MAS220 ALGEBRA, SEMESTER 1

Dr. Neil Dummigan

University of Sheffield

Contents

1. GROUPS

1.1. GROUPS.

Definition 1.1.1. *A non-empty set G is a* **group** *under $\odot$ if*

G1 *(Closure) $\odot : G \times G \to G$, i.e. $\odot$ is a binary operation on G;*

G2 *(Associativity) $(a \odot b) \odot c = a \odot (b \odot c)$, $\forall\, a, b, c \in G$;*

G3 *(Neutral element) $\exists e \in G$ such that, $\forall g \in G$,*

$$e \odot g = g \odot e = g;$$

G4 *(Inverses) $\forall g \in G\ \exists h \in G$ such that*

$$g \odot h = h \odot g = e.$$

h is called the inverse of g (usually denoted g^{-1}).

When the operation is understood, we often omit it, i.e. we write ab instead of $a \odot b$, except where the operation has another name we might use, especially if we need to avoid confusion with multiplication, e.g. when the operation is addition of integers, we write $a + b$ rather than ab, and $-a$ rather than a^{-1}.

Definition 1.1.2. *G (i.e. $(G, \odot)$) is* **abelian** *if $ab = ba$ $\forall a, b \in G$.*

Example 1.1.3. *$G = \{R, Y, B\}$ with operation given by the table*

$\odot$	R	Y	B
R	R	Y	B
Y	Y	B	R
B	B	R	Y

G1 is easy to see. For G3, R is a neutral element, and for G4, R is self-inverse while B and Y are inverse to each other. But to establish G2 would appear to require us to check all $3^3 = 27$ possibilities for (a, b, c). In fact G2 does hold, but it turns out we can avoid all this checking (see Section 1.3 later).

Example 1.1.4. *$(\mathbb{Z}, +)$ is an abelian group. The neutral (identity) element is 0, and the inverse of any $a \in \mathbb{Z}$ is $-a$.*

Example 1.1.5. *Similarly, $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$ and $(\mathbb{C}, +)$ are all abelian groups.*

Example 1.1.6. *$\mathbb{R}$ is not a group under multiplication because 0 does not have a multiplicative inverse. But if we let $\mathbb{R}^\times := \mathbb{R} - \{0\}$ then this is a group under multiplication. Similarly if $\mathbb{Q}^\times := \mathbb{Q} - \{0\}$ and $\mathbb{C}^\times := \mathbb{C} - \{0\}$ then these are (abelian) groups under multiplication.*

Example 1.1.7. *$\mathbb{Z} - \{0\}$ is* not *a group under multiplication. The only integers having multiplicative inverses* also in $\mathbb{Z}$*, rather than just in $\mathbb{Q}$, are 1 and -1. So $\{\pm 1\}$ is a group under multiplication.*

Example 1.1.8. *Let $\mathrm{GL}_2(\mathbb{R})$ be the set of all invertible 2-by-2 matrices with real entries, i.e.*

$$\mathrm{GL}_2(\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in \mathbb{R},\ ad - bc \neq 0 \right\}.$$

This is a group (the general linear group*) under the operation of matrix multiplication. The neutral element is the identity matrix $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, and if $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then $A^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$. (See Section 2.6 below for a proof of the associativity of matrix multiplication.) This is not an abelian group. For an example of invertible 2-by-2 matrices that do not commute, again see Section 2.6.*

Example 1.1.9. *The set S_n of all bijections $\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ (i.e. permutations of $\{1, 2, \ldots, n\}$) is a group under the operation of composition of functions:*

$$\sigma\tau(x) := \sigma(\tau(x)) \ \ \forall \sigma, \tau \in S_n, x \in \{1, \ldots, n\}.$$

Composition of functions is always associative (Proposition 1.8 of MAS114).
The identity element is the identity function id *(such that* $\mathrm{id}(x) = x \quad \forall x \in \{1, \ldots, n\}$*).*
The inverse of a permutation σ is its inverse function σ^{-1}, which exists because σ is a bijection.
For $n \geq 3$, this is not an abelian group, e.g. in S_3, $(12)(13) = (132)$, while $(13)(12) = (123)$.

Example 1.1.10. *The set O_2 of all rotations centred at the origin and reflections in axes through the origin, in the plane $\mathbb{R}^2$, is a group.*

$$O_2 = \{\mathrm{rot}_\alpha, \mathrm{ref}_\beta : \alpha, \beta \in \mathbb{R}\}.$$

What is the operation? A rotation or reflection may be thought of as a special way of taking any point of the plane and mapping it to another point of the plane, i.e. as a function from $\mathbb{R}^2$ to $\mathbb{R}^2$. Doing one rotation or reflection after the other is again composition of functions.
The identity function $\mathrm{id} : \mathbb{R}^2 \to \mathbb{R}^2$ *is* rot_0*, so belongs to O_2.*
Each rotation or reflection is a bijection, so is invertible as a function from $\mathbb{R}^2$ to $\mathbb{R}^2$, and moreover the inverse function is also in O_2:
$\mathrm{rot}_\alpha^{-1} = \mathrm{rot}_{-\alpha}$ *and* $\mathrm{ref}_\beta^{-1} = \mathrm{ref}_\beta$.
One may show (e.g. in MAS114) that

$$\mathrm{rot}_\alpha \mathrm{rot}_\beta = \mathrm{rot}_{\alpha+\beta};$$

$$\mathrm{ref}_\alpha \mathrm{ref}_\beta = \mathrm{rot}_{\alpha-\beta};$$

$$\mathrm{rot}_\alpha \mathrm{ref}_\beta = \mathrm{ref}_{\alpha+\beta};$$

$$\mathrm{ref}_\alpha \mathrm{rot}_\beta = \mathrm{ref}_{\alpha-\beta}$$

(which incidentally demonstrates closure). Note that the second line, or the last two lines, shows that O_2 is not abelian. For example,

$$\mathrm{ref}_0 \mathrm{ref}_{\pi/2} = \mathrm{rot}_{-\pi/2} \neq \mathrm{rot}_{\pi/2} = \mathrm{ref}_{\pi/2} \mathrm{ref}_0.$$

1.2. **SUBGROUPS.**

Definition 1.2.1. *Let G be a group. A subset H of G is a* **subgroup** *(written $H \leq G$) if H is itself a group, using the same binary operation as G.*

This is if and only if (Subgroup Criterion, Thm. 3.19 in MAS114)
SG1 $H \neq \emptyset$;
SG2 $gh \in H \ \ \forall g, h \in H$ (closure);
SG3 $h^{-1} \in H \ \ \forall h \in H$.

We usually check SG1 by showing that $e \in H$, where e is the neutral element.

Example 1.2.2. *For any fixed $m \in \mathbb{Z}$, the subset $m\mathbb{Z} := \{mz : z \in \mathbb{Z}\}$, i.e. the set of multiples of m, is a subgroup of $\mathbb{Z}$. To check this,*
SG1 $0 = m0 \in m\mathbb{Z}$*;*
SG2 *Given $mz_1, mz_2 \in m\mathbb{Z}$, $mz_1 + mz_2 = m(z_1 + z_2) \in m\mathbb{Z}$;*
SG3 *Given $mz \in m\mathbb{Z}$, $-(mz) = m(-z) \in m\mathbb{Z}$.*

In fact, one can show that these are all *the subgroups of* $\mathbb{Z}$*. For example, if we take* $m = 2$ *then we get the subgroup of even integers,*

$$2\mathbb{Z} = \{\ldots, -4, -2, 0, 2, 4, 6, \ldots\}.$$

Example 1.2.3. $\mathbb{Z} \leq \mathbb{Q} \leq \mathbb{R} \leq \mathbb{C}$ *(additive groups) and* $\{\pm 1\} \leq \mathbb{Q}^\times \leq \mathbb{R}^\times \leq \mathbb{C}^\times$ *(multiplicative groups). These were all known groups already, so it is unnecessary to use the subgroup criterion (which is about showing that a given subset is a group). We are just noticing that some are inside others.*

Example 1.2.4. *Inside the group* $\mathrm{GL}_2(\mathbb{R})$ *(cf. Example 1.1.8), we can find various subgroups, such as the* special linear group

$$\mathrm{SL}_2(\mathbb{R}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in \mathbb{R},\ ad - bc = 1 \right\}.$$

SG1 $\det(I) = 1$*, so* $I \in \mathrm{SL}_2(\mathbb{R})$*;*

SG2 *Given* $A, B \in \mathrm{SL}_2(\mathbb{R})$*,* $\det(AB) = \det(A)\det(B) = 1 \times 1 = 1$*, so* $AB \in \mathrm{SL}_2(\mathbb{R})$*;*

SG3 *Given* $A \in \mathrm{SL}_2(\mathbb{R})$*,* $\det(A^{-1}) = (\det(A))^{-1} = 1^{-1} = 1$*, so* $A^{-1} \in \mathrm{SL}_2(\mathbb{R})$*.*

Example 1.2.5. *Inside* S_n *we have a subgroup* A_n *of all even permutations (those that are products of even numbers of transpositions). You can check the subgroup criterion if you like.*

Example 1.2.6. *Inside* O_2 *we have a subgroup* $\mathrm{SO}_2 := \{\mathrm{rot}_\alpha : \alpha \in \mathbb{R}\}$ *comprising just the rotations, no reflections. Let's show that this one is a subgroup.*

SG1 $\mathrm{id} = \mathrm{rot}_0 \in \mathrm{SO}_2$*;*

SG2 *Given any* $\mathrm{rot}_\alpha, \mathrm{rot}_\beta \in \mathrm{SO}_2$*,* $\mathrm{rot}_\alpha \mathrm{rot}_\beta = \mathrm{rot}_{\alpha+\beta} \in \mathrm{SO}_2$*;*

SG3 *Given any* $\mathrm{rot}_\alpha \in \mathrm{SO}_2$*,* $(\mathrm{rot}_\alpha)^{-1} = \mathrm{rot}_{-\alpha} \in \mathrm{SO}_2$*.*

1.3. **ISOMORPHISM.** .

There is an efficiency in enunciating a list of axioms to define what it means to be a group (or a ring or a vector space, for that matter). It means that anything we deduce purely from the axioms applies equally to all examples, and does not need to be proved separately for each one. But a deeper reason for defining the concept of a group is that it allows us to see previously invisible connections between different structures in mathematics. That $\mathbb{Z}$, S_n and O_2 are all groups is something non-obvious that they have in common, which we could not express without having defined what a group is. There is another level to this recognition of commonality. Sometimes two apparently different structures are not only both groups, but essentially *the same* group. Such groups are said to be *isomorphic*.

Definition 1.3.1. *Let G, H be groups. G is* **isomorphic** *to H (written $G \simeq H$) if there exists an* **isomorphism** *$\theta : G \to H$, i.e. a bijection such that $\theta(ab) = \theta(a)\theta(b)$.*

One way to think of this (especially when G and H are finite) is that we see the same pattern in the Cayley table, just with the elements labelled differently. In other words, if we take a Cayley table for G and replace throughout the table each element $g \in G$ by the corresponding $\theta(g) \in H$ then we obtain the Cayley table for H. Down the left hand edge, in a certain row, a is replaced by $\theta(a)$. Along the top, in a certain column, b is replaced by $\theta(b)$. Where that row and column intersect, the entry ab in the Cayley table of G is replaced by $\theta(ab)$, which is correct for the Cayley table of H, because it is the same as $\theta(a)\theta(b)$.

Example 1.3.2. *$D_3 = \{\mathrm{id}, \mathrm{rot}_{2\pi/3}, \mathrm{rot}_{4\pi/3}, \mathrm{ref}_0, \mathrm{ref}_{2\pi/3}, \mathrm{ref}_{4\pi/3}\}$, the group of symmetries of an equilateral triangle with vertices at $(1,0), (-1/2, \sqrt{3}/2), (-1/2, -\sqrt{3}/2)$ (labelled 1,2,3 respectively), is evidently a subgroup of O_2. By looking at how it permutes the vertices of the triangle, we see that it is also isomorphic to S_3. (You can fill in the picture if you like.) The bijection $\theta : D_3 \to S_3$ is as follows:*

g	$\theta(g)$
$\mathrm{id}_{\mathbb{R}^2}$	$\mathrm{id}_{\{1,2,3\}}$
$\mathrm{rot}_{2\pi/3}$	$(1\,2\,3)$
$\mathrm{rot}_{4\pi/3}$	$(1\,3\,2)$
ref_0	$(2\,3)$
$\mathrm{ref}_{2\pi/3}$	$(1\,2)$
$\mathrm{ref}_{4\pi/3}$	$(1\,3)$

Doing one rotation or reflection in D_3 followed by another corresponds to doing one permutation of the vertices followed by another. In fact given $g \in D_3$, viewed as a map $g : \mathbb{R}^2 \to \mathbb{R}^2$, $\theta(g)$ is just its restriction to the subset $\{1,2,3\}$ (labels) of the domain, whose image is again $\{1,2,3\}$. Hence $\theta(gh) = \theta(g)\theta(h)$ $\forall g, h \in D_3$. (On the left, gh is a composition of rotations or reflections. On the right, $\theta(g)\theta(h)$ is a composition of associated permutations.)

In the above example we saw that S_3 is isomorphic to a subgroup of O_2. This kind of situation, where we see one familiar group not actually equal to, but rather isomorphic to, a subgroup of another, is very common. Let us now return to Example 1.1.3, where we had a table describing a binary operation on the set $G = \{R, Y, B\}$. We do not yet know that this is a group. Now let $\omega = e^{2\pi i/3} = \cos(2\pi/3) + i\sin(2\pi/3) = (-1/2) + (\sqrt{3}/2)i$, and consider

the set $U_3 := \{1, \omega, \omega^2\}$ of cube roots of 1 in $\mathbb{C}$. Using $\omega^3 = 1$ it is easy to see that this is a subgroup of $\mathbb{C}^\times$. The Cayley table for U_3 is

$\times$	1	ω	ω^2
1	1	ω	ω^2
ω	ω	ω^2	1
ω^2	ω^2	1	ω

Under the bijection $1 \mapsto R$, $\omega \mapsto Y$, $\omega^2 \mapsto B$, this turns into the table

$\odot$	R	Y	B
R	R	Y	B
Y	Y	B	R
B	B	R	Y

describing the binary operation on G. It follows that $(G, \odot)$ *is* a group, isomorphic to the subgroup U_3 of the multiplicative group $\mathbb{C}^\times$ (with identity R and $Y^{-1} = B$). Since multiplication in $\mathbb{C}^\times$ is associative, we now get the associativity of $\odot$ for free, without having to check 27 instances of $(a \odot b) \odot c = a \odot (b \odot c)$.

Example 1.3.3. *Looking back at Example 1.3.2, more generally, for any $n \geq 3$ we have a subgroup D_n of O_2 comprising the symmetries of a regular n-gon, and looking at the permutations of the vertices again gives us a map from D_n to S_n, turning composition of symmetries into composition of permutations. This is however not an isomorphism once $n > 3$, because it is not a bijection. Although it is always injective (different symmetries can be distinguished by the permutations they induce), it is not surjective. For example, when $n = 4$, no rotation or reflection will swap a single pair of adjacent vertices. In fact $|D_n| = 2n$ while $S_n = n!$, which is bigger for $n > 3$, and sets of different sizes cannot possibly be in bijection. What we do get is D_n isomorphic not to the whole of S_n, but to a* subgroup *of S_n.*

Example 1.3.4. *For any group G,* $\mathrm{id} : G \to G$ *is an isomorphism of G with itself. (It is certainly a bijection, and* $\mathrm{id}(ab) = \mathrm{id}(a)\mathrm{id}(b)$ *is simply $ab = ab$, since* $\mathrm{id}(a) = a \ \ \forall a \in G$*.) But it might not be the only one. For example, ω^2 is another generator of the cyclic group U_3 on the previous page. In other words, U_3, which is the set of powers of ω, is also the set of powers of ω^2. The map sending $1 \mapsto 1$, $\omega \mapsto \omega^2$, $\omega^2 \mapsto (\omega^2)^2 = \omega$ is an isomorphism from U_3 to itself,*

different from id. *In other words, exchange* ω *and* ω^2 *throughout the above table, and it is still correct.*

Notice that $\omega^2 = e^{4\pi i/3} = (-1/2) - (\sqrt{3}/2)i = \overline{\omega}$*, and this isomorphism of* U_3 *to itself is the restriction of the isomorphism* $z \mapsto \overline{z}$ *of* $\mathbb{C}^\times$ *to itself. (We know, or believe, that* $\overline{z_1 z_2} = \overline{z_1}\ \overline{z_2}$*.)*

Incidentally, take care not to confuse the identity map $\mathrm{id} : G \to G$ *with the neutral element* $e \in G$*. The neutral element of* S_n *just happens to be called* id*, being the identity map from* $\{1, 2 \ldots, n\}$ *(not* S_n*) to itself.*

Example 1.3.5. *In* SO_2 *we have* $\mathrm{rot}_\alpha \mathrm{rot}_\beta = \mathrm{rot}_{\alpha+\beta}$*. Composition of rotations turns into addition of angles. So we might suspect that the map from the additive group* $\mathbb{R}$ *to the group* SO_2 *of rotations under composition, given by* $\alpha \mapsto \mathrm{rot}_\alpha$*, is an isomorphism. But it is not an isomorphism because it is not a bijection. This is because it is not injective:* rot_α *and* $\mathrm{rot}_{\alpha+2\pi}$ *are the same.*

1.4. **OTHER WAYS OF GETTING NEW GROUPS FROM OLD.** .

The purpose of this section is to motivate the concept of quotient group, which is dealt with rigorously in the next section. We will see how from one group another group can arise in a very natural way, by throwing away a lot of information but still keeping something important. For example, recall that in O_2,

$$\mathrm{rot}_\alpha \mathrm{rot}_\beta = \mathrm{rot}_{\alpha+\beta};$$

$$\mathrm{ref}_\alpha \mathrm{ref}_\beta = \mathrm{rot}_{\alpha-\beta};$$

$$\mathrm{rot}_\alpha \mathrm{ref}_\beta = \mathrm{ref}_{\alpha+\beta};$$

$$\mathrm{ref}_\alpha \mathrm{rot}_\beta = \mathrm{ref}_{\alpha-\beta}.$$

If we ignore the angles, we see a pattern:

$\cdot$	rot	ref
rot	rot	ref
ref	ref	rot

.

This is the Cayley table of a group, with identity rot.

For another example, in the additive group $\mathbb{Z}$, if we forget everything about a number except whether it is even or odd, we get a consistent pattern:

$+$	even	odd
even	even	odd
odd	odd	even

Again this is the Cayley table of a group, with identity even. As it happens, the tables have the same pattern, and the groups are isomorphic, via rot $\leftrightarrow$ even and ref $\leftrightarrow$ odd. In fact, both are isomorphic to the multiplicative group $\{\pm 1\}$:

$\cdot$	1	-1
1	1	-1
-1	-1	1

.

From the big group O_2 we have obtained the smaller group $\{\text{rot}, \text{ref}\}$, and from the big group $\mathbb{Z}$ we have obtained the smaller group $\{\text{even}, \text{odd}\}$. But these are not subgroups. So what are they? In fact, what kinds of objects are these elements rot, ref, even and odd that inhabit these smaller groups?

In the first case, we have classified elements of O_2 into two types, namely reflections and rotations, i.e. we have partitioned the set O_2 into two subsets (equivalence classes), one, which we could call "rot", containing all the rotations, and another, which we could call "ref", containing all the reflections. The group table works because if you combine two elements and all you want to know is whether the result is a rotation or a reflection, all you need to know is the same information about the elements you combined, not what the particular angles are.

Similarly, in the second case, we have partitioned the set $\mathbb{Z}$ into two equivalence classes, one called "even", containing all the even integers, the other called "odd", containing all the odd integers. If you add two integers and all you want to know is whether the result is odd or even, that is all you need to know about the integers being added, not which particular odd number or which particular even number.

Now we attempt in a similar way to get a smaller group from S_3. Here is the Cayley table for S_3.

$\cdot$	id	(123)	(132)	(12)	(13)	(23)
id	id	(123)	(132)	(12)	(13)	(23)
(123)	(123)	(132)	id	(13)	(23)	(12)
(132)	(132)	id	(123)	(23)	(12)	(13)
(12)	(12)	(23)	(13)	id	(132)	(123)
(13)	(13)	(12)	(23)	(123)	id	(132)
(23)	(23)	(13)	(12)	(132)	(123)	id

We can partition S_3 into two subsets $E = \{\mathrm{id}, (123), (132)\}$ and $O = \{(12), (13), (23)\}$. This corresponds to drawing dividing lines in the table as follows.

$\cdot$	id	(123)	(132)	(12)	(13)	(23)
id	id	(123)	(132)	(12)	(13)	(23)
(123)	(123)	(132)	id	(13)	(23)	(12)
(132)	(132)	id	(123)	(23)	(12)	(13)
(12)	(12)	(23)	(13)	id	(132)	(123)
(13)	(13)	(12)	(23)	(123)	id	(132)
(23)	(23)	(13)	(12)	(132)	(123)	id

Now we observe a remarkable fact, akin to what happened with the partitions $O_2 = \mathrm{rot} \cup \mathrm{ref}$ and $\mathbb{Z} = \mathrm{even} \cup \mathrm{odd}$. If we combine two elements and want to know whether the result is in class E or in class O, we only need to know the classes of what we are combining, not the particular elements. The pattern is

$\cdot$	E	O
E	E	O
O	O	E

For example, if we look at the product $OE := \{ab : a \in O, b \in E\}$, there are 3 choices for a and 3 choices for b, so $3 \times 3 = 9$ products to look at, but we do not get 9 different elements, only the same 3 elements (consistently in O), 3 times each.

This kind of thing does not have to happen if we partition the group a different way. For example, consider the partition $\{A, B, C\}$ of S_3, where $A = \{\mathrm{id}, (12)\}$, $B = \{(13), (123)\}$ and $C = \{(23), (132)\}$. We need to write the elements in the group table in a different order before we can draw lines.

·	id	(12)	(13)	(123)	(23)	(132)
id	id	(12)	(13)	(123)	(23)	(132)
(12)	(12)	id	(132)	(23)	(123)	(13)
(13)	(13)	(123)	id	(12)	(132)	(23)
(123)	(123)	(13)	(23)	(132)	(12)	id
(23)	(23)	(132)	(123)	(13)	id	(12)
(132)	(132)	(23)	(12)	id	(13)	(123)

We have failed in our attempt even to define a binary operation on $\{A, B, C\}$ using the original operation on S_3. Although $AA = A$, $BA = B$ and $CA = C$, we find that $AB = B \cup C$, not a single class, and similarly, for example, $BB = A \cup C$. So the class of the output in general depends on more than just the classes of the inputs. In the next section we describe how to produce a partition of a group that does successfully produce a "quotient" group as in the examples $\{\text{rot}, \text{ref}\}$, $\{\text{even}, \text{odd}\}$ and $\{E, O\}$.

1.5. **COSETS, NORMAL SUBGROUPS AND QUOTIENT GROUPS.** .

Let G be a group, H a subgroup of G. We can partition G into **left cosets** of H, each of the form $gH = \{gh : h \in H\}$ for $g \in G$. In particular, letting $g = e$, H itself is one of these left cosets. Since $e \in H$, $g \in gH$, i.e. gH is the left coset containing g. A given left coset C can be written in many different ways, as gH for *any* $g \in C$. The set of left cosets of H in G is written G/H.

Example 1.5.1. $G = \mathbb{Z}$ *(necessarily under addition),* $H = 2\mathbb{Z}$*,* $G/H = \{2\mathbb{Z}, 1 + 2\mathbb{Z}\} = \{even, odd\}$*.*

Example 1.5.2. $G = O_2, H = \text{SO}_2$*,* $G/H = \{H, \text{ref}_0 H\} = \{\text{rot}, \text{ref}\}$*. To see why* $\text{ref}_0 H = \text{ref}$*, given any* $\text{rot}_\beta \in H$ *we find* $\text{ref}_0 \text{rot}_\beta = \text{ref}_{-\beta}$ *(putting* $\alpha = 0$ *in the last of the four formulas in Example 1.1.10), and by varying* β *we can make this any reflection we choose.*

Example 1.5.3. $G = S_3, H = A_3 = \{\text{id}, (1\,2\,3), (1\,3\,2)\}$*,*

$$G/H = \{A_3, (12)A_3\} = \{\{\text{id}, (1\,2\,3), (1\,3\,2)\}, \{(1\,2), (2\,3), (1\,3)\}\} = \{E, O\} (from Section 1.4).$$

Example 1.5.4. $G = S_3, H = \langle (1\,2) \rangle = \{\text{id}, (1\,2)\}$*,*

$$G/H = \{H, (1\,3)H, (2\,3)H\} = \{\{\text{id}, (1\,2)\}, \{(1\,3), (1\,2\,3)\}, \{(2\,3), (1\,3\,2)\}\} = \{A, B, C\} (from Section 1.4).$$

Thus, all the partitions of groups that we considered in the previous section were partitions into left cosets. In the first three cases we successfully turned the set G/H into a group, using

the group operation on G but failing to distinguish between equivalent elements, i.e. elements belonging to the same equivalence class, to the same left coset of H. In each of these three cases, the identity element was the coset H, namely even, rot or E. But in the fourth case the group operation did not "descend" to G/H. What made the difference?

We could also have considered **right cosets** $Hg = \{hg : h \in H\}$, and the set $H\backslash G$ of right cosets of H in G.

Example 1.5.5. $G = \mathbb{Z}, H = 2\mathbb{Z}$, $H\backslash G = \{2\mathbb{Z}, 2\mathbb{Z}+1\} = \{even, odd\}$.

Example 1.5.6. $G = O_2, H = \mathrm{SO}_2$, $H\backslash G = \{H, H\mathrm{ref}_0\} = \{\mathrm{rot}, \mathrm{ref}\}$.

Example 1.5.7. $G = S_3, H = A_3 = \{\mathrm{id}, (1\,2\,3), (1\,3\,2)\}$,

$$H\backslash G = \{A_3, A_3(12)\} = \{\{\mathrm{id}, (1\,2\,3), (1\,3\,2)\}, \{(1\,2), (1\,3), (2\,3)\}\} = \{E, O\}.$$

Example 1.5.8. $G = S_3, H = \langle(1\,2)\rangle = \{\mathrm{id}, (1\,2)\}$,

$$H\backslash G = \{H, H(1\,3), H(2\,3)\} = \{\{\mathrm{id}, (1\,2)\}, \{(1\,3), (1\,3\,2)\}, \{(2\,3), (1\,2\,3)\}\}.$$

In the first three examples, left and right cosets are the same, whereas in the fourth they are different. (See the first four examples of this section for the left cosets.)

Definition 1.5.9. *Let G be a group. A subgroup N of G is said to be* **normal** *(written $N \lhd G$) if $gN = Ng \ \ \forall g \in G$.*

We have just seen that $2\mathbb{Z} \lhd \mathbb{Z}$, $\mathrm{SO}_2 \lhd O_2$ and $A_3 \lhd S_3$, but $\langle(1\,2)\rangle \not\lhd S_3$. The example $2\mathbb{Z} \lhd \mathbb{Z}$ illustrates the following.

Proposition 1.5.10. *If G is abelian,* any *subgroup N of G is automatically normal.*

Proof. $gN = \{gn \mid \ \ n \in N\} = \{ng \mid \ \ n \in N\} = Ng$. □

Beware that it is not enough for N to be abelian. For example, the abelian subgroup $\{\mathrm{id}, (12)\}$ of S_3 is not normal.

The next proposition shows how the condition $gN = Ng$ (for all $g \in G$) allows us to multiply two cosets of N together to get another coset rather than something else.

Proposition 1.5.11. *If $N \lhd G$ and $a, b \in G$ then $aNbN = abN$.*

Proof. $aNbN = a(Nb)N = a(bN)N$ (because N is normal) $= ab(NN) = abN$ (since N is a subgroup). □

This gives us a well-defined operation on the set G/N of cosets of N in G.

Theorem 1.5.12. *With this operation, G/N is a group*

Proof. The proposition gives us G1 (closure). For G2 (associativity),

$$[(aN)(bN)]cN = [abN]cN = ((ab)c)N = (a(bc))N = aN[bcN] = aN[(bN)(cN)].$$

For G3, N is the neutral element: $N(aN) = (aN)N = aN \ \ \forall aN \in G/N$. For G4, $(aN)(a^{-1}N) = (a^{-1}N)(aN) = N$, so $a^{-1}N$ is inverse to aN. □

G/N is called the **quotient group** of G by N.

Example 1.5.13. *$\mathbb{Z}$ is abelian, so $m\mathbb{Z} \triangleleft \mathbb{Z}$.*

$$\mathbb{Z}/m\mathbb{Z} = \{m\mathbb{Z}, 1 + m\mathbb{Z}, 2 + m\mathbb{Z}, \ldots, (m-1) + m\mathbb{Z}\} = \{\overline{0}, \overline{1}, \overline{2}, \ldots, \overline{m-1}\}.$$

For example, with $m = 3$ the Cayley table is

$+$	$\overline{0}$	$\overline{1}$	$\overline{2}$
$\overline{0}$	$\overline{0}$	$\overline{1}$	$\overline{2}$
$\overline{1}$	$\overline{1}$	$\overline{2}$	$\overline{0}$
$\overline{2}$	$\overline{2}$	$\overline{0}$	$\overline{1}$

Example 1.5.14. *The additive group $\mathbb{R}$ has a subgroup $2\pi\mathbb{Z}$, necessarily normal since $\mathbb{R}$ is abelian. We have $\mathbb{R}/2\pi\mathbb{Z} = \{\alpha + 2\pi\mathbb{Z} : \alpha \in \mathbb{R}\}$.*

$$\alpha + 2\pi\mathbb{Z} = \beta + 2\pi\mathbb{Z} \iff \alpha - \beta \in 2\pi\mathbb{Z}.$$

Recall that $\mathrm{rot}_\alpha \mathrm{rot}_\beta = \mathrm{rot}_{\alpha+\beta}$ yet $\alpha \mapsto \mathrm{rot}_\alpha$ is not an isomorphism from $\mathbb{R}$ to SO_2 (Example 1.3.5). In fact $\alpha + 2\pi\mathbb{Z} \mapsto \mathrm{rot}_\alpha$ gives an isomorphism between $\mathbb{R}/2\pi\mathbb{Z}$ and SO_2. Thus the concept of quotient group very naturally solves our earlier problem.

1.6. **CONJUGACY CLASSES.** .

From 5.1 in MAS114. **Group actions.**

Definition 1.6.1. *A group G acts on a non-empty set X if, for each $g \in G$ and each $x \in X$, there is an element $g * x \in X$ such that*

GA1*: $e * x = x \ \ \forall x \in X$;*

GA2*: $g * (h * x) = (gh) * x \ \ \forall g, h \in G, x \in X$.*

For example, S_n acts on $\{1, 2, \ldots, n\}$, O_2 acts on $\mathbb{R}^2$ and any group G acts on itself by left multiplication: $g * x = gx$.

From 5.8 in MAS114. **Orbits and stabilisers.**

Definition 1.6.2. *Suppose that a group G acts on a non-empty set X. For any $x \in X$,*

$$\mathrm{orb}(x) = \{y \in X : y = g * x \text{ for some } g \in G\} = \{g * x \mid g \in G\}.$$

$$\mathrm{stab}(x) = \{g \in G \mid g * x = x\}.$$

Facts.

(1) The orbits partition X. (MAS114 Corollary 8.2.) If there is only one orbit, the action is said to be **transitive**.
(2) For any $x \in X$, $\mathrm{stab}(x)$ is a subgroup of G. (MAS114, Theorem 5.9.)
(3) $g * x \leftrightarrow g\,\mathrm{stab}(x)$ is a bijection between $\mathrm{orb}(x)$ and $G/\mathrm{stab}(x)$ (the set of left cosets of $\mathrm{stab}(x)$ in G). Hence if G is finite then $|\mathrm{orb}(x)| = \frac{|G|}{|\mathrm{stab}(x)|}$, in particular $|\mathrm{orb}(x)|$ divides $|G|$.

For example, S_n acts transitively on $\{1, 2, \ldots, n\}$: given any $i, j \in \{1, 2, \ldots, n\}$, we can produce a $\sigma \in S_n$ such that $\sigma(i) = j$, for example the transposition (ij). But the action of O_2 on $\mathbb{R}^2$ is not transitive. The orbits are $\{(0,0)\}$ and all the concentric circles centred on $(0,0)$. For the action of S_n on $\{1, 2, \ldots, n\}$, $\mathrm{stab}(n) \simeq S_{n-1}$ (because if we know that σ fixes n, it is determined by how it permutes $\{1, 2, \ldots, n-1\}$), so

$$|\mathrm{orb}(n)| = \frac{|S_n|}{|S_{n-1}|} = \frac{n!}{(n-1)!} = n,$$

which we already knew, since $\mathrm{orb}(n) = \{1, 2, \ldots, n\}$, the whole set.

Conjugation action.

Proposition 1.6.3. *Let G be a group, and $X = G$. Define $g * x := gxg^{-1}$ $\forall g \in G, x \in X$. Then this gives an action of G on itself.*

Proof. .
GA1: $e * x = exe^{-1} = exe = x$ $\forall x \in G$.
GA2: $g * (h * x) = g(hxh^{-1})g^{-1} = (gh)x(h^{-1}g^{-1}) = (gh)x(gh)^{-1} = (gh) * x$ $\forall g, h, x \in G$. □

For this action, the orbits are called **conjugacy classes**, and $\mathrm{orb}(x)$ is also written $\mathrm{conj}_G(x)$. The stabilisers also have a special name; $\mathrm{stab}(x)$ is called the **centraliser** of x in G, denoted $\mathrm{cent}_G(x)$. So
$\mathrm{conj}_G(x) = \{gxg^{-1} \mid g \in G\}$, the set of **conjugates** of x in G, and
$\mathrm{cent}_G(x) = \{g \in G \mid gxg^{-1} = x\} = \{g \in G \mid gx = xg\}$.
(iii) becomes $|\mathrm{conj}_G(x)| = \frac{|G|}{|\mathrm{cent}_G(x)|}$, in particular $|\mathrm{conj}_G(x)|$ divides $|G|$.

Remarks

(1) In any group G, if e is the identity element, for any $g \in G$ we have $geg^{-1} = gg^{-1} = e$. So $\mathrm{conj}_G(e) = \{e\}$ and $\mathrm{cent}_G(e) = G$.
(2) In an *abelian* group G, the same applies to any element x. Since $gx = xg$, $gxg^{-1} = x$, for all $g \in G$, so $\mathrm{conj}_G(x) = \{x\}$ and $\mathrm{cent}_G(x) = G$.

Given an element x in a group G, if $x = e$ or if G is abelian then the above remark tells us the conjugacy class of x. In general the most simple minded way to find $\mathrm{conj}_G(x)$ is to work out

gxg^{-1} for every $g \in G$. This will be laborious except in small groups. But sometimes we can do better than this.

Conjugation in S_n

Suppose we start with some permutation, written as a product of disjoint cycles, e.g. $\alpha = (1963)(248)(57)$ in S_9. If we apply some permutation $\beta \in S_9$ to scramble up the entries, we get a new permutation with the "same cycle structure", e.g. if $\beta = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 9 & 6 & 7 & 5 & 3 & 4 & 1 & 8 & 2 \end{pmatrix}$ (written in tabular form), then $\gamma = (\beta(1)\beta(9)\ldots)\ldots = (9247)(658)(31)$. In the expression for α, adjacent elements in a cycle look like $(\ldots, i, \alpha(i), \ldots)$. The corresponding entries in the expression for γ are $(\ldots, \beta(i), \beta(\alpha(i)), \ldots)$. Hence $\gamma(\beta(i)) = \beta(\alpha(i))$. This is for all $i \in \{1, 2, \ldots, n\}$, so $\gamma\beta = \beta\alpha$, i.e. $\gamma = \beta\alpha\beta^{-1}$. In other words, to conjugate α by β in S_n, we express α as a product of disjoint cycles, then apply β to each entry. For any two permutations in S_n with the same cycle structure, there is some $\beta \in S_n$ that takes the entries in one to the corresponding entries in the other. We have arrived at

Theorem 1.6.4. *Two elements in S_n are conjugate to each other if and only if they have the same cycle structure.*

For example $S_3 = \{\mathrm{id}\} \cup \{(1\,2\,3), (1\,3\,2)\} \cup \{(1\,2), (1\,3), (2\,3)\}$ as a union of conjugacy classes. This illustrates the general idea that sorting elements of a group into conjugacy classes puts together elements that naturally have something in common. The following proposition about a general group points in the same direction. (Recall that the order of an element g in a finite group G is the smallest strictly positive integer n such that $g^n = e$.)

Proposition 1.6.5. *In a finite group G, if elements g_1, g_2 are conjugate then they have the same order.*

Proof. Say $g_2 = gg_1g^{-1}$. For any $n \in \mathbb{N}$,

$$g_2^n = \underbrace{(gg_1g^{-1})(gg_1g^{-1})\ldots(gg_1g^{-1})}_{n \text{ times}} = gg_1(g^{-1}g)g_1(g^{-1}g)\ldots(g^{-1}g)g_1g^{-1} = gg_1^ng^{-1}.$$

Now

$$g_2^n = e \iff gg_1^ng^{-1} = e \iff g_1^n = g^{-1}eg = g^{-1}g = e.$$

Hence g_1 and g_2 have the same order. □

NORMAL SUBGROUPS AND CONJUGACY CLASSES.

Let G be a group, N a subgroup of G. Recall that $N \triangleleft G \iff gN = Ng \ \forall g \in G$. This is $\iff gNg^{-1} = N \ \forall g \in G$. The following is a direct consequence.

Proposition 1.6.6. *A subgroup N of a group G is normal $\iff$ N is a union of conjugacy classes.*

Example 1.6.7. *(cf. Examples 1.5.7,1.5.8.) $G = S_3$ has conjugacy classes $\mathcal{C}_1 = \{\mathrm{id}\}, \mathcal{C}_2 = \{(1\,2\,3), (1\,3\,2)\}$ and $\mathcal{C}_3 = \{(1\,2), (1\,3), (2\,3)\}$. The subgroup $A_3 = \langle(1\,2\,3)\rangle = \{\mathrm{id}, (1\,2\,3), (1\,3\,2)\}$ is $\mathcal{C}_1 \cup \mathcal{C}_2$, so is normal, while the subgroup $\langle(1\,2)\rangle = \{\mathrm{id}, (1\,2)\}$ is not a union of conjugacy classes, so is not normal.*

Example 1.6.8. *In S_4, the subgroup $V_4 = \{\mathrm{id}, (1\,2)(3\,4), (1\,3)(2\,4), (1\,4)(2\,3)\}$ is normal, while the subgroup $K = \{\mathrm{id}, (1\,2\,3), (1\,3\,2)\}$ is not normal, since the conjugate $(1\,2\,4)$ of $(1\,2\,3)$ (by $(3\,4)$) is not in K, so $gKg^{-1} \neq K$, where $g = (3\,4)$.*

Example 1.6.9. *We already know that if G is abelian then any subgroup H is normal, but let's see it a different way. Since G is abelian, every element sits on its own in a one-element conjugacy class (by Remark (2) on p.14). So $H = \cup_{h\in H}\{h\}$ is a union of conjugacy classes, obtained by putting together all the one-element classes containing each of its elements, so is normal.*

1.7. THE CENTRE AND THE CLASS EQUATION (optional).

Definition 1.7.1. *Let G be a group. The* **centre** *$Z(G)$ of G is*

$$Z(G) := \{h \in G \mid hg = gh \ \forall g \in G\}.$$

Example 1.7.2. $Z(\mathrm{GL}_2(\mathbb{R})) = \left\{ \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} : \lambda \in \mathbb{R}^\times \right\} \simeq \mathbb{R}^\times$. *(It is easy to see that all such non-zero scalar matrices belong to the centre, but not so easy to see that there is nothing else.)*

Proposition 1.7.3. $Z(G) \lhd G$.

Proof. $Z(G) = \cap_{g\in G}\mathrm{cent}_G(g)$, an intersection of subgroups, so is a subgroup of G. (Alternatively check **SG1, SG2, SG3**.)

$$h \in Z(G) \iff gh = hg \ \forall g \in G \iff ghg^{-1} = h \ \forall g \in G \iff \{h\} \text{ is a conjugacy class.}$$

In particular, $Z(G) = \cup_{h\in Z(G)}\{h\}$ is a union of conjugacy classes, so is normal. □

Note that since $Z(G)$ is the set of those elements of G that commute with everything else, $Z(G)$ is the whole of G if and only if G is abelian.

Now let G be a finite group. It is partitioned into conjugacy classes: $G = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \ldots \cup \mathcal{C}_k$. By counting the elements of G one class at a time, we get the **class equation**:

$$|G| = |\mathcal{C}_1| + |\mathcal{C}_2| + \ldots + |\mathcal{C}_k|.$$

All the numbers on the right are divisors of $|G|$, and at least one of them is 1, since $\mathrm{conj}_G(e) = \{e\}$.

Example 1.7.4. *If $|G| = 6$, the class equation could be*
$6 = 1+1+1+1+1+1$. *This happens when G is abelian of order 6, e.g $G = \mathbb{Z}/6\mathbb{Z}$.*
$6 = 1+2+3$. *This happens for $G = S_3 = \{\mathrm{id}\} \cup \{(123),(132)\} \cup \{(12),(13),(23)\}$.*
$6 = 1+1+2+2$, *or* $6 = 1+1+1+3$ *or* $6 = 1+1+1+1+2$. *These last three "possibilities" can all be eliminated, by similar arguments. For example, if the class equation is $6 = 1+1+2+2$ then $|Z(G)| = 2$, since $Z(G)$ is the union of the 1-element conjugacy classes. If we take a in one of the 2-element conjugacy classes then $|\mathrm{conj}_G(a)| = 2 \implies |\mathrm{cent}_G(a)| = \frac{|G|}{2} = 3$. Since $Z(G) \leq \mathrm{cent}_G(a)$, Lagrange's Theorem would imply that $2 \mid 3$, which is not true, thus eliminating this case.*

Proposition 1.7.5. *If $|G| = p^r$ with p a prime number and $r \geq 1$ then $Z(G) \neq \{e\}$ (i.e. it is bigger).*

Proof. The class equation $|G| = |\mathcal{C}_1| + |\mathcal{C}_2| + \ldots + |\mathcal{C}_k|$ is

$$p^r = p^{a_1} + p^{a_2} + \ldots + p^{a_k},$$

for certain $0 \leq a_i \leq r$. Now

$$p^{a_i} \equiv \begin{cases} 1 \pmod{p} & \text{if } a_i = 0 \text{ (i.e. if } |\mathcal{C}_i| = 1); \\ 0 \pmod{p} & \text{if } a_i > 0. \end{cases}$$

The class equation mod p is
$0 \equiv \#\{\text{1-element conjugacy classes}\} \pmod{p}$, i.e.
$0 \equiv |Z(G)| \pmod{p}$. Thus $p \mid \#Z(G)$, so $Z(G) \neq \{e\}$. □

Example 1.7.6. $|D_4| = 8 = 2^3$ *and* $Z(D_4) = \{\mathrm{id}, \mathrm{rot}_\pi\}$.

Theorem 1.7.7. *Let p be a prime number. Any group G of order p^2 is abelian.*

Proof. By Proposition 1.7.5, $|Z(G)| = p$ or p^2. If $|Z(G)| = p^2$ then $Z(G) = G$, so G is abelian. If $|Z(G)| = p$ then the quotient group $G/Z(G)$ (recall that $Z(G) \triangleleft G$) has size $\frac{|G|}{|Z(G)|} = \frac{p^2}{p} = p$. Any element of $G/Z(G)$ has order 1 or p, but only the identity element has order 1, so we may choose an element $gZ(G)$ of order p, necessarily generating the group $G/Z(G)$. So $G/Z(G) = \langle gZ(G)\rangle = \{Z(G), gZ(G), \ldots, g^{p-1}Z(G)\}$. Now given any two elements $a_1, a_2 \in G$, we have $a_1 = g^{r_1}z_1, a_2 = g^{r_2}z_2$ for some $r_1, r_2 \in \{0, \ldots, p-1\}, z_1, z_2 \in Z(G)$. Then

$$a_1a_2 = g^{r_1}z_1g^{r_2}z_2 = g^{r_1}g^{r_2}z_1z_2 \text{ (since } z_1 \in Z(G)) = g^{r_1+r_2}z_1z_2,$$

while similarly (swapping 1 and 2)

$$a_2a_1 = g^{r_2+r_1}z_2z_1 = g^{r_1+r_2}z_2z_1 \ \text{(since } z_1, z_2 \in Z(G)).$$

Hence $a_1a_2 = a_2a_1$. This shows that G is abelian in this case too, which is all we need. But notice that this is even a contradiction, since if G is abelian then $Z(G) = G$ so $|Z(G)| = p^2$. Hence the case $|Z(G)| = p$ does not occur. □

Remark. You might like to try to prove the easier proposition that any group of order p is necessarily abelian. (In fact it is even cyclic.)

1.8. HOMOMORPHISMS AND THE FIRST ISOMORPHISM THEOREM (FOR GROUPS). .

Definition 1.8.1. *Let G, H be groups. A map $f : G \to H$ is a (group)* **homomorphism** *if*

$$f(ab) = f(a)f(b) \ \forall a, b \in G.$$

Special case: If f is also a bijection then f is an isomorphism.

Proposition 1.8.2. *Let $f : G \to H$ be a homomorphism.*

(1) $f(e_G) = e_H$;
(2) $f(g^{-1}) = f(g)^{-1} \ \forall g \in G$.

Proof. (1) $f(e_G) = f(e_Ge_G) = f(e_G)f(e_G)$. Combining with $f(e_G)^{-1}$, $e_H = f(e_G)$.
(2) $e_H = f(e_G) = f(gg^{-1}) = f(g)f(g^{-1})$. Hence $f(g^{-1}) = f(g)^{-1}$.

□

EXAMPLES OF HOMOMORPHISMS THAT ARE NOT ISOMORPHISMS. We provide many examples to illustrate what a natural concept group homomorphism is, how it expresses something special about maps we had already considered, which however failed to be isomorphisms.

1. Not surjective (but still injective).

Example 1.8.3. *The inclusion $i : \mathbb{Z} \hookrightarrow \mathbb{R}$ (both additive groups), which takes any integer to itself, is a homomorphism, since for any $n, m \in \mathbb{Z}$, $i(n + m) = n + m = i(n) + i(m)$. $\mathbb{Z}$ is isomorphic not to $\mathbb{R}$, but to the subgroup $i(\mathbb{Z}) = \mathbb{Z}$.*

Example 1.8.4. *$f : D_4 \to S_4$ according to permutations of the vertices of a square (cf. Examples 1.3.2, 1.3.3, in particular $D_4 = \{\mathrm{id}, \mathrm{rot}_{\pi/2}, \mathrm{rot}_{\pi}, \mathrm{rot}_{3\pi/2}, \mathrm{ref}_0, \mathrm{ref}_{\pi/2}, \mathrm{ref}_{\pi}, \mathrm{ref}_{3\pi/2}\}$, the group*

of symmetries of a square). This f is a homomorphism because when one composes two rotations or reflections $g, h \in D_4$ to get $gh \in D_4$, then looks at the corresponding permutation $f(gh) \in S_4$, this is the same as looking at the permutations $f(g), f(h) \in S_4$ corresponding to the rotations or reflections $g, h \in D_4$, and combining them in S_4 to get $f(g)f(h)$. In other words, when you compose symmetries of the square, you compose the corresponding permutations of the vertices.

g	$f(g)$
$\mathrm{id}_{\mathbb{R}^2}$	$\mathrm{id}_{\{1,2,3,4\}}$
$\mathrm{rot}_{\pi/2}$	$(1\,2\,3\,4)$
rot_{π}	$(1\,3)(2\,4)$
$\mathrm{rot}_{3\pi/2}$	$(1\,4\,3\,2)$
ref_{0}	$(1\,3)$
$\mathrm{ref}_{\pi/2}$	$(1\,4)(3\,2)$
ref_{π}	$(2\,4)$
$\mathrm{ref}_{3\pi/2}$	$(1\,2)(3\,4)$

D_4 is isomorphic not to S_4 (which has order 24), but to the 8-element subgroup $f(D_4) = \{\mathrm{id}, (1\,2\,3\,4), \ldots, (1\,2)(3\,4)\}$.

Proposition 1.8.5. *If $f : G \to H$ is a group homomorphism then the image $f(G)$ (i.e. $\mathrm{im}(f)$) is a* **subgroup** *of H.*

Proof. .

SG1 $e_H = f(e_G) \in f(G)$, so $f(G) \neq \emptyset$.

SG2 Given $f(a), f(b) \in f(G)$, $f(a)f(b) = f(ab) \in f(G)$, so $f(G)$ is closed.

SG3 Given $f(g) \in f(G)$, $f(g)^{-1} = f(g^{-1}) \in f(G)$, so $f(G)$ is "closed under inverses". □

2. Not injective (but still surjective).

Example 1.8.6. $f : S_n \to \{\pm 1\}$ *(a multiplicative group on the right)* $(n \geq 3)$,

$$f(\sigma) = \mathrm{sgn}(\sigma) = \begin{cases} +1 \text{ if } \sigma \text{ is even;} \\ -1 \text{ if } \sigma \text{ is odd.} \end{cases}$$

Example 1.8.7. $f : O_2 \to \{\pm 1\}$,

$$f(g) = \begin{cases} +1 \text{ if } g \text{ is a rotation;} \\ -1 \text{ if } g \text{ is a reflection.} \end{cases}$$

It was already noted in Section 1.4 how the pattern when combining elements of $\{\mathrm{rot}, \mathrm{ref}\}$ *mimics the pattern of multiplying elements of* $\{1, -1\}$*. Similarly the pattern for combining permutations, paying attention to whether they are even or odd, follows the addition table for* $\{\mathrm{even}, \mathrm{odd}\}$*, which, again as noted in Section 1.4, is the same pattern as multiplication of* $\{1, -1\}$*. This is why these two examples are homomorphisms.*

Example 1.8.8. $f : \mathrm{GL}_2(\mathbb{R}) \to \mathbb{R}^\times$,

$$f(A) = \det(A).$$

This is a homomorphism because $\det(AB) = \det(A)\det(B)$.

3. Neither surjective nor injective.

Example 1.8.9. $f : \mathbb{R} \to O_2$,

$$f(\alpha) = \mathrm{rot}_\alpha.$$

It is not surjective, because $f(\mathbb{R}) = \mathrm{SO}_2$*, the proper subgroup of just rotations, but surjectivity is easily restored by viewing it as a homomorphism from* $\mathbb{R}$ *to* SO_2 *instead of to* O_2*. But this is still not injective, cf. Example 1.3.5. Injectivity is more difficult to achieve, but see Example 1.5.14 for an idea of what we need to do.*

Definition 1.8.10. *Let* $f : G \to H$ *be a group homomorphism. The* **kernel** *of* f*, denoted* $\ker f$*, is*

$$\ker f := \{g \in G \mid f(g) = e_H\}.$$

Proposition 1.8.11. $\ker f$ *is a normal subgroup of* G*.*

Proof. **SG1:** $f(e_G) = e_H \implies e_G \in \ker f$.
SG2: If $a, b \in \ker f$ then $f(a) = f(b) = e_H$. Now $f(ab) = f(a)f(b) = e_H e_H = e_H$ so $ab \in \ker f$.
SG3: If $g \in \ker f$ then $f(g^{-1}) = f(g)^{-1} = e_H^{-1} = e_H$, so $g^{-1} \in \ker f$.

So far we have $\ker f \leq G$. But if $a \in \ker f$ and $g \in G$ then

$$f(gag^{-1}) = f(g)f(a)f(g^{-1}) = f(g)e_H f(g)^{-1} = e_H,$$

so $gag^{-1} \in \ker f$. So if $\ker f$ contains some element a, it also contains all the conjugates gag^{-1} of a, i.e. $\ker f$ is a union of conjugacy classes. This shows that $\ker f \lhd G$. □

As an exercise in comprehending the definition of kernel, you might like to check the following five statements, the fourth of which I have spelled out a little more than the others.

Example 1.8.12. $i : \mathbb{Z} \hookrightarrow \mathbb{R}$, $\ker i = \{0\}$. *(cf. Example 1.8.3.)*

Example 1.8.13. $f : D_4 \to S_4$, $\ker f = \{\mathrm{id}\}$ *(cf. Example 1.8.4).*

Example 1.8.14. $f : S_n \to \{\pm 1\}$, $f(\sigma) = \mathrm{sgn}(\sigma)$, $\ker f = A_n$ *(cf. Examples 1.2.5, 1.8.6).*

Example 1.8.15. $f : O_2 \to \{\pm 1\}$ *as in Example 1.8.7, then* $\ker f = \{g \in O_2 \mid \ f(g) = 1\} = \{\mathrm{rot}_\alpha \mid \ \alpha \in \mathbb{R}\} = \mathrm{SO}_2$.

Example 1.8.16. $f : \mathrm{GL}_2(\mathbb{R}) \to \mathbb{R}^\times$, $f(A) = \det(A)$, $\ker f = \mathrm{SL}_2(\mathbb{R})$ *(cf. Examples 1.2.4 and 1.8.8).*

Proposition 1.8.17. $\ker f = \{e_G\} \iff f$ *is injective.*

Proof. $f(e_G) = e_H$. If f is injective then there can't be any $g \neq e_G$ with $f(g) = f(e_G) = e_H$. So $\ker f = \{e_G\}$.

Conversely, suppose that $\ker f = \{e_G\}$. If $f(a) = f(b)$ then $f(ab^{-1}) = f(a)f(b^{-1}) = f(a)f(b)^{-1} = e_H$, so $ab^{-1} \in \ker f$. Hence $ab^{-1} = e_G$, so $a = b$. This shows that f is injective. □

If $f : G \to H$ is an injective group homomorphism then $G \simeq \mathrm{im}(f)$ (like in Examples 1.8.3 and 1.8.4). More generally

Theorem 1.8.18. *[***First Isomorphism Theorem***]* *Let $f : G \to H$ be a group homomorphism. There is an* **isomorphism**

$$\overline{f} : G/\ker f \xrightarrow{\sim} \mathrm{im}(f)$$

given by $\overline{f}(\overline{g}) = f(g)$, where $\overline{g}$ stands for the element $g \ker f$ of $G/\ker f$.

Proof. First we must show that $\overline{f}$ is well-defined, i.e that if $\overline{g} = \overline{h}$ then $f(g)$ and $f(h)$ agree. But if $\overline{g} = \overline{h}$ then $g^{-1}h \in \ker f$, so $e_H = f(g^{-1}h) = f(g^{-1})f(h) = f(g)^{-1}f(h)$, so $f(g) = f(h)$, as required. Hence $\overline{f}$ is well-defined. Now

$$\overline{f}(\overline{a})\overline{f}(\overline{b}) = f(a)f(b) = f(ab) = \overline{f}(\overline{ab}) = \overline{f}(\overline{a}\,\overline{b})$$

(note that Proposition 1.5.11 reads $\overline{a}\overline{b} = \overline{ab}$), so $\overline{f}$ is a homomorphism. It is surjective, by definition of $\mathrm{im}(f)$, so it remains to show it is injective. But if $\overline{f}(\overline{a}) = \overline{f}(\overline{b})$ then $f(a) = f(b)$ so $f(a^{-1}b) = e_H$, so $a^{-1}b \in \ker f$ and $\overline{a} = \overline{b}$. □

Example 1.8.19. $i : \mathbb{Z} \hookrightarrow \mathbb{R}$, $\ker f = \{0\}$, $\mathrm{im}(f) = \mathbb{Z}$.
FIT $\implies \mathbb{Z}/\{0\} \simeq \mathbb{Z}$. *Not surprising!*

Example 1.8.20. $f : S_n \to \{\pm 1\}$, $f(\sigma) = \mathrm{sgn}(\sigma)(cf. Example 1.8.6$, $\ker f = A_n$, $\mathrm{im}(f) = \{\pm 1\}$.
FIT $\Longrightarrow$ $S_n/A_n \simeq \{\pm 1\}$. *For the case* $n = 3$*, see Example 1.5.3, and the subdivided table in the section preceding it.*

Example 1.8.21. $f : O_2 \to \{\pm 1\}(cf. Example 1.8.7)$, $\ker f = \mathrm{SO}_2$, $\mathrm{im}(f) = \{\pm 1\}$.
FIT $\Longrightarrow$ $O_2/\mathrm{SO}_2 \simeq \{\pm 1\}$. *Looking back to the beginning of Section 1.4, this is the isomorphism between* $\{\mathrm{rot}, \mathrm{ref}\}$ *and* $\{\pm 1\}$.

Example 1.8.22. $f : \mathbb{R} \to O_2$, $f(\alpha) = \mathrm{rot}_\alpha$, $\ker f = 2\pi\mathbb{Z}$, $\mathrm{im}(f) = \mathrm{SO}_2$.
FIT $\Longrightarrow$ $\mathbb{R}/2\pi\mathbb{Z} \simeq \mathrm{SO}_2$*, via* $\alpha + 2\pi\mathbb{Z} \mapsto \mathrm{rot}_\alpha$*, which we saw already in Example 1.5.14.*

Example 1.8.23. $f : \mathbb{Z} \to \mathbb{C}^\times$, $f(n) = i^n$. *This is a homomorphism because* $i^{m+n} = i^m i^n$. *Since* $f(\mathbb{Z}) = \langle i \rangle = \{\pm 1, \pm i\}$ *and* $\ker f = 4\mathbb{Z}$*, FIT* $\Longrightarrow$ $\mathbb{Z}/4\mathbb{Z} \simeq \langle i \rangle$*, via* $\overline{n} \mapsto i^n$. *This is an instance of the fact that two cyclic groups of the same order must be isomorphic. (In this case they are of order* 4*.)*

1.9. MATRIX REPRESENTATIONS OF GROUPS. .

Another look at O_2.

$$\mathrm{rot}_\alpha : \begin{pmatrix}1\\0\end{pmatrix} \mapsto \begin{pmatrix}\cos\alpha\\ \sin\alpha\end{pmatrix}, \quad \begin{pmatrix}0\\1\end{pmatrix} \mapsto \begin{pmatrix}-\sin\alpha\\ \cos\alpha\end{pmatrix}.$$

$$\mathrm{rot}_\alpha\left(\begin{pmatrix}x\\y\end{pmatrix}\right) = \mathrm{rot}_\alpha\left(x\begin{pmatrix}1\\0\end{pmatrix} + y\begin{pmatrix}0\\1\end{pmatrix}\right) = x\mathrm{rot}_\alpha\left(\begin{pmatrix}1\\0\end{pmatrix}\right) + y\mathrm{rot}_\alpha\left(\begin{pmatrix}0\\1\end{pmatrix}\right)$$

(this is the **linearity** of rot_α)

$$= x\begin{pmatrix}\cos\alpha\\ \sin\alpha\end{pmatrix} + y\begin{pmatrix}-\sin\alpha\\ \cos\alpha\end{pmatrix} = \begin{pmatrix}\cos\alpha & -\sin\alpha\\ \sin\alpha & \cos\alpha\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix} = M(\mathrm{rot}_\alpha)\begin{pmatrix}x\\y\end{pmatrix},$$

where $M(\mathrm{rot}_\alpha) := \begin{pmatrix}\cos\alpha & -\sin\alpha\\ \sin\alpha & \cos\alpha\end{pmatrix}$. Similarly,

$$\mathrm{ref}_\beta\left(\begin{pmatrix}x\\y\end{pmatrix}\right) = M(\mathrm{ref}_\beta)\begin{pmatrix}x\\y\end{pmatrix},$$

where $M(\mathrm{ref}_\beta) := \begin{pmatrix}\cos\beta & \sin\beta\\ \sin\beta & -\cos\beta\end{pmatrix}$. Given any $g \in O_2$, we now have $M(g) \in \mathrm{GL}_2(\mathbb{R})$ such that $g(\underline{v}) = M(g)\underline{v}$, for any $\underline{v} \in \mathbb{R}^2$. Now

$$(g_1g_2)(\underline{v}) = g_1(g_2(\underline{v})) = M(g_1)M(g_2)\underline{v},$$

so $M(g_1g_2) = M(g_1)M(g_2)$, i.e. $M : O_2 \to \mathrm{GL}_2(\mathbb{R})$ is a group homomorphism.

Definition 1.9.1. *A group homomorphism $f : G \to \mathrm{GL}_n(\mathbb{R})$ (or more generally to $\mathrm{GL}_n(\mathbb{C})$) is called a (n-dimensional)* **representation** *of G.*

We have found a 2-dimensional representation of O_2, i.e. $M : O_2 \to \mathrm{GL}_2(\mathbb{R})$. This restricts to a representation of any subgroup, e.g.

$$M(D_4) = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \right\}$$

Composing the homomorphisms $O_2 \overset{M}{\to} \mathrm{GL}_2(\mathbb{R}) \overset{\det}{\to} \mathbb{R}^\times$, we get a homomorphism $\det \circ M : O_2 \to \mathbb{R}^\times = \mathrm{GL}_1(\mathbb{R})$, i.e. $g \mapsto \det(M(g))$. This is a 1-dimensional representation of O_2. It is the familiar homomorphism such that $\mathrm{rot}_\alpha \mapsto 1$, $\mathrm{ref}_\beta \mapsto -1$.

To a great extent, the concept of a group is an attempt to capture mathematically the notion of symmetry. Many molecules have interesting finite symmetry groups, and it turns out that representations of these groups can be used to predict properties such as energy levels, differences of which show up in the wavelengths of radiation absorbed or emitted.

Google "A5 icosahedron". Click on "mathy coolness". The symmetry group of the icosahedron is isomorphic to the Cartesian product $A_5 \times \{\pm 1\}$ (with operation $(a_1, b_1)(a_2, b_2) = (a_1a_2, b_1b_2)$ for $a_1, a_2 \in A_5$ and $b_1, b_2 \in \{\pm 1\}$). The A_5 bit is just the rotations, and they naturally produce even permutations of the five colours in the picture, e.g. a rotation through $2\pi/5$ about an axis through opposite vertices produces a 5-cycle. There exist "irreducible" representations of A_5 of dimensions $1, 3, 3, 4, 5$, from which all the others can be "built".

Google "truncated icosahedron", which has the same symmetry group. Click on Wikipedia. Look at pictures, then click on "Carbon-60" (Buckminsterfullerene). The existence of C_{60} "buckyballs" was conjectured in the late 60s, early 70s. They were discovered in 1985 by Curl, Kroto (Sheffield undergraduate) and Smalley (Nobel Prize 1996). Mass spectroscopy of the results of the laser evaporation of graphite found C_{60} molecules, but it was still necessary to test the hypothesis that the atoms are arranged in a buckyball. Using the representation theory of $A_5 \times \{\pm 1\}$, it was possible to show that if so, the absorption spectrum should have only 4 lines in the infra-red range. This was confirmed experimentally around 1990 by Krätschmer and Huffman. In 2010, this absorption signature was detected in radiation from nebulae, providing evidence for the natural occurrence of buckyballs in outer space.

Infinite groups akin to O_2, whose elements depend on continuous parameters (like rot_α and ref_β) show up as symmetry groups of space and time, and their representations are central to theoretical particle physics.

2. INTRODUCTION TO RINGS

2.1. BASIC DEFINITIONS.

Definition 2.1.1. *A* **ring** *is a set R with binary operations $+, \cdot : R \times R \to R$, such that*

R1 *$(R, +)$ is an abelian group. [Call the neutral element "0" or "0_R".]*

R2 *$a(bc) = (ab)c, \quad \forall\, a, b, c \in R$ [i.e. $\cdot$ is associative. Note that "$a \cdot b$" is abbreviated to "ab".]*

R3 *(Distributive Laws) $(a + b)c = ac + bc$ and $a(b + c) = ab + ac \quad \forall a, b, c \in R$.*

R4 *There exists an element $1_R \in R$ (the multiplicative identity element) such that*

$$1_R \cdot a = a \cdot 1_R = a, \qquad \forall a \in R.$$

Example 2.1.2. *$\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$ are all rings, ultimately thanks to the basic laws of arithmetic in $\mathbb{N}$. But $\mathbb{N}$ itself is not a ring. We had to invent $\mathbb{Z}$ (i.e. create negative numbers) to ensure that every element has an additive inverse (for* **R1***).*

Definition 2.1.3. *A* **commutative ring** *is a ring R such that $ab = ba \ \ \forall a, b \in R$.*

Example 2.1.4. *$\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$ are all commutative rings.*

Definition 2.1.5. *A* **field** *is a non-zero commutative ring R such that*

$$\forall a \in R \text{ with } a \neq 0_R, \ \ \exists b \in R \text{ such that } ab = 1_R.$$

b is called a^{-1}, the multiplicative inverse of a.

Example 2.1.6. *$\mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$ are fields, but $\mathbb{Z}$ is not, since only ± 1 have multiplicative inverses inside $\mathbb{Z}$.*

2.2. SOME SIMPLE CONSEQUENCES OF THE RING AXIOMS. .

We shall use "$a - b$" as shorthand for "$a + (-b)$".

Proposition 2.2.1. *Let R be a ring. Then $0_R \cdot a = a \cdot 0_R = 0_R$ for all $a \in R$.*

Proof. $0_R = 0_R + 0_R$, by $R1$.
Hence $a \cdot 0_R = a \cdot (0_R + 0_R) = a \cdot 0_R + a \cdot 0_R$, by R3
Adding $-(a \cdot 0_R)$ to both sides, $0_R = a \cdot 0_R$ (using R1). Similarly, multiplying by a on the right instead of the left, we may show that $0_R = 0_R \cdot a$. □

Proposition 2.2.2. *Let R be a ring. If $1_R = 0_R$ then $R = \{0_R\}$.*

Proof. Suppose $1_R = 0_R$. For any $a \in R$, $a = 1_R \cdot a$ (by R4)
$= 0_R \cdot a$ (by assumption)
$= 0_R$ (by Proposition 2.2.1). □

Proposition 2.2.3. *Let R be a ring. For all $a, b \in R$, $-(ab) = (-a)b$.*

Proof. $ab + (-a)b = (a + (-a))b$ (by R3)$= 0_R b = 0_R$ (by Proposition 2.2.1).
Adding $-(ab)$ to both sides, $(-a)b = -(ab)$. □

2.3. **AN EXAMPLE: POLYNOMIAL RINGS.** .

Let R be a commutative ring. We can enlarge R to a ring $R[x]$ by adjoining an "indeterminate" x. As a set,

$$R[x] = \{a_0 + a_1 x + \ldots + a_n x^n \mid \ a_0, \ldots, a_n \in R, n \in \mathbb{N}_0\}.$$

These elements should be thought of as formal expressions rather than as functions, but we add and multiply them in the usual way, imagining x to stand for an unknown element of R. In particular, for multiplication, if

$$f(x) = a_0 + a_1 x + \ldots + a_n x^n \text{ and } \ g(x) = b_0 + b_1 x + \ldots + b_m x^m$$

then, expanding out the brackets and collecting together terms involving the same power of x,

$$f(x)g(x) = a_0 b_0 + (a_0 b_1 + a_1 b_0)x + \ldots + a_n b_m x^{n+m} = \sum_{r=0}^{n+m} \left(\sum_{t=0}^{r} a_t b_{r-t} \right) x^r = \sum_{r=0}^{n+m} \left(\sum_{s=0}^{r} b_s a_{r-s} \right) x^r$$

(putting $s = r - t$ and using commutativity of multiplication in R)

$$= g(x)f(x).$$

This shows that multiplication in $R[x]$ is commutative. Looking close-up at the multiplication of two terms,

$$(a_t x^t)(b_{r-t} x^{r-t}) = a_t (x^t b_{r-t}) x^{r-t} = a_t (b_{r-t} x^t) x^{r-t} = (a_t b_{r-t}) x^r.$$

To make this work, we have to decree that the element x we have created commutes with all elements of R. But we impose no other relations (beyond those required by the ring axioms, e.g. $1x = x$).

$R[x]$ is a commutative ring with additive identity 0 (i.e. 0_R) and multiplicative identity 1 (i.e. 1_R). Similarly we may adjoin 2 (or more) indeterminates to get a commutative ring such as $R[x, y]$. Here x and y commute with elements of R, and with each other, but satisfy no further relations (beyond those required by the ring axioms, e.g. $1x = x$).

2.4. **SUBRINGS.** .

Definition 2.4.1. *Let R be a ring. A subset $S \subseteq R$ is a* **subring** *of R if and only if S is a ring with the same $+, \cdot, 0$ and 1 as R.*

Example 2.4.2. .

$\mathbb{Z}$ *is a subring of* $\mathbb{Q}$.

$\mathbb{Q}$ *is a subring of* $\mathbb{R}$.

$\mathbb{R}$ *is a subring of* $\mathbb{C}$.

$\mathbb{Q}$ *is a subring of* $\mathbb{Q}[x]$.

$\mathbb{Q}[x]$ *is a subring of* $\mathbb{R}[x]$.

$\mathbb{R}[x]$ *is a subring of* $\mathbb{R}[x, y]$.

These are all things we already knew to be rings, and we are just observing that sometimes one is inside another. When inside a ring we have a subset which we do not already know to be a ring, we need the following.

Subring Criterion Given a ring R, and a subset $S \subseteq R$, to show that S is a subring of R it suffices to show

SR1 $a - b \in S \ \forall a, b \in S$;

SR2 $ab \in S \ \forall a, b \in S$;

SR3 $1_R \in S$.

Example 2.4.3. *The set of* **Gaussian integers** $\mathbb{Z}[i] = \{a + bi \mid \ a, b \in \mathbb{Z}\}$. *For example* $3 - 2i \in \mathbb{Z}[i]$. *Clearly* $\mathbb{Z}[i] \subseteq \mathbb{C}$. *But if* $a + bi, c + di \in \mathbb{Z}[i]$ *then*

(1) $(a + bi) - (c + di) = (a - c) + (b - d)i \in \mathbb{Z}[i]$;

(2) $(a + bi)(c + di) = (ac - bd) + (ad + bc)i \in \mathbb{Z}[i]$;

(3) $1 = 1 + 0i \in \mathbb{Z}[i]$,

so by the Subring Criterion, $\mathbb{Z}[i]$ *is a subring of* $\mathbb{C}$.

2.5. HAMILTON'S QUATERNION RING: OUR FIRST NONCOMMUTATIVE EXAMPLE.

[Google "William Rowan Hamilton", click on Wikipedia and "1.4 Quaternions".] As a set, $\mathbb{H} = \{a + bi + cj + dk \mid \ a, b, c, d \in \mathbb{R}\}$. They are added in the obvious way, and multiplied using the rules

$$i^2 = j^2 = k^2 = -1, \ ij = k, \ ji = -k, \ jk = i, \ kj = -i, \ ki = j, \ ik = -j.$$

Notice that this multiplication is *not* commutative.

Example 2.5.1. $(2 + i - j) + (3 + j + 2k) = 5 + i + 2k$, *while* $(2 + i - j)(3 + j + 2k) = 6 + 3i - j + 4k - j^2 + ij + 2ik - 2jk$
$= 6 + 3i - j + 4k - (-1) + k + 2(-j) - 2i = 7 + i - 3j + 5k$.

Definition 2.5.2. *Given* $\alpha = a + bi + cj + dk \in \mathbb{H}$ *(with* $a, b, c, d \in R$*), its* **conjugate** $\overline{\alpha} = a - bi - cj - dk$.

Note that

$$\alpha\overline{\alpha} = (a+bi+cj+dk)(a-bi-cj-dk)$$

$$= a^2-b^2i^2-c^2j^2-d^2k^2+a[(bi+cj+dk)-(bi+cj+dk)]-bc(ij+ji)-bd(ik+ki)-cd(jk+kj)$$

$$= a^2+b^2+c^2+d^2 \ (=\overline{\alpha}\alpha \text{ similarly}).$$

If $\alpha \neq 0$ then $\alpha\overline{\alpha} > 0$. Now if we let $\beta = \frac{\overline{\alpha}}{\alpha\overline{\alpha}}$, we have $\alpha\beta = \beta\alpha = 1$. We can write $\beta = \alpha^{-1}$.

Definition 2.5.3. *A (non-zero) ring R is said to be a* **division ring** *if every non-zero element has a multiplicative inverse.*

A *commutative* division ring is the same thing as a field. We have just found that $\mathbb{H}$ is a division ring. But it is noncommutative.

2.5.1. *Quaternions and vectors.* Let $\mathbf{a} = (a_1, a_2, a_3)$, $\mathbf{b} = (b_1, b_2, b_3)$ be two vectors in $\mathbb{R}^3$. Identify them with "pure" quaternions, so $\mathbf{a} = a_1i + a_2j + a_3k$ and $\mathbf{b} = b_1i + b_2j + b_3k$. Now if we multiply them together,

$$\mathbf{ab} = -(a_1b_1 + a_2b_2 + a_3b_3) + (a_2b_3 - a_3b_2)i + (a_3b_1 - a_1b_3)j + (a_1b_2 - a_2b_1)k.$$

We recognise the scalar product and the vector product of $\mathbf{a}$ and $\mathbf{b}$ appearing in the answer. This suggests that $\mathbb{H}$ is not a completely arbitrary construction. Hamilton searched for a long time for a way to multiply together ordered 3-tuples of real numbers, bearing a similar relationship to rotation in 3-dimensional space as that between multiplication of complex numbers and rotation in 2-dimensional space (addition of arguments). He searched fruitlessly for a long time before realising that first, it was better to look at 4-tuples, and second, he needed to abandon commutativity of multiplication, which had always been implicitly assumed without even imagining it could be any different.

2.5.2. *Quaternions and rotations.* As above, identify the vector $\mathbf{a}$ with a quaternion $a_1i+a_2j+a_3k$. Suppose now we want to rotate $\mathbf{a}$ (in some positive sense) through an angle θ about an axis along the unit vector $\mathbf{u} = u_1i+u_2j+u_3k$. What vector $\mathbf{b}$ do we end up with? It is possible to show that if we let $p = \cos(\theta/2) + \sin(\theta/2)\mathbf{u}$ then $\mathbf{b} = p\mathbf{a}\overline{p}$.

Example 2.5.4. *If $\mathbf{a} = i$ and we want to rotate it through $\pi/2$ about k, we know the answer is j. Let's see if the recipe produces the right answer. We have $p = \cos(\pi/4)+\sin(\pi/4)k = \frac{1}{\sqrt{2}}(1+k)$, so*

$$p\mathbf{a}\overline{p} = \frac{1}{\sqrt{2}}(1+k)i\frac{1}{\sqrt{2}}(1-k) = \frac{1}{2}(1+k)(i-ik) = \frac{1}{2}(1+k)(i+j)$$
$$= (1/2)(i+j+ki+kj) = (1/2)(i+j+j-i) = j,$$

as expected.

Google "Rotating Objects Using Quaternions", for applications to computer graphics, flight simulation, etc.

2.6. **ANOTHER NONCOMMUTATIVE EXAMPLE: MATRIX RINGS.** .

Let R be a ring, and $n \geq 1$ a natural number. Let $M_n(R) = \{A = (A_{ij}) \mid \ A_{ij} \in R,\ 1 \leq i, j \leq n\}$. Here, A_{ij} is an element of R and (A_{ij}) is an n-by-n array whose entry in the i^{th} row and j^{th} column is A_{ij}. Define addition by

$$(A_{ij}) + (B_{ij}) = (A_{ij} + B_{ij})$$

(i.e. $A_{ij} + B_{ij}$ is the i, j-entry of the sum) and multiplication by

$$(AB)_{ij} = \sum_{t=1}^{n} A_{it}B_{tj}.$$

We can check that this makes $M_n(R)$ into a ring.

R1 It is easy to check that $(M_n(R), +)$ is an abelian group, with neutral element 0_n (all entries equal to 0_R) and $-(A_{ij}) = (-A_{ij})$.

R2 Associativity:

$$\begin{aligned}
((AB)C)_{ij} &= \sum_{t=1}^{n} (AB)_{it}C_{tj} = \sum_{t=1}^{n} \left(\sum_{s=1}^{n} A_{is}B_{st} \right) C_{tj} \\
&= \sum_{t=1}^{n} \sum_{s=1}^{n} (A_{is}B_{st})C_{tj} \text{ (by the distributive law } \mathbf{R3} \text{ for } R) \\
&= \sum_{t=1}^{n} \sum_{s=1}^{n} A_{is}(B_{st}C_{tj}) \text{ (by associativity } \mathbf{R2} \text{ for } R) \\
&= \sum_{s=1}^{n} \sum_{t=1}^{n} A_{is}(B_{st}C_{tj}) \text{ (by commutativity of addition in } R) \\
&= \sum_{s=1}^{n} A_{is} \left(\sum_{t=1}^{n} B_{st}C_{tj} \right) \text{ (by } \mathbf{R3} \text{ again)} \\
&= \sum_{s=1}^{n} A_{is}(BC)_{sj} = (A(BC))_{ij}.
\end{aligned}$$

Hence $A(BC) = (AB)C \ \forall A, B, C \in M_n(R)$.

R3 Exercise to show that $A(B+C) = AB+AC$ and $(A+B)C = AC+BC \ \forall A, B, C \in M_n(R)$.

R4 The identity matrix $I_n = \text{diag}(1_R, \ldots, 1_R)$ (with 0_R everywhere off the leading diagonal) is the multiplicative identity.

The ring $M_n(R)$ is always noncommutative if $n \geq 2$ and $R \neq \{0\}$. For example, when $n = 2$,

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

Even if R is a field (let's suppose that it is), $M_n(R)$ is not a division ring once $n \geq 2$. A multiplicative inverse A^{-1} exists $\iff$ $\det(A) \neq 0$, for example $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ is a non-zero matrix without a multiplicative inverse.

2.7. **ANOTHER NONCOMMUTATIVE EXAMPLE: THE WEYL ALGEBRA.** .

The Weyl algebra W (or $W_1(\mathbb{C})$) is created by starting with $\mathbb{C}$ and adjoining two new elements P, Q which commute with everything in $\mathbb{C}$ but not with each other:

$$PQ - QP = 1.$$

Elements look a bit like polynomials in two variables, e.g.

$$f = 1 + 2P + P^2Q, \quad g = 2 - 3P,$$

and are added just like polynomials:

$$f + g = 3 - P + P^2Q;$$

but multiplied differently:

$$\begin{aligned} fg &= (1 + 2P + P^2Q)(2 - 3P) \\ &= 2 + P - 6P^2 + 2P^2Q - 3P^2QP \\ &= 2 + P - 6P^2 + 2P^2Q - 3P^2(PQ - 1) \\ &= 2 + P - 3P^2 + 2P^2Q - 3P^3Q. \end{aligned}$$

For multiplication of polynomials, in bringing the Ps to the left and the Qs to the right we would have just replaced QP by PQ, but here we have to use the fact that the difference between PQ and QP is not 0, rather 1. The Weyl algebra is important in quantum mechanics, where Q stands for position, P for momentum, and $PQ - QP = 1$ represents Heisenberg's uncertainty relation.

2.8. **UNITS.**

Definition 2.8.1. *Let R be a ring. A* **unit** *in R is an element $a \in R$ with a* **multiplicative inverse**, *i.e. an element $b \in R$ such that $ab = ba = 1_R$. The set of all units in R is denoted $U(R)$.*

Example 2.8.2. .

(1) $U(\mathbb{Z}) = \{\pm 1\}$;

(2) $U(\mathbb{Q}) = \mathbb{Q}^\times := \mathbb{Q}\backslash\{0\}$;

(3) $U(\mathbb{R}) = \mathbb{R}^\times := \mathbb{R}\backslash\{0\}$ (likewise for any field, cf. Definition 2.1.5);

(4) $U(\mathbb{H}) = \mathbb{H}\backslash\{0\}$ *(likewise for any division ring, cf. Definition 2.5.3);*

(5) $U(M_2(\mathbb{R})) = \mathrm{GL}_2(\mathbb{R}) = \{A \in M_2(\mathbb{R}) \mid \det(A) \neq 0\}$.

Proposition 2.8.3. $U(\mathbb{Z}[i]) = \{\pm 1, \pm i\}$.

Proof. Since $1 = 1 \times 1 = (-1) \times (-1) = i \times (-i)$, they clearly are units. Why the only ones? If α is a unit, say $\alpha\beta = 1$. Then $\overline{\alpha}\overline{\beta} = 1$, so $(\alpha\overline{\alpha})(\beta\overline{\beta}) = 1$. But $\alpha\overline{\alpha}, \beta\overline{\beta} \in \mathbb{Z}[i] \cap \mathbb{R} = \mathbb{Z}$, so are units in $\mathbb{Z}$. Hence $\alpha\overline{\alpha} = \pm 1$. If $\alpha = a + bi$ then $\alpha\overline{\alpha} = a^2 + b^2 > 0$, so $a^2 + b^2 = 1$. The only solutions are $(a, b) = (\pm 1, 0)$ or $(0, \pm 1)$, leading to ± 1 and $\pm i$ respectively. □

It is easy to prove that in general $U(R)$ is a group under multiplication, and this fact is amply illustrated by the above examples.

Challenge question. In W, is P a unit?

2.9. RING HOMOMORPHISMS.

Definition 2.9.1. *Let R, S be rings. A function $f : R \to S$ is said to be a* **ring homomorphism** *if*

(1) $f(a + b) = f(a) + f(b) \quad \forall a, b \in R$;

(2) $f(ab) = f(a)f(b) \quad \forall a, b \in R$;

(3) $f(1_R) = 1_S$

If f is also a bijection, then f is a **ring isomorphism**, $R \simeq S$.

Example 2.9.2. *If R is a subring of S then the inclusion $i : R \hookrightarrow S$ is a ring homomorphism.*

Example 2.9.3. *Define $f : \mathbb{C} \to \mathbb{C}$ by $f(z) = \overline{z}$. This f is a bijection (equal to its own inverse function). In fact, f is an isomorphism: (1),(2) and (3) just say $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$, $\overline{z_1 z_2} = \overline{z_1}\ \overline{z_2}$ and $\overline{1} = 1$, respectively.*

Note that the identity map from any ring to itself (here it would be $\mathrm{id}(z) = z \quad \forall z \in \mathbb{C}$) is always an isomorphism of a ring with itself, not a very interesting one. But in this example we have a non-identity isomorphism from $\mathbb{C}$ to itself. The following example is more typical, a homomorphism from a ring to a different ring.

Example 2.9.4. *Define $f : \mathbb{C} \to M_2(\mathbb{R})$ by* $f(x + iy) = \begin{pmatrix} x & -y \\ y & x \end{pmatrix}$.

Claim: *f is a ring homomorphism.*

Proof. (1)

$$f(x_1 + iy_1) + f(x_2 + iy_2) = \begin{pmatrix} x_1 & -y_1 \\ y_1 & x_1 \end{pmatrix} + \begin{pmatrix} x_2 & -y_2 \\ y_2 & x_2 \end{pmatrix}$$

$$= \begin{pmatrix} x_1 + x_2 & -(y_1 + y_2) \\ y_1 + y_2 & x_1 + x_2 \end{pmatrix} = f[(x_1 + iy_1) + (x_2 + iy_2)].$$

(2)

$$f(x_1 + iy_1)f(x_2 + iy_2) = \begin{pmatrix} x_1 & -y_1 \\ y_1 & x_1 \end{pmatrix} \begin{pmatrix} x_2 & -y_2 \\ y_2 & x_2 \end{pmatrix} = \begin{pmatrix} x_1x_2 - y_1y_2 & -(x_1y_2 + x_2y_1) \\ x_1y_2 + x_2y_1 & x_1x_2 - y_1y_2 \end{pmatrix}$$

$$= f((x_1x_2 - y_1y_2) + i(x_1y_2 + x_2y_1)) = f((x_1 + iy_1)(x_2 + iy_2)).$$

(3) $f(1) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

□

Note that in the preceding example, $f(z)$ is the matrix representing the linear transformation "multiply by z" in the complex plane. e.g. $f(i) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ represents $\mathrm{rot}_{\pi/2}$ (cf. Section 1.9). We have $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$.

Example 2.9.5. *Define $f_i : \mathbb{C} \to \mathbb{H}$ by $f_i(a + ib) = a + bi$. This is obviously a ring homomorphism. But so is*

$$f_j : \mathbb{C} \to \mathbb{H} \quad f_j(a + ib) = a + bj, \text{ and}$$

$$f_k : \mathbb{C} \to \mathbb{H} \quad f_k(a + ib) = a + bk, \quad \text{since } i^2 = j^2 = k^2 = -1.$$

In fact, for any point (x, y, z) on the sphere $x^2 + y^2 + z^2 = 1$, if $\alpha = xi + yj + zk \in \mathbb{H}$ then $\alpha^2 = -1$ and $a + ib \mapsto a + b\alpha$ is a ring homomorphism from $\mathbb{C}$ to $\mathbb{H}$.

Proposition 2.9.6. *Let $f : R \to S$ be a ring homomorphism. Then the image $\mathrm{im}(f) = f(R)$ is a subring of S.*

Proof. We check the Subring Criterion.

SR1: If $a_1, a_2 \in f(R)$, say $a_1 = f(r_1)$ and $a_2 = f(r_2)$ then

$$a_1 - a_2 = f(r_1) - f(r_2) = f(r_1 - r_2) \text{ (since } f \text{ is a homomorphism of additive groups)}$$

$$\in f(R).$$

SR2: $a_1a_2 = f(r_1)f(r_2) = f(r_1r_2) \in f(R)$.

SR3: $1_S = f(1_R) \in f(R)$. □

Note that if $f : R \to S$ is an **injective** ring homomorphism then f gives an **isomorphism** between R and the subring $f(R)$ of S. E.g. $\mathbb{C}$ is isomorphic to the subring $\left\{ \begin{pmatrix} x & -y \\ y & x \end{pmatrix} : \ x, y \in \mathbb{R} \right\}$ of $M_2(\mathbb{R})$ (not the whole of $M_2(\mathbb{R})$, which, unlike $\mathbb{C}$, would be non-commutative), and also to infinitely many subrings of $\mathbb{H}$, such as $\{a + bj \in \mathbb{H} \mid \ a, b \in \mathbb{R}\}$.

Example 2.9.7. *Define (given a fixed $a \in \mathbb{R}$) a function $\mathrm{ev}_a : \mathbb{R}[x] \to \mathbb{R}$ (evaluation at a), by $\mathrm{ev}_a(f(x)) = f(a)$. We have*

$$\mathrm{ev}_a(f(x) + g(x)) = f(a) + g(a) = \mathrm{ev}_a(f(x)) + \mathrm{ev}_a(g(x)),$$

$$\mathrm{ev}_a(f(x)g(x)) = f(a)g(a) = \mathrm{ev}_a(f(x))\mathrm{ev}_a(g(x))$$

and $\mathrm{ev}_a(1) = 1$, so ev_a is a ring homomorphism. It is surjective, since any $r \in \mathbb{R}$ is $\mathrm{ev}_a(r)$, but it is certainly not injective, since $\mathrm{ev}_a(x - a) = \mathrm{ev}_a(0) = 0$.

Example 2.9.8. *Define $\mathrm{ev}_i : \mathbb{R}[x] \to \mathbb{C}$ by $\mathrm{ev}_i(f(x)) = f(i)$. Similarly to the previous example, ev_i is a ring homomorphism. It is surjective, since any $a + ib \in \mathbb{C}$ is $\mathrm{ev}_i(a + bx)$. But ev_i is not injective, since $\mathrm{ev}_i(x^2 + 1) = \mathrm{ev}_i(0) = 0$.*

Example 2.9.9. *Given $(a, b) \in \mathbb{R}^2$, define $\mathrm{ev}_{(a,b)} : \mathbb{R}[x, y] \to \mathbb{R}$ by $\mathrm{ev}_{(a,b)}(f(x, y)) = f(a, b)$. This is a surjective ring homomorphism, but $\mathrm{ev}_{(a,b)}(x - a) = \mathrm{ev}_{(a,b)}(y - b) = \mathrm{ev}_{(a,b)}(0) = 0$, so it is not injective.*

2.10. QUOTIENT RINGS AND THE FIRST ISOMORPHISM THEOREM FOR RING HOMOMORPHISMS. .

Let $F : R \to S$ be a ring homomorphism. Then in particular it is a group homomorphism, viewing R and S as groups under $+$, temporarily ignoring multiplication. By Theorem 1.8.18 (the First Isomorphism Theorem), there is an **isomorphism** of additive groups

$$\overline{f} : R/\ker f \xrightarrow{\sim} \mathrm{im}(f)$$

given by $\overline{f}(\overline{a}) = f(a)$, where $\overline{a}$ stands for the element $a + \ker f$ of $R/\ker f$, and $\ker(f) = \{a \in R \mid \ f(a) = 0_S\}$. We may sometimes use the alternative notation $[a]$ for $\overline{a}$.

We know from Proposition 2.9.6 that $\mathrm{im}(f)$ is a subring of S, in particular, the right-hand-side of the isomorphism is a ring, not just an additive group. This raises two questions:

(1) Is the left-hand-side (i.e. $R/\ker f$) also a ring?
(2) If so, is $\overline{f}$ a ring isomorphism?

The answer to both these questions is "Yes". (**Optional from here until the beginning of Example 2.10.2.**) The most important point is that we can define an operation of multiplication on $R/\ker f$, in the natural way: $\overline{a}\,\overline{b} := \overline{ab}$, where on the right ab is a product in R,

already given since R is a ring, but on the left we are defining how to multiply the elements $\overline{a}$ and $\overline{b}$ of $R/\ker f$. The main thing we have to worry about is whether this is well-defined. If $k_1, k_2 \in \ker f$ then $\overline{a} = \overline{a+k_1}$ and $\overline{b} = \overline{b+k_2}$, so we must show that writing these elements the different way does not produce a different answer, i.e. that $\overline{(a+k_1)(b+k_2)} = \overline{ab}$. This boils down to showing that $ak_2 + k_1b + k_1k_2 \in \ker f$, i.e. that $f(ak_2 + k_1b + k_1k_2) = 0$. But since f is a ring homomorphism,

$$f(ak_2 + k_1b + k_1k_2) = f(ak_2) + f(k_1b) + f(k_1k_2) = f(a)f(k_2) + f(k_1)f(b) + f(k_1)f(k_2) = 0,$$

since $k_1, k_2 \in \ker f \implies f(k_1) = f(k_2) = 0$.

Now for each of the ring axioms, the fact that it holds for $R/\ker f$ is a direct consequence of the fact that it holds for R. For example, $\overline{1_R}$ plays the role of $1_{R/\ker f}$, since for any $\overline{a} \in R/\ker f$, $\overline{1_R}\overline{a} = \overline{1_R a} = \overline{a}$, and similarly $\overline{a}\overline{1_R} = \overline{a}$. It is also easy now to show that $\overline{f}$ is a ring homomorphism. We already had $\overline{f}(\overline{a}+\overline{b}) = \overline{f}(\overline{a}) + \overline{f}(\overline{b})\ \ \forall \overline{a}, \overline{b} \in R/\ker f$, from the First Isomorphism Theorem for groups. Now similarly

$$\overline{f}(\overline{a}\,\overline{b}) = \overline{f}(\overline{ab}) = f(ab) = f(a)f(b) = \overline{f}(\overline{a})\overline{f}(\overline{b}),$$

and $\overline{f}(1_{R/\ker f}) = \overline{f}(\overline{1_R}) = f(1_R) = 1_S$.

Example 2.10.1. *In Example 2.9.7 we had, for any fixed $a \in \mathbb{R}$, a surjective ring homomorphism $\mathrm{ev}_a : \mathbb{R}[x] \to \mathbb{R}$, given by $\mathrm{ev}_a(f(x)) = f(a)$. By the Factor Theorem (Proposition 3.2.4 later), $\ker f = \langle x-a\rangle := \{h(x)(x-a) \mid \ h(x) \in \mathbb{R}[x]\}$, the set of multiples of $(x-a)$ in $\mathbb{R}[x]$. Hence we have a ring isomorphism $\overline{\mathrm{ev}_a} : R/\langle x-a\rangle \xrightarrow{\sim} \mathbb{R}$ given by $\overline{\mathrm{ev}_a}(\overline{g(x)}) = g(a)$.*

Given an additive subgroup I of a ring R, when can we turn the quotient group R/I into a quotient *ring* in the manner that we did when $I = \ker f$? The answer is that we can do it if and only if I is an **ideal**. This means that in addition to being an additive subgroup of R, I is required to satisfy $RI \subseteq I$ (and also $IR \subseteq I$, in case R is not commutative). This is what guarantees that $ak_2 + k_1b + k_1k_2 \in I$ whenever $k_1, k_2 \in I$ and $a, b \in R$, so that $\overline{(a+k_1)(b+k_2)} = \overline{ab}$.

Example 2.10.2. *$3\mathbb{Z}$ is an ideal in $\mathbb{Z}$. The quotient group $\mathbb{Z}/3\mathbb{Z}$ is a quotient ring. To the addition table*

$+$	$\overline{0}$	$\overline{1}$	$\overline{2}$
$\overline{0}$	$\overline{0}$	$\overline{1}$	$\overline{2}$
$\overline{1}$	$\overline{1}$	$\overline{2}$	$\overline{0}$
$\overline{2}$	$\overline{2}$	$\overline{0}$	$\overline{1}$

we may now add a multiplication table

$\cdot$	$\overline{0}$	$\overline{1}$	$\overline{2}$
$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$
$\overline{1}$	$\overline{0}$	$\overline{1}$	$\overline{2}$
$\overline{2}$	$\overline{0}$	$\overline{2}$	$\overline{1}$

(cf. Example 1.5.13).

3. DIVISIBILITY AND FACTORISATION

3.1. SOME DEFINITIONS.

Definition 3.1.1. *A commutative ring R is called an* **integral domain** *if $\forall a, b \in R$, if $ab = 0$ then $a = 0$ or $b = 0$.*

Proposition 3.1.2. *If R is a subring of a field F (cf. Definition 2.1.5), then R is an integral domain.*

Proof. Suppose $a, b \in R$ and $ab = 0$. We must show that if $a \neq 0$ then $b = 0$. If $a \neq 0$ then in F we have a^{-1}. Multiplying both sides of $ab = 0$ by a^{-1} (and using Proposition 2.2.1) gives $b = 0$, as desired. □

Example 3.1.3. *$\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$ are all integral domains. $\mathbb{Z}/4\mathbb{Z}$ is not an integral domain, since $\overline{2}\,\overline{2} = \overline{4} = \overline{0}$, while $\overline{2} \neq \overline{0}$.*

Lemma 3.1.4. *[Cancellation Lemma] Let R be an integral domain. If $a, b, c \in R$ with $ac = bc$ and $c \neq 0$, then $a = b$.*

Proof. $ac = bc \implies ac - bc = 0 \implies (a-b)c = 0$. Since R is an integral domain, if $c \neq 0$ then $a - b = 0$, so $a = b$. □

Definition 3.1.5. *Let R be a commutative ring, $a, b \in R$. We say that a* **divides** *b, written $a \mid b$, if $\exists c \in R$ such that $ac = b$.*

Example 3.1.6. *In $\mathbb{Z}$, $7 \mid 21$ while $3 \nmid 7$.*
In $\mathbb{Q}$, $3 \mid 7$, because $3(7/3) = 7$.
In $\mathbb{Q}[x]$, $(x-2) \mid x^2 - 3x + 2$, while $(x-1) \nmid x^2 + 1$.

Definition 3.1.7. *Given elements a, b of a commutative ring R, we say that a is* **associate** *to b if there is a unit u (recall Definition 2.8.1) such that $a = ub$.*

One may show that this is an equivalence relation, so we may talk of a and b being associate to each other.

Example 3.1.8. *In $\mathbb{Z}$, 7 and -7 are associates.*

Definition 3.1.9. *Let R be a commutative ring, $r \in R$ with $r \neq 0$, r not a unit. We say that r is* **irreducible** *if whenever $r = ab$, either a or b is a unit. (Equivalently, whenever $r = ab$, either a or b is associate to r.)*

Example 3.1.10. *In $\mathbb{Z}$, the irreducibles are the prime numbers and their associates.*
$7 = 1 \times 7 = (-1) \times (-7)$. *Here, 1 and -1 are units, and ± 7 are associates of 7. There is no other way of factorising 7.*

Example 3.1.11. *In $\mathbb{Q}[x]$, $x^2 - 2$ is irreducible. We can only factor it as the product of a unit and an associate e.g. as $1 \times (x^2 - 2)$ or as $(1/7) \times (7x^2 - 14)$. There is no "proper" factorisation. But in $\mathbb{R}[x]$, $x^2 - 2 = (x - \sqrt{2})(x + \sqrt{2})$, so is not irreducible, but $(x - \sqrt{2})$ and $(x + \sqrt{2})$ are now irreducible in $\mathbb{R}[x]$.*

Example 3.1.12. *In $\mathbb{R}[x]$, $x^2 + 1$ is irreducible, but in $\mathbb{C}[x]$, $x^2 + 1 = (x - i)(x + i)$ is reducible (i.e. not irreducible).*

3.2. **EUCLIDEAN DOMAINS.**

Definition 3.2.1. *A* **Euclidean domain** *is an integral domain R with a function $\delta : R - \{0\} \to \mathbb{N}_0$ such that for all $a, b \in R - \{0\}$,*

(1) $\exists q, r \in R$ ("quotient" and "remainder") such that $a = qb + r$, with either $r = 0$ (if $b \mid a$) or $\delta(r) < \delta(b)$;
(2) if $b \mid a$ in R then $\delta(b) \leq \delta(a)$.

Example 3.2.2. *$\mathbb{Z}$ is a Euclidean domain, with $\delta(n) = |n|$ for all non-zero $n \in \mathbb{Z}$.*

Example 3.2.3. *If k is a field then $k[x]$ is a Euclidean domain, with $\delta(f) = \deg(f)$ for all non-zero $f \in k[x]$.*

Proposition 3.2.4. *[Factor Theorem] If k is a field, and if $f(x) \in k[x]$ and $\alpha \in k$, we have $f(\alpha) = 0 \iff (x - \alpha) \mid f(x)$.*

Proof. We may assume that $f \neq 0$ (not the zero polynomial). First, if $(x - \alpha) \mid f(x)$, say $f(x) = (x - \alpha)g(x)$, then clearly $f(\alpha) = 0$. We must show that conversely, if $f(\alpha) = 0$ then $(x - \alpha) \mid f(x)$. But $f(x) = q(x)(x - \alpha) + r(x)$, with $r(x) \equiv 0$ or $\deg(r) < \deg((x - \alpha)) = 1$, so all-in-all, $\deg(r) = 0$ and r is a constant. Plugging in $x = \alpha$ to $f(x) = q(x)(x - \alpha) + r$, $f(\alpha) = r$. So if $f(\alpha) = 0$ then $r = 0$ and $f(x) = q(x)(x - \alpha)$, so $(x - \alpha) \mid f(x)$, as required. □

Example 3.2.5. *The ring of Gaussian integers $\mathbb{Z}[i]$, with $\delta(\alpha) = \alpha\overline{\alpha}$, i.e. $\delta(a + bi) = a^2 + b^2$, is a Euclidean domain.*

Proof. First note that we can extend the function $\delta : \mathbb{Z}[i] - \{0\} \to \mathbb{N}_0$ to $\delta : \mathbb{C} \to \mathbb{R}_{\geq 0}$, by the same formula, $\delta(\alpha) = \alpha\overline{\alpha}$, and we have

$$\delta(\alpha\beta) = \alpha\beta\overline{(\alpha\beta)} = \alpha\beta\overline{\alpha}\,\overline{\beta} = \alpha\overline{\alpha}\beta\overline{\beta} = \delta(\alpha)\delta(\beta) \ \ \forall \alpha, \beta \in \mathbb{C}.$$

Now, take any $\alpha = a + bi, \beta = c + di \in \mathbb{Z}[i] - \{0\}$. Then $\frac{\alpha}{\beta} = \frac{(a+bi)(c-di)}{c^2+d^2}$ exists in $\mathbb{C}$, say $(\alpha/\beta) = s + ti$, with $s, t \in \mathbb{Q}$ but not necessarily in $\mathbb{Z}$. There exist $p, q \in \mathbb{Z}$ with $|p - s| \leq 1/2$ and $|q - t| \leq 1/2$. Let $\gamma = p + qi \in \mathbb{Z}[i]$. (We can think of this as the nearest point of the lattice $\mathbb{Z}[i]$ to the point α/β in the complex plane.) Then

$$\delta\left(\frac{\alpha}{\beta} - \gamma\right) = (p - s)^2 + (q - t)^2 \leq (1/2)^2 + (1/2)^2 < 1.$$

Hence

$$\delta(\alpha - \gamma\beta) = \delta\left(\beta\left(\frac{\alpha}{\beta} - \gamma\right)\right) = \delta(\beta)\delta\left(\frac{\alpha}{\beta} - \gamma\right) < \delta(\beta).$$

(1) Letting $\mathfrak{r} = \alpha - \gamma\beta$, we have $\alpha = \gamma\beta + \mathfrak{r}$, with $\gamma, \mathfrak{r} \in \mathbb{Z}[i]$, and $\delta(\mathfrak{r}) < \delta(\beta)$.

(2) If $\beta \mid \alpha$, say $\alpha = \gamma\beta$, then $\delta(\beta) \leq \delta(\alpha)$, since $\delta(\alpha) = \delta(\gamma)\delta(\beta)$ and $\delta(\gamma) \geq 1$ (because $\delta(\gamma) \in \mathbb{N}$ and is not 0).

□

3.3. **FACTORISATION INTO IRREDUCIBLES IN EUCLIDEAN DOMAINS.**

Lemma 3.3.1. *If (R, δ) is a Euclidean domain, $a \in R$, and $a = bc$ with b, c non-zero, non-units, then $\delta(b) < \delta(a)$ (and similarly $\delta(c) < \delta(a)$).*

Note that (2) in the definition of Euclidean domain only gives $\delta(b) \leq \delta(a)$, not $\delta(b) < \delta(a)$.

Proof. We have $b = qa + r$ with $r = 0$ or $\delta(r) < \delta(a)$. If $r = 0$ then $b = qa = qbc$. By Lemma 3.1.4, cancelling b, $1 = qc$, contrary to c being a non-unit. So we must have $\delta(r) < \delta(a)$. Now $b = qa + r \implies r = b - qa = b - qbc = b(1 - qc)$, so $b \mid r$. Hence $\delta(b) \leq \delta(r) < \delta(a)$, as required. □

Theorem 3.3.2. *Let R be a Euclidean domain. If $a \in R$ is a non-zero non-unit, then there exist irreducibles $p_1, \ldots, p_s \in R$ such that $a = p_1 p_2 \ldots p_s$.*

Proof. If a is irreducible then $s = 1$ and $a = p_1$. If a is not irreducible then $a = bc$ for some non-units bc, i.e. b and c are both *proper* factors of a (neither an associate of a). By Lemma 3.3.1, $\delta(b) < \delta(a)$ and $\delta(c) < \delta(a)$. Now continue factoring b and c (if they are not irreducible), and factoring their factors, etc. This process must terminate in the required kind of factorisation into irreducibles, otherwise we get an infinite sequence of *proper* divisibilities $\ldots a_3 \mid a_2 \mid a_1 \mid a$ (where $a_1 = b$ or c), and a corresponding sequence $\ldots \delta(a_3) < \delta(a_2) < \delta(a_1) < \delta(a)$, contradicting the fact that there is no infinite decreasing sequence of natural numbers. □

Example 3.3.3. *In $\mathbb{Z}$, $462 = 2 \times 231 = 2 \times 11 \times 21 = 2 \times 11 \times 3 \times 7$.*

Example 3.3.4. *In $\mathbb{C}[x]$, if $f(x)$ is non-constant then by the Fundamental Theorem of Algebra (which will be proved in MAS332 Complex Analysis), there exists $\alpha \in \mathbb{C}$ such that $f(\alpha) = 0$. By the Factor Theorem, $f(x) = (x-\alpha)g(x)$ for some $g(x) \in \mathbb{C}[x]$. Starting again with $g(x)$, we eventually arrive at a factorisation of $f(x)$ into linear factors (which are necessarily irreducible). For example*

$$(x^4 - 1) = (x-1)(x^3 + x^2 + x + 1) = (x-1)(x+1)(x^2+1) = (x-1)(x+1)(x-i)(x+i).$$

Example 3.3.5. *In $\mathbb{R}[x]$, irreducibles can be either linear or quadratic, e.g. $x^4 - 1 = (x-1)(x+1)(x^2+1)$ is a factorisation into irreducibles in $\mathbb{R}[x]$.*

Example 3.3.6. *In $\mathbb{Q}[x]$, $x^2 - 2$ is already irreducible, but in $\mathbb{R}[x]$ it factors $x^2 - 2 = (x - \sqrt{2})(x + \sqrt{2})$.*

Example 3.3.7. *In $\mathbb{Z}$, 5 is irreducible, but in $\mathbb{Z}[i]$, $5 = (2-i)(2+i)$.*

Lemma 3.3.8. *If $\alpha \in \mathbb{Z}[i]$ and $\delta(\alpha)$ is a prime number, then α is irreducible.*

Proof. If α is reducible then $\alpha = \beta\gamma$ with $\beta, \gamma \in \mathbb{Z}[i]$ both non-units. In $\mathbb{Z}$, $\delta(\alpha) = \delta(\beta)\delta(\gamma)$ (as in Example 3.2.5) with (by Lemma 3.3.1) $\delta(\beta) < \delta(\alpha)$ and $\delta(\gamma) < \delta(\alpha)$, contrary to $\delta(\alpha)$ being a prime number. □

It follows that $2 \pm i$ are irreducible elements of $\mathbb{Z}[i]$, since $\delta(2 \pm i) = 5$, a prime number.

3.4. **EUCLID'S ALGORITHM.**

Definition 3.4.1. *Let (R, δ) be a Euclidean domain, $a, b \in R - \{0\}$.*

(1) d is a **common divisor** *of a and b if $d \mid a$ and $d \mid b$.*

(2) d is a **greatest common divisor** *of a and b if*

(a) d is a common divisor of a and b;

(b) among common divisors of a and b, $\delta(d)$ is maximal.

We will show that in a Euclidean domain any non-zero a, b have a greatest common divisor (denoted $\gcd(a, b)$), by showing how to find it, as follows (Euclid's algorithm).

Let $a_0 = a, a_1 = b$.

First, $a_0 = q_1 a_1 + a_2$, with $a_2 = 0$ or $\delta(a_2) < \delta(a_1)$.

If $a_2 \neq 0$, $a_1 = q_2 a_2 + a_3$, with $a_3 = 0$ or $\delta(a_3) < \delta(a_2)$.

If $a_3 \neq 0$, $a_2 = q_3 a_3 + a_4$, with $a_4 = 0$ or $\delta(a_4) < \delta(a_3)$.

etc.

We can't have an infinite decreasing sequence of natural numbers

$$\delta(a_1) > \delta(a_2) > \delta(a_3) > \dots,$$

so eventually some $a_{n+1} = 0$ and we may define $d := a_n$, the last non-zero remainder.

Theorem 3.4.2. *(1) d is a* $\gcd(a, b)$*;*

(2) $d = ka + \ell b$, for some $k, \ell \in R$.

Proof. (1) $a_{n-1} = q_n a_n + 0$, so $a_n \mid a_{n-1}$, i.e. $d \mid a_{n-1}$.

$a_{n-2} = q_{n-1}a_{n-1} + a_n$. Given $d \mid a_n$ and $d \mid a_{n-1}$, also $d \mid a_{n-2}$.

$a_{n-3} = q_{n-2}a_{n-2} + a_{n-1}$. Given $d \mid a_{n-1}$ and $d \mid a_{n-2}$, also $d \mid a_{n-3}$.

etc.

Working all the way back up, we find that $d \mid a$ and $d \mid b$, so d is a common divisor of a and b.

Now let e be any common divisor of a and b (i.e. of a_0 and a_1).

$a_0 = q_1 a_1 + a_2 \implies e \mid a_2$ (given already $e \mid a_0$ and $e \mid a_1$).

$a_1 = q_2 a_2 + a_3 \implies e \mid a_3$ (given already $e \mid a_1$ and $e \mid a_2$).

Continuing this way, we find eventually that $e \mid a_n$, i.e. $e \mid d$. Hence $\delta(e) \leq \delta(d)$, as required. (Note that an equivalent definition of $\gcd(a, b)$, avoiding reference to δ or the Euclideanness of R, is

(a) d is a common divisor of a and b;

(b) if e is any common divisor of a and b then $e \mid d$.)

(2) $a_2 = a_0 - q_1 a_1$.

$a_3 = a_1 - q_2 a_2 = a_1 - q_2(a_0 - q_1 a_1) = -q_2 a_0 + (1 + q_2 q_1)a_1$.

$a_4 = a_2 - q_3 a_3 = (a_0 - q_1 a_1) - q_3(-q_2 a_0 + (1 + q_2 q_1)a_1) = (1 + q_3 q_2)a_0 - (q_1 + q_3 + q_3 q_2 q_1)a_1$,

etc. We find that each a_i, in particular $d = a_n$, is of the required form. (Recall that $a_0 = a, a_1 = b$.)

□

Note that I have been writing "a greatest common divisor" rather than "the greatest common divisor". This is because it is not quite unique, but we now show that the greatest common divisors of a and b form a single class of associates.

Lemma 3.4.3. *Let d be the* $\gcd(a, b)$ *produced by Theorem 3.4.2. Let d' be any other* $\gcd(a, b)$*. Then d and d' are associates, and $d' = k'a + \ell' b$ for some $k', \ell' \in R$.*

Proof. Letting $e = d'$ in the proof of Theorem 3.4.2, we see that $d' \mid d$, say $d = ud'$. If u is not a unit then, by Lemma 3.3.1 we have $\delta(d') < \delta(d)$, contrary to the maximality of $\delta(d')$. So u is a unit and d and d' are associates. From $d = ka + \ell b$ we get $d' = k'a + \ell' b$ by multiplying through by u^{-1} and putting $k' = u^{-1}k, \ell' = u^{-1}\ell$. □

It is not difficult to show that, conversely, any associate of a $\gcd(a, b)$ is also a $\gcd(a, b)$.

Example 3.4.4. $R = \mathbb{Z},\ a = 57, b = 15.$

$57 = 3 \times 15 + 12.$

$15 = 1 \times 12 + 3.$

$12 = 4 \times 3 + 0.$

So $d = 3 = 15 - 1 \times 12 = 15 - (57 - 3 \times 15) = (-1)(57) + 4(15).$

3.5. UNIQUE FACTORISATION IN EUCLIDEAN DOMAINS.

Proposition 3.5.1. *Let R be a Euclidean domain, $p \in R$ an irreducible element. If $p \mid ab$ in R then $p \mid a$ or $p \mid b$.*

Proof. Suppose p is irreducible and $p \mid ab$ but that $p \nmid a$. We must show that $p \mid b$. Since p is irreducible but $p \nmid a$, 1 must be a $\gcd(p, a)$, as in the proof of Theorem 3.6.8 below. (The only divisors of p are associates of 1 and associates of p, but p doesn't divide a.) Then $1 = kp + \ell a$ for some $k, \ell \in R$, by Lemma 3.4.3. Multiplying by b, $b = kpb + \ell ab$. The right-hand-side is a multiple of p, since ab is, hence $p \mid b$, as required. □

We saw in Theorem 3.3.2 that in a Euclidean domain, every non-zero, non-unit, is a product of irreducible elements. We might ask, is such an expression unique? In $\mathbb{Z}$ we have $30 = 2 \times 3 \times 5 = 3 \times 5 \times 2 = (-2) \times 3 \times (-5)$, but re-ordering the factors, or multiplying some factor by a unit and dividing another by the same unit (replacing them by associates), are somehow trivial changes. Is there a factorisation that is *really* different? If we look at $2 \times 5 \times 5 = 50$ and $3 \times 17 = 51$, we almost landed on the same number with two products of different irreducibles. It is not so difficult to imagine calculating such products and finding exactly the same number. But the following theorem shows that in fact this does not happen.

Theorem 3.5.2. *Let R be a Euclidean domain, $a \in R$ a non-zero, non-unit. If $a = p_1 p_2 \dots p_r = q_1 q_2 \dots q_s$, where the p_i and q_j are irreducibles, then $r = s$ and there exists a permutation $\sigma \in S_r$ such that q_j is associate to $p_{\sigma(j)}$ for all $1 \leq j \leq r$.*

Proof. Since $p_1 p_2 \dots p_r = q_1 q_2 \dots q_s$, we have $q_1 \mid p_1 p_2 \dots p_r$. If there is more than one p_i, we can repeatedly apply Proposition 3.5.1 (using the irreducibility of q_1) to show that $q_1 \mid p_i$ for some $1 \leq i \leq r$ (which we may call $\sigma(1)$). Since p_i is irreducible, its only divisors are units and associates of itself. Since q_1 is not a unit, it must be associate to p_i, say $q_1 = u p_i$ with u a unit. Cancelling p_i from both sides, $\prod_{t \neq i} p_t = (u q_2) q_3 \dots q_s$. Starting again and applying the same process repeatedly, each q_j can be matched up with an associate $p_{\sigma(j)}$, as required, as we exhaust all the factors on both sides, arriving necessarily at $1 = 1$. (Note that uq_2 is irreducible, just an associate of q_2.) □

Another way of saying Theorems 3.3.2 and 3.5.2 is that any Euclidean domain is a **unique factorisation domain**.

Example 3.5.3. *In* $\mathbb{C}[x]$, $x^4-1=(x-1)(x+1)(x-i)(x+i)$, *a product of irreducibles. We can modify this in trivial ways, e.g.* $x^4-1=(x-i)(x-1)(x+1)(x+i)$, *or* $x^4-1=(x-1)[3(x+1)][(1/3)(x-i)](x+i)$, *but the theorem tells us there is no essentially different factorisation into irreducibles. Actually in* $\mathbb{C}[x]$ *there is an easier way to see the uniqueness of the factorisation. If there was a different irreducible dividing* x^4-1, *say* $(x-\alpha)$ *with* $\alpha\notin\{\pm1,\pm i\}$ *then* $x^4-1=(x-\alpha)g(x)$ *for some* $g(x)\in\mathbb{C}[x]$. *Plugging in* $x=\alpha$ *shows that* $(\alpha-1)(\alpha+1)(\alpha-i)(\alpha+i)=0$, *which cannot be true, since in the integral domain* $\mathbb{C}$ *this product of non-zero factors cannot be* 0.

Example 3.5.4. *In* $\mathbb{Q}[x]$, x^3-2 *is irreducible.*

Proof. If not, there would be a linear factor in $\mathbb{Q}[x]$, hence a root $a/b\in\mathbb{Q}$.
$(a/b)^3=2 \implies a^3=2b^3$. If $a=p_1\dots p_r$ and $b=q_1\dots q_s$ as products of irreducibles, then $p_1^3\dots p_r^3=2q_1^3\dots q_s^3$. The exponent of 2 on the left is a multiple of 3, while on the right it is 1 more than a multiple of 3. So these are distinct factorisations into irreducibles of the same number, contradicting Theorem 3.5.2. □

In a similar manner one may show that the n^{th}-root of any positive integer that is not a perfect n^{th} power is irrational. This is a considerable strengthening of the theorem that $\sqrt{2}$ is irrational.

3.6. RINGS OF CONGRUENCE CLASSES IN EUCLIDEAN DOMAINS. .

Recall that if $a,b,m\in\mathbb{Z}$, we say that $a\equiv b \pmod m$ (a is **congruent** to b modulo m) if $m\mid(a-b)$. Equivalently, a and b belong to the same congruence class modulo m, or put another way, the cosets $a+m\mathbb{Z}$ and $b+m\mathbb{Z}$ are equal, i.e. $\overline{a}=\overline{b}$ in $\mathbb{Z}/m\mathbb{Z}$. We may extend this notion to any commutative ring R.

Definition 3.6.1. *Let* R *be any commutative ring. Given* $a,b,m\in R$, *we say* $a\equiv b \pmod m$ *if* $m\mid(a-b)$.

Example 3.6.2. *In* $\mathbb{R}[x]$, $x^6+3\equiv x^4+1 \pmod{x^2+1}$, *because* $(x^6+3)-(x^4+1)=x^6-x^4+2=(x^2+1)(x^4-2x^2+2)$.

Given a commutative ring R, and a modulus $m\in R$, we may turn the set of congruence classes mod m into a ring, using the rules $\overline{a}+\overline{b}=\overline{a+b}$ and $\overline{a}\,\overline{b}=\overline{ab}$. To justify this more carefully, we can use the following proposition, which you could get away with skipping if you prefer.

Proposition 3.6.3. *Let R be a commutative ring, and fix $m \in R$. Then $mR = Rm = \{rm \mid r \in R\}$ (the set of multiples of m in R) is an ideal of R (cf. end of §2.10).*

Proof. First we prove that Rm is an additive subgroup of R. For SG1, $0 = 0.m \in Rm$ (Proposition 2.2.1). For SG2, given $r_1m, r_2m \in Rm$, $r_1m + r_2m = (r_1 + r_2)m \in Rm$. For SG3, given $rm \in Rm$, $-(rm) = (-r)m \in Rm$ (Proposition 2.2.3). So Rm is an additive subgroup of R. It remains to prove that $R(Rm) \subseteq Rm$. But given $rm \in Rm$ and $s \in R$, $s(rm) = (sr)m \in Rm$. □

It follows that the quotient additive group R/mR, which is the set of congruence classes (mod m), is in fact a quotient ring in a natural way. (See the end of §2.10, and Example 2.10.2 for the example $\mathbb{Z}/3\mathbb{Z}$.) Sometimes when R is understood, we may write $\langle m \rangle$ in place of Rm.

Another way to look at Example 3.6.2: $x^2 + 1 \equiv 0 \pmod{x^2+1}$, so $x^2 \equiv -1 \pmod{x^2+1}$. Hence $x^6+3 \equiv (x^2)^3+3 \equiv (-1)^3+3 \equiv 2 \pmod{x^2+1}$, and also $x^4+1 \equiv (x^2)^2+1 \equiv (-1)^2+1 \equiv 2 \pmod{x^2+1}$.

A way of thinking of congruence in $\mathbb{Z}$, not yet mentioned, is that $a \equiv b \pmod{m}$ if a and b leave the same remainder when divided by m. This viewpoint carries over to any other Euclidean domain.

Proposition 3.6.4. *Let (R, δ) be a Euclidean domain, and non-zero $m \in R$. Any congruence class in R/mR may be represented by either 0 or some r with $\delta(r) < \delta(m)$.*

Proof. Given any $a \in R$, there exist $q, r \in R$ with $a = qm + r$ and either $r = 0$ or $\delta(r) < \delta(m)$. And $a - r = qm$, so $\overline{a} = \overline{r}$. □

Example 3.6.5. *Let $m = (x^2 + 1)$ in $\mathbb{R}[x]$. Then $x^6 + x + 3 = (x^2 + 1)(x^4 - x^2 + 1) + (x + 2)$, so $x^6 + x + 3 \equiv x + 2 \pmod{x^2+1}$, and $\overline{x^6 + x + 3} = \overline{x + 2}$ in $\mathbb{R}[x]/\langle x^2 + 1 \rangle$.*

We have to be a little careful about uniqueness, for example in $\mathbb{Z}/5\mathbb{Z}$ we have $\overline{3} = \overline{-2}$ with both $|3| < |5|$ and $|-2| < |5|$. For uniqueness of remainders representing congruence classes in $\mathbb{Z}/m\mathbb{Z}$, we need to take them non-negative, i.e. $\mathbb{Z}/m\mathbb{Z} = \{\overline{0}, \overline{1}, \ldots, \overline{m-1}\}$. In polynomial rings it is more straightforward.

Proposition 3.6.6. *Let k be a field, and $m(x)$ a fixed non-constant element of $k[x]$. Then the congruence classes $\overline{r(x)}$, as $r(x)$ runs over all elements of $k[x]$ with* $\deg(r) < deg(m)$ *(including $r = 0$), are all distinct, and are* all *the congruence classes modulo $m(x)$.*

Proof. That they are all the congruence classes is what the previous proposition says. We just need to show they are distinct. Suppose that $\deg(r_1) < \deg(m)$ and $\deg(r_2) < \deg(m)$, and that $\overline{r_1} = \overline{r_2}$. We must show that $r_1 = r_2$. But $r_1(x) \equiv r_2(x) \pmod{m(x)} \implies$ there

exists some $a(x) \in k[x]$ with $r_1(x) - r_2(x) = a(x)m(x)$. But $\deg(r_1 - r_2) < \deg(m)$, since $\deg(r_1), \deg(r_2) < \deg(m)$. This is compatible with $r_1 - r_2 = am$ only if $a = 0$, so $r_1 = r_2$, as required. □

Example 3.6.7. *The elements of* $\mathbb{R}[x]/\langle x^2+1\rangle$ *are all of the form* $\overline{a+bx}$ *with* $a, b \in \mathbb{R}$*, and these elements are all different, i.e.* $\overline{a+bx} = \overline{c+dx} \iff a = c$ *and* $b = d$*.*

Theorem 3.6.8. *Let* (R, δ) *be a Euclidean domain, and* $m \in R$ *an irreducible element. Then the quotient ring* $R/\langle m\rangle$ *is a field.*

Proof. We need to show that given any non-zero $\overline{a} \in R/\langle m\rangle$, there is some $\overline{b} \in R/\langle m\rangle$ with $\overline{a}\overline{b} = \overline{1}$, i.e. $ab \equiv 1 \pmod m$, i.e. $ab - 1 = cm$ for some $c \in R$, i.e. $ab + cm = 1$. We can use Lemma 3.4.3 to produce b and c, if we can argue that 1 is a $\gcd(a, m)$. But if d is a $\gcd(a, m)$ then $d \mid a$ and $d \mid m$. Now m is irreducible, so its only divisors are associates of m and associates of 1, so either m or 1 is a $\gcd(a, m)$. But $\overline{a}$ is given to be non-zero, i.e. $m \nmid a$, so m cannot be a $\gcd(a, m)$, and 1 must be. □

Example 3.6.9. *If* p *is any prime number then* $\mathbb{Z}/p\mathbb{Z}$ *is a field, which will also be denoted* $\mathbb{F}_p$*. For example,* $\mathbb{F}_3 = \mathbb{Z}/3\mathbb{Z}$ *is a field, with addition and multiplication tables as in Example2.10.2.*

Example 3.6.10. *Find* $\overline{5}^{-1}$ *in* $\mathbb{F}_{17}$*.*
$17 = 3 \times 5 + 2$*;*
$5 = 2 \times 2 + 1$*.*
$1 = 5 - 2 \times 2 = 5 - (2(17 - 3 \times 5) = 7 \times 5 - 2 \times 17$*.*
So $7 \times 5 \equiv 1 \pmod{17}$*, i.e.* $\overline{7} = \overline{5}^{-1}$ *in* $\mathbb{F}_{17}$*.*

Example 3.6.11. *In* $\mathbb{R}[x]$*, if* $x^2 + 1$ *were reducible, it would have to have a linear factor in* $\mathbb{R}[x]$*, hence a root in* $\mathbb{R}$*, but it has no root in* $\mathbb{R}$*, hence it must be irreducible. Since* $\mathbb{R}$ *is a field,* $\mathbb{R}[x]$ *is a Euclidean domain, hence, since* $x^2 + 1$ *is irreducible, the quotient ring* $\mathbb{R}[x]/\langle x^2+1\rangle$ *is a field. By Proposition 3.6.6,* $\mathbb{R}[x]/\langle x^2+1\rangle = \{[a + bx] \mid \ a, b \in \mathbb{R}\}$*. (Recall that* $[a + bx]$ *means the same as* $\overline{a+bx}$*.) We have*

$$[a + bx] + [c + dx] = [(a + bx) + (c + dx)] = [(a + c) + (b + d)x],$$

and

$$[a + bx][c + dx] = [(a + bx)(c + dx)] = [ac + (ad + bc)x + bdx^2] = [(ac - bd) + (ad + bc)x],$$

since $x^2 \equiv -1 \pmod{x^2+1}$*. This is just like adding and multiplying complex numbers, in fact* $\mathrm{ev}_i : \mathbb{R}[x] \to \mathbb{C}$ *(cf. Example2.9.8) is a ring homomorphism with image* $\mathbb{C}$ *and kernel* $\langle x^2+1\rangle$*, so the First Isomorphism Theorem for rings gives an isomorphism* $\mathbb{R}[x]/\langle x^2+1\rangle \simeq \mathbb{C}$*, with*

$[a+bx] \mapsto \mathrm{ev}_i(a+bx) = a+bi$. *In fact if we didn't already "know"* $\mathbb{C}$*, the quotient ring* $\mathbb{R}[x]/\langle x^2+1\rangle$ *gives a construction of something we can call* $\mathbb{C}$*, with* $[x]$ *providing a square root of* -1*, i.e. we can call* $[x]$ *"*i*". Note that* $a \mapsto [a]$ *gives an isomorphism from* $\mathbb{R}$ *to a subfield of this version of* $\mathbb{C}$*.*

Example 3.6.12. $\mathbb{F}_2 = \{\overline{0}, \overline{1}\}$ *is a field, so* $\mathbb{F}_2[x]$ *is a Euclidean domain. The element* $x^2+x+\overline{1}$ *is irreducible in* $\mathbb{F}_2[x]$*, because neither* $\overline{0}$ *nor* $\overline{1}$ *is a root. Hence* $\mathbb{F}_2[x]/\langle x^2+x+\overline{1}\rangle$ *is a field. By Proposition 3.6.6,* $\mathbb{F}_2[x]/\langle x^2+x+\overline{1}\rangle = \{[a+bx] \mid \ a, b \in \mathbb{F}_2\}$*. There are* 2 *choices for each of* a *and* b*, so this field contains* 4 *elements, and we may call it "*$\mathbb{F}_4$*". (Note that this is not isomorphic to* $\mathbb{Z}/4\mathbb{Z}$*, which is not a field, or even an integral domain.) Just as in the previous example we thought of* $\mathbb{R}$ *as a subfield of* $\mathbb{C}$*, here we can think of* $\mathbb{F}_2$ *as a subfield of* $\mathbb{F}_4$ *via the injective homomorphism* $a \mapsto [a]$*. So we have started from* $\mathbb{F}_2$ *and extended it to a bigger field* $\mathbb{F}_4$*, a bit like we started with* $\mathbb{R}$ *and extended it to the bigger field* $\mathbb{C}$*. In the latter case, we created a square root of* -1 *as* $[x]$*. What have we created here? Let* $\alpha := [x] \in \mathbb{F}_4$*. Then* $\alpha^2+\alpha+[\overline{1}] = [x]^2+[x]+[\overline{1}] = [x^2+x+\overline{1}] = [\overline{0}]$*. So we have created a root of* $x^2+x+\overline{1}$ *in* $\mathbb{F}_4$*, where there wasn't one in* $\mathbb{F}_2$*. Being a bit looser with the notation, we can write* $\mathbb{F}_4 = \{0, 1, \alpha, 1+\alpha\}$*, with* $\alpha^2+\alpha+1 = 0$*.*

The binary digits on which digital technology is based can be thought of as the elements of $\mathbb{F}_2$. Fields such as $\mathbb{F}_4$, extending $\mathbb{F}_2$ like $\mathbb{C}$ extends $\mathbb{R}$, and quotients of polynomial rings with coefficients in such fields, actually have applications to things used by millions everyday, without any awareness of the underlying mathematics. Google "QR codes", click on Wikipedia, then on "Reed-Solomon Error Correction", then on "3.2: The BCH view", for applications of polynomials with coefficients in finite fields. Don't expect to follow it, but you will spot some familiar key words to get an impression that this stuff is applicable to something important.

3.7. SQUARE ROOTS OF -1 IN $\mathbb{F}_p$ (proofs optional, statement of Proposition 3.7.3 used in next section). .

We all know that in $\mathbb{R}$ there is no square root of -1, so we have to invent one and create $\mathbb{C}$. Now we have all these new fields $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$, where p is any prime number. We may ask whether or not $-\overline{1}$ is a square in $\mathbb{F}_p$. For example $\overline{2}^2 = -\overline{1}$ in $\mathbb{F}_5$, but in $\mathbb{F}_3$ the only squares are $\overline{0}^2 = \overline{0}$ and $(\pm\overline{1})^2 = \overline{1}$, so $-\overline{1} = \overline{2}$ is not a square. In $\mathbb{F}_2$, $\overline{1}^2 = \overline{1} = -\overline{1}$, so we only need to consider odd p, which must be congruent to either 1 or 3 (mod 4).

Proposition 3.7.1. *If a prime* $p \equiv 3 \pmod 4$ *then* $-\overline{1}$ *is not a square in* $\mathbb{F}_p$*.*

Proof. If $x^2 \equiv -1 \pmod p$ then $(-1)^{(p-1)/2} \equiv (x^2)^{(p-1)/2} \equiv x^{p-1} \equiv 1 \pmod p$, since $p-1$ is the order of the multiplicative group $\mathbb{F}_p^\times$, and the order of an element divides the order of the

group. But since $p \equiv 3 \pmod 4$, $(p-1)/2$ is odd, so $(-1)^{(p-1)/2} = -1 \not\equiv 1 \pmod p$ (since $p \neq 2$). Contradiction. Hence there is no such x. □

Lemma 3.7.2 (Wilson's Theorem)**.** *Let p be any prime number. Then $(p-1)! \equiv -1 \pmod p$.*

Proof. In $\mathbb{F}_p$, the factors $\overline{1}, \overline{2}, \ldots, \overline{p-2}, \overline{p-1}$ come in inverse pairs that cancel, except for $\overline{1}$ and $\overline{-1}$, which are self-inverse. So the product is $\overline{-1}$. □

Proposition 3.7.3. *If a prime $p \equiv 1 \pmod 4$ then $-\overline{1}$ is a square in $\mathbb{F}_p$.*

Proof. Let $a = \left(\frac{p-1}{2}\right)!$. Then
$a^2 = a(-1)^{(p-1)/2}a$ (since $p \equiv 1 \pmod 4 \implies (p-1)/2$ is even)
$= 1 \cdot 2 \cdot \ldots \cdot \left(\frac{p-1}{2}\right) \cdot \left[-\left(\frac{p-1}{2}\right)\right] \cdot \left[-\left(\frac{p-3}{2}\right)\right] \cdot \ldots \cdot [-2] \cdot [-1]$
$\equiv 1 \cdot 2 \cdot \ldots \cdot (p-2)(p-1) \pmod p$
$\equiv (p-1)! \equiv -1 \pmod p$, by Wilson's Theorem. Hence $\overline{a}^2 = -\overline{1}$ in $\mathbb{F}_p$. □

Example 3.7.4. *In $\mathbb{F}_{41}$, $\overline{9}^2 = -\overline{1}$.*

3.8. THE GAUSSIAN INTEGERS AND THE TWO-SQUARE THEOREM. .

Recall that $\mathbb{Z}[i] = \{a + bi : \ a, b \in \mathbb{Z}\}$, that this is a Euclidean domain (Example 3.2.5) with $\delta(\alpha) = \alpha\overline{\alpha}$ (hence a unique factorisation domain), and that $U(\mathbb{Z}[i]) = \{\pm 1, \pm i\}$ (Proposition 2.8.3).

Some factorisations in $\mathbb{Z}[i]$:

(1) Though 2 is irreducible in $\mathbb{Z}$, in $\mathbb{Z}[i]$ we have $2 = (1+i)(1-i) = (1+i)[-i(1+i)] = -i(1+i)^2$.
$-i$ is a unit. Is $(1+i)$ irreducible?

(2) $1+9i = (5+4i)(1+i) = (-4+5i)(1-i) = (-5-4i)(-1-i) = (4-5i)(-1+i)$. These are equivalent factorisations, the first factor always an associate of $(5+4i)$, the second factor always an associate of $(1+i)$. Is $(5+4i)$ also irreducible?

The answer to both questions is "Yes", by Lemma 3.3.8: $\delta(1+i) = 1^2 + 1^2 = 2$, which is prime, so $1+i$ is irreducible, and $\delta(5+4i) = 5^2 + 4^2 = 41$, which is prime, so $5+4i$ is irreducible.

It is a fact that we get all irreducibles in $\mathbb{Z}[i]$ by factoring prime numbers. (This is a consequence of Proposition 3.5.1.) We have already seen that $2 = -i(1+i)^2$, with $1+i$ irreducible. Now we consider odd primes, which must be congruent to either 1 or 3 (mod 4).

Proposition 3.8.1. *If $p \equiv 3 \pmod 4$ is a prime number, then p stays irreducible in $\mathbb{Z}[i]$.*

Proof. If $p = \alpha\beta$, with neither α nor β a unit, then taking norms, $\delta(p) = \delta(\alpha)\delta(\beta)$ (cf. third line of proof of Example 3.2.5), with neither $\delta(\alpha) = 1$ nor $\delta(\beta) = 1$. (If $\beta\overline{\beta} = 1$ then β is a

unit with inverse $\overline{\beta}$.) But $\delta(p) = p^2$, so this implies that $\delta(\alpha) = \delta(\beta) = p$. (Alternatively use Lemma 3.3.1.) If $\alpha = a + bi$ then $a^2 + b^2 = \delta(\alpha) = p$. But modulo 4 we have $0^2 \equiv 2^2 \equiv 0$ and $1^2 \equiv 3^2 \equiv 1$, i.e. the only squares (mod 4) are 0 and 1. So $a^2 + b^2 \equiv 0{,}1$ or $2 \pmod 4$, contrary to $p \equiv 3 \pmod 4$. Hence α must be irreducible. □

Theorem 3.8.2. *If $p \equiv 1 \pmod 4$ is a prime number, then in $\mathbb{Z}[i]$ we have $p = \alpha\overline{\alpha}$, a product of two irreducibles.*

Proof. By Proposition 3.7.3, there exists $q \in \mathbb{Z}$ with $q^2 \equiv -1 \pmod p$, i.e. $p \mid (q^2 + 1)$. In $\mathbb{Z}[i]$ this gives $p \mid (q + i)(q - i)$. If p were irreducible then by Proposition 3.5.1 either $p \mid (q + i)$ or $p \mid (q - i)$, neither of which is true, since a multiple of p would be of the form $p(c+di) = pc+pdi$, with imaginary part divisible by p, which is not true of $q \pm i$. Hence p is not irreducible in $\mathbb{Z}[i]$, say $p = \alpha\beta$ with neither α nor β a unit. As above, this implies that $\delta(\alpha) = \delta(\beta) = p$. Hence $\alpha\overline{\alpha} = p$, and by Lemma 3.3.8 these factors are irreducible. □

The following consequence was stated by Fermat on 25th December 1640, proved by Euler in 1754.

Theorem 3.8.3 (Two square theorem)**.** *If $p \equiv 1 \pmod 4$ is a prime number then p is a sum of two squares, in a unique way.*

Proof. If $p = \alpha\overline{\alpha}$ with $\alpha = a + bi$, then $p = a^2 + b^2$. If also $p = c^2 + d^2 = (c + di)(c - di)$ then by Theorem 3.5.2, $(c + di)$ is either $\pm(a + bi)$, $\pm i(a + bi)$, $\pm(a - bi)$ or $\pm i(a - bi)$. In all cases $\{c^2, d^2\} = \{a^2, b^2\}$. □

Example 3.8.4. $5 = 1^2 + 2^2$, $13 = 2^2 + 3^2$, $17 = 1^2 + 4^2$, $29 = 2^2 + 5^2$, $41 = 4^2 + 5^2$, $113 = 7^2 + 8^2$.