n(f) + (c_n(g) + c_n(\gamma)))X^n$$

equals

$$\sum_n ((c_n(f) + c_n(g)) + c_n(\gamma))X^n = \sum_n (c_n(f) + c_n(g))X^n + \sum_n c_n(\gamma)X^n.$$

• (R+2) $0 = \sum_n 0X^n = \cdots 0X^n + \cdots + 0X + 0$ is the identity. For every n,

$$c_n(f) + 0 = 0 + c_n(f) = c_n(f)$$

holds by (R+2) for $(R, +, \times)$. Therefore $\sum_n c_n(f)X^n + 0 = 0 + \sum_n c_n(f)X^n = \sum_n c_n(f)X^n$.

• (R+3) If $f = \sum_n c_n(f)X^n$, the inverse is $\sum_n (-c_n(f))X^n$. This is because for every $n \geqslant 0$

$$c_n(f) + (-c_n(f)) = (-c_n(f)) + c_n(f) = 0$$

holds by (R+3) for $(R, +, \times)$.

• (R+4) By (R+4) for $(R, +, \times)$, we have $c_n(f) + c_n(g) = c_n(g) + c_n(f)$ and therefore

$$\sum_n c_n(f)X^n + \sum_n c_n(g)X^n = \sum_n (c_n(f) + c_n(g))X^n$$

equals

$$\sum_n (c_n(g) + c_n(f))X^n = \sum_n c_n(g)X^n + \sum_n c_n(f)X^n.$$

• (R×0) Fix n. It follows from (R×0) for $(R, +, \times)$ that $c_r(f)c_{n-r}(g) \in R$ for every $0 \leqslant r \leqslant n$. By (R+0) for $(R, +, \times)$, we may then deduce that the coefficient $c_n(fg) = \sum_r c_r(f)c_{n-r}(g)$ of X^n lies in R and therefore that $\sum_n (\sum_r c_r(f)c_{n-r}(g))X^n \in R[X]$.

• (R×1) To prove $f(g\gamma) = (fg)\gamma$, it suffices to compare the coefficients of X^n. The coefficient of X^n on the LHS is

$$\sum_r c_r(f)c_{n-r}(g\gamma) = \sum_r c_r(f) \left(\sum_s c_s(g)c_{(n-r)-s}(\gamma) \right) = \sum c_p(f)c_q(g)c_r(\gamma)$$

where the rightmost sum ranges over the set of all non-negative integers p, q and r satisfying $p + q + r = n$, while the coefficient on the RHS is

$$\sum_r c_r(fg)c_{n-r}(\gamma) = \sum_r \left(\sum_s c_s(f)c_{r-s}(g) \right) c_{n-r}(\gamma) = \sum c_p(f)c_q(g)c_r(\gamma).$$

• (R×+) To prove $f(g + \gamma) = fg + f\gamma$, it suffices to compare the coefficient of X^n. The coefficient on the LHS is

$$\sum_r c_r(f)c_{n-r}(g + \gamma) = \sum_r c_r(f) \left(c_{n-r}(g) + c_{n-r}(\gamma) \right)$$

which is equal, by (R+×) for $(R, +, \times)$, to

$$\sum_r c_r(f)c_{n-r}(g) + c_r(f)c_{n-r}(\gamma) = \left(\sum_r c_r(f)c_{n-r}(g) \right) + \left(\sum_r c_r(f)c_{n-r}(\gamma) \right) = c_n(fg) + c_n(f\gamma).$$

• (R+×) A proof of $(g + \gamma)f = gf + \gamma f$ is similar to (R×+) and is left as an exercise. We make appeal to (R+×) for $(R, +, \times)$ instead.

• $R[X]$ is commutative when R is. If $(R, +, \times)$ is commutative, $c_r(f)c_{n-r}(g) = c_{n-r}(g)c_r(f)$ and therefore

$$c_n(fg) = \sum_r c_r(f)c_{n-r}(g) = \sum_r c_{n-r}(g)c_r(f) = \sum_s c_s(g)c_{n-s}(f) = c_n(gf).$$

• $R[X]$ has a multiplicative unit if R does. Let 1 be the multiplicative unit R has and, by slight abuse of notation, let 1 again denote the polynomial $1 = \cdots + 0X^n + \cdots + 0X + 1$ of degree 0 with constant term 1, i.e. the polynomial 1 with $c_n(1) = 0$ for every $n \geqslant 1$ and $c(1) = 1$. To establish $f \times 1 = 1 \times f = 1$, we compare the coefficients of X^n for every $n \geqslant 0$. For $n \geqslant 1$, we have

$$c_n(f \times 1) = \sum_r c_r(f)c_{n-r}(1) = 0 + \cdots + 0 + c_n(f)c(1) = c_n(f) \times 1 = c_n(f)$$

by Proposition 16, (R+2) for $(R, +, \times)$ and the fact that 1 is the multiplicative identity. Similarly,

$$c_n(1 \times f) = \sum_r c_r(1)c_{n-r}(f) = c(1)c_n(f) + 0 + \cdots + 0 = c_n(f).$$

For $n = 0$, we have

$$c(f)c(1) = c(f) \times 1 = c(f)$$

and

$$c(1)c(f) = 1 \times c(f) = c(f).$$

□

Proposition 26. If $(R, +, \times)$ is a ring with identity 1, then $R[X]$ is not a division ring.

Proof. Suppose, firstly, that R consists only of one element– the element is necessarily the additive identity 0 of R. It then follows that $R[X] = \{0\}$, as f with $c_n(f) = 0$ for every n is necessarily the 'polynomial' 0. However, this forces $R[X]$ not to be a division ring as the condition $1 \neq 0$ does not hold.

Having dealt with the case that R consists of one element, we may assume now that $R \neq \{0\}$. In this case, there exists a non-zero element c in R. Consider the polynomial cX of degree 1. It suffices to prove that cX does not have multiplicative inverse (if $R[X]$ were a division ring, then any element would have multiplicative inverse). If cX had a multiplicative inverse, then there should be a polynomial $f = c_n(f)X^n + \cdots + c_1(f)X + c(f)$ such that $f \times cX = 1$. However,

$$f \times cX = (cc_n(f))X^n + \cdots + (cc_1(f))X^2 + (cc(f))X$$

and comparing the constant terms, we deduce that $1 = 0$. However, this would have implied that $R = \{0\}$ which we know should not occur. □

Remark. By definition, $\deg(f)\deg(g) \geqslant \deg(fg)$. Let $f = \sum_n c_n(f)X^n$ and $g = \sum_n c_n(g)X^n$. By definition, $c_n(f) =$ for every $n \geqslant \deg(f)$ while $c_n(f)$ is non-zero when $n = \deg(f)$. Similarly for g. Since

$$fg = \sum_n \left(\sum_r c_r(f)c_{n-r}(g) \right) X^n = c(f)c(g) + (c(f)c_1(g) + c_1(f)c(g))\, X + \cdots + c_{\deg(f)}c_{\deg(g)}X^{\deg(f)+\deg(g)},$$

we see that $\deg(fg) \leqslant \deg(f) + \deg(g)$ where the equality holds exactly when $c_{\deg(f)}c_{\deg(g)}$ is non-zero. For example, if $R = \mathbb{Z}_6$ and $c_{\deg(f)} = [2]$ and $c_{\deg(g)} = [3]$, then $c_{\deg(f)}c_{\deg(g)} = [2][3] = [6] = [0]$ and therefore $\deg(fg) < \deg(f) + \deg(g)$.

Remark (non-examinable). If R is a ring with the property– if any pair of elements a and b of R are non-zero, then their product ab is again non-zero– then $\deg(fg) = \deg(f) + \deg(g)$ always holds. A commutative ring with this property is called an *integral domain*. One of the most important example of an integral domain is $\mathbb{Z}$. Another important example is a field. And it is for this reason, we shall specialised the coefficient ring to be a field from now on.

Proposition 27. Let $(F, +, \times)$ be a field. The units $F[X]^\times$ of $F[X]$ are $F^\times = F - \{0\}$.

Proof. Let f be a unit in $F[X]$. Then there exists a polynomial g in $F[X]$ such that $fg = gf = 1$. By the remark above, $\deg(f) + \deg(g) = \deg(fg) = \deg(1) = 0$. Therefore $\deg(f) = \deg(g) = 0$, i.e. f and g are non-zero constants in F whose product is 1, in other words, f and g are units in F. □

Remark (non-examinable) The assertion of Proposition 27 holds with an integral domain in place of F. If R is (merely) a commutative ring with identity, then the units are the group of polynomials f with the property that $c(f) \in R^\times$ and $c_n(f)$ is nilpotent (i.e. its sufficiently large power is 0) for every $n \geqslant 1$. See for example https://kconrad.math.uconn.edu/blurbs/ringtheory/polynomial-properties.pdf for a proof (and much more).

5.3 Polynomial division

Theorem 28 (Division algorithm in the context of the polynomial ring $F[X]$). Let F be a field. Let f and g be two polynomials in $F[X]$ and assume, in particular, that g is non-zero. Then there exists polynomials q and r in $F[X]$ such that

$$f = gq + r$$

where either $r = 0$ or $\deg(r) < \deg(g)$.

Proof. We prove the theorem by induction on the degree of f.

• Suppose $\deg(f) < \deg(f)$. Then

$$f = g \cdot 0 + f$$

(i.e. $q = 0$ and $r = f$) holds.

• Suppose, for any polynomial f' of degree $< \deg(f)$, the assertion of the theorem holds (with the same g!), i.e., there exists q' and r' in $F[X]$ such that

$$f' = gq' + r'$$

where r' is either 0 or $\deg(r') < \deg(g)$. The goal is to show for f (of degree $\deg(f)$!) there are q and r as above. By the case already dealt with above, we may assume

$$\deg(f) \geqslant \deg(g)$$

and let

$$f' = f'(X) = f(X) - \frac{c_{\deg(f)}(f)}{c_{\deg(g)}(g)} X^{\deg(f)-\deg(g)} g(X).$$

Then $c_n(f') = 0$ for every $n > \deg(f)$ and

$$c_{\deg(f)}(f') = c_{\deg(f)}(f) - \frac{c_{\deg(f)}(f)}{c_{\deg(g)}(g)} c_{\deg(g)}(g) = 0.$$

Therefore $\deg(f') < \deg(f)$. By the inductive hypothesis, there exists q' and r' in $F[X]$ such that

$$f' = q'g + r'$$

where $r' = 0$ or $\deg(r') < \deg(g)$. It therefore follows from the definition of f' that

$$f = \left(q' + \frac{c_{\deg(f)}(f)}{c_{\deg(g)}(g)} X^{\deg(f)-\deg(g)} \right) g + r'$$

as desired. □

5.4 Roots and factors

Definition. Let f and g be polynomials in $F[X]$. We say that g divides f, or g is a factor of f, if there exists a polynomial q in $F[X]$ such that $f = gq$.

Remark. One needs to be careful when it come to polynomial division. Suppose g divides f. Then, for every unit γ in $F[X]$, the product $g\gamma$ also divides f! By Proposition 27, we know that $F[X]^\times = F - \{0\}$, hence this assertions amounts to saying that if g divides f, then any non-zero constant multiple of g also divides f.

The divisibility of a polynomial depends on F:

Examples.

$X+\sqrt{-1}$ divides X^2+1 in $\mathbb{C}[X]$. Indeed, $(X+\sqrt{-1})(X-\sqrt{-1}) = X^2-(\sqrt{-1})^2 = X^2+1$.

On the other hand, no non-trivial polynomial in $\mathbb{Q}[X]$ divides $f(X) = X^2+1$ in $\mathbb{Q}[X]$! Firstly, any degree 0 polynomial in $\mathbb{Q}[X]$ divides $f(X)$ because a polynomial in $\mathbb{Q}$ of degree 0 is nothing other than an element c of $\mathbb{Q} - \{0\}$, hence $f = c(c^{-1}f)$. Similarly, the only degree 2 polynomial of degree 2 that divides f is f itself. Indeed, if g of degree 2 divides f, then there exists γ in $\mathbb{Q}[X]$ such that $g\gamma = f$. Since $\deg(g) + \deg(\gamma) = \deg(f)$, then $\deg(\gamma) = 0$, i.e. γ is an element of $\mathbb{Q} - \{0\}$. Therefore, g is forced to be $\gamma^{-1}f$. To see that no polynomial of degree 1 in $\mathbb{Q}[X]$ divides f, it suffices to establish that $X^2 + 1$ does not factorises as the product $(X + a)(X + b)$ of degree one polynomials, i.e. there are no rational numbers a and b such that $a + b = 0$ and $ab = 1$ (by comparing the coefficients). Suppose for contradiction that it does. It then follows from $a+b = 0$ that $b = -a$ and substituting this into $ab = 1$, we get $-a^2 = 1$. Since $-a^2 \leqslant 0$, this is a contradiction.

Corollary 29. Let F be a field. Let f in $F[X]$ and α be an element of F. Then there exists q in $F[X]$ and r in F such that

$$f = (X - \alpha)q + r.$$

Proof. This follows from the theorem with $g = X - \alpha$. □

Corollary 30. Let f in $F[X]$ and α in F. The remainder of f when divided by $(X - \alpha)$ is $f(\alpha)$. In particular, $f(\alpha) = 0$ if and only if $X - \alpha$ is a factor of $f(X)$ in $F[X]$.

Proof. It follows from the corollary (by letting $X = \alpha$) that $f(\alpha) = r$. If $f(\alpha) = 0$, it therefore follows from the corollary that $f = (X - \alpha)q$ and $X - \alpha$ is a factor of f. Conversely, if $X - \alpha$ is a factor of f, there exists q in $F[X]$ such that $f = (X - \alpha)q$. Letting $X = \alpha$, we deduce that $f(\alpha) = 0$. □

We may use the corollary to check if a given polynomial factorises or not factorises at all.

Example. Consider $f(X) = X^2 + 3$ in $\mathbb{F}_7[X]$. Then $X - 2$ divides $X^2 + 3$, Indeed,

$$[2]^2 + [3] = [4] + [3] = [7] = [0],$$

i.e. $f([2]) = [0]$. In fact, $X + 2$ also divides f as

$$[-2]^2 + [3] = [4] + [3] = [7] = [0],$$

i.e. $f([-2]) = f(-[2]) = [0]$.

Example. The polynomial $f(X) = X^2 + 2$ is irreducible in $\mathbb{F}_5[X]$, i.e. no non-trivial polynomial in $\mathbb{F}_5[X]$ divides f. To see this, we observe that no polynomial of the form $X - \alpha$ divides f in $\mathbb{F}_5[X]$. By Corollary 30, this is equivalent to checking that no α in $\mathbb{F}_5$ satisfy $f(\alpha) = 0$. Indeed,

α	[0]	[1]	[2]	[3]	[4]
$f(\alpha)$	[2]	[3]	[1]	[1]	[3]

Definition. Let N be a non-negative integer. An element α in F is a root of multiplicity N of a polynomial f in $F[X]$, if $(X - \alpha)^N$ is the highest power of $(X - \alpha)$ that divides $f(X)$.

5.5 The fundamental theorem of algebra

Definition. Let F be a field. We say that α is a root, or zero, of the polynomial $f(X) = c_n X^n + \cdots + c_1 X + c$ in $F[X]$ if $f(\alpha) = 0$, i.e. $c_n\alpha^n + c_{n-1}\alpha^{n-1} + \cdots + c_1\alpha + c_1 X + c = 0$.

Theorem 31.(The Fundamental Theorem of Algebra) Let $n \geq 1$. Let $c, c_1, \ldots, c_n$ be complex numbers, where c_n is assumed to be non-zero. Then the polynomial $c_n X^n + \cdots + c$ has at least one root inside $\mathbb{C}$.

Theorem 32.(The Fundamental Theorem of Algebra with multiplicities) Let $n \geq 1$. Let $c, c_1, \ldots, c_n$ be complex numbers, where c_n is assumed to be non-zero. Then the polynomial $f(X) = c_n X^n + \cdots + c$ has exactly n roots in $\mathbb{C}$ counted with multiplicities, i.e. there exist complex numbers $\alpha_1, \ldots, \alpha_n$ such that

$$f(X) = c_n(X - \alpha_n)(X - \alpha_{n-1}) \cdots (X - \alpha_1).$$

These theorems are proved, for example, by complex analysis! Needless to say, proofs are non-examinable (and I won't even try to spell them out either!). Look at H. A. Priestley's 'Introduction to Complex Analysis', Oxford University Press.

5.6 GCDs of polynomials

Theorem 33.

- Any two polynomials f and g have a greatest common divisor in $F[X]$.
- The gcd of two polynomials in $F[X]$ can be found by Euclid's algorithm.
- If $\gcd(f, g) = \gamma$ (a polynomial in $F[X]$), then there exist p, q in $F[X]$ such that

  $$fp + gq = \gamma;$$

  these polynomials p and q can also be found from the extended Euclid's algorithm.

Proof. Non-examinable. Similar to the proof in the setting of $\mathbb{Z}$ though. $\square$

Examples.

• Let $f = X^4 + 1$ and $g = X^2 + X$ in $\mathbb{Q}[X]$. What is $\gcd(f, g)$? Since

$$\begin{aligned} X^4 + 1 &= (X^2 - X + 1)(X^2 + X) + (-X + 1) \\ X^2 + X &= (-X - 2)(-X + 1) + 2 \\ -X + 1 &= \frac{1}{2}(-X + 1) \cdot 2 + 0, \end{aligned}$$

the gcd is 1 (not 2!) since gcd is defined to be monic. Note that if 2 is a common divisor, any $F^\times$-multiple of 2 is also a common divisor. Because gcd is defined to be monic, we are forced to choose 1, instead of 2.

To find p, q such that $fp + gq = \gcd(f, g) = 1$, we do something analogous to what we saw in Euclid's algorithm for $\mathbb{Z}$. Indeed, since

$$\begin{aligned} 2 &= (X^2 + X) - (-X - 2)(-X + 1) \\ &= (X^2 + X) + (X + 2)((X^4 + 1) - (X^2 - X + 1)(X^2 + X)) \\ &= (X + 2)(X^4 + 1) + (-X^3 - X^2 + X - 1)(X^2 + X) \end{aligned}$$

we have

$$\gcd(f, g) = 1 = \frac{1}{2}(X + 2)f + \frac{1}{2}(-X^3 - X^2 + X - 1)g.$$

• Let $f = X^4 + 2X^3 + X^2 - 4$ and $g = X^3 - 1$ in $\mathbb{Q}[X]$. What is gcd?

$$\begin{aligned} X^4 + 2X^3 + X^2 - 4 &= (X + 2)(X^3 - 1) + (X^2 + X - 2) \\ X^3 - 1 &= (X - 1)(X^2 + X - 2) + (3X - 3) \\ X^2 + X - 2 &= \frac{1}{3}(X + 2)(3X - 3) + 0 \end{aligned}$$

and therefore $\gcd(f, g) = X - 3$. As before, as soon as $3X - 3$ is a common divisor of f and g in $\mathbb{Q}[X]$, we know that any $F^\times$-multiple of $3X - 3$ is also a common divisor. Amongst those, the only one is monic and that is $X - 1$ which is the gcd.

To find p and q such that $fp + gq = \gcd(f, g)$, we see that

$$\begin{aligned} 3X - 3 &= (X^3 - 1) - (X - 1)(X^2 + X - 2) \\ &= g - (X - 1)(f - (X + 2)g) \\ &= (-X + 1)f + (X^2 - X - 1)g. \end{aligned}$$

• Let $f = X^4 + [1]$ and $g = X^2 + X$ in $\mathbb{F}_2[X]$. What is gcd in $\mathbb{F}_2[X]$?

Since $X^4 + [1] = (X + [1])^4$ in $\mathbb{F}_2[X]$, we work with $(X + [1])^4$ instead. Since $g(X) = X(X + [1])$, both $f = (X + [1])^4$ and $g = X(X + [1])$ are divisible by $X + [1]$ exactly once. Since

$$\gcd\left(\frac{f}{X + [1]}, \frac{g}{X + [1]}\right) = \gcd((X + [1])^3, X) = 1,$$

the gcd is $X + [1]$. Alternatively, we may follow 'Euclid's algorithm':

$$\begin{aligned} (X + [1])^4 &= ((X + [1])^2 + (X + [1]) + [1])(X^2 + X) + (X + [1]) \\ X^2 + X &= X(X + [1]) + 0. \end{aligned}$$

and conclude that $\gcd(f,g) = X + [1]$ in $\mathbb{F}_2[X]$. To find p, q, we simply see that

$$\gcd(f,g) = X+[1] = 1\cdot(X+[1])^4-((X+[1])^2+(X+[1])+[1])(X^2+X) = 1\cdot f+(X^2+X+1)g.$$

5.7 Power series rings (non-examinable)

Definition. Let R be a ring. A power series f in one variable X with coefficients in R is:

$$f = c + c_1X + \cdots + c_nX^N + \cdots = \sum_n c_nX^n$$

where c_n, for every n, is an element of R.

The set of all power series in one variable X with coefficients in R will be denoted by $R[[X]]$. This is a ring with addition and multiplication defined similarly to the one for $R[X]$.

What is the difference between $R[X]$ and $R[[X]]$? For example, $1 - X$ is not a unit in $R[X]$ and it is a unit in $R[[X]]$ as

$$(1 - X)(1 + X + X^2 + \cdots) = 1.$$

6 Matrices

Let $(R, +, \times)$ be a ring and let $\mathrm{M}_2(R)$ be the set of 'matrices'

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

where a, b, c, d are elements of R, together with addition

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \begin{pmatrix} a+a' & b+b' \\ c+c' & d+d' \end{pmatrix}$$

and multiplication

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \begin{pmatrix} aa'+bc' & ab'+bd' \\ ca'+dc' & cb'+db' \end{pmatrix}.$$

Theorem 34. $\mathrm{M}_2(R)$ is a ring. If R is a ring with identity, then so is $\mathrm{M}_2(R)$.

Proof. Exercise. □

Remark. The additive identity, the identity element with respect to $+$ defined above, is $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, where each entry 0 is the additive identity in R as defined in (R+2). If R is a ring with identity 1, then $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is the identity. Indeed,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a\times 1+b\times 0 & a\times 0+b\times 1 \\ c\times 1+d\times 0 & c\times 0+d\times 1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

The $(1,1)$-entry is a because $a \times 1 = a$ (since 1 *is* the element of R satisfying $a \times 1 = 1 \times a = a$) and $b \times 0$ (by Proposition 16), therefore $a \times 1 + b \times 0 = a + 0 = a$ by (R+2) for $(R, +, \times)$.

Remark. In contrast to Theorem 25, $\mathrm{M}_2(R)$ is never commutative, even if R is commutative. Let us see this in an example. Let $A = \begin{pmatrix} [1] & [1] \\ [0] & [1] \end{pmatrix}$ and $B = \begin{pmatrix} [1] & [1] \\ [1] & [1] \end{pmatrix}$ be matrices in $\mathrm{M}_2(\mathbb{F}_2)$, where $\mathbb{F}_2$ is the field with two elements $[0]$ and $[1]$. Following the formula above, together with $[1] + [1] = [2] = [0]$, we see that

$$AB = \begin{pmatrix} [0] & [0] \\ [1] & [1] \end{pmatrix}$$

while

$$BA = \begin{pmatrix} [1] & [0] \\ [1] & [0] \end{pmatrix}.$$

Proposition 35 If $(R, +, \times)$ is a ring with identity but is not a ring with the property that for every elements a, b in R, the product is always $ab = 0$, then $\mathrm{M}_2(R)$ is neither commutative nor a division ring.

Remarks. An example of those rings *excluded* is the ring $(G, *, \times)$ given by a group $(G, *)$ with multiplication $a \times b = e$ for all a, b in G. A field is an example of those rings considered in the proposition.

Proof. The assumption amounts to the existence of elements a, b in R such that ab is not 0 (the additive identity). By Proposition 16, neither a nor b is 0. We use these two elements to prove the assertions of the proposition.

Following the definition of multiplication in matrices, we see that

$$\begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & ab \\ 0 & 0 \end{pmatrix}$$

and

$$\begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

and therefore the ring is not commutative.

To show that $\mathrm{M}_2(R)$ is not a division ring, we show that $\begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix}$ does not have a multiplicative inverse, i.e. there is no matrix A in $\mathrm{M}_2(R)$ that satisfies the relation $A \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Suppose, for contradiction, that such a matrix A exists. In which case, since $\mathrm{M}_2(R)$ is a ring (Theorem 34), it follows from (R×1) that

$$A \left[\begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix} \right] = \left[A \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} \right] \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}$$

holds. However, the LHS equals

$$A \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

while the RHS equals

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}.$$

Since a is not 0, this is a contradiction. Therefore $\mathrm{M}_2(R)$ is not a division ring.□

7 Permutations

7.1 Definition

Definition. A permutation of a set S is a function $f : S \to S$ which is bijection (one-to-one and onto).

Definition. The set of permutations on the set $\{1, \ldots, n\}$ is denoted S_n and every element f in S_n is written as

$$f = \begin{pmatrix} 1 & \cdots & n \\ f(1) & \cdots & f(n) \end{pmatrix}.$$

Example. The element $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 7 & 3 & 8 & 1 & 5 & 2 & 6 \end{pmatrix}$ in S_8 is a bijection $S_8 \to S_8$ which sends 1 to 4, 2 to 7, 3 to 3, ...

Remark. Since f is a bijection, if a and b are distinct, so are $f(a)$ and $f(b)$, i.e., none of the numbers in the first row appears more than once in the second row, i.e. the second row is obtained by 'shuffling' the numbers in the first row.

Remark. Whilst it is rather rare to come across, strictly speaking, it is not necessary to have the first row in the order $1, 2, \ldots$.

For example,

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 7 & 3 & 8 & 1 & 5 & 2 & 6 \end{pmatrix} = \begin{pmatrix} 2 & 3 & 5 & 1 & 8 & 6 & 7 & 4 \\ 7 & 3 & 1 & 4 & 6 & 5 & 2 & 8 \end{pmatrix}$$

simply because both expressions contain the same set of information that characterises the permutation.

Proposition 36. $|S_n| = n!$.

Proof. This follows by definition. □

7.2 Composition

Definition. If f and g are permutations on S, we define the composition, denoted $f \circ g$ to be the which sends s in S to $f(g(s))$.

Proposition 37. If f and g are elements of S_n, then so is the composite $f \circ g$ is in S_n.

Proof. It suffices to show that $f \circ g$ is bijective. To prove injectivity, suppose $(f \circ g)(x) = (f \circ g)(y)$, and we aim at establishing that $x = y$. Since $f(g(x)) = f(g(y))$, it follows from the injectivity of f that $g(x) = g(y)$ holds. By the injectivity of g, it follows that $x = y$.

To prove the surjectivity, suppose that z is an element of S. By the surjectivity of f, there exists y in S such that $f(y) = z$. It then follows from the surjectivity of g that there exists x such that $y = g(x)$. Combining, $z = f(y) = f(g(x)) = (f \circ g)(x)$. □

Example. Let

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 7 & 3 & 8 & 1 & 5 & 2 & 6 \end{pmatrix}$$

and

$$g = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 6 & 3 & 1 & 8 & 7 & 2 & 5 & 4 \end{pmatrix}$$

Then $f \circ g$ sends

$$\begin{array}{ccccc} & g & & f & \\ 1 & \mapsto & 6 & \mapsto & 5 \\ 2 & \mapsto & 3 & \mapsto & 3 \\ 3 & \mapsto & 1 & \mapsto & 8 \\ 4 & \mapsto & 8 & \mapsto & 6 \\ 5 & \mapsto & 7 & \mapsto & 2 \\ 6 & \mapsto & 2 & \mapsto & 7 \\ 7 & \mapsto & 5 & \mapsto & 1 \\ 8 & \mapsto & 4 & \mapsto & 8 \end{array}$$

or to put it another way,

$$f \circ g = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 5 & 3 & 4 & 6 & 2 & 7 & 1 & 8 \end{pmatrix}.$$

Proposition 38. If f is in S_n, then the inverse function f^{-1} exists and is an element of S_n.

Proof. Given an element y of S_n, define T_y to be the subset $\{x \,|\, f(x) = y\}$ of S_n. Since f is surjective, T_y is non-empty, i.e. $|T_y| \geqslant 1$. Since f is injective, $|T_y| = 1$. Indeed, if x and x' are non-trivial elements of T_y, then $f(x) = y = f(x')$. Since f is injective $x = x'$.

Since $|T_y| = 1$, T_y is an element of S_n (rather than a subset of S_n); and it makes sense to define a function $g : S_n \to S_n$ by decreeing that it sends y to T_y. By definition, $f \circ g = 1$ and $g \circ f = 1$ (exercise!). In fact, the bijectivity of g follows from these, but we deduce it directly from the definition. To prove the injectivity, suppose $g(x) = g(y)$. Applying f to these elements, we see that $x = y$. To prove the surjectivity, let x be an element of S_n. Let $y = f(x)$. Then $x = T_y = g(y)$. □

Example. If $f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 7 & 3 & 8 & 1 & 5 & 2 & 6 \end{pmatrix}$, then one reads 'from the second row to the first row' to work out f^{-1}. For example, to work out where f^{-1} sends 1 to, it follows by the definition of the inverse f^{-1} of f that we need to find an integer (uniquely determined) which gets mapped to 1 by f. Looking at the expression of f, the number we seek necessarily lies above 1. This is 5.

What we get in the end is:

$$\begin{array}{ccc} & f^{-1} & \\ 1 & \mapsto & 5 \\ 2 & \mapsto & 7 \\ 3 & \mapsto & 3 \\ 4 & \mapsto & 1 \\ 5 & \mapsto & 6 \\ 6 & \mapsto & 8 \\ 7 & \mapsto & 2 \\ 8 & \mapsto & 4 \end{array}$$

or to put it another way,

$$f^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 5 & 7 & 3 & 1 & 6 & 8 & 2 & 4 \end{pmatrix}.$$

7.3 Cycles

Consider:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 3 & 4 & 1 & 8 & 6 & 7 & 5 \end{pmatrix}$$

Then $1 \mapsto 2 \mapsto 3 \mapsto 4 \mapsto 1 \mapsto \cdots$ so it is a loop that involves four numbers $1, 2, 3$ and 4. Similarly, $5 \mapsto 8 \mapsto 5 \mapsto \cdots$ is a loop that involves two numbers, while 6 (resp. 7) remains unchanged. So we are seeing four different 'loops' of various length. Each of these loops is called a cycle.

Definition. Let $\gamma_1, \ldots, \gamma_\tau$ denote *distinct* elements of $\{1, \ldots, n\}$ (necessarily, $1 \leqslant \tau \leqslant n$). The cycle $(\gamma_1, \gamma_2, \ldots, \gamma_\tau)$ is the permutation of S_n which sends γ_1 to γ_2, ..., $\gamma_{\tau-1}$ to γ_τ ,and γ_τ to γ_1, while maintaining those elements NOT in $\{\gamma_1, \ldots, \gamma_\tau\}$ unchanged. Following the representation earlier, this is the element

$$\begin{pmatrix} 1 & \cdots & \gamma_1 & \cdots & \gamma_{\tau-1} & \cdots & \gamma_\tau & \cdots & n \\ 1 & \cdots & \gamma_2 & \cdots & \gamma_\tau & \cdots & \gamma_1 & \cdots & n \end{pmatrix}$$

of S_n (if $1 < \gamma_1 < \cdots < \gamma_\tau < n$, of course). By definition,

$$(\gamma_1, \gamma_2, \ldots, \gamma_\tau) = (\gamma_2, \ldots, \gamma_\tau, \gamma_1) = \cdots = (\gamma_\tau, \gamma_1, \ldots, \gamma_{\tau-1}) = \cdots$$

as they all define the same permutation as an element of S_n. By definition, the inverse is given by

$$(\gamma_1, \gamma_2, \ldots, \gamma_\tau)^{-1} = (\gamma_\tau, \gamma_{\tau-1}, \ldots, \gamma_2, \gamma_1) = (\gamma_1, \gamma_\tau, \gamma_{\tau-1}, \ldots, \gamma_2) = \cdots$$

If $(\gamma_1, \ldots, \gamma_\tau)$ and $(\gamma'_1, \ldots, \gamma'_{\tau'})$ share no common element, then we say that these cycles are disjoint.

Example. The example at the beginning may be written as $(1, 2, 3, 4)(5, 8)(6)(7)$ or more commonly $(1, 2, 3, 4)(5, 8)$ (implicitly saying that those not in the list, e.g. 6 and 7, are fixed).

Theorem 39 Any permutation can be written as a composition of disjoint cycles. The representation is unique, up to the facts that

- the cycles can be written in any order,
- each cycle can be started at any point,
- cycles of length 1 can be left out.

Proof. The proof is 'algorithmic', 'constructive' and/or 'inductive'. Starting with the smallest integer in the set, e.g. 1, follow its successive images under f until we return to something we have seen before. This has to be 1 (if we started the process with 1). Indeed, suppose f sends $1 = \gamma_1 \mapsto \gamma_2 \mapsto \cdots \mapsto \gamma_\tau$ sequentially with $\gamma_1, \gamma_2, \ldots, \gamma_{\tau-1}$ all distinct, but γ_τ actually coinciding with one in the string, i.e. there exists $1 < \sigma < \tau$ such that $\gamma_\sigma = \gamma_\tau$, then

$$f(\gamma_{\sigma-1}) = \gamma_\sigma = \gamma_\tau = f(\gamma_{\tau-1}).$$

By the injectivity of f, this forces $\gamma_{\sigma-1} = \gamma_{\tau-1}$ but this contradicts the assumption. The only possibility is $\tau = 1$.

To sum up, as we trace where f sends 1 to (sequentially), the string visits distinct elements of $\{1, \ldots, n\}$ until being forced back to 1 by the argument. Repeat the process. This new string will never have any element that has previously appeared in the old string. □

Example The cycle notation for

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 7 & 3 & 8 & 1 & 5 & 2 & 6 \end{pmatrix}$$

is $(1\,4\,8\,6\,5)(2\,7)(3)$.

Definition. Let f be an element of S_n. The order of f is the smallest number of times we compose f with f itself, $f \circ f \circ f \cdots$, to get the identity.

Remark. The order $|G|$ of a group G is the number of elements in G (if finite). On the other hand, the order of an element g is the smallest N that yields $g^N = (g * \cdots * g) = e$.

Examples. The order of $f = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix} = (13)(24)$ in S_4 is 2, because

$$\begin{array}{ccccccc} & & f & & f & & \\ 1 & \mapsto & 3 & \mapsto & 1 \\ 2 & \mapsto & 4 & \mapsto & 2 \\ 3 & \mapsto & 1 & \mapsto & 3 \\ 4 & \mapsto & 2 & \mapsto & 4 \end{array}$$

while the order of $g = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 1 & 3 \end{pmatrix} = (143)(2)$ is 3, because

$$\begin{array}{ccccccccc} & & g & & g & & g & & \\ 1 & \mapsto & 4 & \mapsto & 3 & \mapsto & 1 \\ 2 & \mapsto & 2 & \mapsto & 2 & \mapsto & 2 \\ 3 & \mapsto & 1 & \mapsto & 4 & \mapsto & 3 \\ 4 & \mapsto & 3 & \mapsto & 1 & \mapsto & 4 \end{array}$$

The order of $\begin{pmatrix}1&2&3&4\\2&3&4&1\end{pmatrix} = (1234)$ in S_4 is 4.

Proposition 40. The order of a permutation is the least common multiple of the lengths of the cycles in the disjoint cycle representation.

Proof. By proposition 39, a permutation is written as the composition of cycles of the form $(\gamma_1, \ldots, \gamma_\tau)$. Each of these has order the length τ of the cycle, i.e.

$$\overbrace{(\gamma_1, \ldots, \gamma_\tau) \circ \cdots \circ (\gamma_1, \ldots, \gamma_\tau)}^{\times\tau}$$

or indeed any positive integer multiple

$$\overbrace{(\gamma_1, \ldots, \gamma_\tau) \circ \cdots \circ (\gamma_1, \ldots, \gamma_\tau)}^{\times\tau} \circ \overbrace{(\gamma_1, \ldots, \gamma_\tau) \circ \cdots \circ (\gamma_1, \ldots, \gamma_\tau)}^{\times\tau} \circ \cdots \circ \overbrace{(\gamma_1, \ldots, \gamma_\tau) \circ \cdots \circ (\gamma_1, \ldots, \gamma_\tau)}^{\times\tau}$$

exactly gives rise to the identity on $\{\gamma_1, \ldots, \gamma_\tau\}$. Doing this in conjunction with all other cycles, the smallest integer 'power' that defines the identity on $\{1, \ldots, n\}$ would therefore be the lcm of the length (or the orders) of the cycles. □

8 Groups revisited

Theorem 41. $(S_n, \circ(\text{composition}))$ is a group.

Proof. Exercise. □

Proposition 42. S_n is an abelian group if $n \leqslant 2$ and is non-abelian if $n > 2$.

Proof. When $n = 1$, S_n consists only of the identity. When $n = 2$, S_n consists of $\begin{pmatrix}1&2\\1&2\end{pmatrix}$ and $\begin{pmatrix}1&2\\2&1\end{pmatrix}$ and is straightforward to see that these two elements commute. When $n \geqslant 2$, S_n contains elements/cycles $f = (12)$ and $g = (23)$ and $f \circ g$ does not equal $g \circ f$ (the former is (123) while the latter is (132)), hence S_n has elements that do not commute. □

Example. $S_3 = \text{Sym}(\{1, 2, 3\}) = S_3 = \{e, r, (r \circ r), s, r \circ s\}$ where $r = \begin{pmatrix}1&2&3\\2&3&1\end{pmatrix}$ and $s = \begin{pmatrix}1&2&3\\1&3&2\end{pmatrix}$. This is not abelian. Do you recognise this example?

8.1 Subgroups

Definition. Let $(G, *)$ be a group and Γ be a subset. We say that Γ is a subgroup of G if $(\Gamma, *)$ is a group.

To recall, it needs to satisfy the following:

(G0) If a and b are elements of Γ, then $a * b$ is an element of Γ.

(G1) If a, b, c are elements of Γ, then $(a * b) * c = a * (b * c)$ holds. Since the equality holds for elements of G, this remains true for elements in Γ.

(G2) Γ contains the identity element e_Γ. In fact, $e_\Gamma = e$ (the identity element of G). To see this, we firstly see that $e_\Gamma * e_\Gamma = e_\Gamma$ (in Γ). On the other hand $e_\Gamma = e_\Gamma * e$ (in G). Combining $e_\Gamma * e_\Gamma = e_\Gamma * e$. It then follows from Proposition 14 that $e = e_\Gamma$ (in G).

(G3) Every element of Γ has an inverse. By the uniqueness, this inverse is the inverse we get when we think of it as an element of G. The content of what this assertion says if that if γ is an element of Γ, then the inverse γ^{-1} (in G) indeed lies in Γ.

Examples. • $\{e\}$ and G (itself) are subgroups of G.
• $(\mathbb{Z}, +) \subset (\mathbb{Q}, +) \subset (\mathbb{R}, +) \subset (\mathbb{C}, +)$.
• Fix a positive integer n. Then $n\mathbb{Z}$ is a subgroup of $\mathbb{Z}$.

Remark. Not every subset is a subgroup. For example, $\mathbb{Z}_6$ has 2^6 distinct subsets but the subgroups are:

$$\{[0]\}, \mathbb{Z}_6, \{[0], [2], [4]\}, \{[0], [3]\}.$$

They are subgroups of order $1, 12, 3$ and 2. Similarly, $\mathbb{Z}_{12}$ has 2^{12} subsets but the subgroups are:

$$\{[0]\}, \mathbb{Z}_{12}, \{[0], [2], [4], [6], [8], [10]\}, \{[0], [3], [6], [9]\}, \{[0], [4], [8]\}, \{[0], [6]\}$$

They are subgroups of order $1, 12, 6, 4, 3$ and 2.

How do we work this out? Let's do this in the case of $\mathbb{Z}_6$. Any subgroup needs to have the identity element (G1). So $[0]$ needs to be a member of any subgroup we seek. Since $\{[0]\}$ is the only subgroup of order 1, let's assume a subgroup Γ we seek is of order > 1 and has an element not equal to $[0]$.

• If $[1]$, then $[1] + [1] = [2]$ needs to be in Γ by (G0). Similarly $[1] + [1] + [1] = [3]$ needs to be in Γ. Repeating this, we see that $\Gamma = \{[0], [1], \ldots, [5]\}$.

• If $[1]$ is not in Γ but $[2]$ is, then so are $[2] + [2] = [4], [2] + [2] + [2] = [6] = [0], \ldots$, so $\Gamma = \{[0], [2], [4]\}$.

• If neither $[1]$ nor $[2]$ is in Γ but $[3]$ is, then repeated application of (G0) shows $\Gamma = \{[0], [3]\}$.

• If we start with $[0]$ and $[4]$, we get $[4] + [4] = [8] = [2], [4] + [4] + [4] = [12] = [0], \ldots$. Hence $\Gamma = \{[0], [2], [4]\}$.

• If we start with $[0]$ and $[5]$, we get $[5] + [5] = [10] = [4], [5] + [5] + [5] = [15] = [3], \ldots$ and $\Gamma = \mathbb{Z}_6$.

It looks like every subgroup of $\mathbb{Z}_n$ has order that is a divisor of n (and in fact all divisors of n show up!). Is this always the case? Indeed, if r is a divisor of n and $rs = n$, then s is an integer and

$$\{[0], [s], [2s], \ldots, [(r-1)s]\}$$

is a subgroup of $\mathbb{Z}_n$ of order r. Indeed, for integers $0 \leq a, b \leq r - 1$, we see that $[as] + [bs] = [(a+b)s] = [cs]$ where we may, and will, choose c to be the integer $0 \leq c \leq r-1$ that is congruent to $a + b$ modulo r. Use this to find the inverse of $[as]$ for $0 \leq a \leq r - 1$.

Proposition 43. A non-empty subset Γ of a group $(G, *)$ is a subgroup if and only if, for every g, γ in Γ, $g * \gamma^{-1}$ is in Γ.

Proof. Suppose that Γ be a subgroup of G. In this case, if g and γ are elements of Γ, so is g^{-1}. Since Γ is closed with respect to $*$ (restricted to Γ), $g * \gamma^{-1}$ is an element of Γ.

Conversely, suppose the condition holds. Since Γ is assumed to be non-empty, there exists an element g in Γ; and it follows from the condition that $g * g^{-1}$, i.e. e lies in Γ (G2).

Letting $(g, \gamma) = (e, \gamma)$, it also follows from the condition that $e * \gamma^{-1} = \gamma^{-1}$ lies in Γ (G3).

Let g and γ be elements of Γ. By (G3) at our disposal, γ^{-1} lies in Γ. It then follows from the condition that $g * (\gamma^{-1})^{-1} = g * \gamma$ lies in Γ (G0).

Finally, (G1) follows for free. □

Theorem 44 (Lagrange's theorem). Let G be a finite group and Γ be a subgroup. Then $|\Gamma|$ divides $|G|$.

Proof. Non-examinable.

9 Topics I could have covered but did not, hence non-examinable, in 2023-2024

Let Γ be a subgroup of $(G, *)$. Define an equivalence relation $\mathcal{R}$ on G by $g \mathcal{R} \gamma$ if $g * \gamma^{-1}$ is an element of Γ. This defines an equivalence relation.

Definition. Let $g\Gamma$ denote the equivalence class represented by g. As a set, this is the subset of elements in G of the form $g\gamma$ for some γ in H, and is often referred to as a (left) H-coset.

Let G/Γ denote the set of all equivalence classes $\{g\Gamma\}$ (as g ranges over G).

If we are to define a group structure on G/Γ, the most sensible thing is to define $*_{G/\Gamma}$ on G/Γ as

$$(g\Gamma) *_{G/\Gamma} (\gamma\Gamma) = (g * \gamma)\Gamma$$

for all g and γ in G. However, this is not well-defined. Indeed, let G be the group $\mathrm{Sym}(\{1, 2, 3\}) = S_3 = \{e, r, (r \circ r), s, r \circ s\}$ where $r = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$ and $s = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$. Let Γ be the subgroup $\{1, s\}$. Then

$$r\Gamma = \{r, r \circ s\} = (r \circ s)\Gamma$$

and

$$(r \circ r)\Gamma = \{(r \circ r), (r \circ r) \circ s\} = (r \circ r \circ s)\Gamma.$$

On one hand,

$$(r\Gamma) *_{G/\Gamma} ((r \circ r)\Gamma) = (r \circ r \circ r)\Gamma = \Gamma$$

on the other hand,

$$((r \circ s)\Gamma) *_{G/\Gamma} ((r \circ r \circ s)\Gamma) = (r \circ s \circ r \circ r \circ s)\Gamma = r^{-1}\Gamma$$

since $r \circ s \circ r \circ r \circ s = r \circ s \circ r \circ e \circ r \circ r \circ s = r \circ s \circ r \circ (s \circ s) \circ r \circ r \circ s = s \circ r \circ s = r^{-1}$ (this follows from $(r \circ s) \circ (r \circ s) = e = (s \circ r) \circ (s \circ r)$ and $s \circ s = e$). In other others, $*_{G/H}$ defined

above depends on representatives and is well-defined!

To circumvent the issue, instead of seeking to redefine $*_{G/\Gamma}$, we qualify Γ further to a *normal* subgroup.

Definition. We say that Γ is a normal subgroup if $g\Gamma g^{-1} = \Gamma$ (or equivalently $g\Gamma = \Gamma g$) holds for every g in G. Note that this is an equality of sets; it does not demand that $g\gamma = \gamma g$ holds for every γ in Γ.

Example. Let n be a non-negative integer. Then $n\mathbb{Z}$ is a normal subgroup of $\mathbb{Z}$.

Definition. If Γ is a normal subgroup, we define $*_{G/\Gamma}$ on G/H by

$$(g\Gamma) *_{G/\Gamma} (\gamma\Gamma) = (g * \gamma)\Gamma.$$

Proposition 45. $(G/H, *_{G/H})$ is a group.

Remark. If G is finite, Lagrange's theorem proves that $|G/H| = |G|/|H|$.

Definition. Let $(G, *_G)$ and $(\Gamma, *_\Gamma)$ be groups. A function $f : G \to \Gamma$ which satisfies the condition

$$f(g *_G \gamma) = f(g) *_\Gamma f(\gamma)$$

is called a group homomorphism.

Let F be a field (e.g. $\mathbb{R}$) and $\mathrm{GL}_2(F)$ be the set of 2-by-2 matrices with non-zero determinant. Then $\mathrm{GL}_2(F)$ is a group under multiplication with identity $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. A group homomorphism

$$\rho : (G, *) \to (\mathrm{GL}_2(F), \times)$$

is called a (two-dimensional) representation of G over F– this 'represents' the (group) structure of G in terms of (2-by-2) matrices.

Representation theory is where group theory and linear algebra meet and is one of the fundamental tools in mathematics and even in physics.