Chapter 6

Appendix

The five previous chapters were designed for a year undergraduate course in algebra. In this appendix, enough material is added to form a basic first year graduate course. Two of the main goals are to characterize finitely generated abelian groups and to prove the Jordan canonical form. The style is the same as before, i.e., everything is right down to the nub. The organization is mostly a linearly ordered sequence except for the last two sections on determinants and dual spaces. These are independent sections added on at the end.

Suppose R is a commutative ring. An R-module M is said to be cyclic if it can be generated by one element, i.e., $M \approx R/I$ where I is an ideal of R. The basic theorem of this chapter is that if R is a Euclidean domain and M is a finitely generated R-module, then M is the sum of cyclic modules. Thus if M is torsion free, it is a free R-module. Since $\mathbf{Z}$ is a Euclidean domain, finitely generated abelian groups are the sums of cyclic groups – one of the jewels of abstract algebra.

Now suppose F is a field and V is a finitely generated F-module. If $T : V \to V$ is a linear transformation, then V becomes an $F[x]$-module by defining $vx = T(v)$. Now $F[x]$ is a Euclidean domain and so $V_{F[x]}$ is the sum of cyclic modules. This classical and very powerful technique allows an easy proof of the canonical forms. There is a basis for V so that the matrix representing T is in Rational canonical form. If the characteristic polynomial of T factors into the product of linear polynomials, then there is a basis for V so that the matrix representing T is in Jordan canonical form. This always holds if $F = \mathbf{C}$. A matrix in Jordan form is a lower triangular matrix with the eigenvalues of T displayed on the diagonal, so this is a powerful concept.

In the chapter on matrices, it is stated without proof that the determinant of the product is the product of the determinants. A proof of this, which depends upon the classification of certain types of alternating multilinear forms, is given in this chapter. The final section gives the fundamentals of dual spaces.

The Chinese Remainder Theorem

On page 50 in the chapter on rings, the Chinese Remainder Theorem was proved for the ring of integers. In this section this classical topic is presented in full generality. Surprisingly, the theorem holds even for non-commutative rings.

Definition Suppose R is a ring and $A_1, A_2, ..., A_m$ are ideals of R. Then the *sum* $A_1 + A_2 + \cdots + A_m$ is the set of all $a_1 + a_2 + \cdots + a_m$ with $a_i \in A_i$. The *product* $A_1A_2 \cdots A_m$ is the set of all finite sums of elements $a_1a_2 \cdots a_m$ with $a_i \in A_i$. Note that the sum and product of ideals are ideals and $A_1A_2 \cdots A_m \subset (A_1 \cap A_2 \cap \cdots \cap A_m)$.

Definition Ideals A and B of R are said to be *comaximal* if $A + B = R$.

Theorem If A and B are ideals of a ring R, then the following are equivalent.

1) A and B are comaximal.
2) $\exists\, a \in A$ and $b \in B$ with $a + b = \underline{1}$.
3) $\pi(A) = R/B$ where $\pi : R \to R/B$ is the projection.

Theorem If $A_1, A_2, ..., A_m$ and B are ideals of R with A_i and B comaximal for each i, then $A_1A_2 \cdots A_m$ and B are comaximal. Thus $A_1 \cap A_2 \cap \cdots \cap A_m$ and B are comaximal.

Proof Consider $\pi : R \to R/B$. Then $\pi(A_1A_2 \cdots A_m) = \pi(A_1)\pi(A_2) \cdots \pi(A_m) = (R/B)(R/B) \cdots (R/B) = R/B$.

Chinese Remainder Theorem Suppose $A_1, A_2, ..., A_n$ are pairwise comaximal ideals of R, with each $A_i \neq R$. Then the natural map $\pi : R \to R/A_1 \times R/A_2 \times \cdots \times R/A_n$ is a surjective ring homomorphism with kernel $A_1 \cap A_2 \cap \cdots \cap A_n$.

Proof There exists $a_i \in A_i$ and $b_i \in A_1A_2 \cdots A_{i-1}A_{i+1} \cdots A_n$ with $a_i + b_i = \underline{1}$. Note that $\pi(b_i) = (0, .., 0, \underline{1}_i, 0, .., 0)$. If $(r_1 + A_1, r_2 + A_2, ..., r_n + A_n)$ is an element of the range, it is the image of $r_1b_1 + r_2b_2 + \cdots + r_nb_n = r_1(\underline{1} - a_1) + r_2(\underline{1} - a_2) + \cdots + r_n(\underline{1} - a_n)$.

Theorem If R is commutative and $A_1, A_2, ..., A_n$ are pairwise comaximal ideals of R, then $A_1A_2 \cdots A_n = A_1 \cap A_2 \cap \cdots \cap A_n$.

Proof for $n = 2$. Show $A_1 \cap A_2 \subset A_1A_2$. $\exists\, a_1 \in A_1$ and $a_2 \in A_2$ with $a_1 + a_2 = \underline{1}$. If $c \in A_1 \cap A_2$, then $c = c(a_1 + a_2) \in A_1A_2$.

Prime and Maximal Ideals and UFD^s

In the first chapter on background material, it was shown that $\mathbf{Z}$ is a unique factorization domain. Here it will be shown that this property holds for any principle ideal domain. Later on it will be shown that every Euclidean domain is a principle ideal domain. Thus every Euclidean domain is a unique factorization domain.

Definition Suppose R is a commutative ring and $I \subset R$ is an ideal.

I is *prime* means $I \neq R$ and if $a, b \in R$ have $ab \in I$, then a or $b \in I$.
I is *maximal* means $I \neq R$ and there are no ideals properly between I and R.

Theorem $\underline{0}$ is a prime ideal of R iff R is ________________
$\underline{0}$ is a maximal ideal of R iff R is ________________

Theorem Suppose $J \subset R$ is an ideal, $J \neq R$.
J is a prime ideal iff R/J is ________________
J is a maximal ideal iff R/J is ________________

Corollary Maximal ideals are prime.

Proof Every field is a domain.

Theorem If $a \in R$ is not a unit, then $\exists$ a maximal ideal I of R with $a \in I$.

Proof This is a classical application of the Hausdorff Maximality Principle. Consider $\{J : J$ is an ideal of R containing a with $J \neq R\}$. This collection contains a maximal monotonic collection $\{V_t\}_{t \in T}$. The ideal $V = \bigcup\limits_{t \in T} V_t$ does not contain $\underline{1}$ and thus is not equal to R. Therefore V is equal to some V_t and is a maximal ideal containing a.

Note To properly appreciate this proof, the student should work the exercise in group theory at the end of this section (see page 114).

Definition Suppose R is a domain and $a, b \in R$. Then we say $a \sim b$ iff there exists a unit u with $au = b$. Note that $\sim$ is an equivalence relation. If $a \sim b$, then a

and b are said to be *associates.*

Examples If R is a domain, the associates of $\underline{1}$ are the units of R, while the only associate of $\underline{0}$ is $\underline{0}$ itself. If $n \in \mathbf{Z}$ is not zero, then its associates are n and $-n$. If F is a field and $g \in F[x]$ is a non-zero polynomial, then the associates of g are all cg where c is a non-zero constant.

The following theorem is elementary, but it shows how associates fit into the scheme of things. An element a divides b $(a|b)$ if $\exists!$ $c \in R$ with $ac = b$.

Theorem Suppose R is a domain and $a, b \in (R - \underline{0})$. Then the following are equivalent.

1) $a \sim b$.
2) $a|b$ and $b|a$.
3) $aR = bR$.

Parts 1) and 3) above show there is a bijection from the associate classes of R to the principal ideals of R. Thus if R is a PID, there is a bijection from the associate classes of R to the ideals of R. If an element of a domain generates a non-zero prime ideal, it is called a prime element.

Definition Suppose R is a domain and $a \in R$ is a non-zero non-unit.

1) a is *irreducible* if it does not factor, i.e., $a = bc \Rightarrow b$ or c is a unit.
2) a is *prime* if it generates a prime ideal, i.e., $a|bc \Rightarrow a|b$ or $a|c$.

Note If a is a prime and $a|c_1c_2 \cdots c_n$, then $a|c_i$ for some i. This follows from the definition and induction on n. If each c_j is irreducible, then $a \sim c_i$ for some i.

Note If $a \sim b$, then a is irreducible (prime) iff b is irreducible (prime). In other words, if a is irreducible (prime) and u is a unit, then au is irreducible (prime).

Note a is prime $\Rightarrow a$ is irreducible. This is immediate from the definitions.

Theorem Factorization into primes is unique up to order and associates, i.e., if $d = b_1b_2 \cdots b_n = c_1c_2 \cdots c_m$ with each b_i and each c_i prime, then $n = m$ and for some permutation σ of the indices, b_i and $c_{\sigma(i)}$ are associates for every i. Note also $\exists$ a unit u and primes $p_1, p_2, \ldots, p_t$ where no two are associates and $du = p_1^{s_1} p_2^{s_2} \cdots p_t^{s_t}$.

Proof This follows from the notes above.

Definition R is a *factorization domain* (FD) means that R is a domain and if a is a non-zero non-unit element of R, then a factors into a finite product of irreducibles.

Definition R is a *unique factorization domain* (UFD) means R is a FD in which factorization is unique (up to order and associates).

Theorem If R is a UFD and a is a non-zero non-unit of R, then a is irreducible $\Leftrightarrow$ a is prime. Thus in a UFD, elements factor as the product of primes.

Proof Suppose R is a UFD, a is an irreducible element of R, and $a|bc$. If either b or c is a unit or is zero, then a divides one of them, so suppose each of b and c is a non-zero non-unit element of R. There exists an element d with $ad = bc$. Each of b and c factors as the product of irreducibles and the product of these products is the factorization of bc. It follows from the uniqueness of the factorization of $ad = bc$, that one of these irreducibles is an associate of a, and thus $a|b$ or $a|c$. Therefore the element a is a prime.

Theorem Suppose R is a FD. Then the following are equivalent.

1) R is a UFD.
2) Every irreducible element of R is prime, i.e., a irreducible $\Leftrightarrow$ a is prime.

Proof We already know 1) $\Rightarrow$ 2). Part 2) $\Rightarrow$ 1) because factorization into primes is always unique.

This is a revealing and useful theorem. If R is a FD, then R is a UFD iff each irreducible element generates a prime ideal. Fortunately, principal ideal domains have this property, as seen in the next theorem.

Theorem Suppose R is a PID and $a \in R$ is non-zero non-unit. Then the following are equivalent.

1) aR is a maximal ideal.
2) aR is a prime ideal, i.e., a is a prime element.
3) a is irreducible.

Proof Every maximal ideal is a prime ideal, so 1) $\Rightarrow$ 2). Every prime element is an irreducible element, so 2) $\Rightarrow$ 3). Now suppose a is irreducible and show aR is a maximal ideal. If I is an ideal containing aR, $\exists\ b \in R$ with $I = bR$. Since b divides a, the element b is a unit or an associate of a. This means $I = R$ or $I = aR$.

Our goal is to prove that a PID is a UFD. Using the two theorems above, it only remains to show that a PID is a FD. The proof will not require that ideals be principally generated, but only that they be finitely generated. This turns out to be equivalent to the property that any collection of ideals has a "maximal" element. We shall see below that this is a useful concept which fits naturally into the study of unique factorization domains.

Theorem Suppose R is a commutative ring. Then the following are equivalent.

1) If $I \subset R$ is an ideal, $\exists$ a finite set $\{a_1, a_2, ..., a_n\} \subset R$ such that $I = a_1R + a_2R + \cdots + a_nR$, i.e., each ideal of R is finitely generated.
2) Any non-void collection of ideals of R contains an ideal I which is maximal in the collection. This means if J is an ideal in the collection with $J \supset I$, then $J = I$. (The ideal I is maximal only in the sense described. It need not contain all the ideals of the collection, nor need it be a maximal ideal of the ring R.)
3) If $I_1 \subset I_2 \subset I_3 \subset ...$ is a monotonic sequence of ideals, $\exists\ t_0 \geq 1$ such that $I_t = I_{t_0}$ for all $t \geq t_0$.

Proof Suppose 1) is true and show 3). The ideal $I = I_1 \cup I_2 \cup \ldots$ is finitely generated and $\exists\ t_0 \geq 1$ such that I_{t_0} contains those generators. Thus 3) is true. Now suppose 2) is true and show 1). Let I be an ideal of R, and consider the collection of all finitely generated ideals contained in I. By 2) there is a maximal one, and it must be I itself, and thus 1) is true. We now have 2)$\Rightarrow$1)$\Rightarrow$3), so suppose 2) is false and show 3) is false. So there is a collection of ideals of R such that any ideal in the collection is properly contained in another ideal of the collection. Thus it is possible to construct a sequence of ideals $I_1 \subset I_2 \subset I_3 \ldots$ with each properly contained in the next, and therefore 3) is false. (Actually this construction requires the Hausdorff Maximality Principle or some form of the Axiom of Choice, but we slide over that.)

Definition If R satisfies these properties, R is said to be *Noetherian*, or it is said to satisfy the *ascending chain condition*. This property is satisfied by many of the classical rings in mathematics. Having three definitions makes this property useful and easy to use. For example, see the next theorem.

Theorem A Noetherian domain is a FD. In particular, a PID is a FD.

Proof Suppose there is a non-zero non-unit element that does not factor as the finite product of irreducibles. Consider all ideals dR where d does not factor. Since R is Noetherian, $\exists$ a maximal one cR. The element c must be reducible, i.e., $c = ab$ where neither a nor b is a unit. Each of aR and bR properly contains cR, and so each of a and b factors as a finite product of irreducibles. This gives a finite factorization of c into irreducibles, which is a contradiction.

Corollary A PID is a UFD. So $\mathbf{Z}$ is a UFD and if F is a field, $F[x]$ is a UFD.

You see the basic structure of UFD^s is quite easy. It takes more work to prove the following theorems, which are stated here only for reference.

Theorem If R is a UFD then $R[x_1, ..., x_n]$ is a UFD. Thus if F is a field, $F[x_1, ..., x_n]$ is a UFD. (This theorem goes all the way back to Gauss.)

If R is a PID, then the formal power series $R[[x_1, ..., x_n]]$ is a UFD. Thus if F is a field, $F[[x_1, ..., x_n]]$ is a UFD. (There is a UFD R where $R[[x]]$ is not a UFD. See page 566 of *Commutative Algebra* by N. Bourbaki.)

Theorem Germs of analytic functions on $\mathbf{C}^n$ form a UFD.

Proof See Theorem 6.6.2 of *An Introduction to Complex Analysis in Several Variables* by L. Hörmander.

Theorem Suppose R is a commutative ring. Then R is Noetherian $\Rightarrow R[x_1, ..., x_n]$ and $R[[x_1, ..., x_n]]$ are Noetherian. (This is the famous *Hilbert Basis Theorem.*)

Theorem If R is Noetherian and $I \subset R$ is a proper ideal, then R/I is Noetherian. (This follows immediately from the definition. This and the previous theorem show that Noetherian is a ubiquitous property in ring theory.)

Domains With Non-unique Factorizations Next are presented two of the standard examples of Noetherian domains that are not unique factorization domains.

Exercise Let $R = \mathbf{Z}(\sqrt{5}) = \{n + m\sqrt{5} : n, m \in \mathbf{Z}\}$. Show that R is a subring of $\mathbf{R}$ which is not a UFD. In particular $2 \cdot 2 = (1 - \sqrt{5}) \cdot (-1 - \sqrt{5})$ are two distinct irreducible factorizations of 4. Show R is isomorphic to $\mathbf{Z}[x]/(x^2 - 5)$, where $(x^2 - 5)$ represents the ideal $(x^2 - 5)\mathbf{Z}[x]$, and $R/(2)$ is isomorphic to $\mathbf{Z}_2[x]/(x^2 - [5]) = \mathbf{Z}_2[x]/(x^2 + [1])$, which is not a domain.

Exercise Let $R = \mathbf{R}[x, y, z]/(x^2 - yz)$. Show $x^2 - yz$ is irreducible and thus prime in $\mathbf{R}[x, y, z]$. If $u \in \mathbf{R}[x, y, z]$, let $\bar{u} \in R$ be the coset containing u. Show R is not a UFD. In particular $\bar{x} \cdot \bar{x} = \bar{y} \cdot \bar{z}$ are two distinct irreducible factorizations of $\bar{x}^2$. Show $R/(\bar{x})$ is isomorphic to $\mathbf{R}[y, z]/(yz)$, which is not a domain. An easier approach is to let $f : \mathbf{R}[x, y, z] \to \mathbf{R}[x, y]$ be the ring homomorphism defined by $f(x) = xy$, $f(y) = x^2$, and $f(z) = y^2$. Then $S = \mathbf{R}[xy, x^2, y^2]$ is the image of f and S is isomorphic to R. Note that xy, x^2, and y^2 are irreducible in S and $(xy)(xy) = (x^2)(y^2)$ are two distinct irreducible factorizations of $(xy)^2$ in S.

Exercise In Group Theory If G is an additive abelian group, a subgroup H of G is said to be maximal if $H \neq G$ and there are no subgroups properly between H and G. Show that H is maximal iff $G/H \approx \mathbf{Z}_p$ for some prime p. For simplicity, consider the case $G = \mathbf{Q}$. Which one of the following is true?

1) If $a \in \mathbf{Q}$, then there is a maximal subgroup H of $\mathbf{Q}$ which contains a.
2) $\mathbf{Q}$ contains no maximal subgroups.

Splitting Short Exact Sequences

Suppose B is an R-module and K is a submodule of B. As defined in the chapter on linear algebra, K is a summand of B provided $\exists$ a submodule L of B with $K + L = B$ and $K \cap L = \underline{0}$. In this case we write $K \oplus L = B$. When is K a summand of B? It turns out that K is a summand of B iff there is a splitting map from B/K to B. In particular, if B/K is free, K must be a summand of B. This is used below to show that if R is a PID, then every submodule of R^n is free.

Theorem 1 Suppose R is a ring, B and C are R-modules, and $g : B \to C$ is a surjective homomorphism with kernel K. Then the following are equivalent.

1) K is a summand of B.
2) g has a right inverse, i.e., $\exists$ a homomorphism $h : C \to B$ with $g \circ h = I : C \to C$. (h is called a *splitting map*.)

Proof Suppose 1) is true, i.e., suppose $\exists$ a submodule L of B with $K \oplus L = B$. Then $(g|L) : L \to C$ is an isomorphism. If $i : L \to B$ is inclusion, then h defined by $h = i \circ (g|L)^{-1}$ is a right inverse of g. Now suppose 2) is true and $h : C \to B$ is a right inverse of g. Then h is injective, $K + h(C) = B$ and $K \cap h(C) = \underline{0}$. Thus $K \oplus h(C) = B$.

Definition Suppose $f : A \to B$ and $g : B \to C$ are R-module homomorphisms. The statement that $0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$ is a *short exact sequence* (s.e.s) means f is injective, g is surjective and $f(A) = \ker(g)$. The canonical split s.e.s. is $A \to A \oplus C \to C$ where $f = i_1$ and $g = \pi_2$. A short exact sequence is said to split if $\exists$ an isomorphism $B \xrightarrow{\approx} A \oplus C$ such that the following diagram commutes.

$$\begin{tikzcd}
0 \arrow[r] & A \arrow[r, "f"] \arrow[dr, "i_1"'] & B \arrow[r, "g"] \arrow[d, "\approx"] & C \arrow[r] & 0 \\
& & A \oplus C \arrow[ur, "\pi_2"'] & &
\end{tikzcd}$$

We now restate the previous theorem in this terminology.

Theorem 1.1 A short exact sequence $0 \to A \to B \to C \to 0$ splits iff $f(A)$ is a summand of B, iff $B \to C$ has a splitting map. If C is a free R-module, there is a splitting map and thus the sequence splits.

Proof We know from the previous theorem $f(A)$ is a summand of B iff $B \to C$ has a splitting map. Showing these properties are equivalent to the splitting of the sequence is a good exercise in the art of diagram chasing. Now suppose C has a free basis $T \subset C$, and $g : B \to C$ is surjective. There exists a function $h : T \to B$ such that $g \circ h(c) = c$ for each $c \in T$. The function h extends to a homomorphism from C to B which is a right inverse of g.

Theorem 2 If R is a domain, then the following are equivalent.

1) R is a PID.
2) Every submodule of R_R is a free R-module of dimension ≤ 1.

This theorem restates the ring property of PID as a module property. Although this theorem is transparent, 1)$\Rightarrow$2) is a precursor to the following classical result.

Theorem 3 If R is a PID and $A \subset R^n$ is a submodule, then A is a free R-module of dimension $\leq n$. Thus subgroups of $\mathbf{Z}^n$ are free $\mathbf{Z}$-modules of dimension $\leq n$.

Proof From the previous theorem we know this is true for $n = 1$. Suppose $n > 1$ and the theorem is true for submodules of R^{n-1}. Suppose $A \subset R^n$ is a submodule.

Consider the following short exact sequences, where $f : R^{n-1} \to R^{n-1} \oplus R$ is inclusion and $g = \pi : R^{n-1} \oplus R \to R$ is the projection.

$$0 \longrightarrow R^{n-1} \xrightarrow{f} R^{n-1} \oplus R \xrightarrow{\pi} R \longrightarrow 0$$

$$0 \longrightarrow A \cap R^{n-1} \longrightarrow A \longrightarrow \pi(A) \longrightarrow 0$$

By induction, $A \cap R^{n-1}$ is free of dimension $\leq n-1$. If $\pi(A) = \underline{0}$, then $A \subset R^{n-1}$. If $\pi(A) \neq \underline{0}$, it is free of dimension 1 and thus the sequence splits by Theorem 1.1. In either case, A is a free submodule of dimension $\leq n$.

Exercise Let $A \subset \mathbf{Z}^2$ be the subgroup generated by $\{(6, 24), (16, 64)\}$. Show A is a free $\mathbf{Z}$-module of dimension 1. Also show the s.e.s. $\mathbf{Z}_4 \xrightarrow{\times 3} \mathbf{Z}_{12} \longrightarrow \mathbf{Z}_3$ splits but $\mathbf{Z} \xrightarrow{\times 2} \mathbf{Z} \longrightarrow \mathbf{Z}_2$ and $\mathbf{Z}_2 \xrightarrow{\times 2} \mathbf{Z}_4 \longrightarrow \mathbf{Z}_2$ do not (see top of page 78).

Euclidean Domains

The ring $\mathbf{Z}$ possesses the Euclidean algorithm and the polynomial ring $F[x]$ has the division algorithm (pages 14 and 45). The concept of *Euclidean domain* is an abstraction of these properties, and the efficiency of this abstraction is displayed in this section. Furthermore the first axiom, $\phi(a) \leq \phi(ab)$, is used only in Theorem 2, and is sometimes omitted from the definition. Anyway it is possible to just play around with matrices and get some deep results. If R is a Euclidean domain and M is a finitely generated R-module, then M is the sum of cyclic modules. This is one of the great classical theorems of abstract algebra, and you don't have to worry about it becoming obsolete. Here $\mathbf{N}$ will denote the set of all non-negative integers, not just the set of positive integers.

Definition A domain R is a *Euclidean domain* provided $\exists\ \phi : (R - \underline{0}) \longrightarrow \mathbf{N}$ such that if $a, b \in (R - \underline{0})$, then

1) $\phi(a) \leq \phi(ab)$.
2) $\exists\ q, r \in R$ such that $a = bq + r$ with $r = \underline{0}$ or $\phi(r) < \phi(b)$.

Examples of Euclidean Domains

$\mathbf{Z}$ with $\phi(n) = |n|$.
A field F with $\phi(a) = 1\ \ \forall\ a \neq \underline{0}$ or with $\phi(a) = 0\ \ \forall\ a \neq \underline{0}$.
$F[x]$ where F is a field with $\phi(f = a_0 + a_1x + \cdots + a_nx^n) = \deg(f)$.
$\mathbf{Z}[i] = \{a + bi : a, b \in \mathbf{Z}\} =$ Gaussian integers with $\phi(a + bi) = a^2 + b^2$.

Theorem 1 If R is a Euclidean domain, then R is a PID and thus a UFD.

Proof If I is a non-zero ideal, then $\exists\ b \in I - \underline{0}$ satisfying $\phi(b) \leq \phi(a)\ \forall\ a \in I - \underline{0}$. Then b generates I because if $a \in I - \underline{0}$, $\exists\ q, r$ with $a = bq + r$. Now $r \in I$ and $r \neq \underline{0} \Rightarrow \phi(r) < \phi(b)$ which is impossible. Thus $r = \underline{0}$ and $a \in bR$ so $I = bR$.

Theorem 2 If R is a Euclidean domain and $a, b \in R - \underline{0}$, then

$\phi(\underline{1})$ is the smallest integer in the image of ϕ.
a is a unit in R iff $\phi(a) = \phi(\underline{1})$.
a and b are associates $\Rightarrow \phi(a) = \phi(b)$.

Proof This is a good exercise. However it is unnecessary for Theorem 3 below.

The following remarkable theorem is the foundation for the results of this section.

Theorem 3 If R is a Euclidean domain and $(a_{i,j}) \in R_{n,t}$ is a non-zero matrix, then by elementary row and column operations $(a_{i,j})$ can be transformed to

$$\begin{pmatrix} d_1 & 0 & \cdots & & & 0 \\ 0 & d_2 & & & & \\ \vdots & & \ddots & & & \\ & & & d_m & & \\ & & & & 0 & \\ & & & & & \\ 0 & & & & & 0 \end{pmatrix}$$

where each $d_i \neq \underline{0}$, and $d_i | d_{i+1}$ for $1 \leq i < m$. Also d_1 generates the ideal of R generated by the entries of $(a_{i,j})$.

Proof Let $I \subset R$ be the ideal generated by the elements of the matrix $A = (a_{i,j})$. If $E \in R_n$, then the ideal J generated by the elements of EA has $J \subset I$. If E is invertible, then $J = I$. In the same manner, if $E \in R_t$ is invertible and J is the ideal generated by the elements of AE, then $J = I$. This means that row and column operations on A do not change the ideal I. Since R is a PID, there is an element d_1 with $I = d_1 R$, and this will turn out to be the d_1 displayed in the theorem.

The matrix $(a_{i,j})$ has at least one non-zero element d with $\phi(d)$ a miminum. However, row and column operations on $(a_{i,j})$ may produce elements with smaller

ϕ values. To consolidate this approach, consider matrices obtained from $(a_{i,j})$ by a finite number of row and column operations. Among these, let $(b_{i,j})$ be one which has an entry $d_1 \neq 0$ with $\phi(d_1)$ a minimum. By elementary operations of type 2, the entry d_1 may be moved to the $(1,1)$ place in the matrix. Then d_1 will divide the other entries in the first row, else we could obtain an entry with a smaller ϕ value. Thus by column operations of type 3, the other entries of the first row may be made zero. In a similar manner, by row operations of type 3, the matrix may be changed to the following form.

$$\begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & c_{ij} & \\ 0 & & & \end{pmatrix}$$

Note that d_1 divides each $c_{i,j}$, and thus $I = d_1 R$. The proof now follows by induction on the size of the matrix.

This is an example of a theorem that is easy to prove playing around at the blackboard. Yet it must be a deep theorem because the next two theorems are easy consequences.

Theorem 4 Suppose R is a Euclidean domain, B is a finitely generated free R-module and $A \subset B$ is a non-zero submodule. Then $\exists$ free bases $\{a_1, a_2, ..., a_t\}$ for A and $\{b_1, b_2, ..., b_n\}$ for B, with $t \leq n$, and such that each $a_i = d_i b_i$, where each $d_i \neq \underline{0}$, and $d_i | d_{i+1}$ for $1 \leq i < t$. Thus $B/A \approx R/d_1 \oplus R/d_2 \oplus \cdots \oplus R/d_t \oplus R^{n-t}$.

Proof By Theorem 3 in the section Splitting Short Exact Sequences, A has a free basis $\{v_1, v_2, ..., v_t\}$. Let $\{w_1, w_2, ..., w_n\}$ be a free basis for B, where $n \geq t$. The composition

$$R^t \xrightarrow{\approx} A \xrightarrow{\subset} B \xrightarrow{\approx} R^n$$

$$e_i \longrightarrow v_i \qquad w_i \longrightarrow e_i$$

is represented by a matrix $(a_{i,j}) \in R_{n,t}$ where $v_i = a_{1,i}w_1 + a_{2,i}w_2 + \cdots + a_{n,i}w_n$. By the previous theorem, $\exists$ invertible matrixes $U \in R_n$ and $V \in R_t$ such that

$$U(a_{i,j})V \;=\; \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \\ \vdots & 0 & \ddots & \\ & & & d_t \\ & & & \\ 0 & \cdots & & 0 \end{pmatrix}$$

with $d_i | d_{i+1}$. Since changing the isomorphisms $R^t \xrightarrow{\approx} A$ and $B \xrightarrow{\approx} R^n$ corresponds to changing the bases $\{v_1, v_2, ..., v_t\}$ and $\{w_1, w_2, ..., w_n\}$, the theorem follows.

Theorem 5 If R is a Euclidean domain and M is a finitely generated R-module, then $M \approx R/d_1 \oplus R/d_2 \oplus \cdots \oplus R/d_t \oplus R^m$ where each $d_i \neq \underline{0}$, and $d_i | d_{i+1}$ for $1 \leq i < t$.

Proof By hypothesis $\exists$ a finitely generated free module B and a surjective homomorphism $B \longrightarrow M \longrightarrow 0$. Let A be the kernel, so $0 \longrightarrow A \xrightarrow{\subset} B \longrightarrow M \longrightarrow 0$ is a s.e.s. and $B/A \approx M$. The result now follows from the previous theorem.

The way Theorem 5 is stated, some or all of the elements d_i may be units, and for such d_i, $R/d_i = \underline{0}$. If we assume that no d_i is a unit, then the elements $d_1, d_2, ..., d_t$ are called *invariant factors*. They are unique up to associates, but we do not bother with that here. If $R = \mathbf{Z}$ and we select the d_i to be positive, they are unique. If $R = F[x]$ and we select the d_i to be monic, then they are unique. The splitting in Theorem 5 is not the ultimate because the modules R/d_i may split into the sum of other cyclic modules. To prove this we need the following Lemma.

Lemma Suppose R is a PID and b and c are non-zero non-unit elements of R. Suppose b and c are relatively prime, i.e., there is no prime common to their prime factorizations. Then bR and cR are comaximal ideals. (See p 108 for comaximal.)

Proof There exists an $a \in R$ with $aR = bR + cR$. Since $a|b$ and $a|c$, a is a unit, so $R = bR + cR$.

Theorem 6 Suppose R is a PID and d is a non-zero non-unit element of R. Assume $d = p_1^{s_1} p_2^{s_2} \cdots p_t^{s_t}$ is the prime factorization of d (see bottom of p 110). Then the natural map $R/d \xrightarrow{\approx} R/p_1^{s_1} \oplus \cdots \oplus R/p_t^{s_t}$ is an isomorphism of R-modules. (The elements $p_i^{s_i}$ are called *elementary divisors* of R/d.)

Proof If $i \neq j$, $p_i^{s_i}$ and $p_j^{s_j}$ are relatively prime. By the Lemma above, they are comaximal and thus by the Chinese Remainder Theorem, the natural map is a ring isomorphism (page 108). Since the natural map is also an R-module homomorphism, it is an R-module isomorphism.

This theorem carries the splitting as far as it can go, as seen by the next exercise.

Exercise Suppose R is a PID, $p \in R$ is a prime element, and $s \geq 1$. Then the R-module R/p^s has no proper submodule which is a summand.

Torsion Submodules This will give a little more perspective to this section.

Definition Suppose M is a module over a domain R. An element $m \in M$ is said to be a *torsion element* if $\exists\ r \in R$ with $r \neq \underline{0}$ and $mr = \underline{0}$. This is the same as saying m is dependent. If $R = \mathbf{Z}$, it is the same as saying m has finite order. Denote by $T(M)$ the set of all torsion elements of M. If $T(M) = \underline{0}$, we say that M is torsion free.

Theorem 7 Suppose M is a module over a domain R. Then $T(M)$ is a submodule of M and $M/T(M)$ is torsion free.

Proof This is a simple exercise.

Theorem 8 Suppose R is a Euclidean domain and M is a finitely generated R-module which is torsion free. Then M is a free R-module, i.e., $M \approx R^m$.

Proof This follows immediately from Theorem 5.

Theorem 9 Suppose R is a Euclidean domain and M is a finitely generated R-module. Then the following s.e.s. splits.

$$0 \longrightarrow T(M) \longrightarrow M \longrightarrow M/T(M) \longrightarrow 0$$

Proof By Theorem 7, $M/T(M)$ is torsion free. By Theorem 8, $M/T(M)$ is a free R-module, and thus there is a splitting map. Of course this theorem is transparent anyway, because Theorem 5 gives a splitting of M into a torsion part and a free part.

Note It follows from Theorem 9 that $\exists$ a free submodule V of M such that $T(M) \oplus V = M$. The first summand $T(M)$ is unique, but the complementary summand V is not unique. V depends upon the splitting map and is unique only up to isomorphism.

To complete this section, here are two more theorems that follow from the work we have done.

Theorem 10 Suppose T is a domain and T^* is the multiplicative group of units of T. If G is a finite subgroup of T^*, then G is a cyclic group. Thus if F is a finite field, the multiplicative group F^* is cyclic. Thus if p is a prime, $(\mathbf{Z}_p)^*$ is cyclic.

Proof This is a corollary to Theorem 5 with $R = \mathbf{Z}$. The multiplicative group G is isomorphic to an additive group $\mathbf{Z}/d_1 \oplus \mathbf{Z}/d_2 \oplus \cdots \oplus \mathbf{Z}/d_t$ where each $d_i > 1$ and $d_i | d_{i+1}$ for $1 \le i < t$. Every u in the additive group has the property that $ud_t = \underline{0}$. So every $g \in G$ is a solution to $x^{d_t} - \underline{1} = \underline{0}$. If $t > 1$, the equation will have degree less than the number of roots, which is impossible. Thus $t = 1$ and so G is cyclic.

Exercise For which primes p and q is the group of units $(\mathbf{Z}_p \times \mathbf{Z}_q)^*$ a cyclic group?

We know from Exercise 2) on page 59 that an invertible matrix over a field is the product of elementary matrices. This result also holds for any invertible matrix over a Euclidean domain.

Theorem 11 Suppose R is a Euclidean domain and $A \in R_n$ is a matrix with non-zero determinant. Then by elementary row and column operations, A may be transformed to a diagonal matrix

$$\begin{pmatrix} d_1 & & & 0 \\ & d_2 & & \\ & & \ddots & \\ 0 & & & d_n \end{pmatrix}$$

where each $d_i \neq \underline{0}$ and $d_i | d_{i+1}$ for $1 \le i < n$. Also d_1 generates the ideal generated by the entries of A. Furthermore A is invertible iff each d_i is a unit. Thus if A is invertible, A is the product of elementary matrices.

Proof It follows from Theorem 3 that A may be transformed to a diagonal matrix with $d_i|d_{i+1}$. Since the determinant of A is not zero, it follows that each $d_i \neq \underline{0}$. Furthermore, the matrix A is invertible iff the diagonal matrix is invertible, which is true iff each d_i is a unit. If each d_i is a unit, then the diagonal matrix is the product of elementary matrices of type 1. Therefore if A is invertible, it is the product of elementary matrices.

Exercise Let $R = \mathbf{Z}$, $A = \begin{pmatrix} 3 & 11 \\ 0 & 4 \end{pmatrix}$ and $D = \begin{pmatrix} 3 & 11 \\ 1 & 4 \end{pmatrix}$. Perform elementary operations on A and D to obtain diagonal matrices where the first diagonal element divides the second diagonal element. Write D as the product of elementary matrices. Find the characteristic polynomials of A and D. Find an elementary matrix B over $\mathbf{Z}$ such that $B^{-1}AB$ is diagonal. Find an invertible matrix C in $\mathbf{R}_2$ such that $C^{-1}DC$ is diagonal. Show C cannot be selected in $\mathbf{Q}_2$.

Jordan Blocks

In this section, we define the two special types of square matrices used in the Rational and Jordan canonical forms. Note that the Jordan block $B(q)$ is the sum of a scalar matrix and a nilpotent matrix. A Jordan block displays its eigenvalue on the diagonal, and is more interesting than the companion matrix $C(q)$. But as we shall see later, the Rational canonical form will always exist, while the Jordan canonical form will exist iff the characteristic polynomial factors as the product of linear polynomials.

Suppose R is a commutative ring, $q = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + x^n \in R[x]$ is a monic polynomial of degree $n \geq 1$, and V is the $R[x]$-module $V = R[x]/q$. V is a torsion module over the ring $R[x]$, but as an R-module, V has a free basis $\{1, x, x^2, \ldots, x^{n-1}\}$. (See the last part of the last theorem on page 46.) Multiplication by x defines an R-module endomorphism on V, and $C(q)$ will be the matrix of this endomorphism with respect to this basis. Let $T : V \to V$ be defined by $T(v) = vx$. If $h(x) \in R[x]$, $h(T)$ is the R-module homomorphism given by multiplication by $h(x)$. The homomorphism from $R[x]/q$ to $R[x]/q$ given by multiplication by $h(x)$, is zero iff $h(x) \in qR[x]$. That is to say $q(T) = a_0I + a_1T + \cdots + T^n$ is the zero homomorphism, and $h(T)$ is the zero homomorphism iff $h(x) \in qR[x]$. All of this is supposed to make the next theorem transparent.

Theorem Let V have the free basis $\{1, x, x^2, ..., x^{n-1}\}$. The companion matrix representing T is

$$C(q) = \begin{pmatrix} 0 & \ldots & \ldots & 0 & -a_0 \\ 1 & 0 & \ldots & 0 & -a_1 \\ 0 & 1 & 0 & & -a_2 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \ldots & \ldots & 1 & -a_{n-1} \end{pmatrix}$$

The characteristic polynomial of $C(q)$ is q, and $|C(q)| = (-1)^n a_0$. Finally, if $h(x) \in R[x]$, $h(C(q))$ is zero iff $h(x) \in qR[x]$.

Theorem Suppose $\lambda \in R$ and $q(x) = (x-\lambda)^n$. Let V have the free basis $\{1, (x-\lambda), (x-\lambda)^2, \ldots, (x-\lambda)^{n-1}\}$. Then the matrix representing T is

$$B(q) = \begin{pmatrix} \lambda & 0 & \ldots & \ldots & 0 \\ 1 & \lambda & 0 & \ldots & 0 \\ 0 & 1 & \lambda & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \ldots & \ldots & 1 & \lambda \end{pmatrix}$$

The characteristic polynomial of $B(q)$ is q, and $|B(q)| = \lambda^n = (-1)^n a_0$. Finally, if $h(x) \in R[x]$, $h(B(q))$ is zero iff $h(x) \in qR[x]$.

Note For $n = 1$, $C(a_0 + x) = B(a_0 + x) = (-a_0)$. This is the only case where a block matrix may be the zero matrix.

Note In $B(q)$, if you wish to have the 1^s above the diagonal, reverse the order of the basis for V.

Jordan Canonical Form

We are finally ready to prove the Rational and Jordan forms. Using the previous sections, all that's left to do is to put the pieces together. (For an overview of Jordan form, read first the section in Chapter 5, page 96.)

Suppose R is a commutative ring, V is an R-module, and $T : V \to V$ is an R-module homomorphism. Define a scalar multiplication $V \times R[x] \to V$ by $v(a_0 + a_1 x + \cdots + a_r x^r) = va_0 + T(v)a_1 + \cdots + T^r(v)a_r$.

Theorem 1 Under this scalar multiplication, V is an $R[x]$-module.

This is just an observation, but it is one of the great tricks in mathematics. Questions about the transformation T are transferred to questions about the module V over the ring $R[x]$. And in the case R is a field, $R[x]$ is a Euclidean domain and so we know almost everything about V as an $R[x]$-module.

Now in this section, we suppose R is a field F, V is a finitely generated F-module, $T : V \to V$ is a linear transformation and V is an $F[x]$-module with $vx = T(v)$. Our goal is to select a basis for V such that the matrix representing T is in some simple form. A submodule of $V_{F[x]}$ is a submodule of V_F which is invariant under T. We know $V_{F[x]}$ is the sum of cyclic modules from Theorems 5 and 6 in the section on Euclidean Domains. Since V is finitely generated as an F-module, the free part of this decomposition will be zero. In the section on Jordan Blocks, a basis is selected for these cyclic modules and the matrix representing T is described. This gives the Rational Canonical Form and that is all there is to it. If all the eigenvalues for T are in F, we pick another basis for each of the cyclic modules (see the second theorem in the section on Jordan Blocks). Then the matrix representing T is called the Jordan Canonical Form. Now we say all this again with a little more detail.

From Theorem 5 in the section on Euclidean Domains, it follows that

$$V_{F[x]} \approx F[x]/d_1 \oplus F[x]/d_2 \oplus \cdots \oplus F[x]/d_t$$

where each d_i is a monic polynomial of degree ≥ 1, and $d_i | d_{i+1}$. Pick $\{1, x, x^2, \ldots, x^{m-1}\}$ as the F-basis for $F[x]/d_i$ where m is the degree of the polynomial d_i.

Theorem 2 With respect to this basis, the matrix representing T is

$$\begin{pmatrix} C(d_1) & & & \\ & C(d_2) & & \\ & & \ddots & \\ & & & C(d_t) \end{pmatrix}$$

The characteristic polynomial of T is $p = d_1 d_2 \cdots d_t$ and $p(T) = \underline{0}$. This is a type of canonical form but it does not seem to have a name.

Now we apply Theorem 6 to each $F[x]/d_i$. This gives $V_{F[x]} \approx F[x]/p_1^{s_1} \oplus \cdots \oplus F[x]/p_r^{s_r}$ where the p_i are irreducible monic polynomials of degree at least 1. The p_i need not be distinct. Pick an F-basis for each $F[x]/p_i^{s_i}$ as before.

Theorem 3 With respect to this basis, the matrix representing T is

$$\begin{pmatrix} C(p_1^{s_1}) & & & \\ & C(p_2^{s_2}) & & 0 \\ & & & \\ 0 & & \ddots & \\ & & & C(p_r^{s_r}) \end{pmatrix}$$

The characteristic polynomial of T is $p = p_1^{s_1} \cdots p_r^{s_r}$ and $p(T) = \underline{0}$. This is called the *Rational canonical form* for T.

Now suppose the characteristic polynomial of T factors in $F[x]$ as the product of linear polynomials. Thus in the Theorem above, $p_i = x - \lambda_i$ and

$$V_{F[x]} \approx F[x]/(x-\lambda_1)^{s_1} \oplus \cdots \oplus F[x]/(x-\lambda_r)^{s_r}$$

is an isomorphism of $F[x]$-modules. Pick $\{1, (x-\lambda_i), (x-\lambda_i)^2, \ldots, (x-\lambda_i)^{m-1}\}$ as the F-basis for $F[x]/(x-\lambda_i)^{s_i}$ where m is s_i.

Theorem 4 With respect to this basis, the matrix representing T is

$$\begin{pmatrix} B((x-\lambda_1)^{s_1}) & & & \\ & & & 0 \\ & B((x-\lambda_2)^{s_2}) & & \\ & & & \\ 0 & & \ddots & \\ & & & \\ & & & B((x-\lambda_r)^{s_r}) \end{pmatrix}$$

The characteristic polynomial of T is $p = (x - \lambda_1)^{s_1} \cdots (x - \lambda_r)^{s_r}$ and $p(T) = \underline{0}$. This is called the *Jordan canonical form* for T. Note that the λ_i need not be distinct.

Note A diagonal matrix is in Rational canonical form and in Jordan canonical form. This is the case where each block is one by one. Of course a diagonal matrix is about as canonical as you can get. Note also that if a matrix is in Jordan form, its trace is the sum of the eigenvalues and its determinant is the product of the eigenvalues. Finally, this section is loosely written, so it is important to use the transpose principle to write three other versions of the last two theorems.

Exercise Suppose F is a field of characteristic 0 and $T \in F_n$ has trace$(T^i) = \underline{0}$ for $0 < i \leq n$. Show T is nilpotent. Let $p \in F[x]$ be the characteristic polynomial of T. The polynomial p may not factor into linears in $F[x]$, and thus T may have no conjugate in F_n which is in Jordan form. However this exercise can still be worked using Jordan form. This is based on the fact that there exists a field $\bar{F}$ containing F as a subfield, such that p factors into linears in $\bar{F}[x]$. This fact is not proved in this book, but it is assumed for this exercise. So $\exists$ an invertible matrix $U \in \bar{F}_n$ so that $U^{-1}TU$ is in Jordan form, and of course, T is nilpotent iff $U^{-1}TU$ is nilpotent. The point is that it sufficies to consider the case where T is in Jordan form, and to show the diagonal elements are all zero.

So suppose T is in Jordan form and trace $(T^i) = \underline{0}$ for $1 \leq i \leq n$. Thus trace $(p(T)) = a_0 n$ where a_0 is the constant term of $p(x)$. We know $p(T) = \underline{0}$ and thus trace $(p(T)) = \underline{0}$, and thus $a_0 n = \underline{0}$. Since the field has characteristic 0, $a_0 = \underline{0}$ and so $\underline{0}$ is an eigenvalue of T. This means that one block of T is a strictly lower triangular matrix. Removing this block leaves a smaller matrix which still satisfies the hypothesis, and the result follows by induction on the size of T. This exercise illustrates the power and facility of Jordan form. It also has a cute corollary.

Corollary Suppose F is a field of characteristic 0, $n \geq 1$, and $(\lambda_1, \lambda_2, .., \lambda_n) \in F^n$ satisfies $\lambda_1^i + \lambda_2^i + \cdots + \lambda_n^i = \underline{0}$ for each $1 \leq i \leq n$. Then $\lambda_i = \underline{0}$ for $1 \leq i \leq n$.

Minimal polynomials To conclude this section here are a few comments on the minimal polynomial of a linear transformation. This part should be studied only if you need it. Suppose V is an n-dimensional vector space over a field F and $T : V \to V$ is a linear transformation. As before we make V a module over $F[x]$ with $T(v) = vx$.

Definition $Ann(V_{F[x]})$ is the set of all $h \in F[x]$ which annihilate V, i.e., which satisfy $Vh = \underline{0}$. This is a non-zero ideal of $F[x]$ and is thus generated by a unique monic polynomial $u(x) \in F(x)$, $Ann(V_{F[x]}) = uF[x]$. The polynomial u is called the *minimal polynomial* of T. Note that $u(T) = \underline{0}$ and if $h(x) \in F[x]$, $h(T) = \underline{0}$ iff h is a multiple of u in $F[x]$. If $p(x) \in F[x]$ is the characteristic polynomial of T, $p(T) = \underline{0}$ and thus p is a multiple of u.

Now we state this again in terms of matrices. Suppose $A \in F_n$ is a matrix representing T. Then $u(A) = \underline{0}$ and if $h(x) \in F[x]$, $h(A) = \underline{0}$ iff h is a multiple of u in $F[x]$. If $p(x) \in F[x]$ is the characteristic polynomial of A, then $p(A) = \underline{0}$ and thus p is a multiple of u. The polynomial u is also called the minimal polynomial of A. Note that these properties hold for any matrix representing T, and thus similar matrices have the same minimal polynomial. If A is given to start with, use the linear transformation $T : F^n \to F^n$ determined by A to define the polynomial u.

Now suppose $q \in F[x]$ is a monic polynomial and $C(q) \in F_n$ is the companion matrix defined in the section Jordan Blocks. Whenever $q(x) = (x - \lambda)^n$, let $B(q) \in F_n$ be the Jordan block matrix also defined in that section. Recall that q is the characteristic polynomial and the minimal polynomial of each of these matrices. This together with the rational form and the Jordan form will allow us to understand the relation of the minimal polynomial to the characteristic polynomial.

Exercise Suppose $A_i \in F_{n_i}$ has q_i as its characteristic polynomial and its minimal polynomial, and $A = \begin{pmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_r \end{pmatrix}$. Find the characteristic polynomial and the minimal polynomial of A.

Exercise Suppose $A \in F_n$.

1) Suppose A is the matrix displayed in Theorem 2 above. Find the characteristic and minimal polynomials of A.

2) Suppose A is the matrix displayed in Theorem 3 above. Find the characteristic and minimal polynomials of A.

3) Suppose A is the matrix displayed in Theorem 4 above. Find the characteristic and minimal polynomials of A.

4) Suppose $\lambda \in F$. Show λ is a root of the characteristic polynomial of A iff λ is a root of the minimal polynomial of A. Show that if λ is a root, its order in the characteristic polynomial is at least as large as its order in the minimal polynomial.

5) Suppose $\bar{F}$ is a field containing F as a subfield. Show that the minimal polynomial of $A \in F_n$ is the same as the minimal polynomial of A considered as a matrix in $\bar{F}_n$. (This funny looking exercise is a little delicate.)

6) Let $F = \mathbf{R}$ and $A = \begin{pmatrix} 5 & -1 & 3 \\ 0 & 2 & 0 \\ -3 & 1 & -1 \end{pmatrix}$. Find the characteristic and minimal polynomials of A.

Determinants

In the chapter on matrices, it is stated without proof that the determinant of the product is the product of the determinants (see page 63). The purpose of this section is to give a proof of this. We suppose R is a commutative ring, C is an R-module, $n \geq 2$, and $B_1, B_2, \ldots, B_n$ is a sequence of R-modules.

Definition A map $f : B_1 \oplus B_2 \oplus \cdots \oplus B_n \to C$ is R-*multilinear* means that if $1 \leq i \leq n$, and $b_j \in B_j$ for $j \neq i$, then $f|(b_1, b_2, \ldots, B_i, \ldots, b_n)$ defines an R-linear map from B_i to C.

Theorem The set of all R-multilinear maps is an R-module.

Proof From the first exercise in Chapter 5, the set of all functions from $B_1 \oplus B_2 \oplus \cdots \oplus B_n$ to C is an R-module (see page 69). It must be seen that the R-multilinear maps form a submodule. It is easy to see that if f_1 and f_2 are R-multilinear, so is $f_1 + f_2$. Also if f is R-multilinear and $r \in R$, then (fr) is R-multilinear.

From here on, suppose $B_1 = B_2 = \cdots = B_n = B$.

Definition

1) f is *symmetric* means $f(b_1, \ldots, b_n) = f(b_{\tau(1)}, \ldots, b_{\tau(n)})$ for all permutations τ on $\{1, 2, \ldots, n\}$.
2) f is *skew-symmetric* if $f(b_1, \ldots, b_n) = \text{sign}(\tau) f(b_{\tau(1)}, \ldots, b_{\tau(n)})$ for all τ.

3) f is *alternating* if $f(b_1, \ldots, b_n) = \underline{0}$ whenever some $b_i = b_j$ for $i \neq j$.

Theorem

i) Each of these three types defines a submodule of the set of all R-multilinear maps.
ii) Alternating $\Rightarrow$ skew-symmetric.
iii) If no element of C has order 2, then alternating $\iff$ skew-symmetric.

Proof Part i) is immediate. To prove ii), assume f is alternating. It sufficies to show that $f(b_1, ..., b_n) = -f(b_{\tau(1)}, ..., b_{\tau(n)})$ where τ is a transposition. For simplicity, assume $\tau = (1, 2)$. Then $\underline{0} = f(b_1 + b_2, b_1 + b_2, b_3, ..., b_n) = f(b_1, b_2, b_3, ..., b_n) + f(b_2, b_1, b_3, ..., b_n)$ and the result follows. To prove iii), suppose f is skew symmetric and no element of C has order 2, and show f is alternating. Suppose for convenience that $b_1 = b_2$ and show $f(b_1, b_1, b_3, \ldots, b_n) = \underline{0}$. If we let τ be the transposition $(1, 2)$, we get $f(b_1, b_1, b_3, \ldots, b_n) = -f(b_1, b_1, b_3, \ldots, b_n)$, and so $2f(b_1, b_1, b_3, \ldots, b_n) = \underline{0}$, and the result follows.

Now we are ready for determinant. Suppose $C = R$. In this case multilinear maps are usually called *multilinear forms*. Suppose B is R^n with the canonical basis $\{e_1, e_2, \ldots, e_n\}$. (We think of a matrix $A \in R_n$ as n column vectors, i.e., as an element of $B \oplus B \oplus \cdots \oplus B$.) First we recall the definition of determinant.

Suppose $A = (a_{i,j}) \in R_n$. Define $d : B \oplus B \oplus \cdots \oplus B \to R$ by $d(a_{1,1}e_1 + a_{2,1}e_2 + \cdots + a_{n,1}e_n, ..., a_{1,n}e_1 + a_{2,n}e_2 + \cdots + a_{n,n}e_n) = \sum_{\text{all } \tau} \text{sign}(\tau)(a_{\tau(1),1}a_{\tau(2),2} \cdots a_{\tau(n),n}) = |A|$.

The next theorem follows from the section on determinants on page 61.

Theorem d is an alternating multilinear form with $d(e_1, e_2, \ldots, e_n) = \underline{1}$.

If $c \in R$, dc is an alternating multilinear form, because the set of alternating forms is an R-module. It turns out that this is all of them, as seen by the following theorem.

Theorem Suppose $f : B \oplus B \oplus \ldots \oplus B \to R$ is an alternating multilinear form. Then $f = df(e_1, e_2, \ldots, e_n)$. This means f is the multilinear form d times the scalar $f(e_1, e_2, ..., e_n)$. In other words, if $A = (a_{i,j}) \in R_n$, then $f(a_{1,1}e_1 + a_{2,1}e_2 + \cdots + a_{n,1}e_n, ..., a_{1,n}e_2 + a_{2,n}e_2 + \cdots + a_{n,n}e_n) = |A|f(e_1, e_2, ..., e_n)$. Thus the set of alternating forms is a free R-module of dimension 1, and the determinant is a generator.

Proof For $n = 2$, you can simply write it out. $f(a_{1,1}e_1 + a_{2,1}e_2, a_{1,2}e_1 + a_{2,2}e_2) = a_{1,1}a_{1,2}f(e_1, e_1) + a_{1,1}a_{2,2}f(e_1, e_2) + a_{2,1}a_{1,2}f(e_2, e_1) + a_{2,1}a_{2,2}f(e_2, e_2) = (a_{1,1}a_{2,2} - a_{1,2}a_{2,1})f(e_1, e_2) = |A|f(e_1, e_2)$. For the general case, $f(a_{1,1}e_1 + a_{2,1}e_2 + \cdots + a_{n,1}e_n, \ldots.., a_{1,n}e_1 + a_{2,n}e_2 + \cdots + a_{n,n}e_n) = \sum a_{i_1,1}a_{i_2,2}\cdots a_{i_n,n}f(e_{i_1}, e_{i_2}, ..., e_{i_n})$ where the sum is over all $1 \le i_1 \le n,\ 1 \le i_2 \le n, ..., 1 \le i_n \le n$. However, if any $i_s = i_t$ for $s \ne t$, that term is 0 because f is alternating. Therefore the sum is just $\sum_{\text{all } \tau} a_{\tau(1),1}a_{\tau(2),2}\cdots a_{\tau(n),n}f(e_{\tau(1)}, e_{\tau(2)}, \ldots, e_{\tau(n)}) = \sum_{\text{all } \tau} \text{sign}(\tau)a_{\tau(1),1}$ $a_{\tau(2),2}\cdots a_{\tau(n),n}f(e_1, e_2, \ldots, e_n) = |A|f(e_1, e_2, ..., e_n)$.

This incredible classification of these alternating forms makes the proof of the following theorem easy. (See the third theorem on page 63.)

Theorem If $C, A \in R_n$, then $|CA| = |C||A|$.

Proof Suppose $C \in R_n$. Define $f : R_n \to R$ by $f(A) = |CA|$. In the notation of the previous theorem, $B = R^n$ and $R_n = R^n \oplus R^n \oplus \cdots \oplus R^n$. If $A \in R_n, A = (A_1, A_2, ..., A_n)$ where $A_i \in R^n$ is column i of A, and $f : R^n \oplus \cdots \oplus R^n \to R$ has $f(A_1, A_2, ..., A_n) = |CA|$. Use the fact that $CA = (CA_1, CA_2, ..., CA_n)$ to show that f is an alternating multilinear form. By the previous theorem, $f(A) = |A|f(e_1, e_2, ..., e_n)$. Since $f(e_1, e_2, ..., e_n) = |CI| = |C|$, it follows that $|CA| = f(A) = |A||C|$.

Dual Spaces

The concept of dual module is basic, not only in algebra, but also in other areas such as differential geometry and topology. If V is a finitely generated vector space over a field F, its dual V^* is defined as $V^* = \text{Hom}_F(V, F)$. V^* is isomorphic to V, but in general there is no natural isomorphism from V to V^*. However there is a natural isomorphism from V to V^{**}, and so V^* is the dual of V and V may be considered to be the dual of V^*. This remarkable fact has many expressions in mathematics. For example, a tangent plane to a differentiable manifold is a real vector space. The union of these spaces is the tangent bundle, while the union of the dual spaces is the cotangent bundle. Thus the tangent (cotangent) bundle may be considered to be the dual of the cotangent (tangent) bundle. The sections of the tangent bundle are called vector fields while the sections of the cotangent bundle are called 1-forms.

In algebraic topology, homology groups are derived from chain complexes, while cohomology groups are derived from the dual chain complexes. The sum of the cohomology groups forms a ring, while the sum of the homology groups does not.

Thus the concept of dual module has considerable power. We develop here the basic theory of dual modules.

Suppose R is a commutative ring and W is an R-module.

Definition If M is an R-module, let $H(M)$ be the R-module $H(M)$=$\text{Hom}_R(M, W)$. If M and N are R-modules and $g : M \to N$ is an R-module homomorphism, let $H(g) : H(N) \to H(M)$ be defined by $H(g)(f) = f \circ g$. Note that $H(g)$ is an R-module homomorphism.

$$\begin{tikzcd}
M \arrow[r, "g"] \arrow[dr, "H(g)(f) = f \circ g"'] & N \arrow[d, "f"] \\
 & W
\end{tikzcd}$$

Theorem

i) If M_1 and M_2 are R-modules, $H(M_1 \oplus M_2) \approx H(M_1) \oplus H(M_2)$.

ii) If $I : M \to M$ is the identity, then $H(I) : H(M) \to H(M)$ is the identity.

iii) If $M_1 \xrightarrow{g} M_2 \xrightarrow{h} M_3$ are R-module homomorphisms, then $H(g) \circ H(h) = H(h \circ g)$. If $f : M_3 \to W$ is a homomorphism, then $(H(g) \circ H(h))(f) = H(h \circ g)(f) = f \circ h \circ g$.

$$\begin{tikzcd}
M_1 \arrow[r, "g"] \arrow[rrd, "f \circ h \circ g"'] & M_2 \arrow[r, "h"] \arrow[rd, "f \circ h"] & M_3 \arrow[d, "f"] \\
 & & W
\end{tikzcd}$$

Note In the language of the category theory, H is a contravariant functor from the category of R-modules to itself.

Theorem If M and N are R-modules and $g : M \to N$ is an isomorphism, then $H(g) : H(N) \to H(M)$ is an isomorphism with $H(g^{-1}) = H(g)^{-1}$.

Proof

$$I_{H(N)} = H(I_N) = H(g \circ g^{-1}) = H(g^{-1}) \circ H(g)$$

$$I_{H(M)} = H(I_M) = H(g^{-1} \circ g) = H(g) \circ H(g^{-1})$$

Theorem

i) If $g : M \to N$ is a surjective homomorphism, then $H(g) : H(N) \to H(M)$ is injective.

ii) If $g : M \to N$ is an injective homomorphism and $g(M)$ is a summand of N, then $H(g) : H(N) \to H(M)$ is surjective.

iii) If R is a field and $g : M \to N$ is a homomorphism, then g is surjective (injective) iff $H(g)$ is injective (surjective).

Proof This is a good exercise.

For the remainder of this section, suppose $W = R_R$. In this case $H(M) = \mathrm{Hom}_R(M, R)$ is denoted by $H(M) = M^*$ and $H(g)$ is denoted by $H(g) = g^*$.

Theorem Suppose M has a finite free basis $\{v_1, ..., v_n\}$. Define $v_i^* \in M^*$ by $v_i^*(v_1 r_1 + \cdots + v_n r_n) = r_i$. Thus $v_i^*(v_j) = \delta_{i,j}$. Then $v_1^*, \ldots, v_n^*$ is a free basis for M^*, called the *dual basis*. Therefore M^* is free and is isomorphic to M.

Proof First consider the case of $R^n = R_{n,1}$, with basis $\{e_1, \ldots, e_n\}$ where $e_i = \begin{pmatrix} 0 \\ \cdot \\ 1_i \\ \cdot \\ 0 \end{pmatrix}$.

We know $(R^n)^* \approx R_{1,n}$, i.e., any homomorphism from R^n to R is given by a $1 \times n$ matrix. Now $R_{1,n}$ is free with dual basis $\{e_1^*, \ldots, e_n^*\}$ where $e_i^* = (0, \ldots, 0, 1_i, 0, \ldots, 0)$. For the general case, let $g : R^n \overset{\approx}{\to} M$ be given by $g(e_i) = v_i$. Then $g^* : M^* \to (R^n)^*$ sends v_i^* to e_i^*. Since g^* is an isomorphism, $\{v_1^*, \ldots, v_n^*\}$ is a basis for M^*.

Theorem Suppose M is a free module with a basis $\{v_1, \ldots, v_m\}$ and N is a free module with a basis $\{w_1, \ldots, w_n\}$ and $g : M \to N$ is the homomorphism given by $A = (a_{i,j}) \in R_{n,m}$. This means $g(v_j) = a_{1,j} w_1 + \cdots + a_{n,j} w_n$. Then the matrix of $g^* : N^* \to M^*$ with respect to the dual bases, is given by A^t.

Proof Note that $g^*(w_i^*)$ is a homomorphism from M to R. Evaluation on v_j gives $g^*(w_i^*)(v_j) = (w_i^* \circ g)(v_j) = w_i^*(g(v_j)) = w_i^*(a_{1,j}w_1 + \cdots + a_{n,j}w_n) = a_{i,j}$. Thus $g^*(w_i^*)$ $= a_{i,1}v_1^* + \cdots + a_{i,m}v_m^*$, and thus g^* is represented by A^t.

Exercise If U is an R-module, define $\phi_U : U^* \oplus U \to R$ by $\phi_U(f, u) = f(u)$. Show that ϕ_U is R-bilinear. Suppose $g : M \to N$ is an R-module homomorphism, $f \in N^*$ and $v \in M$. Show that $\phi_N(f, g(v)) = \phi_M(g^*(f), v)$. Now suppose $M = N = R^n$ and $g : R^n \to R^n$ is represented by a matrix $A \in R_n$. Suppose $f \in (R^n)^*$ and $v \in R^n$. Use the theorem above to show that $\phi : (R^n)^* \oplus R^n \to R$ has the property $\phi(f, Av) = \phi(A^t f, v)$. This is with the elements of R^n and $(R^n)^*$ written as column vectors. If the elements of R^n are written as column vectors and the elements of $(R^n)^*$ are written as row vectors, the formula is $\phi(f, Av) = \phi(fA, v)$. Of course this is just the matrix product fAv. Dual spaces are confusing, and this exercise should be worked out completely.

Definition "Double dual" is a "covariant" functor, i.e., if $g : M \to N$ is a homomorphism, then $g^{**} : M^{**} \to N^{**}$. For any module M, define $\alpha : M \to M^{**}$ by $\alpha(m) : M^* \to R$ is the homomorphism which sends $f \in M^*$ to $f(m) \in R$, i.e., $\alpha(m)$ is given by evaluation at m. Note that α is a homomorphism.

Theorem If M and N are R-modules and $g : M \to N$ is a homomorphism, then the following diagram is commutative.

$$\begin{tikzcd}
M \arrow[r, "\alpha"] \arrow[d, "g"'] & M^{**} \arrow[d, "g^{**}"] \\
N \arrow[r, "\alpha"] & N^{**}
\end{tikzcd}$$

Proof On M, α is given by $\alpha(v) = \phi_M(-, v)$. On N, $\alpha(u) = \phi_N(-, u)$. The proof follows from the equation $\phi_N(f, g(v)) = \phi_M(g^*(f), v)$.

Theorem If M is a free R-module with a finite basis $\{v_1, \ldots, v_n\}$, then $\alpha : M \to M^{**}$ is an isomorphism.

Proof $\{\alpha(v_1), \ldots, \alpha(v_n)\}$ is the dual basis of $\{v_1^*, \ldots, v_n^*\}$, i.e., $\alpha(v_i) = (v_i^*)^*$.

Note Suppose R is a field and C is the category of finitely generated vector spaces over R. In the language of category theory, α is a natural equivalence between the identity functor and the double dual functor.

Note For finitely generated vector spaces, α is used to identify V and V^{**}. Under this identification V^* is the dual of V and V is the dual of V^*. Also, if $\{v_1, \ldots, v_n\}$ is a basis for V and $\{v_i^*, \ldots, v_n^*\}$ its dual basis, then $\{v_1, \ldots, v_n\}$ is the dual basis for $\{v_1^*, \ldots, v_n^*\}$.

In general there is no natural way to identify V and V^*. However for real inner product spaces there is.

Theorem Let $R = \mathbf{R}$ and V be an n-dimensional real inner product space. Then $\beta : V \to V^*$ given by $\beta(v) = (v, -)$ is an isomorphism.

Proof β is injective and V and V^* have the same dimension.

Note If β is used to identify V with V^*, then $\phi_V : V^* \oplus V \to \mathbf{R}$ is just the dot product $V \oplus V \to \mathbf{R}$.

Note If $\{v_1, \ldots, v_n\}$ is any orthonormal basis for V, $\{\beta(v_1), \ldots, \beta(v_n)\}$ is the dual basis of $\{v_1, \ldots, v_n\}$, that is $\beta(v_i) = v_i^*$. The isomorphism $\beta : V \to V^*$ defines an inner product on V^*, and under this structure, β is an isometry. If $\{v_1, \ldots, v_n\}$ is an orthonormal basis for V, $\{v_1^*, \ldots, v_n^*\}$ is an orthonormal basis for V^*. Also, if U is another n-dimensional IPS and $f : V \to U$ is an isometry, then $f^* : U^* \to V^*$ is an isometry and the following diagram commutes.

$$\begin{tikzcd} V \arrow[r, "\beta"] \arrow[d, "f"'] & V^* \\ U \arrow[r, "\beta"] & U^* \arrow[u, "f^*"'] \end{tikzcd}$$

Exercise Suppose R is a commutative ring, T is an infinite index set, and for each $t \in T$, $R_t = R$. Show $(\bigoplus_{t \in T} R_t)^*$ is isomorphic to $R^T = \prod_{t \in T} R_t$. Now let $T = \mathbf{Z}^+$, $R = \mathbf{R}$, and $M = \bigoplus_{t \in T} \mathbf{R}_t$. Show M^* is not isomorphic to M.