ADVANCED LINEAR ALGEBRA

Adair Antonio da Silva Neto

October 23, 2022

Contents

1 Review **2**

1.1 Vector Spaces . . . 2

1.1.1 Subspaces . . . 3

1.1.2 Bases and Dimension . . . 4

1.1.3 Coordinates . . . 6

1.1.4 The Row and Column Spaces of a Matrix . . . 8

1.2 Linear Transformations . . . 9

1.2.1 Basic Definitions . . . 9

1.2.2 The Algebra of Linear Transformations . . . 11

1.2.3 Isomorphisms . . . 13

1.2.4 Matrix Representation . . . 14

1.3 Diagonalization . . . 15

1.3.1 Motivation . . . 15

1.3.2 Characteristic Values . . . 16

1.3.3 Diagonalization . . . 17

1.3.4 Direct Sums . . . 19

1.3.5 Invariant Subspaces . . . 20

2 Dual Spaces **21**

2.1 Category Theory . . . 21

2.2 Quotient Spaces . . . 23

2.3 Dual Space . . . 25

2.4 The Double Dual . . . 29

2.5 The Transpose of a Linear Transformation . . . 32

3 Polynomials **34**

3.1 Algebras . . . 34

3.2 The Algebra of Polynomials . . . 34

3.3 Polynomial Ideals . . . 35

3.4 Prime Factorization 36

4 Canonical Forms **38**

4.1 Annihilating Polynomials 38

4.2 Cyclic subspaces 42

5 Determinants **48**

5.1 Commutative Rings 48

5.2 Determinant Functions 48

5.3 Permutations and the Uniqueness of Determinants 53

5.4 Properties of Determinants 55

5.5 Modules 58

5.6 Multilinear Functions 60

5.7 The Grassman Ring 63

References **64**

Chapter 1

Review

In this chapter, we'll proceed with an overview of elementary linear algebra, covering the definition of vector spaces, bases and coordinates, linear transformations and matrices, rank, nullity, inner product, normal and self-adjoint operators, and diagonalization. The proofs in this chapter will be skipped.

1.1 Vector Spaces

Loosely speaking, linear algebra is that branch of mathematics which treats the common properties of algebraic systems which consist of a set, together with a reasonable notion of a 'linear combination' of elements in the set.

Definition 1.1.1 (Vector Space)**.** A **vector space** (or **linear space**) V over a field $\mathbb{F}$ is a set with a binary operation '+' on V (called **addition**) and an action '$\cdot$' of $\mathbb{F}$ on V (called **scalar multiplication**) such that, for any $x, y \in \mathrm{V}$ and $a, b \in \mathbb{F}$, $x+y \in \mathrm{V}$ (closed under addition) and $a \cdot x \in \mathrm{V}$ (invariant under scalar multiplication) satisfying:

1. $x+y=y+x$.
2. $(x+y)+z=x+(y+z)$.
3. There exists $0 \in \mathrm{V}$ such that $x+0=x$ for all $x \in \mathrm{V}$.
4. For all $x \in \mathrm{V}$, there exists $y \in \mathrm{V}$ such that $x+y=0$.
5. There exists $1 \in \mathbb{F}$ such that $1 \cdot x = x$ for all $x \in \mathrm{V}$.
6. $a \cdot (b \cdot x) = (a \cdot b) \cdot x$.
7. $a \cdot (x+y) = a \cdot x + a \cdot y$.
8. $(a+b) \cdot x = a \cdot x + b \cdot x$.

We'll refer to the elements of V as **vectors** and to the elements of $\mathbb{F}$ as **scalars**.

In the following pages, we'll use V to denote a vector space and $\mathbb{F}$ to denote a field. And 'iff.' means 'if and only if'.

Example 1.1.1 (Some Vector Spaces).

1. The **zero-dimensional space**. The set $V = \{0\}$ under some field $\mathbb{F}$.
2. The **field $\mathbb{F}$ as a one-dimensional coordinate space**. A field (e.g. $\mathbb{C}$) can be interpreted as a vector space of a subfield of it (e.g. $\mathbb{R}$).
3. The **n-tuple space** $\mathbb{F}^n$.
4. The **space of $m \times n$ matrices** $\mathbb{F}^{m\times n}$.
5. **Function spaces** $F(S)$. Which maps S into the field $\mathbb{F}$.
6. The **space of polynomial functions over a field** $\mathbb{F}$.

Some immediate conclusions follow from this definition.

Lemma 1.1.1 (Basic Properties). For all $x \in V$ and $a \in \mathbb{F}$, the following properties hold:

1. $\underset{\in\mathbb{F}}{0} \cdot x = \underset{\in V}{0}$
2. $\underset{\in\mathbb{F}}{(-a)} \cdot x = -(\underset{\in V}{a \cdot x}) = a \cdot \underset{\in V}{(-x)}$
3. $a \cdot \underset{\in V}{0} = \underset{\in V}{0}$

The basic motivation of Linear Algebra is to solve systems of linear equations. The concept of linear combination is of essential character in solving these systems and inspires the definition of matrix multiplication and linear transformations.

Definition 1.1.2 (Linear combinations). Let $S \subseteq V$, $S \neq \varnothing$.

A vector $v \in V$ is a **linear combination** of S if it can be written as

$$v = a_1u_1 + a_2u_2 + \ldots + a_nu_n = \sum_{i=1}^{n} a_iu_i$$

for some vectors $u_1, \ldots, u_n \in S$ and scalars $a_1, \ldots, a_n \in \mathbb{F}$.

1.1.1 Subspaces

Definition 1.1.3 (Subspace). Let V be a vector space over a field $\mathbb{F}$. A subset $W \subseteq V$ is a **subspace** of V if W is itself a vector space with respect to the addition and scalar multiplication on V.

Theorem 1.1.2 (Criteria for Subspaces). Let $W \subseteq V$. Then W is a subspace of V iff.

1. $0 \in W$.
2. $x + y \in W$ for all $x, y \in W$ (closed under addition).
3. $c \cdot x \in W$ for all $c \in \mathbb{F}$ and $x \in W$ (closed under scalar multiplication).

However, we can simplify this check a little more.

Theorem 1.1.3 (New Criteria for Subspaces)**.** Let $W \subseteq V$. Then W is a subspace of V iff. for any $x, y \in W$ and $c \in \mathbb{F}$, we have that $cx + y \in W$.

The conditions that an arbitrary vector in V must satisfy in order to belong to W are called **linear conditions**. A combination of linear conditions is also a linear condition. In other words, we have the next theorem.

Theorem 1.1.4 (Intersection of subspaces is a subspace)**.** If $W_1, \ldots, W_n$ are subspaces of V, then $W = \bigcap_{i=1}^{n} W_i$ is also a subspace of V.

Definition 1.1.4 (Span)**.** Let $S \subseteq V$. The **subspace spanned** by S (or **span** of S), denoted by Span(S), is the intersection of all subspaces of V which contain S.

We define the $\text{Span}(\varnothing) = \{0\}$.

The following theorem gives an equivalent definition.

Theorem 1.1.5 (Equivalent Definition for Span)**.** The **span** of S is the subset of V consisting of all linear combinations of S.

$$\text{Span}(S) = \{a_1 u_1 + \ldots + a_n u_n : n \in \mathbb{N}, a_i \in \mathbb{F}, u_i \in S\}$$

Theorem 1.1.6 (Properties of the Span)**.** Let S be any subset of V, not necessarily a subspace. Then,

1. Span(S) is a subspace of V.
2. Any subspace of V containing S also must contain Span(S).

Definition 1.1.5 (Generation of Spaces)**.** Let $S \subseteq V$. We say that S **generates** (or **spans**) V if $\text{Span}(S) = V$.

1.1.2 Bases and Dimension

Definition 1.1.6 (Linear Dependence)**.** A subset S of V is **linearly dependent** if there exists a finite number of distinct vectors $u_1, \ldots, u_n \in S$ and scalars $a_1, \ldots, a_n \in \mathbb{F}$, with at least one $a_i \neq 0$, such that

$$a_1 u_1 + \ldots + a_n u_n = 0$$

And $S \subseteq V$ is **linearly independent** if it is not linearly dependent, i.e., no non-trivial linear combination of $u_1, \ldots, u_n$ vanishes.

Theorem 1.1.7 (Criteria for Linear Dependence)**.** Let $S_1 \subseteq S_2 \subseteq V$.

1. If S_1 is linearly dependent, then S_2 is also linearly dependent.
2. If S_2 is linearly independent, then S_1 is also linearly independent.

3. Let $S \subseteq V$ be linearly independent, and $v \in V$ such that $v \notin S$. Then, $S \cup \{v\}$ is linearly dependent iff. $v \in \text{Span}(S)$.

Definition 1.1.7 (Basis)**.** A **basis** for V is a subset of V which is both linearly independent and generates V.

Example 1.1.2. Let S be the subset of $\mathbb{F}^n$ containing

$$\begin{aligned} e_1 &= (1, 0, 0, \ldots, 0) \\ e_2 &= (0, 1, 0, \ldots, 0) \\ &\vdots \\ e_n &= (0, 0, 0, \ldots, 1) \end{aligned}$$

Clearly, these vectors span $\mathbb{F}^n$ and are linearly independent. Then this set is a basis for $\mathbb{F}^n$ and is called the **standard basis** of $\mathbb{F}^n$.

An alternative characterization of vector spaces is given by the following theorem.

Theorem 1.1.8. A subset of vectors $\{u_1, \ldots, u_n\}$ of V is a basis iff. every $v \in V$ can be uniquely written in the form

$$v = a_1 u_1 + \ldots + a_n u_n$$

for some $a_i \in \mathbb{F}$.

Theorem 1.1.9 (Replacement Theorem)**.** Let V be a vector space generated by $G \subseteq V$ with $|G| = n$, and L be a linearly independent subset of V, $|L| = m$. Then $m \leq n$, and there exists $H \subseteq G$ such that $|H| = n - m$ and $L \cup H$ generates V.

In other words, if V is a vector space spanned by a finite set of vectors $u_1, \ldots, u_n$, then any independent set of vectors in V is finite and contains no more than n elements.

The next theorem guarantees that every basis has the same cardinality, i.e., the number of elements in the basis does not depend on the basis.

Theorem 1.1.10. If V is a finitely generated vector space, then every basis of V has the same number of elements in it.

Definition 1.1.8 (Dimension)**.** If V is a finitely generated vector space, we define the **dimension** of V, denoted dim(V), as the cardinality of a basis for V.

Corollary 1.1.11. Let $n = \dim V < \infty$. Then

1. Any subset of V which contains more than n vectors is linearly dependent.
2. No subset of V which contains fewer than n vectors can span V.

Lemma 1.1.12. Let S be a linearly independent subset of a vector space V. If $v \in V$ is not in the subspace spanned by S, then the set obtained by adjoining v to S is linearly independent.

Theorem 1.1.13. If W is a subspace of a finite-dimensional vector space V, every linearly independent subset of W is finite and is a part of a finite basis for W.

A corollary of this theorem is that proper subspaces have smaller dimension.

Corollary 1.1.14 (Monotonicity of dimension)**.** Let W be a subspace of V with $\dim(V) < \infty$. Then

$$\dim(W) \leq \dim(V)$$

If the equality $\dim(W) = \dim(V)$ holds, then $V = W$.

Corollary 1.1.15 (Extension of a basis)**.** If $W = \{w_1, \ldots, w_m\}$ is a linearly independent set of vectors in a finite-dimensional vector space V, then there exists a basis of V that contains W.

Corollary 1.1.16. Let $A \in \mathbf{M}_n(\mathbb{F})$ and suppose that the row vectors of A form a linearly independent set of vectors in $\mathbb{F}^n$. Then A is invertible.

Theorem 1.1.17. If W_1 and W_2 are both finite-dimensional subspaces of V, then $W_1 + W_2$ is finite-dimensional and

$$\dim W_1 + \dim W_2 = \dim(W_1 \cap W_2) + \dim(W_1 + W_2)$$

Definition 1.1.9 (Maximal)**.** Let $E = \{v_1, \ldots, v_n\}$ be a set of vectors in V and let $F = \{v_{i_1}, \ldots, v_{i_m}\}$ be a linearly independent subset of E. If every element in E can be expressed as a linear combination of the elements of F, then F is said to be **maximal**.

The number of elements in a maximal subset equals the dimension of the span of E and is called the **rank**.

Definition 1.1.10 (Flags)**.** A sequence of subspaces $V_0 \subset V_1 \subset \ldots \subset V_n$ of the space V is said to be a **flag**.

More generally, a sequence of subsets $S_0 \subset S_1 \subset \ldots \subset S_n$ is called **increasing filtering**.

A flag is said to be **maximal** if $V_0 = \{0\}$, $\bigcup V_i = V$ and there's no subspace between other two, i.e., if $V_i \subset M \subset V_{i+1}$ then either $V_i = M$ or $V_{i+1} = M$.

Notice that given any basis $\{u_1, \ldots, u_n\}$ of V, we can construct a flag by setting $V_0 = \{0\}$ and $V_i = \operatorname{span}(\{u_1, \ldots, u_i\})$ for $i \geq 1$.

Theorem 1.1.18. The dimension of a vector space V equals the length of any maximal flag of V.

The next theorem is an example of application of Zorn's lemma.

Theorem 1.1.19. Every vector space has a basis.

1.1.3 Coordinates

The coordinates of a vector relative to a basis will be the coefficients that are used to represent the vector as a linear combination of the vectors in the basis. For example, if $(v_1, \ldots v_n)$ is an

arbitrary vector in $\mathbb{R}^n$ and $e_1, \ldots, e_n$ is the standard basis for $\mathbb{R}^n$, then we express

$$v = (v_1, \ldots, v_n) = \sum_{i=1}^{n} v_i e_i$$

However, for this expression to be adequately defined, the vectors in the basis must be ordered. To put it another way, we must look at our basis as a sequence instead of a set to distinguish its i-th element.

Definition 1.1.11 (Ordered Basis)**.** Let $\dim(V) < \infty$. An **ordered basis** for V is a basis for V with a fixed order on its vectors.

With this definition, we say that v_i is the ith **coordinate of** v **relative to the ordered basis**. And we use $[v]_\beta$ to denote the coordinates of v concerning the ordered basis β. More precisely,

Definition 1.1.12 (Coordinates)**.** Let $\beta = \{v_1, \ldots, v_n\}$ be an ordered basis for V. Then any vector $x \in V$ can be written uniquely as

$$x = a_1 v_1 + \ldots + a_n v_n$$

for $a_1, \ldots, a_n \in \mathbb{F}$.

We define the **coordinate vector** as

$$[x]_\beta = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \in \mathbb{F}^n$$

Now, what happens with the coordinates when we change from one basis to another?

Let $\beta = \{\beta_1, \ldots, \beta_n\}$ and $\gamma = \{\gamma_1, \ldots, \gamma_n\}$ be two ordered bases for the finite-dimensional space V. And notice that we can write every vector of the basis γ as a linear combination of the vectors of β as follows:

$$\begin{aligned} \gamma_1 &= a_{11} \cdot \beta_1 + a_{21} \cdot \beta_2 + \ldots + a_{n1} \cdot \beta_n \\ \gamma_2 &= a_{12} \cdot \beta_1 + a_{22} \cdot \beta_2 + \ldots + a_{n2} \cdot \beta_n \\ &\vdots \\ \gamma_n &= a_{1n} \cdot \beta_1 + a_{2n} \cdot \beta_2 + \ldots + a_{nn} \cdot \beta_n \end{aligned}$$

where each a_{ij} is a scalar.

Thus, for each $i \in \{1, 2, \ldots, n\}$, the coordinates vector of γ_i in the basis β is given by

$$[\gamma_i]_\beta = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{ni} \end{bmatrix}$$

With this algorithm, we obtain the coordinates of each vector in the basis γ concerning the

basis β. And we form the **transition matrix**, also called **change-of-basis matrix**, from β to γ:

$$P_{\beta\to\gamma} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$

Note that each column is formed by the coordinates of $\gamma_1, \ldots, \gamma_n$ with respect to the basis β.

Theorem 1.1.20. Let V be an n-dimensional vector space and let $\beta = \{u_1, \ldots, u_n\}$ and $\gamma = \{u'_1, \ldots, u'_n\}$ be two ordered bases of V. Then there is a unique and invertible $n \times n$ matrix P such that

1. $[u]_\beta = P[u]_\gamma$,
2. $[u]_\gamma = P^{-1}[u]_\beta$,

for every vector $u \in V$. And the columns of P are given by

$$P_j = [u'_j]_\beta,\ j = 1, \ldots, n$$

Example 1.1.3 (Change of basis). Consider β the standard basis of $\mathbb{R}^3$ and

$$\gamma = \{(1,0,1), (1,1,1), (1,1,2)\}$$

Find the transition matrix $P_{\gamma\to\beta}$.

Solution: The first step is to write each vector of β as a linear combination of the vectors of γ. I.e.,

$$\begin{aligned}(1,0,0) &= a_{11}\cdot(1,0,1) + a_{21}\cdot(1,1,1) + a_{31}\cdot(1,1,2) \\ &= 1\cdot(1,0,1) + 1\cdot(1,1,1) - 1\cdot(1,1,2) \\ (0,1,0) &= a_{12}\cdot(1,0,1) + a_{22}\cdot(1,1,1) + a_{32}\cdot(1,1,2) \\ &= -1\cdot(1,0,1) + 1\cdot(1,1,1) + 0\cdot(1,1,2) \\ (0,0,1) &= a_{13}\cdot(1,0,1) + a_{23}\cdot(1,1,1) + a_{33}\cdot(1,1,2) \\ &= 0\cdot(1,0,1) - 1\cdot(1,1,1) + 1(1,1,2)\end{aligned}$$

With these values, we form the transition matrix:

$$P_{\gamma\to\beta} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix}$$

Another way of solving the above example is using the previous theorem. First, form a matrix P whose ith column is the ith element of the basis γ. Second, find the inverse P^{-1}. Finally, the ith column of P^{-1} gives the coordinate of the ith vector of the standard basis in the basis γ.

1.1.4 The Row and Column Spaces of a Matrix

Before heading to next section, we introduce some useful nomenclature and results.

Definition 1.1.13 (Row Space). Let A be an $m \times n$ matrix over the field $\mathbb{F}$. We define the **row space** as the subspace of $\mathbb{F}^n$ generated by the rows of A. The dimension of the row space is called **row rank**.

Theorem 1.1.21.

1. Row-equivalent matrices have the same row space.
2. The non-zero lines of a row-reduced echelon matrix form a basis for its row space.
3. If W is a subspace of $\mathbb{F}^n$ such that $\dim \mathrm{W} \leq m$, then there exists a unique row-reduced echelon matrix $m \times n$ over $\mathbb{F}$ whose row space is W.
4. Every matrix is row-equivalent to one, and only one, row reduced echelon matrix.
5. Two matrices are row-equivalent iff. they have the same row space.

1.2 Linear Transformations

In plain words, a linear transformation (or linear mapping) is a function from a vector space to another which preserves the structure of a vector space. More precisely,

1.2.1 Basic Definitions

Definition 1.2.1 (Linear Transformation). Let V and W be two vector spaces over the same field $\mathbb{F}$. A **linear transformation** $\mathrm{T} : \mathrm{V} \longrightarrow \mathrm{W}$ is a function satisfying:

1. $\mathrm{T}(x+y) = \mathrm{T}(x) + \mathrm{T}(y)$, for all $x, y \in \mathrm{V}$.
2. $\mathrm{T}(cx) = c\mathrm{T}(x)$, for all $x \in \mathrm{V}$, $c \in \mathbb{F}$.

Put it another way, a linear mapping is a **homomorphism** of additive groups.

Theorem 1.2.1 (Properties).

1. If T is a linear transformation, then $\mathrm{T}(0) = 0$.
2. $\mathrm{T}\left(\sum_{i=1}^{n} a_i x_i\right) = \sum_{i=1}^{n} a_i \mathrm{T}(x_i)$ for all $x_i \in \mathrm{V}$, $a_i \in \mathbb{F}$.
3. A function $\mathrm{T} : \mathrm{V} \longrightarrow \mathrm{W}$ is a linear transformation iff. $\mathrm{T}(cx+y) = c\mathrm{T}(x) + \mathrm{T}(y)$ for all $x, y \in \mathrm{V}$, $c \in \mathbb{F}$.

Example 1.2.1. Let $\mathbb{F}$ be a field and V be the space of polynomial functions $f : \mathbb{F} \longrightarrow \mathbb{F}$ given by

$$f(x) = c_0 + c_1 x + \ldots + c_k x^k$$

Define

$$(\mathrm{D}f)(x) = c_1 + 2c_2 x + \ldots + k c_k x^{k-1}$$

Then D is a linear transformation called the differentiation operator.

Example 1.2.2. Given the field of real numbers $\mathbb{R}$ and $V = \mathscr{C}(\mathbb{R})$, we define

$$T(f(x)) = \int_0^x f(t)\, dt$$

which is a linear transformation.

How can we define linear transformations? The easiest way is to define its values on a basis and then linearly extend it to the whole space. The next theorem says this process returns a well defined linear mapping.

Theorem 1.2.2. Let $\{v_1, \ldots, v_n\}$ be a basis for V. Then for any vectors $w_1, \ldots, w_n \in W$, there exists exactly one linear transformation $T : V \longrightarrow W$ such that

$$T(v_i) = w_i, \text{ for } 1 \leq i \leq n$$

Definition 1.2.2 (Null space and Range)**.** Let $T : V \longrightarrow W$ be a linear transformation.

1. The **null space** (or **kernel**) of T is

$$\ker(T) = \{x \in V : T(x) = 0\}$$

2. The **range** of T is the image V under T, i.e.,

$$\mathrm{Im}(T) = \{y \in W : y = T(x), x \in V\}$$

The dimension of the range is called the **rank** of T and the dimension of the kernel is called the **nullity** of T.

Theorem 1.2.3. The null space of T $\ker(T)$ is a subspace of V and $\mathrm{Im}(T)$ is a subspace of W.

Theorem 1.2.4 (The Dimension Theorem (Rank–Nullity))**.** If $\dim(V) < \infty$, then

$$\dim(V) = \dim(\ker(T)) + \dim(\mathrm{Im}(T))$$

i.e., $\dim(V) = \text{nullity}(T) + \text{rank}(T)$.

Definition 1.2.3 (Injection and Surjection)**.** Let $T : V \longrightarrow W$ be a linear transformation.

1. T is **injective** if $T(v) = T(u)$ implies $v = u$, for all $u, v \in V$.
2. T is **surjective** if for every $w \in W$ there exists $v \in V$ such that $T(v) = w$.
3. T is **bijective** if T is injective and surjective.

Theorem 1.2.5.

- T is injective iff. $\ker(T) = \{0\}$.

- T is surjective iff. $\mathrm{Im}(T) = W$.

Theorem 1.2.6. Assume $\dim(V) = \dim(W)$. Then the following affirmations are equivalent:

1. T is injective.
2. T is surjective.
3. T is bijective.
4. $\dim(\mathrm{Im}(T)) = \dim(V)$.

1.2.2 The Algebra of Linear Transformations

Theorem 1.2.7. Let $T, U : V \longrightarrow W$ be linear transformations. We define, for all $x \in V$ and $a \in \mathbb{F}$,

1. $(T+U)(x) = T(x) + U(x)$;
2. $(aT)(x) = aT(x)$.

Then $T+U$ and $a \cdot U$ are also linear transformations from V to W.

Theorem 1.2.8 (Space of Linear Transformations)**.** Let $\hom_{\mathbb{F}}(V,W)$ be the set of all linear transformations from V to W. Then $\hom_{\mathbb{F}}(V,W)$ is a vector space over the same field $\mathbb{F}$ with respect to the operations defined above.

An alternative notation is $\mathscr{L}(V,W) = \hom_{\mathbb{F}}(V,W)$. When $V = W$, we write $\mathscr{L}(V)$ or $\mathrm{end}_{\mathbb{F}}(V)$.

Theorem 1.2.9. If V is an n-dimensional vector space and W is an m-dimensional vector space, both over $\mathbb{F}$, then the space $\mathscr{L}(V,W)$ has dimension mn.

Definition 1.2.4 (Composition of Linear Transformations)**.** Let V, W, Z be vector spaces. Let $T : V \longrightarrow W$ and $U : W \longrightarrow Z$ be linear transformations. Their **composition** is the function $UT : V \longrightarrow Z$ defined by $(UT)(x) = U(T(x))$ for all $x \in V$.

Theorem 1.2.10 (Composition is also linear)**.** If T and U are both linear transformations, then their composition UT is a linear transformation.

Definition 1.2.5 (Linear Operator)**.** A **linear operator** is a linear transformation from a vector space to itself. It is also called an **endomorphism**. The set of linear operators on a vector space V is denoted by $\mathscr{L}(V)$ or $\mathrm{End}_{\mathbb{F}}(V)$.

Remark that if U and T are linear operators on V, then the composition $U \circ T$ is also a a linear operator on V. The space $\mathscr{L}(V)$ has a 'multiplication' defined on it by composition. The operator $T \circ U$ is also defined, but in general $UT \neq TU$, i.e., the **Lie bracket** $[U,T] = UT - TU \neq 0$.

Lemma 1.2.11. Let U, T_1 and T_2 be linear operators on the vector space V and $c \in \mathbb{F}$. The following affirmations hold.

1. $IU = UI = U$;
2. $U(T_1 + T_2) = UT_1 + UT_2$ and $(T_1 + T_2)U = T_1U + T_2U$;
3. $c(UT_1) = (cU)T_1 = U(cT_1)$.

As a matter of fact, the vector space $\mathscr{L}(V)$, together with the composition operation, is known as a **linear algebra with identity**.

Definition 1.2.6 (Invertibility)**.** Let V, W be vector spaces, and $T : V \longrightarrow W$ a linear transformation.

1. A linear transformation $U : W \longrightarrow V$ is the **inverse** of T if $UT = I_V$ and $TU = I_W$, where I denotes the identity matrix.
2. T is invertible if it has an inverse.

Theorem 1.2.12 (Characterization of Inverses)**.** Let V, W be vector spaces, and $T : V \longrightarrow W$ a linear transformation.

1. If T is invertible, then its inverse is unique, denoted by T^{-1}.
2. T is invertible iff. T is a bijection.

Lemma 1.2.13. Let $T : V \longrightarrow W$ be an invertible linear transformation, and $\dim(V) < \infty$. Then $\dim(V) = \dim(W)$.

To check whether a transformation T is injective, notice that if T is linear, then $T(u - v) = T(u) - T(v)$. Therefore, $T(u) = T(v)$ iff. $T(u - v) = 0$.

Definition 1.2.7 (Non-singular Transformations)**.** A linear mapping T is **non-singular** if $T(v) = 0$ implies $v = 0$, i.e., the null space of T is $\{0\}$.

Hence, T is injective iff. T is non-singular. more than that, non-singular linear transformations are those which preserve linear independence.

Theorem 1.2.14. Let $T : V \longrightarrow W$ be a linear mapping. Then T is non-singular if and only if T carries each linearly independent subset of V onto a linearly independent subset of W.

Theorem 1.2.15. Let V and W be finite-dimensional vector spaces such that $\dim V = \dim W$. If T is a linear mapping from V into W, the following are equivalent:

1. T is invertible;
2. T is non-singular;
3. T is surjective;
4. If $\{v_1, \ldots, v_n\}$ is a basis for V, then $\{T(v_1), \ldots, T(v_n)\}$ is a basis for W;

5. There is some basis for V such that $\{T(v_1), \ldots, T(v_n)\}$ is a basis for W.

The set of invertible linear operators on a given space, with the operation of composition, provides an example of a group.

Definition 1.2.8 (Group). A **group** consists of the following:

1. A set G;
2. A rule (or operation) $\odot$ which associates with each pair of elements $x, y \in G$ an element $x \odot y$ in G satisfying
   - Associativity: $x \odot (y \odot z) = (x \odot y) \odot z$, for all $x, y, z \in G$;
   - Identity: There is an element e in G such that $e \odot x = x \odot e = x$, for every x in G;
   - Inverse: To each element $x \in G$ there corresponds an element x^{-1} in G such that $x \odot x^{-1} = x^{-1} \odot x = e$.

Example 1.2.3. The following are examples of groups.

- **General linear group** $GL(n)$, formed by the set of non-singular $n \times n$ matrices with the operation of function composition.
- **Permutation group** S_n, of permutations of sets of n elements.
- **Special linear group** $SL(n)$, of $n \times n$ matrices with determinant equal to one.
- **Orthogonal group** $O(n)$, of $n \times n$ matrices such that $AA^t = I$, which is the group of isometries of Euclidean space that preserve a fixed point.
- **Special Orthogonal group** $SO(n)$, consisting of orthogonal matrices whose determinant is equal to one.
- **Unitary group** $U(n)$ of all complex $n \times n$ matrices satisfying $AA^* = 1$, where $A^* = \bar{A}^t$.
- **Special unitary group** $SU(n)$ of unitary matrices with determinant one.

1.2.3 Isomorphisms

Definition 1.2.9 (Isomorsphism). Bijective linear mappings $T \in \mathscr{L}(V, W)$ are said to be **isomorphisms**, and the spaces V and W are called **isomorphic** if there exists an isomorphism between them.

If $T \in \mathscr{L}(V)$ is an isomorphism, then T is said to be an **automorphism**.

Remark that isomophism is an equivalence relation in the family of vector spaces.

Theorem 1.2.16. Every n-dimensional vector space over a field $\mathbb{F}$ is isomorphic to $\mathbb{F}^n$.

To convince yourself that this claim is true, it is enough to map every vector to its coordinates in a given basis.

A more general result states that the dimension of a space completely determines the space up to isomorphism. To put it another way, every finite subspace $S \subseteq V$ has the same dimension as

the range T(S), i.e., isomorphisms preserve dimension.

Theorem 1.2.17. Two finite-dimensional spaces V and W are isomorphic iff. they have the same dimension.

If the isomorphism does not depend on arbitrary choices, such as the basis, then it is called a **canonical** or **natural isomorphism**. This will be made precise when the language of categories is introduced.

Finally, note that the isomorphisms from a space to itself form a group with respect to the operation of function composition, which is exactly the general linear group we saw earlier.

1.2.4 Matrix Representation

Definition 1.2.10 (Matrix Representation)**.** Let V, W be vector spaces with ordered basis $\beta = \{v_1, \ldots, v_n\}$ and $\gamma = \{w_1, \ldots, w_m\}$, respectively.

Let $T : V \longrightarrow W$ be a linear transformation. Then the **matrix representation** of T with respect to β and γ is defined as the matrix $[T]_{\beta,\gamma} \in \mathbf{M}_{m\times n}(\mathbb{F})$ given by

$$[T]_{\beta,\gamma} = \begin{pmatrix} | & | & & | \\ [T(v_1)]_\gamma & [T(v_2)]_\gamma & \cdots & [T(v_n)]_\gamma \\ | & | & & | \end{pmatrix}$$

where $[T(v_i)]_\gamma$ are the coordinates of the vector $T(v_i) \in W$ with respect to the ordered basis γ.

If $V = W$ and $\beta = \gamma$, we write $[T]_\beta$.

Theorem 1.2.18. Assume V, W are finite dimensional vector spaces with ordered basis β and γ. Let $T, U : V \longrightarrow W$ be linear transformations. Then,

1. $U = T$ iff. $[U]_{\beta,\gamma} = [T]_{\beta,\gamma}$;
2. $[T + U]_{\beta,\gamma} = [T]_{\beta,\gamma} + [U]_{\beta,\gamma}$;
3. $[aT]_{\beta,\gamma} = a[T]_{\beta,\gamma}$, for all $a \in \mathbb{F}$.

Theorem 1.2.19. Let V, W, Z be vector spaces with ordered basis α, β, γ respectively. Let $T : V \longrightarrow W$ and $U : W \longrightarrow Z$ be linear transformations. Then

$$[UT]_{\alpha,\gamma} = [U]_{\beta,\gamma}[T]_{\alpha,\beta}$$

Corollary 1.2.20. Let V be a finite vector space with ordered basis β. Let $T, U \in \mathscr{L}(V)$. Then $[UT]_\beta = [U]_\beta[T]_\beta$.

Theorem 1.2.21. Let V, W be finite dimensional vector spaces with ordered basis β and γ. Let $T : V \longrightarrow W$ be a linear transformation. For all $u \in V$,

$$[T(u)]_\gamma = [T]_{\beta,\gamma}[u]_\beta$$

Definition 1.2.11 (Invertibility for a Matrix)**.** A matrix $A \in \mathbf{M}_{m\times n}(\mathbb{F})$ is **invertible** if there exists $B \in \mathbf{M}_{m\times n}(\mathbb{F})$ such that $AB = BA = I$.

Theorem 1.2.22. Let V, W be finite dimensional vector spaces with ordered bases β and γ respectively. Let $T : V \longrightarrow W$ be a linear transformation. Then T is invertible iff. $[T]_{\beta,\gamma}$ is invertible. Moreover,

$$[T^{-1}]_{\gamma,\beta} = ([T]_{\beta,\gamma})^{-1}$$

Theorem 1.2.23. Let V be a finite-dimensional vector space and let

$$\beta = \{v_1, \ldots, v_n\} \text{ and } \gamma = \{w_1, \ldots, w_n\}$$

be ordered basis for V. Suppose that $T \in \mathscr{L}(V)$. If P is the matrix with columns $P_j = [w_j]_\beta$ (i.e. the coordinates of the j-th vector on the basis β), then

$$[T]_\gamma = P^{-1}[T]_\beta P$$

Alternatively, if U is the invertible operator defined by $U[v_j] = w_j$, then

$$[T]_\gamma = [U]_\beta^{-1}[T]_\beta[U]_\beta$$

Definition 1.2.12 (Similar Matrices)**.** Let A and B be $n \times n$ matrices. We say that B is **similar** to A if there exists an invertible $n \times n$ matrix P such that

$$B = P^{-1}AP$$

1.3 Diagonalization

1.3.1 Motivation

The question that motivates this section is 'when the matrix of a linear operator assumes a simple form?'

Consider, for example, the following diagonal matrix.

$$D = \begin{bmatrix} c_1 & 0 & 0 & \ldots & 0 \\ 0 & c_2 & 0 & \ldots & 0 \\ 0 & 0 & c_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & c_n \end{bmatrix}$$

And suppose that T is a linear operator on a finite vector space V. If there exists an ordered basis $\beta = \{v_1, v_2, \ldots, v_n\}$ of V in which T is represented as the diagonal matrix D, then it is possible to extract some informations about the linear operator T, such as its rank and determinant, in a simple and direct way.

Since

$$[T]_\beta = D \iff T(v_k) = c_k v_k, \quad k = 1, 2, \ldots, n$$

the range of T is simply the subspace spanned by the vectors v_k in which c_k does not vanish. Analogously, the null space of T is generated by the remaining v_k's.

Is it always possible to represent a linear operator T as a diagonal matrix? If not, what is the simplest type of matrix by which we can represent T?

1.3.2 Characteristic Values

We saw that if T can be represented as a diagonal matrix, then

$$[\mathrm{T}]_\beta = \mathrm{D} \iff \mathrm{T}(v_k) = c_k v_k, k = 1, 2, \ldots, n$$

Motivated by this fact, we'll look for which vectors are mapped by T into scalar multiples of themselves.

Definition 1.3.1 (Eigenvalue and Eigenvector)**.** Given a vector space V over a field $\mathbb{F}$ and $\mathrm{T} \in \mathrm{End}(\mathrm{V})$, we define **eigenvalue** (or **characteristic value**) if T as the scalar $\lambda \in \mathbb{F}$ such that there exists a non-zero vector $v \in \mathrm{V}$ satisfying $(v) = \lambda v$.

If λ is an eigenvalue of T, then

- Any vector v satisfying $\mathrm{T}(v) = \lambda v$ is said to be a **eigenvector** (or **characteristic vector**) of T associated with the eigenvalue λ.
- The set of all eigenvectors is called the **eigenspace** (or characteristic space) associated with λ.

Notice that the eigenspace associated with λ is a subspace of V and it is exactly the null space of the linear transformation $(\mathrm{T} - \lambda\mathrm{I})$. We say that λ is an eigenvalue of T when the eigenspace is different from the zero subspace, i.e., if $(\mathrm{T} - \lambda\mathrm{I})$ is not injective. If $\dim(\mathrm{V}) < \infty$, this happens exactly when its determinant is different from zero.

Theorem 1.3.1. Let $\mathrm{T} \in \mathrm{End}(\mathrm{V})$ and $\lambda \in \mathbb{F}$. The following are equivalent.

1. λ is an eigenvalue of T.
2. The operator $(\mathrm{T} - \lambda\mathrm{I})$ is singular (i.e. not invertible).
3. $\det(\mathrm{T} - \lambda\mathrm{I}) = 0$.

The determinant criterion tell us how to find the eigenvalues of T. Since $det(\mathrm{T} - \lambda\mathrm{I})$ is a polynomial of degree n in the variable λ, we find the eigenvalues as the roots of that polynomial.

Definition 1.3.2 (Eigenvalues for Matrices)**.** If A a matrix $n \times n$ over $\mathbb{F}$, an **eigenvalue** of A in $\mathbb{F}$ is a scalar $\lambda \in \mathbb{F}$ such that the matrix $(\mathrm{A} - \lambda\mathrm{I})$ is singular.

Hence, λ is an eigenvalue of A iff. $\det(\mathrm{A} - \lambda\mathrm{I}) = 0$.

Definition 1.3.3 (Characteristic Polynomial)**.** This leads us to define the **characteristic polynomial** of A as

$$f(x) = \det(\mathrm{A} - x\mathrm{I})$$

The set of all roots of the characteristic polynomial is called the **spectrum** of A.

An important result is that similar matrices have the same characteristic polynomial. Thus, they have the same eigenvalues.

1.3.3 Diagonalization

Definition 1.3.4 (Diagonalization)**.** Let $T \in End(V)$. We say that T is **diagonalizable** if there exists a basis $\beta = \{v_1, v_2, \ldots, v_n\}$ for V formed by eigenvectors of T.

In other words, the linear operator has a diagonal matrix with respect to a V. Since $T(v_i) = \lambda_i v_i$, the representation of T in the ordered basis β is given by:

$$[T]_\beta = \begin{bmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ 0 & 0 & \lambda_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \lambda_n \end{bmatrix}$$

An alternative definition is that T is diagonalizable when the eigenvectors of T span V.

Some important results:

Theorem 1.3.2. Let $\dim(V) < \infty$ and $T \in End(V)$.

1. T is diagonalizable iff. there exists a basis of V formed by eigenvalues of T.
2. If f is any polynomial and $T(v) = \lambda v$, then $f(T(v)) = f(\lambda)v$.
3. If $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues of T and $v_1, \ldots, v_k$ are the eigenvectors associated with $\lambda_1, \ldots, \lambda_k$ respectively, then $\{v_1, \ldots, v_k\}$ is linearly independent.
4. If $\dim(V) = n$ and $\lambda_1, \ldots, \lambda_n$ are distinct eigenvalues of T, then T is diagonalizable. In other words, if all eigenvalues of T are different, then T is diagonalizable.
5. If W_i is the eigenspace associated with λ_i and $W = W_1 + W_2 + \ldots + W_k$, then

   $$\dim(W) = \dim(W_1) + \ldots + \dim(W_k)$$

   Moreover, if β_i is an ordered basis for W_i, then $\beta = \{\beta_1, \ldots, \beta_n\}$ is an ordered basis for W. Thus, the sum of eigenspaces is a direct sum.

With these conclusions, it is possible to guess that there exist more equivalences between diagonalizable transformations and their eigenvalues and eigenspaces.

Theorem 1.3.3. Suppose that $\dim(V) < \infty$ and let $T \in End(V)$, $\lambda_1, \ldots, \lambda_k$ distinct eigenvalues of T and $W_i = \text{Ker}(T - \lambda_i I)$. The following are equivalent.

1. T is diagonalizable.
2. The characteristic polynomial for T is

   $$p_T(\lambda) = (\lambda - \lambda_1)^{d_1}(\lambda - \lambda_2)^{d_2} \ldots (\lambda - \lambda_k)^{d_k}$$

and $\dim(W_i) = d_i$ for $i = 1, 2, \ldots, k$.

3. $\dim(W_1) + \ldots + \dim(W_k) = \dim(V)$.

With this result, given a diagonalizable matrix A, we can find a diagonal matrix Λ, similar to A, such that

$$A = P\Lambda P^{-1} \text{ e } \Lambda = P^{-1}AP$$

where Λ is constructed by the eigenvalues of A, and P is constructed from the eigenvectors of A.

Definition 1.3.5 (Algebraic and Geometric Multiplicities)**.** Consider $T \in \text{End}(V)$ and β any basis for V. We know that its characteristic polynomial is given by

$$p_T(\lambda) = \det([T]_\beta^\beta - \lambda I)$$

If $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the roots of $p_T(\lambda)$, then by the fundamental theorem of algebra

$$p_T(\lambda) = a(\lambda - \lambda_1)^{m_1}(\lambda - \lambda_2)^{m_2} \ldots (\lambda - \lambda_k)^{m_k}$$

Choosing an eigenvalue λ_i, we define:

- The **algebraic multiplicity** of λ_i as the power of the term $(\lambda - \lambda_i)$ in $p_T(\lambda)$.
- The **geometric multiplicity** of λ_i as $\dim \text{Ker}(T - \lambda_i I)$.

Remark. The geometric multiplicity is always lesser or equal to the algebraic multiplicty.

Example 1.3.1 (Diagonalization of a Linear Operator)**.** Let $T \in \mathscr{L}(\mathbb{R}^3)$ defined by

$$T(x, y, z) = (-9x + 4y + 4z, -8x + 3y + 4z, -16x + 8y + 7z)$$

Show that T is diagonalizable and find the eigenvectors that form a basis for $\mathbb{R}^3$.

Solution: Notice that the matrix representation of T in the standard basis β is:

$$[T]_\beta = \begin{bmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{bmatrix}$$

The first step is to find the eigenvalues of $[T]_\beta$. Computing $\det([T]_\beta - \lambda I)$:

$$\begin{vmatrix} -9-\lambda & 4 & 4 \\ -8 & 3-\lambda & 4 \\ -16 & 8 & 7-\lambda \end{vmatrix} = -\lambda^3 + \lambda^2 + 5\lambda + 3 = 0 \iff (\lambda + 1)^2(\lambda - 3) = 0$$

Thus, we have two eigenvalues $\lambda_1 = -1$, with algebraic multiplicity equal to two, and $\lambda_2 = 3$.

Computing the eigenvalue associated with $\lambda_1 = -1$:

$$\begin{bmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = -1 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \iff \begin{cases} -8x_1 + 4x_2 + 4x_3 = 0 \\ -8x_1 + 4x_2 + 4x_3 = 0 \\ -16x_1 + 8x_2 + 8x_3 = 0 \end{cases}$$

Notice that we have only one linearly independent row. I.e., the nullspace of the coefficient matrix have rank equal to two. That means that we can extract two linearly independent eigenvectors.

In fact, we can take $x_1 = 1, x_2 = 2, x_3 = 0$ and $x_1 = 1, x_2 = 0, x_3 = 2$, obtaining the eigenvectors $(1, 2, 0)$ and $(1, 0, 2)$.

For $\lambda_2 = 3$, we have the system:

$$\begin{cases} -12x_1 + 4x_2 + 4x_3 = 0 \\ -8x_1 + 0x_2 + 4x_3 = 0 \\ -16x_1 + 8x_2 + 4x_3 = 0 \end{cases} \iff \begin{cases} x_1 = \frac{1}{2}x_3 \\ x_2 = \frac{1}{2}x_3 \\ x_3 = x_3 \end{cases}$$

Hence, we can choose the vector $(1, 1, 2)$.

Since we obtained three linearly independent eigenvectors, we have that T is a diagonalizable linear operator. Moreover, we obtained the following basis for $\mathbb{R}^3$:

$$\begin{bmatrix} 1 & 2 & 0 \\ 1 & 0 & 2 \\ 1 & 1 & 2 \end{bmatrix}$$

Theorem 1.3.4 (Spectral Theorem)**.** Suppose that T is a linear operator in a finite-dimensional vector space V. If V is defined over $\mathbb{C}$, consider T normal. If V is defined over $\mathbb{R}$, consider T self-adjoint.

Let $\lambda_1, \ldots, \lambda_k$ be distinct eigenvalues of T, W_j the eigenspace associated with λ_j and E_j the orthogonal projection of V in W_j. The following are equivalent:

1. W_j is orthogonal to W_i when $i \neq j$.
2. V is a direct sum of $W_1, \ldots, W_k$.
3. T can be decomposed as
$$T = \lambda_1 E_1 + \ldots + \lambda_k E_k$$
denominated **spectral resolution**.

1.3.4 Direct Sums

Definition 1.3.6 (Direct Sum)**.** A space V is a **direct sum** of its subspaces $V_1, \ldots, V_n$ if every vetor $v \in V$ can be uniquely represented in the form $\sum_{i=1}^{n} v_i$, where $v_i \in V_i$.

When these conditions are satisfied, we write

$$V = V_1 \oplus \ldots \oplus V_n = \bigoplus_{i=1}^{n} V_i$$

A vector space is a direct sum of its subspaces iff. the intersections of subspaces is the zero vector and the sum of each subspace equals the whole space.

What happens when $V_1, \ldots V_n$ are not imbedded in a general space?

Definition 1.3.7 (External Direct Sum)**.** Let $V_1, \ldots, V_n$ be vector spaces. The **external direct sum** V consists of

1. The n-uples $(v_1, \ldots, v_n)$, where $v_i \in V_i$;
2. Addition and multiplication by a scalar performed coordinate-wise

$$(v_1, \ldots, v_n) + (w_1, \ldots, w_n) = (v_1 + w_1, \ldots, v_n + w_n)$$
$$a(v_1, \ldots, v_n) = (av_1, \ldots av_n)$$

Notice that the mapping $f_i : V_i \longrightarrow V$, where $f_i(v) = (0, \ldots, 0, v, 0, \ldots, 0)$ (v is in the ith location) is a linear imbedding of V_i into V.

It follows immediately from the definiton that

$$V = \bigoplus_{i=1}^{n} f_i(V_i)$$

Identifying V_i with $f_i(V_i)$, we obtain a vector space which contains V_i and decomposes into the direct sum of V_i.

1.3.5 Invariant Subspaces

Suppose $T \in \mathrm{End}(V)$. If we decompose V in direct sums

$$V = U_1 \oplus \ldots \oplus U_m$$

where each U_j is a proper subspace of V, then to understand the behavior of T it is sufficient to study the behavior of T in each U_j. To do that, we consider $T|_{U_j}$, i.e., T restricted to U_j.

However, it is not always the case that $T|_{U_j}$ has its image in the same subspace U_j. That's why we need the following.

Definition 1.3.8 (Invariant Subspace)**.** A subspace U of V is said an **invariant subspace** under T if for each $u \in U$ we have that $T(u) \in U$. I.e., U is invariant under T if $T|_U$ is a linear operator in U.

Chapter 2

Dual Spaces

2.1 Category Theory

In this first section, we introduce a useful tool to generalize and see our results in another perspective.

Definition 2.1.1 (Category)**.** A category $\mathfrak{C}$ consists of the following:

1. A collection $\mathrm{Ob}(\mathfrak{C})$ of elements which are called **objects**.
2. For each pair of objects $A, B \in \mathrm{Ob}(\mathfrak{C})$, a set $\hom_{\mathfrak{C}}(A, B)$ whose elements are called **morphisms**, **maps** or **arrows** from A to B. They are denoted by
$$f : A \longrightarrow B \text{ or } A \xrightarrow{f} B$$
The object $A = \mathrm{dom}\,(f)$ is called the **domain** of f and $B = \mathrm{codom}\,(f)$ is called the **codomain** of f.
3. The set of all morphisms of $\mathfrak{C}$ is denoted by $\mathrm{Mor}\,(\mathfrak{C})$ and must satisfy the following property. For every morphism f there exist uniquely defined objects A, B such that $f \in \hom_{\mathfrak{C}}(A, B)$. I.e., $\mathrm{Mor}\,(\mathfrak{C})$ is a disjoint union $\bigcup \hom_{\mathfrak{C}}(A, B)$ for all ordered pairs $A, B \in \mathrm{Ob}(\mathfrak{C})$.
4. For $f \in \hom_{\mathfrak{C}}(A, B)$ and $g \in \hom_{\mathfrak{C}}(B, C)$ there is a morphism $g \circ f \in \hom_{\mathfrak{C}}(A, C)$ called the **composition** or **product** of g with f. Moreover, composition is associative:
$$f \circ (g \circ h) = (f \circ g) \circ h$$
5. For each object $A \in \mathfrak{C}$ there exists a morphism $1_A \in \hom_{\mathfrak{C}}(A, A)$ called the **identity morphism** for A with the property that if $f \in \hom_{\mathfrak{C}}(A, B)$ then
$$1_B \circ f = f \text{ and } f \circ 1_A = f$$

Definition 2.1.2 (Isomorphism for Categories)**.** The morphism $f : A \longrightarrow B$ is said to be an **iso-**

morphism if there exists a morphism $g : B \longrightarrow A$ such that

$$g \circ f = 1_A \text{ and } f \circ g = 1_B$$

But how are categories related? Can we define mappings between them? The following definition, of a 'functor', is precisely that, to perform an operation on two categories. The idea is to take objects into objects and arrows into arrows.

Definition 2.1.3 (Functor). Let $\mathfrak{C}$ and $\mathfrak{D}$ be categories. A **functor** $F : \mathfrak{C} \longrightarrow \mathfrak{D}$ consists of:

1. A mapping from objects in $\mathfrak{C}$ to objects in $\mathfrak{D}$: $\mathrm{Ob}(\mathfrak{C}) \longrightarrow \mathrm{Ob}(\mathfrak{D})$;
2. A mapping from morphisms in $\mathfrak{C}$ to morphisms in $\mathfrak{D}$ such that if $f \in \hom_{\mathfrak{C}}(A, B)$, then $F(f) \in \hom_{\mathfrak{D}}(F(A), F(B))$.

Satisfying:

1. Identity is preserved: $F(1_A) = 1_{F(A)}$;
2. Composition is preserved: $F(g \circ f) = F(g) \circ F(f)$.

The functors defined above are often called **covariant functors**. If we 'invert the arrows', we obtain **contravariant functors**. This notion of inverting arrows leads us to the following definition.

Definition 2.1.4 (Dual Category). For every category $\mathfrak{C}$, we may form a new category $\mathfrak{C}^{\mathrm{op}}$ called the **dual category**. Its objects are the same as those of $\mathfrak{C}$, but its morphisms are 'reversed', i.e.

$$\hom_{\mathfrak{C}^{\mathrm{op}}}(A, B) = \hom_{\mathfrak{C}}(B, A)$$

And the composition $g \circ f$ of morphisms in $\mathfrak{C}$ corresponds to the composition $f \circ g$ of the same morphisms in $\mathfrak{C}^{\mathrm{op}}$.

We generalize it further and define a map between functors.

Definition 2.1.5 (Natural Transformation). Let $\mathfrak{C}$ and $\mathfrak{D}$ be categories and $\mathfrak{C} \overset{F}{\underset{G}{\rightrightarrows}} \mathfrak{D}$ be functors. A **natural transformation** $\lambda : F \longrightarrow G$ is a family

$$\left(F(A) \xrightarrow{\lambda_A} G(A) \right)_{A \in \mathfrak{C}}$$

of morphisms in $\mathfrak{D}$ such that for every map $f : A \longrightarrow A'$ in $\mathfrak{C}$, the square

$$\begin{array}{ccc} F(A) & \xrightarrow{F(f)} & F(A') \\ \downarrow{\scriptstyle \lambda_A} & & \downarrow{\scriptstyle \lambda_{A'}} \\ G(A) & \xrightarrow[G(f)]{} & G(A') \end{array}$$

commutes. The morphisms λ_A are called the **components** of λ.

2.2 Quotient Spaces

Let S be a subspace of a vector space V. Recalling the modular arithmetic, it is easy to see that the binary relation on V defined binary

$$u \equiv v \iff u - v \in \mathrm{S}$$

is an equivalence relation. We say that u and v are **congruent modulo** S.

Now notice that

$$\begin{aligned}[v] &= \{u \in \mathrm{V} : u \equiv v\} \\ &= \{u \in \mathrm{V} : u - v \in \mathrm{S}\} \\ &= \{u \in \mathrm{V} : u = v + s, \text{ for some } s \in \mathrm{S}\} \\ &= \{v + s : s \in \mathrm{S}\} \\ &= v + \mathrm{S}\end{aligned}$$

The set

$$[v] = v + \mathrm{S} = \{v + s : s \in \mathrm{S}\}$$

is called a **coset** or **affine subset** of S in V.

Example 2.2.1. The solution set of a linear system

$$\mathrm{C} = \{x : \mathrm{A}x = b\}$$

is an affine subspace, with $\mathrm{A} \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$.

Definition 2.2.1 (Quotient Space)**.** The set of all cosets (or classes) of S in V, denoted by

$$\mathrm{V}/\mathrm{S} = \{v + s : v \in \mathrm{V}\}$$

is called the **quotient space of** V **modulo** S.

Naturally, we define addition and scalar multiplication as follows

$$(u + \mathrm{S}) + (v + \mathrm{S}) = (u + v) + \mathrm{S} \text{ and } \lambda(u + \mathrm{S}) = \lambda u + \mathrm{S}$$

which are well defined.

Theorem 2.2.1. The quotient space of V modulo S is a vector space over $\mathbb{F}$ with the operations

$$\lambda(u + \mathrm{S}) = \lambda u + \mathrm{S}$$

$$(u + \mathrm{S}) + (v + \mathrm{S}) = (u + v) + \mathrm{S}$$

Definition 2.2.2 (Natural Projection)**.** If S is a subspace of V, we may define the mapping

$$\pi_{\mathrm{S}} : \mathrm{V} \longrightarrow \mathrm{V}/\mathrm{S}$$

which sends every vector to the coset containing it, i.e., the class associated to it. This map is called the **natural projection** or **canonical projection**.

Theorem 2.2.2. The canonical projection π_S is a surjective linear mapping with $\ker(\pi_S) = S$.

Theorem 2.2.3 (The Correspondence Theorem)**.** Let S be a subspace of V. Then the function that assigns each subspace $S \subset T \subset V$, the subspace T/S of V/S is an order-preserving one-to-one correspondence between the set of all subspaces of V containing S and the set of all subspaces of V/S.

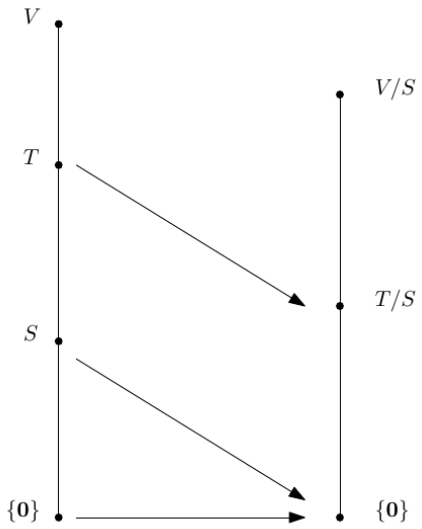


Figure 2.1: Correspondence between V and V/S.

Definition 2.2.3 (Codimension)**.** Let W be a subspace of V. Then the **codimension** of W in V is

$$\operatorname{codim}_V W = \dim V/W$$

Theorem 2.2.4. Suppose $\dim V = n$ and W is a subspace of V with $\dim W = k$. Then

$$\dim V/W = \operatorname{codim}_V W = n - k$$

Definition 2.2.4 (Cokernel and Coimage)**.** Let $T : V \longrightarrow W$ a linear transformation. Then the **cokernel** of T is the quotient space

$$\operatorname{coker}(T) = W/\operatorname{Im}(T)$$

And the **coimage** of T is defined as

$$\operatorname{coim}(T) = V/\operatorname{Ker}(T)$$

Corollary 2.2.5. Let V be a finite-dimensional vector space $T : V \longrightarrow V$ a linear transformation. Then

$$\dim(\ker(T)) = \dim(\operatorname{coker}(T))$$

2.3 Dual Space

A concept that will help us in the study of subspaces, linear equations, and coordinates is the following.

Definition 2.3.1 (Linear Functional and Dual Space)**.** A linear transformation from the vector space V to its scalar field $\mathbb{F}$ is called a **linear functional**.

The set of linear functionals is denoted by V^* and is called **dual space** of V. In other words, $V^* = \mathscr{L}(V, \mathbb{F})$.

Linear functionals are also called a **form** or a **1-form**.

Example 2.3.1. Let $(c_1, \ldots, c_n) \in \mathbb{F}^n$ and define $f : \mathbb{F}^n \longrightarrow \mathbb{F}$ by

$$f(x_1, \ldots, x_n) = c_1 x_1 + \ldots + c_n x_n$$

Then f is a linear functional on $\mathbb{F}^n$.

Example 2.3.2 (Trace)**.** If A is an $n \times n$ matrix, the **trace** of A is the scalar

$$\operatorname{tr} A = A_{11} + A_{22} + \ldots + A_{nn}$$

Remark that the trace function is a linear functional on the matrix space $\mathbf{M}_n$.

Remark. Suppose V is finite-dimensional. Then the dimension of the dual space is equal to the dimension of the space.

$$\dim V^* = \dim V$$

If V is infinite-dimensional, then $\dim(V^*) > \dim(V)$.

Definition 2.3.2 (Dual basis)**.** If $\beta = \{v_1, \ldots, v_n\}$ is a basis of V then the **dual basis** of β is the set $\beta^* = \{f_1, \ldots, f_n\}$, where each f_i is the linear functional on V such that

$$f_i(v_j) = \delta_{ij}$$

where δ is the Kronecker's delta.

Theorem 2.3.1. Let V be a finite-dimensional vector space. Then the dual basis of a basis of V is a basis of V^*.

Proof. Let $\beta = \{v_1, \ldots, v_n\}$ be a basis for V. Then there exists a unique linear functional f_i on V such that

$$f_i(v_j) = \delta_{ij}$$

for each i.

With this process, we obtain n distinct linear functionals $f_1, \ldots, f_n$ on V.

To show that $f_1, \ldots, f_n$ are linearly independent, suppose that $c_1, \ldots, c_n \in \mathbb{F}$ are such that

$$c_1 f_1 + \ldots + c_n f_n = 0$$

Since $(c_1f_1 + \ldots + c_nf_n)(v_j) = c_j$ for each $j = 1, \ldots, n$, we know that $c_1 = \ldots = c_n = 0$. Hence, $f_1, \ldots, f_n$ is linearly independent.

And given that $\dim V^* = n$, the set $\beta^* = \{f_1, \ldots, f_n\}$ is a basis for V^*. □

Theorem 2.3.2. Let $\beta = \{v_1, \ldots, v_n\}$ be a basis for a vector space V. Then there is a unique dual basis $\beta^* = \{f_1, \ldots, f_n\}$ for V^* such that $f_i(v_j) = \delta_{ij}$.

For each linear functional f on V we have

$$f = \sum_{i=1}^{n} f(v_i) f_i$$

and for each vector $v \in V$ we have

$$v = \sum_{i=1}^{n} f_i(v) v_i$$

Proof. The last proof established that there is a unique basis which is 'dual' to β. Let f be a linear functional on V. Then f is a linear combination of the f_i, so the scalars $c_j = f(v_j)$. Now, if

$$v = \sum_{i=1}^{n} x_i v_i$$

is a vector in V, then

$$f_j(v) = \sum_{i=1}^{n} x_i f_j(v_i) = \sum_{i=1}^{n} x_i \delta_{ij} = x_j$$

so v has a unique expression as a linear combination of v_i given by

$$v = \sum_{i=1}^{n} f_i(v) v_i$$

□

Note that f_i are coordinate functions for β, given that f_i assigns to each vector $v \in V$ the ith coordinate of v relative to the ordered basis β.

How are linear functionals and subspaces related? If f is a non-zero linear functional, then the rank of f is one. And if V is finite-dimensional, then by the Rank–Nullity theorem, the null space N_f has dimension

$$\dim N_f = \dim V - 1$$

In a vector space of dimension n, a subspace of of dimension $n-1$ is called a **hyperspace**, which is sometimes called **hyperplanes** or **subspaces of codimension one**. The hyperspace is always the null space of a linear functional.

Definition 2.3.3 (Annihilator)**.** Let V be a vector space over $\mathbb{F}$ and S a subset of V. Then the **annihilator** of S is the set S^0 of linear functionals f on V such that $f(v) = 0$ for every $v \in S$.

$$S^0 = \{f \in V^* : f(v) = 0, \forall v \in S\}$$

S^0 is a subspace of V^*. If $S = \{0\}$, then $S^0 = V^*$. If $S = V$, then S^0 is the zero subspace of V^*.

The next example shows an important procedure in the following proofs.

Example 2.3.3. Let $\{e_1, e_2, e_3, e_4, e_5\}$ be the standard basis of $\mathbb{R}^5$ and $\{f_1, f_2, f_3, f_4, f_5\}$ be the dual basis of $\mathbb{R}^5$. Suppose

$$W = \text{span}(e_1, e_2) = \{(x_1, x_2, 0, 0, 0) \in \mathbb{R}^5 : x_1, x_2 \in \mathbb{R}\}$$

We show that $W^0 = \text{span}(f_3, f_4, f_5)$.

Recall that f_j is the linear functional that selects the jth coordinate, i.e. $f_j(x_1, x_2, x_3, x_4, x_5) = x_j$.

First suppose $f \in \text{span}(f_3, f_4, f_5)$. Then there exist $c_3, c_4, c_5 \in \mathbb{R}$ such that $f = c_3 f_3 + c_4 f_4 + c_5 f_5$. If $(x_1, x_2, 0, 0, 0) \in W$, then

$$f(x_1, x_2, 0, 0, 0) = (c_3 f_3 + c_4 f_4 + c_5 f_5)(x_1, x_2, 0, 0, 0) = 0$$

Hence $f \in W^0$. I.e., $\text{span}(f_3, f_4, f_5) \subset W^0$.

Now suppose $f \in W^0$. Since the dual basis is a basis of $(\mathbb{R}^5)^*$, there exist $c_1, \ldots, c_5 \in \mathbb{R}$ such that $f = c_1 f_1 + c_2 f_2 + c_3 f_3 + c_4 f_4 + c_5 f_5$. Because $e_1 \in W$ and $f \in W^0$, we have

$$0 = f(e_1) = (c_1 f_1 + c_2 f_2 + c_3 f_3 + c_4 f_4 + c_5 f_5)(e_1) = c_1$$

Similarly, $e_2 \in W$ and thus $c_2 = 0$. Since $e_3, e_4, e_5 \notin W$, $f = c_3 f_3 + c_4 f_4 + c_5 f_5$. Thus $f \in \text{span}(f_3, f_4, f_5)$, i.e., $W^0 \subset \text{span}(f_3, f_4, f_5)$.

The next theorem states that each d-dimensional subspace of an n-dimensional space is the intersection of the null spaces of $(n-d)$ linear functionals.

Theorem 2.3.3. Let V be a finite-dimensional vector space and let W be a subspace of V. Then

$$\dim W + \dim W^0 = \dim V$$

Proof. Let $\{v_1, \ldots, v_k\}$ be a basis for W and choose vectors $\{v_{k+1}, \ldots, v_n \in V$ to extend to a basis $\{v_1, \ldots, v_n\}$ of V. And let $\{f_1, \ldots, f_n\}$ be the basis for V^* which is dual to this basis for V. We show that $\{f_{k+1}, \ldots, f_n\}$ is a basis for W^0.

For $i \geq k+1$, since $f_i(v_j) = \delta_{ij}$ and $\delta_{ij} = 0$ if $i \geq k+1$ and $j \leq k$, we know that f_i belongs to W^0. Hence, for $i \geq k+1$, $f_i(v) = 0$ whenever v is a linear combination of $v_1, \ldots, v_k$.

Given that the functionals $f_{k+1}, \ldots, f_n$ are linearly independent, all we need to show is that they span W^*. Suppose $f \in V^*$. Now

$$f = \sum_{i=1}^{n} f(v_i) f_i$$

implies that if $f \in W^0$, we have $f(v_i) = 0$ for $i \leq k$ and

$$f = \sum_{i=k+1}^{n} f(v_i) f_i$$

Therefore, W^0 has dimension $n-k$, as desired. □

The next corollary shows that if we select some select ordered basis for the space, each k-dimensional subspace can be described by specifying $(n-k)$ homogeneous linear conditions on

the coordinates relative to that basis.

Corollary 2.3.4. If W is a k-dimensional subspace of an n-dimensional vector space V, then W is the intersection of $(n-k)$ hyperspaces in V.

Corollary 2.3.5. If W_1 and W_2 are subspaces of a finite-dimensional vector space, then $W_1 = W_2$ iff. $W_1^0 = W_2^0$.

This theory provides a 'dual' point of view on the system of equations, showing how annihilators are related to systems of homogeneous linear equations.

Example 2.3.4. Let W be the subspace of $\mathbb{R}^5$ spanned by the vectors $v_1 = (2,-2,3,4,-1)$, $v_2 = (-1,1,2,5,2)$, $v_3 = (0,0,-1,-2,3)$, and $v_4 = (1,-1,2,3,0)$.

To find the annihilator W^0 of W, we first form a matrix A with row vectors v_1, v_2, v_3, v_4 and find the row-reduced echelon matrix R which is row-equivalent to A.

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

Now, if f is a linear functional on $\mathbb{R}^5$,

$$f(x_1,\ldots,x_5) = \sum_{j=1}^{5} c_j x_j$$

and f is in W^0 iff. $f(v_i) = 0$, for $i = 1,2,3,4$.

This is equivalent to $Ac = 0$, where $c = (c_1,c_2,c_3,c_4,c_5)^t$. Which is, in turn, equivalent to $Rc = 0$. Or simply

$$\begin{aligned} c_1 - c_2 - c_4 &= 0 \\ c_3 + 2c_4 &= 0 \\ c_5 &= 0 \end{aligned}$$

By setting $c_2 = a$ and $c_4 = b$, we have $c_1 = a+b$, $c_3 = -2b$, $c_5 = 0$. So W^0 consists of all linear functionals of the form

$$f(x_1,x_2,x_3,x_4,x_5) = (a+b)x_1 + ax_2 - 2bx_3 + bx_4$$

The dimension of W^0 is two and a basis $\{f_1,f_2\}$ for it can be found by setting $a = 1, b = 0$, and then $a = 0, b = 1$:

$$\begin{aligned} f_1(x_1,\ldots,x_5) &= x_1 + x_2 \\ f_2(x_1,\ldots,x_5) &= x_1 - 2x_3 + x_4 \end{aligned}$$

And the general form of $f \in W^0$ is $f = af_1 + bf_2$.

2.4 The Double Dual

Is every basis for V^* the dual of some basis for V? To answer that question we consider V^{**}, the dual space of V^*.

Let $v \in V$. Then v **induces** a linear functional φ_v on V^* defined by

$$\varphi_v(f) = f(v), \text{ where } f \in V^*$$

It is easy to check that φ_v is linear simply using the definition of linear operations in V^*.

$$\begin{aligned}\varphi_v(cf+g) &= (cf+g)(v) = (cf)(v) + g(v)\\ &= cf(v) + g(v) = c\varphi_v(f) + \varphi_v(g)\end{aligned}$$

Theorem 2.4.1. Let V be a finite-dimensional vector space. For each vector $v \in V$ define

$$\varphi_v(f) = f(v), \text{ where } f \in V^*$$

The mapping $v \longrightarrow \varphi_v$ is an isomorphism of V onto V^{**}.

Proof. Suppose that $v, u \in V$ and $c \in \mathbb{F}$ and let $w = cv + u$. Then for each $f \in V^*$,

$$\begin{aligned}\varphi_w(f) &= f(w) = f(cv+u) = cf(v) + f(u)\\ &= c\varphi_v(f) + \varphi_u(f)\end{aligned}$$

Hence, $v \longrightarrow \varphi_v$ is a linear mapping from V to V^{**}.

Notice that $\varphi_v = 0$ iff. $v = 0$ by linearity, i.e., φ_v is a non-singular linear transformation.

And given that

$$\dim V^{**} = \dim V^* = \dim V$$

we know that φ_v is bijective. Hence, this linear mapping is an isomorphism of V onto V^{**}. □

Corollary 2.4.2. Let V be a finite-dimensional vector space. If L is a linear functional on the dual space V^*, then there is a unique vector $v \in V$ such that $L(f) = f(v)$, for every $f \in V^*$.

Corollary 2.4.3. Let V be a finite-dimensional vector space. Each basis for V^* is the dual of some basis for V.

Proof. Let $\beta^* = \{f_1, \ldots, f_n\}$ be a basis for V^*. Then there exists a basis $\{L_1, \ldots, L_n\}$ for V^{**} such that $L_i(f_j) = \delta_{ij}$.

By the previous corollary, for each index i there exists a vector $v_i \in V$ such that $L_i(f) = f(v_i)$, for every $f \in V^*$. Then $\{v_1, \ldots, v_n\}$ is a basis for V and β^* is the dual of this basis. □

In this corollary we see that V and V^* are naturally in duality with one another. Each is the dual space of the other.

If E is a subset of V^*, then the annihilator E^0 is a subset of V^{**}. We know that each subspace W is determined by its annihilator W^0. This is the case because W is the subspace annihilated by all $f \in W^0$, i.e., the intersection of the null spaces of all $f \in W^0$. We state this fact in the following identity

$$W = (W^0)^0$$

Theorem 2.4.4. If S is any subset of a finite-dimensional vector space V, then $(S^0)^0$ is the subspace spanned by S.

Proof. Let W be the subspace generated by S. Clearly $W^0 = S^0$. We prove that $W = W^{00}$. By a previous theorem,

$$\dim W + \dim W^0 = \dim V$$
$$\dim W^0 + \dim W^{00} = \dim V^*$$

Therefore,

$$\dim W = \dim W^{00}$$

And since W is a subspace of W^{00}, we have that $W = W^{00}$. □

The results of this section hold for arbitrary vector spaces using Axiom of Choice. In particular, we redefine **hyperspaces** in order to include the infinite dimensional case.

The idea is that a space N falls just one dimension short of filling out V, using the concept of maximal.

Definition 2.4.1 (Hyperspace). If V is a vector space, a **hyperspace** in V is a maximal proper subspace of V.

Put another way, let W be a proper subspace of V. If there isn't a subspace U such that $W \subsetneq U \subsetneq V$, then W is a hyperplane.

Theorem 2.4.5. If f is a non-zero linear functional on the vector space V, then the null space of f is a hyperspace in V. Every hyperspace in V is the null space of a linear functional on V.

Proof. Let f be a non-zero linear functional on V and N_f its null space. Let $v \in V$ such that $v \notin N_f$, i.e., $f(v) \neq 0$.

Note that the subspace spanned by N_f and v consists of all vectors of the form $w + cv$, where $w \in N_f$, $c \in \mathbb{F}$.

Let $u \in V$. Define

$$c = \frac{f(u)}{f(v)}$$

Then the vector $w = u - cv \in N_f$, since

$$f(w) = f(u - cv) = f(u) - cf(v) = 0$$

Hence, every vector $u \in V$ is in the subspace spanned by N_f and v, showing that the null space of f is a hyperspace in V.

Let N be a hyperspace in V and fix $v \notin N$. Since N is a maximal proper subspace, the subspace spanned by N and v is the entire space V. Therefore, each vector $u \in V$ has the form $u = w + cv$, where $w \in N, c \in \mathbb{F}$.

Notice that

$$u = w' + c'v \implies (c' - c)v = w - w'$$

and

$$c' - c \neq 0 \implies v \in N$$

Thus, $c' = c$ and $w' = w$. In other words, if u is in V, there is a unique scalar c such that $u - cv \in \mathrm{N}$. Call that scalar $c = g(u)$. Then g is a linear functional on V and N is the null space of g. □

Lemma 2.4.6. If f and g are linear functionals on a vector space V, then g is a scalar multiple of f iff. the null space of g contains the null space of f.

Proof. ($\Rightarrow$) Is immediate.

($\Leftarrow$) If $f = 0$, then $g = 0$ and g is a scalar multiple of f. Suppose that $f \neq 0$ so that its null space N_f is a hyperspace in V. Choose $v \in \mathrm{V}$ with $f(v) \neq 0$ and define

$$c = \frac{g(v)}{f(v)}$$

The linear functional $h = g - cf$ is zero on N_f (since both f and g are zero there) and $h(v) = g(v) - cf(v) = 0$.

Thus, h is zero on the subspace spanned by n_f and v, which is exactly V. Therefore, $h = 0$ and $g = cf$. □

Theorem 2.4.7. Let $g, f_1, \ldots, f_r$ be linear functionals on a vector space V with respective null spaces $\mathrm{N}, \mathrm{N}_1, \ldots, \mathrm{N}_r$. Then g is a linear combination of $f_1, \ldots, f_r$ iff. the null space of g contains the intersection of the null spaces of $f_1, \ldots, f_r$, i.e.,

$$\mathrm{N}_1 \cap \ldots \cap \mathrm{N}_r \subset \mathrm{N}$$

Proof. ($\Rightarrow$) If $g = c_1 f_1 + \ldots + c_r f_r$ and $f_i(v) = 0$ for each i, then clearly $g(v) = 0$. Hence, N contains $\mathrm{N}_1 \cap \ldots \cap \mathrm{N}_r$.

($\Leftarrow$) This proof is by induction on the index r. The preceeding lemma handles the case $r = 1$. Suppose that the claim is true for $r = k - 1$, i.e., $\mathrm{N}_1 \cap \ldots \cap \mathrm{N}_k \subset \mathrm{N}$.

Let $g', f_1', \ldots, f_{k-1}'$ be the restriction of $g, f_1, \ldots, f_{k-1}$ to the subspace N_k. If $v \in \mathrm{N}_k$ and $f_i'(v) = 0$ for i ranging from one to $k - 1$, then $v \in \mathrm{N}_1 \cap \ldots \cap \mathrm{N}_r$ and so $g'(v) = 0$. By the induction hypothesis, there are scalars c_i such that

$$g' = c_1 f_1' + \ldots + c_{k-1} f_{k-1}'$$

And let

$$h = g - \sum_{i=1}^{k-1} c_i f_i$$

Then h is a linear functional on V and $h(v) = 0$ for every $v \in \mathrm{N}_k$. By the previous lemma, h is a scalar multiple of f_k. If $h = c_k f_k$, then

$$g = \sum_{i=1}^{k} c_i f_i$$

□

2.5 The Transpose of a Linear Transformation

Definition 2.5.1 (Transpose). The **transpose** of a linear transformation $T : V \longrightarrow W$ is the mapping $T^t : W^* \longrightarrow V^*$ such that

$$(T^t g)(v) = g(T(v)) = g \circ T$$

for every $g \in W^*$ and $v \in V$.

Theorem 2.5.1. The transpose $T^t : W^* \longrightarrow V^*$ of a linear transformation $T : V \longrightarrow W$ is a linear transformation.

Proof. First we show that $T^t f \in V^*$. In fact, for all $f \in W^*$,

$$\begin{aligned} T^t f(au + bv) &= f(T(au + bv)) = f(aT(u) + bT(v)) \\ &= af(T(u)) + bf(T(v)) = aT^t f(u) + bT^t f(v) \end{aligned}$$

Hence, $T^t f \in V^*$, as desired.

Now we show that T^t is linear.

$$\begin{aligned} T^t(af + bg)(v) &= (af + bg)T(v) = afT(v) + bgT(v) \\ &= a(T^t f)(v) + b(T^t g)(v) \end{aligned}$$

I.e., that $T^t(af + bg) = aT^t f + bT^t g$. □

Theorem 2.5.2. Let $T : V \longrightarrow W$ be a linear transformation. The null space of T^t is the anihhilator of the range of T.

Moreover, if V and W are finite-dimensional, then

1. $\operatorname{rank}(T^t) = \operatorname{rank}(T)$;
2. The range of T^t is the annihilator of the null space of T, i.e., $\operatorname{Im} T^t = (\ker T)^0$.
3. The null space of T^t is the annihilator of the of the range of T, i.e., $\ker T^t = (\operatorname{Im} T)^0$.

Proof. 3. Notice that

$$\ker T^t = \{f \in W^* : T^t f = 0\} = \{f \in W^* : fT = 0\}$$

and

$$(\operatorname{Im} T)^0 = \{f \in W^* : f(\operatorname{Im} T) = 0\} = \{f \in W^* : fT = 0\}$$

Thus, $\ker T^t = (\operatorname{Im} T)^0$.

1. We know that

$$\dim(\operatorname{Im} T) + \dim(\operatorname{Im} T)^0 = \dim W$$

and

$$\dim(\ker T^t) + \dim(\operatorname{Im} T^t) = \dim W^*$$

However, $\dim W = \dim W^*$ and $\dim(\operatorname{Im} T)^0 = \dim(\ker T^t)$. Hence, $\dim(\operatorname{Im} T) = \dim(\operatorname{Im} T^t)$.

2. First, we'll prove that $\text{Im } T^t \subseteq (\ker T)^0$.

Consider $f \in \text{Im } T^t \subseteq V^*$, i.e., $f = T^t h$, $h \in W^*$. For all $v \in \ker T$,

$$f(v) = (T^t h)(v) = h(T(v)) = h(0) = 0$$

I.e., $f \in (\ker T)^0$.

Now, it is sufficient to show that $\dim(\ker T)^0 = \dim(\text{Im } T^t)$. Since

$$\dim(\ker T)^0 + \dim(\ker T) = \dim V$$

and

$$\dim(\text{Im } T) + \dim(\ker T) = \dim V$$

we have that $\dim(\ker T)^0 = \dim(\text{Im } T)$.

Therefore, $\text{Im } T^t = (\ker T)^0$. □

Theorem 2.5.3. Let V and W be finite-dimensional vector spaces. Let β be an ordered basis for V with dual basis β^* and γ be an ordered basis for W with dual basis γ^*. And let $T : V \longrightarrow W$.

If A is the matrix of T relative to β, γ, and B is the matrix of T^t relative to γ^*, β^*, then

$$B_{ij} = A_{ji}$$

Put another way,

$$[T^t]_{\gamma^*,\beta^*} = ([T]_{\beta,\gamma})^t$$

Proof. See [Mou22]'s Proposition 9.2.3. □

Definition 2.5.2 (Transpose). If $A \in \mathbf{M}_{m\times n}(\mathbb{F})$, the **transpose** of A is the $n \times m$ matrix A^t defined by $A^t_{ij} = A_{ji}$.

Theorem 2.5.4. Let $A \in \mathbf{M}_{m\times n}(\mathbb{F})$. Then the row rank of A is equal to the column rank of A.

Chapter 3

Polynomials

3.1 Algebras

Definition 3.1.1 (Linear algebra over a field)**.** Let $\mathbb{F}$ be a field. A **linear algebra over the field** $\mathbb{F}$ is a vector space V over $\mathbb{F}$ with an additional operation called **multiplication of vectors**, which associates with each pair of vectors $u, v \in \mathrm{V}$ a vector $uv \in \mathrm{V}$ in such a way that

1. Multiplication is associative, $u(vw) = (uv)w$;
2. Is distributive with respect to addition, $u(v+w) = uv + uw$ and $(u+v)w = uw + vw$;
3. For each $a \in \mathbb{F}$, $a(uv) = (au)v = a(uv)$.

If there is an element $1 \in \mathrm{V}$ such that $1v = v1 = v$ for all $v \in \mathrm{V}$, then V is a **linear algebra with identity over** $\mathbb{F}$. The algebra V is called **commutative** if $uv = vu$ for all $u, v \in \mathrm{V}$.

Example 3.1.1 (Algebra of formal power series)**.** The algebra $\mathbb{F}^\infty$ is called the **algebra of formal power series**. The element $f = (f_0, f_1, f_2, \ldots)$ is frequentlyt written as

$$f = \sum_{n=0}^{\infty} f_n x^n$$

Notice that $x = (0, 1, 0, \ldots, 0, \ldots)$, and x^n is equal to one at the nth position (recall that the index starts at zero) and zero elsewhere.

3.2 The Algebra of Polynomials

Definition 3.2.1 (Polynomial)**.** Let $\mathbb{F}[x]$ be the subspace of $\mathbb{F}^\infty$ spanned by the vectors $1, x, x^2, \ldots$. An element of $\mathbb{F}[x]$ is called a **polynomial** over $\mathbb{F}$.

The **degree** of a polynomial, denoted by $\deg \mathrm{F}$, is the largest integer n such that $f_n \neq 0$ and such that $f_k = 0$ for all integers $k > n$. If f is a non-zero polynomial of degree n, then it can be written as

$$f = f_0 x^0 + x_1 x + f_2 x^2 + \ldots + f_n x^n$$

Theorem 3.2.1. Let f and g be non-zero polynomials over $\mathbb{F}$. Then,

1. fg is a non-zero polynomial;
2. $\deg(fg) = \deg f + \deg g$;
3. fg is a monic polynomial if both f and g are monic polynomials;
4. fg is a scalar polynomial if and only if both f and g are scalar polynomials;
5. If $f+g \neq 0$, then $\deg(f+g) \leq \max(\deg f, \deg g)$.

Corollary 3.2.2. The set of all polynomials over a given field is a commutative linear algebra with identity.

Corollary 3.2.3. Suppose f, g, and h are polynomials such that $f \neq 0$ and $fg = fh$. Then $g = h$.

Definition 3.2.2. We shall denote the identity of a linear algebra A by 1 and make the convention that $v^0 = 1$ for all $v \in \mathrm{A}$. Then to each polynomial $f = \sum_{i=0}^{n} f_i x^i$ and $v \in \mathrm{A}$, we associate an element $f(v) \in \mathrm{A}$ by the rule

$$f(v) = \sum_{i=0}^{n} f_i v^i$$

Theorem 3.2.4. Let A be a linear algebra with identity. Suppose f and g are polynomials, $v \in \mathrm{A}$ and a is a scalar. Then

1. $(cf+g)(v) = cf(v) + g(v)$;
2. $(fg)(v) = f(v)g(v)$.

3.3 Polynomial Ideals

Definition 3.3.1 (Division and Irreducible Elements)**.** We say that $g\mathbb{F}[x]$ **divides** $f\mathbb{F}[x]$, and denote this by $g \mid f$, if there exists $q\mathbb{F}[x]$ such that $f = qg$.

And g is called **irreducible** (or **prime**) if g is not scalar and it's only monic divisors are 1 and g.

Note that these 'irreducible' and 'prime' mean the same thing in the polynomial ring, but not in general ring theory.

Remark. Every polynomial of degree one is prime.

Example 3.3.1. The polynomial $f = t^2 + 1$ is prime over $\mathbb{R}$ but not over $\mathbb{C}$, since $(t-i) \mid f$.

Definition 3.3.2 (Algebraically Closed)**.** A field $\mathbb{F}$ is **algebraically closed** if every prime of the polynomial ring $\mathbb{F}[x]$ has degree one.

Definition 3.3.3 (Polynomial Ideal)**.** An **ideal** in the polynomial algebra $\mathbb{F}[x]$ is a subspace M of $\mathbb{F}[x]$ such that fg belong to M whenever f is in $\mathbb{F}[x]$ and g is in M.

More generally,

Definition 3.3.4 (Ideal)**.** Let A be a ring. If $\mathrm{I} \subseteq \mathrm{A}$, $\mathrm{I} \neq \emptyset$, then I is called an **ideal** of A if the following properties hold

- Closure under addition: $\forall x, y \in \mathrm{I}, x + y \in \mathrm{I}$.
- Absorption property: $\forall x \in \mathrm{I}, \forall a \in \mathrm{A}, ax \in \mathrm{I}$.

This definition is equivalent to saying that, given I non-empty, a linear combination $a_1x_1 + \ldots a_rx_r$ of elements $x_i \in \mathrm{I}$ with coefficients $a_i \in \mathrm{A}$ is in I.

For example, $n\mathbb{Z} := \{zn \mid z \in \mathbb{Z}\}$ is an ideal of the ring of integers (where n is a non-negative integer).

Definition 3.3.5 (Generated Ideal)**.** The **ideal generated by** a set of elements $a_1, \ldots, a_n \in \mathrm{A}$ is the smallest ideal containing these elements.

The ideal

$$\mathrm{M} = p\mathbb{F}[x]$$

where p is a fixed polynomial, is called the **principal ideal generated by** p.

3.4 Prime Factorization

Can polynomials be factored? If so, how?

Follows from the Euclid's Division Algorithm.

There exists unique q, r such that

$$f = qg + r, \quad \deg(r) < \deg(g)$$

Remark that $g \mid f$ iff. $r \equiv 0$.

Definition 3.4.1 (Prime Factor and Multiplicity)**.** If g is prime, monic and $g \mid f$, we say that g is a **prime factor** of f.

The **multiplicity** of g, as a prime factor of f, is defined as

$$\max\{m \in \mathbf{Z}_{\geq 0} : g^m \mid f\}$$

Theorem 3.4.1. If $f \in \mathbb{F}[x] \setminus \{0\}$, then the set of prime factors is finite.

If $\{g_1, \ldots, g_k\}$ is the set of prime factors of f, there exists a unique $u \in \mathbb{F} \setminus \{0\}$ such that $f = ug_1^{m_1} \ldots g_k^{m_k}$, where m_j is the multiplicity of g_j as a prime factor of f.

A monic polynomial g is the **greatest common divisor (GCD)** if $f_1, \ldots, f_k \in \mathbb{F}[x] \setminus \{0\}$ if

$$g \in \bigcap_{j=1}^{k} \mathrm{Div}(f_j) \text{ and } h \mid g \quad \forall h \in \bigcap_{j=1}^{k} \mathrm{Div}(f_j)$$

Theorem 3.4.2. If $g = \text{mdc}(f_1, \ldots, f_k)$, there exist $g_j \in \mathbb{F}[x]$, $1 \leq j \leq k$ such that $g = g_1 f_1 + \ldots + g_k f_k$.

Chapter 4

Canonical Forms

Our goal in this chapter is to break a vector space into 'important' subspaces with respect to a polynomial associated with a given linear operator.

Here we'll use V to denote a vector space over $\mathbb{F}$, $T \in \text{End}(V)$, and $V_p := \{v \in V : p(T)(v) = 0\}$, i.e., the nullspace of the operator, where $p \in \mathbb{F}[x]$.

4.1 Annihilating Polynomials

Definition 4.1.1 (Annihilator Set)**.** We define $\mathfrak{A}_T := \{p \in \mathbb{F}[x] : p(T) = 0\}$ as the **annihilator set** of T.

It follows that the collection of polynomials p which annihilate T is an ideal in the polynomial algebra $\mathbb{F}[x]$.

We'll state the following theorem and proceed with a discussion before proving each of its parts.

Theorem 4.1.1 (Existence of the Minimal Polynomial)**.**

1. If dim(V) is finite, then $\mathfrak{A}_T \neq \{0\}$.
2. If $\mathfrak{A}_T \neq \{0\}$, then there exists a unique monic polynomial $m_T \in \mathbb{F}[x]$ such that m_T divides every element of $\mathfrak{A}_T$.

Definition 4.1.2 (Minimal Polynomial)**.** The polynomial m_T in the previous theorem is called **minimal polynomial** of T.

If $\mathfrak{A}_T = \{0\}$, we define $m_T = 0$.

Since every polynomial ideal consists of all multiples of some fixed monic polynomial (the generator of the ideal), we may define the minimal polynomial for T as the unique monic generator of the ideal of polynomials over $\mathbb{F}$ which annihilate T.

Summarizing, the minimal polynomial p for the linear operator T is uniquely determined by these three properties:

1. p is a monic polynomial over $\mathbb{F}$;

2. $p(\mathrm{T}) = 0$;

3. No polynomial over $\mathbb{F}$ which annihilates T has smaller degree than p has.

These definitions can be easily extended to a matrix A instead of an operator T. Moreover, it follows from previous remarks that similar matrices have the same minimal polynomial.

Theorem 4.1.2. The characteristic and minimal polynomials have the same roots, except for multiplicities.

Proof. Let $\mathrm{T} \in \mathrm{End}(\mathrm{V})$, where $\dim \mathrm{V} = n$, m_{T} the minimal polynomial for T and λ a scalar. We'll show that $m_{\mathrm{T}}(\lambda) = 0$ iff. λ is a characteristic value of T.

Suppose that $m_{\mathrm{T}}(\lambda) = 0$. Then write $m_{\mathrm{T}} = (x-\lambda)q(x)$. By the minimality of m_{T}, it follows that $q(\mathrm{T}) \neq 0$ and, therefore, there exists $u \in \mathrm{V}$ such that $q(\mathrm{T})(u) \neq 0$.

If $v := q(\mathrm{T})(u)$,

$$0 = m_{\mathrm{T}}(\mathrm{T})(u) = (\mathrm{T} - \lambda \mathrm{I})(q(\mathrm{T})(u)) = (\mathrm{T} - \lambda \mathrm{I})(v)$$

and hence v is an eigenvector of T associated with λ. Thus, $c_{\mathrm{T}}(\lambda) = 0$.

Suppose that λ is an eigenvalue. Then we can write $m_{\mathrm{T}}(\mathrm{T})v = m_{\mathrm{T}}(\lambda)v$, which implies that $m_{\mathrm{T}}(\lambda) = 0$. □

The Cayley-Hamilton Theorem will narrow the search for the minimal polynomial of various operators. Before proving it, we need the next lemma.

Lemma 4.1.3. If T is a diagonalizable linear operator, then the minimal polynomial for T is a product of distinct linear factors.

Proof. Follows from

$$m_{\mathrm{T}} = (\lambda - \lambda_1) \cdots (\lambda - \lambda_k)$$

□

Theorem 4.1.4 (Cayley-Hamilton)**.** Let $\mathrm{T} \in \mathrm{End}(\mathrm{V})$, where V is finite-dimensional. If c_{T} is a characteristic polynomial for T, then

$$c_{\mathrm{T}}(\mathrm{T}) = 0$$

Put another way, the minimal polynomial divides the characteristic polynomial for T.

Proof.

1. Take a basis β of V and write $\mathrm{A} = [\mathrm{T}]_\beta$.
2. Consider $\mathrm{A}' = x\mathrm{I} - \mathrm{A}$ and notice that $c_{\mathrm{T}}(x) = \det \mathrm{A}'$.
3. Let B be the adjoint matrix of A'. Note that its elements are polynomials of x of degree at most $n-1$.
4. Write $b_{ij} = b_{ij}^{(0)} + b_{ij}^{(1)}x + \ldots + b_{ij}^{(n-1)}x^{n-1}$.
5. Let $\mathrm{B}^{(k)}$ be the matrix whose entries are given by $b_{ij}^{(k)}$.
6. Write $c_{\mathrm{T}}(x) = a_0 + a_1 x + \ldots + x^n$ and use that

$$\mathrm{BA}' = \mathrm{adj}(\mathrm{A}')\mathrm{A}' = (\det \mathrm{A}')\mathrm{I} = c_{\mathrm{T}}(x)\mathrm{I}$$

7. It follows that $(B^{(0)} + B^{(1)}x + \ldots + B^{(n-1)}x^{n-1})(xI - A) = (a_0 + a_1x + \ldots + x^n)I$.

8. Comparing each coeficient, we have that

$$\begin{cases} a_0 I & = -B^{(0)}A \\ a_1 I & = B^{(0)} - B^{(1)}A \\ & \vdots \\ a_{n-1}I & = B^{(n-2)} - B^{(n-1)}A \\ I & = B^{(n-1)} \end{cases}$$

9. Multiplying these equations by $I, A, A^2, \ldots, A^n$ respectively, and adding all of them, we have

$$c_T(T) = a_0 I + a_1 A + \ldots + A^n = 0$$

□

The next example shows that it is possible to have $\mathfrak{A}_T = \{0\}$ when the dimension is infinite.

Example 4.1.1. If $V = \mathbb{F}[x]$ and T is the operator $f(t) \mapsto tf(t)$, then $\mathfrak{A}_T = \{0\}$.

Theorem 4.1.5 (Primary Decomposition Theorem (PDT)). Suppose that $\mathfrak{A}_T \neq \{0\}$ and let $p_1, \ldots, p_m$ be the distinct irreducible factors of m_T in $\mathbb{F}[x]$. If k_j is the multiplicity of p_j in m_T, then

$$V = V_{p_1^{k_1}} \oplus \ldots \oplus V_{p_m^{k_m}}$$

The polynomials $p_j^{k_j}$ are called the **primary factors (pf)** of T and $V_{p_j^{k_j}}$ is said to be a T-**primary subspace** of V.

In words, if there is non-zero polynomial in the annihilator, then there exists a minimal polynomial. Considering the prime factors of the minimal polynomial and its multiplicities, we can 'break' the vector space as direct sum of certain subspaces, where each subspace is related to one of the factors of the minimal polynomial to its multiplicity.

It follows immediately from the PDT that a linear operator is diagonalizable iff. its primary factors have degree one.

Recall that a subspace W of V is said to be T-invariant if $T(W) \subseteq W$. In particular, the restriction of T to W induces a linear operator in W given by $w \mapsto T(w)$ for all $w \in W$.

Lemma 4.1.6. If $S \in \mathrm{End}(V)$ satisfies $S \circ T = T \circ S$, then the nullspace of S, V_S, is T-invariant. In particular, V_p is T-invariant for all $p \in \mathbb{F}[x]$.

Proof. Let $v \in V_S$ and notice that $S(T(v)) = T(S(v)) = 0$, which shows that $T(v) \in V_S$. □

Generalizing the notion of eigenspace. To do that, we want to describe the following set

$$\{p \in \mathbb{F}[x] : V_p \neq \{0\}\} \tag{4.1}$$

We already know that the minimal polynomial is inside this set.

For any two polynomials f, p, then $V_p \subseteq V_{fp}$. In particular, $V_p^k \subseteq V_p^{k+1}$ for all $k \geq 0$. Therefore,

the set

$$V_p^\infty := \bigcup_{k>0} V_p^k$$

is a T-invariant subspace of V.

Definition 4.1.3 (Generalized Eigenspace)**.** For $p(t) = t-\lambda$, where $\lambda \in \mathbb{F}$, the space V_p^∞ is called the **generalized eigenspace** associated with λ.

Example 4.1.2. If $T \in \mathrm{End}(\mathbb{F}^2)$ is given by $T(x,y) = (y,0)$ and $p(t) = t$, then

$$V_p = V_T = [e_1] \quad \text{and} \quad V_p^2 = \mathbb{F}^2 \ \ (\text{since } T^2 = 0)$$

Is there any other polynomial such that $V_p \neq \{0\}$?

If $f \in \mathbb{F}[x]$ and $p \nmid f$, then $V_f = \{0\}$. In fact, if $f = pq + r$ is the division of f by p, then $r \in \mathbb{F} \setminus \{0\}$ (since p has degree one and $p \nmid f$) and

$$f(T)(x,y) = q(T)(T(x,y)) + r(x,y) = q(T)(y,0) + r(x,y) = (q(0)y + rx, ry)$$

Therefore, if $r \neq 0$,

$$F(T)(x,y) = (0,0) \iff 0 = y = x$$

The following proof, for the second item of the Theorem 4.1.1, shows the existence of minimal polynomial and how to find it.

Theorem 4.1.7. If $\mathfrak{A}_T \neq \{0\}$, then there exists a unique monic polynomial $m_T \in \mathbb{F}[x]$ such that m_T divides every element of $\mathfrak{A}_T$.

Proof. Let $m = \min\{k : \exists p \in \mathfrak{A}_T \setminus \{0\}, \deg(p) = k\} > 0$, i.e., the smallest degree of a non-zero polynomial in the annihilator.

Fix $f, p \in \mathfrak{A}_T$ with $\deg(p) = m$. By Euclid's Division Algorithm, $f = qp + r$, where $\deg(r) < m$. Notice that $r = f - qp \in \mathfrak{A}_T$. By the minimality of m, $r = 0$ and therefore $p \mid f$.

This shows that any non-constant polynomial with the minimal degree divides every other polynomial in the annihilator. And a polynomial divides another polynomial with the same degree if one is a scalar multiple of the other. □

To describe the set (4.1), it is sufficient to consider the following lemma and proposition.

Lemma 4.1.8. If $\gcd(f,g) = 1$, then the restriction of $f(T)$ to V_g is injective.

Proof. Let $p, q \in \mathbb{F}[x]$ such that $pf + qg = 1$. Then, for every $v \in V_g$,

$$v = (p(T)f(T) + q(T)g(T))(v) = (p(T)f(T))(v)$$

since $g(T)(v) = 0$. It follows that the restriction of $p(T) \circ f(T)$ to V_g is the identity function. Hence, the lemma follows. □

Proposition 4.1.9. If $\mathfrak{A}_T \neq \{0\}$ and $p \in \mathbb{F}[x]$ is irreducible, then $V_p \neq \{0\}$ iff. $p \mid m_T$.

Proof. Suppose that $p|m_T$ and $V_p = \{0\}$, i.e., $p(T)(u) \neq 0$ for all $u \in V \setminus \{0\}$. Consider $f = m_T/p$ and notice that

$$0 = m_T(T)(v) = p(T)(f(T)(v)) \quad \forall\, v \in V$$

If $f(T)(v) \neq 0$, then $p(T)(f(T)(v)) \neq 0$. Thus, $f \in \mathfrak{A}_T$, which is a contradiction since $\deg(f) < \deg(m_T)$.

Reciprocally, if $p \nmid m_T$, then $\gcd(p, m_T) = 1$, since p is irreducible. If follows from the lemma that the restriction of $m_T(T)$ to V_p is injective. Since $m_T(T) = 0$, we can conclude that $V_p = \{0\}$. □

Put another way, a prime polynomial lives in the set iff. it divides the minimal polynomial.

In a finite dimensional vector space, how can we compute the minimal polynomial? Proving the first item of the Theorem 4.1.1, we show an algorithm of how to do it.

Theorem 4.1.10. If $\dim(V) = n < \infty$, then $\mathfrak{A}_T \neq \{0\}$.

Proof. If $T = 0$, then $\mathfrak{A}_T = \{p \in \mathbb{F}[x] : p(0) = 0\} \neq \{0\}$. In this case, the minimal polynomial is $p(t) = t$.

Suppose that $T \neq 0$ and remember that $\dim(\mathrm{End}(V)) = n^2$. Then we can construct the family $(T^k)_k$, where $k = 0, 1, \ldots, n^2$, which is linearly dependent (since it contains $n^2 + 1$ elements).

Since the subfamily given by $\mathfrak{A}_T = I_V$ is linearly independent, there exists $1 \leq m \leq n^2$ minimal such that the subfamily $(T^k)_{k=0,\ldots,m}$ is linearly dependent.

Let $a_0, \ldots, a_{m-1} \in \mathbb{F}$ such that

$$T^m = a_0 I_V + \ldots + a_{m-1} T^{m-1} \text{ and } f(t) = t^m - \sum_{k=0}^{m-1} a_k t^k$$

It follows that $f(T) = 0$ and, therefore, $f \in \mathfrak{A}_T \setminus \{0\}$. □

Exercise. Prove that $f = m_T$, where f is as in the preceeding proof.

How can we use this fact and matricial representations of T to compute m_T?

4.2 Cyclic subspaces

Definition 4.2.1 (T-cycle). Given $v \in V$, the sequence of vectors

$$v_0 = v, v_1 = T(v), \ldots, v_k = T^k(v), \ldots$$

is called a T-**cycle** generated by v. We denote it by $\mathfrak{C}_T^\infty(v) = (v_k)_{k \geq 0}$ and $\mathfrak{C}_T^m(v) = v_0, v_1, \ldots, v_{m-1}$.

We define the T-**cyclic subspace generated by** v as

$$C_T(v) = [\mathfrak{C}_T^\infty(v)]$$

If $\dim(V)$ is finite, there exists $m \geq 0$ minimal such that $\mathfrak{C}_T^{m+1}(v)$ is linearly dependent. Notice

that $m \geq 1$ if $v \neq 0$. In particular, if $\mathfrak{C}_T(v) := \mathfrak{C}_T^m(v)$ is a basis for $C_T(v)$ and v_m is a linear combination of the linearly independent vectors $v_0, \ldots, v_{m-1}$. And

$$v_m = a_0 v_0 + \ldots + a_{m-1} v_{m-1} = \sum_{k=0}^{m-1} a_k T^k(v)$$

Defining $m_{T,v}(t) = t^m - a_{m-1}t^{m-1} - \ldots - a_1 t - a_0$, called the **minimal polynomial of v with respect to** T,

$$m_{T,v}(T)(v) = v_m - \sum_{k=0}^{m-1} a_k T^k(v) = 0$$

which proves that $v \in V_{m_{T,v}}$. Since $V_{m_{T,v}}$ is T-invariant, it follows that

$$C_T(v) \subseteq V_{m_{T,v}}$$

Cyclic subspaces are useful to find divisors of m_T. Since T is fixed, we'll write $m_v = m_{T,v}$.

The next result shows that it is possible to approximate the minimal polynomial of a linear operator by a minimal polynomial of a vector.

Theorem 4.2.1. For all $v \in V$, $m_v \,|\, m_T$.

Proof. Since V_{m_v} is T-invariant, consider the induced operator

$$S : V_{m_v} \longrightarrow V_{m_v}, \quad v \mapsto T(v)$$

Note that $m_v \in \mathfrak{A}_S$. Thus, $m_S \,|\, m_v$. In fact, $m_v = m_S$, since if $\deg(m_S) = m' < m$, then $v_0, \ldots, v^{m'}$ would be linearly dependent, contradicting the minimality of m. □

We proceed to show that coprimality implies direct sum.

Theorem 4.2.2 (Coprimality implies direct sum)**.** If each pair $p_1, \ldots, p_m \in \mathbb{F}[x]$ is relatively prime, then the sum $V_{p_1}^\infty + \ldots + V_{p_m}^\infty$ is direct.

Proof. Let $v_j \in V_{p_j}^\infty \setminus \{0\}$, for $1 \leq j \leq m$. Our goal is to show that $v_1, \ldots, v_m$ is linearly independent (which is equivalent to show that the sum in the theorem is direct). We proceed by induction on m, which is immediate when $m = 1$.

Suppose that $a_1, \ldots, a_m \in \mathbb{F}$ satisfies $a_1 v_1 + \ldots + a_m v_m = 0$. By the induction hypothesis, $v_1, \ldots, v_{m-1}$ is linearly independent.

For each $1 \leq j \leq m$, choose k_j such that $p_j^{k_j}(T)(v_j) = 0$ and consider

$$p = \prod_{j=1}^{m-1} p_j^{k_j}$$

Notice that

$$0 = p(T)(a_1 v_1 + \ldots + a_m v_m) = a_m p(T)(v_m)$$

By the Lemma 4.1.8, the restriction $p_j^{k_j}(T)$ to $V_{p_m^{k_m}}$ is injective for $j \neq m$, it follows that

$p(T)(v_m) \neq 0$ and, therefore, $a_m = 0$. The induction hypothesis implies that $a_j = 0$ for all j. □

Finally, we can decompose V into direct sums.

Theorem 4.2.3. Let $f_1, \ldots, f_m \in \mathbb{F}[x]$, where each pair is relatively prime and $f = f_1 \ldots f_m$. Then, $V_f = V_{f_1} \oplus \ldots \oplus V_{f_m}$.

Proof. We'll prove for $m = 2$. The general case follows from induction on m.

Then $V_{f_1} + V_{f_2} \subseteq V_f$ and we just proved that this sum is direct.

Let $g_1, g_2 \in \mathbb{F}[x]$ such that $g_1 f_1 + g_2 f_2 = 1$, and define $h_j := g_j f_j$ and $P_j := h_j(S)$, where S is the induced linear operator by T on V_f (i.e. the restriction of T to V_f).

By construction, $f \in \mathfrak{A}_S$ and $P_1 + P_2 = I_{V_f}$. Notice that

$$P_1 P_2 = g_1(S) f_1(S) g_2(S) f_2(S) = g_1(S) g_2(S) f_1(S) f_2(S) = 0 = P_2 P_1$$

In particular, $P_j^2 = P_j - P_i P_j = P_j$ if $\{i, j\} = \{1, 2\}$ and $\mathrm{Im}(P_j) \subseteq V_{f_i}$. It follows that $V_f = \mathrm{Im}(P_2) \oplus \mathrm{Im}(P_1) \subseteq V_{f_1} \oplus V_{f_2} \subseteq V_f$. Hence, $V_{f_1} + V_{f_2} = V_f$. □

Taking $f_j = p_j^{k_j}$, we prove the Primary Decomposition Theorem, since $V = V_{m_T}$.

However, the PDT breaks the vector space in direct sums that are still quite complicated. Our goal now is to break the vector space into the sum of cyclic subspaces.

The next theorem states that we can find a finite collection of vectors such that it is possible to break the space into a direct sum of cyclic subspaces. Therefore, we can find a basis formed by the union of cycles, which are easy to operate.

Our main result here, the Cyclic Decomposition Theorem, is a generalization of this result.

Recall that if v is an eigenvector, then $T(v) = \lambda v$. Hence, a basis formed by eigenvectors is a basis formed by union of cycles, where all cycles have size one.

Lemma 4.2.4. If $p \in \mathbb{F}[x]$ is irreducible and $u, v \in V$ satisfying $m_u = m_v = p$. Then one and only one of the following is true:

1. $C_T(u) = C_T(v)$;
2. $C_T(u) \cap C_T(v) = \{0\}$.

Proof. **Exercise.** □

Lemma 4.2.5. $C_T(v) = \{p(T)(v) : p \in \mathbb{F}[x]\}$.

Proof. **Exercise.** □

A subspace W is said to be T-**cyclic** if $W = C_T(v)$ for some $v \in V$.

Theorem 4.2.6. If $p \in \mathbb{F}[x]$ is irreducible and $\dim(V_p)$ is finite, there exist $l \geq 0$ and $v_1, \ldots, v_l \in V_p$ such that $V_p = C_T(v_1) \oplus \ldots \oplus C_T(v_l)$.

Proof. Suppose that we found vectors $v_1, \ldots, v_k \in V_p$ such that $C_T(v_1) + \ldots + C_T(v_k)$ is a direct

sum. We will show that

$$C_T(v) \cap (C_T(v_1) + \ldots + C_T(v_k)) = \{0\} \tag{4.2}$$

for all $v \in V_P \setminus C_T(v_1) + \ldots + C_T(v_k)$.

If this is true, since $\dim(V_p) < \infty$, it is immediate how to choose a sequence of vectors starting with $v_1 \in V_p$ arbitrary.

To show (4.2), suppose that $w \in C_T(v) \cap (C_T(v_1) + \ldots + C_T(v_k))$ is non-zero. Therefore, there exists $w_j \in C_T(v_j)$ such that $w = w_1 + \ldots + w_k$.

By the lemma 4.2.4, $C_T(w) = C_T(v)$ and, by the lemma 4.2.5, there exists $f \in \mathbb{F}[x]$ such that $v = f(T)(w)$. Hence,

$$v = f(T)(w_1) + \ldots + f(T)(w_k) \in C_T(v_1) + \ldots + C_T(v_k)$$

contradicting our hypothesis about v. □

Theorem 4.2.7. Let $p \in \mathbb{F}[x]$ be irreducible and $\dim(V_p^\infty) < \infty$. Then $\deg(p) \mid \dim(V_p^\infty)$. Moreover,

$$\frac{\dim(V_p^\infty)}{\deg(p)} \geq \min\{k : V_p^\infty = V_{p^k}\}$$

where the equality holds iff. V_p^∞ is T-cyclic.

Proof. Define $m = \min\{k : V_p^\infty = V_{p^k}\}$. Suppose that V_p^∞ is T-cyclic. Let S be the operator T restricted to $V_p^\infty = V_{p^m}$ and notice that $m_S = p^m$.

Moreover, if v satisfies $V_p^\infty = C_T(v)$, then $m_v = m_S = p^m$. Since $\dim(C_T(v)) = \deg(m_v) = m \cdot \deg(p)$, the result follows.

We proceed by induction on $n = \dim(V_p^\infty)$. For $n = 1$, it is immediate. Suppose that $n > 1$. Our induction hypothesis is: if S is an endomorphism on W such that $\dim(W_{p(S)}^\infty) < n$, then

$$\deg(p) \mid \dim(W_{p(S)}^\infty) \quad \text{and} \quad \frac{\dim(W_{p(S)}^\infty)}{\deg(p)} \geq \min\{k : W_{p(S)}^\infty = W_{p^k}\}$$

For $m = 1$, the cyclic case with the previous theorem complete the proof.

If $m > 1$, consider $W = V_{p^{m-1}}$, which is a T-invariant, proper and non-trivial subspace of V_p^∞. Let S be the induced operator by T on W and notice that $W_{p^k(S)} = V_{p^k(T)}$ if $k < m$. Hence, $W = W_{p(S)}^\infty = V_{p^{m-1}}$. By induction hypothesis,

$$\deg(p) \mid \dim(W) \quad \text{and} \quad \frac{\dim(W)}{\deg(p)} \geq m - 1$$

Consider R the operator on V_p^∞ induced by $p^{m-1}(T) = p(T)^{m-1}$. Notice that $W = \ker(R)$. Hence, by the Rank-Nullity Theorem,

$$\dim(V_p^\infty) = \dim(W) + \dim(\mathrm{Im}(R))$$

Thus, what remains to be shown is that $\deg(p) \mid \dim(\mathrm{Im}(\mathrm{R}))$. Note that $\mathrm{Im}(\mathrm{R})$ is T-invariant: if $w = \mathrm{R}(v)$, where $v \in \mathrm{V}_p^\infty$, since $\mathrm{R} \circ \mathrm{T} = \mathrm{T} \circ \mathrm{R}$, it follows that

$$\mathrm{T}(w) = \mathrm{T}(\mathrm{R}(v)) = \mathrm{R}(\mathrm{T}(v)) \in \mathrm{Im}(\mathrm{R})$$

More than that, $\mathrm{Im}(\mathrm{R}) \subseteq \mathrm{V}_p$: if $v \in \mathrm{V}_p^\infty$, $p(\mathrm{T})(\mathrm{R}(v)) = p^m(\mathrm{T})(v) = 0$. By the induction hypothesis on $\mathrm{Im}(\mathrm{R})$ instead of W, the result follows. □

How is all of this related to the idea of characteristic polynomial?

If $\dim(\mathrm{V}) = n < \infty$ and $p_j^{k_j}$, where $1 \leq j \leq m$, are the primary factors of T. It follows that.

$$\deg(m_\mathrm{T}) = \sum_{j=1}^{m} k_j \deg(p_j) \leq \sum_{j=1}^{m} \dim(\mathrm{V}_{p_j{}^{k_j}}) = n$$

since

$$k_j = \min\{k : \mathrm{V}_{p_j{}^k} = \mathrm{V}_{p_j{}^{k+1}}\}$$

Define

$$n_j := \frac{\dim(\mathrm{V}_{p_j{}^{k_j}})}{\deg(p_j)} \geq k_j$$

and

$$c_\mathrm{T} := \prod_{j=1}^{m} p_j^{n_j}$$

This polynomial c_T is the **characteristic polynomial** of T. By definition, $\deg(c_\mathrm{T}) = n$ and $c_\mathrm{T} \in \mathfrak{A}_\mathrm{T}$ (Cayley-Hamilton Theorem).

If T_j is the restriction of T on $\mathrm{V}_{p_j{}^{k_j}}$,

$$c_\mathrm{T} = \prod_{j=1}^{m} c_{\mathrm{T}_j}$$

With these tools at hand, we can define determinants and the trace. Let $c_k \in \mathbb{F}$, $1 \leq k \leq n$, such that

$$c_\mathrm{T}(t) = t^n + \sum_{k=1}^{n} (-1)^k c_k t^{n-k}$$

and define $\det(\mathrm{T}) = c_n$ and $\mathrm{tr}(\mathrm{T}) = c_1$.

We will show that $\det(\mathrm{T}) = \det([\mathrm{T}]_\alpha^\alpha)$ for every basis α of V and

$$c_\mathrm{T}(\lambda) = \det([\mathrm{T}]_\alpha^\alpha - \lambda \mathrm{I})$$

This gives us a new method to obtain a the minimal polynomial m_T.

1. Compute c_T and find its irreducible factors (hint: start by row reducing the matrix), say $p_1, \ldots, p_m$. Let n_j be the multiplicity of p_j in the factorization of c_T;

2. Let $A = [T]_\alpha^\alpha$ and, given $k = (k_1, \ldots, k_m) \in (\mathbb{Z}_{>0})^m$ with $k_j \leq n_j$, let $p_k = \sum_{j=1}^m p_j^{k_j}$. From all polynomials of the form p_k such that $p_k(A) = 0$, the one with the minimal k_j for all j is m_T. To find each minimum k_j, we can compute the sequence of kernels $V_{p_j} \subseteq \ldots \subseteq V_{p_j}^{n_j} = V_{p_j}^{n_j+1}$.

Example 4.2.1. Suppose that $\mathrm{char}(\mathbb{F})$, $V = \mathbb{F}^4$ and $T \in \mathrm{End}(V)$ given by

$$T(x_1, x_2, x_3, x_4) = (3x_3 + x_4, 2x_1 + 2x_2 - 4x_3 + 2x_4, x_3 + x_4, 3x_4 - x_3)$$

If α is the standard basis,

$$[T]_\alpha^\alpha = \begin{bmatrix} 0 & 0 & 3 & 1 \\ 2 & 2 & -4 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & 3 \end{bmatrix} \quad \text{and} \quad c_T(\lambda) = t(t-2)^3$$

Therefore, the options for m_T are $t(t-2)^k$ with $1 \leq k \leq 3$. In fact,

$$k = 4 - \dim(V_{t_2})$$

Since

$$[T]_\alpha^\alpha - 2I = \begin{bmatrix} -2 & 0 & 3 & 1 \\ 2 & 0 & -4 & 2 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

the homogeneous linear system associated with this matrix has solution $\{(0, x, 0, 0) : x \in \mathbb{F}\}$, it follows that $k = 3$ and

$$[e_2] = V_{t-2} \not\subseteq V_{(t-2)^2} \not\subseteq V_{(t-2)^3} = V_{t-2}^\infty$$

Squaring the matrix,

$$\begin{bmatrix} -2 & 0 & 3 & 1 \\ 2 & 0 & -4 & 2 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix}^2 = \begin{bmatrix} 4 & 0 & -10 & 2 \\ -4 & 0 & 8 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

with solution $V_{(t-2)^2} = \{(2x, y, x, x) : x, y \in \mathbb{F}\}$.

Taking the cube,

$$\begin{bmatrix} -2 & 0 & 3 & 1 \\ 2 & 0 & -4 & 2 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix}^3 = \begin{bmatrix} -8 & 0 & 20 & -4 \\ 8 & 0 & -20 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

with solution $V_{(t-2)^3} = \{(z, y, x, 5x - 2z) : x, y, z \in \mathbb{F}\}$.

Chapter 5

Determinants

5.1 Commutative Rings

In this chapter, we define determinants, and we do so for a broader range of matrices, in which its entries are not only from a field but of a more general form, such as polynomials. To do that, we need the following definition.

Definition 5.1.1 (Ring)**.** A **ring** is a set K together with two operations + (addition) and · (multiplication) satisfying:

1. K is a commutative group under addition;
2. Multiplication is associative: $(x \cdot y) \cdot z = x \cdot (y \cdot z)$;
3. The two distributive laws hold: $x \cdot (y+z) = x \cdot y + x \cdot z$ and $(y+z) \cdot x = y \cdot x + z \cdot x$.

If $x \cdot y = y \cdot x$, then the ring K is said to be **commutative**.

If there exists an element $e \in \mathrm{K}$ such that $e \cdot x = c \cdot e = x$ for each $x \in \mathrm{K}$, then e is called the **identity** for K and the ring K is said to be a **ring with identity**.

In this text, we use 'ring' to denote a commutative ring with identity.

5.2 Determinant Functions

Intuitively, the determinant has a geometrical meaning: it is an oriented volume. What do we mean by that? [1]

Consider a matrix A and $\mathrm{T_A}$ is the transformation that A represents. Notice that the standard basis is a unitary cube, which we denote by C. We obtain a parallelepiped by applying the transformation $\mathrm{T_A}$ to the unitary cube C (i.e., the standard basis).

So the determinant is the volume of the parallelepiped obtained by the range of the unitary cube by the transformation represented by the matrix.

Algebraically, our goal is to define a function $\mathbf{M}_n(\mathrm{K}) \longrightarrow \mathrm{K}$ such that it is linear in each row of

[1] For more information on the geometrical interpretation of the determinant, please refer to [Die69].

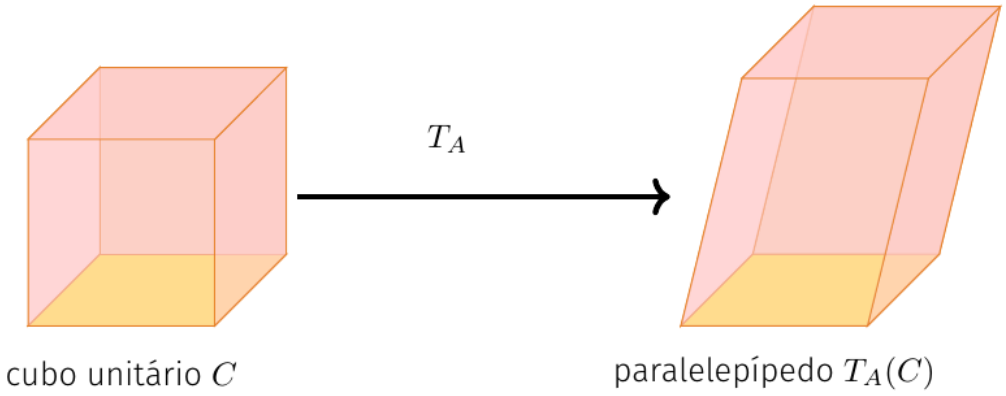


Figure 5.1: The unitary cube C mapped into the parallelepiped $T_A(C)$ [Mir20].

the matrix (hence it is zero if two rows are equal), and the value of this function applied to the identity matrix is one.

Definition 5.2.1 (n-linear). Let K be a ring, n a positive integer, and let D be the following function

$$\begin{aligned} D : \mathbf{M}_n(K) &\longrightarrow K \\ A &\longmapsto D(A) \end{aligned}$$

Then D is called **n-linear** if for each i, $1 \leq i \leq n$, D is a linear function of the ith row when the other $(n-1)$ rows are held fixed.

Notice that if $v_1, \ldots, v_n$ are the rows of A,

$$A = \begin{bmatrix} \text{—} & v_1 & \text{—} \\ \text{—} & v_2 & \text{—} \\ & \vdots & \\ \text{—} & v_n & \text{—} \end{bmatrix}$$

Then D(A) is a function of the rows of A, i.e.

$$D(A) = D(v_1, v_2, \ldots, v_n)$$

Then D is n-linear means that

$$D(v_1, \ldots, cv_i + v_i', \ldots, v_n) = cD(v_1, \ldots, v_i, \ldots, v_n) + D(v_1, \ldots, v_i', \ldots v_n)$$

Example 5.2.1. The product of the diagonal entries of an $n \times n$ matrix is n-linear.

Lemma 5.2.1. A linear combination of n-linear functions is n-linear.

Proof. Let D, E be n-linears. If $a, b \in K$, then

$$(aD + bE)(A) = aD(A) + bE(A)$$

Fixing all rows except the ith row, we can write $D(v_i)$ instead of $D(A)$. Then

$$\begin{aligned}(aD+bE)(cv_i+v_i') &= aD(cv_i+v_i')+bE(cv_i+v_i')\\ &= acD(v_i)+aD(v_i')+bcE(v_i)+bE(v_i')\\ &= c(aD+bE)(v_i)+(aD+bE)(v_i')\end{aligned}$$

□

Example 5.2.2. By the above lemma, the function $D(A) = A_{11}A_{22} - A_{12}A_{21}$ is 2-linear.

Notice that for the D defined in the preceding example, we have

- $D(I) = 1$;
- If two rows are equal, then $D(A) = 0$;
- If A' is obtained from A by interchanging its rows, then $D(A') = -D(A)$, since

$$D(A') = A'_{11}A'_{22} - A'_{12}A'_{21} = A_{21}A_{12} - A_{22}A_{11} = -D(A)$$

This fact (plus the geometrical interpretation of volume) induces the following nomenclature.

Definition 5.2.2 (Alternating)**.** Let D be a n-linear function. We say that D is **alternating** (or **alternated**) if the following two conditions are satisfied:

1. $D(A) = 0$ whenever two rows of A are equal;
2. $D(A') = -D(A)$ if A' is a matrix obtained from A by interchanging two rows of A.

Notice that the second condition fails at fields of characteristic two. But actually, nothing is lost! In fact, the first condition implies the second one. Therefore, we can define alternating functions just using the first condition, which will be equivalent to the previous definition and works at char $\mathbb{F} = 2$.

With these tools at hand, we can finally define the determinant.

Definition 5.2.3 (Determinant)**.** Let K be a ring, and n a positive integer. Suppose that $D : \mathbf{M}_n(K) \longrightarrow K$. Then D is said to be a **determinant function** if D is n-linear, alternating and $D(I) = 1$.

Our task now is to show the existence and uniqueness of the determinant function.

For $n = 1$, it is trivial: $A = [a]$ and $D(A) = a$.

In the case $n = 2$, D(A) is of the form

$$D(A) = D(A_{11}e_1 + A_{12}e_2, A_{21}e_1 + A_{22}e_2)$$

If D is 2-linear, we have that

$$\begin{aligned}D(A) &= A_{11}D(e_1, A_{21}e_1 + A_{22}e_2) + A_{12}D(e_2, A_{21}e_1 + A_{22}e_2)\\ &= A_{11}A_{21}D(e_1,e_1) + A_{11}A_{22}D(e_1,e_2) + A_{12}A_{21}D(e_2,e_1) + A_{12}A_{22}D(e_2,e_2)\end{aligned}$$

Hence, D is completely determined by

$$D(e_1,e_1),\ D(e_1,e_2),\ D(e_2,e_1),\ \text{and } D(e_2,e_2)$$

We've shown that $D(A) = A_{11}A_{22} - A_{12}A_{21}$ is a determinant function. By the preceeding argument, then D is also unique.

Since D is alternating,

$$D(e_1, e_1) = D(e_2, e_2) = 0$$

and

$$D(e_2, e_1) = -D(e_1, e_2) = -D(I)$$

and $D(I) = 1$.

Example 5.2.3. Let D be any alternating 3-linear function on 3×3 matrices over the polynomial ring $\mathbb{F}[x]$. And let

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

Then $D(A) = D(xe_1 - x^2e_3, e_2, e_1 + x^3e_3)$. Since D is linear in each row,

$$\begin{aligned} D(A) &= xD(e_1, e_2, e_1 + x^3e_3) - x^2D(e_3, e_2, e_1 + x^3e_3) \\ &= xD(e_1, e_2, e_1) + x^4D(e_1, e_2, e_3) - x^2D(e_3, e_2, e_1) - x^5D(e_3, e_2, e_3) \end{aligned}$$

By the hypothesis that D is alternating it follows that

$$D(A) = (x^4 + x^2)D(e_1, e_2, e_3)$$

Lemma 5.2.2. Let D be a 2-linear function such that $D(A) = 0$ for all 2×2 matrices A having equal rows. Then D is alternating.

Proof. We show that if $A = (u, v)$, then $D(v, u) = -D(u, v)$. Since D is 2-linear,

$$D(u + v, u + v) = D(u, u) + D(u, v) + D(v, u) + D(v, v) = 0$$

Hence,

$$0 = D(u, v) + D(v, u)$$

□

Lemma 5.2.3. If D is an n-linear function and $D(A) = 0$ whenever two adjacent rows of A are equal, then D is alternating.

Proof. If A' is obtained by interchanging two adjacent rows of A, then by the preceeding lemma, $D(A') = -D(A)$.

We need to show that $D(A) = 0$ when any two rows of A are equal. Let B be the matrix obtained by interchanging rows i and j of A, where $i < j$. We obtain B by a succession of interchanges of pairs of adjacent rows.

First we interchange the rows i and $i + 1$ until the rows are in the order

$$v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_j, v_i, v_{j+1}, \ldots, v_n$$

This process required $k = j - i$ interchanges of adjacent rows. Now, we need to move v_j to

the ith position, requiring more $j-1-i=k-1$ interchanges of adjacent rows.

Since we obtained B from A by $k+(k-1)=2k-1$ interchanges,

$$D(B)=(-1)^{2k-1}D(A)=-D(A)$$

Now let A be a matrix with two equal rows, say $v_i=v_j$, where $i<j$.

If $j=i+1$, then $D(A)=0$, since A has two equal and adjacent rows.

If $j>i+1$, then the batrix B, obtained by exchanging the rows j and $i+1$ of the matrix A has two equal and adjacent rows. I.e., $D(B)=0$.

Since $D(A)=-D(B)$, then it follows that $D(A)=0$. □

The next definition and theorem will be applied when we study the cofactor method.

Definition 5.2.4. If $n>1$ and A is an $n\times n$ matrix, let $A(i\,|\,j)$ denote the $(n-1)\times(n-1)$ matrix obtained by deleting the ith row and the jth row column of A.

If D is an $(n-1)$-linear function, then we define $D_{ij}(A)=D[A(i\,|\,j)]$.

Theorem 5.2.4. Let $n>1$ and D an alternating $(n-1)$-linear function on $(n-1)\times(n-1)$ matrices. For each j, where $1\le j\le n$, the function E_j defined by

$$E_j(A)=\sum_{i=1}^{n}(-1)^{i+j}A_{ij}D_{ij}(A)$$

is an alternating n-linear function on $n\times n$ matrices A. If D is a determinant function, so is each E_j.

Proof. **First step:** Show that E_j is n-linear.

Notice that $D_{ij}(A)$ is independent of the the *ith* row of A. Since D is $(n-1)$-linear, it is linear as a function of any row except row i. Therefore $A_{ij}D_{ij}(A)$ is an n-linear function of A. Hence, E_j is n-linear.

Second step: Show that E_j is alternating.

Suppose that two adjacent rows a_k and a_{k+1} are equal. If $i\neq k$ and $i\neq k+1$, then $A(i\,|\,j)$ has two equal rows, and $D_{ij}(A)=0$. Therefore,

$$E_j(A)=(-1)^{k+j}A_{kj}D_{kj}(A)+(-1)^{k+1+j}A_{(k+1)j}D_{(k+1)j}(A)$$

Using that $a_k=a_{k+1}$,

$$A_{kj}=A_{(k+1)j}\text{ and }A(k\,|\,j)=A(k+1\,|\,j)$$

Therefore,

$$D_{kj}(A)=D[A(k\,|\,j)]=D[A(k+1\,|\,j)]=D_{(k+1)j}(A)$$

which implies that $E_j(A)=0$. Hence, $E_j(A)=0$ whenever A has two equal and adjacent rows. I.e., E_j is alternating.

Third step: Show that $E_j(I)=1$.

Suppose that D is a determinant function. Then

$$E_j(I_n) = D(I_{n-1}) = 1 \implies E_j(I_n) = 1$$

Hence, E_j is a determinant function. □

Corollary 5.2.5. If K is a ring and n is a positive integer, then there exists at least one determinant function on $\mathbf{M}_n(K)$.

5.3 Permutations and the Uniqueness of Determinants

To prove the uniqueness of the determinant function, we proceed in steps. Here, suppose that A is an $n \times n$ matrix over K and D is an alternanting n-linear function on these matrices.

First step. Express every row $a_1, a_2, \ldots, a_n$ of the matrix A in terms of the standard basis $e_1, e_2, \ldots, e_n$.

Second step. Using multilinearity, express D(A) as a linear combination of the matrix entries and the determinant of the standard vectors.

To do that, notice that, for $1 \leq i \leq n$, each row can be expressed as

$$a_i = \sum_{j=1}^{n} A(i,j)e_j$$

Hence,

$$\begin{aligned} D(A) &= D\left(\sum_{j=1}^{n} A(i,j)e_j, a_2, \ldots, a_n\right) \\ &= \sum_{j=1}^{n} A(1,j)D(e_j, a_2, \ldots, a_n) \end{aligned}$$

Replacing a_2 by $\sum_{k=1}^{n} A(2,k)e_k$,

$$D(e_j, a_2, \ldots, a_n) = \sum_{k=1}^{n} A(2,k)D(e_j, e_k, \ldots, a_n)$$

Thus,

$$D(A) = \sum_{j,k} A(1,j)A(2,k)D(e_j, e_k, \ldots, a_n)$$

Repeating this process, we obtain an important expression for D(A),

$$D(A) = \sum_{k_1,k_2,\ldots,k_n} A(1,k_1)A(2,k_2)\ldots A(n,k_n)D(e_{k_1}, e_{k_2}, \ldots, e_{k_n})$$

where $1 \leq k_i \leq n$, $i = 1, 2, \ldots, n$.

Third step. The repeated terms will vanish, and only the permutations remain.

Since D is alternating, $D(e_{k_1}, e_{k_2}, \ldots, e_{k_n}) = 0$ whenever two of the indices k_i are equal. If the sequence $(k_1, k_2, \ldots, k_n)$ of positive integers not exceeding n, with the property that no two k_i are equal, is called a **permutation of degree** n.

With this remark, we can simplify the sum above by only considering those sequences which are permutations of degree n.

In order to do that, notice that a permutation of degree n may be defined as an injection σ from $\{1, 2, \ldots, n\}$ onto itself (hence, a bijection). Put another way, this function corresponds to the n-tuple $(\sigma_1, \sigma_2, \ldots, \sigma_n)$, which is a reordering of $1, 2, \ldots, n$.

Then, we have

$$D(A) = \sum_{\sigma} A(1, \sigma_1) \ldots A(n, \sigma_n) D(e_{\sigma_1}, \ldots, e_{\sigma_n})$$

where the sum extends over the distinct permutations σ of degree n.

From the fact that D is an alternating function, we know that

$$D(e_{\sigma_1}, \ldots, e_{\sigma_n}) = \pm D(e_1, \ldots, e_n)$$

where the sign depends only on the permutation σ.

For example, if we pass from $(1, 2, \ldots, n)$ to $(\sigma_1, \sigma_2, \ldots, \sigma_n)$ by m interchanges, we have that

$$D(e_{\sigma_1}, \ldots, e_{\sigma_n}) = (-1)^m D(e_1, \ldots, e_n)$$

If D is a determinant function,

$$D(e_{\sigma_1}, \ldots, e_{\sigma_n}) = (-1)^m$$

A permutation is called **even** if the number of interchanges used is even, and **odd** otherwise. We define the **sign** of a permutation by

$$\text{sign } \sigma = \begin{cases} 1, & \text{if } \sigma \text{ is even} \\ -1, & \text{if } \sigma \text{ is odd} \end{cases}$$

With this defintion, we may write

$$D(A) = \left[\sum_{\sigma} (\text{sign } \sigma) A(1, \sigma_1) \ldots A(n, \sigma_n) \right] D(I)$$

By this formula, we know that the determinant exists and that there is exactly one determinant function. If we denote this function by det, we can summarize our results in the following.

Theorem 5.3.1. Let K be a ring and n a positive integer. There is precisely one determinant function on the set of $n \times n$ matrices over K, and it is the function det defined by

$$\det(A) = \sum_{\sigma} (\text{sign } \sigma) A(1, \sigma_1) \ldots A(n, \sigma_n)$$

If D is any alternaning n-linear function on $\mathbf{M}_n(K)$, then for each $n \times n$ matrix A we have that

$$D(A) = \det(A) D(I)$$

Remark. We can define a product of permutations σ and τ as the composed function $\sigma \circ \tau$, defined as

$$(\sigma\tau)(i) = \sigma(\tau(i))$$

If e denotes the identity permutation, $e(i) = i$, then each σ has an inverse σ^{-1} such that

$$\sigma\sigma^{-1} = \sigma^{-1}\sigma = e$$

Summarizing, under the operation of composition, the set of permutations of degree n is a group called the **symmetric group of degree** n, denoted by S_n.

A simple property of permutations is that

$$\text{sign } (\sigma\tau) = (\text{sign } \sigma)(\text{sign } \tau)$$

Theorem 5.3.2. Let K be a ring, and let A and B be $n \times n$ matrices over K. Then

$$\det(AB) = (\det A)(\det B)$$

Proof. Let B be a fixed $n \times n$ matrix, and define $D(A) = \det(AB)$.

Since det is n-linear and alternating, D is also n-linear and alternating. Then, by Theorem 5.3.1,

$$D(A) = \det(A)D(I)$$

Since $D(I) = D(IB) = \det B$, we have

$$\det(AB) = D(A) = (\det A)(\det B)$$

□

5.4 Properties of Determinants

Since there is no fundamental difference between rows and columns, the following result is expected.

Theorem 5.4.1. Let A be an $n \times n$ matrix. Then the determinant of the transpose of A equals the determinant of A, i.e.,

$$\det(A^t) = \det(A)$$

Proof. If σ is a permutation of degree n,

$$A^t(i, \sigma_i) = A(\sigma_i, i)$$

And the determinant is given by

$$\det(A^t) = \sum_{\sigma} (\text{sign } \sigma) A(\sigma_1, 1) \dots A(\sigma_n, n)$$

When $i = \sigma_j^{-1}$, we have that $A(\sigma_i, i) = A(j, \sigma_j^{-1})$. Thus

$$A(\sigma_1, 1) \dots A(\sigma_n, n) = A(1, \sigma_1^{-1}) \dots A(n, \sigma_n^{-1})$$

Since $\sigma\sigma^{-1}$ is the identity permutation, $(\text{sign } \sigma)(\text{sign } \sigma^{-1}) = 1$, i.e.,

$$\text{sign } (\sigma^{-1}) = \text{sign } (\sigma)$$

And as σ varies over all permutations of degree n, so does σ^{-1}. Therefore,

$$\det(A^t) = \sum_{\sigma} (\text{sign } \sigma^{-1}) A(1, \sigma_1^{-1}) \dots A(n, \sigma_n^{-1}) = \det(A)$$

□

Theorem 5.4.2. If B is obtained from A by adding a multiple of one row of A to another, then

$$\det B = \det A$$

Proof. If B is obtained from A by adding ca_j to the row a_i, where $i < j$, by the fact that det is n-linear and is alternating,

$$\det B = \det A + c \det(a_1, \dots, a_j, \dots, a_j, \dots, a_n) = \det A$$

□

Theorem 5.4.3. If we have an $n \times n$ matrix of the block form

$$\begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$$

where A is $r\times$, C is $s \times s$, B is $r \times s$ and 0 is the $s \times r$ zero matrix. Then

$$\det \begin{bmatrix} A & B \\ 0 & C \end{bmatrix} = (\det A)(\det C)$$

Proof. Define

$$D(A, B, C) = \det \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$$

Fixing A and B, D is alternating and s-linear as a function of the rows of C. Then, by Theorem 5.3.1

$$D(A, B, C) = (\det C) D(A, B, I)$$

By subtracting multiples of the rows of I from the rows of B, we have

$$D(A, B, I) = D(A, 0, I)$$

Note that $D(A, 0, I)$ is alternating and r-linear as a function of the rows of A. Thus

$$D(A, 0, I) = (\det A) D(I, 0, I)$$

Since $D(I, 0, I) = 1$,

$$D(A, B, C) = (\det C)D(A, B, I) = (\det C)D(A, 0, I) = (\det C)(\det A)$$

□

Remark that the same argument works by taking transposes, i.e.,

$$\det \begin{bmatrix} A & 0 \\ B & C \end{bmatrix} = (\det A)(\det C)$$

By the Theorem 5.2.4, if we fix any column j,

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} A_{ij} \det A(i|j)$$

Definition 5.4.1 (Cofactor)**.** The scalar $C_{ij} = (-1)^{i+j} \det A(i|j)$ is called the i,j **cofactor** of A.

Definition 5.4.2 (Classical adjoint)**.** The transpose of the matrix of cofactors of A is called the **classical adjoint** of A.

$$(\text{adj } A)_{ij} = C_{ji} = (-1)^{i+j} \det A(j|i)$$

Lemma 5.4.4. Let A be an $n \times n$ matrix. Then

$$(\text{adj } A)A = A(\text{adj } A) = (\det A)I$$

Proof. **Step 1.** Show that $(\text{adj } A)A = (\det A)I$.

Replacing the jth column of A by its kth column and calling the resulting matrix B, we have that B has two equal columns and so $\det B = 0$. Given that $B(i|j) = A(i|j)$, we have

$$0 = \det B = \sum_{i=1}^{n} (-1)^{i+j} B_{ij} \det B(j|i) = \sum_{i=1}^{n} (-1)^{i+j} A_{ik} \det B(j|i) = \sum_{i=1}^{n} A_{ik} C_{ij}$$

Thus,

$$A_{ik} C_{ij} = \delta_{jk} \det A$$

By the definition of adjoint, the previous equations gives $(\text{adj } A)A = (\det A)I$.

Step 2. Show that $(\text{adj } A) = (\det A)I$.

Since $A^t(i|j) = A(j|i)^t$, we have

$$(-1)^{i+j} \det A^t(i|j) = (-1)^{i+j} \det A(j|i)$$

i.e., the i,j cofactor of A^t is the j,i cofactor of A. Thus

$$\text{adj}(A^t) = (\text{adj } A)^t$$

Hence,

$$(\text{adj } A^t)A^t = (\det A^t)I = (\det A)I$$

Transposing,

$$A(\text{adj } A^t)^t = (\det A)I$$

And finally,

$$A(\text{adj } A) = (\det A)I$$

□

Theorem 5.4.5. Let A be an $n \times n$ matrix over K. Then A is invertible over K iff. $\det A$ is invertible in K. When A is invertible, the unique inverse for A is

$$A^{-1} = (\det A)^{-1}\text{adj } A$$

In particular, an $n \times n$ matrix over a field is invertible iff. its determinant is different from zero.

Proof. Suppose that $\det A$ is invertible in K. Then, A is invertible and

$$A^{-1} = (\det A)^{-1}\text{adj } A$$

is the unique inverse of A.

Conversely, if A is invertible over K, then there exists a matrix B such that $BA = I$. Then

$$1 = \det I = \det(AB) = (\det A)(\det B)$$

Hence, $\det A$ is invertible in K. □

Remark. For matrices with polynomial entries, the matrix is invertible over $\mathbb{F}[x]$ iff. its determinant is a non-zero scalar polynomial.

Example 5.4.1 (Cramer's Rule). Let $A \in \mathbf{M}_n(\mathbb{F})$ and suppose we wish to solve the system $AX = Y$. Then,

$$(\text{adj } A)AX = (\text{adj } A)Y \iff (\det A)X = (\text{adj } A)Y$$

Thus

$$(\det A)x_j = \sum_{i=1}^{n}(\text{adj } A)_{ji}y_i = \sum_{i=1}^{n}(-1)^{i_j}y_i \det A(i\,|\,j)$$

If $\det A \neq 0$, we obtain **Cramer's Rule** and the unique solution $X = A^{-1}Y$ is given by

$$x_j = \frac{\det B_j}{\det A}$$

where B_j is obtained from A by replacing the jth column of A by Y.

5.5 Modules

A module over a ring K behaves like a vector space with K being used as scalars. More precisely,

Definition 5.5.1 (Module). Let K be a commutative ring with identity. A K-**module** is a nonempty set V with two operations.

The first operation, called **addition** and denoted by '+', assigns each pair $(u,v) \in V \times V$ an element $u+v \in V$. And under this operation, V is an abelian group.

The second operation, called **action** or **multiplication** and denoted by juxtaposition, assigns to each pair $(k,v) \in K \times V$ an element $kv \in V$. This operation must satisfy

$$\begin{aligned} r(u+v) &= ru+rv \\ (r+s)u &= ru+su \\ (rs)u &= r(su) \\ 1u &= u \end{aligned}$$

Example 5.5.1. The following are examples of modules.

1. A vector space is a module over a field.
2. The n-tuple modules of K^n.
3. The matrix modules $K^{m\times n}$.

Note that if $v_1,\dots,v_k$ are linearly dependent, it is not always the case that some v_i is a linear combination of the others. The reason for this is the absence of multiplicative inverse in a ring.

The definition of a basis of a module is the same given for vector spaces.

Definition 5.5.2 (Basis). A **basis** for the module V is a linearly independent subset which spans (or generates) the module.

However, it is not always the case that a basis always exists in any module which is spanned by a finite number of elements.

Definition 5.5.3 (Free Module). The K-module V is called a **free module** if it has a basis. If V has a finite basis containing n elements, then V is called a free module with n generators.

Definition 5.5.4 (Finitely Generated and Rank). A K-module V is said to be **finitely generated** if it contains a finite subset which spans V.

The **rank** of a finitely generated module is the smallest integer k such that some k elements span V.

Remark. If V is a free K-module with n generators, then V is isomorphic to K^n.

Theorem 5.5.1. Let K be a ring. If V is a free K-module with n generators, then the rank of V is n.

Proof. We want to prove that V cannot be spanned by less than n of its elements. Since $V \simeq K^n$, we show that, if $m < n$, then the module K^n is not spanned by n-tuples $v_1,\dots,v_m$. □

Definition 5.5.5 (Dual Module). If V is a module over K, the **dual module** V^* consists of all linear functions f from V into K.

If V is a free module of rank n, then V^* is also a free module of rank n.

And if $\{\beta_1, \ldots, \beta_n\}$ is an ordered basis for V, there exists an associated **dual basis** $\{f_1, \ldots, f_n\}$ for the module V^*, where each f_i assigns to each $v \in V$ its ith coordinate relative to $\{\beta_1, \ldots, \beta_n\}$:

$$v = f_1(v)\beta_1 + \ldots + f_n(v)\beta_n$$

If f is a linear function on V, then

$$f = f(\beta_1)f_1 + \ldots + f(\beta_n)f_n$$

5.6 Multilinear Functions

In this section, we treat alternating multilinear forms on modules. These are the natural generalization of determinants as we presented them.

Definition 5.6.1 (Multilinear Functions). Let K be a commutative ring with identity and let V be a module over K. If r is a positive integer, a function L from V^r into K is called **multilinear** if $L(v_1, \ldots, v_r)$ is linear as a function of each v_i when all other v_j's are held fixed.

A multilinear function on V^r is also called an **r-linear form** on V or a **multilinear form of degree r** on V. Such functions are sometimes called **r-tensors** on V.

The collection of all multilinear functions on V^r is denoted by $M^r(V)$. And a 2-linear form on V is usually called a **bilinear form** on V.

Notice that by defining addition and scalar multiplication as usual, $M^r(V)$ is a submodule of all functions from V^r into K.

If $r = 1$, then $M^1(V) = V^*$, the dual module. Linear functions can be used to construct mutilinear forms of higher order. If $f_1, \ldots, f_r$ are linear functions on V, define

$$L(v_1, \ldots, v_r) = f_1(v_1)f_2(v_2)\ldots f_r(v_r)$$

Example 5.6.1. The determinant function is an n-linear form on K^n.

The next definition provides a canonical mapping of the spaces $V_1 \times \ldots \times V_r$.

Definition 5.6.2 (Tensor Product). Let L be a multilinear function on V^r and M a multilinear function on V^s. We define a function $L \otimes M$ on V^{r+s} by

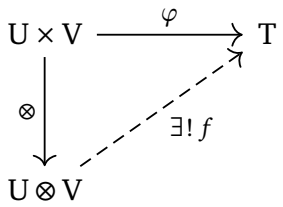
$$(L \otimes M)(v_1, \ldots, v_{r+s}) = L(v_1, \ldots, v_r)M(v_{r+1}, \ldots, v_{r+s})$$

If we think of V^{r+s} as $V^r \times V^s$, then for $v \in V^r$ and w in V^s

$$(L \otimes M)(v, w) = L(v)M(w)$$

Clearly, $L \otimes M$ is multilinear on V^{r+s}. And the function $L \otimes M$ is called the **tensor product** of L and M.

The idea here is to 'linearize' the tensor product. Given any bilinear map $\varphi \in \hom(U \times V, T)$, there exists a unique linear mapping $f : U \otimes V \longrightarrow T$ such that the following diagram commutes.



Put another way, we have the isomorphism

$$\begin{aligned} \hom(U \otimes V, T) &\longrightarrow \hom(U \times V, T) \\ f &\longmapsto f \circ \otimes \end{aligned}$$

In the language of categories, this isomorphism is a universal property.

Lemma 5.6.1 (Properties of Tensoring)**.** Let L, L_1 be r-linear forms on V, M, M_1 be s-linear forms on V, N a t-linear form on V, and let $c \in K$.

1. The tensor product is not commutative. $M \otimes L \neq L \otimes M$ unless $L = 0$ or $M = 0$;
2. $(cL + L_1) \otimes M = c(L \otimes M) + L_1 \otimes M$;
3. $L \otimes (cM + M_1) = c(L \otimes M) + L \otimes M_1$;
4. The tensor product is associative, i.e., $(L \otimes M) \otimes N = L \otimes (M \otimes N)$.

Note that the previous definition can be naturally extended. If $L_1, \ldots, L_k$ are multilinear functions on $V^{r_1}, \ldots, V^{r_k}$, then the tensor product

$$L = L_1 \otimes \ldots \otimes L_k$$

is defined as a multilinear function on V^r, where $r = r_1 + \ldots + r_k$.

Theorem 5.6.2. If V is a free K-module of rank n, then $M^r(V)$ is a free K-module of rank n^r.

In fact, if $\{f_1, \ldots, f_n\}$ is a basis for the dual module V^*, then the n^r tensor products

$$f_{j_1} \otimes \ldots \otimes f_{j_r},\ 1 \leq j_1 \leq n, \ldots, 1 \leq j_r \leq n$$

form a basis for $M^r(V)$.

Definition 5.6.3 (Alternating Linear Form)**.** Let L be an r-linear form on a K-module V. Then L is said to be **alternating** if $L(v_1, \ldots, v_r) = 0$ whwnever $v_i = v_j$ with $i \neq j$.

We denote by $\Lambda^r(V)$ the collection of all alternating r-linear forms on V.

Remark that every permutation σ is a product of transpositions, so

$$L(v_{\sigma_1}, \ldots, v_{\sigma_r}) = (\operatorname{sgn} \sigma) L(v_1, \ldots, v_r)$$

Also notice that $\Lambda^r(V)$ is a submodule of $M^r(V)$.

Remark. We already showed that there is precisely one alternating n-linear form D on the module K^n with the property that $D(e_1,\dots,e_n)=1$. We also showed (Theorem 5.3.1) that if L is any form in $\Lambda^n(K^n)$ then

$$L = L(e_1,\dots,e_n)D$$

Therefore, $\Lambda^n(K^n)$ is a free K-module of rank one. Using the formula for D of the previously cited theorem, we can now write

$$D = \sum_\sigma (\operatorname{sgn}\sigma) f_{\sigma_1}\otimes\dots\otimes f_{\sigma_n}$$

where $f_1,\dots,f_n$ are the coordinate functions on K^n and the sum is extended over the $n!$ different permutations σ of the set $\{1,\dots,n\}$.

If we write the determinant of a matrix A as

$$\det A = \sum_\sigma (\operatorname{sgn}\sigma) A(\sigma_1,1)\dots A(\sigma_n,n)$$

then we obtain the following expression for D:

$$\begin{aligned} D(v_1,\dots,v_n) &= \sum_\sigma (\operatorname{sgn}\sigma) f_1(v_{\sigma_1})\dots f_n(v_{\sigma_n}) \\ &= \sum_\sigma (\operatorname{sgn}\sigma) L(v_{\sigma_1},\dots,v_{\sigma_n}) \end{aligned}$$

where $L = f_1\otimes\dots\otimes f_n$.

A more general method for associating an alternating form with a multilinear form is the following.

Remark. If L is an r-linear form and σ is a permutation of $\{1,\dots,r\}$, we obtain another r-linear function L_σ by defining

$$L_\sigma(v_1,\dots,v_r) = L(v_{\sigma_1},\dots,v_{\sigma_r})$$

If L is alternating, then $L_\sigma = (\operatorname{sgn}\sigma)L$.

For each $L\in M^r(V)$, define a function $\pi_r L\in M^r(V)$ by

$$\pi_r L = \sum_\sigma (\operatorname{sgn}\sigma) L_\sigma$$

Lemma 5.6.3. The function π_r is a linear mapping from $M^r(V)$ into $\Lambda^r(V)$. If $L\in\Lambda^r(V)$, then $\pi_r L = r!L$.

Applying these results to our previous expression for the determinant function $D\in\Lambda^n(K^n)$, we can now write

$$D = \pi_n(f_1\otimes\dots\otimes f_n)$$

Theorem 5.6.4. Let V be a free K-module of rank n. If $r>n$, then $\Lambda^r(V)=\{0\}$. If $1\le r\le n$, then $\Lambda^r(V)$ is a free K-module of rank $\binom{n}{r}$.

Corollary 5.6.5. If V is a free K-module of rank n, then $\Lambda^n(V)$ is a free K-module of rank one. If $T \in \mathrm{End}(V)$, there is a unique element $c \in K$ such that

$$L(T(v_1), \ldots, T(v_n)) = cL(v_1, \ldots, v_n)$$

for every alternating n-linear form L on V. The element c is called the **determinant of** T.

5.7 The Grassman Ring

How can we define a 'natural' multiplication of alternating forms? In order to obtain an associative multiplication, we define a new product.

Definition 5.7.1 (Exterior Product)**.** Let L be an r-liner form and M an s-linear form. We define the **exterior product** (or **wedge product**) by

$$L \wedge M = \frac{1}{r!s!}\pi_{r+s}(L \otimes M)$$

However, a few observations will lead us to a better definition.

Definition 5.7.2 (Exterior Product (Again))**.** Let L be an r-liner form and M an s-linear form. We define the **exterior product** (or **wedge prdocut**) by

$$L \wedge M = \sum_{\sigma} (\mathrm{sign}\ \sigma)(L \otimes M)_{\sigma}$$

Theorem 5.7.1. The exterior product is associative.

Definition 5.7.3 (Grassman Ring)**.** The set of alternating forms $\Lambda(V)$ with the exterior product as multiplication and the addition of the module $\Lambda(V)$ is called the **Grassman Ring**.

Remark that the Grassmann Ring can be seen as a subset of the projective space of the alternating forms.

Bibliography

[Die69] Jean Alexandre Dieudonné. *Linear Algebra and Geometry*. Hermann, 1969. 48

[HK71] Kenneth Hoffman and Ray Kunze. *Linear Algebra*. 1971.

[Mir20] Daniel Miranda. *Álgebra Linear Avançada*. Notas de Aula, 2020. 49

[Mou22] Adriano Adrega de Moura. *Álgebra Linear com Geometria Analítica*. Notas de Aula, 2022. 33

[Rom05] Steven Roman. *Advanced Linear Algebra*. Springer, 2005.

[YIK89] Manin Yu I and Alexei I Kostrikin. *Linear Algebra and Geometry*. CRC Press, 1989.