Linear Algebra for Math 542

JWR

Spring 2001

Contents

Chapter 1

Preliminaries

1.1 Sets and Maps

We assume that the reader is familiar with the language of sets and maps. The most important concepts are the following:

Definition 1.1.1. Let V and W be sets and $T : V \to W$ be a map between them. The map T is called **one-one** iff $x_1 = x_2$ whenever $T(x_1) = T(x_2)$. The map T is called **onto** iff for every $y \in W$ there is an $x \in V$ such that $T(x) = y$. A map is called **one-one onto** iff it is both one-one and onto.

Remark 1.1.2. Think of the equation $y = T(x)$ as a problem to be solved for x. Then:

$$\text{the map } T : V \to W \text{ is } \left\{ \begin{array}{c} \text{one-one} \\ \text{onto} \\ \text{one-one onto} \end{array} \right\}$$

if and only if for every $y \in W$ the equation

$$y = T(x) \text{ has } \left\{ \begin{array}{c} \text{at most} \\ \text{at least} \\ \text{exactly} \end{array} \right\} \text{one solution } x \in V.$$

Example 1.1.3. The map

$$\mathbb{R} \to \mathbb{R} : x \mapsto x^3$$

is both one-one and onto since the equation

$$y = x^3$$

possesses the unique solution $y^{\frac{1}{3}} \in \mathbb{R}$ for every $y \in \mathbb{R}$. In contrast, the map

$$\mathbb{R} \to \mathbb{R} : x \to x^2$$

is not one-one since the equation

$$4 = x^2$$

has *two* distinct solutions, namely $x = 2$ and $x = -2$. It is also not onto since $-4 \in \mathbb{R}$, but the equation

$$-4 = x^2$$

has *no* solution $x \in \mathbb{R}$. The equation $-4 = x^2$ does have a complex solution $x = 2i \in \mathbb{C}$, but that solution is not relevant to the question of whether the map $\mathbb{R} \to \mathbb{R} : x \mapsto x^2$ is onto. The maps $\mathbb{C} \to \mathbb{C} : x \mapsto x^2$ and $\mathbb{R} \to \mathbb{R} : x \mapsto x^2$ are different: they have a different source and target. The map $\mathbb{C} \to \mathbb{C} : x \mapsto x^2$ *is* onto.

Definition 1.1.4. The **composition** $T \circ S$ of two maps

$$S : U \to V, \quad T : V \to W$$

is the map

$$T \circ S : U \to W$$

defined by

$$(T \circ S)(u) = T(S(u))$$

for $u \in U$. For any set V the **identity map**

$$I_V : V \to V$$

is defined by

$$I_V(v) = u$$

for $v \in V$. It satisfies the identities

$$I_V \circ S = S$$

for $S : U \to V$ and

$$T \circ I_V = T$$

for $T : V \to W$.

Definition 1.1.5 (Left Inverse). Let $T : V \to W$. A **left inverse** to T is a map $S : W \to V$ such that

$$S \circ T = I_V V.$$

Theorem 1.1.6 (Left Inverse Principle). *A map is one-one if and only if it has a left inverse.*

Proof. If $S : W \to V$ is a left inverse to $T : V \to W$, then the problem $y = T(x)$ has at most one solution: if $y = T(x_1) = T(x_2)$ then $S(y) = S(T(x_1)) = S(T(x_2))$, hence $x_1 = x_2$ since $S(T(x)) = I_V(x) = x$. Conversely, if the problem $y = T(x)$ has at most one solution, then any map $S : W \to V$ which assigns to $y \in W$ a solution x of $y = T(x)$ (when there is one) is a left inverse to T. (It does not matter what value S assigns to y when there is no solution x.) QED

Remark 1.1.7. If T is one-one but not onto the left inverse is not unique, provided that its source has at least two distinct elements. This is because when T is not onto, there is a y in the target of T which is not in the range of T. We can always make a given left inverse S into a different one by changing $S(y)$.

Definition 1.1.8 (Right Inverse). *Let $T : V \to W$. A* **right inverse** *to T is a map $R : W \to V$ such that*

$$T \circ R = I_W.$$

Theorem 1.1.9 (Right Inverse Principle). *A map is onto if and only if it has a right inverse.*

Proof. If $R : W \to V$ is a right inverse to $T : V \to W$, then $x = R(y)$ is a solution to $y = T(x)$ since $T(R(y)) = I_W(y) = y$. In other words, if T has a right inverse, it is onto. The examples below should convince the reader of the truth of the converse.

Remark 1.1.10. The assertion that there is a right inverse $R : W \to V$ to any onto map $T : V \to W$ may not seem obvious to someone who thinks of a map as a computer program: even though the problem $y = T(x)$ has a solution x, it may have many, and how is a computer program to choose?

If $V \subseteq \mathbb{N}$, one could define $R(y)$ to be the smallest $x \in V$ which solves $y = T(x)$. But this will not work if $V = \mathbb{Z}$; in this case there may not be a smallest x. In fact, this converse assertion is generally taken as an axiom, the so-called **axiom of choice**, and can neither be proved (Cohen showed this in 1963) nor disproved (Gödel showed this in 1939) from the other axioms of mathematics. It can, however, be proved in certain cases; for example, when $V \subseteq \mathbb{N}$ (we just did this). We shall also see that it can be proved in the case of matrix maps, which are the most important maps studied in these notes.

Remark 1.1.11. If T is onto but not one-one, the right inverse is not unique. Indeed, if T is not one-one, then there will be $x_1 \neq x_2$ with $T(x_1) = T(x_2)$. Let $y = T(x_1)$. Given a right inverse R we may change its value at y to produce two distinct right inverses, one which sends y to x_1 and another which sends y to x_2.

Definition 1.1.12 (Inverse). Let $T : V \to W$. A **two-sided inverse** to T is a map $T^{-1} : W \to V$ which is both a left inverse to T and a right inverse to T:

$$T^{-1} \circ T = I_V, \quad T \circ T^{-1} = I_W.$$

The word **inverse** unmodified means two-sided inverse. A map is called **invertible** iff it has a (two-sided) inverse.

As the notation suggests, inverse T^{-1} to T is unique (when it exists). The following easy proposition explains why this is so.

Theorem 1.1.13 (Unique Inverse Principle). *If a map T has both a left inverse and a right inverse, then it has a two-sided inverse. This two-sided inverse is the only one-sided inverse to T.*

Proof. Let $S : W \to V$ be a left inverse to T and $R : W \to V$ be a right inverse. Then $S \circ T = I_V$ and $T \circ R = I_W$. Compose on the right by R in the first equation to obtain $S \circ T \circ R = I_V \circ R$ and use the second to obtain $S \circ I_W = I_V \circ R$. Now composing a map with the identity (on either side) does not change the map so we have $S = R$. This says that S $(= R)$ is a two-sided identity. Now if S_1 is another left inverse to T, then this same argument shows that $S_1 = R$ (that is, $S_1 = S$). Similarly R is the only right inverse to T. QED

Definition 1.1.14 (Iteration). A map $T : V \to V$ from a set to itself can be iterated: for each non-negative integer p define $T^p : V \to V$ by

$$T^p = \underbrace{T \circ T \circ \cdots \circ T}_{p}.$$

The iterate T^p is meaningful for negative integers p as well when T is an isomorphism. Note the formulas

$$T^{p+q} = T^p \circ T^q, \qquad T^0 = I_V, \qquad (T^p)^q = T^{pq}.$$

1.2 Matrix Theory

Throughout $\mathbb{F}$ denotes a field such as the rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$, or the complex numbers $\mathbb{C}$. We assume the reader is familiar with the following operations from matrix theory:

$\mathbb{F}^{p\times q} \times \mathbb{F}^{p\times q} \to \mathbb{F}^{p\times q} : (X, Y) \mapsto X + Y$	(Addition)
$\mathbb{F} \times \mathbb{F}^{p\times q} \to \mathbb{F}^{p\times q} : (a, X) \mapsto aX$	(Scalar Multiplication)
$0 = 0_{p\times q} \in \mathbb{F}^{p\times q}$	(Zero Matrix)
$\mathbb{F}^{m\times n} \times \mathbb{F}^{n\times p} \to \mathbb{F}^{m\times p} : (A, B) \mapsto AB$	(Matrix Multiplication)
$\mathbb{F}^{m\times n} \to \mathbb{F}^{n\times m} : A \mapsto A^*$	(Transpose)
$\mathbb{F}^{m\times n} \to \mathbb{F}^{n\times m} : A \mapsto A^\dagger$	(Conjugate Transpose)
$I = I_n \in \mathbb{F}^{n\times n}$	(Identity Matrix)
$\mathbb{F}^{n\times n} \to \mathbb{F}^{n\times n} : A \mapsto A^p$	(Power)
$\mathbb{F}^{n\times n} \to \mathbb{F}^{n\times n} : A \mapsto f(A)$	(Polynomial Evaluation)

We shall assume that the reader knows the following fact which is proved by Gaussian Elimination:

Lemma 1.2.1. *Suppose that $A \in \mathbb{F}^{m\times n}$ and $n > m$. Then there is an $X \in \mathbb{F}^{m\times n}$ with $AX = 0$ but $X \neq 0$.*

The equation $AX = 0$ represents a homogeneous system of m linear equations in n unknowns so the theorem says that *a homogeneous linear system with more unknowns than equations possesses a non-trivial solution.* Using this lemma we shall prove the all-important

Theorem 1.2.2 (Dimension Theorem). *Let* $A \in \mathbb{F}^{m\times n}$ *and* $\mathbf{A} : \mathbb{F}^{n\times 1} \to \mathbb{F}^{m\times 1}$ *be the corresponding matrix map:*

$$\mathbf{A}(X) = AX$$

for $X \in \mathbb{F}^{n\times 1}$. *Then*

(1) *If* $\mathbf{A}$ *is one-one, then* $n \leq m$.

(2) *If* $\mathbf{A}$ *is onto, then* $m \leq n$.

(3) *If* $\mathbf{A}$ *is invertible, then* $m = n$.

Proof of (1). Assume $n > m$. The lemma gives $X \neq 0$ with $AX = A0$ so $\mathbf{A}$ is not one-one.

Proof of (2). Assume $m > n$. The lemma (applied to A^*) gives $H \neq 0$ with $HA = 0$. Choose $Y \in \mathbb{F}^{m\times 1}$ with $HY \neq 0$. Then for $X \in \mathbb{F}^{n\times 1}$ we have $H\mathbf{A}(X) = HAX = 0$. Hence $\mathbf{A}(X) \neq Y$ for all $X \in \mathbb{F}^{n\times 1}$ so $\mathbf{A}$ is not onto.

Proof of (3).. This follows from (1) and (2). QED

Chapter 2

Vector Spaces

A vector space is simply a space endowed with two operations, *addition* and *scalar multiplication*, which satisfy the same algebraic laws as matrix addition and scalar multiplication. The archetypal example of a vector space is the space $\mathbb{F}^{p\times q}$ of all matrices of size $p \times q$, but there are many other examples. Another example is the space $\mathrm{Poly}_n(\mathbb{F})$ of all polynomials (with coefficients from $\mathbb{F}$) of degree $\leq n$.

The vector space $\mathrm{Poly}_2(\mathbb{F})$ of all polynomials $f = f(t)$ of form $f(t) = a_0+a_1t+a_2t^2$ and the vector space $\mathbb{F}^{1\times 3}$ of all row matrices $A = \begin{bmatrix} a_0 & a_1 & a_2 \end{bmatrix}$ are *not* the same: the elements of the former space are polynomials and the elements of the latter space are matrices, and a polynomial and a matrix are different things. But there is a correspondence between the two spaces: to specify an element of either space is to specify three numbers: a_0, a_1, a_2. This correspondence preserves the vector space operations in the sense that if the polynomial f corresponds to the matrix A and the polynomial g corresponds to the matrix B then the polynomial $f + g$ corresponds to the matrix $A + B$ and the polynomial bf corresponds to the matrix bA. (This is just another way of saying that to add matrices we add their entries and to add polynomials we add their coefficients and similarly for multiplication by a scalar b.) What this means is that calculations involving polynomials can often be reduced to calculations involving matrices. This is why we make the definition of *vector space*: to help us understand what apparently different mathematical objects have in common.

2.1 Vector Spaces

Definition 2.1.1. A **vector space** over [1] $\mathbb{F}$ is a set $\mathbf{V}$ endowed with two operations:

addition	$\mathbf{V} \times \mathbf{V} \to \mathbf{V} : (\mathbf{u}, \mathbf{v}) \mapsto \mathbf{u} + \mathbf{v}$
scalar multiplication	$\mathbb{F} \times \mathbf{V} \to \mathbf{V} : (a, \mathbf{v}) \mapsto a\mathbf{v}$

and having a distinguished element $\mathbf{0} \in \mathbf{V}$ (called the **zero vector** of the vector space) and satisfying the following axioms:

$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$	(additive associative law)
$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$	(additive commutative law)
$\mathbf{u} + \mathbf{0} = \mathbf{u}$	(additive identity)
$a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$	(left distributive law)
$(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$	(right distributive law)
$a(b\mathbf{u}) = (ab)\mathbf{u}$	(multiplicative associative law)
$1\mathbf{v} = \mathbf{v}$	(multiplicative identity)
$0\mathbf{v} = \mathbf{0}$	(zero law)

for $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbf{V}$ and $a, b \in \mathbb{F}$. The elements of a vector space are sometimes called **vectors**. For vectors $\mathbf{u}$ and $\mathbf{v}$ we introduce the abbreviations

$-\mathbf{u} = (-1)\mathbf{u}$	(additive inverse)
$\mathbf{u} - \mathbf{v} = \mathbf{u} + (-\mathbf{v})$	(subtraction)

A great many other algebraic laws follow from the axioms and definitions but we shall not prove any of them. This is because for the vector spaces we study these laws are as obvious as the axioms.

Example 2.1.2. The archetypal example is:

$$\mathbf{V} = \mathbb{F}^{p\times q}$$

[1] A vector space over $\mathbb{R}$ is also called a **real vector space** and a vector space over $\mathbb{C}$ is also called a **complex vector space.**

the space of all $p \times q$ matrices with elements from $\mathbb{F}$ with the operations

$$\mathbb{F}^{p\times q} \times \mathbb{F}^{p\times q} \to \mathbb{F}^{p\times q} : (X, Y) \mapsto X + Y$$

of matrix addition and

$$\mathbb{F} \times \mathbb{F}^{p\times q} \to \mathbb{F}^{p\times q} : (a, X) \mapsto aX$$

of scalar multiplication and zero element

$$\mathbf{0} = 0_{p\times q}$$

the $p \times q$ zero matrix.

2.2 Linear Maps

Definition 2.2.1. Let $\mathbf{V}$ and $\mathbf{W}$ be vector spaces. A **linear map** from $\mathbf{V}$ to $\mathbf{W}$ is a map

$$\mathbf{T} : \mathbf{V} \to \mathbf{W}$$

(defined on $\mathbf{V}$ with values in $\mathbf{W}$) which preserves the operations of addition and scalar multiplication in the sense that

$$\mathbf{T}(\mathbf{u} + \mathbf{v}) = \mathbf{T}(\mathbf{u}) + \mathbf{T}(\mathbf{v})$$

and

$$\mathbf{T}(a\mathbf{u}) = a\mathbf{T}(\mathbf{u})$$

for $\mathbf{u}, \mathbf{v} \in \mathbf{V}$ and $a \in \mathbb{F}$.

The archetypal example is given by the following

Theorem 2.2.2. *A map $\mathbf{A} : \mathbb{F}^{n\times 1} \to \mathbb{F}^{m\times 1}$ is linear if and only if there is a (necessarily unique) matrix $A \in \mathbb{F}^{m\times n}$ such that*

$$\mathbf{A}(X) = AX$$

for all $X \in \mathbb{F}^{m\times n}$. The linear map $\mathbf{A}$ is called the **matrix map** *determined by A.*

Proof. First assume $\mathbf{A}$ is a matrix map. Then

$$\begin{aligned}\mathbf{A}(aX+bY) &= A(aX+bY)\\ &= a(AX)+b(AY)\\ &= a\mathbf{A}(X)+b\mathbf{A}(Y)\end{aligned}$$

where we have used the distributive law for matrix multiplication. This proves that $\mathbf{A}$ is linear.

Assume that $\mathbf{A}$ is linear. We must find the matrix A. Let $I_{n,j}$ be the j-th column of the $n \times n$ identity matrix:

$$I_{n,j} = \mathrm{col}_j(I_n)$$

so that

$$X = x_1 I_{n,1} + x_2 I_{n,2} + \cdots + x_n I_{n,n}$$

for $X \in \mathbb{F}^{n\times 1}$ (where $x_j = \mathrm{entry}_j(X)$ is the j-th entry of X). Let $A \in \mathbb{F}^{n\times m}$ be the matrix whose j-th column is $\mathbf{A}(I_{n,j})$:

$$\mathrm{col}_j(A) = \mathbf{A}(I_{n,j}).$$

(This formula shows the uniqueness of A.) Then for $X \in \mathbb{F}^{n\times 1}$ we have

$$\begin{aligned}\mathbf{A}(X) &= \mathbf{A}(x_1 I_{n,1} + x_2 I_{n,2} + \cdots + x_n I_{n,n})\\ &= x_1\mathbf{A}(I_{n,1}) + x_2\mathbf{A}(I_{n,2}) + \cdots + x_n\mathbf{A}(I_{n,n})\\ &= x_1\mathrm{col}_1(A) + x_2\mathrm{col}_2(A) + \cdots + x_n\mathrm{col}_n(A)\\ &= AX.\end{aligned}$$

QED

Example 2.2.3. For a given linear map $\mathbf{A}$ the proof of the Theorem 2.2.2 shows how to find the matrix A: substitute in the columns $I_{n,k} = \mathrm{col}_k(I_n)$ of the identity matrix. Here's an example. Define $\mathbf{A} : \mathbb{F}^{3\times 1} \to \mathbb{F}^{2\times 1}$ by

$$\mathbf{A}(X) = \begin{bmatrix} 3x_1 + x_3 \\ x_1 - x_2 \end{bmatrix}$$

for $X \in \mathbb{F}^{3\times 1}$ where $x_j = \mathrm{entry}_j(X)$. We find a matrix $A \in \mathbb{F}^{2\times 3}$ such that $\mathbf{A}(X) = AX$:

$$\mathbf{A}\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\right) = \begin{bmatrix}3\\1\end{bmatrix}, \quad \mathbf{A}\left(\begin{bmatrix}0\\1\\0\end{bmatrix}\right) = \begin{bmatrix}0\\-1\end{bmatrix}, \quad \mathbf{A}\left(\begin{bmatrix}0\\0\\1\end{bmatrix}\right) = \begin{bmatrix}1\\0\end{bmatrix},$$

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

Proposition 2.2.4. *The identity map* $\mathbf{I_V} : \mathbf{V} \to \mathbf{V}$ *of a vector space is linear.*

Proposition 2.2.5. *A composition of linear maps is linear.*

Corollary 2.2.6. *The iterates* $\mathbf{T}^p$ *of a linear map* $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ *from a vector space to itself are linear maps.*

Definition 2.2.7. Let $\mathbf{V}$ and $\mathbf{W}$ be vector spaces. An **isomorphism**[2] from $\mathbf{V}$ to $\mathbf{W}$ is a linear map $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ which is invertible. We say that $\mathbf{V}$ is **isomorphic** to $\mathbf{W}$ iff there is an isomorphism from $\mathbf{V}$ to $\mathbf{W}$.

Theorem 2.2.8. *The inverse of an isomorphism is an isomorphism.*

Proof. Exercise.

Proposition 2.2.9. *Isomorphisms satisfy the following properties:*

(identity) *The identity map* $\mathbf{I_V} : \mathbf{V} \to \mathbf{V}$ *of any vector space* $\mathbf{V}$ *is an isomorphism.*

(inverse) *If* $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ *is an isomorphism, then so is its inverse* $\mathbf{T}^{-1} : \mathbf{W} \to \mathbf{V}$.

(composition) *If* $\mathbf{S} : \mathbf{U} \to \mathbf{V}$ *and* $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ *are isomorphisms, then so is the composition* $\mathbf{T} \circ \mathbf{S} : \mathbf{U} \to \mathbf{W}$.

Corollary 2.2.10. *Isomorphism is an equivalence relation. This means that it satisfies the following conditions:*

(reflexivity) *Every vector space is isomorphic to itself.*

(symmetry) *If* $\mathbf{V}$ *is isomorphic to* $\mathbf{W}$, *then* $\mathbf{W}$ *is isomorphic to* $\mathbf{V}$.

(transitivity) *If* $\mathbf{U}$ *is isomorphic to* $\mathbf{V}$ *and* $\mathbf{V}$ *is isomorphic to* $\mathbf{W}$, *then* $\mathbf{U}$ *is isomorphic to* $\mathbf{W}$.

[2] The word *isomorphism* is commonly used in mathematics, with a variety of analogous - but different - meanings. It comes from the Greek: *iso* meaning *same* and *morphos* meaning *structure*. The idea is that isomorphic objects should have the same properties.

2.3 Space of Linear Maps

Let $\mathbf{V}$ and $\mathbf{w}$ be vector spaces. Denote by $\mathcal{L}(\mathbf{V}, \mathbf{W})$ the **space of linear maps** from $\mathbf{V}$ to $\mathbf{W}$. Thus $\mathbf{T} \in \mathcal{L}(\mathbf{V}, \mathbf{W})$ if and only if

(i) $\mathbf{T} : \mathbf{V} \to \mathbf{W}$,

(ii) $\mathbf{T}(\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{T}(\mathbf{v}_1) + \mathbf{T}(\mathbf{v}_2)$ for $\mathbf{v}_1, \mathbf{v}_2 \in \mathbf{V}$,

(iii) $\mathbf{T}(a\mathbf{v}) = a\mathbf{T}(\mathbf{v})$ for $\mathbf{v} \in \mathbf{V}$, $a \in \mathbb{F}$.

Linear operations on maps from $\mathbf{V}$ to $\mathbf{W}$ are defined **point-wise**. This means:

(1) If $\mathbf{T}, \mathbf{S} : \mathbf{V} \to \mathbf{W}$, then $(\mathbf{T} + \mathbf{S}) : \mathbf{V} \to \mathbf{W}$ is defined by

$$(\mathbf{T} + \mathbf{S})(\mathbf{v}) = \mathbf{T}(\mathbf{v}) + \mathbf{S}(\mathbf{v}).$$

(2) If $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ and $a \in \mathbb{F}$, then $(a\mathbf{T}) : \mathbf{V} \to \mathbf{W}$ is defined by

$$(a\mathbf{T})(\mathbf{v}) = a\mathbf{T}(\mathbf{v}).$$

(3) $\mathbf{0} : \mathbf{V} \to \mathbf{W}$ is defined by

$$\mathbf{0}(\mathbf{v}) = \mathbf{0}.$$

Proposition 2.3.1. *These operations preserve linearity. In other words,*

(1) $\mathbf{T}, \mathbf{S} \in \mathcal{L}(\mathbf{V}, \mathbf{W}) \implies \mathbf{T} + \mathbf{S} \in \mathcal{L}(\mathbf{V}, \mathbf{W})$,

(2) $\mathbf{T} \in \mathcal{L}(\mathbf{V}, \mathbf{W}), a \in \mathbb{F} \implies a\mathbf{T} \in \mathcal{L}(\mathbf{V}, \mathbf{W})$,

(3) $\mathbf{0} \in \mathcal{L}(\mathbf{V}, \mathbf{W})$.

(Here $\implies$ means *implies*.)

Hint for proof: For example, to prove (1) assume that $\mathbf{T}$ and $\mathbf{S}$ satisfy (ii) and (iii) above and show that $\mathbf{T} + \mathbf{S}$ also does. By similar methods one can also prove that

Proposition 2.3.2. *These operations make* $\mathcal{L}(\mathbf{V}, \mathbf{W})$ *a vector space.*

The last two propositions make possible the following

Corollary 2.3.3. *The map*

$$\mathbb{F}^{m\times n} \to \mathcal{L}(\mathbb{F}^{n\times 1}, \mathbb{F}^{m\times 1}) : A \mapsto \mathbf{A}$$

(which assigns to each matrix A the matrix map $\mathbf{A}$ determined by A) is an isomorphism.

2.4 Frames and Matrix Representation

The space $\mathbb{F}^{n\times 1}$ of all column matrices of a given size is the standard example of a vector space, but not the only example. This space is well suited to calculations with the computer since computers are good at manipulating arrays of numbers. Now we'll introduce a device for converting problems about vector spaces into problems in matrix theory.

Definition 2.4.1. A **frame** for a vector space $\mathbf{V}$ is an isomorphism

$$\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$$

from the standard vector space $\mathbb{F}^{n\times 1}$ to the given vector space $\mathbf{V}$

The idea is that $\mathbf{\Phi}$ assigns co-ordinates $X \in \mathbb{F}^{n\times 1}$ to a vector $\mathbf{v} \in \mathbf{V}$ via the equation

$$\mathbf{v} = \mathbf{\Phi}(X).$$

These co-ordinates enable us to transform problems about vectors into problems about matrices. The frame is a way of 'naming' the vectors $\mathbf{v}$; the 'names' are the column matrices X. The following propositions are immediate consequences of the Isomorphism Laws and show that there are lots of frames for a vector space.

Let $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$, be a frame for the vector space $\mathbf{V}$, $\mathbf{\Psi} : \mathbb{F}^{m\times 1} \to \mathbf{W}$, be a frame for the vector space $\mathbf{W}$, and $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ be a linear map. These determine a linear map

$$\mathbf{A} : \mathbb{F}^{n\times 1} \to \mathbb{F}^{m\times 1}$$

by

$$\mathbf{A} = \mathbf{\Psi}^{-1} \circ \mathbf{T} \circ \mathbf{\Phi}. \tag{1}$$

According to the Theorem 2.2.2 a linear map for $\mathbb{F}^{n\times 1}$ to $\mathbb{F}^{m\times 1}$ is a matrix map. Thus there is a matrix $A \in \mathbb{F}^{m\times n}$ with

$$\mathbf{A}(X) = AX \tag{2}$$

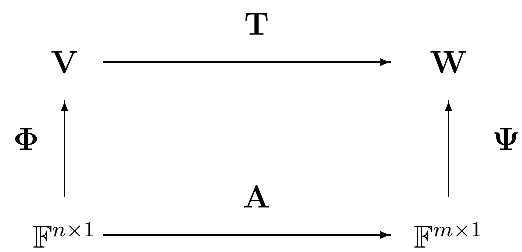
for $X \in \mathbb{F}^{n\times 1}$.

Definition 2.4.2 (Matrix Representation). We call the matrix A determined by (1) and (2) **matrix representing $\mathbf{T}$** in the frames $\mathbf{\Phi}$ and $\mathbf{\Psi}$ and say A **represents $\mathbf{T}$** in the frames $\mathbf{\Phi}$ and $\mathbf{\Psi}$. When $\mathbf{V} = \mathbf{W}$ and $\mathbf{\Phi} = \mathbf{\Psi}$ we also call the matrix A the **matrix representing $\mathbf{T}$** in the frame $\mathbf{\Phi}$ and say that A **represents $\mathbf{T}$** in the frame $\mathbf{\Phi}$.

Equation (1) says that

$$\mathbf{\Psi}(AX) = \mathbf{T}(\mathbf{\Phi}(X))$$

for $X \in \mathbb{F}^{n\times 1}$. The following diagram provides a handy way of summarizing this:



Matrix representation is used to convert problems in linear algebra to problems in matrix theory. The laws in this section justify the use of matrix representation as a computational tool.

Proposition 2.4.3. *Fix frames* $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ *and* $\mathbf{\Psi} : \mathbb{F}^{m\times 1} \to \mathbf{W}$ *as above. Then the map*

$$\mathbb{F}^{m\times n} \to \mathcal{L}(\mathbf{V}, \mathbf{W}) : A \mapsto \mathbf{T} = \mathbf{\Psi} \circ \mathbf{A} \circ \mathbf{\Phi}^{-1}$$

is an isomorphism. The inverse of this isomorphism is the map which assigns to each linear map $\mathbf{T}$ *the matrix* A *which represents* $\mathbf{T}$ *in the frames* $\mathbf{\Phi}$ *and* $\mathbf{\Psi}$.

Proof. This isomorphism is the composition of two isomorphisms. The first is the isomorphism

$$\mathbb{F}^{m\times n} \to \mathcal{L}(\mathbb{F}^{n\times 1}, \mathbb{F}^{m\times 1}) : A \mapsto \mathbf{A}$$

of the Theorem 2.3.3 and the second is the isomorphism

$$\mathcal{L}(\mathbb{F}^{n\times 1}, \mathbb{F}^{m\times 1}) \to \mathcal{L}(\mathbf{V}, \mathbf{W}) : \mathbf{A} \mapsto \mathbf{\Psi} \circ \mathbf{A} \circ \mathbf{\Phi}^{-1}.$$

The rest of the argument is routine. QED

Remark 2.4.4. The theorem asserts two kinds of linearity. In the first place the expression

$$\mathbf{T}(\mathbf{v}) = \mathbf{\Psi} \circ \mathbf{A} \circ \mathbf{\Phi}^{-1}(\mathbf{v})$$

is linear in $\mathbf{v}$ for fixed $\mathbf{A}$. This is the meaning of the assertion that $\mathbf{T} \in \mathcal{L}(\mathbf{V}, \mathbf{W})$. In the second place the expression is linear in $\mathbf{A}$ for fixed $\mathbf{v}$. This is the meaning of the assertion that the map $A \mapsto \mathbf{T}$ is linear.

Exercise 2.4.5. Show that for any frame $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ the identity matrix I_n represents the identity transformation $\mathbf{I_V} : \mathbf{V} \to \mathbf{V}$ in the frame $\mathbf{\Phi}$.

Exercise 2.4.6. Show that for any frame $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ the identity matrix I_n represents the identity transformation $\mathbf{I_V} : \mathbf{V} \to \mathbf{V}$ in the frame $\mathbf{\Phi}$.

Exercise 2.4.7. Suppose

$$\mathbf{\Upsilon} : \mathbb{F}^{p\times 1} \to \mathbf{U}, \quad \mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}, \quad \mathbf{\Psi} : \mathbb{F}^{m\times 1} \to \mathbf{W},$$

are frames for vector spaces $\mathbf{U}$, $\mathbf{V}$, $\mathbf{W}$, respectively and that

$$\mathbf{S} : \mathbf{U} \to \mathbf{V}, \quad \mathbf{T} : \mathbf{V} \to \mathbf{W},$$

are linear maps. Let $A \in \mathbb{F}^{m\times n}$ represent $\mathbf{T}$ in the frames $\mathbf{\Phi}$ and $\mathbf{\Psi}$ and $B \in \mathbb{F}^{n\times p}$ represent $\mathbf{S}$ in the frames $\mathbf{\Upsilon}$ and $\mathbf{\Phi}$. Show that the product $AB \in \mathbb{F}^{p\times n}$ represents the composition

$$\mathbf{T} \circ \mathbf{S} : \mathbf{U} \to \mathbf{W}$$

in the frames $\mathbf{\Upsilon}$ and $\mathbf{\Phi}$. (In other words composition of linear maps corresponds to multiplication of the representing matrices.)

Exercise 2.4.8. Suppose that $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ is a linear map from a vector space to itself, that $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ is a frame, and that $A \in \mathbb{F}^{n\times n}$ represents $\mathbf{T}$ in the frame $\mathbf{\Phi}$. Show that for every non-negative integer p, the power A^p represents the iterate $\mathbf{T}^p$ in the frame $\mathbf{\Phi}$. If $\mathbf{T}$ is invertible (so that A is invertible), then this holds for negative integers p as well.

Exercise 2.4.9. Let

$$f(t) = \sum_{p=0}^{m} b_p t^p$$

be a polynomial. We can evaluate f on a linear map $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ from a vector space to itself. The result is the linear map $f(\mathbf{T}) : \mathbf{V} \to \mathbf{V}$ defined by

$$f(\mathbf{T}) = \sum_{p=0}^{m} b_p \mathbf{T}^p.$$

Suppose that $\mathbf{T}$, $\mathbf{\Phi}$, A, are as in Exercise 2.4.8. Show that the matrix $f(A)$ represents the map $f(\mathbf{T})$ in the frame $\mathbf{\Phi}$.

Exercise 2.4.10. The **dual space** of a vector space $\mathbf{V}$ is the space

$$\mathbf{V}^* = \mathcal{L}(\mathbf{V}, \mathbb{F})$$

of linear maps with values in $\mathbb{F}$. Show that the map

$$\mathbb{F}^{1\times n} \to \left(\mathbb{F}^{n\times 1}\right)^* : H \mapsto \mathbf{H}$$

defined by

$$\mathbf{H}(X) = HX$$

for $X \in \mathbb{F}^{n\times 1}$ is an isomorphism between $\mathbb{F}^{1\times n}$ and the dual space of $\mathbb{F}^{n\times 1}$. (We do not distinguish $F^{1\times 1}$ and $\mathbb{F}$.)

Exercise 2.4.11. A linear map $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ determines a dual linear map $\mathbf{T}^* : \mathbf{W}^* \to \mathbf{V}^*$ via the formula

$$\mathbf{T}^*(\alpha) = \alpha \circ \mathbf{T}$$

for $\alpha \in \mathbf{W}^*$. Suppose that A is the matrix representing $\mathbf{T}$ in the frames $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ and $\mathbf{\Psi} : \mathbb{F}^{m\times 1} \to \mathbf{W}$. Find frames $\mathbf{\Phi'} : \mathbb{F}^{n\times 1} \to \mathbf{V}^*$ and $\mathbf{\Psi'} : \mathbb{F}^{m\times 1} \to \mathbf{W}^*$ such that the matrix representing $\mathbf{T}^*$ in this frames is the transpose A^*.

2.5 Null Space and Range

Let $\mathbf{V}$ and $\mathbf{W}$ be vector spaces and

$$\mathbf{T} : \mathbf{V} \to \mathbf{W}$$

be a linear map. The **null space** of the linear map $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ is the set $\mathcal{N}(\mathbf{T})$ of all vectors $\mathbf{v} \in \mathbf{V}$ which are mapped to $\mathbf{0}$ by $\mathbf{T}$:

$$\mathcal{N}(\mathbf{T}) = \{\mathbf{v} \in \mathbf{V} : \mathbf{T}(\mathbf{v}) = \mathbf{0}\}.$$

(The null space is also called the **kernel** by some authors.) The **range** of $\mathbf{T}$ is the set $\mathcal{R}(\mathbf{T})$ of all vectors $\mathbf{w} \in \mathbf{W}$ of form $w = \mathbf{T}(\mathbf{v})$ for some $\mathbf{v} \in \mathbf{V}$:

$$\mathcal{R}(\mathbf{T}) = \{\mathbf{T}(\mathbf{v}) : \mathbf{v} \in \mathbf{V}\}.$$

To decide if a vector $\mathbf{v}$ is an element of the null space of $\mathbf{T}$ we first check that it lies in $\mathbf{V}$ (if $\mathbf{v}$ fails this test it is *not* in $\mathcal{N}(\mathbf{T})$) and then apply $\mathbf{T}$ to $\mathbf{v}$; if we obtain $\mathbf{0}$ then $\mathbf{v} \in \mathcal{N}(\mathbf{T})$, otherwise $\mathbf{v} \notin \mathcal{N}(\mathbf{T})$.

To decide if a vector $\mathbf{w}$ is an element of the range of $\mathbf{T}$ we first check that it lies in $\mathbf{W}$ (if $\mathbf{w}$ fails this test it is *not* in $\mathcal{R}(\mathbf{T})$) and then attempt to solve the equation $\mathbf{w} = \mathbf{T}(\mathbf{v})$ for $\mathbf{v} \in \mathbf{V}$. If we obtain a solution $\mathbf{v} \in \mathbf{V}$, then $\mathbf{w} \in \mathcal{R}(\mathbf{T})$ otherwise $\mathbf{w} \notin \mathcal{R}(\mathbf{T})$. (Warning: It is conceivable that the formula defining $\mathbf{T}(\mathbf{v})$ makes sense for certain $\mathbf{v}$ which are not elements of $\mathbf{V}$; in this case the equation $\mathbf{w} = \mathbf{T}(\mathbf{v})$ may have a solution $\mathbf{v}$ but *not* a solution with $\mathbf{v} \in \mathbf{V}$. If this happens $\mathbf{w} \notin \mathcal{R}(\mathbf{T})$.)

Theorem 2.5.1 (One-One/NullSpace). *A linear map* $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ *is one-one if and only if* $\mathcal{N}(\mathbf{T}) = \{\mathbf{0}\}$.

Proof. If $\mathcal{N}(\mathbf{T}) = \{\mathbf{0}\}$ and $\mathbf{v}_1$ and $\mathbf{v}_2$ are two solutions of $\mathbf{w} = \mathbf{T}(\mathbf{v})$ then $\mathbf{T}(\mathbf{v}_1) = \mathbf{w} = \mathbf{T}(\mathbf{v}_2)$ so $\mathbf{0} = \mathbf{T}(v_1) - \mathbf{T}(v_2) = \mathbf{T}(\mathbf{v}_1 - \mathbf{v}_2)$ so $\mathbf{v}_1 - \mathbf{v}_2 \in \mathcal{N}(\mathbf{T}) = \{\mathbf{0}\}$ so $\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{0}$ so $\mathbf{v}_1 = \mathbf{v}_2$. Conversely if $\mathcal{N}(\mathbf{T}) \neq \{\mathbf{0}\}$ then there is a $\mathbf{v}_1 \in \mathcal{N}(\mathbf{T})$ with $\mathbf{v}_1 \neq \mathbf{0}$ so the equation $\mathbf{0} = \mathbf{T}(\mathbf{v})$ has two distinct solutions namely $\mathbf{v} = \mathbf{v}_1$ and $\mathbf{v} = \mathbf{0}$. QED

Remark 2.5.2 (Onto/Range). A map $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ is onto if and only if $\mathbf{W} = \mathcal{R}(\mathbf{T})$

2.6 Subspaces

Definition 2.6.1. Let $\mathbf{V}$ be a vector space. A **subspace** of $\mathbf{V}$ is a subset $\mathbf{W} \subseteq \mathbf{V}$ which contains the zero vector of $\mathbf{V}$ and is closed under the operations of addition and scalar multiplication, that is, which satisfies

(zero) $\mathbf{0} \in \mathbf{W}$;

(addition) $\mathbf{u} + \mathbf{v} \in \mathbf{W}$ whenever $\mathbf{u} \in \mathbf{W}$ and $\mathbf{v} \in \mathbf{W}$;

(scalar multiplication) $a\mathbf{u} \in \mathbf{W}$ whenever $a \in \mathbb{F}$ and $\mathbf{u} \in \mathbf{W}$;

Remark 2.6.2. If $\mathbf{W}$ is a subspace of a vector space $\mathbf{V}$, then $\mathbf{W}$ is a vector space in its own right: the vector space operations are those of $\mathbf{V}$. Thus any theorem about vector spaces applies to subspaces.

Theorem 2.6.3. *The null space $\mathcal{N}(\mathbf{T})$ of the linear map $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ is a vector subspace of the vector space $\mathbf{V}$.*

Proof. The space $\mathcal{N}(\mathbf{T})$ contains the zero vector since $\mathbf{T}(\mathbf{0}) = \mathbf{0}$. If $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{N}(\mathbf{T})$ then $\mathbf{T}(\mathbf{v}_1) = \mathbf{T}(\mathbf{v}_2) = \mathbf{0}$ so $\mathbf{T}(\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{T}(\mathbf{v}_1) + \mathbf{T}(\mathbf{v}_2) = \mathbf{0} + \mathbf{0} = \mathbf{0}$ so $\mathbf{v}_1 + \mathbf{v}_2 \in \mathcal{N}(\mathbf{T})$. If $\mathbf{v} \in \mathcal{N}(\mathbf{T})$ and $a \in \mathbb{F}$ then $\mathbf{T}(a\mathbf{v}) = a\mathbf{T}(\mathbf{v}) = a\mathbf{0} = \mathbf{0}$ so that $a\mathbf{v} \in \mathbb{F}$. Hence $\mathcal{N}(\mathbf{T})$ is a subspace. QED

Theorem 2.6.4. *The range $\mathcal{R}(\mathbf{T})$ of the linear map $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ is a subspace of the vector space $\mathbf{W}$.*

Proof. The space $\mathcal{R}(\mathbf{T})$ contains the zero vector since since $\mathbf{T}(\mathbf{0}) = \mathbf{0}$. If $\mathbf{w}_1, \mathbf{w}_2 \in \mathcal{R}(\mathbf{T})$ then $\mathbf{T}(\mathbf{v}_1) = \mathbf{w}_1$ and $\mathbf{T}(\mathbf{v}_2) = \mathbf{w}_2$ for some $\mathbf{v}_1, \mathbf{v}_2 \in \mathbf{V}$ so $\mathbf{w}_1 + \mathbf{w}_2 = \mathbf{T}(\mathbf{v}_1) + \mathbf{T}(\mathbf{v}_2) = \mathbf{T}(\mathbf{v}_1 + \mathbf{v}_2)$ so $\mathbf{w}_1 + \mathbf{w}_2 \in \mathcal{R}(\mathbf{T})$. If $\mathbf{w} \in \mathcal{R}(\mathbf{T})$ and $a \in \mathbb{F}$ then $\mathbf{w} = \mathbf{T}(\mathbf{v})$ for some $\mathbf{v} \in \mathbf{V}$ so $a\mathbf{w} = a\mathbf{T}(\mathbf{v}) = \mathbf{T}(a\mathbf{v})$ so $a\mathbf{w} \in \mathcal{R}(\mathbf{T})$. Hence $\mathcal{R}(\mathbf{T})$ is a subspace. QED

2.7 Examples

2.7.1 Matrices

The spaces $\mathbf{V} = \mathbb{F}^{p\times q}$ are all vector spaces. A frame $\mathbf{\Phi} : \mathbb{F}^{pq\times 1} \to \mathbb{F}^{p\times q}$ can be constructed be taking the first row of $\mathbf{\Phi}(X)$ to be the first q entries of X, the second row to be the second q entries of X and so on. For example, with $p = q = 2$ we get

$$\mathbf{\Phi}\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}\right) = \begin{bmatrix} x_1 & x_1 \\ x_2 & x_4 \end{bmatrix}.$$

In case $p = 1$ and $q = n$ this frame is the transpose map

$$\mathbb{F}^{n\times 1} \to \mathbb{F}^{1\times n} : X \mapsto X^*.$$

More generally, for any p and q the transpose map

$$\mathbb{F}^{p\times q} \to \mathbb{F}^{q\times p} : X \mapsto X^*$$

is an isomorphism. The inverse of the transpose map from $\mathbb{F}^{p\times q}$ to $\mathbb{F}^{q\times p}$ is the transpose map from $\mathbb{F}^{q\times p}$ to $\mathbb{F}^{p\times q}$. (Proof: $(X^*)^* = X$ and $(H^*)^* = H$.)

Suppose $P \in \mathbb{F}^{n\times n}$ and $Q \in \mathbb{F}^{m\times m}$ are invertible. Then the maps

$$\begin{aligned}
&\mathbb{F}^{n\times k} \to \mathbb{F}^{n\times k} : Y \mapsto QY \\
&\mathbb{F}^{k\times n} \to \mathbb{F}^{k\times n} : H \mapsto HP \\
&\mathbb{F}^{m\times n} \to \mathbb{F}^{m\times n} : A \mapsto QDP^{-1}
\end{aligned}$$

are all isomorphisms. The first of these has been called the **matrix map** determined by Q and denoted by **Q**.

Question 2.7.1. What are the inverses of these isomorphisms? (Answer: The inverse of $Y \mapsto QY$ is $Y_1 \mapsto Q^{-1}Y_1$. The inverse of $H \mapsto HP$ is $H_1 \mapsto H_1P^{-1}$. The inverse of $A \mapsto QAP^{-1}$ is $B \mapsto Q^{-1}BP$.)

2.7.2 Polynomials

An important example is the space $\mathrm{Poly}_n(\mathbb{F})$ of all polynomials of degree $\leq n$. This is the space of all functions $f : \mathbb{F} \to \mathbb{F}$ of form

$$f(t) = c_0 + c_1 t + c_2 t^2 + \cdots + c_n t^n$$

for $t \in \mathbb{F}$. Here the coefficients $c_0, c_1, c_2, \ldots, c_n$ are chosen from $\mathbb{F}$. The vector space operations on $\mathrm{Poly}_n(\mathbb{F})$ are defined pointwise meaning that

$$(f + g)(t) = f(t) + g(t), \quad (bf)(t) = b(f(t))$$

for $f, g \in \mathrm{Poly}_n(\mathbb{F})$ and $b \in \mathbb{F}$. This means that the vector space operations are also performed 'coefficientwise', as if the coefficients $c_0, c_1, \ldots, c_n$ were entries in a matrix: If

$$f(t) = c_0 + c_1 t + c_2 t^2 + \cdots + c_n t^n$$

and

$$g(t) = b_0 + b_1 t + b_2 t^2 + \cdots + b_n t^n$$

then

$$f(t) + g(t) = (c_0 + b_0) + (c_1 + b_1)t + (c_2 + b_2)t^2 + \cdots + (c_n + b_n)t^n$$

and

$$bf(t) = (bc_0) + (bc_1)t + (bc_2)t^2 + \cdots + (bc_n)t^n.$$

Question 2.7.2. Suppose $f, g \in \mathrm{Poly}_2(\mathbb{F})$ are given by

$$f(t) = 2 - 6t + 3t^2, \quad g(t) = 4 + 7t.$$

What is $5f - 2g$? (Answer: $5f(t) - 2g(t) = 2 - 44t + 15t^2$.)

If $n \leq m$ the space $\mathrm{Poly}_n(\mathbb{F})$ of all polynomials of degree $\leq n$ is a subspace of the space $\mathrm{Poly}_m(\mathbb{F})$ of all polynomials of degree $\leq m$:

$$\mathrm{Poly}_n(\mathbb{F}) \subseteq \mathrm{Poly}_m(\mathbb{F}) \text{ for } n \leq m.$$

A typical element f of $\mathrm{Poly}_m(\mathbb{F})$ has form

$$f(t) = c_0 + c_1 t + c_2 t^2 + \cdots + c_m t^m$$

and f is an element of the smaller space $\mathrm{Poly}_n(\mathbb{F})$ exactly when $c_{n+1} = c_{n+2} = \cdots = c_m = 0$. For example, $\mathrm{Poly}_2(\mathbb{F}) \subseteq \mathrm{Poly}_5(\mathbb{F})$ since every polynomial f whose degree is ≤ 2 has degree ≤ 5. A frame

$$\mathbf{\Phi} : \mathbb{F}^{(n+1)\times 1} \to \mathrm{Poly}_n(\mathbb{F})$$

for $\mathrm{Poly}_n(\mathbb{F})$ is defined by

$$\mathbf{\Phi}\left(\begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}\right)(t) = c_0 + c_1 t + c_2 t^2 + \cdots + c_n t^n$$

This frame is called the **standard frame** for $\mathrm{Poly}_n(\mathbb{F})$. For example, with $n = 2$:

$$\mathbf{\Phi}\left(\begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix}\right)(t) = c_0 + c_1 t + c_2 t^2$$

Remark 2.7.3. Think about the notation $\mathbf{\Phi}(X)(t)$. The frame $\mathbf{\Phi}$ accepts a input a matrix $X \in \mathbb{F}^{n\times 1}$ and produces as output a polynomial $\mathbf{\Phi}(X)$. The polynomial $\mathbf{\Phi}(X)$ is itself a map which accepts as input a real number $t \in \mathbb{R}$ and produces as output a number $\mathbf{\Phi}(X)(t) \in \mathbb{F}$. The equation $\mathbf{\Phi}(X) = f$ might be expressed in words as *the entries of* X *are the coefficients of* f.

Any $a \in \mathbb{R}$ determines an isomorphism $\mathbf{T}_a : \mathrm{Poly}_n(\mathbb{F}) \to \mathrm{Poly}_n(\mathbb{F})$ via

$$\left(\mathbf{T}_a(f)\right)(t) = f(t+a).$$

The inverse is given by $(\mathbf{T}_a)^{-1} = \mathbf{T}_{-a}$. The composition $\mathbf{T}_{-a} \circ \mathbf{\Phi}\mathbb{F}^{(n+1)\times 1} \to \mathrm{Poly}_n(\mathbb{F})$ of the standard frame $\mathbf{\Phi}$ with the isomorphism $\mathbf{T}_{-a}$ is given by

$$\left(\mathbf{T}_{-a} \circ \mathbf{\Phi}\right)(X)(t) = \sum_{k=0}^{n} b_k (t-a)^k$$

where $b_k = \mathrm{entry}_{k+1}(X)$. The inverse of this new frame is easily computed using **Taylor's Identity**:

$$f(t) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(t-a)k$$

for $f \in \mathrm{Poly}_n(\mathbb{F})$. Here $f^{(k)}(a)$ denotes the k-th derivative of f evaluated at a.

2.7.3 Trigonometric Polynomials

The vector space $\mathrm{Trig}_n(\mathbb{F})$ is the space of all functions $f : \mathbb{R} \to \mathbb{F}$ of form

$$f(t) = a_0 + \sum_{k=1}^{n} a_k \cos(kt) + b_k \sin(kt)$$

for $t \in \mathbb{R}$. Here the coefficients $b_n, \ldots, b_2, b_1, a_0, a_1, a_2, \ldots, a_n$ are arbitrary elements of $\mathbb{F}$. This space is called the space of **trigonometric polynomials** of degree $\leq n$ with coefficients from $\mathbb{F}$. The vector space operations are performed pointwise (and hence coefficientwise) as for polynomials. Two important subspaces of $\mathrm{Trig}_n(\mathbb{F})$ are

$$\mathrm{Cos}_n(\mathbb{F}) = \{f \in \mathrm{Trig}_n(\mathbb{F}) : f(-t) = f(t)\}$$

called the space of **even trigonometric polynomials** and

$$\mathrm{Sin}_n(\mathbb{F}) = \{f \in \mathrm{Trig}_n(\mathbb{F}) : f(-t) = -f(t)\}.$$

called the space of **odd trigonometric polynomials**. The following proposition justifies the notation.

Proposition 2.7.4. (1) *When $\mathbb{F} = \mathbb{C}$ the space $\mathrm{Trig}_n(\mathbb{F})$ is the space of all functions of form*

$$f(t) = \sum_{k=-n}^{n} c_k e^{ikt}.$$

(2) *The subspace $\mathrm{Cos}_n(\mathbb{F})$ is the space of all functions $g : \mathbb{R} \to \mathbb{F}$ of form*

$$g(t) = a_0 + a_1 \cos(t) + a_2 \cos(2t) + \cdots + a_n \cos(nt).$$

(3) *The subspace $\mathrm{Sin}_n(\mathbb{F})$ is the space of all functions $h : \mathbb{R} \to \mathbb{F}$ of form*

$$h(t) = b_1 \sin(t) + b_2 \sin(2t) + \cdots + b_n \sin(nt)$$

for $t \in \mathbb{R}$.

A frame

$$\mathbf{\Phi}_{SC} : \mathbb{F}^{(2n+1)\times 1} \to \mathrm{Trig}_n(\mathbb{F})$$

for $\mathrm{Trig}_n(\mathbb{F})$ is given by

$$\mathbf{\Phi}_{SC}\left(\begin{bmatrix} b_n \\ \vdots \\ b_1 \\ a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}\right)(t) = a_0 + \sum_{k=1}^{n} a_k \cos(kt) + b_k \sin(kt).$$

When $\mathbb{F} = \mathbb{C}$ another frame

$$\mathbf{\Phi}_E : \mathbb{F}^{(2n+1)\times 1} \to \mathrm{Trig}_n(\mathbb{F})$$

is given by

$$\mathbf{\Phi}_E\left(\begin{bmatrix} c_{-n} \\ \vdots \\ c_{-1} \\ c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix}\right)(t) = \sum_{k=-n}^{n} c_k e^{ikt}.$$

A frame

$$\mathbf{\Phi}_C : \mathbb{F}^{(n+1)\times 1} \to \mathrm{Trig}_n(\mathbb{F})$$

for $\mathrm{Cos}_n(\mathbb{F})$ is given by

$$\mathbf{\Phi}_C\left(\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}\right)(t) = a_0 + \sum_{k=1}^{n} a_k \cos(kt).$$

A frame for $\mathrm{Sin}_n(\mathbb{F})$ is given by

$$\mathbf{\Phi}_S : \mathbb{F}^{n\times 1} \to \mathrm{Trig}_n(\mathbb{F})$$

for $\mathrm{Sin}_n(\mathbb{F})$ is given by

$$\mathbf{\Phi}_S\left(\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}\right)(t) = \sum_{k=1}^{n} b_k \sin(kt).$$

If $n \leq m$ then the space $\mathrm{Sin}_n(\mathbb{F})$ is a subspace of $\mathrm{Sin}_m(\mathbb{F})$, the space $\mathrm{Cos}_n(\mathbb{F})$ is a subspace of $\mathrm{Cos}_m(\mathbb{F})$, and the space $\mathrm{Trig}_n(\mathbb{F})$ is a subspace of $\mathrm{Trig}_m(\mathbb{F})$.

Example 2.7.5. The function $f : \mathbb{R} \to \mathbb{F}$ defined by

$$f(t) = \sin^2(t)$$

is an element of $\mathrm{Cos}_2(\mathbb{F})$ because it can be written in the form

$$f(t) = a_0 + a_1 \cos(t) + a_2 \cos(2t)$$

(with $a_0 = -a_2 = 1/2$, $a_1 = 0$) by the half angle formula

$$\sin^2(t) = \frac{1}{2} - \frac{1}{2}\cos(2t)$$

from trigonometry.

2.7.4 Derivative and Integral

Recall from calculus the rules for differentiating and integrating polynomials:

$$f'(t) = a_1 + 2a_2 t + 3a_3 t^2 + \cdots + n a_n t^{n-1}$$

$$\int_c^t f(t)\,dt = -c + a_0 t + \frac{a_1}{2} t^2 + \cdots + \frac{a_n}{n+1} t^{n+1}$$

for

$$f(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n.$$

These operations are linear:

$$(b_1 f_1 + b_2 f_2)'(t) = b_1 f_1'(t) + b_2 f_2'(t),$$

$$\int_c^t (b_1 f_1(t) + b_2 f_2(t))\,dt = b_1 \int_c^t f_1(t)\,dt + b_2 \int_c^t f_2(t)\,dt.$$

Hence the formulas[3]

$$\mathbf{T}(f) = f', \quad \mathbf{S}(f)(t) = \int_0^t f(t)\,dt.$$

define linear maps

$$\mathbf{T} : \mathrm{Poly}_n(\mathbb{F}) \to \mathrm{Poly}_{n-1}(\mathbb{F}), \quad \mathbf{S} : \mathrm{Poly}_n(\mathbb{F}) \to \mathrm{Poly}_{n+1}(\mathbb{F})$$

Beginners find this a bit confusing: the maps **T** and **S** accept polynomials as input and produce polynomials as output. But a polynomial is (among other things) a map. Thus **T** is a map whose inputs are maps and whose outputs are maps.

[3]Changing the lower limit in the integral from 0 to some other number c gives a different linear map **S**.

Question 2.7.6. Is $\mathbf{T}$ one-one? onto? What about $\mathbf{S}$? (Answer: $\mathbf{T}$ is not one-one since $f' = 0$ if f is a constant. $\mathbf{T}$ is onto since $f' = g$ if $g(t) = \int_0^t f(t)\,dt$. $\mathbf{S}$ is not onto since $\mathbf{S}(f)(0) = 0$ for all f so we can never solve $\mathbf{S}(f) = 1$ (the constant polynomial). $\mathbf{S}$ is onto since $\mathbf{S}(f') = f$.)

Remark 2.7.7. Recall that the maps $\mathbf{T}_1 : \mathbf{V}_1 \to \mathbf{W}_1$ and $\mathbf{T}_2 : \mathbf{V}_2 \to \mathbf{W}_2$ are **equal** iff $\mathbf{V}_1 = \mathbf{V}_2$, $\mathbf{W}_1 = \mathbf{W}_2$, and $\mathbf{T}_1(\mathbf{v}) = \mathbf{T}_2(\mathbf{v})$ for all $\mathbf{v} \in \mathbf{V}_1$. By this definition two maps $\mathbf{T}_1 : \mathbf{V}_1 \to \mathbf{W}_1$ and $\mathbf{T}_2 : \mathbf{V}_2 \to \mathbf{W}_2$ are unequal if either the sources $\mathbf{V}_1$ and $\mathbf{V}_2$ are different or the targets $\mathbf{W}_1$ and $\mathbf{W}_2$ are different. For example, differentiation also determines a linear map $\mathrm{Poly}_n(\mathbb{F}) \to \mathrm{Poly}_n(\mathbb{F}) : f \mapsto f'$ and we will distinguish this from the linear map $\mathrm{Poly}_n(\mathbb{F}) \to \mathrm{Poly}_{n-1}(\mathbb{F}) : f \mapsto f'$ since the targets are different. (The latter is onto, the former is not.)

The formula $\mathbf{T}(f) = f'$ can be used to define many other interesting linear maps depending on the choice of the source and target form $\mathbf{T}$. For example, if $f \in \mathrm{Sin}_n(\mathbb{F})$, then $f' \in \mathrm{Cos}_n(\mathbb{F})$. The exercises at the end of the chapter treat some examples like this.

2.8 Exercises

Exercise 2.8.1. Let g_1 and g_2 be the polynomials given by

$$g_1(t) = 6 - 5t + t^2, \quad g_2(t) = 2 + 3t + 4t^2,$$

and define vector spaces

$$\mathbf{V}_1 = \mathbb{F}^{3\times 1}, \quad \mathbf{V}_2 = \mathbb{F}^{4\times 1}, \quad \mathbf{V}_3 = \mathrm{Poly}_2(\mathbb{F}), \quad \mathbf{V}_4 = \mathrm{Poly}_3(\mathbb{F}),$$

and elements

$$\mathbf{v}_1 = \begin{bmatrix} 6 \\ -5 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}, \quad \mathbf{v}_3 = g_1, \quad \mathbf{v}_4 = g_2.$$

For which pairs (i, j) is it true that $\mathbf{v}_i \in \mathbf{V}_j$?

Exercise 2.8.2. In the notation of the previous exercise define subspaces

$$\begin{aligned}
\mathbf{W}_1 &= \{\begin{bmatrix} a & b & c \end{bmatrix} : 6a - 5b + c = 0\} \\
\mathbf{W}_2 &= \{f \in \mathbf{V}_3 : f(2) = 0\} \\
\mathbf{W}_3 &= \{f \in \mathbf{V}_3 : f(1) = f(2) = 0\} \\
\mathbf{W}_4 &= \{f \in \mathbf{V}_4 : f(1) = f(2) = 0\}
\end{aligned}$$

When is $\mathbf{v}_i \in \mathbf{W}_j$?

Exercise 2.8.3. In the notation of the previous exercise which of the set inclusions $\mathbf{W}_i \subseteq \mathbf{W}_j$ are true?

Let us distinguish truth and nonsense. Only a meaningful equation can be true or false. An equation is nonsense if it contains some notation (like 0/0) which has not been defined or if it equates two objects of different types such as a polynomial and a matrix. Mathematicians thus distinguish two levels of error. The equation $2 + 2 = 5$ is false, but at least meaningful. The equation

$$3 + \begin{bmatrix} 4 & 0 \end{bmatrix} = 7 \text{ (nonsense)}$$

is meaningless - *neither true nor false* - since we have not defined how to add a number to a 1×2 matrix. Philosophers sometimes call an error like this a **category error**. Another sort of category error is illustrated by the equation

$$f = \begin{bmatrix} a & b & c \end{bmatrix} \text{ (nonsense)}$$

where $f(t) = a + bt + ct^2$.

Exercise 2.8.4. Continue the notation of the previous exercise and define a map

$$\mathbf{T} : \mathbb{F}^{1\times 3} \to \mathrm{Poly}_2(\mathbb{F})$$

by

$$\mathbf{T}\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right)(t) = a + bt + ct^2.$$

Which of the equations $\mathbf{T}(\mathbf{v}_i) = \mathbf{v}_j$ are meaningful? Which of the equations $\mathbf{T}(\mathbf{W}_i) = \mathbf{W}_j$ are meaningful? Of the meaningful ones which are true?

Exercise 2.8.5. Define $\mathbf{A} : \mathbb{F}^{2\times 1} \to \mathbb{F}^{2\times 1}$ by

$$\mathbf{A}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 5x_1 + 4x_2 \\ 3x_2 \end{bmatrix}.$$

Find the matrix A such that $\mathbf{A}(X) = AX$.

Exercise 2.8.6. Prove that a map

$$\mathbf{T} : \mathbb{F}^{1\times m} \to \mathbb{F}^{1\times n}$$

is a linear map if and only if there is a (necessarily unique) matrix $A \in \mathbb{F}^{m\times n}$ such that

$$\mathbf{T}(H) = HA$$

for all $H \in \mathbb{F}^{1\times m}$.

Exercise 2.8.7. For which of the following pairs $\mathbf{V}$, $\mathbf{W}$ of vector spaces does the formula $\mathbf{T}(f) = f'$ define a linear map $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ with source $\mathbf{V}$ and target $\mathbf{W}$?

(1) $\mathbf{V} = \mathrm{Poly}_3(\mathbb{F})$, $\mathbf{W} = \mathrm{Poly}_5(\mathbb{F})$.
(2) $\mathbf{V} = \mathrm{Poly}_3(\mathbb{F})$, $\mathbf{W} = \mathrm{Poly}_2(\mathbb{F})$.
(3) $\mathbf{V} = \mathrm{Cos}_3(\mathbb{F})$, $\mathbf{W} = \mathrm{Sin}_3(\mathbb{F})$.
(4) $\mathbf{V} = \mathrm{Sin}_3(\mathbb{F})$, $\mathbf{W} = \mathrm{Cos}_3(\mathbb{F})$.
(5) $\mathbf{V} = \mathrm{Cos}_3(\mathbb{F})$, $\mathbf{W} = \mathrm{Trig}_3(\mathbb{F})$.
(6) $\mathbf{V} = \mathrm{Trig}_3(\mathbb{F})$, $\mathbf{W} = \mathrm{Cos}_3(\mathbb{F})$.
(7) $\mathbf{V} = \mathrm{Poly}_3(\mathbb{F})$, $\mathbf{W} = \mathrm{Cos}_3(\mathbb{F})$.

Exercise 2.8.8. In each of the following you are given vector spaces $\mathbf{V}$ and $\mathbf{W}$, frames $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ and $\mathbf{\Psi} : \mathbb{F}^{m\times 1} \to \mathbf{W}$, a linear map $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ and a matrix $A \in \mathbb{F}^{m\times n}$. Verify that the matrix A represents the map $\mathbf{T}$ in the frames $\mathbf{\Phi}$ and $\mathbf{\Psi}$ by proving the identity $\mathbf{\Psi}(AX) = \mathbf{T}(\mathbf{\Phi}(X))$.

(1) $\mathbf{V} = \mathrm{Poly}_2(\mathbb{F})$, $\mathbf{W} = \mathrm{Poly}_1(\mathbb{F})$, $\mathbf{\Phi}(X)(t) = x_1 + x_2 t + x_3 t^2$, $\mathbf{\Psi}(Y)(t) = y_1 + y_2 t$, $\mathbf{T}(f) = f'$,

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

(2) $\mathbf{V}$, $\mathbf{W}$, $\mathbf{\Phi}$, $\mathbf{\Psi}$ as in (1), $\mathbf{T}(f)(t) = (f(t+h) - f(t))/h$,

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

(3) $\mathbf{V} = \mathrm{Cos}_2(\mathbb{F})$, $\mathbf{W} = \mathrm{Sin}_1(\mathbb{F})$, $\mathbf{\Phi}(X)(t) = x_1 + x_2\cos(t) + x_3\cos(2t)$, $\mathbf{\Psi}(Y)(t) = y_1\sin(t) + y_2\sin(2t)$, $\mathbf{T}(f) = f'$,

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

(4) $\mathbf{V}$ and $\mathbf{\Phi}$ as in (1), $\mathbf{W} = \mathbb{F}^{1\times 3}$, $\mathbf{\Psi}(Y) = Y^*$,

$$\mathbf{T}(f)(t) = \begin{bmatrix} f(0) & f(1) & f(2) \end{bmatrix}, \qquad A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 1 & 4 \end{bmatrix}.$$

Here $x_j = \text{entry}_j(X)$ and $y_i = \text{entry}_i(Y)$.

Exercise 2.8.9. In each of the following you are given a vector space $\mathbf{V}$, a frame $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$, a linear map $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ from $\mathbf{V}$ to itself, and a matrix $A \in \mathbb{F}^{n\times n}$. Verify that the matrix A represents the map $\mathbf{T}$ in the frame $\mathbf{\Phi}$ by proving the identity $\mathbf{\Phi}(AX) = \mathbf{T}(\mathbf{\Phi}(X))$.

(1) $\mathbf{V} = \text{Poly}_2(\mathbb{F})$, $\mathbf{\Phi}(X)(t) = x_1 + x_2 t + x_3 t^2$, $\mathbf{T}(f) = f'$,

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

(2) $\mathbf{V}$ and $\mathbf{\Phi}$ as in (1), $\mathbf{T}(f)(t) = (f(t+h) - f(t))/h$,

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

(3) $\mathbf{V} = \text{Trig}_1(\mathbb{F})$, $\mathbf{\Phi}(X)(t) = x_1 + x_2\cos(t) + x_3\sin(t)$, $\mathbf{T}(f) = f'$,

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

(4) $\mathbf{V}$ and $\mathbf{\Phi}$ as in (3), $\mathbf{T}(f)(t) = (f(t+h) - f(t))/h$,

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -h^{-1}(1-\cos h) & h^{-1}\sin h \\ 0 & -h^{-1}\sin h & -h^{-1}(1-\cos h) \end{bmatrix}.$$

Here $x_j = \text{entry}_j(X)$.

Exercise 2.8.10. Which of the following linear maps $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ is one-one? onto?

1. $\mathbf{T} : \mathrm{Poly}_3(\mathbb{F}) \to \mathrm{Poly}_2(\mathbb{F}) : \mathbf{T}(f) = f'$.

2. $\mathbf{T} : \mathrm{Poly}_3(\mathbb{F}) \to \mathrm{Poly}_3(\mathbb{F}) : \mathbf{T}(f) = f'$.

3. $\mathbf{T} : \mathrm{Poly}_2(\mathbb{F}) \to \mathrm{Poly}_3(\mathbb{F}) : \mathbf{T}(f) = \int f$.

4. $\mathbf{T} : \mathrm{Poly}_2(\mathbb{F}) \to \mathrm{Poly}_4(\mathbb{F}) : \mathbf{T}(f) = \int f$.

5. $\mathbf{T} : \mathrm{Sin}_3(\mathbb{F}) \to \mathrm{Cos}_3(\mathbb{F}) : \mathbf{T}(f) = f'$.

6. $\mathbf{T} : \mathrm{Cos}_3(\mathbb{F}) \to \mathrm{Sin}_3(\mathbb{F}) : \mathbf{T}(f) = f'$.

7. $\mathbf{T} : \mathrm{Sin}_3(\mathbb{F}) \to \mathrm{Cos}_3(\mathbb{F}) : \mathbf{T}(f) = \int f$.

Here f' denotes the derivative of f and $\int f$ stands for the function F defined by

$$F(t) = \int_0^t f(\tau)\, d\tau.$$

(If the map is not one-one find a non-zero f with $\mathbf{T}(f) = \mathbf{0}$. If the map is not onto find a g with $\mathbf{T}(f) \neq g$ for all f. If the map is one-one find a left inverse. If the map is onto find a right inverse.)

Question 2.8.11. Conspicuously absent from the list of linear maps in the last problem is a map $\mathrm{Cos}_3(\mathbb{F}) \to \mathrm{Sin}_3(\mathbb{F}) : \mathbf{T}(f) = \int f$. Why? (Answer: The constant function $f(t) = 1$ is in the space $\mathrm{Cos}_3(\mathbb{F})$ but its integral $F(t) = t$ is not in the space $\mathrm{Sin}_3(\mathbb{F})$.)

Exercise 2.8.12. The map $\mathbf{T} : \mathrm{Poly}_3(\mathbb{F}) \to \mathrm{Poly}_3(\mathbb{F})$ defined by

$$\mathbf{T}(f)(t) = f(t+2)$$

is an isomorphism. What is $\mathbf{T}^{-1}$?

Exercise 2.8.13. Let $A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix}$ and let $\mathbf{A} : \mathbb{F}^{4\times 1} \to \mathbb{F}^{2\times 1}$ be the corresponding linear map. Find a frame $\mathbf{\Phi} : \mathbb{F}^{2\times 1} \to \mathcal{N}(\mathbf{A})$.

Exercise 2.8.14. Let $\mathbf{V} = \{f \in \mathrm{Poly}_3(\mathbb{F}) : f(1) = f(-1) = 0\}$. Find a frame $\mathbf{\Phi} : \mathbb{F}^{2\times 1} \to \mathbf{V}$. Hint: This problem is a little bit like the preceding one.

Exercise 2.8.15. Show that the map

$$\mathrm{Poly}_n(\mathbb{F}) \to \mathbb{F}^{1\times 3} : f \mapsto \begin{bmatrix} f(0) & f(1) & f(2) \end{bmatrix}$$

is one-one for $n \leq 2$ and onto for $n \geq 2$. Show that it is not one-one for $n > 2$ and not onto for $n = 1$.

Exercise 2.8.16. Let

$$\mathbf{V} = \{f \in \mathrm{Poly}_n(\mathbb{F}) : f(0) = 0\}$$

and define $\mathbf{T} : \mathbf{V} \to \mathrm{Poly}_{n-1}(\mathbb{F})$ by $\mathbf{T}(f) = f'$. Show that $\mathbf{T}$ is an isomorphism and find its inverse.

Exercise 2.8.17. Show that the map

$$\mathrm{Poly}_n(\mathbb{F}) \to \mathrm{Poly}_n(\mathbb{F}) : f \mapsto F$$

where

$$F(t) = t^{-1} \int_0^t f(t)\, dt$$

is an isomorphism. What is its inverse?

Exercise 2.8.18. For each of the following four spaces $\mathbf{V}$ the formula

$$\mathbf{T}(f) = f''$$

defines a linear map $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ from $\mathbf{V}$ to itself.

(1) $\mathbf{V} = \mathrm{Poly}_3(\mathbb{F})$

(2) $\mathbf{V} = \mathrm{Trig}_3(\mathbb{F})$

(3) $\mathbf{V} = \mathrm{Cos}_3(\mathbb{F})$

(4) $\mathbf{V} = \mathrm{Sin}_3(\mathbb{F})$

In which of these four cases is $\mathbf{T}$ invertible? In which of these four cases is $\mathbf{T}^4 = \mathbf{0}$?

Chapter 3

Bases and Frames

In this chapter we relate the notion of frame to the notion of basis as explained in the first course in linear algebra. The two notions are essentially the same (if you look at them right).

3.1 Maps and Sequences

Let $\mathbf{V}$ be a vector space, $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ be a linear map, and $(\phi_1, \phi_2, \ldots, \phi_n)$ be a sequence of elements of $\mathbf{V}$. We say that the linear map $\mathbf{\Phi}$ and the sequence $(\phi_1, \phi_2, \ldots, \phi_n)$ **correspond** iff

$$\phi_j = \mathbf{\Phi}(I_{n,j}) \tag{1}$$

for $j = 1, 2, \ldots, n$ where $I_{n,j} = \mathrm{col}_j(I_n)$ is the j-th column of the identity matrix.

Theorem 3.1.1. *A linear map $\mathbf{\Phi}$ and a sequence $(\phi_1, \phi_2, \ldots, \phi_n)$ correspond iff*

$$\mathbf{\Phi}(X) = x_1\phi_1 + x_2\phi_2 + \cdots + x_n\phi_n \tag{2}$$

for all $X \in \mathbb{F}^{n\times 1}$. Here $x_j = \mathrm{entry}_j(X)$. Hence, every sequence corresponds to a unique linear map.

Proof. Exercise. (Read the rest of this section first.)

Question 3.1.2. Why is the map $\mathbf{\Phi}$ defined by (2) linear? (Answer: $\mathbf{\Phi}(aX+bY) = \sum_j (ax_j + by_j)\phi_j = a\left(\sum_j x_j\phi_j\right) + b\left(\sum_j y_j\phi_j\right) = a\mathbf{\Phi}(X) + b\mathbf{\Phi}(Y)$.)

Theorem 3.1.3. *Let $\mathbf{V}^n$ denote the set of sequences of length n from the vector space $\mathbf{V}$, and $\mathcal{L}(\mathbb{F}^{n\times 1}, \mathbf{V})$ denote the set of linear maps from $\mathbb{F}^{n\times 1}$ to $\mathbf{V}$. Then the map*

$$\mathcal{L}(\mathbb{F}^{n\times 1}, \mathbf{V}) \to \mathbf{V}^n : \mathbf{\Phi} \to \ (\mathbf{\Phi}(I_{n,1}), \mathbf{\Phi}(I_{n,2}), \ldots, \mathbf{\Phi}(I_{n,n}))$$

is one-one and onto.

Proof. Exercise.

Remark 3.1.4. Thus the sequence $(\phi_1, \phi_2, \ldots, \phi_n)$ and the corresponding linear map $\mathbf{\Phi}$ carry the same information: each determines the other uniquely. We will distinguish them carefully for they are set-theoretically distinct. The sequence is an operation which accepts as input an integer j between 1 and n and produces as output an element ϕ_j in the vector space $\mathbf{V}$. The linear map is an operation which accepts as input an element X of the vector space $\mathbb{F}^{n\times 1}$ and produces as output an element $\mathbf{\Phi}(X)$ in the vector space $\mathbf{V}$.

Example 3.1.5. In the special case $n = 2$

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ x_2 \end{bmatrix} = x_1 I_{2,1} + x_2 I_{2,2}$$

so equation (2) is

$$\phi_1 = \mathbf{\Phi}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right), \quad \phi_2 = \mathbf{\Phi}\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right).$$

and equation (1) is

$$\mathbf{\Phi}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = x_1\phi_1 + x_2\phi_2$$

Example 3.1.6. Suppose $\mathbf{V} = \mathbb{F}^{m\times 1}$ and form the matrix $A \in \mathbb{F}^{m\times n}$ with columns $\phi_1, \phi_2, \ldots, \phi_n$:

$$\phi_j = \mathrm{col}_j(A)$$

for $j = 1, 2, \ldots, n$. Now

$$AX = x_1\phi_1 + x_2\phi_2 \cdots + x_n\phi_n$$

where $x_j = \mathrm{entry}_j(X)$. This says that $\mathbf{\Phi}(X) = AX$. Hence (in this special case) the map $\mathbf{\Phi}$ goes by two names: it is the *map corresponding to the sequence* $(\phi_1, \phi_2, \ldots, \phi_n)$ and it is the *matrix map determined by the matrix A Remember that this is a special case; the map corresponding to a sequence is a matrix map only when* $\mathbf{V} = \mathbb{F}^{m\times 1}$.

Example 3.1.7. Suppose $\mathbf{V} = \mathbb{F}^{1\times m}$ and that

$$\phi_i = \mathrm{row}_i(B), \qquad i = 1, 2, \dots, n$$

are the rows of $B \in \mathbb{F}^{n\times m}$. Then the map $\mathbf{\Phi}$ is given by

$$\mathbf{\Phi}(X) = X^* B$$

where X^* is the transpose of X.

Example 3.1.8. Recall that $\mathrm{Poly}_n(\mathbb{F})$ is the space of polynomials

$$f(t) = x_0 + x_1 t + x_2 t^2 + \cdots + x_n t^n$$

of degree $\leq n$ with coefficients from $\mathbb{F}$. For $k = 0, 1, 2, \dots, n$ define $\phi_k \in \mathrm{Poly}_n(\mathbb{F})$ by

$$\phi_k(t) = t^k.$$

Then the corresponding map

$$\mathbf{\Phi} : \mathbb{F}^{(n+1)\times 1} \to \mathrm{Poly}_n(\mathbb{F})$$

is defined by $\mathbf{\Phi}(X) = f$ where the coefficients of f are the entries of X: $x_k = \mathrm{entry}_{k+1}(X)$ for $k = 0, 1, 2, \dots, n$. For example, with $n = 2$:

$$\mathbf{\Phi}\left(\begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix}\right)(t) = x_0 + x_1 t + x_2 t^2$$

3.2 Independence

Definition 3.2.1. The sequence $(\phi_1, \phi_2, \dots, \phi_n)$ is (linearly) **independent** iff the only solution $x_1, x_2, \dots, x_n \in \mathbb{F}$ of

$$x_1\phi_1 + x_2\phi_2 + \cdots + x_n\phi_n = \mathbf{0} \qquad (\clubsuit)$$

is the trivial solution $x_1 = x_2 = \cdots = x_n = 0$. The sequence $(\phi_1, \phi_2, \dots, \phi_n)$ is called **dependent** iff it is not independent, that is, iff equation ($\clubsuit$) possesses a non-trivial solution, (i.e. one with at least one $x_i \neq 0$).

Remark 3.2.2. It is easy to confuse the words *independent* and *dependent*. It helps to remember the etymology. Equation (♣) asserts a relation among the elements of the sequence. Thus the sequence is *dependent* when its elements satisfy a non-trivial relation. Note also that we have worded the definition in terms of a *sequence* of matrices rather than a *set*: repetitions are relevant. Thus the sequence (ϕ_1, ϕ_1, ϕ_2) is dependent, since $x_1\phi_1 + x_2\phi_1 + x_3\phi_2 = 0$ for $x_1 = 1$, $x_2 = -1$, and $x_3 = 0$.

Question 3.2.3. Is the sequence (ϕ_1, ϕ_2) dependent if $\phi_2 = \mathbf{0}$? (Answer: Yes, because then $0\phi_1 + 1\phi_2 = 0$).

Theorem 3.2.4 (One-One/Independence). *Let $(\phi_1, \ldots, \phi_n)$ be a sequence of vectors in the vector space $\mathbf{V}$ and $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ be the corresponding map $\mathbf{\Phi}$. Then the following are equivalent:*

(1) *The sequence $(\phi_1, \phi_2, \ldots, \phi_n)$ is independent.*

(2) *The corresponding map $\mathbf{\Phi}$ is one-one.*

(3) *The null space of the corresponding linear map consists only of the zero vector:*

$$\mathcal{N}(\mathbf{\Phi}) = \{\mathbf{0}\}.$$

Proof. By the definition of $\mathbf{\Phi}$ we can write equation (♣) in the form

$$\mathbf{\Phi}(X) = \mathbf{0} \text{ where } X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

To say that the sequence $(\phi_1, \phi_2, \ldots, \phi_n)$ is independent is to say that the only solution of $\mathbf{\Phi}(X) = \mathbf{0}$ is $X = 0$; hence parts (1) and (3) are equivalent. According to the Theorem 2.5.1 parts (2) and (3) are equivalent. QED

Example 3.2.5. For $A \in \mathbb{F}^{m\times n}$ let $A_j = \mathrm{col}_j(A) \in \mathbb{F}^{m\times 1}$ be the j-th column of A and $x_j = \mathrm{entry}_j(X)$ be the j-th entry of $X \in \mathbb{F}^{1\times n}$. Then

$$AX = x_1A_1 + x_2A_2 + \cdots + x_nA_n.$$

Hence *the columns of A are independent if and only if the only solution of the homogeneous system $AX = 0$ is $X = 0$.*

Example 3.2.6. Similarly, the rows of A are independent if and only if the only solution of the dual homogeneous system $HA = 0$ is $H = 0$.

3.3 Span

Definition 3.3.1. Let $\mathbf{V}$ be a vector space and $(\phi_1, \phi_2, \ldots, \phi_n)$ be a sequence of vectors from $\mathbf{V}$. The sequence **spans** $\mathbf{V}$ if and only if every element $\mathbf{v}$ of $\mathbf{V}$ is expressible as a linear combination of $(\phi_1, \phi_2, \ldots, \phi_n)$, that is, for every $\mathbf{v} \in \mathbf{V}$ there exist scalars $x_1, x_2, \ldots, x_n$ such that

$$\mathbf{v} = x_1\phi_1 + x_2\phi_2 + \cdots + x_n\phi_n. \qquad (\diamondsuit)$$

Theorem 3.3.2 (Onto/Spanning). *Let $(\phi_1, \phi_2, \ldots, \phi_n)$ be a sequence of vectors from the vector space $\mathbf{V}$ and $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ be the corresponding map $\mathbf{\Phi}$. Then the following are equivalent:*

(1) *The sequence $(\phi_1, \phi_2, \ldots, \phi_n)$ spans the vector space $\mathbf{V}$.*

(2) *The corresponding map $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ is onto.*

(3) $\mathcal{R}(\mathbf{\Phi}) = \mathbf{V}$.

Proof. By the definition of $\mathbf{\Phi}$ we can write equation $(\diamondsuit)$ in the form

$$\mathbf{v} = \mathbf{\Phi}(X) \quad \text{where} \quad X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

To say that the sequence $(\phi_1, \phi_2, \ldots, \phi_n)$ spans is to say that there is a solution of $\mathbf{V} = \mathbf{\Phi}(X)$ no matter what is $\mathbf{v} \in \mathbf{V}$; hence parts (1) and (2) are equivalent. Parts (2) and (3) are trivially equivalent for the range $\mathcal{R}(\mathbf{\Phi})$ of $\mathbf{\Phi}$ is by definition the set of all vectors $\mathbf{v}$ of form $\mathbf{v} = \mathbf{\Phi}(X)$. (See Remark 2.5.2.) QED

Example 3.3.3. For $A \in \mathbb{F}^{m\times n}$ let $A_j = \mathrm{col}_j(A) \in \mathbb{F}^{m\times 1}$ be the j-the column of A and $x_j = \mathrm{entry}_j(X)$ be the j-th entry of $X \in \mathbb{F}^{1\times n}$. Then

$$AX = x_1A_1 + x_2A_2 + \cdots + x_nA_n.$$

Hence *the columns of A span the vector space $\mathbb{F}^{m\times 1}$ if and only if for every column $Y \in \mathbb{F}^{m\times 1}$ the inhomogeneous system $Y = AX$ is has a solution X.*

Example 3.3.4. Similarly,the rows of A span $\mathbb{F}^{1\times n}$ if and only if for every row $K \in \mathbb{F}^{1\times n}$ the dual inhomogeneous system $K = HA$ has a solution $H \in \mathbb{F}^{1\times m}$.

Definition 3.3.5. Every sequence $\phi_1, \phi_2, \ldots, \phi_n$ spans *some* vector space, namely the space

$$\mathrm{Span}(\phi_1, \phi_2, \ldots, \phi_n) = \mathcal{R}(\mathbf{\Phi})$$

which is called the **vector space spanned by** the sequence $(\phi_1, \phi_2, \ldots, \phi_n)$. Here $\phi_1, \phi_2, \ldots, \phi_n \in \mathbf{V}$ where $\mathbf{V}$ is a vector space, and $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ is the linear map corresponding to this sequence. Thus a sequence $(\phi_1, \phi_2, \ldots, \phi_n)$ of elements of $\mathbf{V}$ spans $\mathbf{V}$ if and only if

$$\mathrm{Span}(\phi_1, \phi_2, \ldots, \phi_n) = \mathbf{V}.$$

Remark 3.3.6. Let $\mathbf{V}$ be a vector space and $\mathbf{W}$ be a subspace of $\mathbf{V}$: $\mathbf{W} \subseteq \mathbf{V}$. Let $\phi_1, \phi_2, \ldots, \phi_n$ be elements of $\mathbf{V}$. Then the following are equivalent:

(1) $\phi_j \in \mathbf{W}$ for $j = 1, 2, \ldots, n$;

(2) $\mathrm{Span}(\phi_1, \phi_2, \ldots, \phi_n) \subseteq \mathbf{W}$.

Exercise 3.3.7. Prove this.

3.4 Basis and Frame

Definition 3.4.1. A **basis** for the vector space $\mathbf{V}$ is a sequence of vectors in $\mathbf{V}$ which is both independent and spans $\mathbf{V}$. Recall (see Definition 2.4.1 that a **frame** for the vector space $\mathbf{V}$ is an isomorphism

$$\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}.$$

Theorem 3.4.2 (Frame and Basis). *The sequence* $(\phi_1, \ldots, \phi_n)$ *of vectors in* $\mathbf{V}$ *is a basis for* $\mathbf{V}$ *if and only the corresponding linear map*

$$\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$$

is a frame.

Proof. The sequence $(\phi_1, \phi_2, \ldots, \phi_n)$ is a basis iff it is independent and spans $\mathbf{V}$. By Theorem 3.2.4 the sequence $(\phi_1, \phi_2, \ldots, \phi_n)$ is independent iff the map $\mathbf{\Phi}$ is one-one. By Theorem 3.3.2 the sequence $(\phi_1, \phi_2, \ldots, \phi_n)$ spans $\mathbf{V}$ iff map $\mathbf{\Phi}$ is onto. According to the definition of *isomorphism*, the map $\mathbf{\Phi}$ is a frame iff it is invertible. QED

One should think of the vector space $\mathbf{V}$ as a "geometric space" and of the basis $(\phi_1, \phi_2, \ldots, \phi_n)$ as a vehicle for introducing co-ordinates in $\mathbf{V}$. The correspondence $\mathbf{\Phi}$ between the "numerical space" $\mathbb{F}^{n\times 1}$ and the geometric space $\mathbf{V}$ constitutes a co-ordinate system on $\mathbf{V}$. This means that the entries of the column

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

should be viewed as the "co-ordinates" of the vector

$$\mathbf{v} = x_1\phi_1 + x_2\phi_2 + \ldots + x_n\phi_n = \mathbf{\Phi}(X).$$

When $\mathbf{v} = \mathbf{\Phi}(X)$ we say that the matrix X **represents** the vector $\mathbf{v}$ in the frame $\mathbf{\Phi}$.

In any particular problem we try to choose the basis $(\phi_1, \phi_2, \ldots, \phi_n)$ (that is, the frame $\mathbf{\Phi}$) so that numerical description of the problem is as simple as possible. The notation just introduced can (if used systematically) be of great help in clarifying our thinking.

3.5 Examples and Exercises

Definition 3.5.1. The columns of the identity matrix

$$I_{n,1} = \mathrm{col}_1(I_n),\ I_{n,2} = \mathrm{col}_2(I_n),\ \ldots,\ I_{n,n} = \mathrm{col}_n(I_n)$$

form a basis for $F^{n\times 1}$ called the **standard basis** for $F^{n\times 1}$.

The standard basis for $\mathbb{F}^{3\times 1}$ is

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Note the obvious equation

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

This equation shows that every $X \in \mathbb{F}^{3\times 1}$ has a unique expression as a linear combination of the vectors $I_{3,j}$; the coefficients x_1, x_2, x_3 are precisely the entries in the column matrix x. Thus $(I_{n,1}, I_{n,2}, \ldots, I_{n,n})$ is a basis for $\mathbb{F}^{3\times 1}$ as claimed. (The same argument works for arbitrary n to show that the standard basis is a basis.)

Question 3.5.2. What is the frame corresponding to the standard basis? (Answer: The identity map of $\mathbb{F}^{n\times 1}$.)

Proposition 3.5.3. *Let $B_1, B_2, \ldots, B_n \in F^{n\times n}$ and let $B \in \mathbb{F}^{n\times n}$ be matrix having these as columns:*

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

Then the sequence $(B_1, B_2, \ldots, B_n)$ is a basis for $\mathbb{F}^{n\times 1}$ if and only if the matrix B is invertible. The frame corresponding to this basis is the isomorphism the matrix map $\mathbf{B}$ determined by B.

Proof. We have

$$\mathbf{B}(X) = BX = x_1 B_1 + x_2 B_2 \cdots + x_n B_n$$

where $x_j = \text{entry}_j(X)$. Hence (in this special case) the map $\mathbf{B}$ goes by two names: it is the *map corresponding to the sequence* $(B_1, B_2, \ldots, B_n)$, and it is the *matrix map determined by the matrix* B. The map $\mathbf{B}$ is an isomorphism iff the matrix B is invertible. By Theorem 3.4.2, the sequence is a basis iff the corresponding map $\mathbf{B}$ is an isomorphism. QED

Exercise 3.5.4. The vectors

$$B_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

form a basis for $\mathbb{F}^{2\times 1}$ since the matrix $B = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$ is invertible. Find the unique numbers x_1, x_2 such

$$\begin{bmatrix} 1 \\ 9 \end{bmatrix} = x_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Example 3.5.5. The set $\{\mathbf{0}\}$ consisting of the single element $\mathbf{0} \in \mathbf{V}$ is a subspace of the vector space $\mathbf{V}$. It is called the **zero subspace**. By convention the **empty sequence** () is a basis for the zero vector space.

Example 3.5.6. Suppose that the numbers a, b, c are not all zero. Let $\mathbf{V}$ be the set of all $\begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{F}^{1\times 3}$ such that $ax + by + cz = 0$. Geometrically, $\mathbf{V}$ is a plane through the origin. If $c \neq 0$, a basis is given by by

$$\phi_1 = \begin{bmatrix} c \\ 0 \\ -a \end{bmatrix}, \quad \phi_2 = \begin{bmatrix} 0 \\ c \\ -b \end{bmatrix}.$$

To prove this we must show three things: (1) that $\phi_1, \phi_2 \in \mathbf{V}$, (2) that the sequence (ϕ_1, ϕ_2) is independent, and (3) that the sequence (ϕ_1, ϕ_2) spans $\mathbf{V}$. Part (1) follows from the calculations

$$a(c) + b(0) + c(-a) = 0, \quad a(0) + b(c) + c(-b) = 0.$$

Part (2) follows from the equation

$$x_1\phi_1 + x_2\phi_2 = \begin{bmatrix} cx_1 \\ cx_2 \\ -ax_1 - bx_2 \end{bmatrix}$$

so that (as $c \neq 0$) the equation $x_1\phi_1 + x_2\phi_2 = 0$ implies $x_1 = x_2 = 0$. Part (3) follows from the observation that if $ax + by + cz = 0$, then

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{x}{c}\phi_1 + \frac{y}{c}\phi_2.$$

Example 3.5.7. Let

$$R = \begin{bmatrix} 1 & 0 & c_{13} & c_{14} & c_{15} \\ 0 & 1 & c_{23} & c_{24} & c_{25} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

A basis for the null space of the matrix map determined by R is (ϕ_1, ϕ_2, ϕ_3) where

$$\phi_1 = \begin{bmatrix} -c_{13} \\ -c_{23} \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \phi_2 = \begin{bmatrix} -c_{14} \\ -c_{24} \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \phi_3 = \begin{bmatrix} -c_{15} \\ -c_{25} \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Example 3.5.8. Let

$$R = \begin{bmatrix} 1 & c_{11} & 0 & c_{12} & 0 & c_{13} & c_{14} \\ 0 & c_{21} & 1 & c_{22} & 0 & c_{23} & c_{24} \\ 0 & c_{31} & 0 & c_{32} & 1 & c_{33} & c_{34} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

A basis $(\phi_1, \phi_2, \phi_3, \phi_4)$ for the null space of matrix map determined by R is

$$\phi_1 = \begin{bmatrix} -c_{11} \\ 1 \\ -c_{21} \\ 0 \\ -c_{31} \\ 0 \\ 0 \end{bmatrix}, \quad \phi_2 = \begin{bmatrix} -c_{12} \\ 0 \\ -c_{22} \\ 1 \\ -c_{32} \\ 0 \\ 0 \end{bmatrix}, \quad \phi_3 = \begin{bmatrix} -c_{13} \\ 0 \\ -c_{23} \\ 0 \\ -c_{33} \\ 1 \\ 0 \end{bmatrix}, \quad \phi_4 = \begin{bmatrix} -c_{14} \\ 0 \\ -c_{24} \\ 0 \\ -c_{34} \\ 0 \\ 1 \end{bmatrix}.$$

Example 3.5.9. Recall that $\mathrm{Poly}_n(\mathbb{F})$ is the space of all polynomials of degree $\leq n$. This is the space of all functions $f : \mathbb{F} \to \mathbb{F}$ of form

$$f(t) = a_0 +_1 t + a_2 t^2 + \cdots + a_n t^n$$

for $t \in \mathbb{F}$. Here the coefficients $a_0, a_1, a_2, \ldots, a_n$ are chosen from $\mathbb{F}$. A frame

$$\mathbf{\Phi} : \mathbb{F}^{(n+1)\times 1} \to \mathrm{Poly}_n(\mathbb{F})$$

is given by $\mathbf{\Phi}(X) = f$ where

$$X = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

where the coefficients $a_0, a_1, a_2, \ldots, a_n$ of $f \in \mathrm{Poly}_n(\mathbb{F})$ are the entries of $X \in \mathbb{F}^{(n+1)\times 1}$. In other words, the polynomials

$$\phi_k(t) = t^k \text{ for } k = 0, 1, 2, \ldots, n$$

form a basis for $\mathrm{Poly}_n(\mathbb{F})$.

Exercise 3.5.10. Verify the formula

$$a_k = \frac{f^k(0)}{k!}$$

for a polynomial $f \in \mathrm{Poly}_n(\mathbb{F})$ and $k = 0, 1, 2, \ldots, n$. Here the numerator $f^k(0)$ is the k-th derivative of $f = f(t)$ with respect to t evaluated at $t = 0$. (This formula proves that the frame $\mathbf{\Phi}$ is one-one.)

Example 3.5.11. Recall that $\mathrm{Sin}_n(\mathbb{F})$ is the space of all functions $f : \mathbb{R} \to \mathbb{F}$ of form

$$f(t) = b_1 \sin(t) + b_2 \sin(2t) + \cdots + b_n \sin(nt)$$

for $t \in \mathbb{R}$. Here the coefficients $b_1, b_2, \ldots, b_n$ are arbitrary elements of $\mathbb{F}$. The n functions

$$\phi_k(t) = \sin(kt) \text{ for } k = 1, 2, \ldots, n$$

span $\mathrm{Sin}_n(\mathbb{F})$ by definition. The corresponding map

$$\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathrm{Sin}_n(\mathbb{F})$$

is given by $\mathbf{\Phi}(X) = f$ where

$$X = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

is the column of coefficients. The map $\mathbf{\Phi}$ is onto because the sequence $(\phi_1, \ldots, \phi_n)$ spans $\mathrm{Sin}_n(\mathbb{F})$. The following exercise shows that it is one-one and hence a frame.

Exercise 3.5.12. Show that for $f \in \mathrm{Sin}_n(\mathbb{F})$ and $k = 1, 2, \ldots, n$ we have

$$b_k = \frac{2}{\pi} \int_0^\pi f(t) \sin(kt)\, dt.$$

(Hint: Show

$$\int_0^\pi \sin(mt)\sin(kt)\ dt = 0$$

if $k \neq m$.)

Example 3.5.13. Recall that $\mathrm{Cos}_n(\mathbb{F})$ is the space of all functions $f : \mathbb{R} \to \mathbb{F}$ of form

$$f(t) = a_0 + a_1\cos(t) + a_2\cos(2t) + \cdots + a_n\cos(nt)$$

for $\tau \in \mathbb{R}$. Here the coefficients $a_0, a_1, a_2, \ldots, a_n$ are arbitrary elements of $\mathbb{F}$. The $n+1$ functions

$$\phi_k(t) = \cos(kt) \text{ for } k = 0, 1, 2, \ldots, n$$

span $\mathrm{Cos}_n(\mathbb{F})$ by definition. The corresponding map $\mathbf{\Phi} : \mathbb{F}^{(n+1)\times 1} \to \mathrm{Cos}_n(\mathbb{F})$ is given by $\mathbf{\Phi}(X) = f$ where

$$X = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

is the column of coefficients. The map $\mathbf{\Phi}$ is onto because the sequence $(\phi_0, \ldots, \phi_n)$ spans $\mathrm{Cos}_n(\mathbb{F})$. The following exercise shows that it is one-one and hence a frame.

Exercise 3.5.14. Express each of the coefficient a_k $k = 0, 1, 2, \ldots, n$ of $\cos(kt)$ in $f \in \mathrm{Cos}_n(\mathbb{F})$ in terms of an integral involving f, thus verifying that the correspondence $\mathbf{\Phi}$ is one-one.

Example 3.5.15. Recall that $\mathrm{Trig}_n(\mathbb{F})$ is the space of all functions $f : \mathbb{R} \to \mathbb{F}$ of form

$$f(t) = a_0 + \sum_{k=1}^{n} a_k\cos(kt) + b_k\sin(kt)$$

for $t \in \mathbb{R}$. Here the coefficients $b_n, \ldots, b_2, b_1, a_0, a_1, a_2, \ldots, a_n$ are arbitrary elements of $\mathbb{F}$. The $2n+1$ functions

$$\phi_{-k}(t) = \sin(kt) \text{ for } k = 1, 2, \ldots, n$$

$$\phi_k(t) = \cos(kt) \text{ for } k = 0, 1, 2, \ldots, n$$

span for $\mathrm{Trig}_n(\mathbb{F})$ by definition. The map $\mathbf{\Phi}$ is onto since the sequence $(\phi_{-n}, \ldots, \phi_n)$ spans $\mathrm{Trig}_n(\mathbb{F})$. The following exercise shows that it is one-one and hence a frame.

Exercise 3.5.16. Express each of the coefficients a_k $(k = 0, 1, 2, \ldots, n)$ of $\cos(kt)$ and each of the coefficients b_k $(k = 1, 2, \ldots, n$ of $\sin(kt)$ of $f \in \mathrm{Trig}_n(\mathbb{F})$ in terms of an integral involving f, thus verifying that the correspondence $\mathbf{\Phi}$ is one-one. You will need to verify the following identities:

$$\int_{-\pi}^{\pi} \cos(mt)\sin(kt)\,dt = 0 \text{ for all integers } m, k$$

$$\int_{-\pi}^{\pi} \cos(mt)\cos(kt)\,dt = 0 \text{ for all integers } m \neq k$$

$$\int_{-\pi}^{\pi} \sin(mt)\sin(kt)\,dt = 0 \text{ for all integers } m \neq k$$

Definition 3.5.17. The basis constructed in each of the preceding examples is called the **standard basis** for the corresponding vector space and the corresponding frame is called the **standard frame**. For example, the standard basis for $\mathrm{Poly}_2(\mathbb{F})$ is the sequence (ϕ_0, ϕ_1, ϕ_2) given by $\phi_j(t) = t^j$. Note the discrepancy between the subscript and the place in the sequence: the second element of the sequence is ϕ_1 (not ϕ_2).

3.6 Cardinality

In the next section we shall define the *dimension* of a vector space $\mathbf{V}$. It is the analog of the *cardinality* of a finite set. A set X is **finite** iff for some n there is an invertible map $\phi : \{1, 2, \ldots, n\} \to X$; the number n is therefore the cardinality of the set X. The number n is called the **cardinality** of the finite set X; it is the number of elements in the set X. For an invertible map

$$f : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, m\}$$

we have that $m = n$. *If* $\phi : \{1, 2, \ldots, m\} \to X$ *and* $\psi : \{1, 2, \ldots, m\} \to X$ *are both invertible, then* $\psi^{-1} \circ \phi : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, m\}$ *is also invertible, so* $m = n$. This little argument shows that the cardinality of the set X as defined above is legally defined, that is, that the number n is independent of the choice of ϕ. The definition of *dimension* of a vector space given in the next section proceeds in an analogous fashion.

3.7 The Dimension Theorem

Just as the cardinality of a finite set is the number of its elements, so the dimension of a vector space is the length of a basis for that vector space. To be sure that this is a legal definition we need the

Theorem 3.7.1 (Dimension Theorem). *Let $(\psi_1, \dots, \psi_m)$ be a basis for the vector space* $\mathbf{V}$ *and $(\phi_1, \phi_2, \dots, \phi_n)$ be a sequence of vectors from* $\mathbf{V}$. *Then*

(1) *If $(\phi_1, \phi_2, \dots, \phi_n)$ is independent, then $n \le m$.*

(2) *If $(\phi_1, \phi_2, \dots, \phi_n)$ spans $\mathbf{V}$. then $m \le n$.*

(3) *If $(\phi_1, \phi_2, \dots, \phi_n)$ is a basis for $\mathbf{V}$. then $m = n$.*

Proof. Let $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ correspond to $(\phi_1, \dots, \phi_n)$ and $\mathbf{\Psi} : \mathbb{F}^{m\times 1} \to \mathbf{V}$ correspond to $(\psi_1, \dots, \psi_m)$. Then $\mathbf{\Psi}$ is linear isomorphism so we may form the composition

$$\mathbf{A} = \mathbf{\Psi}^{-1} \circ \mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbb{F}^{m\times 1}.$$

By Theorem 2.2.2 the linear map $\mathbf{A}$ determines a matrix $A \in \mathbb{F}^{m\times n}$ satisfying

$$\mathbf{A}(X) = AX$$

for $X \in \mathbb{F}^{n\times 1}$. Now

(1) $\mathbf{A}$ is one-one iff $\mathbf{\Phi}$ is,

(2) $\mathbf{A}$ is onto iff $\mathbf{\Phi}$ is, and

(3) $\mathbf{A}$ is invertible iff $\mathbf{\Phi}$ is,

so the result follows from the Theorem 1.2.2. QED

Part (3) of the Dimension Theorem says that any two bases for a vector space $\mathbf{V}$ have the same number of elements. This justifies the following

Definition 3.7.2. A vector space $\mathbf{V}$ is **finite dimensional** iff it has a basis $(\psi_1, \psi_2, \dots, \psi_m)$. The number m of vectors in a basis for $\mathbf{V}$ is called the **dimension** of $\mathbf{V}$.

Example 3.7.3. The dimension of $\mathbb{F}^{2\times 2}$ is 4. A basis is given by

$$\phi_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \phi_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \phi_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad \phi_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Question 3.7.4. What is the dimension of $\mathbb{F}^{n\times 1}$? of $\mathbb{F}^{p\times q}$? of $\mathrm{Poly}_n(\mathbb{F})$? of $\mathrm{Trig}_n(\mathbb{F})$? (Answer: $\dim(\mathbb{F}^{n\times 1}) = n$, $\dim(\mathbb{F}^{p\times q}) = pq$, $\dim(\mathrm{Poly}_n(\mathbb{F})) = n+1$, $\dim(\mathrm{Trig}_n(\mathbb{F})) = 2n+1$.)

Parts (1) and (2) of the Dimension Theorem may be phrased as follows: Suppose that the vector space $\mathbf{V}$ has dimension m. Then any independent sequence of vectors from $\mathbf{V}$ has length $\leq m$ and any sequence which spans $\mathbf{V}$ has length $\geq m$. Hence

Corollary 3.7.5. *Suppose that* $\mathbf{V}$ *has dimension* n *and that* $(\phi_1, \phi_2, \dots, \phi_n)$ *is a sequence of vectors from* $\mathbf{V}$*. Then the following are equivalent:*

(1) *The sequence is independent.*

(2) *The sequence spans* $\mathbf{V}$*.*

(3) *The sequence is a basis for* $\mathbf{V}$*.*

Question 3.7.6. Suppose that $\phi_1, \phi_2 \in \mathbb{F}^{1\times 3}$. Is it true that the sequence (ϕ_1, ϕ_2) is a basis for $\mathbb{F}^{1\times 3}$ if and only if it is independent? (Answer: No. In fact, a sequence of length 2 can never be a basis for a vector space of dimension 3 by the Dimension Theorem. It might however be independent, for example, the first two elements of a basis.

Remark 3.7.7. For a vector space $\mathbf{V}$ the following conditions have the same meaning:

(1) $\mathbf{V}$ has dimension n.

(2) $\mathbf{V}$ has a basis $(\phi_1, \phi_2, \dots, \phi_n)$ of length n.

(3) There is an isomorphism (frame) $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$.

3.8 Isomorphism

Theorem 3.8.1. *If* $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ *is an isomorphism, and if the sequence* $(\phi_1, \phi_2, \dots, \phi_n)$ *is a basis for* $\mathbf{V}$*, then the sequence* $(\mathbf{T}(\phi_1), \mathbf{T}(\phi_2), \dots, \mathbf{T}(\phi_n))$ *is a basis for* $\mathbf{W}$*.*

Proof. In other words the composition $\mathbf{T} \circ \mathbf{\Phi}$ of the isomorphism $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ with the frame $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ corresponding to the basis $(\phi_1, \phi_2, \ldots, \phi_n)$ is a frame $\mathbf{T} \circ \mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{W}$. QED

Corollary 3.8.2. *If two finite dimensional vector spaces are isomorphic, then they have the same dimension.*

Question 3.8.3. Is the converse of this corollary true? (Answer: Yes. If $\mathbf{V}$ and $\mathbf{W}$ both have dimension n, then they are each isomorphic to $\mathbb{F}^{n\times 1}$ and hence to each other.)

Example 3.8.4. The sequence of polynomials $(1, t, t^2, \ldots, t^n)$ forms a basis for the $(n+1)$-dimensional vector space $\mathrm{Poly}_n(\mathbb{F})$ of polynomials of degree $\leq n$. Each number a determines an isomorphism $\mathbf{T}$ from $\mathrm{Poly}_n(\mathbb{F})$ to itself via the formula

$$\mathbf{T}(f)(t) = f(t-a);$$

the inverse isomorphism is defined by

$$\mathbf{T}^{-1}(g)(t) = g(t+a).$$

Hence the sequence of polynomials $(1, t-a, (t-a)^2, \ldots, (t-a)^n)$ forms another basis for $\mathrm{Poly}_n(\mathbb{F})$. A polynomial f may be expressed in terms of this basis using **Taylor's formula**:

$$f(t) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(t-a)^k$$

where $f^{(k)}(a)$ is the k-th derivative of f evaluated at a.

3.9 Extraction

Lemma 3.9.1. *Assume that the sequence* $(\phi_1, \ldots, \phi_k, \phi_{k+1})$ *spans* $\mathbf{V}$ *and that* ϕ_{k+1} *is a linear combination of* $(\phi_1, \ldots, \phi_k)$*:*

$$\phi_{k+1} = a_1\phi_1 + \cdots + a_k\phi_k.$$

Then the shorter sequence $(\phi_1, \ldots, \phi_k)$ *also spans* $\mathbf{V}$*.*

Proof. Choose $\mathbf{v} \in \mathbf{V}$. Then there are constants $b_1, \dots, b_k, b_{k+1}$ such that

$$\mathbf{v} = b_1\phi_1 + b_2\phi_2 + \cdots + b_k\phi_k + b_{k+1}\phi_{k+1}$$

since $(\phi_1, \dots, \phi_k, \phi_{k+1})$ spans $\mathbf{V}$. Into this equation substitute the expression for ϕ_{k+1} to obtain

$$\mathbf{v} = (b_1 + b_{k+1}a_1)\phi_1 + (b_2 + b_{k+1}a_2)\phi_2 + \cdots + (b_k + b_{k+1}a_k)\phi_k$$

showing that $\mathbf{v}$ is a linear combination of $\phi_1, \dots, \phi_k$. Thus $(\phi_1, \dots, \phi_k)$ spans $\mathbf{V}$. QED

Theorem 3.9.2 (Extraction Theorem). *Assume that the sequence*

$$(\phi_1, \phi_2, \dots, \phi_m)$$

spans a vector space $\mathbf{V}$ *of dimension* n. *Then there is a subsequence*

$$(\phi_{i_1}, \phi_{i_2}, \dots, \phi_{i_n})$$

which is a basis for $\mathbf{V}$.

Proof. The sequence $(\phi_1, \phi_2, \dots, \phi_m)$ spans $\mathbf{V}$. If it is not a basis, then there is a relation

$$c_1\phi_1 + c_2\phi_2 + \cdots + c_m\phi_m = 0$$

where not all of the coefficients $c_1, c_2, \dots, c_m$ are zero. Suppose for example that $c_1 \neq 0$. Then we may express ϕ_1 as a linear combination of $\phi_2, \dots, \phi_m$:

$$\phi_1 = -\frac{c_2}{c_1}\phi_2 - \cdots - \frac{c_m}{c_1}\phi_m$$

and so $(\phi_2, \dots, \phi_m)$ also spans $\mathbf{V}$. Repeat this process until you get a sequence which is independent. QED

Corollary 3.9.3. *Let* $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ *be a linear map and* $(\phi_1, \phi_2, \dots, \phi_n)$ *be a basis for* $\mathbf{V}$. *Then there is a subsequence* $(\phi_{i_1}, \phi_{i_2}, \dots, \phi_{i_r})$ *such that* $(\mathbf{T}(\phi_{i_1}), \mathbf{T}(\phi_{i_2}), \dots, \mathbf{T}(\phi_{i_r}))$ *forms a basis for* $\mathcal{R}(\mathbf{T}) \subseteq \mathbf{W}$.

Proof. In order to apply Lemma 3.9.2 we must prove that

$$\mathcal{R}(\mathbf{T}) = \text{Span}(\mathbf{T}(\phi_1), \mathbf{T}(\phi_2), \ldots, \mathbf{T}(\phi_n)).$$

This is seen as follows. Choose $\mathbf{w} \in \mathcal{R}(\mathbf{T})$. Then $\mathbf{w} = \mathbf{T}(\mathbf{v})$ for some $\mathbf{v} \in \mathbf{V}$ by the definition of the range. But then $\mathbf{v} = \sum_j c_j \phi_j$ for some numbers c_j since $(\phi_1, \ldots, \phi_n)$ is a basis for $\mathbf{V}$. Then

$$\mathbf{w} = \mathbf{T}(\mathbf{v}) = \mathbf{T}\left(\sum_{j=1}^{n} c_j \phi_j\right) = \sum_{j=1}^{n} c_j \mathbf{T}(\phi_j) \in \text{Span}(\mathbf{T}(\phi_1), \mathbf{T}(\phi_2), \ldots, \mathbf{T}(\phi_n))$$

as required.

Example 3.9.4. The first, third, and fourth columns of the matrix

$$R = \begin{bmatrix} 1 & c_{11} & 0 & 0 & c_{12} \\ 0 & c_{21} & 1 & 0 & c_{22} \\ 0 & c_{31} & 0 & 1 & c_{32} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

form a basis the range of the map

$$\mathbb{F}^{5\times 1} \to \mathbb{F}^{4\times 1} : X \mapsto RX.$$

3.10 Extension

Lemma 3.10.1. *If the sequence $(\phi_1, \phi_2, \ldots, \phi_k)$ is independent and $\phi_{k+1} \notin$ Span$(\phi_1, \phi_2 \ldots, \phi_k)$, then the longer sequence $(\phi_1, \phi_2 \ldots, \phi_k, \phi_{k+1})$ is independent.*

Proof. If the sequence $(\phi_1, \ldots, \phi_{k+1})$ were not independent there would be a non-trivial relation

$$c_1\phi_1 + c_2\phi_2 + \cdots + c_k\phi_k + c_{k+1}\phi_{k+1} = 0.$$

In this relation we must have $c_{k+1} \neq 0$, since $(\phi_1, \phi_2, \ldots, \phi_k)$ is independent. But then

$$\phi_{k+1} = -\frac{c_1}{c_{k+1}}\phi_1 - \frac{c_2}{c_{k+1}}\phi_2 - \cdots - \frac{c_k}{c_{k+1}}\phi_k,$$

contradicting the hypothesis that ϕ_{k+1} is not in Span$(\phi_1, \phi_2, \ldots, \phi_k)$. QED

Theorem 3.10.2 (Extension Theorem). *Let* $\mathbf{V}$ *be a vector space of dimension* n*. Any independent sequence*

$$(\phi_1, \phi_2, \ldots, \phi_m)$$

of elements of $\mathbf{V}$ *may be extended to a basis*

$$(\phi_1, \phi_2, \ldots, \phi_m, \phi_{m+1}, \phi_{m+2}, \ldots, \phi_n)$$

for $\mathbf{V}$*.*

Proof. The sequence $(\phi_1, \phi_2, \ldots, \phi_m)$ is independent. If it is not a basis for $\mathbf{V}$ then it must fail to span, so there must be an element $\phi_{m+1} \in \mathbf{V}$ which is not in the span of the sequence:

$$\phi_{m+1} \notin \mathrm{Span}(\phi_1, \phi_2, \ldots, \phi_m).$$

We may append ϕ_{m+1} to the sequence and, by the lemma, the result

$$(\phi_1, \phi_2, \ldots, \phi_m, \phi_{m+1})$$

is still independent. Repeat this process until you get a sequence which spans $\mathbf{V}$. The process must terminate within $n - m$ steps by the Dimension Theorem. QED

3.11 One-sided Inverses

A map between sets is one-one if and only if it has a left inverse; it is onto if and only if it has a right inverse. Analogs of these statements hold for linear maps between finite dimensional vector spaces. These analogs say more: namely that there exist *linear inverses*. To prove this we need the following

Lemma 3.11.1. *Let* $(\psi_1, \psi_2, \ldots, \psi_m)$ *be a basis for a vector space* $\mathbf{W}$ *and let* $\mathbf{V}$ *be another vector space. Then for any sequence* $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m)$ *there is a unique linear map* $\mathbf{S} : \mathbf{W} \to \mathbf{V}$ *such that* $\mathbf{S}(\psi_i) = \mathbf{v}_i$ *for* $i = 1, 2, \ldots, m$*.*

Proof. To prove this simply choose $\mathbf{w} \in \mathbf{W}$ and write it as a linear combination of the ψ_i:

$$\mathbf{w} = y_1\psi_1 + y_2\psi_2 + \cdots + y_m\psi_m$$

where $y_i \in \mathbb{F}$. If $\mathbf{S}$ is linear and satisfies $\mathbf{S}(\psi_i) = \mathbf{v}_i$ then applying $\mathbf{S}$ to the equation for $\mathbf{w}$ gives

$$\mathbf{S}(\mathbf{w}) = y_1\mathbf{v}_1 + y_2\mathbf{v}_2 + \cdots + y_m\mathbf{v}_m.$$

This shows the uniqueness of $\mathbf{S}$. To show existence use this formula to define $\mathbf{S}$. The definition is legal since the representation of $\mathbf{w}$ is unique. We leave it to the reader to show that $\mathbf{S}$ defined in this way is linear. QED

Remark 3.11.2. This lemma is a generalization of the concept of the Theorem 3.1.1. It may be restated as follows:

$$\mathcal{L}(\mathbf{W}, \mathbf{V}) \to \mathbf{V}^m : \mathbf{S} \mapsto (\mathbf{S}(\psi_1), \mathbf{S}(\psi_2), \ldots, \mathbf{S}(\psi_m))$$

is a one-one onto correspondence. Here $\mathcal{L}(\mathbf{W}, \mathbf{V})$ *denotes of the set of linear maps of* $\mathbf{W}$ *to* $\mathbf{V}$*, and* $\mathbf{V}^n$ *denotes the set of sequences of elements from the vector space* $\mathbf{V}$*.*

Corollary 3.11.3 (Left Inverse Theorem). *A linear map* $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ *between finite dimensional vector spaces is one-one if and only if it has a linear left inverse* $\mathbf{S} : \mathbf{W} \to \mathbf{V}$*.*

Proof. Assume that $\mathbf{T}$ is one-one. Let $(\phi_1, \phi_2, \ldots, \phi_n)$ be a basis for $\mathbf{V}$ and $\mathbf{\Phi}$ denote the corresponding frame. Then $\mathbf{T} \circ \mathbf{\Phi}$ is one-one, so the sequence $(\mathbf{T}(\phi_1), \mathbf{T}(\phi_2), \ldots, \mathbf{T}(\phi_n))$ is linearly independent. Extend to a basis

$$(\mathbf{T}(\phi_1), \mathbf{T}(\phi_2), \ldots, \mathbf{T}(\phi_n), \psi_{m+1}, \psi_{m+2}, \ldots, \psi_n)$$

for $\mathbf{W}$. Now let $\mathbf{S}$ be any linear map satisfying $\mathbf{S}(\mathbf{T}(\phi_j)) = \phi_i$ for $i = j, 2, \ldots, m$. (If $m > n$, then $\mathbf{S}(\psi_i)$ can be anything: there is more than one left inverse.) QED

Corollary 3.11.4 (Right Inverse Theorem). *A linear map* $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ *between finite dimensional vector spaces is onto if and only if it has a linear right inverse* $\mathbf{S} : \mathbf{W} \to \mathbf{V}$*.*

Proof. Assume that $\mathbf{T}$ is onto, Let $(\phi_1, \phi_2, \ldots, \phi_n)$ be a basis for $\mathbf{V}$ and $\mathbf{\Phi}$ denote the corresponding frame. Then $\mathbf{T} \circ \mathbf{\Phi}$ is onto, so the sequence $(\mathbf{T}(\phi_1), \mathbf{T}(\phi_2), \ldots, \mathbf{T}(\phi_n))$ spans $\mathbf{W}$. Extract a basis

$$(\mathbf{T}(\phi_{j_1}), \mathbf{T}(\phi_{j_2}), \ldots, \mathbf{T}(\phi_{j_m})$$

for $\mathbf{W}$. Then define $\mathbf{S}$ by $\mathbf{S}(\mathbf{T}(\phi_{j_1}) = \phi_{j_1}$

3.12 Independence and Span

The notion of *linear independence* can be defined in terms of the operation

$$(\phi_1, \phi_2, \ldots, \phi_n) \mapsto \mathrm{Span}(\phi_1, \phi_2, \ldots, \phi_n)$$

which assigns to a sequence the space which it spans. This is the content of the next proposition.

Proposition 3.12.1. *The sequence $(\phi_1, \phi_2, \ldots, \phi_n)$ is dependent if and only if some element ϕ_j of the sequence is in the space spanned by the remaining elements:*

$$\phi_j \in \mathrm{Span}(\phi_1, \ldots, \phi_{j-1}, \phi_{j+1}, \ldots, \phi_n).$$

Exercise 3.12.2. Prove this.

Example 3.12.3. Let

$$\phi_1 = \begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix}, \quad \phi_2 = \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix}, \quad \phi_3 = \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}.$$

Then the sequence (ϕ_1, ϕ_2, ϕ_3) is dependent since

$$\phi_1 - 2\phi_2 + \phi_3 = 0$$

and $\phi_1 \in \mathrm{Span}(\phi_2, \phi_3)$ since

$$\phi_1 = 2\phi_2 - \phi_3.$$

3.13 Rank and Nullity

Definition 3.13.1. The **rank** of a linear map is the dimension of its range. The **nullity** of a linear map is the dimension of its null space. The rank (or nullity) of a matrix is the rank (or nullity) of the corresponding matrix map.

Theorem 3.13.2 (Rank Nullity Relation). *The rank and nullity of a linear map*

$$\mathbf{T} : \mathbf{V} \to \mathbf{W}$$

are related by

$$\dim(\mathcal{R}(\mathbf{T})) + \dim(\mathcal{N}(\mathbf{T})) = \dim(\mathbf{V}).$$

Proof. Extend a basis $(\phi_1, \ldots, \phi_k)$ for $\mathcal{N}(\mathbf{T})$ to a basis $(\phi_1, \ldots, \phi_n)$ for $\mathbf{V}$. Then $(\mathbf{T}(\phi_{k+1}), \ldots, \mathbf{T}(\phi_n))$ is a basis for $\mathcal{R}(\mathbf{T})$. QED

3.14 Exercises

Exercise 3.14.1. Let the column vectors $\phi_1, \phi_2, \phi_3 \in \mathbb{F}^{3\times 1}$ be defined by

$$\phi_1 = \begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix}, \quad \phi_2 = \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix}, \quad \phi_3 = \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}$$

and let $\mathbf{\Phi} : \mathbb{F}^{3\times 1} \to \mathbb{F}^{3\times 1}$ be the linear map corresponding to the sequence ϕ_1, ϕ_2, ϕ_3. Find a matrix $A \in \mathbb{F}^{3\times 3}$ such that $\mathbf{\Phi}(X) = AX$ for $X \in \mathbb{F}^{3\times 1}$.

Exercise 3.14.2. Let the row vectors $\phi_1, \phi_2, \phi_3 \in \mathbb{F}^{1\times 3}$ be defined by

$$\phi_1 = \begin{bmatrix} 1 & 4 & 7 \end{bmatrix},$$
$$\phi_2 = \begin{bmatrix} 2 & 5 & 8 \end{bmatrix},$$
$$\phi_3 = \begin{bmatrix} 3 & 6 & 9 \end{bmatrix}$$

and let $\mathbf{\Phi} : \mathbb{F}^{3\times 1} \to \mathbb{F}^{1\times 3}$ be the linear map corresponding to the sequence (ϕ_1, ϕ_2, ϕ_3). Find a matrix $A \in \mathbb{F}^{3\times 3}$ such that $\mathbf{\Phi}(X) = X^*A$ for $X \in \mathbb{F}^{3\times 1}$ where X^* is the transpose of X.

Exercise 3.14.3. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 3 & 3 & 3 \end{bmatrix}$. Show that the columns of A are dependent by finding x_1, x_2, x_3, not all zero, such that

$$x_1\mathrm{col}_1(A) + x_2\mathrm{col}_2(A) + x_3\mathrm{col}_3(A) = 0.$$

Exercise 3.14.4. Let A be as in the previous problem. Show that the rows of A are dependent by finding x_1, x_2, x_3, not all zero, such that

$$x_1\mathrm{row}_1(A) + x_2\mathrm{row}_2(A) + x_3\mathrm{row}_3(A) = 0.$$

Exercise 3.14.5. Are there numbers x_1, x_2, x_3 (not all zero) which simultaneously solve both of the previous two problems?

Exercise 3.14.6. Let $\phi_1, \phi_2, \phi_3 \in \mathrm{Poly}_2(\mathbb{F})$ be given by

$$\begin{aligned} \phi_1(t) &= 1 + 2t + 3t^2, \\ \phi_2(t) &= 4 + 5t + 6t^2, \\ \phi_3(t) &= 3 + 3t + 3t^2. \end{aligned}$$

Show that ϕ_1, ϕ_2, ϕ_3 are dependent. Which of the previous problems is this most like?

Exercise 3.14.7. Let $W_1, W_2 \in \mathbb{F}^{2\times 1}$. When is the sequence (W_1, W_2) independent?

Exercise 3.14.8. When does $\mathrm{Span}(W_1, W_2) = \mathbb{F}^{2\times 1}$?

Exercise 3.14.9. When does $\mathrm{Span}(W_1, W_2, W_3) = \mathbb{F}^{2\times 1}$?

Exercise 3.14.10. Let $W_1, W_2, W_3 \in \mathbb{F}^{2\times 1}$. When is (W_1, W_2, W_3) independent?

Exercise 3.14.11. Let $\phi_1, \phi_2, \phi_3 \in \mathrm{Cos}_2(\mathbb{F})$ be given by

$$\begin{aligned}
\phi_1(t) &= 1 + 2\cos(t) + 3\cos(2t), \\
\phi_2(t) &= 4 + 5\cos(t) + 6\cos(2t), \\
\phi_3(t) &= 3 + 3\cos(t) + 3\cos(2t).
\end{aligned}$$

Show that ϕ_1, ϕ_2, ϕ_3 are dependent. Which of the previous problems is this most like?

Exercise 3.14.12. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 3 & 3 & 3 \end{bmatrix}$. Show that the columns of A do not span $\mathbb{F}^{3\times 1}$ by finding $Y \in \mathbb{F}^{3\times 1}$, such that the inhomogeneous system

$$Y = x_1 \mathrm{col}_1(A) + x_2 \mathrm{col}_2(A) + x_3 \mathrm{col}_3(A)$$

has no solution x_1, x_2, x_3.

Exercise 3.14.13. Let A be as in the previous problem. Show that the rows of A do not span $\mathbb{F}^{1\times 3}$ by finding $K \in \mathbb{F}^{1\times 3}$, such that the inhomogeneous system

$$K = x_1 \mathrm{row}_1(A) + x_2 \mathrm{row}_2(A) + x_3 \mathrm{row}_3(A)$$

has no solution x_1, x_2, x_3.

Exercise 3.14.14. Let $\phi_1, \phi_2, \phi_3 \in \mathrm{Poly}_2(\mathbb{F})$ be given by

$$\begin{aligned}
\phi_1(t) &= 1 + 2t + 3t^2, \\
\phi_2(t) &= 4 + 5t + 6t^2, \\
\phi_3(t) &= 7 + 8t + 9t^2.
\end{aligned}$$

Show that ϕ_1, ϕ_2, ϕ_3 do not span $\mathrm{Poly}_2(\mathbb{F})$ by exhibiting a polynomial

$$f(t) = a_0 + a_1 t + a_2 t^2$$

which can not be written in the form

$$f(t) = x_1\phi_1(t) + x_2\phi_2(t) + x_3\phi_3(t).$$

Which of the previous problems is this most like?

Exercise 3.14.15. Verify that

$$f(t) = f(a) + f'(a)(t-a) + \frac{f''(a)}{2}(t-a)^2 + \frac{f'''(a)}{6}(t-a)^3$$

for $f(t) = c_0 + c_1 t + c_2 t^2 + c_3 t^3$.

Exercise 3.14.16. Let $D \in \mathbb{F}^{m\times n}$ be of form:

$$D = \begin{bmatrix} I_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix}$$

where I_r is the $r\times r$ identity matrix. When are the columns of D independent? When do they span $\mathbb{F}^{m\times 1}$?

Exercise 3.14.17. Let $R_j = \mathrm{col}_j(R)$ be the j-th column of the matrix

$$R = \begin{bmatrix} 1 & c_{11} & 0 & 0 & c_{12} \\ 0 & c_{21} & 1 & 0 & c_{22} \\ 0 & c_{31} & 0 & 1 & c_{32} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and $A = QR$ where Q is invertible. Let $A_j = \mathrm{col}_j(A)$ be the j-th column of A. Show that (A_1, A_3, A_4) is a basis for $\mathrm{Span}(A_1, A_2, A_3, A_4, A_5)$.

Exercise 3.14.18 (Lagrange Interpolation). Let $\lambda_0, \ldots, \lambda_n$ be distinct numbers and $(\phi_0, \ldots, \phi_n)$ be the sequence of polynomials given by

$$\phi_k(t) = \frac{\prod_{j\neq k}(t-\lambda_j)}{\prod_{j\neq k}(\lambda_k-\lambda_j)}.$$

Show that this sequence is a basis for $\mathrm{Poly}_n(\mathbb{F})$. Given $b_0, b_1, b_2, \ldots, b_n$ there is a unique polynomial $f \in \mathrm{Poly}_n(\mathbb{F})$ such that

$$f(\lambda_j) = b_j, \quad \text{for } j = 0, 1, 2, \ldots, n.$$

Express f as a linear combination of $\phi_0, \phi_1, \ldots, \phi_n$. Hint: What is $\phi_k(\lambda_i)$?

Exercise 3.14.19 (Transitivity Lemma). Suppose $\mathbf{V}$ is a vector space and that $\phi_1, \phi_2, \dots, \phi_n$, $\psi_1, \psi_2, \dots, \psi_m$, and $\mathbf{v}$ are elements of $\mathbf{V}$. Assume

$$\psi_i \in \mathrm{Span}(\phi_1, \phi_2, \dots, \phi_n)$$

for $i = 1, 2, \dots, m$ and

$$\mathbf{v} \in \mathrm{Span}(\psi_1, \psi_2, \dots, \psi_m)$$

Show that

$$\mathbf{v} \in \mathrm{Span}(\phi_1, \phi_2, \dots, \phi_n).$$

Exercise 3.14.20. Assume

$$\phi_{m+j} \in \mathrm{Span}(\phi_1, \phi_2, \dots, \phi_m)$$

for $j = 1, 2, \dots, n - m$. Show that

$$\mathrm{Span}(\phi_1, \phi_2, \dots, \phi_m) = \mathrm{Span}(\phi_1, \phi_2, \dots, \phi_n).$$

Exercise 3.14.21. For $j = 1, 2, 3, 4, 5$, let $R_j = \mathrm{col}_j(R)$ be the j-th column of the matrix

$$R = \begin{bmatrix} 1 & c_{12} & 0 & c_{14} & c_{15} \\ 0 & c_{22} & 1 & c_{24} & c_{25} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Show that $\mathrm{Span}(R_1, R_3) = \mathrm{Span}(R_1, R_2, R_3, R_4, R_5)$.

Exercise 3.14.22. Prove that if $\sigma : \{1, 2, \dots, n\} \to \{1, 2, \dots, n\}$ is a permutation of $1, 2, \dots, n$, then

$$\mathrm{Span}(\phi_1, \phi_2, \dots, \phi_n) = \mathrm{Span}(\phi_{\sigma(1)}, \phi_{\sigma(2)}, \dots, \phi_{\sigma(n)}).$$

Exercise 3.14.23. Let

$$B_1 = \begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix}.$$

Extend the sequence (B_1, B_2) to a basis (B_1, B_2, B_3) for $F^{3\times 1}$

Chapter 4

Matrix Representation

A matrix $A \in \mathbb{F}^{m\times n}$ determines a matrix map $\mathbf{A} : \mathbb{F}^{n\times 1} \to \mathbb{F}^{m\times 1}$ (see Theorem 2.2.2) and the isomorphism

$$\mathbb{F}^{m\times n} \to \mathcal{L}(\mathbb{F}^{n\times 1}, \mathbb{F}^{m\times 1}) : A \mapsto \mathbf{A}$$

(see Corollary 2.3.3) says that a matrix and a linear map from $\mathbb{F}^{n\times 1}$ to $\mathbb{F}^{m\times 1}$ are essentially the same thing. We have seen (Theorem 3.4.2) that a frame $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ and a basis for the vector space $\mathbf{V}$ are essentially the same thing and that the map

$$\mathbb{F}^{m\times n} \to \mathcal{L}(\mathbf{V}, \mathbf{W}) : A \mapsto \mathbf{\Psi}\mathbf{A} \circ \mathbf{\Phi}^{-1}$$

determined by two frames $\mathbf{\Phi}$ and $\mathbf{\Psi}$ is an isomorphism. In this chapter we see how this isomorphism relates the vector space theory to matrix theory.

4.1 The Representation Theorem

Assume

(1) $\mathbf{V}$ is a finite dimensional vector space of dimension n.

(2) $\mathbf{W}$ is a finite dimensional vector space of dimension m.

(3) $I_{n,j} = \mathrm{col}_j(I_n)$ is the j-th column of the $n \times n$ identity matrix.

(4) $I_{m,i} = \mathrm{col}_i(I_m)$ is the i-th column of the $m \times m$ identity matrix.

(5) $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ is a frame for $\mathbf{V}$.

(6) $(\phi_1, \phi_2, \dots, \phi_n)$ is the basis corresponding to the frame $\mathbf{\Phi}$. Thus $\phi_j = \mathbf{\Phi}(I_{n,j})$ for $j = 1, 2, \dots, n$.

(7) $\mathbf{\Psi} : \mathbb{F}^{m\times 1} \to \mathbf{W}$ is a frame for $\mathbf{W}$.

(8) $(\psi_1, \psi_2, \dots, \psi_m)$ is the basis corresponding to the frame $\mathbf{\Psi}$. Thus $\psi_i = \mathbf{\Psi}(I_{m,i})$ for $i = 1, 2, \dots, m$.

Proposition 4.1.1 (Representation Theorem). *Let $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ be a linear map. The matrix A representing the map $\mathbf{T}$ in the frames $\mathbf{\Phi}$ and $\mathbf{\Psi}$ is characterized by the equations*

$$\mathbf{T}(\phi_j) = \sum_{i=1}^{m} a_{ij}\psi_i \tag{3}$$

for $j = 1, 2, \dots, n$. Here ϕ_j is the j-th element of the basis corresponding to the frame $\mathbf{\Phi}$, ψ_i is the i-th element of the basis corresponding to the frame $\mathbf{\Psi}$, and $a_{ij} = \mathrm{entry}_{ij}(A)$.

Proof. The equation

$$AI_{n,j} = \sum_{i=1}^{m} a_{ij} I_{m,i} \tag{3$'$}$$

is analogous to equation (3); it says that $AI_{n,j} = \mathrm{col}_j(A)$. Note also that

$$\phi_j = \mathbf{\Phi}(I_{n,j}), \quad \psi_i = \mathbf{\Psi}(I_{m,i}).$$

The matrix A characterized by the equation

$$\mathbf{T}(\mathbf{\Phi}(X)) = \mathbf{\Psi}(AX) \tag{4}$$

for $X \in \mathbb{F}^{n\times 1}$. (Equation (4) is obtained by rewriting equation (1) as $\mathbf{T} \circ \mathbf{\Phi} = \mathbf{\Psi} \circ \mathbf{A}$ and evaluating at X.) Now take $X = I_{n,j}$ in equation (4) to obtain

$$\begin{aligned} \mathbf{T}(\phi_j) &= \mathbf{\Psi}(AI_{n,j}) \\ &= \mathbf{\Psi}\left(\sum_{i=1}^{m} a_{ij} I_{m,i}\right) \\ &= \sum_{i=1}^{m} a_{ij}\mathbf{\Psi}(I_{m,i})) \\ &= \sum_{i=1}^{m} a_{ij}\psi_i \end{aligned}$$

as required. QED

Remark 4.1.2. When $\mathbf{V} = \mathbf{W}$ and $\boldsymbol{\Psi} = \boldsymbol{\Phi}$ the matrix A representing the map $\mathbf{T}$ in the frame $\boldsymbol{\Phi}$ is characterized by the equations

$$\mathbf{T}(\phi_j) = \sum_{i=1}^{m} a_{ij}\phi_i$$

for $j = 1, 2, \ldots, n$ where $a_{ij} = \text{entry}_{ij}(A)$.

Example 4.1.3. We take

$$\mathbf{V} = \text{Poly}_3(\mathbb{F}), \quad \mathbf{W} = \mathbb{F}^{1\times 3},$$

define $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ by

$$\mathbf{T}(f) = \begin{bmatrix} f(1) & f(-1) & f'(0) \end{bmatrix}.$$

Let the frame $\boldsymbol{\Phi} : \mathbb{F}^{4\times 1} \to \mathbf{V}$ be the standard frame given by

$$\phi_1(t) = 1, \quad \phi_2(t) = t, \quad \phi_3(t) = t^2, \quad \phi_4(t) = t^3,$$

and the frame $\boldsymbol{\Psi} : \mathbb{F}^{3\times 1} \to \mathbb{F}^{1\times 3}$ be defined by $\boldsymbol{\Psi}(Y) = Y^*$ so that

$$\begin{aligned} \psi_1 &= \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \\ \psi_2 &= \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \\ \psi_3 &= \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

We find the first column of A:

$$\begin{aligned} \mathbf{T}(\phi_1) &= \begin{bmatrix} \phi_1(1) & \phi_1(-1) & \phi_1'(0) \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \\ &= 1\psi_1 + 1\psi_2 + 0\psi_3. \end{aligned}$$

We find the second column of A:

$$\begin{aligned} \mathbf{T}(\phi_2) &= \begin{bmatrix} \phi_2(1) & \phi_2(-1) & \phi_2'(0) \end{bmatrix} \\ &= \begin{bmatrix} 1 & -1 & 1 \end{bmatrix} \\ &= 1\psi_1 - 1\psi_2 + 1\psi_3. \end{aligned}$$

We find the third column of A:

$$\begin{aligned} \mathbf{T}(\phi_3) &= \begin{bmatrix} \phi_3(1) & \phi_3(-1) & \phi_3'(0) \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & 2 \end{bmatrix} \\ &= 1\psi_1 + 1\psi_2 + 2\psi_3. \end{aligned}$$

We find the fourth column of A:

$$\begin{aligned} \mathbf{T}(\phi_4) &= \begin{bmatrix} \phi_4(1) & \phi_4(-1) & \phi_4'(0) \end{bmatrix} \\ &= \begin{bmatrix} 1 & -1 & 3 \end{bmatrix} \\ &= 1\psi_1 - 1\psi_2 + 3\psi_3. \end{aligned}$$

Thus A is given by

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

This example required very little calculation because of the simple nature of the frame $\mathbf{\Psi}$. In general we will have to solve an inhomogeneous linear system of m equations in m unknowns to find the j-th column of A. As we must solve such a system for each value of $j = 1, 2, \ldots, n$ this can lead to quite a bit of work. The next example requires us to invert an $m \times m$ matrix to find A. It still isn't too bad since we take $m = 2$.

Example 4.1.4. We take

$$\mathbf{V} = \mathrm{Poly}_3(\mathbb{F}), \quad \mathbf{W} = \mathbb{F}^{1\times 2},$$

define $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ by

$$\mathbf{T}(f) = \begin{bmatrix} f(1) & f(2) \end{bmatrix}.$$

Let the frame $\mathbf{\Phi} : \mathbb{F}^{4\times 1} \to \mathbf{V}$ be the standard frame given by

$$\phi_1(t) = 1, \quad \phi_2(t) = t, \quad \phi_3(t) = t^2, \quad \phi_4(t) = t^3,$$

and the frame $\mathbf{\Psi} : \mathbb{F}^{2\times 1} \to \mathbb{F}^{1\times 2}$ be defined by

$$\begin{aligned} \psi_1 &= \begin{bmatrix} 7 & 3 \end{bmatrix} \\ \psi_2 &= \begin{bmatrix} 2 & 1 \end{bmatrix} \end{aligned}$$

We find the first column of A:

$$\begin{aligned}\mathbf{T}(\phi_1) &= \begin{bmatrix} \phi_1(1) & \phi_1(2) \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 \end{bmatrix} \\ &= a_{11} \begin{bmatrix} 7 & 3 \end{bmatrix} + a_{21} \begin{bmatrix} 2 & 1 \end{bmatrix}.\end{aligned}$$

This leads to the 2×2 system

$$\begin{aligned} 1 &= 7a_{11} + 2a_{21} \\ 1 &= 3a_{11} + 1a_{21} \end{aligned}$$

which has the solution $a_{11} = -1$, $a_{21} = 4$. We repeat this for columns two, three, and four to obtain

$$A = \begin{bmatrix} -1 & -3 & -7 & -15 \\ 4 & 11 & 25 & 52 \end{bmatrix}.$$

4.2 The Transition Matrix

Let $(\phi_1, \phi_2, \ldots, \phi_n)$ be a basis for a vector space $\mathbf{V}$ and $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ be the corresponding frame. Let $(\tilde{\phi}_1, \tilde{\phi}_2, \ldots, \tilde{\phi}_n)$ be another basis for $\mathbf{V}$ with corresponding frame $\widetilde{\mathbf{\Phi}} : \mathbb{F}^{n\times 1} \to \mathbf{V}$. Then the composition

$$\widetilde{\mathbf{\Phi}}^{-1} \circ \mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbb{F}^{n\times 1}$$

is a linear isomorphism from $\mathbb{F}^{n\times 1}$ to itself and is thus given by a an invertible matrix P:

$$\widetilde{\mathbf{\Phi}}^{-1}(\mathbf{\Phi}(X)) = PX$$

for $X \in \mathbb{F}^{n\times 1}$.

Definition 4.2.1. This matrix P is called the **transition matrix** from the basis $(\phi_1, \phi_2, \ldots, \phi_n)$ to the basis $(\tilde{\phi}_1, \tilde{\phi}_2, \ldots, \tilde{\phi}_n)$. (One also calls P the transition matrix from the frame $\mathbf{\Phi}$ to the frame $\widetilde{\mathbf{\Phi}}$.)

Remark 4.2.2. Note that P is the matrix representing the identity transformation in the frames $\mathbf{\Phi}$ and $\widetilde{\mathbf{\Phi}}$, but it is less confusing to have a separate name in this context since it plays a different role.

The equation defining P may be written in the form

$$\widetilde{\mathbf{\Phi}}(PX) = \mathbf{\Phi}(X).$$

If we plug in $X = \mathrm{col}_j(I_n)$ the j-the column of the identity matrix we obtain

$$\sum_{i=1}^{n} p_{ij}\tilde{\phi}_i = \phi_j$$

where $p_{ij} = \mathrm{entry}_{ij}(P)$ the (i,j)-entry of P. Thus the matrix P enables us to express the vectors ϕ_j as a linear combination of the vectors $\tilde{\phi}_i$, $i = 1, 2, \ldots, n$. On the other hand suppose that $\mathbf{v} \in \mathbf{V}$. Then $\mathbf{v} = \mathbf{\Phi}(X)$ for some $X \in \mathbb{F}^{n\times 1}$ and $\mathbf{v} = \widetilde{\mathbf{\Phi}}(\widetilde{X})$ for some $\widetilde{X}$:

$$\mathbf{v} = \sum_{i=1}^{n} x_i\phi_i = \sum_{i=1}^{n} \tilde{x}_i\tilde{\phi}_i.$$

Since $\mathbf{\Phi}(X) = \widetilde{\mathbf{\Phi}}(\widetilde{X})$ we have $\widetilde{X} = PX$ so that P transforms the column vector X which represents $\mathbf{v}$ in the frame $\mathbf{\Phi}$ to the column vector $\widetilde{X}$ which represents the same vector $\mathbf{v}$ in the frame $\widetilde{\mathbf{\Phi}}$.

Example 4.2.3. Here is a basis for $\mathrm{Poly}_2(\mathbb{F})$:

$$\phi_1(t) = 1, \quad \phi_2(t) = t, \quad \phi_3(t) = t^2,$$

and here is another basis:

$$\tilde{\phi}_1(t) = 1, \quad \tilde{\phi}_2(xy) = t + 1, \quad \tilde{\phi}_3(t) = (t+1)^2.$$

We find the transition matrix P from the first basis to the second. The columns of P are given by

$$\mathrm{col}_j(P) = \widetilde{\mathbf{\Phi}}^{-1}(\mathbf{\Phi}(I_{n,j}))$$

for $j = 1, 2, 3$, where $I_{n,j} = \mathrm{col}_j(I_3)$ is the j-th column of the identity matrix. We apply $\widetilde{\mathbf{\Phi}}_j$ to both sides and use the formula $\mathbf{\Phi}(I_{n,j}) = \phi_j$ to rewrite this in the form

$$p_{1j}p\tilde{h}i_1 + p_{2j}\tilde{\phi}_2 + p_{3j}\tilde{\phi}_3 = \phi_j$$

or

$$p_{1j}1 + p_{2j}(t+1) + p_{3j}(t+1)^2 = t^{j-1}$$

where $p_{ij} = \text{entry}_{ij}(P)$. For each $j = 1, 2, 3$ we must thus solve three equations in three unknowns. By equating coefficients of t^0, t^1, t^2 we get

$$\begin{array}{lll} p_{11} = 1, & p_{12} = 1, & p_{13} = 1, \\ p_{21} = 0, & p_{22} = 1, & p_{23} = -2, \\ p_{31} = 0, & p_{32} = 0, & p_{33} = 1. \end{array}$$

Question 4.2.4. Let (ϕ_1, ϕ_2, ϕ_3) and $(\tilde{\phi}_1, \tilde{\phi}_2, \tilde{\phi}_3)$ be bases for a vector space $\mathbf{V}$ and $P \in \mathbb{F}^{3\times 3}$ be the transition matrix from the former to the latter. Suppose that a matrix $B \in \mathbb{F}^{3\times 3}$ is defined by $\text{entry}_{ij}(B) = b_{ij}$ where

$$\begin{aligned} \tilde{\phi}_1 &= b_{11}\phi_1 + b_{12}\phi_2 + b_{13}\phi_3 \\ \tilde{\phi}_2 &= b_{21}\phi_1 + b_{22}\phi_2 + b_{23}\phi_3 \\ \tilde{\phi}_3 &= b_{31}\phi_1 + b_{32}\phi_2 + b_{33}\phi_3. \end{aligned}$$

Which of the following is necessarily true?

(1) B is P.

(2) B is the transpose of P.

(3) B is P^{-1}.

(4) B is the transpose of P^{-1}.

(Answer: (4).)

4.3 Change of Frames

Let $\mathbf{T} : \mathbf{V} \to \mathbf{W}$, $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$, $\mathbf{\Psi} : \mathbb{F}^{m\times 1} \to \mathbf{W}$, be as in Section 4.1 and let $A \in \mathbb{F}^{m\times n}$ be the matrix representing the map $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ in the frames $\mathbf{\Phi}$ and $\mathbf{\Psi}$, and $\mathbf{A} : \mathbb{F}^{n\times 1} \to \mathbb{F}^{m\times 1}$ be the matrix map corresponding to A.

Proposition 4.3.1. *Changing frames has the effect of replacing the matrix A representing $\mathbf{T}$ by an equivalent matrix $\widetilde{A}$. More precisely, for $\widetilde{A} \in \mathbb{F}^{m\times n}$ the following conditions are equivalent:*

(1) *There are frames $\widetilde{\mathbf{\Phi}} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ and $\widetilde{\mathbf{\Psi}} : \mathbb{F}^{m\times 1} \to \mathbf{W}$ so that $\widetilde{A}$ is the matrix representing $\mathbf{T}$ in the frames $\widetilde{\mathbf{\Phi}}$ and $\widetilde{\mathbf{\Psi}}$.*

(2) *The matrices A and $\widetilde{A}$ are equivalent in the sense that there are invertible matrices $P \in \mathbb{F}^{n\times n}$ and $Q \in \mathbb{F}^{m\times m}$ such that*

$$\widetilde{A} = QAP^{-1}.$$

Proof. Assume (1). Let $\widetilde{\mathbf{A}}$ be the matrix map corresponding to $\widetilde{A}$:

$$\widetilde{\mathbf{A}} = \widetilde{\mathbf{\Psi}}^{-1} \circ \mathbf{T} \circ \widetilde{\mathbf{\Phi}}.$$

Then

$$\widetilde{\mathbf{\Psi}} \circ \widetilde{\mathbf{A}} \circ \widetilde{\mathbf{\Phi}}^{-1} = \mathbf{T} = \mathbf{\Psi} \circ \mathbf{A} \circ \mathbf{\Phi}^{-1}$$

so

$$\widetilde{\mathbf{A}} = \mathbf{Q} \circ \mathbf{A} \circ \mathbf{P}^{-1} \tag{5}$$

where $\mathbf{Q} : \mathbb{F}^{m\times 1} \to \mathbb{F}^{m\times 1}$ and $\mathbf{P} : \mathbb{F}^{n\times 1} \to \mathbb{F}^{n\times 1}$ are the transition matrices given by

$$\mathbf{Q} = \widetilde{\mathbf{\Psi}}^{-1} \circ \mathbf{\Psi}, \quad \mathbf{P} = \widetilde{\mathbf{\Phi}}^{-1} \circ \mathbf{\Phi}.$$

Then $\mathbf{Q}$ is a matrix map corresponding to a matrix $Q \in \mathbb{F}^{m\times m}$ and $\mathbf{P}$ is a matrix map corresponding to a matrix $P \in \mathbb{F}^{n\times n}$. Equation (5) implies $\widetilde{A} = QAP^{-1}$.

Assume (2). Define frames $\widetilde{\mathbf{\Psi}}$ and $\widetilde{\mathbf{\Phi}}$ by

$$\widetilde{\mathbf{\Psi}} = \mathbf{\Psi} \circ \mathbf{Q}^{-1}, \quad \widetilde{\mathbf{\Phi}} = \mathbf{\Phi} \circ \mathbf{P}^{-1}.$$

Then

$$\begin{aligned} \widetilde{\mathbf{A}} &= \mathbf{Q} \circ \mathbf{A} \circ \mathbf{P}^{-1} \\ &= (\widetilde{\mathbf{\Psi}}^{-1} \circ \mathbf{\Psi}) \circ \mathbf{A} \circ (\widetilde{\mathbf{\Phi}}^{-1} \circ \mathbf{\Phi})^{-1} \\ &= \widetilde{\mathbf{\Psi}}^{-1} \circ (\mathbf{\Psi} \circ \mathbf{A} \circ \mathbf{\Phi}^{-1}) \circ \widetilde{\mathbf{\Phi}}^{-1} \\ &= \widetilde{\mathbf{\Psi}}^{-1} \circ \mathbf{T} \circ \widetilde{\mathbf{\Phi}} \end{aligned}$$

which proves (1). QED

Corollary 4.3.2. *Changing the frame $\mathbf{\Psi}$ at the target has the effect of replacing the matrix A representing $\mathbf{T}$ by a left equivalent matrix $\widetilde{A}$. More precisely, for $\widetilde{A} \in \mathbb{F}^{m\times n}$ the following conditions are equivalent:*

(1) *There is a frame $\widetilde{\mathbf{\Psi}} : \mathbb{F}^{m\times 1} \to \mathbf{W}$ so that $\widetilde{A}$ is the matrix representing $\mathbf{T}$ in the frames $\mathbf{\Phi}$ and $\widetilde{\mathbf{\Psi}}$.*

(2) *The matrices A and $\widetilde{A}$ are equivalent in the sense that there is an invertible matrix $Q \in \mathbb{F}^{m\times m}$ such that $\widetilde{A} = QA$.*

Proof. Take $\mathbf{\Phi} = \widetilde{\mathbf{\Phi}}$ in Theorem 4.3.1 so that $P = I_n$ is the identity matrix. QED

Corollary 4.3.3. *Changing the frame $\mathbf{\Phi}$ has the effect of replacing the matrix A representing $\mathbf{T}$ by a right equivalent matrix $\widetilde{A}$. More precisely, for $\widetilde{A} \in \mathbb{F}^{m\times n}$ the following conditions are equivalent:*

(1) *There is a frame $\widetilde{\mathbf{\Phi}} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ so that $\widetilde{A}$ is the matrix representing $\mathbf{T}$ in the frames $\widetilde{\mathbf{\Phi}}$ and $\mathbf{\Psi}$.*

(2) *The matrices A and $\widetilde{A}$ are right equivalent in the sense that there is an invertible matrix $P \in \mathbb{F}^{n\times n}$ such that $\widetilde{A} = AP^{-1}$.*

Proof. Take $\mathbf{\Psi} = \widetilde{\mathbf{\Psi}}$ in Theorem 4.3.1 so that $Q = I_m$ is the identity matrix. QED

Corollary 4.3.4 (Similarity). *Now assume that $\mathbf{V} = \mathbf{W}$ so that $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ is a linear map from a vector space to itself. Let $\mathbf{\Phi} : \mathbb{F}^{n\times 1}$ be a frame for $\mathbf{V}$. Then changing frames has the effect of replacing the matrix representing $\mathbf{T}$ by a similar matrix. More precisely, for $\widetilde{A} \in \mathbb{F}^{n\times n}$ the following conditions are equivalent:*

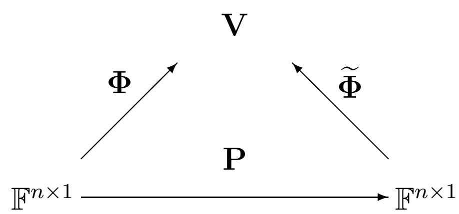
(1) *There is a frame $\widetilde{\mathbf{\Phi}} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ such that $\widetilde{A}$ is the matrix representing $\mathbf{T}$ in the frame $\widetilde{\mathbf{\Phi}}$.*

(2) *The matrices A and $\widetilde{A}$ are* **similar**, *i.e. there is an invertible matrix $P \in \mathbb{F}^{n\times n}$ such that*

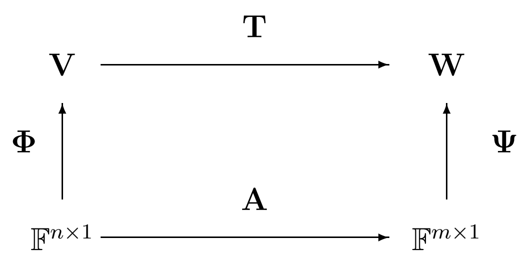
$$\widetilde{A} = PAP^{-1}.$$

Proof. Take $\mathbf{\Psi} = \mathbf{\Phi}$ and $\widetilde{\mathbf{\Psi}} = \widetilde{\mathbf{\Phi}}$ in Theorem 4.3.1 so that $Q = P$. QED

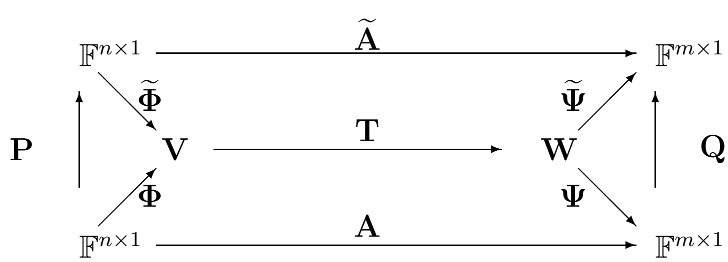
Diagrams can be useful for remembering formulas. The formula $\widetilde{\mathbf{\Phi}}\circ\mathbf{P} = \mathbf{\Phi}$ which says that P is the transition matrix from $\mathbf{\Phi}$ to $\widetilde{\mathbf{\Phi}}$ can be represented by the triangle:



The formula $\mathbf{\Psi}\circ\mathbf{A} = \mathbf{T}\circ\mathbf{\Phi}$ which says that A is the matrix representing $\mathbf{T}$ in the frames $\mathbf{\Phi}$ and $\mathbf{\Psi}$ can be represented by the rectangle:



The Change of Frames Theorem is represented by the following diagram:



4.4 Flags

The following terminology will be used in the next section.

Definition 4.4.1. A **flag** in a vector space $\mathbf{V}$ is an increasing sequence of subspaces

$$\{0\} = \mathbf{V}_0 \subseteq \mathbf{V}_1 \subseteq \mathbf{V}_2 \subseteq \cdots \subseteq \mathbf{V}_n = \mathbf{V}$$

where $\dim(\mathbf{V}_j) = j$. The **standard flag**

$$\{0\} = E_{n,0} \subseteq E_{n,1} \subseteq E_{n,2} \subseteq \cdots \subseteq E_{n,n} = \mathbf{V}$$

in $\mathbb{F}^{n\times 1}$ is defined by

$$E_{n,k} = \mathrm{Span}(I_{n,1}, I_{n,2}, \ldots, I_{n,k})$$

where $I_{n,j} = \mathrm{col}_j(I_n)$ is the j-th column of the $n \times n$ identity matrix. For example,

$$E_{3,2} = \mathrm{Span}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right) = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} \in \mathbb{F}^{3\times 1} : x_1, x_2 \in \mathbb{F} \right\}.$$

Now any basis $(\phi_1, \phi_2, \ldots, \phi_n)$ for a vector space $\mathbf{V}$ determines a flag by

$$\mathbf{V}_k = \mathrm{Span}(\phi_1, \phi_2, \ldots, \phi_k).$$

We call this the **flag determined by** the basis. (Thus the standard basis for $\mathbb{F}^{n\times 1}$ is determines the standard flag.) If $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ is the frame corresponding to the basis $(\phi_1, \phi_2, \ldots, \phi_n)$ we also say that the flag is determined by the frame. Note that

$$\mathbf{\Phi}(E_{n,k}) = \mathbf{V}_k.$$

Different bases can determine the same flag. For example, if we replace each ϕ_j by a non-zero multiple of itself we do not change $\mathbf{V}_k$. Our next task is to determine when two different bases determine the same flag.

Proposition 4.4.2. *Two bases determine the same flag if and only if the transition matrix P from one to the other preserves the standard flag i.e. if and only if*

$$PE_{n,k} = E_{n,k}$$

for $k = 1, 2, \ldots, n$.

Proof. Let $\mathbf{\Phi}$ and $\widetilde{\mathbf{\Phi}}$ be two frames for $\mathbf{V}$ which determine the same flag and let $P \in \mathbb{F}^{n\times n}$ be the transition matrix from $\mathbf{\Phi}$ to $\widetilde{\mathbf{\Phi}}$. Thus

$$\widetilde{\mathbf{\Phi}}^{-1} \circ \mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbb{F}^{n\times 1}$$

and

$$\widetilde{\mathbf{\Phi}}^{-1}(\mathbf{\Phi}(X)) = PX$$

for $X \in \mathbb{F}^{n\times 1}$. Since

$$\mathbf{\Phi}(E_{n,k} = \widetilde{\mathbf{\Phi}}(E_{n,k}$$

we conclude that

$$PE_{n,k} = E_{n,k}$$

for $k = 1, 2, \ldots, n$. QED

4.5 Normal Forms

We are already accustomed to the idea that to solve a problem involving a matrix we should transform it to an equivalent problem involving a simpler matrix. *Simpler* generally means having a special form where many of the entries vanish. We can now express this idea in a new way: *To solve a problem involving a linear map we should choose frames so that the matrix representation is simple.* A matrix in **normal form** is one which is simple (according to some notion of *simple*.)

Our purpose in this section is to understand what frames give normal forms. Most of these definitions are familiar (diagonal, reduced row echelon form etc.); some are new and will be used later on. The pattern in each case is the same: first we state (or restate) the definition of the simple form in matrix theoretic language, then we give an equivalent formulation in terms of the standard basis and flag, and finally we apply the Representation Theorem 4.1.1 to say when a matrix representation has the simple form.

Notation 4.5.1. Throughout we will use the notations

$$\begin{aligned} \mathbf{V}_k &= \operatorname{Span}(\phi_1, \phi_2, \ldots, \phi_k) \\ \mathbf{W}_k &= \operatorname{Span}(\psi_1, \psi_2, \ldots, \psi_k) \end{aligned}$$

for the (elements of the) flags determined by $\mathbf{\Phi}$ and $\mathbf{\Psi}$ respectively as well as the notation $E_{n,k}$ for the standard flag introduced Definition 4.4.1. Recall also that

$$I_{n,j} = \mathrm{col}_j(I_n)$$

denotes the jth column of the $n \times n$ identity matrix I_n. Also for $A \in \mathbb{F}^{m\times n}$, and subspaces $V \subseteq \mathbb{F}^{n\times 1}$ and $W \subseteq \mathbb{F}^{m\times 1}$ $A(V) \subseteq \mathbb{F}^{m\times 1}$ denotes the image of V and $A^{-1}(W) \subseteq \mathbb{F}^{n\times 1}$ denotes the preimage of W under the matrix map corresponding to A, i.e.

$$A(V) = \{AX \in \mathbb{F}^{m\times 1} : X \in V\}.$$

and

$$A^{-1}(W) = \{X \in \mathbb{F}^{n\times 1} : AX \in W\}.$$

By Theorems 2.6.4 and nullspace-subspace, these are again subspaces.

4.5.1 Zero-One Normal Form

A matrix $D \in \mathbb{F}^{m\times n}$ is in **zero-one normal form** iff

$$D = \begin{bmatrix} I_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix}$$

where I_r is the $r \times r$ identity matrix. Here's how to say this definition in the language of this chapter.

Proposition 4.5.2. *The matrix $D \in \mathbb{F}^{m\times n}$ is in zero-one normal form iff*

$$\begin{array}{lll} DI_{n,j} = & I_{m,j} & \textit{for } j = 1, 2, \ldots, r; \\ DI_{n,j} = & 0 & \textit{for } j = r+1, r+2, \ldots, n. \end{array}$$

where $I_{n,j}$ is as in 4.5.1.

For example, the matrix

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

satisfies

$$DI_{4,1} = I_{3,1}, \quad DI_{4,2} = I_{3,2} \quad DI_{4,3} = DI_{4,4} = 0.$$

Corollary 4.5.3. *The matrix representing the linear map $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ in the frames $\mathbf{\Phi}$ and $\mathbf{\Psi}$ is in zero-one normal form iff there is a number $r \leq n, m$ such that*

$$\begin{array}{rcl} \mathbf{T}(\phi_j) = & \psi_j & \text{for } j = 1, 2, \ldots, r; \\ \mathbf{T}(\phi_j) = & \mathbf{0} & \text{for } j = r+1, r+2, \ldots, n. \end{array}$$

Theorem 4.5.4. *For any linear map $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ there are frames $\mathbf{\Phi}$ and $\mathbf{\Psi}$ such that the matrix representing $\mathbf{T}$ in the frames $\mathbf{\Phi}$ and $\mathbf{\Psi}$ is in zero-one normal form.*

Proof. Let $(\phi_{n-r+1}, \phi_{n-r+2}, \ldots, \phi_n)$ be a basis for $\mathcal{N}(\mathbf{T})$ and extend it to a basis $(\phi_1, \phi_2, \ldots, \phi_n)$ for $\mathbf{V}$. For $j = 1, 2, \ldots, r$ let $\psi_j = \mathbf{T}(\phi_j)$. We claim that $(\psi_1, \psi_2, \ldots, \psi_r)$ is a basis for the range $\mathcal{R}(\mathbf{T})$ of $\mathbf{T}$. We must verify three things:

(1) $\psi_j \in \mathcal{R}(\mathbf{T})$ for $j = 1, 2, \ldots, r$.

(2) $\mathcal{R}(\mathbf{T}) = \mathrm{Span}(\psi_1, \psi_2, \ldots, \psi_r)$.

(3) The sequence $(\psi_1, \psi_2, \ldots, \psi_r)$ is independent.

Part (1) is immediate from the definition of the range and the fact that $\psi_j = \mathbf{T}(\phi_j)$. For part (2) choose $\mathbf{w} \in \mathcal{R}(\mathbf{T})$. Then $\mathbf{w} = \mathbf{T}(\mathbf{v})$ for some $\mathbf{v} \in \mathbf{V}$. As $(\phi_1, \phi_2, \ldots, \phi_n)$ is a basis for $\mathbf{V}$ there are numbers $x_1, x_2, \ldots, x_n$ with

$$\mathbf{v} = \sum_{j=1}^{n} x_j \phi_j.$$

Hence

$$\begin{aligned} \mathbf{w} &= \mathbf{T}(\mathbf{v}) \\ &= \mathbf{T}\left(\sum_{j=1}^{n} x_j \phi_j\right) \\ &= \sum_{j=1}^{n} x_j \mathbf{T}(\phi_j) \\ &= \sum_{j=1}^{r} x_j \mathbf{T}(\phi_j) \\ &= \sum_{j=1}^{r} x_j \psi_j. \end{aligned}$$

For part (3) assume that the numbers $y_1, y_2, \ldots, y_r$ satisfy

$$\sum_{j=1}^{r} y_j \psi_j = \mathbf{0};$$

we must show they vanish. Let

$$\mathbf{u} = \sum_{j=1}^{r} y_j \phi_j \qquad (i)$$

so that

$$\begin{aligned} \mathbf{T}(u) &= \mathbf{T}\left(\sum_{j=1}^{r} y_j \phi_j\right) \\ &= \sum_{j=1}^{r} y_j \mathbf{T}(\phi_j) \\ &= \sum_{j=1}^{r} y_j \psi_j \\ &= \mathbf{0} \end{aligned}$$

so $\mathbf{u} \in \mathcal{N}(\mathbf{T})$. Hence there are numbers $y_{n-r+1}, y_{n-r+2}, \ldots, y_n$ with

$$\mathbf{u} = \sum_{j=n-r+1}^{n} y_j \phi_j. \qquad (ii)$$

Combining (i) and (ii) gives

$$\sum_{j=1}^{r} y_j \phi_j - \sum_{j=n-r+1}^{n} y_j \phi_j = \mathbf{0}$$

so the coefficients y_j vanish as $(\phi_1, \phi_2, \ldots, \phi_n)$ is a basis for $\mathbf{V}$.

Now extend $(\psi_1, \psi_2, \ldots, \psi_r)$ to a basis $(\psi_1, \psi_2, \ldots, \psi_n)$ for $\mathbf{W}$. The conclusion of the theorem follows immediately from the previous corollary. QED

4.5.2 Row Echelon Form

An $m \times n$ matrix R is in **row echelon form** iff

(1) All the rows which vanish identically (if any) appear below the other (non-zero) rows.

(2) The leading entry in any row appears to the left of the leading entry of any non-zero row below.

(Here the **leading entry** in any row is the first non-zero entry in that row.) Here's how to say this definition in the language of this chapter.

Proposition 4.5.5. *The matrix $R \in \mathbb{F}^{m\times n}$ is in row echelon form iff there are indices $j_0 = 0 < 1 \leq j_1 < j_2 < \cdots < j_r \leq n$ such that*

$$R\big(E_{n,j}\big) = E_{m,i} \quad \text{for } j_i \leq j < j_{i+1}$$

for $i = 0, 1, 2, \ldots, r-1$. (See 4.5.1.) The leading entry in the i-th row occurs in the j_i-th column.

For example, if $a_1a_2a_3 \neq 0$, then the matrix

$$R = \begin{bmatrix} 0 & a_1 & b_1 & c_{12} & b_2 & c_{13} & c_{14} \\ 0 & 0 & a_2 & c_{22} & b_3 & c_{23} & c_{24} \\ 0 & 0 & 0 & 0 & a_3 & c_{33} & c_{34} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

is in row echelon form with $j_1 = 2$, $j_2 = 3$, $j_3 = 5$, since $RE_{7,1} = E_{5,0}$, $RE_{7,2} = E_{5,1}$, $RE_{7,3} = RE_{7,4} = E_{5,2}$, $RE_{7,5} = RE_{7,6} = RE_{7,7} = E_{5,3}$. The leading entries are a_1, a_2, a_3.

By the Representation Theorem 4.1.1, the matrix representing the linear map $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ in the frames $\mathbf{\Phi}$ and $\mathbf{\Psi}$ is in Row Echelon Form iff there are indices $j_0 = 0 < 1 \leq j_1 < j_2 < \cdots < j_r \leq n$ such that

$$\mathbf{T}(\mathbf{V}_j) = \mathbf{W}_i \quad \text{for } j_i \leq j < j_{i+1}$$

for $i = 0, 1, 2, \ldots, r-1$ where

$$\begin{aligned} \mathbf{V}_j &= \mathrm{Span}(\phi_1, \phi_2, \ldots, \phi_j) \\ \mathbf{W}_i &= \mathrm{Span}(\psi_1, \psi_2, \ldots, \phi_i) \end{aligned}$$

are the flags determined by the frames $\mathbf{\Phi}$ and $\mathbf{\Psi}$.

4.5.3 Reduced Row Echelon Form

An $m \times n$ matrix R is in **reduced row echelon form** iff it is in row echelon form and, in addition, satisfies

(3) The leading entry in any non-zero row is a 1,

(4) All other entries in the column of a leading entry are 0.

Here's how to say this definition in the language of this chapter.

Proposition 4.5.6. *A matrix $R \in \mathbb{F}^{m\times n}$ is in reduced row echelon form iff there are indices $j_0 = 0 < 1 \le j_1 < j_2 < \cdots < j_r \le n$ such that*

$$\begin{array}{ll} RI_{n,j_i} = I_{m,i} & \textit{for } i = 1, 2, \dots, r, \\ RI_{n,j} \in E_{m,i} & \textit{for } j_i < j < j_{i+1}. \end{array}$$

(See 4.5.1.)

For example, the matrix

$$R = \begin{bmatrix} 0 & 1 & 0 & c_{12} & 0 & c_{13} & c_{14} \\ 0 & 0 & 1 & c_{22} & 0 & c_{23} & c_{24} \\ 0 & 0 & 0 & 0 & 1 & c_{33} & c_{34} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

is in reduced row echelon form since with $j_1 = 2$, $j_2 = 3$, $j_3 = 5$, we have

$$\begin{array}{rcl} RI_{7,1} &=& 0 \in E_{5,0} \\ RI_{7,2} &=& I_{5,1} \\ RI_{7,3} &=& I_{5,2} \\ RI_{7,4} &=& c_{12}I_{5,1} + c_{22}I_{5,2} \in E_{5,2} \\ RI_{7,5} &=& I_{5,3} \\ RI_{7,6} &=& c_{13}I_{5,1} + c_{23}I_{5,2} + c_{33}I_{5,3} \in E_{5,3} \\ RI_{7,7} &=& c_{14}I_{5,1} + c_{24}I_{5,2} + c_{34}I_{5,3} \in E_{5,3} \end{array}$$

Corollary 4.5.7. *The matrix representing the linear map $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ in the frames $\mathbf{\Phi}$ and $\mathbf{\Psi}$ is in reduced row echelon form iff there are indices $j_0 = 0 < 1 \le j_1 < j_2 < \cdots < j_r \le n$ such that*

$$\begin{array}{ll} \mathbf{T}(\phi_{j_i}) = \psi_i & \textit{for } i = 1, 2, \dots, r, \\ \mathbf{T}(\phi_j) \in \mathbf{W}_i & \textit{for } j_i < j < j_{i+1}, \end{array}$$

where $\mathbf{W}_i = \mathrm{Span}(\psi_1, \psi_2, \dots, \psi_i)$.

Theorem 4.5.8. *For any* $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ *and frame* $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ *there is a frame* $\mathbf{\Psi} : \mathbb{F}^{n\times 1} \to \mathbf{W}$ *such that the matrix representing* $\mathbf{T}$ *in the frames* $\mathbf{\Phi}$ *and* $\mathbf{\Psi}$ *is in reduced row echelon form.*

Proof. The indices $j_1, j_2, \ldots, j_r$ are precisely those values of j for which

$$\mathbf{T}(\phi_j) \notin \mathbf{T}(\mathbf{V}_{j-1})). \qquad (\sharp)$$

The fact that the j_i-th column of the representing matrix must be $I_{m,i}$, the i-th column of the identity forces us to defined ψ_i by the equation

$$\psi_i = \mathbf{T}(\phi_{j_i}). \qquad (\flat)$$

Then the sequence $(\psi_1, \ldots, \psi_r)$ is independent since

$$\psi_i \notin \mathrm{Span}(\psi_1, \psi_2, \ldots, \psi_{i-1})$$

by definition. Extend this sequence to a basis $(\psi_1, \ldots, \psi_m)$ of $\mathbf{W}$. QED

Corollary 4.5.9. *The matrix of Theorem 4.5.8 is unique.*

Proof. Here's what the statement means. Assume that $\mathbf{\Psi}$ and $\widetilde{\mathbf{\Psi}}$ are two frames for $\mathbf{W}$, that $R \in \mathbb{F}^{m\times n}$ is the matrix representing $\mathbf{T}$ in the frames $\mathbf{\Phi}$ and $\mathbf{\Psi}$, and that $\widetilde{R} \in \mathbb{F}^{m\times n}$ is is the matrix representing $\mathbf{T}$ in the frames $\mathbf{\Phi}$ and $\widetilde{\mathbf{\Psi}}$. The corollary asserts that if both R and $\widetilde{R}$ are in reduced row echelon form, then $R = \widetilde{R}$. But this is clear from the proof of the RREF Theorem: equations ($\sharp$) and ($\flat$) determine $\psi_1, \psi_2, \ldots, \psi_r$ uniquely. We are free to extend the basis in any way we like, but this will not affect the matrix representing $\mathbf{T}$ since $(\psi_1, \psi_2, \ldots, \psi_r)$ is a basis for $\mathcal{R}(\mathbf{T})$ of $\mathbf{T}$. QED

4.5.4 Diagonalization

A square matrix $D \in \mathbb{F}^{n\times n}$ is called **diagonal** iff $\mathrm{entry}_{ij}(D) = 0$ for $i \neq j$, that is, iff all the off-diagonal entries vanish. Here's how to say this definition in the language of this chapter.

Proposition 4.5.10. *A matrix* $D \in \mathbb{F}^{n\times n}$ *is diagonal iff the columns* $I_{n,j}$ *from the standard basis iff for* $j = 1, 2, \ldots, n$ *we have*

$$DI_{n,j} = \lambda_j I_{n,j}$$

where $\lambda_j = \mathrm{entry}_{jj}(D)$. *(See 4.5.1.)*

A number λ is called an **eigenvalue** of a linear map $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ iff there is a non-zero vector $\mathbf{v} \in \mathbf{V}$ such that

$$\mathbf{T}(\mathbf{v}) = \lambda \mathbf{v}.$$

Any vector $\mathbf{v}$ satisfying this equation is called an **eigenvector** for the eigenvalue λ.

Corollary 4.5.11. *Let $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ be a linear map from $\mathbf{V}$ to itself, $(\phi_1, \dots, \phi_n)$ be a basis for $\mathbf{T}$, and $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ be the corresponding frame. The matrix representing $\mathbf{T}$ in the frame $\mathbf{\Phi}$ is diagonal iff the vectors ϕ_j are eigenvectors of $\mathbf{T}$:*

$$\mathbf{T}(\phi_j) = \lambda_j \phi_j \tag{$\natural$}$$

for $j = 1, 2, \dots, n$.

Definition 4.5.12. When $\mathbf{T}$ and $\mathbf{\Phi}$ are related by equation ($\natural$), we say that **$\mathbf{\Phi}$ diagonalizes $\mathbf{T}$**. A linear map $\mathbf{T}$ is called **diagonalizable** iff there is a frame which diagonalizes it and a square matrix A is called diagonalizable iff the corresponding matrix map is, i.e. iff there is an invertible matrix P such that $P^{-1}AP$ is diagonal.

4.5.5 Triangular Matrices

A square matrix B is **triangular** iff all the entries below the diagonal vanish, i.e. $\text{entry}_{ij}(B) = 0$ for $i > j$. For example, the matrix

$$\begin{bmatrix} a & b & d \\ 0 & d & e \\ 0 & 0 & f \end{bmatrix}$$

is triangular. Here's how to say this definition in the language of this chapter.

Proposition 4.5.13. *A matrix $B \in \mathbb{F}^{n\times n}$ is triangular iff*

$$B\big(E_{n,k}\big) \subseteq E_{n,k}.$$

A matrix $B \in \mathbb{F}^{n\times n}$ is invertible and triangular iff

$$B\big(E_{n,k}\big) = E_{n,k}.$$

(See 4.5.1.)

Proof. Since $I_{n,k} \in E_{n,k}$ the set inclusion means that

$$\mathrm{col}_k(B) = BI_{n,k} = \sum_{i=1}^{k} b_{ik} I_{n,i}$$

where $b_{ik} = \mathrm{entry}_{ik}(B)$. This says that $\mathrm{entry}_{ik}(B) = 0$ for $i > k$, that is, that B is triangular. If B is invertible and triangular, then $B\big(E_{n,k}\big)$ and $E_{n,k}$ have the same dimension and so must be equal. If B is not invertible, then $B\big(E_{n,n}\big) \neq E_{n,n}$. QED

Corollary 4.5.14. *The matrix representing the linear map* $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ *in the frame* $\mathbf{\Phi}$ *is triangular iff*

$$\mathbf{T}(\mathbf{V}_k) \subseteq \mathbf{V}_k$$

for $k = 1, 2, \ldots, n$ *where*

$$\mathbf{V}_j = \mathrm{Span}(\phi_1, \phi_2, \ldots, \phi_j)$$

is the flag determined by the frame $\mathbf{\Phi}$.

4.5.6 Strictly Triangular Matrices

A matrix $N \in \mathbb{F}^{n\times n}$ is called **strictly triangular** iff $\mathrm{entry}_{ij}(N) = 0$ for $i \geq j$, that is, iff all its entries on or below the diagonal vanish. For example, the matrix

$$\begin{bmatrix} 0 & a & b \\ 0 & 0 & c \\ 0 & 0 & 0 \end{bmatrix}$$

is strictly triangular. Here's how to say this definition in the language of this chapter.

Proposition 4.5.15. *A matrix* $N \in \mathbb{F}^{n\times n}$ *is strictly triangular iff*

$$N\big(E_{n,k}\big) \subseteq E_{n,k-1}$$

for $k = 1, 2, \ldots, n$. *(See 4.5.1.)*

Proof. Exercise.

Corollary 4.5.16. *The matrix representing the linear map* $\mathbf{N} : \mathbf{V} \to \mathbf{V}$ *is strictly triangular iff*

$$\mathbf{N}(\mathbf{V}_k) \subseteq \mathbf{V}_{k-1}$$

for $k = 1, 2, \ldots, n$ *where*

$$\mathbf{V}_j = \mathrm{Span}(\phi_1, \phi_2, \ldots, \phi_j)$$

is the flags determined by the frame $\mathbf{\Phi}$*.*

4.6 Exercises

Exercise 4.6.1. In each of the following you are given vector spaces $\mathbf{V}$ and $\mathbf{W}$, frames $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ and $\mathbf{\Psi} : \mathbb{F}^{m\times 1} \to \mathbf{W}$, and a linear map $\mathbf{T} : \mathbf{V} \to \mathbf{W}$. Find the matrix $A \in \mathbb{F}^{m\times n}$ which represents the map $\mathbf{T}$ in the frames $\mathbf{\Phi}$ and $\mathbf{\Psi}$.

(1) $\mathbf{V} = \mathrm{Poly}_2(\mathbb{F})$, $\mathbf{W} = \mathrm{Poly}_1(\mathbb{F})$, $\mathbf{\Phi}(X)(t) = x_1 + x_2 t + x_3 t^2$, $\mathbf{\Psi}(Y)(t) = y_1 + y_2 t$, $\mathbf{T}(f) = f'$.

(2) $\mathbf{V}$, $\mathbf{W}$, $\mathbf{\Phi}$, $\mathbf{\Psi}$ as in (1), $\mathbf{T}(f)(t) = (f(t+h) - f(t))/h$.

(3) $\mathbf{V} = \mathrm{Cos}_2(\mathbb{F})$, $\mathbf{W} = \mathrm{Sin}_1(\mathbb{F})$, $\mathbf{\Phi}(X)(t) = x_1 + x_2\cos(t) + x_3\cos(2t)$, $\mathbf{\Psi}(Y)(t) = y_1\sin(t) + y_2\sin(2t)$, $\mathbf{T}(f) = f'$.

(4) $\mathbf{V}$, $\mathbf{\Phi}$ as in (1), $\mathbf{W} = \mathbb{F}^{1\times 3}$, $\mathbf{\Psi}(Y) = Y^*$,

$$\mathbf{T}(f)(t) = \begin{bmatrix} f(0) & f(1) & f(2) \end{bmatrix}.$$

Here $x_j = \mathrm{entry}_j(X)$ and $y_i = \mathrm{entry}_i(Y)$.

Exercise 4.6.2. In each of the following you are given a vector space $\mathbf{V}$, a frame $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$, and a linear map $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ from $\mathbf{V}$ to itself. Find the matrix $A \in \mathbb{F}^{n\times n}$. which represents the map $\mathbf{T}$ in the frame $\mathbf{\Phi}$.

(1) $\mathbf{V} = \mathrm{Poly}_2(\mathbb{F})$, $\mathbf{\Phi}(X)(t) = x_1 + x_2 t + x_3 t^2$, $\mathbf{T}(f) = f'$.

(2) $\mathbf{V}$ and $\mathbf{\Phi}$ as in (1), $\mathbf{T}(f)(t) = (f(t+h) - f(t))/h$.

(3) $\mathbf{V} = \mathrm{Trig}_1(\mathbb{F})$, $\mathbf{\Phi}(X)(t) = x_1 + x_2\cos(t) + x_3\sin(t)$, $\mathbf{T}(f) = f'$.

(4) $\mathbf{V}$ and $\mathbf{\Phi}$ as in (3), $\mathbf{T}(f)(t) = (f(t+h) - f(t))/h$.

Here $x_j = \text{entry}_j(X)$.

Exercise 4.6.3. What is the dimension of the vector space $\mathcal{L}(\mathbf{V}, \mathbf{W})$ of linear maps from $\mathbf{V}$ to $\mathbf{W}$?

Exercise 4.6.4. Let

$$\begin{array}{ll} \phi_1(t) = 1 & \psi_1(t) = (t-2)(t-3)/2 \\ \phi_2(t) = t & \psi_2(t) = -(t-1)(t-3) \\ \phi_3(t) = t^2 & \psi_3(t) = (t-1)(t-2)/2 \end{array}$$

Each of the sequences (ϕ_1, ϕ_2, ϕ_3) and (ψ_1, ψ_2, ψ_3) is a basis for $\text{Poly}_2(\mathbb{F})$. Find the transition matrix from (ψ_1, ψ_2, ψ_3) to (ϕ_1, ϕ_2, ϕ_3). Find the transition matrix from (ϕ_1, ϕ_2, ϕ_3) to (ψ_1, ψ_2, ψ_3).

Exercise 4.6.5. Let $(\phi_1, \phi_2, \phi_3, \phi_4, \phi_5)$ be as basis for a vector space $\mathbf{V}$. Find the transition matrix from this basis to the basis $(\phi_3, \phi_5, \phi_2, \phi_1, \phi_4)$.

Exercise 4.6.6. In each of the following, you are given a linear map $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ and frames $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ and $\mathbf{\Psi} : \mathbb{F}^{m\times 1} \to \mathbf{W}$. Find the matrix A representing $\mathbf{T}$ in the frames $\mathbf{\Phi}$ and $\mathbf{\Psi}$. Also say if $\mathbf{T}$ is one-one and if it is onto.

(1) $\mathbf{V} = \text{Poly}_3(\mathbb{F})$, $\mathbf{W} = \text{Poly}_2(\mathbb{F})$, $\mathbf{T}(f) = f'$, $\psi_i(t) = t^{i-1}$ for $i = 1, 2, 3$.

(2) $\mathbf{V} = \text{Poly}_3(\mathbb{F})$, $\mathbf{W} = \mathbb{F}^{1\times 3}$, $T(f) = \begin{bmatrix} f(1) & f(2) & f(3) \end{bmatrix}$, $\phi_j(t) = t^{j-1}$ for $j = 1, 2, 3, 4$, $\psi_i = \text{row}_i(I_3)$.

(3) $\mathbf{V} = \mathbb{F}^{3\times 1}$, $\mathbf{W} = \mathbb{F}^{2\times 1}$, $\mathbf{T}(X) = \begin{bmatrix} 3x_1 + x_3 \\ x_2 + 6x_3 \end{bmatrix}$, $\phi_j = \text{col}_j(I_3)$, $\psi_i = \text{col}_i(I_2)$. (Here $x_j = \text{entry}_j(X)$.)

(4) $\mathbf{V} = \mathbb{F}^{3\times 1}$, $\mathbf{W} = \mathbb{F}^{2\times 1}$, $\mathbf{T}(X) = \begin{bmatrix} 3x_1 + x_3 \\ x_2 + 6x_3 \end{bmatrix}$, $\phi_j = \text{col}_j(P)$, $\psi_i = \text{col}_i(Q)$, where

$$P = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 0 & 0 & 1 \end{bmatrix}, \quad Q = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

(5) $\mathbf{V} = \text{Cos}_n(\mathbb{F})$, $\mathbf{W} = \text{Sin}_n(\mathbb{F})$, $\mathbf{T}(f) = f'$, $\phi_j(t) = \cos(j-1)t$, $\psi_k(t_= \sin(kt)$.

(6) $\mathbf{V} = \{\begin{bmatrix} x & y & z \end{bmatrix} : x + 2y + 3z = 0\}$, $\mathbf{W} = \mathbb{F}^{1\times 2}$, $\mathbf{T}(\begin{bmatrix} x & y & z \end{bmatrix}) = \begin{bmatrix} x & y \end{bmatrix}$, $\phi_1 = \begin{bmatrix} -3 & 0 & 1 \end{bmatrix}$, $\phi_1 = \begin{bmatrix} 0 & -3 & 2 \end{bmatrix}$, $\psi_1 = \begin{bmatrix} 1 & 0 \end{bmatrix}$, $\psi_2 = \begin{bmatrix} 0 & 1 \end{bmatrix}$.

(7) $\mathbf{V} = \mathrm{Poly}_3(\mathbb{F})$, $\mathrm{Poly}_2(\mathbb{F})$, $T(f)(t) = f'(t+1)$, $\phi_j(t) = t^{j-1}$, $\psi_j(t) = t^{j-1}$.

Exercise 4.6.7. For each of the map $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ of the previous problem find a frame $\tilde{\mathbf{\Psi}} : \mathbb{F}^{m\times 1} \to \mathbf{W}$ such that the matrix representing $\mathbf{T}$ in the frames $\mathbf{\Phi}$ and $\tilde{\mathbf{\Psi}}$ is in reduced row echelon form.

Exercise 4.6.8. For each of the map $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ of the previous problem find frames $\tilde{\mathbf{\Phi}} : \mathbb{F}^{n\times 1} \to \mathbf{V}$ and $\tilde{\mathbf{\Psi}} : \mathbb{F}^{m\times 1} \to \mathbf{W}$ such that the matrix representing $\mathbf{T}$ in the frames $\tilde{\mathbf{\Phi}}$ and $\tilde{\mathbf{\Psi}}$ is in zero-one diagonal form.

Exercise 4.6.9. Let $\mathbf{T} : \mathrm{Poly}_3(\mathbb{F}) \to \mathbb{F}^{1\times 3}$ be defined by

$$\mathbf{T}(f) = \begin{bmatrix} f(1) & f'(1) & f(1) \end{bmatrix}.$$

(1) Find a basis for the null space of $\mathbf{T}$ and extend it to a basis for $\mathrm{Poly}_3(\mathbb{F})$.

(2) Find a basis for the range of $\mathbf{T}$ and extend it to a basis for $\mathbb{F}^{1\times 3}$.

(3) Find the matrix representing T in these frames.

Is T one-one? onto?

Exercise 4.6.10. In each of the following, you are given a linear map $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ and a frame $\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$. Find the matrix A representing the map $\mathbf{T}$ in the frame $\mathbf{\Phi}$.

(1) $\mathbf{V} = \mathrm{Poly}_n(\mathbb{F})$, $\mathbf{T}(f)(t) = f(t+a)$, $\phi_j(t) = t^{j-1}$.

(2) $\mathbf{V} = \mathrm{Trig}_n(\mathbb{F})$, $\mathbf{T}(f)(t) = f(t+a)$, $\phi_j(t) = e^{(n+1-j)it}$.

(3) $\mathbf{V} = \mathrm{Poly}_n(\mathbb{F})$, $\mathbf{T}(f)(t) = f'(t)$, $\phi_j(t) = t^{j-1}$.

(4) $\mathbf{V} = \mathrm{Trig}_n(\mathbb{F})$, $\mathbf{T}(f)(t) = f'(t)$, $\phi_j(t) = e^{(n+1-j)it}$.

(5) $\mathbf{V} = \mathrm{Poly}_n(\mathbb{F})$, $\mathbf{T}(f)(t) = f''(t)$, $\phi_j(t) = t^{j-1}$.

(6) $\mathbf{V} = \mathrm{Trig}_n(\mathbb{F})$, $\mathbf{T}(f)(t) = f''(t)$, $\phi_j(t) = e^{(n+1-j)it}$.

Exercise 4.6.11. For each of the maps $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ of the previous problem, find its eigenvalues and eigenvectors.

Exercise 4.6.12. Suppose that $\mathbf{V}$ is a vector space of dimension n and that the linear map $\mathbf{T}: \mathbf{V} \to \mathbf{V}$ has n distinct eigenvalues. Show there is a basis of $\mathbf{V}$ consisting of eigenvectors of $\mathbf{T}$. Hint: The key point is that the sequence of eigenvectors is independent. This can be proved by assuming a linear relation and applying $f(\mathbf{T})$ for various polynomials $f(t)$. See Exercise 3.14.18.

Exercise 4.6.13. Show that the matrix

$$N = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

is not diagonalizable, i.e. there is no invertible matrix P such that $P^{-1}AP$ is a diagonal matrix.

Exercise 4.6.14. Define $\mathbf{T}: \mathrm{Poly}_n(\mathbb{F}) \to \mathrm{Poly}_n(\mathbb{F})$ by

$$\mathbf{T}(f)(t) = f(t+b)$$

where b is a constant. Find the eigenvalues of $\mathbf{T}$. Is $\mathbf{T}$ diagonalizable? Hint: Find the matrix representing $\mathbf{T}$ in the standard basis $\phi_j(t) = t^{j-1}$. If you can't do the general case try the case $n = 1$ first.

Exercise 4.6.15. Define $\mathbf{T}: \mathrm{Poly}_n(\mathbb{F}) \to \mathrm{Poly}_n(\mathbb{F})$ by

$$\mathbf{S}(f)(t) = f(bt)$$

where b is a constant. Find the eigenvalues of $\mathbf{T}$. Is $\mathbf{T}$ diagonalizable?

Exercise 4.6.16. Define $\mathbf{T}: \mathrm{Trig}_n(\mathbb{F}) \to \mathrm{Trig}_n(\mathbb{F})$ by

$$\mathbf{T}(f)(t) = f(t+b)$$

where b is a constant. Find the eigenvalues of $\mathbf{T}$. Is $\mathbf{T}$ diagonalizable?

Exercise 4.6.17. The matrix

$$A = \begin{bmatrix} 0 & a_{12} & a_{13} & a_{14} \\ 0 & 0 & a_{23} & a_{24} \\ 0 & 0 & 0 & a_{34} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

satisfies $\mathrm{entry}_{ij}(A) = 0$ for $j < i + 1$ and the matrix

$$B = \begin{bmatrix} 0 & 0 & b_{13} & b_{14} \\ 0 & 0 & 0 & b_{24} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

satisfies $\text{entry}_{jk}(B) = 0$ for $k < j + 2$. Compute AB and conclude that it satisfies $\text{entry}_{ik}(AB) = 0$ for $k < i + 3$.

Exercise 4.6.18. A square matrix $A \in \mathbb{F}^{n\times n}$ is called **p-triangular** iff

$$\text{entry}_{ij}(A) = 0 \text{ for } j < i + p.$$

Thus the terms 0-*triangular* and *triangular* are synonymous, and the terms 1-*triangular* and *strictly triangular* are synonymous. Show that if A is p-triangular matrix and B is q-triangular, then AB is $(p+q)$-triangular. Hint: You can, of course, simply calculate $\text{entry}_{ik}(AB)$ and show that it is zero for $k < i + p + q$. However, it is more elegant to express the property of being p-triangular in terms of the standard flag.

Exercise 4.6.19. A matrix $N \in \mathbb{F}^{n\times n}$ is called **nilpotent** iff $N^p = 0$ for some positive integer p. Show that a strictly triangular matrix N is nilpotent.

Exercise 4.6.20. Let $U = I - N$ where $I = I_3$ is the 3×3 identity matrix and

$$N = \begin{bmatrix} 0 & a & b \\ 0 & 0 & c \\ 0 & 0 & 0 \end{bmatrix}.$$

Show that $N^3 = 0$ and $U^{-1} = I + N + N^2$.

Exercise 4.6.21. A square matrix U is called **unipotent** iff it is the sum of the identity matrix and a nilpotent matrix. Show that a unipotent matrix is invertible. (Hint: Factor $I - N^n$ to find a formula for the inverse of $U = I - N$.)

Exercise 4.6.22. Call a square matrix **uni-triangular** iff it is triangular and all its diagonal entries are one. Show that a uni-triangular matrix is invertible.

Exercise 4.6.23. A triangular matrix $A \in \mathbb{F}^{3\times 3}$ may be written as $A = DU$ where

$$A = \begin{bmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{bmatrix}, \quad D = \begin{bmatrix} a & 0 & 0 \\ 0 & d & 0 \\ 0 & 0 & f \end{bmatrix}, \quad U = \begin{bmatrix} 1 & a^{-1}b & a^{-1}c \\ 0 & 1 & d^{-1}e \\ 0 & 0 & 1 \end{bmatrix}.$$

Find A^{-1}. (Don't forget that $(DU)^{-1} = U^{-1}D^{-1}$.)

Exercise 4.6.24. Suppose that A is invertible and triangular. Show that $A = DU$ where D is invertible diagonal and U is a uni-triangular. Use this to find a formula for A^{-1}.

Exercise 4.6.25 (Important). Let $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ be a linear map between finite dimensional vector spaces. Show that $\mathbf{T}$ is one-one if and only if $\mathbf{T}^*$ is onto and that $\mathbf{T}$ is onto if and only if $\mathbf{T}^*$ is one-one. (See Exercises 2.4.10 and 2.4.11.)

Exercise 4.6.26 (Important). Let $\mathbf{A} : \mathbf{V}_1 \to \mathbf{W}_1$ and $\mathbf{B} : \mathbf{V}_2 \to \mathbf{W}_2$ be a linear maps between finite dimensional vector spaces. Say that that $\mathbf{A}$ and $\mathbf{B}$ are **equivalent** iff there exist isomorphisms $\mathbf{P} : \mathbf{V}_2 \to \mathbf{V}_1$ and $\mathbf{Q} : \mathbf{W}_2 \to \mathbf{W}_1$ such that

$$\mathbf{A} = \mathbf{Q} \circ \mathbf{B} \circ \mathbf{P}^{-1}.$$

Show that $\mathbf{A}$ and $\mathbf{B}$ are equivalent if and only if $\mathbf{V}_1$ and $\mathbf{V}_2$ have the same dimension, $\mathbf{W}_1$ and $\mathbf{W}_2$ have the same dimension, and $\mathbf{A}$ and $\mathbf{B}$ have the same rank.

Chapter 5

Block Diagonalization

Not every square matrix can be diagonalized. In this chapter we will see that every square matrix can be "block diagonalized"

5.1 Direct Sums

Let $\mathbf{V}$ be a vector space. The notation

$$\mathbf{V} = \mathbf{W} \oplus \mathbf{U}$$

says that $\mathbf{V}$ is the **direct sum** of $\mathbf{W}$ and $\mathbf{U}$. This means that $\mathbf{W}$ and $\mathbf{U}$ are subspaces of $\mathbf{V}$ and that for every $\mathbf{v} \in \mathbf{V}$ there are unique $\mathbf{w} \in \mathbf{W}$ and $\mathbf{u} \in \mathbf{U}$ such that

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

More generally, the notation

$$\mathbf{V} = \mathbf{V}_1 \oplus \mathbf{V}_2 \oplus \cdots \oplus \mathbf{V}_m$$

means that the spaces $\mathbf{V}_i$ $(i = 1, 2, \ldots, m)$ are subspaces of $\mathbf{V}$ and for for every $\mathbf{v} \in \mathbf{V}$ there are unique vectors $\mathbf{v}_i \in \mathbf{V}_i$ such that

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_m.$$

Another notation for the direct sum, analogous to the sigma notation for ordinary sums, is

$$\mathbf{V} = \bigoplus_{j=1}^{m} \mathbf{V}_j.$$

When $\mathbf{V} = \bigoplus_{j=1}^{m} \mathbf{V}_j$ we say the subspaces $\mathbf{V}_j$ give a **direct sum decomposition** of $\mathbf{V}$. When $\mathbf{V} = \mathbf{W} \oplus \mathbf{U}$, one says that the subspace $\mathbf{U}$ of $\mathbf{V}$ is a **complement** to the subspace $\mathbf{W}$ in the vector space $\mathbf{V}$.

To prove the equation $\mathbf{V} = \mathbf{W} \oplus \mathbf{U}$ we must show four things:

(1) $\mathbf{W}$ is a subspace of $\mathbf{V}$.

(2) $\mathbf{U}$ is a subspace of $\mathbf{V}$.

(3) $\mathbf{V} = \mathbf{W} + \mathbf{U}$ which means that every $\mathbf{v} \in \mathbf{V}$ has form $\mathbf{v} = \mathbf{w} + \mathbf{u}$ for some $\mathbf{w} \in \mathbf{W}$ and $\mathbf{u} \in \mathbf{U}$.

(4) $\mathbf{W} \cap \mathbf{U} = \{\mathbf{0}\}$ which means that the only $\mathbf{v} \in \mathbf{V}$ which is in both $\mathbf{W}$ and $\mathbf{U}$ is $\mathbf{v} = \mathbf{0}$.

Remark 5.1.1 (Uniqueness Remark). Part (4) relates to the uniqueness of the decomposition. If $\mathbf{w}_1, \mathbf{w}_2 \in \mathbf{W}$ and $\mathbf{u}_1, \mathbf{u}_2 \in \mathbf{U}$ satisfy

$$\mathbf{w}_1 + \mathbf{u}_1 = \mathbf{w}_2 + \mathbf{u}_2,$$

then $\mathbf{w}_1 - \mathbf{w}_2 = \mathbf{u}_2 - \mathbf{u}_1 \in \mathbf{W} \cap \mathbf{U}$. Then part (4) implies that $\mathbf{w}_1 - \mathbf{w}_2 = \mathbf{u}_2 - \mathbf{u}_1 = \mathbf{0}$, that is, that $\mathbf{w}_1 = \mathbf{w}_2$ and $\mathbf{u}_1 = \mathbf{u}_2$, so that the representation is unique. On the other hand, if part (4) fails, then there is a non-zero $\mathbf{v} \in \mathbf{W} \cap \mathbf{U}$. Then $\mathbf{0} \in \mathbf{V}$ has two distinct representations, $\mathbf{0} = \mathbf{0} + \mathbf{0}$ and $\mathbf{0} = \mathbf{v} + (-\mathbf{v})$, as the sum of an element of $\mathbf{W}$ and an element of $\mathbf{U}$, so that the representation is *not* unique.

The first thing to understand is that a subspace has many complements. For example, take $\mathbf{V} = \mathbb{F}^{2\times 1}$ and let $\mathbf{W}$ be the horizontal axis:

$$\mathbf{W} = \left\{ \begin{bmatrix} x_1 \\ 0 \end{bmatrix} : x_1 \in \mathbb{F} \right\}.$$

Then for any $b \in \mathbb{F}$ the space

$$\mathbf{U} = \left\{ \begin{bmatrix} bx_2 \\ x_2 \end{bmatrix} : x_2 \in \mathbb{F} \right\}$$

is a complement to $\mathbf{W}$ since any $X \in \mathbf{V} = \mathbb{F}^{2\times 1}$ can be decomposed as

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 - bx_2 \\ 0 \end{bmatrix} + \begin{bmatrix} bx_2 \\ x_2 \end{bmatrix}.$$

Note that different values of b give different complements $\mathbf{U}$ to $\mathbf{W}$. Geometrically, *any line through the origin and distinct from* $\mathbf{W}$ *is a complement to* $\mathbf{W}$ *in* $\mathbf{V} = \mathbb{F}^{2\times 1}$.

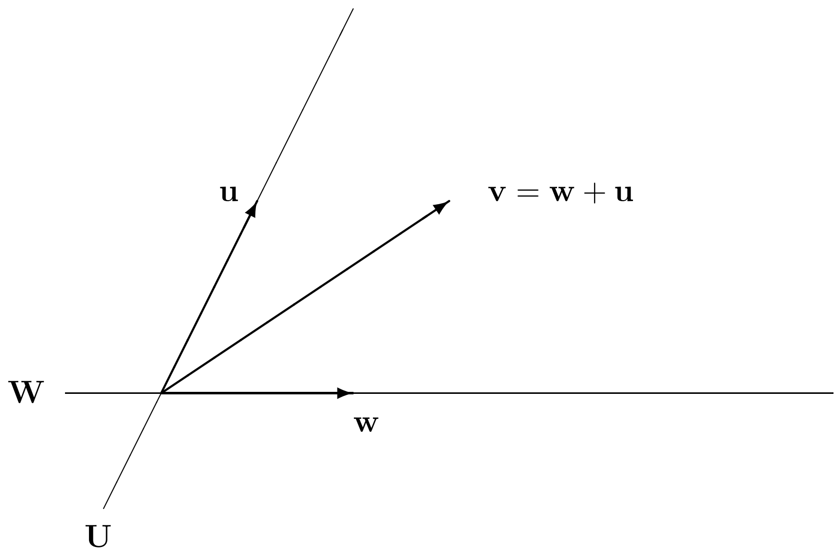


Figure 5.1: $\mathbf{V} = \mathbf{W} \oplus \mathbf{U}$

Proposition 5.1.2. *Let* $\mathbf{V}$ *be a vector space and* $\mathbf{W}, \mathbf{U} \subseteq \mathbf{V}$ *be subspaces of* $\mathbf{V}$. *Suppose that*

(1) $(\phi_1, \phi_2, \ldots, \phi_m)$ *is a basis for* $\mathbf{W}$,

(2) $(\phi_{m+1}, \phi_{m+2}, \ldots, \phi_n)$ *is a basis for* $\mathbf{U}$,

(3) $(\phi_1, \phi_2, \ldots, \phi_n)$ *is a basis for* $\mathbf{V}$,

Then $\mathbf{V} = \mathbf{W} \oplus \mathbf{U}$.

Proof. To show that $\mathbf{V} = \mathbf{W} + \mathbf{U}$ choose $\mathbf{v} \in \mathbf{V}$. By (3) there are numbers $x_1, x_2 \ldots, x_n$ with

$$\mathbf{v} = \sum_{j=1}^{n} x_j \phi_j.$$

Then $\mathbf{v} = \mathbf{w} + \mathbf{u}$ where

$$\mathbf{w} = \sum_{j=1}^{m} x_j \phi_j, \quad \mathbf{u} = \sum_{j=m+1}^{n} x_j \phi_j.$$

By (1) we have that $\mathbf{w} \in \mathbf{W}$ and by (2) we have that $\mathbf{u} \in \mathbf{U}$. To show that $\mathbf{W} \cap \mathbf{U} = \{\mathbf{0}\}$ choose $\mathbf{v}$ in this intersection. Then by (1) and (2) there are numbers $x_1, x_2, \ldots, x_n$ with

$$\mathbf{v} = \sum_{j=1}^{m} x_j \phi_j = \sum_{j=m+1}^{n} x_j \phi_j.$$

Hence

$$\mathbf{0} = \sum_{j=1}^{m} x_j \phi_j - \sum_{j=m+1}^{n} x_j \phi_j.$$

so $x_1 = x_2 = \cdots = x_n = 0$ by (3). Hence $\mathbf{v} = \mathbf{0}$. QED

Corollary 5.1.3. *Let* $\mathbf{W}$ *be a subspace of* $\mathbf{V}$*. To find a complement* $\mathbf{U}$ *to* $\mathbf{W}$ *in* $\mathbf{V}$ *proceed as follows:*

- *Find a basis* $(\phi_1, \phi_2, \ldots, \phi_m)$ *for* $\mathbf{W}$.
- *Extend it to a basis* $(\phi_1, \phi_2, \ldots, \phi_n)$ *for* $\mathbf{V}$.
- *Define* $\mathbf{U} = \mathrm{Span}(\phi_{m+1}, \phi_{m+2}, \ldots, \phi_n)$.

Corollary 5.1.4. *Suppose* $\mathbf{V} = \mathbf{W} \oplus \mathbf{U}$ *with* $\dim(\mathbf{V}) = n$*. Then there is a frame*

$$\mathbf{\Phi} : \mathbb{F}^{n\times 1} \to \mathbf{V}$$

such that

$$\begin{aligned} \mathbf{\Phi}^{-1}(\mathbf{W}) &= \{X \in \mathbb{F}^{n\times 1} : x_{m+1} = x_{m+2} = \cdots x_n = 0\} \\ \mathbf{\Phi}^{-1}(\mathbf{U}) &= \{X \in \mathbb{F}^{n\times 1} : x_1 = x_2 = \cdots x_m = 0\}. \end{aligned}$$

For each pair (m, n) of integers with $0 \le m \le n$ there is a **standard direct sum**

$$\mathbb{F}^{n\times 1} = \mathbf{W}_m^n \oplus \mathbf{U}_m^n$$

where

$$\mathbf{W}_m^n = \left\{ \begin{bmatrix} X_1 \\ 0 \end{bmatrix} X_1 \in \mathbb{F}^{m\times 1}, \;\; 0 = 0_{(n-m)\times 1} \right\},$$

$$\mathbf{U}_m^n = \left\{ \begin{bmatrix} 0 \\ X_2 \end{bmatrix} X_2 \in \mathbb{F}^{(n-m)\times 1}, \;\; 0 = 0_{m\times 1} \right\}.$$

The decomposition of $X \in \mathbb{F}^{n\times 1}$ into an element of $\mathbf{W}^n_m$ and an element of $\mathbf{U}^n_m$ is given by

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ X_2 \end{bmatrix}.$$

The corollary says that any direct sum decomposition is isomorphic to a standard one: If $\mathbf{V} = \mathbf{W} \oplus \mathbf{U}$, then there is a frame $\mathbf{\Phi}$ for $\mathbf{V}$ with

$$\mathbf{W} = \mathbf{\Phi}(\mathbf{W}^n_m), \quad \mathbf{U} = \mathbf{\Phi}(\mathbf{U}^n_m).$$

5.2 Idempotents

Definition 5.2.1. An **idempotent** on a vector space $\mathbf{V}$ is a linear map

$$\mathbf{\Pi} : \mathbf{V} \to \mathbf{V}$$

from $\mathbf{V}$ to itself which is its own square:

$$\mathbf{\Pi} \circ \mathbf{\Pi} = \mathbf{\Pi}.$$

A square matrix $\Pi \in \mathbb{F}^{n\times n}$ is called an **idempotent** iff the corresponding matrix map is an idempotent, that is, iff $\Pi^2 = \Pi$. The word *idempotent* means *same power* and comes from the obvious fact that for an idempotent we have

$$\mathbf{\Pi}^p = \mathbf{\Pi}$$

for all positive integers p. We also call a square matrix $\Pi \in \mathbb{F}^{n\times n}$ an **idempotent** if the corresponding matrix map is an idempotent, that is, if $\Pi^2 = I_n$.

The simplest examples of idempotent matrices are square matrices in zero-one diagonal form. Thus the matrix

$$\Pi = \begin{bmatrix} I_r & 0_{r\times(n-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(n-r)} \end{bmatrix}$$

satisfies $\Pi^2 = \Pi$ so the corresponding matrix map is an idempotent. Note that

$$\Pi = D^* D$$

where

$$D = \begin{bmatrix} I_r & 0_{r\times(n-r)} \end{bmatrix}, \quad D^* = \begin{bmatrix} I_r \\ 0_{(n-r)\times r} \end{bmatrix}.$$

Of course, if Π is an idempotent, and $P \in \mathbb{F}^{n\times n}$ is invertible, then $P\Pi P^{-1}$ is an idempotent. This is because

$$\begin{aligned}(P\Pi P^{-1})^2 &= P\Pi P^{-1}P\Pi P^{-1} \\ &= P\Pi^2 P^{-1} \\ &= P\Pi P^{-1}.\end{aligned}$$

Remark 5.2.2. A map $\mathbf{\Pi} : \mathbf{V} \to \mathbf{V}$ is an idempotent iff its range is its fixed point set, that is, iff

$$\mathcal{R}(\mathbf{\Pi}) = \{\mathbf{w} \in \mathbf{V} : \mathbf{\Pi}(\mathbf{w}) = \mathbf{w}\}.$$

Indeed, this equation clearly implies that $\mathbf{\Pi}^2(\mathbf{v}) = \mathbf{\Pi}(\mathbf{v})$ for $\mathbf{v} \in \mathbf{V}$ since $\mathbf{w} = \mathbf{\Pi}(\mathbf{v}) \in \mathcal{R}(\mathbf{\Pi})$. Conversely, any fixed point is clearly in the range: if $\mathbf{w} = \mathbf{\Pi}(\mathbf{w})$, then $\mathbf{w} \in \mathcal{R}(\mathbf{\Pi})$, and, if $\mathbf{\Pi}^2 = \mathbf{\Pi}$, then any vector $\mathbf{w} = \mathbf{\Pi}(\mathbf{v}) \in \mathcal{R}(\mathbf{\Pi})$ in the range is a fixed point.

Theorem 5.2.3 (Direct Sums and Idempotents). *There is a one-one onto correspondence between the set of idempotents* $\mathbf{V}$ *and the set of direct sum decompositions of* $\mathbf{V}$. $\mathbf{V} = \mathbf{W} \oplus \mathbf{U}$. *The idempotent* $\mathbf{\Pi}$ *and the direct sum decomposition* $\mathbf{V} = \mathbf{W} \oplus \mathbf{U}$ **correspond** *iff*

$$\begin{aligned}\mathbf{W} &= \mathcal{R}(\mathbf{\Pi}), \\ \mathbf{U} &= \mathcal{N}(\mathbf{\Pi}),\end{aligned}$$

that is, $\mathbf{W}$ *and* $\mathbf{U}$ *are range and null space of* $\mathbf{\Pi}$ *respectively.*

Proof. Exercise. Do Exercise 5.8.2 first.

Question 5.2.4. What is the idempotent corresponding to the direct sum decomposition in the example (with $\mathbf{V} = \mathbb{F}^{2\times 1}$) after Remark 5.1.1? (Answer: The matrix (map determined by) $\Pi = \begin{bmatrix} 1 & -b \\ 0 & 0 \end{bmatrix}$.)

Proposition 5.2.5. *Suppose* $\mathbf{V} = \mathbf{W} \oplus \mathbf{U}$ *and let* $\mathbf{\Pi}$ *be the corresponding idempotent:*

$$\mathbf{W} = \mathcal{R}(\mathbf{\Pi}), \quad \mathbf{U} = \mathcal{N}(\mathbf{\Pi}).$$

Then $\mathbf{I} - \mathbf{\Pi}$ *is an idempotent and the corresponding direct sum decomposition is* $\mathbf{V} = \mathbf{U} \oplus \mathbf{W}$*:*

$$\mathbf{U} = \mathcal{R}(\mathbf{I} - \mathbf{\Pi}), \quad \mathbf{W} = \mathcal{N}(\mathbf{I} - \mathbf{\Pi}).$$

Here $\mathbf{I} = \mathbf{I_V}$ *is the identity map of* $\mathbf{V}$.

Proof. Note that

$$(\mathbf{I} - \mathbf{\Pi}) \circ \Pi = \mathbf{\Pi} - \mathbf{\Pi}^2 = \mathbf{\Pi} - \mathbf{\Pi} = \mathbf{0}$$

so

$$(\mathbf{I} - \mathbf{\Pi})^2 = (\mathbf{I} - \mathbf{\Pi})$$

which show that $\mathbf{I} - \mathbf{\Pi}$ is an idempotent. For the rest note that

$$\begin{aligned} \mathbf{w} \in \mathcal{R}(\mathbf{\Pi}) &\iff \mathbf{\Pi}(\mathbf{w}) = \mathbf{w} \\ &\iff (\mathbf{I} - \mathbf{\Pi})(\mathbf{w}) = \mathbf{0} \\ &\iff \mathbf{w} \in \mathcal{N}(\mathbf{I} - \mathbf{\Pi}) \end{aligned}$$

so that $\mathcal{R}(\mathbf{\Pi}) = \mathcal{N}(\mathbf{I} - \mathbf{\Pi})$ and similarly (reading $\mathbf{I} - \mathbf{\Pi}$ for $\mathbf{\Pi}$) $\mathcal{R}(\mathbf{I} - \mathbf{\Pi}) = \mathcal{N}(\pi)$. QED

Two idempotents $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ of $\mathbf{V}$ are called **disjoint** iff $\mathbf{\Pi}_1 \circ \mathbf{\Pi}_2 = \mathbf{\Pi}_2 \circ \mathbf{\Pi}_1 = \mathbf{0}$. A **splitting** of $\mathbf{V}$ is a sequence of pairwise disjoint idempotents of $\mathbf{V}$ which sum to the identity. Thus a given sequence $\mathbf{\Pi}_1, \mathbf{\Pi}_2, \ldots, \mathbf{\Pi}_m$ of linear maps from $\mathbf{V}$ to itself is a splitting iff it satisfies

(1) $\mathbf{I} = \mathbf{\Pi}_1 + \mathbf{\Pi}_2 + \cdots + \mathbf{\Pi}_m$,

(2) $\mathbf{\Pi}_i \circ \mathbf{\Pi}_j = \mathbf{0}$ for $i \neq j$,

(3) $\mathbf{\Pi}_i^2 = \mathbf{\Pi}_i$ for $i = 1, 2, \ldots, m$.

where $\mathbf{I} = \mathbf{I}_{\mathbf{V}}$ the identity map of $\mathbf{V}$.

Theorem 5.2.6 (Decompositions and Splittings). *There is a one-one onto correspondence between direct sum decompositions and splittings. The direct sum decomposition* $\mathbf{V} = \bigoplus_{i=1}^m \mathbf{V}_i$ *and the splitting* $\mathbf{I} = \sum_{i=1}^m \mathbf{\Pi}_i$ **correspond** *iff*

$$\mathbf{V}_i = \mathcal{R}(\mathbf{\Pi}_i)$$

for $i = 1, 2, \ldots, m$.

Proof. Three things are asserted.

(i) If $\mathbf{I} = \sum_i \mathbf{\Pi}_i$ is a splitting and $\mathbf{V}_i = \mathcal{R}(\mathbf{\Pi}_i)$, then $\mathbf{V} = \bigoplus_{i=1}^m \mathbf{V}_i$.

(ii) Every direct sum decomposition arises this way.

(iii) If $\mathbf{I} = \sum_i \mathbf{\Pi}_i^{(1)}$ and $\mathbf{I} = \sum_i \mathbf{\Pi}_i^{(2)}$ are splittings and $\mathcal{R}(\mathbf{\Pi}_i^{(1)}) = \mathcal{R}(\mathbf{\Pi}_i^{(2)})$ for $i = 1, 2, \ldots, m$, then $\mathbf{\Pi}_i^{(1)} = \mathbf{\Pi}_i^{(2)}$ for $i = 1, 2, \ldots, m$,

Proof of (i). We show that any $\mathbf{v} \in \mathbf{V}$ has a unique decomposition

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_m \tag{$\heartsuit$}$$

with $\mathbf{v}_i \in \mathcal{R}(\Pi_i)$. Condition (1) gives the existence of this decomposition: we simply define $\mathbf{v}_i = \mathbf{\Pi}_i(\mathbf{v})$. Conditions (2) and (3) gives the uniqueness of the decomposition. To see this, apply $\mathbf{\Pi}_i$ to ($\heartsuit$). We obtain

$$\mathbf{\Pi}_i(\mathbf{v}) = \mathbf{\Pi}_i(\mathbf{v}_i)$$

by (2) and hence

$$\mathbf{\Pi}_i(\mathbf{v}) = \mathbf{v}_i$$

by (3).

Proof of (ii). Define

$$\mathbf{\Pi}_i(\mathbf{v}) = \mathbf{v}_i$$

where $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ are defined by ($\heartsuit$). The maps $\mathbf{\Pi}_i$ are well-defined since the decomposition ($\heartsuit$) is unique. The reader can check that the maps $\mathbf{\Pi}_i$ are linear and satisfy conditions (1)-(3).

Proof of (iii). If the decomposition $\mathbf{V} = \bigoplus_i \mathbf{V}_i$ and the splitting $\mathbf{I} = \sum_i \mathbf{\Pi}_i$ correspond, then

$$\begin{aligned} \mathbf{\Pi}_i(\mathbf{v}) &= \mathbf{v} \text{ for } \mathbf{v} \in \mathbf{V}_i \\ &= \mathbf{0} \text{ for } \mathbf{v} \in \mathbf{V}_j,\ i \neq j. \end{aligned}$$

These conditions determine $\mathbf{\Pi}_i$ uniquely since ever $\mathbf{v} \in \mathbf{V}$ is a sum of elements in the various $\mathbf{V}_j$. QED

A sequence of square matrices of the same size, say $n \times n$, is called a **splitting** of I_n iff the corresponding sequence of matrix maps is a splitting of $\mathbb{F}^{n\times 1}$. Thus the sequence $(\Pi_1, \Pi_2, \ldots, \Pi_m)$ is a splitting of I_n iff $\Pi_i \in \mathbb{F}^{n\times n}$ for $i = 1, 2, \ldots, m$ and

(1) $I = \Pi_1 + \Pi_2 + \cdots + \Pi_m$,

(2) $\Pi_i \Pi_j = 0$ for $i \neq j$,

(3) $\Pi_i^2 = \Pi_i$ for $i = 1, 2, \ldots, m$.

where $I = I_n$ the $n \times n$ identity matrix.

It is easy to make examples. For any sequence

$$\nu = (n_1, n_2, \ldots, n_m)$$

of positive integers which sums to n:

$$n_1 + n_2 + \cdots + n_m = n,$$

we define the **standard splitting** of I_n determined by ν by the equations

$$\begin{aligned} \text{entry}_{jj}(\Pi_i) &= 1 \text{ for } s_{i-1} < j \leq s_i \\ &= 0 \text{ for } j \leq s_{i-1} \text{ or } s_i < j \\ \text{entry}_{kj}(\Pi_i) &= 0 \text{ for } k \neq j \end{aligned}$$

where

$$s_i = n_1 + n_2 + \cdots + n_i$$

(with $s_0 = 0$). For example, with $n = 8$, $m = 4$, and $\nu = (3, 2, 2, 1)$ we have

$$\begin{aligned} \Pi_1 &= \text{diag}(1, 1, 1, 0, 0, 0, 0, 0) \\ \Pi_2 &= \text{diag}(0, 0, 0, 1, 1, 0, 0, 0) \\ \Pi_3 &= \text{diag}(0, 0, 0, 0, 0, 1, 1, 0) \\ \Pi_4 &= \text{diag}(0, 0, 0, 0, 0, 0, 0, 1) \end{aligned}$$

There are many other splittings of I_n besides the standard ones: given one splitting we can make another via

$$I_n = Q\Pi_1 Q^{-1} + Q\Pi_2 Q^{-1} + \cdots + Q\Pi_m Q^{-1}.$$

5.3 Invariant Decomposition

Let $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ be a linear map from a vector space to itself. A subspace $\mathbf{W} \subseteq \mathbf{V}$ is called **T-invariant** iff $\mathbf{T}(\mathbf{W}) \subseteq \mathbf{W}$. A direct sum decomposition $\mathbf{V} = \sum_{i=1}^m \mathbf{V}_i$ is called **T-invariant** iff each of the summands $\mathbf{V}_i$ is **T**-invariant, that is, iff

$$\mathbf{T}(\mathbf{V}_i) \subseteq \mathbf{V}_i$$

for $i = 1, 2, \ldots, m$. A splitting $\mathbf{I} = \sum_{i=1}^m \mathbf{\Pi}_i$ is called **T-invariant** iff the corresponding direct sum decomposition is.

Proposition 5.3.1 (Invariance Theorem). *Let* $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ *be a linear map from a vector space* $\mathbf{V}$ *to itself. Then a splitting*

$$\mathbf{I} = \mathbf{\Pi}_1 + \mathbf{\Pi}_2 + \cdots \mathbf{\Pi}_m$$

is $\mathbf{T}$*-invariant if and only if* $\mathbf{T}$ *commutes with each of the summands:*

$$\mathbf{T} \circ \mathbf{\Pi}_i = \mathbf{\Pi}_i \circ \mathbf{T}$$

for $i = 1, 2, \ldots, m$.

Proof. Assume the commutation equations; we prove that $\mathbf{T}(\mathbf{V}_i) \subseteq \mathbf{V}_i$. We need the fact that

$$\mathbf{v} \in \mathbf{V}_i \iff \mathbf{\Pi}_i(\mathbf{v}) = \mathbf{v}.$$

Choose $\mathbf{w} \in \mathbf{V}_i$. Then

$$\mathbf{\Pi}(\mathbf{T}(\mathbf{w})) = \mathbf{T}(\mathbf{\Pi}_i(\mathbf{w})) = \mathbf{T}(\mathbf{w}).$$

This shows that $\mathbf{T}(\mathbf{w}) \in \mathbf{V}_i$ as required. The converse is just as easy. If $\mathbf{T}(\mathbf{V}_i) \subseteq \mathbf{V}_i$, then certainly

$$\mathbf{T} \circ \mathbf{\Pi}_i(\mathbf{v}) = \mathbf{\Pi}_i \circ \mathbf{T}(\mathbf{v})$$

for $\mathbf{v} \in \mathbf{V}_i$ since both sides equal $\mathbf{T}(\mathbf{v})$. Similarly this holds for $\mathbf{v} \in \mathbf{V}_j$ with $j \neq i$ since then both sides are $\mathbf{0}$. This means that it must hold for all $\mathbf{v} \in \mathbf{V}$ since every $\mathbf{v}$ is a sum $\mathbf{v} = \mathbf{w}_1 + \mathbf{w}_2 + \cdots + ww_m$ where the formula is true for $\mathbf{v} = \mathbf{w}_i$. QED

Example 5.3.2. Let $\mathbf{V} = \mathbb{F}^{2\times 1}$ and

$$\mathbf{V}_1 = \left\{ \begin{bmatrix} x_1 \\ 0 \end{bmatrix} : x_1 \in \mathbb{F} \right\}, \quad \mathbf{V}_2 = \left\{ \begin{bmatrix} 0 \\ x_2 \end{bmatrix} : x_2 \in \mathbb{F} \right\},$$

and $\mathbf{T}(X) = AX$ the matrix map corresponding to the matrix $A \in \mathbb{F}^{2\times 2}$ given by

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

we have that $\mathbf{V}_1$ is $\mathbf{T}$-invariant iff $a_{21} = 0$ and the decomposition $\mathbf{V} = \mathbf{V}_1 \oplus \mathbf{V}_2$ is $\mathbf{T}$-invariant iff $a_{12} = a_{21} = 0$. The splitting corresponding to this direct sum decomposition is given by (the matrix maps determined by) the matrices

$$\Pi_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \Pi_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Note that

$$\Pi_1 A = \begin{bmatrix} a_{11} & a_{12} \\ 0 & 0 \end{bmatrix}, \quad A\Pi_1 = \begin{bmatrix} a_{11} & 0 \\ A_{21} & 0 \end{bmatrix},$$

so that $\Pi_1 A = A\Pi_1$ iff $a_{12} = a_{21} = 0$.

5.4 Block Diagonalization

An invariant direct sum decomposition should be viewed as a generalization of diagonalization. We now explain this point. Let

$$\mathbf{V} = \mathbf{V}_1 \oplus \mathbf{V}_2 \oplus \cdots \oplus \mathbf{V}_m$$

be a direct sum of the vector space $\mathbf{V}$. Given any linear maps

$$\mathbf{T}_i : \mathbf{V}_i \to \mathbf{V}_i$$

from the i-th summand of a direct sum decomposition to itself, there is a unique map $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ from $\mathbf{V}$ to itself characterized by the following two properties:

(1) $\mathbf{T}(\mathbf{w}) = \mathbf{T}_i(\mathbf{w})$ for $\mathbf{w} \in \mathbf{V}_i$, $i = 1, 2, \ldots, m$;

(2) The decomposition $\mathbf{V} = \bigoplus_{i=1}^m \mathbf{V}_i$ is $\mathbf{T}$-invariant.

We express these conditions with the formula:

$$\mathbf{T} = \mathbf{T}_1 \oplus \mathbf{T}_2 \oplus \cdots \mathbf{T}_m.$$

This formula establishes a one-one onto correspondence between two sets: the set of all linear maps $\mathbf{T}$ for which the direct sum decomposition $\mathbf{V} = \bigoplus_i \mathbf{V}_i$ is $\mathbf{T}$-invariant and the set of all sequences $(\mathbf{T}_1, \mathbf{T}_2, \ldots, \mathbf{T}_m)$ of linear maps with $\mathbf{T}_i : \mathbf{V}_i \to \mathbf{V}_i$ for $i = 1, 2, \ldots, m$. We call $\mathbf{T}_i$ the **restriction** of $\mathbf{T}$ to the invariant summand $\mathbf{V}_i$.

Here is a similar notation for matrices. If $A_i \in \mathbb{F}^{n_i \times n_i}$ for $i = 1, 2, \ldots, m$ and $n = n_1 + n_2 + \cdots + n_m$, then the notation

$$A = \operatorname{diag}(A_1, A_2, \ldots, A_m)$$

means that $A \in \mathbb{F}^{n\times n}$ is the **block diagonal** matrix

$$A = \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_m \end{bmatrix}$$

with the indicated blocks on the diagonal. (The blank entries denote 0.) Thus, for example, if

$$A_1 = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad A_2 \begin{bmatrix} e \end{bmatrix},$$

then

$$\operatorname{diag}(A_1, A_2) = \begin{bmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & e \end{bmatrix}.$$

The relation between these concepts is given by

Theorem 5.4.1 (Block Representation). *Assume that a direct sum decomposition is $\mathbf{T}$-invariant. Then the matrix representing $\mathbf{T}$ in any basis which respects this decomposition is block diagonal.*

Proof. The assertion that the basis $(\phi_1, \phi_2, \dots, \phi_n)$ respects the direct sum decomposition means that for each i the subsequence

$$(\phi_{s_{i-1}+1}, \phi_{s_{i-1}+2}, \dots, \phi_{s_i}) \qquad (\clubsuit_i)$$

is a basis for the summand $\mathbf{V}_i$. For $s_{i-1} < k \le s_i$ we have $\phi_k \in \mathbf{V}_i$. Hence $\mathbf{T}(\phi_k) \in \mathbf{V}_i$ by $\mathbf{T}$-invariance. Since $(\clubsuit_i)$ is a basis for $\mathbf{V}_i$ we obtain

$$\mathbf{T}(\phi_k) = \sum_{j=s_{i-1}+1}^{s_i} a_{jk}\phi_j \qquad (\#)$$

where $a_{jk} = \text{entry}_{jk}(A)$ and A represents $\mathbf{T}$ in the basis $(\phi_1, \phi_2, \dots, \phi_n)$. The equations ($\sharp$) show that A is block diagonal since they assert that $\text{entry}_{jk}(A) = 0$ unless j and k lie in the same block of integers $s_{i-1}+1, s_{i-1}+2, \dots, s_i$. Note that A_i represents the linear map $\mathbf{T}_i : \mathbf{V}_i \to \mathbf{V}_i$ in the basis $(\clubsuit_i)$. QED

5.5 Eigenspaces

Let $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ be a linear map from a vector space $\mathbf{V}$ to itself. For each $\lambda \in \mathbb{F}$ let $\mathcal{E}_\lambda(\mathbf{T})$ be the subspace of $\mathbf{V}$ defined by

$$\mathcal{E}_\lambda(\mathbf{T}) = \{\phi \in \mathbf{V} : \mathbf{T}(\phi) = \lambda\phi\}.$$

This is the null space of $\mathbf{T} - \lambda\mathbf{I}$:

$$\mathcal{E}_\lambda(\mathbf{T}) = \mathcal{N}(\mathbf{T} - \lambda\mathbf{I}),$$

where $\mathbf{I} = \mathbf{I}_\mathbf{V}$ is the identity map of $\mathbf{V}$. As in Section 4.5.4 λ is an **eigenvalue** of $\mathbf{T}$ iff $\mathcal{E}_\lambda(\mathbf{T}) \neq \{\mathbf{0}\}$ and the elements of $\mathcal{E}_\lambda(\mathbf{T})$ are the **eigenvectors** of $\mathbf{T}$ for this eigenvalue. We also call $\mathcal{E}_\lambda(\mathbf{T})$ the **eigenspace** of $\mathbf{T}$ for the eigenvalue λ.

Proposition 5.5.1. The eigenspaces are $\mathbf{T}$-invariant.

Proof. $\phi \in \mathcal{E}_\lambda(\mathbf{T}) \implies \mathbf{T}(\phi) = \lambda\phi \implies \mathbf{T}^2(\phi) = \lambda\mathbf{T}(\phi) \implies \mathbf{T}(\phi) \in \mathcal{E}_\lambda(\mathbf{T})$. QED

Theorem 5.5.2 (Eigenspace Decomposition). *The map* $\mathbf{T}$ *is diagonalizable iff*

$$\mathbf{V} = \bigoplus_\lambda \mathcal{E}_\lambda(\mathbf{T})$$

where the direct sum is over all eigenvalues λ *of* $\mathbf{T}$.

Proof. Recall (see Definition 4.5.12) that a linear map $\mathbf{T}$ is called **diagonalizable** iff there is a basis $(\phi_1, \phi_2, \ldots, \phi_n)$ consisting of eigenvectors of $\mathbf{T}$. Suppose that $\mu_1, \mu_2, \ldots, \mu_m$ are the distinct eigenvalues of $\mathbf{T}$ and that the indexing is chosen so that

$$\mathbf{T}(\phi_j) = \mu_i\phi_j \text{ for } s_{i-1} < j \leq s_i.$$

Then

$$\mathcal{E}_{\mu_i}(\mathbf{T}) = \mathrm{Span}\left(\phi_{s_{i-1}+1}, \phi_{s_{i-1}+2}, \ldots, \phi_{s_i}\right)$$

which shows both that

$$\mathbf{V} = \mathcal{E}_{\mu_1}(\mathbf{T}) \oplus \mathcal{E}_{\mu_2}(\mathbf{T}) \oplus \cdots \oplus \mathcal{E}_{\mu_m}(\mathbf{T})$$

(as required) and that the basis $(\phi_1, \phi_2, \dots, \phi_n)$ respects this direct sum decomposition as in Theorem 5.4.1. Conversely, if this eigenspace decomposition is valid, then any basis which respects this decomposition will consist of eigenvectors of $\mathbf{T}$. In particular, $\mathbf{T}$ will be diagonalizable. QED

Corollary 5.5.3. *Suppose that* $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ *is diagonalizable. Then*

$$\mathbf{T} = \sum_{i=1}^{m} \mu_i \mathbf{\Pi}_i$$

where $\mu_1, \mu_2, \dots, \mu_m$ *are the distinct eigenvalues of* $\mathbf{T}$ *and*

$$\mathbf{I} = \sum_{i=1}^{m} \mathbf{\Pi}_i$$

is the splitting corresponding to the direct sum decomposition

$$\mathbf{V} = \sum_{i=1}^{m} \mathcal{E}_{\mu_i}(\mathbf{T}).$$

5.6 Generalized Eigenspaces

Let $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ be a linear map from a vector space $\mathbf{V}$ to itself. For each $\lambda \in \mathbb{F}$ define a subspace

$$\mathcal{G}_\lambda(\mathbf{T}) = \mathcal{N}((\mathbf{T} - \lambda\mathbf{I})^n).$$

Here n is the dimension of $\mathbf{V}$ and $\mathbf{I} = \mathbf{I}_\mathbf{V}$ is the identity map of $\mathbf{V}$. The space $\mathcal{G}_\lambda(\mathbf{T})$ is called the **generalized eigenspace** of $\mathbf{T}$ for the eigenvalue λ and its elements are called **generalized eigenvectors**.

Our first step is to show that the integer n in the definition of $\mathcal{G}_\lambda(\mathbf{T})$ may be replaced by any integer $p \geq \dim(\mathbf{V})$ without affecting the definition. We need the following

Lemma 5.6.1. *Let* $\mathbf{N} : \mathbf{V} \to \mathbf{V}$ *be a linear map and* $\mathbf{v} \in \mathbf{V}$*. Suppose that* p *is a positive integer with*

$$\mathbf{N}^p(\mathbf{v}) = \mathbf{0}, \quad \mathbf{N}^{p-1}(\mathbf{v}) \neq \mathbf{0}.$$

Then $p \leq n$*.*

Proof. By the Dimension Theorem it is enough to show that the sequence of iterates

$$(\mathbf{v}, \mathbf{N}(\mathbf{v}), \mathbf{N}^2(\mathbf{v}), \ldots, \mathbf{N}^{p-1}(\mathbf{v}))$$

is independent. Suppose that the numbers $c_0, c_1, c_2, \ldots, c_{p-1}$ satisfy

$$c_0\mathbf{v} + c_1\mathbf{N}(\mathbf{v}) + c_2\mathbf{N}^2(\mathbf{v}) + \cdots + c_{p-1}\mathbf{N}^{p-1}(\mathbf{v}) = \mathbf{0}; \tag{1}$$

we must show that $c_0 = c_1 = c_2 = \cdots = c_{p-1} = 0$. Apply $\mathbf{N}^{p-1}$ to (1) gives $c_0\mathbf{N}^{p-1}(\mathbf{v})$
0 from which we conclude that $c_0 = 0$ so that (1) simplifies to

$$c_1\mathbf{N}(\mathbf{v}) + c_2\mathbf{N}^2(\mathbf{v}) + \cdots + c_{p-1}\mathbf{N}^{p-1}(\mathbf{v}) = \mathbf{0}. \tag{2}$$

Now we repeat the argument. Applying $\mathbf{N}^{p-2}$ to (2) gives $x_1 = 0$ and so on. QED

Corollary 5.6.2. *If* $(\mathbf{T} - \lambda\mathbf{I})^p(\mathbf{v}) = \mathbf{0}$ *for some positive integer* p*, then* $\mathbf{v}$ *is a generalized eigenvector.*

Proof. Take $\mathbf{N} = \mathbf{T} - \lambda\mathbf{I}$ in the lemma. QED

Proposition 5.6.3. *The generalized eigenspaces are* $\mathbf{T}$*-invariant.*

Proof. The equation

$$\mathbf{T} \circ (\mathbf{T} - \lambda\mathbf{I})^n(\phi) = (\mathbf{T} - \lambda\mathbf{I}))^n(\mathbf{T}(\phi))$$

implies that

$$\phi \in \mathcal{G}_\lambda(\mathbf{T}) \implies \mathbf{T}(\phi) \in \mathcal{G}_\lambda(\mathbf{T}).$$

QED

Note that an ordinary eigenvector is a generalized eigenvector:

$$(\mathbf{T} - \lambda\mathbf{I})(\phi) = \mathbf{0} \implies (\mathbf{T} - \lambda\mathbf{I})^n(\phi) = \mathbf{0}.$$

(Here $\implies$ means *implies.*) The converse is not true. For example, if $\mathbf{V} = \mathbb{F}^{2\times 1}$ and $\mathbf{T}$ is the matrix map corresponding to the matrix

$$L = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$$

then λ is the only eigenvalue of $\mathbf{T}$, the eigenspace is given by

$$\mathcal{E}_\lambda(\mathbf{T}) = \left\{ \left[\begin{array}{c} x \\ 0 \end{array} \right] : x \in \mathbb{F} \right\}$$

whereas every vector is a generalized vector

$$\mathcal{G}_\lambda(\mathbf{T}) = \mathbb{F}^{2\times 1}$$

since

$$(L - \lambda I)^2 = \left[\begin{array}{cc} 0 & 1 \\ 0 & 0 \end{array} \right]^2 = 0.$$

There is however no distinction between eigenvalues and generalized eigenvalues.

Theorem 5.6.4. *The number λ is an eigenvalue for $\mathbf{T}$ iff the corresponding generalized eigenspace $\mathcal{G}_\lambda(\mathbf{T})$ is not the zero space:*

$$\mathcal{E}_\lambda(\mathbf{T}) \neq \{\mathbf{0}\} \iff \mathcal{G}_\lambda(\mathbf{T}) \neq \{\mathbf{0}\}.$$

Proof. One direction is easy since $\mathcal{E}_\lambda(\mathbf{T}) \subseteq \mathcal{G}_\lambda(\mathbf{T})$. For the converse suppose $\phi \in \mathcal{G}_\lambda(\mathbf{T})$ is non-zero. Then

$$\begin{array}{rcl} (\mathbf{T} - \lambda)^k(\phi) & = & \mathbf{0} \text{ for } k = n, \text{ but} \\ & \neq & \mathbf{0} \text{ for } k = 0, \end{array}$$

so there is a largest value of k with $\psi = (\mathbf{T} - \lambda)^{k-1}(\phi) \neq \mathbf{0}$. Then

$$(\mathbf{T} - \lambda)\psi = (\mathbf{T} - \lambda)^k(\phi) = \mathbf{0}$$

so $\psi \in \mathcal{E}_\lambda(\mathbf{T})$ and hence $\mathcal{E}_\lambda(\mathbf{T}) \neq \{\mathbf{0}\}$ as required. QED

Corollary 5.6.5. *The only eigenvalue of the linear map*

$$\mathcal{G}_\lambda(\mathbf{T}) \to \mathcal{G}_\lambda(\mathbf{T}) : \mathbf{v} \mapsto \mathbf{T}(\mathbf{v})$$

is λ.

Proof. Suppose that $\phi \in \mathcal{G}_\lambda(\mathbf{T})$ satisfies $\mathbf{T}(\phi) = \mu\phi$. Then $\psi = (\mathbf{T} - \lambda\mathbf{I})^{k-1}(\phi)$ (from the last proof) also satisfies $\mathbf{T}(\psi) = \mu\psi$. But the last proof showed that $\mathbf{T}(\psi) = \lambda\psi$ and $\psi \neq \mathbf{0}$ so $\lambda = \mu$. QED

Question 5.6.6. Show that

$$\mathcal{G}_\lambda(\mathbf{T}) \cap \mathcal{G}_\mu(\mathbf{T}) = \{\mathbf{0}\}$$

for $\lambda \neq \mu$. (Answer: Otherwise (as in the proof) the intersection would contain an eigenvector for $\mathbf{T}$. The corresponding eigenvalue would be both λ and μ which is impossible.)

Theorem 5.6.7 (Generalized Eigenspace Decomposition). *Assume*

$$\mathbb{F} = \mathbb{C}$$

the field of complex numbers. Then any linear map

$$\mathbf{T} : \mathbf{V} \to \mathbf{V}$$

has an $\mathbf{T}$*-invariant direct sum decomposition*

$$\mathbf{V} = \bigoplus_\lambda \mathcal{G}_\lambda(\mathbf{T})$$

where the direct sum is over all eigenvalues λ *of* $\mathbf{T}$*.*

This theorem is an improvement over the Eigenspace Decomposition of of Theorem 5.5.2 in that it works for *any* linear map, not just diagonalizable ones. We have already proved in Proposition 5.6.3 that the decomposition is $\mathbf{T}$-invariant. We shall postpone the rest of the proof to the next section. For the moment we recast this theorem in the language of matrix theory.

Theorem 5.6.8 (Block Diagonalization). *Any matrix* $A \in \mathbb{C}^{n\times n}$ *is similar to a block diagonal matrix where each of the blocks has a single eigenvalue. More precisely, suppose* $\mu_1, \mu_2, \ldots, \mu_m$ *are the distinct eigenvalues of* A*. Then there is an invertible matrix* $P \in \mathbb{C}^{n\times n}$ *such that*

$$P^{-1}AP = \operatorname{diag}(B_1, B_2, \ldots, B_m)$$

where the matrix $B_i - \mu_i I$ *is nilpotent for* $i = 1, 2, \ldots, m$*.*

Proof. We deduce this as a corollary of the Generalized Eigenspace Decomposition. We take $\mathbf{V} = \mathbb{C}^{n\times 1}$ and $\mathbf{T} = \mathbf{A}$ the matrix map determined by A. Choose any basis $(P_1, P_2, \ldots, P_n)$ which respects the Generalized Eigenspace Decomposition, that is,

$$\mathcal{N}((A - \mu_i I)^n) = \operatorname{Span}\left(P_{s_{i-1}+1}, P_{s_{i-1}+2}, \ldots, P_{s_i}\right)$$

where $0 = s_0 < s_1 < \cdots < s_m = n$. Define P by $\mathrm{col}_j(P) = P_j$. Then $P^{-1}AP$ is the matrix representing $\mathbf{T} = \mathbf{A}$ in the basis $(P_1, P_2, \ldots, P_n)$. By Theorem 5.4.1, this matrix is block diagonal. Since B_i is the matrix representing the restriction to the Generalized Eigenspace $\mathcal{N}((A - \mu_i I)^n)$ it follows that $B_i - \mu_i I$ is nilpotent. QED

Remark 5.6.9. We deduced the Block Diagonalization Theorem from the Generalized Eigenspace Decomposition but it is just as easy to do the reverse. Let A represent $\mathbf{T}$ in any basis. By the Block Diagonalization Theorem A is similar to $P^{-1}AP$ which is in block diagonal form. By Theorem 4.3.4 there is a basis for $\mathbf{V}$ so that the matrix $P^{-1}AP$ represents the map $\mathbf{T}$ in this basis. The elements of this basis are the generalized eigenvectors required by Generalized Eigenspace Decomposition. We omit the details.

5.7 Minimal Polynomial

Let $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ be a linear map from a finite-dimensional vector space $\mathbf{V}$ to itself. The space $\mathcal{L}(\mathbf{V}, \mathbf{V})$ of all linear maps from $\mathbf{V}$ to itself is a vector space of dimension n^2 where n is the dimension of $\mathbf{V}$. Hence for some $m \le n^2$ the sequence

$$(\mathbf{I}, \mathbf{T}, \mathbf{T}^2, \mathbf{T}^3, \ldots, \mathbf{T}^m)$$

of powers of $\mathbf{T}$ must be dependent. Thus there are numbers $c_0, c_1, c_2, \ldots, c_m$, not all zero, such that

$$c_0\mathbf{I} + c_1\mathbf{T} + c_2\mathbf{T}^2 + \cdots + c_m\mathbf{T}^m = 0. \qquad (\sharp)$$

Take the smallest value of m for which the system $(\sharp)$ has a non-trivial solution and form the polynomial

$$f(t) = c_0 + c_1 t + c_2 t^2 + \cdots + c_m t^m.$$

Then equation $(\sharp)$ can be written as

$$f(\mathbf{T}) = 0.$$

Notice that since m is smallest we must have $c_m \neq 0$ (else a smaller value of m would work) so we can divide through by it and assume that $c_m = 1$. The resulting polynomial is called the **minimal polynomial** for $\mathbf{T}$. Since m is smallest, it follows that $g(\mathbf{T}) \neq 0$ for any non-zero polynomial of degree less than m.

Theorem 5.7.1 (Minimal Polynomial Theorem). *Assume that $\mathbb{F} = \mathbb{C}$. Then the eigenvalues of $\mathbf{T}$ are the roots of the minimal polynomial f of $\mathbf{T}$*

Proof. Choose any number λ. Divide the polynomial $f(t)$ by the polynomial $t - \lambda$ to obtain a quotient $g(t)$ of degree $m - 1$:

$$f(t) = (t - \lambda)g(t) + c$$

Here c is a number (that is, a polynomial of degree zero). Note that $c = 0$ iff $f(\lambda) = 0$, that is, iff λ is a root of f.

First assume that $f(\lambda) = 0$. Then $c = 0$ so when we substitute $\mathbf{T}$ for t we get Substitute $\mathbf{T}$ for t:

$$\mathbf{0} = f(\mathbf{T}) = (\mathbf{T} - \lambda\mathbf{I})g(\mathbf{T}).$$

As $g(t)$ has smaller degree than $f(t)$ we have that $g(\mathbf{T}) \neq 0$. Hence there is a $\mathbf{w} \in \mathbf{V}$ with $g(\mathbf{T})(\mathbf{w}) \neq \mathbf{0}$. Let $\mathbf{v} = g(\mathbf{T})(\mathbf{w})$. Then

$$\mathbf{0} = f(\mathbf{T})\mathbf{w} = (\mathbf{T} - \lambda\mathbf{I})g(\mathbf{T})(\mathbf{w}) = (\mathbf{T} - \lambda\mathbf{I})(\mathbf{v})$$

which shows that λ is an eigenvalue of $\mathbf{T}$ with eigenvector $\mathbf{v}$.

Conversely assume that λ is an eigenvalue for $\mathbf{T}$. Then there is a non-zero $\mathbf{v} \in \mathbf{V}$ with $(\mathbf{T} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$. Hence

$$\mathbf{0} = f(\mathbf{T})(\mathbf{v}) = g(\mathbf{T})(\mathbf{T} - \lambda\mathbf{I})(\mathbf{v}) + c\mathbf{v} = \mathbf{0} + c\mathbf{v} = c\mathbf{v}$$

so $c = 0$ and hence $f(\lambda) = 0$ as required. QED

Corollary 5.7.2 (Eigenvalues Exist). *Assume that $\mathbb{F} = \mathbb{C}$. Then any linear map $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ has an eigenvector.*

Proof. By the Fundamental Theorem of Algebra any complex polynomial has a complex root. QED

Corollary 5.7.3. *The minimal polynomial f of $\mathbf{T}$ has the form*

$$f(t) = (t - \mu_1)^{p_1}(t - \mu_2)^{p_2} \cdots (t - \mu_m)^{p_m}$$

where $\mu_1, \mu_2, \ldots, \mu_m$ be the distinct eigenvalues of $\mathbf{T}$ and the exponents p_k are positive integers.

Proof of Theorem 5.6.7. We now prove the Generalized Eigenspace Decomposition Theorem. Assume that $\mathbb{F} = \mathbb{C}$, that $\mathbf{V}$ is a finite dimensional vector space, and that $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ is a linear map from $\mathbf{V}$ to itself. Let $\mu_1, \mu_2, \ldots, \mu_m$ be the distinct eigenvalues of $\mathbf{T}$ and denote by $\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_m$ the corresponding generalized eigenspaces:

$$\mathbf{V}_k = \mathcal{G}_{\mu_k}(\mathbf{T})$$

for $k = 1, 2, \ldots, m$.

Let $f_k(t)$ be the minimal polynomial of the linear map

$$\mathbf{V}_k \to \mathbf{V}_k : \mathbf{v} \mapsto \mathbf{T}(\mathbf{v}). \qquad (\natural)$$

Let $g_k(t) = \prod_{j \neq k} f_j(t)$ be the product of all the $f_j(t)$ with $j \neq k$:

$$g_k(t) = f_1(t) \cdots f_{k-1}(t) f_{k+1}(t) \cdots f_m(t).$$

Lemma 5.7.4. *The map*

$$\mathbf{V}_k \to \mathbf{V}_k : \mathbf{v} \mapsto g_k(\mathbf{T})(\mathbf{v})$$

is an isomorphism, but

$$g_k(\mathbf{T})(\mathbf{v}) = \mathbf{0} \textit{ for } \mathbf{v} \in \mathbf{V}_j \textit{ with } j \neq k.$$

Proof. In the last section we noted that the only eigenvalue of this map is μ_k so f_k must have the form

$$f_k(t) = (t - \mu_k)^{p_k}.$$

For $j \neq k$ the map

$$\mathbf{V}_k \to \mathbf{V}_k : \mathbf{v} \mapsto (\mathbf{T} - \mu_j \mathbf{I})(\mathbf{v})$$

is an isomorphism, else μ_j would be an eigenvalue for the map ($\natural$). If we raise this map to the p_j-th power and then multiply the results together for $j \neq k$ we obtain the first part of the lemma. (a composition of isomorphism is an isomorphism). The second part of the lemma is trivial, since $f_j(\mathbf{T})(\mathbf{v}) = \mathbf{0}$ for $\mathbf{v} \in \mathbf{V}_j$ and $f_j(t)$ is a factor of $g_k(t)$. QED

We resume the proof of Theorem 5.6.7. Let

$$\mathbf{W} = \mathbf{V}_1 + \mathbf{V}_2 + \cdots + \mathbf{V}_m$$

be the sum of all these spaces $\mathbf{V}_k$; that is, $\mathbf{w} \in \mathbf{W}$ if and only if there exist vectors $\mathbf{v}_k \in \mathbf{V}_k$ with

$$\mathbf{w} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_m.$$

We must show two things:

$$\mathbf{W} = \mathbf{V}_1 \oplus \mathbf{V}_2 \oplus \cdots \oplus \mathbf{V}_m \tag{1}$$

and

$$\mathbf{W} = \mathbf{V}. \tag{2}$$

We prove (1). Suppose that

$$\mathbf{0} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_m$$

where $\mathbf{v}_k \in \mathbf{V}_k$. Apply $g_k(\mathbf{T})$ to both sides. By the second part of the lemma $\mathbf{0} = g_k(\mathbf{T})(\mathbf{v}_k)$. Hence $\mathbf{v}_k = \mathbf{0}$ by the first part of the lemma.

We prove (2). Assume (2) is false, that is, that $\mathbf{W} \neq \mathbf{V}$. Choose any complement $\mathbf{U}$ to $\mathbf{W}$ in $\mathbf{V}$,

$$\mathbf{V} = \mathbf{W} \oplus \mathbf{U}$$

and let $\iota : \mathbf{U} \to \mathbf{V}$ denote the inclusion and $\pi : \mathbf{V} \to \mathbf{U}$ the projection onto $\mathbf{U}$ along $\mathbf{W}$, i.e.

$$\iota(\mathbf{u}) = \mathbf{u}, \qquad \pi(\mathbf{u} + \mathbf{w}) = \mathbf{w}$$

for $\mathbf{u} \in \mathbf{U}$ and $\mathbf{w} \in \mathbf{W}$. Let λ be an eigenvalue for

$$\pi \circ \mathbf{T} \circ \iota : \mathbf{U} \to \mathbf{U}$$

and let $\mathbf{u} \in \mathbf{U}$ be the corresponding eigenvector. Then

$$\begin{aligned} \pi \circ \mathbf{T} \circ \iota(\mathbf{u}) &= \lambda \mathbf{u} \qquad \text{so} \\ \pi(\mathbf{T}(\mathbf{u}) - \lambda \mathbf{u}) &= \mathbf{0} \qquad \text{so} \\ \mathbf{T}(\mathbf{u}) - \lambda \mathbf{u} \in \mathcal{N}(\pi) &= \mathbf{W} \end{aligned}$$

where we have used $\iota(\mathbf{u}) = \pi(\mathbf{u}) = \mathbf{u}$ which follows from $\mathbf{u} \in \mathbf{U}$. From the definition of $\mathbf{W}$ we obtain

$$\mathbf{T}(\mathbf{u}) - \lambda \mathbf{u} = \mathbf{w}_1 + \mathbf{w}_k + \cdots + \mathbf{w}_m \tag{3}$$

where $\mathbf{w}_k \in \mathbf{V}_k$.

We distinguish two cases. In case λ is not an eigenvalue then the linear map

$$\mathbf{V}_k \to \mathbf{V}_k : \mathbf{v} \mapsto (\mathbf{T} - \lambda\mathbf{I})(\mathbf{v})$$

is invertible for each $k = 1, 2, \ldots, m$ so we may choose $\mathbf{v}_k \in \mathbf{V}_k$ satisfying

$$(\mathbf{T} - \lambda\mathbf{I})(\mathbf{v}_k) = \mathbf{w}_k \tag{4}$$

so (3) may be written as

$$(\mathbf{T} - \lambda\mathbf{I})(\mathbf{u} - \mathbf{v}_1 - \mathbf{v}_2 - \cdots - \mathbf{v}_m) = \mathbf{0}.$$

As λ is not an eigenvalue, $(\mathbf{T} - \lambda\mathbf{I})$ is invertible so we may cancel it in the last equation and obtain

$$\mathbf{u} - \mathbf{v}_1 - \mathbf{v}_2 - \cdots - \mathbf{v}_m = \mathbf{0}.$$

But $\mathbf{u} \neq \mathbf{0}$ so this contradicts

$$\mathbf{V} = \mathbf{U} \oplus \mathbf{W} = \mathbf{U} \oplus \mathbf{V}_1 \oplus \mathbf{V}_2 \oplus \cdots \oplus \mathbf{V}_m. \tag{5}$$

The second case is that λ *is* an eigenvalue of $\mathbf{T}$, say $\lambda = \mu_1$. We may still find $\mathbf{v}_k \in \mathbf{V}_k$ satisfying (4) for $k = 2, 3, \ldots, m$ so we may write (3) as

$$(\mathbf{T} - \lambda\mathbf{I})(\mathbf{u} - \mathbf{v}_2 - \cdots - \mathbf{v}_m) = \mathbf{w}_1.$$

As $\mathbf{w}_1 \in \mathbf{V}_1$ we obtain

$$(\mathbf{T} - \lambda\mathbf{I})^p(\mathbf{u} - \mathbf{v}_2 - \cdots - \mathbf{v}_m) = \mathbf{0}$$

for sufficiently large p and hence that

$$\mathbf{u} - \mathbf{v}_2 - \cdots - \mathbf{v}_m \in \mathbf{V}_1.$$

But this also contradicts (5). QED

5.8 Exercises

Exercise 5.8.1. Suppose that $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ is an isomorphism and that $\mathbf{V} = \mathbf{V}_1 \oplus \mathbf{V}_2$. Show that $\mathbf{W} = \mathbf{W}_1 \oplus \mathbf{W}_2$ where $\mathbf{W}_1 = \mathbf{T}(\mathbf{V}_1)$ and $\mathbf{W}_2 = \mathbf{T}(\mathbf{V}_2)$.

Exercise 5.8.2. Given two vector spaces $\mathbf{W}$ and $\mathbf{U}$, the **direct product** $\mathbf{W} \times \mathbf{U}$ of $\mathbf{W}$ and $\mathbf{U}$ is the set of all pairs $(\mathbf{w}, \mathbf{u})$ with $\mathbf{w} \in \mathbf{W}$ and $\mathbf{u} \in \mathbf{U}$:

$$\mathbf{W} \times \mathbf{U} = \{(\mathbf{w}, \mathbf{u}) : \mathbf{w} \in \mathbf{W},\ \mathbf{u} \in \mathbf{U}\}.$$

We make $\mathbf{W} \times \mathbf{U}$ into a vector space by defining the vector space operations via the following rules:

$$\begin{aligned}
(\mathbf{w}_1, \mathbf{u}_1) + (\mathbf{w}_2, \mathbf{u}_2) &= (\mathbf{w}_1 + \mathbf{w}_2, \mathbf{u}_1 + \mathbf{u}_2) \\
(a\mathbf{w}, \mathbf{u}) &= (a\mathbf{w}, a\mathbf{u}) \\
\mathbf{0}_{\mathbf{W}\times\mathbf{U}} &= (\mathbf{0}_{\mathbf{W}}, \mathbf{0}_{\mathbf{U}}).
\end{aligned}$$

Suppose that $\mathbf{W}$ and $\mathbf{U}$ are subspaces of $\mathbf{W}$. Show that $\mathbf{V} = \mathbf{W} \oplus \mathbf{U}$ if and only if the linear map

$$\mathbf{W} \times \mathbf{U} \to \mathbf{V} : (\mathbf{w}, \mathbf{u}) \mapsto \mathbf{w} + \mathbf{u}$$

is an isomorphism.

Exercise 5.8.3. Let $\mathbf{W}$ and $\mathbf{U}$ be subspaces of a vector space $\mathbf{V}$. Define the **sum** $\mathbf{W} + \mathbf{U}$ and **intersection** $\mathbf{W} \cap \mathbf{U}$ of $\mathbf{W}$ and $\mathbf{U}$ by

$$\begin{aligned}
\mathbf{W} + \mathbf{U} &= \{\mathbf{w} + \mathbf{u} : \mathbf{w} \in \mathbf{W},\ \mathbf{U} \in \mathbf{U}\} \\
\mathbf{W} \cap \mathbf{U} &= \{\mathbf{v} \in \mathbf{V} : \mathbf{v} \in \mathbf{W} \text{ and } \mathbf{v} \in \mathbf{U}\}.
\end{aligned}$$

Show that

(1) $\mathbf{W} + \mathbf{U}$ and $\mathbf{W} \cap \mathbf{U}$ are subspaces of $\mathbf{V}$.

(2) $\mathbf{W} + \mathbf{U} = \mathbf{W} \oplus \mathbf{U}$ iff $\mathbf{W} \cap \mathbf{U} = \{\mathbf{0}\}$.

(3) $\dim(\mathbf{W} + \mathbf{U}) + \dim(\mathbf{W} \cap \mathbf{U}) = \dim(\mathbf{W}) + \dim(\mathbf{U})$.

Exercise 5.8.4. Let $A, B \in \mathbb{F}^{2\times 4}$ be defined by

$$\begin{aligned}
A &= \begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{bmatrix} \\
B &= \begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{bmatrix}
\end{aligned}$$

Let $\mathbf{V} = \mathbb{F}^{4\times 1}$. Find $\mathbf{W}+\mathbf{U}$ and $\mathbf{W}\cap\mathbf{U}$ if $\mathbf{W} = \mathcal{N}(A)$ and $\mathbf{U} = \mathcal{N}(B)$. (Here $\mathcal{N}$ denotes *null space*.)

Exercise 5.8.5. Suppose that $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ is a linear map and that $\mathbf{S} : \mathbf{W} \to \mathbf{V}$ is a right inverse to $\mathbf{T}$:

$$\mathbf{T}\circ\mathbf{S} = \mathbf{I}_{\mathbf{W}}.$$

Show that $\mathbf{S}\circ\mathbf{T}$ is an idempotent on $\mathbf{V}$ and that the corresponding direct sum decomposition is given by

$$\mathbf{V} = \mathbf{W}\oplus\mathbf{U}$$

where

$$\begin{aligned} \mathbf{W} &= \mathcal{R}(\mathbf{S}\circ\mathbf{T}) = \mathcal{R}(\mathbf{S}) \\ \mathbf{U} &= \mathcal{N}(\mathbf{S}\circ\mathbf{T}) = \mathcal{N}(\mathbf{T}). \end{aligned}$$

Exercise 5.8.6. For any subset $K \subseteq \{1, 2, \ldots, n\}$ define a matrix $I_{n,K} \in \mathbb{F}^{n\times n}$ by

$$I_{n,K} = \operatorname{diag}(e_1, e_2, \ldots, e_n)$$

where

$$e_j = \begin{cases} 1 & \text{if } j \in K, \\ 0 & \text{if } j \notin K. \end{cases}$$

For example

$$I_{n,K} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

when $n = 3$ and $K = \{1, 3\}$.

(1) Show that $I_{n,K}$ is an idempotent.

(2) Show that the rank of $I_{n,K}$ is the cardinality of K.

(3) Show that $I_{n,K}I_{n,H} = I_{n,K\cap H}$.

(4) Show that $I_{n,K}$ and $I_{n,H}$ are disjoint idempotents iff H and K are disjoint sets, that is, $H \cap K = \emptyset$.

(5) Prove that $I_{n,K\cup H} + I_{n,K\cap H} = I_{n,K} + I_{n,H}$.

Chapter 6

Jordan Normal Form

In this chapter we will find a complete system of invariants that characterize similarity. This means a collection of nonnegative integers $\rho_{\lambda,k}(A)$ – defined for each square matrix A, each positive integer k, and each complex number λ – such that for $A, B \in \mathbb{C}^{n\times n}$, we have that A and B are similar if and only if

$$\rho_{\lambda,k}(A) = \rho_{\lambda,k}(B) \qquad \text{for all } \lambda \in \mathbb{C} \text{ and all } k = 1, 2, \ldots.$$

We will prove a normal form theorem for similarity called the *Jordan Normal Form Theorem.*

6.1 Similarity Invariants

Definition 6.1.1. Let $A \in \mathbb{C}^{n\times n}$, $\lambda \in \mathbb{C}$, and $k = 1, 2, 3, \ldots$. Define

$$\rho_{\lambda,k}(A) = \operatorname{rank}(\lambda I - A)^k$$

where $I = I_n$ is the $n \times n$ identity matrix. The integer $\rho_{\lambda,k}(A)$ is called the kth **eigenrank** of A for the eigenvalue λ.

Remark 6.1.2. If λ is not an eigenvalue of A, then $\rho_{\lambda,k}(A) = n$. If $k \geq n$, $\rho_{\lambda,k}(A) = \rho_{\lambda,n}(A)$. (See Exercise 6.1.8 below.) Thus only finitely many of these numbers are of interest.

Definition 6.1.3. The **eigennullities** $\nu_{\lambda,k}(A)$ of the matrix A are defined by

$$\nu_{\lambda,k}(A) = \operatorname{nullity}((\lambda I - A)^k) = \dim \mathcal{N}((\lambda I - A)^k)$$

From the Rank Nullity Relation 3.13.2 (*rank + nullity = n)*, we obtain

$$\rho_{\lambda,k}(A) + \nu_{\lambda,k}(A) = n \tag{$*$}$$

for $A \in \mathbb{C}^{n\times n}$. Hence, the eigennullities and eigenranks contain the same information.

Remark 6.1.4. The eigennullity

$$\nu_{\lambda,1}(A) = \dim \mathcal{N}(\lambda I - A) = \dim \mathcal{E}_\lambda(A)$$

is called the **geometric multiplicity** of the eigenvalue λ. It is the dimension of the eigenspace $\mathcal{E}_\lambda(A)$. The eigennullity

$$\nu_{\lambda,n}(A) = \dim \mathcal{N}(\lambda I - A)^n = \dim \mathcal{G}_\lambda(A)$$

is called the **algebraic multiplicity** of λ for A. It is the dimension of the generalized eigenspace $\mathcal{G}_\lambda(A)$. For a diagonalizable matrix these two multiplicities are the same.

Theorem 6.1.5 (Invariance). *Similar matrices have the same eigenranks.*

Proof. There are three key points: (1) Similar matrices are *a fortiori* equivalent (see Exercise 4.6.26), for if $A = PBP^{-1}$, then $A = QBP^{-1}$ where $Q = P$. (2) Similar matrices have similar powers, for $(PBP^{-1})^k = PB^kP^{-1}$. (3) If A and B are similar so are $\lambda I - A$ and $\lambda I - B$ since $P(\lambda I - B)P^{-1} = \lambda I - PBP^{-1}$.

Now assume that A and B are similar. Then $A = PBP^{-1}$ where P is invertible. Choose $\lambda \in \mathbb{C}$. Then $\lambda I - A = P(\lambda I - B)P^{-1}$. Hence, $(\lambda I - A)^k = P(\lambda I - B^kP^{-1}$ for $k = 1, 2, \ldots$. By Exercise 4.6.26, the matrices $(\lambda I - A)^k$ and $(\lambda I - B)^k$ have the same rank. By the definition of $\rho_{\lambda,k}$, we have $\rho_{\lambda,k}(A) = \rho_{\lambda,k}(B)$, as required. QED

Remark 6.1.6. Of course, by equation $(*)$ of Definition 6.1.3, similar matrices have the same eigennullities as well. Below (Corollary 6.9.4), we will prove the converse to Theorem 6.1.5.

Exercise 6.1.7. Prove that a matrix $A \in \mathbb{C}^{n\times n}$ is diagonalizable if and only if $\rho_{\lambda,k}(A) = \rho_{\lambda,1}(A)$ for all eigenvalues λ of A and all $k = 1, 2, 3, \ldots$.

Exercise 6.1.8. Prove that $\rho_{\lambda,k}(A) = \rho_{\lambda,n}(A)$ if $k \geq n$.

6.2 Jordan Normal Form

We can improve the Block Diagonalization Theorem 5.6.8 considerably by making further similarity transformations within each block. The resulting blocks will be almost diagonal except for a few nonzero entries above the diagonal. Here are the precise definitions.

The entries $\text{entry}_{ii}(A)$ of a matrix A are called the **diagonal entries**, and said to be *on the diagonal*. The entries $\text{entry}_{i,i+1}(A)$ are called the **superdiagonal** entries, and said to lie on the *on the superdiagonal*. The superdiagonal entries lie just above the diagonal. A **Jordan block** is a square matrix Λ having all its diagonal entries equal, zeros or ones on the superdiagonal, and zeros elsewhere. Thus Λ is a Jordan block iff

$$\begin{aligned}\text{entry}_{ii}(\Lambda) &= \lambda, \\ \text{entry}_{i,i+1}(\Lambda) &= 0 \text{ or } 1, \\ \text{entry}_{ij}(\Lambda) &= 0 \text{ if } j \neq i, i+1.\end{aligned}$$

Definition 6.2.1. *Jordan Normal Form A matrix J is in* **Jordan normal form** *iff it is in block diagonal form*

$$J = \text{diag}(\Lambda_1, \Lambda_2, \ldots, \Lambda_m)$$

where each Λ_k is a Jordan block.

Example 6.2.2. The 6×6 matrix

$$J = \begin{bmatrix} \lambda_1 & e_1 & 0 & & & \\ 0 & \lambda_1 & e_2 & & & \\ 0 & 0 & \lambda_1 & & & \\ & & & \lambda_2 & & \\ & & & & \lambda_3 & e_3 \\ & & & & 0 & \lambda_3 \end{bmatrix}$$

is in Jordan normal form provided that each of the superdiagonal entries e_1, e_2, e_3 is either zero or one.

Theorem 6.2.3 (Jordan Normal Form). *Every square matrix A is similar to a matrix J in Jordan normal form.*

In other words, any square matrix A may be written in the form

$$A = PJP^{-1}$$

where P is invertible and J is in Jordan normal form. By the Block Diagonalization Theorem 5.6.8, we can assume that the matrix A is block diagonal. We can work a block at a time, so it is enough to prove the theorem for matrices with only one eigenvalue. As the matrices $\lambda I + V_1$ and $\lambda I + V_2$ are similar if and only if the matrices V_1 and V_2 are, it is enough to prove the theorem for nilpotent (in fact, strictly upper triangular) matrices. The proof will occupy most of the rest of this chapter.

6.3 Indecomposable Jordan Blocks

In this section we'll prove a special case of the Jordan Normal Form Theorem 6.2.3 as a warmup. The ideas in the general case are similar. We'll make a preliminary definition.

An **indecomposable Jordan block** is one where all the entries on the superdiagonal are one. It has the form $\lambda I + W$ where

$$\mathrm{entry}_{ij}(W) = \begin{cases} 1 & \text{if } j = i+1, \\ 0 & \text{otherwise.} \end{cases}$$

Notice that W is itself an indecomposable Jordan block (with eigenvalue zero). A Jordan block has form

$$\Lambda = \mathrm{diag}(\lambda I + W_1,\ \lambda I + W_2,\ \ldots,\ \lambda I + W_k)$$

where the matrices $\lambda I + W_1, \lambda I + W_2, \ldots, \lambda I + W_k$ are indecomposable Jordan blocks.[1] For example, the Jordan block

$$\Lambda = \begin{bmatrix} \lambda & 1 & 0 & & & \\ 0 & \lambda & 1 & & & \\ 0 & 0 & \lambda & & & \\ & & & \lambda & & \\ & & & & \lambda & 1 \\ & & & & 0 & \lambda \end{bmatrix}$$

has form

$$\Lambda = \mathrm{diag}(\lambda I + W_1,\ \lambda I + W_2,\ \lambda I + W_3)$$

[1]The terminology here is at slight variance with the general usage. Most authors call *Jordan block* what we have called *indecomposable Jordan block*.

where the constituent indecomposable Jordan blocks are

$$\lambda I + W_1 = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}, \quad \lambda I + W_2 = \begin{bmatrix} \lambda \end{bmatrix}, \quad \lambda I + W_3 = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}.$$

Question 6.3.1. What are the eigenranks of this last matrix Λ? (Answer: $\rho_{\mu,k}(\Lambda) = 6$ for $\mu \neq \lambda$, $\rho_{\lambda,1}(\Lambda) = 3$, $\rho_{\lambda,2}(\Lambda) = 1$, and $\rho_{\lambda,k}(\Lambda) = 0$ for $k > 2$.)

Theorem 6.3.2. *Let $N \in \mathbb{F}^{n\times n}$ be a matrix of size $n \times n$ and degree of nilpotence n, i.e. that $N^n = 0$ but $N^{n-1} \neq 0$. Then N is similar to the indecomposable $n \times n$ Jordan block W.*

Proof. Since $N^n = 0$ but $N^{n-1} \neq 0$, there is a vector $X \in \mathbb{F}^{n\times 1}$ such that $N^n X = 0$ but $N^{n-1}X \neq 0$. Form the matrix P whose jth column is $N^{n-j}X$. We will prove that

$$NP = PW.$$

Then we will show that P is invertible. Multiplying on the right by P^{-1} gives

$$N = PWP^{-1}.$$

We prove that $NP = PW$, i.e. that

$$\mathrm{col}_j(NP) = \mathrm{col}_j(PW)$$

for $j = 1, 2, \ldots, n$. By the definition of P,

$$P = \begin{bmatrix} N^{n-1}X & N^{n-2}X & \cdots & NX & X \end{bmatrix},$$

so

$$NP = \begin{bmatrix} 0 & N^{n-1}X & \cdots & N^2X & NX \end{bmatrix} = PW,$$

so

$$\begin{aligned} \mathrm{col}_1(NP) &= 0, \\ \mathrm{col}_j(NP) &= \mathrm{col}_{j-1}(P) \text{ for } j = 2, 3, \ldots, n. \end{aligned}$$

On the other hand, the first column of W is zero, and the jth column of W is the $(j-1)$st column of the identity matrix. Thus

$$\begin{aligned} \mathrm{col}_1(PW) &= 0, \\ \mathrm{col}_j(PW) &= \mathrm{col}_{j-1}(P) \text{ for } j = 2, 3, \ldots, n. \end{aligned}$$

This proves that $NP = PW$.

We prove that P is invertible. It is enough to show that its columns are independent. Suppose

$$0 = c_1 N^{n-1} X + c_2 N^{n-2} X + \cdots + c_{n-1} N X + c_n X.$$

Since $N^k = 0$ for $k \geq n$ we may apply N^{n-1} to both sides and obtain that $c_n N^{n-1} X = 0$. But $N^{n-1} X \neq 0$ so $c_n = 0$. Now apply N^{n-2} to both sides to prove that $c_{n-1} = 0$. Repeating in this way we obtain that $c_1 = c_2 = \cdots = c_n = 0$, as required. QED

6.4 Partitions

A little terminology from number theory is useful in describing the relations among the various eigennullities of a nilpotent matrix.

A **partition** of a positive integer n is a nonincreasing sequence π of positive integers which sum to n, that is,

$$\pi = (n_1, n_2, \ldots, n_m)$$

where

$$n_1 \geq n_2 \geq \cdots \geq n_m \geq 1$$

and

$$n_1 + n_2 + \cdots + n_m = n.$$

A partition $\pi = (n_1, n_2, \ldots, n_m)$ can be used to construct a diagram of stars called a **tableau**. The tableau consists of $n = n_1 + n_2 + \cdots + n_m$ stars arranged in m rows with the kth row having n_k stars. The stars in a row are left justified so that the jth columns align. The jth column of the tableau intersects the kth row exactly when $j \leq n_k$. The **dual partition** π^* of π is obtained by forming the transpose of this tableau. Thus $\pi^* = (\ell_1, \ell_2, \ldots, \ell_p)$ where ℓ_j is the number of indices k with $j \leq n_k$. For example, if

$$\pi = (5, 5, 4, 3, 3, 3, 1),$$

then the tableau is

$$\begin{array}{ccccc} \star & \star & \star & \star & \star \\ \star & \star & \star & \star & \star \\ \star & \star & \star & \star & \\ \star & \star & \star & & \\ \star & \star & \star & & \\ \star & \star & \star & & \\ \star & & & & \end{array}$$

and the dual partition

$$\pi^* = (7, 6, 6, 3, 2)$$

is obtained by counting the number of entries in successive columns. The dual of the dual is the original partition:

$$\pi^{**} = \pi.$$

6.5 Weyr Characteristic

Let $N \in \mathbb{F}^{n\times n}$ be a nilpotent matrix, and let p be the **degree of nilpotence** of N. This is the least integer for which $N^p = 0$:

$$N^p = 0, \qquad N^{p-1} \neq 0.$$

Recall that the kth **eigenrank** of N is the integer

$$\rho_k(N) = \operatorname{rank}(N^k) = \dim \mathcal{R}(N^k).$$

We have dropped the subscript λ since N is nilpotent: its only eigenvalue is zero. The sequence of integers $\omega = (\ell_1, \ell_2, \ldots, \ell_p)$ defined by

$$\ell_k = \rho_{k-1}(N) - \rho_k(N)$$

for $k = 1, 2, \ldots, p$ is called the **Weyr characteristic** of of the nilpotent matrix N.

Theorem 6.5.1. *The Weyr characteristic of a nilpotent matrix $N \in \mathbb{F}^{n\times n}$ is a partition of n.*

Proof. Successive terms ℓ_k and ℓ_{k+1} in the sum $\ell_1 + \cdots + \ell_p$ contain $\rho_k(N)$ with opposite signs. Hence, the sum "telescopes":

$$\ell_1 + \ell_2 + \cdots + \ell_p = \rho_0(N) - \rho_p(N) = n - 0 = n$$

as $N^0 = I$ and $N^p = 0$. To show that $\ell_k \geq \ell_{k+1}$, first note the obvious inclusion of ranges

$$\mathcal{R}(N^k) \subseteq \mathcal{R}(N^{k-1}).$$

This holds because $N^k X = N^{k-1}(NX)$. Let Φ be a frame for the subspace $\mathcal{R}(N^k)$, and extend it to a frame Ψ for $\mathcal{R}(N^{k-1})$ by adjoining additional columns Υ:

$$\Psi = \begin{bmatrix} \Phi & \Upsilon \end{bmatrix}.$$

Then Ψ has $\rho_{k-1}(N)$ columns, Φ has $\rho_k(N)$ columns, and Υ has ℓ_k columns. Now

$$\mathcal{R}(N^{k-1}) = \mathcal{R}(\Psi), \qquad \mathcal{R}(\mathbb{N}^k) = \mathcal{R}(\Phi),$$

so

$$\mathcal{R}(N^k) = \mathcal{R}(N\Psi), \qquad \mathcal{R}(N^{k+1}) = \mathcal{R}(N\Phi).$$

Discard some columns from $N\Phi$ to make a basis $\tilde{\Phi}$ for $\mathcal{R}(N^{k+1})$, and then discard some columns from

$$N\Psi = \begin{bmatrix} N\Phi & N\Upsilon \end{bmatrix},$$

so that

$$\tilde{\Psi} = \begin{bmatrix} \tilde{\Phi} & \tilde{\Upsilon} \end{bmatrix}$$

is a basis for $\mathcal{R}(N^k)$. Then $\tilde{\Upsilon}$ has ℓ_{k+1} columns. Since the discarded columns were taken from Υ, it follows that $\ell_{k+1} \leq \ell_k$, as required. QED

6.6 Segre Characteristic

For each k let $W_k \in \mathbb{F}^{k\times k}$ denote the $k \times k$ indecomposable Jordan block with eigenvalue zero:

$$\begin{aligned} \text{entry}_{ij}(W_k) &= 0 \quad \text{if } j \neq i+1 \\ \text{entry}_{i,i+1}(W_k) &= 1 \quad \text{for } i = 1, 2, \ldots, k-1. \end{aligned}$$

The subscript k indicates the size of the matrix W_k. For each partition

$$\pi = (n_1, n_2, \ldots, n_m)$$

denote by W_π the Jordan block given by the block diagonal matrix

$$W_\pi = \mathrm{diag}(W_{n_1}, W_{n_2}, \ldots W_{n_m}).$$

A matrix of form W_π is called a **Segre matrix**.

A Segre matrix is in Jordan normal form. Conversely, any nilpotent matrix in Jordan normal form can be transformed to a Segre matrix by permuting the blocks so that their sizes decrease along the diagonal. (This can be accomplished by replacing J by PJP^{-1} for a suitable permutation matrix P.)

Now define the **Segre characteristic** of a nilpotent matrix to be the dual partition

$$\pi = \omega^*$$

of the Weyr characteristic ω. The key to understanding the Jordan Normal Form Theorem is the following

Theorem 6.6.1. *The Segre characteristic of the Segre matrix W_π is π.*

An example is more convincing than a general proof. Let

$$\pi = (3, 2, 2, 1), \qquad \omega = \pi^* = (4, 3, 1),$$

so

$$W = W_\pi = \mathrm{diag}(W_3, W_2, W_2, W_1).$$

Written in full this is

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

(The blank entries represent 0; they have been omitted to make the block structure more evident.) In the notation of the definition

$$\pi = (n_1, n_2, n_3, n_4), \qquad \omega = (\ell_1, \ell_2, \ell_3),$$

where $n_1 = 3$, $n_2 = n_3 = 2$, $n_4 = 1$, $\ell_1 = 4$, $\ell_2 = 3$, $\ell_3 = 1$ and

$$n = n_1 + n_2 + n_3 + n_4 = \ell_1 + \ell_2 + \ell_3 = 8.$$

For $j = 1, 2, \ldots, 8$ let $E_j = \mathrm{col}_j(I_8)$ denote the jth column of the 8×8 identity matrix so that

$$WE_j = \begin{cases} 0 & \text{for } j = 1, 4, 6, 8; \\ E_{j-1} & \text{for } j = 2, 3, 5, 7. \end{cases}$$

Arrange these columns in a tableau

$$\begin{array}{lll} E_1 & E_2 & E_3 \\ E_4 & E_5 & \\ E_6 & E_7 & \\ E_8 & & \end{array}$$

so that n_i is the number of entries in the ith row of the tableau and ℓ_j be the number of entries in the jth column. We can decorate the tableau with arrows to indicate the effect of applying W:

$$\begin{array}{ccccccc} 0 & \leftarrow & E_1 & \leftarrow & E_2 & \leftarrow & E_3 \\ 0 & \leftarrow & E_4 & \leftarrow & E_5 & & \\ 0 & \leftarrow & E_6 & \leftarrow & E_7 & & \\ 0 & \leftarrow & E_8 & & & & \end{array}$$

We now see a general principle:

> *Applying W^k to the tableau annihilates the elements in the first k columns and transforms the remaining elements into the columns of a basis for $\mathcal{R}(N^k)$.*

Thus the kth eigenrank is the number

$$\rho_k(W) = \ell_{k+1} + \ell_{k+2} + \cdots + \ell_p$$

of elements to the right of the kth column. This equation says precisely that $\omega = (\ell_1, \ell_2, \ldots, \ell_p)$ is the Weyr characteristic of $W = W_\pi$, as required.

6.7 Jordan-Segre Basis

Continue the notation of the last section. Let π be a partition of n and W_π be the corresponding Segre matrix. For $j = 1, 2, \cdots, n$ let

$$E_j = \mathrm{col}_j(I_n)$$

denote the jth column of the identity matrix I_n. Then $W_\pi E_j$ is either E_{j-1} or 0 depending on π. We'll use a double subscript notation to specify for which values of j the former alternative holds. Let

$$E_{1,1}, \ldots, E_{1,n_1}, E_{2,1}, \ldots, E_{2,n_2}, \ldots$$

denote the columns $E_1, E_2, \ldots, E_n$ in that order. Then

$$W_\pi E_{i,1} = 0 \qquad \text{for } i = 1, 2, \ldots, m,$$

$$W_\pi E_{i,j} = E_{i,j-1} \quad \text{for } j = 2, 3, \ldots, n_i.$$

These relations say that the doubly indexed sequence E_{ij} forms a *Jordan-Segre Basis* for $(\mathbb{F}^{n\times 1}, W_\pi)$. Here's the definition.

Let $N \in \mathbb{F}^{n\times n}$ be a matrix and $\mathbf{V} \subseteq \mathbb{F}^{n\times 1}$ be a subspace. A **Jordan-Segre Basis** for $(\mathbf{V}, N)$ $N \in \mathbb{F}^{n\times n}$ is a doubly indexed sequence

$$X_{i,j} \in \mathbf{V}, \qquad (i = 1, 2, \ldots, m; \;\; j = 1, 2, \ldots, n_i)$$

of columns which forms a basis for $\mathbf{V}$ and satisfies

$$NX_{i,1} = 0 \qquad \text{for } i = 1, 2, \ldots, m,$$

$$NX_{i,j} = X_{i,j-1} \quad \text{for } j = 2, 3, \ldots, n_i.$$

The sequence $\pi = (n_1, n_2, \ldots, n_m)$ is called the **associated partition** of the basis; it is a partition of the dimension of $\mathbf{V}$:

$$\dim(\mathbf{V}) = n_1 + n_2 + \cdots + n_m.$$

The condition that the elements $X_{i,j} \in \mathbf{V}$ form a basis for $\mathbf{V}$ means that every $X \in \mathbf{V}$ may be written uniquely as a linear combination of these $X_{i,j}$, in other words, that the inhomogeneous system

$$X = \sum_{i=1}^{m} \sum_{j=1}^{n_m} c_{ij} X_{ij}$$

(in which the $c_{i,j}$ are the unknowns) has a unique solution. Throughout most of these notes we would have said instead that the matrix formed from these columns is a *basis* for $\mathbf{V}$, but the present terminology is more conventional. The matrix whose columns are

$$X_{1,1}, \ldots, X_{1,n_1}, X_{2,1}, \ldots, X_{2,n_2}, \ldots, X_{n,n_m}$$

(in that order) is called the **basis corresponding** to the Jordan-Segre basis. In case $\mathbf{V} = \mathbb{F}^{n\times 1}$, this is an invertible matrix.

Theorem 6.7.1. *Suppose that $P \in \mathbb{F}^{n\times n}$ is the basis (matrix) corresponding to a Jordan-Segre basis for $(\mathbb{F}^{n\times 1}, N)$. Then*

$$N = PW_\pi P^{-1}$$

where π is the associated partition.

Proof. Since P is invertible, the conclusion may be written as $NP = PW_\pi$. Let $E_{i,j}$ be the kth column of the identity matrix where $X_{i,j}$ be the kth column of P. Then

$$\mathrm{col}_k(NP) = N\mathrm{col}_k(P) = NX_{i,j} = \begin{cases} X_{i,j-1} = \mathrm{col}_{k-1}(P) & \text{if } j > 1, \\ 0 & \text{if } j = 1, \end{cases}$$

while

$$\mathrm{col}_k(PW_\pi) = P\mathrm{col}_k(I) = \begin{cases} PE_{i,j-1} = \mathrm{col}_{k-1}(P) & \text{if } j > 1, \\ 0 & \text{if } j = 1, \end{cases}$$

so

$$\mathrm{col}_k(NP) = \mathrm{col}_k(PW_\pi).$$

As k is arbitrary this shows that

$$NP = PW_\pi,$$

as required. QED

6.8 Improved Rank Nullity Relation

The Rank Nullity Relation 3.13.2 says that for $A \in \mathbb{F}^{m\times n}$ we have

$$\operatorname{rank}(A) + \operatorname{nullity}(A) = n.$$

For the proof of the Jordan Normal Form Theorem, we'll need a slight generalization.

Lemma 6.8.1. *Suppose that* $\mathbf{V} \subseteq \mathbb{F}^{n\times 1}$ *is a subspace and that* $A \in \mathbb{F}^{m\times n}$*. Then*

$$\dim(A\mathbf{V}) + \dim\big(\mathbf{V} \cap \mathcal{N}(A)\big) = \dim(\mathbf{V})$$

where

$$A\mathbf{V} = \{AX \in \mathbb{F}^{m\times 1} : X \in \mathbf{V}\}$$

and

$$\mathbf{V} \cap \mathcal{N}(A) = \{X \in \mathbf{V} : AX = 0\}.$$

Proof. Exercise.

6.9 Proof of the Jordan Normal Form Theorem

To prove the Jordan Normal Form Theorem 6.2.3, it is enough to prove it for nilpotent matrices. For this, by Theorem 6.7.1, it is enough to prove that if N is a nilpotent matrix, there is a Jordan-Segre basis for $(\mathbb{F}^{n\times 1}, N)$. We shall prove this inductively.

Let $N \in \mathbb{F}^{n\times n}$ be nilpotent, and let p be the degree of nilpotence of N. This means that

$$N^p = 0, \qquad N^{p-1} \neq 0.$$

Let $\mathbf{V}_k$ denote the range $\mathcal{R}(N^k)$ of N^k:

$$\mathbf{V}_k = N^k\mathbb{F}^{n\times 1} = N\mathbf{V}_{k-1}.$$

Clearly, $\mathbf{V}_k \subseteq \mathbf{V}_{k-1}$. (Proof: Choose $X \in \mathbf{V}_k$. Then $X = N^kY$ for some Y, so $X = N^{k-1}Z$ where $Z = NY$, so $X \in \mathbf{V}_{k-1}$.) Hence, we have an increasing sequence

$$\{0\} = \mathbf{V}_k \subseteq \mathbf{V}_{p-1} \subseteq \cdots \subseteq \mathbf{V}_1 \subseteq \mathbf{V}_0 = \mathbb{F}^{n\times 1}$$

of subspaces of $\mathbb{F}^{n\times 1}$. The theorem follows by taking $k = 0$ in the following

Lemma 6.9.1. *There is a Jordan-Segre basis for* $(\mathbf{V}_k, N)$.

Proof. This is proved by reverse induction on k. This means that first we prove it for $k = p$, then for $k = p - 1$, then for $k = p - 2$, and so on. At the $(p - k)$th stage of the proof, we use the basis constructed for $\mathbf{V}_{k+1}$ to construct a basis for $\mathbf{V}_k$.

For $k = p$, the basis is empty, as $\mathbf{V}_k = \{0\}$. For $k = p - 1$, any basis for $\mathbf{V}_{p-1}$ is a Jordan-Segre basis, since $NX = 0$ for $X \in \mathbf{V}_{p-1}$. Now assume that we have constructed a Jordan-Segre basis

$$\begin{array}{llllll} X_{1,1} & X_{1,2} & \dots & \dots & \dots & X_{1,m_1} \\ & & \vdots & & & \\ X_{i,1} & X_{i,2} & \dots & \dots & X_{i,m_i} & \\ & & \vdots & & & \\ X_{h,1} & X_{h,2} & \dots & X_{h,m_h} & & \end{array}$$

for $(\mathbf{V}_{k+1}, N)$. We shall extend it to a Jordan-Segre basis for $(\mathbf{V}_k, N)$ by adjoining an additional element to the end of every row and (possibly) some additional elements at the bottom of the first column.

As the elements of the basis lie in $\mathbf{V}_{k+1} = N\mathbf{V}_k$, each has the form NX for some $X \in \mathbf{V}_k$. In particular, this is true for these elements on the right edge of the tableau, so there are elements $X_{i,m_i+1} \in \mathbf{V}_k$ satisfying

$$X_{i,m_i} = NX_{i,m_i+1}.$$

We adjoin this element X_{i,m_i+1} to the right end of the ith row. The elements in the first column form a basis for $\mathbf{V}_{k+1} \cap \mathcal{N}(N)$. As $\mathbf{V}_{k+1} \subseteq \mathbf{V}_k$, these elements form an independent sequence in $\mathbf{V}_k \cap \mathcal{N}(N)$. Hence, we may extend to a basis

$$X_{1,1}, X_{2,1}, \dots, X_{h,1}, X_{h+1,1}, \dots, X_{g,1}$$

for $\mathbf{V}_k \cap \mathcal{N}(N)$.

We claim that this is a Jordan-Segre basis for $(\mathbf{V}_k, N)$. The elements $NX_{i,j}$ with $j > 1$ are precisely the elements of the Jordan-Segre basis for $\mathbf{V}_{k+1} = N\mathbf{V}_k$, while the elements $\mathbf{V}_{i,1}$ form a basis for $\mathbf{V}_k \cap \mathcal{N}(N)$ by construction. Thus by the Rank Nullity Relation 3.13.2, the elements $X_{i,j}$ $(i = 1, 2, \dots, g,\ j \geq 1)$ form a basis for $\mathbf{V}_k$, as required. This completes the proof of the lemma and hence of the Jordan Normal Form Theorem 6.2.3. QED

Example 6.9.2. *Suppose that the Segre characteristic of the nilpotent matrix N is the partition $\pi = (3, 2, 2, 1)$ of the example in the proof of Theorem 6.6.1. We follow the steps in the proof of 6.9.1 to construct a Jordan-Segre basis. Note that $N^3 = 0$.*

- *Let X_1 be a basis for $\mathcal{R}(N^2)$*
- *Extend to a basis $\begin{bmatrix} X_1 & X_2 & X_4 & X_6 \end{bmatrix}$ for $\mathcal{R}(N)$ by solving the inhomogeneous system $NX_2 = X_1$ for X_2 and extending $\begin{bmatrix} X_1 \end{bmatrix}$ to a basis $\begin{bmatrix} X_1 & X_4 & X_6 \end{bmatrix}$ of $\mathcal{R}(N) \cap \mathcal{N}(N)$.*
- *Extend to a basis*

$$P = \begin{bmatrix} X_1 & X_2 & X_3 & X_4 & X_5 & X_6 & X_7 & X_8 \end{bmatrix}.$$

of $\mathbb{F}^{8\times 1}$ by solving the inhomogeneous systems

$$NX_3 = X_2, \qquad NX_5 = X_4, \qquad NX_7 = X_6,$$

for X_3, X_5, and X_7, and then extending $\begin{bmatrix} X_1 & X_4 & X_6 \end{bmatrix}$ to a basis $\begin{bmatrix} X_1 & X_4 & X_6 & X_8 \end{bmatrix}$ for $\mathcal{N}(N)$.

Theorem 6.9.3. *For two nilpotent matrices of the same size, the following conditions are equivalent:*

(1) *they are similar;*

(2) *they have the same eigenranks;*

(3) *they have the same eigennullities;*

(4) *they have the same Segre characteristic;*

(5) *they have the same Weyr characteristic.*

Proof. The eigennullities and the Weyr characteristic are related by the two equations

$$\begin{aligned} \nu_k(N) &= \ell_1 + \ell_2 + \cdots + \ell_k, \\ \ell_k &= \nu_k(N) - \nu_{k-1}(N), \end{aligned}$$

and so they determine one another. By the Rank Nullity Relation 3.13.2,

$$\nu_k(N) + \rho_k(N) = n$$

the Weyr characteristic and the eigenranks determine one another. By duality, the Weyr characteristic and the Segre characteristic determine one another. This shows that conditions (2) through (5) are equivalent. We have seen that (1) $\implies$ (2) in Theorem 6.1.5. We have proved that every nilpotent matrix is similar to some Segre matrix W_π (Theorems 6.7.1 and 6.9.1), and that the Segre characteristic of W_π is π (Theorem 6.6.1). Hence, (4) $\implies$ (1). QED

Corollary 6.9.4. *The eigenranks*

$$\rho_{\lambda,k}(A) = \operatorname{rank}(\lambda I - A)^k$$

form a complete system of invariants for similarity. This means that two square matrices $A, B \in \mathbb{F}^{n\times n}$ *are similar if and only if*

$$\rho_{\lambda,k}(A) = \rho_{\lambda,k}(B)$$

for all $\lambda \in \mathbb{C}$ *and all* $k = 1, 2, \ldots$.

Proof. We have already proved "only if" as Theorem 6.1.5. In the nilpotent case, "if" is Theorem 6.9.3, just proved. The general case follows from the nilpotent case as indicated in the discussion just after the statement of Theorem 6.2.3.

6.10 Exercises

Exercise 6.10.1. Calculate the eigenranks $\rho_{\lambda,k}(A)$ where

$$A = \begin{bmatrix} 5 & 1 & 0 & & & \\ 0 & 5 & 1 & & & \\ 0 & 0 & 5 & & & \\ & & & 7 & 0 & 0 \\ & & & 0 & 7 & 1 \\ & & & 0 & 0 & 7 \end{bmatrix}.$$

Exercise 6.10.2. A 24×24 matrix N satisfies $N^5 = 0$ and

$$\operatorname{rank}(N^4) = 2, \quad \operatorname{rank}(N^3) = 5, \quad \operatorname{rank}(N^2) = 11, \quad \operatorname{rank}(N) = 17.$$

Find its Segre characteristic.

Exercise 6.10.3. For a fixed eigenvalue λ there are 8 matrices in Jordan normal form of size 4×4 having λ as the only eigenvalue. (Each of the three entries on the superdiagonal can be either 0 or 1.) Which of these are similar? Hint: Compute the invariants $\rho_{\lambda,k}$.

Exercise 6.10.4. Show that a matrix and its transpose are similar.

Exercise 6.10.5. Suppose that N is nilpotent, that W is invertible, and that $WN = NW$. Show that N and NW are similar.

Exercise 6.10.6. Prove that if N is nilpotent, then $I+N$ and e^N are similar.

Exercise 6.10.7 (Chevalley Decomposition). *Show that a square matrix $A \in \mathbb{F}^{n\times n}$ may be written uniquely in the form*

$$A = S + N$$

where S is diagonalizable, N is nilpotent, and S and N commute. Moreover, if A is real, then so are S and N (although S might have nonreal eigenvalues and thus not be diagonalizable over $\mathbb{R}$). Hint: In the complex case we may assume that A is in Jordan Normal Form. Then S is diagonal and N is strictly triangular. Find polynomials f and g such that $S = f(A)$ and $N = g(A)$.

Chapter 7

Groups and Normal Forms

7.1 Matrix Groups

Definition 7.1.1. A **matrix group** is a set

$$G \subseteq \mathbb{F}^{n\times n}$$

of invertible matrices such that

- G contains the identity matrix: $I_n \in G$.
- G is closed under taking inverses: $P \in G \implies P^{-1} \in G$.
- G is closed under multiplication: $P, Q \in G \implies PQ \in G$.

Theorem 7.1.2. *The set of all invertible matrices in* $\mathbb{F}^{n\times n}$ *is a matrix group. (It is called the* **general linear group***.)*

Theorem 7.1.3. *The set of all matrices in* $\mathbb{F}^{n\times n}$ *of determinant one is a matrix group. (It is called the* **special linear group***.)*

Definition 7.1.4. A matrix P is called **unitary** iff its conjugate transpose is its inverse:

$$P^{\dagger} = P^{-1}.$$

Theorem 7.1.5. *The set of* all *unitary matrices in* $\mathbb{F}^{n\times n}$ *is a matrix group. (It is called the* **unitary group***.)*

Definition 7.1.6. A matrix P is called **orthogonal** iff its transpose is its inverse:

$$P^* = P^{-1}.$$

(Thus a real matrix is unitary if and only if its orthogonal.)

Theorem 7.1.7. *The set of all orthogonal matrices in* $\mathbb{F}^{n\times n}$ *is a matrix group. (It is called the* **orthogonal group***.)*

Theorem 7.1.8. *The set of all invertible diagonal matrices in* $\mathbb{F}^{n\times n}$ *is a matrix group.*

Theorem 7.1.9. *The set of all invertible triangular (see 4.5.5) matrices in* $\mathbb{F}^{n\times n}$ *is a matrix group.*

Theorem 7.1.10. *The set of all uni-triangular (see 4.6.22) matrices in* $\mathbb{F}^{n\times n}$ *is a matrix group.*

Definition 7.1.11. A matrix is called **lower triangular** iff its transpose is triangular.

Theorem 7.1.12. *The set of all invertible lower triangular matrices in* $\mathbb{F}^{n\times n}$ *is a matrix group.*

7.2 Matrix Invariants

Each of the theorems in this section has the form

> *Two matrices of the same size are "equivalent" if and only if they have the same "invariant".*

The equivalence relations involve the matrix groups of the previous section. Some of these theorems have been proved in the text or can be easily be deduced from theorems in the text and elementary matrix algebra. Theorems 7.2.16, 7.2.14, and 7.2.20 use material not explained in these notes.

Definition 7.2.1. Two matrices $A, B \in \mathbb{F}^{m\times n}$ are called **equivalent** iff there exists an invertible matrix $Q \in \mathbb{F}^{m\times m}$ and an invertible matrix $P \in \mathbb{F}^{n\times n}$ such that

$$A = QBP^{-1}.$$

Theorem 7.2.2. *Two matrices of the same size are equivalent if and only if they have the same rank.*

Definition 7.2.3. Two matrices $A, B \in \mathbb{F}^{m\times n}$ are called **left equivalent** iff there is an invertible matrix $Q \in \mathbb{F}^{m\times m}$ such that

$$A = QB.$$

Theorem 7.2.4. *Two matrices of the same size are left equivalent if and only if they have the null space.*

Definition 7.2.5. Two matrices $A, B \in \mathbb{F}^{m\times n}$ are called **right equivalent** iff there is an invertible matrix $P \in \mathbb{F}^{n\times n}$ such that

$$A = BP^{-1}.$$

Theorem 7.2.6. *Two matrices of the same size are right equivalent if and only if they have the same range.*

Definition 7.2.7. For any matrix A the rank $\delta_{pq}(A)$ of the $p \times q$ submatrix in the upper left hand corner of A is called the (p, q)th **corner rank** of A. Two matrices $A, B \in \mathbb{F}^{m\times n}$ are called **lower upper equivalent** iff there exists an invertible lower triangular matrix $Q \in \mathbb{F}^{m\times m}$ and a uni-triangular matrix $P \in \mathbb{F}^{n\times n}$ such that

$$A = QBP^{-1}.$$

Theorem 7.2.8. *Two matrices of the same size are lower upper equivalent if and only if they have the same corner ranks.*

Definition 7.2.9. Two matrices $A, B \in \mathbb{F}^{m\times n}$ are called **lower equivalent** iff $A = QB$ where $Q \in \mathbb{F}^{m\times m}$ is invertible lower triangular. Let $E_{m,k}$ denote the span of the last $k - 1$ columns of the $m \times m$ identity matrix, i.e. for $Y \in \mathbb{F}^{m\times 1}$

$$Y \in E_{m,k} \iff \text{entry}_{k+1}(Y) = \cdots = \text{entry}_m(Y) = 0.$$

Compare with 4.4.1. For $V \subseteq \mathbb{F}^{m\times 1}$ and $A \in \mathbb{F}^{m\times n}$ define

$$A^{-1}(V) = \{X \in \mathbb{F}^{n\times 1} : AX \in V\}.$$

Theorem 7.2.10. *Two matrices A and B are lower equivalent if and only if*

$$A^{-1}\big(E_{m,k}\big) = B^{-1}\big(E_{m,k}\big)$$

for $k = 0, 1, 2, \ldots, m$.

Definition 7.2.11. Two square matrices $A, B \in \mathbb{F}^{n\times n}$ are called **similar** iff there exists an invertible matrix $P \in \mathbb{F}^{n\times n}$ such that

$$A = PBP^{-1}.$$

We restate Corollary 6.9.4 so the reader can see the pattern.

Theorem 7.2.12. *Two square matrices of the same size are similar if and only if they have the same eigenranks (see 6.1.1).*

Definition 7.2.13. Two square matrices $A, B \in \mathbb{F}^{n\times n}$ are called **unitarily similar** iff there exists a unitary matrix $P \in \mathbb{F}^{n\times n}$ such that

$$A = PBP^{-1}.$$

A square matrix $A \in \mathbb{F}^{n\times n}$ is called **Hermitean** iff it is equal to its conjugate transpose:

$$A = A^\dagger.$$

Theorem 7.3.11 below states that a Hermitean matrix is diagonalizable so that for each eigenvalue the algebraic multiplicity and the geometric multiplicity are the same.

Theorem 7.2.14. *Two Hermitean matrices of the same size are unitarily similar if and only if they have the same eigenvalues each with the same multiplicity.*

Definition 7.2.15. Two matrices $A, B \in \mathbb{F}^{m\times n}$ of the same size are called **unitarily left equivalent** iff there exists a unitary matrix $Q \in \mathbb{F}^{m\times m}$ such that

$$A = QB.$$

Theorem 7.2.16. *Two matrices A and B are unitarily left equivalent if and only if $A^\dagger A = B^\dagger B$.*

Remark 7.2.17. Note that the matrices $A^\dagger A$ and $B^\dagger B$ are Hermitean.

Definition 7.2.18. Suppose $A \in \mathbb{F}^{m\times n}$ and $m \geq n$. A number σ is called **singular value** for a matrix A iff $\sigma \geq 0$, and σ^2 is an eigenvalue of the Hermitean matrix $A^\dagger A$, i.e. there is a nonzero vector $X \in \mathbb{F}^{n\times 1}$ satisfying the condition

$$A^\dagger AX = \sigma^2 X.$$

Any X satisfying this condition is called a **singular vector** of A corresponding to the singular value σ. The **multiplicity** of a singular value σ of A is the dimension

$$\dim \mathcal{E}_{\sigma^2}(A^\dagger A) = \text{nullity}(\sigma^2 I - A^\dagger A)$$

of the corresponding space of singular vectors. If $m < n$, the singular values and multiplicities of A are, by definition, the same as those of $A^\dagger$.

Definition 7.2.19. Two matrices A and B of the same size are called **unitarily equivalent** iff there exist unitary matrices $Q \in \mathbb{F}^{m\times m}$ and $P \in \mathbb{F}^{n\times n}$ such that

$$A = QBP^{-1}.$$

Theorem 7.2.20. *For two matrices A and B of the same size the following are equivalent:*

(1) *A and B are unitarily equivalent.*

(2) *$A^\dagger A$ and $B^\dagger B$ are unitarily similar.*

(3) *A and B have the same singular values each with the same multiplicity.*

7.3 Normal Forms

The theorems of this section all have the form

> *Every matrix is "equivalent" to a matrix in "normal form".*

The notion of equivalence is one of the equivalence relations of the previous section. Often (but not always) the normal form is unique. Not all the theorems in this section can be easily proved from the material in these notes.

Theorem 7.3.1 (Gauss Jordan Decomposition). *Any $A \in \mathbb{F}^{m\times n}$ may be written in the form*

$$A = QR$$

where $Q \in \mathbb{F}^{m\times m}$ is invertible and $R \in \mathbb{F}^{m\times n}$ is in reduced row echelon form. (See 4.5.3) If $A = Q'R'$ is another such decomposition, then $R = R'$.

Theorem 7.3.2. *Any matrix $A \in \mathbb{F}^{m\times n}$ may be written in the form*

$$A = TP^{-1}$$

where $P \in \mathbb{F}^{n\times n}$ is invertible and $T \in \mathbb{F}^{m\times n}$ has a reduced row echelon form. If $A = T'P'^{-1}$ is another such decomposition, then $T = T'$.

Theorem 7.3.3. *Any matrix $A \in \mathbb{F}^{m\times n}$ may be written in the form*

$$A = QDP^{-1}$$

where $Q \in \mathbb{F}^{m\times m}$ and $P \in \mathbb{F}^{n\times n}$ are invertible and $R \in \mathbb{F}^{m\times n}$ is in zero-one normal form. (See 4.5.1) If $A = Q'R'P'^{-1}$ is another such decomposition, then $D = D'$.

Definition 7.3.4. A matrix is in **rook normal form** iff all its entries are either zero or one and it has at most one nonzero entry in every row and at most one nonzero entry in every column.

Theorem 7.3.5. *Any matrix $A \in \mathbb{F}^{m\times n}$ may be written in the form*

$$A = QDP^{-1}$$

where $Q \in \mathbb{F}^{m\times m}$ is invertible lower triangular, and $P \in \mathbb{F}^{n\times n}$ is uni-triangular, and $D \in \mathbb{F}^{m\times n}$ is in rook normal form. If $A = Q'D'P'^{-1}$ is another such decomposition, then $D = D'$.

Definition 7.3.6. A matrix $R \in \mathbb{F}^{m\times n}$ is said to be in in **leading entry normal form**, iff there is a matrix $D \in \mathbb{F}^{m\times n}$ in rook normal form, such that for each pair (p, q) of indices for which $\mathrm{entry}_{pq}(D) \neq 0$ we have

$$\begin{array}{lll} \mathrm{entry}_{p,q}(R) & = 1, & \\ \mathrm{entry}_{p,j}(R) & = 0 & \text{for } j < q, \\ \mathrm{entry}_{i,q}(R) & = 0 & \text{for } p < i. \end{array}$$

For example, the 4×5 matrix

$$R = \begin{bmatrix} 0 & 0 & 1 & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 1 & * \\ 1 & * & 0 & 0 & * \end{bmatrix}$$

is in leading entry normal form.

Theorem 7.3.7. *Any matrix $A \in \mathbb{F}^{m\times n}$ may be written in the form*

$$A = LR$$

where $L \in \mathbb{F}^{m\times m}$ is invertible lower triangular and $R \in \mathbb{F}^{m\times n}$ is in leading entry normal form. If $A = L'R'$ is another such decomposition, then $R = R'$.

Theorem 7.3.8 (Jordan Normal Form). *Any matrix $A \in \mathbb{C}^{n\times n}$ may be be written in the form*

$$A = PJP^{-1}$$

where $P \in \mathbb{C}^{n\times n}$ is invertible and J is in Jordan normal form.

Remark 7.3.9. The normal form J is obviously not unique; if J is diagonal and Q is a permutation matrix, then QJQ^{-1} is again diagonal with the diagonal entries occurring in a different order.

Theorem 7.3.10 (Gram Schmidt Decomposition). *Any $A \in \mathbb{F}^{m\times n}$ with independent columns may be written in the form*

$$A = BP^{-1}$$

where $P \in \mathbb{F}^{m\times m}$ is positive triangular and $B \in \mathbb{F}^{m\times n}$ satisfies $B^\dagger B = I_n$. If $A = B'P'^{-1}$ is another such decomposition, then $B = B'$.

Theorem 7.3.11 (Spectral Theorem). *Assume $\mathbb{F} = \mathbb{C}$ or $\mathbb{F} = \mathbb{R}$. Any Hermitean matrix $A \in \mathbb{F}^{n\times n}$ may be written in the form*

$$A = PDP^{-1}$$

where $P \in \mathbb{F}^{n\times n}$ is unitary and $D \in \mathbb{R}^{n\times n}$ is real and diagonal.

Definition 7.3.12. An $m \times n$ matrix R is in **positive row echelon form** iff it is in row echelon form (see 4.5.2) and in addition all the leading entries are positive.

Theorem 7.3.13 (Householder Decomposition). *Assume $\mathbb{F} = \mathbb{C}$ or $\mathbb{R}$. Then any matrix $A \in \mathbb{F}^{m\times n}$ may be written in the form*

$$A = QR$$

where $Q \in \mathbb{F}^{m\times m}$ is unitary and $R \in \mathbb{F}^{m\times n}$ is in positive row echelon form. If $A = Q'R'$ is another such decomposition, then $R = R'$.

Definition 7.3.14. A matrix D is in **singular normal form** iff

$$D = \begin{bmatrix} \Delta & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix}$$

where

$$\Delta = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r)$$

is an $r \times r$ diagonal matrix with positive entries σ_j on the diagonal. (Note that the diagonal entries of Δ (and 0 if $r < m$) are the singular values of D.)

Theorem 7.3.15 (Singular Value Decomposition). *Assume* $\mathbb{F} = \mathbb{C}$ *or* $\mathbb{F} = \mathbb{R}$. *Then any matrix* $A \in \mathbb{F}^{m\times n}$ *may be written in the form*

$$A = QDP^{-1}$$

where $Q \in \mathbb{F}^{m\times m}$ *and* $P \in \mathbb{F}^{n\times n}$ *are unitary and* $D \in \mathbb{F}^{m\times n}$ *is in singular normal form.*

7.4 Exercises

Exercise 7.4.1. Show that if $c = \cos\theta$ and $s = \sin\theta$, then the matrix

$$Q = \begin{bmatrix} c & s \\ -s & c \end{bmatrix}$$

is orthogonal and of determinant one.

Exercise 7.4.2. Show that the set of matrices $T \in \mathbb{F}^{(n+1)\times(n+1)}$ of form

$$T = \begin{bmatrix} L & X_0 \\ 0_{1\times n} & 1 \end{bmatrix} \in \mathbb{F}^{(n+1)\times(n+1)}$$

where $L \in \mathbb{F}^{n\times n}$ is invertible and $X_0 \in \mathbb{F}^{n\times 1}$ is a matrix group. (It is called the **affine group**.)

Exercise 7.4.3. Show that the set of all matrices T of form

$$T = \begin{bmatrix} L & X_0 \\ 0_{1\times n} & 1 \end{bmatrix} \in \mathbb{R}^{(n+1)\times(n+1)}$$

where $L \in \mathbb{R}^{n\times n}$ is orthogonal and $X_0 \in \mathbb{R}^{n\times 1}$ is a matrix group. (It is called the **Euclidean group**.)

Chapter 8

Index