Dilla University

Department of Mathematics

Advanced Linear Algebra Lecture Note

by

Dereje Kifle (PhD)

Contents

Chapter 1

Vector Spaces

1.1 Vector Spaces

Definition 1.1.1. Let F be a field. A *vector space* over F is a nonempty set V together with two operations:

- addition: assigns to each pair $(u, v) \in V \times V$ a vector $u + v \in V$.
- scalar multiplication: assigns to each pair $(r, u) \in F \times V$ a vector ru in V.

Furthermore, the following properties must be satisfied:

- *Associativity of addition*: For all vectors $u, v, w \in V$, $u + (v + w) = (u + v) + w$.
- *Commutativity of addition*: For all vectors $u, v \in V$, $u + v = v + u$.
- *Existence of zero*: There is a zero vector $0 \in V$ with the property that $0 + u = u + 0 = u$ for all vectors $u \in V$.
- *Existence of additive inverses*: For each vector $u \in V$, there is a vector in V, denoted by $-u$, with the property that $u + (-u) = (-u) + u = 0$.
- *Properties of scalar multiplication*: For all scalars $a, b \in F$ and for all vectors $u, v \in V$,

$$a(u + v) = au + av$$
$$(a + b)u = au + bu$$
$$(ab)u = a(bu)$$
$$1u = u$$

In the above definition

- Elements of F (resp. V) are referred to as *scalars* (resp. *vectors*).
- The first four properties are equivalent to $(V, +)$ is an abelian group.
- V is sometimes called an *F-space*.
- If $F = \mathbb{R}$ (resp. $\mathbb{C}$), then V is a *real* (resp. *complex*) vector space.

1.2 Examples of a vector space

1) Let F be a field. The set V_F of all functions from F to F is a vector space over F, under the operations of ordinary addition and scalar multiplication of functions:

$$(f+g)(x) = f(x) + g(x), \text{ and } (af)(x) = a(f(x)).$$

2) The set $M_{m\times n}(F)$ of all $m \times n$ matrices with entries in a field F is a vector space over F, under the operations of matrix addition and scalar multiplication.

1.3 Subspaces, Linear combinations and Generators

Most algebraic structures contain substructures.

Definition 1.3.1. A *subspace* of a vector space V is a subset S of V that is a vector space in its own right under the operations obtained by restricting the operations of V to S. To indicate that S is a subspace of V, we use the notation $S \leq V$. If S is a subspace of V but $S \neq V$, we say that S is a proper subspace of V and it is denoted by $S < V$. The zero subspace of V is $\{0\}$.

Definition 1.3.2. Let S be a nonempty subset of a vector space V. A *linear combination* (L.C) of vectors in S is an expression of the form

$$a_1v_1 + \ldots + a_nv_n$$

where $v_1 \ldots v_n \in S$ and $a_1, \ldots, a_n \in F$. The scalars a_i are called the *coefficients* of the linear combination. A L.C is trivial if every coefficient a_i is zero. Otherwise, it is non trivial.

Theorem 1.3.3. *A non-empty subset S of a vector space V is a subspace of V if and only if S is closed under addition and scalar multiplication or equivalently, S is closed under linear combinations, that is,*

$$a, b \in F, u, v \in S \Longrightarrow au + bv \in S.$$

Example 1.3.4. Consider the vector space $V(n,2)$ of all binary n-tuples, that is, n-tuples of 0's and 1's. The weight $\mathcal{W}(v)$ of a vector $v \in V(n,2)$ is the number of non-zero coordinates in v. Let E_n be the set of all vectors in V of even weight. Then $E_n \leq V(n,2)$.

Proof. For vectors $u, v \in V(n,2)$, show that

$$\mathcal{W}(u+v) = \mathcal{W}(u) + \mathcal{W}(v) - 2\mathcal{W}(u \cap v) \tag{1.1}$$

where $u \cap v$ is the vector in $V(n,2)$ whose i^{th} component is the product of the i^{th} components of u and v, that is, $(u \cap v)_i = u_i \cdot v_i$. Let u and v be elements of E_n. Then by definition $\mathcal{W}(u)$ and $\mathcal{W}(v)$ are even which by (1.1) implies $\mathcal{W}(u+v)$ is even, that is, $u + v \in E_n$. Let $a \in \mathbb{F}_2$ and let $u \in E_n$. Clearly, $\mathcal{W}(au)$ is even which implies $au \in E_n$. Thus $E_n \leq V(n,2)$, known as the even weight subspace of $V(n,2)$. □

Definition 1.3.5. The subspace *spanned (or generated)* by a nonempty set S of vectors in V is the set of all linear combinations of vectors from S:

$$\langle S \rangle = \text{Span}(S) = \left\{ \sum_{i=1}^{n} r_i v_i \;\middle|\; r_i \in F, v_i \in S \right\}.$$

When $S = \{v_1, \ldots, v_n\}$ is a finite set, we use the notation $\langle v_1, \ldots, v_n \rangle$ or $\text{span}(v_1, \ldots, v_n)$. A set S of vectors in V is said to be span V, or generates V, if $V = \text{Span}(S)$.

Any superset of a spanning set is also a spanning set and all vector spaces have spanning set since V spans itself.

1.4 Linear Dependence and Independence of Vectors

Definition 1.4.1. Let V be a vector space. A nonempty set S of vectors in V is linearly independent (L.I) if for any distinct vectors $s_1, \ldots, s_n$ in S

$$a_1 s_1 + \ldots + a_n s_n = 0 \Rightarrow a_i = 0 \text{ for all } i.$$

In other words, S is L.I if the only L.C of vectors from S that is equal to 0 is the trivial L.C, all of whose coefficients are 0. If S is not L.I, it is said to be linearly dependent (LD).

A L.I set of vectors cannot contain the zero vector, since $1 \cdot 0 = 0$ violates the condition of linear independence.

Definition 1.4.2. Let S be a nonempty set of vectors in V. To say that a nonzero vector $v \in V$ is an *essentially unique L.C* of the vectors in S is to say that, up to the order of terms, there is one and only one way to express v as a L.C $\quad v = \sum_{i=1}^{n} a_i s_i$ where the s_i's are distinct vectors in S and the coefficients a_i are nonzero. More explicitly, $v \neq 0$ is an essentially unique L.C of vectors in S if $v \in \langle S \rangle$ and if whenever

$$v = a_1 s_1 + \ldots + a_n s_n \text{ and } v = b_1 t_1 + \ldots + b_m t_m$$

where the s_i's are distinct and t_i's are distinct and all coefficients are nonzero, then $m = n$ and after a reindexing of the $b_i t_i$'s if necessary, we have $a_i = b_i$ and $s_i = t_i$ for all $i = 1, \ldots, n$.

Theorem 1.4.3. *Let $S \neq \{0\}$ be a nonempty set of vectors in V. The following are equivalent:*

(a) S is L.I.

(b) Every nonzero vector $v \in \operatorname{span}(S)$ is an essentially unique L.C of the vectors in S.

(c) No vector in S is a L.C of other vectors in S.

Proof. (a) $\Rightarrow$ (b) Suppose that

$$0 \neq v = a_1 s_1 + \ldots + a_n s_n \text{ and } v = b_1 t_1 + \ldots + b_m t_m$$

where the s_i's are distinct and t_i's are distinct and the coefficients are nonzero. By subtracting and grouping s's and t's that are equal, we can write

$$\begin{aligned} 0 &= (a_{i_1} - b_{i_1})\, s_{i_1} + \ldots + (a_{i_k} - b_{i_1})\, s_{i_k} \\ &\quad + a_{i_{k+1}} s_{i_{k+1}} + \ldots + a_{i_n} s_{i_n} - b_{i_{k+1}} t_{i_{k+1}} - \ldots - b_{i_m} t_{i_m} \end{aligned}$$

(a) $\Rightarrow m = n = k$ and $a_{i_u} = b_{i_u}$ and $s_{i_u} = t_{i_u}$ for all $u = 1, \ldots, k$.

(b) $\Rightarrow$ (c) and (c) $\Rightarrow$ (a) is left as an exercise. $\square$

1.5 Direct sum and direct product of subspaces

Definition 1.5.1. Let $V_1, \ldots, V_n$ be vector spaces over a field F. The *external direct sum* of $V_1, \ldots, V_n$, denoted by $V_1 \boxplus \ldots \boxplus V_n$ is the vector space V whose elements are ordered n-tuples:

$$V = \{(v_1, \ldots, v_n) \mid v_i \in V_i, i = 1, \ldots, n\}$$

with componentwise operations

$$\begin{aligned}(u_1, \ldots, u_n) + (v_1, \ldots, v_n) &= (u_1 + v_1, \ldots, u_n + v_n) \text{ and} \\ r(u_1, \ldots, u_n) &= (ru_1, \ldots, ru_n) \qquad \text{for all } r \in F.\end{aligned}$$

Example 1.5.2. The vector space F^n is the external direct sum of n copies of F, that is, $F^n = F \boxplus \ldots \boxplus F$ where there are n summands on the right hand side.

The above construction can be generalized to any collection of vector spaces by generalizing the idea that an ordered n-tuple $(v_1, \ldots, v_n)$ is just a function

$$\begin{aligned}f : \{1, \ldots, n\} &\to \bigcup V_i, \\ i &\mapsto f(i)\,.\end{aligned}$$

Definition 1.5.3. Let $\mathcal{F} = \{V_i \mid i \in I\}$ be any family of vector spaces over F. The *direct product* of $\mathcal{F}$ is the vector space

$$\prod_{i \in I} V_i = \left\{ f \;:\; I \to \bigcup V_i \;\middle|\; f(i) \in V_i \right\}$$

thought of as a subspace of the vector space of all functions from I to $\bigcup V_i$.

Note that

$$\prod_{i \in I} V_i = \{v = (v_i)_{i \in I} \;\mid\; v_i \in V_i\} = \left\{ f \;:\; I \to \bigcup V_i \;\middle|\; f(i) \in V_i \right\}.$$

If we define addition and scalar multiplication by

$$\begin{aligned}v + w &= \left(f \;:\; I \to \bigcup V_i\right) + \left(g \;:\; I \to \bigcup V_i\right) \\ &= \left(f + g \;:\; I \to \bigcup V_i\right) \text{ and} \\ av &= a\left(f \;:\; I \to \bigcup V_i\right) \\ &= \left(af \;:\; I \to \bigcup V_i\right)\end{aligned}$$

or by

$$\begin{aligned}(v_i)_{i \in I} + (w_i)_{i \in I} &= (v_i + w_i)_{i \in I} \text{ and} \\ a(v_i)_{i \in I} &= (av_i)_{i \in I}\end{aligned}$$

Then the direct product $\prod_{i \in I} V_i$ is a vector space over F.

Definition 1.5.4. Let $\mathcal{F} = \{V_i \mid i \in I\}$ be any family of vector spaces over F. The support of a function $f : I \to \bigcup V_i$ is the set

$$\text{support}(f) = \{i \in I \mid f(i) \neq 0\}.$$

We say that f has *finite support* if $f(i) = 0$ for all but a finite number of $i \in I$.

Definition 1.5.5. Let $\mathcal{F} = \{V_i \mid i \in I\}$ be any family of vector spaces over F. The *external direct sum* of the family $\mathcal{F}$ is the vector space

$$\bigoplus_{i \in I}^{\text{ext}} V_i = \left\{ f : I \to \bigcup V_i \mid f(i) \in V_i, f \text{ has finite support} \right\}.$$

thought of as a subspace of the vector space of all functions from I to $\bigcup V_i$.

If $V_i = V$ for all $i \in I$,

- we denote the set of all functions from I to V by V^I, and
- we denote the set of all functions in V^I that have finite support by $\left(V^I\right)_0$.

In this case, we have

$$\prod_{i \in I} V = V^I \quad \text{and} \quad \bigoplus_{i \in I}^{\text{ext}} V = \left(V^I\right)_0.$$

Definition 1.5.6. A vector space V is the *internal direct sum* of a family $\mathcal{F} = \{S_i \mid i \in I\}$ of subspaces of V, written

$$V = \bigoplus \mathcal{F} \text{ or } V = \bigoplus_{i \in I} S_i$$

if the following hold:

(1) (*Join of the family*) V is the sum (join) of the family $V = \sum_{i \in I} S_i$

(2) (*Independence of the family*) For each $i \in I$,

$$S_i \bigcap \left(\sum_{j \neq i} S_j \right) = \{0\}.$$

In this case,

- each S_i is called a *direct summand* of V.
- if $\mathcal{F} = \{S_1, \ldots, S_n\}$ is a finite family, the direct sum is often written $V = S_1 \oplus \ldots \oplus S_n$.

- if $V = S \oplus T$, then T is called a *complement* of S in V.

If S and T are subspaces of V, then we may always say that the sum $S + T$ exists. However, to say that the direct sum of S and T exists or to write $S \oplus T$ is to imply that $S \cap T = \{0\}$. Thus, while the sum of two subspaces always exists, the direct sum of two subspaces does not always exist. Similar statements apply to families of subspaces of V.

Theorem 1.5.7. *Let $\mathcal{F} = \{V_i \mid i \in I\}$ be any family of vector spaces over F. The following are equivalent:*

*(1) (*Independence of the family*) For each $i \in I$,*

$$S_i \bigcap \left(\sum_{j \neq i} S_j \right) = \{0\}.$$

*(2) (*Uniqueness of expression for 0*) The zero vector cannot be written as a sum of nonzero vectors from distinct subspaces of $\mathcal{F}$.*

*(3) (*Uniqueness of expression*) Every nonzero vector $v \in V$ has a unique, except for order of terms, expression as a sum*

$$v = s_1 + \ldots + s_n$$

of nonzero vectors from distinct subspaces in $\mathcal{F}$.

Hence, a sum

$$V = \sum_{i \in I} S_i$$

is direct if and only if any one of (1)-(3) holds.

Proof. (1) $\Rightarrow$ (2) Suppose that (2) fails, that is,

$$0 = s_{j_1} + \ldots + s_{j_n}$$

where the nonzero vectors s_{j_i}'s are from distinct subspaces of S_{j_i}. Then $n > 1$ and, hence,

$$-s_{j_1} = s_{j_2} \ldots + s_{j_n}$$

which violates (1).
(2) $\Rightarrow$ (3) If (2) holds and

$$v = s_1 + \ldots + s_n = t_1 + \ldots + t_n$$

where the terms are nonzero and both the s_i's and the t_i's belong to distinct subspaces in $\mathcal{F}$. Then

$$0 = s_1 + \ldots + s_n - t_1 - \ldots - t_n.$$

Now, by collecting terms from the same subspaces, we may write

$$\begin{aligned} 0 &= (s_{i_1} - t_{i_1}) + \ldots + (s_{i_k} - t_{i_k}) \\ &\quad + s_{i_{k+1}} + \ldots + s_{i_n} - t_{i_{k+1}} - \ldots - t_{i_m}. \end{aligned}$$

Then (2) implies that $m = n = k$ and $s_{i_u} = t_{i_u}$ for all $u = 1, \ldots, k$.
(3) $\Rightarrow$ (1)

$$0 \neq v \in S_i \bigcap \left(\sum_{j \neq i} S_j \right) \Rightarrow v = s_i \in S_i \text{ and } s_i = s_{j_1} + \ldots + s_{j_n}$$

where $s_{j_k} \in S_{j_k}$ are nonzero which violates (3). □

Example 1.5.8. Let $A = \{(x, 0) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$ and let $B = \{(0, y) \in \mathbb{R}^2 \mid y \in \mathbb{R}\}$. Then $\mathbb{R}^2 = A \oplus B$ since $A \cap B = \{0\}$ and $\mathbb{R}^2 = A + B$. Any element (x, y) of $\mathbb{R}^2$ can be written as

$$(x, y) = (x, 0) + (0, y).$$

Proposition 1.5.9. *Suppose U and W are subspaces of the vector space V over a field F. Consider the map*

$$\alpha : U \oplus W \to V$$

defined by $\alpha(u, w) = u + w$. Then

- *α is injective if and only if $U \cap W = \{0\}$.*
- *α is surjective if and only if $U \cup W$ spans V.*

Example 1.5.10. Let $A = \{(x, 0) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$ and let $C = \{(y, y) \in \mathbb{R}^2 \mid y \in \mathbb{R}\}$. Then $\mathbb{R}^2 = A \oplus C$. To see this, note that the map

$$\begin{aligned} \alpha : A \oplus B &\to \mathbb{R}^2 \\ (x, y) &\mapsto x + y \end{aligned}$$

is injective since $A \cap C = \{0\}$. Moreover, α is a surjective map since any element (x, y) of $\mathbb{R}^2$ can be written as

$$(x, y) = \underbrace{(x - y, 0)}_{\in A} + \underbrace{(y, y)}_{\in C}.$$

Thus, by the above proposition $A \cup C$ spans $\mathbb{R}^2$.

Example 1.5.11. Let $A \in \mathcal{M}_n$ be a matrix. Then A can be written in the form

$$A = \frac{1}{2}(A + A^t) + \frac{1}{2}(A - A^t) = B + C \tag{1.2}$$

where A^t is the transpose of A. Verify that B is symmetric and C is skew-symmetric. Thus (1.2) is a decomposition of A as a sum of a symmetric matrix ($A^t = A$) and a skew-symmetric matrix ($A^t = -A$).

Exercise 1.5.12. Show that the sets Sym and SkewSym of all symmetric and skew-symmetric matrices in $\mathcal{M}_n$ are subspaces of $\mathcal{M}_n$.

Thus, we have

$$\mathcal{M}_n = \mathrm{Sym} + \mathrm{SkewSym}.$$

Furthermore, if $S, S' \in \mathrm{Sym}$ and $T, T' \in \mathrm{SkewSym}$ such that $S + T = S' + T'$, then the matrix

$$U = S - S' = T - T' \in \mathrm{Sym} \cap \mathrm{SkewSym}.$$

Hence, provided that $\mathrm{char}(F) \neq 2$, we must have $U = 0$. Thus,

$$\mathcal{M}_n = \mathrm{Sym} \oplus \mathrm{SkewSym}.$$

1.6 Bases of a Vector Space

Theorem and Definition 1.6.1. Let S be a set of vectors in V. The following are equivalent:

(i) S is L.I and spans V.

(ii) Every nonzero vector $v \in V$ is an essentially unique L.C of vectors in S.

(iii) S is a minimal spanning set, that is, S spans V but any proper subset of S does not span V.

(iv) S is a maximal L.I set, that is, S is L.I but any proper superset of S is not L.I.

A set of vectors in V that satisfies any (and hence all) of these conditions is called a basis for V.

Proof. (i) $\longleftrightarrow$ (ii) by Theorem 1.4.3.

(i) $\Rightarrow$ (iii) By given S is L.I and a spanning set, $V = \mathrm{span}(S)$. Suppose that any proper subset S' of S spans V. Let $s \in S - S'$. Since $s \in V$, s is a L.C of the vectors in S' which is a contradiction to the fact that S is L.I.

(iii) $\Rightarrow$ (i) If S is a minimal spanning set, then it must be L.I. For if not, some vector $s \in S$ would be a L.C of the other vectors in S, $S - \{s\}$. Then $S - \{s\}$ would be a proper spanning subset of S which is not possible.

(i) $\Leftrightarrow$ (iv): **exercise** □

Example 1.6.2.

(1) Find a basis of the subspace of $\mathbb{R}^3$ given by

$$V = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \;\middle|\; x - 2y + 5z = 0 \right\}.$$

Solution: Let $v = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ be any vector in V. Then

$$v = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2y - 5z \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2y \\ y \\ 0 \end{pmatrix} + \begin{pmatrix} -5z \\ 0 \\ z \end{pmatrix}$$
$$= y \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + z \begin{pmatrix} -5 \\ 0 \\ 1 \end{pmatrix}, \; y, z \in \mathbb{R}.$$

This shows that the set

$$\{u, v\} = \left\{ \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -5 \\ 0 \\ 1 \end{pmatrix} \right\}$$

spans V. It is easy to see that the set $\{u, v\}$ is L.I. Thus it is a basis for the subspace V of $\mathbb{R}^3$.

(2) The set $S = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$ is a basis of $\mathbb{R}^2$.

(3) The i^{th} standard vector in F^n is the vector e_i that has 0's in all coordinate positions except the i^{th}, where it has a 1. Thus,

$$e_1 = (1, \ldots, 0),\; e_2 = (0, 1, \ldots, 0), \ldots, e_n = (0, \ldots, 0, 1).$$

The set $\{e_1, \ldots, e_n\}$ is called the standard basis for F^n.

Theorem 1.6.3. *Let V be a nonzero vector space. Let I be a L.I set in V and let S be a spanning set in V containing I. Then there is a basis $\mathcal{B}$ for V for which $I \subset \mathcal{B} \subset S$. In particular,*

(1) Any vector space, except the zero space $\{0\}$, has a basis.

(2) Any L.I set in V is contained in a basis.

(3) Any spanning set in V contains a basis.

1.7 Dimension of a Vector Space

The following theorem says that if a vector space V has a finite spanning set S, then the size of any linearly independent set cannot exceed the size of S.

Theorem 1.7.1. *Let V be a vector space and assume that the vectors $v_1, \ldots, v_n$ are L.I and the vectors $s_1, \ldots, s_m$ span V. Then $n \leq m$.*

Corollary 1.7.2. *If V has a finite spanning set, then any two bases of V have the same size.*

Theorem 1.7.3. *If V is a vector space, then any two bases for V have the same cardinality.*

Definition 1.7.4. A vector space V is *finite-dimensional* if it is the zero space or if it has a finite basis. All other vector spaces are *infinite-dimensional*. The *dimension* of the a non-zero vector space V is the cardinality of any basis for V.

(a) The dimension of the zero space is 0.

(b) If a vector space V has a basis of cardinality k, we say that V is *k-dimensional* and write $\dim(V) = k$.

(c) If S is a subspace of V, then $\dim(S) \leq \dim(V)$. If in addition $\dim(S) = \dim(V) < \infty$, then $S = V$.

Theorem 1.7.5. *Let V be a vector space.*

1) If $\mathcal{B}$ is a basis for V and if $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ and $\mathcal{B}_1 \cap \mathcal{B}_2 = \emptyset$, then $V = \langle \mathcal{B}_1 \rangle \oplus \langle \mathcal{B}_2 \rangle$.

2) Let $V = S \oplus T$. If $\mathcal{B}_1$ is a basis for S and $\mathcal{B}_2$ is a basis for T, then $\mathcal{B}_1 \cap \mathcal{B}_2 = \emptyset$ and $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ is a basis for V.

Proof. 1) If $\mathcal{B}_1 \cap \mathcal{B}_2 = \emptyset$ and $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ is a basis for V, then $0 \notin \mathcal{B}_1 \cup \mathcal{B}_2$. But, if a nonzero vector $v \in \langle \mathcal{B}_1 \rangle \cap \langle \mathcal{B}_2 \rangle$, then $\mathcal{B}_1 \cap \mathcal{B}_2 \neq \emptyset$, a contradiction. Hence, $\{0\} = \langle \mathcal{B}_1 \rangle \cap \langle \mathcal{B}_2 \rangle$. Furthermore, since $\mathcal{B}_1 \cup \mathcal{B}_2$ is a basis for V and for $\langle \mathcal{B}_1 \rangle + \langle \mathcal{B}_2 \rangle$, we must have $V = \langle \mathcal{B}_1 \rangle + \langle \mathcal{B}_2 \rangle$. Thus, $V = \langle \mathcal{B}_1 \rangle \oplus \langle \mathcal{B}_2 \rangle$.

2) If $V = S \oplus T$, then $S \cap T = \{0\}$. Since $0 \notin \mathcal{B}_1 \cup \mathcal{B}_2$, we have $\mathcal{B}_1 \cap \mathcal{B}_2 = \emptyset$. Let $v \in V$. Then v has the form

$$a_1 u_1 + \ldots + a_n u_n + b_1 v_1 + \ldots + b_m v_m$$

for $u_1, \ldots, u_n \in \mathcal{B}_1$ and $v_1, \ldots, v_m \in \mathcal{B}_2$ which implies $v \in \langle \mathcal{B}_1 \cup \mathcal{B}_2 \rangle$ and thus $\mathcal{B}_1 \cup \mathcal{B}_2$ is a basis for V by Theorem 1.6.1. □

Theorem 1.7.6. *Let S and T be subspaces of a vector space V. Then*

$$\dim(S) + \dim(T) = \dim(S + T) + \dim(S \cap T).$$

In particular, if T is any complement of S in V, then

$$\dim(S) + \dim(T) = \dim(V) = \dim(S \oplus T).$$

Proof. Suppose that $\mathcal{B} = \{v_i \mid i \in I\}$ is a basis for $S \cap T$. Extend this to a basis $\mathcal{A} \cup \mathcal{B}$ for S and to a basis $\mathcal{B} \cup \mathcal{C}$ for T, where $\mathcal{A} = \{u_j \mid j \in J\}$ and $\mathcal{C} = \{w_k \mid k \in K\}$, $\mathcal{A} \cap \mathcal{B} = \emptyset$ and $\mathcal{C} \cap \mathcal{B} = \emptyset$.

Claim: $\mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$ is a basis for $S + T$.

Clearly, $\langle \mathcal{A} \cup \mathcal{B} \cup \mathcal{C} \rangle = S + T$. It remains to show that the set $\mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$ is L.I. To see this, suppose to the contrary that

$$\alpha_1 v_1 + \ldots + \alpha_n v_n = 0$$

where $v_i \in \mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$ and $\alpha_i \neq 0$ for all i. Then there must be vectors $v_i \in \mathcal{A} \cap \mathcal{C}$ since $\mathcal{A} \cup \mathcal{B}$ and $\mathcal{B} \cup \mathcal{C}$ are L.I. Now, isolating the terms involving the vectors from $\mathcal{A}$, say $v_1, \ldots, v_k$ without loss of generality, on one side of the equality shows that there is a nonzero vector in $x \in \mathcal{A} \cap \langle \mathcal{B} \cup \mathcal{C} \rangle$.

That is,

$$\begin{aligned} x = &\underbrace{a_1 v_1 + \ldots + a_k v_k}_{\in \text{ span}(A)} = \underbrace{a_{k+1} v_{k+1} + \ldots + a_n v_n}_{\in \text{ span}(\mathcal{B} \cup \mathcal{C})} \\ &\Rightarrow x \in \text{span}(\mathcal{A}) \cap \text{span}(\mathcal{B} \cup \mathcal{C}) \subset S \cap T = \langle \mathcal{B} \rangle \quad (\text{span}(\mathcal{A}) \subset S) \\ &\Rightarrow x \in \langle \mathcal{A} \rangle \cap \langle \mathcal{B} \rangle = \{0\} \\ &\Rightarrow x = 0, \text{ a contradiction.} \end{aligned}$$

Hence, $\mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$ is L.I and a basis for $S+T$. Now,

$$\begin{aligned}\dim(S) + \dim(T) &= |\mathcal{A} \cup \mathcal{B}| + |\mathcal{B} \cup \mathcal{C}| \\ &= |\mathcal{A}| + |\mathcal{B}| + |\mathcal{B}| + |\mathcal{C}| \\ &= |\mathcal{A}| + |\mathcal{B}| + |\mathcal{C}| + \dim(S \cap T) \\ &= \dim(S+T) + \dim(S \cap T),\end{aligned}$$

as desired. □

Chapter 2

Linear Transformations

2.1 Linear Transformations

Roughly speaking, a linear transformation is a function from one vector space to another that preserves the vector space operations.

Definition 2.1.1. Let V and W be vector spaces over a field F. A function $\tau : V \to W$ is a *linear transformation (L.T)* if

$$\tau(u+v) = \tau(u) + \tau(v) \text{ and } \tau(ru) = r\tau(u)$$

for all scalars $r \in F$ and vectors $u, v \in V$. The set of all linear transformations from $V \to W$ is denoted by $\mathcal{L}(V, W)$.

- A L.T from V to V is called a *linear operator* on V. The set of all linear operators on V is denoted by $\mathcal{L}(V)$.
- A linear operator on a real vector space is called a *real operator* and a linear operator on a complex vector space is called a *complex operator*.
- A L.T from V to the base field F (thought of as a vector space over itself) is called a *linear functional* on V. The set of all linear functions on V is denoted by V^* and called the *dual space* of V.

Definition 2.1.2. The following terms are also employed:

- **homomorphism** for L.T denoted also by $\mathrm{Hom}(V, W)$;
- **endomorphism** for L. operator denoted also by $\mathrm{End}(V)$;
- **monomorphism (embedding)** for injective L.T;
- **epimorphism** for surjective L.T;

- **isomorphism (invertible L.T)** for bijective L.T $\tau \in \mathcal{L}(V, W)$. In this case, we write $V \cong W$ to say that V and W are isomorphic. The set of all linear isomorphisms from V to W is denoted $\mathrm{GL}(V, W)$.

- **automorphism** for bijective L. operator. The set of all automorphisms of V is denoted $\mathrm{Auto}(V)$ or $\mathrm{GL}(V)$.

Example 2.1.3.

⊚ The derivative $D : V \to V$ is a linear operator on the vector space V of all infinitely differentiable functions on $\mathbb{R}$.

⊚ Let $V = \mathbb{R}^2$ and let $W = \mathbb{R}$. Define $L : V \to W$ by $f(v, w) = vw$. Is L a L.T?

⊚ The integral operator $\tau : F[x] \to F[x]$ defined by

$$\tau(f) = \int_0^x f(t)dt$$

is a linear operator on $F[x]$.

⊚ Let $V = \mathbb{R}^2$ and let $W = \mathbb{R}^3$. Define $L : V \to W$ by $L(v, w) = (v, w - v, w)$. Is L a L.T?

⊚ Let A be an $m \times n$ matrix over F. The function

$$\begin{aligned} \tau_A : F^n &\to F^m, \\ v &\mapsto Av, \end{aligned}$$

where all vectors are written as column vectors, is a L.T from $F^n \to F^m$.

Note:

⊖ The set $\mathcal{L}(V, W)$ is a vector space in its own right.

⊖ The *identity transformation*, $I_V : V \to V$, given by $I_V(x) = x$ for all $x \in V$. Clearly, since $I_V(av + bu) = av + bu = aI_V(u) + bI_V(v)$, I_V is L.T.

⊖ The *zero transformation*, $\tau_0 : V \to W$, given by $\tau_0(x) = 0$ for all $x \in V$, is a L.T.

⊖ If $\tau \in \mathcal{L}(V)$ such that $\tau^2 = \tau$, we call τ an *idempotent operator*.

2.2 Basic properties of Linear Transformations

In the following we collect a few facts about linear transformations:

Theorem 2.2.1. *Let τ be a L.T from a vector space V into a vector space W. Then*

i) $\tau(0) = 0$.

ii) $\tau(-v) = -\tau(v)$ *for all* $v \in V$.

iii) $\tau(u - v) = \tau(u) - \tau(v)$ *for all* $u, v \in V$.

iii) $\tau\left(\sum_{k=1}^{n} a_k v_k\right) = \sum_{k=1}^{n} a_k \tau(v_k)$ *for all* $v_1, \ldots, v_k \in V$.

Theorem 2.2.2. *Let V and W be vector spaces over over a field F and let $\mathcal{B} = \{v_i \mid i \in I\}$ is a basis for V. Then for any $\tau \in \mathcal{L}(V, W)$, we have $\operatorname{im}(\tau) = \langle \tau(\mathcal{B}) \rangle$.*

Theorem 2.2.3.

a) *The set $\mathcal{L}(V, W)$ is a vector space under ordinary addition of functions and scalar multiplication of functions by elements of F.*

b) *If $\sigma \in \mathcal{L}(U, V)$ and $\tau \in \mathcal{L}(V, W)$, then the composition $\tau\sigma$ is in $\mathcal{L}(U, W)$.*

c) *If $\tau \in \mathcal{L}(V, W)$ is bijective, then $\tau^{-1} \in \mathcal{L}(W, V)$.*

Proof. b) Since for all scalars $r, s \in F$ and vectors $u, v \in U$

$$\begin{aligned} \tau\sigma(ru + sv) &= \tau(r\sigma(u) + s\sigma(v)) && (\sigma \in \mathcal{L}(U, V)) \\ &= r(\tau\sigma(u)) + s(\tau\sigma(v)) && (\tau \in \mathcal{L}(V, W)) \\ &\Rightarrow \ \tau\sigma \in \mathcal{L}(U, W). \end{aligned}$$

c) Let $\tau : V \to W$ be a bijective L.T. Then $\tau^{-1} : W \to V$ is a well-defined function and since any two vectors w_1 and w_2 in W have the form $w_1 = \tau v_1$ and $w_2 = \tau v_2$, we have

$$\begin{aligned} \tau^{-1}(rw_1 + sw_2) &= \tau^{-1}(r\tau v_1 + s\tau v_2) \\ &= \tau^{-1}(\tau(rv_1 + sv_2)) \\ &= rv_1 + sv_2 \\ &= r\tau^{-1}(w_1) + s\tau^{-1}(w_2) \\ &\Rightarrow \ \tau^{-1} \in \mathcal{L}(V, W). \end{aligned}$$

□

One of the easiest ways to define a L.T is to give its values on a basis.

Theorem 2.2.4. *Let V and W be vector spaces and let $\mathcal{B} = \{v_i \mid i \in I\}$ be a basis for V. Then we can define a L.T $\tau \in \mathcal{L}(V, W)$ by specifying the values of $\tau(v_i)$ arbitrarily for all $v_i \in \mathcal{B}$ and extending τ to V by linearity, that is,*

$$\tau(a_1 v_1 + \ldots + a_n v_n) = a_1 \tau(v_1) + \ldots + a_n \tau(v_n).$$

This process defines a unique L.T, that is, if $\tau, \sigma \in \mathcal{L}(V, W)$ satisfying $\tau(v_i) = \sigma(v_i)$ for all $v_i \in \mathcal{B}$, then $\tau = \sigma$.

Note that if $\tau \in \mathcal{L}(V, W)$ and if S is a subspace of V, then the restriction $\tau|S$ of τ to S is a L.T from S to W.

2.3 The Kernel and Image of a L.T

Definition 2.3.1. Let $\tau \in \mathcal{L}(V, W)$.

- ⊙ The subspace
  $$\ker(\tau) = \{v \in V \mid \tau(v) = 0\}$$
  is called the *kernel* of τ.
- ⊙ The subspace
  $$\operatorname{im}(\tau) = \{\tau(v) \in W \mid v \in V\}$$
  is called the *image* of τ.
- ⊙ The dimension of $\ker(\tau)$ is called the *nullity* of τ and is denoted by $\operatorname{null}(\tau)$.
- ⊙ The dimension of $\operatorname{im}(\tau)$ is called the *rank* of τ and is denoted by $\operatorname{rk}(\tau)$.

Remark and Exercise 2.3.2.

- $\ker(\tau)$ is a subspace of V.
- $\operatorname{im}(\tau)$ is a subspace of W.

Theorem 2.3.3. *Let $\tau \in \mathcal{L}(V, W)$. Then*

1) τ is surjective if and only if $\operatorname{im}(\tau) = W$.

2) τ is injective if and only if $\ker(\tau) = \{0\}$.

Proof. 1) is clear. 2) Observe that,

$$\tau(v) = \tau(u) \Leftrightarrow \tau(v-u) = 0 \Leftrightarrow u - v \in \ker(\tau) = \{0\}$$

which implies $u = v$ and, hence, τ is injective. Conversely, suppose τ is injective and $u \in \ker(\tau)$. Then $\tau(u) = 0 = \tau(0)$ and, hence, $u = 0$. □

Theorem 2.3.4. *Let $\tau \in \mathcal{L}(V, W)$ be an isomorphism. Let $S \subset V$. Then*

a) S spans V if and only if $\tau(S) = \{\tau(u) \mid u \in S\}$ spans W.

b) S is L.I in V if and only if $\tau(S)$ is L.I in W.

c) S is a basis for V if and only if $\tau(S)$ is a basis for W.

Proof. a) $V = \langle S \rangle \Leftrightarrow W = \operatorname{im}(\tau) = \tau(\langle S \rangle) = \langle \tau(S) \rangle$ (since $\tau \in \mathrm{GL}(V, W)$).

b) By given S is L.I. For any $s_1, \ldots, s_n \in S$, we have

$$\sum_{i=1}^{n} a_i s_i = 0 \Leftrightarrow a_i = 0 \text{ for all } i,$$

which implies

$$\begin{aligned}
\tau\left(\sum_{i=1}^{n} a_i s_i\right) &= \sum_{i=1}^{n} a_i \tau(s_i) = 0 = \tau(0) \\
&\Rightarrow \sum_{i=1}^{n} a_i s_i = 0 \quad (\tau \in \mathrm{GL}(V, W)) \\
&\Rightarrow a_1 = \ldots = a_n = 0 \quad (S \text{ is L.I }) \\
&\Rightarrow \tau(S) \text{ is L.I} \quad (\text{ since this is true for all } s_i \in S).
\end{aligned}$$

Conversely, if $\tau(S)$ is L.I we have for any $\tau(s_1), \ldots, \tau(s_n) \in \tau(S)$

$$\begin{aligned}
0 = \sum_{i=1}^{n} a_i \tau(s_i) = \tau\left(\sum_{i=1}^{n} a_i s_i\right) &= \tau(0) \\
&\Rightarrow \sum_{i=1}^{n} a_i s_i = 0 \quad (\tau \in \mathrm{GL}(V, W)) \\
&\Rightarrow a_1 = \ldots = a_n = 0 \quad (\tau(S) \text{ is L.I }) \\
&\Rightarrow S \text{ is L.I}.
\end{aligned}$$

c) S is a basis for V iff, by a) and b), $\tau(S)$ is L.I in W and $W = \langle \tau(S) \rangle$ which implies $\tau(S)$ is a basis for W.

□

Isomorphisms Preserve Bases

An isomorphism can be characterized as a L.T $\tau : V \to W$ that maps a basis for V to a basis for W.

Theorem 2.3.5. *A L.T $\tau \in \mathcal{L}(V, W)$ is an isomorphism if and only if there is a basis $\mathcal{B}$ for V for which $\tau(\mathcal{B})$ is a basis for W. In this case, τ maps any basis of V to a basis of W.*

Proof. $\tau \in \mathrm{GL}(V, W) \Rightarrow \tau$ is bijective. Thus by Theorem 2.2.2 $\tau(\mathcal{B})$ is a basis for W. Conversely, if $\tau(\mathcal{B})$ is a basis for W, then for all $v \in V$, there exist unique elements $a_1, \ldots, a_n \in F$ and $u_1, \ldots, u_n$ such that $u = a_1u_1 + \ldots + a_nu_n$. Therefore,

$$\begin{aligned} 0 = \tau(u) &= a_1\tau(u_1) + \ldots + a_n\tau(u_n) \\ &\Rightarrow a_1 = \ldots = a_n = 0 \\ &\Rightarrow \ker(\tau) = \{0\} \\ &\Rightarrow \tau \text{ is injective.} \end{aligned}$$

Since $W = \langle \tau(\mathcal{B}) \rangle$, we have for all $w \in W$ there exist unique elements $a_1, \ldots, a_n \in F$ such that

$$w = a_1\tau(u_1) + \ldots + a_n\tau(u_n) = \tau(a_1u_1 + \ldots + a_nu_n).$$

So there exists $u = a_1u_1 + \ldots + a_nu_n \in V$ such that $w = \tau(u) \in \tau(V) = \mathrm{im}(\tau)$ which implies $W \subset \mathrm{im}(\tau)$. Clearly, $\mathrm{im}(\tau) \subset W$ and, hence, τ is surjective. Thus τ is bijective. □

Isomorphisms Preserve Dimension

The following theorem says that, upto isomorphism, there is only one vector space of any given dimension over a given field.

Theorem 2.3.6.

(i) *Let V and W be vector spaces over F. Then $V \cong W$ if and only if dim$(V) = \dim(W)$.*

(ii) *If n is a natural number, then any n-dimensional vector space over F is isomorphic to F^n.*

Proof. (i) $V \cong W \Rightarrow \exists \tau \in \mathrm{GL}(V, W)$. Thus $\mathcal{B}$ is a basis for V implies $\tau(\mathcal{B})$ is a basis for W and $\dim(V) = |\mathcal{B}| = |\tau(\mathcal{B})| = \dim(W)$. Conversely, if $\dim(V) = |\mathcal{B}_1| = |\mathcal{B}_2| = \dim(W)$, where $\mathcal{B}_1$ (resp. $\mathcal{B}_2$) is a basis for V (resp. W), then $\exists \tau \in \mathrm{GL}(\mathcal{B}_1, \mathcal{B}_2)$.

Extending τ to V by linearity defines a unique $\tau \in \mathcal{L}(V, W)$ by Theorem 2.2.4 and τ is an isomorphism because it is surjective and injective, that is, $\operatorname{im}(\tau) = W$ and $\ker(\tau) = \{0\}$.

(ii) Clear by (i). □

2.4 The Rank-Nullity Theorem

Lemma 2.4.1. *If V and W are vector spaces over a field F and $\tau \in \mathcal{L}(V, W)$, then any complement of the kernel τ is isomorphic to the range of τ, that is,*

$$V = \ker(\tau) \oplus \ker(\tau)^c \Rightarrow \ker(\tau)^c \cong \operatorname{im}(\tau)$$

where $\ker(\tau)^c$ is any complement of $\ker(\tau)$.

Proof. $V = \ker(\tau) \oplus \ker(\tau)^c \Rightarrow \dim(V) = \dim(\ker(\tau)) + \dim(\ker(\tau)^c)$. Let τ^c be the restriction of τ to $\ker(\tau)^c$. That is,

$$\tau^c : \ker(\tau)^c \to \operatorname{im}(\tau).$$

We claim that the map τ^c is bijective. □

To see this, note that the map τ^c is injective since

$$\ker(\tau^c) = \ker(\tau) \cap \ker(\tau)^c = \{0\}.$$

Clearly, $\operatorname{im}(\tau^c) \subset \operatorname{im}(\tau)$. For the reverse inclusion, if $\tau(v) \in \operatorname{im}(\tau)$, then since $v = u + w$ for $u \in \ker(\tau)$ and $w \in \ker(\tau)^c$, we have

$$\tau(v) = \tau(u) + \tau(w) = \tau(w) = \tau^c(w) \in \operatorname{im}(\tau^c).$$

Thus $\operatorname{im}(\tau^c) = \operatorname{im}(\tau)$ which implies

$$\tau^c : \ker(\tau)^c \to \operatorname{im}(\tau)$$

is an isomorphism.

Theorem 2.4.2 (Rank-Nullity Theorem). *Let V and W be vector spaces over a field F and let $\tau \in \mathcal{L}(V, W)$. Then*

$$\dim(\ker(\tau)) + \dim(\operatorname{im}(\tau)) = \dim(\mathrm{V})$$

or in other notation

$$\operatorname{rk}(\tau) + \operatorname{null}(\tau) = \dim(\mathrm{V})$$

Proof.

$$\begin{aligned}\dim(V) &= \dim(\ker(\tau)) + \dim(\ker(\tau)^c) \\ &= \dim(\ker(\tau)) + \dim(\operatorname{im}(\tau)) \quad (\text{Lemma } 2.4.1) \\ &= \operatorname{null}(\tau) + \operatorname{rk}(\tau)\end{aligned}$$

which completes the proof. □

Corollary 2.4.3. *Let V and W be vector spaces over a field F and $\tau \in \mathcal{L}(V, W)$. If $\dim(V) = \dim(W)$, then the following are equivalent:*

i) τ is injective.

ii) τ is surjective.

iii) $rk(\tau) = \dim(V)$.

Proof. By the Rank-Nullity Theorem, $\operatorname{rank}(\tau) + \operatorname{null}(\tau) = \dim(V)$ and , we have

$$\begin{aligned}\tau \text{ is 1-1} &\overset{Thm\ 2.3.3}{\Leftrightarrow} \ker(\tau) = \operatorname{null}(\tau) = \{0\} \\ &\overset{\text{R-N Thm}}{\Leftrightarrow} \dim(\operatorname{im}(\tau)) = \operatorname{rk}(\tau) = \dim(V) \overset{\text{assu.}}{=} \dim(V) \\ &\Leftrightarrow \operatorname{im}(\tau) = W \\ &\Leftrightarrow \tau \text{ is onto which completes the proof.}\end{aligned}$$

□

2.5 Linear Transformations from F^n to F^m

Recall that for any $m \times n$ matrix A over F the multiplication map

$$\tau_A(v) = A\tau$$

is a L.T. In fact, any L.T $\tau \in \mathcal{L}(F^n, F^m)$ has this form, that is, τ is just multiplication by a matrix, for we have

$$(\tau(e_1)|\cdots|\tau(e_n))\, e_i = (\tau(e_1)|\cdots|\tau(e_n))^{(i)} = \tau(e_i)$$

and so $\tau = \tau_A$, where $A = (\tau(e_1)|\cdots|\tau(e_n))$. Here any vector $v \in F^n$ can be written as

$$v = (v_1, \ldots, v_n) = v_1 e_1 + \ldots + v_n e_n.$$

The map τ sends any vector $(v_1, \ldots, v_n)$ to $(w_1, \ldots, w_m)$, that is,

$$\begin{aligned} \tau : F^n &\to F^m \\ v &\mapsto Av := w, \end{aligned}$$

where

$$\begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}.$$

The matrix A is called the *standard matrix* of τ.

Example 2.5.1. Consider the linear transformation

$$\begin{aligned} \tau : F^3 &\to F^3 \\ (x, y, z) &\mapsto (x - 2y, z, x + y + z). \end{aligned}$$

Then we have in column form

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & -2 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

So the standard matrix of τ is

$$A = \begin{pmatrix} 1 & -2 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Theorem 2.5.2.

a) If A is an $m \times n$ matrix over F, then $\tau_A \in \mathcal{L}(F^n, F^m)$.

b) If $\tau \in \mathcal{L}(F^n, F^m)$, then $\tau = \tau_A$, where $A = (\tau e_1 | \cdots | \tau e_n)$. The matrix A is called the matrix *of τ.*

f $A \in M_{m,n}$, then since the image of τ_A is the column space of A, we have

$$\dim(\ker(\tau_A)) + \operatorname{rk}(A) = \dim(F^n).$$

Theorem 2.5.3. *Let A be an $m \times n$ matrix over F.*

1) $\tau_A : F^n \to F^m$ is injective iff $rk(A) = n$.

2) $\tau_A : F^n \to F^m$ is surjective iff $rk(A) = m$.

Example 2.5.4. Find $\mathrm{rk}(A)$ and $\mathrm{null}(A)$ for

$$A = \begin{pmatrix} -2 & -5 & 8 & 0 & -17 \\ 1 & 3 & -5 & 1 & 5 \\ 3 & 11 & -19 & 7 & 1 \\ 1 & 7 & -13 & 5 & -3 \end{pmatrix}.$$

Note that

◎ the rank of A equals the number of nonzero rows in the row echelon form, which equals the number of leading entries.

◎ the nullity of A equals the number of free variables in the corresponding system, which equals the number of columns without leading entries.

The reduced echelon form of A is

$$\begin{pmatrix} 1 & 3 & -5 & 1 & 5 \\ 0 & 1 & -2 & 2 & -7 \\ 0 & 0 & 0 & 1 & -5 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

and thus $\mathrm{rk}(A) = 3$.

Note that the reduced echelon form of the above matrix is

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & -2 & 0 & 3 \\ 0 & 0 & 0 & 1 & -5 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

To determine nullity, we need to find a basis for the solution set of $Ax = 0$, that is, to find the solution set of the system of equations

$$\begin{aligned} x_1 + x_2 + x_5 &= 0 \\ x_2 - 2x_3 + 3x_5 &= 0 \\ x_4 - 5x_5 &= 0\,. \end{aligned}$$

In this equation, the leading variables are x_1, x_2 and x_4. Hence, the free variables are x_3 and x_5 which implies $\mathrm{null}(A) = 2$. In fact, we can write the solution in parametric form as follows:

$$x_1 = -s - t,\ x_2 = 2s - 3t,\ x_3 = s,\ x_4 = 5t,\ x_5 = t.$$

From this we have,

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = s \begin{pmatrix} -1 \\ 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ -3 \\ 0 \\ 5 \\ 1 \end{pmatrix}.$$

Thus, $\text{null}(A) = 2$, as desired. Moreover, we have

$$\dim(\ker(\tau_A)) + \text{rk}(A) = 5 = \dim(\mathbb{R}^5).$$

Example 2.5.5. Find the L.T $\tau : \mathbb{R}^2 \to \mathbb{R}^2$ that perpendicularly projects both of the vectors e_1 and e_2 onto the line $x_1 = x_2$.

Solution:

2.6 Matrix Representation of a L.T

Definition 2.6.1. Let V be a vector space of dimension n. An *ordered basis* for V is an ordered n-tuple $(v_1, \dots, v_n)$ of vectors for which the set $\{v_1, \dots, v_n\}$ is a basis for V.

Definition 2.6.2. Let $\mathcal{B} = (v_1, \dots, v_n)$ be an ordered basis for V and let $v \in V$. Then there exist unique scalars $r_1, \dots, r_n \in F$ such that $v = \sum_{i=1}^{n} r_i v_i$. These scalars are called the *coordinates* of v w.r.t. $\mathcal{B}$. Define the map $\phi_{\mathcal{B}} : V \to F^n$ by

$$\phi_{\mathcal{B}}(v) = [v]_{\mathcal{B}} = \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix}.$$

The vector $[v]_{\mathcal{B}} = (r_1, \dots, r_n) \in F^n$ is called the *coordinate vector* of v w.r.t. $\mathcal{B}$ (or *coordinate matrix* when the vector $[v]_{\mathcal{B}}$ is viewed as column matrix) and the map $\phi_{\mathcal{B}}$ is called the *coordinate map.*

Note that the coordinate map $\phi_{\mathcal{B}}$ is a bijective L.T. τ is linear since

$$[r_1 v_1 + \ldots + r_n v_n]_{\mathcal{B}} = r_1 [v_1]_{\mathcal{B}} + \ldots + r_n [v_n]_{\mathcal{B}}.$$

Example 2.6.3.

1) Let $v \in V$. If $\mathcal{B} = (e_1, \dots, e_n) \in V$, then $[v]_{\mathcal{B}} = v$.

2) If $v = (3, 4)$ and $\mathcal{B} = ((1, -1), (1, 1))$, then $[v]_{\mathcal{B}} = \begin{pmatrix} -1/2 \\ 7/2 \end{pmatrix}$.

3) If $v = (3, 4)$ and $\mathcal{B} = (v, (0, 1))$, then $[v]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

Definition 2.6.4. Let $\tau \in \mathcal{L}(V, W)$ with $\dim(V) = n$ and $\dim(W) = m$. Let $\mathcal{B} = (v_1, \ldots, v_n)$ be an ordered basis for V and $\mathcal{C}$ an ordered basis for W. Then the map

$$\theta : [v]_{\mathcal{B}} \to [\tau(v)]_{\mathcal{C}}$$

is a *representation* of τ as a L.T from F^n to F^m, in the sense that knowing θ along with $\mathcal{B}$ and $\mathcal{C}$ is equivalent to knowing τ.

In this definition, the representation θ of τ depends on the choice of ordered bases $\mathcal{B}$ and $\mathcal{C}$. Since θ is a L.T from F^n to F^m, it is just multiplication by an $m \times n$ matrix A, that is, $[\tau(v)]_{\mathcal{C}} = A[v]_{\mathcal{B}}$. Indeed, since $[v_i]_{\mathcal{B}} = e_i$, we get the columns of A as follows:

$$A^{(i)} = Ae_i = A[v_i]_{\mathcal{B}} = [\tau(v_i)]_{\mathcal{C}}\,.$$

Theorem and Definition 2.6.5. Let $\tau \in \mathcal{L}(V, W)$ and let $\mathcal{B} = (v_1, \ldots, v_n)$ and $\mathcal{C}$ be ordered bases for V and W, respectively. Then τ can be represented w.r.t. $\mathcal{B}$ and $\mathcal{C}$ as matrix multiplication, that is,

$$[\tau(v)]_{\mathcal{C}} = [\tau]_{\mathcal{B},\mathcal{C}}[v]_{\mathcal{B}}$$

where

$$[\tau]_{\mathcal{B},\mathcal{C}} = \Big[[\tau(v_1)]_{\mathcal{C}}|\cdots|[\tau(v_n)]_{\mathcal{C}}\Big]$$

is called the *matrix of* τ *w.r.t. the bases* $\mathcal{B}$ and $\mathcal{C}$. When $V = W$ and $\mathcal{B} = \mathcal{C}$, we denote $[\tau]_{\mathcal{B},\mathcal{B}}$ by $[\tau]_{\mathcal{B}}$ and so $[\tau(v)]_{\mathcal{B}} = [\tau]_{\mathcal{B}}[v]_{\mathcal{B}}$.

Example 2.6.6.

1) Let $D : P_2 \to P_2$ be the derivative operator, defined on the vector space of all polynomials of degree at most 2. Let $\mathcal{B} = \mathcal{C} = (1, x, x^2)$. Then

$$[D(1)_{\mathcal{C}}] = [0]_{\mathcal{C}} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, [D(x)_{\mathcal{C}}] = [1]_{\mathcal{C}} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},$$

$$[D(x^2)]_{\mathcal{C}} = [2x]_{\mathcal{C}} = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} \text{ and so}$$

$$[D]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

For example, if $p(x) = 5 + x + 2x^2$, then

$$[D(p(x))]_{\mathcal{C}} = [D]_{\mathcal{B},\mathcal{C}}[p(x)]_{\mathcal{B}} = [D]_{\mathcal{B}}[p(x)]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 0 \end{bmatrix}$$

and so $D(p(x)) = 1 + 4x$.

2) Consider the map

$$D : \mathbb{R}_3[x] \to \mathbb{R}_2[x], \qquad f \mapsto D(f) = f'.$$

Let $\mathcal{B} = (1, x, x^2)$ and $\mathcal{C} = (1, x, x^2, x^3)$ are the standard ordered bases for $\mathbb{R}_2[x]$ and $\mathbb{R}_3[x]$, respectively. Then

$$[D]_{\mathcal{B},\mathcal{C}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

3) If $\tau \in \mathcal{L}(\mathbb{R}^2, \mathbb{R}^3)$ is given by $\tau(x_1, x_2) = (x_1 + 3x_2, 0, 2x_1 - 4x_2)$, and $\mathcal{B} = (e_1, e_2)$ and $\mathcal{C} = (e_1, e_2, e_3)$ are the standard bases for $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively. Then

$$[\tau]_{\mathcal{B},\mathcal{C}} = \begin{bmatrix} 1 & 3 \\ 0 & 0 \\ 2 & -4 \end{bmatrix}.$$

Theorem 2.6.7. *Let V and W be finite-dimensional vector spaces over F, with ordered bases $\mathcal{B} = (b_1, \ldots, b_n)$ and $\mathcal{C} = (c_1, \ldots, c_m)$, respectively.*

1) The map $\mu : \mathcal{L}(V, W) \to \mathcal{M}_{m,n}(F)$ defined by $\mu(\tau) = [\tau]_{\mathcal{B},\mathcal{C}}$ is an isomorphism and so $\mathcal{L}(V, W) \cong \mathcal{M}_{m,n}(F)$. Hence,

$$\dim(\mathcal{L}(V, W)) = \dim(\mathcal{M}_{m,n}(F)) = m \times n\,.$$

2) If $\sigma \in \mathcal{L}(U, V)$ and $\tau \in \mathcal{L}(V, W)$ and if $\mathcal{B}, \mathcal{C}, \mathcal{D}$ are ordered bases for U, V and W respectively, then

$$[\tau\sigma]_{\mathcal{B},\mathcal{D}} = [\tau]_{\mathcal{C},\mathcal{D}}[\sigma]_{\mathcal{B},\mathcal{C}}\,.$$

Thus, the matrix of the product (composition) $\tau\sigma$ is the product of the matrices of τ and σ.

Proof. Since for all i,

$$\begin{aligned}
[s\sigma + t\tau]_{\mathcal{B},\mathcal{C}}[b_i]_{\mathcal{B}} &= [(s\sigma + t\tau)(b_i)]_{\mathcal{C}} \\
&= [s\sigma(b_i) + t\tau(b_i)]_{\mathcal{C}} \\
&= s[\sigma(b_i)]_{\mathcal{C}} + t[\tau(b_i)]_{\mathcal{C}} \\
&= s[\sigma]_{\mathcal{B},\mathcal{C}}[b_i]_{\mathcal{B}} + t[\tau]_{\mathcal{B},\mathcal{C}}[b_i]_{\mathcal{B}} \\
&= (s[\sigma]_{\mathcal{B},\mathcal{C}} + t[\tau]_{\mathcal{B},\mathcal{C}})\,[b_i]_{\mathcal{B}}
\end{aligned}$$

and since $[b_i]_{\mathcal{B}} = e_i$ is the standard basis vector, we conclude that

$$[s\sigma + t\tau]_{\mathcal{B},\mathcal{C}} = s[\sigma]_{\mathcal{B},\mathcal{C}} + t[\tau]_{\mathcal{B},\mathcal{C}},$$

that is, μ is linear.

Let $A \in \mathcal{M}_{m,n}$, then there exists $\tau \in \mathcal{L}(V, W)$ defined by $[\tau(b_i)]_{\mathcal{C}} = A^{(i)}$ such that $\mu(\tau) = A$ and, hence, μ is surjective. Also, $\ker(\mu) = \{0\}$ since $[\tau]_{\mathcal{B}} = 0$.

Proof of part 2)

$$[\tau\sigma]_{\mathcal{B},\mathcal{D}}[\tau]_{\mathcal{B}} = [\tau(\sigma(v))]_{\mathcal{D}} = [\tau]_{\mathcal{C},\mathcal{D}}[\sigma(v)]_{\mathcal{C}} = [\tau]_{\mathcal{C},\mathcal{D}}[\sigma]_{\mathcal{B},\mathcal{C}}[v]_{\mathcal{B}}.$$

□

Chapter 3

Diagonalization and Inner Product Spaces

3.1 Eigenvalues and Eigenvectors, Diagonalization

Definition 3.1.1. Let $\tau : V \to V$ be a L.T of the vector space V to itself. Let v be a non-zero vector in V. We say that v is said to be an *eigenvector* of *tau* if there exists $\lambda \in F$ with $\tau(v) = \lambda v$. The scalar λ is called the *eigenvalue* of τ corresponding to v.

Example 3.1.2. Let V be the set of all infinitely differentiable functions and let

$$\tau = D : V \to V \text{ derivative operator}\,.$$

Then, $v(x) = e^{\lambda x}$ is an eigenvector of τ because

$$\tau(v) = D(v) = \lambda e^{\lambda x} = \lambda v\,.$$

Recall from Chapter 2 that $\tau(v) = Av$, where A is the matrix of τ. Thus $Av = \lambda v$.

Example 3.1.3. Let

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 4 & 0 \\ 6 & 0 & 0 \end{pmatrix}\,.$$

Then $v = \begin{pmatrix} 3 \\ 0 \\ 6 \end{pmatrix}$ is an eigenvector of A with eigenvalue $\lambda = 3$. To see this,

$$Av = \begin{pmatrix} 9 \\ 0 \\ 18 \end{pmatrix} = 3 \begin{pmatrix} 3 \\ 0 \\ 6 \end{pmatrix} = 3v\,.$$

Remark 3.1.4.

a) The eigenvector corresponding to a given eigenvalue is not unique. In fact, any multiple of an eigenvector is an eigenvector with the same eigenvalue.

b) It is possible for multiple L.I eigenvectors to have the same eigenvalue.

Definition 3.1.5. The characteristic equation of a square matrix of order n is the nth order (or possibly lower order) polynomial $\det(A - \lambda I) = 0$.

Example 3.1.6. Consider the matrix

$$A = \begin{pmatrix} 1 & 0 & 3 \\ 4 & 5 & 0 \\ 0 & 3 & 1 \end{pmatrix} .$$

The characteristic equation of A is

$$\begin{aligned} 0 &= \det(A - \lambda I) \\ &= \begin{vmatrix} 1-\lambda & 0 & 3 \\ 4 & 5-\lambda & 0 \\ 0 & 3 & 1-\lambda \end{vmatrix} \\ &= (1-\lambda)(5-\lambda)(1-\lambda) + 3 \cdot 4 \cdot 3 \\ &= -\lambda^3 + 7\lambda^2 - 11\lambda + 41. \end{aligned}$$

The characteristic polynomial of a given square matrix A is denoted by χ_A. From the above example, we have

$$\chi_A(\lambda) = -\lambda^3 + 7\lambda^2 - 11\lambda + 41.$$

Theorem 3.1.7. *The eigenvalues of a square matrix A are the roots of its characteristic polynomial.*

Proof. By definition, λ is an eigenvalue of A if and only if there is a non-zero vector v such that $Av = \lambda v$. But

$$\begin{aligned} Av = \lambda v &\iff Av - \lambda v = 0 \\ &\iff (A - \lambda I) \cdot v = 0 \\ &\iff \det(A - \lambda I) = \chi_A(\lambda) = 0 \,. \end{aligned}$$

Hence, λ is a root of the characteristic polynomial. □

Example 3.1.8. The characteristic polynomial of the matrix

$$A = \begin{pmatrix} 2 & -2 & 3 \\ 1 & 1 & 1 \\ 1 & 3 & -1 \end{pmatrix}$$

is

$$\chi_A(\lambda) = -(\lambda + 2)(\lambda - 3)(\lambda - 1)\,.$$

Thus, by Theorem 3.1.7 the eigenvalues of A are the roots of $\chi_A = 0$. These are $-2, 1$ and 3.

Example 3.1.9. Find eigenvectors of the matrix given in Example 3.1.8.

Solution: The eigenvalues of A are $-2, 1$ and 3, see Example 3.1.8. In the following we see that how to find eigenvectors of A corresponding to $\lambda = -2$. One can then similarly find eigenvectors of A corresponding to the other eigenvalues. To find the eigenvectors of A corresponding to $\lambda = -2$, we solve the system

$$Av = -2v.$$

That is,

$$\begin{pmatrix} 2 & -2 & 3 \\ 1 & 1 & 1 \\ 1 & 3 & -1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = -2 \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \tag{3.1}$$

for v_1, v_2, v_3. This system is equivalent to

$$\begin{aligned} 4v_1 - 2v_2 + 3v_3 &= 0 \\ v_1 + 3v_2 + v_3 &= 0\,. \end{aligned}$$

From these system of equations, we obtain $v_3 = -14v_2$ and $v_1 = 11v_2$. Thus

$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 11v_2 \\ v_2 \\ -14v_2 \end{pmatrix} = v_2 \begin{pmatrix} 11 \\ 1 \\ -14 \end{pmatrix}.$$

This implies the vector $v = \begin{pmatrix} 11 \\ 1 \\ -14 \end{pmatrix}$ is an eigenvector of A w.r.t. $\lambda = -2$.

Note that kv where $k \in F$ is also an eigenvector. That is the eigenvector is never unique. For example, if v is an eigenvector of A with eigenvalue λ, then

$$A(kv) = kAv = k\lambda v$$

which implies $k\lambda$ is also an eigenvalue of A.

Definition 3.1.10. A *diagonal* matrix is a square matrix that all of its non-zero entries are on the main diagonal.

Theorem 3.1.11. *Let A be an $n \times n$ diagonal matrix. Then the eigenvalues of A are the elements of the diagonal.*

Proof. Let the diagonal elements be $d_1, \ldots, d_n$. By Theorem 3.1.7 the eigenvalues of A are the roots of the characteristic polynomial

$$\begin{aligned}\chi_A(\lambda) &= \det(A - \lambda I) \\ &= \begin{vmatrix} d_1 - \lambda & 0 & \cdots & 0 \\ 0 & d_2 - \lambda & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n - \lambda \end{vmatrix} \\ &= \prod_{i=1}^{n} (d_i - \lambda)\,.\end{aligned}$$

of A. Thus, $d_1, \ldots, d_n$ are eigenvalues of A. □

Theorem 3.1.12. *The determinant of an $n \times n$ diagonal matrix A is the product of the eigenvalues:*

$$\det(A) = \prod_{i=1}^{n} \lambda_i\,.$$

Proof. Let $\lambda_1, \ldots, \lambda_n$ be diagonal elements of A. Then the characteristic polynomial of A is

$$\begin{aligned}\det(\lambda I - A) = \chi_A(\lambda) &= \prod_{i=1}^{n} (\lambda - \lambda_i) \\ &\Rightarrow \chi_A(0) = (-1)^n \prod_{i=1}^{n} \lambda_i = \det(-A) \\ &\Rightarrow \chi_A(0) = (-1)^n \det(A) \\ &\Rightarrow \det(A) = \prod_{i=1}^{n} \lambda_i\,.\end{aligned}$$

□

Definition 3.1.13. The *trace* of a square matrix A, denoted by $\mathrm{tr}(A)$, is the sum of its diagonal elements.

Theorem 3.1.14. *Let A be an $n \times n$ square matrix. The trace of A is equal to the sum of the eigenvalues:*

$$\operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i \,.$$

Definition 3.1.15. An *upper* (resp. *lower*) triangular matrix a square matrix that only has non-zero entries on above (resp. below) the main diagonal.

Theorem 3.1.16. *The eigenvalues of an upper (resp. lower) triangular matrix lie on the main diagonal.*

Proof. Let A be an upper triangular matrix with $d_1, d_2, \ldots$ on the diagonal. The characteristic equation is

$$\begin{aligned} 0 &= \det(A - \lambda I) \\ &= \begin{vmatrix} d_1 - \lambda & 0 & \cdots & 0 \\ 0 & d_2 - \lambda & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n - \lambda \end{vmatrix} \\ &= \prod_{i=1}^{n} (d_i - \lambda) \end{aligned}$$

where the determinant is expanded by the first column. Hence, the roots are $d_1, d_2, \ldots$. A similar calculation holds for lower triangular matrices, expanding the determinant by the first row. □

Definition 3.1.17. A square matrix A is invertible if and only if $\det(A) \neq 0$.

Theorem 3.1.18. *If k is a positive integer, λ is an eigenvalue of a matrix A, and v is a corresponding eigenvector, then λ^k is an eigenvalue of A^k and v is a corresponding eigenvector.*

Example 3.1.19. Consider the matrix

$$A = \begin{pmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{pmatrix} .$$

Verify that $\lambda_1 = 2$ and $\lambda_2 = 1$ are eigenvalues of A. Then the eigenvalues of A^3, by Theorem 3.1.18, are 8 and 1, respectively.

Exercise 3.1.20. Show that

i) the vectors

$$v = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \text{ and } u = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

are eigenvectors of A and A^3 corresponding to $\lambda_1 = 2$.

ii) the vector

$$v = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}$$

is an eigenvector of A and A^3 corresponding to $\lambda_2 = 1$.

Theorem 3.1.21. *A square matrix A is invertible if and only if $\lambda = 0$ is not an eigenvalue of A.*

Proof. Assume that A is an $n \times n$ matrix. Suppose $\lambda = 0$ is a solution of the characteristic equation $\lambda^n + c_1\lambda^{n-1} + \ldots + c_n = 0$ if and only if the constant term c_n is zero. Thus it suffices to show that A is invertible if and only if $c_n \neq 0$. Now

$$\begin{aligned} \chi_A(\lambda) &= \det(\lambda I - A) \\ &= \lambda^n + c_1\lambda^{n-1} + \ldots + c_n \\ &\Longrightarrow c_n = \chi_A(0) = \det(-A) = (-1)^n \det(A) \\ \det(A) = 0 &\Longleftrightarrow c_n = 0 \end{aligned}$$

□

Example 3.1.22. The matrix given in Example 3.1.9 is invertible since $\lambda_1 = 2$ and $\lambda_2 = 1$ are both non-zero.

Theorem 3.1.23. *Let $\tau : \mathbb{R}^n \to \mathbb{R}^n$ be a L.T and let A be an $n \times n$ matrix of τ. Then the following are equivalent:*

a) *A is invertible.*

b) $\ker\tau = \{0\}$.

c) $\det(A) \neq 0$.

d) $\mathrm{Im}\tau = \mathbb{R}^n$.

e) *τ is 1-1.*

f) The column (row) vectors of A are L.I.

g) The column (row) vectors of A span $\mathbb{R}^n$.

h) The column (row) vectors of A form a basis for $\mathbb{R}^n$.

i) $\operatorname{rk}(A) = n$ *and* $\operatorname{null}(A) = 0$.

j) $A^T A$ *is invertible.*

k) $\lambda = 0$ *is not an eigenvalue of* A.

Definition 3.1.24. A square matrix A is diagonalizable if and only if there exists an invertible matrix P such that $P^{-1}AP = D$, a diagonal matrix. In this case, the matrix P is said to diagonalize A.

Let

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{2n} & \cdots & p_{nn} \end{bmatrix} \text{ and } D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

Since $P^{-1}AP = D \Longrightarrow AP = PD$,

$$PD = \left[\lambda_1 \begin{pmatrix} p_{11} \\ \vdots \\ p_{n1} \end{pmatrix} \cdots \lambda_n \begin{pmatrix} p_{1n} \\ \vdots \\ p_{nn} \end{pmatrix} \right].$$

But

$$AP = \left[A \begin{pmatrix} p_{11} \\ \vdots \\ p_{n1} \end{pmatrix} \cdots A \begin{pmatrix} p_{1n} \\ \vdots \\ p_{nn} \end{pmatrix} \right].$$

comparing with sides, we have

$$A \begin{pmatrix} p_{1i} \\ \vdots \\ p_{ni} \end{pmatrix} = \lambda_i \begin{pmatrix} p_{1i} \\ \vdots \\ p_{ni} \end{pmatrix}$$

Letting

$$p_i = \begin{pmatrix} p_{1i} \\ \vdots \\ p_{ni} \end{pmatrix}$$

we have

$$Ap_i = \lambda_i p_i \,. \tag{3.2}$$

Thus, the i-th column of P is an eigenvector in F^n with eigenvalue λ_i.

Example 3.1.25. Let $A = \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix}$. Find, if possible, an invertible matrix P with $P^{-1}AP = D$ is a diagonal matrix.

Solution:

$$\begin{aligned} 0 &= \det(A - \lambda I) \\ &= \begin{vmatrix} -2-\lambda & 1 \\ 1 & -2-\lambda \end{vmatrix} \\ &= \lambda^2 + 4\lambda + 3 = (\lambda+1)(\lambda+3) \end{aligned}$$

Thus, $\lambda = -1$ and $\lambda = -2$ are eigenvalues of A. If $\lambda = -3$, then

$$\begin{aligned} (A - \lambda I)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} &\Longleftrightarrow \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ &\Longleftrightarrow x_1 = -x_2 \\ &\Longleftrightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -x_2 \\ x_2 \end{pmatrix} = x_2 \begin{pmatrix} -1 \\ 1 \end{pmatrix}. \end{aligned}$$

Therefore, corresponding to $\lambda = -3$, we have an eigenvector $v_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

If $\lambda = -1$, then

$$\begin{aligned} (A - \lambda I)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} &\Longleftrightarrow \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ &\Longleftrightarrow x_1 = x_2 \\ &\Longleftrightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ x_2 \end{pmatrix} = x_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \end{aligned}$$

Therefore, corresponding to $\lambda = -1$, we have an eigenvector $v_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Set $P = [v_1\ v_2]$. Then

$$\begin{aligned} P = \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} &\Longrightarrow P^{-1} = \begin{pmatrix} -1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \\ &\Longrightarrow P^{-1}AP = \begin{pmatrix} -3 & 0 \\ 0 & -1 \end{pmatrix}. \end{aligned}$$

Theorem 3.1.26. *Let A be an $n \times n$ matrix. Then the following are equivalent:*

a) A is diagonalizable.

b) A has n L.T eigenvectors.

Proof. *a)* $\Longrightarrow$ *b)* Since A is assumed diagonalizable, there is an invertible matrix

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{2n} & \cdots & p_{nn} \end{bmatrix}$$

such that $P^{-1}AP$ is diagonal, say $P^{-1}AP = D$, where

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

Since $AP = PD$, by (3.2), we have $Ap_i = \lambda_i p_i$ where p_i the ith column vector which is also an eigenvector corresponding to λ_i. By Theorem 3.1.23, $p_1, \ldots, p_n$ are L.I since P is invertible. Thus, A has n L.I eigenvectors.

b) $\Longrightarrow$ *a)* Assume that A has n L.I eigenvectors $p_1, \ldots, p_n$ with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Let $P = [p_1| \cdots p_n]$ be the matrix with $p_1, \ldots, p_n$ are column vectors. But the column vectors of AP are $Ap_1, \ldots, Ap_n$ and

$$\begin{aligned} Ap_i &= \lambda_i p_i \forall i = 1, \ldots, n \\ \Longrightarrow AP &= [Ap_1| \cdots |Ap_n] \\ &= [\lambda_1 p_1| \cdots |\lambda_n p_n] \\ &= [p_1| \cdots |p_n]D \text{ with } D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \\ &= PD \end{aligned}$$

Thus by definition P diagonalizes A or A is diagonalizable. □

Procedure for Diagonalizing a Matrix

Given an $n \times n$ diagonalizable matrix A, the constructive proof of the above theorem provides the following method to diagonalize A:

Step 1 Find n L.I eigenvectors of A, say $p_1, \ldots, p_n$.

Step 2 Form the matrix P having $p_1, \ldots, p_n$ as its column vectors.

Step 3 The matrix $P^{-1}AP$ will then be diagonal with $\lambda_1, \ldots, \lambda_n$ as its successive diagonal entries, where λ is the eigenvalue corresponding to p_i for $i = 1, \ldots, n$.

Example 3.1.27. Find a matrix P that diagonalizes

$$A = \begin{pmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{pmatrix}.$$

Soulution: Verify that $\chi_A\lambda = (\lambda - 1)(\lambda - 2)^2$. Thus $\lambda = 1$ and $\lambda = 2$ are eigenvalues of A. Recall from Example 3.1.19 that

$$p_1 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \text{ and } p_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \text{ and } p_3 = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}$$

are eigenvectors corresponding to $\lambda = 2$ and $\lambda = 1$, respectively.

Exercise 3.1.28. Show that the vectors p_1, p_2 and p_3 are L.I.

Thus, the matrix $P = [p_1\ p_2\ p_3]$ diagonalizes A by Theorem 3.1.26 and hence

$$P^{-1}AP = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Note that changing the order of P just changes the order of the eigenvalues on the main diagonal of $P^{-1}AP$. Thus if we had written $P = \begin{pmatrix} -1 & -2 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$, we would have obtained

$$P^{-1}AP = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Example 3.1.29. Given the matrix

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ -3 & 5 & 2 \end{pmatrix},$$

there does not exist a matrix P that diagonalizes A. Because $\chi_A(\lambda) = (\lambda - 1)(\lambda - 2)^2$ and the bases for the eigenspaces are

$$\lambda = 1 : p_1 = \begin{pmatrix} 1/8 \\ -1/8 \\ 1 \end{pmatrix}, \ \lambda = 2 : p_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Since A is 3×3 and there are only two basis vectors in total, by Theorem 3.1.26, A is not diagonalizable.

Theorem 3.1.30. *Let $v_1, \ldots, v_k$ be eigenvectors of a matrix A and let $\lambda_1, \ldots, \lambda_k$ be the corresponding distinct eigenvalues. Then the set $\{v_1, \ldots, v_k\}$ is L.I.*

Proof. Suppose the $\{v_1, \ldots, v_k\}$ is L.D. Since an eigenvector is non-zero by definition, $\{v_1\}$ is L.I. Let r be the largest integer such that $\{v_1, \ldots, v_r\}$ is L.I. Since we are assuming that $\{v_1, \ldots, v_k\}$ is L.D, r satisfies $1 \leq r \leq k$. Moreover, by definition of r, $\{v_1, \ldots, v_{r+1}\}$ is L.D. Thus there are scalars $c_1, \ldots, c_{r+1}$, not all zero such that

$$c_1 v_1 + \ldots + c_{r+1} v_{r+1} = 0\,. \tag{3.3}$$

But

$$\begin{aligned} 0 &= c_1 v_1 + \ldots + c_{r+1} v_{r+1} \\ \implies A \cdot 0 = 0 &= c_1 A v_1 + \ldots + c_{r+1} A v_{r+1}\,. \end{aligned}$$

Since $Av_i = \lambda_i v_i$, we have

$$c_1 \lambda_1 v_1 + \ldots + c_{r+1} \lambda_{r+1} v_{r+1} = 0\,. \tag{3.4}$$

Multiply (3.3) both sides by λ_{r+1} and subtracting the resulting equation from (3.4) yields

$$c_1(\lambda_1 - \lambda_{r+1}) v_1 + \ldots + c_r(\lambda_r - \lambda_{r+1}) v_r = 0\,.$$

Since $v_1, \ldots, v_r$ are L.I, we have

$$c_1(\lambda_1 - \lambda_{r+1}) = \cdots = c_r(\lambda_r - \lambda_{r+1}) = 0\,.$$

Now

$$\begin{aligned} c_1(\lambda_1-\lambda_{r+1}) = \cdots = c_r(\lambda_r-\lambda_{r+1}) = 0 \\ \Longrightarrow & c_1 = \cdots = c_r = 0 \text{ since } \lambda_i \neq \lambda_j \ \forall i \neq j \\ \Longrightarrow & c_{r+1}v_{r+1} = 0 \quad \text{see Equation (3.3)} \\ \Longrightarrow & c_{r+1} = 0\,. \end{aligned}$$

This contradicts the fact that not all c_i's are zero. Thus the set $\{v_1, \ldots, v_k\}$ is L.I. □

Theorem 3.1.31. *Let A be an $n \times n$ matrix such that A has n distinct eigenvalues. Then A is diagonalizable.*

Proof. Let $v_1, \ldots, v_n$ be eigenvectors corresponding to $\lambda_1, \ldots, \lambda_n$.

$$\begin{aligned} & \lambda_i \neq \lambda_j \ \forall i \neq j \\ \Longrightarrow\ & v_1, \ldots, v_n \text{ are L.I by Theorem 3.1.30} \\ \Longrightarrow\ & A \text{ is diagonalizable by Theorem 3.1.26}\,. \end{aligned}$$

□

Example 3.1.32. The matrix

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 4 & -17 & 8 \end{pmatrix}$$

has three distinct eigenvalues (**verify**) $\lambda = 4$ and $\lambda = 2 \pm \sqrt{3}$. Thus by Theorem 3.1.31, A is diagonalizable and

$$P^{-1}AP = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 2+\sqrt{3} & 0 \\ 0 & 0 & 2-\sqrt{3} \end{pmatrix}.$$

Remark 3.1.33. The eigenvalues of a triangular matrix are the entries on its main diagonal. Thus, a triangular matrix with distinct entries on the main diagonal is diagonalizable.

Example 3.1.34. The matrix

$$A = \begin{pmatrix} -1 & 2 & 4 & 0 \\ 0 & 3 & 1 & 7 \\ 0 & 0 & 5 & 8 \\ 0 & 0 & 0 & -2 \end{pmatrix}$$

is diagonalizable by the above remark.

Remark 3.1.35. Theorem 3.1.31 says only that if a matrix has all distinct eigenvalues (whether real or complex), then it is diagonalizable. In other words, only matrices with repeated eigenvalues might be non-diagonalizable.

Example 3.1.36. a) The matrix $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ has repeated eigenvalues $\lambda = 1$ with multiplicity 3 but is diagonalizable since any non-zero vector in $\mathbb{R}^3$ is an eigenvector of A (**verify**). In particular, we can find 3 L.I eigenvectors.

b) The matrix $B = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ has repeated eigenvalues $\lambda = 1$ with multiplicity 3 but is diagonalizable but solving for its eigenvectors leads to the system

$$(\lambda I - B) \cdot x = 0$$

the solution of which is $x_1 = t, x_2 = 0, x_3 = 0$. Thus every eigenvector of B is a multiple of

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

which means that the eigenspace has dimension 1 and that is B is non-diagonalizable.

3.2 Block Matrices and Their Properties

Let A be an $n \times m$ matrix and B be an $m \times p$ matrix. Suppose $r < m$. Then, we can decompose the matrices A and B as

$$A = \begin{bmatrix} P & Q \end{bmatrix} \text{ and } B = \begin{bmatrix} H \\ K \end{bmatrix}$$

where P order $n \times r$ and H has order $r \times p$. That is, the matrices P and Q are submatrices of A and P consists of the first r columns of A and Q consists of the last $m - r$ columns of A. Similarly, H and K are submatrices of B and H consists of the first r rows of B and K consists of the last $m - r$ rows of B.

Theorem 3.2.1. *Let $A = [a_{ij}] = \begin{bmatrix} P & Q \end{bmatrix}$ and $B = [a_{ij}] = \begin{bmatrix} H \\ K \end{bmatrix}$ be defined as above. Then*

$$AB = PH + QK\,.$$

Proof. First note that the matrices PH and QK are each of order $n \times p$. The matrix products PH and QK are valid as the order of the matrices P, H, Q and K are respectively, $n \times r$, $r \times p$, $n \times (m-r)$ and $(m-r) \times p$. Let $P = [P_{ij}], Q = [Q_{ij}], H = [H_{ij}]$ and $K = [K_{ij}]$. Then, for $1 \leq i \leq n$ $1 \leq j \leq p$, we have

$$
\begin{aligned}
(AB) &= \sum_{k=1}^{m} a_{ik} b_{kj} \\
&= \sum_{k=1}^{r} a_{ik} b_{kj} + \sum_{k=r+1}^{m} a_{ik} b_{kj} \\
&= \sum_{k=1}^{r} P_{ik} H_{kj} + \sum_{k=r+1}^{m} Q_{ik} K_{kj} \\
&= (PH)_{ij} + (QK)_{ij} \\
&= (PH + QK)_{ij}
\end{aligned}
$$

□

Remark 3.2.2. Theorem 3.2.1 is very useful due to the following reasons:

- The order of the matrices P, Q, H and K are smaller that that of A or B.
- It may be possible to block the matrix in such a way that a few blocks are either identify matrices or zero matrices. In this case, it may be easy to handle the matrix product using the block form.
- When we want to prove results using induction, then we may assume that the result for $r \times r$ submatrices and then look for $(r+1) \times (r+1)$ submatrices, etc.

Example 3.2.3. If

$$
A = \begin{bmatrix} 0 & -1 & 2 \\ 3 & 1 & 4 \\ -2 & 5 & -3 \end{bmatrix},
$$

then A can be decomposed as follows:

$$
A = \left[\begin{array}{c|cc} 0 & -1 & 2 \\ 3 & 1 & 4 \\ \hline -2 & 5 & -3 \end{array}\right], \quad A = \left[\begin{array}{cc|c} 0 & -1 & 2 \\ \hline 3 & 1 & 4 \\ -2 & 5 & -3 \end{array}\right]
$$

$$
A = \left[\begin{array}{cc|c} 0 & -1 & 2 \\ 3 & 1 & 4 \\ \hline -2 & 5 & -3 \end{array}\right], \quad A = \left[\begin{array}{c|cc} 0 & -1 & 2 \\ \hline 3 & 1 & 4 \\ -2 & 5 & -3 \end{array}\right]
$$

Definition 3.2.4. Let $A = \begin{bmatrix} P & Q \\ R & S \end{bmatrix}$ be a decomposition of a matrix A. Then the submatrices P, Q, R and S are called the *blocks* of the matrix A.

Example 3.2.5. In Example 3.2.3, if

$$A = \left[\begin{array}{cc|c} 0 & -1 & 2 \\ \hline 3 & 1 & 4 \\ -2 & 5 & -3 \end{array}\right],$$

, then the block matrices P, Q, R and S of A respectively are

$$P = \begin{bmatrix} 0 & -1 \end{bmatrix}, Q = \begin{bmatrix} 2 \end{bmatrix}, R = \begin{bmatrix} 3 & 1 \\ -2 & 5 \end{bmatrix} \text{ and } S = \begin{bmatrix} 4 \\ -3 \end{bmatrix}.$$

Remark 3.2.6. Let $A = \begin{bmatrix} P & Q \\ R & S \end{bmatrix}$ and $B = \begin{bmatrix} E & F \\ G & H \end{bmatrix}$ be matrices with block matrices P, Q, R, S and E, F, G, H.

1. Even if $A + B$ is defined, the order of P and E may not be same and, hence we may not be able to add A and B in the block form. But if $A + B$ and $P + E$ are defined, then
$$A = \begin{bmatrix} P+E & Q+F \\ R+G & S+H \end{bmatrix}$$

2. If the product AB is defined the product PE need not be defined. Therefore, if both the products AB and PE are defined, then the product of block matrices is defined. In this case, we have
$$AB = \begin{bmatrix} PE+QG & PF+QH \\ RE+SG & RF+SH \end{bmatrix}$$

Note that once a partition of A is fixed, the partition of B has to be properly chosen for the purposes of addition or multiplication.

Exercise 3.2.7.

1. Compute the matrix product AB using the block matrix multiplication for the matrices
$$A = \begin{bmatrix} 1 & -2 & 1 & 2 \\ -1 & 3 & 2 & 0 \\ 7 & 0 & 1 & 4 \\ 0 & 1 & 4 & 2 \end{bmatrix}, B = \begin{bmatrix} 1 & -1 & 0 & 7 \\ 0 & -2 & 1 & 3 \\ 1 & 4 & 2 & 1 \\ 2 & 2 & 0 & 4 \end{bmatrix}$$

2. Let $A = \begin{bmatrix} P & Q \\ R & S \end{bmatrix}$. If P, Q, R, and S are symmetric, what can you say about A? Are P, Q, R and S symmetric, when A is symmetric?

3. Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be two matrices. Suppose $a_1, a_2, \ldots, a_n$ are the rows of A and $b_1, b_2, \ldots, b_m$ are the columns of B. If the product AB is defined, then show that
$$AB = \begin{bmatrix} Ab_1 & Ab_2 & \cdots & Ab_m \end{bmatrix} = \begin{bmatrix} a_1B \\ a_2B \\ \vdots \\ a_nB \end{bmatrix}$$

3.3 Determinants of 2 by 2 Block Matrices

Lemma 3.3.1. *If* $A = \begin{bmatrix} E & F \\ G & H \end{bmatrix}$, *then*
$$\det(A) = \det(EH - FG)$$
whenever at least one of the blocks E, F, G *and* H *is equal to zero.*

3.4 Jordan Canonical Forms

Recall that not every $n \times n$ matrix A can be diagonalized. However, we can always put matrices of type A into something called *Jordan canonical form*, which means that A can be written as
$$A = B^{-1} \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_k \end{bmatrix} B$$
where the J_i are certain block matrices of the form
$$J_i = \begin{bmatrix} \lambda \end{bmatrix} \text{ or } \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} \text{ or } \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix} \text{ etc}$$
with λ an eigenvalue of A.

How do we determine the J_i blocks?

Fix an eigenvalue λ. To determine the size of the Jordan blocks J_i that are associated to λ, it turns out that all we need to know are the numbers

$$\text{null}(A-\lambda I), (\text{null}(A-\lambda I))^2, \text{etc}.$$

Moreover, we have

- $\text{null}(A-\lambda I)$ is the number of Jordan blocks J_i associated to λ.
- The number
$$s_j = (\text{null}(A-\lambda I))^j - (\text{null}(A-\lambda I))^{j-1}$$
is the number of Jordan blocks associated to λ that are of size at least $j \times j$.

Let J_i be an $n_i \times n_i$ block matrix. Then one can easily check that

$$A - \lambda I = B^{-1} C_k B$$

where

$$C_k = \begin{bmatrix} J_1 - \lambda I_{n_1} & 0 & \cdots & 0 \\ 0 & J_2 - \lambda I_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_k - \lambda I_{n_k} \end{bmatrix}.$$

It is not hard to see that the jth power of the matrix $A - \lambda I$ is

$$(A-\lambda I)^j = B^{-1} C_k' B$$

where

$$C_k' = \begin{bmatrix} (J_1 - \lambda I_{n_1})^j & 0 & \cdots & 0 \\ 0 & (J_2 - \lambda I_{n_2})^j & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (J_k - \lambda I_{n_k})^j \end{bmatrix}.$$

Remark 3.4.1.

$$\text{null}(C_k') = \sum_{i=1}^{k} (\text{null}(A - \lambda I_{n_i}))^j \text{ , and}$$
$$\text{rank}(C_k') = \sum_{i=1}^{k} (\text{rank}(A - \lambda I_{n_i}))^j \text{ .}$$

How do we determine s_j?

Until now we know that how many Jordan blocks there are, but we would like to also determine their various sizes s_j. To do this, we need to understand what the various powers of the blocks H_i look like:

Let

$$H_i = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} .$$

Then

$$H_i^1 = H_i \,, H_i^2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} , H_i^3 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} , H_i^4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} .$$

Thus $H_i^j = 0$ for all $j \geq 4$. By studying this example, one should be convinced that the following is true:

$$\text{null}\left(H_i^j\right) = \begin{cases} j & \text{if } n_i \geq j \\ n_i & \text{if } n_i < j \end{cases} \implies \text{null}\left(H_i^j\right) - \text{null}\left(H_i^{j-1}\right) = \begin{cases} 1 & \text{if } n_i \geq j \\ 0 & \text{if } n_i < j \end{cases}$$

Note that the only blocks that could possibly contribution to the nullity (when we sum up the nullities of the $(J_i - \lambda I_{n_i})^j$ blocks) are those whose eigenvalues equal to λ, because otherwise $(J_i - \lambda I_{n_i})^j$ is an $n_i \times n_i$ upper triangular matrix whose diagonal contains non-zero entries, making it invertible which is equivalent to $\text{null}(J_i - \lambda I_{n_i})^j = 0$.

We thus only need to focus on blocks corresponding to the same eigenvalue λ. Let $J_1, \ldots, J_t$ be the reordered Jordan blocks corresponding to λ. Here, $J_{t+1}, \ldots, J_k$ do not contribute to the nullity since their nullities are zero. Then, $H_i = J_i - \lambda I_{n_i} \,, i = 1, \ldots, t$, might look like.

$$\begin{bmatrix} 0 \end{bmatrix} \text{ or } \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \text{ or } \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \text{ or } \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ or}$$

$$\text{null}(H_i) = 1 \implies \text{rk}(H_i) = n_i - 1 \,, i = 1, \ldots, t$$

$$\implies \sum_{i=1}^{t} \text{null}(H_i) = t = \text{null}(C_k) = \text{null}(A - \lambda I) \,.$$

We thus have t Jordan blocks associated to the eigenvalue λ. If we sum this up over all the blocks H_i, we get a sum of 1's when $n_i \geq j$ which means that

$$\text{null}\left((A - \lambda I)^j - (A - \lambda I)^{j-1}\right)$$

equals the number of blocks of size at least $j \times j$ corresponding to the eigenvalue λ.

Remark 3.4.2. The number of possible sizes of the Jordan blocks of an $n \times n$ matrix having a single λ is the number of integer partitions of n, denoted by $P(n)$. For example, $P(4) = 5$, namely

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

Example 3.4.3. Find the Jordan blocks of the matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -3 & 4 & 0 & 0 \\ 2 & -1 & 2 & 0 \\ -1 & 1 & 1 & 2 \end{bmatrix}.$$

In the first step, we determine its characteristic polynomial. It is easy to see that the characteristic polynomial of A is:

$$\chi_A(\lambda) = (\lambda - 2)^4$$

and thus $\lambda = 2$ is an eigenvalue of A. Since A has a single eigenvalue, the number of Jordan blocks, by the above remark, is $P(4) = 5$. These are

$$C_1 = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}, C_2 = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}, C_3 = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}, C_4 = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

and

$$C_5 = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

Now in order to reduce the possibilities, we will need to first compute the number of Jordan blocks by computing nullity of $A - 2I$, since $\lambda = 2$. To do this, by row reducing, we get

$$A-2I = \begin{bmatrix} -2 & -1 & 0 & 0 \\ -3 & 2 & 1 & 0 \\ 2 & -1 & 0 & 0 \\ -1 & 1 & 1 & 0 \end{bmatrix} \to \begin{bmatrix} -1 & 1 & 1 & 0 \\ -3 & 2 & 1 & 0 \\ 2 & -1 & 0 & 0 \\ -2 & 1 & 1 & 0 \end{bmatrix} \to \begin{bmatrix} -1 & 1 & 1 & 0 \\ 0 & -1 & -2 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & -1 & -2 & 0 \end{bmatrix} \to \begin{bmatrix} -1 & 1 & 1 & 0 \\ 0 & -1 & -2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

This implies that $\mathrm{null}(A - 2I) = 2$ and, hence, we have two Jordan blocks. Thus either we have

$$C_3 = \left[\begin{array}{ccc|c} 2 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ \hline 0 & 0 & 0 & 2 \end{array}\right] \text{ or } C_4 = \left[\begin{array}{cc|cc} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ \hline 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{array}\right].$$

To determine which it is, we must compute $\text{null}(A-2I)^2$. First

$$(A-2I)^2 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 2 & 0 & 2 & 0 \\ -1 & 0 & -1 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix}.$$

Thus $\text{null}((A-2I)^2) = 3$ since $\text{rk}((A-2I)^2) = 1$. That is, only the first column is a basis for the column space. Moreover, we have

$$\text{null}((A-2I)^2) - \text{null}((A-2I)) = 3 - 2 = 1\,.$$

This implies that there is exactly 1 matrix having size at least 2×2. Since in C_4 there are two 2×2 block matrices,

$$A = B^{-1} C_3 B\,.$$

Why C_3? Again consider the matrix $(A-2I)^3 = 0$. Now

$$\text{null}((A-2I)^3) = 4 \Longrightarrow \text{null}((A-2I)^3) - \text{null}((A-2I)^2) = 4 - 3 = 1\,.$$

This implies that there is exactly 1 matrix having size at least 3×3 which is of course C_3.

3.5 Inner Products

Definition 3.5.1. Let V be a real vector space. An *inner product* on V is a function

$$\langle\ ,\ \rangle : V \times V \to \mathbb{R}$$

that associates each pair (u, v) of elements in V to a real number $\langle u, v\rangle$ such that for all u, v, w in V and scalar λ, we have:

- *symmetric*: $\langle u, v\rangle = \langle v, u\rangle$.
- *bilinear* (that is linear in both factors):

- $\langle \lambda u, v\rangle = \lambda\langle u, v\rangle$ for all scalars λ and
- $\langle u + w, v\rangle = \langle u, v\rangle + \langle w, v\rangle$

- *positive*: $\langle v, v\rangle \geq 0$ and $\langle v, v\rangle = 0$ if and only if $v = 0$.

A vector space V together with an inner product $\langle , \rangle$ is called a *real inner product space.*

Definition 3.5.2. Let $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_n)$ be vectors in $\mathbb{R}^n$.

- A *Euclidean inner product* on R^n is defined as

$$\langle u, v\rangle = \sum_{i=1}^{n} u_i v_i \,, \text{ and}$$

- A *weighted Euclidean inner product* on R^n is defined as

$$\langle u, v\rangle = \sum_{i=1}^{n} w_i u_i v_i$$

where $w_1, \ldots, w_n$ are positive real numbers, which we call *weights.*

Example 3.5.3. The function $\langle , \rangle$ defined by

$$\begin{aligned} \langle , \rangle : \mathbb{R}^2 &\to \mathbb{R} \\ (u, v) &\mapsto 3u_1v_1 + 2u_2v_2 \end{aligned}$$

where $u = (u_1, u_2)$ and $v = (v_1, v_2)$ are vectors in $\mathbb{R}^2$ defines an inner product on $\mathbb{R}^2$. In fact, this inner product is a weighted inner product on $\mathbb{R}^2$. It is easy to see that the above function satisfies the axioms of inner product space. Let us check it these axioms.

- Symmetric: $\langle u, v\rangle = \langle v, u\rangle$

$$\langle u, v\rangle = 3u_1v_1 + 2u_2v_2 = 3v_1u_1 + 2v_2u_2 = \langle v, u\rangle \,.$$

- Bilinear: Let $z = (z_1, z_2) \in \mathbb{R}^2$. Then

$$\begin{aligned} \langle u + z, v\rangle &= \langle (u_1 + z_1, u_2 + z_2), (v_1, v_2)\rangle \\ &= 3(u_1 + z_1)v_1 + 2(u_2 + z_2)v_2 \\ &= 3u_1v_1 + 2u_2v_2 + 3z_1v_1 + 2z_2v_2 \\ &= \langle u, v\rangle + \langle z, v\rangle \,. \end{aligned}$$

and

$$\begin{aligned} \langle \lambda u, v\rangle &= 3\lambda(u_1v_1) + 2\lambda(u_2v_2) \\ &= \lambda(3u_1v_1 + 2u_2v_2) \\ &= \lambda\langle u, v\rangle \,. \end{aligned}$$

- positive:

$$\begin{aligned}\langle v, v\rangle &= 3v_1v_1 + 2v_2v_2 \\ &= 3v_1^2 + 2v_2^2 \geq 0 \text{ and}\end{aligned}$$

$$\langle v, v\rangle = 0 \Longleftrightarrow v_1 = v_2 = 0 \Longleftrightarrow v = 0\,.$$

Definition 3.5.4. Let V be an inner product space. The *norm* (or *length*) of a vector u in V is denoted by $\|u\|$ and is defined by $\|u\| = \sqrt{\langle u, v\rangle}$. The *distance* between two points (vectors) u and v is denoted by $d(u, v)$ and is defined by $d(u, v) = \|u - v\|$.

Note that if a vector has norm 1, then we say that it is a *unit* vector.

Example 3.5.5. Let $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_n$ are vectors in $\mathbb{R}^n$ with the Euclidean inner product , then

$$\|u\| = \sqrt{u, u}^{1/2} = (u \cdot u)^{1/2} = \sqrt{u_1^2 + \ldots + u_n^2}$$

and

$$d(u, v) = \|u - v\| = (u - v) \cdot (u - v) = \sqrt{(u_1 - v_1)^2 + \ldots + (u_n - v_n)^2}\,.$$

Remark 3.5.6. Norm and distance depend on the inner product being used. That is, if the inner product is changed, then the norms and distances between vectors also change.

Example 3.5.7. For the vectors $u = e_1$ and $v = e_2$ in $\mathbb{R}^2$ with

a) the Euclidean inner product, we have

$$\|u\| = \sqrt{1^2 + 0^2} = 1 \text{ and } d(u, v) = \sqrt{(1 - 0)^2 + (0 - 1)^2} = \sqrt{2}\,.$$

b) the weighted Euclidean inner product of Example 3.5.3, we obtain

$$\|u\| = \sqrt{3(1)^2 + 2(0)^2} = \sqrt{3} \text{ and } d(u, v) = \sqrt{3(1 - 0)^2 + 2(0 - 1)^2} = \sqrt{5}\,.$$

Inner Products Generated by Matrices

The Euclidean inner product (EIP) and weighted Euclidean inner product (WEIP) are special cases of a general class of inner products on $\mathbb{R}^n$, which we shall now describe: Let $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_n)$ be vectors in $\mathbb{R}^n$ and let A be an invertible $n \times n$ matrix. If $u \cdot v$ is the EIP on $\mathbb{R}^n$, then the formula

$$\langle u, v\rangle = Au \cdot Av \tag{3.5}$$

defines an inner product (**exercise**), it is called the *inner product* on $\mathbb{R}^n$ generated by A. Since $v^T \cdot u$, (3.5) can be written as

$$\langle u, v\rangle = (Av)^T Au = v^T A^T Au\,.$$

Example 3.5.8 (Inner Product generated by the identity matrix). The inner product on $\mathbb{R}^n$ generated by the $n \times n$ identity matrix is the EIP, since substituting $A = I$ in (3.5) yields

$$\langle u, v \rangle = Iu \cdot Iv = u \cdot v\,.$$

The WEIP $3u_1v_1 + 2u_2v_2$ discussed in Example 3.5.3 is the IP on $\mathbb{R}^2$ generated by

$$\begin{bmatrix} \sqrt{3} & 0 \\ 0 & \sqrt{3} \end{bmatrix} = A \qquad\qquad [A \text{ is symmetric }]$$

In fact, it is easy to check that

$$\langle u, v \rangle = \begin{bmatrix} v_1 & v_2 \end{bmatrix} A^T A \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = 3u_1v_1 + 2u_2v_2$$

In general, the WEIP $\langle u, v \rangle = \sum_{i=1}^{n} w_i u_i v_i$ is the IP on $\mathbb{R}^n$ generated by

$$A = \begin{bmatrix} \sqrt{w_1} & 0 & \dots & 0 \\ 0 & \sqrt{w_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{w_n} \end{bmatrix}$$

Example 3.5.9 (Inner Product on $M_{2\times 2}$). If

$$u = \begin{bmatrix} u_1 & u_2 \\ u_3 & u_4 \end{bmatrix} \text{ and } v = \begin{bmatrix} v_1 & v_2 \\ v_3 & v_4 \end{bmatrix}$$

are any two 2×2 matrices, then the following formula defines an IP on $M_{2\times 2}$ (**easy exercise**):

$$\langle u, v \rangle = \operatorname{tr}(u^T v) = \operatorname{tr}(v^T)u = u_1v_1 + u_2v_2 + u_3v_3 + u_4v_4\,.$$

For example, if

$$u = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \text{ and } v = \begin{bmatrix} -1 & 0 \\ 3 & 2 \end{bmatrix},$$

then

$$\langle u, v \rangle = 1(-1) + 2(0) + 3(3) + 4(2) = 16\,.$$

- The norm of a matrix u relative to this IP is

$$\|u\| = \langle u, u \rangle^{1/2} = \sqrt{u_1^2 + u_2^2 + u_3^2 + u_4^2}\,.$$

- the unit sphere in this space consists of all matrices $u = \begin{bmatrix} u_1 & u_2 \\ u_3 & u_4 \end{bmatrix}$ in $M_{2\times 2}$ whose that satisfy the equation $\|u\| = 1$, that is, $u_1^2 + u_2^2 + u_3^2 + u_4^2 = 1$.

Example 3.5.10 (**An IP on** $P_2 = k[x]_2 = \{a + bx + cx^2 \mid a, b, c \in k$)**.** If $p = a_0 + a_1x + a_2x^2$ and $q = b_0 + b_1x + b_2x^2$ are any two vectors in P_2, then the following formula defines an IP on P_2 (**easy exercise**)

$$\langle p, q\rangle = a_0b_0 + a_1b_1 + a_2b_2\,.$$

- $\|p\| = \langle p, p\rangle^{1/2} = \sqrt{a_0^2 + a_1^2 + a_2^2}$.
- the unit sphere in this space consists of all polynomials p in P_2 whose coefficients satisfy the equation $\|p\| = 1$, that is, $a_0^2 + a_1^2 + a_2^2 = 1$.

Example 3.5.11 (**An IP on** $C[a, b]$**, the space of all continuous functions**)**.** Let f and g are two functions in $C[a, b]$ and define

$$\langle f, g\rangle = \int_a^b f(x)g(x)\mathrm{d}x\,.$$

This is well defined since the functions in $C[a, b]$ are continuous. Let $p = a_0 + a_1x + a_2x^2$ and $q = b_0 + b_1x + b_2x^2$ are any two vectors in P_2, then the following formula defines an IP on P_2 (**easy exercise**)

$$\langle p, q\rangle = a_0b_0 + a_1b_1 + a_2b_2\,.$$

- $\|p\| = \langle p, p\rangle^{1/2} = \sqrt{a_0^2 + a_1^2 + a_2^2}$.
- the unit sphere in this space consists of all polynomials p in P_2 whose coefficients satisfy the equation $\|p\| = 1$, that is, $a_0^2 + a_1^2 + a_2^2 = 1$.

Theorem 3.5.12 (**Properties of Inner Products**)**.** *If u, v and w are vectors in a real IP space, and k is any scalar, then*

a) $\langle 0, v\rangle = \langle v, 0\rangle = 0$.

b) $\langle u, v + w\rangle = \langle u, v\rangle + \langle uwv\rangle$.

c) $\langle u, kv\rangle = k\langle u, v\rangle$.

d) $\langle u - v, w\rangle = \langle u, w\rangle - \langle v, w\rangle$.

e) $\langle u, v - w\rangle = \langle u, v\rangle - \langle u, w\rangle$.

3.6 Angle and Orthogonality in Inner Product Spaces

Theorem 3.6.1 (Cauchy-Schwarz Inequality)**.** *If u and v are vectors in a real inner product space, then*

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

Proof.

- **Case 1:** Assume $u = 0$. Then $|u \cdot | = |0 \cdot | = 0$ and $\|u\|\|v\| = 0\|v\| = 0$.
- **Case 2:** Assume $u \neq 0$. This implies $a = u \cdot u = \|u\|^2 \geq 0$. Let t be any real number. By the positivity axiom,

  $$0 \leq \langle tu + v, tu + v\rangle = \langle u, u\rangle t^2 + 2\langle u, v\rangle t + \langle v, v\rangle = at^2 + bt + c =: f(t)$$

  with $a = \langle u, u\rangle, b = 2\langle u, v\rangle$ and $c = \langle v, v\rangle$. Note that $f(t) \geq 0$ implies that either f has no a real root or has a repeated root. This is true only if $b^2 - 4ac \leq 0$, by the quadratic formula. Now

  $$b^2 - 4ac \leq 0 \Rightarrow 4\langle u, v\rangle^2 \leq 4\langle u, u\rangle\langle v, v\rangle = 4\|u\|^2\|v\|^2 .$$

  Taking square root on both sides yields

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

  as desired.

□

Theorem 3.6.2 (**Properties of Length**)**.** *If u and v are vectors in an inner product space V, and if k is any scalar, then*

i) $\|u\| \geq 0$

ii) $\|u\| = 0 \Leftrightarrow u = 0$.

iii) $\|ku\| = |k|\|u\|$.

iv) $\|u + v\| \leq \|u\| + \|v\|$ *(triangle inequality).*

Theorem 3.6.3 (**Properties of Distance**)**.** *If u and v are vectors in an inner product space V, and if k is any scalar, then*

i) $d(u, v) \geq 0$

ii) $d(u, v) = 0 \Leftrightarrow u = 0$.

iii) $d(u, v) = d(v, u)$.

iv) $d(u, v) \leq d(u, w) + d(w, v)$ *(triangle inequality).*

Proof. By definition,

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

Thus we have

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

□

Angle Between Vectors

Cauchy-Schwarz inequality can be used to define angles in general inner product spaces. From Theorem 3.6.1, we obtain

$$\left[\frac{\langle u, v \rangle}{\|u\|\|v\|}\right]^2 \leq 1 \Rightarrow -1 \leq \frac{\langle u, v \rangle}{\|u\|\|v\|} \leq 1 \tag{3.6}$$

If θ is an angle between radian measure varies from 0 to π, then $\cos\theta$ assumes every value between -1 and 1 inclusive once. From (3.6), there is a unique angle θ such that

$$\cos\theta = \frac{\langle u, v \rangle}{\|u\|\|v\|} \text{ and } 0 \leq \theta \leq \pi \,.$$

Formally, the angle between two vectors u and v is defined as follows:

Definition 3.6.4. Let u and v are vectors in an inner product space V. Then the angle between two vectors u and v is defined as

$$\cos\theta = \frac{\langle u, v \rangle}{\|u\|\|v\|} \text{ and } 0 \leq \theta \leq \pi \,.$$

Definition 3.6.5. Two vectors u and v in an IP space are called *orthogonal* if $\langle u, v \rangle = 0$.

Example 3.6.6. Show that the matrices $u = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ and $v = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}$ are orthogonal.

Since

$$\langle u, v \rangle = \operatorname{tr}(u^T v) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} = 0,$$

they are orthogonal.

Theorem 3.6.7. *If u and v are orthogonal vectors in an IP space, then $\|u+v\|^2 = \|u\|^2 + \|v\|^2$.*

Proof. **Left as an exercise** □

Example 3.6.8. Let $p = x$ and $q = x^2 \in P_2 = \{f \in F[x] \mid \deg f \leq 2\}$. Find $\|u+v\|^2$.

Solution: Verify that p and q are orthogonal. Thus by the above theorem, we have

$$\begin{aligned} \|p+q\|^2 &= \|p\|^2 + \|q\|^2 \\ &= \left(\frac{\sqrt{2}}{\sqrt{3}}\right)^2 + \left(\frac{\sqrt{2}}{\sqrt{5}}\right)^2 \\ &= \frac{2}{3} + \frac{2}{5} = \frac{16}{15}. \end{aligned}$$

Definition 3.6.9. Let W be a subspace of an IP space V. A vector u in V is said to be *orthogonal* to W if it is orthogonal to every vector in W, and the set of all vectors in V that are orthogonal to W is called the *orthogonal complement* of W and denoted by $W^\perp$ (read W perp).

Theorem 3.6.10 (**Properties of Orthogonal Complements**). *If W is a subspace of a finite-dimensional IP space V, then*

a) $W^\perp$ is a subspace of V.

b) The only vector common to W and $W^\perp$ is 0, that is, $W \cap W^\perp = \{0\}$.

c) $(W^\perp)^\perp = W$.

Proof.

a) $W^\perp = \{v \in V \mid \langle v, w \rangle \forall w \in W\}$. Let u and v be elements of $W^\perp$. Then

$$\langle u+v, w \rangle = \langle u, w \rangle + \langle v, w \rangle = 0 + 0 = 0 \text{ and } \langle ku, w \rangle = k\langle u, w \rangle = k \cdot 0 = 0.$$

This implies that $u + v$ and ku are in $W^\perp$.

b) and c) (non-trivial) are left as an exercise.

□

Theorem 3.6.11. *Let A be an $m \times n$ matrix. Then*

a) The null space of A and the row space of A are orthogonal complements in $\mathbb{R}^n$ with respect to the EIP.

b) The null space of A^T and the column space of A are orthogonal complements i $n\mathbb{R}^m$ with respect to the EIP.

c) Moreover, if $m = n$, the following statements are equivalent:

i) A is invertible

ii) The orthogonal complement of the null space of A is $\mathbb{R}^n$.

iii) The orthogonal complement of the row space of A is $\{0\}$.

3.7 Orthonormal Bases

Definition 3.7.1. A set of vectors in an IP space is called an *orthogonal set* if all pairs of distinct vectors in the set are orthogonal. An orthogonal set in which each vector has norm 1 is called *orthonormal*.

Example 3.7.2. The standard vectors e_1, e_2, e_3 in $\mathbb{R}^3$ are orthonormal. That is, the set $S = \{e_1, e_2, e_3\}$ is orthogonal (since $e_i \cdot e_j = 0$ for all $i \neq j$) and also orthonormal since $\|e_i\| = 1$ for $i = 1, 2, 3$.

Note that for any non-zero vector v in an IP space, the vector $\frac{v}{\|v\|}$ has norm 1. Moreover, if $S = \{v_1, \ldots, v_n\}$ is orthogonal set, then the set $S' = \{u_1, \ldots, u_n\}$ with $u_i = \frac{v_i}{\|v_i\|}$ is orthonormal.

How do we construct an Orthonormal set?

Given orthogonal vectors $v_1, \ldots, v_n$ in an IP space, we construct an orthonormal set as follows: First we normalize the vectors $v_1, \ldots, v_n$ to obtain the vectors

$$u_1 = \frac{v_1}{\|v_1\|}, \ldots, u_n = \frac{v_n}{\|v_n\|}.$$

Then the set $S = \{u_1, \ldots, u_n\}$ is an orthonormal set since each pair of these vectors satisfies the conditions in the above definition.

Example 3.7.3. Consider the vectors $v_1 = e_2, v_2 = (1,0,1)$ and $v_3 = (1,0,-1)$ in $\mathbb{R}^3$. Assume that $\mathbb{R}^3$ has the EIP. Since $\langle v_i, v_j\rangle = 0$ for all $i \neq j$ (1,2,3) the vectors v_1, v_2, v_3 are orthogonal. Are they orthonormal? No since $\|v_2\| = \sqrt{2} \neq 1$. In order to construct an orthonormal set with respect to the above vectors we consider first the unit vector

$$u_1 = \frac{v_1}{\|v_1\|}, u_2 = \frac{v_2}{\|v_2\|} \text{ and } u_3 = \frac{v_3}{\|v_3\|}$$

$$\Rightarrow u_1 = e_2, u_2 = \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right) \text{ and } u_3 = \left(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}\right).$$

Thus the set $\{u_1, u_2, u_3\}$ is orthonormal.

Definition 3.7.4. In an IP space, a basis consisting of orthonormal vectors is called an *orthonormal basis*, and a basis consisting of orthogonal vectors is called an *orthogonal basis*.

Coordinates Relative to Orthonormal Bases

Theorem 3.7.5. *If $S = \{v_1, \ldots, v_n\}$ is an orthonormal basis for an IP space V, and u is any vector in V, then*

$$u = \langle u, v_1\rangle v_1 + \langle u, v_2\rangle v_2 + \ldots + \langle u, v_n\rangle v_n .$$

Proof. Since $S = \{v_1, \ldots, v_n\}$ is a basis, a vector u can be expressed in the form $u = k_1 v_1 + \ldots + k_n v_n$. We want to show that $k_i = \langle u, v_i\rangle$ for all $i = 1, \ldots, n$. For each vector v_i in S, we have

$$\begin{aligned}
\langle u, v_i\rangle &= \langle k_1 v_1 + \ldots + k_n v_n, v_i\rangle \\
&= k_1\langle v_1, v_i\rangle + \ldots + k_{i-1}\langle v_{i-1}, v_i\rangle + k_i\langle v_i, v_i\rangle + k_{i+1}\langle v_{i+1}, v_i\rangle + \ldots + \langle v_n, v_i\rangle \\
&= k_i \text{ since } \langle v_i, v_j\rangle = 0 \text{ for } i \neq j \text{ and } \langle v_i, v_j\rangle = 1 \text{ for } i = j .
\end{aligned}$$

□

Definition 3.7.6. Let $S = \{v_1, \ldots, v_n\}$ be an orthonormal basis for an IP space V. The coordinates of a vector u relative to the orthonormal basis S are $\langle u, v_1\rangle, \ldots, \langle u, v_n\rangle$ and the vector $(\langle u, v_1\rangle, \ldots, \langle u, v_n\rangle)$ is called the *coordinate vector* of u relative to the basis S and is denoted by $(u)_S$.

Example 3.7.7. Consider the vectors

$$v_1 = e_2 , v_2 = \left(-\frac{4}{5}, 0, \frac{3}{5}\right) \text{ and } v_3 = \left(\frac{3}{5}, 0, \frac{4}{5}\right).$$

With EIP, the set $S = \{v_1, v_2, v_3\}$ is an orthonormal basis (check!). Express the vector $u = (1,1,1)$ as a L.C of the vectors in S, and find the coordinate vector $(u)_S$.

Solution: Since $\langle u, v_1\rangle = 1$, $\langle u, v_2\rangle = -\frac{1}{5}$ and $\langle u, v_3\rangle = \frac{7}{5}$, we have, by Theorem 3.7.5, $u = v_1 - \frac{1}{5}v_2 + \frac{7}{5}v_3$. The coordinate vector of u relative to S is $(u)_S = \left(1, -\frac{1}{5}, \frac{7}{5}\right)$.

Theorem 3.7.8. *If S is an orthogonal basis for an n-dimensional IP space and if $(u)_S = (u_1, \ldots, u_n)$ and $(v)_S = (v_1, \ldots, v_n)$, then*

a) $\|u\| = \sqrt{u_1^2 + \ldots + u_n^2} = \|(u)_S\|$.

b) $d(u,v) = \sqrt{(u_1 - v_1)^2 + \ldots + (u_n - v_n)^2}$.

c) $\langle u, v\rangle = u_1v_1 + \ldots + u_nv_n = \langle (u)_S, (v)_S\rangle$.

Proof.

a)

$$\begin{aligned}
\|u\| &= \|\sum_{s\in S}\langle u, s\rangle\| \\
&= \left\langle \sum_{s\in S}\langle u, s\rangle, \sum_{s\in S}\langle u, s\rangle \right\rangle^{\frac{1}{2}} \\
&= \left(\sum_{s\in S}\langle u, s\rangle^2\langle s, s\rangle + \sum_{s\in S, t\neq s}\langle u, t\rangle^2\langle s, t\rangle\right)^{\frac{1}{2}} \\
&= \left(\sum_{s\in S}\langle u, s\rangle^2\langle s, s\rangle\right)^{\frac{1}{2}} \quad \text{since } \langle s, t\rangle = 0 \text{ for } t \neq s \\
&= \left(\sum_{s\in S}\langle u, s\rangle^2\right)^{\frac{1}{2}} \quad \text{since } \langle s, s\rangle = 1 \\
&= \|(u)_S\|
\end{aligned}$$

b) and c) are exercise. □

Example 3.7.9. In Example 3.7.7, $\|u\| = 3 = \|(u)_S\|$.

Coordinates Relative to Orthogonal Bases

If $S = \{v_1, \ldots, v_n\}$ is an orthogonal basis for a vector space V, then normalizing each of these vectors yields the orthonormal basis

$$S' = \left\{\frac{v_1}{\|v_1\|}, \frac{v_2}{\|v_2\|}, \ldots, \frac{v_n}{\|v_n\|}\right\}.$$

If u is any vector in V, then by Theorem 3.7.5

$$\begin{aligned} u &= \left\langle u, \frac{v_1}{\|v_1\|} \right\rangle \frac{v_1}{\|v_1\|} + \ldots + \left\langle u, \frac{v_n}{\|v_n\|} \right\rangle \frac{v_n}{\|v_n\|} \\ &= \left\langle u, \frac{v_1}{\|v_1\|^2} \right\rangle v_1 + \ldots + \left\langle u, \frac{v_n}{\|v_n\|^2} \right\rangle v_n \end{aligned}$$

which implies that the coordinates of u relative to S are

$$\left\langle u, \frac{v_1}{\|v_1\|^2} \right\rangle, \ldots, \left\langle u, \frac{v_n}{\|v_n\|^2} \right\rangle .$$

Theorem 3.7.10. *If $S = \{v_1, \ldots, v_n\}$ is an orthogonal set of non-zero vectors in an IP space, then S is L.I.*

Proof. Suppose $k_1 v_1 + \ldots + k_n v_n = 0$. We want to show that each k_i is zero.

$$\begin{aligned} 0 &= \langle 0, v_i \rangle \\ &= \langle k_1 v_1 + \ldots + k_n v_n, v_i \rangle \\ &= k_1 \langle v_1, v_i \rangle + \ldots + k_{i-1} \langle v_{i-1}, v_i \rangle + k_i \langle v_i, v_i \rangle + k_{i+1} \langle v_{i+1}, v_i \rangle + \ldots + \langle v_n, v_i \rangle \\ &= k_i \text{ since } \langle v_i, v_j \rangle = 0 \text{ for } i \neq j \text{ and } \langle v_i, v_j \rangle = 1 \text{ for } i = j \,. \end{aligned}$$

□

3.8 Orthogonal Projections

Theorem 3.8.1 (Projection Theorem)**.** *If W is a finite-dimensional subspace of an IP space V, then every vector u in V can be expressed in exactly one way as*

$$u = w_1 + w_2, w_1 \in W, w_2 \in W^{\perp} \,. \tag{3.7}$$

Definition 3.8.2. The vector w_1 is called the *orthogonal projection of* u on W and is denoted by $\mathrm{Pro}_{\mathrm{W}}^{\mathrm{u}}$. The vector w_2 is called the *component of u orthogonal to W* and is denoted by $\mathrm{Pro}_{\mathrm{W}^{\perp}}^{\mathrm{u}}$. In this case, (3.7) can be expressed as

$$u = \mathrm{Pro}_{\mathrm{W}}^{\mathrm{u}} + \mathrm{Pro}_{\mathrm{W}^{\perp}}^{\mathrm{u}} \,.$$

Since $w_2 = u - w_1$, it follows that

$$\mathrm{Pro}_{\mathrm{W}^{\perp}}^{\mathrm{u}} = u - \mathrm{Pro}_{\mathrm{W}}^{\mathrm{u}} \Rightarrow u = \mathrm{Pro}_{\mathrm{W}}^{\mathrm{u}} + (u - \mathrm{Pro}_{\mathrm{W}}^{\mathrm{u}}) \,.$$

Theorem 3.8.3. *Let W be a finite-dimensional subspace of an IP space V.*

a) If $\{v_1, \dots, v_r\}$ is an orthonormal basis for W, and u is any vector in V, then

$$\mathrm{Pro}_{\mathrm{W}}^{\mathrm{u}} = \langle u, v_1 \rangle v_1 + \ldots + \langle u, v_r \rangle v_r$$

b) If $\{v_1, \dots, v_r\}$ is an orthogonal basis for W, and u is any vector in V,

$$\mathrm{Pro}_{\mathrm{W}}^{\mathrm{u}} = \frac{\langle u, v_1 \rangle}{\|v_1\|^2} v_1 + \ldots + \frac{\langle u, v_r \rangle}{\|v_r\|^2} v_r \,.$$

Example 3.8.4. Consider the subspace $W = \langle v_1, v_2 \rangle$ of $\mathbb{R}^3$ where $v_1 = e_2$ and $v_2 = \left(-\frac{4}{5}, 0, \frac{3}{5}\right)$ and $u = (1, 1, 1) \in \mathbb{R}^3$. Find, in a EIP,

a) $\mathrm{Pro}_{\mathrm{W}}^{\mathrm{u}}$ b) $\mathrm{Pro}_{\mathrm{W}^\perp}^{\mathrm{u}}$

Solution: Note that the vectors v_1 and v_2 are orthonormal (check!). Thus by Theorem 3.8.3,

$$\mathrm{Pro}_{\mathrm{W}}^{\mathrm{u}} = \langle u, v_1 \rangle v_1 + \langle u, v_2 \rangle v_2 = \left(\frac{4}{25}, 1, -\frac{3}{25}\right)$$

and

$$\mathrm{Pro}_{\mathrm{W}^\perp}^{\mathrm{u}} = u - \mathrm{Pro}_{\mathrm{W}}^{\mathrm{u}} = \left(\frac{21}{25}, 0, \frac{28}{25}\right)$$

3.9 Gram-Schmidt Orthogonalization Process

Theorem 3.9.1. *Every non-zero finite-dimensional IP space has an orthonormal basis.*

Proof. Let V be any non-zero finite-dimensional IP space and suppose that $\{u_1, \dots, u_n\}$ is any basis for V. It suffices to show that V has an orthogonal basis, since the vectors in the orthogonal basis can be normalized to produce an orthonormal basis for V. To see this, we follow the following steps:

step 1: Let $v_1 = u_1$

step 2: $W_1 = \langle v_1 \rangle$ and

$$v_2 = \mathrm{Pro}_{\mathrm{W}_1^\perp}^{\mathrm{u}_2} = u_2 - \mathrm{Pro}_{\mathrm{W}_1}^{\mathrm{u}_2} = u_2 - \frac{\langle u_2, v_1 \rangle}{\|v_1\|^2} v_1 \,.$$

The set containing only v_1 is L.I.

step 3: $W_2 = \langle v_1, v_2 \rangle$ and

$$v_3 = \mathrm{Pro}_{\mathrm{W}_2^\perp}^{\mathrm{u}_3} = u_3 - \frac{\langle u_3, v_1 \rangle}{\|v_1\|^2} v_1 - \frac{\langle u_3, v_2 \rangle}{\|v_2\|^2} v_2 \,.$$

Clearly, W_2 is not in W_1 since v_2 is not in W_1. Thus the set containing only v_1 and v_2 is L.I. Continuing in this way we have

step $n-2$**:** $W_{n-1} = \langle v_1, v_2, \ldots, v_{n-1}\rangle$ and

$$v_n = \mathrm{Pro}^{\mathrm{u_n}}_{\mathrm{W^{\perp}_{n-1}}} = u_n - \sum_{j=1}^{n-1} \frac{\langle u_n, v_j\rangle}{\|v_j\|^2} v_j\,.$$

Now set $W_n := \langle v_1, v_2, \ldots, v_n\rangle$. Clearly, W_n is not in W_{n-1} thus by the independence of $\{v_1, v_2, \ldots, v_n\}$, v_n is orthogonal to $\{v_1, v_2, \ldots, v_{n-1}\}$. The set $\{v_1, v_2, \ldots, v_n\}$ is orthogonal set and then we normalize it to obtain an orthonormal set. □

Definition 3.9.2. The preceding step-by-step construction for converting an arbitrary basis into an orthonormal basis is called the *Gram-Schmidt process.*

Example 3.9.3. Consider the vectors $u_1 = (1,1,1), u_2 = (0,1,1)$ and $u_3 = e_3$ in a EIP space $\mathbb{R}^3$. Using Gram-Schmidt process convert the above vectors into an orthogonal and orthonormal basis.

We start with $v_1 = u_1$ and $W_1 = \langle v_1\rangle$. Then

$$v_2 = u_2 - \frac{\langle u_2, v_1\rangle}{\|v_1\|^2} v_1 = \left(-\frac{2}{3}, \frac{1}{3}, \frac{1}{3}\right)\,.$$

Since $v_2 \neq 0$, then $v_2 \perp v_1$ and $\{v_1, v_2\}$ is L.I. Now we set $W_2 = \langle v_1, v_2\rangle$. Then

$$v_3 = u_3 - \frac{\langle u_3, v_1\rangle}{\|v_1\|^2} v_1 - \frac{\langle u_3, v_2\rangle}{\|v_2\|^2} v_2 = \left(0, -\frac{1}{2}, \frac{1}{2}\right)\,.$$

Therefore, the set $\{v_1, v_2, v_3\}$ is an orthogonal basis for $\mathbb{R}^3$. Since $\|v_1\| = \sqrt{3}, \|v_2\| = \frac{\sqrt{6}}{3}$ and $\|v_3\| = \frac{1}{\sqrt{2}}$, an orthonormal basis for $\mathbb{R}^3$ is

$$\left\{\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right), \left(-\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right), \left(0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)\right\}\,.$$

3.10 Orthogonal Matrices

Definition 3.10.1. A square matrix A with the property $A^{-1} = A^T$ is said to be an *orthogonal matrix*

By the above definition, a square matrix A is orthogonal if and only if $AA^T = A^TA = I$ where I is an identity matrix.

Example 3.10.2. The matrices

$$A = \frac{1}{7}\begin{bmatrix} 3 & 2 & 6 \\ -6 & 3 & 2 \\ 2 & 6 & -3 \end{bmatrix} \text{ and } B = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

are orthogonal (verify!).

Theorem 3.10.3. *The following are equivalent for an $n \times n$ matrix A:*

a) A is orthogonal

b) The row vectors of A form an orthonormal set in $\mathbb{R}^n$ with the EIP.

c) The column vectors of A form an orthonormal set in $\mathbb{R}^n$ with the EIP.

Example 3.10.4. Consider the matrix A given in Example 3.10.2. Since the row or column vectors of this matrix form an orthonormal set (verify) in $\mathbb{R}^3$, A is an orthogonal matrix by the above theorem.

Theorem 3.10.5.

a) The inverse of an orthogonal matrix is orthogonal.

b) A product of orthogonal matrices is orthogonal.

c) If A is orthogonal, then $\det(A) = 1$ or $\det(A) = -1$

3.11 Complex Inner Product Spaces

Definition 3.11.1. An inner product on a complex vector space V is a function that associate a complex number $\langle u, v \rangle$ with each pair of vectors u and v in V in such a way that the following axioms are satisfied for all vectors u, v, and w in V and for all scalars k:

a) $\langle u, v \rangle = \overline{\langle v, u \rangle}$

b) $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$.

c) $\langle ku, v \rangle = k\langle u, v \rangle$

d) $\langle v, v \rangle \geq 0$ and $\langle v, v \rangle = 0$ if and only if $v = 0$.

A complex vector space with an IP is called a *complex IP space.*

Remark 3.11.2. The following additional properties follow immediately from the four IP axioms:

i) $\langle 0, v \rangle = \langle v, 0 \rangle = 0$

ii) $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$.

iii) $\langle u, kv\rangle = \overline{k}\langle u, v\rangle$.

Proof.

$$\begin{aligned}\langle u, kv\rangle &= \overline{\langle ku, v\rangle} && a)\\ &= \overline{k\langle v, u\rangle} && c)\\ &= \overline{k}\,\overline{\langle v, u\rangle} && (\text{ properties of conjugate})\\ &= \overline{k}\langle u, v\rangle && a)\end{aligned}$$

□

Exercise 3.11.3. Let $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_n)$ be vectors in $\mathbb{C}^n$. Show that the EIP $\langle u, v\rangle = u_1\overline{v_1} + \ldots + u_n\overline{v_n}$ satisfies all the IP axioms.

Definition 3.11.4. Let V be a complex IP space. The *norm* (or *length*) of a vector u in V is defined as

$$\|u\| = \langle u, u\rangle^{\frac{1}{2}} = \sqrt{|u_1|^2 + \ldots + |u_n|^2}$$

and the distance between two vectors u and v is defined by

$$d(u, v) = \|u - v\| = \sqrt{|u_1 - v_1|^2 + \ldots + |u_n - v_n|^2}$$

Remark 3.11.5. The definitions of terms like orthogonal vectors, orthogonal set, orthonormal set, and orthonormal basis carry over to complex IP spaces without change. Moreover, Theorems 3.6.1, 3.7.5, 3.7.10, 3.8.1, 3.8.3, 3.9.1 remain valid in complex IP spaces, and the Gram-Schmidt process can be used to convert an arbitrary basis for a complex IP space into an orthonormal basis.

Example 3.11.6. Show that the vectors $u = (i, 1)$ and $v = (1, i)$ in $\mathbb{C}^2$ are orthogonal w.r.t. EIP.

Solution: Since

$$\langle u, v\rangle = u_1\overline{v_1} + u_2\overline{v_2} = i \cdot \overline{1} + 1 \cdot \overline{i} = 0\,,$$

they are orthogonal.

Example 3.11.7. Consider the vector space $\mathbb{C}^3$ with the EIP. Apply the Gram-Schmidt process to transform the basis vectors $u_1 = (i, i, i)$, $u_2 = (0, i, i)$ and $u_3 = (0, 0, i)$ into an orthonormal basis.

Solution:

step 1: $v_1 = u_1$ and set $W_1 = \langle v_1 \rangle$.

step 2: $v_2 = u_2 - \mathrm{Proj}_{W_1}^{u_2} = u_2 - \frac{\langle u_2, v_1 \rangle}{\|v_1\|^2} v_1 = \left(-\frac{2}{3}i, \frac{1}{3}i, \frac{1}{3}i\right)$.

step 3: Set $W_2 := \langle v_1, v_2 \rangle$ and

$$v_3 = \mathrm{Proj}_{W_2}^{u_3} = u_3 - \frac{\langle u_3, v_1 \rangle}{\|v_1\|^2} v_1 - \frac{\langle u_3, v_2 \rangle}{\|v_2\|^2} v_2 = \left(0, -\frac{1}{2}i, \frac{1}{2}i\right).$$

Thus the vectors v_1, v_2 and v_3 form an orthogonal basis for $\mathbb{C}^3$. The norm of these vectors are $\|v_1\| = \sqrt{3}$, $\|v_2\| = \frac{\sqrt{6}}{3}$ and $\|v_3\| = \frac{1}{\sqrt{2}}$. So an orthonormanl basis for $\mathbb{C}^3$ is

$$\left\{ \left(\frac{i}{\sqrt{3}}, \frac{i}{\sqrt{3}}, \frac{i}{\sqrt{3}}\right), \left(-\frac{2i}{\sqrt{3}}, \frac{i}{\sqrt{6}}, \frac{i}{\sqrt{6}}\right), \left(0, -\frac{i}{\sqrt{2}}, \frac{i}{\sqrt{2}}\right) \right\}.$$

3.12 Unitary Matrices

Definition 3.12.1. Let A be a matrix with complex entries. Then the conjugate transpose of A, denoted by A^*, is defined by $A^* = \overline{A}^T$ where $\overline{A}$ is the matrix whose entries are the complex conjugates of the corresponding entries in A and $\overline{A}^T$ is the transpose of $\overline{A}$. The conjugate transpose is also called the *Hermitian matrix*.

Example 3.12.2.

$$A = \begin{bmatrix} 1+i & -i & 0 \\ 2 & 3-2i & i \end{bmatrix} \Rightarrow \overline{A} = \begin{bmatrix} 1-i & i & 0 \\ 2 & 3+2i & -i \end{bmatrix} \Rightarrow \overline{A}^T = \begin{bmatrix} 1-i & 2 \\ i & 3-2i \\ 0 & -i \end{bmatrix}.$$

Properties of the conjugate Transpose

Theorem 3.12.3. *If A and B are matrices with complex entries and k is any complex number, then*

a) $(A^*)* = A$;

b) $(A+B)^* = A^* + B^*$;

c) $(kA)^* = \overline{k}A^*$.

Recall that if u and v are column vectors in $\mathbb{R}^n$, then the EIP on $\mathbb{R}^n$ can be expressed as $u \cdot v = u^T \cdot v$. However, if u and v are column vectors in $\mathbb{C}^n$, then the EIP on $\mathbb{C}^n$ can be expressed as $u \cdot v = u^* v$.

Definition 3.12.4. A square matrix A with complex entries is *unitary* if $A^{-1} = A^*$.

Theorem 3.12.5. *If A is an $n \times n$ matrix with complex entries, then the following are equivalent:*

a) A is unitary.

b) The row (resp. column) vectors of A form an orthonormal set in $\mathbb{C}^n$ with the EIP.

Example 3.12.6. Show that the matrix

$$A = \begin{bmatrix} \frac{1+i}{2} & \frac{1+i}{2} \\ \frac{1-i}{2} & \frac{-1+i}{2} \end{bmatrix}.$$

is unitary w.r.t. the EIP on $\mathbb{C}^2$.

Solution: Verify that the row vectors have norm 1 w.r.t. the EIP. Moreover, these vectors are orthogonal and they form an orthonormal set in $\mathbb{C}^2$. Thus A is unitary by the equivalent conditions given in the above theorem and, hence, we have

$$A^{-1} = A^* = \overline{A}^T = \begin{bmatrix} \frac{1-i}{2} & \frac{1+i}{2} \\ \frac{1-i}{2} & \frac{-1-i}{2} \end{bmatrix}$$

since this can easily be verfied by showing that $AA^* = I$.

Definition 3.12.7. A square matrix A with complex entries is called *uniratarly diagonalizable* if there is a unitary matrix P such that $P^{-1}AP(P^*AP)$ is a diagonal matrix.

Definition 3.12.8. A square matrix A with complex entries is called *Hermitian* if $A = A^*$.

Example 3.12.9. Is the matrix $A = \begin{bmatrix} 1 & i & 1+i \\ -i & -5 & 2-i \\ 1-i & 2+i & 3 \end{bmatrix}$ Hermitian?

Solution: Since

$$\overline{A} = \begin{bmatrix} 1 & -i & 1-i \\ i & -5 & 2+i \\ 1+i & 2-i & 3 \end{bmatrix} \Rightarrow A^* = \overline{A}^T = \begin{bmatrix} 1 & i & 1+i \\ -i & -5 & 2-i \\ 1-i & 2+i & 3 \end{bmatrix} = A,$$

it is Hermitian.

One can easily recognize Hermitian matrices by inspection. Consider the following remark:

Remark 3.12.10. A matrix A is Hermitian if

- the entries on the main diagonal are real numbers.

- the mirror image of each entry across the main diagonal is its complex conjugate.

Definition 3.12.11.

a) A square matrix A is *orthogonal* w.r.t. EIP if $A^{-1} = A^T$.

b) A square matrix A is *symmetric* if $A = A^T$.

Definition 3.12.12. An $n \times n$ matrix A is said to be *orthogonally diagonalizable* if there exists an orhtogonal matrix P such that the matrix $P^{-1}AP = P^T AP$ is a diagonal matrix. In this case, P is said to orthogonally diagonalizes A.

Theorem 3.12.13. *Let A be an $n \times n$ matrix. The following statements are equivalent:*

a) A is orthogonally diagonalizable.

b) A has an orhtonormal set of n eigenvectors.

c) A is symmetric.

Theorem 3.12.14. *Let A be a symmetric matrix. Then*

i) The eigenvalues of A are all real numbers.

ii) Eigenvectors from different eigenspaces are orthogonal.

Steps to Diagonalize Symmetric Matrices

Thus to diagonalize symmetric matrices we follow the following steps:

step 1: Find a basis for each eigenspace of A.

step 2: Apply the Gram-Schmidt process to each of these bases to obtain an orthonormal basis for each eigenspace.

step 3: Form the matrix P whose columns are the basis vectors constructed in step 2. This matrix orthogonally diagonalizes A.

Remark 3.12.15. Theorem 3.12.14 ensures that eigenvectors from different eigenspaces are orthogonal, whereas the application of the Gram-Schmidt process ensures that the eigenvectors within the same eigenspace are orthonormal. Therefore, the entire set of eigenvectors obtained by this procedure is orthonotrmal.

Example 3.12.16. Find an orhtogonal matrix P that diagonalizes

$$A = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 24 & 2 \\ 2 & 2 & 4 \end{bmatrix}.$$

Solution: $\chi_A(\lambda) = (\lambda - 2)^2(\lambda - 8) = 0 \Rightarrow \lambda_1 = 2$ and $\lambda_2 = 8$ are eigenvalues. The vectors $u_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$ and $u_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$ form a basis for the eigenspace corresponding to $\lambda_1 = 2$. Applying the gram-Schmidt process to $\{u_1, u_2\}$ yields the following orthonormal eigenvectors (verify!):

$$v_1 = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} \text{ and } v_2 = \begin{pmatrix} -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix}$$

The eigenspace corresponding to $\lambda_2 = 8$ has $u_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ as a basis. Apply Gram-schmidt process to $\{u_3\}$ to obtain

$$v_3 = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}.$$

Finally, using u_1, u_2 and u_3 as column vectors, we obtain

$$P = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}.$$

which orthogonally diagonalizes A.

3.13 Normal Matrices

Definition 3.13.1. A square matrix A with complex entries ic called *normal* if $AA^* = A * A$.

Example 3.13.2.

a) Every Hermitian matrix is normal since $AA^* = AA = A^*A$.

b) Every unitary matrix $AA^* = I = A^*A$.

Theorem 3.13.3. *Let A be $n \times n$ a square matrix with complex entries. The following statements are equivalent:*

i) A is unitarly diagonalizable.

ii) A has an orthonormal set of n eigenvectors.

iii) A is normal.

Theorem 3.13.4. *If A is a normal matrix, then eigenvectors from different eigenspaces of A are orthogonal.*

Theorem 3.13.4 is the key to construct a matrix that unitarly diagonalizes a normal matrix.

Steps to Diagonalize Normal Matrices

Recall that a symmetric matrix is orthogonally diagonalized by any orthogonal matrix whose column vectors are eigenvectors of A. Similarly, a normal matrix A is diagonalized by any unitary matrix whose column vectors are eigenvectors of A. Thus to diagonalize normal matrices we follow the following steps:

step 1: Find a basis for each eigenspace of A.

step 2: Apply the Gram-Schmidt process to each of these bases to obtain an orthonormal basis for each eigenspace.

step 3: Form the matrix P whose columns are the basis vectors constructed in step 2. This matrix unitarly diagonalizes A.

Remark 3.13.5. Theorem 3.13.4 ensures that eigenvectors from different eigenspaces are orthogonal, and the application of the Gram-Schmidt process ensures that the eigenvectors within the same eigenspace are orthonormal. Thus the entire set of eigenvectors obtained by this procedure is orthonormal and, hence, is a basis by Theorem 3.13.3.

Example 3.13.6. Show that the matrix $A = \begin{bmatrix} 2 & 1+i \\ 1-i & 3 \end{bmatrix}$ is unitarly diagonalizable. Find a matrix P that unitarily diagonalizes A.

Solution: Since the matrix A is Hermitian, it is normal see Example 3.13.2 a). Thus A is unitarily diagonalizable by Theorem 3.13.3. To find the matrix P, first we find eigenvalues of A. It is easy to see that $\lambda_1 = 1$ and $\lambda_2 = 4$ are eigenvalues of A. By definition, $x = (x_1, x_2)^T$ is an eiegenvector of A corresponding to λ if and only if x is a non-trivial solution of

$$\begin{bmatrix} \lambda - 2 & -1-i \\ - + i & \lambda - 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

If $\lambda_1 = 1$, we have

$$\begin{bmatrix} -1 & -1-i \\ -1+i & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

and

$$\begin{bmatrix} -1 & -1-i \\ -1+i & -2 \end{bmatrix} \to \begin{bmatrix} 1 & 1+i \\ -1+i & -2 \end{bmatrix} \to \begin{bmatrix} 1 & 1+i \\ 0 & 0 \end{bmatrix}.$$

This implies that

$$\begin{bmatrix} 1 & 1+i \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow x_2 \text{ is free variable}.$$

Let $x_2 = s$. Then $x_1 + x_2(1+i) = 0 \Rightarrow x_1 = -(1+i)s$. Thus we have

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = s \begin{bmatrix} -1-i \\ 1 \end{bmatrix} \Rightarrow u_1 = \begin{pmatrix} -1-i \\ 1 \end{pmatrix}$$

is an eigenvector corresponding to $\lambda_1 = 1$. Gram-schmidt process involves only one step. After normalizing this vector, we obtain an orthonormal basis $p_1 = \begin{pmatrix} \frac{-1-i}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}$ for the eigenspace corresponding to $\lambda_1 = 1$. Similarily, $u_2 = \begin{pmatrix} \frac{1+i}{2} \\ 1 \end{pmatrix}$ is an eigenvector corresponding to $\lambda_2 = 4$. Moreover,

$$v_2 = \begin{pmatrix} \frac{1+i}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix}$$

is an orthormal basis obtained by applying Gram-Schmidt process for the eigenspace corresponding to $\lambda_2 = 4$. Thus $P = \begin{bmatrix} v_1 & v_2 \end{bmatrix}$ diagonalizes A and

$$P^{-1}AP = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}.$$

Chapter 4

Quadratic Forms

4.1 Quadratic Forms

Consider the linear equation $a_1x_1 + \ldots + a_nx_n = b$. The expression on the left side of this equation,

$$a_1x_1 + \ldots + a_nx_n$$

is a function of n variables, called a *linear form.* In a linear form, all variables occur to the first power, and there are no products of variables in the expression.

Definition 4.1.1. Functions of the form

$$a_1x_1^2 + \ldots + a_nx_n^2 + (\text{ all possible terms of the form} a_kx_ix_j \text{ for } i < j)$$

is called *quadratic forms.*

Example 4.1.2. The most general quadratic form in the variables

a) x_1 and x_2 is

$$a_1x_1^2 + a_2x_2^2 + a_3x_1x_2 \tag{4.1}$$

b) x_1, x_2 and x_3 is

$$a_1x_1^2 + a_2x_2^2 + a_3x_3^2 + a_4x_1x_2 + a_5x_1x_3 + a_6x_2x_3 \tag{4.2}$$

Note that the terms in a quadratic form that involve products of different variables are called the *cross-product* terms. In the above example,

a) the cross-product term is $a_3x_1x_3$;

b) the cross-product terms are $a_4x_1x_2 + a_5x_1x_3 + a_6x_2x_3$.

The quadratic form in (4.1) and (4.2) can be written in matrix form,respectively, as:

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_1 & \frac{a_3}{2} \\ \frac{a_3}{2} & a_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ and } \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} a_1 & \frac{a_4}{2} & \frac{a_5}{2} \\ \frac{a_4}{2} & a_2 & \frac{a_6}{2} \\ \frac{a_5}{2} & \frac{a_6}{2} & a_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

The above matrix forms are of the form $x^T Ax$, where x is the column vector of variables, and A is a symmetric matrix

- whose diagonal entries are the coefficients of the squared terms.
- whose entries off the main diagonal are half the coefficients of the cross-product.

More precisely,

a) the diagonal entry in row i and column i is the coefficients of x_i^2.

b) the off diagonal entry in row i and column j is half the coefficient of the product $x_j x_i$.