Lecture Notes in Linear Algebra and Matrix Analysis: “E212”

Bharath B. N
Electrical Communication Engineering Department
Indian Institute of Science
Bangalore-560012.

August 2012

Contents

1 Vector Space **1**

1.1 Basic Notions 1

2 Linear Transformation **9**

2.1 Linear Transformation and its Properties 9

2.2 Matrices and Linear transformations 13

2.3 Isomorphisms and Homomorphisms 23

2.4 Dual Space 25

3 Inner Product and Normed Spaces **31**

3.1 Orthogonality and Orthogonal Projection 35

3.1.1 Gram-Schmidt Orthogonalization 36

3.2 Norms on Linear Transformations 39

3.2.1 Bounded Linear Functions 39

This is an introduction to some of the concepts and results in linear algebra that supplements the course “E2 212: Matrix Theory” offered in the department of ECE at the Indian Institute of Science, Bangalore during fall 2012. The document is not a comprehensive study of linear algebra. Unlike any of the standard text book, I will not attempt to prove every theorem that is stated in the document. I recommend the reader to refer to the class notes for a more rigorous coverage of the subject.

Chapter 1

Vector Space

1.1 Basic Notions

Consider the following set:

$$\mathcal{R}^2 := \{(x_1, x_2) : x_1 \in \mathcal{R}, x_2 \in \mathcal{R}\}\,. \tag{1.1}$$

The above set is the set of all vectors in a two dimensional real space. Now, let us investigate some of the properties of the set $\mathcal{R}^2$. If $(x_1, x_2) \in \mathcal{R}^2$ and $(y_1, y_2) \in \mathcal{R}^2$, then the sum defined by $(x_1, x_2) + (y_1, y_2) := (x_1 + y_1, x_2 + y_2) \in \mathcal{R}^2$. Further, $(x_1, x_2)+(y_1, y_2) = (y_1, y_2)+(x_1, x_2)$, i.e., the elements of $\mathcal{R}^2$ satisfy the commutativity property. If a vector is enlarged or contracted, it still remains in $\mathcal{R}^2$, i.e., if $(x_1, x_2) \in \mathcal{R}^2$, $\alpha \in \mathcal{R}$, then $\alpha(x_1, x_2) = (x_1, x_2)\alpha := (\alpha x_1, \alpha x_2) \in \mathcal{R}^2$. Obviously, the *zero vector* $\mathbf{0} := (0, 0) \in \mathcal{R}^2$. This along with the definition of vector addition, it is easy to see that the zero vector is an *additive identity* element of the vector space $\mathcal{R}^2$, i.e., adding any vector to it will not change the vector. For every vector $(x_1, x_2) \in \mathcal{R}^2$, there is a vector $(-x_1, -x_2)$ such that $(x_1, x_2) + (-x_1, -x_2) = 0$, the identity element. Take three vectors $(x_1, x_2), (y_1, y_2), (z_1, z_2)$ in $\mathcal{R}^2$. Then $[(x_1, x_2) + (y_1, y_2)] + (z_1, z_2) = (x_1, x_2) + [(y_1, y_2) + (z_1, z_2)]$.

Now, it is interesting to see if there are any other sets with these properties. We expect that the three dimensional space that we live in should also have these properties. But the way we add these vectors are slightly different. For example $(x_1, x_2, x_3), (y_1, y_2, y_3) \in \mathcal{R}^3$, then the sum is defined as $(x_1, x_2, x_3) +$

$(y_1, y_2, y_3) := (x_1 + y_1, x_2 + y_2, x_3 + y_3)$. Note that the "plus" here is quite different from the "plus" in the case of $\mathcal{R}^2$. Thus, while defining a vector space it is crucial to define the "plus" that makes the space a vector space. I will leave it for the reader to convince themselves that by properly defining the addition, additive identity and scalar multiplication, the space $\mathcal{R}^3$ obeys pretty much like $\mathcal{R}^2$.

Consider the following set

$$\mathcal{R}^n := \{(x_1, x_2, \ldots, x_n) : x_i \in \mathcal{R}, i = 1, 2, \ldots, n\}. \tag{1.2}$$

Can be think[1] of objects of the form $x := (x_1, x_2, \ldots, x_n)$ as vectors? This motivates us to abstract all the properties of $\mathcal{R}^2$. The process of abstraction requires the following two operations:

- Vector addition (the "plus").
- Multiplication of vectors with scalars.

Now, we state the definition of a vector space.

Definition (Vector space) A set V is said to be a vector space over $\mathcal{R}$ if there exist maps (the "plus") $+ : (V \times V) \to \mathcal{R}$ defined by $(x, y) \to x + y$, and multiplication $(\alpha, x) : \mathcal{R} \times V \to \mathcal{R}$ defined by $(\alpha, x) := \alpha x$, satisfying the following properties:

- $\forall x, y \in V$, $x + y \in V$
- There exist a 0 such that $\forall x \in V$, $0 + x = x$
- $\forall x \in V$ there is a $y \in V$ such that $x + y = 0 = y + x$
- For all $x, y, z \in V$, we have $(x + y) + z = x + (y + z)$
- For all $x, y \in V$ and for all $\alpha \in \mathcal{R}$, $\alpha(x + y) = \alpha x + \alpha y$
- $1x = x$
- For all $\alpha, \beta \in \mathcal{R}$, and $x \in V$, we have $(\alpha\beta)x = \alpha(\beta x)$

[1] I will leave it for you to see that the space $\mathcal{R}^n$ behaves like $\mathcal{R}^2$.

It is an easy exercise to see that $\mathcal{R}^n$ is a vector space over $\mathcal{R}$ by appropriately defining the above two maps. One might ask whether can we replace the $\mathcal{R}$ in the definition above by some other set? The answer is yes if the set that we replace with should satisfies the property of a field. In general, while talking about a vector space V, we say that V is a vector space over a field $\mathbb{F}$. In the initial part of this notes, we consider the underlying field to be $\mathcal{R}$ (or $\mathcal{C}$ in some cases).

Up to this point, we have been giving examples of a vector space that seems to be a natural extension of $\mathcal{R}^2$. However, the following provides an example of some objects that can be viewed as vectors that are not an obvious extension of $\mathcal{R}^n$.

Example: Consider the set of all continuous functions defined as $F := \{f : X \to \mathcal{R}, f \text{ is continuous}\}$, where $X := [0, 1]$ is a non-empty compact set. Supposing that the set F is a vector space, then we can visualize the functions in F as vectors. The geometrical viewpoint helps us to understand these strange looking objects in a better way! Now, we will see whether the set F is a vector space or not. In fact, we should also mention the field over which the vector space is defined.

Now, we will look at the first property in the definition of a vector space. Let $f_1, f_2 \in F$, then we need to find whether $f_1 + f_2 \in F$ or not. Towards this, let us define the addition as $(f_1 + f_2)(x) := f_1(x) + f_2(x)$ for all $x \in X$. With this definition, and the property that the sum of continuous function is a continuous function, it is clear that the sum of two functions belong to F. Taking the underlying field as $\mathcal{R}$, we see that for all $\alpha \in \mathcal{R}$ and $f \in F$, we have $(\alpha f)(x) := \alpha f(x) \in F$. Now, we define the zero function **0** as $f(x) = 0$ for all $x \in X$. It is easy to see that this function is the additive identity. An easy exercise also shows $\forall f_1, f_2 \in F$, $f_1 + f_2 = f_2 + f_1$, and $\forall f_1, f_2, f_3 \in F$, $(f_1 + f_2) + f_3 = f_1 + (f_2 + f_3)$. This shows that the set F is a vector space over $\mathcal{R}$.

Consider a vector $x := (x_1, x_2, \ldots, x_n)$ in $\mathcal{R}^n$. This vector can be written as $x = x_1(1, 0, \ldots, 0) + x_2(0, 1, 0, \ldots, 0) + \ldots + x_n(0, 0, \ldots, 0, 1)$. We call the set of vectors $e_i := (0, 0, \ldots, 1, 0, \ldots, 0)$, 1 in the i^{th} position, $i = 1, 2, \ldots, n$ as standard vectors. This motivates us to have the following definition.

Definition Let $x_1, x_2, \ldots, x_n$ be any set of vectors in a vector space V over $\mathcal{R}$, and let $\alpha_i \in \mathcal{R}$, $i = 1, 2, \ldots, n$. Then the vector $\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n \in V$ is called the linear combination of the vectors $x_1, x_2, \ldots, x_n$.

Another interesting fact about the standard vectors is that $\alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_n e_n = 0$ implies that all the coefficients have to be zero. Geometrically, it means that no more than two vectors lie in a plane! This can be generalized as follows.

Definition We say that the vectors $x_1, x_2, \ldots, x_n$ in a vector space V over $\mathcal{R}$ are linearly independent if for $\alpha_i \in \mathcal{R}$, $i = 1, 2, \ldots, n$,

$$\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n = 0$$

implies that $\alpha_i = 0$ for all $i = 1, 2, \ldots, n$.

The standard vectors have another interesting property that any vector in $\mathcal{R}^n$ can be written as a linear combinations of it, i.e., for all $x := (x_1, x_2, \ldots, x_n)$ in $\mathcal{R}^n$,

$$x = x_1(1, 0, \ldots, 0) + x_2(0, 1, 0, \ldots, 0) + \ldots + x_n(0, 0, \ldots, 0, 1).$$

Definition We say that the vectors $x_1, x_2, \ldots, x_n$ in a vector space V over $\mathcal{R}$ spans the vector space V if for all $x \in V$, $\exists\ \alpha_i \in \mathcal{R}$, $i = 1, 2, \ldots, n$ such that $x = \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n$.

Exercise Prove that if $x_1, x_2, \ldots, x_n$ spans the vector space then $x_1, x_2, \ldots, x_n, x$ also spans the vector space V for all $x \in V$.

The above exercise indicates that there could be redundancies in the spanning set of vectors. This, however, can be removed one by one until we get a spanning set from which removing even a single vector from it will make the set loose the property of a spanning set.

Definition A set $x_1, x_2, \ldots, x_n$ is called a bases vector of the vector space V if the set is linearly independent and spans the vector space V. The number n is called the *dimension* of the vector space.

Now, we ask the following question: Is the dimension unique? This requires us to prove an important lemma called the replacement lemma:

Lemma 1 ***Replacement Lemma*** *Let* $v_1, \dots, v_n$ *be a set of bases vectors in* V. *Let* v *be any non-zero vector in* V. *Then, there exists a vector* v_i *such that* $(v_1, v_2, \dots, v, \dots, v_n)$ *forms a bases.*

Proof: Since $v \in V$ and $v_1, \dots, v_n$ is a bases vector, we have $v = \sum_{i=1}^{n} \alpha_i v_i$ with at least one $\alpha_i \neq 0$. Without Loss Of Generality (WLOG), let $\alpha_1 \neq 0$. This implies that v_1 can be written as

$$v_1 = \frac{v}{\alpha_1} - \sum_{j=2}^{n} \frac{\alpha_j}{\alpha_1} v_j,$$

which is a linear combination of $v, v_2, \dots, v_n$. It easily follows that this set of vectors span the vector space V. Next, we will prove that it is a linearly independent set, i.e.,

$$\sum_{k=2}^{n} \beta_k v_k + \beta_1 v = 0$$

implies $\beta_i = 0$ for all $i = 1, 2, \dots, n$. Substituting for $v = \sum_{i=1}^{n} \alpha_i v_i$, we get $\beta_2 v_2 + \beta_3 v_3 + \dots, +\beta_n v_n + \beta_1(\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n) = \alpha_1 \beta_1 v_1 + (\beta_2 + \alpha_2 \beta_1) v_2 + (\beta_3 + \alpha_3 \beta_1) v_3 + \dots, +(\beta_n + \beta_1 \alpha_n) v_n = 0$. By linear independence of $v_1, \dots, v_n$, we have $\alpha_1 \beta_1 = 0$, $\beta_2 + \alpha_2 \beta_1 = \dots = \beta_n + \beta_1 \alpha_n = 0$. From $\alpha_1 \beta_1 = 0$ implies $\beta_1 = 0$ since $\alpha_1 \neq 0$. Now, using $\beta_1 = 0$, we have $\beta_2 + \alpha_2 \beta_1 = 0$ implies $\beta_2 = 0$, and so on. Thus, all the coefficients have to be zero. Therefore, the set of vectors $v, v_2, \dots, v_n$ are linearly independent. □

In the following, using the replacement lemma, we will prove that the dimension of a vector space is unique.

Theorem 1 *The dimension of the vector space is unique.*

Proof: Suppose for the sake of contradiction let us assume that there are two set of bases, say $v_1, \dots, v_n$ and $u_1, \dots, u_m$, $m \neq n$. Further, WLOG, let $m < n$. Since $u_1 \neq 0$, by using the replacement lemma, we can replace one of the bases vector in $v_1, \ldots, v_n$, say v_1 with u_1. This results in $u_1, v_2, \ldots, v_n$, which is linearly independent. Similarly, WLOG, replacing v_2 by u_2, we get $u_1, u_2, \ldots, v_n$. By repeatedly applying the replacement lemma until all the first $v_1, \ldots, v_m$ are replaced by $u_1, \ldots, u_m$, we get $u_1, \ldots, u_m, v_{m+1}, \ldots, v_n$. Since $u_1, \ldots, u_m$ are assumed to be linearly independent, the vector $u_1, \ldots, u_m, v_{m+1}, \ldots, v_n$ cannot be linearly independent, a contradiction. Therefore, $m = n$. □

Exercise Let V be a finite dimensional vector space of dimension n. Then, prove that any set of vectors having more than n elements are linearly dependent.

Now, we state and prove the following lemma:

Lemma 2 ***Bases Completion Lemma (BCL)*** *Any linearly independent set of vectors in an n dimensional vector space can be extended to form a bases.*

Proof: Let $v_1, \ldots, v_m$, $m < n$ be an independent set of vectors in V. By *assuming that the bases exists*, let $u_1, \ldots, u_n$ be any bases vector. By replacement lemma, WLOG, we can replace the first m elements of the bases by $v_1, \ldots, v_m$ resulting in $v_1, \ldots, v_m, u_{m+1}, \ldots, u_n$ retaining the bases property. This is indeed an extension of the $v_1, \ldots, v_m$ to a bases vector. □

The above theorem relied on the fact that the bases exists, which seems questionable. However, thanks to the following remarkable theorem which proves the existence of bases.

Theorem 2 *In any finite dimensional vector space, there exists a bases.*

Proof: The proof is omitted for the time being. In fact, the proof involves using the Zorn's lemma in set theory.

Exercise Prove that any vector in a finite dimensional vector space can be uniquely represented as a linear combination of bases vectors.

From the title, one may wonder what a vector space has got to do with matrices. Recall from your undergraduate matrix theory that the ij^{th} element of the matrix $A \in \mathcal{R}^{n\times m}$ denoted $a_{ij} \in \mathcal{R}$. Denote the set of all matrices of dimension $m \times n$ by $\mathcal{M}_{m,n}$. It is an easy exercise to show that it is a vector space over $\mathcal{R}$ of dimension mn. However, this turns out to be a not so elegant way of looking at matrices in vector space theory. In the next chapter, we will show that the study of vector spaces is important by viewing matrices as a representation of a linear map for a given bases.

Chapter 2

Linear Transformation

Consider the following set of linear equations:

$$\mathbf{y} = Ax, \tag{2.1}$$

where $x \in \mathcal{R}^n$ and $A \in \mathcal{R}^{n \times n}$. Naturally, in these problems, one is interested in finding the solution for x. The existence of the solution to the above problem depends on the invertibility of the matrix A. If at all the solution exists, one way to solve the above problem is to reduce the matrix to a simpler form such as diagonal, upper/lower triangle form etc. Now, the following questions arise:

- When is the matrix A invertible?
- Is it possible to convert the matrix A to a diagonal form?
- Is it possible to convert the matrix A to an upper/lower triangle form?

To answer these questions, we will take a slightly general standpoint of viewing the matrices as linear transformations, which is done in the following section.

2.1 Linear Transformation and its Properties

First, we give a definition for linear transformation.

Definition A map $T : V \rightarrow W$ between two vector spaces V and W is said to be linear if the following property is satisfied:

- $T(\alpha v_1 + \beta v_2) = \alpha T v_1 + \beta T v_2$ for all $\alpha, \beta \in \mathcal{R}$

Now, we return to the question that we posed in the beginning of this chapter: when does the inverse for T exists? Intuitively, for all vector $w \in W$, there should be a corresponding element $v \in V$ that the linear transformation maps to, and it should be unique. First of all, the question makes sense if the space W is as big/small as V. Otherwise, there is no hope of finding the inverse. The above intuition brings in the notions of surjective mapping and one-one mapping, as defined below.

Definition A map $T : V \to W$ is said to be surjective if for every $w \in W$, there exists an element $v \in V$ such that $Tv = w$.

Definition A map $T : V \to W$ is said to be injective (or one-one) if for all $v_1 \in V$ and $v_2 \in V$, $Tv_1 = Tv_2$ implies $v_1 = v_2$.

Definition The image of a map $T : V \to W$ is defined as

$$Imag(T) := \{Tv : v \in V\}.$$

Definition The kernel or Null of a map $T : V \to W$ is defined as

$$Null(T) := \{v \in V : Tv = 0 \in W\}.$$

It is an easy exercise to show that $Imag(T)$ is a vector space (Exercise). But note that $Imag(T) \subseteq W$. This motivates us to define another notion called a subspace.

Definition Let V be a vector space. A space $U \subseteq V$ is said to be a subspace if for all $u_1, u_2 \in U \Rightarrow \alpha u_1 + \beta u_2 \in U$ for all $\alpha, \beta \in \mathcal{R}$.

Exercise: Check that the above is a valid definition for subspaces.

As noted earlier, the inverse of a map exists if and only if the map covers the entire range and the mapping is unique, which is the essence of the following theorem.

Theorem 3 *A map $T : V \to V$ is invertible if and only if it is surjective and injective, i.e., it is bijective.*

Proof: Directly follows from the definition of surjective and injective mappings. □

Now, one may wonder what is the use of the above theorem. In fact, it turns out that given a transformation, it is hard to directly verify these properties. This motivates us to investigate some other properties of a map that implies invertibility and it is easily verifiable. Instead of trying out different things, let us see whether there are any other properties of a map that implies that the map is surjective and injective. Let us first investigate the property of a map being injective.

Suppose let the map be injective. Then, for all for all $v_1 \in V$ and $v_2 \in V$, $Tv_1 = Tv_2$ implies $v_1 = v_2$. Let us also assume that the map is linear, we have $Tv_1 = Tv_2 \Rightarrow T(v_1 - v_2) = 0 \in W$. This implies that $v_1 - v_2 = 0 \in V \Rightarrow v_1 = v_2$; the map is injective then the Kernel contains only the zero vector. We state this result as a theorem below.

Theorem 4 *If the linear map $T : V \to W$ is injective then $Null(T) = 0 \in V$.*

Now, let us use the above argument in the reverse direction, i.e., let $Null(T) = 0 \in V$. Let $v_1, v_2 \in V$ be such that $Tv_1 = Tv_2$. From linearity, this implies $T(v_1 - v_2) = 0 \in W$. By the assumption that $Null(T) = 0 \in V$, we have $v_1 = v_2$. This proves that if $Null(T) = 0 \in V$, then the map is injective, which is the essence of the following theorem.

Theorem 5 *The linear map $T : V \to W$ is injective if and only if $Null(T) = 0 \in V$.*

Since $Null(T)$ is a subspace of V, one way to investigate the invertibility of the map is to see how big is $Null(T)$, i.e., what is $dim(Null(T))$? Now, we will answer this question in the general situation.

Since $Null(T)$ is a subspace of V, let $v_1, \ldots, v_m$ be a bases vector of $Null(T)$. By the BCL, this can be extended to a bases of the entire space V. WLOG, let this be $v_1, \ldots, v_m, v_{m+1}, \ldots, v_{m+n}$, i.e., the dimension of V is $m+n$. Now, we know that $Tv_i = 0 \in W$ for all $i = 1, 2, \ldots, m$. Consider

$$Tv_{m+1}, Tv_{m+2}, \ldots, Tv_{m+n},$$

which is in the range space of T. Since range space is a subspace, we expect that the bases should be related to $Tv_{m+1}, Tv_{m+2}, \ldots, Tv_{m+n}$. Now, let $w \in Imag(T)$. Then, there exists a vector $v := \sum_{i=m+1}^{n+m} \alpha_i v_i \in V$ such that $Tv = w$. This implies that $Tv = \sum_{i=m+1}^{n+m} \alpha_i T v_i$, which is a linear combination of $Tv_{m+1}, Tv_{m+2}, \ldots, Tv_{m+n}$. Since every vector in the range space can be written as a linear combination of $\{Tv_{m+1}, Tv_{m+2}, \ldots, Tv_{m+n}\}$,

$$\{Tv_{m+1}, Tv_{m+2}, \ldots, Tv_{m+n}\}$$

spans $Imag(T)$. Naturally, we ask whether this set of vectors form a bases? Only condition that we need to check is the linear independency condition. Let

$$\sum_{i=m+1}^{m+n} \beta_i T v_i = 0.$$

By linearity,

$$\sum_{i=m+1}^{m+n} \beta_i T v_i = 0 \Rightarrow T \sum_{i=m+1}^{m+n} \beta_i v_i = 0.$$

This implies that $\sum_{i=m+1}^{m+n} \beta_i v_i = 0$ (why?). By linear independency of the set

$$\{v_{m+1}, \ldots, v_{m+n}\},$$

we have $\beta_i' s = 0$. This proves that the vector $\{Tv_{m+1}, Tv_{m+2}, \ldots, Tv_{m+n}\}$ forms a bases of the image of T. Now, from the above, we have that the dimension of $Imag(T)$ is n, the dimension of the Kernel of T is m, and the dimension of V is $m+n$. Thus, we have the following theorem:

Theorem 6 *For every linear map $T : V \longrightarrow W$, we have*

$$dim(Imag(T)) + dim(Null(T)) = dim(V).$$

Now, if $Ker(T) = 0 \in V$, then the above theorem implies that $dim(V) = dim(Imag(V))$. If $V = W$, then $dim(Imag(V)) = dim(V)$, the entire space. Thus, the map is both injective and surjective if $Ker(T) = 0 \in V$ and $W = V$! Thus, we have:

Theorem 7 *A linear map $T : V \to V$ is invertible if and only if $Ker(T) = 0$.*

Remark:We will define the dimension of the image of a linear map as its rank, denoted $rank(T)$. The above theorem can be restated as rank plus nullity of a map T is equal to the dimension of the vector space V. Although, we promised to arrive at a condition that is easily verifiable, it looks like the condition $Ker(T) = 0$ is hard to check. Instead, let us check if we can say something about $Ker(T) \neq 0$. This implies that there exists at least one vector $v \in V$, $v \neq 0$ such that $Tv = 0$. This can be written in a slightly different form $T(v - 0v) = 0$. Those who are already familiar with the notions of eigenvectors and eigenvalues would immediately recognize that the above is a problem of finding whether a map has zero as its eigenvalue or not. This seems promising as it amounts to solving a polynomial! At least now, we have some hope that the $Ker(T)$ is computable, and we can hope to answer whether the inverse of a map exists or not. With this hope, we continue to study some additional properties of a linear map and relegate the study of eigenvectors and eigenvalues to the next chapter.

Note that all matrix transformation of the form Ax comes under linear transformation. Is the converse true? In the following, we show that this is indeed true!

2.2 Matrices and Linear transformations

In this section, I will excuse myself by giving a “not so” rigorous explanation of why a matrix can be thought of as a representation for a linear transformation in a vector space with a fixed basis. Consider a linear map $T : V \to W$. Let $\{v_1, \dots, v_n\}$ and $\{w_1, \dots, w_m\}$ be a set of bases vectors for V and W, respectively. Now, consider any vector $v \in V$. Now, let us investigate the action of T on

v. Since $v \in V$, we have $v := \sum_{i=1}^{n} \alpha_i v_i$ for some $\alpha_i' s \in \mathcal{R}$. Now, by linearity, we have

$$Tv = T \sum_{i=1}^{n} \alpha_i v_i = \sum_{i=1}^{n} \alpha_i T v_i.$$

Note that $Tv_i \in W$, and therefore $Tv_i := \sum_{j=1}^{m} \beta_{ij} w_j$ for some $\beta_{ij} \in \mathcal{R}$. Upon substitution, we have

$$Tv = \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{m} \beta_{ij} w_j = \sum_{i,j} \alpha_i \beta_{ij} w_j.$$

Now, since $Tv \in W$, we have $Tv := \sum_{j=1}^{m} \gamma_i w_j$ for some $\gamma_i \in \mathcal{R}$, $i = 1, 2, \ldots, m$. Equating both, we get

$$\sum_{i,j} \alpha_i \beta_{ij} w_j = \sum_{j=1}^{m} \gamma_i w_j.$$

This implies that

$$\sum_{j} \alpha_i \beta_{ij} = \sum_{j} \gamma_j.$$

This in matrix form becomes $B\mathcal{A} = \Gamma$, where β_{ij} is the $ij-$th entry of $B \in \mathcal{R}^{m \times n}$, α_i is the i^{th} entry of $\mathcal{A}$, and γ_j is the j^{th} entry of Γ. For a fixed bases vector, the variables that depend on the vector v is $\mathcal{A}$ and Γ, and not the matrix B. Thus, for a given bases, any linear transformation seems to have a matrix representation. Now on, we can think of linear transformations as matrices with a fixed bases. We state this result as a theorem. We leave it for the reader to use the above discussion as a hint and rigourously prove the following theorem.

Theorem 8 *There is a one-one correspondence between the set of all linear maps from V to W of dimensions n and m, respectively, and the set of all $m \times n$ matrices.*

With this remarkable theorem, all the properties mentioned in this book thus far holds true for the corresponding matrices also. In other words, we can replace linear transformation everywhere with matrices in this book! Now, let us return to the question that we posed in the beginning of this chapter, i.e., when does the inverse of a matrix exists? From theorem 7, it amounts to checking if there is a nonzero vector $x \in \mathcal{R}^{n\times n}$ such that $Ax = 0$. By stacking the columns of the matrix A as $A := [a_1\ a_2\ \ldots\ a_n]$, and writing $x := [x_1, \ldots, x_n]$, the equation $Ax = 0$ can be rewritten as $\sum_{i=1}^{n} x_i a_i = 0$. This is just the linear combination of the columns of the matrix. Thus, the matrix inverse exists if and only if the columns of the matrix are linearly independent. This linear independency of the columns is defined as column rank. Similarly, one can define the row rank. Is there any relationship between column rank and row rank? In the sequel, we address this question.

First, we observe the following interesting fact. For any matrix A, the linear map $T : \mathcal{R}^{n\times m} \rightarrow \mathcal{R}^{n\times m}$ defined by $T(e_i) := a_i$, where e_i is the standard bases vector[1], has one-one correspondence with the matrix for a fixed standard bases (check!). Now, we can define the rank of the matrix A as the rank of the corresponding linear transformation, which is equal to the dimension of the image of T. Note that Te_i, $i = 1, 2, \ldots, n$ must span the $Imag(T)$. But $Te_i = a_i$. This implies that the rank of A is equal to the number of linearly independent columns of the matrix A. Now, is this equal to the number of linearly independent rows of A? The answer is yes:

Theorem 9 *Row rank of any matrix $A \in \mathcal{R}^{m\times n}$ is equal to the column rank of A.*

Proof: Let the column rank of A be $r > 0$. Let $c_1, \ldots, c_r$ be a bases for the column space, and let $C := [c_1, \ldots, c_r] \in \mathcal{R}^{n\times r}$. Then, each columns of A can be written as a linear combination of the bases. In the matrix form, A can be written as $A = CR$, where $R \in \mathcal{R}^{r\times n}$ contains the coefficient of the bases expansion. Note that the column rank of C is r. Now, each rows of the matrix A can be written as a linear combination of the rows of R with coefficients being the elements from C. Thus, the row space of A is contained in the row space of R. Thus, row rank of A is less than or equal to the row rank of R which is at most r. This implies that the row rank of A is at most equal to r which is equal to the column rank of A by assumption. Now, applying the same argument to the transpose of A completes the proof. □

[1]This consists of one in the i^{th} position and zeros in the rest of the positions

Remark: This proof seems a little constructive in nature. There is a more elegant alternative proof of the above theorem which will be introduced in the next chapter.

Since we now know that the column rank and row rank are equal, we can pose the following questions:

- what happens to the rank of a matrix when it is multiplied by another matrix of full rank?
- what happens to the rank of a matrix when it is multiplied by another matrix which is rank deficient?
- what happens to the rank of a matrix by additive perturbation?

We answer these questions in a more general fashion in the subsequent theorems.

Theorem 10 *Let $A \in \mathcal{R}^{n\times m}$ and $B \in \mathcal{R}^{m\times p}$. Then,*

$$rank(AB) \leq \min\{rank(A), rank(B)\}.$$

Proof: Consider $C := AB$. From the proof of theorem 9, the column rank of C is at most equal to the column rank of A, which is equal to $rank(A)$. On the other hand, the row rank of C is at most equal to the row rank of B, i.e., $rank(B)$. Combining the two, we get the desired inequality. □

The above theorem says that by multiplying a matrix A with another matrix can only reduce the rank of the matrix A. Now, we will answer the last question posed above.

Theorem 11 *(Rank Inequality Theorem (RIT)) Let $A \in \mathcal{R}^{m\times n}$ and $B \in \mathcal{R}^{m\times n}$. Then,*

$$rank(A+B) \leq rank(A) + rank(B).$$

Proof: We prove this result in stages. First, consider

$$rank(A+B) := dim\{Imag(A+B)\} = dim\{Ax + Bx : x \in \mathcal{R}^n\}.$$

Now, let us investigate the set

$$\{Ax + Bx : x \in \mathcal{R}^n\}.$$

Note that $Imag\{A\}$ and $Imag\{B\}$ are subspaces. The set $\{Ax + Bx : x \in \mathcal{R}^n\}$ can be viewed as the sum of two subspaces. This occurs frequently in linear algebra and it deserves a definition.

Definition Let U and W be subspaces of V. Then the direct sum $U \oplus W$ is defined as

$$U \oplus W := \{u + w : u \in U, w \in W\}.$$

Verify that the direct sum is indeed a subspace. Also, note that the intersection of subspaces is again a subspace. Let us denote the intersection by $U \cap W := \{x : x \in U \cap W\} \subseteq V$.

Let us denote the image of A and B by U and W, respectively. Now, consider the bases $x_1, \ldots, x_l$ of $U \cap W$. This bases can be extended to the subspace U or W or $U \oplus W$. Let us denote the extension of $x_1, \ldots, x_l$ to U and W by $B_1 := \{x_1, \ldots, x_l, x_{l+1}, \ldots, x_m\}$, and $B_2 := \{x_1, \ldots, x_l, \bar{x}_{l+1}, \ldots, \bar{x}_n\}$, respectively. Now, consider the union

$$B_1 \bigcup B_2 := \{x_1, \ldots, x_l, x_{l+1}, \ldots, x_m, \bar{x}_{l+1}, \ldots, \bar{x}_n\}.$$

We claim that this is a bases of $U \oplus W$. Supposing that this is true, then the proof is complete by a simple observation that $|B_1 \bigcup B_2| = |B_1| + |B_2| - |B_1 \bigcap B_2|$, which implies that $|B_1 \bigcup B_2| \leq |B_1| + |B_2|$.

Clearly, $B_1 \bigcup B_2$ spans the direct sum. Therefore, we need to prove that it is linearly independent. Consider the linear combination

$$\sum_{i=1}^{m} \alpha_i x_i + \sum_{j=l+1}^{n} \beta_j \bar{x}_j = 0.$$

Now, we prove that all the coefficients have to be zero. For the sake of contradiction, let us assume that some $\alpha_j \neq 0$. Then, we can write the vector $x_j = \frac{-1}{\alpha_j}\left[\sum_{i=1, i\neq j}^{m} \alpha_i x_i + \sum_{j=l+1}^{n} \beta_j \bar{x}_j\right]$. This means that the vector x_j is in the span of $B_1 \bigcup B_2/x_j$. Thus, $B_1 \bigcup B_2/x_j$ still spans $U \oplus W$. However, by removing x_j from B_1, any vector of the form $u + 0 \in U \oplus W$ cannot be written as a linear combination of $\{x_1, \ldots, x_l, x_{l+1}, \ldots, x_m, \bar{x}_{l+1}, \ldots, \bar{x}_n\}/x_j$, a contradiction. Therefore, none of the $\alpha_i's$ can be nonzero. By a similar argument, it is easy to see that none of the $\beta_i's$ can be nonzero. Thus, all the coefficients have to be zero, which proves linear independency. □

Now, as a relatively straight forward extension of the rank inequality theorem, we have:

Theorem 12 *Let A and B be two matrices over $\mathcal{R}$ of same dimensions. Then,*

$$|rank(A) - rank(B)| \leq rank(A - B) \tag{2.2}$$

Proof: Writing (2.2) in its glory, we have

$$-rank(A - B) \leq rank(A) - rank(B) \leq rank(A - B) \tag{2.3}$$

Let us first prove the second inequality, i.e., $rank(A) \leq rank(A-B)+rank(B)$. Note that $A = A+B-B$. From theorem 11, the rank of A can be upper bounded as

$$rank(A) = rank(A + B - B) \leq rank(A - B) + rank(B).$$

This proves the second inequality above. Now, writing $B = B + A - A$, the rank of B can be upper bounded as

$$rank(B) \leq rank(B - A) + rank(A).$$

Since $rank(B - A) = rank(A - B)$, the first inequality follows. □

The following is a simple result which follows directly by using $A = A+E-E$, and then using the RIT.

Corollary 1 *Let $A \in \mathcal{R}^{m\times n}$ with $rank(A) = r$ and $E \in \mathcal{R}^{m\times n}$ with*

$rank(E) = k$, $r \leq n$, *then*

$$r - k \leq rank(A + E) \leq r + k. \tag{2.4}$$

Yet another theorem.

Theorem 13 *Let $A \in \mathcal{R}^{m \times n}$, and let $B \in \mathcal{R}^{n \times p}$ be such that $AB = 0$. Then,*

$$rank(A) + rank(B) \leq n.$$

Proof: The equality $AB = 0$ implies that the columns of the matrix B are in the null space of A. This implies that the image space of B is in the null space of A. Thus, $Null(A) \subseteq Imag(B)$ implies that

$$dim(Null(A)) \leq dim(Imag(B)) = rank(B).$$

Applying the rank-nullity theorem to the map $A : \mathcal{R}^n \to \mathcal{R}^p$, we get

$$rank(A) + dim(Null(A)) = n.$$

Using $dim(Null(A)) \leq rank(B)$, we get $rank(A) + rank(B) \leq n$. □

Let me state another theorem mainly to illustrate some useful proof techniques in linear algebra.

Theorem 14 *Let $A \in \mathcal{R}^{m \times n}$, and let $\mathcal{S}$ be a subspace of $\mathcal{R}^n$. Let us denote the image of A under $\mathcal{S}$ as*

$$A(\mathcal{S}) := \{Ax : x \in \mathcal{S}\}.$$

If $Null(A) \bigcap \mathcal{S} = 0$, then $dim(A(\mathcal{S})) = dim(\mathcal{S})$.

Proof: First, it is easy to see that $A(\mathcal{S})$ is a subspace. Therefore, there exists a bases of $\mathcal{S}$, say $x_1, \ldots, x_n$. Operating A on these bases vector, we get

$Ax_1, Ax_2, \ldots, Ax_n$. Now, the claim is that these set of vectors form the bases of $A(\mathcal{S})$. First, we prove that the above set of vectors spans $A(\mathcal{S})$. Consider any vector $y \in A(\mathcal{S})$. Since $y \in A(\mathcal{S})$, there exists a vector $x \in \mathcal{S}$ such that $y = Ax$. But $x := \sum_{i=1}^{n} \alpha_i x_i$ for some $\alpha_i' s \in \mathcal{R}$. This implies that $y = Ax = \sum_{i=1}^{n} \alpha_i A x_i$. Note this is a linear combination of $Ax_1, Ax_2, \ldots, Ax_n$. Further, any vector $y \in A(\mathcal{S})$ can be written in this form. Thus, $Ax_1, Ax_2, \ldots, Ax_n$ spans $A(\mathcal{S})$.

Next, we will show that this set of vectors are linearly independent. Consider the following linear combination:

$$\sum_{i=1}^{n} \beta_i A x_i = 0 \tag{2.5}$$

$$\Rightarrow \quad A \sum_{i=1}^{n} \beta_i x_i = A\bar{y} = 0 \tag{2.6}$$

for some $\bar{y} = \sum_{i=1}^{n} \beta_i x_i$. Thus, all β_i's are zeros provided the vector $\bar{y}$ is zero which happens only when the null space is zero. That is if $Null(A) \bigcap \mathcal{S} = 0$, then $\sum_{i=1}^{n} \beta_i x_i = 0$, which implies that $\beta_i = 0$ for all $i = 1, 2, \ldots, n$ by linear independency of $x_1, \ldots, x_n$. Thus, $Ax_1, \ldots, Ax_n$ are linearly independent, and therefore it forms a bases. □

For any given matrices $A \in \mathcal{R}^{m \times n}$ and $B \in \mathcal{R}^{n \times p}$, consider the following matrix:

$$M := \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \tag{2.7}$$

It is interesting to see what is the rank of M in terms of the ranks of A and B. We will state this result as a theorem below:

Theorem 15 *For any given two matrices $A \in \mathcal{R}^{m \times n}$ and $B \in \mathcal{R}^{n \times p}$, we have*

$$rank \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = rank(A) + rank(B). \tag{2.8}$$

Proof: Easy exercise. □

Next, let me illustrate the use of the above theorem. Consider the following:

$$rank \begin{pmatrix} I_n & 0 \\ 0 & AB \end{pmatrix} = n + rank(AB). \tag{2.9}$$

Now, supposing that we carry out a transformation of the matrix using a set of rank invariant transformation that results in a different matrix but with a similar structure as above, we get a new set of inequalities.[2] Let us try this on the above matrix itself.

$$\begin{pmatrix} I_n & 0 \\ 0 & AB \end{pmatrix} \rightarrow \begin{pmatrix} I_n & 0 \\ A & AB \end{pmatrix} \rightarrow \begin{pmatrix} I_n & -B \\ A & 0 \end{pmatrix} \rightarrow \begin{pmatrix} B & I_n \\ 0 & A \end{pmatrix} \tag{2.10}$$

Since the above is a rank invariant transformation, we have

$$rank \begin{pmatrix} I_n & 0 \\ 0 & AB \end{pmatrix} = n+rank(AB) = rank \begin{pmatrix} B & I_n \\ 0 & A \end{pmatrix} \geq rank(A)+rank(B). \tag{2.11}$$

This leads to the following theorem:

Theorem 16 *(Frobenius inequality) For any two matrices $A \in \mathcal{R}^{m\times n}$ and $B \in \mathcal{R}^{n\times p}$, we have the following rank inequality:*

$$n + rank(AB) \geq rank(A) + rank(B).$$

Now, you see how to prove some of the not so trivial rank inequalities. Let us see if we can give a more sophisticated inequality. Towards this consider

$$\begin{pmatrix} B & 0 \\ 0 & ABC \end{pmatrix}. \tag{2.12}$$

The rank of the above matrix is $rank(B)+rank(ABC)$. Let us do some elementary transformation on the above matrix as follows:

$$\begin{pmatrix} B & 0 \\ 0 & ABC \end{pmatrix} \rightarrow \begin{pmatrix} B & 0 \\ AB & ABC \end{pmatrix} \rightarrow \begin{pmatrix} B & -BC \\ AB & 0 \end{pmatrix} \rightarrow \begin{pmatrix} BC & B \\ 0 & AB \end{pmatrix} \tag{2.13}$$

[2] Can you figure out the transformation that is done below?

Thus, we have

$$\begin{aligned} rank\begin{pmatrix} B & 0 \\ 0 & ABC \end{pmatrix} &= rank(B) + rank(ABC) \\ &= rank\begin{pmatrix} BC & B \\ 0 & AB \end{pmatrix} \text{ (the last matrix in (2.13))} \\ &\geq rank(BC) + rank(AB). \end{aligned}$$

This is summarized in the following lemma:

Lemma 3 *Let* $A \in \mathcal{R}^{m\times n}$, $B \in \mathcal{R}^{n\times p}$, *and* $C \in \mathcal{R}^{p\times q}$. *Then,*

$$rank(B) + rank(ABC) \geq rank(BC) + rank(AB).$$

Lemma 4 *Let* $A \in \mathcal{R}^{m\times n}$. *Then, prove that*

$$rank(I_m - AA^T) - rank(I_n - A^T A) = m - n.$$

Proof: Consider the following matrix

$$M := \begin{pmatrix} I_m - AA^T & 0 \\ 0 & I_n \end{pmatrix}. \tag{2.14}$$

Note that $\text{rank}(M) = n + \text{rank}(I_m - AA^T)$. Now, we will carry out the following elementary transformation on the matrix M:

$$M := \begin{pmatrix} I_m - AA^T & 0 \\ 0 & I_n \end{pmatrix} \to \begin{pmatrix} I_m - AA^T & A \\ 0 & I_n \end{pmatrix}$$

$$\to \begin{pmatrix} I_m & A \\ A^T & I_n \end{pmatrix} \to \begin{pmatrix} I_m & 0 \\ A^T & I_n - A^T A \end{pmatrix} \to \begin{pmatrix} I_m & 0 \\ 0 & I_n - A^T A \end{pmatrix} := N.$$

Note that the matrix N has the same rank as that of M, and therefore $\mathrm{rank}(M) = n + \mathrm{rank}(I_m - AA^T) = \mathrm{rank}(N) = m + \mathrm{rank}(I_n - A^T A)$. This completes the proof. □

The following section can be skipped in the first reading.

2.3 Isomorphisms and Homomorphisms

Consider two vector spaces U and W both of finite dimensions over the same field $\mathbb{F}$. Note that till now we have been considering a real field. However, extending the study of linear operators to any other field is not difficult.

Now, let us consider the special case of the dimensions of U and W being equal to n. By the existence of bases theorem, there are two sets of bases vectors $B_1 := \{u_1, \ldots, u_n\}$ and $B_2 := \{w_1, \ldots, w_n\}$ of U and W, respectively. Now, one can define a map as follows:

$$f : U \to W \tag{2.15}$$

such that

- it maps bases to bases, i.e., $f(u_i) := w_i$, $i = 1, 2, \ldots, n$, and
- it preserves the structure of a vector space, i.e., $f(\alpha u_1 + \beta u_2) := \alpha f(u_1) + \beta f(u_2)$, $u_1, u_2 \in U$ for any $\alpha, \beta \in \mathbb{F}$.

With the above map, consider any vector $u \in U$. This can be written as $u := \sum_{i=1}^{n} \alpha_i u_i$ for some $\alpha_i' s \in \mathbb{F}$. With this it is easy to see that the inverse of the map exists, which is explained as follows.

For any vector $w \in W$, we have

$$\begin{aligned} w &:= \sum_{i=1}^{n} \beta_i w_i && (2.16)\\ &= \sum_{i=1}^{n} \beta_i f(u_i) && (2.17)\\ &= f\left(\sum_{i=1}^{n} \beta_i u_i\right) && (2.18)\\ &= f(u), && (2.19) \end{aligned}$$

where $u := \sum_{i=1}^{n} \alpha_i u_i \in U$. In other words, given any vector in $w \in W$, there is a corresponding vector $u \in U$ that the function maps to. This implies that the map is surjective! Is the map one-one? For any two vectors u and $\bar{u}$ in U, $f(u) = f(\bar{u})$ implies $u = \bar{u}$ (prove this!). This implies that the map in one-one. Thus, the map is bijective. It is not just bijective but also preserves the structure of the spaces. To put it differently, all that the map f is doing is to in some sense relabel the vectors in U. The key property that preserves the structure is the second property of the map. We give a name to such mappings, which is defined as follows for two vector spaces U and W.

Definition A map $f : U \rightarrow W$ is said to be an isomorphism if

- it is one-one and surjective, and
- $f(\alpha u_1 + \beta u_2) := \alpha f(u_1) + \beta f(u_2)$, $u_1, u_2 \in U$ for any $\alpha, \beta \in \mathbb{F}$.

We say that the two vector spaces U and W are isomorphic if there exist a map from U to W that is an isomorphism.

Exercise: Prove that any finite dimensional vector space V of dimension n over a field $\mathbb{F}$ is isomorphic to $\mathbb{F}^n$.

Exercise: Prove that if an isomorphism exists between two finite dimensional vector spaces $\mathbb{F}^n$ and $\mathbb{F}^m$ over a field $\mathbb{F}$, then $m = n$.

The concept of isomorphism helps us to visualize any finite dimensional vector space as a bunch of elements contained in the field over which the space is defined. Note that the above definition of isomorphism relies on the fact that the operator is bijective. However, in most cases, this may not be true. Therefore, relaxing the definition by removing the condition of the map being bijective results in the following.

Definition (Homomorphism) A map $f : U \rightarrow W$ is said to be a homomorphism if $f(\alpha u_1 + \beta u_2) := \alpha f(u_1) + \beta f(u_2)$, $u_1, u_2 \in U$ for any $\alpha, \beta \in \mathbb{F}$.

Note that the above preserves algebraic structure. Also, if a map is isomorphic, then it is also a homomorphic. Let us denote the set of all homomorphisms from U into V by Hom(U,V). In the following section, we shall study more about this set.

2.4 Dual Space

Intuitively, one possible way to learn more about a vector space under consideration is to take an operator and operate on the vector space and see the result. In some sense, each operator will give us different information about the space. If we have enough number of such operators, we expect that we should be able to say a lot about the space. Also, it provides a convenient tool where one can deduce the property of the space by studying its operators provided the operators are more amenable to analysis.

Consider for instance the set Hom(U,V). Now, we shall see that this can be given a vector space structure. In order to do so, we should define the binary operator $+$ over it. The "plus" is defined as $(T_1 + T_2)(u) := T_1(u) + T_2(u)$ for all $u \in U$, and for all $T_1, T_2 \in$ Hom(U,V). Let us define the scalar multiplication as $(\alpha T)(u) := \alpha T(u)$ for all $T \in$ Hom(U,V), and $u \in U$. With this definition, it is easy to see that Hom(U,V) is a vector space over $\mathbb{F}$, which is stated as a theorem below.

Theorem 17 *The set Hom(U,V) is a vector space over $\mathbb{F}$ under the binary and scalar operations defined above.*

Since Hom(U,V) is a vector space, a natural thing to do is to construct a bases for it. In order to understand the construction of bases, we will restrict to the following special cases of U and V.

Let the bases of U and V be $\{u_1, u_2, u_3\}$, and $\{v_1, v_2\}$, respectively. Now, we say that $T_1, \ldots, T_N$ is a bases of Hom(U,V), if for all $T \in$ Hom(U,V), we have

$$T = \sum_{i=1}^{N} \alpha_i T_i,$$

and $T_1, \ldots, T_N$ is a linearly independent set. This means that

$$Tu = (\sum_{i=1}^{N} \alpha_i T_i)u = \sum_{i=1}^{N} \alpha_i T_i u$$

for all $u \in U$, and

$$\sum_{i=1}^{N} \beta_i T_i u = 0$$

for all $u \in U$ implies $\beta_i = 0$ for all $i = 1, 2, \ldots, N$. Now, we will construct a set of bases that spans Hom(U,V). Let $B := \{T_1, \ldots, T_N\}$, and see what is it that is required for this to be a basis. First of all, we need

- $Tu_i \subseteq span\{B\}$, $i = 1, 2, 3$ for all $T \in$ Hom(U,V), and
- B should be linearly independent.

Now, let us investigate the first requirement for $i = 1, 2, 3$ as follows.

- $T \in$ Hom(U,V), $Tu_1 \subseteq span\{B\}$, which can be written as

$$Tu_1 = \sum_{i=1}^{2} \beta_{1i} v_i = \sum_{i=1}^{N} \alpha_{1i} T_i u_1. \tag{2.20}$$

  Now, the equality above is possible if $T_1 u_1 = v_1$, $T_2 u_1 = v_2$, $T_i u_1 = 0$ for all $i = 3, \ldots, N$, and $\beta_{11} = \alpha_{11}$, $\beta_{12} = \alpha_{12}$.

- $T \in$ Hom(U,V), $Tu_2 \subseteq span\{B\}$, which can be written as

$$Tu_2 = \sum_{i=1}^{2} \beta_{2i} v_i = \sum_{i=1}^{N} \alpha_{2i} T_i u_2. \tag{2.21}$$

  Now, the equality above is possible if $T_3 u_2 = v_1$, $T_4 u_2 = v_2$, $T_i u_2 = 0$ for all $i = 5, \ldots, N$, and $\beta_{21} = \alpha_{21}$, $\beta_{22} = \alpha_{22}$.

- $T \in$ Hom(U,V), $Tu_3 \subseteq span\{B\}$, which can be written as

$$Tu_3 = \sum_{i=1}^{2} \beta_{3i} v_i = \sum_{i=1}^{N} \alpha_{3i} T_i u_3. \tag{2.22}$$

  Now, the equality above is possible if $T_5 u_3 = v_1$, $T_6 u_3 = v_2$, $T_i u_3 = 0$ for all $i = 7, \ldots, N$, and $\beta_{31} = \alpha_{31}$, $\beta_{32} = \alpha_{32}$.

From the above, it is easy to see that $T_1, \ldots, T_6$ is sufficient to span Hom(U,V). Therefore, let $N = 6$. Now, from the above discussion, let us recall the conditions that are required for the set B with $N = 6$ to be a bases:

- $T_1 u_1 = v_1$, and $T_1 u_1 = 0$,
- $T_2 u_1 = v_2$, and $T_2 u_1 = 0$,
- $T_3 u_2 = v_1$, and $T_3 u_2 = 0$,
- $T_4 u_2 = v_2$, and $T_4 u_2 = 0$,
- $T_5 u_3 = v_1$, and $T_5 u_3 = 0$,
- $T_6 u_2 = v_2$, and $T_6 u_3 = 0$.

Thus, assuming the above conditions on B, it is easy to see that it spans Hom(U,V). We will now investigate the linear independence of $T_1, \ldots, T_6$. Here, we need to prove that $\sum_{i=1}^{6} \beta_i T_i u = 0$ for all $u \in U$ implies $\beta_i = 0$ for all $i = 1, 2, \ldots, 6$. Pick $u = u_1$, then, from the conditions on $T_1, \ldots, T_6$, we have

$$\sum_{i=1}^{6} \beta_i T_i u_1 = 0 \Rightarrow \beta_1 v_1 + \beta_2 v_2 = 0. \tag{2.23}$$

By linear independency of v_1, v_2, we have $\beta_1 = \beta_2 = 0$. Thus, the above equation becomes

$$\sum_{i=3}^{6} \beta_i T_i u = 0,$$

for all $u \in U$. Now, pick $u = u_2$. Then,

$$\sum_{i=1}^{6} \beta_i T_i u_2 = 0 \Rightarrow \beta_3 v_1 + \beta_4 v_2 = 0.$$

This implies that $\beta_3 = \beta_4 = 0$. Continuing this process further, it is easy to see that all $\beta_i' s$ have to be zero. This proves linear independency. Thus, $T_1, \ldots, T_6$ forms a basis of Hom(U,V) for $\dim(U) = 3$, and $\dim(V) = 2$. Note that the number $6 = 3 \times 2$ is the product of the dimensions of each vector spaces. This can be easily generalized to any U and V of finite dimensions, which is the essence of the following theorem.

Theorem 18 *Let U and V be two vector spaces of dimensions m and n, respectively. Then, Hom(U,V) is a vector space of dimension mn*

Proof: Exercise. *Hint*: Try to imitate the proof for $m = 3$, and $n = 2$ case described above.

As an important corollary, we have:

Corollary 2 *Let V be a vector space over $\mathbb{F}$ of dimension n. Then, $\mathsf{Hom}(V, \mathbb{F})$ is a vector space of dimension n. Further, V and $\mathsf{Hom}(V, \mathbb{F})$ are isomorphic to each other.*

Proof: From Theorem 18, the dimension of $\mathsf{Hom}(V, \mathbb{F})$ is equal to n. The isomorphism follows from a previous exercise. $\square$

Now, let us investigate the above corollary even further. Any vector $v \in V$ can be written as $v = \sum_{i=1}^{n} \alpha_i v_i$, where $v_1, \ldots, v_n$ is a bases vector of V. Let $F \in \mathsf{Hom}(V, \mathbb{F})$. Consider

$$F(v) = F \sum_{i=1}^{n} \alpha_i v_i \tag{2.24}$$

$$= \sum_{i=1}^{n} \alpha_i F(v_i) \tag{2.25}$$

$$= \Phi \bullet \mathbf{F}, \tag{2.26}$$

where $\Phi := (\alpha_1, \ldots, \alpha_n)$, $\mathbf{F} := (F(v_1), \ldots, F(v_n))^T$, and $\bullet$ represents the usual dot product on $\mathcal{R}^n$. Note that the vector $\mathbf{F}$ is fixed for a given bases. Also, it is easy to check that this representation is unique given the bases vectors. Thus, not only that all the linear operators in $\mathsf{Hom}(V, \mathbb{F})$ can be written as a dot product but also there is a one-one correspondence between the elements of V and $\mathsf{Hom}(V, \mathbb{F})$.

Therefore, it is interesting to see if this can be generalized even further. This requires us to generalize the notions of the dot product, which is done in the next chapter. The space $\mathsf{Hom}(V, \mathbb{F})$ is special in linear algebra, and has a name to it. It is called the dual space, as it behaves like V, but the objects appears to be completely different.

Definition (Dual space) If V is a vector space over $\mathbb{F}$, then its dual space is $\mathsf{Hom}(V, \mathbb{F})$. The elements of the dual space are called *linear functionals*.

In the following chapter, we will generalize the notions of dot product and the distance to an arbitrary finite dimensional vector spaces.

Chapter 3

Inner Product and Normed Spaces

Consider two vectors $x := (x_1, x_2)$ and $y := (y_1, y_2)$ in $\mathcal{R}^2$. The inner-product is defined as

$$\langle x, y \rangle := x_1 y_1 + x_2 y_2. \tag{3.1}$$

Now, let us investigate the properties of $\langle x, y \rangle$, which are listed below:

1. $\langle x, x \rangle \geq 0$, for all $x \in \mathcal{R}^n$, and $\langle x, x \rangle = 0$ if and only if $x = 0$.

2. $\langle \alpha x_1 + \beta x_2, y \rangle := \alpha \langle x_1, y \rangle + \beta \langle x_2, y \rangle$ for all $\alpha, \beta \in \mathcal{R}$.

3. $\langle x, y \rangle = \langle y, x \rangle$.

In the case where the underlying field is complex, the above definition of the dot product of two vectors $x := (x_1, x_2)$ and $y := (y_1, y_2)$ in $\mathcal{C}^2$ is modified as

$$\langle x, y \rangle := x_1 \bar{y}_1 + x_2 \bar{y}_2. \tag{3.2}$$

Now, let us investigate the properties of $\langle x, y \rangle$, which are listed below:

1. $\langle x, x \rangle \geq 0$, for all $x \in \mathcal{C}^n$, and $\langle x, x \rangle = 0$ if and only if $x = 0$.

2. $\langle \alpha x_1 + \beta x_2, y \rangle := \bar{\alpha} \langle x_1, y \rangle + \bar{\beta} \langle x_2, y \rangle$ for all $\alpha, \beta \in \mathcal{C}$.

3. $\langle x, y \rangle = \overline{\langle y, x \rangle}$.

Consider two vectors $x := (x_1, x_2)$ and $y := (y_1, y_2)$ in $\mathcal{R}^2$. The distance between x and y denoted $dist(x, y)$ is calculated using the following formula.

$$dist(x, y) := \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}. \tag{3.3}$$

From this the length of any vector x is defined as $len(x) := dist(0, x)$. It is easy to see that $len(*) : \mathcal{R}^2 \to \mathcal{R}^+$ satisfies the following properties:

1. $len(x) \geq 0$ for all $x \in \mathcal{R}^2$ with equality if and only if $x = 0$.

2. $len(x) + len(y) \geq len(x + y)$ for all $x, y \in \mathcal{R}^2$.

3. $len(\alpha x) = |\alpha| len(x)$, for all $x \in \mathcal{R}^2$.

The above can be easily generalized to $\mathcal{R}^n$ as follows. The function $len(*) : \mathcal{R}^n \to \mathcal{R}^+$ is a length function on $\mathcal{R}^n$ if it satisfies the following properties:

1. $len(x) \geq 0$ for all $x \in \mathcal{R}^n$ with equality if and only if $x = 0$.

2. $len(x) + len(y) \geq len(x + y)$ for all $x, y \in \mathcal{R}^n$.

3. $len(\alpha x) = |\alpha| len(x)$, for all $x \in \mathcal{R}^n$.

Generalizing the notions of distance and inner product to an arbitrary vector space over a real or a complex field is done below.

Definition (Norm) A function on the vector space V over $\mathbb{F}$ denoted $\| * \| : V \to \mathcal{R}^+$ is said to be a norm if it satisfies the following properties:

1. $\|x\| \geq 0$ for all $x \in \mathcal{R}^n$ with equality if and only if $x = 0$.

2. $\|x\| + \|y\| \geq \|x + y\|$ for all $x, y \in \mathcal{R}^n$.

3. $\|\alpha x\| = |\alpha| \|x\|$, for all $x \in \mathcal{R}^n$.

Definition (Inner-product) A function on the vector space V over $\mathbb{F}$ ($\mathbb{F} = \mathcal{R}$ or $\mathcal{C}$) denoted $\langle *, * \rangle : V \times V \to \mathbb{F}$ is said to be an inner-product if it satisfies the following properties:

1. $\langle x, x \rangle \geq 0$, for all $x \in V$, and $\langle x, x \rangle = 0$ if and only if $x = 0$.

2. $\langle \alpha x_1 + \beta x_2, y\rangle := \bar{\alpha} \langle x_1, y\rangle + \bar{\beta} \langle x_2, y\rangle$ for all $\alpha, \beta \in \mathbb{F}$.

3. $\langle x, y\rangle = \overline{\langle y, x\rangle}$.

Note that in the above definition, we have used the fact that the conjugate of any real number is the number itself. Recall that the dot product in $\mathcal{R}^n$ can be used to measure the distance between two vectors x and y by simply taking the dot product of the difference, i.e., $(x - y) \bullet (x - y)$. In other words, the dot product induces the Euclidean distance notion in the real space. Now, consider a general inner product space $(V, \langle *, *\rangle)$. A natural question to ask is if $\sqrt{\langle *, *\rangle}$ is a norm. Let us check all the properties of the norm.

The first property of the norm directly follows from the definition of the inner product. Now, let us check for the triangle inequality. Let $x, y \in V$. Then, we need to prove the following:

$$\|x + y\|^2 \leq \|x\|^2 + \|y\|^2, \tag{3.4}$$

which is equivalent to proving the following:

$$\langle x + y, x + y\rangle^2 \leq \langle x, x\rangle^2 + \langle y, y\rangle^2 . \tag{3.5}$$

Using the definitions of the inner product, the right hand side in the above equation can be simplified as follows

$$\begin{aligned} \langle x + y, x + y\rangle^2 &= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle^2 & (3.6)\\ &= \|x\|^2 + 2\langle x, y\rangle + \|y\|^2. & (3.7) \end{aligned}$$

The proof would be complete if $\langle x, y\rangle \leq \|x\|\|y\|$ is true. This is because

$$\begin{aligned} \|x\|^2 + 2\langle x, y\rangle + \|y\|^2 &\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 & (3.8)\\ &= (\|x\| + \|y\|)^2. & (3.9) \end{aligned}$$

Now, the goal is to check if $\langle x, y\rangle \leq \|x\|\|y\|$ is true. This is the famous *Cauchy-Schwartz* inequality.

Theorem 19 *Let* $(V, \langle *, * \rangle)$ *be an inner product space. Then, for all* $x, y \in V$*, we have*

$$\langle x, y \rangle \leq \|x\| \|y\|. \tag{3.10}$$

Proof: Let us first assume that x and y are unit norm vectors. Then, we need to prove that $\langle x, y \rangle \leq 1$. Note that there is nothing to prove if one of the vector is a zero vector. The inequality $\langle x, y \rangle \leq 1$ can be proved as follows:

$$\begin{aligned} 0 &\leq \langle x - y, x - y \rangle & (3.11) \\ &= \|x\| + \|y\| - 2 \langle x, y \rangle & (3.12) \\ &= 2 - 2 \langle x, y \rangle & (3.13) \\ &\Rightarrow \langle x, y \rangle \geq 1. & (3.14) \end{aligned}$$

Consider any pair x, y in V, not necessarily unit vectors. Consider $\bar{x} := \frac{x}{\|x\|}$ and $\bar{y} := \frac{y}{\|y\|}$. Since $\bar{x}$ and $\bar{y}$ are unit vectors, using $\langle \bar{x}, \bar{y} \rangle \leq 1$ for unit vectors, we get

$$\langle x, y \rangle \leq \|x\| \|y\|. \ \square \tag{3.15}$$

Now, we know that in any inner product space, the inner product induces a norm. One natural question to ask is if we can say whether a norm is induced by an inner product or not? The answer is the following:

Theorem 20 *Let* $(V, \langle *, * \rangle)$ *be an inner product space. The norm* $(\| * \|)$ *is induced by the inner product if and only if it satisfies the following parallelogram identity:*

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2) \tag{3.16}$$

for all $x, y \in V$*.*

Proof: If norm is induced by an inner product, then the above equation is valid, which is easy to verify. However, the proof of the converse is non-trivial and is left as a reading exercise. $\square$

3.1 Orthogonality and Orthogonal Projection

Inner product not only induces a norm that brings in the notion of length but also enables us to talk about the angles between two vectors. For example, in $\mathcal{R}^2$ with the usual dot product, the angle between the two vectors $x := (x_1, x_2)$ and $y := (y_1, y_2)$ is measured using

$$\text{angle} := \cos^{-1} \frac{x \bullet y}{len(x)len(y)}. \tag{3.17}$$

Two vectors are said to be orthogonal if $\cos\theta = 0$, which implies that $x \bullet y = 0$. Now, we generalize this in the following definition.

Definition Let $(V, \langle *, * \rangle)$ be an inner product space. We say that $x, y \in V$ are orthogonal if and only if $\langle x, y \rangle = 0$.

In this chapter, it is understood that the vector space is an inner product space. Consider a vector space V. Let $x \in V$ be a vector. Then, consider the following set:

$$A_x^{\perp} := \{y \in V : \langle x, y \rangle = 0\}. \tag{3.18}$$

It is easy to see that the set $A_x^{\perp}$ is a subspace. Now consider any vector $v \in V$. Intuitively, we see that the difference $w := v - \alpha x$ should be orthogonal to the $\text{span}\{x\}$ for some properly chosen $\alpha \in \mathcal{R}$. Now, let us investigate the value of α for which this is true. We need to check if $\langle v - \alpha x, x \rangle = 0$:

$$\begin{aligned} 0 &= \langle v - \alpha x, \alpha x \rangle & (3.19) \\ &= \alpha \langle v, x \rangle - \alpha^2 \langle x, x \rangle & (3.20) \\ \Leftrightarrow \alpha^* &= \frac{\langle v, x \rangle}{\langle x, x \rangle}. & (3.21) \end{aligned}$$

Thus, $w = v - \alpha^* x \in A_x^{\perp}$, which implies that $v = \alpha^* x + w$. This can interpreted as any vector in V can be written as the sum of a vector in the span of x and a vector in the orthogonal compliment of the span of x.

Now, we shall see if we can generalize this to an arbitrary subspace of V. Towards, this we need the notion of projection of a vector onto a subspace. First, let us look at the projection of one vector, say $u \in V$ onto another vector, say $v \in V$, denoted $\mathsf{Proj}_v(u)$. This makes sense only when the two vectors are different in the sense that $\mathsf{span}\{u\} \neq \mathsf{span}\{v\}$. Our intuition in $\mathcal{R}^2$ suggests the following simple definition:[1]

$$\mathsf{Proj}_v(u) := \frac{\langle u, v\rangle}{\langle u, u\rangle} u. \tag{3.22}$$

Now, it is easy to see that the vector $\bar{u} := \mathsf{Proj}_v(u) - u$ is orthogonal to v. Thus, beginning with two vectors, we have found two different vectors $\bar{u}$ and v that are orthogonal to each other, and they span the same space as that of v and u. General version of this method is called Gram-Schmidt orthogonalization procedure, which is explained below.

3.1.1 Gram-Schmidt Orthogonalization

Consider a vector space V with a set of bases $\{v_1, \ldots, v_n\}$. Now, the procedure is as follows:

- Let $\bar{v}_1 := v_1$.
- Obtain an orthogonal vector $\bar{v}_2$ by using the vector v_2 as

$$\bar{v}_2 := v_2 - \mathsf{Proj}_{\bar{v}_1}(v_2). \tag{3.23}$$

- The third vector $\bar{v}_3$ which is orthogonal to both $\bar{v}_1$ and $\bar{v}_2$ is obtained as

$$\bar{v}_3 := v_3 - \mathsf{Proj}_{\bar{v}_1}(v_3) - \mathsf{Proj}_{\bar{v}_2}(v_3). \tag{3.24}$$

- ....
- ....
- ....

[1] An interesting observation is that the vector $\alpha^* x = \frac{\langle v,x\rangle}{\langle x,x\rangle} x$ is a projection of v onto x.

- The i-th vector, $i = 1, 2, \ldots, n$ is obtained as follows:

$$\bar{v}_i := v_i - \sum_{j=1}^{i-1} \mathsf{Proj}_{\bar{v}_j}(v_i). \tag{3.25}$$

It is easy to see that the above procedure leads to a set of orthogonal vectors starting from a set of bases vectors. The following theorem states that the orthogonal vector thus obtained still retains the bases property.

Theorem 21 *Let the bases vector of a vector space V be $\{v_1, \ldots, v_n\}$. Let $\{\bar{v}_1, \ldots, \bar{v}_n\}$ be the corresponding orthogonal vectors obtained by applying Gram-Schmidt orthogonalization procedure. Then, $\{\bar{v}_1, \ldots, \bar{v}_n\}$ is also a bases.*

Proof: First, let us prove linear independency of $\{\bar{v}_1, \ldots, \bar{v}_n\}$, i.e., we need to prove that

$$\sum_{i=1}^{n} \alpha_i \bar{v}_i = 0$$

implies $\alpha_i = 0$ for all $1 \leq i \leq n$. Taking the inner product of the above equation with $\bar{v}_i$, we get $\alpha_i \langle \bar{v}_i, \bar{v}_i \rangle = 0$. Since $\langle \bar{v}_i, \bar{v}_i \rangle \geq 0$, we have $\alpha_i = 0$, and this is true for all $1 \leq i \leq n$. Thus, all of α_i's $= 0$. This proves linear independency. Now, since the vector space is of dimension n, and the set $\{\bar{v}_1, \ldots, \bar{v}_n\}$ has n linearly independent vectors, it should span the space, which proves the theorem. □

As a simple consequence of the above, we have:

Lemma 5 *Let $\{\bar{v}_1, \ldots, \bar{v}_n\}$ be an orthogonal bases vector. Then, $\{\tilde{v}_1, \ldots, \tilde{v}_n\}$, where*

$$\tilde{v}_i := \frac{\bar{v}_i}{\|\bar{v}_i\|}$$

is an orthonormal bases.

Proof: Easy exercise. □

Exercise Let V be a vector space, and let W be a subspace. Prove that any vector $v \in V$ can be written as $v = w + w^{\perp}$, for some $w \in W$ and $w^{\perp} \in W^{\perp}$.

It is easy to see that the inner product is a linear map on the vector space. In fact, we have seen in the previous chapter that it is called the linear functionals. In the following theorem, we show that other than the inner products, there are no other linear functionals.

Theorem 22 ***(Riesz-Representation theorem)*** *Given any linear map $T : V \longrightarrow \mathbb{F}$, there exists a unique $y \in V$ such that*

$$Tx = \langle x, y \rangle \tag{3.26}$$

for all $x \in v$.

Proof: Suppose that $Tx = 0$ for all $x \in V$, then choose $y \equiv 0$. Let us assume that there exists an $x \neq 0$ such that $Tx \neq 0$. Consider,

$$N := \{z \in V : Tz = 0\}.$$

Clearly, N is a subspace (why?). If at all there exists a representation as in (3.26), then $y \perp N$, i.e.,

$$\langle z, y \rangle = 0$$

for all $z \in N$. So, we have to search for a $y \in N^{\perp}$. Towards this, choose any $x \in V$ and $Tx \neq 0$. Consider,

$$\bar{z} := (Ty)x - (Tx)y.$$

Since $T\bar{z} = 0$, we have $\bar{z} \in N$. Thus, we need to choose a y such that $\langle \bar{z}, y \rangle = 0$ for all $x \in V$. Substituting for $\bar{z}$ in $\langle \bar{z}, y \rangle = 0$, we get

$$\langle (Ty)x - (Tx)y, y \rangle = 0 \tag{3.27}$$

$$\Rightarrow Tx = \frac{Ty}{\langle y, y \rangle} \langle x, y \rangle . \tag{3.28}$$

The proof is complete by choosing $\frac{Ty}{\langle y,y \rangle} y$ as our new y. □

Remark: It is worthwhile noting that the above proof is general and can be easily extended to infinite dimensional Hilbert spaces.

3.2 Norms on Linear Transformations

In the previous section, we studied the norm of a vector in a vector space. The main motivation for studying norms on linear transformations is justified by the following simple example. Let us consider a simple model which typically arises in communication and signal processing problems; the one of estimating the unknown matrix from a given set of noisy measurements:

$$Y = HS + W, \tag{3.29}$$

where all the matrices above belong to $\mathcal{C}^{n\times n}$. Given Y and H, one is interested in finding S. This can be posed as the following optimization problem:

$$\min_{X\in\mathcal{C}^{n\times n}} \|Y - HX\|^2, \tag{3.30}$$

where $\|*\|$ is a norm on the set of matrices. The problem above naturally demands for a distance notion on the set of matrices, which is the main topic of study of this section.

3.2.1 Bounded Linear Functions

In this subsection, we shall use the definition of a norm of a vector to define a norm on linear functions. One way to measure the length of a linear transformation is to see the magnitude of the amplification that T results in relative to the length of the vector. This is possible only if the length of the vector is finite. This motivates the following definition.

Definition Let $T : V \to W$ be a linear map from the vector space $(V, \|\|_V)$ to a vector space $(W, \|\|_W)$. We say that the map T is bounded if there exists $M < \infty$ such that

$$\|Tx\| \leq M\|x\|$$

for all $x \in V$.

Note that the set of matrices are bounded linear transformations. Now, consider the following set

$$N_L := \left\{ \frac{\|Tx\|}{\|x\|} : x \in V \right\}. \tag{3.31}$$

In the above, we have dropped the subscript in the norm definition as it is evident from the context. It is easy to see that $0 \leq \frac{\|Tx\|}{\|x\|} \leq M < \infty$. Therefore, the set N_L is bounded, which implies that there exists a supremum element. Thus, we define the supremum of N_L as the "induced norm" of T:

Definition The induced norm or the operator norm of a linear transformation $T : V \to W$ is defined as

$$\|T\| := \sup_{x \in V} \frac{\|Tx\|}{\|x\|}. \tag{3.32}$$

Theorem 23 *The function of $T : V \to W$ defined by*

$$\|T\| := \sup_{x \in V} \frac{\|Tx\|}{\|x\|} \tag{3.33}$$

is a norm.

Proof: It follows from the definition of the norm on the vector space that $\|T\| \geq 0$. Suppose if $\|T\| = 0$, then

$$\sup_{x \in V} \frac{\|Tx\|}{\|x\|} = 0 \Rightarrow \frac{\|Tx\|}{\|x\|} = 0 \tag{3.34}$$

for all $x \in V$, which is possible only when $Tx = 0$ for all $x \in V$. This implies that $T = 0$. If $T = 0$, then $\|T\| = 0$ (trivial!). Consider any two linear transformations

T_1 and T_2. Then,

$$\begin{aligned}
\|T_1 + T_2\| &:= \sup_{x \in V} \frac{\|(T_1 + T_2)x\|}{\|x\|} && (3.35)\\
&= \sup_{x \in V} \frac{\|T_1 x + T_2 x\|}{\|x\|} && (3.36)\\
&\leq \sup_{x \in V} \frac{\|T_1 x\| + \|T_2 x\|}{\|x\|} && (3.37)\\
&\leq \sup_{x \in V} \frac{\|T_1 x\|}{\|x\|} + \sup_{x \in V} \frac{\|T_2 x\|}{\|x\|} \text{ (why?)} && (3.38)\\
&= \|T_1\| + \|T_2\|, && (3.39)
\end{aligned}$$

which verifies the triangle inequality. □