Linear Algebra

Nishant Panda

7205 Hart Lane, Apt 2026, Austin Tx
E-mail address: nishant.panda@gmail.com

2010 *Mathematics Subject Classification.* Primary

Key words and phrases. vector spaces, duality, linear transformation, bi-linear

Contents

Nomenclature

$\mathcal{U}^0$ Annhilator of $\mathcal{U}$
$\mathbb{C}$ Complex Numbers
$\dim \mathcal{V}$ Dimension of $\mathcal{V}$
V^* Dual Vector space of $\mathcal{V}$
$\mathbb{F}$ Field
$\hom(V, W)$ Set of all linear transformations from $\mathcal{V}$ to $\mathcal{W}$
$\operatorname{Im} T$ Image of the linear transformation T
$\ker T$ Kernal of the linear transformation T
$\mathcal{M}_{m\times n}(\mathbb{F})$ Set of all matrices of size $m \times n$ whose elements belong to $\mathbb{F}$
$\mathcal{U}^\perp$ Orthogonal complement of $\mathcal{U}$
$\mathbb{R}$ Reals
$\boldsymbol{v}$ Element of a vector space
$V/\mathbb{F}$ Vector Space over field
$\mathcal{V}$ Vector Space
$\mathbb{Z}$ Integers
$\mathbf{0}_\mathcal{V}$ zero vector of a vector space $\mathcal{V}$

Chapter 1

Basic group theory

In this chapter we will cover the essentials of abstract algebra. The notion of algebra is very fundamental and we will study this in an abstract setting. Our main motivation comes from the number systems that are familiar to us. We all know how the integers (set of integers is denoted by $\mathbb{Z}$) behave. If we add two integers we get a new integer. If we add any integer to 0 we get the same integer and for any integer, we can subtract it from itself to get back 0. Adding in this sense is an algebraic operation. Our goal in this chapter will be to study ways in which we can abstract this notion.

1.1. Groups: fundamental algebraic structure

Definition 1.1.1. *An operation $\star$ on a set A is a function*

$$\star : A \times A \to A$$

For example if $A = \mathbb{Z}$, then $+, *$ are operations. Few characteristics are noted below:

- If for any $a, b \in A$, $a \star b = b \star a$ then the operation is commuatative.
- If for any $a, b, c \in A$ $a \star (b \star c) = (a \star b) \star c$ then the operation is associative.
- If for any $a \in A$ there is an element $e \in A$ such that $a \star e = e \star a = a$, then e is called the identity element of A w.r.t the operation $\star$.
- If for every $x \in A$ there is an element $a \in A$ such that $a \star x = x \star a = e$, then a is the inverse of x w.r.t the operation $\star$.

An *algebraic structure* is a set with one or more operations defined on it. One of the most simplest and useful algebraic structure is the group.

Definition 1.1.2. *A group is a* triple $(G, \star, e)$ *consisting of a set* G*, an operation* $\star$ *defined on it and an identity element* e *such that*

(1) G *is* closed *under* $\star$ *i. e. for every* $a, b \in G$ $a \star b \in G$.
(2) G *is associative.*
(3) For any element $a \in G$ *there is an element* $e \in G$ *such that* e *is an identity element.*
(4) For every element $a \in G$ *there is an inverse element denoted by* a^{-1} *such that* $a \star a^{-1} = a^{-1} \star a = e$.

When the operation is addition we will denote it by $+$ and the identity element by 0. When the operaton is multiplication we will denote it by $\cdot$ and the identity element by 1. Thus $(\mathbb{Z}, +, 0)$ is a group. This is called the **additive group of integers**. Note that $(\mathbb{Z}^+, \cdot, 1)$, where $\mathbb{Z}^+ = \{n \in \mathbb{Z} : n > 0\}$, is not a group because any integer greater than 1 does not have a multiplicative inverse. Another group is the congruence class of integers for example $(\mathbb{Z}_3, +, [0]_3)$ is a group. See A.1 for this definition and some basic properties of integers. We will define congruence class later in the chapter.

Usually, with a group of finite elements we denote the operation with a *group table*. As an example the group table denoting modular arithmetic in the group $\mathbb{Z}_2$ is given below.

$+$	$[0]$	$[1]$
$[0]$	$[0]$	$[1]$
$[1]$	$[1]$	$[0]$

In general for an operation $\star$, the group table will be:

$\star$	$\dots$	y	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
x	$\dots$	$x \star y$	$\dots$

Note that *commutative* property is not required as an axiom of group theory. For example consider the set of $n \times n$ matrices $\mathcal{M}_{n\times n}(\mathbb{F})$ where $\mathbb{F} = \mathbb{R}$ that are invertible, with the operation being matrix multiplication. This is an important set and we denote it by $GL_n(\mathbb{F})$. We will talk more about $GL_n(\mathbb{F})$ when we discuss invertible matrices. It is easy to check that with the Identity matrix such a set is a group, however it is not commutative.

Whenever commutative property holds we denote the group as an *Abelian* group.

Definition 1.1.3 (Abelian group)**.** *A commutative group is called an Abelian group.*

For example both $(\mathbb{Z}, +, 0)$ and $(\mathbb{Z}_m, +, [0]_m)$ are Abelian groups.

As an example of an Abelian group on sets, consider the *symmetric difference* between two sets A, B defined as $A + B = (A - B) \bigcup (B - A)$. Here, $A - B$ means all the elements of A that are not in B. Consider any set D and the power set of D defined by $\mathscr{P}(D) = \{A : A \subset D\}$. Note the $A + \emptyset = A$. Thus we can form the group $(\mathscr{P}(D), +, \emptyset)$. Note that for every $A \in \mathscr{P}(D)$, $A^{-1} = A$. Next, we will consider some elementary properties of groups.

Proposition 1.1.1 (Uniqueness of identity and inverse elements)**.** *In any group $(G, \star, e)$, e is the unique identity element. Also, for any $a \in G$, there is an unique inverse element a^{-1} such that $a \star a^{-1} = e$.*

Proof. Let e_1 also be an identity element. Then $e_1 \star e = e$ because e is an identity element of G and similarly $e_1 \star e = e_1$ since e_1 is also an identity element of G. But this means $(e_1, e) \mapsto e$ and e_1. Thus $e = e_1$.

Consider an element $a \in G$. Let a^{-1} and a_1^{-1} be two inverse element of a. Consider $\left(a^{-1} \star a\right) \star a_1^{-1}$. Since a^{-1} is an inverse element we get $e \star a_1^{-1} = a_1^{-1}$. Also since a group is associative $\left(a^{-1} \star a\right) \star a_1^{-1}$ is equal to $a^{-1} \star \left(a \star a_1^{-1}\right)$ which is equal to $a^{-1} \star e = a^{-1}$ since a_1^{-1} is also an inverse element of a. Thus $a^{-1} = a_1^{-1}$. □

Remark 1.1.1. *From now on we will not explicitly denote the operation by $\star$. Thus $a \star b$ will be written as ab. We will call this the **product** of two elements a, b in the group. This doesn't mean that the operation is multiplication, rather we are being implicit in defining the operation in the group. If the operation is addition $(+)$ then ab will mean $a + b$. Similarly if the opereation is compostion, ab will mean $a \circ b$.*

The most basic rule of calculation in groups is the cancellation law, which allows us to cancel common factors in equations.

Theorem 1.1.1 (Cancellation Laws in Groups)**.** If G is a group and $a, b, c \in G$, then
(1) $ab = ac$ implies $b = c$ and
(2) $ba = ca$ implies $b = c$

Proof. *Multiply* (operate) both sides of the equation (1) $ab = ac$ by a^{-1} on the left to get $b = c$.

$$\begin{aligned} ab &= ac \\ a^{-1}(ab) &= a^{-1}(ac) \\ (a^{-1}a)b &= (a^{-1}a)c \\ eb &= ec \\ b &= c \end{aligned}$$

Similarly multiply both sides of the equation (2) $ba = ca$ by a^{-1} on the right to get $b = c$. □

Note that we cannot apply cancellation laws to the equation $ab = ca$ because we may not have a commutative group.

The next theorem will be stated without proof.

Theorem 1.1.2. If G is a group and a, b are elements of G then,

$$ab = e \quad \text{implies} \quad a = b^{-1} \quad \text{and} \quad b = a^{-1}$$

The next theorem states some important facts about inverses. The first being that the product of the inverse is the inverse of the products and the second being

that the inverse of the inverse of an element is the element. Again it is to be noted that the term *product* here just means the operation.

Theorem 1.1.3. If G is a group and a, b, c are elements of G, then
(1) $(ab)^{-1} = b^{-1}a^{-1}$ and
(2) $(a^{-1})^{-1} = a$

Proof. For (1) we have to show that either (ab) is the inverse of $b^{-1}a^{-1}$ of vice versa. Thus let us look at the product $(ab)b^{-1}a^{-1}$.

$$\begin{aligned}(ab)b^{-1}a^{-1} &= a(bb^{-1})a^{-1}\\ &= aea^{-1}\\ &= aa^{-1}\\ &= e\end{aligned}$$

Hence, $(ab)^{-1} = b^{-1}a^{-1}$.

For (2) we know that (a^{-1}) is the inverse of a ie $a(a^{-1}) = e$. Thus from the Theorem above $(a^{-1})^{-1} = a$. □

Definition 1.1.4. *If G is a finite group, the number of elements in G is called the order of G and is denoted by $|G|$.*

With these theorems we can solve elementary equations in Groups. Note that a^n is meant to be the product of n $a's$ i. e. $aaa \ldots a$ n times.

To solve for x in the equation $axb = c$, we multiply on the right by b^{-1} to get $ax = cb^{-1}$ and then we multiply by a^{-1} on the left to get $x = a^{-1}cb^{-1}$. To solve the simultaneous equation given by $x^2a = bxc^{-1}$ and $acx = xac$, from the first we see that $b^{-1}x^2a = xc^{-1}$ and multiplying by a^{-1} on the right we see that $b^{-1}x^2 = xc^{-1}a^{-1}$. Now multiplying by x^{-1} on the right we see $b^{-1}xxx^{-1} = c^{-1}a^{-1}$ and thus $b^{-1}x = c^{-1}a^{-1}$ i. e. $x = bc^{-1}a^{-1}$. This satisfies the second equation.

With two groups we can form a new group given by the direct product. Let G, H be two groups with operations $\star_1$ and $\star_2$. Then $G \times H$ is the direct product of G and H and is given by

$$G \times H = \{(x, y) : \quad x \in G \quad \textit{and} \quad y \in H\}$$

For elements (x_1, y_1) and (x_2, y_2) in G the operation in $G \times H$ is defined as

$$(x_1, y_1) \star (x_2, y_2) = (x_1 \star_1 x_2, y_1 \star_2 y_2).$$

As an example consider the direct product of the (additive) groups $\mathbb{Z}_2$ and $\mathbb{Z}_3$ given by $\mathbb{Z}_2 \times Z_3$. This group will consist of 6 elements and the group table (omitting the subscripts 2 and 3) is given by

+	([0],[0])	([0],[1])	([0],[2])	([1],[0])	([1],[1])	([1],[2])
([0],[0])	([0],[0])	([0],[1])	([0],[2])	([1],[0])	([1],[1])	([1],[2])
([0],[1])	([0],[1])	([0],[2])	([0],[0])	([1],[1])	([1],[2])	([1],[0])
([0],[2])	([0],[2])	([0],[0])	([0],[1])	([1],[2])	([1],[0])	([1],[1])
([1],[0])	([1],[0])	([1],[1])	([1],[2])	([0],[0])	([0],[1])	([0],[2])
([1],[1])	([1],[1])	([1],[2])	([1],[0])	([0],[1])	([0],[2])	([0],[0])
([1],[2])	([1],[2])	([1],[0])	([1],[1])	([0],[2])	([0],[0])	([0],[1])

One of the most important groups are the permutation groups.

Definition 1.1.5. *Given a set A, a permutation of the set A is a* bijective *function from A onto A.*

Given a set A, the set of all permutations on A is denoted by Σ_A or $\mathrm{Sym}(A)$. It is easy to see that $(\mathrm{Sym}(A), \circ, e_A)$ is a group. Here $\circ$ is the *composition* operation given by

$$(f \circ g)(x) = f(g(x)) \quad \text{for all } x \in A \text{ and } f, g \in \mathrm{Sym}(A).$$

The identity element e_A is the identity function given by $e_A(x) = x$. When A is a finite set of cardinality n we can regard the set of permuations on A as the set of permutations on $1, 2, \ldots, n$ and denote it by Σ_n or S_n where the later symbol is used to denote the *symmetric* group on n elements.

What are the elements of S_1? The only function that is possible is one that maps 1 to 1. We denote this by,

$$\epsilon = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

The notation is such that the bottom row shows how the function acts on the top row, i.e. $\epsilon(1) = 1$. Similarly, we can write the elements of S_2 as follows,

$$\epsilon = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix} \quad \alpha = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

We can see that $\epsilon \circ \alpha = \alpha \circ \epsilon = \alpha$ and $\alpha \circ \alpha = \epsilon$. Thus S_2 is an Abelian group. We will see that S_3 is not Abelian. From now on we will denote the operation (composition) as product i.e $\alpha \circ \beta$ will be denote as $\alpha\beta$.

Let the elements of S_3 be $\epsilon, \alpha, \beta, \gamma, \delta, \kappa$ given by

$$\epsilon = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} \quad \alpha = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \quad \beta = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$$

$$\gamma = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \quad \delta = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \quad \kappa = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$

Let us check if S_3 is Abelian. It is easy to observe that,

$$\beta\alpha = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \quad \alpha\beta = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$$

Thus $\alpha\beta \neq \beta\alpha$. This is the simplest non-Abelian group we can find. The nex proposition shows that all symmetric groups of n elements where n is greater than 2 must be non-Abelian.

Proposition 1.1.2. *Every symmetric group of $n > 2$ elements is non-Abelian.*

Proof. We have shown that S_3 is non-Abelian. For any S_n where $n > 3$ we fix the letters $4, 5 \ldots, n$ and only permute the first 3. There are 6 such possibilities that correspond to permutations in S_3 and since S_3 is non-Abelian, one of such permutation must not commute and thus S_n when $n > 2$ is a non-Abelian group. □

Every permutation in S_n can be broken down into simple parts called cycles. As an example let us look at the following permutation,

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

We can observe that $f(1) = 3$, while $f(3) = 6$, $f(6) = 2$ and $f(2) = 1$. Thus we started with 1 and ended at 1. We can consider this *chain or cycle* $1, 3, 6, 2$ and observe that f acts on the left element to yield the right element giving back the first element when acting on the last. Similarly the other chains are $4, 9, 7$ and $5, 8$.

Definition 1.1.6. *Let $a_1, a_2, \ldots, a_s$ be distinct elements of the set $\{1, 2, \ldots, n\}$. By the cycle $(a_1 a_2 \ldots a_s)$ we mean the permutation*

$$a_1 \to a_2 \to \ldots \to a_{s-1} \to a_s \to a_1$$

of $\{1, 2, \ldots, n\}$ which carries a_1 to a_2, a_2 to a_3 and so on and a_s to a_1, while leaving all the remaining elements of $\{1, 2, \ldots, n\}$ fixed.

For example in S_5, the cycle (254) is the permutation

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 5 & 3 & 2 & 4 \end{pmatrix}$$

Let us see how composition works on cycles. Let $\alpha, \beta \in S_5$ be given by (245) and (124). To get $\alpha\beta$ we must operate β and then operate α on the result. The permutations are shown below,

$$\beta = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 3 & 1 & 5 \end{pmatrix}$$

$$\alpha = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 3 & 5 & 2 \end{pmatrix}$$

Thus to get $\alpha\beta$ we have to see how α acts on $(2, 4, 3, 1, 5)$. From the permuation above we observe that $\alpha\beta = (4, 5, 3, 1, 2)$. If two cycles have no elements in common then they are *disjoint*. For example the two cycles introduced above are not disjoint whereas (123) and (45) are disjoint cycles. It is easy to see that disjoint cycles commute i. e. $(123)(45) = (45)(123)$. The next theorem shows that any permuation can be written as a product of disjoint cycles. We will give an informal explaination of the proof.

Theorem 1.1.4. Every (finite) permutation is either the identity, a single cycle or the product of disjoint cycles.

Let us look at an example. Consider the permuatation,

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 4 & 5 & 2 & 1 & 6 \end{pmatrix}$$

We begin with the first element $i_1 \in \{1, 2, \ldots, 6\}$ such that $f(i_1) \neq i_1$. If no such i_1 exists then f must be identity. In this case we get $i_1 = 1$. Next we look at $f(i_1) = i_2$ which in this case is 3. If $f(i_2) = i_1$ then we stop and our first cycle will be (i_1, i_2). However, in this example $f(3) = 5$ and so let us denote $i_3 = 5$ and again check if $f(i_3) = i_1$. Here we get $f(5) = 1 = i_1$ and so we stop and our first cycle is (135). Let us now pick the smallest element i_4 which is in $\{1, 2, \ldots, n\} - \{i_1, i_2, i_3\}$ (If this were not possible, the we would have exhuasted all elements and will end up with a single cycle.). This gives us $i_4 = 2$. We repeat the steps above to get $i_5 = 4$ and this is where are cycle stops. Now we pick the smallest element $i_6 \in \{1, 2, \ldots, n\} - \{i_1, i_2, i_3, i_4, i_5\}$. The only one left is 6 and is unmoved by f. Thus $f = (135)(24)$.

Definition 1.1.7 (Transposition)**.** *A cycle of s elements has length s. A cycle of length 2 is called a* transposition.

Note that a cycle (abc) can be written in atleast two different ways as product of transpositions given by $(abc) = (ab)(ac)$ and $(abc) = (cb)(ca)$. Any cycle $(a_1 a_2 \ldots a_s)$ can be written as product of transpositions given by

$$(a_s a_{s-1})(a_s a_{s-2}) \ldots (a_s a_1)$$

For example consider the following permutation,

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 1 & 2 & 5 & 3 \end{pmatrix}$$

f can be written as (14532), which can also be written as (23)(25)(24)(21). To see how this is true we will work out the composition from *right* to *left* step by step. Let us look at (24)(21).

$$(21) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 3 & 4 & 5 \end{pmatrix}$$

$$(24) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 3 & 2 & 5 \end{pmatrix}$$

And thus (24)(21) is

$$(24)(21) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 1 & 3 & 2 & 5 \end{pmatrix}$$

Again let us work out (25) composed with this.

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 1 & 3 & 2 & 5 \end{pmatrix}$$

$$(25) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 5 & 3 & 4 & 2 \end{pmatrix}$$

And thus we get,

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 1 & 3 & 5 & 2 \end{pmatrix}$$

Thus when we compose with (23) we get

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 1 & 3 & 5 & 2 \end{pmatrix}$$

$$(23) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 2 & 4 & 5 \end{pmatrix}$$

we get,

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 1 & 2 & 5 & 3 \end{pmatrix}$$

which was our original permuatation.

Definition 1.1.8. *The parity of permuatation is* even *if it can be written as an even number of transpostions. It is odd if the number of transpostions are odd.*

For example the parity of (14532) from above is even.

Definition 1.1.9. *The signature (sign) of a permuation is* -1 *if the parity is odd and* 1 *if the parity is even and for any* $\pi \in S_n$ *the signuture is denoted by* $sign(\pi)$.

Every permutation can be written as product of transpositions. However the number of transposition is not unique. What is unique is the parity. Every permutation can either be written as an even number of transposition or odd.

Theorem 1.1.5. The identity permuation has signature 1.

Proof. To prove this we will need to show that the identity permuation can only be written as an even number of transposition. Another way to look at it is by saying that if the identity permutation is written as a transpostion of m terms then it can be reduced to $m-2$ terms.

To see this let $\epsilon = t_1 t_2 t_3 \dots t_k t_{k+1} \dots t_m$. Let $x \in \mathbb{Z}^+$ be a numeral that appears in t_k and moreover it is the last appearance of x. Let $t_k = (xa)$. There are four cases to consider for t_{k-1}.

(1) CASE 1 $t_{k-1} = (xa)$. Then $(xa)(xa)$ is just the identity and we can remove both t_{k-1} and t_k leaving ϵ unchanged. Thus we have reduce the transpostions to $m-2$ terms.

(2) CASE 2 $t_{k-1} = (xb)$. Then $(xb)(xa) = (xab) = (xa)(ab)$ and so we have moved the last occurance of x to t_{k-1}. Applying this repeatedly we will end up with CASE 1.

(3) CASE 3 $t_{k-1} = (ba)$. Then $(ba)(xa) = (xba) = (xb)(ba)$ and is similar to CASE 2.

(4) CASE 4 $t_{k-1} = (bc)$. Then $(bc)(xa) = (xa)(bc)$ and is similar to CASE 2.

Hence we can reduce a transposition of m terms to that of $m-2$. Thus m has to be even other wise if m is odd repeated reduction will lead to transposition of single term and that is not possible since we have an identity permutation. □

Theorem 1.1.6. If $\pi \in S_n$ then π cannot be both an odd permutation and an even permutation.

Proof. If π can be written as both as an odd permutation and an even permutation then so can π^{-1}. But that would mean that $\epsilon = \pi \circ \pi^{-1}$ can be written as an odd permutation by chosing π as even and π^{-1} as odd. This will violate the theorem above. □

Some useful facts about parity can be seen from the following viewpoint. Consider the symmetric group S_n and construct a polynomial with factors $(t_i - t_j)$ as follows

$$f = \prod_{1 \le i < j \le n} (t_i - t_j).$$

For example when $n = 4$, then $f = (t_1 - t_2)(t_1 - t_3)(t_1 - t_4)(t_2 - t_3)(t_2 - t_4)(t_3 - t_4)$. In all there will be $n * (n-1)/2$ terms. Each permutation $\pi \in S_n$ converts the polynomial f into a new polynomial given by

$$\pi f = \prod_{1 \le i < j \le n} (t_{\pi(i)} - t_{\pi(j)}).$$

For example if $\tau \in S_4 = (2, 4)$ then $\tau f = (t_1 - t_4)(t_1 - t_3)(t_1 - t_2)(t_4 - t_3)(t_4 - t_2)(t_3 - t_2)$. Thus $\tau f = -f$.

Note that for any $\pi \in S_n$ we have that $\pi f = f$ except possibly a $\pm$ sign.

Lemma 1.1.1. *Consider the symmetric group S_n. Then*

- *If $\sigma \in S_n$ is a transposition then $\sigma f = -f$.*
- *For any $\sigma, \tau \in S_n$ $(\sigma\tau)(f) = \sigma(\tau f)$.*
- *For any $\sigma \in S_n$ we have $\sigma(f) = sign(\sigma) f$.*

Proof.

- If $\sigma = (a, b)$ then either $a < b$ or $b < a$. WLOG let $a < b$ then the factor $(t_a - t_b)$ will change to $(t_b - t_a)$. For all other factors that don't have a, b nothing will change. The factors containing only one of a or b come in pairs that is $(t_a - t_k)$ and $(t_k - t_b)$ and so their changes will cancel sign. Thus $\sigma f = -1 f$. Note that the sign of a transposition is -1. Thus $\sigma f = sign(\sigma) f$ in this case.
- Note that

$$(\sigma\tau)(f) = \prod_{1 \le i < j \le n} (t_{\sigma\circ\tau(i)} - t_{\sigma\circ\tau(j)}) = \prod_{1 \le i < j \le n} (t_{\sigma(\tau(i))} - t_{\sigma(\tau(j))}) = \sigma(\tau(f)).$$

- $\sigma(i), \sigma(j) \le n$ and so $\sigma(f)$ doesn't change f except possibly the sign. If σ is even permutation then the sign changes will be in pairs and $\sigma f = f$ else $\sigma f = -1 f$. This can be seen by repeated application of the results above. Hence $\sigma f = sign(\sigma) f$.

□

Theorem 1.1.7. Let S_n be the symmetric group of n elements. Then for any σ, τ in S_n, we have $sign(\sigma\tau) = sign(\sigma) sign(\tau)$.

Proof. This follows immediately from the above Lemma. □

One other way to get groups from a given group G is to look at subsets of G that preserve the algebraic structure.

Definition 1.1.10. *Let $(G, \star, e)$ be a group and S be a non-empty subset of G. Then S is called a subgroup of G iff*
(1) S is closed w.r.t $\star$.
(2) S is closed w.r.t inverses.

Again dropping $\star$ the first condition means that for any $a, b \in S$ it must be that $ab \in S$ while the second condition states that for any $a \in S$, $a^{-1} \in S$. It is easy to see that $(S, \star, e)$ is also a group.

Given a set A, the set of all functions defined on A is denoted by $\mathcal{F}[A]$ i. e.

$$\mathcal{F}[A] = \{f| \quad f : A \to A\}$$

When $A = \mathbb{R}$, then $\mathcal{F}[\mathbb{R}]$ is the set of all *real valued* functions. Addition of two functions $f, g \in \mathcal{F}[\mathbb{R}]$ is given by

$$(f + g)(x) = f(x) + g(x) \quad \text{for all } x \in \mathbb{R}.$$

Thus we can consider $(\mathcal{F}[\mathbb{R}], +, 0_f)$ as a group where the identity element is the *zero function* $0_{\mathbb{R}} : \mathbb{R} \to \mathbb{R}$ given by

$$0_{\mathbb{R}}(x) = 0 \quad \text{for all } x \in \mathbb{R}.$$

Some important subgroups of $(\mathcal{F}[\mathbb{R}], +, 0_{\mathbb{R}})$ are the set of real valued *continuous* functions, real valued *differentiable* functions etc.

For any group G, the two trivial subgroups are the singleton $\{e\}$, and the entire set G. All other subgroups of G are called *proper* subgroups.

Given a group, it is usually not possible to identify all the subgroups. However the next proposition shows that for the group $(\mathbb{Z}, +, 0)$, we can identify all the subgroups.

Proposition 1.1.3. *For any integer $b \geq 0$, the subset $b\mathbb{Z}$ is a subgroup of $(\mathbb{Z}, +, 0)$, where*

$$b\mathbb{Z} = \{n \in \mathbb{Z} \ : \ n = bk \quad \text{for some } k \in \mathbb{Z}\}.$$

Moreover, every subgroup H of $\mathbb{Z}$ is of they type $H = b\mathbb{Z}$ for some $b \in \mathbb{Z}$.

Proof. First we will show that $b\mathbb{Z}$ is a subgroup of $\mathbb{Z}$. Note that $0 = b0$ and hence 0 is in $b\mathbb{Z}$. Let $x, y \in b\mathbb{Z}$. Then $x = bm$ and $y = bn$ for some $m, n \in \mathbb{Z}$. Thus $x + y = b(m + n)$ and hence $x + y \in b\mathbb{Z}$. For any $x = bm \in bZ$, we have $bm + b(-m) = b(0) = 0$. Hence $b(-m)$ is the additive inverse of x. Now let H be a subgroup of $\mathbb{Z}$. If H contains only 0 then $H = 0\mathbb{Z}$. Assume there is an $m \in H$ such that $m \neq 0$. Either $m < 0$ or $m > 0$. In both cases H contains a positive number ($-m$ or m) and hence it must contain a smallest positive integer (well defined property). Let $b > 0$ must be the smallest integer in H. Since H is a subgroup of Z it must contain $b + b + b \ldots$, $b + (-b) \ldots$ i.e $H \supset b\mathbb{Z}$. Let m be any element of H. By the division algorithm we can write, $m = bq + r$ such that $0 \leq r < b$. But this means that $r \in H$ since m, bq are in H. But b was the smallest positive integer in H and this must mean that $r = 0$. Hence $H = b\mathbb{Z}$. $\square$

Finite subgroups can be thought of subsets that contain finite elements of a group along with all their products and the products of their inverses.

Definition 1.1.11. *If $a_1, a_2, \ldots, a_n$ are any finite number of elements of a group G, then the subgroup generated by $a_1, a_2, \ldots, a_n$ is the set formed by including all the the possible* products *of $a_1, a_2, \ldots, a_n$ along with their inverses.*

For example if a, b are elements of G, then the subgroup generated by a, b is the set that contains elements like $abab^{-1}$, $b^{-1}a^{-1}b^{-1}$, $aaabbbaa$ etc.

Definition 1.1.12. *A subgroup of a group G is calle a cyclic subgroup if it generated by a single element $a \in G$ and is denoted by $\langle a \rangle$.*

Thus $\langle a \rangle$ contains $a, aa, aaa, aaaa, \ldots$ and $a^{-1}, a^{-1}a^{-1}, a^{-1}a^{-1}a^{-1}, \ldots$ and $aa^{-1} = e$.

We can see that $b\mathbb{Z}$ is a subgroup of $\mathbb{Z}$ that contains all multiples of $\mathbb{Z}$ i.e. all the integers divisible by b. Hence $b\mathbb{Z}$ is the subgroup generated by a single element b. Thus $b\mathbb{Z}$ is a cyclic subgroup of $\mathbb{Z}$. What is the subgroup of $\mathbb{Z}$ that is generated by two integers $a \geq 0, b \geq 0$. Let us assume that both a, b are not simultaneously zero. Then the set,

$$a\mathbb{Z} + b\mathbb{Z} = \{n \in \mathbb{Z} \ : \ n = ar + bs, \quad \text{for some } r, s \in \mathbb{Z}\},$$

is a subgroup of Z. By 1.1.3, this must be equal to $d\mathbb{Z}$ for some $d > 0 \in Z$. Since $d\mathbb{Z}$ is the set of all integers divisibe by d, it must be the case that d divides both a, b. This is because the set $d\mathbb{Z}$ contains all the multiples of a (with $b0$) and similarly all the multiples of b (with $a0$). If e is any integer that divides a, b it must divide any combination $ar + bs$, which means that $e\mathbb{Z} \subset d\mathbb{Z}$. Hence, $d = \gcd(a, b)$.

In these examples it will be useful to introduce some shorthand. Let $(G, *, e)$ be a group and let x be any element of G. For any $n \in \mathbb{Z}^+$ let us identify the following,

$$x^n = \underbrace{x*x*x*\ldots*x}_{n \text{ times}}$$
$$x^{-n} = \underbrace{x^{-1}*x^{-1}*x^{-1}*\ldots*x^{-1}}_{n \text{ times}}$$
$$x^0 = e$$

Again following 1.1.1 we will omit the $*$. It is easy to see that the following hold for any $m, n \in \mathbb{Z}$

$$x^{m+n} = x^m x^n$$
$$(x^m)^n = x^{mn}$$
$$x^{-n} = (x^{-1})^n = (x^n)^{-1}$$

Observation 1.1.1. *If there is a non-zero integer m such that $x^m = e$, then there must be a positive integer n such that $x^n = e$. This is easy to see if $m > 0$, in which case $n = m$. If $m < 0$, then $n = -m > 0$ and $x^n = x^{-m} = (x^m)^{-1} = e^{-1} = e$.*

This leads us to a definition,

Definition 1.1.13 (Order of an element)**.** *Let G be a group with the identity e. For any element $x \in G$ if there exist a non-zero integer m such that $x^m = e$, then the order of the element x is the* ***least positive integer*** *n such that $x^n = e$. If there is no such n such that $x^n = e$ then we say that x has infinite order. We denote the order of x by* $\text{ord}(x)$.

For example in page 6, $\alpha \in S^n$ has order 2 since $\alpha^2 = \epsilon$. Similarly the order of $\delta = 3$ since $\delta^3 = \epsilon$. For any non zero element b in $\mathbb{Z}$ with the additive group $(\mathbb{Z}, +, 0)$, $b^n = \underbrace{b + b + \ldots b}_{n \text{ times}}$. And hence b has infinite order.

The following proposition collects some important facts about the order of an element.

Proposition 1.1.4. *Let G be a group with the identity element e. Let a be an element of G. Then,*

(1) If $\text{ord}(a) = n$ *is finite, then there are exactly n different powers of a given by,*

$$a^0, a^1, \ldots, a^{n-1}.$$

(2) If $\text{ord}(a)$ *is infinite, then all powers of a are different.*

Proof. We prove in order.

(1) Let $m \in \mathbb{Z}$ be any power of a. By division theorem $m = nq + r$ for $0 \leq r < n$. Thus,

$$a^m = a^{nq}a^r = (a^n)^q a^r = ea^r = a^r.$$

Hence any power of a is of the form $a^0, a^1, \ldots, a^{n-1}$. To show that these powers are different, let us assume that for some $r, s \in Z$ such that r, s are one of the integers in $0, 1, \ldots, n-1$, $a^r = a^s$. WLOG assume $r < s$ i.e $0 \leq r < s \leq n-1$. Then $0 < s - r < n$, and hence,

$$a^r = a^s \implies a^r(a^s)^{-1} = e \implies a^{r-s} = e,$$

but since n is the smallest positive integer that makes $a^n = e$, it must be the case that $r = s$ which gives us a contradiction. Hence the powers in $0, 1, \ldots, n-1$ are different.

(2) Assume $r < s$ and $a^r = a^s$ to get $a^{r-s} = e$ which is a contradiction.

□

Definition 1.1.14 (Cyclic group)**.** *If G is a group and $a \in G$ such that $G = \langle a \rangle$,then we call G a cyclic group. Thus a cyclic group G is given by,*

$$G = \{a^n \ : \ n \in \mathbb{Z}\}.$$

We call a the generator of G. If no finite n exists, then G is a cyclic group of infinite order.

For example $\mathbb{Z}_6$ is a cyclic group that is generated by $[1]_6$. Is the additive group of integers $\mathbb{Z}$ cyclic? We can easily see that $(Z, +, 0)$ is generated by 1. Moreover it is of infinite order. In general what is the order of a cyclic group? Since we have shown in 1.1.4, that if an element a of group G has order n, there are n distinct element of $\langle a \rangle$ if n is finite and all distinct powers if n is infinite. Hence, if $G = \langle a \rangle$,

then $|G| = n = \text{ord}(a)$. If a group is cyclic, is it the case that any of its subgroup is cyclic? The answer isn't obvious. If G is a cyclic group of order n such that $G = \langle a \rangle$, and H is a subgroup of G, then it must have as its elements some powers of a. What powers of a generate H? Intuition suggest that we must take the smallest such power. This result is proved in the following proposition.

Proposition 1.1.5. *Every subgroup of a cyclic group is cyclic.*

Proof. Let G be a cyclic group generated by a. Let m be the smallest positive integer such that $a^m \in H$. We claim that $H = \langle a^m \rangle$. Let t be any integer such that $a^t \in H$. By division theorem there is a $r \in \mathbb{Z}^+$ such that $0 \leq r < m$ and,

$$\begin{aligned} & t = mq + r \\ \implies & a^t = (a^m)^q a^r \\ \implies & a^t((a^m)^q)^{-1} = a^r. \end{aligned}$$

Since $a^m \in H$, we have $((a^m)^q)^{-1} \in H$. Since we assumed $a^t \in H$, this means that a^r is also in H. But this contradicts our assumption that m was the smallest positive integer for which $a^m \in H$. Hence, $r = 0$. This means that if $a^t \in H$ then $t = mq$. Hence H is generated by a^m. □

1.2. Mappings on groups: isomorphism and homomorphism

The notion of congruence in geometry is well understood. Two shapes are congruent if they have the same structure, i.e. a plane motion consisting of rotation, translation and dilitation (magnifying or shrinking) can make one figure coincide with the other. Such a motion can be thought of as a transformation, i.e a function that takes one shape and produces an image shape that coincides with the second shape. In terms of groups, we have an algebraic structure and we want to find out when two groups are essentially representing the same structure. A precise characterization of this notion leads to the following definition,

Definition 1.2.1 (Isomorphism)**.** *Let G_1, G_2 be groups. A **bijective** function $f : G_1 \to G_2$ is called an isomorphism if,*

$$f(\underbrace{ab}_{op.\ in\ G_1}) = \underbrace{f(a)f(b)}_{op.\ in\ G_2},$$

for any $a, b \in G_1$. If G_1, G_2 are isomorphic, we denote it by $G_1 \cong G_2$.

Let us consider a few examples.

Example 1.2.1. *The following are all isomorphic groups.*

(1) Let $G_1 = \{\pm 1, \pm i\} \subset (\mathbb{C} - (0,0), \cdot, 1)$. Let $G_2 = \langle \sigma \rangle \subset S_4$, where σ is given by,

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

Note that G_1 is a cyclic group generated by i since $i^0 = 1$, $i^1 = i$, $i^2 = -1$ and $i^3 = -1$. The order of $\langle \sigma \rangle$ is also 4. If we define a function, $f : G_1 \to G_2$

given by $f(i^k) = \sigma^k$ for $k = 0, 1, 2, 3$ we get a bijection. Moreover for $a, b \in G_1$, $a = i^k, b = i^l$ for some $k, l \in 0, 1, 2, 3$. If $a \neq b$ then $k \neq l$ and hence,

$$f(ab) = f(i^k i^l) = f(i^{k+l}) = \sigma^{k+l} = \sigma^k \sigma^l = f(a)f(b).$$

*Hence $G_1 \cong G_2$. We can see that an isomorphism is a re-labelling of a groups elements. Here f just **re-labels** element of G_1 given by elements of G_2.*

(2) Let G_1 be the group $(\mathbb{R}, +, 0)$ and let G_2 be the group $(\mathbb{R} - 0, \cdot, 1)$. Define the function $f : G_1 \to G_2$ to be $f(x) = e^x$ for all $x \in \mathbb{R}$. Then f is bijective. Moreover for any $x, y \in \mathbb{R}$,

$$f(x + y) = e^{(x+y)} = e^x e^y = f(x) \cdot f(y).$$

Hence $G_1 \cong G_2$.

Example (1) above gives us a useful result.

Proposition 1.2.1. *Any two cyclic groups of the same order are isomorphic.*

Proof. Let G_1 be a cyclic group of order n generated by a and let G_2 be the cyclic group of order n generated by b. The function $f : G_1 \to G_2$ given by $f(a^i) = b^i$ for $i = 0, 1, n - 1$ is bijective and for any $x \in G_1, y \in G_2$, $f(xy) = f(x)f(y)$. Hence $G_1 \cong G_2$. If both G_1, G_2 are of infinite order then pick a, b to be any non-identity element in G_1 and G_2 respectively. □

When can we fail to have an isomorphism between groups G_1, G_2? A few simple cases to check will be:

- $|G_1| \neq |G_1|$
- G_1 is abelian but G_2 is not (or vice-versa).
- G_1 has element of different order than G_2.

It turns out that any group is isomorphic to a group of permutations. This is an important theorem since it tells us that the permutation group are in some sense the most fundamental group. A precise statement is given in the following theorem due to A. Cayley.

Theorem 1.2.1 (Cayley's theorem)**.** Every group G is isomorphic to a subgroup of the symmetric group $\mathrm{Sym}(G)$.

There is some complexity in proving this statement, since we need to find an isomorphic function f which maps every element of G to a bijective function defined on G. Fix an $a \in G$ and consider the map $\tau : G \to G$ given by, $\tau(x) = ax$ for every $x \in G$. We can make the following observations:

- If $\tau(x) = \tau(y)$, then $ax = ay$ which means $x = y$.
- For any $y \in G$, we can write $y = aa^{-1}y$ and hence there is an element $x = a^{-1}y \in G$ such that $y = \tau(x)$.

These two observations show that for a fixed a this function τ is bijective defined on G. Let us denote this function as τ_a to show the dependence of a. Clearly τ_a is an element of $\mathrm{Sym}(G)$. Let $G^* = \{\tau_a : a \in G\}$ be the collection of all such τ, i.e. as we vary a we get a new function and we take all such functions. Then

$G^* \subset \mathrm{Sym}(G)$, but is it a subgroup? For any τ_a, τ_b in G^*, the composition $\tau_a\tau_b$ is given by $(ab)x = \tau_{ab}(x)$ for any $x \in G$. Since ab is an element of G, this means that G^* is closed under composition. Similarly for any τ_a in G^* if we define $(\tau_a)^- = \tau_{a^{-1}}$, then $\tau_a\tau_{a^{-1}}(x)$ is equal to $aa^{-1}x = x$ which is the identity function. Hence G^* is a subgroup. Now we have our isomorphism.

Proof. For any $a \in G$ let $\tau_a : G \to G$ be defined as,

$$\tau_a(x) = ax,$$

for any $x \in G$. Then $\tau_a \in \mathrm{Sym}(G)$. Let $G^* = \{\tau_a : a \in G\}$. Then $G^* \subset G$ is a subgroup of G. Define $f : G \to G^*$ as,

$$f(a) = \tau_a.$$

If $f(a) = f(b)$ then $\tau_a = \tau_b$ which means that $a = b$ since τ is injective. Thus f is injective. Fix a $b \in G$, then since τ_b is surjective, there is an $a \in G$ such that $\tau_b(a) = b$ which means that $f(a) = b$. Thus f is surjective. Hence, f is bijective. Also for any $a, b \in G$,

$$f(ab) = \tau_{ab} = \tau_a\tau_b,$$

becuase for any $x \in G$, $\tau_{ab}(x) = abx = \tau_a(bx) = \tau_a\tau_b(x)$. Hence, f is an isomorphic function. Thus, $G \cong G^*$. □

While an isomorphism tells us which groups are **similar**, we would want to know which maps preserve the underlying algebraic structure. Certainly any isomorphic function does that. What other functions have this property? Such maps are very important, in fact a case can be made for them being more important than the groups themselves. As with any important concept we give them a name.

Definition 1.2.2 (Homomorphism)**.** *Let G_1, G_2 be two groups. A function $f : G_1 \to G_2$ is called a homomorphism if,*

$$f(\underbrace{ab}_{op.\ in\ G_1}) = \underbrace{f(a)f(b)}_{op.\ in\ G_2},$$

for any $a, b \in G_1$.

Let us give a few examples of homomorphisms.

Example 1.2.2. *The following are all homomorphism,*

(1) Any $f : G_1 \to G_2$ that is an isomorphism is also a homomorphism.

(2) The trivial homomorphism given by $f(x) = e_2$ where e_2 is the identity element in G_2. This is because $f(xy) = e_2 = e_2e_2 = f(x)f(y)$.

(3) Let f be the function $f : S_n \to \{\pm 1\}$ given by $f(\sigma) = sign(\sigma)$, for any $\sigma \in S_n$. Then, by 1.1.7, f is a homomorphism. While f is surjective for S_n when $n > 1$, it is by no means injective. Hence, f is NOT an isomorphism.

(4) Let G_1 be the additive group of integers $(\mathbb{Z}, +, 0)$ and let $G_2 = S_2 = \{\epsilon, \alpha\}$ where α is the transposition (12). Then the function, $f : G_1 \to G_2$ given by,

$$f(x) = \begin{cases} \epsilon & \text{if } x \text{ is even} \\ \alpha & \text{if } x \text{ is odd,} \end{cases}$$

is a homomorphism, because if x, y are even then $f(x+y) = \epsilon = \epsilon\epsilon = f(x)f(y)$, and when x is even and y is odd, then $f(x+y) = \alpha = \epsilon\alpha = f(x)f(y)$.

Proposition 1.2.2. *Let G_1, G_2 be groups with identity elements e_1, e_2 respectively. If $f : G_1 \to G_2$ is a homomorphism, then*

(1) $f(e_1) = e_2$,

(2) $f(x^{-1}) = (f(x))^{-1}$.

Proof. For any $x \in G_1$, $f(x)e_2 = f(x) = f(xe_1) = f(x)f(e_1)$. Thus by the cancellation law of groups we get $e_2 = f(e_1)$.

For the second statement, $f(x^{-1})f(x) = f(xx^{-1}) = f(e_1) = e_2$. Hence, $f(x^{-1}) = (f(x))^{-1}$. □

A group homomorphism induces two natural subgroups that are extremely important.

Definition 1.2.3 (Kernel)**.** *Let G_1, G_2 be groups and let $f : G_1 \to G_2$ be a homomorphism. The kernel of f is the set* $\ker(f)$ *of all the elements of G_1 which are mapped by f to the identity element e_2 of G_2. That is,*

$$\ker(f) = \{x \in G_1 \ : \ f(x) = e_2\}.$$

Note that $\ker(f) = f^{-1}(\{e_2\})$, where the latter is the inverse image of the set $\{e_2\}$. Since e_1 is mapped by a homomorphism to e_2, a kernel is never empty.

Definition 1.2.4 (Image or Range)**.** *Let G_1, G_2 be groups and let $f : G_1 \to G_2$ be a homomorphism. The image or range of f is the set* $\mathrm{Im}(f)$ *of all the element of G_2 that have been mapped by some elements of G_1. That is,*

$$\mathrm{Im}(f) = \{y \in G_2 \ : \ \exists\,(x \in G_1) \text{ such that } f(x) = y\}.$$

The following proposition states that these two sets are in fact subgroups.

Proposition 1.2.3. *Let G_1, G_2 be groups and let $f : G_1 \to G_2$ be a homomorphism. The kernel and image of f are subgroups of G_1, G_2 respectively.*

Proof. To see that $\ker(f)$ is a subgroup of G_1, we need to check if it is closed under operation and inverses in G_1. Let $a, b \in \ker(f)$, then $f(a) = f(b) = e_2$. Thus $e_2 = f(a)f(b) = f(ab)$. Hence $ab \in \ker(f)$. Let $a \in \ker(f)$ which means $f(a) = e_2$. But $e_2 = e_2^{-1} = (f(a))^{-1} = f(a^{-1})$. Hence $a^{-1} \in \ker(f)$.

To see that image is a subgroup of G_2, let $c, d \in \mathrm{Im}(f)$. Hence there are $x_c, x_d \in G_1$ such that $f(x_c) = c$ and $f(x_d) = d$. Since G_2 is a group $f(x_c)f(x_d) \in G_2$ and hence $cd = f(x_c)f(x_d) = f(x_c x_d)$ is in G_2. But since $x_c, x_d \in G_1$, this means that there is an $x = x_c x_d$ such that $f(x) = cd \in G_2$. Hence $cd \in \mathrm{Im}(f)$. Let $c \in \mathrm{Im}(f)$. Thus there is an $x_c \in G_1$ such that $f(x_c) = c$. But $c^{-1} = (f(x_c))^{-1} = f(x_c^{-1})$. Hence there is an $x = x_c^{-1}$ in G_1 such that $f(x) = c^{-1} \in G_2$. Thus $c \in \mathrm{Im}(f)$. □

The kernel is a special subgroup. To see this, take any $x \in G_1$. For any $a \in \ker(f)$, let us consider $f(xax^{-1})$. Since $a, x^{-1} \in G_1$, ax^{-1} is also in G_1 and since f is a homomorphism we have, $f(xax^{-1}) = f(x)f(ax^{-1}) = f(x)f(a)f(x^{-1})$.

Since $f(a) = e_2$, we have $f(xax^{-1}) = f(x)f(x^{-1}) = f(xx^{-1}) = f(e_1) = e_2$. Thus xax^{-} is also in the kernel. We isolate this important property in the definition below.

Definition 1.2.5 (conjugate)**.** *Let G be a group and let a, x be element of G. The element $xax^{-1} \in G$ is called a conjugate of a. If H is a subgroup of G, we say that H is* ***closed under conjugation*** *if for any $a \in H$, all the conjugates of a are also in H.*

Definition 1.2.6. *A subgroup H of G which is closed under conjugation is called a normal subgroup of G and is denoted by $H \triangleleft G$.*

Example 1.2.3. *Let G_1, G_2 be groups and let $f : G_1 \to G_2$ be a homomorphism. Then $\ker(f)$ is a normal subgroup of G_1 i.e. $\ker(f) \triangleleft G_1$.*

Example 1.2.4. *Let G be a group and let us define the following set, called the center of a group. Then,*

Definition 1.2.7. *the set,*

$$Z(G) = \{z \in G \ : \ zx = xz \, \forall \, (x \in G)\},$$

is a normal subgroup of G. This is easy to see since for any $z \in G$, $zx = xz$ which means that $z = xzx^{-1}$. The center of group G is the collection of all those elements of G that commute with every element of G.

We collect some elementary properties of homomorphism.

Proposition 1.2.4. *Let G, H, K be groups. The following are some of the basic properties of homomorphism.*

- ***Property 1*** *If $f : G \to H$ and $g : H \to K$ are homomorphism, then the composition map $g \circ f : G \to K$ is also a homomorphism.*
- ***Property 2*** *If $f : G \to H$ is a homomorphism, then f is injective if and only if $\ker f = \{e\}$, where e is the identity element of G.*
- ***Property 3*** *If $f : G \to H$ is a homomorphism and $K \subset G$ is a subgroup of G then the direct image of K, i.e. $f(K)$ is a subgroup of H.*
- ***Property 4*** *If $f : G \to H$ is a homomorphism and $J \subset H$ is a subgroup of H then the inverse image of J, i.e. $f^{-1}(J)$ is a subgroup of G. Moreover $\ker(f) \subset f^{-1}(J)$.*

1.3. Cosets and Quotient groups

In this section we will explore some of the most important concepts in group theory. We have studied the fundamentals of group theory by analyzing groups and maps that preserve group structure. Now we will build on it to realize some of the most astonishing facts about group theory. We will need a few facts from set theory concerning equivalence relations and partitions. See A.2. Let A, B be any arbitrary sets and let $f : A \to B$ be any function. Then f induces a natural partition of A give by the equivalence relation $\sim$ on A,

$$a \sim b \quad \text{if } f(a) = f(b).$$

What are the equivalence class of this relation? The main idea is to think about the inverse image of f. If we fix a $y \in \mathrm{Im}(f)$, then $f^{-1}(\{y\})$ is the collection of all those elements in A that are mapped to y. Since y is in the image of f, the inverse image $f^{-1}(\{y\})$ is not empty. If $a, b \in f^{-1}(\{y\})$, then $f(a) = f(b) = y$ and thus $a \sim b$. For any $y \in \mathrm{Im}(f)$, let $x_y \in A$ be such that $f(x_y) = y$. Then the equivalence class of x_y is given by,

$$[x_y] = \{a \in A \ : \ f(a) = f(x_y)\} = f^{-1}(\{y\}).$$

So we can see that A is partitioned as follows,

$$A = \bigcup_{y \in \mathrm{Im}(f)} [x_y].$$

We denote the set of all equivalent classes in A as $\overline{A}$. Thus we can see that the inverse image partitions the domain in the way described above. In other words we can identify the image with the set of all equivalence classes as follows: There is a bijective map $\overline{f} \ : \overline{A} \to \mathrm{Im}\, f$ given by,

$$f([a]) = f(a).$$

Clearly this is injective because we know that if $[a] \cap [b] \neq \emptyset$ then they represent the same equivalent class. In other words if $f([a]) = f([b])$ then $f(a) = f(b)$, which means that $a \sim b$ which means that $[a] = [b]$. The map is surjective since for any $y \in \mathrm{Im}\, f$, the inverse image of y is one of the equivalence classes. Thus we can identify the image with the set of all equivalence classes on the domain.

We can make some important observations when $A = G_1$ and $B = G_2$ are groups.

Observation 1.3.1. *Let G_1, G_2 be groups with identity element e_1, e_2 respectively, and let $f \ : G_1 \to G_2$ be a homomorphism. Since a homomorphism always carries e_1 to e_2 we know that $e_2 \in \mathrm{Im}(f)$. Thus from the discussion above $f^{-1}(\{e_2\})$ is one of the equivalence classes of A namely $[e_1]$. But this is precisely the Kernel of f! Hence $\ker f$ is one of the equivalence classes of A. What are the other equivalence classes? The next proposition gives a remarkable answer to this question.*

Proposition 1.3.1. *Let G_1, G_2 be groups with identity element e_1, e_2 respectively, and let $f \ : G_1 \to G_2$ be a homomorphism. Let $\sim$ be the equivalence relation on A given by $a \sim b$ iff $f(a) = f(b)$. Any equivalence class of A is given by $a \ker f = \{an \ : \ n \in \ker f\}$. In other words $f(a) = f(b)$ if and only if $b = an$ for some $n \in \ker f$.*

Proof. If $b = an$ for some $n \in \ker f$, then since f is a homomorphism and using the properties of the kernel,

$$f(b) = f(a)f(n) = f(a)e_2 = f(a).$$

For the other implication, let $f(a) = f(b)$. Then

$$e_2 = (f(a))^{-1} f(b) = f(a^{-1})f(b) = f(a^{-1}b).$$

Hence, $a^{-1}b \in \ker f$, i.e. there is some $n \in \ker f$ such that $n = a^{-1}b$ which means that $an = b$. □

This motivates a very important idea in group theory, namely that of a **coset**.

Definition 1.3.1 (Left coset)**.** *Let $(G, *, e)$ be a group and let $H \subset G$ be any subgroup of G. The left coset of H is given by the set,*

$$a*H = \{b \in G \ : \ \exists\, (h \in H)\,, \text{ such that } \quad b = a*h\}\,,$$

where a is an element of G.

In short $a * H$ contains elements of the form $a * h$ for some h in H. Again following our convention in 1.1.1, we denote the left coset of H by aH. Is aH a subgroup of G? Let $x, y \in aH$. Then $x = ah_1$ and $y = ah_2$ for some $h_1, h_2 \in H$. Thus $xy = (ah_1)(ah_2)$ which is not necessarily ah_1h_2. Thus cosets are not necessarily subgroups of G. The only coset that is a subgroup of G is when $a = e$ and in that case $eH = H$.

Proposition 1.3.2. *The left cosets are equivalence classes on G for the relation given by,*

$$a \sim b \, if b = ah,$$

for some $h \in H$.

Proof. Since e is in H, $a = ae$ and hence $a \sim a$. If $a \sim b$, then $b = ah$ which means that $a = bh^{-1}$. Since H is a subgroup h^{-1} is in H and hence $b \sim a$. If $a \sim b$ and $b \sim c$ then $b = ah_1$ and $c = bh_2$ for some $h_1, h_2 \in H$. Thus $c = ah_1h_2 = ah$ where $h = h_1h_2 \in H$. Hence, $a \sim c$. □

Thus we see that the cosets aH **partition** G. The next proposition states that the cosets of H have the same cardinality of H. This has an important ramification for finte Groups. This means that if H is a finite subgroup, then each coset will have the same number of elements as in H.

Proposition 1.3.3. *If H is an subgroup of G, then there is a one-to-one correspondence from H onto aH.*

Proof. Let $f \ : H \to aH$ be given by $f(b) = ab$ for any $b \in H$. Let $f(b) = f(c)$ which means $ab = ac$, which means that $b = c$. Hence, f is injective. Let x be any element of aH. Then $x = ah$ for some $h \in H$ i.e. there is a h in H such that $f(h) = ah = x$. Hence f is surjective. □

Observation 1.3.2. *If H has n elements, then every coset aH has n elements.*

Theorem 1.3.1 (Lagrange theorem)**.** Let G be a finite group, and H any subgroup of G. The order of G is a multiple of the order of H.

Proof. Each coset of H has the same number of elements as H that is $|aH| = |H|$. Let the number of distinct cosets be m. Then $|G| = m * |H|$. Thus $|H|$ divides $|G|$. □

This is an important result. It says that if a group G has 15 elements, it have can have subgroups with elements $1, 3, 5, 15$. Thus there will be two non-trivial subgroups of elements 3 and 5. If a group has 7 elements, then it cannot have any non-trivial subgroups!

Corollary 1.3.1.1. *If G is a group with prime number of elements, then G is cyclic group generated by any element of G that is not an identity.*

Proof. Let $a \in G$ be any element of G which is not equal to e. Consider the cyclic subgroup $\langle a \rangle$ of order m. Then by 1.3.1, m must divide p. Since p is prime and $m \neq 1$, this means that $m = p$. Hence $G = \langle a \rangle$. Since any cyclic groups are isomorphic, this means that there is only one type of group with prime number of elements. □

We usually denote the number of distinct cosets of H by $[G : H]$ and call it the index of H in G.

Now we will look at another important idea in group theory. We have seen that any homomorphic image induces a normal subgroup given by its kernel. Is there a coverse? That is given any normal subgroup can be construct a homomorphic image that has the normal subgroup as its kernel? Before we proceed to accomplish this, we will construct the congruence class of integers. This will serve as an example of what we are going to show.

Let n be any positive integer and let $H = n\mathbb{Z}$ be a subroup of $(\mathbb{Z}, +, 0)$. It is easy to check that H is normal. Infact since $\mathbb{Z}$ is Abelian, H is Abelian and any subgroup of an Abelian group must be normal (easy to prove). What are the cosets of H? Any coset of H is given by the set $a * H$. Since we are in an additive group, $*$ here is (addition) $+$. Thus if b is in $a + H$, then b must be of the form $a + nk$ for some integer k i.e $(b - a)$ is a multiple of n. How many cosets are there? In other words what is $[\mathbb{Z} : H]$? Let b be any integer, by the division theorem $b = qn + r$, hence b, r are in the same coset. There are n distinct remainders $0, 1, 2, \ldots, n - 1$ and hence there are n distinct cosets given by $0 + H, 1 + H, \ldots, (n - 1) + H$. Since each coset is an equivalence class, we denote it by

$$[0], [1], \ldots, [n].$$

The set of all these equivalent classes partition $\mathbb{Z}$ and we denote them by $\overline{H}$. However, we will given an alternate notation for the above. We denote the set of cosets for $H = n\mathbb{Z}$ as $\mathbb{Z}/n\mathbb{Z}$. A very remarkable observation is that $\mathbb{Z}/n\mathbb{Z}$ can be given a group structure. Let us define addition over $\mathbb{Z}/n\mathbb{Z}$ as follows,

$$[a] + [b] = [a + b],$$

for any $[a], [b] \in \mathbb{Z}/n\mathbb{Z}$. Is this well defined. The left hand side is a set, the right hand side is a set and $+$ is an operation. Does the operation give the same result if we pick different elements from the set? In other words, let $a_1, a_2 \in [a]$ and $b_1, b_2 \in [b]$. Is it the case that $[a_1 + b_1] = [a_2 + b_2]$? If $a_1 \in [a]$, then $a_1 = a + nk_1$. Similarly $b_1 = b + nl_1$. Thus $a_1 + b_1 = (a + b) + nh_1$ where $h_1 = k_1 + l_1$. Hence $a_1 + b_1 \in [a + b]$. Similarly $a_2 + b_2 \in [a + b]$ and thus the notion is well defined. Hence we can state that

$$(\mathbb{Z}/n\mathbb{Z}, +, [0]),$$

is a group. Consider the map

$$\phi : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$$

given by $\phi(a) = [a]$. Note that ϕ is surjective. Moreover, $\phi(a + b) = \phi(a) + \phi(b)$ and hence ϕ is a homomorphism. What is $\ker(\phi)$? It is precisely $n\mathbb{Z}$, since for any

$b \in n\mathbb{Z}$, $b \in 0 + n\mathbb{Z}$ and the other way too, and hence $[b] = [0]$ i.e. $\phi(b) = [0]$. Hence $\ker(f) = n\mathbb{Z}$. Thus, starting with a normal subgroup $H = n\mathbb{Z}$, we were able to construct a homomorphism which had H as its kernel. This is an important result and we pursue it in the following paragraphs. First, an important proposition.

Proposition 1.3.4. *If H is a normal subgroup of G, then $aH = Ha$ for every $a \in G$.*

Proof. Let h be any element of H. Then for any a in G, aha^{-1} is in H. Let x be an element of aH. Then $x = ah$ for some $h \in H$. Now $xa^{-1} = aha^{-1}$ which is in H because H is normal. Let $xa^{-1} = h'$. Then $x = h'a \in Ha$. Thus $aH \subset Ha$. A similar argument leads to $Ha \subset aH$, hence $aH = Ha$. □

Proposition 1.3.5. *Let $H \triangleleft G$. Then $(aH) * (bH) := (ab)H$ is a well defined operation.*

Proof. Let $aH = a'H$ and let $bH = b'H$. We need to show that $(ab)H = (a'b')H$. Let $d \in (ab)H$. Then $d = abh$ for some $h \in H$. Thus $d \in a(bH)$ which means that $d \in a(Hb)$. Hence $d = ah_1b$ for some $h_1 \in H$. But $ah_1 = a'h_2$ for some $h_2 \in H$ and hence $d = a'h_2b$ wich means that $d \in a'(Hb)$ and hence $d \in a'(bH)$ i.e $d = a'bh_3$ and using the fact that $bh_3 = b'h_4$ for some $h_4 \in H$ we get $d = a'b'h_4$ and hence $d \in (a'b')H$. The other inclusion is analogous. □

Example 1.3.1. *The assumption that H be a normal subgroup is crucial. For example let $G = S_3$ and let $H = \{\epsilon, \beta\}$. Then $\delta H = \{\delta, \alpha\} = \alpha H$ and $\kappa H = \{\kappa, \beta\} = \gamma H$. However, $(\delta \circ \kappa)H = H \neq (\alpha \circ \gamma)H = \kappa H$.*

Example 1.3.2. *Let $H = n\mathbb{Z}$. Then $(a + H) + (b + H) = (a + b) + H$ is a well defined operation. Note that $a + H = [a]$ and $b + H = [b]$.*

Definition 1.3.2 (Quotient Group)**.** *Let $H \triangleleft G$ with the group $(G, *, e)$, and consider the set of all cosets of H,*

$$G/H := \{a * H \ : \ a \in G\},$$

called the quotient set. The triple $(G/H, *, H)$ is called the quotient group with the operation $*$ defined as,

$$(a * H) * (b * H) = (a * b) * H.$$

The fact that it is a group is due to the proposition above. Following our discussion in 1.1.1, we will omit the operation $*$.

Proposition 1.3.6. *Let $H \triangleleft G$. The quotient group G/H is a homomorphic image of the group G where the homomorphism is given by,*

$$\phi(a) = aH,$$

for all $a \in G$. Moreover, for any normal subgroup $H \triangleleft G$, the homomorphic map $\phi : G \to G/H$ has H as its kernel.

Proof. The proof uses the same ideas in our discussion of $\mathbb{Z}/n\mathbb{Z}$. □

Let us recapitulate what we have observed so far.

- If G is any group and if $H \triangleleft G$, then we can construct a homomorphic map from G that has H as its kernel. Note that any homomorphic map f on a group G induces a normal subgroup on G given by $\ker f$.
- A normal subgroup $H \triangleleft G$ enables us to form a quotient group G/H whose elements are called cosets of H. The canonical projection map $f : G \to G/H$ given by $f(a) = aH$ is a surjective homomorphism from G onto G/H. The kernel of f is H.
- Any homomorphic map $f : G \to \operatorname{Im} f$ is a surjective map. The image of f partitions G into its cosets given by $\ker f$.

Now we will see every homomorphic image of a group G is a quotient group of G. More exactly, every homomorphic image of G is isomorphic to a quotient group of G.

Theorem 1.3.2 (Fundamental homomorphism theorem)**.** Let $f : G \to H$ be a surjective homomorphism. If $K = \ker f$, then

$$H \cong G/K.$$

Note that if f is not surjective we can replace H by $\operatorname{Im} f$.

Proof. Since f is surjective $H = \operatorname{Im} f$, and hence H partitions G by the following equivalence relation,

$$a \sim b \iff f(a) = f(b).$$

Now, G/K also partions G given by the relation,

$$a \sim b \iff b = ak,$$

for some $k \in K$. Thus we need to find a correspondence between these two partitions. But this is easy, define

$$\phi : H \to G/K,$$

given by $\phi(y) = aK$, where $y = f(a)$. Is ϕ injective? Let $\phi(y_1) = \phi(y_2)$ where $y_1 = f(a_1)$ and $y_2 = f(a_2)$. Then we have $a_1K = a_2K$ which means that $a_2 = a_1k$ for some $k \in K$. Hence $f(a_2) = f(a_1k) = f(a_1)f(k) = f(a_1)e_2 = f(a_1)$. Thus $y_2 = y_1$.

Is ϕ surjective? Any element of G/K is given by aK for some $a \in G$ and so $f(a) = aK$. Finally, we need to show that ϕ is a homomorphism. Let $y_1 = f(a)$ and $y_2 = f(b)$ for some $a, b \in G$. Note that $f(a)f(b) = f(ab)$ because f is a homorphism. Then,

$$\phi(y_1y_2) = \phi(f(a)f(b)) = \phi(f(ab)) = (ab)K = aKbK = \phi(y_1)\phi(y_2).$$

□

1.4. Rings and Fields

In this section, we will will look at sets that have two (algebraic) operations defined on it. We have studied sets with a single operation in some detail in the preceeding sections. Such sets were called Groups. If in a group we can identify another operation, then we are increasing the complexity of its algebraic structure. In the simplest form, the only things we would need in order to understand the effect of

the second operation are that the second operation be associative and there is a well defined interplay between the two operations.

Traditionally we call the first operation addition and the second operation multiplication. These don't have to correspond to the notion of additon and multiplication in number system and infact can be quite bizzare. We define the simplest algebraic structure with two operations as an object with 4 things, $(G, +, 0, \cdot)$ where G is a non-empty set, $+$ is a well defined operation called addition, 0 is the identity element w.r.t. addition and $\cdot$ is a well defined operation called multiplication.

Definition 1.4.1 (Ring)**.** *By a ring $(G, +, 0, \cdot)$, we mean a set G with operation $+, \cdot$ called addition and multiplication which satisfy the following axioms:*

- *$(G, +, 0)$ is an Abelian group.*
- *Multiplication $(\cdot)$ is associative i.e. $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all a, b, c in G.*
- *Multiplication is* distributive *over addition. That is, for all a, b, c in G,*
  - *$a \cdot (b + c) = a \cdot b + a \cdot c$,*
  - *$(b + c) \cdot a = b \cdot a + c \cdot a$.*

Example 1.4.1. *The simplest examples are the familiar number system $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$.*

Example 1.4.2. *We saw that $\mathbb{Z}_n$ is the quotient group $(\mathbb{Z}/n\mathbb{Z}, +, [0])$. We can also define $\cdot$ as $[a] \cdot [b] := [a \cdot b]$, where $a \cdot b$ is the multiplication of integers, which is well defined. It is easy to check that $(\mathbb{Z}/n\mathbb{Z}, +, [0], \cdot)$ is a ring.*

Remark 1.4.1. *We will omit the $\cdot$ for multiplication from here on and just denote $a \cdot b$ as ab for any a, b in G.*

Remark 1.4.2. *Since $(G, +, 0)$ is an additive group in the ring, we must be careful in denoting the additive inverses and group operation. Thus for any a, b, the addition operation will be denoted by $a+b$. Similarly, for any a the additive inverse will be denoted by $(-a)$ and hence $a + (-a) = 0 = (-a) + a$. The cancellation property will read $a + b = a + c$ implies $b = c$ and $a + b = 0$ implies $a = (-b)$ and $b = (-a)$. a^n will mean $a + a + a + \cdots + a$, n times. The additive inverse of any element a will be called the* ***negative*** *of a*

What happens in a ring when we multiply element in G with 0 and with additive inverses i.e. negatives.

Proposition 1.4.1. *Let G be a ring (here we identify a set with the ring instead of being technically correct) and let a, b be any element in G. Then,*

(1) $a0 = 0a = 0$,

(2) $a(-b) = -(ab) = -(a)b$,

(3) $(-a)(-b) = ab$.

Proof. We prove in order.

(1) First, observe that

$$0 + a0 = a0 = a(0 + 0) = a0 + a0.$$

Thus, from cancellation law in $0 + a0 = a0 + a0$

$$0 = a0.$$

(2) This follows from,

$$a(-b) + ab = a(-b + b) = a0 = 0.$$

(3) Using the above twice,

$$(-a)(-b) = -(a)\,[-b] = -\,[-(ab)]\,.$$

Since $(-(-a)) = a$ we get the result.

□

We have catalogued the basic properties of a ring. Usually we see rings with some additional properties and we will now describe them. By definition, a ring is commutative w.r.t. to addition. When multiplication is also commutative, we give it a different name.

Definition 1.4.2 (Commutative ring)**.** *A ring where multiplication is commutative is called a commutative ring.*

Note that we have not demanded that there be a identity element w.r.t multiplication in a ring.

Definition 1.4.3 (Ring with unity)**.** *When a ring has an identity element w.r.t multiplication, we call the ring a ring with unity and denote the identity element by* 1*. If in addition to an identity element, the ring is commutative we call it a commutative ring with unity.*

Example 1.4.3. *The number systems* $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ *all are commutative rings with unity.*

How are 0 and 1 related?

Definition 1.4.4. *A trivial ring is a ring whose only element is* 0*.*

For any non-trivial ring with unity $1 \neq 0$. This is because if $1 = 0$, then $x = 1x = 0x = 0$ for any x which is a contradiction since we stated that we have a non-trivial ring. Once we have an identity element w.r.t. multiplication, we can ask if there are multiplicative inverses. First observe that 0 doesnot have a multiplicative inverse in a non-trival ring. For if x was the multiplicative inverse of 0, then, $0 = 0x = 1$ which we showed is not possible in non-trivial rings. This leads to a very important algebraic structure.

Definition 1.4.5 (Field)**.** *A field* $\mathbb{F}$ *is a commutative ring with unity in which every non-zero element is invertible w.r.t. multiplication. Thus a field is an object* $(\mathbb{F}, +, 0, \cdot, 1)$*, with a non-empty set* $\mathbb{F}$ *and two operations* $+, \cdot$ *called addition and multiplication respectively, with their respective identity elements* $0, 1$ *such that:*
(1) $(\mathbb{F}, +, 0)$ *is an Abelian group and,*
(2) $(\mathbb{F} - \{0\}, \cdot, 1)$ *is an Abelian group.*

Example 1.4.4. $\mathbb{Z}$ *is not a field. The only elements that are invertible w.r.t. multiplication are* $-1, 1$*. The other number systems* $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ *are fields.*

Example 1.4.5. *Let us consider the ring* $(\mathbb{Z}/4\mathbb{Z}, +, [0], \cdot)$. *The multiplication table is given by,*

$\cdot$	$[0]$	$[1]$	$[2]$	$[3]$
$[0]$	$[0]$	$[0]$	$[0]$	$[0]$
$[1]$	$[0]$	$[1]$	$[2]$	$[3]$
$[2]$	$[0]$	$[2]$	$[0]$	$[2]$
$[3]$	$[0]$	$[3]$	$[2]$	$[1]$

Note that $[2]$ *doesn't have a multiplicative inverse and hence* $\mathbb{Z}/4\mathbb{Z}$ *is not a field.* ($\mathbb{Z}/p\mathbb{Z}$ *where* p *is prime will be a field. See A.1 for details.)*

We have learnt that when the product of two numbers say ab is 0 then one of the factors must be 0 i.e.

$$ab = 0 \implies a = 0 \text{ or } b = 0.$$

This is true in any number system, however for a general ring this is generally true. From the table above we see that $[2]\,[2] = [0]$, but neither of them are $[0]$.

Definition 1.4.6 (Divisors of zero)**.** *In any ring, a non-zero element* a *is called a divisor of zero if there is a* non-zero *element* b *in the ring such that* $ab = 0$.

We also know from number systems that if $ax = ay$ and if $a \neq 0$ then $x = y$. This is the cancellation property. In a field, this is certainly true. However in general rings this is false. Again, from the table above, note that $[2]\,[1] = [2]\,[3]$, but $[1] \neq [3]$.

Definition 1.4.7 (Cancellation property)**.** *A ring is said to have the cancellation propery if, for any* a, b, c *in the ring* $ab = ac$ *or* $ba = ca$ *implies* $b = c$, *provided* $a \neq 0$.

Proposition 1.4.2. *A ring has the cancellation property iff it has no divisors of zero.*

Proof. Let the ring have the cancellation property and let $ab = 0$. Assume $a \neq 0$. Now $ab = 0 = a0$ and so by cancellation $b = 0$. Thus there are no divisors of zero. Let the ring have no divisors of zero and let $ab = ac$ such that $a \neq 0$. Then $ab - ac = 0$ and so $a(b - c) = 0$ which means that $b - c = 0$ and hence $b = c$. □

Definition 1.4.8 (Integral domain)**.** *A ring is said to be an integral domain if it is a commutative ring with unity having the cancellation property.*

Example 1.4.6. *The integers* $\mathbb{Z}$ *is an integral domain because it has no divisors of zero. Note that an integral domain need not be a field since we saw that* $\mathbb{Z}$ *is not a field. However, every field is an integral domain since it has the cancellation property.*

We end our discussion on basic group theory. In the next chapter we will start with the study of linear algebra.

Chapter 2

Basic Linear Algebra

2.1. Vector spaces

Just as groups and rings were an abstraction of number systems, vector spaces are an abstraction of geometry. In mordern geometry, a point in 3 dimensional space is abstracted as a vector. Typically, we identify a co-ordinate system and an origin to represent the $3D$ space around as. A **point** in space is identified with a **vector** whose end is rooted in the origin and whose tip is placed at the point in question. Given a vector in a $3D$ space we can multiply it by a real number c called a **scalar** to stretch or shrink the vector. See 2.1. When c is negative we reverse the direction of the vector. We can add two vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ to get a new vector $\boldsymbol{v}_1 + \boldsymbol{v}_2$. Addition of two vectors is certainly commutative as is shown in 2.2. Note

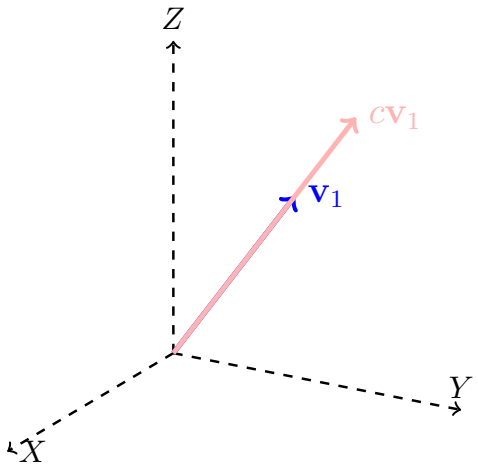

Figure 2.1. Scalar multiplying a vector

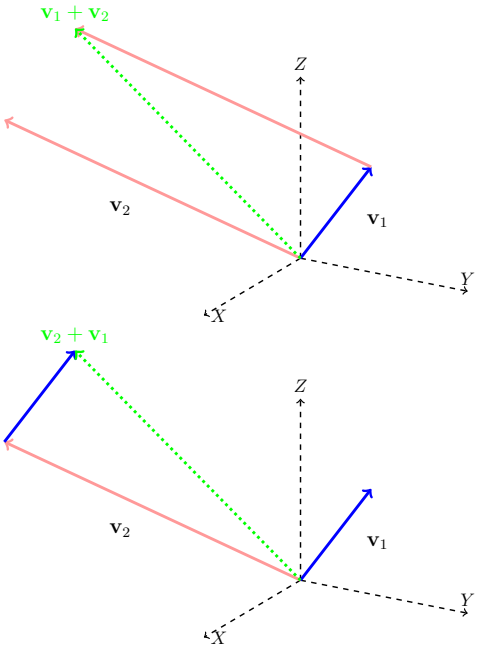

Figure 2.2. Vector addition

that whene we add $\boldsymbol{v}_1$ to $-\boldsymbol{v}_1$ we get the zero vector which is the origin of our co-ordinate system.

We can abstract this notion by defininining an algebraic structure that mimics the algebra of vectors in $3D$ space. Let $\mathbb{F}$ be any field.

Definition 2.1.1. *A vector space over a field $\mathbb{F}$ is a triple $(\mathcal{V}, +, \alpha)$ where,*

- *$\mathcal{V}$ is a set called the set of vectors.*
- *$+ \ : \mathcal{V} \times \mathcal{V} \to \mathbb{F}$ is a binary operation.*
- *$\alpha \ : \mathbb{F} \times \mathcal{V} \to \mathbb{F}$ is a scalar mulitiplication map such that $\alpha(\lambda, \boldsymbol{v}) := \lambda \cdot \boldsymbol{v}$,*

and,

(1) $(\mathcal{V}, +, \mathbf{0}_\mathcal{V})$ is a (commutative) Abelian group.
(2) For any $\beta, \gamma \in \mathbb{F}$ and $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{V}$

- *$\beta \cdot (\boldsymbol{u} + \boldsymbol{v}) = \beta \cdot \boldsymbol{u} + \beta \cdot \boldsymbol{v}$.*
- *$(\beta + \gamma) \cdot \boldsymbol{u} = \beta \cdot \boldsymbol{u} + \gamma \cdot \boldsymbol{u}$.*
- *$\beta \cdot (\gamma \cdot \boldsymbol{u}) = (\beta\gamma) \cdot \boldsymbol{u}$.*
- *$1 \cdot \boldsymbol{u} = \boldsymbol{u}$.*

We usually do not denote the scalar multiplication map and in short a vector space over a field is denoted by $\mathcal{V}/\mathbb{F}$.

In this definition we have used **boldface** letters for arbitrary vectors and greek letters for scalars (field elements). We can immediately observe two useful properties from the definition,

Observation 2.1.1. *Let $\mathcal{V}/\mathbb{F}$ be a vector space. For any $\boldsymbol{v} \in \mathcal{V}$,*

(1) $0 \cdot \boldsymbol{v} = \mathbf{0}_{\mathcal{V}}$

Proof. We can write $0 \cdot \boldsymbol{v}$ as $(0+0) \cdot \boldsymbol{v}$ which is equal to $0 \cdot \boldsymbol{v} + 0 \cdot \boldsymbol{v}$. There is an additive inverse $-0 \cdot \boldsymbol{v}$ such that $-0 \cdot \boldsymbol{v} + 0 \cdot \boldsymbol{v} = \mathbf{0}_{\mathcal{V}}$. Thus,

$$\begin{aligned} & 0 \cdot \boldsymbol{v} = 0 \cdot \boldsymbol{v} + 0 \cdot \boldsymbol{v} \\ \iff & 0 \cdot \boldsymbol{v} + -0 \cdot \boldsymbol{v} = 0 \cdot \boldsymbol{v} + 0 \cdot \boldsymbol{v} + -0 \cdot \boldsymbol{v} \\ \iff & \mathbf{0}_{\mathcal{V}} = 0 \cdot \boldsymbol{v} + \mathbf{0}_{\mathcal{V}} \\ \iff & \mathbf{0}_{\mathcal{V}} = 0 \cdot \boldsymbol{v}. \end{aligned}$$

□

(2) $-1 \cdot \boldsymbol{v} = -\boldsymbol{v}$.

Proof. We can write $0 \cdot \boldsymbol{v}$ as $(1 + (-1)) \cdot \boldsymbol{v}$ and use the above observation along with the fact that $1 \cdot \boldsymbol{v} = \boldsymbol{v}$ □

The two observations allow us to write $\beta \cdot \boldsymbol{v}$ as $\beta \boldsymbol{v}$ for any $\beta \in \mathbb{F}$ and $\boldsymbol{v} \in \mathcal{V}$.

Example 2.1.1. *We list a few vector spaces.*

(1) Let $\mathbb{F} = \mathbb{R}$ and $\mathcal{V} = \mathbb{R}^n$. For any $\boldsymbol{u} \in \mathcal{V}$ we can write $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)$. Define $+ \ : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ to be $\boldsymbol{u} + \boldsymbol{v} = (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n)$ and $\alpha \ : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ to be $\beta \cdot \boldsymbol{u} = (\beta u_1, \beta u_2, \ldots, \beta u_n)$. Then $\mathbb{R}^n/\mathbb{R}$ is a vector space. We can think of $\mathbb{R}^1$ as a vector space over $\mathbb{R}$. When $n = 3$, we get our usual vector algebra in the 3D space that motivated our discussion of vector spaces.

(2) Let X be a non-empty set and let $\mathbb{F}$ be a field. We define,

$$\mathbb{F}(X) := \{f \ : \ f \ : X \to \mathbb{F}\},$$

with the following maps, $+ \ : \mathbb{F}(X) \times \mathbb{F}(X) \to \mathbb{F}$ as follows,

$$\forall\, (f, g \in \mathbb{F}(X)), \quad (f+g)(x) := f(x) + g(x),$$

for any $x \in X$ and, $\alpha \ : \mathbb{F} \times \mathbb{F}(X) \to \mathbb{F}$ as

$$\forall\, (\beta \in \mathbb{F}), \quad (\beta \cdot f)(x) := \beta f(x),$$

for any $x \in X$. Then $\mathbb{F}(X)/\mathbb{F}$ is a vector space where the $\mathbf{0}_{\mathbb{F}(X)}$ is the zero function that maps all points in X to $0 \in \mathbb{F}$.

A particular vector space of the kind discussed in (2) above is the set of all polynomials.

Definition 2.1.2. *A polynomial defined on a set X with coefficients from a field $\mathbb{F}$ is an expression of the form,*

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n,$$

where n is a non-negative integer. The degree of a polynomial is the largest exponent k of x such that $a_k \neq 0$. We denote the space of polynomials as $\mathbb{F}[X]$ or sometimes when X is obvious from the context as $P(\mathbb{F})$.

Remark 2.1.1. *Note that $\mathbb{F}^n = \mathbb{F} \times F \times \cdots \times \mathbb{F}$, where the product is taken n times. It is customary to express elements of $\mathbb{F}^n$ as column vectors. Thus if $\boldsymbol{v} \in \mathbb{F}^3$, then we express $\boldsymbol{v} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}$ for some α, β, γ in $\mathbb{F}$ which are called the **co-ordinates** of $\boldsymbol{v}$.*

Just as in groups, we want to examine subsets of a given set that have the same algebraic structure.

Definition 2.1.3. *Let $\mathcal{V}/\mathbb{F}$ be a vector space. A subset $\mathcal{W} \subset \mathcal{V}$ is called a subspace of $\mathcal{V}$ if $\mathcal{W}$ is a vector space over $\mathbb{F}$.*

The following theorem shows that we only need to check 3 conditions on a subset to determine if it is a subspace.

Theorem 2.1.1. Let $\mathcal{V}/\mathbb{F}$ be a vector space and let $\mathcal{W} \subset \mathcal{V}$ be a subset of $\mathcal{V}$. Then $\mathcal{W}/\mathbb{F}$ is a subspace of $\mathcal{V}$ if and only if,

- $\mathbf{0}_{\mathcal{V}} \in \mathcal{W}$.
- (closed under addition) $\boldsymbol{x} + \boldsymbol{y} \in \mathcal{W}$, whenever $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{W}$.
- (closed under scalar multiplication) $\beta \cdot \boldsymbol{x} \in \mathcal{W}$, whenever $\boldsymbol{x} \in \mathcal{W}$ for any $\beta \in \mathbb{F}$.

Example 2.1.2. *Let $C(\mathbb{F})$ be the set of all continuous functions defined on a set X. Clearly $C(\mathbb{F}) \subset \mathbb{F}(X)$ and it satisfies all the three properties above and so is a subspace of $\mathbb{F}(X)$.*

Example 2.1.3. *Let $P_n(\mathbb{F})$ be the space of polynomials of degree less than or equal to n. Then $P_n(\mathbb{F})$ is a subspace of $P(\mathbb{F})$.*

Given a collection of subspaces of a vector space, we can construct a new subspace.

Theorem 2.1.2. Any intersection of subspaces of a vector space $\mathcal{V}$ is a subspace of $\mathcal{V}$. It is the largest subspace that is contained in each subspace.

Proof. Let $\mathscr{V}$ be a collection of subspaces of $\mathcal{V}$ and let

$$\mathcal{W} = \bigcap_{\mathcal{U} \in \mathscr{V}} \mathcal{U} = \{\boldsymbol{x} \in \mathcal{V} \ : \ \boldsymbol{x} \in \mathcal{U}, \quad \forall\, (\mathcal{U} \in \mathscr{V})\}\,.$$

To show that $\mathcal{W}$ is a subspace of $\mathcal{V}$ we need to show all the conditions in 2.1.1 are satisfied. Since $\mathbf{0}_{\mathcal{V}}$ is in all the $\mathcal{U}$, it is also in $\mathcal{W}$. Consider $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}$ such that $\boldsymbol{x}, \boldsymbol{y}$ are in $\mathcal{W}$. Then by definition $\boldsymbol{x}, \boldsymbol{y}$ are in all $\mathcal{U}$ and hence $\boldsymbol{x} + \boldsymbol{y}$ is in all $\mathcal{U}$ and thus $\boldsymbol{x} + \boldsymbol{y}$ is in $\mathcal{W}$. Similary, it is closed under scalar multiplication. □

We can ask what is the smallest subspace that contains a collection of subspaces of a vector space. We might think the union of such a collection would give us an answer, but it is easy to check that given two subspaces $\mathcal{U}, \mathcal{W}$, their union may fail to be a subspace. To answer this, let us start with a definition.

Definition 2.1.4. *Suppose $\mathcal{V}$ is a vector space and let $\mathcal{U}_1, \mathcal{U}_2, \ldots, \mathcal{U}_m$ be subspaces of $\mathcal{V}$. Then sum of $\mathcal{U}_1, \mathcal{U}_2, \ldots, \mathcal{U}_m$ is a subset of $\mathcal{V}$ that contains all possible sums of elements of $\mathcal{U}_1, \mathcal{U}_2, \ldots, \mathcal{U}_m$. We denote it as,*

$$\mathcal{U}_1 + \mathcal{U}_2 + \cdots + \mathcal{U}_m = \{\boldsymbol{u}_1 + \boldsymbol{u}_2 + \cdots + \boldsymbol{u}_m \ : \ \boldsymbol{u}_i \in \mathcal{U}_i,\, 1 \leq i \leq m\}\,.$$

Example 2.1.4. *Suppose $\mathcal{V}$ is a subset of $\mathbb{R}^3$ containing all those vectors whose second and third co-ordinate are zero and $\mathcal{W}$ is a subset of $\mathbb{R}^3$ containing all those vectors whose first and third co-odinates are zero. Then*

$$\mathcal{V}+\mathcal{W}=\left\{\begin{pmatrix}x\\y\\0\end{pmatrix}\in\mathbb{R}^3\ :\ x,y\in\mathbb{R}\right\}$$

is a subset of $\mathbb{R}^3$.

Theorem 2.1.3. Suppose $\mathcal{U}_1,\mathcal{U}_2,\ldots,\mathcal{U}_m$ are subspaces of $\mathcal{V}$. Then $\mathcal{U}_1+\mathcal{U}_2+\cdots+\mathcal{U}_m$ is the smallest subspace of $\mathcal{V}$ containing $\mathcal{U}_1,\mathcal{U}_2,\ldots,\mathcal{U}_m$.

Proof. Since $\mathbf{0}_{\mathcal{V}}$ is in each $\mathcal{U}_i$, the first condition of 2.1.2 is satisfied. Let $\beta\in\mathbb{F}$ be an arbitrary scalar and let $\boldsymbol{x},\boldsymbol{y}$ be in the sum space. Then there are vectors $\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots,\boldsymbol{u}_m$ and $\boldsymbol{v}_1,\boldsymbol{v}_2,\ldots,\boldsymbol{v}_m$ such that $\boldsymbol{u}_i,\boldsymbol{v}_i\in\mathcal{U}_i$ and,

$$\boldsymbol{x}=\boldsymbol{u}_1+\boldsymbol{u}_2+\cdots+\boldsymbol{u}_m,$$
$$\boldsymbol{y}=\boldsymbol{v}_1+\boldsymbol{v}_2+\cdots+\boldsymbol{v}_m.$$

Thus, $\beta\cdot\boldsymbol{x}+\boldsymbol{y}$ is given by,

$$(\beta\cdot\boldsymbol{u}_1+\boldsymbol{v}_1)+(\beta\cdot\boldsymbol{u}_2+\boldsymbol{v}_2)+\ldots+(\beta\cdot\boldsymbol{u}_m+\boldsymbol{v}_m),$$

which is in the sum space. Suppose $\mathcal{W}$ is any subspace of $\mathcal{V}$ such that $\mathcal{U}_i\subset\mathcal{W}$ for $1\le i\le m$. For any $\boldsymbol{x}$ in the sum space $\boldsymbol{x}=\boldsymbol{u}_1+\boldsymbol{u}_2+\cdots+\boldsymbol{u}_m$, $\boldsymbol{u}_i\in\mathcal{U}_i$ and hence $\boldsymbol{x}\in\mathcal{W}$ and thus the sum space is a subset of $\mathcal{W}$. □

Thus the sum space is the smallest space that contains a collection of **finite** subspaces of a vector space. The next example show that a vector in the sum space doesn't have a unique representation.

Example 2.1.5. *Let,*

$$\mathcal{U}_1=\left\{\begin{pmatrix}x\\y\\0\end{pmatrix}\in\mathbb{R}^3\ :\ x,y\in\mathbb{R}\right\}$$
$$\mathcal{U}_2=\left\{\begin{pmatrix}0\\0\\z\end{pmatrix}\in\mathbb{R}^3\ :\ z\in\mathbb{R}\right\}$$
$$\mathcal{U}_3=\left\{\begin{pmatrix}0\\y\\y\end{pmatrix}\in\mathbb{R}^3\ :\ y\in\mathbb{R}\right\}$$

Any $\boldsymbol{v}=\begin{pmatrix}x\\y\\z\end{pmatrix}\in\mathbb{R}^3$ can be written as $\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}x\\y\\0\end{pmatrix}+\begin{pmatrix}0\\0\\z\end{pmatrix}+\begin{pmatrix}0\\0\\0\end{pmatrix}$ and hence $\mathcal{U}_1+\mathcal{U}_2+\cdots+\mathcal{U}_m=\mathbb{R}^3$. However, the representation is not unique. For example we can write

$$\begin{pmatrix}0\\0\\0\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}+\begin{pmatrix}0\\0\\0\end{pmatrix}+\begin{pmatrix}0\\0\\0\end{pmatrix},$$

and

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

When we can write a vector in the sum space uniquely as the sum of vectors in each subspace comprising the sum space we call it a **Direct sum**.

Definition 2.1.5 (Direct Sum)**.** *Suppose $\mathcal{V}$ is a vector space and let $\mathcal{U}_1, \mathcal{U}_2, \ldots, \mathcal{U}_m$ be subspaces of $\mathcal{V}$. The sum $\mathcal{U}_1 + \mathcal{U}_2 + \cdots + \mathcal{U}_m$ space is called a direct sum if each element in the sum space can be written uniquely as a sum $\boldsymbol{u}_1 + \boldsymbol{u}_2 + \cdots + \boldsymbol{u}_m$, where each $\boldsymbol{u}_j$ is in $\mathcal{U}_j$. We denote the direct sum as,*

$$\mathcal{U}_1 \oplus \cdots \oplus \mathcal{U}_m.$$

Example 2.1.6. *Suppose $\mathcal{U}$ is a the subspace of $\mathbb{R}^3$ consisting of those vectors whose last co-ordinate equals 0 and $\mathcal{W}$ is the subspace of $\mathbb{R}^3$ of those vectors whose first two co-ordinate equals 0. Then $\mathcal{U} \oplus \mathcal{W} = \mathbb{R}^3$. This is easily seen as any $\begin{pmatrix}x\\y\\z\end{pmatrix}$ in $\mathbb{R}^3$ is equal to $\begin{pmatrix}x\\y\\0\end{pmatrix} + \begin{pmatrix}0\\0\\z\end{pmatrix}$. Moreover, this is unique since we don't get any contribution from the first two co-ordinates from $\mathcal{W}$ and from the last co-ordinate in $\mathcal{U}$.*

Proposition 2.1.1. *Suppose $\mathcal{U}_1, \mathcal{U}_2, \ldots, \mathcal{U}_m$ are subspaces of $\mathcal{V}$. Then $\mathcal{U}_1 + \mathcal{U}_2 + \cdots + \mathcal{U}_m$ is a direct sum if and only if the only way to write $\mathbf{0}_\mathcal{V}$ as a sum of vectors $\boldsymbol{u}_1 + \boldsymbol{u}_2 + \cdots + \boldsymbol{u}_m$, where each $\boldsymbol{u}_j$ is in $\mathcal{U}_j$, is by taking each $\boldsymbol{u}_j$ equal to $\mathbf{0}_\mathcal{V}$.*

Proof. Note that $\mathbf{0}_\mathcal{V} = \mathbf{0}_\mathcal{V} + \cdots + \mathbf{0}_\mathcal{V}$, where we take the m sums. If $\mathcal{U}_1 + \mathcal{U}_2 + \cdots + \mathcal{U}_m$ is a direct sum then this is the only way to write $\mathbf{0}_\mathcal{V}$ as the sum of $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m$ vectors such that $\boldsymbol{u}_j$ is in $\mathcal{U}_j$.

Now, let $\mathcal{U}_1 + \mathcal{U}_2 + \cdots + \mathcal{U}_m$ be a direct such that if $\mathbf{0}_\mathcal{V} = \boldsymbol{u}_1 + \boldsymbol{u}_2 + \cdots + \boldsymbol{u}_m$, then each $\boldsymbol{u}_j = \mathbf{0}_\mathcal{V}$. Assume $\boldsymbol{v}$ in the sum space can be written as,

$$\boldsymbol{v} = \boldsymbol{u}_1 + \boldsymbol{u}_2 + \cdots + \boldsymbol{u}_m,$$

and

$$\boldsymbol{v} = \boldsymbol{w}_1 + \boldsymbol{w}_2 + \cdots + \boldsymbol{w}_m,$$

where $\boldsymbol{u}_j, \boldsymbol{w}_j$ is in $\mathcal{U}_j$. Now $\boldsymbol{v} - \boldsymbol{v} = \mathbf{0}_\mathcal{V}$ is in the sum space and so,

$$\mathbf{0}_\mathcal{V} = (\boldsymbol{u}_1 - \boldsymbol{w}_1) + (\boldsymbol{u}_2 - \boldsymbol{w}_2) + \ldots + (\boldsymbol{u}_m - \boldsymbol{w}_m).$$

According to our hypothesis, each $(\boldsymbol{u}_j - \boldsymbol{w}_j) = \mathbf{0}_\mathcal{V}$ which means that the representation of $\boldsymbol{v}$ is unique and hence the sum space is a direct sum. □

For a pair of subspaces, the condition to check if their sum space is a direct sum becomes simpler.

Proposition 2.1.2. *Suppose $\mathcal{U}, \mathcal{W}$ are subspaces of $\mathcal{V}$. Then $\mathcal{U} + \mathcal{W}$ is a direct sum if and only if $\mathcal{U} \cap \mathcal{W} = \{\mathbf{0}_\mathcal{V}\}$.*

Proof. First assume we have $\mathcal{U} \oplus \mathcal{W}$. Let $\boldsymbol{v}$ be a vector in $\mathcal{U} \cap \mathcal{W}$. Then $-\mathcal{V}$ is also in $\mathcal{U} \cap \mathcal{W}$. Hence $\mathbf{0}_{\mathcal{V}} = \boldsymbol{v} + -\boldsymbol{v}$. This means that $\boldsymbol{v} = \mathbf{0}_{\mathcal{V}}$ because in a direct sum the zero vector can only be written as sum of zero vector.

Assume $\mathcal{U} \cap \mathcal{W} = \{\mathbf{0}_{\mathcal{V}}\}$. Let $\mathbf{0}_{\mathcal{V}} = \boldsymbol{u} + \boldsymbol{w}$. This means that $\boldsymbol{u} = -\boldsymbol{w}$ and hence $\boldsymbol{u}$ is in $\mathcal{W}$. Thus $\boldsymbol{u}$ is in $\mathcal{U} \cap \mathcal{W}$. But this implies that $\boldsymbol{u} = \mathbf{0}_{\mathcal{V}}$. Hence $\boldsymbol{w} = \mathbf{0}_{\mathcal{V}}$. □

Our next definition is again motivated by lines and planes in $3D$ geometry.

Definition 2.1.6 (Linear combination)**.** *Let $\mathcal{V}/\mathbb{F}$ be a vector space over the field $\mathbb{F}$ and let S be an non-empty set of $\mathcal{V}$. A vector $\boldsymbol{v} \in \mathcal{V}$ is called a linear combination of the elements of S if there is a **finite** number of elements $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ in S and $\beta_1, \beta_2, \ldots, \beta_n$ in $\mathbb{F}$ such that,*

$$\boldsymbol{v} = \beta_1 \boldsymbol{u}_1 + \cdots + \beta_n \boldsymbol{u}_n.$$

If S has finite vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$, then we say that $\boldsymbol{v}$ is a linear combination of $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$.

Logically, a vector $\boldsymbol{v}$ in $\mathcal{V}$ is a linear combination of vectors in S if,

$$\exists\, (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n;\ \boldsymbol{u}_i \in S)\, \exists\, (\beta_1, \beta_2, \ldots, \beta_n;\ \beta_i \in \mathbb{F})\, [\boldsymbol{v} = \ \beta_1 \boldsymbol{u}_1 + \cdots + \beta_n \boldsymbol{u}_n]\,.$$

Thus a vector $\boldsymbol{v}$ is **NOT** a linear combination of vectors in S if no matter what finite set of vectors we take in S and no matter what corresponding finite set of scalars we take, we cannot write $\boldsymbol{v}$ as a linear combination of those vectors and scalars. Note that the zero vector $\mathbf{0}_{\mathcal{V}}$ is a linear combination of any non-empty subset of $\mathcal{V}$ since $\mathbf{0}_{\mathcal{V}} = 0\boldsymbol{u}$ for any $\boldsymbol{u} \in \mathcal{V}$.

Definition 2.1.7 (Span)**.** *Let S be a non-empty subset of a vector space $\mathcal{V}$. The span of S, denoted by* $\operatorname{Span} S$*, is the set of all the linear combinations of the elements of S. For convenience we define* $\operatorname{Span} S = \{\emptyset\}$.

Example 2.1.7. *Let $\mathcal{V} = \mathbb{R}^3$ over the field $\mathbb{R}$. Let $S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right\}$. Then* $\operatorname{Span} S$ *is the set of all vectors of the form $\boldsymbol{u} = \begin{pmatrix} \beta \\ \beta \\ 0 \end{pmatrix}$ which is seen as the line passing through origin and containing the point $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$.*

If $S = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$, then the $\operatorname{Span} S$ *is the set of all vectors of the form $\boldsymbol{u} = \begin{pmatrix} \beta \\ \gamma \\ 0 \end{pmatrix}$ for some $\beta, \gamma \in \mathbb{R}$. This is seen as the XY plane.*

Proposition 2.1.3. *The span of any subset S of a vector space $\mathcal{V}$ is a subspace of $\mathcal{V}$. Moreover, any subspace $\mathcal{U}$ that contains S must contain* $\operatorname{Span} S$.

Proof. First note that $\mathbf{0}_{\mathcal{V}}$ is contained in $\operatorname{Span} S$. Let $\boldsymbol{v}_1$ be a vector in $\operatorname{Span} S$. Then there are finite vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ in S such that $\boldsymbol{v}_1 = \beta_1\boldsymbol{u}_1 + \cdots + \beta_n\boldsymbol{u}_n$. Let $\boldsymbol{v}_2$ be a vector in $\operatorname{Span} S$. Then ther are finite vectors $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m$ in S such that $\boldsymbol{v}_2 = \gamma_1\boldsymbol{w}_1 + \cdots + \gamma_m\boldsymbol{w}_m$. Then $\boldsymbol{v}_1 + \boldsymbol{v}_2 = \beta_1\boldsymbol{u}_1 + \cdots + \beta_n\boldsymbol{u}_n + \gamma_1\boldsymbol{w}_1 + \cdots + \gamma_m\boldsymbol{w}_m$ is a linear combination of finite vectors in S. The case of scalar multiplication is similar. □

Definition 2.1.8 (Spanning set)**.** *A subset S of vector space $\mathcal{V}$ is called a spanning set of $\mathcal{V}$ if $\operatorname{Span} S = \mathcal{V}$. Thus every vector in V can be written as a linear combination of vectors in S. We say that S **generates** or **spans** $\mathcal{V}$.*

Definition 2.1.9 (Linear dependence)**.** *A subset S of a vector space $\mathcal{V}$ is called linearly dependent if there exists a finite number of **distinct** vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ in S and scalars $\beta_1, \beta_2, \ldots, \beta_n$ in $\mathbb{F}$ such that there is atleast one $\beta_j \neq 0$ and*

$$\beta_1\boldsymbol{u}_1 + \cdots + \beta_n\boldsymbol{u}_n = \mathbf{0}_{\mathcal{V}}.$$

Logically this means that S is linearly dependent set if,

$$\exists\,(\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n)\,\exists\,(\beta_1, \beta_2, \ldots, \beta_n)\,[\exists\,(1 \leq j \leq n)\,; \beta_j \neq 0 \text{ and } \beta_1\boldsymbol{u}_1 + \cdots + \beta_n\boldsymbol{u}_n = \mathbf{0}_{\mathcal{V}}]\,,$$

where each u_i is in S. When is a set S **NOT** linearly dependent? For that to happen we must negate the logical statement above which says that if S is NOT linearly dependent then,

$$\forall\,(\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n)\,\forall\,(\beta_1, \beta_2, \ldots, \beta_n)\,[\beta_1\boldsymbol{u}_1 + \cdots + \beta_n\boldsymbol{u}_n = \mathbf{0}_{\mathcal{V}} \implies \beta_j = 0;\, \forall\,(1 \leq j \leq n)]\,.$$

This is an important notion and needs a definition,

Definition 2.1.10 (Linear independence)**.** *A subset S of vector space $\mathcal{V}$ is Linearly independent if it is not linearly dependent. Thus for any finite set of distinct vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ in S and scalars $\beta_1, \beta_2, \ldots, \beta_n$ in $\mathbb{F}$, if $\beta_1\boldsymbol{u}_1 + \cdots + \beta_n\boldsymbol{u}_n = \mathbf{0}_{\mathcal{V}}$, then each $\beta_j = 0$.*

A method to determine if a given set of finite vectors is linearly dependent will be shown later when we discuss solution of linear systems of equation.

Proposition 2.1.4. *Let $\mathcal{V}/\mathbb{F}$ be a vector field over the field $\mathbb{F}$ and let $S_1 \subset S_2 \subset \mathcal{V}$. Then,*

(1) If S_1 is linearly dependent, then S_2 is also linearly dependent.

(2) If S_2 is linearly independent, then S_1 is also linearly independent.

Proof. We prove (1) since (2) is just the contrapositive of (1). If S_1 is linearly dependent then there are vectors and scalars $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ in S_1 and $\beta_1, \beta_2, \ldots, \beta_n$ in $\mathbb{F}$ respectively, such that atleast one such β_j is not 0 and

$$\beta_1\boldsymbol{u}_1 + \cdots + \beta_n\boldsymbol{u}_n = \mathbf{0}_{\mathcal{V}}.$$

But since $S_1 \subset S_2$, the vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ are in S_2 and they satisfy the condition of linear dependence. Thus S_2 is linearly dependent. □

There is a crucial interplay between linear independence and span. We start with a few useful observation.

Proposition 2.1.5. *Let $S = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$ be a subset of a vector space $\mathcal{V}$ over field $\mathbb{F}$, then S is linearly dependent if and only if there is a vector $\boldsymbol{u}_j \in S$ such that $\boldsymbol{u}_j$ is in* $\operatorname{Span}(S - \{\boldsymbol{u}_j\})$. *Moreover* $\operatorname{Span} S = \operatorname{Span}(S - \{\boldsymbol{u}_j\})$

Proof. Let S be linearly independent. Then there are scalars $\beta_1, \beta_2, \ldots, \beta_n$ in $\mathbb{F}$ and vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ such that,

$$\beta_1 \boldsymbol{u}_1 + \cdots + \beta_n \boldsymbol{u}_n = \boldsymbol{0}_{\mathcal{V}},$$

and there is atleast one β_j such that $\beta_j \neq 0$. Since we have a **Field** $\mathbb{F}$, we know that every non-zero element of a field has a multiplicative inverse and so β_j^{-1} exists and hence,

$$\boldsymbol{u}_j = \frac{1}{\beta_j}(\beta_1 \boldsymbol{u}_1 + \cdots + \beta_{j-1}\boldsymbol{u}_{j-1} + \beta_{j+1}\boldsymbol{u}_{j+1} + \cdots + \beta_n \boldsymbol{u}_n).$$

Hence $\boldsymbol{u}_j$ is in $\operatorname{Span}(\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\} - \{\boldsymbol{u}_j\})$.

If there is a $\boldsymbol{u}_j$ such that $\boldsymbol{u}_j$ is in $\operatorname{Span}(\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\} - \{\boldsymbol{u}_j\})$, then there are scalars such that,

$$\boldsymbol{u}_j = \beta_1 \boldsymbol{u}_1 + \cdots + \beta_{j-1}\boldsymbol{u}_{j-1} + \beta_{j+1}\boldsymbol{u}_{j+1} + \cdots + \beta_n \boldsymbol{u}_n.$$

Thus,

$$\beta_1 \boldsymbol{u}_1 + \cdots + \beta_{j-1}\boldsymbol{u}_{j-1} + (-1)\boldsymbol{u}_j + \beta_{j+1}\boldsymbol{u}_{j+1} + \cdots + \beta_n \boldsymbol{u}_n = \boldsymbol{0}_{\mathcal{V}}.$$

Hence S is linearly independent. It is easy to check that $\operatorname{Span} S = \operatorname{Span}(S - \{\boldsymbol{u}_j\})$. □

Proposition 2.1.6. *Let S be a linearly independent subset of a vector space $\mathcal{V}$, and let $\boldsymbol{v}$ be an element of V that is not in S. Then $S \cup \{\boldsymbol{v}\}$ is linearly dependent if and only if $\boldsymbol{v} \in \operatorname{Span} S$.*

Proof. Let $\boldsymbol{v}$ be in $\operatorname{Span} S$. Then there are finite vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ in S and scalars $\beta_1, \beta_2, \ldots, \beta_n$ in $\mathbb{F}$ such that,

$$\boldsymbol{v} = \beta_1 \boldsymbol{u}_1 + \cdots + \beta_n \boldsymbol{u}_n.$$

Without loss of generality, we can assume $\boldsymbol{u}$'s to be distinct. This means that,

$$(-1)\boldsymbol{v} + \beta_1 \boldsymbol{u}_1 + \cdots + \beta_n \boldsymbol{u}_n = \boldsymbol{0}_{\mathcal{V}}.$$

Since $\boldsymbol{v}$ is not in S, none of the vectors $\boldsymbol{u}_j$ is equal to $\boldsymbol{v}$. Thus we have shown that the set $S \cup \{\boldsymbol{v}\}$ is linearly dependent.

Assume $S \cup \{\boldsymbol{v}\}$ is a linearly dependent set. Then there is a set of finite distinct vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ in $S \cup \{\boldsymbol{v}\}$ and scalars $\beta_1, \beta_2, \ldots, \beta_n$ such that,

$$\beta_1 \boldsymbol{u}_1 + \cdots + \beta_n \boldsymbol{u}_n = \boldsymbol{0}_{\boldsymbol{v}},$$

and atleast one such β_j is not equal to 0. One of the $\boldsymbol{u}_j$'s must be $\boldsymbol{v}$. If not all the vectors are from S and S being linearly independent will contradict our assertion. Re-ordering, let $\boldsymbol{u}_n$ be $\boldsymbol{v}$, then

$$\beta_n \boldsymbol{v} + \beta_1 \boldsymbol{u}_1 + \cdots + \beta_n \boldsymbol{u}_n = \boldsymbol{0}_{\mathcal{V}}.$$

Again, with similar reasoning β_n cannot be 0. Since we have a **Field** $\mathbb{F}$, we know that every non-zero element of a field has a multiplicative inverse and so β_n^{-1} exists and hence,

$$\boldsymbol{v} = \frac{1}{\beta_n^{-1}}(\beta_1 \boldsymbol{u}_1 + \cdots + \beta_n \boldsymbol{u}_n).$$

Thus $\boldsymbol{v}$ is in Span S. □

The proposition can be equivalently restated as follows:

Remark 2.1.2. *Let S be a linearly independent subset of a vector space $\mathcal{V}$, and let $\boldsymbol{v}$ be an element of V that is not in S. Then $S \cup \{\boldsymbol{v}\}$ is linearly independent if and only if $\boldsymbol{v} \notin$ Span S.*

Theorem 2.1.4 (Exchange Lemma)**.** Let $\mathcal{V}/\mathbb{F}$ be a vector space over a field $\mathbb{F}$. Consider the following sets:

- $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$,
- $L = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$,

such that Span $S = V$ and L is linearly independent. Then $m \geq n$.

Proof. We prove by contradiction. Assume $n > m$. The proof will use a construction of sets S_i and L_i in each stage i, where, at the end of the stage we replace one element in S_i by $\boldsymbol{u}_i$. We will get a contradiction at stage m.

Stage 1 Since Span $S = \mathcal{V}$, we can write $\boldsymbol{u}_1$ as,

$$\boldsymbol{u}_1 = \beta_1 \boldsymbol{v}_1 + \cdots + \beta_m \boldsymbol{v}_m.$$

There must be one β_j such that β_j is not equal to 0. Let us re-order so that the corresponding $\boldsymbol{v}_j$ is $\boldsymbol{v}_m$. Hence,

$$\boldsymbol{u}_1 = \beta_m \boldsymbol{v}_m + \beta_1 \boldsymbol{v}_1 + \cdots + \beta_{m-1} \boldsymbol{v}_{m-1},$$

and thus,

$$\boldsymbol{v}_m = \frac{1}{\beta_m}(\boldsymbol{u}_1 + -\beta_1 \boldsymbol{v}_1 + \cdots + -\beta_{m-1} \boldsymbol{v}_{m-1}).$$

Replace $\boldsymbol{v}_m$ by $\boldsymbol{u}_1$.

$$S_1 = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{m-1}, \boldsymbol{u}_1\},$$
$$L_1 = \{\boldsymbol{u}_2, \ldots, \boldsymbol{u}_m, \ldots, \boldsymbol{u}_n\}.$$

Claim:

- Span $S_1 = V$,
- L_1 is linearly independent.
- Number of $\boldsymbol{v}_j$ in S_1 is $m - 1$.

Since $L_1 \subset L$ and L was linearly independent, L_1 is linearly independent. For any $\boldsymbol{w}$ in $\mathcal{V}$, $\boldsymbol{w} = \gamma_1 \boldsymbol{v}_1 + \cdots + \gamma_m \boldsymbol{v}_m$. Subsituting for $\boldsymbol{v}_m$ we see that $\boldsymbol{w}$ is in Span S_1. The fact that we have $m - 1$ $\boldsymbol{v}$'s in S_1 is by replacing one such $\boldsymbol{v}$'s (namely $\boldsymbol{v}_m$) from S to get S_1.

Stage 2 Since Span $S_1 = \mathcal{V}$, we can write $\boldsymbol{u}_2$ as,

$$\boldsymbol{u}_2 = \beta_1 \boldsymbol{v}_1 + \cdots + \beta_{m-1} \boldsymbol{v}_{m-1} + \beta_m \boldsymbol{u}_m.$$

First note that β_j's for $1 \leq j \leq m-1$ cannot be zero, since that would mean that $\boldsymbol{u}_2$ is in Span $\{\boldsymbol{u}_1\}$. Thus there exists a j such that $1 \leq j \leq m-1$ and β_j is not equal to 0. Let us re-order such that the corresponding $\boldsymbol{v}_j$ is $\boldsymbol{v}_{m-1}$. Hence,

$$\boldsymbol{u}_2 = \beta_1 \boldsymbol{v}_1 + \cdots + \beta_{m-2}\boldsymbol{v}_{m-2} + \beta_{m-1}\boldsymbol{v}_{m-1} + \beta_m \boldsymbol{u}_m.$$

Replace $\boldsymbol{v}_{m-1}$ by $\boldsymbol{u}_2$ and set,

$$S_2 = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{m-2}, \boldsymbol{u}_2, \boldsymbol{u}_1\},$$
$$L_1 = \{\boldsymbol{u}_3, \ldots, \boldsymbol{u}_m, \ldots, \boldsymbol{u}_n\}.$$

Again:

- Span $S_2 = V$,
- L_2 is linearly independent.

Since we have assumed that $n > m$, at the end of the m^{th} stage, we would have exhausted all of $\boldsymbol{v}$'s, that is

$$S_m = \{\boldsymbol{u}_m, \boldsymbol{u}_{m-1}, \ldots, \boldsymbol{u}_2, \boldsymbol{u}_1\},$$
$$L_m = \{\boldsymbol{u}_{m+1}, \ldots, \boldsymbol{u}_n\}.$$

But note that L_m is a subset of L and hence should be linearly independent. But that is not the case, since S_m spans $\mathcal{V}$, we can write,

$$\boldsymbol{u}_{m+1} = \beta_1 \boldsymbol{u}_1 + \cdots + \beta_m \boldsymbol{u}_m,$$

and this is not possible since all β_j cannot be zero for that would mean $\boldsymbol{u}_{m+1} = \boldsymbol{0}_{\mathcal{V}}$, but in that case $-\boldsymbol{u}_{m+1} + \beta_1 \boldsymbol{u}_1 + \cdots + \beta_m \boldsymbol{u}_m = \boldsymbol{0}_{\mathcal{V}}$ with non-zero coefficients. □

These two propositions are extremely useful.

Definition 2.1.11 (Basis)**.** *Let $\mathcal{V}$ be a vector space over field $\mathbb{F}$. A non-empty subset $\mathcal{B}$ of $\mathcal{V}$ is called a basis of $\mathcal{V}$ if* Span $\mathcal{B} = \mathcal{V}$ *and $\mathcal{B}$ is linearly independent. If $\mathcal{B}$ is **finite** then $\mathcal{V}$ is called a **finite dimensional** vector space.*

Note that, by 2.1.6, in a finite dimensional vector space the number of elements in a spanning set is always greater than or equal to the number of elements in a linearly independent set. The crucial property about basis in a finite dimensional vector space that we are going to see, is that every vector will have a unique representation w.r.t a given basis. First a few important observations.

Proposition 2.1.7. *If $\mathcal{V}$ is finite dimensional and S is a subset of $\mathcal{V}$ such that* Span $S = \mathcal{V}$*, then S can be reduced to give a basis for $\mathcal{V}$.*

Proof. Let $S = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$. If S is also linearly independent then S is a basis. If not, then by 2.1.5, there is a vector $\boldsymbol{u}_j \in S$ such that $\boldsymbol{u}_j$ is in the span of the rest of the vectors and the set formed by removing $\boldsymbol{u}_j$ from S doesn't change its span. Let $S_1 = S - \{\boldsymbol{u}_j\}$. If S_1 is linearly independent the we have found a basis. IF not we repeat this process. Since $\mathcal{V}$ is finite dimensional this process will stop eventually yielding a set of distinct finite vectors that is a basis for $\mathcal{V}$. □

Proposition 2.1.8. *If $\mathcal{V}$ is finite dimensional and L is a subset of $\mathcal{V}$ such that L is linearly independent, then L can be extended to a basis for $\mathcal{V}$.*

Proof. Let $L = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$. Why does this set have to be finite? Since $\mathcal{V}$ is finite dimensional there is a set $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$ such that $\operatorname{Span} S = \mathcal{V}$ and by the Exchange lemme we know that $n \leq m$. If L spans $\mathcal{V}$ we have found our basis. If not, there is a $\boldsymbol{w} \in \mathcal{V}$ such that $\boldsymbol{w} \notin \operatorname{Span} L$. But by 2.1.6, $L \cup \{\boldsymbol{w}\}$ is also linearly independent. If this spans $\mathcal{V}$ we are done, if not we continue. This process will stop eventually since we cannot have a linearly independent set of more than m vectors. □

These two propositions enable us to give a unique number to a finite dimensional vector space called its **dimension**

Proposition 2.1.9. *If $\mathcal{V}$ is a finite dimensional vector space, then $\mathcal{V}$ admits a basis of finite vectors. Any two basis in $\mathcal{V}$ must have the same number of elements. Moreover, any vector $\boldsymbol{v}$ in $\mathcal{V}$ can be written uniquely as a linear combination of vectors in the basis for $\mathcal{V}$.*

Proof. If $\mathcal{V}$ is finite dimensional then it has a finite spanning set S. Since S can be reduced to a basis, every finite dimensional vector space admits a basis. Let $\mathcal{B}_1$ be a basis for $\mathcal{V}$ with m elements and let $\mathcal{B}_2$ be a basis for $\mathcal{V}$ with n elements. Since $\mathcal{B}_1$ is spanning and $\mathcal{B}_2$ is linearly independent, $n \leq m$ by the exchange theorem. Reversing the roles we get $m \leq n$. Thus $m = n$.

Let $\mathcal{B} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$ be a basis for $\mathcal{V}$ and let

$$\boldsymbol{v} = \beta_1 \boldsymbol{u}_1 + \cdots + \beta_n \boldsymbol{u}_n,$$

and

$$\boldsymbol{v} = \gamma_1 \boldsymbol{u}_1 + \cdots + \gamma_n \boldsymbol{u}_n$$

be two representation of $\boldsymbol{v}$ in $\mathcal{V}$. Then $\mathbf{0}_{\mathcal{V}} = (\beta - \gamma)_1 \boldsymbol{u}_1 + \cdots + (\beta - \gamma)_n \boldsymbol{u}_n$. Since $\mathcal{B}$ is linearly independent, this must mean that each $\beta_j = \gamma_j$ and hence the representation of $\boldsymbol{v}$ in $\mathcal{V}$ is unique w.r.t. basis $\mathcal{B}$. □

Definition 2.1.12 (Dimension)**.** *If $\mathcal{V}$ is finite dimensional vector space then the number of vectors in a basis $\mathcal{B}$ for $\mathcal{V}$ is called the dimension of the vector space $\mathcal{V}$.*

Proposition 2.1.10. *Let $\mathcal{V}$ be a vector space over the field $\mathbb{F}$ such that $\dim V = n$. Then,*

(1) If S is a subset of $\mathcal{V}$ such that $\operatorname{Span} S = \mathcal{V}$, then the number of elements in S is greater than or equal to n. If number of elements in S is equal to n, then S is a basis for $\mathcal{V}$.

(2) If L is a linearly independent subset of $\mathcal{V}$ that contains n elements then L is a basis for $\mathcal{V}$.

Proof. We prove in order.

(1) Let $\mathcal{B}$ be a basis for $\mathcal{V}$. Then $\mathcal{B}$ has n elements. Since a spanning set (S) cannot have fewer elements than a linearly independent set ($\mathcal{B}$) in a finite dimensional vector space, the number of elements in S is greater than or equal to n. If S has n elements, S can be reduced to a basis. But each basis must have the same number of elements and so S cannot be reduced any further. Hence S is a basis.

(2) L can be extended to a basis for $\mathcal{V}$, but any basis must have the same number of elements (n) and so L must be a basis for $\mathcal{V}$.

□

Proposition 2.1.11. *Let $\mathcal{W}$ be a subspace of a finite dimensional vector space $\mathcal{V}$. Then $\mathcal{W}$ is also finite dimensional and $\dim \mathcal{W} \leq \dim \mathcal{V}$. Moreover, if $\dim \mathcal{W} = \dim \mathcal{V}$, then $V = W$.*

Proof. Let $\dim \mathcal{V} = n$ and let $\mathcal{B}_\mathcal{V}$ be a basis for $\mathcal{V}$. If $\mathcal{W} = \{\mathbf{0}_\mathcal{V}\}$, then $\dim \mathcal{W} = 0 \leq n$. Assume there is a non-zero element $\boldsymbol{v}_1$ in $\mathcal{W}$. If there is a $\boldsymbol{v}_2 \in \mathcal{V}$ such that $\boldsymbol{v}_2$ is not in the span of $\{\boldsymbol{v}_1\}$ and $\boldsymbol{v}_2$ is in $\mathcal{W}$, then $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ is linearly independent subset of $\mathcal{W}$. Continuing this way we can get a linearly independent set in $\mathcal{W}$,

$$\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}.$$

However k cannot be larger than n. Suppose $k = n$, if $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ doesn't span $\mathcal{W}$, then there must be an element $\boldsymbol{v}$ such that $\{\boldsymbol{v}\} \cup \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ is also linearly independent in $\mathcal{W}$. But this cannot happen if $k = n$. Hence $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ also spans $\mathcal{W}$, with $k \leq n$, and so is a basis for $\mathcal{W}$. Thus dimension of $\mathcal{W}$ is less than dimension of $\mathcal{V}$.

If $\dim \mathcal{W} = n$, then the basis of $\mathcal{W}$ is a linearly independent subset of n vectors in $\mathcal{V}$ and hence must be a basis for $\mathcal{V}$. Thus $\mathcal{V} \subset \mathcal{W}$. Hence $\mathcal{V} = \mathcal{W}$. □

The 2.1.4 and the subsequent propositions are extremely useful. We give a few applications of them for the case of sums and direct sums.

Proposition 2.1.12. *If $\mathcal{W}_1, \mathcal{W}_2$ are finite dimensional subspaces of vector space $\mathcal{V}$, then the subspace $\mathcal{W}_1 + \mathcal{W}_2$ is finite dimensional, and*

$$\dim(\mathcal{W}_1 + \mathcal{W}_2) = \dim \mathcal{W}_1 + \dim \mathcal{W}_2 - \dim(\mathcal{W}_1 \cap \mathcal{W}_2).$$

Proof. Let $\dim V = n$. Since $\mathcal{W}_1 + \mathcal{W}_2$ is the smallest subspace containing both $\mathcal{W}_1$ and $\mathcal{W}_2$, it is a subspace of $\mathcal{V}$ and hence finite dimensional. Also, $(\mathcal{W}_1 \cap \mathcal{W}_2)$ is the subspace contained in both $\mathcal{W}_1$ and $\mathcal{W}_2$ and hence it is finite dimensional. Let $\mathcal{B}_{(\mathcal{W}_1 \cap \mathcal{W}_2)} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ where $k \leq \dim \mathcal{W}_1$ and $k \leq \dim \mathcal{W}_2$. Then $\mathcal{B}_{(\mathcal{W}_1 \cap \mathcal{W}_2)}$ can be extended to a basis for $\mathcal{W}_1$, given by,

$$\mathcal{B}_{\mathcal{W}_1} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\} \cup \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p\}.$$

Similarly, $\mathcal{B}_{(\mathcal{W}_1 \cap \mathcal{W}_2)}$ can be extended to a basis for $\mathcal{W}_2$, given by,

$$\mathcal{B}_{\mathcal{W}_2} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\} \cup \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_q\}.$$

Claim:

$$\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\} \cup \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p\} \cup \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_q\}$$

is a basis of $\mathcal{W}_1 + \mathcal{W}_2$. It is easy to see that the above spans $(\mathcal{W}_1 + \mathcal{W}_2)$. To check linear independence, Let

$$\alpha_1 \boldsymbol{v}_1 + \cdots + \alpha_k \boldsymbol{v}_k + \beta_1 \boldsymbol{x}_1 + \cdots + \beta_p \boldsymbol{x}_p + \gamma_1 \boldsymbol{y}_1 + \cdots + \gamma_q \boldsymbol{y}_q = \mathbf{0}_\mathcal{V}.$$

Then,

$$\alpha_1 \boldsymbol{v}_1 + \cdots + \alpha_k \boldsymbol{v}_k + \beta_1 \boldsymbol{x}_1 + \cdots + \beta_p \boldsymbol{x}_p = -\gamma_1 \boldsymbol{y}_1 + \cdots + \gamma_q \boldsymbol{y}_q.$$

Since the vector on the left hand side is an element of $\mathcal{W}_1$ and the vector on the right hand side is an element of $\mathcal{W}_2$, they must be an element of both, i.e. $(\mathcal{W}_1 \cap \mathcal{W}_2)$. Hence,

$$\gamma_1 \boldsymbol{y}_1 + \cdots + \gamma_q \boldsymbol{y}_q = \delta_1 \boldsymbol{v}_1 + \cdots + \delta_k \boldsymbol{v}_k.$$

Thus,

$$(\alpha - \delta)_1 \boldsymbol{v}_1 + \cdots + (\alpha - \delta)_k \boldsymbol{v}_k + \beta_1 \boldsymbol{x}_1 + \cdots + \beta_p \boldsymbol{x}_p = \boldsymbol{0}_{\mathcal{V}}.$$

These vectors are a basis for $\mathcal{W}_1$ and so $\beta_j = 0$ for all $1 \leq j \leq p$. Thus we see that α_i's and γ_l's are all zero since they are co-efficients for basis in $\mathcal{W}_2$. Thus,

$$\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\} \cup \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p\} \cup \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_q\}$$

is a basis for $(\mathcal{W}_1 + \mathcal{W}_2)$. Hence,

$$\begin{aligned} \dim(\mathcal{W}_1 + \mathcal{W}_2) &= k + p + q \\ &= (k + p) + (k + q) - q \\ &= \dim \mathcal{W}_1 + \dim \mathcal{W}_2 - \dim(\mathcal{W}_1 \cap \mathcal{W}_2). \end{aligned}$$

□

Remark 2.1.3. *If $\mathcal{V} = \mathcal{W}_1 + \mathcal{W}_2$, then we get that,*

$$\dim \mathcal{V} = \dim \mathcal{W}_1 + \dim \mathcal{W}_2 - \dim(\mathcal{W}_1 \cap \mathcal{W}_2).$$

If we have a direct sum, that is if $\mathcal{V} = \mathcal{W}_1 \oplus \mathcal{W}_2$, then the above formula becomes,

$$\dim \mathcal{V} = \dim \mathcal{W}_1 + \dim \mathcal{W}_2.$$

2.2. Homomorphic mappings between vector spaces: Linear Transformations

Just as we did in the case of groups, where we first studied the group structure and then the functions that preserve the structure, we now look at functions that preserve the vector space structure. Structure preserving maps are called homomorphism. In the case of linear algebra, homomorphisms between vector spaces are called linear transformation.

Definition 2.2.1 (Linear transformation)**.** *Let $\mathcal{V}/\mathbb{F}, \mathcal{W}/\mathbb{F}$ be two vector spaces over the **same** field $\mathbb{F}$. A function, $T : \mathcal{V} \to \mathcal{W}$ is called a linear transformation if:*

(1)

$$T(\underbrace{\boldsymbol{x} + \boldsymbol{y}}_{op.\ in\ \mathcal{V}}) = \underbrace{T(\boldsymbol{x}) + T(\boldsymbol{y})}_{op.\ in\ \mathcal{W}},$$

(2)

$$T(\underbrace{\beta \boldsymbol{x}}_{op.\ in\ \mathcal{V}}) = \underbrace{\beta T(\boldsymbol{x})}_{op.\ in\ \mathcal{W}},$$

for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}$.

Remark 2.2.1. *It is easily observable that $T(\boldsymbol{0}_{\mathcal{V}}) = \boldsymbol{0}_{\mathcal{W}}$ and that $T(\beta_1 \boldsymbol{x}_1 + \cdots + \beta_n \boldsymbol{x}_n) = \beta_1 T(\boldsymbol{x})_1 + \cdots + \beta_n T(\boldsymbol{x})_n$.*

Example 2.2.1. *The following are all linear transformations.*

(1) *Let* $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ *be an element of* $\mathbb{R}^2$. *Then* $T : \mathbb{R}^2 \to \mathbb{R}^2$ *defined by,* $T(\boldsymbol{x}) = \begin{pmatrix} 2x_1 + x_2 \\ x_2 \end{pmatrix}$ *is a linear transformation.*

(2) *Let* $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ *be an element of* $\mathbb{R}^2$. *For any angle* θ *we define* $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ *by,*

$$T(\boldsymbol{x}) = \begin{pmatrix} x_1 \cos\theta - x_2 \sin\theta \\ x_1 \sin\theta + x_2 \cos\theta \end{pmatrix},$$

is a linear transformation called the ***rotation*** *by* θ.

(3) *Let* $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ *be an element of* $\mathbb{R}^2$. *Define* $T : \mathbb{R}^2 \to \mathbb{R}^2$ *by* $T(\boldsymbol{x}) = \begin{pmatrix} x_1 \\ -x_2 \end{pmatrix}$. *Then,* T *is the linear transformation called the* ***reflection*** *about the* x *axis.*

(4) *Let* $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ *be an element of* $\mathbb{R}^2$. *Define* $T : \mathbb{R}^2 \to \mathbb{R}^2$ *by* $T(\boldsymbol{x}) = \begin{pmatrix} x_1 \\ 0 \end{pmatrix}$. *Then,* T *is the linear transformation called the* ***projection*** *on the* x *axis.*

As in the case of groups, any linear transformation induces two important subspaces.

Definition 2.2.2 (Kernel)**.** *Let* $\mathcal{V}, \mathcal{W}$ *be vector spaces over the field* $\mathbb{F}$ *and let* $T : \mathcal{V} \to \mathcal{W}$ *be a linear transformation. We define the Kernel or* ***null space*** *as the subset* $\ker T$ *of* $\mathcal{V}$ *as the inverse image* $T^{-1}(\mathbf{0}_{\mathcal{W}})$ *that is* $\ker T = T^{-1}(\mathbf{0}_{\mathcal{W}})$.

Definition 2.2.3 (Image)**.** *Let* $\mathcal{V}, \mathcal{W}$ *be vector spaces over the field* $\mathbb{F}$ *and let* $T : \mathcal{V} \to \mathcal{W}$ *be a linear transformation. We define the image or* ***range*** *as the subset* $\operatorname{Im} T$ *of* $\mathcal{W}$ *as the direct image* $T(\mathcal{V})$, *that is* $\operatorname{Im} T = T(\mathcal{V})$.

The following proposition is analogous to the case of groups.

Proposition 2.2.1. *The kernel and image of a linear transformation* $T : \mathcal{V} \to \mathcal{W}$ *are subspaces of* $\mathcal{V}$ *and* $\mathcal{W}$ *respectively.*

The next proposition gives us a method to determine a spanning set for the image of a linear transformation.

Proposition 2.2.2. *Let* $\mathcal{V}, \mathcal{W}$ *be vector spaces over the field* $\mathbb{F}$, *and let* $T : \mathcal{V} \to \mathcal{W}$ *be a linear transformation. If* $\mathcal{B}_{\mathcal{V}} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ *is a basis for* $\mathcal{V}$, *then*

$$\operatorname{Im} T = \operatorname{Span}\{T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \ldots, T(\boldsymbol{v}_n)\}.$$

Proof. Let $\boldsymbol{w}$ be in $\operatorname{Im} T$. Then there is a vector $\boldsymbol{x} \in \mathcal{V}$ such that $T(\boldsymbol{x}) = \boldsymbol{w}$. But, $\boldsymbol{x} = \beta_1 \boldsymbol{v}_1 + \cdots + \beta_n \boldsymbol{v}_n$ for some scalars $\beta_1, \beta_2, \ldots, \beta_n$ in $\mathbb{F}$. Thus,

$$\boldsymbol{w} = T(\beta_1 \boldsymbol{v}_1 + \cdots + \beta_n \boldsymbol{v}_n) = \beta_1 T(\boldsymbol{v}_1) + \cdots + \beta_n T(\boldsymbol{v}_n),$$

which means that $\boldsymbol{w}$ in $\operatorname{Span}\{T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \ldots, T(\boldsymbol{v}_n)\}$. Thus, $\operatorname{Im} T \subset \operatorname{Span}\{T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \ldots, T(\boldsymbol{v}_n)\}$.

Let $\boldsymbol{w}$ be in $\operatorname{Span}\{T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \ldots, T(\boldsymbol{v}_n)\}$ which means that,

$$\boldsymbol{w} = \beta_1 T(\boldsymbol{v}_1) + \cdots + \beta_n T(\boldsymbol{v}_n) = T(\beta_1 \boldsymbol{v}_1 + \cdots + \beta_n \boldsymbol{v}_n).$$

Let $\boldsymbol{x} = \beta_1 \boldsymbol{v}_1 + \cdots + \beta_n \boldsymbol{v}_n$. Then $\boldsymbol{x}$ is a vector in $\mathcal{V}$ such that $T(\boldsymbol{x}) = \boldsymbol{w}$ and hence $\operatorname{Span}\{T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \ldots, T(\boldsymbol{v}_n)\} \subset \operatorname{Im} T$. □

We measure the *size* of a subspace by its dimension. The kernel and image are so important that we attach special names to their dimensions.

Definition 2.2.4 (Nullity)**.** *The dimension of the kernel of a linear transformation T is called its nullity and is denoted by nullity(T).*

Definition 2.2.5 (Rank)**.** *The dimension of the image of a linear transformation T is called its rank and is denoted by rank(T).*

One of the most fundamental theorem of (finite dimensional) linear algebra is that of the relation between rank and nullity.

Theorem 2.2.1 (Dimension theorem)**.** Let $\mathcal{V}, \mathcal{W}$ be vector spaces over the field $\mathbb{F}$ and let $T : \mathcal{V} \to \mathcal{W}$ be a linear transformation. If $\mathcal{V}$ is finite dimensional then,

$$\dim \mathcal{V} = \dim(\ker T) + \dim(\operatorname{Im} T).$$

Proof. Since $\ker T$ is a subspace of $\mathcal{V}$, it is finite dimensional. Let $\mathcal{B}_{\ker T} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ be a basis for $\ker T$. Then $\mathcal{B}_{\ker T}$ can be extended to a basis for $\mathcal{V}$ given by,

$$\mathcal{B}_{\mathcal{V}} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\} \cup \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p\},$$

such that $k + p = n$.

Claim: $\{T(\boldsymbol{u}_1), T(\boldsymbol{u}_2), \ldots, T(\boldsymbol{u}_p)\}$ is a basis for $\operatorname{Im} T$. First note that

$$\{T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \ldots, T(\boldsymbol{v}_k)\} \cup \{T(\boldsymbol{u}_1), T(\boldsymbol{u}_2), \ldots, T(\boldsymbol{u}_p)\}$$

spans $\operatorname{Im} T$. Let $\boldsymbol{w}$ be any vector in $\operatorname{Im} T$. Then,

$$\boldsymbol{w} = \underbrace{\beta_1 T(\boldsymbol{v}_1) + \cdots + \beta_k T(\boldsymbol{v}_k)}_{\text{gives } \mathbf{0}_{\mathcal{W}}} + \gamma_1 T(\boldsymbol{u}_1) + \cdots + \gamma_p T(\boldsymbol{u}_p),$$

which means that $\operatorname{Im} T = \operatorname{Span}\{T(\boldsymbol{u}_1), T(\boldsymbol{u}_2), \ldots, T(\boldsymbol{u}_p)\}$ (we showed that $\operatorname{Im} T \subset \{T(\boldsymbol{u}_1), T(\boldsymbol{u}_2), \ldots, T(\boldsymbol{u}_p)\}$. The other inclusion is similar.) We need to show that $\{T(\boldsymbol{u}_1), T(\boldsymbol{u}_2), \ldots, T(\boldsymbol{u}_p)\}$ is linearly independent subset of $\mathcal{W}$. Let,

$$\beta_1 T(\boldsymbol{u}_1) + \cdots + \beta_p T(\boldsymbol{u}_p) = \mathbf{0}_{\mathcal{W}}.$$

This means that,

$$\beta_1 \boldsymbol{u}_1 + \cdots + \beta_p \boldsymbol{u}_p \in \ker T,$$

that is,

$$\beta_1 \boldsymbol{u}_1 + \cdots + \beta_p \boldsymbol{u}_p = \alpha_1 \boldsymbol{v}_1 + \cdots + \alpha_k \boldsymbol{v}_k,$$

which implies that,

$$\alpha_1 \boldsymbol{v}_1 + \cdots + \alpha_k \boldsymbol{v}_k + (-\beta)_1 \boldsymbol{u}_1 + \cdots + (-\beta)_p \boldsymbol{u}_p = \mathbf{0}_{\mathcal{V}}.$$

Since this a linear combination of basis vectors in $\mathcal{V}$, each β_j must be 0. Thus, $\{T(\boldsymbol{u}_1), T(\boldsymbol{u}_2), \ldots, T(\boldsymbol{u}_p)\}$ is a basis for $\mathcal{W}$. Thus, $\dim \operatorname{Im} T = p$. Now,

$$\begin{aligned} \dim \mathcal{V} &= k + p \\ &= \dim \ker T + \dim \operatorname{Im} T. \end{aligned}$$

□

For a linear transformation, being $1-1$ is intimately connected to being onto. The next proposition is analogous to one for groups.

Proposition 2.2.3. *A linear transformation $T : \mathcal{V} \to \mathcal{W}$ is injective if and only if* $\ker T = \{\mathbf{0}_\mathcal{V}\}$.

Proof. Let T be injective. Assume $\boldsymbol{v}_1$ in $\ker T$. Then, $T(\boldsymbol{v}_1) = \mathbf{0}_\mathcal{W}$. But since $\mathbf{0}_\mathcal{V}$ is also in $\ker T$, this implies $T(\boldsymbol{v}_1) = T(\mathbf{0}_\mathcal{V})$. However, since T is injective we get $\boldsymbol{v}_1 = \mathbf{0}_\mathcal{V}$. Thus $\ker T = \{\mathbf{0}_\mathcal{V}\}$.

Assume $\ker T = \{\mathbf{0}_\mathcal{V}\}$. Let $\boldsymbol{v}_1, \boldsymbol{v}_2$ be elements of $\mathcal{V}$ such that $T(\boldsymbol{v}_1) = T(\boldsymbol{v}_2)$. This means that $T(\boldsymbol{v}_1 - \boldsymbol{v}_2) = \mathbf{0}_\mathcal{W}$ and thus $(\boldsymbol{v}_1 - \boldsymbol{v}_2)$ is an element of $\ker T$. But only $\mathbf{0}_\mathcal{V}$ is an element of $\ker T$ and hence $\boldsymbol{v}_1 = \boldsymbol{v}_2$.

Thus T is $1-1$ if and only if $\dim \ker T = 0$. □

Proposition 2.2.4. *Let $\mathcal{V}, \mathcal{W}$ be finite dimensional vector spaces over field $\mathbb{F}$ such that $\dim \mathcal{V} = \dim \mathcal{W}$. Let $T : \mathcal{V} \to \mathcal{W}$ be a linear transformation. Then T is one-to-one if and only if T is onto.*

Proof. Let T be $1-1$. Then, by the above proposition, $\dim \ker T = 0$. To show that T is onto, we need to show that $\dim \operatorname{Im} T = \dim \mathcal{W}$ since that would mean that $\operatorname{Im} T = \mathcal{W}$ (because $\operatorname{Im} T \subset \mathcal{W}$). By the dimension theorem,

$$\begin{aligned}\dim \operatorname{Im} T &= \dim \mathcal{V} - \dim \ker T,\\ &= \dim \mathcal{V} - 0,\\ &= \dim \mathcal{W}.\end{aligned}$$

Let T be onto. Hence $\dim \operatorname{Im} T = \dim \mathcal{W}$. By the dimension theorem,

$$\begin{aligned}\dim \ker T &= \dim \mathcal{V} - \dim \operatorname{Im} T,\\ &= \dim \mathcal{V} - \dim \mathcal{W},\\ &= 0.\end{aligned}$$

Thus, T is $1-1$. □

One of the most important properties of linear transformatios is that they are completely determined by their action on a basis.

Proposition 2.2.5. *Let $\mathcal{V}, \mathcal{W}$ be vector spaces over field $\mathbb{F}$. Let $\mathcal{B}_\mathcal{V} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ be a basis for $\mathcal{V}$. If $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n$ are vectors in $\mathcal{W}$, then there is a unique linear transformation $T : \mathcal{V} \to \mathcal{W}$ such that $T(\boldsymbol{v}_i) = \boldsymbol{w}_i$ for each i.*

Proof. We will first show existence. Note that $\boldsymbol{v}_i = \delta_{i1}\boldsymbol{v}_1 + \cdots + \delta_{in}\boldsymbol{v}_n$ where $\delta_{ij} = 1$ if $i = j$ and 0 if $i \neq j$. For any $\boldsymbol{u}$ in $\mathcal{V}$, we can write $\boldsymbol{u} = \alpha_1\boldsymbol{v}_1 + \cdots + \alpha_n\boldsymbol{v}_n$. Let $T : \mathcal{V} \to \mathcal{W}$ be such that,

$$T(\boldsymbol{u}) = \alpha_1\boldsymbol{w}_1 + \cdots + \alpha_n\boldsymbol{w}_n.$$

Thus, we see that $T(\boldsymbol{v}_i) = \boldsymbol{w}_i$ for each i. Let $\boldsymbol{x} = \alpha_1\boldsymbol{v}_1 + \cdots + \alpha_n\boldsymbol{v}_n$ and $\boldsymbol{y} = \beta_1\boldsymbol{v}_1 + \cdots + \beta_n\boldsymbol{v}_n$. Then, $\boldsymbol{x} + \boldsymbol{y} = (\alpha + \beta)_1\boldsymbol{v}_1 + \cdots + (\alpha + \beta)_n\boldsymbol{v}_n$. Thus,

$$T(\boldsymbol{x}+\boldsymbol{y}) = (\alpha + \beta)_1\boldsymbol{w}_1 + \cdots + (\alpha + \beta)_n\boldsymbol{w}_n = \alpha_1\boldsymbol{w}_1 + \cdots + \alpha_n\boldsymbol{w}_n + \beta_1\boldsymbol{w}_1 + \cdots + \beta_n\boldsymbol{w}_n = T(\boldsymbol{x}) + T(\boldsymbol{y}).$$

Similarly we can show $T(\beta\boldsymbol{x}) = \beta T(\boldsymbol{x})$ for any $\beta \in \mathbb{F}$. Thus T is a linear transformation.

Let $U : \mathcal{V} \to \mathcal{W}$ be another linear transformation such that $U(\boldsymbol{v}_i) = \boldsymbol{w}_i$ for each i. Then for any $\boldsymbol{x} = \alpha_1\boldsymbol{v}_1 + \cdots + \alpha_n\boldsymbol{v}_n$,

$$U(\boldsymbol{x}) = \alpha_1 U(\boldsymbol{v}_1) + \cdots + \alpha_n U(\boldsymbol{v}_n) = \alpha_1\boldsymbol{w}_1 + \cdots + \alpha_n\boldsymbol{w}_n = T(\boldsymbol{x}).$$

□

2.3. Matrices

In this section we will show that there is a very useful way to look at linear transformations between finite dimensional vector spaces. In particular, we will show that such linear transformation are matrices. First we define what an ordered basis means.

Definition 2.3.1 (Ordered Basis)**.** *Let $\mathcal{V}/\mathbb{F}$ be a finite dimensional vector space over the field $\mathbb{F}$. An ordered basis for $\mathcal{V}$ is a a basis endowed with a specific order, that is an ordered basis for $\mathcal{V}$ is a finite sequence of linearly independent elements of $\mathcal{V}$ that spans $\mathcal{V}$.*

Example 2.3.1. *If $\mathcal{B}_\mathcal{V} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ is a basis for $\mathcal{V}$, then there are $n!$* different *ordered bases of $\mathcal{V}$ from the given unordered basis $\mathcal{B}_\mathcal{V}$. We denote an ordered basis as $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$. Thus $(\boldsymbol{v}_2, \boldsymbol{v}_3, \ldots, \boldsymbol{v}_n, \boldsymbol{v}_1)$ is an ordered basis for $\mathcal{V}$ where the **first** basis is $\boldsymbol{v}_2$.*

Example 2.3.2. *For any field $\mathbb{F}$, $\mathbb{F}^n$ is a vector space of $\mathbb{F}$. What is a basis for $\mathbb{F}^n$? Note that any α in $\mathbb{F}$ can be written as $\alpha \cdot 1$, where 1 is the identity element of $\mathbb{F}$. Thus any $\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}$ can be written as, $\boldsymbol{\alpha} = \alpha_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \alpha_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$. The **vectors** $\boldsymbol{e}_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ 0 \end{pmatrix}$ with 1 in the i^{th} slots are called **standard** basis vectors for $\mathbb{F}^n$. Each field $\mathbb{F}^n$ is equipped with its standard basis vectors. Thus $\mathcal{B}_{\mathbb{F}^n} = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ is the standard unordered basis for $\mathbb{F}^n$. This is an important observation. To reiterate, any $\boldsymbol{x} \in \mathbb{F}^n$ is represented by a column vector*

$$\boldsymbol{x} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix},$$

if and only if

$$\boldsymbol{x} = \alpha_1\boldsymbol{e}_1 + \cdots + \alpha_n\boldsymbol{e}_n.$$

Let us explore the structure of linear transformation between fields (as vector spaces) in the following examples. For any $\mathbb{F}^n$ we write $(\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n)$ as the standard ordered basis for $\mathbb{F}^n/\mathbb{F}$.

Example 2.3.3. *Let $T : \mathbb{F} \to \mathbb{F}$ be a linear transformation. How can T be described? Any $\boldsymbol{\alpha} \in \mathbb{F}$ can be written as $\alpha \cdot \mathbf{1}$. Thus $T(\boldsymbol{\alpha}) = \alpha T(\mathbf{1})$. Note that $\mathbf{1} = \boldsymbol{e}_1$ is the standard basis for $\mathbb{F}$. Thus T can be described as $T(\boldsymbol{x}) = ax$ for any $\boldsymbol{x} \in \mathbb{F}$ where $a = T(\boldsymbol{e}_1)$.*

Example 2.3.4. *Let $T : \mathbb{F}^2 \to \mathbb{F}$ be a linear transformation. How can T be described? Any $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ can be written as $\boldsymbol{x} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2$. Thus T can be described as $T(\boldsymbol{x}) = a_{11}x_1 + a_{12}x_2$, where $a_{11} = T(\boldsymbol{e}_1)$ and $a_{12} = T(\boldsymbol{e}_2)$.*

Example 2.3.5. *We might be noticing a pattern. Let $T : \mathbb{F}^n \to \mathbb{F}$ be a linear transformation. How can T be described? Any $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{F}^n$ can be written as*

$$\boldsymbol{x} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + x_3\boldsymbol{e}_3 + \cdots + x_n\boldsymbol{e}_n.$$

Thus,

$$T(\boldsymbol{x}) = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n,$$

where $a_{1j} = T(\boldsymbol{e}_j)$.

Example 2.3.6. *Let $T : \mathbb{F}^2 \to \mathbb{F}^2$ be linear transformation. How can T be described? We can think of T as $\begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$ where $T_i : \mathbb{F}^2 \to F$ is a linear transformation. We have seen how these behave, that is for any $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$,*

$$T_i(\boldsymbol{x}) = a_{i1}x_1 + a_{i2}x_2.$$

Thus,

$$T(\boldsymbol{x}) = \begin{pmatrix} T_1(\boldsymbol{x}) \\ T_2(\boldsymbol{x}) \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{pmatrix}.$$

Example 2.3.7. *Let $T : \mathbb{F}^n \to \mathbb{F}^m$ be linear transformation. How can T be described? As seen above, we can think of T as $\begin{pmatrix} T_1 \\ T_2 \\ \vdots \\ T_m \end{pmatrix}$ where each $T_i : \mathbb{F}^n \to F$ is a linear transformation for $1 \leq i \leq m$. Let $\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. For any i,*

$$T_i(\boldsymbol{x}) = a_{i1}x_1 + \cdots + a_{in}x_n.$$

Thus,

$$T(\boldsymbol{x}) = \begin{pmatrix} T_1(\boldsymbol{x}) \\ T_2(\boldsymbol{x}) \\ \vdots \\ T_m(\boldsymbol{x}) \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ a_{21}x_1 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix}.$$

Now we are ready to describe a matrix.

Definition 2.3.2 (Matrix)**.** *A matrix M is a table of size $m \times n$ where m is the number of rows and n is the number of columns. The elements of this table are elements of a field $\mathbb{F}$. We denote by $M[:, j]$, the j^{th} column of the matrix M and by $M[i, :]$ the i^{th} row of the matrix. The set of all matrices of size $m \times n$ with elements from field $\mathbb{F}$ is denoted by $\mathcal{M}_{m\times n}(\mathbb{F})$.*

By the notation $[a_{ij}]_{m\times n}$ we mean a matrix of size $m \times n$ whose elements are given by a_{ij} for $1 \leq i \leq m$ and $1 \leq j \leq n$.

Looking at our examples above, we see that any linear transformation $T : \mathbb{F}^n \to \mathbb{F}^m$ can be described by a matrix. For example the matrix of T as given in 2.3.7 is given by,

$$M_T = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

To see the precise relation, let us consider the action of T on a standard basis vector $\boldsymbol{e}_j$. From 2.3.7, we see that,

$$T_i(\boldsymbol{e}_j) = a_{i1}0 + \cdots + a_{ij}1 + \cdots + a_{in}0 = a_{ij}.$$

Thus,

$$T(\boldsymbol{e}_j) = \begin{pmatrix} T_1(\boldsymbol{e}_j) \\ T_2(\boldsymbol{e}_j) \\ \vdots \\ T_m(\boldsymbol{e}_j) \end{pmatrix} = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix},$$

which is the j^{th} column of the matrix M_T.

Definition 2.3.3. *Given a linear transformation $T : \mathbb{F}^n \to \mathbb{F}^m$, the matrix of T, denoted by M_T is a table of size $m \times n$, with elements from field $\mathbb{F}$ such that,*

$$M_T[:, j] = T(\boldsymbol{e}_j),$$

where $\boldsymbol{e}_j$ is the j^{th} standard basis of the oderered basis, $(\{e_1, e_2, \ldots, e_n\})$ for $\mathbb{F}^n$.

We can impart a group structure on the set of matrices.

Proposition 2.3.1. *Let $\mathbb{F}$ be a field. The set $\mathcal{M}_{m\times n}(\mathbb{F})$ is an additive Abelian group, where addition $(+)$ is given by,*

$$[a_{ij}]_{m\times n} + [b_{ij}]_{m\times n} = [a_{ij} + b_{ij}]_{m\times n},$$

with the identity element $[0]_{m\times n}$ consiting of 0 in each row and column.

Thus, we can regard matrices as a separate entity without any relation to the underlying linear transformation. However, it would be remiss to not discuss this deep connection.

Proposition 2.3.2. *Let $\mathcal{V}, \mathcal{W}$ be vector spaces over the field $\mathbb{F}$. We denote by $\hom(\mathcal{V}, \mathcal{W})$ as the set of all linear transformations from $\mathcal{V}$ to $\mathcal{W}$, that is,*

$$\hom(\mathcal{V}, \mathcal{W}) = \{T \ : \mathcal{V} \to \mathcal{W} \ : \ T \text{ is a linear transformation}\}.$$

Then, $\left(\hom(\mathcal{V}, \mathcal{W}), +, \mathbf{0}_{\hom(\mathcal{V},\mathcal{W})}\right)$ is a vector space, where

$$(T_1 + T_2)(\boldsymbol{v}) := T_1(\boldsymbol{v}) + T_2(\boldsymbol{v}),$$

for any $T_1, T_2 \in \hom(\mathcal{V}, \mathcal{W})$ and any $\boldsymbol{v} \in \mathcal{V}$; and

$$(\beta T_1)(\boldsymbol{v}) := \beta(T_1(\boldsymbol{v})),$$

for any $\beta \in \mathbb{F}$.

We will not prove the proposition.

An immediate consequence of this proposition is that $\hom(\mathbb{F}^n, \mathbb{F}^m)$ is a vector space. In fact the group $\mathcal{M}_{m\times n}(\mathbb{F})$ can be given a vector space structure.

Proposition 2.3.3. *$\mathcal{M}_{m\times n}(\mathbb{F})$ is a vector space over the field $\mathbb{F}$, with addition defined by,*

$$[a_{ij}]_{m\times n} + [b_{ij}]_{m\times n} = [a_{ij} + b_{ij}]_{m\times n},$$

and scalar multiplication by,

$$\beta[a_{ij}]_{m\times n} = [\beta a_{ij}]_{m\times n}.$$

What is the relation between $\hom(\mathbb{F}^n, \mathbb{F}^m)$ and $\mathcal{M}_{m\times n}(\mathbb{F})$? Intiuition suggests that these two are equivalent notions.

Given two linear transformations $T_1 \ : \mathbb{F}^n \to \mathbb{F}^m$ and $T_2 \ : \mathbb{F}^n \to \mathbb{F}^m$, what is the matrix of $T_1 + T_2$? From the definition,

$$M_{(T_1+T_2)}[:, j] = (T_1 + T_2)(\boldsymbol{e}_j) = T_1(\boldsymbol{e}_j) + T_2(\boldsymbol{e}_j) = M_{T_1} + M_{T_2}.$$

In other words, if $M_{T_1} = [a_{ij}]_{m\times n}$ and $M_{T_2} = [b_{ij}]_{m\times n}$, then $M_{(T_1+T_2)} = [a_{ij} + b_{ij}]_{m\times n}$. Similarly, we can easily see that $M_{(\beta T)}[:, j] = (\beta T)(\boldsymbol{e}_j) = \beta T(\boldsymbol{e}_j)$. Thus if $M_T = [a_{ij}]_{m\times n}$, then $M_{(\beta T)} = [\beta a_{ij}]_{m\times n}$.

Thus addition of two linear transformations amounts to addition of their matrices and similarly for scalar multiplication. These two vector spaces seem to have the same algebraic structure. To make precise this notion we need to show that they are isomorphic.

Recall that two algebraic structure are equivalent if they are isomorphic. In the context of vector space, we define isomorphism as,

Definition 2.3.4. *Two vector spaces $\mathcal{V}$ and $\mathcal{W}$ are isomorphic if there is a bijective linear transformation mapping $\mathcal{V}$ onto $\mathcal{W}$.*

Before we show that $\hom(\mathbb{F}^n, \mathbb{F}^m)$ and $\mathcal{M}_{m\times n}(\mathbb{F})$ are isomorphic, we will need to understand what it means to mulitply a matrix with a vector.

Note that if $T \in \hom(\mathbb{F}^n, \mathbb{F}^m)$, then $T(\boldsymbol{x}) = T(\alpha_1 \boldsymbol{e}_1 + \cdots + \alpha_n \boldsymbol{e}_n) = \alpha_1 T(\boldsymbol{e}_1) + \cdots + \alpha_n T(\boldsymbol{e}_n)$. Thus,

$$T(\boldsymbol{x}) = \sum_{i=1}^{n} \alpha_i M_T[:, i].$$

This gives us a way to define a product of a matrix with a vector.

Definition 2.3.5. *Let A be any $m \times n$ matrix. For any $\boldsymbol{x} \in \mathbb{F}^n$ such that*

$$\boldsymbol{x} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix},$$

let us define

$$A\boldsymbol{x} := \sum_{i=1}^{n} \alpha_i M[:, i],$$

*as the **matrix-vector** product.*

Proposition 2.3.4. *Let A be a $m \times n$ matrix and let*

$$\boldsymbol{x} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix},$$

be a vector in $\mathbb{F}^n$. Then $L_A(\boldsymbol{x}) := A\boldsymbol{x}$ is a linear transformation from $\mathbb{F}^n$ to $\mathbb{F}^m$ and is called the linear transformation induced by A.

Proof. Let $\boldsymbol{x} = \alpha_1 \boldsymbol{e}_1 + \cdots + \alpha_n \boldsymbol{e}_n$ and $\boldsymbol{y} = \beta_1 \boldsymbol{e}_1 + \cdots + \beta_n \boldsymbol{e}_n$ be two vectors in $\mathbb{F}^n$. Then, $\boldsymbol{x} + \boldsymbol{y} = \sum_{i=1}^{n} (\alpha_i + \beta_i) \boldsymbol{e}_i$ and thus,

$$(\boldsymbol{x} + \boldsymbol{y}) = \begin{pmatrix} (\alpha_1 + \beta_1) \\ (\alpha_2 + \beta_2) \\ \vdots \\ (\alpha_n + \beta_n) \end{pmatrix}$$

$$\begin{aligned} L_A(\boldsymbol{x} + \boldsymbol{y}) &= A(\boldsymbol{x} + \boldsymbol{y}) \\ &= \sum_{j=1}^{n} (\alpha_j + \beta_j) A[:, j] \\ &= \sum_{j=1}^{n} \alpha_j A[:, j] + \sum_{j=1}^{n} \beta_j A[:, j] \\ &= A\boldsymbol{x} + A\boldsymbol{y} \\ &= L_A(\boldsymbol{x}) + L_A(\boldsymbol{y}), \end{aligned}$$

where the 3^{rd} equalilty is due to the fact that $A[:, j]$'s are vectors in $\mathbb{F}^m$. Similarly we can show scalar multiplication. □

Theorem 2.3.1. The vector spaces $\hom(\mathbb{F}^n, \mathbb{F}^m)$ and $\mathcal{M}_{m\times n}(\mathbb{F})$ are isomorphic. Hence, any linear transformation from $\mathbb{F}^n$ to $\mathbb{F}^m$ is equivalent to a matrix of size $m \times n$.

Proof. WE need to find a linear transformation $U : \hom(\mathbb{F}^n, \mathbb{F}^m) \to \mathcal{M}_{m\times n}(\mathbb{F})$ such that U is bijective. Let us define U to be,

$$U(T) = M_T.$$

Clearly U is a linear transformation. To show that U is bijective we must find a linear transformation $S : \mathcal{M}_{m\times n}(\mathbb{F}) \to \hom(\mathbb{F}^n, \mathbb{F}^m)$ such that,

(1) $(S \circ U)(T) = T$, for any T in $\hom(\mathbb{F}^n, \mathbb{F}^m)$ and,

(2) $(U \circ S)(A) = A$, for any A in $\mathcal{M}_{m\times n}(\mathbb{F})$.

Let $S(A) := L_A$. We will show the first one since the second is analogous.

$$\begin{aligned}(S \circ U)(T) &= S(U(T)) \\ &= S(M_T) \\ &= L_{M_T}\end{aligned}$$

To show that $T = L_{M_T}$, we need to just check if they agree on the basis vectors $(\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n)$. For any $1 \le i \le n$,

$$\begin{aligned}L_{M_T}(\boldsymbol{e}_i) &= M_T \boldsymbol{e}_i \\ &= \sum_{j=1}^{n} \delta_{ij} M_T[:, j] \\ &= M_T[:, i] \\ &= T(\boldsymbol{e}_i).\end{aligned}$$

Here $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. □

Let us now look at the matrix of a composition. Let $T_1 : \mathbb{F}^n \to \mathbb{F}^p$ and $T_2 : \mathbb{F}^p \to \mathbb{F}^m$. Then the composite $T_2 \circ T_1 : \mathbb{F}^n \to \mathbb{F}^m$ is easily seen to be a linear transformation. Note that $T_1 \circ T_2$ may fail to be a function. What is the matrix of $T_2 \circ T_1$? Let $(\boldsymbol{e}^n{}_1, \boldsymbol{e}^n{}_2, \ldots, \boldsymbol{e}^n{}_n)$ be the standard ordered basis of $\mathbb{F}^n$, $(\boldsymbol{e}^p{}_1, \boldsymbol{e}^p{}_2, \ldots, \boldsymbol{e}^p{}_p)$ be the standard ordered basis for $\mathbb{F}^p$ and $(\boldsymbol{e}^m{}_1, \boldsymbol{e}^m{}_2, \ldots, \boldsymbol{e}^m{}_m)$ be the standard ordered basis for $\mathbb{F}^m$. Let $M_{T_1} = [a_{ij}]_{p\times n}$ and let $M_{T_2} = [b_{ij}]_{m\times p}$.

By definition, $M_{T_2\circ T_1}[:,j] = (T_2 \circ T_1)(\boldsymbol{e}_j^n)$. Thus,

$$\begin{aligned}
M_{T_2\circ T_1}[:,j] &= (T_2 \circ T_1)(\boldsymbol{e}_j^n). \\
&= T_2(T_1(\boldsymbol{e}_j^n)) \\
&= T_2(M_{T_1}[:,j]) \\
&= T_2(\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{pj} \end{pmatrix}) \\
&= T_2(\sum_{k=1}^{p} a_{kj}\boldsymbol{e}_k^p) \\
&= \sum_{k=1}^{p} a_{kj} T_2(\boldsymbol{e}_k^p) \\
&= \sum_{k=1}^{p} a_{kj} M_{T_2}[:,k] \\
&= \sum_{k=1}^{p} a_{kj} \begin{pmatrix} b_{1k} \\ \vdots \\ b_{mk} \end{pmatrix} \\
&= \sum_{k=1}^{p} a_{kj} (\sum_{i=1}^{m} b_{ik}\boldsymbol{e}_i^m) \\
&= \sum_{i=1}^{m} (\sum_{k=1}^{p} b_{ik}a_{kj})\boldsymbol{e}_i^m.
\end{aligned}$$

Hence, if $M_{T_2\circ T_1} = [c_{ij}]_{m\times n}$, then,

$$c_{ij} = \sum_{k=1}^{p} b_{ik}a_{kj}.$$

We can use this to define a matrix multiplication operation.

Definition 2.3.6. *Let $A = [a_{ij}]_{m\times p}$ be a matrix of size $m \times p$ and let $B = [b_{ij}]_{p\times n}$, then we define a matrix $C = [c_{ij}]_{m\times n}$ of size $m \times n$ as the **product** of the matrix A and B denoted by $C = AB$ whose elements are given by,*

$$c_{ij} = \sum_{k=1}^{p} a_{ik}b_{kj}.$$

Now that we have seen that matrices can be identified with linear transformations, we will abstract this notion to linear transformations between general finite dimensional vector spaces. The crucial idea that will help us in this regard is the notion of a co-ordinate. A co-ordinate map makes a finite dimensional vector space, of dimension n, look like $\mathbb{F}^n$.

Definition 2.3.7 (Co-ordinate)**.** *Let $\mathcal{V}/\mathbb{F}$ be a finite dimensional vector space over $\mathbb{F}$ and let $\mathcal{B}_{\mathcal{V}} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ be an ordered basis for $\mathcal{V}$. For any $\boldsymbol{x} \in \mathcal{V}$, we can write $\boldsymbol{x}$ uniquely as,*

$$\boldsymbol{x} = \beta_1 \boldsymbol{v}_1 + \cdots + \beta_n \boldsymbol{v}_n.$$

The co-ordinate of $\boldsymbol{x}$ w.r.t $\mathcal{B}_{\boldsymbol{v}}$, denoted by $[\boldsymbol{x}]_{\mathcal{B}_{\boldsymbol{v}}}$, is an element of $\mathbb{F}^n$ given by,

$$[\boldsymbol{x}]_{\mathcal{B}_{\boldsymbol{v}}} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}.$$

Definition 2.3.8 (Co-ordinate map)**.** *Let $\mathcal{V}/\mathbb{F}$ be a finite dimensional vector space over $\mathbb{F}$ and let $\mathcal{B}_{\mathcal{V}} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ be an ordered basis for $\mathcal{V}$. For any $\boldsymbol{x} \in \mathcal{V}$, we can write $\boldsymbol{x}$ uniquely as,*

$$\boldsymbol{x} = \beta_1 \boldsymbol{v}_1 + \cdots + \beta_n \boldsymbol{v}_n.$$

The co-ordinate map $\phi_{\mathcal{B}_{\mathcal{V}}} : \mathcal{V} \to \mathbb{F}^n$, is given by:

$$\varphi_{\mathcal{B}_{\mathcal{V}}}(\boldsymbol{x}) = [\boldsymbol{x}]_{\mathcal{B}_{\boldsymbol{v}}} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}.$$

Remark 2.3.1. *Easy to see that a co-ordinate map is a linear transformation from $\mathcal{V}$ to $\mathbb{F}^n$.*

Remark 2.3.2. *Note that for any $\boldsymbol{v}_j$ in $\mathcal{B}_{\mathcal{V}}$, $\phi_{\mathcal{B}_{\boldsymbol{v}}}(\boldsymbol{v}_j) = [\boldsymbol{v}_j]_{\mathcal{B}_{\boldsymbol{v}}} = \boldsymbol{e}_j$ where $\boldsymbol{e}_j$ is the j^{th} standard basis for $\mathbb{F}^n$.*

Example 2.3.8. *Let $\mathcal{V} = P_2(\mathbb{R})$ be the vector space of polynomials in $\mathbb{R}$ of degree less than or equal to 2. It is easy to check that $\mathcal{B}_{\mathcal{V}} = \left\{1, x, x^2\right\}$ is a basis for $\mathcal{V}$. Let $f \in \mathcal{V}$ be given by $4 + 6x - 7x^2$, then $f = 4(1) + 6(x) + (-7)x^2$ and hence $[f]_{\mathcal{B}_{\boldsymbol{v}}} = \begin{pmatrix} 4 \\ 6 \\ -7 \end{pmatrix}$.*

We said that an ordered basis induces a co-ordinate map that makes a n dimensional vector space look like $\mathbb{F}^n$. We make this precise.

Proposition 2.3.5. *Let $\mathcal{V}$ be a finite dimensional vector space with dimension n. Then $\mathcal{V}$ is isomorphic to $\mathbb{F}^n$.*

Proof. Let $\mathcal{B}_{\mathcal{V}} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ be an ordered basis for $\mathcal{V}$. We will show that the co-ordinate map $\varphi_{\mathcal{B}\mathcal{V}}$ is an isomorphism from $\mathcal{V}$ onto $\mathbb{F}^n$. First, observe that $\varphi_{\mathcal{B}_{\mathcal{V}}}$ is a linear transformation. Define $\Phi_{\mathcal{B}_{\mathcal{V}}} : \mathbb{F}^n \to \mathcal{V}$ as follows:

$$\Phi_{\mathcal{B}_{\mathcal{V}}}\left(\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}\right) = \beta_1 \boldsymbol{v}_1 + \cdots + \beta_n \boldsymbol{v}_n.$$

Clearly $\Phi_{\mathcal{B}_\mathcal{V}}$ is a linear transformation. It is easy to see that:

(1) $(\varphi_{\mathcal{B}_\mathcal{V}} \circ \Phi_{\mathcal{B}_\mathcal{V}}) \left(\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \right) = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$ for any $a_i \in \mathbb{F}$ and,

(2) $(\Phi_{\mathcal{B}_\mathcal{V}} \circ \varphi_{\mathcal{B}_\mathcal{V}})(\boldsymbol{u}) = \boldsymbol{u}$ for any $\boldsymbol{u} \in \mathcal{V}$.

Thus, $\Phi_{\mathcal{B}_\mathcal{V}} = \varphi^{-1}_{\mathcal{B}_\mathcal{V}}$. □

The above Proposition is extremely useful. We will now see that given any linear transformation between abstract finite dimensional vector spaces of dimension n, m we can find the **equivalent** linear transformation between $\mathbb{F}^n, \mathbb{F}^m$.

Definition 2.3.9 (Induced linear transformation)**.** *Let $\mathcal{V}, \mathcal{W}$ be finite dimensional vector spaces of dimension n and m respectively. Let $\mathcal{B}_\mathcal{V} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ and $\mathcal{B}_\mathcal{W} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m)$ be ordered basis for $\mathcal{V}$ and $\mathcal{W}$ respectively. If $T : \mathcal{V} \to \mathcal{W}$ is any linear transformation, we call $\check{T} : \mathbb{F}^n \to \mathbb{F}^m$ the induced linear transformation of T w.r.t. bases $\mathcal{B}_\mathcal{V}$ and $\mathcal{B}_\mathcal{W}$ given by:*

$$\check{T} = \varphi_{\mathcal{B}_\mathcal{W}} \circ T \circ \varphi^{-1}_{\mathcal{B}_\mathcal{V}}.$$

This makes it easier to define the matrix of a linear transformation between abstract finite dimensional vector spaces.

Definition 2.3.10 (Matrix)**.** *Let $\mathcal{V}, \mathcal{W}$ be finite dimensional vector spaces of dimension n and m respectively. Let $\mathcal{B}_\mathcal{V} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ and $\mathcal{B}_\mathcal{W} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m)$ be ordered basis for $\mathcal{V}$ and $\mathcal{W}$ respectively. If $T : \mathcal{V} \to \mathcal{W}$ is any linear transformation, we denote by $[T]^{\mathcal{B}_\mathcal{W}}_{\mathcal{B}_\mathcal{V}}$ the matrix of T, given by,*

$$[T]^{\mathcal{B}_\mathcal{W}}_{\mathcal{B}_\mathcal{V}} := M_{\check{T}}.$$

Let $A = [a_{ij}]_{m \times n}$ be the matrix of T i.e. $A = [T]^{\mathcal{B}_\mathcal{W}}_{\mathcal{B}_\mathcal{V}}$. What is the j^{th} column of A, $A[:, j]$? By definition, $A[:, j] = M_{\check{T}}[:, j] = \check{T}(\boldsymbol{e}^n_j)$, where $\mathcal{B}_{\mathbb{F}^n} = (\boldsymbol{e}^n{}_1, \boldsymbol{e}^n{}_2, \ldots, \boldsymbol{e}^n{}_n)$ is the standard ordered basis for $\mathbb{F}^n$. Thus,

$$\begin{aligned} A[:, j] &= \check{T}(\boldsymbol{e}^n_j) \\ &= (\varphi_{\mathcal{B}_\mathcal{W}} \circ T \circ \varphi^{-1}_{\mathcal{B}_\mathcal{V}})(\boldsymbol{e}^n_j) \\ &= \varphi_{\mathcal{B}_\mathcal{W}} \left(T \left(\varphi^{-1}_{\mathcal{B}_\mathcal{V}}(\boldsymbol{e}^n_j) \right) \right) \\ &= \varphi_{\mathcal{B}_\mathcal{W}} \left(T(\boldsymbol{v}_j) \right) \\ &= [T(\boldsymbol{v}_j)]_{\mathcal{B}_\mathcal{W}} \end{aligned}$$

Thus, $T(\boldsymbol{v}_j) = \sum_{i=1}^{n} a_{ij} \boldsymbol{w}_i$.

Let $\boldsymbol{u}$ be a vector in $\mathcal{V}$ such that $\boldsymbol{u} = \alpha_1 \boldsymbol{v}_1 + \cdots + \alpha_n \boldsymbol{v}_n$. What is $T(\boldsymbol{u})$ in terms of its matrix $A = [a_{ij}]_{m \times n} = [T]^{\mathcal{B}_\mathcal{W}}_{\mathcal{B}_\mathcal{V}}$? Note that,

$$T(\boldsymbol{u}) = \sum_{i=1}^{m} \beta_i \boldsymbol{w}_i.$$

By definition,

$$\varphi_{\mathcal{B}_\mathcal{W}} \circ T = \check{T} \circ \varphi_{\mathcal{B}_\mathcal{W}}.$$

Thus,

$$\begin{aligned}(\varphi_{\mathcal{B}_\mathcal{W}} \circ T)(\boldsymbol{u}) &= \check{T}\left(\varphi_{\mathcal{B}_\mathcal{V}}(\boldsymbol{u})\right) \\ &= \check{T}\left([\boldsymbol{u}]_{\mathcal{B}_\mathcal{V}}\right) \\ &= M_{\check{T}}\,[\boldsymbol{u}]_{\mathcal{B}_\mathcal{V}} \\ &= A\,[\boldsymbol{u}]_{\mathcal{B}_\mathcal{V}}\,.\end{aligned}$$

Hence if $T(\boldsymbol{u}) = \sum_{i=1}^{m} \beta_i \boldsymbol{w}_i$, then each β_i is given by,

$$\beta_i = \sum_{l=1}^{n} \alpha_l a_{il}.$$

To re-iterate if $T \in \hom(\mathcal{V}, \mathcal{W})$ where $\dim \mathcal{V} = n$ and $\dim \mathcal{W} = m$, and if A is the matrix of T, i.e. $A = [T]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{W}}$, then for any vector $\boldsymbol{u} \in \mathcal{V}$,

$$[T(\boldsymbol{u})]_{\mathcal{B}_\mathcal{W}} = A\,[\boldsymbol{u}]_{\mathcal{B}_\mathcal{V}}\,.$$

Note that for any $\boldsymbol{x} \in Fn$, $[\boldsymbol{x}]_{\mathcal{B}_{\mathbb{F}^n}} = \boldsymbol{x}$, when $\mathcal{B}_{\mathbb{F}^n}$ is the standard ordered basis for $\mathbb{F}^n$. Now let us generalize the definition of a linear transformation induced by a matrix.

Definition 2.3.11. *Let $A = [a_{ij}]_{m\times n}$ be a matrix of size $m \times n$. The linear transformation L_A induced by A is given as,*

$$L_A(\boldsymbol{x}) := A\,[\boldsymbol{x}]_{\mathcal{B}_\boldsymbol{v}}\,,$$

for any $\boldsymbol{x}$ in $\mathbb{F}^n$ and any ordered basis $\mathcal{B}_\boldsymbol{v}$ of $\mathbb{F}^n$.

Example 2.3.9. *For any linear transformation $T : \mathcal{V} \to \mathcal{W}$, if $A = [T]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{W}}$, then $L_A = \check{T}$. This is easy to see. Let $\boldsymbol{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$.*

$$\begin{aligned}L_A(\boldsymbol{a}) &= A \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \\ &= \sum_{j=1}^{n} a_j A[:j] \\ &= \sum_{j=1}^{n} a_j \check{T}(\boldsymbol{e}_j) \\ &= \check{T}(\sum_{j=1}^{n} a_j \boldsymbol{e}_j) \\ &= \check{T}(\boldsymbol{a}).\end{aligned}$$

The next theorem states the equivalency of the vector space of matrices and linear transformation in finite dimensional vector spaces. The proof is similar to the particular case of mappings between $\mathbb{F}^n$ and $\mathbb{F}^m$.

Theorem 2.3.2. Let $\mathcal{V}, \mathcal{W}$ be finite dimensional vector spaces (over a field $\mathbb{F}$) of dimensions n, m respectively. Then $\hom(\mathcal{V}, \mathcal{W})$ and $\mathcal{M}_{m\times n}(\mathbb{F})$ are isomorphic.

Next, we define the matrix of addition of two linear transformation and composition of two linear transformations.

Theorem 2.3.3. Let $\mathcal{V}, \mathcal{W}$ be finite dimensional vector spaces over the field $\mathbb{F}$ and with dimensions n, m respectively. Let $\mathcal{B}_\mathcal{V}$ and $\mathcal{B}_\mathcal{W}$ be ordered bases for $\mathcal{V}$ and $\mathcal{W}$ respectively. Let $T, U : \mathcal{V} \to \mathcal{W}$ be a linear transformation. Then

(1)
$$[T+U]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{W}} = [T]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{W}} + [U]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{W}}.$$

(2)
$$[\beta T]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{W}} = \beta [T]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{W}}.$$

Proof. We will show for (1). Let $A = [T]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{W}}$, $B = [U]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{W}}$ and let $C = [T+U]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{W}}$. Let $\mathcal{B}_\mathcal{V} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$. Then,

$$\begin{aligned} C[:, j] &= \varphi_{\mathcal{B}_\mathcal{W}}(T+U)(\boldsymbol{v}_j) \\ &= \varphi_{\mathcal{B}_\mathcal{W}}(T(\boldsymbol{v}_j) + U(\boldsymbol{v}_j)) \\ &= \varphi_{\mathcal{B}_\mathcal{W}} T(\boldsymbol{v}_j) + \varphi_{\mathcal{B}_\mathcal{W}} U(\boldsymbol{v}_j) \\ &= A[:, j] + B[:, j]. \end{aligned}$$

□

Theorem 2.3.4. Let $\mathcal{V}, \mathcal{W}, \mathcal{Z}$ be finite dimensional vector spaces over field $\mathbb{F}$ and of dimensions n, p, m respectively. Let $\mathcal{B}_\mathcal{V}, \mathcal{B}_\mathcal{W}$ and $\mathcal{B}_\mathcal{Z}$ be ordered bases for $\mathcal{V}, \mathcal{W}$ and $\mathcal{Z}$ respectively. If $T : \mathcal{V} \to \mathcal{W}$ and $U : \mathcal{W} \to \mathcal{Z}$ are linear transformations, then
$$[(U \circ T)]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{Z}} = [U]_{\mathcal{B}_\mathcal{W}}^{\mathcal{B}_\mathcal{Z}} [T]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{W}}.$$

Proof. Let $A = [T]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{W}}$, $B = [U]_{\mathcal{B}_\mathcal{W}}^{\mathcal{B}_\mathcal{Z}}$ and let $C = [(U \circ T)]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{Z}}$. Let $\mathcal{B}_\mathcal{V} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$. Let $\mathcal{B}_\mathcal{W} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_p)$. Let $\mathcal{B}_\mathcal{Z} = (\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_m)$. Let us denote by $\left[\varphi_{\mathcal{B}_\mathcal{Z}}(\cdot)\right]_i$ as the i^{th} co-ordinate (slot) of the co-ordinate map $\varphi_{\mathcal{B}_\mathcal{Z}}$.

Thus,

$$\begin{aligned}
C[i,j] &= \left[\varphi_{\mathcal{B}_\mathcal{Z}}(U\circ T)(\boldsymbol{v}_j)\right]_i \\
&= \left[\varphi_{\mathcal{B}_\mathcal{Z}} U\left(T(\boldsymbol{v}_j)\right)\right]_i \\
&= \left[\varphi_{\mathcal{B}_\mathcal{Z}} U\left(\sum_{l=1}^{p} a_{lj}\boldsymbol{w}_l\right)\right]_i \\
&= \left[\varphi_{\mathcal{B}_\mathcal{Z}}\left(\sum_{l=1}^{p} a_{lj} U(\boldsymbol{w}_l)\right)\right]_i \\
&= \left[\varphi_{\mathcal{B}_\mathcal{Z}}\left(\sum_{l=1}^{p} a_{lj} \sum_{i=1}^{m} b_{il}\boldsymbol{z}_i\right)\right]_i \\
&= \left[\varphi_{\mathcal{B}_\mathcal{Z}}\left(\sum_{i=1}^{m}\left(\sum_{l=1}^{p} b_{il}a_{lj}\right)\boldsymbol{z}_i\right)\right]_i \\
&= \sum_{l=1}^{p} b_{il}a_{lj},
\end{aligned}$$

which by definition means that $C = BA$. □

For any function f, it is important to ask if there is a y such that $f(x) = y$ is meaningful. If f were invertible, then this would mean that $x = f^{-1}\circ f = f^{-1}(y)$. We will see shortly that when f is a linear transformation, such a question amounts to a solution of a system of linear equations. We have already seen invertible transformations when dealing with isomorphisms. We make this notion precise and discuss a few consequences in the following paragraphs.

Definition 2.3.12. *Let $\mathcal{V},\mathcal{W}$ be vector spaces and let $T:\mathcal{V}\to\mathcal{W}$ be a linear transformation. Let $I_\mathcal{V}$ and $I_\mathcal{W}$ be the identity linear transformations on $\mathcal{V}$ and $\mathcal{W}$ respectively. A function, $U:\mathcal{W}\to\mathcal{V}$ is said to be an inverse of T if,*

(1) $U\circ T = I_\mathcal{V}$ and,
(2) $T\circ U = I_\mathcal{W}$.

Remark 2.3.3. *The definition for invertibility of T described above is equivalent to stating that T is one-one and onto. If U is the inverse of T, then it is unique and we denote U by T^{-1}.*

Remark 2.3.4. *If T,U are invertible then $(U\circ T)^{-1} = T^{-1}\circ U^{-1}$. Morevore, $(T^{-1})^{-1} = T$.*

Theorem 2.3.5. Let $\mathcal{V},\mathcal{W}$ be vector spaces and let $T:\mathcal{V}\to\mathcal{W}$ be an invertible linear transformation. Then T^{-1} is also a linear transformation.

Proof. First we will show that for any $\boldsymbol{w}_1,\boldsymbol{w}_2\in\mathcal{W}$,

$$T^{-1}(\boldsymbol{w}_1+\boldsymbol{w}_2) = T^{-1}(\boldsymbol{w}_1) + T^{-1}(\boldsymbol{w}_2).$$

Let $\boldsymbol{u}_1 \in \mathcal{V}$ be equal to $T^{-1}(\boldsymbol{w}_1 + \boldsymbol{w}_2)$ and let $\boldsymbol{u}_2 \in \mathcal{V}$ be equal to $T^{-1}(\boldsymbol{w}_1) + T^{-1}(\boldsymbol{w}_2)$. Then,

$$T(\boldsymbol{u}_1) = T \circ T^{-1}(\boldsymbol{w}_1 + \boldsymbol{w}_2) = \boldsymbol{w}_1 + \boldsymbol{w}_2.$$

And,

$$T(\boldsymbol{u}_2) = T(T^{-1}(\boldsymbol{w}_1) + T^{-1}(\boldsymbol{w}_2)) = T(T^{-1}(\boldsymbol{w}_1)) + T(T^{-1}(\boldsymbol{w}_2)) = \boldsymbol{w}_1 + \boldsymbol{w}_2.$$

Since, T is injective $\boldsymbol{u}_1 = \boldsymbol{u}_2$. Hence,

$$T^{-1}(\boldsymbol{w}_1 + \boldsymbol{w}_2) = T^{-1}(\boldsymbol{w}_1) + T^{-1}(\boldsymbol{w}_2).$$

Similary, we can show scalar multiplication. □

Similarly, we can define invertible matrices.

Definition 2.3.13. *The $n \times n$ identity matrix, denoted by I_n, is defined by*

$$I_n = [\delta_{ij}]_{n\times n},$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

Remark 2.3.5. *It is easy to see that if $\mathcal{V}$ is a vector space of dimension n and if $\mathcal{B}_\mathcal{V}$ is an ordered basis for $\mathcal{V}$, then $[I]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{V}} = I_n$.*

Definition 2.3.14. *Let $A \in \mathcal{M}_{m\times n}(\mathbb{F})$. Then A is invertible if there is a matrix $B \in \mathcal{M}_{m\times n}(\mathbb{F})$ such that*

$$AB = BA = I_n.$$

We say that $B = A^{-1}$ and $A = B^{-1}$.

Proposition 2.3.6. *Let $T : \mathcal{V} \to \mathcal{W}$ be a linear transformation, where $\mathcal{V}, \mathcal{W}$ are finite dimensional vector spaces. If T is invertible, then* $\dim(\mathcal{V}) = \dim(\mathcal{W})$.

Proof. By the rank-nullity theorem, $\dim(\mathcal{V}) = \dim(\ker T) + \dim(\operatorname{Im} T)$. Since, T is injective, $\dim(\ker T) = 0$ and since T is surjective, $\dim \operatorname{Im} T = \dim \mathcal{W}$. Thus, $\dim(\mathcal{V}) = \dim(\mathcal{W})$. □

Theorem 2.3.6. Let $\mathcal{V}, \mathcal{W}$ be finite dimensional vector spaces with ordered basis $\mathcal{B}_\mathcal{V}, \mathcal{B}_\mathcal{W}$ respectively. Let $T : \mathcal{V} \to \mathcal{W}$ be a linear transformation. Then T is invertible if and only if $[T]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{W}}$ is invertible.

Proof. Suppose T is invertible. Then $\dim \mathcal{V} = \dim \mathcal{W} = n$ say. Let, $\mathcal{B}_\mathcal{V} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ and $\mathcal{B}_\mathcal{W} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n)$ be ordered bases for $\mathcal{V}, \mathcal{W}$ respectively. Let $[T]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{W}}$ be the matrix of T and let $\left[T^{-1}\right]_{\mathcal{B}_\mathcal{W}}^{\mathcal{B}_\mathcal{V}}$ be the matrix of T^{-1}. Then,

$$I_n = [I_\mathcal{V}]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{V}} = [T \circ T]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{V}} = [T]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{W}} \left[T^{-1}\right]_{\mathcal{B}_\mathcal{W}}^{\mathcal{B}_\mathcal{V}}.$$

Thus, $[T]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{W}}$ is ivertible.

Suppose $A = [T]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{W}}$ is invertible. Thus there is a matrix B such that $BA = I_n = AB$. The goal is to show that T is invertible i.e. there is a $U : \mathcal{W} \to \mathcal{V}$ such

that $U \circ T = I_{\mathcal{V}}$ and $T \circ U = I_{\mathcal{W}}$. The idea is to use B to construct U. Let $\check{U} = L_B$ i.e. $\check{U}(\boldsymbol{x}) = B\boldsymbol{x}$ for any $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ in $\mathbb{F}^n$. Then we can construct U as follows:

$$U = \varphi_{\mathcal{B}_{\mathcal{V}}}^{-1} \circ \check{U} \circ \varphi_{\mathcal{B}_{\mathcal{W}}}.$$

Then for any $\boldsymbol{v}_j \in \mathcal{B}_{\mathcal{V}}$,

$$\begin{aligned} U \circ T &= \left(\varphi_{\mathcal{B}_{\mathcal{V}}}^{-1} \circ \check{U} \circ \varphi_{\mathcal{B}_{\mathcal{W}}}\right)(T(\boldsymbol{v}_j)) \\ &= \varphi_{\mathcal{B}_{\mathcal{V}}}^{-1}\left(\check{U} \circ \varphi_{\mathcal{B}_{\mathcal{W}}} \circ T(\boldsymbol{v}_j)\right) \\ &= \varphi_{\mathcal{B}_{\mathcal{V}}}^{-1}\left(\check{U}(A[:,j])\right) \\ &= \varphi_{\mathcal{B}_{\mathcal{V}}}^{-1}\left(BA[:,j]\right) \\ &= \varphi_{\mathcal{B}_{\mathcal{V}}}^{-1}(\boldsymbol{e}_j) \\ &= \boldsymbol{v}_j. \end{aligned}$$

Thus, $U \circ T = I_{\mathcal{V}}$. Here, we used the fact that $\varphi_{\mathcal{B}_{\mathcal{W}}} \circ T(\boldsymbol{v}_j) = A[:,j]$ and that $BA = I_n$. Similarly, we can show $T \circ U = I_{\mathcal{W}}$. □

We saw that $\hom(\mathcal{V},\mathcal{W})$ and $\mathcal{M}_{m\times n}(\mathbb{F})$ are isomorphic vector spaces where $\dim \mathcal{V} = n$ and $\dim \mathcal{W} = m$. What is the dimension of $\hom(\mathcal{V},\mathcal{W})$? Note, that it is easy to see that the dimension of $\mathcal{M}_{m\times n}(\mathbb{F})$ is mn. This is because we can find a basis $\{A_{ij} \ : \ 1 \le i \le m, 1 \le j \le n\}$ of size mn where A_{ij} has all zeros except for the ij entry which is 1. We saw that any invertible linear transformation must be a map from vector spaces of equal dimensions and hence $\dim \hom(\mathcal{V},\mathcal{W})$ must be mn. We make this precise.

Proposition 2.3.7. *Let $\mathcal{V},\mathcal{W}$ be finite-dimensional vector spaces over a field $\mathbb{F}$. Then $\mathcal{V}$ is isomorphic to $\mathcal{W}$ if and only if $\dim \mathcal{V} = \dim \mathcal{W}$.*

Proof. If $\mathcal{V},\mathcal{W}$ are isomorphic, then there is an invertible linear transformation $T : \mathcal{V} \to \mathcal{W}$. Thus, $\dim \mathcal{V} = \dim \mathcal{W}$. If $\dim \mathcal{V} = \dim \mathcal{W}$, then $\mathcal{V}$ and $\mathcal{W}$ have the same number of basis vectors. Let, $\mathcal{B}_{\mathcal{V}} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ and $\mathcal{B}_{\mathcal{W}} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n)$ be ordered bases for $\mathcal{V}$ and $\mathcal{W}$ respectively. Then there is a unique linear transformation T such that $T(\boldsymbol{v}_j) = \boldsymbol{w}_j$. But this means that $\operatorname{Im} T = span T(\mathcal{B}_{\mathcal{V}}) = span \mathcal{W} = \mathcal{W}$ and so T is onto. But since $\dim \mathcal{V} = \dim \mathcal{W}$, it means that T is also injective. Hence T is invertible. □

We can make an important observation about isomorphic linear transformation.

Proposition 2.3.8. *Let $T : \mathcal{V} \to \mathcal{W}$ be an isomorphism. Then for any subspace $\mathcal{U}$ of $\mathcal{V}$, $\dim T(\mathcal{U}) = \dim \mathcal{U}$.*

Proof. Let $\mathcal{B}_{\mathcal{U}} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r)$ be a basis for $\mathcal{U}$. Then, $T(\mathcal{U}) = span(T(\boldsymbol{u}_1), T(\boldsymbol{u}_2), \ldots, T(\boldsymbol{u}_r))$. Thus, $\dim T(\mathcal{U}) \le r$. Let $\alpha_1 T(\boldsymbol{u}_1) + \cdots + \alpha_r T(\boldsymbol{u}_r) = \boldsymbol{0}_{\mathcal{W}}$. This means that, $T(\alpha_1 \boldsymbol{u}_1 + \cdots + \alpha_r \boldsymbol{u}_r) = \boldsymbol{0}_{\mathcal{W}}$. Since, T is injective, $\alpha_1 \boldsymbol{u}_1 + \cdots + \alpha_r \boldsymbol{u}_r = \boldsymbol{0}_{\mathcal{V}}$ and since we have basis vectors, each $\alpha_i = 0$. Thus $\dim T(\mathcal{U}) = \dim \mathcal{U}$. □

Proposition 2.3.9. *Let $T : \mathcal{V} \to \mathcal{W}$ be a linear transformation between two finite dimensional vector spaces. Then, rank T is equal to rank $\check{T}$.*

Proof. We know that $\varphi_{\mathcal{B}_\mathcal{W}} \circ T = \check{T} \circ \varphi_{\mathcal{B}_\mathcal{V}}$. From the left side,

$$(\varphi_{\mathcal{B}_\mathcal{W}} \circ T)(\mathcal{V}) = \varphi_{\mathcal{B}_\mathcal{W}}(\operatorname{Im} T).$$

Hence, from the proposition above, since co-ordinate maps are isomorphic, $\dim(\varphi_{\mathcal{B}_\mathcal{W}} \circ T)(\mathcal{V}) = \dim \operatorname{Im} T$. From the right side,

$$(\check{T} \circ \varphi_{\mathcal{B}_\mathcal{V}})(\mathcal{V}) = \check{T}(\varphi_{\mathcal{B}_\mathcal{V}}(\mathcal{V})) = \check{T}(\mathbb{F}^n) = \operatorname{Im} \check{T}.$$

Thus, rank T is equal to rank $\check{T}$. □

We have seen that linear transformations can be viewed as matrices when we specify an ordered bases. However, we can have different ordered bases which will give rise to different matrices for the **same** linear transformation. This concept leads to the definition of similar matrices. We will look into similar matrices when we describe eigenvalues. For now, we will define a transition matrix or a change of co-ordinate basis which arise when we want to see how the co-ordinates of a vector change when we change the basis.

Definition 2.3.15 (Change of Basis)**.** *Let $\mathcal{V}$ be a vector space over the field $\mathbb{F}$ and let $\dim \mathcal{V} = n$. Let $\mathcal{B}_\mathcal{V}$ and $\mathcal{C}_\mathcal{V}$ be two ordered bases for $\mathcal{V}$. We denote the change of basis map by $T_{\mathcal{B}_\mathcal{V} \to \mathcal{C}_\mathcal{V}}$, given by $T_{\mathcal{B}_\mathcal{V} \to \mathcal{C}_\mathcal{V}} = \varphi_{\mathcal{C}_\mathcal{V}} \circ \varphi_{\mathcal{B}_\mathcal{V}}^{-1}$. The change of basis matrix or the transition matrix $M_{\mathcal{B}_\mathcal{V} \to \mathcal{C}_\mathcal{V}}$ is given by $M_{\mathcal{B}_\mathcal{V} \to \mathcal{C}_\mathcal{V}} := M_{T_{\mathcal{B}_\mathcal{V} \to \mathcal{C}_\mathcal{V}}}$.*

Remark 2.3.6. *Let $\mathcal{B}_\mathcal{V} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$. Then,*

$$M_{\mathcal{B}_\mathcal{V} \to \mathcal{C}_\mathcal{V}}[:, j] = \varphi_{\mathcal{C}_\mathcal{V}} \circ \varphi_{\mathcal{B}_\mathcal{V}}^{-1}(\boldsymbol{e}_j) = [\boldsymbol{v}_j]_{\mathcal{C}_\mathcal{V}}.$$

The change of basis matrix enables us to write the co-ordinates of a vector in one basis w.r.t another basis. Thus if $\boldsymbol{v}$ is a vector in $\mathcal{V}$, then

$$[\boldsymbol{v}]_{\mathcal{C}_\mathcal{V}} = M_{\mathcal{B}_\mathcal{V} \to \mathcal{C}_\mathcal{V}} [\boldsymbol{v}]_{\mathcal{B}_\mathcal{V}}.$$

2.4. Dual spaces

Definition 2.4.1 (Dual vector space)**.** *Let $\mathcal{V}$ be a vector space over $\mathbb{F}$. We denote by $\mathcal{V}^*$, the dual vector space of $\mathcal{V}$ which is the set of all linear transformations that map $\mathcal{V}$ into $\mathbb{F}$, i.e. $\mathcal{V}^* = \hom(\mathcal{V}, \mathbb{F})$. We denote the elements of $\mathcal{V}^*$ by f, g, h etc. , thus $f \in \mathcal{V}^*$ if $f : \mathcal{V} \to \mathbb{F}$ is a linear transformation. When $\mathbb{F} = \mathbb{R}$, the elements of $\mathcal{V}^*$ are called linear functionals.*

Example 2.4.1. *Let $\mathcal{V}$ be a vector space over $\mathbb{F}$ and let $\mathcal{B}_\mathcal{V}$ be an ordered basis for $\mathcal{V}$. Let $f_i : \mathcal{V} \to \mathbb{F}$ be given by $f_i(\boldsymbol{v}) = \left([\boldsymbol{v}]_{\mathcal{B}_\mathcal{V}}\right)_i$ be the i^{th} co-ordinate of $\boldsymbol{v}$ w.r.t $\mathcal{B}_\mathcal{V}$. Then $f_i \in \mathcal{V}^*$ is is called the i^{th} co-ordinate function w.r.t $\mathcal{B}_\mathcal{V}$.*

Remark 2.4.1. *If $\mathcal{B}_\mathcal{V} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$, then it is easy to see that*

$$f_i(\boldsymbol{v}_j) = \delta_{ij}.$$

Remark 2.4.2. *Note that $\dim \mathcal{V}^* = \dim \mathcal{V} \dim \mathbb{F} = \dim \mathcal{V}$ if $\mathcal{V}$ is finite dimensional. This suggests that $\mathcal{V}$ and $\mathcal{V}^*$ are isomorphic to each other.*

Theorem 2.4.1. Suppose that $\mathcal{V}$ is a finite dimensional vector space with the ordered basis $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$. Let f_i be the i^{th} co-ordinate function w.r.t $\mathcal{B}_\mathcal{V}$. Let, $\mathcal{B}^* = (f_1, f_2, \ldots, f_n)$ be an ordered list of co-ordinate functions. Then, $\mathcal{B}^*$ is an ordered basis for $\mathcal{V}^*$ and for any $f \in \mathcal{V}^*$,

$$f = \sum_{i=1}^{n} f(\boldsymbol{v}_i) f_i.$$

Proof. Let $f \in \mathcal{V}^*$. We need to show that $\mathcal{B}^*$ spans $\mathcal{V}^*$ that is $f = \beta_1 f_1 + \cdots + \beta_n f_n$, where $\beta_i = f(\boldsymbol{v}_i)$ and $\mathcal{B}^*$ is linearly independent. Note that if we show that $\mathcal{B}^*$ is a linearly independent list, then it must also span $\mathcal{V}^*$ since $\dim \mathcal{V} = n$. Let $\alpha_1 f_1 + \cdots + \alpha_n f_n = \mathbf{0}_{\mathcal{V}^*}$. Then evaluating both sides on $\boldsymbol{v}_i$, we get $\alpha_i \boldsymbol{v}_i = 0$ for all $1 \leq i \leq n$. Thus each α_i is zero. Hence, $\mathcal{B}^*$ is linearly independent and also spans $\mathcal{V}^*$. Thus, $f = \beta_1 f_1 + \cdots + \beta_n f_n$. Again, evaluating both sides on $\boldsymbol{v}_i$ we get $\beta_i = f(\boldsymbol{v}_i)$. □

Remark 2.4.3. *We have the useful duality formula. For any $\boldsymbol{u} \in \mathcal{V}$ and any $f \in \mathcal{V}^*$*

$$\boldsymbol{u} = \sum_{i=1}^{n} f_i(\boldsymbol{u}) \boldsymbol{v}_i,$$

$$f = \sum_{i=1}^{n} f(\boldsymbol{v}_i) f_i,$$

where $\mathcal{B}_\mathcal{V} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ and $\mathcal{B}_{\mathcal{V}^*} = (f_1, f_2, \ldots, f_n)$ are ordered bases for $\mathcal{V}$ and $\mathcal{V}^*$ respectively.

Definition 2.4.2 (Transpose map)**.** *Given any linear transformation $T : \mathcal{V} \to \mathcal{W}$, there is an induced linear transformation $T^t : \mathcal{W}^* \to \mathcal{V}^*$ given by,*

$$T^t(g) := g \circ T,$$

for any $g \in \mathcal{W}^*$.

The property of this induced map is that its matrix is *transpose* of the matrix of T. We define what a transpose matrix means.

Definition 2.4.3. *If $A = [a_{ij}]_{m \times n}$ is a $m \times n$ matrix, then the transpose of A, denote by A^t, is a matrix of size $n \times m$ given by,*

$$A^t = [b_{ij}]_{n \times n},$$

where, $b_{ij} = a_{ji}$.

Thus, a transposed matrix interchanges the row and columns in the original matrix.

Theorem 2.4.2. Let $\mathcal{V}, \mathcal{W}$ be vector spaces over $\mathbb{F}$ with ordered basis $\mathcal{B}_\mathcal{V}, \mathcal{B}_\mathcal{W}$ respectively. For any linear transformation $T : \mathcal{V} \to \mathcal{W}$, the mapping T^t has the property

$$\left[T^t\right]_{\mathcal{B}_{\mathcal{V}^*}}^{\mathcal{B}_{\mathcal{W}^*}} = \left([T]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{W}}\right)^t.$$

Proof. Let $\mathcal{B}_{\mathcal{V}} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ and $\mathcal{B}_{\mathcal{W}} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m)$ be the ordered basis for $\mathcal{V}, \mathcal{W}$ respectively. Let $\mathcal{B}_{\mathcal{V}^*} = (f_1, f_2, \ldots, f_n)$ and $\mathcal{B}_{\mathcal{W}^*} = (g_1, g_2, \ldots, g_m)$ be the corresponding dual basis vectors. Finally, let $(\boldsymbol{e}^n{}_1, \boldsymbol{e}^n{}_2, \ldots, \boldsymbol{e}^n{}_n)$ and $(\boldsymbol{e}^m{}_1, \boldsymbol{e}^m{}_2, \ldots, \boldsymbol{e}^m{}_m)$ be the standard basis vectors for $\mathbb{F}^n$ and $\mathbb{F}^m$ respectively. Let $[T]^{\mathcal{B}_{\mathcal{W}}}_{\mathcal{B}_{\mathcal{V}}} = A = [a_{ij}]_{m\times n}$ and let $[T^t]^{\mathcal{B}_{\mathcal{W}^*}}_{\mathcal{B}_{\mathcal{V}^*}} = B$. Then,

$$\begin{aligned} B[:, j] &= \left(\varphi_{\mathcal{B}_{\mathcal{V}^*}} \circ T^t \circ \varphi^{-1}_{\mathcal{B}_{\mathcal{W}^*}}\right)(\boldsymbol{e}^n_j), \\ &= \left(\varphi_{\mathcal{B}_{\mathcal{V}^*}} \circ T^t\right)(g_j), \\ &= \varphi_{\mathcal{B}_{\mathcal{V}^*}}(g_j \circ T). \end{aligned}$$

Since $g_j \circ T \in \mathcal{V}^*$, we can use the duality formula to write,

$$g_j \circ T = \sum_{i=1}^{n} g_j \circ T(\boldsymbol{v}_i)\, f_i.$$

Thus,

$$b_{ij} = B[i, j] = g_j \circ T(\boldsymbol{v}_i).$$

Now,

$$T(\boldsymbol{v}_i) = \sum_{k=1}^{m} a_{ki} \boldsymbol{w}_k.$$

Thus,

$$g_j(T(\boldsymbol{v}_i)) = a_{ji}.$$

Hence, $B = A^t$.

□

An extremely important theorem is the following:

Theorem 2.4.3. Let $\mathcal{V}, \mathcal{W}$ be finite dimensional vector spaces of $\dim n$ and $\dim m$ respectively and let $T : \mathcal{V} \to \mathcal{W}$ be a linear transformation and let $T^t : \mathcal{W}^* \to \mathcal{V}^*$ be the corresponding transpose map (linear transformation) of T. Then,

(1) T is injective if and only if T^t is surjective.
(2) T is surjective if and only if T^t is injective.

We will need a few more concepts and Lemmas to prove the theorem above.

Definition 2.4.4 (Annihilator)**.** *Let $\mathcal{V}$ be a vector space of $\dim n$ and let $\mathcal{V}^*$ be the dual vector space of $\mathcal{V}$. Let $\mathcal{U} \subset \mathcal{V}$ be a subspace of $\mathcal{V}$ of dimension $\dim r \leq n$. The Annihilator of $\mathcal{U}$, denoted by U^0 is the set,*

$$\mathcal{U}^0 = \{f \in \mathcal{V}^* \ : \ f(\boldsymbol{u}) = 0 \text{ for any } \boldsymbol{u} \in \mathcal{U}\}.$$

Lemma 2.4.1. *$\mathcal{U}^0$ is a subspace of $\mathcal{V}^*$.*

Proof. Clearly $\mathbf{0}_{\mathcal{V}^*}$ is in $\mathcal{U}^0$. Let f, g be elements of $\mathcal{V}^*$. Then, f, g are elements of $\mathcal{V}^*$ and so $(f + g) \in \mathcal{V}^*$. Now for any $\boldsymbol{u} \in \mathcal{U}$, $(f + g)(\boldsymbol{u}) = f(\boldsymbol{u}) + g(\boldsymbol{u}) = 0$. Thus, $(f + g) \in U^0$. Similarly, we can show that $\mathcal{U}^0$ is closed under scalar multiplication. □

Lemma 2.4.2. $\dim \mathcal{U} + \dim \mathcal{U}^0 = \dim \mathcal{V}$.

Proof. Let $\mathcal{B}_{\mathcal{U}} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r)$ be an ordered basis for $\mathcal{U}$. We can extend this to create an ordered basis for $\mathcal{V}$, given by,

$$\mathcal{B}_{\mathcal{V}} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r, \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k),$$

where $r + k = n$. Consider the corresponding dual basis for $\mathcal{V}^*$ given by,

$$\mathcal{B}_{\mathcal{V}^*} = (f_1, f_2, \ldots, f_r, g_1, g_2, \ldots, g_k).$$

Claim: $(g_1, g_2, \ldots, g_k)$ is an ordered basis for $\mathcal{U}^0$. Clearly, $(g_1, g_2, \ldots, g_k)$ is linearly independent being a subset of a basis. Let $h \in \mathcal{U}^0$. Then since $h \in \mathcal{V}^*$, we can write,

$$h = \sum_{i=1}^{r} h(\boldsymbol{u}_i) f_i + \sum_{j=1}^{k} h(\boldsymbol{v}_j) g_j.$$

Since, $h \in \mathcal{U}^0$, $\sum_{i=1}^{r} h(\boldsymbol{u}_i) f_i = 0$ and thus, $(g_1, g_2, \ldots, g_k)$ spans $\mathcal{U}^0$. □

Lemma 2.4.3. *Let $T : \mathcal{V} \to \mathcal{W}$ be a linear transformation and let $T^t : \mathcal{W}^* \to \mathcal{V}^*$ be the corresponding transpose map (linear transformation) of T. Then,*

(1) $\operatorname{Im} T^t = (\ker T)^0$.

(2) $\ker T^t = (\operatorname{Im} T)^0$.

Proof. Showing that something is in the annihilator is easier. Hence, first we will show

(1) $\operatorname{Im} T^t \subset (\ker T)^0$ and,

(2) $\ker T^t \subset (\operatorname{Im} T)^0$.

For the first statement, let $f \in \mathcal{V}^*$ be such that there is a $g \in \mathcal{W}^*$ such that $T^t(g) = f$. Thus, $g \circ T = f$. Let $\boldsymbol{v}$ be an element of $\ker T$. Then,

$$f(\boldsymbol{v}) = g(T(\boldsymbol{v})) = g(\mathbf{0}_{\mathcal{W}}) = 0.$$

Hence, f is in $(\ker T)^0$.

Let $g \in \mathcal{W}^*$ be such that $T^t(g) = \mathbf{0}_{\mathcal{V}^*}$. Thus, $g \circ T = \mathbf{0}_{\mathcal{V}^*}$. Fix a vector $\boldsymbol{w}$ in $\operatorname{Im} T$. Hence, there is a vector $\boldsymbol{v} \in \mathcal{V}$ such that $T(\boldsymbol{v}) = \boldsymbol{w}$. Thus,

$$g(\boldsymbol{w}) = (g \circ T)(\boldsymbol{v}) = \mathbf{0}_{\mathcal{V}^*}(\boldsymbol{v}) = 0.$$

Hence, g is in $(\operatorname{Im} T)^0$.

We have shown,

$$\dim \operatorname{Im} T^t \leq \dim (\ker T)^0$$
$$\dim \ker T^t \leq \dim (\operatorname{Im} T)^0$$

Adding and using the Lemma above along with the rank nullity theorem, we get

$$\begin{aligned}\dim \mathcal{W}^* &\leq \dim (\ker T)^0 + \dim (\operatorname{Im} T)^0 \\ &= (\dim \mathcal{V} - \dim \ker T) + (\dim \mathcal{W} - \dim \operatorname{Im} T) \\ &= \dim \mathcal{V} + \dim \mathcal{W} - (\dim \ker T + \dim \operatorname{Im} T) \\ &= \dim \mathcal{W}.\end{aligned}$$

Thus, we get $\dim \mathcal{W}^* \leq \dim \mathcal{W}$. But we know that $\dim \mathcal{W} = \dim \mathcal{W}^*$. Hence, it must be the case that $\dim \operatorname{Im} T^t = \dim (\ker T)^0$ and $\dim \ker T^t = \dim (\operatorname{Im} T)^0$. Thus, we get the desired result.

□

Now we are ready to prove 2.4.3.

Proof.

$$\begin{aligned}T \text{ is injective.} &\iff \dim \ker T = 0 \\ &\iff \dim (\ker T)^0 = \dim \mathcal{V} = \dim \mathcal{V}^* \\ &\iff \dim \operatorname{Im} T^t = \dim \mathcal{V}^* \\ &\iff T^t \text{ is surjective.}\end{aligned}$$

$$\begin{aligned}T \text{ is surjective.} &\iff \dim \operatorname{Im} T = \dim \mathcal{W} \\ &\iff \dim (\operatorname{Im} T)^0 = 0 \\ &\iff \dim \ker T^t = 0 \\ &\iff T^t \text{ is injective.}\end{aligned}$$

□

Chapter 3

Systems of Linear equations

In this chapter we will study certain rank-preserving operations on matrices and view them in the theory of solving systems of linear equations. Before we proceed further, we will look at different ways of looking at matrix-matrix multiplication.

Let $A = [a_{ij}]_{m\times p}$, let $B = [b_{ij}]_{p\times n}$ and $C = [c_{ij}]_{m\times n}$ such that $C = AB$. Then we can make certain observations about the rows and column of C.

Remark 3.0.4. *The i^{th} row of C can be viewed as the linear combination of the matrix B with the i^{th} row of A. Thus,*

$$C[i,:] = \sum_{k=1}^{p} a_{ik}B.$$

The j^{th} column of C can be viewed as the linear combination of the matrix A with the j^{th} column of B. Thus,

$$C[:,j] = \sum_{k=1}^{p} b_{kj}A.$$

Definition 3.0.5 (Elementary operations)**.** *Let $A = [a_{ij}]_{m\times n}$ be a matrix of size $m \times n$. Any one of the following operations on the rows (columns) of A is called an elementary row (column) operation:*

(1) (TYPE 1) interchanging any two rows (columns) of A,
(2) (TYPE 2) multiplying any row (column) of A by a nonzero constant.
(3) (TYPE 3) adding any constant multiple of row (column) of A to another row (column) of A.

Definition 3.0.6 (Elementary matrix)**.** *An $n \times n$ elementary matrix is a matrix obtained by performing an elementary operation on I_n. To A TYPE 1 elementary matrix is given by $E_1(i,j)$, a TYPE 2 elementary matrix is given by $E_2(i,c)$ and a TYPE 3 elementary matrix is given by $E_3(i,j,c)$, where:*

(1) $E_1(i,j)$ is I_n with rows i,j swapped.
(2) $E_2(i,c)$ is I_n with the row i multiplied by c.
(3) $E_3(i,j,c)$ is I_n with row i equal to row (i) + $c*$row (j).

Example 3.0.2. *Consider I_3. We show a few examples of elementary matrices. $E_1(1,2)$ is given by*

$$E_1(1,2) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$E_2(2,-5)$ is given by

$$E_2(2,-5) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$E_3(1,3,2)$ is given by

$$E_3(1,3,2) = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The following theorem concerns elementary matrices.

Theorem 3.0.4. Let E be an elementary matrix.

(1) E is invertible and is an elementary matrix of the same type as E.
(2) If A is a matrix of size $m \times n$, then an elementary row operation on A corresponds to left multiplication of A by an elementary matrix of size $m \times m$, and an elementary column operation on A corresponds to right multiplication of A by an elementary matrix of size $n \times n$.

Definition 3.0.7. *If A is a matrix of size $m \times n$, we define the rank of A, denoted by rank A, as the rank of the corresponding linear transformation $L_A : \mathbb{F}^n \to \mathbb{F}^m$.*

Remark 3.0.5. *From the definition, it immediately follows that an $n \times n$ matrix is invertible if and only if its rank is equal to n.*

Proposition 3.0.1. *Let $T : \mathcal{V} \to \mathcal{W}$ be a linear transformation between finite dimensional vector spaces and let $\mathcal{B}_\mathcal{V}$ and $\mathcal{B}_\mathcal{W}$ be ordered bases for $\mathcal{V}$ and $\mathcal{W}$ respectively. Then, rank T = rank $[T]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{W}}$.*

Proof. This follows from the fact that if $A = [T]_{\mathcal{B}_\mathcal{V}}^{\mathcal{B}_\mathcal{W}}$, then, $L_A = \check{T}$ and rank T is equal to rank $\check{T}$. □

Proposition 3.0.2. *Let A be an $m \times n$ matrix. If P, Q are invertible $m \times m$ and $n \times n$ matrices, respectively, then*

(1) rank (AQ) = rank A.
(2) rank (PA) = rank A.

(3) rank $(PAQ) =$ *rank* A.

Corollary 3.0.4.1. *Elementary row and column operations are rank preserving.*

Theorem 3.0.5. The rank of a matrix equals the number of linearly independent columns.

Proof. □

Theorem 3.0.6. Let A be a matrix of size $m \times n$ of rank n. Then, $r \leq m, r \leq n$ and by means of elementary row and column transformation, A can be transformed into the matrix,

$$D = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

Chapter 4

Systems of Linear equations

In this chapter we will study certain rank-preserving operations on matrices and view them in the theory of solving systems of linear equations. Before we proceed further, we will look at different ways of looking at matrix-matrix multiplication.

Let $A = [a_{ij}]_{m\times p}$, let $B = [b_{ij}]_{p\times n}$ and $C = [c_{ij}]_{m\times n}$ such that $C = AB$. Then we can make certain observations about the rows and column of C.

Remark 4.0.6. *The i^{th} row of C can be viewed as the linear combination of the matrix B with the i^{th} row of A. Thus,*

$$C[i,:] = \sum_{k=1}^{p} a_{ik}B.$$

The j^{th} column of C can be viewed as the linear combination of the matrix A with the j^{th} column of B. Thus,

$$C[:,j] = \sum_{k=1}^{p} b_{kj}A.$$

Definition 4.0.8 (Elementary operations)**.** *Let $A = [a_{ij}]_{m\times n}$ be a matrix of size $m \times n$. Any one of the following operations on the rows (columns) of A is called an elementary row (column) operation:*

(1) (TYPE 1) interchanging any two rows (columns) of A,
(2) (TYPE 2) multiplying any row (column) of A by a nonzero constant.
(3) (TYPE 3) adding any constant multiple of row (column) of A to another row (column) of A.

Definition 4.0.9 (Elementary matrix)**.** *An $n \times n$ elementary matrix is a matrix obtained by performing an elementary operation on I_n. To A TYPE 1 elementary matrix is given by $E_1(i,j)$, a TYPE 2 elementary matrix is given by $E_2(i,c)$ and a TYPE 3 elementary matrix is given by $E_3(i,j,c)$, where:*

(1) $E_1(i,j)$ is I_n with rows i,j swapped.
(2) $E_2(i,c)$ is I_n with the row i multiplied by c.
(3) $E_3(i,j,c)$ is I_n with row i equal to row (i) + $c*$row (j).

Example 4.0.3. *Consider I_3. We show a few examples of elementary matrices. $E_1(1,2)$ is given by*

$$E_1(1,2) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$E_2(2,-5)$ is given by

$$E_2(2,-5) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$E_3(1,3,2)$ is given by

$$E_3(1,3,2) = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Corollary 4.0.6.1. *Elementary row and column operations are rank preserving.*

Theorem 4.0.7. The rank of a matrix equals the number of linearly independent columns.

Proof. □

Theorem 4.0.8. Let A be a matrix of size $m \times n$ of rank n. Then, $r \leq m, r \leq n$ and by means of elementary row and column transformation, A can be transformed into the matrix,

$$D = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

Appendix A

Preliminary Concepts: Set theory

A.1. Some basic properties of integers

In this section we will list some of the basic properties of Integers without including the proofs of most of them. We denote the integers by $\mathbb{Z}$ and the positive integers by $\mathbb{Z}^+$.

Definition A.1.1. *For any two integers a, b, we say that a divides b iff b is a multiple of a i.e $a|b \equiv b = ak$ for some $k \in \mathbb{Z}$.*

There are times when, for two integers a, b, *exact* division is not possible. In that case we will appeal to Euclid's division theorem.

Theorem A.1.1 (Division Theorem)**.** Let $a, b \in \mathbb{Z}$ with $a \in \mathbb{Z}^+$. Then there are unique integers q, r such that

$$b = a * q + r \quad \text{and } 0 \leq r < a.$$

For two integers a, b we can find all the integers that divide a and all the integers that divide b. We'll be interested in those integers that divide both a and b, particularly the greatest integer that divides them both. This integer is called the greatest common divisor and is denoted by $\gcd(a, b)$ or (a, b).

Definition A.1.2. *Greatest common divisor Let a, b be two integers, at least one of which is non-zero. Then the greatest common divisor of a, b is the unique positive integer d such that*

- *d is a common divisor, i.e $d|a$ and $d|b$.*
- *d is greater than every other common divisor, i.e if $c|a$ and $c|b$ then $c \leq d$.*

To compute the $\gcd(a, b)$ we will use Euclid's recursive algorithm. Two important properties of gcd are:

- If $a|b$ then $\gcd(a, b) = a$.
- For two integers a, b if $b = a * q + r$ then $\gcd(a, b) = \gcd(r, a)$.

As an example of the second property consider the $\gcd(30, 72)$. We can write $72 = 30 * 2 + 12$. So $\gcd(30, 72) = \gcd(12, 30)$. We can continue further by writing $30 = 12 * 2 + 6$, so $\gcd(12, 30) = \gcd(6, 12) = 6$. The last result came from the first property.

Theorem A.1.2 (Euclidean Algorithm)**.** Suppose that $a, b \in \mathbb{Z}^+$. The following procedure defines a finite sequence of positive integers $a_0, a_1, \ldots, a_n$ such that $a_n = \gcd(a, b)$.

(1) Put $a_0 = b$ and $a_1 = a$.
(2) At stage $k \geq 1$ suppose $a_0, a_1, \ldots, a_k$ have been defined. Then using the division theorem we can write
$$a_{k-1} = a_k * q_k + r_k. \quad \text{(A.1.1)}$$
(3) If $r_k = 0$ then stop, $\gcd(a, b) = a_k$ else continue.

The $\gcd(a, b)$ can also be written as an integral combination of a, b. Let us define a linear combination of two integers a, b.

Definition A.1.3. *For any two integers a, b, any integer of the form $m * a + n * b$ where $m, n \in \mathbb{Z}$ is called a linear combination of a, b.*

For example given $-5, 2$ then $30 = 2 * (-5) + 20 * 2$ is a linear combination of $-5, 2$. Such a linear combination is by no means unique. A very important fact (theorem) is that the $\gcd(a, b)$ can be written as a linear combination of a, b.

Theorem A.1.3 (gcd as linear combination)**.** For any two integers a, b, the $\gcd(a, b)$ is a linear combination of a and b.

Let us see how to get the linear combination. Let $a = 136$ and $b = 232$. From Euclid's algorithm (A.1.2) we get the following:

$$\begin{aligned} 232 &= 136 * 1 + 96 \\ 136 &= 96 * 1 + 40 \\ 96 &= 40 * 2 + 16 \\ 40 &= 16 * 2 + 8 \\ 16 &= 8 * 2 + 0 \end{aligned}$$

This tells us that $\gcd(136, 232) = 8$. To get 8 as a linear combination of $136, 232$ we will write each of the numbers towards the left of the equations in the Euclidean Algorithm above as linear combination of $136, 232$.

$$\begin{aligned} 232 &= (232) * 1 + (136) * 0 \\ 136 &= (232) * 0 + (136) * 1 \end{aligned}$$

$$\begin{aligned} 96 &= 232 - 136 * 1 \\ &= (232) * 1 + (136) * -1 \end{aligned}$$

Now $40 = 136 - 96 * 1$. And thus,

$$\begin{aligned} 40 &= 136 - [(232) * 1 + (136) * -1] \\ &= (232) * -1 + (136) * 2 \end{aligned}$$

Similarly $16 = 96 - 40 * 2$ and so $16 = (232) * 3 + (136) * -5$. Thus we get

$$8 = (232) * -7 + (136) * 12$$

Definition A.1.4. *Two integers a, b are co-prime iff 1 can be written as linear combination of a and b i.e iff* $\gcd(a, b) = 1$.

A very interesting problem known due to *Diophantus* is that given three integers a, b and c, find all possible linear combinations of a and b that yield c i . e . find all the integers m, n such that $c = m * a + n * b$.

Theorem A.1.4. For positive integers a, b, c there exists integers m, n such that $m * a + n * b = c$ iff $\gcd(a, b)|c$

As an example consider the puzzle made famous in the movie Die Hard 3. Can 4 gallon be poured using on $3, 5$ gallon jugs? Note that $\gcd(3, 5) = 1$ which divides 4. Thus we can find m, n such that $m * 3 + n * 5 = 4$. In this case $m = 3, n = -1$.

As another example consider the equation $m * 140 + n * 63 = 35$. Does this equation have any solution? The $\gcd(63, 140) = 7$ and we know that $7|35$ so the equation does have a solution. How do we find it. Note that we can write 7 as a linear combination of $63, 140$ as $7 = (63) * 9 + (140) * -4$ and thus since $35 = 5 * 7$ we get

$$35 = 140 * (-20) + (63) * 45$$

How do we get all the solutions? Consider the *homogenous* equation as above i.e $m * 140 + n * 63 = 0$. To find its solution we can observe

$$\begin{aligned} m * 140 + 63 * n = 0 \; &\Leftrightarrow m * 20 + n * 9 = 0 \\ &\Leftrightarrow 20m = -9n \\ &\Leftrightarrow 9 \text{ divides } 20m. \end{aligned}$$

Since $\gcd(9, 20) = 1$, this means that 9 divides m and so $m = 9 * q$ for some $q \in \mathbb{Z}$. Thus plugging it in the equation we get $n = -20 * q$. And so we get the solution pair $(m, n) = (9 * q, -20 * q)$. The following proposition sums up what we did.

Proposition A.1.1. *Suppose a, b are co-prime non-zero integers, then for $(m, n) \in \mathbb{Z}^2$,*

$$a * m + b * n = 0 \Leftrightarrow (m, n) = (b * q, -a * q) \quad \textit{for some } q \in \mathbb{Z}.$$

To find all the solutions for a linear Diophantine equation we will appeal to the following propostion,

Proposition A.1.2. *Suppose that $(m_0, n_0) \in \mathbb{Z}^2$ is a solution to the Diophantine equation $a * m + b * n = c$. Then, for $(m, n) \in \mathbb{Z}^2$,*

$$a * m + b * n = c \Leftrightarrow a * (m - m_0) + b * (n - n_0) = 0.$$

It can be easily proven by using $c = a*m_0+b*n_0$. And using A.1.1 we can find the solution of $a*(m-m_0)+b*(n-n_0) = 0$ given by $(m-m_0, n-n_0) = (b*q, -a*q)$ and so we can find $(m, n) = (m_0 + b * q, n_0 - a * q)$.

As an example consider again the equation $140 * m + 63 * n = 35$. We found a particular solution given by $(m_0, n_0) = (-20, 45)$ and so the solution set by A.1.2 is given by $(m, n) = (-20 + 9 * q, 45 - 20 * q)$. Note that we divide by 7 to get the homogenous case to be co-prime numbers. Thus we get $9 * q$ and $20 * q$ instead of $63 * q$ and $140 * q$.

Congruence class. Another very useful characterization of integers is done by the congruence or modulo construction.

Definition A.1.5. *Let $m \in \mathbb{Z}^+$. Two integers a, b are congruent modulo m whenever $a - b$ is divisible by m. Equivalently a and b are congruent modulo m whenver the remainders when a, b is divided by m are equal. We denote this by*

$$a \equiv b \pmod{m}.$$

For example $20 = 34 \pmod 7$.

The congruence modulo exhibit the following important properties.

Proposition A.1.3. *Let $m \in \mathbb{Z}^+$. Then*

- *For all $a \in \mathbb{Z}$ $a \equiv a \pmod m$.*
- *If $a, b \in \mathbb{Z}$ such that $a \equiv b \pmod m$ then $b \equiv a \pmod m$.*
- *If $a, b, c \in \mathbb{Z}$ such that $a \equiv b \pmod m$ and $b \equiv c \pmod m$ then $a \equiv c \pmod m$.*

Proposition A.1.4. *Cancellation. Let $m \in \mathbb{Z}^+$ and let $a * b_1 \equiv a * b_2 \pmod m$. Then,*

- *$b_1 \equiv b_2 \pmod m$ whenever $\gcd(a, m) = 1$.*
- *$b_1 \equiv b_2 \pmod{m/a}$ whenever $a|m$.*

Proposition A.1.5. *Arithmetic. Let $m \in \mathbb{Z}^+$ and let $a_1, a_2, b_1, b_2 \in \mathbb{Z}$ such that $a_1 \equiv a_2 \pmod m$ and $b_1 \equiv b_2 \pmod m$. Then*

- *$a_1 \pm b_1 \equiv a_2 \pm b_2 \pmod m$.*
- *$a_1 * b_1 \equiv a_2 * b_2 \pmod m$.*

An important question is to determine if $a * x \equiv b \pmod m$ exists for certain values of x? The following theorem tells us that there is a bijective correspondence between solving the linear Diophantine equation and the congruence equation i.e,

Theorem A.1.5. For integers a, b and positive integer $m \in \mathbb{Z}^+$, there is a bijection

$$f : \{(x, y) \in \mathbb{Z}^2 | a * x + m * y = b\} \to \{x \in \mathbb{Z} | a * x \equiv b \pmod{m}\}$$

Thus to get the solution for $a * x \equiv b \pmod{m}$ we must have

- $\gcd(a, m)|b$.
- x is such that (x, y) must be solutions to the linear Diophantine equation $a * x + m * y = b$.

Since arithmetic is allowed in the modulo framework we can define two inverse operations of addition and multiplication. We say that $-a$ is the additive inverse of a modulo m whenever $a + (-a) = 0 \pmod{m}$. Similarly a^{-1} is the multiplicative inverse of a modulo m whenever $a * a^{-1} = 1 \pmod{m}$. Note that for a multiplicative inverse to exist we must have that $\gcd(a, m)|1$. This is only possible when a and m are co-prime and a is non-zero.

Definition A.1.6. *Given $m \geq 2 \in \mathbb{Z}^+$ and an integer a, the congruence class of a modulo m is the set of all integers which are congruent to a modulo m. We denote this congruence class by $[a]_m$. Thus,*

$$[a]_m = \{x \in \mathbb{Z} \,|\, x \equiv a \pmod{m}\}$$

Since $x \equiv a \pmod{m}$ is the same as $a \equiv x \pmod{m}$, the equivalence class $[a]_m$ can be thought of as the set of all integers that give a remainder a when divided by m.

As an example consider $m = 2$. There are only two remainders possible 0 and 1 and hence there are two equivalent classes $[0]_2$ and $[1]_2$. Similarly when $m = 6$ we get $[0]_6$,$[1]_6$,$[2]_6$,$[3]_6$, $[4]_6$ and $[5]_6$.

Definition A.1.7. *Given $m \geq 2 \in \mathbb{Z}^+$. The set of congruence class $[0]_m$, $[1]_m$, $\ldots [m-1]_m$ is denoted by $\mathbb{Z}_m$ i.e*

$$\mathbb{Z}_m = \{[0]_m, [1]_m, \ldots, [m-1]_m\}$$

Note that $\mathbb{Z}_m$ is a finite set. We can define arithmetic in $\mathbb{Z}_m$ as follows: $[a]_m + [b]_m = [a + b]_m$ and $[a]_m * [b]_m = [a * b]_m$.

For example consider $\mathbb{Z}_3$. Then

- $[1]_3 + [2]_3 = [3]_3 = [0]_3$.
- $[2]_3 * [2]_3 = [4]_3 = [1]_3$.

Since arithmetic has been defined for $\mathbb{Z}_m$ we can consider the additive inverse and multiplicative inverse of $[a]_m \in \mathbb{Z}_m$. Thus,

- $[b]_m \in \mathbb{Z}_m$ is the additive inverse of $[a]_m$ whenever $[a]_m + [b]_m = [0]_m$.
- $[b]_m \in \mathbb{Z}_m$ is the multiplicative inverse of $[a]_m$ whenever $[a]_m * [b]_m = [1]_m$.

Again note that for multiplicative inverse of $[a]_m$ to exist it would mean that a and m are co-prime and a is non-zero. As an example in $\mathbb{Z}_{10}$ only $[1]_m$, $[3]_m$, $[7]_m$ and $[9]_m$ will have multiplicative inverses.

Prime factorization.

Definition A.1.8. *A positive integer p is said to be prime iff the only positive divisors of p are p and 1. If a number is not prime then it is composite.*

Thus if a number $q > 1$ is composite then $q = a * b$ for $1 < a < q$ and $1 < b < q$.

We state two important propositions without proof.

Proposition A.1.6.
- *Every integer greater than 1 can be written as a product of primes.*
- *Suppose that p is prime and $p|a * b$, then $p|a$ or $p|b$.*

A very important theorem about prime numbers is that every positive integer can be uniquely factored as a product of primes.

Theorem A.1.6 (Fundamental theorem of Arithmetic)**.** Every positive integer greater than 1 can be uniquely written as a product of prime numbers with the prime factors in the product written in non-decreasing order.

For example $72 = 2^3 * 3^2$.

Note that if $m = p \geq 2 \in \mathbb{Z}^+$ is prime then every non-zero element of $\mathbb{Z}_p$ will have an multiplicative inverse.

This completes an introduction to the integers. We will revisit them in the context of Group theory.

A.2. Foundations of set theory

In this section, we will cover the basics of axiomatic set theory and will conclude with the Axiom of Choice. Most proofs will be omitted. The axiomatic theory of set is just based two (undefined) notions *class* and the *membership relation* denoted by $\in$. All objects are classes. However there are two kinds of classes:

- Sets
- Proper Classes

If x, A are classes then the expression $x \in A$ means that x is an element of A. This leads to our first definition

Definition A.2.1. *Let x be a class. If x is belongs to some class A then x is called an element.*

All elements are denoted by lower case letters. Hence whenever we write x, y, z we mean classes that belong to some class. Whenever we denote classes by capital letters A, B, C then such a class may be an element of some other class or may not be an element at all.

Definition A.2.2. *Let A, B be classes. We define $A = B$ to mean that every class that has A as its element must have B as an element and vice versa. Logically,*

$$A = B \text{ iff } (\forall X)\,[A \in X \implies B \in X \wedge B \in X \implies A \in X]\,.$$

The first axiom is called the **axiom of Extent** and is an equivalent statement of the above definition.

Axiom 1: $A = B$ iff $x \in A \iff x \in B$.

Definition A.2.3. *Let A, B be classes; we define $A \subseteq B$ to mean that every element of A is an element of B. A is called a subclass of B.*

The second axiom is called the **axiom of class construction** and defines way to construct sets from elements. For this we define a *property* as,

Definition A.2.4. *A property $P(x)$ is a mathematical statement involving an element x such that it can be expressed entirely in terms of the logical symbols $\in, \wedge, \vee, \neg, \exists, \forall$ and variables $x, y, z, A, B \ldots$.*

Axiom 2: If $P(x)$ is a property then there exists a class C whose elements are precisely those that satisfy $P(x)$. Logicall we denote C as,

$$C = \{x : P(x)\} .$$

If A and B are classes then the following properties gives very important classes,

(1) $P(x)$ is $x \in A \vee x \in B$.

(2) $P(x)$ is $x \in A \wedge x \in B$.

The class satisfying the first property is called the union of A, B and is denoted by $A \cup B$. The class satisfying the second property is called the intersection of A, B and is denoted by $A \cap B$.

Definition A.2.5. *The universal class $\mathcal{U}$ is the class of all the elements. Thus it contains classes that belong to some class. Thus,*

$$\mathcal{U} = \{x : x = x\}$$

Definition A.2.6. *The empty class $\emptyset$ is the class that has no elements. Thus,*

$$\emptyset = \{x : x \neq x\}$$

Definition A.2.7. *If two classes A, B have no elements in common, they are said to be disjoint. Thus, A, B are disjoint if,*

$$A \cap B = \emptyset.$$

Definition A.2.8. *The complement of a class A, A^{c} is the class of all elements that do not belong to A. Thus,*

$$A^{\mathsf{c}} = \{x : x \notin A\}$$

Note that for any class A, $\emptyset \subseteq A$, $A \subseteq \mathcal{U}$ and $A \cup A^{\mathsf{c}} = \mathcal{U}$ and $A \cap A^{\mathsf{c}} = \emptyset$. The Demorgan Laws provide a duality about union and intersection,

- $(A \cup B)^{\mathsf{c}} = A^{\mathsf{c}} \cap B^{\mathsf{c}}$.
- $(A \cap B)^{\mathsf{c}} = A^{\mathsf{c}} \cup B^{\mathsf{c}}$.

It is very important fact that union, intersection and complement describe an algebra of classes which is summarized below (easily proven),

- Identity Laws:
  - $A \cup A = A$.

  - $A \cap A = A$.

- Associative Laws:
  - $A \cup (B \cup C) = (A \cup B) \cup C$.
  - $A \cap (B \cap C) = (A \cap B) \cap C$.
- Commutative Laws:
  - $A \cup B = B \cup A$.
  - $A \cap B = B \cap A$.
- Distributive Laws:
  - $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
  - $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Definition A.2.9. *The difference of two classes A, B, $A - B$ is the class of those elements that are elements of A but not of B. Thus,*

$$A - B = \{x : x \in A \wedge x \notin B\}$$

Note that $A - B = A \cap B^{\mathsf{c}}$.

If a is an element then from Axiom 2 we can construct the class that contains only a. Such a class is called a *singleton* and is given by:

$$\{a\} = \{x : x = a\}.$$

Similarly we can create the *un-ordered* pair $\{a, b\}$ which is,

$$\{a, b\} = \{x : x = a \vee x = b\}.$$

It is easy to see that $\{a, b\} = \{c, d\}$ when $(a = c) \vee (b = d)$ or $(a = d) \vee (b = c)$. Something that is very important is the notion of ordered pair which we denote by (a, b). What is important about ordered pairs is that if $(a, b) = (c, d)$ the $a = c$ and $b = d$. Thus the *order* in which they appear in the set is important.

Definition A.2.10. *If a, b are elements, then the ordered pair is the class given by,*

$$(a, b) = \{\{a\}, \{a, b\}\}$$

Definition A.2.11. *The Cartesian product of two classes A and B denoted by $A \times B$ is the class of all ordered pairs (x, y) where $x \in A$ and $y \in B$. Thus,*

$$A \times B = \{(x, y) : x \in A \wedge y \in B\}.$$

A class of ordered pairs is called a *Graph*. Thus any subclass of $\mathcal{U} \times \mathcal{U}$ is a graph. If G is a graph, we denote G^{-1} to be the inverse graph given by,

$$G^{-1} = \{(y, x) : (x, y) \in G\}.$$

Definition A.2.12. *If G, H are graphs, then $G \circ H$ is the graph defined as follows:*

$$G \circ H = \{(x, y) \ : \ \exists z \ni (x, z) \in H \ \wedge \ (z, y) \in G\}$$

Theorem A.2.1. If G, H, J are graphs, then the following hold:
(1) $(G \circ H) \circ J = G \circ (H \circ J)$.
(2) $(G^{-1})^{-1} = G$.
(3) $(G \circ H)^{-1} = H^{-1} \circ G^{-1}$.

Proof. We prove in order,

(1) let $(x, y) \in (G \circ H) \circ J$. Then there is a $z \ni (x, z) \in J$ and $(z, y) \in (G \circ H)$. Thus there is a $u \ni (z, u) \in H$ and $(u, y) \in G$. Thus $(x, u) \in H \circ J$. And so $(x, y) \in G \circ (H \circ J)$. The argument can be reversed and so we get the equality of classes.

(2) Let $(x, y) \in (G^{-1})^{-1}$. Hence $(y, x) \in G^{-1}$. Thus $(x, y) \in G$. The other direction is similar.

(3) Let $(x, y) \in (G \circ H)^{-1}$. Hence $(y, x) \in (G \circ H)$. Thus there is a $z \ni (y, z) \in H$ and $(z, x) \in G$. Thus $(z, y) \in H^{-1}$ and $(x, z) \in G^{-1}$, which is just $(x, z) \in G^{-1}$ and $(z, y) \in H^{-1}$. Hence $(x, y) \in H^{-1} \circ G^{-1}$.

□

Definition A.2.13. *Let G be a graph. By the domain of G we mean the class*

$$dom\, G = \{x \ : \ \exists y \ni (x, y) \in G\}.$$

Definition A.2.14. *Let G be a graph. By the range of G we mean the class*

$$range\, G = \{y \ : \ \exists x \ni (x, y) \in G\}.$$

Theorem A.2.2. IF G, H are graphs then
(1) $dom\, G = range\, G^{-1}$.
(2) $dom\, G^{-1} = range\, G$.
(3) $dom\, (G \circ H) \subseteq dom\, H$.
(4) $range\, (G \circ H) \subseteq range\, G$.

The proof of the third statement is as follows,

Proof. Let $x \in dom\, (G \circ H)$. Thus there is a $y \ni (x, y) \in G \circ H$. Thus there is a $z \ni (x, z) \in H$ and $(z, y) \in G$. Thus the existence of $z \ni (x, z) \in H$ means that $x \in dom\, H$. □

An important corollary of the above theorem is that if $range\, H = dom\, G$ then $dom\, G \circ H = dom\, H$. To prove the equality we just need to show that $dom\, H \subseteq dom\, G \circ H$. Consider an element $x \in dom\, H$. Thus there is a $z \ni (x, z) \in H$. Since range of H equal to $dom\, G$, this means that $z \in dom\, G$. Hence there is a $y \ni (z, y) \in G$ and so $(x, y) \in G \circ H$. Thus $x \in dom\, G \circ H$.

Definition A.2.15. *An indexed class is the class denoted by $\{A_i : i \in I\}$ where I is the class whose elements are called indices.*

Formally an indexed class is a graph G and each $A_i = \{x \ : \ (i, x) \in G\}$. Thus if $I = \{1, 2\}$ and $A_1 = \{a, b\}$ and $A_2 = \{e, f\}$ then the indexed class $\{A_i : i \in I\}$ is the graph $G = \{(1, a), (1, b), (2, e), (2, f)\}$.

Definition A.2.16. *Let $\{A_i : i \in I\}$ be an indexed family. Then,*

(1) The union of the classes A_i consists of all those elements x that are contained in atleast one A_i.

$$\bigcup_{i\in I} A_i = \{x \ : \ \exists j \ni x \in A_j\}.$$

(2) The intersection of the classes A_i consists of all those elements x that are contained in each A_i.

$$\bigcap_{i\in I} A_i = \{x \ : \ \forall j, x \in A_j\}.$$

Theorem A.2.3. Let $\{A_i : i \in I\}$ be an indexed class and B be any class. Then,
(1) If $B \subseteq A_i$ for every $i \in I$ then $B \subseteq \bigcap_{i\in I} A_i$.
(2) If $A_i \subseteq B$ for every $i \in I$ then $\bigcup_{i\in I} A_i \subseteq B$.

The Generalized DeMorgan's Laws and Distributive laws can be restated as:

Theorem A.2.4 (DeMorgan's Laws)**.** Let $\{A_i : i \in I\}$ be an indexed class. Then,
(1) $(\bigcup_{i\in I} A_i)^{\mathsf{c}} = \bigcap_{i\in I} {A_i}^{\mathsf{c}}$
(2) $(\bigcap_{i\in I} A_i)^{\mathsf{c}} = \bigcup_{i\in I} {A_i}^{\mathsf{c}}$

Theorem A.2.5 (Distributive Laws)**.** Let $\{A_i : i \in I\}$ and $\{B_j : j \in J\}$ be indexed classes. Then,
(1) $(\bigcup_{i\in I} A_i) \bigcap (\bigcup_{j\in J} B_j) = \bigcup_{(i,j)\in I\times J} A_i \cap B_j$.
(2) $(\bigcap_{i\in I} A_i) \bigcup (\bigcap_{j\in J} B_j) = \bigcap_{(i,j)\in I\times J} A_i \cup B_j$.

Theorem A.2.6. Let $\{G_i : i \in I\}$ be a family of graphs. Then,
(1) $dom\,(\bigcup_{i\in I} G_i) = \bigcup_{i\in I}(dom\, G_i)$.
(2) $range\,(\bigcup_{i\in I} G_i) = \bigcup_{i\in I}(range\, G_i)$.

In the begining we noted that there were two kinds of classes. We now define the most important kind of class called *Set*.

Definition A.2.17. *A class X is called a set if there is a class Y such that $X \in Y$.*

If for all class Y, $X \notin Y$ then X is called a proper class. The remaining axioms all concern sets.

Axiom 3: Every subclass of a set is itself a set.

Such a subclass is called a *subset*. Note that for any class B, if A is a set then $A \cap B \subseteq A$. Thus *intersections* are sets. The next axiom gives the existence of sets.

Axiom 4: The empty class $\emptyset$ is a set.

Axiom 5: If a, b are sets then the un-ordered pair $\{a, b\}$ is a set.

Note that $\emptyset$ is a set and the set containing the emptyset $\{\emptyset\}$ is also a set i.e we used $b = a$ in the above axiom to construct this set. Thus we can form a new set $\{\emptyset, \{\emptyset\}\}$. We can continue this forever.

Definition A.2.18. *Let A be a set. By the power set of A we mean the class which contains all subsets of A. Thus,*

$$\mathscr{P}(A) = \{X \ : \ X \subseteq A\}.$$

The next two axioms concerns sets of sets.

Axiom 6: If $\mathcal{A}$ is a set of sets then $\bigcup \mathcal{A}$ is also a set.

Note that $\bigcup \mathcal{A}$ is the set $\{x \ : \ \exists A \in \mathcal{A} \ni x \in A\}$.

Axiom 7: If A is a set then $\mathscr{P}(A)$ is also a set.

The following theorem shows that the cartesian product of two sets is also a set.

Theorem A.2.7. If A, B are sets then $A \times B$ is also a set.

Proof. We will show that $A \times B$ is an subset of $\mathscr{P}(\mathscr{P}(A \cup B))$ and thus by Axiom 3 is a set. Let $(x, y) \in A \times B$. Note that $(x, y) = \{\{x\}, \{x, y\}\}$. But $\{x\} \in \mathscr{P}(A \cup B)$ and $\{y\} \in \mathscr{P}(A \cup B)$. Thus $\{\{x\}, \{x, y\}\}$ is a subset of $\mathscr{P}(A \cup B)$. That is $(x, y) \in \mathscr{P}(\mathscr{P}(A \cup B))$. Hence $A \times B$ is a subset of $\mathscr{P}(\mathscr{P}(A \cup B))$. □

Now we will look into the set-theoretic definition of functions. A function f is a *triple* (A, B, f) where A, B are sets and $f \subseteq A \times B$ is a graph satisfying the following conditions:

F 1: For every $x \in A$ there is a $y \in B$ such that $(x, y) \in f$.

F 2: For every $x \in A$, if $y_1, y_2 \in B$ such that $(x, y_1) \in f$ and $(x, y_2) \in f$ then $y_1 = y_2$.

We usually denote (A, B, f) as $f : A \to B$ and write $(x, y) \in f$ as $f(x) = y$. Thus the above conditions become:

F 1: For every $x \in A$ there is a $y \in B$ such that $f(x) = y$.

F 2: For every $x \in A$, $f(x) = y_1$ and $f(x) = y_2$ implies $y_1 = y_2$.

We note that if A, B are sets then any graph $f \subseteq A \times B$ is a function iff

(1) $dom\, f = A$.

(2) $range\, f \subseteq B$.

(3) $F2$ is satisfied.

Some important functions are:

(1) INJECTIVE, A function $f : A \to B$ is said to be injective iff for $x_1, x_2 \in A$, $f(x_1) = f(x_2)$ implies that $x_1 = x_2$.

(2) SURJECTIVE, A function $f : A \to B$ is surjective iff for every $y \in B$ there is a $x \in A$ such that $y = f(x)$, i.e range $f = B$.

(3) BIJECTIVE, A function $f : A \to B$ is bijective iff it is both surjective and injective.

Some examples are as follows:

(1) Identity function. The function $i_A : A \to A$ given by $i_A(x) = x$ for every $x \in A$ is called the identity function.

(2) Inclusion function. Let A, B be sets such that $B \subseteq A$. The function $i_B : B \to A$ is called the inclusion function. Note that $i_B(x) = x$ for every $x \in B$.

(3) Characteristic function. Let 2 designate the class of all functions with two elements, say the class $\{0, 1\}$. If $B \subset A$, then the characteristic function of B is given by $\chi_B : B \to 2$ such that whenever $x \in B$ then $\chi_B(x) = 1$, otherwise $\chi_B(x) = 0$.

(4) Restriction function. Let $C \subseteq A$ and $f : A \to B$. Then the restriction of f to C is the function given by $f_{|_C} : C \to B$ such that $f_{|_C}(x) = f(x)$ for every $x \in C$.

The next theorem concerns composition.

Theorem A.2.8. If $f : A \to B$ and $g : B \to C$ are functions then $g \circ f : A \to C$ is a function and $(g \circ f)(x) = g(f(x))$ for every $x \in A$.

The proof is easy once we notice that since g is a fuction $dom\, g = B$ and so range of f is subset of domain of g thus $dom\, g \circ f = A$. Also we observe that $range\, g \circ f \subseteq range\, g$ for any graph. That $F2$ is satisfied is easy.

Definition A.2.19. *A function $f : A \to B$ is invertible if the inverse graph is a function $f^{-1} : B \to A$.*

Theorem A.2.9. A function is invertible iff it is bijective. Furthermore, if a function is invertible then the inverse is a bijective function.

A useful characterization of invertible function is the following:

Theorem A.2.10. A function $f : A \to B$ is invertible iff there is a function $g : B \to A$ such that $g \circ f = i_A$ and $f \circ g = i_B$. If such a function g exists then $g = f^{-1}$.

Theorem A.2.11. A function $f : A \to B$ is injective iff there is a function $g : B \to A$ such that $g \circ f = i_A$.

Proof. Let a function $g : B \to A$ be such that $g \circ f(x) = x$ for every $x \in A$. Consider x_1, x_2 such that $f(x_1) = f(x_2)$. Thus $x_1 = (g \circ f)(x_1) = g(f(x_1)) = g(f(x_2)) = x_2$. Hence f is injective. Consider an injective function f and fix an element $a \in A$. Construct $g : B \to A$ as follows. If $y \in range\, f$ let $g(y) = f(x)$. If $y \notin range\, f$ then $g(y) = a$. Easy to see that $g \circ f = i_A$. □

Theorem A.2.12. Let $f : A \to B$ and $g : B \to C$ be functions. Then we can state the following about the composition $g \circ f$,

(1) If f, g are injective then so is $g \circ f$.
(2) If f, g are surjective then so is $g \circ f$.
(3) If f, g are bijective then so is $g \circ f$.

Now we will define Direct and Inverse images of sets under a function.

Definition A.2.20. *Let $f : A \to B$ be a function and consider a set $C \subseteq A$. Then the direct image of C under f is the set of all images of elements of C,*

$$f(C) = \{y \in B \ : \ \exists x \in C \ni f(x) = y\}.$$

Definition A.2.21. *Let $f : A \to B$ be a function and consider a set $D \subseteq B$. Then the inverse image of D under f is the set of all elements in A whose images are elements of D,*

$$f^{-1}(D) = \{x \in A \ : \ \exists y \in D \ni f(x) = y\}.$$

We can alternatively write the inverse image of D as the set

$$f^{-1}(D) = \{x \in A \ : \ f(x) \in D\}.$$

It is important to see how direct images and inverse images act on generalized unions and intersections. The next theorem summarizes there actions.

Theorem A.2.13. Let $f : A \to B$ and let $\{C_i\}_{i\in I}$ and $\{D_i\}_{i\in I}$ be sub-families in A and B respectively. Then,

(1) $f\left(\bigcup_{i\in I} C_i\right) = \bigcup_{i\in I} f(C_i)$,
(2) $f\left(\bigcap_{i\in I} C_i\right) \subseteq \bigcap_{i\in I} f(C_i)$,
(3) $f^{-1}\left(\bigcup_{i\in I} D_i\right) = \bigcup_{i\in I} f^{-1}(D_i)$,
(4) $f^{-1}\left(\bigcap_{i\in I} D_i\right) = \bigcap_{i\in I} f^{-1}(D_i)$,

Thus, we see that the inverse image is well behaved w.r.t unions and intersections (and complements). Note that the inverse and direct image are functions that maps powersets, i.e $C(\in \mathscr{P}(A)) \mapsto f(C)\,(\in \mathscr{P}(B))$ and $D(\in \mathscr{P}(B)) \mapsto f^{-1}(D)\,(\in \mathscr{P}(A))$. This is easy to see if we observe that if $C_1 = C_2$ then $f(C_1) = f(C_2)$. Similarly for inverse images. However, the converse is not true in general.

We have defined the *product* of two classes as $A \times B$ as the class of all the ordered pairs in A and B. We can extend this idea to the product of finite classes $A_1, A_2, \ldots, A_n$ as the class of all *n-tuple* $(a_1, a_2, \ldots, a_n)$ such that $a_i \in A_i$ for all $1 \leq i \leq n$. However, we have a potential problem if we an arbitrary indexed family, $\{A_i \ : \ i \in I\}$. In such a case we have to redefine what a product of classes mean.

Definition A.2.22. *Let $\{A_i \ : \ i \in I\}$ be an indexed family of classes; let*

$$A = \bigcup_{i \in I} A_i.$$

The product of the classes A_i is defined to be the class

$$\prod_{i \in I} A_i = \{f \ : \ f \ : I \to A \text{ and } f(i) \in A_i \, \forall \, (i \in I)\}.$$

We will adopt the following notational convention: we designate elements of a product $\prod_{i \in I} A_i$ with boldface letters $\boldsymbol{a}, \boldsymbol{b}$ etc. If $\boldsymbol{a}$ is an element of $\prod_{i \in I} A_i$, we will denote by a_j as $\boldsymbol{a}(j)$. We call a_j as the j^{th} co-ordinate of $\boldsymbol{a}$. Let $A = \prod_{i \in I} A_i$, corresponding to each index we define a function $\Pi_i \ : A \to A_i$ by $\Pi_i(\boldsymbol{a}) = a_i$. We call Π_i as the $i^{th} - projection$ of A to A_i.

Definition A.2.23. *If A, B are classes, we denote by B^A as the class of all functions whose domain is A and whose co-domain is B.*

In particular if $2 = \{0, 1\}$ denotes the class of two elements, then 2^A is the class of all functions from A to $\{0, 1\}$.

Theorem A.2.14. If A is a set, then $\mathscr{P}(A)$ and 2^A are in $1-1$ correspondence.

Proof. We will show that there is a function $f \ : \mathscr{P}(A) \to 2^A$, such that f is injective. For any $B \in \mathscr{P}(A)$ define $f(B) = \boldsymbol{\chi}_B$. Easy to see that f is injective. Infact there is a bijection. Let $g \in 2^A$. Define $B = g^{-1}(\{1\})$. Then $g = \boldsymbol{\chi}_B$. □

Let us list a couple more axioms that will *almost* complete the set construction axiom.

Axiom 9: If A is a non-empty set, there is an element $a \in A$ such that $a \cap A = \emptyset$.

The above axiom states that a set is disjoint from its elements. Hence if A is a set, the singleton $\{A\} \neq A$.

Axiom 10: If A is a set and $f \ : A \to B$ is a surjective function, then B is a set.

Next we define relations on sets.

Definition A.2.24. *Let A be a class, by a relation R in A we mean an arbitrary subclass of $A \times A$.*

Let R be a relation in A, then

(1) (Reflexive) R is reflexive if for every $a \in A$, $(a, a) \in R$.

(2) (Irreflexive) R is irreflexive if for every $a \in A$, $(a, a) \notin R$.

(3) (Symmetric) R is symmetric if $(a, b) \in R \implies (b, a) \in R$.

(4) (Asymmetric) R is asymmetric if $(a, b) \in R \implies (b, a) \notin R$.

(5) (Anti-symmetric) R is anti-symmetric if $(a, b), (b, a) \in R \implies a = b$.

(6) (Transitive) R is transitive if $(a, b), (b, c) \in R \implies (a, c) \in R$.

Definition A.2.25. *A relation R in A is called an equivalence relation if it is Reflexive, Transitive and Symmetric.*

Definition A.2.26. *A relation R in A is called a partial order relation if it is Reflexive, Transitive and Anti-symmetric.*

Definition A.2.27. *A relation R in A is called a strict order relation if it is Irreflexive, Transitive and Asymmetric.*

A.3. Countability