# INTRODUCTION TO ALGEBRA: FALL 2017

MENG HSUAN HSIEH

## Contents

## 1. Basic Ideas from Linear Algebra

Let's recall a couple of definitions that were studied in depth in linear algebra.

**Definition 1.1.** Matrices containing $m$ rows and $m$ columns are denoted as having dimensions $m \times n$.

**Definition 1.2.** Matrix multiplication $AB$ can only happen when # of columns of $A =$ # of rows of $B$.

*Date*: July 25, 2018.

Based on undergraduate Algebra course taught by Robert Ji Wai Young. Please contact `menghsuan.hsieh@nyu.edu` for any typos, maths errors, and the like.

**Properties 1.3** (Matrix multiplication)**.** (a) Distributive laws apply, i.e.

$$A(B+B') = AB + AB'$$
$$(B+B')A = BA + B'A$$

(b) Associative law applies, i.e.

$$A(BC) = (AB)C$$

(c) Commutative law does not *usually* apply, i.e.

$$AB \neq BA$$

**Definition 1.4.** A **right inverse** is a matrix $R$ s.t. $AR = I$, if $A$ is a square matrix.

**Definition 1.5.** A **left inverse** is a matrix $L$ s.t. $LA = I$, if $A$ is a square matrix.

**Lemma 1.6.** Let $A$ be a square matrix that has both right and left inverses (using notations as given above). Then, $R = L$. Furthermore, $A$ is invertible and $R$ is its inverse.

*Proof.* $R = IR = (LA)R = L(AR) = LI = L$. ■

**Lemma 1.7.** Inverse of a 2 × 2 matrix of the form

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

denoted $A^{-1}$, is given by the formula

$$A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

One can show this by applying Cramer's rule.

**Lemma 1.8.** A square matrix that has either a row of zeros or a column of zeros is not invertible.

*Proof.* This is relatively straightforward. Suppose $A$ is a square matrix with a row of zeros, and $B$ is any other square matrix (both of the same dimensions, say, $n \times n$). Then, $AB$ will have a row of zero (corresponding to the row of zeros in $A$). As such, $AB$ is not the identity matrix, hence no right inverse exists. If we let $B$ have a column of zeros, then we observe that there is no left inverse. ■

**Definition 1.9.** A sequence of **row reductions** reduce the matrix $M$ to the form

$$\begin{bmatrix} \mathbf{1} & 0 & -1 & 0 & 0 & a \\ & \mathbf{1} & c & 0 & 0 & b \\ & & & \mathbf{1} & 0 & d \\ & & & & \mathbf{1} & e \end{bmatrix}$$

which we call **row reduced echelon form**.

**Definition 1.10.** A **homogeneous linear equation** $AX = [0]$ always admit a trivial solution, i.e. $X = 0$. This notation is faulty, because $X$ is a column vector.

**Corollary 1.11.** Observe from the definition of RREF form that if there are more unknowns than there are equations, then the homogeneous equation $AX = 0$ admits a nontrivial solution.

*Proof.* Contained in textbook. ■

**Theorem 1.12.** Let $A$ be a square matrix. The following conditions are equivalent:

(a) $A$ can be reduced to the identity by a sequence of elementary row operations.
(b) $A$ is a product of elementar matrices.
(c) $A$ is invertible.

*Proof.* Omitted. ■

**Definition 1.13.** Determinant is a map defined

$$\det : \mathbb{R}^n \mapsto \mathbb{R}$$

where $n \times n$ describes the size of the matrix.

It is not of interest to us to prove the formulae of computing determinants, but rather some of its algebraic properties.

**Theorem 1.14.** There is a unique function $\delta$ on the space of $n \times n$ matrices with the properties below (the formula is that using cofactor matrices):

(i) With $i$ denoting the identity matrix, $\delta(I) = 1$
(ii) $\delta$ is linear in the rows of matrix $A$
(iii) If two adjacent rows of matrix $A$ are equal, then $\delta(A) = 0$.

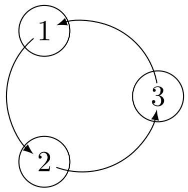
**Theorem 1.15.** For any $n \times n$ matrices $A$ and $B$, $\det(AB) = \det(A)\det(B)$.

Determinants, in general, can be difficult to work with in $n$ dimensions; we omit the rest of the materials here. Consult textbook for more details.

**Definition 1.16.** A **permutation** of a set $S$ is a bijective map $p$ from set $S$ to itself; i.e.

$$p : S \mapsto S.$$

**Remark 1.17.** A common way to represent permutation is by drawing a function table. But that gets cumbersome really quickly, because we need to always keep track of permutations on a set. Take a simple example:



This means 1 permutes to 2, 2 to 3 and 3 to 1. An easy and rather not cumbersome notation we tend to work with is the cyclic notation, which is as follows:

$$(1\,2\,3),$$

which we will work with in the remainder of this course.

**Example 1.18** (Product of Permutations)**.** Suppose we are dealing with the product of

$$(1\,4\,5\,2)[(3\,4\,1)(2\,5)].$$

The order of operations is to be read backwards. Therefore, we deal with the square bracket first, then go "forward" to the first bracket. Take the element 1 first; the permutation is as follows:

$$1 \to 3$$

so we first write (1 3. Then, starting with 3,

$$3 \to 4 \to 5$$

so we then write (1 3 5. Then, starting with 5,

$$5 \to 2 \to 1$$

which means we are done—we have 1 cycle. Therefore, we close the bracket, (1 3 5).

Then we verify if 2 and 4 are independent cycles, or jointly form a 2-cycle:

$$2 \to 5 \to 2$$

$$4 \to 1 \to 4$$

therefore, these elements form 1-cycle with itself. We denote the entire product as the set of permutations

$$(1\,3\,5)(2)(4).$$

**Proposition 1.19.** Permutations and matrices are both bijective objects. We have a **permutation matrix** defined as

$$P = \sum_i e_{p(i),i},$$

where we can switch the positions of elements by imposing $e_{i,j} = 1$.

This concludes a brief review of linear algebra. We will work largely with groups, rings, and perhaps fields in the following sections.

## 2. Groups

This is the easiest object of great complexity to study. Much of the course will be devoted to this.

**Definition 2.1.** A **law of composition** on a set $S$ is any rule for combining pairs $a, b \in S$ to get another element in $S$.

**Corollary 2.2.** A more formal definition of law of composition is

$$S \times S \to S,$$

where $S \times S$ denotes the **product set**.

**Remark 2.3.** There are many function operators we can use to denote multiplication:

$$p = ab, a \times b, a \circ b,$$

but in this course we will mostly be defining a law of composition by ourselves, so we will work with the notation $aRb$ to indicate some law of composition $R$.

Addition, of course, follows the notation $a + b$.

**Properties 2.4.**
1. With the multiplicative notation, a law of composition is **associative** if the rule
$$(ab)c = a(bc)$$
holds $\forall a, b, c \in S$.
2. A law of composition is **commutative** if the rule
$$ab = ba$$
holds $\forall a, b, c \in S$. Groups whose elements are commutative are **abelian**.

**Observation 2.5.** Note that the associative law is more fundamental, because compositions of functions is associative. Let's see this with an example: suppose there exists a set $T$, and let $f$ and $g$ be maps from $T \to T$. As such,
$$g, f \rightsquigarrow g \circ f$$
and it is easy to observe that this law of composition is associative.

**Definition 2.6.** A **multiplication table** represents the law of composition on a particular set of elements of maps $T \to T$. Note that the table is to be read this way:

$$\begin{array}{c|ccc} & & & f \\ \hline & & & \vdots \\ & & & \vdots \\ g & \cdots & \cdots & g \circ f \\ & & & \end{array}$$

**Proposition 2.7.** Let an associative law of composition be given on a set $S$. There is a unique way to define a product of $n$ elements, $a_1, \dots, a_n$, for any integer $n$. We denote this by $[a_1 \dots a_n]$, with the following properties:
- (i) The product $[a_1]$ of one element is the element itself.
- (ii) The product $[a_1\, a_2]$ of two elements is given by the law of composition.
- (iii) For any $i \in [1, n), i \in \mathbb{N}$, $[a_1 \dots a_n] = [a_1 \dots a_i][a_{i+1} \dots a_n]$.

*Proof.* We can prove this by inducting on $n$. ■

## 2.1. Group and Subgroups.

**Definition 2.8.** A **group** is a set $G$ with a law of composition, satisfying the following properties:
1. *Associativity*: $\forall a, b, c \in G$, $(ab)c = a(bc)$.
2. *Identity*: there is an identity element 1 or $e$,such that $ea = a$, $a = ea$, $\forall a \in G$.
3. *Inverse*: $\forall a \in G$, then there exists an element $b$ such that $ab = ba = 1$.

**Definition 2.9.** A subset $H$ of group $G$ is a **subgroup** if the following properties are satisfied:
1. *Closure*: $a, b \in H$ then $ab \in H$.
2. *Identity*: $1 \in H$.
3. *Inverse*: $a \in H \implies a^{-1} \in H$.

It does not make sense to define subgroups without specifying the types of groups that we will be working with. In particular, we will examine a few types of groups in this class. The first is *symmetric group*, ie. the group with composition operation that allows one to transpose/permute elements. We denote *the symmetric group of order*[1] *$n$ as $S_n$*.

**Remark 2.10.** $S_n$ is nonabelian for $n \geq 3$. This is perhaps why $S_3$ is so interesting to study: any composition of elements in $S_3$ can be tricky.

To make studying $S_3$ easier, the multiplication table of $S_3$ is as follows:

| $\circ$ | $e$ | $(123)$ | $(321)$ | $(12)$ | $(23)$ | $(13)$ |
|---|---|---|---|---|---|---|
| $e$ | $e$ | $(123)$ | $(321)$ | $(12)$ | $(23)$ | $(13)$ |
| $(123)$ | $(123)$ | $(321)$ | $e$ | $(13)$ | $(12)$ | $(23)$ |
| $(321)$ | $(321)$ | $e$ | $(123)$ | $(23)$ | $(13)$ | $(12)$ |
| $(12)$ | $(12)$ | $(23)$ | $(13)$ | $e$ | $(123)$ | $(321)$ |
| $(23)$ | $(23)$ | $(13)$ | $(12)$ | $(132)$ | $e$ | $(123)$ |
| $(13)$ | $(13)$ | $(12)$ | $(23)$ | $(123)$ | $(321)$ | $e$ |

Some other groups we will use often are

(1) *General linear group*: $\mathrm{GL}_n(\mathbb{R}) = \{M \in \mathcal{M}_{n\times n} : \det M \neq 0\}$
(2) *Special linear group*: $\mathrm{SL}_n(\mathbb{R}) = \big\{M \in GL_n(\mathbb{R}) : \det M = 1\big\}$

Of course, we are always ready to study *number groups*; we will focus on $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{C}$.

### 2.2. Additive Group of Integers.

In general, when we refer to group of integers $\mathbb{Z}$, we refer to $(\mathbb{Z}, +)$. Why? Because we do not have multiplicative inverses in $\mathbb{Z}$ (except for 1 and $-1$); think $1/2 \notin \mathbb{Z}$. In what follows, we denote additive group of integers as $\mathbb{Z}^+$. It turns out that this is an interesting group to study, not the least of which because it is quite tractable to do so. The only theorem/result we will prove is as follows.

**Theorem 2.11.** Let $S \subset (\mathbb{Z}, +)$ be a subgroup (ie. $S \leq \mathbb{Z}$). Then *either*

(1) $S = \{0\}$
(2) $S = \mathbb{Z}\, a$, $a$ is the smallest positive integer contained in $S$.

Before we prove the theorem, we need a lemma.

**Lemma 2.12** (Division)**.** If $a \in \mathbb{Z}$, $n \in \mathbb{Z}$, and $a > 0$, then $\exists q, r \in \mathbb{Z}$ s.t.

$$n = qa + r, \quad 0 \leq r < a$$

*Proof.* Let $q$ be the largest integer such that $qa \leq n$. Let $r = n - qa \implies r \geq 0$. We now proceed to prove the second inequality by contradiction. Suppose $r \geq a$. Then, $n - qa \geq a \implies (q+1)a \leq n$, which implies that $q$ is the largest integer. Hence, $r < a$. ■

Now we're ready to prove the result.

*Proof.* Suppose $S \leq \mathbb{Z}$.

(1) Then $0 \in S$ because it is the identity element in $S$. If this is the *only element in* $S$, then $S = \{0\}$ is the trivial subgroup.

[1] To be defined later.

(2) If 0 is not the only element of $S$, then $\exists n \in S, n \neq 0$ s.t. $n > 0$ or $-n > 0$, and $n, -n \in S$. By well-ordering principle, then $S$ contains a least positive integer $n$. We claim both inclusions to prove $S = \mathbb{Z}\,a$.

$\underline{(S \subset \mathbb{Z}\,a)}$ Let $n \in S$. Then by division lemma, $\exists q, r \in \mathbb{Z}$, $0 \leq r < a$ s.t. $n = qa + r$. Then, $n + (-qa) = qa + r + (-qa) = r \in S$. Since both $n$ and $qa$ are in the set $S$, and $r < a$, then this implies $r = 0$ ($r$ is not positive). So $n = qa \in S$. Since this is true for every $n \in S$, this means $S \subset \mathbb{Z}\,a$.

$\underline{(S \supset \mathbb{Z}\,a)}$ This is slightly easier. Let $k \in \mathbb{Z}\,a$. Then $\mathbb{Z}\,a$ contins the smallest positive integer, and $ka = \underbrace{a + a + \cdots + a}_{k \text{ times}}$. Its inverse, $-ka \in S$ as well. Finally, $0 \in S \implies 0a \in S$, so $S$ is a subgroup, ie. $\mathbb{Z}\,a \subset S$.

as desired. ■

**Definition 2.13.** We call $\mathbb{Z}\,a + \mathbb{Z}\,b$ the **subgroup generated by** $a$ **and** $b$**.**

### 2.3. Cyclic Groups and Subgroups.

**Definition 2.14.** If $x \in G$ and $G$ is a group, let $n \in \mathbb{Z}_+$, and $x^n = \underbrace{x \cdot x \cdot \cdots \cdot x}_{n \text{ times}}$. Let $x^0 = e$. Then,

$$\langle x \rangle = \{\dots, x^{-3}, x^{-2}, x^{-1}, e, x, x^2, x^3, \dots\}$$

**Remark 2.15.** $e = 1$ in the case of cyclic subgroups.

**Remark 2.16.** Exponentiation in group can yield the same elements. Because we are working with *sets of elements*, we *can* omit the elements that are repeated.

**Properties 2.17.** $\forall r, s, n \in \mathbb{Z}$, then

$$\begin{aligned} x^{r+s} &= x^r x^s \\ x^{rs} &= (x^r)^s \\ (x^{-1})^n &= (x^n)^{-1} \end{aligned}$$

**Definition 2.18.** The **order of a group** $G$ is the number of elements in $G$, ie. $|G|$.

**Definition 2.19.** The **order of an element** $x \in G$ is the smallest positive integer $n$ s.t. $x^n = 1$.

**Definition 2.20.** We say a group $G$ is **cyclic** if $\exists x \in G$ s.t. $G =$ cyclic subgroup generated by $x$, ie. $G = \langle x \rangle$.

We now encounter an important proposition.

**Proposition 2.21.** Let $\langle x \rangle \leq G$, where $x \in G$. Let $S = \left\{k \in \mathbb{N} : \, x^k = e\right\}$.

(1) $S \leq \mathbb{Z}^+$.
(2) $x^r = x^s$ with $r \geq s$ are equal iff $x^{r-s} = e$, ie. iff $r - s \in S$.
(3) Suppose $S$ is not the trivial subgroup. Then $S = \mathbb{Z}\,n$ for some $n \in \mathbb{N}$, and $\left\{1, x, x^2, \dots, x^{n-1}\right\}$ are distinct elements of the subgroup $\langle x \rangle$. Obviously, the order of $\langle x \rangle$ is $n$.

*Proof.*
1. Let $k, l$ be such that $x^k = x^l = e$, then $x^k x^l = x^{k+l} = e$, hence $k, l \in S$. This shows closure. Obviously, $x^0 = e$. Finally, if $x^k = e$, then $x^{-k} = (x^k)^{-1} = e$. This shows inverse.
2. This follows from cancellation law.
3. Since all subgroups of $\mathbb{Z}^+$ are either trivial or of the form $\mathbb{Z}\,a$, $S = \mathbb{Z}\,n$ for some $n \in \mathbb{N}$. Since $k$ is arbitrary, divide $k$ by $n$, so we write $k = qn + r$ where $r$ is the remainder. Note that $x^{qn} = e^q = e$, and $x^k = x^{qn}x^r = x^r$, so $x^k$ is equal to one of the elements in $\left\{1, x, x^2, \dots, x^{n-1}\right\}$. By the previous part, these elements are all distinct, because $x^n$ is the smallest positive power equal to $e$.

This concludes the proof. ■

### 2.4. **Homomorphism.**

**Definition 2.22.** A **homomorphism** $f : G \mapsto H$, $(G, \star), (H, *)$ are groups, is defined $\forall a, b \in G$
$$f(a \star b) = f(a) * f(b).$$

**Example 2.23.** Absolute value functions, exponentiation in $\mathbb{R}$, complex numbers, linear transformations, and determinants are all examples of homomorphisms.

**Example 2.24.** There are two examples that are *always true*, for any groups $G, H$ and homomorphism $f : G \mapsto H$:
1. Take $f(g) = e_H$. This is the *trivial homomorphism.*
2. In particular, if $H \leq G$, then $i : H \mapsto G$ is called *inclusion homomorphism of G into H* if $i(h) = h$.

**Proposition 2.25** (Properties of Homomorphism)**.** For any groups $G, H$ and homomorphism $f : G \mapsto H$:
1. $\forall a_1, a_2, \dots, a_k \in G$, $k \in \mathbb{Z}_+$, $f(a_1 a_2 \dots a_k) = f(a_1)f(a_2)\dots f(a_k)$.
2. $f(e_G) = e_H$.
3. $\forall a \in G$, $f(a^{-1}) = [f(a)]^{-1}$.

*Proof.*
1. Follows from induction on $n$.
2. By the definition of homomorphism: note that
$$f(e_G)f(e_G) = f(e_G e_G) = f(e_G)$$
and by cancellation law we then have $f(e_G) = e_H$.
3. By definition, for every $a \in G$, $a^{-1} \in G$. Then, notice that
$$e_H = f(e_G) = f(a^{-1}a) = f(a^{-1})f(a)$$
premultiply $[f(a)]^{-1}$ on both sides gives the result.

This concludes the proof. ■

**Definition 2.26.** The *image of homomorphism* $f : G \mapsto H$ is defined as
$$\operatorname{im} f = \left\{h \in H :\ f(g) = h,\ \forall h \in H,\ g \in G\right\} \tag{2.1}$$

**Definition 2.27.** The *kernel of homomorphism* $f : G \mapsto H$ is defined as
$$\ker f = \left\{g \in G :\ f(g) = e_H,\ \forall g \in G\right\} \tag{2.2}$$

There are some important results to remember and prove (by first principles).

**Proposition 2.28.** Let $f : G \mapsto H$ be a homomorphism. Then,

(1) $\operatorname{im} f$ forms a *subgroup of* $H$.
(2) $\ker f$ forms a *subgroup of* $G$.

*Proof.* Check both by using the definition of subgroup: closure, identity, inverse ■

Just like with functions, homomorphisms do not necessarily need to have good properties—in particular, as maps, they do not have to be injective or surjective. However, these properties are extremely nice to have, and give rise to good properties, which we will explore in later chapters.

**Proposition 2.29.** Let $f : G \mapsto H$ be a homomorphism. Then,

(1) $f$ is *surjective/onto* $\iff \operatorname{im} f = H$.
(2) $f$ is *injective/one-to-one* $\iff \ker f = \{e_G\}$.

**Remark 2.30.** Instead of saying an injective map, we sometimes use *embedding* instead.

**Proposition 2.31.** Let $f : G \mapsto H$ be a homomorphism. Then $K = \ker f$ is a subgroup of $G$.

*Proof.* It suffices to prove that the kernel is non-empty and closed under products and inverses. ■

**Proposition 2.32.** Let $f : G \mapsto H$ be a homomorphism from $G$ to $H$ and let $K = \ker f$. Let $a, b \in G$. The following conditions are equivalent:

(1) $f(a) = f(b)$
(2) $a^{-1}b \in K$
(3) $b \in aK$
(4) $bK = aK$

*Proof.* We proceed by showing that (1) = (2), and (2) $\implies$ (4) $\implies$ (3) $\implies$ (2).

Suppose (1) is true. Then, $f(a) = f(b)$ implies, by properties of homomorphism,

$$\begin{aligned} f^{-1}(a)f(a) &= f^{-1}(a)f(b) \\ \implies f(a^{-1}a) &= f(a^{-1}b) \\ \implies f(e_G) &= f(a^{-1}b) \\ \implies e_H &= f(a^{-1}b) \end{aligned}$$

and the definition of kernel is

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

so the last line can be written as (2).

Of course, (2) is equivalent to (1). We can rewrite the statement as

$$a^{-1}b \in \{g \in G \mid f(g) = e_H\}$$

and tracing back our steps yield the calculations

$$\begin{aligned} &\implies f(a^{-1}b) = e_H = f(e_G) \\ &\implies a^{-1}b \in e_G \\ &\implies b \in a \\ &\therefore f(b) = f(a) \end{aligned}$$

Let's show that (2) $\implies$ (4). We use the fact that (1) = (2) established above, and make a couple of computations:

$$\forall x \in K, \quad bx = a\underbrace{a^{-1}b}_{\in K}x = ax \in aK$$

so $bK \subset aK$. To show the other inclusion, we say that

$$\forall x \in K, \quad ax = b \quad \underbrace{b^{-1}a}_{\text{inverse of } a^{-1}b} \quad x = bx \in bK$$

and we know such inverse must exist by property of (2), and must be an element of $K$. Hence, $aK \subset bK$.

Now, we need to show that (4) $\implies$ (3). We need to note that $k^{-1}k'$, $k, k' \in K$ is *not* the identity element. If we take $k, k'$ s.t.

$$bk = ak', \quad \forall k, k' \in K,$$

then we can say

$$b = a\underbrace{k'k^{-1}}_{\in K} = a\bar{k}, \quad \forall \bar{k} \in K$$

hence this can be written as $b \in aK$.

Finally, we just need to show that (3) $\implies$ (2). This is easy; by definition,

$$\begin{aligned} b = ak \quad \forall k \in K \\ \implies a^{-1}b = k \quad \forall k \in K \\ \therefore a^{-1}b \in K \end{aligned}$$

as desired. ■

We will circle back to this proposition once we learn more about cosets.

2.5. **Isomorphism and Normal Subgroups.** It might make no sense (at first) to talk about these two ideas in conjunction, but we hope that this will be a little clearer. We need to define a couple of things.

**Definition 2.33.** Let $G$ be a group. Then $\forall a, g \in G$, the **conjugate of $a$ by $g$** is given by $gag^{-1}$.

**Definition 2.34.** If $N$ is a **normal subgroup** of $G$, denoted $N \lhd G$, then $\forall a \in N, g \in G$, $gag^{-1} \in N$.

**Definition 2.35.** Let $f : G \mapsto H$ be a *bijective* homomorphism. Then $f$ is an **isomorphism**.

**Proposition 2.36.** If $f : G \mapsto H$ is an isomorphism, then $f^{-1} : H \mapsto G$ is also an isomorphism.

*Proof.* The proof is constructive and pedagogically friendly. Take any $a, b \in H$. We need to verify two things:

Bijectivity: $f^{-1}(a) = g \ \forall g \in G$, meaning the inverse of a bijective map is bijective.

Homomorphism: Check if $f^{-1}(ab) = f^{-1}(a)f^{-1}(b)$. We perform the following calculations; let

$$x = f^{-1}(a)$$
$$y = f^{-1}(b)$$
$$z = f^{-1}(ab)$$

then $a = f(x)$, $b = f(y)$, $ab = f(z)$ (this operation is well-defined because we have a bijective homomorphism). Hence,

$$f(xy) = f(x)f(y) = ab = f(z)$$

using injectivity, $xy = z$, hence $f^{-1}(ab) = f^{-1}(a)f^{-1}(b)$.

This concludes the proof. ■

**Definition 2.37.** If $G$ and $H$ are **isomorphic**, then there exists an isomorphism $f : G \mapsto H$ (and the other direction is also defined). Notation-wise, we write $G \approx H$, or $f : G \overset{\approx}{\longmapsto} H$.

**Remark 2.38.** If $G$ and $H$ are groups, and are isomorphic (to each other), then all group theoretic properties are the same. In other words, these include, but not limited to,

1. $|G| = |H|$
2. $G$ cyclic $\iff$ $H$ cyclic
3. $G$ abelian $\iff$ $H$ abelian
4. $K$ is a subgroup (normal/cyclic) and $f : G \overset{\approx}{\longmapsto} H$, then $f(K)$ is a subgroup (normal/cyclic) of $H$.

**Proposition 2.39.** If $G$ and $H$ are cyclic groups, and $|G| = |H|$, then $G$ and $H$ are isomorphic.

Why do we care about isomorphisms? One of the motivational problems to the study of isomorphism is to classify different looking groups, but which are actually the same.

**Definition 2.40.** If $G$ is a group, then the set of groups that are isomorphic to $G$ forms the **isomorphism class of** $G$.

**Definition 2.41.** An isomorphism from (group) $G$ to $G$ is called an **automorphism**.

**Remark 2.42.** The identity map is an automorphism, but it is often *not* the only one. In particular, for all nonabelian groups, there always exists a nontrivial automorphism. In particular, conjugation is an automorphism.

**Theorem 2.43.** Let $\gamma_x(g) = xgx^{-1}$ (a map $\gamma_x : G \mapsto G$). This map is an automorphism.

*Proof.* This map is obviously a homomorphism. Equally obviously, $h = xgx^{-1} \iff g = x^{-1}hx$, so an inverse $\gamma_{x^{-1}}$ exists! As such, $\gamma_x$ is an isomorphism from itself to itself. ■

**Remark 2.44.** An *injective homomorphism* is called an *embedding*, denoted $f : G \hookrightarrow H$. Explicitly, this means $f$ is an isomorphism defined $f : G \to f(G)$, where $f(G) \subset H$.

**Proposition 2.45.** Let $\varphi : G \mapsto H$ be a group homomorphism. Then $\ker \varphi$ is a normal subgroup of $G$.

*Proof.* We know that $\ker \varphi \leq G$. Let $k \in \ker \varphi$, and $x \in G$. Then,

$$\begin{aligned}\varphi(xkx^{-1}) &= \varphi(x)\varphi(k)\varphi(x^{-1}) \\ &= \varphi(x)\varphi(x^{-1}) \\ &= \varphi(x)[\varphi(x)]^{-1} \\ &= e_H\end{aligned}$$

so $xkx^{-1} \in \ker \varphi$. Hence, $\ker \varphi \lhd G$. ■

### 2.6. **Equivalence Relation and Partitions.**

**Definition 2.46.** An **equivalence relation** $\sim$ is a relation that satisfies three properties:

(1) *Reflexive*: $\forall a \in S$, $a \in a$
(2) *Symmetric*: $\forall a, b \in S$, if $a \sim b$, then $b \sim a$
(3) *Transitive*: $\forall a, b, c \in S$, if $a \sim b$ and $b \sim c$, then $a \sim c$.

**Remark 2.47.** Group isomorphism is an *equivalence relation.*

**Example 2.48.** Take $x, y \in \mathbb{Z}$. Let $\sim$ be a relation defined

$$x \sim y \iff x - 2y = 2n, \ \ n \in \mathbb{Z}$$

then $\sim$ is an equivalence relation.

(1) *Reflexive*: $\forall a \in S$, $a \in a$. In particular, $x - x = 0 = 2(0)$.
(2) *Symmetric*: $\forall a, b \in S$, if $a \sim b$, then $b \sim a$. In particular, $x - y = 2n \implies y - x = -2n$. Hence, $x \sim y \implies y \sim x$.
(3) *Transitive*: $\forall a, b, c \in S$, if $a \sim b$ and $b \sim c$, then $a \sim c$. In particular,

$$x \sim z = (x - y) + (y - z) = x - z \quad (= 2n + 2n = 4n \in 2\,\mathbb{Z})$$

We now prove a really important theorem in equivalence relation, vis-à-vis *partitions.*

**Definition 2.49.** A **partition** of set $S$ is a way of subdividing $S$ into nonempty, nonoverlapping subsets. In particular,

$$S = S_1 \cup S_2 \cup \cdots \cup S_n$$

where every element of $S$ is contained in exactly one subset $S_i$.

**Theorem 2.50.** Equivalence relation on set $S$ and partitions on $S$ are one-to-one correspondent.

We prove this in two propositions.

**Proposition 2.51.** Every partition of set $S$ corresponds to an equivalence relation.

*Proof.* Let $S = S_1 \cup S_2 \cup \ldots$. Let $\sim$ be an equivalence relation on $S$ s.t. $\forall x, y \in S$, $x \sim y \iff \exists i \in \mathbb{N}$ s.t. $x, y \in S_i$.

(1) *Reflexive*: $\forall a \in S$, $a \in a$. In particular, $x, x \in S$.
(2) *Symmetric*: $\forall a, b \in S$, if $a \sim b$, then $b \sim a$. In particular, $\exists i \in \mathbb{N}$ s.t. $x, y \in S_i$. Then $y \sim x$ follows.
(3) *Transitive*: $\forall a, b, c \in S$, if $a \sim b$ and $b \sim c$, then $a \sim c$. In particular, $a, b \in S_i$, $b, c \in S_j$. But $b \in S_i \cap S_j$. Since we imposed that $i \neq j$, and partitions by definition are nonoverlapping, $S_i = S_j$. Hence, $x, y, z \in S_i$, so $x \sim z$.

so partitions correspond to an equivalence relation. ■

**Proposition 2.52.** Every equivalence relation defined on $S$ induces a partition on $S$.

We will need the following definition to prove the proposition directly above.

**Definition 2.53.** If $\sim$ is an equivalence relation on $S$, and if $a \in S$, then we define the **equivalence classes** of $a$ to be the set

$$E_a = \{b \in S : \ b \sim a\}$$

*Proof.* In particular, we want to show that every equivalence relation defined on $S$ induces equivalence classes on $S$, and these are partitions on $S$. We need to prove three things:

(1) $\underline{\bigcup_{i \in \mathbb{N}} E_i = S}$. This is easy to show, since, by reflexivity, $a \in E_a$ is guaranteed (hence each equivalence class is nonempty). And, obviously, $\bigcup_{a \in S} [a] \in S$
(2) Nonempty. Explained above.
(3) Nonoverlapping. We show the contrapositive: that if $E_a$ and $E_b$ have an element in common, then $E_a = E_b$. In particular, take $c \in E_a \cap E_b$. Then $c \sim a$ and $c \sim b$, then $a \sim b$. This means that $\forall x \in E_a$, then by definition, $x \sim a$ and $a \sim b$ implies $x \sim b$ by transitivity; $x \in E_b$, so $E_a \subset E_b$.

    In the same way, we say that $\forall y \in E_b$, $y \sim b$, and $b \sim a$ implies $y \sim a$, by transitivity. This means that $y \in E_a$, so $E_b \subset E_a$.

So we've proven both inclusions, hence equivalence classes form partitions on $S$. ■

**Definition 2.54** (Equivalence classes of modulo $\sim$)**.** If $\sim$ is an equivalence relation on $S$, we let $S/\sim$ or $\bar{S}$ (which reads $S \mod \sim$) be the set of equivalence classes of $\sim$.

**Example 2.55.** Define $\sim$ as the integer relation $x \sim y \iff x - y \equiv 0 \mod 2$. Then this defines a set of equivalence classes on the integers, denoted $\overline{\mathbb{Z}}$. In particular, it admits elements of both parity, ie.

$$\overline{\mathbb{Z}} = \{\{2n\}, \{2n+1\}\} \quad \forall n \in \mathbb{Z}$$

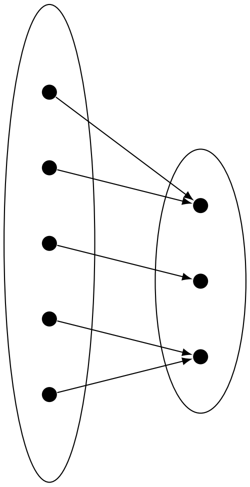
but this notation needs some refinement: it is a little tedious to refer to sets of sets. If we want to emphasise a particular element of the set $E$—to refer to a set as an element—we use the notation $[A]$. In particular, we can say that

$$\overline{\mathbb{Z}} = \{[2n], [2n+1]\} \quad \forall n \in \mathbb{Z}$$

**Remark 2.56.** There are generally a lot of ways to represent the same set of equivalence classes. For example,

$$\overline{\mathbb{Z}} = \{\overline{0}, \overline{13}\} = \{\overline{391}, \overline{24}\} = \dots$$

**Example 2.57.** Let us illustrate a particular map.



Define this map $f : A \mapsto B$ as an equivalence relation on $A$ s.t. $x \sim y \iff f(x) = f(y)$. We skip the verification here. We want to focus on the characterisation of equivalence classes:

$$E_s = \big\{ r \in A : \ f(r) = f(s) \big\}$$

which implies that the *pre-images of* $f(s)$, $f^{-1}(\{f(s)\})$, is the set of elements of the equivalence class of $s$. So

$$\overline{A} = \{f^{-1}(\{f(a)\}), f^{-1}(\{f(c)\}), f^{-1}(\{f(d)\})\}$$

which is not *the representation of the entire set A*.

**Definition 2.58.** A set of preimages of such $f : A \to B$ is called the **fiber** of $f$.

Below is an important proposition for the class.

**Proposition 2.59.** There exists a bijective correspondence between *equivalence classes* and the *image* of a map. In other words, if $f : A \mapsto B$ and $\sim$ as defined (an equivalence relation), then $\exists \varphi : \overline{A} \mapsto \operatorname{im} f$, where $\varphi$ is a bijection.

*Proof.* If $x \in \operatorname{im} f$ and $x \xmapsto{\psi} [f^{-1}(\{x\})]$, then we can define $[E] \xmapsto{\varphi} f(E)$. Alternatively, we can always write $[E] = \overline{a}$ for some $a \in A$, and check that $\psi(a) = \varphi(\overline{a})$. ■

2.7. **Cosets, Congruences, and Modulo Arithmetic.** A particularly useful objective that is used all the time in the coset.

**Definition 2.60.** If $H \leq G$, and $a \in G$, the set

$$aH = \{b \in G \mid b = ah \text{ for some } h \in H\} = \{ah \mid h \in H\}$$

is called the **left coset of $G$ in $H$**.

**Proposition 2.61.** If $a, b \in G$, then we say $a$ is **congruent** to $b$ ($a \equiv b$) if $b = ah$ for some $h \in H$.

*Proof.* We need to show that this is an equivalence relation.

(1) *Reflexive*: if $a \equiv a$, and $a = ah \implies h = e$.
(2) *Symmetric*: if $a \equiv b$ then $b = ah$. Then, $bh^{-1} = a$ and $h^{-1} \in H$, so $b \equiv a$.
(3) *Transitive*: if $a \equiv b \implies b = ah_1$ and $b \equiv c \implies c = bh_2$, then $c = bh_2 = a(h_1h_2) \implies a \equiv c$.

And the equivalence classes defined by such equivalence relation is

$$\begin{aligned} E_a &= \{b \in G \mid a \equiv b\} \\ &= \{b \in G \mid b = ah \text{ for some } h \in H\} \\ &= \{ah \in G \mid h \in H\} \end{aligned}$$

which is the left coset of $G$ in $H$. Therefore, the equivalence classes induced by congruence (which, as we checked, is an equivalence relation) are *left cosets*. ■

**Proposition 2.62.** Let $G$ be a group, and $H \leq G$. Then, the left cosets of $H$ form partitions of $G$.

**Proposition 2.63.** The cardinalities (orders) of any two cosets are equal.

*Proof.* We can construct bijection $\varphi$ between cosets. Let $a \in G$. Then the map $\varphi_a : H \mapsto aH$ where $\varphi_a(h) = ah$ has an inverse. It is defined by $(\varphi_a)^{-1}(g) = a^{-1}g \ \forall g \in aH$ (remember $(\varphi_a)^{-1} = \varphi_{a^{-1}}$ is defined as $(\varphi_a)^{-1} : aH \mapsto H$). As such, $\varphi_a(h)$ is a bijection, hence $|H| = |aH|$. ■

This lets us establish the following proposition, which we call the counting formula.

**Proposition 2.64** (Counting formula)**.** Let $H \leq G$. Since $|H| = |aH|$, we can use the order of cosets of $H$ in $G$ to find the order of $G$. In particular,

$$|G| = |H| \,(\text{number of cosets of } H \text{ in } G)$$

**Definition 2.65.** The number of cosets of $H$ in $G$ is called the **index of $H$ in $G$.**

**Theorem 2.66** (Lagrange's Theorem)**.** The order of any subgroup of group $G$ divides the order of $G$.

*Proof.* Already established above. This is the direct consequence of the counting formula:

$$|G| = |H| \, [G : H]$$

where $[G : H]$ is the number of (left) cosets of $H$ in $G$. ■

**Remark 2.67.** The reverse of Theorem 2.66 does not work. Take $S_4$, the symmetric group of order 4. Just because it has 24 elements *does not mean* there is a subgroup of order 12. There can be divisors of $G$ that are *not* the orders of subgroups of $G$.

**Proposition 2.68.** Let $G$ be a group of order $p$, where $p > 1$ is a prime number. Then $G$ is cyclic.

*Proof.* Elements in such group $G$ has either order 1 or $p$. So, $|G| \geq 2$ because 1 is not a prime number. Therefore, $\exists$ a non-identity element $x \in G$. Hence, $\big|\langle x \rangle\big| = p$, and $G = \langle x \rangle$. ■

A particularly powerful interpretation comes when we talk about cosets in conjunction with equivalence classes. We note that the counting formula can be used to identify the following fact: that the left cosets of $\ker\varphi$ are the nonempty fibres of the map $\varphi$. Hence, there exists abijective correspondence between fibres and the elements of the image. This implies

$$[G : \ker\varphi] = |\operatorname{im}\varphi|$$

A few results follow immediately.

**Corollary 2.69.** Let $\varphi : G \to \mathcal{G}$ be a homomorphism between finite groups. Then,

(1) $|G| = |\ker\varphi||\operatorname{im}\varphi|$,
(2) $|\ker\varphi|$ divides $|G|$,
(3) $|\operatorname{im}\varphi|$ divides $|G|$ and $|\mathcal{G}|$.

*Proof.* These proofs are elementary. We omit them here. ■

**Remark 2.70.** By definition, equivalence classes are fibers fo $f$, which are sets $f^{-1}(\{y\})$, $y \in \operatorname{im} f$. There exists a bijection between cosets and fibers. In particular, we say that if $f : G \mapsto H$ is a homomorphism, and $K = \ker f$, then $|G| = |\ker f\ |\ |\operatorname{im} f|$.

Another observation from this is $|\operatorname{im} f|$ divides $|G|$ (meaning $|\operatorname{im} f|$ is a factor of $|G|$). Furthermore, since $\operatorname{im} f \leq H$, this implies $|\operatorname{im} f|$ divides $|H|$.

**Remark 2.71.** If $\gcd(|G|, |H|) = 1$, then $|\operatorname{im} f\ | = 1$. This implies $\operatorname{im} f \subset H$, so $\operatorname{im} f$ is a *trivial subgroup of $H$*.

Now, there is a particularly useful characterisation of index. We call this the *multiplication index theorem.*

**Theorem 2.72** (Multiplication index theorem)**.** Let $K \leq H \leq G$ (as groups; as sets, $K \subset H \subset G$). Then $[G : K] = [G : H][H : K]$.

*Proof.* The idea is as follows. If $H \leq G$, and $[G : H] = 2$, then we can say that a coset takes up "half" of $G$. Of course, this is not well-defined if we are talking about infinite groups, so we refrain from doing so. In the finite case (ie. if orders of the groups are finite), this follows from applying the counting formula.

In general, we prove the theorem as follows. Let $m := [G : H]$ and $n := [H : K]$. Suppose $m, n$ are finite. Then by definition of cosets, we can write both

$$\begin{aligned} G &= g_1H \cup g_2H \cup \dots g_mH \\ H &= h_1K \cup h_2K \cup \dots h_nK \end{aligned}$$

where $g_i \in G$ and $h_i \in H\ \forall i$. Then, $\forall g \in G$,

$$\begin{aligned} gH &= g \cup (h_1K \cup h_2K \cup \dots h_nK) \\ &= gh_1K \cup gh_2K \cup \dots gh_nK \end{aligned}$$

and we want to prove that these parts are still partitions.

We proceed by contradiction. Pick $i \neq j$ s.t. $x \in gh_iK \cap gh_jK$. Then, $g^{-1}x \in h_iK$ and $g^{-1}x \in h_jK$. But we've assumed at first that $h_iK \cap h_jK = \emptyset$. Contradiction.

Hence we can write

$$G = (g_1h_1K \cup g_1h_2K \cup \dots g_1h_nK) \cup (g_2h_1K \cup g_2h_2K \cup \dots g_2h_nK) \cup \dots \\ \dots \cup (g_mh_1K \cup g_mh_2K \cup \dots g_mh_nK)$$

so that $G$ has $m \times n$ cosets of $K$. Hence, $[G : K] = [G : H][H : K] = mn$.

We will prove the infinite case once we cover quotient groups. ■

Until now, we've worked with *left cosets*. We can and should work with right cosets.

**Definition 2.73.** If $H \leq G$, and $a \in G$, the set

$$Ha = \{b \in G \mid b = ha \text{ for some } h \in H\} = \{ha \mid h \in H\}$$

is called the **right coset of $G$ in $H$**.

**Proposition 2.74.** If $a, b \in G$, then $a, b$ are in the right coset $\iff a \in Hb \iff ab^{-1} \in H$.

**Remark 2.75.** In general, the left and right cosets of subgroup $H$ are *different*.

**Remark 2.76.** Kernels are normal subgroups of homomorphisms.

**Proposition 2.77.** $H \lhd G \iff \forall h \in H,\ hG = Gh$. To be more specific, let $H \leq G$. The following conditions are equivalent:

(1) $H$ is normal
(2) $\forall g \in G$, $gHg^{-1} = H$
(3) $\forall g \in G$, $gH = Hg$ (the sets are equal)
(4) Every left coset is a right coset

The reason why we talk about cosets is so we can apply them to studying number theory, and, later, quotient groups.

**Definition 2.78.** Let $a, b \in \mathbb{Z}$, $n \in \mathbb{Z}_{>0}$. Then, $a \equiv b \mod n \iff b - a = m(n)$ for some $m \in \mathbb{Z}_{>0}$.

**Remark 2.79.** We can use the notion of equivalence classes to perform construction of modulo classes of $a$, denoted $\overline{a}$. In particular,

$$\begin{aligned} \overline{a} &= \{b \mid b - a \in \mathbb{Z}\,n\} \\ &= \{a + k \mid k \in \mathbb{Z}\,n\} = a + \mathbb{Z}\,n \end{aligned}$$

so $\overline{a}$ is a coset of $\mathbb{Z}\,n$.

**Remark 2.80.** Note that

$$\begin{aligned} \mathbb{Z}\,10 = 10\,\mathbb{Z} &= \{\dots, -10, 0, 10, \dots\} \\ 3 + \mathbb{Z}\,10 &= \{\dots, -7, 3, 13, \dots\} \\ \implies a + \mathbb{Z}\,n &= \{\dots, a - 2n, a - n, a, a + n, a + 2n, \dots\} \end{aligned}$$

so $a + \mathbb{Z}\,n$ has $n$ many *distinct cosets*, namely,

$$a + \mathbb{Z}\,n = \{a, a + \mathbb{Z}\,1, a + \mathbb{Z}\,2, \dots, a + \mathbb{Z}(n-1)\} = \{\overline{1}, \overline{2}, \dots, \overline{n-1}\}$$

are the congruence classes of modulo $n$.

**Definition 2.81.** Operations on these cosets are as follows:

(1) $\overline{a} + \overline{b} = \overline{a+b}$
(2) $\overline{a} \cdot \overline{b} = \overline{a \cdot b}$

**Remark 2.82.** Note that this definition is not entirely "error-proof". Since there is generally no unique representation of equivalence classes, it calls for the following proposition.

**Proposition 2.83.** If $\overline{a} = \overline{a'}$ and $\overline{b} = \overline{b'}$, then $\overline{a+b} = \overline{a'+b'}$ and $\overline{ab} = \overline{a'b'}$.

*Proof.* WLOG, $\forall r, s \in \mathbb{Z}$, let

$$\begin{aligned} a' &= a + rn \\ b' &= b + sn \end{aligned}$$

then

$$\begin{aligned} a' + b' &= (a + rn) + (b + sn) \\ &= (a + b) + (r + s)n \\ \implies \overline{a' + b'} &= \overline{a + b} \end{aligned}$$

Similarly,

$$\begin{aligned} a'b' &= (a + rn)(b + sn) \\ &= (ab) + (as + br + srn)n \\ \implies \overline{a'b'} &= \overline{ab} \end{aligned}$$

as desired. ■

**Example 2.84.** We show that these operations may not be well-defined in a particular scenario. Take modulo 6. In particular, take congruence classes $\overline{1}$ and $\overline{5}$. Then, take

$$\begin{aligned} 1 \cdot 5 = 5 &\equiv 5 \mod 6 \\ 7 \cdot 5 = 35 &\equiv 5 \mod 6 \\ 13 \cdot 11 = 143 &\equiv 5 \mod 6 \end{aligned}$$

but the following operation:

$$\forall a, b \in \mathbb{Z}, \overline{a} < \overline{b} \iff a < b$$

is not well-defined; pick $13 \in \overline{1}$ and $5 \in \overline{5}$ to see why.

The reason why we defined multiplication and addition of cosets is because these operations are the same for the integers.

(1) Additive identity in modulo $n$: $\overline{0}$, which is analogous to 0 in the integers.
(2) Multiplicative identity in modulo $n$: 1, or $\mathbb{Z}/n\mathbb{Z} = 1 \mod n$.
(3) Additive inverse in modulo $n$: $\overline{-a} = -a \mod n$.

And we can use associativity in groups to show a number of results.

**Corollary 2.85.** $(\mathbb{Z}/\mathbb{Z}n, +)$ is an abelian group.

*Proof.* Show that this satisfies group properties, and commutativity. ■

**Remark 2.86.** $(\mathbb{Z}/\mathbb{Z}n, +)$ is a cyclic group generated by $\overline{1}$, with addition as operation.

2.8. **Correspondence Theorem.** Now, we can, and will, describe a particular application of all the materials we have built. We set the scene as follows. Let $\varphi : G \mapsto \mathcal{G}$ be a group homomorphism, and $H \leq G$. We can always restrict $\varphi$ to $H$, ie.

$$\varphi|_H : H \mapsto \mathcal{G}$$

in particular, this is still a homomorphism (this is not hard to check at all), with the kernel being the intersection of $\ker \varphi$ with $H$:

$$\ker(\varphi|_H) = \ker \varphi \cap H$$

These facts are clear by definition. The image of the restriction is the same as $\operatorname{im} \varphi(H)$ (ie. image of $H$ under the map $\varphi$). By Corollary 2.69, we have the obvious fact that $\left|\operatorname{im} \varphi|_H\right|$ divides both $|H|$ and $|\mathcal{G}|$; hence, if $\gcd(|H|, |\mathcal{G}|) = 1$, then it is obviously true that $\varphi(H) = \{e\}$, so $H \in \ker \varphi$.

To make the ideas in this example more concrete, we introduce the following proposition—which adds another item to the list of properties of homomorphism.

**Proposition 2.87.** Let $\varphi : G \mapsto \mathcal{G}$ be a group homomorphism. Denote its kernel by $K$. Let $\mathcal{H} \leq \mathcal{G}$. Let $\varphi^{-1}(\mathcal{H}) := H$. Then,

(1) $K \subset H \leq G$ (kernel is contained in $H$, while $H$ is a subgroup of $G$).
(2) if $\mathcal{H} \lhd \mathcal{G}$, then $H \lhd G$.
(3) if $\varphi$ is surjective, and $H \lhd G$, then $\mathcal{H} \lhd \mathcal{G}$.

*Proof.* (1) This is simple enough to check, as long as we keep in mind that $\varphi^{-1}$ is not a map: by definition, the symbolic $\varphi^{-1}(\mathcal{H}) = H$ is the preimage of $\mathcal{H}$. Obviously, if $x \in \ker \varphi$, then $\varphi(x) = e$. Since $e \in \mathcal{H}$, $x \in H$, then it follows that $H \supset K$ (since $K \leq H$). We leave the subgroup verification as an exercise.

(2) Let $\mathcal{H} \lhd \mathcal{G}$. Pick $h \in H$ and $g \in G$. Note that the image of the conjugate of $h$ by $g$ is $\varphi(ghg^{-1}) = \varphi(g)\varphi(h)\varphi(g)^{-1} = \varphi(h)$, ie. the conjugation results in the image of $h$. Obviously, by construction, $\varphi(h) \in \mathcal{H}$. Since $\mathcal{H}$ is normal, we then have $\varphi(ghg^{-1}) \in \mathcal{H}$, implying $ghg^{-1} \in H$.

(3) This should be intuitively true. To see why this is the case, pick $a \in \mathcal{H}$ and $b \in \mathcal{G}$. Then, $a = \varphi(x)$ and $b = \varphi(y)$ is guaranteed for some $x \in H$ and $y \in G$, by surjectivity of the map $\varphi$. In particular, since $H$ is normal, $yxy^{-1} \in H$, so $\varphi(yxy^{-1}) = \varphi(y)\varphi(x)\varphi(y)^{-1} = bab^{-1} \in \mathcal{H}$.

This concludes the proof. ■

The following is a profound connection between subgroups of domain and range of a group homomorphism.

**Theorem 2.88** (Correspondence Theorem)**.** Let $\varphi : G \mapsto \mathcal{G}$ be a surjective group homomorphism. Denote its kernel as $K$. There exists a bijective correspondence between subgroups of $\mathcal{G}$ and subgroups of $G$ that contains $K$. We define the correspondence as

$$\text{a subgroup } H \text{ of } G \text{ that contains } K \rightsquigarrow \text{its image } \varphi(H) \in \mathcal{G}$$
$$\text{a subgroup } \mathcal{H} \text{ of } \mathcal{G} \rightsquigarrow \text{its preimage } \varphi^{-1}(\mathcal{H}) \in G$$

then,

(1) if $H$ and $\mathcal{H}$ are corresponding subgroups, then $H \lhd G \iff \mathcal{H} \lhd \mathcal{G}$.

(2) if $H$ and $\mathcal{H}$ are corresponding subgroups, then $|H| = |\mathcal{H}||K|$.

2.9. **Quotient Groups.** With the materials we have built, we can construct *quotient groups.*

**Definition 2.89.** Let $N \lhd G$. Furthermore, let $G/N = \{$ cosets of $N$ in $G\}$. Let $x \in G/N$, which is represented as either $\overline{a}$ or $[aN]$. We can define a map $\pi : G \mapsto G/N$ where $\pi(a) = \overline{a}$, $\forall a \in G$.

**Theorem 2.90.** There exists a law of composition on $G/N$ that makes it a group, with the properties that $\pi$ is a *surjective homomorphism* with $\ker \pi = N$.

The fact that $\pi(a)$ is surjective shouldn't be surprising; a minute or two of staring should make this apparent. But we need to find such law of composition to make the theorem hold. We need a few results.

**Definition 2.91.** If $A, B \leq G$, then we define $A \times B$ as the **product set**

$$A \times B = \big\{(a \times b) : \ a \in A, \ b \in B\big\}$$

**Remark 2.92.** The product set is *generally not a subgroup of $A$ or $B$ unless there are some "good" properties with the subgroups.*

**Lemma 2.93.** If $N \lhd G$, and if $a, b \in G$, then $(aN)(bN) = (ab)(N)$.

*Proof.* If $N$ is normal, then we have that $aN = Na$, $\forall a \in G$. Then,

$$\begin{aligned}(aN)(bN) &= (aN)(Nb) \\ &= a(Nb) = a(bN) = (ab)N\end{aligned}$$

as desired. ■

**Definition 2.94.** Let the law of composition on $G/N$ be such that $\forall [c_1], [c_2] \in G/N$, $[c_1c_2] = [c_1][c_2]$. By lemma, $c_1c_2$ is in the coset of $N$, and $[c_1c_2]$ is an element of $G/N$.

*Check: $G/N$ equipped with Definition 2.94 forms a group.* We check all the conditions.

(1) Associativity: $\forall a, b, c \in G/N$,

$$\overline{a}\overline{b}\overline{c} = \overline{ab}\overline{c} = \overline{(ab)c} = \overline{abc}$$

(2) Identity: $\overline{1}$ is obviously the identity element.

(3) Inverse: recall that $\overline{a} \cdot \overline{a^{-1}} = \overline{a \cdot a^{-1}} = \overline{1}$ and $\overline{a^{-1}} \cdot \overline{a} = \overline{a^{-1} \cdot a} = \overline{1}$.

This concludes the proof. ■

**Lemma 2.95.** $\pi : G \mapsto G/N$ is a surjective homomorphism with $K = \ker \pi = N$.

*Proof.* First, $\pi : G \mapsto G/N$ is indeed a homomorphism (check). Surjectivity requires us to check the image of the homomorphism, ie.

$$\begin{aligned}\operatorname{im} \pi &= \big\{x \in G/N : \ \exists g \in G \ \text{ s.t. } \ \pi(g) = x\big\} \\ &= \big\{x : \ x \in G/N\big\} = G/N\end{aligned}$$

Also,

$$\ker \pi = \Big\{y \in G : \ \pi(y) = e_{G/N}\Big\}$$

$$= \{y : y \in N\} = N$$

as desired. ■

With this, we can talk about a lot about integer groups—in particular, as we said before, every subgroup of the form $\mathbb{Z}n$, $n \in Z$, is normal in $(\mathbb{Z}, +)$. Therefore, we can define $G/N = \mathbb{Z}/\mathbb{Z}n = \{\overline{0}, \overline{1}, \ldots, \overline{n-1}\}$. By Theorem 2.90, we can define $\varphi : \mathbb{Z} \mapsto \mathbb{Z}/n$ where $\varphi(a) = \overline{a} = \overline{a + \mathbb{Z}n}$[2]. This is indeed a surjective homomorphism.

Let us see why normal subgroup is a really important criterion for Theorem 2.90 to hold.

**Example 2.96.** Let $G = (S_3, \circ)$. Define generators $x = (123)$ and $y = (12)$, and $\langle x \rangle$ is a normal subgroup of $G$ (check). Let us find the cosets (partitions) of the this generator: since $y$ is of order 2, $\langle x \rangle$ is index 2 in $G$. The cosets are

$$\begin{aligned} 1\langle x \rangle &= \{1, x, x^2\} = \langle x \rangle 1 \\ y\langle x \rangle &= \{y, yx, yx^2\} = \{y, x^2y, xy\} \\ &= \langle x \rangle y \end{aligned}$$

so $\langle x \rangle$ is indeed a normal subgroup of $G$. The quotient group $G/N = S_3/\langle x \rangle = \{1\langle x \rangle, y\langle x \rangle\} = \{\overline{1}, \overline{y}\}$. In particular, because the order of the group is prime, we can use Proposition 2.68 to say that $S_3/\langle x \rangle \cong C_2$.

We now check that $\langle y \rangle$ is *not* a normal subgroup of $G$. Just use the fact that if $\langle y \rangle \lhd S_3$, then $\forall g \in G$, $gyg^{-1} \in \langle y \rangle$. But, take $x = (123)$ as defined above

$$\begin{aligned} xyx^{-1} &= (x(1)x(2)) \\ &= (23) \notin \langle y \rangle \end{aligned}$$

and, take $S_3/\langle y \rangle = \{[1, y], [x, xy], [x^2, x^2y]\} = \{\overline{1}, \overline{x}, \overline{x}^2\}$:

$$\begin{aligned} \{1, y\} \circ \{x, xy\} &= \{1x, 1xy, yx, yxy\} \\ &= \{x, xy, x^2y, x^2\} \end{aligned}$$

which is *not* a coset. Hence, the law of composition does not make $S_3/\langle y \rangle$ a group.

**Definition 2.97.** The **canonical map** $\pi : G \mapsto G/N$, where $N \lhd G$, is a surjective homomorphism.

Now, we need to explore a full example to arrive at a really important theorem.

**Example 2.98.** We can easily prove that for some homomorphism $f : G \mapsto H$, $K = \ker f \lhd G$. We can therefore construct a map $\beta : G \mapsto G/K$. In particular, we want to use this property to construct the example

$$\text{abs} : \mathbb{C}^\times \mapsto \mathbb{R}^\times, \quad \text{abs}(z) = |z|$$

where $|xy| = |x||y|$, $\forall x, y \in \mathbb{C}$. This is indeed a homomorphism: define $x = re^{i\theta}$ and $y = se^{i\tau}$. Then,

$$\text{abs}(xy) = |xy| = rse^{i(\theta+\tau)} = \text{abs}(x)\text{abs}(y)$$

[2] An equivalent notation I use is $[a]$.

and $K = \ker(\text{abs}) = \Big\{g \in \mathbb{C} : \text{abs}(g) = 1\Big\}$ = the unit circle. Therefore, the cosets $nK$, $n > 0$, and $n \in \mathbb{C}$, ie. $\Big\{\overline{nK} : n \in \mathbb{C}\Big\}$ is the set of all cosets.

Then, we can define

$$\mathbb{C}^\times /K = \{rK : r > 0, r \in \mathbb{R}\}$$

with law of composition satisfying $\overline{r} \cdot \overline{s} = [rK][sK] = \overline{rs}$. This is indeed a *surjective homomorphism*. Then, $\forall r, s \in \mathbb{R}_{>0}$ and $\forall z \in \mathbb{C}^\times$, then

$$\begin{aligned} \text{abs} : \mathbb{C}^\times \mapsto \mathbb{R}^\times, \quad \text{abs}(z) = &|z| \\ \overline{r} : [rK] \mapsto r& \\ \implies \overline{\text{abs}} : \overline{z} \mapsto &|z| \end{aligned}$$

is an *isomorphism*. It is a homomorphism, which we've checked many times by now. Let us see if it is bijective:

(a) *Injective*: $\ker(\overline{\text{abs}}) = \Big\{\overline{z} : \text{abs}(\overline{z}) = 1\Big\} = \{\overline{z} : \overline{z} = 1\} = \{1\}$.
(b) *Surjective*: $\text{im}(\overline{\text{abs}}) = \Big\{y \in R : \exists z \in \mathbb{C} \text{ s.t. } \overline{r}(z) = y\Big\} = \mathbb{R}$

This is a special case of the *First Isomorphism Theorem.*

**Theorem 2.99** (First Isomorphism Theorem)**.** Let $\nu : G \mapsto H$ be a *surjective homomorphism.* Let $K = \ker \nu$. Since $K \lhd G$, $G/K \cong H$. In fact, there exists an isomorphism $\overline{\nu} : G/K \mapsto H$ such that

$$(*) \qquad \overline{\nu}(\overline{g}) = \nu(g), \quad \forall g \in G$$

*Proof.* So we need to prove that there *exists* a map $\nu : G \mapsto H$ such that $(*)$ holds true. It suffices to show that $\forall a, b \in G$ s.t. $\overline{a} = \overline{b}$ then $\nu(a) = \nu(b)$.

If $\overline{a} = \overline{b}$, then $b \in aK$. Then, there exists $k \in K$ s.t. $b = ak$. So

$$\nu(b) = \nu(ak) = \nu(a)\nu(k) = \nu(a)$$

Then we define $\overline{\nu}$ as follows: $\forall [C] \in G/K$, choose an element $a \in [C]$ where $\overline{\nu}([C]) = \nu(a)$. We can check that this is a homomorphism (which satisfies $(*)$), and that it is bijective:

(a) *Injective*: $\ker(\overline{\nu}) = \Big\{\overline{a} : \overline{\nu}(\overline{a}) = 1\Big\} = \{\overline{a} : a \in K\} = \Big\{[K]\Big\} = \{1\}$.
(b) *Surjective*: $\forall h \in H$, $\exists g \in G$ s.t. $\overline{\nu}(\overline{g}) = \nu(h) = h \implies \nu(g) = h$.

as desired. ■

This concludes the brief introduction to finite group theory. We will talk more about group theory in a few chapters.

## 3. Linear Algebra

Linear algebra over $\mathbb{R}$ constituted an introductory undergraduate course, and there were a lot to talk about. It is now of interest to study linear algebra over algebraic structures.

**Definition 3.1.** A **field** is a set $\mathcal{F}$ equipped with *two* laws of composition

$$\begin{cases} + & : & \mathcal{F} \times \mathcal{F} \mapsto \mathcal{F} \\ \times & : & \mathcal{F} \times \mathcal{F} \mapsto \mathcal{F} \end{cases}$$

where *a lot* properties are satisfied:

(1) $+$ and $\times$ are *associative* and *commutative*
(2) *Distributive*: $\forall a, b, c \in \mathcal{F}$, $a \times (b + c) = a \times b + a \times c$
(3) There is an additive identity, denoted 0, and multiplicative identity, denoted 1, such that $0 \neq 1$
(4) $\forall a \in \mathcal{F}$, $-a$ is the additive inverse; if $a \neq 0$, then $a^{-1} = \frac{1}{a}$ is the multiplicative inverse
(5) $\forall a \in \mathcal{F}$, $0 \times a = a \times 0 = 0$

Equivalently, the conditions above can also be restated as

(1) $(\mathcal{F}, +)$ forms an abelian group with identity 0
(2) $(\mathcal{F} \setminus \{0\}, \times)$ is an abelian group with identity 1
(3) Distributive property holds: $\forall a, b, c \in \mathcal{F}$, $a \times (b + c) = a \times b + a \times c$

A natural inclination is to define a **subfield**.

**Definition 3.2.** A **subfield** of field $\mathcal{F}$ is a subset $F \subset \mathcal{F}$ s.t. $1 \in F$, and $F$ is closed under $+, -, \times, \div$. This means $\forall a, b \in \mathcal{F}$, $a \pm b \in \mathcal{F}$, $a \times b \in \mathcal{F}$, $a \div b \in \mathcal{F}$.

**Example 3.3.** Let $\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$ be a polynomial (called *extension field*). Check that this is a subfield of $\mathbb{R}$.

Intuitively, $\mathbb{Q}[\sqrt{2}] \subset \mathbb{Q} \subset \mathbb{R}$, so this is indeed a subset of $\mathbb{R}$. To check if it a subfield, we need to verify the conditions. Indeed, $\forall a, b \in \mathbb{Q}[\sqrt{2}]$, closure in subtraction and addition are easy to check. For multiplication and division, if we let $x = a + b\sqrt{2}$, $y = c + d\sqrt{2}$, we compute

$$\begin{aligned} x \times y &= (a + b\sqrt{2}) \times (c + d\sqrt{2}) \\ &= (ac + 2bd) + (ad + bc)\sqrt{2} \\ \frac{x}{y} &= \frac{a + b\sqrt{2}}{c + d\sqrt{2}} \times \frac{c - d\sqrt{2}}{c - d\sqrt{2}} \\ &= \frac{(ac - 2bd) + (bc - ad)\sqrt{2}}{c - 2d^2} \end{aligned}$$

since all coefficients are rational numbers, and that operations in $\mathbb{Q}$ are closed in $\mathbb{Q}$, we have that $\mathbb{Q}[\sqrt{2}]$ is a subfield of $\mathbb{R}$.

In general, we can get at something much more interesting.

**Theorem 3.4.** Let $p$ be a prime number. Then, $\mathbb{Z} / \mathbb{Z}p$ equipped with $+$ and $\times$ under mod$p$ is a field.

*Proof.* Check the conditions by Definition 3.1. ■

We need a couple more definitions so we can better identify fields from one another—as unique as they are, we still need to be able to pinpoint finite fields from one another.

**Definition 3.5.** Let $\overline{1}$ denote the multiplicative identity, and $\overline{0}$ be the additive identity. The **characteristic** of field $\mathcal{F}$, denoted $\operatorname{char}\mathcal{F}$, is the smallest $n > 0$ s.t.

$$\underbrace{\overline{1}+\overline{1}+\cdots+\overline{1}}_{n\text{ times}}=\overline{0}$$

**Remark 3.6.** If $\overline{1}$ has finite order in a group $G$, then $\operatorname{char} 1 = |1|$ (the order of the element). Note that $\operatorname{char}\mathbb{R} = \operatorname{char}\mathbb{C} = \operatorname{char}\mathbb{Z} = 0$, while $\operatorname{char}\mathbb{Z}/\mathbb{Z}p = p$.

**Lemma 3.7.** If $\mathcal{F}$ is a finite field, then $\operatorname{char}\mathcal{F} = 0$ or $\operatorname{char}\mathcal{F} = p$, where $p$ is a prime number.

*Proof.* Let $\overline{1}$ denote the multiplicative identity, and $\overline{0}$ be the additive identity[3]. There are two cases:

(1) $\operatorname{char}\mathcal{F} = 0$: then we're done.

(2) Suppose that $\operatorname{char}\mathcal{F} \neq 0$, but $\operatorname{char}\mathcal{F} = k$, where $k$ is not a prime number. We prove that this cannot be true by contradiction. If $k$ is not a prime number, then $\exists r, s \in \mathbb{Z}$ s.t. $1 < r, s < k$, $k = rs$. Then,

$$\begin{aligned} rs &= \underbrace{(\overline{1}+\overline{1}+\cdots+\overline{1})}_{r\text{ times}}\underbrace{(\overline{1}+\overline{1}+\cdots+\overline{1})}_{s\text{ times}} \\ &= \underbrace{\overline{1}+\overline{1}+\cdots+\overline{1}}_{r+s\text{ times}} \end{aligned}$$

where second line follows from distributive property of $\mathcal{F}$ (field). Moreover, $rs = k = \overline{0}$ by definition of characteristic $k$. Since $r, s > 1$, it must be that they are multiplicative inverses of one another. WLOG, suppose $r \neq \overline{0}$, then $\exists r^{-1}$ s.t. $r^{-1}r = \overline{1}$, so

$$\begin{aligned} (r^{-1}r)s &= r^{-1}\overline{0} \\ \implies s &= \overline{0} \end{aligned}$$

which is a contradiction.

Hence, $k$ is a prime number. ■

We skip the review of linear algebra; but it is important to remember that these concepts are integral what will be developed next. Row reduction, determinants, cofactor expansion, Cramer's rule, and solving systems of linear equations (now, in $\mathbb{Z}/\mathbb{Z}n$) are all important to recall. Now, we turn to an important object from linear algebra over the real numbers.

**Definition 3.8.** A **vector space** is a set $V$ over field $\mathcal{F}$ where $\forall v, w \in V$, $\forall a, b \in \mathcal{F}$, equipped with two laws of composition

$$\begin{cases} + & : \; V \times V \mapsto V \\ \cdot & : \; \mathcal{F} \times V \mapsto V \end{cases}$$

[3] From here onwards, unless otherwise stated, these are the notations for the identity elements under different law of compositions.

such that the following conditions hold:

1. $(V, +)$ is an abelian group with identity $\vec{0}$[4]
2. $1_{\mathcal{F}} \cdot v = v$, $\forall v \in V$
3. Associativity: $(ab)v = a(bv)$
4. Distributive: $(a + b)v = av + bv$ and $a(v + w) = av + aw$

We can now define a couple of things familiar from linear algebra over $\mathbb{R}$.

**Definition 3.9.** **Dimension** of a vector space can be thought of as the number of coordinates required to describe the space (of interest).

**Theorem 3.10** (From linear algebra over $\mathbb{R}$)**.** *If $V$ is a finite dimensional vector space over $\mathbb{R}$, then the following conditions are equivalent:*

1. Maximum number of linearly independent vectors
2. Number of vectors in a basis
3. Minimum number of vectors needed to *span* the vector space $V$

This is the **dimension** of $V$.

**Remark 3.11.** We can translate this theorem to linear algebra of field $\mathcal{F}$: use elements in $\mathcal{F}$ instead of elements in $\mathbb{R}$, whenever such definition comes up.

**Definition 3.12.** A **linear combination** of $v_1, v_2, \dots, v_n \in V$ is a vector of the form

$$a_1v_1 + a_2v_2 + \cdots + a_nv_n, \quad \forall a_i \in \mathcal{F},\ i = 1, 2, \dots, n$$

**Definition 3.13.** The **span** of the vectors $\{v_1, \dots, v_n\}$, denoted $\operatorname{span}\{v_1, v_2, \dots, v_n\}$, is the linear combination of the vectors.

**Definition 3.14.** A set of vectors $\{v_1, \dots, v_n\}$ is **linearly independent** if there *does not exist* nontrivial solution to

$$a_1v_1 + a_2v_2 + \cdots + a_nv_n = 0_{\mathcal{F}}, \quad \forall a_i \in \mathcal{F},\ i = 1, 2, \dots, n$$

**Definition 3.15.** A **basis** is a set of vectors that are *linearly independent* and *span* the vector space.

Thinking in this way requires a bit of paradigm shift.

**Example 3.16.** Take $\mathcal{F} = \mathbb{Z} / \mathbb{Z}\,5$, which has 5 elements (5 distinct cosets of $(\mathbb{Z}\,5, +)$ in $(\mathbb{Z}, +)$). Take $V = (\mathbb{Z} / \mathbb{Z}\,5)^2$, which is the set

$$\left\{ \begin{bmatrix} a \\ b \end{bmatrix} \,:\, a, b \in \mathbb{Z} / \mathbb{Z}\,5 \right\}$$

has 25 elements. Note that $\dim V = 2$, because it only takes 2 “coordinates” to describe every $v \in V$. A basis of $V$ is

$$\left\{ \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$$

[4] I will only use this notation here; from here on, it will be simply 0, but one will need to be cognizant of what context this identity element is used in.

Note that, if the field of interest is $\mathbb{R}^2$, we can describe other sets of basis, too: in fact, in $\mathbb{R}^2$, as long as the two vectors of choice are *not* parallel, we will be able to describe every vector in $\mathbb{R}^2$ with those vectors.

**Remark 3.17** (Should be familiar from linear algebra over $\mathbb{R}$)**.** Choice of basis is not unique.

**Theorem 3.18** (Should be familiar from linear algebra over $\mathbb{R}$)**.** If $V$ is a vector space over scalar field $K$, and $\dim V = n$, then $V \cong \mathbb{R}^n$.

*Proof.* Let $v_1, \dots, v_n \in V$ be a basis. If we have a basis in $V$, take any $x \in V$ and $x = \sum_{i=1}^n a_i v_i$ is unique (proven as a lemma below).

Take the map

$$\varphi : \underbrace{\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}}_{\in \mathbb{R}^n} \mapsto \underbrace{\sum_{i=1}^n a_i v_i}_{\in V}$$

then $\varphi$ is

(1) **surjective**, because every $x_i \in V$ can be written as a linear combination of the basis we've chosen
(2) **injective**, requires the uniqueness of linear combination

**Lemma 3.19.** Let $\{v_1, \dots, v_n\}$ be a basis of the vector space $V$. Then, every $x \in V$ can be written as a *unique* linear combination of the basis vectors we've chosen.

*Proof.* Proceed by contradiction: suppose the linear combination is not unique. WLOG, take $x \in V$ and write them as

$$x = a_1 v_1 + \cdots + a_n v_n$$
$$x = b_1 v_1 + \cdots + b_n v_n$$

subtracting one from another yields $x = (a_1 - b_1)v_1 + (a_2 - b_2) + \cdots + (a_n - b_n)v_n$. Since we've written $x$ as a of combination of basis vectors (which are linearly independent by definition), we conclude that $a_1 - b_1 = a_2 - b_2 = \cdots = a_n - b_n = 0 \implies a_1 = b_1, a_2 = b_2, \dots, a_n = b_n$. Contradiction on non-uniqueness. ■

Returning to the proof: using the proof of Lemma 3.19, we can conclude that our map is injective. Hence, the map is bijective, resulting in an isomorphism. ■

**Theorem 3.20.** If $V$ is a vector space over field $\mathcal{F}$, and $\dim_{\mathcal{F}} V = n$, then $V \cong \mathcal{F}^n$.

*Proof.* Change every instance of $\mathbb{R}$ into $\mathcal{F}$ from the proof above. ■

**Theorem 3.21.** If $\mathcal{F}$ is a finite field, $|\mathcal{F}| = p^n$ for some $n$.

*Proof.* The point is to express field as a vector space. Since $\mathcal{F}$ is a finite field, then $\overline{1} \in \mathcal{F}$ has a finite order. In particular, that means $\operatorname{char} \mathcal{F} = p$ (see Lemma 3.7 for why), for some prime $p$. We can show that $\{\overline{0}, \overline{1}, \dots, \overline{p-1}\}$ is a subfield of $\mathcal{F}$, and that $\{\overline{0}, \overline{1}, \dots, \overline{p-1}\} \cong \mathbb{Z}/\mathbb{Z}p$. So we've just defined a prime subfield that has a prime order

$p$, and that the field $\mathcal{F}$ is a vector space of a subfield (represent this using $V$ over subfield). Now define operations of $V$:

- Vector addition in $V$ = addition in $\mathcal{F}$
- Scalar multiplication in $V$ = multiplication in $\mathcal{F}$

Now we just need to count the elements in the field, if the dimension of the vector space we've defined is $n$. That is easy, because this is simply $p^n$, because $\dim_{\mathbb{Z}/\mathbb{Z}p} \mathcal{F} = n$ by construction above. ■

Now it is of interest to discuss *linear groups*, and, specifically, *linear operators*. We will bring back something again from linear algebra over $\mathbb{R}$.

**Definition 3.22.** Let $V$ and $W$ be vector spaces over field $\mathcal{F}$. $\varphi : V \mapsto W$ is a **linear transformation** if, $\forall a, b \in \mathcal{F}$, $\forall v, w \in V$,

$$\varphi(av + bw) = a\varphi(v) + b\varphi(w)$$

The following proposition is quite important, and it describes (in part) a correspondence between matrices and linear transformations (functions).

**Proposition 3.23.** If $\varphi : \mathcal{F}^n \mapsto \mathcal{F}^m$ is a linear transformation, then $\exists!$ matrix $A = \mathcal{M}_{m\times n}(\mathcal{F})$ s.t. $\varphi = Av$, $\forall v \in \mathcal{F}^n$. The construction is as follows: such $m \times n$ matrix has colummns

$$A = \begin{pmatrix} | & | & | & | \\ a_1 & a_2 & \dots & a_n \\ | & | & | & | \end{pmatrix}$$

where $a_i = e_i$, and

$$e_i \coloneqq [0\,0\,\dots\,1\,\dots\,0\,\dots]^T$$

where 1 appears at the $i$-th coordinate, $\forall a_i \in \mathcal{F}^m$.

*Proof.* We need to prove the uniqueness, which comes from the choice of basis. If we choose $\{e_i\}_{i=1}^n$, then our matrix is (by construction stated in Proposition 3.23)

$$A = \begin{pmatrix} | & | & | & | \\ \varphi(e_1) & \varphi(e_2) & \dots & \varphi(e_n) \\ | & | & | & | \end{pmatrix}$$

and now, take $c \in \mathcal{F}^n$, and calculate

$$\begin{aligned} Ac &= \begin{pmatrix} | & | & | & | \\ \varphi(e_1) & \varphi(e_2) & \dots & \varphi(e_n) \\ | & | & | & | \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \\ &= \varphi(e_1)c_1 + \varphi(e_2)c_2 + \cdots + \varphi(e_n)c_n \\ &= \varphi(e_1c_1 + e_2c_2 + \cdots + e_nc_n) \\ &= \varphi\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \end{aligned}$$

as desired. ■

By extension, Proposition 3.23 lets us generalise the following fact about changing bases.

**Theorem 3.24.** *If $\varphi : V \mapsto W$, where $\dim V = n$, $\dim W = m$, and both are vector spaces over $\mathcal{F}$, is a linear transformation, and we choose the bases $\{v_1, v_2, \ldots, v_n\}$ for $V$, and $\{w_1, w_2, \ldots, w_m\}$ for $W$, then $\exists! A \in \mathcal{M}_{m\times n}(\mathcal{F})$ s.t. $\forall v \in V$, if*

$$\begin{cases} v & = a_1 v_1 + \cdots + a_n v_n \\ \varphi(v) & = b_1 w_1 + \cdots + b_m v_m \end{cases}$$

*then*

$$(\star) \qquad \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} = A \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

*Proof.* We seek to apply $(\star)$ to $\{v_i\}$ we have chosen. Write a particular $v_i$ as a linear combination, i.e.

$$\begin{aligned} v_i &= 0v_1 + 0v_2 + \cdots + 1v_i + \cdots + 0v_n \\ \varphi(v_i) &= c_{1i}w_1 + c_{2i}w_2 + \cdots + c_{mi}w_m \end{aligned}$$

and the reason we've written $c_{mi}$ is to indicate that we've calculated $\varphi(v_i)$ based on the basis we've chosen for $V$. Then, by $(\star)$, this set of equations has to satisfy

$$\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = A \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

and by Proposition 3.23, we know that such $A$ is unique, and by construction it is

$$A = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{bmatrix}$$

as desired. ■

**Remark 3.25.** We call $A$ in Theorem 3.24 the **matrix of $\varphi$ w.r.t. $\{v_i\}$ and $\{w_i\}$**.

We can study the connections between matrices of linear transformations and *isomorphisms.*

**Definition 3.26.** Suppose $\{v_1, \ldots, v_n\} \in V$. Let $T : V \mapsto W$ be a linear transformation. A **hypervector** is the image of an ordered set of vectors $\mathbf{B} = (v_1, v_2, \ldots, v_n)$ under linear transformation, ie. if we define $\forall a_i \in \mathcal{F}$, $i = 1, 2, \ldots, n$, then

$$T(\mathbf{B}) = (v_1\, v_2\, \ldots\, v_n) \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = a_1 v_1 + \cdots + a_n v_n = T(v_1) + \cdots + T(v_n)$$

**Remark 3.27.** By extension, let $v = \mathbf{B}\, X = \mathbf{B} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = v_1x_1 + \cdots + v_nx_n$. Then,

$$\begin{aligned} T(v) &= T(v_1)x_1 + T(v_2)x_2 + \cdots + T(v_n)x_n \\ &= T(\mathbf{B})X \end{aligned}$$

**Remark 3.28.** This definition is very *pliable*, in the sense that hypervectors are defined for any set of basis. Let $\mathbf{B} = (v_1, v_2, \ldots, v_n)$ denote hypervector.

**Observation 3.29.** Let $\mathbf{B} = (v_1, v_2, \ldots, v_n)$. Assume $\{v_1, v_2, \ldots, v_n\}$ is a basis of vector space $V$. Define the map

$$\eta : \begin{cases} \mathcal{F}^n & \mapsto V \\ \quad a & \mapsto \mathbf{B}\, a \end{cases}$$

we verify that

- $\eta$ is injective: by *uniqueness* of fact that vectors can be written as linear combinations of vectors in basis;
- $\eta$ is surjective: by the fact that *every* vector in $V$ can be written as a linear combination of vectors in basis.

Therefore, $\eta$ is an isomorphism.

**Proposition 3.30.** Let $T : V \mapsto W$ be a linear transformation, and let $\mathbf{B} = (v_1, \ldots, v_n)$ and $\mathbf{C} = (w_1, \ldots, w_m)$ be bases of $V$ and $W$, respectively. Let $X$ be the coordinate vector of an arbitrary vector $v$ with respect to the basis $\mathbf{B}$, and let $Y$ be the coordinate vector of its image $T(v)$. So $v = \mathbf{B}\, X$ and $T(v) = \mathbf{C}\, Y$. There is an $m \times n$ matrix $A$ with the dual properties

$$T(\mathbf{B}) = \mathbf{C}\, A \quad \text{and} \quad AX = Y.$$

*Proof.* The trick is to write $T(v_j)$ as a linear combination of the vectors in $\mathbf{C}$, ie.

$$T(v_j) = w_1a_{1j} + \cdots + w_m + a_{mj}$$

and we assemble the coefficients $a_j$ into a column vector $A_j = (a_{1j}, \ldots, a_{mj})$, so that $T(v_j) = \mathbf{C}\, A_j$. Then if $A$ is the matrix whose columns are $A_1, \ldots, A_n$, then

$$T(\mathbf{B}) = (T(v_1), T(v_2), \ldots, T(v_n)) = (w_1, w_2, \ldots, w_m)A = \mathbf{C}\, A$$

as desired. Now, if $v = \mathbf{B}\, X$, then

$$T(v) = T(\mathbf{B})X = \mathbf{C}\, AX$$

Therefore, the coordinate vector of $T(v)$, which we named $Y$, is equal to $AX$. ■

With this proposition, we can state the following observation:

$$\begin{tikzcd}
\mathcal{F}^n \arrow[r, dashed, "\mathbf{C}^{-1} \circ \varphi \circ \mathbf{B}"] \arrow[d, "\mathbf{B}(\cong)"'] & \mathcal{F}^m \arrow[d, "\mathbf{C}(\cong)"] \\
V \arrow[r, "\varphi"'] & W
\end{tikzcd}$$

This map can only be described based on some choice of bases. The matrix of linear transformation is unique based on the choice of bases.

An important proposition is the following.

**Proposition 3.31.** Let $V$ and $W$ be finite-dimensional vector spaces over a field $\mathbb{F}$ and let $\varphi : V \to W$ be a linear transformation. Let $r = \operatorname{rank} \varphi$, $n = \dim V$, $m = \dim W$. Show that there are bases $v_1, \dots, v_n \in V$ and $w_1, \dots, w_m \in W$ such that

$$\varphi(v_i) = \begin{cases} w_i & \text{if } i = 1, \dots, r \\ 0 & \text{if } i = r+1, \dots, n. \end{cases}$$

The story gets a little more complicated if we have two vector spaces of the *same dimension*. Changing the commutative diagram so that $W$ is replaced by $V$ yields that the new matrix is $A = \mathbf{B}^{-1}\,\varphi\,\mathbf{B}$. This severely restricts the matrix of linear transformation. In particular, consider the following. Let $\mathbf{B}' = (v_1', \dots, v_n')$ be a basis for $\Vdash$. Then the matrix of linear transformation of $\varphi$ and $\mathbf{B}'$ is $A' = (\mathbf{B}')^{-1}\varphi\,\mathbf{B}'$. Here, note that $A$ and $A'$ are conjugate of one another, ie.

$$\begin{aligned} A' &= [(\mathbf{B}')^{-1}\,\mathbf{B}]\varphi[\mathbf{B}^{-1}\,\mathbf{B}]' \\ &= [(\mathbf{B}')^{-1}\,\mathbf{B}]\varphi[(\mathbf{B}')^{-1}\,\mathbf{B}]^{-1} \end{aligned}$$

Let us generalise this finding.

**Proposition 3.32.** Two matrices represent the same linear transformation in different bases *iff* they are conjugates.

To see Proposition 3.32 at work, it requires us to understand that two matrices can represent the same linear transformation *but in different coordinate systems.* In particular, we use the following example.

**Example 3.33.** Let

$$A = \begin{bmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}, \quad B = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 2 \end{bmatrix}$$

which, obviously, represent the same linear transformation, but one requires you to see the $\mathbb{R}^2$ plane from a different orientation. These matrices are conjugates of one another, ie. if we take

$$M = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

then we can verify that $B = MAM^{-1}$ (note that $M$ above is the counterclockwise rotation matrix).

**Proposition 3.34.** The general idea is that if we let $A, B, M \in \mathcal{M}_{m\times n}(\mathcal{F})$, and suppose $M$ is invertible, $B = MAM^{-1}$. Then,

$$\begin{cases} \quad v & \overset{A}{\longmapsto} & Av \\ Mv & \overset{B}{\longmapsto} & BMv \\ & & = (MAM^{-1})Mv = MAv \end{cases}$$

Some consequences of conjugate matrices follow directly.

**Remark 3.35.**
- $Av = 0 \iff BMv = MAv = 0$, which implies $\mathbf{N}(B) = M \times \mathbf{N}(A)$
- (Set of *fixed points*) $Av = v \iff BMv = Mv$, which implies $\operatorname{fix}(B) = M \operatorname{fix}(A)$.

What is interesting is beyond what we can directly observe. In particular, we see that the coordinate system is preserved by *eigenvectors.*

**Definition 3.36. Eigenvectors** are vectors $v \in \mathcal{F}^n$ s.t. $Av = \lambda v$ for some $\lambda \in \mathcal{F}$, where $v \neq 0$.

**Definition 3.37. Eigenvalue** is $\lambda$ given in the previous definition.

**Definition 3.38.** The **eigenspace** of $\lambda$ is the set

$$E_\lambda = \{v \in \mathcal{F}^n \mid Av = \lambda v\}$$

**Remark 3.39.** An eigenspace is a vector subspace of a vector space $A$.

**Remark 3.40.** $\mathbf{N}(A) = E_0(A)$, and $\operatorname{fix}(A) = E_1(A)$.

**Remark 3.41.** If $B = MAM^{-1}$, $\lambda \in \mathcal{F}$, then $Av = \lambda v \iff BMv = MAv = M\lambda v = \lambda Mv$. Hence,

$$\lambda \in E_\lambda(A) \iff M\lambda \in E_\lambda(B)$$

Therefore, we reach the conclusion that $E_\lambda(B) = ME_\lambda(A)$, $\forall \lambda \in \mathcal{F}$.

With these remarks, we can construct the following proposition.

**Proposition 3.42.** If $A, B \in \mathcal{M}_{n\times n}(\mathcal{F})$, and they are conjugates of one another, then eigenvalues of $A$ and $B$ are the same. Furthermore, $\forall \lambda \in \mathcal{F}$, $\dim(E_\lambda(A)) = \dim(E_\lambda(b))$.

**Remark 3.43.** The converse generally does not hold true, *unless we are working over the field fo complex numbers.* Let us see the concept in the following example.

**Example 3.44.** Again, take the rotation matrix

$$M = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

and this has no eigenvalues and eigenvectors over $\mathbb{R}$: note that the characteristic polynomial is

$$\det(M - \lambda I) = \lambda^2 + 1 = 0$$
$$\implies \lambda = \pm i$$

Let $\chi_A$ denote the characteristic polynomial of matrix $A$. In this case, we say $\chi_A = 0$ has solutions $x_1 = i$, and $x_2 = -i$.

The following idea is the most profound that we've come across so far. This is provided sans proof in this case, but is germane to this discussion.

**Theorem 3.45** (Fundamental Theorem of Algebra)**.** Let $f(\lambda)$ be a polynomial of the form

$$f(\lambda) = a_n\lambda^n + \cdots + a_1\lambda + a_0$$

where $n > 0$, $a_n \neq 0$, and $\{a_k\}_{k=1}^{n} \in \mathbb{C}$. Then, $\exists \lambda_0 \in \mathbb{C}$ s.t. $f(\lambda_0) = 0$.

**Remark 3.46.** Complex number $\mathbb{C}$ is not the only field that Theorem 3.45 holds; in general, if we replace $\mathbb{C}$ with algebraic relations, then we still get the same result.

**Remark 3.47.** A natural question to ask is *when are two matrices over* $\mathbb{C}$ *conjugates of one another*? We note two direct consequences:

- If $A$ and $B$ are conjugates, then the eigenvalues are the same, and $\dim(E_\lambda(A)) = \dim(E_\lambda(B))$.
- If $A$ and $B$ are conjugates, then $\chi_A = \chi_B$.

**Remark 3.48.** Similar to Remark 3.43, we note that Remark 3.47 contains *only necessary conditions.* In general, they are not sufficient. The following makes the discussion of conjugate matrices a lot clearer.

**Theorem 3.49.** If $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ and $\chi_A$ has $n$ distinct roots, then the following are equivalent:

(1) $A$ is diagonalisable
(2) $A$ is conjugate to the diagonal matrix

$$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \tag{$\triangle$}$$

*Proof.* The proof of this theorem is due to linear algebra over $\mathbb{R}$. ■

**Remark 3.50.** Therefore, if $A$ and $B$ are such that $\chi_A = \chi_B$, and these characteristic polynomials have distinct roots, then $A$, $B$ are conjugate to ($\triangle$).

**Example 3.51.** Take a strictly upper triangular matrix

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

where the blank entries are zeros. Then, $\chi_D = \lambda^3 = 0 \implies \lambda_1 = \lambda_2 = \lambda_3 = 0$, so the eigenvalues are 0, with multiplicity 3. What is the simplest matrix that is conjugate to $D$? Well, take $E = 3 \times 3$ zero matrix and we're done. However, the zero matrix is only conjugate to itself, ie. let $\mathbf{0}$ denote the $3 \times 3$ zero matrix; then,

$$M\mathbf{0}M^{-1} = \mathbf{0} \implies M \in GL_3(\mathbb{C})$$

Diagonalisation is generally quite difficult to achieve (save for applying singular value decomposition when working over $\mathbb{R}$); we have a close relative to diagonalisation that we will introduce briefly, sans proof, below. It is best motivated by an example.

**Example 3.52.** Let $v = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} \in \mathbb{C}^3$. Furthermore, let $Av = \begin{pmatrix} 3 \\ 3 \\ 0 \end{pmatrix}, A^2v = \begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix}, A^3v = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$. We can verify that $Av, A^2v, A^3v$ form a basis, and that

$$v \xmapsto{A} Av \xmapsto{A} A^2v \xmapsto{A} A^3v = 0$$

so the matrix $A$ w.r.t. this basis is

$$\begin{array}{c|ccc} & v & Av & A^2v \\ \hline v & 0 & 0 & 0 \\ Av & 1 & 0 & 0 \\ A^2v & 0 & 1 & 0 \end{array}$$

This is a very nice result, one we refer to by a proper name.

**Theorem 3.53** (Jordan Decomposition Theorem)**.** If $A \in \mathcal{M}_{n\times n}(\mathbb{C})$, then $A$ is conjugate to a matrix of the form

$$B = \begin{bmatrix} B_1 & & & \\ & B_2 & & \\ & & \ddots & \\ & & & B_n \end{bmatrix}$$

where $B_i$, $\forall i$, is a **Jordan block**.

**Definition 3.54.** A **Jordan block** is of the following form:

$$\begin{bmatrix} \lambda \end{bmatrix}, \begin{bmatrix} \lambda & \\ 1 & \lambda \end{bmatrix}, \begin{bmatrix} \lambda & & \\ 1 & \lambda & \\ & 1 & \lambda \end{bmatrix}, \dots, \begin{bmatrix} \lambda & & & & \\ 1 & \lambda & & & \\ & 1 & \lambda & & \\ & & \ddots & \ddots & \\ & & & 1 & \lambda \end{bmatrix}$$

**Remark 3.55.** Every matrix $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ is conjugate to the form $B$ in Theorem 3.53. Take a simple example: $(\triangle)$ matrix. Then, this is diagonalisable iff all Jordan blocks are single blocks, $1 \times 1$ matrix.

**Remark 3.56.** Why is this theorem useful? Well, if we have figured out the number of distinct eigenvalues, and the multiplicities of those eigenvalues, we can figure out the appropriate Jordan block matrix that the matrix of interest is conjugate to.

## 4. Orthogonal Groups

The idea of this section is fairly simple: we want to study more about matrix groups. So far, we've encountered the following:

$$\begin{aligned} GL_n(\mathcal{F}) &= \{ \text{ invertible matrix of size } n \times n\} \\ T_n(\mathcal{F}) &= \{ \text{ upper-triangle matrix of size } n \times n\} \end{aligned}$$

where these matrices admit coefficients from field $\mathcal{F}$.

**Example 4.1.** If we have $v\ in\, \mathcal{F}^n$, then we define **stabilizer of** $v$ as

$$\operatorname{Stab} v = \{A \in GL_n(\mathcal{F}) \mid Av = v\}$$

We can verify that this is a subgroup of $GL_n(\mathcal{F})$:

- Contains the identity matrix $I_n$
- Closed under multiplication: if $A$, $B \in \operatorname{Stab}(v)$, then $ABv = Av = v$
- Closed under inverses: if $A \in \operatorname{Stab}(v)$, then $Av = v \implies A^{-1}Av = A^{-1}v = v$.

**Definition 4.2** (Basic vector operations)**.** If $\vec{v}, \vec{w} \in \mathcal{F}^n$, and, WLOG, let

$$\begin{aligned} \vec{v} &= (v_1, v_2, \dots, v_n) \\ \vec{w} &= (w_1, w_2, \dots, w_n) \end{aligned}$$

then define

- $\langle \vec{v}, \vec{w} \rangle = v_1 w_1 + \cdots + v_n w_n$
- $\|\vec{v}\| = \sqrt{\langle \vec{v}, \vec{v} \rangle} = \sqrt{v_1^2 + \cdots + v_n^2}$
- These vectors are **orthogonal** if $\langle \vec{v}, \vec{w} \rangle = 0$
- These vectors are **orthonormal** if $\langle \vec{v}, \vec{w} \rangle = 0$, and $\|\vec{v}\| = \|\vec{w}\| = 1$.

**Remark 4.3.** A set of vectors $V = \{\vec{v}_i\}_{i=1}^n$ is orthonormal if, $\forall i, j$,

$$\langle \vec{v}_i, \vec{v}_j \rangle = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

so the set $V$ has linearly independent vectors, and $V$ is an orthonormal basis.

**Theorem 4.4** (From linear algebra over $\mathbb{R}$)**.** If $A \in \mathcal{M}_{n\times n}(\mathbb{R})$, then the following are equivalent:

(1) $\forall \vec{v} \in \mathbb{R}^n, \|\vec{v}\| = \|A\vec{v}\|$ ($A$ preserves the length of $\vec{v}$)
(2) $\forall \vec{v}, \vec{w} \in \mathbb{R}^n$, $\langle A\vec{v}, A\vec{w} \rangle = \langle \vec{v}, \vec{w} \rangle$
(3) $A^T A = I \iff AA^T = I$
(4) Columns of $A$ form an orthonormal basis
(5) Rows of $A$ form an orthonormal basis

**Definition 4.5.** Any matrix $B$ satisfying the conditions in Theorem 4.4 is an orthogonal matrix.

**Properties 4.6.** The following are properties of an orthogonal matrix $A$:

(1) $\det(A) = \pm 1$, because

$$\begin{aligned} \det(A^T A) &= \det(I) \\ \det(A^T)\det(A) &= 1 \\ [\det(A)]^2 &= 1 \\ \implies \det(A) &= \pm 1 \end{aligned}$$

(2) $A^{-1}$ is orthogonal: since $A$ is orthogonal, the columns of $A$ are orthonormal, which

$$\begin{aligned} \implies A^{-1} &= A^T \\ &\implies \text{rows of } A^{-1} \text{ are orthonormal} \\ &\implies A^{-1} \text{ is an orthogonal matrix} \end{aligned}$$

(3) $A$, $B$ orthogonal implies $AB$ is orthogonal matrix, because

$$\forall v \in \mathcal{F}^n, \quad \big\|A(Bv)\big\| = \|Bv\| = \|v\|$$

by condition 1 in Theorem 4.4.

**Definition 4.7.** The set of orthogonal $n \times n$ matrices admitting coefficients in field $\mathcal{F}$ is denoted $O_n(\mathcal{F})$ (the **orthogonal (matrix) group**). Furthermore, $O_n(\mathcal{F}) \leq GL_n(\mathcal{F})$.

**Definition 4.8.** The **special orthogonal group** admitting coefficients in field $\mathcal{F}$, denoted $SO_n(\mathcal{F})$, is defined

$$SO_n(\mathcal{F}) = \{A \in O_n(\mathcal{F}) \mid \det(A) = 1\}$$

We want to first characterise the set of actions we can perform in lower dimensions. We begin with $n = 2$, ie. studying $O_2$. In particular, we want to know what transformations preserve the lengths in two dimensions.

**Definition 4.9.** A **rotation** in $\mathbb{R}^2$ is characterised by matrix $R_\theta$, where $\theta$ is the angle of rotation.

**Definition 4.10.** The **rotation matrix** in two dimensions is constructed by choosing the “canonical” set of vectors as our basis, ie. take

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

and coordinates of $R_\theta$, under these basis vectors, are

$$R(e_1) = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \quad R(e_2) = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$$

the matrix of linear transformation (rotation) is therefore

$$T = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

**Definition 4.11. Reflection** is a little more difficult to characterise. Specially, we want our choices of *basis vectors/eigenvectors to be orthogonal*—even though we know that, in

$\mathbb{R}^2$, as long as we have two non-parallel direction vectors, we have a basis. We can then define two reflections (of basis vectors $\vec{w}_1$ and $\vec{w}_2$) as

$$\begin{aligned} S_\theta w_1 &= w_1 \\ S_\theta w_2 &= -w_2 \end{aligned}$$

where $S_\theta$ is the reflection through line with angle $\theta$ with $x$ axis. Then, we get that these have eigenvalues 1 and $-1$. In general, we construct

$$\begin{aligned} &w_1 \quad \text{with angle } \theta \\ &w_2 \quad \text{with angle } \theta + \frac{\pi}{2} \end{aligned}$$

**Remark 4.12.** Such reflection matrix $S_\theta$ can also be characterised w.r.t. conjugation. To be clear, take the rotation matrix $R_\theta$, and define $S'_\theta$ as

$$\begin{aligned} S'_\theta &= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix} := \bar{A} \end{aligned}$$

and we notice that $\bar{A} = \bar{A}^{-1}$, so this is indeed a reflection matrix.

**Theorem 4.13.** $M \in O_2$ is either

- rotation matrix ($\det M = 1$); or,
- reflection matrix ($\det M = -1$).

*Proof.* If $M$ is orthogonal, then it has orthonormal columns. Then, $\exists \theta \in [0, 2\pi]$ s.t.

$$M = \begin{bmatrix} \cos\theta & x \\ \sin\theta & y \end{bmatrix}$$

where we want $\|(x, y)\| = 1$, and $\langle (x, y), (\cos\theta, \sin\theta) \rangle = 0$. Hence, we want

$$\begin{aligned} x\cos\theta + y\sin\theta &= 0 \\ \implies (x, y) &= t(\sin\theta, -\cos\theta) \\ \implies \langle (x, y) \rangle &= |t| \\ \implies t &= \pm 1 \end{aligned}$$

hence, $(x, y) = \pm(\sin\theta, -\cos\theta)$. Then we get either of the following:

$$M = \begin{cases} \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix} & = S_{\frac{\pi}{2}} \\ \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} & = R_\theta \end{cases}$$

■

**Corollary 4.14.**

- Composition of rotations is rotation
- Composition of reflections is rotation.

- Composition of rotation and reflection is a reflection.

This is easily deduced by multiplying two matrices and looking at their determinants.

## 5. Isometries and Isometry Group

From Definition 4.7, we see that any orthogonal matrix $A \in \mathcal{M}(\mathbb{R})$ satisfies the property that $\forall v \in \mathbb{R}^n$, $\|Av\| = \|v\|$.

**Definition 5.1.** An **isometry** of $\mathbb{R}^n$ is a map

$$\varphi : \mathbb{R}^n \mapsto \mathbb{R}$$

s.t. $\forall v, w \in \mathbb{R}^n$, $\|v - w\| = \big\|\varphi(w) - \varphi(w)\big\|$.

**Remark 5.2.** A map that preserves lengths automatically preserves distances.

**Remark 5.3.** A few objects we've studied are isometries:

- Orthogonal linear transformations are isometries:
$$\begin{aligned} \forall v, w \in \mathbb{R}^n, &\quad \forall M \in O_n \\ \|Mv - Mw\| &= \big\|M(v - w)\big\| \\ &= \|v - w\| \end{aligned}$$
- Translations are isometries: let $a \in \mathbb{R}$; let
$$\begin{array}{cccc} t_a & : & \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ & & x & \longmapsto & x + a \end{array}$$
then this
$$\begin{aligned} \implies t_a(v) - t_a(w) &= \big\|v + a - (w + a)\big\| \\ &= \|v - w\| \end{aligned}$$
- Compositions of isometries are isometries: let $f, g$ be isometries, defined
$$f, g : \mathbb{R}^n \mapsto \mathbb{R}$$
then it must be the case, for $f \circ g$, that $\forall v, w \in \mathbb{R}^n$,
$$\begin{aligned} \big\|(f \circ g)(v) - (f \circ g)(w)\big\| &= \big\|f(g(v)) - f(g(w))\big\| \\ &= \big\|g(v) - g(w)\big\| = \|v - w\| \end{aligned}$$

We state the following in full; the proof is in Artin, which is omitted because it is less than instructive.

**Theorem 5.4.** Every isometry of $\mathbb{R}^n$ is a composition of an orthogonal matrix and a translation. In particular, if $f : \mathbb{R}^n \mapsto \mathbb{R}$ is an isometry, then $\exists! a \in \mathbb{R}^n$, $M \in O_n$ s.t. $f = t_a \circ M$.

There are a few direct consequences of Theorem 5.4; we state a couple of them.

**Corollary 5.5.** The set of isometries in $\mathbb{R}^n$ forms a group; let us denote it $\mathrm{Isom}_n$, we law of composition as the composition of maps.

*Proof.* Check the usual conditions:

1. Closure under group operation
2. Identity element: compose identity matrix and identity function (translation)

(3) Inverse: if $f = t_a \circ M$, for some $a \in \mathbb{R}^n$, $M \in O_n$, then

$$\begin{aligned} f^{-1} &= (t_a \circ M)^{-1} \\ &= M^{-1} \circ t_{-a} \end{aligned}$$

and we know such inverse matrix exists because the definition of orthogonal matrix says that such $M$ satisfies $M^T M = I$.

$\blacksquare$

**Properties 5.6** (of the isometry group).

- If $M \in O_n$, $a \in \mathbb{R}^n$, then

$$\begin{aligned} M \circ t_a(x) &= M(x+a) = Mx + Ma \\ &= (t_{Ma} \circ M)(x) \\ \implies M \circ t_a &= t_{Ma} \circ M \end{aligned}$$

- $\forall a, b \in \mathbb{R}^n$, $t_a t_b = t_{a+b}$
- If $a \in \mathbb{R}^n$, $f \in \mathrm{Isom}_n$ (in particular, $f = t_b \circ M$), then

$$\begin{aligned} f \circ t_a \circ f^{-1} &= t_b \circ M \circ t_a \circ M^{-1} \circ t_b^{-1} \\ &= t_b \circ (t_{Ma} \circ M) \circ M^{-1} \circ t_b^{-1} \\ &= t_b \circ t_{Ma} \circ (M \circ M^{-1}) \circ t_b^{-1} \\ &= t_b \circ t_{Ma} \circ t_b^{-1} = t_{Ma} \end{aligned}$$

which is just a translation; hence, conjugation of translation by isometry is still a translation.
- Denote the group of translation $T$. $T \lhd \mathrm{Isom}_n$.

**Remark 5.7.** Of course, we can prove that $T \lhd \mathrm{Isom}_n$ if we construct a homomorphism $\pi$, with the property $\ker \pi = T$. Let us preserve the linear part of the translation—the derivative of the translation, ie.

$$\begin{array}{rccc} \pi & : & \mathrm{Isom}_n & \longrightarrow & O_n \\ & & t_a \circ M & \longmapsto & M \end{array}$$

check that this is indeed a homomorphism:

$$\begin{aligned} \pi(t_a M M' t_b) &= \pi(t_a t_{Mb} M t_b) \\ &= \pi(t_{a+Mb} M M') = M M' \end{aligned}$$

with $\ker \pi = \{\text{translations}\}$.

**Definition 5.8.** Let $f \in \mathrm{Isom}_n$. Then, we say

(1) $f$ is **orientation-preserving** if $\det(\pi(f)) = 1$; and
(2) $f$ is **orientation-reversing** if $\det(\pi(f)) = -1$,

where

$$\begin{array}{rccc} \pi & : & \mathrm{Isom}_n & \longrightarrow & O_n \\ & & t_a \circ M & \longmapsto & M \end{array}$$

is the derivative map we've defined above.

Again, we study only isometries of lower dimension spaces. It is particularly interesting to study $\text{Isom}_2$. We have a nice classification theorem as follows.

**Theorem 5.9.** Every isometry of $\mathbb{R}^2$ falls into either one of the four categories:

1. **Translation**
2. **Rotation** around some point in the plane
3. **Reflection** through some line $\ell$
4. **Glide reflection**: reflection through $\ell$, then translation parallel to $\ell$

*Proof.* In order to prove Theorem 5.9, we need to prove the following lemmas; we prove them one by one.

**Lemma 5.10.** *All* orientation-preserving isometries are either translations or rotations.

*Proof of Lemma 5.10.* The idea is as follows. Let $f : \mathbb{R}^2 \mapsto \mathbb{R}^2$ be a rotation of a point $x$ by angle $\theta$, around some point $P$. Let us denote distances: $(x-p)$ for the first coordinate, and $f(x) - p$ for the second coordinate. But, with some calculations, we can show that

$$\begin{aligned} f(x) - p &= R_\theta(x-p) \\ \implies f(x) &= R_\theta(x-p) + p \\ &= t_p \circ R_\theta \circ t_{-p} \end{aligned}$$

Then, let $f = t_a \circ M_2$, where $a \in \mathbb{R}^2$, $M \in O_2$. Then, $\det M = 1$, and $M = R_\theta$ for some $\theta \in [0, 2\pi)$. Then, there are two subcases:

1. $\theta = 0 \implies M = R_0 = id$. Hence, $f$ is a translation.
2. $\theta \neq 0 \implies f$ is a rotation. How do we see this? We need to first solve for a fixed point $p$, ie. the point of rotation. Hence,

$$\begin{aligned} t_a \circ M &= p \\ \implies t_a \circ R_\theta p &= p \\ \implies a &= (I - R_\theta)p \end{aligned}$$

and $(I - R_\theta)$ needs to be invertible in order for us to solve $p$; a direct computation yields

$$\begin{aligned} \det(I - R_\theta) &= \det \begin{bmatrix} 1-\cos\theta & -\sin\theta \\ \sin\theta & 1-\cos\theta \end{bmatrix} \\ &= 1 - 2\cos\theta + \cos^2\theta + \sin^2\theta \\ &= 2 - 2\cos\theta \neq 0 \end{aligned}$$

since $\theta \neq 0, 2\pi, \dots, n\pi$, where $n \in \mathbb{Z}2$. Hence, $\exists! p \in \mathbb{R}^2$ s.t. $f(p) = p$. Hence, $f$ is a rotation by $\theta$ around some point $p$. Lastly, we just want to show that $f$ is well-defined:

$$\begin{aligned} t_p \circ R_\theta \circ t_{-p} &= t_p \circ t_{R_\theta(-p)} \circ R_\theta \\ &= t_{p(I-R_\theta)} \circ R_\theta \\ &= t_a \circ R_\theta \end{aligned}$$

as desired.

■

**Lemma 5.11.** Every orientation-reversing $f \in \mathrm{Isom}_2$ is either a reflection or a glide-reflection.

*Proof of Lemma 5.11.* The idea is fairly simple. Assume, WLOG, $f = t_a \circ S_\ell$, where $S_\ell$ is a reflection through line $\ell$. Define $u$, $v$ by choosing a particular coordinate, and make sure such vector passes through the origin. Then, change the coordinates such that line $\ell$ is the $u$-axis. Then, we can say that

$$\begin{aligned} S_\ell(u,v) &= (u,-v) \\ \implies t_a \circ S_\ell(u,v) &= (u,-v) + (a_1,a_0) \\ &= (u+a_1, a_0 - v) \end{aligned}$$

and this is a glide reflection through $\ell' = \left\{v = \frac{a_2}{2}\right\}$, and translation by $(a_1, 0)$. Hence, we can (algebraically) say that

$$\begin{aligned} S_{\ell'}(u,v) &= (u, a_2 - v) \\ \implies t_{(a,0)} \circ S_{\ell'} &= (u, a_2 - v) + (a_1, 0) \\ &= (u + a_1, a_2 - v) \end{aligned}$$

which demonstrates the fact that any such $f$ is well-defined.

Now, if $a_1 = 0$, then we have a translation; if $a_1 \neq 0$, we have a glide reflection. ■

Combining both Lemma 5.10 and Lemma 5.11 proves the theorem we want to prove. ■

A particularly powerful way to study groups of lower dimensions, such as $O_2$, is to classify the *finite subgroups* of such groups. Let us consider classifying finite subgroups of $O_2$—which can be understood as symmetry groups of some geometric figures. We are ultimately interested in proving the following theorem.

**Theorem 5.12.** Let $G$ be a finite subgroup of $O_2$. Then, $\exists n > 0$ s.t. either

(1) $G$ is a cyclic group of order $n$, generated by rotation $R_{\frac{2\pi}{n}}$; or,
(2) $G = D_n$, the dihedral group of order $2n$.

We cannot begin to prove this theorem, unless we have more structures to work with. In particular, we want to define a dihedral group, and describe a couple of properties of such group.

**Definition 5.13. Dihedral group** of order $n$, denoted $D_n$, is a group generated by 2 elements:

$$\begin{aligned} R_{\frac{2\pi}{n}} &\coloneqq R \\ S_\ell &\coloneqq S \end{aligned}$$

**Example 5.14.** Let us examine $D_3$. This is the group of symmetries of an equilateral triangle. Such triangle has three lines of symmetry—each from a vertex to the midpoint of its opposite side. This is indeed a group: we can define all the elements in this group from the collection of symmetries on equilateral triangles,

$$\left\{id, R_{\frac{2\pi}{3}}, R_{\frac{4\pi}{3}}, S_{\ell_1}, S_{\ell_2}, S_{\ell_3}\right\}$$

in particular, we let $R = R_{\frac{2\pi}{3}}$, $R^2 = R_{\frac{4\pi}{3}}$, and $S = S_{\ell_1}$ in the computations below. Then, we can see that

$$S \circ R = \left\{ \begin{array}{ccc} B & \longmapsto & C \\ A & \longmapsto & A \\ C & \longmapsto & B \end{array} \right. = S_{\ell_2}$$

$$R \circ S = \left\{ \begin{array}{ccc} B & \longmapsto & A \\ A & \longmapsto & B \\ C & \longmapsto & C \end{array} \right. = S_{\ell_3}$$

so compositions of reflection with rotation (in any order) gives reflection. Hence, $D_3$ is the set

$$\left\{ id, R, R^2, S, SR, SR^2 \right\}$$

equipped with composition.

**Remark 5.15.** We claim that $D_3 \cong S_3$. Recall that we can always define $S_3$ using two generators: $x = (123)$ and $y = (12)$, for example. And, we can then define $SR^{-1} = RS$, resulting in a one-to-one correspondence between the elements generated by these elements.

Another way of seeing the existence of this isomorphism is to construct a homomorphism: namely,

$$\begin{array}{rcccl} \varphi & : & D_3 & \longrightarrow & S_3 \\ & & f & \longmapsto & \text{permutations of vertices induced by } f \end{array}$$

and map all elements in $D_3$ to exactly one element in $S_3$.

**Remark 5.16.** The size of $D_n$ is generally much smaller than the size of $S_n$; in particular, $|D_n| = 2n$ while $|S_n| = n!$.

**Proposition 5.17.** The generators of $D_n$ satisfy the following relations:

- $R^n = id$
- $S^2 = id$
- $RS = SR^{-1}$

and $|D_n| = 2n$, while the set of the group $D_n$ is

$$\left\{ \underbrace{id, R, R^2, \ldots, R^{n-1}}_{n \text{ rotations}}, \underbrace{S, SR, SR^2, \ldots, SR^{n-1}}_{n \text{ reflections}} \right\} \tag{5.1}$$

*Proof.* The idea is to prove that the elements in this set are distinct, and that there are exactly $2n$ elements.

For the first part, we know that rotations and reflections are distinct, hence $R^i \neq SR^j$, $\forall i, j$.

For the second part, it suffices to prove that elements in (5.1) are closed under multiplication, and have inverses that are in the group. It follows that it suffices to prove closure by multiplication by the two generators of $D_n$,

- Multiplication by $R$: we have that

$$\begin{aligned} R\cdot R^j &= R^{j+1}\in D_n\\ SR^i\cdot R &= S\cdot R^{i+1}\in D_n\\ R\cdot SR^i &= SR^{-1}R^i = SR^{i-1}\in D_n \end{aligned}$$

- Multiplication by $S$: similarly,

$$\begin{aligned} S\cdot R^j &\in D_n\\ R^iS\cdot R &= S\cdot R^{-i}\in D_n\\ SR^i\cdot S &= \underbrace{SS}_{=S^2=id}R^{-i} = R^{-i}\in D_n \end{aligned}$$

and that inverses are closed:

$$(R^i)^{-1} = R^{-i}$$

which is defined because $R^{-i}$ would just be a backward/reverse rotation. ■

Before we can prove the finite subgroups classification of $O_2$, we need to define one more piece of convention.

**Definition 5.18.** A subgroup $\Gamma\leq\mathbb{R}$ is **discrete** if $\exists\varepsilon>0$ s.t. $\forall x\in\Gamma$, either $|x|>\varepsilon$, or $x=0$.

**Example 5.19.** $\mathbb{Z}\leq\mathbb{R}$ is a discrete subgroup. In particular, recall that subgroups of $\mathbb{Z}^+$ is either trivial or of the form $\mathbb{Z}\,a$, for some least positive integer $a>0$. $\mathbb{Z}\,a$ is also a discrete subgroup of $\mathbb{R}$.

**Example 5.20.** $\mathbb{Q}\leq\mathbb{R}$ is *not* a discrete subgroup, because $\mathbb{Q}$ is dense in $\mathbb{R}$.

We want to prove a quick lemma, as follows.

**Lemma 5.21.** Let $\Gamma\leq\mathbb{R}$ be a discrete, non-trivial subgroup. Then, $\exists\alpha>0$ s.t. $\Gamma=\mathbb{Z}\,\alpha$.

*Proof.* The proof follows largely by definition. In particular, we need to prove in two steps:

(1) That $\Gamma$ has a least positive element,
(2) That $\Gamma$ is generated by some $a>0$.

The first step is nontrivial because if we define $\{x_n\}\subset\mathbb{Q}$, where $x_n=\frac{1}{n}$, then we have the case that, $\forall n\in\mathbb{N}$,

$$x_n=\frac{1}{n}>\frac{1}{n+1}=x_{n+1}$$

But we can always define something as follows: $\forall a,b\in\Gamma$, $a-b\in\Gamma$, so by definition of discrete subgroup, $a-b=0$ or $|a-b|>\varepsilon$, for some $\varepsilon>0$. Hence, two consequences can happen:

(1) $a-b=0\implies a=b$
(2) $b-\varepsilon<a<b+\varepsilon$

If we're in the first case, we're done. If we're in the second case, then we can always say that any interval of length $\varepsilon$ contains at most one element of $\Gamma$. In other words, $\exists c$ s.t. $c \in \Gamma$, and if $c > 0 \implies$ we're done; $c < 0 \implies$ pick $-c$ to do the trick.

Finally, pick $(0, c]$ as our interval, and cut it into finitely many intervals of length $\varepsilon$. Take $[a, b]$ as teh smallest interval among those intervals (where $a$, $b$ are the least possible values in the intervals we've cleaved).

The second step is to prove that $G = \mathbb{Z}\,a$, where $a$ is the least positive element of $G$. One inclusion is obvious: $\mathbb{Z}\,a \subset G$, because $a \in G$. Hence, $\langle a \rangle = \mathbb{Z}\,a \subset G$. The other inclusion comes from a little less obviously: take $g \in G$; then $\exists N \in \mathbb{N}$ s.t. $g = na$, which implies $n \leq \frac{g}{a} < n + 1$. Hence, multiplying through the inequalities through by $a$

$$\begin{array}{ccccccc} & an & \leq & g & < & (n+1)a \\ \implies & 0 & \leq & g - na & < & a \end{array}$$

but $a$ can be arbitrarily small, hence $g - na = 0 \implies g = na$. Hence, $G \subset \mathbb{Z}\,a$.

Another way to prove second step is as follows: again, one inclusion is obvious—$\mathbb{Z}\,a \subset G$, because $a \in G$. For the other inclusion, let $g \in G$, and $g = ra$ where $r \in \mathbb{R}$. We write $r = m + r'$, where $m \in \mathbb{Z}$ and $0 \leq r' < 1$. Since $G$ is a group, $g' = g - ma \in G$ and $g' = r'a$. Since $0 \leq g' < a$, and $a$ is the least positive integer in $G$, $g' = 0 \implies g - ma = 0 \implies g = ma$. Hence, $G \subset \mathbb{Z}\,a$. ■

Now we're ready to prove Theorem 5.12.

*Proof of Theorem 5.12.* There are two different cases we must consider.

(1) G contains rotations only. Claim: $G \cong C_n$ for some $n$. Consider the set

$$A = \{\theta \in \mathbb{R} \mid R_\theta \in G\}$$

which is a subgroup of $\mathbb{R}^+$. This contains $R_{2\pi} = R_0 = id$. This is a discrete subgroup because $A$ contains finitely many numbers between 0 and $2\pi$. Hence, by Lemma 5.21, $A = \mathbb{Z}\,a$, where $a > 0$. We know $2\pi$ is in the set; hence,

$$\begin{aligned} 2\pi &= na \\ \implies a &= \frac{2\pi}{n} \end{aligned}$$

hence

$$G = \left\{R_{\frac{2\pi}{n}}, R_{2\times\frac{2\pi}{n}}, \ldots, R_{(n-1)\times\frac{2\pi}{n}}\right\} = \left\langle R_{\frac{2\pi}{n}} \right\rangle \implies G \cong \left\langle R_{\frac{2\pi}{n}} \right\rangle$$

(2) G contains rotations and a reflection. Let $S$ be a reflection through a line $\ell$. Choose coordinates so that $\ell$ is the $x$-axis. Let

$$H = \{\text{rotations in } G\} = SO_2 \cap G$$

and $H$ is finite as a subset of $G$, which is a finite group. Hence, by case 1, $\exists n > 0$ s.t. $H = C_n = \left\langle R_{\frac{2\pi}{n}} \right\rangle$. Hence

$$G \supset \left\{id, R_{\frac{2\pi}{n}}, R_{2\times\frac{2\pi}{n}}, \ldots, R_{(n-1)\times\frac{2\pi}{n}}, S, R_{\frac{2\pi}{n}}S, R_{2\times\frac{2\pi}{n}}S, \ldots, R_{(n-1)\times\frac{2\pi}{n}}S\right\} = D_n$$

The other inclusion is more difficult to show. Suppose $M \in G$. If $M$ is a rotation, then $M \in H \subset D_n$; if $M$ is a reflection, then $MS$ is a rotation $\implies MS = R_{\frac{2\pi}{n}k}$ for some $k \implies M = R_{\frac{2\pi}{n}k}S^{-1} \in D_n$. Hence, $G \cong D_n$.

■

A much more profound question is to characterise the finite subgroups of $\text{Isom}_2$. We define

$$\text{Isom}_2 = \left\{ x \mapsto Mx + a : \ M \in O_2, a \in \mathbb{R}^2 \right\}$$

and we can see that finite subgroups of $\text{Isom}_2$ contains no translations or glide-reflections. This is simply because finite subgroups have elements of finite order.

**Theorem 5.22.** A finite subgroup of $\text{Isom}_2$ is isomorphic to $C_n$ or $D_n$ for some $n \in \mathbb{N}$.

To prove Theorem 5.22, we need the following lemma.

**Lemma 5.23.** If $G \subset \text{Isom}_2$ is a finite subgroup, then it has a fixed point $p$. In other words, $\exists p \in \mathbb{R}^2$ s.t. $g(p) = p$, $\forall g \in G$.

We need to define the following object:

**Definition 5.24.** Let $v \in \mathbb{R}^2$. The **orbit** of $v$ is the set

$$Gv = \left\{ g(v) \in \mathbb{R}^2 : \ g \in G \right\}$$

where $Gv$ is symmetric in $G$.

*Proof of Lemma 5.23.* Let $f \in G$. Then,

$$\begin{aligned} f(Gv) &= \left\{ f(g(v)) : \ g \in G, v \in R^2 \right\} \\ &= \left\{ (f \circ g)(v) : \ g \in G, v \in R^2 \right\} \\ &\subset Gv \end{aligned}$$

Similarly, we know

$$\begin{aligned} &f^{-1}(Gv) \subset Gv \\ &\implies Gv \subset f(Gv) \end{aligned}$$

therefore, $f(Gv) = Gv$. We conclude that orbit is preserved by an isometry. ■

*Proof of Theorem 5.22.* We define the *centroid* as the average of the $Gv$, where, since $G$ is finite, we can write

$$Gv = \{s_1, \ldots, s_n\}$$

and the centroid $c$ can therefore be expressed as

$$c = \frac{1}{n}(s_1 + \cdots + s_n)$$

We claim that $\forall g \in G$, $g(c) = c$. Let $g \in G$. Then, $\exists M \in O_2, v \in \mathbb{R}^2$ s.t. $\forall x \in \mathbb{R}^2$, $g(x) = Mx + v$. Then,

$$g(c) = Mc + v$$

$$
\begin{aligned}
&= M\left(\frac{1}{n}\left(s_1 + \cdots + s_n\right)\right) + v \\
&= \frac{1}{n}MS_1 + \frac{1}{n}MS_2 + \cdots + \frac{1}{n}MS_n + v \\
&= \frac{1}{n}(MS_1 + v) + \frac{1}{n}(MS_2 + v) + \cdots + \frac{1}{n}(MS_n + v) \\
&= \frac{1}{n}(g(s_1) + \cdots + g(s_n)) \\
&= \frac{1}{n}\left(s_1 + \cdots + s_n\right) = c
\end{aligned}
$$

where the penultimate equality sign in the last line follows from the fact that $\forall g \in G, g(Gv) = Gv$. Since $c$ is a fixed point, ie. $g(c) = c$ for all $g \in G$, rotation about point $c$ *and* reflection about a line passing through $c$ with angle $\theta$ would be a decent coordinate system to use.

To that end, let $c$ be the origin of the coordinate system, ie. $c = 0$. Hence, $g(0) = 0$. Then, $G \subset O_2$. And, since $G$ is finite, use Theorem 5.12 to say that $G$ is either $C_n$ or $D_n$. Hence, any finite subgroup of $\text{Isom}_2$ is either $C_n$ or $D_n$, as desired. ■

Now, we know it is possible to characterise all finite subgroups of $\text{Isom}_2$. Let us try to construct a characterisation of *infinite subgroups of* $\text{Isom}_2$. In particular, let us characterise the possible groups of symmetry of a figure in the plane. If I have a rectangular brick tiling, it is possible to define translations, rotations about some point(s), and reflections/glide-reflections if we want; *but* it is not always the case that, if we take a subgroup $G$, we can construct a figure that is symmetric under $G$. If I take a subgroup $G$ generated by two translations and a reflection about a line $\ell$ passing through the origin, with angle $\theta$, then I have exactly

$$G = \langle t_a, t_b, s_\ell \rangle$$

where $a, b \in \mathbb{R}^2$, and $s_\ell$ is a reflection about the line $\ell$; the claim is elements in this group is symmetric under $G$. We know this because everything plays out "nicely": translations are easy to manipulate, and reflections always produce one corresponding image. But some rather un-nice sets are the following: take

$$
\begin{aligned}
H &= \langle R_1 \rangle \\
J &= \left\{(x, y) : \; x, y \in \mathbb{Q}\right\}
\end{aligned}
$$

where we have an infinite amount of translations in the case of $J$, and an infinite amount of rotations in the case of $H$. There is no way to get back to the identity element $R_0$ if we have a group generated by rotations of 1 radian (as no multiple of $\pi$ is ever an integer). Hence, we need to define the notion of *discreteness of subgroups of* $\text{Isom}_2$ to make sense of this dichotomy.

**Definition 5.25.** A **discrete subgroup** $G$ **of** $\text{Isom}_2$ is where $\exists \varepsilon > 0$ s.t.

(1) $G$ contains no translations $\vec{a}$ with $\|\vec{a}\| < \varepsilon$, except the identity element;
(2) $G$ contains no rotations of angle $\theta$ with $\theta < \varepsilon$, except the identity element.

Now, a fair amount of work goes into characterising all discrete subgroups of isometry group $\mathrm{Isom}_2$; there are two dozens of those subgroups. We will characterise two of them, and explain the connection between them.

**Observation 5.26.** From here on, let $G \leq \mathrm{Isom}_2$ be a discrete subgroup. Then, we can describe $G$ using

- translation subgroups (subgroup consisting only of translations);
- point group (subgroup of $O_2$ consisting of linear parts of elements of $G$).

**Translation Subgroup:** Let $T \subset \mathrm{Isom}_2$ where $T = \{\text{translations}\} = \left\{t_a \mid a \in \mathbb{R}^2\right\}$. Define **lattice** $L = T \cap G = \{\text{set of translations in } G\}$. Hence, we know $L \cong (\mathbb{R}^2, +)$. A particularly powerful theorem can therefore be stated as follows:

**Theorem 5.27.** The translation subgroup $L = T \cap G$ is a discrete subgroup of $T \cong (\mathbb{R}^2, +)$. There are only three possible types of $L$:
(1) $L = \langle t_a, t_b \rangle$
(2) $L = \langle t_a \rangle$
(3) $L = \{e\}$
where $a, b \in \mathbb{R}^2$.

**Remark 5.28.** The first part of Theorem 5.27 is extremely simple: $L$ is discrete as a subgroup of a discrete subgroup. It is much harder to prove that these are the only cases of subgroups of $T$.

*Proof of Theorem 5.27. (Developing)* ■

**Point Group:** Recall that every element $a \in \mathrm{Isom}_2$ is a composition of $M \in O_2$ and translation $a \in \mathbb{R}^2$. Then, it is the case that we can define such $f : \mathbb{R}^2 \mapsto \mathrm{Isom}_2$ as $f(x) = Mx + a$. Again, we can preserve the linear part of this map, ie. define

$$\begin{array}{rccc} \pi & : & \mathrm{Isom}_2 & \longrightarrow & O_2 \\ & & Mx + a & \longmapsto & M \end{array}$$

where $\pi\left(t_a \circ M\right) = M$. Then, we can define the following:

**Definition 5.29.** The **point group** $\overline{G}$ of $G$ is defined $\overline{G} = \pi(G)$.

**Lemma 5.30.** With $G$, as given, $\overline{G}$ is discrete and finite.

**Idea 5.31.** If $\overline{G}$ contains no small rotations, then it has elements of finite order. Hence, Lemma 5.30 makes sense.

With these in mind, we want to now develop some connections between translation subgroups and point groups. We continue to denote $G$ as a discrete subgroup of $\mathrm{Isom}_2$, and $L$ as the lattice of $G$ as defined above. We make a couple of remarks below.

**Remark 5.32.** $L \lhd G$. In particular, with $\pi$ as above, we see that $L = \ker \pi$.

*Proof of Remark 5.32.* By definition, $\ker \pi = \left\{f \in L \text{ s.t. } \pi(f) = e_G\right\}$, and $e_G = I_2$. Hence, any $f \in L$ in the form $I_2 x + a$ is in the kernel of $\pi$. This is precisely the set of translations in $G$, ie. $L$. ■

**Remark 5.33.** Restrict $\pi$ to $G$. Then, we see that

$$\pi|_G : G \mapsto \overline{G}$$

meaning we have a surjective homomorphism. By First Isomorphism Theorem, we can then write

$$\operatorname{im}\pi \cong G/\ker\pi \implies \overline{G} \cong G/L$$

**Remark 5.34.** Take the tiling diagram in Artin. Observe that translations always work: take two linearly independent vectors, ie. $(0,2)$ and $(1,1)$, to generate the set of translations. Suppose rotation around the centres of these squares, in multiples of $\pi/2$ radians, are the only viable actions. Suppose further that there are reflections about some line $\ell$ passing through the centre of the squares. Then, we can define the point group

$$\overline{G} = \left\{ id, R_{\frac{\pi}{2}}, R_{\pi}, R_{\frac{3\pi}{2}}, S_0, S_{\frac{\pi}{4}}, S_{\frac{\pi}{2}}, S_{\frac{3\pi}{4}} \right\} \cong D_4$$

**Remark 5.35.** As in Remark 5.34, and by Theorem 5.22, we have that $\overline{G} \cong C_n$ or $\overline{G} \cong D_n$. In this case, since we have reflections, it must be the case that $\overline{G} \cong D_n$.

**Remark 5.36.** Obviously, we can denote each element in $\overline{G}$ using $f_i$, where $i = 1, 2, \ldots, 8$, and conclude that, by the quotient group definition of $\overline{G}$,

$$[G : L] = \big|G/L\big| = 8$$

and, for any $M \in \overline{G}$, $\pi^{-1}(M)$ is a coset of $L$, and each $f_i$ is in a different coset $Lf_i$. Hence,

$$G = Lf_1 \cup Lf_2 \cup \cdots \cup Lf_8$$

With these remarks, we can get at a fairly beautiful lemma, as follows.

**Lemma 5.37.** The point group is a symmetry group of the translation subgroup, ie. $\forall a \in L, f \in G$, $\pi(f)a \in L$, using $\pi$ as defined earlier.

**Observation 5.38.** If we have a plane of squares of 1 unit lengths, then the set of translations is exactly

$$L = \big\{(x, y) : \ x, y \in \mathbb{Z}\big\} = \mathbb{Z}^2$$

which admits reflections and rotations about some centres, hence $\overline{G} \cong D_4$, as before. If we have have a plane of rectangles, of size 2 units by 1 unit, then we have

$$L = \big\langle (2,0), (0,1) \big\rangle \implies \overline{G} \cong D_2$$

where reflections over horizontal and vertical lines are well-defined. This makes sense for us to conclude that the point group is a group of symmetries on $L$.

*Proof of Lemma 5.37.* Recall $L \lhd G$. Hence, $\forall f \in G$, $t_a \in L$, we have that $f \circ t_a \circ f^{-1} \in L$. In particular, let $f = t_b \circ M$; then, a simple calculation yields

$$\begin{aligned} f \circ t_a \circ f^{-1} &= t_b \circ M \circ t_a \circ M^{-1} \circ t_b^{-1} \\ &= t_b \circ (t_{Ma} \circ M) \circ M^{-1} \circ t_b^{-1} \\ &= t_b \circ t_{Ma} \circ (M \circ M^{-1}) \circ t_b^{-1} \\ &= t_b \circ t_{Ma} \circ t_b^{-1} = t_{Ma} \in L \end{aligned}$$

hence $\forall M \in \overline{G}$, $ML = L$. ■

**Theorem 5.39** (The Crystallographic Restriction)**.** If $L$ is a discrete, non-trivial subgroup of $(\mathbb{R}^2, +)$, and if $H \leq O_2$, where $M \in H$, $ML = L$, then $H \cong C_n$ or $H \cong D_n$, where $n = 1, 2, 3, 4, 6$.

*Proof.* The proof is extremely elegant. We suppose, by contradiction, that $a \in H$ is the shortest, nonzero vector in $L$. Suppose $n > 6$, where $n \in \mathbb{N}$. Then, $H$ contains $\rho = R_{\frac{2\pi}{n}}$, $\forall n > 6$. But, we see immediately that $\big\|\rho(a) - a\big\| < \|a\|$, which contradicts our assumption. Hence, $n \leq 6$ is the condition on $n$.

Why do we exclude $n = 5$? Suppose $\tau = R_{\frac{2\pi}{5}}$. Then, we see that $\big\|a + \tau^2(a)\big\| < \|a\|$, hence $n = 5$ is not possible. ■

## 6. More Group Theory: Abstract Group Actions and Sylow Theorems

Now, we are ready to define an array of abstract group theoretic notions. We will get into classifying groups of different orders, which sounds easy, but are extremely difficult and need extreme good care.

Up to now, we construct groups that are products in the sense of law of compositions. Recall the definition of the product group: that if $(G, \star)$ and $(H, *)$ are groups, the **(direct) product group** is the set $G \times H$ with elements in the form of Cartesian products:

$$G \times H = \big\{(g, h) : \ g \in G \text{ and } h \in H\big\}$$

where group operation takes place component-wise, ie. $(g_1, h_1) \cdot (g_2, h_2) = (g_1 \star g_2, h_1 * h_2)$. But this is not the most general setting we want. Instead, a slightly more generalised version of such product is that between a group and a set. By way of example, think of the group of isometries acting on a set of points/lines in the plane; this is a group acting on a set of points in $\mathbb{R}^2$. Hence, our goal is to define maps of the form $\pi : G \times S \to S$, where $G$ is a group and $S$ is a set, such that it makes sense to speak of $\pi$ as a map[5]!

We begin with a new set of axioms. For $\pi$ (as defined above) to make sense, we need the following facts:

(1) $e_G * s = s$ for all $s \in S$;
(2) $(gg') * s = g(g' * s)$ for all $g, g' \in G$, and $s \in S$.

We will drop $*$ as group operation from hereon, and write such action multiplicatively. These axioms actually make sense in contexts that we are interested in—for instance, $(S_n, \circ)$ acts on the set of indices $N := \{1, 2, \dots, n\}$.

### 6.1. The Basics: Group Actions.

It is important to note that for a fixed action $g \in G$, the **left multiplication map**

$$m_g : S \mapsto S$$

where $m_g(s) = gs$ is a sensible thing to write down. Why? It is a *permutation* of the set $S$, and, in particular, it is a bijective map (since $G$ is a group, $g^{-1}$ and $m_{g^{-1}}(s) = g^{-1}s$ both exist. It is worth noting that for different $g \in G$, $m_g$ differs significantly, but once we fix a $g$, $s \to s'$ is guaranteed! For this reason, we define the following objects:

**Definition 6.1.** The **orbit of** $s \in S$ is

$$O_s = \big\{s' \in S : \ s' = gs \text{ for some } g \in G\big\}$$

It is obvious that Definition 6.1 (of group actions) are equivalence classes for the following equivalence relation:

$$s \sim s' \iff s' = gs \text{ for some } g \in G$$

This relation implies that if $s \sim s'$, $O_s = O_{s'}$. Hence, the orbits of $s \in S$ partition the group, ie. orbits are disjoint sets, whose union is the entire set. There is a particular class of group actions that we encounter more often than we would think.

**Definition 6.2.** A group action on set $S$ is **transitive** if the set is nonempty, and there is exactly one orbit.

[5] It is not clear at all that $\pi$ of this form should be a map, but we will prove it is so.

**Example 6.3.**

(1) For all $n \geq 1$, the action of $S_n$ on $\{1, 2, \ldots, n\}$ is transitive: there is a permutation sending 1 to any number.

(2) For any $n$-gon with $n \geq 3$, the action is transitive: there exists rotations in $D_n$ that sends every vertex to all other vertices.

(3) The usual action of $\mathrm{SL}_2(\mathbb{R})$ on $\mathbb{R}^2 \setminus \{(0,0)\}$ is transitive: for all $a, b \neq 0$, write the matrices

$$\begin{bmatrix} a & 0 \\ b & \frac{1}{a} \end{bmatrix}, \begin{bmatrix} a & -\frac{1}{b} \\ b & 0 \end{bmatrix}$$

These are matrices of determinant 1 that sends $(1,0) \to (a,b)$ when $(a,b) \neq (0,0)$.

Another object of interest address the following concern: are there actions that fix a particular element?

**Definition 6.4.** For all actions ($g \in G$) that fixes $s \in S$, we call them **stabilizers**. Notation-wise, we write

$$G_s = \{g \in G : \; gs = s\}$$

Now we have an important proposition regading the notions of subgroups, and the classification of groups, by group actions.

**Proposition 6.5.** Let group $G$ act on set $S$, with $s \in S$, $H := \mathrm{Stab}(s)$.

(1) For all $a, b \in G$, $as = bs \iff a^{-1}b \in H \iff b \in aH$.

(2) $as = s'$. Then, $\mathrm{Stab}(s')$ is the *conjugate subgroup of* $G$:

$$H' := aHa^{-1} = \Big\{ g \in G : \; g = aha^{-1} \text{ for some } h \in H \Big\}$$

*Proof.* (1) This result is trivially true, since $a, b$ fixes $s \in S$.

(2) We show both inclusions.

If $g \in aHa^{-1}$, then $g = aha^{-1}$ for some $h \in H$. This implies $gs' = (aha^{-1})s' = (aha^{-1})as = ahs = as = s'$, since $h$ stabilizes $s$; ie. $H' \supset aHa^{-1}$.

$a \in G$, so $a^{-1} \in G$ by definition. Then, $s = a^{-1}s'$. Reversing the roles of $s$ and $s'$ gives $a^{-1}H'a \subset H$, ie. $H' \subset aHa^{-1}$.

This concludes the proof. ■

A particularly interesting class of group actions is those on *cosets*. This naturally arises because a large class of groups consists of quotient groups, of the form $(G/H, *)$, where $H \lhd G$. But, regardless of the normality of $H$ in $G$, $G$ always acts naturally on the set $G/H$. In particular, if $g \in G$, and $[C] \in G/H$ is a coset, then $g[C] = [gC]$ is another coset! Therefore, we say that if $[C] = [aH]$, then $g[C] = [gC] = [gaH]$. These facts are generalised in the following proposition (with a formal proof that follows).

**Proposition 6.6.** Let $H \leq G$.

(1) The left multiplication of $G$ on $G/H$ is transitive.

(2) $\mathrm{Stab}([H]) = H$.

*Proof.* (1) Since $gH = g \cdot H$ (the RHS being the left multiplication operation)[6], so the action of $G$ by left multiplication on the coset space $G/H$ has one orbit.

[6] Some people write the left action as $g.(xH) = (gx)H$; the choice is entirely a preference.

(2) We prove the equivalent statement: $gH = H \iff g \in H$. One way is obvious: assuming $gH = H$, then $g = g \cdot e_H \in gH = H$. Another way is a little less so; we claim that $g \sim e$ as above is an equivalence relation. Since equivalent elements represent the same equivalence class, we have $gH = eH = H$. *Why* is $g \sim e$ an equivalence relation? Rewrite the equivalence relation as

$$g \sim f \iff g^{-1}f \in H$$

and this is indeed an equivalence relation! Check

- *Reflexive*: $g \sim g$ since $g^{-1}g = e \in H$!
- *Symmetric*: if $g \sim f$, then $g^{-1}f \in H$. Notice $f^{-1}g = (g^{-1}f)^{-1} \in H$, since inverses are closed in $H$ by definition. Hence, $f \sim g$.
- *Transitive*: Suppose $g \sim i$ and $i \sim j$. Then,

$$(g^{-1}i)(i^{-1}j) = g^{-1}(ii^{-1})j = g^{-1}j \in H$$

by associativity.

It remains to prove that this equivalence relation induces left cosets as its equivalence classes. Let $[x] = \{g \in G : x \sim g\}$. Pick $y \in [x]$; then, by definition, $x \sim y$ means $x^{-1}y \in H$. So

$$y = x(x^{-1}y) \in xH$$

ie. $xH \supset [x]$.

Now, pick $y \in xH$. Then, $y = xh$ for some $h \in H$. Finally, by the fact that $x \in G$, $x^{-1}$ exists and is in $G$! Hence,

$$x^{-1}y = h \in H \implies xH \subset [x]$$

As such, $[x] = xH$.

This concludes the proof. ■

**Remark 6.7.** Part (1) of Proposition 6.6 has a banal, but instructive, special case: suppose $G$ acts on itself by left multiplication. Since $g = g \cdot e$, it is then the case that if $H$ is the trivial subgroup, we conclude that the action is transitive.

Now, we encounter an extremely important theorem that describes the relationship between arbitrary group operation and the operations on cosets. This is called the **orbit-stabilizer theorem**.

**Theorem 6.8** (Orbit-Stabilizer Theorem)**.** Let $S$ be a set on which group $G$ acts, $s \in S$, $H \coloneqq \mathrm{Stab}(s)$, $O_s$ is the orbit of $s$. Then, there exists a bijective map $\varepsilon : G/H \mapsto O_s$ such that $[aH] \to as$. Furthermore, this map is compatible with the operations of the group: $\varepsilon(g[C]) = g\varepsilon([C])$ for every coset $C$ and $g \in G$.

*Proof.* It is clear that $\varepsilon$ as defined in the theorem is compatible with the operation of the group. It is also clear that if such map exists at all, it is injective as well as surjective by definitions. We need to prove that such map is *actually* a map: in particular, for all $g \in G$, we need to verify if the rule $[gH] \to gs$ defines a map. Hence, if $a, b \in G$, we need to show that if $aH = bH$, then $a = b$. By an equivalence result introduced earlier in the class, we see that $aH = bH \iff a^{-1}b \in H$. Since $H$ stabilizes $s$, we have $a^{-1}bs = s \implies bs = as$.

This map is obviously injective. For surjectivity, notice that, by definition of the map $\varepsilon : [gH] \to gs$, there is a coset for every $s$ mapped to by the map. ■

From here, we get the important counting formula for groups, far more powerful than the version we were given in the earlier section. By the previous counting formula, we replace the index of $H$ in $G$ by $\big|G/H\big|$, and the formula then becomes $|G| = |H|\big|G/H\big|$. In this case, though, we require finiteness on the set. Then, we have the following result.

**Proposition 6.9.** Let $S$ be a finite set on which group $G$ acts, and $G_s$, $O_s$ be stabilizer and orbit of $s \in S$, respectively. Then,

$$|G| = |G_s||O_s|$$

*Proof.* Follows from the earlier formula and Theorem 6.8. ■

**Remark 6.10.** Two more notions of counting:

(1) $|O_s| = [G : G_s]$
(2) $|S| = |O_1| + \cdots + |O_n|$

obviously, these counting formulae are only meaningful if $S$ is finite, and $\big|\{O_s\}_{s\in S}\big| < \infty$. The second formula is intuitively true: that if the set can be decomposed into cycles, it should be the size of their sum. It remains to prove that these orbits are *disjoint*. But, the tricky part is the first formula: proving that $|O_s| = [G : G_s]$ takes some time. Keeping the same notation, we prove these facts in the next theorem.

**Theorem 6.11** (Fundamental Theorem of Group Actions)**.** Let $G$ act on $X$.

(1) Different orbits of the action are disjoint.
(2) For each $x \in X$, $G_x \leq G$, and $G_{gx} = gG_xg^{-1}$ for all $g \in G$.
(3) For $x \in X$, $gx = g'x \iff g, g'$ lie in the same left coset of $G_x$. In particular, if $x$ and $y$ are in the same orbit, then $\{g \in G : gx = y\}$ is a left coset of $G_x$, and different left cosets of $G_x$ correspond to different points in $O_x$, hence

$$|O_s| = [G : G_s]$$

*Proof.* ■

6.2. **The Basics: Permutation Representation.** There are various ways in which a group $G$ can act on a set $S$; in particular, we are interested in permutations, since it is a left multiplication of $G$ on the set $G/H$, as we've stated before.

**Definition 6.12.** A **permutation representation of a group** $G$ is a homomorphism from the group to the symmetric group, defined

$$\psi : G \mapsto S_n$$

The following is a shocking result (at least, at first).

**Proposition 6.13.** Let $G$ be a group. There exists a bijection between operations of $G$ on the set $N := \{1, 2, \ldots, n\}$ and the permutation representations $G \mapsto S_n$.

*Proof.* For a given $G$ and its associated action, define $\psi(g) = m_g = gs$, ie. left multiplication map just as before. The associative property shows that

$$m_g(m_hj) = g(hj) = (gh)j = m_{gh}j$$

so the map is a homomorphism. Conversely, given $\psi$, a permutation representation, the same formula defines an operation of $G$ on $S$. ■

Note that Proposition 6.13 has nothing to do with the actual set it acts on. Let Perm($S$) be the group of permutations over some arbitrary set $S$. The map $\psi : G \mapsto \text{Perm}(S)$ is also a permutation representation of $G$. A direct consequence of this observation is the following:

**Corollary 6.14.** Let Perm($S$) be as defined above, and $G$ be a group. There exists a bijection between actions of $G$ on $S$ and the permutation representations $\psi : G \mapsto \text{Perm}(S)$.

A few more words on actions and permutation representation:

- In general, permutation representations do not need to be injective or surjective.
- In general, permutation representations are rarely surjective, because the order of Perm($S$) tends to be *very* large.

But from here arises two interesting facts.

**Definition 6.15.** An injective group action is called a **faithful** action.

Equivalently, an operation must have the property that left multiplication by $g$ is not the identity map unless $g = e_G$! In other words, the only element $g \in G$ such that $gs = s$ for all $s \in S$ is $g = e_G$. For example, the operation of the group of isometries $M$ on the set $S$ of equilateral triangles in the plane is faithful, since the only isometry that maps equilateral triangles to themselves is the identity map.

Below is an example of a permutation representation that is surjective. Notice the group $GL_2(\mathcal{F}_2)$ of invertible matrices with mod 2 coefficients is isomorphic to $S_3$. Let $G := GL_2(\mathcal{F}_2)$ and $F := \mathcal{F}_2$. Then, it is the case that $F^2$ contains four vectors: the canonical basis, 0 vector, and $(1, 1)$. The group then acts on the set of nonzero vectors $\{e_1, e_2, e_1 + e_2\}$. This gives us the permutation representation $\psi : G \to S_3$. Obviously, the identity is the only matrix that fixes both $e_1$ and $e_2$, so the operation of $G$ on $S$ is *faithful.* The columns of an invertible matrix must be an ordered pair of distinct elements of $S$, and there are six such pairs. Since $|S_3| = 6$ as well, $\psi$ is an isomorphism.

Now, as we can see, the basic idea in any group action is that the elements of a group are viewed as permutations of a set in such a way that composition of the corresponding permutations matches multiplication in the original group

We summarise the reason for studying group actions to begin with. The symmetric groups $S_n := (\text{Sym}(X), \circ)$, alternating groups $A_n := (\text{Alt}(X), \circ)$ (which is a subgroup of $S_n$), and (for $n \geq 3$) dihedral groups $(D_n, \circ)$ behave, by their very definition, as permutations on certain sets. However, more abstractly, if we are given any set $X$ (not necessarily the set of vertices of a square), then the set Sym($X$) of all permutations of $X$ is a group under composition, and the subgroup Alt($X$) of even permutations of $X$ is a group under composition. If we list the elements of $X$ in a definite order, say as $X = \{x_1, x_2, \ldots, x_n\}$, then we can think about Sym($X$) as $S_n$ and Alt(X) as $A_n$, but a listing in a different order leads to different identifications of Sym($X$) with $S_n$ and Alt($X$) with $A_n$. This notion of "abstract" symmetric groups lead us to a myriad of interesting results.

6.3. **Cayley's Theorem.** That a group $G$ acts on a set is interesting, but how about on itself? This turns out to be surprisingly large field of study. There are two ways can analyse such action: one by left multiplication, and another, far more subtly and importantly, by conjugation.

As we have seen, if we define the left multiplication map $\varphi : G \times G \mapsto G$ by sending $(g, x) \to gx$, then this is a transitive, faithful action. In particular, $\mathrm{Stab}(g)$ is the trivial subgroup of $G$, and the permutation representation is injective. An obvious comparison of such finite group is then the permutation group. This leads us to an important result in finite group theory.

**Theorem 6.16** (Cayley)**.** Every finite group is isomorphic to a subgroup of permutation group. Every group of order $n$ is isomorphic to a subgroup of $S_n$; in other words, every finite group $G$ can be embedded in a symmetric group.

*Proof.* Since the action is faithful, $G$ is isomorphic to its image in $\mathrm{Perm}(G)$. If $|G| = n$, then $\mathrm{Perm}(G) \cong S_n$ (specifically, its image in $S_n$).

The embedding part is more difficult to prove. To each $g \in G$, define the left multiplication function $m_g : G \mapsto G$, where $m_g(x) = gx$ for each $x \in G$. Each $m_g$ is a permutation of $G$ as a *set*, with inverse $m_{g^{-1}}$, hence $m_g$ belongs to $\mathrm{Sym}(G)$. Since $m_{g_1} \circ m_{g_2} = m_{g_1 g_2}$ for all $x \in G$, ie. $g_1(g_2 x) = (g_1 g_2)x$ for all $x \in G$, associating $g$ to $m_g$ gives a homomorphism of groups, $G \to S_n$. This homomorphism is one-to-one, since $m_g$ determines $g$ (after all, $m_g(e) = g$). Therefore the correspondence $g \to m_g$ is an embedding of $G$ as a subgroup of $S_n$. ■

As interesting as Theorem 6.16 is, it is often not useful because $|S_n| = n!$, which grow exponentially as $n$ grows slightly. Hence, the left multiplication map becomes some sort of a chore to use. Like we said earlier, there is a far better map to use: the conjugation map. We introduce the concept in full here.

6.4. **Conjugation and the Class Equation.** Sans the basic definitions (conjugation by elements, conjugate subgroup, conjugacy class), we make direct links between conjugates and abstract group theoretic concepts. Here are some facts (some provided sans proofs) about conjugates.

**Fact 6.17.** Conjugacy defines an equivalence relation on the group $G$. This is not difficult to check by definition.

**Fact 6.18.** Conjugacy works with subgroups: let $G$ be a group and $H \leq G$. The conjugacy subgroup is as defined earlier.

**Fact 6.19.** In a group $G$, for all $n \in \mathbb{Z}$, $(xgx^{-1})^n = xg^n x^{-1}$.

*Proof.* By induction. ■

**Fact 6.20.** Any two elements in a conjugacy class have the same order.

*Proof.* It suffices to prove that $g$ and $xgx^{-1}$ have the same order.

If $g^n = e$: then $(xgx^{-1})^n = xg^n x^{-1} = xx^{-1} = e$, where first equality follows from Fact 6.19.

If $(xgx^{-1})^n = e$: then $(xgx^{-1})^n = e \implies g^n = x^{-1}ex = e$. ■

**Fact 6.21.** The converse of Fact 6.20 is just not true in general!

**Fact 6.22.** Let $H$ be a cyclic subgroup of $G$. Then, every conjugate subgroup to $H$ is cyclic.

*Proof.* Let $H = \langle y \rangle$ for some $y \in G$. Then,

$$\begin{aligned} gHg^{-1} &= \left\{ ghg^{-1} :\ h \in H \right\} \\ &= \left\{ gy^n g^{-1} :\ n \in \mathbb{Z}^+ \right\} \\ &= \left\{ (gyg^{-1})^n :\ n \in \mathbb{Z}^+ \right\} \end{aligned}$$

so the generator of $gHg^{-1}$ is the conjugate of the generator of $H$ by $g$. ■

**Fact 6.23.** Just like orbits are disjoint, conjugacy classes are disjoint.

*Proof.* We prove the equivalent: that if two conjugacy classes overlap, they are the same.

Pick arbitrary $g, h \in G$; then, it suffices to prove that an element conjugate to $g$ is also conjugate to $h$. Since conjugacy classes overlap, for some $x, y \in G$,

$$xgx^{-1} = yhy^{-1} \implies g = x^{-1}yhy^{-1}x = (x^{-1}y)h(x^{-1}y)^{-1}$$

so $g$ is conjugate to $h$. Furthermore, for any element $z \in G$ conjugate to $g$, we have

$$\begin{aligned} zgz^{-1} &= z(x^{-1}y)h(x^{-1}y)^{-1}z^{-1} \\ &= (zx^{-1}y)h(zx^{-1}y)^{-1} \end{aligned}$$

so any element conjugate to $g$ is also conjugate to $h$.

For the other direction, set $h = (y^{-1}x)g(y^{-1}x)^{-1}$ and similar calculations follow. ■

**Consequence 6.24.** The direct consequence to Fact 6.23 is that every element in a conjugacy class is a *representative* of that class. As such, instead of looking for conjugates to $g$, we look for all the $x \in G$ that are conjugate too $g$ that we fix. The set $\left\{xgx^{-1}\right\}$ is therefore interesting upon fixing an action $g$. Furthermore, by Theorem 2.43, such map is an automorphism.

**Definition 6.25.** The map $\gamma_x$ defined in Theorem 2.43 is called an **inner automorphism**.

Here is an important property of inner automorphisms—that knowing something about it tells us something about the automorphism. We fix the notation $\mathrm{Aut}(G)$ for the automorphism associated with group $G$, and define the following object:

**Definition 6.26.** The **centre** of a group $G$, denoted $C(x)$ for some $x \in G$, is the set of elements that commute with every other element in the group:

$$C(x) = \{g \in G :\ xg = gx\}$$

**Definition 6.27.** The stabilizer of conjugation of $x$ by $g$ is called the **centralizer** of $x$, denoted $Z(x)$, defined as

$$Z(x) = \left\{ g \in G :\ gxg^{-1} = x \right\} = \{g \in G :\ gx = xg\}$$

**Theorem 6.28.** If $G$ is a group with trivial centre, then $\mathrm{Aut}(G)$ has a trivial centre.

*Proof.* Let $\varphi \in \mathrm{Aut}(G)$. Suppose $\varphi$ commutes with all other automorphisms, and $\gamma_x$ be an inner automorphism. Then,

$$\begin{aligned}(\varphi \circ \gamma_x)(g) &= \varphi(\gamma_x(g)) \\ &= \varphi(xgx^{-1}) = \varphi(x)\varphi(g)\varphi(x)^{-1} \\ (\gamma_x \circ \varphi)(g) &= \gamma_x(\varphi(g)) \\ &= x\varphi(g)x^{-1}\end{aligned}$$

and note that $\varphi$ commutes means

$$\varphi(x)\varphi(g)\varphi(x)^{-1} = x\varphi(g)x^{-1} \iff x^{-1}\varphi(x)\varphi(g) = \varphi(g)x^{-1}\varphi(x)$$

so $x^{-1}\varphi(x)$ commutes with $\varphi(g)$. In other words, $x^{-1}\varphi(x) \in Z(x)$, the *centraliser of* $x$, which is guaranteed by the fact that $\varphi$ is onto. Hence, $x^{-1}\varphi(x) = e \implies \varphi(x) = x$, ie. $\varphi$ is the identity map. ■

**Remark 6.29.** With these definitions, we arrive at yet another way to count the order of a finite group:

$$|G| = \big|Z(x)\big|\big|C(x)\big| = |\text{ centraliser }||\text{ centre }|$$

**Remark 6.30.** For finite group $G$, by the formula above, we have yet another way to count $|G|$:

$$|G| = \sum_{\text{all conjugacy classes } C} |C| \tag{6.1}$$

Note that, by orbit-stabilizer theorem, every $|C|$ divides $|G|$. (6.1) is so important that it goes by a name: it is the **class equation**. Below is a worked example, an application of (6.1).

**Example 6.31.** Let $SL_2(\mathcal{F}_3)$ be the set of invertible matrices with mod3 coefficients, with determinant 1. Listing the elements of this group and writing down their centralisers is fine, but there is a far easier way of finding the class equation. We start by a matrix $A = \begin{bmatrix} & -1 \\ 1 & \end{bmatrix}$, and, by definition, if $A$ is the centraliser, then $AB = BA$ for all $B \in SL_2(\mathcal{F}_3)$. By the multiplicative linearity of determinants, $\det AB = \det A \det B = \det B = 1$. Hence, by the conditions, we have the linear system $a^2 + c^2 = 1$ for such $A$. Therefore, $\big|Z(A)\big| = 4$, $\big|C(A)\big| = 6$, giving us the order of the group. But, if we are following the class equation, we can decompose this further by looking for matrices of different characteristic polynomial[7]. At the end, we have the class equation for $SL_2(\mathcal{F}_3)$:

$$\big|SL_2(\mathcal{F}_3)\big| = 1 + 1 + 4 + 4 + 4 + 4 + 6 = 24$$

[7] This is because different eigenvalues are associated to matrices from different conjugacy classes.

6.5. **$p$-groups, Sylow Theorems.** Up to this point, we have come across a few classifications of groups of different orders. For instance, we know groups of order 12 (which we will explore in a separate section) are isomorphic to $C_3 \times C_4$. Now, **$p$-groups** are groups of order prime $p$. We will introduce $p$-groups, and move onto one of the most fundamental results in late nineteenth century: the Sylow theorems, which concerns classifying a certain type of $p$-(sub)groups.

**Theorem 6.32.** The centre of $p$-group is not trivial.

*Proof.* By definition of the $p$-group $X$, suppose $|X| = p^e$, where $e \geq 1$. By the class equation, and the orbit-stabilizer theorem, every term on the RHS divides $|X| = p$. Suppose the group has a trivial centre; then, the class $C_1$ with only the identity element has order 1. However, rest of the classes must have orders divisible by $p$, yielding the class equation

$$|X| = p^e = 1 + \sum (\text{multiples of } p)$$

which yields an immediate contradiction. ■

There is something special about groups of order $p^2$.

**Proposition 6.33.** Every group of order $p^2$ is abelian.

*Proof.* Recall that a group is abelian iff group equals its centre. By Theorem 6.32, the centre of $G$ (a group of order $p^2$) is nontrivial; by a variation of the counting formula for finite groups, the order of the centre divides the order of the group. This implies the centre is either of order $p$ or $p^2$. Two cases then arise:

(1) <u>If $|Z| = p^2$</u>: we're done—since $|G| = |Z||C| \implies |G| = |Z| \implies G = Z$.
(2) <u>If $|Z| = p$</u>: we must either show that $|C| = p$, or $|Z| = p$ is impossible. Suppose $|Z| = p$, and let $x \in G$ but $x \notin Z$. Obviously, $x, Z \in Z(x)$, so $|Z| < \big|Z(x)\big|$ or $Z(x) \supset Z$. Since $|G| = \big|Z(x)\big|\big|C(x)\big|$, it implies $\big|Z(x)\big| \mid |G|$, ie. $\big|Z(x)\big| = p^2$. But this means $Z(x) = G$, hence $x \in Z$. This results in a contradiction.

This concludes the proof. ■

**Corollary 6.34.** Every group of order $p^2$ is either cyclic or the product of two cyclic groups of order $p$.

*Proof.* Let $G$ be a group of order $p^2$. Two possibilities:

(1) <u>If $G$ contains an element of order $p^2$</u>: $G = |x|$, where $x \in G$ satisfies $|x| = p^2$.
(2) <u>If $G$ does not contain an element of order $p^2$</u>: by corollary to Lagrange's theorem, every non-identity element in $G$ has order $p$. Let $x, y$ be elements of order $p$ such that $x \notin \langle y \rangle$; by an earlier proposition on product groups, we have that $G \cong \langle x \rangle \times \langle y \rangle$.

This concludes the proof. ■

Finally, we need a side remark on **normalisers**. Let $G$ be a group and $H \leq G$. Let us consider the orbit of $[H]$ under conjugation by $G$; this is exactly the set of conjugate subgroups $[gHg^{-1}]$. In this scenario, the *stabiliser of* $[H]$ is called the **normaliser of $H$**, defined

$$N(H) = \Big\{ g \in G : \, gHg^{-1} = H \Big\} \tag{6.2}$$

and for all $|G| < \infty$, we have the counting formula

$$|G| = \big|N(G)\big|\,[G : N(H)] \tag{6.3}$$

Now we are ready to study Sylow $p$-subgroups.

**Definition 6.35.** Let $|G| = n$. Let $p$ be a prime number with $p \mid n$. Let $p^e$ be the largest power of $p$ such that $p^e \mid n$ (ie. $n = p^e m$ for some $m \not| p$). Then, all subgroups $H \leq G$ such that $|G| = p^e$ are called the **Sylow $p$-subgroups of $G$**.

Then, we have the following famous theorems, called **Sylow theorems**.

**Theorem 6.36** (First Sylow Theorem)**.** *Suppose $G$ is a finite group. Let $p \mid |G|$ ($p$ divides the order of $G$). Then, $G$ contains a Sylow $p$-subgroup.*

**Theorem 6.37** (Second Sylow Theorem)**.** *Let $G$ be a finite group with $p \mid |G|$.*

(1) *Sylow $p$-subgroups are conjugate subgroup for any fixed $p$.*
(2) *Every $H \leq G$ that is a $p$-group is contained in a Sylow $p$-subgroup.*

**Theorem 6.38** (Third Sylow Theorem)**.** *Let $G$ be a finite group of order $k$, where $p \mid k$ ($k = p^e m$ and $m \not| p$). Let $s$ be the number of Sylow $p$-subgroups of $G$. Then, $s \mid m$ and $s \equiv 1 \mod p$.*

In stating these theorems, we will come up with ways to prove them distinctly. The proofs of Theorem 6.37 and Theorem 6.38 are self-contained; we need two lemmas for proving Theorem 6.36.

**Lemma 6.39.** *Let $[U]$ is the set of cosets from left multiplication fof $G$ on the set of $U \subset G$, where $G$ is a finite group. Then, $\big|\mathrm{Stab}([U])\big|$ divides $|U|$ and $|G|$.*

*Proof.* Let $H \leq G$. Any $H$-orbit element $u \in G$ for multiplication by $H$ is the right coset $Hu$. Let $H$ be the stabiliser of $[U]$. This means $H$ permutes $[U]$, ie. $U$ is partitioned into $H$-orbits, which are the aforementioned right cosets. Each coset has order $|H|$, implying $|H| \mid |U|$. Since $H \leq G$ and $|G| < \infty$, $|H| \mid |G|$ by corollary to Lagrange's theorem. ■

**Lemma 6.40.** *Let $n = p^e m$ for all $e > 0$ and $p \mid m$. Then, $N = \#$ of subsets of order $p^e$, does not divide $p$.*

*Proof.* Note that

$$N = \binom{n}{p^e} = \frac{n!}{(n-p^e)!(p^e)!} = \frac{n(n-1)\cdots(n-p^e+1)}{p^e(p^e-1)\cdots(p^e-k)\cdots 1}$$

it remains to prove that $N \not\equiv 0 \mod p$, because $p$ divides $(n-k)$ in the numerator of $N$; in addition, $(n-k)$ is divided by $(p^e-k)$ the same number of times. In other words, for $k = p^i \ell$ where $p \not| \ell$. It implies $i < e$, and $(m-k) = (p^e-k)$ and $(n-k) = (p^e m - k)$ are divisible by $p^i$, but not $p^{i+1}$. ■

Now we begin to prove the Sylow theorems.

*Proof of Theorem 6.36.* To show the existence of a Sylow $p$-subgroup, we will find a $H$-orbit whose left multiplication by $G$ gives a stabiliser that has exactly order $p^e$.

Let $\mathcal{S}$ be a collection of subsets of $G$, each of order $p^e$. The order of $S$ can be decomposed into orbits for left multiplication by $G$ (by the class equation):

$$N = |S| = \sum_{O \in O_s} |O|$$

By Lemma 6.40, $N$ does not divide $p$, ie. there exists one orbit whose order is not divisible by $p$. Without loss, let such orbit be denoted $O_{[U]}$, and its order $|O|_{[U]} = m$, where $p \not| m$. Let $H \coloneqq \text{Stab}([U])$, and $|H| \mid |U|$ and $|H| \mid |G|$ by Lemma 6.39. This implies $|H| = p^e$. By the counting formula, $|G| = |H| \left| O_{[U]} \right| = p^e m$. Such $H$ is a Sylow $p$-subgroup of $G$. ■

*Proof of Theorem 6.37.* From Theorem 6.36, we know for any finite group $G$ where $p \mid |G|$, Sylow $p$-subgroups exist; take $H, K \leq G$ as the Sylow $p$-subgroups. We show that $H'$ (the conjugate subgroup of $H$) contains $K$, which proves part (b); furthermore, we will show that if $K$ is also a Sylow $p$-subgroup, then it equals the conjugate subgroup $H'$, which proves part (a).

To start, we choose a set $\mathcal{C}$ such that the following properties hold:

- $p \not| |\mathcal{C}|$,
- Group action is transitive,
- There exists a $c \in \mathcal{C}$ such that $\text{Stab}(c) = H$.

Such a set definitely exists, since $H \leq G$ and $H =$ left cosets of $H$ in $G$ (which obviously satisfy all of the above).

Furthermore, because $p \not| |\mathcal{C}|$, so restricting group $G$ on $\mathcal{C}$ to $p$-subgroup $K$ is possible. By a fixed point theorem, there exists a $c' \in \mathcal{C}$ that fixes the operation on $K$.

Since group action is transitive by assumption, $c' = cg = gc$ for some $g \in G$. Hence, $\text{Stab}(c') = [gHg^{-1}]$ is the conjugate subgroup. Since $K$ fixes $c'$, $\text{Stab}(c') \supset K$ as the result. ■

*Proof of Theorem 6.38.* Let $|G| = p^e m$ and $S = \#$ Sylow $p$-subgroups in $G$. From Theorem 6.37, we know the operation of $G$ is transitive. Hence, $\text{Stab}([H]) = N(H)$ (the normaliser of $H$). Counting formula gives $|S| = \left|N(H)\right| [G : N(H)] = [G : N(H)]$. Furthermore, since $N \supset H$, and $[G : H] = m$ by assumption, $S \mid m$. Now, let

$$\mathscr{S} \coloneqq \bigcup_{O_{[H]} \in O_s} O_{[H]}$$

where orbits $O_{[H]}$ are for conjugation by $H$. The $H$-orbits have order 1. Since $H$ is a $p$-subgroup, it remains to prove that $S \equiv 1 \mod p$, which we do by showing that no element in $S$ except $[H]$ is fixed by $H$.

Suppose $H'$ is a $p$-Sylow subgroup, and conjugation by $H$ fixes $[H']$. Then, $H \subset N'(H')$ by construction. This implies that $H$ and $H'$ are $p$-subgroups of $N'$. By Theorem 6.37, Sylow $p$-subgroups $H$ and $H'$ are conjugate $p$-subgroups of $N'$. In addition, by the fact that the group action is transitive, $H' \lhd N'$. This means $H' = nH'n^{-1}$, which is equivalent to $H' = H$. ■