LECTURE NOTES FOR 417

CHARLES REZK

1. Introduction

Look at appendices A (logic),B (sets) ,C (induction), of textbook. **Lecture 01**

You first encounter *algebra* as the study of solving equations by means of certain manipulations. In abstract algebra, we study the structures that allow such manipulations. Key actors are:

- **Fields:** these admit the usual "arithmetic" operations of addition, subtraction, multiplication, division. Familiar examples are $\mathbb{Q}, \mathbb{R}, \mathbb{C}$.
- **Rings:** these admit most of the usual operations except division.

  A key example is $\mathbb{R}[x]$, the ring of polynomials in one variable with coefficients in $\mathbb{R}$; this is where a lot of high school level algebra lives.

  Another example: $\mathbb{Z}$ the ring of integers. This is the natural place for questions about divisibility.

  Another: $M_{n\times n}(\mathbb{R})$, square matrices with real coefficients. Multiplication is not commutative in this one.
- **Groups:** these admit just one operation, instead of two. Any ring with identity R gives you two examples of groups: $(R, +)$, the elemets of R under addition, and $(R^\times, \cdot)$, the *invertible* elements of R under multiplication (e.g., if $R = M_{n\times n}(F)$, then $R^\times$ is the set of matrices with non-zero determinant.)

  Another source of groups is as symmetries, which we describe soon.
- **Monoids:** these admit one operation just like groups, but the operation is not required to have inverses. In addition to groups (which are always monoids), we have for any ring the monoid $(R, \cdot)$ of elements under multiplication.

2. Symmetry

See sections 1.1–4.

2.1. Symmetries of an equalateral triangle.

An equilateral triangle, thought of as a flat sheet sitting in 3-space. The symmetry group is the collection of all rotations which take the triangle to itself, possibly permuting points.

Let A, B, C be vertices of the triangle: put it in the xy-plane, with $A = (1, 0, 0)$, $B = (\cos 120°, \sin 120°, 0)$, $C = (\cos 240°, \sin 240°, 0)$. We have the following symmetries:

- e, the identity symmetry.
- r_1, rotation by 120° (counterclockwise from above) around the z-axis.
- r_2, rotation by $-120°$ (clockwise from above) around the z-axis.
- a, rotation by 180° around the x-axis (i.e., around the line through A and the opposite midpoint).
- b, rotation by 180° around the line through B and the opposite midpoint.
- c, rotation by 180° around the line through C and the opposite midpoint.

Date: February 7, 2023.

I claim that the collection $G = \{e, r_1, r_2, a, b, c\}$ forms a group, where the "product" is composition of symmetries. That is, if $x, y \in G$, then

$$xy = \text{first do } y\text{, then do } x.$$

Here are a few products.

- Clearly, e is an identity element.
- We have $r_1 r_1 = r_2$, and $r_2 r_2 = r_1$.
- We have $r_1 r_2 = e = r_2 r_1$. This means that r_1 and r_2 are inverses of each other.

It is convenient to write $r = r_1$, and then write $r^2 = r_2$, so that $G = \{e, r, r^2, a, b, c\}$. We have $r^3 = e$.

- We have $a^2 = e$, $b^2 = e$, and $c^2 = e$. That is, the elements a, b, c are their own inverses.

Here is another product.

- $ra = c$. That is, if we *first* rotate 180° around the x-axis, *then* rotate 120° (counterclockwise from above) around the z-axis, it is the same as *rotating* 180° *around the line connecting* C *and the opposite midpoint.*
- $ar = b$. That is, if we *first* rotate 120° (counterclockwise from above) around the z-axis, *then* rotate 180° around the x-axis, it is the same as *rotating* 180° *around the line connecting* B *and the opposite midpoint.*

We can actually write all these elements as functions $\mathbb{R}^3 \to \mathbb{R}^3$, given by matrix multiplication:

$$r(x,y,z) = \begin{bmatrix} \cos 120° & -\sin 120° & \\ \sin 120° & \cos 120° & \\ & & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \qquad a(x,y,z) = \begin{bmatrix} 1 & & \\ & -1 & \\ & & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix},$$

So

$$ra(x,y,z) = \begin{bmatrix} \cos 120° & -\sin 120° & \\ \sin 120° & \cos 120° & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & -1 & \\ & & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \cos 120° & \sin 120° & \\ \sin 120° & -\cos 120° & \\ & & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix},$$

which turns out to be c. (What are the eigenvectors?)

- $ab = r$, and $ba = r^2$.

We end up with the multipication table for computing xy:

		$y =$					
		e	r	r^2	a	b	c
$x =$	e	e	r	r^2	a	b	c
	r	r	r^2	e	c	a	b
	r^2	r^2	e	r	b	c	a
	a	a	b	c	e	r	r^2
	b	b	c	a	r^2	e	r
	c	c	a	b	r	r^2	e

Note: every element appears exactly once in each row and in each column. This is not an accident.

2.2. **Symmetries of a rectangle.** Use a non-square rectangle in the xy-plane, with corners at $(\pm a, \pm b)$, $a \neq \pm b$. Let

- e, identity symmetry.
- r_1, 180° rotation around x-axis.
- r_2, 180° rotation around y-axis.
- r_3, 180° rotation around z-axis.

Write multiplication table. (Refer to book for more detail.)

2.3. **Symmetries of a square.** I'll just set it up: read the book for the rest. Use the square in the xy-plane, with corners at $(\pm 1, \pm 1, 0)$. Let

- e, identity symmetry.
- r, 90° rotation around z-axis.
- r^2, 180° rotation aronud z-axis.
- r^3, 270° rotation around z-axis.
- a, 180°-rotation around x-axis.
- b, 180°-rotation around y-axis.
- c, 180°-rotation around line through $(1, -1)$, $(-1, 1)$.
- d, 180°-rotation around line through $(1, 1)$, $(-1, -1)$.

The collection $G = \{e, r, r^2, r^3, a, b, c, d\}$ forms a group.

Here is a partial multiplication table.

	e	r	r^2	r^3	a	b	c	d
e	e	r	r^2	r^3	a	b	c	d
r	r	r^2	r^3	e				
r^2	r^2	r^3	e	r				
r^3	r^3	e	r	r^2				
a	a	c	b		e	r^2	r	
b	b					e		
c	c						e	
d	d							e

Fill in the rest (the full table is in the book). For instance,

- $ar = c$ (compute for each of the points $(\pm 1, \pm 1, 0)$.)
- $ar^2 = b$.
- $ab = r^2$.
- $ac = r$.

Symmetry groups. A collection G of symmetries of an object for which all composites are in the collection is called a symmetry group. We note the following:

- Composition $x, y \mapsto xy$ gives a well-defined binary operation on G.
- This operation is associative: $(xy)z = x(yz)$, because composition of functions is associative.
- There is an identity symmetry e, with the property that $ex = xe = x$.
- Each x has an inverse x^{-1}, so that $xx^{-1} = x^{-1}x = e$. (It is possible for $x = x^{-1}$.)

Symmetrices of three dimensional objects. We can ask about the rotational symmetries of a 3 dimensional object. For instance, a cube has many rotational symmetries, including the following.

- e, identity symmetry.
- For each of the three axes through the centers of opposite sides, two 90-degree rotations (in either direction) around the axis.
- For each of these axes, one 180-degree rotation.
- ...

Exercise: find all the rotational symmetrices of the cube. There are 24 in total.

If you are feeling very ambitious, compute its multiplication table.

3. Rotations of space

Definition of group. Let me pause here to give the formal definition of a group. **Lecture 02**

A **group** consists of: a set G together with a binary operation $G \times G \to G$ (often written as $x, y \mapsto xy$, and called a *product*), satisfying the following axioms: **group**

(1) The product is **associative**: $(ab)c = a(bc)$ for all $a, b, c \in G$. **associative**

(2) There is an **identity**: there exists an element $e \in G$ with the property that $ae = a = ea$ for all $a \in G$. **identity**

(3) Every element has an **inverse**: for each $a \in G$, there exists an element $a^{-1} \in G$ such that $aa^{-1} = e = a^{-1}a$, where e is the identity element of (2). **inverse**

When you refer to a group, you should in principle mention both the set G and the product operation. Thus, we would speak of the group $(\mathbb{Z}, +)$ of integers under addition, or the group $(\mathbb{R}^\times, \cdot)$ of non-zero reals under multiplication.

In practice, people will just name a group by its set, and this is usually ok because you can infer the product from the context (so I might talk about the groups $\mathbb{Z}$ or $\mathbb{R}^\times$).

3.1. *Example.* Define an operation $\star\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ by the formula

$$x \star y := \sqrt[3]{x^3 + y^3}.$$

Show that $(\mathbb{R}, \star)$ satisfies the axioms for a group. In particular, determine the identity element and inverses of any element.

Note that this is not the same as the group $(\mathbb{R}, +)$, because the product operation is different.

3.2. *Example.* Let F be a field (like $\mathbb{R}$ or $\mathbb{C}$), and let $GL_n(F) \subseteq \mathrm{Mat}_{n\times n}(F)$ be the set of *invertible* $n \times n$ matrices. We can equip this set with a product, which is just product of matrices: $A, B \mapsto AB$. The resulting group $(GL_n(F), \cdot)$ (or just $GL_n(F)$) is called a **general linear group**. **general linear group**

Rotations. Rotations of the plane around the origin can be expressed by certain 2×2-matrices:

$$\mathrm{Rot}(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

The action is by multiplication by a column vector.

Note the following identities.

- $\mathrm{Rot}(\alpha)\,\mathrm{Rot}(\beta) = \mathrm{Rot}(\alpha + \beta)$,
- $\mathrm{Rot}(0) = I$,
- $\mathrm{Rot}(\alpha)^{-1} = \mathrm{Rot}(-\alpha)$,
- $\mathrm{Rot}(\alpha + 2\pi n) = \mathrm{Rot}(\alpha)$ for any $n \in \mathbb{Z}$.

Exercise. Prove these using trigonometric identities.

So this notation gives us multiple names for the same rotation.

By adding an extra coordinate, we get rotations of space around the z-axis:

$$\mathrm{Rot}_{e_3}(\theta) := \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Here e_3 is the unit vector along the positive z-axis This rotation "counterclockwise" if you are looking down from the head of the vector e_3.

More generally, there is a rotation by angle θ around the axis along a unit vector u:

$$\mathrm{Rot}_u(\theta) = \left\{ \begin{matrix} \text{matrix describing rotation of angle } \theta \text{ along the axis} \\ \text{through } u \text{ (counterclockwise viewed from endpoint of } u\text{)} \end{matrix} \right\}.$$

Recall that an **orthogonal matrix** is a real square matrix P such that $P^\top P = I$. Equivalently, a real square matrix is orthogonal if: **orthogonal matrix**

- $P^\top P = I$.
- $PP^\top = I$.
- The columns of P are an orthonormal basis of $\mathbb{R}^n$
- The rows of P are an orthonormal basis of $\mathbb{R}^n$.

Note that any orthogonal matrix is invertible, since $P^{-1} = P^\top$, and that $\det P \in \{\pm 1\}$ since $\det P^\top = \det P$.

Now we have the following useful formula for a general rotation:

$$P \operatorname{Rot}_u(\theta) P^{-1} = \operatorname{Rot}_{Pu}(\theta),$$

where P is an 3×3 orthogonal matrix. I wont't try to prove it, but you can think about it in terms of change-of-basis.

For example, to compute $\operatorname{Rot}_u(\theta)$, use Gram-Schmidt to find any orthonormal basis u_1, u_2, u_3 with $u_3 = u$, and then

$$\operatorname{Rot}_u(\theta) = P \operatorname{Rot}_{e_3} P^{-1}, \qquad \text{where } P = [u_1\ u_2\ u_3].$$

The idea is you are taking a standard rotation around the z-axis, but then changing the basis by replacing e_1, e_2, e_3 with u_1, u_2, u_3, so you get a rotation around the u_3-axis.

3.3. *Example.* Let $u = (e_1 + e_2)/\sqrt{2}$, a unit vector in the xy-plane. Here's how you can compute $\operatorname{Rot}_u(\pi/3)$. First, we need to construct a orthonormal basis u_1, u_2, u_3 with $u_3 = u$. For instance, since u is in the xy-plane, we can take $u_1 = e_3$ since these are orthogonal. A vector orthogonal to both is $u_2 = (e_1 - e_2)/\sqrt{2}$.

Now let $P = [e_1\ e_2\ e_3] = \begin{bmatrix} 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \\ 1 & 0 & 0 \end{bmatrix}$. Then

$$\begin{aligned}
\operatorname{Rot}_{(e_1+e_2)/\sqrt{2}}(\pi/3) &= P \operatorname{Rot}_{e_3}(\pi/3) P^\top \\
&= \begin{bmatrix} 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/2 & -\sqrt{3}/2 & 0 \\ \sqrt{3}/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \end{bmatrix} \\
&= \begin{bmatrix} 3/4 & 1/4 & \sqrt{3/8} \\ 1/4 & 3/4 & -\sqrt{3/8} \\ -\sqrt{3/8} & \sqrt{3/8} & 1/2 \end{bmatrix}.
\end{aligned}$$

3.4. *Example* (Symmetries of the square as matrices). These are given by

$$\begin{gathered}
I, \quad R = \operatorname{Rot}_{e_3}(\pi/2), \quad R^2 = \operatorname{Rot}_{e_3}(\pi), \quad R^3 = \operatorname{Rot}_{e_3}(3\pi/2), \\
A = \operatorname{Rot}_{e_1}(\pi), \quad B = \operatorname{Rot}_{e_2}(\pi), \quad C = \operatorname{Rot}_{(e_1-e_2)/\sqrt{2}}(\pi), \quad D = \operatorname{Rot}_{(e_1+e_2)/\sqrt{2}}(\pi).
\end{gathered}$$

We have the following basic facts.

(1) $\operatorname{Rot}_u(0) = I$ for any u.
(2) $\operatorname{Rot}_u(\theta + 2\pi n) = \operatorname{Rot}_u(\theta)$ for all $n \in \mathbb{Z}$.
(3) $\operatorname{Rot}_u(\theta) = \operatorname{Rot}_{-u}(-\theta)$, i.e., a counterclockwise rotation viewed from the opposite side is a clockwise rotation.
(4) $\operatorname{Rot}_u(\theta)^{-1} = \operatorname{Rot}_u(-\theta) = \operatorname{Rot}_{-u}(\theta)$.
(5) $\operatorname{Rot}_u(\alpha) \operatorname{Rot}_v(\beta)$ is a rotation matrix.

This last one is intuitively clear, but is not obvious from the formulas. In fact, there is really no convenient formula for the product in general, though we do know that $\operatorname{Rot}_u(\alpha) \operatorname{Rot}_u(\beta) = \operatorname{Rot}_u(\alpha + \beta)$. You can prove it using the following.

3.5. **Proposition.** *A 3×3 matrix A is rotation matrix iff it is* special orthogonal*, i.e., $A^\top A = I$ and $\det A = 1$.*

Note: this is basically a consquence of spectral theory/"the principal axis theorem", which you learned about in linear algebra. I won't prove it now, but I'll give the idea of the proof below.

Given this, it is clear that if A, B are rotation matrices, then so is $C := AB$, since

$$C^\top C = (AB)^\top (AB) = B^\top A^\top AB = B^\top B = I, \qquad \det C = (\det A)(\det B) = 1.$$

Proof of characterization of rotation matrices. First note the following: if A is special orthogonal, i.e., $A^\top A = I$ and $\det A = 1$, and if P is any orthogonal matrix ($P^\top P = I$), then the matrix $B := PAP^\top$ is also special orthogonal: We have

$$B^\top B = (PAP^\top)^\top (PAP^\top) = PA^\top P^\top PAP^\top = PA^\top AP^\top$$

and

$$\det B = \det(PAP^{-1}) = \det P \det A (\det P)^{-1} = \det A.$$

It is easy to check by hand that $\mathrm{Rot}_{e_3}(\theta)$ is special orthogonal, and therefore so is $A = P\,\mathrm{Rot}_{e_3}(\theta)P^\top = \mathrm{Rot}_{Pe_3}(\theta)$.

Conversely, suppose A is special orthogonal. I will show that A has 1 as an eigenvalue, and hence a corresponding real eigenvector u_3 with $Au_3 = u_3$ (this vector will lie along the axis of the rotation, except in the special case that $A = I$). Using this, here's how we can show that A is a rotation:

Assume the eigenvector u_3 is a unit vector, and choose an orthogonal matrix $P = [u_1\ u_2\ u_3]$ with u_3 as the third column. Let $B = P^\top AP$, which is also special orthogonal, and whose third column is e_3 since $Be_3 = P^{-1}APe_3 = P^{-1}Au_3 = P^{-1}u_3 = e_3$. Since B is special orthogonal, its columns are an orthonormal basis and it has determinant 1, so we must have

$$B = \begin{bmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad (a, c) \cdot (b, d) = 0, \quad a^2 + c^2 = 1 = b^2 + d^2, \quad ad - bc = 1.$$

The only two unit vectors perpendicular to (a, c) are $\pm(-c, a)$, and the condition that $ad - bc = 1$ forces $(b, d) = (-c, a)$. So $B = \mathrm{Rot}_{e_3}(\theta)$ where $(\cos\theta, \sin\theta) = (a, c)$. Thus $A = P\,\mathrm{Rot}_{e_3}(\theta)P^{-1} = \mathrm{Rot}_{u_3}(\theta)$ is a rotation matrix as desired.

Finally, here is a sketch of the proof that a 3×3 special orthogonal matrix has 1 as an eigenvalue.

Consider the characteristic polynomial of A, and factor it over the complex numbers to get:

$$p(x) = \det(xI - A) = (x - \lambda_1)(x - \lambda_2)(x - \lambda_3).$$

Since A is a real matrix, this is a real polynomial, so either:

(1) all three roots $\lambda_1, \lambda_2, \lambda_3$ are real, or
(2) one root is real (say λ_3), and the other two are non-real, but come as a conjugate pair, so $\lambda_2 = \overline{\lambda_1}$.

Also, note that $1 = \det A = \lambda_1\lambda_2\lambda_3$.

I claim that if $A^\top A = I$, then each eigenvalue is a length 1 complex number: $|\lambda_i| = 1$. If this is true, then in case (1) we have $\lambda_1, \lambda_2, \lambda_3 \in \{\pm 1\}$, so at least one must be $= 1$ since their product is 1, while in case (2) we have $1 = |\lambda_1|^2 = \lambda_1\overline{\lambda_1} = \lambda_1\lambda_2$, so $\lambda_3 = 1$.

So why do the eigenvalues have length 1? Because $A^\top$ commutes with A, it is a "normal" matrix. The most general form of the principal axis theorem says that A is diagonalizable over the complex numbers, with a basis of eigenvectors v_1, v_2, v_3 which are orthonormal in $\mathbb{C}^3$.

Let $\lambda_1, \lambda_2, \lambda_3$ be the eignevalues for these three eigenvectors. I write A^* for the conjugate transpose of a matrix A, and note that if v, w are column vectors then w^*v is the Hermitian inner product of v and w.

Thus we have $(Av_i)^*(Av_i) = v_i^*A^*Av_i = v_i^*v_i = 1$, and also $(Av_i)^*(Av_i) = (\lambda_i v_i)^*(\lambda_i v_i) = \overline{\lambda_i}\lambda_i$. Thus each λ_i has length 1, as desired.

3.6. **The group of rotations of space.** We write $SO(n)$ for the collection of $n \times n$ special orthogonal matrices. That is, elements of $SO(n)$ are $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ such that $A^\top A = I$ and $\det A = 1$. For each $n \geq 1$, this set is an example of a group.

(1) If $A, B \in SO(n)$, then $AB \in SO(n)$ (check this).
(2) The $n \times n$ identity matrix $I \in SO(n)$.
(3) If $A \in SO(n)$, then $A^{-1} \in SO(n)$.

When $n = 3$, this is the group of rotations of space around the origin. When $n = 2$, this is the group of rotations of the plane around the origin.

When $n \geq 2$, these are examples of infinite groups.

This lives inside a larger group $O(n)$, the collection of $n \times n$ orthogonal matrices (these only satisfy $A^\top A = I$, but can have $\det A \in \{\pm 1\}$).

4. Permutations

See section 1.5. **Lecture 03**

Let X be a set. A map $\phi\colon X \to X$ which is a bijection is called a **permutation** of the set. **permutation**

4.1. *Example.* Consider symmetries of the square, and label the vertices A, B, C, D. Each symmetry gives a permutation of the set $\{A, B, C, D\}$.

We are mainly interested in the case when X is a finite set, which we can take to be $\underline{n} := \{1, 2, \ldots, n\}$.

We have the following "two-line" notation for permutations of $\underline{n}$, as a $2 \times n$ matrix. For instance, $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 1 & 3 \end{pmatrix}$ is notation for the function $\phi\colon \{1, 2, 3, 4\} \to \{1, 2, 3, 4\}$ given by

$$\phi(1) = 4, \quad \phi(2) = 2, \quad \phi(3) = 1, \quad \phi(4) = 3.$$

The matrix functions as a "look-up table" for the function ϕ.

We can compose these:

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

Notice that this is not matrix multiplication. Note also that composition in the opposite order is different:

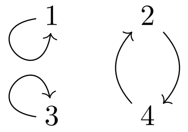
$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 1 & 3 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 3 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 1 & 2 \end{pmatrix}.$$

4.2. **Permutation group.** Let $\mathrm{Sym}(X)$ be the set of all bijections $\phi\colon X \to X$. Composition of functions defines a binary operation of $\mathrm{Sym}(X)$, which makes it into a group. (You can think of $\mathrm{Sym}(X)$ as the "symmetries" of the set X.)

We write $S_n := \mathrm{Sym}(\underline{n})$. Note that S_n has $n!$ elements.

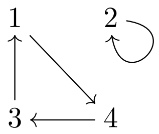
4.3. **Cycles.** We can represent each permutation pictorially as a cycle picture:

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



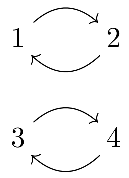
(2 4)

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



(1 4 3)

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



$$(1\ 2)(3\ 4)$$

This suggests cycle notation. Thus $(1\ 4\ 3) \in S_4$ is notation for the permutation given by $1 \to 4$, $4 \to 3$, $3 \to 1$, and which keeps all other elements fixed.

The above identity is

$$(2\ 4)(1\ 4\ 3) = (1\ 2\ 4\ 3).$$

Some facts:

- The order of entries in a cycle $(a_1\ a_2\ \cdots\ a_k)$ usually matters. However, the permutation is unchanged if we move the front to the end:
$$(a_1\ a_2\ \cdots a_k) = (a_2\ \cdots\ a_k\ a_1).$$
That is, there is no preferred "first element" in a cycle (which is obvious from the pictures). We think of these as two names for the same cycle.
Example: in S_7,
$$(1\ 3\ 7\ 2\ 5) = (3\ 7\ 2\ 5\ 1) = (7\ 2\ 5\ 1\ 3) = (2\ 5\ 1\ 3\ 7) = (5\ 1\ 3\ 7\ 2).$$
- Two cycles $\sigma = (a_1\ \cdots\ a_k)$ and $\tau = (b_1\ \cdots\ b_\ell)$ are **disjoint** if they have no two elements in common. *Disjoint cycles commute:* $\sigma\tau = \tau\sigma$. **disjoint**
Example: $(173)(2645) = (2645)(173)$.
Thus, if $\sigma_1, \dots, \sigma_d$ are cycles which are pairwise disjoint, then it doesn't matter what order we multiply them in. (*Disjoint* is very important here.)
- Any cycle of length one (or zero) represents the identity element. We usually don't write these.

4.4. **Theorem.** *Every permutation in S_n can be written uniquely (up to reordering) as a product of pairwise disjoint cycles.*

Note: we regard the identity map e as being written by an "empty product" of cycles.

I'll sketch a proof. Let $\phi\colon X \to X$ be a permutation of a finite set X. Pick some element $x \in X$. Let

$$C_x := \{\, \phi^k(x) \mid k \geq 0 \,\},$$

where $\phi^k = \phi \cdots \phi$, k-fold composition. (When $k = 0$, we have $\phi^0(x) = x$.)

4.5. **Lemma.** *There exists a $d \geq 1$ such that $|C_x| = d$, and $C_x = \{x, \phi(x), \dots, \phi^{d-1}(x)\}$.*

Proof. Since X is finite, $C_x \subseteq X$ must be finite. So there exist some $0 \leq m < n$ such that $\phi^m(x) = \phi^n(x)$. Apply ϕ^{-m} to get $x = \phi^{n-m}(x)$. Thus, there exists $k > 0$ such that $\phi^k(x) = x$.

Let $d \geq 1$ be the *smallest* positive integer such that $\phi^d(x) = x$. Then clearly every element of C_x is of the form $\phi^j(x)$, for some $0 \leq j < d$. Furthermore, $x, \phi(x), \dots, \phi^{d-1}(x)$ must be distinct: if $\phi^i(x) = \phi^j(x)$ then $\phi^{j-i}(x) = x$, contradicting minimality of d. □

4.6. **Lemma.** *Let $Y := X \smallsetminus C_x$. Then $\phi(C_x) \subseteq C_x$ and $\phi(Y) \subseteq Y$, and in fact ϕ restricts to a permutation of C_x and a permutation of Y.*

Proof. That $\phi(C_x) \subseteq C_x$ is clear. If $y \in Y$, and if $\phi(y) \in C_x$, then $\phi(y) = \phi^j(x)$. Then $y = \phi^{j-1}(x)$ (or $\phi^{d-1}(x)$ if $j = 0$), whence $y \in C_x$ a contradiction.

Because ϕ is injective, each restricted map $\phi_{C_x}\colon C_x \to C_x$ and $\phi_Y\colon Y \to Y$ is injective. Since the sets are finite, they are bijections. □

Now work inductively. By induction, ϕ_Y is a product of disjoint cycles, say $\phi_Y = \sigma_1 \cdots \sigma_{r-1}$ where the σ_i are pairwise disjoint (or $\phi_Y = \mathrm{id}$), and we know that ϕ_{C_x} is a single cycle, say σ_r, which must be disjoint to each of σ_i for $1 \leq i < n$. So $\phi_Y = \sigma_1 \cdots \sigma_{r-1}\sigma_r$.

4.7. **Cycle type.** This means we can classify permutations by "cycle type". I'll just give examples.

S_2 :	2	(12)
	1+1	e

S_3 :	3	(123), (132)
	2+1	(12), (13), (23)
	1+1+1	e

S_4 :	4	(1234), (1243), (1324), (1342), (1423), (1432)
	3+1	(123), (132), (124), (142), (134), (143), (234), (243)
	2+2	(12)(34), (13)(24), (14)(23)
	2+1+1	(12), (13), (14), (23), (24), (34)
	1+1+1+1	e

Question. For a given n, how many elements of each cycle type are in S_n.

5. Order and parity of permutations

Order of elements in a group. If $(G, \cdot)$ is a group, and $a \in G$, the **order** of a is the smallest positive integer n such that $a^n = e$. **Lecture 04** **order**

5.1. *Example.* In a symmetric group, any k-cycle has order k.

5.2. *Example.* The permutation $\sigma = (1\ 2)(3\ 4\ 5)$ (in S_5) has order 6.

In the above example, σ is a product of a permutation of order 2 and order 3, and has order 6. It is *not* the case that "orders multiply" in general.

5.3. *Example.* The permutation $\tau = (1\ 2)(2\ 3\ 4)$ has order 4.

5.4. *Exercise.* Show that the order of a permutation only depends on its cycle type. *Harder:* give a formula describing the order of a permutation in terms of its cycle type.

Parity of a permutation. A **transposition** is a 2-cycle. **transposition**

5.5. **Proposition.** *Every permutation of a finite set is either the identity, or can be written (non uniquely) as a product of a finite list of transpositions (not necessarily disjoint).*

Proof. This is left as a homework exercise. The idea is to first prove that any cycle $(a_1\ \cdots\ a_k)$ is equal to a finite product of transpositions. □

The set $\mathrm{Sym}(X)$ of permutations of a finite set can be divided into two types, because of the following result.

5.6. **Proposition.** *If σ is a permutation of a finite set which can be written as a product of transpositions (not necessarily disjoint) in two different ways:*

$$\sigma = \tau_1 \cdots \tau_k = \upsilon_1 \cdots \upsilon_\ell,$$

where τ_i, υ_j are transpositions, then k and ℓ are either both odd or both even, i.e., $(-1)^k = (-1)^\ell$.

Because of this, we say that a permutation is **even** if it is equal to an even number product of transpositions, and **odd** if it is equal to an odd number product of transpositions. **even** **odd**

5.7. *Example.* You can write any k-cycle as a product of $(k-1)$ transpositions. Therefore, a cycle of *even length* is an *odd permutation*, and a cycle of *odd* length is an *even permutation.*

5.8. *Exercise.* Show that the parity of a permutation in S_n depends only on its cycle type. *Harder:* give a formula describing the parity of a permutation in terms of its cycle type.

Proof. WLOG we can assume $X = \{1, \dots, n\}$, so $\mathrm{Sym}(X) = S_n$.

I will define a function $\mathrm{sgn}\colon S_n \to \{\pm 1\}$, called the "sign" function, and prove the following properties:

1. $\mathrm{sgn}(\sigma\tau) = \mathrm{sgn}(\sigma)\,\mathrm{sgn}(\tau)$ for all $\sigma, \tau \in S_n$.

(2) $\operatorname{sgn}(\mathrm{id}) = +1$.
(3) $\operatorname{sgn}(\tau) = -1$ if τ is a transposition.

Given this, the claim follows, since we can use these properties to compute that

$$\operatorname{sgn}(\tau_1 \cdots \tau_k) = (-1)^k, \qquad \operatorname{sgn}(\upsilon_1 \cdots \upsilon_\ell) = (-1)^\ell,$$

but since these are both equal to $\operatorname{sgn}(\sigma)$, we must have $(-1)^k = (-1)^\ell$.

Here is how we define sgn. Given $\sigma \in S_n$, define the **permutation matrix** $A_\sigma \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ by

permutation matrix

$$A_\sigma = [e_{\sigma(1)} \;\cdots\; e_{\sigma(n)}],$$

so that the columns are standard basis vectors "permuted" by σ. Here are some permutation matrices for (1 2) and (1 2 3) in S_3, and (1 3)(2 4) in S_4:

$$A_{(1\ 2)} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad A_{(1\ 2\ 3)} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \qquad A_{(1\ 3)(2\ 4)} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

Note that left-multiplication by A actually is a permutation of the standard basis vectors, and works exactly like σ:

$$A_\sigma e_k = e_{\sigma(k)}.$$

Therefore $A_\sigma A_\tau e_k = A_\sigma e_{\tau(k)} = e_{\sigma\tau(k)}$. This proves that

$$A_\sigma A_\tau = A_{\sigma\tau}.$$

Now define $\operatorname{sgn}(\sigma) := \det A_\sigma$. It is now easy to check all the properties of sgn, using properties of determinant.

In fact, all you need to know is that $\det(AB) = \det(A)\det(B)$, proving (1), that $\det I = 1$, proving (2), and that if A and B differ by switching two columns, then $\det B = -\det A$. Using this you can show that $\det A_\sigma \in \{\pm 1\}$ for any $\sigma \in S_n$, and that $\operatorname{sgn}(\tau) = -1$ if τ is a transposition, proving (3). □

Second proof. Here is another construction of the sign function which I wrote up before the one above. I won't do it in class, and you don't need to know about it, but I'll keep it here.

I'm going to use a trick involving a polynomial in n variables $x_1, \ldots, x_n$:

$$\Delta_n := \prod_{1 \le i < j \le n} (x_i - x_j).$$

For example:

$$\begin{aligned} \Delta_2 &= (x_1 - x_2), \\ \Delta_3 &= (x_1 - x_3)(x_1 - x_3)(x_2 - x_3), \\ \Delta_4 &= (x_1 - x_2)(x_1 - x_3)(x_1 - x_4)(x_2 - x_3)(x_2 - x_4)(x_3 - x_4), \end{aligned}$$

Given any polynomial $f = f(x_1, \ldots, x_n)$ in these variables and any $\sigma \in S_n$, define

$$\sigma(f(x_1, \ldots, x_n)) = f(x_{\sigma(1)}, \ldots, x_{\sigma(n)}),$$

so $\sigma(f)$ is a polynomial obtained by permuting the variables according to σ. We will be interested in applying this with $f = \Delta_n$:

$$\sigma(\Delta_n) := \prod_{1 \le i < j \le n} (x_{\sigma(i)} - x_{\sigma(j)}).$$

For example, if $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}$, then

$$\begin{aligned}\sigma(\Delta_4) &= \prod_{1\leq i<j\leq 4} (x_{\sigma(i)} - x_{\sigma(j)}) \\ &= (x_2 - x_4)(x_2 - x_1)(x_2 - x_3)(x_4 - x_1)(x_4 - x_3)(x_1 - x_3) \\ &= (-1)^3(x_2 - x_4)(x_1 - x_2)(x_2 - x_3)(x_1 - x_4)(x_3 - x_4)(x_1 - x_3) \\ &= -\Delta_4.\end{aligned}$$

Notice that $\sigma(\Delta_n) \in \{+\Delta_n, -\Delta_n\}$. Let $\operatorname{sgn}(\sigma) \in \{\pm 1\}$ be the sign in this formula, so $\sigma(\Delta_n) = \operatorname{sgn}(\sigma)\Delta_n$ defines a function $\operatorname{sgn}\colon S_n \to \{\pm 1\}$.

Claim. If $\sigma, \tau \in S_n$ with product $\upsilon = \sigma\tau$, then

$$\sigma(\tau(\Delta_n)) = \upsilon(\Delta_n),$$

and so

$$\operatorname{sgn}(\sigma\tau) = \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau).$$

This is clear, since $\tau(\Delta_n) = \pm\Delta_n$, so $\sigma(\tau(\Delta_n)) = \pm\sigma(\Delta_n)$.

Finally, we compute that any transposition τ has $\operatorname{sgn}(\tau) = -1$: the idea is that if $\tau = (a\ b)$ with $a < b$, then the number of factors in Δ_n which "change sign" is equal to $2(b - a - 1) + 1$.

□

We say that $\sigma \in S_n$ is **even** if it can be written as a product of an even number of transpositions, **even**
or equivalently if $\operatorname{sgn}(\sigma) = +1$, and **odd** if it can be written as a product of an odd number of **odd**
transpositions, or equivalently if $\operatorname{sgn}(\sigma) = -1$.

Note that the product of two even permutations is even, the product of two odd permutations is even, and the product of an even with an odd is odd.

6. Divisibility

See section 1.6.

Integers. Integers $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$. Has operations of $+$ and $\cdot$.

Let natural numbers $\mathbb{N} = \{1, 2, 3, \dots\}$ be positive integers. (In our book, 0 is not a natural number.) I'll also sometimes write $\mathbb{Z}_{>0} := \mathbb{N}$.

We can add and multiply integers. These operations have the following properties.

(1) Addition is associative: $(a + b) + c = a + (b + c)$ for all $a, b, c \in \mathbb{Z}$.
(2) Addition is commutative: $a + b = b + a$ for all $a, b \in \mathbb{Z}$.
(3) 0 is an additive identity: $a + 0 = a = 0 + a$ for all $a \in \mathbb{Z}$.
(4) Every $a \in \mathbb{Z}$ has an additive inverse $-a$, so that $a + (-a) = 0$. Notation: "$a - b$" means "$a + (-b)$".
(5) Multiplication is associative: $(ab)c = a(bc)$ for all $a, b, c \in \mathbb{Z}$.
(6) Multiplication is commutative: $ab = ba$ for all $a, b \in \mathbb{Z}$.
(7) 1 is a multiplicative identity: $a1 = a = 1a$ for all $a \in \mathbb{Z}$.
(8) Distributive law: $a(b + c) = (ab) + (ac)$ for all $a, b, c \in \mathbb{Z}$.
(9) Non-zero integers are closed under multiplication: If $a, b \in \mathbb{Z} \setminus \{0\}$, then $ab \in \mathbb{Z} \setminus \{0\}$.
(10) $\mathbb{N}$ is closed under addition and multiplication: If $a, b \in \mathbb{N}$, then $a + b, ab \in \mathbb{N}$.
(11) For $a, b \in \mathbb{Z}_{>0}$ we have $|ab| \geq \max\{|a|, |b|\}$.

Note some similar looking rules are missing, because they are false: e.g., $\mathbb{Z} \setminus \{0\}$ is not closed under addition, and elements in $\mathbb{Z}$ need not have multiplicative inverses.

Note that (9) can be stated another way: if $ab = 0$, then either $a = 0$ or $b = 0$.

We write "$a > b$" or "$b < a$" if $a - b \in \mathbb{N}$, and thus write "$a \geq b$" or "$b \leq a$" if $a - b \in \mathbb{Z}_{\geq 0}$.

Divisibility. For $a, b \in \mathbb{Z}$, say a **divides** b if there exists $m \in \mathbb{Z}$ such that $b = ma$. The notation is "$a \mid b$". **a divides b**

On way to read this: "a divides b" if $b/a \in \mathbb{Z}$. (Actually, this is not quite right when $a = 0$; according to our definition, 0 divides 0. Goodman is not careful about this.)

Another way to read this: "a divides b" is the same as "b is a multiple of a".

If $a \mid b$ we also so "a is a factor of b". Given any b, there is a set of factors. Example: the factors of 6 are

$$\{\, a \in \mathbb{Z} \mid a \mid 6 \,\} = \{\pm 1, \pm 2, \pm 3, \pm 6\}.$$

Basic facts:

6.1. **Proposition.** (1) *If $a \mid b$ and $b \mid a$, then $b = \pm a$.*
(2) *If $u \mid 1$, then $u \in \{\pm 1\}$.*
(3) *If $a \mid b$ and $b \mid c$, then $a \mid c$.*
(4) *If $a \mid b$ and $a \mid c$, then $a \mid (b + c)$.*
(5) *More generally, if $a \mid b$ and $a \mid c$, then $a \mid (sb + tc)$ for all $s, t \in \mathbb{Z}$.*

Proof. (1) If $am = b$ and $bn = a$, then $a = (mn)a$, i.e., $a(1 - mn) = 0$. If $a = 0$ then $b = 0$ and the claim is immediate, while if $a \neq 0$ then $1 - mn = 0$ so $mn = 1$. Then $|mn| \geq \max\{|m|, |n|\}$ implies $|m|, |n| \leq 1$, so we must have $m = n \in \{\pm 1\}$.

(2) Is immediate from (1).

(3) If $ma = b$ and $nb = c$, then $(mn)a = c$.

(4) If $ma = b$ and $na = c$, then $(sm + tn)a = sb + tc$. □

Given $a \in \mathbb{Z}$, let

$$\mathbb{Z}a := \{\, ma \mid m \in \mathbb{Z} \,\} = \text{set of integer multiples of } a = \{\, n \mid a \text{ divides } n \,\}.$$

Then $a \mid b$ is the same as $b \in \mathbb{Z}a$. That is, $\mathbb{Z}a$ is exactly the set of multiples of a.

Note: $a \mid b$ if and only if $\mathbb{Z}b \subseteq \mathbb{Z}a$.

Primes. A $p \in \mathbb{N}$ is **prime** if $p \neq 1$, and $n \mid p$ for $n \in \mathbb{N}$ implies $n \in \{1, p\}$. **prime**

Note that if $a, b \in \mathbb{N}$, then $a \mid b$ implies $a \leq b$: this is because $am = b$ implies $|b| \geq \max\{|a|, |m|\}$, so $b \geq a$.

This allows you to prove the following by induction.

6.2. **Proposition.** *Every $n \in \mathbb{N}$ with $n > 1$ can be written as a product of one or more primes.*

Proof. If n is not prime, it is $n = ab$ with $1 < a, b < n$. Proof by induction on n, with basis step $n = 2$ is prime. □

Note: you can also think of 1 as a product of the empty set of primes.

Also recall Euclid's theorem.

6.3. **Proposition** (Euclid)**.** *There are infinitely many primes.*

Proof. Proof by contradiction. If there are finitely many primes $p_1, \dots, p_n$, consider $m := (p_1 \cdots p_n) + 1$. Since $m \geq 2$ it is a product of primes, but if $p_i \mid m$ then we must have $p_i \mid 1$ which is impossible since only ± 1 divide 1. □

Division with remainder. You have probably learned an algorithm which given $a, d \in \mathbb{N}$ lets you write $a/d = q + r/d$, where $q, r \in \mathbb{Z}$ and $0 \leq r/d < 1$. This is the "division algorithm". Let's just prove that there is a unique solution to this.

6.4. **Proposition.** *Given $a, d \in \mathbb{Z}$ with $d \geq 1$, there exist unique $q, r \in \mathbb{Z}$ such that*

$$a = qd + r, \qquad 0 \leq r < d.$$

Proof. (Omit in class?) To prove existence I'll use the "well-ordering principle", which says that any non-empty subset of $\mathbb{Z}_{\geq 0}$ has a least element.

First suppose $a \geq 0$ and let $S = \{\, a - qd \mid q \in \mathbb{Z},\ a - qd \geq 0 \,\}$. This is a subset of $\mathbb{Z}_{\geq 0}$ by construction. It is non-empty because $a = a - 0q \in S$. By the well-ordering principle there is a smallest element r of S, and by construction we have $a = qd + r$ for some $q \in \mathbb{Z}$, and $r \geq 0$. It suffices to show $r < d$. If not, then $r \geq d$, but then it is easy to show that $r' = r - d \in S$ contradicting minimality (since $r' = a - (q-1)d$ and $r' \geq 0$).

If $a < 0$, apply the above to $-a > 0$, giving $-a = q'd + r'$ for some $q', r' \in \mathbb{Z}$, $0 \leq r' < d$. There are two cases:

- If $r' = 0$, then take $r = 0$ and $q' = -q$, so $a = -q'd = qd + r$.
- If $0 < r' < d$, then write
$$a = (-q')d + (-r') = (-q'-1)d + (-r'+d) = qd + r, \qquad q = -q' - 1,\ r = -r' + d,$$
which works since $-d < -r' < 0$ so $0 < r < d$.

Now we prove uniqueness. If $a = qd + r = q'd + r'$ where both $0 \leq r, r' < d$, then
$$0 = a - a = (qd + r) - (q'd + r') = (q - q')d + (r - r') \qquad \Rightarrow \qquad r - r' = (q' - q)d,$$
so $r - r'$ is divisible by d. Since $0 \leq r, r' < d$, we have $|r - r'| < d$. Since $d \mid r - r'$ this implies $r - r' = 0$, so $r = r'$, and thus $q = q'$.

□

Note. If $d \in \mathbb{Z} \setminus \{0\}$, there is unique $q, r \in \mathbb{Z}$ such that
$$a = qd + r, \qquad 0 \leq r < |d|\,.$$
The idea is that if $d < 0$, then the proposition gives q', r' with $a = q'(-d) + r'$ with $0 \leq r' < d$. Just set $q = -q'$ and $r = r'$.

Greatest common divisor. For $m, n \in \mathbb{Z}$, a **greatest common divisor (gcd)** is $d \in \mathbb{Z}_{\geq 0}$ such that

greatest common divi (gcd)

(a) d is a *common divisor* of m and n, and
(b) every common divisor of m and n also divides d.

That is:

(a) $d \mid m$ and $d \mid n$.
(b) if $e \in \mathbb{Z}$ such that $e \mid m$ and $e \mid n$, then $e \mid d$.

Note: if a gcd exists, it is unique. If d and d' are both gcds of m, n, then they divide each other: $d \mid d'$ and $d' \mid d$, so $d = d'$ (since both are positive). Thus, we will write $\gcd(m, n)$ for the GCD of m and n, assuming it exists.

Example: gcd of 18 and 30 is 6.

6.5. *Remark.* This is not exactly the definition used in the textbook. There, m and n are required to be non-zero, and d is assumed to be strictly positive. But in fact the definition makes sense when you allow 0. To see this, just remember that *every integer divides* 0.

For instance, if $m \neq 0$ but $n = 0$, then the common divisors of $\{m, 0\}$ are exactly the divisors of m. Then $d = |m|$ is $\gcd(m, 0)$, since it a divisor of m and any divisor of m divides $|m|$.

And if both $m = n = 0$, then every integer is a common divisor, and therefore $d = 0$ is $\gcd(0, 0)$! This is because 0 is a common divisor, and any integer divides 0.

The only issue here is that if $m = n = 0$, then the GCD is not actually the *largest* common divisor (in fact, there is no largest divisor in this case), which can be a little confusing. But the definition actually says nothing about being largest.

Lecture 05

Here is the big theorem: gcds always exist, and you can compute the gcd of two integers without knowing how to factor either. Proving this involves the notion of **integer combination**.

integer combination

Given $b, c \in \mathbb{Z}$, let

$$I(b,c) := \{\, sb + tc \mid s,t \in \mathbb{Z} \,\} = \text{set of integer combinations of } b \text{ and } c.$$

Example: $I(4,6) = \mathbb{Z}2$. Example: $I(m,0) = \mathbb{Z}m$ for any integer m.

A consequence of this definition: if $a \mid b$ and $a \mid c$, then $I(b,c) \subseteq \mathbb{Z}a$. *Proof.* If $b = am$ and $c = an$, then for any $xb + yc \in I(b,c)$ we have

$$x(am) + y(an) = (xm + yn)a \in \mathbb{Z}a.$$

We will need the following fact, which has a similar proof.

6.6. **Lemma.** *If $d, e \in I(b,c)$, then $I(d,e) \subseteq I(b,c)$.*

Proof. Left as an exercise. □

Note: another notation I might use for $I(b,c)$ is "$\mathbb{Z}b + \mathbb{Z}c$".

6.7. **Theorem.** *Let m, n be integers.*

1. *m and n have a gcd d.*
2. *d is computed by the Euclidean algorithm.*
3. *$I(m,n) = \mathbb{Z}d$.*

First let me explain the **Euclidean algorithm**. Let $\mathbb{Z}^2 := \{\, (m,n) \mid m,n \in \mathbb{Z} \,\}$. Define a function $F\colon \mathbb{Z}^2 \to \mathbb{Z}^2$ by

Euclidean algorithm

$$F(m,n) := \begin{cases} (n,r) \text{ where } 0 \le r < |n| \text{ st. } m = qn + r, & \text{if } n \neq 0, \\ (|m|, 0) & \text{if } n = 0. \end{cases}$$

The **Euclidean algorithm** is the process of iteratively applying F to (m,n), until we reach a value of the form $(d,0)$ with $d \ge 0$. The claim is that d is the GCD.

Euclidean algorithm

Example:

$$(42,-24) \xrightarrow{F} (-24,18) \xrightarrow{F} (18,12) \xrightarrow{F} (12,6) \xrightarrow{F} (6,0) \xrightarrow{F} (6,0) \xrightarrow{F} \cdots.$$

Note that after the second step, the second entry is always ≥ 0. Furthermore, after that point the second entry always strictly decreases if it is positive: if $F(m,n) = (n,r)$ with $r > 0$, then $0 \le r < |n|$. Therefore *the Euclidean algorithm is guaranteed to halt at some finite step.*

6.8. **Lemma.** *For any $(m,n) \in \mathbb{Z}^2$, if $F(m,n) = (m',n')$ then $I(m,n) = I(m',n')$.*

Proof. There are two cases (because F is defined in terms of two cases).

1. If $n \neq 0$, then $(m',n') = (n,r)$, where $0 \le r < |n|$ and $m = qn + r$. It is clear that $n, r = m - qn \in I(m,n)$, so $I(n,r) \subseteq I(m,n)$, while $m = qn + r, n \in I(n,r)$, so $I(m,n) \subseteq I(n,r)$. Therefore $I(m,n) = I(n,r)$ as desired.
2. If $n = 0$, then $m' = |m| = \pm m$, and it is clear that $I(m,0) = \mathbb{Z}m = I(|m|, 0)$.

□

Proof of the theorem. Thus, given a pair of integers (m,n), we can apply the Euclidean algorithm to obtain an integer $d \ge 0$ such that $F^k(m,n) = (d,0)$ for some $k \ge 0$. By the lemma we have that $I(m,n) = I(d,0) = \mathbb{Z}d$. Since $m, n \in \mathbb{Z}d$, it is clear that d is a common divisor of m and n. If e is another common divisor of m and n, then $m, n \in \mathbb{Z}e$ and thus $I(m,n) \subseteq \mathbb{Z}e$. So $\mathbb{Z}d = I(m,n) \subseteq \mathbb{Z}e$, so $e \mid d$. Thus d must be the gcd. □

Note: the Euclidean algorithm is an algorithm both for computing $d = \gcd(m, n)$, and also for finding s, t such that $d = sm + tn$. The idea is that when you compute division with the remainder, you get a formula $r = 1m - qn$ which computes r as an integer combination of m, n, and by repeating this you can get what you want. For instance, recall that $6 = \gcd(42, -24)$. Then we have:

$$\begin{aligned}
\boxed{18} &= 1\boxed{42} + 1\boxed{\text{-24}},\\
\boxed{12} &= 1\boxed{\text{-24}} + 2\boxed{18} = 1\boxed{\text{-24}} + 2\big(1\boxed{42} + 1\boxed{\text{-24}}\big) = 2\boxed{42} + 3\boxed{\text{-24}},\\
\boxed{6} &= 1\boxed{18} + (-1)\boxed{12} = 1\big(1\boxed{42} + 1\boxed{\text{-24}}\big) + (-1)\big(2\boxed{42} + 3\boxed{\text{-24}}\big) = (-1)\boxed{42} + (-2)\boxed{\text{-24}}.
\end{aligned}$$

Prime factorization. Say that $a, b \in \mathbb{N}$ are **relatively prime** if their gcd is 1. Sometimes people say **coprime**.

relatively prime
coprime

Warning: "relatively prime" is a *relation* between two natural numbers, not a property of a single natural number. Neither of the relatively prime integers needs to be prime.

Example: 4 and 9 are relatively prime, but neither are prime.

The above discussion gives us the following very powerful fact:

6.9. **Theorem.**

$$a, b \textit{ are relatively prime iff } 1 = sa + tb \textit{ for some } s, t \in \mathbb{Z}.$$

Proof. Let $d := \gcd(a, b)$. We have shown that $\mathbb{Z}d = I(a, b)$. If $d = 1$ then $1 \in I(a, b)$ so $1 = sa + tb$ for some $s, t \in \mathbb{Z}$. Conversely, if $1 \in I(a, b)$ then $1 \in \mathbb{Z}d$, so $d = 1$. □

6.10. **Proposition.** *If a, b are relatively prime, $a \mid n$ and $b \mid n$, then $ab \mid n$.*

Proof. By the hypotheses, $\exists s, t, u, v \in \mathbb{Z}$ such that $1 = sa + tb$, and $n = au = bv$. Then

$$n = (sa + tb)n = san + tbn = sa(bv) + tb(au) = (sv + tu)ab.$$

□

6.11. **Proposition.** *If p is prime and $p \mid ab$, then either $p \mid a$ or $p \mid b$.*

Proof. Suppose $p \mid ab$ and $p \nmid a$. Then $\gcd(a, p) = 1$, since the only positive divisors of p are $\{1, p\}$. By the big theorem we can write $1 = sa + tp$ for some $s, t \in \mathbb{Z}$. Then

$$b = (sa + tp)b = s(ab) + tb(p).$$

since $p \mid ab$ and $p \mid p$, we have that $p \mid b$ as desired. □

As a consequence: if $p \mid a_1 \cdots a_r$, then $p \mid a_i$ for at least one $i \in \{1, \ldots, r\}$ (inductive argument).

6.12. **Theorem.** *The prime factorization of a natural number $n \geq 2$ is unique (up to reordering).*

Proof. Let $n = p_1 \cdots p_r = q_1 \cdots q_s$, $r \leq s$; we want to show $r = s$ and that the lists of primes are same up to reordering. Induction on r.

If $r = 1$, then $n = p_1 = q_1 \cdots q_s$. Then $p_1 \mid q_i$ for some i, whence $p_1 = q_i$, so removing this factor from both sides gives

$$1 = q_1 \cdots q_{i-1} q_{i+1} \cdots q_s,$$

which is impossible (unless $s = 1$).

Since $p_r \mid n$ we must have $p_r \mid q_i$ for some i, which means $p_r = q_i$ since q_i is also prime. Let $n' := n/p_r$, then

$$n' = p_1 \cdots p_{r-1} = q_1 \cdots q_{i-1} q_{i+1} \cdots q_s.$$

Since $n' < n$, by induction we have $r - 1 = s - 1$ and the two lists of primes are the same up to reordering. □

Lecture 06

Least common multiple. Given $a, b \in \mathbb{Z}$, the $\mathbb{Z}a \cap \mathbb{Z}b$ is the set of **common multiples** of a and b.

common multiples

6.13. **Lemma.** *If $a, b \in \mathbb{Z}$ there is a unique $m \geq 0$ such that $\mathbb{Z}m = \mathbb{Z}a \cap \mathbb{Z}b$.*

Proof. If $a = 0$ or $b = 0$ then $J := \mathbb{Z}a \cap \mathbb{Z}b = \{0\} = \mathbb{Z}0$.

So suppose both a, b are nonzero. Clearly J contains a positive element, e.g., $|ab|$. Let m be the smallest positive element of J (guaranteed by the *well-ordering property* of natural numbers). I claim $J = \mathbb{Z}m$. It is clear that $\mathbb{Z}m \subseteq J$.

Suppose $n \in J$. Use division by remainder $n \div m$ to get

$$n = qm + r, \qquad q, r \in \mathbb{Z}, \quad 0 \leq r < m.$$

Then $r = n - qm \in J = \mathbb{Z}a \cap \mathbb{Z}b$. The condition that m be the minimal positive element of J and $0 \leq r < m$ implies $r = 0$, so $qm = n$. That is, n is a multiple of n as desired, so $J \subseteq \mathbb{Z}m$. □

The element m of the lemma is the unique $m \geq 0$ such that (i) $a \mid m$ and $b \mid m$, and (ii) if $a \mid n$ and $b \mid n$ then $m \mid n$. This is called the **least common multiple** of a, b, and denoted $\operatorname{lcm}(a, b) = m$. (As with GCD, it is common to restrict this definition to the case when a, b are both non-zero.)

least common multiple

We have the following well-known formula.

6.14. **Proposition.** *Suppose $a, b \in \mathbb{N}$ with $m = \operatorname{lcm}(a, b)$ and $d = \gcd(a, b)$. Then $m = ab/d$.*

Proof. We need two facts.

(1) If a, b are relatively prime, then $m = ab$.
(2) If $e \mid a$ and $e \mid b$, then $\operatorname{lcm}(a/e, b/e) = m/e$.

Then since $a/d, b/d$ are relatively prime, we take $e = d$ to get

$$(a/d)(b/d) = \operatorname{lcm}(a/d, b/d) = \operatorname{lcm}(a, b)/d, \qquad \text{so} \qquad ab/d = m.$$

To prove (1) note that ab is certainly a common multiple, and we have proved that if a, b are relatively prime, then $a \mid n$ and $b \mid n$ imply $ab \mid n$, so m divides n so it is the LCM.

To prove (2) note that x is a common multiple of $a/e, b/e$ iff xe is a common multiple of a, b. □

Pairwise relative primeness. A list of non-zero integers is **pairwise relatively prime** if each $\gcd(a_i, a_j) = 1$ for $i \neq j$.

pairwise relatively prime

Example: the list $9, 14, 25$ is pairwise relatively prime.

6.15. **Lemma.** *If $a_1, \dots, a_n$ are pairwise relatively prime, then $a_1 \cdots a_{n-1}$ and a_n are relatively prime.*

Proof. If $a_1 \cdots a_{n-1}$ and a_n have a common factor, then they have a common prime factor p, which must divide a_i for some $i = 1, \dots, n-1$, contradicting pairwise relative primeness for a_i and a_n. □

6.16. **Proposition.** *If $a_1, \dots, a_n$ are* pairwise *relatively prime, and $a_i \mid m$ for all i, then $a_1 \cdots a_n \mid m$.*

Proof. Use induction on n, the special case of $n = 2$ which we already know, and the lemma to show $a_1 \cdots a_k \mid m$. □

Warning. There is another kind of relative primeness of a set of integers, usually just called "relatively prime": a list $a_1, \dots, a_n$ of non-zero integers is relatively prime if the only positive integer wich divides all of them is 1. A pairwise relative prime set is relatively prime, but not conversely. (E.g., $2, 4, 3, 9$ is relatively prime but not pairwise relatively prime.)

7. Modular arithmetic

See section 1.7.

Fix $n \in \mathbb{N}$. This number is called the **modulus**. **modulus**

For $a, b \in \mathbb{Z}$, say "a and b are **congruent modulo** n", write **congruent modulo**

$$a \equiv b \mod n,$$

if $n \mid a - b$; equivalently, if there exists $k \in \mathbb{Z}$ such that $a - b = kn$.

Properties:

- $a \equiv a \mod n$ for all $a \in \mathbb{Z}$. **(Reflexive)**
- $a \equiv b \mod n$ implies $b \equiv a \mod n$, for all $a, b \in \mathbb{Z}$. **(Symmetric)**
- $a \equiv b \mod n$ and $b \equiv c \mod n$ implies $a \equiv c \mod n$, for all $a, b, c \in \mathbb{Z}$. **(Transitive)**

That is, "$\equiv \mod n$" is an **equivalence relation** on $\mathbb{Z}$. **equivalence relation**

The **residue class** (or **congruence class**) modulo n of a is **residue class** **congruence class**

$$[a] := \{\, b \in \mathbb{Z} \mid a \equiv b \mod n \,\} = \{\, a + kn \mid k \in \mathbb{Z} \,\}.$$

This is just the set of all integers which are congruent to a modulo n. Sometimes I'll write $[a]_n$ to make clear what the modulus is.

Example. If $n = 3$, then $[7] = \{\, 7 + 3k \mid k \in \mathbb{Z} \,\} = \{\dots, -5, -2, 1, 4, 7, \dots\}$. Note that any element of this set gives the same congruence class: e.g., $[-2] = [7]$.

The **remainder function** $\mathrm{rem}_n \colon \mathbb{Z} \to \{0, \dots, n-1\}$ computes remainder on division by n: it is defined by $\mathrm{rem}_n(a) = r$, where $a = qn + r$, $r \in \{0, \dots, n-1\}$. **remainder function**

Fact. $r := \mathrm{rem}_n(a)$ satisfies $a \equiv r \mod n$, and $[a] \cap \{0, \dots, n-1\} = \{r\}$.

7.1. **Proposition.** *TFAE:*

- $a \equiv b \mod n$ *(congruence modulo n).*
- $[a] = [b]$ *(equality of sets).*
- $\mathrm{rem}_n(a) = \mathrm{rem}_n(b)$ *(equality of integers).*
- $[a] \cap [b] \neq \varnothing$.

The last says, if $[a]$ and $[b]$ have any element in common, they are the same set.

7.2. **Proposition.** *There exist exacty n congruence classes modulo n, namely* $[0], [1], \dots, [n-1]$.

Note. You don't have to use $0, \dots, n-1$ as standard representatives of congruence classes. For instance, sometimes it is convenient to use some negative numbers. For instance, the congruence classes modulo 7 are:

$$[-3], [-2], [-1], [0], [1], [2], [3].$$

Note that $[-3] = [4]$, $[-2] = [5]$, etc.

7.3. Operations on residue classes.

7.4. **Proposition.** *If $a \equiv a' \mod n$ and $b \equiv b' \mod n$, then*

$$a + b \equiv a' + b' \mod n \quad \textit{and} \quad ab \equiv a'b' \mod n.$$

Proof. We have $a' = a + sn$ and $b' = b + tn$ for some $s, t \in \mathbb{Z}$. Then

$$\begin{aligned} a' + b' &= (a + sn) + (b + tn) = (a + b) + (s + t)n, \\ a'b' &= (a + sn)(b + tn) = ab + (sb + at + stn)n. \end{aligned}$$

□

Let $\mathbb{Z}_n$ = set of residue classes modulo n. We want to define operations on $\mathbb{Z}_n$ by

$$[a] + [b] := [a + b], \qquad [a][b] = [ab].$$

We *must check* that these are well-defined. The problem is that an element of $\mathbb{Z}_n$ can have many names (e.g., $[2]_6 = [-4]_6 = [32]_6$.) The definitions I just gave are based on *the names* of the residue classes. We need to show that if we use different names, we get the same result.

Example. If I say: all students with first name starting with "L" must do only problem 1, while all students with last name starting with "C" do only problem 2.

Proof of well-definedness of $+$ and $\cdot$. The names of $[a]$ are exactly the elements of $[a]$. So we need to show that if $a' \in [a]$ and $b' \in [b]$, then $a' + b' \in [a + b]$. That is, if $a \equiv a' \mod n$ and $b \equiv b' \mod n$, then $a + b \equiv a' + b' \mod n$, which is exactly what was proved above.

The proof for multiplication is the same. □

Thus, $\mathbb{Z}_n$ has addition and multiplication. These follow most of the standard rules that happen in $\mathbb{Z}$:

- Addition and multiplication are commutative and associative.
- Have additive identity $[0]$ and additive inverses.
- Have multiplicative identity.
- Distributive law.

Give proofs in some cases.

7.5. *Remark.* What I've described is how mathematicians set up modular arithmetic. It is tempting to just *define* $\mathbb{Z}_n$ to be the set of symbols $[0], \dots, [n - 1]$, and then define operations by

$$[x] + [y] := [\mathrm{rem}_n(x + y)], \qquad [x][y] := [\mathrm{rem}_n(xy)],$$

using the remainder function after the operation on integers, since these give correct answers.

One problem: how do you see that these operations are associative?

7.6. **Multiplicative inverses.** Some things don't work. **Lecture 07**

- Non-zero elements can multiply to 0. For instance, in $\mathbb{Z}_6$ we have that $[4] \neq [0]$ and $[3] \neq [0]$, but $[4][3] = [12] = [0]$.
- Many elements can have multiplicative inverses. For instance, in $\mathbb{Z}_9$, we have

  $$[1] = [1][1] = [2][5] = [4][7] = [8][8].$$

  Thus, $[1], [2], [4], [5], [7], [8]$ have multiplicative inverses in $\mathbb{Z}_9$, but $[0], [3], [6]$ don't.

7.7. **Proposition.** *$[a] \in \mathbb{Z}_n$ has a multiplicative inverse iff a and n are relatively prime.*

Proof. $[a]$ has a multiplicative inverse in $\mathbb{Z}_n$ iff $ab \equiv 1 \mod n$ for some $b \in \mathbb{Z}$, iff $ab - 1 = kn$ for some $b, k \in \mathbb{Z}$. This last equation can be written

$$1 = ba + (-k)n.$$

Thus, $[a]$ has a multiplicative inverse in $\mathbb{Z}_n$ iff there exist $b, -k \in \mathbb{Z}$ solving this equation, i.e., iff a and n are relatively prime. □

This even gives an algorithm for computing a multiplicative inverse in $\mathbb{Z}_n$, using the Euclidean algorithm. E.g., $n = 19$, $a = 5$, so $1 = (15)(5) + (-4)(19)$.

We write $\Phi(n) \subset \mathbb{Z}_n$ for the set of elements with multiplicative inverses.

For instance, $\Phi(9) = \{[1], [2], [4], [5], [7], [8]\}$. Note that all of these have the form $[2]^k$ for $k \in \mathbb{Z}$.

Another example: $\Phi(8) = \{[1], [3], [5], [7]\}$. All these elements have the property that $[a]^2 = [1]$. Exercise: make a multiplication table for $\Phi(8)$. ($\Phi(8)$ is "isomorphic" to the group of symmetries of the rectangle.)

8. Fermat's little theorem

Recall the binomial theorem.

8.1. **Proposition.** *If a, b are elements of $\mathbb{R}$, and if $n \geq 0$, then*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k},$$

where

$$\binom{n}{k} := \frac{n!}{k!\,(n-k)!}$$

is a positive integer for $0 \leq k \leq n$.

Proof. This can be proved by induction on n, using "Pascal's identity": $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$ when $0 < k \leq n$. □

The integer $\binom{n}{k}$ is called a **binomial coefficient**.

binomial coefficient

We can use this to prove "Fermat's little theorem". The key observation we need is that if p is prime and $0 < k < p$, then p divides $\binom{p}{k}$, using the equation

$$p! = \binom{p}{k} k!\,(p-k)!$$

and the observation that $p \mid p!$, but $p \nmid k!$ and $p \nmid (p-k)!$.

8.2. **Proposition.** *If $a, b \in \mathbb{Z}$ and if p is a prime number, then*

$$(a+b)^p \equiv a^p + b^p \mod p.$$

Proof. The binomial theorem gives

$$(a+b)^p = \sum_{k=0}^{p} \binom{p}{k} a^k b^{p-k} = \binom{p}{p} a^p + \binom{p}{p-1} a^{p-1} b + \cdots + \binom{p}{1} a b^{p-1} + \binom{p}{0} b^p.$$

Since $p \mid \binom{p}{k}$ if $0 < k < p$, the claim follows. □

8.3. **Theorem** (Fermat's Little Theoremc)**.** *Let p be a prime number.*

1. *For all $a \in \mathbb{Z}$ we have $a^p \equiv a \pmod{p}$.*
2. *For all $a \in \mathbb{Z}$ not divisible by p, we have $a^{p-1} \equiv 1 \pmod{p}$.*

Proof. I'll show (1) for $a \geq 0$ using induction. The case of $a = 0$ is obvious. For $a > 0$ we have

$$(a+1)^p \equiv a^p + 1^p \equiv a^p + 1 \mod p.$$

By induction, $a^p \equiv a \pmod{p}$, so we are done.

We can get $a \leq 0$ using downward induction, based on

$$(a-1)^p \equiv a^p + (-1)^p \equiv a^p - 1 \mod p.$$

This is relies on one more fact:

$$(-1)^p \equiv -1 \mod p \qquad \text{for any prime } p.$$

This is easy to prove, but note that the proof is *different* depending on whether $p = 2$ or p is odd!

For (2), note that (1) really says that $[a]^p = [a]$ in $\mathbb{Z}_p$. If $p \nmid a$ then $[a]$ has a multiplicative inverse $[b]$, so we can multiply through by $[b]$ to get $[a]^{p-1} = [1]$, i.e., $a^{p-1} \equiv 1 \pmod{p}$. □

It turns out we can generalize this a bit to allow for some non-prime moduli.

8.4. **Theorem** (Two-prime Fermat)**.** *Let $n = pq$ where p and q are distinct primes, and let $m = \operatorname{lcm}(p-1, q-1) = (p-1)(q-1)/\gcd(p-1, q-1)$. Then for any integer a, if $h \equiv 1 \pmod{m}$, then*

$$a^h \equiv a \mod n.$$

Proof. Write $h = tm + 1$, so $a^h - a = a(a^{tm} - 1)$. We need to show this is divisible by n.

I claim that *either* $p \mid a$ or $p \mid (a^{tm} - 1)$. If $p \nmid a$, then $a^{tm} \equiv 1$ by Fermat's theorem since $p - 1 \mid m$. Therefore p divides their product $a^h - a$.

Likewise, *either* $q \mid a$ or $q \mid (a^{tm} - 1)$, so q also divides their product $a^h - a$.

Since p, q are distinct and thus relatively prime, this means $n = pq$ divides $a^h - a$ as desired. □

9. RSA encryption

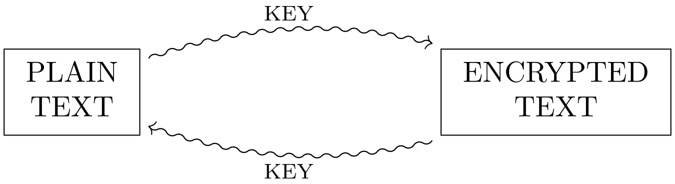
The two-prime version of Fermat's little theorem is the basis of one of the most important encryption algorithms used today (e.g., for secure internet communication), called the **RSA cryptosystem**.

RSA cryptosystem

Before I describe the mathematics involved, here is some context to see how cryptography works in practice.

The following diagram describes a **symmetric encryption** system:

symmetric encryption

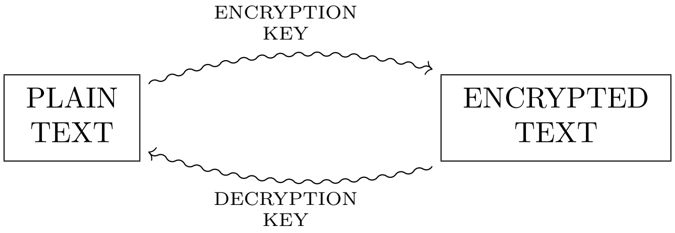


The "plain text" is encoded as a number in a certain fixed range $0 \leq X < n$, which we can think of as an element of $\mathbb{Z}_n$. The "key" k is another integer. The *encryption function* takes as input X and k and gives an encrypted text $Y \in \mathbb{Z}_n$. The *decryption function* takes Y and k and gives back the original X.

The secrecy of this relies on keeping the key secret: without the key, you cannot decrypt an encrypted text Y back to the plain text X. Such systems exist which are very secure, but have a problem: if two people (or computers) in different locations want to communicate securely, they need to exchange a key, and this can't be done securely.

The following diagram describes an **asymmetric encryption** system:

asymmetric encryption



This is almost the same, except there is both an encryption key e and a separate decryption key d. The important feature of such a system is that knowledge of the encryption key e does not tell you anything about the decryption key d and vice versa.

For instance, this allows for **public key encryption**: If I want to receive secure messages from everyone, I generate a key pair d, e, and send e to all my friends (or just post it publicly), while I keep d secret. They send me encrypted messages using e, which only I can decrypt because only I know d.

public key encryption

One of the very first asymmetric encryption systems is **RSA**, first introduced publically in 1977,

RSA

and still the most common system used for secure internet transactions. (It was actually developed independently several years earlier at GCHQ, but this was not known publically until much later.)

Here is how the RSA system works.

(1) Choose two very large prime numbers p, q (on the order of 100s of digits long).
(2) Compute $n = pq$. The "text" we encode with be a number in the range $0 \le X < n$, which we can think of as an element of $\mathbb{Z}_n$.
(3) Compute $m = \operatorname{lcm}(p-1, q-1) = \frac{(p-1)(q-1)}{\gcd(p-1,q-1)}$.
(4) Pick $1 \le e, d < m$ such that $ed \equiv 1 \pmod{m}$. For instance, pick any e relatively prime to m, then solve for a d using the Euclidean algorithm.

The number m is kept completely secret, while n is public. The number e is the encryption key, and d is the decryption key.

Given a plain text X, the corresponding encrypted text is $Y = X^e$ modulo n. To decrypt Y, compute Y^d modulo n, which is the original plain text since

$$(X^e)^d = X^{de} \equiv X \pmod{n}, \qquad \text{since } de \equiv 1 \pmod{m}.$$

That is the core of the system.

The reason this works as a method for *asymmetric encryption* is that, even if n and e are public knowledge, it is infeasible to compute m and d from it. If you could factor n into its primes p and q, then you could easily compute m, and then compute d (modulo m) from that and e. But in fact, there is no known efficient way to factor an integer into primes much better than "trial division", and this is spectacularly inefficient when the prime factors are large.

Fine print: things are actually a bit more complicated. You need to be careful to avoid using certain values for X or p, q or d, e. For instance, $X = 0$ encrypts to $Y = 0$ no matter what the key is, and $d = e = 1$ is a possible choice for keys which is not very secret. Consult a more detailed reference for more. The wikipedia page[1] is a good start.

Note: you still need to find the very large primes p and q. You might think this is hard because factoring is hard. But in fact, it turns out there are relatively efficient primality tests, which don't require doing brute force factor checking.

10. Groups

See section 1.10. Let's review some definitions. **Lecture 08**

A **binary operation** on a set G is a function $G \times G \to G$, often called a *product*, and written by juxtaposition: **binary operation**

$$a, b \quad \mapsto \quad ab.$$

10.1. *Definition.* A **group** G is a set equipped with a binary operation $a, b \mapsto ab$ satisfying the properties: **group**

(a) *Associativity.* For all $a, b, c \in G$, we have $(ab)c = a(bc)$.
(b) *Identity.* There exists an element $e \in G$ with the property that $ae = a = ea$ for all $a \in G$. (An "identity element".)
(c) *Inverse.* For each $a \in G$ there exists an element $b \in G$ such that $ab = e = ba$, where e is an identity element.

Note: a group is necessarily nonempty by (b), which guarantees at least one element.

I'll note several variant definitions:

- If $(G, \cdot)$ satisfies (a) and (b), it is called a **monoid**. **monoid**
- If $(G, \cdot)$ satisfies (a), it is called a **semigroup**. **semigroup**
- An arbitrary pair $(G, \cdot)$ of a set and binary operation is called a **magma**. **magma**

[1] `https://en.wikipedia.org/wiki/RSA_(cryptosystem)`

Also,

- a group is **commutative** or **abelian** if $ab = ba$ for all $a, b \in G$.

commutative
abelian

We have already seen many examples. Here are two which come from modular arithmetic.

10.2. *Example.* Given $n \geq 1$, we have an abelian group $(\mathbb{Z}_n, +)$. Here, the operation is called a *sum*, the identity element is $[0]$, and inverses are given by $[a] \mapsto [-a]$.

Note: the operation is written with "+", then we don't use the notations "e" and "a^{-1}" for identity and inverse, but rather "0" and "$-a$".

10.3. *Example.* Given $n \geq 1$, we have a pair $(\mathbb{Z}_n, \cdot)$, where the operation is multiplication. This is a monoid, but not a group. However, we can instead consider

$$\Phi(n) := \{\, [a] \in \mathbb{Z}_n \mid [a] \text{ has a multiplicative inverse in } \mathbb{Z}_n \,\}.$$

Then $(\Phi(n), \cdot)$ is a group! The operation is well-defined, since if $[a]$ and $[b]$ have multiplicative inverses, so does $[ab]$. (For instance, if $[a][a'] = [1]$ and $[b][b'] = [1]$ then $[ab][a'b'] = [1]$.) Clearly $[1]$ is an identity element, and inverses exist by definition.

The elements of $\Phi(n)$ correspond to $a \in \{0, 1, \dots, n-1\}$ which are relatively prime to a. Example: $\Phi(12) = \{[1], [5], [7], [11]\}$. Exercise: compute the multiplication table, i.e., the grid showing all 16 possible products.

11. Basic properties of groups

Section 2.1.

11.1. **Basic properties of identities and inverses.** First, we check the issues we noticed in the definition of a group: certain things which are required to exist actually turn out to be *unique*.

The second axiom for a group requires an identity element. It is necessarily unique, so we speak of *the* identity element.

11.2. **Proposition.** *A group has only one identity element.*

Proof. If e and e' are identity elements, then we have

$$xe = x = ex, \qquad ye' = y = e'y, \qquad \text{for all } x, y \in G.$$

Take $x = e'$ and $y = e$ to get $e = ee' = e'$. □

The third axiom for a group requires inverses. These are also unique, so we speak of *the* inverse of an element.

11.3. **Proposition.** *An element in a group has only one inverse.*

Proof. Suppose g and h are both inverses of a. The idea is to "reduce" gah in two different ways, using associativity. That is,

$$h = eh = (ga)h = g(ah) = ge = g.$$

□

We write a^{-1} for the unique inverse of the element a.

Another nice fact: although an inverse is required to work on both sides, we can recognize an inverse by checking only one side.

11.4. **Proposition.** *If G is a group and $a, b \in G$, and if $ab = e$, then $b = a^{-1}$ and $a = b^{-1}$.*

Proof. The idea is to "reduce" $a^{-1}ab$ in two different ways, using associativity.

If $ab = e$, then

$$a^{-1} = a^{-1}e = a^{-1}(ab) = (a^{-1}a)b = eb = b.$$

The other identity comes from abb^{-1}. (Note: this proof requires the existence of inverses.) □

Warning. In a monoid, elements are not required to have inverses, but some elements might have inverses. It is still true that an inverse of an element in a monoid is unique if it exists. However, the above proposition is false for a monoid: the proof makes use of the existence of inverses.

11.5. *Exercise.* Let G be the set of all functions $\mathbb{N} \to \mathbb{N}$.

(1) Show that $(G, \circ)$, where $g, f \mapsto g \circ f$ is the operation of composition of functions, is a monoid.
(2) Give an example of elements $g, f \in G$ such that $g \circ f = \mathrm{id}$, but $f \circ g \neq \mathrm{id}$.

It falls out that the inverse of your inverse is you.

11.6. **Corollary.** $(a^{-1})^{-1} = a$.

Another fact: the inverse of a product is the same as the product of inverses, except *in the opposite order.*

11.7. **Corollary.** $(ab)^{-1} = b^{-1}a^{-1}$.

Proof. Check $(b^{-1}a^{-1})(ab) = e$. □

Note: it is also possible that you are your own inverse. Notably, $e^{-1} = e$. Also, the other element in a 2-element group.

11.8. **Cancellation.** Given $a \in G$, define a function $L_a\colon G \to G$ by $L_a(g) = ag$. "Left multiplication by a." Likewise, define $R_a\colon G \to G$ by $R_a(g) = ga$; "right multiplication by a.

11.9. **Proposition.** *L_a and R_a are bijections.*

Proof. L_a has an inverse map, namely $L_{a^{-1}}$! Check

$$L_{a^{-1}}(L_a(g)) = L_{a^{-1}}(ag) = a^{-1}(ag) = (a^{-1}a)g = eg = g,$$

and

$$L_a(L_{a^{-1}}(g)) = L_a(a^{-1}g) = a(a^{-1}g) = (aa^{-1})g = eg = g.$$

Likewise, the inverse of R_a is $R_{a^{-1}}$. □

If we list the elements of the group as $g_1, g_2, \dots, g_n$ (if finite), then $L_a(g_1), L_a(g_2), \dots$ is one of the rows of the multiplication table. Thus, each element appears in the row exactly once; likewise column.

We can also state this the following way.

11.10. **Corollary.** *Let $a, b \in G$. Then the equation $ax = b$ has a unique solution $x \in G$, and the equation $ya = b$ has a unique solution $y \in G$.*

Note: the solutions of the two equations can be different: $x = a^{-1}b$ and $y = ba^{-1}$ need not be the same.

11.11. **Corollary** (Cancellation)**.** *If $ac = bc$ then $a = b$. Likewise, if $ca = cb$, then $a = b$.*

11.12. **The general associative law.** The general associative law says that any way you stick parentheses in an expression $a_1a_2\cdots a_n$, it comes out the same.

Associativity says $(ab)c = a(bc)$, and we celebrate this by simply writing abc for this expression, since the parentheses don't matter.

Given four elements, we see that there are five different products of them (without changing their order):

$$(a(bc))d, \quad ((ab)c)d, \quad (ab)(cd), \quad a(b(cd)), \quad a((bc)d).$$

11.13. *Remark.* You can represent these by "binary trees" with four leaves.

These are all equal. (Proof: $(a(bc))d = ((ab)c)d = (ab)(cd) = a(b(cd)) = a((bc)d)$.) If we remember 3-fold associativity, this is really just three different expressions:

$$(abc)d = (ab)(cd) = a(bcd).$$

We write $abcd$ for any of these.

Given five elements, there are 14 different products of them (I won't list them).

11.14. *Exercise.* For any $n \geq 2$, show that there are $\frac{1}{n}\binom{2n-2}{n-1}$ different ways to parethesize an n-fold product.

Since we have shown that the way we parenthesize 3 or 4 fold doesn't matter, we can reduce the possible 5-fold products to:

$$a(bcde) = (ab)(cde) = (abc)(de) = (abcd)e.$$

Each of these can be proved by using associativity to slide the rightmost element in the left bracket to be the leftmost element in the right bracket. E.g.,

$$(abc)(de) = ((ab)c)(de) = (ab)(c(de)) = (ab)(cde).$$

We can thus write $abcde$ for any of these 5-fold products. We can now state this in arbitrary generality, and the proof will not even require all the structure of the group: just associativity of the product.

11.15. **Proposition.** *Let $(G, \cdot)$ be a set equipped with an associative product (i.e., a semi-group). There is a unique collection of functions $G^n \to G$, defined for each $n \geq 1$, (which we write as $(a_1, \ldots, a_n) \mapsto a_1 a_2 \cdots a_n$), such that*

1. *For $n = 1$, the function $G \to G$ sends $a \mapsto a$.*
2. *For $n = 2$, the function $G \times G \to G$ is the product $(a_1, a_2) \mapsto a_1 a_2$.*
3. *For all $1 \leq k < n$, we have $a_1 a_2 \cdots a_n = (a_1 \cdots a_k)(a_{k+1} \cdots a_n)$.*

Proof. Let's define "$a_1 \cdots a_n$" inductively (on n) by

$$a_1 \cdots a_n := (a_1 \cdots a_{n-1})a_n.$$

Thus,

$$a_1a_2a_3 := (a_1a_2)a_3, \quad a_1a_2a_3a_4 := ((a_1a_2)a_3)a_4, \quad a_1a_2a_3a_4a_5 := (((a_1a_2)a_3)a_4)a_5, \quad \ldots$$

These formulas are forced on us by (3), in the case $k = n - 1$. So if there is a solution, it must be these functions. We still need to check the other cases of (3) to know that it is correct.

We now show (3): that $(a_1 \cdots a_k)(a_{k+1} \cdots a_n) = a_1 \cdots a_n$ for all $1 \leq k < n$ by induction on n. If $k = n - 1$, it's just the definition, while if $k < n - 1$ we have:

$$\begin{aligned}
(a_1 \cdots a_k)(a_{k+1} \cdots a_n) &= (a_1 \cdots a_k)\big((a_{k+1} \cdots a_{n-1})a_n\big) && \text{definition,} \\
&= \big((a_1 \cdots a_k)(a_{k+1} \cdots a_{n-1})\big)a_n && \text{associativity,} \\
&= (a_1 \cdots a_{n-1})a_n && \text{induction,} \\
&= a_1 \cdots a_n && \text{definition.}
\end{aligned}$$

□

From now on, I will leave out parentheses when I can.

Note that the above proof did not use inverses or identity, so it is true in any semigroup.

11.16. **General powers.** You raise a number to an integer power by “repeated multiplication”. We can generalize this rule to an arbitrary group.

Define $a^1 = a$, and for $n > 1$ inductively define $a^{n+1} := a^n a$. Likewise, for $n > 1$ inductively define $a^{-n-1} := a^{-n}a^{-1}$. Set $a^0 = e$. Thus, we have defined a^n for all $n \in \mathbb{Z}$.

11.17. **Proposition.** *For all $m, n \in \mathbb{Z}$, we have $a^m a^n = a^{m+n}$.*

Proof. Prove this on a case by case basis, depending on the signs of $m, n, m + n$. Because of all the cases it is a little tedious: I will prove the case of $m, n > 0$, using induction on n.

Base case. If $n = 1$, then $a^{m+1}a^m a^1$ by definition.

Induction case. Supose $n \geq 1$. Then

$$a^m a^{n+1} = a^m(a^n a) = (a^m a^n)a = a^{m+n}a = a^{m+n+1},$$

using the associativity, the definition $a^n a = a^{n+1}$, and the induction hypothesis $a^m a^n = a^{m+n}$. □

11.18. *Exercise* (Important). Show that $(a^m)^n = a^{mn}$.

11.19. *Remark.* When a group is written with addition, such as $(\mathbb{Z}, +)$ or $(\mathbb{Z}_n, +)$ or $(\mathbb{R}^n, +)$, we write “multiply by integer” instead of “raise to an integer power”. Thus, if $(G, +)$ is a group written with addition, with identity element 0 and inverses $-a$, we define $n, a\colon \mathbb{Z} \times G \to G$ by

$$0a := 0, \quad 1a := a, \quad (-1)a := -a, \quad (n+1)a := na+a \quad \text{if } n > 1, \qquad (n-1)a := na+(-a) \quad \text{if } n < -1.$$

11.20. **Isomorphism.** An **isomorphism** between two groups G and H is a bijection $\phi\colon G \to H$ such that $\phi(g_1g_2) = \phi(g_1)\phi(g_2)$ for all $g_1, g_2 \in G$. (That is, a bijection which *preserves* the product.) **isomorphism**

Say groups G and H are **isomorphic** if there exists an isomorphism $\phi\colon G \to H$. **isomorphic**

11.21. *Remark.* Note that the inverse of an isomorphism is an isomorphism, and composites of isomorphisms are isomorphisms. Being “isomorphic” is an equivalence relation on groups.

We write $G \approx H$ if these groups are isomorphic.

11.22. *Example.* $D_3 = \{e, r, r^2, a, b, c\}$, symmetry group of an equalateral triangle. $S_3 = \{e, (123), (132), (23), (13), (12)\}$ group of permutations of $\{1, 2, 3\}$. These groups are isomorphic.

We can describe an isomorphism $\phi\colon D_3 \to S_3$ by labelling the vertices of the triangle: 1,2,3, and observing that each symmetry $g \in D_3$ permutes the set of vertices. The function ϕ is given by the table

D_3	e	r	r^2	a	b	c
S_3	e	(123)	(132)	(23)	(13)	(12)

Draw pictures.

11.23. *Example.* Let $G = \{e, r_1, r_2, r_3\}$ be the symmetry group of the rectangle. Recall $\mathbb{Z}_4$ is group of congruence classes modulo 4 under additions. Both groups have 4 elements, but they are not isomorphic.

Proof. Every element $x \in G$ has $x^2 = e$, but this is not true in $\mathbb{Z}_4$: $a^2 \neq e$ (really, $[1] + [1] = [2] \neq [0]$). *Exercise.* If $\phi\colon G \to H$ is an isomorphism of groups, and $g^n = e$, then $h^n = e$ for $h = \phi(g)$.

11.24. *Exercise.* Here are five groups which each have exactly 4 elements.

(1) $(\mathbb{Z}_4, +)$.
(2) The group G of symmetries of a rectangle.
(3) $(\Phi(8), \cdot)$.
(4) $(\Phi(5), \cdot)$.
(5) The subset $K = \{I, A, A^2, A^3\}$ of $GL_2(\mathbb{R})$, where $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

Which (if any) of these five groups are isomorphic to each other?

Lecture 09

order

11.25. **Groups of small order.** The **order** $|G|$ of a group is the cardinality of its set.

For any $n \in \mathbb{N}$, $(\mathbb{Z}_n, +)$ is a group of order n.

The group $(\mathbb{Z}, +)$ has infinite (countable) order.

Here is the classification of all groups of order ≤ 5, up to isomorphism.

Order 0. There are no groups of order 0. The axioms for a group require that a group contains an identity element, so a group can never be empty.

Order 1. $\mathbb{Z}_1$ is the unique group of order 1, up to isomorphism.

Order 2. $\mathbb{Z}_2$ is the unique group of order 2, up to isomorphism.

Order 3. $\mathbb{Z}_3$ is the unique group of order 3, up to isomorphism.

Order 4. There are two groups of order 4, up to isomorphism.

- $\mathbb{Z}_4$.
- The group of symmetries of a (non-square) rectangle. (See section 1.2.) (Sometimes called the "Klein 4 group", and written $\mathbb{Z}_2 \times \mathbb{Z}_2$.)

Order 5. $\mathbb{Z}_5$ is the unique group of order 5, up to isomorphism.

The book gives these as exercises. We'll soon have other ways to prove these, but it is worth trying to hack it out by hand.

11.26. *Exercise.* Classify the groups of order 6 up to isomorphism. (This is hard now, but will soon be easy. We already know some examples: $\mathbb{Z}_6$ and S_3.)

isomorphism invariant

11.27. **Abelian groups.** We say that a property of a group is **isomorphism invariant** if whenever $G \approx H$, if G has the property then H also has the property. For instance, the property of having *order n* is an isomorphism invariant, as is the property of being finite.

abelian
commutative

Recall that a group G is **abelian** (or **commutative**) if $ab = ba$ for all $a, b \in G$. Examples: $\mathbb{Z}$, every $\mathbb{Z}_n$, every $\Phi(n)$c, symmetries of a rectangle. All groups of order ≤ 5 are abelian.

We have met some non-abelian groups. For instance, the group D_3 of symmetries of an equalateral triangle is a non-abelian group of order 6. Also $GL_n(\mathbb{R})$ for any $n \geq 2$.

The property of being abelian is isomorphism invariant.

11.28. **Proposition.** *If G and H are isomorphic groups, then G is abelian if and only if H is abelian.*

Proof. Suppose $\phi\colon G \to H$ is an isomorphism. First let's show that if G is abelian then so is H. Given any elements $a', b' \in H$, let $a = \phi^{-1}(a'), b = \phi^{-1}(b') \in G$. Then

$$\phi(ab) = \phi(a)\phi(b) = a'b', \qquad \phi(ba) = \phi(b)\phi(a) = b'a'.$$

Since $ab = ba$, then $a'b' = b'a'$.

Now let's show that if H is abelian so is G. Given $a, b \in G$, let $a' = \phi(a), b' = \phi(b) \in H$. Then as before

$$\phi(ab) = \phi(a)\phi(b) = a'b', \qquad \phi(ba) = \phi(b)\phi(a) = b'a'.$$

Since $a'b' = b'a'$, we have $\phi(ab) = \phi(ba)$, and thus $ab = ba$ since ϕ is injective. (Note: we could instead derive this second part from the first part and the fact that the inverse function $\phi^{-1}\colon H \to G$ is also an isomorphism, which you can easily prove.) □

Thus, D_3 is not isomorphic to $\mathbb{Z}_6$, which is an abelian group of order 6.

The group D_4 of symmetries of a square has 8 elements. It is not isomorphic to a symmetric group, because no symmetric group has order 8. The group G does permute the four corners of a square, and it turns out that this defines an isomorphism between G and a *subgroup* of S_4.

12. Subgroups and cyclic groups

Section 2.2.

12.1. **Subgroups.** A nonempty subset $H \subseteq G$ of a group G is a **subgroup** if it is itself a group with operation inherited from G. We have a special notation for this: we write "$H \leq G$" if the subset H is a subgroup. **subgroup**

12.2. *Example.* Recall the group $GL_n(\mathbb{R})$ of invertible $n \times n$ matrices under matrix multiplication. The subsets $SO(n)$ and $O(n)$ are also groups with operation of matrix multiplication, so these are both subgroups of $GL_n(\mathbb{R})$.

12.3. *Example.* The subset $\Phi(n) \subseteq \mathbb{Z}_n$ is not a subgroup, since the operations are different: the operation on $\mathbb{Z}_n$ is addition, but on $\Phi(n)$ it is multiplication. In fact, the subset $\Phi(n)$ is not even closed under addition.

Other examples of subsets which are not subgroups include $(\mathbb{R}, +)$ and $(\mathbb{R}^\times, \cdot)$, and $\mathrm{Mat}_{n\times n}(\mathbb{R})$ and $GL_n(\mathbb{R})$.

12.4. **Proposition.** *A subset H is a subgroup of G if*

(1) *it is nonempty,*
(2) *for $a, b \in H$, we have $ab \in H$ (using the operation on G) (i.e., H* is closed under multiplication in *G),*
(3) *for $a \in H$, we have $a^{-1} \in H$ (where "a^{-1}" is the inverse of a in G) (i.e., H* is closed under inverses*).*

Proof. (2) says that the group operation restricts to a well-defined function $H \times H \to H$. This restricted operation is still associative.

(1) says that there is an element $a \in H$. (3) says $a^{-1} \in H$ (this means the inverse in the group G). Then $e = aa^{-1} \in H$, so H has an idenity element.

Now (3) implies that the multiplication on H has inverses (which turn out to be the same as inverses in G). □

Note that these imply that e (the identity element of G) is contained in H. In fact, we can just replace (1) with

(1') $e \in H$,

and probably you should.

Some basic examples.

- If G is a group, then G and $\{e\}$ are subgroups.
- The subset H of S_4 consisting of $\{\, \sigma \in S_4 \mid \sigma(4) = 4 \,\}$ is a subgroup. It is straightforward to check this from the definition. The group H has order 6; it is isomorphic to S_3. (*Exercise:* describe an isomorphism $H \to S_3$.)
- The subset H' of S_4 consisting of e and the three disjoint unions of two 2-cycles: (12)(34), (13)(24), (14)(23), is a subgroup. It is isomorphic to V = the Klein 4-group = the symmetries of a rectangle. (*Exercise:* describe an isomorphism $V \to H'$.)
- The subset H'' of S_4 consisting of
  - e,
  - two 4-cycles (1234) and (4321),
  - the three products of disjoint 2-cycles (13)(24), (12)(34), (14)(23),
  - the two 2-cycles (13) and (24),

  form a subgroup of order 8. This is not obvious from the description: you have to check lots of products.

  However, we have met this group H'' before: it is isomorphic to the group D_4 of symmetries of the square. The isomorphism is obtained by labelling the vertices of the square $1, 2, 3, 4$, and seeing how the symmetries permute them. If we label (1,1) as 1, (-1,1) as 2, (-1,-1) as 3, (1,-1) as 4, we get:

D_4	e	r	r^2	r^3	a	b	c	d
H''	e	(1234)	$(13)(24)$	(1432)	$(14)(23)$	$(12)(34)$	(13)	(24)

Draw picture.

12.5. **Subgroups generated by a subset.** Any intersection of subgroups of a group is a subgroup.

12.6. **Proposition.** *For any collection $\{H_i\}_{i\in I}$ of subgroups of G, the intersection $H := \bigcap_i H_i$ is also a subgroup.*

Proof. Check that H (1) contains $e \in G$, (2) is closed under multiplication in G, (3) is closed under inverses in G.

1. Since every H_i is a subgroup, $e \in H_i$, so $e \in H = \bigcap H_i$.
2. If $a, b \in H$, then $a, b \in H_i$ for each i, so $ab \in H_i$ since each H_i is a subgroup, so $ab \in H = \bigcap H_i$.
3. If $a \in H$, thenh $a \in H_i$ for each i, so $a^{-1} \in H_i$ since each H_i is a subgroup, so $a^{-1} \in H = \bigcap H_i$.

□

Given a *subset* S of a group G, there is a "smallest" *subgroup* of G containing S, called the **subgroup generated by** S. We define this by

subgroup generated by S

$$\langle S \rangle := \bigcap_i H_i,$$

where $\{H_i\}$ is the collection of all subgroups $H_i \leq G$ with the property that $S \subseteq H_i$. Observe that

1. $\langle S \rangle$ is a subgroup, since it is an intersection of subgroups,
2. $S \subseteq \langle S \rangle$,
3. If $K \leq G$ is any subgroup such that $S \subseteq K$, then $\langle S \rangle \leq K$. (This is the sense in which $\langle S \rangle$ is the "smallest" subgroup containing the subset S.)

There is another description of $\langle S \rangle$, which tells you exactly what elements are in it.

12.7. **Proposition.** *Given a subset S of a group G, the subgroup $\langle S \rangle$ is equal to the set of elements in G of the form*

$$g_1 \cdots g_k, \qquad \text{either } g_i \in S \text{ or } g_i^{-1} \in S \text{ for all } i.$$

That is, the set of all products of elements which are in S or inverse to an element of S (together with e, which can be thought of as the "empty product").

Proof. Write $T := \{e\} \cup \{\, g_1 \cdots g_k \mid k \geq 1,\ g_i \in S \text{ or } g_i^{-1} \in S \text{ for all } i = 1, \dots, k \,\}$. We need to show

1. T is a subgroup.
2. $S \subseteq T$.
3. If $K \leq G$ is any subgroup with $S \subseteq K$, then $T \subseteq K$.

Then (1) and (2) imply $\langle S \rangle \subseteq T$, and (1) and (3) imply $T \subseteq \langle S \rangle$.

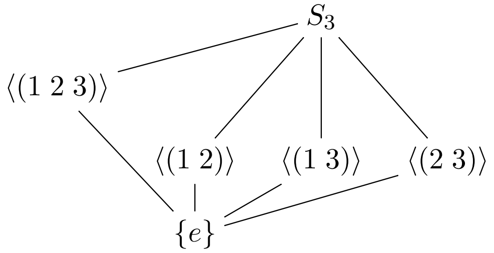
It is easy to check that T is a subgroup. (Note that $(g_1 \cdots g_k)^{-1} = g_k^{-1} \cdots g_1^{-1}$.) We have $S \subseteq T$ by construction. If K is any subgroup which has S as a subset, it is clear that $T \subseteq K$. □

12.8. **Lattice of subgroups.** Example: $S_3 = \{e,\ (1\,2),\ (1\,3),\ (2\,3),\ (1\,2\,3),\ (1\,3\,2)\}$. We can list various subgroups in terms of generating sets.

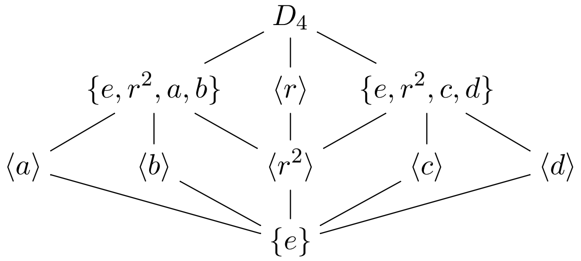
- $\langle (1\,2) \rangle = \{e,\ (1\,2)\}$.
- $\langle (1\,3) \rangle = \{e,\ (1\,3)\}$.
- $\langle (2\,3) \rangle = \{e,\ (2\,3)\}$.
- $\langle (1\,2\,3) \rangle = \{e,\ (1\,2\,3),\ (1\,3\,2)\} = \dots$
- $\dots = \langle (1\,3\,2) \rangle = \{e,\ (1\,3\,2),\ (1\,2\,3)\}$.
- $\langle e \rangle = \{e\} = \dots$

- $\cdots = \langle\rangle = \{e\}$.
- $\langle(1\ 2),\ (1\ 2\ 3)\rangle = S_3 = \ldots$
- $\cdots = \langle(1\ 2),\ (2\ 3)\rangle = \langle S_3\rangle$.

We can picture these in terms of a *lattice of subgroups*. **Lecture 10**



(I probably won't get to this example in class.) Consider symmetries of square (call it $D_4 = \{e, r, r^2, r^3, a, b, c, d\}$.) We have



One of the three order 4 subgroups can be generated by a single element, but the other two *cannot*. Instead, we have:

$$\{e, r^2, a, b\} = \langle r^2, a\rangle = \langle a, b\rangle = \langle r^2, b\rangle, \qquad \{e, r^2, c, d\} = \langle r^2, c\rangle = \langle c, d\rangle = \langle r^2, d\rangle,$$
$$D_4 = \langle r, a\rangle = \langle r, b\rangle = \langle r, c\rangle = \langle r, d\rangle = \langle a, c\rangle = \langle a, d\rangle = \langle b, c\rangle = \langle b, d\rangle = \langle r^3, a\rangle = \cdots .$$

12.9. *Example.* The subgroup $H'' = \langle(1\ 2\ 3\ 4),\ (2\ 4)\rangle$ of S_4 has exactly 8 elements.

12.10. **Cyclic subgroups.** When $S = \{a\}$, the subgroup it generates is called a **cyclic** subgroup, **cyclic**
and denoted $\langle a\rangle$. The element a is a **generator** of $\langle a\rangle$. **generator**
If $G = \langle a\rangle$ for some $a \in G$, then we say that G is a **cyclic group**. **cyclic group**

12.11. **Proposition.** *If $a \in G$, then $\langle a\rangle = \{\, a^k \mid k \in \mathbb{Z}\,\}$. It is always an abelian group.*

Proof. We have observed that $a^m a^n = a^{m+n}$ in any group, so the set is closed under product. Likewise $(a^m)^{-1} = a^{-m}$. It is obviously abelian, since $a^m a^n = a^{m+n} = a^n a^m$. □

12.12. *Example.* Each $\mathbb{Z}_n$ is a cyclic group, since $\mathbb{Z}_n = \langle[1]\rangle$.

Note that there can be more than one choice for the cyclic generator. For instance, $\mathbb{Z}_9 = \langle[1]\rangle = \langle[2]\rangle = \langle[4]\rangle = \ldots$.

12.13. *Example.* Consider $\mathbb{Z}$, under addition. For any $a \in \mathbb{Z}$, there is a cyclic subgroup $\langle a\rangle$. Because we must translate from multiplication to addition, "a^k" gets written as "ka". Thus

$$\langle a\rangle = \{\, ka \mid k \in \mathbb{Z}\,\} = \mathbb{Z}a.$$

If $a \neq 0$, this is an infinite subgroup: $\{\ldots, -3a, -2a, -a, 0, a, 2a, 3a, \ldots\}$. Note that $\langle a\rangle = \langle -a\rangle$.

In fact, $\langle a\rangle$ is isomorphic to $\mathbb{Z}$ (if $a \neq 0$). What is the isomorphism $\mathbb{Z} \to \langle a\rangle$?

Since $\mathbb{Z} = \langle 1\rangle$, it is a cyclic group.

On the other hand, $\langle 0\rangle = \{0\}$.

12.14. *Example.* Given $a, b \in \mathbb{Z}$, we have the subgroup $\langle a, b\rangle$ generated by these two integers. In fact, this is exactly the set

$$\langle a, b\rangle = I(a, b) = \{\, ma + nb \mid m, n \in \mathbb{Z} \,\}$$

of integer combinations of a and b.

Although these subgroups are described by a generating set of two elements, they are actually *cyclic subgroups*, since $I(a, b) = \mathbb{Z}d$ with $d = \gcd(a, b)$.

12.15. **Subgroups of $\mathbb{Z}$.** Recall that $\mathbb{Z}$ is the group of integers under addition. Identity is 0, inverse of a is $-a$.

All subgroups of $\mathbb{Z}$ are cyclic groups, and are all infinite except for the trivial subgroup $\{0\}$.

Notation. Given an integer d, we write $\mathbb{Z}d = \{\, kd \mid k \in \mathbb{Z} \,\}$ for the set of all integer multiples of d. This is clearly the same as $\langle d\rangle$.

12.16. **Proposition.** *All subgroups of $\mathbb{Z}$ are cyclic. Except for the trivial subgroup $\{0\}$, they are infinite. Each non-trivial subgroup is* equal *to $\mathbb{Z}d$ for a unique $d \geq 1$. Each non-trivial subgroup is* isomorphic *to $\mathbb{Z}$.*

Proof. Consider a non-trivial subgroup $H \subseteq \mathbb{Z}$, which being non-trivial has a non-zero element, which can be assumed positive. Therefore there is a *least* positive element $d \in H$, so $\mathbb{Z}d \subseteq H$.

I will show $H \subseteq \mathbb{Z}d$. If $x \in H$, then $r := \operatorname{rem}_d(x)$ is also an element of H (because $r = x - qd$ for some $q \in \mathbb{Z}$). Since $0 \leq r < d$, the minimality of d in $H \cap \mathbb{N}$ implies $r = 0$, so $x = qd$ and thus $x \in \mathbb{Z}d$ as desired.

If $d \neq 0$, an isomorphism $\mathbb{Z} \to \mathbb{Z}d$ is given by $x \mapsto dx$. □

12.17. *Remark* (**Important**). Note that different subgroups H, H' of a group G can be *isomorphic*, but *not equal.* For instance, $\mathbb{Z}2 \neq \mathbb{Z}3$, even though $\mathbb{Z}2 \approx \mathbb{Z}3$.

12.18. **Proposition.** *For integers a, b, we have $\mathbb{Z}a \subseteq \mathbb{Z}b$ if and only if b divides a.*

Proof. Straightforward. □

As a consequence, the *lattice of subgroups of $\mathbb{Z}$* is the basically the reverse of the *divisibility lattice of non-negative integers.*

12.19. **Corollary.** *If a, b are non-zero integers, then $\langle a, b\rangle = \langle d\rangle$, where $d = \gcd(a, b)$.*

Proof. Note that $\langle a, b\rangle = \{\, ma + nb \mid m, n \in \mathbb{Z} \,\} = I(a, b)$. We have proved that this set $= \mathbb{Z}d = \langle d\rangle$. □

In particular, $\langle a, b\rangle = \mathbb{Z}$ iff a and b are relatively prime. E.g., $\mathbb{Z}$ is generated by the set $\{4, 9\}$.

12.20. **Order of an element.** Let's define the **order** of an element $a \in G$ is defined to be the order (= number of elements) in the cyclic subgroup $\langle a\rangle$. It can be either finite (and ≥ 1) or infinity (countable infinite). We have notation $o(a)$ for the order. **order**

Question. When does an element have order 1?

Earlier, I gave a different definition of order of an element earlier, but it is actually the same as this one.

12.21. **Proposition.** *If $a \in G$ has* finite *order n, then n is the least positive integer such that $a^k = e$. If a has* infinite order*, then $a^k \neq e$ for all positive k.*

Proof. Recall that $\langle a\rangle = \{\, a^m \mid m \in \mathbb{Z} \,\}$.

First suppose $\langle a\rangle$ is finite, and write $n = o(a) = |\langle a\rangle|$. Since it is finite, the list of elements $a^1, a^2, a^3, \ldots$ must contain at least one repetition, i.e., there exists some positive $i < j$ such that $a^i = a^j$, and therefore $a^k = e$ with $k = j - i > 0$. Thus by the well-ordering principle there there is a *least positive* k such that $a^k = e$. We need to show that this least positive k is equal to the order n of $\langle a \rangle$.

First note that if $a^k = e$, then $a^m = a^r$ where $r = \mathrm{rem}_k(m)$. This is because $m = qk + r$ with $q \in \mathbb{Z}$ and $r \in \{0, 1, \ldots, k-1\}$, so $a^m = (a^k)^q a^r = e^q a^r = a^r$. Thus $a^0, a^1, \ldots, a^{k-1}$ is a complete list of elements of $\langle a \rangle$, possibly with repetitions. Suppose there is a repetition, i.e., $a^i = a^j$ for some $0 \le i < j < k$, so $a^{j-i} = e$. Since $0 < j - i < k$, this is impossible because k was defined to be minimal with this property. Therefore $k = n$ as desired.

Now suppose $\langle a \rangle$ is infinite. Then we cannot have $a^k = e$ for any positive k, since otherwise the same argument would show that $a^0, a^1, \ldots, a^{k-1}$ is a complete list of elements, contradicting the infinitude of $\langle a \rangle$. □

12.22. **Proposition.** *If $a \in G$ has*

- *$o(a) = n$ finite, then $\langle a \rangle \approx \mathbb{Z}_n$.*
- *$o(a) = \infty$ then $\langle a \rangle \approx \mathbb{Z}$.*

Proof. If the order is finite n, then an isomorphism $\phi\colon \mathbb{Z}_n \to \langle a \rangle$ is given by $\phi([j]) := a^j$; the idea of the proof is in the previous proposition.

If the order is infinite then an isomorphism $\phi\colon \mathbb{Z} \to \langle a \rangle$ is given by $\phi(j) = a^j$. This preserves products. This is obviously surjective, and is injective since if $\phi(i) = \phi(j)$ for $i \neq j$, then $a^{i-j} = e$, which would imply that $\langle a \rangle$ has finite order. □

12.23. *Example.* In the group $D_4 = \{e, r, r^2, r^3, a, b, c, d\}$ of symmetries of a square, e has order 1, r^2, a, b, c, d have order 2, and r, r^3 have order 4.

12.24. **Subgroups of finite cyclic groups.** Consider $n \geq 1$. We consider the subgroups of the finite cyclic group $\mathbb{Z}_n$. The short answer is that all subgroups are cyclic, and the subgroup lattice of $\mathbb{Z}_n$ is the same as the lattice of divisors of n.

12.25. *Example.* Lattice of subgroups of $\mathbb{Z}_{12}$. First note that there are potentially 12 cyclic subgroups, corresponding to the 12 elements. However, many of these coincide.

$$\begin{tikzcd}
 & & \langle[1]\rangle = \langle[5]\rangle = \langle[7]\rangle = \langle[11]\rangle \arrow[dll, dash] \arrow[drr, dash] & & \\
\langle[2]\rangle = \langle[10]\rangle \arrow[d, dash] \arrow[drrrr, dash] & & & & \langle[3]\rangle = \langle[9]\rangle \arrow[d, dash] \\
\langle[4]\rangle = \langle[8]\rangle \arrow[drr, dash] & & & & \langle[6]\rangle \arrow[dll, dash] \\
 & & \langle[0]\rangle = \langle[12]\rangle & &
\end{tikzcd}$$

where

$$\langle[1]\rangle = \mathbb{Z}_{12}, \quad \langle[2]\rangle = \{[0],[2],[4],[6],[8],[10]\},$$
$$\langle[3]\rangle = \{[0],[3],[6],[9]\}, \quad \langle[4]\rangle = \langle[0],[4],[8]\rangle, \quad \langle[6]\rangle = \{[0],[6]\}, \quad \langle[12]\rangle = \{[0]\}.$$

For instance, there are four different elements which generate $\mathbb{Z}_{12}$. (This does not even count the fact that each of these elements has many different names: for instance, it is correct to say that $\mathbb{Z}_{12} = \langle[131]\rangle$.)

Warning. Just as elements of $\mathbb{Z}_n$ have multiple names, so do cyclic subgroups. We are going to show that each cyclic subgroup has a *canonical* name, given by a positive divisor of n. Thus in $\mathbb{Z}_{12}$ we have $\langle[2]\rangle = \langle[10]\rangle = \langle[122]\rangle = \ldots$.

12.26. **Lemma.** *Fix $n \geq 1$. Let $H \subseteq \mathbb{Z}_n$ be a subset, and define the subset*

$$\widetilde{H} := \{\, a \in \mathbb{Z} \mid [a] \in H \,\} = \bigcup_{S \in \mathbb{Z}_n} S,$$

the union of the congruence classes mod n which are elements of H. Then H is a subgroup of $\mathbb{Z}_n$ iff $\widetilde{H}$ is a subgroup of $\mathbb{Z}$.

Proof. This is a straightforward application of the proposition which lets us identify subgroups. I'll write this out here:

Suppose $H \leq \mathbb{Z}_n$.

(1) Since H is non-empty there is some $[a] \in H$, so $a \in \widetilde{H}$.
(2) Suppose $a, b \in \widetilde{H}$. Then $[a+b] = [a] + [b] \in H$ so $ab \in \widehat{H}$.
(3) Suppose $a \in \widetilde{H}$. Then $-[a] = [-a] \in H$ so $-a \in H$.

Thus $\widetilde{H} \leq \mathbb{Z}$.

Suppose $\widetilde{H} \leq \mathbb{Z}$.

(1) Since $\widetilde{H}$ is non-empty there is some $a \in \widetilde{H}$, so $[a] \in H$.
(2) Suppose $[a], [b] \in H$. Then $a + b \in \widetilde{H}$ so $[a] + [b] = [a+b] \in H$.
(3) Suppose $[a] \in H$. Then $-a \in \widetilde{H}$ so $-[a] = [-a] \in H$.

Thus $H \leq \mathbb{Z}_n$. □

12.27. **Proposition.** *Fix $n \geq 1$.*

(1) *Every subgroup of $\mathbb{Z}_n$ is cyclic.*
(2) *For any integer $a \in \mathbb{Z}$, we have that $\langle [a] \rangle = \langle [d] \rangle$ where $d = \gcd(a, n)$.*
(3) *If a, b are divisors of n, then $\langle [a] \rangle \subseteq \langle [b] \rangle$ iff $b \mid a$.*
(4) *Every subgroup of $\mathbb{Z}_n$ is of the form $\langle [d] \rangle$ for a* unique *positive divisor of n. This subgroup has order n/d.*

Proof. Let $H \leq \mathbb{Z}_n$. Then $\widehat{H} = \{\, a \in \mathbb{Z} \mid [a] \in \mathbb{Z}_n \,\}$ is a subgroup of $\mathbb{Z}$, so we have $\widetilde{H} = \langle d \rangle$ for some $d \geq 0$. Therefore $H = \langle [d] \rangle$, which proves (1).

Suppose $H = \langle [a] \rangle$, and let $d = \gcd(a, n)$. To show $\langle [a] \rangle = \langle [d] \rangle$, I'll show (i) $[a] \in \langle [d] \rangle$, and (ii) $[d] \in \langle [a] \rangle$. For (i), note that since $d \mid a$, we have $a = sd$ for some $s \in \mathbb{Z}$, so $[a] = s[d] \in \langle [d] \rangle$. For (ii), note that $d = sa + tn$ for some $s, t \in \mathbb{Z}$. Thus $[d] = s[a] + t[n] = s[a] + t[0] = s[a] \in \langle [a] \rangle$, so $d \in \langle [a] \rangle$. This proves (2).

To prove (3), first notice that for arbitrary $a, b \in \mathbb{Z}$, we have: $\langle [a] \rangle \subseteq \langle [b] \rangle$ iff $[a] \in \langle [b] \rangle$, iff $[a] = s[b]$ for some $s \in \mathbb{Z}$, iff $a = sb + tn$ for some $s, t \in \mathbb{Z}$. Now if $b \mid a$ then clearly $a = sb + 0n$. Conversely, if $a = sb + tn$ and $b \mid n$, then $b \mid a$.

To prove (4), first note that by (1) any subgroup H has the form $\langle [a] \rangle$, and so by (2) has the form $\langle [d] \rangle$ with $d = \gcd(a, n)$, which is a divisor of n. This gives existence of such a d.

For uniqueness, suppose $H = \langle [a] \rangle = \langle [b] \rangle$ for two positive divisors a, b of n. Then by (3) we have $b \mid a$ and $a \mid b$, so $a = b$.

Finally, note that the element $[d]$ must have order n/d in $\mathbb{Z}_n$, since $0 < 1d, 2d, 3d, \dots, (\frac{n}{d} - 1)d < n$. □

So, cyclic subgroups of $\mathbb{Z}_n$ correspond to *positive divisors* of n.

(Note: under this rule, the trivial subgroup $\{[0]\}$ of $\mathbb{Z}_n$ corresponds to the positive divisor n, i.e., $\{[0]\} = \langle [n] \rangle$.)

12.28. **Corollary.** *For any $n \geq 1$ and $a \in \mathbb{Z}$, the order of $[a]$ in $\mathbb{Z}_n$ is n/d, where $d = \gcd(a, n)$.*

Proof. By (2) of the previous proposition, $\langle [a] \rangle = \langle [d] \rangle$, so the order of $[a]$ is the same as the order of $[d]$. By (4) this order is n/d. □

12.29. **Corollary.** *$\mathbb{Z}_n = \langle [a] \rangle$ if and only if a and n are relatively prime.*

12.30. *Example.* The group $\mathbb{Z}_{12}$ has four elements which can be cyclic generators: $\langle [1] \rangle = \langle [5] \rangle = \langle [7] \rangle = \langle [11] \rangle = \mathbb{Z}_{12}$.

12.31. **Corollary.** *For $n \geq 1$, the lattice of subgroups of $\mathbb{Z}_n$ is equivalent to the (opposite of) the lattice of positive divisors of n.*

12.32. **Multiplicative formulation.** What we have proved for $\mathbb{Z}_n$ is true for *any* cyclic group of order n. Let's restate some of the above proposition for a cyclic group $G = \langle g \rangle$ with generator g, which we now write multiplicatively. **Lecture 11**

12.33. **Proposition.** *Let $G = \langle g \rangle$ be a cyclic group of finite order n.*

(1) *Every subgroup of G is cyclic.*
(2) *For every integer $a \in \mathbb{Z}$, we have that $\langle g^a \rangle = \langle g^d \rangle$ where $d = \gcd(a, n)$.*
(3) *If a, b are divisors of n, then $\langle g^a \rangle \subseteq \langle g^b \rangle$ iff $b \mid a$.*
(4) *Every subgroup of G is of the form $\langle g^d \rangle$ for a unique positive divisor of n. This subgroup has order n/d.*

The proof is exactly the same as the one I gave for $\mathbb{Z}_n$, except everything is written with multiplication instead of addition.

13. Dihedral groups

Section 2.3.

13.1. **Symmetries of the disk.** Think of the unit disk in the xy-plane, which I'll think of as a flat disk inside 3-space: $\{ (x, y, 0) \mid x^2 + y^2 \leq 1 \}$. I'll describe its symmetry group D. This will be the set of rotations of 3-space which carry the disk to itself; it is an infinite group, which is a subgroup of $SO(3)$.

There are two kinds of elements in D: any $g \in D$ must send the z-axis to itself. So either $g(e_3) = e_3$ (a *rotation* of the disk), or $g(e_3) = -e_3$ (a *flip* of the disk).

- **Rotations around the z-axis.** For each angle θ, write $r_\theta := \mathrm{Rot}_{e_3}(\theta)$ for rotation by angle θ around the z-axis. These are all elements of D. We have
$$r_\alpha r_\beta = r_{\alpha+\beta}, \qquad r_0 = e, \qquad (r_\theta)^{-1} = r_{-\theta}.$$
Thus, $N := \{ r_\theta \mid \theta \in \mathbb{R} \}$ is a subgroup of D, which I'll call the subgroup of *rotations* of the disk.

  Note that $r_\theta = r_{\theta+2n\pi}$ if $n \in \mathbb{Z}$. Thus, elements in N can be put in one-to-one correspondence with elements of the set $[0, 2\pi)$.
- **Rotations around a line in the xy-plane.** For each angle θ, let ℓ_θ be the line through the origin and $(\cos\theta, \sin\theta, 0)$. Write j_θ for rotation by angle $\pi = 180°$ around the line ℓ_θ. Explicitly, this is
$$j_\theta = \mathrm{Rot}_{u_\theta}(\pi), \qquad u_\theta = (\cos\theta)e_1 + (\sin\theta)e_2.$$
The j_θ are also elements of D. We clearly have
$$j_\theta j_\theta = e,$$
so $(j_\theta)^{-1} = j_\theta$. Each j_θ generates a cyclic subgroup $\langle j_\theta \rangle$ of order 2.

  I will call such elements *flips*.

  Note that $j_\theta = j_{\theta+n\pi}$, if $n \in \mathbb{Z}$. This is because ℓ_θ and $\ell_{\theta+\pi}$ are names for the same line, so $u_{\theta+\pi} = -u_\theta$. The set $\{ j_\theta \mid \theta \in \mathbb{R} \}$ can be put into bijective correspondence with $[0, \pi)$; it is *not* a subgroup, as we will see.

As we have seen, all these elements can be described by multiplication by $3x3$-matrices. For instance, r_θ is given by $R_\theta = \left(\begin{smallmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right)$, while j_0 is given by $J_0 = \left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{smallmatrix}\right)$. You can use the formulas below to work out the matrix J_θ explicitly.

The claim is that

$$D = \{\, r_\theta, j_\theta \mid \theta \in \mathbb{R} \,\} = \{\, r_\theta \mid \theta \in [0, 2\pi) \,\} \cup \{\, j_\theta \mid \theta \in [0, \pi) \,\},$$

and that in fact this exactly enumerates the elements of D: if $g \in D$, then either $g = r_\theta$ for a unique $\theta \in [0, 2\pi)$, or $g = j_\theta$ for a unique $\theta \in [0, \pi)$, but never both.

We need to know how to multiply with flips.

- The composite of two rotations is a rotation. In fact,
$$r_\alpha r_\beta = r_{\alpha+\beta}.$$
- A flip followed by a rotation is a flip. In fact,
$$r_\alpha j_\beta = j_{\beta+\frac{\alpha}{2}}.$$
To see this, picture what happens to the point $u_{\beta+\frac{\alpha}{2}}$: the flip j_β sends this to $u_{\beta-\frac{\alpha}{2}}$, and then r_α rotates this back to $u_{\beta+\frac{\alpha}{2}}$. Thus this point is fixed by the composite, so the axis of the composite will go through it.
- A rotation followed by a flip is also flip. In fact,
$$j_\alpha r_\beta = j_{\alpha-\frac{\beta}{2}}.$$
This time, think about $u_{\alpha-\frac{\beta}{2}}$.
- The composite of two flips is a rotation. In fact,
$$j_\alpha j_\beta = r_{2(\alpha-\beta)}.$$
To see this, first note that each flip switches e_3 and $-e_3$, so a composite of two flips fixes e_3, so the composite is a rotation around the z-axis. There are several ways to figure out the angle:
  - Note that j_α sends $u_{\alpha+\beta} \mapsto u_{\alpha-\beta}$, which you can rewrite as $j_\alpha(u_\theta) = u_{2\alpha-\theta}$. Then $j_\alpha(j_\beta(u_\theta)) = j_\alpha(u_{2\beta-\theta}) = u_{2\alpha-(2\beta-\theta)} = u_{\theta+2(\alpha-\beta)}$, so it is a rotation through the angle $2(\alpha-\beta)$.
  - Alternately, we can use the formulas we've already prove, which give $j_\alpha = r_{2\alpha} j_0$ and $j_\beta = j_0 r_{-2\beta}$, and the fact that flips have order 2:
$$j_\alpha j_\beta = (r_{2\alpha} j_0)(j_0 r_{-2\beta}) = r_{2(\alpha-\beta)}.$$

Thus, the group structure on D is entirely determined by the following "multiplication table":

$$\begin{array}{ll} r_\alpha r_\beta = r_{\alpha+\beta} & \qquad r_\alpha j_\beta = j_{\beta+\frac{\alpha}{2}} \\ j_\alpha r_\beta = j_{\alpha-\frac{\beta}{2}} & \qquad j_\alpha j_\beta = r_{2(\alpha-\beta)} \end{array}$$

together with the identities $r_{\theta+2\pi n} = r_\theta$ and $j_{\theta+\pi n} = j_\theta$ for $n \in \mathbb{Z}$.

There is another way to list the elements, using the formula

$$j_\theta = r_{2\theta} j_0.$$

I will write $j := j_0$. Note that we have an identity $j r_\theta = r_{-\theta} j$. Thus, every element in D can be written uniquely as one of

$$r_\theta,\ \theta \in [0, 2\pi), \qquad r_\theta j,\ \theta \in [0, 2\pi),$$

and the only formulas we need to compute any product are

$$r_\alpha r_\beta = r_{\alpha+\beta}, \qquad r_{\theta+2\pi n} = r_\theta, \qquad j^2 = r_0, \qquad j r_\theta = r_{-\theta} j.$$

13.2. *Example.* Compute the product of $r_{\pi/2}j$ with $r_{\pi/3}j$:

$$(r_{\pi/2}j)(r_{\pi/3}j) = r_{\pi/2}jr_{\pi/3}j = r_{\pi/2}r_{-\pi/3}jj = r_{\pi/6}.$$

13.3. **Dihedral groups.** Write D_n for the symmetry group of a regular n-gon. We put the n-gon inside the disk of radius 1, with a vertex at $(1, 0, 0)$. Thus, D_n is a subgroup of D.

- **Rotations around the z-axis.** Let $r := r_{2\pi/n}$. Then $\langle r\rangle \subseteq D_n$, and $r^n = e$ so $o(r) = n$.
- **Rotations around a line in the xy-plane.** Let $j := j_0$. Then $j \in D_n$, and $j^2 = e$.

  We have
  $$r^k a = r_{2\pi k/n}j = j_{k(\pi/n)} \in D_n.$$
  This is a flip around the line $\ell_{k(\pi/n)}$. The elements $j, rj, \ldots, r^{n-1}j$ are distinct.
- We have that $jr = r^{-1}j$.
- To summarize: D_n consists of $2n$ distinct elements
  $$D_n = \{e,\, r,\, \ldots,\, r^{n-1},\, j,\, rj,\, \ldots,\, r^{n-1}j\} = \{r_0,\, r_{2\pi/n},\, \ldots,\, r_{(n-1)(2\pi/n)},\, j_0,\, j_{\pi/n},\, \ldots,\, j_{(n-1)(\pi/n)}\}.$$
  For instance,
  $$D_4 = \{e,\, r,\, r^2,\, r^3,\, j,\, rj,\, r^2j,\, r^3j\} = \{r_0,\, r_{\pi/2},\, r_\pi,\, r_{3\pi/2},\, j_0,\, j_{\pi/4},\, j_{2\pi/4},\, j_{3\pi/4}\}.$$
  Thus $|D_n| = 2n$. All products in D_n can be derived from the idenities
  $$r^n = e, \qquad j^2 = e, \qquad jr = r^{-1}j.$$
  Note that we can iterate the last identity to get $jr^k = r^{-k}j$ for any $k \in \mathbb{Z}$.

The geometry of D_n is a little different, depending on whether n is even or odd.

- If n is odd, the lines $\ell_{k(\pi/n)}$ connect a vertex to the midpoint of an opposite edge.
- If n is even, then the lines $\ell_{k(\pi/n)}$ connect
  - opposite vertices if k is even,
  - opposite midpoints if k is odd.

Thus, for D_{odd} (e.g., D_3), the "flips" $\{j, rj, \ldots, r^{n-1}j\}$ all have axis through a vertex and an opposite midpoint and so have the "same type", while for D_{even} (e.g., D_4), there are two different types of flip: $\{j, r^2j, \ldots, r^{2n-2}j\}$ and $\{rj, r^3j, \ldots, r^{2n-1}j\}$, whose axes are either along a line connecting vertices, or a line connecting midpoints.

Warning. Many people write "D_{2n}" for the symmetries of a regular n-gon instead of D_n (it is the "dihedral group with $2n$ elements"). You just have to pay attention.

14. Homomorphisms

Section 2.4.

A **homomorphism** is a function $\phi\colon G \to H$ between two groups which preserves multiplication; i.e., $\phi(g_1g_2) = \phi(g_1)\phi(g_2)$ for all $g_1, g_2 \in G$. **homomorphism**

An **isomorphism** is thus a homomorphism which is also a bijection. **isomorphism**

Easy to check consequences. If ϕ is a homomorphism, then

$$\phi(e_G) = e_H$$

and

$$\phi(g^{-1}) = (\phi(g))^{-1}.$$

(To prove the first one, use the fact that in a group the only element g such that $gg = g$ is the identity element. The second fact follows from the first and the definition of inverse.)

14.1. **Image, preimage, and kernel.** Let $\phi\colon G \to H$ be a homomorphism.

- The subset
$$\phi(G) := \{\, \phi(g) \in H \mid g \in G \,\}$$
of H is called the **image** of ϕ. When ϕ is a homomorphism, $\phi(G)$ is a subgroup of H. **image**
- The subset
$$\ker(\phi) := \{\, g \in G \mid \phi(g) = e \,\}$$
is called the **kernel** of ϕ. It is a subgroup of G. **kernel**

We can generalize these to subgroups. Thus, for a subgroup $A \le G$, we have its **image** subgroup **image**

$$\phi(A) := \{\, \phi(a) \mid a \in A \,\} \le H$$

under ϕ. For a subgroup $B \le H$, we have its **preimage** subgroup **preimage**

$$\phi^{-1}(B) := \{\, g \in G \mid \phi(g) \in B \,\} \le G.$$

So $\ker(\phi) = \phi^{-1}(\{e\})$. (Don't confuse this use of "ϕ^{-1}" with an inverse function to ϕ, which might not even exist.)

14.2. **Examples.** Some examples. (In each case, think about image and kernel.)

- The map $\pi\colon \mathbb{Z} \to \mathbb{Z}_n$ sending $\pi(k) = [k]_n$.
- For any group G and element $a \in G$, define $\phi\colon \mathbb{Z} \to G$ by $\phi(k) := a^k$.
- Let V be the set of vertices of a regular n-gon. The dihedral group D_n acts on a regular n-gon, and therefore permutes the vertices. This defines a function $\phi\colon D_n \to \mathrm{Sym}(V)$ to the symmetric group of permutations of the set V. This function is a homomorphism. If we number the vertices $1, 2, \dots, n$, then we can write this as $\phi\colon D_n \to S_n$.

  This homomorphism is always *injective.* It is not usually surjective, since $|D_n| = 2n$ but $|S_n| = n!$, which is certainly larger if $n \ge 4$.

  The subset $\phi(D_n) \subseteq S_n$ is a subgroup of S_n, which is isomorphic to D_n.
- Let $n = 2d$ (even), and consider the set X which consists of *diagonals* of the regular n-gon. The set X has exactly d elements. A symmetry of the n-gon takes a diagonal to a diagonal, and thus we get a homomorphism $\phi\colon D_n \to \mathrm{Sym}(X)$. If we label the diagonals by $1, \dots, d$, then we get a homomorphism $D_{2d} \to S_d$.

  When $n = 4$, so $d = 2$, we have $|D_4| = 8$ but $|\mathrm{Sym}(X)| = 2! = 2$. In this case, the homomorphism is not injective. It is surjective.

  When $n = 6$, so $d = 3$, we have $|D_6| = 12$ and $|\mathrm{Sym}(X)| = 3! = 6$. Is $\phi\colon D_6 \to \mathrm{Sym}(X)$ injective? surjective?

  In fact, it is surjective, but not injective. The *kernel* of ϕ is the set $\phi^{-1}(e) = \{e, r^3\}$, which is a subgroup.

  In general, this $\phi\colon D_{2d} \to S_d$ is never injective, and is not surjective once $d \ge 4$.
- Write $GL(n, \mathbb{R})$ for the set of $n \times n$ real matrices which are invertible. This is a group, under matrix multiplication.

  Write $\mathbb{R}^\times = (\mathbb{R} \smallsetminus \{0\}, \cdot)$ for the group of non-zero reals under multiplication.

  $\det\colon GL(n, \mathbb{R}) \to \mathbb{R}^\times$ is a homomorphism.
- Define a homomorphism $\phi\colon S_n \to GL(n, \mathbb{R})$ as follows. Given $\sigma \in S_n$, let $\phi(\sigma)$ be the **permutation matrix** $P_\sigma = (a_{ij})$ with **permutation matrix**
$$a_{ij} = \begin{cases} 1 & \text{if } i = \sigma(j), \\ 0 & \text{otherwise.} \end{cases}$$
(Give $n = 3$ examples.)

  The matrix P_σ is characterized by the property $P_\sigma e_k = e_{\sigma(k)}$, where e_k represents the kth standard basis vector. Thus, the kth row of P_σ is $e_{\sigma(k)}$, so left multiplication by P_σ permutes the standard basis vectors according to σ.

This is a homomorphism because $P_\tau P_\sigma e_k = P_\tau e_{\sigma(k)} = e_{\tau(\sigma k)} = e_{(\tau\sigma)(k)} = P_{\tau\sigma} e_k$.

- If $\phi\colon G \to H$ and $\pi\colon H \to K$ are homomorphisms, so is the composite function $\pi \circ \phi\colon G \to K$.
- Let $\epsilon\colon S_n \to \mathbb{R}^\times$ be the composite $\epsilon = \det \circ \phi$, where these were defined above. Thus

  $$\epsilon(\sigma) = \det P_\sigma.$$

  Note that the image of ϵ is contained in $\{\pm 1\} \subseteq \mathbb{R}^*$.

  The number $\epsilon(\sigma) \in \{\pm 1\}$ is called the **sign** (or **parity**) of the permutation. A permutation is said to be **even** ($\epsilon = +1$) or **odd** ($\epsilon = -1$).

  sign
  parity
  even
  odd

  Examples: $\epsilon(\text{any 2-cycle}) = -1$, so ϵ is surjective.

  A consequence: any permutation can be written as a product of some number of 2-cycles. The number of 2-cycles isn't fixed, but whether it is even or odd depends exactly on the parity of σ.

  We write $A_n := \ker \epsilon \leq S_n$ for the set of even permutations. This subgroup is called the **nth alternating group**.

  nth alternating group
- Any vector space V is in particular a group under addition. Any linear map $T\colon V \to W$ is then in particular a homomorphism of groups.

Briefly review definition of homomorphism, and image, kernel, and preimage subgroups. **Lecture 12**

14.3. **Kernels and normal subgroups.** The kernel is always a subgroup of a particular type.

A subgroup $N \leq G$ is **normal** if for all $g \in G$ we have $gNg^{-1} = N$, where $gNg^{-1} := \{\, gng^{-1} \mid n \in N \,\}$.

normal

14.4. **Proposition.** *If $N \leq G$ is a subgroup such that $gNg^{-1} \subseteq N$ for all $g \in G$, then N is a normal subgroup.*

Proof. We need to show $N \subseteq gNg^{-1}$. We suppose $n \in N$ and $g \in G$ and try to show $n \in gNg^{-1}$. Here is a "standard trick": we can write n as

$$n = (gg^{-1})n(g^{-1}g) = g(g^{-1}ng)g^{-1} = gxg^{-1}, \qquad x := g^{-1}ng.$$

By hypothesis, $x = g^{-1}n(g^{-1})^{-1} \in g^{-1}N(g^{-1})^{-1} \subseteq N$, and thus we are done.

□

14.5. *Example.* In S_3, $\langle(123)\rangle$ is normal (check), but $\langle(12)\rangle$ is not normal.

14.6. *Example.* If G is an abelian group, then *every* subgroup is normal, since $gxg^{-1} = x$ for all $g, x \in G$, so $gHg^{-1} = H$ for every subgroup H.

14.7. **Proposition.** *If $\phi\colon G \to H$ is a homomorphism, then* $\ker(\phi)$ *is a normal subgroup of G.*

You can check whether a homomorphism is injective by checking its kernel.

14.8. **Proposition.** *A homomorphism $\phi\colon G \to H$ is injective iff* $\ker(\phi) = \{e_G\}$.

Proof. Claim. For $a, b \in G$, we have $\phi(a) = \phi(b)$ if and only if $ab^{-1} \in \ker \phi$. *Proof.* Compute

$$\phi(ab^{-1}) = \phi(a)\phi(b^{-1}) = \phi(a)\phi(b)^{-1}.$$

So (i) if $\phi(a) = \phi(b)$, then $\phi(ab^{-1}) = e_H$ so $ab^{-1}\ker(\phi)$, and (ii) if $ab^{-1} \in \ker(\phi)$, then $e_H = \phi(a)\phi(b)^{-1}$, so $\phi(a) = \phi(b)$.

The result follows easily. □

15. Cosets

Given a group G and a subgroup H, and an element $g \in G$, we write

$$gH := \{\, gh \mid h \in H \,\}, \qquad Hg := \{\, hg \mid h \in H \,\}.$$

Subsets of G of the form gH are called **left cosets**, of the form Hg are called **right cosets**.

left cosets
right cosets

15.1. *Example.* $D_4 = \{e, r, r^2, r^3, j, rj, r^2j, r^3j\}$, with $r^4 = e = j^2$, $jr = r^{-1}j$. Let $H = \langle j \rangle = \{e, j\}$. The left cosets are the four subsets

$$\begin{aligned} eH = jH &= \{e, j\}, \\ rH = rjH &= \{r, rj\}, \\ r^2H = r^2jH &= \{r^2, r^2j\}, \\ r^3H = r^3jH &= \{r^3, r^3j\}. \end{aligned}$$

Note that these four subsets all have the same size, are pairwise disjoint, with union equal to D_4. Note that one of the cosets is actually the subgroup: $eH = H$. The other cosets are not subgroups.

The right cosets are the four subsets

$$\begin{aligned} He = Hj &= \{e, j\}, \\ Hr = Hr^3j &= \{r, r^3j\}, \\ Hr^2 = Hr^2j &= \{r^2, r^2j\}, \\ Hr^3 = Hrj &= \{r^3, rj\}. \end{aligned}$$

Note these are not the same as the left cosets. (More precisely, two of the right cosets are also left cosets, but the other two aren't.)

15.2. *Example.* Let $G = \mathbb{Z}$, and $H = 3\mathbb{Z} = \langle 3 \rangle$. Then the left cosets of H are the sets $k + 3\mathbb{Z} := \{\, k + m \mid m \in 3\mathbb{Z} \,\}$. Note that $k + 3\mathbb{Z} = [k]_3$, the mod 3 congruence class. Thus, the left cosets are precisely $[0]_3, [1]_3, [2]_3$. In this case, they are also the right cosets: $k + 3\mathbb{Z} = 3\mathbb{Z} + k$ since $\mathbb{Z}$ is abelian.

I'll concentrate on left cosets right now. Basically, for every fact about left cosets, there is an analogous one about right cosets.

15.3. **Proposition.** *Let $H \leq G$ be a subgroup. If $X, Y \subseteq G$ are two left H-cosets, then either*

- *$X = Y$, or*
- *$X \cap Y = \varnothing$.*

Furthermore, the set G is the union of all left H-cosets.

Proof. A restatement of the first claim is: *if* X and Y have an element in common, *then* they are the same set.

Since X and Y are cosets, we can write $X = xH$ and $Y = yH$ for some $x, y \in G$. Suppose $a \in X \cap Y = xH \cap yH$ is a common element. Then there exist $h_1, h_2 \in H$ such that $a = xh_1 = yh_2$. This implies

$$x = yh_2h_1^{-1}, \qquad y = xh_1h_2^{-1}.$$

Now consider an arbitrary element $xh \in xH$. Then $xh = yh_2h_1^{-1}h \in yH$. That is, I showed $xH \subseteq yH$. A similar argument gives that $yH \subseteq xH$. Thus $xH = yH$ as desired.

For the second claim, just note that every $g \in G$ is contained in its own coset gH. □

In other words, left cosets partition G into pairwise disjoint subsets. A **partition** of a set X is a collection of subsets $\{S_i\}$ which (i) are pairwise disjoint and (ii) their union is X. (So, every element $x \in X$ is an element of *exactly one* of the S_i.) **partition**

When we write aH, we are *naming* a left coset with the element a. But remember a coset can have many names.

15.4. **Proposition.** *The following are equivalent. ($H \leq G$, $a, b \in G$.)*

1. *$a \in bH$.*
2. *$b \in aH$.*

(3) $aH = bH$.
(4) $b^{-1}a \in H$.
(5) $a^{-1}b \in H$.

Proof. Easy given the previous proposition. □

Review normal subgroups. All subgroups of an abelian group are normal. **Lecture 13**

Review cosets. Example: cosets of $\mathbb{Z}n \leq \mathbb{Z}$ are congruence classes $a + \mathbb{Z}n = [a]_n = \mathbb{Z}n + a$.

The usefulness of this comes from the following.

15.5. **Proposition.** *Any two left cosets of H in G have the same size, which is the size of H. Same for right cosets.*

Proof. Let aH be a left coset. Consider the function $H \to aH$ sending $x \mapsto ax$. This is a bijection, with inverse function $aH \to H$ given by $y \mapsto a^{-1}y$. □

15.6. **Theorem** (Lagrange)**.** *If G is a finite group, and H a subgroup, then $|H|$ divides $|G|$, and $|G| / |H| =$ the number of left cosets of H in $G =$ the number of right cosets of H in G.*

Proof. The set G is partitioned into pairwise disjoint left H-cosets, let's call them $a_1H, a_2H, \dots, a_mH$. Thus, $|G| = \sum_{k=1}^m |a_kH|$. But each coset has size equal to $|H|$, so

$$|G| = \sum_{k=1}^{m} |H| = m\,|H|\,.$$

Thus $|H|$ divides $|G|$ (if finite). □

This immediately gives the consequence

15.7. **Corollary** (Order theorem)**.** *For any finite group G and $a \in G$, we have that $o(a)$ divides $|G|$.*

Proof. $o(a) = |\langle a \rangle|$ divides $|G|$. □

We define the **index** of a subgroup H in G to be the number of left cosets. It is denoted $[G : H]$. **index** It can be finite or infinite. When G is finite, Lagrange's formula says

$$[G : H] = \frac{|G|}{|H|}.$$

Note that index can be finite for infinite groups. *Example.* For $n \neq 0$, $[\mathbb{Z} : \mathbb{Z}n] = n < \infty$, even though $|Z| = |\mathbb{Z}n| = \infty$.

Another example: $D = \{r_\theta, j_\theta\}$ the symmetries of a disk, and $H = \{r_\theta\}$ the subgroup of rotations. Both D and H are infinite, but $[D : H] = 2$: the left cosets are H and j_0H.

What about the number of right cosets? It is always the same, because ofthe following.

15.8. **Proposition.** *There is a bijection between the set of left H-cosets and the set of right H-cosets, defined by the formula*

$$X \quad \mapsto X^{-1} := \{\, x^{-1} \mid x \in X \,\}.$$

Proof. First I check that if we plug in a left coset, then we get a right coset. In fact, if $X = aH$, then

$$X^{-1} = \{\, (ah)^{-1} \mid h \in H \,\} = \{\, h^{-1}a^{-1} \mid h \in H \,\} = \{\, h'a^{-1} \mid h \in H \,\} = Ha,$$

since $h^{-1} \in H$ runs over all elements of H. Thus the left coset aH is sent to the right coset Ha^{-1}.

There are several ways to see this is a bijection. For instance, you can check that the operation $Y \mapsto Y^{-1}$ also defines an inverse function, from right cosets to left cosets. □

15.9. *Exercise.* Why can't I use the formula $aH \mapsto Ha$ to define a bijection between the collections of left and right cosets?

Here's another characterization of normal subgroup.

15.10. **Proposition.** *Let $H \leq G$ be a subgroup. The following are equivalent.*

(1) *H is a normal subgroup.*
(2) *Left cosets are the same as right cosets. That is, $aH = Ha$ for all $a \in G$.*
(3) *All left cosets are contained in right cosets. That is, $aH \subseteq Ha$ for all $a \in G$.*

Proof. (1) ⇒ (2). Suppose H is a normal subgroup. For any $a \in G$, I want to show that $aH \subseteq Ha$ and $Ha \subseteq aH$. First note that if $ah \in aH$ with $h \in H$, then

$$ah = ah(a^{-1}a) = (aha^{-1})a$$

which is in Ha since $aha^{-1} \in aHa^{-1} = H$ since H is normal. Thus $aH \subseteq Ha$. The proof that $Ha \subseteq aH$ is similar:

$$ha = a(a^{-1}ha) \in aH.$$

(2) ⇒ (3) is immediate.

(3) ⇒ (1). Assume $aH \subseteq Ha$ for all $a \in H$. We want to show $aHa^{-1} \subseteq H$ for all $a \in G$. By hypothesis, for any $a \in G$ and $h \in H$, we have $aH \subseteq Ha$, and therefore

$$ah = h'a \qquad \text{for some } h' \in H.$$

This means

$$aha^{-1} = h' \in H.$$

Thus, we have shown that $a \in G$ and $h \in H$ imply $aha^{-1} \in H$, so $aHa^{-1} \subseteq H$ for all $a \in H$. So H is a normal subgroup. □

16. Applications of Lagrange's theorem and the order theorem

Some immediate consequences of the Order Theorem.

16.1. **Proposition.** *Every group G of prime order p is cyclic.*

Proof. Elements in G can only have order 1 or p. Only e can have order 1, so any $a \in G \setminus \{e\}$ has order p, so $G = \langle a \rangle \approx \mathbb{Z}_p$. □

16.2. *Example* (Classification of groups of order 4)*.* Isomorphic to either $\mathbb{Z}_4$ or the "Klein 4-group" V (= symmetries of a non-square rectangle).

If $|G| = 4$, then e is the only element of order 1. By Lagrange, non-identity elements $g \in G \setminus \{e\}$ have $o(g) = 2$ or $o(g) = 4$. If there exists at least one element g of order 4, then $G = \langle g \rangle \approx \mathbb{Z}_4$. If not, then all non-identity elements have order 2. In this case let $a, b, c \in G$ be the distinct non-identity elements. Since they each have order 2 we can already fill in part of the multiplication table:

		$y =$			
		e	a	b	c
$x =$	e	e	a	b	c
	a	a	e		
	b	b		e	
	c	c			e

We can fill in the rest using the fact that every element appears exactly once in each row and each column.

16.3. *Example* (Subgroups of D_5)*.* We have $|D_5| = 10$, so subgroups can have only orders 1,2,5,10. The only subgroup of order 1 is the trivial subgroup $\{e\}$, and the only subgroup of order 10 is the whole group 10. Since 2 and 5 are prime, subgroups of these orders are cyclic groups. So a complete list of subgroups is:

Order 1: $\{e\}$.

Order 2: $\langle j\rangle$, $\langle rj\rangle$, $\langle r^2j\rangle$, $\langle r^3j\rangle$, $\langle r^4j\rangle$.
Order 5: $\langle r\rangle$.
Order 10: D_5.

16.4. *Example* (Subgroups of D_9). We have $|D_9| = 18$, so subgroups can have only orders 1,2,3,6,9,18. There is only one subgroup of order 1 and of order 18. For the primes (2 and 3), we can apply the order theorem to such subgroups: they will have to be cyclic subgroups. For the remaining orders (6 and 9) there might be cyclic subgroups of that order, but also possibly non-cyclic subgroups.

I did this at length in class. First collect the orders of all the elements of the group:

- Order 1: e.
- Order 2: $j, rj, r^2j, r^3j, r^4j, r^5j, r^6j, r^7j, r^8j$.
- Order 3: r^3, r^6.
- Order 9: $r, r^2, r^4, r^5, r^7, r^8$.

So we have some obvious and cyclic subgroups:

- Order 1: $\{e\}$.
- Order 2: $\langle j\rangle, \langle rj\rangle, \langle r^2j\rangle, \langle r^3j\rangle, \langle r^4j\rangle, \langle r^5j\rangle, \langle r^6j\rangle, \langle r^7j\rangle, \langle r^8j\rangle$.
- Order 3: $\langle r^3\rangle$.
- Order 9: $\langle r\rangle$.
- Order 18: $D_9 = \langle r, j\rangle$.

The subgroup $\langle r\rangle$ is the only one of order 9: by the order theorem, a subgroup of order 9 can only contain elements of order 1, 3, or 9, and there are only nine of these altogether.

It remains to determine whether there are subgroups of order 6 (which we know will not be cyclic since there are no elements of order 6). By the order theorem, if $H \leq D_9$ with $|H| = 6$, then H can only have elements of orders 1,2,3,6. The easiset thing to do is to consider subgroups of the form $\langle a_1, \dots, a_k\rangle$ where the a_i are chosen from elements of order 2 and 3, and see what we can get.

For instance, any set of elements from $\{r^k, k \in \mathbb{Z}\}$ can only generate a subgroup of $\langle r\rangle$, and such a subgroup cannot have order 9.

If we take a pair consisting of any element of order 2 and one of order 3, then we check "by hand" that the group they generate has order 6. For instance, starting with r^3 and j, we easily find 6 elements we can write as words in these:

$$e, r^3, r^6, j, r^3j, r^6j,$$

but we can also check that any product of two elements of this list is also in the list (e.g., $jr^3 = r^6j$). Thus $\langle r^3, j\rangle$ is a subgroup of order 6.

Sometimes a pair of elements of order 2 also generates an element of order 6, for instance, $\langle j, r^3j\rangle$ (which is the same as $\langle r^3, j\rangle$), and sometimes they will generate the whole group D_9, for instance $\langle j, rj\rangle$. Using more than two elements can only give larger subgroups, so we are basically done.

Altogether, there are three distinct subgroups of order 6:

- Order 6: $\langle r^3, j\rangle, \langle r^3, rj\rangle, \langle r^3, r^2j\rangle$.

Recall the definition of index, and that $|G| = [G : H]\,|H|$ for finite groups. Note that $|H| = [H : \{e\}]$. **Lecture 14**

We have a generalization of Lagrange's theorem to "chains" of subgroups.

16.5. **Proposition** (Generalized Lagrange). *If G is a group, $H \leq G$ a subgroup of G, and $K \leq H$ a subgroup of H, then $K \leq G$ and $[G : K] = [G : H][H : K]$.*

If $|G| < \infty$, we can just apply Lagrange's theorem: $[G : H]\,[H : K] = \frac{|G|}{|H|}\,\frac{|H|}{|K|} = \frac{|G|}{|K|} = [G : K]$. But we can't use this formula if the groups are infinite. In that case we will need to count cosets directly.

16.6. *Example.* Consider $D_4 \geq H \geq K$ with $H = \langle r^2, j\rangle = \{e, r^2, j, r^2j\}$ and $K = \langle r^2\rangle$, so that $[D_4 : H] = 2$ and $[H : K] = 2$. The following table shows both left H-cosets and all four left K-cosets in G. Notice that there are two left K-cosets contained in eH, namely eK and jK. If we multiply on the left by r, we see that we get exactly two left K-cosets contained in rH.

$$\begin{array}{r|ll} eH = \{e, r^2, j, r^2j\} & eK = \{e, r^2\} & jK = \{j, r^2j\} \\ rH = \{r, r^3, rj, r^3j\} & rK = \{r, r^3\} & rjK = \{rj, r^3j\} \end{array}$$

Proof. Let $m := [G : H]$, the number of left H-cosets in G, and let $n := [H : K]$, the number of left K-cosets in H. The claim that $[G : K] = mn$ is immediate from the following observation.

Claim. Each left H-coset aH is partitioned into exactly n pairwise disjoint left K-cosets.

Proof of claim. Consider the bijection $\phi\colon H \to aH$ defined by $h \mapsto ah$. I claim that the image of any left coset bK which is contained in H is a left K-coset contained in aH. In fact, the image of bK under ϕ is the subset abK, which is obviously a left H-coset.

Thus, if $b_1K, \dots, b_nK \subseteq H$ is the partition of H into pairwise disjoint left K-cosets, then $ab_1K, \dots, ab_nK$ is the partition into pairwise disjoint left K-cosets. □

16.7. *Example.* Let $G = \mathbb{Z}$, $H = \mathbb{Z}3$, and $K = \mathbb{Z}6$. These are all infinite groups, but we have $[G : H] = 3$, $[H : K] = 2$, and $[G : K] = 6$. If we make a table of cosets it looks like this:

$$\begin{array}{r|ll} 0 + H = [0]_3 & 0 + K = [0]_6 & 3 + K = [3]_6 \\ 1 + H = [1]_3 & 1 + K = [1]_6 & 4 + K = [4]_6 \\ 2 + H = [2]_3 & 2 + K = [2]_6 & 5 + K = [5]_6 \end{array}$$

17. Even order theorem

We have shown that the order of an element divides the order of a (finite) group. We can also ask about the converse: if m divides the order of G, must G contain an element of order m? Sometimes the answer is yes.

17.1. **Lemma.** *$o(a) \mid 2$ if and only if $a = a^{-1}$.*

Proof. If $a^2 = e$, then multiplying on the left by a^{-1} gives $a = ea = a^{-1}aa = a^{-1}e = a^{-1}$. If $a = a^{-1}$, then multiplying on the left by a gives $a^2 = aa = aa^{-1} = e$. □

Thus, for every element $a \in G$, the subset $\{a, a^{-1}\}$ has either 1 or 2 elements. It has 1 element exactly if $a^2 = e$, i.e., if either $a = e$ or $o(a) = 2$.

Note that for any two such subsets $C = \{a, a^{-1}\}$ and $C' = \{b, b^{-1}\}$, either (i) $C \cap C' = \varnothing$, or (ii) $C = C'$. This is because, if there is any element in common, then we must have either $a = b$ or $a = b^{-1}$, and either way we have $C = C'$.

Furthermore, every element of G is contained in one of these subsets, since $a \in \{a, a^{-1}\}$. Thus the collection $\{\, C = \{a, a^{-1}\} \mid a \in G \,\}$ is a partition of G into pairwise disjoint subsets (note: these are not cosets).

17.2. **Proposition.** *If G is a finite group with an* even *number of elements, then G contains an element of order 2.*

Proof. Let $n = |G|$. Partition G into subsets of the form $\{a, a^{-1}\}$. Let $r =$ number of such subsets of size 1, and $s =$ number of such subsets of size 2. Then

$$n = r + 2s.$$

Thus n even implies r is even. We also know that $\{e, e^{-1}\} = \{e\}$ has size 1, so $r > 0$ and thus $r \geq 2$. Therefore, there exists a non-identity element $a \in G$ such that $\{a, a^{-1}\} = \{a\}$, so $o(a) = 2$ as desired. □

It is not true that G must contain an element of order 4 if 4 divides $|G|$.

17.3. *Exercise.* Give an example of a group of order $n = |G|$, such that $4 \mid n$, but there is no element of order 4 in G.

This result is actually a special case of *Cauchy's theorem* (which we will prove later): if p is a *prime* which divides $|G|$, then G has an element of order p.

18. Equivalence relations

I'll review the notion of equivalence relation.

By a **relation** on a set X, we mean a subset $R \subseteq X \times X$. We might use notation like "$a \sim b$" to mean "$(a, b) \in R$". **relation**

18.1. *Example.* Relations on the set $\mathbb{Z}$ which have appeared in this course include "$=$", "$\leq$", "$<$", "$|$", "coprime", and "congruence modulo n".

An **equivalence relation** on a set X is a relation $\sim$ which is **equivalence relation**

- **Reflexive:** $a \sim a$ for all $a \in X$,
- **Symmetric:** $a \sim b$ implies $b \sim a$ for all $a, b \in X$,
- **Transitive:** $a \sim b$ and $b \sim c$ implies $a \sim c$ for all $a, b, c \in X$.

18.2. *Example.* The relations "$=$" and "congruence modulo n" are equivalence relations, but the other relations on $\mathbb{Z}$ listed above are not.

18.3. *Example.* On a group G, define a relation so that "$x \sim y$" if and only if *either* $x = y$ or $x = y^{-1}$. This is an equivalence relation.

Given a set X with an equivalence relation $\sim$, an **equivalence class** is a subset of X of the form **equivalence class**

$$[x] := \{\, y \in X \mid y \sim x \,\}$$

for some $x \in X$.

Here are some facts, whose proof is left as an exercise.

- For $x, y \in X$, we have $x \sim y$ iff $[x] = [y]$.
- For $x, y \in X$, either $[x] \cap [y] = \varnothing$ or $[x] = [y]$.
- The union $\bigcup [x]$ of all equivalence classes is X.
- Each $[x]$ is non-empty.

Thus, the collection $\{\, [x] \mid x \in X \,\}$ of all equivalence classes is a **partition** of X: i.e., a set of pairwise disjoint and non-empty subsets whose union is X. In fact, an equivalence relation presents the same information as a partition. **partition**

18.4. **Proposition.** *There is a bijective correpondence between the set of equivalence relations on X, and the set of partitions of X.*

Proof. To an equivalence relation $\sim$, the corresponding partition is the set $\{\, [x] \mid x \in X \,\}$ of equivalence classes. Conversely, given a partition $\{S_i\}$ of X, define a relation so that $x \sim y$ iff there exists an S_i such that $x, y \in S_i$. □

Given a set X with an equivalence relation $\sim$, write Y for the set of equivalence classes $\overline{X} = \{\, [x] \mid x \in X \,\}$. Often this is denoted "$X/\sim$". There is a function $\pi\colon X \to \overline{X}$ called a **quotient function**, defined by $\pi(x) := [x] \in \overline{X}$. **quotient function**

18.5. **Lemma.** *The quotient function is surjective, and for any $x, y \in X$ we have $\pi(x) = \pi(y)$ iff $x \sim y$.*

18.6. **Cosets and equivalence relations.** Given a group G and a subgroup $H \leq G$, we can define a relation on G by

$$a \sim_H b \qquad \text{iff} \qquad b = ah \text{ for some } h \in H.$$

Equivalently, we can say

$$a \sim_H b \qquad \text{iff} \qquad a^{-1}b \in H,$$

or

$$a \sim_H b \qquad \text{iff} \qquad aH = bH.$$

This is an equivalence relation on G, and the equivalence classes are precisely the left H-cosets.

Given a group G and a subgroup H, we write G/H for the set of left cosets. The **quotient function** is the surjective function

quotient function

$$\pi \colon G \to G/H, \qquad \pi(a) := aH$$

which sends an element to the coset it is contained in.

We can do the same thing with right H-cosets. There is an equivalence relation "$a \sim'_H b$" defined by $b^{-1}a \in H$, whose equivalence classes are the right H-cosets. We write $H\backslash G$ for the set of right cosets, with a quotient map

$$\pi' \colon G \to H\backslash G, \qquad \pi'(a) := Ha.$$

Goal. Given a group G and a subgroup H, produce a group structure on the set G/H, so that the quotient function $\pi \colon G \to G/H$ is a homomorphism.

18.7. *Example.* $G = \mathbb{Z}$, $H = n\mathbb{Z}$. Then the quotient $\mathbb{Z}/n\mathbb{Z}$ is the group $\mathbb{Z}_n$. The map $\pi \colon \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ sending $x \mapsto [x]$ is a homomorphism.

Note that if $\pi \colon G \to G/H$ is going to be a homomorphism, then $\pi(e) = eH$ needs to be the identity element of G/H. But we have $\pi(g) = eH$ if and only if $g \in H$, i.e., we would have $\ker(\pi) = H$.

Since kernels are always normal subgroups, we see that we are not going to need to require that H be normal.

19. Quotient groups

Section 2.7.

Recall that a subgroup H is *normal* if $gHg^{-1} = H$ for all $g \in G$.

Let G be a group and $N \leq G$ a *normal subgroup.* We want to define a product on the set of left cosets G/N by

$$aN \cdot bN := (ab)N.$$

19.1. **Proposition.** *This is well-defined.*

Proof. The problem is that the formula for a product that I gave depends on making a choice of *name* (or *representative*) for a coset. We need to show that names don't matter, i.e., that

$$aN = a'N \quad \text{and} \quad bN = b'N \qquad \text{implies} \qquad abN = a'b'N.$$

The hypotheses $aN = a'N$ and $bN = b'N$ imply that there exist $n_1, n_2 \in N$ such that

$$a' = an_1, \qquad b' = bn_2.$$

Then

$$\begin{aligned} a'b' &= an_1bn_2 \\ &= abb^{-1}n_1bn_2 \\ &= (ab)(b^{-1}n_1b)n_2. \end{aligned}$$

Since N is a normal subgroup, $b^{-1}n_1b \in N$, so $(b^{-1}n_1b)n_2 \in N$. Therefore we have proved $a'b' \in abN$, so $a'b'N = abN$ as desired. □

19.2. **Proposition.** *The product we have defined on G/N makes it into a group, and the quotient map $\pi\colon G \to G/N$ is a homomorphism, surjective with* $\ker(\pi) = N$. **Lecture 15**

Proof. To show that G/N is a group:

- The product associative, because it is constructed from the one on G, which is associative: $(aNbN)cN = abNcN = (ab)cN = a(bc)N = aNbcN = aN(bNcN)$.
- There is an identity element eN.
- $g^{-1}N$ is an inverse of gN.

The quotient map is defined by $\pi(g) = gN$, and we have $\pi(a)\pi(b) = aNbN = abN = \pi(ab)$ (i.e., the group structure on G/N was constructed so that π would be a homomorphism).

If $\pi(g) = e_{G/N}$, then $gN = eN$, which is equivalent to $g \in eN = N$.

□

20. Examples of quotient groups

The quotient group construction gives us a way to construct new groups out of old ones.

20.1. *Example.* If $G = \mathbb{Z}$ and $N = \mathbb{Z}n$ and $n > 0$, then G/N is exactly the construction of the group $\mathbb{Z}_n$, with addition.

20.2. *Example.* Condsider $G = D_4$. The subgroup $N = \langle r^2 \rangle$ is normal. To see this, note that $rr^2r^{-1} = r^2$ and $jr^2j^{-1} = r^{-2} = r^2$, so in fact $gr^2g^{-1} = r^2$ for every $g \in G$.

The quotient group $G/N = \{eN, rN, jN, rjN\}$ is isomorphic to the Klein 4 group. To see this, just note that $g^2 \in N$ for all $g \in G$, so $(gN)^2 = g^2N = eN$ for all elements of G/N.

20.3. *Example.* Consider $G = D_6$. The subgroup $N = \langle r^3 \rangle$ is normal. The quotient group $G/N = \{eN, rN, r^2N, jN, rjN, r^2jN\}$ has order 6. What other group of order 6 is it isomorphic to?

20.4. *Example.* $G = \mathbb{R}$ (with addition), and $H = \mathbb{Z}$. Then we get a quotient group $\mathbb{R}/\mathbb{Z}$, whose elements are cosets $[a] = a + \mathbb{Z}$ with $a \in \mathbb{R}$. The quotient map $\pi\colon \mathbb{R} \to \mathbb{R}/\mathbb{Z}$ is a surjective homomorphism with $\ker(\pi) = \mathbb{Z}$.

We actually know another surjective homomorphism from $\mathbb{R}$ with this kernel. Let

$$T = \{\, r_\theta = \mathrm{Rot}_{e_3}(\theta) \in SO(3) \mid \theta \in \mathbb{R} \,\},$$

the group of rotations around the z-axis. We have a homomorphism $\phi\colon \mathbb{R} \to T$ by

$$\phi(x) := \mathrm{Rot}_{e_3}(2\pi x).$$

Note that if $[x] = [y]$, then $y = x + n$ for some $n \in \mathbb{Z}$, so $\phi(x) = \phi(y)$. So ϕ gives the same answer on any element of the same coset of $\mathbb{Z}$. Thus we can define a function $\overline{\phi}\colon \mathbb{R}/\mathbb{Z} \to T$ by the rule

$$\overline{\phi}([x]) := \phi(x) = \mathrm{Rot}_{e_3}(2\pi x).$$

This is actually an isomorphism. We get a diagram of homomorphisms

$$\begin{tikzcd}
\mathbb{R} \arrow[r, two heads, "\phi"] \arrow[d, two heads, "\pi"'] & T \\
\mathbb{R}/\mathbb{Z} \arrow[ur, "\approx", "\overline{\phi}"'] &
\end{tikzcd}$$

where $\overline{\phi} \circ \pi = \phi$ and $\overline{\phi}$ is an isomorphism.

20.5. *Example.* Let $G = GL_n(F)$ for some field F and $n \geq 1$. Let

$$H = \{\, cI \mid c \in F^\times \,\} \subseteq G.$$

It is easy to see that this is a subgroup of G. As a group it is clearly isomorphic to $F^\times$, by $F^\times \to H$ defined by $c \mapsto cI$. In fact, it is a normal subgroup, since

$$A(cI)A^{-1} = cAIA^{-1} = cI.$$

The quotient group $G/H = GL_n(F)/H$ is called a **projective linear group**, and called $PGL_n(F)$. **projective linear group**

21. Normal subgroups of symmetric groups

How do we determine if a subgroup of S_n is normal? We have the following formula for conjugation in a symmetric group.

21.1. **Proposition** (Cycle conjugation formula)**.** *Let $\sigma \in S_n$ be any element, and let $\tau = (a_1\ \cdots\ a_k) \in S_n$ be a k-cycle, where $a_1, \dots, a_k \in \{1, \dots, n\}$. Then*

$$\sigma(a_1\ \cdots\ a_k)\sigma^{-1} = (\sigma(a_1)\ \cdots\ \sigma(a_k)).$$

21.2. *Example.* Let $\sigma = (1\ 4\ 3\ 5\ 2) \in S_5$ and $\tau = (1\ 2\ 3)$. Then

$$(1\ 4\ 3\ 5\ 2)\,(1\ 2\ 3)\,(1\ 4\ 3\ 5\ 2)^{-1} = (4\ 1\ 5).$$

Proof. This is just a calculation. Suppose $x \in \{1, \dots, n\}$. There are two cases:

- If $x = \sigma(a_j)$ for some $1 \le j \le k$, then
$$\begin{aligned}\sigma(a_1\ \cdots\ a_k)\sigma^{-1}(x) &= \sigma(a_1\ \cdots\ a_k)(\sigma^{-1}(\sigma(a_j))) \\ &= \sigma(a_1\ \cdots\ a_k)(a_j) \\ &= \begin{cases} a_{j+1} & \text{if } j < k, \\ a_1 & \text{if } j = k. \end{cases}\end{aligned}$$
- If $x \notin \{\sigma(a_1), \dots, \sigma(a_k)\}$, then $\sigma^{-1}(x) \notin \{a_1, \dots, a_k\}$, so
$$\sigma(a_1\ \cdots\ a_k)\sigma^{-1}(x) = \sigma(\sigma^{-1}(x)) = x.$$

Thus $\sigma\tau\sigma^{-1}$ is the k-cycle $(\sigma(a_1)\ \cdots\ \sigma(a_k))$. □

21.3. *Remark.* This implies a similar formula for products of cycles. For instance:

$$\begin{aligned}\sigma(a_1\ \cdots\ a_k)(b_1\ \cdots\ b_\ell)\sigma^{-1} &= \sigma(a_1\ \cdots\ a_k)\sigma^{-1}\sigma(b_1\ \cdots\ b_\ell)\sigma^{-1} \\ &= (\sigma(a_1)\ \cdots\ \sigma(a_k))(\sigma(b_1)\ \cdots\ \sigma(b_\ell)).\end{aligned}$$

Thus, we get a condiition for a subgroup of a symmetric group to be normal.

21.4. **Proposition.** *Let $H \le S_n$ be a subgroup. Then H is normal iff whenever H contains a permutation σ of a certain cycle type, then it contains* every *permutation of that cycle type.*

21.5. *Example.* Let $G = S_3$. Then $H = \langle (1\ 2\ 3) \rangle = \{e, (1\ 2\ 3), (1\ 3\ 2)\}$ is a normal subgroup. The quotient group has two elements:

$$G/H = \{eH,\ (1\ 2)H\}.$$

It is a cyclic group of order 2, since $[G : H] = 6/3 = 2$.

21.6. *Example.* More generally, let $G = S_n$ with $n \ge 2$. The we have the alternating subgroup

$$A_n = \{\text{even permutations in } S_n\} = \ker[\mathrm{sgn}\colon S_n \to \{\pm 1\}].$$

This is clearly a normal subgroup, since it is the kernel of a homomorphism. We know that $S_n = eA_n \cup (1\ 2)A_n$, so $[S_n : A_n] = 2$, and therefore S_n/A_n is a cyclic group of order 2.

The case of $n = 3$ I gave earlier is a special case of this.

21.7. *Example.* Let $K := \{e,\ (1\,2)(3\,4),\ (1\,3)(2\,4),\ (1\,4)(2\,3)\} \subseteq S_4$. You can check that this is a subgroup, and by the proposition it is a normal subgroup, since it consists of every permutation of cycle types $1+1+1+1$ and $2+2$, and nothing of other cycle types.

The quotient group S_4/K has order 6. What group is it?

21.8. *Remark.* It turns out that symmetric groups have very few normal subgroups. In fact, when $n \neq 4$, the only normal subgroups of S_n are: $\{e\}, A_n, S_n$. (We will prove this for $n = 5$ later.) The only exception is $n = 4$, in which the example $K \leq S_4$ is another normal subgroup.

22. Constructing homomorphisms from quotient groups

Lecture 16

We have shown that given a group G and a normal subgroup $N \leq G$, we obtain a quotient group G/N and a homomorphism $\pi\colon G \to G/N$. Note that the homomorphism π is surjective, and $\ker(\pi) = N$.

The following theorem tells us how to construct homomorphisms out of the quotient group G/N.

22.1. **Theorem.** *Consider*

- *a homomorphism of groups $\phi\colon G \to H$, with kernel* $\ker(\phi) = K$, *and*
- *a normal subgroup $N \leq G$.*

Write $\pi\colon G \to G/N$ for the quotient homomorphism.

If $N \subseteq K$, then *there exists a group homomorphism $\phi'\colon G/N \to H$ such that*

$$\phi'(gN) = \phi(g) \qquad \text{for all } g \in G.$$

The kernel of ϕ' is a subgroup

$$K/N \subseteq G/N.$$

Finally, the images of ϕ and ϕ' are the same: $\phi(G) = \phi'(G/N) \leq H$.

Proof. We are given a potential formula for ϕ' so we just have to check that it is well-defined. That is, we need to check that if $aN = bN$, then $\phi(a) = \phi(b)$. But if $aN = bN$ then $b = an$ for some $n \in N$, and thus

$$\phi(b) = \phi(an) = \phi(a)\phi(n) = \phi(a)e = \phi(a), \qquad \text{since } n \in N \subseteq \ker(\phi).$$

Next we check that ϕ' is a group homomorphism, but this is easy:

$$\phi'(aN \cdot bN) = \phi'(abN) = \phi(ab), \qquad \phi'(aN)\phi'(bN) = \phi(a)\phi(b),$$

and these are equal because ϕ is a homomorphism. Finally, note that $aN \in \ker(\phi')$ iff $a \in K$, iff $aN \subseteq K$. Thus

$$\ker(\phi') = \{\, aN \mid a \in K \,\} = K/N \subseteq G/N.$$

Note that N is also a normal subgroup of K, so that it makes sense to talk about the quotiont group K/N, which is also the subgroup of G/N consisting of cosets aN which are contained in K.

Finally note that for any $h \in H$, we have $h = \phi(g)$ iff $h = \phi'(gN)$, so the images of ϕ and ϕ' are the same. □

Note: this theorem is not actually stated in the book, but it is the form that I actually remember and use most.

There is another way to understand the condition that $\phi'(gN) = \phi(g)$. It is equivalent to saying that $\phi' \circ \pi = \phi$, i.e., that both ways of following an element around the following diagram are the same.

$$N \subseteq \ker(\phi) \qquad \begin{tikzcd} G \arrow[r, "\phi"] \arrow[d, "\pi"'] & H \\ G/N \arrow[ur, dotted, "\phi'"'] & \end{tikzcd}$$

(People say that this diagram of homomorphisms “commutes”.) In fact, ϕ' is *unique*: there is only one homomorphism ϕ' with the property that $\phi' \circ \pi = \phi$. That’s because that property tells us the formula for ϕ'.

Here’s the quick way to state the theorem: A homomorphism $\phi\colon G \to H$ “factors through” the quotient G/N whenever $\phi(N) = \{e\}$.

23. Constructing isomorphisms from quotient groups

The book calls the following a “homomorphism theorem”, but it makes more sense to call it an “isomorphism theorem”. It is the special case of the previous theorem where the constructed homomorphism ϕ' is an isomorphism. You can think of it as saying that all surjective homomorphisms “look like” quotient homomorphisms $G \to G/N$. Or in other words, whenever you have a surjective homomorphism $G \to H$, the group H is isomorphic to a quotient group of G.

23.1. **Theorem** (Isomorphism theorem (called “Homomorphism theorem” in Goodman))**.** *Let $\phi\colon G \twoheadrightarrow H$ be a* surjective *homomorphism, and let $N := \ker(\phi)$. Let $\pi\colon G \to G/N$ be the quotient homomorphism. Then there exists a group isomorphism $\phi'\colon G/N \to H$ such that $\phi' \circ \pi = \phi$.*

$$N = \ker(\phi) \qquad \begin{tikzcd} G \arrow[r, two heads, "\phi"] \arrow[d, "\pi"'] & H \\ G/N \arrow[ur, dotted, "\approx", "\phi'"'] & \end{tikzcd}$$

In fact, ϕ' is unique*: there is only one isomorphism ϕ' satisfying $\overline{\phi} \circ \pi = \phi$.*

Proof. We can apply the previous theorem, since the kernel $K = \ker(\phi)$ is equal to N here, and get a homomorphism $\phi'\colon G/N \to H$. The kernel is $K/N = N/N = \{eN\}$, the trivial subgroup of G/N, so ϕ' is injective. The image is $\phi'(G/N) = \phi(G) = H$, so ϕ' is surjective. □

Examples.

- Let $G = \mathbb{Z}$, and let $H = \langle a \rangle$ be a finite cyclic group of order n, with generator a. Let $\phi\colon G \to H$ be defined by $\phi(k) := a^k$. This is a homomorphism because $a^{k+\ell} = a^k a^\ell$. Then $N = \ker(\phi) = \mathbb{Z}n$. The isomorphism theorem constructs an isomorphism
$$\mathbb{Z}_n = \mathbb{Z}/\mathbb{Z}n \xrightarrow{\approx} \overline{G}.$$
- Let $G = GL_n(R)$, and let $\phi = \det\colon GL_n(\mathbb{R}) \to \mathbb{R}^*$, which is surjective. Recall that $N = \ker(\det)$ is denoted $SL_n(\mathbb{R})$, the group of matrices of determinant 1. The isomorphism theorem constructs an isomorphism
$$GL_n(\mathbb{R})/SL_n(\mathbb{R}) \xrightarrow{\approx} \mathbb{R}^*.$$

23.2. *Example* (S_3 as a quotient of S_4). Start with the set $X := \{1, 2, 3, 4\}$. Make a new set Y, which has 3 elements:

$$S_1 := 12|34, \qquad S_2 := 13|24, \qquad S_3 := 14|23.$$

This is the set of ways to divide X into two equal subsets. We can write the same element muliple ways: $42|13$ is the same as $13|24$, etc.

(Imagine that there are four students: Ashley, Brian, Caitlin, David, and they have to pair off into two teams. There are three ways to do it: (i) Ashley-Brian and Caitlin-David, Ashely-Caitlin and Brian-David, and Ashley-David and Brian-Caitlin.)

We use this to define a homomorphism $\phi\colon S_4 \to S_3$ as follows. An element $g \in S_4$ is a permutation of $\{1, \dots, 4\}$. Let $\phi(g)$ be the permutation of $\{S_1, S_2, S_3\}$ defined by

$$\phi(g)(ab|cd) := g(a)\, g(b) \,|\, g(c)\, g(d).$$

- For instance, if $g = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix} = (1\ 2\ 3)$, then $\phi(g)$ sends
$$\begin{aligned} S_1 = 12|34 &\mapsto 23|14 = 14|23 = S_3, \\ S_2 = 13|24 &\mapsto 21|34 = 12|34 = S_1, \\ S_3 = 14|23 &\mapsto 24|31 = 13|24 = S_2. \end{aligned}$$
That is, $\phi((1\ 2\ 3)) = (1\ 3\ 2)$.
- If $g' = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix} = (1\ 2)(3\ 4)$, then $\phi(g')$ sends
$$\begin{aligned} S_1 = 12|34 &\mapsto 21|43 = 12|34 = S_1 \\ S_2 = 13|24 &\mapsto 24|13 = 13|24 = S_2, \\ S_3 = 14|23 &\mapsto 23|14 = 14|23 = S_3. \end{aligned}$$
That is, $\phi((12)(34)) = e$.

Claims:

- The homomorphism ϕ is surjective.
- Write $K := \ker \phi$. By the isomorphism theorem, we will get an isomorphism $S_4/K \approx S_3$.
- From this it follows that $|K| = 4$.
- The kernel is the subgroup K={e, (1 2)(3 4), (1 3)(2 4), (1 4)(2 3) }.

These claims are left as exercises.

23.3. *Exercise.* Consider a regular polygon with $2n$ vertices, $n \geq 3$. Let X be the set of diagonals of the polygon, so $|X| = n$. Show how the dihedral group D_{2n} (of order $4n$) permutes the set of diagonals, and therefore gives a homomorphism $\phi\colon D_{2n} \to S_n$. Show that $N = \ker(\phi)$ has 2 elements, and therefore D_{2n}/N is isomorphic to a subgroup of S_n. of order $4n/2 = 2n$.

Note: This is not quite what happens when $n = 2$. In that case, $|N| = 4$, so D_4/N has order 2.

24. Correspondence theorem

I'm going to give a simplified statement of the correspondence theorem, compared to the one in the book. **Lecture 17**

24.1. **Theorem** (Correspondence theorem). *Let G be a group with normal subgroup N, and write $\pi\colon G \twoheadrightarrow G/N$ for the quotient homomorphism.*

- *The map $B \mapsto \pi^{-1}B := \{\, g \in G \mid gN \in B \,\}$ gives a bijection*
$$\{\textit{subgroups of } G/N\} \xrightarrow{\sim} \{\textit{subgroups of } G \textit{ containing } N.\}$$
- *Under this bijection, normal subgroups of G/N correspond to normal subgroups of G which contain N.*

Proof. Check that:

- The preimage $\pi^{-1}B$ of a subgroup $B \leq G/N$ is a subgroup of G which contains N.
- We have that $\pi(\pi^{-1}B) = B$, so that the assignment $B \mapsto \pi^{-1}B$ is injective.
- If $A \leq G$ and $N \subseteq A$, then $B := \pi(A)$ is a subgroup of G/N, and $\pi^{-1}B = A$. Thus the assignment $B \mapsto \pi^{-1}B$ is surjective.
- B is normal in G/N iff $A := \pi^{-1}B$ is normal in G.

□

24.2. *Example.* Let $G = \mathbb{Z}$ and $N = \mathbb{Z}4$, so that $G/N = \mathbb{Z}_4$ and $\pi\colon \mathbb{Z} \to \mathbb{Z}_4$ sends $\phi(a) := [a]_4$. The correspondence theorem gives a bijection between subgroups of $\mathbb{Z}_4$ and subgroups of $\mathbb{Z}$ which contain $\mathbb{Z}4$.

$$\begin{tikzcd}
\mathbb{Z} \arrow[d, dash] & & \mathbb{Z}_4 \arrow[d, dash] \\
\mathbb{Z}2 \arrow[d, dash] & \Longleftrightarrow & \{[0]_4, [2]_4\} \arrow[d, dash] \\
\mathbb{Z}4 & & \{[0]_4\}
\end{tikzcd}$$

If we have an isomorphism $\phi\colon H \approx H'$ of groups, then the subgroups of H correspond exactly to the subgroups of H', and under this correspondence the normal subgroups of H correspond to the normal subgroups of H'. (Exercise: prove this.) Given this, we can use the isomorphism theorem to state a more general version of the correspondence theorem (which is the version the book states).

24.3. **Theorem** (Correspondence theorem, more general)**.** *Let $\phi\colon G \to \overline{G}$ be a surjective homomorphism with kernel $N = \ker\phi$.*

- *The map $B \mapsto \phi^{-1}B := \{\, g \in G \mid \phi(g) \in B \,\}$ gives a bijection*
  $$\{\textit{subgroups of } \overline{G}\} \xrightarrow{\sim} \{\textit{subgroups of } G \textit{ containing } N.\}$$
- *Under this bijection, normal subgroups of $\overline{G}$ go to normal subgroups of G which contain N.*

25. Factorization theorem

Here is a consequence of the homomorphism theorem: if you have two nested normal subgroup of G, you get a surjective homomorphism from the "larger quotient" to the "smaller quotient".

25.1. **Corollary.** *Let $N \subseteq K \subseteq G$ be subgroups which are both normal in G. Then $xN \mapsto xK$ defines a surjective homomorphism $G/N \to G/K$, whose kernel is $K/N = \{\, kN \mid k \in K \,\}$. In particular, we get an isomorphism $(G/N)/(K/N) \approx G/K$.*

Proof. We use the homomorphism theorem. Consider

$$\begin{tikzcd}
G \arrow[r, "\pi_{G/K}"] \arrow[d, "\pi_{G/N}"'] & G/K \\
G/N \arrow[ur, dotted, "\psi"'] &
\end{tikzcd}$$

Since $N \subseteq K = \ker(\pi_{G/k})$, the homomorphism ψ exists. It is easy to check that $\ker(\psi) = K/N$, so the isomorphism theorem implies that $(G/N)/(K/N) \approx G/K$. □

26. Product subsets which are subgroups

Recall that if $A, B \leq G$ are subgroups, the intersection $A \cap B$ is also a subgroup.

Given these subgroups, we can also define subsets of products:

$$AB := \{\, ab \in G \mid a \in A,\ b \in B \,\}, \qquad BA := \{\, ba \in G \mid b \in B,\ a \in A \,\}.$$

Note that in general we would not expect $AB \neq BA$, unless G is commutative. It is *not true* in general that the product subsets AB or BA are subgroups.

26.1. *Example.* For instance, take $A = \langle(1\ 2)\rangle$ and $B = \langle(2\ 3)\rangle$, subgroups of S_3. You can check that the set of products AB has order 4, and so cannot be a subgroup of S_3 by Lagrange.

26.2. *Example.* Take $A = \langle r^3 \rangle$ and $B = \langle j \rangle$ in $G = D_6$. Although D_6 is not abelian, it turns out that $AB = BA$ as sets, and in fact this subset is a subgroup, namely $\langle r^3, j \rangle$.

26.3. **Proposition.** *If $A, B \leq G$ are subgroups, then AB is a subgroup of G iff $AB = BA$.*

Proof. $\Longrightarrow$. Suppose AB is a subgroup of G. Therefore $a \in A$ and $b \in B$, we have $ba = (eb)(ae) \in (AB)(AB) = AB$, so $BA \subseteq AB$. Likewise, for any $x \in AB$, we have $x^{-1} \in AB$, so $x^{-1} = ab$ for some $a \in A$, $b \in B$, and therefore $x = b^{-1}a^{-1} \in BA$. This shows $AB \subseteq BA$.

$\Longleftarrow$. Suppose $AB = BA$. Then (i) $e = ee \in AB$, (ii) if $ab \in AB$ then $(ab)^{-1} = b^{-1}a^{-1} \in BA = AB$, and (iii) $ABAB = AABB = AB$, so AB is closed under multiplication. So AB is a subgroup. □

26.4. **Corollary.** *If $A, B \leq G$ and at least one of A or B is a normal subgroup of G, then AB is a subgroup.*

Proof. If B is normal in G, then $gB = Bg$ for all $g \in G$. In particular, $aB = Ba$ for all $a \in A$, and this implies that $AB = BA$. □

26.5. *Exercise.* Suppose A, B are subgroups of G (but AB is not necessarily a subgroup). Let $C = A \cap B$, which is also a subgroup of G. Show that there is a bijection

$$\{\text{left } C\text{-cosets contained in } A\} \leftrightarrow \{\text{left } B\text{-cosets contained in } AB\},$$

defined by $aC \mapsto aB$. Conclude that if A, B are finite subgroups, then

$$|AB| = |B|\,[A : A \cap B] = |A|\,|B| \,/\, |A \cap B|\,.$$

You can sometimes use this to show that subgroups are not normal: if $|AB| = |A|\,|B| \,/\, |A \cap B|$ doesn't divide $|G|$, then neither A nor B is a normal subgroup.

26.6. *Example.* Consider $G = D_9$ with $A = \langle j \rangle$ and $B = \langle rj \rangle$. We can check directly that $AB = \{e, j, rj, r^3\}$ has order 4 (or use the formula). This does not divide $|G|$, so AB is not a subgroup, and so neither A nor B is a normal subgroup.

27. Diamond isomorphism theorem

Let's note an easy to prove fact about intersections of subgroups with normal subgroups.

27.1. **Proposition.** *If $N \trianglelefteq G$ is a normal subgroup, and $H \leq G$ any subgroup, then $N \cap H$ is a normal subgroup of H.*

In particular, if $N \subseteq H$, then N is a normal subgroup of H.

Proof. We know $N \cap H$ is a subgroup of G, and as it is a subset of the subgroup H it is also a subgroup of H.

Need to show: if $x \in N \cap H$ and $h \in H$, then $hxh^{-1} \in N \cap H$. We have $hxh^{-1} \in H$ because H is a subgroup, and $hxh^{-1} \in N$ because N is normal in G.

The final statement is just because if $N \subseteq H$, then $N \cap H = N$. □

Note: I've simplified the statement slightly from the form given in the book.

27.2. **Theorem** (Diamond isomorphism theorem)**.** *Let G be a group with subgroups A and N, and suppose that N is Then*

(a) *$AN = NA$.*
(b) *AN is a subgroup of G containing N as a normal subgroup, and $A \cap N$ is a normal subgroup of A.*
(c) *$AN/N \approx A/(A \cap N)$.*

The idea is that we get a "diamond" of subgroups in G of the form:

$$\begin{tikzcd}
 & AN \arrow[dl, dash] \arrow[dr, dash] & \\
A \arrow[dr, dash] & & N \arrow[dl, dash] \\
 & A \cap N &
\end{tikzcd}$$

Furthermore, $N \trianglelefteq AN$ and $A \cap N \trianglelefteq A$ are normal subgroups, and we get an isomorphism $AN/N \approx A/(A \cap N)$.

Proof. We have already proved (a) before. As noted, this implies that AN is a subgroup of G.

(b) From part (a) we know that $AN = NA$, so this must be a subgroup by the result we proved earlier.

The previous result implies that N is a normal subgroup of AN, and that $A \cap N$ is a normal subgroup of A. To prove (c), we produce an isomorphism as follows. There is a homomorphism $\phi\colon A \to AN/N$ defined by $a \mapsto aN$. Check that $\ker \phi = A \cap N$. Apply the isomorphism theorem to this:

$$\begin{tikzcd}
A \arrow[r, "\phi"] \arrow[d, "\pi"'] & AN/N \\
A/(A \cap N) \arrow[ur, dotted, "\sim"'] &
\end{tikzcd}$$

$\square$

27.3. *Example.* Let $\phi\colon \mathbb{Z} \to \mathbb{Z}_4$ as before with $N = \mathbb{Z}4$, and let $A = \mathbb{Z}6$. These are groups under addition, so we write $A + N$ instead of AN. We have

$$A + N = \mathbb{Z}6 + \mathbb{Z}4 = \mathbb{Z}2,$$
$$A \cap N = \mathbb{Z}6 \cap \mathbb{Z}4 = \mathbb{Z}12.$$

Then

$$\begin{tikzcd}
 & A+N \arrow[dl, dash] \arrow[dr, dash] & & & & \mathbb{Z}2 \arrow[dl, dash] \arrow[dr, dash] & \\
A \arrow[dr, dash] & & N \arrow[dl, dash] & = & \mathbb{Z}6 \arrow[dr, dash] & & \mathbb{Z}4 \arrow[dl, dash] \\
 & A \cap N & & & & \mathbb{Z}12 &
\end{tikzcd}$$

Note: we already know that $\mathbb{Z}a + \mathbb{Z}b = \mathbb{Z}\gcd(a, b)$, and it turns out that $\mathbb{Z}a \cap \mathbb{Z}b = \mathbb{Z}\operatorname{lcm}(a, b)$.

27.4. *Example.* Let $G = S_4$, and let $V = \{e, (1\ 2)(3\ 4), (1\ 3)(2\ 4), (1\ 4)(2\ 3)\}$, a normal subgroup of order 4. Let $A = \langle\sigma\rangle \leq G$ where σ is any 3-cycle, e.g., $\sigma = (2\ 3\ 4)$. Then we get a diamond of subgroups

$$\begin{tikzcd}
 & \langle\sigma\rangle V \arrow[dl, dash] \arrow[dr, dash] & \\
\langle\sigma\rangle \arrow[dr, dash] & & V \arrow[dl, dash] \\
 & \langle\sigma\rangle \cap V &
\end{tikzcd}$$

Since $o(\sigma) = 3$ and V has no elements of order 3, we have that $\langle\sigma\rangle \cap V = \{e\}$. Thus we get an isomorphism

$$\langle\sigma\rangle V/V \approx \langle\sigma\rangle.$$

Note that this implies

$$|\langle\sigma\rangle V| = [\langle\sigma\rangle V : V]\ |V| = |\langle\sigma\rangle|\ |V| = 12.$$

Since σ and V consist only of even permutations, we see that $\langle\sigma\rangle V = A_4$.

27.5. *Example.* Let $G = GL_n(F)$ for some field F and $n \geq 1$. We have subgroups

$$H := SL_n(F) = \{\, A \in G \mid \det A = 1 \,\}, \qquad Z := \{\, cI \mid c \in F^\times \,\}.$$

These are both normal subgroups of G.

We thus get a "diamond" of subgroups

$$\begin{tikzcd}
 & SL_n(F)Z \arrow[dl, dash] \arrow[dr, dash] & \\
SL_n(F) \arrow[dr, dash] & & Z \arrow[dl, dash] \\
 & SL_n(F) \cap Z &
\end{tikzcd}$$

all of which are normal in $GL_n(F)$, and we therefore get isomorphisms

$$SL_n(F)Z/Z \approx SL_n(F)/SL_n(F) \cap Z, \qquad SL_n(F)Z/SL_n(F) \approx Z/SL_n(F) \cap Z.$$

The interesting question is to identify these. In particular, what are $SL_n(F)Z$ and $SL_n(F) \cap Z$? The answer will depend on n and F.

27.6. *Exercise.* Show that if n is odd, then $SL_n(\mathbb{R})Z = GL_n(\mathbb{R})$ and $SL_n(\mathbb{R}) \cap Z = \{I\}$, and therefore that $PGL_n(\mathbb{R}) \approx SL_n(\mathbb{R})$.

27.7. *Exercise.* Show that if n is even, then $SL_n(\mathbb{R})Z = GL_n^+(\mathbb{R})$, the subgroup of real matrices A with $\det A > 0$, and that $SL_n(\mathbb{R}) \cap Z = \{\pm I\}$. Conclude that $PGL_n(\mathbb{R})$ contains an index 2 subgroup which is isomorphic to $SL_n(\mathbb{R})/\{\pm I\}$.

28. Products of groups

28.1. **Direct products of groups.** Goodman 3.1. **Lecture 18**

Given two groups G and H, the *direct product* (or just *product*) $G \times H$ is the group defined as follows: $G \times H$ is the set of ordered pairs (g, h) with $g \in G$ and $h \in H$, and the product is defined by

$$(g, h)(g', h') := (gg', hh').$$

In other words, the product is "component-wise".

Verify that this is a group: $e_{G\times H} = (e_G, e_H)$ and $(g, h)^{-1} = (g^{-1}, h^{-1})$.

If we have a list of groups $G_1, \dots, G_n$, we can form a product group $G_1 \times \cdots \times G_n$ in the same way, where elements are $(g_1, \dots, g_n)$ with $g_i \in G_i$.

28.2. *Example.* Let $G = \langle a\rangle$ with $o(a) = 2$, and let $H = \langle b\rangle$ with $o(b) = 2$. Then $G \times H = \{(e_G, e_H), (a, e_H), (e_G, b), (a, b)\}$. We have $(a, e)^2 = (e, b)^2 = (a, b)^2 = (e, e)$, and $(a, e)(e, b) = (a, b)$. It is not hard to see that $G \times H$ is isomorphic to the Klein 4 group.

In other words, the Klein 4-group (i.e., symmetries of a rectangle) is isomorphic to this product. Since $G \approx \mathbb{Z}_2$ and $H \approx \mathbb{Z}_2$, we can say that this is isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$.

Note that in general $|G \times H| = |G| \cdot |H|$.

Also, note the following.

28.3. **Proposition.** *If $G_1 \approx G_2$ and $H_1 \approx H_2$, then $G_1 \times H_1 \approx G_2 \times H_2$.*

Proof. If $\phi_1 \colon G_1 \to H_1$ and $\phi_2 \colon G_2 \to H_2$ are isomorphisms, consider $\phi \colon G_1 \times G_2 \to H_1 \times H_2$ by $\phi(g_1, g_2) := (\phi_1(g_1), \phi_2(g_2))$, and show ϕ is an isomorphism. (Leave as exercise.) □

28.4. *Exercise.* Show that there is an isomorphism $G \times H \to H \times G$.

Note that a product group $G \times H$ contains two special subgroups.

$$G' := \{\, (g, e_H) \mid g \in G \,\}, \qquad H' := \{\, (e_G, h) \mid h \in H \,\}.$$

Exercise: Both of these are normal subgroups of $G \times H$, and there are isomorphisms $G \simeq G'$ and $H \simeq H'$.

Note that $G \times H$ has other subgroups, including ones not contained in either G' or H'.

28.5. *Exercise.* Let G be a group, and consider the product group $P := G \times G$. Show that the subset $D := \{\, (g, g) \mid g \in G \,\}$ is a subgroup of P. Give an example in which D is not a normal subgroup of P.

28.6. *Exercise.* D_6 (of order 12) is isomorphic to a direct product of two groups, of order 2 and 6 respectively. Construct such an isomorphism.

29. Products of finite cyclic groups

We have seen that $\mathbb{Z}_2 \times \mathbb{Z}_2$ is not a cyclic group. However

29.1. *Example.* Let $G = \langle a \rangle$ with $o(a) = 2$ and $H = \langle b \rangle$ with $o(b) = 3$. Then a complete list of elements in $G \times H$ is given by

$$(e, e), (a, e), (e, b), (a, b), (a^2, e), (a^2, b).$$

Set $x := (a, b)$. Then

$$x^0 = (e, e), x^1 = (a, b), x^2 = (e, b^2), x^3 = (a, e), x^4 = (e, b), x^5 = (a, b^2),$$

and $x^6 = x^0$. In other words, $\mathbb{Z}_2 \times \mathbb{Z}_3 \approx \mathbb{Z}_6$.

29.2. **Proposition.** *Let $a, b \geq 1$. The formula*

$$\phi(x) := ([x]_a, [x]_b)$$

defines a homomorphism $\phi \colon \mathbb{Z} \to \mathbb{Z}_a \times \mathbb{Z}_b$ to the product group, i.e., $\phi(x + y) = \phi(x) + \phi(y)$. Furthermore, $\ker(\phi) = m\mathbb{Z}$, where m is the least common multiple of a and b, i.e., the smallest positive integer which is a multiple of a and b.

Proof. We check the homomorphism condition:

$$\begin{aligned} \phi(x) + \phi(y) &= ([x]_a, [x]_b) + ([y]_a, [y]_b) \\ &= ([x]_a + [y]_a, [x]_b + [y]_b) \\ &= ([x + y]_a, [x + y]_b) \\ &= \phi(x + y). \end{aligned}$$

By definition, $x \in \ker(\phi)$ if and only if $[x]_a = [0]_a$ and $[x]_b = [0]_b$. In other words, $\ker(\phi) = a\mathbb{Z} \cap b\mathbb{Z}$, the set of all common multiples of a and b. We know that all subgroups of $\mathbb{Z}$ are cyclic, so $\ker(\phi) = m\mathbb{Z}$; this m is the least common multiple □

Using the homomorphism theorem, we obtain an *injective* homomorphism $\overline{\phi}\colon \mathbb{Z}_m \to \mathbb{Z}_a \times \mathbb{Z}_b$:

$$\begin{tikzcd}
\mathbb{Z} \arrow[r, "{\phi\colon x\mapsto([x]_a,[x]_b)}"] \arrow[d] & \mathbb{Z}_a \times \mathbb{Z}_b \\
\mathbb{Z}_m = \mathbb{Z}/\mathbb{Z}m \arrow[ur, dotted, tail, "{[x]_m\mapsto([x]_a,[x]_b)}"'] &
\end{tikzcd}$$

because $\mathbb{Z}m = \ker\phi$. We know the orders of these groups: $|\mathbb{Z}_m| = m$ and $|\mathbb{Z}_a \times \mathbb{Z}_b| = ab$. Therefore, the map $\widetilde{\phi}$ is an isomorphism if and only if $m = ab$.

The statement $m = ab$ is equivalent to saying that a and b are relatively prime. We get the following.

29.3. **Theorem.** *If a and b are relatively prime, then $\mathbb{Z}_{ab} \to \mathbb{Z}_a \times \mathbb{Z}_b$ defined by $[x]_{ab} \mapsto ([x]_a, [x]_b)$ is an isomorphism of groups.*

The converse is also true (if a and b are not relatively prime, then $\mathbb{Z}_{ab}$ is not isomorphic to $\mathbb{Z}_a \times \mathbb{Z}_b$), but we have not proved it yet.

For instance, $\mathbb{Z}_6 \approx \mathbb{Z}_2 \times \mathbb{Z}_3$. On the other hand, $\mathbb{Z}_4 \not\approx \mathbb{Z}_2 \times \mathbb{Z}_2$.

The above is equivalent to what is called the "Chinese remainder theorem":

29.4. **Theorem.** *Suppose $a, b \geq 1$ are relatively prime. Then for any integers α, β, there exists $x \in \mathbb{Z}$ such that $x \equiv \alpha \mod a$ and $x \equiv \beta \mod b$. Furthermore, such a solution x is unique modulo $m = ab$.*

Proof. Since $\phi\colon \mathbb{Z}_m \to \mathbb{Z}_a \times \mathbb{Z}_b$ defined by $\phi([x]_m) := ([x]_a, [x]_b)$ is a bijection, given α, β there is a unique $[x]_m \in \mathbb{Z}_m$ such that $([x]_a, [x]_b) = ([\alpha]_a, [\beta]_b)$. □

More generally, we have

29.5. **Theorem.** *Suppose $a_1, \dots, a_n$ are pairwise relatively prime, and let $a = a_1 \cdots a_n$. Then the map $\mathbb{Z}_a \to \mathbb{Z}_{a_1} \times \cdots \times \mathbb{Z}_{a_n}$ given by $[x]_a \mapsto ([x]_{a_1}, \dots, [x]_{a_n})$ is an isomorphism of groups.*

Proof. As in the case of two factors, the formula $x \mapsto ([x]_{a_1}, \dots, [x]_{a_n})$ defines a homomorphism $\phi\colon \mathbb{Z} \to \mathbb{Z}_{a_1} \times \cdots \times \mathbb{Z}_{a_n}$ with kernel $K = \mathbb{Z}a_1 \cap \cdots \cap \mathbb{Z}a_n$. Note that $a \in K$, and that the elements of K are precisely the integers x such that $a_i \mid x$ for all $i = 1, \dots, n$. Since $a_1, \dots, a_n$ are pairwise relatively prime, we know that $a_i \mid x$ for all i implies $a \mid x$ for all i, and therefore $a \mid x$. Thus we have proved $K = \mathbb{Z}a$, and so by the isomorphism theorem, we get an isomorphism

$$\phi\colon \mathbb{Z}a \to \mathbb{Z}_{a_1} \times \cdots \times \mathbb{Z}_{a_n}.$$

□

29.6. **The group of multiplicative units mod n.** The set $\mathbb{Z}_n$ is not merely a group, but a *ring*. That is, we can not merely add congruence classes, but multiply them, by the rule $[x]_n[y]_n = [xy]_n$.

The multiplication operation on $\mathbb{Z}_n$ does not define a group structure: it is associative, commutative, and has unit $[1]$, but inverses may not exist. In fact, we showed the following:

29.7. **Proposition.** *Let $n \geq 1$ and $a \in \mathbb{Z}$. The following are equivalent.*

(1) *a and n are relatively prime.*
(2) *There exists $b \in \mathbb{Z}$ such that $ab \equiv 1 \mod n$.*

Recall that $\Phi(n) \subseteq \mathbb{Z}_n$ is the set of congruence classes $[a]$ which admit multiplicative inverses, or equivalently which are such that a is relatively prime to n. It is not a subgroup of $\mathbb{Z}_n$, but does admit a group structure given by *multiplication.*

Warning. $\Phi(n)$ is a *subset* of $\mathbb{Z}_n$, but not a *subgroup*, since the products are defined differently.

29.8. **Theorem.** *If $a, b \geq 1$ are relatively prime, then there is an isomorphism $\Phi(ab) \to \Phi(a) \times \Phi(b)$ defined by $[x]_{ab} \mapsto ([x]_a, [x]_b)$.*

Proof. We already have a bijection of sets $\phi' \colon \mathbb{Z}_{ab} \to \mathbb{Z}_a \times \mathbb{Z}_b$ defined by this formula.

I claim that this function restricts to a bijection of sets $\phi \colon \Phi(ab) \to \Phi(a) \times \Phi(b)$. This amounts to observing that if $x \in \mathbb{Z}$, then $\gcd(x, ab) = 1$ iff $\gcd(x, a) = \gcd(x, b) = 1$. (Prove that $\gcd(x, ab) > 1$ iff either $\gcd(x, a) > 1$ or $\gcd(x, b) > 1$.

Now note that ϕ is also a homomorphism of groups, because multiplication "makes sense" modulo an integer:

$$\begin{aligned}\phi([x]_{ab})\phi([y]_{ab}) &= ([x]_a, [x]_b)([y]_a, [y]_b) \\ &= ([x]_a[y]_a, [x]_b[y]_b) \\ &= ([xy]_a, [xy]_b) \\ &= \phi([xy]_{ab}).\end{aligned}$$

□

We can generalize this to arbitrarily many factors.

29.9. **Theorem.** *Suppose $a_1, \dots, a_n$ are pairwise relatively prime, and let $a = a_1 \cdots a_n$. Then the map $\Phi(a) \to \Phi(a_1) \times \cdots \times \Phi(a_n)$ given by $[x]_a \mapsto ([x]_{a_1}, \dots, [x]_{a_n})$ is an isomorphism of groups.*

29.10. *Example.* We have an isomorphism

$$\Phi(24) \approx \Phi(8) \times \Phi(3),$$

since 8 and 3 are relatively prime. The theorem doesn't let us reduce this any further, but we can verify by hand that $\Phi(8) \approx \mathbb{Z}_2 \times \mathbb{Z}_2$, and $\Phi(3) \approx \mathbb{Z}_2$, so that $\Phi(24) \approx \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$.

29.11. **Euler ϕ-function.** Given $n \in \mathbb{N}$, we write $\phi(n) := |\Phi(n)|$ for the order of the group of modular units. This is exactly the number of integers $d \in \{1, 2, \dots, n\}$ which are relatively prime to n. The function $\phi \colon \mathbb{N} \to \mathbb{N}$ is called the **Euler ϕ-function**. (The function came first, so the group $\Phi(n)$ is named after the function.)

Euler ϕ-function

29.12. **Proposition.**

(1) *If $a_1, \dots, a_s \geq 1$ are pairwise relatively prime, then $\phi(a_1 \cdots a_s) = \phi(a_1) \cdots \phi(a_s)$.*
(2) *If p is prime, then $\phi(p^k) = p^k - p^{k-1} = p^{k-1}(p-1) = p^k(1 - \frac{1}{p})$.*

Proof. Statement (1) is immediate from the bijection $\Phi(a_1 \cdots a_s) \to \Phi(a_1) \times \cdots \times \Phi(a_s)$. Statement (2) is proved by direct counting: the elements of $\{1, 2, \dots, p^k\}$ which are *not* relatively prime to p^k are the ones which are multiples of p, i.e., $\{p, 2p, \dots, (p^{k-1} - 1)p, p^{k-1}p\}$. There are exactly p^{k-1} such elements. □

29.13. **Corollary.** *Let n be a natural number with factorization $n = p_1^{k_1} p_2^{k_2} \cdots p_s^{k_s}$ into distinct primes $p_1, p_2 \dots, p_s$. Then:*

$$\phi(n) = n\big(1 - \frac{1}{p_1}\big)\big(1 - \frac{1}{p_2}\big) \cdots \big(1 - \frac{1}{p_s}\big).$$

29.14. *Example.* We have $\phi(30) = 30 \cdot \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{4}{5} = 8$.

30. Recognizing direct products

How can we recognize that G is isomorphic to a product of groups? We can do this if we have a pair of subgroups with nice properties.

Lecture 19

30.1. **Lemma.** *Let G be a group with normal subgroups N and M, and suppose that $M \cap N = \{e\}$.*

(1) *Then $mn = nm$ for all $m \in M$ and $n \in N$.*

(2) *The set $MN = \{\, mn \mid m \in M, n \in N \,\}$ is a subgroup of G, and the function $M \times N \to MN$ given by $(m,n) \mapsto mn$ is an isomorphism of groups.*
(3) *If $G = MN$, then $G \approx M \times N$.*

Proof. For (1), consider $mnm^{-1}n^{-1}$. This element is in $(mNm^{-1})N = NN = N$ since N is normal, and is also in $M(nMn^{-1}) = MM = M$ since M is normal. Thus $mnm^{-1}n^{-1} \in M \cap N$, so by hypothesis $mnm^{-1}n^{-1} = e$.

For (2), that MN is subgroup is already a consequence of the diamond isomorphism theorem. Now check that the function $\phi(m,n) \mapsto mn$ is a homomorphism, a proof which uses (1):

$$\phi(m,n)\phi(m',n') = mnm'n' = mm'nn' = \phi(mm',nn').$$

Clearly this homomorphism surjects onto the subgroup MN. If $(m,n) \in \ker\phi$, then $mn = e$ and therefore $m = n^{-1} \in M \cap N = \{e\}$, whence $m = n = e$. So the kernel is trivial.

Claim (3) is immediate. □

Warning! Statement (1) in the lemma *does not* say that MN is a commutative group. It says that "an element from M commutes with an element from N"; it may still happen that two elements from M do not commute with each other.

When $G = MN$ for normal subgroups M, N with $M \cap N = \{e\}$, we say that G is an *internal direct product* of the subgroups M and N.

Note the case of $G := A \times B$, with

$$M := A \times \{e_B\} = \{\, (a, e_B) \mid a \in A \,\}, \qquad N := \{e_A\} \times B = \{\, (e_A, b) \mid b \in B \,\}.$$

There are theorems for recognizing internal products of more than two groups, but they are harder to use. Here is one.

30.2. **Proposition.** *Let $N_1, \dots, N_r$ be normal subgroups of G such that (i) $N_1 \cdots N_r = G$, and (ii) given $x_i \in N_i$ for $i = 1, \dots, r$ such that $x_1 \cdots x_r = e$, we must have $x_i = e$. Then $\phi\colon N_1 \times \cdots \times N_r \to G$ given by $(x_1, \dots, x_r) \mapsto x_1 \cdots x_r$ is an isomorphism of groups.*

Proof. A consequence of (ii) is that $N_i \cap N_j = \{e\}$ if $i \neq j$, and the same proof as before shows that $xy = yx$ for $x \in N_i$ and $y \in N_j$. It's now easy to check that ϕ is a homomorphism, which is clearly surjective, and has $\ker(\phi) = \{e\}$. □

31. Automorphism groups

An **automorphism** of a group G is an isomorphism $\phi\colon G \to G$ from the group to itself. **automorphism**

31.1. *Example.* Let $K = \{e, r_1, r_2, r_3\}$ be the symmetry group of a rectangle. Consider the function $\phi\colon K \to K$ defined by $\phi(e) = e$, $\phi(r_1) = r_2$, $\phi(r_2) = r_3$, and $\phi(r_3) = r_1$. This ϕ is an automorphism.

To check this: (i) clearly ϕ is a bijective function. (ii) We need to check $\phi(xy) = \phi(x)\phi(y)$ for all $x, y \in K$, so there are 16 things to check. But it's not so bad if you remember

- $xe = x = ex$ for any $x \in K$.
- $xx = e$ for any $x \in K$.
- The remaining kind of product is xy where $x, y \in \{r_1, r_2, r_3\}$ with $x \neq y$. In this case we always have $xy = z$ where z is the element of $\{r_1, r_2, r_3\}$ not equal to either x or y.

The collection $\mathrm{Aut}(G)$ of all automorphisms of G itself forms a group under composition of functions: $\phi \circ \psi$.

31.2. *Example.* If K is the group in the previous example, then $\mathrm{Aut}(K) \approx S_3$. Any permutation $\sigma \in S_3$ gives an automorphism defined by $\sigma(r_k) = r_{\sigma(k)}$.

31.3. *Example.* The automorphism group of $\mathbb{Z}_n$ is $\Phi(n)$. Proof: each $[a] \in \Phi(n)$ determines an automorphism $\phi\colon \mathbb{Z}_n \to \mathbb{Z}_n$ by the formula $\phi([x]) := [a][x] = [ax]$ (to show this, check that $\phi([x]+[y]) = \phi([x]) + \phi([y])$, and that ϕ is a bijection).

If $\phi\colon \mathbb{Z}_n \to \mathbb{Z}_n$ is any automorphism, then write $[a] := \phi([1])$. Check that we must have $\phi([x]) = [ax]$, so this automorphism is one of the ones given above.

31.4. *Example.* Let S_n be the symmetric group of an n-element set. Then $\mathrm{Aut}(S_n) \approx S_n$ for all n *except* $n = 2$ or $n = 6$. This is *not* an obvious fact.

Most of the basic facts about $\mathrm{Aut}(G)$ are developed in exercises. Here is a summary.

- For every $g \in G$, the function $c_g\colon G \to G$ defined by $c_g(x) := gxg^{-1}$ is an automorphism of G; this kind of automorphism is called an **inner** automorphism. **inner**
- The function $c\colon G \to \mathrm{Aut}(G)$ defined by $c\colon g \mapsto c_g$ is a homomorphism of groups. (That is, $c_{gh} = c_g \circ c_h$.)
- The kernel of $c\colon G \to \mathrm{Aut}(G)$ is the **center** of G, defined by **center**
$$Z(G) := \{\, a \in G \mid ag = ga \text{ for all } g \in G \,\}.$$
In particular, the center $Z(G) \leq G$ is a normal subgroup.
- We write $\mathrm{Inn}(G) := c(G) \leq \mathrm{Aut}(G)$ for the image of c, called the **group of innner automorphisms**. (Our textbook calls this subgroup "Int(G)". The factorization theorem tells us that $\mathrm{Inn}(G) \approx G/Z(G)$. **group of innner autom phisms**
- The subgroup $\mathrm{Inn}(G)$ is normal in $\mathrm{Aut}(G)$. The quotient group $\mathrm{Out}(G) := \mathrm{Aut}(G)/\mathrm{Inn}(G)$ is sometimes called the group of **outer automorphisms**. **outer automorphisms**

32. Semidirect products

Goodman 3.2. **Lecture 20**

It's possible to have groups that sort of look like products, but aren't really.

32.1. *Example.* Consider D_n, with $n \geq 3$. It has a subgroup $N = \langle r \rangle$ of order n, and a subgroup $A = \langle j \rangle$ of order 2. Every element of D_n can be written uniquely as a product xy where $x \in N$ and $y \in A$ (i.e., as $r^k e$ or $r^k j$).

Furthermore, N is normal and $D_n/N \approx A$.

However, D_n is *not* a product. The subgroup A is *not* normal. The map $N \times A \to D_n$ given by $(x, y) \mapsto xy$ is *not* a homomorphism. The elements of N do not commute with elements from A.

In addition to identifying the subgroups N and A, we know the group once we also know the "commutation relation" $jr = r^{-1}j$. Note that the function
$$\gamma\colon A \to \mathrm{Aut}(N), \qquad x \mapsto (n \mapsto xnx^{-1})$$
is a homomorphism.

Here is the general picture. Suppose we have a group G with two subgroups $A, N \leq G$, such that $A \cap N = \{e\}$ and $G = AN$. We have shown that if both A and N are normal, then G is a product group, i.e., there is an isomorphism $A \times N \to G$.

Suppose we only know that N is normal. Then we may not have a product group. However, we do get a homomorphism $\gamma\colon A \to \mathrm{Aut}(N)$ defined by $\gamma(x)(n) := xnx^{-1}$. It turns out that we can reconstruct G from the three pieces of data: A, N, and $\gamma\colon A \to \mathrm{Aut}(N)$. That is, G is isomorphic to something called the *semi-direct product of A and N by γ*.

32.2. **Definition of semi-direct product.** Consider groups N and A, and a homomorphism $\gamma\colon a \mapsto \gamma_a\colon A \to \mathrm{Aut}(N)$. Note this means γ satisfies:
$$\gamma_a(n_1 n_2) = \gamma_a(n_1)\gamma_a(n_2) \quad \text{for all } a \in A,\ n_1, n_2 \in N,$$
$$\gamma_{a_1}(\gamma_{a_2}(n)) = \gamma_{a_1 a_2}(n) \quad \text{for all } a_1, a_2 \in A,\ n \in N.$$

The first line means "each function $\gamma_a\colon N \to N$ is a homomorphism", and the second line means "$\gamma_{a_1} \circ \gamma_{a_2} = \gamma_{a_1 a_2}$".

We define a new group, called the **semidirect product** of N and A by γ, as follows. **semidirect product**

$$\begin{aligned} N \rtimes_\gamma A &:= \{\, (n,a) \mid n \in N, a \in A \,\}, \\ (n_1,a_1)(n_2,a_2) &:= (n_1\, \gamma_{a_1}(n_2), a_1 a_2). \end{aligned}$$

Verify this is a group.

- *Associativity.* We have
$$\begin{aligned} \big[(n_1,a_1)(n_2,a_2)\big](n_3,a_3) &= (n_1\,\gamma_{a_1}(n_2), a_1a_2)(n_3,a_3) \\ &= (n_1\, \underline{\gamma_{a_1}(n_2)\,\gamma_{a_1a_2}(n_3)}, (a_1a_2)a_3), \end{aligned}$$
and
$$\begin{aligned} (n_1,a_1)\big[(n_2,a_2)(n_3,a_3)\big] &= (n_1,a_1)(n_2\,\gamma_{a_2}(n_3), a_2a_3) \\ &= (n_1\, \underline{\gamma_a\big(n_2\,\gamma_{a_2}(n_3)\big)}, a_1(a_2a_3)). \end{aligned}$$
The reason these are equal is that $\gamma\colon A \to \operatorname{Aut}(N)$ is a homomorphism, and that elements $\phi \in \operatorname{Aut}(N)$ such as γ_a are homomorphisms. Thus
$$\begin{aligned} \gamma_{a_1}\big(n_2\,\gamma_{a_2}(n_3)\big) &= \gamma_{a_1}(n_2)\,\gamma_{a_1}\big(\gamma_{a_2}(n_3)\big) && (\gamma_{a_1} \text{ is a homomorphism}), \\ &= \gamma_{a_1}(n_2)\,\gamma_{a_1a_2}(n_3), && (\gamma \text{ is a homomorphism.}) \end{aligned}$$
The second one is saying that $\gamma_{a_1} \circ \gamma_{a_2} = \gamma_{a_1a_2}$.
- *Identity.* The element $e := (e_N, e_A)$ is an identity element.
- *Inverse.* $(n,a)^{-1} = (\gamma_a(n^{-1}), a^{-1})$.

Finish proof that this defines a group law

32.3. *Example.* Suppose $\gamma\colon A \to \operatorname{Aut}(N)$ is the **trivial homomorphism**, which sends every $a \in A$ to the identity map of N. In other words, **trivial homomorphism**

$$\gamma_a(n) = n, \qquad \text{for all } a \in A,\ n \in N.$$

Then the product rule simplifies to $(n_1,a_1)(n_2,a_2) = (n_1n_2, a_1a_2)$, and we are just looking at the product group $N \times A$.

32.4. *Example.* Let $N = \mathbb{Z}_n$ and recall that $\operatorname{Aut}(\mathbb{Z}_n) = \Phi(n)$.

Let $A = \mathbb{Z}_2$ and let $\gamma\colon \mathbb{Z}_2 \to \operatorname{Aut}(\mathbb{Z}_n) = \Phi(n)$ be the function

$$[0]_2 \mapsto [1]_n, \qquad [1]_2 \mapsto [-1]_n.$$

You could just write this as $[k]_2 \mapsto [(-1)^k]_n$. Then γ is a homomorphism. The resulting semidirect product group $N \rtimes_\gamma A$ is isomorphic to D_n.

To see this, it is easier to write everything with multiplicative group laws. So identify

$$\mathbb{Z}_n \approx \langle r \rangle = \{e, r, \dots, r^{n-1}\}, \qquad \mathbb{Z}_2 \approx \langle a \rangle = \{e, a\},$$

i.e., $r^i \leftrightarrow [i]_n$ and $a^j \leftrightarrow [j]_2$. Then the rule for $\gamma\colon A \to \operatorname{Aut}(N)$ is

$$\gamma_e(r^k) = r^k, \qquad \gamma_a(r^k) = r^{-k}.$$

If we write $\underline{r} := (r,e)$ and $\underline{a} := (e,a)$, we see that

$$\underline{a}\underline{r} = \underline{r}^{-1}\underline{a}, \qquad \underline{r}^n = e, \qquad \underline{a}^2 = e.$$

Note that in a general semidirect product $N \rtimes_\gamma A$ we can identify A with the subgroup $\{\,(e_N, a) \mid a \in A\,\}$, and N with the subgroup $\{\,(n, e_A) \mid n \in N\,\}$. Sometimes we do this, so that every element in $G = N \rtimes_\gamma A$ can be written uniquely as na, with $n \in N \leq G$ and $a \in A \leq G$. The product rule implies

$$an = \gamma_a(n)a, \qquad \text{or equivalently} \qquad ana^{-1} = \gamma_a(n).$$

32.5. *Example.* The group $\mathbb{Z}_7$ admits an automorphism defined by $\phi([x]) = [2x]$. You can check this has order 3. Use this to define $\gamma\colon \mathbb{Z}_3 \to \operatorname{Aut}(\mathbb{Z}_7)$, and thus $\mathbb{Z}_7 \rtimes_\gamma \mathbb{Z}_3$. If $a =$ generator of $\mathbb{Z}_7$, and $b =$ generator of $\mathbb{Z}_3$, then we have

$$a^7 = e = b^3, \qquad bab^{-1} = a^2.$$

Thus, we have constructed a new non-abelian group of order 21.

33. Recognizing semi-direct products

We now have a theorem for recognizing semidirect products. The idea is that the "abstract" semidirect product $G = N \rtimes_\gamma A$ has "coordinate axis subgroups"

$$N' := N \times \{e_A\}, \qquad A' := \{e_N\} \times A,$$

which are isomorphic to N and A. Furthermore, we have $G = N'A'$ and $N' \cap A' = \{(e, e)\}$, and N' is normal in G.

33.1. **Theorem.** *Let G be a group with*

- *subgroup $A \leq G$,*
- *normal subgroup $N \trianglelefteq G$, such that*
- *$G = NA$, and*
- *$A \cap N = \{e\}$.*

Then there exists a homomorphism $\gamma\colon A \to \operatorname{Aut}(N)$ such that there is an isomorphism of groups

$$N \rtimes_\gamma A \approx G.$$

Proof. Each element $g \in G$ gives a conjugation homomorphism $\operatorname{cnj}_g\colon G \to G$. Since N is normal, we have that $\operatorname{cnj}_g(N) = N$, so it restricts to an isomorphism

$$\operatorname{cnj}_g|N \in \operatorname{Aut}(N).$$

So for $a \in A$ and $n \in N$, define

$$\gamma_a(n) := \operatorname{cnj}_a(n) = ana^{-1}.$$

We have that $\gamma_a(n) \in N$ since N is normal, and that $\gamma_a\colon N \to N$ is an isomorphism (its inverse map is $\gamma_{a^{-1}}$). Thus we can define a semidirect product group $N \rtimes_\gamma A$.

Now consider the function

$$\phi\colon N \rtimes_\gamma A \to G$$

defined by

$$\phi(n, a) := na.$$

Check that ϕ is a homomorphism:

$$\begin{aligned} \phi((n_1, a_1)(n_2, a_2)) &= \phi(n_1\gamma_{a_1}(n_2), a_1a_2) \\ &= n_1(a_1n_2a_1)^{-1}a_1a_2 \\ &= n_1a_1n_2a_2, \\ \phi((n_1, a_1))\phi((n_2, a_2)) &= (n_1a_1)(n_2a_2) \end{aligned}$$

give the same answer in each case.

Finally, ϕ is surjective because $G = NA$, while ϕ is injective because $\ker \phi = \{e\}$, as shown by

$$\phi((n,a)) = e \;\Rightarrow\; na = e \;\Rightarrow\; n = a^{-1},$$

whence $n = a^{-1} \in N \cap A$ so $n = a^{-1} = e$, so $(n,a) = (e,e)$. □

34. Examples of semi-direct products

34.1. *Example.* Suppose $\gamma\colon A \to \mathrm{Aut}(N)$ is the **trivial homomorphism**, which sends every $a \in A$ to the identity map of N. In other words, **trivial homomorphism**

$$\gamma_a(n) = n, \qquad \text{for all } a \in A,\, n \in N.$$

Then the product rule simplifies to $(n_1, a_1)(n_2, a_2) = (n_1 n_2, a_1 a_2)$, and we are just looking at the product group $N \times A$.

34.2. *Example* (The affine group). The group $\mathrm{Aff}(n)$ is the set of all transformations of $\mathbb{R}^n$ of the form $f(x) := Ax + b$, where $A \in GL(n, \mathbb{R})$ and $b \in \mathbb{R}^n$. (Verify that this is a group.)

The subgroup $N = \{\, T_b \mid b \in \mathbb{R}^n \,\}$ where $T_b(x) = b + x$ is a normal subgroup of $\mathrm{Aff}(n)$ (isomorphic to $\mathbb{R}^n$ with $+$). The subgroup $A = \{\, A \mid A \in GL(n, \mathbb{R}) \,\}$ is not normal. We have that $\mathrm{Aff}(n)/N \approx A$, with "commutation relation" $AT_b = T_{Ab}A$. Note that the function

$$\gamma\colon A \to \mathrm{Aut}(N), \qquad A \mapsto (T_b \mapsto AT_bA^{-1} = T_{Ab})$$

is a homomorphism.

34.3. *Example.* The group $\mathbb{Z}_7$ admits an automorphism defined by $\phi([x]) = [2x]$. You can check this has order 3. Use this to define $\gamma\colon \mathbb{Z}_3 \to \mathrm{Aut}(\mathbb{Z}_7)$, and thus $G = \mathbb{Z}_7 \rtimes_\gamma \mathbb{Z}_3$. We have constructed new non-abelian group of order 21.

Since the group is non-abelian, it is convenient to rewrite things multiplicatively. If $a =$ generator $[1]_7$ of $\mathbb{Z}_7$, and $b =$ generator $[1]_3$ of $\mathbb{Z}_3$, and we write the group law on G multiplicatively then:

- Every element of G can be written uniquely as $a^i b^j$ with $i \in \{0, 1, \ldots, 6\}$ and $j \in \{0, 1, 2\}$.
- We have identities

$$a^7 = e = b^3, \qquad ba = a^2 b.$$

Using these you can compute any product in G.

34.4. *Example* (Infinite dihedral group). Let $N = \mathbb{Z}$ and note that $\mathrm{Aut}(\mathbb{Z}) \approx \{\pm 1\}$. (The functions $x \mapsto \pm x$ are both automorphisms of $\mathbb{Z}$.) **Lecture 21**

Let $A = \mathbb{Z}_2$ and $\gamma\colon \mathrm{Aut}(\mathbb{Z}) = \{\pm 1\}$ by

$$[k]_2 \mapsto (-1)^k.$$

The resulting semidirect product $G = \mathbb{Z} \rtimes_\gamma \mathbb{Z}_2$ is called the **infinite dihedral group**, and sometimes denoted D_∞. **infinite dihedral group**

It is conventional to set

$$r := (1, [0]), \quad j := (0, [1]) \quad \in \mathbb{Z} \rtimes_\gamma \mathbb{Z}_2,$$

and to write the group law multiplicatively.

34.5. *Exercise.* Show that in D_∞ (written multiplicatively), we have identities

$$j^2 = e = jr = r^{-1} j.$$

Show that every element of D_∞ can be written exactly one way in the forms

$$r^k, \quad k \in \mathbb{Z} \qquad \text{or} \qquad r^k j, \quad k \in \mathbb{Z}.$$

34.6. *Exercise.* Show that every finite dihedral group is isomorphic to a quotient group of D_∞.

34.7. *Exercise.* Let $N = \mathbb{Z}_2 \times \mathbb{Z}_2$, and recall that $\operatorname{Aut}(N) \approx S_3$. Let $A := \operatorname{Aut}(N)$, and consider

$$G := N \rtimes_{\mathrm{id}} A, \qquad \mathrm{id}\colon A \to \operatorname{Aut}(N).$$

Show that G is isomorphic to S_4.

Given a group G with a normal subgroup N, it is not necessarily the case that G is a semi-direct product: there might be no suitable subgroup $A \leq G$ with $NA = G$ and $N \cap A = \{e\}$.

34.8. *Example.* The quaternion group $Q = \{\pm 1, \pm i, \pm j, \pm k\}$ contains a normal subgroup $N = \langle i \rangle$ of order 4. But it is not a semi-direct product involving N. Although there is one subgroup $\langle -1 \rangle$ of order 2, it does not have trivial intersection with N.

34.9. *Example.* Let K be the Klein 4-group (isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$), and recall that $\operatorname{Aut}(K) \approx S_3$. Thus we can form a semidirect product $G = K \rtimes_\gamma S_3$ where $\gamma\colon S_3 \to \operatorname{Aut}(K)$ is the isomorphism. This gives a group of order 24; it turns out to be isomorphic to S_4.

In fact, we have seen these subgroups of S_4:

$$K' = \{e, (12)(34), (13)(24), (14)(23)\} \leq S_4, \qquad A' = \{e, (123), (132), (12), (13), (23)\} \leq S_4.$$

35. Classification of finite abelian groups: Theorems

These are stated in Goodman 3.6. I'll state the theorems now, but prove them later.

As we've seen, it is difficult to classify all finite groups. However, there is an excellent classification theorem for finite *abelian* groups, up to isomorphism.

The key fact (which is not so easy to prove) is the following:

Every finite abelian group is isomorphic to a direct product of finitely many cyclic groups.

However, this is not a classification, because a finite abelian group can be isomorphic to a cyclic group in more than one way, using that $\mathbb{Z}_{ab} \approx \mathbb{Z}_a \times \mathbb{Z}_b$ when $\gcd(a, b) = 1$, e.g.,

$$\mathbb{Z}_{60} \approx \mathbb{Z}_{20} \times \mathbb{Z}_3 \approx \mathbb{Z}_{12} \times \mathbb{Z}_5 \approx \mathbb{Z}_4 \times \mathbb{Z}_3 \times \mathbb{Z}_5.$$

One form of the classification is as follows.

35.1. **Theorem** (Classification of finite abelian groups (elementary divisor form))**.** *Every finite abelian group G is isomorphic to a finite direct product of cyclic groups of prime power order:*

$$G \approx \mathbb{Z}_{p_1^{r_1}} \times \cdots \times \mathbb{Z}_{p_k^{r_k}},$$

where $k \geq 0$, $p_1, \dots, p_k$ are primes, and each $r_j \geq 1$. The "elementary divisors" $p_1^{r_1}, \dots, p_k^{r_k}$ which appear here are unique, up to reordering. (They can be repeated, however.)

35.2. *Example.* Every abelian group of order $24 = 2^2 \times 3$ is isomorphic to exactly one in the following list in elementary divisor form:

$$\mathbb{Z}_{2^3} \times \mathbb{Z}_3, \qquad \mathbb{Z}_{2^2} \times \mathbb{Z}_2 \times \mathbb{Z}_3, \qquad \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3.$$

Some remarks.

- When $k = 0$, we understand this as the case $G = \{e\}$.
- We are allowed to repeat numbers in the list $p_1^{r_1}, \dots, p_k^{r_k}$ of elementary divisors.
- Every finite abelian groups is a product on cyclic groups. However, this does not mean that it is a product of cyclic groups in only one way, even considering reordering, because of the theorem
$$\mathbb{Z}_m \times \mathbb{Z}_n \approx \mathbb{Z}_{mn} \qquad \text{if } m, n \text{ are relatively prime.}$$

35.3. *Example.* We have $\mathbb{Z}_{2^2} \times \mathbb{Z}_2 \times \mathbb{Z}_3 \approx \mathbb{Z}_{12} \times \mathbb{Z}_2 \approx \mathbb{Z}_6 \times \mathbb{Z}_4$.

There can be lots of ways to group the factors in the decomposition into bigger cyclic groups. One way to do this leads to another form of the classification, which is what we will prove first.

35.4. **Theorem** (Classification of finite abelian groups (invariant factor form))**.** *Every finite abelian group G is isomorphic to exactly one of the form*

$$G \approx \mathbb{Z}_{a_1} \times \cdots \times \mathbb{Z}_{a_s},$$

where $s \geq 0$, all $a_i \geq 2$, and a_i divides a_{i+1} for $1 \leq i \leq s-1$.

35.5. *Example.* Every abelian group of order 24 is isomorphic to exactly one in the following list.

$$\mathbb{Z}_{24}, \qquad \mathbb{Z}_2 \times \mathbb{Z}_{12}, \qquad \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_6.$$

Sketch proof of elementary divisor form from invariant factor form. We assume the theorem about existence and uniqueness of invariant factor decomposition, and use it to prove the one about elementary divisor decomposition

Let G be a finite abelian group, with a particular invariant factor decomposition. We proceed by reorganizing the factors in a particular way, which is best demonstrated by example. For instance, suppose the invariant factors of G are

$$5, \quad 25, \quad 50, \quad 36000.$$

Factor each of these:

$$5 = 5^1, \quad 25 = 5^2, \quad 50 = 2^1 \times 5^2, \quad 36000 = 2^5 \times 3^2 \times 5^2.$$

It is convenient to put this in a table, so the invariant factors are listed on the left, and the later columns show the prime-power factors of each, grouped according to prime:

5			5^1
25			5^2
50	2^1		5^2
36000	2^5	3^2	5^3

By the Chinese remainder theorem, we get that

$$G \approx \mathbb{Z}_5 \times \mathbb{Z}_{25} \times \mathbb{Z}_{50} \times \mathbb{Z}_{36000} \approx \mathbb{Z}_{5^1} \times \mathbb{Z}_{5^2} \times \mathbb{Z}_2 \times \mathbb{Z}_{5^2} \times \mathbb{Z}_{2^5} \times \mathbb{Z}_{3^2} \times \mathbb{Z}_{5^2},$$

an elementary divisor decomposition of G.

You can go backwards. Given the list of elementary divisors

$$2^5, \quad 2^1, \quad 3^2, \quad 5^3, \quad 5^2, \quad 5^2, \quad 5^1,$$

stack them in columns so that each column corresponds to a different prime, and so the terms in each column decrease as you go up: Again, the Chinese remainder theorem gives an isomorphism between G and the product of cyclic groups of orders $5, 25, 50, 36000$.

Thus, invariant factors can be recovered from elementary divisors, and vice versa. In fact, what I have just illustrated gives a bijective correspondence between

> *Invariant factors lists:* The set of sequences $a_1, \dots, a_s$ of integers with $s \geq 0$, $a_i \geq 2$, and $a_i \mid a_{i+1}$,

and

> *Elementary divisor lists:* The set of sequences of integers *up to reordering* of the form $p_1^{r_1}, \dots, p_k^{r_k}$, where $k \geq 0$, $p_1, \dots, p_k$ are primes, and $r_i \geq 1$.

Furthermore, if G has an invariant factor decomposition, then G also has an elementary divisor decomposition using the corresponding elementary divisor list.

So if G had two distinct invariant factor decompositions, it would also have two distinct elementary divisor decompositions, which contradicts the uniqueness of elementary divisors. This shows that invariant factors are also unique, assuming that elementary divisors are unique. □

36. Uniqueness of elementary divisors

Now I will show uniqueness of elementary divisor decompositions, which as we have seen also implies uniqueness of invariant factor decompositions. **Lecture 22**

Suppose G is a group. For each $m \geq 1$, define a subset of $G[m]$ of G by

$$G[m] := \{\, a \in G \mid a^m = e \,\}.$$

This is the subset of elements whose mth power is the idenity. Note: this is not the same as the subset of elements of order m, but it is related to it.

36.1. *Exercise.* Show that $G[m] = \{\, a \in G \mid o(a) \mid m \,\}$.

Note that $G[m]$ might not be a subgroup. For instance, $S_3[2] = \{e, (1\ 2), (2\ 3), (1\ 3)\}$, which is not a subgroup of S_3. However, it does turn out to be a subgroup when G is abelian.

36.2. *Exercise.* Show that if G is *abelian*, then $G[m]$ is a subgroup of G.

For a finite group G and any $m \geq 1$, define

$$\alpha_m(G) := |G[m]| \in \mathbb{N},$$

the size of the set $G[m]$.

36.3. **Proposition.** *The function α_m is an isomorphism invariant of finite groups. That is, if G, H are finite groups and $G \simeq H$, then $\alpha_m(G) = \alpha_m(H)$.*

Proof. Let $\phi\colon G \to H$ be an isomorphism. Then I claim that ϕ restricts to a bijection $\phi'\colon G[m] \to H[m]$, which implies that $\alpha_m(G) = \alpha_m(H)$.

First note that if $a \in G[m]$, then $\phi(a) \in H[m]$, since $\phi(a)^m = \phi(a^m) = \phi(e) = e$. Thus we get a function $\phi'\colon G[m] \to H[m]$. We can do the same with the inverse isomorphism $\phi^{-1}\colon H \to G$, which will give us an inverse function to ϕ'. □

It is also easy to compute α_m for a direct product.

36.4. **Proposition.** *Let $G_1, \dots, G_k$ be finite groups, and let $G := G_1 \times \cdots \times G_k$ be the direct product group. Then*

$$\alpha_m(G) = \alpha_m(G_1) \cdots \alpha_m(G_k).$$

Proof. I claim that for any $g := (a_1, \dots, a_k) \in G$, we have that $g \in G[m]$ if and only if $a_i \in G_i[m]$ for all $i = 1, \dots, k$. This is straightforward, since $g^m = (a_1^m, \dots, a_k^m)$. □

Finally, let's compute $\alpha_m(G)$ when G is a finite cycilc group.

36.5. *Example.* Let $G = \mathbb{Z}_8$. We have

$$\alpha_1(G) = 1, \quad \alpha_2(G) = 2, \quad \alpha_4(G) = 4, \quad \alpha_8(G) = 8, \quad \alpha_{16}(G) = 8, \quad \alpha_{32}(G) = 8, \quad \dots$$

Also, if m is odd, then $\alpha_m(G) = 1$.

36.6. **Proposition.** *Let $G = \mathbb{Z}_n$. Then $\alpha_m(\mathbb{Z}_n) = \gcd(m, n)$.*

Proof. Let $d := \gcd(m, n)$. *Step 1.* Show that if G is any group of order n, then $G[m] = G[d]$. Since $d \mid m$, we have that $g^d = e$ implies $g^m = e$, so $G[d] \subseteq G[m]$.

Conversely, suppose $g^m = e$. By Lagrange we know $g^n = e$, and since d is an integer combination of m and n we have $g^d = e$. Thus $G[m] \subseteq G[d]$.

Step 2. We describe $G[d]$ when $G = \mathbb{Z}_n$ and d divides n. This is the set of congruence classes $[k]_n$ such that $[dk]_n = [0]_n$, i.e., such that $n \mid dk$. There are exactly d integers $k \in \{1, 2, \dots, n\}$ with this property, namely $e, 2e, \dots, de = n$ where $e = n/d$. □

We will only need this when n and m are prime powers.

36.7. **Corollary.** *Let $G = \mathbb{Z}_{p^r}$ where p is prime. Then for any $j, r \geq 0$ we have*

$$\alpha_{p^j}(\mathbb{Z}_{p^r}) = \min(p^j,\, p^r) = \begin{cases} p^j & \text{if } j \leq r, \\ p^r & \text{if } i \geq r. \end{cases}$$

If q is a prime such that $q \neq p$, then for any $j \geq 0$ we have

$$\alpha_{q^j}(\mathbb{Z}_{p^r}) = 1.$$

Notice that we can now compute α_{q^j} on any elementary divisor decomposition $G = \mathbb{Z}_{p_1^{r_1}} \times \cdots \times \mathbb{Z}_{p_k^{r_k}}$. Note that for such groups, $\alpha_{q^j}(G)$ will always be an integer, and in fact a power of q. Thus it makes sense to think about $\log_q \alpha_{q^j}(G)$.

Now I will use this to define a new function β_{q^j} for q prime and $j \geq 1$, which will have similar properties to α_{q^j}, but is designed to be easy to compute on a product of cyclic groups of prime power order. The formula is

$$\begin{aligned} \beta_{q^j}(G) &:= \log_q \left[\frac{\alpha_{q^j}(G)^2}{\alpha_{q^{j-1}}(G)\, \alpha_{q^{j+1}}(G)} \right] \\ &= 2 \log_q \alpha_{q^j}(G) - \log_q \alpha_{q^{j-1}}(G) - \log_q \alpha_{q^{j+1}}(G) \\ &= \left[\log_q \alpha_{q^j}(G) - \log_q \alpha_{q^{j-1}}(G)\right] - \left[\log_q \alpha_{q^{j+1}}(G) - \log_q \alpha_{q^j}(G)\right]. \end{aligned}$$

36.8. *Example.* Let $G = \mathbb{Z}_{16}$. Then if q is an odd prime, we have $\alpha_{q^j}(G) = 1$ for any j, so $\beta_{q^j}(G) = 0$. Here are the values of $\alpha_{2^j}(G)$.

j	0	1	2	3	4	5	6
2^j	1	2	4	8	16	32	64
$\alpha_{2^j}(G)$	1	2	4	8	16	16	16
$\log_2 \alpha_{2^j}(G)$	0	1	2	3	4	4	4
$\beta_{2^j}(G)$		0	0	0	1	0	0

36.9. *Example.* Let $H = \mathbb{Z}_2 \times \mathbb{Z}_8$. Then if q is an odd prime, we have $\alpha_{q^j}(H) = 1$ for any j, so $\beta_{q^j}(H) = 0$. Here are the values of $\alpha_{2^j}(H)$.

j	0	1	2	3	4	5	6
2^j	1	2	4	8	16	32	64
$\alpha_{2^j}(H)$	1	4	8	16	16	16	16
$\log_2 \alpha_{2^j}(H)$	0	2	3	4	4	4	4
$\beta_{2^j}(H)$		1	0	1	0	0	0

36.10. **Proposition.**

1. *If G and H are finite groups and $G \simeq H$, then $\beta_{q^j}(G) = \beta_{q^j}(H)$.*
2. *If $G = G_1 \times \cdots \times G_k$, then $\beta_{q^j}(G) = \sum_{i=1}^k \beta_{q^j}(G_i)$.*
3. *If $G = \mathbb{Z}_{p^r}$, $r \geq 0$ and p prime, then*

$$\beta_{p^j}(\mathbb{Z}_{p^r}) = \begin{cases} 1 & \text{if } j = r, \\ 0 & \text{if } j \neq r, \end{cases}$$

and

$$\beta_{q^j}(\mathbb{Z}_{p^r}) = 0 \qquad \text{if } q \neq p.$$

36.11. **Corollary.** *If $G \simeq \mathbb{Z}_{p_1^{r_1}} \times \cdots \times \mathbb{Z}_{p_k^{r_k}}$ where $p_1, \ldots, p_k$ are prime and $r_1, \ldots, r_k \geq 1$, then*

$$\beta_{q^j}(G) = \text{the number of times } q^j \text{ appears in the list } p_1^{r_1}, \ldots, p_k^{r_k}.$$

36.12. **Corollary.** *Any two elementary divisor decompositions of G have the same list of elementary divisors, up to reordering.*

Proof. The previous Corollary, and the fact that β_{q^j} is an isomorphism invariant. □

37. Finitely generated abelian groups

A group G is **finitely generated** if there *exists* a finite subset $S \subseteq G$ such that $\langle S\rangle = G$.

Lecture 23
finitely generated

37.1. *Example.* The group $\mathbb{Z}$ is finitely generated. One example of a finite generating set is $\{1\}$. (There are others though: $\{4, 9\}$ is also a finite generating set, for instance.)

37.2. *Example.* Every *finite* group is automatically *finitely generated*, since $\langle G\rangle = G$. (This is usually not the smallest generating set of course.)

37.3. *Example.* The group $\mathbb{Q}$ under addition is not finitely generated. Proof: show that any finitely generated subgroup of $\mathbb{Q}$ is contained in a cyclic subgroup, by "clearing denominators".

37.4. *Exercise.* If $G \simeq H$ and G is finitely generated, then H is finitely generated.

37.5. *Exercise.* If G and H are finitely generated, so is $G \times H$.

We are going to show the following, which will imply existence of invariant factor decomposition for finite abelian groups.

37.6. **Proposition.** *Every finitely generated abelian group G is isomorphic to a direct product of the form*

$$G \simeq \mathbb{Z}/\mathbb{Z}d_1 \times \mathbb{Z}/\mathbb{Z}d_2 \times \cdots \times \mathbb{Z}/\mathbb{Z}d_s,$$

for some sequence $d_1, \dots, d_s$ of integers such that (i) each $d_i \geq 0$ and (ii) we have $d_i \mid d_{i+1}$ for all i.

For instance, we might find that for a given G the list of d_is is: 1,3,12,12,720,0,0, so that

$$G \simeq\ \mathbb{Z}/\mathbb{Z}1\ \times\ \mathbb{Z}/\mathbb{Z}3\ \times\ \mathbb{Z}/\mathbb{Z}12\ \times\ \mathbb{Z}/\mathbb{Z}12\ \times\ \mathbb{Z}/\mathbb{Z}720\ \times\ \mathbb{Z}/\mathbb{Z}0\ \times\ \mathbb{Z}/\mathbb{Z}0,$$

since $1 \mid 3 \mid 12 \mid 12 \mid 720 \mid 0 \mid 0$. Notice that if $d = 1$ then $\mathbb{Z}/\mathbb{Z}1 = \mathbb{Z}_1$ is a trivial group, so we can remove it from the direct product if we want. Likewise, $\mathbb{Z}/\mathbb{Z}0 \simeq \mathbb{Z}$, and $\mathbb{Z}/\mathbb{Z}d = \mathbb{Z}_d$ if $d \geq 2$, so we can also write this as

$$G \simeq\ \mathbb{Z}_3 \times \mathbb{Z}_{12} \times \mathbb{Z}_{12} \times \mathbb{Z}_{720} \times \mathbb{Z} \times \mathbb{Z}.$$

Note that because of the $\mathbb{Z}$ factors this group is not finite. When G is finite we cannot have a $d_i = 0$, so in that case we must get an invariant factor decomposition.

Actually, these ideas lead to a complete classification of finitely generated abelian groups.

37.7. **Theorem.** *Every finitely generated abelian group G is isomorphic to a direct product of the form $G \approx F \times \mathbb{Z}^k$, where $k \geq 0$ and F is a finite abelian group. The integer k is unique, and the finite group F is unique up to isomorphism.*

You can combine this with either of the statements of the classification of finite abelian groups to get invariant factor and elementary divisor forms.

Let me state the invariant factor form, since it's what will actually prove.

37.8. **Theorem** (Goodman 3.6.2)**.** *If G is a finitely generated abelian group, then*

$$G \approx \mathbb{Z}_{a_1} \times \cdots \times \mathbb{Z}_{a_s} \times \mathbb{Z}^k,$$

where $s, k \geq 0$, each $a_i \geq 2$ and $a_i | a_j$ for $i \leq j$.

Furthermore, this decomposition is unique, in the sense that if $G \approx \mathbb{Z}_{b_1} \times \cdots \times \mathbb{Z}_{b_t} \times \mathbb{Z}^\ell$, with $\ell, t \geq 0$, $b_i \geq 2$, and $b_i \mid b_j$ if $i \leq j$, then $k = \ell$, $s = t$, and $a_i = b_i$.

Here is a sketch of the ideas of the proof of the proposition.

1. s If G is finitely generated and abelian, there exists a surjective homomorphism $\pi\colon \mathbb{Z}^m \to G$ for some $m \geq 0$, and thus an isomorphism $G \simeq \mathbb{Z}^m/N$ for some subgroup $N \leq \mathbb{Z}^m$.
2. Every subgroup N of $\mathbb{Z}^m$ is also finitely generated and abelian. Thus we can find a homomorphism $\alpha\colon \mathbb{Z}^n \to \mathbb{Z}^m$ for some $n \geq 0$ such that $\alpha(\mathbb{Z}^n) = N$. Thus $G \simeq \mathbb{Z}^m/\alpha(\mathbb{Z}^n)$ for some homomorphism α.

(3) We can represent elements of $\mathbb{Z}^m$ as column vectors with entries of $\mathbb{Z}$, and represent every homomorphism $\alpha\colon \mathbb{Z}^n \to \mathbb{Z}^m$ as left-multiplication by a matrix $A \in \mathrm{Mat}_{m\times n}(\mathbb{Z})$ with integer entries. Thus $G \simeq \mathbb{Z}^m/A\mathbb{Z}^n$ for some integer matrix A.

(4) A square matrix P is *integer invertible* if both P and P^{-1} have integer entries. Given $P \in \mathrm{Mat}_{m\times m}(\mathbb{Z})$ and $Q \in \mathrm{Mat}_{n\times n}(\mathbb{Z})$ which are integer invertible, form $B := PAQ$. In this case there is an isomorphism $\mathbb{Z}^m/A\mathbb{Z}^n \simeq \mathbb{Z}^m/B\mathbb{Z}^n$.

(5) A *Smith matrix* is an integer matrix $D \in \mathrm{Mat}_{m\times n}(\mathbb{Z})$ of the form $D = \mathrm{diag}(d_1, \dots, d_s)$ (with $s = \min(m, n)$) where each $d_i \in \mathbb{Z}_{\geq 0}$ and $d_i \mid d_{i+1}$. If D is this Smith matrix, then there is an isomorphism
$$\mathbb{Z}^m/D\mathbb{Z}^n \approx \mathbb{Z}/\mathbb{Z}d_1 \times \cdots \times \mathbb{Z}/\mathbb{Z}d_s \times \mathbb{Z}^{m-s}.$$

(6) *Smith normal form:* For any $A \in \mathrm{Mat}_{m\times n}(\mathbb{Z})$ there are integer invertible matrices P and Q such that $D = PAQ$ is a Smith matrix. Thus
$$G \simeq \mathbb{Z}^m/A\mathbb{Z}^n \simeq \mathbb{Z}^m/D\mathbb{Z}^n \simeq \mathbb{Z}/\mathbb{Z}d_1 \times \cdots \times \mathbb{Z}/\mathbb{Z}d_s \times \mathbb{Z}^{m-s}.$$

Thus, we are going to end up thinking about "linear algebra over $\mathbb{Z}$".

38. Linear algebra over $\mathbb{Z}$

Section 3.5.

We try to apply the concepts of basic linear algebra for abelian groups, using $\mathbb{Z}$ in place of the scalar field F. We assume all groups G are abelian, and for this discussion we will always write the group law for such G using *addition*.

Here is a table translating some concepts between multiplicative and additive notation: here $a, b \in G$ and $n \in \mathbb{Z}$.

	multiplicative	additive
group law	ab	$a+b$
identity element	e	0
inverse	a^{-1}	$-a$
nth power	a^n	na
left coset	aH	$a+H$
$a_i \in G,\ c_i \in \mathbb{Z}$	$a_1^{c_1}a_2^{c_2}\cdots a_k^{c_k}$	$c_1a_1 + c_2a_2 + \cdots + c_ka_k$

In other words, abelian groups written with addition have "scalar multiplication" by integers. Thus, such abelian groups are "like vector spaces", except that scalars are integers, rather than elements of a field.

39. Finitely generated free abelian groups

We will assume G is an abelian group, written with addition.

Say a subset $S \subseteq G$ is **linearly independent** if for every list of pairwise distinct elements $x_1, \dots, x_n$ of S, if $c_1x_1 + \cdots + c_nx_n = 0$ for some $c_i \in \mathbb{Z}$, then all $c_i = 0$. **linearly independent**

39.1. *Exercise.* If $S = \{x\}$ has one element, then S is linearly independent if and only if $o(x) = \infty$.

A subset $S \subseteq G$ is a **basis** if linearly independent and $G = \langle S\rangle$. Note: for an abelian group G written with addition, we have **basis**
$$\langle S\rangle = \{\, c_1x_1 + \cdots + c_nx_n \mid c_i \in \mathbb{Z},\ x_i \in S \,\}.$$
We often write "$\mathbb{Z}S$" instead of $\langle S\rangle$ for this subset.

A **free abelian group** is one for which there exists a basis. Note that many abelian groups are not free, e.g., the finite cyclic groups. This is unlike the situation for vector spaces, which always admit a basis. **free abelian group**

Warning. There is also a notion of "free group", which we will not talk about. Free abelian groups are not the same as free groups, which is why I keep saying "free abelian".

39.2. *Example.* If $G = \mathbb{Z}_n$, then no non-empty subset is linearly independent. Thus this cannot be a free abelian group (unless $n = 1$, in which case $\mathbb{Z}_1$ is the trivial group).

39.3. *Exercise.* Explain why the trivial group is a free abelian group.

39.4. *Example.* Let $G = \mathbb{Q}^\times/\{\pm 1\}$, the quotient of the group $\mathbb{Q}^\times = \mathbb{Q} \smallsetminus \{0\}$ of non-zero rationals under multiplication, by the subgroup consisting of ± 1. Then it turns out that G is a free abelian group, with an infinite set of generators $S \subseteq G$ corresponding to the set of all prime numbers.

The point is that (because of unique prime factorization of integers) every non-zero rational number can be written uniquely in the form

$$\pm p_1^{k_1} \cdots p_r^{k_r}, \qquad k_i \in \mathbb{Z}, \quad p_1, \dots, p_r \text{ distinct primes.}$$

In the quotient group G, this implies that the collection $S = \{\, \{\pm p\} \mid p \text{ prime}\,\}$ of cosets of $\{\pm 1\}$ is a basis of G.

The moral is that (i) not all abelian groups are free-abelian, but (ii) the free-abelian groups have properties similar to vector spaces.

39.5. *Example.* $\mathbb{Z}^n$ has $\{e_1, \dots, e_n\}$ as a basis, where $e_i = (0, \dots, 1, \dots, 0)$, so is a free abelian group. This is called the *standard basis* of $\mathbb{Z}^n$.

39.6. **Proposition.** *Let G be an abelian group and let S be the list of elements $e_1, \dots, e_n$ of G. Then the function $\phi\colon \mathbb{Z}^n \to G$ defined by $(r_1, \dots, r_n) \mapsto \sum r_i e_i$ is a homomorphism of groups.*

Furthermore, ϕ is injective iff S is linearly independent, and ϕ is surjective iff $G = \mathbb{Z}S$. Thus, ϕ is an isomorphism iff S is a basis.

Proof. Verify that ϕ is a homomorphism:

$$\phi\big((r_1, \dots, r_n) + (s_1, \dots, s_n)\big) = \phi\big((r_1 + s_1, \dots r_n + s_n\big) = (r_1 + s_1)e_1 + \cdots + (r_n + s_n)e_n,$$

$$\phi\big((r_1, \dots, r_n)\big) + \big((s_1, \dots, s_n)\big) = (r_1 e_1 + \cdots + r_n e_n) + (s_1 e_1 + \cdots + s_n e_n),$$

which are equal.

The statement about injectivity is immediate from the definition: B is linearly independent iff $\ker(\phi) = \{0\}$. Likewise the statement about surjectivity, since $\phi(\mathbb{Z}^n) = \mathbb{Z}S$. □

39.7. **Corollary.** *Every free abelian group with a finite basis is isomorphic to $\mathbb{Z}^n$ for some $n \geq 0$.*

Proof. If $S \subseteq G$ is a finite basis for G, then the previous proposition gives an isomorphism $\mathbb{Z}^n \to G$. □

39.8. **Corollary.** *Every finitely generated abelian group is isomorphic to $\mathbb{Z}^n/N$ for some $n \geq 0$ and subgroup $N \leq \mathbb{Z}^n$.*

Proof. If $S \subseteq G$ is a finite generating set for G, then the previous proposition gives a surjective homomorphism $\phi\colon \mathbb{Z}^n \to G$. The isomorphism theorem says this induces an isomorphism $\mathbb{Z}^n/N \simeq G$, where $N = \ker(\phi)$. □

40. Subgroups of finitely generated groups

Every quotient group of a finitely generated group is finitely generated. **Lecture 24**

40.1. *Exercise.* Suppose G is a group and $G = \langle a_1, \dots, a_n \rangle$. Show that for any normal subgroup $N \trianglelefteq G$, we have that $G/N = \langle a_1 N, \dots, a_n N \rangle$.

We will also need the following.

40.2. *Example.* If $G = \langle a_1, \dots, a_n \rangle$ and if $N = \langle a_n \rangle$ is a normal subgroup, then G/N is generated by a subset of $n - 1$ elements: $\{a_1 N, \dots, a_{n-1} N\}$.

It is different for subgroups. The subgroup of a finitely generated group can fail to be finitely generated!

40.3. *Example.* Let $G = \langle A, B \rangle$ be the subgroup of $GL_2(\mathbb{R})$ generated by the matrices $A := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $B := \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$. Show that the subset $H \subseteq G$ consisting of matrices of the form $\begin{bmatrix} 1 & * \\ 0 & 1 \end{bmatrix}$ is a subgroup of G, and that H is not finitely generated.

However, if we assume the group is abelian, this does work.

40.4. **Proposition.** *Let G be a finitely generated abelian group. Then any subgroup $H \leq G$ is finitely generated.*

To prove this we will need a lemma.

40.5. **Lemma.** *Let H be an abelian group with subgroup $N \leq H$. If both N and H/N are finitely generated, then K is finitely generated.*

Proof. Straightforward; in fact, "abelian" is not actually necessary here, as long as we require that N is a normal subgroup of H, though it is a little more difficult to write the proof out in the nonabelian case.

By hypothesis we can write $N = \mathbb{Z}\{x_1, \dots, x_m\}$ and $H/N = \mathbb{Z}\{\overline{y}_1, \dots, \overline{y}_n\}$. Choose $y_1, \dots, y_n \in H$ such that $\pi\colon (y_k) = \overline{y}_k$, where $\pi\colon H \to H/N$ is the quotient homomorphism. (That is, $\overline{y}_i = y_i + N$.) The claim is that $\{x_1, \dots, x_m, y_1, \dots, y_n\}$ generate H.

Suppose $v \in H$, and consider $\pi(v) \in H/N$. By hypothesis, there are $b_1, \dots, b_n \in \mathbb{Z}$ such that

$$\pi(v) = b_1\overline{y}_1 + \cdots + b_n\overline{y}_n.$$

Set

$$u := v - (b_1 y_1 + \cdots + b_n y_n).$$

We have

$$\pi(u) = \pi(v) - (b_1\overline{y}_1 + \cdots + b_n\overline{y}_n) = 0,$$

so $u \in \ker(\pi) = N$. Thus by hypothesis there exist $a_1, \dots, a_m \in \mathbb{Z}$ such that

$$u = a_1 x_1 + \cdots + x_m x_m.$$

Therefore

$$v = u + (b_1 y_1 + \cdots + b_n y_n) = (a_1 x_1 + \cdots + a_m x_m) + (b_1 y_1 + \cdots + b_n y_n),$$

as desired. □

Proof of theorem. For each $n \geq 0$ we will show "if G can be generated by a subset with n elements, then every subgroup of G is finitely generated". I'll work by induction on n.

If $n = 0$, this means G is the trivial group $G = \{0\}$, and the claim is obvious. If $n = 1$, then G is a cyclic group, and we have shown that every subgroup of a cyclic group is cyclic, so finitely generated.

Suppose G is abelian and generated by n elements $a_1, \dots, a_n$, with $n \geq 2$. Then the quotient group $G/\mathbb{Z}a_n$ is generated by the $n-1$ elements $\pi(a_1), \dots, \pi(a_{n-1})$. Write $a := a_n$.

Now suppose $H \leq G$ is any subgroup. We have a "diamond" associated to the subgroups H and $\mathbb{Z}a$, both of which are normal since G is abelian:

$$\begin{tikzcd}
& G \arrow[rr, two heads] & & G/\mathbb{Z}a & \\
& H+\mathbb{Z}a \arrow[u, tail] \arrow[rr, two heads] & & H+\mathbb{Z}a/\mathbb{Z}a \arrow[u, tail] & \\
\mathbb{Z}a \arrow[ur, tail] & & H \arrow[ul, tail] \arrow[rr, two heads] & & H/H\cap\mathbb{Z}a \arrow[ul, "\simeq"'] \\
& H\cap\mathbb{Z}a \arrow[ul, tail] \arrow[ur, tail] & & &
\end{tikzcd}$$

The diamond isomorphism theorem says that $H/H \cap \mathbb{Z}a \simeq H + \mathbb{Z}a/\mathbb{Z}a$. By the correspondence theorem, $H + \mathbb{Z}a/\mathbb{Z}a \leq G/\mathbb{Z}a$.

To show H is finitely generated, by the previous proposition it is enough to show that both $N := H \cap \mathbb{Z}a$ and $H/N = H/H \cap \mathbb{Z}a$ are finitely generated. In fact:

- $H \cap \mathbb{Z}a$ is a subgroup of the cyclic group $\mathbb{Z}a$, so is cyclic.
- $H/H \cap \mathbb{Z}a$ is isomorphic to $H + \mathbb{Z}a/\mathbb{Z}a$, which is a subgroup of $G/\mathbb{Z}a$. Since $G/\mathbb{Z}a$ is generated by $n-1$ elements, by induction any subgroup of can be generated by $n-1$ elements.

Thus H can be generated by n elements. □

We can now prove step 2 of our plan.

40.6. **Proposition.** *If G is a finitely generated abelian group, then $G \simeq \mathbb{Z}^m/\alpha(\mathbb{Z}^n)$ for some $m, n \geq 0$ and homomorphism $\alpha\colon \mathbb{Z}^n \to \mathbb{Z}^m$.*

Proof. We know we have some isomorphism $G \simeq \mathbb{Z}^m/N$, where N is some subgroup of $\mathbb{Z}^m$. The subgroup N is finitely generated by the theorem, so there exists a surjective homomorphism $\overline{\alpha}\colon \mathbb{Z}^n \to N$. □

41. Integer matrices and maps between finitely generated free-abelian groups

We now know that every finitely generated free-abelian group G is isomorphic to $\mathbb{Z}^n$ for some $n \geq 0$. n. We can represent any homomorphism between such by matrices.

41.1. **Proposition.** *Every homomorphism $\phi\colon \mathbb{Z}^n \to \mathbb{Z}^m$ has the form*

$$\phi(x_1, \dots, x_n) = \Big(\sum a_{1j}x_j, \dots, \sum a_{mj}x_j\Big)$$

for a unique matrix $A = (a_{ij}) \in \operatorname{Mat}_{m\times n}(\mathbb{Z})$ with integer coefficents.

Conversely, any matrix $A \in M_{m\times n}(\mathbb{Z})$ deterimes a homomorphism $L_A\colon \mathbb{Z}^n \to \mathbb{Z}^m$, to be thought of as left multiplication of a column vector over $\mathbb{Z}$.

Furthermore, we have that $L_A L_B = L_{AB}$ whenever this makes sense.

Proof. Write $e_1 = (1, 0, \dots, 0), \dots, e_n = (0, \dots, 0, 1)$ for the "standard basis" of $\mathbb{Z}^n$, and similarly $f_1, \dots, f_m$ for the standard basis of $\mathbb{Z}^m$. Thus

$$(x_1, \dots, x_n) = x_1e_1 + \cdots + x_ne_n, \qquad x_i \in \mathbb{Z}.$$

Define $a_{ij} \in \mathbb{Z}$ by the formulas

$$\phi(e_j) = (a_{1j}, \dots, a_{mj}) = \sum_i a_{ij} f_i.$$

Then for a homomorphism $\phi\colon \mathbb{Z}^n \to \mathbb{Z}^m$ we have

$$\begin{aligned}\phi((x_1,\dots,x_n)) &= \phi(x_1e_1+\cdots+x_ne_n)\\ &= \sum_j x_j\phi(e_j)\\ &= \sum_j\sum_i x_ja_{ij}f_i\\ &= \sum_i\big(\sum_j a_{ij}x_j\big)f_i.\end{aligned}$$

So $\phi(x) = L_A(x)$ where $A = (a_{ij})$. The converse is easy: any A gives a homomorphism L_A.

It is easy to check that $L_{AB} = L_AL_B$; it's exactly the same as for linear maps of vector spaces. □

Given finitely generated free-abelian groups G and G' with bases $B = \{v_1,\dots,v_n\}$ and $B' = \{w_1,\dots,w_m\}$, for any homomorphism $\phi\colon G \to G'$ we get an $m\times n$ $\mathbb{Z}$-matrix

$$A = [\phi]_{B',B} = (a_{ij})$$

characterized by

$$\phi(v_j) = \sum_i a_{ij}w_i.$$

In particular, if $G = G'$, but we have two different bases, we get "change of basis matrices"

$$P = [\mathrm{id}]_{B',B}, \qquad Q = [\mathrm{id}]_{B,B'}.$$

Exercise: $PQ = I_n$ and $QP = I_m$.

In the theory of vector spaces, a big result is that finite dimensional vector spaces have a well-defined dimension. More precisely: if V is spanned by a finite set, then V has at least one basis, and any two bases have the same size. We have the same result for free abelian groups.

41.2. **Theorem** (Classification of finitely generated free-abelian groups)**.** *If G is a free abelian group, which has two finite bases B and B', then $|B| = |B'|$. Thus any free-abelian group with a finite basis is isomorphic to one of the form $G \approx \mathbb{Z}^r$ for a unique $r \geq 0$.*

Proof. Let B, B' be bases of G with sizes n and m. Then we have $\mathbb{Z}$-matrices P and Q ($m\times n$ and $n\times m$ respectively) such that $PQ = I_n$ and $QP = I_m$. But these $\mathbb{Z}$-matrices are also $\mathbb{R}$-matrices, and we know that this cannot happen unless $m = n$. □

It will be a consequence of the classification theorem will prove that any *finitely generated* free-abelian group has a finite basis, so this result is usually stated as: "every finitely generated free-abelian group is isomrphic to $\mathbb{Z}^r$ for a unique $r \geq 0$". This number r is called the **rank** of the free-abelian group. **rank**

41.3. **Proposition.** *Every finitely generated abelian group G is isomorphic to $\mathbb{Z}^m/L_A(\mathbb{Z}^n)$ for some m, n and some $\mathbb{Z}$-matrix $A \in M_{m\times n}(\mathbb{Z})$.*

Proof. Choose a finite generating set S of G (size m). This gives a surjective homomorphism $\psi\colon \mathbb{Z}^m \to G$, hence by the isomorphism theorems an isomorphism $\mathbb{Z}^m/N \approx G$, where $N = \operatorname{Ker}\psi$. By what we just proved, H also has a finite generating set T (size n, which can actually be chosen so $n \geq m$). Construct a surjective homomorphism $\chi\colon \mathbb{Z}^n \to H$. Taken together, we get a composite

$$\mathbb{Z}^n \xrightarrow{\chi} H \xrightarrow{\text{incl.}} \mathbb{Z}^m$$

which we call α, with $\alpha(\mathbb{Z}^n) = N$. The result follows, using the fact that any such $\phi = L_A$ for some $A \in \mathrm{Mat}_{m\times n}(\mathbb{Z})$. □

42. Smith normal form

The upshot of the above discussion is that to understand the structure of finitely generated abelian groups, we need to understand the structure of homomorphisms between finitely generated free groups, which we can describe using matrices with integer entries.

Smith normal form

An integer matrix $A \in M_{m\times n}(\mathbb{Z})$ is in **Smith normal form** if it is of the form

$$A = \begin{pmatrix} d_1 & 0 & 0 & \cdots \\ 0 & d_2 & 0 & \cdots \\ 0 & 0 & d_3 & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix}$$

where each $d_i \geq 0$ and $d_i \mid d_j$ if $i \leq j$. (Give examples.)

Note: we allow d_i to be 0. Since 0 only divides itself, this means that if any $d_i = 0$, then $d_j = 0$ for all $i \geq j$.

Note: A is not required to be square. Thus there are $\min(m, n)$ diagonal entries.

Note: it is possible for some $d_i = 1$, in which case $d_j = 1$ for all $j \leq i$.

Example:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 300 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

42.1. **Proposition.** *Let $L_A \colon \mathbb{Z}^n \to \mathbb{Z}^m$ be given by left multiplication by a matrix A in Smith normal form. Then*

$$\mathbb{Z}^m / L_A(\mathbb{Z}^n) \approx \mathbb{Z}_{d_1} \times \cdots \times \mathbb{Z}_{d_k} \times \mathbb{Z}^{m-k},$$

where $d_1, \dots, d_k$ are the non-zero *diagonal entries of B, satisfying $d_1 \mid d_2 \mid \cdots \mid d_k$.*

Proof. Let $G := \mathbb{Z}_{d_1} \times \cdots \times \mathbb{Z}_{d_k} \times \mathbb{Z}^{m-k}$, and let $\phi \colon \mathbb{Z}^m \to G$ be the "obvious" homomorphism:

$$\phi(c_1, \dots, c_n) := \big([c_1]_{d_1}, \dots, [c_k]_{d_k}, c_{k+1}, \dots, c_n\big).$$

This is certainly surjective, so $G \approx \mathbb{Z}^m/N$ where $N = \ker(\phi)$. Explicitly, the elements of N are exactly those of the form $(d_1x_1, \dots, d_kx_k, 0, \dots, 0)$, where $x_1, \dots, x_k \in \mathbb{Z}$, and these are exactly the elements in the image of L_A. □

Note that we might have some $d_i = 1$, which give a trivial group factor $\mathbb{Z}_1$, which we can discard from the statement of the isomorphism.

$\mathbb{Z}$-invertible

Say that a square matrix $P \in M_{n\times n}(\mathbb{Z})$ is **$\mathbb{Z}$-invertible** if there exists $Q \in M_{n\times n}(\mathbb{Z})$ such that $PQ = I = QP$. (Exercise: show that if such Q if it exists is actually unique; we can call it P^{-1}.)

Warning: $\mathbb{Z}$ is contained in the field $\mathbb{Q}$ (or $\mathbb{R}$). An integer matrix can be $\mathbb{Q}$-invertible but not $\mathbb{Z}$ invertible. (Give exmaple: $\left[\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right]^{-1} = \begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix}$.)

42.2. **Theorem** (Proposition 3.5.9)**.** *For every $A \in M_{m\times n}(\mathbb{Z})$, there exist $\mathbb{Z}$-invertible $P \in M_{n\times n}(\mathbb{Z})$, $Q \in M_{m\times m}(\mathbb{Z})$, such that $B = PAQ$ is in Smith normal form.*

Before we prove this theorem, we use it to get invariant factor form.

Easy fact: if $P \in M_{n\times n}(\mathbb{Z})$ is $\mathbb{Z}$-invertible, then the induced homomorphism $L_P \colon \mathbb{Z}^n \to \mathbb{Z}^n$ is an isomorphism of groups, with inverse map $(L_P)^{-1} = L_{P^{-1}}$.

42.3. **Lemma.** *If $A, B \in M_{m\times n}(\mathbb{Z})$ and $B = PAQ$ with $P \in M_{n\times n}(\mathbb{Z})$ and $Q \in M_{m\times m}(\mathbb{Z})$ both $\mathbb{Z}$-invertible, then there is an isomorphism $\mathbb{Z}^m/L_A(\mathbb{Z}^n) \approx \mathbb{Z}^m/L_B(\mathbb{Z}^m)$.*

Proof. We use the isomorphism theorem. Let $\pi_A\colon \mathbb{Z}^m \to \mathbb{Z}^m/L_A(\mathbb{Z}^n)$ be the quotient homomorphism associated to the subgroup $L_A(\mathbb{Z}^n) \le \mathbb{Z}^m$, and likewise $\pi_B\colon \mathbb{Z}^m \to \mathbb{Z}^m/L_B(\mathbb{Z}^n)$. Consider the diagram

$$\begin{tikzcd}
\mathbb{Z}^m \arrow[r, "\phi:=\pi_A L_{P^{-1}}"] \arrow[d, "\pi_B"'] & \mathbb{Z}^m/L_A(\mathbb{Z}^n) \\
\mathbb{Z}^m/L_B(\mathbb{Z}^n) \arrow[ur, dotted, "\overline{\phi}"'] &
\end{tikzcd}$$

Because P^{-1} is $\mathbb{Z}$-invertible, $L_{P^{-1}}\colon \mathbb{Z}^m \to \mathbb{Z}^m$ is an isomorphism. Thus the composite $\phi := \pi_A L_{P^{-1}}$ is a surjective homomorphism.

Claim. $\ker\phi = L_B(\mathbb{Z}^n)$.

Proof of claim. We have that

$$x \in \ker\phi \quad \Leftrightarrow \quad \pi_A(P^{-1}x) = 0 \quad \Leftrightarrow \quad P^{-1}x \in L_A(\mathbb{Z}^n).$$

That is, $x \in \ker\phi$ iff there exists $y \in \mathbb{Z}^n$ such that $P^{-1}x = Ay$, or equivalently $x = PAy$. Since Q is invertible we can form $z := Q^{-1}y$, so $y = Qz$, so we have that

$$x \in \ker\phi \quad \Leftrightarrow \quad \textit{exists } z \in \mathbb{Z}^n \textit{ st } x = PAQz = Bz.$$

We have proved $\ker\phi = L_B(\mathbb{Z}^n)$.

Now the isomorphism theorem tells us that there is an isomorphism $\overline{\phi}$. □

The theorem is equivalent to the following, which together with what we have shown proves the existence part of invariant factor decomposition.

42.4. **Corollary.** *Every finitely generated abelian group is isomorphic to one in invariant factor form.*

Proof, using the theorem and lemma. We have already shown that every finitely generated abelian group is isomorphic to one of the form $G \approx \mathbb{Z}^m/L_A(\mathbb{Z}^n)$ for some integer matrix A. The theorem says that there exists $B = PAQ$ in Smith normal form with P and Q $\mathbb{Z}$-invertible. The lemma says that $G \approx \mathbb{Z}^m/L_B(\mathbb{Z}^n)$, and we have already seen that if B is in Smith normal form, then $\mathbb{Z}^m/L_B(\mathbb{Z}^n)$ has an invariant factor form. □

43. Proof of existence of Smith normal form

Lecture 25
equivalent

Say that two matrices $A, B \in \mathrm{Mat}_{m\times n}(\mathbb{Z})$ are **equivalent** if there exist $\mathbb{Z}$-invertible P, Q such that $B = QAP$. *Exercise:* this is an equivalence relation.

We show that Smith normal form exists for any $A \in \mathrm{Mat}_{m\times n}(\mathbb{Z})$, by an induction on $s = \min(m, n)$, using the following reduction step.

43.1. **Lemma.** *Every $A \in \mathrm{Mat}_{m\times n}(\mathbb{Z})$ is equivalent to one of the form*

$$A' = \left[\begin{array}{c|ccc} d & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array}\right]$$

with $d \in \mathbb{Z}_{\ge 0}$, $B \in \mathrm{Mat}_{(m-1)\times(n-1)}(\mathbb{Z})$, such that d divides every entry of B.

Note: when $m = 1$ or $n = 1$, this says that A' will be a row or column matrix with only one non-zero entry at $(1, 1)$.

Assuming the lemma, we obtain Smith normal form by induction on s. So by induction we have a Smith normal form $P'BQ' = D'$ with $D' = \operatorname{diag}(d_2, \dots, d_s)$ and $d_i \mid d_j$ when $i \leq j$, so that

$$A \sim A' = \left[\begin{array}{c|c} d_1 & \\ \hline & B \end{array}\right] \sim \left[\begin{array}{c|c} 1 & \\ \hline & P' \end{array}\right] \left[\begin{array}{c|c} d_1 & \\ \hline & B \end{array}\right] \left[\begin{array}{c|c} 1 & \\ \hline & Q' \end{array}\right] = \left[\begin{array}{c|c} d_1 & \\ \hline & D' \end{array}\right].$$

Furthermore, since d_1 divides every entry of B, it divides every entry of $D' = P'BQ'$, so $d_1 \mid d_i$ for all i.

43.2. *Exercise.* If P, B, Q are integer matrices, and $d \in \mathbb{Z}$ divides every entry of B, then d divides every entry of PBQ.

The key idea is that there are basic examples of $\mathbb{Z}$-invertible matrices called *elementary matrices*, so that $A \mapsto PA$ performs a row operation, and $A \mapsto AP$ performs a column operation.

We note three kinds of "elementary" square integer matrices in $M_{n\times n}(\mathbb{Z})$:

- $I + cE_{i,j}$, where $E_{i,j}$ is the matrix with 1 in position (i, j) and all other entries 0, and $i \neq j$.
  Example:
  $$P = I + 7E_{1,2} = \begin{pmatrix} 1 & 7 \\ 0 & 1 \end{pmatrix},$$
  so $A \mapsto PA$ is the operation $\mathrm{Row}_1 \to \mathrm{Row}_1 + 7\,\mathrm{Row}_2$, and $A \mapsto AP$ is $\mathrm{Col}_2 \to \mathrm{Col}_2 + 7\,\mathrm{Col}_1$.
  Exercise: $(I + cE_{i,j})^{-1} = I - cE_{i,j}$, another elementary matrix.
- The permutation matrix $P_{i,j}$ correpsonding to the 2-cycle (i, j).
  Example:
  $$P = P_{1,2} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$
  so $A \mapsto PA$ is $\mathrm{Row}_1 \leftrightarrow \mathrm{Row}_2$, and $A \mapsto AP$ is $\mathrm{Col}_1 \leftrightarrow \mathrm{Col}_2$.
  Note that $P_{i,j}^{-1} = P_{i,j}$.
- The matrix M_i with diagonal entries $= 1$ except for $a_{ii} = -1$, and other entries 0.
  Example:
  $$P = M_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$
  so $A \mapsto PA$ is $\mathrm{Row}_2 \to -\mathrm{Row}_2$, and $A \mapsto AP$ is $\mathrm{Col}_2 \to -\mathrm{Col}_2$.
  Note that $M_i^{-1} = M_i$.

Left multiplication corresponds to row operations on integer matrices which are reversible. Right multiplication corresponds to column operations which are reversible.

Thus, if we operate on A by a series of row and column operations, the resulting matrix has the form $B = PAQ$ for invertible P and Q, where P and Q are each products of elementary matrices.

We will show that any $\mathbb{Z}$-matrix can be converted into Smith normal form by a sequence of row and column operations.

43.3. *Example* (Euclidean algorithm). Here is a very simple example, with a 1×2 matrix. Note that a 1×2 matrix is in Smith normal form iff it looks like $[d\ 0]$ with $d \geq 0$.

We will only use column operations.

$$\begin{aligned}
[42\ 15] &\xrightarrow{\mathrm{Col}_1 \leftrightarrow \mathrm{Col}_2} [15\ 42] \xrightarrow{\mathrm{Col}_2 \to \mathrm{Col}_2 - 2\mathrm{Col}_1} [15\ 12] \\
&\xrightarrow{\mathrm{Col}_1 \leftrightarrow \mathrm{Col}_2} [12\ 15] \xrightarrow{\mathrm{Col}_2 \to \mathrm{Col}_2 - \mathrm{Col}_1} [12\ 3] \\
&\xrightarrow{\mathrm{Col}_1 \leftrightarrow \mathrm{Col}_2} [3\ 12] \xrightarrow{\mathrm{Col}_2 \to \mathrm{Col}_2 - 4\mathrm{Col}_1} [3\ 0].
\end{aligned}$$

44. End of the proof of Smith normal form

Finally, I prove the Lemma.

Proof of Lemma. If $A = 0$, then it is already in the appropriate form, so assume $A \neq 0$.

If $A \neq 0$, then by switching rows and/or columns, and possibly changing the sign on a row or column, we can assume that A is equivalent to a matrix whose $(1,1)$ entry is a positive integer a.

We now show by induction on a, that if A has $a_{11} = a > 0$, then the statement of the lemma is true for A. We have three cases.

(a) If a divides every entry of A, then we can use row and column operations to transform A into a block matrix of the form $A' = \left[\begin{array}{c|c} a & 0 \\ \hline 0 & B \end{array}\right]$. Furthermore, since $A' = PAQ$, we see that a also divides every entry of A', so we are done.

In particular, this case always applies when $a = 1$, giving us the base case of the induction.

(b) Suppose a does not divide some entry of the top row or left column. Then $A \sim A'$, where the $(1,1)$ entry of A' is a positive integer strictly less that a, whence the claim of the Lemma follows by induction. We obtain A' from A by row and or column operations, which carry out a Euclidean algorith.

For instance, suppose $b = a_{1j}$ is not divisible by a. Then by a sequence of column operations involving only the 1st and jth columns, we get

$$A = \left[\begin{array}{c|c|c|c} a & * & b & * \\ \hline * & * & * & * \end{array}\right] \xrightarrow{\mathrm{Col}_j \to \mathrm{Col}_j - q\mathrm{Col}_1} \left[\begin{array}{c|c|c|c} a & * & r & * \\ \hline * & * & * & * \end{array}\right] \xrightarrow{\mathrm{Col}_1 \leftrightarrow \mathrm{Col}_j} \left[\begin{array}{c|c|c|c} r & * & a & * \\ \hline * & * & * & * \end{array}\right]$$

where $r = \mathrm{rem}_a(b) = b - qa$ for some $q \in \mathbb{Z}$. Since $a \nmid b$ we have $0 < r < a$, as desired.

(c) Suppose a does divide every entry of the top row and left column, but does not divide some entry a_{ij} with $i, j > 1$. By the same argument as in (a), we can use row and column operations to transform A into one of the form $A' = \left[\begin{array}{c|c} a & 0 \\ \hline 0 & B \end{array}\right]$. There will still be an entry $b = a_{ij}$ with $i, j > 1$ which a does not divide. Perform the row operation which adds the ith row to the 1st row, obtaining $A'' = \left[\begin{array}{c|c|c|c} a & 0 & b & 0 \\ \hline 0 & * & * & * \end{array}\right]$. Then as in (b) we have that A'' is equivalent to one with a smaller entry in position $(1,1)$, whence the claim of the Lemma follows by induction. (Note: this case is relevant only if $s = \min(m, n) > 1$, i.e., if the matrix is just a row or column vector you can never get to this step.)

□

44.1. *Example.* Here's an example of a reduction to Smith normal form.

$$\begin{bmatrix} 4 & 0 & 2 \\ 3 & 1 & 0 \end{bmatrix} \xrightarrow{\mathrm{Col}_1 \leftrightarrow \mathrm{Col}_3} \begin{bmatrix} 2 & 0 & 4 \\ 0 & 1 & 3 \end{bmatrix} \xrightarrow{\mathrm{Col}_3 \to \mathrm{Col}_3 - 2\mathrm{Col}_1} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 3 \end{bmatrix}$$

$$\xrightarrow{\mathrm{Row}_1 \to \mathrm{Row}_1 + \mathrm{Row}_2} \begin{bmatrix} 2 & 1 & 3 \\ 0 & 1 & 3 \end{bmatrix} \xrightarrow{\mathrm{Col}_1 \leftrightarrow \mathrm{Col}_2} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 3 \end{bmatrix}$$

$$\xrightarrow{\mathrm{Col}_2 \to \mathrm{Col}_2 - 2\mathrm{Col}_1} \begin{bmatrix} 1 & 0 & 3 \\ 1 & -2 & 3 \end{bmatrix} \xrightarrow{\mathrm{Col}_3 \to \mathrm{Col}_3 - 3\mathrm{Col}_1} \begin{bmatrix} 1 & 0 & 0 \\ 1 & -2 & 0 \end{bmatrix}$$

$$\xrightarrow{\mathrm{Row}_2 \to \mathrm{Row}_2 - \mathrm{Row}_1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \end{bmatrix} \xrightarrow{\mathrm{Row}_2 \to -\mathrm{Row}_2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix}$$

45. Regular polyhedra

A **regular polyhedron** is a polyhedron all of whose faces are congruent regular polygons, and which have the same number of faces at each vertex. (More or less: it is surprisingly hard to give a precise definition of "regular polyhedron".)

regular polyhedron

There are five of these: tetrahedron, cube and octahedron, dodecahedron and icosahedron.

	vertices	edges	faces	vertices/edges per face	faces/edges per vertex
tetrahedron	4	6	4	3	3
cube	8	12	6	4	3
octahedron	6	12	8	3	4
dodecahedron	20	30	12	5	3
icosahedron	12	30	20	3	5

Note that some of these come in "dual pairs". For every polyhedron P, there is a **dual polyhedron** P^*, whose vertex set is the set of centroids of the faces of P. The dual of a cube is an octahedron and vice versa, and the dual of a dodecahedron is an icosahedron, and vice versa. The dual of a tetrahedron is another tetrahedron.

dual polyhedron

Each of these has a *rotational symmetry group*, consisting of rotations of the polyhedron about its center of mass. We might as well the center of mass is at the origin, so that the rotational symmetry group is a subgroup of $SO(3)$. We can identify this with the set of symmetries of the vertices of the polyhedron, so the rotation group of a polyhedron with vertex set V is

$$G := \{\, A \in SO(3) \mid v \in V \iff Av \in V \,\}.$$

This G is a finite group: if A fixes every vertex v, then $A = I$, so we can identify G with a subgroup of $\mathrm{Sym}(V)$, which is finite.

Because every element of G is a rotation, it must take vertices to vertices, edges to edges, and faces to faces.

Here is one way to compute the order of the symmetry group G. Fix an edge, and label its endpoints A and B. (I like to think of attaching a "handle" to this edge.) Any symmetry must take the "vector" $\overrightarrow{AB}$ to another vector $\overrightarrow{A'B'}$, where A' and B' are also the endpoints of an edge. This information completely determines the symmetry, and for every directed edge $\overrightarrow{A'B'}$ there is a unique symmetry taking $\overrightarrow{AB}$ to $\overrightarrow{A'B'}$. Thus we have

$$|G| = 2 \times (\text{number of edges}).$$

	order of symmetry group
tetrahedron	12
cube	24
octahedron	24
dodecahedron	60
icosahedron	60

All non-identity symmetries will be *rotations* through an axis which passes through the center of the polyhedron (which we will assume is the origin), through an angle of the form $2\pi(k/n)$ for some integers k, n (since the rotation must be of finite order). So we can organize symmetries according to (i) the axis of the symmetry, and (ii) their order.

46. Rotational symmetries of the tetrahedron

Here are the symmetries of the tetrahedron.

axis of symmetry	number of axes	angle of rotation	order of symmetry	number of symmetries
vertex/centroid of opposite faces	4	$+2\pi/3$	3	4
vertex/centroid of opposite face	4	$-2\pi/3$	3	3
midpoint of opposite edges	3	$2\pi/2$	2	3
*	*	0	1	1

I'm thinking of the first class as being rotations of the form $\mathrm{Rot}_v(+2\pi/3)$, where v is a *vertex* of the tetrahedron (so that $-v$ points in the direction of the centroid of the opposite face). So these are *counterclockwise* as viewed from the vertex v. Then the second class consists of *clockwise* rotations viewed from the vertex v, of the form $\mathrm{Rot}_v(-2\pi/3)$. The third class is $\mathrm{Rot}_u(\pi)$, where u is the midpoint of some edge.

The rotational symmetry group G faithfully permutes the set of four vertices ("faithfully" means only the identity symmetry acts as identity on the set of vertices). Thus G is isomorphic to a subgroup of S_4.

Two elements $x, y \in G$ in a group are **conjugate** if $y = gxg^{-1}$ for some $g \in G$. **conjugate**

46.1. **Proposition.** *The group G of rotational symmetries of the tetrahedron is isomorphic to A_4, the subgroup of even permutations in S_4.*

Furthermore, two elements of G are in the same class in the above chart iff they are conjugate in G.

Proof. Elements of G are determined by how the permute the set V of vertices. Label the vertices $1, 2, 3, 4$. Then we can identify G with a subgroup of S_4 of order 12 which has 8 elements of order 3. Since A_4 is generated by the set of all elements of order 3 in S_4, and $|A_4| = 12$, this must be the answer. (The three order 2 elements correspond to the products of disjoint transpositions.)

The statement about conjugacy is just about when elements of A_4 are conjugate, and I'll leave you to think about it. Note that although any two 3-cycles are conjugate in S_4, a 3-cycle g is *not* conjugate to $g^2 = g^{-1}$ in A_4. □

46.2. *Remark.* A convenient example of the vertices of a tetrahedron is the set of points of the form $(\pm 1, \pm 1, \pm 1)$ where there are an even-number of minus signs. That is:

$$(1,1,1), \quad (1,-1,-1), \quad (-1,1,-1), \quad (-1,-1,1).$$

From this you can work out matrix representations of every element of G: they are the ones of the form

$$\begin{bmatrix} \pm 1 & & \\ & \pm 1 & \\ & & \pm 1 \end{bmatrix}, \quad \begin{bmatrix} & & \pm 1 \\ \pm 1 & & \\ & \pm 1 & \end{bmatrix}, \quad \begin{bmatrix} & \pm 1 & \\ & & \pm 1 \\ \pm 1 & & \end{bmatrix},$$

which have $\det = 1$, i.e., such that the matrix has an even number of negative entries.

47. Rotational symmetries of the cube and octohedron

Here are the rotational symmetries of the cube and the octahedron, which are "dual" polyhedra. **Lecture 26**

axis of symmetry (cube)	number of axes	angle of rotation	order of symmetry	number of symmetries	axis of symmetry (octahedron)
vertex/opposite vertex (diagonal)	4	$\pm 2\pi/3$	3	8	centroids of opposite faces
midpoints of opposite edges	6	$2\pi/2$	2	6	midpoints of opposite edges
centroids of opposite faces	3 3	$\pm 2\pi/4$ $2\pi/2$	4 2	6 3	vertex/opposite vertex (diagonal)
*	*	0	1	1	*

Note that all these axes always pass through points of the same type (two vertices, two midpoints, or two centroids). So there is no way to distinguish "clockwise" vs "counterclockwise" versions of these: what is clockwise viewed from one direction is counterclockwise as viewed from the opposite direction.

47.1. **Proposition.** *The group G of rotational symmetries of the cube/octahedron is isomorphic to S_4. Furthermore, two elements of G are in the same class in the above chart iff they are conjugate in G.*

Proof. It suffices to consider symmetries of the cube, since the octahedron is dual to it.

The trick is to find four things that G permutes faithfully. Such a set is given by the set D of diagonals of the cube. A diagonal is a line segment between a pair of opposite vertices, which we can represent by its set of endpoints $\{v, -v\} \subseteq V$. There are exactly four diagonals, so we get a homomorphism $\phi\colon G \to \mathrm{Sym}(D) \approx S_4$, sending ϕ to the permutation of diagonals defined by $\{\pm v\} \mapsto \{\pm\phi(v)\}$.

The claim is that ϕ is an isomorphism. Since both groups have order 24, it is enough to show that ϕ is injective, i.e., that $\ker(\phi) = \{I\}$, or that $\phi(g) = \mathrm{id}$ implies $g = \mathrm{id}$.

Actually, it's easiest to show the contrapositive: if $g \neq \mathrm{id}$ then $\phi(g) \neq \mathrm{id}$, because we have an explicit list of all the elements.

- Each order 3 rotation g along an axis through opposite sides fixes one diagonal, but moves the other three, so $\phi(g)$ is a 3-cycle. other three diagonals.
- Each order 2 rotation along an axis through opposite edges transposes two diagonals (the ones which touch the edges the axis goes through), but fixes the other two, so $\phi(g)$ is a 2-cycle.
- Each order 4 rotation g along an axis through opposite vertices gives a 4-cycle $\phi(g)$.
- Each order 2 rotation g along an axis through opposite vertices is a square of one of the previous type, so $\phi(g)$ is a square of a 4-cycle, i.e., a product of disjoint 2-cycles.

This also gives the relation to conjugacy in S_4: each class of symmetries in G is exactly sent to permutations with a single cycle type. □

47.2. *Remark.* A convenient set of the vertices of the cube is the set of all points of the form $(\pm 1, \pm 1, \pm 1)$ (so the vertices of the cube can be split up into two dual tetrahedra). The vertices of octahedron dual to this cube have the form $(\pm 1, 0, 0), (0, \pm 1, 0), (0, 0, \pm 1)$.

You can use these to give explicit matrix representations of every element of G: they are all of the form $A_\sigma D$, where A_σ is a permutation matrix for some $\sigma \in S_3$, and $D = \mathrm{diag}(\pm 1, \pm 1, \pm 1)$ with $\det D = 1$.

48. Symmetries of the dodecahedron and icosahedron

Here are the symmetries of the dodecahedron and icosahedron, which are dual polyhedra.

axis of symmetry (dodecahedron)	number of axes	angle of rotation	order of symmetry	number of symmetries	axis of symmetry (icosahedron)
vertex/opposite vertex (diagonal)	10	$\pm 2\pi/3$	3	20	centroids of opposite faces
midpoints of opposite edges	15	$2\pi/2$	2	15	midpoints of opposite edges
centroids of opposite faces	6 6	$\pm 2\pi(1/5)$ $\pm 2\pi(2/5)$	5 5	12 12	vertex/opposite vertex diagonal
*	*	0	1	1	*

48.1. **Proposition.** *The group of rotational symmetries of the dodecahedron/icosahedron is isomorphic to A_5, the subgroup of even permutations in S_5. Two symmetries are in the same class iff they are conjugate in G.*

Proof. It suffices to consider symmetrices of the dodecahedron, since the icosahedron is dual to it.

We need to find five things that G permutes faithfully. There are several choices. The book suggests using "inscribed cubes", which can be hard to visualize without a model.

We can partition the 30 edges of the dodecahedron (or icosahedron) into 5 sets of 6 each. For any edge e, there is one other edge which is parallel to it, and four other edges which are perpendicular edges. Furthermore, any two edges in this set of six are either parallel or perpendicular to each other. (You can orient the polyhedron so that these six edges have their midpoints along the standard coordinate axes.)

Thus we can partition the 30 edges into five sets of six each. Let $X = \{U_1, \dots, U_5\}$ be the five sets.

Exercise. For any face of the dodecahedron, each of its five edges is in a different set U_i.

Then any $g \in G$ permutes these batches, so we get a homomorphism $\phi \colon G \to \mathrm{Sym}(X) \approx S_5$. I claim (i) ϕ is injective, and (ii) all $\phi(g)$ are even permutations. Since $|G| = 60 = |A_5|$, this must give an isomorphism $G \approx A_5$.

Again, I'll go through the non-identity symmetries g by class and show (i) $\phi(g) \neq \mathrm{id}$, and (ii) $\phi(g)$ is an even permutation.

Before starting, it is useful to note: any five edges $E_1, \dots, E_5$ which are the sides of one face of the dodecahedron correspond to the five distinct groups of six.

- Each order 3 rotation g along an axis through opposite vertices permutes three groups (the ones which have an edge touching the axis vertices), and fixes the other two groups. So $\phi(g)$ is a 3-cycle.
- Each order 2 rotation g along an axis through opposite edges fixes only one group (the one which has the axis edges). Because the permutation has order 2, we see that $\phi(g)$ must be a product of disjoint 2-cycles.
- Each order 5 rotation g along an axis through opposite faces gives a cyclic permutation of the groups, so $\phi(g)$ is a five cycle.

Thus, every non-identity element of G goes to a non-identity even permutation in S_5. The statement about conjugacy follows, once we understand which elements are conjugate to each other in A_5.

Note: if $g = (a\ b\ c\ d\ e) \in A_5$ is a 5-cycle, then g^2 is also a 5-cycle, and so g and g^2 are conjugate in S_5. But they are *not* conjugate in A_5. □

48.2. *Remark.* A convenient set of the vertices of an icosahedron is the set of all points of the forms $(0, \pm 1, \pm\varphi)$, $(\pm 1, \pm\varphi, 0)$, $(\pm\varphi, 0, \pm 1)$, where $\varphi = (1+\sqrt{5})/2$. Vertices for a dual dodecahedron are $(\pm(\varphi+1), \pm(\varphi+1), \pm(\varphi+1))$, $(0, \pm(2\varphi+1), \pm\varphi)$, $(\pm(2\varphi+1), \pm\varphi, 0)$, $(\pm\varphi, 0, \pm(2\varphi+1))$.

49. Group actions

An **action** of a group G on a set X is a homomorphism $\phi\colon G \to \mathrm{Sym}(X)$. **action**

To spell this out, this means that for each $g \in G$ and $x \in X$ we get $\phi(g)(x) \in X$. The fact that ϕ is a homomorphism means that this satisfies

$$\phi(g_1 g_2)(x) = \phi(g_1)\big(\phi(g_2)(x)\big), \qquad \text{for all } x \in X,\ g_1, g_2 \in G.$$

Notation. When an action $\phi\colon G \to \mathrm{Sym}(X)$ is chosen, we sometimes use the following abbreviated notation:

$$gx \quad \text{for} \quad \phi(g)(x).$$

That is, we notationally confuse the group element $g \in G$ with the function $\phi(g) \in \mathrm{Sym}(X)$. Then the identity that this has to satisfy becomes

$$(g_1 g_2)x = g_1(g_2 x).$$

This can potentially be confusing: it is possible that different elements $g_1 \neq g_2$ can give the same functions $X \to X$ (e.g., if $\phi\colon G \to \mathrm{Sym}(X)$ is not injective). We will work with this.

Remark. The above notation says we can think of an action of G on X as being given by a "product"

$$(g, x) \mapsto gx \colon G \times X \to X,$$

which satisfies a kind of associative law. Note that this "product" is not any kind of group structure, since it involves elements from two different sets.

Note that since $\phi\colon G \to \mathrm{Sym}(X)$ is a homomorphism, we have

$$\phi(e) = \mathrm{id}, \qquad \phi(g^{-1}) = \phi(g)^{-1}.$$

Here are some examples of group actions.

49.1. *Example* (Tautological action). For any set X, there is a **tautological action** by $G = \mathrm{Sym}(X)$ **tautological action** on X, given by $\iota\colon \mathrm{Sym}(X) \to \mathrm{Sym}(X)$ sending $\iota(\sigma) := \sigma$. Thus the rule for this action is

$$\iota(\sigma)(x) = \sigma(x).$$

In this case, the abbreviated form of the notation is natural.

More generally, any subgroup $H \leq \mathrm{Sym}(X)$ acts on X.

49.2. *Example* (Trivial action). Every group G can be made to act on any set X, by $\phi\colon G \to \mathrm{Sym}(X)$ sending $\phi(g) := \mathrm{id}$ for all $g \in G$. This is called the **trivial action**. **trivial action**

49.3. *Example.* The dihedral group D_n acts on the set V of vertices of the regular n-gon. This action is represented by a homomorphism $D_n \to \mathrm{Sym}(V) \simeq S_n$.

49.4. *Example.* Let V be the set of vertices of a regular polyhedron, and let G be the symmetry group of that polyhedron. Then there is an action by G on V which we can write $\phi\colon G \to \mathrm{Sym}(V)$.

For instance, the symmetry group of the cube (of order 24) acts on the set V of vertices of size 8.

49.5. *Example.* There is also an action by the symmetry group G of a polyhedron on its sets E of edges, and set F of faces.

49.6. *Example.* The group $SO(3)$ of rotations of space around the origin acts on $X = \mathbb{R}^3$, by matrix multiplication: $(A, x) \mapsto Ax$.

49.7. *Example.* We can regard the dihedral group D_n as a subgroup of $SO(3)$, so that it is the set of rotations preserving the regular n-gon $V = \{v_0, \dots, v_{n-1}\}$ in the xy-plane, where $v_k = (\cos 2\pi/k, \sin 2\pi/k, 0)$. This D_n also acts on $\mathbb{R}^3$.

50. Orbits

Given an action $G \to \mathrm{Sym}(X)$, define an equivalence relation on X by $x \sim y$ if there exists $g \in G$ such that $y = gx$. *Exercise.* This is an equivalence relation on X.

I'm going to call this the **action relation** on X.

action relation

Given $x \in X$, let $\mathcal{O}(x) := \{\, gx \mid g \in G \,\}$ be the equivalence class under the action equivalence relation. It is called an *orbit* of the action.

50.1. **Proposition.** *If G acts on a set X, then the orbits of the action give a partition of X into pairwise disjoint subsets.*

Proof. This is just because the orbits are exactly the equivalence classes for an equivalence relation on X. □

An action is **transitive** if there is only one orbit.

transitive

50.2. *Example* (Tautological symmetric group action). The symmetric group S_n comes, by definition, with an action $S_n \curvearrowright X = \{1, 2, \dots, n\}$; this is an example of a tautological action, since $S_n = \mathrm{Sym}(\{1, \dots, n\})$.

This action has a single orbit, since for any element $x \in X$ there is a permutation σ such that $\sigma 1 = x$, namely $\sigma = (1x)$.

50.3. *Exercise.* It is possible for an action to have zero orbits. How can this be?

50.4. *Example.* Consider the standard action of $SO(3)$ on $\mathbb{R}^3$. The orbits of this action are the subsets

$$S_R := \{\, (x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = R^2 \,\}, \qquad R \geq 0.$$

The point is that any vector of length R can be rotated into any other vector of the same length. Thus, the orbits form a family of concentric spheres centered at the origin (together with $S_0 = \{(0, 0, 0)\}$ consisting of just the origin itself).

50.5. *Exercise.* Show that for a regular polyhedron, its group of symmetrices acts transitively on each of its sets of vertices, edges, and faces.

50.6. *Example.* The group $D = \{r_\theta, j_\theta\}$ (symmetries of a disk, as a subgroup of $SO(3)$) acts on $\mathbb{R}^3$. The orbits of D can be described as follows:

- The singleton set $\{(0, 0, 0)\}$ containing the origin is an orbit.
- Each set $C_R^0 := \{\, (x, y, 0) \mid x^2 + y^2 = R^2 \,\}$ for a fixed $R > 0$ (a circle in the xy-plane centered at the origin, is an orbit.
- Each set $\{(0, 0, z), (0, 0, -z)\}$ with $z > 0$ is an orbit.
- Each set $C_R^z := \{\, (x, y, \pm z) \mid x^2 + y^2 = R^2 \,\}$ for $R, z > 0$ (two circles above and below xy-plane) is an orbit.

50.7. *Exercise.* Think of a dihedral group like D_4 as the group of symmetries of the square with vertices $\{v_0, v_1, v_2, v_3\}$ in the xy-plane, where $v_k = (\cos \pi k/2, \sin \pi k/2, 0)$. This identifies D_4 as a subgroup of $SO(3)$, so D_4 acts on $\mathbb{R}^3$. Determine the orbits of this action $D_4 \curvearrowright \mathbb{R}^3$.

51. Stabilizers

Suppose $G \curvearrowright X$, an action of G on some set X. For $x \in X$, define

$$\mathrm{Stab}(x) = \mathrm{Stab}_G(x) := \{\, g \in G \mid gx = x \,\}.$$

51.1. **Proposition.** $\mathrm{Stab}_G(x)$ *is a subgroup of G. (Exercise.)*

Different elements of X can have different stabilizer groups. However, there is one relation to consider: if elements are in the same orbit, then their stabilizer groups are *conjugate.* In particular, this implies that they are isomorphic.

51.2. **Proposition.** *Suppose $G \curvearrowright X$, and $x, y \in X$ are such that $y = gx$. Then* $\operatorname{Stab}(y) = g\operatorname{Stab}(x)g^{-1}$.

Proof. Exercise. □

The **kernel** of an action is just the kernel of the homomorphism $\phi\colon G \to \operatorname{Sym}(X)$. An action is said to be **faithful** if the kernel is the trivial subgroup.

kernel

faithful

51.3. *Exercise.* If $\phi\colon G \to \operatorname{Sym}(X)$ is an action, then $\ker(\phi) = \bigcap_{x\in X}\operatorname{Stab}_G(x)$.

51.4. *Example.* Consider the rotation action by $SO(3)$ on $\mathbb{R}^3$. The stabilizer of $e_3 = (0,0,1)$ is the set of matrices $A \in SO(3)$ such that $Ae_3 = e_3$. This means that $\operatorname{Stab}(e_3)$ consists of all

$$A = \operatorname{Rot}_{e_3}(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} =$$

for some $\theta \in \mathbb{R}$. (This group is isomorphic to $\mathbb{R}/\mathbb{Z}$, as we have seen.)

More generally, if v is any non-zero vector, $\operatorname{Stab}(v) = \{\operatorname{Rot}_u(\theta) \mid \theta \in \mathbb{R}\}$ where $u = v/\|v\|$, while $\operatorname{Stab}(0) = SO(3)$.

51.5. *Example.* Consider the action by D_n on the set of vertices $V = \{v_0, \dots, v_{n-1}\}$ of a regular n-gon. With our usual notation, we have $D_n = \langle r, j\rangle$ with $r \cdot v_k = v_{k+1}$ and $j \cdot v_k = v_{-k}$, where I will set $v_i = v_j$ if $i \equiv j \pmod{n}$.

Then we can compute stabilizer groups:

$$\operatorname{Stab}(v_0) = \langle j\rangle, \quad \operatorname{Stab}(v_1) = r\langle j\rangle r^{-1} = \langle r^2 j\rangle, \quad \operatorname{Stab}(v_2) = r^2\langle j\rangle r^{-2} = \langle r^4 j\rangle, \quad \dots$$

The intersection of these is the trivial subgroup, so D_n acts faithfully on V.

51.6. *Example.* Consider the action by a group G of symmetries of a regular polyhedron on its vertex set V. The stabilizer of any $v \in V$ will be a cyclic group, consisting of rotations around the axis through that vertex. The order of $\operatorname{Stab}(v)$ will depend on the type of polyhedron.

51.7. *Example.* Action of D_4 on $\mathbb{R}^3$. Here are some examples of stabilizers.

- $\operatorname{Stab}((0,0,0)) = D_4$.
- $\operatorname{Stab}((1,0,0)) = \langle j\rangle$.
- $\operatorname{Stab}((1,1,0)) = \langle rj\rangle$.
- $\operatorname{Stab}((2,1,0)) = \{e\}$.
- $\operatorname{Stab}((0,0,1)) = \langle r\rangle$.

52. Orbit-stabilizer theorem

52.1. **Theorem** (Orbit-stabilizer theorem)**.** *Let $G \curvearrowright X$ and $x \in X$. Let $H = \operatorname{Stab}(x)$. The function $\psi(gH) := gx$ defines a bijection $\psi\colon G/\operatorname{Stab}(x) \to \mathcal{O}(x)$ between the left cosets of H and the orbit containing x.*

Proof. Because ψ is defined on cosets, we must check it is well-defined. If $gH = g'H$, then $g' = gh$ for some $h \in H = \operatorname{Stab}(x)$, so $g'x = g(hx) = gx$. (Because $hx = x$.)

ψ is surjectve: if $y \in \mathcal{O}$, then $y = gx$ for some $g \in G$. Then $\psi(gH) = gx = y$.

ψ is injective: if $\psi(gH) = \psi(g'H)$, then $gx = g'x$, so $x = g^{-1}g'x$. Thus $g^{-1}g' \in H$, which means exactly $gH = g'H$. □

Note that the function ψ satisfies $g\,\psi(aH) = \psi(gaH)$. In other words, the function ψ is "compatible" with the actions of G on $G/\operatorname{Stab}(x)$ and on $\mathcal{O}(x)$.

52.2. **Corollary.** *We have that $|\mathcal{O}(x)| = [G : \operatorname{Stab}(x)]$. In particular, if G is a finite group, then $|\mathcal{O}(x)| = |G| / |\operatorname{Stab}(x)|$.*

53. Actions by a group on a group

There are many ways in which we can get an action on the underlying set of a group G, sometimes even by the group itself.

53.1. **Inversion action.** Let $G = \{e, a\}$ be the cyclic group of order 2. Let H be any group. Define an action by G on the set $X = H$ by $\phi\colon G \to \mathrm{Sym}(H)$, where

$$\rho(e)(h) := h, \qquad \rho(a)(h) := h^{-1}, \qquad h \in H.$$

We have seen this action before. The orbit of any $h \in H$ under this action is

$$\mathcal{O}(h) = \{h, h^{-1}\},$$

which can have 1 or 2 elements depending on whether $h^2 = e$.

53.2. **Left regular action.** We make a group G act on itself.

The **left regular action** $\lambda\colon G \to \mathrm{Sym}(G)$ is given by

left regular action

$$\lambda(g)(x) := gx \qquad \text{for } g, x \in G.$$

Check that this satisfies $\lambda(g_1)\big(\lambda(g_2)(x)\big) = \lambda(g_1g_2)(x)$:

$$\lambda(g_1)(\lambda(g_2)(x)) = \lambda(g_1)(g_2x) = g_1(g_2x) = (g_1g_2)x = \lambda(g_1g_2)(x).$$

The abbreviated notation for action causes no confusion: $gx = gx$ either way.

Exercise. This action has only one orbit (it is transitive), and $\mathrm{Stab}(g) = \{e\}$ for any $g \in G$.

53.3. **Restricted left regular action.** Now suppose $H \leq G$ is a subgroup. We can use the same formua to give an action by H on $X = G$, so $\lambda\colon H \to \mathrm{Sym}(G)$ is given by

$$\lambda(h)(x) := hx \qquad \text{for } h \in H,\ x \in G.$$

An orbit of this action has the form $\mathcal{O}(x) = Hx$. So the orbits are exactly the *right cosets.*

53.4. **Right regular action.** The **right regular action** $\rho\colon G \to \mathrm{Sym}(G)$ is given by

right regular action

$$\rho(g)(x) := xg^{-1} \qquad \text{for } g, x \in G.$$

53.5. *Exercise.* Check that this satisfies $\rho(g_1)(\rho(g_2)(x)) = \rho(g_1g_2)(x)$.

The abbreviated notation for this action would be really confusing: gx would be the same as xg^{-1}, so you should avoid using it.

Warning. The inverse is necessary.

The function $\tau\colon G \to \mathrm{Sym}(G)$ defined by $\tau(g)(x) := xg$ is *not* an action, unless G is an abelian group.

53.6. *Exercise.* Show that τ defined this way is an action if and only if G is an abelian group.

Let G be a group and $H \leq G$ be a subgroup. Then we can use the same formulas as left regular action to define a **restricted right regular action** action by H on G, by

restricted right regular tion

$$\lambda'(h)(x) := hx \qquad \text{for } h \in H,\ x \in G.$$

Exercise. The orbits for the restricted right regular action are exactly the left H-cosets.

53.7. **Left coset action.** Let $H \leq G$ be a subgroup. Then G acts on $X = G/H$ (left cosets) by $\lambda\colon G \to \mathrm{Sym}(G/H)$ defined by

$$\lambda(g)(xH) := gxH.$$

We need to check that this is well defined. (Why?)

This is compatible with the abbreviated notation for action: $g(xH) = gxH$.

Note, if $H = \{e\}$, then $G/\{e\}$ is in bijective correspondence with G, so the left coset action in this case is the same thing as the left regular action.

Exercise. Show that for this action, $\mathrm{Stab}(eH) = H$.

Exercise. Show that $\mathrm{Stab}(aH) = aHa^{-1}$.

54. Conjugation action

There is a much more complex action by G on itself, called **conjugation action**. **conjugation action**

Let $c\colon G \to \mathrm{Sym}(G)$ be the function defined by

$$c(g)(x) := gxg^{-1}.$$

Verify the identity $c(g_1)(c(g_2)(x)) = c(g_1g_2)(x)$:

$$c(g_1)(c(g_2)(x)) = c(g_1)(g_2xg_2^{-1}) = g_1(g_2xg_2^{-1})g_1^{-1} = (g_1g_2)x(g_1g_2)^{-1} = c(g_1g_2)(x).$$

Remark. The set of automorphisms $\mathrm{Aut}(G)$ is a subgroup of $\mathrm{Sym}(G)$, and the image of c is contained in $\mathrm{Aut}(G)$.

The conjugation action of G is almost never transitive.

54.1. *Exercise.* Show that the conjugation action by G on G is transitive if and only if G is a trivial group.

The orbits of the conjugation action are called **conjugacy classes**, and written **conjugacy classes**

$$\mathrm{Cl}(x) = \mathrm{Cl}_G(x) := \{\, gxg^{-1} \mid g \in G \,\}.$$

Conjugacy classes were discussed in Goodman §2.5.

54.2. *Example* (Conjugation action, abelian group). If G is an abelian group, the conjugacy classes are all singleton: $\mathrm{Cl}(x) = \{\, gxg^{-1} \mid g \in G \,\} = \{x\}$.

54.3. *Example* (Conjugation action, D_3). Let $G = D_3 = \{e, r, r^2, j, rj, r^2j\}$. Draw picture of the conjugation action.

The conjugacy classes are

$$\{e\}, \qquad \{r, r^2\}, \qquad \{j, rj, r^2j\}.$$

Note that $r^2 = jrj^{-1}$, while $rjr^{-1} = r^2j$ and $r(r^2j)r^{-1} = rj$.

54.4. *Example* (Conjugation action, D_4). Let $G = D_4 = \{e, r, r^2, r^3, j, rj, r^2j, r^3j\}$. The conjugacy classes are

$$\{e\}, \qquad \{r^2\}, \qquad \{r, r^3\}, \qquad \{j, r^2j\}, \qquad \{rj, r^3j\}.$$

The stabilizers of the the conjugation action are called **centralizers**, and are described by **centralizers**

$$\mathrm{Cent}(a) = \mathrm{Cent}_G(a) := \{\, g \in G \mid gag^{-1} = a \,\} = \{\, g \in G \mid ga = ag \,\}.$$

Exercise. $\mathrm{Cent}(a)$ is actually a subgroup of G, so we call it the *centralizer subgroup*.

The centralizer subgroup $\mathrm{Cent}(a)$ is the set of all elements that "commute with a". Note that this does *not* imply that $\mathrm{Cent}(a)$ is abelian: the elements of $\mathrm{Cent}(a)$ commute with a, but don't have to commute with each other.

The orbit-stabilizer gives us:

54.5. **Proposition.** *For any $x \in G$, we have*

$$|\mathrm{Cl}(x)| = [G : \mathrm{Cent}(x)].$$

This is most useful when G is finite, in which case, we learn that the size of any conjugacy class must divide $|G|$. Here is another useful and easy fact, which gives a "lower bound" for a centralizer subgroup.

54.6. **Proposition.** *For any $x \in G$, we have $\langle x \rangle \subseteq \mathrm{Cent}(x)$,*

Proof. Because $x \in \mathrm{Cent}(x)$ and $\mathrm{Cent}(x)$ is a subgroup. □

54.7. *Example.* The centralizer of $e \in G$ is $\mathrm{Cent}(e) = G$. The conjugacy class of e is $\mathrm{Cl}(e) = \{e\}$. **Lecture 28**

54.8. *Example.* More generally, if $g \in Z(G)$ is in the center of G, then $\mathrm{Cl}(g) = \{g\}$, and $\mathrm{Cent}(g) = G$. Conversely, $\mathrm{Cl}(g) = \{g\}$ or $\mathrm{Cent}(g) = G$ implies $g \in Z(G)$.

In particular, if G is an abelian group, then every conjugacy class in G has size 1.

54.9. *Example.* The centralizer of $\sigma = (12)(34)$ in S_4 must contain $\langle \sigma \rangle$ (order 2). Furthermore, we can easily check that $\mathrm{Cl}(\sigma)$ contains exactly three elements (all the products of two disjoint 2-cycles). Therefore $|\mathrm{Cent}(\sigma)| = |S_4| \,/\, |\mathrm{Cl}(\sigma)| = 24/3 = 8$. A little bit of trial and error finds the other elements of the centralizer.

$$\mathrm{Cent}(\sigma) = \{e, (12), (34), (12)(34), (13)(24), (14)(23), (1324), (1423)\}.$$

In general, when computing a centralizer group, remember the following facts.

- $\langle x \rangle \leq \mathrm{Cent}(x)$.
- $Z(G) \leq \mathrm{Cent}(x)$.
- $|\mathrm{Cent}(x)| = |G| \,/\, |\mathrm{Cl}(x)|$ if the group is finite.

55. Conjugacy classes in $SO(3)$

Let $G = SO(3)$ be the group of rotations of space around the origin. Remember that every element of G has the form $\mathrm{Rot}_u(\theta)$, i.e., a rotation by angle θ around the axis through the unit vector u. I think of this as counterclockwise when viewed from "above" the head of the vector u. (For example, $\mathrm{Rot}_{e_3}(\pi/2)$ sends e_1 to e_2.)

Also remember that elements get multiple names under this notation:

- $\mathrm{Rot}_u(0) = I$ for any unit vector u.
- $\mathrm{Rot}_u(\theta + 2\pi n) = \mathrm{Rot}_u(\theta)$ for any $n \in \mathbb{Z}$.
- $\mathrm{Rot}_{-u}(\theta) = \mathrm{Rot}_u(-\theta)$, i.e., a rotation seen as clockwise from one direction is seen as counter-clockwise as seen from the opposite direction.

Finally, recall the *conjugation formula* for rotations:

$$A \,\mathrm{Rot}_u(\theta) A^{-1} = \mathrm{Rot}_{Au}(\theta).$$

The following describes conjugacy classes and centralizers in $SO(3)$ for every element. Note that because of identities (2) and (3) above, we can always assume that $\theta \in [0, \pi]$.

55.1. **Proposition.** *Conjugacy classes and centralizers in $G = SO(3)$ have the following form. Let $P = \mathrm{Rot}_u(\theta)$ with $\theta \in [0, \pi]$.*

1. *If $P = I$ (i.e., $\theta = 0$), $\mathrm{Cl}(I) = \{I\}$ and $\mathrm{Cent}(I) = SO(3)$.*
2. *If $P = \mathrm{Rot}_u(\theta)$ is a rotation which is not the identity (i.e., $\theta \neq 0$), then*

$$\mathrm{Cl}(P) = \{\, \mathrm{Rot}_v(\theta) \mid v \in \mathbb{R}^3,\ \|v\| = 1 \,\},$$

the set of all rotations with angle θ, but with any unit vector v.

Note that if $\theta \notin \{0, \pi\}$, then $\mathrm{Cl}(P)$ contains two *rotations around the* same *axis as P: the rotation $P = \mathrm{Rot}_u(\theta)$ itself and its inverse $P^{-1} = \mathrm{Rot}_u(-\theta) = \mathrm{Rot}_{-u}(\theta)$. However, if $\theta = \pi$, then $P = P^{-1}$, so there is only one such rotation.*

(3) *If* $P = \mathrm{Rot}_u(\theta)$ *is a rotation which is not the identity and is not* 180° *(i.e.,* $\theta \notin \{0,\pi\}$*) then*
$$\mathrm{Cent}(P) = \{\, \mathrm{Rot}_u(\alpha) \mid \alpha \in \mathbb{R} \,\},$$
the subgroup of all rotations through the same axis, but with any angle.

(4) *If* $P = \mathrm{Rot}_u(\pi)$*, then*
$$\mathrm{Cent}(P) = \{\, \mathrm{Rot}_u(\alpha) \mid \alpha \in \mathbb{R} \,\} \cup \{\, \mathrm{Rot}_v(\pi) \mid u \cdot v = 0 \,\},$$
the subgroup consisting of (i) all rotations around the axis through u*, but any angle, and (ii) all* 180° *rotations around any axis perpendicular to* u*.*

55.2. *Remark.* For instance, $\mathrm{Cent}(\mathrm{Rot}_{e_3}(\pi)$ is exactly the group D of symmetries of the unit disk we described earlier.

Proof. Applying the conjugation formula tells us that for any $P = \mathrm{Rot}_u(\theta)$, we have
$$\mathrm{Cl}(\mathrm{Rot}_u(\theta)) = \{\, \mathrm{Rot}_{Au}(\theta) \mid A \in SO(3) \,\}, \qquad \mathrm{Cent}(\mathrm{Rot}_u(\theta)) = \{\, A \in SO(3) \mid \mathrm{Rot}_{Au}(\theta) = \mathrm{Rot}_u(\theta) \,\}.$$
When $P = I$, we already know what the answer is.

For $P = \mathrm{Rot}_u(\theta)$ not the identity, note that for any two unit vectors u, v, there is always some rotation A such that $Au = v$. This gives the formula for the set of conjugacy classes. The only rotations $\mathrm{Rot}_v(\theta)$ with the same axis of rotation as P are the ones such that u and v are parallel, i.e., $v = \pm u$. So P and $P^{-1} = \mathrm{Rot}_{-u}(\theta)$ are the only ones in $\mathrm{Cl}(P)$, and $P = P^{-1}$ iff $\theta = \pi$.

If $P = \mathrm{Rot}_u(\theta)$ with $\theta \in (0,\pi)$, then $\mathrm{Rot}_{Au}(\theta) = \mathrm{Rot}_u(\theta)$ only if $Au = u$, i.e., if A fixes the vector u. The only rotations which fix u are ones with axis of rotation through u, so this gives the centralizer in this case. (Note that $\mathrm{Rot}_{-u}(\theta) \neq \mathrm{Rot}_u(\theta)$ in this case, since θ is not 0 or π.)

If $P = \mathrm{Rot}_u(\pi)$, then $\mathrm{Rot}_{Au}(\pi) = \mathrm{Rot}_u(\theta)$ iff $Au = \pm u$ (because $\theta = \pi$ so $\mathrm{Rot}_{-u}(\pi) = \mathrm{Rot}_u(\pi)$). Thus $A \in \mathrm{Cent}(P)$ iff either

(1) $Au = u$, so A is a rotation around an axis through u, or
(2) $Au = -u$, so A is a rotation by angle π around axis perpendicular to u.

This gives the centralizer in this case.

□

We can use this to help us describe conjugacy classes in symmetry groups of polyhedra. But we have to be careful. If H is a subgroup of G, then two elements $h, h' \in H$ can be conjugate in G but fail to be conjugate in H.

55.3. *Example* (Conjugacy classes in the symmetry group of a cube). Let $G \leq SO(3)$ be the symmetry group of a cube. Recall our classification of these, which I have arbitrarily named C_1 through C_5:

	axis	angle	number of elements
C_1	vertex/vertex	$\pm 2\pi/3$	8
C_2	edge/edge	$2\pi/2$	6
C_3	face/face	$\pm 2\pi/4$	6
C_4	face/face	$2\pi/2$	3
C_5	*	0	1

Each of these sets C_i forms a conjugacy class in G. This may be a little surprising, because all elements of C_2 ad C_4 are conjugate to each other in $SO(3)$, since they are all rotations by angle π. But, for instance,
$$P = \mathrm{Rot}_u(\pi) \in C_2, \quad u = (e_1 + e_2)/\sqrt{2}, \qquad Q = \mathrm{Rot}_v(\pi) \in C_4, \quad v = e_3,$$
are not conjugate in G. To be conjugate, we would need an $A \in G$ such that $Au = v$ or $Au = -v$. But although there is such a rotation in $SO(3)$, there is none in G, since no symmetry of G can take a midpoint of an edge (like $e_1 + e_2$) to a center of a face (like e_3). (I am putting the vertices of the cube at $(\pm 1, \pm 1, \pm 1)$.)

Also note that any two rotations of the form $P = \mathrm{Rot}_u(2\pi/3)$ and $Q = \mathrm{Rot}_u(-2\pi/3) = \mathrm{Rot}_{-u}(2\pi/3)$ in C_1 are conjugate to each other in G, since in this case there is a rotation A which sends u to $-u$. For instance, if

$$P = \mathrm{Rot}_u(2\pi/3), \quad Q = \mathrm{Rot}_u(-2\pi/3), \qquad u = (e_1 + e_2 + e_3)/\sqrt{3},$$

then we can take $A = \mathrm{Rot}_v(\pi)$ where $v = (e_2 - e_3)/\sqrt{2}$, which is an element of G (in fact, in C_2). Since $u \cdot v = 0$, we have $Au = -u$.

Similarly, note that any two rotations of the form $P = \mathrm{Rot}_u(2\pi/4)$ and $Q = \mathrm{Rot}_u(-2\pi/4) = \mathrm{Rot}_{-u}(2\pi/4)$ in C_3 are conjugate to each other in G: for instance, if

$$P = \mathrm{Rot}_u(2\pi/4), \quad Q = \mathrm{Rot}_u(-2\pi/4), \qquad u = e_3,$$

then we can take $A = \mathrm{Rot}_v(\pi)$ where $v = (e_2 + e_3)/\sqrt{2}$, which is an element of G (in fact, in C_2). Or we could take $A = \mathrm{Rot}_w(\pi)$ with $w = e_2$, which is an element of C_4.

56. Conjugacy classes in permutation groups

Consider the k-cycle $g = (1, 2, \ldots, k)$ in S_n. Suppose $\sigma \in S_n$. Then

$$\sigma(1, 2, \ldots, k)\sigma^{-1} = (\sigma(1), \sigma(2), \ldots, \sigma(k)).$$

In other words, the conjugate of the k-cycle another k-cycle, with its entries replaced according to σ.

56.1. *Example.* $g = (1234)$, $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 5 & 1 & 4 & 3 \end{pmatrix}$. Then $\sigma g \sigma^{-1} = (2514)$.

Proof. Let $h = \sigma g \sigma^{-1}$. Given $x \in \{1, \ldots, n\}$, we have the following calculation:

$$h(\sigma(x)) = \sigma g \sigma^{-1}(\sigma(x)) = \sigma g(x).$$

Thus, if $x \in \{1, \ldots, k-1\}$, we have $h(\sigma(x)) = \sigma(x+1)$, while $h(\sigma(k)) = \sigma(1)$, so h cycles the list $\sigma(1), \sigma(2), \ldots, \sigma(k)$. On the other hand, if $x > k$, then $h(\sigma(x)) = \sigma(x)$.

More generally, if $g = c_1 \cdots c_r$ is a product of *disjoint* cycles, then

$$\sigma g \sigma^{-1} = (\sigma c_1 \sigma^{-1}) \cdots (\sigma c_r \sigma^{-1}).$$

For instance,

$$\sigma(123)(4567)(89)\sigma^{-1} = (\sigma(1), \sigma(2), \sigma(3))(\sigma(4), \sigma(5), \sigma(6), \sigma(7))(\sigma(8), \sigma(9)).$$

56.2. **Proposition.** *Two elements of S_n are conjugate if and only if they have the same cycle type.*

Notation for cycle types: The book uses the notation $5^3 2^4 1^6$ to mean: three 5-cycles, four 2-cycles, all pairwise disjoint, and also three elements not in any cycle. In this case the element must be in S_n with $n = 3 \times 5 + 4 \times 2 + 6 \times 1 = 15 + 8 + 6 = 29$. (Note: I have prefered to write this as $5 + 5 + 5 + 2 + 2 + 2 + 2 + 1 + 1 + 1 + 1 + 1 + 1$, because this reminds you that these numbers need to add up to n.)

56.3. *Example.* S_4 has five conjugacy classes.

- 1^4: $\{e\}$.
- $2^1 1^2$: $\{(12), (13), (14), (23), (24), (34)\}$.
- 2^2: $\{(12)(34), (13)(24), (14)(23)\}$.
- $3^1 1^1$: $\{(123), (132), (124), (142), (134), (143), (234), (243)\}$.
- 4^1: $\{(1234), (1243), (1324), (1342), (1423), (1432)\}$.

I'll compute a centralizer in each case of one representative of each conjugacy class.

- $\mathrm{Cent}(e) = S_4$.

- Since $|\mathrm{Cl}((12))| = 6$, we have $|\mathrm{Cent}((12))| = |S_4| \,/\, |\mathrm{Cl}((12))| = 4$. Any element of the centralizer must leave the set $\{1,2\}$ invariant, and therefore also the set $\{3,4\}$. In fact,
$$\mathrm{Cent}((12)) = \{e, (12), (34), (12)(34)\}.$$
If I chose a different 2-cycle, I may get a different centralizer: $\mathrm{Cent}((23)) = \{e, (23), (14), (14)(23)\}$.
- We have $|\mathrm{Cent}((12)(34))| = 24/3 = 8$. Elements of this centralizer can either (i) leave both sets $\{1,2\}$ and $\{3,4\}$ invariant, or (ii) switch the two sets with each other (which can happen in several ways). We have
$$\mathrm{Cent}((12)(34)) = \{e, (12), (34), (12)(34), (13)(24), (14)(23), (1324), (1423)\}.$$
- We have $|\mathrm{Cent}((123))| = 24/8 = 3$. Since the centralizer subgroup must contain the element it centralizes, we see that
$$\mathrm{Cent}((123)) = \{e, (123), (132)\}.$$
- We have $|\mathrm{Cent}((1234))| = 24/6 = 4$. Again, the centralizer must contain the element it centralizes, which in this case has order 4, so
$$\mathrm{Cent}((1234)) = \{e, (1234), (13)(24), (1432)\}.$$

57. Counting orbits

Lecture 29

Here is an example of a "necklace problem": *How many different necklaces can we make from two red beads and two white beads?*

The necklace is a closed loop, with four beads on it. The first idea is to draw the necklace with the four beads at the vertices of a square. This gives six possible arrangements. Let's write X for the set of such arrangements. If I list these as a string, where the first position corresponds to the bead on the positive x-axis, then the arrangements are:

1	2	3	4	5	6
RRWW	RWRW	RWWR	WRRW	WRWR	WWRR

There are $6 = \binom{4}{2}$ arrangemenst.

However, these arrangements do not all count as different necklaces, as we can rotate or flip one to get another one. Instead, we have an action by D_4 on the set of bead arrangements. When we take this into account, there are really only two arrangements: these correspond to the two orbits of the action.

For instance, arrangements 1,4,6 are really the same, since we can just rotate the necklace to get from one to another. We can also get from one to another of these by flips. But you can't get 2,3,5 from 1,4,6 by these transformations. So there are really two types:

(1) Alternating color beads (2,3,5).
(2) Same colors adjacent (1,4,6).

57.1. *Example* (Necklace problem). *How many different necklaces can we make from four red beads, three white beads, two yellow beads?*

We can represent an arrangement of beads on a string by a list, like $WRRYWWRYR$. There are $9!/(4!\,3!\,2!) = 1260$ such arrangements. We should picture these as labels on the vertices of a regular 9-gon.

Because we can slide the beads around, this is the same as any cyclic permutation of the list. We can also turn the necklace upside down, which corresponds to a "flip". Thus, there is an action of D_9 on the set X of lists of size 1260. The number of necklaces is the number of orbits of this action.

We are going to compute this using a formula for counting orbits of a group action. In this example, the group is the dihedral group $G = D_9$. The X is the set of ways to arrange the beads on the loop, where the necklace is never moved: equvialently, this is the set of *linear* arrangements of the 9 beads. In the example, $|X| = 1260$.

57.2. **Burnside's formula.** We give a formula for counting orbits. Suppose $G \curvearrowright X$ where both G and X are finite, and consider $g \in G$. Let

$$\mathrm{Fix}(g) := \{\, x \in X \mid gx = x \,\}.$$

This is the set of elements in X which are *fixed* by the element g. It is a subset of X.

57.3. **Theorem** (Burnside lemma)**.** *The number of orbits of the action is*

$$\frac{1}{|G|}\sum_{g\in G}|\mathrm{Fix}(g)|\,.$$

Proof. This a going to be a "combinatorial proof", where we count a set F in several different ways. Consider

$$F = \{\, (g,x) \in G \times X \mid gx = x \,\},$$

the set of pairs of g and x so that $gx = x$.

We can add this up as either

$$|F| = \sum_{x\in X}|\{\, g\in G \mid gx = x \,\}| = \sum_{x\in X}|\mathrm{Stab}(x)|\,,$$

or

$$|F| = \sum_{g\in G}|\{\, x\in X \mid gx = x \,\}| = \sum_{g\in G}|\mathrm{Fix}(g)|\,.$$

Dividing by $|G|$ gives the identity

$$\frac{1}{|G|}\sum_{g\in G}|\mathrm{Fix}(g)| = \frac{|F|}{|G|} = \sum_{x\in X}\frac{|\mathrm{Stab}(x)|}{|G|} = \sum_{x\in X}\frac{1}{|\mathcal{O}(x)|},$$

using the fact that $|\mathcal{O}(x)| = [G : \mathrm{Stab}(x)]$. Now we remember that X is a dijoint union of its orbits:

$$\sum_{x\in X}\frac{1}{|\mathcal{O}(x)|} = \sum_{\mathcal{O}}\sum_{x\in\mathcal{O}}\frac{1}{|\mathcal{O}(x)|} = \sum_{\mathcal{O}} 1 = \#\text{orbits}.$$

□

58. Application to necklace problem

We return to the problem: necklaces with 4 red, 3 white, 2 yellow beads; 9 beads total. We decided this was the same as the number of orbits of $D_9 \curvearrowright X$, where X is the set of all *linear* arrangements of beads: $|X| = 9!/(4!3!2!) = 1260$.

By the Burnside lemma, we need to count, for each $g \in D_9$, the sets $\mathrm{Fix}(g)$. An arrangement is fixed by g iff every bead in each component of the cycle decomposition of g consists of beads of one color.

Here are the basic cases.

- e. Every necklace is fixed by e, so $|Fix(e)| = |X| = 1260$.
- r. This permutes the beads cyclically: $(123\ldots9)$. A necklace is fixed by r iff every bead has the same color. So in our example, $\mathrm{Fix}(r) = \varnothing$. Every rotation of order 9 has the same property, so this accounts for $r, r^2, r^4, r^5, r^7, r^8$.
- r^3. As a permutation, this is $(147)(258)(369)$. There is no way to arrange beads to that each cycle has a single color (since 4 and 2 are not $\equiv 0 \mod 3$.) For instance, for an arrangement in $\mathrm{Fix}(r^3)$, beads 1,4,7 must have the same color, beads 2,5,8 must have the same color, and beads 3,6,9 must have the same color. This is impossible since we want some colors to appear exactly two or four times. Thus $\mathrm{Fix}(r^3) = \varnothing$. This accounts for r^3, r^6.

- j. As a permutation, this fixes exactly one bead and transposes four pairs: $(1)(29)(38)(47)(56)$. We have an odd number of W, so let bead 1 be W. Pick one of the four pairs to consist of the two other W beads; then pick another of the pairs to consist of the two Y beads. The rest (four) will be R. The number of such choices is $4 \times 3 = 12$, or
$$\frac{4!}{2!}1!1! = 12.$$
Thus $|\mathrm{Fix}(j)| = 12$.

  The same argument works for all the flips $(j, rj, \dots, r^8 j)$, whose associated permutations all have the same type of cycle decompositions.

Thus,
$$\begin{aligned}\#\text{orbits} &= \frac{1}{|G|}\sum_g |\mathrm{Fix}(g)| \\ &= \frac{1}{|G|}\Big(\sum_e 1260 + \sum_{r,r^2,r^4,r^5,r^7,r^8} 0 + \sum_{r^3,r^6} 0 + \sum_{a,ra,\dots,r^8a} 12\Big) \\ &= \frac{1}{18}\big(1260 + 9(12)\big) = 76.\end{aligned}$$

Note: there are two kinds of actions showing up in this problem:

- The action by $G = D_n$ on $X =$ the set of linear arrangements of beads of a given color.
- The action by a cyclic subgroup $\langle g\rangle \le G$ on the set of "vertices" $1, 2, \dots, n$ (corresponding to locations of beads).

The set $\mathrm{Fix}(g)$ is a subset of X. But to count $|\mathrm{Fix}(g)|$, we need to think about the action of g on the set of vertices.

Another problem. Now consider a necklace of nine beads, with three colors, but no restriction on how many of each color are used. Now the set X of linear arrangements has $|X| = 3^9$.

An element of X is fixed by $g \in D_9$ exactly if each cycle of g consists of elements of a uniform color. If we write $N(g)$ for the number of orbits of the action $\langle g\rangle \curvearrowright \{1, 2, \dots, 9\}$, then $|\mathrm{Fix}(g)| = 3^{N(g)}$ (choose a color for each cycle of g).

elements of D_9	$N(g)$	$3^{N(g)}$
$\mathrm{Cl}(e)$	9	3^9
$\mathrm{Cl}(r), \mathrm{Cl}(r^2), \mathrm{Cl}(r^4), \mathrm{Cl}(r^5), \mathrm{Cl}(r^7), \mathrm{Cl}(r^8)$	1	3
$\mathrm{Cl}(r^3), \mathrm{Cl}(r^6)$	3	3^3
$\mathrm{Cl}(a), \mathrm{Cl}(ra), \dots, \mathrm{Cl}(r^8a)$	5	3^5

Thus,
$$\#\text{orbits} = \frac{1}{|G|}\sum_{g\in G}|\mathrm{Fix}(g)| = \frac{1}{18}\left(1\cdot 3^9 + 6\cdot 3 + 2\cdot 3^3 + 9\cdot 3^5\right) = 1219.$$

58.1. **The general formula for Polya enumeration.** Suppose G acts on a finite set X via $\phi\colon G \to \mathrm{Sym}(X)$, and let $m \ge 1$. For each $g \in G$, write $N(g)$ for the number of cycles in the cycle decomposition of the permutation $N(g)$

The set of m-**colorings** of X is the set F of functions $f\colon X \to \{1, \dots, m\}$. There is an induced action $\psi\colon G \to \mathrm{Sym}(F)$, defined by

m-**colorings**

$$\psi(g)(f)(x) = f(g^{-1}x), \qquad g \in G, \quad f \in F, \quad x \in X.$$

Let's verify this is an action: $\psi(e)(f)(x) = f(e^{-1}x) = f(x)$ so $\psi(e)(f) = f$, and
$$\psi(g_1)\big(\psi(g_2)(f)\big)(x) = \psi(g_2)(f)(g_1^{-1}x) = f(g_2^{-1}g_1^{-1}x) = f((g_1g_2)^{-1}x) = \psi(g_1g_2)(f)(x).$$

58.2. **Proposition.** *The number of orbits of the action by G on F is*

$$\frac{1}{|G|}\sum_{g\in G} m^{N(g)}.$$

59. Finite subgroups of $SO(3)$

I am going to sketch the classification of finite subgroups of $SO(3)$. Note that if $H \leq SO(3)$ is a finite subgroup, so is any conjugate gHg^{-1} for any $g \in SO(3)$. Thus, we will really classify finite subgroups *up to conjugacy.* **Lecture 30**

59.1. **Theorem.** *Every finite subgroup of $SO(3)$ is conjugate to exactly one of the following list.*

1. *The trivial group.*
2. *The group Z_m of order $m \geq 2$, as a cyclic subgroup of rotations around one axis through angles which are multiples of $2\pi/m$.*
3. *The group D_m of order $2m$ with $m \geq 2$, as the group of rotational symmetries of a regular m-gon. (For $m = 2$ see below.)*
4. *The rotational symmetry group of a tetrahedron (isomorphic to A_4).*
5. *The rotational symmetry group of a cube/octahedron (isomorphic to S_4).*
6. *The rotational symmetry group of a dodecahedron/icosahedron (isomorphic to A_5).*

Note that D_2 is not really a dihedral group, but is a Klein 4-group, viewed as the symmetries of a "brick", or a rectangle. For instance, one copy of D_2 in $SO(3)$ looks like

$$D_2 = \{I,\ \mathrm{Rot}_{e_1}(\pi),\ \mathrm{Rot}_{e_2}(\pi),\ \mathrm{Rot}_{e_3}(\pi)\}.$$

I'm not going to prove this completely, but I will do the main part. The idea is to use the action by a finite subgroup $G \leq SO(3)$ on its set of *pole.*

Given a subgroup $G \leq SO(3)$, a **pole** is a unit vector $u \in \mathbb{R}^3$, $\|u\| = 1$, such that there exists $A \in G$ with $A \neq I$ and $Au = u$. I'll write X for the set of poles, so **pole**

$$X = \{\, x \in \mathbb{R}^3 \mid \|x\| = 1,\ |\mathrm{Stab}_G(x)| \geq 2 \,\}.$$

Two observations:

- The poles of G are the unit vectors which lie on the axis of rotation of some non-identity element in G. This means a finite G has only finitely many poles: in fact, $|X| \leq 2(|G| - 1)$.
- If x is a pole and $g \in G$, then gx is also a pole: this is because $\mathrm{Stab}_G(gx) = g\,\mathrm{Stab}_G(x)g^{-1}$, so if a non-identity element a stabilizes x, then the non-identity element gag^{-1} stabilizers x. Thus we have an action by G on X. The classification will come from studying this action.

For each of the groups in our list, we can identify the poles, and in fact describe all the orbits of the G action on the set of poles. In fact, I'll make a table, where $n = |G|$, $k =$ number of orbits in X, $\mathcal{O}_1, \dots, \mathcal{O}_k$ are the orbits, and $c_i = |\mathrm{Stab}(x)|$ for any $x \in \mathcal{O}_i$, so $|O_i| = n/c_i$. I'm going to list the orbits in order of decreasing size, and thus the c_i in increasing order.

group (up to iso)	$n = \lvert G\rvert$	$k = \#$ of orbits	$\lvert O_1\rvert \geq \cdots \geq \lvert O_k\rvert$	$c_1 \leq \cdots \leq c_k$
{e}	1	0		
$\mathbb{Z}_n,\ n \geq 2$	n	2	$1, 1$	n, n
$D_m,\ m \geq 2$	$2m$	3	$m, m, 2$	$2, 2, m$
A_4	12	3	$6, 4, 4$	$2, 3, 3$
S_4	24	3	$12, 8, 6$	$2, 3, 4$
A_5	60	3	$30, 20, 12$	$2, 3, 5$

Here is a description of these.

- The trivial subgroup has no poles.
- For $\mathbb{Z}_m$, there are only two poles, which lie on the axis of rotation.

- For D_m, there are $2+2m$ poles: 2 for the rotations r^k of the m-gon, and $2m$ for each of the m "flips" $r^k j$ of the m-gon.
- For symmetries of regular polyhedra, the poles are in the direction of: vertices, midpoints of edges, centroids of faces. So in these cases there are: $4+6+4=14$, $8+12+6=26$, or $30+20+12=62$ poles.

Let's apply Burnside's formula to the action by G on X. The formula says that

$$k = \frac{1}{n}\sum_{g\in G} |\mathrm{Fix}(g)|\,.$$

We see that:

- If $g=e$, then $\mathrm{Fix}(e)=X$, so $|\mathrm{Fix}(e)| = |X| = |\mathcal{O}_1| + \cdots + |\mathcal{O}_k|$.
- If $g \neq e$, then $\mathrm{Fix}(g) = \{\pm u\}$, where u lies on the axis of rotation of g, i.e., $g = \mathrm{Rot}_u(\theta)$. So $|\mathrm{Fix}(g)| = 2$.

So Burnside's formula simplifies to

$$k = \frac{1}{n}\Big[|\mathcal{O}_1| + \cdots + |\mathcal{O}_k| + 2(n-1)\Big]$$

Using that $|\mathcal{O}_i| = n/c_i$, we rewrite this as the **key equation**:

key equation

$$k = \frac{1}{c_1} + \cdots + \frac{1}{c_k} + 2 - \frac{2}{n}.$$

Solutions to this equation are integers, which also satisfy the constraints:

$$k \geq 0, \qquad n \geq 1, \qquad 2 \leq c_1 \leq \cdots \leq c_k, \qquad c_i \mid n.$$

Remember that $c_i = |\mathrm{Stab}(x)|$ for some $x \in \mathcal{O}_i \subseteq X$, and since x is a pole $c_i \geq 2$.

The key equation and the constraints don't involve anything about groups or polyhedra: it is just an equation about numbers which we can try to solve.

59.2. **Proposition.** *The only solutions to the key equation subject to the constraints are the ones in our table.*

Proof. Possible values of k. First I'll show that a solution must have $k \in \{0,2,3\}$. Rwrite the equation as

$$\big(1-\frac{1}{c_1}\big) + \cdots + \big(1-\frac{1}{c_k}\big) = 2 - \frac{2}{n}.$$

The RHS is in the set $\{0, 1, \frac{3}{2}, \frac{5}{3}, \frac{7}{4}, \dots\} \subset \{0\}\cup[1,2)$, since $n \geq 1$. Each $1-\frac{1}{c}$ with $c \geq 2$ is in the set $\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots\} \subset [\frac{1}{2}, 1)$.

In particular, we cannot have $k=1$, since in $1-\frac{1}{c_1} = 2-\frac{2}{n}$ the LHS is in $[\frac{1}{2},1)$, and the RHS cannot have any of these values.

Furthermore, we cannot have $k \geq 4$, since in that case the LHS $\geq \frac{k}{2} \geq 2$, and the RHS is < 2.

Case of $k=0$. In this case, the key equation becomes

$$0 = 2 - \frac{2}{n} \qquad \Rightarrow n = 1.$$

This is the first line on our table.

Case of $k=2$. In this case, the key equation reduces to

$$\frac{2}{n} = \frac{1}{c_1} + \frac{1}{c_2}, \qquad n \geq 1, \quad 2 \leq c_1 \leq c_2, \quad c_i \mid n.$$

Multiply through by n to get $2 = \frac{n}{c_1} + \frac{n}{c_2}$, and remember the constraint $c_i \mid n$. Since $n > 0$ we must have $\frac{n}{c_1} = \frac{n}{c_2} = 1$, so $n = c_1 = c_2 \geq 2$. This is exactly the second line of our table.

Case of $k = 3$, possible values of c_1. In this case, the key equation reduces to

$$1 + \frac{2}{n} = \frac{1}{c_1} + \frac{1}{c_2} + \frac{1}{c_3}, \qquad n \geq 1, \quad 2 \leq c_1 \leq c_2 \leq c_3.$$

I claim that $c_1 = 2$. If not, then $3 \leq c_1 \leq c_2 \leq c_3$, so $\frac{1}{3} \geq \frac{1}{c_1} \geq \frac{1}{c_2} \geq \frac{1}{c_3}$, which implies that the RHS ≤ 1. But the LHS is clearly > 1, so this is impossible.

Case of $k = 3$, possible values of c_2. We have reduced the key equation to

$$1 + \frac{2}{n} = \frac{1}{2} + \frac{1}{c_2} + \frac{1}{c_3}, \qquad n \geq 1, \quad 2 \leq c_2 \leq c_3.$$

I claim that $c_2 \in \{2, 3\}$. If not, then $4 \leq c_2 \leq c_3$, so $\frac{1}{4} \geq \frac{1}{c_2} \geq \frac{1}{c_3}$, which implies that the RHS ≤ 1, which is again impossible since the LHS > 1.

Case of $k = 3$, $c_1 = c_2 = 2$. We have reduced the key equation to

$$1 + \frac{2}{n} = \frac{1}{2} + \frac{1}{2} + \frac{1}{c_3}, \qquad n \geq 1, \quad 2 \leq c_3.$$

We can solve this to get $n = 2c_3$ with $c_3 \geq 2$. This is exactly the third line of our table.

Case of $k = 3$, $c_1 = 2$, $c_2 = 3$. We have reduced the key equation to

$$1 + \frac{2}{n} = \frac{1}{2} + \frac{1}{3} + \frac{1}{c_3}, \qquad n \geq 1, \quad 3 \leq c_3.$$

I claim that $c_3 \in \{3, 4, 5\}$. If not, then $c_3 \geq 6$, but then the RHS ≤ 1, which is impossible. For each of the three cases, we can use the formula to compute n, and we get the last three lines of our table: we have $n = 2/(\frac{1}{c_3} - \frac{1}{6})$, so $c_3 = 3$ gives $n = 12$, $c_3 = 4$ gives $n = 24$, and $c_3 = 5$ gives $n = 60$. □

These cases are exactly compatible with the symmetry subsgroups of $SO(3)$ that we have already listed, so all of these possibilities are realized by finite subgroups of $SO(3)$.

So far we have:

- Solved for all possible data $(n, k, c_1, \ldots, c_k)$ satisfying the key equation.
- For each solution $(n, k, c_1, \ldots, c_k)$, we have identified at least one finite subgroup $G \leq SO(3)$ which has this data (i.e., $|G| = n$, X has k orbits, with sizes $n/c_k, \ldots, n/c_k$).

It remains to show that these subgroups are the *only* possibilities. That is, we need to show:

- Given any finite subgroup $G \leq SO(3)$ with data $(n, k, c_1, \ldots, c_k)$, show that it conjugate to the ones we have already identified.

Here is how we do this: given a finite $G \leq SO(3)$, we identify a subset $V \leq \mathbb{R}^3$ such that G is a group of symmetries of V, so that this V is either (i) a pair of opposite unit vectors, (ii) a regular n-gon centered at the origin, or (iii) a regular polyhedron centered at the origin. Then G vill be a subgroup of the symmetry group G' of V, and we can use this to identify G with one of the groups on our original list.

I'm not going to work carefully through every case. It's entertaining to work through it, so I'll most of the details to you if you are interested. **I won't go through the following in class very much.**

First a note: poles come in antipodal pairs $\{\pm x\}$. The two opposite poles do *not* need to be in the same orbit of G acting on X. For instance, this is so in the cases (2) (cyclic group), (3) (dihedral group) when $4 \nmid n$, and (4) (tetrahedron group).

However, if they are in different orbits, those two orbits must have the same size (since $\text{Stab}(x) = \text{Stab}(-x)$).

(Case 1) $n = 1$, $k = 0$. This is the trivial subgroup.

(Case 2) $n \geq 2$, $k = 2$, $(c_1, c_2) = (n, n)$. We have $|X| = 1 + 1 = 2$, so $X = \{\pm x\}$, and all elements of G are rotations about this one axis. Thus G must be a cyclic group, consisting of rotations about a single axis.

(Case 3) $n = 2m$, $m \geq 2$, $k = 3$, $(c_1, c_2, c_3) = (2, 2, m)$. First suppose $m \neq 2$. Then $\mathcal{O}_3$ is the only orbit of size 2, so $\mathcal{O}_3 = \{\pm x\}$ for some x. Then $\mathrm{Stab}(x) = \langle r \rangle$ is cyclic of order m, where $r = \mathrm{Rot}_x(2\pi/m)$.

For any $g \in G \smallsetminus \mathrm{Stab}(x)$ we have $g(x) \in \mathcal{O}_3$ but $g(x) \neq x$, so $g(x) = -x$. So every $g \notin \mathrm{Stab}(x)$ is an angle π-rotation around an axis perpendicular to $\pm x$. The poles of g must lie in either $\mathcal{O}_1$ or $\mathcal{O}_2$, and must lie in the plane perpendicular to $\pm x$.

No non-identity element of $\mathrm{Stab}(x) = \langle r \rangle$ can fix any element of either $\mathcal{O}_1$ or $\mathcal{O}_2$. Thus, if $y \in \mathcal{O}_1$ then $y, r(y), \dots, r^{m-1}(y)$ are *distinct* elements of $\mathcal{O}_1$, so $\mathcal{O}_1 = \{y, r(y), r^2(y), \dots, r^{m-1}(y)\}$ are the points of a regular m-gon lying in the plane perpendicular to x, and similarly for $\mathcal{O}_2$. Thus G is the symmetry group of either of these regular m-gons, i.e., $G \approx D_m$.

When $m = 2$, you need to argue a little differently, since in this case $c_1 = c_2 = c_3 = 2$. Since the c_is are all 2, all elements of G have order at most 2, so $G \approx \mathbb{Z}_2 \times \mathbb{Z}_2$. Write $G \smallsetminus \{e\} = \{r, s, t\}$. Each of r, s, t is a rotation through angle π. Since they pairwise commute, the axes of rotation of any two must be perpendicular. Thus, the group is D_2 = Klein 4-group.

(Case 4) $n = 12$, $k = 3$, $(c_1, c_2, c_3) = (2, 3, 3)$. Consider $\mathcal{O}_3$ which has size 4 (the same argument will work with $\mathcal{O}_2$). Pick an $x \in \mathcal{O}_3$, so $\mathrm{Stab}(x) = \langle r \rangle$, a cyclic subgroup of order 3 with $r = \mathrm{Rot}_x(2\pi/3)$. Because $\mathcal{O}_3$ is a G-orbit, the subgroup $\langle r \rangle$ acts on $\mathcal{O}_3$. Since $o(r) = 3$, the orbits of $\langle r \rangle$ have size either 1 or 3. We know that $r(x) = x$, and that there are only two elements of X fixed by r, namely $\pm x$. Thus we must have

$$O_3 = \{x, y, r(y), r^2(y)\},$$

where y is any other element of $\mathcal{O}_3$. (And note that $-x \notin \mathcal{O}_3$ since $r(-x) = -x$.)

Because $r \in SO(3)$, it preserves dot product of vectors, so it preserves distances, so

$$\|x - y\| = \|x - r(y)\| = \left\|x - r^2(y)\right\|.$$

Thus, x is equidistant to the other three points in $\mathcal{O}_3$. But x was an *arbitrary* element of $\mathcal{O}_3$, so this argument shows that all points in $\mathcal{O}_3$ are equidistant, i.e., it is the set of vertices of a regular tetrahedron, and G acts as symmetries of it. (See exercise G 11.3.5.)

(Exercise: show that $x \cdot y = -1/3$.)

(Case 5) $n = 24$, $k = 3$, $(c_1, c_2, c_3) = (2, 3, 4)$. Here we can show that $\mathcal{O}_2$ are the vertices of a regular octahedron, and $\mathcal{O}_3$ are the vertices of a cube. (See exercise G 11.3.6.)

(Case 6) $n = 60$, $k = 3$, $(c_1, c_2, c_3) = (2, 3, 5)$. Here we can show that $\mathcal{O}_2$ are the vertices of a regular dodecahedron, and $\mathcal{O}_3$ are the vertices of a regular icosahedron. (See exercise G 11.3.7.)

60. The p-group fixed point theorem

Let G act on a set X. Write **Lecture 31**

$$X^G := \{\, x \in X \mid gx = x \text{ for all } g \in G \,\}.$$

This is called the set of **fixed points** of the action. **fixed points**

Note: this looks sort of like our definiton of $\mathrm{Fix}(g)$, except that here we want the set of $x \in X$ fixed by *every* element of the group. So we have

$$X^G = \bigcap_{g \in G} \mathrm{Fix}(g) = \{\, x \in X \mid \mathrm{Stab}(x) = G \,\} = \{\, x \in X \mid |\mathcal{O}(x)| = 1 \,\}$$

If $G \curvearrowright X$ is a group action, then G decomposes the set X into disjoint orbits. Then we can count the elements of X by adding up the orbits separately:

$$|X| = \sum_{\text{orbits } \mathcal{O}_i \subseteq X} |\mathcal{O}_i|$$

where the sum is over distinct *orbits*. We might call this the **orbit equation**. Note there is only one term for each orbit, not for each element of X. **orbit equation**

We also have the formula that counts the size of an orbit in terms of a stabilizer group:

$$|O(x)| = [G : \mathrm{Stab}(x)].$$

This gives

$$|X| = \sum_{\text{orbits } \mathcal{O}_i \subseteq X} [G : \mathrm{Stab}(x_i)], \qquad x_i \in \mathcal{O}_i.$$

Here, x_i is *any* choice of element of the orbit $\mathcal{O}_i$.

As we have seen, the orbits of size 1 correspond to elements of the fixed set X^G. So we can break these off into a separate term in the orbit equation:

$$|X| = \left|X^G\right| + \sum_{\substack{\mathcal{O}(x)\subset X\\ |\mathcal{O}(x)|\geq 2}} [G : \mathrm{Stab}(x)],$$

where the sum is over orbits which are not singleton.

Let p be a *prime* number. A ***p*-group** is a finite group G such that $|G| = p^n$ for some $n \geq 1$. ***p*-group**

60.1. **Theorem.** *If G is a p-group acting on a finite set X, then $\left|X^G\right| \equiv |X| \mod p$.*

Proof. By Lagrange, all subgroups have order dividing $|G| = p^n$, and thus have index dividing p^n. If $[G : \mathrm{Stab}(x)] \neq 1$, then p divides $[G : \mathrm{Stab}(x)] = |\mathcal{O}(x)|$. Thus, the term $\sum_{|\mathcal{O}|\geq 2} |\mathcal{O}| \equiv 0 \mod p$, so $|X| \equiv \left|X^G\right| \mod p$. □

60.2. *Example.* Let $G = \langle\phi\rangle$ with $o(\phi) = p$, a cyclic group of order p. Then the orbits of an action $G \curvearrowright X$ have size either 1 or p.

As an application, we can show the following theorem.

60.3. **Proposition.** *Every finite p-group has a non-trivial center.*

Proof. Suppose $|G| = p^n$, $n \geq 1$. Apply the above to the conjugation action by G on $X = G$. The fixed set of the action is

$$X^G = \{\, x \in X \mid gxg^{-1} = x \text{ for all } g \in G \,\} = \{\, x \in G \mid gx = xg \text{ for all } g \in G \,\},$$

which is exactly the center $Z(G) \subseteq G$. So $|Z(G)| \equiv |G| \equiv 0 \pmod p$. But $Z(G)$ is a subgroup, so $e \in Z(G)$, so $|Z(G)| \geq p$. □

60.4. **Corollary.** *Let p be a prime. Then every group of order p^2 is abelian.*

Proof. Suppose $|G| = p^2$. Since G is a p-group, the center is non-trivial, so $|Z(G)| \in \{p, p^2\}$. If $|Z(G)| = p^2$ then $G = Z(G)$ and so G is abelian.

If instead $|Z(G)| = p$, then since $Z(G)$ is a normal subgroup, the quotient group $G/Z(G)$ has prime order p, so is cyclic. But this was the subject of a homework assigment: if $G/Z(G)$ is cyclic, then $G = Z(G)$ and so G is abelian. □

61. Cauchy's theorem

This is in §5.4 in the book. Assume G is a finite group.

61.1. **Theorem** (Cauchy's theorem)**.** *If a prime p divides $|G|$, then G contains an element of order p.*

The case of $p = 2$ was the "even-order theorem" which we proved earlier. We can think of the proof as an application of a group actions. The idea was to let a cyclic group $C = \langle \phi \rangle \simeq \mathbb{Z}_2$ of order 2 act on the set $X = G$ of a group, so that the generator ϕ acts by the rule $\phi(g) := g^{-1}$. Since $|C| = 2$, the orbits of this action have size 1 or 2, and look like either

$$\{g\} \qquad \text{if } g^2 = e \text{ so } g = g^{-1},$$

or

$$\{g, g^{-1}\} \qquad \text{if } g^2 \neq e \text{ so } g \neq g^{-1}.$$

Thus $\left|X^C\right| \equiv |X| \pmod 2$. Since $|X| = |G|$ is even, this means $\left|X^C\right|$ is even.

But we know there is at least one element in X^C, namely e. So there are at least two elements in X^C, so at least one element of order 2.

We now give a proof of Cauchy's theorem which generalizes this, using an action by a cyclic group $C = \langle \phi \rangle \simeq \mathbb{Z}_p$ of prime order p on a suitable set X, and using the fact that $\left|X^C\right| \equiv |X| \pmod p$.

Proof of Cauchy's theorem (due to McKay). Consider the set

$$X = \{\, (a_1, a_2, \dots, a_p) \in G^p \mid a_1 a_2 \cdots a_p = e \,\}$$

the set of p-tuples of elements which multiply (in the given order) to the identity element. Note that $a_p = (a_1 a_2 \cdots a_{p-1})^{-1}$. Thus, if we choose $a_1, \dots, a_{p-1}$ arbitrarily, there is a unique a_p giving an element of X. Thus $|X| = |G|^{p-1}$, which is divisible by p.

Define a function $\phi \colon X \to X$ by

$$\phi(a_1, \dots, a_p) := (a_p, a_1, \dots, a_{p-1}).$$

We need to check that the output of ϕ is actually contained in X:

$$a_p a_1 \cdots a_{p-1} = a_p (a_1 \dots a_{p-1}) a_p a_p^{-1} = a_p e a_p^{-1} = e.$$

Clearly, iterating the function ϕ p times is the identity map on X. Thus, we have defined an action of the cyclic group $C = \langle \phi \rangle \approx \mathbb{Z}_p$ of order p on the set X.

What are elements of the fixed set X^C? These are $(a_1, \dots, a_p) \in X$ such that $a_1 \cdots a_p = e$ and

$$\phi(a_1, \dots, a_p) = (a_p, a_1, \dots, a_{p-1}) \quad \text{is equal to} \quad (a_1, \dots, a_p).$$

But this happens only if $a_1 = \cdots = a_p$. So there is a bijective correspondence $a \mapsto (a, \dots, a)$ of the form

$$\{\, a \in G \mid a^p = e \,\} \xrightarrow{\sim} X^C.$$

So elements of X^C corresponds to elements of order 1 or p in G.

Now we use our congruence:

$$\left|X^C\right| \equiv |X| \mod p.$$

Since p divides $|X|$, we have that p divides $\left|X^C\right|$. However, $(e, \dots, e) \in X^C$ so $\left|X^C\right| \geq p$. □

61.2. *Remark.* The proof of Cauchy's theorem actually shows a little more: for any finite group G and prime p, the subset $\{\, g \in G \mid g^p = e \,\}$ has size divisible by p. This can be useful to know sometimes.

61.3. *Example* (Classify groups of order 6). If $|G|$ has order 6, then G must have an element a of order 2, and an element b of order 3, by Cauchy. The cyclic subgroup $N := \langle b \rangle$ has index 2, so must be normal. If we let $A := \langle a \rangle$, then we have $N \cap A = \{e\}$. We can deduce from this that $NA = G$. (For instance, NA must be a subgroup so Lagrange applies, so $|N|$ and $|A|$ both divide $|NA|$, so $|NA| = 6 = |G|$.

Thus $G = N \rtimes_\gamma A$ for some $\gamma \colon A \to \mathrm{Aut}(N)$, a semi-direct product. We know $\mathrm{Aut}(N) \simeq \Phi(3)$, a group of order two. So there are only two possibilities for γ:

- $\gamma(a) = \mathrm{id}$, or
- $\gamma(a)$ sends $b \mapsto b^{-1}$.

The first case is $\mathbb{Z}_3 \times \mathbb{Z}_2 \simeq \mathbb{Z}_6$, the second case is $D_3 \simeq S_3$.

61.4. *Example* (Classify groups of order $2p$). The same idea works if $|G| = 2p$ where p is an *odd prime.* By Cauchy's theorem, we have subgroups

$$N = \langle b\rangle, \quad A = \langle a\rangle, \qquad o(b) = p, \quad o(a) = 2.$$

Since $[G : N] = 2$, N is a normal subgroup. Clearly $N \cap A = \{e\}$, and since NA is a subgroup we must have $G = NA$. So $G = N \rtimes_\gamma A$ for some $\gamma \to \mathrm{Aut}(N)$.

Recall that $\mathrm{Aut}(N) \simeq \Phi(p)$. On a problem set, you showed that for p an odd prime there are only two elements $a \in \Phi(p)$ such that $a^2 = e$, namely $[1]$ and $[-1]$. Thus there are only two possibilities for γ:

- $\gamma(a) = \mathrm{id}$, so $G \simeq N \times A \simeq \mathbb{Z}_p \times \mathbb{Z}_2$.
- $\gamma(a)(b) = b^{-1}$, so $G \simeq D_p$.

So groups of order $2p$ with p an odd prime are either abelian or dihedral.

61.5. *Example* (Classify groups of order pq). Here is a partial classification of groups G of order pq, where $p > q$ are *distinct primes.* Again by Cauchy we have $a \in G$ with $o(a) = q$, and $b \in G$ with $o(b) = p$, so consider the subgroups $N := \langle b\rangle$ and $A := \langle a\rangle$. Note that $A \cap N = \{e\}$. Also, since $[G : N] = q$ is the smallest prime dividing $|G| = pq$, the subgroup N is normal (proved on PS 11).

Thus as before we have $NA = G$, and so $G = N \rtimes_\gamma A$ for some $A \to \mathrm{Aut}(N)$. We know that $A \simeq \mathbb{Z}_q$ and $N \simeq \mathbb{Z}_p$, and that $\mathrm{Aut}(N) \simeq \Phi(p)$ has order $p - 1$.

So every group of order pq is isomorphic to a semidirect product of the above type. There are two cases:

- If $q \nmid p - 1$, then $\Phi(p)$ has no elements of order q, so γ must be the trivial homomorphism, with $\gamma(a) = \mathrm{id}$. Thus in this case $G \simeq \mathbb{Z}_p \times \mathbb{Z}_q \simeq \mathbb{Z}_{pq}$. Thus, all groups of order $15 = 5 \cdot 3$, $33 = 11 \cdot 3$, $35 = 7 \cdot 5$, $77 = 11 \cdot 7$, etc., are abelian.
- If $q \mid p - 1$, then $\Phi(p)$ contains elements of order p. So in this case we will get at least one non-trivial homomorphism $A \to \mathrm{Aut}(N)$, and thus at least one non-abelian group of order pq. Thus, there are non-abelian groups of order $21 = 7 \cdot 3$, $39 = 13 \cdot 3$, $55 = 11 \cdot 5$, etc.

This is just here in case you are wondering. In fact, it turns out that $\mathrm{Aut}(N) \simeq \Phi(p)$ is always *cyclic* when p is prime. This implies that if $q \mid p - 1$, then there are exactly q different homomorphisms $A \to \mathrm{Aut}(N)$ corresponding to q distinct elements $\phi \in \mathrm{Aut}(N)$ such that $\phi^p = \mathrm{id}$, so $\gamma(a) = \phi$. One of these is $\phi = \mathrm{id}$, which gives the product group $N \times A \simeq \mathbb{Z}_p \times \mathbb{Z}_q \simeq \mathbb{Z}_{pq}$. The others give non-abelian groups, and it turns out that these non-abelian semi-direct products are all isomorphic to each other. So if $q \mid p - 1$ there are exactly *two* groups of order pq up to isomorphism.

61.6. *Remark.* **This is just here in case you are wondering.** Let's see why all the non-abelian groups G of order 21 are isomorphic to each other. We have seen these are isomorphic to semidirect products $G = N \rtimes_\gamma A$, where $N = \langle b\rangle$ has order 7, $A = \langle a\rangle$ has order 3, and $\gamma\colon A \to \mathrm{Aut}(N) \simeq \Phi(7)$ is a non-trivial homomorphism.

There are only two non-trivial homomorphisms $\gamma_1, \gamma_2\colon A \to \mathrm{Aut}(N)$, defined by

$$\gamma_1(a)(b) = b^2, \qquad \gamma_2(a)(b) = b^4,$$

since $[2], [4] \in \Phi(7)$ are the elements of order 3.

The group $G_1 = N \rtimes_{\gamma_1} A$ can be described as:

$$G_1 = \{e, b, \dots, b^6, a, ba, \dots, b^6a, a^2, ba^2, \dots, ba^6\}, \qquad b^7 = e = a^3, \quad ab = b^2a^{-1}.$$

Suppose we set $a' := a^2$ in G_1. Then we can use a' in place of a in the description, and everything is almost the same, except we would replace the last relation with $a'b = b^4a'$. Thus, we can also describe G_1 as

$$G_1 = \{e, b, \dots, b^6, a', ba', \dots, b^6a', a'^2, ba'^2, \dots, ba'^6\}, \qquad b^7 = e = a'^3, \quad a'b = b^2a'^{-1}.$$

But this is exactly how we can describe G_2. So this actually gives an isomorphism $G_1 \simeq G_2$.

A similar kind of argument works for any pq.

These results, together with our classification of finite abelian groups, completely classify finite groups of order ≤ 11 up to isomorphism, except for order 8.

order	groups
1	$\{e\}$
2	$\mathbb{Z}_2$
3	$\mathbb{Z}_3$
4	$\mathbb{Z}_4$, $\mathbb{Z}_2 \times \mathbb{Z}_2$
5	$\mathbb{Z}_5$
6	$\mathbb{Z}_6 \simeq \mathbb{Z}_2 \times \mathbb{Z}_3$, $D_3 \simeq S_3$
7	$\mathbb{Z}_7$
8	$\mathbb{Z}_8$, $\mathbb{Z}_4 \times \mathbb{Z}_2$, $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$, D_4, Q
9	$\mathbb{Z}_9$, $\mathbb{Z}_3 \times \mathbb{Z}_3$
10	$\mathbb{Z}_{10} \simeq \mathbb{Z}_2 \times \mathbb{Z}_5$, D_5
11	$\mathbb{Z}_{11}$

The new group Q of order 8 on this table is the **quaternion group**. It is harder to get to that the others because it is not a semi-direct product of any pair of its proper subgroups. **quaternion group**

61.7. *Exercise.* Let $Q = \{\pm I, \pm A, \pm B, \pm C\} \subseteq GL_2(\mathbb{C})$, where

$$A := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad B := \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \quad C := AB = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}.$$

Verify the formulas

$$A^2 = B^2 = C^2 = -I, \qquad AB = C,\ BC = A,\ CA = B, \quad BA = -C,\ CB = -A,\ AC = -B,$$

and conclude that Q is a subgroup of $GL_2(\mathbb{C})$. This is the quaternion group.

This is just here in case you are wondering. Here is a roadmap for classifying groups of order 8.

(1) Show that if every element of G has order 1 or 2, then G is abelian, and so $G \simeq \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$.

(2) If not all elements of G have order 2, then there is at least one element b of order 4. Set $N = \langle b \rangle$, and note that N is a normal subgroup since $[G : N] = 2$.

Now consider elements of $G \smallsetminus N$. Note that for any $a \in G \smallsetminus N$, we have $G = N \cup aN$, so $G = \langle a, b \rangle$.

(3) If there is $a \in G \smallsetminus N$ with $o(a) = 2$, then show $G \simeq N \rtimes_\gamma A$, so that $G \simeq \mathbb{Z}_4 \times \mathbb{Z}_2$ or $G \simeq D_4$.

(4) If there is $a \in G \smallsetminus N$ with $o(a) = 8$, then $G \simeq \mathbb{Z}_8$.

(5) If $a \in G \smallsetminus N$ with $o(a) = 4$, since $N \trianglelefteq G$ we have either $aba^{-1} = b$, in which case G is abelian and it is one of $\mathbb{Z}_4 \times \mathbb{Z}_2$ or $\mathbb{Z}_8$, or $aba^{-1} = b^{-1}$.

If $aba^{-1} = b^{-1}$, show that there is an isomorphism $G \to Q$ to the quaternion group, sending $a \mapsto A$ and $b \mapsto B$.

62. Rings

I'll briefly recall the notion of a ring. Rings have two operations: addition and multiplication. **Lecture 32**

A **ring** is a data $(R, +, \cdot)$ consisting of a set R, with two binary operations **ring**

$$+\colon R \times R \to R, \qquad \cdot\colon R \times R \to R,$$

such that

- R is an abelian group under $+$,
- multiplication is associative: $(ab)c = a(bc)$

- multiplication distributes over addition:

$$a(b+c) = (ab) + (ac), \qquad (a+b)c = ac + bc.$$

The identity for + is conventionally called 0, and the inverse of a under + is called $-a$.

Remark on precedence. If R is a ring and $a, b, c \in R$, then an expression like

$$a + bc$$

would seem to be ambiguous: it could mean either $a + (bc)$ or $(a + b)c$. In practice, we make the assumption we are already used to making: multiplication has "higher precedence" than addition, so we interpret this as meaning $a + (bc)$.

There are several modifications to this definition.

- A **ring with identity** is a ring such that there exists $1 \in R$ with the property that $1a = a = a1$ for all $a \in R$. **ring with identity**
- A **commutative ring** is a ring such that multiplication is commutative: $ab = ba$ for all $a, b \in R$. **commutative ring**

Warning. Like many basic textbooks, our textbook does not *require* rings to have a multiplicative identity. However, in many fields of mathematics, it is standard to require a multiplicative identity. I'll follow the textbook, but be warned you will see other conventions.

62.1. *Example.* The integers $(\mathbb{Z}, +, \cdot)$ is a commutative ring with identity. The operations restrict to the subset $\mathbb{Z}2$ of even integers. Then $(2\mathbb{Z}, +, \cdot)$ is a commutative ring, but does not have identity.

When R has a multiplicative identity, you can talk about inverses: an element $a \in R$ is a **unit** if there exists $b \in R$ such that $ab = 1 = ba$. The element b is called the **multiplicative inverse** of a, and is usually written b^{-1}. **unit** **multiplicative inverse**

A few basic facts which are easy to prove.

(1) In any ring, $0a = 0 = a0$ for any $a \in R$. (Because $a0 + a0 = a(0 + 0) = a0$.)
(2) If R has 1, then $-a = (-1)a$. (Because $a + (-1)a = 1a + (-1)a = (1 + (-1))a = 0a = 0$.
(3) The multiplicative identity in a ring is unique, if it exists.
(4) If an element of a ring with identity has an inverse, that inverse is unique.
(5) The subset $R^{\times} \subseteq R$ of units in a ring with identity is a group, via multiplication.

A **field** F is a commutative ring with identity such that every non-zero element is a unit, and such that $1 \neq 0$. This says exactly that $F^{\times} = F \smallsetminus \{0\}$. **field**

62.2. *Remark.* The **zero ring** is the set $R = \{0\}$, equipped with the only possible operations of + and $\cdot$. Under our definitions this is a ring, and it even has identity: $1 = 0$, and every element of R is a unit. However, we exclude this from being a field. **zero ring**

The notion of a ring is a lot more complicated than that of a group. However, it is also a lot more familiar: you already know many important examples.

Examples.

- The set $\mathbb{Z}$ of integers is a ring, commutative with identity.
- The set $\mathbb{R}$ of real numbers is a field.
- The rationals $\mathbb{Q}$ and complex numbers $\mathbb{C}$ are fields.
- The set $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$ of integers modulo n is a commutative ring with identity.
- The set $\mathbb{Z}_p$ is a field if p is a prime number.
- The set $\mathrm{Mat}_{n\times n}(R)$ of matrices with entries in a ring R is a ring, with identity; the product is matrix multiplication. It is not commutative (unless $n = 1$ and R is commutative.) If R has an identity so does $\mathrm{Mat}_{n\times n}(R)$.
- The set $F(\mathbb{R}, \mathbb{R})$ of all functions $f\colon \mathbb{R} \to \mathbb{R}$ is a commutative ring with identity, under "pointwise" addition and multiplication:

$$(f + g)(x) := f(x) + g(x), \qquad (fg)(x) := f(x)g(x).$$

What are the zero and identity elements?

- The set $C(\mathbb{R},\mathbb{R})$ of **continuous** functions $f\colon \mathbb{R} \to \mathbb{R}$ is a subring of $F(\mathbb{R},\mathbb{R})$, commutative with identity. This is because of the fact that sums and products of continuous functions are continuous; it has identity because constant functions are continuous. **continuous**
- The set $C_0(\mathbb{R},\mathbb{R})$ of continuous functions f such that the limits $\lim_{x\to\pm\infty} f(x)$ exist and are $= 0$ is a subring of $F(\mathbb{R},\mathbb{R})$, commutative but without identity (since the function $f \equiv 1$ is not in this set).

A **subring** of a ring R is a subset $S \subseteq R$ which is a ring with the two operations inherited from R. **subring**

This means that the operations on R must *restrict* to S, so that $a, b \in S$ implies $a + b, ab \in S$, and that these make S into a ring.

62.3. **Proposition.** *A subset S of a ring R is a subring if and only if*

(1) *$0 \in S$ (you can replace this condition with "S is non-empty").*
(2) *For all $a, b \in S$, we have $a + b, ab \in S$.*
(3) *If $a \in S$ then $-a \in S$.*

Note that a subring of a commutative ring must be commutative. However, a subring of a ring with identity need not have identity.

62.4. *Example.* $\mathbb{Z}2$ is a subring of $\mathbb{Z}$, although it does not have identity.

62.5. *Example.* We have subrings $\mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$.

62.6. *Example.* Let $S \subset \mathrm{Mat}_{2\times 2}(\mathbb{R})$ be the subset consisting of matrices of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$, $a, b \in \mathbb{R}$. Then S is a subring of the matrix ring. (As we will see, it is *isomorphic* to $\mathbb{C}$.)

Warning. It is possible to have a subring $S \subseteq R$, where both R and S have multiplicative identity, but they are *different elements*! (However, the additive identity will always be the same.) An example is given on a homework assignment.

This is different from what we saw for groups: the identity element of a subgroup is always the same as that of the larger group. The difference is because groups have inverses of all elements, and we needed that to prove that the identity elements must be the same.

63. Complex numbers and quaternions

A **complex number** is a "formal expression" of the form $a + bi$, where $a, b \in \mathbb{R}$, and i is a special symbol. The set $\mathbb{C}$ of complex numbers has the structure of a field, where **complex number**

- Addition is defined in the "obvious" way, so that
$$(a + bi) + (c + di) = (a + b) + (c + d)i;$$
- Multiplication is defined in the "obvious" way, together with the stipulation that $i^2 = -1$. Thus
$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i.$$

The additive identity is $0 := 0+0i$, while the multiplicative identity is $1 := 1+0i$. It is straightforward to check all the axioms for a commutative ring with identity, so I leave that to you. (Proving associativity of multiplication is a bit tedious, but it's something you just have to grind through.)

In fact, $\mathbb{C}$ is not just a commutative ring, but a *field.* To prove this, note that if $z = a + bi \in \mathbb{C}$ is not equal to 0, then the number
$$w = \frac{a}{a^2 + b^2} + \frac{-b}{a^2 + b^2} i$$
is defined (since $a + bi \neq 0$ implies $a^2 + b^2 \neq 0$, because a and b are *real* numbers, and squares of non-zero reals are positive), and you can check that $zw = 1$ as desired.

63.1. *Remark.* If you don't like using "formal expressions", you can think of a complex number as an ordered pair (a, b) of real numbers, with operations defined by $(a, b) + (c, d) := (a + c, b + d)$ and $(a, b) \cdot (c, d) := (ac - bd, ad + bc)$.

63.2. *Example.* Let $R = \mathbb{Z}_3[i]$ denote the set of "formal expressions" $a + bi$, where $a, b \in \mathbb{Z}_3$ (instead of $\mathbb{R}$), with $i^2 = -1$. Then R is a field, where operations are defined just as they are for $\mathbb{C}$. The proof is the roughly same: the hard part is for multiplicative inverses, where we need to use the fact that if $a, b \in \mathbb{Z}_3$ are not both 0, then $a^2 + b^2 \neq 0$ (because $a^2, b^2 \in \{[0], [1]\}$ by direct calculation).

So $\mathbb{Z}_3[i]$ is a field with 9 elements. It is often denoted $\mathbb{F}_9$ or $GF(9)$.

63.3. *Example.* Let $R' = \mathbb{Z}_5[i]$ be the set of formal expressions $a + bi$ with $a, b \in \mathbb{Z}_5$, with $i^2 = -1$. This becomes a commutative ring, but not a field. The formula for multiplicative inverse doesn't work because for instance $[1]^2 + [2]^2 = [0]^2$.

Basically, the point is that $\mathbb{Z}_5$ already has a "square root of -1", namely ± 2, so adding an additional one messing things up a bit.

63.4. *Exercise.* Let $R'' = \mathbb{Z}_5[\gamma]$ be the set of formal expressions $a + b\gamma$ with $a, b \in \mathbb{Z}_5$, and $\gamma^2 = 2$. Show that this is a field. (Hint: $a^2 + 2b^2 = 0$ in $\mathbb{Z}_5$ only if $a = b = 0$.)

63.5. *Example* (Gaussian numbers). Let $\mathbb{Q}[i]$ be the set of formal expressions $a + bi$ with $a, b \in \mathbb{Q}$. This is a field, for the same reason as $\mathbb{C}$ is. Note that this ring is actually a subring of $\mathbb{C}$. It is called the **field of Gaussian numbers**. **field of Gaussian number**

Likewise we can form $\mathbb{Z}[i]$, a subring of $\mathbb{Q}[i]$ and thus of $\mathbb{C}$. This is the **ring of Gaussian integers**. **ring of Gaussian integer**

63.6. *Exercise.* What is the group of units $\mathbb{Z}[i]^\times$ in the ring of Gaussian integers?

63.7. *Remark.* For a prime nmuber p, we always get a commutative ring $\mathbb{Z}_p[i]$, but it is a field if and only if $p \equiv 3 \mod 4$. (This is not an obvious fact.)

The **quaternions** $\mathbb{H}$ are the set of formal expressions **quaternions**

$$a + bi + cj + dk, \qquad a, b, c, d \in \mathbb{R},$$

where i, j, k are formal symbols. Addition is defined in the obvious way, while multiplication is determined by the table

$$\begin{array}{lll} ii = -1 & ij = k & ik = -j \\ ji = -k & jj = -1 & jk = i \\ ki = j & kj = -i & kk = -1 \end{array}$$

The quaternions form a ring with identity, but not a commutative ring. We cannot call $\mathbb{H}$ a field, but it does have the property that non-zero elements have a multiplicative inverse: verify that

$$(a + bi + cj + dk)^{-1} = \frac{a}{L} + \frac{-b}{L}i + \frac{-c}{L}j - \frac{-j}{L}k, \qquad L := a^2 + b^2 + c^2 + d^2.$$

63.8. *Remark.* We can think of the quaternion $a + bi + cj + dk$ as an ordered pair $(a, \mathbf{u})$ consisting of a scalar $a \in \mathbb{R}$ and a vector $\mathbf{u} = a_1 i + a_2 j + a_3 k \in \mathbb{R}^3$. Then the formula for multiplication has the form

$$(a, \mathbf{u})(b, \mathbf{v}) = (ab - \mathbf{u} \cdot \mathbf{v}, a\mathbf{v} + b\mathbf{u} + \mathbf{u} \times \mathbf{v}).$$

63.9. *Exercise.* Show that the quaternions does satisfy the axioms for a ring. (The hard part is associativity of multiplication.)

64. Polynomial rings

Fix a commutative ring R with identity (for instance, a field).

Let $R[x]$ be the set of polynomials in x with coefficients in R. This is the set of finite "formal expressions" of the form

$$\sum_{i=0}^{n} a_i x^i = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0,$$

with $a_0, \dots, a_n \in K$. When $a_n = 0$, we treat the expressions

$$0x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \quad \text{and} \quad a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

as the same element of $R[x]$. So you can also think of a polynomial as a formal expression $\sum_i a_i x^i$ where only finitely many a_i are non-zero. In particular there is a 0 polynomial.

64.1. *Remark.* An more correct way to say it: an element f of $R[x]$ corresponds to an infinite sequence $(a_k)_{k\geq 0}$ of elements of R, with the property that there exists N such that $a_k = 0$ for all $k \geq N$. But we choose to write such a sequence as $\sum a_k x^k$.

For instance, x corresponds to the sequence $(0, 1, 0, 0, 0, \dots)$, i.e., $a_1 = 1$ and all other $a_k = 0$.

$R[x]$ is a commutative ring with idenity. We add and multiply by

$$\sum_i a_i x^i + \sum_j b_j x^j = \sum_k (a_k + b_k) x^k,$$

$$\Big(\sum_i a_i x^i\Big)\Big(\sum_j b_j x^j\Big) = \sum_k \Big(\sum_{i+j=k} a_i b_j\Big)\, x^k.$$

The 0 element is the 0 polynomial. The 1 element is the constant polynomial 1.

64.2. *Remark.* In terms of sequences of elements of R, this says

$$(a_k) + (b_k) = (a_k + b_k), \qquad (a_k)(b_k) = (a_k b_0 + a_{k-1} b_1 + \cdots + a_0 b_k).$$

For instance, if $f = x^2 + 3$, $g = x - 1$, $h = 2x + 7$ are elements of $\mathbb{R}[x]$, then

$$fg + h = (x^2 + 3)(x - 1) + (2x + 7) = x^3 - x^2 + 5x + 4.$$

Verify the distributive law: $f(x)(g(x) + h(x)) = (f(x)g(x)) + (f(x)h(x))$. (See Goodman 1.8.)

The ring of polynomials encodes the algebra of polynomials we all know and love. Note that we do not here think of polynomials as *functions*, but merely as formal expressions that can be manipulated.

A **constant polynomial** is a polynomial $f = \sum a_k x^k$ such that $a_k = 0$ whenever $k \geq 1$. **constant polynomial**

64.3. **Proposition.** *The subset $C \subseteq R[x]$ of constant polynomials is a subring.*

Proof. Straightforward: check that constant polynomials are closed under taking sums, products, inverses. □

Note that there is an obvious bijection $R \to C$, sending an element $a \in R$ to the constant polynomial $f = \sum a_k x^k$ with $a_0 = 0$. In fact, this bijection is compatible with the ring operations: adding or multiplying constant polynomials amounts to adding or multiplying elements of R. For this reason, it is typical to "identify" elements of R with the corresponding constant polynomials. This is compatible with the way we usually write these: a number a is also the way we write the constant polynomial a.

64.4. *Remark.* There are also rings of polynomials in several variables. For instance, $R[x, y]$ consists of formal expressions $\sum_{i,j} a_{ij} x^i y^j$. I won't spell out how to do this, but it turns out that one possibly definition is $R[x, y] := (R[x])[y]$.

65. Degree of a polynomial in one variable

There is a notion of degree of a polynomial, which works best when the coefficients are a field.

Fix a field K.

The **degree** of a *non-zero* polynomial $f(x) = \sum_k a_k x^k$ in $K[x]$ is the largest integer k such that $a_k \neq 0$. We write $\deg(f)$ for this integer. **degree**

The above makes no sense for the zero polynomial, so we instead make the convention that $\deg(0) := -\infty$. Thus, degree is a function

$$\deg\colon R[x] \to \mathbb{Z}_{\geq 0} \cup \{-\infty\}.$$

A **constant** polynomial is one with degree ≤ 0. (Not $= 0$, since the 0 polynomial is constant.) **constant**

65.1. *Exercise.* The subset $C \subseteq R[x]$ of constant polynomials is a subring, which is isomorphic to the ring R. (We haven't actually defined isomorphism of rings yet.)

Degree has the following properties.

65.2. **Proposition.** *For $f, g \in K[x]$, we have*

(1) $\deg(fg) = \deg(f) + \deg(g)$. *(We make the convention that $(-\infty) + (\text{anything}) = -\infty$.)*
(2) $\deg(f \pm g) \leq \max\{\deg(f), \deg(g)\}$. *(We make the convention that $-\infty \leq$ anything.)*

Proof. For (1), suppose $\deg(f) = m$ and $\deg(g) = n$, where both $m, n \in \mathbb{Z}_{\geq 0}$. So $f = \sum a_i x^i$ with $a_m \neq 0$, $a_i = 0$ if $i > n$, and $g = \sum b_j x^j$ with $b_n \neq 0$, $b_j = 0$ if $j > n$. Then if we write $fg = \sum c_k x^k$, then $c_k = \sum_{i=0}^k a_i b_{k-i}$. If $k > m + n$ then this is 0 since for each i either $i > m$ or $k - i > n$. If $k = m + n$, then $c_{m+n} = a_m b_n \neq 0$, since the product of non-zero elements in a field is non-zero.

The proof of (2) is similar, but easier: the largest possible non-zero coefficient of $f + g = \sum d_k x^k$ is at $k = \max(m, n)$, but note that if $m = n$ this can possibly be 0 if $a_k = -b_k$. □

Note: in a field, non-zero elements are exactly units, so their product is also a unit and thus non-zero. (E.g., if a^{-1} and b^{-1} exist, then $(ab)^{-1} = b^{-1}a^{-1}$.)

The first line includes the fact that $f \neq 0$ and $g \neq 0$ implies $fg \neq 0$.

65.3. **Corollary.** *If K is a field, the group $K[x]^\times$ of units in the polynomial ring consists of constant non-zero polynomials, and is isomorphic to the group $K^\times$ of units in K.*

The following is a statement of "polynomial long division".

65.4. **Proposition** (Division algorithm for polynomials)**.** *Let K be a field. Let $p, d \in K[x]$ with $\deg(d) \geq 0$. Then there exist unique $q, r \in K[x]$ such that*

- *$p = dq + r$ and*
- *$\deg(r) < \deg(d)$.*

Proof. I'll write $m = \deg(p)$ and $n = \deg(d)$. Note that $n \neq -\infty$.

First I'll prove an easier statement: under the hypotheses of the proposition, and if $m \geq n$, there is a "monomial" cx^k with $c \in R$ and $k \geq 0$, and a polynomial p' such that

$$p = (cx^k)d + p', \qquad \deg(p') < \deg(p).$$

To do this, write $p = a_m x^m + (\text{lower deg terms})$ and $d = b_n x^n + (\text{lower deg terms})$, and set $cx^k := (a_m b_n^{-1})^{-1} x^{m-n}$. Then

$$\begin{aligned} p' := p - (cx^k)d &= a_m x^m + (\text{lower deg terms}) - a_m b_n^{-1} x^{m-n}\big(b_n x^n + (\text{lower deg terms})\big) \\ &= (a_m x^m - a_m x^m) + (\text{lower degree terms}), \end{aligned}$$

so $\deg(p') < m$.

Now to prove the proposition, we work by induction on $m = \deg(p)$. If $m < n = \deg(d)$, just set $q = 0$ and $r = p$, so that

$$p = 0d + p.$$

In particular, this handles the base case of $\deg(p) = -\infty$, i.e., if $p = 0$.

Now suppose $m \geq n$, and assume the inductive hypothesis that the proposition holds for all p with degree $< m$. By the above observation, we can write

$$p = (cx^k)d + p', \qquad \deg(p') < m,$$

for some monomial cx^k. Since $\deg(p') < m$, the inductive hypothesis gives $p' = q'd + r$ for some $q', r \in K[x]$ with $\deg(r) < n$. Then

$$p = (cx^k)d + p' = (cx^k)d + q'd + r = (cx^k + q')d + r,$$

so $q = cx^k + q'$.

The values of q and r are unique: if $p = qd + r = q'd + r'$ with $\deg(r), \deg(r') < n$, then

$$0 = p - p = (q - q')d + (r - r'), \qquad \Longrightarrow \qquad r' - r = (q - q')d.$$

If $q \neq q'$, then $\deg((q - q')d) = \deg(q - q') + \deg(d) \geq n$, but this contradicts $\deg(r' - r) \leq \max\{\deg(r), \deg(r')\} < n$. Thus we must have $q = q'$ and thus $r = r'$. □

Note: you know this as the "division algorithm for polynomials", and probably think of it as computing

$$\frac{p}{d} = q + \frac{r}{d}, \qquad \deg(r) < \deg(d).$$

A *monic* polynomial is a polynomial with leading coefficient 1. That is, $f = x^n + a_{n-1}x^{n-1} + \cdots + a_0$. Note that the zero polynomial is not a monic polynomial; however, the constant polynomial 1 is monic.

66. Homomorphisms of rings

Lecture 34

A **homomorphism** of rings is a function $\phi \colon R \to S$ satisfying

homomorphism

$$\phi(x + y) = \phi(x) + \phi(y), \quad \phi(xy) = \phi(x)\phi(y)$$

for all $x, y \in R$.

Note that in particular, ϕ gives a map of abelian groups $(R, +) \to (S, +)$, so we must have $\phi(0) = 0$. It is *not* necessarily the case that $\phi(1) = 1$. I will say that a homomorphism is a **unital homomorphism** if also $\phi(1) = 1$.

unital homomorphism

66.1. *Example.* The complex conjugation function $\phi \colon \mathbb{C} \to \mathbb{C}$, which sends $\alpha(a + bi) = a - bi$, is a unital homomorphism.

66.2. *Example.* Consider the projection map $\phi \colon \mathbb{Z} \to \mathbb{Z}_n$, defined by $\phi(x) := [x]_n$. This is a ring homomorphism. It is unital, since $\phi(1) = [1]_n$ is a multiplicative identity in $\mathbb{Z}_n$.

66.3. *Example.* Let R be any ring with identity. There is a unique homomorphism of abelian groups $\phi \colon \mathbb{Z} \to R$ which sends the generator 1 of $\mathbb{Z}$ to $1 \in R$. It maps $\mathbb{Z}$ onto the cyclic subgroup $\langle 1 \rangle$ of $(R, +)$.

I claim that this homomorphism of groups is actually a ring homomorphism. That is, for any two integers $m, n \in \mathbb{Z}$, we have $\phi(mn) = \phi(m)\phi(n)$.

Remember the notation in an abelian group: given an integer n, we write

$$na = \underbrace{a + \cdots + a}_{n \text{ times}},$$

$$(-n)a \underbrace{(-a) + \cdots + (-a)}_{n \text{ times}},$$

when $n > 0$, and $0a = 0$. Thus $\phi(n) = n1$ with this notation.

Claim 1: in a ring, we always have $(n1)a = na$, where $a \in R$.

Claim 2: we always have $m(na) = (mn)a$ where $a \in R$. (This is just because R is an abelian group.)

Therefore, $(m1)(n1) = m(n1) = (mn)1$, so $\phi(m)\phi(n) = \phi(mn)$.

66.4. **Isomorphism of rings.** An **isomorphism** between rings R and S is a bijection $\phi\colon R \to S$ such that

isomorphism

$$\phi(a+b) = \phi(a) + \phi(b), \qquad \phi(ab) = \phi(a)\phi(b),$$

for all $a, b \in R$. Note that if ϕ is an isomorphism of rings, its inverse map $\phi^{-1}\colon S \to R$ is an isomorphism of rings.

66.5. *Example.* Let $\phi\colon \mathbb{C} \to \mathbb{C}$ be the complex conjugation function $\phi(z) := \overline{z}$, i.e., $\phi(a+bi) := a - bi$ for $a, b \in \mathbb{R}$. Then ϕ is an isomorphism of rings (in fact, an *automorphism* of the field $\mathbb{C}$).

66.6. *Example.* Let $S \subset M_{2\times 2}(\mathbb{R})$ be the subring consisting of matrices of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$. Define a function $\phi\colon S \to \mathbb{C}$ by

$$\phi(\left[\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right]) := a + bi.$$

This is an isomorphism of rings.

67. The substitution principle

There is a special recipe for describing homomorphisms from a polynomial ring to some other (commutative) ring.

67.1. **Proposition** (Substitution principle)**.** *Let R and S be commutative rings with identity, and suppose $\phi\colon R \to S$ is a unital ring homomorphism. Then for each $c \in S$, there exists a unique ring homomorphism $\phi_c\colon R[x] \to S$, such that*

(1) *$\phi_c(r) = \phi(r)$ if $r \in R \subseteq R[x]$ (i.e., if r is a constant polynomial), and*
(2) *$\phi_c(x) = c$.*

The formula for ϕ_c is

$$\phi_c\bigl(\sum a_k x^k\bigr) = \sum a_k c^k.$$

This is called the "substitution principle" because the formula for ϕ_c is literally: "substitute c for x".

67.2. *Example.* Suppose $R = S$ and $\phi\colon R \to S$ is the identity function. Then we might write ev_c for ϕ_c, and call it the **evaluation function**: it is defined by

evaluation function

$$\mathrm{ev}_c(a_0 + a_1 x + \cdots + a_n x^n) = a_0 + a_1 c + \cdots + a_n c^n.$$

We have another notation for this: if $f = \sum a_k x^k$, we write $f(c) := \mathrm{ev}_c(f)$.

The point is that this function ev_c is a unital ring homomorphism:

$$\mathrm{ev}_c(1) = 1, \qquad \mathrm{ev}_c(f+g) = \mathrm{ev}_c(f) + \mathrm{ev}_c(g), \qquad \mathrm{ev}_c(fg) = \mathrm{ev}_c(f)\,\mathrm{ev}_c(g).$$

These are just obscure ways of writing the identities:

$$s(0) = 1, \qquad (f+g)(c) = f(c) + g(c), \qquad (fg)(c) = f(c)g(c),$$

where I'm writing s for the constant polynomial corrresponding to $1 \in R$.

Proof of Substitution Principle. First check that the formula I gave actually does what is claimed: that it is a unital ring homomorphism ($\phi_c(1) = 1$, $\phi_c(f+g) = \phi_c(f) + \phi_c(g)$, $\phi_c(fg) = \phi_c(f)\phi_c(g)$) which satisfies (1) and (2) ($\phi_c(r) = \phi(r)$ if $r \in R$ and $\phi_c(x) = c$.).

The last two properties are immediate, and also imply $\phi_c(1) = 1$. So you just need to check sum and product, which is a calculation:

Let $f = \sum_i a_i x^i$ and $g = \sum_j b_j x^j$. Since $f + g = \sum_k (a_k + b_k)x^k$, we have

$$\phi_c(f + g) = \sum_k (a_k + b_k)c^k, \qquad \phi_c(f) + \phi_c(g) = \sum_i a_i x^i + \sum_j b_j x^j,$$

which are equal, so $\phi_c(f + g) = \phi_c(f) + \phi_c(g)$. Since $fg = \sum_k (\sum_{i=0}^k a_i b_{k-i})x^k$, we have

$$\phi_c(fg) = \sum_k (\sum_{i=0}^k a_i b_{k-i})c^k, \qquad \phi_c(f)\phi_c(g) = \big(\sum_i a_i c^i\big)\big(\sum_j b_j c^j\big),$$

and by expanding out the right-hand formula and rearranging terms you can check these are equal, so $\phi_c(fg) = \phi_c(f)\phi_c(g)$.

Conversely, if $\psi\colon R[x] \to S$ is any ring homomorphism which agrees with ϕ on R and $\psi(x) = c$, you can easily verify

$$\begin{aligned}\psi\big(\sum a_k x^k\big) &= \sum \psi(a_k x^k)\\ &= \sum \psi(a_k)\psi(x)^k\\ &= \sum \phi(a_k)c^k \qquad\qquad \text{using } \psi|_R = \phi \text{ and } \psi(x) = c.\end{aligned}$$

□

Note that any polynomial $f \in R[x]$ gives you a function $R \to R$, by sending $c \mapsto f(c) = \mathrm{ev}_c(f)$.

67.3. *Exercise.* Show that the function $\epsilon\colon R[x] \to \mathcal{F}(R, R)$ sending f to this function is a homorphism of rings.

Warning. A polynomial might not be determined by its function.

67.4. *Example.* For instance, suppose $R = \mathbb{Z}_p$ with p prime. Consider the polynomials $f = x^p$ and $g = x$ in $\mathbb{Z}_p[x]$. Then we have that

$$\mathrm{ev}_c(f) = c^p, \qquad \mathrm{ev}_c(g) = c \qquad \text{for all } c \in \mathbb{Z}_p,$$

but these are always *equal* by Fermat's Little Theorem. So both f and g give you the same function $\mathbb{Z}_p \to \mathbb{Z}_p$ (actually, the identity function). So you cannot recover the polynomial just from its function.

Similarly, $f - g = x^p - x$ gives the constant function 0.

Note: it turns out to be true that a polynomial in $K[x]$ is determined by its function when K is an *infinite field*, such as $\mathbb{Q}$ or $\mathbb{R}$. (We will be able to prove this soon.)

68. Ideals

Recall that given a group G and a *normal subgroup* $N \leq G$, we could form a quotient group G/N. In ring theory, the analogue of a normal subgorup is an ideal.

An **ideal** of a ring R is a subset $I \subseteq R$ such that **ideal**

(1) I is a subgroup of $(R, +)$,
(2) if $r \in R$ and $x \in I$, then $rx, xr \in I$.

Note that an ideal, under our definition of ring, is a subring. However, it is not required that an ideal contain an identity element, and it usually won't.

Warning. The notion of ideal we have defined is sometimes called a **two-sided ideal**, to distinguish it from notions of "left-ideal" and "right-ideal", which I won't discuss. **two-sided ideal**

68.1. *Example* (Unit ideal and trivial ideal). In any ring R, the subsets R and $\{0\}$ are always ideals of R.

68.2. *Example* (Ideals in a field). If K is a field, then the *only* ideals are $\{0\}$ and K. This follows from the fact that all non-zero elements of K are units, and the following exercise.

68.3. *Exercise.* Show that if R is a ring with 1 and if $I \subseteq R$ is an ideal, then if I contains any unit $a \in R^\times$, then $I = R$. (For this reason, the ideal $I = R$ is sometimes called the "unit ideal".)

68.4. *Example.* In $R = \mathbb{Z}$, the subsets $n\mathbb{Z} = \{\, nx \mid x \in \mathbb{Z}\,\}$ are ideals for every n.

68.5. *Example.* Let K be a field, and $n \geq 1$. Then the matrix ring $R = \operatorname{Mat}_{n\times n}(K)$ only has two ideals, namely K and (0). The proof of this is a little tricky, and is outlined in the exercise (Goodman 6.2.11).

There's another proof, using the following fact: if $A \in \operatorname{Mat}_{n\times n}(K)$ is any matrix, then there exist $P, Q \in R^\times = GL_n(K)$ such that $PAQ = B_r := \left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right]$, where $r = \operatorname{rank} A$, and I_r is the $r \times r$ identity matrix. (This is easy to prove from things you learned in linear algebra. You can also think of it as "Smith normal form over a field".)

Thus, if $J \subseteq R$ is not the trivial ideal, there exists an $A \in J$ with $r > 0$, and thus $B_r \in J$ for some $r > 0$. With a little more work you can produce, for any $k = 1, \dots, n$ matrices P', Q' such that $E_k = P'B_rQ'$ has 1 in the (k,k)-entry, and 0 in all other entries. Thus J contains $E_1 + \cdots + E_n$, which is the identity matrix, so $J = R$.

69. Kernel of a ring homomorphism

The **kernel** of a ring homomorphism $\phi\colon R \to S$ is the set **kernel**

$$\ker(\phi) := \{\, r \in R \mid \phi(r) = 0\,\}.$$

In other words, it is the same as the kernel of ϕ thought of as a homomorphism of abelian groups.

69.1. **Proposition.** *The kernel* $\ker(\phi) \subset R$ *of a ring homomorphism* $\phi\colon R \to S$ *is an ideal of* R.

Proof. Straightforward. □

69.2. *Example.* Let $R[x]$ be a polynomial ring over some commutative ring R with 1. Choose $c \in R$ and let $\mathrm{ev}_c\colon R[x]$ denote the evaluation homomorphism, defined by $\mathrm{ev}_c(f) := f(c)$. Then the kernel of ev_c is

$$\ker(\phi) = \{\, f \in R[x] \mid f(c) = 0\,\},$$

the subset of polynomials which have c as a root.

70. Ideals generated by subsets

Here are a few facts about ideals we will need. **Lecture 35**

70.1. **Proposition.** *Let R be a ring.*

(1) *If $\{I_\alpha\}$ is any collection of ideals in R, then $J := \bigcap I_\alpha$ is an ideal.*
(2) *If $\{I_n\}_{n\in\mathbb{N}}$ is a chain of ideals (so $I_k \subseteq I_{k+1}$ for all k), then $J := \bigcup_n I_n$ is an ideal.*

Proof. (1) is easy to check. For (2), suppose $a, b \in J$. Then $a \in I_i$ and $b \in I_j$ for some i, j, and therefore $a, b \in I_k$ where $k = \max(i, j)$. Thus $a \pm b \in I_k \subseteq J$. A similar argument shows that if $a \in J$, then $a \in I_i$ for some i, so $ra \in I_i \subseteq J$ for any $r \in R$.

(Warning: a union of an arbitrary collection of ideals is almost never an ideal.) □

(I'm going to assume that rings have identity here. The book doesn't assume this, and it makes things more complicated.)

Let S be a *subset* of R. Define a subset of R by

$$(S) := \{\, a_1s_1b_1 + \cdots + a_ns_nb_n \mid a_i, b_i \in R,\ s_i \in S,\ n \geq 0\,\}.$$

We always assume that $0 \in (S)$ (e.g., $0 = 0s_1 0$). If S is itself empty, then we define $(S) = \{0\}$ (e.g., corresponding to the empty sum).

70.2. **Proposition.** *Let R be a ring with identity. The subset $(S) \subseteq R$ is a (two-sided) ideal of R, and is the intersection of all ideals containing the set S.*

Proof. First note that $S \subseteq (S)$, since $s = 1s1 \in (S)$.

Note that (S) is a subgroup of $(R, +)$. It contains 0. and is closed under addition. It has additive inverses, since $-(a_i s_i b_i) = (-a_i)s_i b_i$, so

$$-\sum a_i s_i b_i = \sum (-a_i) s_i b_i.$$

We also have that (S) is closed by left- and right-multiplication by arbitrary elements of the ring:

$$r \sum a_i s_i b_i = \sum (r a_i) s_i b_i, \qquad (\sum a_i s_i b_i) r = \sum a_i s_i (b_i r).$$

If I is any ideal containing S, then it is clear that any element of the form $\sum a_i s_i b_i \in (S)$ must be in I. Therefore (S) is the intersection of all ideals contanining S. □

Note: if R does not have 1, things get more complicated, since then it is not automatic that $S \subseteq (S)$. Since we don't care about rings without 1, we won't worry about how to do this in that case.

If R is commutative, then $asb = (ab)s$, and we can simplify the above formula to

$$(S) = \{\, a_1 s_1 + \cdots + a_n s_n \mid a_i \in R,\ s_i \in S,\ n \geq 0 \,\}.$$

71. Principal ideals

A **principal ideal** is an ideal which can be generated by a single element. For $x \in R$ we write $(x) := (\{x\})$. Thus **principal ideal**

$$(x) = \{\, a_1 x b_1 + \cdots + a_n x b_n \mid a_i, b_i \in R \,\}.$$

When R is *commutative*, we can always rearrange

$$a_1 x b_1 + \cdots + a_n x b_n = (a_1 b_1) x + \cdots + (a_n b_n) x = (a_1 b_1 + \cdots + a_n b_n) x.$$

Thus, for commutative R, principal ideals have the form

$$(x) = \{\, rx \mid r \in R \,\}.$$

Thus principal ideals in commutative rings are especially important.

There are some rings for which every ideal is principal.

71.1. *Exercise.* Show that if K is a field, then there are only two ideals: $\{0\} = (0)$ and $K = (1)$, both of which are principal.

71.2. **Proposition.** *The ideals in $\mathbb{Z}$ are precisely the subsets $(n) = \mathbb{Z}n = \{\, xn \mid x \in \mathbb{Z} \,\}$, for $n \geq 0$. In particular, all ideals of $\mathbb{Z}$ are principal.*

Proof. We already know that subgroups of $(\mathbb{Z}, +)$ must have this form. These are clearly ideals as well.

Note: it is worthwhile to remember how we proved this. If $I \leq \mathbb{Z}$ is a subgroup of I, then either $I = (0)$, or there is a smallest positive $d \in I$. Clearly $\mathbb{Z}d \subseteq I$. To show $I \subseteq \mathbb{Z}d$, note that for any $a \in I$ we can use the division algorithm, so

$$a = bd + r \qquad \text{for some } b, r \in \mathbb{Z}, \text{ with } r \in \{0, 1, \dots, d-1\}.$$

Since I is an ideal, we must have $r = a - bd \in I$, and since d is the smallest *positive* element in I, we must have $r = 0$. So $a = bd \in \mathbb{Z}d$. □

A polynomial $f \in R[x]$ is **monic** if $f \neq 0$, and the coefficient of the highest power of x is 1. **monic**

71.3. **Proposition.** *Let K be a field. The ideals in $K[x]$ are precisely the subsets $(f) = \{ gf \mid g \in K[x] \}$, where f is either a monic polynomial or is 0. In particular, all ideals of $K[x]$ are principal.*

Proof. The subset $(0) = \{0\}$ is an ideal. If $I \subseteq K[x]$ is not equal to $\{0\}$, it contains a *non-zero* element f of minimal degree n. We are going to show that *every element of I is divisible by this f*, i.e., that $I = (f)$.

Now suppose $p \in I$. The division algorithm for "$p \div f$" gives q, r with $p = fq + r$ with $\deg(r) < \deg(f) = n$. Since $p, f \in I$, it follows that $r = p - fq = (1)p + (-f)q \in I$. Since f has minimal degree among non-zero elements of I, it follows that $r = 0$, so $p = fq$.

Thus, every non-zero ideal in $K[x]$ is of the form (f) for some non-zero polynomial f. We can always write $f = cg$ where g is monic and $c \in K \smallsetminus \{0\}$, and it is easy to see that $(f) = (g)$, since $f = cg$ and $g = c^{-1}f$. □

71.4. *Example* (Kernel of evaluation). Consider a field K, and an element $c \in K$. Consider the evaluation homomorphism

$$\mathrm{ev}_c \colon K[x] \to K,$$

and let $I = \ker(\mathrm{ev}_c) = \{ f \in K[x] \mid f(c) = 0 \}$. I claim that

$$I = (x - c) = \{ (x - c)h \mid h \in K[x] \}.$$

One direction is easy: it is clear that $g := x - c$ is in the kernel of ev_c, since $g(c) = c - c = 0$.

Suppose $f \in K[x]$. Use the division algorithm for $f \div g$: this says there exist $q, r \in K[X]$ such that

$$f = (x - c)q + r, \qquad \deg(r) < \deg(x - c) = 1.$$

Thus, r must be a constant polynomial, so corresponds to an element of K. If we feed this identity into ev_c, we get

$$\mathrm{ev}_c(f) = \mathrm{ev}_c(x - c)\,\mathrm{ev}_c(q) + \mathrm{ev}_c(r) \qquad \Rightarrow \qquad f(c) = (0 - 0)q(c) + r.$$

In other words, $r = f(c)$.

In particular, if $f \in I = \ker(\mathrm{ev}_c)$, then $r = f(0) = 0$, and therefore $f = (x - c)q$, so $f \in (g)$.

The moral is: for a polynomial over field, the kernel of evaluation at c is the principal ideal $(x - c)$. Another way to say this: $f(c) = 0$ *if and only if $x - c$ is a factor of f.*

71.5. **Corollary.** *If K is a field, then any non-zero $f \in K[x]$ has only finitely many roots in K.*

Proof. If f has a root c, then $f = (x - c)g$ for some $g \in K[x]$, and $\deg(g) = \deg(f) - 1$. By inductively applying this observation, we find that we can write

$$f = (x - c_1) \cdots (x - c_k)g$$

for some $c_1, \dots, c_k \in K$ and $g \in K[x]$ such that g has no roots in K. (If $\deg(g) = 0$ then g is a non-zero constant polynomial, which certainly has no roots.) But now we see that $c_1, \dots, c_k$ are the only roots of f in K, since $f(c) = 0$ implies $(c - c_1) \cdots (c - c_k)g(c) = 0$, with $g(c) \neq 0$ since g has no roots. □

71.6. *Example* (Non-principal ideals). Not every ideal is principal. For instance, let $R = \mathbb{Z}[x]$, and consider the evaluation $\mathrm{ev}_2 \colon \mathbb{Z}[x] \to \mathbb{Z}$. Let $I = \ker(\mathrm{ev}_2)$. Then I is not a principal ideal. I'll explain why soon.

72. Quotient rings

Given an ideal I of a ring R, there is a **quotient ring** R/I. Elements of R/I are cosets

quotient ring

$$a + I \subset R$$

of the additive group $(R, +)$ with respect to the subgroup $(I, +)$. Thus, it is automatic that R/I is an abelian group, with

$$(a + I) + (b + I) = (a + b) + I.$$

72.1. **Proposition.** *If $I \subseteq R$ is an ideal, then R/I is a ring, with addition as above and multiplication defined by*

$$(a + I)(b + I) := ab + I.$$

If R has multiplicative identity 1*, then R/I has multiplicative identity $1 + I$. If R is commutative, so is R/I.*

Proof. Check that the formula for product is well defined. Suppose $a + I = a' + I$ and $b + I = b' + I$, which implies

$$a' = a + x, \quad b' = b + y, \qquad x, y \in I.$$

Then

$$a'b' = (a + x)(b + y) = ab + (ay + xb + xy) \in ab + I.$$

Therefore $a'b' + I = ab + I$.

Checking the various axioms for R/I to be a ring is straightforward, using that they are true for R. □

The projection map $\pi \colon R \to R/I$ is a ring homomorphism, unital if R has 1.

72.2. *Example.* $\mathbb{Z}_n = \mathbb{Z}/(n)$, the quotient of integers by the ideal $(n) = \mathbb{Z}n$. We have already noticed this is a quotient group. In fact, it is a quotient ring.

72.3. *Example.* $R = \mathbb{Q}[x]/(f)$, where $f = x^2 - 2$.

Lecture 36

Write $I = (x^2 - 2) = \{\, (x^2 - 2)g \mid g \in \mathbb{Q}[x] \,\}$. An element of R has the form

$$g + I = a_n x^n + \cdots + a_1 x + a_0 + I, \qquad a_n, \ldots, a_0 \in \mathbb{Q},$$

and that a given element can be written like this in more than one way. We can use the division algorithm to find a "canonical form": for instance,

$$x^3 - 2x^2 + 5x - 3 = (x)(x^2 - 2) + (5x - 3),$$

so

$$x^3 - 2x^2 + 5x - 3 + I = 5x - 3 + I.$$

In general, every element of R is represented by

$$bx + a + I$$

for *unique* $a, b \in \mathbb{Q}$.

Addition of canonical forms gives a canonical form:

$$\big((ax + b) + I\big) + \big((a'x + b') + I\big) = (a + a')x + (b + b') + I.$$

Multiplication is more complicated:

$$(bx + a + I)(dx + c + I) = bcx^2 + (ad + bc)x + ac + I = (ad + bc)x + (ac + 2bc) + I.$$

Basically, we have the rule: $(x + I)(x + I) = x^2 + I = 2 + I$. In general, whenever we see x^2, we can replace it by 2.

Finish example.

72.4. *Example.* Consider the function $\phi\colon R \to \mathbb{R}$ defined by "evaluation at $\sqrt{2}$":

$$\phi(g + I) := g(\sqrt{2}).$$

We need to check that this is well-defined, i.e., if $g + I = h + I$, then $g(\sqrt{2}) = h(\sqrt{2})$. We know that $g + I = h + I$ means $g - h \in I = (x^2 - 2)$, so there exists a polynomial q such that

$$g(x) - h(x) = (x^2 - 2)q(x).$$

Plugging in $x = \sqrt{2}$ gives

$$g(\sqrt{2}) - h(\sqrt{2}) = ((\sqrt{2})^2 - 2)q(\sqrt{2}) = 0.$$

Thus $g(\sqrt{2}) = h(\sqrt{2})$.

Claim. The homomorphism ϕ is injective.

We need to show $\phi(g + I) = 0$ implies $g + I = 0 + I$, i.e., if $g(\sqrt{2}) = 0$ then $g = (x^2 - 2)q$ for some $q \in \mathbb{Q}[x]$. We know that because $\sqrt{2}$ is a root, we can factor off $(x - \sqrt{2})$ (why do we know this?), but of course this is only as real polynomials, not as rational ones, since $\sqrt{2}$ is irrational. You probably know that $-\sqrt{2}$ also has to be a root (but why do you know this?)

Let's think about the composite of homomorphisms

$$\mathbb{Q}[x] \xrightarrow{\pi} \mathbb{Q}[x]/(x^2 - 2) \xrightarrow{\phi} \mathbb{R}, \qquad g(x) \mapsto g(\sqrt{2})$$

which is exactly the *evaluation* homomorphism $\operatorname{eval}_{\sqrt{2}}\colon \mathbb{Q}[x] \to \mathbb{R}$.

Let $J := \operatorname{Ker}(\operatorname{eval}_{\sqrt{2}}) \subset \mathbb{Q}[x]$. By what we said above, we need to show that $J = I$. It is obvious that $I = (x^2 - 2) \subseteq J$, since $\operatorname{eval}_{\sqrt{2}}(x^2 - 2) = 0$.

By the classification of ideals in a $\mathbb{Q}[x]$, $J = (h)$ for some monic polynomial $h \in \mathbb{Q}[x]$. Because $x^2 - 2 \in J$, we know $x^2 - 2 = hq$ for some polynomial $q \in \mathbb{Q}[x]$. So h must be a factor of $x^2 - 2$ which has rational coefficients. In particular, $\deg(h) \in \{0, 1, 2\}$.

If $\deg(h) = 0$ it is constant, e.g., $h = 1$. But $\operatorname{eval}_{\sqrt{2}}(1) = 1 \neq 0$, so this is not possible.

If $\deg(h) = 1$, write $h(x) = x + b$. Then $q(x) = cx + d$ and

$$x^2 - 2 = (x + b)(cx + d) = cx^2 + (d + bc)x + bd.$$

Then $c = 1$, $d = -b$, and $-2 = bd = -b^2$. But this is not possible since 2 is irrational.

72.5. *Remark.* Let $S \subset \mathbb{R}$ be the subring

$$S = \{\, a + b\sqrt{2} \mid a, b \in \mathbb{Q} \,\} \subset \mathbb{R}.$$

You can check that S is really a subring: it is closed under addition and multiplication, and also has a multiplicative identity. The map ϕ gives an isomorphism between R and this subring S.

72.6. *Example.* (Important) $R = K[x]/I$, where $I = (f)$ for $f \in K[x]$ is any polynomial, with $\deg f = n \geq 0$.

Any element of R can be written *uniquely* as

$$a_{n-1}x^{n-1} + \cdots + a_1x + a_0 + I, \qquad a_i \in K.$$

We think of this as a *canonical form* of an element of R.

Addition in R then amounts to "vector addition" of the tuple $(a_0, \ldots, a_{n-1})$. Multiplication of canonical forms in R can be computed using polynomial long division: given

$$p + I, \qquad p' + I,$$

form the usual product $pp' \in K[x]$, then carry out $pp' \div f$ to get

$$pp' = qf + r, \qquad \deg r < n.$$

Then

$$(p + I)(p' + I) = r + I,$$

and $r + I$ is in canonical form.

72.7. *Example.* Consider $R = \mathbb{Q}[x]/(x^3+1)$. Every element of R can be written uniquely as $a + bx + cx^2 + (f)$ with $a, b, c \in \mathbb{Q}$.

73. Homomorphism theorems

73.1. **Theorem** (Homomorphism theorem for rings)**.** *Let $\phi\colon R \to S$ be a homomorphism of rings, and let $I \subseteq R$ be an ideal such that $I \subseteq \operatorname{Ker}\phi$. Then there exists a unique ring isomorphism $\overline{\phi}\colon R/I \to S$ such that $\overline{\phi}(a+I) = \phi(a)$ (i.e., such that $\overline{\phi}\circ\pi = \phi$, where $\pi\colon R \to R/I$ is the quotient homomorphism.)*

Proof. The proof is straightforward. The first step is to check that the formula $a + I \mapsto \phi(a)$ is well-defined. □

73.2. **Theorem** (Isomorphism theorem for rings)**.** *Let $\phi\colon R \to S$ be a surjective homomorphism of rings with $I = \operatorname{Ker}\phi$. Let $\pi\colon R \to R/I$ be the quotient homomorphism. Then there exists a unique ring isomorphism $\overline{\phi}\colon R/I \to S$ such that $\overline{\phi}\circ\pi = \phi$.*

73.3. *Example.* Let $\phi\colon \mathbb{R}[x] \to \mathbb{C}$ be the homomorphism defined by evaluation at $i \in \mathbb{C}$, and using the usual inclusion $\mathbb{R} \subset \mathbb{C}$. Thus $\phi(g) := g(i)$. This is surjective because $a + bi = \phi(a + bx)$.

$\operatorname{Ker}\phi$ = set of all polynomials g such that $g(i) = 0$. Clearly $x^2 + 1 \in \operatorname{Ker}(\phi)$, therefore $(x^2+1) \subseteq \operatorname{Ker}(\phi)$.

Given an arbitrary $g \in \mathbb{R}[x]$, by polynomial division we have $g = (x^2+1)q + (a + bx)$ for some $q \in \mathbb{R}[x]$. We have

$$\phi(g) = \phi(x^2+1)\phi(q) + (a + bx) = a + bi.$$

This is 0 iff $a = 0 = b$, so we see that $\operatorname{Ker}(\phi) = (x^2+1)$.

The homomorphism theorem gives a unital isomorphism $\mathbb{R}[x]/(x^2+1) \approx \mathbb{C}$ of rings.

73.4. *Example.* (I didn't do this one in class.) Let $\psi\colon \mathbb{R}[x] \to \mathbb{C}$ be the homomorphism defined by evaluation at $\omega = e^{2\pi i/3} = (-1 + i\sqrt{3})/2$.

This is surjective. Both $\mathbb{R}[x]$ and $\mathbb{C}$ are, in particular, real vector spaces over $\mathbb{R}$, and ψ is, in particular, an $\mathbb{R}$-linear map. The set $\{\psi(1) = 1, \psi(x) = \omega\}$ in $\mathbb{C}$ is linearly independent over $\mathbb{R}$. Since $\dim_{\mathbb{R}}\mathbb{C} = 2$, this means ψ is surjective.

Write $\operatorname{Ker}(\psi) = (f)$. The homomorphism theorem gives an isomorphism

$$\mathbb{R}[x]/(f) \xrightarrow{\sim} \mathbb{C}$$

of rings. By counting dimensions, we see that $\deg f = 2$. The set (f) is the collection of all polynomials over $\mathbb{R}$ which have ω as a root.

There is a degree 2 polynomial with ω as a root, for instance $f = x^2 + x + 1$.

We get an isomorphism $\mathbb{R}[x]/(x^2+x+1) \approx \mathbb{C}$ of rings.

74. Domains

Let R be a commutative ring with 1.

An **integral domain** (or **domain**) is a commutative ring with 1 such that $1 \neq 0$, and such that $xy = 0$ impies either $x = 0$ or $y = 0$. **integral domain** **domain**

74.1. *Example.* Examples of domains are $\mathbb{Z}$, fields K.

$K[x]$ with K a field (because $\deg(fg) = \deg(f)\deg(g)$.

Any subring (with 1) of a domain is a domain. For instance, the **Gaussian integers** **Gaussian integers**

$$\mathbb{Z}[i] := \{\, a + bi \mid a, b \in \mathbb{Z} \,\} \subset \mathbb{C}.$$

74.2. *Example.* Let $f_1, f_2 \in K[x]$ be two non-constant polynomials, and let $g = f_1 f_2$ and $I = (g)$. The ring $K[x]/I$ is not a domain, since $\overline{f}_1 = f_1 + I, \overline{f}_2 = f_2 + I$ are non-zero, but $\overline{f}_1\overline{f}_2 = 0$.

Domains have cancellation.

74.3. **Proposition** (Cancellation). *If R is a domain, and $ab = ac$, then $b = c$.*

75. Fields of fractions

Every integral domain is a subring of a field, called its field of fractions.

Given a domain R, consider the set

$$S = \{\, (a, b) \in R \times R \mid b \neq 0 \,\}.$$

We'll write "a/b" instead of (a, b) for this pair: right now we think of "a/b" as a "formal symbol", rather than an actual fraction. Say that

$$a/b \sim a'/b' \qquad \text{iff} \qquad ab' = ba'.$$

(Remember that $\frac{a}{b} - \frac{a'}{b'} = \frac{ab' - a'b}{bb'}$.)

75.1. **Lemma.** *This is an equivalence relation on S.*

Proof.

- *Reflexive.* To see that $a/b \sim a/b$, note that $ab = ab$.
- *Symmetric.* If $a/b \sim a'/b'$, then $ab' = ba'$. But this means $a'b = b'a$, so $a'/b' \sim a/b$.
- *Transitive.* If $a/b \sim a'/b'$ and $a'/b' \sim a''/b''$, then

  $$ab' = ba', \qquad a'b'' = b'a''.$$

  We can combine these to get

  $$ab'b'' = (ab')b'' = (ba')b'' = b(a'b'') = b(b'a'') = bb'a''.$$

  Because the ring is commutative, we can rewrite this as $b'(ab'') = b'(ba'')$. Because $b' \neq 0$ and R is a domain, we can cancel to get $ab'' = ba''$, which implies $a/b \sim a''/b''$.

□

Let $\mathrm{Frac}(R) :=$ the set of equivalence classe under this relation. We write "$[a/b]$" in $\mathrm{Frac}(R)$ for the equivalence class of (a, b). Note that the equivalence relation then says that $[a/b] = [a'/b']$ iff $ab' = ba'$.

Recall the definition of domain (and add condition that $1 \neq 0$). Recall the definition of the set $\mathrm{Frac}(R)$ for a domain R.

Define operations

$$[a/b] + [c/d] := [(ad + bc)/bd], \qquad [a/b][c/d] = [ac/bd].$$

Check that these are compatible with the equivalence relation (Exercise!), and so are well-defined operations on $\mathrm{Frac}(R)$.

75.2. **Proposition.** *If R is a domain, then* $\mathrm{Frac}(R)$ *is a field. The function $a \mapsto [a/1]$ is a unital homomorphism, injective.*

Proof. This is just straightforward, and is partially an exercise. Note: the 0 element is $[0/1]$, the 1 element is $[1/1]$.

Do multiplicative inverses. If $x \in \mathrm{Frac}(R)$ is not equal to 0, write $x = [a/b]$. Because $x \neq 0$, we have $[a/b] \neq [0/1]$, i.e., $a1 \neq b0$, i.e., $a \neq 0$. So set $y = [b/a] \in \mathrm{Frac}(R)$ (this makes sense exactly because $a \neq 0$, and check that $[a/b][b/a] = [ab/ab] = [1/1] = 1$. □

We can identify R with its image, a subring in $\mathrm{Frac}(R)$: we have a ring homomorphism $R \to \mathrm{Frac}(R)$ by $a \mapsto [a/1]$.

75.3. *Example.* If $R = \mathbb{Z}$, then $\mathrm{Frac}(\mathbb{Z}) = \mathbb{Q}$.

75.4. *Example.* If $R = K[x]$, then $\mathrm{Frac}(K[x])$ is called the field of **rational functions in one variable**. Elements are $f(x)/g(x)$ where f, g are polynomials. **rational functions in o variable**

I defined degree for polynomials over a field, but in fact this works for polynomials over a domain. We define it the same way: for $f \in R[x]$ with $f = \sum_{k=0}^{n} a_k$ with $a_n \neq 0$, we have $\deg(f) = n$, while $\deg(0) = -\infty$. **Lecture 37**

75.5. **Proposition.** *Let R be a domain, and $f, g \in R[x]$. Then*

1. $\deg(f \pm g) \leq \max\{\deg(f), \deg(g)\}$.
2. $\deg(fg) = \deg(f)\deg(g)$.

Proof. Proved exactly as over a field. Note that we need R to be a domain to prove (2). □

76. Irreducible elements of a domain

Let R be a domain.

Given $a, b \in R$, we say a **divides** b and a **is a factor of** b if there exists $c \in R$ such that $b = ac$. (I.e., if "$b/a \in \mathrm{Frac}(R)$" is actually contained in the subring R.) We write $a \mid b$. **a divides b** **a is a factor of b**

76.1. *Example.* When $R = \mathbb{Z}$, this coincides exactly with the notion of divisibility we discussed before.

76.2. *Example.* When $R = K[x]$ with K a field, this is the usual notion of divisibility of polynomials that you may be familiar with. For instance, in $\mathbb{R}[x]$ we have that $x - 2$ divides $x^2 - 4$, since $x^2 - 4 = (x - 2)(x + 2)$.

Note: I'm going to write $Ra := \{\, ra \mid r \in R \,\}$. Because R is commutative, this is also the principal ideal (a).

Important observation. $a \mid b$ iff $b \in Ra$ iff $Rb \subseteq Ra$ (inclusion of principal ideals). This is often the best way to think of this.

Here is a categorization of elements in a domain R.

- The zero element $0 \in R$.
- A **unit** $a \in R$ is an element which has a multiplicative inverse. Recall that we write the subset of units as $R^\times \subseteq R$, and that $R^\times$ is a group under multiplication. **unit**
- A **reducible** $a \in R$ is a non-zero non-unit such that there exist $b, c \in R$ such that $a = bc$ and neither b or c is a unit. (Neccesarily b and c are non-zero, since R is a domain.) **reducible**
- An **irreducible** $a \in R$ is a non-zero non-unit which is not reducible. That is, if $a = bc$, then either b or c must be a unit. **irreducible**

This is a *partition* of the domain R: an element is in exactly one of these four classes.

76.3. *Example.* For $R = \mathbb{Z}$, we have

- 0.
- $\mathbb{Z}^\times = \{1, -1\}$.
- Irreducible elements $= \{\, \pm p \mid p \in \mathbb{N}$ a prime number $\}$.
- Reducible elements: the rest, i.e., the composite integers (both positive and negative).

76.4. *Example.* For $R = K$ a field, we have

- 0.
- $K^\times = K \smallsetminus \{0\}$.
- Irreducible elements = none.
- Reducible elements = note.

For instance, this is tha case if $K = \mathbb{Q}$.

76.5. *Example.* For $R = K[x]$ with K a field, we have

- 0.
- $K[x]^\times = \{\, f(x) = c \in K[x] \mid c \in K \smallsetminus \{0\} \,\}$. I.e., units are the non-zero constant polynomials.

- Irreducible elements = set of *irreducible polynomials*, i.e., polynomials of degree ≥ 1 which do not factor into product of polys of smaller degree.
- Reducible elements = set of polynomials of degree ≥ 1 which do factor into polys of smaller degree.

For instance, if $K[x] = \mathbb{R}[x]$, then the four classes are

- 0.
- Units = $c \in \mathbb{R} \smallsetminus \{0\}$.
- Irreducible = $\{\, ax+b \mid a,b \in \mathbb{R},\ a \neq 0 \,\} \cup \{\, ax^2+bx+c \mid a,b,c \in \mathbb{R},\ a \neq 0,\ b^2-4ac < 0 \,\}$.
- Reducible = Everything else.

This is a fact you are probably aware of, though the proof is not completely obvious. The proof amounts to the fact that (i) every real polynomial splits into degree 1 factors over $\mathbb{C}$, and (ii) any non-real roots of a real polynomial come in conjugate pairs.

76.6. *Remark.* These notions are *relative* to the domain you are in. For instance, we have the sequence of domains

$$\mathbb{Q}[x] \subseteq \mathbb{R}[x] \subseteq \mathbb{R}(x),$$

where $\mathbb{R}(x) = \operatorname{Frac}(\mathbb{R}[x])$ is the field of frations of $\mathbb{R}[x]$. Consider $f = x^2 - 2$. In $\mathbb{Q}[x]$ this is irreducible, while in $\mathbb{R}[x]$ it is reducible, and in $\mathbb{R}(x)$ it is a unit.

We can define an equivalence relation on the non-zero elements: say $a, b \in R \smallsetminus \{0\}$ are **related up-to-units**, or are **associate**, and write $a \sim b$, if there exists $u \in R^\times$ such that $b = ua$. (*Exercise:* check that this is an equivalence relation.

related up-to-units
associate

76.7. *Example.* In $\mathbb{Z}$, the up-to-units classes are pairs $\{n, -n\}$ for $n \neq 0$. We can always pick the positive element if we want a standard representative.

In $K[x]$, the equivalence class of f is $\{\, cf \mid c \in \mathbb{R} \smallsetminus \{0\} \,\}$. Each up-to-units class contains exactly one monic polynomial.

Easy fact: if $a \sim b$, then a and b are in the same class (i.e., are both units, both reducible, or both irreducible).

77. Gaussian integers

Let $R = \mathbb{Z}[i] = \{\, a+bi \in \mathbb{C} \mid a,b \in \mathbb{Z} \,\}$, the ring of **Gaussian integers**. This is a subring of $\mathbb{C}$ (with 1), so is a domain. (Check this.) Thought of as a subset of the plane, it is the set of lattice points.

Gaussian integers

There is a function $N\colon R \to \mathbb{Z}$ defined by

$$N(a+bi) := (a+bi)(a-bi) = a^2+b^2 = \|a+bi\|^2 .$$

This is called a **norm** function. The following are easy to check.

norm

(1) $N(x) \in \mathbb{Z}_{\geq 0}$ for all $x \in R$.
(2) $N(x) = 0$ iff $x = 0$.
(3) $N(xy) = N(x)N(y)$, i.e., N is multiplicative.
(4) $N(x) = 1$ iff $x \in R^\times$.

Here's a proof of (4): $N(a+bi) = 1$ exactly means $(a+bi)(a-bi) = 1$, so $(a+bi)^{-1} = a-bi$. Conversely, if $x \in R^\times$ then $xx^{-1} = 1$ then $1 = N(1) = N(x)N(x^{-1})$, so $N(x) \in \mathbb{Z}^\times$. By (1) this implies $N(x) = 1$.

We deduce from this

- $R^\times = \{1, -1, i, -i\}$. (Only solutions to $1 = N(a+bi) = a^2+b^2$ with $a, b \in \mathbb{Z}$.)
- The element 2 is reducible in R, since $2 = (1+i)(1-i)$.

- The elements $x = 1 + i$ and $x' = 1 - i$ are irreducible. (Proof: If $1 \pm i = yz$ for $y, z \in R$, then $2 = N(1 \pm i) = N(y)N(z)$. Because 2 is prime in $\mathbb{Z}$, one of $N(y)$ or $N(z)$ must be 1, so either y or z is a unit.)
  The up-to-units class of $1 + i$ is $\{\pm 1 \pm i\}$ (four elements).
- The element 3 is irreducible in R. (Proof: If $3 = yz$ then $9 = N(y)N(z)$. But there are *no* $y \in R$ with $N(y) = 3$, because $y = a + bi$ implies $n(y) = a^2 + b^2$, and $a^2, b^2 \in \{0, 1, 4, 9, \dots\}$.)
  The up-to-units class of 3 is $\{\pm 3, \pm 3i\}$.
- The element 4 is reducible in R, since $4 = 2^2$. In R it actually reduces furthere: $4 = (1 + i)^2(1 - i)^2$.
- The element 5 is reducible, because $5 = (2 + i)(2 - i)$.
  The elmements $x = 2 + i$ and $y = 2 - i$ are irreducible, because $N(x) = N(y) = 5$ which is prime in $\mathbb{Z}$.

Note that an prime integer $p \in \mathbb{Z}$ may or may not still be irreducible in $\mathbb{Z}[i]$. In fact, we have the following.

77.1. **Proposition.** *A prime number $p \in \mathbb{Z}$ is reducible in $\mathbb{Z}[i]$ if and only if there exist $a, b \in \mathbb{Z}$ such that $p = a^2 + b^2$.*

Proof. First, suppose $p = a^2 + b^2$. Then $p = (a + bi)(a - bi)$, and since $N(a \pm bi) = p > 1$, we see that both $a \pm bi$ are not units in $\mathbb{Z}[i]$. Thus p is reducible in $\mathbb{Z}[i]$.

Now, suppose $p = xy$ for some non-units $x, y \in \mathbb{Z}[i]$. Then $p^2 = N(p) = N(x)N(y)$, and since x, y are not units we must have $N(x) = N(y) = p$. So if we write $x = a + bi$ we have $p = N(x) = a^2 + b^2$. □

Going down the list of prime numbers, we find:

reducible in $\mathbb{Z}[i]$	irreducible in $\mathbb{Z}[i]$
$2 = 1^2 + 1^2$	3
$5 = 2^2 + 1^2$	7
$13 = 3^2 + 2^2$	11
$17 = 4^2 + 1^2$	19
$29 = 5^2 + 2^2$	23
$37 = 6^2 + 1^2$	31
$41 = 5^2 + 4^2$	43
$53 = 7^2 + 2^2$	47

Do you see the pattern here?

78. Prime elements in a domain

Lecture 38
prime

There is also the notion of **prime** element of a domain. Say $p \in R$ is prime if it is a non-unit non-zero with the property that if $p \mid ab$ then p divides at least one of a or b. (Equivalently, if $p \nmid a$ and $p \nmid b$, then $p \nmid ab$.)

Primes are a subset of irreducibles.

78.1. **Proposition.** *Every prime element is irreducible.*

Proof. Suppose p is prime and $p = ab$; I'll show that at least one of a or b is a unit. Then $p \mid ab$ so p divides one of the factors, say $p \mid a$. Thus $a = pc$ for some c, so $p = ab = pcb$, so $1 = cb$, so b is a unit. □

The usual definition of "prime" in the integers (no positive factors other than 1 and self), is more-or-less the same as what we have called *irreducible.* This can be a little confusing. It turns out that all irreducible integers are also prime in the above sense, so it is ok. (We proved this in the first week, and we'll prove it again.)

Began to give example of $R = \mathbb{Z}[\sqrt{-5}]$, where 2 is irreducible but not prime.

78.2. *Example.* Let $R = \mathbb{Z}[\sqrt{5}] = \{ a + b\sqrt{-5} \in \mathbb{C} \mid a, b \in \mathbb{Z} \}$.

Fact. 2 is irreducible in R, but not prime.

Idea. There is a function $N\colon R \to \mathbb{Z}$ defined by

$$N(a + b\sqrt{-5}) := (a + b\sqrt{-5})(a - b\sqrt{-5}) = a^2 + 5b^2.$$

This is called a **norm** function, and it has the following properties:

norm

(1) $N(x) \in \mathbb{Z}_{\geq 0}$ for any $x \in R$.
(2) $N(x) = 0$ iff $x = 0$.
(3) $N(xy) = N(x)N(y)$, i.e., N is multiplicative. (This is just a verification.)
(4) $N(x) = 1$ if and only if $x \in R^\times$.

Note that if $a, b \in \mathbb{Z}$, then $a^2, b^2 \in \{0, 1, 4, 9, \dots\}$. So possible values of $N(a + b\sqrt{-5}) = a^2 + 5b^2$ are: $0, 1, 4, 5, 6, 9, \dots$. In particular, $N(x) \neq 2$ and $N(x) \neq 3$.

Use this to show 2 is irreducible: if $2 = xy$, then

$$4 = N(2) = N(x)N(y),$$

since N can't give 2, one of the factors is a unit.

In fact, all four elements 2, 3, $1 + \sqrt{5}$, $1 - \sqrt{5}$ are irreducible, using this argument.

But we have

$$6 = (2)(3) = (1 + \sqrt{-5})(1 - \sqrt{-5}),$$

so $2 \mid 6$. But $2 \nmid 1 \pm \sqrt{-5}$. (Proof: suppose $1 + \sqrt{-5} = 2x$ for some $x \in R$. Take norms to get $6 = 4N(x)$, which is impossible.)

79. Principal ideal domains

A **principal ideal domain (PID)** is a domain such that every ideal is principal.

principal ideal doma (PID)

Examples:

- Any field K.
- $\mathbb{Z}$.
- $K[x]$ where K is a field.

The fact that both $\mathbb{Z}$ and $K[x]$ are PIDs means we can prove some facts about both using the same proof.

79.1. *Example.* The ring $R = \mathbb{Z}[x]$ of polynomials with coefficients in integers is *not* a PID. For instance, $I = (2, x) = \{ 2g + xh \mid g, h \in \mathbb{Z}[x] \}$ is an ideal, and it turns out there is no $f \in \mathbb{Z}[x]$ such that $(f) = (2, x)$. (Idea of proof: think about the gcd of the coefficients of f.)

Here's a reformulation of some our basic notions in terms of principal ideals.

79.2. **Lemma.** *Let R be a domain and $a \in R$ with $a \neq 0$.*

(1) *$a \in R^\times$ iff $Ra = R$.*
(2) *a is associate to b iff $Ra = Rb$.*
(3) *a is prime iff $bc \in Ra$ implies either $b \in Ra$ or $c \in Ra$.*
(4) *a is irreducible iff (i) $Ra \neq R$, (ii) if $b \in R$ is such that $Ra \subseteq Rb \subseteq R$, then either $Ra = Rb$ or $Rb = R$.*

Proof. Exercise. □

79.3. **Proposition.** *If R is a PID, then a is prime iff it is irreducible.*

Proof. We've shown prime implies irreducible always.

Conversely, suppose p is irreducible; we show p is prime. Suppose $ab \in Rp$ and $a \notin Rp$; we want to show $b \in Rp$. Consider the ideal $I := (p, a) = Rp + Ra$. We have $Rp \subsetneq I \subseteq R$ (since $a \notin Rp$), so $I = R$. Thus $1 = up + va$ for some $u, v \in R$. Now consider

$$b = 1b = (up + va)b = upb + v(ab).$$

The right hand side is in Rp, since $ab \in pR$. Thus $b \in Rp$. □

Many domains are not PIDs: for instance, $\mathbb{Z}[x]$ or $K[x, y]$ are not PIDs. Another example is $R = \mathbb{Z}[\sqrt{-5}]$, which is not a PID because $I = (2, 1 + \sqrt{-5})$ is not a principal ideal. (Note: this is a fact which requires proof.)

79.4. *Example.* Let $I = (2, 1 + \sqrt{-5}) \subseteq \mathbb{Z}[\sqrt{-5}]$. Then I is not a principal ideal.

To show this, we suppose $I = (z)$ for some $z \in \mathbb{Z}[i]$ and derive a contradiction. If this is true, then since $2, 1 + \sqrt{-5} \in I$, there must exist $x, y \in \mathbb{Z}[\sqrt{-5}]$ such that

$$2 = zx, \qquad 1 + \sqrt{-5} = zy.$$

Taking norms gives

$$4 = N(z)N(x), \qquad 6 = N(z)N(y).$$

Since these are positive integers, we must have $N(z) \mid 2$, i.e., either $N(z) = 2$ or $N(z) = 1$.

But $a^2 + 5b^2$ is never 2 if $a, b \in \mathbb{Z}$. So we must have $N(z) = 1$, i.e., z is a unit so $I = R$.

So I just have to show that $1 \notin I$. If $1 \in I$, then we can write

$$1 = 2(a + b\sqrt{-5}) + (1 + \sqrt{-5})(c + d\sqrt{-5}), \qquad a, b, c, d \in \mathbb{Z}.$$

This simplifies to

$$1 = (2a + c - 5d) + (2b + c + d)\sqrt{-5}.$$

So we need to solve the pair of equations $1 = 2a + c - 5d$, $0 = 2b + c + d$ in integers. But there is no such solution: if we add the equations, we get $1 = 2(a + b + c - 2d)$ which is impossible.

80. $\mathbb{Z}[i]$ is a PID

80.1. **Theorem.** *The Gaussian integers $\mathbb{Z}[i]$ are a PID.*

We prove this by showing that $\mathbb{Z}[i]$ admits a "division algorithm". Recall the norm function $N : \mathbb{Z}[i] \to \mathbb{Z}_{\geq 0}$ by $N(a + bi) = a^2 + b^2 = \|a + bi\|^2$.

80.2. **Proposition.** *If $u, v \in \mathbb{Z}[i]$ with $v \neq 0$, there exist $q, r \in \mathbb{Z}[i]$ such that*

$$u = qv + r, \qquad N(r) < N(v).$$

Proof. We use the fact that $\mathbb{Z}[i] \subseteq \mathbb{C}$, so it actually makes sense to divide elements. Also note that in this plane, elements of $\mathbb{Z}[i]$ correspond to the "integer lattice" in the plane: i.e., to points whose x and y coordinates are integers. We are going to use the following "geometric" fact: every $z \in \mathbb{C}$ is within a distance $\leq \sqrt{2}/2 < 1$ from some $q = a + bi \in \mathbb{Z}[i]$.

In fact, we apply this to $z := u/v \in \mathbb{C}$, so we can choose $q = a + bi \in \mathbb{Z}[i]$ so that $\|z - q\| < 1$. Thus

$$\left\|\frac{u}{v} - q\right\| = \left\|\frac{u - qv}{v}\right\| = \frac{\|u - qv\|}{\|v\|}$$

is < 1, so $\|u - qv\| < \|v\|$. If we write $r := u - qv$ and take squares, this gives

$$N(r) < N(v)$$

as desired. □

80.3. *Remark.* This proof fails for $\mathbb{Z}[\sqrt{-5}]$, which does not admit such division with remainder, and in fact is not a PID. The reason it fails is because there are points in $\mathbb{C}$ which can be a distance > 1 from points in $\mathbb{Z}[\sqrt{-5}]$. For instance, when $z = (1 + \sqrt{-5})/2$, the closest points in $\mathbb{Z}[\sqrt{-5}]$ are $0, 1, \sqrt{-5}, 1 + \sqrt{-5}$ and all are a distance $\sqrt{6}/2 > 1$ from z.

80.4. **Corollary.** *$\mathbb{Z}[i]$ is a PID.*

Proof. Let $I \subseteq \mathbb{Z}[i]$ be a non-trivial ideal. Choose $v \in I \smallsetminus \{0\}$ which has minimal $N(v) \in \mathbb{Z}_{>0}$ among elements in this set. Then show that $I = (v)$, by the same pattern of proof we have seen for $\mathbb{Z}$ and $K[x]$. (I.e., if $u \in I$, then write $u = qv + r$ with $q, r \in \mathbb{Z}[i]$ and $N(r) < N(v)$, so by minimality $r = 0$ so $u \in (v)$.) □

81. Unique factorization domains

A **unique factorization domain (UFD)** is a domain R such that every non-zero non-unit has factorization into irreducibles, unique up to order and multiplication by units.

Lecture 39
unique factorizati domain (UFD)

That is, if $a \in R \smallsetminus \{0\}$, then if f is not a unit, we can write

$$a = p_1 \cdots p_m, \qquad m \geq 1$$

for some sequence of irreducible elements p_i in R. Furthermore, if

$$a = p_1 \cdots p_m = q_1 \cdots q_n, \qquad m, n \geq 1,$$

with all p_i, q_j irreducible, then $m = n$ and you can reorder the qs so that $q_i = p_i u_i$ for some $u_i \in \text{Units}(R)$.

81.1. *Example.* The integers are a UFD: For instance, we have factorizations

$$-15 = (-3)(5) = (5)(-3) = \dots$$

These all count as "the same" factorization of -15.

81.2. *Example.* A ring $K[x]$ of polynomials over a field is a UFD. (We will prove this soon.)

81.3. *Example.* The ring $R = \mathbb{Z}[\sqrt{-5}]$ is *not* a UFD. As we noted, we have

$$6 = (2)(3) = (1 + \sqrt{-5})(1 - \sqrt{-5}),$$

where $2, 3, 1 \pm \sqrt{-5}$ all irreducible. But $2 \not\sim (1 \pm \sqrt{-5})$ (for instance, because $R^\times = \{1, -1\}$). So factorization into irreducibles is not unique in $\mathbb{Z}[\sqrt{-5}]$. (It is true that factorization into irreducibles *exists* in this domain.)

81.4. *Example.* Let $R \subseteq \mathbb{Q}[x]$ be the subset consisting of polynomials whose constant term is an integer, i.e., $f = \sum_k a_k x^k$ with all $a_k \in \mathbb{Q}$ and $a_0 \in \mathbb{Z}$. Then the element $x \in R$ has no irreducible factorizations in R, so R is not a UFD. To see this, note that:

- $R^\times = \mathbb{Z}^\times = \{\pm 1\}$.
- x is not irreducible, since in R we can always write $x = (2)(\frac{1}{2}x)$.
- More generally, for any $a \in \mathbb{Q}$ the element ax is not irreducible in R, for the same reason.
- Using properties of the degree function on $\mathbb{Q}[x]$, you can see that any factorization of x must have the form

  $$x = a_1 \cdots a_{m-1} f, \qquad a_1, \dots, a_{m-1} \in \mathbb{Z}, \qquad f = bx, \quad b \in \mathbb{Q}.$$

  In particular, such a factorization always involves a reducible element f, so there is no irreducible factorization.

82. PIDs are UFDs

The big theorem (6.5.19) in the book, is the following:

82.1. **Theorem.** *Every PID is a UFD.*

This theorem implies, from what we already know, that both $\mathbb{Z}$ and $K[x]$ are UFDs. It also implies that $\mathbb{Z}[i]$ is a UFD (because it is a PID.) The proof is actually very much like the proof that $\mathbb{Z}$ is a UFD, which we proved at the beginning of the course.

Proof. We need to show (i) that every non-zero non-unit of of R is equal to a product of irreducibles, and (ii) that this factorization is unique up to units.

(i) Say $a \in R$ is *bad* if a is a non-zero non-unit which is not a product of irreducible. We want to show that there are no bad elements, which I'll do by contradiction. First note that if a is bad it cannot be irreducible, so $a = bc$ for some non-units b, c. Furthermore at least one of b or c must be bad, since otherwise they would both be products of irreducible and thus so is a.

So we can inductively choose factorizations

$$a = a_1 b_1, \quad a_1 = a_2 b_2, \quad a_2 = a_3 b_3,$$

where all a_k, b_k are non-units and all a_k are bad. we get a chain of ideals

$$Ra \subseteq Ra_1 \subseteq Ra_2 \subseteq Ra_3 \subseteq \cdots .$$

Note that $Ra_k \neq Ra_{k+1}$ because b_k is a non-unit.

Let $I := \bigcup_{k=1}^{\infty} Ra_k$, which we have shown is an ideal. Since R is a PID, we have $I = (c)$ for some $c \in R$. But since I is a union, $c \in Ra_k$ for some k, but then $I \subseteq Ra_k \subseteq Ra_{k+1} \subseteq I$, so $Ra_k = Ra_{k+1}$, a contradiction. We have proved (i).

(ii) This is exactly like the proof for $\mathbb{Z}$. We use an induction on the length of the shortest factorization. It relies on the fact we proved that in a PID, every irreducible is prime. Thus if

$$a = p_1 \cdots p_r = q_1 \cdots q_s$$

are two factorizations into irreducibles, then $p_r \mid q_1 \cdots q_s$ implies $p_r \mid q_i$ for some i because p_r is prime, and therefore $q_i = cp_r$ with $c \in R^\times$ (since q_i is irreducible). Now cancel the p_rs to get

$$b = p_1 \cdots p_{r-1} = q_1 \cdots q_{i-1} q_{i+1} \cdots (cq_s).$$

By induction these are the same up to reordering and units.

□

83. Fermat's theorem on sums of two squares

Question. Which integers m are the sum of two squares of integers? That is, given m, does there exist $a, b \in \mathbb{Z}$ such that $m = a^2 + b^2$?

As we have seen, this question is related to the Gaussian integers. This is because $m = a^2 + b^2$ implies that you can factor m in $\mathbb{Z}[i]$ as $m = (a + bi)(a - bi) = a^2 + b^2$.

Because of the relation to factorization, it turns out it is easiest to answer the question for integers which are prime.

Question (special case). Which *prime* integers p are the sum of two squares?

In this case, we know that these are *exactly* the integer primes p which are reducible in $\mathbb{Z}[i]$.

Here are the easy facts.

- For $p = 2$, we have $2 = 1^2 + 1^2$.
- For odd p with $p \equiv 3 \mod 4$ this is *not* possible, because $a^2, b^2 \in [0]_4 \cup [1]_4$ but $p \in [3]_4$. (In fact, no integer in $[3]_4$ is a sum of two squares.)
- The remaining case is odd primes p such that $p \equiv 1 \mod 4$. In this case, it turns out that all such primes are sums of two squares, by a theorem of Fermat.

83.1. **Theorem** (Fermat). *If $p = 4n + 1$ is a prime integer, then $p = a^2 + b^2$ for some $a, b \in \mathbb{Z}$.*

We are going to prove this using the fact that the Gaussian integers $\mathbb{Z}[i]$ are a PID, and so irreducibles in $\mathbb{Z}[i]$ are also prime. (This proof is due to Dedekind (1877).) These ideas will prove the following proposition.

83.2. **Proposition.** *If $p = 4n + 1$ is a prime integer, then it is reducible in the Gaussian integers $\mathbb{Z}[i]$.*

Given this, the proof of the Theorem is not hard.

Proof of Theorem using the Proposition. We have already shown that if a prime number $p \in \mathbb{Z}$ is reducible in $\mathbb{Z}[i]$, then p is a sum of two squares. Explicitly, if $p = xy$ for two non-unit $x, y \in \mathbb{Z}[i]$, then $N(p) = p^2$ implies $N(x) = N(y) = p$, so if we write $x = a+bi$ with $a, b \in \mathbb{Z}$, then $p = N(x) = a^2+b^2$.

Note that if $x = a + bi$, then we must have $y = a - bi$, since $xy = a^2 + b^2 = p$. □

To prove the proposition, we need a lemma, which was in the homework. It turns out to be equivalent to a fact about the field $\mathbb{Z}_p$.

83.3. **Lemma** (Lagrange)**.** *If $p = 4n + 1$ is a prime integer, then there exists $m \in \mathbb{Z}$ such that $p \mid (m^2 + 1)$. (Divisibility in $\mathbb{Z}$.)*

Proof. Another way to say this: for such primes $p = 4n + 1$, we want an integer m such that $m^2 \equiv -1 \mod p$. Or said another way: If p is a prime congruent to 1 mod 4, then $\mathbb{Z}_p$ contains a square root of -1. Left as a (non-obvious) exercise, which appeared on PS 6. □

83.4. *Remark.* There is another way to prove this. On the optional problem set, exercise (7) asks you to prove that if K is a field, then any finite subgroup $G \leq K^\times$ of its unit group is cyclic. In particular, this means that if $K = \mathbb{Z}_p$, then $\mathbb{Z}_p^\times$ is cyclic, or order $p - 1$. Therefore if $4 \mid p - 1$ there will be an element in this group of order 4, which is therefore a square-root of -1.

Before giving the proof of the proposition, let's note something about ideals in $R = \mathbb{Z}[i]$. If $p \in \mathbb{Z}$ is an integer, then the principal ideal that it generates in $\mathbb{Z}[i]$ has the form

$$(p) = Rp = \{\, a + bi \mid a, b \in \mathbb{Z}p \,\}.$$

This is just because for $c+di \in \mathbb{Z}[i]$ (with $c, d \in \mathbb{Z}$) we have $(c+di)p = (cp)+(dp)i$, so if $a+bi \in (p)$ (with $a, b \in \mathbb{Z}$) then both integers a and b are divisible by p as integers.

Proof of Proposition. Let $p = 4n + 1$ be a prime integer. By the lemma, there exists $m \in \mathbb{Z}$ such that $p \mid (m^2 + 1)$ (divisibility in $\mathbb{Z}$). This implies that p also divides $m^2 + 1$ in the Gaussian integers $\mathbb{Z}[i]$. (Given $m^2 + 1 = pk$ with $k \in \mathbb{Z}$, this is also a factorization of $m^2 + 1$ in $\mathbb{Z}[i]$.)

But in $\mathbb{Z}[i]$ we can factor $m^2 + 1 = (m + i)(m - i)$. We suppose p irreducible and derive a contradiction.

Because $\mathbb{Z}[i]$ is a PID, irreducibles are primes, so we must have $p \mid m + i$ or $p \mid m - i$ in $\mathbb{Z}[i]$. But either of these is impossible: e.g., as noted above, $p \mid m + i$ implies $p \mid 1$ as integers.

Thus, p cannot be irreducible in $\mathbb{Z}[i]$. □

84. General sums of two squares

We can use this to solve the general question about which integers are a sum of two squares. **Lecture 40**

84.1. **Proposition.** *A positive integer m is a sum of two squares of integers if and only if its prime factorization (in $\mathbb{Z}$) $m = p_1^{k_1} \cdots p_r^{k_r}$ is such that: if $p_i \equiv -1 \mod 4$, then k_i is even.*

In other words, a positive m is *not* a sum of two squares if it has an odd number of factors of some prime p such that $p \equiv -1 \mod 4$. Thus, the integers ≤ 50 which are not a sum of two squares are:

3, 6, 7, 11, 12, 14, 15, 19, 21, 22, 23, 24, 27, 28, 30, 31, 33, 35, 38, 39, 42, 43, 44, 46, 47, 48.

On the other hand, we have

$$\begin{array}{rrrrr}
1 = 1^2 + 0^2, & 9 = 3^2 + 0^2, & 18 = 3^2 + 3^2, & 32 = 4^2 + 4^2, & 41 = 5^2 + 4^2, \\
2 = 1^2 + 1^2, & 10 = 3^2 + 1^2, & 20 = 4^2 + 2^2, & 34 = 5^2 + 3^2, & 45 = 6^2 + 3^2, \\
4 = 2^2 + 0^2, & 13 = 3^2 + 2^2, & 25 = 5^2 + 0^2, & 36 = 6^2 + 0^2, & 49 = 7^2 + 0^2, \\
5 = 2^2 + 1^2, & 16 = 4^2 + 0^2, & 26 = 5^2 + 1^2, & 37 = 6^2 + 1^2, & 50 = 7^2 + 1^2. \\
8 = 2^2 + 2^2, & 17 = 4^2 + 1^2, & 29 = 5^2 + 2^2, & 40 = 6^2 + 2^2, &
\end{array}$$

Proof. Let

$$S = \{\, m \in \mathbb{Z}_{>0} \mid m = a^2 + b^2 \text{ for some } a, b \in \mathbb{Z} \,\}$$

and

$$T = \{\, m \in \mathbb{Z}_{>0} \mid m = p_1^{k_1} \cdots p_r^{k_r},\ p_i \text{ distinct primes, } k_i \text{ even if } p_i \equiv -1 \mod 4 \,\}.$$

We want to show $S = T$.

First note that S is closed under multiplication. In fact, S is the image of $N(\mathbb{Z}[i] \smallsetminus \{0\}) \subseteq \mathbb{Z}$, since $m = a^2 + b^2$ iff $m = N(a + bi)$. So the claim follows because N preserves multiplication.

Note also that T is closed under multiplication: if m and n each have an even number of factors of a prime p, then so does their product mn.

Proof that $T \subseteq S$. It suffices to show for each prime integer p, that (i) $p \in S$ if $p \not\equiv -1 \pmod 4$, and (ii) $p^2 \in S$ if $p \equiv -1 \pmod 4$. Case (i) is just Fermat's theorem, while case to is obvious, since $p^2 = p^2 + 0^2$.

Proof that $S \subseteq T$. Suppose $m \in S$. We prove $m \in T$ by induction on m. Note that $1, 2 \in T$, so we have a base case.

Write $m = a^2 + b^2$. If m has no prime factors (in $\mathbb{Z}$) which are $\equiv -1 \mod 4$ then $m \in T$. So suppose there is a $p \equiv -1 \mod 4$ which divides m in $\mathbb{Z}$, and therefore divides it in $\mathbb{Z}[i]$. Then since $m = (a + bi)(a - bi)$ we have $p \mid (a + bi)(a - bi)$ in $\mathbb{Z}[i]$. We know that p is irreducible in $\mathbb{Z}[i]$, and thus prime since $\mathbb{Z}[i]$ is a PID. Therefore p must divide one of the factors, either $p \mid a + bi$ or $p \mid a - bi$. In either case, this implies that $p \mid a$ and $p \mid b$ in $\mathbb{Z}$, since p is actually an integer as we have noted above. Thus $a + bi = p(c + di)$ for some $c, d \in \mathbb{Z}$, so

$$m = (a + bi)(a - bi) = p^2(c + di)(c - di) = p^2(c^2 + d^2).$$

Now $n = c^2 + d^2 \in S$ and $n < m$, so by the inductive hypothesis $n \in T$. We know $p^2 \in T$, and therefore $m = np^2$ is also in T. □

85. Primes in Gaussian integers

We now know that $R = \mathbb{Z}[i]$ is a UFD. We can give a complete classification of the irreducible (=prime) elements in R, up-to-units.

85.1. **Lemma.** *Let $u \in R = \mathbb{Z}[i]$ be an irreducible element. Then $I := Ru \cap \mathbb{Z}$ is generated by a prime number p.*

Proof. It is straightforward to prove that I is an ideal in $\mathbb{Z}$. Since $\mathbb{Z}$ is a PID, we have $I = \mathbb{Z}p$ for some $p \in \mathbb{Z}$. Note also that $I \neq 0$, since $N(u) = u\overline{u} \in I$ is non-zero. Thus we can assume $p > 0$.

To show that p is a prime number, it suffices to show that it is a prime element of $\mathbb{Z}$, i.e., that $p \mid ab$ and $p \nmid a$ imply $p \mid b$. Equivalently, we want to show $ab \in I$, $a \notin I$ imply $b \in I$. But $I \subseteq Ru$, and u is irreducible and hence prime in $\mathbb{Z}[i]$. So we have $u \mid ab$ and $u \mid a$, whence $u \mid b$. This means $b \in I \cap \mathbb{Z} = Rp$, so $p \mid b$ as desired. □

Thus, for any irreducible $u \in R$, we have $Ru \cap \mathbb{Z} = \mathbb{Z}p$ for a *unique* prime number p. This p is the only prime number which is divisible by u in $\mathbb{Z}[i]$. We say that u **lies over** p. **lies over**

Furthermore, since $N(u) = u\overline{u} \in Ru \cap \mathbb{Z}$, we know that the prime p which u lies over must divide $N(u)$. Using this, we can easily compute examples of irreducibles in $\mathbb{Z}[i]$ and the prime integers they lie over.

$$\begin{tikzcd}
\mathbb{Z}[i]: & 1+i \arrow[d, dash] & 3 \arrow[d, dash] & \begin{matrix}2+i\\2-i\end{matrix} \arrow[d, dash] & 7 \arrow[d, dash] & 11 \arrow[d, dash] & \begin{matrix}3+2i\\3-2i\end{matrix} \arrow[d, dash] & \begin{matrix}4+i\\4-i\end{matrix} \arrow[d, dash] & 19 \arrow[d, dash] \\
\mathbb{Z}: & 2 & 3 & 5 & 7 & 11 & 13 & 17 & 19
\end{tikzcd}$$

Note that $1 + i \sim 1 - i$, but that the other pairs $a + bi, a - bi$ in the diagram are not associate to each other. (Though there are other elements associate to these: the associates of $a + bi$ are $-a - bi$, $-b + ai$, $b - ai$.)

85.2. **Proposition.** *The following is a complete list of irreducibles in $\mathbb{Z}[i]$ up-to-units.*

(1) $1+i$. *(This lies over 2.)*
(2) *For each prime number p with $p \equiv -1 \pmod 4$, the element $p \in \mathbb{Z}[i]$. (This lies over p.)*
(3) *For each prime number p with $p \equiv 1 \pmod 4$, the elements $a+bi$ and $a-bi$ in $\mathbb{Z}[i]$, with $p = a^2+b^2$ and $a > b > 0$. (This lies over p.)*

Proof. Using Fermat's theorem on sums-of-squares and N, we see that each of the listed elements is irreducible in $\mathbb{Z}[i]$. So we just have to show they are the only possibilities up-to-units.

Suppose $u = a+bi$ is any irreducible in $R = \mathbb{Z}[i]$. Let p be the prime number u lies over, so $Ru \cap \mathbb{Z} = \mathbb{Z}p$, and thus $p = uv$ for some $v \in \mathbb{Z}[i]$. We thus have

$$p^2 = N(p) = N(u)N(v).$$

We consider cases according to $N(u)$.

- $N(u) = 1$ is not possible, since then u is a unit.
- $N(u) = p$, $N(v) = p$. Then $p = a^2+b^2$, and by Fermat we know that either $p = 2$ or $p \equiv 1 \pmod 4$.

  If $p = 2$ then $u \in \{1+i, 1-i, -1+i, -1-i\}$, by considering the possible solutions of $2 = a^2+b^2$ in integers. All four of these are associate.

  If p odd, then $a \neq \pm b$ (since otherwise we would have $2 \mid a^2+b^2 = p$). WLOG we can assume $a > b > 0$, since if not then we can replace u with $\pm iu$ to get this. As we have seen, the two possibilites are not associate.
- $N(u) = p^2$, $N(v) = 1$, so $v \in \{\pm 1, \pm i\}$ and thus $u \in \{\pm p, \pm ip\}$. By Fermat we know that if $p = 2$ or $p \equiv -1 \pmod 4$ then p is reducible in $\mathbb{Z}[i]$, so in this case we must have $p \equiv -1 \pmod 4$.

□

85.3. *Example.* Let's find an irreducible factorization of $u = 3+9i$ in $\mathbb{Z}[i]$. Note that $N(u) = 3^2+9^2 = 90 = 2 \cdot 3^2 \cdot 5$. This means that candidate irreducible factors of u are: $1+i$, 3, $2 \pm i$, and in fact both $1+i$ and 3 *must* be irreducible factors, as these are the only primes in $\mathbb{Z}[i]$ (up-to-units) with norm 2 or 3. Only one of $2 \pm i$ can be a factor, and it's not immediately clear which, but dividing through by the factors we know, we find that

$$3+9i = (1+i)3(2+i).$$

86. Another example: Eisenstein integers

Let $\omega = e^{2\pi i/3} = (-1+i\sqrt{3})/2$. This has the property that $\omega^3 = 1$, and in fact $\omega^3 - 1 = (\omega-1)(\omega^2+\omega+1)$, we have $\omega^2+\omega+1 = 0$, so $\omega^2 = -1-\omega$.

The **Eisenstein integers** is the set

Eisenstein integers

$$R = \mathbb{Z}[\omega] := \{\, a+b\omega \mid a, b \in \mathbb{Z} \,\} \subseteq \mathbb{C}.$$

This is a subring of $\mathbb{C}$, so a domain.

86.1. *Remark.* It is helpful to draw a picture of how $\mathbb{Z}[\omega]$ sits inside $\mathbb{C}$ viewed as a plane. Eisenstein integers correspond to points in a hexagonal lattice inside $\mathbb{C}$.

We have a norm function $N\colon \mathbb{Z}[\omega] \to \mathbb{Z}$, defined by

$$N(a+b\omega) = \|a+b\omega\|^2 = (a+b\omega)(a+b\omega^2) = a^2 - ab + b^2.$$

This satisfies the same properties we have seen:

- $N(z) \geq 0$ for $z \in \mathbb{Z}[\omega]$, with $N(z) = 0$ iff $z = 0$.
- $N(zw) = N(z)N(w)$ and $N(1) = 1$.
- $N(z) = 1$ iff $z \in \mathbb{Z}[\omega]$.

86.2. **Proposition.** *We have $\mathbb{Z}[\omega]^\times = \{\pm 1, \pm\omega, \pm\omega^2\} = \langle -\omega\rangle$, a cyclic group of order 6.*

Proof. We just have to solve $a^2 - ab + b^2 = 1$ in integers. To do this, complete the square:

$$1 = a^2 - ab + b^2 = a^2 - ab + \tfrac{1}{4}b^2 + \tfrac{3}{4}b^2 = (a - \tfrac{1}{2}b)^2 + \tfrac{3}{4}b^2.$$

It is immediate that we must have $b^2 \le 1$, and using this we can work out the solutions case by case:

- If $b = 0$, then $a = \pm 1$.
- If $b = 1$, then $a = 0, 1$.
- If $b = -1$, then $a = 0, -1$.

□

86.3. **Proposition.** *If $u, v \in \mathbb{Z}[\omega]$ with $v \neq 0$, then there exist $q, r \in \mathbb{Z}[\omega]$ with $u = qv + r$ and $N(r) < N(v)$.*

Proof. This is proved just as in $\mathbb{Z}[i]$, using the fact that every point in $\mathbb{C}$ has a distance $\le \sqrt{3}/3 < 1$ from some point in $\mathbb{Z}[\omega]$. □

86.4. **Corollary.** *$\mathbb{Z}[\omega]$ is a PID.*

Proof. Proved exactly as for $\mathbb{Z}[i]$. □

86.5. **Proposition.** *If $p \in \mathbb{N}$ is a prime number, and if p is reducible in $\mathbb{Z}[\omega]$, then $p = N(a + b\omega) = a^2 - ab + b^2$ for some $a, b \in \mathbb{Z}$.*

Proof. Write $p = uv$ with u, v non-units, so $p^2 = N(p) = N(u)N(v)$. Then $N(u) = N(v) = p$, so if $u = a + b\omega$ with $a, b \in \mathbb{Z}$ we have $a^2 - ab + b^2 = p$. (Note that this implies $v = a + b\omega^2 = a - b - b\omega$.) □

Here are some examples.

- 2 is irreducible in $\mathbb{Z}[\omega]$, since $2 = a^2 - ab + b^2$ has no solution in $\mathbb{Z}$.
- $3 = (2 + \omega)(2 + \omega^2) = (2 + \omega)(1 - \omega)$, so 3 is reducible in $\mathbb{Z}[\omega]$. Note that the two factors are associate: $\omega(2 + \omega) = 1 - \omega$.
- 5 is irreducible in $\mathbb{Z}[\omega]$, since $5 = a^2 - ab + b^2$ has no solution in $\mathbb{Z}$.
- $7 = (3 + \omega)(3 + \omega^2) = (3 + \omega)(2 - \omega)$, so 7 is reducible in $\mathbb{Z}[\omega]$. In this case the factors are not associate.

86.6. *Remark.* For every $a, b \in \mathbb{Z}$, the expression $a^2 - ab + b^2$ is never $\equiv -1 \pmod 3$. I don't know a clever way to prove this, but it is enough to verify 9 cases:

$$\begin{aligned} &0^2 - 0\cdot 0 + 0^2 \equiv 0, \quad 1^2 - 1\cdot 0 + 0^2 \equiv 1, \quad 2^2 - 2\cdot 0 + 0^2 \equiv 1,\\ &0^2 - 0\cdot 1 + 1^2 \equiv 1, \quad 1^2 - 1\cdot 1 + 1^2 \equiv 1, \quad 2^2 - 2\cdot 1 + 1^2 \equiv 0,\\ &0^2 - 0\cdot 2 + 2^2 \equiv 1, \quad 1^2 - 1\cdot 2 + 2^2 \equiv 0, \quad 2^2 - 2\cdot 2 + 2^2 \equiv 1. \end{aligned}$$

86.7. **Lemma.** *Let p be a prime number such that $p = 3n + 1$. Then there exists $m \in \mathbb{Z}$ such that $p \mid (m^2 + m + 1)$.*

Proof. Consider the field $\mathbb{Z}_p$. On the optional homework, we have shown that its group of units $\mathbb{Z}_p^\times = \Phi(p)$ is cyclic. Since this has order $p - 1$, which by hypothesis is divisible by 3, there exists $[m] \in \mathbb{Z}_p$ which has order 3 in $\mathbb{Z}_p^\times$. Thus $p \mid m^3 - 1$, and also $p \nmid m - 1$ (since $[m]$ does not have order 1 in $\mathbb{Z}_p^\times$), so since $p \mid (m-1)(m^2 + m + 1)$ we must have $p \mid m^2 + m + 1$ as desired. □

86.8. **Proposition.** *If $p = 3n + 1$ is a prime integer, then it is reducible in the Eisenstein integers $\mathbb{Z}[\omega]$.*

Proof. Let $p = 3n + 1$ a prime integer. By the lemma, there exists $m \in \mathbb{Z}$ such that $p \mid (m^2 + m + 1)$ in $\mathbb{Z}$, but therefore also $p \mid (m^2 + m + 1)$ in $\mathbb{Z}[\omega]$. But we can factor $m^2 + m + 1 = (m - \omega)(m - \omega^2) = (m - \omega)(m + 1 + \omega)$ in $\mathbb{Z}[\omega]$. We suppose p irreducible in $\mathbb{Z}[\omega]$ and derive a contradiction.

Because $\mathbb{Z}[\omega]$ is a PID, irreducibles are primes, so we must have $p \mid m - \omega$ or $p \mid m + 1 + \omega$. But either of these are impossible, since either would imply $p \mid 1$ in $\mathbb{Z}$. So p is not irreducible in $\mathbb{Z}[\omega]$. □

86.9. **Theorem.** *If p is a prime number, then $p = a^2 - ab + b^2$ for some $a, b \in \mathbb{Z}$ if and only if $p \equiv 1 \pmod 3$.*

Proof. We have seen that if $p \equiv -1 \pmod 3$, then we cannot write $p = a^2 - ab + b^2$. Also $p = 3$ cannot be written this way. So the only primes where this can be possible are the ones $\equiv 1 \pmod 3$.

But we have shown in this case that p is reducible in $\mathbb{Z}[\omega]$, so $p = uv$ for some $u, v \in \mathbb{Z}[\omega]$ with $N(u) = p$, which gives p in this form. □

As in $\mathbb{Z}[i]$, each irreducible $u \in R = \mathbb{Z}[\omega]$ "lies over" some prime number in $\mathbb{Z}$, in the sense that $Ru \cap \mathbb{Z} = \mathbb{Z}p$ for a unique prime number p (same proof). Using this, we can classify irreducibles in $\mathbb{Z}[\omega]$ up-to-units.

86.10. **Proposition.** *The following is a complete list of irreducibles in $\mathbb{Z}[\omega]$ up-to-units.*

1. $2 + \omega$ *(which lies over* 3*). (This is associate to $2 + \omega^2 = 3 - \omega$.)*
2. *For each prime number with $p \equiv -1 \pmod 3$, the element $p \in \mathbb{Z}[\omega]$. (This lies over p.)*
3. *For each prime number with $p \equiv 1 \pmod 3$, the elements $a + b\omega$ and $a + b\omega^2 = a - b - b\omega$, where $p = a^2 - ab + b^2$ and $a > b > 0$. (These are not associate to each other.)*

87. Ideal numbers

The domain $R = \mathbb{Z}[\sqrt{-5}]$ looks similar to the Gauss and Eisenstein integers, but does not have unique factorization. For instance:

$$6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$$

are two distinct irreducible factorizations of 6. (Note that 2 and 3 are not the same as $1 \pm \sqrt{-5}$ up-to-units, since $R^\times = \{\pm 1\}$.)

However, there is a kind of way to "force" unique factorization in R, by introducing additional "ideal numbers", which are not actually numbers, but can be made to make sense for the purposes of factorizations. (But only up-to-units, and you cannot define addition of ideal numbers so that they behave like elements of a ring.)

In the example of $R = \mathbb{Z}[\sqrt{-5}]$, let's introduce three new symbols: P, Q, Q', which will thought of as irreducible, and specify that

$$2 \sim P^2, \qquad 3 \sim QQ', \qquad 1 + \sqrt{-5} \sim PQ, \qquad 1 - \sqrt{-5} \sim PQ',$$

where "$\sim$" means "same up-to-units". These imply that

$$6 \sim P^2QQ'$$

With this new "ideal numbers", $2, 3, 1 \pm \sqrt{-5}$ are no longer irreducible, and instead 6 gets a unique factorization into irreducible ideal numbers.

This gives a kind of solution, but may seem kind of arbitrary. Dedekind gave an interpretation of "ideal numbers", which led to the definition of ideal in a ring. Dedekind observed that the only thing you ever need to know about an ideal number is the set of elements of R that it divides. For instance, in the above example, it turns out that

$$\{\, a \in R \mid P \text{ divides } a \,\} = \{\, 2x + (1 + \sqrt{-5})y \mid x, y \in R \,\}.$$

Notice that this is actually an ideal in R. Thus, the correct definition of "ideal number" is exactly: non-zero ideal in R.

So for a domain R, let $\mathrm{Ideal}(R)$ be the set of non-zero ideals in R. We can define a multiplication on this set by

$$IJ := (\{\, ab \mid a \in I,\ b \in I \,\}),$$

the ideal generated by all products ab with $a \in I$ and $b \in J$. This is easy to compute if you have generating sets for your ideals. For instance

$$I = (a_1, \dots, a_p), \quad J = (b_1, \dots, b_q) \qquad \Longrightarrow \qquad IJ = (a_1b_1, \dots, a_pb_q).$$

There is an obvious function

$$R \smallsetminus \{0\} \to \text{Ideal}(R), \qquad a \mapsto (a)$$

which sends an element to its principal ideal, and this function is multiplicative: $(a)(b) = (ab)$. Note: it is not injective, but it is true that $(a) = (b)$ iff $a \sim b$.

For a certain class of rings called **Dedekind domains**, it turns out that Ideal(R) has unique factorization, in the sense that every ideal is a finite product of "irreducibles", uniquely up-to-reordering. An ideal is irreducible in Ideal(R) if it is not a product of two non-unit ideals. (It turns out that $P \in \text{Ideal}(R)$ is irreducible in this sense if and only if the quotient ring R/P is a field.)

Dedekind domains

The ring $R = \mathbb{Z}[\sqrt{-5}]$ is a Dedekind domain, so we have unique factorization for ideals. In fact, if we write

$$P := (2,\, 1+\sqrt{-5}) = (2,\, 1-\sqrt{-5}), \qquad Q := (3,\, 1+\sqrt{-5}), \qquad Q' := (3, 1-\sqrt{-5}),$$

then you can compute:

$$\begin{aligned}
&(2,\, 1+\sqrt{-5})(2,\, 1-\sqrt{-5}) = (4,\, 2-2\sqrt{-5},\, 2+2\sqrt{-5},\, 6) = (2),\\
&(3,\, 1+\sqrt{-5})(3,\, 1-\sqrt{-5}) = (9,\, 3+3\sqrt{-5},\, 3-3\sqrt{-5},\, 6) = (3),\\
&(2,\, 1-\sqrt{-5})(3,\, 1+\sqrt{-5}) = (6,\, 2+2\sqrt{-5},\, 3-3\sqrt{-5},\, 6) = (1+\sqrt{-5}),\\
&(2,\, 1+\sqrt{-5})(3,\, 1-\sqrt{-5}) = (6,\, 2-2\sqrt{-5},\, 3+3\sqrt{-5},\, 6) = (1-\sqrt{-5}).
\end{aligned}$$

Department of Mathematics, University of Illinois, Urbana, IL
E-mail address: rezk@illinois.edu