ADVANCED LINEAR ALGEBRA

Adair Antonio da Silva Neto

January 16, 2023

Contents

1 Canonical Forms **2**

1.1 Annihilating Polynomials . . . 2
1.2 Invariant Subspaces . . . 4
1.3 Simultaneous Triangulation / Diagonalization . . . 5
1.4 Direct-Sum Decompositions . . . 6
1.5 Invariant Direct Sums . . . 7
1.6 Primary Decomposition Theorem . . . 9
1.7 Alternative proofs for minimal polynomials . . . 10
1.8 Cyclic subspaces . . . 12
1.9 Jordan and Frobenius Bases . . . 17

2 Bilinear Forms and Geometry **27**

2.1 Multilinear Functions . . . 27
2.2 Duality . . . 29
2.3 Bilinear Pairings and Orthogonality . . . 38
2.4 Hyperbolic and Orthogonal Bases . . . 41
2.5 Orthogonal and Sympletic Transformations . . . 49
2.6 Self-Adjoint Operators . . . 54

3 Tensor Algebra **60**

3.1 Category Theory . . . 60
3.2 Quotient Spaces and Universal Properties . . . 62
3.3 Tensor Product of Vector Spaces . . . 66
3.4 Tensor Product of Linear Mappings . . . 75
3.5 Exterior and Symmetric Powers . . . 78
3.6 Associative Algebras . . . 83

A Review **86**

A.1 Vector Spaces . . . 86

A.2 Linear Transformations 93
A.3 Diagonalization 99

B Polynomials **105**
B.1 Algebras 105
B.2 The Algebra of Polynomials 105
B.3 Polynomial Ideals 106
B.4 Prime Factorization 107

C Determinants **109**
C.1 Commutative Rings 109
C.2 Determinant Functions 109
C.3 Permutations and the Uniqueness of Determinants 114
C.4 Properties of Determinants 116
C.5 Modules 119
C.6 Multilinear Functions 121
C.7 The Grassman Ring 123

References **125**

Chapter 1

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 $\mathbb{F}$, $\mathrm{T} \in \mathrm{End}(\mathrm{V})$, and $\mathrm{V}_p := \{v \in \mathrm{V} : p(\mathrm{T})(v) = 0\}$, where $p \in \mathbb{F}[x]$.

1.1 Annihilating Polynomials

Definition 1.1.1 (Annihilator Set)**.** We define $\mathfrak{A}_\mathrm{T} := \{p \in \mathbb{F}[x] : p(\mathrm{T}) = 0\}$ 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 1.1.1 (Existence of the Minimal Polynomial)**.**

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

Definition 1.1.2 (Minimal Polynomial)**.** The polynomial m_T in the previous theorem is called **minimal polynomial** of T.

If $\mathfrak{A}_\mathrm{T} = \{0\}$, we define $m_\mathrm{T} = 0$.

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.

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

Example 1.1.1. If $\mathrm{V} = \mathbb{F}[x]$ and T is the operator $f(t) \mapsto tf(t)$, then $\mathfrak{A}_\mathrm{T} = \{0\}$.

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 1.1.2. 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.

Theorem 1.1.3 (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 $b_{ij} = b_{ij}^{(0)} + b_{ij}^{(1)}x + \cdots + b_{ij}^{(n-1)}x^{n-1}$.
5. Let $\mathrm{B}^{(k)}$ be the matrix whose entries are given by $b_{ij}^{(k)}$.
6. Write $c_\mathrm{T}(x) = a_0 + a_1x + \cdots + x^n$ and use that
$$\mathrm{B}\mathrm{A}' = \mathrm{adj}(\mathrm{A}')\mathrm{A}' = (\det \mathrm{A}')\mathrm{I} = c_\mathrm{T}(x)\mathrm{I}$$
7. It follows that $(\mathrm{B}^{(0)} + \mathrm{B}^{(1)}x + \cdots + \mathrm{B}^{(n-1)}x^{n-1})(x\mathrm{I} - \mathrm{A}) = (a_0 + a_1x + \cdots + x^n)\mathrm{I}$.

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 $I, A, A^2, \dots, A^n$ respectively, and adding all of them, we have

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

□

1.2 Invariant Subspaces

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

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

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 1.2.1 (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.

Example 1.2.1. Each of the following subspaces is invariant under T:

1. $\{0\}$;
2. V;
3. $\ker T$;
4. range T.

With the idea of invariant subspaces, the minimal polynomial gives us characterizations of diagonalizable and triangulable operators.

Lemma 1.2.1. Let W be an invariant subspace of T, c_T denote the characteristic polynomial for T and m_T the minimal polynomial for T. Then

$$c_{T|_W} | c_T \quad \text{and} \quad m_{T|_W} | m_T$$

Lemma 1.2.2. Let W be an invariant subspace of T. Then, W is invariant under every polynomial in T. And, for every $v \in \mathrm{V}$, $\mathrm{S}(v, \mathrm{W})$ is an ideal in $\mathbb{F}[x]$.

Definition 1.2.2 (Triangulable)**.** A linear operator T is **triangulable** if there exists a basis of V under which T is represented as a triangular matrix.

Lemma 1.2.3. Let V be finite-dimensional and that the minimal polynomial m_{T} for $\mathrm{T} \in \mathrm{End}(\mathrm{V})$ is a product of linear factors

$$m_{\mathrm{T}} = (\lambda - \lambda_1)^{r_1} \cdots (\lambda - \lambda_k)^{r_k}$$

And let W be a proper subspace of V invariant under T. Then there exists a vector $v \in \mathrm{V}$ such that

1. $v \notin \mathrm{W}$;
2. $(\mathrm{T} - \lambda \mathrm{I})v$ is in W, for some eigenvalue λ of T.

Theorem 1.2.4. Let V be finite-dimensional. Then $\mathrm{T} \in \mathrm{End}(\mathrm{V})$ is triangulable iff. the minimal polynomial m_{T} is a product of linear polynomials.

Proof. ($\Rightarrow$) The idea here is to apply the lemma repeatedly.

For $\mathrm{W}_0 = \{0\}$, there exists $v_0 \neq 0$ such that $(\mathrm{T} - \lambda_0 \mathrm{I})v_0 \in \mathrm{W}_0$.

For $\mathrm{W}_1 = \mathrm{span}\{v_0\}$, there exists $v_1 \notin \mathrm{W}_1$ such that $(\mathrm{T} - \lambda_1 \mathrm{I})v_1 \in \mathrm{W}_1$.

Repeating this process, we obtain that $\mathrm{T}(v_i) \in \mathrm{W}_i$, for all i, and hence we constructed the desired basis.

($\Leftarrow$) If T is triangulable, then c_{T} can be written as

$$c_{\mathrm{T}} = (\lambda - \lambda_1)^{d_1} \cdots (\lambda - \lambda_k)^{d_k}$$

Hence, the elements in the diagonal $a_{11}, \ldots, a_{nn}$, are the eigenvalues. Since $m_{\mathrm{T}} | c_{\mathrm{T}}$, the minimal polynomial can be factored by the eigenvalues λ_i. □

Corollary 1.2.5. If $\mathbb{F}$ is algebraically closed, then every square matrix is similar to a triangular matrix.

Theorem 1.2.6. Let V be finite-dimensional. Then T is diagonalizable iff. the minimal polynomial for T is a product of distinct linear factors.

$$m_{\mathrm{T}} = (\lambda - \lambda_1) \cdots (\lambda - \lambda_k) \qquad \lambda_i \neq \lambda_j,\ \forall\ i \neq j$$

1.3 Simultaneous Triangulation / Diagonalization

In this section, suppose that V is finite-dimensional and let $\mathfrak{F}$ denote a family of linear operators on V. The main question here is how to diagonalize or triangulize every $\mathrm{T} \in \mathfrak{F}$ simultaneously?

Notice that for diagonalization, it is sufficient and necessary that $\mathfrak{F}$ is commuting, since diagonal matrices commute.

Lemma 1.3.1. Let $\mathfrak{F}$ be a family of commuting and triangulable operators, and $W \subsetneq V$ invariant under $\mathfrak{F}$ (i.e. invariant under each operator in $\mathfrak{F}$). There exists $v \in V$ such that

1. $v \notin W$;
2. For all $T \in \mathfrak{F}$, $T(v) \in \operatorname{span}\{v, W\}$.

Theorem 1.3.2. Let $\mathfrak{F}$ be a family of commuting and triangulable operators. There exists an ordered basis for V such that every operator in $\mathfrak{F}$ is represented by a triangular matrix in that basis.

Corollary 1.3.3. Let $\mathfrak{F}$ be a family of commuting and triangulable operators. If $\mathbb{F}$ is algebraically closed, there exists a matrix P non-singular such that $P^{-1}AP$ is upper triangular, for all $A \in \mathfrak{F}$.

Theorem 1.3.4. Let $\mathfrak{F}$ be a family of commuting and diagonalizable operators. There exists an ordered basis for V such that every operator in $\mathfrak{F}$ is represented in that basis by a diagonal matrix.

The proofs in this section are analogous to the corresponding results in the previous one.

1.4 Direct-Sum Decompositions

To continue our studies, we'll think less in terms of matrices and more in terms of subspaces. Our goal is to decompose a vector space into the sum of invariant subspaces where the restriction is simple.

Definition 1.4.1 (Independent Subspaces)**.** Let $W_1, \dots, W_k$ be subspaces of V. Then $W_1, \dots, W_k$ are **independent** if

$$v_1 + \cdots + v_k = 0, \quad v_i \in W_i$$

implies that each $v_i = 0$.

Lemma 1.4.1. Let V be finite-dimensional, $W_1, \dots, W_k$ subspaces of V and $W = W_1 + \cdots + W_k$. The following are equivalent.

1. $W_1, \dots, W_k$ are independent.
2. For all $2 \le j \le k$, we have
$$W_j \cap (W_1 + \cdots + W_{j-1}) = \{0\}$$
3. If β_i is an ordered basis for W_i, $1 \le i \le k$, then the sequence $\beta = (\beta_1, \dots, \beta_k)$ is an ordered basis for W.

Remark that if the conditions of the lemma above hold, then $W = W_1 \oplus \cdots \oplus W_k$. For example, if W_i is the eigenspace associated with the eigenvalue λ_i of T and T is diagonalizable, then V can be decomposed as $V = W_1 \oplus \cdots \oplus W_k$.

Theorem 1.4.2. If $V = W_1 \oplus \cdots \oplus W_k$, then there exist k endomorphisms $P_1, \ldots, P_k$ on V such that

1. Each P_i is a projection (i.e. $P_i^2 = P_i$);
2. If $i \neq j$, then $P_i P_j = 0$;
3. $I = P_1 + \cdots + P_k$;
4. The range of P_i is W_i.

Conversely, if $P_1, \ldots, P_k$ are k endomorphisms on V satisfying conditions (1), (2), and (3), and if we let W_i be the range of P_i, then $V = W_1 \oplus \cdots \oplus W_k$.

Proof. Certainly, $V = W_1 + \cdots + W_k$.

By condition (3), $\alpha = P_1\alpha + \cdots + P_k\alpha$. But since each $P_i\alpha$ is in W_i, we have that $\alpha = \alpha_1 + \cdots + \alpha_k$, with each $\alpha_i \in W_i$.

Hence, using (1) and (2) we see that V is the direct sum of the W_i. □

1.5 Invariant Direct Sums

The most interesting decomposition $V = W_1 \oplus \cdots \oplus W_k$ is when each subspace W_i is invariant under a given endomorphism T. In this case, T induces another endomorphism T_i on each W_i, which is simply the restriction of T on W_i.

If $v \in V$, then

$$v = v_1 + \cdots + c_k, \quad v_i \in W_i$$

and

$$T(v) = T_1(v_1) + \cdots + T_k(v_k)$$

Theorem 1.5.1. Let $T, V, W_1, \ldots, W_k, P_1, \ldots, P_k$ be as in Theorem 1.4.2. Then each subspace W_i is invariant under T iff. T commutes with each of the projection P_i, i.e., $T \circ P_i = P_i \circ T$, $i = 1, \ldots, k$.

Proof. ($\Leftarrow$) Suppose that T commutes with each P_i and let $w_j \in W_j$. Then

$$T(w_j) = T(P_j(w_j)) = P_j(T(w_j)) \implies T(w_j) \in \text{range}(P_j)$$

i.e., W_j is invariant under T.

($\Rightarrow$) Suppose that each W_i is invariant under T. Let $v \in V$. Then

$$v = P_1(v) + \cdots + P_k(v) \implies T(v) = T(P_1(v)) + \cdots + T(P_k(v))$$

Since $P_i(v) \in W_i$, which is invariant under T, we have $T(P_i(v)) = P_i(u_i)$ for some u_i. Thus

$$P_j(T(v)) = P_j(T(P_1(v))) + \cdots + P_j(T(P_k(v))) = P_j(u_j) = T(P_j(v))$$

Hence, $P_j(T) = T(P_j)$. □

With this result that invariant direct sum decompositions can be described as projections which commute with T, we can start studying deeper decomposition theorems.

Theorem 1.5.2. Let $T \in \mathrm{End}(V)$, $\dim(V) < \infty$. If T is diagonalizable and $\lambda_1, \dots, \lambda_k$ are the distinct eigenvalues of T, then there exist endomophisms $P_1, \dots, P_k$ on V such that

1. $T = \lambda_1 P_1 + \cdots + \lambda_k P_k$;
2. $I = P_1 + \cdots + P_k$;
3. $P_i P_j = 0$, $i \neq j$;
4. $P_i^2 = P_i$ (P_i is a projection);
5. The range of P_i is the eigenspace for T associated with λ_i.

Conversely, if there exist $\lambda_1, \dots, \lambda_k \in \mathbb{F}$ and non-zero endomorphisms $P_1, \dots, P_k$ satisfying conditions (1), (2), and (3), then T is diagonalizable, $\lambda_1, \cdots, \lambda_k$ are the distinct eigenvalues of T, and conditions (4) and (5) also hold.

Proof. ($\Rightarrow$) Let W_i be the space of eigenvectors associated with the eigenvalue λ_i. Then $V = W_1 \oplus \cdots \oplus W_k$ and let P_i be the projections associated with this decomposition. By Theorem 1.4.2 the conditions (2), (3), (4), and (5) are satisfied. To verify (1), notice that for each $v \in V$,

$$T(v) = T(P_1(v)) + \cdots + T(P_k(v)) = \lambda_1 P_1(v) + \cdots + \lambda_k P_k(v)$$

($\Leftarrow$) Since $P_i P_j = 0$ for $i \neq j$, multiplying both sides of $I = P_1 + \cdots + P_k$ by E_i, we have that $P_i^2 = P_i$.

Multiplying $T = \lambda_1 P_1 + \cdots + \lambda_k P_k$ by E_i implies that $T(P_i) = \lambda_i P_i$ and therefore any vector in the range of P_i is in $\ker(T - \lambda_i I)$. Since $P_i \neq 0$, there exists a non-zero vector in $\ker(T - \lambda_i I)$, i.e., λ_i is an eigenvalue of T.

More than that, λ_i are all the eigenvalues of T. If x is any scalar then $(T - xI)v = 0$ implies that $(\lambda_i - x)P_i(v) = 0$. If $v \neq 0$, then $P_i(v) \neq 0$ for some i.

Since every non-zero vector is in the range of P_i is an eigenvector of T and $I = P_1 + \cdots + P_k$, we have that these eigenvectors span V.

Now notice that if $T(v) = \lambda_i v$ then

$$\sum_{j=1}^{k} (\lambda_j - \lambda_i) P_j(v) = 0 \iff (\lambda_j - \lambda_i) P_j(v) = 0, \quad \forall\, j$$

Thus

$$P_j(v) = 0, \quad j \neq i$$

Since $v = P_1(v) + \cdots + P_k(v)$ and $P_j(v) = 0$ for $j \neq i$, then $v = P_i(v)$, which shows that $v \in \mathrm{range}(P_i)$. □

1.6 Primary Decomposition Theorem

Suppose that the minimal polynomial for T is of the form

$$m_{\mathrm{T}} = (x-\lambda_1)^{r_1}\cdots(x-\lambda_k)^{r_k}$$

where $\lambda_1,\dots,\lambda_k$ are distinct scalars. The next theorem shows that the whole space V is the direct sum of the null spaces of $(\mathrm{T}-\lambda_i\mathrm{I})^{r_i}$, for $i=1,\dots,k$. In fact, the theorem is more general, as we'll see.

Theorem 1.6.1 (Primary Decomposition Theorem (PDT))**.** Let $\mathrm{T}\in\mathrm{End}(\mathrm{V})$, $\dim(\mathrm{V})<\infty$, and $p_1,\dots,p_m$ be the distinct irreducible factors of m_{T} in $\mathbb{F}[x]$. If r_i is the multiplicity of p_i in m_{T}, then

1. $\mathrm{V}=\mathrm{V}_{p_1^{k_1}}\oplus\cdots\oplus\mathrm{V}_{p_m^{k_m}}$;
2. Each $\mathrm{V}_{p_i^{k_i}}$ is invariant under T;
3. If $\mathrm{T}_i=\mathrm{T}|_{\mathrm{W}_i}$, i.e., T_i is the operator induced on W_i by T, then the minimal polynomial for T_i is $p_i^{r_i}$.

The polynomials $p_i^{k_i}$ are called the **primary factors (pf)** of T and $\mathrm{V}_{p_i^{k_i}}$ 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.

Corollary 1.6.2. If $\mathrm{P}_1,\dots,\mathrm{P}_k$ are projections associated with the primary decomposition of T, then each P_i is a polynomial in T, and if an endomorphism U commutes with T, then U commutes with each P_i, i.e., each subspace W_i is invariant under U.

Recall that

Definition 1.6.1 (Nilpotent Operator)**.** Let $\mathrm{N}\in\mathrm{End}(\mathrm{V})$. Then N is **nilpotent** if there exists some positive integer r such that $\mathrm{N}^r=0$.

Under an algebraically closed field, the PDT allows us to 'decompose' an endomophism into a **diagonalizable part** D and a nilpotent operator N, i.e., $\mathrm{T}=\mathrm{D}+\mathrm{N}$. Moreover, D and N commute.

Theorem 1.6.3. Let $\mathrm{T}\in\mathrm{End}(\mathrm{V})$, $\dim(\mathrm{V})<\infty$. Suppose that the minimal polynomial for T decomposes into a product of linear polynomials. Then there exists a diagonalizable operator D and a nilpotent operator N such that

1. $\mathrm{T}=\mathrm{D}+\mathrm{N}$;

2. DN = ND.

D and N are uniquely determined by (1) and (2) and each of them is a polynomial in T.

Corollary 1.6.4. Let V be a finite-dimensional vector space over an algebraically closed field. Then every endomorphism T on V can be written as the sum of a diagonalizable operator D and a nilpotent operator N which commute. These operators D and N are unique and each is a polynomial in T.

This fact is important because it reduces our work to the study of nilpotent operators.

1.7 Alternative proofs for minimal polynomials

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 1.7.1. 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 $T(v) \in V_S$. □

Now, let us try to describe the following set

$$\{p \in \mathbb{F}[x] : V_p \neq \{0\}\} \tag{1.1}$$

We already know that the minimal polynomial is inside this set. Besides, 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 1.7.1 (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 1.7.1. 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 1.1.1, shows the existence of minimal polynomial and how to find it.

Theorem 1.7.2. 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 $\mathfrak{A}_T$.

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 $p \,|\, f$.

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. $\square$

To describe the set (1.1), it is sufficient to consider the following lemma and proposition.

Lemma 1.7.3. 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. $\square$

Proposition 1.7.4. If $\mathfrak{A}_T \neq \{0\}$ and $p \in \mathbb{F}[x]$ is irreducible, then $V_p \neq \{0\}$ iff. $p \,|\, m_T$.

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 $V_p = \{0\}$. $\square$

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 1.1.1, we show an algorithm of how to do it.

Theorem 1.7.5. 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

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

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?

1.8 Cyclic subspaces

Definition 1.8.1 (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

$$\mathfrak{C}_T^{\infty}(v) = (v_k)_{k \geq 0} \quad \textbf{and} \quad \mathfrak{C}_T^{m}(v) = v_0, v_1, \ldots, v_{m-1}$$

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

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

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 1.8.1. For all $v \in \mathrm{V}$, $m_v \,|\, m_{\mathrm{T}}$.

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

$$\mathrm{S} : \mathrm{V}_{m_v} \longrightarrow \mathrm{V}_{m_v}, \quad v \mapsto \mathrm{T}(v)$$

Note that $m_v \in \mathfrak{A}_{\mathrm{S}}$. Thus, $m_{\mathrm{S}} \,|\, m_v$. In fact, $m_v = m_{\mathrm{S}}$, since if $\deg(m_{\mathrm{S}}) = m' < m$, 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 1.8.2 (Coprimality implies direct sum)**.** If each pair $p_1, \ldots, p_m \in \mathbb{F}[x]$ is relatively prime, then the sum $\mathrm{V}^{\infty}_{p_1} + \cdots + \mathrm{V}^{\infty}_{p_m}$ is direct.

Proof. Let $v_j \in \mathrm{V}^{\infty}_{p_j} \setminus \{0\}$, 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 $a_1 v_1 + \cdots + a_m v_m = 0$. 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 $p_j^{k_j}(\mathrm{T})(v_j) = 0$ and consider

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

Notice that

$$0 = p(\mathrm{T})(a_1 v_1 + \cdots + a_m v_m) = a_m p(\mathrm{T})(v_m)$$

By the Lemma 1.7.3, the restriction $p_j^{k_j}(\mathrm{T})$ to $\mathrm{V}_{p_m^{k_m}}$ is injective for $j \neq m$, it follows that $p(\mathrm{T})(v_m) \neq 0$ 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 1.8.3. Let $f_1, \ldots, f_m \in \mathbb{F}[x]$, where each pair is relatively prime and $f = f_1 \ldots f_m$. Then, $\mathrm{V}_f = \mathrm{V}_{f_1} \oplus \cdots \oplus \mathrm{V}_{f_m}$.

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

Then $\mathrm{V}_{f_1} + \mathrm{V}_{f_2} \subseteq \mathrm{V}_f$ 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 $\mathrm{P}_j := h_j(\mathrm{S})$, where S is the induced linear operator by T on V_f (i.e. the restriction of T to V_f).

By construction, $f \in \mathfrak{A}_{\mathrm{S}}$ and $\mathrm{P}_1 + \mathrm{P}_2 = \mathrm{I}_{\mathrm{V}_f}$. Notice that

$$\mathrm{P}_1 \mathrm{P}_2 = g_1(\mathrm{S}) f_1(\mathrm{S}) g_2(\mathrm{S}) f_2(\mathrm{S}) = g_1(\mathrm{S}) g_2(\mathrm{S}) f_1(\mathrm{S}) f_2(\mathrm{S}) = 0 = \mathrm{P}_2 \mathrm{P}_1$$

In particular, $\mathrm{P}_j^2 = \mathrm{P}_j - \mathrm{P}_i \mathrm{P}_j = \mathrm{P}_j$ if $\{i, j\} = \{1, 2\}$ and range $(\mathrm{P}_j) \subseteq \mathrm{V}_{f_i}$. It follows that $\mathrm{V}_f = \text{range } (\mathrm{P}_2) \oplus \text{range } (\mathrm{P}_1) \subseteq \mathrm{V}_{f_1} \oplus \mathrm{V}_{f_2} \subseteq \mathrm{V}_f$. Hence, $\mathrm{V}_{f_1} + \mathrm{V}_{f_2} = \mathrm{V}_f$. □

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 1.8.4. 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 1.8.5. $C_T(v) = \{p(T)(v) : p \in \mathbb{F}[x]\}$.

Proof. **Exercise.** □

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

Theorem 1.8.6. 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 $V_p = C_T(v_1) \oplus \cdots \oplus C_T(v_l)$.

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

$$C_T(v) \cap (C_T(v_1) + \cdots + C_T(v_k)) = \{0\} \tag{1.2}$$

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

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

By the lemma 1.8.4, $C_T(w) = C_T(v)$ and, by the lemma 1.8.5, there exists $f \in \mathbb{F}[x]$ such that $v = f(T)(w)$. Hence,

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

contradicting our hypothesis about v. □

Theorem 1.8.7. 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,

$$\dim(V_p^\infty) = \dim(W) + \dim(\text{range }(R))$$

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

$$T(w) = T(R(v)) = R(T(v)) \in \text{range }(R)$$

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

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

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

$$\deg(m_T) = \sum_{j=1}^{m} k_j \deg(p_j) \leq \sum_{j=1}^{m} \dim(V_{p_j^{k_j}}) = n$$

since

$$k_j = \min\{k : V_{p_j^k} = V_{p_j^{k+1}}\}$$

Define

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

and

$$c_T := \prod_{j=1}^{m} p_j^{n_j}$$

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

If T_j is the restriction of T on $V_{p_j^{k_j}}$,

$$c_T = \prod_{j=1}^{m} c_{T_j}$$

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

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

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

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

$$c_T(\lambda) = \det([T]_\alpha^\alpha - \lambda I)$$

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 1.8.1. Suppose that 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

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

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

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

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 $\mathrm{V}_{(t-2)^2} = \{(2x, y, x, x) : x, y \in \mathbb{F}\}$.

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 $\mathrm{V}_{(t-2)^3} = \{(z, y, x, 5x - 2z) : x, y, z \in \mathbb{F}\}$.

Hence, $m_\mathrm{T} = t(t-2)^3$.

1.9 Jordan and Frobenius Bases

In this section, we assume that V is a finite-dimensional space.

Frobenius basis: always exist. Jordan basis: only when the prime factors have degree one.

Suppose that $m_v(t) = (t–\lambda)^k$, for $k \in \mathbb{Z}_{\geq 0}$ and recall that $\mathrm{C_T}(v)$ is T-invariant and $\dim(\mathrm{C_T}(v)) = k$.

Define $w_j = (\mathrm{T} - \lambda \mathrm{I})^j(v)$, $\beta = w_0, \ldots, w_{k-1}$, and notice that $w_k = 0 = w_{k+1} = \ldots$. Moreover, $m_{w_j} = (t-\lambda)^{k-j}$ for $j < k$. In particular, β is linearly independent and, therefore, a basis of $\mathrm{C_T}(v)$.

Notice that $\mathrm{T}(w_j) = \lambda w_j + w_{j+1}$, and therefore,

$$[\mathrm{S}]_\beta^\beta = \mathrm{J}_k(\lambda) := \begin{bmatrix} \lambda & 0 & \cdots & \cdots & 0 \\ 1 & \lambda & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & \lambda \end{bmatrix}$$

in which $\mathrm{S} \in \mathrm{End}(\mathrm{C_T}(v))$ is given by $\mathrm{S}(w) = \mathrm{T}(w)$ for all $w \in \mathrm{C_T}(v)$. The matrix $\mathrm{J}_k(\lambda)$ is said to be a **Jordan block** of size k and eigenvalue λ.

It is sufficient to reorder the basis β to change to an upper triangular matrix.

The sequence β is a **Jordan** T-**cycle** of size k and eigenvalue λ. A basis formed by the union of Jordan T-cycles is a **Jordan basis** (w.r.t. T).

If β is a Jordan basis of V with respect to T with l T-cycles, we have

$$[\mathrm{T}]_\beta^\beta = \begin{bmatrix} \mathrm{J}_{k_1}(\lambda_1) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathrm{J}_{k_l}(\lambda_l) \end{bmatrix}$$

This matrix is called the **Jordan Canonical Form** of T.

Theorem 1.9.1 (Jordan Decomposition Theorem (JDT))**.** There exists a Jordan basis w.r.t. T for V iff. the prime factors of m_T have degree one. For whatever two bases of Jordan of V w.r.t. T, the quantity of Jordan T-cycles of size k and eigenvalue λ are the same whatever λ and k.

How do we find a Jordan basis? If the T-primary subspaces of V are T-cyclic and we chose a generator for each T-cycle, the PDT and the previous discussion show how to find it.

What to do if any T-primary subspace is not T-cyclic? Or if some prime factor of m_T has a degree greater than one? We use the Cyclic Decomposition Theorem since a cyclic decomposition always exists.

Theorem 1.9.2 (Cyclic Decomposition Theorem (CDT))**.** There exist $m \in \mathbb{Z}, m \geq 1$, and $v_1, \ldots, v_m \in \mathrm{V} \setminus \{0\}$ such that

$$\mathrm{V} = \bigoplus_{j=1}^{m} \mathrm{C}_\mathrm{T}(v_j)$$

and $m_{v_{j+1}} | m_{v_j}$ for all $1 \leq j < m$.

If $u_1, \ldots, u_l \in \mathrm{V} \setminus \{0\}$ satisfying $\mathrm{V} = \bigoplus_{j=1}^{l} \mathrm{C}_\mathrm{T}(u_j)$ and $m_{u_{j+1}} | m_{u_j}$ for all $1 \leq j < l$, then $l = m$ and $m_{u_j} = m_{v_j}$ for all $1 \leq j \leq m$.

The polynomials m_{v_j} are called the **invariant factors** of T.

Recall that if $m_v(t) = t^k - \sum_{j=0}^{k-1} a_j t^j$, then $\mathfrak{C}_\mathrm{T}(v) = v_0, \ldots, v_{k-1}$ with $v_j = \mathrm{T}^j(v)$ is a basis of $\mathrm{C}_\mathrm{T}(v)$.

If S is the induced operator by T on $\mathrm{C}_\mathrm{T}(v)$, we have

$$[\mathrm{S}]_{\mathfrak{C}_\mathrm{T}(v)}^{\mathfrak{C}_\mathrm{T}(v)} = \begin{bmatrix} 0 & \cdots & \cdots & 0 & a_0 \\ 1 & \ddots & & \vdots & a_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & a_{k-1} \end{bmatrix}$$

This matrix is called the **Frobenius matrix** (or **companion matrix**) of the polynomial m_v. A basis formed by the union of sets of the form $\mathfrak{C}_\mathrm{T}(v)$ is said to be a **rational basis** (or **Frobenius basis**) of V w.r.t. T.

Notice that $m_\mathrm{T} = m_{v_1}$. In fact, we know that $m_{v_1} | m_\mathrm{T}$ and, hence, it is sufficient to show that $m_{v_1} \in \mathfrak{A}_\mathrm{T}$. By the CDT, given $w \in \mathrm{V}$, there exist unique $w_j \in \mathrm{C}_\mathrm{T}(v_j)$, $1 \leq j \leq m$ such that

$w = w_1 + \cdots + w_m$. We need to show that $m_{v_1}(\mathrm{T})(w_j) = 0$ for all j. However, $w_j \in \mathrm{C_T}(v_j)$ implies that $m_{v_j}(\mathrm{T})(w_j) = 0$. Therefore, the result follows since $m_{v_j} | m_{v_1}$ for all j.

Lemma 1.9.3. If $\mathrm{V} = \bigoplus_{j=1}^m \mathrm{V}_j$, V_j is T-invariant for all j and T_j is induced by T on V_j. Show that $c_\mathrm{T} = \prod_{j=1}^m c_{\mathrm{T}_j}$.

Proof. Exercise. □

Thus, c_T is the product of invariant factors.

If T has a JCF, its distinct eigenvalues are λ_j, $1 \leq j \leq l$, and k_j is the sum of the sizes of all Jordan blocks with eigenvalue λ_j, then

$$c_\mathrm{T} = \prod_{j=1}^{l} (t - \lambda_j)^{k_j}$$

The amount of Jordan blocks with eigenvalue λ is equal to $\dim(\mathrm{V}_{t-\lambda})$.

How are cyclic decomposition and Jordan decomposition related?

Lemma 1.9.4. If S is the induced operator by T on $\mathrm{V}_{t-\lambda}^{\infty}$, a Frobenius basis w.r.t. $\mathrm{S} - \lambda\mathrm{I}$ is a Jordan basis w.r.t. S.

Proof. Exercise. □

Now, given a Jordan basis, how do we find a rational basis? Let λ_j, $1 \leq j \leq l$, the distinct eigenvalues, and m_j the amount of blocks with eigenvalue λ_j, and $m = \max\{m_j : 1 \leq j \leq l\}$. Let $v_{j,1}, \ldots, c_{j,m_j}$ be vectors that start the Jordan T-cycles with eigenvalue λ_j of a Jordan basis and suppose that they are ordered in such a way that $m_{v_{j,i+1}} | m_{v_{j,i}}$.

If $m_j < i \leq m$, define $v_{j,i} = 0$. Also

$$v_i = v_{1,i} + \cdots + v_{l,i}$$

Lemma 1.9.5.

$$\mathrm{V} = \bigoplus_{i=1}^{m} \mathrm{C_T}(v_i) \quad \text{and} \quad m_{v_{i+1}} | m_{v_i}, \quad \forall\ 1 \leq i \leq m$$

Proof. Exercise. □

Exercise. How can we reverse the result? I.e., given a Rational Form, how can we find a Jordan Canonical Form?

Definition 1.9.1 (Similar Operators). Given two operators $\mathrm{S}, \mathrm{T} \in \mathrm{End}(\mathrm{V})$. We say that S and T are **similar** if there exist α and β bases of V such that $[\mathrm{S}]_\alpha^\alpha = [\mathrm{T}]_\beta^\beta$.

Remark that similarity is an equivalence relation.

Theorem 1.9.6. S and T are similar iff. they have the same Rational Form. If T has a Jordan Form, then $\mathrm{S} \sim \mathrm{T}$ iff. they have the same Jordan Form.

Example 1.9.1. Let us continue the example 1.8.1. To generate the Jordan cycle associated with the eigenvalue $\lambda = 2$, we choose $v \in V_{(t-2)^3} \setminus V_{(t-2)^2}$, say $v = (1,0,0,-2)$. In this case, the Jordan cycle is

$$v_1 = v, \quad v_2 = T(v_1) - 2v_1 = -2(2,1,1,1), \quad v_3 = T(v_2) - 2v_2 = -4e_2$$

To complete the basis, notice that $V_t = [v_4]$, in which $v_4 = (1,-1,0,0)$. Hence, $\beta = \{v_1, v_2, v_3, v_4\}$ is a Jordan basis and

$$[T]_\beta^\beta = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Since $c_T = m_T$, it follows that V is T-cyclic and $C_T(w) = V$, in which $w = v_1 - v_4 = (0,1,0,-2)$. Hence, the vectors $w_1 = w$,

$$w_2 = T(w_1) = -2(1,1,1,3), \quad w_3 = T(w_2) = -4(3,3,2,4), \quad w_4 = T(w_3) = -8(5,6,3,5)$$

form a rational basis for V w.r.t. T.

Given that $m_w(t) = c_T(t) = t^4 - 6t^3 + 12t^2 - 8t$, it follows that

$$T(w_4) = 6w_4 - 12w_3 + 8w_2$$

Thus, taking the basis $\gamma = \{w_1, w_2, w_3, w_4\}$

$$[T]_\gamma^\gamma = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 8 \\ 0 & 1 & 0 & -12 \\ 0 & 0 & 1 & 6 \end{bmatrix}$$

How do we choose the initial vectors of each cycle? Jordan and Frobenius bases are formed by a union of cycles. Each cycle has an initial vector. How do we find it? This choice cannot be random.

Example 1.9.2. Consider $T \in \mathrm{End}(\mathbb{R}^3)$ given by $T(x,y,z) = (0,x,y)$. If α is the standard basis, we have

$$[T]_\alpha^\alpha = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

which is a Jordan block. Thus we can take e_1 as the initial vector of the (unique) cycle.

Notice that $v = e_1 + e_2$ can also be used as an initial vector and we have the cycle $\beta = \{v, e_2 + e_3, e_3\}$. However, if the x-coordinate of the initial vector w is zero, then this vector cannot be chosen, since $\mathfrak{C}_T(w)$ has at most two vectors and the final one is a multiple of e_3.

Finding bases for V_t and V_{t^2} we can see that there doesn't exist T-invariant subspace W satisfying $\mathbb{R}^3 = C_T(w) \oplus W$. I.e., $C_T(w)$ doesn't admit a complementary T-invariant subspace.

Our task, then, is to pick v_1 such that $C_T(v_1)$ has a complementary subspace that can be written as a direct sum of cycles, hence T-invariant.

After that, we find v_2 such that $C_T(v_1) \cap C_T(v_2) = \{0\}$ and $C_t(v_1) \oplus C_T(v_2)$ has a complementary subspace such that it can be written as a direct sum of cycles, hence T-invariant.

Repeating this procedure, we find a rational basis.

Definition 1.9.2 (T-admissible subspace)**.** A T-invariant subspace W of V is said to be T-**admissible** if there exists a T-invariant subspace W' such that $V = W \oplus W'$.

Example 1.9.3. V and $\{0\}$ are T-admissible. If V is T-cyclic and v satisfies $m_v = m_T$, then $V = C_T(v)$ and we can find a decomposition as in the CDT.

To prove the Cyclic Decomposition Theorem, we need the following concept.

Definition 1.9.3 (Conductor)**.** Let W be an invariant subspace of T and $v \in V$. The set

$$\mathfrak{C}_{T,v}(W) = \{p \in \mathbb{F}[x] : p(T)(v) \in W\}$$

is called the T-**conductor of** v **into** W.

If $W = \{0\}$, the conductor is called the T-**annihilator of** v.

Notice that m_v divides every element of $\mathfrak{A}_{T,v}$ and there exists an unique monic polynomial $c_{v,W}$ that divides every element of $\mathfrak{C}_{T,v}(W)$.

It follows that $c_{v,W} | c_{v,W'}$ if $W' \subseteq W$. Therefore, $c_{v,W} | m_v$. If $v \in W$, then $c_{v,W} = 1$.

Definition 1.9.4. A T-invariant subspace W of V is said to be T-**admissible** iff. for all $v \in V$ and $f \in \mathfrak{C}_{T,v}(W)$, there exists $w \in W$ such that $f(T)(w) = f(T)(v)$.

Lemma 1.9.7. Let v be as in the second item of the Theorem 1.9.10. Then, for all $u \in V$ satisfying $C_T(u) \cap (W \oplus C_T(v)) = \{0\}$, $m_u | m_v$.

Lemma 1.9.8. If $v \in V$, then $C_T(v) \cap W = \{0\}$ (i.e. is a direct sum) iff. $m_v = c_{v,W}$.

Lemma 1.9.9. Let W be a T-invariant subspace of V and suppose that $u, v \in V$ such that $v - u \in W$. Then $\mathfrak{C}_{T,v}(W) = \mathfrak{C}_{T,u}(W)$.

Proof. Let $w = v - u$ and notice that

$$f(T)(v) - f(T)(u) = f(T)(w) \in W \quad \forall f \in \mathbb{F}[x]$$

Hence, $f \in \mathfrak{C}_{T,u}(W)$ iff. $f \in \mathfrak{C}_{T,v}(W)$. □

Theorem 1.9.10. Let W be a proper subspace and T-admissible of V. Then the following are true:

1. For all $v \in V \setminus W$, there exists $v' \in V$ such that $m_{v'} = c_{v,W}$ and

$$W + C_T(v) = W \oplus C_T(v')$$

2. $W + C_T(v)$ is T-admissible if $v \in V$ satisfies $C_T(v) \cap W = \{0\}$ and

$$\deg(m_v) = \max\{\deg(m_u) : u \in V, C_T(u) \cap W = \{0\}\}$$

Proof. (1) Using the definition of T-admissible, let $w \in W$ such that $c_{v,W}(T)(v) = c_{v,W}(T)(w)$ and consider $v' = v - w$.

Since $W, C_T(v)$, and $C_T(v')$ are T-invariant,

$$W + C_T(v) = W + C_T(v')$$

and we need to show that $W \cap C_T(v') = \{0\}$.

By the Lemma 1.9.9, $c_{v,W} = c_{v',W}$ and

$$c_{v',W}(T)(v') = c_{v,W}(T)(v - w) = 0$$

Hence, $c_{v',W} \in \mathfrak{A}_{T,v'}$ and thus $c_{v',W} = m_{v'}$. By the Lemma 1.9.8, $C_T(v') \cap W = \{0\}$. Remark that $c_{v,W}$ is the minimal polynomial in the space V/W.

(2) Let $W' = W + C_T(v)$. If $W' = V$, there is nothing to do. Suppose that $W \subsetneq V$ and take $u \in V \setminus W'$. Let $w \in W$ and $p \in \mathbb{F}[x]$ such that

$$c_{u,W'}(T)(u) = w + p(T)(v)$$

We're going to show that

$$c_{u,W'} | p \quad \text{and} \quad w = c_{u,W'}(T)(w') \text{ for some } w' \in W \tag{1.3}$$

Writing the division of p by $c_{u,W'}$, we have that $p = qc_{u,W'} + r$. Define $u' := u - q(T)(v) \in V \setminus W'$. By the Lemma 1.9.9, $c_{u,W'} = c_{u',W'}$.

By the part (1) and choice of v, $\deg(m_v) \geq \deg(c_{u',W})$, and

$$\begin{aligned} c_{u,W'}(T)(u') &= c_{u,W'}(T)(u - q(T)(v)) = c_{u,W'}(T)(u) - (p(T) - r(T))(v) \\ &= w + p(T)(v) - (p(T) - r(T))(v) = w + r(T)(v) \end{aligned}$$

On the other hand, there exists $h \in \mathbb{F}[x]$ such that $c_{u',W} = hc_{u',W'}$. Hence,

$$c_{u',W}(T)(u') = h(T)(c_{u,W'}(T(u'))) = h(T)(w + r(T)(v))$$

and thus $hr \in \mathfrak{C}_{T,v}(W)$.

If $r \neq 0$, then $\deg(h) + \deg(r) \geq \deg(c_{v,W}) = \deg(m_v) \geq \deg(c_{u',W}) = \deg(h) + \deg(c_{u,W'})$ and, therefore, $\deg(r) \geq \deg(c_{u,W'})$, which is impossible. Hence, $r = 0$ and the first part of (1.3) follows.

Now notice that

$$c_{u,W'}(T)(u') = w + r(T)(v) = w \in W$$

shows that $c_{u,W'} \in \mathfrak{C}_{T,u'}(W)$. Given that W is T-admissible, there exists $w' \in W$ such that $c_{u,W'}(T)(w') = c_{u,W'}(T)(u') = w$, proving the second part (1.3).

Take $q \in \mathbb{F}[x]$ such that $p = qc_{u,\mathrm{W}'}$ and $w'' := w' + q(\mathrm{T})(v) \in \mathrm{W}'$. Then

$$c_{u,\mathrm{W}'}(\mathrm{T})(u) = c_{u,\mathrm{W}'}(\mathrm{T})(w') + c_{u,\mathrm{W}'}(\mathrm{T})(q(\mathrm{T})(v)) = c_{u,\mathrm{W}'}(\mathrm{T})(w'')$$

and hence W' is T-admissible. □

Theorem 1.9.11 (Cyclic Decomposition Theorem (Existence Part)). If W is the proper subspace and T-admissible of V, there exist $m \in \mathbb{Z}_{\geq 1}$ and $v_1, \ldots, v_m \in \mathrm{V} \setminus \{0\}$ satisfying

1. The sum $\mathrm{W}' = \sum_{j=1}^{m} \mathrm{C_T}(v_j)$ is direct and $\mathrm{V} = \mathrm{W} \oplus \mathrm{W}'$;
2. $m_{v_{j+1}} | m_{v_j}, \quad \forall\ 1 \leq j < m.$

Proof. We proceed by induction on $k := \dim(\mathrm{V}) - \dim(\mathrm{W}) \geq 1$. If $k = 1$, it is immediate by the previous result. In this case, $m = 1$ and $v_1 = v'$ given by the Theorem 1.9.10 is an eigenvector.

Suppose that $k > 1$ and, by induction hypothesis, the theorem is true for any subspace T-admissible U of V satisfying $\dim(\mathrm{V}) - \dim(\mathrm{U}) < k$. Choose v_1 as in the second item of the Theorem 1.9.10 and let $\mathrm{U} = \mathrm{W} \oplus \mathrm{C_T}(v_1)$, which satisfies our conditions.

Thus, there exist $v_2, \ldots, v_m$ such that $\mathrm{V} = \mathrm{U} \oplus \mathrm{C_T}(v_2) \oplus \cdots \oplus \mathrm{C_T}(v_m)$ and $m_{v_{j+1}} | m_{v_j}$ for all $2 \leq j < m$. The Lemma 1.9.7 gives that $m_{v_2} | m_{v_1}$. □

Lemma 1.9.12. Let $\mathrm{S}, \mathrm{T} \in \mathrm{End}(\mathrm{V})$ and suppose that S is invertible and $\mathrm{S} \circ \mathrm{T} = \mathrm{T} \circ \mathrm{S}$. Then, $m_{\mathrm{T},\mathrm{S}(v)} = m_{\mathrm{T},v}$ for all $v \in \mathrm{V}$.

Proof. Notice that

$$m_v(\mathrm{T})(\mathrm{S}(v)) = \mathrm{S}(m_v(\mathrm{T})(v)) = 0$$

shows that $m_{\mathrm{S}(v)} | m_v$.

On the other hand, let $u = \mathrm{S}(v)$ and recall that $\mathrm{S}^{-1} \circ \mathrm{T} = \mathrm{T} \circ \mathrm{S}^{-1}$. Then

$$m_{\mathrm{S}(v)}(\mathrm{T})(v) = m_u(\mathrm{T})(\mathrm{S}^{-1}(u)) = \mathrm{S}^{-1}(m_u(\mathrm{T})(u)) = 0$$

□

Lemma 1.9.13. If $f \in \mathbb{F}[x], v \in \mathrm{V}, u = f(\mathrm{T})(v)$, and $p = \gcd(f, m_v)$, then $m_v = pm_u$.

Proof. Suppose wlog that $\mathrm{V} = \mathrm{C_T}(v)$. Otherwise, apply the following argument to the induced operator T on $\mathrm{C_T}(v)$.

If f and m_v are coprime, then we shows that $m_u = m_v$. If $\mathrm{S} = f(\mathrm{T})$, then $\mathrm{S} \circ \mathrm{T} = \mathrm{T} \circ \mathrm{S}$ and S is bijective (since $\dim(\mathrm{V}) < \infty$). The conclusion follows from the Lemma 1.9.12.

Now suppose that f is prime. if f does not divide m_v, then $\gcd(f, m_v) = 1$ and we're back to the previous case. If $f = \gcd(f, m_v)$, we show that $m_u = g$, where $g = m_v/f$.

$$g(\mathrm{T})(u) = g(\mathrm{T})(f(\mathrm{T})(v)) = (gf)(\mathrm{T})(v) = m_v(\mathrm{T})(v) = 0$$

showing that $m_u | g$. If $m_u \neq g$, then

$$\deg(m_u) < \deg(g) = \deg(m_v) - \deg(f)$$

However,

$$(m_u f)(\mathrm{T})(v) = m_u(\mathrm{T})(f(\mathrm{T})(v)) = m_u(\mathrm{T})(u) = 0$$

shows that $m_v | m_u f$. Thus

$$\deg(m_v) \leq \deg(m_u) + \deg(f)$$

gives a contradiction. Hence, $m_u = g$.

In the general case, we proceed by induction on $\deg(f) \geq 1$, which starts in the preceeding case. Suppose that $\deg(f) > 1$ and let h be an irreducible factor of f, and g be such that $hg = f$, $u' = g(\mathrm{T})(v)$, $q = \gcd(g, m_v)$, and $r = \gcd(h, m_{u'})$.

Notice that $u = h(\mathrm{T})(u')$. By induction hypothesis,

$$m_v = q m_{u'} \quad \text{and} \quad m_{u'} = r m_u$$

By the properties of greatest common divisor, it follows that $qr = p$. □

Theorem 1.9.14 (Cyclic Decomposition Theorem (Uniqueness Part))**.** If W is a proper and T-admissible subspace of V, and suppose that $v_1, \ldots, v_m \in \mathrm{V} \setminus \{0\}$ and $u_1, \ldots, u_l \in \mathrm{V} \setminus \{0\}$ satisfy

1. The sums $\mathrm{W}' := \sum_{j=1}^{m} \mathrm{C_T}(v_j)$ and $\mathrm{W}'' := \sum_{j=1}^{l} \mathrm{C_T}(u_j)$ are direct;
2. $\mathrm{V} = \mathrm{W} \oplus \mathrm{W}' = \mathrm{W} \oplus \mathrm{W}''$;
3. $m_{v_{j+1}} | m_{v_j}, \quad \forall\ 1 \leq j < m$, and $m_{u_{j+1}} | m_{u_j}, \quad \forall\ 1 \leq j < l$.

Then $l = m$ and $m_{v_j} = m_{u_j}$ for all $1 \leq j \leq m$.

Proof. Suppose $m \leq l$ and notice that $l = m$ will follow if we show that

$$m_{u_j} = m_{v_j} \quad \forall\ 1 \leq j \leq m \tag{1.4}$$

In fact, these identities imply that

$$\dim(\mathrm{V}) = \dim(\mathrm{W}) + \sum_{j=1}^{m} \dim(\mathrm{C_T}(v_j)) = \dim(\mathrm{W}) + \sum_{j=1}^{m} \dim(\mathrm{C_T}(u_j))$$

and therefore there can't be more elements at the decomposition.

Now consider $\mathrm{S}_j := m_{v_j}(\mathrm{T})$, for $1 \leq j \leq m$. Applying S_1 to both decompositions,

$$\mathrm{S}_1(\mathrm{W}) \oplus \mathrm{S}_1(\mathrm{C_T}(u_1)) \oplus \cdots \oplus \mathrm{S}_1(\mathrm{C_T}(u_l)) = \mathrm{S}_1(\mathrm{W})$$

by definition of S_1 and the fact that $m_{v_{j+1}} | m_{v_j}$.

Hence, $m_{v_1}(\mathrm{T})(u_1) = 0$, showing that $m_{u_1} | m_{v_1}$. In a similar way, we show that $m_{v_1} | m_{u_1}$ and, therefore, $m_{u_1} = m_{v_1}$.

If $m = 1$, then (1.4) is proved. Otherwise, apply $\mathrm{S}_1 = m_{v_2}(\mathrm{T})$ to obtain

$$\mathrm{S}_2(\mathrm{W}) \oplus \mathrm{S}_2(\mathrm{C_T}(u_1)) \oplus \cdots \oplus \mathrm{S}_2(\mathrm{C_T}(u_l)) = \mathrm{S}_2(\mathrm{W}) \oplus \mathrm{S}_2(\mathrm{C_T}(v_1))$$

Since $f(\mathrm{T})(\mathrm{C_T}(v)) = \mathrm{C_T}(f(\mathrm{T})(v))$, for all $f \in \mathbb{F}[x]$ and $v \in \mathrm{V}$,

$$\mathrm{S}_2(\mathrm{W}) \oplus \mathrm{C_T}(\mathrm{S}_2(u_1)) \oplus \cdots \oplus \mathrm{C_T}(\mathrm{S}_2(u_l)) = \mathrm{S}_2(\mathrm{W}) \oplus \mathrm{C_T}(\mathrm{S}_2(v_1))$$

By the Lemma 1.9.13, $m_{\mathrm{S}_2(v_1)} = \frac{m_{v_1}}{m_{v_2}} = \frac{m_{u_1}}{m_{v_2}} = m_{\mathrm{S}_2(u_1)}$. Hence,

$$\dim(\mathrm{C_T}(\mathrm{S}_2(v_1))) = \dim(\mathrm{C_T}(\mathrm{S}_2(u_1)))$$

and $\mathrm{C_T}(\mathrm{S}_2(u_j)) = \{0\}$, for all $j \geq 2$. Thus, $m_{v_2}(\mathrm{T})(u_2) = \mathrm{S}_2(u_2) = 0$, showing that $m_{u_2} | m_{v_2}$. Analogously, $m_{v_2} | m_{u_2}$ and, therefore, $m_{u_2} = m_{v_2}$.

Proceding recursively using S_j, the (1.4) is proved. □

Using the Primary Decomposition Theorem and the Cyclic Decomposition Theorem, how can we obtain the Jordan Decomposition Theorem?

Consider the T-primary decomposition of V, $\mathrm{V} = \mathrm{V}_1 \oplus \cdots \oplus \mathrm{V}_m$, and obtain a decomposition of each V_i as in the CDT:

$$\mathrm{V}_i = \mathrm{C_T}(v_{i,1}) \oplus \cdots \oplus \mathrm{C_T}(v_{i,l_i})$$

Then $m_{v_{i,j}} = p_i^{k_{i,j}}$, where $k_{i,1} \geq \ldots \geq k_{i,l_i}$, in which p_i is the corresponding prime factor of m_T. It follows that

$$\mathrm{V} = \bigoplus_{i=1}^{m} \bigoplus_{j=1}^{l_i} \mathrm{C_T}(v_{i,j})$$

If $p_i = t - \lambda_i$ for all i, and for each $v_{i,j}$, construct the corresponding Jordan T-cycle, which we'll call $\mathfrak{J}_{i,j}$. Thus,

$$\mathfrak{J} := \bigcup_{i,j} \mathfrak{J}_{i,j}$$

is a Jordan basis of V w.r.t. T.

Reciprocally, supposing that there exists a Jordan basis w.r.t. T for V, and using such a basis to compute c_T, we conclude that $\deg(p_i) = 1$, for all i.

Theorem 1.9.15. S and T are similar if, and only if, they have the same rational canonical form. If T has a Jordan canonical form, then $\mathrm{S} \sim \mathrm{T}$ if, and only if, they have the same Jordan canonical form.

Proof. Consequence of the Cyclic Decomposition Theorem. □

If $\mathbb{K}$ is a subfield $\mathbb{F}$, it may happen that A and B are similar over $\mathbb{F}$ but not over $\mathbb{K}$.

Example 1.9.4. Let

$$\mathrm{A} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$$

Then over $\mathbb{C}$, $c_\mathrm{A}(t) = t^2 + 1 = (t-i)(t+i)$, and A has the Jordan form

$$\mathrm{J_A} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} = \mathrm{B}$$

Thus, A and B are similar over $\mathbb{C}$.

Notice that the rational form of A is $R_A = A$. Since the characteristic polynomial for B is $c_B(t) = (t-i)(t+i) = c_A(t)$, the rational form of B is

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

Notice, however, that over $\mathbb{R}$ the matrices A and B are not similar

Chapter 2

Bilinear Forms and Geometry

In this chapter, we study geometries that emerge from a generalization of the inner product.

2.1 Multilinear Functions

Definition 2.1.1 (Multilinear Function)**.** Given vector spaces $V_1, \dots, V_k$ and W, consider the vector space

$$\mathfrak{F}(V_1 \times \cdots \times V_k, W)$$

A function $\varphi \in \mathfrak{F}$ is said to be ***k*-linear** if it is linear in each entry. More precisely,

$$\begin{aligned}\varphi(v_1, \dots, v_{i-1}, v + \lambda v', v_{i+1}, \dots, v_k) = \varphi(v_1, \dots, v_{i-1}, v, v_{i+1}, \dots, v_k)\\ + \lambda\varphi(v_1, \dots, v_{i-1}, v', v_{i+1}, \dots, v_k)\end{aligned}$$

for all $1 \leq i \leq k$.

We denote by $\hom_{\mathbb{F}}^k(V_1, \dots, V_k, W)$ the subset of $\mathfrak{F}$ formed by k-linear functions. It can be easily verified that $\hom_{\mathbb{F}}^k$ is a subspace of $\mathfrak{F}$.

If $W = \mathbb{F}$, then an element of $\hom_{\mathbb{F}}^k$ is called a ***k*-linear form**.

Example 2.1.1 (Inner Product)**.** If $\mathbb{F} = \mathbb{R}$, an inner product on V is a bilinear form on V, i.e., an element of $\mathrm{Hom}^2(V, \mathbb{R})$.

Example 2.1.2 (Determinant)**.** A determinant function is a multilinear form.

The next example shows that, in general, the range is not closed under vector sum.

Example 2.1.3. Consider $V = \mathbb{F}^2$, $W = \mathbb{F}^4$, and $\varphi \in \hom_{\mathbb{F}}^2(V, W)$ given by

$$\varphi(v_1, v_2) = (x_1x_2, x_1y_2, y_1x_2, y_1y_2), \quad v_1 = (x_1, y_1),\ v_2 = (x_2, y_2)$$

Notice that φ is a bilinear function. An also

$$(a_1, a_2, a_3, a_4) \in \mathrm{range}(\varphi) \iff a_1a_4 = a_2a_3$$

Now consider $w = (2,2,1,1)$ and $w' = (1,0,1,0)$. Then $w, w' \in \operatorname{range}(\varphi)$, but $w + w' = (3,2,2,1) \notin \operatorname{range}(\varphi)$.

We know that every vector space has a basis. Can we find a basis for $V_1 \times \cdots \times V_k$ using the bases for each V_i? The next result shows that every k linear function is completely determined by a cartesian product of bases.

Theorem 2.1.1. If $\alpha_j = (v_{i,j})_{i \in I_j}$ is a basis for V_j, $1 \leq j \leq k$, $I = I_1 \times \cdots \times I_k$, and $(w_i)_{i \in I}$ is a family on the vector space W, then there exists an unique function $\varphi \in \hom_{\mathbb{F}}^k(V_1, \ldots, V_k, W)$ such that $\varphi(v_{i_1,1}, \ldots, v_{i_k,k}) = w_i$ for all $i = (i_1, \ldots, i_k) \in I$.

Proof. The proof is the same as for linear transformations, just adding the indices. □

The next theorem gives a procedure to find a basis.

Theorem 2.1.2. Let V_j, α_j, $1 \leq j \leq k$, I and W be as in the previous theorem, and let $\beta = (w_s)_{s \in S}$ a basis for W. Given $i = (i_1, \ldots, i_k) \in I$, define $v_i = (v_{i_1,1}, \ldots, v_{i_k,k})$, and given $(i,s) \in I \times S$, let $\varphi_{i,s} \in \hom_{\mathbb{F}}^k(V_1, \ldots, V_k, W)$ be the element satisfying

$$\varphi_{i,s}(v_{i'}) = \delta_{i,i'} w_s, \quad \forall\ i' \in I$$

Then $(\varphi_{i,s})_{(i,s) \in I \times S}$ is a linearly independent family on $\hom_{\mathbb{F}}^k(V_1, \ldots, V_k, W)$ and is a basis if $\dim(V_j) < \infty$ for all $1 \leq j \leq k$. In this case,

$$\dim(\hom_{\mathbb{F}}^k(V_1, \ldots, V_k, W)) = \dim(W) \prod_{j=1}^{k} \dim(V_j)$$

Proof. For each finite subset $\Gamma = \{\gamma_1, \ldots, \gamma_m\} \subseteq I \times S$, we need to show that

$$a_1 \varphi_{\gamma_1} + \cdots + a_m \varphi_{\gamma_m} = 0 \iff a_1 = \cdots = a_m = 0$$

Write $\gamma_j = (i_j, s_j)$ and define

$$\Omega_j = \{l \in \mathbb{Z} : 1 \leq l \leq m, i_l = i_j\}, \quad 1 \leq j \leq m$$

Given $a_1, \ldots, a_m \in \mathbb{F}$, then

$$\varphi = a_1 \varphi_{\gamma_1} + \cdots + a_m \varphi_{\gamma_m} \implies \varphi(v_{i_j}) = \sum_{l \in \Omega_j} a_l w_{s_l}, \quad \forall\ 1 \leq j \leq m$$

Since $l \neq l'$ implies $\gamma_l \neq \gamma_{l'}$, and $l, l' \in \Omega_j$ implies $i_l = i_{l'}$, we must have $s_l \neq s_l'$ under these conditions.

Hence, the family $(w_{s_l})_{l \in \Omega_j}$ is a subfamily of β and, therefore, linearly independent. Thus, $\varphi(v_{i_j}) = 0$ only if $a_l = 0$ for all $l \in \Omega_j$.

Now suppose that $\dim(V_j) < \infty$ for all suitable j and, therefore, I is finite. Given $\varphi \in \operatorname{Hom}^k(V_1, \ldots, V_k, W)$, we need to find a family of scalars $(a_\gamma)_{\gamma \in I \times S}$ such that $a_\gamma \neq 0$ for finite values of γ And

$$\varphi = \sum_{\gamma \in I \times S} a_\gamma \varphi_\gamma$$

To define such a family, notice that for each $i \in I$, there exists a family of scalars $(a_{i,s})_{s\in S}$ with $a_{i,s} \neq 0$ for finite values of s and

$$\varphi(v_i) = \sum_{s \in S} a_{i,s} w_s$$

Thus we have defined the desired family of scalars. Since I is finite, $a_\gamma \neq 0$ for finite values of γ. Moreover, for each $i \in I$, we have that

$$\left(\sum_{\gamma \in I \times S} a_\gamma \varphi_\gamma\right)(v_i) = \sum_{\substack{\gamma=(i,s): \\ s \in S}} a_\gamma \varphi_\gamma(v_i) = \sum_{s \in S} a_{i,s} w_s = \varphi(v_i)$$

□

2.2 Duality

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

Definition 2.2.1 (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 2.2.1. 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 2.2.2 (Evaluation Function)**.** Notice that the function

$$\begin{aligned} V^* \times V &\longrightarrow \mathbb{F} \\ (f, v) &\longmapsto f(v) \end{aligned}$$

is a 2-linear function called **evaluation function**. In general, an element $\mathrm{Hom}(W, V, \mathbb{F})$ is called a **bilinear pairing** between W and V.

Example 2.2.3 (Trace)**.** If A is an $n \times n$ matrix, the **trace** of A is the scalar

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

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 2.2.2 (Dual basis)**.** If $\beta = \{v_1, \dots, v_n\}$ is a basis of V then the **dual basis** of β is the set $\beta^* = \{f_1, \dots, f_n\}$, 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 2.2.1. Let V be a finite-dimensional vector space. Then the dual basis of a basis of V is a basis of V^*.

Proof. Let $\beta = \{v_1, \dots, v_n\}$ 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 $f_1, \dots, f_n$ on V.

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

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

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

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

Theorem 2.2.2. Let $\beta = \{v_1, \dots, v_n\}$ be a basis for a vector space V. Then there is a unique dual basis $\beta^* = \{f_1, \dots, f_n\}$ 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 $v \in \mathrm{V}$ 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

$$\dim \mathrm{N}_f = \dim \mathrm{V} - 1$$

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 2.2.3 (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 $v \in \mathrm{S}$.

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

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

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

Example 2.2.4. 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

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

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

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 $f \in \mathrm{span}(f_3, f_4, f_5)$. 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 $(x_1, x_2, 0, 0, 0) \in \mathrm{W}$, 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 $f \in \mathrm{W}^0$. I.e., $\mathrm{span}(f_3, f_4, f_5) \subset \mathrm{W}^0$.

Now suppose $f \in \mathrm{W}^0$. 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 $e_1 \in \mathrm{W}$ and $f \in \mathrm{W}^0$, 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 $f \in \operatorname{span}(f_3, f_4, f_5)$, i.e., $W^0 \subset \operatorname{span}(f_3, f_4, f_5)$.

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 2.2.3. 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 2.2.4. 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 2.2.5. 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 2.2.5. 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 $\mathbb{R}^5$,

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

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 $a = 0, b = 1$:

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

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

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 2.2.6. 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 2.2.7. 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 2.2.8. Let V be a finite-dimensional vector space. Each basis for V^* is the dual of some basis for V.

Proof. Let $\beta^* = \{f_1, \dots, f_n\}$ be a basis for V^*. Then there exists a basis $\{L_1, \dots, L_n\}$ 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 $\{v_1, \dots, v_n\}$ 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 2.2.9. 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 2.2.4 (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 2.2.10. 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 $u - cv \in N$. Call that scalar $c = g(u)$. Then g is a linear functional on V and N is the null space of g. □

Lemma 2.2.11. 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 $v \in V$ 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 $g = cf$. $\square$

Theorem 2.2.12. 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

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

$\square$

Definition 2.2.5 (Transpose)**.** The **transpose** of a linear transformation $\mathrm{T} : \mathrm{V} \longrightarrow \mathrm{W}$ is the mapping $\mathrm{T}^t : \mathrm{W}^* \longrightarrow \mathrm{V}^*$ such that

$$(\mathrm{T}^t g)(v) = g(\mathrm{T}(v)) = g \circ \mathrm{T}$$

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

Theorem 2.2.13. The transpose $\mathrm{T}^t : \mathrm{W}^* \longrightarrow \mathrm{V}^*$ of a linear transformation $\mathrm{T} : \mathrm{V} \longrightarrow \mathrm{W}$ is a linear transformation.

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

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

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

Now we show that T^t is linear.

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

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

Theorem 2.2.14. 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

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

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 $\operatorname{Im} T^t \subseteq (\ker T)^0$.

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

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

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

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

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

and

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

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

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

Theorem 2.2.15. Let V and W be finite-dimensional vector spaces. Let α be an ordered basis for V with dual basis α^* and β be an ordered basis for W with dual basis β^*. And let $T : V \longrightarrow W$.

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

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

Put another way,

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

Proof. Write $\alpha = v_1, \dots, v_n$, $\beta = w_1, \dots, w_m$, $\alpha^* = f_1, \dots, f_n$, and $\beta^* = g_1, \dots, g_m$.

If $[T]_{\alpha,\beta} = (a_{i,j})$, then

$$T(v_j) = \sum_{i=1}^{m} a_{i,j} w_i$$

On the other hand, the entry (j,i) of $[T^t]_{\beta^*,\alpha^*}$ is the jth coordinate of $T^t(g_i)$ with respect to α^*.

Notice that if $f = a_1 f_1 + \cdots + a_n f_n \in V^*$, we have that $f(v_j) = a_j$ for all $1 \le j \le n$. Hence,

$$T^t(g_i)(v_j) = g_i(T(v_j)) = g_i\left(\sum_{k=1}^{m} a_{k,j} w_k\right) = \sum_{k=1}^{m} a_{k,j} g_i(w_k) = a_{i,j}$$

□

Definition 2.2.6 (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 2.2.16. Let $A \in \mathbf{M}_{m\times n}(\mathbb{F})$. Then the row rank of A is equal to the column rank of A.

2.3 Bilinear Pairings and Orthogonality

Recall that given two vector spaces V and W, and

$$B(V,W) = \hom^2_{\mathbb{F}}(V,W,\mathbb{F}) \quad \text{and} \quad B(V) = B(V,V)$$

an element of $B(V,W)$ is said to be a **bilinear pairing** between V and W. In the case $V = W$, the bilinear pairing is said to be a **bilinear form** on V.

If $\mathbb{F} = \mathbb{R}$, an inner product on V is an example of a bilinear form. However, if $\mathbb{F} = \mathbb{C}$, an inner product is not a bilinear form since it is not linear on the second entry.

Example 2.3.1. Suppose that $\dim(V) = m$ and $\dim(W) = n$ are finite, and let α and β be bases for V and W respectively. Then for any matrix $A \in M_{m,n}(\mathbb{F})$,

$$\varphi(v,w) = [v^t]_\alpha \, A \, [w]_\beta$$

defines an element $\varphi \in B(V,W)$.

Given that, we can construct a matrix representation of a bilinear pairing analogous to the Gram matrix.

Definition 2.3.1. Let $\varphi \in \mathrm{B}(\mathrm{V},\mathrm{W})$ and families $\alpha = v_1,\dots,v_m \in \mathrm{V}$ and $\beta = w_1,\dots,w_n \in \mathrm{W}$.

The matrix of φ with respect to α and β, denoted by ${}_\alpha[\varphi]_\beta$, is defined as the matrix whose entry in the (i,j) position is $\varphi(v_i,w_j)$, i.e.,

$$ {}_\alpha[\varphi]_\beta = \begin{bmatrix} \varphi(v_1,w_1) & \varphi(v_1,w_2) & \cdots & \varphi(v_1,w_n) \\ \varphi(v_2,w_1) & \varphi(v_2,w_2) & \cdots & \varphi(v_2,w_n) \\ \vdots & \vdots & \cdots & \vdots \\ \varphi(v_m,w_1) & \varphi(v_m,w_2) & \cdots & \varphi(v_m,w_n) \end{bmatrix} $$

Remark that if α and β are bases,

$$ \varphi(v,w) = [v^t]_\alpha \; {}_\alpha[\varphi]_\beta \; [w]_\beta, \quad \forall v \in \mathrm{V},\ w \in \mathrm{W} $$

From these facts, the next result follows.

Theorem 2.3.1. If $\dim(\mathrm{V}) = m$ and $\dim(\mathrm{W}) = n$, and α and β are bases for V and W respectively, then the following function is an isomorphism:

$$ \begin{aligned} \mathrm{B}(\mathrm{V},\mathrm{W}) &\longrightarrow \mathrm{M}_{m,n}(\mathbb{F}) \\ \varphi &\longmapsto {}_\alpha[\varphi]_\beta \end{aligned} $$

In particular,

$$ \dim(\mathrm{B}(\mathrm{V},\mathrm{W})) = \dim(\mathrm{V})\dim(\mathrm{W}) $$

The same question from canonical forms now appears. Given a bilinear pairing, how can we find bases in which its matrix representation is as simple as possible?

Definition 2.3.2 (Radical)**.** Let $\varphi \in \mathrm{B}(\mathrm{V},\mathrm{W})$ and consider the functions

$$ {}_\varphi\mathrm{D} : \mathrm{V} \longrightarrow \mathrm{W}^* \quad \text{and} \quad \mathrm{D}_\varphi : \mathrm{W} \longrightarrow \mathrm{V}^* $$

defined as

$$ {}_\varphi\mathrm{D}(v)(w) = \varphi(v,w) = \mathrm{D}_\varphi(w)(v) \quad \forall\ v \in \mathrm{V},\ w \in \mathrm{W} $$

Notice that both ${}_\varphi\mathrm{D}$ and D_φ are linear mappings.

The kernel of ${}_\varphi\mathrm{D}$ is called the **left radical** of φ, while $\ker(\mathrm{D}_\varphi)$ is the **right radical** of φ.

Definition 2.3.3 (Degenerate and Singular)**.** We say that φ is **left nondegenerate** if ${}_\varphi\mathrm{D}$ is injective. Otherwise, it is said to be **left singular**. Analogously, we define **right nondegenerate** and **right singular** for D_φ.

The vectors on $\ker(\mathrm{D}_\varphi)$ are said to be **right degenerate**, while the vectors on $\ker({}_\varphi\mathrm{D})$ are called **left degenerate** w.r.t. φ.

If $\alpha = v_1,\dots,v_m$ and $\beta = w_1,\dots,w_n$ are bases for V and W respectively, then

$$ [\mathrm{D}_\varphi]_{\beta,\alpha^*} = {}_\alpha[\varphi]_\beta \quad \text{and } [{}_\varphi\mathrm{D}]_{\alpha,\beta^*} = ([\mathrm{D}_\varphi]_{\beta,\alpha^*})^t $$

Hence,

Theorem 2.3.2. If $\dim(V)$ and $\dim(W)$ are finite, then the rank of D_φ and ${}_\varphi D$ are the same.

Corollary 2.3.3. Suppose that φ is left (right) nondegenerate and that $\dim(V)$ or $\dim(W)$ is finite. Then the following are equivalent:

1. φ is right (left) nondegenerate;
2. $\dim(V) = \dim(W)$;
3. ${}_\varphi D$ is an isomorphism;
4. D_φ is an isomorphism.

From now on, let $\varphi \in B(V)$. If $\dim(V) < \infty$, then φ is left degenerate iff. it is right degenerate. Thus, we only say that φ is (not) degenerate.

Definition 2.3.4 (Symmetric, Antisymmetric and Alternating)**.** The bilinear form φ is said to be

- **Symmetric** if $\varphi(v, w) = \varphi(w, v)$;
- **Antisymmetric** (or **skew-symmetric**) if $\varphi(v, w) = -\varphi(w, v)$;
- **Alternating** if $\varphi(v, v) = 0$

for all $v, w \in V$.

The geometry of spaces with a symmetric bilinear form is called an **orthogonal geometry**. If the bilinear form is alternating, it is called a **sympletic geometry**.

If $\operatorname{char}(\mathbb{F}) = 2$, then φ is symmetric iff. it is antisymmetric.

How to find if a bilinear form is symmetric, antisymmetric, or alternating?

Theorem 2.3.4. Let $\alpha = (v_i)_{i \in I}$ a basis for V.

1. φ is symmetric iff. ${}_\alpha[\varphi]_\alpha$ is symmetric.
2. φ is antisymmetric iff. ${}_\alpha[\varphi]_\alpha$ is antisymmetric.
3. φ is alternating iff. ${}_\alpha[\varphi]_\alpha$ is antisymmetric and $\varphi(v_i, v_i) = 0$ for all $i \in I$.

Definition 2.3.5 (Orthogonality)**.** Let $v, w \in V$, we say that v is **orthogonal** to w if $\varphi(v, w) = 0$. I.e.,

$$v \perp_\varphi w \iff \varphi(v, w) = 0$$

Thus, $\perp_\varphi$ defines a binary relation on V. In general, $\perp_\varphi$ is not symmetric.

Theorem 2.3.5. The orthogonality relation $\perp_\varphi$ is symmetric iff. φ is symmetric or alternating.

Definition 2.3.6 (Isotropic)**.** A vector v is **isotropic** w.r.t. φ if $v \perp_\varphi v$, i.e., its length is zero.

With this terminology we see that φ is alternating iff. every vector is isotropic.

Definition 2.3.7 (Orthogonal Complement)**.** Given a family α of vectors in V, we define

$$\alpha^{\perp_\varphi} = \{w \in \mathrm{V} : v \perp_\varphi w,\ v \in \alpha\} \quad \text{and} \quad {}^{\perp_\varphi}\alpha = \{v \in \mathrm{V} : v \perp_\varphi w,\ w \in \alpha\}$$

These sets are subspaces, however $\alpha^{\perp_\varphi} \neq {}^{\perp_\varphi}\alpha$ in general. The equality holds only if φ is symmetric or antisymmetric. Moreover,

$$\ker(\mathrm{D}_\varphi) = \mathrm{V}^{\perp_\varphi} \quad \text{and} \quad \ker({}_\varphi\mathrm{D}) = {}^{\perp_\varphi}\mathrm{V}$$

Theorem 2.3.6. If $\dim(\mathrm{V}) < \infty$ then the following are equivalent.

1. φ is degenerate;
2. $\mathrm{V}^{\perp_\varphi} \neq \{0\}$;
3. ${}^{\perp_\varphi}\mathrm{V} \neq \{0\}$.

Also

$$\mathrm{rank}(\varphi) = \dim(\mathrm{V}) - \dim(\mathrm{V}^{\perp_\varphi}) = \dim(\mathrm{V}) - \dim({}^{\perp_\varphi}\mathrm{V})$$

2.4 Hyperbolic and Orthogonal Bases

Denote by $\mathrm{B}_s(\mathrm{V})$ and $\mathrm{B}_a(\mathrm{V})$ the subspaces of $\mathrm{B}(\mathrm{V})$ formed by all symmetric and alternating bilinear forms on V, respectively. Also define $\mathrm{B}_{as}(\mathrm{V}) = \mathrm{B}_s(\mathrm{V}) \cup \mathrm{B}_a(\mathrm{V})$.

In this section, suppose that $\dim(\mathrm{V})$ is finite and $\varphi \in \mathrm{B}_{as}(\mathrm{V})$. By the Theorem 2.3.5, $\perp_\varphi$ is symmetric and then ${}^{\perp_\varphi}\alpha = \alpha^{\perp_\varphi}$. For simplicity, we will write $\perp$ instead $\perp_\varphi$.

Our goal in this section is to find a basis α of V such that ${}_\alpha[\varphi]_\alpha$ be the 'simplest' as possible. We also will write $[\varphi]_\alpha$ instead of ${}_\alpha[\varphi]_\alpha$.

Definition 2.4.1 (Degenerate, Radical for Subspaces)**.** A subspace W of V is said to be **degenerate** (or **singular**) w.r.t. φ if $\varphi|_\mathrm{W}$ is degenerate. And we define the **radical** of W as

$$\mathrm{rad}(\mathrm{W}) := \mathrm{W} \cap \mathrm{W}^\perp$$

Notice that the rank of $\varphi|_\mathrm{W}$ is

$$\mathrm{rank}(\varphi|_\mathrm{W}) = \dim(\mathrm{W}) - \dim(\mathrm{rad}(\mathrm{W}))$$

Therefore,

Theorem 2.4.1. W is degenerate iff. $\mathrm{rad}(\mathrm{W}) \neq \{0\}$.

More than that,

Theorem 2.4.2.

$$\mathrm{V} = \mathrm{V}^\perp \oplus \mathrm{W} \implies \mathrm{rad}(\mathrm{W}) = \{0\}$$

Proof. In fact, if $w \in \mathrm{rad}(W)$, then $v \perp w$ for all $v \in V^\perp$ and $w' \perp w$ for all $w' \in W$. Hence, $w \in V^\perp \cap W = \{0\}$. $\square$

The next result shows some properties of the orthogonal complement.

Theorem 2.4.3. Suppose that φ is nondegenerate and let W be a subspace of V. The following are true.

1. $\dim(V) = \dim(W) + \dim(W^\perp)$;
2. $V = W + W^\perp$ iff. $W \cap W^\perp = \{0\}$;
3. $(W^\perp)^\perp = W$;
4. $\mathrm{rad}(W) = \mathrm{rad}(W^\perp)$. In particular, W is nondegenerate iff. $W^\perp$ is nondegenerate.

Proof. Let $T : V \longrightarrow W^*$ defined by $v \longmapsto {}_\varphi D(v)|_W$.

Since every element of W^* is the restriction on W of an element of V^*, and ${}_\varphi D$ is surjective (since the domain and codomain have the same finite dimension and φ is injective), T is also surjective and we have that

$$\dim(V) = \dim(W^*) + \dim(\ker(T))$$

On the other hand, $v \in \ker(T)$ iff. $\varphi(v, w) = 0$ for all $w \in W$. I.e., $\ker(T) = W^\perp$. This proves the first item of the theorem. The second item follows immediately from the first.

Notice that $W \subseteq (W^\perp)^\perp$ is always true, even when φ is degenerate. The first item gives that $\dim((W^\perp)^\perp) = \dim(W)$. Since the dimension is finite, the third item follows.

Finally,

$$\mathrm{rad}(W^\perp) = W^\perp \cap (W^\perp)^\perp = W^\perp \cap W = \mathrm{rad}(W)$$

$\square$

The next example shows that if we remove the hypothesis that φ is nondegenerate, then the previous result may not be the case.

Example 2.4.1. Suppose that $V = \mathbb{F}^2$ and $[\varphi]_\alpha = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ where α is the standard basis. Notice that this bilinear form is symmetric and degenerate.

Let $W = [e_2]$ and note that $V^\perp = W$ and $W^\perp = V$. This shows that the first item of the preceding theorem does not hold.

Since $V = V + V^\perp$ and $V \cap V^\perp = [e_2]$, the second item is also not true.

Also,

$$([e_1]^\perp)^\perp = [e_2]^perp = V$$

and then the third item is not true.

And since $\mathrm{rad}([e_1]) = \{0\}$ and $\mathrm{rad}([e_1]^\perp) = \mathrm{rad}([e_2]) = [e_2]$, the last item also does not hold.

What happens if we change the hypothesis to the subspace? I.e., in which cases the orthogonal complement is, in fact, a complement?

Theorem 2.4.4. Let W be a subspace of V. Then $V = W \oplus W^{\perp}$ iff. W is nondegenerate.

Proof. Suppose that $V = W \oplus W^{\perp}$. Then since $\mathrm{rad}(W) = W \cap W^{\perp}$, it follows that $\mathrm{rad}(W) = 0$.

Reciprocally, suppose that W is nondegenerate and define $\psi := \varphi|_W$. Then the linear mapping $T : V \longrightarrow W^*$, where $v \longmapsto {}_{\varphi}D(v)|_W$ is surjective, since

$${}_{\varphi}D|W = {}_{\psi}D : W \longrightarrow W^*$$

is an isomorphism (it is injective since W is nondegenerate and surjective since the dimension is equal and finite).

The conclusion follows the first and second items of the previous theorem. □

Hyperbolic bases generalize orthogonal bases for the inner product in the case that φ is an alternating bilinear form. To do that, suppose that φ is alternating.

Definition 2.4.2 (Hyperbolic)**.** An ordered pair (v, w) of vectors in V is said to be a **hyperbolic pair** if $\varphi(v, w) = -1$. In this case, the subspace spanned by these vectors is called an **hyperbolic plane**.

We say that the space V is **hyperbolic** if it is a direct sum of hyperbolic planes mutually orthogonal.

Suppose that V is hyperbolic, say $V = V_1 \oplus \cdots \oplus V_m$, in which V_j is a hyperbolic plane spanned by v_j, w_j with (v_j, w_j) a hyperbolic pair for all $1 \leq j \leq m$.

Then, if $\alpha = v_1, w_1, \ldots, v_m, w_m$ is a basis for V, we have that

$$[\varphi]_\alpha = H_m := \begin{bmatrix} H & 0 & 0 & \cdots & 0 \\ 0 & H & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \ddots & \ddots & H \end{bmatrix}, \quad H = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

Remark that φ is nondegenerate.

Theorem 2.4.5. If $\varphi \in B_a(V)$, then every subspace W of V complementary to $V^{\perp}$ is hyperbolic. In particular, there exists a basis α of V such that

$$[\varphi]_\alpha = \begin{bmatrix} 0 & 0 \\ 0 & H_m \end{bmatrix}, \qquad m = \mathrm{rank}(\varphi)/2$$

Proof. By Theorem 2.4.2, W is nondegenerate. Suppose, w.l.o.g., that φ is nondegenerate. We proceed by induction on $\dim(V)$.

If $V = \{0\}$, there is nothing to do. Otherwise, there exists $v, u \in V$ such that $\varphi(v, u) = a$, $a \in \mathbb{F} \setminus \{0\}$.

Defining $w = -a^{-1}u$, then (v, w) is a hyperbolic pair.

Let $V_1 = [v, w]$ and $V' = V_1^{\perp}$. Since V_1 is nondegenerate (it is a hyperbolic plane), it follows from the Theorem 2.4.4 that $V = V_1 \oplus V_1^{\perp}$. And since φ is nondegenerate, from the

Theorem 2.4.3 we have that $V_1^\perp$ is nondegenerate.

Using the induction hypothesis, $V_1^\perp$ is hyperbolic and thus V is also hyperbolic. □

Example 2.4.2. Suppose that $\operatorname{char}(\mathbb{F}) = 0$, $V = \mathbb{F}^4$, $v = (x_1, x_2, x_3, x_4)$ and $w = (y_1, y_2, y_3, y_4)$. Then

$$\varphi(v, w) = x_2y_1 - x_1y_2 - 2x_1y_4 + 2x_4y_1 + 3x_3y_4 - 3x_4y_3$$

If α is the standard basis,

$$[\varphi]_\alpha = \begin{bmatrix} 0 & -1 & 0 & -2 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 \\ 2 & 0 & -3 & 0 \end{bmatrix}$$

Notice that φ is antisymmetric and alternating. Our task now is to find a hyperbolic basis.

Take (e_1, e_2), which is a hyperbolic pair and let $V_1 = [e_1, e_2]$.

To find $V_1^\perp$, we have that $v \in V_1^\perp$ iff. $v \perp e_1$ and $v \perp e_2$. Computing $[\varphi]_\alpha \cdot e_1 = 0$ we find that $x_2 + 2x_4 = 0$. Computing $[\varphi]_\alpha \cdot e_2 = 0$, we have that $x_1 = 0$. Thus,

$$V_1^\perp = \{(0, -2x_4, x_3, x_4) : x_3, x_4 \in \mathbb{F}\}$$

and $e_3, u = (0, -2, 0, 1)$ form a basis for $V_1^\perp$.

Since $\varphi(e_3, u) = 3$, we have that $(e_3, -u/3)$ is a hyperbolic pair.

Hence, $e_1, e_2, e_3, -u/3$ is a hyperbolic basis for V w.r.t. φ.

Remark that if we use an orthogonal basis for an alternating form, then the matrix is zero and the bilinear form is also zero.

In the symmetric case we have orthogonal bases, which is the direct generalization of the inner product.

Definition 2.4.3 (Orthogonal Family/Basis)**.** A family α is said to be **orthogonal** (w.r.t. φ) if, for all distinct $v, v' \in \alpha$, we have that $v \perp v'$.

If $\alpha = v_1, \dots, v_n$ is an orthogonal basis for $\varphi \in B(V)$, then

$$\#\{i : \varphi(v_i, v_i) \neq 0\} = \operatorname{rank}(\varphi)$$

and the vectors v_i such that $\varphi(v_i, v_i) = 0$ form a basis for $V^\perp$, i.e.,

$$[\{v_i : 1 \leq i \leq n,\ \varphi(v_i, v_i) = 0\}] = V^\perp$$

Theorem 2.4.6 (Existence of orthogonal basis for symmetric forms)**.** If $\operatorname{char}(\mathbb{F}) \neq 2$ and $\varphi \in B_s(V)$, there exists an orthogonal basis for V.

Proof. If $\varphi = 0$ or $V = \{0\}$, there is nothing to do. Otherwise, since $\operatorname{char}(\mathbb{F}) \neq 2$, we know that $B_s(V) \cap B_a(V) = \{0\}$. Therefore, φ is not alternating and there exists $v_1 \in V$ such that $\varphi(v_1, v_1) \neq 0$.

Define $V_1 = [v_1]$. It follows from Theorem 2.4.4 that $V = V_1 \oplus V_1^\perp$.

Since the restriction of φ to $V_1^\perp$ is symmetric, by induction hypothesis on the dimension of the space, there exists an orthogonal basis for $V_1^\perp$ w.r.t. φ. Complementing v_1 with this basis, we form a basis for V which is orthogonal w.r.t. φ. □

Note that if $\mathrm{char}(\mathbb{F}) = 2$, this is not always possible, since $B_a(V) \subseteq B_s(V)$.

The question now is: is there an orthonormal basis? This is important because the inner product with any orthonormal basis is the usual inner product of $\mathbb{R}^n$.

Let W be a complementary subspace to $V^\perp$ and let $\beta = w_1, \dots, w_p$ an orthogonal basis for W w.r.t. a symmetric bilinear form φ.

Suppose that, for all $1 \le i \le p$, there exists $a_i \in \mathbb{F}$ such that

$$a_i^2 = \frac{1}{\varphi(w_i, w_i)}$$

Then, given $v_i = a_i w_i$,

$$\varphi(v_i, v_i) = 1 \quad \forall\ 1 \le i \le p$$

In this case, if γ is a basis for $V^\perp$, we define $\alpha = \gamma \cup \{v_1, \dots, v_p\}$ and have that

$$[\varphi]_\alpha = \begin{bmatrix} 0 & 0 \\ 0 & I_p \end{bmatrix}$$

Definition 2.4.4 (Orthonomal basis). If φ is nondegenerate, then α is an **orthonormal basis** w.r.t. φ.

Such normalization is only possible if $\mathbb{F}$ has the square roots of the elements $\varphi(w_i, w_i)$. If $\mathbb{F} \subseteq \mathbb{R}$, this may not be the case, since it is possible to have $w \in V$ such that $\varphi(w, w) < 0$. In this case, we may choose $a \in \mathbb{R}$ such that

$$a^2 = \frac{-1}{\varphi(w, w)}$$

Taking $v = aw$, then $\varphi(v, v) = -1$. This shows that if $\mathbb{F} = \mathbb{R}$ and $\mathrm{rank}(\varphi) = p$, there exists $0 \le i \le p$ and a basis $\alpha = v_1, \dots, v_n$ of V such that

$$[\varphi]_\alpha = \begin{bmatrix} -I_i & 0 & 0 \\ 0 & I_{p-i} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Definition 2.4.5 (Sylvester Basis). When $\mathbb{F} \subseteq \mathbb{R}$, this basis is called a **Sylvester basis** for V w.r.t. φ. However, this is not an usual terminology.

Definition 2.4.6 (Positive definite). Let $\mathbb{F} \subseteq \mathbb{R}$. We say that $\varphi \in B_s(V)$ is **positive semidefinite** if $\varphi(v, v) \ge 0$ for all $v \in V$.

We say that φ is **positive definite** if $\varphi(v, v) > 0$ for all $v \in V \setminus \{0\}$.

The concepts of negative (semi)definite are analogous.

In this language, an inner product over $\mathbb{F} \subseteq \mathbb{R}$ is a symmetric bilinear form which is positive definite.

Recall that $A \in M_n(\mathbb{C})$ is positive definite if $X^*AX \in \mathbb{R}_{>0}$, for all $X \in M_{n,1}(\mathbb{C}) \setminus \{0\}$.

Given a basis α for V, φ is symmetric iff. $[\varphi]_\alpha$ is symmetric (and hence diagonalizable over $\mathbb{R}$). It follows that φ is an inner product iff. φ is positive definite or, equivalently, $[\varphi]_\alpha$ is positive definite (and hence all its eigenvalues are positive, by the Spectral Theorem). This allows us to use results from eigenvalue theory to discover whether a bilinear form is an inner product.

Definition 2.4.7 (Negativity Index)**.** The **negativity index** of φ is defined as

$$i(\varphi) = \max\{\dim(W) : W \text{ is a subspace such that } \varphi|_W < 0\}$$

Clearly, $i(\varphi) \le \mathrm{rank}(\varphi)$, $\varphi \ge 0$ (is positive semidefinite) iff. $i(\varphi) = 0$, and $\varphi < 0$ iff $i(\varphi) = \dim(V)$.

Definition 2.4.8 (Signature)**.** The **signature** of φ is

$$\mathrm{sign}(\varphi) = \mathrm{rank}(\varphi) - 2i(\varphi)$$

Remark that in the Equation 2.4, the signature is the trace of the matrix.

Theorem 2.4.7 (Sylvester's Inertial Law)**.** Suppose that $\mathbb{F} \subseteq \mathbb{R}$ and let $\varphi \in B_s(V)$ and $\alpha = v_1, \dots v_n$ is an orthogonal basis for V w.r.t. φ. Then

$$i(\varphi) = \#\{j : \varphi(v_j, v_j) < 0\}$$

Proof. Let $i = \#\{j : \varphi(v_j, v_j) < 0\}$. Clearly, $i \le p = \mathrm{rank}(\varphi)$.

Suppose w.l.o.g. that $\varphi(v_j, v_j) = 0$ if $j > p$ and that $\varphi(v_j, v_j) < 0$ for $j \le i$ (just ordering the vectors). Let $W^- = [v_1, \dots, v_i]$ and $W^+ = [v_{i+1}, \dots, v_p]$.

Since $\varphi|_{W^-} < 0$, we have that $i \le i(\varphi)$. We need to show that

$$\varphi|_W < 0 \implies \dim(W) \le i$$

which is equivalent to

$$\varphi|_W < 0 \implies W \cap (V^\perp \oplus W^+) = \{0\} \tag{2.1}$$

Since $V^\perp = [v_{p+1}, \dots, v_n]$, we have that $V = W^- \oplus W^+ \oplus V^\perp$. Supposing that 2.1 holds,

$$\begin{aligned} n &\ge \dim(W) + \dim(V^\perp) + \dim(W^+) = \dim(W) + (n-p) + (p-i) \\ &= \dim(W) + n - i \end{aligned}$$

Showing that $\dim(W) \le i$.

To show that 2.1 is true, let $w \in W \cap (V^\perp \oplus W^+)$, say, $w = u + v$, $u \in W^+$ and $v \in V^\perp$. Note that

$$\varphi(w, w) = \varphi(u, u) + 2\varphi(u, v) + \varphi(v, v) = \varphi(u, u) \ge 0$$

Since $\varphi|_W < 0$, it follows that $w = 0$. □

Example 2.4.3. Let $V = \mathbb{F}^3$ and

$$[\varphi]_\alpha = \begin{bmatrix} 2 & -2 & 1 \\ -2 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

where α is the standard basis.

First step: Take a vector v_i such that $\varphi(v_i, v_i) \neq 0$. Notice that $\varphi(e_3, e_3) = 1$.

Second step: Find the orthogonal complement of the chosen vector. Computing $[\varphi]_\alpha \cdot [x, y, z]^t = 0$,

$$W := [e_3]^\perp = \{(x, y, z) \in V : x + z = 0\} = [w, e_2] \quad w = (1, 0, -1)$$

Third step: Find a vector in the orthogonal complement which is non-zero. Note that $\varphi(w, w) = 1$, $\varphi(e_2, e_2) = 0$, and $\varphi(w, e_2) = -2$. Thus, defining $\psi = \varphi|_W$ and $\beta = \{w, e_2\}$,

$$[\psi]_\beta = \begin{bmatrix} 1 & -2 \\ -2 & 0 \end{bmatrix}$$

Fourth step: Find the orthogonal complement of the chosen vector inside the previous orthogonal complement, i.e., find $[w]^\perp \cap W = [w]^{\perp_\psi}$.

Given $u = aw + be_2$,

$$\psi(w, u) = a - 2b = 0 \iff u = 2w + e_2 = (2, 1, -2)$$

Hence, $[w]^{\perp_\psi} = [u] = [(2, 1, -2)]$.

Notice that $\varphi(u, u) = -4$ and that $\gamma = \{e_3, w, u\}$ is an orthogonal basis for V. It follows that

$$[\varphi]_\gamma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -4 \end{bmatrix}$$

Moreover, $\operatorname{rank}(\varphi) = 3$, $i(\varphi) = 1$ and $\operatorname{sign}(\varphi) = 1$.

Replacing u by $(1, 1/2, -1)$ we obtain a Sylvester basis for V.

Notice that to describe the isotropic vectors w.r.t. φ is easier using an orthogonal basis like γ:

$$[v]_\gamma = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \implies \varphi(v, v) = [v]_\gamma^t [\varphi]_\gamma [v]_\gamma = y_1^2 + y_2^2 - 4y_3^2$$

Therefore, v is isotropic iff.

$$v = y_1 e_3 + y_2 w + y_3 u, \quad |y_3| = \frac{1}{2}\sqrt{y_1^2 + y_2^2}$$

A model of space with negative length vectors, such as in the previous example (but with one more dimension), is the **Minkowski space**, which has a bilinear form that, restricted to the physical space, is an inner product, but in the time axis, the vectors have negative length. The **light cone** in the preceding example is the collection of isotropic vectors.

To end this section, we turn our attention into quadratic forms.

Definition 2.4.9. A **quadratic form** on n variables with values in $\mathbb{F}$ is a polynomial function derived from a homogeneous polynomial with degree two and coefficients in $\mathbb{F}$.

One way of producing such polynomial is from a bilinear form $\varphi \in \mathrm{B}(\mathrm{V})$ and a basis α of a vector space V:

$$q_{\varphi}(x_1,\ldots,x_n) = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} [\varphi]_{\alpha} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Reciprocally, every quadratic form can be derived from a bilinear form. Given a homogeneous polynomial of degree two, say

$$q(x_1,\ldots,x_n) = \sum_{1\leq i\leq j\leq n} c_{i,j}x_ix_j$$

Choose the matrix $\mathrm{A} = (a_{i,j})_{1\leq i\leq j\leq n}$ such that $a_{i,j}+a_{j,i} = c_{i,j}$ for all $1 \leq i < j \leq n$ and $a_{i,i} = c_{i,i}$. And let $\varphi \in \mathrm{B}(\mathrm{V})$ such that $[\varphi]_{\alpha} = (a_{i,j})_{1\leq i\leq j\leq n}$.

If $\mathrm{char}(\mathbb{F}) \neq 2$, we can choose A symmetric: $a_{i,j} = a_{j,i} = c_{i,j}/2$. Thus, the study of quadratic forms is related to the study of symmetric bilinear forms.

Let us find a linear change of variables $(x_1,\ldots,x_n) \longleftrightarrow (y_1,\ldots,y_n)$ such that

$$q(y_1,\ldots,y_n) = \sum_{i=1}^{n} b_i y_i^2$$

The axis of this next coordinate system is often called a **principal axis system** for q.

To find such coordinate system is equivalent to find a basis of $\mathbb{F}^n$ which is orthogonal w.r.t. symmetric bilinear form.

In the case $\mathbb{F} = \mathbb{R}$, it is interesting that this new basis be orthonormal w.r.t. the usual inner product. This is always possible thanks to the Spectral Theorem. Recall that

Theorem 2.4.8 (Spectral Theorem for Matrices)**.** A matrix $\mathrm{A} \in \mathrm{M}_n(\mathbb{R})$ is orthogonally diagonalizable iff. A is symmetric.

Definition 2.4.10 (Orthogonally Diagonalizable)**.** A matrix $\mathrm{P} \in \mathrm{M}_n(\mathbb{R})$ is **orthogonal** if $\mathrm{P}^t\mathrm{P} = \mathrm{I}$.

A matrix $\mathrm{A} \in \mathrm{M}_n(\mathbb{R})$ is **orthogonally diagonalizable** if there exists $\mathrm{P} \in \mathrm{M}_n(\mathbb{R})$ orthogonal such that $\mathrm{P}^t\mathrm{AP}$ is diagonal.

Theorem 2.4.9. If α and β are basis for V,

$$[\varphi]_{\beta} = ([\mathrm{I}]_{\beta,\alpha})^t[\varphi]_{\alpha}[\mathrm{I}]_{\beta,\alpha}$$

Example 2.4.4. Suppose that $\mathrm{char}(\mathbb{F}) = 0$ and consider the quadratic form

$$q(x_1,x_2,x_3) = 2x_1^2 - 4x_1x_2 + 2x_1x_3 + x_3^2$$

The symmetric bilinear form on $\mathrm{V} = \mathbb{F}^3$ associated with q is the bilinear form φ from the example 2.4.3.

Using the basis γ found there, we have a principal axis system for q. In this coordinate system, $q(y_1, y_2, y_3) = y_1^2 + y_2^2 - 4y_3^2$.

The concept of orthogonal matrix is related to the concept of orthogonal linear operator. In the next section, we study the generalization of this concept to the context of alternating and symmetric bilinear forms. After that section, we study the theory of self-adjoint operators in the same context.

2.5 Orthogonal and Sympletic Transformations

Definition 2.5.1 (Compatibility). Let V and W be vector spaces, $\varphi \in \mathrm{B}(\mathrm{V})$ and $\psi \in \mathrm{B}(\mathrm{W})$. Then $\mathrm{T} \in \mathrm{Hom}(\mathrm{V}, \mathrm{W})$ is said to be **compatible** with the pair (φ, ψ) if

$$\psi(\mathrm{T}(u), \mathrm{T}(v)) = \varphi(u, v), \quad \forall\ u, v \in \mathrm{V}$$

If φ and ψ are inner products, this definition coincides with the definition of orthogonal transformation. Motivated by this, if φ and ψ are symmetric and nondegenerate, we say that T is an **orthogonal linear transformation**. If both are alternating, we say that T is **sympletic**.

Theorem 2.5.1. The following are equivalent.

1. T is compatible with (φ, ψ).
2. For every basis α of V, $[\varphi]_\alpha = [\psi]_{\mathrm{T}(\alpha)}$.
3. There exists a basis α of V such that $[\varphi]_\alpha = [\psi]_{\mathrm{T}(\alpha)}$.

Proof. Similar to the analogous result in elementary linear algebra. □

How can we describe all compatible transformations? Is there any compatible transformation?

Note that this result does not impose any condition on $\psi(w_1, w_2)$ if w_1 or w_2 are not in the range of T. Thus, we can suppose that T is surjective.

In this case, the radical of φ is entirely taken into the radical of ψ:

$$v \in \mathrm{V}^{\perp_\varphi} \implies \mathrm{T}(v) \in \mathrm{W}^{\perp_\psi}$$

To see this, let $w \in \mathrm{W}$, say $w = \mathrm{T}(u)$. Then

$$\psi(\mathrm{T}(v), w) = \varphi(v, u) = 0$$

Therefore, it is sufficient to study the case in which φ is nondegenerate. In this case, $\det([\varphi]_\alpha) \neq 0$ for every finite and linearly independent subset α of V.

Since $[\psi]_{\mathrm{T}(\alpha)} = [\varphi]_\alpha$, we have that $\mathrm{T}(\alpha)$ is linearly independent and hence T is injective.

With these facts, the following result holds.

Theorem 2.5.2. If $\varphi \in \mathrm{B}(\mathrm{V})$ and $\psi \in \mathrm{B}(\mathrm{W})$ with φ nondegenerate, there exists $\mathrm{T} \in \mathrm{Hom}(\mathrm{V}, \mathrm{W})$

surjective and compatible with (φ, ψ) iff. $\dim(V) = \dim(W)$ and there exist bases α of V and β of W such that $[\varphi]_\alpha = [\psi]_\beta$. In this case, T is an isomorphism.

Let us move to the context of linear operators. I.e., here $V = W$ and $\varphi = \psi$ is nondegenerate. Our goal is to describe the set of linear operators on V compatible with (φ, φ) in the case $\varphi \in B_{as}(V)$ is nondegenerate and $\dim(V)$ is finite.

Consider the subset of compatible operators

$$\mathrm{End}^\varphi(V) := \{T \in \mathrm{End}(V) : \varphi(T(u), T(v)) = \varphi(u, v),\ u, v \in V\}$$

It follows from the previous result that every element of $\mathrm{End}^\varphi(V)$ is bijective.

We can easily verify that $\mathrm{End}^\varphi(V)$ is closed under composition, i.e.,

$$T, S \in \mathrm{End}^\varphi(V) \implies T \circ S \in \mathrm{End}^\varphi(V)$$

and

$$T^{-1} \in \mathrm{End}^\varphi(V)$$

Thus, the pair $(\mathrm{End}^\varphi(V), \circ)$ is a group. If φ is symmetric, this is called an **orthogonal group**. If φ is alternating, this is called an **sympletic group**.

Given a basis α of V, $\beta = T(\alpha)$ is also a basis for V. We have that $[\varphi]_\alpha = [\varphi]_\beta$ and it follows that

$$[\varphi]_\alpha = ([I]_{\beta,\alpha})^t [\varphi]_\alpha [I]_{\beta,\alpha}$$

By definition of β, $[I]_{\beta,\alpha} = [T]_{\alpha,\alpha}$. Hence,

$$\det([\varphi]_\alpha) = \det(([T]_{\alpha,\alpha})^t) \det([\varphi]_\alpha) \det([T]_{\alpha,\alpha})$$

Since, $\det([\varphi]_\alpha) \neq 0$, it follows that $\det(T) = \pm 1$.

Definition 2.5.2 (Rotation and Reflection)**.** Suppose that $\varphi \in B_s(V)$ and $T \in \mathrm{End}^\varphi(V)$. T is said to be a **rotation** if $\det(T) = 1$, and is said to be a **reflection** if $\det(T) = -1$.

The subset of rotations forms a subgroup of $\mathrm{End}^\varphi(V)$. The composition of two reflections is a rotation.

Definition 2.5.3 (Simple Reflection)**.** Let $w \in V$ be a non-isotropic vector and $W = [w]$. The function

$$\begin{aligned} R_W^\varphi : V &\longrightarrow V \\ v &\longmapsto v - 2\frac{\varphi(v, w)}{\varphi(w, w)} w \end{aligned}$$

is called a **simple reflection**.

It can be easily verified that R_W^φ is linear and depends only on W.

To show that R_W^φ is a reflection, notice that

$$R_W^\varphi(v) = \begin{cases} -v, & \text{if } v \in W, \\ v, & \text{if } v \in W^{\perp_\varphi} \end{cases}$$

Since w is non-isotropic, we have $V = W \oplus W^{\perp_\varphi}$ and we can choose a basis $\alpha = v_1, \dots, v_n$ of V such that $v_1 \in W$ and $v_j \in W^{\perp_\varphi}$ for $j > 1$. With this choice, we have the following matrix

$$[R_W^\varphi]_{\alpha,\alpha} = \begin{bmatrix} -1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}$$

Thus, $\det(R_W^\varphi) = -1$ and

$$\varphi(R_W^\varphi(v_i), R_W^\varphi(v_j)) = (-1)^{\delta_{i,1}+\delta_{j,1}}\varphi(v_i, v_j) = \varphi(v_i, v_j), \quad \forall\ 1 \le i \le j \le n$$

Hence, $[\varphi]_{T(\alpha)} = [\varphi]_\alpha$ and it follows from the theorem 2.5.1 that R_W^φ is compatible with φ. Also

$$R_W^\varphi \circ R_W^\varphi = \mathrm{Id}_V$$

If φ is an inner product, the set of simple reflections coincides with the set of orthogonal reflections w.r.t. hyperplanes.

Lemma 2.5.3. If $\varphi \in B_{as}(V)$, $T \in \mathrm{End}^\varphi(V)$ and W is a nondegenerate subspace w.r.t. φ and T-invariant, then $W^{\perp_\varphi}$ is also T-invariant.

Lemma 2.5.4. Suppose that $\mathrm{char}(\mathbb{F}) \neq 2$, $\varphi \in B_s(V)$ is nondegenerate, and that $u, v \in V$ satisfy $\varphi(u,u) = \varphi(v,v) \neq 0$. Then there exists a simple reflection T such that $R(v) \in \{u, -u\}$.

Proof. Consider $w_\pm = v \pm u$ and $W_\pm = [w_\pm]$. We show that at least one of the vectors w_+ and w_- is not isotropic. In fact,

$$\varphi(w_\pm, w_\pm) = 2(\varphi(u,u) \pm \varphi(u,v))$$

Thus

$$\varphi(w_+, w_+) = \varphi(w_-, w_-) = 0 \implies \varphi(u,u) = \pm\varphi(u,v)$$

contradicting the hypothesis that $\varphi(u,u) = 0$.

Notice that $\varphi(w_+, w_-) = 0$. If w_+ is non-isotropic, then

$$R_{W_+}^\varphi(w_\pm) = \mp w_\pm$$

and, therefore,

$$R_{W_+}^\varphi(v) = \frac{1}{2}R_{W_+}^\varphi(w_+ + w_-) = \frac{1}{2}(w_- - w_+) = -u$$

Analogously, if w_- is non-isotropic, it follows that $R_{W_-}^\varphi(v) = u$. □

With the language of reflections and these lemmas, it is possible to characterize orthogonal operators.

Theorem 2.5.5. Suppose that $\mathrm{char}(\mathbb{F}) \neq 2$ and $0 \neq \dim(V) < \infty$. Let $\varphi \in B_s(V)$ nondegenerate and $T \in \mathrm{End}(V)$. T is orthogonal iff. T is a composition of simple reflections.

Proof. If $\mathrm{End}^{\varphi}(V)$ is a group, then the composition of simple reflection are orthogonal operators. We'll prove the reciprocal by induction on $n = \dim(V) \geq 1$.

If $n = 1$, then $V = [v]$ for all $v \in V \setminus \{0\}$ and $T(v) = \lambda v$. Since $\det(T) = \pm 1$, it follows that $T = \pm \mathrm{Id}_V$. And since $-\mathrm{Id}_V = R_V^{\varphi}$ and $\mathrm{Id}_V = (R_V^{\varphi})^2$, the case $n = 1$ is proved.

Suppose that $n > 1$ and choose a non-isotropic vector u (which exists since $\mathrm{char}(\mathbb{F}) \neq 2$ and φ is nondegenerate). Take $v = T(u)$ and let R be a simple reflection such that $R(v) = \pm u$, which exists by the lemma 2.5.4.

In particular, $U = [u]$ is $(R \circ T)$-invariant. And since U is nondegenerate, it follows from the lemma 2.5.3 that $U^{\perp_\varphi}$ is also $(R \circ T)$-invariant.

Let S be the linear operator on $U^{\perp_\varphi}$ induced by $R \circ T$. By induction hypothesis, S is a composition of simple reflections, say $S = S_1 \circ \cdots \circ S_m$.

For each $1 \leq j \leq m$, let R_j be the unique linear operator on V satisfying

$$R_j(u) = u \text{ and } R_j(w) = S_j(w), \quad \forall\, w \in U^{\perp_\varphi}$$

Also consider

$$R_0 = \begin{cases} \mathrm{Id}_V, & \text{if } R(v) = u \\ R_U^{\varphi}, & \text{if } R(v) = -u \end{cases} \tag{2.2}$$

To finish the proof, we will verify that

$$T = R \circ R_0 \circ R_1 \circ \cdots \circ R_m \tag{2.3}$$

and that R_j is a simple reflection on V for all $1 \leq j \leq m$.

Start by noticing that $R_0(w) = w$ for all $w \in U^{\perp_\varphi}$ and, therefore, $R \circ T$ coincides with $R_0 \circ \cdots \circ R_m$ on $U^{\perp_\varphi}$.

On the other hand, since $R_j(u) = u$ for all $1 \leq j \leq m$,

$$(R_0 \circ \cdots \circ R_m)(u) = R_0(u) \overset{2.2}{=} R(T(u))$$

Since $V = U \oplus U^{\perp_\varphi}$, it follows that $R \circ T = R_0 \circ \cdots \circ R_m$. With the fact that $R^{-1} = R$, the equation 2.3 is proved.

The last step is to show that each R_j is a simple reflection. Let ψ be the restriction of φ to $U^{\perp_\varphi}$ and, given $1 \leq j \leq m$, let $w_j \in U^{\perp_\varphi}$ such that

$$S_j = R_{W_j}^{\psi}, \quad W_j = [w_j]$$

Let us show that $R_j = R_{W_j}^{\varphi}$. Since $\varphi(u, w_j) = 0$, we have

$$R_{W_j}^{\varphi}(u) = u = R_j(u)$$

On the other hand, if $w \in U^{\perp_\varphi}$, we have

$$R^{\varphi}_{W_j}(w) = w - 2\frac{\varphi(w, w_j)}{\varphi(w_j, w_j)} w_j = w - 2\frac{\psi(w, w_j)}{\psi(w_j, w_j)} w_j = S_j(w) = R_j(w)$$

Since $R^{\varphi}_{W_j}$ coincides with R_j on U and $U^{\perp_\varphi}$, it follows that $R^{\varphi}_{W_j} = R_j$. $\square$

Example 2.5.1. Let $V = \mathbb{R}^3$ and φ be the usual inner product. Take $T = \mathrm{Rot}^w_\theta$ with $\|w\| = 1$ and $\theta \in [0, 2\pi]$.

Choose $w_1 \in [w]^\perp$ with $\|w_1\| = 1$ and define $w_2 = w \times w_1$ such that $\beta = \{w_1, w_2, w\}$ be an orthonormal basis. Thus

$$[T]_{\beta,\beta} = \begin{bmatrix} a & -b & 0 \\ b & a & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \text{with } a = \cos(\theta) \text{ and } \ b = \sin(\theta)$$

If $b = 0$ and $a = 1$, then $T = \mathrm{Id}_V = (R^{\varphi}_W)^2$ for any line W.

Otherwise, choose $u = w_1$ and $v = T(u)$ as in the proof. It follows that

$$w_- = v - u = (a-1)w_1 + bw_2 \neq 0$$

and thus w_- is non-isotropic.

Taking $R = R^{\varphi}_{W_-}$ with $W_- = [w_-]$ and $W = [u]^\perp = [w_2, w]$, it follows from the proof that $R(T(u)) = u$ and that W is $(R \circ T)$-invariant.

Let S_1 be the linear operator induced by $R \circ T$ on W and $\psi = \varphi|_W$. Since T and R fix w, we have $S_1(w) = w$. Since $[w]$ is nondegenerate w.r.t. ψ, $[w]^{\perp_\psi} = [w_2]$ is also S_1-invariant.

Thus, $S_1(w_2) = \lambda w_2$ for some scalar λ. It follows that $\det(S_1) = \lambda$ and, therefore, $\lambda = \pm 1$.

Let $R_1 \in \mathrm{End}(V)$ such that $R_1|_W = S_1$ and $R_1(u) = u$, as in the proof. It follows from the previous argument that $R \circ T = R_1$, i.e., $T = R \circ R_1$.

In fact, $R_1(u) = R(T(u))$ and $(R \circ T)|_W = S_1 = R_1|_W$.

Since $\det(T) = 1$, $\det(R) = -1$ and $\det(R_1) = \lambda$, it follows that $\lambda = -1$ and, therefore, $R_1 = R^{\varphi}_{[w_2]}$.

Notice that these conclusions follow from the fact verified above:

$$R(T(w_1)) = w_1, \quad R(T(w)) = w, \quad \text{and } R(T(w_2)) = -w_2$$

Hence, $\mathrm{Rot}^w_\theta = R^{\varphi}_{W_-} \circ R^{\varphi}_{[w_2]}$.

For alternating forms, instead of simple reflections, we'll have sympletic transvections.

Definition 2.5.4 (Sympletic Transvection)**.** Suppose that $\varphi \in B_a(V)$ is nondegenerate. Given

$w \in V$, $a \in \mathbb{F}$, consider the function

$$\begin{aligned} T_{w,a} : V &\longrightarrow V \\ v &\longmapsto v + a\varphi(v,w)w \end{aligned}$$

The function $T_{w,a}$ is called a **sympletic transvection**.

Clearly, $T_{w,a}$ is linear. Let us verify that it is sympletic.

$$\begin{aligned} \varphi(T_{w,a}(v_1), T_{w,a}(v_2)) &= \varphi(v_1 + a\varphi(v_1,w)w, v_2 + a\varphi(v_2,w)w) \\ &= \varphi(v_1,v_2) + a\varphi(v_1,w)\varphi(v_2,w) + a\varphi(v_1,w)\varphi(w,v_2) \\ &= \varphi(v_1,v_2) \end{aligned}$$

Analogously to the theorem 2.5.5, we have the following result.

Theorem 2.5.6. Suppose that $0 \neq \dim(V) < \infty$ and let $\varphi \in B_a(V)$ nondegenerate and $T \in \operatorname{End}(V)$. Then T is sympletic iff. T is a composition of sympletic transvections.

2.6 Self-Adjoint Operators

Definition 2.6.1 (Adjoint). Let $\varphi \in B(V)$, $\psi \in B(W)$ and $T \in \operatorname{Hom}(V,W)$. A function $S : W \longrightarrow V$ is said to be a **right adjoint** of T w.r.t. (φ,ψ) if

$$\varphi(v, S(w)) = \psi(T(v), w), \quad \forall\, v \in W,\ w \in W$$

And S is called a **left adjoint** of T w.r.t. (φ,ψ) if

$$\varphi(S(w), v) = \psi(w, T(v)), \quad \forall\, v \in W,\ w \in W$$

Notice that if φ and ψ are symmetric and alternating, both conditions are equivalent.

We'll denote by T^{φ}_{ψ} the right adjoint of T w.r.t. (φ,ψ) when it exists and φ is right nondegenerate.

The following lemma states that the adjoint is linear and unique if it exists.

Lemma 2.6.1. Suppose that φ is right nondegenerate.

1. If S is right adjoint of T, then S is linear.
2. If S_1 and S_2 are right adjoint of T, then $S_1 = S_2$.

Proof. Since φ is right nondegenerate, the first item follows if we prove that

$$S(w_1 + \lambda w_2) - S(w_1) - \lambda S(w_2) \in \ker(D_\varphi), \quad \forall\, w_1, w_2 \in W, \lambda \in \mathbb{F}$$

i.e. for any $v \in V$

$$\varphi(v,\ S(w_1 + \lambda w_2) - S(w_1) - \lambda S(w_2)) = 0$$

In fact,

$$\begin{aligned}\varphi(v,\ S(w_1+\lambda w_2)) &= \psi(T(v), w_1+\lambda w_2)\\ &= \psi(T(v), w_1) + \lambda\psi(T(v), w_2)\\ &= \varphi(v, S(w_1)) + \lambda\varphi(v, S(w_2))\\ &= \varphi(v, S(w_1) + \lambda S(w_2))\end{aligned}$$

To show the second item, we'll prove that $S_1(w) - S_2(w) \in \ker(D_\varphi)$ for all $w \in W$:

$$\begin{aligned}\varphi(v,\ S_1(w) - S_2(w)) &= \varphi(v, S_1(w)) - \varphi(v, S_2(w))\\ &= \psi(T(v), w) - \psi(T(v), w) = 0\end{aligned}$$

□

The concept of adjoint gives a new interpretation to transposition.

Lemma 2.6.2. If α is a basis for V and β is a basis for W, then

$$[T_\psi^\varphi]_{\beta,\alpha} = [\varphi]_\alpha^{-1}([T]_{\alpha,\beta})^t[\psi]_\beta$$

When the adjoint exists?

Theorem 2.6.3. If $\dim(V) < \infty$ and φ is nondegenerate, then there exists a right adjoint of T.

Proof. The hypothesis guarantee that D_φ is bijective. Thus, we can consider

$$S = D_\varphi^{-1} \circ T^t \circ D_\psi$$

which gives us the following diagram

$$\begin{tikzcd} W \arrow[r, dashed, "S"] \arrow[d, "D_\psi"'] & V \\ W^* \arrow[r, "T^t"'] & V^* \arrow[u, "D_\varphi^{-1}"'] \end{tikzcd}$$

Let us verify that S is adjoint of T. By definition of D_φ, given $f \in V^*$,

$$u = D_\varphi^{-1}(f) \iff f(v) = \varphi(v,u), \quad \forall\ v \in V$$

I.e.,

$$f(v) = \varphi(v, D_\varphi^{-1}(f)), \quad \forall\ v \in V,\ f \in V^*$$

Thus,

$$\begin{aligned}\varphi(v, S(w)) &= \varphi(v,\ D_\varphi^{-1}(T^t(D_\psi(w)))) = (T^t(D_\psi(w)))(v)\\ &= (D_\psi(w))(T(v)) = \psi(T(v), w)\end{aligned}$$

□

The next result shows some properties of adjoints.

Theorem 2.6.4. Let $\varphi \in B(V)$, $\psi \in B(W)$, $\xi \in B(U)$ and suppose that V and W are finite-dimensional, and that φ and ψ are nondegenerate. The following are true.

1. If $S, T \in \mathrm{Hom}(W, U)$ and $\lambda \in \mathbb{F}$, then $(S + \lambda T)^{\psi}_{\xi} = S^{\psi}_{\xi} + \lambda T^{\psi}_{\xi}$. In words, adjunction is a linear operation.
2. If $T \in \mathrm{Hom}(V, W)$ and $S \in \mathrm{Hom}(W, U)$, then $(S \circ T)^{\varphi}_{\xi} = T^{\varphi}_{\psi} \circ S^{\psi}_{\xi}$. In words, adjunction is contravariant.
3. If $T \in \mathrm{Hom}(V, W)$ is invertible, then $(T^{-1})^{\psi}_{\varphi} = (T^{\varphi}_{\psi})^{-1}$. In words, the adjoint of the inverse is the inverse of the adjoint.
4. If $T \in \mathrm{Hom}(V, W)$, then $(T^{\varphi}_{\psi})^{\psi}_{\varphi} = T$.

Lemma 2.6.5. Suppose that $\dim(V) < \infty$, $\varphi \in B_{as}(V)$ is nondegenerate and $\psi \in B_{as}(W)$.

1. If ψ is nondegenerate, then $\ker(T) = \mathrm{range}(T^{\varphi}_{\psi})^{\perp_\varphi}$. In particular, T is injective iff. T^{φ}_{ψ} is surjective. Moreover, if $\ker(T)$ is nondegenerate, then $V = \ker(T) \oplus \mathrm{range}(T^{\varphi}_{\psi})$.
2. $\ker(T^{\varphi}_{\psi}) = \mathrm{range}(T)^{\perp_\psi}$. If ψ is nondegenerate, then T^{φ}_{ψ} is injective iff. T is surjective. More than that, if $\dim(W) < \infty$ and both W and $\ker(T^{\varphi}_{\psi})$ are nondegenerate, $W = \ker(T^{\varphi}_{\psi}) \oplus \mathrm{range}(T)$.

Proof. If ψ is nondegenerate, $v \in \ker(T)$ iff. $\psi(T(v), w) = 0$ for all $w \in W$.

Using the concept of adjoint, the first affirmation of 1. follows. The second one follows by noticing that, since φ in nondegenerate, for a subspace $U \subseteq V$, then $U^{\perp_\varphi} = \{0\}$ iff. $U = V$.

For the last affirmation, if $\ker(T)$ is nondegenerate, then $V = \ker(T) \oplus \ker(T)^{\perp_\varphi}$, by the theorem 2.4.4. Using the first affirmation and the theorem 2.4.3, the result follows.

The item 2. is analogous. □

Our goal now is to review the spectral theorem in the context of symmetric bilinear forms. Suppose that $\varphi \in B_s(V)$ and $T \in \mathrm{End}(V)$. We denote the adjoint of T w.r.t. φ, φ simply as T^{φ}.

Definition 2.6.2 (Self-Adjoint)**.** T is said to be **self-adjoint** w.r.t. φ if $T^{\varphi} = T$, i.e.

$$\varphi(T(v), w) = \varphi(v, T(w)), \quad \forall\, v, w \in V$$

Definition 2.6.3 (Anisotropic Space)**.** A vector space V is called **anisotropic** if it does not have any isotropic vectors.

Lemma 2.6.6. Suppose that T is self-adjoint w.r.t. φ.

1. The eigenspaces of T are mutually orthogonal.
2. If v is an eigenvector of T, then $\{v\}^{\perp_\varphi}$ is T-invariant.

Proof. Follows, without any circularity, from the lemma 2.6.8. □

Theorem 2.6.7 (Spectral Theorem (First Version))**.** Suppose that $\mathbb{F}$ is algebrically closed, $\mathrm{char}(\mathbb{F}) \neq 2$ and that V is anisotropic w.r.t. φ. There exists an orthogonal basis for V w.r.t. φ formed by eigenvectors of T iff. T is self-adjoint.

Proof. ($\Rightarrow$) Suppose that α is an orthogonal basis of V w.r.t. φ formed by eigenvectors of T such that

$$[\mathrm{T}]_{\alpha,\alpha} = \mathrm{diag}(\lambda_1, \dots, \lambda_n) \quad \text{and} \quad [\varphi]_\alpha = \mathrm{diag}(\mu_1, \dots, \mu_n)$$

Since $[\mathrm{T}^\varphi]_{\alpha,\alpha} = [\varphi]_\alpha^{-1}([\mathrm{T}]_{\alpha,\alpha})^t[\varphi]_\alpha$, it follows that $[\mathrm{T}^\varphi]_{\alpha,\alpha} = [\mathrm{T}]_{\alpha,\alpha}$, i.e., $\mathrm{T}^\varphi = \mathrm{T}$.

($\Leftarrow$) Suppose that T is self-adjoint. Since $\mathbb{F}$ is algebraically closed, there exists an eigenvector w for T. Let $\mathrm{W} := [w]$.

Since V is anisotropic, W is nondegenerate, and thus $\mathrm{V} = \mathrm{W} \oplus \mathrm{W}^{\perp_\varphi}$.

Using that φ is nondegenerate, $\mathrm{W}^{\perp_\varphi}$ is nondegenerate and, by the second item of the lemma 2.6.6, $\mathrm{W}^{\perp_\varphi}$ is T-invariant.

By induction on the dimension of V (which starts trivially for dimension one), there exists a basis α of $\mathrm{W}^{\perp_\varphi}$ which is orthogonal w.r.t. φ and formed by eigenvectors of T.

Hence, $\beta = \alpha \cup \{w\}$ is an orthogonal basis for V w.r.t. φ and formed by eigenvectors of T. □

Notice that if $\mathbb{F} = \mathbb{R}$ and φ is an inner product, the spectral theorem remains valid.

Definition 2.6.4 (Complex Inner Product)**.** If V is a vector space over $\mathbb{C}$, then the function $\varphi : \mathrm{V} \times \mathrm{V} \longrightarrow \mathbb{C}$ is an **inner product** on V if

1. It is linear on the first entry;
2. It is **sesquilinear** on the second entry, i.e., $\varphi(u,v) = \overline{\varphi(v,u)}$, for all $u, v \in \mathrm{V}$;
3. $\varphi(v,v) \in \mathbb{R}_{>0}$, if $v \neq 0$.

Notice that if $\alpha = v_1, \dots, v_n$ is a basis for V and $[\varphi]_\alpha = (\varphi(v_j, v_i))_{(i,j)}$, we have

$$\varphi(u,v) = [v]_\alpha^*[\varphi]_\alpha[u]_\alpha$$

Also $[\varphi]_\alpha = ([\varphi]_\alpha)^*$ and $[\varphi]_\beta = ([\mathrm{I}]_{\beta,\alpha})^*[\varphi]_\alpha[\mathrm{I}]_{\beta,\alpha}$.

Analogously,

$$[\mathrm{T}^\varphi]_{\alpha,\alpha} = [\varphi]_\alpha^{-1}([\mathrm{T}]_{\alpha,\alpha})^*[\varphi]_\alpha$$

If α is an orthonormal basis for V w.r.t. φ formed by eigenvectors of T, we have $[\mathrm{T}]_{\alpha,\alpha} = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$ and $[\varphi]_\alpha = \mathrm{I}$. Hence,

$$[\mathrm{T}]_{\alpha,\alpha}[\mathrm{T}^\varphi]_{\alpha,\alpha} = [\mathrm{T}^\varphi]_{\alpha,\alpha}[\mathrm{T}]_{\alpha,\alpha}$$

Definition 2.6.5 (Normal Operator)**.** T is said to be **normal** if $\mathrm{T} \circ \mathrm{T}^\varphi = \mathrm{T}^\varphi \circ \mathrm{T}$.

Every self-adjoint operator is a normal operator.

Lemma 2.6.8. Suppose that T is normal.

1. If $T(v) = \lambda v$, then $T^{\varphi}(v) = \overline{\lambda} v$.
2. The eigenspaces of T are mutually orthogonal.
3. If v is an eigenvector of T, then $\{v\}^{\perp_{\varphi}}$ is T-invariant.

Proof. Denote $\varphi(v,w) = \langle v,w\rangle$ and $\|v\|^2 = \varphi(v,v)$. We have that

$$\begin{aligned}
\|T^{\varphi}(v) - \overline{\lambda} v\|^2 &= \langle T^{\varphi}(v), T^{\varphi}(v)\rangle - \lambda\langle T^{\varphi}(v), v\rangle - \overline{\lambda}\langle v, T^{\varphi}(v)\rangle + |\lambda|^2\|v\|^2 \\
&= \langle v, (T\circ T^{\varphi})(v)\rangle - \lambda\langle v, T(v)\rangle - \overline{\lambda}\langle T(v), v\rangle + |\lambda|^2\|v\|^2 \\
&= \langle v, (T^{\varphi}\circ T)(v)\rangle - \lambda\langle v, \lambda v\rangle - \overline{\lambda}\langle \lambda v, v\rangle + |\lambda|^2\|v\|^2 \\
&= \langle T(v), T(v)\rangle - |\lambda|^2\langle v, v\rangle \\
&= |\lambda|^2\langle v, v\rangle - |\lambda|^2\langle v, v\rangle = 0
\end{aligned}$$

Suppose that $v \in V_{t-\lambda}, w \in V_{t-\mu}$ and $\lambda \neq \mu$. We show that $v \perp w$.

$$\lambda\langle v,w\rangle = \langle T(v), w\rangle = \langle v, T^{\varphi}(w)\rangle = \langle v, \overline{\mu} w\rangle = \mu\langle v,w\rangle \iff \langle v,w\rangle = 0$$

Finally, if $w \perp v$,

$$\langle T(w), v\rangle = \langle w, T^{\varphi}(v)\rangle = \lambda\langle w, v\rangle = 0$$

which proves the last item. □

Lemma 2.6.9. If T is self-adjoint, then every root of the characteristic polynomial, as an element of $\mathbb{C}[x]$, is real.

Proof. The strategy is to study the operator over $\mathbb{C}$ and conclude that all eigenvalues are real.

We have that

$$c_T(t) = \prod_{j=1}^{n}(t-\lambda_j), \quad \lambda_j \in \mathbb{C}$$

However, λ_j is an eigenvalue of T iff. $\lambda_j \in \mathbb{F}$.

Let α be an orthonormal basis of V. Consider $W = \mathbb{C}^n$ and let $\psi = \langle\cdot,\cdot\rangle$ be usual inner product on W. And also consider the unique linear operator S on W satisfying $[S]_{\beta}^{\beta} = [T]_{\alpha}^{\alpha}$, where β is the standard basis.

Since T is self-adjoint and α is orthonormal, we know that $([T]_{\alpha}^{\alpha})^* = [T]_{\alpha}^{\alpha}$. Then, by definition of S and ψ, it follows that $([S]_{\beta}^{\beta})^* = [S]_{\beta}^{\beta}$. Thus, S is self-adjoint.

Using that $c_T = c_S$, it follows that λ_j is an eigenvalue of S for all $1 \le j \le n$.

Let λ be an eigenvalue of S and $w \in W_{(S-\lambda I_W)} \setminus \{0\}$. Given that S is self-adjoint,

$$\lambda\langle w,w\rangle = \langle S(w), w\rangle = \langle w, S(w)\rangle = \overline{\lambda}\langle w,w\rangle$$

Thus, $\lambda = \overline{\lambda}$, showing that $\lambda \in \mathbb{R}$. □

Theorem 2.6.10 (Spectral Theorem (Second Version))**.** If V is a vector space over $\mathbb{C}$ and φ is an inner product on V, then there exists an orthogonal basis for V w.r.t. φ formed by eigenvectors of T iff. T is normal.

Proof. Analogous to the proof of the theorem 2.6.7 using lemmas 2.6.8 and 2.6.9. □

Corollary 2.6.11. If $\mathbb{F} = \mathbb{R}$, T is normal, and c_T factors in a product of terms with degree one, then there exists an orthogonal basis for V formed by eigenvectors of T.

Notice that if we change the hypothesis of T being normal for T being self-adjoint, then by the lemma 2.6.9 it follows that c_T factors in a product of terms with degree one. Thus the theorem 2.6.7 holds in the case $\mathbb{F} = \mathbb{R}$.

Chapter 3

Tensor Algebra

This chapter is dedicated to the concept of the tensor product, which relates multilinearity with linearity. This theory is applied in differential geometry, the theory of representations of groups, differential equations, and quantum mechanics.

3.1 Category Theory

This first section introduces a valuable tool to generalize and see our results from another perspective.

Definition 3.1.1 (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 $\mathrm{A}, \mathrm{B} \in \mathrm{Ob}(\mathfrak{C})$, a set $\hom_{\mathfrak{C}}(\mathrm{A}, \mathrm{B})$ whose elements are called **morphisms**, **maps** or **arrows** from A to B. They are denoted by

$$f : \mathrm{A} \longrightarrow \mathrm{B} \text{ or } \mathrm{A} \xrightarrow{f} \mathrm{B}$$

   The object $\mathrm{A} = \mathrm{dom}\ (f)$ is called the **domain** of f and $\mathrm{B} = \mathrm{codom}\ (f)$ is called the **codomain** of f.
3. The set of all morphisms of $\mathfrak{C}$ is denoted by Mor $(\mathfrak{C})$ and must satisfy the following property. For every morphism f, there exist uniquely defined objects A, B such that $f \in \hom_{\mathfrak{C}}(\mathrm{A}, \mathrm{B})$. I.e., Mor $(\mathfrak{C})$ is a disjoint union $\bigcup \hom_{\mathfrak{C}}(\mathrm{A}, \mathrm{B})$ for all ordered pairs $\mathrm{A}, \mathrm{B} \in \mathrm{Ob}(\mathfrak{C})$.
4. For $f \in \hom_{\mathfrak{C}}(\mathrm{A}, \mathrm{B})$ and $g \in \hom_{\mathfrak{C}}(\mathrm{B}, \mathrm{C})$ there is a morphism $g \circ f \in \hom_{\mathfrak{C}}(\mathrm{A}, \mathrm{C})$ 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 $\mathrm{A} \in \mathfrak{C}$ there exists a morphism $1_{\mathrm{A}} \in \hom_{\mathfrak{C}}(\mathrm{A}, \mathrm{A})$ called the **identity morphism** for A with the property that if $f \in \hom_{\mathfrak{C}}(\mathrm{A}, \mathrm{B})$ then

$$1_{\mathrm{B}} \circ f = f \text{ and } f \circ 1_{\mathrm{A}} = f$$

Definition 3.1.2 (Isomorphism for Categories). The morphism $f : A \longrightarrow B$ is said to be an **isomorphism** 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 3.1.3 (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 3.1.4 (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 3.1.5 (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

$$\begin{tikzcd}
F(A) \arrow[r, "F(f)"] \arrow[d, "\lambda_A"'] & F(A') \arrow[d, "\lambda_{A'}"] \\
G(A) \arrow[r, "G(f)"'] & G(A')
\end{tikzcd}$$

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

3.2 Quotient Spaces and Universal Properties

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 as

$$u \equiv v \iff u - v \in S$$

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

Now notice that

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

The set

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

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

Example 3.2.1. The solution set of a linear system

$$C = \{x : Ax = b\}$$

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

Definition 3.2.1 (Quotient Space)**.** The set of all cosets (or classes) of S in V, denoted by

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

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

Naturally, we define addition and scalar multiplication as follows

$$\begin{aligned}(u + S) + (v + S) = (u + v) + S \quad &\text{i.e.} \quad [u] + [v] = [u + v] \\ \lambda(u + S) = \lambda u + S \quad &\text{i.e.} \quad \lambda[u] = [\lambda u]\end{aligned}$$

Theorem 3.2.1. The quotient space of V modulo S is a vector space over $\mathbb{F}$ with the operations

$$\lambda(u + S) = \lambda u + S$$

$$(u + S) + (v + S) = (u + v) + S$$

Proof. To prove that the addition is well defined, consider $v_1 \sim v_1'$ and $v_2 \sim v_2'$. Then

$$(v_1 + v_2) - (v_1' + v_2') = (v_1 - v_1') + (v_2 - v_2') \in S$$

In order to show that multiplication is well defined, consider $v \sim v'$. Then

$$(\lambda v) - (\lambda v') = \lambda(v - v') \in S$$

□

Definition 3.2.2 (Natural Projection)**.** If S is a subspace of V, we may define the mapping

$$\begin{aligned} \pi_S : V &\longrightarrow V/S \\ v &\longmapsto [v] \end{aligned}$$

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

Theorem 3.2.2. The canonical projection π_S is a surjective linear mapping with $\ker(\pi_S) = S$.

Proof. Notice that

$$\pi_S(v_1 + \lambda v_2) = [v_1 + \lambda v_2] = [v_1] + \lambda[v_2] = \pi_S(v_1) + \lambda \pi_S(v_2)$$

Since $\pi_S(v) = [0] \iff v \in S$, it follows that $\ker(\pi_S) = S$.

The fact that π_S is surjective is immediate. □

Theorem 3.2.3 (The Correspondence Theorem)**.** Let S be a subspace of V. Then the function that assigns each subspace $S \subseteq T \subseteq V$, 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.

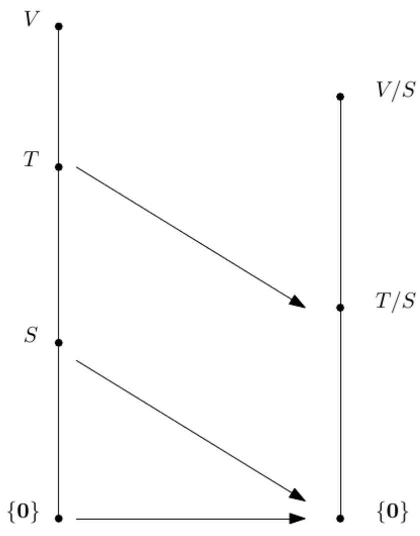

Figure 3.1: Correspondence between V and V/S [Mir20].

Theorem 3.2.4 (Universal Property of Quotient)**.** Let V, W vector spaces over $\mathbb{F}$ and $T \in \mathrm{Hom}(V, W)$. If U is a subspace of V such that $U \subseteq \ker(T)$, then there exists a unique linear mapping

$$S : V/U \longrightarrow W$$

Proof. We need to show that $S([v]) = T(v)$ for all $v \in V$.

Since $U \subseteq \ker(T)$, if $v - v' \in U$, then $T(v - v') = 0$. Thus,

$$T(v') = T(v), \quad \forall\, v \in V,\ v' \in [v]$$

Hence, there exists a unique function

$$\begin{aligned} S : V/U &\longrightarrow W \\ [v] &\longmapsto T(v) \end{aligned}$$

Given that

$$S([v_1] + \lambda[v_2]) = S([v_1 + \lambda v_2]) = T(v_1 + \lambda v_2) = T(v_1) + \lambda T(v_2) = S([v_1]) + \lambda S([v_2])$$

it follows that S is linear. □

The 'universal' here means that for all objects in the category, there is exactly one map from this object to 1, i.e., the **terminal object**.

Theorem 3.2.5 (First Theorem of Isomorphism)**.** If $U = \ker(T)$, then the linear mapping S given by the Universal Property of Quotient induces an isomorphism between V/U and $\mathrm{range}(T)$.

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "\varphi"'] & T(V) \\ V/\ker(T) \arrow[ur, dashed, "\cong"'] & \end{tikzcd}$$

Proof. Consider the linear mapping $R : V/U \longrightarrow \mathrm{range}(T)$ given by $R([v]) = S([v]) = T(v)$ for all $v \in V$. Then R is surjective by definition.

On the other hand,

$$R([v]) = 0 \iff T(v) = 0 \iff v \in U \iff [v] = 0$$

i.e. R is injective. □

It will be necessary to have some language to talk about universal properties. To do that, we introduce the following definition.

Definition 3.2.3 (Universal Pair)**.** Consider two 'functional' properties P_1 and P_2, i.e., properties that a function may satisfy. Given two sets X and U and a function $\varphi : X \longrightarrow U$, we say that the pair (φ, U) is **universal** over X w.r.t. P_1 and P_2 if, for all functions $\psi : X \longrightarrow A$ satisfying P_1, there exists a unique function $\tilde{\psi} : U \longrightarrow A$ satisfying P_2 such that

$$\tilde{\psi} \circ \varphi = \psi$$

We say that $\tilde{\psi}$ is the function **induced** by ψ on U w.r.t. P_2 and represent its existence by the following commutative diagram

$$\begin{tikzcd} X \arrow[r, "\psi"] \arrow[d, "\varphi"'] & A \\ U \arrow[ur, dashed, "\tilde{\psi}"'] & \end{tikzcd}$$

The immediate question is: are universal pairs unique?

Theorem 3.2.6 (Cancellation Law)**.** Suppose that (φ, U) is universal over X w.r.t. P_1 and P_2 and $\psi_1, \psi_2 : U \longrightarrow A$ satisfy P_2. If $\psi_1 \circ \varphi$ satisfies P_1 and coincide with $\psi_2 \circ \varphi$, then $\psi_1 = \psi_2$.

Proof. Let $\psi := \psi_1 \circ \varphi$ satisfying P_1 and consider the following diagram

$$\begin{tikzcd} U \arrow[dr, "\psi_1"'] & X \arrow[l, "\varphi"'] \arrow[r, "\varphi"] \arrow[d, "\psi"] & U \arrow[dl, "\psi_2"] \\ & A & \end{tikzcd}$$

The hypothesis $\psi_2 \circ \varphi = \psi_1 \circ \varphi = \psi$ and the universality of (φ, U), (i.e. $\tilde{\psi}$ is unique) imply $\psi_1 = \psi_2$. □

Definition 3.2.4 (Compatible with compositions)**.** We say that a property P is **compatible with compositions** if $f \circ g$ satisfies P whenever f and g are functions satisfying P.

Lemma 3.2.7. Suppose that

1. The pairs (φ, U) and (ψ, V) are universal over X w.r.t. P_1 and P_2;
2. The functions φ and ψ satisfy P_1;
3. The identity functions Id_U and Id_V satisfy P_2;
4. The property P_2 is compatible with compositions.

Then there exists a unique function $f : U \longrightarrow V$ satisfying P_2 such that $\psi = f \circ \varphi$. More than that, f is bijective.

Proof. Using the universal properties and the hypothesis that φ and ψ satisfy P_1, we know that there exist unique $\tilde{\psi} : U \longrightarrow V$ and $\tilde{\varphi} : V \longrightarrow U$ satisfying P_2 such that

$$\tilde{\psi} \circ \varphi = \psi \quad \text{and} \quad \tilde{\varphi} \circ \psi = \varphi$$

Thus,

$$(\tilde{\varphi} \circ \tilde{\psi}) \circ \varphi = \tilde{\varphi} \circ (\tilde{\psi} \circ \varphi) = \tilde{\varphi} \circ \psi = \varphi = \mathrm{Id}_U \circ \varphi$$

and, by the cancellation law, $\tilde{\varphi} \circ \tilde{\psi} = \mathrm{Id}_U$.

Analogously,
$$(\tilde{\psi} \circ \tilde{\varphi}) \circ \psi = \tilde{\psi} \circ (\tilde{\varphi} \circ \psi) = \tilde{\psi} \circ \varphi = \psi = \mathrm{Id}_V \circ \psi$$
shows that $\tilde{\psi} \circ \tilde{\varphi} = \mathrm{Id}_V$.

Taking $f = \tilde{\psi}$, we finish the proof.

$$\begin{tikzcd}
X \arrow[r, "\psi"] \arrow[d, "\varphi"'] & V \arrow[dl, dashed, shift left, "\tilde{\varphi}"] \\
U \arrow[ur, dashed, shift left, "\tilde{\psi}"] &
\end{tikzcd}$$

□

To end this section, we introduce some usual terminology.

Definition 3.2.5 (Codimension)**.** Let W be a subspace of V. Then the **codimension** of W in V is

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

Definition 3.2.6 (Cokernel and Coimage)**.** Let $T : V \longrightarrow W$ be a linear transformation. Then the **cokernel** of T is the quotient space

$$\mathrm{coker}(T) = W/\mathrm{range}(T)$$

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

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

Theorem 3.2.8. Suppose $\dim V = n$ and W is a subspace of V with $\dim W = k$. Then

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

I.e.,
$$\dim(V) = \dim(W) + \dim(V/W)$$

Proof. Follows from the theorem 3.2.2 and the rank-nullity theorem. □

Corollary 3.2.9. Let V be a finite-dimensional vector space $T : V \longrightarrow V$ a linear transformation. Then
$$\dim(\ker(T)) = \dim(\mathrm{coker}(T))$$

3.3 Tensor Product of Vector Spaces

The idea here is to 'linearize' the tensor product. We will see the tensor product as a universal property, which relates multilinear functions with linear functions.

Definition 3.3.1 (Tensor Product)**.** Consider the functional properties P_1 = 'be k-linear' and P_2 = 'be linear'. Given a family $V_1, \ldots, V_k$ of vector spaces and $\varphi \in \mathrm{Hom}^k(V_1, \ldots, V_k, V)$, a pair (φ, V) is a **tensor product** of this family if it is universal over $X = V_1 \times \cdots \times V_k$ w.r.t. P_1 and P_2.

Put another way, (φ, V) is a tensor product for $V_1, \ldots, V_k$ if for all $\psi \in \mathrm{Hom}^k(V_1, \ldots, V_k, W)$ (W is a vector space), there exists a unique $\tilde{\psi} \in \mathrm{Hom}(V, W)$ such that $\tilde{\psi} \circ \varphi = \psi$:

$$
\begin{tikzcd}
V_1 \times \cdots \times V_k \arrow[r, "\psi"] \arrow[d, "\varphi"'] & W \\
V \arrow[ur, dashed, "\tilde{\psi}"'] &
\end{tikzcd}
$$

Theorem 3.3.1 (Existence of tensor product)**.** For all families of vector spaces $V_1, \ldots, V_k$, there exists a tensor product. Moreover, if (φ, V) is a tensor product for $V_1, \ldots, V_k$, and α_j is a basis for V_j, for all $1 \leq j \leq k$, then $\varphi(\alpha_1 \times \cdots \times \alpha_k)$ is a basis for V.

Proof. Notice that the conditions of the Lemma 3.2.7 are satisfied. Thus, it is sufficient to prove the second claim of the theorem for a specific tensor product. We start the proof by constructing a tensor product for which we can also verify the second claim. The fact that a multilinear function is completely determined by its action on a cartesian product of bases will appear thoroughly.

Let $\alpha = \alpha_1 \times \cdots \alpha_k$ such that $\dim(V) = \#\alpha$, $\iota : \alpha \longrightarrow V$ such that $\iota(\alpha)$ is a basis of V, and $\varphi \in \mathrm{Hom}^k(V_1, \ldots, V_k, V)$ be the unique function satisfying $\varphi|_\alpha = \iota$.

Now it will be sufficient to show that (φ, V) is a tensor product for $V_1, \ldots, V_k$, since the claim that $\varphi(\alpha)$ is a basis for V is immediate from the definitions of V and φ.

Let us show that (φ, V) satisfies the required universal property.

Given $\psi \in \mathrm{Hom}^k(V_1, \ldots, V_k, W)$, and since $\iota(\alpha)$ is a basis for V, there exists a unique $\tilde{\psi} \in \mathrm{Hom}(V, W)$ such that

$$\tilde{\psi}(\iota(v)) = \psi(v), \quad \forall\, v \in \alpha$$

In particular, since φ and ψ are k-linear, it follows that $\tilde{\psi} \circ \varphi = \psi$. More than that, if $\xi \in \mathrm{Hom}(V, W)$ satisfies $\xi \circ \varphi = \psi$, then for all $v \in \alpha$,

$$\xi(\iota(v)) = \xi(\varphi(v)) = \psi(v)$$

and, therefore, $\xi = \tilde{\psi}$. □

In a certain sense, the tensor product is the 'price' of linearization. An alternative proof of the Theorem 3.3.1 is given at the end of this section.

Theorem 3.3.2. If (φ, V) is a tensor product for $V_1, \ldots, V_k$, then for all vector space W, there exists an isomorphism

$$
\begin{aligned}
\Gamma : \mathrm{Hom}^k(V_1, \ldots, V_k, W) &\longrightarrow \mathrm{Hom}(V, W) \\
\psi &\longmapsto \tilde{\psi}
\end{aligned}
$$

Proof. The universality of φ defines Γ uniquely. To show that Γ is surjective, given $\tau \in \mathrm{Hom}(V, W)$, take $\psi = \tau \circ \varphi$, which is k-linear and, therefore, $\tau = \tilde{\psi} = \Gamma(\psi)$.

To prove injectivity, suppose that $\psi, \xi \in \mathrm{Hom}^k(V_1, \ldots, V_k, W)$ satisfy $\tilde{\psi} = \tilde{\xi}$. Then

$$\psi = \tilde{\psi} \circ \varphi = \tilde{\xi} \circ \varphi = \xi$$

Let us show that Γ is linear. Given a scalar λ,

$$\Gamma(\psi) + \lambda\Gamma(\xi) = \tilde{\psi} + \lambda\tilde{\xi}$$

On the other hand, $\Gamma(\psi+\lambda\xi)$ is the only element of $\mathrm{Hom}(V, W)$ such that $\Gamma(\psi+\lambda\xi)\circ\varphi = \psi + \lambda\xi$. Thus, we need to verify that

$$(\tilde{\psi} + \lambda\tilde{\xi}) \circ \varphi = \psi + \lambda\xi$$

In fact, given $v_j \in V_j$, for all $1 \le j \le k$, we have

$$\begin{aligned}(\tilde{\psi} + \lambda\tilde{\xi})(\varphi(v_1, \ldots, v_k)) &= \tilde{\psi}(\varphi(v_1, \ldots, v_k)) + \lambda\tilde{\xi}(\varphi(v_1, \ldots, v_k)) \\ &= \psi(v_1, \ldots, v_k) + \lambda\xi(v_1, \ldots, v_k) \\ &= (\psi + \lambda\xi)(v_1, \ldots, v_k)\end{aligned}$$

Hence, Γ is an isomorphism. □

The isomorphism in the previous theorem is a canonical isomorphism.

For the sake of simplicity, we introduce the following notation.

Definition 3.3.2 (Tensor Notation)**.** We will denote by $V_1 \otimes \cdots \otimes V_k$ the vector space of the universal pair of a tensor product for $V_1, \ldots, V_k$ and the corresponding k-linear function of the pair by $\otimes$.

Given $v_j \in V_j$, for all $1 \le j \le k$, we will use the notation

$$v_1 \otimes \cdots \otimes v_k = \otimes(v_1, \ldots, v_k)$$

When it is necessary to specify the field, we'll use the symbol $\otimes_{\mathbb{F}}$.

Notice that the k-linearity of $\otimes$ can be rewritten as

$$\begin{aligned}&v_1 \otimes \cdots \otimes v_{j-1} \otimes (v_j + \lambda v_j') \otimes v_{j+1} \otimes \cdots \otimes v_k = \\ &\qquad (v_1 \otimes \cdots \otimes v_k) + \lambda(v_1 \otimes \cdots \otimes v_{j-1} \otimes v_j' \otimes v_{j+1} \otimes \cdots \otimes v_k)\end{aligned}$$

for any $v_j, v_j' \in V_j$, $1 \le j \le k$, and $\lambda \in \mathbb{F}$.

Theorem 3.3.3 (Associativity of the Tensor Product)**.** Given $1 \le l < k$, there exists a unique isomorphism of vector spaces

$$\Gamma : V_1 \otimes \cdots \otimes V_k \longrightarrow (V_1 \otimes \cdots \otimes V_l) \otimes (V_{l+1} \otimes \cdots \otimes V_k)$$

satisfying $\Gamma(v_1 \otimes \cdots \otimes v_k) = (v_1 \otimes \cdots \otimes v_l) \otimes (v_{l+1} \otimes \cdots \otimes v_k)$ for any $v_j \in V_j$ and $1 \le j \le k$.

Proof. Let $V = V_1 \otimes \cdots \otimes V_k$ and $W = (V_1 \otimes \cdots \otimes V_l) \otimes (V_{l+1} \otimes \cdots \otimes V_k)$.

Consider the function $\psi : \mathrm{V}_1 \times \cdots \times \mathrm{V}_k \longrightarrow \mathrm{W}$ given by

$$(v_1, \dots, v_k) \longmapsto (v_1 \otimes \cdots \otimes v_l) \otimes (v_{l+1} \otimes \cdots \otimes v_k)$$

for any $v_j \in \mathrm{V}_j$ and $1 \leq j \leq k$, which k-linear.

By the universal property of V, there exists a unique $\Gamma \in \mathrm{Hom}(\mathrm{V}, \mathrm{W})$ satisfying

$$\Gamma(v_1, \dots, v_k) = (v_1 \otimes \cdots \otimes v_l) \otimes (v_{l+1} \otimes \cdots \otimes v_k)$$

for any $v_j \in \mathrm{V}_j$ and $1 \leq j \leq k$.

$$\begin{tikzcd}
\mathrm{V}_1 \times \cdots \times \mathrm{V}_k \arrow[r, "\psi"] \arrow[d, "\otimes"'] & \mathrm{W} \\
\mathrm{V} \arrow[ur, "\Gamma"'] &
\end{tikzcd}$$

By the second part of the Theorem 3.3.1, Γ takes basis into basis and, therefore, is an isomorphism. □

How does the tensor product of subspaces relate to the tensor product of the original spaces?

Theorem 3.3.4. If V'_j is a subspace of V_j, for all $1 \leq j \leq k$, there exists a unique linear transformation

$$\begin{aligned} \Gamma : \mathrm{V}'_1 \otimes \cdots \otimes \mathrm{V}'_k &\longrightarrow \mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k \\ v_1 \otimes \cdots \otimes v_k &\longmapsto v_1 \otimes \cdots \otimes v_k \end{aligned}$$

for any $v_j \in \mathrm{V}'_j$ and $1 \leq j \leq k$. Moreover, Γ is injective.

Using identity functions, we may interpret $\mathrm{V}'_1 \otimes \cdots \otimes \mathrm{V}'_k$ as a subspace of $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$.

As a consequence of the theorem 3.3.4, $v_1 \otimes \cdots \otimes v_k = 0$ iff. $v_j = 0$ for some $1 \leq j \leq k$.

Definition 3.3.3 (Pure Tensors)**.** Vectors of the form $v_1 \otimes \cdots \otimes v_k$ are called **homogeneous vectors** or **pure tensors**.

By the second part of the Theorem 3.3.1, the homogeneous vectors span $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$. More than that, since a scalar multiple of a pure tensor is again a pure tensor, every vector of $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$ can be represented as a sum of homogeneous vectors:

$$v = \sum_{i=1}^{m} v_{i,1} \otimes \cdots \otimes v_{i,k} \tag{3.1}$$

for some choice of $m \in \mathbb{Z}_{\geq 0}$ and $v_{i,j} \in \mathrm{V}_j \setminus \{0\}$.

Definition 3.3.4 (Tensor Rank)**.** The minimal amount of parts in (3.1) is called the **rank** of v and is denoted by $\mathrm{rank}(v)$.

The rank of the null vector is zero and the rank of nonzero homogeneous vectors is one.

Example 3.3.1 (Impure tensor). Let $V = \mathbb{F}^2$ and let us show that the rank of the following vector of $V^{\otimes 3}$ is two:

$$v = e_1 \otimes e_1 \otimes e_1 + e_2 \otimes e_1 \otimes e_2$$

Notice that the parts on the definition of v form a linearly independent family on $V^{\otimes 3}$ and, therefore, $0 < \operatorname{rank}(v) \leq 2$.

If $\operatorname{rank}(v) = 1$, there must exist a family v_1, v_2, v_3 such that $v = v_1 \otimes v_2 \otimes v_3$. Each v_i would be of the form

$$v_i = a_{i,1} e_1 + e_{i,2} e_2$$

Using the 3-linearity of $\otimes$, we can write $v_1 \otimes v_2 \otimes v_3$ in the basis $(e_i \otimes e_j \otimes e_k)$, where $1 \leq i, j, k \leq 2$. This gives us a nonlinear system without solution.

Hence, $\operatorname{rank}(v) = 2$.

Lemma 3.3.5. If $u = v_1 \otimes w_1 + \cdots + v_m \otimes w_m \in V \otimes W$ has rank m, then $(v_i)_i$ and $(w_i)_i$ are linearly independent. Hence,

$$\operatorname{rank}(u) \leq \min\{\dim(V), \dim(W)\}, \quad \forall\ u \in V \otimes W$$

Proof. Suppose that

$$w_m = a_1 w_1 + \cdots + a_{m-1} w_{m-1}, \quad a_j \in \mathbb{F},\ 1 \leq j < m$$

It follows that

$$v_m \otimes w_m = \sum_{j=1}^{m-1} (a_j v_m) \otimes w_j$$

and, therefore,

$$u = \sum_{j=1}^{m-1} (v_j + a_j v_m) \otimes w_j$$

contradicting the fact that $\operatorname{rank}(u) = m$. □

Lemma 3.3.6. Let $\alpha = v_1, \ldots, v_m$ and $\alpha' = v'_1, \ldots, v'_p$ be families on V, and $\beta = w_1, \ldots, w_m$ and $\beta' = w'_1, \ldots, w'_p$ be families on W such that

$$\sum_{j=1}^{m} v_j \otimes w_j = \sum_{i=1}^{p} v'_i \otimes w'_i$$

If α, α' and β' are linearly independent, then $[\beta] \subseteq [\beta']$. In particular, if $p = 0$, then $w_j = 0$ for all $1 \leq j \leq m$.

Proof. We proceed by induction on p. For $p = 0$, we use induction on $m \geq 1$, which starts when $m = 1$ since $v_1 \otimes \cdots \otimes v_k = 0$ iff. $v_j = 0$ for some $1 \leq j \leq k$ (see Theorem 3.3.4).

If $w_1, \ldots, w_m$ were linearly independent, then each part $v_j \otimes w_j$ would be a part of a basis for $V \otimes W$, contradicting $v_1 \otimes w_1 + \cdots + v_m \otimes w_m = 0$.

Thus, we can suppose w.l.o.g. that $w_m = a_1 w_1 + \cdots + a_{m-1} w_{m-1}$ and, using the lemma

3.3.5, it follows that

$$\sum_{j=1}^{m-1}(v_j + a_j v_m) \otimes w_j = 0$$

Notice that $(v_j + a_j v_m)_{1\le j<m}$ is linearly independent. Using the induction hypothesis, $w_j = 0$ for $1 \le j < m$. Then $v_m \otimes w_m = 0$ and, by the case $m = 1$, $w_m = 0$.

For $p > 0$, if $\alpha \cup \alpha'$ is linearly independent, the case $p = 0$ implies that $w_j = w_i' = 0$, contradicting the fact that β' is linearly independent. We can rewrite

$$\sum_{j=1}^{m} v_j \otimes w_j + \sum_{i=1}^{p} v_i' \otimes (-w_i') = 0$$

Therefore, $\alpha \cup \alpha'$ is linearly dependent. We can suppose, w.l.o.g. that

$$v_p' = a_1 v_1 + \cdots + a_m v_m + a_1' v_1' + \cdots + a_{p-1}' v_{p-1}'$$

Using calculations similar to the lemma 3.3.5, we obtain

$$\sum_{j=1}^{m} v_j \otimes (w_j - a_j w_p') = \sum_{i=1}^{p-1} v_i' \otimes (w_i' + a_i' w_p')$$

Since the family $\beta'' = (w_i' + a_i' w_p')_{1\le i<p}$ is linearly independent, by the induction hypothesis on p, we have that $w_j - a_j w_p' \in [\beta'']$ for all $1 \le j \le m$.

Since $[\beta''] \subseteq [\beta']$, it follows that $w_j \in [\beta']$ for all $1 \le j \le m$. □

Theorem 3.3.7. If $\alpha = v_1, \ldots, v_m$ is a linearly independent family on V and $\beta = w_1, \ldots, w_m$ is a linearly independent family on W, then the rank of $u = v_1 \otimes w_1 + \cdots + v_m \otimes w_m$ is m.

Proof. Let $p = \operatorname{rank}(u)$. In particular, $m \ge p$.

Let $\alpha' = v_1', \ldots, v_p'$ and $\beta' = w_1', \ldots, w_p'$ such that

$$u = v_1' \otimes w_1' + \cdots + v_p' \otimes w_p'$$

Notice that α' and β' are linearly independent by the lemma 3.3.5. Thus, it follows from the lemma 3.3.6 that $[\beta] \subseteq [\beta']$. Since β is linearly independent, we have that $m \le p$. □

Example 3.3.2. If $V = W = \mathbb{Q}^2$, $\operatorname{rank}(5e_1 \otimes e_1 - 3e_2 \otimes e_2) = 2$.

Let us compute the rank of

$$u = (e_1 + e_2) \otimes (e_1 - e_2) + (e_1 + 2e_2) \otimes e_2 + (e_1 - e_2) \otimes (e_1 + e_2)$$

Since $(e_1 + e_2) = (e_1 - e_2) + 2e_2$,

$$\begin{aligned} u &= (e_1 + e_2 + e_1 - e_2) \otimes (e_1 - e_2) + (e_1 + 2e_2 + 2(e_1 - e_2)) \otimes e_2 \\ &= (2e_1) \otimes (e_1 - e_2) + (3e_1) \otimes e_2 = e_1 \otimes (2(e_1 - e_2) + e_2) \\ &= e_1 \otimes (2e_1 + e_2) \end{aligned}$$

Hence, $\operatorname{rank}(u) = 1$.

The next result shows a 'translation' between the language of linear transformations and the language of tensors.

Theorem 3.3.8. There exists a unique linear mapping $\Gamma : \mathrm{V}^* \otimes \mathrm{W} \longrightarrow \mathrm{Hom}(\mathrm{V}, \mathrm{W})$ satisfying

$$\Gamma(f \otimes w)(v) = f(v)w, \quad \forall\ v \in \mathrm{V},\ w \in \mathrm{W},\ f \in \mathrm{V}^*$$

Furthermore,

1. Γ is injective;
2. For all $u \in \mathrm{V}^* \otimes \mathrm{W}$, $\operatorname{rank}(\Gamma(u)) = \operatorname{rank}(u)$;
3. $\mathrm{T} \in \operatorname{range}(\Gamma)$ iff. $\operatorname{rank}(\mathrm{T})$ is finite;
4. Γ is surjective iff. $\dim(\mathrm{V})$ or $\dim(\mathrm{W})$ is finite.

Proof. Notice that the function $\psi : \mathrm{V}^* \otimes \mathrm{W} \longrightarrow \mathrm{Hom}(\mathrm{V}, \mathrm{W})$ given by $\psi(f, w)(v) = f(v)w$ is well defined. In fact,

$$\begin{aligned} \psi(f, w)(v_1 + \lambda v_2) &= f(v_1 + \lambda v_2)w \\ &= (f(v_1)w) + \lambda(f(v_2)w) \\ &= \psi(f, w)(v_1) + \lambda\psi(f, w)(v_2) \end{aligned}$$

It can be easily verified that ψ is bilinear. Thus, the existence of $\Gamma = \tilde{\psi}$ follows from the universal property of $\mathrm{V}^* \otimes \mathrm{W}$.

Since $\operatorname{rank}(\mathrm{T}) \leq \min\{\dim(\mathrm{V}), \dim(\mathrm{W})\}$, the third item implies the fourth.

To show the first item, let $u \in \ker(\Gamma)$ and write

$$u = \sum_{j=1}^{m} f_j \otimes w_j, \quad \text{with } (f_j),\ (w_j) \text{ l.i.} \tag{3.2}$$

If $m > 0$, since $f_1 \neq 0$, then there exists $v \in \mathrm{V}$ such that $f_1(v) \neq 0$ and thus

$$0 = \Gamma(u)(v) = \sum_{j=1}^{m} f_j(v) w_j \tag{3.3}$$

contradicts the fact that (w_j) is linearly independent. Hence, $m = 0$ and Γ is injective.

To prove the second item, using an expression like (3.2), let $W' = [w_1, \dots, w_m]$. Using the Theorem 3.3.7, we have that $\dim(W') = \operatorname{rank}(u)$. Thus it suffices to show that $\operatorname{range}(\Gamma(u)) = W'$.

Looking at the equation (3.3) it is immediate that $\operatorname{range}(\Gamma(u)) \subseteq W'$ and we need to prove that $w_j \in \operatorname{range}(\Gamma(u))$ for all $1 \le j \le m$. Since, for all $1 \le j \le m$, there exists $v \in V$ such that $f_i(v) = \delta_{i,j}$, for all $1 \le i \le m$, we have that $\Gamma(u)(v) = w_j$.

Now we prove the third item. Let $F = \{T \in \operatorname{Hom}(V, W) : \operatorname{rank}(T) < \infty\}$. The second item implies that every element in the range has a finite rank, since the tensor rank is always finite, i.e., $\operatorname{range}(\Gamma) \subseteq F$.

Reciprocally, if $T \in F$ with $\operatorname{rank}(T) = m$, choose a basis $\{w_1, \dots, w_m\}$ for $\operatorname{range}(T)$. Given a basis $\alpha = (v_i)_{i \in I}$ for V, we have

$$T(v_i) = \sum_{j=1}^{m} a_{i,j} w_j, \quad a_{i,j} \in \mathbb{F}$$

For each $1 \le j \le m$ let f_j be the only element of V^* satisfying $f_j(v_i) = a_{i,j}$ for all $i \in I$. It follows that

$$\Gamma\left(\sum_{j=1}^{m} f_j \otimes w_j\right)(v_i) = \sum_{j=1}^{m} f_j(v_i) \otimes w_j = \sum_{j=1}^{m} a_{i,j} \otimes w_j = T(v_i)$$

□

This result motivates a simple way to compute the tensor rank using matrices.

Let $\alpha = v_1, \dots, v_m$ be a basis for V and $\beta = w_1, \dots, w_m$ be a basis for W. And also let $\alpha^* = \{f_1, \dots, f_n\}$ be the dual basis of α and Γ be the function in the Theorem 3.3.8. Then

$$\Gamma(f_j \otimes w_i)(v_k) = \delta_{k,j} w_i$$

and, therefore,

$$[\Gamma(f_j \otimes w_i)]_{\beta}^{\alpha} = E_{i,j}$$

Hence, if $A = (a_{i,j}) \in M_{m,n}(\mathbb{F})$ and $T \in \operatorname{Hom}(V, W)$ is given by $[T]_{\beta}^{\alpha} = A$, then

$$T = \Gamma\left(\sum_{i=1}^{m}\sum_{j=1}^{n} a_{i,j} f_j \otimes w_i\right) \tag{3.4}$$

From this fact, the rank of each element of $V^* \otimes W$ coincides with the rank of the corresponding matrix w.r.t. the basis α^* and β.

We also obtain that to compute the rank of a tensor, it suffices to write the tensor in a given basis and row-reduce the matrix of the coordinates.

How can we interpret the trace in the context of tensors?

Theorem 3.3.9 (Tensor Trace)**.** There exists a unique $\tau \in (V^* \otimes V)^*$ satisfying

$$\tau(f \otimes v) = f(v), \quad \forall f \in V^*, v \in V$$

Additionally, if $\dim(V) < \infty$ and $\Gamma : V^* \otimes V \longrightarrow \mathrm{End}(V)$ as in the Theorem 3.3.8, then $\tau(\Gamma^{-1}(T)) = \mathrm{tr}(T)$.

Proof. The existence and uniqueness part follows from the universal property of the tensor product.

To justify $\tau(\Gamma^{-1}(T)) = \mathrm{tr}(T)$, suppose that $\alpha = \{v_1, \ldots, v_n\}$ is a basis for V and $\alpha^* = \{f_1, \ldots, f_n\}$ is the dual basis of α.

Using (3.4), if $[T]_\alpha^\alpha$, then

$$\Gamma^{-1}(T) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{i,j}\, f_j \otimes v_i$$

Thus,

$$\tau(\Gamma^{-1}(T)) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{i,j}\ \tau(f_j \otimes v_i) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{i,j}\, f_j(v_i) = \sum_{i=1}^{n} a_{i,i}$$

□

We finish this section with an alternative proof for the existence of tensor product (Theorem 3.3.1).

Proof. Let $X = V_1 \times \cdots \times V_k$ and (α, L) be a universal pair over X w.r.t. properties P_1 = 'is a vector space over $\mathbb{F}$' and P_2 = 'is linear', i.e., $\alpha : X \longrightarrow L$ indexes a basis for L.

Since α is a basis, it is injective and we can identify X with its range on L. Thus the element $(v_1, \ldots, v_k)$ will be interpreted as a vector on L, where each $v_j \in V_j$, i.e., we're looking to the image of $(v_1, \ldots, v_k)$ by α.

Now consider the subspace $K \subseteq L$ spanned by all vectors of the form

$$\begin{gathered}(v_1, \ldots, v_{j-1}, \lambda v_j, v_{j+1}, \ldots, v_k) - \lambda(v_1, \ldots, v_k), \quad \text{and} \\ (v_1, \ldots, v_{j-1}, v_j + v_j', v_{j+1}, \ldots, v_k) - (v_1, \ldots, v_k) - (v_1, \ldots, v_{j-1}, v_j', v_{j+1}, \ldots, v_k)\end{gathered} \tag{3.5}$$

with $v_j, v_j' \in V_j$ and $\lambda \in \mathbb{F}$. Define

$$V = L/K \quad \text{and} \quad \varphi = \pi \circ \alpha$$

where $\pi : L \longrightarrow L/K$ is the canonical projection.

Let us show that (φ, V) is a tensor product for $V_1, \ldots, V_k$. By the equation (3.5), we have that

$$\begin{gathered}\varphi(v_1, \ldots, v_{j-1}, \lambda v_j, v_{j+1}, \ldots, v_k) = \lambda \varphi(v_1, \ldots, v_k), \quad \text{and} \\ \varphi(v_1, \ldots, v_{j-1}, v_j + v_j', v_{j+1}, \ldots, v_k) = \varphi(v_1, \ldots, v_k) + \varphi(v_1, \ldots, v_{j-1}, v_j', v_{j+1}, \ldots, v_k)\end{gathered}$$

i.e., φ is k-linear. Now we need to show that (φ, V) satisfies the desired universal property.

Given $\psi \in \mathrm{Hom}^k(V_1, \ldots, V_k, W)$, since α is a basis for L, there exists a unique $T \in \mathrm{Hom}(L, W)$ such that

$$T(v_1, \ldots, v_k) = \psi(v_1, \ldots, v_k), \quad \forall\ v_j \in V_j,\ 1 \leq j \leq k$$

Since ψ is k-linear, the generating vectors of K, described in (3.5), are in the nullspace of T, i.e., $K \subseteq \ker(T)$. By the Universal Property of the Quotient, there exists a unique $\tilde{\psi} \in \operatorname{Hom}(V, W)$ satisfying

$$\tilde{\psi}(\pi(u)) = T(u), \quad \forall\, u \in L$$

In particular, $\tilde{\psi} \circ \varphi = \psi$.

What remains to be proved now is the uniqueness of $\tilde{\psi}$, i.e., that if $\xi \in \operatorname{Hom}(V, W)$ satisfies $\xi \circ \varphi = \psi$, then $\xi = \tilde{\psi}$. Notice that the properties

$$\tilde{\psi} \circ \varphi = \psi \quad \text{and} \quad \xi \circ \varphi = \psi$$

imply that $\tilde{\psi}$ and ξ coincide in range(φ). However, range(φ) spans V, since α spans L and π is surjective. Hence, $\xi = \tilde{\psi}$, completing the proof that (φ, V) is a tensor product. □

3.4 Tensor Product of Linear Mappings

In this section we study the tensor product of linear transformations, which can be interpreted as linear transformation.

Given $T_j \in \operatorname{Hom}(V_j, W_j)$, for $1 \le j \le k$, consider the k-linear function

$$\begin{aligned} \psi : V_1 \times \cdots \times V_k &\longrightarrow W_1 \otimes \cdots \otimes W_k \\ (v_1, \ldots, v_k) &\longmapsto T_1(v_1) \otimes \cdots \otimes T_k(v_k) \end{aligned}$$

Notice that this function is a generalization of the evaluation function. Now consider the induced linear mapping

$$\begin{aligned} \tilde{\psi} : V_1 \otimes \cdots \otimes V_k &\longrightarrow W_1 \otimes \cdots \otimes W_k \\ v_1 \otimes \cdots \otimes v_k &\longmapsto T_1(v_1) \otimes \cdots \otimes T_k(v_k) \end{aligned}$$

Thus, we can define a function

$$\begin{aligned} \varphi : \operatorname{Hom}(V_1, W_1) \times \cdots \times \operatorname{Hom}(V_k, W_k) &\longrightarrow \operatorname{Hom}(V_1 \otimes \cdots \otimes V_k,\ W_1 \otimes \cdots \otimes W_k) \\ (T_1, \ldots, T_k)(v_1 \otimes \cdots \otimes v_k) &\longmapsto T_1(v_1) \otimes \cdots \otimes T_k(v_k) \end{aligned}$$

Note that φ is k-linear and we have the induced linear transformation

$$\begin{aligned} \tilde{\varphi} : \operatorname{Hom}(V_1, W_1) \otimes \cdots \otimes \operatorname{Hom}(V_k, W_k) &\longrightarrow \operatorname{Hom}(V_1 \otimes \cdots \otimes V_k,\ W_1 \otimes \cdots \otimes W_k) \\ T_1 \otimes \cdots \otimes T_k &\longmapsto \varphi(T_1, \ldots, T_k) \end{aligned}$$

Definition 3.4.1 (Tensor Product of Linear Transformations)**.** The linear transformation

$$\tilde{\psi} = \varphi(T_1, \ldots, T_k)$$

will be denoted by $T_1 \otimes \cdots \otimes T_k$ and will be called the **tensor product** of the family $T_1, \ldots, T_k$.

Lemma 3.4.1. The following are true.

1. $\mathrm{range}(T_1 \otimes \cdots \otimes T_k) = \mathrm{range}(T_1) \otimes \cdots \otimes \mathrm{range}(T_k)$.
2. Let $N_j = V_j \otimes \cdots \otimes V_{j-1} \otimes \ker(T_j) \otimes V_{j+1} \otimes \cdots \otimes V_k$, for $1 \le j \le k$. Then $\ker(T_1 \otimes \cdots \otimes T_k) = N_1 + \cdots + N_k$. In particular, $T_1 \otimes \cdots \otimes T_k$ is injective if T_j is injective for all $1 \le j \le k$.

Proof. (1.) **Exercise.**

(2.) The second claim is immediate from the first since $N_j = \{0\}$.

The fact that $N_j \subseteq \ker(T_1 \otimes \cdots \otimes T_k)$ is evident e, and therefore, if $N := N_1 + \cdots + N_k$, we have $N \subseteq \ker(T_1 \otimes \cdots \otimes T_k)$.

Let $V + V_1 \otimes \cdots \otimes V_k$ and $W = \mathrm{range}(T_1) \otimes \cdots \otimes \mathrm{range}(T_k) = \mathrm{range}(T_1 \otimes \cdots \otimes T_k)$. Consider the canonical projection $\pi : V \longrightarrow V/N$. We'll show that there exists $S \in \mathrm{Hom}(W, V/N)$ such that

$$S \circ (T_1 \otimes \cdots \otimes T_k) = \pi \tag{3.6}$$

If this is the case,

$$\ker(T_1 \otimes \cdots \otimes T_k) \subseteq \ker(S \circ (T_1 \otimes \cdots \otimes T_k)) \overset{(3.6)}{=} N$$

and the lemma follows.

We need to show how to define S satisfying (3.6). For each $1 \le j \le k$, let σ_j be a right inverse for T_j, i.e., $\sigma_j : \mathrm{range}(T_j) \longrightarrow V_j$ satisfies

$$T_j(\sigma_j(w)) = w, \quad \forall\, w \in \mathrm{range}(T_j)$$

Since $T_j(\sigma_j(T_j(v)) - v) = T_j(\sigma_j(T_j(v))) - T_j(v) = 0$, this implies that

$$\sigma_j(T_j(v)) - v \in \ker(T_j), \quad \forall\, v \in V_j \tag{3.7}$$

Define

$$\begin{aligned} \mathrm{range}(T_1) \times \cdots \times \mathrm{range}(T_k) &\longrightarrow V/N \\ (w_1, \ldots, w_k) &\longmapsto \pi(\sigma_1(w_1) \otimes \cdots \otimes \sigma_k(w_k)) \end{aligned}$$

We now show that σ is k-linear and that $S := \tilde{\sigma}$ satisfies (3.6).

Given $v_j \in V_j$, (3.7) implies that there exists $v_j \in \ker(T_j)$ such that $\sigma_j(T_j(v_j)) = v_j + v_j'$. Then

$$\begin{aligned} \tilde{\sigma}((T_1 \otimes \cdots \otimes T_k)(v_1 \otimes \cdots \otimes v_k)) &= \tilde{\sigma}(T_1(v_1) \otimes \cdots \otimes T_k(v_k)) \\ &= \sigma(T_1(v_1), \ldots, T_k(v_k)) \\ &= \pi(\sigma_1(T_1(v_1)) \otimes \cdots \otimes \sigma_k(T_k(v_k))) \\ &= \pi((v_1 + v_1') \otimes \cdots \otimes (v_k + v_k')) \\ &= \pi(v_1 \otimes \cdots \otimes v_k) + \pi(v_1' \otimes \cdots \otimes v_k') \end{aligned}$$

where $v_1' \otimes \cdots \otimes v_k'$ is a sum of elements on N, proving that $\tilde{\sigma}$ satisfies (3.6).

Recall that for every surjective linear transformation, there exists a right inverse which is also linear. Using this fact, we prove that σ is k-linear. $\square$

Theorem 3.4.2. The linear mapping $\tilde{\varphi}$ is injective. If V_j and W_j are finite-dimensional for all $1 \leq j \leq k$, then $\tilde{\varphi}$ is an isomorphism.

Proof. The second claim follows form the first one, since domain and codomain have the same finite dimension.

We proceed by induction on $k \geq 2$. For $k = 2$, let $\Gamma \in \ker(\tilde{\varphi})$ and choose an expression for Γ of the form

$$\Gamma = \sum_{j=1}^{m} T_j \otimes S_j$$

where $T_j \in \mathrm{Hom}(V_1, W_1)$ and $S_j \in \mathrm{Hom}(V_2, W_2)$ such that $T_1, \dots, T_m$ and $S_1, \dots, S_m$ are linearly independent families. In particular,

$$\tilde{\varphi}(\Gamma)(v, v') = \sum_{j=1}^{m} T_j(v) \otimes S_j(v') = 0, \quad \forall\ v \in V_1,\ v' \in V_2 \tag{3.8}$$

Our goal is to show that $m = 0$. If $m \neq 0$, then $T_1 \neq 0$ and we can choose $v \in V_1$ such that $T_1(v) \neq 0$.

Let r be the dimension of the subspace of W_1 spanned by $T_1(v), \dots, T_m(v)$ and suppose, w.l.o.g., that $T_1(v), \dots, T_r(v)$ is linearly independent.

Thus, for each $r < l \leq m$, there exist scalars $a_{i,l}$, where $1 \leq i \leq r$, such that

$$T_l(v) = \sum_{i=1}^{r} a_{i,l} T_i(v)$$

Applying these expressions in (3.8),

$$\sum_{j=1}^{r} T_j(v) \otimes \left(S_j(v') + \sum_{l=r+1}^{m} a_{j,l} S_l(v') \right) = 0$$

Since $T_1(v), \dots, T_r(v)$ is linearly independent, it follows from the Lemma 3.3.6 that

$$S_j + \sum_{l=r+1}^{m} a_{j,l} S_l = 0, \quad \forall\ 1 \leq j \leq r$$

contradicting the fact that $S_1, \dots, S_m$ is linearly independent. Hence, $m = 0$ and $\Gamma = 0$, completing the proof for $k = 2$.

Suppose that $k > 2$ and define

$$V = V_1 \otimes \cdots \otimes V_{k-1}, \quad W + W_1 \otimes \cdots \otimes W_{k-1}$$

and

$$\tilde{\varphi}_{k-1} : \mathrm{Hom}(V_1, W_1) \otimes \cdots \otimes \mathrm{Hom}(V_{k-1}, W_{k-1}) \longrightarrow \mathrm{Hom}(V, W)$$

which is injective by induction hypothesis.

Now consider

$$\tilde{\varphi}' : \mathrm{Hom}(V, W) \otimes \mathrm{Hom}(V_k, W_k) \longrightarrow \mathrm{Hom}(V \otimes V_k, W \otimes W_k)$$

which is injective by the case $k = 2$.

The associativity of the tensor product induces the following isomorphisms:

$$\begin{aligned}\mathrm{Hom}(V_1, W_1) \otimes \cdots \otimes \mathrm{Hom}(V_k, W_k) &\xrightarrow{\psi} \\ &(\mathrm{Hom}(V_1, W_1) \otimes \cdots \otimes \mathrm{Hom}(V_{k-1}, W_{k-1})) \otimes \mathrm{Hom}(V_k, W_k), \\ T_1 \otimes \cdots \otimes T_k &\longmapsto \mathrm{Hom}(T_1 \otimes \cdots \otimes T_{k-1}) \otimes T_k\end{aligned}$$

and

$$\mathrm{Hom}(V \otimes V_k, W \otimes W_k) \xrightarrow{\tilde{\psi}} \mathrm{Hom}(V_1 \otimes \cdots \otimes V_k, W_1 \otimes \cdots \otimes W_k)$$

It can be easily verified that $\tilde{\varphi} = \psi' \circ \tilde{\varphi}' \circ (\tilde{\varphi}_{k-1} \otimes \mathrm{Id}) \circ \psi$. Since ψ', $\tilde{\varphi}'$, and ψ are injective, and, by the Lemma 3.4.1, $\tilde{\varphi}_{k-1} \otimes \mathrm{Id}$ is also injective, we have that $\tilde{\varphi}$ is injective. □

3.5 Exterior and Symmetric Powers

Consider the set of pairs (φ, U) with $\varphi \in \mathrm{Hom}^k(V_1, \ldots, V_k, U)$ satisfying that for all $\psi \in \mathrm{Hom}^k(V_1, \ldots, V_k, W)$ there exists $\tilde{\psi} \in \mathrm{Hom}(U, W)$ such that $\tilde{\psi} \circ \varphi = \psi$.

In fact, (φ, U) is a tensor product for $V_1, \ldots, V_k$ iff. $\dim(U)$ is minimal. Intuitively, the tensor product produces the smallest vector space in which the multilinear problem is linearized.

Suppose that $V_j = V$ for all $1 \leq j \leq k$ and define

$$T^k(V) = V^{\otimes k} = \underbrace{V \otimes \cdots \otimes V}_{k \text{ times}}$$

The subset of $\mathrm{Hom}^k(V_1, \ldots, V_k, W)$ formed by the k-linear symmetric or alternating functions is a subspace. We denote these spaces by $S^k(V, W)$ and $A^k(V, W)$ respectively.

If $\psi \in S^k(V, W)$, then there exists a unique $\tilde{\psi} \in \mathrm{Hom}(V^{\otimes k}, W)$ that 'linearizes' ψ. However, $\dim(V^{\otimes k})$ is not minimal in the vector spaces U satisfying the following property: there exists $\varphi \in S^k(V, U)$ such that, for all $\psi \in S^k(V, W)$, there exists $\tilde{\psi} \in \mathrm{Hom}(U, W)$ satisfying $\psi = \tilde{\psi} \circ \varphi$. A vector space U with the minimal dimension will be called a k-th **symmetric power** of V. Using $A^k(V, W)$ instead of $S^k(V, W)$, we have the k-th **exterior power** of V.

Exterior Powers

The previous discussion motivates the following definition.

Definition 3.5.1 (Exterior Power). A k-th **exterior power** for V is a pair (φ, U), with $\varphi \in A^k(V, U)$, which is universal over V^k w.r.t. the properties P_1 = 'be alternating and k-linear' and P_2 = 'be linear'. In other words, for all $\psi \in A^k(V, W)$ there exists a unique $\tilde{\psi} \in \mathrm{Hom}(U, W)$ such that $\tilde{\psi} \circ \varphi = \psi$.

With this definition in mind, our task is to recover results similar to those of the 'usual' tensor product.

Lemma 3.5.1. Let α be a basis for V and $\varphi \in \mathrm{Hom}^k(V, W)$. If φ is alternating (and hence antisymmetric) with $v_j \in \alpha$ for all $1 \le j \le k$, then $\varphi \in A^k(V, W)$.

Theorem 3.5.2 (Existence and Basis of Exterior Powers)**.** For all $k \ge 1$ there exists a k-th exterior power for V.

Let $\alpha = (v_i)_{i \in I}$ be a basis for V, $\le$ a total order relation on I and $\alpha^k_< = \{(v_{i_1}, \ldots, v_{i_k}) \in \alpha^k : i_1 < \ldots < i_k\}$. If (φ, U) is a k-th exterior power for V, then $\varphi(\alpha^k_<)$ is a basis for U.

Proof. As in the proof of the Theorem 3.3.1, it suffices to show the second claim for a specific k-th exterior power. The overall argument here will be similar to that proof.

Consider a vector space U with $\dim(U) = \#\alpha^k_<$ and let $\iota : \alpha^k_< \longrightarrow U$ be a basis for U indexed by $\alpha^k_<$.

Notice that every element of α^k with distinct entries can be obtained from a single element of $\alpha^k_<$ by reordering its entries. We now construct φ using this reordering algorithm.

Let $\varphi \in \mathrm{Hom}^k(V, U)$ given by $\varphi|_{\alpha^k_<} = \iota$ (in the already ordered case), $\varphi(v_{i_1}, \ldots, v_{i_k}) = 0$ if there exists $1 \le j < l \le k$ such that $i_j = i_l$ (i.e. there is repetition), and

$$\varphi(v_{i_1}, \ldots, v_{i_j}, \ldots, v_{i_l}, \ldots, v_{i_k}) = -\varphi(v_{i_1}, \ldots, v_{i_l}, \ldots, v_{i_j}, \ldots, v_{i_k})$$

for all $1 \le j < l \le k$ (here the reordering algorithm is used). It follows from the Lemma 3.5.1 that $\varphi \in A^k(V, U)$.

We now need to show that (φ, U) satisfies the universal property. Given $\psi \in A^k(V, W)$, since ι is a basis for U, there exists a unique $\tilde{\psi} \in \mathrm{Hom}(U, W)$ such that $\tilde{\psi}(\iota(\mathbf{v})) = \psi(\mathbf{v})$, for all $\mathbf{v} \in \alpha^k_<$.

Since φ and ψ are alternating and k-linear, we have that $\tilde{\psi} \circ \varphi$ is alternating. Besides, $\tilde{\psi}(\varphi(\mathbf{v})) = \psi(\mathbf{v})$ for all $\mathbf{v} \in \alpha^k$. Hence, $\tilde{\psi} \circ \varphi = \psi$.

Suppose that $\xi \in \mathrm{Hom}(U, W)$ satisfies $\xi \circ \varphi = \psi$. Then, for all $\mathbf{v} \in \alpha^k$, we have that $\xi(\iota(\mathbf{v})) = \xi(\varphi(\mathbf{v})) = \psi(\mathbf{v})$. Thus, $\xi = \tilde{\psi}$, completing our proof. □

Theorem 3.5.3. If (φ, U) is a k-th exterior power for V, then for all vector space W, the function

$$\begin{aligned} \Gamma : A^k(V, W) &\longrightarrow \mathrm{Hom}(U, W) \\ \psi &\longmapsto \tilde{\psi} \end{aligned}$$

is an isomorphism.

Proof. Analogous to the proof of the Theorem 3.3.2. □

Definition 3.5.2 (Wedge Notation)**.** In the case of exterior product (also called **wedge product**), we use the symbol $\wedge$ instead of φ and $\bigwedge^k V$ to denote the k-th exterior product of V.

We also denote $v_1 \wedge \cdots \wedge w_k = \wedge(v_1, \ldots, v_k)$.

In this notation,

$$v_1 \wedge \cdots \wedge v_i \wedge \cdots \wedge v_j \wedge \cdots \wedge v_k = -v_1 \wedge \cdots \wedge v_j \wedge \cdots \wedge v_i \wedge \cdots \wedge v_k$$

for any $1 \leq i < j \leq k$.

Notice that $v_1 \wedge \cdots \wedge = 0$ if $v_1, \ldots, v_k$ is linearly dependent.

Let us compute the dimension of exterior power. By the Theorem 3.5.2, given a basis $\alpha = (v_i)_{i \in \mathrm{I}}$ for V and a total order relation on I, the vectors of the form

$$v_{i_1} \wedge \cdots \wedge v_{i_k}, \quad i_1 < \cdots < i_k$$

form a basis for $\bigwedge^k \mathrm{V}$. In particular, if $\dim(\mathrm{V}) = n \in \mathbb{Z}_{>0}$, we have

$$\dim(\wedge^k \mathrm{V}) = 0 \ \text{ if } k > n, \quad \text{ and } \dim(\wedge^k \mathrm{V}) = \binom{n}{k} \ \text{ if } 1 \leq k \leq n$$

It follows from the Theorem 3.5.3 that

$$\dim(\mathrm{A}^k(\mathrm{V},\mathrm{W})) = 0 \quad \text{ if } k > n, \quad \text{ and } \dim(\mathrm{A}^k(\mathrm{V},\mathrm{W})) = \binom{n}{k} \dim(\mathrm{W}) \ \text{ if } 1 \leq k \leq n$$

We can define the rank of the exterior product in an analogous way to the 'usual' tensor product.

Symmetric Powers

Now we turn our attention to the symmetric case.

Definition 3.5.3 (Symmetric Power). A k-th **symmetric power** for V is a pair (φ, U), with $\varphi \in \mathrm{S}^k(\mathrm{V},\mathrm{U})$, which is universal over V^k w.r.t. the properties $\mathrm{P}_1 =$ 'be symmetric and k-linear' and $\mathrm{P}_2 =$ 'be linear'. In other words, for all $\psi \in \mathrm{S}^k(\mathrm{V},\mathrm{W})$ there exists a unique $\tilde{\psi} \in \mathrm{Hom}(\mathrm{U},\mathrm{W})$ such that $\tilde{\psi} \circ \varphi = \psi$.

And again we have our existence result.

Theorem 3.5.4 (Existence and Basis of Symmetric Powers). For all $k \geq 1$ there exists a k-th symmetric power for V.

Let $\alpha = (v_i)_{i \in \mathrm{I}}$ be a basis for V, $\leq$ a total order relation on I and $\alpha^k_{\leq} = \{(v_{i_1}, \ldots, v_{i_k}) \in \alpha^k : i_1 \leq \ldots \leq i_k\}$. If (φ, U) is a k-th symmetric power for V, then $\varphi(\alpha^k_{\leq})$ is a basis for U.

Proof. Analogous to the Theorem 3.5.2. □

Theorem 3.5.5. If (φ, U) is a k-th symmetric power for V, then for all vector space W, the function

$$\begin{aligned} \Gamma : \mathrm{S}^k(\mathrm{V},\mathrm{W}) &\longrightarrow \mathrm{Hom}(\mathrm{U},\mathrm{W}) \\ \psi &\longmapsto \tilde{\psi} \end{aligned}$$

is an isomorphism.

Proof. Analogous to the proof of the Theorem 3.3.2. □

Definition 3.5.4 (Symmetric Power Notation)**.** We use $S^k V$ to denote the vector space of the universal pair of a k-th symmetric product of V. The corresponding k-linear function of the pair will be denoted by $\odot$.

We also denote $v_1 \odot \cdots \odot w_k = \odot(v_1, \ldots, v_k)$.

By the symmetric property,

$$v_1 \odot \cdots \odot v_i \odot \cdots \odot v_j \odot \cdots \odot v_k = v_1 \odot \cdots \odot v_j \odot \cdots \odot v_i \odot \cdots \odot v_k$$

for any $1 \leq i < j \leq k$.

Let us compute the dimension of a symmetric power. By the Theorem 3.5.4, given a basis $\alpha = (v_i)_{i \in I}$ for V and a total order relation on I, the vectors of the form

$$v_{i_1} \odot \cdots \odot v_{i_k}, \quad i_1 \leq \cdots \leq i_k$$

form a basis for $S^k V$. In particular, if $\dim(V) = n \in \mathbb{Z}_{>0}$, we have

$$\dim(S^k V) = \binom{n-1+k}{k} = \binom{n-1+k}{n-1}, \quad \forall\, k \geq 1$$

It follows from the Theorem 3.5.5 that

$$\dim(S^k(V, W)) = \binom{n+k-1}{k} \dim(W)$$

Determinants

Now we turn our attention to the determinant and how it can be defined using the theory of exterior products. Thus, we'll possess three viewpoints of determinants. The first was available only for matrices and used the Laplace expansion. The second construction used characteristic polynomials and generalized eigenvectors (recall it here). The third viewpoint will be built here. This construction is similar to the one in Appendix C.

For this, we'll consider alternating k-linear forms, i.e., $A^k(V) = A^k(V, \mathbb{F})$, and start by generalizing the idea of transpose. Given $T \in \mathrm{Hom}(V, W)$, consider

$$\begin{aligned} T^{\times k} : V^k &\longrightarrow W^k \\ (v_1, \ldots, v_k) &\longmapsto (T(v_1), \ldots, T(v_k)) \end{aligned}$$

and

$$\begin{aligned} T^{\wedge k} : A^k(W) &\longrightarrow A^k(V) \\ \varphi &\longmapsto \varphi \circ T^{\times k} \end{aligned}$$

It can be easily verified that $T^{\wedge k}(\varphi) \in A^k(V)$, for all $\varphi \in A^k(W)$ and that $T^{\wedge k}$ is linear.

Let us consider the particular case in which $W = V$ and $k = n = \dim(V)$. It follows that $\dim(A^n(V)) = 1$ and, therefore, every linear operator on $A^n(V)$ is a factor of a fixed scalar. Since $T^{\wedge n}$ is a linear operator on $A^n(V)$, there exists $\delta(T) \in \mathbb{F}$ such that

$$T^{\wedge n}(\varphi) = \delta(T)\varphi, \quad \forall\, \varphi \in A^n(V) \tag{3.9}$$

To compute $\delta(\mathrm{T})$, we need to evaluate $\mathrm{T}^{\wedge n}$ on a single nonzero element of $\mathrm{A}^n(\mathrm{V})$. Fixed a basis $\alpha = \{v_1, \dots, v_n\}$ for V, we take φ as the unique element of $\mathrm{A}^n(\mathrm{V})$ satisfying $\varphi(v_1, \dots, v_n) = 1$. Since $\mathrm{T}^{\wedge n}(\varphi)$ is determined by its value on $(v_1, \dots, v_n)$, we have

$$\delta(\mathrm{T}) = \delta(\mathrm{T})\varphi(v_1, \dots, v_n) \overset{(3.9)}{=} \mathrm{T}^{\wedge n}(\varphi)(v_1, \dots, v_n) = \varphi(\mathrm{T}(v_1), \dots, \mathrm{T}(v_n))$$

In particular, $\delta(\mathrm{Id_V}) = 1$. And if $\mathrm{S}, \mathrm{T} \in \mathrm{End}(\mathrm{V})$, then

$$\delta(\mathrm{T} \circ \mathrm{S}) = \varphi(\mathrm{T}(\mathrm{S}(v_1)), \dots, \mathrm{T}(\mathrm{S}(v_n))) \overset{(3.9)}{=} \delta(\mathrm{T})\varphi(\mathrm{S}(v_1), \dots, \mathrm{S}(v_n)) = \delta(\mathrm{T})\delta(\mathrm{S})$$

Theorem 3.5.6. For all $\mathrm{T} \in \mathrm{End}(\mathrm{V})$, $\delta(\mathrm{T}) = \det(\mathrm{T})$.

Proof. First, we'll prove for $\dim(\mathrm{V}) = 2$. Let $\alpha = \{v_1, v_2\}$ be a basis for V and $[\mathrm{T}]_\alpha^\alpha = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

If $\varphi \in \mathrm{A}^2(\mathrm{V})$ is the element satisfying $\varphi(v_1, v_2) = 1$, we have

$$\begin{aligned} \delta(\mathrm{T}) = \varphi(\mathrm{T}(v_1), \mathrm{T}(v_2)) &= \varphi(av_1 + cv_2, bv_1 + dv_2) \\ &= \varphi(av_1, dv_2) + \varphi(cv_2, bv_1) = ad - bc \end{aligned}$$

For the general case, let

1. $\alpha = \{v_1, \dots, v_n\}$ be a basis for V;
2. $\mathrm{W} = \mathrm{M}_{n,1}(\mathbb{F})$;
3. $\nu : \mathrm{W}^n \longrightarrow \mathrm{M}_n(\mathbb{F})$ such that $\mathrm{C}_j(\nu(\mathrm{A}_1, \dots, \mathrm{A}_n)) = \mathrm{A}_j$ for all $1 \le j \le n$;
4. $\rho_\alpha : \mathrm{W}^n \longrightarrow \mathrm{End}(\mathrm{V})$ such that $\rho_\alpha(\mathrm{A}_1, \dots, \mathrm{A}_n) = \mathrm{T}$, where $\mathrm{T} \in \mathrm{End}(\mathrm{V})$ is such that $[\mathrm{T}]_\alpha^\alpha = \nu(\mathrm{A}_1, \dots, \mathrm{A}_n)$ (ρ_α takes the matrix to the linear operator represented by it);
5. $\mu_\alpha : \mathrm{End}(\mathrm{V}) \longrightarrow \mathrm{M}_n(\mathbb{F})$, where $\mathrm{T} \longmapsto [\mathrm{T}]_\alpha^\alpha$ (μ_α takes the linear operator into its matrix representation);
6. $\varphi, \psi : \mathrm{W}^n \longrightarrow \mathbb{F}$ given by $\varphi = \det \circ \nu$ and $\psi = \delta \circ \rho_\alpha$.

In summary, we have the following diagram

$$\begin{tikzcd}
 & \mathrm{End}(\mathrm{V}) \arrow[dr, "\delta"] \arrow[dd, "\mu_\alpha"] & \\
\mathrm{W}^n \arrow[ur, "\rho_\alpha"] \arrow[dr, "\nu"'] \arrow[rr, dashed, bend left, "\psi"] \arrow[rr, dashed, bend right, "\varphi"'] & & \mathbb{F} \\
 & \mathrm{M}_n(\mathbb{F}) \arrow[ur, "\det"'] &
\end{tikzcd}$$

Notice that ρ_α and ν are bijective and consider the basis $\beta = \{E_{1,1}, \ldots, E_{n,1}\}$ for W. We know that $\varphi \in A^n(W)$. In fact, φ is the only element of $A^n(W)$ satisfying $\varphi(E_{1,1}, \ldots, E_{n,1}) = \det(I_n) = 1$, i.e., the determinant is a basis of $A^n(W)$.

We will show that

$$\psi \in A^n(W) \quad \text{and} \quad \psi(E_{1,1}, \ldots, E_{n,1}) = 1 \tag{3.10}$$

which implies that $\psi = \varphi$. Supposing that this is true, we complete the proof as follows.

We know that that $\det(T) = \det(\mu_\alpha(T))$. By definition of ρ_α, we have $\nu = \mu_\alpha \circ \rho_\alpha$, i.e., the left triangle is commutative. Thus,

$$\delta(T) = \psi(\rho_\alpha^{-1}(T)) = \varphi(\rho_\alpha^{-1}(T)) = \det(\nu(\rho_\alpha^{-1}(T))) = \det(\mu_\alpha(T)) = \det(T)$$

Let us prove (3.10). Since $\rho_\alpha(E_{1,1}, \ldots, E_{n,1}) = \mathrm{Id}_V$ and $\delta(\mathrm{Id}) = 1$, the second claim follows.

Let $1 \le k \le n$, $A, A_j \in W$, for all $1 \le j \le n$, and $\lambda \in \mathbb{F}$. And consider

$$\begin{aligned} T = \rho_\alpha(A_1, \ldots, A_n), \quad S &= \rho_\alpha(A_1, \ldots, A_{k-1}, A, A_{k+1}, \ldots, A_n), \\ R &= \rho_\alpha(A_1, \ldots, A_{k-1}, A_k + \lambda A, A_{k+1}, \ldots, A_n) \end{aligned}$$

Notice that $R(v_j) = T(v_j) = S(v_j)$ if $j \neq k$, and $R(v_k) = T(v_k) + \lambda S(v_k)$. Thus, if $\varphi \in A^n(V)$ satisfies $\varphi(v_1, \ldots, v_n = 1)$, then

$$\begin{aligned} \psi(A_1, \ldots, A_{k-1}, A_k + \lambda A, A_{k+1}, \ldots, A_n) &= \delta(R) = \varphi(R(v_1), \ldots, R(v_n)) \\ &= \varphi(R(v_1), \ldots, R(v_{k-1}), T(v_k) + \lambda S(v_k), R(v_{k+1}), \ldots, R(v_n)) \\ &= \varphi(T(v_1), \ldots, T(v_n)) + \lambda\varphi(S(v_1), \ldots, S(v_n)) \\ &= \psi(A_1, \ldots, A_n) + \lambda\psi(A_1, \ldots, A_{k-1}, A, A_{k+1}, \ldots, A_n) \end{aligned}$$

showing that ψ is n-linear. Finally, let us prove that ψ is alternating.

Given $A_1, \ldots, A_n \in W$, suppose that there exist $1 \le j < k \le n$ such that $A_j = A_k$. Thus, if $T = \rho_\alpha(A_1, \ldots, A_n)$, then $T(v_j) = T(v_k)$ and

$$\psi(A_1, \ldots, A_n) = \delta(T) = \varphi(T(v_1), \ldots, T(v_n)) = 0$$

since φ is alternating. □

3.6 Associative Algebras

In this section, we use the concept of the tensor product to build algebras, particularly the algebra of polynomials.

Definition 3.6.1 (Associative Algebra)**.** An **algebra** over a field $\mathbb{F}$ is a pair (A, m) in which A is a vector space over $\mathbb{F}$ and $m \in \mathrm{Hom}^2(A, A)$. In particular, m is a binary operation on A. We'll denote $m(a, b)$ simply by ab.

If m is associative, we say that (A, m) (or simply A) is an **associative algebra**. If there exists a neutral element for m, then A is said to be an **algebra with identity** and this element is usually denoted by ‘1’.

Notice that the bilinearity of m is equivalent to

$$(a+\lambda b)c = (ac)+\lambda(bc) \quad \text{and} \quad a(b+\lambda c) = (ab)+\lambda(ac), \quad \forall a,b,c \in \mathrm{A},\ \lambda \in \mathbb{F}$$

In particular, m distributes over the addition of vectors. Thus, we can think of m as multiplication on A.

More than that, if $\mathrm{A} \neq \{0\}$, then $1 \neq 0$.

Example 3.6.1. The following are examples of associative algebras.

1. Polynomials: $\mathrm{A} = \mathfrak{P}_n(\mathbb{F})$.
2. Square matrices: $\mathrm{A} = \mathrm{M}_n(\mathbb{F})$.
3. Linear operators: $\mathrm{A} = \mathrm{End}_{\mathbb{F}}(\mathrm{V})$ and m is composition.

Our goal here is to use the concept of tensors to fabricate algebras. Remark that instead of (A, m), we may consider the pair (A, μ) with $\mu := \tilde{m} \in \mathrm{Hom}(\mathrm{A}^{\otimes 2}, \mathrm{A})$.

Given a vector space V, consider $\mathrm{T(V)} = \bigoplus_{k\geq 0} \mathrm{T}^k(\mathrm{V})$, in which $\mathrm{T}^0(\mathrm{V}) := \mathbb{F}$. We'll define a structure of associative algebra with identity on T(V) such that $1 \in \mathbb{F}$ is the neutral element of multiplication. To do that, let $k, l \in \mathbb{Z}_{\geq 0}$ and consider

$$\begin{aligned} m_{k,l} : \mathrm{T}^k(\mathrm{V}) \times \mathrm{T}^l(\mathrm{V}) &\longrightarrow \mathrm{T}^{k+l}(\mathrm{V}) \\ (a,b) &\longmapsto a \otimes b \end{aligned}$$

Note that $m_{0,l}$ is the scalar multiplication of $\mathrm{T}^l(\mathrm{V})$. In particular, $m_{0,l}(1,b) = b$ for all $b \in \mathrm{T}^l(\mathrm{V})$.

Let $\mu_{k,l} = \tilde{m}_{k,l} \in \mathrm{Hom}(\mathrm{T}^k(\mathrm{V}) \otimes \mathrm{T}^l(\mathrm{V}),\ \mathrm{T}^{k+l}(\mathrm{V}))$ and recall that

$$\mathrm{T(V)}^{\otimes 2} = \bigoplus_{k,l\geq 0} \mathrm{T}^k(\mathrm{V}) \otimes \mathrm{T}^l(\mathrm{V})$$

Define $\mu \in \mathrm{Hom}(\mathrm{T(V)}^{\otimes 2},\ \mathrm{T(V)})$ by $\mu = \sum_{k,l\geq 0} \mu_{k,l}$. It can be verified that the pair $(\mathrm{T(V)}, \mu)$ is an associative algebra with identity.

Definition 3.6.2 (Tensor Algebra). The algebra $(\mathrm{T(V)}, \mu)$ constructed above is called the **tensor algebra** over V.

Definition 3.6.3 (Homomorphism of Algebras). Suppose that A and B are associative algebras with identity. We say that $f \in \mathrm{Hom}(\mathrm{A}, \mathrm{B})$ is an **algebra homomorphism** if

$$f(aa') = f(a)f(a'), \quad \forall\ a, a' \in \mathrm{A}, \quad \text{and } f(1) = 1$$

Remark (Universal Property of Tensor Algebras). Since $\mathrm{V} = \mathrm{T}^1(\mathrm{V})$, we may consider the inclusion $\varphi : \mathrm{V} \longrightarrow \mathrm{T(V)}$ given by $\varphi(v) = v$. The pair $(\mathrm{T(V)}, \varphi)$ is universal over V w.r.t. the properties $\mathrm{P}_1 =$ ‘be linear with an associative algebra with identity as codomain’ and $\mathrm{P}_2 =$ ‘be an algebra homomorphism’.

Thus, if A is an algebra with identity and $\mathrm{T} \in \mathrm{Hom}(\mathrm{V}, \mathrm{A})$, there exists a unique algebra homomorphism $f : \mathrm{T}(\mathrm{V}) \longrightarrow \mathrm{A}$ such that $f|_{\mathrm{V}} = \mathrm{T}$.

Definition 3.6.4 (Symmetric and Exterior Algebras)**.** Replacing $\mathrm{T}^k(\mathrm{V})$ by $\mathrm{S}^k\mathrm{V}$ or $\bigwedge^k \mathrm{V}$ on the construction above, we get to the concepts of **symmetric algebra** and **exterior algebra** (also called **Grassmann Algebra**), denoted by $\mathrm{S}(\mathrm{V})$ and $\bigwedge(\mathrm{V})$ respectively.

If $\dim(\mathrm{V}) = n$, then

$$\dim\left(\bigwedge(\mathrm{V})\right) = \sum_{k=0}^{n} \binom{n}{k} = 2^n$$

The question now is how to use these concepts in the context of polynomials.

Suppose that $\{v_1, \ldots, v_n\}$ is a basis for V. The tensors

$$v_{i_1} \odot \cdots \odot v_{i_k}, \quad k \geq 0 \quad \text{and} \quad i_1 \leq i_2 \leq \cdots \leq i_k$$

form a basis for $\mathrm{S}(\mathrm{V})$. Note that we obtain 1 for $k = 0$.

Omitting the symbol $\odot$ and changing the notation from v_i to x_i, this basis can be written as $x_{i_1} x_{i_2} \cdots x_{i_k}$. Since $v_i \cdot v_j = v_j \cdot v_i$, i.e. $x_i x_j = x_j x_i$, it follows that $\mathrm{S}(\mathrm{V})$ is isomorphic to the algebra $\mathfrak{P}_n(\mathbb{F})$ of the polynomials on n 'variables' and coefficients in $\mathbb{F}$.

Note that the variables here are elements of a basis. We may interpret $\mathrm{T}(\mathrm{V})$ as the algebra of polynomials on 'noncommutative variables'. Using the exterior algebra may be interpreted as the algebra of polynomials on 'alternating variables', i.e., $x_i^2 = 0$ and $x_i x_j = -x_j x_i$.

Example 3.6.2. If $\dim(\mathrm{V}) = 1$ and $x \in \mathrm{V} \setminus \{0\}$, then

$$\mathrm{T}(\mathrm{V}) \cong \mathrm{S}(\mathrm{V}) \cong \mathfrak{P}(\mathbb{F})$$

For the exterior algebra, $\bigwedge(\mathrm{V}) = \{a + bx : a, b \in \mathbb{F}\}$ with multiplication given by

$$(a + bx)(c + dx) = ac + (ad + bc)x$$

Appendix A

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.

A.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 A.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 '·' of $\mathbb{F}$ on V (called **scalar multiplication**) such that, for any $x, y \in V$ and $a, b \in \mathbb{F}$, $x+y \in V$ (closed under addition) and $a \cdot x \in V$ (invariant under scalar multiplication) satisfying:

1. $x+y = y+x$.
2. $(x+y)+z = x+(y+z)$.
3. There exists $0 \in V$ such that $x+0 = x$ for all $x \in V$.
4. For all $x \in V$, there exists $y \in V$ such that $x+y = 0$.
5. There exists $1 \in \mathbb{F}$ such that $1 \cdot x = x$ for all $x \in V$.
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 A.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 A.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 A.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 $a_1, \ldots, a_n \in \mathbb{F}$.

Subspaces

Definition A.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 A.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 A.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 A.1.4 (Intersection of subspaces is a subspace). If $W_1, \dots, W_n$ are subspaces of V, then $W = \bigcap_{i=1}^{n} W_i$ is also a subspace of V.

Definition A.1.4 (Span). Let $S \subseteq V$. The **subspace spanned** by S (or **span** of S), denoted by Span(S) or $[s_1, \dots, s_k]$ (where s_i is each vector of 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 A.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 A.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 A.1.5 (Generation of Spaces). Let $S \subseteq V$. We say that S **generates** (or **spans**) V if $\text{Span}(S) = V$.

Bases and Dimension

Definition A.1.6 (Linear Dependence). A subset S of V is **linearly dependent** if there exists a finite number of distinct vectors $u_1, \dots, u_n \in S$ and scalars $a_1, \dots, a_n \in \mathbb{F}$, 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 $u_1, \dots, u_n$ vanishes.

Theorem A.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. $v \in \mathrm{Span}(S)$.

Definition A.1.7 (Basis). A **basis** for V is a subset of V which is both linearly independent and generates V.

Example A.1.2. Let S be the subset of $\mathbb{F}^n$ containing

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

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 A.1.8. A subset of vectors $\{u_1,\dots,u_n\}$ of V is a basis iff. every $v \in V$ can be uniquely written in the form

$$v = a_1u_1 + \ldots + a_nu_n$$

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

Theorem A.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 $u_1,\dots,u_n$, 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 A.1.10. If V is a finitely generated vector space, then every basis of V has the same number of elements in it.

Definition A.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 A.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 A.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 A.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 A.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 A.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 A.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 A.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 A.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 A.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 $V_i = \mathrm{span}(\{u_1, \ldots, u_i\})$ for $i \geq 1$.

Theorem A.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 A.1.19. Every vector space has a basis.

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 $e_1, \dots, e_n$ is the standard basis for $\mathbb{R}^n$, then we express

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

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 A.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 A.1.12** (Coordinates). Let $\beta = \{v_1, \dots, v_n\}$ 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 $a_1, \dots, a_n \in \mathbb{F}$.
>
> 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 $\beta = \{\beta_1, \dots, \beta_n\}$ and $\gamma = \{\gamma_1, \dots, \gamma_n\}$ 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 $i \in \{1, 2, \dots, n\}$, 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 A.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 A.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.,

$$\begin{aligned}(1,0,0) &= a_{11}\cdot(1,0,1) + a_{21}\cdot(1,1,1) + a_{31}\cdot(1,1,2) \\ &= 1\cdot(1,0,1) + 1\cdot(1,1,1) - 1\cdot(1,1,2)\end{aligned}$$

$$\begin{aligned}(0,1,0) &= a_{12}\cdot(1,0,1) + a_{22}\cdot(1,1,1) + a_{32}\cdot(1,1,2) \\ &= -1\cdot(1,0,1) + 1\cdot(1,1,1) + 0\cdot(1,1,2)\end{aligned}$$

$$\begin{aligned}(0,0,1) &= a_{13}\cdot(1,0,1) + a_{23}\cdot(1,1,1) + a_{33}\cdot(1,1,2) \\ &= 0\cdot(1,0,1) - 1\cdot(1,1,1) + 1(1,1,2)\end{aligned}$$

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.

The Row and Column Spaces of a Matrix

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

Definition A.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 A.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 $\dim W \leq m$, 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.

A.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,

Basic Definitions

Definition A.2.1 (Linear Transformation)**.** Let V and W be two vector spaces over the same field $\mathbb{F}$. A **linear transformation** $T : V \longrightarrow W$ is a function satisfying:

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

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

Theorem A.2.1 (Properties)**.**

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

Example A.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

$$(Df)(x) = c_1 + 2c_2 x + \ldots + kc_k x^{k-1}$$

Then D is a linear transformation called the differentiation operator.

Example A.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 A.2.2. Let $\{v_1, \dots, v_n\}$ be a basis for V. Then for any vectors $w_1, \dots, w_n \in W$, there exists exactly one linear transformation $T : V \longrightarrow W$ such that

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

Definition A.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.,

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

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

Theorem A.2.3. The null space of T ker(T) is a subspace of V and Im(T) is a subspace of W.

Theorem A.2.4 (The Dimension Theorem (Rank–Nullity))**.** If $\dim(V) < \infty$, then

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

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

Definition A.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 A.2.5.

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

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

Theorem A.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. $\dim(\mathrm{Im}(T)) = \dim(V)$.

The Algebra of Linear Transformations

Theorem A.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 A.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 A.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 A.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 A.2.10 (Composition is also linear)**.** If T and U are both linear transformations, then their composition UT is a linear transformation.

Definition A.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 A.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 A.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 A.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 A.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 A.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 A.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 A.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 A.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 A.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.

Isomorphisms

Definition A.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 A.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 A.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.

Matrix Representation

Definition A.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

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

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 A.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 A.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 A.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 A.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 A.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 A.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 A.2.23. Let V be a finite-dimensional vector space and let

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

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 A.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

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

A.3 Diagonalization

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.

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

And suppose that T is a linear operator on a finite vector space V. If there exists an ordered basis $\beta = \{v_1, v_2, \dots, v_n\}$ 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

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

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?

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 A.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 A.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 A.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 A.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.

Diagonalization

Definition A.3.4 (Diagonalization). Let $T \in \mathrm{End}(V)$. 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 A.3.2. Let $\dim(V) < \infty$ and $T \in \mathrm{End}(V)$.

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 A.3.3. Suppose that $\dim(V) < \infty$ and let $T \in \mathrm{End}(V)$, $\lambda_1, \ldots, \lambda_k$ distinct eigenvalues of T and $W_i = \mathrm{Ker}(T - \lambda_i I)$. 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 $i = 1, 2, \dots, k$.

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

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 A.3.5 (Algebraic and Geometric Multiplicities)**.** Consider $T \in \mathrm{End}(V)$ 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 $\lambda_1, \lambda_2, \dots, \lambda_k$ are the roots of $p_T(\lambda)$, then by the fundamental theorem of algebra

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

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 $\dim \mathrm{Ker}(T - \lambda_i I)$.

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

Example A.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 eigenvector 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:

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

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 $\mathbb{R}^3$:

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

Theorem A.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**.

Direct Sums

Definition A.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 A.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 embedding 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.

Appendix B

Polynomials

B.1 Algebras

Definition B.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 B.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.

B.2 The Algebra of Polynomials

Definition B.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 B.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 B.2.2. The set of all polynomials over a given field is a commutative linear algebra with identity.

Corollary B.2.3. Suppose f, g, and h are polynomials such that $f \neq 0$ and $fg = fh$. Then $g = h$.

Definition B.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 B.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. $(fg)(v) = f(v)g(v)$.

B.3 Polynomial Ideals

Definition B.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 B.3.1. The polynomial $f = t^2 + 1$ is prime over $\mathbb{R}$ but not over $\mathbb{C}$, since $(t - i) \mid f$.

Definition B.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 B.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 B.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 $a_1x_1 + \dots a_rx_r$ 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 B.3.5 (Generated Ideal). The **ideal generated by** a set of elements $a_1, \dots, a_n \in \mathrm{A}$ 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.

B.4 Prime Factorization

This section shows how polynomials can be factored following Euclid's Division Algorithm.

Start by noticing that there exists a unique q, r such that
$$f = qg + r, \qquad \deg(r) < \deg(g)$$

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

Definition B.4.1 (Prime Factor and Multiplicity). If g is prime, monic and $g|f$, we say that g is a **prime factor** of f.

The **multiplicity** of g, as a prime factor of f, is defined as
$$\max\{m \in \mathbf{Z}_{\geq 0} : g^m | f\}$$

Theorem B.4.1. If $f \in \mathbb{F}[x] \setminus \{0\}$, then the set of prime factors is finite.

If $\{g_1, \dots, g_k\}$ is the set of prime factors of f, there exists a unique $u \in \mathbb{F} \setminus \{0\}$ such that $f = ug_1^{m_1} \dots g_k^{m_k}$, 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 $f_1, \dots, f_k \in \mathbb{F}[x] \setminus \{0\}$ if
$$g \in \bigcap_{j=1}^{k} \mathrm{Div}(f_j) \text{ and } h|g \quad \forall h \in \bigcap_{j=1}^{k} \mathrm{Div}(f_j)$$

Theorem B.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 $g = g_1 f_1 + \ldots + g_k f_k$.

Appendix C

Determinants

C.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 C.1.1 (Ring)**.** A **ring** is a set K together with two operations $+$ (addition) and $\cdot$ (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.

C.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].

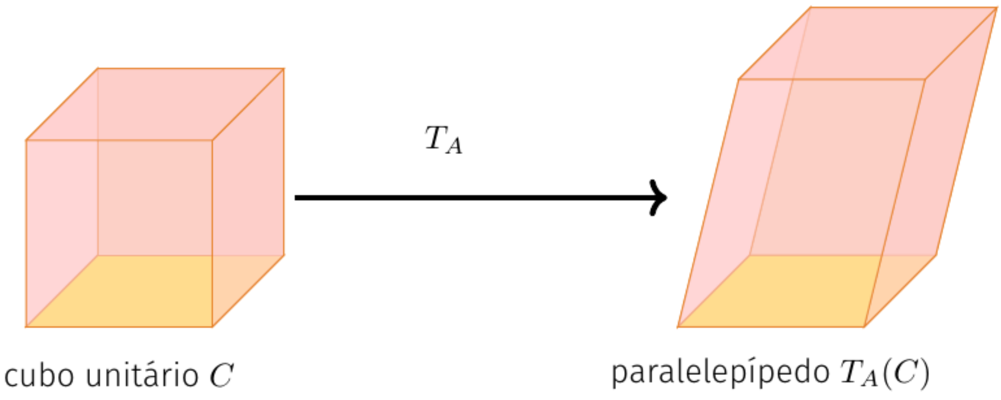

Figure C.1: The unitary cube C mapped into the parallelepiped $T_A(C)$ [Mir20].

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 C.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 C.2.1. The product of the diagonal entries of an $n \times n$ matrix is n-linear.

Lemma C.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 C.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 C.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 C.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 C.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 C.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 C.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,

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

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

If $j = i+1$, then $\mathrm{D(A)} = 0$, 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., $\mathrm{D(B)} = 0$.

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

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

Definition C.2.4. If $n > 1$ and A is an $n \times n$ matrix, let $\mathrm{A}(i|j)$ 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 $\mathrm{D}_{ij}(\mathrm{A}) = \mathrm{D}[\mathrm{A}(i|j)]$.

Theorem C.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

$$\mathrm{E}_j(\mathrm{A}) = \sum_{i=1}^{n} (-1)^{i+j}\mathrm{A}_{ij}\mathrm{D}_{ij}(\mathrm{A})$$

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 $\mathrm{D}_{ij}(\mathrm{A})$ 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 $\mathrm{A}_{ij}\mathrm{D}_{ij}(\mathrm{A})$ 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 $\mathrm{A}(i|j)$ has two equal rows, and $\mathrm{D}_{ij}(\mathrm{A}) = 0$. Therefore,

$$\mathrm{E}_j(\mathrm{A}) = (-1)^{k+j}\mathrm{A}_{kj}\mathrm{D}_{kj}(\mathrm{A}) + (-1)^{k+1+j}\mathrm{A}_{(k+1)j}\mathrm{D}_{(k+1)j}(\mathrm{A})$$

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

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

Therefore,

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

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

Third step: Show that $\mathrm{E}_j(\mathrm{I}) = 1$.

Suppose that D is a determinant function. Then

$$\mathrm{E}_j(\mathrm{I}_n) = \mathrm{D}(\mathrm{I}_{n-1}) = 1 \implies \mathrm{E}_j(\mathrm{I}_n) = 1$$

Hence, E_j is a determinant function. □

Corollary C.2.5. If K is a ring and n is a positive integer, then there exists at least one determinant function on $\mathbf{M}_n(\mathrm{K})$.

C.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 $a_1, a_2, \dots, a_n$ of the matrix A in terms of the standard basis $e_1, e_2, \dots, e_n$.

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 $1 \le i \le n$, each row can be expressed as

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

Hence,

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

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

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

Thus,

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

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

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

where $1 \le k_i \le n$, $i = 1, 2, \dots, n$.

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 C.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 C.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 C.3.1,

$$D(A) = \det(A)D(I)$$

Since $D(I) = D(IB) = \det B$, we have

$$\det(AB) = D(A) = (\det A)(\det B)$$

□

C.4 Properties of Determinants

Since there is no fundamental difference between rows and columns, the following result is expected.

Theorem C.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 C.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 C.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 C.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 C.2.4, if we fix any column j,

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

Definition C.4.1 (Cofactor)**.** The scalar $C_{ij} = (-1)^{i+j} \det A(i|j)$ is called the i, j **cofactor** of A.

Definition C.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 C.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,

$$A(\text{adj } A) = (\det A)I$$

$\square$

Theorem C.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. $\square$

Remark. For matrices with polynomial entries, the matrix is invertible over $\mathbb{F}[x]$ iff. its determinant is a non-zero scalar polynomial.

Example C.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.

C.5 Modules

A module over a ring K behaves like a vector space with K being used as scalars. More precisely,

Definition C.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 C.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 $v_1, \ldots, v_k$ 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 C.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 C.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 C.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 C.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 $v_1, \ldots, v_m$. □

Definition C.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

$$f = f(\beta_1)f_1 + \ldots + f(\beta_n)f_n$$

C.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 C.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 C.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 C.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.

Lemma C.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 C.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 C.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 $D(e_1, \ldots, e_n) = 1$. We also showed (Theorem C.3.1) that if L is any form in $\Lambda^n(K^n)$ then

$$L = L(e_1, \ldots, e_n) D$$

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

$$D = \sum_{\sigma} (\operatorname{sgn} \sigma) f_{\sigma_1} \otimes \ldots \otimes f_{\sigma_n}$$

where $f_1, \ldots, f_n$ are the coordinate functions on K^n and the sum is extended over the $n!$ different permutations σ of the set $\{1, \ldots, n\}$.

If we write the determinant of a matrix A as

$$\det A = \sum_{\sigma} (\operatorname{sgn} \sigma) A(\sigma_1, 1) \ldots A(\sigma_n, n)$$

then we obtain the following expression for D:

$$\begin{aligned} D(v_1, \ldots, v_n) &= \sum_\sigma (\text{sgn } \sigma) f_1(v_{\sigma_1}) \ldots f_n(v_{\sigma_n}) \\ &= \sum_\sigma (\text{sgn } \sigma) L(v_{\sigma_1}, \ldots, v_{\sigma_n}) \end{aligned}$$

where $L = f_1 \otimes \ldots \otimes f_n$.

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 $\{1, \ldots, r\}$, we obtain another r-linear function L_σ by defining

$$L_\sigma(v_1, \ldots, v_r) = L(v_{\sigma_1}, \ldots, v_{\sigma_r})$$

If L is alternating, then $L_\sigma = (\text{sgn } \sigma)L$.

For each $L \in M^r(V)$, define a function $\pi_r L \in M^r(V)$ by

$$\pi_r L = \sum_\sigma (\text{sgn } \sigma) L_\sigma$$

Lemma C.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

$$D = \pi_n(f_1 \otimes \ldots \otimes f_n)$$

Theorem C.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 C.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 $T \in \text{End}(V)$, 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.

C.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 C.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 C.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 product**) by

$$L \wedge M = \sum_{\sigma} (\text{sign } \sigma)(L \otimes M)_{\sigma}$$

Theorem C.7.1. The exterior product is associative.

Definition C.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.

[HK71] Kenneth Hoffman and Ray Kunze. *Linear Algebra*. 1971.

[Mir20] Daniel Miranda. *Álgebra Linear Avançada*. Notas de Aula, 2020.

[Mou22] Adriano Adrega de Moura. *Álgebra Linear com Geometria Analítica*. Notas de Aula, 2022.

[Rom05] Steven Roman. *Advanced Linear Algebra*. Springer, 2005.

[YIK89] Manin Yu I and Alexei I Kostrikin. *Linear Algebra and Geometry*. CRC Press, 1989.