Advanced Linear Algebra

Renko Usami

Contents

Preface

These notes are written as a structural version of advanced linear algebra. The first part develops vector spaces, quotients, duality, universal properties, tensor products, and exterior powers. The second part introduces modules over principal ideal domains and uses this language to reinterpret the structure theory of linear operators. The guiding idea is that many familiar constructions are best understood not by coordinates, but by the mapping properties that characterize them.

Throughout the first five chapters, vector spaces are taken over a fixed field $\mathbb{F}$. From Chapter 6 onward, R denotes a commutative ring with identity, and later a principal ideal domain when explicitly stated.

Chapter 1

Vector Spaces and Bases

1.1 Vector Spaces and Linear Combinations

Definition 1.1. A vector space over a field $\mathbb{F}$ is an abelian group $(V, +)$ together with a scalar multiplication

$$\mathbb{F} \times V \longrightarrow V, \qquad (a, v) \longmapsto av,$$

satisfying the usual distributive, associative, and unit axioms:

$$a(v + w) = av + aw, \quad (a + b)v = av + bv, \quad a(bv) = (ab)v, \quad 1v = v.$$

Definition 1.2. Let V be a vector space and let $S \subseteq V$. A vector $v \in V$ is called a linear combination of elements of S if there exist $s_1, \ldots, s_n \in S$ and $a_1, \ldots, a_n \in \mathbb{F}$ such that

$$v = a_1 s_1 + \cdots + a_n s_n.$$

The set of all linear combinations of elements of S is denoted by $\operatorname{span}(S)$.

Proposition 1.3. *For every subset $S \subseteq V$, the set $\operatorname{span}(S)$ is a subspace of V and is the smallest subspace of V containing S.*

Proof. First $0 \in \operatorname{span}(S)$, since 0 is the empty linear combination, or equivalently $0 = 0s$ if S is nonempty. If $u, v \in \operatorname{span}(S)$, then there exist finite expressions

$$u = \sum_{i=1}^{m} a_i s_i, \qquad v = \sum_{j=1}^{n} b_j t_j,$$

where $s_i, t_j \in S$. Then

$$u + v = \sum_{i=1}^{m} a_i s_i + \sum_{j=1}^{n} b_j t_j$$

is again a finite linear combination of elements of S. If $c \in \mathbb{F}$, then

$$cu = \sum_{i=1}^{m} (ca_i) s_i,$$

so $cu \in \operatorname{span}(S)$. Thus $\operatorname{span}(S)$ is a subspace.

It clearly contains S, since every $s \in S$ equals $1s$. If $W \subseteq V$ is any subspace containing S, then W is closed under finite linear combinations. Hence every element of $\operatorname{span}(S)$ lies in W. Therefore $\operatorname{span}(S) \subseteq W$, proving that $\operatorname{span}(S)$ is the smallest subspace containing S. □

Definition 1.4. A subset $S \subseteq V$ is called spanning if $\operatorname{span}(S) = V$. It is called linearly independent if for every finite subset $\{s_1, \dots, s_n\} \subseteq S$, the equality

$$a_1 s_1 + \cdots + a_n s_n = 0$$

implies $a_1 = \cdots = a_n = 0$.

Proposition 1.5. *Let $S \subseteq V$ be linearly independent and let $v \in V$. Then $S \cup \{v\}$ is linearly independent if and only if $v \notin \operatorname{span}(S)$.*

Proof. Suppose first that $v \in \operatorname{span}(S)$. Then there exist $s_1, \dots, s_n \in S$ and $a_1, \dots, a_n \in \mathbb{F}$ such that

$$v = a_1 s_1 + \cdots + a_n s_n.$$

Thus

$$v - a_1 s_1 - \cdots - a_n s_n = 0$$

is a nontrivial linear relation among elements of $S \cup \{v\}$, since the coefficient of v is 1. Hence $S \cup \{v\}$ is linearly dependent.

Conversely, suppose $S \cup \{v\}$ is linearly dependent. Then there is a nontrivial relation

$$cv + a_1 s_1 + \cdots + a_n s_n = 0$$

with $s_i \in S$. If $c = 0$, then the relation would be a nontrivial relation among elements of S, contradicting the linear independence of S. Hence $c \neq 0$, and

$$v = -c^{-1}(a_1 s_1 + \cdots + a_n s_n) \in \operatorname{span}(S).$$

This proves the equivalence. □

1.2 Bases and Dimension

Definition 1.6. A subset $B \subseteq V$ is called a basis of V if B is linearly independent and spans V.

Proposition 1.7. *Let $B \subseteq V$. Then B is a basis of V if and only if every vector $v \in V$ can be written uniquely in the form*

$$v = a_1 b_1 + \cdots + a_n b_n$$

with $a_i \in \mathbb{F}$ and $b_i \in B$ distinct.

Proof. Suppose B is a basis. Since B spans V, every $v \in V$ admits at least one expression as a finite linear combination of elements of B. Suppose

$$v = \sum_{i=1}^{m} a_i b_i = \sum_{j=1}^{n} c_j d_j.$$

After collecting all basis elements appearing in either expression, this gives a relation

$$\sum_{e \in E} \lambda_e e = 0$$

where E is a finite subset of B. Since B is linearly independent, all λ_e are zero. Hence the two expressions have the same coefficients for every basis element. Thus the expression is unique.

Conversely, if every vector has a unique expression as a finite linear combination of elements of B, then B spans V. If

$$a_1 b_1 + \cdots + a_n b_n = 0,$$

then this is an expression of 0. But 0 also has the empty expression, in which all coefficients are zero. By uniqueness, $a_1 = \cdots = a_n = 0$. Hence B is linearly independent. Therefore B is a basis. □

Lemma 1.8 (Zorn's lemma argument)**.** *Let $S \subseteq V$ be linearly independent. Then there exists a basis B of V such that $S \subseteq B$.*

Proof. Let $\mathcal{P}$ be the set of all linearly independent subsets of V containing S, ordered by inclusion. This set is nonempty because it contains S. Let $\{S_\alpha\}_{\alpha \in A}$ be a chain in $\mathcal{P}$. Put

$$S_\infty = \bigcup_{\alpha \in A} S_\alpha.$$

We claim that S_∞ is linearly independent. Indeed, any finite subset of S_∞ is contained in one of the S_α, because the family is totally ordered by inclusion. Since that S_α is linearly independent, the finite subset is linearly independent. Thus $S_\infty \in \mathcal{P}$ and is an upper bound for the chain.

By Zorn's lemma, $\mathcal{P}$ has a maximal element B. We show that B spans V. If not, choose $v \in V \setminus \operatorname{span}(B)$. By Proposition 1.3, $B \cup \{v\}$ is linearly independent, contradicting the maximality of B. Hence $\operatorname{span}(B) = V$, so B is a basis containing S. □

Theorem 1.9. *Every vector space has a basis.*

Proof. The empty set is linearly independent. Applying the preceding lemma to $S = \varnothing$ gives a basis of V. □

Lemma 1.10 (Replacement lemma)**.** *Let V be a vector space generated by n vectors. Then any linearly independent subset of V has at most n elements.*

Proof. Let $V = \operatorname{span}(v_1, \ldots, v_n)$ and let $w_1, \ldots, w_m$ be linearly independent. We prove by induction on r that, after replacing r of the v_i by $w_1, \ldots, w_r$, we still have a spanning set of V. For $r = 0$ this is clear.

Assume $0 \leq r < m$ and that

$$\{w_1, \ldots, w_r, v'_{r+1}, \ldots, v'_n\}$$

spans V. Since w_{r+1} lies in its span, we may write

$$w_{r+1} = a_1 w_1 + \cdots + a_r w_r + b_{r+1} v'_{r+1} + \cdots + b_n v'_n.$$

At least one coefficient b_j is nonzero; otherwise w_{r+1} would lie in $\operatorname{span}(w_1, \ldots, w_r)$, contradicting the linear independence of the w_i. After reindexing the remaining v' terms, suppose $b_{r+1} \neq 0$. Then v'_{r+1} can be solved as a linear combination of $w_1, \ldots, w_{r+1}, v'_{r+2}, \ldots, v'_n$. Hence replacing v'_{r+1} by w_{r+1} preserves the span.

If $m > n$, after n replacements we obtain that $w_1, \ldots, w_n$ span V. Then $w_{n+1} \in \operatorname{span}(w_1, \ldots, w_n)$, contradicting linear independence. Therefore $m \leq n$. □

Theorem 1.11. *Any two finite bases of a vector space have the same cardinality.*

Proof. Let B and C be finite bases. Since B spans V and C is linearly independent, the replacement lemma gives $|C| \leq |B|$. Since C spans V and B is linearly independent, the same lemma gives $|B| \leq |C|$. Hence $|B| = |C|$. □

Definition 1.12. If V has a finite basis, the common cardinality of all finite bases is called the dimension of V and is denoted by $\dim V$. If V has no finite basis, we say V is infinite-dimensional.

1.3 Subspaces and Direct Sums

Definition 1.13. Let $\{W_i\}_{i\in I}$ be a family of subspaces of V. The sum of the family is

$$\sum_{i\in I} W_i = \left\{ w_{i_1} + \cdots + w_{i_n} : n \geq 0,\ w_{i_j} \in W_{i_j} \right\}.$$

The sum is called direct if every vector in $\sum_i W_i$ has a unique expression as a finite sum of elements from the W_i. In that case we write

$$\bigoplus_{i\in I} W_i.$$

Proposition 1.14. *For subspaces $W_1, \ldots, W_n \subseteq V$, the sum $W_1 + \cdots + W_n$ is direct if and only if whenever*

$$w_1 + \cdots + w_n = 0, \qquad w_i \in W_i,$$

one has $w_1 = \cdots = w_n = 0$.

Proof. Suppose the sum is direct. Since 0 has the expression $0 + \cdots + 0$, uniqueness implies that any expression $w_1 + \cdots + w_n = 0$ has all $w_i = 0$.

Conversely, suppose the stated condition holds. If a vector v has two expressions

$$v = w_1 + \cdots + w_n = w_1' + \cdots + w_n',$$

then

$$(w_1 - w_1') + \cdots + (w_n - w_n') = 0.$$

Since $w_i - w_i' \in W_i$, the condition gives $w_i - w_i' = 0$ for all i. Hence $w_i = w_i'$ for all i, so the expression is unique. Thus the sum is direct. □

Corollary 1.15. *For two subspaces $U, W \subseteq V$, the sum $U+W$ is direct if and only if $U \cap W = \{0\}$.*

Proof. If $U + W$ is direct and $x \in U \cap W$, then $x + (-x) = 0$ with $x \in U$ and $-x \in W$. By the preceding proposition, $x = 0$. Hence $U \cap W = \{0\}$.

Conversely, suppose $U \cap W = \{0\}$. If $u + w = 0$, then $u = -w \in U \cap W$, so $u = 0$ and $w = 0$. By the preceding proposition, $U + W$ is direct. □

Proposition 1.16. *Let $W_1, \ldots, W_n \subseteq V$ be finite-dimensional subspaces. If B_i is a basis of W_i for each i, then $W_1 + \cdots + W_n$ is direct if and only if $B_1 \cup \cdots \cup B_n$ is a basis of $W_1 + \cdots + W_n$.*

Proof. The union $B_1 \cup \cdots \cup B_n$ always spans $W_1 + \cdots + W_n$. Thus it remains to discuss linear independence.

Suppose the sum is direct. If

$$\sum_{i=1}^{n} \sum_{b\in B_i} a_{i,b} b = 0,$$

then for each i, the vector $w_i = \sum_{b\in B_i} a_{i,b} b$ lies in W_i, and $w_1 + \cdots + w_n = 0$. Directness gives $w_i = 0$ for every i. Since B_i is a basis, all coefficients $a_{i,b}$ are zero. Hence the union is linearly independent.

Conversely, suppose the union is a basis. If $w_1 + \cdots + w_n = 0$ with $w_i \in W_i$, express each w_i in the basis B_i. This gives a linear relation among vectors in $B_1 \cup \cdots \cup B_n$. Since this union is linearly independent, all coefficients are zero, so every $w_i = 0$. Hence the sum is direct. □

Corollary 1.17. *If $V = W_1 \oplus \cdots \oplus W_n$ and each W_i is finite-dimensional, then*

$$\dim V = \sum_{i=1}^{n} \dim W_i.$$

Proof. Choose a basis B_i of each W_i. By the preceding proposition, $B_1 \cup \cdots \cup B_n$ is a basis of V. Therefore

$$\dim V = |B_1 \cup \cdots \cup B_n| = \sum_{i=1}^{n} |B_i| = \sum_{i=1}^{n} \dim W_i.$$

□

1.4 External Direct Sums and Products

Definition 1.18. Let $\{V_i\}_{i \in I}$ be a family of vector spaces. The external direct product is

$$\prod_{i \in I} V_i = \{(v_i)_{i \in I} : v_i \in V_i\}.$$

The external direct sum is

$$\bigoplus_{i \in I} V_i = \left\{ (v_i)_{i \in I} \in \prod_{i \in I} V_i : v_i = 0 \text{ for all but finitely many } i \right\}.$$

Proposition 1.19. *Both $\prod_i V_i$ and $\bigoplus_i V_i$ are vector spaces under componentwise addition and scalar multiplication. If I is finite, then*

$$\bigoplus_{i \in I} V_i = \prod_{i \in I} V_i.$$

Proof. Componentwise addition and scalar multiplication satisfy the vector space axioms because each V_i does. Thus $\prod_i V_i$ is a vector space. The direct sum is closed under addition because the union of two finite supports is finite, and it is closed under scalar multiplication because scalar multiplication does not enlarge support. Hence it is a subspace of the product.

If I is finite, every element of the product has finite support automatically. Thus the product and direct sum are equal. □

Chapter 2

Quotients, Linear Maps, and Isomorphism Theorems

2.1 Quotient Spaces

Definition 2.1. Let $W \subseteq V$ be a subspace. Define an equivalence relation on V by

$$v \sim v' \quad \iff \quad v - v' \in W.$$

The equivalence class of v is denoted by $v + W$ and is called a coset of W.

Proposition 2.2. *The relation $\sim$ is an equivalence relation.*

Proof. For every $v \in V$, $v - v = 0 \in W$, so $v \sim v$. If $v \sim v'$, then $v - v' \in W$, hence $v' - v = -(v - v') \in W$, so $v' \sim v$. If $v \sim v'$ and $v' \sim v''$, then $v - v' \in W$ and $v' - v'' \in W$, so

$$v - v'' = (v - v') + (v' - v'') \in W.$$

Thus $v \sim v''$. Hence $\sim$ is reflexive, symmetric, and transitive. □

Definition 2.3. The quotient set V/W is the set of cosets $v + W$. Define addition and scalar multiplication by

$$(v + W) + (v' + W) = (v + v') + W,$$
$$a(v + W) = av + W.$$

Proposition 2.4. *The operations above are well-defined, and V/W is a vector space.*

Proof. Suppose $v + W = \tilde{v} + W$ and $v' + W = \tilde{v}' + W$. Then $v - \tilde{v} \in W$ and $v' - \tilde{v}' \in W$. Hence

$$(v + v') - (\tilde{v} + \tilde{v}') = (v - \tilde{v}) + (v' - \tilde{v}') \in W,$$

so $(v + v') + W = (\tilde{v} + \tilde{v}') + W$. Thus addition is well-defined. Similarly, for $a \in \mathbb{F}$,

$$av - a\tilde{v} = a(v - \tilde{v}) \in W,$$

so scalar multiplication is well-defined.

The vector space axioms follow from the corresponding axioms in V. For example, associativity of addition follows because

$$((u + W) + (v + W)) + (w + W) = ((u + v) + w) + W = (u + (v + w)) + W.$$

The zero vector is $0 + W = W$, and the additive inverse of $v + W$ is $(-v) + W$. Therefore V/W is a vector space. □

2.2 Linear Transformations

Definition 2.5. Let V, W be vector spaces. A map $T : V \to W$ is called linear if

$$T(av + bv') = aT(v) + bT(v')$$

for all $a, b \in \mathbb{F}$ and all $v, v' \in V$. The vector space of all linear maps from V to W is denoted by $\mathrm{Hom}_{\mathbb{F}}(V, W)$ or $\mathcal{L}(V, W)$.

Proposition 2.6. *The set $\mathcal{L}(V, W)$ is a vector space under pointwise addition and scalar multiplication.*

Proof. For $S, T \in \mathcal{L}(V, W)$ and $a \in \mathbb{F}$, define

$$(S + T)(v) = S(v) + T(v), \qquad (aT)(v) = aT(v).$$

If $v, v' \in V$ and $c, d \in \mathbb{F}$, then

$$(S + T)(cv + dv') = S(cv + dv') + T(cv + dv')$$

$$= cS(v) + dS(v') + cT(v) + dT(v') = c(S + T)(v) + d(S + T)(v').$$

Thus $S + T$ is linear. Similarly,

$$(aT)(cv + dv') = aT(cv + dv') = acT(v) + adT(v') = c(aT)(v) + d(aT)(v'),$$

so aT is linear. The vector space axioms are inherited pointwise from W. □

Definition 2.7. For a linear map $T : V \to W$, define

$$\ker T = \{v \in V : T(v) = 0\}, \qquad \mathrm{im}\, T = \{T(v) : v \in V\}.$$

Proposition 2.8. *Let $T : V \to W$ be linear. Then $\ker T$ is a subspace of V and $\mathrm{im}\, T$ is a subspace of W.*

Proof. If $v, v' \in \ker T$ and $a, b \in \mathbb{F}$, then

$$T(av + bv') = aT(v) + bT(v') = 0,$$

so $av + bv' \in \ker T$. Thus $\ker T$ is a subspace.

If $w, w' \in \mathrm{im}\, T$, then $w = T(v)$ and $w' = T(v')$ for some $v, v' \in V$. For $a, b \in \mathbb{F}$,

$$aw + bw' = aT(v) + bT(v') = T(av + bv') \in \mathrm{im}\, T.$$

Thus $\mathrm{im}\, T$ is a subspace. □

Proposition 2.9. *A linear map $T : V \to W$ is injective if and only if $\ker T = \{0\}$.*

Proof. Suppose T is injective. If $v \in \ker T$, then $T(v) = 0 = T(0)$, so $v = 0$. Thus $\ker T = \{0\}$.

Conversely, suppose $\ker T = \{0\}$. If $T(v) = T(v')$, then $T(v - v') = 0$, so $v - v' \in \ker T = \{0\}$. Hence $v = v'$, and T is injective. □

2.3 The First Isomorphism Theorem

Theorem 2.10 (First isomorphism theorem)**.** *Let $T : V \to W$ be linear. Then there is a canonical isomorphism*

$$V/\ker T \cong \operatorname{im} T$$

given by

$$v + \ker T \longmapsto T(v).$$

Proof. Define $\overline{T} : V/\ker T \to \operatorname{im} T$ by $\overline{T}(v + \ker T) = T(v)$. We first check that this is well-defined. If $v + \ker T = v' + \ker T$, then $v - v' \in \ker T$, so $T(v - v') = 0$, hence $T(v) = T(v')$.

The map is linear because

$$\overline{T}(a(v + \ker T) + b(v' + \ker T)) = \overline{T}((av + bv') + \ker T)$$

$$= T(av + bv') = aT(v) + bT(v') = a\overline{T}(v + \ker T) + b\overline{T}(v' + \ker T).$$

It is surjective by definition of $\operatorname{im} T$. If $\overline{T}(v + \ker T) = 0$, then $T(v) = 0$, so $v \in \ker T$, and hence $v + \ker T = 0 + \ker T$. Thus $\overline{T}$ is injective. Therefore $\overline{T}$ is an isomorphism. □

2.4 Rank-Nullity

Definition 2.11. If $T : V \to W$ is linear and V is finite-dimensional, define

$$\operatorname{rank} T = \dim \operatorname{im} T, \qquad \operatorname{nullity} T = \dim \ker T.$$

Theorem 2.12 (Rank-nullity theorem)**.** *Let $T : V \to W$ be a linear map and assume V is finite-dimensional. Then*

$$\dim V = \dim \ker T + \dim \operatorname{im} T.$$

Proof. Choose a basis $u_1, \ldots, u_r$ of $\ker T$. Extend it to a basis

$$u_1, \ldots, u_r, v_1, \ldots, v_s$$

of V. We claim that $T(v_1), \ldots, T(v_s)$ is a basis of $\operatorname{im} T$.

First, it spans $\operatorname{im} T$. If $w \in \operatorname{im} T$, then $w = T(v)$ for some $v \in V$. Write

$$v = \sum_{i=1}^{r} a_i u_i + \sum_{j=1}^{s} b_j v_j.$$

Applying T gives

$$w = T(v) = \sum_{j=1}^{s} b_j T(v_j),$$

because $T(u_i) = 0$. Thus the $T(v_j)$ span $\operatorname{im} T$.

They are linearly independent. If

$$\sum_{j=1}^{s} b_j T(v_j) = 0,$$

then $T(\sum_j b_j v_j) = 0$, so $\sum_j b_j v_j \in \ker T$. Hence

$$\sum_{j=1}^{s} b_j v_j = \sum_{i=1}^{r} a_i u_i$$

for some coefficients a_i. Since $u_1, \dots, u_r, v_1, \dots, v_s$ is a basis, all $b_j = 0$. Thus $T(v_1), \dots, T(v_s)$ is a basis of $\operatorname{im} T$.

Therefore $\dim \ker T = r$ and $\dim \operatorname{im} T = s$, while $\dim V = r + s$. □

2.5 Matrices and Change of Basis

Definition 2.13. Let V, W be finite-dimensional vector spaces with ordered bases

$$A = (v_1, \dots, v_n), \qquad B = (w_1, \dots, w_m).$$

For $T : V \to W$ linear, the matrix of T with respect to A and B is the $m \times n$ matrix $[T]_{B,A} = (a_{ij})$ determined by

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

Proposition 2.14. *Let $S : U \to V$ and $T : V \to W$ be linear maps between finite-dimensional vector spaces with ordered bases A for U, B for V, and C for W. Then*

$$[T \circ S]_{C,A} = [T]_{C,B}[S]_{B,A}.$$

Proof. Write $A = (u_1, \dots, u_r)$, $B = (v_1, \dots, v_n)$, and $C = (w_1, \dots, w_m)$. Suppose

$$S(u_j) = \sum_{k=1}^{n} b_{kj} v_k, \qquad T(v_k) = \sum_{i=1}^{m} a_{ik} w_i.$$

Then

$$T(S(u_j)) = T\left(\sum_{k=1}^{n} b_{kj} v_k\right) = \sum_{k=1}^{n} b_{kj} T(v_k) = \sum_{k=1}^{n} b_{kj} \sum_{i=1}^{m} a_{ik} w_i.$$

Thus the coefficient of w_i in $T(S(u_j))$ is

$$\sum_{k=1}^{n} a_{ik} b_{kj},$$

which is precisely the (i, j)-entry of the product $[T]_{C,B}[S]_{B,A}$. □

Proposition 2.15. *Let $T : V \to V$ be linear and let A, B be two ordered bases of V. Let $P = [\mathrm{id}_V]_{A,B}$ be the change-of-basis matrix from B-coordinates to A-coordinates. Then*

$$[T]_{B,B} = P^{-1}[T]_{A,A}P.$$

Proof. By the composition rule,

$$[T]_{A,B} = [T]_{A,A}[\mathrm{id}_V]_{A,B} = [T]_{A,A}P.$$

Also

$$[T]_{A,B} = [\mathrm{id}_V]_{A,B}[T]_{B,B} = P[T]_{B,B}.$$

Therefore $P[T]_{B,B} = [T]_{A,A}P$. Since P is invertible, multiplying by P^{-1} on the left gives

$$[T]_{B,B} = P^{-1}[T]_{A,A}P.$$

□

Chapter 3

Dual Spaces and Annihilators

3.1 Dual Spaces

Definition 3.1. The dual space of a vector space V is

$$V^* = \mathrm{Hom}_{\mathbb{F}}(V, \mathbb{F}).$$

Elements of V^* are called linear functionals.

Definition 3.2. Let V be finite-dimensional with ordered basis $B = (v_1, \ldots, v_n)$. The dual basis $B^* = (v_1^*, \ldots, v_n^*)$ of V^* is defined by

$$v_i^*(v_j) = \delta_{ij}.$$

Proposition 3.3. *If $B = (v_1, \ldots, v_n)$ is a basis of V, then $B^* = (v_1^*, \ldots, v_n^*)$ is a basis of V^*.*

Proof. First, the functionals v_i^* are linearly independent. If

$$a_1 v_1^* + \cdots + a_n v_n^* = 0,$$

then evaluating at v_j gives

$$a_j = \sum_{i=1}^{n} a_i v_i^*(v_j) = 0.$$

Thus all $a_j = 0$.

Now let $f \in V^*$. We claim

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

Indeed, the two sides agree on each basis vector v_j:

$$\left(\sum_{i=1}^{n} f(v_i) v_i^* \right)(v_j) = \sum_{i=1}^{n} f(v_i)\delta_{ij} = f(v_j).$$

Since two linear maps that agree on a basis are equal, the claim follows. Therefore B^* is a basis of V^*. □

Corollary 3.4. *If V is finite-dimensional, then*

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

Proof. A basis of V with n elements gives a dual basis of V^* with n elements. Hence $\dim V^* = n = \dim V$. □

Definition 3.5. The canonical map from V to its double dual is

$$\eta_V : V \longrightarrow V^{**}, \qquad \eta_V(v)(f) = f(v)$$

for $v \in V$ and $f \in V^*$.

Proposition 3.6. *The map $\eta_V : V \to V^{**}$ is linear and injective. If V is finite-dimensional, then it is an isomorphism.*

Proof. For $a, b \in \mathbb{F}$ and $v, w \in V$,

$$\eta_V(av + bw)(f) = f(av + bw) = af(v) + bf(w) = a\eta_V(v)(f) + b\eta_V(w)(f).$$

Thus η_V is linear.

If $v \neq 0$, extend $\{v\}$ to a basis $v, v_2, \ldots, v_n$ of the subspace it generates inside a basis of V; in the infinite-dimensional case, use Zorn's lemma to extend $\{v\}$ to a basis of V. Define f on that basis by $f(v) = 1$ and $f(b) = 0$ for all other basis vectors. Then $f \in V^*$ and $\eta_V(v)(f) = 1$. Hence $\eta_V(v) \neq 0$, so η_V is injective.

If V is finite-dimensional, then $\dim V = \dim V^* = \dim V^{**}$. An injective linear map between finite-dimensional spaces of the same dimension is an isomorphism. □

3.2 Dual Maps

Definition 3.7. Let $T : V \to W$ be linear. The dual map, or adjoint map, is

$$T^* : W^* \longrightarrow V^*, \qquad T^*(f) = f \circ T.$$

Proposition 3.8. *The map T^* is linear. Moreover, if $S : U \to V$ and $T : V \to W$, then*

$$(T \circ S)^* = S^* \circ T^*.$$

Also $\mathrm{id}_V^* = \mathrm{id}_{V^*}$.

Proof. For $f, g \in W^*$ and $a, b \in \mathbb{F}$,

$$T^*(af + bg) = (af + bg) \circ T = a(f \circ T) + b(g \circ T) = aT^*(f) + bT^*(g).$$

Thus T^* is linear.

For $h \in W^*$,

$$(T \circ S)^*(h) = h \circ T \circ S = S^*(h \circ T) = S^*(T^*(h)).$$

Thus $(T \circ S)^* = S^* \circ T^*$. Finally,

$$\mathrm{id}_V^*(f) = f \circ \mathrm{id}_V = f,$$

so $\mathrm{id}_V^* = \mathrm{id}_{V^*}$. □

3.3 Annihilators

Definition 3.9. Let $S \subseteq V$. The annihilator of S is

$$S^0 = \{f \in V^* : f(s) = 0 \text{ for all } s \in S\}.$$

If $W \subseteq V$ is a subspace, W^0 is also called the annihilator of W.

Proposition 3.10. *For every subset $S \subseteq V$, the annihilator S^0 is a subspace of V^*.*

Proof. Let $f, g \in S^0$ and $a, b \in \mathbb{F}$. For every $s \in S$,

$$(af + bg)(s) = af(s) + bg(s) = 0.$$

Thus $af + bg \in S^0$, so S^0 is a subspace. □

Proposition 3.11. *Let V be finite-dimensional and let $W \subseteq V$ be a subspace. Then*

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

Proof. Choose a basis $w_1, \ldots, w_r$ of W and extend it to a basis

$$w_1, \ldots, w_r, v_{r+1}, \ldots, v_n$$

of V. Let

$$w_1^*, \ldots, w_r^*, v_{r+1}^*, \ldots, v_n^*$$

be the dual basis. We claim that $v_{r+1}^*, \ldots, v_n^*$ form a basis of W^0.

Each v_j^* with $j > r$ vanishes on $w_1, \ldots, w_r$, hence lies in W^0. If

$$f = \sum_{i=1}^{r} a_i w_i^* + \sum_{j=r+1}^{n} b_j v_j^* \in W^0,$$

then evaluating at w_i gives $a_i = f(w_i) = 0$ for $1 \le i \le r$. Hence f lies in the span of $v_{r+1}^*, \ldots, v_n^*$. These functionals are linearly independent because they are part of a basis of V^*. Thus they form a basis of W^0. Therefore

$$\dim W^0 = n - r = \dim V - \dim W.$$

□

3.4 Duality and Quotients

Theorem 3.12. *Let $W \subseteq V$ be a subspace. There is a canonical isomorphism*

$$(V/W)^* \cong W^0.$$

Proof. Let $\pi : V \to V/W$ be the quotient map. Define

$$\Phi : (V/W)^* \longrightarrow V^*, \qquad \Phi(f) = f \circ \pi.$$

If $w \in W$, then $\pi(w) = 0$, so $(f \circ \pi)(w) = 0$. Thus $\Phi(f) \in W^0$, and we may view Φ as a map $(V/W)^* \to W^0$.

The map Φ is linear by the linearity of composition. It is injective: if $f \circ \pi = 0$, then for every coset $v + W \in V/W$,

$$f(v + W) = f(\pi(v)) = (f \circ \pi)(v) = 0,$$

so $f = 0$.

It is surjective. Let $g \in W^0$. Define $\overline{g} : V/W \to \mathbb{F}$ by

$$\overline{g}(v + W) = g(v).$$

This is well-defined: if $v + W = v' + W$, then $v - v' \in W$, so $g(v - v') = 0$, hence $g(v) = g(v')$. It is linear because g is linear. Finally, $\Phi(\overline{g}) = \overline{g} \circ \pi = g$. Hence Φ is an isomorphism. □

Proposition 3.13. *Let $T : V \to W$ be linear. Then*

$$\ker T^* = (\operatorname{im} T)^0.$$

If V and W are finite-dimensional, then

$$\operatorname{im} T^* = (\ker T)^0.$$

Proof. For $f \in W^*$, we have $f \in \ker T^*$ if and only if $f \circ T = 0$, which holds if and only if f vanishes on every vector of the form $T(v)$. This is exactly the condition $f \in (\operatorname{im} T)^0$.

Assume now that V and W are finite-dimensional. If $f \in \operatorname{im} T^*$, then $f = g \circ T$ for some $g \in W^*$. For $v \in \ker T$, $f(v) = g(T(v)) = 0$, so $f \in (\ker T)^0$. Hence $\operatorname{im} T^* \subseteq (\ker T)^0$.

To prove equality, compare dimensions. By rank-nullity applied to T^*,

$$\dim \operatorname{im} T^* = \dim W^* - \dim \ker T^*.$$

Using $\ker T^* = (\operatorname{im} T)^0$ and the dimension formula for annihilators,

$$\dim \ker T^* = \dim W - \dim \operatorname{im} T.$$

Thus

$$\dim \operatorname{im} T^* = \dim W - (\dim W - \dim \operatorname{im} T) = \dim \operatorname{im} T.$$

On the other hand,

$$\dim(\ker T)^0 = \dim V - \dim \ker T = \dim \operatorname{im} T$$

by rank-nullity. Therefore the inclusion $\operatorname{im} T^* \subseteq (\ker T)^0$ is an equality. □

Chapter 4

Universal Properties in Linear Algebra

4.1 Categories, Functors, and Universal Properties

Definition 4.1. A category $\mathcal{C}$ consists of the following data:

(i) a class of objects $\mathrm{Ob}(\mathcal{C})$;

(ii) for every pair of objects X, Y, a set $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ whose elements are called morphisms from X to Y;

(iii) for every object X, an identity morphism $\mathrm{id}_X \in \mathrm{Hom}_{\mathcal{C}}(X, X)$;

(iv) for every triple of objects X, Y, Z, a composition law

$$\mathrm{Hom}_{\mathcal{C}}(Y, Z) \times \mathrm{Hom}_{\mathcal{C}}(X, Y) \longrightarrow \mathrm{Hom}_{\mathcal{C}}(X, Z), \qquad (g, f) \mapsto g \circ f.$$

These data are required to satisfy associativity

$$h \circ (g \circ f) = (h \circ g) \circ f$$

whenever both sides are defined, and the identity laws

$$f \circ \mathrm{id}_X = f, \qquad \mathrm{id}_Y \circ f = f$$

for every morphism $f : X \to Y$.

Example 4.2. The category $\mathbf{Vec}_{\mathbb{F}}$ has vector spaces over $\mathbb{F}$ as objects and linear maps as morphisms. The category R-**Mod** has R-modules as objects and R-linear maps as morphisms. The category **Set** has sets as objects and functions as morphisms.

Definition 4.3. A morphism $f : X \to Y$ in a category $\mathcal{C}$ is called an isomorphism if there exists a morphism $g : Y \to X$ such that

$$g \circ f = \mathrm{id}_X, \qquad f \circ g = \mathrm{id}_Y .$$

In this case g is called the inverse of f.

Proposition 4.4. *The inverse of an isomorphism is unique.*

Proof. Suppose $g, h : Y \to X$ are both inverses of $f : X \to Y$. Then

$$g = g \circ \mathrm{id}_Y = g \circ (f \circ h) = (g \circ f) \circ h = \mathrm{id}_X \circ h = h.$$

Hence the inverse is unique. □

Definition 4.5. Let $\mathcal{C}$ and $\mathcal{D}$ be categories. A covariant functor $F : \mathcal{C} \to \mathcal{D}$ assigns to every object X of $\mathcal{C}$ an object $F(X)$ of $\mathcal{D}$, and to every morphism $f : X \to Y$ a morphism $F(f) : F(X) \to F(Y)$, such that

$$F(\mathrm{id}_X) = \mathrm{id}_{F(X)}, \qquad F(g \circ f) = F(g) \circ F(f).$$

A contravariant functor from $\mathcal{C}$ to $\mathcal{D}$ is a covariant functor $\mathcal{C}^{\mathrm{op}} \to \mathcal{D}$, where $\mathcal{C}^{\mathrm{op}}$ is the opposite category.

Example 4.6. The dual construction is a contravariant functor

$$(-)^* : \mathbf{Vec}_{\mathbb{F}} \to \mathbf{Vec}_{\mathbb{F}}.$$

It sends a vector space V to V^* and a linear map $T : V \to W$ to $T^* : W^* \to V^*$. The identities

$$(T \circ S)^* = S^* \circ T^*, \qquad \mathrm{id}_V^* = \mathrm{id}_{V^*}$$

are precisely the functoriality axioms.

Definition 4.7. Let $\mathcal{C}$ be a category. An object I is called initial if for every object X there exists exactly one morphism $I \to X$. An object T is called terminal if for every object X there exists exactly one morphism $X \to T$.

Proposition 4.8 (Uniqueness of initial and terminal objects)**.** *Initial objects are unique up to unique isomorphism. Terminal objects are also unique up to unique isomorphism.*

Proof. Let I and I' be initial objects. Since I is initial, there is a unique morphism $f : I \to I'$. Since I' is initial, there is a unique morphism $g : I' \to I$. The composite $g \circ f : I \to I$ is a morphism from I to itself. But id_I is also such a morphism. By the uniqueness in the definition of initial object, $g \circ f = \mathrm{id}_I$. Similarly $f \circ g = \mathrm{id}_{I'}$. Thus f is an isomorphism.

If $h : I \to I'$ is any other isomorphism compatible with the initial structure, it is in particular a morphism from I to I', hence by uniqueness $h = f$. Therefore the isomorphism is unique.

The proof for terminal objects is dual: if T and T' are terminal, there are unique morphisms $T \to T'$ and $T' \to T$, and their composites must be the identity morphisms by terminal uniqueness. □

Definition 4.9. A universal property characterizes an object by a mapping property. More precisely, one often constructs a category of possible solutions to a problem; a universal solution is an initial or terminal object in that category.

Remark 4.10. In practice, this means that an object is not characterized by its elements or by a coordinate description, but by how maps into or out of it correspond to simpler data. Quotient spaces, direct sums, direct products, tensor products, kernels, and cokernels are all examples of this principle.

4.2 Quotients

Theorem 4.11 (Universal property of quotient spaces)**.** *Let $W \subseteq V$ be a subspace and let $\pi : V \to V/W$ be the quotient map. If $T : V \to U$ is linear and $W \subseteq \ker T$, then there exists a unique linear map*

$$\overline{T} : V/W \to U$$

such that

$$T = \overline{T} \circ \pi.$$

Proof. Define $\overline{T}(v + W) = T(v)$. If $v + W = v' + W$, then $v - v' \in W \subseteq \ker T$, so $T(v - v') = 0$ and $T(v) = T(v')$. Thus $\overline{T}$ is well-defined. It is linear because

$$\overline{T}(a(v + W) + b(v' + W)) = \overline{T}((av + bv') + W) = T(av + bv')$$

$$= aT(v) + bT(v') = a\overline{T}(v + W) + b\overline{T}(v' + W).$$

For every $v \in V$,

$$(\overline{T} \circ \pi)(v) = \overline{T}(v + W) = T(v),$$

so $T = \overline{T} \circ \pi$.

For uniqueness, suppose $S : V/W \to U$ is linear and $T = S \circ \pi$. Then for every coset $v + W$,

$$S(v + W) = S(\pi(v)) = T(v) = \overline{T}(v + W).$$

Thus $S = \overline{T}$. □

Remark 4.12. Categorically, V/W is an initial object in the category whose objects are pairs (U, T) with $T : V \to U$ linear and $W \subseteq \ker T$, and whose morphisms $(U, T) \to (U', T')$ are linear maps $a : U \to U'$ such that $T' = a \circ T$.

4.3 Direct Sums as Coproducts

Theorem 4.13 (Universal property of direct sums)**.** *Let $\{V_i\}_{i \in I}$ be a family of vector spaces and let $\iota_i : V_i \to \bigoplus_{i \in I} V_i$ be the canonical inclusion. For every vector space W and every family of linear maps $f_i : V_i \to W$, there exists a unique linear map*

$$f : \bigoplus_{i \in I} V_i \to W$$

such that $f \circ \iota_i = f_i$ for all i.

Proof. For $v = (v_i)_{i \in I} \in \bigoplus_i V_i$, only finitely many v_i are nonzero. Define

$$f(v) = \sum_{i \in I} f_i(v_i),$$

where the sum is finite because v has finite support. This map is linear by componentwise addition and scalar multiplication. For $v_i \in V_i$, the element $\iota_i(v_i)$ has v_i in the i-th component and zero elsewhere, so

$$f(\iota_i(v_i)) = f_i(v_i).$$

Thus $f \circ \iota_i = f_i$.

If $g : \bigoplus_i V_i \to W$ is another linear map with $g \circ \iota_i = f_i$, then every $v \in \bigoplus_i V_i$ can be written as a finite sum

$$v = \sum_{i \in I} \iota_i(v_i).$$

Therefore

$$g(v) = \sum_i g(\iota_i(v_i)) = \sum_i f_i(v_i) = f(v).$$

Hence $g = f$, proving uniqueness. □

Corollary 4.14. *There is a natural isomorphism*

$$\mathrm{Hom}_{\mathbb{F}}\left(\bigoplus_{i \in I} V_i, W\right) \cong \prod_{i \in I} \mathrm{Hom}_{\mathbb{F}}(V_i, W).$$

Proof. Define

$$\Phi : \mathrm{Hom}_{\mathbb{F}}\left(\bigoplus_i V_i, W\right) \to \prod_i \mathrm{Hom}_{\mathbb{F}}(V_i, W)$$

by $\Phi(f) = (f \circ \iota_i)_i$. The universal property says precisely that for every family $(f_i)_i$ there is a unique f with $f \circ \iota_i = f_i$. Hence Φ is bijective. It is linear because composition with ι_i is linear in f for each i. Therefore Φ is an isomorphism. □

4.4 Direct Products as Products

Theorem 4.15 (Universal property of direct products)**.** *Let $\{V_i\}_{i \in I}$ be a family of vector spaces and let $\pi_i : \prod_{i \in I} V_i \to V_i$ be the canonical projection. For every vector space W and every family of linear maps $g_i : W \to V_i$, there exists a unique linear map*

$$g : W \to \prod_{i \in I} V_i$$

such that $\pi_i \circ g = g_i$ for all i.

Proof. Define

$$g(w) = (g_i(w))_{i \in I}.$$

This is an element of the product. The map g is linear because for $a, b \in \mathbb{F}$ and $w, w' \in W$,

$$g(aw + bw') = (g_i(aw + bw'))_i = (ag_i(w) + bg_i(w'))_i = ag(w) + bg(w').$$

Clearly $\pi_i(g(w)) = g_i(w)$, so $\pi_i \circ g = g_i$.

If $h : W \to \prod_i V_i$ also satisfies $\pi_i \circ h = g_i$ for all i, then for every $w \in W$, the i-th component of $h(w)$ equals $g_i(w)$, which is the i-th component of $g(w)$. Hence $h(w) = g(w)$ for all w, so $h = g$. □

Corollary 4.16. *There is a natural isomorphism*

$$\mathrm{Hom}_{\mathbb{F}}\left(W, \prod_{i \in I} V_i\right) \cong \prod_{i \in I} \mathrm{Hom}_{\mathbb{F}}(W, V_i).$$

Proof. The map sends $g : W \to \prod_i V_i$ to $(\pi_i \circ g)_i$. By the universal property of the product, every family $(g_i)_i$ arises uniquely in this way. The map is linear componentwise. Hence it is a natural isomorphism. □

4.5 Pullbacks and Pushouts

Definition 4.17. Given linear maps $f : X \to Z$ and $g : Y \to Z$, the pullback is

$$X \times_Z Y = \{(x, y) \in X \oplus Y : f(x) = g(y)\}.$$

Proposition 4.18 (Universal property of pullbacks)**.** *Let $p_X : X \times_Z Y \to X$ and $p_Y : X \times_Z Y \to Y$ be the projections. If A is a vector space with maps $a_X : A \to X$ and $a_Y : A \to Y$ satisfying*

$$f \circ a_X = g \circ a_Y,$$

then there exists a unique linear map $u : A \to X \times_Z Y$ such that

$$p_X \circ u = a_X, \qquad p_Y \circ u = a_Y.$$

Proof. Define $u(a) = (a_X(a), a_Y(a))$. The compatibility condition gives

$$f(a_X(a)) = g(a_Y(a)),$$

so $u(a) \in X \times_Z Y$. The map is linear because a_X and a_Y are linear. It satisfies the desired projection identities by definition.

If u' is another such map, then for every $a \in A$, the first component of $u'(a)$ is $a_X(a)$ and the second component is $a_Y(a)$. Hence $u'(a) = u(a)$ for all a, so $u' = u$. □

Definition 4.19. Given linear maps $f : Z \to X$ and $g : Z \to Y$, the pushout is

$$X \amalg_Z Y = (X \oplus Y)/N,$$

where

$$N = \operatorname{span}\{(f(z), -g(z)) : z \in Z\}.$$

Proposition 4.20 (Universal property of pushouts)**.** *Let $j_X : X \to X \amalg_Z Y$ and $j_Y : Y \to X \amalg_Z Y$ be induced by the inclusions into $X \oplus Y$. If A is a vector space with maps $a_X : X \to A$ and $a_Y : Y \to A$ satisfying*

$$a_X \circ f = a_Y \circ g,$$

then there exists a unique linear map $u : X \amalg_Z Y \to A$ such that

$$u \circ j_X = a_X, \qquad u \circ j_Y = a_Y.$$

Proof. Define $h : X \oplus Y \to A$ by

$$h(x, y) = a_X(x) + a_Y(y).$$

For every $z \in Z$,

$$h(f(z), -g(z)) = a_X(f(z)) - a_Y(g(z)) = 0.$$

Thus $N \subseteq \ker h$. By the universal property of the quotient, h factors through a unique linear map

$$u : (X \oplus Y)/N \to A$$

such that $u((x, y) + N) = h(x, y)$. Then $u \circ j_X = a_X$ and $u \circ j_Y = a_Y$.

If u' is another such map, then u' and u agree on all elements $j_X(x)$ and $j_Y(y)$. These elements span the pushout because every coset of (x, y) equals $j_X(x) + j_Y(y)$. Therefore $u' = u$. □

4.6 Exact Sequences

Definition 4.21. A sequence of vector spaces and linear maps

$$\cdots \longrightarrow V_{i-1} \xrightarrow{d_{i-1}} V_i \xrightarrow{d_i} V_{i+1} \longrightarrow \cdots$$

is exact at V_i if $\operatorname{im} d_{i-1} = \ker d_i$. It is exact if it is exact at every term.

Proposition 4.22. *A sequence*

$$0 \longrightarrow U \xrightarrow{f} V \xrightarrow{g} W \longrightarrow 0$$

is exact if and only if f is injective, g is surjective, and $\operatorname{im} f = \ker g$.

Proof. Exactness at U says that $\ker f = \operatorname{im}(0 \to U) = 0$, so f is injective. Exactness at V says $\operatorname{im} f = \ker g$. Exactness at W says $\operatorname{im} g = \ker(W \to 0) = W$, so g is surjective. Conversely, these three conditions are exactly the exactness conditions at U, V, W. □

Proposition 4.23. *A short exact sequence*

$$0 \longrightarrow U \xrightarrow{f} V \xrightarrow{g} W \longrightarrow 0$$

splits if and only if there exists a linear map $s : W \to V$ such that $g \circ s = \mathrm{id}_W$. In that case

$$V \cong U \oplus W.$$

Proof. Suppose the sequence splits, meaning $V \cong U \oplus W$ with f identified with inclusion into the first factor and g with projection onto the second factor. Then the inclusion $s : W \to U \oplus W$ into the second factor satisfies $g \circ s = \mathrm{id}_W$.

Conversely, suppose $s : W \to V$ satisfies $g \circ s = \mathrm{id}_W$. Define

$$\Phi : U \oplus W \to V, \qquad \Phi(u, w) = f(u) + s(w).$$

We show that Φ is an isomorphism. If $v \in V$, set $w = g(v)$. Then

$$g(v - s(w)) = g(v) - g(s(w)) = w - w = 0.$$

Thus $v - s(w) \in \ker g = \operatorname{im} f$, so $v - s(w) = f(u)$ for some $u \in U$. Hence $v = \Phi(u, w)$, proving surjectivity.

If $\Phi(u, w) = 0$, then applying g gives

$$0 = g(f(u) + s(w)) = 0 + w = w.$$

Thus $w = 0$, and then $f(u) = 0$. Since f is injective, $u = 0$. Hence Φ is injective. Therefore Φ is an isomorphism. □

Chapter 5

Tensor Products, Exterior Powers, and Determinants

5.1 Tensor Products

Definition 5.1. Let V, W, U be vector spaces. A map $b : V \times W \to U$ is bilinear if it is linear in each variable separately. The vector space of bilinear maps is denoted by $\mathrm{Bil}(V, W; U)$.

Theorem 5.2 (Construction of tensor products)**.** *For vector spaces V and W, there exists a vector space $V \otimes W$ and a bilinear map*

$$\tau : V \times W \to V \otimes W, \qquad (v, w) \mapsto v \otimes w,$$

such that for every vector space U, every bilinear map $b : V \times W \to U$ factors uniquely through a linear map $\widetilde{b} : V \otimes W \to U$:

$$b = \widetilde{b} \circ \tau.$$

Proof. Let $F(V \times W)$ be the free vector space with basis symbols $[v, w]$ for $(v, w) \in V \times W$. Let N be the subspace generated by all elements

$$[v + v', w] - [v, w] - [v', w],$$

$$[av, w] - a[v, w],$$

$$[v, w + w'] - [v, w] - [v, w'],$$

$$[v, aw] - a[v, w],$$

where $v, v' \in V$, $w, w' \in W$, and $a \in \mathbb{F}$. Define

$$V \otimes W = F(V \times W)/N$$

and let $v \otimes w$ be the coset of $[v, w]$. The defining relations imply that $(v, w) \mapsto v \otimes w$ is bilinear.

Let $b : V \times W \to U$ be bilinear. There is a unique linear map $\widehat{b} : F(V \times W) \to U$ with $\widehat{b}([v, w]) = b(v, w)$. Since b is bilinear, $\widehat{b}$ vanishes on all generators of N. Hence $N \subseteq \ker \widehat{b}$. By the universal property of quotient spaces, $\widehat{b}$ factors uniquely through a linear map

$$\widetilde{b} : V \otimes W \to U$$

with $\widetilde{b}(v \otimes w) = b(v, w)$. The uniqueness follows because the pure tensors $v \otimes w$ span $V \otimes W$. $\square$

Corollary 5.3. *For every vector space U, there is a natural isomorphism*

$$\mathrm{Hom}_{\mathbb{F}}(V \otimes W, U) \cong \mathrm{Bil}(V, W; U).$$

Proof. The map sends a linear map $L : V \otimes W \to U$ to the bilinear map $(v, w) \mapsto L(v \otimes w)$. The universal property of the tensor product says that every bilinear map arises uniquely in this way. Linearity of the correspondence is immediate from pointwise addition and scalar multiplication. Therefore it is a natural vector space isomorphism. □

5.2 Basic Isomorphisms

Proposition 5.4. *There is a canonical isomorphism*

$$V \otimes W \cong W \otimes V$$

given by $v \otimes w \mapsto w \otimes v$.

Proof. The map $V \times W \to W \otimes V$ defined by $(v, w) \mapsto w \otimes v$ is bilinear. By the universal property of $V \otimes W$, it induces a unique linear map

$$\Phi : V \otimes W \to W \otimes V$$

with $\Phi(v \otimes w) = w \otimes v$. Similarly, the bilinear map $W \times V \to V \otimes W$ given by $(w, v) \mapsto v \otimes w$ induces a linear map

$$\Psi : W \otimes V \to V \otimes W$$

with $\Psi(w \otimes v) = v \otimes w$. Then $\Psi \circ \Phi$ and $\mathrm{id}_{V \otimes W}$ agree on all pure tensors $v \otimes w$, which span $V \otimes W$. Hence $\Psi \circ \Phi = \mathrm{id}$. Similarly $\Phi \circ \Psi = \mathrm{id}$. Thus Φ is an isomorphism. □

Proposition 5.5. *There is a canonical isomorphism*

$$(U \otimes V) \otimes W \cong U \otimes (V \otimes W)$$

given on pure tensors by

$$(u \otimes v) \otimes w \longmapsto u \otimes (v \otimes w).$$

Proof. For fixed W', linear maps $(U \otimes V) \otimes W \to W'$ correspond to bilinear maps $(U \otimes V) \times W \to W'$. Such maps correspond to functions trilinear in $u \in U$, $v \in V$, and $w \in W$, because a bilinear map out of $U \otimes V$ is the same as a bilinear map out of $U \times V$. Similarly, linear maps $U \otimes (V \otimes W) \to W'$ also correspond to trilinear maps $U \times V \times W \to W'$.

Both tensor products therefore represent the same functor of trilinear maps. By uniqueness of universal objects, they are canonically isomorphic. Explicitly, the trilinear map $(u, v, w) \mapsto u \otimes (v \otimes w)$ induces a map $(U \otimes V) \otimes W \to U \otimes (V \otimes W)$, and the trilinear map $(u, v, w) \mapsto (u \otimes v) \otimes w$ induces its inverse. The two composites agree with the identity on pure tensors, hence are identities. □

Proposition 5.6. *There is a canonical isomorphism*

$$(V \oplus W) \otimes U \cong (V \otimes U) \oplus (W \otimes U).$$

Proof. Define a bilinear map $(V \oplus W) \times U \to (V \otimes U) \oplus (W \otimes U)$ by

$$((v, w), u) \mapsto (v \otimes u, w \otimes u).$$

By the universal property of tensor products, it induces a linear map

$$\Phi : (V \oplus W) \otimes U \to (V \otimes U) \oplus (W \otimes U).$$

Conversely, the bilinear maps $V \times U \to (V \oplus W) \otimes U$ and $W \times U \to (V \oplus W) \otimes U$ given by

$$(v, u) \mapsto (v, 0) \otimes u, \qquad (w, u) \mapsto (0, w) \otimes u$$

induce linear maps $V \otimes U \to (V \oplus W) \otimes U$ and $W \otimes U \to (V \oplus W) \otimes U$. By the universal property of direct sums, these give a linear map

$$\Psi : (V \otimes U) \oplus (W \otimes U) \to (V \oplus W) \otimes U.$$

The composite $\Phi \circ \Psi$ is the identity on pure tensors in each summand, and $\Psi \circ \Phi$ is the identity on pure tensors $(v, w) \otimes u$. Since these elements span the relevant tensor products, the two composites are identities. □

5.3 Bases of Tensor Products

Theorem 5.7. *Let $\{v_i\}_{i \in I}$ be a basis of V and $\{w_j\}_{j \in J}$ a basis of W. Then*

$$\{v_i \otimes w_j : (i, j) \in I \times J\}$$

is a basis of $V \otimes W$.

Proof. The pure tensors span $V \otimes W$. Since every $v \in V$ and $w \in W$ can be written as finite linear combinations of the v_i and w_j, bilinearity gives

$$v \otimes w = \left(\sum_i a_i v_i \right) \otimes \left(\sum_j b_j w_j \right) = \sum_{i,j} a_i b_j (v_i \otimes w_j).$$

Thus the tensors $v_i \otimes w_j$ span.

To prove linear independence, suppose

$$\sum_{k=1}^{n} c_k (v_{i_k} \otimes w_{j_k}) = 0$$

with distinct pairs (i_k, j_k). For each pair (i, j), let v_i^* and w_j^* be the coordinate functionals on V and W associated to the bases. The map

$$b_{ij} : V \times W \to \mathbb{F}, \qquad b_{ij}(v, w) = v_i^*(v) w_j^*(w)$$

is bilinear. Hence it induces a linear map $\widetilde{b}_{ij} : V \otimes W \to \mathbb{F}$ satisfying

$$\widetilde{b}_{ij}(v_{i'} \otimes w_{j'}) = \delta_{ii'} \delta_{jj'}.$$

Applying $\widetilde{b}_{i_k j_k}$ to the relation gives $c_k = 0$. Thus all coefficients are zero. Therefore the tensors form a basis. □

Corollary 5.8. *If V and W are finite-dimensional, then*

$$\dim(V \otimes W) = \dim V \cdot \dim W.$$

Proof. If $\dim V = m$ and $\dim W = n$, choose bases with m and n elements. By the theorem, $V \otimes W$ has a basis indexed by pairs of basis elements, hence has mn basis elements. □

5.4 Tensor Products of Linear Maps

Proposition 5.9. *Let $S : V \to V'$ and $T : W \to W'$ be linear. There is a unique linear map*

$$S \otimes T : V \otimes W \to V' \otimes W'$$

satisfying

$$(S \otimes T)(v \otimes w) = S(v) \otimes T(w).$$

Proof. The map $V \times W \to V' \otimes W'$ defined by

$$(v, w) \mapsto S(v) \otimes T(w)$$

is bilinear because S, T, and the tensor map are linear in each variable. By the universal property of $V \otimes W$, it induces a unique linear map $S \otimes T$ satisfying the displayed formula. □

Proposition 5.10. *Tensor products of linear maps are functorial: if $S_1 : V \to V'$, $S_2 : V' \to V''$, $T_1 : W \to W'$, and $T_2 : W' \to W''$, then*

$$(S_2 \circ S_1) \otimes (T_2 \circ T_1) = (S_2 \otimes T_2) \circ (S_1 \otimes T_1).$$

Moreover $\mathrm{id}_V \otimes \mathrm{id}_W = \mathrm{id}_{V \otimes W}$.

Proof. Both sides send a pure tensor $v \otimes w$ to

$$S_2(S_1(v)) \otimes T_2(T_1(w)).$$

Since pure tensors span $V \otimes W$, the two maps are equal. Similarly, $\mathrm{id}_V \otimes \mathrm{id}_W$ sends $v \otimes w$ to $v \otimes w$ for every pure tensor, hence equals the identity. □

5.5 Exterior Powers

Definition 5.11. A multilinear map $a : V^k \to U$ is alternating if $a(v_1, \ldots, v_k) = 0$ whenever $v_i = v_j$ for some $i \neq j$. The vector space of alternating k-linear maps from V^k to U is denoted by $\mathrm{Alt}^k(V, U)$.

Definition 5.12. The k-th exterior power $\Lambda^k V$ is the quotient of $V^{\otimes k}$ by the subspace generated by tensors

$$v_1 \otimes \cdots \otimes v_k$$

for which $v_i = v_j$ for some $i \neq j$. The image of $v_1 \otimes \cdots \otimes v_k$ is denoted by

$$v_1 \wedge \cdots \wedge v_k.$$

Theorem 5.13 (Universal property of exterior powers)**.** *For every vector space U, there is a natural isomorphism*

$$\mathrm{Hom}_{\mathbb{F}}(\Lambda^k V, U) \cong \mathrm{Alt}^k(V, U).$$

Proof. Let $q : V^{\otimes k} \to \Lambda^k V$ be the quotient map. The map

$$V^k \to \Lambda^k V, \qquad (v_1, \ldots, v_k) \mapsto v_1 \wedge \cdots \wedge v_k$$

is multilinear and alternating by construction.

Given a linear map $L : \Lambda^k V \to U$, the composite

$$V^k \to \Lambda^k V \xrightarrow{L} U$$

is alternating and k-linear. This gives a map

$$\mathrm{Hom}_{\mathbb{F}}(\Lambda^k V, U) \to \mathrm{Alt}^k(V, U).$$

Conversely, let $a : V^k \to U$ be alternating and multilinear. By repeated use of the universal property of tensor products, a induces a unique linear map

$$\widehat{a} : V^{\otimes k} \to U$$

with $\widehat{a}(v_1 \otimes \cdots \otimes v_k) = a(v_1, \ldots, v_k)$. Since a is alternating, $\widehat{a}$ vanishes on every generator of the subspace used to define $\Lambda^k V$. Therefore $\widehat{a}$ factors uniquely through a linear map

$$\widetilde{a} : \Lambda^k V \to U.$$

The two constructions are inverse to one another by their defining formulas on wedges. Hence the claimed natural isomorphism holds. □

Proposition 5.14. *For vectors $v_1, \ldots, v_k \in V$, if two adjacent vectors are interchanged, then*

$$v_1 \wedge \cdots \wedge v_i \wedge v_{i+1} \wedge \cdots \wedge v_k = -v_1 \wedge \cdots \wedge v_{i+1} \wedge v_i \wedge \cdots \wedge v_k.$$

Consequently, if $\mathrm{char}\,\mathbb{F} \neq 2$*, exchanging any two arguments changes the sign.*

Proof. In $\Lambda^k V$, the wedge with two equal adjacent vectors is zero. Hence

$$0 = v_1 \wedge \cdots \wedge (v_i + v_{i+1}) \wedge (v_i + v_{i+1}) \wedge \cdots \wedge v_k.$$

Expanding by multilinearity gives four terms. The terms with $v_i \wedge v_i$ and $v_{i+1} \wedge v_{i+1}$ vanish. Thus

$$v_1 \wedge \cdots \wedge v_i \wedge v_{i+1} \wedge \cdots \wedge v_k + v_1 \wedge \cdots \wedge v_{i+1} \wedge v_i \wedge \cdots \wedge v_k = 0,$$

which is the desired identity. Any transposition can be written as a product of adjacent transpositions, so the last statement follows. □

Theorem 5.15. *Let V have basis $v_1, \ldots, v_n$. Then the elements*

$$v_{i_1} \wedge \cdots \wedge v_{i_k}, \qquad 1 \leq i_1 < \cdots < i_k \leq n,$$

form a basis of $\Lambda^k V$. In particular,

$$\dim \Lambda^k V = \binom{n}{k}.$$

Proof. The tensors $v_{i_1} \otimes \cdots \otimes v_{i_k}$ span $V^{\otimes k}$, hence their images span $\Lambda^k V$. If two indices are equal, the corresponding wedge is zero. If all indices are distinct, repeated adjacent swaps rewrite the wedge as a sign times one with strictly increasing indices. Thus the displayed wedges span $\Lambda^k V$.

To prove linear independence, let $v_1^*, \ldots, v_n^*$ be the dual basis. For each increasing k-tuple $I = (i_1 < \cdots < i_k)$, define an alternating k-linear form $\omega_I : V^k \to \mathbb{F}$ by

$$\omega_I(x_1, \ldots, x_k) = \det(v_{i_a}^*(x_b))_{1 \leq a, b \leq k}.$$

This map is alternating because a determinant with two equal columns is zero. By the universal property of $\Lambda^k V$, it induces a linear functional $\widetilde{\omega}_I : \Lambda^k V \to \mathbb{F}$.

For increasing $J = (j_1 < \cdots < j_k)$,

$$\widetilde{\omega}_I(v_{j_1} \wedge \cdots \wedge v_{j_k}) = \det(v_{i_a}^*(v_{j_b}))_{a,b}.$$

This determinant is 1 if $I = J$ and 0 otherwise, because the matrix has two equal zero patterns or a zero row when the sets differ. Therefore the functionals $\widetilde{\omega}_I$ pick out the coefficients of the displayed wedges. Any linear relation among those wedges has all coefficients zero. Hence they are linearly independent. The dimension formula follows by counting increasing k-tuples. □

5.6 Determinants

Definition 5.16. Let V be an n-dimensional vector space and let $T : V \to V$ be linear. Since $\Lambda^n V$ is one-dimensional, the map $\Lambda^n T : \Lambda^n V \to \Lambda^n V$ is multiplication by a unique scalar. This scalar is called the determinant of T and is denoted by $\det T$.

Proposition 5.17. *For linear endomorphisms S, T of a finite-dimensional vector space,*

$$\det(ST) = \det(S)\det(T).$$

Proof. Functoriality of exterior powers gives

$$\Lambda^n(ST) = (\Lambda^n S)(\Lambda^n T).$$

Since $\Lambda^n V$ is one-dimensional, $\Lambda^n S$ is multiplication by $\det S$ and $\Lambda^n T$ is multiplication by $\det T$. Their composite is multiplication by $(\det S)(\det T)$. On the other hand, $\Lambda^n(ST)$ is multiplication by $\det(ST)$. Therefore $\det(ST) = \det(S)\det(T)$. □

Proposition 5.18. *Let $A = (a_{ij})$ be the matrix of T with respect to a basis $v_1, \ldots, v_n$. Then*

$$T(v_1) \wedge \cdots \wedge T(v_n) = \det(A)\, v_1 \wedge \cdots \wedge v_n,$$

where $\det(A)$ is the usual Leibniz determinant

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1),1} \cdots a_{\sigma(n),n}.$$

Proof. Write

$$T(v_j) = \sum_{i=1}^{n} a_{ij} v_i.$$

Using multilinearity of the wedge product,

$$T(v_1) \wedge \cdots \wedge T(v_n) = \sum_{i_1, \ldots, i_n} a_{i_1,1} \cdots a_{i_n,n} v_{i_1} \wedge \cdots \wedge v_{i_n}.$$

Every term with repeated indices vanishes. The remaining terms are indexed by permutations $\sigma \in S_n$, with $i_j = \sigma(j)$. For such a term,

$$v_{\sigma(1)} \wedge \cdots \wedge v_{\sigma(n)} = \operatorname{sgn}(\sigma) v_1 \wedge \cdots \wedge v_n.$$

Therefore

$$T(v_1) \wedge \cdots \wedge T(v_n) = \left(\sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1),1} \cdots a_{\sigma(n),n} \right) v_1 \wedge \cdots \wedge v_n.$$

This is the claimed formula. □

Chapter 6

Modules over Rings and PIDs

6.1 Modules and Submodules

Definition 6.1. Let R be a commutative ring with identity. An R-module is an abelian group $(M, +)$ equipped with a scalar multiplication $R \times M \to M$ satisfying the same formal axioms as scalar multiplication in a vector space.

Definition 6.2. A subset $N \subseteq M$ is an R-submodule if it is a subgroup of M and $rn \in N$ for every $r \in R$ and $n \in N$.

Proposition 6.3. *If $S \subseteq M$, then*

$$RS = \{r_1 s_1 + \cdots + r_n s_n : r_i \in R,\ s_i \in S\}$$

is the smallest submodule of M containing S.

Proof. The proof is identical to the proof for spans of vector spaces. The set RS is closed under addition and under multiplication by elements of R, hence is a submodule. It contains S because $s = 1s$. Any submodule containing S is closed under finite R-linear combinations, hence contains RS. Therefore RS is the smallest submodule containing S. □

Definition 6.4. An R-module M is finitely generated if there exists a finite subset $S \subseteq M$ such that $M = RS$.

6.2 Free Modules and Presentations

Definition 6.5. An R-module F is free with basis B if every element of F can be written uniquely as a finite R-linear combination of elements of B.

Proposition 6.6. *Let B be a set. The direct sum*

$$R^{(B)} = \bigoplus_{b \in B} R$$

is a free R-module with basis given by the standard elements e_b.

Proof. An element of $R^{(B)}$ is a family $(r_b)_{b \in B}$ with finite support. It can be written as the finite sum

$$\sum_{b \in B} r_b e_b,$$

where only finitely many r_b are nonzero. This proves spanning. If

$$\sum_{b \in B} r_b e_b = 0,$$

then every coordinate of the left-hand side is zero, so every $r_b = 0$. Hence the expression is unique. Therefore $R^{(B)}$ is free with basis $\{e_b\}_{b \in B}$. □

Proposition 6.7. *An R-module M is finitely generated if and only if there exists a surjective homomorphism*

$$R^n \twoheadrightarrow M$$

for some $n \geq 0$.

Proof. If $R^n \twoheadrightarrow M$ is surjective, the images of the standard basis vectors generate M. Conversely, if M is generated by $m_1, \ldots, m_n$, define

$$\varphi : R^n \to M, \qquad (r_1, \ldots, r_n) \mapsto r_1 m_1 + \cdots + r_n m_n.$$

This is a module homomorphism and is surjective by the assumption that the m_i generate M. □

Definition 6.8. A presentation of an R-module M is an exact sequence

$$R^m \longrightarrow R^n \longrightarrow M \longrightarrow 0.$$

6.3 Quotients and Isomorphism Theorems

Definition 6.9. If $N \subseteq M$ is a submodule, the quotient module M/N is the quotient abelian group with scalar multiplication

$$r(m + N) = rm + N.$$

Proposition 6.10. *The scalar multiplication on M/N is well-defined and makes M/N an R-module.*

Proof. If $m + N = m' + N$, then $m - m' \in N$. Since N is a submodule, $r(m - m') \in N$, so $rm + N = rm' + N$. Thus scalar multiplication is well-defined. The module axioms follow immediately from the corresponding axioms in M. □

Theorem 6.11 (First isomorphism theorem for modules)**.** *Let $\varphi : M \to N$ be an R-module homomorphism. Then*

$$M/\ker\varphi \cong \operatorname{im}\varphi.$$

Proof. Define $\overline{\varphi} : M/\ker\varphi \to \operatorname{im}\varphi$ by

$$\overline{\varphi}(m + \ker\varphi) = \varphi(m).$$

The same proof as in the vector space case shows that the map is well-defined, R-linear, surjective, and injective. Therefore it is an isomorphism. □

6.4 Torsion and Annihilators

Definition 6.12. Let R be a domain and let M be an R-module. The torsion submodule of M is

$$M_{\text{tor}} = \{m \in M : rm = 0 \text{ for some } 0 \neq r \in R\}.$$

The module M is torsion-free if $M_{\text{tor}} = 0$ and torsion if $M = M_{\text{tor}}$.

Proposition 6.13. *If R is a domain, then M_{tor} is a submodule of M.*

Proof. Let $m, n \in M_{\text{tor}}$. Then there exist nonzero $r, s \in R$ such that $rm = 0$ and $sn = 0$. Since R is a domain, $rs \neq 0$, and

$$rs(m+n) = s(rm) + r(sn) = 0.$$

Thus $m + n$ is torsion. If $a \in R$, then $r(am) = a(rm) = 0$, so am is torsion. Hence M_{tor} is a submodule. □

Definition 6.14. For an R-module M and an element $m \in M$, the annihilator of m is

$$\text{Ann}_R(m) = \{r \in R : rm = 0\}.$$

The annihilator of M is

$$\text{Ann}_R(M) = \{r \in R : rM = 0\}.$$

Proposition 6.15. *For every $m \in M$, the cyclic submodule Rm is isomorphic to $R/\operatorname{Ann}_R(m)$.*

Proof. Define $\varphi : R \to Rm$ by $\varphi(r) = rm$. This map is an R-module homomorphism and is surjective by definition of Rm. Its kernel is

$$\ker \varphi = \{r \in R : rm = 0\} = \text{Ann}_R(m).$$

By the first isomorphism theorem,

$$R/\operatorname{Ann}_R(m) \cong Rm.$$

□

6.5 Noetherian Modules

Definition 6.16. An R-module M is noetherian if every ascending chain of submodules

$$N_1 \subseteq N_2 \subseteq N_3 \subseteq \cdots$$

stabilizes.

Proposition 6.17. *An R-module M is noetherian if and only if every submodule of M is finitely generated.*

Proof. Suppose M is noetherian and let $N \subseteq M$ be a submodule. If N is not finitely generated, choose $x_1 \in N$. Since $Rx_1 \neq N$, choose $x_2 \in N \setminus Rx_1$. Continuing inductively, if $Rx_1 + \cdots + Rx_n \neq N$, choose x_{n+1} outside this submodule. This gives a strictly increasing chain

$$Rx_1 \subsetneq Rx_1 + Rx_2 \subsetneq Rx_1 + Rx_2 + Rx_3 \subsetneq \cdots,$$

contradicting the noetherian condition. Hence N is finitely generated.

Conversely, suppose every submodule of M is finitely generated. Let

$$N_1 \subseteq N_2 \subseteq \cdots$$

be an ascending chain and set $N = \bigcup_i N_i$. Then N is a submodule. By assumption, N is generated by finitely many elements $x_1, \ldots, x_r$. Each x_j lies in some N_{i_j}. Let N_N be one of the chain terms containing all x_j, say N_m with $m \geq i_j$ for all j. Then $N \subseteq N_m$, and since $N_m \subseteq N$, we have $N = N_m$. Hence $N_i = N_m$ for all $i \geq m$, so the chain stabilizes. □

Proposition 6.18. *Every principal ideal domain is noetherian.*

Proof. Let R be a PID. Every ideal of R is principal, hence finitely generated. Applying the preceding proposition to R as a module over itself, R is noetherian. □

Proposition 6.19. *If R is noetherian, then every finitely generated R-module is noetherian.*

Proof. First, R^n is noetherian. This follows by induction on n. The case $n = 1$ is the assumption that R is noetherian as a module over itself. For the induction step, use the exact sequence

$$0 \to R^{n-1} \to R^n \to R \to 0.$$

A standard argument shows that if A and C are noetherian in a short exact sequence $0 \to A \to B \to C \to 0$, then B is noetherian: any ascending chain in B gives ascending chains in A by intersection and in C by image, both of which stabilize; after that the original chain stabilizes.

Now let M be finitely generated. There is a surjection $R^n \twoheadrightarrow M$. A quotient of a noetherian module is noetherian, because submodules of the quotient correspond to submodules of R^n containing the kernel, and ascending chains lift to ascending chains. Hence M is noetherian. □

6.6 Modules over a PID

Lemma 6.20. *Let R be a PID and let N be a submodule of a finite free module R^n. Then N is free of rank at most n.*

Proof. We prove this by induction on n. The case $n = 0$ is clear. Let $\pi : R^n \to R$ be projection onto the first coordinate. Then $\pi(N)$ is an ideal of R, so $\pi(N) = (a)$ for some $a \in R$. If $a = 0$, then $N \subseteq 0 \oplus R^{n-1}$, and the result follows by induction.

If $a \neq 0$, choose $x \in N$ with $\pi(x) = a$. Let

$$K = N \cap (0 \oplus R^{n-1}).$$

By induction, K is free of rank at most $n - 1$. We claim that

$$N = Rx \oplus K.$$

For any $y \in N$, $\pi(y) \in (a)$, so $\pi(y) = ra$ for some $r \in R$. Then $y - rx \in K$, so $y \in Rx + K$. If $rx \in K$, then $0 = \pi(rx) = ra$. Since R is a domain and $a \neq 0$, $r = 0$. Hence $Rx \cap K = 0$. Thus $N \cong R \oplus K$ is free of rank at most n. □

Proposition 6.21. *Let R be a PID. Every finitely generated torsion-free R-module is free.*

Proof. Let M be finitely generated and torsion-free, generated by $m_1, \dots, m_n$. There is a surjection $R^n \twoheadrightarrow M$. We show that M embeds into a finite-dimensional vector space over the fraction field $K = \mathrm{Frac}(R)$.

Define

$$\iota : M \to K \otimes_R M, \qquad m \mapsto 1 \otimes m.$$

If $1 \otimes m = 0$, then by the construction of localization there exists nonzero $r \in R$ such that $rm = 0$. Since M is torsion-free, $m = 0$. Hence ι is injective. The K-vector space $K \otimes_R M$ is generated by $1 \otimes m_1, \dots, 1 \otimes m_n$, hence is finite-dimensional. Choose a K-basis among these generators, say $1 \otimes u_1, \dots, 1 \otimes u_s$. Then the submodule $Ru_1 + \cdots + Ru_s$ is free. For each generator m_i, there exist coefficients $\lambda_{ij} \in K$ with

$$1 \otimes m_i = \sum_j \lambda_{ij}(1 \otimes u_j).$$

Choose a nonzero common denominator $d \in R$ for all λ_{ij}. Then $dm_i \in Ru_1 + \cdots + Ru_s$ for all i. Hence dM is a submodule of the free module $Ru_1 + \cdots + Ru_s \cong R^s$, so dM is free by the preceding lemma. Since M is torsion-free, multiplication by d gives an isomorphism $M \cong dM$. Therefore M is free. □

Proposition 6.22. *Let R be a PID and let M be a finitely generated R-module. Then*

$$M \cong M_{\mathrm{tor}} \oplus R^r$$

for some $r \geq 0$.

Proof. The quotient M/M_{tor} is finitely generated and torsion-free. Indeed, if $0 \neq r \in R$ and $r(m + M_{\mathrm{tor}}) = 0$, then $rm \in M_{\mathrm{tor}}$, so there exists $0 \neq s \in R$ with $srm = 0$. Since $sr \neq 0$, $m \in M_{\mathrm{tor}}$, hence $m + M_{\mathrm{tor}} = 0$.

By the preceding proposition, M/M_{tor} is free, say isomorphic to R^r. Choose elements $m_1, \dots, m_r \in M$ whose images form a basis of this quotient. The map $R^r \to M$ sending the standard basis to the m_i gives a section of the quotient map $M \to M/M_{\mathrm{tor}}$. Hence the short exact sequence

$$0 \to M_{\mathrm{tor}} \to M \to M/M_{\mathrm{tor}} \to 0$$

splits, and $M \cong M_{\mathrm{tor}} \oplus R^r$. □

Chapter 7

Associated Primes and Primary Decomposition

The goal of this chapter is not to reproduce commutative algebra in full generality, but to isolate the part of primary decomposition that will be used for linear operators. We first introduce primary ideals and associated prime ideals in a general commutative ring, and then specialize to finitely generated modules over a principal ideal domain.

7.1 Primary Ideals and Associated Primes

Definition 7.1. Let A be a commutative ring. An ideal $q \subseteq A$ is called primary if $q \neq A$ and whenever $xy \in q$, either $x \in q$ or $y^n \in q$ for some $n > 0$.

Proposition 7.2. *An ideal $q \subseteq A$ is primary if and only if every zero divisor in A/q is nilpotent.*

Proof. Suppose q is primary. Let $\overline{y} \in A/q$ be a zero divisor. Then there exists $\overline{x} \neq 0$ in A/q such that $\overline{x}\,\overline{y} = 0$. This means that $xy \in q$ and $x \notin q$. Since q is primary, $y^n \in q$ for some $n > 0$, so $\overline{y}^n = 0$ in A/q. Hence $\overline{y}$ is nilpotent.

Conversely, suppose every zero divisor in A/q is nilpotent. If $xy \in q$ and $x \notin q$, then $\overline{x} \neq 0$ and $\overline{x}\,\overline{y} = 0$ in A/q. Therefore $\overline{y}$ is a zero divisor, hence nilpotent. Thus $y^n \in q$ for some $n > 0$. This proves that q is primary. □

Proposition 7.3. *If q is a primary ideal, then* $\operatorname{rad}(q)$ *is a prime ideal.*

Proof. Let $xy \in \operatorname{rad}(q)$. Then $(xy)^n \in q$ for some $n > 0$, so $x^n y^n \in q$. If $x^n \in q$, then $x \in \operatorname{rad}(q)$. If $x^n \notin q$, then the primary property applied to $x^n y^n \in q$ gives $(y^n)^m \in q$ for some $m > 0$. Hence $y \in \operatorname{rad}(q)$. Therefore $\operatorname{rad}(q)$ is prime. □

Definition 7.4. If q is primary and $\operatorname{rad}(q) = \mathfrak{p}$, then q is called $\mathfrak{p}$-primary.

Definition 7.5. Let M be an A-module. A prime ideal $\mathfrak{p} \subseteq A$ is called an associated prime of M if

$$\mathfrak{p} = \operatorname{Ann}_A(x)$$

for some nonzero element $x \in M$. The set of associated primes of M is denoted by $\operatorname{Ass}_A(M)$.

Proposition 7.6. *Let $I \subseteq A$ be an ideal. Then*

$$\operatorname{Ass}_A(A/I) = \{\mathfrak{p} \in \operatorname{Spec} A : \mathfrak{p} = (I : x) \text{ for some } x \in A\}.$$

Proof. An element of A/I has the form $\overline{x} = x + I$. Its annihilator is

$$\mathrm{Ann}_A(\overline{x}) = \{a \in A : a\overline{x} = 0 \text{ in } A/I\} = \{a \in A : ax \in I\} = (I : x).$$

If $\overline{x} \neq 0$, then $x \notin I$. Thus the associated primes of A/I are precisely the prime ideals that occur as $(I : x)$ for some $x \in A$ with $x \notin I$. Allowing $x \in I$ only gives the ideal A, which is not prime in the usual sense, so the displayed set is exactly $\mathrm{Ass}_A(A/I)$. □

Lemma 7.7. *Let q be a $\mathfrak{p}$-primary ideal. If $x \notin q$, then $(q : x)$ is $\mathfrak{p}$-primary.*

Proof. First $(q : x) \neq A$ because $x \notin q$. Suppose $ab \in (q : x)$ and $a \notin (q : x)$. Then $abx \in q$ and $ax \notin q$. Since q is primary, this implies $b^n \in q$ for some $n > 0$. Hence $b^n x \in q$, so $b^n \in (q : x)$. Thus $(q : x)$ is primary.

It remains to compute its radical. If $a \in \mathrm{rad}(q) = \mathfrak{p}$, then $a^n \in q$ for some $n > 0$, so $a^n x \in q$, and hence $a \in \mathrm{rad}(q : x)$. Thus $\mathfrak{p} \subseteq \mathrm{rad}(q : x)$. Conversely, if $a \in \mathrm{rad}(q : x)$, then $a^n x \in q$ for some $n > 0$. Since $x \notin q$ and q is primary, we get $(a^n)^m \in q$ for some $m > 0$, hence $a \in \mathrm{rad}(q) = \mathfrak{p}$. Therefore $\mathrm{rad}(q : x) = \mathfrak{p}$. □

Theorem 7.8 (Associated primes from primary decomposition)**.** *Let A be noetherian, and let*

$$I = q_1 \cap \cdots \cap q_r$$

be a minimal primary decomposition, where each q_i is $\mathfrak{p}_i$-primary and the primes $\mathfrak{p}_i$ are distinct. Then

$$\mathrm{Ass}_A(A/I) = \{\mathfrak{p}_1, \ldots, \mathfrak{p}_r\}.$$

Proof. First let $\mathfrak{p} \in \mathrm{Ass}_A(A/I)$. Then $\mathfrak{p} = (I : x)$ for some $x \in A$. Since

$$(I : x) = \bigcap_{i=1}^{r} (q_i : x),$$

we have

$$\prod_{i=1}^{r} (q_i : x) \subseteq (I : x) = \mathfrak{p}.$$

Because $\mathfrak{p}$ is prime, one of the ideals $(q_i : x)$ is contained in $\mathfrak{p}$. On the other hand, since $(I : x) = \bigcap_i (q_i : x)$, we have $(I : x) \subseteq (q_i : x)$. Hence

$$\mathfrak{p} \subseteq (q_i : x) \subseteq \mathfrak{p},$$

so $(q_i : x) = \mathfrak{p}$. Since $(q_i : x)$ is a proper ideal, $x \notin q_i$. By the preceding lemma, $(q_i : x)$ is $\mathfrak{p}_i$-primary. But it is also the prime ideal $\mathfrak{p}$, so its radical is $\mathfrak{p}$. Therefore $\mathfrak{p} = \mathfrak{p}_i$.

Conversely, fix i. Since the decomposition is irredundant, there exists

$$y \in \bigcap_{j \neq i} q_j \quad \text{with} \quad y \notin q_i.$$

Then

$$(I : y) = \left(\bigcap_{j=1}^{r} q_j : y \right) = (q_i : y),$$

because $y \in q_j$ for every $j \neq i$, so $(q_j : y) = A$ for $j \neq i$. By the preceding lemma, $(q_i : y)$ is $\mathfrak{p}_i$-primary.

Since A is noetherian and $\operatorname{rad}(q_i : y) = \mathfrak{p}_i$, there exists $n > 0$ such that $\mathfrak{p}_i^n \subseteq (q_i : y)$. Choose n minimal with this property. Then $\mathfrak{p}_i^{n-1} \not\subseteq (q_i : y)$, so choose $z \in \mathfrak{p}_i^{n-1}$ with $z \notin (q_i : y)$. We claim that

$$((q_i : y) : z) = \mathfrak{p}_i.$$

Indeed, if $a \in \mathfrak{p}_i$, then $az \in \mathfrak{p}_i^n \subseteq (q_i : y)$, so $a \in ((q_i : y) : z)$. Conversely, if $a \in ((q_i : y) : z)$ and $a \notin \mathfrak{p}_i$, then $az \in (q_i : y)$ while $z \notin (q_i : y)$. Since $(q_i : y)$ is $\mathfrak{p}_i$-primary, the condition $a \notin \mathfrak{p}_i = \operatorname{rad}(q_i : y)$ forces a contradiction: if $az \in (q_i : y)$ and $z \notin (q_i : y)$, primaryness implies $a^m \in (q_i : y)$ for some m, hence $a \in \mathfrak{p}_i$. Therefore $a \in \mathfrak{p}_i$. Thus the equality holds.

Finally,

$$(I : yz) = ((I : y) : z) = ((q_i : y) : z) = \mathfrak{p}_i.$$

Hence $\mathfrak{p}_i \in \operatorname{Ass}_A(A/I)$. This proves the theorem. □

7.2 Primary Decomposition of Modules

Definition 7.9. Let M be an A-module and let $N \subseteq M$ be a submodule. We say that N is a $\mathfrak{p}$-primary submodule of M if $\operatorname{Ass}_A(M/N) = \{\mathfrak{p}\}$. A primary decomposition of 0 in M is an expression

$$0 = N_1 \cap \cdots \cap N_r$$

where each N_i is a primary submodule of M.

Proposition 7.10. *Let $0 = N_1 \cap \cdots \cap N_r$ be a primary decomposition of 0 in an A-module M. Then the natural map*

$$M \longrightarrow \bigoplus_{i=1}^{r} M/N_i, \qquad m \longmapsto (m + N_1, \ldots, m + N_r)$$

is injective.

Proof. The kernel consists of all $m \in M$ such that $m \in N_i$ for every i. Hence

$$\ker = \bigcap_{i=1}^{r} N_i = 0.$$

Therefore the map is injective. □

Proposition 7.11. *Let M be a finitely generated torsion module over a PID R. If*

$$M = M_1 \oplus \cdots \oplus M_r,$$

then

$$\operatorname{Ann}_R(M) = \bigcap_{i=1}^{r} \operatorname{Ann}_R(M_i).$$

Proof. Let $a \in R$. The element a annihilates M if and only if $a(m_1 + \cdots + m_r) = 0$ for all $m_i \in M_i$. Since the sum is direct, this is equivalent to $am_i = 0$ for every $m_i \in M_i$ and every i. Thus $a \in \operatorname{Ann}_R(M)$ if and only if $a \in \operatorname{Ann}_R(M_i)$ for every i, which is exactly

$$a \in \bigcap_{i} \operatorname{Ann}_R(M_i).$$

□

Remark 7.12. For the rest of the chapter, the module-theoretic primary decomposition we actually use is the explicit one over a PID. The preceding definitions explain why the primes occurring in the decomposition should be regarded as associated primes of the module, rather than as accidental factors of a polynomial.

7.3 Specialization to Principal Ideal Domains

We now specialize the language of primary ideals and associated primes to a principal ideal domain. The point is not to redo primary decomposition by elementary polynomial tricks, but to identify the general notions from Sections 7.1 and 7.2 in the case where all ideals are principal.

Proposition 7.13. *Let R be a PID. The nonzero prime ideals of R are exactly the ideals (p), where $p \in R$ is irreducible.*

Proof. If p is irreducible, then (p) is prime. Indeed, suppose $p \mid ab$. If $p \nmid a$, then $\gcd(p, a) = 1$, so there exist $u, v \in R$ such that

$$up + va = 1.$$

Multiplying by b gives

$$upb + vab = b.$$

Both terms on the left are divisible by p, hence $p \mid b$. Therefore (p) is prime.

Conversely, let $\mathfrak{p} \neq 0$ be a prime ideal. Since R is a PID, $\mathfrak{p} = (a)$ for some nonzero nonunit a. If $a = bc$ with neither b nor c a unit, then $bc \in (a) = \mathfrak{p}$, so $b \in \mathfrak{p}$ or $c \in \mathfrak{p}$. If $b \in (a)$, then $b = ad$ for some $d \in R$, hence $a = bc = adc$. Since R is a domain and $a \neq 0$, we get $1 = dc$, so c is a unit. Similarly, if $c \in (a)$, then b is a unit. This contradiction proves that a is irreducible. □

Proposition 7.14. *Let R be a PID and let $p \in R$ be irreducible. The (p)-primary ideals of R are precisely (p^n) for $n \geq 1$.*

Proof. First, (p^n) is (p)-primary. Suppose $ab \in (p^n)$ and $a \notin (p^n)$. Write the exponent of p in the factorization of a as $s < n$. Since $p^n \mid ab$, the exponent of p in b is at least $n - s > 0$. Hence $p \mid b$, so $b^n \in (p^n)$. Thus (p^n) is primary, and its radical is (p).

Conversely, let q be a (p)-primary ideal. Since R is a PID, $q = (a)$ for some nonzero a. Factor

$$a = up_1^{e_1} \cdots p_t^{e_t}$$

into irreducibles. Then

$$\operatorname{rad}((a)) = (p_1 \cdots p_t).$$

Since $\operatorname{rad}(q) = (p)$, every p_i is associated to p. Therefore a is associated to p^n for some $n \geq 1$, and $q = (p^n)$. □

Lemma 7.15 (Coprime intersections)**.** *Let $I_1, \ldots, I_r$ be pairwise comaximal ideals in a commutative ring A. Then*

$$I_1 \cap \cdots \cap I_r = I_1 \cdots I_r.$$

Moreover, the natural map

$$A/(I_1 \cdots I_r) \longrightarrow \prod_{i=1}^{r} A/I_i$$

is an isomorphism.

Proof. For two comaximal ideals I and J, it is clear that $IJ \subseteq I \cap J$. Conversely, choose $u \in I$ and $v \in J$ with $u + v = 1$. If $x \in I \cap J$, then

$$x = x(u + v) = xu + xv.$$

Here $xu \in JI$ because $x \in J$ and $u \in I$, while $xv \in IJ$ because $x \in I$ and $v \in J$. Hence $x \in IJ$, so $I \cap J = IJ$.

The finite statement follows by induction, since the product $I_1 \cdots I_{r-1}$ is comaximal with I_r when each I_i is comaximal with I_r.

It remains to prove the product statement. The natural map

$$A \longrightarrow \prod_i A/I_i$$

has kernel $\bigcap_i I_i = \prod_i I_i$. To show surjectivity, fix i. Since I_i is comaximal with $\prod_{j \neq i} I_j$, one can choose an element whose image is 1 in A/I_i and 0 in every A/I_j for $j \neq i$. These elements generate the standard coordinate vectors in the product. Hence the map is surjective, and the first isomorphism theorem gives the desired isomorphism. □

Theorem 7.16 (Principal ideals have primary decompositions over a PID)**.** *Let R be a PID and let $0 \neq a \in R$ be a nonunit. Write*

$$a = up_1^{e_1} \cdots p_r^{e_r},$$

where u is a unit and the p_i are pairwise nonassociated irreducibles. Then

$$(a) = (p_1^{e_1}) \cap \cdots \cap (p_r^{e_r})$$

is a minimal primary decomposition of the ideal (a). Consequently, by the associated-prime theorem from Section 7.1,

$$\mathrm{Ass}_R(R/(a)) = \{(p_1), \ldots, (p_r)\}.$$

Proof. By the previous proposition, each $(p_i^{e_i})$ is (p_i)-primary. Since the irreducibles p_i are pairwise nonassociated, the ideals $(p_i^{e_i})$ are pairwise comaximal. Hence the coprime-intersection lemma gives

$$(p_1^{e_1}) \cap \cdots \cap (p_r^{e_r}) = (p_1^{e_1} \cdots p_r^{e_r}) = (a).$$

The radicals of the components are the distinct prime ideals $(p_1), \ldots, (p_r)$.

The decomposition is irredundant. Indeed, if the i-th component is omitted, then the product

$$b_i = \prod_{j \neq i} p_j^{e_j}$$

lies in every remaining component but does not lie in (p_i), hence does not lie in $(p_i^{e_i})$. Thus the intersection becomes strictly larger after omitting the i-th component.

Therefore this is a minimal primary decomposition. Applying the associated-prime theorem from Section 7.1, namely the theorem that the radicals in a minimal primary decomposition are exactly the associated primes of the quotient, gives

$$\mathrm{Ass}_R(R/(a)) = \{(p_1), \ldots, (p_r)\}.$$

□

Proposition 7.17. *Let $m \in M$ be an element of an R-module. Then the cyclic submodule Rm is isomorphic to*

$$R/\operatorname{Ann}_R(m).$$

Proof. Define $\varphi : R \to Rm$ by $\varphi(r) = rm$. This map is an R-module homomorphism and is surjective by definition of Rm. Its kernel is

$$\ker\varphi = \{r \in R : rm = 0\} = \operatorname{Ann}_R(m).$$

By the first isomorphism theorem for modules,

$$R/\operatorname{Ann}_R(m) \cong Rm.$$

□

Corollary 7.18. *Let R be a PID and let $0 \neq a \in R$. If $a = u\prod_i p_i^{e_i}$, then the cyclic torsion module $R/(a)$ has associated primes*

$$\operatorname{Ass}_R(R/(a)) = \{(p_i) : p_i \mid a\}.$$

In particular, the associated primes of a cyclic torsion module are exactly the prime factors of its annihilator.

Proof. This is precisely the conclusion of the preceding theorem applied to the principal ideal (a). □

7.4 Primary Decomposition over a PID

This section is the main tool for the next chapter. We now use the primary-decomposition language of Section 7.2 together with the associated-prime calculation of Section 7.3. The construction is best viewed through the Chinese remainder decomposition of the quotient ring; no explicit Bezout idempotents are needed in the statement.

Definition 7.19. Let M be a torsion module over a PID R, and let $p \in R$ be irreducible. The p-primary component of M is

$$M[p^\infty] = \{m \in M : p^n m = 0 \text{ for some } n > 0\}.$$

Proposition 7.20. *For a torsion module M over a PID, $M[p^\infty]$ is a submodule of M.*

Proof. If $m, n \in M[p^\infty]$, then $p^a m = 0$ and $p^b n = 0$ for some $a, b > 0$. Let $c = \max(a, b)$. Then

$$p^c(m + n) = p^c m + p^c n = 0,$$

so $m + n \in M[p^\infty]$. If $r \in R$, then

$$p^a(rm) = r(p^a m) = 0,$$

so $rm \in M[p^\infty]$. Hence $M[p^\infty]$ is a submodule. □

Lemma 7.21 (Modules over a finite product)**.** *Let $A = A_1 \times \cdots \times A_r$ be a finite product of rings. If M is an A-module, then*

$$M = e_1 M \oplus \cdots \oplus e_r M,$$

where $e_i = (0, \ldots, 1, \ldots, 0) \in A$ is the i-th coordinate idempotent.

Proof. The elements e_i satisfy

$$e_i^2 = e_i, \qquad e_i e_j = 0 \quad (i \neq j), \qquad e_1 + \cdots + e_r = 1.$$

For every $m \in M$,

$$m = 1m = (e_1 + \cdots + e_r)m = e_1 m + \cdots + e_r m,$$

so $M = e_1 M + \cdots + e_r M$. If

$$m_1 + \cdots + m_r = 0, \qquad m_i \in e_i M,$$

then applying e_k gives $m_k = 0$, because $e_k m_k = m_k$ and $e_k m_i = 0$ for $i \neq k$. Hence the sum is direct. □

Theorem 7.22 (Primary decomposition controlled by the annihilator)**.** *Let M be an R-module over a PID R. Suppose*

$$\operatorname{Ann}_R(M) = (a), \qquad a = u p_1^{e_1} \cdots p_r^{e_r},$$

where u is a unit and the p_i are pairwise nonassociated irreducibles. Then

$$M = M[p_1^\infty] \oplus \cdots \oplus M[p_r^\infty]$$

and

$$M[p_i^\infty] = \{m \in M : p_i^{e_i} m = 0\}.$$

Moreover, if $N_i = \bigoplus_{j \neq i} M[p_j^\infty]$, then

$$0 = N_1 \cap \cdots \cap N_r$$

is a primary decomposition of 0 *in the sense of Section 7.2, after omitting the zero primary components.*

Proof. Since $aM = 0$, the R-module structure on M factors through the quotient ring $R/(a)$. By the principal-ideal primary decomposition from Section 7.3 and the Chinese remainder part of the coprime-intersection lemma,

$$R/(a) \cong \prod_{i=1}^{r} R/(p_i^{e_i}).$$

Thus M is a module over this finite product. By the module decomposition for finite product rings,

$$M = e_1 M \oplus \cdots \oplus e_r M,$$

where e_i is the i-th coordinate idempotent under the product-ring decomposition.

We claim that

$$e_i M = \{m \in M : p_i^{e_i} m = 0\}.$$

The inclusion $e_i M \subseteq \{m : p_i^{e_i} m = 0\}$ holds because $p_i^{e_i} = 0$ in the i-th factor $R/(p_i^{e_i})$. Conversely, suppose $p_i^{e_i} m = 0$ and write $m = \sum_j e_j m$. For $j \neq i$, the element $p_i^{e_i}$ is a unit in $R/(p_j^{e_j})$, because p_i and p_j are nonassociated irreducibles. Hence multiplication by $p_i^{e_i}$ is an automorphism of $e_j M$. Since

$$0 = p_i^{e_i} m = \sum_j p_i^{e_i} e_j m,$$

and the sum is direct, we get $p_i^{e_i} e_j m = 0$ for every j. For $j \neq i$, invertibility of $p_i^{e_i}$ on $e_j M$ gives $e_j m = 0$. Hence $m = e_i m \in e_i M$. This proves the claim.

The same argument gives $e_iM = M[p_i^\infty]$. Indeed, e_iM is killed by $p_i^{e_i}$, so $e_iM \subseteq M[p_i^\infty]$. Conversely, if $p_i^N m = 0$, then p_i^N is a unit on e_jM for $j \neq i$, and the direct product decomposition forces all components e_jm with $j \neq i$ to vanish. Hence $m \in e_iM$.

It remains to interpret this as a primary decomposition in the sense of Section 7.2. If $e_iM \neq 0$, then $M/N_i \cong e_iM$. The module e_iM is killed by $p_i^{e_i}$, so any associated prime of e_iM contains (p_i). Since the only nonzero prime ideal of a PID containing (p_i) is (p_i) itself, every associated prime of e_iM is (p_i). Conversely, choose a nonzero element $x \in e_iM$ with $p_i^n x = 0$ and n minimal. Then $y = p_i^{n-1}x \neq 0$ and $p_i y = 0$. If $ry = 0$ and $p_i \nmid r$, then r is a unit modulo (p_i), so $y = 0$, a contradiction. Thus $\mathrm{Ann}_R(y) = (p_i)$, and $(p_i) \in \mathrm{Ass}_R(e_iM)$. Hence

$$\mathrm{Ass}_R(M/N_i) = \mathrm{Ass}_R(e_iM) = \{(p_i)\}.$$

By the definition of primary submodule in Section 7.2, N_i is (p_i)-primary. Since the decomposition of M is direct, the intersection of the submodules $N_i = \bigoplus_{j\neq i} e_jM$ is zero. Therefore

$$0 = N_1 \cap \cdots \cap N_r$$

is a primary decomposition of 0 in M, with the zero components omitted. □

Theorem 7.23 (Associated primes over a PID)**.** *Let M be a finitely generated torsion module over a PID R. Then*

$$\mathrm{Ass}_R(M) = \{(p) : M[p^\infty] \neq 0\}.$$

Proof. Because M is finitely generated and torsion, there exists a nonzero element $a \in R$ with $aM = 0$. Thus $\mathrm{Ann}_R(M) = (d)$ for some nonzero $d \in R$. Factor

$$d = up_1^{e_1} \cdots p_r^{e_r}.$$

By the preceding theorem,

$$M = M[p_1^\infty] \oplus \cdots \oplus M[p_r^\infty].$$

Write $M_i = M[p_i^\infty]$.

First suppose $M_i \neq 0$. In the proof of the preceding theorem we showed, using the definition of associated primes from Section 7.1, that

$$\mathrm{Ass}_R(M_i) = \{(p_i)\}.$$

Hence $(p_i) \in \mathrm{Ass}_R(M)$, since an element of M_i may also be regarded as an element of the direct sum M.

Conversely, let $\mathfrak{p} \in \mathrm{Ass}_R(M)$. Then $\mathfrak{p} = \mathrm{Ann}_R(m)$ for some nonzero $m \in M$. Write

$$m = m_1 + \cdots + m_r, \qquad m_i \in M_i.$$

Using the same annihilator-intersection principle as in Section 7.2, we have

$$\mathrm{Ann}_R(m) = \bigcap_{m_i \neq 0} \mathrm{Ann}_R(m_i).$$

Since $\mathfrak{p}$ is prime and the product of the ideals $\mathrm{Ann}_R(m_i)$ is contained in their intersection, one of the ideals $\mathrm{Ann}_R(m_i)$ is contained in $\mathfrak{p}$. On the other hand, $\mathfrak{p} = \mathrm{Ann}_R(m)$ is contained in every $\mathrm{Ann}_R(m_i)$. Hence $\mathrm{Ann}_R(m_i) = \mathfrak{p}$ for some i. Therefore $\mathfrak{p} \in \mathrm{Ass}_R(M_i) = \{(p_i)\}$, and $M_i \neq 0$.

Thus the associated primes of M are exactly the primes attached to the nonzero primary components:

$$\mathrm{Ass}_R(M) = \{(p_i) : M[p_i^\infty] \neq 0\}.$$

This is the desired formula. □

Theorem 7.24 (Primary decomposition for finitely generated torsion modules over a PID)**.** *Let M be a finitely generated torsion module over a PID R. Then*

$$M = \bigoplus_{(p)\in \mathrm{Ass}_R(M)} M[p^\infty].$$

Proof. Let $\mathrm{Ann}_R(M) = (d)$ and factor d into irreducible powers. The primary-decomposition theorem controlled by the annihilator gives a direct sum decomposition into the corresponding p-primary components. The theorem on associated primes over a PID identifies precisely which of these components are nonzero: they are exactly the components indexed by primes in $\mathrm{Ass}_R(M)$. Therefore

$$M = \bigoplus_{(p)\in \mathrm{Ass}_R(M)} M[p^\infty].$$

□

7.5 The Structure Theorem over a PID

In this final section we record the classification theorem over a PID. The primary decomposition tool has already been separated from this classification result; its role here is to reduce the torsion part to primary torsion modules. Smith normal form appears only as the matrix shadow of the invariant factor form, rather than as the proof of it.

Lemma 7.25 (An extension lemma for primary modules)**.** *Let R be a PID, let $p \in R$ be irreducible, and let M be a finitely generated R-module such that $p^e M = 0$. If $N \subseteq M$ is a submodule, then every homomorphism*

$$f : N \longrightarrow R/(p^e)$$

extends to a homomorphism

$$F : M \longrightarrow R/(p^e).$$

Proof. Since M is finitely generated, it is enough to show that a homomorphism defined on a submodule N can be extended to a submodule of the form $N + Ry$. Repeating the argument for finitely many generators of M/N then gives an extension to all of M.

Let $f : N \to R/(p^e)$ be given and fix $y \in M$. Put

$$I = \{r \in R : ry \in N\}.$$

Since $p^e y = 0$, we have $p^e \in I$. Hence I is a nonzero ideal of the PID R, say $I = (p^s)$ with $0 \le s \le e$. The element $p^s y$ lies in N, so $f(p^s y)$ is defined. Moreover,

$$p^{e-s} f(p^s y) = f(p^e y) = 0.$$

In the module $R/(p^e)$, the elements killed by p^{e-s} are exactly the multiples of p^s: indeed, if $p^{e-s}\overline{a} = 0$, then $p^e \mid p^{e-s}a$, so $p^s \mid a$. Therefore there exists $\overline{t} \in R/(p^e)$ such that

$$p^s \overline{t} = f(p^s y).$$

Define

$$F : N + Ry \longrightarrow R/(p^e), \qquad F(n + ay) = f(n) + a\overline{t}.$$

We check that this is well-defined. Suppose

$$n + ay = n' + a'y.$$

Then $(a - a')y = n' - n \in N$, so $a - a' \in I = (p^s)$. Write $a - a' = bp^s$. Then

$$f((a - a')y) = bf(p^s y) = bp^s \bar{t} = (a - a')\bar{t}.$$

Since $n - n' = -(a - a')y$, we get

$$f(n) - f(n') = -(a - a')\bar{t}.$$

Thus

$$f(n) + a\bar{t} = f(n') + a'\bar{t}.$$

So F is well-defined. It is plainly R-linear and extends f. This proves the one-step extension, and hence the lemma. □

Lemma 7.26 (Splitting off a maximal cyclic summand)**.** *Let M be a finitely generated module over a PID R such that $p^e M = 0$. Suppose there exists $x \in M$ with*

$$\mathrm{Ann}_R(x) = (p^e).$$

Then the cyclic submodule $Rx \cong R/(p^e)$ is a direct summand of M.

Proof. The map

$$\theta : Rx \longrightarrow R/(p^e), \qquad \theta(rx) = r + (p^e),$$

is well-defined because $rx = r'x$ if and only if $(r - r')x = 0$, equivalently $r - r' \in (p^e)$. It is an isomorphism.

By the preceding lemma, θ extends to an R-module homomorphism

$$\Theta : M \longrightarrow R/(p^e).$$

Let

$$\iota : R/(p^e) \longrightarrow Rx$$

be the inverse isomorphism to θ. Then

$$P = \iota \circ \Theta : M \longrightarrow Rx$$

is an R-module homomorphism and P restricts to the identity on Rx. Hence P is a projection onto Rx. For every $m \in M$,

$$m = P(m) + (m - P(m)),$$

where $P(m) \in Rx$ and $m - P(m) \in \ker P$. Also $Rx \cap \ker P = 0$, because P is the identity on Rx. Thus

$$M = Rx \oplus \ker P.$$

□

Theorem 7.27 (Cyclic decomposition of a primary torsion module)**.** *Let R be a PID, let $p \in R$ be irreducible, and let M be a finitely generated module such that $p^e M = 0$ for some $e > 0$. Then*

$$M \cong \bigoplus_j R/(p^{e_j})$$

for some positive integers e_j.

Proof. We use induction on the integer

$$\ell_p(M) = \sum_{i \geq 0} \dim_{R/(p)} p^i M / p^{i+1} M.$$

This is finite because $p^e M = 0$ and each quotient $p^i M / p^{i+1} M$ is generated over the field $R/(p)$ by the images of a finite set of generators of M.

If $M = 0$, there is nothing to prove. Otherwise let e_0 be minimal such that $p^{e_0} M = 0$. Then $p^{e_0 - 1} M \neq 0$, so there exists $x \in M$ with

$$p^{e_0} x = 0, \qquad p^{e_0 - 1} x \neq 0.$$

It follows that $\operatorname{Ann}_R(x) = (p^{e_0})$. Indeed, the annihilator of x contains (p^{e_0}), and since R is a PID and M is killed by a power of p, it must be of the form (p^s) for some $s \leq e_0$; the condition $p^{e_0 - 1} x \neq 0$ forces $s = e_0$.

By the preceding lemma, Rx is a direct summand:

$$M = Rx \oplus M'$$

for some submodule M'. Since $Rx \neq 0$, we have $\ell_p(M') < \ell_p(M)$. By the induction hypothesis,

$$M' \cong \bigoplus_j R/(p^{e_j}).$$

Also $Rx \cong R/(p^{e_0})$. Therefore

$$M \cong R/(p^{e_0}) \oplus \bigoplus_j R/(p^{e_j}),$$

which proves the theorem. □

Theorem 7.28 (Structure theorem over a PID, elementary divisor form)**.** *Let R be a PID and let M be a finitely generated R-module. Then*

$$M \cong R^s \oplus \bigoplus_p \bigoplus_j R/(p^{e_{p,j}}),$$

where p runs through irreducibles up to associates and only finitely many torsion summands occur.

Proof. By the splitting theorem from Chapter 6,

$$M \cong M_{\mathrm{tor}} \oplus R^s$$

for some $s \geq 0$. Thus it remains to decompose the finitely generated torsion module M_{tor}.

If $M_{\mathrm{tor}} = 0$, there is nothing more to prove. Otherwise M_{tor} has a nonzero annihilator (a), since it is finitely generated and torsion. By the primary decomposition theorem over a PID proved in the preceding section,

$$M_{\mathrm{tor}} = \bigoplus_{(p) \in \operatorname{Ass}_R(M_{\mathrm{tor}})} M_{\mathrm{tor}}[p^\infty].$$

Each summand is finitely generated and killed by a power of the single irreducible p. By the cyclic decomposition theorem for primary torsion modules,

$$M_{\mathrm{tor}}[p^\infty] \cong \bigoplus_j R/(p^{e_{p,j}}).$$

Combining these decompositions gives

$$M \cong R^s \oplus \bigoplus_p \bigoplus_j R/(p^{e_{p,j}}),$$

which is the claimed elementary divisor form. □

Corollary 7.29 (Invariant factor form)**.** *Let R be a PID and let M be a finitely generated R-module. Then*

$$M \cong R^s \oplus R/(d_1) \oplus \cdots \oplus R/(d_t),$$

where $d_i \neq 0$, d_i is not a unit, and

$$d_1 \mid d_2 \mid \cdots \mid d_t.$$

Remark 7.30 (Smith normal form)**.** The usual Smith normal form of a matrix over a PID is the presentation-matrix version of the invariant factor form above. If a homomorphism of finite free modules is represented by a matrix A, then the invariant factors of $\operatorname{coker} A$ are the diagonal entries in the Smith form. In these notes, the module structure theorem is the main result; the matrix normal form is best regarded as a coordinate expression of it.

Proposition 7.31 (Uniqueness of elementary divisors)**.** *Let*

$$M_p = \bigoplus_j R/(p^{e_j})$$

be a finite p-primary torsion module over a PID. The multiset of exponents $\{e_j\}$ is determined by M_p.

Proof. For $k \geq 1$, consider the quotient vector space over the field $R/(p)$

$$Q_k = p^{k-1}M_p/p^kM_p.$$

In a summand $R/(p^e)$, the quotient

$$p^{k-1}R/(p^e) \,/\, p^kR/(p^e)$$

is one-dimensional over $R/(p)$ if $e \geq k$, and zero if $e < k$. Therefore

$$\dim_{R/(p)} Q_k = \#\{j : e_j \geq k\}.$$

The numbers on the left are invariants of M_p, so they determine the number of exponents at least k for every k. Hence they determine exactly how many exponents equal k, namely

$$\#\{j : e_j = k\} = \dim Q_k - \dim Q_{k+1}.$$

Thus the multiset $\{e_j\}$ is determined by M_p. □

Corollary 7.32. *The elementary divisor form of a finitely generated module over a PID is unique up to associates and reordering.*

Proof. The free rank is determined by

$$\dim_{\operatorname{Frac}(R)}(\operatorname{Frac}(R) \otimes_R M).$$

The torsion submodule is the kernel of the natural map $M \to \operatorname{Frac}(R) \otimes_R M$, so it is intrinsic. For each irreducible p, the p-primary component $M[p^\infty]$ is intrinsic, and by the preceding proposition the elementary divisors in each p-primary component are uniquely determined. Therefore the whole elementary divisor decomposition is unique up to associates and reordering. □

Corollary 7.33 (Classification of finite abelian groups)**.** *Every finite abelian group G is isomorphic to a finite direct sum*

$$\bigoplus_p \bigoplus_j \mathbb{Z}/(p^{e_{p,j}})\mathbb{Z}.$$

The prime powers $p^{e_{p,j}}$ are uniquely determined up to reordering.

Proof. A finite abelian group is the same thing as a finite, hence finitely generated torsion, $\mathbb{Z}$-module. Since $\mathbb{Z}$ is a PID, the elementary divisor form applied to $R = \mathbb{Z}$ gives the displayed decomposition. Uniqueness follows from uniqueness of elementary divisors over a PID. □

Chapter 8

Linear Operators as $\mathbb{F}[x]$-Modules

The point of this chapter is that the structure theory of a linear operator is not a separate collection of polynomial tricks. It is the structure theory of a finitely generated torsion module over the PID $\mathbb{F}[x]$.

8.1 The Associated Module

Definition 8.1. Let V be a vector space over $\mathbb{F}$ and let $T \in \mathrm{End}_{\mathbb{F}}(V)$. Define an $\mathbb{F}[x]$-module structure on the underlying abelian group of V by

$$f(x) \cdot v = f(T)v.$$

This $\mathbb{F}[x]$-module is denoted by V_T.

Proposition 8.2. *The formula above makes V into an $\mathbb{F}[x]$-module. Moreover, the $\mathbb{F}[x]$-submodules of V_T are exactly the T-invariant subspaces of V.*

Proof. Let $f, g \in \mathbb{F}[x]$, $a \in \mathbb{F}$, and $v, w \in V$. Polynomial addition and multiplication give

$$(f+g)(T)v = f(T)v + g(T)v,$$

$$(fg)(T)v = f(T)(g(T)v),$$

and $1(T)v = v$. Also

$$f(T)(v+w) = f(T)v + f(T)w, \qquad f(T)(av) = af(T)v,$$

because $f(T)$ is an $\mathbb{F}$-linear endomorphism of V. Hence the axioms of an $\mathbb{F}[x]$-module are satisfied.

Let $W \subseteq V_T$ be an $\mathbb{F}[x]$-submodule. Since $x \cdot w = T(w)$, every $w \in W$ satisfies $T(w) \in W$. Thus W is T-invariant as a subspace of V.

Conversely, suppose $W \subseteq V$ is a T-invariant subspace. Then $T(W) \subseteq W$, and by induction $T^n(W) \subseteq W$ for every $n \geq 0$. Hence for every polynomial $f(x) = a_0 + a_1 x + \cdots + a_n x^n$ and every $w \in W$,

$$f(T)w = a_0 w + a_1 T(w) + \cdots + a_n T^n(w) \in W.$$

Thus W is stable under the $\mathbb{F}[x]$-action, so it is an $\mathbb{F}[x]$-submodule of V_T. □

Proposition 8.3. *If V is finite-dimensional over $\mathbb{F}$, then V_T is a finitely generated torsion $\mathbb{F}[x]$-module.*

Proof. Let $v_1, \dots, v_n$ be an $\mathbb{F}$-basis of V. Since $\mathbb{F} \subseteq \mathbb{F}[x]$, the same set generates V_T as an $\mathbb{F}[x]$-module: every vector $v = \sum_i a_i v_i$ can be written as $\sum_i a_i(x) \cdot v_i$ with constant polynomials $a_i(x) = a_i$.

It remains to show torsion. The vector space $\mathrm{End}_{\mathbb{F}}(V)$ has finite dimension n^2 over $\mathbb{F}$. Hence the $n^2 + 1$ endomorphisms

$$I, T, T^2, \dots, T^{n^2}$$

are linearly dependent. Therefore there exists a nonzero polynomial $f \in \mathbb{F}[x]$ such that $f(T) = 0$. This means $f \cdot v = 0$ for every $v \in V_T$. Thus every element of V_T is killed by the same nonzero polynomial f, and V_T is a torsion $\mathbb{F}[x]$-module. □

8.2 Minimal Polynomial and Associated Primes

Definition 8.4. Assume V is finite-dimensional. The annihilator ideal of V_T is

$$\mathrm{Ann}_{\mathbb{F}[x]}(V_T) = \{f \in \mathbb{F}[x] : f(T) = 0\}.$$

Since $\mathbb{F}[x]$ is a PID, this ideal is generated by a unique monic polynomial $m_T(x)$. This polynomial is called the minimal polynomial of T.

Proposition 8.5. *The minimal polynomial m_T is the monic polynomial of least positive degree such that $m_T(T) = 0$. Moreover, if $f(T) = 0$, then m_T divides f.*

Proof. By definition,

$$\mathrm{Ann}_{\mathbb{F}[x]}(V_T) = (m_T).$$

Thus $m_T(T) = 0$. If $f(T) = 0$, then $f \in \mathrm{Ann}_{\mathbb{F}[x]}(V_T) = (m_T)$, so $f = m_T g$ for some $g \in \mathbb{F}[x]$; hence m_T divides f.

If h is a nonzero polynomial satisfying $h(T) = 0$, then $h \in (m_T)$, so $h = m_T g$ for some nonzero $g \in \mathbb{F}[x]$. Hence

$$\deg h = \deg m_T + \deg g \geq \deg m_T.$$

Therefore no nonzero annihilating polynomial has degree smaller than m_T. Since m_T is chosen monic, it is exactly the monic annihilating polynomial of least positive degree. □

Proposition 8.6. *Similar matrices have the same minimal polynomial.*

Proof. Suppose $B = P^{-1}AP$. For every $k \geq 0$,

$$B^k = (P^{-1}AP)^k = P^{-1}A^k P,$$

which follows by induction on k. Therefore, for every polynomial f,

$$f(B) = P^{-1} f(A) P.$$

Thus $f(B) = 0$ if and only if $f(A) = 0$. Hence the annihilator ideals of A and B in $\mathbb{F}[x]$ are equal. Their monic generators are therefore equal, so the minimal polynomials are the same. □

Proposition 8.7 (Associated primes of a linear operator)**.** *Let V be finite-dimensional and let $T \in \mathrm{End}_{\mathbb{F}}(V)$. Suppose*

$$m_T(x) = p_1(x)^{e_1} \cdots p_r(x)^{e_r},$$

where the p_i are distinct monic irreducible polynomials. Then

$$\mathrm{Ass}_{\mathbb{F}[x]}(V_T) = \{(p_1), \dots, (p_r)\}.$$

Proof. The module V_T is a finitely generated torsion module over the PID $\mathbb{F}[x]$, and

$$\mathrm{Ann}_{\mathbb{F}[x]}(V_T) = (m_T).$$

By the theorem on associated primes over a PID from Chapter 7,

$$\mathrm{Ass}_{\mathbb{F}[x]}(V_T) = \{(p) : V_T[p^\infty] \neq 0\}.$$

The theorem on primary decomposition controlled by the annihilator from Chapter 7 says that the only nonzero primary components are those corresponding to the irreducible factors of the generator of the annihilator. Since the irreducible factors of m_T are exactly $p_1, \ldots, p_r$, the associated primes are precisely

$$(p_1), \ldots, (p_r).$$

□

8.3 Primary Decomposition of a Linear Operator

Theorem 8.8 (Primary decomposition of an operator)**.** *Let V be finite-dimensional and let $T \in \mathrm{End}_{\mathbb{F}}(V)$. Suppose*

$$m_T(x) = p_1(x)^{e_1} \cdots p_r(x)^{e_r},$$

where the p_i are distinct monic irreducible polynomials. Then

$$V = \ker p_1(T)^{e_1} \oplus \cdots \oplus \ker p_r(T)^{e_r}.$$

Each summand is T-invariant.

Proof. By the preceding section, V_T is a finitely generated torsion module over the PID $\mathbb{F}[x]$, and

$$\mathrm{Ann}_{\mathbb{F}[x]}(V_T) = (m_T).$$

The theorem on primary decomposition controlled by the annihilator from Chapter 7 applies to the factorization of m_T. It gives

$$V_T = V_T[p_1^\infty] \oplus \cdots \oplus V_T[p_r^\infty]$$

and, more precisely,

$$V_T[p_i^\infty] = \{v \in V_T : p_i(x)^{e_i} \cdot v = 0\}.$$

Under the $\mathbb{F}[x]$-module structure on V_T, the equality $p_i(x)^{e_i} \cdot v = 0$ is exactly

$$p_i(T)^{e_i} v = 0.$$

Hence

$$V_T[p_i^\infty] = \ker p_i(T)^{e_i}.$$

Therefore

$$V = \ker p_1(T)^{e_1} \oplus \cdots \oplus \ker p_r(T)^{e_r}.$$

Finally, each summand is an $\mathbb{F}[x]$-submodule of V_T, hence it is T-invariant by the first proposition of this chapter. □

Corollary 8.9 (Generalized eigenspace decomposition)**.** *Assume $\mathbb{F}$ is algebraically closed. If*

$$m_T(x) = \prod_{\lambda} (x - \lambda)^{e_\lambda},$$

then

$$V = \bigoplus_{\lambda} \ker(T - \lambda I)^{e_\lambda}.$$

Proof. Over an algebraically closed field, every irreducible polynomial in $\mathbb{F}[x]$ is linear. Thus each irreducible factor of m_T has the form $p_\lambda(x) = x - \lambda$. Applying the preceding theorem gives

$$V = \bigoplus_{\lambda} \ker p_\lambda(T)^{e_\lambda} = \bigoplus_{\lambda} \ker(T - \lambda I)^{e_\lambda}.$$

□

8.4 Cyclic Decomposition and Rational Canonical Form

Definition 8.10. For a monic polynomial

$$f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0,$$

the companion matrix of f is

$$C(f) = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{pmatrix}.$$

Proposition 8.11. *Let $M = \mathbb{F}[x]/(f)$, where f is monic of degree n. With respect to the $\mathbb{F}$-basis*

$$1, x, \ldots, x^{n-1},$$

the linear map given by multiplication by x has matrix $C(f)$.

Proof. The quotient $\mathbb{F}[x]/(f)$ has $\mathbb{F}$-basis $1, x, \ldots, x^{n-1}$ by the division algorithm. Multiplication by x sends

$$1 \mapsto x, \quad x \mapsto x^2, \quad \ldots, \quad x^{n-2} \mapsto x^{n-1}.$$

Thus the first $n-1$ columns of the matrix have a single 1 immediately below the diagonal.

Since $f = 0$ in the quotient, we have

$$x^n = -a_{n-1}x^{n-1} - \cdots - a_1 x - a_0.$$

Therefore the image of x^{n-1} under multiplication by x has coordinate column

$$(-a_0, -a_1, \ldots, -a_{n-1})^t.$$

This gives exactly the displayed companion matrix. □

Theorem 8.12 (Rational canonical form)**.** *Let V be finite-dimensional and let $T \in \mathrm{End}_{\mathbb{F}}(V)$. Then there exist monic polynomials*

$$f_1 \mid f_2 \mid \cdots \mid f_r$$

such that, for a suitable basis of V, the matrix of T is block diagonal with companion blocks

$$C(f_1), \ldots, C(f_r).$$

The polynomials f_i are uniquely determined by T.

Proof. View V as the $\mathbb{F}[x]$-module V_T. We have proved that V_T is a finitely generated torsion module over the PID $\mathbb{F}[x]$. Applying the invariant factor form of the structure theorem from Chapter 7 gives an isomorphism of $\mathbb{F}[x]$-modules

$$V_T \cong \mathbb{F}[x]/(f_1) \oplus \cdots \oplus \mathbb{F}[x]/(f_r),$$

where $f_1, \ldots, f_r$ are monic nonconstant polynomials satisfying

$$f_1 \mid f_2 \mid \cdots \mid f_r.$$

For each summand $\mathbb{F}[x]/(f_i)$, choose the $\mathbb{F}$-basis

$$1, x, \ldots, x^{\deg f_i - 1}.$$

The $\mathbb{F}[x]$-module structure says that the operator T corresponds to multiplication by x. By the preceding proposition, multiplication by x on this summand has matrix $C(f_i)$. Taking the union of these bases over all summands gives a basis of V in which the matrix of T is block diagonal with companion blocks $C(f_i)$.

The uniqueness of the polynomials f_i is exactly the uniqueness of the invariant factor decomposition in the structure theorem over a PID, applied to the module V_T. If T is represented in two bases, the resulting $\mathbb{F}[x]$-modules are the same module V_T, so the invariant factors must agree. □

Proposition 8.13. *If T has invariant factors $f_1 \mid \cdots \mid f_r$, then*

$$m_T = f_r.$$

Proof. By the rational canonical form proof,

$$V_T \cong \bigoplus_{i=1}^{r} \mathbb{F}[x]/(f_i).$$

The annihilator of the cyclic module $\mathbb{F}[x]/(f_i)$ is the ideal (f_i). Hence the annihilator of the direct sum is

$$(f_1) \cap \cdots \cap (f_r).$$

In a PID this intersection is generated by the least common multiple of $f_1, \ldots, f_r$. Since $f_1 \mid \cdots \mid f_r$, this least common multiple is f_r. Therefore

$$\mathrm{Ann}_{\mathbb{F}[x]}(V_T) = (f_r).$$

By definition, the monic generator of $\mathrm{Ann}_{\mathbb{F}[x]}(V_T)$ is m_T, so $m_T = f_r$. □

8.5 Jordan Normal Form

Definition 8.14. For $\lambda \in \mathbb{F}$, the Jordan block of size n with eigenvalue λ is

$$J_n(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & \cdots & 0 & \lambda \end{pmatrix}.$$

Proposition 8.15. *The cyclic $\mathbb{F}[x]$-module $\mathbb{F}[x]/((x-\lambda)^n)$ corresponds to a single Jordan block $J_n(\lambda)$.*

Proof. Put $y = x - \lambda$. Then

$$\mathbb{F}[x]/((x-\lambda)^n) \cong \mathbb{F}[y]/(y^n).$$

Use the ordered basis

$$y^{n-1}, y^{n-2}, \ldots, y, 1.$$

Multiplication by x is multiplication by $\lambda + y$. Hence

$$x \cdot y^k = (\lambda + y)y^k = \lambda y^k + y^{k+1},$$

where $y^n = 0$ in the quotient.

In the ordered basis above, the term λy^k gives the diagonal entry λ, and the term y^{k+1} moves one step to the previous basis vector, giving a 1 on the superdiagonal. For $k = n-1$, the term y^n is zero. Therefore the matrix is precisely $J_n(\lambda)$. □

Theorem 8.16 (Jordan normal form)**.** *Assume $\mathbb{F}$ is algebraically closed. Let V be finite-dimensional and let $T \in \mathrm{End}_{\mathbb{F}}(V)$. Then there exists a basis of V in which the matrix of T is block diagonal with Jordan blocks. The multiset of Jordan blocks is uniquely determined by T.*

Proof. Since $\mathbb{F}$ is algebraically closed, every irreducible polynomial in $\mathbb{F}[x]$ is of the form $x - \lambda$. Apply the elementary divisor form from Chapter 7 to the finitely generated torsion $\mathbb{F}[x]$-module V_T. We obtain

$$V_T \cong \bigoplus_{\lambda} \bigoplus_{j} \mathbb{F}[x]/((x-\lambda)^{e_{\lambda,j}}).$$

Each summand is a cyclic primary module. By the preceding proposition, the operator induced by multiplication by x on the summand $\mathbb{F}[x]/((x-\lambda)^{e_{\lambda,j}})$ is represented by the Jordan block $J_{e_{\lambda,j}}(\lambda)$ in a suitable basis. Taking the union of these bases gives a basis of V in which the matrix of T is block diagonal with Jordan blocks.

The elementary divisors $((x-\lambda)^{e_{\lambda,j}})$ are unique by the uniqueness theorem for elementary divisors in Chapter 7. Since a Jordan block $J_e(\lambda)$ corresponds exactly to the elementary divisor $(x-\lambda)^e$, the multiset of Jordan blocks is uniquely determined by T. □

8.6 Characteristic Polynomial and Cayley–Hamilton

Definition 8.17. If A is a square matrix over $\mathbb{F}$, its characteristic polynomial is

$$\chi_A(t) = \det(tI - A).$$

For an endomorphism T of a finite-dimensional vector space, define $\chi_T(t)$ to be the characteristic polynomial of any matrix of T.

Proposition 8.18. *The characteristic polynomial of an endomorphism is independent of the chosen basis.*

Proof. If two matrices of T are related by $B = P^{-1}AP$, then

$$tI - B = tI - P^{-1}AP = P^{-1}(tI - A)P.$$

Taking determinants gives

$$\det(tI - B) = \det(P^{-1})\det(tI - A)\det(P) = \det(tI - A).$$

Thus the characteristic polynomial is basis-independent. □

Lemma 8.19. *If $f(t)$ is monic, then the companion matrix $C(f)$ has characteristic polynomial $f(t)$.*

Proof. Write

$$f(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0.$$

The companion matrix $C(f)$ represents multiplication by x on $\mathbb{F}[x]/(f)$ in the basis $1, x, \ldots, x^{n-1}$. Therefore $f(C(f)) = 0$, because $f(x) = 0$ in the quotient module. Hence the minimal polynomial of $C(f)$ divides f.

On the other hand, the vectors

$$e_1, C(f)e_1, C(f)^2e_1, \ldots, C(f)^{n-1}e_1$$

are the standard basis vectors corresponding to $1, x, \ldots, x^{n-1}$, hence are linearly independent. If a polynomial g of degree $< n$ satisfied $g(C(f)) = 0$, then applying it to e_1 would give a nontrivial linear relation among these n vectors unless $g = 0$. Thus the minimal polynomial of $C(f)$ has degree n. Since it divides the monic polynomial f of degree n, it is equal to f.

Now the characteristic polynomial of the $n \times n$ matrix $C(f)$ is monic of degree n, and the minimal polynomial divides the characteristic polynomial once Cayley–Hamilton is known. To avoid circularity, we compute the determinant directly. We have

$$tI - C(f) = \begin{pmatrix} t & 0 & \cdots & 0 & a_0 \\ -1 & t & \cdots & 0 & a_1 \\ 0 & -1 & \cdots & 0 & a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & t + a_{n-1} \end{pmatrix}.$$

Expanding along the first row gives

$$\det(tI - C(f)) = tD_{n-1} + a_0,$$

where D_{n-1} is the determinant of the same form associated to

$$t^{n-1} + a_{n-1}t^{n-2} + \cdots + a_1.$$

The sign in the a_0 term is positive because the minor is lower triangular with diagonal entries all -1, contributing $(-1)^{n-1}$, and the cofactor sign contributes another $(-1)^{n+1}$. These signs multiply to 1.

By induction on n, $D_{n-1} = t^{n-1} + a_{n-1}t^{n-2} + \cdots + a_1$. Hence

$$\det(tI - C(f)) = t(t^{n-1} + a_{n-1}t^{n-2} + \cdots + a_1) + a_0 = f(t).$$

□

Proposition 8.20. *If T has invariant factors $f_1, \ldots, f_r$, then*

$$\chi_T(t) = f_1(t) \cdots f_r(t)$$

and

$$m_T(t) = f_r(t).$$

Proof. By rational canonical form, T has a block diagonal matrix with companion blocks $C(f_1), \ldots, C(f_r)$. The characteristic polynomial of a block diagonal matrix is the product of the characteristic polynomials of its diagonal blocks. By the preceding lemma, the characteristic polynomial of $C(f_i)$ is f_i. Therefore

$$\chi_T(t) = f_1(t) \cdots f_r(t).$$

The equality $m_T = f_r$ was proved above from the annihilator of the invariant factor decomposition. □

Corollary 8.21. *For every finite-dimensional endomorphism T,*

$$m_T(t) \mid \chi_T(t).$$

Proof. By the preceding proposition, $m_T = f_r$ and $\chi_T = f_1 \cdots f_r$. Since f_r is one of the factors in the product, m_T divides χ_T. □

Theorem 8.22 (Cayley–Hamilton theorem)**.** *For every finite-dimensional endomorphism T,*

$$\chi_T(T) = 0.$$

Proof. By the preceding corollary, the minimal polynomial m_T divides the characteristic polynomial χ_T. Write

$$\chi_T = m_T g$$

for some $g \in \mathbb{F}[x]$. Since $m_T(T) = 0$ by definition of the minimal polynomial,

$$\chi_T(T) = m_T(T) g(T) = 0.$$

Therefore T satisfies its characteristic polynomial. □