M2PM2: ALGEBRA 2

LECTURES BY PROF. KEVIN BUZZARD; NOTES BY ALEKSANDER HORAWA

These are notes from the course M2PM2: Algebra 2 at Imperial College London taught by Professor Kevin Buzzard in Autumn 2014. They were LaTeX'd by Aleksander Horawa.

If you find a mistake, please let me know at aleksander.horawa13@ic.ac.uk. Please check if a new version is available at https://sites.google.com/site/aleksanderhorawa/.

Contents

1. More examples of groups

The examples will be symmetry groups of objects in 2 or 3 dimensions. Let us start with an example in 2-d.

Consider the distance structure on $\mathbb{R}^2$, i.e. the distance function $d\colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}_{\geq 0} = \{x \in \mathbb{R}\colon x \geq 0\}$ that assigns to two vectors the distance between them. For $\mathbf{x} = (x_1, x_2)$, $\mathbf{y} = (y_1, y_2)$, we have

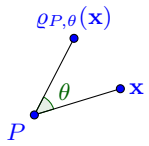
$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}.$$

An *isometry* of $\mathbb{R}^2$ is a bijection $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that

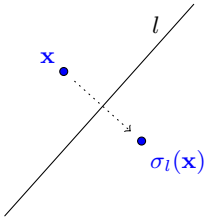
$$d(f(\mathbf{x}), f(\mathbf{y})) = d(\mathbf{x}, \mathbf{y})$$

for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$. In other words, an isometry is a distance-preserving bijection.

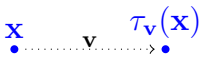
Examples. (1) Rotations. If $P \in \mathbb{R}^2$ and θ is an angle, then define $\varrho_{P,\theta}$ to be the rotation through angle θ anticlockwise with center P.



(2) Reflections. If $l \subseteq \mathbb{R}^2$ is a line, then let σ_l be the reflection about the line l.



(3) Translations. If $\mathbf{v} \in \mathbb{R}^2$, then define $\tau_{\mathbf{v}}$ by $\tau_{\mathbf{v}}(\mathbf{x}) = \mathbf{x} + \mathbf{v}$.



δ

Remark. There are isometries that are not of type (1), (2), or (3), for example a glide-reflection $f = \sigma_l \circ \tau_{\mathbf{v}}$, $\mathbf{v}$ a vector parallel to line l. On the other hand, every isometry can be built by composing reflections, rotations, and translations.

(Tricky exercise. Step 1: an isometry that fixes 2 points, fixes the line also.
Step 2: an isometry that fixes 2 points must be the identity or a reflection.
Step 3: an isometry that fixes 3 non-collinear points is the identity.
Step 4: any isometry is determined by its value on 3 non-collinear points.)

Definition. Let $I(\mathbb{R}^2)$ be the set of all isometries of $\mathbb{R}^2$.

Proposition 1.1. *The set $I(\mathbb{R}^2)$ under the composition of functions is a group.*

Proof. (Closure) Say $f, g \in I(\mathbb{R}^2)$. The composite of two bijections is a bijection, so $f \circ g$ is a bijection. Furthermore, if $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$, then

$$\begin{aligned} d((f \circ g)(\mathbf{x}), (f \circ g)(\mathbf{y})) &= d(f(g(\mathbf{x})), f(g(\mathbf{y}))) && \text{definition of } f \circ g \\ &= d(g(\mathbf{x}), g(\mathbf{y})) && f \text{ is an isometry} \\ &= d(\mathbf{x}, \mathbf{y}) && g \text{ is an isometry} \end{aligned}$$

(Associativity) Composition of functions is always associative.

(Identity) Define $e\colon \mathbb{R}^2 \to \mathbb{R}^2$ by $e(\mathbf{x}) = \mathbf{x}$. Then for all $f \in I(\mathbb{R}^2)$

$$f \circ e = f = e \circ f$$

and $e \in I(\mathbb{R}^2)$ (easy).

(Inverses) Say $f \in I(\mathbb{R}^2)$. Then f is a bijection, so there exists an inverse function $f^{-1}\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that $f \circ f^{-1} = f^{-1} \circ f = e$. Need to check that $f^{-1} \in I(\mathbb{R}^2)$. It is a bijection—does it preserve lengths?

Take any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$. Then

$$d(f^{-1}(\mathbf{x}), f^{-1}(\mathbf{y})) = d(f(f^{-1}(\mathbf{x})), f(f^{-1})(\mathbf{y})) = d(\mathbf{x}, \mathbf{y}),$$

since f is an isometry, and thus $f^{-1} \in I(\mathbb{R}^2)$. Therefore, inverses exist. □

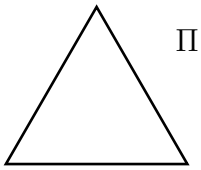
Of course, $I(\mathbb{R}^2)$ is hugely infinite. Therefore, we will consider subgroups of isometries which preserve particular subsets of $\mathbb{R}^2$.

Say $\Pi \subseteq \mathbb{R}^2$ is a subset. For $g \in I(\mathbb{R}^2)$, define

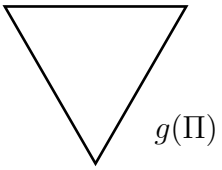
$$g(\Pi) := \{g(x) \mid x \in \Pi\}.$$

(Abuse of notation. In computing, function overloading.)

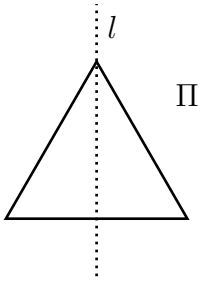
Example. Let Π be an equilateral triangle centred at the origin pointing up:



Let $g = \varrho_{0,\pi}$ be the rotation by π centred at 0. Then $g(\Pi)$ is the triangle upside down:



On the other hand, if l is the y-axis and σ_l the reflection about l, then $\sigma_l(\Pi) = \Pi$



Definition. Say $\Pi \subseteq \mathbb{R}^2$. Then the *symmetry group of* Π is

$$G(\Pi) := \{g \in I(\mathbb{R}^2) \mid g(\Pi) = \Pi\} = \{\text{the isometries sending } \Pi \text{ to itself}\}.$$

Example. If Π is the triangle above, then

$$G(\Pi) = \{e, \sigma_l, \sigma_{0,2\pi/3}, \ldots\}.$$

Proposition 1.2. *The set $G(\Pi)$ is a subgroup of $I(\mathbb{R}^2)$.*

Proof. (1) If e is the identity map, then $e(\Pi) = \Pi$ by definition, so $e \in G(\Pi)$.

(2) If $x, y \in G(\Pi)$, then

$$\begin{array}{lll} (x \circ y)(\Pi) & = x(y(\Pi)) & \\ & = x(\Pi) & \text{since } y \in G(\Pi) \\ & = \Pi & \text{since } x \in G(\Pi) \end{array}$$

so $x \circ y \in G(\Pi)$.

(3) If $x \in G(\Pi)$, then

$$x \in G(\Pi) \Rightarrow x(\Pi) = \Pi \Rightarrow x^{-1}(x(\Pi)) = x^{-1}(\Pi) \Rightarrow \Pi = x^{-1}\ (\Pi),$$

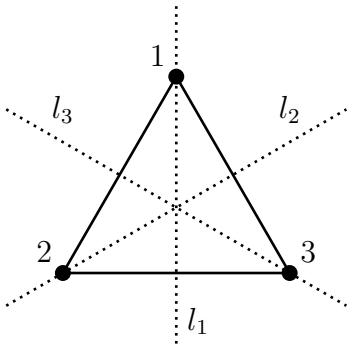
so $x^{-1} \in G(\Pi)$. □

Strategy to work out $G(\Pi)$:

- Write down all the elements we can think of.
- Prove there are no more.

Examples.

(1) Equilateral triangle, centred at the origin.



Note that $G(\Pi)$ contains (at least): 3 rotations: e, $\varrho = \varrho_{0,\pi/3}$, ϱ^2, 3 reflections: σ_1, σ_2, σ_3 about (respectively) l_1, l_2, l_3.
Next, note that each symmetry of Π sends corners to corners, so it induces a permutation of the corners. If we label the corners by 1, 2, 3 (anticlockwise), then

$$\begin{array}{lll} e & \to & e \\ \varrho & \to & (123) \\ \varrho & \to & (132) \\ \sigma_1 & \to & (23) \\ \sigma_2 & \to & (13) \\ \sigma_3 & \to & (12) \end{array}$$

Now say $g \in G(\Pi)$ is any element. Then g permutes the corners, but all 6 permutations of corners are listed above, so g has to agree with one of our elements on corners. Thus g must be one of the elements we already thought of (an isometry is determined by what it does to 3 non-collinear points). Therefore

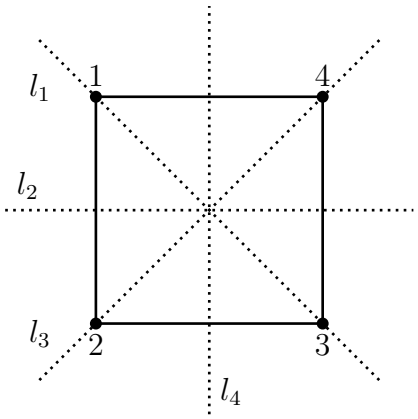
$$G(\Pi) = \{e, \varrho, \varrho^2, \sigma_1, \sigma_2, \sigma_3\}.$$

To compute the group law, work in S_3 instead, for example

$$\varrho\sigma_1 \to (123)(23) = (12) \leftarrow \sigma_3.$$

We will see later that $G(\Pi)$ is 'isomorphic' to S_3.

(2) Square, centred at the origin.



We note that $G = G(\Pi)$ contains 4 reflections (σ_i about l_i for $i = 1, 2, 3, 4$) and 4 rotations e, $\varrho = \varrho_{0,\pi/2}$, ϱ^2, ϱ^3. Therefore, $|G| \geq 8$.

We claim that $|G| = 8$. To show that, take $g \in G$. Since g preserves distance, it also preserves the corners. Furthermore, g sends opposite corners to opposite corners and adjacent corners to adjacent corners. Say the corners are labeled 1, 2, 3, 4 (as above) and suppose $g(1) = i$, $1 \leq i \leq 4$. Then $g(2)$ is a neighbour of i, i.e. "$i + 1$ or $i - 1$" (where $4 + 1 = 1$, $1 - 1 = 4$), call it j. Then:

$$g(1) = i,$$
$$g(2) = j,$$
$$g(3) = \text{corner opposite } g(1) = i,$$
$$g(4) = \text{corner opposite } g(2) = j,$$

and g is determined by what it does to corners. Total number of possibilities for g is hence at most

$$(\text{no. of } i) \times (\text{no. of } j) = 4 \times 2 = 8,$$

and $|G| \leq 8$.

But we already found 8 elements of G, so $|G| \geq 8$, and thus $|G| = 8$. So we have shown

$$G = \{e, \varrho, \varrho^2, \varrho^3, \sigma_1, \ldots, \sigma_4\}.$$

This group is called D_8, the dihedral group of order 8.

To work out multiplication in D_8, let us look at what our elements do to corners

$$\begin{array}{lcl} e & \to & e \\ \varrho & \to & (1234) \\ \varrho^2 & \to & (13)(24) \\ \varrho^3 & \to & (1432) \\ \sigma_1 & \to & (24) \\ \sigma_2 & \to & (12)(34) \\ \sigma_3 & \to & (13) \\ \sigma_4 & \to & (14)(23) \end{array}$$

We can use this to multiply elements, for example

$$\sigma_2 \circ \sigma_3 = (12)(34)(13)$$

which sends 1 to 4, 4 to 3, 3 to 2, 2 to 1, and thus

$$\sigma_2 \circ \sigma_3 = (1432) = \varrho_3.$$

Let $H \subseteq D_8$ be the cyclic subgroup generated by ϱ. Then

$$H = \{e, \varrho, \varrho^2, \varrho^3\}.$$

Let $\sigma = \sigma_1$. Then $\sigma \notin H$, so the right coset $H\sigma$ of H is not H. Therefore, $H \cap H\sigma = \emptyset$ and $|H| = |H\sigma| = 4$ which shows

$$H\sigma = \{\sigma_1, \sigma_2, \sigma_3, \sigma_4\}.$$

Finally, we can write

$$D_8 = H \cup H\sigma = \{e, \varrho, \varrho^2, \varrho^3, \sigma, \varrho\sigma, \varrho^2\sigma, \varrho^3\sigma\},$$

a better notation for the elements of D_8.

What equations do σ and ϱ satisfy?
Easy ones: $\varrho^4 = e$ and $\sigma^2 = e$. What is $\sigma\varrho$?

$$\sigma\varrho = (24)(1234) = (14)(23) = \varrho^?\sigma$$

for some ?. Let us try $\varrho^3\sigma$:

$$\varrho^3\sigma = (1432)(24) = (14)(23).$$

Finally, we get the equation

$$\sigma\varrho = \varrho^3\sigma = \varrho^{-1}\sigma.$$

We now claim that these are the only 3 equations we need to figure out any product of 2 elements of D_8.

For example, consider $(\varrho\sigma)(\varrho^2\sigma)$. We have

$$\begin{aligned} \varrho\sigma\varrho^2\sigma &= \varrho(\sigma\varrho)\varrho\sigma \\ &= \varrho(\varrho^{-1}\sigma)\varrho\sigma \\ &= \varrho\varrho^{-1}\ (\sigma\varrho)\sigma \\ &= (\varrho^{-1}\sigma)\sigma \\ &= \varrho^{-1} \\ &= \varrho^3 \end{aligned}$$

This will work in general, because we have a method of getting the ϱs to the left of the σs. Hence D_8 is generated by ϱ and σ subject to the relations $\varrho^4 = \sigma^2 = e$, $\sigma\varrho = \varrho^{-1}\sigma$.

(3) Regular polygons. Let Π be a regular n-gon with $n \geq 3$ and $G = G(\Pi)$. Then G contains n rotations

$$e, \varrho = \varrho_{2\pi/n}, \varrho^2, \ldots, \varrho^{n-1}$$

and n reflections

$$\sigma = \sigma_1, \sigma_2, \ldots, \sigma_n.$$

Hence $|G| \geq 2n$. We claim that $|G| \leq 2n$. If $g \in G$, then g sends corners to corners. Say $g(1) = i$, n choices. Then $g(2) = j$ has to be a corner neighbouring i ("$j = i \pm 1$"), 2 choices for j. Now $g(3)$ is a neighbour of j that is not i, 1 choice, etc. This means

$$(\text{no. of possibilities for } g) \leq \big(\text{ no. of choices for } (i,j)\big) = 2n,$$

so $|G| \leq 2n$. Hence $G = |2n|$.

Definition. The *dihedral group* D_{2n} is the group of symmetries of a regular n-gon.

Part (3) of the example showed that $|D_{2n}| = 2n$ and

$$D_{2n} = \{e, \varrho, \varrho^2, \ldots, \varrho^{n-1}, \sigma_1, \sigma_2, \ldots, \sigma_n\}.$$

Same trick as for D_8: $\sigma = \sigma_1$, $H = \langle \varrho \rangle$. Then $H\sigma \neq H$ and we check that

$$D_{2n} = \{e, \varrho, \varrho^2, \ldots, \varrho^{n-1}, \sigma, \varrho\sigma, \varrho^2\sigma, \ldots, \varrho^{n-1}\sigma\}.$$

Clearly, $\varrho^n = e$, $\sigma^2 = e$, and one checks that $\sigma\varrho = \varrho^{-1}\sigma$. This is on Problem Sheet 1, but three ways of doing it are:

(1) Think about permuting corners.
(2) Since ϱ, σ are linear maps, we can check using matrices.

$$\varrho_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

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

(3) Pretend $\mathbb{R}^2 = \mathbb{C}$ and use complex numbers:

$$\varrho(z) = e^{2\pi i/n} z,$$

$$\sigma(z) = \overline{z}.$$

Because $\sigma\varrho = \varrho^{-1}\sigma$, we can "move σ's to the right of ϱ's" and compute the multiplication table of D_{2n}.

Example. Let Π be the string

$$\ldots\ R\ R\ R\ R\ R\ R\ R\ R\ R\ \ldots \subseteq \mathbb{R}^2.$$

What is $G(\Pi)$? There is a translation τ sending each "R" to the next one. So

$$G(\Pi) \supseteq \langle \tau \rangle = \{\tau^n \mid n \in \mathbb{Z}\},$$

an infinite cyclic group. Note that τ^n sends the j^{th} R to the $(j+n)^{th}$ R.

We claim that $G(\Pi) = \langle \tau \rangle$. To see that, take $g \in G = G(\Pi)$ and suppose g sends $R_0 = 0^{th}R$ to $R_n = n^{th}R$. Then $\tau^{-n}g \in G$ and it sends R_0 to itself. But "R" has no symmetries, so $\tau^{-n}g = e$, proving $G = \langle \tau \rangle$.

Example. The above all works in $\mathbb{R}^n$. Example in $\mathbb{R}^3$: set $I(\mathbb{R}^3)$ = isometries of $\mathbb{R}^3$ and Π = regular tetrahedron (or any platonic solid). A symmetry of Π preserves the corners and 4 corners of Π will determine the isometry. Can check $G(\Pi) \cong S^4$, any permutation of corners works. This is a tricky exercise.

2. Isomorphisms

Recall that a map preserving the structure of vector spaces was simply a linear map. In this section, we will see maps between groups, but they will be a special kind of map—an isomorphism.

Basic idea: the symmetry group of the equilateral triangle was somehow 'equal' to S_3 without actually being S_3. The notion of isomorphism will formalise this.

Consider the following two groups:

- $G = C_3 = \{x \in \mathbb{C} \mid x^3 = 1\} = \{1, \omega, \omega^2\}$ with $\omega = e^{2\pi i/3}$,
- $H =$ subgroup of S_3 generated by a 3-cycle $a = (123)$, i.e. $H = \langle a \rangle = \{e, a, a^2\}$.

Multiplication tables.

G	1	ω	ω^2
1	1	ω	ω^2
ω	ω	ω^2	1
ω^2	ω^2	1	ω

H	e	a	a^2
e	1	a	a^2
a	a	a^2	1
a^2	a^2	1	a

The multiplication tables are the same, except that the elements have different labels. Formally, there is a bijection $G \to H$ such that if $g_1 \mapsto h_1$ and $g_2 \mapsto h_2$, then $g_1 g_2 \mapsto h_1 h_2$.

Definition. Let G and H be groups. A function $f\colon G \to H$ is an *isomorphism* if

(1) f is a bijection,
(2) for all $x, y \in G$, $f(xy) = f(x)f(y)$.

Notation. If there exists an isomorphism $G \to H$, then we write $G \cong H$, and say G is isomorphic to H.

Remark. Group isomorphism is an equivalence relation; indeed, we have the following:

- $G \cong G$;
- if $G \cong H$, then $H \cong G$;
- if $G \cong H$ and $H \cong K$, then $G \cong K$.

We saw an example of this in Section 1: $D_6 \cong S_3$.

Question. Given 2 groups G and H, how can we tell if they are isomorphic?

Example. Let $G_1 = C_4 = \{\pm 1, \pm i\}$;
$G_2 =$ symmetry group of a non-square rectangle $= \{e, \varrho_\pi, \sigma_1, \sigma_2\}$;
$G_3 = \langle \varrho \rangle =$ cyclic subgroup of D_8 generated by $\varrho = \varrho_{\pi/2}$, $G_3 = \{e, \varrho, \varrho^2, \varrho^3\}$.

Which groups in that list are isomorphic?

First of all, we show that $G_1 \cong G_3$. Define $f\colon G_1 \to G_3$ by

$$\begin{array}{rcl} 1 & \mapsto & e \\ i & \mapsto & \varrho \\ -1 & \mapsto & \varrho^2 \\ -i & \mapsto & \varrho^3 \end{array}$$

i.e. $i^n \mapsto \varrho^n$ for all $n \in \{0,1,2,3\}$ or all $n \in \mathbb{Z}$. It is a bijection and

$$\begin{array}{rll} f(i^m i^n) & = f(i^{m+n}) & \\ & = \varrho^{m+n} & \text{definition of } f \\ & = \varrho^m \varrho^n & \\ & = f(i^m) f(i^n) & \text{definition of } f \end{array}$$

We now claim that $G_1 \ncong G_2$ (even though both have size 4). We prove this by contradiction. Suppose there exists an isomorphism $f\colon G_1 \to G_2$. Every element $x \in G_2$ satisfies $x^2 = e$ (apart from e, all elements have order 2). Say

$$f(i) = x \in G_2$$

$$f(1) = y \in G_2$$

Then

$$\begin{array}{rll} f(-1) & = f(i^2) & \\ & = f(i)^2 & f \text{ is an isomorphism} \\ & = x^2 & \\ & = e & \end{array}$$

and

$$\begin{array}{rll} f(1) & = f(1^2) & \\ & = f(1)^2 & f \text{ is an isomorphism} \\ & = y^2 & \\ & = e & \end{array}$$

Therefore, $f(-1) = f(1)$, so f is not a bijection, a contradiction. Hence $G_1 \ncong G_2$.

Then also $G_3 \ncong G_2$, as $G_3 \cong G_1$ and $\cong$ is an equivalence relation.

General strategy for isomorphism questions:

- If you think 2 groups are isomorphic, try and find an isomorphism.
- If you think they are not, try and find a "group-theoretic thing" that one group has and the other does not.

Here are some examples.

Proposition 2.1. *Let G, H be groups.*

(1) *If $|G| \neq |H|$, then $G \ncong H$.*
(2) *If G is abelian and H is not, then $G \ncong H$.*
(3) *Suppose there exists $k \in \mathbb{Z}_{\geq 1}$ such that G and H have different numbers of elements of order k. Then $G \ncong H$.*

However, the converse of the proposition does not hold. Note that (1)–(3) are just examples: we cannot use them to prove that 2 groups are isomorphic.

We precede the proof of Proposition 2.1 with a demonstration of its use.

Examples.

(1) $G = D_{10}$, $H = S_5$. Then $|G| = 10$ and $|H| = 5! = 120$, so $G \ncong H$.
(2) $G = S_4$, $H = C_{24}$ = cyclic group of order 24. Then $|G| = |H| = 24$, but G is not abelian and H is, so $G \ncong H$.
(3) $G = C_4$, $H =$ symmetric group of a non-square rectangle. Then G has 1 element of order 2, but H has 3 elements of order 2, so $G \ncong H$.
(4) $G = (\mathbb{R}, +)$, $H = (\mathbb{R}^\times, \times)$. Then $(\mathbb{R}, +)$ has no element of order 2, but $(\mathbb{R}^\times, \times)$ has one, namely -1, so $G \ncong H$.

To prove Proposition 2.1, we will need the following lemma.

Lemma 2.2.

(1) *If $f\colon G \to H$ is an isomorphism, then $f(e_G) = e_H$.*
(2) *If $f\colon G \to H$ is an isomorphism and $g \in G$ has order n, then $f(g)$ has order n.*

Proof. For (1), let $f(e_G) = h \in H$. Then

$$\begin{aligned} h &= f(e_G) && \text{definition} \\ &= f(e_G \times e_G) \\ &= f(e_G)^2 && f \text{ isomorphism} \\ &= h^2 \end{aligned}$$

so $h = h^2$ and multiplying both sides by h^{-1} we get $h = e_H$.

For (2), assume $f\colon G \to H$ be an isomorphism. By (1), if $g = e_G$, then $f(g) = e_H$. Because f is a bijection, we can deduce that if $g \neq e_G$, then $f(g) \neq e_H$. So say g has order $n > 1$. Then

$$g \neq e_G, g^2 \neq e_G, \ldots, g^{n-1} \neq e_G$$

but $g^n = e_G$, so taking images of these under f,

$$f(g) \neq e_H, f(g)^2 = f(g^2) \neq e_H, \ldots, f(g)^{n-1} = f(g^{n-1}) \neq e_H,$$

but $f(g)^n = f(g^n) = f(e_G) = e_H$ by (1). So $f(g)$ has order exactly n. □

We are now ready to prove Proposition 2.1.

Proof of Proposition 2.1. All the proofs go by contradiction, i.e. suppose that $G \cong H$ and $f\colon G \to H$ is an isomorphism.

Since $f\colon G \to H$ is a bijection, $|G| = |H|$, so (1) is clear.

In (2) we want to show that if G is abelian, then H is abelian as well. So suppose that G is abelian. To show H is abelian, let $h_1, h_2 \in H$. Since f is a bijection, for some $g_1, g_2 \in G$:

$$h_1 = f(g_1),$$
$$h_2 = f(g_2).$$

Then

$$\begin{aligned} h_1h_2 &= f(g_1)f(g_2) && \text{definition} \\ &= f(g_1g_2) && f \text{ is an isomorphism} \\ &= f(g_2g_1) && G \text{ abelian} \\ &= f(g_2)f(g_1) && f \text{ isomorphism} \\ &= h_2h_1 && \text{definition} \end{aligned}$$

so H is abelian.

For (3), we will show that for all $k \in \mathbb{Z}_{\geq 1}$, G and H have the same number of elements of order k. So fix $k \in \mathbb{Z}_{\geq 1}$ and let

$$G_k = \{x \in G \mid o(x) = k\},$$
$$H_k = \{y \in H \mid o(y) = k\}.$$

We claim that f sends G_k bijectively onto H_k.

If $x \in G$ and $o(x) = k$, then $o(f(x)) = k$ by Lemma 2.2(b). Hence $\overline{f}\colon G_k \to H_k$ given by $\overline{f}(x) = f(x)$ for $x \in G_k$ is well-defined. Consider the inverse function $f^{-1}\colon H \to G$, which is also an isomorphism. Then $\overline{f^{-1}}\colon H_k \to G_k$ is also well-defined and because $f \circ f^{-1} = f^{-1} \circ f =$ identity, we see that G_k bijects with H_k, so $|G_k| = |H_k|$. □

Cyclic groups. Recall that a group is *cyclic* if it is generated by a single element.

Proposition 2.3.

(1) *If G is cyclic and $|G| = n$, then $G \cong C_n$.*
(2) *If G is cyclic and infinite, then $G \cong (\mathbb{Z}, +)$.*

Proof. For (1), suppose G is cyclic, say $G = \langle x \rangle$ and $o(x) = n$. This means

$$G = \{e, x, x^2, \ldots, x^{n-1}\}.$$

Recall that $C_n = \{1, \omega, \omega^2, \ldots, \omega^{n-1}\}$ with $\omega = e^{2\pi i/n}$. So let us define $f\colon G \to C_n$ by

$$f(x^r) = \omega^r \text{ for } 0 \leq r \leq n-1.$$

Because x and ω both have order n, we could even write

$$f(x^r) = \omega^r$$

for all $r \in \mathbb{Z}$. Then f is a bijection and, furthermore, by the definition of f and exponentiation:

$$f(x^r x^s) = f(x^{r+s}) = \omega^{r+s} = \omega^r\omega^s = f(x^r)f(x^s).$$

Thus f is an isomorphism and $G \cong C_n$.

For (2), say $G = \langle x \rangle$ but suppose G is infinite. Then $o(x) = \infty$, so

$$G = \{\ldots, x^{-2}, x^{-1}, e, x, x^2, x^3, \ldots\}.$$

Define $f\colon G \to \mathbb{Z}$ by

$$f(x^r) = r$$

for all $r \in \mathbb{Z}$. This is clearly a bijection and, futhermore, by the definition of f and exponentiation:

$$f(x^r x^s) = f(x^{r+s}) = r + s = f(x^r) + f(x^s).$$

Thus f is an isomorphism and $G \cong \mathbb{Z}$. □

Remark. Proposition 2.3 (1) says that all cyclic groups of size n are isomorphic to C_n. We say that *up to isomorphism* (which means: *count isomorphic things as the same*) there is only one cyclic group of order n. Similarly, (2) says that up to isomorphism, there is only one infinite cyclic group.

As a consequence: if G is a group, p is a prime number, and $|G| = p$, then $G \cong C_p$ by Lagrange's theorem. Hence, up to isomorphism, there is only one group of order p.

This is very useful, because without the "up to isomorphism" trick, there are infinitely many groups of every order.

Example. There are infinitely many groups of order 1. For if $x \in \mathbb{R}$, let $G = \{x\}$ and let us define a group law $\star$ by $x \star x = x$. Then $(G, \star)$ is a group and $|G| = 1$. While G is probably not equal to C_1, $G \cong C_1$ trivially.

Example. Let $G = (\mathbb{Z}, +)$, take $g = 2 \in G$ and $H = \langle g \rangle$. Then

$$H = \{\ldots, -6, -4, -2, 0, 2, 4, \ldots\},$$

so $H \subseteq G$ but $H \neq G$. But we can prove that $H \cong G$ either by noticing that H is infinite and cyclic, so $H \cong \mathbb{Z}$ by Proposition 2.3(2), or simply defining $f\colon G \to H$ by $f(n) = 2n$, an isomorphism.

3. Parity of permutations

Recall that if $n \in \mathbb{Z}_{\geq 1}$, then S_n is the group of all permutations of $\{1, 2, \ldots, n\}$. From 1st year: every element of S_n is a product of disjoint cycles. What we are going to do in this section is assign a sign ($+1$ or -1) to each permutation: the permutations with sign $+1$ will be *even* and the permutations with sign -1 will be *odd*.

Exercise (hard, for now). A *transposition* in S_n is a permutation of the form (ij) with $i \neq j$, i.e. "swap two around".
Q1. Is the identity element of S_3 equal to a product of an odd number of transpositions?
Q2. What about S_n?

We will attach a sign to a permutation (an element of S_n). We start with the case S_3. Set

$$\Delta = (x_1 - x_2)(x_1 - x_3)(x_2 - x_3).$$

A permutation $\sigma \in S_3$ is going to permute the x_i in the obvious way (permutation representation of S_3), for example, if $\sigma = (13)$, then $\sigma(x_1) = x_3$, $\sigma(x_2) = x_2$, $\sigma(x_3) = x_1$. What is $\sigma(\Delta)$?

$$\begin{aligned}\sigma(\Delta) &= (13)\big((x_1 - x_2)(x_1 - x_3)(x_2 - x_3)\big)\\ &= (x_3 - x_2)(x_3 - x_1)(x_2 - x_1)\\ &= (x_2 - x_1)(x_3 - x_1)(x_3 - x_2)\\ &= -\Delta\end{aligned}$$

Another example, $\sigma = (123)$:

$$\begin{aligned}\sigma(\Delta) &= (123)\big((x_1 - x_2)(x_1 - x_3)(x_2 - x_3)\big)\\ &= (x_2 - x_3)(x_2 - x_1)(x_3 - x_1)\\ &= (x_2 - x_1)(x_3 - x_1)(x_2 - x_3)\\ &= \Delta\end{aligned}$$

Each permutation will send Δ to $\pm\Delta$. That is the sign! By the above, the sign of (13) is $-$ and the sign of (123) is $+$.

Boring exercise. Check that sign is $+$ for e, (123), (132) and sign is $-$ for (13), (12), (23).

General case. Let $n \in \mathbb{Z}_{\geq 1}$, $x_1, \ldots, x_n$ variables and define

$$\Delta = \prod_{1 \leq i < j \leq n} (x_i - x_j).$$

For $\sigma \in S_n$, let σ permute the x_i in the obvious way (by permuting the indices) to get

$$\sigma(\Delta) = \pm\Delta.$$

Definition. The *signature* or *sign* of σ is

$$\begin{array}{ll} +1 & \text{if } \sigma(\Delta) = \Delta \\ -1 & \text{if } \sigma(\Delta) = -\Delta \end{array}$$

Another way of thinking about it. If $\sigma \in S_n$, then the signature of σ is $(-1)^d$, where d is the number of sign changes in Δ, i.e.

$$d = \#\{(i,j) \mid 1 \leq i < j \leq n \text{ and } \sigma(i) > \sigma(j)\}$$

Notation. We write $\operatorname{sgn}(\sigma)$ for the signature of σ, i.e. sgn is a function

$$\operatorname{sgn}\colon S_n \to \{\pm 1\}.$$

We call σ an *even* permutation if $\operatorname{sgn}(\sigma) = +1$ and *odd* if $\operatorname{sgn}(\sigma) = -1$.

Note. This definition is unusable for big n. How can we work out $\operatorname{sgn}(g)$ for $g \in S_{10}$?

Here is a great first step.

Proposition 3.1.

(1) *For any* $x, y \in S_n$, $\operatorname{sgn}(xy) = \operatorname{sgn}(x)\operatorname{sgn}(y)$,
(2) $\operatorname{sgn}(e) = +1$ *and* $\operatorname{sgn}(x^{-1}) = \operatorname{sgn}(x)$,
(3) *If* $t = (rs)$ *is a 2-cycle, then* $\operatorname{sgn}(t) = -1$.

Proof. For (1), we note that $x(\Delta) = \operatorname{sgn}(x)\Delta$ and $y(\Delta) = \operatorname{sgn}(y)\Delta$. Now

$$\begin{aligned} xy(\Delta) &= x(y(\Delta)) && \text{definition}\\ &= x(\operatorname{sgn}(y)\Delta) && \text{definition}\\ &= \operatorname{sgn}(y)(x(\Delta)) && x \text{ does not move signs}\\ &= \operatorname{sgn}(y)\operatorname{sgn}(x)\Delta && \text{definition}\end{aligned}$$

so $\operatorname{sgn}(xy)\Delta = \operatorname{sgn}(y)\operatorname{sgn}(x)\Delta$ which yields $\operatorname{sgn}(xy) = \operatorname{sgn}(x)\operatorname{sgn}(y)$.

We now use (1) to prove (2):

$$\mathrm{sgn}(x)\mathrm{sgn}(x^{-1}) = \mathrm{sgn}(xx^{-1}) = \mathrm{sgn}(e) = 1$$

where $\mathrm{sgn}(e) = 1$ because $e(\Delta) = \Delta$. Then $\mathrm{sgn}(x^{-1}) = \mathrm{sgn}(x)^{-1} = \mathrm{sgn}(x)$, as requested.

Finally, to show (3) let $t = (rs)$ with $1 \leq r, s \leq n$ and assume without loss of generality that $r < s$. We need to count the number of brackets $(x_i - x_j)$ which contribute a sign change. The question becomes: how many (i, j) have $i < j$ but $t(i) > t(j)$?

A careful count shows that the brackets that change sign are:

$$(x_r - x_{r+1}), (x_r - x_{r+2}), \ldots, (x_r - x_{s-1})$$

similarly

$$(x_{r+1} - x_s), (x_{r+2} - x_s), \ldots, (x_{s-1} - x_s)$$

and finally

$$(x_r - x_s).$$

Total number of brackets here is:

$$d = (s - 1 - r) + (s - 1 - r) + 1 = 2(s - 1 - r) + 1$$

which is an odd number, so $(-1)^d = -1$ and we get $\mathrm{sgn}(t) = -1$. □

As we will see later, (1) actually shows that $\mathrm{sgn}\colon S_n \to C_2$ is a *homomorphism*.

This proposition now gives a strategy for computing the sgn of a permutation without ever thinking about Δ. We can express the permutation σ as a product

$$\sigma = t_1 t_2 \ldots t_n$$

of n 2-cycles, and then (1) and (3) give

$$\mathrm{sgn}(\sigma) = (-1)^n.$$

Proposition 3.2. *Let $c = (a_1 a_2 \ldots a_r)$ be an r-cycle. Then*

$$c = (a_1 a_r)(a_1 a_{r-1}) \ldots (a_1 a_3)(a_1 a_2).$$

Hence c is a product of $r - 1$ 2-cycles.

Proof. Compute! Check the product on RHS sends:

$$\begin{array}{ccc} a_1 & \to & a_2 \\ a_2 & \to & a_3 \\ \vdots & \vdots & \vdots \\ a_{r-1} & \to & a_r \\ a_r & \to & a_1 \end{array}$$

so it is exactly c. □

Proposition 3.3. *If c is an r-cycle, then* $\mathrm{sgn}(\sigma) = (-1)^{r-1}$.

Proof. By Proposition 3.2, c is a product of $r - 1$ 2-cycles, so by Proposition 3.1 (1) and (3) we get $\mathrm{sgn}(\sigma) = (-1)^{r-1}$. □

Proposition 3.4.

(1) *Every $g \in S_n$ is a product of 2-cycles.*
(2) *If $g = c_1 c_2 \dots c_m$ with c_i a cycle of length r_i, then*

$$\operatorname{sgn}(g) = (-1)^{(r_1-1)+(r_2-1)+\cdots+(r_m-1)}.$$

Proof. By 1st year, any $g \in S_n$ is a product of (disjoint) cycles, so (1) follows from Proposition 3.2. Then (2) follows from Propositions 3.3 and 3.1 (1) □

Example. Let $\sigma = (123)(4567)$. To compute $\operatorname{sgn}(\sigma)$, we note that (by Proposition 3.2):

$$(123)(4567) = (13)(12)(47)(46)(45),$$

so $\operatorname{sgn}(\sigma) = (-1)^5 = -1$. Alternatively, Proposition 3.3 (2) yields

$$\operatorname{sgn}((123)(4567)) = (-1)^{(3-1)+(4-1)} = (-1)^5 = -1.$$

Alternatively, use pro method: $\operatorname{sgn}(123) = +1$, $\operatorname{sgn}(4567) = -1$ by Proposition 3.3, so

$$\operatorname{sgn}((123)(4567)) = -1$$

Example. Let $\sigma = (12)(345)(6789)$. Then

$$\operatorname{sgn}(\sigma) = (-1) \times (+1) \times (-1) = +1.$$

The sgn function breaks S_n up into 2 disjoint pieces:

$$S_n = \{\sigma \colon \operatorname{sgn}(\sigma) = +1\} \cup \{\sigma \colon \operatorname{sgn}(\sigma) = -1\}.$$

What can we say about these pieces?

Definition. For $n \in \mathbb{Z}_{n\geq 1}$, let the *alternating group* be

$$A_n = \{\sigma \in S_n \colon \operatorname{sgn}(\sigma) = +1\}.$$

Proposition 3.5. *The subset $A_n \subseteq S_n$ is a subgroup and if $n \geq 2$ then $|A_n| = (n!)/2 = |S_n|/2$.*

We will see later that A_n is actually the kernel of sgn, so it is immediately a subgroup of S_n. However, we also give a direct proof below.

Proof. We will show that A_n is a subgroup:

(1) $sgn(e) = +1$ by Proposition 3.1 (1), so $e \in A_n$.
(2) If $x, y \in A_n$, then $\operatorname{sgn}(x) = \operatorname{sgn}(y) = +1$ by definition, so $\operatorname{sgn}(xy) = +1$ and $xy \in A_n$.
(3) If $x \in A_n$, then $\operatorname{sgn}(x) = +1$, so $\operatorname{sgn}(x^{-1}) = +1$ by Proposition 3.1 and $x^{-1} \in A_n$.

Now say $n \geq 2$ and set $\tau = (12)$. Consider the coset $A_n\tau$. We claim that:

$$A_n\tau = \{\sigma \in S_n \colon \operatorname{sgn}(\sigma) = -1\}.$$

In general, a good strategy to prove sets X and Y are equal is to show $X \subseteq Y$ and $Y \subseteq X$. So first, suppose that $x \in A_n\tau$. Then $x = \sigma\tau$ for some σ with $\operatorname{sgn}(\sigma) = +1$. But $\operatorname{sgn}(\tau) = -1$

by Proposition 3.1 (3), so

$$\begin{aligned}\operatorname{sgn}(x) &= \operatorname{sgn}(\sigma\tau)\\ &= \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau) \quad \text{by Proposition 3.1 (1)}\\ &= (+1)\times(-1)\\ &= -1\end{aligned}$$

For the other inclusion, consider $y \in S_n$ and $\operatorname{sgn}(y) = -1$ and set $\sigma = y\tau$. Then

$$\operatorname{sgn}(\sigma) = \operatorname{sgn}(y)\operatorname{sgn}(\tau) = (-1)\times(-1) = +1$$

by Proposition 3.1, so $\sigma \in A_n$ and $y = y\tau^2 = (y\tau)\tau = \sigma\tau \in A_n\tau$.

We just proved that

$$S_n = A_n \cup A_n\tau$$

and since both cosets have the same cardinality, $|A_n| = |S_n|/2 = n!/2$. □

Examples.

($n = 0$) $A_0 = S_0 = \{\text{empty function}\}$, so $|S_0| = |A_0| = 1$.
($n = 1$) $A_1 = S_1 = \{e\}$.
($n = 2$) $S_2 = \{e, (12)\}$ and $A_2 = \{e\}$.
($n = 3$) In S_3: e, (123), (132) all have sign $+1$, (12), (13), (23) all have sign -1. Thus

$$A_3 = \{e, (123), (132)\} \cong C_3.$$

($n = 4$) S_4 cycle types: e $(+1)$, 2-cycle (-1), 3-cycle $(+1)$, 4-cycle (-1), and "2+2"-cycle $(+1)$. Therefore:

$$A_4 = \{e, (12)(34), (13)(24), (14)(23), \text{ all 8 of the 3-cycles}\}$$

Remark. For $n \geq 5$, the group A_n is a so-called *simple group*.

4. Direct products

Idea: let G_1 and G_2 be groups. We can form the product set (cartesian product)

$$G_1 \times G_2 = \{(g_1, g_2) \colon g_1 \in G_1, g_2 \in G_2\}.$$

Define a multiplication on $G_1 \times G_2$.

Definition. For $g_1, h_1 \in G_1$ and $g_2, h_2 \in G_2$, we define

$$(g_1, g_2) \times (h_1, h_2) = (g_1h_1, g_2h_2).$$

Clearly a well-defined product. Check the axioms.

Proposition 4.1. *$G_1 \times G_2$ becomes a group with this multiplication.*

Proof. Associativity: If $g_1, h_1, k_1 \in G_1$ and $g_2, h_2, k_2 \in G_2$, then

$$\begin{aligned}(g_1, g_2)((h_1, h_2)(k_1, k_2)) &= (g_1, g_2)(h_1k_1, h_2k_2)\\ &= (g_1(h_1k_1), g_2(h_2k_2))\\ &= ((g_1h_1)k_1, (g_2h_2)(k_2)) \qquad \text{by associativity of } G_1 \text{ and } G_2\\ &= (g_1h_1, g_2h_2)(k_1, k_2)\\ &= ((g_1, g_2)(h_1, h_2))(k_1, k_2)\end{aligned}$$

TL;DR version: G_1, G_2 associative $\Rightarrow G_1 \times G_2$ associative.

It is easy check to check that inverse of (g_1, g_2) is (g_1^{-1}, g_2^{-1}) and the identity is (e_1, e_2) for e_i identity of G_i. $\square$

If $|G| = a$, $|H| = b$, then $|G \times H| = ab$. For example, $S_4 \times S_7$, $C_{10} \times \mathbb{Z}$.

Basic observations about $G \times H$.

(1) If $G = \{e\}$, then $H \cong G \times H$ via the obvious map $h \mapsto (e, h)$. Similarly, $G \times \{e\} \cong G$.
In general, for any groups G, H, there is a subgroup $G \times \{e_H\} \subseteq G \times H$, i.e.
$$G \times \{e_H\} = \{(g, e_H) \mid g \in G\}$$
and this subgroup is isomorphic to G. Similarly, $\{e_G\} \times H \subseteq G \times H$ is a subgroup isomorphic to H. ("x-axis and y-axis").
(2) $G \times H \cong H \times G$ (isomorphism sends (g, h) to (h, g))

Example. Let $G = H = C_2 = \{+1, -1\}$. Then
$$G \times H = C_2 \times C_2 = \{(1,1) = e, (1,-1) = a, (-1,1) = b, (-1,-1)\}.$$
Note that
$$ab = (1,-1) \times (-1,1) = (-1,-1)$$
which is the 4th element.

	e	a	b	ab
e	e	a	b	ab
a	a	e	ab	b
b	b	ab	e	a
ab	ab	b	a	e

It is easy to check that $C_2 \times C_2$ is isomorphic to the symmetries of a non-square rectangle.

We can do more than 2 groups. Start with groups G_1, G_2, ..., G_r. Define
$$G_1 \times G_2 \times \cdots \times G_r = \{(g_1, g_2, \ldots, g_r) \colon g_i \in G_i\}.$$
with the group law:
$$(g_1, g_2, \ldots, g_r)(h_1, h_2, \ldots, h_r) = (g_1 h_1, g_2 h_2, \ldots, g_r h_r).$$
It is easy to check that the group axioms are satisfied.

The group $G_1 \times G_2 \times \cdots \times G_r$ is called the *direct product* of the G_i.

Remark. For $r = 0$, we get $\{\emptyset\} \cong C_1$. For $r = 1$, we just get G_1 again. We can even do $r = \infty$:
$$(g_1, g_2, \ldots) \in \prod_{i=1}^{\infty} G_i.$$
with the obvious multiplication.

Example ($r = 3$)**.** The direct product $C_2 \times C_2 \times C_2$ is a group of size 8 whose elements are triples $(\pm 1, \pm 1, \pm 1)$. Note that $x^2 = e$ for all $x \in C_2 \times C_2 \times C_2$.

Proposition 4.2.

(1) *The order of* $G_1 \times G_2 \times \cdots \times G_r$ *is* $|G_1| \times |G_2| \times \cdots \times |G_r|$.
(2) *If all the* G_i *are abelian, then so is* $G_1 \times \cdots \times G_r$.
(3) *If* $x = (x_1, x_2, \ldots, x_r) \in G_1 \times \cdots \times G_r$, *then*

$$o(x) = \mathrm{lcm}(o(x_i)).$$

Proof. (1) is clear. For (2), say $x = (x_1, \ldots, x_r)$ and $y = (y_1, \ldots, y_r)$ in the product. Assume all G_i are abelian and let us prove $xy = yx$. We have

$$\begin{array}{lll} xy & = (x_1y_1, x_2y_2, \ldots, x_ry_r) & \text{definition} \\ & = (y_1x_1, y_2x_2, \ldots, y_rx_r) & \text{all } G_i \text{ abelian} \\ & = yx & \text{definition} \end{array}$$

For (3), recall that the order of $x \in G$ is the smallest $k \in \mathbb{Z}$, $k \geq 1$, such that $x^k = e$ (or $+\infty$ if no such k exists). If $o(x) = n$ then

$$x^t = e \quad \Leftrightarrow \quad t = \text{multiple of } n$$

Assume $o(x_i) = n_i < +\infty$ for all i. Let $n = \mathrm{lcm}(n_1, n_2, \ldots, n_r)$. First note

$$x^n = (x_1^n, x_2^n, \ldots, x_r^n) = (e_1, e_2, \ldots, e_r)$$

because n is a multiple of n_i for $1 \leq i \leq r$.

Conversely, if $0 < m < n$, then we claim that $x^m \neq e$. Indeed, since $m < \mathrm{lcm}(n_i)$, there exists i such that n_i does not divide m, so $x_i^m \neq e_i$. Therefore

$$x^m = (x_1^m, \ldots, x_i^m, \ldots, x_r^m) \neq (e_1, \ldots, e_i, \ldots, e_r)$$

and we have shown that the order of x is n. □

Example. Consider the group

$$G = D_{10} \times S_3 \times C_2.$$

Then for $g = (\varrho, (123), -1)$, where ϱ is a rotation of order 5, we have

$$o(g) = \mathrm{lcm}(5, 3, 2) = 30.$$

Let us think of some groups of order 8. Let us even think of **abelian** groups of order 8:

$$C_8$$

$$C_4 \times C_2$$

$$C_2 \times C_2 \times C_2$$

Are any two of these isomorphic? No! Let us look at elements of order 2.

Group	elements of order 2	number of elements of order 2
C_8	-1	1
$C_4 \times C_2$	$(-1, 1)$, $(1, -1)$, $(-1, -1)$	3
$C_2 \times C_2 \times C_2$	all $(\pm 1, \pm 1, \pm 1)$ except $(1, 1, 1)$	7

Structure theorem for finite abelian groups.

Theorem 4.3. *Every finite abelian group is isomorphic to a direct product of cyclic groups.*

Proof. See for example R. Allenby, *Rings, Fields and Groups* (p. 254). □

So to compute all abelian groups of order n, we must write down all factorisations $n = n_1 n_2 \dots n_r$ and then the groups

$$C_{n_1} \times C_{n_2} \times \cdots \times C_{n_r}$$

give us a complete list. For example, for $n = 4$, every abelian group of order 4 is isomorphic to C_4 or $C_2 \times C_2$. For $n = 8$, every abelian group of order 8 is isomorphic to one of

$$C_8,\ C_4 \times C_2,\ C_2 \times C_2 \times C_2.$$

Now let us try $n = 6$:

$$C_6, C_2 \times C_3,$$

so any abelian group of order 6 is isomorphic to either C_6 or $C_2 \times C_3$. However

$$C_2 \times C_3 \cong C_6.$$

Indeed, $-1 \in C_2$ has order 2 and $\omega = e^{2\pi i/3} \in C_3$ has order 3, so $x = (-1, \omega) \in C_2 \times C_3$ has order $\operatorname{lcm}(2,3) = 6$. Therefore $\langle x \rangle$ has size 6 and so must be all of $C_2 \times C_3$.

We can conclude that up to isomorphism, there is only one abelian group of order 6, namely C_6.

In general, the following proposition holds.

Proposition 4.4. *If* $\operatorname{hcf}(m,n) = 1$*, then*

$$C_m \times C_n \cong C_{mn}.$$

Proof. If $g \in C_m$ has order m and $h \in C_n$ has order n, then $(g,h) \in C_m \times C_n$ has order $\operatorname{lcm}(m,n) = mn/\operatorname{hcf}(m,n) = mn$. Thus $\langle (g,h) \rangle$ has order mn and must generate $C_m \times C_n$ which is thus cyclic. □

Exercise. What are all abelian groups of size 12?

5. Groups of small order

In this section, we consider the problem of finding all groups of size n up to isomorphism.

Remark. For $n = 2048$, this is an open problem! No one knows how many groups of order 2048 there are, up to isomorphism. The largest lower bound I found is 1774274116992170.

In this chapter, we will solve this problem for $n \leq 7$. Let us start with the simplest cases first:

($n = 1$) The only group is C_1.

($n = 2$) Since 2 is prime, the only group is C_2.

In general, if p is prime and G has order p, then G is cyclic. This solves the problem for $n = 2, 3, 5, 7$ and we are left with the cases $n = 4, 6$.

Proposition 5.1. *Up to isomorphism, the only groups of size 4 are* C_4 *and* $C_2 \times C_2$*.*

Proof. Say G is a group and $|G| = 4$. What are the orders of the elements of G? Firstly, $o(g) = 1$ if and only if $g = e$. If there is a $g \in G$ such that $o(g) = 4$, then $|\langle g \rangle| = 4 = |G|$ and $G \cong C_4$. So suppose that $G \ncong C_4$. Then we can write

$$G = \{e, x, y, z\} \text{ with } o(x) = o(y) = o(z) = 2.$$

To check that $xy = z$, note that $x^2 = y^2 = e$, but $x, y \neq e$ and $x \neq y$. Now, if $xy = e$, then $xy = x^2 = xx$ and $y = x$, a contradiction. Similarly, if $xy = x$ ($xy = y$), then $y = e$ ($x = e$), a contradiction. Thus $xy = z$ and the same argument shows $yx = z$.

Therefore, we can write out the multiplication table of G:

	e	x	y	z
e	e	x	y	z
x	x	e	z	y
y	y	z	e	x
z	z	y	x	e

and note that this is the same as the multiplication table of $C_2 \times C_2$, i.e. $G \cong C_2 \times C_2$ via $e \mapsto (1, 1)$, $x \mapsto (-1, 1)$, $y \mapsto (1, -1)$, $z \mapsto (-1, -1)$. □

Groups of order 6. Can think of C_6, D_6, S_3. Now, $S_3 \cong D_6$ and also $C_6 \cong C_2 \times C_3$ but $C_6 \ncong D_6$. So is the answer that $|G| = 6$ implies $G \cong C_6$ or D_6?

Lemma 5.2. *If G is a finite group and $|G|$ is even, then there exists $g \in G$ such that $o(g) = 2$.*

Proof. Proof by contradiction: assume there is no $g \in G$ such that $o(g) = 2$. Then for all $g \in G$ either $g = e$ or $g \neq g^{-1}$. Then we can list all elements of G to be

$$G = \{e, g_1, g_1^{-1}, g_2, g_2^{-1}, \ldots, g_r, g_r^{-1}\}$$

and $|G| = 2r + 1$ is odd, a contradiction. □

Remark (Cauchy's Theorem)**.** If p is prime and $|G| =$ multiple of p, then there exists $g \in G$ such that $o(g) = p$. (This will not be proved in this course.)

Proposition 5.3. *Up to isomorphism, the only groups of order 6 are C_6 and D_6.*

Proof. Let G be a group with $|G| = 6$. Note that $e \in G$ has order 1 and all the other elements of G have orders $2, 3, 6$. If some element has order 6, then $G \cong C_6$, so suppose no element has order 6. By Lemma 5.2, there is an element of order 2 in G. But not all the elements can have order 2—if they did, then (by Problem Sheet 2) either $|G| = 2$ or $|G|$ is a multiple of 4.

Therefore, we know that if $|G| = 6$ and $G \ncong C_6$, then there exist elements $x \in G$ of order 3 and $y \in G$ of order 2. Set $H = \langle x \rangle \cong C_3 \subseteq G$, $|H| = 3$, $|G| = 6$. Then $y \notin H$, since the elements of H have order 1 or 3. Thus

$$G = H \cup Hy = \{e, x, x^2, y, xy, x^2y\}.$$

What is yx? If $yx \in H$, then $y \in Hx^{-1} = H$ which is impossible. Thus $yx \in Hy$. First, $yx = y$ implies $x = e$, so $yx \neq y$. Moreover, if $yx = xy$, then:

$$(yx)^1 = xy \neq e,$$
$$(yx)^2 = xyxy = x^2y^2 = x^2 \neq e,$$
$$(yx)^3 = xyxyxy = x^3y^3 = y \neq e,$$

so $o(xy) \neq 1, 2, 3$, but $xy \in G$, a contradiction. Therefore, $yx = x^{-1}y$ and

$$G = H \cup Hy = \{e, x, x^2, y, xy, x^2y\}$$

with the relations $x^3 = e$, $y^2 = e$, $yx = x^{-1}y$. Thus $G \cong D_6$ via $x \mapsto \varrho$, $y \mapsto \sigma$. □

Summary of this chapter:

Size	Groups
1	C_1
2	C_2
3	C_3
4	C_4, $C_2 \times C_2$
5	C_5
6	C_6, D_6
7	C_7

Remark. For some chosen higher orders:

Size	Groups
8	C_8, $C_4 \times C_2$, $C_2 \times C_2 \times C_2$, D_8, and one other
9	C_9, $C_3 \times C_3$
10	C_{10}, D_{10}
32	51 groups
1024	49,487,365,422 groups
2048	unknown

Interestingly, $\approx 99.2\%$ of all groups of order at most 2000 have order 1024.

6. Homomorphisms, normal subgroups, and factor groups

Homomorphisms are functions between groups which preserve the group multiplication.

Definition. Let G, H be groups. A function $\varphi\colon G \to H$ is a *homomorphism* if

$$\varphi(xy) = \varphi(x)\varphi(y)$$

for all $x, y \in G$.

Note that an isomorphism is a homomorphism which is also a bijection.

Examples.

(1) Let G, H be any groups. Define $\varphi\colon G \to H$ by $\varphi(x) = e_H$ for all $x \in G$. This is the *trivial homomorphism*.

(2) The signature function $\mathrm{sgn}\colon S_n \to C_2$ is a homomorphism, because for all $x, y \in S_n$ we have $\mathrm{sgn}(xy) = \mathrm{sgn}(x)\mathrm{sgn}(y)$ by Proposition 3.1 (1).
(3) Define $\varphi\colon (\mathbb{R}, +) \to (\mathbb{C}^\times, \times)$ by $\varphi(x) = e^{2\pi i x}$ for $x \in \mathbb{R}$. Check it is a group homomorphism: if $x, y \in \mathbb{R}$, then

$$\varphi(x+y) = e^{2\pi i(x+y)} = e^{2\pi i x} \cdot e^{2\pi i y} = \varphi(x)\varphi(y).$$

(4) Define $\varphi\colon D_{2n} \to C_2$ by $\varphi(\varrho^i\sigma^j) = (-1)^j$ for all $i, j \in \mathbb{Z}$.
Dodgy definition: need to check it is well-defined, i.e. if $\varrho^i\sigma^j = \varrho^k\sigma^l$ we had better make sure $(-1)^j = (-1)^l$. (Otherwise φ would not be a function.) Let us check we are ok: if $\varrho^i\sigma^j = \varrho^k\sigma^l$, then

$$\varrho^{-k}\varrho^i\sigma^j = \sigma^l$$
$$\varrho^{-k}\varrho^i = \sigma^l\sigma^{-j}$$
$$\sigma^{l-j} = \varrho^{i-k} \in \langle\varrho\rangle,$$

so $l - j$ had better be even, $(-1)^l = (-1)^j$, and φ is well defined.
In words, φ sends rotations to $+1$ and reflections to -1.
Finally, check ϱ is a homomorphism: take $x = \varrho^i\sigma^j$, $y = \varrho^k\sigma^l$. Then $\varphi(x) = (-1)^j$, $\varphi(y) = (-1)^l$. Moreover

$$xy = \varrho^i\sigma^j\varrho^k\sigma^l.$$

Note that $\sigma\varrho^k = \varrho^{-k}\sigma$ and $\sigma^0\varrho^k = \varrho^k = \varrho^k\sigma^0$, so we can write $\sigma^j\varrho^k = \varrho^{\pm k}\sigma^j$ (where the sign depends on the value of j). Thus

$$xy = \varrho^i\varrho^{\pm k}\sigma^j\sigma^l = \varrho^{i\pm k}\sigma^{j+l}$$

and

$$\varphi(xy) = (-1)^{j+l} = (-1)^j(-1)^l = \varphi(x)\varphi(y).$$

(5) (Exercise.) If G, H are groups, then there are projection maps

$$G \times H \to G$$
$$(g, h) \mapsto g$$

and

$$G \times H \to H$$
$$(g, h) \mapsto h$$

These are both group homomorphisms!

Proposition 6.1. *Say $\varphi\colon G \to H$ is a group homomorphism. Then*

(1) $\varphi(e_G) = e_H$,
(2) $\varphi(x^{-1}) = \varphi(x)^{-1}$ *for all* $x \in G$,
(3) *for $x \in G$, the order of $\varphi(x)$ divides $o(x)$.*

Proof. For (1), set $h = \varphi(e_G)$. Then

$$\begin{array}{lll} h & = \varphi(e_G e_G) & e_G^2 = e_G \\ & = \varphi(e_G)\varphi(e_G) & \varphi \text{ is a homomorphism} \\ & = h^2 & \text{definition} \end{array}$$

so $h = e_H$, canceling h.

For (2), take $x \in G$. Then $xx^{-1} = e_G$. Apply φ to get $\varphi(xx^{-1}) = \varphi(e_G)$. Thus

$$\begin{aligned} \varphi(x)\varphi(x^{-1}) &= \varphi(e_G) && \varphi \text{ group homomorphism} \\ &= e_H && \text{by (1)} \end{aligned}$$

so $\varphi(x^{-1}) =$ inverse of $\varphi(x) = \varphi(x)^{-1}$.

Finally, for (3) set $o(x) = k$. Then

$$\underbrace{xx\dots x}_{k \text{ times}} = e_G.$$

Apply φ to get

$$\varphi(x)^k = \varphi(e_G) = e_H$$

since φ is a group homomorphism and using (1). Thus $k =$ multiple of order of $\varphi(x)$, so the order of $\varphi(x)$ divides k. □

Image and kernel. Recall that two important subspaces of vector spaces are the image and the kernel of a linear map. In this section, we discuss the analogous notions for groups and group homomorphisms.

Definition. Let $\varphi\colon G \to H$ be a group homomorphism. The *image of* φ is

$$\operatorname{im}(\varphi) = \{\varphi(x) \mid x \in G\} \subseteq H.$$

Proposition 6.2. *The image* $\operatorname{im}(\varphi)$ *of a homomorphism* $\varphi\colon G \to H$ *is a subgroup of* H.

Proof. First, $\varphi(e_G) = e_H$ by Proposition 6.1, so $e_H \in \operatorname{im}(\varphi)$. Next, if $\varphi(x), \varphi(y) \in \operatorname{im}(\varphi)$, then

$$\varphi(x)\varphi(y) = \varphi(xy) \in \operatorname{im}(\varphi).$$

Finally, say $\varphi(x) \in \operatorname{im}(\varphi)$. Then

$$\varphi(x)^{-1} = \varphi(x^{-1}) \in \operatorname{im}(\varphi)$$

by Proposition 6.1. Thus the image is a subgroup of H. □

Question. There is a non-trivial homomorphism $S_3 \to C_2$ (namely, sgn). Is there a non-trivial homomorphism

$$S_3 \to C_3?$$

Note that

$$S_3 \supseteq \langle (123) \rangle \cong C_3,$$
$$S_3 \supseteq \langle (12) \rangle \cong C_2.$$

We also have

$$\operatorname{sgn}\colon S_3 \to C_2$$

and $\operatorname{sgn}(e) = \operatorname{sgn}((123)) = \operatorname{sgn}((132)) = 1$, $\operatorname{sgn}((i,j)) = -1$ and sgn is a group homomorphism. However, we claim that the only group homomorphism $S_3 \to C_3$ is the trivial one (i.e. the one that sends all $g \in S_3$ to $1 \in C_3$).

Suppose $\varphi\colon S_3 \to C_3$ is a group homomorphism. Say $\sigma = (12)$ is a 2-cycle. Then $o(\sigma) = 2$, so $o(\varphi(\sigma))$ divides 2 by Proposition 6.1. Thus $o(\varphi(\sigma))$ is 1 or 2, but $\varphi(\sigma) \in C_3$, which only has elements of order 1 and 3, so $o(\varphi(\sigma)) = 1$ and $\varphi(\sigma) = 1 \in C_3$. Similarly,

$$\varphi((23)) = \varphi((13)) = 1.$$

However, note $(123) = (13)(12)$, so

$$\begin{aligned}\varphi((123)) &= \varphi((13))\varphi((12)) \quad \varphi \text{ is a group homomorphism} \\ &= 1 \times 1 = 1\end{aligned}$$

Similarly, $\varphi((132)) = \varphi((12)(13)) = 1 \times 1 = 1$. Finally, by Proposition 6.1, $\varphi(e) = 1$, and we have shown φ is trivial.

Definition. Let $\varphi\colon G \to H$ be a group homomorphism. The *kernel of* φ is

$$\ker(\varphi) := \{x \in G\colon \varphi(x) = e_H\} \subseteq G.$$

Proposition 6.3. *The kernel* $\ker(\varphi)$ *of a homomorphism* $\varphi\colon G \to H$ *is a subgroup of* G.

Proof. First, $\varphi(e_G) = e_H$ by Proposition 6.1, so $e_G \in \ker(\varphi)$. Next, suppose $x, y \in \ker(\varphi)$, i.e. $\varphi(x) = \varphi(y) = e_H$. Hence

$$\varphi(xy) = \varphi(x)\varphi(y) = e_H e_H = e_H$$

so $xy \in \ker(\varphi)$. Finally, say $x \in \ker(\varphi)$. Then $\varphi(x) = e_H$, so

$$\begin{aligned}\varphi(x^{-1}) &= \varphi(x)^{-1} \quad \text{Proposition 6.1} \\ &= e_H^{-1} \\ &= e_H\end{aligned}$$

so $x^{-1} \in \ker(\varphi)$ by definition. □

Examples.

(0) Let $\varphi\colon G \to H$ be the trivial homomorphism, $\varphi(g) = e_H$ for all $g \in G$. Then $\mathrm{im}(\varphi) = \{e_H\}$ and $\ker(\varphi) = G$.

(1) Suppose $\varphi\colon G \to H$ an isomorphism. Then $\mathrm{im}(\varphi) = H$ as φ is a bijection (so a surjection). Moreover, $\varphi(e_G) = e_H$ and φ is a bijection (so an injection), so $\ker(\varphi) = \{e_G\}$.

(2) Consider the signature homomorphism $\mathrm{sgn}\colon S_n \to C_2$. If $n \geq 2$, then

$$\mathrm{im}(\mathrm{sgn}) = C_2.$$

Moreover

$$\ker(\mathrm{sgn}) = \{g \in S_n\colon \mathrm{sgn}(g) = +1\} = A_n$$

by definition of A_n. (This gives a new proof that A_n is a subgroup of S_n, but this is really the same as the old proof.)

(3) Define $\varphi\colon D_{2n} \to C_2$ by $\varphi(\varrho^i\sigma^j) = (-1)^j$. We checked earlier that this is well-defined. Since $\varphi(\sigma) = -1$ and $\varphi(e) = +1$, so $\mathrm{im}(\varphi) = C_2$. The kernel of φ will be

$$\ker(\varphi) = \{\varrho^i\sigma^j\colon j \text{ even}\}$$

but $\sigma^2 = e$, so

$$\ker(\varphi) = \{\varrho^i\colon i \in \mathbb{Z}\} = \langle\varrho\rangle = \{\text{rotations}\}.$$

(4) Let $\varphi\colon (\mathbb{R}, +) \to (\mathbb{C}^\times, \times)$ be defined by $\varphi(x) = e^{2\pi i x}$. Here

$$\ker(\varphi) = \{x \in \mathbb{R}\colon e^{2\pi i x} = 1\} = \mathbb{Z},$$

$$\mathrm{im}(\varphi) = S^1 = \text{unit circle} = \{z\colon |z| = 1\}.$$

Geometrically: think of $\mathbb{R}$ as a spiral which we project down onto S^1.

Normal subgroups. These are *special* subgroups of a group.

For example, $\langle(123)\rangle \subseteq S_3$ is a normal subgroup but $\langle(12)\rangle$ is not! We will see this later.

Say G is a group and $N \subseteq G$ is a subgroup. If $g \in G$, then by $g^{-1}Ng$ we mean the set

$$\{g^{-1}ng \colon n \in N\}.$$

Note that for an abelian group $g^{-1}Ng = N$, trivially.

Definition. Let G be a group, $N \subseteq G$ be a subgroup. We say that N is a *normal subgroup* of G if $g^{-1}Ng = N$ for all $g \in G$.

Notation: we write $N \lhd G$ for a normal subgroup $N \subseteq G$.

Examples.

(1) Let G be any group. The identity subgroup $\{e_G\} \subseteq G$ is normal, because
$$g^{-1}\{e_G\}g = \{g^{-1}e_G g\} = \{e_G\},$$
i.e. $\{e_G\} \lhd G$. Similarly, $G \lhd G$ as (easy exercise) for any $g \in G$:
$$g^{-1}Gg = G.$$

(2) Let G be any abelian group. Then any subgroup N of G is normal because
$$g^{-1}Ng = \{g^{-1}ng \colon n \in N\} = \{ng^{-1}g \colon n \in N\} = \{n \colon n \in N\} = N.$$

A non-example in S_3: take $g = (13)$, $n = (12)$; then
$$g^{-1}ng = (13)(12)(13) = (23).$$
Now, setting $G = S_3$ and $N = \langle(12)\rangle \subseteq G$ we get
$$g^{-1}Ng = \{g^{-1}eg, g^{-1}\ (12)g\} = \{e, (23)\} = \langle(23)\rangle \neq N.$$
Thus N is not a normal subgroup of G.

Remark. Even if $\sigma^{-1}N\sigma = N$ for **some** $\sigma \in S_3$, for example $\sigma = e$ or $\sigma = (12)$, N may not be a normal subgroup! For $\sigma = (13)$, we may have $\sigma^{-1}N\sigma \neq N$, and this is enough to say that N is not normal.

Lemma 6.4. *Say G is a group and N is a subgroup. Then $N \lhd G$ if and only if $g^{-1}Ng \subseteq N$ for all $g \in G$.*

Proof. The 'only if' implication is obvious. For the 'if' implication, suppose $g^{-1}Ng \subseteq N$ for all $g \in G$ and take an arbitrary $\sigma \in G$. We will show that
$$\sigma^{-1}N\sigma = N.$$
Setting $g = \sigma$, we get
$$\sigma^{-1}N\sigma \subseteq N$$
and setting $g = \sigma^{-1}$, we get
$$\sigma N\sigma^{-1} \subseteq N.$$
Multiplying the second equation from the left by σ^{-1} and from the right by σ, we deduce that
$$N = \sigma^{-1}\sigma N\sigma^{-1}\sigma \subseteq \sigma^{-1}N\sigma$$

so combining this with $\sigma^{-1}N\sigma \subseteq N$ we get equality, as requested. □

We can apply this to show that A_n is a normal subgroup of S_n. Take an arbitrary $g \in S_n$. All we need to check is that if $h \in A_n$, then $g^{-1}hg \in A_n$, but

$$\mathrm{sgn}(g^{-1}hg) = \mathrm{sgn}(g^{-1})\mathrm{sgn}(h)\mathrm{sgn}(g) = \mathrm{sgn}(g^{-1})\mathrm{sgn}(g) = \mathrm{sgn}(e) = 1$$

since $\mathrm{sgn}(h) = +1$. Thus $A_n \lhd S_n$.

Proposition 6.5. *Suppose $\varphi\colon G \to H$ is a group homomorphism. Then $\ker(\varphi) \lhd G$.*

Remark.

(1) Note that $\mathrm{im}(\varphi)$ may not be a normal subgroup of H. For example, if $i\colon \langle(12)\rangle \hookrightarrow S_3$ is the inclusion, then $\mathrm{im}(i) = \langle(12)\rangle$ is not a normal subgroup of S_3.
(2) Proposition 6.5 also has a converse. If $N \lhd G$, we will later on see how to build $\varphi\colon G \to H$ with kernel N.

Proof of Proposition 6.5. Let $\varphi\colon G \to H$ be a homomorphism and write $N = \ker(\varphi)$. Choose arbitrary $g \in G$ and $n \in N$. By Lemma 6.4, we just need to check that $g^{-1}ng \in N$, i.e. prove that $\varphi(g^{-1}ng) = e_H$. We have

$$\begin{aligned}
\varphi(g^{-1}ng) &= \varphi(g^{-1})\varphi(n)\varphi(g) && \varphi \text{ is a group homomorphism}\\
&= \varphi(g^{-1})e_H\varphi(g) && \text{as } n \in N\\
&= \varphi(g^{-1})\varphi(g)\\
&= \varphi(g^{-1}g) && \varphi \text{ is a group homomorphism}\\
&= e_H
\end{aligned}$$

which shows that $N \lhd G$. □

Examples.

(1) Take $\varphi\colon G \to H$, the trivial homomorphism ($\varphi(g) = e_H$ for all $g \in G$). Then $\ker(\varphi) = G \lhd G$. Similarly, $\varphi\colon G \to G$ the isomorphism $\varphi(g) = g$ for all $g \in G$ gives $\ker(\varphi) = \{e_G\} \lhd G$.
(2) Take $\mathrm{sgn}\colon S_n \to C_2$ to get
$$\ker(\mathrm{sgn}) = A_n \lhd S_n.$$
(3) Take $\varphi\colon D_{2n} \to C_2$, $\varphi(\varrho^i\sigma^j) = (-1)^j$. Then
$$\ker(\varphi) = \langle\varrho\rangle \lhd D_{2n},$$
a new example of a normal subgroup.

Factor groups. In this section, given a group G and a subgroup $H \subseteq G$, we want to *divide G by H*, i.e. get a *quotient group* or a *factor group* G/H.

Consider $G/H = X = \{Hg\colon g \in G\}$ and let us try to make X into a group. For Hy and Hz in X, try to define

$$(Hy)(Hz) = Hyz.$$

What could go wrong? *Elements of X can have more than one name!*

Example. Take $G = S_3$ and $H = \langle(12)\rangle$. The issue: we can have distinct elements $y_1, y_2 \in G$ such that $Hy_1 = Hy_2$. For example, take $y_1 = (23)$, $y_2 = (123)$. Then

$$Hy_1 = \{y_1, (12)y_1\} = \{(23), (123)\}.$$
$$Hy_2 = \{y_2, (12)y_2\} = \{(123), (23)\}.$$

(Reminder from 1st year: in general, $H\gamma = H\delta$ if and only if $\gamma\delta^{-1} \in H$.) So an element of X will have $|H|$ names in general. Let us go back to the multiplication. For the above H, y_1, y_2, we have:

$$(Hy_1)^2 = Hy_1 \cdot Hy_1 = H(y_1)^2 = He = H$$
$$(Hy_2)^2 = Hy_2 \cdot Hy_2 = H(y_2)^2 = H(132) \neq H$$

because $(132) \notin H$. Hence Hy_1 is an element of order 2, but Hy_2 is not. But $Hy_1 = Hy_2$.

What went wrong? The multiplication on X was not well-defined!

Problem. In general, we can find $y_1, y_2, z_1, z_2 \in G$ such that

$$Hy_1 = Hy_2$$
$$Hz_1 = Hz_2$$

but

$$Hy_1z_1 \neq Hy_2z_2.$$

Indeed, we just saw an explicit example with

$$H = \langle(12)\rangle, \quad y_1 = (23) = z_1, \quad y_2 = (123) = z_2.$$

If the above happens, then we will never manage to put a group structure on X in a natural way.

Lemma 6.6. *Suppose $N \lhd G$. Then $gN = Ng$ for all $g \in G$.*

Proof. If $g \in G$, then by normality of N, we know $g^{-1}Ng = N$, so multiplying from the left by g, we get $Ng = gN$. □

In general, if $H \subseteq G$, then the cosets we saw last year are $Hg =$ right cosets, but we can also talk about left cosets $gH = \{gh \colon h \in H\}$ and in general right cosets are not left cosets, $gH \neq Hg$.

But for normal subgroups N, we just saw that $gN = Ng$ for all $g \in G$.

Lemma 6.7. *Suppose $N \lhd G$. If $Ny_1 = Ny_2$ and $Nz_1 = Nz_2$, then $Ny_1z_2 = Ny_2z_2$.*

Proof. We have

$$\begin{aligned}
Ny_1z_1 &= (Ny_1)z_1 && \text{definition}\\
&= (y_1N)z_1 && \text{by Lemma 6.6}\\
&= y_1(Nz_1) && \text{associativity}\\
&= y_1(Nz_2) && \text{assumption}\\
&= (y_1N)z_2 && \text{associativity}\\
&= (Ny_1)z_2 && \text{by Lemma 6.6}\\
&= (Ny_2)z_2 && \text{assumption}\\
&= Ny_2z_2 && \text{definition}
\end{aligned}$$

which completes the proof. □

Theorem 6.8. *Suppose $N \lhd G$ and let G/N denote the set of right cosets of N in G. Define a multiplication by $(Ny)(Nz) = Nyz$. Then this is well-defined and makes G/N into a group.*

Remark. By Lemma 6.6, G/N is also the set of left cosets of N in G.

Proof. The multiplication is well-defined by Lemma 6.7. It is enough to check the group axioms:

(Associativity) We have

$$(NxNy)Nz = NxyNz = N(xy)z = Nx(yz) = NxNyz = Nx(NyNz).$$

(Identity) If $e \in G$ identity, then $He = H$ is the identity in G/N, because

$$(Hx)(He) = Hxe = Hx = Hex = (He)(Hx).$$

(Inverses) The inverse of Hx is Hx^{-1}, because

$$(Hx)(Hx^{-1}) = Hxx^{-1} = He = H = He = Hx^{-1}x = (Hx^{-1})(Hx)$$

This completes the proof. □

Definition. We call G/N the *factor group* of G by N or the *quotient* of G by N.

Examples.

(1) $G = \mathbb{Z}$, $N = 3\mathbb{Z}$ = multiples of 3. Since G is abelian and N is a subgroup, N is necessarily normal. The cosets are:

$$N = \{\ldots, -6, -3, 0, 3, 6, \ldots\},$$
$$N + 1 = \{\ldots, -5, -2, 1, 4, 7, \ldots\} = \{3k + 1 \colon k \in \mathbb{Z}\},$$
$$N + 2 = \{3k + 2 \colon k \in \mathbb{Z}\}.$$

Set of cosets G/N has 3 elements $G/N = \{N, N + 1, N + 2\}$ and the group law is

$$(N + y) + (N + z) = N + (y + z).$$

The group is generated by $x = N + 1$ of order 3:

$$x + x = (N + 1) + (N + 1) = N + 2$$
$$x + x + x = (N + 1) + (N + 2) = N + 3 = N,$$

so $\mathbb{Z}/3\mathbb{Z} \cong C_3$.

(2) $A_n \lhd S_n$, $N \geq 2$. The factor group S_n/A_n has size 2, so it must be isomorphic to C_2. The identity is A_n and the other element is $A_n(12)$ with order 2:

$$A_n(12) \times A_n(12) = A_n(12)^2 = A_n e = A_n.$$

Examples. Take $G \lhd G$. Then

$$G/G = \text{set of cosets of } G \text{ in } G = \{Ge\} = \{G\}$$

has size 1, i.e. $G/G \cong C_1$.

Now, take $\{e\} \lhd G$. What is $G/\{e\}$? In general

$$Ng = \{ng \colon n \in N\},$$

so in this case $Ng = \{g\}$ and strictly speaking

$$G/\{e\} = \big\{\{g_1\}, \{g_2\}, \dots\big\}$$

if $G = \{g_1, g_2, \dots\}$ and the obvious map

$$G \to G/\{e\}$$
$$g \mapsto \{g\}$$

is an isomorphism.

Take $G = D_{2n}$, $n \geq 3$ and let $N = \langle \varrho \rangle$. We know that $N \lhd G$. What is G/N? Well, $|G/N| = |G|/|N| = 2$ and thus $G/N \cong C_2$. Explicitly:

$$G/N = \{N, N\sigma\} = \{\text{set of rotations}, \text{set of reflections}\}.$$

Note that this shows again that rotation $\times$ reflection = reflection and reflection $\times$ reflection = rotation.

Fact. *If $G = D_{2n}$ and $N = \langle \varrho^j \rangle$ for $0 \leq j < n$, then $N \lhd G$.*

(The proof is Q6 from Problem Sheet 5.)

Example. $G = D_{12}$, $N = \langle \varrho^2 \rangle$. Then $|G| = 12$, $N = 3$, so $|G/N| = 4$. Set $F = G/N$, which is a group of size 4. Which one is it?

Solution: what are the cosets of N in G?

$$N = \{e, \varrho^2, \varrho^4\}$$
$$N\varrho = \{\varrho, \varrho^3, \varrho^5\}$$
$$N\sigma = \{\sigma, \varrho^2\sigma, \varrho^4\sigma\}$$
$$N\varrho\sigma = \{\varrho\sigma, \varrho^3\sigma, \varrho^5\sigma\}$$

Thus

$$G/N = \{N, N\varrho, N\sigma, N\varrho\sigma\}.$$

Let us check that if $f \in G/N$, then $f^2 = e_F$:

$$N^2 = (Ne)^2 = Ne^2 = Ne = N$$
$$(N\varrho)^2 = N\varrho N\varrho = N\varrho^2 = N$$
$$(N\sigma)^2 = N\sigma N\sigma = N\sigma^2 = Ne = N$$
$$(N\varrho\sigma)^2 = N(\varrho\sigma)^2 = Ne = N$$

and indeed G/N cannot be C_4, so it must be $C_2 \times C_2$.

Example. Let $G = D_{12}$ and $N = \langle \varrho^3 \rangle = \{e, \varrho^3\}$. Then $N \lhd G$, so $F = G/N$ is a group of size 6, i.e. S_3 or C_6. Which one?

$$F = \{N, N\varrho, N\varrho^2, N\sigma, N\varrho\sigma, N\varrho^2\sigma\}$$

Set $x = N\varrho$, $y = N\sigma$. Then what is $o(x)$? We have

$$x = N\varrho \neq N$$
$$x^2 = N\varrho^2 \neq N$$

$$x^3 = N\varrho^3 = N$$

(ϱ has order 6 but $\varrho^3 \in N$, so $x^3 = e_F$) so $o(x) = 3$. Moreover, $y \neq N$ and $y^2 = (N\sigma)^2 = N\sigma^2 = N$, so $o(y) = 2$. Finally, is $yx = xy$ or not?

$$\begin{array}{lll} yx & = (N\sigma)(N\varrho) & \text{definition} \\ & = N\sigma\varrho & \text{definition} \\ & = N\varrho^{-1}\sigma & \varrho, \sigma \in D_{12} \\ & = (N\varrho^{-1})(N\sigma) & \text{definition} \\ & = x^{-1}y & \end{array}$$

Therefore, $F \cong D_6$.

Musings on finite group theory. We can now try to factor groups into "prime factors".

Analogue: big number g, n = divisor of g, then $g = n \times g/n$ and if $1 < n < g$, we have made progress. If g has no factors other than 1 or g, g is prime.

In group theory, analogue of "prime number" must be "group $G \neq \{e\}$ such that the only normal subgroups are $\{e\}$ and G."

Such a group is called a *simple group* and the strategy for understanding all groups is:

(1) understand all simple groups,
(2) understand how to glue them together to make all groups.

(1) is apparently "done". Here is the answer: cyclic groups of prime order, A_n for $n \geq 5$, various matrix groups, 26 sporadic examples (including the Monster Group).

(2) is hard :-(The problem is

$$A_3 \lhd S_3 \quad \text{and} \quad S_3/A_3 = C_2$$

but $C_2 \times A_3 \cong C_2 \times C_3 \cong C_6 \ncong S_3$. In general, $N \lhd G$, $F = G/N$ but G might not be isomorphic to $F \times N$.

For a general finite group G, we can do the following:
Choose $G_1 \neq G$, $G_1 \lhd G$, G_1 large as possible.
Choose $G_2 \neq G_1$, $G_2 \lhd G_1$, G_2 large as possible.
Get $G_n \lhd \cdots \lhd G_2 \lhd G_1 \lhd G$, where G_n is simple.

Then G_i/G_{i+1} are all simple groups and G is built from these simple groups in a weird and complicated way.

The first isomorphism theorem. A group theory version of rank-nullity: say G, H are groups and $\varphi\colon G \to H$ is a group homomorphism. What new groups can we build? We have groups

$$\mathrm{im}(\varphi) \subseteq H$$

$$\ker(\varphi) \lhd G$$

and a factor group $G/\ker(\varphi)$.

Theorem 6.9 (The first isomorphism theorem)**.** *Let $\varphi\colon G \to H$ be a group homomorphism. Then*

$$G/\ker(\varphi) \cong \operatorname{im}(\varphi)$$

Corollary 6.10. *If $\varphi\colon G \to H$ is a group homomorphism and $|G|$ is finite, then*

$$|G| = |\ker(\varphi)| \times |\operatorname{im}(\varphi)|.$$

Note that this looks like rank-nullity theorem.

Proof of Theorem 6.9. We want to write down an isomorphism $\alpha\colon G/\ker(\varphi) \to \operatorname{im}(\varphi)$. Write $N = \ker(\varphi)$. We would like to say $\alpha(Ng) = \varphi(g)$. To prove this is well-defined, we will show that if $\varphi\colon G \to H$ is a group homomorphism and $N = \ker(\varphi)$, then $Nx = Ny$ if and only if $\varphi(x) = \varphi(y)$ for all $x, y \in G$. We have

$$\begin{array}{lll} Nx = Ny & \Leftrightarrow xy^{-1} \in N & \text{first year} \\ & \Leftrightarrow xy^{-1} \in \ker(\varphi) & \text{definition} \\ & \Leftrightarrow \varphi(xy^{-1}) = e & \text{definition} \\ & \Leftrightarrow \varphi(x)\varphi(y)^{-1} = e & \varphi \text{ group homomorphism} \\ & \Leftrightarrow \varphi(x) = \varphi(y) & \end{array}$$

Hence we can define $\alpha\colon G/N \to \operatorname{im}(\varphi)$ by $\alpha(Nx) = \varphi(x)$.

We will now show that α is a group isomorphism. Firstly, α is clearly surjective, because if $h \in \operatorname{im}(\varphi)$, then by definition, $h = \varphi(g)$ for some $g \in G$ and then $h = \alpha(Ng)$. To show injectivity, suppose $\alpha(Nx) = \alpha(Ny)$ for some $x, y \in G$. Then, by definition, $\varphi(x) = \varphi(y)$. But we showed above that this is equivalent to $Nx = Ny$.

Finally, we check that α is a group homomorphism:

$$\alpha(Nx)\alpha(Ny) = \varphi(x)\varphi(y) = \varphi(xy) = \alpha(Nxy) = \alpha((Nx)(Ny))$$

for any $x, y \in G$. □

Examples.

(1) $\operatorname{sgn}\colon S_n \to C_2$, $n \geq 2$. Then $\ker(\operatorname{sgn}) = A_n$ and $\operatorname{im}(\operatorname{sgn}) = C_2$, so we get

$$S_n/A_n \cong C_2.$$

(2) $\varphi\colon (\mathbb{R}, +) \to (\mathbb{C}^\times, \times)$, $\varphi(x) = e^{2\pi i x}$. Then

$$\operatorname{im}(\varphi) = \{z \in \mathbb{C}\colon |z| = 1\} = S^1$$

$$\ker(\varphi) = \{x\colon e^{2\pi i x} = 1\} = \mathbb{Z}$$

so the theorem shows $\mathbb{R}/\mathbb{Z} \cong S^1$.

(3) Is there a non-trivial group homomorphism $S_3 \to C_3$? New answer: Say $\varphi\colon S_3 \to C_3$ was a non-trivial group homomorphism. Then $\operatorname{im}(\varphi) \neq \{1\}$, so $\operatorname{im}(\varphi) = C_3$ and by the theorem $S_3/\ker(\varphi) \cong C_3$. But them $|\ker(\varphi)| = 2$ and $\ker(\varphi) = \{e, (ij)\}$ for some 2-cycle (ij). But this subgroup cannot be normal! For example, $(13)\langle(12)\rangle(13) \neq \langle(12)\rangle$, so no such φ exists.

Recall that the kernel of a group homomorphism is a normal subgroup. The converse is also true—every normal subgroup is a kernel of a group homomorphism.

Proposition 6.11. *Say $N \lhd G$. Define $\varphi\colon G \to G/N$ by $\varphi(x) = Nx$. Then φ is a group homomorphism and $\ker(\varphi) = N$.*

Proof. We have $\varphi(xy) = Nxy$ and

$$\varphi(x)\varphi(y) = (Nx)(Ny) = N(xy)$$

so φ is a group homomorphism. Moreover:

$$\ker(\varphi) = \{x\colon \varphi(x) = N\} = \{x\colon Nx = N\}$$

and $Nx = N$ if and only if $x \in N$, so $\ker(\varphi) = N$. □

Finally, given a group G, we will explain how to find all groups H such that there exists a surjective homomorphism $\varphi\colon G \to H$.

If $\varphi\colon G \to H$ is a surjection, then by the first isomorphism theorem $H = \mathrm{im}(\varphi) \cong G/\ker(\varphi)$. But $\ker(\varphi)$ is a normal subgroup and by Proposition 6.11 any normal subgroup is a kernel of a homomorphism. Therefore, all we need to do is:

(1) List all normal subgroups of G.
(2) For each $N \lhd G$, compute G/N to get the possible H's (up to isomorphism).

Example. Take $G = S_3$. Normal subgroups: $\langle e\rangle$ and G. Any other normal subgroups have size 2, 3. Easy to check S_3 has no normal subgroups of size 2 and A_3 is the only normal subgroup of size 3, $A_3 = \langle(123)\rangle$. Then

$$G/\{e\} \cong G = S_3$$

$$G/G \cong C_1$$

$$G/A_3 \cong C_2$$

is the complete list of groups H such that there exists a surjective homomorphism $S_3 \to H$.

7. Determinants

Consider the matrix

$$M = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

Recall that the determinant of M is:

$$\det(M) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}.$$

6 terms: each a product of 3 entries, 1 from each row, 1 from each column. Each product has a $\pm$ sign.

Here is an interesting way to think about this determinant! Each of the 6 terms in the sum $\det(M)$ gives us a permutation of $\{1, 2, 3\}$ in the following way: send $i \to j$ if a_{ij} is present in the term. (This is a permutation because each term has one entry in each row and one entry in each column.) Explicitly:

term	permutation	sign
$a_{11}a_{22}a_{33}$	e	$+$
$a_{12}a_{23}a_{31}$	(123)	$+$
$a_{13}a_{21}a_{32}$	(132)	$+$
$a_{13}a_{22}a_{31}$	(13)	$-$
$a_{12}a_{21}a_{33}$	(12)	$-$
$a_{11}a_{23}a_{32}$	(23)	$-$

Note that each sign is the signature of the corresponding permutation!

Hence, if $M = (a_{ij})$ as above, $1 \leq i, j \leq 3$, then we can write the determinant succinctly as

$$\det(M) = \sum_{\pi \in S_3} \operatorname{sgn}(\pi) a_{1\pi(1)} a_{2\pi(2)} a_{3\pi(3)}.$$

For example, if $\pi = (12)$ then $\pi(1) = 2$, $\pi(2) = 1$, $\pi(3) = 3$, so we get the term $-a_{12}a_{21}a_{33}$.

Let $A = (a_{ij})$ be an $n \times n$ matrix:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

Definition. The *determinant* of $A = (a_{ij})$ is

$$\det(A) = |A| = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{1\pi(1)} a_{2\pi(2)} \dots a_{n\pi(n)}$$

Examples.

($n = 1$) If $A = (a_{11})$, then $\det(A) = a_{11}$.
($n = 2$) If $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, then $\det(A) = +a_{11}a_{22} - a_{12}a_{21}$.
($n = 3$) We did this above already!
($n = 4$) Here $\det(A)$ is the sum of $4! = 24$ terms.

Remark. Where do a_{ij}'s live?

Answer 1: For applied purposes, they are all real or complex numbers.

Answer 2: They are all elements of an underlying *ground field* E. Some examples: $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}(i)$, $\mathbb{Z}/2\mathbb{Z}$, $\mathbb{Z}/p\mathbb{Z}$ etc.

The theory of vector spaces holds over a general field!

Our first main aim is to figure out how $|A|$ changes when we apply row operations.

Recall that if $A = (a_{ij})$ is an $n \times n$ matrix, its *transpose* is $A^T = (b_{ij})$ with $b_{ij} = a_{ji}$.

Proposition 7.1. *If A is $n \times n$, then $|A| = |A^T|$.*

Proof. Let $A = (a_{ij})$, $B = (b_{ij})$ with $b_{ij} = a_{ji}$. Then

$$\begin{aligned} |B| &= \sum_{\pi \in S_n} \mathrm{sgn}(\pi) b_{1\pi(1)} b_{2\pi(2)} \dots b_{n\pi(n)} \\ &= \sum_{\pi \in S_n} \mathrm{sgn}(\pi) a_{\pi(1)1} a_{\pi(2)2} \dots a_{\pi(n)n} \end{aligned}$$

Now the permutation that sends $\pi(i)$ to i for each i is the inverse of π. Therefore

$$|B| = \sum_{\pi \in S_n} \mathrm{sgn}(\pi) a_{1\sigma(1)} a_{2\sigma(2)} \dots a_{n\sigma(n)}$$

where $\sigma = \pi^{-1}$ and we have rearranged the a_{ij} in the product. Moreover, $\mathrm{sgn}(\pi) = \mathrm{sgn}(\pi^{-1})$, so we can sum over σ instead of π to get

$$|B| = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \dots a_{n\sigma(n)} = |A|$$

which completes the proof. □

(As a consequence, any result involving determinants and row operations will have an analogue with determinants and column operations.)

Proposition 7.2. *If B is obtained from A by swapping 2 rows (or 2 columns), then $|B| = -|A|$.*

Proof. Let us prove it for columns. The result for rows will follow from Proposition 7.1. Say $\tau = (rs)$ and let us swap the rth and sth column in A to get B. Set $B = (b_{ij})$, $A = (a_{ij})$. Then

$$b_{ij} = a_{i\tau(j)},$$

so

$$|B| = \sum_{\pi \in S_n} \mathrm{sgn}(\pi) b_{1\pi(1)} b_{2\pi(2)} \dots b_{n\pi(n)} = \sum_{\pi \in S_n} \mathrm{sgn}(\pi) a_{1\tau\pi(1)} a_{2\tau\pi(2)} \dots a_{n\tau\pi(n)}$$

Write $\sigma = \tau\pi$. then $\mathrm{sgn}(\sigma) = \mathrm{sgn}(\tau)\mathrm{sgn}(\pi) = -\mathrm{sgn}(\pi)$ and

$$|B| = \sum_{\sigma \in S_n} (-\mathrm{sgn}(\sigma)) b_{1\sigma(1)} b_{2\sigma(2)} \dots b_{n\sigma(n)} = -|A|,$$

as requested. □

Proposition 7.3.

(1) *If A has a row of 0's then* $\det(A) = 0$.
(2) *If A has 2 identical rows (or columns) then* $\det(A) = 0$.
(3) *If A is upper-triangular, i.e.*

$$A = \begin{pmatrix} a_{11} & & & \star \\ & a_{22} & & \\ & & \ddots & \\ 0 & & & a_{nn} \end{pmatrix}$$

(or A is lower triangular), then $\det(A) = \prod_{i=1}^{n} a_{ii} = a_{11} a_{22} \dots a_{nn}$.

Proof. For (1), $|A| = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{1\pi(1)} a_{2\pi(2)} \dots a_{n\pi(n)}$ but each term contains an element from the row of zeroes, so their product is 0 and $|A| = 0$.

For (2), say rows r and s of A are equal. If we swap those rows, then by Proposition 7.2 we change the sign of $\det(A)$, but we do not change A. Thus

$$\det(A) = -\det(A)$$

and $\det(A) = 0$ (if the field E has characteristic $\neq 2$).

For (3), if A is upper triangular, then $a_{ij} = 0$ if $i > j$ and it is not hard to check that every term in the sum must then vanish apart from the one corresponding to $\pi =$ identity. Hence

$$|A| = \operatorname{sgn}(e) a_{11} a_{22} \dots a_{nn} = a_{11} a_{22} \dots a_{nn}.$$

If A is lower triangular, then A^T is upper triangular, so we are done. □

We can now see the effect that elementary rows operations have on $|A|$.

Theorem 7.4.

(1) *If a row of A is multiplied by a scalar λ to get B, then $|B| = \lambda |A|$.*
(2) *If two rows of A are swapped to get B, then $|B| = -|A|$.*
(3) *If a scalar multiple of one row of A is added to another row to get B, then $|B| = |A|$.*
(4) *If B is obtained from A by any elementary row operation [i.e. part (1) with $\lambda \neq 0$ or (2) or (3) above], then $|A| \neq 0 \Leftrightarrow |B| \neq 0$.*

Proof. In (1), every term of $|A|$ gets multiplied by a factor of λ, so the sum gets multiplied by a factor of λ as well, and $|B| = \lambda |A|$.

Part (2) was Proposition 7.2.

For (3), suppose $\mu \times$ (row k) is added to (row l) for a scalar μ. Then $A = (a_{ij})$, $B = (b_{ij})$, and $b_{ij} = a_{ij}$ unless $i = l$ in which case $b_{lj} = a_{lj} + \mu a_{kj}$. Therefore,

$$\begin{aligned} |B| &= \sum_{\pi \in S_n} \operatorname{sgn}(\pi) b_{1\pi(1)} b_{2\pi(2)} \dots b_{n\pi(n)} \\ &= \sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{1\pi(1)} a_{2\pi(2)} \dots (a_{l\pi(l)} + \mu a_{k\pi(l)}) \dots a_{n\pi(n)} \\ &= |A| + \mu \times (\text{determinant of a matrix with rows } k \text{ and } l \text{ the same}) \\ &= |A| \end{aligned}$$

by Proposition 7.3 (2).

Finally, (4) follows immediately from (1)–(3). □

Remark. We can also deduce that if B is obtained from A by applying elementary column operations, then again

$$|A| \neq 0 \;\Leftrightarrow\; |B| \neq 0.$$

(Just apply $|A| = |A^T|$).

Expansion by rows (or columns). If $A = (a_{ij})$, $n \times n$, and if we now fix $1 \leq i, j \leq n$, we say that the (i, j)th *minor* of A is the $(n-1) \times (n-1)$ matrix that you get by removing the ith row and jth column of A.

Example. Take

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

Then (taking $i = 2$, $j = 3$), the $(2, 3)$th minor will be

$$\begin{pmatrix} 1 & 2 \\ 7 & 8 \end{pmatrix}$$

Notation: $A_{i,j}$ is the (i, j)th minor.

Proposition 7.5 (Laplace expansion of determinants)**.** *Let A be $n \times n$, $A = (a_{ij})$. Then:*

- *"Expansion by 1st row":*
$$|A| = a_{11}|A_{11}| - a_{12}|A_{12}| + a_{13}|A_{13}| + \cdots \pm a_{1n}|A_{1n}|,$$
- *"Expansion by ith row":*
$$|A| = (-1)^{i-1}a_{i1}|A_{i1}| - a_{i2}|A_{i2}| + a_{i3}|A_{i3}| + \cdots \pm a_{in}|A_{in}|,$$
- *"Expansion by jth column":*
$$|A| = (-1)^{j-1}a_{1j}|A_{1j}| - a_{2j}|A_{2j}| + a_{3j}|A_{3j}| + \cdots \pm a_{nj}|A_{ij}|,$$

Proof. We will first prove (1). We have:

$$|A| = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{1\pi(1)} a_{2\pi(2)} \dots a_{n\pi(n)}.$$

Strategy: each term is $\pm a_{1i} a_{2\pi(2)} \dots a_{n\pi(n)}$ for some i, so we can group together the a_{1i} terms and show that they sum to $a_{1i}|A_{1i}|$.

Terms featuring a_{11}. These are $a_{11} \sum_{\pi \in X} a_{2\pi(2)} \dots a_{n\pi(n)}$ where $X = \{\pi \in S_n \colon \pi(1) = 1\}$. Now:

$$A_{11} = \begin{pmatrix} a_{22} & a_{23} & \dots & a_{2n} \\ a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & & \vdots \\ a_{n2} & a_{n3} & \dots & a_{nn} \end{pmatrix}$$

and if $\pi(1) = 1$, then we can think of π as a permutation of $\{2, 3, \dots, n\}$, so the terms featuring a_{11} sum to $a_{11}|A_{11}|$.

Terms featuring a_{12}. Trick: swap columns 1 and 2 of A to get

$$B = \begin{pmatrix} a_{12} & a_{11} & a_{13} & \dots & a_{1n} \\ a_{22} & a_{21} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{n2} & a_{n1} & a_{n3} & \dots & a_{nn} \end{pmatrix}$$

which changes the determinant to $-|A|$. Now, use the previous part with matrix B instead to get that the terms mentioning a_{12} sum to $a_{12}|B_{11}| = -a_{12}|A_{12}|$.

Terms featuring a_{13}. Trick: first swap columns 2 and 3, then swap columns 1 and 2. This does not change the determinant and shows that the terms featuring a_{13} sum to $a_{13}|A_{13}|$.

Continue in this way to get (1).

For (2), use the same trick but with rows: swap rows i and $i-1$, then $i-1$ and $i-2$, etc., finally, 2 and 1. The ith row is now at the top. All other rows are still in order. Now apply (1) to the new matrix to get $(-1)^{i-1}|A| = a_{i1}|A_{i1}| - a_{i2}|A_{i2}| + a_{i3}|A_{i3}| + \cdots \pm a_{in}|A_{in}|$, as requested.

Finally, (3) follows from (2) using $|A^T| = |A|$ (Proposition 7.1). □

Remark. In a_{13} terms if we had only swapped columns 1 and 3, we would not have gotten A_{13} as a (1,1) minor of the new matrix! Hence we also swap columns 2 and 3.

Theorem 7.6. *For an $n \times n$ matrix A, the following are equivalent:*

(1) $|A| \neq 0$,
(2) *A is invertible,*
(3) *The system of equation $Ax = 0$ has only the solution $x = 0$,*
(4) *A can be reduced to I_n by elementary row operations.*

Proof. We know that (2), (3), and (4) are all equivalent by M1GLA (Geometry and Linear Algebra).

Assume (1), $|A| \neq 0$. Put A into echelon form, using elementary row operations to get a new matrix A'. By Theorem 7.4 (4), $|A'| \neq 0$. Hence last row of A' cannot be all 0 by Proposition 7.3, so

$$A' = \begin{pmatrix} 1 & & & \bigstar \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix}$$

and now it is easy to reduce A' to I_n using elementary row operations, so we have shown (4).

So now suppose (4), i.e. A can be reduced to I_n by elementary row operations. As $|I_n| = 1 \neq 0$, by Theorem 7.4, $|A| \neq 0$, so (1) holds. □

Corollary 7.7. *Suppose A is $n \times n$. If $Ax = 0$ has a non-zero solution x, then $|A| = 0$.*

We will now state the main theorem of this chapter.

Theorem 7.8. *If A, B are $n \times n$, then* $\det(AB) = \det(A)\det(B)$.

To prove it, we will need *elementary matrices*, which we define below.

If $r \neq 0$ and $1 \leq i \leq n$, define $A_i(r)$ = identity with ith row multiplied by r:

$$A_i(r) = \begin{pmatrix} 1 & & & & & \\ & 1 & & & 0 & \\ & & \ddots & & & \\ & & & r & & \\ & 0 & & & \ddots & \\ & & & & & 1 \end{pmatrix}$$

If $1 \leq i, j \leq n$, $i \neq j$, define $B_{i,j}$ = identity with ith and jth rows swapped (there is no way I am going to write this one down nicely).
Finally, if s is a scalar, $1 \leq i, j \leq n$, $i \neq j$, define $C_{i,j}(s)$ = identity with $s \times j$th row added to ith row:

$$C_{i,j}(s) = \begin{pmatrix} 1 & & & & & & \\ & 1 & & & & 0 & \\ & & \ddots & & & & \\ & s & & 1 & & & \\ & & & & 1 & & \\ & 0 & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

where the s is in the (i, j) entry of the matrix.

Examples. If $n = 2$:

$$A_1(7) = \begin{pmatrix} 7 & 0 \\ 0 & 1 \end{pmatrix}$$

$$A_1(-3) = \begin{pmatrix} 1 & 0 \\ 0 & -3 \end{pmatrix}$$

$$B_{12} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$C_{12}(3) = \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}$$

What happens when we multiply an elementary matrix by a matrix? Set

$$M = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

Then:

$$A_1(7)M = \begin{pmatrix} 7 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 7 & 14 \\ 3 & 4 \end{pmatrix}$$

which is M with 1st row multiplied by 7,

$$B_{12}M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix}$$

which is M with 1st and 2nd rows swapped,

$$C_{12}(3)M = \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 10 & 14 \\ 3 & 4 \end{pmatrix}$$

which is M with 3 times 2nd row added to 1st row.

Proposition 7.9. *Let M be an $n \times n$ matrix. An elementary row operation on M changes it to EM where E is an elementary matrix.*

Proof. Exercise! □

Proposition 7.10. *The determinants and inverses of elementary matrices are:*

matrix E	*determinant* $\det(E)$	*inverse* E^{-1}
$A_i(r)$	r	$A_i(1/r)$
B_{ij}	-1	B_{ij}
$C_{ij}(s)$	1	$C_{ij}(-s)$

Proof. For the determinants, you could use Theorem 7.4 (with I_n). For inverses, you could use Proposition 7.9.

Or you could check everything directly. □

The following result is very useful.

Proposition 7.11. *Every invertible matrix can be expressed as a product of elementary matrices.*

Proof. Let A be invertible. By Theorem 7.6, (2) implies (4), can reduce A to I_n using elementary row operations. By Proposition 7.9, the first row operation sends A to E_1A with E_1 an elementary matrix. The seconds row operations sends it to

$$E_2E_1A$$

where E_2 is a second elementary matrix. Continuing, we deduce:

$$I_n = E_kE_{k-1}\dots E_3E_2E_1A$$

and all the E_i are elementary matrices. Hence

$$A = (E_kE_{k-1}\dots E_2E_1)^{-1} = E_1^{-1}E_2^{-1}\dots E_{k-1}^{-1}E_k^{-1}$$

but by Proposition 7.10 the inverse of an elementary matrix is also an elementary matrix, so we are done! □

Example. Let $A = \begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix}$. Since $\det A \neq 0$, A is invertible.

By reducing A to the identity, we find:

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 0 & 2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

(where the consecutive matrices correspond to consecutive row operations we apply to reduce A to the identity).

Therefore, we have

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}^{-1}\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}^{-1}\begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

and we are done.

The following Proposition is a key step towards proving $\det(AB) = \det(A)\det(B)$ (Theorem 7.8).

Proposition 7.12. *If A is an $n \times n$ matrix and E is an elementary matrix, then* $\det(EA) = \det(E)\det(A)$.

Proof. Let $B = EA$. 3 cases:

(1) If $E = A_i(r)$, then $B = A$ with ith row multiplied by r, so $\det(B) = r\det(A) = \det(E)\det(A)$.
(2) If $E = B_{ij}$, then $B = A$ with ith and jth rows swapped, so $\det(B) = -\det(A) = \det(E)\det(A)$.
(3) If $E = C_{ij}(s)$, then similarly $\det(EA) = \det(A) = \det(E)\det(A)$.

□

Proposition 7.13. *Let A be invertible. By Proposition 7.11, can write $A = E_1 \dots E_k$, each E_i elementary. Then*

$$|A| = |E_1||E_2| \dots |E_k|.$$

Proof. Induction on k with 7.12 as the base step. □

We can finally prove that $\det(AB) = \det(A)\det(B)$.

Proof of Theorem 7.8. **Case 1.** If $|A| = 0$ or $|B| = 0$, then $|AB| = 0$ by Q6 from Problem Sheet 7.

Case 2. Say $|A| \neq 0$ and $|B| \neq 0$. By Proposition 7.11,

$$A = E_1E_2 \dots E_k$$
$$B = F_1F_2 \dots F_l$$

for E_i, F_j elementary matrices. Then

$$AB = E_1 \dots E_kF_1 \dots F_l$$

so 7.13 implies

$$|AB| = |E_1| \dots |E_k||F_1| \dots |F_l| = |A||B|$$

and we are done! □

Proposition 7.14. *Let P be an invertible $n \times n$ matrix. Then*

(1) $\det(P^{-1}) = 1/\det(P)$
(2) *If A is an $n \times n$ matrix, then* $\det(P^{-1}AP) = \det(A)$.

Proof. For (1), note that $PP^{-1} = I_n$ and so

$$1 = \det(I_n) = \det(PP^{-1}) = \det(P)\det(P^{-1}).$$

For (2):

$$\det(P^{-1}AP) = \det(P^{-1})\det(A)\det(P) = \det(A)\det(P^{-1})\det(P) = \det(A)$$

by part (1). □

8. Matrices and linear transformations

Reminder from M1P2 (Algebra 1). Say V is a finite-dimensional vector space (over a ground field E, where $E = \mathbb{R}$ is a fine choice). Say $T\colon V \to V$ is a linear map.

Let $B = \{v_1, v_2, \ldots, v_n\}$ be a basis of V. Say

$$\begin{aligned} Tv_1 &= a_{11}v_1 + a_{21}v_2 + \cdots + a_{n1}v_n, \\ Tv_2 &= a_{12}v_1 + a_{22}v_2 + \cdots + a_{n2}v_n, \\ &\vdots \\ Tv_n &= a_{1n}v_1 + a_{2n}v_2 + \cdots + a_{nn}v_n, \end{aligned}$$

where $a_{ij} \in E$.

The *matrix of T with respect to B* is

$$[T]_B = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix}$$

Proposition 8.1. *If $S, T\colon V \to V$ are linear maps, and $ST = S \circ T$ is the composition, then*

$$[ST]_B = [S]_B[T]_B.$$

(Even easier: $[S+T]_B = [S]_B + [T]_B$ and $[\lambda S]_B = \lambda[S]_B$.)

Consequences:

(1) Notation as above, and say $[T]_B = A$. Then $T^2 = T \circ T$ and $[T \circ T]_B = A^2$. More generally,

$$[T^k]_B = A^k.$$

Most general setting: say $p(x) = c_m x^m + c_{m-1}x^{m-1} + \cdots + c_1 x + c_0$ is a polynomial with coefficients $c_i \in E$. Let us define

$$p(T) = c_m T^m + \cdots + c_1 T + c_0 I$$

where $T\colon V \to V$ as before and $I\colon V \to V$ is the identity function $I(v) = v$. Note that $[I]_B = I_n$, the identity matrix. Then

$$p(T)\colon V \to V$$

is a linear map. Also

$$p(A) = c_m A^m + \cdots + c_1 A + c_0 I_n$$

and one checks $[p(T)]_B = p(A)$.

(2) Notation as above. Define $GL(V)$ to be the set of invertible linear maps $V \to V$. This is in fact a group under composition of functions. This is a group and the map

$$GL(V) \to GL(n, E)$$

defined by $T \mapsto [T]_B$ is an isomorphism of groups.

Change of basis. Say V has 2 bases $B = \{e_1, e_2, \ldots, e_n\}$, $C = \{f_1, f_2, \ldots, f_n\}$, and

$$f_1 = p_{11}e_1 + p_{21}e_2 + \cdots + p_{n1}e_n$$

$$\vdots$$

$$f_n = p_{1n}e_1 + p_{2n}e_2 + \cdots + p_{nn}e_n.$$

The change of basis matrix P from B to C is

$$P = \begin{pmatrix} p_{11} & \cdots & p_{1n} \\ p_{21} & \cdots & p_{2n} \\ \vdots & & \vdots \\ p_{n1} & \cdots & p_{nn} \end{pmatrix}$$

Proposition 8.2. *If $T\colon V \to V$ is a linear map, then $[T]_C = P^{-1}[T]_B P$.*

As an amazing consequence,

$$\det([T]_C) = \det([T]_B)$$

(by Proposition 7.14) i.e. *a linear map $T\colon V \to V$ has a well-defined determinant.*

Definition. Two $n \times n$ matrices A and B are *similar* if there exists an invertible P such that $B = P^{-1}AP$.

Remark.

(1) If we define $A \sim B \Leftrightarrow A$ is similar to B, then $\sim$ is an equivalence relation.
(2) The matrices $[T]_B$ and $[T]_C$ are similar.
(3) By Proposition 7.14, similar matrices have the same determinant.

Definition. Let $T\colon V \to V$ be a linear map. The *determinant* $\det(T)$ of T is defined to be $\det([T]_B)$ for B any basis of V.

This is well-defined by remark (2) above!

Example. Let V be a vector space of polynomials over R of degree ≤ 2. Define $T\colon V \to V$ by $T(p(x)) = p(3x+1)$. Find $\det(T)$.

Solution. Set $B = \{1, x, x^2\}$. Need $[T]_B$. We have

$$T(1) = 1$$
$$T(x) = 3x + 1$$
$$T(x^2) = (3x+1)^2 = 9x^2 + 6x + 1$$

so we have

$$[T]_B = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 3 & 6 \\ 0 & 0 & 9 \end{pmatrix}$$

and we have $\det(T) = 1 \times 3 \times 9 = 27$.

9. Characteristic polynomial

Let $T\colon V \to V$ be a linear transformation where V is a finite-dimensional vector space.

Definition.

(1) We say $v \in V$ is an *eigenvector of* T if $v \neq 0$ and $T(v) = \lambda v$ for some $\lambda \in E$ (ground field).
(2) We say λ as above is an eigenvalue of T.
(3) The characteristic polynomial of T is the determinant

$$\det(xI - T).$$

Proposition 9.1.

(1) *The eigenvalues of T are the zeroes of the characteristic polynomial of T.*
(2) *If λ is an eigenvalue of T and we define $E_\lambda \subseteq V$*

$$E_\lambda = \{v \in V\colon T(v) = \lambda v\}.$$

Then E_λ is a vector subspace of V.
(3) *The matrix $[T]_B$ is diagonal if and only if the basis B consists of eigenvectors of T.*

Proof. (1) and (3) were proved in M1P2 (Algebra I).

To prove (2):

$$\begin{aligned} E_\lambda &= \{v \in V\colon T(v) = \lambda v\} \\ &= \{v \in V\colon (T - \lambda I)(v) = 0\} \\ &= \text{kernel of the linear map } T - \lambda I\colon V \to V \end{aligned}$$

so it is a subspace. □

Definition. The subspace E_λ is called the λ-*eigenspace* of T.

Corollary 9.2. *Let V be a non-zero finite-dimensional vector space over E and say $T\colon V \to V$ is a linear transformation. If E is algebraically closed, then T has an eigenvalue $\lambda \in E$.*

Proof. The characteristic polynomial of T is a non-constant polynomial over E, so it has a root λ. So use 9.1 (1). □

Note that $\mathbb{C}$ is algebraically closed, so this holds for $E = \mathbb{C}$.

Example (of an eigenspace)**.** Let $E = \mathbb{R}$, $V = \mathbb{R}^2$. Let $T\colon V \to V$ be

$$T\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}$$

Take $B = \left\{\binom{1}{0}, \binom{0}{1}\right\}$ so that $[T]_B$ is the matrix above. The eigenvalues of T are just 2, characteristic polynomial $= (x-2)^2$. Then

$$E_2 = \left\{\begin{pmatrix} a \\ b \end{pmatrix} \colon \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = 2\begin{pmatrix} a \\ b \end{pmatrix}\right\}$$

Need

$$\begin{cases} 2a + b &= 2a \\ 2b &= 2b \end{cases}$$

The solution is $b = 0$, so the eigenspace is

$$E_2 = \left\{\begin{pmatrix} a \\ 0 \end{pmatrix}\right\}$$

Example (of an eigenspace)**.** For $T = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ the 2-eigenspace is just $\mathbb{R}^2$!

Diagonalisation. Recall that some matrices are diagonalisable and some are not. The aim of this subsection is to determine when we can diagonalise a matrix.

Proposition 9.3. *Let $T\colon V \to V$ a linear map, V finite-dimensional vector space. Suppose $v_1, v_2, \ldots, v_k \in V$ are all eigenvectors for T with corresponding eigenvalues $\lambda_1, \ldots, \lambda_k$, i.e.*

$$Tv_i = \lambda_i v_i$$

Suppose all the λ_i's are distinct. Then $v_1, v_2, \ldots, v_k$ are linearly independent.

Proof. Induction on k. True for $k = 1$, because $v_1 \neq 0$. Suppose that

$$v_1, v_2, \ldots, v_k$$

are linearly independent. Assume for a contradiction that

$$r_1 v_1 + r_2 v_2 + \cdots + r_k v_k = 0 \tag{1}$$

for $r_i \in E$ and not all equal to 0. Applying T to the equation, we get

$$r_1\lambda_1 v_1 + r_2\lambda_2 v_2 + \cdots + r_k\lambda_k r_k = 0. \tag{2}$$

Thus subtracting (2) from λ_k times (1):

$$r_1(\lambda_k - \lambda_1)v_1 + r_2(\lambda_k - \lambda_2)v_2 + \cdots + r_{k-1}(\lambda_k - \lambda_{k-1})v_{k-1} = 0$$

so by linear independence of $v_1, \ldots, v_{k-1}$, we get

$$r_i(\lambda_k - \lambda_i) = 0$$

for $i = 1, 2, \ldots, k-1$. But the λ_j's are all distinct, so $\lambda_k - \lambda_i \neq 0$ for $i = 1, 2, \ldots, k-1$, and hence

$$r_1 = r_2 = \cdots = r_{k-1} = 0.$$

Substituting back into (2), we deduce that $r_k v_k = 0$. But v_k is an eigenvector, so $v_k \neq 0$, and $r_k = 0$. Thus we have shown that $r_i = 0$ for $i = 1, 2, \ldots, k$, which contradicts our assumption. □

Corollary 9.4.

(1) *Say* $\dim V = n$ *and suppose* $T\colon V \to V$ *has a characteristic polynomial with distinct roots in* E*. Then there exists a basis* B *of* V *such that* $[T]_B$ *is diagonal.*
(2) *Let* A *be an* $n \times n$ *matrix and suppose the characteristic polynomial of* A *has* n *distinct roots in* E*. Then there exists an invertible* P *such that* $P^{-1}AP$ *is diagonal.*

Proof. For (1), say the roots are $\lambda_1, \ldots, \lambda_n$ are distinct. Choose $v_1, \ldots, v_n$ eigenvectors such that $T(v_i) = \lambda_i v_i$. Then by Proposition 9.3, the v_i are linearly independent. But there is n of them, so they form a basis B and $[T]_B$ is diagonal.

For (2), set $V = E^n$ and define $T\colon V \to V$ by $T(v) = Av$. The result follows from (1). □

Example. Say A is an upper triangular matrix, and diagonal entries are distinct. Then A is diagonalisable. For example:

$$\begin{pmatrix} -3 & 2 & 9 \\ 0 & 4 & 6 \\ 0 & 0 & 1 \end{pmatrix}$$

is diagonalisable, i.e. there exists P such that $P^{-1}AP$ is diagonal.

This is because the characteristic polynomial of

$$A = \begin{pmatrix} \lambda_1 & & & \bigstar \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}$$

is $(x - \lambda_1)(x - \lambda_2) \ldots (x - \lambda_n)$. Now, apply 9.4, since λ_i are distinct.

Problem with our toolkit so far:

$$\begin{pmatrix} -3 & 2 & 9 \\ 0 & 4 & 6 \\ 0 & 0 & -3 \end{pmatrix}$$

has 2 equal eigenvalues. Is that matrix diagonalisable or not?

Algebraic and geometric multiplicities of eigenvalues. Set-up: V is a finite-dimensional vector space over E and $T\colon V \to V$ is a linear map. Moreover, $p(x)$ is the characteristic polynomial of T and $\lambda \in E$ is an eigenvalue for T. Since $p(\lambda) = 0$, we can factor $(x - \lambda)$ out of $p(x)$, and write:

$$p(x) = (x - \lambda)^{a(\lambda)} q(x)$$

with $a(\lambda) \in \mathbb{Z}_{\geq 1}$.

We call the number $a(\lambda)$ the *algebraic multiplicity* of λ. By convention: if $\lambda \in E$ is not a root of $p(x)$, i.e. not an eigenvalue, then $a(\lambda) = 0$.

If $n = \dim V$, then n is the degree of $p(x)$. Therefore:

$$\sum_{\lambda \in E} a(\lambda) \leq n$$

with equality if E is algebraically closed or, more generally, if $p(x)$ factors as a product of linear factors.

Also, recall the λ-eigenspace

$$E_\lambda = \{v \in V : T(v) = \lambda v\} \subseteq V.$$

Set $g(\lambda) = \dim E_\lambda$. Note that $g(\lambda) \in \mathbb{Z}_{\geq 1}$, because λ is an eigenvalue, so there exists an eigenvector v with eigenvalue λ, and $0 \neq v \in E_\lambda$.

We call the number $g(\lambda)$ the *geometric multiplicity* of λ. Again, by convention $g(\lambda) = 0$ if $\lambda \in E$ is not an eigenvalue.

Note that if A is an $n \times n$ matrix, define $a(\lambda)$ and $g(\lambda)$ in the same way, using $T : E^n \to E^n$ given by $T(v) = Av$.

Example. Let $A = \begin{pmatrix} 3 & 1 \\ 0 & 4 \end{pmatrix}$ whose characteristic polynomial is $p(x) = (x-3)(x-4)$. Therefore, $a(3) = a(4) = 1$ (eigenvalues are 3 and 4).

Check that $g(3) = 1$ (this is clear as $g(3) \geq 1$ and conversely if $g(3) = 2$, then $E_3 = \mathbb{R}^2$ and A would be $3I$) and similarly $g(4) = 1$.

Now, let $A = \begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}$ whose characteristic polynomial is $p(x) = (x-3)^2$. Therefore, $a(3) = 2$ (eigenvalue is 3). But $g(3) = 1$, because:

$$E_3 = \left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\rangle.$$

Proposition 9.5. *Say V is a finite dimensional vector space and $T : V \to V$ is a linear map with λ an eigenvalue. Then $g(\lambda) \leq a(\lambda)$.*

Proof. Say $g(\lambda) = r$. Then the eigenspace $E_\lambda \subseteq V$ is r-dimensional, so pick a basis $\{v_1, v_2, \ldots, v_r\}$ for E_λ. Extend it to a basis for V:

$$B = \{v_1, \ldots, v_r, w_1, \ldots, w_s\}.$$

What can we say about $[T]_B$?

$$\begin{aligned}
T(v_1) &= \lambda v_1 \\
&\vdots \\
T(v_r) &= \lambda v_r \\
T(w_1) &= a_{11}v_1 + \cdots + a_{r1}v_r + b_{11}w_1 + b_{21}w_2 + \cdots + b_{s1}w_s \\
&\vdots \\
T(w_s) &= a_{1s}v_1 + \cdots + a_{rs}v_r + b_{1s}w_1 + b_{2s}w_2 + \cdots + b_{ss}w_s
\end{aligned}$$

Therefore:

$$[T]_B = \left(\begin{array}{ccc|ccc} \lambda & & 0 & a_{11} & \dots & a_{1s} \\ & \ddots & & \vdots & & \vdots \\ 0 & & \lambda & a_{r1} & \dots & a_{rs} \\ \hline & & & b_{11} & \dots & b_{1s} \\ & 0 & & \vdots & & \vdots \\ & & & b_{s1} & \dots & b_{ss} \end{array}\right)$$

So if we let $A = (a_{ij})$ and $B = b_{ij}$, then

$$[T]_B = \left(\begin{array}{c|c} \lambda I_r & A \\ \hline 0 & B \end{array}\right)$$

We claim that the characteristic polynomial of $[T]_B$ is $(x-\lambda)^r \times$(characteristic polynomial of matrix B). Indeed, the characteristic polynomial of $[T]_B$ is the determinant of

$$xI - [T]_B = \left(\begin{array}{c|c} (x-\lambda)I_r & -A \\ \hline 0 & xI_s - B \end{array}\right)$$

which is exactly $(x-\lambda)^n \det(xI_s - B)$ by Q5 from Problem Sheet 7. □

Theorem 9.6. *Say* $\dim V = n$ *and* $T\colon V \to V$ *linear map and* $\lambda_1, \dots, \lambda_m$ *are distinct eigenvalues and* $p(x)$ *is the characteristic polynomial of* T*,*

$$p(x) = (x-\lambda_1)^{a(\lambda_1)} \dots (x-\lambda_m)^{a(\lambda_m)}.$$

Then the following are equivalent:

(1) *There exists a basis* B *of* V *consisting of eigenvectors (i.e.* T *is diagonalisable, i.e.* $[T]_B$ *is diagonal);*
(2) $\sum_{i=1}^{m} g(\lambda_i) = n$*;*
(3) $g(\lambda_i) = a(\lambda_i)$ *for all* i*.*

Proof. It is easy to show that (2) is equivalent to (3):

$$\sum_{i=1}^{m} a(\lambda_i) = \deg p(x) = n$$

and $g(\lambda_i) \leq a(\lambda_i)$ for all i, so

$$\sum_{i=1}^{m} g(\lambda_i) \leq \sum_{i=1}^{m} a(\lambda_i) = n$$

with equality if and only if $g(\lambda_i) = a(\lambda_i)$.

We will now show that (1) is equivalent to (2). First, assume (1) and say B is a basis of eigenvectors for T. We will show (2). If $B_i \subseteq B$ consists of elements with eigenvalues λ_i, then

$$B_i \subseteq E_{\lambda_i} = \lambda_i\text{-eigenspace}$$

and B is the disjoint union of the B_i's. Thus

$$n = |B| = \sum_{i=1}^{m} |B_i|.$$

Now, $B_i \subseteq E_{\lambda_i}$ is a linearly independent subset, so

$$|B_i| \leq \dim(E_{\lambda_i}) = g(\lambda_i).$$

Therefore:

$$\begin{aligned} n &= \sum_{i=1}^{m} |B_i| \\ &\leq \sum_{i=1}^{m} g(\lambda_i) \\ &\leq \sum_{i=1}^{m} a(\lambda_i) \\ &= n \end{aligned}$$

so equalities hold everywhere and (2) holds.

Finally, we suppose (2), $\sum_{i=1}^{m} g(\lambda_i) = n$, and show (1). For each i, $1 \leq i \leq m$, let B_i be a basis for E_{λ_i}, and let $B = \bigcup_{i=1}^{m} B_i$. Note that one eigenvector cannot have 2 eigenvalues, so the B_i are disjoint. Therefore:

$$|B| = \sum_{i=1}^{m} g(\lambda_i) = n$$

and it suffices to prove that B is a linearly independent set. So set $|B_i| = n_i$, $B_i = \{b_{i1}, b_{i2}, \ldots, b_{in_i}\}$, and say we have a linear combination

$$\sum_{i=1}^{m} \left(\sum_{j=1}^{n_i} \lambda_{ij} b_{ij} \right) = \sum_{j=1}^{n_1} \lambda_{1j} b_{1j} + \cdots + \sum_{j=1}^{n_m} \lambda_{mj} b_{mj} = 0.$$

Now, set $v_i = \sum_{j=1}^{n_i} \lambda_{ij} b_{ij}$ for $i = 1, 2, \ldots, m$. Then

$$v_1 + v_2 + \cdots + v_m = 0$$

with each $v_i \in E_{\lambda_i}$ and, in particular, $T(v_i) = \lambda_i v_i$. If any of the v_i were non-zero, then we would have a non-zero linear relation $\lambda_1 v_1 + \cdots + \lambda_m v_m = 0$ on eigenvectors with distinct eigenvalues, contradicting Proposition 9.3. Thus for $i = 1, 2, \ldots, m$ we have $v_i = 0$ and

$$\sum_{j=1}^{n_i} \lambda_{ij} b_{ij} = v_i = 0,$$

so since B_i is a linearly independent set, $\lambda_{ij} = 0$ for $j = 1, \ldots, n_i$.

Thus B is a linearly independent set of size n in V and (1) follows. □

We can use this theorem to check whether explicit linear maps or matrices can be diagonalised.

Example. Consider the matrix:

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

Its characteristic polynomial is $(x+2)^2(x-4)$, so $a(-2) = 2$, $a(4) = 1$. We know that $1 \leq g(4) \leq a(4) = 1$, so $g(4) = 1$.

What about $g(-2)$? Note that E_{-2} is the kernel of the map given by the matrix:

$$A + 2I = \begin{pmatrix} -1 & 1 & -1 \\ -7 & 7 & -1 \\ -6 & 6 & 0 \end{pmatrix}$$

and the rank of this matrix is 2 (to see this: look at the columns). Therefore, by the Rank-Nullity Theorem, its nullity is 1 and $g(-2) = 1 < a(-2)$. Hence the matrix cannot by diagonalised by Theorem 9.6.

10. Upper-triangularisation

A matrix is *upper triangular* if it is of the form:

$$A = \begin{pmatrix} \lambda_1 & & & \bigstar \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}$$

In this chapter, we will show that any matrix over $\mathbb{C}$ is similar to an upper triangular matrix. We first prove some basic properties of upper triangular matrices.

Proposition 10.1. *Suppose*

$$A = \begin{pmatrix} \lambda_1 & & & \bigstar \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix} \qquad B = \begin{pmatrix} \mu_1 & & & \bigstar \\ & \mu_2 & & \\ & & \ddots & \\ 0 & & & \mu_n \end{pmatrix}$$

Then:

(1) *the characteristic polynomial of A is $(x-\lambda_1)(x-\lambda_2)\dots(x-\lambda_n)$ and the determinant is $\lambda_1 \dots \lambda_n$*

(2) $AB = \begin{pmatrix} \lambda_1\mu_1 & & & \bigstar \\ & \lambda_2\mu_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n\mu_n \end{pmatrix}$

$$(3)\quad A^k = \begin{pmatrix} \lambda_1^k & & & \star \\ & \lambda_2^k & & \\ & & \ddots & \\ 0 & & & \lambda_n^k \end{pmatrix}$$

Proof. Easy exercise. □

We have seen that there exist matrices that you cannot diagonalise, for example

$$\begin{pmatrix} 7 & 1 \\ 0 & 7 \end{pmatrix}$$

has $g(7) = 1$, $a(7) = 2$, $g(7) < a(7)$.

However

Theorem 10.2. *Say $E = \mathbb{C}$ (or any algebraically closed field).*

1. *If V is an n-dimensional vector space over E and $T\colon V \to V$ is a linear map, then there exists a basis B of V such that $[T]_B$ is upper triangular.*
2. *If A is an $n \times n$ matrix over E, then there exists invertible $n \times n$ matrix P such that $P^{-1}AP$ is upper-triangular.*

Proof. First, note that (1) for $n \Leftrightarrow$ (2) for n. We will prove them both by induction on n. More precisely, we will show that (2) for $n-1$ implies (1) for n.

Base case $n = 1$**.** Suppose $\dim V = 1$ and pick any $0 \neq v \in V$; this is a basis and $[T]_B = (\lambda)$ is uppertriangular.

Inductive step. Assume that (2) is true for $\dim V < n$ and prove (1) for $\dim V = n$. Take $T\colon V \to V$ linear, $\dim(V) = n$. By Proposition 9.1, T has an eigenvalue λ (take any root of the characteristic polynomial). Choose an eigenvector $0 \neq v \in V$ such that $Tv = \lambda v$ and extend to a basis of V, $\{v, e_2, e_3, \ldots, e_n\} = C$. What is $[T]_C$? We know that $Tv = \lambda v$, but we do not know anything about Te_i. So the matrix looks like:

$$[T]_C = \left(\begin{array}{c|ccccc} \lambda & ? & ? & \ldots & ? & ? \\ \hline 0 & & & & & \\ 0 & & & & & \\ \vdots & & & M & & \\ 0 & & & & & \\ 0 & & & & & \end{array}\right)$$

where M is some matrix, $(n-1)\times(n-1)$. By the inductive hypothesis, there exists invertible $(n-1) \times (n-1)$ matrix Q such that $Q^{-1}MQ$ is upper triangular. Set

$$P = \left(\begin{array}{c|ccccc} 1 & 0 & 0 & \dots & 0 & 0 \\ \hline 0 & & & & & \\ 0 & & & & & \\ \vdots & & & Q & & \\ 0 & & & & & \\ 0 & & & & & \end{array}\right)$$

then

$$P^{-1} = \left(\begin{array}{c|ccccc} 1 & 0 & 0 & \dots & 0 & 0 \\ \hline 0 & & & & & \\ 0 & & & & & \\ \vdots & & & Q^{-1} & & \\ 0 & & & & & \\ 0 & & & & & \end{array}\right)$$

and

$$P^{-1}[T]_C P = \left(\begin{array}{c|ccccc} \lambda & ? & ? & & \dots & ? & ? \\ \hline 0 & & & & & & \\ 0 & & & & & & \\ \vdots & & & Q^{-1}MQ & & & \\ 0 & & & & & & \\ 0 & & & & & & \end{array}\right)$$

which is upper triangular. This shows (2) for n. To see (1), take B to be the basis you get from C after applying P. □

In practice, how do we actually upper-triangularise a matrix?

Basic strategy: find as many linearly independent eigenvectors as you can—use those as the first few elements, and extend to a basis. Will it work?

Say $\dim V = n$. If you find n linearly independent eigenvectors, then $[T]_B$ is diagonal! If you only find $n-1$, say e_1, e_2, …, e_{n-1} eigenvectors, then extend to a basis $e_1, \dots, e_n$ of V. Then

$$[T]_B = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 & ? \\ 0 & \lambda_2 & \dots & 0 & ? \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda_{n-1} & ? \\ 0 & 0 & \dots & 0 & ? \end{pmatrix}$$

because $Te_i = \lambda_i e_i$ if $i < n$ and e_n is anything. Thus $[T]_B$ is upper-triangular.

If you can only find $n-2$, then you need to look at the proof, and decide where to go next.

Cayley–Hamilton theorem. If A is an $n \times n$ matrix, $A = (a_{ij})$ and if $f(x) = a_m x^m + \dots + a_1 x + a_0$ is any polynomial (with coefficients in the ground field), then it makes sense to talk about

$$f(A) = a_m A^m + \dots + a_1 A + a_0 I_n.$$

Theorem 10.3 (Cayley–Hamilton)**.** *Let A be an $n \times n$ matrix and let $p(x)$ be the characteristic polynomial, $p(x) = \det(xI - A)$. Then $p(A) = 0$, the zero matrix.*

Example. Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. Then $p(x) = x^2 - 5x - 2$, so

$$p(A) = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^2 - \begin{pmatrix} 5 & 10 \\ 15 & 20 \end{pmatrix} - \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 7 & 10 \\ 15 & 22 \end{pmatrix} - \begin{pmatrix} 7 & 10 \\ 15 & 22 \end{pmatrix} = 0$$

as expected.

A bogus proof the Cayley–Hamilton Theorem 10.3 is to say $p(x) = \det(xI - A)$, so

$$p(A) = \det(AI - A) = \det(A - A) = \det(0) = 0.$$

However, this makes no sense, since $p(A)$ should be a matrix not a number. The reason this is bogus is that x is supposed to be a number, so you cannot substitute A for x in the determinant. What does $x - a_{11}$ mean if $x = A$?

The following proposition is a preparation for the proof.

Proposition 10.4. *Let A and B be similar $n \times n$ matrices, i.e. $B = P^{-1}AP$ for some invertible $n \times n$ matrix P. Then*

(1) *A and B have the same characteristic polynomial,*
(2) *if $f(x)$ is any polynomial, then*

$$f(B) = P^{-1}f(A)P,$$

(3) *if $f(B) = 0$, then $f(A) = 0$.*

Proof. For (1), we have:

$$\begin{array}{lll} \det(xI - B) & = \det(xI - P^{-1}AP) & \text{definition} \\ & = \det(P^{-1}xIP - P^{-1}AP) & P^{-1}IP = I \\ & = \det(P^{-1}(xI - A)P) & \\ & = \det(xI - A) & \text{by 7.14(2)} \end{array}$$

For (2), first note:

$$B^k = P^{-1}A^kP$$

for all $k \geq 0$. (Easy proof by induction on k.) Now, if $f(x) = a_m x^m + \cdots + a_1 x + a_0$, then

$$f(B) = a_m P^{-1}A^m P + \cdots + a_1 P^{-1}AP + a_0 I = P^{-1}(a_m A^m + \cdots + a_1 A + a_0)P = P^{-1}f(A)P.$$

Finally, (3) follows from (2). □

Proof of Cayley–Hamilton Theorem 10.3. Let A be $n \times n$, with characteristic polynomial $p(x)$. By Theorem 10.2, there exists an invertible P such that $P^{-1}AP = B$ is upper triangular (with diagonal entries $\lambda_1, \ldots, \lambda_n$). [Note that we pretend the ground field is algebraically closed in this part of the proof.] We will show that $P(B) = 0$. That suffices by Proposition 10.4 (3). Since A and B are similar, $p(x)$ is the characteristic polynomial of B by Proposition 10.4 (1). Since $xI - B$ is upper-triangular:

$$p(x) = (x - \lambda_1)(x - \lambda_2)\ldots(x - \lambda_n).$$

To show that $p(B) = 0$ (the zero matrix), we will show that $p(B)v = 0$ (the zero vector) for any vector v. We have:

$$p(B) = (B - \lambda_1 I)(B - \lambda_2 I) \dots (B - \lambda_n I),$$

so for any vector $v = (\alpha_1, \alpha_2, \dots, \alpha_n)^T$:

$$(B - \lambda_n I)v = \begin{pmatrix} \lambda_1 - \lambda_n & & & & \bigstar \\ & \lambda_2 - \lambda_n & & & \\ & & \ddots & & \\ & & & \lambda_{n-1} - \lambda_n & \\ \mathbf{0} & & & & 0 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{n-1} \\ \alpha_n \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{n-1} \\ 0 \end{pmatrix}$$

Next step:

$$(B-\lambda_{n-1}I)(B-\lambda_n I)v = \begin{pmatrix} \lambda_1 - \lambda_{n-1} & & & & \bigstar \\ & \lambda_2 - \lambda_{n-1} & & & \\ & & \ddots & & \\ & & & 0 & \star \\ \mathbf{0} & & & & \lambda_n - \lambda_{n-1} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{n-1} \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_{n-2} \\ 0 \\ 0 \end{pmatrix}$$

It is easy to check that this pattern continues and hence deduce that:

$$(B - \lambda_2 I) \dots (B - \lambda_n I)v = \begin{pmatrix} \delta_1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

The final step shows that

$$p(B)v = (B-\lambda_1 I)(B-\lambda_2 I) \dots (B-\lambda_n I)v = \begin{pmatrix} 0 & & & \bigstar \\ & \lambda_2 - \lambda_1 & & \\ & & \ddots & \\ \mathbf{0} & & & \lambda_n - \lambda_1 \end{pmatrix} \begin{pmatrix} \delta_1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

and this completes the proof. □

11. Jordan Canonical Form

Idea: try to find out when two matrices are similar. Say A and B are similar matrices, i.e. $B = P^{-1}AP$. What do they have in common?

Proposition 11.1. *Let A and B be similar matrices. Then A and B have the same:*

(1) *determinant*
(2) *characteristic polynomial*
(3) *eigenvalues and algebraic multiplicities*
(4) *geometric multiplicities*

(5) *rank and nullity*
(6) *trace*

Proof. Firstly, (1) is 7.14 (2), (2) is 10.4, and (3) follows from (2).

For (4), say $B = P^{-1}AP$ and $\lambda \in E$. We will show that $\dim E_\lambda(A) = \dim E_\lambda(B)$, where $E_\lambda(M)$ is the λ-eigenspace for matrix M. Take $v \in E_\lambda(B)$ so that $Bv = \lambda v$. Since $B = P^{-1}AP$, $PB = AP$ and

$$APv = PBv = P(Bv) = P(\lambda v) = \lambda Pv$$

so $Pv \in E_\lambda(A)$. Therefore, $v \mapsto Pv$ is a map $E_\lambda(B) \to E_\lambda(A)$. It is easy to check it is an invertible linear map with the inverse given by $v \mapsto P^{-1}v$. Therefore, $\dim E_\lambda(B) = \dim E_\lambda(A)$.

For (5), we first check that the nullities coincide:

$$\begin{aligned} \mathrm{null}A &= \dim \ker A \\ &= \dim E_0(A) \quad \text{definition} \\ &= \dim E_0(B) \quad \text{by (4)} \\ &= \dim \ker B \quad \text{definition} \\ &= \mathrm{null}B \end{aligned}$$

and by the rank-nullity theorem the ranks also coincide.

Finally, to show (6) we note that the trace is the coefficient of x^{n-1} in the characteristic polynomial:

$$\det(xI - A) = \begin{pmatrix} x - a_{11} & -a_{12} & -a_{13} & \dots & -a_{1n} \\ -a_{21} & x - a_{22} & & & \\ -a_{31} & & \ddots & & \\ \vdots & & & & \\ -a_{n1} & & & & x - a_{nn} \end{pmatrix} = x^n - (a_{11} + \cdots + a_{nn})x^{n-1} + \cdots .$$

But by (2) the characteristic polynomials of A and B coincide. □

Is the converse true? Say A and B are $n \times n$ matrices with the same characteristic polynomial, rank, nullity etc. Is A similar to B? No! Take the following matrices:

$$A = \begin{pmatrix} 7 & 1 & 0 & 0 \\ 0 & 7 & 0 & 0 \\ 0 & 0 & 7 & 1 \\ 0 & 0 & 0 & 7 \end{pmatrix} \qquad B = \begin{pmatrix} 7 & 1 & 0 & 0 \\ 0 & 7 & 1 & 0 \\ 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 7 \end{pmatrix}$$

We have $\det A = \det B = 7^4$ and their characteristic polynomials are $(x-7)^4$. The only eigenvalue is 7 and its algebraic multiplicity is 4. Moreover, $g(7) = 2$ for both A and B:

$$E_7(A) = \langle e_1, e_3 \rangle, \quad E_7(B) = \langle e_1, e_4 \rangle.$$

Since $\det A = \det B \neq 0$, their rank is 4 and their nullity is 0. Finally, their traces are both 28.

However, A and B are not similar! For example,

$$(A-7I)^2 = 0$$

$$(B-7I)^2 \neq 0$$

so they cannot be similar by Proposition 10.4 (3).

In general, how do we check 2 matrices are similar? One way—put them both into *Jordan Canonical Form*. The rest of this chapter will explain this.

Jordan blocks. Ingredients: integer $n \geq 1$, $\lambda \in E$. We define a *Jordan block* to be

$$J_n(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & & & 0 \\ & \lambda & 1 & 0 & & \\ & & \lambda & 1 & & \\ & & & \ddots & & \\ & & & & \lambda & 1 \\ 0 & & & & & \lambda \end{pmatrix}$$

Examples.

$$J_2(7) = \begin{pmatrix} 7 & 1 \\ 0 & 7 \end{pmatrix} \quad J_4(0) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad J_1\left(3\frac{1}{2}\right) = \left(3\frac{1}{2}\right)$$

Proposition 11.2 (Basic properties of Jordan blocks)**.** *Set $J = J_n(\lambda)$. Then*

(1) *characteristic polynomial of J is $(x-\lambda)^n$.*
(2) *λ is the only eigenvalue of J, $a(\lambda) = n$, $g(\lambda) = 1$.*
(3) *Set $K = J - \lambda I = J_n(0)$. The linear map $\mathbb{C}^n \to \mathbb{C}^n$ defined by K sends*

$$e_n \mapsto e_{n-1} \mapsto e_{n-2} \mapsto \cdots \mapsto e_2 \mapsto e_1 \mapsto e_0.$$

(4) *$(J-\lambda I)^n = 0$ and for $1 \leq i \leq n-1$:*

$$(J-\lambda I)^i = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 & & & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 & & \\ & & & & & & \ddots & \\ & & & & & & & 1 \\ & & & & & & & 0 \\ & & & & & & & 0 \\ & & & 0 & & & & \vdots \\ & & & & & & & 0 \end{pmatrix}$$

where the 1 in the first row appears in the $(i+1)$th column etc.

(5) *If $\mu \neq \lambda$, then $J - \mu I$ is invertible.*

Proof. Since $J_n(\lambda)$ is upper-triangular, (1) is clear, and the only thing to prove in (2) is $g(\lambda) = 1$. Need to solve

$$\begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & & \\ & & & 0 & 1 \\ & & & & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_{n-1} \\ v_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

but $LHS = (v_2, v_3, \ldots, v_n, 0)^T$, so $E_\lambda(J_n(\lambda)) = \{(v_1, 0, \ldots, 0)^T : v_1 \in E\}$ is 1-dimensional.

(3) is clear.

For (4), first note that $(J - \lambda I)^n$ kills every basis vector by (3), so $(J - \lambda I)^n = 0$. For any $1 \leq i \leq n-1$, note that by (3), $(J - \lambda I)^i$ sends

$$\begin{aligned} e_n &\mapsto e_{n-i} \\ e_{n-1} &\mapsto e_{n-i-1} \\ &\vdots \\ e_{i+1} &\mapsto e_1 \\ e_i &\mapsto 0 \\ &\vdots \\ e_1 &\mapsto 0 \end{aligned}$$

and (4) follows from writing this linear map in the basis $\{e_1, \ldots, e_n\}$.

For (5), note that $J - \mu I = J_n(\lambda - \mu)$ which has determinant $(\lambda - \mu)^n \neq 0$. □

Remark. Since $a(\lambda) = n$, $g(\lambda) = 1$, $J_n(\lambda)$ is diagonalisable if and only if $n = 1$.

Block diagonal matrices. If A is an $n \times n$ matrix and B is an $m \times m$ matrix, define

$$A \oplus B = \left(\begin{array}{c|c} A & 0 \\ \hline 0 & B \end{array}\right)$$

an $(n+m) \times (n+m)$ matrix.

For example:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \oplus (5) = \begin{pmatrix} 1 & 2 & 0 \\ 3 & 4 & 0 \\ 0 & 0 & 5 \end{pmatrix}$$

More generally, if A_1, …, A_k are matrices and A_i is $n_i \times n_i$, then

$$A_1 \oplus A_2 \oplus \cdots \oplus A_k = \left(\begin{array}{c|c|c|c} A_1 & 0 & 0 & 0 \\ \hline 0 & A_2 & 0 & 0 \\ \hline 0 & 0 & \ddots & 0 \\ \hline 0 & 0 & 0 & A_k \end{array}\right) = \begin{pmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_k \end{pmatrix}$$

Proposition 11.3 (Basic properties of $\oplus$)**.** *Let $A = A_1 \oplus \cdots \oplus A_k$. Then*

(1) *If $p_i(x)$ is the characteristic polynomial of A_i, then the characteristic polynomial of A is*
$$\prod_{i=1}^{k} p_i(x).$$
(2) *The set eigenvalues of A is the union of the sets of eigenvalues of A_i.*
(3) *For any λ:*
$$\dim E_\lambda(A) = \sum_{i=1}^{k} E_\lambda(A_i)$$
(4) *For any polynomial $f(x)$:*
$$f(A) = f(A_1) + \cdots + f(A_k).$$

Proof. For (1), we have that
$$\det(xI - A) = \det \left(\begin{array}{c|c|c|c} xI - A_1 & 0 & 0 & 0 \\ \hline 0 & xI - A_2 & 0 & 0 \\ \hline 0 & 0 & \ddots & 0 \\ \hline 0 & 0 & 0 & xI - A_k \end{array} \right) = \det(xI - A_1) \dots \det(xI - A_k)$$
by Q5 from Problem Sheet 7 and (2) follows from (1) immediately. Moreover, (3) is Q5 from Problem Sheet 6.

For (4), it is enough to check that $A^r = A_1^r \oplus \cdots \oplus A_k^r$ which one can do by induction. □

Jordan Canonical Form theorem. We are now ready to state the biggest theorem of this chapter.

Theorem 11.4 (Jordan Canonical Form)**.** *Every $n \times n$ matrix over an algebraically closed field E is similar to a matrix of the form*
$$J_{n_1}(\lambda_1) \oplus J_{n_2}(\lambda_2) \oplus \cdots \oplus J_{n_k}(\lambda_k)$$
Furthermore, this "JCF matrix" is unique up to reordering of the factors.

We will prove this later. For now, assume the theorem temporarily. By uniqueness, the matrices
$$J_3(7) \oplus J_1(7) = \begin{pmatrix} 7 & 1 & 0 & 0 \\ 0 & 7 & 1 & 0 \\ 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 7 \end{pmatrix} \qquad J_2(7) \oplus J_2(7) = \begin{pmatrix} 7 & 1 & 0 & 0 \\ 0 & 7 & 0 & 0 \\ 0 & 0 & 7 & 1 \\ 0 & 0 & 0 & 7 \end{pmatrix}$$
are not similar, as we have seen before!

Remark.

(1) The λ_i in JCF Theorem 11.4 might not be distinct!
(2) JCF Theorem 11.4 is similar to the fundamental theorem of arithmetic (any positive integer is uniquely a product of primes).

(3) We work over an algebraically closed field to have all the eigenvalues.
(4) The characteristic polynomial of $J_{n_1}(\lambda_1) \oplus \cdots \oplus J_{n_k}(\lambda_k)$ is

$$(x-\lambda_1)^{n_1}\ (x-\lambda_2)^{n_2} \dots (x-\lambda_k)^{n_k}.$$

(5) $A \oplus B$ is similar to $B \oplus A$. To see this, let $C = A \oplus B$ and consider the linear map $E^N \to E^N$ given by $v \mapsto Cv$. Then if A is $n \times n$, B is $(N-n) \times (N-n)$, then we can write the linear map $v \mapsto Cv$ in the basis

$$\{e_{n+1}, e_{n+2}, \dots, e_N, e_1, \dots, e_n\}$$

(where $\{e_1, e_2, \dots, e_N\}$ is the standard basis of E^N) to get $B \oplus A$.

Let us stop assuming JCF Theorem 11.4 and prove it!

Say A is any $n \times n$ matrix and assume that A is similar to

$$J = J_{n_1}(\lambda_1) \oplus \cdots \oplus J_{n_k}(\lambda_k).$$

By rearranging the blocks, we can assume that

$$J = J_{n_1}(\lambda) \oplus J_{n_2}(\lambda) \oplus \cdots \oplus J_{n_g}(\lambda) \oplus J_{n_{g+1}}(\mu_1) \oplus \cdots \oplus J_k(\mu_{k-g})$$

and none of the $\mu_1, \dots, \mu_{k-g}$ are equal to λ.

Proposition 11.5. *In the above setting:*

(1) *the algebraic multiplicity of λ in A is $n_1 + \cdots + n_g$,*
(2) *the geometric multiplicity of λ in A is g,*
(3) *the geometric multiplicity of λ in A is $n - \operatorname{rank}(A - \lambda I) = n - \operatorname{rank}(J - \lambda I)$,*
(4) *for all $i \geq 1$ we have $\operatorname{rank}(A-\lambda I)^i = \operatorname{rank}(J-\lambda I)^i$.*

The Proposition follows from combining the previous results of this section: Propositions 11.1, 11.2, and 11.3. We will use them in the proof without further reference.

Proof. For (1), we have that the characteristic polynomial of A is the characteristic polynomial of J which is

$$(x-\lambda)^{n_1+\cdots+n_g}\ (x-\mu_1)^{n_{g+1}} \dots (x-\mu_{k-g})^{n_k}.$$

To show (2), recall that $J_{n_i}(\lambda)$ has the geometric multiplicity of λ equal to 1 (and if $\mu \neq \lambda$, $J_{n_i}(\mu)$ has the geometric multiplicity of λ equal to 0), so J has geometric multiplicity of λ equal to $\underbrace{1+1+\cdots+1}_{g \text{ times}} = g$.

Part (3) follows from the definition of the geometric multiplicity and the rank-nullity theorem:

$$g(\lambda) = \dim \ker(A-\lambda I) = n - \operatorname{rank}(A-\lambda I) = n - \operatorname{rank}(J-\lambda I).$$

Finally, for (4) note that 10.4 (2) implies $(A-\lambda I)^i$ and $(J-\lambda I)^i$ are similar, so we are done. □

Example. (Assume JCF theorem 11.4.)

Find the JCF of

$$A = \begin{pmatrix} 3 & 5 & 0 & 0 & 1 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 9 & 0 & -1 \\ 0 & 0 & 0 & 9 & 0 \\ 0 & 0 & 0 & 0 & 9 \end{pmatrix}$$

Since A is upper-triangular, we easily see that its characteristic polynomial is

$$(x-3)^2(x-9)^3.$$

Therefore, we can write

$$J = \bigoplus_{i=1}^{r} J_{n_i}(3) \oplus \bigoplus_{j=1}^{s} J_{m_j}(9).$$

First, consider the eigenvalue $\lambda = 3$. Since $a(3) = 2$, $n_1 + \cdots + n_r = 2$. Moreover, the rank of

$$A - 3I = \begin{pmatrix} 0 & 5 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 6 & 0 & -1 \\ 0 & 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 0 & 6 \end{pmatrix}$$

is readily 4, so $g(3) = 5 - 4 = 1$. Thus $r = 1$ and $n_1 = 2$.

Now, consider the eigenvalue $\lambda = 9$. Since $a(9) = 3$, $m_1 + \cdots + m_s = 3$. Moreover, the rank of

$$A - 9I = \begin{pmatrix} -6 & 5 & 0 & 0 & 1 \\ 0 & -6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

is readily 3, so $g(9) = 5 - 3 = 2$. Thus $s = 2$ and $m_1 + m_2 = 3$. So $m_1 = 1$, $m_2 = 2$ (up to permutation of factors).

Therefore:

$$J = J_2(3) \oplus J_1(9) \oplus J_2(9).$$

Proposition 11.6. *Suppose*

$$A = J_{a_1}(\lambda_1) \oplus J_{a_2}(\lambda_2) \oplus \cdots$$

$$B = J_{b_1}(\mu_1) \oplus J_{b_2}(\mu_2) \oplus \cdots$$

are two similar matrices in JCF. Then, after re-ordering blocks of B if necessary, A and B have the same number of blocks, and

$$J_{a_1}(\lambda_1) = J_{b_1}(\mu_1),$$

$$J_{a_2}(\lambda_2) = J_{b_2}(\mu_2),$$

$$\vdots$$

Strategy of the proof: need to check that we can figure out numbers like a_1, λ_1, a_2, λ_2 etc. using only properties which are unchanged if we replace J_1 by a similar matrix (invariant under conjugation).

The kind of questions we now have to answer is: Why is $J_1(\lambda) \oplus J_4(\lambda)$ not similar to $J_2(\lambda) \oplus J_3(\lambda)$?
We will illustrate the idea of the proof by showing these matrices are not similar. We have

$$A = J_1(\lambda) \oplus J_4(\lambda) = \begin{pmatrix} \lambda & 0 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 \\ 0 & 0 & \lambda & 1 & 0 \\ 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{pmatrix}$$

$$B = J_2(\lambda) \oplus J_3(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & \lambda & 1 & 0 \\ 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{pmatrix}$$

Suppose A is similar to B. Then $A - \lambda I$ is similar to $B - \lambda I$ by Proposition 10.4 (2). We have:

$$A - \lambda I = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$B - \lambda I = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

They both have rank 3. Also, $(A - \lambda I)^2$ and $(B - \lambda I)^2$ must be similar. We have:

$$(A - \lambda I)^2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$(B - \lambda I)^2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

One has rank 1 and one has rank 2. This contradicts 11.1 (5). Thus A and B are not similar.

Proof of Proposition 11.6. First, let us assume that all the λ_i and all the μ_j are equal to one number, λ.

New notation: Let us reorder the blocks in A so that

$$a_1 \leq a_2 \leq a_3 \leq \dots$$

Say that m_1 of the a_i's equal 1, m_2 of the a_i's equal 2, etc. Then

$$A = J_1(\lambda)^{m_1} \oplus J_2(\lambda)^{m_2} \oplus \cdots,$$

where $J_1(\lambda)^{m_1} = J_1(\lambda) \oplus \cdots \oplus J_1(\lambda)$, m_1 times. We have:

$$A = \left(\begin{array}{c|cc|cc|c} \lambda & & & & & \\ & \lambda & & & & \\ & & \ddots & & & \\ & & \lambda & & & \\ \hline & & & \lambda & 1 & \\ & & & 0 & \lambda & \\ & & & & & \ddots \\ \hline & & & & & \ddots \end{array}\right)$$

where the first part has dimension $m_1 \times m_1$, second part has dimension $(2m_2) \times (2m_2)$, etc.

Say the largest block that occurs is a block of size r. So $m_r \neq 0$ and $m_{r+1} = m_{r+2} = \cdots = 0$. Then $(A - \lambda I)^r = 0$ (raise $A - \lambda I$ to the power r). (This is because $(J_n(\lambda) - \lambda I)^n = 0$ by Proposition 11.2 (4)).

What is $(A-\lambda I)^{r-1}$? Well, raising $J_n(\lambda)-\lambda I$ to power $r-1$ will kill it if and only if $n \leq r-1$. But $(J_r(\lambda) - \lambda I)^{r-1} \neq 0$ and by Proposition 11.2 (4), it has rank 1. By Proposition 11.3 (3), we deduce that the rank of $(A - \lambda I)^{r-1}$ equal number of blocks of size r which is m_r. So if A and B are similar, they have the same m_r.

What about $(A - \lambda I)^{r-2}$? Each block of size r contributes 2 to the rank, each block of size $r - 1$ contributes 1, all others contribute nothing. Thus:

$$\operatorname{rank}(A - \lambda I)^{r-2} = 2m_r + m_{r-1}.$$

So if A and B are similar, they have the same m_{r-1}.

Continuing in this way, we show that if A and B are similar, then they have the same r and the same m_1, …, m_r, so blocks are the same after reordering.

In general, the similar matrices A and B may have lots of eigenvalues. Since A and B have the same characteristic polynomials, they have the same eigenvalues with algebraic multiplicities. Now, observe that if $\mu \neq \lambda$

$$(J_a(\mu) - \lambda I)^n$$

is an invertible matrix, so it has rank a ($J_a(\mu)$ is an $a \times a$ matrix). Therefore, if we look at the ranks of $(A - \lambda I)^n$ and $(B - \lambda I)^n$, the contributions from the Jordan blocks with eigenvalue μ will be the same (the dimension of the block). Thus we can repeat the same

trick to show that, for a fixed eigvenvalue λ, the Jordan blocks of A and B with eigenvalues λ are the same (up to reordering). $\square$

The conclusion so far is that any matrix is similar to at most one matrix in JCF (regarding permutations of blocks as the same).

We still need to prove that any matrix is similar to at least one matrix in JCF.

Let A be an $n \times n$ matrix . How to find P such that $P^{-1}AP$ is in JCF? Set $V = \mathbb{C}^n$ and let us define $T\colon V \to V$ by $T(v) = Av$. If $E =$ standard basis of V, then

$$[T]_E = A.$$

If we can find a new basis F such that $[T]_F = J$ and

$$J = J_{a_1}(\lambda_1) \oplus J_{a_2}(\lambda_2) \oplus \cdots$$

then we are done, as if P is an approporiate change of basis matrix, then $[T]_F = P^{-1}[T]_E P$, so $J = P^{-1}AP$.

Say $T\colon V \to V$ is a linear map. We need to find a basis, F, such that $[T]_F$ is a JCF matrix.

There are 3 ingredients in the proof:

(A) Method of braking V up as a *direct sum* of subspaces $V = V_1 \oplus \cdots \oplus V_k$. This is an abstract analogue of $A = A_1 \oplus \cdots \oplus A_k$.
(B) Break V up as above such that if $T\colon V \to V$, then

$$T(V_1) \subseteq V_1,$$
$$T(V_2) \subseteq V_2,$$
$$\vdots$$
$$T(V_k) \subseteq V_k.$$

and such that T restricted to V_i only has 1 eigenvalue.
(C) Prove the existence of JCF for linear maps with just 1 eigenvalue.

We will start with (C), the existence of JCF matrices that have one eigenvalue only.

Theorem 11.7. *Say V is a vector space of dimension n, and suppose $S\colon V \to V$ is linear and the characteristic polynomial of S is x^n (equivalently, only eigenvalue of S is 0). Then V has a basis B such that*

$$[S]_B = J_{n_1}(0) \oplus J_{n_2}(0) \oplus \cdots \oplus J_{n_g}(0).$$

Before we start the proof, let us think about what B will look like. Say

$$B = \{e_1, e_2, e_3, \ldots, e_{n_1}, e_{n_1+1}, \ldots, e_{n_1+n_2}, e_{n_1+n_2+1}, \ldots\}$$

where n_1 is the size of the first block, n_2 is the size of the second block etc. By definition of $[S]_B$:

$$S(e_{n_1}) = e_{n_1-1}$$
$$S(e_{n_1-1}) = e_{n_1-2}$$
$$\vdots$$

$$S(e_2) = e_1$$
$$S(e_1) = 0$$

Similarly for the next Jordan block:

$$S(e_{n_1+n_2}) = e_{n_1+n_2-1}$$
$$\vdots$$
$$S(e_{n_2+2}) = e_{n_1+1}$$
$$S(e_{n_1+1}) = 0$$

With this motivation, we introduce a temporary definition (it will simplify the proof; however, it is not a standard notion that you will find in a textbook).

Definition. A *chain* is a finite sequence of vectors

$$v, S(v), S^2(v), \ldots, S^{d-1}(v)$$

such that $S^d(v) = 0$ and $S^{d-1}(v) \neq 0$.

For example, $e_{n_1}, e_{n_1-1}, \ldots, e_1$ is a chain (as above).

Remark. If $0 \neq v \in V$ and $S(v) = 0$, then $\{v\}$ is a perfectly good chain (length 1).

We can now restate Theroem 11.7 using chains: If the only evalue of S is 0, then V has a basis which is a finite disjoint union of chains (because any permutation of a basis is a basis).

Proof of Theorem 11.7. Induction on $n = \dim V$. Base case $n = 1$: $S = (0)$ and any basis will do!

Inductive step: $\dim V = n \geq 2$ and $S\colon V \to V$, only eigenvalue is 0. If $0 \neq v \in V$ is an eigenvector with eigenvalue 0, then $Sv = 0v = 0$, so $\dim \ker(S) > 0$ as it contains v, so $\dim(\mathrm{im}(S)) < n$ by rank-nullity theorem.

So let us set $W = \mathrm{im}(S) = S(V)$. Then $\dim W < n$. Restriction of S to W is a linear map $S_{|W}\colon W \to V$, but the image of this linear map is $\subseteq \mathrm{im}(S) = W$, so the restriction $S_{|W}$ of S is a linear map $W \to W$.

Remark: if λ is an eigenvalue for $S_{|W}$ and $0 \neq w$ is an eigenvector with eigenvalue λ, then $S(w) = \lambda w$. But $w \in V$, so $\lambda = 0$.

Therefore, we can apply the inductive hypothesis to $S_{|W}\colon W \to W$. Hence W has a basis which is a union of chains. Say a basis of W is

$$u_1,\ S(u_1),\ S^2(u_1),\ \ldots,\ S^{m_1-1}(u_1),$$
$$u_2,\ S(u_2),\ S^2(u_2),\ \ldots,\ S^{m_2-1}(u_2),$$
$$\vdots$$
$$u_h,\ S(u_h),\ S^2(u_h),\ \ldots,\ S^{m_h-1}(u_h).$$

and $S^{m_1}(u_1) = S^{m_2}(u_2) = \cdots = S^{m_h}(u_h) = 0$.

Now we need a basis for V. Recall that $W = \mathrm{im}(S) = \{S(v) \colon v \in V\}$. In particular, for $u_i \in \mathrm{im}(S)$, we can choose $v_i \in V$ such that

$$S(v_i) = u_i$$

and we have h slightly longer chains:

$$\begin{array}{c} v_1,\ S(v_1),\ S^2(v_1),\ \ldots,\ S^{m_1}(v_1), \\ v_2,\ S(v_2),\ S^2(v_2),\ \ldots,\ S^{m_2}(v_2), \\ \vdots \\ v_h,\ S(v_h),\ S^2(v_h),\ \ldots,\ S^{m_h}(v_h). \end{array}$$

Note that $\ker S$ already contains $S^{m_1-1}(u_1), \ldots, S^{m_h-1}(u_h)$ which are linearly independent (but may not span $\ker S$). We can extend to a basis of $\ker S$ by throwing in vectors

$$w_1, w_2, \ldots, w_t$$

so that $\dim \ker S = h + t$.

We claim that the following chains

$$\begin{array}{c} v_1,\ S(v_1),\ S^2(v_1),\ \ldots,\ S^{m_1}(v_1), \\ v_2,\ S(v_2),\ S^2(v_2),\ \ldots,\ S^{m_2}(v_2), \\ \vdots \\ v_h,\ S(v_h),\ S^2(v_h),\ \ldots,\ S^{m_h}(v_h). \\ w_1 \\ w_2 \\ \vdots \\ w_t \end{array}$$

form a basis. Let us first check linear independence. Suppose that

$$\begin{aligned} 0 \ = \ & \lambda_{11}v_1 + \lambda_{12}S(v_1) + \cdots + \lambda_{1(m_1+1)}S^{m_1}(v_1) \\ & + \ \cdots \\ & + \lambda_{h1}v_h + \lambda_{h2}S(v_h) + \cdots + \lambda_{1(m_h+1)}S^{m_h}(v_h) \\ & + \mu_1 w_1 + \mu_2 w_2 + \cdots + \mu_t w_t \end{aligned} \tag{1}$$

We will show that all the λ_{ij}'s and μ_i's are 0. First, apply S to equation (1) to get an equation in W:

$$\lambda_{11}u_1 + \lambda_{12}S(u_1) + \cdots + \lambda_{1m_1}S^{m_1-1}(u_1) + \cdots + \lambda_{h1}u_h + \lambda_{h2}S(u_h) + \cdots + \lambda_{hm_h}S^{m_h-1}(u_h) = 0.$$

This is a linear relation on our basis for W, so

$$\lambda_{11} = \lambda_{12} = \cdots = \lambda_{1m_1} = \ldots = \lambda_{h1} = \lambda_{h2} = \ldots = \lambda_{hm_h} = 0$$

We substitute this back to (1) to get

$$\lambda_{1(m_1+1)}S^{m_1}(v_1) + \cdots + \lambda_{h(m_h+1)}S^{m_h}(v_h) + \mu_1 w_1 + \cdots + \mu_t w_t = 0$$

which is a linear relation on our basis for $\ker S$, so

$$\lambda_{1(m_h+1)} = \ldots = \lambda_{h(m_h+1)} = \mu_1 = \ldots = \mu_t = 0.$$

Therefore, our chains in V are linearly independent.

We now claim that the total number of elements in the proposed basis of chains is n. The total number is:

$$(m_1 + 1) + (m_2 + 1) + \cdots + (m_h + 1) + t = (m_1 + \cdots + m_h) + h + t.$$

Our basis for W had size $m_1 + \cdots + m_h$ and the dimension of the kernel of S was $h + t$, so the total number of vectors in the chains is

$$\dim W + h + t = \dim(\mathrm{im}(S)) + \dim(\ker(S)) = n$$

by rank-nullity. Hence the set must be a basis. □

Corollary 11.8. *If $S\colon V \to V$ has one eigenvalue λ then there exists a basis B for which $[S]_B$ is in JCF.*

Proof. Apply Theorem 11.7 to $S - \lambda I$. □

Direct sums of vector spaces. We will now show a method of braking up V as a direct sum of subspaces, an abstract analogue of the direct product of matrices. (This is point (A) of the outline before.)

Definition. Let V be a vector space, $V_1, \ldots, V_s \subseteq V$. We say that V is the *direct sum* of $V_1, \ldots, V_s$ and write

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_s,$$

if every $v \in V$ can be written uniquely as $v = v_1 + v_2 + \cdots + v_s$ with $v_i \in V_i$.

Examples. Let $V = \mathbb{R}^3$, e_1, e_2, e_3 standard basis vectors. Then

$$V = \mathrm{sp}(e_1) \oplus \mathrm{sp}(e_2) \oplus \mathrm{sp}(e_3),$$
$$V = \mathrm{sp}(e_1, e_2) \oplus \mathrm{sp}(e_3).$$

More exotic example:

$$\mathbb{R}^2 = \mathrm{sp}\left(\begin{pmatrix}1\\0\end{pmatrix}\right) \oplus \mathrm{sp}\left(\begin{pmatrix}1\\1\end{pmatrix}\right).$$

Proposition 11.9. *The following are equivalent:*

(1) $V = V_1 \oplus \cdots \oplus V_s$,
(2) $\dim V = \sum_{i=1}^{s} \dim V_i$ *and if B_i is a basis for V_i, then $B = B_1 \cup \ldots \cup B_s$ is a basis for V.*

Proof. Throughout the proof, we will use the notation $|B_i| = n_i$ and $B_i = \{b_{i1}, b_{i2}, \ldots, b_{in_i}\}$ for $i = 1, 2, \ldots, s$.

We will first show (1) implies (2). Let B_i be a basis for V_i. We will check that $B = B_1 \cup \ldots \cup B_s$ is a basis for V.

We will first show that B spans V. Say $v \in V$ and write $v = v_1 + \cdots + v_s$ for $v_i \in V_i$. Each v_i is a linear combination of elements of B_i, so v is a linear combination of elements of B.

We will now show that B is linearly independent. Say we have a linear combination

$$\sum_{i=1}^{s} \left(\sum_{j=1}^{n_i} \lambda_{ij} b_{ij} \right) = \sum_{j=1}^{n_1} \lambda_{1j} b_{1j} + \cdots + \sum_{j=1}^{n_s} \lambda_{sj} b_{sj} = 0.$$

Now, set $v_i = \sum_{j=1}^{n_i} \lambda_{ij} b_{ij}$ for $i = 1, 2, \ldots, s$. Then

$$v_1 + v_2 + \cdots + v_s = 0$$

But note that $0 = 0 + 0 + \cdots + 0$ since $0 \in V_i$ for all i. By definition of a direct sum, 0 can only be written as a sum of elements of the V_i in one way. Thus $v_1 = v_2 = \cdots = v_s = 0$. But $B_1, B_2, \ldots, B_s$ are bases, so they are linearly independent and:

$$\lambda_{11} = \lambda_{12} = \cdots = \lambda_{1n_1} = 0,$$
$$\lambda_{21} = \lambda_{22} = \cdots = \lambda_{2n_2} = 0,$$
$$\vdots$$
$$\lambda_{s1} = \lambda_{s2} = \cdots = \lambda_{sn_s} = 0.$$

Linear independence follows.

Therefore, $|B| = \sum_{i=1}^{s} |B_i|$ and $B = \bigcup_{i=1}^{s} B_i$ is a basis for V.

We will now show that (2) implies (1). Say $v \in V$ and write it as a linear combination of elements of $B = \bigcup_{i=1}^{s} B_i$:

$$v = \sum_{i=1}^{s} \left(\sum_{j=1}^{n_i} \lambda_{ij} b_{ij} \right)$$

so setting

$$v_i = \sum_{j=1}^{n_i} \lambda_{ij} b_{ij} \in V_i$$

for $i = 1, 2, \ldots, s$, we get that $v = v_1 + \cdots + v_s$.

For uniqueness, say $v = v_1' + \cdots + v_s'$ for $v_i' \in V_i$. Write v_i' as a linear combination of B_i, the fixed basis for V_i:

$$v_i' = \sum_{j=1}^{n_i} \lambda_{ij}' b_{ij} \in V_i$$

Then $v_1 + \ldots + v_s = v = v_1' + \cdots + v_s'$, so

$$0 = v - v = \sum_{i=1}^{s} \left(\sum_{j=1}^{n_i} (\lambda_{ij} - \lambda_{ij}') b_{ij} \right)$$

so we have a linear relation on B. But B is a basis, so we get that $\lambda_{ij} = \lambda_{ij}'$, and $v_i = v_i'$ for $i = 1, 2, \ldots, s$, as requested. □

Exercise: Let V be a vector space, and V_1, V_2 be subspaces. Then

$$V = V_1 \oplus V_2 \quad \Leftrightarrow \quad V = V_1 + V_2 \ \text{ and } \ V_1 \cap V_2 = 0$$

where $V_1 + V_2 = \{v_1 + v_2 \colon v_i \in V_i\}$.

(Hint: $\dim(V_1 + V_2) = \dim V_1 + \dim V_2 - \dim(V_1 \cap V_2)$.)

Lemma 11.10. *Say $V = V_1 \oplus V_2 \oplus \cdots \oplus V_s$ and say B_i basis for V_i and B = union of the B_i's. Suppose $T\colon V \to V$ is a linear map and $T(V_i) \subseteq V_i$ for all i. Write $T_i\colon V_i \to V_i$ for the restriction of T to V_i. Then*

$$[T]_B = [T_1]_{B_1} \oplus [T_2]_{B_2} \oplus \cdots \oplus [T_s]_{B_s}.$$

Proof. Say $B_1 = \{e_1, e_2, \ldots, e_n\}$ and $T(e_1) \in V_1$ because $e_1 \in V_1$. Thus

$$T(e_1) = a_{11}e_1 + \cdots + a_{n1}e_n + 0f_1 + 0f_2 + \cdots$$

$$\vdots$$

$$T(e_n) = a_{1n}e_1 + \cdots + a_{nn}e_n + 0f_1 + 0f_2 + \cdots$$

and so

$$[T]_B = \left(\begin{array}{ccc|c} a_{11} & \ldots & a_{1n} & \\ \vdots & & \vdots & \\ a_{n_1} & \ldots & a_{nn} & \\ \hline & 0 & & \end{array}\right)$$

and the lemma follows if you keep going. □

Theorem 11.11. *Suppose $T\colon V \to V$, V a vector space over an algebraically closed field E. Suppose the characteristic polynomial of T is*

$$\prod_{i=1}^{s}(x - \lambda_i)^{a_i}$$

with λ_i distinct eigenvalues. Define $V_i \subseteq V$ by

$$V_i = \ker(T - \lambda_i I)^{a_i}.$$

Then:

(1) *$V = V_1 \oplus \cdots \oplus V_s$,*
(2) *$T(V_i) \subseteq V_i$ for all i,*
(3) *the characteristic polynomial of $T_{|V_i}$, the restriction of T to V_i, is $(x - \lambda_i)^{a_i}$.*

Corollary 11.12. *Existence of Jordan Canonical Form.*

Proof of Corollary 11.12. Because Corollary 11.8 applies to each V_i (only eigenvalue is λ_i), there exists a basis B_i such that $[T_i]_{B_i}$ is in JCF. So set $B = \bigcup B_i$ and use Lemma 11.10 to finish the proof. □

Proof of Theorem 11.11. We will prove all three parts together by induction on s.

In the base case $s = 1$ so $V_1 = V$ and there is nothing to prove.

Inductive step. First, let us write characteristic polynomial of $T\colon V \to V$ as

$$(x - \lambda)^a \times q(x)$$

where λ is not a root of $q(x)$, and let $p(x) = (x - \lambda)^a$. The two key facts we will use are:

(i) $p(x)$, $q(x)$ are coprime,
(ii) $p(T)q(T) = 0$.

[Since λ is not a root of $q(x)$, (i) is trivial. Since $p(x)q(x)$ is the characteristic polynomial of T, (ii) follows from Cayley–Hamilton Theorem 10.3.]

By (i) and Euclid's algorithm for polynomials, there exist polynomials $\lambda(x)$ and $\mu(x)$ such that $\lambda(x)p(x) + \mu(x)q(x) = 1$ and, substituting $x = T$, we get

$$\lambda(T)p(T) + \mu(T)q(T) = \mathrm{Id}. \tag{2}$$

Set:

$$V_1 = \ker(T - \lambda I)^a = \ker p(T),$$
$$W = \ker q(T).$$

We will show that $V = V_1 \oplus W$.

Firstly, if $v \in V$, then by equation (2) we have

$$\lambda(T)p(T)v + \mu(T)q(T)v = \mathrm{Id}v = v.$$

Set $w = \lambda(T)p(T)v$ and $v_1 = \mu(T)q(T)v$. We have $w \in W$ as

$$q(T)w = q(T)\lambda(T)p(T)v = \lambda(T)[p(T)q(T)]v = 0$$

where the last equality follows from (ii), and similarly $v_1 \in V_1$ as

$$p(T)v_1 = p(T)\mu(T)q(T)v = \mu(T)[p(T)q(T)]v = 0.$$

Therefore, $V = V_1 + W$.

We now claim that $V_1 \cap W = 0$. If $v \in V_1 \cap W$, then

$$p(T)v = q(T)v = 0$$

and thus by equation (2)

$$v = \mathrm{Id}v = \lambda(T)p(T)v + \mu(T)q(T)v = 0.$$

Therefore $V_1 \cap W = 0$, which together with $V = V_1 + V_2$ shows that $V = V_1 \oplus W$.

Note that the only eigenvalue of $T_{|V_1} \colon V_1 \to V_1$, T restricted to V_1, is λ. Indeed, suppose contrary that some $\mu \neq \lambda$ is an eigenvalue of $T_{|V_1}$ with eigenvector $v \in V_1$. Then $T(v) = \mu v$ so μ is also an eigenvalue of T, and $p(\mu)q(\mu) = 0$. But $\mu \neq \lambda$ implies $p(\mu) \neq 0$, so $q(\mu) = 0$, and hence $v \in \ker(q(T)) = W$. However, we already showed that $V_1 \cap W = 0$, so $v = 0$, a contradiction. Therefore, the characteristic polynomial of $T_{|V_1}$ is $p(x) = (x - \lambda)^a$.

To check that $T(V_1) \subseteq V_1$, note that if $p(T)v_1 = 0$ then

$$p(T)Tv_1 = Tp(T)v_1 = T(0) = 0$$

and similarly $T(W) \subseteq W$.

Therefore, we have shown that $V = V_1 \oplus W$ and V_1 satisfies (2) and (3).

In order to use the inductive hypotheses, we have to check that the characteristic polynomial of $T_{|W} \colon W \to W$ is $q(x)$. Let the characteristic polynomial be $f(x)$. Since $V = V_1 \oplus W$, $T = T_{|V_1} \oplus T_{|W}$, and we have

$$p(x)q(x) = f(x)q(x)$$

so $f(x) = p(x)$. Therefore, we can decompose W in the same way by the inductive hypothesis to finish the proof. $\square$