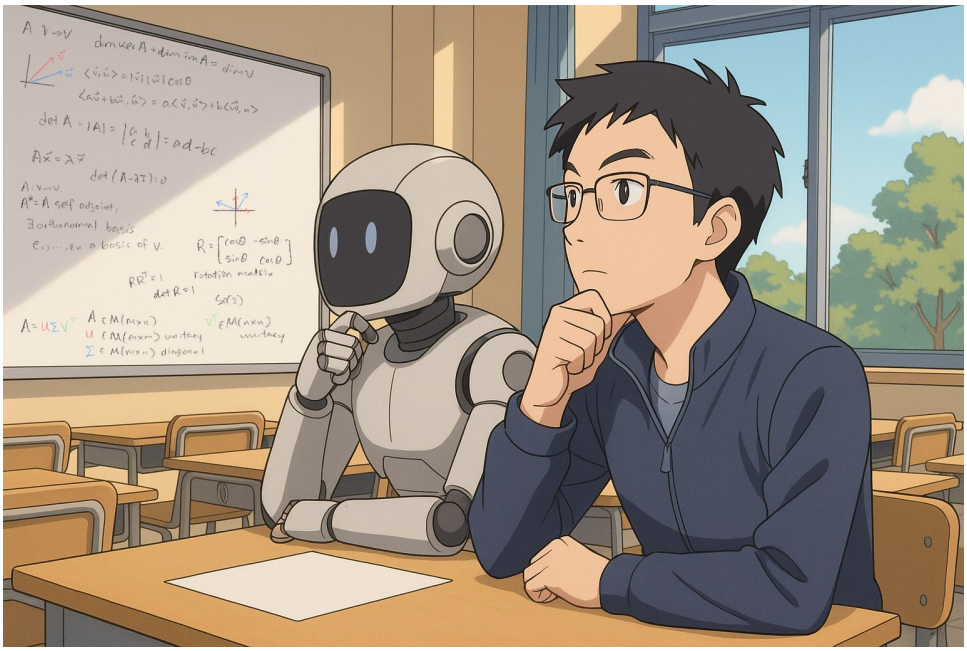
NOTES ON LINEAR ALGEBRA

WRITTEN BY

TIANYI WANG



PUBLISHED

IN THE WILD

Contents

Preface

This is my second LaTeXnotes project and the subject is linear algebra. Many people suggested me to first study linear algebra before entering analysis in $\mathbb{R}^k$, which I think is a good idea.

The text is *Linear Algebra Done Right* by Sheldon Axler (3rd edition). My notes will be based on this text and Professor Jason Morton's lecture on linear algebra at Penn State. Meanwhile, I may refer to Axler's own video lectures, which serves as a supplement of his book, and Gilbert Strang's 18.06 lecture at MIT.

My first encounter with linear algebra has a heavily applied focus. Though my purpose is quite different now, I had a very happy time learning linear algebra, thanks to Professor Gilbert Strang and his 18.06. This notes is also for him, as he is one of the best teachers I have ever met.

This notes will be much more concise compared with my analysis notes. The content will focus heavily on theoretical components, instead of applied linear algebra. This is my second attempt to study serious mathematics, wish me good luck.

Tianyi
March 11, 2020

Chapter 1

Vector Spaces

You can buy the textbook from me for a discount and I'll even personally sign it. And if you pay a bit more I won't sign it.

Gilbert Strang

In this section we will talk about vector spaces over fields and subspaces. The reader should already be familiar with the n dimensional real and complex spaces $\mathbb{R}^n$ and $\mathbb{C}^n$. Readers with no such experience should refer to my *Notes on Mathematical Analysis* and section 1.A in Axler's book.

Although the standard notations are $\mathbb{R}$ and $\mathbb{C}$ and $\mathbb{F}$, we will mostly use the bold letter $\mathbf{R}$ and $\mathbf{C}$ and $\mathbf{F}$ to represent the same thing (Axler use the latter).

Notation 1.0.1. Throughout this notes, $\mathbf{F}$ stands for either $\mathbf{R}$ or $\mathbf{C}$, since they are examples of fields. Thus if we prove a theorem involving $\mathbf{F}$, we know it will hold if $\mathbf{F}$ is replaced by either $\mathbf{R}$ or $\mathbf{C}$.

Definition 1.0.1. $\mathbf{F}^n$ is the set of all lists of length n of elements of $\mathbb{F}$. In other words,

$$\mathbf{F}^n = \left\{(x_1, ..., x_n) : x_j \in \mathbf{F} \text{ for } j = 1, ...n\right\}.$$

A *field* is a set containing at least two distinct elements called 0 and 1, along with operations of addition and multiplication satisfying the field axioms (refer to my analysis note). We often denote such field by $(F, +, \times)$. If we write $(F, +, \times, <)$, it means that we are talking about an *ordered field*.

1.1 Definition of Vector Spaces

The motivation for the definition of a vector space comes from the properties of addition and scalar multiplication in $\mathbf{F}^n$.

Definition 1.1.1. (Addition) An addition on a set V is a function $+ : V \times V \to V$ that assigns an element $u + v \in V$ to each pair of elements $u, v \in V$.

Definition 1.1.2. (Scalar Multiplication) A scalar multiplication on a set V is a function $\times : \mathbf{F} \times V \to V$ that assigns an element $\lambda v \in V$ to each $\lambda \in \mathbf{F}$ and each $v \in V$.

Definition 1.1.3. (Vector Space) A vector space is a set V along with an addition and multiplication on V, defined as above, such that the following properties hold:

1. **commutativity**: $u + v = v + u$ for all $u, v \in V$;
2. **associativity**: $(u+v)+w = u+(v+w)$ and $(ab)v = a(bv)$ for all $u, v, w \in V$ and $a, b \in \mathbf{F}$;
3. **additive identity**: there exists an element $\mathbf{0} \in V$ such that $v + \mathbf{0} = v = \mathbf{0} + v$ for all $v \in V$;
4. **additive inverse**: for every $v \in V$, there exists $w \in V$ such that $v + w = 0$;
5. **multiplicative identity**: $\mathbf{1_F} v = v$ for all $v \in V$, where $\mathbf{1_F}$ is the multiplicative identity in $\mathbf{F}$;
6. **distributive properties**: $a(u + v) = au + av$ and $(a + b)v = av + bv$ for all $a, b \in \mathbf{F}$ and for all $u, v \in V$.

Remark 1.1.1. Elements of a vector space are called vectors or points. There is really no difference. But they are used in different context.

The scalar multiplication in a vector space depends on $\mathbf{F}$. Thus when we need to be precise, we will say that V *is a vector space over* $\mathbf{F}$. Example: $\mathbf{R}^n$ is a vector space over $\mathbf{R}$ and $\mathbf{C}^n$ is a vector space over $\mathbf{C}$. A vector space over $\mathbf{R}$ is called a *real vector space*. Likewise, a vector space over $\mathbf{C}$ is called a *complex vector space*.

Example 1.1.1. Let $\mathbf{F}^\infty$ be defined as the set of all sequences of elements of $\mathbf{F}$(or $\mathbb{F}$):

$$\mathbf{F}^\infty = \{(x_1, x_2, \ldots) : x_j \in \mathbf{F} \text{ for } j = 1, 2, \ldots\}.$$

Addition and multiplication are defined just as you expect (i.e. pointwise addition and multiplication). Then $\mathbf{F}^\infty$ becomes a vector space over $\mathbf{F}$.

Definition 1.1.4. If S is a set, then $\mathbf{F}^S$ denotes the set of functions $f : S \to \mathbf{F}$.

For $f, g \in \mathbf{F}^S$, the sum $f + g \in \mathbf{F}^S$ is defined by the function

$$(f + g)(x) = f(x) + g(x)$$

for all $x \in S$. For $\lambda \in \mathbf{F}$ and $f \in \mathbf{F}^S$, the product $\lambda f \in \mathbf{F}^S$ is the function defined by

$$(\lambda f)(x) = \lambda f(x)$$

for all $x \in S$.

Remark 1.1.2. As an example, if $S = [0, 1]$ and $\mathbf{F} = \mathbf{R} = \mathbb{R}$, then $\mathbf{R}^{[0,1]}$ is the set of real-valued function defined on the interval $[0, 1]$.

Example 1.1.2. If S is a nonempty set, then $\mathbf{F}^S$ is a vector space over $\mathbf{F}$ with addition and scalar multiplication defined above.

The additive identity is $0 : S \to \mathbf{F}$ defined by $0(x) = 0$ for all $x \in S$. For any $f \in \mathbf{F}^S$, the additive inverse of f is the function $-f : S \to \mathbf{F}$ defined by $(-f)(x) = -f(x)$ for all $x \in S$.

Remark 1.1.3. $\mathbf{F}^\infty$ is really just a special of $\mathbf{F}^S$. Looking at the definition of $\mathbf{F}^\infty$, we can think of $x : \mathbb{N} \to \mathbf{F}$ as a function that takes every natural number to an element in $\mathbf{F}$.

Theorem 1.1.1. *A vector space has a unique additive identity and a unique additive inverse.*

Proof. The additive identity is unique. Suppose 0 and $0'$ are both additive identities for some vector space V. Then

$$0' = 0' + 0 = 0 + 0' = 0.$$

For the additive inverse, suppose $v \in V$ and both w and w' are additive inverses of v. Then

$$w = w + 0 = w + \big(v + w'\big) = (w + v) + w' = 0 + w' = w'.$$

□

Notation 1.1.1. Let $v, w \in V$. Then let $-v$ denotes the additive inverse of v. Let $w - v$ be defined as $w + (-v)$. For the rest of this notes, V denotes a vector space over $\mathbf{F}$.

Theorem 1.1.2. $0v = \mathbf{0}$ *for every* $v \in V$. *Also,* $a0 = \mathbf{0}$ *for every* $a \in \mathbf{F}$. *Here, the boldface* $\mathbf{0} \in V$ *means the zero vector. And* 0 *is a scalar in* $\mathbb{F}$.

Proof. For $v \in V$, we have

$$0v = (0 + 0)v = 0v + 0v.$$

Adding the additive inverse of $0v$ on both sides gives the desired result, that is, $0v = \mathbf{0}$.

For the second part we apply similar strategy. We have

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

And adding the additive inverse of $a0$ on both sides gives the desired result. □

Theorem 1.1.3. $(-1)v = -v$ *for every* $v \in V$.

Proof. For $v \in V$ we have

$$v + (-1)v = 1v + (-1)v = (1 + (-1))v = 0v = \mathbf{0}.$$

This proves the theorem. □

1.2 Subspaces, Sums, Direct Sums

Definition 1.2.1. Given a vector space V, a subset U of V is a subspace of V if U is also a vector space using the same addition and multiplication as on V.

Remark 1.2.1. The set $U = \{\mathbf{0}\}$ is a subspace of any vector space V. Conversely, if U is a subspace of V that contains only one element, then $U = \{\mathbf{0}\}$.

Theorem 1.2.1. ***(Conditions for a subspace)*** *A subset $U \subset V$ is a vector space of V if and only if V satisfies the following properties:*

1. ***additive identity**: $0 \in U$;*
2. ***closed under addition**: $u, w \in U$ implies $u + w \in U$;*
3. ***closed under scalar multiplication**: $a \in \mathbf{F}$ and $u \in U$ implies $au \in U$.*

Proof. Let U be a subspace of V, then by the definition of a subspace it is easy to verify that U satisfies all the properties above. Hence the first half of the proof is finished.

Conversely, suppose U satisfies the three properties above, we will verify that it satisfies the axioms in Definition.1.1.3.

Since U is a subspace of V, every element in U is in V. And since V satisfies commutativity(1) and associativity(2), it follows immediately that U also satisfies (1) and (2). Since U satisfies the first property in the theorem, the additive identity is in U, hence (3) is satisfied. Now, if $u \in U$, then since U is closed under multiplication, $-u = (-1)u \in U$, hence (4) is satisfied. (5) follows from the third property. The distributive law (6) follows again from $U \subset V$. □

Remark 1.2.2. Clearly, $\{\mathbf{0}\}$ is the smallest vector space of V and V itself is the largest subspace of V.

Definition 1.2.2. (Sums) Suppose $U_1, ..., U_m$ are subsets of V. The sum of $U_1, ..., U_m$, denoted by $U_1 + ... + U_m$, is the set of all possible sums of elements of $U_1, ..., U_m$. More precisely,

$$U_1 + \cdots + U_m = \{u_1 + \cdots + u_m : u_1 \in U_1, \ldots, u_m \in U_m\}.$$

Example 1.2.1. Let V be the set of all real-valued functions $f : [0, 1] \to \mathbf{R}$. Let

$$U = \left\{ f \in V : f\left(\frac{1}{3}\right) = f\left(\frac{1}{2}\right) = 0 \right\}, \quad W = \left\{ f \in V : f\left(\frac{2}{3}\right) = 0 \right\}.$$

Then if $f \in U$ and $g \in W$, the sum $U + W$ is the set of all functions with the form $f + g$. At $x = 1/3$ and $x = 1/2$, although $f(x) = 0$, but $g(x)$ can be anything. Conversely on $x = 2/3$, although $g(x) = 0$ but $f(x)$ can be anything. Hence

$$U + W = V.$$

Example 1.2.2. Let $V = \mathbf{R}^3$. Let us define

$$U = \left\{ \begin{pmatrix} 2y \\ y \\ 0 \end{pmatrix} : y \in \mathbf{R} \right\}, \; W = \left\{ \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} : x, y \in \mathbf{R} \right\}, \; Z = \left\{ \begin{pmatrix} 0 \\ 0 \\ z \end{pmatrix} : z \in \mathbf{R}. \right\}$$

Then U is a line with slope 2 on the $x - y$ plane. And it is obvious that $U + W = W$. And $U + Z$ is a slant plane with one direction of its slope being 2. Finally, $W + Z = V$. Also, note that if we define, alternatively, that

$$Z = \left\{ \begin{pmatrix} 0 \\ z \\ 2z \end{pmatrix} : z \in \mathbf{R} \right\},$$

Then $U + Z$, although having three components, is a two dimensional plane in $\mathbf{R}^3$.

Now if we consider three vectors defined by

$$U_1 = \begin{pmatrix} x \\ 2x \end{pmatrix}, \; U_2 = \begin{pmatrix} x \\ x \end{pmatrix}, \; U_3 = \begin{pmatrix} 2x \\ x \end{pmatrix}$$

where $x \in \mathbf{R}$, then we can express a vector in $\mathbf{R}^3$ using vectors in the form of $U_1 + U_2 + U_3$. For example:

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

However, this expression is not unique. We notice an alternate expression:

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} -2 \\ -2 \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

And sometimes we may want to only focus on the situations where the expression is unique, which inspires the following definition of the direct sum.

Definition 1.2.3. Suppose $U_1, ..., U_m$ are subspaces of V. The sum $U_1 + ... + U_m$ is called a **direct sum** if each element of $U_1 + ... + U_m$ can be written in only one way as a sum

$$u_1 + u_2 + ... + u_m \qquad u_j \in U_j \quad \text{for all} \quad 1 \leq j \leq m.$$

If $U_1 + ... + U_m$ is a direct sume, then we denote this direct sum with

$$U_1 \oplus U_2 \oplus \cdots \oplus U_m.$$

Example 1.2.3. Look at examples in Axler's book: Example 1.41, 1.42 and 1.43. Note that for a collection of subspaces to be a direct sum, these subspaces need not to be perpendicular.

Remark 1.2.3. Consider a vector space V, and a collection of its subspaces $U_1, ..., U_m$ that forms a direct sum. By definition the expression

$$\mathbf{0} = u_1 + u_2 + ... + u_m \qquad u_j \in U_j$$

is unique. Since $\mathbf{0} \in U_1 \cap U_2 \cap ... \cap U_m$, we know *one way* to write this expression is $\mathbf{0} = \mathbf{0} + ... + \mathbf{0}$. Hence this expression must be unique. It then follows that

$$u_1 = u_2 = ... = u_m = \mathbf{0} \qquad \text{for all } 1 \leq j \leq m.$$

Theorem 1.2.2. ***(Conditions for a direct sum)*** *Suppose $U_1, ...U_m$ are subspaces of V. Then $U_1 + ... + U_m$ is a direct sum if and only if the only way to write $\mathbf{0}$ as a sum $u_1 + ... + u_m$, where each $u_j \in U_j$, is by taking each $u_j = 0$.*

Proof. Suppose $U_1 + ... + U_m$ is a direct sum, then from Remark.1.2.3 we know that $u_j = 0$ for all j must be true. Hence we are already halfway through.

Conversely, suppose the only way to write $\mathbf{0}$ as a sum $u_1 + ... + u_m$, where each $u_j \in U_j$, is by taking each $u_j = \mathbf{0}$. To show that $U_1 + ... + U_m$ is a direct sum, we take $v \in U_1 + ... + U_m$. Then we can write

$$v = u_1 + \cdots + u_m.$$

We want to show that this expression is unique. Suppose this expression is not unique. That is, we also have

$$v = v_1 + \cdots + v_m$$

where $v_j \in U_j$. Then subtracting these two equalities we have

$$\mathbf{0} = (u_1 - v_1) + \cdots + (u_m - v_m).$$

Because $u_j - v_j \in U_j$ for all j, and each U_j is a subspace, then according to Remark.1.2.3, we must have $u_j - v_j = \mathbf{0}$ for all j. Hence $u_j = v_j$ for all j. Therefore, this expression of zero vector is unique. □

Corollary 1.2.1. ***(Direct sum of two subspaces)*** *Suppose U and W are subspaces of V. Then $U + W$ is a direct sum if and only if $U \cap W = \{\mathbf{0}\}$.*

Proof. First, suppose that $U + W$ is a direct sum. If $v \in U \cap W$, then $\mathbf{0} = v + (-v)$, where $v \in U$ and $-v \in W$. Again, we must have $v = \mathbf{0}$. And thus we claim $U \cap W = \{\mathbf{0}\}$.

Conversely, suppose $U \cap W = \{\mathbf{0}\}$. Suppose $u \in U$ and $w \in W$ and

$$\mathbf{0} = u + w.$$

Then to show that $U + W$ is a direct sum, Theorem.1.2.2 tells us that we should show $u = w = \mathbf{0}$. The equation above implies that $u = -w \in W$. Thus $u \in U \cap W$. Hence $u = \mathbf{0}$. Then $w = \mathbf{0}$ immediately follows from the equation above. □

Remark 1.2.4. Note that in the theorem above, the assertion only holds for two subspaces. When we are judging whether the sum of more than two subspaces is a direct sum, it is **insufficient** to show that the pairwise intersection is $\mathbf{0}$. For instance, consider Example 1.43 in the book:

$$\begin{aligned} U_1 &= \left\{(x, y, 0) \in \mathbf{F}^3 : x, y \in \mathbf{F}\right\} \\ U_2 &= \left\{(0, 0, z) \in \mathbf{F}^3 : z \in \mathbf{F}\right\} \\ U_3 &= \left\{(0, y, y) \in \mathbf{F}^3 : y \in \mathbf{F}\right\}. \end{aligned}$$

It is clear that $U_1 + U_2 + U_3$ is not a direct sum (proved in the book). However, we have

$$U_1 \cap U_2 = U_1 \cap U_3 = U_2 \cap U_3 = \{\mathbf{0}\}.$$

Chapter 2

Finite-Dimensional Vector Spaces

It is not worth an intelligent man's time to be the majority. By definition, there are already enough people to do that.

-G.H.Hardy

2.1 Span

Definition 2.1.1. (List) Let n be a nonnegative integer. A list of length n is an ordered collection of n elements with the form $(x_1, ..., x_n)$. Two lists are equal if and only if they have the same length and the same elements in the same order. Note that the elements in the list can be repetitive, as oppose to those of a set.

Notation 2.1.1. We usually write lists of vectors without the sorrounding parentheses.

Definition 2.1.2. (Linear Combination) A linear combination of a list $v_1, ..., v_m$ of vectors in V is a vector of the form

$$a_1 v_1 + \cdots + a_m v_m,$$

where $a_1, ..., a_m \in \mathbf{F}$.

Definition 2.1.3. (Span) The set of all linear combinations of a list of vectors $v_1, ..., v_m$ in V is called the span of $v_1, ..., v_m$. We write

$$\text{span}\,(v_1, \ldots, v_m) = \{a_1 v_1 + \cdots + a_m v_m : a_1, \ldots, a_m \in \mathbf{F}\}\,.$$

Remark 2.1.1. The span of an empty list is defined to be the zero vector. In other words, $\text{span}() = \{\mathbf{0}\}$.

Theorem 2.1.1. *The span of a list of vectors in V is the smallest subspace of V containing all the vectors in the list.*

Proof. Let $v_1, ..., v_m$ be a list of vectors in V. We can immediately verify that $\text{span}(v_1, ..., v_m)$ is a subspace of V since it contains the zero element:

$$\mathbf{0} = 0v_1 + \cdots + 0v_m,$$

and it is closed under addition and multiplication (by the definition of a span). That is, we have

$$(a_1v_1 + \cdots + a_mv_m) + (c_1v_1 + \cdots + c_mv_m) = (a_1 + c_1)\, v_1 + \cdots + (a_m + c_m)\, v_m,$$

which is in the span, and

$$\lambda\,(a_1v_1 + \cdots + a_mv_m) = \lambda a_1v_1 + \cdots + \lambda a_mv_m,$$

which is also in the span. Hence $\text{span}(v_1, ..., v_m)$ is a subspace of V by the subspace criterion.

Now we prove that the span is the smallest subspace. First, notice that each v_j where $1 \leq j \leq m$ is in the span, by taking $a_j = 1$ and $a_i = 0$ for all $i \neq j$.

Conversely, since *every subspace is closed under addition and multiplication*, every subspace that contains $v_1, ..., v_m$ will also contain their linear combination. Hence $\text{span}(v_1, ..., v_m)$ is the smallest subspace of V that contains this list of vectors. □

Notation 2.1.2. In the following text we may occasionally use the simplified notation below:

$$\text{span}(v_1, v_2, ..., v_{j-1}) = \text{span}(v_{<j}).$$

Definition 2.1.4. If $\text{span}(v_1, ..., v_m)$ equals V, we say that $v_1, ..., v_m$ spans V.

Definition 2.1.5. (Polynomial) A function $p : \mathbf{F} \to \mathbf{F}$ is called a Polynomial with coefficients in $\mathbf{F}$ if there exists $a_0, a_1, ..., a_m \in \mathbf{F}$ such that

$$p(z) = a_0 + a_1z + a_2z^2 + \cdots + a_mz^m$$

for all $z \in \mathbf{F}$. $\mathcal{P}(\mathbf{F})$ is the set of all polynomials with coefficients in $\mathbf{F}$.

Proposition 2.1.1. *$\mathcal{P}(\mathbf{F})$ is a vector space over $\mathbf{F}$. It is also a subspace of $\mathbf{F}^{\mathbf{F}}$, the vector space of functions $f : \mathbf{F} \to \mathbf{F}$.*

Proof. Left as exercise. □

Definition 2.1.6. (Degree of polynomial) A polynomial $p \in \mathcal{P}(\mathbf{F})$ is said to have a degree m if there exist scalars $a_0, a_1, ..., a_m \in \mathbf{F}$ with $a_m \neq 0$ such that

$$p(z) = a_0 + a_1z + \cdots + a_mz^m$$

for all $z \in \mathbf{F}$. We write $\deg p = m$.

Definition 2.1.7. For m a nonnegative integer, $\mathcal{P}_m(\mathbf{F})$ denotes the set of all polynomials with coefficients in $\mathbf{F}$ and degree at most m.

Definition 2.1.8. (Finite and infinite dimensional vector space) A vector space is called finite dimensional if some list of vectors in it spans the space. A vector space is called infinite dimensional if it is not finite dimensional.

Example 2.1.1. $\mathcal{P}_m(\mathbf{F})$ is a finite dimensional vector space since

$$\mathcal{P}_m(\mathbf{F}) = \text{span}(1, z, ..., z^m),$$

whereas $\mathcal{P}(\mathbf{F})$ is infinite dimensional.

2.2 Linear Independence

Sometimes we want to focus on those cases where $v \in \text{span}(v_{<m+1})$ can be written *uniquely* as a linear combination of $v_1, ..., v_m$. In that case, given two expressions

$$v = a_1 v_1 + \cdots + a_m v_m,$$

$$v = c_1 v_1 + \cdots + c_m v_m,$$

subtracting these two equations will give

$$\mathbf{0} = (a_1 - c_1)\, v_1 + \cdots + (a_m - c_m)\, v_m.$$

Thus we have written the zero vector into the linear combination of $v_1, ..., v_m$. And since there is only one way to express v in terms of the linear combination of $v_1, ..., v_m$. The only way to do this is to make $a_j - c_j = 0$ for all $1 \leq j \leq m$. Hence, for v to be expressed uniquely as a linear combination of the spanning vectors is equivalent to saying the only way to express zero vector as a linear combination of the spanning vector is the trivial way, that is:

$$\mathbf{0} = 0v_1 + 0v_2 + ... + 0v_m.$$

This situation is so important that we give it a name – linear independence.

Notation 2.2.1. In the remaining text, we will write $\mathbf{0}$, the zero vector, as 0 for convinience.

Definition 2.2.1. (Linear Independence) A list $v_1, ..., v_m$ of vectors in V is linearly independent if the only choice of $a_1, ..., a_m \in \mathbf{F}$ that makes

$$a_1 v_1 + a_2 v_2 + ... + a_m v_m = 0$$

is to let

$$a_1 = a_2 = ... = a_m = 0.$$

Theorem 2.2.1. *$v_1, ..., v_m$ is linearly independent if and only if each vector $v \in \text{span}(v_{<m+1})$ has one and only one representation as a linear combination of $v_1, ..., v_m$.*

Proof. Already shown in the beginning of this section. You can do it in reverse order yourself. □

Definition 2.2.2. A list of vectors in V is called linearly dependent if it is not linearly independent. That is, there exists $a_1, ..., a_m \in \mathbf{F}$, not all 0, such that

$$a_1 v_1 + ... + a_m v_m = 0.$$

Lemma 2.2.1. ***(Linear dependence lemma)*** *Suppose $v_1, ..., v_m$ is a linearly dependent list in V. Then there exists $j \in \{1, 2, ..., m\}$ such that the following hold:*

1. *$v_j \in \text{span}(v_{<j})$;*
2. *If the j^{th} term is removed from $v_1, ..., v_m$, the span of the remaining list equals $\text{span}(v_1, ..., v_m)$.*

Proof. Because the list is linearly dependent, there exist $a_1, ..., a_m \in \mathbf{F}$, not all 0, such that

$$a_1 v_1 + \cdots + a_m v_m = 0.$$

Let j be the **largest** element of $\{1, ..., m\}$ such that $a_j \neq 0$. Then we can write

$$v_j = -\frac{a_1}{a_j} v_1 - \cdots - \frac{a_{j-1}}{a_j} v_{j-1}. \tag{2.1}$$

This proves the first claim.

Now suppose $u \in \text{span}(v_1, ..., v_m)$, then there exists a set of coefficients $c_1, ..., c_m$ in $\mathbf{F}$ such that

$$u = c_1 v_1 + \cdots + c_m v_m.$$

Substituting (2.1) in and we have

$$u = c_1 v_1 + \cdots + \left(-\frac{a_1}{a_j} v_1 - \cdots - \frac{a_{j-1}}{a_j} v_{j-1} \right) + \cdots + c_m v_m,$$

and the second claim immediately follows. □

Theorem 2.2.2. *In a finite dimensional vector space V, the length of every linearly independent list of vectors is less than or equal to the length of every spanning list of vectors.*

Proof. Suppose $u_1, \cdots, u_m$ is linearly independent in V. Suppose that $w_1, \cdots, w_n$ spans V. We will prove that $m \leq n$ with a series of steps defined as the following:

(Step 1) Let B be the list $w_1, ..., w_n$ that spans V. Thus adjoining any vector in V to this list B will produce a linearly dependent list. In particular, the list $u_1, w_1, ..., w_m$ is linearly dependent. Thus by the Lemma.2.2.1, we can delete one of the w's so that the new list B (of length n) consisting of u_1 and the remaining w's spans V.

(Step j) The list B from step $j-1$ spans V. Adding u_j to this list will create a linearly dependent list. By Lemma.2.2.1, one of the vectors vectors in the list can be deleted without changing the span. But since $u_1, \cdots, u_j$ are linearly independent by our assumption, this vector is one of w's, not one of u's. We can remove that w from B and the new list B consisting of $u_1, \cdots, u_j$ and the remaining w's spans V.

After step m, we have added all of $u_1, \cdots, u_m$ to the list B and the process stops. At each setp we add a u to B, the linear dependence lemma tells us there is some w to remove. Thus there are at least as many w's as u's, this concludes the proof. □

Example 2.2.1. See Examples2.24 and 2.25 to find out why this theorem is helpful.

Corollary 2.2.1. *Every subspace U of a finite-dimensional vector space V is finite-dimensional.*

Proof. **(Step 1)** If $U = \{0\}$, then we are done. If not, then choose a nonzero vector $v_1 \in U$.

(Step j) If $U = \text{span}(v_{<j})$, then U is finite-dimensional and we are done. If not, choose a vector $v_j \in U$ such that $v_j \notin \text{span}(v_{<j})$.

After each step, we have constructed a linearly independent list by Lemma.2.2.1. This linearly independent list cannot be longer than the spanning list of V by Theorem.2.2.2. And since V is a finite-dimensional vector space, its spanning list has a finite length. Thus our constructed list also has a finite length, as desired. □

2.3 Bases

Definition 2.3.1. A basis of V is a list of vectors in V that is linearly independent and spans V.

Theorem 2.3.1. *A list $v_1, \cdots, v_n$ of vectors in V is a basis of V if and only if every $v \in V$ can be written uniquely in the form*

$$v = a_1 v_1 + \cdots + a_n v_n \tag{2.2}$$

where $a_1, \cdots, a_n \in \mathbf{F}$.

Proof. First, suppose that $v_1, \cdots, v_n$ is a basis of V. Let $v \in V$. By definition we know that $\text{span}(v_1, \cdots, v_n) = V$, then there exists $a_1, \cdots, a_n$ such that (2.2) holds. Suppose $c_1, \cdots, c_n$ are also scalars such that

$$v = c_1 v_1 + \cdots + c_n v_n.$$

Then subtracting this equation from (2.2) we get

$$0 = (a_1 - c_1)\, v_1 + \cdots + (a_n - c_n)\, v_n.$$

Since $v_1, \cdots, v_n$ are linearly independent, we must have $a_i = c_i$ for all i. Hence the representation is unique.

Now, suppose every $v \in V$ can be written uniquely in the form of (2.2). Clearly $\text{span}(v_1, \cdots, v_n) = V$. Suppose $a_1, \cdots, a_n \in \mathbb{F}$ such that

$$0 = a_1 v_1 + \cdots + a_n v_n.$$

Taking $a_1 = a_2 = \cdots = a_n = 0$ we get one expression, and by assumption it is unique. Hence the list is linearly independent. □

Theorem 2.3.2. *Given a spanning list $B = \{v_1, \cdots, v_n\}$, we can always reduce it to a basis of a vector space.*

Proof. Consider the following steps:

Step 1: If $v_1 = 0$, delete v_1 from B, otherwise leave B unchanged.

Step j: If v_j is in $\text{span}(v_{<j})$m delete v_j from B, otherwise, leave B unchanged.

Stop the process after n steps, the obtained list B is a basis of V. Note that the process does not change the span of the original list B. By the Linear Dependence Lemma (Lem. 2.2.1), the obtained list is linearly independent. □

Corollary 2.3.1. *Every finite dimensional vector space has a basis.*

Proof. By definition every finite dimensional vector space has a spanning list. And by Thm.2.3.2 we can reduce it to a basis. □

Theorem 2.3.3. *Every linearly independent list of vectors in a finite dimensional vector space can be extended to a basis of the vector space.*

Proof. Let $u_1 \cdots, u_m$ be a linearly independent list of vectors in a finite dimensional vector space V. Let $w_1, \cdots, w_n$ be a basis of V. Then the list

$$u_1, \cdots, u_m, w_1, \cdots, w_n$$

spans V. By Thm.2.3.2, this list can be reduced to a basis of V without deleting any us since they are linearly independent. Hence we will get a list of vectors including all the us and some of ws, which is the desired extension. □

Theorem 2.3.4. *Suppose V is finite dimensional and U is a subspace of V. Then there is a subspace W of V such that $V = U \oplus W$.*

Proof. Since V is finite dimensional, so is U. Then there is a basis for U, which we call $u_1, \cdots, u_m$. Of course, $u_1 \cdots, u_m$ is linearly independent, hence by Thm.2.3.3 it can be extended to a basis

$$u_1, \cdots, u_m, w_1, \cdots, w_n$$

of V. Let $\text{span}(w_1, \cdots, w_n) = W$, we claim that this is the desired W.

To prove that $V = U \oplus W$, we need to show that

$$V = U + W \quad \text{and} \quad U \cap W = \{0\}.$$

Since $u_1, \cdots, u_m, w_1, \cdots, w_n$ is a basis of V, for any $v \in V$ we have

$$v = \underbrace{a_1u_1 + \cdots + a_mu_m}_{u \in U} + \underbrace{b_1w_1 + \cdots + b_nw_n}_{w \in W}.$$

Thus $V = U + W$ is proven.

Suppose $v \in U \cap W$. Then there exists $a_1, \cdots, a_m, b_1, \cdots, b_n \in \mathbb{F}$ such that

$$v = a_1u_1 + \cdots + a_mu_m = b_1w_1 + \cdots + b_nw_n.$$

Thus

$$a_1u_1 + \cdots + a_mu_m - b_1w_1 - \cdots - b_nw_n = 0.$$

Since $u_1, \cdots, u_m, w_1, \cdots, w_n$ are linearly independent, we have

$$a_1 = \cdots = a_m = b_1 = \cdots = b_n = 0 \quad \text{or} \quad v = 0.$$

Thus $U \cap W = \{0\}$. And the theorem follows. □

2.4 Dimension

Lemma 2.4.1. *Any two bases of a finite dimensional vector space have the same length.*

Proof. Let the two bases be B_1 and B_2. Then both B_1 and B_2 spans the vector space V. Then by Thm.2.2.2, we have

$$\text{length}(B_1) \leq \text{length}(B_2) \quad \text{and} \quad \text{length}(B_2) \leq \text{length}(B_1).$$

This immediately gives $\text{length}(B_1) = \text{length}(B_2)$ and we are done. □

Definition 2.4.1. The dimension of a finite dimensional vector space is the length of any basis if the vector space.

Example 2.4.1. $\dim(\mathbf{F}^n) = n$. This is obvious since the standard basis have length n.

Example 2.4.2. $\dim \mathcal{P}_m(\mathbf{F}) = m + 1$.

Theorem 2.4.1. *If V is finite dimensional and U is a subspace of V, then*

$$\dim U \leq \dim V.$$

Proof. Any basis of U is a linearly independent list in V and any basis of V is a spanning list in V. Then the theorem immediately follows from Thm.2.2.2. □

The next theorem gives a very convinient way of checking whether a list of vectors is a basis. It says any linearly independent list of the right length is a basis.

Theorem 2.4.2. *Suppose V is a finite dimensional. Then every linearly independent list of vectors in V with length* $\dim V$ *is a basis of V.*

Proof. Suppose $\dim V = n$ and $v_1 \cdots, v_n$ is linearly independent in V. By Thm.2.3.3, this list can be extended to a basis. But every basis in V has length n. Hence this list itself must be a basis. □

Here is a very good exercise from the book:

Exercise 2.4.1. Show that $1, (x-5)^2, (x-5)^3$ is a basis of the subspace U of $\mathcal{P}_3(\mathbf{R})$, defined by

$$U = \left\{p \in \mathcal{P}_3(\mathbf{R}) : p'(5) = 0\right\}.$$

Solution 2.4.1. It is clear that $1, (x-5)^2, (x-5)^3$ are in U. Take their linear combination and set it to be zero we have

$$a + b(x-5)^2 + c(x-5)^3 = 0.$$

Simple observation gives $a = b = c = 0$. Hence this list is linearly independent. Thus $\dim U \geq 3$.

Also, since U is a subspace of $\mathcal{P}_3(\mathbf{R})$, by the previous theorem we have

$$\dim U \leq \dim \mathcal{P}_3(\mathbf{R}) = 3 + 1 = 4.$$

But $\dim U \neq 4$, otherwise $p'(5) = 0$ will not always be true. Hence $\dim U = 3$, and by Thm.2.4.2 we are done.

Theorem 2.4.3. *Similarly, every spanning list of a finite dimensional vector space V with length* $\dim V$ *is a basis of V.*

Proof. By Thm.2.3.2, we can construct a similar proof to that of Thm.2.4.2. □

Theorem 2.4.4. *If U_1, U_2 are subspaces of a finite dimensional vector space, then*

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2).$$

Proof. Let $u_1 \cdots, u_m$ be a basis of $U_1 \cap U_2$, that is, $\dim(U_1 \cap U_2) = m$. Then this list is clearly a independent list in both U_1 and U_2. Let $u_1, \cdots, u_m, v_1, \cdots, v_j$ be a basis of U_1 and $u_1, \cdots, u_m, w_1, \cdots, w_k$ be a basis of U_2. Hence $\dim U_1 = m + j$ and $\dim U_2 = m + k$. We will show that

$$u_1, \ldots, u_m, v_1, \ldots, v_j, w_1, \ldots, w_k$$

is a basis of $U_1 + U_2$. This will give

$$\begin{aligned}\dim(U_1 + U_2) &= m + j + k \\ &= (m + j) + (m + k) - m \\ &= \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2)\end{aligned}$$

and the theorem immediately follows.

Clearly, for any vector $u \in U_1 + U_2$, we can write it as

$$u = \underbrace{\sum_{i=1}^{m} \alpha_i u_i + \sum_{i=1}^{j} \beta_i v_i}_{\in U_1} + \underbrace{\sum_{i=1}^{m} \gamma_i u_i + \sum_{i=1}^{k} \eta_i w_i}_{\in U_2},$$

which is a linear combination of $u_1, \ldots, u_m, v_1, \ldots, v_j, w_1, \ldots, w_k$. Hence we know that

$$\operatorname{span}(u_1, \ldots, u_m, v_1, \ldots, v_j, w_1, \ldots, w_k) = U_1 + U_2.$$

Next, we want to show that this list is linearly independent. Suppose

$$\sum_{i=1}^{m} a_i u_i + \sum_{i=1}^{j} b_i v_i + \sum_{i=1}^{k} c_i w_i = 0, \tag{2.3}$$

we want to show that $a_i = b_i = c_i = 0$ for all i. Write

$$\sum_{i=1}^{k} c_i w_i = -\sum_{i=1}^{m} a_i u_i - \sum_{i=1}^{j} b_i v_i$$

Note that the vector on the left hand side is in U_2 and the vector on the right hand side is in U_1. Hence $\sum c_i w_i \in U_1 \cap U_2$. We can write then

$$c_1 w_1 + \cdots + c_k w_k = d_1 u_1 + \cdots + d_m u_m.$$

But since $u_1, \cdots, u_m, w_1, \cdots, w_k$ is a basis of U_2, they are linearly independent. This implies that $c_i = d_i = 0$ for all i. Hence (2.3) becomes

$$a_1 u_1 + \cdots + a_m u_m + b_1 v_1 + \cdots + b_j v_j = 0.$$

But since $u_1, \ldots, u_m, v_1, \ldots, v_j$ is a basis of U_1, they are linearly independent. This yields $a_i = b_i = 0$ for all i

This concludes the proof. □

We conclude this chapter with a very simple proposition.

Proposition 2.4.1. *Let $U_1, \cdots, U_m$ be subspaces of a finite dimensional vector space V. If $V = U_1 + U_2 + \cdots + U_m$ and at the same time,* $\dim V = \dim U_1 + \dim U_2 + \cdots \dim U_m$, *Then we have*

$$V = U_1 \oplus U_2 \oplus \cdots \oplus U_m.$$

Proof. Choose a basis for each U_i (each with length $\dim U_i$) and combine them into one list with length $\sum_i \dim U_i = \dim V$. This combined list spans V since $V = \sum_i U_i$. Because it is a spanning list, it can always be reduced into a basis of V. But since it already has length $\dim V$, it is, by itself, a basis of V.

Then by definition of a basis, every $v \in V$ can be written uniquely as a linear combination of the vectors in the basis of V, which is also a unique combination of each $u_i \in U_i$. This is the definition of a direct sum. □

Chapter 3

Linear Maps

Though I had success in my research both when I was mad and when I was not, eventually I felt that my work would be better respected if I thought and acted like a normal person.

-John Forbes Nash, Jr.

Notation 3.0.1. In this chapter, **F** denotes **R** or **C**.

Notation 3.0.2. In this chapter, V and W denotes vector spaces over **F**.

3.1 Linear Transformations

Definition 3.1.1. A linear transformation (or a linear map) from V to W is a function $T : V \to W$ with the following properties:

1. **Additivity:** $T(u + v) = Tu + Tv$ for all $u, v \in V$;
2. **Homogeneity:** $T(\lambda v) = \lambda(Tv)$ for all $\lambda \in \mathbf{F}$ and for all $v \in V$.

Notation 3.1.1. $\mathcal{L}(V, W)$ denotes the set of all linear transformations from V to W.

Now that we defined the set of all linear transformations, we can define the addition and scalar multiplication on this set.

Definition 3.1.2. Suppose $S, T \in \mathcal{L}(V, W)$ and $\lambda \in \mathbf{F}$. The sum $S + T$ and the product λT are the linear maps from V to W defined by

$$(S + T)(v) = Sv + Tv \quad \text{and} \quad (\lambda T)(v) = \lambda(Tv)$$

for all $v \in V$.

Theorem 3.1.1. *With the operations of addition and scalar multiplication defined above, $\mathcal{L}(V, W)$ is a vector space.*

The proof is omitted in Axler's book (left to readers), here we give a brief outline of the proof.

Proof. We need to verify the axioms of the vector spaces one by one. We use the capital letters $S, T, K \in \mathcal{L}(V, W)$ to denote linear transformations from V to W, and lower case letters and greek letters such as $a, b, c, \mu, \eta \in \mathbf{F}$ to denote the field elements.

1. **Commutativity:** This should be obvious from the definition of the addition. We have
$$(S + T)v = Sv + Tv = Tv + Sv = (T + S)v$$
since the addition $Sv + Tv$ in vector space W is commutative.

2. **Associativity:** Again by definition we have
$$\begin{aligned}[(S + T) + K]v &= (S + T)v + Kv \\ &= Sv + Tv + Kv \\ &= Sv + (Tv + Kv) \\ &= [S + (T + K)]v.\end{aligned}$$
Also, given $\mu, \eta \in \mathbf{F}$, we have
$$\mu[\eta T(v)] = \mu\eta T(v) = (\mu\eta)T(v).$$

3. **Additive Identity:** Define the zero function $\mathbf{0} : V \to W$ to be the function that takes every element $v \in V$ and gives the zero $\mathbf{0}_W$ in W. Note that it is also linear:
$$\mathbf{0}(au + bv) = \mathbf{0}_W = \mathbf{0}_W + \mathbf{0}_W = a\mathbf{0}_W + b\mathbf{0}_W = a\mathbf{0}(u) + b\mathbf{0}(v).$$
for $a, b \in \mathbf{F}$. Then it is clear that $\mathbf{0} \in \mathcal{L}(V, W)$.

4. **Additive Inverse:** Define the additive inverse of T to be
$$-T = \mathbf{0} + (-1)T.$$
Since both $\mathbf{0}$ and T are linear transformations from V to W, so is their linear combination (as you should verify). Hence the additive inverse is also in $\mathcal{L}(V, W)$.

5. **Multiplicative Inverse:** for all $T \in \mathcal{L}(V, W)$, we have $1 \in \mathbf{F}$ such that $1T = T$.

6. **Distributive Properties:** Given $a, b \in \mathbf{F}$ and $S, T \in \mathcal{L}(V, W)$, we have

$$a(S+T)v = a(Sv+Tv) = aSv + aTv = (aS+aT)v$$

and

$$(aT+bT)v = (aT)v + (bT)v = a(Tv) + b(Tv) = (a+b)Tv.$$

Hence the set of all linear transformations from V to W is itself a vector space. This concludes the proof. □

Now given two linear maps $T : U \to V$ and $S : V \to W$, we define their composite map $S \circ T$ by their product ST as the following.

Definition 3.1.3. If $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, then the product $ST \in \mathcal{L}(U, W)$ is defined by

$$(ST)(u) = S(Tu)$$

for all $u \in U$.

It is very easy to verify that ST is indeed a linear transformation. Note that ST is defined only when T maps into the domain of S.

Theorem 3.1.2. *Here are some basic algebraic properties of products of linear maps.*

1. ***Associativity:*** *whenever T_1, T_2, T_3 are linear maps such that the product make sense, we have*
$$(T_1T_2)T_3 = T_1(T_2T_3);$$

2. ***Identity:*** *whenever $T \in \mathcal{L}(V, W)$, we have*
$$TI_V = I_WT,$$
where I_V is the identity map on V and I_W is the identity map on W.

3. ***Distributive Properties:*** *whenever $T, T_1, T_2 \in \mathcal{L}(U, V)$ and $S, S_1, S_2 \in \mathcal{L}(V, W)$, we have*
$$(S_1+S_2)\,T = S_1T + S_2T \quad \text{and} \quad S\,(T_1+T_2) = ST_1 + ST_2.$$

Proof. All three properties are very easy to prove.

1. $(TI_V)v = Tv = I_W(Tv) = (I_WT)v$.
2. $[(T_1T_2)T_3]v = (T_1T_2)(T_3v) = T_1(T_2T_3v)$.

3. From direct computation we have:

$$[(S_1 + S_2)T]v = (S_1 + S_2)Tv = S_1Tv + S_2Tv = (S_1T + S_2T)v.$$

$$S(T_1 + T_2)v = S[T_1v + T_2v] = ST_1v + ST_2v = (ST_1 + ST_2)v.$$

□

Remark 3.1.1. Multiplication of linear map is not commutative, It is not necessarily true that $ST = TS$, even if both sides of the equation make sense.

Corollary 3.1.1. *If $T \in \mathcal{L}(V, W)$, then $T(0) = 0$.*

Proof. $T(0) = T(0 + 0) = T(0) + T(0)$. Done. □

The next theorem gives us an interesting piece of information. It says given a basis $v_1, \cdots, v_n$ of V, we can have a linear transformation that maps these basis vectors to any vectors we want. Also, the theorem says this linear transformation is unique.

Theorem 3.1.3. *Suppose $v_1, \cdots, v_n$ is a basis of V and $w_1, \cdots, w_n \in W$. Then there exists a unique linear map $T : V \to W$ such that*

$$Tv_j = w_j$$

for each $j = 1, \cdots, n$.

Proof. We first need to give a map that works. Define $T : V \to W$ to have the following property:

$$T(c_1v_1 + \cdots + c_nv_n) = c_1w_1 + \cdots + c_nw_n, \tag{3.1}$$

where $c_1, \cdots, c_n \in \mathbf{F}$. Since every $v \in V$ can be written uniquely as a linear combination of $v_1, \cdots, v_n$, the definition above does define a function $T : V \to W$.

Next, this map satisfy $Tv_j = w_j$ by taking $c_j = 1$ and all other c's equal to 0.

Also, this map is linear. Given any $u, v \in V$ with $u = \sum_{i=1}^{n} a_iv_i$ and $v = \sum_{i=1}^{n} c_iv_i$, we have

$$\begin{aligned} T(u + v) &= T((a_1 + c_1)v_1 + \cdots + (a_n + c_n)v_n) \\ &= (a_1 + c_1)w_1 + \cdots + (a_n + c_n)w_n \\ &= (a_1w_1 + \cdots + a_nw_n) + (c_1w_1 + \cdots + c_nw_n) \\ &= Tu + Tv. \end{aligned}$$

Also, given any $\lambda \in \mathbf{F}$, we have

$$\begin{aligned} T(\lambda v) &= T(\lambda c_1v_1 + \cdots + \lambda c_nv_n) \\ &= \lambda c_1w_1 + \cdots + \lambda c_nw_n \\ &= \lambda(c_1w_1 + \cdots + c_nw_n) \\ &= \lambda Tv. \end{aligned}$$

Finally, we claim that this linear map is unique. Because any linear map T that satisfies $Tv_j = w_j$ must also satisfy (3.1) by additivity and homogeneity. □

Remark 3.1.2. The transformed vectors $Tv_j = w_j$ need not to be a basis for W. Only in special cases will they be a basis for W. Here are two counter-examples given.

1. The linear transformation that maps every vector in V to zero vector in W.
2. The linear transformation $f : \mathbf{R}^2 \to \mathbf{R}^3$. The basis v_1 and v_2 and transformed into w_1 and w_2, which is not a basis since it does not have length $\dim \mathbf{R}^3 = 3$.

We conclude this section by giving an example of a very important type of linear transformation. Consider any linear transformation from $\mathbf{F}^n$ to $\mathbf{F}^m$.

Let $A_{j,k} \in \mathbf{F}$ for $j = 1, \cdots, m$ and $k = 1, \cdots, n$. Any linear transformation $T \in \mathcal{L}(\mathbf{F}^n, \mathbf{F}^m)$ is of the form

$$T(x_1, \ldots, x_n) = (A_{1,1}x_1 + \cdots + A_{1,n}x_n, \ldots, A_{m,1}x_1 + \cdots + A_{m,n}x_n).$$

This gives rise to the definition of a matrix. An $n \times m$ matrix defines a linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^m$.

For example, consider the linear transformation $T : \mathbf{R}^3 \to \mathbf{R}^2$ defined by

$$T(x, y, z) = (2x - y + 3z, 7x + 5y - 6z).$$

This transformation corresponds to the matrix

$$T = \begin{pmatrix} 2 & -1 & 3 \\ 7 & 5 & -6 \end{pmatrix}$$

since

$$\begin{pmatrix} 2 & -1 & 3 \\ 7 & 5 & -6 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2x - y + 3z \\ 7x + 5y - 6z \end{pmatrix}.$$

3.2 Nullspace, Injective and Surjective Maps, Range

Definition 3.2.1. For $T \in \mathcal{L}(V, W)$, the null space of T, denoted by $\text{null } T$, is the subset of V consisting of those vectors that T maps to 0:

$$\text{null } T = \{v \in V : Tv = 0\}.$$

It is easy to verify that nullspace itself is a subspace of V.

Theorem 3.2.1. *Suppose $T \in \mathcal{L}(V, W)$, then* $\text{null } T$ *is a subspace of V.*

Proof. We use the subspace criterion. First, since T is a linear map, we know $T(0) = 0$. Thus $0 \in \text{null}\, T$. Suppose $u, v \in \text{null}\, T$, then

$$T(u + v) = Tu + Tv = 0 + 0 = 0.$$

Hence $u + v \in \text{null}\, T$. Also, given $\lambda \in \mathbf{F}$ and $u \in \text{null}\, T$, we have

$$T(\lambda u) = \lambda Tu = \lambda 0 = 0.$$

Hence the nullspace is closed under addition and multiplication. □

Now we define the injective map, also known as the one-to-one or 1-1 map.

Definition 3.2.2. A function $T : V \to W$ is called injective if $Tu = Tv$ implies $u = v$.

Theorem 3.2.2. *Let $T \in \mathcal{L}(V, W)$. Then T is injective if and only if* $\text{null}\, T = \{0\}$.

Proof. First, suppose T is injective. We already know that $\{0\} \subset \text{null}\, T$ by the previous theorem. Suppose $v \in \text{null}\, T$, then

$$T(v) = 0 = T(0),$$

and $v = 0$ follows immediately since T is injective.

Now, suppose $\text{null}\, T = \{0\}$. We want to show that T is injective. Take $u, v \in V$ such that $Tu = Tv$. Then

$$0 = Tu - Tv = T(u - v).$$

Hence $u - v \in \text{null}\, T$, implying $u = v$. □

Definition 3.2.3. Suppose $T : V \to W$ is a function, the range of T is the subset of W consisting of thos vectors that are of the form Tv for some $v \in V$:

$$\text{range}\, T = \{Tv : v \in V\}.$$

Now we define the surjective map, or onto map.

Definition 3.2.4. A function $T : V \to W$ is called surjective if its range equals W.

Theorem 3.2.3. *If $T \in \mathcal{L}(V, W)$, then* range T *is a subspace of W.*

Proof. Suppose $T \in \mathcal{L}(V, W)$. Then $T(0) = 0$, which implies $0 \in \text{range}\, T$. If $w_1, w_2 \in \text{range}\, T$, then there exists $v_1, v_2 \in V$ such that $Tv_1 = w_1$ and $Tv_2 = w_2$. Thus

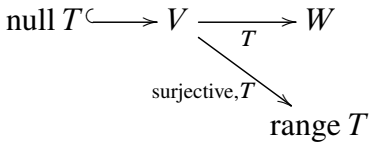
$$T(v_1 + v_2) = Tv_1 + Tv_2 = w_1 + w_2.$$

Hence $w_1 + w_2 \in \text{range}\, T$. Also, if $w \in \text{range}\, T$ and $\lambda \in \mathbf{F}$, then there exists $v \in V$ such that $Tv = w$. Thus

$$T(\lambda v) = \lambda Tv = \lambda w.$$

Hence $\lambda w \in \text{range}\, T$. □

The diagram below shows the relation between null T and range T. There exists a injective map from null T to V, which is another way of saying that the nullspace of T is a subset of V. Also, there exists a surjective map from V to the image of T, which is range T.



Now we come to a very important result. Axler calls it "Fundamental Theorem of Linear Maps". But it has a more standard name. It is called the Rank-Nullity theorem. Nullity, of course, refers to the dimension of nullspace. We define the rank to be the dimension of the range.

Theorem 3.2.4. ***(Rank-Nullity Theorem)*** *Suppose V is finite dimensional and $T \in \mathcal{L}(V, W)$. Then* range T *is finite dimensional and*

$$\dim V = \dim \text{null } T + \dim \text{range } T.$$

Proof. Let $u_1, \cdots, u_m$ be a basis of null T. Then $\dim \text{null } T = m$. The linear independent list $u_1, \cdots, u_m$ can be extended to a basis

$$u_1, \ldots, u_m, v_1, \ldots, v_n$$

of V. Thus $\dim V = m + n$. We need to prove that range T is finite dimensional and $\dim \text{range } T = n$. We show this by proving that $Tv_1, \cdots, Tv_n$ is a basis of range T.

Let $v \in V$, since u's and v's span V, we can write, for some $a_i, b_i \in \mathbf{F}$,

$$v = a_1u_1 + \cdots + a_mu_m + b_1v_1 + \cdots + b_nv_n.$$

Applying T to both sides of this equation we get

$$Tv = b_1Tv_1 + \cdots + b_nTv_n,$$

where the terms Tu_j becomes zero because $u_j \in \text{null } T$ for all j. The last equation implies that $Tv_1, \cdots, Tv_n$ spans range T.

Now we show $Tv_1, \cdots, Tv_n$ is linearly independent. Suppose for some $c_1, \cdots, c_n \in \mathbf{F}$ we have

$$c_1Tv_1 + \cdots + c_nTv_n = 0.$$

Then

$$T(c_1v_1 + \cdots + c_nv_n) = 0.$$

This implies that $\sum c_i v_i \in \text{null}\, T$. Thus we can write

$$c_1 v_1 + \cdots + c_n v_n = d_1 u_1 + \cdots + d_m u_m.$$

But since u's and v's are linearly independent, we have $c_i = d_i = 0$ for all i. This shows that Tv's are linearly independent, as desired. □

This theorem has two very important consequences.
The first corollary says a map to a smaller dimensional space is not injective.

Corollary 3.2.1. *Suppose V, W are finite dimensional vector spaces such that* $\dim V > \dim W$. *Then no linear map from V to W is injective.*

Proof. Let $T \in \mathcal{L}(V, W)$. Then

$$\begin{aligned} \dim \text{null}\, T &= \dim V - \dim \text{range}\, T \\ &\geq \dim V - \dim W \\ &> 0 \end{aligned}$$

by Thm.3.2.4. The inequality above says $\dim \text{null}\, T > 0$. This means that $\text{null}\, T$ contains vectors other than 0. Recall that $\dim\{0\} = 0$ since the zero vector has an empty spanning list. Thus T is not injective by Thm.3.2.2. □

The next corollary says a map to larger dimensional space is not surjective.

Corollary 3.2.2. *Suppose V, W are finite dimensional vector spaces such that* $\dim V < \dim W$. *Then no linear map from V to W is surjective.*

Proof. Let $T \in \mathcal{L}(V, W)$. Then

$$\begin{aligned} \dim \text{range}\, T &= \dim V - \dim \text{null}\, T \\ &\leq \dim V \\ &< \dim W \end{aligned}$$

by Thm.3.2.4. Thus $\text{range}\, T \neq W$, since otherwise we will have equality instead of inequality. Hence by the definition of a surjective map (Def.3.2.4), T is not surjective. □

These two results become very important when we are solving systems of linear equations. Given a homogeneous susyem of linear equations:

$$\begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \cdots & \vdots \\ A_{m1} & \cdots & A_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \mathbf{0},$$

if the system has more variables than equations ($n > m$), then we are doing a linear transformation $T : \mathbf{R}^n \to \mathbf{R}^m$, that is, from a bigger space into a smaller space. By Cor.3.2.1, the transformation is not injective, meaning there exists other elements in the nullspace than the zero vector by Thm.3.2.2. This gives the following theorem:

Theorem 3.2.5. *A homogeneous system of linear equations with more variables than equations has nonzero solutions.*

Also, consider the inhomogeneous system of linear equations:

$$\begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \cdots & \vdots \\ A_{m1} & \cdots & A_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix}.$$

If the system has more equations than variables ($n < m$), then there are some choice of $c_1, \cdots, c_m$ such that the system has no solutions. Because we are doing a linear mapping $T : \mathbf{R}^n \to \mathbf{R}^m$, that is, from a bigger space to a smaller space. By Thm.3.2.2, T is not surjective. This gives the following theorem:

Theorem 3.2.6. *An inhomogeneous system of linear equations with more equations than variables has no solution for some choice of the constant terms.*

We conclude this section with a theorem that states when can a linear map be surjective. In some sense, this is similar to the negation of Cor.3.2.2. This is a supplementary theorem from Prof.Morton.

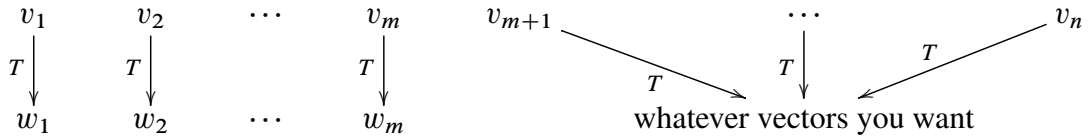
Theorem 3.2.7. *Suppose V, W are finite dimensional vector spaces. Then there exists a surjective linear map $T : V \to W$ if and only if* $\dim V \geq \dim W$.

Proof. First, suppose there exists a surjective linear map. Then we have range $T = W$. This implies that dim range $T = \dim V$. By Thm.3.2.4, we have

$$\dim W = \dim V - \underbrace{\dim \operatorname{null} T}_{\geq 0} \leq \dim V.$$

This completes the first half of the proof.

Next, suppose $\dim V \geq \dim W$. We need to construct a surjective linear map T. Choose a basis for V, say $v_1, \cdots, v_n$, and a basis of W, say $w_1, \cdots, w_m$. By our assumption $n \geq m$. Consider the following linear map T:



That is, we map $v_1, \cdots, v_m$ to $w_1, \cdots, w_m$ respectively, and map $v_{m+1} \cdots, v_n$ to whatever vectors in W we want. Thm.3.1.3 ensures that there exists a unique linear map that satisfies this. It is easy to check that this linear map is surjective. For any $w \in W$ we can express it in terms of the linear combination of the basis of W:

$$w = \sum_{i=1}^{m} a_i w_i = \sum_{i=1}^{m} a_i T(v_i).$$

Thus, for all $w \in W$, there exists $v \in V$ such that $T(v) = w$, where

$$v = \sum_{i=1}^{m} a_i v_i$$

for some chosen $a_i \in \mathbf{F}$. This concludes the proof. □

3.3 Matrices

We assume the reader is familiar with basic matrix operations (e.g. multiplying a matrix to a vector). We will slightly deviate from the structure of the book and the lecture. We will now explain why the concept of matrix comes up.

We associate each linear transformation with a matrix, because a matrix is very efficient in describing a linear transformation. Suppose now you have a linear transformation $T : V \to W$. Choose a basis of V:

$$\text{Basis } V = v_1, \cdots, v_n$$

and a basis of W:

$$\text{Basis } W = w_1, \cdots, w_m.$$

It is obvious that as long as we know $T(v_1), \cdots, T(v_n)$, we know $T(v)$ for any $v \in V$, since any $v \in V$ can be written as a linear combination of the basis vectors of V. Now suppose

$$v = a_1 v_1 + a_2 v_2 + \cdots + a_n v_n \tag{3.2}$$

and we also know that $T(v_1), \cdots, T(v_n) \in W$ can be written as linear combinations of w's.

$$\begin{aligned}
T(v_1) &= d_{11} w_1 + d_{21} w_2 + \cdots + d_{m1} w_m \\
T(v_2) &= d_{12} w_1 + d_{22} w_2 + \cdots + d_{m2} w_m \\
&\vdots \\
T(v_n) &= d_{1n} w_1 + d_{2n} w_2 + \cdots + d_{mn} w_m.
\end{aligned}$$

Now, we can compute $T(v)$ by applying T to both sides of (3.2):

$$\begin{aligned} T(v) &= a_1 T(v_1) + a_2 T(v_2) + \cdots + a_n T(v_n) \\ &= a_1 \sum_{i=1}^{m} d_{i1} w_i + a_2 \sum_{i=1}^{m} d_{i2} w_i + \cdots + a_n \sum_{i=1}^{m} d_{in} w_i \\ &= \sum_{j=1}^{n} a_j \sum_{i=1}^{m} d_{ij} w_i = \sum_{i=1}^{m} \underbrace{\left(\sum_{j=1}^{n} a_j d_{ij} \right)}_{\eta_i} w_i \\ &= \eta_1 w_1 + \eta_2 w_2 + \cdots + \eta_n w_n. \end{aligned}$$

The equation above tells us that given v express as a linear combination of the basis of V in (3.2), and with the knowledge of $T(v_1), \cdots, T(v_n)$, in particular d_{ij} for $1 \leq i \leq m$ and $1 \leq j \leq n$, we can compute $T(v)$, the transformed vector using the procedure above. The computation above should remind you of the following:

$$\begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & & & \vdots \\ d_{m1} & d_{m2} & \cdots & d_{mn} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_m \end{pmatrix}. \tag{3.3}$$

where the coefficients of the linear combination of v (Eqn.3.2) is the input vector, and the output vector is the coefficients of the linear combination of $T(v)$ in terms of w's, the basis of W.

This gives us a way to associate each linear transformation T with a matrix, which is an very efficient way of describing this transformation. That is, write across the top of the matirx the basis after transformation $Tv_1, \cdots, Tv_n$, and write along the left the basis vectors $w_1, \cdots, w_m$ for the vector space into which T maps (see book pg71).

Example 3.3.1. Suppose $T \in \mathcal{L}(\mathbf{R}^2, \mathbf{R}^3)$. We use the standard basis in $\mathbf{R}^2$: $(1, 0)$ and $(0, 1)$. Suppose $T(1, 0) = (1, 2, 7)$ and $T(0, 1) = (3, 5, 9)$, then we can write the associate matrix.

$$\begin{pmatrix} 1 & 3 \\ 2 & 5 \\ 7 & 9 \end{pmatrix}.$$

This is because we can write the transformed basis in $\mathbf{R}^2$ as linear combinations of the standard basis of $\mathbf{R}^3$ in the following way:

$$\begin{aligned} T(1, 0) &= 1(1, 0, 0) + 2(0, 1, 0) + 7(0, 0, 1) \\ T(0, 1) &= 3(1, 0, 0) + 5(0, 1, 0) + 9(0, 0, 1). \end{aligned}$$

As you can see, we first choose our basis in $\mathbf{R}^2$ and $\mathbf{R}^3$, we write down the "linear combination coefficients" for $T(k\text{th basis vector of } \mathbf{R}^2)$ in the kth column of the matrix.

Example 3.3.2. $D \in \mathcal{L}(\mathcal{P}_3(\mathbf{R}), \mathcal{P}_2(\mathbf{R}))$ is the differentiation map. The basis for $\mathcal{P}_3(\mathbf{R})$ is $1, x, x^2, x^3$, and the basis for $\mathcal{P}_2(\mathbf{R})$ is $1, x, x^2$. Given a vector $p \in \mathcal{P}_3(\mathbf{R})$:

$$p = c_0 1 + c_1 x + c_2 x^2 + c_3 x^3$$

We can write the differentiation matrix as

$$\mathcal{M}(D) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

By simple computation we have

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} c_1 \\ 2c_2 \\ 3c_3 \end{pmatrix}.$$

This gives the transformed vector in terms of the linear combination of basis of $\mathcal{P}_2(\mathbf{R})$

$$D(p) = c_1 1 + 2c_2 x + 3c_3 x^2.$$

Now we have some intuition, we can formally introduce the definition of a matrix.

Definition 3.3.1. Suppose $T \in \mathcal{L}(V, W)$ and $v_1, \cdots, v_n$ is a basis of V and $w_1, \cdots, w_m$ is a basis of W. The matrix of T with respect to these bases is the $m \times n$ matrix $\mathcal{M}(T)$ whose entries $A_{j,k}$ are defined by

$$T v_k = A_{1,k} w_1 + \cdots + A_{m,k} w_m.$$

If the bases are not clear from the context, then the notation

$$\mathcal{M}\left(T, (v_1, \ldots, v_n), (w_1, \ldots, w_m)\right)$$

is used.

Remark 3.3.1. It is very important to realize that $\mathcal{M}(T)$ depends on which bases you choose. Even for the same linear transformation T, different bases will associate with different matrices. Usually we use the standard basis.

Definition 3.3.2. Reader should already be familiar with matrix addition and scalar multiplication.

$$(A + C)_{j,k} = A_{j,k} + C_{j,k} \qquad (\lambda A)_{j,k} = \lambda A_{j,k}$$

Two theorems that are very easy to verify:

Theorem 3.3.1. *Suppose $S, T \in \mathcal{L}(V, W)$, then $\mathcal{M}(S+T) = \mathcal{M}(S) + \mathcal{M}(T)$.*

Theorem 3.3.2. *Suppose $S, T \in \mathcal{L}(V, W)$ and $\lambda \in \mathbf{F}$, then $\mathcal{M}(\lambda T) = \lambda \mathcal{M}(T)$.*

Notation 3.3.1. $\mathbf{F}^{m,n}$ denotes the set of all $m \times n$ matirx whose entries are in $\mathbf{F}$.

Now comes an interesting theorem.

Theorem 3.3.3. *$\mathbf{F}^{m,n}$ is a vector space with dimension mn.*

Proof. The proof that $\mathbf{F}^{m,n}$ is a vector space is left to the reader. Note that the additive identity of $\mathbf{F}^{m,n}$ is the matirx whose entires are all zeros.

It is obvious that the matrices that have 0 in all entires except a 1 in one entry is a basis of $\mathbf{F}^{m,n}$. There are mn such matrices. Thus $\dim \mathbf{F}^{m,n} = mn$. □

Though we assume the reader is already familiar with matirx multiplication, we still want to know the motivation for defining multiplication in such a way. The goal is to let the following theorem to be true:

Theorem 3.3.4. *Suppose $T \in \mathcal{L}(U, V)$ and Suppose $S \in \mathcal{L}(V, W)$. Then $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T)$.*

From the definition of the matrix, we know that the kth column of a matrix $\mathcal{M}(T)$ consists of the scalars needed to write Tv_k as a linear combination of $(w_1, \cdots, w_m)$:

$$Tv_k = \sum_{j=1}^{m} A_{j,k} w_j.$$

Now we define the matrix multiplication $\mathcal{M}(S)\mathcal{M}(T)$. Suppose $\mathcal{M}(S) = A$ and $\mathcal{M}(T) = C$. We have

$$\begin{aligned}
(ST)u_k &= S\left(\sum_{r=1}^{n} C_{r,k} v_r\right) \\
&= \sum_{r=1}^{n} C_{r,k} S v_r \\
&= \sum_{r=1}^{n} C_{r,k} \sum_{j=1}^{m} A_{j,r} w_j \\
&= \sum_{j=1}^{m} \left(\sum_{r=1}^{n} A_{j,r} C_{r,k}\right) w_j.
\end{aligned}$$

Thus $\mathcal{M}(ST)$ is the matrix whose entry in row j, colum k, equals

$$\mathcal{M}(ST)_{j,k} = \sum_{r=1}^{n} A_{j,r} C_{r,k}.$$

And the theorem above immediately follows.

3.4 Matrix of a Vector

In the last section we talked about the matrix representation of a linear transformation. Now we formalize what we said in the last section by defining matrix of a vector.

Definition 3.4.1. Suppose $v_1, \cdots, v_n$ is a basis of V. The matrix of v with respect to this basis is the $n \times 1$ matrix

$$\mathcal{M}(v) = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix},$$

where $c_1, \cdots, c_n$ are teh scalars such that

$$v = c_1 v_1 + \cdots + c_n v_n.$$

Remark 3.4.1. $\mathcal{M}(v)$ of a vector $v \in V$ depends on the basis $v_1, \cdots, v_n$ of V.

Example 3.4.1. The matrix of $2 - 7x + 5x^3$ with respect tothe standard basis of $\mathcal{P}_3(\mathbf{R})$ is

$$\begin{pmatrix} 2 \\ -7 \\ 0 \\ 5 \end{pmatrix}.$$

The matrix of a vector $x \in \mathbf{F}^n$ with respect to the standard basis is obtained by writting the coordinates of x as the entires in an $n \times 1$ matrix. In other words, if $x = (x_1, \cdots, x_n) \in \mathbf{F}^n$, then

$$\mathcal{M}(x) = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

From the definition of $\mathcal{M}(T)$ and $\mathcal{M}(v_k)$, we immediately have the following proposition.

Proposition 3.4.1. *Suppose $T \in \mathcal{L}(V, W)$ and $v_1 \cdots, v_n$ is a basis of V and $w_1, \cdots, w_m$ is a basis of W. Let $1 \leq k \leq n$. Then the kth column of $\mathcal{M}(T)$, which is denoted by $\mathcal{M}(T)_{\cdot,k}$, equals $\mathcal{M}(v_k)$.*

The following theorem should also be immediately obvious. It says linear maps act like matrix multiplication.

Theorem 3.4.1. *Suppose $T \in \mathcal{L}(V, W)$ and $v \in V$. Suppose $v_1 \cdots, v_n$ is a basis of V and $w_1, \cdots, w_m$ is a basis of W. Then*

$$\mathcal{M}(Tv) = \mathcal{M}(T)\mathcal{M}(v).$$

Proof. Suppose $v = \sum_{i=1}^{n} c_i v_i$, then

$$Tv = c_1 Tv_1 + \cdots + c_n Tv_n.$$

Since $\mathcal{M}$ is linear (from last section), we have

$$\begin{aligned}\mathcal{M}(Tv) &= c_1\mathcal{M}(Tv_1) + \cdots + c_n\mathcal{M}(Tv_n) \\ &= c_1\mathcal{M}(T)\cdot_1 + \cdots + c_n\mathcal{M}(T)\cdot_n \\ &= \mathcal{M}(T)\mathcal{M}(v).\end{aligned}$$

□

3.5 Invertibility and Isomorphic Vector Spaces

Definition 3.5.1. A linear map $T \in \mathcal{L}(V, W)$ is called invertible if there exists a linear map $S \in \mathcal{L}(W, V)$ such that $ST = I_V$ and $TS = I_W$. where I_V, I_W are identity maps on V, W respectively. S is called the inverse of T. We denote the inverse of T to be T^{-1}.

Theorem 3.5.1. *An invertible linear map has an unique inverse.*

Proof. Suppose the invertible map $T \in \mathcal{L}(V, W)$ has two inverses S_1, S_2. Then

$$S_1 = S_1 I = S_1 (TS_2) = (S_1 T) S_2 = IS_2 = S_2.$$

□

Theorem 3.5.2. *A linear map is invertible if and only if it is injective and surjective.*

Proof. Suppose $T \in \mathcal{L}(V, W)$ is invertible. Suppose $u, v \in V$ such that $Tu = Tv$. Then

$$u = T^{-1}(Tu) = T^{-1}(Tv) = v.$$

Hence T is injective. Let $w \in W$. Then $w = T(T^{-1}w)$. This means for all $w \in W$ there exists a $v = T^{-1}w \in V$ whose image is w. This implies that T is surjective. Thus we are halfway through.

Now suppose T is injective and surjective. We want to prove that T is invertible. Define Sw to be the unique element of V such that $T(Sw) = w$ (the uniqueness follows from surjectivity) for each $w \in W$. From the definition it is clear that $T \circ S$ is the identity map on W.

We also want to prove that $S \circ T$ is the identity map on V. Let $v \in V$, then

$$T((S \circ T)v) = (T \circ S)(Tv) = I(Tv) = Tv \Rightarrow (S \circ T)v = v.$$

The last step follows from the fact that T is injective. Hence $S \circ T$ is the identity map on V.

Finally, we show that S is linear. Suppose $w_1, w_2 \in W$. Then

$$T(Sw_1 + Sw_2) = T(Sw_1) + T(Sw_2) = w_1 + w_2.$$

Thus $Sw_1 + Sw_2$ is the unique element of V that T maps to $w_1 + w_2$. But since TS is the identity map, we know $T(S(w_1 + w_2)) = w_1 + w_2$. Hence $S(w_1 + w_2) = Sw_1 + Sw_2$.

Similarly, if $w \in W$ and $\lambda \in \mathbf{F}$,

$$T(\lambda Sw) = \lambda T(Sw) = \lambda w.$$

This implies $S(\lambda w) = \lambda Sw$, which follows from the injective and surjective properties of T. This shows that S is linear.

Hence S is the inverse of T, and T is thus invertible. This completes the proof. □

Remark 3.5.1. Thm.3.5.2 implies that only square matrices have inverses. Or equivalently, non-square matrices are not invertible. Consider an $m \times n$ matrix, from our discussion in section 3.2, we know that if $m > n$, the corresponding linear map is not surjective. If $m < n$, the corresponding linear map is not injective. Hence only when $m = n$ can matrices have inverses.

There is a easier way of arguing this point. Take $A \in \mathbf{F}^{m,n}$ and $A^{-1} \in \mathbf{F}^{n,p}$ Then $AA^{-1} \in \mathbf{F}^{m,p}$ and $A^{-1}A \in \mathbf{F}^{n,n}$ only when $m = p$. But $AA^{-1} = A^{-1}A = I$, and since the identity matrix is a square matrix, the only possible way is $m = p = n$.

This does not mean that all square matrices are invertible. Non-invertible square matrices are called singular.

Example 3.5.1. For two important examples of linear maps that are not invertible, see Example 3.57 in the book (pg81).

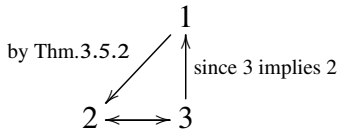
Definition 3.5.2. A linear map $T \in \mathcal{L}(V, V)$ from a vector space to itself is called an operator, or an endomorphism.

In Thm.3.5.2 we learned that for a linear map to be invertible it has to be both injective and surjective. But the next theorem tells us that if this linear map is an endomorphism, then only one condition of the two is needed for this map to be invertible.

Theorem 3.5.3. *If $T \in \mathcal{L}(V, V)$. Then the following are equivalent:*

1. *T is invertible;*
2. *T is injective;*
3. *T is surjective.*

Proof. We know that 1 implies 2. If we can show that 2 implies 3 and 3 implies 2, then we if we know 3 is true, then 2 is true. 2 and 3 together implies 1 by Thm.3.5.2. In other words, we have the following chain:



First, suppose 2 holds, so T is injective. Thus $\text{null } T = \{0\}$. From Thm.3.2.4, we have

$$\begin{aligned} \dim \text{range } T &= \dim V - \dim \text{null } T \\ &= \dim V. \end{aligned}$$

This implies that T is surjective since $\text{range } T = V$. Hence 2 implies 3.

Now, suppose 3 holds, so T is surjective. Again by Thm.3.2.4, we have

$$\begin{aligned} \dim \text{null } T &= \dim V - \dim \text{range } T \\ &= 0. \end{aligned}$$

Thus $\text{null } T = \{0\}$. Hence T is injective. Then T is both injective and surjective, by Thm.3.5.2, it is invertible. Thus 3 implies 1.

Finally, we already proved in Thm.3.5.2 that 1 implies 2. The full chain is achieved. □

Now we come to the concept of isomorphic vector spaces.

Definition 3.5.3. An isomorphism is an invertible linear map.

Definition 3.5.4. Two vector spaces are called isomorphic if there is an isomorphism from one vector space onto the other one.

The next theorem says dimension shows whether vector spaces are isomorphic. Note that from Thm.3.2.1 and Thm.3.2.2, we know that linear transformation from one vector space to another vector space with unequal dimensions are either not injective or not surjective, hence not invertible. So this theorem makes intuitive sense.

Theorem 3.5.4. *Two finite dimensional vector spaces over* $\mathbf{F}$ *are isomorphic if and only if they have the same dimension.*

Proof. First suppose V, W are isomorphic finite dimensional vector spaces. Thus there exists a isomorphism T from V onto W. Because T is invertible, we have $\text{null}\, T = \{0\}$ and $\text{range}\, T = W$.Then the formula

$$\dim V = \dim \text{null}\, T + \dim \text{range}\, T \Rightarrow \dim V = \dim W$$

proves the first half of the theorem (by Thm.3.2.4).

Next, suppose V, W are finite dimensional vector spaces with $\dim V = \dim W$. Let $v_1, \cdots, v_n$ be a basis of V and $w_1, \cdots, w_n$ be a basis of W. Let $T \in \mathcal{L}(V, W)$ be defined by

$$T(c_1 v_1 + \cdots + c_n v_n) = c_1 w_1 + \cdots + c_n w_n.$$

That is, $T(v_j) = w_j$. This can be done by Thm.3.1.3. Also, T is surjective since $\text{span}(w_1, \cdots, w_n) = W$. Furthermore, $\text{null}\, T = \{0\}$ since $w_1, \cdots, w_n$ is linearly independent. Thus T is injective. Since T is both injective and surjective, it is an isomorphism (Thm.3.5.2). This proves the theorem. □

The next theorem brings us up to another level of abstraction. It says $\mathcal{L}(V, W)$ and $\mathbf{F}^{m,n}$ is isomorphic.

Theorem 3.5.5. *Let* $v_1, \cdots, v_n$ *be a basis of* V *and* $w_1, \cdots, w_n$ *be a basis of* W. *Then* $\mathcal{M}$ *is an isomorphism between* $\mathcal{L}(V, W)$ *and* $\mathbf{F}^{m,n}$.

Proof. We already know that $\mathcal{M}$ is linear. Now we want to prove that it is both injective and surjective.

Suppose $T \in \mathcal{L}(V, W)$ and $\mathcal{M}(T) = 0$ (the zero matrix), then $Tv_k = 0$ for all $1 \leq k \leq n$. Since w's is a basis of V, this implies $T = 0$. Hence $\text{null}\, \mathcal{M} = \{0\}$. Thus $\mathcal{M}$ is injective.

To prove that $\mathcal{M}$ is surjective, take any $A \in \mathbf{F}^{m,n}$. Then we can have a linear transformation T whose representation matrix is A by defining $T : V \to W$ such that

$$Tv_k = \sum_{j=1}^{m} A_{j,k} w_j,$$

where $A_{j,k}$ is the entry in jth row and kth column of A. Obviously $\mathcal{M}(T) = A$. Thus $\text{range}\, \mathcal{M} = \mathbf{F}^{m,n}$. Thus $\mathcal{M}$ is surjective.

By Thm.3.5.2, $\mathcal{M}$ is an isomorphism. □

The theorem above has a immediate corollary:

Corollary 3.5.1. *Suppose V, W are finite dimensional. Then $\mathcal{L}(V, W)$ is finite dimensional and*

$$\dim \mathcal{L}(V, W) = (\dim V)(\dim W).$$

Proof. From Thm.3.5.5, we know that $\mathcal{L}(V, W)$ and $\mathbf{F}^{m,n}$ are isomorphic. Since we already know that $\mathbf{F}^{m,n}$ is finite dimensional (Thm.3.3.3), and Thm.3.5.4 implies $\mathcal{L}(V, W)$ is finite dimensional.

Also, from Thm.3.5.4, we know

$$\dim \mathcal{L}(V, W) = \dim \mathbf{F}^{m,n} = mn = (\dim V)(\dim W).$$

□

Chapter 4

Polynomials

Only the mediocre are supremely confident of their ability. The better you are, the higher the standards you set yourself — you can see beyond your immediate reach.

-Michael Atiyah

This chapter is about univariate polynomials, a topic that most readers are fairly familiar with even in high school. Polynomials, however, is a profound subject in algebra and algebraic geometry. But in this chapter we will only prove some seemingly obvious properties of a univariate polynomial.

First recall the following definition.

Definition 4.0.1. A function $p : \mathbf{F} \to \mathbf{F}$ is called a polynomial with coefficients in $\mathbf{F}$ if there exists $a_0 \cdots a_m \in \mathbf{F}$ such that

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m$$

for all $z \in \mathbf{F}$.

Theorem 4.0.1. *Suppose $a_0 \cdots a_m \in \mathbf{F}$. If*

$$a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m = 0$$

for every $z \in \mathbf{F}$, then $a_0 = \cdots = a_m = 0$.

Proof. We will prove that contrapositive: If not all the coefficients are 0, then we can find a $z \in \mathbf{F}$ such that the polynomial is not 0. Let

$$z = \frac{|a_0| + |a_1| + \cdots + |a_{m-1}|}{|a_m|} + 1.$$

Note that $z \geq 1$. And thus $z^j \leq z^{m-1}$ for $j \leq m-1$. Using the triangle inequality, we have

$$\begin{aligned}\left|a_0 + a_1 z + \cdots + a_{m-1} z^{m-1}\right| &\leq \left(|a_0| + |a_1| + \cdots + |a_{m-1}|\right) z^{m-1} \\ &< \left|a_m z^m\right|.\end{aligned}$$

This implies that $a_0 + a_1 z + \cdots + a_{m-1} z^{m-1} \neq -a_m z^m$. Hence $a_0 + a_1 z + \cdots + a_{m-1} z^{m-1} + a_m z^m \neq 0$. □

Remark 4.0.1. Recall that the degree of a zero polynomial is defined to be $-\infty$.

Now we come to the division algorithm of polynomials, which is similar to the division algorithm of integers.

Lemma 4.0.1. *Suppose $V_1, \cdots, V_m$ are finite dimensional vector spaces. Then $V_1 \times V_2 \times \cdots \times V_m$ is finite dimensional and*

$$\dim(V_1 \times V_2 \times \cdots \times V_m) = \dim V_1 + \dim V_2 + \cdots \dim V_m.$$

Proof. Choose a basis for each V_j. Take $v \in V_1 \times V_2 \times \cdots \times V_m$. Then we have

$$v = \sum_j \text{basis vector in the } j^{th} \text{ slot and 0 in other slots.}$$

This is clearly a basis of $V_1 \times V_2 \times \cdots \times V_m$ and has length $\sum_{j=1}^m \dim(V_j)$. □

Theorem 4.0.2. *Suppose $p, s \in \mathcal{P}(\mathbf{F})$ with $s \neq 0$. Then there exist unique polynomials $q, r \in \mathcal{P}(\mathbf{F})$ such that*

$$p = sq + r$$

and $\deg r < \deg s$. In integer division, q is called the quotient and r is called the remainder.

Proof. We use a linear algebraic proof. Let $n = \deg p$ and $m = \deg s$. If $n < m$, then take $q = 0$ and $r = p$, then we are done. Thus we assume that $n \geq m$.

Define a transformation

$$\mathcal{P}_{n-m}(\mathbf{F}) \times \mathcal{P}_{m-1}(\mathbf{F}) \to \mathcal{P}_n(\mathbf{F})$$

by

$$T(q, r) = sq + r.$$

It is easy to verify that T is a linear transformation. To prove the unique part, we prove that T is 1-1(injective), and to prove the exist part, we prove that T is onto (surjective).

If $(q, r) \in \text{null } T$, then $sq + r = 0$ or $sq = -r$. This implies that $q = 0$ and $r = 0$. Otherwise,

$$\deg -r \neq \deg sq = \deg s + \deg q \geq \deg s.$$

Thus $\dim \operatorname{null} T = 0$. This proves the unique part.

For the second part, according to Lem.4.0.1, it is easy to find that

$$\dim (P_{n-m}(\mathbf{F}) \times \mathcal{P}_{m-1}(\mathbf{F})) = (n-m+1) + (m-1+1) = n+1.$$

But Thm.3.2.4 suggests that $\operatorname{range} T = n+1$. Hence $\operatorname{range} T = \dim \mathcal{P}_n(\mathbf{F})$. Hence there exists $q \in \mathcal{P}_{n-m}(\mathbf{F})$ and $r \in \mathcal{P}_{m-1}(\mathbf{F})$ such that $p = T(q,r) = sq + r$ for all $p \in \mathcal{P}_n(\mathbf{F})$. □

Now we talk about zeros of polynomials.

Definition 4.0.2. $\lambda \in \mathbf{F}$ is called a zero of a polynomial $p \in \mathcal{P}(\mathbf{F})$ if $p(\lambda) = 0$.

Definition 4.0.3. $s \in \mathcal{P}(\mathbf{F})$ is called a factor of $p \in \mathbf{P}(\mathbf{F})$ if there exists a polynomial $q \in \mathcal{P}(\mathbf{F})$ such that $p = sq$.

Theorem 4.0.3. *Suppose $p \in \mathbf{P}(\mathbf{F})$ and $\lambda \in \mathbf{F}$. Then $p(\lambda) = 0$ if and only if there is a polynomial $q \in \mathcal{P}(\mathbf{F})$ such that*

$$p(z) = (z-\lambda)q(z)$$

for every $z \in \mathbf{F}$.

Proof. Suppose there is a polynomial $q \in \mathcal{P}(\mathbf{F})$ such that $p(z) = (z-\lambda)q(z)$ for all $z \in \mathbf{F}$. Then

$$p(\lambda) = (\lambda - \lambda)q(z) = 0,$$

which is obvious.

To prove the other direction, suppose $p(\lambda) = 0$. The polynomial $z - \lambda$ is of degree 1. Then by Thm.4.0.2 there exist a polynomial $q \in \mathcal{P}(\mathbf{F})$ and a number (of degree 0) $r \in \mathbf{F}$ such that

$$p(z) = (z-\lambda)q(z) + r$$

for every $z \in \lambda$. This equation, together with $p(\lambda) = 0$, implies that $r = 0$. Thus $p(z) = (z-\lambda)q(z)$ for all z, as desired. □

The next theorem says that a polynomial do not have too many zeros.

Theorem 4.0.4. *Suppose $p \in \mathcal{P}(\mathbf{F})$ is a polynomial with degree $m \geq 0$. Then p has at most m distinct zeros in $\mathbf{F}$.*

Proof. If $m = 0$, then $p(z) = a_0 \neq 0$, thus p has no zeros. If $m = 1$, $p(z) = a_0 + a_1 z$ with $a_1 \neq 0$. Thus p has exactly one zero, namely $-a_0/a_1$.

Suppose $m > 1$. Use induction on m. Assume that every polynomial with degree $m-1$ has at most $m-1$ distinct zeros. If p has no zeros in $\mathbf{F}$, we are done. If p has a zero $\lambda \in \mathbf{F}$, then there is a polynomial q of degree $m-1$ such that $p(z) = (z-\lambda)q(z)$. Since q has at most $m-1$ zeros, p has at most m zeros. □

We are fairly familiar with polynomials over $\mathbf{R}$, but now we talk about polynomials over $\mathbf{C}$. The next theorem is called the Fundamental Theorem of Algebra.

Theorem 4.0.5. *Every nonconstant polynomial with complex coefficients has a zero.*

Proof. The proof uses tools from complex analysis and is therefore omitted. See pg.124 on Axler. □

Theorem 4.0.6. *If $p \in \mathcal{P}(\mathbf{C})$ is a nonconstant polynomial, then p has a unique factorization of the form*

$$p(z) = c\,(z - \lambda_1) \cdots (z - \lambda_m),$$

where $c, \lambda_1, \ldots, \lambda_m \in \mathbf{C}$.

Proof. Let $p \in \mathcal{P}(\mathbf{C})$ and let $m = \deg p$. We use induction on m. If $m = 1$, the factorization clearly exists and is unique. Assume that $m > 1$ and the desired factorization exists and is unique for all polynomials of degree $m - 1$.

First we show that the desired factorization of p exists. By Thm.4.0.5, p has a zero λ. By Thm.4.0.3, we have a polynomial q such that

$$p(z) = (z - \lambda)q(z)$$

for all $z \in \mathbf{C}$. Because $\deg q = m - 1$, our induction hypothesis implies that q has desired factorization. Plug it into the equation above gives the desired factorization of p.

Next we show that the factorization is unique. Clearly c is uniquely determined as the coefficient of z^m in p. So we only need to show that except for the order, there is only one way to choose $\lambda_1 \cdots \lambda_m$. If

$$(z - \lambda_1) \cdots (z - \lambda_m) = (z - \tau_1) \cdots (z - \tau_m)$$

for all $z \in \mathbf{C}$, the left hand side will be zero when $z = \lambda_1$. Hence one of the τ's on the right hand side must be equal to λ_1. Without loss of generality, let $\tau_1 = \lambda_1$. For $z \neq \lambda_1$, we can divide both sides of the equation and get

$$(z - \lambda_2) \cdots (z - \lambda_m) = (z - \tau_2) \cdots (z - \tau_m)$$

for all $z \in \mathbf{C}$ except possibly $z = \lambda_1$. Actually, this holds for all $z \in \mathbf{C}$ because otherwise if we subtract the right hand side from the left hand side, we will get a nonzero polynomial that has infinitely many zeros (i.e. all $z \in \mathbf{C} \setminus \{\lambda_1\}$). We can repeat this process and set $z = \lambda_2, \lambda_3, \cdots, \lambda_m$ and we will see that except for the order, the λ's are the same as the τ's. This proves the uniqueness. □

Now we come back to the factorization of polynomials over $\mathbf{R}$.

Theorem 4.0.7. *Suppose $p \in \mathcal{P}(\mathbf{C})$ is a polynomial with real coefficients. If $\lambda \in \mathbf{C}$ is a zero of p, then so is $\bar{\lambda}$.*

Proof. Take the conjugate, and we find that

$$a_0 + a_1\lambda + \cdots + a_m\lambda^m = 0 \Rightarrow a_0 + a_1\bar{\lambda} + \cdots + a_m\bar{\lambda}^m = 0.$$

This completes the proof. □

Theorem 4.0.8. *Suppose $b, c \in \mathbf{R}$, then there is a polynomial factorization of the form*

$$x^2 + bx + c = (x - \lambda_1)(x - \lambda_2)$$

with $\lambda_1, \lambda_2 \in \mathbf{R}$ if and only if $b^2 \geq 4c$.

Proof. Completing the square:

$$x^2 + bx + c = \left(x + \frac{b}{2}\right)^2 + \left(c - \frac{b^2}{4}\right).$$

First suppose $b^2 < 4c$, then clearly the right hand side of the equation above is positive for every $x \in \mathbf{R}$. Hence this polynomial has no real zeros and thus cannot be factored in the form above with $\lambda_1, \lambda_2 \in \mathbf{R}$.

Conversely, suppose $b^2 \geq 4c$. Then there exists a real number d such that $d^2 = b^2/4 - c$. Then manipulating the equation gives

$$\begin{aligned} x^2 + bx + c &= \left(x + \frac{b}{2}\right)^2 - d^2 \\ &= \left(x + \frac{b}{2} + d\right)\left(x + \frac{b}{2} - d\right), \end{aligned}$$

which is the desired factorization. □

We conclude this chapter with one last theorem about factorization of a polynomial over the real field.

Theorem 4.0.9. *Suppose $p \in \mathcal{P}(\mathbf{R})$ is a nonconstant polynomial. Then p has a unique factorization of the form*

$$p(x) = c(x - \lambda_1)\cdots(x - \lambda_m)\left(x^2 + b_1x + c_1\right)\cdots\left(x^2 + b_Mx + c_M\right),$$

where $c, \lambda_1, \ldots, \lambda_m, b_1, \ldots, b_M, c_1, \ldots, c_M \in \mathbf{R}$, and $b_j^2 < 4c_j$ for each j (that is, the quadratics in the equation cannot be factored again).

Proof. Think of p as an element of $\mathcal{P}(\mathbf{C})$ since $\mathbf{R} \subset \mathbf{C}$. If all the zeros are real, then it means there are no quadratic factors. Hence we are done by Thm.4.0.6. Hence suppose p has a zero $\lambda \in \mathbf{C}$ with $\lambda \notin \mathbf{R}$. Then by the previous result $\bar{\lambda}$ is also a zero of p. Thus we can write

$$\begin{aligned} p(x) &= (x-\lambda)(x-\bar{\lambda})q(x) \\ &= \left(x^2 - 2(\operatorname{Re}\lambda)x + |\lambda|^2\right) q(x) \end{aligned}$$

for some $q \in \mathcal{P}(\mathbf{C})$ with degree two less than the degree of p. If we can prove that q has real coefficients, then by using induction on the degree of p, we can conclude that $(x-\lambda)$ appears in the factorization exactly as many times as $(x-\bar{\lambda})$.

To prove that q has real coefficients, rearrange the equation above we get

$$q(x) = \frac{p(x)}{x^2 - 2(\operatorname{Re}\lambda)x + |\lambda|^2}$$

for all $x \in \mathbf{R}$. The equation implies that $q(x) \in \mathbf{R}$ for all $x \in \mathbf{R}$. Thus we can write

$$q(x) = a_0 + a_1 x + \cdots + a_{n-2}x^{n-2},$$

where $n = \deg p$ and $a_0, \cdots, a_{n-2} \in \mathbf{C}$. We thus have

$$0 = \operatorname{Im} q(x) = (\operatorname{Im} a_0) + (\operatorname{Im} a_1)\, x + \cdots + (\operatorname{Im} a_{n-2})\, x^{n-2}$$

for all $x \in \mathbf{R}$. Since $1, x, x^2, \cdots$ is linearly independent, the corresponding coefficients must all be zero, which means $\operatorname{Im} a_0, \ldots, \operatorname{Im} a_{n-2} = 0$, and hence the coefficients of $q(x)$ are real, and the desired factorization exists.

Now we prove the uniqueness of this factorization. A factor of p of the form $x^2 + b_j x + c_j$ with $b_j^2 < 4c_j$ can be written uniquely as $(x-\lambda_j)(x-\bar{\lambda_j})$ with $\lambda_j \in \mathbf{C}$. Two different factorizations of p as an element of $\mathcal{P}(\mathbf{R})$ would lead to two different factorizations of p as an element of $\mathcal{P}(\mathbf{C})$, contradicting Thm.4.0.6. □

Chapter 5

Eigenvalues, Eigenvectors, and Invariant Subspaces

The definition of a good mathematical problem is the mathematics it generates rather than the problem itself.

-Andrew Wiles

In this chapter we are mainly concerned with endomorphism, $T \in \mathcal{L}(V, V)$. That is, the set of linear transformations that maps a vector space to itself. The main part of this chapter is about finding a matrix $\mathcal{M}(T)$ that is reasonably nice and efficient for describing T. As we will see, using eigenvectors as basis will typically give us really nice matrices.

Notation 5.0.1. The set of operators or endomorphisms $\mathcal{L}(V, V)$, is written as $\mathcal{L}(V)$ in Axler's book.

5.1 Invariant Subspaces

Given a finite dimensional vector space V, suppose we have a direct sum decomposition

$$V = U_1 \oplus U_2 \oplus \cdots \oplus U_m.$$

The endomorphism $T \in \mathcal{L}(V)$ will send V back to itself. Since $U_j \subset V$, if T maps each U_j back to U_j, then we only need to study $T|_{U_j}$, which denotes the linear map T restricted to the smaller domain U_j. Since each $U_j \subset V$ is a smaller vector space. This gives the following definition.

Definition 5.1.1. Suppose $T \in \mathcal{L}(V)$. A subspace U of V is called invariant under T if $u \in U$ implies $Tu \in U$.

Example 5.1.1. $\{0\}$, V, null T, range T are invariant subspaces of V under T.

5.2 Eigenvalues and Eigenvectors

Take any $v \in V$ such that $v \neq 0$, consider the set

$$U = \{\lambda v : \lambda \in \mathbf{F}\} = \text{span}(v).$$

Then U is a one dimensional subspace of V. Suppose U is a invariant subspace of V under T, then we have

$$Tv = \lambda v.$$

Conversely, if $Tv = \lambda v$ for some $\lambda \in \mathbf{F}$, then $\text{span}(v)$ is a one dimensional subspace of V invariant under T. Vectors v and scalars λ satisfying it are given special names:

Definition 5.2.1. Suppose $T \in \mathcal{L}(V)$. A number $\lambda \in \mathbf{F}$ is called an eigenvalue of T if there exists $v \in V$ such that $v \neq 0$ and $Tv = \lambda v$. v is called an eigenvector of T.

Proposition 5.2.1. *Let V be finite dimensional and $T \in \mathcal{L}(V)$. Let $\lambda \in \mathbf{F}$. Then the following are equivalent:*

1. *λ is an eigenvalue of T;*
2. *$T - \lambda I$ is not injective;*
3. *$T - \lambda I$ is not surjective;*
4. *$T - \lambda I$ is not invertible.*

Proof. If 1 is true, then $Tv = \lambda v$ for some nonzero v. Thus $(T - \lambda I)v = 0$ for some nonzero v, this implies 2. The converse is also true, hence $1, 2$ are equivalent. For an endomorphism, $2, 3, 4$ are equivalent by Thm.3.5.3. □

Theorem 5.2.1. *Suppose $T \in \mathcal{L}(V)$. Let $\lambda_1, \cdots, \lambda_m$ be distinct Eigenvalues of T and $v_1, \cdots, v_m$ be corresponding eigenvectors. Then $v_1, \cdots, v_m$ is linearly independent.*

Proof. Suppose eigenvectors is linearly dependent. Let k be the smallest positive integer such that

$$v_k \in \text{span}\left(v_1, \ldots, v_{k-1}\right).$$

The existence of k is gauranteed by Thm.2.2.1. Thus there exists $a_1, \cdots, a_{k-1} \in \mathbf{F}$ such that

$$v_k = a_1 v_1 + \cdots + a_{k-1} v_{k-1}. \tag{5.1}$$

Apply T on both sides:

$$\lambda_k v_k = a_1 \lambda_1 v_1 + \cdots + a_{k-1} \lambda_{k-1} v_{k-1}.$$

Multiply both sides of (5.1) by λ_k and subtract the equation above we get

$$0 = a_1 (\lambda_k - \lambda_1) v_1 + \cdots + a_{k-1} (\lambda_k - \lambda_{k-1}) v_{k-1}.$$

Since $v_1, \cdots, v_{k-1}$ is linearly independent and λ's are distinct, we have $a_i = 0$ for $1 \leq i \leq k-1$. But then (5.1) implies that $v_k = 0$, contradicting the definition of an eigenvector. □

Theorem 5.2.2. *Each endomorphism $T \in \mathcal{L}(V)$ on a finite dimensional vector space V has at most* $\dim V$ *distinct eigenvalues.*

Proof. Let $\lambda_1, \cdots \lambda_m$ be distinct eigenvalues of T. Then by Thm.5.2.1 the corresponding eigenvectors $v_1, \cdots, v_m$ is linearly independent. Thus $m \leq \dim V$ by Thm.2.2.2. □

Given a linear map T we can make more complicated transformations by using $p(T)$. Let $a_i \in \mathbf{F}$, then $p(T)$ is defined by

$$p(T) = a_0 I + a_1 T + a_2 T^2 + \cdots + a_m T^m.$$

We can also define the multiplicative properties of such an operator. Since polynomial multiplication is commutative, we can define

$$(pq)(T) = p(T)q(T) = q(T)p(T) = (qp)(T).$$

For the proof see Axler pg.144.

Now we have the central results about operators on complex vector spaces.

Theorem 5.2.3. *Every operator on a finite dimensional, nonzero, complex vector space has an eigenvalue.*

Remark 5.2.1. Note that this need not to be true for real vector spaces like $\mathbf{R}^2$. For example, the rotation transformation does not have any eigenvalues and eigenvectors (every vector gets rotated).

Proof. Take V as a complex vector space with $\dim V = n > 0$. Let T be an endomorphism. Choose $v \in V$ with $v \neq 0$. Then

$$v, Tv, T^2 v, \ldots, T^n v$$

is not linearly independent since its length is $n + 1 > n$ by Thm.2.2.2. Thus there exists $a_0, \cdots, a_n \in \mathbf{C}$, not all 0, such that

$$0 = a_0 v + a_1 T v + \cdots + a_n T^n v.$$

By the Fundamental Theorem of Algebra (Thm.4.0.5), there exists a factorization such that

$$\begin{aligned} 0 &= a_0 v + a_1 T v + \cdots + a_n T^n v \\ &= (a_0 I + a_1 T + \cdots + a_n T^n)\, v \\ &= c\,(T - \lambda_1 I) \cdots (T - \lambda_m I)\, v. \end{aligned}$$

Thus $T - \lambda_j I$ is not surjective for at least one j, which gives the desired eigenvalues. □

5.3 Upper-Triangular Matrices

The following proposition demonstrates a useful connection between upper-triangular matrices and invariant subspaces.

Proposition 5.3.1. *Suppose $T \in \mathcal{L}(V)$ and $v_1, \cdots, v_n$ is a basis of V. Then the following are equivalent:*

1. *the matrix of T with respect to $v_1, \cdots, v_n$ is upper triangular;*
2. *$Tv_j \in \operatorname{span}(v_1, \cdots, v_j)$ for each $j = 1, \cdots, n$;*
3. *$\operatorname{span}(v_1, \cdots, v_j)$ is invariant under T for each $j = 1, \cdots, n$.*

Proof. The equivalence of 1 and 2 follows immediate from the definition of a matrix. Obviously 3 implies 2. Now we prove that 2 implies 3.

Suppose 2 holds. Fix $j \in \{1, \cdots, n\}$. From 2 we know that

$$\begin{aligned} Tv_1 &\in \operatorname{span}(v_1) \subset \operatorname{span}(v_1, \ldots, v_j) \\ Tv_2 &\in \operatorname{span}(v_1, v_2) \subset \operatorname{span}(v_1, \ldots, v_j) \\ &\vdots \\ Tv_j &\in \operatorname{span}(v_1, \ldots, v_j). \end{aligned}$$

Thus if v is a linear combination of $v_1, \cdots, v_j$, then $Tv \in \operatorname{span}(v_1, \cdots, v_j)$, as desired. □

Theorem 5.3.1. *Let V be a finite dimensional complex vector space and $T \in \mathcal{L}(V)$. Then T has an upper-triangular matrix with respect to some basis of V.*

Proof. Use induction on **the dimension of** V. Clearly when $\dim V = 1$, scalar multiplication (λ) is upper-triangular, thus the desired result holds. Now suppose that $\dim V > 1$ and **the desired result holds for all complex vector spaces whose dimension is less than the dimension of** V.

Let λ be an eigenvalue of T, whose existence is gauranteed by Thm.5.2.3. Let

$$U = \text{range}(T - \lambda I).$$

Obviously $T - \lambda I$ is not injective, hence it is not surjective by Thm.3.5.3. Therefore, $\dim U < \dim V$.

We claim that U is invariant under T. To prove this, let $u \in U$. Then

$$Tu = (T - \lambda I)u + \lambda u.$$

Obviously $(T - \lambda I)u \in U$, which is the range of $T - \lambda I$. Also, since U is a subspace, $\lambda u \in U$. Thus their linear combination $(T - \lambda I)u + \lambda u \in U$. This proves the claim.

Thus $T|_U$ is an operator on U. By the induction hypothesis (note $\dim U < \dim V$), there is a basis $u_1, \cdots, u_m$ of U with respect to which $T|_U$ has an upper triangular matrix. Thus for each j we have

$$Tu_j = (T|_U)\left(u_j\right) \in \text{span}\left(u_1, \ldots, u_j\right)$$

by Prop.5.3.1. Extend this to a basis $u_1, \cdots, u_m, v_1, \cdots, v_n$ of V. For each k, we have

$$Tv_k = (T - \lambda I)v_k + \lambda v_k.$$

Since $(T - \lambda I)v_k \in U = \text{span}(u_1, \cdots, u_m)$, we claim that

$$Tv_k \in \text{span}\left(u_1, \ldots, u_m, v_1, \ldots, v_k\right)$$

for each k. Thus, T has an upper triangular matrix with respect to the basis $u_1, \cdots, u_m, v_1, \cdots, v_n$ of V. This concludes the proof. □

Proposition 5.3.2. *Suppose $T \in \mathcal{L}(V)$ has an upper-triangular matrix with respect to some basis of V. Then T is invertible if and only if all the entires on the diagonal of that upper-triangular matrix are nonzero.*

Proof. Suppose $v_1, \cdots, v_n$ is a basis of V with respect to T has an upper-triangular matrix

$$\mathcal{M}(T) = \begin{pmatrix} \lambda_1 & & & * \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}. \tag{5.2}$$

First suppose all the diagonal entires are nonzero. This implies that $Tv_1 = \lambda_1 v_1$. Because $\lambda_1 \neq 0$, we have $T(v_1/\lambda_1) = v_1$. Thus $v_1 \in \text{range}\, T$.

Now,

$$T\left(v_2/\lambda_2\right) = av_1 + v_2$$

for some $a \in \mathbf{F}$. The left hand side of the equation above and av_1 are both in range T. Hence $v_2 \in$ range T.

Similarly,

$$T(v_3/\lambda_3) = bv_1 + cv_2 + v_3,$$

and we can conclude that $v_3 \in$ range T. In fact, using the same procedure we can conclude that $v_1 \cdots, v_n \in$ range T. Since range is a subspace, their linear combination must also be in the range, giving range $T = V$. Thus T is surjective, implying the Invertibility by Thm.3.5.3.

Now suppose that T is invertible. This implies that $\lambda_1 \neq 0$, cause other wise $Tv_1 = 0$. Suppose $\lambda_j = 0$ for some $1 < j \leq n$. Then (5.2) implies that T maps span$(v_1, \cdots, v_j)$ into span$(v_1, \cdots, v_{j-1})$, which is one dimension less. Hence T is not injective and cannot be invertible, a contradiction, as desired. □

Theorem 5.3.2. *Suppose $T \in \mathcal{L}(V)$ has an upper-triangular matrix with respect to some basis of V. Then the eigenvalues of T are precisely the entires on the diagonal of that upper-triangular matrix.*

Proof. Say we have the upper-triangular matrix

$$\mathcal{M}(T) = \begin{pmatrix} \lambda_1 & & & * \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}.$$

Then

$$\mathcal{M}(T - \lambda I) = \begin{pmatrix} \lambda_1 - \lambda & & & * \\ & \lambda_2 - \lambda & & \\ & & \ddots & \\ 0 & & & \lambda_n - \lambda \end{pmatrix}.$$

Hence $T - \lambda I$ is not invertible if and only if $\lambda = \lambda_j$ for some j by Prop.5.3.2. Thus λ is an eigenvalue of T if and only if $\lambda = \lambda_j$ for some j. □

5.4 Eigenspaces and Diagonal Matrices

Definition 5.4.1. A diagonal matrix is a *square matrix* that is 0 everywhere except possibly along the diagonal.

Definition 5.4.2. Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$. The eigenspace of T corresponding to λ, denoted by $E(\lambda, T)$, is defined by

$$E(\lambda, T) = \text{null}(T - \lambda I),$$

which is the set of all eigenvectors of T corresponding to λ, along with the zero vector.

Example 5.4.1. This is a diagonal matrix:

$$\begin{pmatrix} 8 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix}.$$

Suppose this matrix is with respect to basis v_1, v_2, v_3. Then

$$E(8,T) = \operatorname{span}(v_1), \quad E(5,T) = \operatorname{span}(v_2, v_3).$$

Lemma 5.4.1. $\dim(U_1 \oplus \cdots \oplus U_n) = \dim U_1 + \cdots + \dim U_n$.

Proof. Choose a basis $u_i j$ for each U_j, where $1 \leq i \leq \dim U_j$. Then concatenate these bases into a list with length $\sum \dim U_i$. We claim that this is a basis of $U_1 \oplus \cdots \oplus U_n$. Clearly $u_i j \neq c u_k j$ for any $c \in \mathbf{F}$ for each j since we chose them as a basis of U_j. Also, $u_i p \neq c u_k q$ for any $c \in \mathbf{F}$ and for $u_i p \in U_p$ and $u_k q \in U_q$. Otherwise, there will be two ways of expressing a vector in $U_1 \oplus \cdots \oplus U_n$, one involving a vector $(1+c)u_k q \in U_q$, and one involving $u_i p + u_k q$, contradicting the fact that $U_1 \oplus \cdots \oplus U_n$ is a direct sum. □

Theorem 5.4.1. *Suppose V is finite dimensional and $T \in \mathcal{L}(V)$. Suppose $\lambda_1, \cdots, \lambda_m$ are distinct eigenvectors of T. Then the sum*

$$E(\lambda_1, T) + \cdots + E(\lambda_m, T)$$

is a direct sum. Further more,

$$\dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) \leq \dim V.$$

Proof. To show that $E(\lambda_1, T) + \cdots + E(\lambda_m, T)$ is a direct sum, suppose

$$u_1 + \cdots + u_m = 0,$$

where each $u_j \in E(\lambda, T)$ for some eigenvalue. Since eigenvectors corresponding to distinct eigenvalues are independent, this implies that each $u_j = 0$. By the criterion of a direct sum, this implies that $E(\lambda_1, T) + \cdots + E(\lambda_m, T)$ is a direct sum.

Now, according to the previous lemma, we have

$$\begin{aligned} \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) &= \dim(E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)) \\ &\leq \dim V. \end{aligned}$$

The inequality holds because the direct sum forms a subspace of V, which has a dimension less than that of V, by what we discovered in Chapter 2. □

Definition 5.4.3. An operator $T \in \mathcal{L}(V)$ is diagonalizable if the operator has a diagonal matrix with respect to some basis of V.

Theorem 5.4.2. *Suppose V is finite dimensional and $T \in \mathcal{L}(V)$. Let $\lambda_1, \cdots, \lambda_m$ denote the distinct eigenvalues of T. Then the following are equivalent:*

1. *T is diagonalizable;*
2. *T has a basis consisting of eigenvectors of T;*
3. *there exists 1-dimensional subspaces $U_1, \cdots, U_n$ of V, each invariant under T, such that*
$$V = U_1 \oplus \cdots \oplus U_n;$$
4. $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$;
5. $\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T)$.

Proof. 1 and 2 are clearly equivalent.

Suppose 2 holds and V has a basis $v_1, \cdots, v_n$ consisting of eigenvectors of T. Let $U_j = \text{span}(v_j)$ for each j. Clearly U_j is a one dimensional subspace invariant under T for each j. Because $v_1, \cdots, v_n$ is a basis of V, each vector in V can be written uniquely as a linear combination of v's, that is, it can be written uniquely as a sum $u_1 + \cdots + u_n$ where each $u_j \in U_j$. Thus $V = U_1 \oplus \cdots \oplus U_n$. Hence 2 implies 3.

Now suppose 3 holds. Thus there exist one dimensional subspaces $U_1, \cdots, U_n$ of V, each invariant under T such that $V = U_1 \oplus \cdots \oplus U_n$. Let $v_j \in U_j$ be a nonzero vector for each j. Then each v_j is an eigenvector of T. Because each vector V can be written uniquely as a sum $u_1 + \cdots + u_n$, where $u_j \in U_j$, we see that $v_1, \cdots, v_n$ is a basis of V. Thus 3 implies 2.

At this stage, we know that $1, 2, 3$ are equivalent. We finish the proof by showing 2 implies 4, 4 implies 5, and 5 implies 2.

Suppose 2 holds, then V has a basis consisting of eigenvectors of T. This directly implies that

$$V = E(\lambda_1, T) + \cdots + E(\lambda_m, T). \tag{5.3}$$

Hence 4 holds.

Now suppose 4 holds, (5.3) and the previous lemma immediately implies 5.

Finally, suppose 5 holds. That is,

$$\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T).$$

Choose a basis of each eigenspace. Put them together to form a list $v_1, \cdots, v_n$ of eigenvectors of T, where $n = \dim V$. To show that the list is linearly independent, suppose

$$a_1 v_1 + \cdots + a_n v_n = 0.$$

For each j, let u_j denote the sum of all the terms $a_k v_k$ such that $v_k \in E(\lambda_j, T)$. Thus each $u_j \in E(\lambda_j, T)$. We have

$$u_1 + \cdots + u_m = 0.$$

Because the eigenvectors corresponding to distinct eigenvalues are independent, this implies that each $u_j = 0$. Because each u_j is a sum of terms $a_k v_k$, where v_k's were chosen to be a basis of $E(\lambda_j, T)$, they are independent. Hence $a_k = 0$ for all k. Thus $v_1, \cdots, v_n$ are linearly independent and hence is a basis of V. Thus 5 implies 2.

Now we are done. □

Example 5.4.2. Not all operators are diagonalizable. For example, $T(w, z) = (z, 0)$ is not diagonalizable.

We conclude this chapter with one last theorem.

Theorem 5.4.3. *If $T \in \mathcal{L}(V)$ has $\dim V$ distinct eigenvalues, then T is diagonalizable.*

Proof. eigenvectors corresponding to different eigenvalues are linearly independent. And a linearly independent list of $\dim V$ vectors in V is a basis of V. With respect to this basis, T has a diagonal matrix. □

Chapter 6

Inner Product Spaces

If people do not believe that mathematics is simple, it is only because they do not realize how complicated life is.

-John von Neumann

The inner product was inspired by the properties of the dot product and can be seen as a generalization of the dot product. In this chapter we study the inner product and inner product spaces.

6.1 Inner Products and Norms

Definition 6.1.1. An inner product on V is a function that takes each ordered pair (u, v) of elements of V to a number $\langle u, v\rangle \in \mathbf{F}$ and has the following properties:

1. **positivity:** $\langle v, v\rangle \geq 0$ for all $v \in V$;
2. **definiteness:** $\langle v, v\rangle = 0$ if and only if $v = 0$;
3. **additivity in the first slot:** $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ for all $u, v, w \in V$;
4. **homogeneity in the first slot:** $\langle \lambda u, v\rangle = \lambda\langle u, v\rangle$ for all $\lambda \in \mathbf{F}$ and all $u, v \in V$;
5. **conjugate symmetry:** $\langle u, v\rangle = \overline{\langle v, u\rangle}$ for all $u, v \in V$.

Remark 6.1.1. The last property is worth noting. When doing a dot product, we have $x \cdot x = \|x\|^2$. But when $z \in \mathbf{C}$, we also want to think of $\|z\|^2$ as the inner product of z with itself. Since

$$\|z\| = \sqrt{|z_1|^2 + \cdots + |z_n|^2} = z_1\overline{z_1} + \cdots + z_n\overline{z_n},$$

Thus we want to define the inner product of $w, z \in \mathbf{C}^n$ to be

$$\langle w, z \rangle = w_1\overline{z_1} + \cdots + w_n\overline{z_n}.$$

Then it is clear that $\langle z, w \rangle$ is the complex conjugate of the equation above. Hence we have the fifth property. Every real number is equal to its complex conjugate. Thus if we are dealing with a real vector space, then this condition simply says that $\langle u, v \rangle = \langle v, u \rangle$ for all $v, w \in V$.

Example 6.1.1. Here are some examples of inner products. The *Euclidean inner product* on $\mathbf{F}^n$ is defined by

$$\langle (w_1, \ldots, w_n), (z_1, \ldots, z_n) \rangle = w_1\overline{z_1} + \cdots + w_n\overline{z_n}.$$

An inner product on $\mathbf{F}^n$ can also be defined as:

$$\langle (w_1, \ldots, w_n), (z_1, \ldots, z_n) \rangle = c_1 w_1\overline{z_1} + \cdots + c_n w_n\overline{z_n},$$

for some positive numbers $c_1, \cdots, c_n$.

An inner product can be defined on the vector space of real-valued continuous functions on the interval $[-1, 1]$ by

$$\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)dx.$$

An inner product can be defined on $\mathcal{P}(\mathbf{R})$ by

$$\langle p, q \rangle = \int_{0}^{\infty} p(x)q(x)e^{-x}dx.$$

Definition 6.1.2. An inner product space is a vector space V along with an inner product on V.

Notation 6.1.1. For the rest of this chapter, V denotes an inner product space over $\mathbf{F}$.

Proposition 6.1.1. *Here are some basic properties of an inner product.*

1. *For each fixed $u \in V$, the function that takes v to $\langle v, u \rangle$ is a linear map from V to $\mathbf{F}$.*
2. $\langle 0, u \rangle = 0$ *for every* $u \in V$.
3. $\langle u, 0 \rangle = 0$ *for every* $u \in V$.
4. $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$ *for all* $u, v, w \in V$.

5. $\langle u, \lambda v\rangle = \bar{\lambda}\langle u, v\rangle$ *for all* $\lambda \in \mathbf{F}$ *and* $u, v \in V$.

The proof of all these properties are not difficult and can be found in the book (pg.167-168).

Now we define the norm of a vector.

Definition 6.1.3. For $v \in V$, the norm of v, denoted by $\|v\|$, is defined by

$$\|v\| = \sqrt{\langle v, v\rangle}.$$

Example 6.1.2. With the Euclidean inner product, we have

$$\|(z_1, \ldots, z_n)\| = \sqrt{|z_1|^2 + \cdots + |z_n|^2}.$$

In the vector space of continuous real-valued functions on $[-1, 1]$, along with the inner product defined on it, we have

$$\|f\| = \sqrt{\int_{-1}^{1} (f(x))^2 dx}.$$

Proposition 6.1.2. *Here are some properties of the norm:*

1. $\|v\| = 0$ *if and only if* $v = 0$;
2. $\|\lambda v\| = |\lambda|\|v\|$ *for all* $\lambda \in \mathbf{F}$.

Again the proof is trivial according to the definition, hence it is omitted. It can be found on Axler pg.169.

Now wed come to a crutial definition.

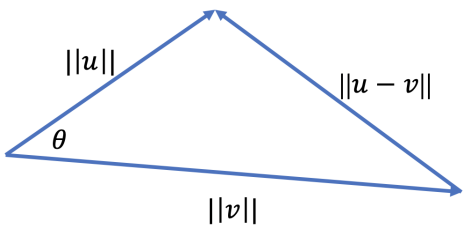
Definition 6.1.4. Two vectors $u, v \in V$ are called orthogonal if $\langle u, v\rangle = 0$.

Theorem 6.1.1. *If* u, v *are two nonzero vectors in* $\mathbf{R}^n$*, then*

$$\langle u, v\rangle = \|u\|\|v\| \cos\theta,$$

where θ *is the angle between* u *and* v*. Thus two vectors in* $\mathbf{R}^n$ *are orthogonal if and only if the cosine of the angle between them is 0, that is, if and only if they are perpendicular.*

Proof. Consider the following picture:



Using the law of cosine we get:

$$\|u - v\|^2 = \|u\|^2 + \|v\|^2 - 2\|u\|\|v\| \cos\theta.$$

Hence,

$$\begin{aligned} -2\|u\|\|v\| \cos\theta &= \langle u, u - v\rangle - \langle v, u - v\rangle - \langle u, u\rangle - \langle v, v\rangle \\ &= -\langle u, v\rangle - \langle v, u\rangle. \end{aligned}$$

Since in real vector space we have $\langle u, v\rangle = \langle v, u\rangle$, we immediately get the desired result. □

Proposition 6.1.3. *We consider the orthogonality of 0:*

1. *0 is orthogonal to every vector in V;*
2. *0 is the only vector in V that is orthogonal to itself.*

Proof. Once again, the proof is trivial (pg.170). □

Theorem 6.1.2. ***(Pythagorean Theorem)*** *Suppose $u, v \in V$ are orthogonal vectors. Then*

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2.$$

Proof.

$$\begin{aligned} \|u + v\|^2 &= \langle u + v, u + v\rangle \\ &= \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle \\ &= \|u\|^2 + \|v\|^2. \end{aligned}$$

□

Remark 6.1.2. The proof above holds if and only if

$$\langle u, v\rangle + \langle v, u\rangle = 2\,\mathrm{Re}\langle u, v\rangle = 0.$$

Thus the converse of the Pythagorean Theorem holds in real inner product spaces.

Now given a vector $u \in V$, we want to write it as a scalar multiple of another vector $v \in V$ plus a vector orthogonal to v. This gives rise to the orthogonal decomposition.

Theorem 6.1.3. ***(Orthogonal Decomposition)*** *Suppose $u, v \in V$ and $v \neq 0$. Let*

$$c = \frac{\langle u, v\rangle}{\|v\|^2} \qquad \textit{and} \qquad w = u - \frac{\langle u, v\rangle}{\|v\|^2} v.$$

Then

$$\langle w, v\rangle = 0 \quad \textit{and} \quad u = cv + w.$$

Proof. Straightforwardly,

$$\langle w, v\rangle = \langle u - cv, v\rangle = \langle u, v\rangle - c\|v\|^2.$$

Substituting the value of c immediately gives the desired result. □

Using the orthogonal decomposition, we can prove the Cauchy-Schwarz inequality, one of the most important inequalities in mathematics.

Theorem 6.1.4. ***Cauchy-Schwarz Inequality*** *Suppose $u, v \in V$, then*

$$|\langle u, v\rangle| \leq \|u\|\|v\|.$$

The inequality becomes an equality if and only if one of u, v is a scalar multiple of the other.

Proof. If $v = 0$, then both sides give 0. Thus we can assume that $v \neq 0$. Consider the orthogonal decomposition

$$u = \frac{\langle u, v\rangle}{\|v\|^2} v + w$$

given by Thm.6.1.3, where w is orthogonal to v. By the Pythagorean Theorem,

$$\begin{aligned}\|u\|^2 &= \left\| \frac{\langle u, v\rangle}{\|v\|^2} v \right\|^2 + \|w\|^2 \\ &= \frac{|\langle u, v\rangle|^2}{\|v\|^2} + \|w\|^2 \\ &\geq \frac{|\langle u, v\rangle|^2}{\|v\|^2}.\end{aligned}$$

This immediate gives the desired result. □

Example 6.1.3. Here are some examples of Cauchy-Schwarz inequality:

If $x, y \in \mathbf{R}^n$, where $x = (x_1, \cdots, x_n)$, $y = (y_1, \cdots, y_n)$, then

$$|x_1 y_1 + \cdots + x_n y_n|^2 \leq \left(x_1^2 + \cdots + x_n^2\right)\left(y_1^2 + \cdots + y_n^2\right).$$

If f, g are continuous real-valued functions on $[-1, 1]$, then

$$\left|\int_{-1}^{1} f(x)g(x)dx\right|^2 \leq \left(\int_{-1}^{1} (f(x))^2 dx\right)\left(\int_{-1}^{1} (g(x))^2 dx\right).$$

The next inequality is used extensively in analysis:

Theorem 6.1.5. ***Triangle Inequality*** *Suppose $u, v \in V$, then*

$$\|u + v\| \leq \|u\| + \|v\|.$$

This will become an inequality if and only if one of u, v is a nonnegative multiple of the other.

Proof.

$$\begin{aligned}
\|u + v\|^2 &= \langle u + v, u + v\rangle \\
&= \langle u, u\rangle + \langle v, v\rangle + \langle u, v\rangle + \langle v, u\rangle \\
&= \langle u, u\rangle + \langle v, v\rangle + \langle u, v\rangle + \overline{\langle u, v\rangle} \\
&= \|u\|^2 + \|v\|^2 + 2\,\mathrm{Re}\langle u, v\rangle \\
&\leq \|u\|^2 + \|v\|^2 + 2|\langle u, v\rangle| \\
&\leq \|u\|^2 + \|v\|^2 + 2\|u\|\|v\| \\
&= (\|u\| + \|v\|)^2,
\end{aligned}$$

where the second-to-last step follows from the Cauchy-Schwarz inequality. □

The final theorem is called the parallelogram equality. It states that in every parallelogram, the sum of the squares of the length of the diagonals equals the sum of squares of the lengths of the four sides.

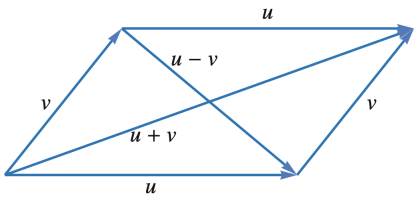


Figure 6.1: The parallelogram equality.

Theorem 6.1.6. *Suppose $u, v \in V$, then*

$$\|u + v\|^2 + \|u - v\|^2 = 2\left(\|u\|^2 + \|v\|^2\right).$$

Proof.

$$\begin{aligned}
\|u + v\|^2 + \|u - v\|^2 =& \langle u + v, u + v\rangle + \langle u - v, u - v\rangle \\
=& \|u\|^2 + \|v\|^2 + \langle u, v\rangle + \langle v, u\rangle \\
&+ \|u\|^2 + \|v\|^2 - \langle u, v\rangle - \langle v, u\rangle \\
=& 2\left(\|u\|^2 + \|v\|^2\right).
\end{aligned}$$

□

6.2 Orthonormal Bases

Definition 6.2.1. A list of vectors is called orthonormal if each vector in the list has norm 1 and is orthogonal to all the other vectors in the list. In other words, $e_1, \cdots, e_m$ in V is orthonormal if

$$\langle e_j, e_k \rangle = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k. \end{cases}$$

Proposition 6.2.1. *If $e_1, \cdots, e_m$ is an orthonormal list of vectors in V, then*

$$\|a_1 e_1 + \cdots + a_m e_m\|^2 = |a_1|^2 + \cdots + |a_m|^2 .$$

Proof. Each e_j has norm 1, this follows easily from repeated applications of the Pythagorean Theorem. □

Corollary 6.2.1. *Every orthonormal list of vectors is linearly independent.*

Proof. Suppose $e_1, \cdots, e_m$ is an orthonormal list of vectors in V such that

$$a_1 e_1 + \cdots + a_m e_m = 0.$$

Then by the previous proposition we see that $|a_1|^2 + \cdots + |a_m|^2 = 0$. Consider $A = (a_1, \cdots, a_m)$, Then $\langle A, A \rangle = \sum |a_i|^2$. This suggests that $A = 0$ and $a_i = 0$ for all i. Thus $e_1, \cdots, e_m$ is linearly independent. □

For any choice of basis $v_1, \cdots, v_n \in V$, we can write $v = \sum a_i v_i$ given any $v \in V$. The number a_i's, however, can be difficult to calculate. But the next theorem tells us that for an orthonormal basis, it is easy.

Theorem 6.2.1. *Suppose $e_1, \cdots, e_n$ is an orthonormal basis of V and $v \in V$. Then*

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n$$

and

$$\|v\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2 .$$

Proof. Because e's is a basis of V, we can write

$$v = a_1 e_1 + \cdots + a_n e_n.$$

Because $e_1, \cdots, e_n$ is orthonormal, taking the inner product of this equation with e_j gives $\langle v, e_j \rangle = a_j$ for all j. Thus the first equation holds.

The second equation immediately follows from the previous proposition. □

The next theorem tells us that given a linearly independent list, there exists an algorithm to convert them into an orthonormal list with the same span.

Theorem 6.2.2. ***(Gram-Schmidt Procedure)*** *Suppose $v_1, \cdots, v_m$ is a linearly independent list of vectors in V. Let $e_1 = v_1/\|v_1\|$. For $j = 2, \cdots, m$, define e_j inductively by*

$$e_j = \frac{v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}}{\left\| v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}\right\|}.$$

Then $e_1, \cdots, e_m$ is an orthonormal list of vectors in V such that

$$\operatorname{span}\left(v_1, \ldots, v_j\right) = \operatorname{span}\left(e_1, \ldots, e_j\right).$$

Proof. Use induction on j. For $j = 1$, $\operatorname{span}(v_1) = \operatorname{span}(e_1)$ since v_1 is a positive multiple of e_1.

Suppose $1 < j < m$, and suppose we have verified that

$$\operatorname{span}\left(v_1, \ldots, v_{j-1}\right) = \operatorname{span}\left(e_1, \ldots, e_{j-1}\right).$$

Note that $v_j \notin \operatorname{span}\left(v_1, \ldots, v_{j-1}\right)$. Thus $v_j \notin \operatorname{span}\left(e_1, \ldots, e_{j-1}\right)$. Hence we are not dividing zero in the definition of e_j given in the theorem. Dividing a vector by its norm produces a new vector with norm 1. Hence $\|e_j\| = 1$.

Let $1 \leq k < j$. Then

$$\begin{aligned}
\langle e_j, e_k\rangle &= \left\langle \frac{v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}}{\left\| v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}\right\|}, e_k \right\rangle \\
&= \frac{\langle v_j, e_k\rangle - \langle v_j, e_k\rangle}{\left\| v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}\right\|} \\
&= 0.
\end{aligned}$$

Thus $e_1, \cdots, e_j$ is an orthonormal list.

From the information given above, we see that

$$v_1, \cdots, v_{j-1} \in \operatorname{span}(v_1, \cdots, v_{j-1}) = \operatorname{span}(e_1, \cdots, e_{j-1}) \subset \operatorname{span}(e_1, \cdots, e_j).$$

And since $v_j \in \operatorname{span}(e_1, \cdots, e_j)$, we see that $\operatorname{span}(v_1, \cdots, v_j) \subset \operatorname{span}(e_1, \cdots, e_j)$. Both lists are linearly independent, thus both subspaces have dimension j, and hence they are equal. This completes the proof. □

Example 6.2.1. For an very interesting and important example of the application of the Gram-Schmidt procedure, see Axler pg.184.

Theorem 6.2.3. *Every finite dimensional inner product space has an orthonormal basis.*

Proof. Choose a basis of this space V and apply the Gram-Schmidt procedure to it, producing an orthonormal list with length $\dim V$. Since this orthonormal list is an independent list with the right length, it is a basis. □

Proposition 6.2.2. *Suppose V is finite dimensional. Then every orthonormal list of vectors in V can be extended to an orthonormal basis of V.*

Proof. Suppose $e_1, \cdots, e_m$ is an orthonormal list of vectors in V. They are linearly independent. Hence this list can be extended to a basis $e_1, \cdots, e_m, v_1, \cdots, v_n$ of V. Now apply the Gram-Schmidt procedure to this concatenated list, producing an orthonormal list $e_1, \cdots, e_m, f_1, \cdots, f_n$. Here the procedure leave the first m vectors unchanged since they are already orthonormal. Thus this is an orthonormal basis of V. □

Proposition 6.2.3. *Suppose $T \in \mathcal{L}(V)$. If T has an upper-triangular matrix with respect to some basis of V, then T has an upper-triangular matrix with respect to some orthonormal basis of V.*

Proof. Suppose T has an upper-triangular matrix with respect to some basis $v_1, \cdots, v_n$ of V. Then $\operatorname{span}(v_1, \cdots, v_j)$ is invariant under T for each j.

Apply the Gram-Schmidt procedure to this basis and produce an orthonormal basis $e_1, \cdots, e_n$ of V. Because

$$\operatorname{span}\left(e_1, \ldots, e_j\right) = \operatorname{span}\left(v_1, \ldots, v_j\right),$$

for each j, we conclude that $\operatorname{span}(e_1, \cdots, e_j)$ is invariant under T for each j. This is the desired basis with respect to which T has an upper-triangular matrix. □

Corollary 6.2.2. *Suppose V is a finite dimensional complex vector space and $T \in \mathcal{L}(V)$. Then T has an upper-triangular matrix with respect to some orthonormal basis of V.*

Proof. Recall that T has an upper-triangular matrix with respect to some basis of V. Now apply the proposition above. □

Definition 6.2.2. A linear functional on V is an element of $\mathcal{L}(V, \mathbf{F})$.

Theorem 6.2.4. *Suppose V is finite dimensional and φ is a linear functional on V. Then there is a unique vector $u \in V$ such that*

$$\varphi(v) = \langle v, u \rangle$$

for every $v \in V$.

Proof. First we show the existence of such a vector $u \in V$. Let $e_1, \cdots, e_n$ be an orthonormal basis of V. Then

$$\begin{aligned}
\varphi(v) &= \varphi\left(\langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n\right) \\
&= \langle v, e_1 \rangle \varphi\left(e_1\right) + \cdots + \langle v, e_n \rangle \varphi\left(e_n\right) \\
&= \left\langle v, \overline{\varphi\left(e_1\right)} e_1 + \cdots + \overline{\varphi\left(e_n\right)} e_n \right\rangle
\end{aligned}$$

for every $v \in V$. Thus setting

$$u = \overline{\varphi(e_1)}e_1 + \cdots + \overline{\varphi(e_n)}e_n$$

and we find the desired u. Now we prove that this u is unique. Suppose $u_1, u_2 \in V$ such that

$$\varphi(v) = \langle v, u_1 \rangle = \langle v, u_2 \rangle$$

for every $v \in V$. Then

$$0 = \langle v, u_1 \rangle - \langle v, u_2 \rangle = \langle v, u_1 - u_2 \rangle$$

for every $v \in V$. This implies that $u_1 - u_2 = 0$ or $u_1 = u_2$. This completes the proof. □

Example 6.2.2. Axler pg.188.

Remark 6.2.1. u is uniquely determined by φ in the theorem above. This u is the same regardless of what orthonormal basis of V is chosen.

6.3 Orthogonal Complements and Minimization

Definition 6.3.1. If U is a subset of V, then the orthogonal complement of U, denoted by $U^\perp$, is the set of all vectors in V that are orthogonal to every vector in U:

$$U^\perp = \{v \in V : \langle v, u \rangle = 0 \text{ for every } u \in U\}.$$

Proposition 6.3.1. *Some basic properties of orthogonal complement:*

1. *If U is a subset of V, then $U^\perp$ is a subspace of V.*
2. $\{0\}^\perp = V$.
3. $V^\perp = \{0\}$.
4. *If U is a subset of V, then $U \cap U^\perp = \{0\}$.*
5. *If $U, W \subset V$ and $U \subset W$, then $W^\perp \subset U^\perp$.*

Proof. The first three properties are easy to prove. We only prove the last two properties. For 4, take $v \in U \cap U^\perp$, then $\langle v, v \rangle = 0$, which implies that $v = 0$.

For 5, suppose $v \in W^\perp$, then $\langle v, u \rangle = 0$ for every $u \in W$, which implies that $\langle v, u \rangle = 0$ for every $u \in U$. Thus $v \in U^\perp$, and this concludes the proof. □

Theorem 6.3.1. *Suppose U is a finite dimensional subspace of V. Then*

$$V = U \oplus U^{\perp}.$$

Proof. We first show that $V = U + U^{\perp}$. Let $e_1, \cdots, e_m$ be a orthonormal basis of U. Obviously

$$v = \underbrace{\langle v, e_1\rangle e_1 + \cdots + \langle v, e_m\rangle e_m}_{u} + \underbrace{v - \langle v, e_1\rangle e_1 - \cdots - \langle v, e_m\rangle e_m}_{w},$$

where $u \in U$. Now, since

$$\langle w, e_j\rangle = \langle v, e_j\rangle - \langle v, e_j\rangle = 0,$$

it follows that w is orthogonal to every vector in $\text{span}(e_1, \cdots, e_m)$. Hence $w \in U^{\perp}$. This proves our frist claim. To show that this is a direct sum, recall from the previous proposition that $U \cap U^{\perp} = \{0\}$. This proves the theorem by Corollary 1.2.1. □

Corollary 6.3.1. *Suppose V is finite dimensional and U is a subspace of V. Then*

$$\dim U^{\perp} = \dim V - \dim U.$$

Theorem 6.3.2. *Suppose U is a finite dimensional subspace of V. Then*

$$U = (U^{\perp})^{\perp}.$$

Proof. We first show that $U \subset (U^{\perp})^{\perp}$. Suppose $u \in U$, then $\langle u, v\rangle = 0$ for every $v \in U^{\perp}$, this implies that $u \in (U^{\perp})^{\perp}$, completing the proof in one direction.

Now we prove that $U \supset (U^{\perp})^{\perp}$. Take $v \in (U^{\perp})^{\perp}$. By Thm.6.3.1, we can write $v = u + w$ where $u \in U$ and $w \in U^{\perp}$ (recall that $U \subset (U^{\perp})^{\perp}$). Then $v - u = w \in U^{\perp}$, but since both $v, u \in (U^{\perp})^{\perp}$, we have $v - u \in U^{\perp} \cap (U^{\perp})^{\perp}$. This means $v = u = 0$ or $v = u$. Since $u \in U$ we conclude $v \in U$ and $(U^{\perp})^{\perp} \subset U$. This proves the theorem. □

Definition 6.3.2. Suppose U is a finite dimensional vector subspace of V. The orthogonal projection of V onto U is the operator $P_U \in \mathcal{L}(V)$ defined as follows:

For every $v \in V$, write $v = u + w$, where $u \in U$ and $w \in U^{\perp}$. Then $P_U(v) = u$.

Proposition 6.3.2. *Suppose U is a finite dimensional subspace of V and $v \in V$. Then*

1. *$P_U \in \mathcal{L}(V)$;*
2. *$P_U(u) = u$ for every $u \in U$;*
3. *$P_U(w) = 0$ for every $w \in U^{\perp}$;*

4. range $P_U = U$;

5. null $P_U = U^\perp$;

6. $v - P_U(v) \in U^\perp$;

7. $P_U^2 = P_U$;

8. $\|P_U(v)\| \leq \|v\|$;

9. *for every orthonormal basis* $e_1, \cdots, e_m$ *of* U,

$$P_U v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m.$$

Proof. All proofs of these properties are easy and are thus omitted. For the last property, recall that

$$v = \underbrace{\langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m}_{u} + \underbrace{v - \langle v, e_1 \rangle e_1 - \cdots - \langle v, e_m \rangle e_m}_{w}$$

is a unique decomposition by Thm.6.3.1. And the claim immediately follows. □

The next theorem, which is very critical in Minimization problems (see Strang for an example about the least square).

Theorem 6.3.3. *Suppose* U *is a finite dimensional subspace of* V, $v \in V$, *and* $u \in U$. *Then*

$$\|v - P_U v\| \leq \|v - u\|.$$

The above is an equality if and only if $u = P_U v$.

Proof.

$$\begin{aligned} \|v - P_U v\|^2 &\leq \|v - P_U v\|^2 + \|P_U v - u\|^2 \\ &= \|(v - P_U v) + (P_U v - u)\|^2 \\ &= \|v - u\|^2. \end{aligned}$$

The second step holds because $v - P_U \in U^\perp$ and $P_U v - u \in U$. Thus we can use the Pythagorean Theorem. □

Example 6.3.1. See Axler pg.199 for an example of approximating $\sin x$ with a polynomial.

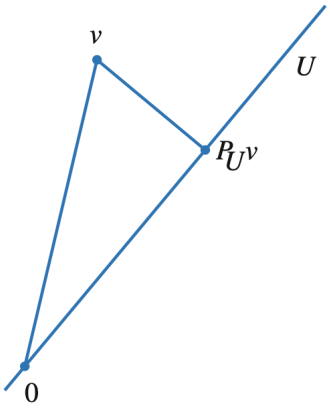


Figure 6.2: $P_U v$ is the closest point in U to v.