ALGEBRAIC COMBINATORICS II, SUMMER 2024

INSTRUCTOR: YU-WEI FAN

Course Description. Group theory is a fundamental mathematical concept that elucidates the symmetry of objects, finding applications across various fields of modern science. The course will start with an exploration of the symmetries exhibited by common geometric objects in two-dimensional and three-dimensional Euclidean spaces. We will then progress to introduce the definition and basic principles of groups, subsequently applying these concepts to concrete combinatorial and geometric objects.

1. Overview of the course

Lecture 1

We will explore the *symmetries* of various *geometric spaces* in this course. The spaces that we will consider include: the Euclidean spaces $\mathbb{R}^2$, $\mathbb{R}^3$, the spheres S^1, S^2, the hyperbolic space $\mathbb{H}^2$, and some of their interesting subsets.

Question 1.1. *Which of the following shapes is more "symmetric"?*

Question 1.2. *How to define "symmetries"?*

Each of the geometric spaces that we will consider ($\mathbb{R}^2$, $\mathbb{R}^3$, S^1, S^2, $\mathbb{H}^2$, etc.) has a natural metric (i.e. distance $d(x, y)$ between any two points x, y). The symmetries that we are interested in are the *isometries* (i.e. distance-preserving functions) of these spaces. For instance, an isometry of $\mathbb{R}^2$ is a function $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that $d(f(x), f(y)) = d(x, y)$ for any $x, y \in \mathbb{R}^2$.

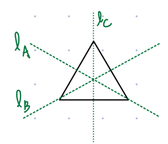
Definition 1.3. Let $S \subseteq \mathbb{R}^2$ be a subset of $\mathbb{R}^2$. An isometry $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ is called a *symmetry of S* if we have $f(S) = S$, i.e.

- for any $p \in S$, we have $f(p) \in S$; and

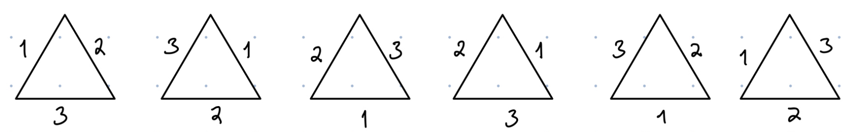
- for any $q \in S$, there exists $p \in S$ such that $f(p) = q$.

Example. Let us look at an easy example: an equilateral triangle. It has two kinds of symmetries:

- Rotational symmetries: one can rotate the triangle by $\frac{2\pi}{3}$, $\frac{4\pi}{3}$, or 2π without changing its appearance.
- Reflection symmetries: there are three "mirror lines" through which we can reflect the shape without changing its appearance.

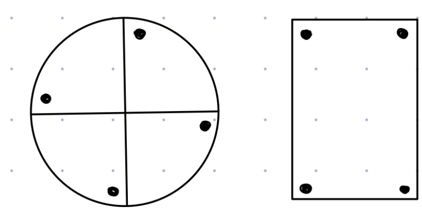


The easiest way to study the symmetries of a shape is by *counting*. In this example, it's easy to check that there are 6 symmetries. If we put labels on the edges of the triangle, then the effect of these symmetries look like:



However, counting alone is usually not good enough.

Example. Both of the following shapes have 4 symmetries. The shape on the



left has 4 rotational symmetries (by $\frac{\pi}{2}$, π, $\frac{3\pi}{2}$, 2π), but no reflection symmetries. In contrast, the shape on the right has 2 rotational symmetries and 2 reflection symmetries. How can we distinguish them?

As we'll see later in this course, *group theory* provides rigorous tools to describe the symmetries of shapes. For any shape (or any geometric object), the set of its symmetries has a natural *group structure*. In the example above, although the sets of symmetries of both shapes have 4 elements, but their underlying group structures are different, and that's how we can tell them apart (e.g. consider the *orders* of elements in these two groups).

Another important tool that we will encounter is basic *linear algebra*, in particular *matrices* or *matrix groups*. The reason is that certain matrix groups ($\mathrm{O}(2,\mathbb{R})$, $\mathrm{O}(3,\mathbb{R})$, $\mathrm{SL}(2,\mathbb{R})$, $\mathrm{SL}(2,\mathbb{C})$, etc.) act naturally as isometries on the spaces that we are interested in like $\mathbb{R}^2$, $\mathbb{R}^3$, S^1, S^2, $\mathbb{H}^2$. For instance, you'll show in the homework that any isometry of the Euclidean space $\mathbb{R}^n$ is a composition of a translation and a linear transformation.

Now we mention some examples that we'll be studying in this course.

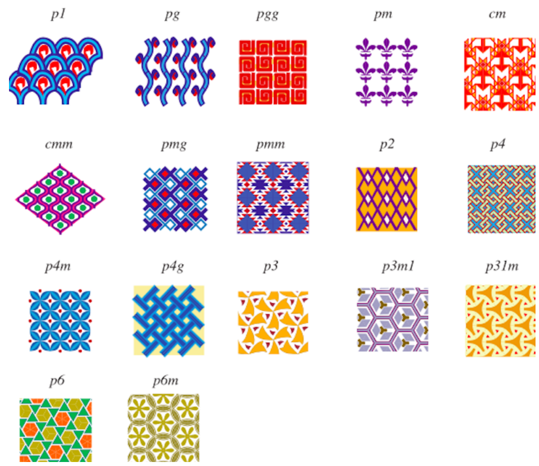
Example. Consider a regular n-gon P_n in $\mathbb{R}^2$. It is not hard to show that P_n has $2n$ symmetries. We'll:

- discuss the group structure of the symmetry group of P_n (the resulting group is called the *dihedral group* D_n;
- prove that any finite subgroup of $\mathrm{O}(2,\mathbb{R})$ is either a cyclic group or the symmetry group of a regular n-gon.

Example. An important class of examples of subsets in $\mathbb{R}^3$, that possess many symmetries, are the Platonic solids (regular polyhedrons). We will:

- prove that there are only five of them;
- study their symmetry groups;
- more importantly, prove that any finite subgroup of $\mathrm{SO}(3,\mathbb{R})$ is either a cyclic group, a dihedral group, or the (rotational) symmetry group of one of the Platonic solids.

Example. A *wallpaper* is a mathematical object that covers the whole $\mathbb{R}^2$ by repeating a pattern indefinitely. The symmetry group of a wallpaper is called a *wallpaper group* (we will provide more precise definition later on). We will show that there are exactly 17 different wallpaper groups.



In later parts of this course, we will discuss the isometries (or more generally, conformal maps) of some important non-Euclidean spaces, including the Riemann sphere S^2 and the hyperbolic plane $\mathbb{H}$.

After that, we will discuss more general notion of *topological spaces*, which have less structures and with more flexibility. In order to study the symmetries of topological spaces, we will introduce some notions from *algebraic topology*, which associated *groups* to topological spaces as invariants. As applications, we will give topological proofs of some interesting theorems, including the fundamental theorem of algebra.

2. A crash course on basic group theory

2.1. **Binary operators.** Before discussing the actual definition of a *group*, let us first consider a more general notion of *binary operators*.

Definition 2.1. Let S be a set. A *binary operator* on S is a function

$$\circ\colon S \times S \to S.$$

Example. Addition on the set of positive integers (denoted by $\mathbb{N}$), or the set of integers (denoted by $\mathbb{Z}$), or the set of rational numbers (denoted by $\mathbb{Q}$) or the set of real numbers (denoted by $\mathbb{R}$), is a binary operator. Same for multiplication.

Non-example. Subtraction on the set of positive integers is *not* a binary operator. Division on the set of integers is *not* a binary operator.

Definition 2.2. Let $(S, \circ)$ be a set with a binary operator. We say an element $e \in S$ is an *identity element* if $e \circ a = a \circ e = a$ for any $a \in S$.

Example. The element $0 \in \mathbb{Z}$ is an identity element of $(\mathbb{Z}, +)$. The element $1 \in \mathbb{Z}$ is an identity element of $(\mathbb{Z}, \times)$.

Non-example. $(\mathbb{N}, +)$ has no identity element.

Exercise. Prove that any set with a binary operator $(S, \circ)$ has at most one identity element.

Definition 2.3. Let $(S, \circ, e)$ be a set with a binary operator and an identity element. We say an element $a' \in S$ is an *inverse* of $a \in S$ if $a \circ a' = a' \circ a = e$.

Example. For $(\mathbb{Z}, +)$, the inverse of $a \in \mathbb{Z}$ is given by $-a$. For $(\mathbb{R}, \times)$, the inverse of $a \in \mathbb{R}$ is given by $1/a$, provided that $a \neq 0$.

Non-example. For $(\mathbb{Z}, \times)$, any element $a \in \mathbb{Z}$ has no inverse unless $a = \pm 1$.

Definition 2.4. Let $(S, \circ)$ be a set with a binary operator. We say $(S, \circ)$ is *associative* if $(a \circ b) \circ c = a \circ (b \circ c)$ holds for any $a, b, c \in S$.

Exercise. Let $(S, \circ, e)$ be a set with an associative binary operator and an identity element. Prove that any element in S has at most one inverse.

Most of the examples that we'll be discussing are associative. Here is a non-example (which we will not encounter in this course):

Non-example. The cross product $\times$ on $\mathbb{R}^3$ is *not* associative. Rather, it satisfies the *Jacobi identity*

$$\vec{v}_1 \times (\vec{v}_2 \times \vec{v}_3) + \vec{v}_2 \times (\vec{v}_3 \times \vec{v}_1) + \vec{v}_3 \times (\vec{v}_1 \times \vec{v}_2) = 0$$

Definition 2.5. Let $(S, \circ)$ be a set with a binary operator. We say $(S, \circ)$ is *commutative* if $a \circ b = b \circ a$ for any $a, b \in S$.

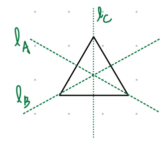
Warning. Many of the examples that we'll consider are *not* commutative.

Non-example. Consider the set of all six geometric transformations that give the symmetries of an equilateral triangle:

$$S = \left\{\text{rotate } 0, \text{ rotate } \frac{2\pi}{3}, \text{ rotate } \frac{4\pi}{3}, \text{ reflect along } \ell_A, \text{ reflect along } \ell_B, \text{ reflect along } \ell_C\right\}.$$

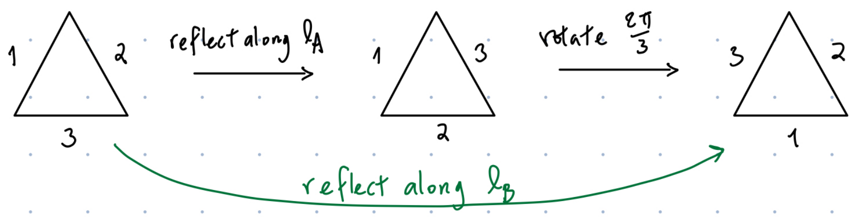
(note: rotations are typically assumed to be counterclockwise)
There is a binary operation on S given by *composing* these geometric transformations:

$$\circ \colon S \times S \to S,$$



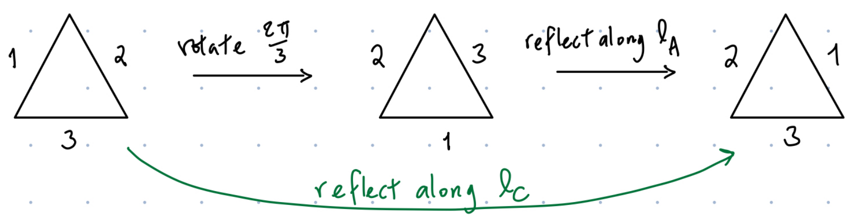
where $a \circ b \in S$ is the transformation given by "do b, and then do a". For instance, we have

$$\left(\text{rotate } \frac{2\pi}{3}\right) \circ \left(\text{reflect along } \ell_A\right) = \text{reflect along } \ell_B.$$



On the other hand, by reversing the order one gets

$$\left(\text{reflect along } \ell_A\right) \circ \left(\text{rotate } \frac{2\pi}{3}\right) = \text{reflect along } \ell_C.$$



This shows that $(S, \circ)$ is *not* commutative.

Non-example. Another important class of groups that we will discuss is the *matrix groups*. They are *not* commutative in most cases.

2.2. **Groups.**

Definition 2.6. Let $(G, \circ)$ be a set with a binary operator. It is called a *group* if it satisfies the following conditions:

(1) It is associative.

(2) It has the identity element (which will usually be denoted by e, e_G, 1, or 1_G).
(3) Any element $a \in G$ has an inverse (which will be denoted by $a^{-1} \in G$).

Remark 2.7. Here are some notions that we will be using frequently:

- If a group $(G, \circ)$ is commutative, then it is called an *abelian group*.
- We'll use $|G|$ to denote the number of elements in the set G, and will call it the *order* of G. Note that the order of a group could be infinite in general.
- We quite often would omit "$\circ$", and simply denote $a \circ b$ by ab, denote $a \circ a$ by a^2, denote $a \circ a \circ a$ by a^3, and so on.

Example. Consider the set of integers modulo n

$$\mathbb{Z}/n\mathbb{Z} \coloneqq \left\{\bar{0}, \bar{1}, \ldots, \overline{n-1}\right\}.$$

Addition and multiplication are well-defined on $\mathbb{Z}/n\mathbb{Z}$. It's not hard to show that $(\mathbb{Z}/n\mathbb{Z}, +)$ is an abelian group of order n, with the identity given by $\bar{0}$.

Example. Consider the subset of $\mathbb{Z}/n\mathbb{Z}$ consisting of elements that are coprime with n:

$$(\mathbb{Z}/n\mathbb{Z})^* \coloneqq \left\{\overline{m} \in \mathbb{Z}/n\mathbb{Z} \colon \gcd(m, n) = 1\right\}.$$

It's not hard to show that $((\mathbb{Z}/n\mathbb{Z})^*, \times)$ is an abelian group, with the identity given by $\bar{1}$.

Example. The set of all integers $\mathbb{Z}$ under addition is an example of an abelian group with infinite order.

Example. The set $\{0\}$ under addition is an example of a group with only one element (a trivial group).

Example. Let G_1 and G_2 be two groups. Consider the set

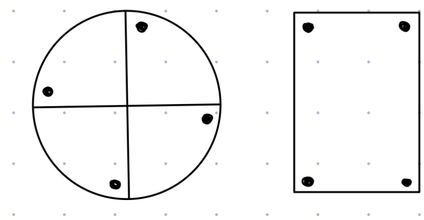
$$G_1 \times G_2 \coloneqq \left\{(g_1, g_2) \colon g_1 \in G_1 \text{ and } g_2 \in G_2\right\}.$$

Define a binary operator on $G_1 \times G_2$ as follows:

$$(g_1, g_2) \circ (g_1', g_2') \coloneqq (g_1 \circ g_1', g_2 \circ g_2').$$

It's not hard to show that $(G_1 \times G_2, \circ)$ is also a group. It's called the *direct product* of G_1 and G_2.

Example. Let's come back to the following examples again. As discussed ear-



lier, the symmetries of a shape form a group, where the binary operation is given by composition. The symmetry group of the first shape is

$$G_1 \coloneqq \left\{\text{rotate } 0,\ \text{rotate } \frac{\pi}{2},\ \text{rotate } \pi,\ \text{rotate } \frac{3\pi}{2}\right\}.$$

One thing we might notice about this group is that all elements of the group can be obtained by taking one element of the set, and combining it different number of times. Let's denote rotate $\frac{\pi}{2}$ by a. Then G_1 can be rewritten as

$$G_1 = \{e, a, a^2, a^3\}.$$

Notice that $a^4 = e$ since rotate 2π is the same as rotate 0, i.e. the identity map. The same is true for $\mathbb{Z}/4\mathbb{Z}$ (under addition) if one lets $a = \bar{1}$ and note that $a^4 = \bar{4} = \bar{0} = e$ in $\mathbb{Z}/4\mathbb{Z}$. In fact, we'll see that the symmetry group of the first shape and $\mathbb{Z}/4\mathbb{Z}$ are *isomorphic*, which means that they are essentially the same group.

On the other hand, the symmetry group of the second shape is

$$G_2 \coloneqq \left\{\text{rotate } 0,\ \text{rotate } \pi,\ \text{reflect along } \ell_1,\ \text{reflect along } \ell_2\right\}.$$

It's not hard to see that there is no element $a \in G_2$ such that $G_2 = \{e, a, a^2, a^3\}$. Therefore, G_2 and G_1 are not isomorphic. In fact, one can show that G_2 is isomorphic to the direct product $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$.

2.3. **Homomorphisms.** For any mathematical structure (like groups), it is crucially important to understand how two structures of the same type (like two groups) are related in a meaningful way. Functions that bridge such two structures are called *homomorphisms*. (In the Ancient Greek language, "homos" means "same", and "morphe" means "form" or "shape".) In general, a homomorphism is a function between two mathematical structures of the same type, that preserves the operations of the structures.

Definition 2.8. Let G and H be two groups. A function $f\colon G \to H$ is called a *homomorphism* if for any $g_1, g_2 \in G$ we have

$$f(g_1 g_2) = f(g_1) f(g_2)$$

Furthermore, a homomorphism that is both injective and surjective is called an *isomorphism*. In this case, we'll use the notation "$G \cong H$".

In other words, a homomorphism is a function that is compatible with the binary operations on the two groups.

Exercise. Let $f\colon G \to H$ be a group homomorphism. Prove that

- It preserves the identity: $f(e_G) = e_H$.
- It preserves the inverses: $f(g^{-1}) = f(g)^{-1}$ for any $g \in G$.

Example. We considered the symmetry group

$$G_1 := \left\{\text{rotate } 0,\ \text{rotate } \frac{\pi}{2},\ \text{rotate } \pi,\ \text{rotate } \frac{3\pi}{2}\right\} = \{e, a, a^2, a^3\}$$

where $a^4 = e$. One can define a function

$$G_1 \to \mathbb{Z}/4\mathbb{Z}$$

by sending $e \mapsto \bar{0}$, $a \mapsto \bar{1}$, $a^2 \mapsto \bar{2}$, and $a^3 \mapsto \bar{3}$. It's an easy exercise to show that this function is an isomorphism.

Remark 2.9. A convenient way to present a group is by choosing elements that *generate* the group (which means that any element of the group can be written as a product of some of these generators and their inverses), and a set of *relations* among these generators. For instance, $\mathbb{Z}/4\mathbb{Z}$ can be presented by

$$\mathbb{Z}/4\mathbb{Z} = \left\langle a\colon a^4 = e \right\rangle,$$

which means that one can find an element $a \in \mathbb{Z}/4\mathbb{Z}$ such that any element in $\mathbb{Z}/4\mathbb{Z}$ can be written as a power of a, and it satisfies $a^4 = e$ (it's not hard to see that a can be chosen to be $\bar{1}$ or $\bar{3}$ in this case).

Definition 2.10. A group G that can be generated by a single element g is called a *cyclic* group (i.e. any element of G is of the form g^k for some $k \in \mathbb{Z}$).

Definition 2.11. Let g be an element in a group G. If there exists a positive integer n such that $g^n = e$, then the smallest possible n satisfying $g^n = e$ is called the *order* of g. If such n does not exist, then we say g is of infinite order.

Exercise. Let G be a cyclic group, and say it can be generated by an element $g \in G$.

- If g is of finite order, say $\mathrm{order}(g) = n$. Prove that $G \cong \mathbb{Z}/n\mathbb{Z}$.
- If g is of infinite order, then prove that $G \cong \mathbb{Z}$.

Therefore, any cyclic group is isomorphic to either $\mathbb{Z}$ or $\mathbb{Z}/n\mathbb{Z}$ for some positive integer n.

Exercise. Prove that $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ is not a cyclic group.

Example. Let D_n be the symmetry group of a regular n-gon. It is not hard to show that D_n is generated by rotation by $2\pi/n$ (which we'll denote by r), and a reflection (which we'll denote by s). The group D_n is of order $2n$, with elements given by

$$D_n = \left\{1, r, r^2, \dots, r^{n-1}, s, sr, sr^2, \dots, sr^{n-1}\right\}.$$

The generators r and s satisfy the relations $r^n = s^2 = 1$ and $s^{-1}rs = r^{-1}$.

$$\begin{aligned} D_n &= \left\langle r, s \mid r^n = s^2 = 1,\ s^{-1}rs = r^{-1} \right\rangle \\ &= \left\langle r, s \mid r^n = s^2 = (rs)^2 = 1 \right\rangle. \end{aligned}$$

Remark 2.12. Since D_n is not commutative, it is not isomorphic to the direct product $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. On the other hand, it is isomorphic to the *semi-direct product* of its order 2 *subgroup* $\langle s \rangle$ and its order n *normal subgroup* $\langle r \rangle$: $D_n \cong \mathbb{Z}/2\mathbb{Z} \ltimes \mathbb{Z}/n\mathbb{Z}$. We'll introduce these notations later on.

2.4. **Subgroups.**

Definition 2.13. Let G be a group. We say a subset $H \subseteq G$ is a *subgroup* if:

(1) it is closed under the binary operation of G: for any $a, b \in H$, we have $ab \in H$;
(2) it contains the identity element of G: $e_G \in H$;
(3) it is closed under taking inverse: for any $a \in H$, we have $a^{-1} \in H$.

Exercise. A subgroup $H \subseteq G$ is itself a group, with the binary operator and the identity element inherit from G.

Example. For any group G, the subset $\{e_G\} \subseteq G$ is always a subgroup, called the *trivial* subgroup of G. Also, the group G itself is a subgroup of G.

Example. For any positive integer n, the subset $n\mathbb{Z} \subseteq (\mathbb{Z}, +)$ is a subgroup.

Exercise. Let G be a group and $g_1, \dots, g_n$ be elements of G. Prove that the following two statements are equivalent:

- any $g \in G$ can be written as $g = g_{i_1}^{a_1} g_{i_2}^{a_2} \cdots g_{i_k}^{a_k}$ for some $i_1, \dots, i_k \in \{1, \dots, n\}$ and $a_1, \dots, a_k \in \mathbb{Z}$;
- the smallest subgroup of G that contains $g_1, \dots, g_n$ is the group G itself.

In this case, we say $\{g_1, \dots, g_n\}$ *generates* the group G.

If H is a subgroup of G, then one can break G up into pieces, each of which looks like H. These pieces are called *cosets* of H, and they arise by "multiplying" H by elements of G.

Definition 2.14. Let G be a group and $H \subseteq G$ be a subgroup. A *left coset* of H in G is a subset of the form

$$gH = \{gh \mid h \in H\} \text{ for some } g \in G.$$

The element g is called a *representative* of the coset gH. The collection of all left cosets is denoted by G/H. Its order $|G/H|$ is called the *index* of H in G, and will sometimes be denoted by $[G : H]$.

Similarly, a *right coset* is a subset of the form

$$Hg = \{hg \mid h \in H\} \text{ for some } g \in G.$$

The collection of all right cosets is denoted by $H\backslash G$.

Example. Consider the subgroup $n\mathbb{Z} \subseteq (\mathbb{Z}, +)$. Since the group $(\mathbb{Z}, +)$ is abelian, its left cosets and right cosets are identical. It is clear that the subgroup has exactly n cosets $\bar{0}, \bar{1}, \dots, \overline{n-1}$, where $\bar{i} = i + n\mathbb{Z}$ consists of integers $\equiv i$ modulo n. Hence $n\mathbb{Z} \subseteq \mathbb{Z}$ is a subgroup of index n.

Lecture 2

Exercise. The representative of a coset is *not* unique. In fact, show that a coset gH can be represented by any element of the form gh where $h \in H$.

Exercise. Consider the subgroup $\mathbb{Z} \subseteq (\mathbb{R}, +)$. The set of cosets $\mathbb{R}/\mathbb{Z}$ can be identified with S^1, the unit circle in $\mathbb{R}^2$: Points of the circle are of the form $e^{2\pi i\theta}$ where $\theta \in \mathbb{R}$. Show that the map $t \mapsto e^{2\pi it}$ gives a bijection between $\mathbb{R}/\mathbb{Z}$ and S^1.

Proposition 2.15. *Let $H \subseteq G$ be a subgroup. Prove that for any two cosets aH and bH, we have:*

- *either aH and bH are disjoint: $aH \cap bH = \emptyset$,*

- *or aH and bH are exactly the same: $aH = bH$.*

Proof. Suppose aH and bH are not disjoint. Then there exists $h_1, h_2 \in H$ such that $ah_1 = bh_2$. For any $h \in H$, we have

$$ah = a(h_1 h_1^{-1})h = b(h_2 h_1^{-1} h) \in bH.$$

Hence $aH \subseteq bH$. Similarly, one can show that $bH \subseteq aH$. Therefore $aH = bH$. $\square$

Theorem 2.16 (Lagrange)**.** *Let G be a finite group, and $H \subseteq G$ be a subgroup. Then $|G|$ is divisible by $|H|$. Moreover, we have $|G| = |H|[G : H]$.*

Proof. Since $g \in gH$, any element of G belongs to a left coset of H. Then the previous proposition shows that G is the disjoint union of the left cosets of H. Since each coset has exactly $|H|$ elements, we can conclude that $|G| = |H|[G : H]$. $\square$

Exercise. Let G be a finite group and g be an element of G. Prove that the order of g divides the order of G.

Remark 2.17. In the example $n\mathbb{Z} \subseteq (\mathbb{Z}, +)$, one can notice that the set of all cosets $\{\bar{0}, \bar{1}, \dots, \overline{n-1}\}$ also has a natural group structure inherits from the group structure on $(\mathbb{Z}, +)$: one defines $\bar{i} + \bar{j}$ to be $\overline{i+j}$.

However, the set of all left cosets does *not* always admit a group structure! Let $H \subseteq G$ be a subgroup and $a, b \in G$ be two elements in G. It is tempting to define a group structure on G/H simply by declaring "$aH \circ bH = (ab)H$". In order for this definition to make sense, we need to show that, if a' is a representative of aH and b' is a representative of bH, then $a'b'H = abH$. This is equivalent to, for any $a, b \in G$ and $h_1, h_2 \in H$, one needs $ah_1bh_2H = abH$, or equivalently, $b^{-1}h_1b \in H$. This is equivalent to the condition that for any $g \in G$ one needs $gH = Hg$, i.e. the left and right cosets of H in G coincide, which is *not* true in general.

Definition 2.18. A subgroup $H \subseteq G$ is called *normal* if $gH = Hg$ for any $g \in G$.

By the previous remark, if $H \subseteq G$ is a normal subgroup, then the set of (left) cosets G/H admits a group structure inherit from G: let aH and bH be two cosets, then $aH \circ bH := (ab)H$ gives a well-defined group structure on G/H. The resulting group G/H is called the *quotient group*.

Theorem 2.19 (First isomorphism theorem)**.** *Let $f\colon G \to H$ be a group homomorphism. Define*

$$\mathrm{Ker}(f) \coloneqq \{g \in G \mid f(g) = 1_H\} \subseteq G$$

and

$$\mathrm{Im}(f) \coloneqq \{h \in H \mid h = f(g) \textit{ for some } g \in G\} \subseteq H.$$

Then

(1) $\mathrm{Ker}(f)$ *is a normal subgroup of* G*.*
(2) $\mathrm{Im}(f)$ *is a subgroup of* H*.*
(3) There is an isomorphism between $G/\mathrm{Ker}(f)$ *and* $\mathrm{Im}(f)$*.*

Proof. It is not hard to show that $\mathrm{Ker}(f) \subseteq G$ and $\mathrm{Im}(f) \subseteq H$ are subgroups (homework). To show that $\mathrm{Ker}(f) \subseteq G$ is normal, one needs to show that for any $g \in \mathrm{Ker}(f)$ and $g' \in G$, we have $g'gg'^{-1} \in \mathrm{Ker}(f)$. This is true because

$$f(g'gg'^{-1}) = f(g')f(g)f(g'^{-1}) = f(g')f(g')^{-1} = 1_H.$$

Now we define a map $\overline{f}$ from $G/\mathrm{Ker}(f)$ to $\mathrm{Im}(f)$: For any coset $g\mathrm{Ker}(f)$, we define $\overline{f}(g\mathrm{Ker}(f)) \coloneqq f(g)$. This is a well-defined function on the set of cosets $G/\mathrm{Ker}(f)$, because any representative of $g\mathrm{Ker}(f)$ is of the form gg' for some $g' \in \mathrm{Ker}(f)$, and we have $f(gg') = f(g)f(g') = f(g)$. It is not hard to check that $\overline{f}\colon G/\mathrm{Ker}(f) \to \mathrm{Im}(f)$ is a surjective group homomorphism. It is also injective: if $\overline{f}(g_1\mathrm{Ker}(f)) = \overline{f}(g_2\mathrm{Ker}(f))$, then we have $f(g_1) = f(g_2)$, or equivalently $g_2^{-1}g_1 \in \mathrm{Ker}(f)$. Hence the cosets $g_1\mathrm{Ker}(f) = g_2\mathrm{Ker}(f)$ coincide. □

2.5. **Symmetry groups.** For any set X, a *permutation* of X is a bijective function $f\colon X \to X$. The *symmetric group* S_X *defined over* X is the set of all permutations of X, equipped with the group structure given by compositions. In particular, when X is a finite set of n elements $\{1, 2, \dots, n\}$, its symmetric group would be denoted by S_n. It is not hard to see that $|S_n| = n!$.

Remark 2.20. One of the reasons that symmetric groups are important is that, any group is isomorphic to a subgroup of a symmetric group (Cayley's theorem). More specifically, one can show that any group G is isomorphic to a subgroup of the symmetric group S_G whose elements are the permutations of the underlying set of G. Explicitly, for each $g \in G$, we define a permutation of G (called left multiplication) $\ell_g\colon G \to G$ by $\ell_g(x) \coloneqq gx$. It is an easy exercise to check that the map $G \to S_G$ given by $g \mapsto \ell_g$ is an injective group homomorphism. Hence G is isomorphic to the image of $G \to S_G$, which is a subgroup of S_G. In particular, if G is a finite group of order n, then this argument shows that G is isomorphic to a subgroup of S_n.

Remark 2.21. Symmetric groups will also arise naturally when we discuss the symmetry groups of Platonic solids. Let G be the symmetry group of a tetrahedron T. It is not hard to see that any symmetry of T sends a vertex of T to a vertex (not necessarily the same one); in other words, it gives rise to a permutation of the four vertices of T. This gives a group homomorphism $\rho\colon \operatorname{Aut}(T) \to S_4$. Note that ρ is injective (why?), hence the symmetry group $\operatorname{Aut}(T)$ is isomorphic to a subgroup of the symmetric group S_4. (In fact, one can use the *orbit-stabilizer theorem* to show that $\operatorname{Aut}(T) \cong S_4$.)

Any element of S_n can be represented by Cauchy's "two-line notation". Let $\sigma \in S_n$ be a permutation of the set $\{1, 2, \ldots, n\}$. Then we'll write

$$\sigma = \begin{bmatrix} 1 & 2 & \cdots & n \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n). \end{bmatrix}$$

As usual, the composition $\sigma_1\sigma_2 \in S_n$ is given by $k \mapsto \sigma_1(\sigma_2(k))$, i.e. first apply σ_2 then apply σ_1. For instance, verify that

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

Permutations are also often written in *cycle notation* ("decomposition into disjoint cycles"). To write down $\sigma \in S_n$ in cycle notation, one proceeds as follows:

- Write an open bracket then select an arbitrary element $x \in \{1, \ldots, n\}$, and write down: $(x$
- Then trace the orbit of x: write down its value under successive applications of σ: $(x\ \sigma(x)\ \sigma^2(x) \cdots$
- Repeat until the value return to x, and write down a closing parenthesis rather than x: $(x\ \sigma(x)\ \sigma^2(x) \cdots)$
- Continue with any element y that is not yet written down, and proceed in the same way: $(x\ \sigma(x)\ \sigma^2(x) \cdots)(y\ \sigma(y) \cdots)$
- Repeat until all elements of $\{1, \ldots, n\}$ are written in one of the cycles.

For instance,

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 4 & 6 & 2 & 3 & 5 \end{bmatrix} = (1)(24)(365) = (24)(365).$$

Here $\sigma(1) = 1$ forms an 1-cycle, which is often omitted.

A 2-cycle is called a *transposition*. An important fact is that any element $\sigma \in S_n$ can be written as a product of transpositions. To see this, it suffices to show that any cycle can be written as a product of transpositions, as any σ is a product of cycles. This can be easily verified:

$$(i_1 i_2 \cdots i_k) = (i_1 i_k)(i_1 i_{k-1}) \cdots (i_1 i_2).$$

It is not hard to see that there is no unique way to represent a permutation by a product of transpositions. For instance, $(123) = (13)(12) = (12)(23) = (12)(23)(13)(13)$. However, the *parity* (i.e. even or odd) of the numbers of transpositions of such representations is unique. (For instance, (123) can not be written as the product of odd number of transpositions.) This permits the *parity of a permutation* to be a well-defined notion.

The key idea of the proof is to define a group homomorphism

$$\mathrm{sgn}\colon S_n \to \{+1, -1\} \text{ (under multiplication)}$$

so that all transpositions map to -1. Indeed, if we can find such a homomorphism, then for any representation $\sigma = \tau_1 \cdots \tau_k$ where τ_i's are transpositions, we have

$$\mathrm{sgn}(\sigma) = \mathrm{sgn}(\tau_1) \cdots \mathrm{sgn}(\tau_k) = (-1)^k.$$

This shows that the parity of k is independent of the choice of the decomposition.

Now, to define such group homomorphism sgn, we consider the Vandermonde polynomial

$$P(x_1, \ldots, x_n) = \prod_{1 \le i < j \le n} (x_i - x_j).$$

For $\sigma \in S_n$, define

$$\mathrm{sgn}(\sigma) := \frac{P(x_{\sigma(1)}, \ldots, x_{\sigma(n)})}{P(x_1, \ldots, x_n)}.$$

Observe that the polynomials $P(x_1, \ldots, x_n)$ and $P(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ have the same factors except for the signs, therefore $\mathrm{sgn}(\sigma) = \pm 1$. It defines a group homomorphism $\operatorname{sgn}\colon S_n \to \{\pm 1\}$ since

$$\begin{aligned}\operatorname{sgn}(\sigma_1\sigma_2) &= \frac{P(x_{\sigma_1(\sigma_2((1))}, \dots, x_{\sigma_1(\sigma_2((n))})}{P(x_1, \dots, x_n)} \\ &= \frac{P(x_{\sigma_1(\sigma_2((1))}, \dots, x_{\sigma_1(\sigma_2((n))})}{P(x_{\sigma_2(1)}, \dots, x_{\sigma_2(n)})} \cdot \frac{P(x_{\sigma_2(1)}, \dots, x_{\sigma_2(n)})}{P(x_1, \dots, x_n)} \\ &= \operatorname{sgn}(\sigma_1)\operatorname{sgn}(\sigma_2).\end{aligned}$$

Also, it is easy to check that sgn sends any transposition to -1. This finishes the proof.

Definition 2.22. The subset of S_n consisting of all *even* permutations will be denoted by A_n. It is a *normal subgroup* of S_n since it is the kernel of the group homomorphism sgn. The group A_n is called the *alternating group* (of n elements).

Exercise. Show that $A_n \subseteq S_n$ is a normal subgroup of index 2; it has two cosets, one of them consists of all even permutations, the other consists of all odd permutations.

Remark 2.23. The *center* $Z(G)$ of a group G is defined to be

$$Z(G) = \{g \in G \mid gh = hg \text{ for any } h \in G\} \subseteq G.$$

It is not hard to show that $Z(G)$ is a subgroup of G. The center measures the *commutativity* of the group: for instance, if G is abelian then $Z(G) = G$. In the homework, you'll show that the symmetric group S_n has trivial center $Z(S_n) = \{e\}$ if $n \geq 3$.

Lecture 3

2.6. **Group actions.**

We will be interested in groups G that act as symmetries of a set X (for instance, the symmetry group of a tetrahedron acting on the set of its vertices). Let us introduce the formal definition of group actions.

Definition 2.24. We say that a group G *acts on a set* X if there is a map

$$G \times X \to X; \quad (g, x) \mapsto g \cdot x$$

satisfying:

- $e_G \cdot x = x$ for any $x \in X$,
- $g \cdot (h \cdot x) = (gh) \cdot x$ for any $g, h \in G$ and $x \in X$.

The dot "$\cdot$" is sometimes omitted when the context is clear.

Exercise. Show that to give a group action of G on X is equivalent to give a group homomorphism $\rho\colon G \to S_X$. (Hint: Relate them by $g \cdot x = \rho(g)(x)$.)

Example. The symmetric group S_n acts on the set $\{1, \ldots, n\}$.

Example. $\mathrm{Isom}(\mathbb{R}^n)$ acts on $\mathbb{R}^n$.

Example. $\mathrm{O}(n, \mathbb{R})$ acts on the unit sphere $S^{n-1} \subseteq \mathbb{R}^n$, where

$$S^{n-1} = \{\vec{x} \in \mathbb{R}^n \mid ||\vec{x}|| = 1\}\,.$$

Example. The dihedral group D_n acts on the set of vertices of a regular n-gon, which gives a group homomorphism $D_n \to S_n$. Similarly, the symmetry group of a Platonic solid P acts on the set of its vertices.

Example. Let G be a group. The left multiplication action of G on itself is defined to be

$$G \times G \to G; \quad (g, h) \mapsto g \cdot h \coloneqq gh.$$

Equivalently, it's a group homomorphism

$$G \to S_G; \quad g \mapsto L_g,$$

where $L_g(h) \coloneqq gh$.

Exercise. Check that the right multiplication $g \cdot h \coloneqq hg$ is *not* an action of G on itself. Instead, $g \cdot h \coloneqq hg^{-1}$ is an action of G on itself.

Example. Let G be a group. An important action of G on itself is the *conjugacy action*:

$$G \times G \to G; \quad (g, h) \mapsto g \cdot h \coloneqq ghg^{-1}.$$

Equivalently, the conjugacy action is given by

$$G \to S_G; \quad g \mapsto \mathrm{Ad}_g,$$

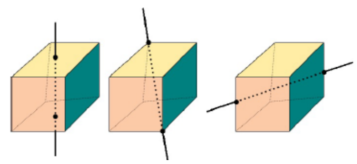
where $\mathrm{Ad}_g(h) \coloneqq ghg^{-1}$. Note that the permutation $\mathrm{Ad}_g\colon G \to G$ is in fact a group isomorphism.

Two elements h and h' of G are said to be *conjugate* if there exists $g \in G$ such that $h' = ghg^{-1}$. In this case, we say h and h' belong to the same conjugacy class. The study of conjugacy classes of non-abelian groups is fundamental for the study of their structure.

Exercise. Elements in the same conjugacy class behave similarly in many ways. For instance, prove that $\mathrm{order}(h) = \mathrm{order}(ghg^{-1})$.

Exercise. Let G be a group and $g \in G$. Show that g is in the center $Z(G)$ of G if and only if the conjugacy class of g consists of a single element $\{g\}$.

Example. Let C be a cube in $\mathbb{R}^3$ centered at the origin. Denote $\operatorname{Aut}^+(C)$ the *rotational symmetric group* of C. Each element of $\operatorname{Aut}^+(C)$ is a rotation that fixes a line through the origin, and sends the cube C to itself. For instance:



- identity map;
- rotate $\pi/2, \pi, 3\pi/2$ along the first (left-most) line: there are 3 such lines, so this gives in total 9 elements of $\operatorname{Aut}^+(C)$;
- rotate $2\pi/3, 4\pi/3$ along the second line: there are 4 such lines, so this gives in total 8 elements of $\operatorname{Aut}^+(C)$;
- rotate π along the third line: there are 6 such lines, so this gives in total 6 elements of $\operatorname{Aut}^+(C)$.

Hence $|\operatorname{Aut}^+(C)|$ is at least 24.

On the other hand, observe that $\operatorname{Aut}^+(C)$ gives an action on the set of the four main diagonals of C, therefore induces a group homomorphism

$$\rho\colon \operatorname{Aut}^+(C) \to S_4.$$

One can show that ρ is injective (this is not a trivial observation: one needs to show that the antipodal map $(x_1, x_2, x_3) \mapsto (-x_1, -x_2, -x_3)$ is *not* a rotation). Now, combining with the fact that $|\operatorname{Aut}^+(C)| \geq 24$, we can conclude that ρ is an isomorphism $\operatorname{Aut}^+(C) \cong S_4$.

Definition 2.25. Let X be a set admitting a group action by G. For any $x \in X$, define its *orbit* to be

$$\operatorname{orb}(x) \coloneqq \{g \cdot x \mid g \in G\} \subseteq X.$$

It sometimes is also denoted by Gx.

The subset of G fixing x

$$\operatorname{Stab}(x) \coloneqq \{g \in G \mid g \cdot x = x\}$$

is called the *stabilizer* of x, which is a subgroup of G (why?).

Exercise. Determine the orbits and stabilizers of the examples of group actions we mentioned above.

Exercise. Let X be a set admitting a group action by G. Let $\operatorname{orb}(x)$ and $\operatorname{orb}(y)$ be two orbits of the action. Prove that either $\operatorname{orb}(x) = \operatorname{orb}(y)$ or $\operatorname{orb}(x) \cap \operatorname{orb}(y) = \emptyset$.

In other words, a group G acting on a set X decomposes X into disjoint union of the orbits of the action. The set of all orbits is denoted by X/G.

Theorem 2.26. *Let X be a set admitting a group action by G. Let $g \in G$ and $x \in X$.*

(1) $\operatorname{Stab}(gx) = g\operatorname{Stab}(x)g^{-1}$. *In other words, the stabilizers of points on the same orbit are conjugate to each other.*

(2) (Orbit-stabilizer theorem) There is a bijective map between the orbit $\operatorname{orb}(x)$ *and the set of left cosets* $G/\operatorname{Stab}(x)$. *In particular, if $|G|$ is finite then* $|G| = |\operatorname{Stab}(x)||\operatorname{orb}(x)|$.

Proof. The first statement follows from

$$h \in \operatorname{Stab}(gx) \Leftrightarrow hgx = gx \Leftrightarrow g^{-1}hgx = x \Leftrightarrow g^{-1}hg \in \operatorname{Stab}(x).$$

To prove the second statement, consider the map

$$f\colon G \to \operatorname{orb}(x); \quad g \mapsto gx.$$

The map is clearly surjective. For any two elements $g_1, g_2 \in G$,

$$f(g_1) = f(g_2) \Leftrightarrow g_1x = g_2x \Leftrightarrow g_2^{-1}g_1x = x \Leftrightarrow g_2^{-1}g_1 \in \operatorname{Stab}(x) \Leftrightarrow g_1 \in g_2\operatorname{Stab}(x).$$

Hence $f(g_1) = f(g_2)$ if and only if g_1 and g_2 lie in the same coset for the stabilizer subgroup $\operatorname{Stab}(x) \subseteq G$. This proves the second statement. □

Example. Given a cube, we would like to put $\{1, 2, 3, 4, 5, 6\}$ on its faces to make it a dice. How many different dice can we build (up to rotational symmetry)?

In terms of group actions, let X be the set of all possible ways of labeling $\{1, 2, 3, 4, 5, 6\}$ on a dice, and G be the rotational symmetric group of the cube $\operatorname{Aut}^+(C)$ which acts naturally on X. The question is equivalent to counting the number of orbits of this group action. Observe that $\operatorname{Stab}(x) = \{e\}$ for any $x \in X$. By the orbit-stabilizer theorem, we have $|\operatorname{orb}(x)| \cdot 1 = |\operatorname{Aut}^+(C)| = 24$ for each $x \in X$, i.e. each orbit has exactly 24 elements. The set X has $6!$ elements, so the number of orbits is $6!/24 = 30$.

Theorem 2.27 (Burnside's lemma)**.** *Let X be a finite set admitting a group action by a finite group G. For any $g \in G$, denote $X^g = \{x \in X \mid gx = x\}$ the collection of points fixed by g. Then the number of disjoint orbits satisfies*

$$|X/G| = \frac{1}{|G|}\sum_{g\in G}|X^g|.$$

Proof. Consider the set of pairs

$$Z = \{(g,x) \in G \times X \mid gx = x\}.$$

On the one hand, for each $g_0 \in G$ there exists $|X^{g_0}|$ many elements in X such that $(g_0, x) \in Z$. Hence $|Z| = \sum_{g\in G}|X^g|$. On the other hand, for each $x_0 \in X$, there are $\mathrm{Stab}(x_0)$ many elements in G such that $(g, x_0) \in Z$. Hence

$$|Z| = \sum_{x\in X}|\mathrm{Stab}(x)| = \sum_{x\in X}\frac{|G|}{|\mathrm{orb}(x)|}.$$

Denote $O_1, \dots, O_k$ the orbits of X under the G-action, where $k = |X/G|$. Then

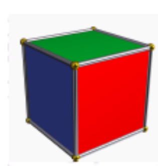
$$\sum_{x\in X}\frac{|G|}{|\mathrm{orb}(x)|} = |G|\sum_{i=1}^{k}\sum_{x\in O_i}\frac{1}{|\mathrm{orb}(x)|} = |G|\sum_{i=1}^{k}1 = |G||X/G|.$$

Therefore, we have

$$|G||X/G| = |Z| = \sum_{g\in G}|X^g|.$$

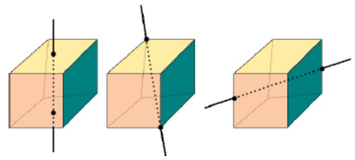
□

Example. How many different ways are there to color the faces of a cube with three colors (up to rotational symmetry)? Let X be the set of all possible



colorings of the cube, and let $G = \mathrm{Aut}^+(C)$. The problem is equivalent to calculating the number of orbits $|X/G|$. By Burnside's lemma, it suffices to compute the size of the fixed point sets for each element of G.

- identity map: fixes all colorings, there are 3^6 of them;



- rotate $\pi/2, 3\pi/2$ along lines of the first type (6 such rotations): each fixes 3^3 colorings;
- rotate π along lines of the first type (3 such rotations): each fixes 3^4 colorings;
- rotate $2\pi/3, 4\pi/3$ along lines of the second type (8 such rotations): each fixes 3^2 colorings;
- rotate π along lines of the third type (6 such rotations): each fixes 3^3 colorings.

By Burnside's lemma, we have

$$|X/G| = \frac{1}{24}\left(1\cdot 3^6 + 6\cdot 3^3 + 3\cdot 3^4 + 8\cdot 3^2 + 6\cdot 3^3\right) = 57.$$

3. Platonic solids and finite subgroups of $\mathrm{SO}(3,\mathbb{R})$

3.1. **Classification of the Platonic solids.**

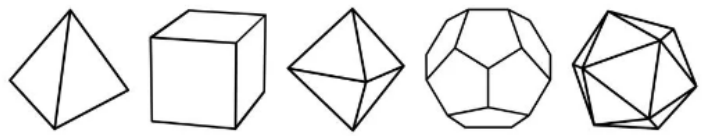
Definition 3.1. A *Platonic solid* is a convex polyhedron satisfying the following conditions:

(1) all its faces are convex regular polygons, and are congruent (identical in shape and size);
(2) none of its faces intersect except at their edges;
(3) the same number of faces meet at each of its vertices.

Each Platonic solid is completely determined by two numbers p and q, where

- p is the number of edges (or equivalently, vertices) of each face;
- q is the number of faces (or equivalently, edges) that meet at each vertex.

Fact 3.2. *There are only five Platonic solids.*



Polyhedron	*Vertices V*	*Edges E*	*Faces F*	(p,q)
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	$(3,5)$

Proof. We would like to show that there is no other possible (p,q) that can be used to form a Platonic solid.

A topological proof: It is not hard to see that $pF = 2E$ and $qV = 2E$. By the Euler's formula $V - E + F = 2$, one obtains

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{2} + \frac{1}{E} > \frac{1}{2}.$$

Also, note that p and q must both be at least 3. One can then check that there are only 5 possibilities for (p,q).

A geometric proof: Consider any vertex of a Platonic solid. The angle between two edges that meet at a vertex is $\pi - \frac{2\pi}{p}$. Now we use the fact that at each vertex of a convex polyhedron, the total (among the adjacent faces) of the angles between their respective adjacent sides is strictly less than 2π. Therefore, we have

$$q\left(\pi - \frac{2\pi}{p}\right) < 2\pi,$$

which is equivalent to $\frac{1}{p} + \frac{1}{q} > \frac{1}{2}$ that we obtained in the previous proof. $\square$

3.2. **Symmetry groups of the Platonic solids.** In order to study the symmetry group $\mathrm{Aut}(P)$ of a Platonic solid P, one can move the solid so that its center is located at the origin $\vec{0} = (0,0,0) \in \mathbb{R}^3$. Then, any isometry of $\mathbb{R}^3$ that fixes P must also fixes the origin.

Definition 3.3. The *orthogonal group* $\mathrm{O}(3,\mathbb{R})$ is defined as:

$$\mathrm{O}(3,\mathbb{R}) := \left\{ f \in \mathrm{Isom}(\mathbb{R}^3) \mid f(\vec{0}) = \vec{0} \right\},$$

which consists of isometries of $\mathbb{R}^3$ that fix the origin $\vec{0} = (0,0,0)$ of $\mathbb{R}^3$.

Exercise. Prove that $\mathrm{O}(3,\mathbb{R})$ is a subgroup of $\mathrm{Isom}(\mathbb{R}^3)$.

Definition 3.4. A *rotation* of $\mathbb{R}^3$ about the origin is a map $f \in \mathrm{O}(3,\mathbb{R})$ such that

- f fixes a line ℓ through the origin (called the *axis of rotation*), and
- f rotates the two-dimensional plane through the origin orthogonal to ℓ.

Exercise (Hard)*.* Prove that the composition of two rotations is still a rotation. Therefore, the set of all rotations about the origin forms a subgroup of $\mathrm{O}(3,\mathbb{R})$.

Notation. The group of all rotations about the origin of $\mathbb{R}^3$ will be denoted by $\mathrm{SO}(3,\mathbb{R})$. We will call it the *rotation group* for now. (Its official name is the *special orthogonal group* in dimension 3. Both $\mathrm{O}(3,\mathbb{R})$ and $\mathrm{SO}(3,\mathbb{R})$ are important examples of *matrix groups*. We will discuss their further properties later.)

Given a Platonic solid P centered at the origin $\vec{0} \in \mathbb{R}^3$, we would like to study:

- the symmetry group $\mathrm{Aut}(P)$, which is a subgroup of the orthogonal group $\mathrm{O}(3,\mathbb{R})$;
- the intersection $\mathrm{Aut}(P) \cap \mathrm{SO}(3,\mathbb{R})$, consisting of rotations that fix the solid P, will be called the *rotational symmetry group* of P, and will be denoted by $\mathrm{Aut}^+(P)$.

We will prove that any finite subgroup of $\mathrm{SO}(3,\mathbb{R})$ is either cyclic, dihedral, or $\mathrm{Aut}^+(P)$ for some Platonic solid P. Therefore, the Platonic solids not only classify the regular polyhedrons in $\mathbb{R}^3$, but also provide a classification of finite subgroups of the rotation groups in dimension 3.

The tetrahedron: Let us first study the symmetry group of the tetrahedron T (centered at the origin). Any symmetry of T permutes its vertices, hence we have a group homomorphism

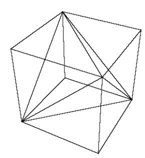
$$\rho\colon \operatorname{Aut}(T) \to S_4.$$

Note that ρ is injective: if a symmetry fixes all four vertices, then it is the identity. On the other hand, consider the action of $\operatorname{Aut}(T)$ on the set of vertices $\{1, 2, 3, 4\}$. For any vertex, it is clear that its orbit consists of four elements, and its stabilizer consists of six elements. Hence $|\operatorname{Aut}(T)| = 24$ by the orbit-stabilizer theorem, and ρ is therefore an isomorphism.

Lecture 4

Exercise. Prove that $|\operatorname{Aut}^+(T)| = 12$ and $\operatorname{Aut}^+(T) \cong A_4$. Can you identify the corresponding rotations?

The cube: There are two tetrahedra embedded in a cube C, with vertices at the vertices of the cube, denoted by T^+ and T^- (see the figure below). Here



are some useful observations:

- Any symmetry of C either maps each of the two tetrahedra onto itself, or interchanges the tetrahedra.
- The symmetry $J\colon \mathbb{R}^3 \to \mathbb{R}^3$ which sends $(x_1, x_2, x_3) \mapsto (-x_1, -x_2, -x_3)$ would interchange T^+ and T^-. It is an exercise to show that J commutes with any other symmetry of the cube (i.e. $J \circ f = f \circ J$ for any symmetry f); or equivalently, J is in the center of $\operatorname{Aut}(C)$.

One can then define a map (which can be checked is a group homomorphism)

$$\rho\colon \operatorname{Aut}(C) \to \operatorname{Aut}(T^+) \times {}^{\mathbb{Z}}/_{2\mathbb{Z}}$$

where:

- $\rho(f) = (f, \overline{0})$ if $f(T^+) = T^+$;
- $\rho(f) = (J \circ f, \overline{1})$ if $f(T^+) = T^-$.

It is again easy to show that ρ is injective, and moreover is isomorphic by the orbit-stabilizer theorem. Hence $\operatorname{Aut}(C) \cong S_4 \times (\mathbb{Z}/2\mathbb{Z})$.

Exercise. Verify that the map ρ is indeed a group homomorphism.

Exercise. Prove that $\mathrm{Aut}^+(C) \cong S_4$. Can you identify the corresponding rotations?

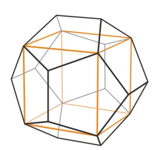
The dodecahedron: By the orbit-stabilizer theorem, it is not hard to show that the rotational symmetry group of a dodecahedron has order $|\mathrm{Aut}^+(D)| = 60$. Let us try to identify these 60 rotations that fix the dodecahedron D:

- the identity map;
- for each pair of opposite faces (there are 6 such pairs), there are 4 rotations (by multiples of $2\pi/5$) about their centers;
- for each pair of opposite edges (there are 15 such pairs), there is 1 rotation (by π) about their centers;
- for each pair of opposite vertices (there are 10 such pairs), there are 2 rotations (by multiples of $2\pi/3$) about the line connecting them.

These add up to $1 + 6 \times 4 + 15 \times 1 + 10 \times 2 = 60$, which therefore are all the elements of $\mathrm{Aut}^+(D)$.

Consider the alternating group A_5. It has 24 elements that are 5-cycles, 20 elements that are 3-cycles, and 15 elements that are of the form $(\star\star)(\star\star)$. This suggests that we might have $\mathrm{Aut}^+(D) \cong A_5$. Let us try to prove it.

Similar to the two tetrahedra embedded in a cube, there are five cubes embedded in a dodecahedron D in such a way that any symmetry of the



dodecahedron permutes these five embedded cubes, therefore gives a group homomorphism

$$\rho\colon\ \mathrm{Aut}(D) \to S_5.$$

However, the homomorphism ρ is *not* surjective: in fact, one can check that any symmetry of D gives an *even* permutations of the set of the five cubes. So we actually get a homomorphism

$$\rho\colon\ \mathrm{Aut}(D) \to A_5.$$

Moreover, one can check that the kernel of ρ, namely symmetries of D that fix each of the five cubes, consists of the identity and the map J that sends $(x_1, x_2, x_3) \mapsto (-x_1, -x_2, -x_3)$. Hence ρ induces an injective homomorphism

$\operatorname{Aut}(D) \big/ \{1, J\} \hookrightarrow A_5$. By the orbit-stabilizer theorem, one can show that $|\operatorname{Aut}(D)| = 120$. Therefore by counting the orders of these groups, we have

$$\operatorname{Aut}(D) \big/ \{1, J\} \cong A_5.$$

On the other hand, consider the composition τ of group homomorphisms

$$\tau\colon \operatorname{Aut}^+(D) \hookrightarrow \operatorname{Aut}(D) \to \operatorname{Aut}(D) \big/ \{1, J\}.$$

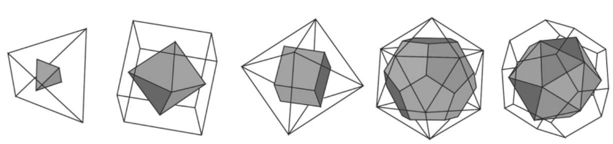
Since J is not a rotation, the composition τ is injective. Again by counting the orders of these groups, we obtain

$$\operatorname{Aut}^+(D) \cong \operatorname{Aut}(D) \big/ \{1, J\} \cong A_5.$$

Now, let us study the full symmetry group $\operatorname{Aut}(D)$. Since $|\operatorname{Aut}(D)| = 120$, we know that $\operatorname{Aut}(D)$ contains $\operatorname{Aut}^+(D) \cong A_5$ as an index two subgroup. It is tempting to believe that $\operatorname{Aut}(D)$ is isomorphic to S_5. However, this is *not* true! One way to see this is by observing that $J \neq 1$ is in the center $\operatorname{Aut}(D)$, but we also know that the center of S_5 is trivial, hence $\operatorname{Aut}(D) \ncong S_5$. In fact, it is not hard to show the following.

Exercise. Prove that $\operatorname{Aut}(D) \cong A_5 \times (\mathbb{Z}/2\mathbb{Z})$.

Remark 3.5. Every polyhedron has a *dual polyhedron* with faces and vertices interchanged. One can construct the dual polyhedron by taking the vertices of the dual to be the centers of the faces of the original figure. Connecting the centers of adjacent faces in the original forms the edges of the dual and thereby interchanges the number of faces and vertices while maintaining the number of edges. The dual of every Platonic solid is another Platonic solid, so we can arrange the five solids into dual pairs (where the tetrahedron is self-dual).



Exercise. The symmetry group of any polyhedron coincides with the symmetry group of its dual. (This is not hard to show by examining the construction of the dual polyhedron.) Therefore, there are only three symmetry groups associated with the Platonic solids rather than five.

We summarize the symmetry groups of the Platonic solids as follows.

Polyhedron	Aut	Aut^+
Tetrahedron	S_4	A_4
Cube/Octahedron	$S_4 \times (\mathbb{Z}/2\mathbb{Z})$	S_4
Dodecahedron/Icosahedron	$A_5 \times (\mathbb{Z}/2\mathbb{Z})$	A_5

3.3. **Finite subgroups of the rotation group** $\mathrm{SO}(3,\mathbb{R})$**.** Before proving the classification result regarding the rotation group in three dimension, let us begin with a two-dimensional result.

Theorem 3.6. *Let G be a finite subgroup of $\mathrm{O}(2,\mathbb{R})$. Then G is isomorphic to either a cyclic group or a dihedral group.*

Proof. Any element of $\mathrm{O}(2,\mathbb{R})$ acts naturally on the unit circle $S^1 \subseteq \mathbb{R}^2$. Let $g \in G$ be a non-identity element. It is not hard to show that g is either a rotation (when g does not fix any point of S^1), or a reflection (when g fixes at least a point of S^1).

First, suppose that all elements of G are rotations. Write $r_\theta \in \mathrm{O}(2,\mathbb{R})$ for the counterclockwise rotation by θ, where $0 \leq \theta < 2\pi$. Choose $r_\phi \in G$ with the smallest positive ϕ (it is possible since G is finite). We claim that G is the cyclic group generated by r_ϕ. Let $r_\theta \in G$, and write $\theta = m\phi + \psi$ where $m \in \mathbb{N}$ and $0 \leq \psi < \phi$. Then $r_\psi = (r_\phi)^{-m} r_\theta \in G$. Therefore $\psi = 0$ by the minimality of ϕ. Hence $r_\theta = (r_\phi)^m$.

Second, suppose G contains a reflection s. Let $H \subseteq G$ be the subgroup consisting of rotations (including the identity). By the first case, we have $H = \{1, r, \cdots, r^{n-1}\}$ for some positive integer n. Consider any other reflection $s' \in G$. One can show that the composition of any two reflections is a rotation, hence $ss' \in H$. So $s' = sr^k$ for some $0 \leq k \leq n-1$. This shows that

$$G = \{1, r, \ldots, r^{n-1}, s, sr, \ldots, sr^{n-1}\}.$$

It is easy to show that a rotation r and a reflection s satisfy the relation $sr = r^{-1}s$. Hence we get $G \cong D_n$. □

Exercise. Prove (geometrically) that $\mathrm{O}(2,\mathbb{R})$ is a subgroup of $\mathrm{SO}(3,\mathbb{R})$.

Theorem 3.7. *Let G be a finite subgroup of* $\mathrm{SO}(3,\mathbb{R})$*. Then G is isomorphic to precisely one of the following groups:*

- *cyclic group $\mathbb{Z}/n\mathbb{Z}$,*
- *dihedral group D_n,*
- *A_4: the rotational symmetry group of a tetrahedron,*
- *S_4: the rotational symmetry group of a cube (or a octahedron),*
- *A_5: the rotational symmetry group of a dodecahedron (or a icosahedron).*

Proof. Observe that $G \subseteq \mathrm{SO}(3,\mathbb{R})$ acts on the unit sphere $S^2 \subseteq \mathbb{R}^3$. Each rotation (other than $e \in \mathrm{SO}(3,\mathbb{R})$) gives two poles on the unit sphere which are the intersection of the axis of rotation with the unit sphere. Let $X \subseteq S^2$ denote the set of all poles of all the elements in $G\backslash\{e\}$. We claim that G acts on the set X. To see this, let $g \in G$ and $x \in X$. Say x is a pole for $h \in G\backslash\{e\}$ (i.e. $h(x) = x$). Then $(ghg^{-1})(gx) = ghx = gx$. Hence gx is a pole for $ghg^{-1} \neq e$, so $gx \in X$.

Now the idea of the proof is to apply Burnside's lemma to the action of G on X, and show that X has to be a particularly nice configuration of points on the sphere.

Let N be the number of orbits of the G-action on X. Choose a representative from each orbit, say $x_1, \ldots, x_N \in X$. Observe that the identity e fixes every pole, and each $g \neq e$ fixes exactly two poles. By Burnside's lemma, we have

$$\begin{aligned} N &= \frac{1}{|G|}\left(|X| + (|G|-1)\cdot 2\right) \\ &= \frac{1}{|G|}\left(2(|G|-1) + \sum_{i=1}^{N} |\mathrm{orb}(x_i)|\right) \end{aligned}$$

By orbit-stabilizer theorem, we have

$$\begin{aligned} 2\left(1 - \frac{1}{|G|}\right) &= N - \sum_{i=1}^{N} \frac{|\mathrm{orb}(x_i)|}{|G|} \\ &= N - \frac{1}{|\mathrm{Stab}(x_i)|} \\ &= \sum_{i=1}^{N}\left(1 - \frac{1}{|\mathrm{Stab}(x_i)|}\right) \end{aligned}$$

Since $|\mathrm{Stab}(x_i)| \geq 2$ for each i, it is then easy to deduce that $N \leq 3$. Clearly there is no solution with $N = 1$, so N is either 2 or 3.

Suppose $N = 2$. Then

$$2 - \frac{2}{|G|} = 2 - \frac{1}{|\mathrm{Stab}(x_1)|} - \frac{1}{|\mathrm{Stab}(x_2)|} \leq 2 - \frac{2}{|G|}.$$

Hence $\mathrm{Stab}(x_1) = \mathrm{Stab}(x_2) = G$ and $|\mathrm{orb}(x_1)| = |\mathrm{orb}(x_2)| = 1$. In other words, X consists of two unit vectors that are fixed by all elements of G. Suppose that one of the vectors is u, then the other must be $-u$. Therefore, any element of G has its axis of rotation given by u (or equivalently $-u$), and rotates the two-dimensional plane orthogonal to u. By what we discussed earlier about finite subgroups of $\mathrm{O}(2, \mathbb{R})$, G is a cyclic group.

Suppose $N = 3$. Let $|\mathrm{Stab}(x_1)| \geq |\mathrm{Stab}(x_2)| \geq |\mathrm{Stab}(x_3)| \geq 2$. Then

$$\frac{1}{|\mathrm{Stab}(x_1)|} + \frac{1}{|\mathrm{Stab}(x_2)|} + \frac{1}{|\mathrm{Stab}(x_3)|} = 1 + \frac{2}{|G|}.$$

This implies that $3/|\mathrm{Stab}(x_3)| > 1$, hence $|\mathrm{Stab}(x_3)| = 2$. Therefore

$$\frac{1}{|\mathrm{Stab}(x_1)|} + \frac{1}{|\mathrm{Stab}(x_2)|} = \frac{1}{2} + \frac{2}{|G|}.$$

This implies that $2/|\mathrm{Stab}(x_2)| > 1/2$, hence $|\mathrm{Stab}(x_2)|$ is either 2 or 3. There are four possible cases:

(a) $|\mathrm{Stab}(x_2)| = 2$.
(b) $|\mathrm{Stab}(x_2)| = 3$ and $|\mathrm{Stab}(x_1)| = 3$. ($|G| = 12$)
(c) $|\mathrm{Stab}(x_2)| = 3$ and $|\mathrm{Stab}(x_1)| = 4$. ($|G| = 24$)
(d) $|\mathrm{Stab}(x_2)| = 3$ and $|\mathrm{Stab}(x_1)| = 5$. ($|G| = 60$)

Case (a): Suppose $|\mathrm{Stab}(x_2)| = 2$. If $|\mathrm{Stab}(x_1)| = 2$, then we get $|G| = 4$. It is an easy exercise to show that any group of four elements is isomorphic to either $\mathbb{Z}/4\mathbb{Z}$ or $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \cong D_2 = \langle r, s \mid r^2 = s^2 = (rs)^2 = 1\rangle$.

If $|\mathrm{Stab}(x_1)| = M \geq 3$, then $|G| = 2M$ and the orbit of x_1 consists of two elements. The stabilizer $\mathrm{Stab}(x_1)$ is a finite group of M rotations about the line ℓ_{x_1} through x_1, so it is a cyclic group of M elements generated by a rotation $R \in G$. Denote the plane orthogonal to ℓ_{x_1} and passes through the origin by P_{x_1}.

First, one observes that $\mathrm{orb}(x_1) = \{x_1, -x_1\}$, since they are the only two elements of X that have M elements in their stabilizers. Now, we consider the action of $\mathrm{Stab}(x_1) = \langle R \rangle$ on x_2. The points $\{x_2, Rx_2, \ldots, R^{M-1}x_2\}$ would form a regular M-gon in a plane orthogonal to ℓ_{x_1}. We claim they actually lie in P_{x_1}.

Lecture 5

Consider any $R^i x_2$. Let $g \neq e$ be a stabilizer of $R^i x_2$. Since g is not a stabilizer of x_1, it must exchange x_1 and $-x_1$. Therefore, $R^i x_2$ must lie in P_{x_1}, and g is a rotation (by π) with axis on P_{x_1}. This proves the claim. Also, this argument shows that g acts on the regular M-gon $\{x_2, Rx_2, \ldots, R^{M-1}x_2\}$ as a reflection, with the mirror line passes through $R^i x_2$.

Note that $x_2, Rx_2, \ldots, R^{M-1}x_2$ are M distinct points of $\mathrm{orb}(x_2)$; and since $|\mathrm{orb}(x_2)| = M$, we have $\mathrm{orb}(x_2) = \{x_2, Rx_2, \ldots, R^{M-1}x_2\}$. The group G acts naturally on $\mathrm{orb}(x_2)$, which is the set of vertices of a regular M-gon, therefore induces a group homomorphism

$$\rho \colon G \to D_M.$$

Now let us consider the image subgroup $\mathrm{Im}(\rho) \subseteq D_M$. It contains all M rotations induced by $\langle R \rangle$; it also contains the reflections along lines passing through $R^i x_2$ induced by stabilizer of $R^i x_2$. Therefore $\mathrm{Im}(\rho) = D_M$, i.e. the homomorphism ρ is surjective. Since $|G| = |D_M| = 2M$, we have $G \cong D_M$.

Case (b): Suppose $|\mathrm{Stab}(x_2)| = 3$ and $|\mathrm{Stab}(x_1)| = 3$. Then $|G| = 12$ and $|\mathrm{orb}(x_1)| = 4$. Choose $v \in \mathrm{orb}(x_1)$ such that $v \neq \pm x_1$. Consider the stabilizer $\mathrm{Stab}(x_1) = \{1, R, R^2\}$ acting on v. We get three distinct elements $\{v, Rv, R^2v\} \subseteq \mathrm{orb}(x_1)$. They are all different from x_1, and they form an equilateral triangle. The same argument works if one replaces x_1 by another element in the same orbit. Therefore $\mathrm{orb}(x_1)$ forms a tetrahedron. Since G acts naturally on $\mathrm{orb}(x_1)$, we get a group homomorphism

$$\rho \colon G \to \mathrm{Aut}^+(T).$$

Since no rotation (other than e) fixes T, the homomorphism ρ is injective. Now as $|G| = |\mathrm{Aut}^+(T)| = 12$, we have that ρ is an isomorphism and $G \cong \mathrm{Aut}^+(T) \cong A_4$.

Note that one can also consider the orbit of x_2, which turns out to be the vertices of the *dual* tetrahedron of $\operatorname{orb}(x_1)$, consisting of $-x_1, -v, -Rv, -R^2v$.

Case (c): Suppose $|\operatorname{Stab}(x_2)| = 3$ and $|\operatorname{Stab}(x_1)| = 4$. Then $|G| = 24$ and $|\operatorname{orb}(x_1)| = 6$. Choose $v \in \operatorname{orb}(x_1)$ such that $v \neq \pm x_1$. Consider the stabilizer $\operatorname{Stab}(x_1) = \{1, R, R^2, R^3\}$ acting on v. By the same argument as above, $\{v, Rv, R^2v, R^3v\} \subseteq \operatorname{orb}(x_1)$ form a square equidistant from x_1. As $-x_1 \in X$ and $-x_1 \notin \operatorname{orb}(x_2) \cup \operatorname{orb}(x_3)$ (since the sizes of their stabilizers are different), we have

$$\operatorname{orb}(x_1) = \{x_1, -x_1, v, Rv, R^2v, R^3v\}.$$

Now, consider $-v \in X$. We have $-v \in \operatorname{orb}(x_1)$ (since $-v \notin \operatorname{orb}(x_2) \cup \operatorname{orb}(x_3)$) and $-v \neq \pm x_1$. Since $\{v, Rv, R^2v, R^3v\}$ forms a square, one can conclude that $-v = R^2v$. Similarly, we have $R^3v = -Rv$. This shows that $\operatorname{orb}(x_1)$ forms the vertices of a regular octahedron. We get a group homomorphism

$$\rho\colon G \to \operatorname{Aut}^+(O).$$

Since no rotation (other than e) fixes O, the homomorphism ρ is injective. Now as $|G| = |\operatorname{Aut}^+(O)| = |\operatorname{Aut}^+(C)| = 24$, we have that ρ is an isomorphism and $G \cong \operatorname{Aut}^+(O) \cong \operatorname{Aut}^+(C) \cong S_4$.

Case (d): Suppose $|\operatorname{Stab}(x_2)| = 3$ and $|\operatorname{Stab}(x_1)| = 5$. Then $|G| = 60$ and $|\operatorname{orb}(x_1)| = 12$. Consider the stabilizer $\operatorname{Stab}(x_1) = \{1, R, R^2, R^3, R^4\}$. It can be shown that we can pick $u \in \operatorname{orb}(x_1)\backslash\{\pm x_1\}$, and $v \in \operatorname{orb}(x_1)$ with $v \neq \pm x_1, u, Ru, R^2u, R^3u, R^4u$, and check that

$$\{x_1, -x_1, u, Ru, R^2u, R^3u, R^4u, v, Rv, R^2v, R^3v, R^4v\}$$

forms a regular icosahedron. Using the same argument as before, we obtain an isomorphism $G \cong \operatorname{Aut}^+(I) \cong \operatorname{Aut}^+(D) \cong A_5$.

$\square$

4. A crash course on basic linear algebra

4.1. **Matrix products, invertibility, determinants.** Elements of the vector space $\mathbb{R}^n$ are of the form

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

where $x_1, \ldots, x_n \in \mathbb{R}$. To save space, we sometimes write down the *transpose* of $\vec{x}$ instead: $\vec{x} = [x_1 \cdots x_n]^T$. There are two important operations on the vector space $\mathbb{R}^n$:

- *Addition*: Let $\vec{x} = [x_1 \cdots x_n]^T$ and $\vec{y} = [y_1 \cdots y_n]^T$ be two vectors in $\mathbb{R}^n$. Define $\vec{x} + \vec{y} \coloneqq [x_1 + y_1 \cdots x_n + y_n] \in \mathbb{R}^n$.
- *Scalar multiplication*: Let $\vec{x} = [x_1 \cdots x_n]^T$ and $\lambda \in \mathbb{R}$. Define $\lambda\vec{x} \coloneqq [\lambda x_1 \cdots \lambda x_n] \in \mathbb{R}^n$.

Definition 4.1. Let $\vec{v}_1, \ldots, \vec{v}_k$ be vectors in $\mathbb{R}^n$. Then, for any $c_1, \ldots, c_k \in \mathbb{R}$, the vector

$$c_1\vec{v}_1 + \cdots + c_k\vec{v}_k \in \mathbb{R}^n$$

is called a *linear combination* of the set of vectors $\vec{v}_1, \ldots, \vec{v}_k$ (with weights $c_1, \ldots, c_k$). The *span* of the set of vectors $\vec{v}_1, \ldots, \vec{v}_k$ is defined to be the collection of all of their linear combinations:

$$\begin{aligned} \operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_k\} &= \{\text{linear combinations of } \vec{v}_1, \ldots, \vec{v}_k\} \\ &= \{c_1\vec{v}_1 + \cdots + c_k\vec{v}_k \mid c_1, \ldots, c_k \in \mathbb{R}\}. \end{aligned}$$

Remark 4.2. The most fundamental question in linear algebra is to determine whether a linear system of equations has a solution:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

Consider the vectors $\vec{v_i} = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{bmatrix}$ $(1 \leq i \leq n)$ and $\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$. Then the system has a solution is equivalent to the statement that

$$\vec{b} \in \operatorname{Span}\{\vec{v_1}, \dots, \vec{v_n}\}.$$

Definition 4.3. Let $A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix}$ be an $m \times n$ matrix with column vectors given by $\vec{a}_1, \dots, \vec{a}_n \in \mathbb{R}^m$. Let $\vec{x} = [x_1 \cdots x_n]^T \in \mathbb{R}^n$. Define the *matrix-vector product* of A and $\vec{x}$ to be the linear combination:

$$A\vec{x} := x_1\vec{a}_1 + \cdots + x_n\vec{a}_n \in \mathbb{R}^m.$$

Remark 4.4. For any $m \times n$ matrix A, the matrix-vector product gives rise to a function

$$T_A \colon \mathbb{R}^n \to \mathbb{R}^m; \quad T_A(\vec{x}) := A\vec{x}.$$

The core of linear algebra is to study such a function. It is easy to check that the function T_A is *linear*, i.e. it is compatible with the additions and scalar multiplications on $\mathbb{R}^n$ and $\mathbb{R}^m$:

- $T_A(\vec{v} + \vec{w}) = T_A(\vec{v}) + T_A(\vec{w})$,
- $T_A(\lambda\vec{v}) = \lambda T_A(\vec{v})$.

Exercise. Let $T \colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation (i.e. compatible with the additions and scalar multiplications on $\mathbb{R}^n$ and $\mathbb{R}^m$). Then there exists a unique $m \times n$ matrix A such that $T_A = T$. In fact, the i-th column of A is given by $T(\vec{e_i})$, where $\vec{e_i} = [0 \cdots 010 \cdots 0]^T$ with the only nonzero entry at the i-th coordinate.

Therefore, there is a one-to-one correspondence between $m \times n$ matrices and linear transformations $\mathbb{R}^n \to \mathbb{R}^m$.

Exercise. Show that under the correspondence described above, the $n \times n$ matrix corresponds to the identity transformation $\mathrm{id} \colon \mathbb{R}^n \to \mathbb{R}^n$ ($\mathrm{id}(\vec{x}) = \vec{x}$ for all $\vec{x}$) is

$$I_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \\ 0 & & \cdots & & 1 \end{bmatrix}$$

and is called the *identity matrix*.

Definition 4.5. Let A be an $m \times n$ matrix and B be an $n \times p$ matrix. We would like to define the *matrix product* AB, which will be an $m \times p$ matrix.

The matrices A and B correspond to linear transformations $T_A\colon \mathbb{R}^n \to \mathbb{R}^m$ and $T_B\colon \mathbb{R}^p \to \mathbb{R}^n$. Consider the composition

$$T_A \circ T_B\colon \mathbb{R}^p \to \mathbb{R}^m; \quad \vec{x} \mapsto T_A(T_B(\vec{x})).$$

One can check the composition of linear maps is still linear, hence there exists a unique $m \times p$ matrix, of which we define to be the matrix product AB, such that $T_{AB} = T_A \circ T_B$.

Exercise. Write down the entries of AB explicitly in terms of the entries of A and B.

Remark 4.6. The definition of matrix product we give here is more conceptual. It has many advantages: for instance, the *associativity* of matrix product $A(BC) = (AB)C$ follows immediately from the associativity of compositions of functions.

Exercise. Let A be an $m \times n$ matrix. Then $A = AI_n = I_m A$.

Notation. Let A and B be $m \times n$ matrices. Let $\lambda \in \mathbb{R}$.

- (Addition) $A + B$ is an $m \times n$ matrix given by entry-wise addition.
- (Scalar multiplication) λA is an $m \times n$ matrix given by entry-wise scalar multiplication by λ.
- (Transpose) A^T is an $n \times m$ matrix given by $(A^T)_{ij} = A_{ji}$.
- If A is a square matrix (i.e. $m = n$), then $A \cdot A$ makes sense and we denote $A^2 = A \cdot A$. Similarly, $A^3 = A \cdot A \cdot A$, and so on.

Definition 4.7. Let A be an $n \times n$ matrix. We say A is *invertible* (or *non-singular*) if there exists $n \times n$ matrices B and C such that

$$AB = I_n = CA.$$

In fact, such B and C must coincide since $B = I_n B = (CA)B = C(AB) = CI_n = C$. Moreover, one can easily show that such B is unique if it exists. When A is invertible, the matrix B such that $AB = I_n = BA$ is called the *inverse* of A, and is denoted by A^{-1}.

Exercise. Prove the following statements.

- If A is invertible, then so is A^{-1}, and $(A^{-1})^{-1} = A$.
- If A, B are invertible matrices of the same size, then AB also is invertible, and $(AB)^{-1} = B^{-1}A^{-1}$.
- If A is invertible, then so is A^T, and $(A^T)^{-1} = (A^{-1})^T$.

A consequence of the first two statements is that, the set of all *invertible* $n \times n$ matrices form a *group*, with operation given by matrix multiplication and identity element given by I_n. The group is called the *general linear group* and denoted by $\mathrm{GL}(n, \mathbb{R})$.

Remark 4.8. Here is a basic fact on characterizing invertible matrices. Let A be an $n \times n$ matrix. The following statements are equivalent:

- A is invertible.
- T_A is bijective.
- T_A is injective.
- T_A is surjective.
- There exists an $n \times n$ matrix B such that $AB = I_n$.
- There exists an $n \times n$ matrix C such that $CA = I_n$.

Note that these equivalences do *not* hold in general: they only hold for square matrices.

Definition 4.9. Let A be an $n \times n$ matrix. Its *determinant* is defined to be

$$\det(A) \coloneqq \sum_{\sigma \in S_n} (-1)^{\mathrm{sgn}(\sigma)} a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} \in \mathbb{R}.$$

For instance, when $n = 2$, we have $\det(A) = a_{11}a_{22} - a_{12}a_{21}$.

Theorem 4.10. *Here are some important results of the determinants.*

- *A square matrix A is invertible if and only if* $\det(A) \neq 0$.
- *For any two square matrices A and B of the same size, we have* $\det(AB) = \det(A)\det(B)$.
- $\det(A) = \det(A^T)$.

- *Geometrically,* $|\det(A)|$ *coincides with the volume of the (n-dimensional) parallelogram spanned by the column (or row) vectors of A.*

Denote $\mathbb{R}^* = \mathbb{R}\backslash\{0\}$ with the group structure given by multiplications. Then the determinants give a group homomorphism

$$\det\colon \mathrm{GL}(n, \mathbb{R}) \to \mathbb{R}^*$$

Its kernel (i.e. matrices with determinant one) is called the *special linear group* and denoted by $\mathrm{SL}(n, \mathbb{R})$.

4.2. **Change of basis, eigenvectors and eigenvalues.**

Example. Consider the matrix $A = \begin{bmatrix} 11 & -2 \\ -2 & 14 \end{bmatrix}$. As before, we would like to understand the linear map $T_A\colon \mathbb{R}^2 \to \mathbb{R}^2$. We have

$$T_A\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \left(\begin{bmatrix} 11 \\ -2 \end{bmatrix}\right) \quad \text{and} \quad T_A\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \left(\begin{bmatrix} -2 \\ 14 \end{bmatrix}\right).$$

If we consider a different *basis* of $\mathbb{R}^2$, say $\mathcal{B} = \left\{\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \end{bmatrix}\right\}$, then we have

$$T_A\left(\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right) = 10\begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad \text{and} \quad T_A\left(\begin{bmatrix} 1 \\ -2 \end{bmatrix}\right) = 15\begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

By considering this new basis $\mathcal{B}$, we can understand the geometry of the linear map T_A easily: it expands in the $[2, 1]^T$-direction by the factor of 10, and expands in the $[1, -2]^T$-direction by the factor of 15. Moreover, since $\mathcal{B}$ is a basis of $\mathbb{R}^2$, any vector $\vec{x} \in \mathbb{R}^2$ can be uniquely written as $\vec{x} = c_1\begin{bmatrix} 2 \\ 1 \end{bmatrix} + c_2\begin{bmatrix} 1 \\ -2 \end{bmatrix}$ for some $c_1, c_2 \in \mathbb{R}$. Since T_A is linear, one can easily deduce that $T_A(\vec{x}) = 10c_1\begin{bmatrix} 2 \\ 1 \end{bmatrix} + 15c_2\begin{bmatrix} 1 \\ -2 \end{bmatrix}$. In other words, the map T_A is easier to understand if we write vectors in $\mathbb{R}^2$ in the $\mathcal{B}$-coordinates.

A more conceptual way to understand this change of basis process is that, starting with the matrix $A = \begin{bmatrix} 11 & -2 \\ -2 & 14 \end{bmatrix}$, we consider an invertible matrix $P_{\mathcal{B}} = \begin{bmatrix} 2 & 1 \\ 1 & -2 \end{bmatrix}$ (where the columns are the vectors from $\mathcal{B}$). Then in terms of the $\mathcal{B}$-coordinates, T_A becomes

$$\mathbb{R}^2 \xrightarrow{T_{P_{\mathcal{B}}}} \mathbb{R}^2 \xrightarrow{T_A} \mathbb{R}^2 \xrightarrow{T_{P_{\mathcal{B}}^{-1}}} \mathbb{R}^2.$$

This linear map sends

$$\begin{bmatrix}1\\0\end{bmatrix} \mapsto \begin{bmatrix}2\\1\end{bmatrix} \mapsto 10\begin{bmatrix}2\\1\end{bmatrix} \mapsto 10\begin{bmatrix}1\\0\end{bmatrix} \quad \text{and} \quad \begin{bmatrix}0\\1\end{bmatrix} \mapsto \begin{bmatrix}1\\-2\end{bmatrix} \mapsto 15\begin{bmatrix}1\\-2\end{bmatrix} \mapsto 15\begin{bmatrix}0\\1\end{bmatrix}.$$

So the matrix corresponds to the composition is simply the diagonal matrix $D = \begin{bmatrix}10 & 0\\ 0 & 15\end{bmatrix} = P_{\mathcal{B}}^{-1} A P_{\mathcal{B}}$.

Note that if A is also invertible, then the change of basis process $A \mapsto P_{\mathcal{B}}^{-1} A P_{\mathcal{B}}$ is the *conjugation* in the group $\mathrm{GL}(n, \mathbb{R})$ we discussed before. This process is called a *diagonalization* of the matrix A.

Definition 4.11. Let A be a square matrix. A nonzero vector $\vec{v} \neq \vec{0}$ is called an *eigenvector* of A if $A\vec{v} = \lambda\vec{v}$ for some λ, and the scalar λ is called an *eigenvalue* of A.

Observe that λ is an eigenvalue of A if and only if $A - \lambda I_n$ is not invertible, which is equivalent to

$$\det(A - \lambda I_n) = 0.$$

This is a degree n polynomial in λ, which is called the *characteristic polynomial* of A.

Remark 4.12. We allow eigenvalues and eigenvectors to take values in *complex* numbers, since the characteristic polynomial has complex roots in general.

4.3. **Inner products, orthogonal matrices.** In Homework 1, we showed that any element $T \in \mathrm{O}(n, \mathbb{R})$ of origin-preserving isometries of $\mathbb{R}^n$ is *linear*, and that T preserves the standard inner product on $\mathbb{R}^n$. There exists a unique $n \times n$ matrix A such that $T_A = T$. Since T preserves the inner product, for any $\vec{x}, \vec{y} \in \mathbb{R}^n$ we have

$$\langle \vec{x}, \vec{y} \rangle = \langle A\vec{x}, A\vec{y} \rangle, \text{ or equivalently } \vec{x}^T \vec{y} = \vec{x}^T A^T A \vec{y}.$$

It is an easy exercise to show that this would imply that $A^T A = I_n$. In fact, the converse is true, namely, if we have a matrix A satisfying $A^T A = I_n$, then T_A is an origin-preserving isometry of $\mathbb{R}^n$. Therefore, we can identify the origin-preserving isometries $\mathrm{O}(n, \mathbb{R})$ with the matrices satisfying $A^T A = I_n$.

Definition 4.13. An $n \times n$ matrix A is called *orthogonal* if $A^T A = I_n$. We will also denote the group of orthogonal matrices by $\mathrm{O}(n, \mathbb{R})$.

Remark 4.14. Let $\{\vec{v}_1, \dots, \vec{v}_n\}$ be the columns of A. Then A is orthogonal if and only if $\{\vec{v}_1, \dots, \vec{v}_n\}$ is an *orthonormal* set, i.e. $\langle \vec{v}_i, \vec{v}_i \rangle = 1$ for all i and $\langle \vec{v}_i, \vec{v}_j \rangle = 0$ for all $i \neq j$.

Observe that if A is orthogonal, then $\det(A)^2 = \det(A^T A) = 1$, hence $\det(A) = \pm 1$. The subgroup of $\mathrm{O}(n, \mathbb{R})$ with determinant one is called the *special orthogonal group*, and denoted by

$$\mathrm{SO}(n, \mathbb{R}) = \{A \in \mathrm{O}(n, \mathbb{R}) \mid \det(A) = 1\}.$$

There is a surjective group homomorphism $\det \colon \mathrm{O}(n, \mathbb{R}) \to \{\pm 1\}$ with kernel given by $\mathrm{SO}(n, \mathbb{R})$, hence we have $[\mathrm{O}(n, \mathbb{R}) : \mathrm{SO}(n, \mathbb{R})] = 2$.

In the previous section, we used $\mathrm{SO}(3, \mathbb{R})$ to denote the *rotation group* in three dimension. Now, we would like to show that these two notations actually coincide, i.e. $A \in \mathrm{SO}(3, \mathbb{R})$ is special orthogonal if and only if T_A is a rotation in $\mathbb{R}^3$.

First, we show that if T_A is a rotation then $A \in \mathrm{SO}(3, \mathbb{R})$. The rotation T_A would fix two unit vectors in $\mathbb{R}^3$, say $\pm \vec{v}_3$. One can find two more unit vectors $\vec{v}_1, \vec{v}_2$ so that $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ forms an orthonormal set. After possibly switching $\vec{v}_1$ and $\vec{v}_2$, we can assume that the matrix $B = [\vec{v}_1 \vec{v}_2 \vec{v}_3]$ with columns given by these three vectors is orthogonal and $\det(B) = 1$, i.e. $B \in \mathrm{SO}(3, \mathbb{R})$. We know that T_A fixes $\vec{v}_3$, and acts on the plane spanned by $\{\vec{v}_1, \vec{v}_2\}$ as a rotation, say by angle θ. So we have

$$\begin{aligned} T_A(\vec{v}_1) &= \cos\theta \vec{v}_1 + \sin\theta \vec{v}_2 \\ T_A(\vec{v}_2) &= -\sin\theta \vec{v}_1 + + \cos\theta \vec{v}_2 \\ T_A(\vec{v}_3) &= \vec{v}_3 \end{aligned}$$

therefore

$$B^{-1} A B = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \in \mathrm{SO}(3, \mathbb{R}).$$

Thus we conclude that $A \in \mathrm{SO}(3, \mathbb{R})$.

Conversely, we would like to show that for any $A \in \mathrm{SO}(3, \mathbb{R})$, the linear map T_A is a rotation. First, we need to show that T_A fixes a unit vector, or equivalently, 1 is an eigenvalue of A. First, one can show that any 3×3 matrix has a real eigenvalue (the characteristic polynomial is of odd degree, one can use the intermediate value theorem to conclude that it must has a real root). Suppose $A\vec{v}=\lambda\vec{v}$ where $\lambda\in\mathbb{R}$ and $\vec{v}\neq\vec{0}$. Since A is orthogonal, we have

$$\lambda^2||\vec{v}||^2=\langle\lambda\vec{v},\lambda\vec{v}\rangle=\langle A\vec{v},A\vec{v}\rangle=\langle\vec{v},\vec{v}\rangle=||\vec{v}||^2.$$

Hence $\lambda=\pm 1$. If $\lambda=1$, then we have shown that 1 is an eigenvalue of A. Now suppose $\lambda=-1$, and let $\vec{v}_3$ be a unit eigenvector of λ. By the same argument as before, we can find $\vec{v}_1,\vec{v}_2$ such that the matrix $B=[\vec{v}_1,\vec{v}_2,\vec{v}_3]\in\mathrm{SO}(3,\mathbb{R})$. Since A is orthgonal, the vectors $A\vec{v}_1,A\vec{v}_2$ are orthogonal to $A\vec{v}_3=-\vec{v}_3$, hence they both lie in $\mathrm{Span}\{\vec{v}_1,\vec{v}_2\}$. So we have

$$B^{-1}AB=\begin{bmatrix}a & b & 0\\ c & d & 0\\ 0 & 0 & -1\end{bmatrix}$$

for some $a,b,c,d\in\mathbb{R}$. Since $A,B\in\mathrm{SO}(3,\mathbb{R})$, the upper 2×2 block matrix lies in $\mathrm{O}(2,\mathbb{R})$ with determinant -1. This implies that it is a reflection. Thus, there exists a nonzero vector $\vec{v}\in\mathrm{Span}\{\vec{v}_1,\vec{v}_2\}$ such that $B^{-1}AB\vec{v}=\vec{v}$, and 1 is an eigenvalue of $B^{-1}AB$. Note that A and $B^{-1}AB$ have identical characteristic polynomials, hence 1 is also an eigenvalue of A.

Now we have shown that any $A\in\mathrm{SO}(3,\mathbb{R})$ has an eigenvalue 1. Choose a unit vector $\vec{v}_3$ such that $A\vec{v}_3=\vec{v}_3$. By the same argument again, there exists $B\in\mathrm{SO}(3,\mathbb{R})$ such that

$$B^{-1}AB=\begin{bmatrix}a & b & 0\\ c & d & 0\\ 0 & 0 & 1\end{bmatrix}$$

for some $a,b,c,d\in\mathbb{R}$. The upper 2×2 block lies in $\mathrm{SO}(2,\mathbb{R})$, which means that it is a rotation. Therefore T_A fixes $\vec{v}_3$ and is rotation on the plane $\vec{v}_3^{\perp}=\mathrm{Span}\{\vec{v}_1,\vec{v}_2\}$. Hence T_A is a rotation in $\mathbb{R}^3$.

5. Classification of plane crystallographic groups

Lecture 6

So far, we considered the symmetry groups of shapes in $\mathbb{R}^2$ and $\mathbb{R}^3$ that are *bounded*, which do not contain any *translation*. In this section, we discuss the symmetry groups of certain *unbounded* shapes/patterns in $\mathbb{R}^2$ (*frieze* patterns and *wallpaper* patterns), which do contain certain translations in their symmetry groups. First, let us take a closer look at the group of isometries $\mathrm{Isom}(\mathbb{R}^n)$ and its subgroups $T(n,\mathbb{R})$ (the group of translations in $\mathbb{R}^n$) and $\mathrm{O}(n,\mathbb{R})$ (the group of orthogonal linear transformations of $\mathbb{R}^n$, or equivalently, the group of origin-preserving isometries of $\mathbb{R}^n$).

5.1. **Translation subgroups and point groups.** From Homework 1 and the previous section, we know that for any $f \in \mathrm{Isom}(\mathbb{R}^n)$, there exists a unique pair of an orthogonal matrix A and a vector $\vec{v}$ such that

$$f(\vec{x}) = A\vec{x} + \vec{v} \text{ for any } \vec{x} \in \mathbb{R}^n.$$

This gives a function

$$\pi\colon \mathrm{Isom}(\mathbb{R}^n) \to \mathrm{O}(n,\mathbb{R}), \quad f \mapsto A.$$

The function π is in fact a group homomorphism: suppose $f_1(\vec{x}) = A_1\vec{x} + \vec{v}_1$ and $f_2(\vec{x}) = A_2\vec{x} + \vec{v}_2$, then

$$f_1(f_2(\vec{x})) = A_1(A_2\vec{x} + \vec{v}_2) + \vec{v}_1 = (A_1A_2)\vec{x} + (A_1\vec{v}_2 + \vec{v}_1).$$

Hence $\pi(f_1f_2) = A_1A_2$. The kernel of π is an isometry of the form $f(\vec{x}) = \vec{x}+\vec{v}$, which is simply the translation by $\vec{v}$. Hence $\mathrm{Ker}(\pi) = T(n,\mathbb{R}) \cong \mathbb{R}^n$. This gives an alternative proof of a problem in Homework 2 that the group of translations $T(n,\mathbb{R})$ is normal in $\mathrm{Isom}(\mathbb{R}^n)$. The homomorphism π is clearly surjective, so we have an isomorphism

$$\mathrm{Isom}(\mathbb{R}^n)/T(n,\mathbb{R}) \cong \mathrm{O}(n,\mathbb{R}).$$

Remark 5.1. Although $\mathrm{Isom}(\mathbb{R}^n) \ncong T(n,\mathbb{R}) \times \mathrm{O}(n,\mathbb{R})$, the two subgroups $T(n,\mathbb{R})$ and $\mathrm{O}(n,\mathbb{R})$ of $\mathrm{Isom}(\mathbb{R}^n)$ do possess some nice properties:

- Any $f \in \mathrm{Isom}(\mathbb{R}^n)$ can be uniquely written as $f = hk$ for some $h \in T(n,\mathbb{R})$ and $k \in \mathrm{O}(n,\mathbb{R})$.
- $T(n,\mathbb{R}) \cap \mathrm{O}(n,\mathbb{R}) = \{1\}$.
- $T(n,\mathbb{R})$ is a normal subgroup of $\mathrm{Isom}(\mathbb{R}^n)$.

As we'll see later, these properties would imply that $\mathrm{Isom}(\mathbb{R}^n)$ is isomorphic to a *semidirect product* $T(n,\mathbb{R}) \rtimes \mathrm{O}(n,\mathbb{R})$. We will introduce this notion later, and it will be necessary for describing the wallpaper groups.

Definition 5.2. Let G be a subgroup of $\mathrm{Isom}(\mathbb{R}^n)$.

- Its image under π will be denoted by $\overline{G} = \pi(G) \subseteq \mathrm{O}(n, \mathbb{R})$, and will be called the *point group* of G.
- The kernel of the composition $G \subseteq \mathrm{Isom}(\mathbb{R}^n) \to \mathrm{O}(n, \mathbb{R})$, which is the intersection $G \cap T(n, \mathbb{R})$, will be called the *translation subgroup* of G, and be denoted by L_G.

Note that since $T(n, \mathbb{R}) \cong \mathbb{R}^n$, the translation subgroup L_G can also be considered as a subgroup of $\mathbb{R}^n$.

The following proposition is a key observation for our later discussions.

Proposition 5.3. *The point group $\overline{G} \subseteq \mathrm{O}(n, \mathbb{R})$ sends $L_G \subseteq \mathbb{R}^n$ to itself, therefore gives an action on L_G.*

Proof. For any $A \in \overline{G}$ and $\ell \in L_G$ (i.e. the translation $T_\ell \in G$), we would like to show that $A\ell \in L_G$ (i.e. $T_{A\ell} \in G$). By the definition of the point group, there exists $g \in G$ such that $\pi(g) = A$, say $g(\vec{x}) = A\vec{x} + \vec{v}$ for all $\vec{x} \in \mathbb{R}^n$. We have

$$gT_\ell g^{-1}(\vec{x}) = gT_\ell(A^{-1}(\vec{x} - \vec{v})) = g(A^{-1}(\vec{x} - \vec{v}) + \ell) = \vec{x} - \vec{v} + A\ell + \vec{v} = \vec{x} + A\ell.$$

Hence $T_{A\ell} = gT_\ell g^{-1} \in G$. $\square$

The shapes/patterns that we'll be considering, like frieze patterns or wallpaper patterns, satisfy the property that the translation subgroups of their symmetry groups are *discrete*. Intuitively, it means that their symmetry groups do not contain any *continuous family* of isometries. Let us try to make it more precise.

Definition 5.4. A subgroup L of $\mathbb{R}^n$ is called *discrete* if there exists $\epsilon > 0$ such that $d(\ell_1, \ell_2) > \epsilon$ for any distinct vectors $\ell_1, \ell_2 \in \mathbb{R}^n$.

Definition 5.5. A subgroup $G \subseteq \mathrm{Isom}(\mathbb{R}^2)$ is called a *plane crystallographic group* if its translation subgroup L_G is *discrete*, and its point group $\overline{G}$ is finite.

The goal of this section is to classify all plane crystallographic groups. Let us begin with classifying discrete subgroups of $\mathbb{R}^2$.

Proposition 5.6. *Let $L \subseteq \mathbb{R}^2$ be a discrete subgroup. Then L is either $\{\vec{0}\}$, $\mathbb{Z}\omega_1$, or $\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ for a pair of linearly independent vectors $\omega_1, \omega_2 \in \mathbb{R}^2$.*

Proof. Suppose $L \neq \{\vec{0}\}$. Since L is discrete, there exists a vector $\omega_1 \in L \backslash \{\vec{0}\}$ with $|\omega_1|$ minimal. Since $L \subseteq \mathbb{R}^2$ is a subgroup, we have $\mathbb{Z}\omega_1 \subseteq L$. Moreover, by the minimality of $|\omega_1|$, it is not hard to see that $t\omega_1 \notin L$ for any $t \in \mathbb{R}\backslash\mathbb{Z}$.

Now, if $\mathbb{Z}\omega_1 = L$ then we're finished. Otherwise, choose an $\omega_2 \in L\backslash\mathbb{Z}\omega_1$ with the minimum length among vectors in $L\backslash\mathbb{Z}\omega_1$. We claim that $L = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$. Any $\vec{v} \in L$ can be written as $\vec{v} = t_1\omega_1 + t_2\omega_2$ where $t_1, t_2 \in \mathbb{R}$. The real numbers t_i can be written as $t_i = a_i + b_i$ where $a_i \in \mathbb{Z}$ and $-\frac{1}{2} \leq b_i < \frac{1}{2}$. Then

$$\vec{v} - a_1\omega_1 - a_2\omega_2 = b_1\omega_1 + b_2\omega_2 \in L.$$

If $b_2 \neq 0$, then $b_1\omega_1 + b_2\omega_2 \notin \mathbb{Z}\omega_1$, and by the minimality of ω_2 we have

$$|\omega_2| \leq |b_1\omega_1 + b_2\omega_2| < \frac{1}{2}(|\omega_1| + |\omega_2|) \leq |\omega_2|,$$

contradiction. Therefore $b_2 = 0$ and $b_1\omega_1 \in L$. By the minimality of ω_1, we have $b_1 = 0$, hence $\vec{v} \in \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$. □

Definition 5.7. We say that the discrete subgroup $L \subseteq \mathbb{R}^2$ is a *lattice* with *rank* 0, 1, or 2, depending on $L = \{\vec{0}\}$, $\mathbb{Z}\omega_1$, or $\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$

Let G be a plane crystallographic group.

(1) Suppose the translation subgroup L_G is of rank 0, i.e. $L_G = 0$. Then $G \cong \overline{G}$ is a finite subgroup of $\mathrm{O}(2, \mathbb{R})$. We proved before that such G must be isomorphic to either a cyclic group or a dihedral group.
(2) Suppose the translation subgroup L_G is of rank 1, i.e. $L_G = \mathbb{Z}\omega$ for some nonzero $\omega \in \mathbb{R}^2$. Such group G is called a *frieze* group.
(3) Suppose the translation subgroup L_G is of rank 2, i.e. $L_G = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ for a pair of linearly independent vectors $\omega_1, \omega_2 \in \mathbb{R}^2$. Such group G is called a *wallpaper* group.

We will be classifying the frieze groups and the wallpaper groups in the remainder of this section. Before we proceed, let us take a detour and discuss some useful facts about isometries of $\mathbb{R}^2$.

Lecture 8

First, we claim that the composition of a counterclockwise rotation around the origin (say by angle $0 < \theta < 2\pi$) $R_\theta \in \mathrm{SO}(2, \mathbb{R})$ followed by a translation $T_{\vec{v}}$ is again a rotation by angle θ, but the center of the rotation would be different. This can be proved by direct computations: on the one hand, we have

$$T_{\vec{v}} \circ R_\theta \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} x\cos\theta - y\sin\theta + v_1 \\ x\sin\theta + y\cos\theta + v_2 \end{bmatrix}$$

On the other hand, the rotation by angle θ centered at $\vec{w} \in \mathbb{R}^2$ is the same as $T_{\vec{w}} \circ R_\theta \circ T_{-\vec{w}}$. So, to prove our claim, it suffices to show that there exists $\vec{w}$ such that

$$T_{\vec{w}} \circ R_\theta \circ T_{-\vec{w}} \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} x\cos\theta - y\sin\theta + v_1 \\ x\sin\theta + y\cos\theta + v_2 \end{bmatrix} \text{ holds for all } \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2.$$

One can directly compute the left hand side, it is:

$$\begin{bmatrix} (x-w_1)\cos\theta - (y-w_2)\sin\theta + w_1 \\ (x-w_1)\sin\theta + (y-w_2)\cos\theta + w_2 \end{bmatrix}.$$

Finding the vector $\vec{w} = [w_1, w_2]^T$ then becomes a basic linear-algebraic problem; such $\vec{w}$ should be:

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \frac{1}{4\sin^2\frac{\theta}{2}} \begin{bmatrix} 1-\cos\theta & -\sin\theta \\ \sin\theta & 1-\cos\theta \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

Second, we claim that any non-trivial isometry of $\mathbb{R}^2$ is either a translation, a rotation, a reflection, or a *glide reflection*. A *glide reflection* is the composition of a reflection along a line ℓ followed by a translation parallel to ℓ. To show the claim, recall that any isometry of $\mathbb{R}^2$ can be written as a composition $T_{\vec{v}} \circ A$, where $A \in \mathrm{O}(2,\mathbb{R})$ is an orthogonal transformation and $T_{\vec{v}}$ is a translation. Also recall that any element of $\mathrm{O}(2,\mathbb{R})$ is either the identity, a rotation, or a reflection. If A is the identity, then $T_{\vec{v}} \circ A$ is a translation. If A is a rotation, then $T_{\vec{v}} \circ A$ is also a reflection by the previous claim. Finally, if A is a reflection, say along the mirror line ℓ. One can decompose $\vec{v} = \vec{v}_1 + \vec{v}_2$ where $\vec{v}_1$ is parallel to ℓ and $\vec{v}_2$ is perpendicular to ℓ, and therefore have $T_{\vec{v}} = T_{\vec{v}_1} \circ T_{\vec{v}_2}$. One can check that $T_{\vec{v}_2} \circ A$ coincides with the reflection in the line $\frac{\vec{v}_2}{2} + \ell$. So, when $\vec{v}_1 = \vec{0}$ the isometry $T_{\vec{v}} \circ A$ is a reflection; and when $\vec{v}_1 \neq \vec{0}$ the isometry $T_{\vec{v}} \circ A$ is a glide reflection.

5.2. **Classification of frieze groups.** Let G be a plane crystallographic group. Suppose the translation subgroup L_G is of rank 1, i.e. $L_G = \mathbb{Z}\omega$ for some nonzero $\omega \in \mathbb{R}^2$. Let us denote the translation $\vec{x} \mapsto \vec{x} + \omega$ by $T \in G$. Recall that the point group $\overline{G} \subseteq \mathrm{O}(2,\mathbb{R})$ sends $L_G = \mathbb{Z}\omega$ to itself. Any such orthogonal map must be either the identity, a rotation R of angle π around the origin, a reflection M in the line $\mathbb{R}\omega$, or a reflection N in the line through the origin orthogonal to ω. They form the Klein four group $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, and $\overline{G}$ must be one of the following:

$$\{I\}, \{I, R\}, \{I, M\}, \{I, N\}, \{I, R, M, N\}.$$

Case (a): $\overline{G} = \{I\}$.

In this case, $G = L_G = \mathbb{Z}\omega \cong \mathbb{Z}$ is a cyclic group of infinite order, which consists entirely of translations. Such G is the symmetry group of a *frieze pattern* such as:

In terms of the IUC notation (short for International Union of Crystallography), this case is denoted by **(p1)**.

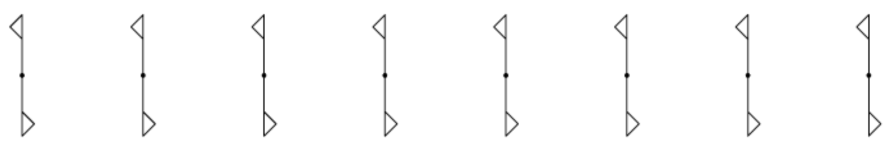
Case (b): $\overline{G} = \{I, R\}$.

Then there exists $A \in G$ such that $\pi(G) = R$. Suppose $A = T_{\vec{v}} \circ R$ for some $\vec{v} \in \mathbb{R}^2$. One can check that A is the rotation of angle π around the point $\vec{v}/2$. By choosing $\vec{v}/2$ as the new origin, one can assume that $A = R \in G$.

Since $G/L_G \cong \overline{G}$, the lattice $L_G = \mathbb{Z}\omega = \langle T \rangle$ is an index two subgroup of G. Hence any element of G is either T^k or $T^k R$ for some $k \in \mathbb{Z}$. Observe that any rotation R and translation T are related by $RTR^{-1} = T^{-1}$. (This can be proved easily by first observing that $RTR^{-1} \in \operatorname{Ker}(\pi)$, hence is a translation; then plug in the zero vector to find the amount of translation.) Therefore, the group G is isomorphic to the *infinite dihedral group* D_∞

$$D_\infty = \{R, T \mid R^2 = 1 \text{ and } RTR^{-1} = T^{-1}\}.$$

Such G is the symmetry group of a frieze pattern such as:



The IUC notation of this case is **(p2)**.

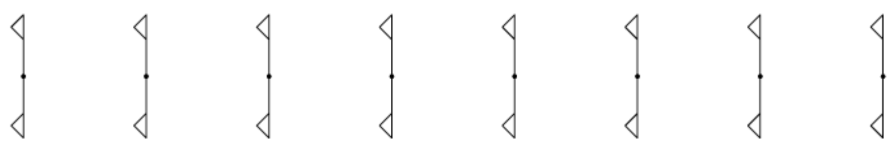
Exercise. Geometrically, $T^k R$ is the rotation by angle π around the point $k\omega/2$.

Lecture 7: Midterm exam

Case (c): $\overline{G} = \{I, M\}$.

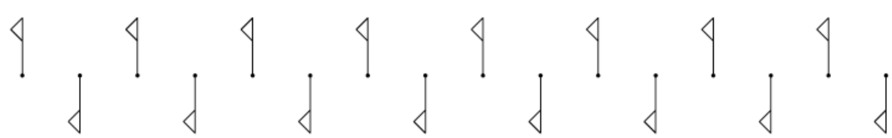
Then there exists $A \in G$ such that $\pi(G) = M$. The element A can be written as $A = T_{a\omega+\vec{x}} \circ M$ for some $a \in \mathbb{R}$ and $\vec{x} \in \omega^\perp$. One can check that $T_{\vec{x}} \circ M$ coincides with the reflection in the line $\vec{x}/2 + \mathbb{R}\omega$. By choosing the new origin to be $\vec{x}/2$, one can assume that $A = T_{a\omega} \circ M$, i.e. A is either a reflection in $\mathbb{R}\omega$ (when $a = 0$), or a *glide reflection* in $\mathbb{R}\omega$ (when $a \neq 0$). (A *glide reflection* is a reflection followed by a translation parallel to the reflection axis.)

In the first case, G contains the translations T^k, the reflection A, and the glide reflections $T^k A$. Note that $AT = TA$, hence $G \cong \mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$.



The IUC notation of this case is **(p11m)**.

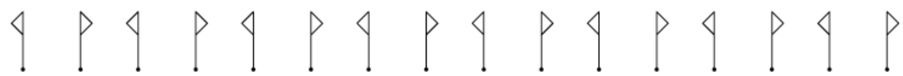
In the second case, note that A^2 is a translation, so $A^2 = T^r$ for some $r \in \mathbb{Z}\backslash\{0\}$. Then A is the reflection followed by translation by $r\omega/2$. If $r = 2k$ is even, then $T^{-k}A \in G$ is the reflection, so we're back in the previous case. If $r = 2k+1$ is odd, then $T^{-k}A \in G$ is the reflection followed by translation by $\omega/2$. This generates the group $G = \langle T^{-k}A \rangle \cong \mathbb{Z}$.



The IUC notation of this case is **(p11g)**.

Case (d): $\overline{G} = \{I, N\}$.

Then there exists $A \in G$ such that $\pi(G) = N$. By the same argument as above, this means that A is either a reflection in a line orthogonal to ω, or a *glide reflection* orthogonal to ω. In the second case, A^2 would be a translation orthogonal to ω, which is impossible. So A is a reflection in a line orthogonal to ω. Choose a coordinate so that the origin is on the mirror. Observe that $ATA^{-1} = T^{-1}$, hence the group G is isomorphic to the infinite dihedral group D_∞.



The IUC notation of this case is **(p1m1)**.

Case (e): $\overline{G} = \{I, R, M, N\}$.

As in Case (b), we choose coordinate so that $R \in G$. There exists $A \in G$ so that $\pi(A) = M$. As in Case (c), A is a (glide) reflection in a line $\omega' + \mathbb{R}\omega$ where ω' is perpendicular to ω. We claim that, now with $R \in G$, we must have $\omega' = 0$, i.e. the mirror (glide) reflection line is exactly $\mathbb{R}\omega$. First, we observe that the element $ARA^{-1}R^{-1} \in G$ is a translation, since

$$\pi(ARA^{-1}R^{-1}) = \pi(A)\pi(R)\pi(A^{-1})\pi(R^{-1}) = MRM^{-1}R^{-1} = RMM^{-1}R^{-1} = 1.$$

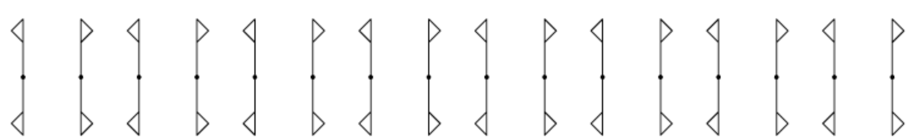
Recall that A is the composition of the reflection in a line $\omega' + \mathbb{R}\omega$ where ω' is perpendicular to ω, followed by a translation by $k\omega$ where $k \in \mathbb{R}$ (when $k = 0$, A is a reflection). Hence

$$ARA^{-1}R^{-1}(\vec{0}) = ARA^{-1}(\vec{0}) = AR(2\omega' - k\omega) = A(-2\omega' + k\omega) = 4\omega' + 2k\omega.$$

Therefore $ARA^{-1}R^{-1} \in G$ is the translation by $4\omega' + 2k\omega$, thus we have $4\omega' + 2k\omega \in L_G$. Now since $L_G = \mathbb{Z}\omega$, one can conclude that $\omega' = 0$, i.e. A is indeed a (glide) reflection in the line $\mathbb{R}\omega$.

First, suppose A is the reflection in $\mathbb{R}\omega$. Then G is generated by T, R, A. Recall that T, R generates the infinite dihedral group, and observe that A commutes with both T and R. Hence

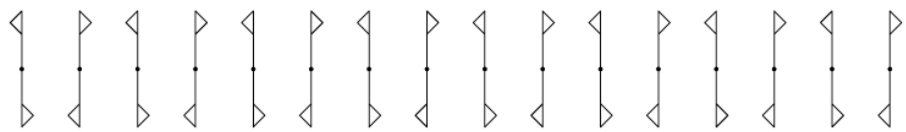
$$\begin{aligned} G &= \langle T, R, A \mid R^2 = A^2 = 1,\ RTR^{-1} = T^{-1},\ AT = TA,\ AR = RA \rangle \\ &\cong D_\infty \times \mathbb{Z}/2\mathbb{Z}. \end{aligned}$$



The IUC notation of this case is **(p2mm)**.

Second, suppose A is a glide reflection in the line $\mathbb{R}\omega$. Then $A^2 = T^r$ for some nonzero $r \in \mathbb{Z}$. If $r = 2k$ is even, then $T^{-k}A \in G$, so we're back in the previous case. If $r = 2k+1$ is odd, then $C := T^{-k}A$ is the reflection followed by translation by $\omega/2$, which generates the cyclic group $\langle T, A\rangle$. Observe that $RCR^{-1} = C^{-1}$. Hence

$$G = \left\langle C, R \mid R^2 = 1 \text{ and } RCR^{-1} = C^{-1}\right\rangle \cong D_\infty.$$



The IUC notation of this case is **(p2mg)**.

5.3. **Semidirect products.** In order to classify the wallpaper groups, which are plane crystallographic groups whose translation subgroups are of rank 2, we have to first introduce the notion of *semidirect products*.

Let us begin with recalling some of the basic properties of *direct products*. Let $(G_1, \cdot_1)$ and $(G_2, \cdot_2)$ be two groups. The direct product $G = G_1 \times G_2$ is a group with the binary operation given by $(g_1, g_2) \cdot (g_1', g_2') = (g_1 \cdot_1 g_1', g_2 \cdot_2 g_2')$. Observe that G_1 is isomorphic to the subgroup $G_1 \times \{e_2\}$ of G; similarly, G_2 is isomorphic to the subgroup $\{e_1\} \times G_2$ of G. Here are three properties of these two subgroups of G:

- They generate $G_1 \times G_2$: $(g_1, g_2) = (g_1, 1)(1, g_2)$.
- They intersect trivially: If $(g_1, 1) = (1, g_2)$, then $g_1 = e_1$ and $g_2 = e_2$.
- They commute with each other: $(g_1, 1)(1, g_2) = (1, g_2)(g_1, 1)$.

It turns out that these three properties *characterize* direct products. More precisely, it is not hard to show the following.

Exercise. Let G be a group with subgroups H and K. Suppose that

- $G = HK$, i.e. for any $g \in G$, there exists $h \in H$ and $k \in K$ such that $g = hk$;
- $H \cap K = \{1\}$ in G;
- $hk = kh$ for any $h \in H$ and $k \in K$.

Then the map $H \times K \to G$ defined by $(h, k) \mapsto hk$ is a group isomorphism.

We have encountered several examples of such G, H, K where the first two conditions are satisfied, but elements of H and K do not commute. For instance, let $G = D_n$ be a dihedral group, and consider $H = \{1, r, \ldots, r^{n-1}\}$ the subgroup of rotations, and $K = \{1, s\}$ where s is a reflection. One can easily check that $G = HK$ and $H \cap K = \{1\}$. But elements of H and K do not commute: $sr = r^{-1}s$. Indeed, in this case $G \ncong H \times K$. Rather, G is isomorphic to a *semidirect product* of H and K. Another example we have seen is when $G = \mathrm{Isom}(\mathbb{R}^n)$, $H = T(n, \mathbb{R})$, and $K = \mathrm{O}(n, \mathbb{R})$.

Let us take a step back, and examine the first condition. Suppose H and K are subgroups of G. Is the product $HK = \{hk \mid h \in H, k \in K\}$ always a subgroup of G? The answer is no. For instance, let $G = S_3$, and let $H = \langle (12) \rangle$ and $K = \langle (13) \rangle$. Then one can check that HK consists of four elements, therefore cannot be a subgroup of S_3. However, if H or K is *normal* in G, then HK would be a subgroup. Say H is a normal subgroup of G. Then we have

$$(hk)(h'k') = (hkh'k^{-1})(kk') \in HK \quad \text{and} \quad (hk)^{-1} = (k^{-1}h^{-1}k)k^{-1} \in HK.$$

Exercise. Let $G = G_1 \times G_2$ be the direct product of two groups G_1 and G_2. Prove that the subgroups $G_1 \times \{e_2\}$ and $\{e_1\} \times G_2$ of G are both normal.

Observe that in the formula "$(hk)(h'k') = (hkh'k^{-1})(kk')$", it involves the element $kh'k^{-1} \in H$, which can be regarded as the element $k \in K$ acts on $h' \in H$ by conjugation, which is an action by *automorphisms* of H (i.e. the map $\mathrm{Ad}_k \colon H \to H$ defined by conjugation $h' \mapsto kh'k^{-1}$ is an isomorphism). This motivates the following definition.

Definition 5.8. Let H and K be any two groups (here we don't assume that they lie inside a common group G), and let $\varphi \colon K \to \mathrm{Aut}(H)$ be an action of K on H by automorphisms (i.e. $\varphi_k \colon H \to H$ is a group isomorphism for any $k \in K$). We define the corresponding *semidirect product* $H \rtimes_\varphi K$ as follows:

- as a set, it is $\{(h, k) \mid h \in H, k \in K\}$;
- its group law is given by

$$(h, k)(h', k') = (h\varphi_k(h'), kk').$$

Exercise. Verify that the semidirect product defined above is a group. (This is a quite non-trivial exercise!)

Exercise. Prove that both $\{(h,1) \mid h \in H\}$ and $\{(1,k) \mid k \in K\}$ are subgroups of $H \rtimes_\varphi K$, which isomorphic to H and K, respectively. Also, show that the map $H \rtimes_\varphi K \to K$ defined by $(h,k) \mapsto (1,k)$ is a group homomorphism, with kernel $\{(h,1) \mid h \in H\} \cong H$. Therefore H is isomorphic to a *normal* subgroup of the semidirect product. (On the other hand, the subgroup K is usually *not* normal in $H \rtimes_\varphi K$.)

Example. When $\varphi\colon K \to \mathrm{Aut}(H)$ is the trivial action, i.e. $\varphi_k = \mathrm{id}_H$ for all $k \in K$, the group law of the semidirect product becomes

$$(h,k)(h',k') = (h\varphi_k(h'), kk') = (hh', kk'),$$

which simply gives the direct product $H \times K$. Hence the direct product is a special case of semidirect products.

Theorem 5.9. *Let G be a group with subgroups H and K, such that*

- $G = HK$,
- $H \cap K = \{1\}$,
- *H is normal in G.*

Then:

(1) the map $\varphi\colon K \to \mathrm{Aut}(H)$ defined by conjugacy actions $\varphi_k(h) = khk^{-1}$ is a group homomorphism,
(2) the map $f\colon H \rtimes_\varphi K \to G$ defined by $f(h,k) = hk$ is a group isomorphism.

Proof. The first statement can be checked straightforwardly. As for the second statement, the map f is surjective by $G = HK$, is injective by $H \cap K = \{1\}$ (why?), and is a group homomorphism since

$$\begin{aligned} f((h,k)(h',k')) &= f(h\varphi_k(h'), kk') \\ &= f(hkh'k^{-1}, kk') \\ &= hkh'k^{-1}kk' \\ &= hkh'k' \\ &= f(h,k)f(h',k'). \end{aligned}$$

□

This theorem implies that we have semidirect products

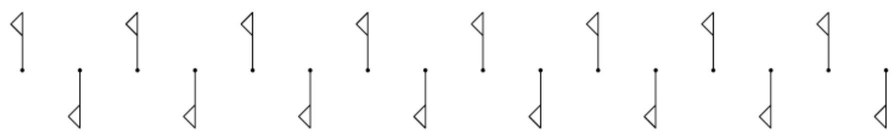
$$D_n = \{1, r, \dots, r^{n-1}\} \rtimes_\varphi \{1, s\} \cong (\mathbb{Z}/n\mathbb{Z}) \rtimes_\varphi (\mathbb{Z}/2\mathbb{Z})$$

and

$$\mathrm{Isom}(\mathbb{R}^n) = T(n, \mathbb{R}) \rtimes_\varphi \mathrm{O}(n, \mathbb{R}),$$

where φ is given by the conjugacy action described in the theorem.

Remark 5.10. Let $G \subseteq \mathrm{Isom}(\mathbb{R}^n)$. Although there is an isomorphism $G/L_G \cong \overline{G}$ induced by $\pi\colon G \to \overline{G}$, the group G is in general *not* isomorphic to a semidirect product of its translation subgroup L_G and its point group $\overline{G}$. For instance, consider the symmetry group of the following frieze pattern. In this



example, G is an infinite cyclic group generated by the reflection M followed by the translation by $\omega/2$; L_G is an infinite cyclic group generated by the translation by ω; and $\overline{G} = \{1, M\} \cong \mathbb{Z}/2\mathbb{Z}$. Observer that G cannot be isomorphic to a semidirect product of L_G and $\overline{G}$: any semidirect product of L_G and $\overline{G}$ would contain a subgroup isomorphic to $\overline{G} \cong \mathbb{Z}/2\mathbb{Z}$, but $G \cong \mathbb{Z}$ does not contain any element of order 2.

Lecture 9

5.4. **Classification of wallpaper groups.** Let G be a plane crystallographic group. Suppose the translation subgroup L_G is of rank 2, i.e. $L_G = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ for a pair of linearly independent vectors $\omega_1, \omega_2 \in \mathbb{R}^2$. Recall that the point group $\overline{G} \subseteq \mathrm{O}(2, \mathbb{R})$ sends the lattice $L_G \subseteq \mathbb{R}^2$ to itself. We can use this fact to classify all possible point groups.

Lemma 5.11. *Let G be a wallpaper group, and let $\overline{G}$ be its point group.*

- *Any element $g \in \overline{G}$ has order either $1, 2, 3, 4,$ or 6.*
- *$\overline{G}$ is isomorphic to one of the following groups*

$$\{1\},\ \mathbb{Z}/2\mathbb{Z},\ \mathbb{Z}/3\mathbb{Z},\ \mathbb{Z}/4\mathbb{Z},\ \mathbb{Z}/6\mathbb{Z},\ D_2,\ D_3,\ D_4,\ D_6.$$

Proof. Let us prove the first statement. Suppose $g \in \overline{G} \subseteq \mathrm{O}(2, \mathbb{R})$ is a reflection, then it has order 2. Otherwise, g is a rotation, say of order n; then G

would contain the counterclockwise rotation by $2\pi/n$ (denoted by $R_{2\pi/n}$). Let $\vec{v}$ be a shortest nonzero vector in the lattice L_G. It is not hard to show:

- If $n > 6$, then $R_{2\pi/n}\vec{v} - \vec{v} \in L_G$ is shorter than $\vec{v}$, contradiction.
- If $n = 5$, then $R^2_{2\pi/n}\vec{v} + \vec{v} \in L_G$ is shorter than $\vec{v}$, contradiction.

Therefore, the order n can only be either $1, 2, 3, 4,$ or 6.

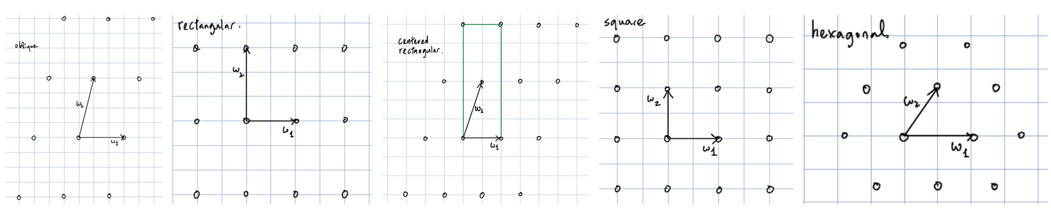
Recall that any finite subgroup of $\mathrm{O}(2, \mathbb{R})$ is either isomorphic to a cyclic group or a dihedral group, the second statement then follows immediately. □

In order to classify all wallpaper patterns, one needs to first classify rank two lattices $L = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ in $\mathbb{R}^2$. (The symmetries that L_G possess would directly impact the wallpaper group G.) Recall that one can choose $\omega_1 \in L$ to be a vector with minimum length in $L\backslash\{\vec{0}\}$, and choose $\omega_2 \in L$ to be a vector with minimum length in $L\backslash\mathbb{Z}\omega_1$. By possibly replacing ω_2 by $-\omega_2$, we can assume that $|\omega_1 - \omega_2| \leq |\omega_1 + \omega_2|$. We thus have

$$|\omega_1| \leq |\omega_2| \leq |\omega_1 - \omega_2| \leq |\omega_1 + \omega_2|.$$

We can then classify rank two lattices according to the inequalities into the following types.

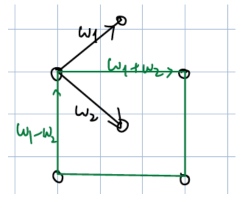
(a) oblique: $|\omega_1| < |\omega_2| < |\omega_1 - \omega_2| < |\omega_1 + \omega_2|$;
(b) rectangular: $|\omega_1| < |\omega_2| < |\omega_1 - \omega_2| = |\omega_1 + \omega_2|$;
(c) centered rectangular: $|\omega_1| < |\omega_2| = |\omega_1 - \omega_2| < |\omega_1 + \omega_2|$;
(d) square: $|\omega_1| = |\omega_2| < |\omega_1 - \omega_2| = |\omega_1 + \omega_2|$;
(e) hexagonal: $|\omega_1| = |\omega_2| = |\omega_1 - \omega_2| < |\omega_1 + \omega_2|$.



There are three more possibilities, in which two of them,

$$|\omega_1| < |\omega_2| = |\omega_1 - \omega_2| = |\omega_1 + \omega_2| \quad \text{and} \quad |\omega_1| = |\omega_2| = |\omega_1 - \omega_2| = |\omega_1 + \omega_2|$$

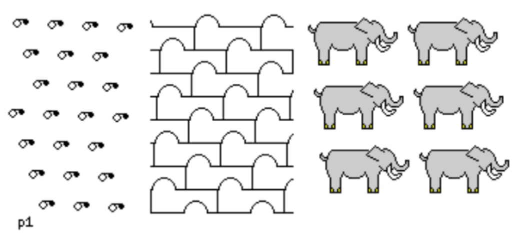
have no corresponding lattices so that the inequalities hold. The remaining one $|\omega_1| = |\omega_2| < |\omega_1 - \omega_2| < |\omega_1 + \omega_2|$ actually has a centered rectangular structure whose rectangles are based on the vectors $\omega_1 - \omega_2$ and $\omega_1 + \omega_2$. It turns out that the wallpaper groups we get from this case would be isomorphic



to the centered rectangular case. Therefore, we only need to consider the five types of lattices classified above.

Case (a): L_G **is oblique.** The only orthogonal transformations which preserve L_G are $\{\pm I\}$. Hence the point group $\overline{G}$ is a subgroup of $\{\pm I\}$.

(p1) $\overline{G} = \{I\}$. In this case, we have $G \cong L_G \cong \mathbb{Z}^2$. It is the symmetry group of a wallpaper pattern like:



Before discussing the case $\overline{G} = \{\pm I\}$, let us state a general lemma.

Lemma 5.12. *Let $G \subseteq \mathrm{Isom}(\mathbb{R}^n)$ be a subgroup, and let $L_G \subseteq G$ and $\overline{G} \subseteq \mathrm{O}(n, \mathbb{R})$ be its translation subgroup and point group, respectively. Assume that $\overline{G} \subseteq \mathrm{O}(n, \mathbb{R}) \subseteq \mathrm{Isom}(\mathbb{R}^n)$ is a subgroup of G. Then $G \cong L_G \rtimes_\varphi \overline{G}$, where $\varphi \colon \overline{G} \to \mathrm{Aut}(L_G)$ is given by the conjugacy action.*

Proof. It suffices to show that L_G and $\overline{G}$ satisfy the conditions in Theorem 5.9. Let g be an element of G. Then there exists an orthgonal transformation $A \in \overline{G}$ and a vector $\vec{v} \in \mathbb{R}^n$ such that $g = T_{\vec{v}} \circ A$. Now, since we assume that $\overline{G}$ is a subgroup of G, we have $A \in G$ and therefore $T_{\vec{v}} \in G$. Hence $G = L_G\overline{G}$. The remaining two conditions of Theorem 5.9 are not hard to check. $\square$

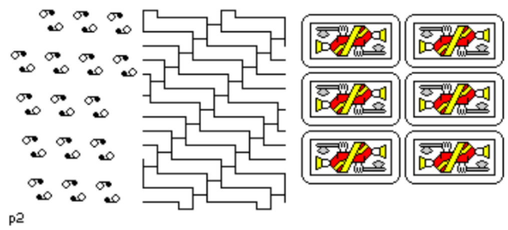
Remark 5.13. In general, $\overline{G}$ is *not* a subgroup of G. For instance, consider the frieze pattern of type (p11g): its symmetry group $G \cong \mathbb{Z}$, while $\overline{G} \cong \mathbb{Z}/2\mathbb{Z}$ cannot be a subgroup of G since $\mathbb{Z}$ does not contain any element of order two.

Remark 5.14. Recall that the translation subgroup L_G can also be considered as a subgroup of $\mathbb{R}^2$, say denoted by $\widetilde{L_G} \subseteq \mathbb{R}^2$. ($T_\ell \in L_G$ if and only if $\ell \in \widetilde{L_G}$.) Since $L_G \cong \widetilde{L_G}$, the group G is also isomorphic to a semidirect product $\widetilde{L_G} \rtimes_\varphi \overline{G}$ under the assumption that $\overline{G}$ happens to be a subgroup of G. Here, the action $\varphi\colon \overline{G} \to \mathrm{Aut}(\widetilde{L_G})$ is given by $\varphi_A\colon \ell \mapsto A\ell$, since $AT_\ell A^{-1} = T_{A\ell}$.

(p2) $\overline{G} = \{\pm I\}$. By choosing a new origin, we can assume that $-I \in G$. Then $\overline{G}$ is a subgroup of G. By Lemma 5.12 and Remark 5.14, we have $G \cong \widetilde{L_G} \rtimes_\varphi \{\pm I\}$, where $\varphi_{-I}\colon \ell \mapsto -\ell$. Hence $\varphi_{-I}(m\omega_1 + n\omega_2) = -m\omega_1 - n\omega_2$. Thus

$$G \cong \widetilde{L_G} \rtimes_\varphi \{\pm I\} \cong \mathbb{Z}^2 \rtimes_\varphi (\mathbb{Z}/2\mathbb{Z}),$$

where the action of the non-identity element of $\mathbb{Z}/2\mathbb{Z}$ on $\mathbb{Z}^2$ is given by $(m,n) \mapsto (-m,-n)$.



Case (b): L_G is rectangular. In this case, there are four orthogonal transformations which preserve L_G, namely: the identity, rotation by π (i.e. $-I$), reflection in the x-axis (denoted M_0), and reflection in the y-axis (denoted M_π).

Notation. We will use $R_\theta \in \mathrm{O}(2,\mathbb{R})$ to denote the counterclockwise rotation of angle θ. We will use M_θ to denote the reflection in the line through the origin which subtends an angle $\theta/2$ with the positive x-axis. Also, for any $f \in \mathrm{Isom}(\mathbb{R}^2)$, there exists a unique pair of $B \in \mathrm{O}(2,\mathbb{R})$ and $\vec{v} \in \mathbb{R}^2$ such that $f = T_{\vec{v}} \circ B$; it is convenient to denote f by $(\vec{v}, B)$.

The point group $\overline{G}$ is then a subgroup of $\{I, -I, M_0, M_\pi\} \cong (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$. Note that if $\overline{G}$ is $\{I\}$ or $\{I, -I\}$, then it becomes Case (p1) or (p2), respectively. In order to find wallpaper groups which we have not seen before, we only need to consider the cases whether $\overline{G}$ is $\{I, M_0\}$, $\{I, M_\pi\}$, or

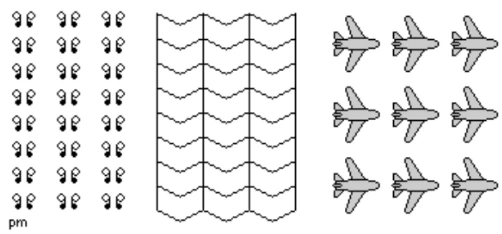
$\{I, -I, M_0, M_\pi\}$. Also, note that taking $\{I, M_\pi\}$ as point group instead of $\{I, M_0\}$ is equivalent to interchanging the roles of "horizontal" and "vertical", which will lead to isomorphic wallpaper group since our lattice L_G is rectangular, so we only need to discuss one of them.

First, suppose $\overline{G} = \{I, M_0\}$. Then there exists $A \in G$ such that $\pi(A) = M_0$. Hence A is a reflection or a glide reflection in a line parallel to the x-axis. By choosing a new origin, one can assume that A is a reflection or a glide reflection in the x-axis.

(pm) $A \in G$ is the reflection in x-axis, i.e. $M_0 \in G$. Again by Lemma 5.12 and Remark 5.14, we have $G \cong \widetilde{L_G} \rtimes_\varphi \langle M_0 \rangle$, where $\varphi_{M_0} \colon [\ell_1, \ell_2]^T \mapsto [\ell_1, -\ell_2]^T$. Hence $\varphi_{M_0}(m\omega_1 + n\omega_2) = m\omega_1 - n\omega_2$. Thus

$$G \cong \widetilde{L_G} \rtimes_\varphi \langle M_0 \rangle \cong \mathbb{Z}^2 \rtimes_\varphi (\mathbb{Z}/2\mathbb{Z}),$$

where the action of the non-identity element of $\mathbb{Z}/2\mathbb{Z}$ on $\mathbb{Z}^2$ is given by $(m, n) \mapsto (m, -n)$.



(pg) $A \in G$ is a glide reflection in the x-axis. Then A^2 is a translation in the x-direction, hence $A^2 = k\omega_1$ for some $k \in \mathbb{Z}\backslash\{0\}$. Suppose k is even, then $M_0 = T_{-k\omega_1/2} \circ A \in G$, and we're back in the previous case. Suppose k is odd, then

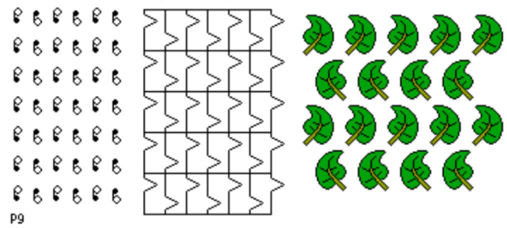
$$\left(\frac{\omega_1}{2}, M_0\right) = T_{-(k-1)\omega_1/2} \circ A \in G.$$

Therefore, any element of G is of the form

$$G = \left\{(m\omega_1 + n\omega_2, I)\,, \left((m + \frac{1}{2})\omega_1 + n\omega_2, M_0\right) \mid m, n \in \mathbb{Z}\right\}.$$

Exercise. Verify that $G \subseteq \mathrm{Isom}(\mathbb{R}^2)$ above forms a group.

Second, suppose $\overline{G} = \{I, -I, M_0, M_\pi\}$. There exists $A_0, A_\pi \in G$ such that $\pi(A_0) = M_0$ and $\pi(A_\pi) = M_\pi$. By choosing a new origin, one can assume that



A_0 is the reflection or a glide reflection in the x-axis, and A_π is the reflection or a glide reflection in the y-axis.

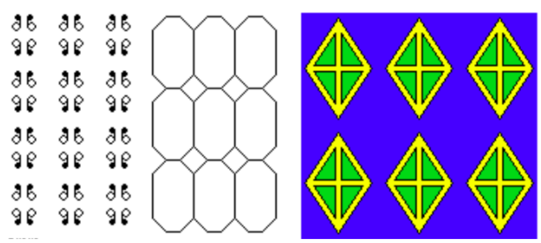
(p2mm) $A_0, A_\pi \in G$ are reflections in the x- and y-axis, respectively. Then we also have $-I = A_0 A_\pi \in G$, hence $\{I, -I, A_0, A_\pi\} \subseteq G$. Hence

$$G = \widetilde{L_G} \rtimes_\varphi \{I, -I, A_0, A_\pi\}.$$

As computed before, the action of A_0 on $\widetilde{L_G}$ is given by: $(\ell_1, \ell_2) \to (\ell_1, -\ell_2)$. Similarly, the action of A_π is given by: $(\ell_1, \ell_2) \to (-\ell_1, \ell_2)$. Hence

$$G = \widetilde{L_G} \rtimes_\varphi \{I, -I, A_0, A_\pi\} \cong \mathbb{Z}^2 \rtimes_\varphi ((\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})),$$

where the action of $(\pm 1, \pm 1) \in (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$ on $\mathbb{Z}^2$ is given by $(m, n) \mapsto (\pm m, \pm n)$.

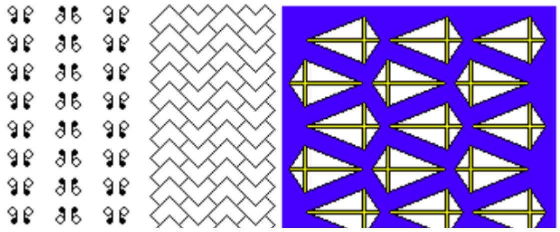


(p2mg) A_0 is the reflection in the x-axis, but A_π is a glide reflection in the y-axis. Then $A_\pi^2 = k\omega_2$ for some $k \in \mathbb{Z}\backslash\{0\}$. If k is even, then $M_\pi = T_{-k\omega_2/2} \circ A_\pi \in G$ and we're back in the previous case. If k is odd, then

$$\left(\frac{\omega_2}{2}, M_\pi\right) = T_{-(k-1)\omega_2/2} \circ A_\pi \in G.$$

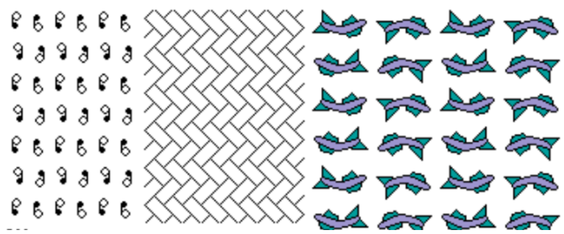
We have found three representatives of three distinct cosets of $L_G \subseteq G$: $(0, I)$, $(0, M_0)$, and $\left(\frac{\omega_2}{2}, M_\pi\right)$. The last coset can be represented by

$$\left(\frac{\omega_2}{2}, M_\pi\right)(0, M_0) = \left(\frac{\omega_2}{2}, -I\right).$$



(p2gg) Both $A_0, A_\pi \in G$ are not reflection in x- or y-axis. By the same argument as above, one can assume that three distinct cosets of $L_G \subseteq G$ can be represented by: $(0, I)$, $(\frac{\omega_1}{2}, M_0)$, and $(\frac{\omega_2}{2}, M_\pi)$. The remaining coset can then be represented by

$$\left(\frac{\omega_2}{2}, M_\pi\right)\left(\frac{\omega_1}{2}, M_0\right) = \left(\frac{-\omega_1 + \omega_2}{2}, -I\right).$$



Case (c): L_G is centered rectangular. In this case, the set of all orthogonal transformations which preserve L_G is the same as the rectangular case, they are $\{I, -I, M_0, M_\pi\}$.

Again, if $\overline{G}$ is $\{I\}$ or $\{I, -I\}$, then it becomes Case (p1) or (p2), respectively. In order to find wallpaper groups which we have not seen before, we only need to consider the cases whether $\overline{G}$ is $\{I, M_0\}$, $\{I, M_\pi\}$, or $\{I, -I, M_0, M_\pi\}$.

(cm) First, suppose $\overline{G} = \{I, M_0\}$. Then there exists $A \in G$ such that $\pi(A) = M_0$. Hence A is a reflection or a glide reflection in a line parallel to the x-axis. By choosing a new origin, one can assume that A is a reflection or a glide reflection in the x-axis. In both cases, A^2 is a translation by $k\omega_1$ for some $k \in \mathbb{Z}$.

If k is even, then $M_0 = T_{-k\omega_1/2} \circ A \in G$. In this case, we have $G \cong \widetilde{L_G} \rtimes_\varphi \langle M_0 \rangle$, where $\varphi_{M_0} \colon [\ell_1, \ell_2]^T \mapsto [\ell_1, -\ell_2]^T$. Hence

$$\varphi_{M_0}(m\omega_1 + n\omega_2) = m\omega_1 + n(\omega_1 - \omega_2) = (m+n)\omega_1 - n\omega_2.$$

Thus

$$G \cong \widetilde{L_G} \rtimes_\varphi \langle M_0 \rangle \cong \mathbb{Z}^2 \rtimes_\varphi (\mathbb{Z}/2\mathbb{Z}),$$

where the action of the non-identity element of $\mathbb{Z}/2\mathbb{Z}$ on $\mathbb{Z}^2$ is given by $(m, n) \mapsto (m + n, -n)$.

If k is odd, then $T_{-\omega_2} \circ T_{-(k-1)\omega_1/2} \circ A \in G$ is a reflection in a line parallel to the x-axis. So we can choose a new origin so that $M_0 \in G$, which reduces to the previous case.

The same argument can be used to study the case when $\overline{G} = \{I, M_\pi\}$. In this case, the semidirect product is given by $\varphi_{M_\pi}([\ell_1, \ell_2]^T) = [-\ell_1, \ell_2]^T$. Hence

$$\varphi_{M_0}(m\omega_1 + n\omega_2) = m(-\omega_1) + n(-\omega_1 + \omega_2) = (-m - n)\omega_1 + n\omega_2.$$

Thus

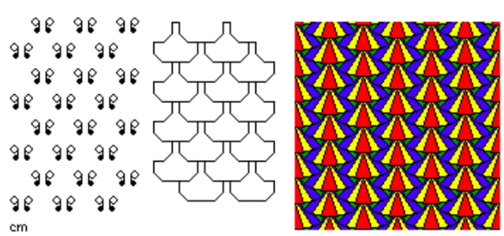
$$G \cong \widetilde{L_G} \rtimes_\varphi \langle M_0 \rangle \cong \mathbb{Z}^2 \rtimes_\varphi (\mathbb{Z}/2\mathbb{Z}),$$

where the action of the non-identity element of $\mathbb{Z}/2\mathbb{Z}$ on $\mathbb{Z}^2$ is given by $(m, n) \mapsto (-m - n, n)$. The following exercise shows that we don't get new wallpaper groups from this case.

Exercise. Let $\varphi_1, \varphi_2 \colon \mathbb{Z}/2\mathbb{Z} = \{\bar{0}, \bar{1}\} \to \mathrm{Aut}(\mathbb{Z}^2)$ be two actions of $\mathbb{Z}/2\mathbb{Z}$ on $\mathbb{Z}^2$ by automorphisms, where the actions of the non-identity element $\bar{1} \in \mathbb{Z}/2\mathbb{Z}$ on $\mathbb{Z}^2$ are given by:

- $\varphi_{1,\bar{1}} \colon (m, n) \mapsto (m + n, -n)$,
- $\varphi_{2,\bar{1}} \colon (m, n) \mapsto (-m - n, n)$.

Prove that the semidirect products $\mathbb{Z}^2 \rtimes_{\varphi_1} (\mathbb{Z}/2\mathbb{Z})$ and $\mathbb{Z}^2 \rtimes_{\varphi_2} (\mathbb{Z}/2\mathbb{Z})$ are isomorphic.

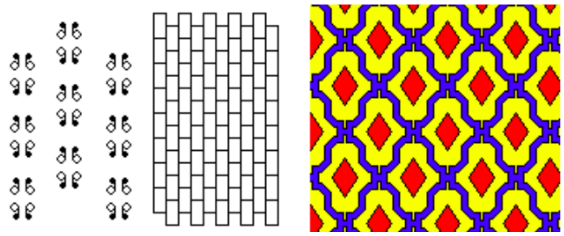


(c2mm) Second, suppose $\overline{G} = \{I, -I, M_0, M_\pi\}$. By the same argument as in the previous case, one can choose a new origin so that $M_0, M_\pi \in G$. Hence

G is a semidirect product of $\widetilde{L_G}$ and $\{I, -I, M_0, M_\pi\} \cong (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$. Moreover, the action φ can also be computed as in the previous case. Hence

$$G \cong \widetilde{L_G} \rtimes_\varphi \{I, -I, M_0, M_\pi\} \cong \mathbb{Z}^2 \rtimes_\varphi \Big((\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})\Big),$$

where the actions of the two non-identity elements from each copy of $\mathbb{Z}/2\mathbb{Z}$ on $\mathbb{Z}^2$ are given by $(m, n) \mapsto (m + n, -n)$ and $(m, n) \mapsto (-m - n, n)$.



Case (d): L_G is a square. The group of orthogonal transformations that preserve L_G is the dihedral group D_4, generated by $R_{\pi/2}$ and M_0. The following exercise shows that, if the point group $\overline{G}$ doesn't contain $R_{\pi/2}$, then we would not obtain new wallpaper groups.

Exercise. Let G be a wallpaper group where L_G is a square lattice. Suppose $\overline{G}$ doesn't contain $R_{\pi/2}$, then G must be isomorphic to the wallpaper groups in the cases of either (p1), (p2), (pm), (pg), (p2mm), (p2mg), (p2gg), (cm), or (c2mm).

Therefore, in order to obtain new wallpaper groups, we only need to consider the case where the point group contains $R_{\pi/2}$. Then, $\overline{G}$ is either the cyclic group of order four generated by $R_{\pi/2}$, or the full dihedral group D_4.

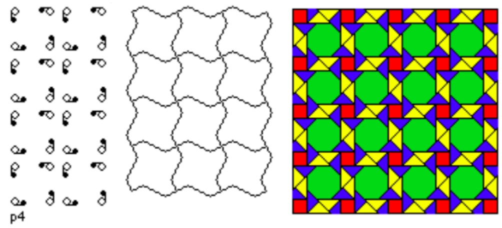
(p4) First, suppose $\overline{G} = \langle R_{\pi/2} \rangle$. By choosing a new origin, one can assume that $R_{\pi/2} \in G$. Hence G can be written as a semidirect product of $\widetilde{L_G}$ and $\langle R_{\pi/2} \rangle$, where the action is given by $\varphi_{R_{\pi/2}} \colon [\ell_1, \ell_2]^T \mapsto [-\ell_2, \ell_1]$. Then

$$\varphi_{R_{\pi/2}}(m\omega_1 + n\omega_2) = -n\omega_1 + m\omega_2.$$

Hence

$$G \cong \mathbb{Z}^2 \rtimes_\varphi (\mathbb{Z}/4\mathbb{Z}),$$

where the action of the generator of $\mathbb{Z}/4\mathbb{Z}$ on $\mathbb{Z}^2$ is given by $(m, n) \mapsto (-n, m)$.



Second, suppose $\overline{G} = \langle R_{\pi/2}, M_0 \rangle$. Again by choosing a new origin, we can assume that the rotation $R_{\pi/2} \in G$. There exists $A \in G$ such that $\pi(A) = M_0$. Such A is either a reflection or a glide reflection in a line parallel to the x-axis.

(p4mm) Suppose A is a reflection in a line parallel to the x-axis. We claim that in this case $M_0 \in G$. Suppose A is the reflection in a line whose intersection with the y-axis is $k\omega_2$, where $k \in \mathbb{R}$. Observe that $T_{\omega_2} \circ A \in G$ is then the reflection in a line also parallel to the x-axis, whose intersection with the y-axis is $(k + \frac{1}{2})\omega_2$. Therefore, we can move the mirror line by an integral multiple of $\frac{\omega_2}{2}$, hence can assume that $0 \le k < \frac{1}{2}$. To prove the claim that $M_0 \in G$, it suffices to show that such k must be zero. Consider the element $AR_{\pi/2}A^{-1}R_{\pi/2} \in G$. It is a translation since it lies in the kernel of π:

$$\pi(AR_{\pi/2}A^{-1}R_{\pi/2}) = M_0 R_{\pi/2} M_0^{-1} R_{\pi/2} = R_{\pi/2}^{-1} R_{\pi/2} = 1.$$

Its amount of translation can be computed by:

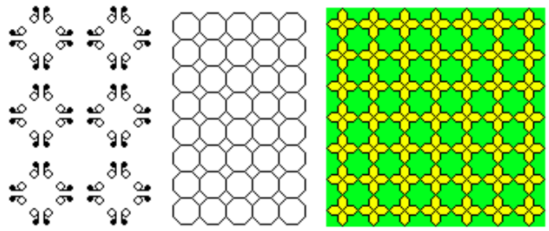
$$AR_{\pi/2}A^{-1}R_{\pi/2}(\vec{0}) = AR_{\pi/2}A^{-1}(\vec{0}) = AR_{\pi/2}(2k\omega_2) = A(-2k\omega_1) = -2k\omega_1 + 2k\omega_2.$$

Hence $-2k\omega_1 + 2k\omega_2 \in L_G$. Since $0 \le k < \frac{1}{2}$, therefore k can only be zero. This shows that $\langle R_{\pi/2}, M_0 \rangle \subseteq G$. Therefore G can be written as a semidirect product of $\widetilde{L_G}$ and $\langle R_{\pi/2}, M_0 \rangle$. We have already computed the action of $R_{\pi/2}$ on $\widetilde{L_G}$ in Case (p4), and the action of $\widetilde{M_0}$ on L_G was computed in Case (pm). We have

$$G \cong \widetilde{L_G} \rtimes_\varphi \langle R_{\pi/2}, M_0 \rangle \cong \mathbb{Z}^2 \rtimes_\varphi D_4 = \mathbb{Z}^2 \rtimes_\varphi \langle r, s \mid r^4 = s^2, srs^{-1} = r^{-1} \rangle,$$

where the action of r on $\mathbb{Z}^2$ is given by $(m, n) \mapsto (-n, m)$, and the action of s on $\mathbb{Z}^2$ is given by $(m, n) \mapsto (m, -n)$.

(p4gm) Now, suppose A is a glide reflection in a line parallel to the x-axis. Then $A^2 = T_{k\omega_1}$ for some integer k. If k is even, then $T_{-k\omega_1/2} \circ A$ is a reflection in a line parallel to the x-axis, and we're back in the previous case. Therefore, we



can assume that k is odd. Furthermore, one can assume that A is a reflection in a line parallel to the x-axis (say its intersection with the y-axis is $k'\omega_2$), followed by the translation by $\omega_1/2$.

Next, we claim that we can choose $A \in G$ so that $k' = 1/4$. By the same argument as in Case (p4mm), we can assume that $0 \leq k' < 1/2$, and compute the translation $AR_{\pi/2}A^{-1}R_{\pi/2} \in G$:

$$\begin{aligned} AR_{\pi/2}A^{-1}R_{\pi/2}(\vec{0}) &= AR_{\pi/2}A^{-1}(\vec{0}) \\ &= AR_{\pi/2}\left(-\frac{1}{2}\omega_1 + 2k'\omega_2\right) \\ &= A\left(-2k'\omega_1 - \frac{1}{2}\omega_2\right) \\ &= \left(-2k' + \frac{1}{2}\right)\omega_1 + \left(2k' + \frac{1}{2}\right)\omega_2. \end{aligned}$$

Hence we must have $-2k' + \frac{1}{2}, 2k' + \frac{1}{2} \in \mathbb{Z}$, and thus $k' = 1/4$. Note that the reflection in a line parallel to the x-axis whose intersection with the y-axis is $\omega_2/4$, is the same as $T_{\omega_2/2} \circ M_0$. This proves that

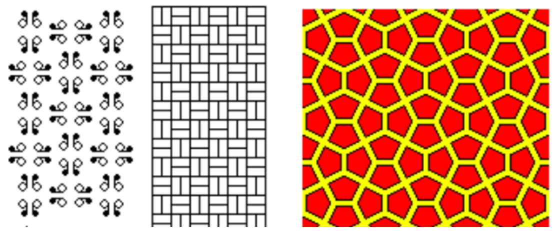
$$\left(\frac{\omega_1 + \omega_2}{2}, M_0\right) = T_{(\omega_1+\omega_2)/2} \circ M_0 \in G.$$

It is then not hard to find representatives of the 8 cosets of $L_G \subseteq G$:

$$(\vec{0}, I), (\vec{0}, R_{\pi/2}), (\vec{0}, R_{\pi}), (\vec{0}, R_{3\pi/2}),$$

$$\left(\frac{\omega_1 + \omega_2}{2}, M_0\right), \left(\frac{\omega_1 + \omega_2}{2}, M_{\pi/2}\right), \left(\frac{\omega_1 + \omega_2}{2}, M_{\pi}\right), \left(\frac{\omega_1 + \omega_2}{2}, M_{3\pi/2}\right).$$

Case (e): L_G is hexagonal. The group of orthogonal transformations that preserve L_G is the dihedral group D_6, generated by $R_{\pi/3}$ and M_0. The following



exercise shows that, if the point group $\overline{G}$ doesn't contain $R_{2\pi/3}$ (rotation of order 3), then we would not obtain new wallpaper groups.

Exercise. Let G be a wallpaper group where L_G is a hexagonal lattice. Suppose $\overline{G}$ doesn't contain $R_{2\pi/3}$, then G must be isomorphic to the wallpaper groups in the cases of either (p1), (p2), (cm), or (c2mm).

Therefore, we only need to consider the case where the point group contains the order 3 rotation $R_{2\pi/3}$.

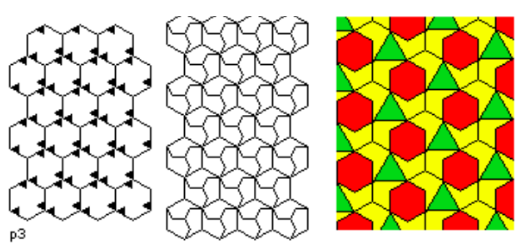
(p3) Suppose $\overline{G} = \langle R_{2\pi/3} \rangle$. By choosing a new origin, we can assume that $R_{2\pi/3} \in G$, hence G is a semidirect product of $\widetilde{L_G}$ and $\langle R_{2\pi/3} \rangle$, where

$$\varphi_{R_{2\pi/3}}(m\omega_1 + n\omega_2) = m(-\omega_1 + \omega_2) + n(-\omega_1) = (-m-n)\omega_1 + m\omega_2.$$

Hence

$$G \cong \mathbb{Z}^2 \rtimes_\varphi (\mathbb{Z}/3\mathbb{Z}),$$

where the action of the generator of $\mathbb{Z}/3\mathbb{Z}$ on $\mathbb{Z}^2$ is given by $(m, n) \mapsto (-m - n, m)$.



(p3m1) Suppose $\overline{G} = \langle R_{2\pi/3}, M_0 \rangle$. It is a (highly) nontrivial exercise to show that there exists a point $p \in \mathbb{R}^2$ such that

- the rotation of angle $2\pi/3$ centered at p lies in G, and
- the reflection along the line passing through p parallel to the x-axis lies in G.

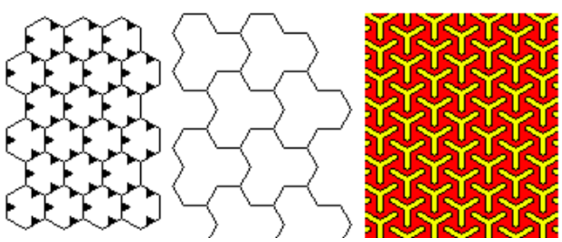
Then, by setting p to be the new origin, one can assume that $\left\langle R_{2\pi/3}, M_0 \right\rangle \subseteq G$. The action of $R_{2\pi/3}$ on $\widetilde{L_G}$ was computed in Case (p3). The action of M_0 on $\widetilde{L_G}$ is given by

$$\varphi_{M_0}(m\omega_1 + n\omega_2) = m\omega_1 + n(\omega_1 - \omega_2) = (m+n)\omega_1 - n\omega_2.$$

Hence we have

$$G \cong \mathbb{Z}^2 \rtimes_\varphi D_3 = \mathbb{Z}^2 \rtimes_\varphi \left\langle r, s \mid r^3 = s^2 = 1, srs^{-1} = r^{-1} \right\rangle,$$

where $\varphi_r \colon (m,n) \mapsto (-m-n, m)$ and $\varphi_s \colon (m,n) \mapsto (m+n, -n)$.



(p31m) Suppose $\overline{G} = \left\langle R_{2\pi/3}, M_{\pi/3} \right\rangle$. It is a nontrivial exercise (easier than the previous case) to show that there exists a point $p \in \mathbb{R}^2$ such that

- the rotation of angle $2\pi/3$ centered at p lies in G, and
- the reflection along the line passing through p parallel to the reflection axis of $M_{\pi/3}$ lies in G.

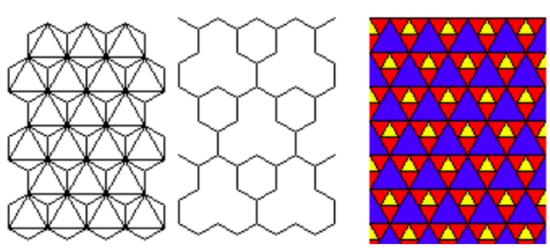
Then, by setting p to be the new origin, one can assume that $\left\langle R_{2\pi/3}, M_{\pi/3} \right\rangle \subseteq G$. The action of $M_{\pi/3}$ on $\widetilde{L_G}$ is given by

$$\varphi_{M_0}(m\omega_1 + n\omega_2) = m\omega_2 + n\omega_1$$

Hence we have

$$G \cong \mathbb{Z}^2 \rtimes_\varphi D_3 = \mathbb{Z}^2 \rtimes_\varphi \left\langle r, s \mid r^3 = s^2 = 1, srs^{-1} = r^{-1} \right\rangle,$$

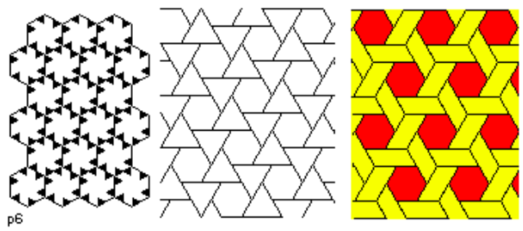
where $\varphi_r \colon (m,n) \mapsto (-m-n, m)$ and $\varphi_s \colon (m,n) \mapsto (n, m)$.



(p6) Suppose $\overline{G} = \langle R_{\pi/3} \rangle$. By choosing a new origin, one can assume that $R_{\pi/3} \in G$. Hence $G \cong \widetilde{L_G} \rtimes_\varphi \langle R_{\pi/3} \rangle$, where the action is given by

$$\varphi_{R_{\pi/3}}(m\omega_1 + n\omega_2) = m\omega_2 + n(-\omega_1 + \omega_2) = -n\omega_1 + (m+n)\omega_2.$$

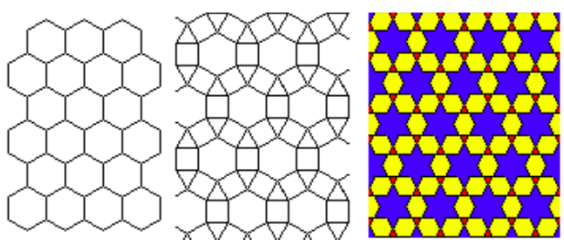
Thus we have $G \cong \mathbb{Z}^2 \rtimes_\varphi (\mathbb{Z}/6\mathbb{Z})$, where the action of the generator of $\mathbb{Z}/6\mathbb{Z}$ on $\mathbb{Z}^2$ is given by $(m, n) \mapsto (-n, m+n)$.



(p6mm) Suppose $\overline{G} = \langle R_{\pi/3}, M_0 \rangle$. By doing the same exercise as in Case (p3m1), one can choose a new origin so that $\overline{G} \subseteq G$. Hence

$$G \cong \mathbb{Z}^2 \rtimes_\varphi D_6 = \mathbb{Z}^2 \rtimes_\varphi \langle r, s \mid r^6 = s^2 = 1, srs^{-1} = r^{-1} \rangle,$$

where $\varphi_r \colon (m, n) \mapsto (-n, m+n)$ and $\varphi_s \colon (m, n) \mapsto (m+n, -n)$.



We would like to show that the 17 types of wallpaper groups we found above are not isomorphic to each other. The following is a key theorem for distinguishing different wallpaper groups.

Theorem 5.15. *An isomorphism between wallpaper groups must take translations to translations, rotations to rotations, reflections to reflections, and glide reflections to glide reflections.*

Proof. Let $f \colon G \to G'$ be an isomorphism between wallpaper groups. Let $T \in G$ be a translation. Note that rotations and reflections are of finite order, and translations and glide reflections are of infinite order. Hence $f(T)$ must be either a translation or a glide reflection. Suppose $f(T)$ is a glide reflection.

Choose a translation $T' \in G'$ whose direction is not parallel to the line of the glide of $f(T)$. Then $f(T)$ and T' do not commute. There exists a unique $g \in G$ such that $f(g) = T'$, and such g must be either a translation or a glide reflection. In any case, g^2 is a translation. Since both T and g^2 are translation, they commute: $Tg^2 = g^2T$. Hence $f(T)T'^2 = T'^2 f(T)$. However, T'^2 is still a translation whose direction is not parallel to the line of the glide of $f(T)$, hence $f(T)T'^2 \neq T'^2 f(T)$. Contradiction. This shows that under isomorphisms between wallpaper groups, translations are mapped to translations, and glide reflections are mapped to glide reflections.

Let $M \in G$ be a reflection. Since M has order two, its image $f(M)$ is either a reflection or a rotation by π. Assume that $f(M)$ is a rotation by π. Choose a translation $T \in G$ in a direction which is not perpendicular to the mirror line of M. Then TM is a glide reflection, which is of infinite order. On the other hand, $f(TM) = f(T)f(M)$ is the composition of a translation and a rotation by π, which is another rotation by π. Hence $f(TM)$ is of order two. Contradiction. Thus f must take reflections to reflections, and therefore takes rotations to rotations. □

Corollary 5.16. *If two wallpaper groups are isomorphic, then their point groups are also isomorphic.*

Proof. Suppose $f\colon G \to G'$ is an isomorphism between wallpaper groups. By the previous theorem, we have $f(L_G) = L_{G'}$. The corollary then follows from the fact that $\overline{G} \cong G/L_G$. □

Below is the list of all 17 wallpaper groups we found above, sorted by their point groups.

- $\overline{G} = \{e\}$: (p1).
- $\overline{G} \cong \mathbb{Z}/2\mathbb{Z}$: (p2), (pm), (pg), (cm).
- $\overline{G} \cong (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$: (p2mm), (p2mg), (p2gg), (c2mm).
- $\overline{G} \cong \mathbb{Z}/4\mathbb{Z}$: (p4).
- $\overline{G} \cong D_4$: (p4mm), (p4gm).
- $\overline{G} \cong \mathbb{Z}/3\mathbb{Z}$: (p3).
- $\overline{G} \cong D_3$: (p3m1), (p31m).
- $\overline{G} \cong \mathbb{Z}/6\mathbb{Z}$: (p6).
- $\overline{G} \cong D_6$: (p6mm).

Proposition 5.17. *No two of (p2), (pm), (pg), (cm) are isomorphic.*

Proof. Among these only (p2) contains rotations, so it can not be isomorphic to any of the others. Of the remaining three groups, (pg) is the only one which does not contain any reflection, so it is not isomorphic to (pm) or (cm). It remains to show that (pm) and (cm) are not isomorphic.

Let us consider the reflection M_0, the translations T_{ω_1} and T_{ω_2} in Case (cm). They satisfy the following properties:

- $M_0 T_{\omega_1} = T_{\omega_1} M_0$.
- $T_{\omega_1}^{-1} T_{\omega_2}^2 M_0$ is a reflection.
- There does not exist a translation $g \in G_{cm}$ such that $T_{\omega_1} = g^2$.

Assume that there is an isomorphism $f\colon G_{cm} \to G_{pm}$, then by Theorem 5.15, $M := f(M)$ is a reflection in G_{pm} and $T_1 := f(T_{\omega_1}), T_2 := f(T_{\omega_2})$ are translations in G_{pm}, with the properties that

- $MT_1 = T_1 M$.
- $T_1^{-1} T_2^2 M$ is a reflection.
- There does not exist a translation $g \in G_{pm}$ such that $T_1 = g^2$.

Let us write $T_1 = T_{m_1\omega_1 + n_1\omega_2}$ and $T_2 = T_{m_2\omega_1 + n_2\omega_2}$. Since M and T_1 commute, we have $n_1 = 0$. Since $T_1^{-1} T_2^2 M$ is a reflection, we have $-m_1 + 2m_2 = 0$, hence m_1 is even, say $m_1 = 2k$ for some $k \in \mathbb{Z}$. This contradicts with the third condition since $T_1 = T_{k\omega_1}^2$ and $T_{k\omega_1} \in G_{pm}$. This proves that G_{cm} and G_{pm} are not isomorphic. □

Proposition 5.18. *No two of (p2mm), (p2mg), (p2gg), (c2mm) are isomorphic.*

Proof. Among these groups, (p2gg) is the only one which does not contain a reflection, so it can not be isomorphic to any of the others. In (p2mg), any reflection has mirror line parallel to the x-axis, hence the composition of any two reflections in (p2mg) is a translation. However, in both (p2mm) and (c2mm), there exists two reflections such that their composition is a rotation (take a horizontal mirror and a vertical mirror). Hence (p2mg) is not isomorphic to (p2mm) or (c2mm). Finally, the same argument as in the proof of the previous proposition can be used to show that (p2mm) is not isomorphic to (c2mm). □

The next two exercises would conclude the classification of wallpaper groups.

Exercise. In this exercise, you'll show that (p4mm) is not isomorphic to (p4gm).

- Show that any rotation of order 4 in (p4mm) can be written as M_1M_2, where M_1 and M_2 are both reflections in (p4mm).
- Show that $(\omega_1, R_{\pi/2})$ in (p4gm) cannot be written as a product of two reflections in (p4gm).

Exercise. In this exercise, you'll show that (p3m1) is not isomorphic to (p31m).

- Show that any rotation of order 3 in (p31m) can be written as M_1M_2, where M_1 and M_2 are both reflections in (p31m).
- Show that $(\omega_1, R_{2\pi/3})$ in (p3m1) cannot be written as a product of two reflections in (p3m1).

Lecture 10

6. Riemann sphere and Möbius transformations

6.1. **Riemann sphere; affine transformations and inversion.** In terms of complex numbers $\mathbb{C} \cong \mathbb{R}^2$, the composition of a rotation and a translation can be expressed as

$$z \mapsto e^{i\theta} z + w$$

where $\theta \in \mathbb{R}$ gives the angle of rotation (around the origin), and $w \in \mathbb{C}$ gives the amount of translation. One can consider a slightly more general notion, the *(complex) affine transformations*, which takes the form

$$T\colon z \mapsto Az + B$$

for some $A, B \in \mathbb{C}$. Note that affine transformations are not necessarily isometries on $\mathbb{C} \cong \mathbb{R}^2$ (it does not preserve distances if $|A| \neq 1$). However, it does *preserve angles*: for any three points $z_1, z_2, z_3 \in \mathbb{C}$, the angle from $\overrightarrow{z_1z_2}$ to $\overrightarrow{z_1z_3}$ coincides with the angle from $\overrightarrow{T(z_1)T(z_2)}$ to $\overrightarrow{T(z_1)T(z_3)}$.

Remark 6.1. Angle-preserving maps (also called *conformal maps*) on $\mathbb{C}$ or subsets $U \subseteq \mathbb{C}$ are crucially important in *complex analysis*. In fact, if $f\colon U \to \mathbb{C}$ is conformal, then (under some mild assumptions on the partial derivatives of f) f is *holomorphic*, i.e. the complex derivative

$$f'(z) := \lim_{w \to z} \frac{f(w) - f(z)}{w - z}$$

exists for any $z \in U$, and $f'(z) \neq 0$. The converse is also true, if f is holomorphic with nonvanishing derivatives, then it is angle-preserving. To see this (intuitively), near a point z we have

$$f(w) - f(z) \approx f'(z)(w - z).$$

Hence up to translations, the map near z is roughly given by multiplying the factor $f'(z) \neq 0$, which is an angle-preserving map.

Example. There is another important conformal map, the *inversion*:

$$T\colon \mathbb{C}\backslash\{0\} \to \mathbb{C}\backslash\{0\}; \quad z \mapsto \frac{1}{z}.$$

It is not hard to verify that the inversion is conformal by direct calculations; we will provide a more geometric proof later.

Remark 6.2. It is very useful to *compactify* the complex plane by adding an extra element ∞ to form the extended complex plane

$$\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}.$$

There is a natural *topology* on $\hat{\mathbb{C}}$, where subsets of the form $\{z \in \mathbb{C} \mid |z| > R\} \cup \{\infty\}$ are open neighborhood of ∞ for any $R > 0$. It is easy to see that one can extend both affine transformations and the inversion to bijective continuous maps $\hat{\mathbb{C}} \to \hat{\mathbb{C}}$: The affine transformation $T\colon z \mapsto Az + B$ can be extended continuously to a map $\hat{\mathbb{C}} \to \hat{\mathbb{C}}$ by setting $T(\infty) = \infty$; the inversion $T\colon z \mapsto 1/z$ can be extended continuously to a map $\hat{\mathbb{C}} \to \hat{\mathbb{C}}$ by setting $T(0) = \infty$ and $T(\infty) = 0$.

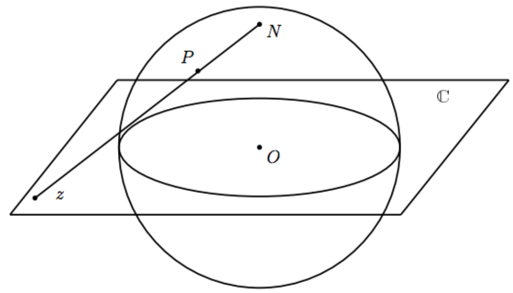
It seems that the point at infinity is quite different from the other finite points in $\mathbb{C}$, but Riemann showed that this is not the case. He did this by representing all of the points of $\hat{\mathbb{C}}$ by points of the unit sphere $S^2 \subseteq \mathbb{R}^3$. This sphere is also called the *Riemann sphere.*
Write

$$S^2 = \{(z, t) \in \mathbb{C} \times \mathbb{R} \mid |z|^2 + t^2 = 1\}.$$

The north pole of this sphere will be denoted by $N = (0, 1)$.

Definition 6.3. Define the *stereographic projection*

$$\mathrm{SP}\colon S^2\backslash\{N\} \to \mathbb{C}$$



by sending a point P on the sphere to the intersection of the line connecting N and P with the complex plane $\{(z,t) \in \mathbb{C} \times \mathbb{R} \mid t = 0\}$. Observe that as P approaches the north pole, its image $\mathrm{SP}(P)$ would tends to ∞ (i.e. $|\mathrm{SP}(P)| \to \infty$). Hence it makes sense to extend the stereographic projection to the whole sphere, and define

$$\mathrm{SP}\colon S^2 \to \hat{\mathbb{C}},$$

where $\mathrm{SP}(N) \coloneqq \infty$. The map SP is a *homeomorphism*, which means that SP is invertible and both SP and SP^{-1} are continuous maps.

It is easy to give a formula for the stereographic projection and its inverse:

$$\mathrm{SP}(z,t) = \frac{z}{1-t}, \ \mathrm{SP}^{-1}(w) = \left(\frac{2w}{1+|w|^2}, \frac{-1+|w|^2}{1+|w|^2}\right).$$

Note that as $t \to 1$ we have $\mathrm{SP}(z,t) \to \infty$, and as $w \to \infty$ we have $\mathrm{SP}^{-1}(w) \to (0,1)$. Hence the above formula makes sense on the whole sphere S^2 and on the whole extended complex plane $\hat{\mathbb{C}}$.

Remark 6.4. An important result proved by Gauss in 1827, called the Gauss' *Theorema Egregium* (Latin for "remarkable theorem" or "totally awesome theorem"), states that the *Gaussian curvature* is an *intrinsic invariant* of a surface. In particular, it would imply that there is no map from a region of the sphere (which is of constant curvature 1) onto a plane (which is of constant curvature 0) that preserves both distances and angles. In particular, it is not possible for the stereographic projection SP to preserve both distances and angles. However, one can show by direct computations that SP does preserve angles, i.e. it is a conformal map.

Proposition 6.5. $\mathrm{SP}\colon S^2 \to \hat{\mathbb{C}}$ *takes circles in* S^2 *to circles in* $\hat{\mathbb{C}}$. *Note that circles in* $\mathbb{C}$ *and straight lines in* $\mathbb{C}$ *are both considered as circles in* $\hat{\mathbb{C}}$.

Proof. Let C be a circle in S^2, which is the intersection of a plane P with S^2. First, suppose C passes through the north pole N. Then one can check that $\mathrm{SP}(C)$ would be the intersection of the plane P with the complex plane $\mathbb{C}$, which is a line in $\mathbb{C}$.

Second, suppose the circle does not pass through the north pole N. Let $Ax_1 + Bx_2 + Cx_3 + D = 0$ be the defining equation of P. The fact that the circle does not pass through N implies that $C+D \neq 0$. A point $z = x+iy \in \mathbb{C}$ is in the image the circle $\mathrm{SP}(C)$ if and only if $\mathrm{SP}^{-1}(z)$ lies in the plane P, i.e.

$$\frac{2x}{x^2+y^2+1}A + \frac{2y}{x^2+y^2+1}B + \frac{x^2+y^2-1}{x^2+y^2+1}C + D = 0,$$

or equivalently

$$x^2 + y^2 + \frac{2A}{C+D}x + \frac{2B}{C+D}y + \frac{-C+D}{C+D} = 0.$$

The set of such points forms a circle in $\mathbb{C}$. $\square$

Remark 6.6. One can use the stereographic projection to give a more geometric proof of the fact that the inversion $z \mapsto 1/z$ is conformal. First, observe that the points

$$\mathrm{SP}^{-1}(w) = \left(\frac{2w}{1+|w|^2}, \frac{-1+|w|^2}{1+|w|^2}\right) \text{ and } \mathrm{SP}^{-1}(\bar{w}) = \left(\frac{2\bar{w}}{1+|w|^2}, \frac{-1+|w|^2}{1+|w|^2}\right)$$

are related by a reflection in $\mathbb{R}^3$, say denoted by R_1. Then the composition $\mathrm{SP} \circ R_1 \circ \mathrm{SP}^{-1}$ corresponds to complex conjugation. Second, the points

$$\mathrm{SP}^{-1}(w) = \left(\frac{2w}{1+|w|^2}, \frac{-1+|w|^2}{1+|w|^2}\right) \text{ and } \mathrm{SP}^{-1}\left(\frac{w}{|w|^2}\right) = \left(\frac{2w}{1+|w|^2}, \frac{1-|w|^2}{1+|w|^2}\right)$$

are also related by a reflection in $\mathbb{R}^3$, say denoted by R_2. Then the composition $\mathrm{SP} \circ R_2 \circ \mathrm{SP}^{-1}$ would send w to $\frac{w}{|w|^2}$. Since the complex conjugate of $\frac{w}{|w|^2}$ is precisely $\frac{1}{w}$, we have

$$T = \mathrm{SP} \circ R_1 \circ R_2 \circ \mathrm{SP}^{-1}.$$

Since SP, SP^{-1} are conformal, and reflections are clearly conformal, therefore the inversion T is also conformal.

6.2. **Möbius transformations.** Now we consider a simultaneous generalization of affine transformations and the inversion, the *Möbius transformations*.

Definition 6.7. Let $a, b, c, d \in \mathbb{C}$ be complex numbers satisfying $ad - bc \neq 0$. Then we can define a *Möbius transformation* $T \colon \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ by

$$T(z) = \frac{az+b}{cz+d} \ \text{ if } z \neq -\frac{d}{c}, \infty$$

and define $T(\infty) = \frac{a}{c}$, $T(-\frac{d}{c}) = \infty$. The set of all Möbius transformations will be denoted by $\text{Möb}(\hat{\mathbb{C}})$.

Proposition 6.8. *Any Möbius transformation is a bijective continuous map.*

Proof. Observe that any Möbius transformation is a composition of affine transformations and inversions:

$$\frac{az+b}{cz+d} = \frac{\frac{a}{c}(cz+d) - \frac{ad}{c} + b}{cz+d} = \frac{a}{c} + \frac{b - \frac{ad}{c}}{cz+d}.$$

It is not hard to check that each affine transformation and inversion is bijective and continuous. This proves the proposition. □

Exercise. Show that affine maps and the inversion both takes circles in $\hat{\mathbb{C}}$ to circles in $\hat{\mathbb{C}}$. Hence any Möbius transformation also has the same property.

Exercise. Show that the set of all Möbius transformations $\text{Möb}(\hat{\mathbb{C}})$ form a group, with binary operation given by composition. In other words, show that if $T_1, T_2 \colon \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ are Möbius transformations, then so are $T_1 \circ T_2$ and T_1^{-1}.

Moreover, show that the map

$$\rho \colon \text{GL}(2, \mathbb{C}) \to \text{Möb}(\hat{\mathbb{C}});\ g = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mapsto \rho_g \text{ where } \rho_g(z) = \frac{az+b}{cz+d}$$

is a surjective group homomorphism, with kernel given by $\left\{ \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} : \lambda \in \mathbb{C}^* \right\}$.

Therefore, the group of Möbius transformations is isomorphic to the *projective linear group* (or *projective general linear group*)

$$\text{PGL}(2, \mathbb{C}) \coloneqq \text{GL}(2, \mathbb{C})/\mathbb{C}^*.$$

The next theorem relates the rotation group $\text{SO}(3, \mathbb{R})$ we discussed before with the Möbius group.

Theorem 6.9. *Let $A \in \mathrm{SO}(3,\mathbb{R})$ be a rotation in $\mathbb{R}^3$. Then*

$$T_A \coloneqq \mathrm{SP} \circ A \circ \mathrm{SP}^{-1} \colon \hat{\mathbb{C}} \to \hat{\mathbb{C}}$$

is a Möbius transformation.

Proof. Consider the standard basis $\vec{e}_1 = (1,0,0)$, $\vec{e}_2 = (0,1,0)$, $\vec{e}_3 = (0,0,1)$ of $\mathbb{R}^3$, and denote $R_i(\theta)$ the counterclockwise rotation of angle θ with respect to the $\vec{e}_i$-axis. It is an exercise to show that any rotation $A \in \mathrm{SO}(3,\mathbb{R})$ can be written as a composition $R_1(\theta_1)R_2(\theta_2)R_3(\theta_3)$. Therefore, it suffices to prove the theorem for $R_i(\theta)$ for each $i = 1,2,3$.

First, let us consider $R_3(\theta)$. Since $\vec{e}_3$ passes through the north pole N of the Riemann sphere, it is easy to check that $\mathrm{SP} \circ R_3(\theta) \circ \mathrm{SP}^{-1}$ is the rotation of angle θ centered at $0 \in \mathbb{C}$, which can be realized as the Möbius transformation associated to

$$\begin{bmatrix} e^{i\theta} & 0 \\ 0 & 1, \end{bmatrix} \text{ or equivalently } \begin{bmatrix} e^{i\theta/2} & 0 \\ 0 & e^{-i\theta/2}. \end{bmatrix}$$

Second, let us consider $R_1(\theta)$ (the case of $R_2(\theta)$ can be proved similarly). Observe that $R_1(\theta) = R_2(\frac{\pi}{2})R_3(\theta)R_2(\frac{\pi}{2})^{-1}$ since

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Hence, it suffices to show that $\mathrm{SP} \circ R_2(\frac{\pi}{2}) \circ \mathrm{SP}^{-1}$ is a Möbius transformation. Observe that

$$R_2(\frac{\pi}{2}) \colon (x_1, x_2, x_3) \mapsto (x_3, x_2, -x_1).$$

Hence

$$\begin{aligned}
\mathrm{SP} \circ R_2(\frac{\pi}{2}) \circ \mathrm{SP}^{-1}(z) &= \mathrm{SP} \circ R_2(\frac{\pi}{2}) \left(\frac{2\mathrm{Re}(z)}{1+|z|^2}, \frac{2\mathrm{Im}(z)}{1+|z|^2}, \frac{-1+|z|^2}{1+|z|^2} \right) \\
&= \mathrm{SP}\left(\frac{-1+|z|^2}{1+|z|^2}, \frac{2\mathrm{Im}(z)}{1+|z|^2}, -\frac{2\mathrm{Re}(z)}{1+|z|^2} \right) \\
&= \frac{|z|^2 - 1 + 2i\mathrm{Im}(z)}{|z|^2 + 2\mathrm{Re}(z) + 1} \\
&= \frac{(z-1)(\bar{z}+1)}{(z+1)(\bar{z}+1)} = \frac{z-1}{z+1}
\end{aligned}$$

is a Möbius transformation. □

Theorem 6.10. *Let $g \in \mathrm{GL}(2, \mathbb{C})$. Then $\mathrm{SP}^{-1} \circ \rho_g \circ \mathrm{SP}\colon S^2 \to S^2$ is a rotation if and only if g can be written as*

$$g = \begin{bmatrix} \alpha & -\bar{\beta} \\ \beta & \bar{\alpha} \end{bmatrix} \text{ for some } |\alpha|^2 + |\beta|^2 = 1.$$

Remark 6.11. The subset of matrices in $\mathrm{GL}(2, \mathbb{C})$ that can be written as

$$g = \begin{bmatrix} \alpha & -\bar{\beta} \\ \beta & \bar{\alpha} \end{bmatrix} \text{ for some } |\alpha|^2 + |\beta|^2 = 1$$

actually forms a subgroup of $\mathrm{GL}(2, \mathbb{C})$, which is called the *special unitary group* and is denoted by $\mathrm{SU}(2)$. The theorems would imply that there is an isomorphism

$$\mathrm{SO}(3, \mathbb{R}) \cong \mathrm{SU}(2)/\{\pm I\}.$$

We defer the discussions on unitary groups to the next subsection.

Proof. Suppose $\mathrm{SP}^{-1} \circ \rho_g \circ \mathrm{SP}\colon S^2 \to S^2$ is a rotation. Let $g = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. One may assume that $\det(g) = 1$. Let p and $-p$ be any two antipodal points on S^2. Then we should have

$$\mathrm{SP}^{-1} \circ \rho_g \circ \mathrm{SP}(p) = -\mathrm{SP}^{-1} \circ \rho_g \circ \mathrm{SP}(-p).$$

Observe that if one writes $z = \mathrm{SP}(p)$, then $\mathrm{SP}(-p) = \frac{-1}{\bar{z}}$. Hence

$$\begin{aligned} \rho_g(z) &= \rho_g \circ \mathrm{SP}(p) \\ &= \mathrm{SP}(-\mathrm{SP}^{-1} \circ \rho_g \circ \mathrm{SP}(-p)) \\ &= \mathrm{SP}(-\mathrm{SP}^{-1}(\rho_g(\frac{-1}{\bar{z}}))) \\ &= \frac{-1}{\overline{\rho_g(\frac{-1}{\bar{z}})}}. \end{aligned}$$

Therefore, for any $z \in \mathbb{C}$ we have

$$\frac{b\bar{z} - a}{d\bar{z} - c} = \rho_g(\frac{-1}{\bar{z}}) = \frac{-1}{\overline{\rho_g(z)}} = -\frac{\bar{c}\bar{z} + \bar{d}}{\bar{a}\bar{z} + \bar{b}}.$$

This shows that the matrices

$$\begin{bmatrix} b & -a \\ d & -c \end{bmatrix} \text{ and } \begin{bmatrix} -\bar{c} & -\bar{d} \\ \bar{a} & \bar{b} \end{bmatrix}$$

give rise to the same Möbius transformation. Any two such matrices can only be differed by a scalar multiplication; and since they both have determinant one, we have

$$\begin{bmatrix} b & -a \\ d & -c \end{bmatrix} = \pm \begin{bmatrix} -\bar{c} & -\bar{d} \\ \bar{a} & \bar{b}. \end{bmatrix}$$

If one takes the minus sign, then $c = \bar{b}$ and $d = -\bar{a}$, thus we have $\det(g) = -|a|^2 - |b|^2 \neq 1$, contradiction. So only the positive sign in the above equation can hold, hence we have $c = -\bar{b}$, $d = \bar{a}$, and $\det(g) = |a|^2 + |b|^2 = 1$.

Conversely, suppose $g = \begin{bmatrix} \alpha & -\bar{\beta} \\ \beta & \bar{\alpha} \end{bmatrix}$ for some $|\alpha|^2 + |\beta|^2 = 1$. We would like to show that $\mathrm{SP}^{-1} \circ \rho_g \circ \mathrm{SP} \colon S^2 \to S^2$ is a rotation. Denote $z = \rho_g(0)$. Choose a rotation R such that

$$R(\mathrm{SP}^{-1}(z)) = \mathrm{SP}^{-1}(0).$$

By what we proved before, there exists $h \in \mathrm{SU}(2)$ such that $R = \mathrm{SP}^{-1} \circ \rho_h \circ \mathrm{SP}$. We have

$$\begin{aligned} \rho_{g^{-1}h^{-1}}(0) &= \rho_{g^{-1}}\rho_{h^{-1}}(0) = \rho_{g^{-1}}\rho_{h^{-1}}\mathrm{SP}(\mathrm{SP}^{-1}(0)) \\ &= \rho_{g^{-1}}\rho_{h^{-1}}\mathrm{SP} \circ R \circ \mathrm{SP}^{-1}(z) \\ &= \rho_{g^{-1}}(z) = 0. \end{aligned}$$

Hence $\rho_{hg}(0) = 0$. Since $hg \in \mathrm{SU}(2)$, we have

$$hg = \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix} \text{ for some } \theta \in \mathbb{R}.$$

Then $\mathrm{SP}^{-1} \circ \rho_{hg} \circ \mathrm{SP} \colon S^2 \to S^2$ is a rotation along the x_3-axis of angle 2θ. Therefore

$$\mathrm{SP}^{-1} \circ \rho_g \circ \mathrm{SP} = R(\mathrm{SP}^{-1} \circ \rho_{hg} \circ \mathrm{SP})$$

is a rotation. □

Remark 6.12. SU(2) is *diffeomorphic* to the three-sphere S^3, which therefore endows S^3 with the structure of a *Lie group*. It also plays an important role in the study of *quaternions*, since SU(2) is isomorphic to the group of *unit quaternions*.

6.3. **Hermitian inner product and unitary matrices.**

Definition 6.13. Let $\vec{v}$ and $\vec{w}$ be two vectors in $\mathbb{C}^n$. The (standard) *Hermitian inner product* on $\mathbb{C}^n$ between them is defined to be

$$\langle \vec{v}, \vec{w} \rangle = \sum_{i=1}^{n} v_i \bar{w}_i.$$

It satisfies the following properties.

- $\langle \vec{v}, \vec{w} \rangle = \overline{\langle \vec{w}, \vec{v} \rangle}$.
- $\langle \lambda_1 \vec{v}_1 + \lambda_2 \vec{v}_2, \vec{w} \rangle = \lambda_1 \langle \vec{v}_1, \vec{w} \rangle + \lambda_2 \langle \vec{v}_2, \vec{w} \rangle$.
- $\langle \vec{v}, \lambda_1 \vec{w}_1 + \lambda_2 \vec{w}_2 \rangle = \overline{\lambda_1} \langle \vec{v}, \vec{w}_1 \rangle + \overline{\lambda_2} \langle \vec{v}, \vec{w}_2 \rangle$.
- $\langle \vec{v}, \vec{v} \rangle \geq 0$, where the equality holds if and only if $\vec{v} = \vec{0}$.

Definition 6.14. A linear transformation $T \colon \mathbb{C}^n \to \mathbb{C}^n$ is called a *unitary transformation* if

$$\langle T\vec{v}, T\vec{w} \rangle = \langle \vec{v}, \vec{w} \rangle \text{ holds for any } \vec{v}, \vec{w} \in \mathbb{C}^n.$$

Any linear transformation $T \colon \mathbb{C}^n \to \mathbb{C}^n$ can be represented by a (complex) $n \times n$ matrix. It is not hard to show that T is unitary if and only if its associated matrix A satisfies $\overline{A^T} A = I$. Matrices satisfying this property are called *unitary matrices*.

$$\mathrm{U}(n) = \{A \in M_n(\mathbb{C}) \mid \overline{A^T} A = I_n\}.$$

One can easily see that the determinant of an unitary matrix has absolute value 1. Those of which with determinant one are called *special unitary matrices*.

$$\mathrm{SU}(n) = \{A \in M_n(\mathbb{C}) \mid \overline{A^T} A = I_n,\ \det(A) = 1\}.$$

Example. Elements of $\mathrm{U}(1) = \{z \in \mathbb{C} \mid |z| = 1\}$ correspond to points on the unit circle in $\mathbb{C}$. There is an isomorphism $\mathrm{U}(1) \cong \mathrm{SO}(2, \mathbb{R})$ given by

$$e^{i\theta} \mapsto \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Exercise. Show that

$$\mathrm{SU}(2) = \left\{ \begin{bmatrix} \alpha & -\bar{\beta} \\ \beta & \bar{\alpha} \end{bmatrix} \;\middle|\; \alpha, \beta \in \mathbb{C},\ |\alpha|^2 + |\beta|^2 = 1 \right\}.$$

6.4. **Cross ratios.**

Question 6.15. *Given two circles in $\hat{\mathbb{C}}$, can one find a Möbius transformation which takes one circle to the other?*

The answer is yes. In fact, we can prove the following stronger statement.

Theorem 6.16. *Let z_1, z_2, z_3 be three distinct points of $\hat{\mathbb{C}}$, and w_1, w_2, w_3 be another three distinct points of $\hat{\mathbb{C}}$ (which may or may not coincides with z_i's). There exists a unique Möbius transformation ρ such that $\rho(z_i) = w_i$ for $i =$* $1, 2, 3$. *In particular, a Möbius transformation is uniquely determined by its images of three distinct points.*

Proof. Observe that the Möbius transformation

$$f(z_1, z_2, z_3) \coloneqq \frac{(z - z_1)(z_2 - z_3)}{(z - z_3)(z_2 - z_1)}$$

maps $z_1 \mapsto 0$, $z_2 \mapsto 1$, and $z_3 \mapsto \infty$. Then the composition

$$\rho \coloneqq f(w_1, w_2, w_3)^{-1} \circ f(z_1, z_2, z_3),$$

which is also a Möbius transformation, would take $z_i \mapsto w_i$ as desired.

Assume that both ρ_1 and ρ_2 have the desired property. Then $\rho_2^{-1} \circ \rho_1$ is a Möbius transformation with three distinct fixed points: $\rho_2^{-1} \circ \rho_1(z_i) = z_i$. It is an easy exercise to show that any non-identity Möbius transformation has at most two distinct fixed points. Therefore, we have $\rho_2^{-1} \circ \rho_1 = 1$ and thus $\rho_1 = \rho_2$. $\square$

Definition 6.17. Let z_1, z_2, z_3, z_4 be four distinct points in $\hat{\mathbb{C}}$. Their *cross ratio* is defined to be

$$[z_1, z_2, z_3, z_4] \coloneqq \frac{(z_1 - z_3)(z_2 - z_4)}{(z_1 - z_4)(z_2 - z_3)}.$$

Lecture 11

Theorem 6.18. *Let z_1, z_2, z_3, z_4 be four distinct points of $\hat{\mathbb{C}}$, and ρ be a Möbius transformation. Then*

$$[\rho(z_1), \rho(z_2), \rho(z_3), \rho(z_4)] = [z_1, z_2, z_3, z_4].$$

Moreover, let w_1, w_2, w_3, w_4 be four distinct point of $\hat{\mathbb{C}}$. Then $[w_1.w_2.w_3, w_4] =$ $[z_1, z_2, z_3, z_4]$ if and only if there exists a Möbius transformation such that $\rho(z_i) = w_i$ for each i.

Proof. Let $\rho(z) = \frac{az+b}{cz+d}$ be a Möbius transformation with $ad - bc = 1$. One can easily check that

$$\rho(z_i) - \rho(z_j) = \frac{z_i - z_j}{(cz_i + d)(cz_j + d)}$$

and therefore $[\rho(z_1), \rho(z_2), \rho(z_3), \rho(z_4)] = [z_1, z_2, z_3, z_4]$.

Now we prove the second statement. By the previous theorem, there exists a unique Möbius transformation ρ such that $\rho(z_i) = w_i$ for each $i = 1, 2, 3$. Moreover, we know that

$$\rho = [z, w_1, w_2, w_3]^{-1} \circ [z, z_1, z_2, z_3],$$

where

$$[z, z_1, z_2, z_3] = \frac{(z - z_1)(z_2 - z_3)}{(z - z_3)(z_1 - z_2)} \text{ and } [z, w_1, w_2, w_3] = \frac{(z - w_1)(w_2 - w_3)}{(z - w_3)(w_1 - w_2)}.$$

The Möbius transformation also satisfies $\rho(z_4) = w_4$ if and only if the cross ratios coincide $[w_1.w_2.w_3, w_4] = [z_1, z_2, z_3, z_4]$. □

Remark 6.19. Consider the set of distinct quadruples in $\mathbb{C}$:

$$X = \{(z_1, z_2, z_3, z_4) \in \mathbb{C}^4 \mid z_i \neq z_j \text{ for all } i \neq j\}.$$

The group of Möbius transformations $\text{Möb}(\hat{\mathbb{C}})$ acts naturally on X:

$$\rho \cdot (z_1, z_2, z_3, z_4) = (\rho(z_1), \rho(z_2), \rho(z_3), \rho(z_4)).$$

The cross ratio defines a map

$$[-]\colon X \to \mathbb{C};\ (z_1, z_2, z_3, z_4) \mapsto [z_1, z_2, z_3, z_4].$$

The theorem we proved implies that the cross ratio map $[-]$ descends to an injective map

$$[-]\colon X/\text{Möb}(\hat{\mathbb{C}}) \to \mathbb{C}.$$

Exercise. Show that the image of the cross ratio map is $\mathbb{C}\backslash\{0, 1\}$. Therefore, the cross ratio map gives a one-to-one correspondence between the set of orbits $X/\text{Möb}(\hat{\mathbb{C}})$ and $\mathbb{C}\backslash\{0, 1\}$.

Remark 6.20. The orbit space $X/\text{Möb}(\hat{\mathbb{C}})$ is called the *moduli space* of four points on the Riemann sphere, and is also denoted by $\mathcal{M}_{0,4}$ in algebraic geometry. Its generalizations $\mathcal{M}_{0,n}$, $\mathcal{M}_{g,n}$, $\mathcal{M}_{g,n}(X)$, etc. play important roles in algebraic geometry, symplectic geometry and mathematical physics among others.

6.5. **Conjugacy classes of Möbius transformations.** We show in this subsection that tr^2 is a complete *invariant* of the conjugacy classes of $\text{Möb}(\hat{\mathbb{C}})$.

Definition 6.21. Let $\rho \in \text{Möb}(\hat{\mathbb{C}})$ be a Möbius transformation, where $\rho(z) = \frac{az+b}{cz+d}$ and $ad - bc = 1$. Define a map

$$\mathrm{tr}^2 \colon \text{Möb}(\hat{\mathbb{C}}) \to \mathbb{C};\ \rho \mapsto \left(\mathrm{tr}\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right)^2 = (a+d)^2.$$

Remark 6.22. Note that for any $\rho \in \text{Möb}(\hat{\mathbb{C}})$, there are two ways to represent it by an element of $\mathrm{SL}(2,\mathbb{C})$, namely $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $\begin{bmatrix} -a & -b \\ -c & -d \end{bmatrix}$. Therefore tr is not a well-defined function on $\text{Möb}(\hat{\mathbb{C}})$, but tr^2 is.

Exercise. Suppose $\rho_1, \rho_2 \in \text{Möb}(\hat{\mathbb{C}})$ are conjugate with each other, i.e. $\rho_2 = \rho\rho_1\rho^{-1}$ for some $\rho \in \text{Möb}(\hat{\mathbb{C}})$. Show that $\mathrm{tr}^2(\rho_1) = \mathrm{tr}^2(\rho_2)$.

Exercise. The translation $T_1 \colon z \mapsto z+1$ has $\mathrm{tr}^2(T_1) = 4$. The scaling action $M_\lambda \colon z \mapsto \lambda z$ $(\lambda \neq 0, 1)$ has $\mathrm{tr}^2(M_\lambda) = \lambda + \lambda^{-1} + 2$.

Theorem 6.23. *Let $\rho \in \text{Möb}(\hat{\mathbb{C}})$ be a non-identity Möbius transformation. Then ρ has either one or two fixed points. Moreover,*

- *if ρ has one fixed point, then it is conjugate to the translation T_1, and therefore has $\mathrm{tr}^2(\rho) = 4$;*
- *if ρ has two fixed points, then it is conjugate to the scaling M_λ for some $\lambda \neq 0, 1$, and therefore has $\mathrm{tr}^2(\rho) = \lambda + \lambda^{-1} + 2$.*

Proof. Suppose ρ has a unique fixed point z_0. Choose $\eta \in \text{Möb}(\hat{\mathbb{C}})$ so that $\eta(z_0) = \infty$. Then $\eta\rho\eta^{-1}$ has a unique fixed point at ∞. This would imply that $\eta\rho\eta^{-1} \colon z \mapsto z + b$ for some $b \neq 0$. Then $M_b^{-1}\eta\rho\eta^{-1}M_b = T_1$.

Suppose ρ has two fixed points z_1 and z_2. Choose $\eta \in \text{Möb}(\hat{\mathbb{C}})$ so that $\eta(z_1) = 0$ and $\eta(z_2) = \infty$. Then $\eta\rho\eta^{-1}$ fixes 0 and ∞. This would imply that $\eta\rho\eta^{-1} = M_\lambda$ for some $\lambda \neq 0, 1$. □

Theorem 6.24. *$\rho_1, \rho_2 \in \text{Möb}(\hat{\mathbb{C}})$ are conjugate to each other if and only if $\mathrm{tr}^2(\rho_1) = \mathrm{tr}^2(\rho_2)$.*

Proof. Suppose $\mathrm{tr}^2(\rho_1) = \mathrm{tr}^2(\rho_2)$. We would like to show that ρ_1 and ρ_2 are in the same conjugacy class. First, suppose $\mathrm{tr}^2(\rho_1) = \mathrm{tr}^2(\rho_2) = 4$. By the previous theorem, both ρ_1 and ρ_2 are conjugate to T_1. Second, suppose $\mathrm{tr}^2(\rho_1) = \mathrm{tr}^2(\rho_2) \neq 4$. Then there exists $\lambda \neq 0, 1$ such that ρ_1 and ρ_2 are conjugate to either M_λ or $M_{1/\lambda}$. One concludes the proof by observing that M_λ and $M_{1/\lambda}$ are conjugate to each other: $M_\lambda = T \circ M_{1/\lambda} \circ T^{-1}$ where T is the inversion. $\square$

6.6. **Geometric classification of conjugacy classes.** In this subsection, we examine the geometric behaviors of Möbius transformations under large iterations.

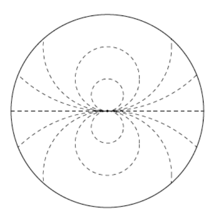
Suppose $\rho \in \mathrm{Möb}(\hat{\mathbb{C}})$ has $\mathrm{tr}^2(\rho) = 4$, or equivalently, suppose it is conjugate to the translation T_1, say $\rho = \eta T_1 \eta^{-1}$. The translation T_1 has a unique fixed point $\infty \in \hat{\mathbb{C}}$. Moreover, for any non-fixed point $z \in \mathbb{C}$, we have

$$\lim_{n\to\infty} T_1^n z = \lim_{n\to\infty} z + n = \infty.$$

In other words, by applying T_1 repeatedly, all points in $\mathbb{C}$ are moved towards the fixed point ∞. Let $\eta(\infty) = z_0$. Then z_0 is the fixed point of ρ, and for any $z \in \hat{\mathbb{C}}$ we have

$$\lim_{n\to\infty} \rho^n z = \lim_{n\to\infty} \eta T_1^n(\eta^{-1} z) = \eta\left(\lim_{n\to\infty} T_1^n(\eta^{-1} z)\right) = \eta(\infty) = z_0.$$

Hence by applying ρ repeatedly, all points are moved towards the fixed point of ρ. Such ρ is called *parabolic*.



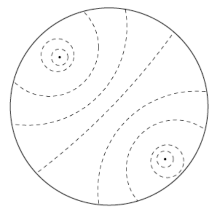
Parabolic

Suppose ρ has two fixed points, or equivalently, is conjugate to M_λ for some $\lambda \neq 0, 1$, say $\rho = \eta M_\lambda \eta^{-1}$. Then the fixed points of ρ are $z_1 = \eta(0)$ and $z_2 = \eta(\infty)$.

If $|\lambda| = 1$, then $\lambda = e^{i\theta}$ for some $\theta \notin 2\pi\mathbb{Z}$. For any $z \in \mathbb{C}\backslash\{0\}$, the limit

$$\lim_{n\to\infty} M_\lambda^n z$$

does not exist, hence neither does the limit $\lim_{n\to\infty} \rho^n z$ for any $z \in \hat{\mathbb{C}}\backslash\{z_1, z_2\}$. Such ρ is called *elliptic*. In this case we have $\operatorname{tr}^2(\rho) = e^{i\theta} + e^{-i\theta} + 2 \in [0, 4)$.



Elliptic

If $|\lambda| < 1$, then for any $z \in \mathbb{C}$, we have

$$\lim_{n\to\infty} M_\lambda^n z = 0.$$

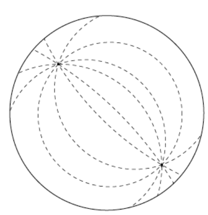
Hence for any $z \in \hat{\mathbb{C}}\backslash\{z_2\}$ we have

$$\lim_{n\to\infty} \rho^n z = \lim_{n\to\infty} \eta M_\lambda^n(\eta^{-1} z) = \eta \left(\lim_{n\to\infty} M_\lambda^n(\eta^{-1} z) \right) = \eta(0) = z_1.$$

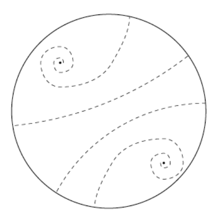
In other words, ρ progressively moves any non-fixed point away from one of the fixed points (z_2) and towards the other one (z_1). Similarly, if $|\lambda| > 1$, then for any $z \in \hat{\mathbb{C}}\backslash\{0\}$ we have

$$\lim_{n\to\infty} M_\lambda^n z = \infty.$$

By the same argument, one can see that ρ progressively moves any non-fixed point away from z_1 and towards z_2. If $\lambda \neq 1$ is a positive real number, then such a map is called *hyperbolic*. Otherwise, it is called *loxodromic*.



Hyperbolic



Loxodromic

Exercise. Let $\rho \in \text{Möb}(\hat{\mathbb{C}})$ be a non-identity Möbius transformation.

- ρ is parabolic if and only if $\operatorname{tr}^2(\rho) = 4$.
- ρ is elliptic if and only if $0 \leq \operatorname{tr}^2(\rho) < 4$.
- ρ is hyperbolic if and only if $\operatorname{tr}^2(\rho) > 4$.
- ρ is loxodromic if and only if $\operatorname{tr}^2(\rho) < 0$ or $\operatorname{tr}^2(\rho) \notin \mathbb{R}$.

6.7. **Finite subgroups of** $\text{Möb}(\hat{\mathbb{C}})$**.**

Theorem 6.25. *Suppose* $\rho \in \text{Möb}(\hat{\mathbb{C}})$ *is of finite order. Then* ρ *is either elliptic or the identity.*

Proof. Suppose ρ is not the identity. then ρ is conjugate to either T_1 or M_λ for some $\lambda \neq 0, 1$. Since T_1 is of infinite order, ρ cannot be conjugate to T_1. Suppose M_λ is of finite order for some $\lambda \in \mathbb{C}$. Then $\lambda^n = 1$ for some positive integer n, hence we must have $|\lambda| = 1$. Therefore, if $\rho \in \text{Möb}(\hat{\mathbb{C}})$ is of finite order, then it is conjugate to M_λ for some $|\lambda| = 1$, hence ρ is elliptic. $\square$

Theorem 6.26. *Let* $G \subseteq \text{Möb}(\hat{\mathbb{C}}) \cong \text{PGL}(2, \mathbb{C})$ *be a finite subgroup. Then* G *is conjugate to a subgroup of* $\text{PSU}(2)$*. In other words, there exists* $\rho \in \text{PGL}(2, \mathbb{C})$ *such that* $\rho G \rho^{-1} \subseteq \text{PSU}(2) \subseteq \text{PGL}(2, \mathbb{C})$*.*

Proof. Let $g \in G$ be a non-identity element of G. Then g is elliptic, i.e. there exists $\rho \in \text{PGL}(2, \mathbb{C})$ such that $\rho g \rho^{-1} = \begin{bmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{bmatrix}$ for some $|\lambda| = 1$ and $\lambda \neq \pm 1$. By replacing G with $\rho G \rho^{-1}$, we may assume that $g = \begin{bmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{bmatrix}$.

Let $h = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be any element of $G \backslash \{e, g\}$. Since both h and gh are elliptic, we have $a + d$ and $\lambda a + \bar{\lambda} d$ are real numbers with absolute values strictly less than 2. Hence

$$\lambda a + \bar{\lambda} d = \bar{\lambda} \bar{a} + \lambda \bar{d}, \text{ thus } \lambda(a - \bar{d}) = \bar{\lambda}(\bar{a} - d).$$

Observe that $a - \bar{d} = a + d - (d + \bar{d})$ is a real number, hence

$$(\lambda - \bar{\lambda})(a - \bar{d}) = 0.$$

Since $|\lambda| = 1$ and $\lambda \neq \pm 1$, we have $\lambda - \bar{\lambda} \neq 0$. Thus we get $a = \bar{d}$, so any element of G is of the form $\begin{bmatrix} a & b \\ c & \bar{a} \end{bmatrix}$.

Let $h = \begin{bmatrix} a & b \\ c & \bar{a} \end{bmatrix}$ be a non-identity element of G. One can assume that $\det(h) = |a|^2 - bc = 1$. We claim that b and c must be both zero or both nonzero. Consider the element

$$ghg^{-1}h^{-1} = \begin{bmatrix} 1+(1-\lambda^2)bc & -ab(1-\lambda^2) \\ -\bar{a}c(1-\bar{\lambda}^2) & 1+(1-\bar{\lambda}^2)bc \end{bmatrix} \in G.$$

Suppose $b = 0$. Then $\operatorname{tr}(ghg^{-1}h^{-1}) = 4$, so $ghg^{-1}h^{-1}$ is the identity element since any non-identity element of G is elliptic. Hence $-\bar{a}c(1-\bar{\lambda}^2) = 0$, thus $c = 0$. Similarly, one can show that $c = 0$ would imply $b = 0$.

Suppose $b \neq 0$, then $c \neq 0$ and $c = r\bar{b}$ for some nonzero real number r. Consider two elements of G

$$h_1 = \begin{bmatrix} a_1 & b_1 \\ r_1\bar{b}_1 & \bar{a}_1 \end{bmatrix} \text{ and } h_2 = \begin{bmatrix} a_2 & b_2 \\ r_2\bar{b}_2 & \bar{a}_2 \end{bmatrix}$$

where $b_1b_2 \neq 0$ and $r_1, r_2 \in \mathbb{R}$. We claim that $r_1 = r_2$ for any two such elements. We have

$$h_1h_2 = \begin{bmatrix} a_1a_2 + r_2b_1\bar{b}_2 & \star \\ \star & \bar{a}_1\bar{a}_2 + r_1\bar{b}_1b_2 \end{bmatrix}.$$

The two diagonal entries are complex conjugate to each other, thus $r_1 = r_2$. This proves that there exists a nonzero real number r such that any element of G can be written as $\begin{bmatrix} a & b \\ r\bar{b} & \bar{a} \end{bmatrix}$.

We claim that $r < 0$. Assume the contrary that r is a positive real number. Since

$$\begin{bmatrix} r^{1/4} & 0 \\ 0 & r^{-1/4} \end{bmatrix} \begin{bmatrix} a & b \\ r\bar{b} & \bar{a} \end{bmatrix} \begin{bmatrix} r^{1/4} & 0 \\ 0 & r^{-1/4} \end{bmatrix}^{-1} = \begin{bmatrix} a & r^{1/2}b \\ r^{1/2}\bar{b} & \bar{a} \end{bmatrix},$$

up to conjugation, one can assume that any element of G is of the form $\begin{bmatrix} a & b \\ \bar{b} & \bar{a} \end{bmatrix}$.

Let $g = \begin{bmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{bmatrix}$ and $h = \begin{bmatrix} a & b \\ \bar{b} & \bar{a} \end{bmatrix}$. Consider the element

$$ghg^{-1}h^{-1} = \begin{bmatrix} 1+(1-\lambda^2)|b|^2 & -ab(1-\lambda^2) \\ -\bar{a}\bar{b}(1-\bar{\lambda}^2) & 1+(1-\bar{\lambda}^2)|b|^2 \end{bmatrix} \in G.$$

Then

$$\operatorname{tr}(ghg^{-1}h^{-1}) = 2 + (2-\lambda^2-\bar{\lambda}^2)|b|^2 > 2$$

since $\lambda \neq \pm 1$. Therefore $ghg^{-1}h^{-1} \in G$ is neither elliptic nor the identity element. Contradiction. This proves that the factor r is a negative real number.

Similarly, by considering

$$\begin{bmatrix} (-r)^{1/4} & 0 \\ 0 & (-r)^{-1/4} \end{bmatrix} \begin{bmatrix} a & b \\ r\bar{b} & \bar{a} \end{bmatrix} \begin{bmatrix} (-r)^{1/4} & 0 \\ 0 & (-r)^{-1/4} \end{bmatrix}^{-1} = \begin{bmatrix} a & (-r)^{1/2}b \\ -(-r)^{1/2}\bar{b} & \bar{a} \end{bmatrix},$$

one can assume that any element of G is of the form $\begin{bmatrix} a & b \\ -\bar{b} & \bar{a} \end{bmatrix} \in \mathrm{SU}(2)$. This proves the theorem. □

Corollary 6.27. *Let G be a finite subgroup of $\text{Möb}(\hat{\mathbb{C}})$. Then G is isomorphic to either a cyclic group, a dihedral group, or the symmetry group of a Platonic solid.*

Proof. By the previous theorem, G is isomorphic to a subgroup of $\mathrm{PSU}(2) \cong \mathrm{SO}(3,\mathbb{R})$. The corollary then follows from the classification of finite subgroups of $\mathrm{SO}(3,\mathbb{R})$. □

Lecture 12

6.8. **The upper half plane.** Let us consider the upper half plane

$$\mathbb{H} = \{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\}.$$

Theorem 6.28. *A Möbius transformation ρ maps $\mathbb{H}$ to itself if and only if $\rho(z) = \frac{az+b}{cz+d}$ for some $a, b, c, d \in \mathbb{R}$ and $ad - bc > 0$.*

Proof. Suppose $\rho(z) = \frac{az+b}{cz+d}$ with $a, b, c, d \in \mathbb{R}$ and $ad - bc > 0$. Then

$$\mathrm{Im}(\rho(z)) = \frac{(ad-bc)\mathrm{Im}(z)}{|cz+d|^2} > 0 \text{ if } \mathrm{Im}(z) > 0.$$

Conversely, suppose ρ maps $\mathbb{H}$ to $\mathbb{H}$. Then it also maps $\mathbb{R}_\infty = \mathbb{R} \cup \{\infty\}$ to itself. Let $r_1 = \rho^{-1}(0)$, $r_2 = \rho^{-1}(1)$, and $r_3 = \rho^{-1}(\infty)$. Then

$$\rho(z) = \frac{(\rho(z)-0)(1-\infty)}{(1-0)(\rho(z)-\infty)} = [\rho(z), 1, 0, \infty] = [z, r_2, r_1, r_3] = \frac{(z-r_1)(r_2-r_3)}{(z-r_3)(r_2-r_1)}.$$

Hence $\rho(z) = \frac{az+b}{cz+d}$ for some $a, b, c, d \in \mathbb{R}$. For any $z \in \mathbb{H}$ we have

$$\mathrm{Im}(\rho(z)) = \frac{(ad-bc)\mathrm{Im}(z)}{|cz+d|^2} > 0.$$

Therefore $ad - bc > 0$. □

Therefore, we have

$$\text{Möb}(\mathbb{H}) \cong \text{PSL}(2, \mathbb{R}) = \text{SL}(2, \mathbb{R})/\{\pm I\},$$

and we will call elements of Möb($\mathbb{H}$) *real* Möbius transformations.

Remark 6.29. Recall that Möb($\hat{\mathbb{C}}$) consists of all conformal maps (i.e. biholomorphic maps) $\hat{\mathbb{C}} \to \hat{\mathbb{C}}$. We are also interested in conformal maps on subsets of $\hat{\mathbb{C}}$. For instance, one can show that all biholomorphic maps of $\mathbb{C}$ are affine transformations $Az + B$. One can also show that all biholomorphic maps of $\mathbb{H}$ are Möbius transformations. In fact, there is an important theorem, the *Riemann mapping theorem*, which states that any open, connected, simply connected proper subset D of $\mathbb{C}$ is biholomorphic to $\mathbb{H}$. Therefore, the group of biholomorphic self-maps $D \to D$ of any such domain is isomorphic to $\text{PSL}(2, \mathbb{R})$. The proofs of these statements require some basic knowledge of *complex analysis*, which is beyond the scope of this course.

Remark 6.30. Any Riemann surface of genus $g \geq 2$ is biholomorphic to a *quotient* of the upper half plane $\mathbb{H}$ by a (discrete) subgroup of Möb($\mathbb{H}$). Therefore the upper half plane model is crucially important in various mathematical fields, including complex geometry, algebraic geometry, number theory, etc.

We now define a *metric* on $\mathbb{H}$ so that Möb($\mathbb{H}$) are isometries of $\mathbb{H}$ with respect to this metric. First, let us recall the definition of a metric.

Definition 6.31. A *metric* on a set X is a function

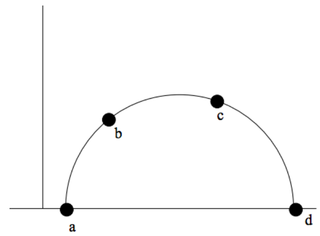
$$d\colon X \times X \to \mathbb{R}$$

satisfying:

- (positivity) $d(x, y) \geq 0$, with equality holds if and only if $x = y \in X$;
- (symmetric) $d(x, y) = d(y, x)$ for any $x, y \in X$.
- (triangle inequality) $d(x, z) \leq d(x, y) + d(y, z)$ for any $x, y, z \in X$.

Example. The standard distance function on $\mathbb{R}^n$ gives a metric.

We now attempt to define a metric on $\mathbb{H}$ so that the *real* Möbius transformations Möb($\mathbb{H}$) act on $\mathbb{H}$ as isometries. For any two points $b, c \in \mathbb{H}$, there is a unique circle in $\hat{\mathbb{C}}$ that passes through these two points and is perpendicular to the x-axis. Note that when b, c have the same x-coordinate, then the "circle in $\hat{\mathbb{C}}$" is actually the straight line passing through them (which is perpendicular



to the x-axis). Let $a, d \in \mathbb{R}_\infty = \mathbb{R} \cup \{\infty\}$ be the intersections of the circle with $\mathbb{R}_\infty$, where a is closer to b and d is closer to c ("closer" in the standard Euclidean distance). The *hyperbolic distance* between b and c is then defined to be

$$d_H(b, c) \coloneqq \log \frac{|a-c||b-d|}{|a-b||c-d|}.$$

It is not hard to see that $(\mathbb{H}, d_H)$ satisfies the positivity and the symmetric properties.

Proposition 6.32. *d_H is invariant under real Möbius transformations, i.e. for any $\rho \in \mathrm{M\ddot{o}b}(\mathbb{H})$ and any distinct two points b, c in $\mathbb{H}$, we have*

$$d_H(\rho(b), \rho(c)) = d_H(b, c)$$

Proof. Since a, b, c, d lies in the same circle, so are $\rho(a), \rho(b), \rho(c), \rho(d)$. Also, since $a, d \in \mathbb{R}_\infty$ and a real Möbius transformation sends $\mathbb{R}_\infty$ to itself, we have $\rho(a), \rho(d) \in \mathbb{R}_\infty$. Moreover, since Möbius transformations are angle preserving, the circle passing through $\rho(a), \rho(b), \rho(c), \rho(d)$ intersects perpendicularly with the x-axis at both intersection points $\rho(a)$ and $\rho(d)$. Therefore, the distance $d_H(\rho(b), \rho(c))$ is

$$\text{either } \log \frac{|\rho(a)-\rho(c)||\rho(b)-\rho(d)|}{|\rho(a)-\rho(b)||\rho(c)-\rho(d)|} \text{ or } \log \frac{|\rho(d)-\rho(c)||\rho(b)-\rho(a)|}{|\rho(d)-\rho(b)||\rho(c)-\rho(a)|},$$

depending whether $\rho(a)$ is closer to $\rho(b)$ or $\rho(c)$. Recall that we have an equality between cross ratios

$$\frac{(\rho(a)-\rho(c))(\rho(b)-\rho(d))}{(\rho(a)-\rho(b))(\rho(c)-\rho(d))} = \frac{(a-c)(b-d)}{(a-b)(c-d)}.$$

In particular,

$$\frac{|\rho(a)-\rho(c)||\rho(b)-\rho(d)|}{|\rho(a)-\rho(b)||\rho(c)-\rho(d)|} = \frac{|a-c||b-d|}{|a-b||c-d|} > 1.$$

Hence $\rho(a)$ is closer to $\rho(b)$ and $\rho(d)$ is closer to $\rho(c)$, and we have

$$d_H(\rho(b), \rho(c)) = \frac{|\rho(a) - \rho(c)||\rho(b) - \rho(d)|}{|\rho(a) - \rho(b)||\rho(c) - \rho(d)|} = \frac{|a - c||b - d|}{|a - b||c - d|} = d_H(b, c).$$

$\square$

Example. Suppose b and c both lie on the y-axis, say $b = iu$ and $c = iv$ for some $u > v > 0$. Then

$$d_H(b, c) = \log \frac{|iu|}{|iv|} = \log \frac{u}{v}.$$

It would be useful to have a more straightforward formula for computing the distance $d_H(b, c)$.

Exercise. Prove that $d_H(b, c)$ defined above coincides with the following formula

$$d_H(b, c) = 2 \log \frac{|b - c| + |b - \bar{c}|}{2\sqrt{\mathrm{Im}(b)\mathrm{Im}(c)}}.$$

Proposition 6.33. *d_H is a metric on the upper half plane $\mathbb{H}$.*

Proof. Since the positivity and symmetric properties are clear, it suffices to show that d_H satisfies the triangle inequality. Let x, y, z be three distinct points in $\mathbb{H}$. We would like to show that

$$d(x, z) \leq d(x, y) + d(y, z).$$

Consider the circle in $\hat{\mathbb{C}}$ that passes through x, z and perpendicular to the x-axis. Say $\alpha \in \mathbb{R}$ is one of the intersections of the circle with the x-axis. Consider a real Möbius transformation of the form

$$\rho(z) = \frac{1}{z - \alpha} + \beta.$$

The transformation ρ would send the circle to a line perpendicular to the x-axis. By choosing an appropriate β, one can assume that ρ sends the circle to the y-axis. Since ρ preserves d_H, by applying ρ simultaneous to x, y, z, one may assume that x, z lie on the y-axis.

Let $x = iu$ and $z = iv$ for some $u, v > 0$, and write $y = a + ib$ where $a, b \in \mathbb{R}$ and $b > 0$. Observe that

$$\frac{|(iu) - (a + ib)| + |(iu) - (a - ib)|}{2\sqrt{ub}} \geq \frac{|(iu) - (ib)| + |(iu) - (-ib)|}{2\sqrt{ub}}$$

Hence we have

$$\begin{aligned} d_H(x,y) + d_H(z,y) &\geq d_H(x,ib) + d_H(z,ib) \\ &= \left|\log \frac{u}{b}\right| + \left|\log \frac{v}{b}\right| \\ &\geq \left|\log \frac{u}{v}\right| = d_H(x,z). \end{aligned}$$

□

Remark 6.34. From the above proof, one can observe that for any two points x and z on the y-axis, the shortest path (with respect to the hyperbolic metric d_H) connecting these two points is the line segment between them. In general, for any two points $b, c \in \mathbb{H}$, there exists a unique circle C passing through them and intersects perpendicularly with the x-axis. There exists a real Möbius transformation $\rho \in \text{Möb}(\mathbb{H})$ such that $\rho(C)$ is the y-axis. Therefore, the shortest path between $\rho(b)$ and $\rho(c)$ is the line segment on the y-axis connecting them. Since ρ preserves the distance d_H, this proves that the shortest path connecting b and c is the arc of C connecting them. Therefore, the "straight lines" in $(\mathbb{H}, d_{\mathbb{H}})$ (or more precisely, the *geodesics*) are either:

- straight vertical rays orthogonal to the x-axis, or
- half-circles whose origin is on the x-axis.

$(\mathbb{H}, d_{\mathbb{H}})$ is not Euclidean. For instance, in Euclidean geometry, given any straight line ℓ and any point $p \notin \ell$, there exists a unique line passing through p that does not intersect with ℓ. In $(\mathbb{H}, d_{\mathbb{H}})$ however, there exists infinitely many such lines.

7. Homology groups of topological spaces

7.1. **Ideas of topological spaces.** In this subsection, we give a brief introduction to topological spaces and continuous maps between them. Let us first recall some relevant notions for *metric spaces* as motivation for the more general definitions for topological spaces.

Definition 7.1. Let (X, d) be a metric space. A subset $U \subseteq X$ is called an *open subset* of X if for any $x \in U$, there exists $\epsilon > 0$ such that

$$B(x, \epsilon) := \{y \in X \mid d(x,y) < \epsilon\} \subseteq U.$$

Remark 7.2. Note that what the notion of open subsets actually captures is the concept of *nearness*. Open subsets are the sets that, with every point x in the set, contain all points that are sufficiently near to x.

Exercise. Let τ be the collection of all open subsets of X. Show that:

- The empty set and X itself belong to τ.
- Any arbitrary (finite or infinite) union of members of τ belongs to τ.
- The intersection of any finite number of members of τ belongs to τ.

Moreover, give an example of infinitely many open subsets whose intersection is not open.

Recall the "ϵ-δ" definition of continuous maps between metric spaces.

Definition 7.3. Let (X, d_X) and (Y, d_Y) be two metric spaces. A map $f\colon X \to Y$ is said to be *continuous* if for any $x_0 \in X$ and $\epsilon > 0$, there exists $\delta > 0$ (which depends on both x_0 and δ) such that

$$d_X(x, x_0) < \delta \implies d_Y(f(x), f(x_0)) < \epsilon.$$

Exercise. Prove that $f\colon (X, d_X) \to (Y, d_Y)$ is continuous if and only if

$$\text{“}U \subseteq Y \text{ is an open subset} \implies f^{-1}(U) \subseteq X \text{ is an open subset”}.$$

Here $f^{-1}(U)$ is defined to be $f^{-1}(U) \coloneqq \{x \in X \mid f(x) \in U\}$.

The concept of *topological space* can be regarded as a generalization of metric space. A topological space is, very roughly speaking, a geometrical space in which *nearness* is defined, but can not necessarily be measured by a distance function. It is the most general type of a mathematical space that allows for the definition of *limits*, *continuity*, *connectedness*, *compactness*, etc. Very roughly speaking, two topological spaces are considered equivalent, or *homeomorphic*, if there is a way to continuously stretching and bending one of the spaces into the other. For instance, a tiny circle and a large circle in $\mathbb{R}^2$ are distinct metric spaces, but they are considered the same topologically. On the other hand, a sphere and a donut are not topologically equivalent, since a donut has a "hole" but a sphere does not.

Definition 7.4. A *topology* on a set X is a collection τ of subsets of X satisfying the following axioms:

- The empty set and X itself belong to τ.
- Any arbitrary (finite or infinite) union of members of τ belongs to τ.

- The intersection of any finite number of members of τ belongs to τ.

Members of τ are called *open subsets* of X (with respect to this topology).

Definition 7.5. A map $f\colon X \to Y$ between topological spaces is called *continuous* if

$$U \subseteq Y \text{ is an open subset } \implies f^{-1}(U) \subseteq X \text{ is an open subset.}$$

The map f is called a *homeomorphism* if it is bijective, and both f and f^{-1} are continuous. In this case, X and Y are said to be *homeomorphic*.

Example. Here are some examples of topological spaces that are homeomorphic.

- The open interval (a, b) is homeomorphic to $\mathbb{R}$.
- The unit square in $\mathbb{R}^2$ is homeomorphic to the unit disc.

Non-example. Here are some non-examples.

- $(0, 1)$ is not homeomorphic to $[0, 1]$. The concept of *compactness* makes sense for topological spaces, which homeomorphisms preserve. Note that $[0, 1]$ is compact but $(0, 1)$ is not.
- Similarly, $\mathbb{R}$ is not homeomorphic to the unit circle, since $\mathbb{R}$ is not compact but the unit circle is.
- $\mathbb{R}^n$ is not homeomorphic to $\mathbb{R}^m$ if $n \neq m$. (Why?) This is actually a deep result, for which a proof of it usually requires certain *homology* theory.

7.2. **Ideas of homology groups of topological spaces.** One of the most fundamental questions in topology is to classify topological spaces up to homeomorphisms. In other words, given two topological spaces X and Y, can we tell whether they are homeomorphic or not? Things are usually easier if X and Y are in fact homeomorphic: one simply needs to construct an explicit homeomorphism between X and Y. On the other hand, it is much harder to prove that two spaces X and Y are not homeomorphic, i.e. prove that there does not exist any homeomorphism between them.

In order to distinguish non-homeomorphic topological spaces, mathematicians introduced various *invariants* that one can associate to topological spaces, like fundamental groups $\pi_1(X)$, homology groups $H_k(X)$, cohomology groups $H^k(X)$, homotopy groups $\pi_k(X)$, etc. Take homology groups for instance. For any topological space X, one can define its homology groups $H_0(X)$, $H_1(X)$,

$H_2(X), \ldots$. One of the many important properties that they satisfy is that if X and Y are homeomorphic, then $H_k(X) \cong H_k(Y)$ for all $k \geq 0$. Therefore, if one can show that $H_k(X) \ncong H_k(Y)$ for some k, then this would imply that X and Y are not homeomorphic.

The original motivation for defining homology groups was the observation that two topological spaces can be distinguished by examining their "holes". For instance, a circle is not homeomorphic to a disk because the circle has a "hole" through it while the disk is solid; and the sphere S^2 is not homeomorphic to a circle S^1 because the sphere encloses a "two-dimensional hole" while the circle encloses a "one-dimensional hole". Homology provides a rigorous mathematical method for defining and categorizing holes. For each $k \geq 0$, the k-th homology group $H_k(X)$ captures the information of holes in X with a k-dimensional boundary.

Here is the general strategy for constructing homology groups. One first chooses a *chain complex* $(C_\bullet(X), \partial)$, which is a sequence of abelian groups $C_0(X), C_1(X), C_2(X), \ldots$, where each $C_k(X)$ encodes certain data of k-dimensional strata of X; together with the *boundary operators*, which are group homomorphisms $\partial_n \colon C_n(X) \to C_{n-1}(X)$, to encode how the n-dimensional strata are glued together along their boundaries. In summary, one associates the space X with a sequence of group homomorphisms

$$\cdots \xrightarrow{\partial_{n+1}} C_n(X) \xrightarrow{\partial_n} C_{n-1}(X) \to \cdots \xrightarrow{\partial_2} C_1(X) \xrightarrow{\partial_1} C_0(X) \xrightarrow{\partial_0} 0,$$

with the important property that this complex is *exact*, namely, $\partial_n \circ \partial_{n+1} = 0$ for all n. In other words, *the boundary of a boundary is empty.* Then the homology groups are defined to be

$$H_n(X) := \frac{\operatorname{Ker}(\partial_n)}{\operatorname{Im}(\partial_{n+1})}.$$

There are (at least) two equivalent ways of defining the homology of X: the *simplicial homology* and the *singular homology.* The simplicial homology is much more computable, but it requires X to be a *simplicial complex* (which is true in all cases that we will consider); on the other hand, the definition of singular homology works for any topological space and is easier to work with for theoretical purposes, but it is hard to compute in practice. Lecture 13

7.3. **Simplicial homology.**

Definition 7.6. The *standard n-simplex* is defined to be the following subset of $\mathbb{R}^{n+1}$

$$\Delta_n \coloneqq \{x \in \mathbb{R}^{n+1} \mid x_0 + \cdots + x_n = 1,\ \text{each } x_i \geq 0\}.$$

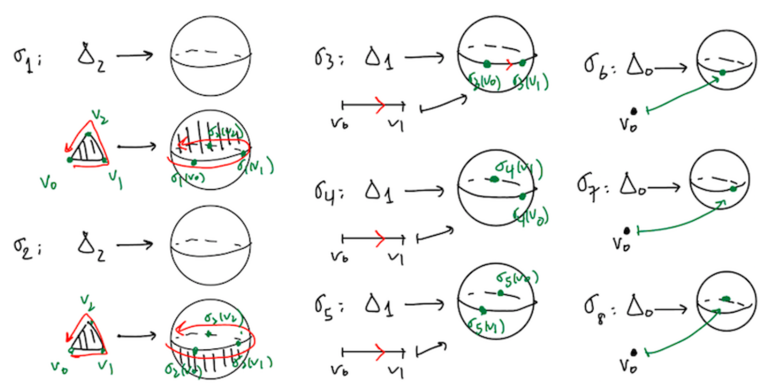
Definition 7.7. A *simplicial complex structure* of a topological space X is a collection of continuous functions

$$\{\sigma_\alpha \colon \Delta_{n_\alpha} \to X\}_{\alpha \in I}$$

such that

- $X = \cup_{\alpha \in I} \sigma_\alpha(\Delta_{n_\alpha})$,
- $\sigma_\alpha|_{\Delta^\circ_{n_\alpha}} \colon \Delta^\circ_{n_\alpha} \to X$ is a homeomorphism onto its image for each $\alpha \in I$,
- each Δ_{n_α} has $n_\alpha + 1$ boundary faces $\Delta^{(0)}_{n_\alpha - 1}, \ldots, \Delta^{(n_\alpha)}_{n_\alpha - 1}$, each of the faces can be canonically identified with $\Delta_{n_\alpha - 1}$; we require that the restriction to each boundary face $\sigma_\alpha|_{\Delta^{(k)}_{n_\alpha - 1}} \colon \Delta_{n_\alpha - 1} \to X$ also belongs to the collection.

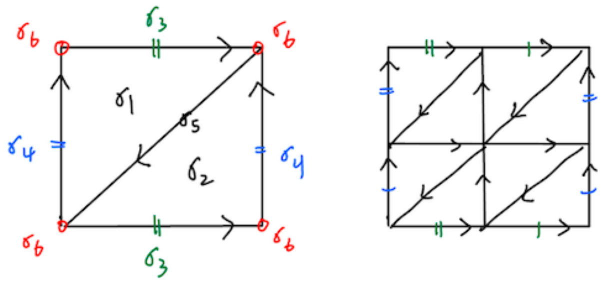
Example. Here is an example of a simplicial complex structure on a sphere. σ_1



and σ_2 are 2-simplices of the sphere whose interiors are mapped to the north (resp. south) hemisphere, and the three boundary edges are mapped to three segments of the equator. $\sigma_3, \sigma_4, \sigma_5$ are 1-simplices of the sphere whose images are precisely these segments on the equator, and $\sigma_6, \sigma_7, \sigma_8$ are 0-simplices of the sphere that mapped to the intersections of these three segments.

Example. Here are two examples of simplicial complex structures on a *torus.*

It should be clear from the examples that there are infinitely many possible simplicial complex structures on a topological space, and the *number* of complexes is *not* a topological invariant.



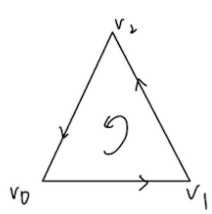
Given a simplicial complex structure of X, one would like to associate to it a *chain complex*

$$\cdots \xrightarrow{\partial_{n+1}} C_n(X) \xrightarrow{\partial_n} C_{n-1}(X) \to \cdots \xrightarrow{\partial_2} C_1(X) \xrightarrow{\partial_1} C_0(X) \xrightarrow{\partial_0} 0.$$

A key concept in defining the *boundary operators* ∂_n is the notion of an *orientation* of a simplex. By definition, an orientation of an n-simplex is given by an ordering of the vertices, written as $[v_0, \ldots, v_n]$, with the rule that two orderings define the same orientation if and only if they differ by an *even* permutation. Intuitively, one might wish to define the boundary of $[v_0, \ldots, v_n]$ to be the sum of its faces $[v_0, \ldots, \hat{v}_i, \ldots, v_n]$. However, it turns out to be better to insert certain *signs* and instead let the boundary of $[v_0, \ldots, \hat{v}_i, \ldots, v_n]$ be

$$\partial[v_0, \ldots, v_n] \coloneqq \sum_{i=0}^{n} (-1)^i [v_0, \ldots, \hat{v}_i, \ldots, v_n].$$

Here, we define the negation of an oriented complex to be the same complex but with the opposite orientation. Heuristically, the signs are inserted to take orientations into account, so that all the faces of a simplex are coherently oriented. For instance,



$$\begin{aligned}\partial[v_0,v_1,v_2] &= [v_1,v_2] - [v_0,v_2] + [v_0,v_1] \\ &= [v_1,v_2] + [v_2,v_0] + [v_0,v_1]\end{aligned}$$

Definition 7.8. Let $\{\sigma_\alpha \colon \Delta_{n_\alpha} \to X\}_{\alpha \in I}$ be a simplicial complex structure of X. We associate to it a chain complex $(C_\bullet(X), \partial)$ as follows.

For each $k \geq 0$, define $C_k(X)$ to be the free abelian group with basis the oriented k-simplices. Equivalently, any element of $C_k(X)$ is a formal sum

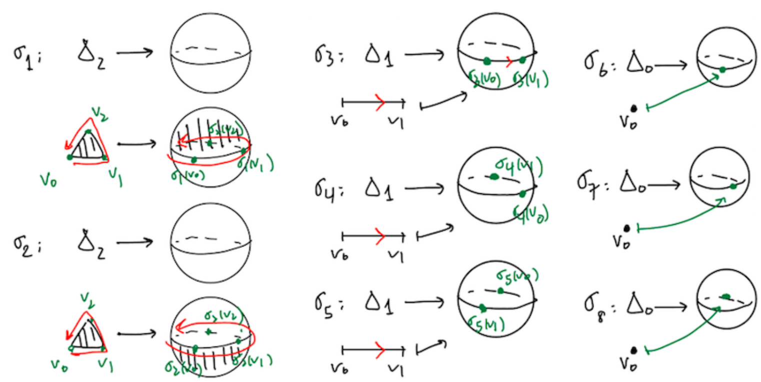
$$\sum_{j=1}^{n} m_j \sigma_{\alpha_j},$$

where each m_j is an integer and each $n_{\alpha_j} = k$ (i.e. $\sigma_{\alpha_j}\colon \Delta_k \to X$ is an oriented k-simplex of X).

The *boundary operator* $\partial_k\colon C_k(X) \to C_{k-1}(X)$ can be defined by specifying its values on basis elements: Let $\sigma_\alpha\colon \Delta_k \to X$ be an oriented k-simplex of X, define

$$\partial_k(\sigma_\alpha) \coloneqq \sum_{i=0}^{k} (-1)^i \sigma_\alpha|_{[v_0,\ldots,\hat{v}_i,\ldots,v_k]}.$$

Example. Let us consider the following simplicial structure on S^2. We have



$$\partial_2(\sigma_1) = \sigma_1|_{[v_1,v_2]} - \sigma_1|_{[v_0,v_2]} + \sigma_1|_{[v_0,v_1]} = \sigma_1|_{[v_1,v_2]} + \sigma_1|_{[v_2,v_0]} + \sigma_1|_{[v_0,v_1]} = \sigma_4 + \sigma_5 + \sigma_3.$$

Similarly, we also have $\partial_2(\sigma_2) = \sigma_3 + \sigma_4 + \sigma_5$. It is easy to see that

$$\partial_1(\sigma_3) = \sigma_7 - \sigma_6;\ \partial_1(\sigma_4) = \sigma_8 - \sigma_7;\ \partial_1(\sigma_5) = \sigma_6 - \sigma_8.$$

Let us verify that $\partial_1 \circ \partial_2 = 0$. Indeed,

$$\partial_1(\partial_2(\sigma_1)) = \partial_1(\sigma_3 + \sigma_4 + \sigma_5) = (\sigma_7 - \sigma_6) + (\sigma_8 - \sigma_7) + (\sigma_6 - \sigma_8) = 0.$$

Similarly, one can show that $\partial_1(\partial_2(\sigma_2)) = 0$. Hence we have an *exact* sequence

$$0 \xrightarrow{\partial_3} C_2(S^2)(\cong \mathbb{Z}^2) \xrightarrow{\partial_2} C_1(S^2)(\cong \mathbb{Z}^3) \xrightarrow{\partial_1} C_0(S^2)(\cong \mathbb{Z}^3) \xrightarrow{\partial_0} 0.$$

The second homology group of a sphere is

$$H_2(S^2) = \frac{\ker(\partial_2)}{\operatorname{Im}(\partial_3)} = \frac{\langle \sigma_1 - \sigma_2 \rangle}{\{0\}} \cong \mathbb{Z}.$$

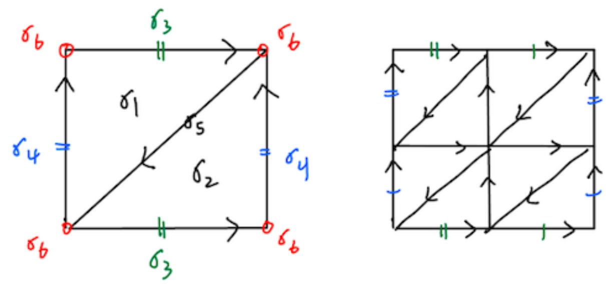
The first homology group of a sphere is

$$H_1(S^2) = \frac{\ker(\partial_1)}{\operatorname{Im}(\partial_2)} = \frac{\langle \sigma_3 + \sigma_4 + \sigma_5 \rangle}{\langle \sigma_3 + \sigma_4 + \sigma_5 \rangle} \cong \{0\}.$$

The zeroth homology group of a sphere is

$$H_0(S^2) = \frac{\ker(\partial_0)}{\operatorname{Im}(\partial_1)} = \frac{\langle \sigma_6, \sigma_7, \sigma_8 \rangle}{\langle \sigma_6 - \sigma_7, \sigma_7 - \sigma_8 \rangle} \cong \mathbb{Z}.$$

Example. Let's compute the simplicial homology group of a torus T^2 by the simplicial structure on the left. We have



$$\partial_2(\sigma_1) = \partial_2(\sigma_2) = \sigma_3 + \sigma_4 + \sigma_5.$$
$$\partial_1(\sigma_3) = \partial_1(\sigma_4) = \partial_1(\sigma_5) = \sigma_6 - \sigma_6 = 0.$$
$$\partial_0(\sigma_6) = 0.$$

Hence

$$H_2(T^2) = \frac{\ker(\partial_2)}{\operatorname{Im}(\partial_3)} = \frac{\langle \sigma_1 - \sigma_2 \rangle}{\{0\}} \cong \mathbb{Z}.$$
$$H_1(T^2) = \frac{\ker(\partial_1)}{\operatorname{Im}(\partial_2)} = \frac{\langle \sigma_3, \sigma_4, \sigma_5 \rangle}{\langle \sigma_3 + \sigma_4 + \sigma_5 \rangle} \cong \mathbb{Z}^2.$$
$$H_0(T^2) = \frac{\ker(\partial_0)}{\operatorname{Im}(\partial_1)} = \frac{\langle \sigma_6 \rangle}{\{0\}} \cong \mathbb{Z}.$$

Exercise. Use the simplicial structure on the right (which consists of 8 triangles) to compute the homology group of T^2, and verify that one gets the same result.

Let $\{\sigma_\alpha \colon \Delta_{n_\alpha} \to X\}_{\alpha \in I}$ be a simplicial complex structure of X, and let $(C_\bullet(X), \partial)$ be the associated chain complex. It is not hard to show that the complex is *exact*, i.e. $\partial_k \circ \partial_{k+1} = 0$, or in words, the boundary of a boundary is empty.

Lemma 7.9. $\partial_k \circ \partial_{k+1} = 0$.

Proof. Let σ be a $(k+1)$-simplex of X. Then

$$\begin{aligned}
\partial_k(\partial_{k+1}(\sigma)) &= \partial_k\Big(\sum_{i=0}^{k+1} (-1)^i \sigma|_{[v_0,\dots,\hat{v_i},\dots,v_{k+1}]} \Big) \\
&= \sum_{i=0}^{k+1} (-1)^i \partial_k \Big(\sigma|_{[v_0,\dots,\hat{v_i},\dots,v_{k+1}]} \Big) \\
&= \sum_i (-1)^i \Big(\sum_{j<i} (-1)^j \sigma|_{[v_0,\dots,\hat{v_j},\dots,\hat{v_i},\dots,v_{k+1}]} + \sum_{j>i} (-1)^{j-1} \sigma|_{[v_0,\dots,\hat{v_i},\dots,\hat{v_j},\dots,v_{k+1}]} \Big) \\
&= 0.
\end{aligned}$$

□

The *homology groups* of X is then defined to be the quotient groups

$$H_k(X) \coloneqq \frac{\operatorname{Ker}(\partial_k)}{\operatorname{Im}(\partial_{k+1})}.$$

A priori this definition depends on the choice of the simplicial complex structure of X. What's amazing about this construction is that the homology groups defined this way is actually *independent* of the choice of the simplicial structure that one starts with, so it is indeed a topological invariant.

7.4. **Singular homology.** The definition of singular homology does not depend on a choice of a simplicial structure, so it can be defined for any topological space X.

A *singular n-simplex* in X is by definition just a continuous map $\sigma \colon \Delta_n \to X$ (it does not have to be injective in Δ_n°; it can even be a constant map). Define $C_n(X)$ to be the free abelian group with basis the set of singular n-simplices in X. Elements of $C_n(X)$ can be written as $\sum_{i=1}^k m_i\sigma_i$ where $m_i \in \mathbb{Z}$ and σ_i's are singular n-simplices in X. One can define the boundary operator $\partial_n \colon C_n(X) \to C_{n-1}(X)$ as before, and we still have $\partial_n \circ \partial_{n+1} = 0$. Thus there is an exact sequence

$$\cdots \xrightarrow{\partial_{n+1}} C_n(X) \xrightarrow{\partial_n} C_{n-1}(X) \to \cdots \xrightarrow{\partial_2} C_1(X) \xrightarrow{\partial_1} C_0(X) \xrightarrow{\partial_0} 0,$$

and we define the *singular homology* of X to be

$$H_n(X) \coloneqq \frac{\operatorname{Ker}(\partial_n)}{\operatorname{Im}(\partial_{n+1})}.$$

It is a theorem in *algebraic topology* that the singular homology groups coincide with the simplicial homology groups that we defined in the previous subsection.

Remark 7.10. Here are some comparisons between simplicial and singular homology groups. Simplicial homology is often quite computable; in contrast, the number of singular n-simplices in X usually is uncountable, which makes singular homology hard to compute in practice. On the other hand, singular homology is easier to work with for theoretical purposes; if one wants to prove certain statements about homology groups, it is often easier to consider the singular homology instead of the simplicial homology. For instance, it is evident from the definition that homeomorphic spaces have isomorphic singular homology groups, in contrast with the situation for simplicial homology groups.

By using the definition of singular homology groups, one can show the following nice properties.

Theorem 7.11. *Each continuous map $f \colon X \to Y$ between topological spaces induces group homomorphisms between their homology groups $H_k(f) \colon H_k(X) \to H_k(Y)$. Moreover,*

- *if $X \xrightarrow{f} Y \xrightarrow{g} Z$ are two continuous maps, then $H_k(g \circ f) = H_k(g) \circ H_k(f)$;*
- *$H_k(\mathrm{id}_X) = \mathrm{id}_{H_k(X)}$;*
- *if $f \colon X \to Y$ is a homeomorphism, then $H_k(f) \colon H_k(X) \to H_k(Y)$ is a group isomorphism.*

Proof. First, we show that $f \colon X \to Y$ induces group homomorphisms between their chain complexes $C_k(f) \colon C_k(X) \to C_k(Y)$. It is clear from the definition of singular homology: since $C_k(X)$ consists of a basis of continuous maps $\sigma\colon \Delta_k \to X$, one can simply define the map by $C_k(f)\colon \sigma \mapsto f \circ \sigma$. The following statements are easy exercises:

- We have $C_{k-1}(f) \circ \partial_{k,X} = \partial_{k,Y} \circ C_k(f)\colon C_k(X) \to C_{k-1}(Y)$.
- Then one can use this to show that $C_k(f)(\mathrm{Ker}(\partial_{k,X})) \subseteq \mathrm{Ker}(\partial_{k,Y})$ and $C_k(f)(\mathrm{Im}(\partial_{k+1,X})) \subseteq \mathrm{Im}(\partial_{k+1,Y})$.
- Therefore, the continuous map f induces a group homomorphism $H_k(X) \to H_k(Y)$.

The remaining statements follow easily from the construction. $\square$

Remark 7.12. This theorem tells us that the homology groups give a *functor* between the *category of topological spaces* and the *category of abelian groups*: homology not only is an invariant of individual topological spaces, but also can be used to study *maps between topological spaces*.

Below are some nice applications of these properties of homology groups.

Theorem 7.13 (Invariance of dimension). *$\mathbb{R}^n$ is not homoemorphic to $\mathbb{R}^m$ if $n \neq m$.*

Proof. Assume that $\mathbb{R}^n$ is homoemorphic to $\mathbb{R}^m$. Let $f\colon \mathbb{R}^n \to \mathbb{R}^m$ be a homeomorphism. Choose any point $x \in \mathbb{R}^m$. Then the map $\bar{f}\colon \mathbb{R}^n\backslash\{x\} \to \mathbb{R}^m\backslash\{f(x)\}$ is still a homeomorphism. Note that $\mathbb{R}^n\backslash\{x\}$ has a *deformation retract* to a sphere S^{n-1}, hence $H_k(\mathbb{R}^n\backslash\{x\}) \cong H_k(S^{n-1})$ is $\mathbb{Z}$ for $k = 0, n-1$ and is 0 otherwise. Similarly, $H_k(\mathbb{R}^m\backslash\{f(x)\}) \cong H_k(S^{m-1})$ is $\mathbb{Z}$ for $k = 0, m-1$ and is 0 otherwise. The homoemorphism $\bar{f}\colon \mathbb{R}^n\backslash\{x\} \to \mathbb{R}^m\backslash\{f(x)\}$ induces isomorphisms between their homology groups, thus we obtain $n = m$. $\square$

Remark 7.14. In general, two continuous functions $f, g\colon X \to Y$ are called *homotopic* if there exists a continuous map $F\colon X \times [0,1] \to Y$ such that $F(x,0) = f(x)$ and $F(x,1) = g(x)$ hold for all $x \in X$. It is a theorem in *algebraic topology* that if f and g are homotopic, then they induce the same group homomorphisms on homology groups $H_k(f) = H_k(g)\colon H_k(X) \to H_k(Y)$.

We say a space X has a *deformation retract* to its subspace A if there exists a continuous map $F\colon X \times [0,1] \to X$ such that

- $F(x,0) = x$ for all $x \in X$;
- $F(x,1) \in A$ for all $x \in X$;

- $F(a, 1) = a$ for all $a \in A$.

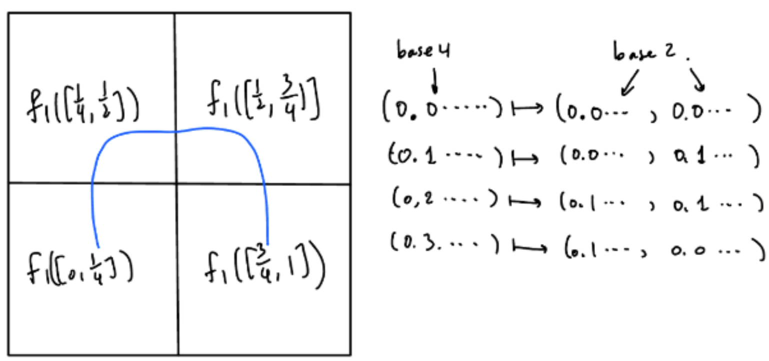
Denote $r \coloneqq F(-, 1)\colon X \to A$, and $i\colon A \to X$ the inclusion. Then

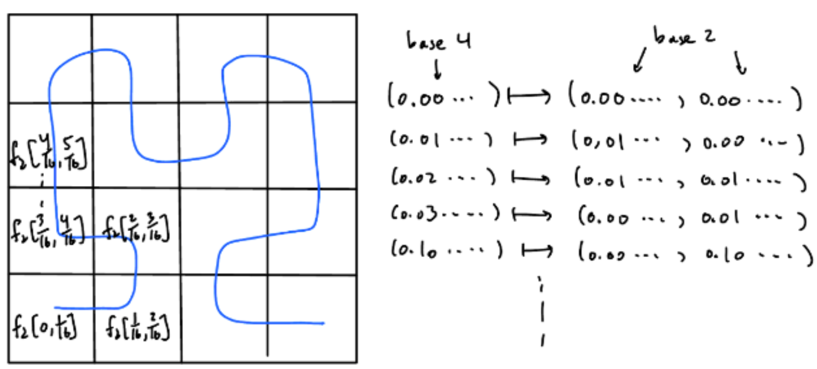
- $i \circ r\colon X \to X$ is homotopic to id_X;
- $r \circ i\colon A \to A$ is the identity map id_A.

This proves that $H_k(A) \cong H_k(X)$.

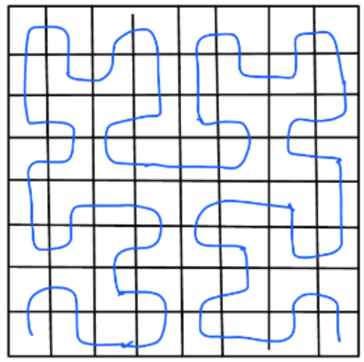
Remark 7.15. Invariance of domain is not as intuitive as it appears to be. For instance, there exists a *space filling curve* constructed by Peano, which is a *continuous* and *surjective* map $[0, 1] \to [0, 1] \times [0, 1]$ (!). Lecture 14

Such "unusual" functions are usually constructed by taking limit of certain sequence of functions. Let us sketch a construction of a space filling curve. Choose a sequence of functions $f_n\colon [0, 1] \to [0, 1] \times [0, 1]$ as follows. It is





not hard to show that the sequence of functions (f_n) converges uniformly to a continuous function $f\colon [0, 1] \to [0, 1] \times [0, 1]$ (using the Cauchy criterion). To show that f is surjective, for any $(x_1, x_2) \in [0, 1] \times [0, 1]$, one can consider



the *base 2* decimal expansion of x_1 and x_2, then the definition of (f_n) gives a way to identify a preimage of (x_1, x_2) in terms of *base 4* decimal expansion of points in $[0, 1]$.

Theorem 7.16 (Hairy ball theorem)**.** *S^n has a continuous field of* nonzero *tangent vectors if and only if n is odd.*

Remark 7.17. When $n = 2$, the theorem can be interpreted as: "there is always at least one point on Earth where the wind is not blowing".

Proof. Suppose $x \mapsto v(x) \in \mathbb{R}^{n+1}$ is a nonzero tangent vector field on S^n. The tangency condition means that x and $v(x)$ are perpendicular in $\mathbb{R}^{n+1}$. Then

$$f_t(x) = (\cos t)x + (\sin t)\frac{v(x)}{|v(x)|}$$

gives a homotopy between the identity map of S^n (when $t = 0$) and the antipodal map $x \mapsto -x$ of S^n (when $t = \pi$).

Now we introduce the notion of *degree* of a map $f : S^n \to S^n$. Recall that $H_n(S^n) = \langle \alpha \rangle \cong \mathbb{Z}$. The induced map $H_n(f) : H_n(S^n) \to H_n(S^n)$ sends the generator α to an integer multiple of it, say $d\alpha$. Then $d \in \mathbb{Z}$ is defined to be the *degree* of f. It is not hard to show the following statements:

- A reflection of S^n has degree -1.
- The antipodal map of S^n is the composition of $n + 1$ reflections.
- $\deg(fg) = \deg(f) \deg(g)$.

These imply that the degree of the antipodal map is $(-1)^{n+1}$. On the other hand, the degree of the identity map is 1. Since homotopy preserves degrees, we have $(-1)^{n+1} = 1$, hence n is odd.

Conversely, when n is odd, it is easy to construct a nonzero tangent vector field on $S^n \subseteq \mathbb{R}^{n+1}$, for instance:

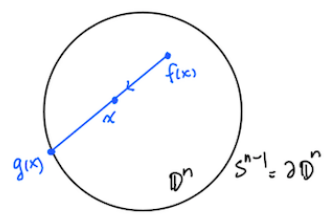
$$v(x_1, \ldots, x_{n+1}) = (-x_2, x_1, -x_4, x_3, \ldots, -x_{n+1}, x_n).$$

□

7.5. **Brouwer fixed point theorem.** Here is a motivation example: Drop a map on the ground. There must exists a point where it matches with the corresponding point on the map. *Brouwer fixed point theorem* is a far reaching generalization of this fact. It has many interesting and important applications, which will be discussed in the next subsection.

Theorem 7.18 (Brouwer fixed point theorem)**.** *Let $\mathbb{D}^n = \{x \in \mathbb{R}^n \mid ||x|| \leq 1\}$ be the closed unit ball in $\mathbb{R}^n$. Any continuous map $f\colon \mathbb{D}^n \to \mathbb{D}^n$ has at least one fixed point, i.e. there exists $x \in \mathbb{D}^n$ such that $f(x) = x$.*

The idea of the proof is to prove by contradiction. Suppose there exists a continuous map $f\colon \mathbb{D}^n \to \mathbb{D}^n$ with no fixed points. Then one can define a map



$g\colon \mathbb{D}^n \to S^{n-1}$ as shown in the figure, with the following properties:

- $g\colon \mathbb{D}^n \to S^{n-1}$ is continuous;
- if $x \in S^{n-1} = \partial\mathbb{D}^n$, then $g(x) = x$.

Denote the inclusion map by $i\colon S^{n-1} \to \mathbb{D}^n$, which is obviously continuous. So, under the assumption that f has no fixed points, we obtain two continuous maps

$$S^{n-1} \xrightarrow{i} \mathbb{D}^n \xrightarrow{g} S^{n-1}$$

such that $g \circ i = \mathrm{id}_{S^{n-1}}$. However, this is not possible. There are several ways to see this.

First, we can prove that this is not possible by computing the *homology groups*. It is an exercise to show that

$$H_0(\mathbb{D}^n) \cong \mathbb{Z}, \text{ and } H_k(\mathbb{D}^n) = 0 \text{ for any } k > 0;$$

while

$$H_0(S^n) \cong H_n(S^n) \cong \mathbb{Z}, \text{ and } H_k(\mathbb{D}^n) = 0 \text{ for any } k \neq 0, n.$$

Now, if we consider the induced group homomorphisms of the $(n-1)$-th homology groups of the continuous maps $S^{n-1} \xrightarrow{i} \mathbb{D}^n \xrightarrow{g} S^{n-1}$, one obtains

$$H_{n-1}(S^{n-1}) \xrightarrow{H_{n-1}(i)} H_{n-1}(\mathbb{D}^n) \xrightarrow{H_{n-1}(g)} H_{n-1}(S^{n-1}).$$

On the one hand, since $H_{n-1}(\mathbb{D}^n) = 0$, the composition of these two group homomorphisms must be the zero morphism. On the other hand, since $g \circ i = \mathrm{id}_{S^{n-1}}$, we have $H_{n-1}(g) \circ H_{n-1}(i) = H_{n-1}(\mathrm{id}_{S^{n-1}}) = \mathrm{id}_{H_{n-1}(S^{n-1})}$ is not the zero morphism since $H_{n-1}(S^{n-1}) \cong \mathbb{Z} \neq 0$. This concludes the proof.

Let us sketch an alternative proof, which uses the concepts of *differential forms* and the *generalized Stokes theorem*. First, we recall the generalized Stokes theorem. It is a simultaneous generalization of the fundamental theorem of calculus, Green's theorem, and the Stokes theorem that one learned from (multivariable) calculus. Let B be an n-dimensional solid, and denote ∂B its $(n-1)$-dimensional boundary. Denote the inclusion by $i \colon \partial B \to B$. Let ω be a *differential* $(n-1)$*-form* on B. One can then obtain:

- $d\omega$, a differential n-form on B, and
- $i^*\omega$, a differential $(n-1)$-form on ∂B (the *pullback* of ω via i).

The *(generalized) Stokes theorem* states that

$$\int_B d\omega = \int_{\partial B} i^*\omega.$$

Example. Let's consider the case when the dimension $n = 1$. Then B is a closed interval $[a, b]$, with boundary $\partial B = \{a, b\}$. Let f be a differentiable function and consider the differential 0-form $\omega = f(x)$ on B. Then $d\omega = f'(x)dx$ is a differential 1-form on B, and $i^*\omega$ is a differential 0-form on ∂B (which takes value $f(a)$ at the point a and $f(b)$ at the point b). Then the Stokes theorem in this case is simply the fundamental theorem of calculus

$$\int_a^b f'(x)dx = f(b) - f(a).$$

Note that the minus sign on the right hand side arises from the fact that the boundary points a and b have opposite *orientations*.

Example. Let B be a two-dimension solid, and let $\omega = Fdx + Gdy$ be a differential 1-form on B. Then $d\omega = \left(\frac{\partial G}{\partial x} - \frac{\partial F}{\partial y}\right) dxdy$ is a differential 2-form on B, and the Stokes theorem gives

$$\iint_B \left(\frac{\partial G}{\partial x} - \frac{\partial F}{\partial y}\right) dxdy = \oint_{\partial B} Fdx + Gdy,$$

which recovers the Green's theorem.

Let us return to the alternative proof of the Brouwer fixed point theorem. We would like to show that there does not exist g such that the composition

$$S^{n-1} \xrightarrow{i} \mathbb{D}^n \xrightarrow{g} S^{n-1}$$

is the identity map on S^{n-1}. Now we use a fact from *differential geometry*, which states that for any compact manifold M of dimension k, there exists a differential k-form ω on M such that $\int_M \omega > 0$. Such ω is called a *volume form* of M. Let us choose a volume form $\omega_{S^{n-1}}$ of the sphere S^{n-1}. We have

$$\begin{aligned}
0 &< \int_{S^{n-1}} \omega_{S^{n-1}} \\
&= \int_{S^{n-1}} i^* g^* \omega_{S^{n-1}} \text{ (since } g \circ i = \mathrm{id}_{S^{n-1}}) \\
&= \int_{\mathbb{D}^n} d(g^* \omega_{S^{n-1}}) \text{ (by Stokes theorem)} \\
&= \int_{\mathbb{D}^n} g^*(d\omega_{S^{n-1}}) \text{ (commutativity between pullbacks and differentials)} \\
&= 0 \text{ (since there is no diffenertial } n\text{-forms on } S^{n-1}).
\end{aligned}$$

This proves the Brouwer fixed point theorem by contradiction.

7.6. **Applications of the fixed point theorem.** We discuss some important applications of the Brouwer fixed point theorem.

Theorem 7.19 (Fundamental theorem of algebra)**.** *Let $f(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_0$ be a non-constant complex polynomial. Then there exists $z \in \mathbb{C}$ such that $f(z) = 0$.*

Proof. Let $R = 2 + |a_{n-1}| + \cdots + |a_0| > 0$. Define a function $g\colon \mathbb{C} \to \mathbb{C}$ as follows:

- if $|z| \leq 1$, write $z = re^{i\theta}$ where $0 \leq r \leq 1$ and $0 \leq \theta < 2\pi$, define
$$g(z) \coloneqq z - \frac{f(z)}{Re^{i\theta(n-1)}};$$
- if $|z| \geq 1$, define
$$g(z) \coloneqq z - \frac{f(z)}{Rz^{n-1}}.$$

It is easy to see that g is continuous. Moreover, we claim that $|g(z)| \leq R$ for any $|z| \leq R$.

- If $|z| \leq 1$, then
$$|g(z)| \leq |z| + \frac{|f(z)|}{R} \leq 1 + \frac{1 + |a_{n-1} + \cdots + a_0|}{R} < 2 \leq R.$$
- If $1 \leq |z| \leq R$, then
$$\begin{aligned} |g(z)| &= \left| \frac{(R-1)z^n - a_{n-1}z^{n-1} - \cdots - a_0}{Rz^{n-1}} \right| \\ &\leq \frac{(R-1)|z|^n + |a_{n-1}||z|^{n-1} + \cdots + |a_0|}{R|z|^{n-1}} \\ &= \frac{(R-1)|z|}{R} + \frac{|a_{n-1}|}{R} + \frac{|a_{n-2}|}{R|z|} + \cdots + \frac{|a_0|}{R|z|^{n-1}} \\ &\leq (R-1) + \frac{|a_{n-1}| + \cdots + |a_0|}{R} \\ &< (R-1) + 1 = R. \end{aligned}$$

Then, by the Brouwer fixed point theorem, there exists $|z| \leq R$ such that $g(z) = z$, which implies that $f(z) = 0$ by the definition of the function g. $\square$

The following is a linear-algebraic statement that can be proved by the fixed point theorem.

Theorem 7.20. *Let $A \in M_{n\times n}(\mathbb{R})$ be a square matrix whose entries are all* positive. *Then A has a positive eigenvalue $r > 0$. Moreover, there exists an eigenvector v of A with eigenvalue r such that all components of v are non-negative real numbers.*

Proof. Consider the set
$$K = \{x \in \mathbb{R}^n \mid ||x|| = 1,\ x_i \geq 0\},$$

which is a closed subset of S^{n-1}. Define the map

$$f_A\colon K \to K; \quad x \mapsto \frac{Ax}{||Ax||}.$$

Here the image of f_A lies in K follows from the assumption that the matrix A is positive. Note that K is homeomorphic to $\mathbb{D}^{n-1}$, hence by the fixed point theorem, the continuous map $f_A\colon K \to K$ has a fixed point. Say $v \in K$ satisfies $f_A(v) = v$. Then v is an eigenvalue with non-negative components, with eigenvalue given by $r := ||Av|| > 0$. $\square$

Remark 7.21. The theorem above is a much weaker version of the *Perron–Frobenius theorem*, which has a wide range of applications, including Markov chains, dynamical systems, etc. Here are some further statements of the Perron–Frobenius theorem: Under the same assumptions as in the previous theorem, we have:

- any other eigenvalue of A (possibly complex) has absolute value strictly less than r;
- the eigenspace of r is one-dimensional;
- there exists an eigenvector v of A with eigenvalue r such that all components of v are positive;
- all eigenvectors with eigenvalue$\neq r$ have at least one negative or non-real component.

The proofs of these statements require some further arguments.

7.7. **Lefschetz fixed point theorem.** For a simplicial complex X, one can define the chain complex over $\mathbb{R}$ in exactly the same way

$$\to C_n(X,\mathbb{R}) \xrightarrow{\partial_n} C_{n-1}(X,\mathbb{R}) \to \cdots \to C_1(X,\mathbb{R}) \xrightarrow{\partial_1} C_0(X,\mathbb{R}) \to 0,$$

then define the homology groups $H_*(X,\mathbb{R})$ over real numbers. Let $f\colon X \to X$ be a continuous map. The *Lefschetz number* of f is defined to be

$$L(f) := \sum_{i\geq 0}(-1)^i \mathrm{tr}\left(H_n(f)\colon H_i(X,\mathbb{R}) \to H_i(X,\mathbb{R})\right).$$

Here tr denotes the *trace* of the linear maps $H_*(f)$.

Theorem 7.22 (Lefschetz fixed point theorem)**.** *Let X be a compact triangulable space, and $f\colon X \to X$ be a continuous map. If $L(f) \neq 0$, then f has a fixed point.*

Remark 7.23. This is a far reaching generalization of the Brouwer fixed point theorem. To see this, take $X = \mathbb{D}^n$ (or more generally, any *contractible* space). One can show that $L(f) = 1$ for any $f\colon \mathbb{D}^n \to \mathbb{D}^n$, since $H_0(f) = \mathrm{id}_{H_0(X,\mathbb{R})}$ and $H_k(f) = 0$ for any $k > 0$. Therefore f has a fixed point by Lefschetz's fixed point theorem.

Before proving the theorem, let us introduce some notations. Denote $Z_n(X,\mathbb{R}) = \mathrm{Ker}(\partial_n)$ and $B_n(X,\mathbb{R}) = \mathrm{Im}(\partial_{n+1})$. Any continuous map $f\colon X \to X$ induces linear maps $Z_n(f)$ and $B_n(f)$ on the corresponding vector spaces.

Lemma 7.24. $L(f) = \sum_{i\geq 0}(-1)^i \mathrm{tr}(H_i(f)) = \sum_{i\geq 0}(-1)^i \mathrm{tr}(C_i(f))$.

Proof. There are exact sequences

$$0 \to B_i(X,\mathbb{R}) \to Z_i(X,\mathbb{R}) \to H_i(X,\mathbb{R}) \to 0$$

and

$$0 \to Z_i(X,\mathbb{R}) \to C_i(X,\mathbb{R}) \to B_{i-1}(X,\mathbb{R}) \to 0.$$

It is an easy linear algebra exercise to show that these imply $\mathrm{tr}Z_i(f) = \mathrm{tr}B_i(f) + \mathrm{tr}H_i(f)$ and $\mathrm{tr}C_i(f) = \mathrm{tr}Z_i(f) + \mathrm{tr}B_{i-1}(f)$. Hence

$$\begin{aligned} L(f) &= \sum_{i\geq 0}(-1)^i \mathrm{tr}(H_i(f)) \\ &= \sum_i (-1)^i \left(\mathrm{tr}Z_i(f) - \mathrm{tr}B_i(f)\right) \\ &= \sum_i (-1)^i \left(\mathrm{tr}C_i(f) - \mathrm{tr}B_{i-1}(f) - \mathrm{tr}B_i(f)\right) \\ &= \sum_i (-1)^i \mathrm{tr}C_i(f). \end{aligned}$$

□

Sketch of proof of Lefschetz fixed point theorem. Suppose f has no fixed point. One can choose a triangulation fine enough so that f sends each simplex to a different simplex (here one needs the compactness assumption on X). This implies that $\mathrm{tr}(C_i(f)) = 0$ for all $i \geq 0$. Hence $L(f) = 0$ by the previous lemma. □

Remark 7.25. Here is an alternative proof of the fundamental theorem of algebra, using the Lefschetz fixed point theorem. First, observe that any (complex) polynomial can be expressed as the characteristic polynomial of a (complex) matrix. Therefore, to show that any polynomial has a (complex) root, it suffices to show that any complex matrix has a complex eigenvalue (and eigenvectors). Let $T\colon \mathbb{C}^{n+1} \to \mathbb{C}^{n+1}$ be a linear map. The *complex projective space* is defined to be

$$\mathbb{CP}^n \coloneqq (\mathbb{C}^{n+1} \backslash \{0\})/\mathbb{C}^*,$$

or equivalently, the space of 1-dimensional subspaces of $\mathbb{C}^{n+1}$. Then T has an eigenvector is equivalent to its induced map $\overline{T}\colon \mathbb{CP}^n \to \mathbb{CP}^n$ has a fixed point. One can show the following:

- $H_k(\mathbb{CP}^n) \cong \mathbb{Z}$ if $k = 0, 2, 4, \ldots, 2n$ and $H_k(\mathbb{CP}^n) = 0$ otherwise.
- $\mathrm{GL}_{n+1}(\mathbb{C})$ is a connected *Lie group*, hence any invertible linear map of $\mathbb{C}^{n+1}$ is homotopic to the identity map, so is the induced map on $\mathbb{CP}^n$.

These would imply that the Lefschetz number $L(\overline{T}) = n + 1 \neq 0$, thus $\overline{T}$ has a fixed point.

This argument also hinted on why the fundamental theorem of algebra fails for real numbers (i.e. not every real polynomial has a real root): The homology groups of the real projective space is $H_k(\mathbb{RP}^n) \cong \mathbb{Z}$ for $k = 0, 1, \ldots, n$ and is zero otherwise. The Lefschetz number $L(\overline{T})$ would vanish if n is odd. This corresponds to the fact that there exists real polynomial without any real root for degree $n + 1$ polynomial when $n + 1$ is even.

Lecture 15

8. Where to go from here

f