

Structure theorems, Canonical forms over a P.I.D.

Ravi A. Rao

Tata Institute of Fundamental Research, Mumbai

Contents.

- §1 Introduction
- §2 Historical comments
- §3 Structure of finitely generated abelian groups
- §4 Structure of finitely generated modules over a P.I.D.
- §5 Elementary divisors, Structure theorem revisited
- §6 Finiteness of Class Number
- §7 Equivalent class of matrices over a P.I.D.
- §8 Similarity class of matrices over a P.I.D.

1 Introduction

An integral domain in which every ideal is generated by one element (i.e. it is principal) is called a Principal Ideal Domain (P.I.D.). \mathbb{Z} , $k[X]$, where k is a field, $\mathbb{Z}[i]$ are examples. All these are Euclidean domains (E.D.), which are P.I.D. But the class of P.I.D. is larger than the class of E.D.

A P.I.D. is a Unique Factorization Domain (U.F.D.). Any two nonzero elements a, b in it have a greatest common divisor (g.c.d.). The g.c.d of a and b is of the form $ra + sb$, for some r, s .

We are going to study the structure of finitely generated modules over a P.I.D. Over any commutative ring R , one has the basic modules, R, I , where I is an ideal of R , and R/I . One can also take the direct sums of these modules to form more modules. One observes that if R is a P.I.D. then the basic modules are all cyclic modules, i.e. they are generated by one element. The interesting fact is that every finitely generated module over a P.I.D. is formed by taking a finite direct sum of these basic modules. This is what we aim to prove. Moreover, one can find invariants, which determine the module upto isomorphism.

The theory began with the study of finitely generated abelian groups; and to classify all finitely generated abelian groups upto isomorphism. In this case we are studying the case when R is \mathbb{Z} ; and the theory is somewhat easier to establish. This theorem was first proved by Leopold Kronecker in 1858. However, we will retain our perspective over P.I.D.'s by proving, whenever it is possible, the intermediary steps over P.I.D.'s with the same effort.

Finally, the theory also evolved via the study of canonical forms of matrices over a ring. A *normal form* or *canonical form* for a matrix is a class of expressions

such that each matrix can be transformed to precisely one matrix with the given class. The reader would have seen the echelon form in his study of linear equations over a field. The three forms studied here are the Smith normal form, the rational canonical form, and the Jordan canonical form. As a consequence, one can reprove the Structure theorem for finitely generated modules over a P.I.D.

We have indicated a few applications to finitely generated abelian group theory, Linear algebra of transformations, Number theory (finiteness of Class Number), Combinatorics (studying Incidence matrix), Differential equations (reducing it to a problem in linear algebra, which can be easily resolved by using these canonical forms).

2 Historical comments

This section has been culled from N. Bourbaki's Algebra II, Chapters 4-7. We have included it to give the reader a clear idea of how this topic evolved from quite extraordinary beginnings. This will also give a clear picture of the deficiency in our presentation in this article; and encourage the reader to delve into the texts, to better his understanding of the material. (For instance, we have not gone into the Jordan-Chevalley decomposition; which is accesible from the material presented here.)

H.J.S. Smith in 1861 defined the invariant factors of an integer matrix, and obtained the Smith normal form of these matrices. Previously Echelon forms over fields (one-sided reduction) were studied by J.C.F. Gauss. The normal form for similarity classes was described by C. Jordan in his 'Traite des substitutions' in 1870.

Independently, following its introduction by J.C.F. Gauss, and the important part it played in the subsequent development of Number theory, the notion of an abelian group was gradually being made precise. In his *Disquisitiones* Gauss showed that there was an element of order equal to the exponent - in other words he got the existence of the largest invariant factor of the group. In a manuscript dating from 1801, but not published in his lifetime, Gauss sketched a general proof of the decomposition of a finite abelian group into a direct sum of p groups. The theory of elliptic functions and abelian integrals, developed by Gauss, Abel and Jacobi, let to attention being paid to the structure of torsion-free abelian groups. The first and best known example of a decomposition of an infinite abelian group into a direct sum of cyclic groups was given in 1846 by Dirichlet in his paper on the units of an algebraic number field. But it was only in 1879 that the connection between the theory of finitely generated abelian groups and Smith's theorem was recognized and explicitly used by Frobenius and Stickelberger.

The notion of an eigenvalue of a linear transformation appeared explicitly in the theorem of systems of linear differential equations without constant coefficients applied by Lagrange to the theorem of small oscillations and by Lagrange and Laplace to the the "secular" perturbations of planets. Euler used it in find-

ing the axes of a conic or a quadric, and to study the principal axes of inertia of a solid body. It is also involved in the beginnings of the theory of partial differential equations, in particular the equation of a vibrating string. It was Cauchy who recognized the relationships between these various problems being studied in 1820's. The classification of "pencils" of conics or quadrics is the study of the elementary divisors of $\alpha + X\beta$, where α, β are symmetric matrices - which Sylvester studied in 1851, and Weierstrass, extending his methods, obtained a canonical form for a "pencil" $\alpha + X\beta$, with $\det(\alpha + X\beta)$ not identically zero. From this he deduced the definition of an elementary divisors of a square matrix (!), and proved that they characterize the matrix upto similarity. These results were partially, and independently recovered by Jordan two years later. Frobenius showed in 1879 that Weierstrass' theorem can be deduced from Smith's theorem extended to polynomial rings.

3 Structure of finitely generated abelian groups

Lemma 3.1 *Let M be a finitely generated module over a P.I.D. R. If m_1, \dots, m_n are generators of M , and if a_1, \dots, a_n are integers with greatest common divisor 1, then the element $m = a_1m_1 + \dots + a_nm_n$ is one of a set of n generators of M .*

Proof: For $n = 1$, the lemma is clear. For $n = 2$, there are integers c_1, c_2 , such that $a_1c_1 + a_2c_2 = 1$. Then

$$\begin{aligned} m_1 &= c_1(a_1m_1 + a_2m_2) + a_2(c_2m_1 - c_1m_2) \\ m_2 &= c_2(a_1m_1 + a_2m_2) - a_1(c_2m_1 - c_1m_2) \end{aligned}$$

and $M = \{m_1, m_2\} = \{m, c_2m_1 - c_1m_2\}$.

For $n > 2$, let d be $\text{g.c.d.}(a_1, \dots, a_{n-1})$ and let $a_i = b_id$, for $i < n$. Since $\text{g.c.d.}(b_1, \dots, b_{n-1}) = 1$, if $m' = b_1m_1 + \dots + b_{n-1}m_{n-1}$, then by induction there are elements m'_2, \dots, m'_{n-1} such that $M' = \{m_1, \dots, m_{n-1}\} = \{m', m'_2, \dots, m'_{n-1}\}$. Hence, $M = \{m', m'_2, \dots, m'_{n-1}, m_n\}$. But $m = dm' + a_nm_n$, and $\text{g.c.d.}(d, a_n) = 1$. Hence, $\{m, m''\} = \{m', m_n\}$, for some $m'' \in M$. Hence, $M = \{m, m'_2, \dots, m'_{n-1}, m''\}$, as required. \square

Theorem 3.2 (Basis Theorem for finitely generated abelian groups)

Let G be a finitely generated abelian group. If n is the least integer such that G is generated by n elements, then G is a direct sum of n cyclic submodules.

Proof: Let g_1, \dots, g_n be generators of G so chosen that g_n has order k , with k minimal, i.e. no other set of n generators of G has an element of smaller order. Let $H = \{g_1, \dots, g_{n-1}\}$. By induction on n , H is a direct sum of $n - 1$ cyclic subgroups. We show that $H \cap \{g_n\} = 0$. If not, then for some positive integer $a_n < k$, and integers a_1, \dots, a_{n-1} , we have an equation

$$a_1g_1 + \dots + a_{n-1}g_{n-1} = a_ng_n$$

Let $d = \text{g.c.d.}\{a_1, \dots, a_n\}$. By Lemma 3.1, $(a_1/d)g_1 + \dots + (a_{n-1}/d)g_{n-1} - (a_n/d)g_n$ is a member of a set of n generators of G . Its order is at most d , and $d \leq a_n < k$. This contradicts the minimality of k in the choice of g_n . Hence $H \cap \{g_n\} = 0$, and $G = H \oplus \{g_n\}$ as required. \square

Theorem 3.3 (Elementary divisors for finitely generated abelian groups)

Let G be the direct sum of n infinite cyclic subgroups, and let H be a non-trivial subgroup of G . Then a basis $\{g_1, \dots, g_n\}$ of G may be chosen, so that for appropriate integers d_i , with $d_i | d_{i+1}$, for all i , the set of non-zero elements of $\{d_i g_i\}$ is a basis of H .

Proof: If $n = 1$, G is cyclic, and the theorem is clear. We induct on n . We show that if $0 \neq h \in H$ then there is an element $g \in G$, which is part of a basis of G of n elements, such that $h = dg$, for some $d \in \mathbb{Z}$: Take any basis $\{f_1, \dots, f_n\}$ of G , and let $h = a_1 f_1 + \dots + a_n f_n$. Let $d = \text{g.c.d.}(a_1, \dots, a_n)$, and let $a_i = db_i$. By Lemma 3.1 $g = b_1 f_1 + \dots + b_n f_n$ is one element of a basis of n elements of G ; and $h = dg$ as claimed.

Among all such n -basis for G , let us assume that the n -basis $\{g_1, g'_i\}$ has been chosen with d minimal, and $dg_1 \in H$. Claim: If $h = b_1 g_1 + b_2 g'_2 + \dots + b_n g'_n \in H$, then $d | b_1$. If not, let $b_1 = ld + r$, with $0 \neq r < d$. Then $h' = h - ldg_1 \in H$. Hence, there is a n -basis $\{g''_i\}$ of G such that $h' = d'g''_1$ with $d' \leq r < d$, contrary to the minimality of d . Hence $d | b_1$.

Hence, $H = \{dg_1, H \cap \{g'_2, \dots, g'_n\}\}$. By induction, $H \cap \{g'_2, \dots, g'_n\} = \{d_2 g_2, \dots, d_n g_n\}$, with $\{g_2, \dots, g_n\}$ a basis of $\{g'_2, \dots, g'_n\}$, and with $d_i | d_{i+1}$, for $i \geq 2$. By above observation, $d | d_2$, and so $\{g_1, g_2, \dots, g_n\}$ is a n -basis of G as required. \square

Remark 3.4 In Exercise 14 there is an example to show that the above theorem does not hold when G is a finite group.

Corollary 3.5 Let G be a finitely generated abelian group.

(i) $G \simeq \mathbb{Z}^r \oplus \mathbb{Z}/m_1\mathbb{Z} \oplus \dots \oplus \mathbb{Z}/m_k\mathbb{Z}$, for some $r, m_1, \dots, m_k \in \mathbb{Z}$.

(Here r is called the **rank** of G , \mathbb{Z}^r is called the **torsion-free** part of G , and $\oplus \mathbb{Z}/m_i\mathbb{Z}$ is called the **torsion** part of M , also denoted by $\text{Tor}(M)$.)

(Note that this decomposition is not in a unique manner, for e.g. $\mathbb{Z}/6\mathbb{Z} \simeq (\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/3\mathbb{Z})$.)

(ii) $G \simeq \mathbb{Z}^r \oplus (\oplus_{i=1}^k \oplus_{j=1}^{k_i} \mathbb{Z}/p_i^{n_{ij}})$

(iii) (Invariant structure theorem)

$G \simeq \mathbb{Z}^r \oplus \mathbb{Z}/q_1\mathbb{Z} \oplus \dots \oplus \mathbb{Z}/q_l\mathbb{Z}$, with $q_i | q_{i-1}$, for all $i > 1$, where $l = \max\{k_i\}$.

(The numbers r, q_1, \dots, q_l are called the **invariant factors** of M .)

The invariant factors determine G uniquely upto isomorphism.

Proof: (i) is clear from Theorem 3.2. We prove (ii): Let $m_i = \prod_{j=1}^{k_i} p_i^{n_{ij}}$, p_i distinct primes. Now use the Chinese Remainder Theorem.

To prove (iii): Organize the elementary divisors exponents as an array:

$$\begin{array}{ccccccc} p_1 & : & n_{11} & \geq & \cdots & \geq & n_{1k_1} \\ p_2 & : & n_{21} & \geq & \cdots & \geq & n_{2k_2} \\ \vdots & : & \cdots & \geq & \cdots & \geq & \cdots \\ p_k & : & n_{k1} & \geq & \cdots & \geq & n_{kk_1} \end{array}$$

For each $h = 1, \dots, l = \max\{k_1, \dots, k_1, \dots, k_m\}$, let $q_h = p_1^{n_{1h}} \dots p_k^{n_{kh}}$. Then $q_h | q_{h-1}$. By Chinese Remainder Theorem, $\mathbb{Z}/(q_h) \simeq \bigoplus_{i=1}^k \mathbb{Z}/p_i^{n_{ih}}$. Hence,

$$G \simeq \mathbb{Z}^r \oplus (\bigoplus \mathbb{Z}/q_h \mathbb{Z}).$$

Uniqueness part: Suppose that $G \simeq \mathbb{Z}^r \oplus \mathbb{Z}/q_1 \mathbb{Z} \oplus \dots \oplus \mathbb{Z}/q_l \mathbb{Z}$, with $q_i | q_{i-1}$, for all $i > 1$, and also $G \simeq \mathbb{Z}^s \oplus \mathbb{Z}/q'_1 \mathbb{Z} \oplus \dots \oplus \mathbb{Z}/q'_m \mathbb{Z}$, with $q'_i | q'_{i-1}$, for all $i > 1$. We show that $r = s$, $l = m$, and that $q_i \mathbb{Z} = q'_i \mathbb{Z}$, for all $i > 1$.

$G/\text{Tor}(G) \simeq \mathbb{Z}^r \simeq \mathbb{Z}^s$. Tensoring with \mathbb{Q} , or going mod a prime (p), gives $r = s$.

For a prime p , let $\text{Syl}_p(G) = \{g \in G \mid p^n g = 0 \text{ for some } n \geq 0\}$. Note that, if G_1, G_2 are groups then $\text{Syl}_p(G_1 \oplus G_2) = \text{Syl}_p(G_1) \oplus \text{Syl}_p(G_2)$. Now

$$\begin{aligned} \text{Syl}_p(G) &= \text{Syl}_p(\text{Tor}(G)) \\ &= \text{Syl}_p(\bigoplus_i \mathbb{Z}/q_i \mathbb{Z}) = \bigoplus_i \text{Syl}_p(\mathbb{Z}/q_i \mathbb{Z}) \\ &= \text{Syl}_p(\bigoplus_j \mathbb{Z}/q_j \mathbb{Z}) = \bigoplus_j \text{Syl}_p(\mathbb{Z}/q'_j \mathbb{Z}) \end{aligned}$$

Suppose that the first p -group is of **type** $(p^{r_1}, \dots, p^{r_s})$, and the second is of type $(p^{m_1}, \dots, p^{m_k})$, with $r_1 \geq \dots \geq r_s \geq 1$, and $m_1 \geq \dots \geq m_k \geq 1$.

Now $p\text{Syl}_p(\text{Tor}(G))$ is also a p -group, of order strictly less than the order of $\text{Syl}_p(\text{Tor}(G))$, and is of type

$$(p^{r_1-1}, \dots, p^{r_s-1}) \& (p^{m_1-1}, \dots, p^{m_k-1}),$$

it being understood that if some exponent r_i or m_j is equal to 1, then the factor corresponding to p^{r_i-1} or p^{m_j-1} is the trivial group.

By induction, we have $r_i - 1 = m_i - 1$ for all those integers i such that $r_i - 1$ or $m_i - 1 \geq 1$. Hence $r_i = m_i$ for all these integers i , and the two series can only differ in their last components which can be equal to p . These correspond to factors of type (p, \dots, p) occurring say a times in the first sequence and say b times in the second sequence. But comparing the orders on both sides, which equal order of the group, one gets $a = b$. The uniqueness is thus established. \square

4 Structure of finitely generated modules over a P.I.D.

We begin with the Chinese Remainder Theorem over a commutative ring.

Definition 4.1 *Two-sided ideals I, J of a ring A are called **comaximal**, if $I + J = A$ or equivalently if there are $a \in I, b \in J$ with $a + b = 1$.*

Lemma 4.2 (Chinese Remainder Theorem)

Let R be a ring, I_1, \dots, I_n be two-sided ideals of R and $\eta : R \rightarrow \prod_{i=1}^n R/I_i$ be the natural homomorphism defined by $\eta(r) = (r + I_1, \dots, r + I_n)$. Then η is surjective if and only if the I_1, \dots, I_n are mutually pairwise comaximal. In this case we have

$$R/(I_1 \cap \dots \cap I_n) \cong \prod_{i=1}^n (R/I_i).$$

If the I_i pair-wise commute, especially if R is commutative, then we have further that $I_1 \cdots I_n = I_1 \cap \dots \cap I_n$ in the above case.

Proof: Assume $I_i + I_j = R$ for $i \neq j$. Let $1 = u_i + v_i$, $u_i \in I_1, v_i \in I_i$ for $i \neq 1$, and let $r_1 = \prod_{i=1}^n v_i = \prod_{i=2}^n (1 - u_i) = 1 + u$ for some $u \in I_1$. Then $\eta(r_1) = e_1 := (1 + I_1, 0 + I_2, \dots, 0 + I_n)$. Analogously, $e_i \in \text{Im}(\eta)$ for all i , and so η is onto.

Conversely, assume η is onto. If $\eta(a_i) = e_i$ then $1 - a_i \in I_i$ and $a_i \in I_j$ if $j \neq i$. Therefore, $1 = (1 - a_i) + a_i \in I_i + I_j$.

Clearly $\ker(\eta) = I_1 \cap \dots \cap I_n$.

We prove the last assertion by induction on n , the case $n = 1$ being trivial. If $I + J = R$ we compute

$$I \cap J = (I + J)(I \cap J) = I(I \cap J) + J(I \cap J) \subset IJ,$$

whence $I \cap J = IJ$. The induction step is

$$I_1 \cap \dots \cap I_n = (I_1 \cdots I_{n-1}) \cap I_n = (I_1 \cdots I_{n-1}) \cap I_n = I_1 \cdots I_n.$$

□

We now prove the structure theorem (weak form) for finitely generated modules over a P.I.D. The proof is similar to that of Theorem 3.1, but the crucial point is to have the right definition for the ‘order’ $o(a)$ of an element a ; and to attach a ‘weight’ to $a \in R$. We first answer this:

In a P.I.D. the generators of an ideal is unique upto associates. If $a \in R$, then the generator of $\text{ann}(a)$ ($= \{r \in R \mid ra = 0\}$) is called the **order** of a , denoted by $o(a)$. This notation is consistent, upto sign, with the notion of order of an element of a group.

We attach a **weight** $P(a)$ to $a \in R$ next. Since R is a U.F.D., we denote by $P(a)$ the number of prime factors (counting multiplicity) of a . By convention,

$P(0) = \infty$. Thus, $a|b$ in R implies that $P(a) \leq P(b)$, with equality if and only if a, b are associates.

Lemma 4.3 *Let M be a finitely generated module over a P.I.D. R , say $M = \{m_1, \dots, m_n\}$. Suppose that there is a relation $a_1m_1 + \dots + a_nm_n = 0$, where not all the a_i are zero. Then there are elements $m'_1, \dots, m'_n \in M$, such that $M = \{m'_1, \dots, m'_n\}$, and the order of m'_i divides every a_i .*

Proof: If one of the a_i is a unit then this is easy: Say a_1 is a unit, then m_1 is a linear combination of the other m_i . So take $m'_1 = 0, m'_i = m_i, i > 1$.

Let $s = \sum P(a_i)$, where $a_i \neq 0$. We induct on s . If $s = 0$, every a_i is zero or a unit, and atleast one a_i is a unit. So done.

If only one a_i is nonzero, the result is easy to establish, so let us assume a_1, a_2 are nonzero, non-unit. Let $b = \text{g.c.d.}(a_1, a_2)$, $a_1 = bc_1, a_2 = bc_2$, and $b_1c_1 + b_2c_2 = 1$. Now

$$\begin{aligned} M &= \{m_1, m_2, \dots, m_n\} \\ &= \{(m_1, m_2) \begin{pmatrix} c_2 & b_1 \\ -c_1 & b_2 \end{pmatrix}, m_3, \dots, m_n\} \\ 0 &= b(b_1m_1 + b_2m_2) + a_3m_3 + \dots + a_nm_n \end{aligned}$$

Now $P(b) \leq P(a_1) < P(a_1) + P(a_2)$. By induction, $M = \{m'_1, \dots, m'_n\}$, with $o(m'_1)|b$, and $o(m'_i)|a_i$, for $i \geq 3$. But $b|a_1, b|a_2$, hence $o(m'_1)|a_i$, for all i . \square

Theorem 4.4 (Structure theorem for finitely generated modules over a P.I.D.)

Every n -generated module M over a P.I.D. R is a direct sum of n cyclic modules $M \simeq \bigoplus_{i=1}^n Rm_i$. Equivalently, $M = \{m_1, \dots, m_n\}$, and $\sum a_i m_i = 0$ implies $a_i m_i = 0$, for all i .

(We say that the set $\{m_1, \dots, m_n\}$ is a **basis** of M .)

Proof: If $n = 1$, this is easy, as R is a P.I.D.; so let $n > 1$. We induct on n .

Amongst all possible set of generators of M having n elements choose one which has an element m with least $P(m)$. Let $M = \{m = m_1, m'_2, \dots, m'_n\}$.

If $M = Rm \oplus \sum_{i \geq 2} Rm'_i$, then by induction the submodule $\sum_{i \geq 2} Rm'_i$ has a basis $\{m_2, \dots, m_n\}$. But then $\{m_1, \dots, m_n\}$ is a basis of M .

We show that Rm is indeed a direct summand of M : If not, one has a relation $a_1m_1 + \dots + a_nm_n = 0$, with $a_1m_1 \neq 0$. Let $b = \text{g.c.d.}(a_1, o(m_1)) = c_1a_1 + c_2o(m_1)$. Since $a_1m_1 \neq 0$, a_1 and $o(m_1)$ are not associates. Hence, $P(b) < P(o(m_1))$.

Note that $bm_1 + c_1a_2m_2 + \dots + c_1a_nm_n = 0$. By Lemma 4.3 $M = \{m'_1, \dots, m'_n\}$, with $o(m'_1)|b, o(m'_i)|c_1a_i$, for $i \geq 2$. Since $P(o(m'_1)) \leq P(b) < P(o(m_1))$, this contradicts the minimality of $\{m_1, \dots, m_n\}$. Thus, Rm is a summand of M . and the result follows. \square

Corollary 4.5 *Let M be a finitely generated module over a P.I.D. R . Then*

- (i) $M \simeq R^r \oplus R/a_1R \oplus \cdots \oplus R/a_nR$, for some $r, a_i \in R$.
 ($r = \mathbf{rank}(M)$, R^r is the **torsion-free** part of M , $\text{Tor}(M) = \oplus R/a_iR$ is the **torsion** part of M)
- (ii) $M \simeq R^r \oplus (\oplus_{i=1}^k \oplus_{j=1}^{k_i} R/p_i^{n_{ij}}R)$, where $a_i = \prod_{j=1}^{k_i} p_i^{n_{ij}}$, p_i are distinct primes.
 (The rank r , and $p_i^{n_{ij}}$ are called the **elementary divisors** of M)
- (iii) $M \simeq R^r \oplus R/a_1R \oplus \cdots \oplus R/a_lR$, where $(a_1) \subset (a_2) \subset \cdots \subset (a_l)$, $l = \max\{k_i\}$.
 (The rank r , and the ideals (a_i) are called the **invariant factors** of M .)
 The invariant factors determine M upto isomorphism.

Proof: Clear from Theorem 4.4, and uniqueness argument in Corollary 3.5.

(An intrinsic characterisation of the ideals (a_i) , which shows that they are uniquely determined, is given in the Exercise 32.) \square

5 Elementary divisors, Structure theorem revisited

In this section we prove the Elementary divisors theorem for finitely generated modules over a P.I.D., and use it to give an alternative proof of the Structure theorem.

Proposition 5.1 *Let R^n be a free module of rank n over a P.I.D. R and $M \subset R^n$ be a submodule. Then M is a free module of rank $\leq n$.*

Proof: We show this by induction on n , the statement being true for $n \leq 1$. Let pr_n be the projection onto the last factor and denote its kernel by R^{n-1} . Then we get an exact sequence

$$0 \longrightarrow M \cap R^{n-1} \longrightarrow M \longrightarrow \text{pr}_n(M) \longrightarrow 0.$$

Since $I_n := \text{pr}_n(M) \subset R$ is an ideal, hence it is either (0) or principal $aR \simeq R$, the sequence splits. So $M \simeq (M \cap R^{n-1}) \oplus I_n$. By the induction hypothesis we are done. \square

Corollary 5.2 *Let M be a n -generated module over a P.I.D. R . If N is a submodule of M then N can be generated by $m \leq n$ elements.*

Proof: One has a commutative diagram

$$\begin{array}{ccc} F_1 = \varphi^{-1}N & \xrightarrow{i} & F \\ \varphi \downarrow & & \downarrow \varphi \\ N & \xrightarrow{i} & M \end{array}$$

By Proposition 5.1 $\varphi^{-1}N = F_1$ is a free module of rank $m \leq n$. \square

Definition 5.3 A R -module is **torsion free** in case for every $m \in M$ and $r \in R$ if $rm = 0$, then $r = 0$ or $m = 0$.

Any free module over an integral domain is torsion free. The converse is only partially true.

The following is clear from the structure theorem. We prove it independently.

Corollary 5.4 A torsion free finitely generated module M over a P.I.D. R is free.

Proof: Let $S = R \setminus \{0\}$, and $K = S^{-1}R$ be the quotient field of R . Since M is torsion free, the natural map $M \rightarrow S^{-1}M$ is injective; so we can regard M as a R -submodule of $S^{-1}M$. The latter is a finitely generated vector space over K . We can choose a basis m_1, \dots, m_n of $S^{-1}M$ over K with each $m_i \in M$. Then $F = \sum Rm_i$ is a free R -module.

Since M is finitely generated, there is a $b \in S$, with $bM \subset F$: if n_1, \dots, n_k generate M , and $n_j = \sum_{i=1}^n (a_{ij}/s_{ij})m_i$, take $b = \prod_{i,j} s_{ij}$. Since $b \neq 0$, and M is torsion free. The map $h_b : m \rightarrow bm$ is injective R -module map from $M \hookrightarrow F$. By Proposition 5.1, M is free. \square

Definition 5.5 An element $m \in M$ is said to be a **torsion element** if $\text{ann}(m) \neq 0$. The set of torsion elements of a R -module M is a submodule denoted by $\text{Tor}(M)$, and called the **torsion submodule** of M .

Lemma 5.6 For every module M , $\text{Tor}(M)$ is a torsion submodule and $M/\text{Tor}(M)$ is torsion free.

Proof: Easy exercise. \square

Corollary 5.7 If M is a finitely generated module over a P.I.D. R , then there is a free module F of rank r such that $M \simeq F \oplus \text{Tor}(M)$.

Proof: One has an exact sequence of R -modules:

$$0 \longrightarrow \text{Tor}(M) \longrightarrow M \longrightarrow M/\text{Tor}(M) \longrightarrow 0.$$

Since $M/\text{Tor}(M)$ is torsion free, it is free by Corollary 5.4. But then the above sequence splits, and $M \simeq M/\text{Tor}(M) \oplus \text{Tor}(M)$. \square

Corollary 5.8 Every finitely generated abelian group is the direct sum of a finite group and a free abelian group.

Theorem 5.9 (Elementary divisors theorem)

Let R be a P.I.D. and let M be a nonzero finitely generated submodule of R^n . Then there is a basis f_1, \dots, f_n of R^n and non-zero elements d_1, \dots, d_m , $m \leq n$, such that

$$(1) \{d_1 f_1, \dots, d_m f_m\} \text{ form a basis of } M.$$

(2) $d_i | d_{i+1}$ for $i = 1, \dots, m-1$.

The sequence of ideals $(d_m) \subset \dots \subset (d_1)$ is uniquely determined by (1), (2). Moreover, $(d_m) = \text{ann}(M)$.

Proof: For $f \in (R^n)^*$ (the dual space of linear functionals on R^n), we attach the ideal $O(f) = f(R^n)$. Select f_1 such that $f_1(M)$ is maximal in the set $\{O(f)\}$, where f varies over all the linear functionals on R^n .

Now $f_1(M) = (d_1) \subsetneq R$. Let $f_1(m_1) = d_1$. If $g \in (R^n)^*$, then $g(m_1) \in (d_1)$, as (d_1) is maximal. (Justify this step!)

Therefore, if $m_1 = \sum b_i f_i$, then $d_1 | b_i$ for all i . Hence $m_1 = d_1 e_1$ for some $e_1 \in R^n$. Moreover, $f_1(e_1) = 1$.

This allows us to conclude that $R^n = R e_1 \oplus \ker(f_1)$: Clearly, $R e_1 \cap \ker(f_1) = 0$. Moreover, if $m \in R^n$, then $m - f_1(m) e_1 \in \ker(f_1)$.

By Proposition 5.1, $\ker(f_1) \simeq R^k$, for some k . Let $M_1 = M \cap \ker(f_1)$. Then $M = R m_1 \oplus M_1$.

From the maximal condition on $f_1(M)$ it follows that for any $g : R^n \rightarrow R$, $g(M) \subset f_1(M)$ (otherwise a suitable linear combination will give a bigger ideal as image of a functional). Now use induction. \square

Theorem 5.10 (Structure theorem over P.I.D.)

Let R be a P.I.D. Every finitely generated R -module M is of the form

$$M \simeq R/Rd_1 \oplus R/Rd_2 \oplus \dots \oplus R/Rd_r,$$

for some $d_i \in R$. ($d_i = 0$ is not excluded.) Moreover, $d_i | d_{i+1}$, for $1 \leq i \leq r-1$.

Proof: We may assume that M is not free. Then $M \simeq F/G$, where F is a free module of finite rank, and G is a nonzero submodule of F . By the Elementary divisors theorem, choose f_1, \dots, f_n of F , and $g_1 = d_1 f_1, \dots, g_k = d_k f_k$ of G , for some $d_i \in R$, satisfying the divisibility condition. Hence,

$$M \simeq F/G \simeq \bigoplus_{i=1}^n R f_i / \bigoplus_{j=1}^k R d_j f_j \simeq (\bigoplus_{j=1}^k R f_j / d_j R f_j) \oplus R^{n-k}.$$

For $1 \leq j \leq k$, $R f_j / d_j R f_j \simeq R/Rd_j$ is cyclic. Discarding those d_j (in the beginning) which are units, and renaming the remaining as d_1, \dots, d_r . we get the required decomposition. \square

We next prove another important decomposition theorem for finitely generated modules over a P.I.D. - the Primary decomposition theorem. We end the section with an application, where we calculate the structure of the multiplicative group of units of the integers modulo an integer a .

We have seen that the torsion submodule $\text{Tor}(M)$ is a direct summand of M , when M is a finitely generated R -module, with R a P.I.D. Hence, the complete analysis of finitely generated modules over a P.I.D. depends on a description of the torsion ones. We first decompose these into their “ p -primary-components”:

Definition 5.11 Let M be a R -module. For each prime $p \in R$. let

$$\begin{aligned} M(p) &= \{m \in M \mid p^n m = 0 \text{ for some } n \geq 0\} \\ &= \{m \in M \mid o(m) = (p^n) \text{ for some } n \geq 0\} \end{aligned}$$

Each $M(p)$ is a submodule of M .

Our next decomposition theorem holds for arbitrary torsion modules, not just finitely generated ones.

Theorem 5.12 (The Primary decomposition theorem)

Let M be a torsion module over a P.I.D. R . Then $M = \bigoplus M(p)$, where p runs over the primes of R . (If M is finitely generated then $M(p) = 0$, for all but finitely many primes $p \in R$).

Proof: Let $0 \neq m \in M$. Since R is a P.I.D., $o(m) = (a)$, with $a = p_1^{r_1} \dots p_n^{r_n}$, for distinct primes $p_1, \dots, p_n \in R$. For each i let $q_i p_i^{r_i} = a$. Then $q_i m \in M(p_i)$. Since $\text{g.c.d.}(q_1, \dots, q_n) = 1$, we have $r_1 q_1 + \dots + r_n q_n = 1$, for some $r_i \in R$. Consequently $m = \sum_i r_i q_i m \in \sum_i M(p_i)$. Thus the submodules $M(p_i)$ generate M .

We show that the $M(p)$ are linearly independent. Let p_1, \dots, p_n, q be distinct primes and let

$$m \in (M(p_1) + \dots + M(p_n)) \cap M(q).$$

Then $o(m) = (q^m) = (p_1^{r_1} \dots p_n^{r_n})$, for some $m, r_1, \dots, r_n \geq 0$. This forces $m = r_1 = \dots = r_n = 0$, and so $o(m) = R$; whence $m = 0$. \square

Corollary 5.13 Let M be a finitely generated torsion module over a P.I.D. R . Every submodule N of M is the direct sum of the submodules $N \cap M(p)$, where p is an irreducible element of R .

Proof: This follows from the fact that $N \cap M(p)$ is $N(p)$. \square

Corollary 5.14 The submodule N of the torsion R -module M over a P.I.D. R , is a direct factor if and only if $N(p)$ is a direct factor of $M(p)$ for every irreducible element p of R .

Proof: Clear from Corollary 5.12. \square

Corollary 5.15 Let N be a submodule of a torsion module M over a P.I.D. R . If, for every irreducible element π , either $N(p) = 0$, or $(M/N)(p) = 0$, then N is a direct factor of M .

Proof: If $(M/N)(p) = 0$, then $N(p) = M(p)$. Now apply Corollary 5.13. \square

The multiplicative group of units of the integers modulo a

Let a be an integer > 1 , and let $(\mathbb{Z}/a\mathbb{Z})^*$ be the multiplicative group of invertible elements of the ring $\mathbb{Z}/a\mathbb{Z}$. If $a = \prod_i p_i^{n(i)}$ is the decomposition of a into prime factors, then by the Chinese Remainder theorem, the ring $\mathbb{Z}/a\mathbb{Z}$ is isomorphic to the product of the rings $\mathbb{Z}/p_i^{n(i)}\mathbb{Z}$ and the group $(\mathbb{Z}/a\mathbb{Z})^*$ is isomorphic to the product of the groups $(\mathbb{Z}/p_i^{n(i)}\mathbb{Z})^*$. We are thus reduced to the study of the groups $(\mathbb{Z}/p^n\mathbb{Z})^*$, where p is a prime number. Recall that the order $\varphi(p^n)$ of $(\mathbb{Z}/p^n\mathbb{Z})^*$ is $p^n - p^{n-1} = p^{n-1}(p-1)$. (Here φ denotes the Euler totient function.)

Suppose first of all that $p > 2$; the natural homomorphism $\mathbb{Z}/p^n\mathbb{Z} \rightarrow \mathbb{Z}/p\mathbb{Z}$ restricts to a homomorphism of groups from $(\mathbb{Z}/p^n\mathbb{Z})^*$ onto $(\mathbb{Z}/p\mathbb{Z})^*$, whose kernel we denote $U(p^n)$. The residue class mod p^n of an integer m is invertible if and only if m is coprime to p , that is if and only if the residue class of m mod p is invertible. It follows that $U(p^n)$ consists of all the residue classes mod p^n of integers congruent to 1 mod p , so has p^{n-1} elements, and that there is an exact sequence

$$\{1\} \rightarrow U(p^n) \rightarrow (\mathbb{Z}/p^n\mathbb{Z})^* \rightarrow (\mathbb{Z}/p\mathbb{Z})^* \rightarrow \{1\}. \quad (1)$$

Similarly, for $n \geq 2$ let $U(2^n)$ denote the kernel of the natural homomorphism from $(\mathbb{Z}/2^n\mathbb{Z})^*$ to $(\mathbb{Z}/4\mathbb{Z})^*$; this is a group of order 2^{n-2} , consisting of all the residue classes mod 2^n of integers congruent to 1 mod 4, and there is an exact sequence

$$\{1\} \rightarrow U(2^n) \rightarrow (\mathbb{Z}/2^n\mathbb{Z})^* \rightarrow (\mathbb{Z}/4\mathbb{Z})^* \rightarrow \{1\}. \quad (2)$$

Lemma 5.16 *Let x, y, k be integers with $k \geq 0$ and let $p > 2$ be a prime number. If $x \equiv 1 + py \pmod{p^2}$, then $x^{p^k} \equiv 1 + p^{k+1}y \pmod{p^{k+2}}$. If $x \equiv 1 + 4y \pmod{8}$ then $x^{2^k} \equiv 1 + 2^{k+2}y \pmod{2^{k+3}}$.*

Proof: To prove the first assertion, it is enough to show that, if $k \geq 1$ and $x \equiv 1 + p^k y \pmod{p^{k+1}}$, then $x^p \equiv 1 + p^{k+1}y \pmod{p^{k+2}}$, and then to argue by induction on the integer k . For all $a \in \mathbb{Z}$ and $k \geq 1$, it is immediate that

$$(1 + p^k a)^p \equiv 1 + p^{k+1} a \pmod{p^{k+2}},$$

hence

$$\begin{aligned} (1 + p^k y + p^{k+1} z)^p &= (1 + p^k (y + pz))^p \\ &\equiv 1 + p^{k+1} (y + pz) \equiv 1 + p^{k+1} y \pmod{p^{k+2}}. \end{aligned}$$

Similarly, for $k \geq 1$ we have $(1 + 2^{k+1} a)^2 \equiv 1 + 2^{k+2} a \pmod{2^{k+3}}$, so $(1 + 2^{k+1} y + 2^{k+2} z)^2 \equiv 1 + 2^{k+2} y \pmod{2^{k+3}}$, whence the second assertion by induction on k . \square

Proposition 5.17 *Let $p > 2$ be a prime number and let $n > 0$ be an integer; then the group $U(p^n)$ is cyclic of order p^{n-1} ; if $n \geq 2$ then the residue class*

mod p^n of an integer x congruent to 1 mod p is a generator of $U(p^n)$ if and only if x is not congruent to 1 mod p^2 . Let $m > 1$ be an integer; then the group $U(2^m)$ is cyclic of order 2^{m-2} ; if $m \geq 3$ then the residue class mod 2^m of an integer x congruent to 1 mod 4 is a generator of $U(2^m)$ if and only if x is not congruent to 1 mod 8.

Proof: Since $U(p^n)$ has order p^{n-1} , the order of every element u of $U(p^n)$ is a power of p , and u is a generator of $U(p^n)$ if and only if $u^{p^{n-2}} \neq 1$. Now if u is the class of $x = 1 + py$, then $u^{p^{n-2}}$ is the class of $1 + p^{n-1}y$, by Lemma 5.16, whence u generates $U(p^n)$ if and only if $y \not\equiv 0 \pmod{p}$, in other words $x \not\equiv 1 \pmod{p^2}$. For example, the class $1 + p$ generates $U(p^n)$. Similarly, the class u of $x \pmod{2^n}$ generates $U(2^n)$ if and only if $u^{2^{n-3}} \neq 1$, which means that x is not congruent to 1 mod 8, by Lemma 5.16; this is satisfied by $x = 5$. \square

Lemma 5.18 *Let A be a principal ideal domain and let $0 \rightarrow N \rightarrow M \rightarrow P \rightarrow 0$ be an exact sequence of A -modules. Suppose that there exist coprime elements $a, b \in A$ such that $aN = 0$ and $bP = 0$. Then the exact sequence splits. If in addition N and P are both cyclic, then M is cyclic.*

Proof: The module M is torsion, since $abM = 0$. The first assertion follows from Corollary 5.14. If N and P are cyclic, then they are finitely generated, and hence so is M ; since each p -primary component of M is isomorphic to a p -primary component either of N or of P , it follows that M is cyclic. \square

Theorem 5.19 *If $a = \prod_i p_i^{n(i)}$ is the prime decomposition of the integer $a > 1$, then the group $(\mathbb{Z}/a\mathbb{Z})^*$ of invertible elements of the ring $\mathbb{Z}/a\mathbb{Z}$ is isomorphic to the product of the groups $(\mathbb{Z}/p_i^{n(i)}\mathbb{Z})^*$. If $p > 2$ is a prime number and $n \geq 1$ an integer, then the group $(\mathbb{Z}/p^n\mathbb{Z})^*$ is cyclic of order $p^{n-1}(p-1)$. The group $(\mathbb{Z}/2\mathbb{Z})^*$ is trivial; for $n \geq 2$ the group $(\mathbb{Z}/2^n\mathbb{Z})^*$ is the direct product of the cyclic group of order 2^{n-2} generated by the residue class of 5 mod 2^n and the cyclic group of order 2 consisting of the residue class of 1 and $-1 \pmod{2^n}$.*

Proof: The orders p^{n-1} of $U(p^n)$ and $p-1$ of $(\mathbb{Z}/p\mathbb{Z})^*$ are coprime; since $U(p^n)$ and $(\mathbb{Z}/p\mathbb{Z})^*$ are cyclic, the group $(\mathbb{Z}/p^n\mathbb{Z})^*$ is cyclic (apply Lemma 5.18 to the exact sequence (1)). If $n \geq 2$ then the restriction of the homomorphism $v : (\mathbb{Z}/2^n\mathbb{Z})^* \rightarrow (\mathbb{Z}/4\mathbb{Z})^*$ to the subgroup $\{1, -1\}$ is bijective; the group $(\mathbb{Z}/2^n\mathbb{Z})^*$ is thus the direct product of this subgroup and the kernel $U(2^n)$ of v ; the result follows from Proposition 5.17. \square

6 Finiteness of Class Number

Definition 6.1 Ideals I, J of a domain A are said to be isomorphic, if and only if there is an $x \in \mathbb{Q}(A)^\times$ with $xI = J$. (Here $\mathbb{Q}(A)$ denotes the quotient field of A .)

We will be interested in domains A which are finitely generated as a module over \mathbb{Z} . For example one knows that the set of integral elements in a finite field extension of \mathbb{Q} is a finitely generated module over \mathbb{Z} .

Our goal is to prove the following important classical result, via the Structure theorem for finitely generated modules over a P.I.D.

Theorem 6.2 *Let A be a domain of characteristic zero, which is a finitely generated module over \mathbb{Z} . Then there are only finitely many isomorphism classes of ideals of A .*

(This number is called the **class number** of A .)

As a \mathbb{Z} -module the ring A is torsion-free and finitely generated, hence free. Since $K = \mathbb{Q}(A) = \{a/s \mid a \in A, s \in R \setminus (0)\}$, clearly A is of rank $n := [\mathbb{Q}(A) : \mathbb{Q}(R)]$. Fix an \mathbb{Z} -basis a_1, \dots, a_n of A . This is also a basis of $\mathbb{Q}(A)$ as a vector space over \mathbb{Q} .

Now let $I \neq (0)$ be an ideal of A . It is also a finitely generated torsion-free \mathbb{Z} -module, hence free. Let $a \in I \setminus (0)$. Then the \mathbb{Q} -linear map ‘multiplication by a ’ m_a on the \mathbb{Q} -vector-space $\mathbb{Q}(A)$ is an automorphism. Therefore aa_1, \dots, aa_n are linearly independent over \mathbb{Q} , hence over \mathbb{Z} . It follows that I , contained in A and containing aa_1, \dots, aa_n , is a free \mathbb{Z} -module of rank n and that A/I is a finite ring. We write $\|I\| := \#(A/I)$.

Lemma 6.3 *Let $a \in A$ and m_a denote ‘multiplication by a ’ on A , regarded as a free \mathbb{Z} -module. Then:*

- a) $\det(m_a) \in Aa$,
- b) $\|Aa\| = |\det(m_a)|$.

Proof: The case $a = 0$ being clear, assume $a \neq 0$. By the Elementary Divisor Theorem there is a \mathbb{Z} -basis a'_1, \dots, a'_n of A and $d_1, \dots, d_n \in \mathbb{Z}$ with $aa'_i = d_i a'_i$. So $\det(m_a) = d := d_1 \cdots d_n \in Aa$. (By a Nakayama type argument.)

Further both sides of b) are equal to $|d|$. □

As a consequence of a) we see that $I \cap \mathbb{Z} \neq (0)$ for every non-zero ideal I of A , in other words, that A/I is a torsion module over \mathbb{Z} .

Lemma 6.4 *There is an integer $C > 0$ such that in every non-zero ideal I of A there is a $\gamma \neq 0$ with $\|A\gamma\| \leq C\|I\|$, i.e. $\#(I/A\gamma) \leq C$.*

Proof: Let $m > 0$ be such that $m^n \leq \|I\| < (m+1)^n$. At least two of the following $(m+1)^n$ elements $\sum_{j=1}^n b_j a_j$, $b_j \in \mathbb{Z}$, $0 \leq b_j \leq m$ must be congruent modulo I , since $\#(A/I) < (m+1)^n$. Their difference will be our choice of C . C has the properties:

- (i) $C \in I \setminus (0)$,

$$(ii) \ C = \sum_{j=1}^n m_j a_j, \quad m_j \in \mathbb{Z}, \quad |m_j|^n \leq \|I\|,$$

The latter means $|m_j| \leq r$, where we set $r := \sqrt[n]{\|I\|}$.

Let as above a_1, \dots, a_n be a basis of A over \mathbb{Z} . Then one can write $a_i a_j = \sum_{i,j,k} a_{ijk} a_k$ with $a_{ijk} \in \mathbb{Z}$. For $\gamma = \sum_j m_j a_j$ (with $m_j \in \mathbb{Z}$) we have $\gamma a_i = \sum_{j,k} m_j a_{ijk} a_k$. So $\det(h_\gamma)$ is a homogeneous polynomial of degree n in the m_j over \mathbb{Z} . Therefore there is a $C > 0$ such that $|m_j| \leq r$ implies $|\det(h_\gamma)| \leq Cr^n = C\|I\|$. \square

Proof of Theorem 6.2: Let $c \in \mathbb{Z}$ be the product of all $a \in \mathbb{Z} \setminus (0)$ with $|a| \leq C$. (There are only finitely many of them.) We will show that every ideal $I \neq (0)$ of A is isomorphic to one between A and Ac . Since A/Ac is a finite ring there are only finitely many of the latter.

Choose γ as in Lemma 6.4. Then $I\gamma^{-1}/A \cong I/A\gamma$ is of order $\leq C$. By the Structure theorem, There is a \mathbb{Z} -module isomorphism $I\gamma^{-1}/A \cong R/(d_1) \oplus \dots \oplus R/(d_m)$ with suitable $d_i \in \mathbb{Z} \setminus (0)$. Then $|d_i| \leq C$, whence $d_i | c$. So $c(I\gamma^{-1}/A) = 0$, i.e. $c\gamma^{-1}I \subset A$. On the other hand $cA \subset c\gamma^{-1}I$, and we are done. \square

7 Equivalent class of matrices over a P.I.D.

In this section we study the class of (two-sided) equivalent matrices over a P.I.D. Here the canonical forms are called Smith normal form. (In the one-sided equivalent situation, they are called Hermite normal form; which have their own interesting features.)

We restrict to the two-sided equivalent situation as it enables us to reprove the Structure theorem for finitely generated modules over a P.I.D. (Equivalently, the Structure theorem can also be used to prove the matrix theorem.)

Definition 7.1 Let R be a commutative ring with 1. Two matrices $\alpha, \beta \in M_{r,s}(R)$ are said to be **equivalent** if there are invertible matrices γ, δ , of appropriate sizes, such that $\alpha = \gamma\beta\delta$. If $\delta = \gamma^{-1}$, we say that α and β are **similar**.

Equivalence, and similarity define an equivalence relation on $M_{r,s}$. The similarity class will be studied in the next section.

For example, two matrices over a field K are equivalent if and only if they have the same rank. But over a commutative ring, one does not have such a simple criterion.

Definition 7.2 Let α be a nonzero $m \times n$ matrix of rank r over a unique factorisation domain R . For $1 \leq t \leq r$, the t -rowed minors of α are not all zero; call their g.c.d. Δ_t . The r -elements $\Delta_1, \dots, \Delta_r$ are defined upto unit factors, and are called **the determinantal divisors** of α .

For $i < r$, every minor of order $i + 1$, is a linear combination of certain minors of order i , and so is divisible by Δ_i ; hence, $\Delta_i | \Delta_{i+1}$, for $1 \leq i \leq r$.

Definition 7.3 The elements $d_1 = \Delta_1, d_2 = \Delta_2/\Delta_1, \dots, d_r = \Delta_r/\Delta_{r-1}$, which are defined upto unit factors, are called **the invariant factors** of α .

Note one can recover the determinantal factors if one knows the invariant factors, and vice versa.

Example: Let $D = \text{diag}(d_1, d_2, \dots, d_r)$, with $d_i | d_{i+1}$, for $1 \leq i \leq r-1$. Then the determinantal divisors of D are $d_1, d_1 d_2, \dots, d_1 d_2 \dots d_r$; and the invariant factors are d_i .

We shall use the following notation below: For t -tuples $\mathbf{i} = (i_1, \dots, i_t)$, $\mathbf{j} = (j_1, \dots, j_t)$ of natural numbers $i_1 < \dots < i_t \leq m$ and $j_1 < \dots < j_t \leq n$, call $m_{\mathbf{i}, \mathbf{j}}$ the t -rowed minor of $\alpha \in M_{m,n}(R)$ with row indices i_1, \dots, i_t and column indices j_1, \dots, j_t .

Lemma 7.4 Let $\alpha = (\alpha_{ij}), \beta = (\beta_{jk})$, be $m \times n$ and $n \times p$ matrices over a commutative ring R , and let $\gamma = \alpha\beta$. Then $\gamma_{\mathbf{i}, \mathbf{k}} = \sum_{\mathbf{j}} \alpha_{\mathbf{i}, \mathbf{j}} \beta_{\mathbf{j}, \mathbf{k}}$, where \mathbf{j} runs over all t -couplets of numbers satisfying $1 \leq j_1 < \dots < j_t \leq n$.

Proof: Left to the reader. □

Proposition 7.5 Equivalent nonzero matrices α, β , over a unique factorization domain R have the same determinantal factors.

Proof: Let $\beta = \gamma\alpha\delta$, for suitable invertible matrices γ, δ . Let r be the common rank of α, β . By Lemma 7.4, for $t \leq r$, the t -rowed minors of $\beta = \gamma(\alpha\delta)$ is a R -linear combination of the t -rowed minors of $\alpha\delta$ and hence also of the t -rowed minors of α . Therefore they are all divisible by $\Delta_t(\alpha)$, and hence their g.c.d. $\Delta_t(\beta)$ is divisible by $\Delta_t(\alpha)$. Similarly, $\Delta_t(\alpha)$ is divisible by $\Delta_t(\beta)$. □

The converse of above proposition is false: The matrices

$$\begin{pmatrix} X-1 & -1 \\ 0 & X-1 \end{pmatrix}, \begin{pmatrix} X-1 & -2 \\ 0 & X-1 \end{pmatrix},$$

over $\mathbb{Z}[X]$ have the same determinantal divisors 1 and $(X-1)^2$ but they are not equivalent over $\mathbb{Z}[X]$. (Why?)

However, over P.I.D.'s the converse is true; this follows from

Theorem 7.6 (Smith Normal Form) Let A be a P.I.D. and $\alpha \in M_{r \times s}(A)$. Then there are $\varepsilon \in \text{Sl}_r(A), \varepsilon' \in \text{Sl}_s(A)$ such that $\varepsilon\alpha\varepsilon' = \text{diag}(d_1, \dots, d_t)$ with $d_i | d_{i+1}$, for $1 \leq i \leq t-1$, where $t := \text{Min}\{r, s\}$.

(We do not exclude the case that there is a $j \geq 1$ with $d_j = d_{j+1} = \dots = d_t = 0$. It should be clear what we mean by $\text{diag}(d_1, \dots, d_t)$ also if $r \neq s$. It's not the geometric diagonal.)

(The elements d_i are called the **invariant factors** of α . The set of prime powers of which occur as divisors of some invariant factor, including repetitions, is called the set of **elementary divisors** of α).

Proof: Consider the set $S := \{\sigma\alpha\tau \mid \sigma \in \text{Sl}_r(A), \tau \in \text{Sl}_s(A)\}$. In S choose a matrix $\alpha' = (a'_{ij})$ such that Aa'_{11} is maximal among all ideals Aa''_{11} for $\alpha'' = (a''_{ij}) \in S$.

We claim that a'_{11} divides all entries in the first row and first column of α' . Assume e.g., it does not divide a'_{12} , so that $Ad \supsetneq Aa'_{11}$, if d is a g.c.d. of a'_{11} and a'_{12} . Let $d = b_1a'_{11} + b_2a'_{12}$. Then for

$$\tau := \begin{pmatrix} b_2 & -a'_{11}/d & 0 \\ b_1 & a'_{12}/d & 0 \\ 0 & 0 & I_{s-2} \end{pmatrix} \in \text{Sl}_s(A) \quad \text{we get} \quad \alpha'\tau = \begin{pmatrix} d & * \\ * & * \end{pmatrix}.$$

This contradicts the maximality condition on Aa'_{11} .

Then we proceed in the obvious way. □

Corollary 7.7 *Two nonzero $m \times n$ matrices α, β over a P.I.D. R are equivalent if and only if they have the same invariant factors, or equivalently, the same determinantal divisors.*

Proof: If they are equivalent, they have the same determinantal factors, and so the same invariant factors. Conversely, they are both equivalent to the same diagonal matrix $\text{diag}\{d_1, \dots, d_r, 0, \dots, 0\}$. □

Example.

Companion matrix of a monic polynomial f . Let $f(X) = X^n + a_1X^{n-1} + \dots + a_n$ be any monic polynomial over a commutative ring R . The matrix $C(f)$ given by

$$C(f) = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ -a_n & -a_{n-1} & -a_{n-1} & \cdots & -a_1 \end{pmatrix}$$

is known as the companion matrix of f . This is the matrix of the linear transformation $m_{\bar{X}}$ of the vector space $K[X]/(f(X))$ with respect to the basis $\{\bar{1}, \bar{X}, \dots, \bar{X}^{n-1}\}$

Lemma 7.8 *The characteristic polynomial of $C(f)$ is $\det(XI - C(f))$ which equals f , and its invariant factors d_i is 1 for $i < n$.*

Proof: The first assertion is shown by induction on $n = \deg(f)$. It is easy to check for $n = 2$. Let $C(f)_{n-1}$ be the submatrix of $C(f)$ consisting of the last $n - 1$ rows and columns. By expanding along the first column

$$\det(XI - C(f)) = X\det(XI - C(f)_{n-1}) + a_n.$$

Now use induction hypothesis to calculate $\det(XI - C(f)_{n-1})$, and resolve.

For each k , $1 \leq k \leq n-1$, consider the $k \times k$ submatrix of $(XI - C(f))$ obtained by choosing rows $1, 2, \dots, k$, and columns $2, 3, \dots, k+1$. Its diagonal elements are all -1 , the elements on the subdiagonal are all X , and the remaining elements are zero. The determinant of this matrix is therefore $(-1)^k$. Therefore, $\Delta_k(XI - C(f)) = 1$, for $1 \leq k \leq n-1$, and so the invariant factors coincide with it. \square

Some applications of the Smith normal form

1. Let R be a P.I.D. and let $a \in R$. Let $\alpha \in M_n(R)$ such that $\det(\alpha) = 1 \pmod{a}$. Then there is a matrix $\beta \in M_n(R)$ such that $\beta \equiv \alpha \pmod{a}$, and with $\det(\beta) = 1$.

Proof: Let $\alpha = \gamma D \delta$, with $\det(\gamma) = 1 = \det(\delta)$, and with D a diagonal matrix. It suffices to show that there is a $\theta \in M_n(R)$ such that $\theta \equiv D \pmod{a}$ and $\det(\theta) = 1$: for then take $\beta = \gamma \theta \delta \equiv \gamma D \delta \equiv \alpha \pmod{a}$, and $\det(\beta) = 1$.

Let $D = \text{diag}\{d_1, \dots, d_n\}$. Then $d_1 d_2 \dots d_n = 1 + ad$, say. Consider

$$\theta = \begin{pmatrix} d_1 + aX & a & & \\ & d_1 & a & \\ & & \dots & \\ aY & & & d_n \end{pmatrix}$$

Then $\theta \equiv D \pmod{a}$, and

$$\begin{aligned} \det(\theta) &= (d_1 + aX)d_2 \dots d_n + a^n Y \\ &= 1 + a(d + d_2 + \dots d_n X + a^{n-1} Y) \end{aligned}$$

Since $\text{g.c.d.}(d_2, \dots, d_n, a) = 1$, one can solve the equation $(d_1 + aX)d_2 \dots d_n + a^n Y = 0$. For this choice of X, Y , $\det(\theta) = 1$.

2. An interesting application to diophantine analysis. Let $\alpha \in M_{r,n}(\mathbb{Z})$, $r \leq n$. Assume that $\text{rank}(\alpha) = r$. Let v an $r \times 1$ vector over \mathbb{Z} . Consider the problem of finding all $n \times 1$ vectors x over \mathbb{Z} which satisfy $\alpha x = v$.

One can find γ, δ of determinant one such that $\gamma \alpha \delta = (D \ 0)$ in Smith normal form, where $D = \text{diag}(s_1, s_2, \dots, s_r)$, $s_i | s_{i+1}$, $1 \leq i \leq r-1$. Set $\delta^{-1}x = y$, $\gamma v = w$. Then $\alpha x = v$ implies $(D \ 0)y = w$. If $y = (y_1, \dots, y_n)^t$, $w = (d_1, \dots, d_r)^t$, then the variables y_{r+1}, \dots, y_n may be chosen in arbitrary fashion, but the variables y_1, \dots, y_r must satisfy $s_i y_i = d_i$, for $1 \leq i \leq r$. So all the solutions can be found in terms of the $n-r$ parameters y_{r+1}, \dots, y_n , if there is a solution of the auxiliary equations $s_i y_i = d_i$, for $1 \leq i \leq r$.

3. Given objects O_1, O_2, \dots, O_p and sets S_1, S_2, \dots, S_q containing these objects, put $a_{ij} = 1$ if O_i belongs to S_j , and 0 otherwise. The $p \times q$ matrix $\alpha = (a_{ij})$ so obtained is known as an *incidence* matrix. A fundamental combinatorial problem is to determine whether a given incidence matrix can be obtained from another by permuting rows and columns. This problem is clearly a finite one, since it is only necessary to form all $p!q!$ products and check to see if the matrix is in the equivalence class. However this procedure is not feasible, in general, because of the size of $p!q!$. A useful negative criterion is that if the two matrices are not equivalent then they certainly cannot be ‘permutation equivalent’.

4. We now derive the Elementary divisors theorem from the Smith Normal Form (also called Elementary Divisors Theorem for matrices over a P.I.D.). As before, the Structure theorem will then follow.

Corollary 7.9 (Structure theorem over a P.I.D.)

Let R be a P.I.D.

- a) *Every submodule M of R^r is free of some rank $m \leq r$, and there are basis f_1, \dots, f_r of R^r and m_1, \dots, m_k of M , such that $m_i = d_i f_i$ for suitable $d_i \in R$, $i \leq k$, satisfying $d_i | d_{i+1}$, for $1 \leq i \leq r-1$.*
- b) *Every finitely generated R -module M is of the form*

$$R/Rd_1 \oplus R/Rd_2 \oplus \cdots \oplus R/Rd_r$$

for some $d_i \in R$. ($d_i = 0$ is not excluded.) Moreover, $d_i | d_{i+1}$, for $1 \leq i \leq r-1$. Moreover, $(d_k) = \text{ann}(M)$. For $1 \leq j \leq k$, $d_1 \dots d_j = \text{g.c.d.}$ of the j -th order minors of the matrix whose columns are the coordinate vectors of the m_i with respect to any basis of R^r .

Proof: a) Choosing a finite generating system of M one gets a linear map

$$A^s \rightarrow A^r \text{ given by a matrix } \alpha.$$

Changing bases means multiplying α on both sides by invertible matrices. One can therefore assume that α is a diagonal matrix $\text{diag}\{d_1, \dots, d_k, 0, \dots, 0\}$. The rest is clear.

b) We have already seen how the Structure theorem follows from the Elementary divisor theorem in Theorem 5.9. The assertions about the d_i : In view of Proposition 7.5, we need to only check it for the Smith normal form; where it is easily established, due to the divisibility conditions on the d_i . \square

Corollary 7.10 *Let $M \simeq F/G$ be a finitely generated module over the P.I.D. R , with $F \simeq R^n$, and $G \simeq R^r$. Let $\mathbf{f} = (f_1, \dots, f_n)$, $\mathbf{g} = (g_1, \dots, g_r)$ be basis of F , G respectively, and write $\mathbf{g} = \mathbf{f}\alpha$, for some $\alpha \in M_{n,r}(R)$. Then $\text{rank}(\alpha) = r$, $\text{rank}(M) = n - r$, the invariants of M are the non-units among the invariant factors of α , and the elementary divisors of M are the elementary divisors of α .*

Proof: Suppose that $\text{rank}(\alpha) = s$ and invariant factors of α are d_1, \dots, d_s . Then there are invertible matrices γ, δ such that $\gamma\alpha\delta = \beta \perp \{0\}$, with $\beta = \text{diag}(d_1, \dots, d_s)$. Set $\mathbf{f}' = \mathbf{f}\gamma^{-1}$, $\mathbf{g}' = \mathbf{g}\delta$; these are basis of F , G , respectively, and we have $\mathbf{g}' = \mathbf{g}\delta = \mathbf{f}\alpha\delta = \mathbf{f}'\gamma\alpha\delta = \mathbf{f}'(\beta \perp \{0\})$.

If $s < r$, then $\mathbf{g}'_{s+1}, \dots, \mathbf{g}'_r$ are all zero; this being impossible, $s = r$ and $\mathbf{g}'_1 = d_1 \mathbf{f}'_1, \dots, \mathbf{g}'_r = d_r \mathbf{f}'_r$.

If d_k, \dots, d_r are the non-units among d_1, \dots, d_r , then M is the direct sum of a free R -module of rank $n - r$ and cyclic R -modules having order ideals $d_k R, \dots, d_r R$. Therefore M has rank $n - r$ and invariant factors d_k, \dots, d_r .

Finally, the elementary divisors of M , being the prime powers occurring in d_k, \dots, d_r , are the prime powers $\neq 1$ occurring in d_1, \dots, d_r , and are therefore the elementary divisors of α . \square

5. We now apply the above theory to a linear operator T on a vector space V over a field F ; equivalently, as we shall see below, to modules over the Euclidean domain $F[X]$. Before we do that, we recall the well-known

Theorem 7.11 (Cayley-Hamilton)

Let R be a commutative ring with 1. Let $\alpha \in M_n(R)$. Let

$$\chi_\alpha(X) = \det(XI_n - \alpha) = a_0 + a_1X + \cdots + a_{n-1}X^{n-1} + X^n$$

be the characteristic polynomial of α . Then α satisfies its characteristic polynomial, i.e.

$$a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} + \alpha^n = 0.$$

Proof: One has a division algorithm over $M_n(R[X])$, with unique quotient and remainder, if one wishes to divide by a monic polynomial in $M_n(R[X])$.

Let us use this to divide the element $\chi_\alpha(X)I_n = a_0I_n + a_1XI_n + \cdots + a_{n-1}X^{n-1}I_n + X^nI_n$ by the linear polynomial $XI_n - \alpha$: we find that there are unique elements $q(X) \in M_n(R[X])$, $r \in M_n(R)$, such that

$$a_0I_n + a_1XI_n + \cdots + a_{n-1}X^{n-1}I_n + X^nI_n = q(X)(XI_n - \alpha) + r.$$

But $\text{adj}(XI_n - \alpha)(XI_n - \alpha) = \det(XI_n - \alpha)I_n$. Hence, the uniqueness gives $r = 0$. Now substitute $X = \alpha$! \square

Definition 7.12 *The monic polynomial of minimal degree satisfied by α is called its **minimal polynomial** $m_\alpha(X)$.*

If K is a field, the minimal polynomial $m_\alpha(X)$ is the monic generator of the ideal $\{f(X) \in K[X] \mid f(\alpha) = 0\}$. By Cayley-Hamilton theorem, $m_\alpha(X) \mid \chi_\alpha(X)$.

Remark. In the theory of canonical forms for linear transformation, the characteristic polynomial plays the role of the order of a finite abelian group and the minimal polynomial plays the role of the exponent.

The $K[X]$ -module structure on V

Let V be a vector space over a field K . Regard the field K as a subfield of the ring $\text{End}_K(V)$ by identifying the elements a of K with the elements a_V of $\text{End}_K(V)$. Since the element α of $\text{End}_K(V)$ commutes with the elements of the subfield K of $\text{End}_K(V)$, therefore there is a (unique) K -homomorphism from the polynomial ring $K[X]$ into the ring $\text{End}_K(V)$ which maps X to α , and hence maps $f(X) \in K[X]$ to $f(\alpha) \in \text{End}_K(V)$. (Note that this homomorphism is not injective, because the powers of X are linearly independent over K , whereas the powers of α are linearly dependent, by Cayley-Hamilton theorem.)

As V has the structure of a module over $\text{End}_K(V)$, therefore, by restriction of scalars, V has the structure of a module over $K[X]$ in which

$$f(X)v = (f(\alpha))(v) = c_0v + c_1\alpha(v) + c_2\alpha^2(v) + \cdots + c_r\alpha^r(v),$$

for every polynomial $f(X) = c_0 + c_1X + \cdots + c_rX^r$ in $K[X]$ and every $v \in V$. This $K[X]$ -module will be denoted by V_α .

Conversely, if M is a module over $K[X]$, and m_X denotes the $K[X]$ -module map ‘multiplication by X ’, then one has $M \simeq M_{m_X}$. Thus, the theory of linear operators on a vector space over K , and $K[X]$ -modules are interchangeable concepts.

Now suppose that V is a finite dimension vector space, and let v_1, \dots, v_n generate V over K . Therefore they generate V over $K[X]$; in particular, the $K[X]$ -module V_α is finitely generated. Moreover, the following holds:

Lemma 7.13 *The $K[X]$ -module V_α is a torsion module having order ideal the ideal of $K[X]$ generated by the minimal polynomial $m(X)$ of α .*

Proof: Indeed, $f(X)V_\alpha = 0$ if and only if $f(\alpha)(V_\alpha) = 0$, i.e. if and only if $f(\alpha) = 0$; hence $f(X)$ is divisible by $m(X)$. Thus $(0 : V_\alpha) = (m(X))$. \square

The following theorem describes the structure of the $K[X]$ -module V_α .

Theorem 7.14 *The invariant factors of the $K[X]$ -module V_α are the non-units $d_k(X), \dots, d_n(X)$ among the invariant factors $d_1(X), \dots, d_n(X)$ of the matrix $XI_n - \alpha$ over the Euclidean domain $K[X]$; hence the $K[X]$ -module V_α (being a torsion module) is the direct sum of cyclic submodules having order ideals $(d_k(X)), \dots, (d_n(X))$.*

Proof: Let F be a free $K[X]$ -module having a basis of n elements u_1, \dots, u_n , and let $\varphi : F \rightarrow V_\alpha$ be the $K[X]$ -linear map which maps u_i to v_i ($1 \leq i \leq n$). Since v_1, \dots, v_n generate V , therefore φ is surjective, and hence $V_\alpha \simeq F/G$ with $G = \ker(\varphi)$. Let $\mathbf{t} = (t_1, \dots, t_n)$ be the n -tuple of elements of F given by $\mathbf{t} = \mathbf{u}(XI_n - \alpha)$, or equivalently, by $t_j = Xu_j - \sum_{i=1}^n a_{ij}u_i$. Since

$$\varphi(t_j) = Xv_j - \sum_{i=1}^n a_{ij}v_i = \alpha(v_j) - \sum_{i=1}^n a_{ij}v_i = 0,$$

therefore, t_1, \dots, t_n belong to G . We claim that $\mathbf{t} = \mathbf{u}(XI_n - \alpha)$ is a basis of G ; then apply Corollary 7.10. The $K[X]$ -independence of t_1, \dots, t_n is obvious since $\det(XI_n - A) \neq 0$. Let now $G' = K[X]t_1 + \cdots + K[X]t_n$ be the submodule of F generated by t_1, \dots, t_n . Since $t_i \in G$ ($1 \leq i \leq n$), therefore $G' \subset G$. Let $U = Ku_1 + \cdots + Ku_n$. By above equation, $Xu_j \in G' + U$, hence $XU \subset G' + U$, and it follows that $G' + U$ is a submodule of the $K[X]$ -module F . Since $G' + U$ contains the $K[X]$ -basis u_1, \dots, u_n of F , therefore $G' + U = F$. Let $x \in G$, and write $x = t' + \sum_i c_i u_i$ with $t' \in G'$ and $c_i \in K$ ($1 \leq i \leq n$). Apply φ and noticing that $\varphi(x) = 0$, $\varphi(t') = 0$, we obtain $\sum_i c_i v_i = 0$. Hence each c_i is zero, and therefore $x = t' \in G'$. Thus $G = G' = K[X]t_1 + \cdots + K[X]t_n$, and the proof is complete. \square

Theorem 7.15 *The elementary divisors of the $K[X]$ -module V_α are the elementary divisors of the matrix $XI_n - \alpha$; if these are $p(X)^a, q(X)^b, \dots$, then V_α is a direct sum of cyclic submodules having order ideals $(p(X)^a), (q(X)^b), \dots$*

Proof: This is obvious from Theorem 7.14. □

In order to avoid unnecessary constants, we shall take the minimal polynomial $m(X)$ of α , and the invariant factors, the determinantal divisors and the elementary divisors of the matrix $XI_n - \alpha$ to be all monic.

The conclusion below is a composite of results due to Hamilton, Cayley, and Frobenius.

Corollary 7.16 *Let K be a field, and $\alpha \in M_n(K)$. The minimal polynomial $m_\alpha(X)$ of α is the n^{th} invariant factor $d_n(X)$ of the characteristic matrix $XI_n - \alpha$:*

$$m(X) = d_n(X) = \Delta(X)/\Delta_{n-1}(X) \quad (3)$$

where $\Delta(X) = \det(XI_n - \alpha) = \chi_\alpha(X)$ is the characteristic polynomial of α (or of A), and $\Delta_{n-1}(X)$ the (monic) g.c.d. of the minors of order $n - 1$ of the matrix $XI_n - \alpha$. Moreover, $\chi_\alpha(X)$ and $m_\alpha(X)$ have the same prime factors in $K[X]$. All the invariant factors divide $m(X)$.

Proof: By Lemma 7.11, the $K[X]$ -module V_α has order ideal $(d_n(X))$; but as noted earlier, this order ideal is $(m(X))$. Therefore, $m(X) = d_n(X) = \Delta(X)/\Delta_{n-1}(X)$. One has $\chi_\alpha(X) = d_1(X) \dots d_n(X)$. Since each $d_i(X)$ divides $m_\alpha(X)$, the last assertion follows. □

Corollary 7.17 *The $K[X]$ -module V_α is cyclic if and only if $m(X) = \Delta(X)$.*

Proof: The $K[X]$ -module V is cyclic if and only if it has only one invariant factor, i.e. if and only if $d_1(X), \dots, d_{n-1}(X)$ are all 1. This means that $\Delta_{n-1}(X) = 1$ or that $m(X) = \Delta(X)$. □

Companion matrices and the rational canonical forms.

Let $W = K[X]w$ be a non-zero cyclic submodule of the $K[X]$ -module V_α , and suppose that the order ideal of W is generated by the (non-constant) polynomial

$$f(X) = X^m + a_1X^{m-1} + \dots + a_m, \text{ i.e. } W \simeq K[X]/(f(X)).$$

Thus the subspace W of the vector space V over K is stable under α , and by Corollary 7.15 the characteristic and the minimal polynomials of the restriction $\beta = \alpha|_W$ of α to W are equal. Since the minimal polynomial of β is clearly $f(X)$ (because $g(\beta) = 0$ means that $g(X)$ annihilates W and hence is a multiple of $f(X)$), therefore $f(X)$ is also the characteristic polynomial of β . In particular, $\dim_K(W) = m$. Since the m elements

$$(\overline{X}) = w, \beta(w), \dots, \beta^{m-1}(w) = (\overline{X}^{m-1}) \quad (4)$$

of W are K -independent, they constitute a K -basis of W . With respect to this basis, β has matrix $C(f(X))$, the companion matrix of $f(X)$. The characteristic and minimal polynomials of $C(f(X))$, being the characteristic and minimal polynomials of $\beta = \alpha|_W$, are both equal to $f(X)$.

If we decompose the $K[X]$ -module V into cyclic summands W_1, W_2, \dots , having order ideals $(f_1(X)), (f_2(X)), \dots$, and choose in each summand a basis as explained above, then, putting these bases together, we obtain a basis of V with respect to which α has matrix $C(f_1(X)) \perp C(f_2(X)) \perp \dots$. In particular, decomposing the $K[X]$ -module V_α according to Theorem 7.14 (respectively Theorem 7.15), we obtain:

Proposition 7.18 *If $C(f_1(X)), C(f_2(X)), \dots, C(f_k(X))$ are the companion matrices of the invariant factors $\neq 1$ (resp: elementary divisors) of the matrix $XI_n - \alpha$, then α is similar to the matrix $\{C(f_1(X)) \perp C(f_2(X)) \perp \dots \perp C(f_k(X))\}$. \square*

8 Similarity class of matrices over a P.I.D.

The similarity classes of the square matrices $M_n(R)$ can be reduced to a study of equivalent classes over $M_n(R[X])$ in view of the following

Lemma 8.1 *Let α, β be $n \times n$ matrices over a commutative ring R . Then $XI_n - \alpha$ and $XI_n - \beta$ are equivalent over $R[X]$ if and only if they are equivalent over R .*

Proof: Let $XI_n - \beta = Q(X)(XI_n - \alpha)P(X)$, for some invertible matrices $P(X), Q(X)$. Let

$$\begin{aligned} P(X) &= P_1(X)(XI_n - \beta) + P_0, \\ Q(X) &= (XI_n - \beta)Q_1(X) + Q_0. \end{aligned}$$

(Note that we have reversed the order in the second division). With a little jugglery (Do it!), one shows that

$$(XI_n - \beta) - Q_0(XI_n - \alpha)P_0 = (XI_n - \beta)S(XI_n - \beta)$$

with $S = Q_1Q^{-1} + P^{-1}P_1 - Q_1(XI_n - \alpha)P_1$.

Since the right hand side has degree at most one, $S = 0$. \square

Corollary 8.2 (The fundamental theorem on similarity)

Two $n \times n$ matrices α, β over a commutative ring R are similar if and only if their characteristic polynomials $XI_n - \alpha, XI_n - \beta$ are equivalent over $R[X]$.

Proof: If α, β are similar then so are $XI_n - \alpha$, and $XI_n - \beta$. If $XI_n - \beta, XI_n - \alpha$ are equivalent over $R[X]$, then they are equivalent over R by Lemma 8.1. Let $(XI_n - \beta) = \gamma(XI_n - \alpha)\delta$. Hence, on comparing coefficients, one has $\gamma\delta = I_n$, and $\gamma\alpha\delta = \beta = \gamma\alpha\gamma^{-1}$. \square

Corollary 8.3 *Two square matrices of order n over a field K are similar if and only if their characteristic matrices have the same invariant factors, or equivalently, the same determinantal divisors.*

Proof: Clear from Corollary 8.2. \square

Corollary 8.4 *Let $f \in K[X]$ be a monic polynomial of degree n , where K is a field. Then the characteristic matrix $XI_n - C(f)$ of the companion matrix $C(f)$ of f is similar to $I_{n-1} \perp \{f\}$.*

Proof: We have computed the invariant factors of $C(f)$ earlier, which are precisely $d_i = 1$, $1 \leq i \leq n-1$, and $d_n = f$. Now apply Corollary 8.3. \square

Corollary 8.5 *Let $\alpha \in M_n(K)$, where K is a field. Suppose that minimal polynomial and characteristic polynomial of α coincide. Then the characteristic matrix $XI_n - \alpha$ is equivalent to $I_{n-1} \perp \{f\}$, where f is the minimal polynomial of α .*

Proof: By Corollary 7.7, Corollary 7.9 the n -th invariant factor $d_n(XI_n - \alpha)$ is the minimal polynomial $m(X)$ of α , that is the monic polynomial of least degree which annihilates α . Moreover, since similar matrices have the same characteristic polynomial, the characteristic polynomial $d(X) = \det(XI_n - \alpha) = \prod_{k=1}^n d_k(XI_n - \alpha)$.

Therefore, $m(X) = d(X)$, if and only if $d_k(XI_n - \alpha) = 1$, for $i < n$, and $d_n(XI_n - \alpha) = m(X)$. By Corollary 8.4 this is if and only if $XI_n - \alpha$ is equivalent to $I_{n-1} \perp \{f\}$, \square

Corollary 8.6 *Let $\alpha \in M_n(K)$, where K is a field. Suppose that minimal polynomial and characteristic polynomial of α coincide. Then α is similar to the companion matrix of its minimum polynomial.*

Proof: Let f be the minimum polynomial of α . Since it coincides with the characteristic polynomial it has degree n . By Corollary 7.14, or otherwise, the characteristic and minimal polynomial of $C(f)$ coincide. Hence, by Corollary 8.5, $XI_n - C(f)$ is equivalent to $I_{n-1} \perp \{f\}$, which is equivalent to $XI_n - \alpha$. Now apply Corollary 8.2. \square

The next theorem is one of the classical results on canonical forms for similarity over a field.

Theorem 8.7 (Frobenius normal form)

Let $\alpha \in M_n(K)$, where K is a field. Let d_1, \dots, d_k be the invariant factors of $XI_n - \alpha$. Then α is similar to $C(d_1) \perp \dots \perp C(d_k)$.

Proof: We have $XI - C(d_i)$ equivalent to $\text{diag}(1, \dots, 1, d_i)$, for $1 \leq i \leq k$.

Hence $XI - \alpha$, and $XI - \{C(d_1) \perp \dots \perp C(d_k)\}$ both have invariant factors d_1, \dots, d_k ; and so by Corollary 7.7 are equivalent. By the fundamental theorem of similarity, α and $\{C(d_1) \perp \dots \perp C(d_k)\}$ are similar. \square

This normal form is also referred to as the **rational normal form**. The rational canonical form of α is unique (via Corollary 8.3). (In particular, the rational canonical form over an extension field L is the same as that over K .)

An immediate consequence of the existence of the rational normal form is the Cayley-Hamilton theorem.

Remark. The rational canonical form is analogous to decomposing a finite abelian group as a direct product of cyclic groups.

The Jordan canonical form. This form is only available when $\alpha \in M_n(K)$ has all its eigenvalues in the field K . We shall assume this in the sequel. (Or just assume that we are working over the field \mathbb{C} of complex numbers).

Jordan matrix $J_n(a)$ of $a \in R$. The $n \times n$ matrix

$$J_n(a) = \begin{pmatrix} a & 1 & 0 & \cdot & \cdots & \cdot & \cdot & 0 \\ 0 & a & 1 & \cdot & \cdots & \cdot & \cdot & 0 \\ \cdot & \cdot & a & 1 & \cdots & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & a & 1 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & a & 1 \\ 0 & 0 & 0 & \cdot & \cdots & \cdot & 0 & a \end{pmatrix}$$

is known as n -Jordan matrix of $a \in R$. This is the matrix of the linear transformation $m_{\overline{X}}$ of the vector space $K[X]/(X - \lambda)^n$ with respect to the basis $\{(\overline{X} - \lambda)^{n-1}, (\overline{X} - \lambda)^{n-2}, \dots, (\overline{X} - \lambda), \overline{1}\}$.

The characteristic polynomial of $J_n(a)$ is $(X - a)^n$. Therefore its minimal polynomial must be of the form $(X - a)^k$, for some k , such that $1 \leq k \leq n$. But the least positive integer k such that $(J_n(a) - aI_n)^k = 0$ is n . Hence, $k = n$, and the minimal and characteristic polynomial of $J_n(a)$ coincide. Hence, by Corollary 8.5, $XI_n - J_n(a)$ is equivalent to $I_{n-1} \perp \{(X - a)^n\}$.

Theorem 8.8 (Jordan normal form)

Let $\alpha \in M_n(K)$ have all its eigenvalues in the field K . Let $p_i(X) = (X - d_i)^{e_i}$, $1 \leq i \leq k$, denote the elementary divisors of $XI_n - \alpha$, where the elements a_1, a_2, \dots, a_k belong to K , but are not necessarily distinct. Then α is similar to $\beta = J_{e_1}(p_1(X)) \perp \cdots \perp J_{e_k}(p_k(X))$, with $k = \dim \ker(XI - \alpha)$. The Jordan canonical form is unique up to a permutation of the Jordan blocks along the diagonal.

Proof: The characteristic matrix of a Jordan matrix is also a Jordan matrix. Hence its minimal and characteristic polynomial coincide. Hence, $XI - J_{e_i}(p_i(X))$ has $p_i(X)$ as its only elementary divisor, for $1 \leq i \leq k$. Thus $XI_n - \alpha$ and $XI_n - \beta$ have the same elementary divisors, and so are equivalent. By the fundamental theorem of similarity, α and β are similar.

The last (but one) assertion is clear as each Jordan block contributes 1 to the dimension of $\ker(XI - \alpha)$. (Exercise 54(a)).

Uniqueness: Apply Corollary 8.3. □

Corollary 8.9 (Jordan decomposition)

Let K be a field. Let $\alpha \in M_n(K)$. If the characteristic polynomial of α splits into linear factors, then $\alpha = D + N$, where $DN = ND$, and N is nilpotent, while D is a diagonalizable matrix.

Proof: Let $\beta = \gamma\alpha\gamma^{-1}$ be in Jordan normal form. If we can prove the theorem for β , then it will follow for α . But it is clear that every Jordan matrix has this type of decomposition; whence so will β have it. \square

Corollary 8.10 A matrix $\alpha \in M_n(K)$, which has all its eigenvalues in K , is diagonalizable if and only if each elementary divisor is linear.

Proof: α is similar to $J_{e_1}(p_1(X)) \perp \cdots \perp J_{e_k}(p_k(X))$. If α is diagonalizable, then by uniqueness, each $J_{e_i}(p_i(X))$ must be of size 1. \square

We reprove Corollary 7.7.

Corollary 8.11 Two matrices α and β over a field K are similar if and only if they have the same invariant factors or, equivalently, the same elementary divisors.

Proof: The Jordan normal form is entirely determined by the invariant factors and these are similarity invariants. So the result holds when K is algebraically closed. For the general case, go to the algebraic closure \overline{K} of K . \square

For the convenience of the reader we list (without proof):

The structure of a Jordan matrix.

The Jordan matrix $J_{n_1}(\lambda_1) \perp \cdots \perp J_{n_k}(\lambda_1)$, $n = n_1 + \dots + n_k$, has a definite structure:

1. The number k of Jordan blocks (counting multiple occurrences of the same block) is the number of linearly independent eigenvectors of J .
2. The matrix J is diagonalizable if and only if $k = n$.
3. The number of Jordan blocks corresponding to a given eigenvalue is the **geometric multiplicity** of the eigenvalue, which is the dimension of the associated eigenspace $E(\lambda)$. The sum of the orders of all the Jordan blocks corresponding to a given eigenvalue is the **algebraic multiplicity** of the eigenvalue, which is the number of times it occurs as a root of the characteristic polynomial $\chi_\alpha(X)$.
4. A Jordan matrix is **not** completely determined in general by a knowledge of the eigenvalues and their algebraic and geometric multiplicity. One must also know the sizes of the Jordan blocks corresponding to each eigenvalue. The size of the largest Jordan block corresponding to an eigenvalue λ is the multiplicity of λ as a root of the minimal polynomial $m_\alpha(X)$.

5. The sizes of the Jordan blocks corresponding to a given eigenvalue are determined by a knowledge of the ranks of certain powers.

Let J_λ is the direct sum of Jordan blocks of any size all the blocks corresponding to the same eigenvalue λ . Then the smallest integer k_1 such that $(J_\lambda - \lambda I)^{k_1} = 0$ is the size of the largest block. The rank of $(J_\lambda - \lambda I)^{k_1-1}$ is the number of blocks of order k_1 , the rank of $(J_\lambda - \lambda I)^{k_1-2}$ is twice the number of blocks of order k_1 plus the number of blocks of size $K_1 - 1$, and so forth. The sequence of ranks of $(J_\lambda - \lambda I)^{k_1-i}$, $i = 0, 1, \dots, k_1 - 1$, recursively determines the orders of all the blocks in J_λ .

6. The sizes of all the Jordan blocks in a general Jordan matrix are determined by a knowledge of the ranks of certain powers. If λ is an eigenvalue of a Jordan matrix $J \in M_n(R)$, then only the Jordan blocks corresponding to λ will be annihilated when one forms $(J - \lambda I)$, $(J - \lambda I)^2$, \dots because the other blocks in $(J - \lambda I)$ all have nonzero diagonal entries. Eventually, the rank of $(J - \lambda I)^k$ will stop decreasing (one need not consider any $k > n$); the smallest value of k for which the rank of $(J - \lambda I)^k$ attains its minimum value is called the **index** of the eigenvalue λ . An analysis of the ranks of the sequence of powers of $(J - \lambda I)$ is sufficient to determine the sizes and number of Jordan blocks corresponding to λ . By doing this procedure successively for each eigenvalue λ of J , one determines the entire Jordan structure of J .

All the above observations also apply to any matrix that is similar to J .

An application to solving linear Differential equations

This is a very quick sketch, and the reader will have to work out details, or refer to standard texts to actually solve the equations. We have also not introduced the exponential of a matrix, which can be defined via the Jordan canonical form. (Please refer to M. Artin's Algebra for more details.)

Consider the linear differential equation with constant coefficients which involves a derivative higher than the first; for example

$$s'' + as' + bs = 0 \tag{5}$$

By introducing new variables we are able to reduce (5) to a first order system of two equations. Let $x_1 = s$ and $x_2 = x'_1 = s'$. Then (5) becomes equivalent to the system:

$$\begin{aligned} x'_1 &= x_2, \\ x'_2 &= -bx_1 - ax_2. \end{aligned} \tag{6}$$

Thus if $x(t) = (x_1(t), x_2(t))$ is a solution of (6), then $s(t) = x_1(t)$ is a solution of (5); if $s(t)$ is a solution of (5), then $x(t) = (s(t), s'(t))$ is a solution of (6).

This procedure of introducing new variables works very generally to reduce higher order equations to first order ones. Thus consider

$$s^{(n)} + a_1s^{(n-1)} + \dots + a_{n-1}s' + a_ns = 0 \tag{7}$$

Here s is a real function of t and $s^{(n)}$ is the n^{th} derivative of s , while a_1, \dots, a_n are constants.

In this case the new variables are $x_1 = s$, $x_2 = x'_1, \dots, x_n = x'_{n-1}$ and the equation (7) is equivalent to the system

$$\begin{aligned} x'_1 &= x_2, \\ x'_2 &= x_3, \\ &\vdots \\ x'_n &= -a_n x_1 - a_{n-1} x_2 - \dots - a_1 x_n. \end{aligned} \tag{8}$$

In vector notation (8) has the form $x' = C(f)x$, where $C(f)$ is the companion matrix of $f = X^n + a_1 X^{n-1} + \dots + a_{n-1} X + a_n$.

Let us now return to our first equation $s'' + as' + bs = 0$. Denote the roots of the polynomial equation $\lambda^2 + a\lambda + b = 0$ by λ_1, λ_2 . Suppose at first that these roots are real and distinct. Then (5) reduces to the equation of first order (6); one can find a diagonalizing system of coordinates (y_1, y_2) . Every solution of (6) for these coordinates is then $y_1(t) = K_1 e^{\lambda_1 t}$, $y_2(t) = K_2 e^{\lambda_2 t}$, with arbitrary constants K_1, K_2 . Thus $x_1(t)$ or $s(t)$ is a certain linear combination $s(t) = p_{11} K_1 e^{\lambda_1 t} + p_{12} K_2 e^{\lambda_2 t}$. We conclude that if λ_1, λ_2 are real and distinct then every solution of (5) is of the form $s(t) = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t}$ for some (real) constants C_1, C_2 . These constants can be found if initial values $s(t_0), s'(t_0)$ are given.

Next, suppose that $\lambda_1 = \lambda_2 = \lambda$ and that these eigenvalues are real. In this case the 2×2 matrix in (6) is similar to a matrix of the Jordan form

$$\begin{pmatrix} \lambda & 0 \\ \beta & \lambda \end{pmatrix}, \quad \beta \neq 0.$$

In the new coordinates the equivalent first-order system is

$$\begin{aligned} y'_1 &= \lambda y_1 \\ y'_2 &= \beta y_1 + \lambda y_2. \end{aligned}$$

It can be shown that the general solution to such a system is

$$\begin{aligned} y_1(t) &= K_1 e^{\lambda t}, \\ y_2(t) &= K_1 \beta t e^{\lambda t} + K_2 e^{\lambda t}, \end{aligned}$$

K_1 and K_2 being arbitrary constants. In the original coordinates the solutions to the equivalent first order system are linear combinations of these. Thus we conclude that if the characteristic polynomial of (5) has only one root $\lambda \in \mathbb{R}$, the solution have the form

$$s(t) = C_1 e^{\lambda t} + C_2 t e^{\lambda t}.$$

The values of C_1 and C_2 can be determined from initial conditions.

Example. Solve the initial-value problem

$$s'' + 2s' + s = 0, \quad s(0) = 1, \quad s'(0) = 2. \quad (9)$$

The characteristic polynomial is $\lambda^2 + 2\lambda + 1$; the only root is $\lambda = -1$. Therefore the general solution is

$$s(t) = C_1 e^{-t} + C_2 t e^{-t}.$$

We find that

$$s'(t) = (-C_1 + C_2)e^{-t} - C_2 t e^{-t}.$$

From the initial conditions in (9) we get, setting $t = 0$ in the last two formulas

$$\begin{aligned} C_1 &= 1, \\ -C_1 + C_2 &= 2. \end{aligned}$$

Hence $C_2 = 3$ and the solution to (9) is

$$s(t) = e^{-t} + 3t e^{-t}.$$

The reader may verify that this actually is a solution to (9)!

The final case to consider is that when λ_1, λ_2 are non-real complex conjugate numbers. Suppose $\lambda_1 = u + iv$, $\lambda_2 = u - iv$. Then we get a solution

$$\begin{aligned} y_1(t) &= e^{ut}(K_1 \cos vt - K_2 \sin vt), \\ y_2(t) &= e^{ut}(K_1 \sin vt + K_2 \cos vt). \end{aligned}$$

Thus we obtain $s(t)$ as a linear combination of $y_1(t)$ and $y_2(t)$, so that finally,

$$s(t) = e^{ut}(C_1 \cos vt + C_2 \sin vt)$$

for some constants C_1, C_2 .

A special case of the last equation is the “harmonic oscillator”: $s'' + b^2 s = 0$; the eigenvalues are $\pm ib$, and the general solution is $C_1 \cos bt + C_2 \sin bt$. We summarize what we have found.

Theorem 8.12 *Let λ_1, λ_2 be the roots of the polynomial $\lambda^2 + a\lambda + b$. Then every solution of the differential equation*

$$s'' + as' + bs = 0$$

is of the following type:

Case (a). λ_1, λ_2 are real distinct: $s(t) = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t}$;

Case (b). $\lambda_1 = \lambda_2 = \lambda$ is real: $s(t) = C_1 e^{\lambda t} + C_2 t e^{\lambda t}$;

Case (c). $\lambda_1 = \bar{\lambda}_2 = u + iv$, $v \neq 0$: $s(t) = e^{ut}(C_1 \cos vt + C_2 \sin vt)$.

In each case C_1, C_2 are (real) constants determined by initial conditions of the form $s(t_0) = \alpha$, $s'(t_0) = \beta$.

Epilogue.

The three canonical forms - Smith, rational normal, Jordan - may be regarded as a complete answer to the question of similarity over a field. The general problem is still unsolved over a P.I.D. In the case of \mathbb{Z} , the problem has been reduced to the determination of the ideal classes in certain rings of algebraic integers. Thus, one can now say that there are only finitely many similarity classes of matrices α of $M_n(\mathbb{Z})$, such that $f(\alpha) = 0$, where $f(X)$ is a monic polynomial of degree n with integral coefficients which is irreducible over \mathbb{Q} . The number of similarity classes is equal to the class number of the ring $\mathbb{Z}[\theta]$, where θ is any root of $f(X)$. A good exposition can be found in Olga Taussky's paper, On a theorem of Latimer and MacDuffee, Canadian J. Math. **1**, 300-302 (1949).

Exercises

1. List the six distinct isomorphism classes of abelian groups of order 1176.
2. If G is an abelian group of order n , and $m|n$, then show that G has a subgroup H of order m .
3. Let G be an abelian group of order $p^k m$, where $p \nmid m$. Show that $G \cong H \oplus K$, where H, K are subgroups of G and $|H| = p^k$.
4. We say G is of type $(1, 2, 4, 27, 27, 25)$ if

$$G \cong \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/4\mathbb{Z} \oplus \mathbb{Z}/27\mathbb{Z} \oplus \mathbb{Z}/27\mathbb{Z} \oplus \mathbb{Z}/25\mathbb{Z}.$$

Find the type of $(\mathbb{Z}/464600\mathbb{Z})^*$. (Ans: $(2, 2, 2, 2^2, 2^2, 5, 5^2, 11)$).

5. Let G be a finite abelian group and let a be an element of maximal order. Show that for any element b in G , $o(b)|o(a)$.
6. Prove that an abelian group of order 2^n , $n \geq 1$, has an odd number of elements of order 2.
7. Suppose that G is an abelian group of order 120 which has exactly 3 elements of order 2. Determine G .
8. Let G be a finite non-cyclic abelian group. Show that $\text{Aut}(G)$ is non abelian.
9. Characterize those integers n for which
 - (a) the only abelian groups of order n are cyclic.
 - (b) there are precisely 4 non-isomorphic abelian groups of order n .
10. Let G be an abelian group of order p^m . Let a be an element of maximal order in G . Show that $G \cong \langle a \rangle \oplus K$, for some subgroup K of G .
11. Let H be a subgroup of a finite abelian group G . Show that G has a subgroup that is isomorphic to G/H .
12. If the elementary divisors of a finite abelian group is known then calculate the invariant factors of the group. Show that if the cyclic decomposition of a finite abelian group G is known, then its elementary divisors and their multiplicity can be determined.

13. Let the abelian group G be the direct sum of n infinite cyclic subgroups. If y_1, \dots, y_n are a set of generators of G , then show that $G \cong \bigoplus_{i=1}^n \langle y_i \rangle$.
14. Let $G = \mathbb{Z}_{p^2} \oplus \mathbb{Z}_p$. Let H be the subgroup of G generated by $\bar{p} + \bar{1}$. Show that there is no basis $\{x_1, x_2\}$ of G such that $m_1x_1 = \bar{p} + \bar{1}$, $m_2x_2 = 0$.
15. Dirichlet's theorem says that, for every pair of relatively prime integers a and b , there are infinitely many primes of the form $at + b$. Use Dirichlet's theorem to prove that every finite abelian group is isomorphic to a subgroup of $U(n) = \text{units of } \mathbb{Z}_n^*$, for some n .
16. Prove that a non-zero row of n elements of a principal ideal domain R having g.c.d 1 is the first row of some matrix in $\text{Gl}_n(R)$. Deduce Lemma 3.1.
17. If the submodules N_i , $i \in I$, of the R -module M have pairwise relatively prime order ideals p_i , $i \in I$, show that the sum $\sum_i N_i$ is direct.
18. If the R -module M is the sum of the cyclic submodules N_1, \dots, N_r having pairwise relatively prime order ideals I_1, \dots, I_r , show that M is cyclic and has order ideal $I = I_1 \dots I_r$.
19. If the finitely generated module M over the principal ideal domain R has invariant factors d_1, \dots, d_r , show that the invariant factors of every non-zero submodule of M are certain divisors d'_1, \dots, d'_r of d_1, \dots, d_r for some $t \leq r$. Prove a similar result for non-zero quotient modules of M assuming in addition that M is a torsion module.
20. Let $M = N \oplus P$ be a finitely generated module over the principal ideal domain R . If N, P have elementary divisors p^2, p^3 and p^3, p^4, q^5, q^6 respectively, determine the invariant factors of M .
21. Let $K = \mathbb{Z}/11\mathbb{Z}$, $R = K[X]$, and M the direct sum of cyclic R -modules whose order ideals are generated respectively by $(X+1)^2(X^2+2)^3$, $(X^3-2X)^2$, $(X^3+X)^2(X^2+2)^2$ and $(X^2+1)^2(X^2-1)^3$. Determine the elementary divisors and invariant factors of M .
22. Let $\alpha, \beta \in M_n(R)$, where R is a P.I.D. Suppose that $\alpha\beta = \mu I_n$, for some $\mu \in R$. Show that the invariant factors of α and β are divisors of μ .
23. Give an example of
 - (i) a torsion module having order ideal zero;
 - (ii) a free module which is not torsion free;
 - (iii) a torsion-free module over a P.I.D which is not free;
 - (iv) a free module which has a non-free submodule.
24. If $J \subset I$ are proper ideals of the commutative ring R , show that the R -module $R/I \oplus R/J$ is not cyclic.
25. If M, N, P are finitely generated modules over the principal ideal domain R such that $M \oplus N \simeq M \oplus P$, show that $N \simeq P$.
26. Let M be a torsion module of exponent p^r , for some prime element p , i.e. $p^r M = 0$, but $p^{r-1} M \neq 0$. Let $m \in M$ be an element of order p^r . Let $\bar{M} = M/Rm$. Let $\bar{m}_1, \dots, \bar{m}_n$ be independent elements of \bar{M} . Then for each i there exists a representation m_i of \bar{m}_i , such that the order of m_i is precisely the order of \bar{m}_i . In this case, the elements m, m_1, \dots, m_n are independent, i.e. $am + \sum_i a_i m_i = 0$ implies each $a_i m_i = 0 = am$.

Definition 8.13 Let M be a finitely generated module with $p^n M = 0$, for some prime $p \in R$. The **socle of M** is

$$\text{Soc}(M) = \{m \in M \mid pm = 0\}.$$

27. Show that $\text{Soc}(M)$ is a finitely generated submodule of M , which is a finite dimensional vector space over $R/(p)$. If $M = M_1 \oplus M_2$, then show that $\text{Soc}(M) \simeq \text{Soc}(M_1) \oplus \text{Soc}(M_2)$.
28. Let M be of exponent p^k , and $m \in M$ be of order p^k . Then $\dim \text{Soc}(M/Rm) < \dim \text{Soc}(M)$. Deduce, by induction on $\dim \text{Soc}(M)$, the Structure theorem for finitely generated p^n -torsion modules M over a P.I.D.
29. Let K be a field, $R = K[X]$, and F a free R -module with basis u_i , $1 \leq i \leq 4$; let G be the submodule of F generated by

$$\begin{aligned} v_1 &= u_1, \\ v_2 &= Xu_1 + (X^2 + X)u_2 + (X^2 + 2X)u_3 \\ v_3 &= (X^2 + X)u_1 + (X^2 + 2X)u_2 + (X^2 + 3X)u_3 \end{aligned}$$

Determine the rank and the invariant factors of F/G .

30. Let M be a finitely generated torsion module over a P.I.D. R . Let k be the exponent of M . Show that M has an element m of order k . Further show that Rm is a direct summand of M .
31. Let M be a module over a commutative ring R with 1. Suppose that M is a direct sum of n cyclic modules R/I_i ($1 \leq i \leq n$), where the I_i are ideals of R . Then for each integer $p > 0$, the R -module $\wedge^p M$ is isomorphic to the direct sum of the modules R/I_K , where for each p -element subset $K = \{k_1, \dots, k_p\}$ of $[1, n]$, the ideal I_K is $\sum_{j=1}^p I_{k_j}$.
32. (Characterization of the invariant factors)
Let M be a module over a commutative ring R with 1. Suppose that M is a direct sum of n cyclic modules R/I_i ($1 \leq i \leq n$), where the I_i are ideals of R . Assume further that $I_1 \subset I_2 \subset \dots \subset I_n$. Then, for $1 \leq p \leq n$, the ideal $I_p = \text{ann}(\wedge^p M)$. If $I_n \neq R$ then $\wedge^p M \neq 0$ for $1 \leq p \leq n$ and $\wedge^k M = (0 : Rm)$ for $k > n$.
33. Let A be an $m \times n$ matrix of rank m over a principal ideal domain R . If $m < n$, show that A can be completed to an $n \times n$ matrix over R having determinant $\Delta_m(A)$.
34. If A is an $m \times n$ matrix over the principal ideal domain R , $m \geq n$, show that there exists $Q \in \text{Gl}_m(R)$ such that QA has the form $\begin{pmatrix} \Delta \\ 0 \end{pmatrix}$ with Δ an $n \times n$ lower triangular matrix.
35. Let R be a principal ideal domain. Prove that every left ideal Λ in the ring $M_n(R)$ is principal. (Hint: Λ is generated by A_1, \dots, A_r and $Q(A_1^t \ \dots \ A_r^t)^t = (H^t \ 0^t)^t$, then $\Lambda = M_n(R)H$.)

36. Find the minimal polynomials and the three canonical forms over \mathbb{C} of the matrices

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}, \begin{pmatrix} 3 & 1 & -1 \\ 0 & 2 & 0 \\ 1 & 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

37. An endomorphism φ of a vector space V of dimension n is nilpotent if and only if its characteristic polynomial is X^n .
38. If φ is a diagonalizable endomorphism of a finite dimensional vector space V , then show that its eigen spaces are the primary component of V_φ .
39. Let V be a vector space of dimension n over a field K . Let φ be an endomorphism of V . Let its characteristic polynomial be $C(\varphi) = \prod_{i=1}^n (X - a_i)$, with $a_i \in \overline{K}$. Let $p(X) \in K[X]$. Show that the characteristic polynomial of $p(\varphi) = \prod_{i=1}^n (X - p(a_i))$. Deduce that
- (i) $p(\varphi)$ is invertible if and only if $\text{g.c.d}(p(\varphi), C(\varphi)) = 1$ if and only if $\text{g.c.d}(p(\varphi), m(\varphi)) = 1$, where $m(\varphi)$ is the minimal polynomial of φ .
 - (ii) Let characteristic $K = 0$. Show that φ is nilpotent if and only if $\text{Tr}(\varphi^s) = 0$ for $1 \leq s \leq n$. (Use Newton's identities.)
40. Let V be a vector space of dimension n over a field K . Let φ be an endomorphism of V . Show that the following statements are equivalent.
- (i) φ is diagonalizable.
 - (ii) V is a direct sum of eigenspaces of φ .
 - (iii) All the roots of the minimal polynomial of φ are in K , and these roots are all simple.

Moreover. prove that if these conditions are satisfied then every subspace of V closed under φ is the direct sum of its intersection with the eigenspaces of φ .

41. Show that the sum and composite of two commuting diagonalizable endomorphism of a finite dimensional vector space V , are diagonalizable.
42. An endomorphism φ of a finite dimensional vector space V over a field K is said to be **semi-simple** if every subspace W of V which is closed under φ has a complement W' (i.e. $V = W \oplus W'$) which is closed under φ . (Equivalently, V_φ is a semi-simple $K[X]$ -module.)

Show that φ is semi-simple if and only if the minimal polynomial of φ has no multiple factors. Deduce

- (1) If L is an extension of K , $\varphi \otimes_K L$ is semi-simple, then φ is semi-simple.
- (2) If L is separable over K , and φ is semi-simple, then $\varphi \otimes_K L$ is semi-simple.

43. Let V be a finite dimensional vector space over a field K , and φ be an endomorphism of V . Let $m(X)$ be its minimal polynomial. Then the following are equivalent:

- (i) For every extension L of K , $\varphi \otimes_K L$ is semi-simple.
- (ii) There exists an extension L of K such that $\varphi \otimes_K L$ is diagonalizable.
- (iii) The polynomial $m(X)$ is separable over K .

(An endomorphism φ satisfying (i) - (iii) is said to be **absolutely semi-simple**.)

Deduce that a necessary and sufficient condition for φ to be absolutely semi-simple is that there exists an extension L of K such that L is perfect and $\varphi \otimes_K L$ is semi-simple.

44. Let V be an $F[X]$ -module, and let $B = (v_1, \dots, v_n)$ be a basis of V , as F -vector space. Let B be the matrix of T with respect to this basis. Prove that $A = XI - B$ is a presentation matrix for the module.
45. If $2a \neq 0$, prove that the Jordan form of the square of the Jordan matrix $J_e(a)$ is the Jordan matrix $J_e(a^2)$.
46. Prove that the following conditions on α are equivalent:

- (a) the R -module V is cyclic;
- (b) $m(X) = \Delta(X)$;
- (c) all invariant factors of $XI_n - \alpha$, except the last, are 1;
- (d) α is similar to the companion matrix of some monic polynomial of degree n ;
- (e) every endomorphism of the vector space E which commutes with α is a polynomial in α with coefficients from K .

47. Suppose that the characteristic matrix of α over the field K has two invariant factors $\neq 1$, viz. $(X - 3)(X^2 + 1)$, $(X - 3)(X^2 + 1)^2$. Find the rational form and Jordan canonical form of α .
48. Find the Jordan normal form and the matrix of transformation, for each of the following:

$$\begin{pmatrix} -2 & 0 \\ -14 & 5 \end{pmatrix}, \begin{pmatrix} 12 & -7 \\ 14 & -9 \end{pmatrix}, \begin{pmatrix} 4 & 2 & -4 \\ 2 & 3 & -3 \\ 3 & 2 & -3 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 1 & 1 & 1 \end{pmatrix}$$

49. Find all possible Jordan forms for 8×8 matrices whose minimal polynomial is $x^2(x - 1)^3$.
50. Prove that the ranks of $(A - \alpha I)^k$ distinguish all Jordan forms, and hence that the Jordan form depends only on the operator and not on the basis.
51. Let V be a complex vector space of dimension 5, and let T be a linear operator on V which has characteristic polynomial $(X - \alpha)^5$. Suppose that the rank of the operator $T - \alpha I$ is 2. What are the possible Jordan forms for T ?
52. Find all possible Jordan forms for a matrix whose characteristic polynomial is $(X + 2)^2(X - 5)^3$.

53. What is the Jordan form of a matrix whose characteristic polynomial is $(X - 2)^2(X - 5)^3$ and such that the space of eigenvectors with eigenvalue 2 is one-dimensional, while the space of eigenvectors with eigenvalue 5 is two-dimensional?
54. (a) Prove that a Jordan block has a one-dimensional space of eigenvectors.
 (b) Prove that, conversely, if the eigenvectors of a complex matrix A are multiples of a single vector, then the Jordan form for A consists of one block.
55. Determine all invariant subspaces of a linear operator whose Jordan form consists of one block.
56. Solve the differential equation $dX/dt = AX$, when A is

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \begin{pmatrix} 0 & -1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 1 & 1 & 1 \end{pmatrix}$$

57. Let α be a $n \times n$ matrix having n distinct eigenvalues $\lambda_1, \dots, \lambda_n$. Explicitly construct an invertible matrix β such that $\beta\alpha\beta^{-1} = \text{diag}\{\lambda_1, \dots, \lambda_n\}$. (Hint: Take $\beta = (f_1, \dots, f_n)$, where f_j is a basis of eigenvectors of α .)
58. Let $\alpha \in M_n(\mathbb{R})$ have n distinct, real eigenvalues. Then for all $x_0 \in \mathbb{R}^n$, the linear differential equation $x' = \alpha x$; $x(0) = x_0$, has a unique solution. (Hint: Use the Smith normal form, and consider the case when α is a diagonal matrix.)
59. Find the general solution to the equations $x'_1 = x_1$, $x'_2 = x_1 + 2x_2$, $x'_3 = x_1 - x_3$.
60. In each case, solve the differential equation $dX/dt = AX$ when A is the Jordan block given.

$$(a) \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix} \quad (b) \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \quad (c) \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

61. Under what conditions on the constants a, b is there a non-trivial solution to $s'' + as + b = 0$ such that the equation $s(t) = 0$ has
- (a) no solution;
 (b) a positive finite number of solutions;
 (c) infinitely many solutions?
62. Find all solutions to $s''' - s'' + 4s' - 4s = 0$.
63. Let $q(t)$ be a polynomial of degree m . Show that any equation $s^{(n)} + a_1s^{(n-1)} + \dots + a_ns = q(t)$ has a solution which is a polynomial of degree $\leq m$.
64. Prove that two 3×3 matrices over K are similar if and only if they have the same minimal and the same characteristic polynomial. Show that such an assertion is false for 4×4 matrices over K .
65. Let T be a linear operator whose matrix is $\begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$. Is the corresponding $\mathbb{C}[X]$ -module cyclic?

66. Let $R = F[X]$ be the polynomial ring in one variable over a field F , and let V be the R -module generated by an element v which satisfies the relation $(X^3 + 3X + 2)v = 0$. Choose a basis for V as F -vector space, and find the matrix of the operator multiplication by t with respect to this basis.
67. Distribute into similarity classes the following 3×3 matrices over \mathbb{Q} :

$$\begin{pmatrix} 1 & 2 & 1 \\ 3 & 1 & 1 \\ 4 & 3 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 3 & 4 \\ 2 & 1 & 3 \\ 1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} -8 & 2 & 1 \\ -13 & 3 & 2 \\ -55 & 17 & 7 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 2 \\ 1 & 0 & -5 \\ 0 & 1 & 4 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

68. If the matrices A, B in $M_n(K)$ are similar in $M_n(L)$, where L is an overfield of K , show that they are already similar in $M_n(K)$.
69. Prove that every matrix in $M_n(K)$ is similar to its transpose, where K is a field.
70. Prove that the matrices $I_2 \perp T, T \perp T$, where $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, of $M_4(K)$ have the same characteristic and the same minimal polynomial without being similar.
71. Prove or disprove: A complex matrix α such that $\alpha^2 = \alpha$ is diagonalizable.
72. A complex $n \times n$ matrix α such that $\alpha^k = I_n$ for some k has diagonal Jordan form.
73. Show that the following concepts are equivalent:

- (i) R -module, where $R = \mathbb{Z}[i]$;
- (ii) abelian group V , together with a homomorphism $\varphi : V \rightarrow V$ such that $\varphi \circ \varphi = -$ identity.

74. (Real Jordan form of $\alpha \in M_n(\mathbb{R})$) Let $\alpha \in M_n(\mathbb{R})$. Show that there is a basis of \mathbb{R}^n corresponds to α has a diagonal blocks of the form $J_{e_1}(\lambda)$, $\lambda \in \mathbb{R}, e_1 > 0$; $D = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$, $a, b \in \mathbb{R}$ or a block matrix with blocks of D type on the diagonal, and blocks I_2 on the lower diagonal.
75. Find the real canonical form of $\alpha \in M_r(\mathbb{R})$, where $e_1\alpha = \begin{pmatrix} 0 & 0 & 0 & -8 \end{pmatrix}$, $e_2\alpha = \begin{pmatrix} 1 & 0 & 0 & 16 \end{pmatrix}$, $e_3\alpha = \begin{pmatrix} 0 & 1 & 0 & -14 \end{pmatrix}$, $e_4\alpha = \begin{pmatrix} 0 & 0 & 1 & 6 \end{pmatrix}$.
76. Let T be a linear operator whose matrix is $\begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$. Is the corresponding $\mathbb{C}[X]$ -module cyclic?
77. Let α be a square matrix of order n over a field K , with minimal polynomial m , and characteristic polynomial c . Show that $m|c|m^h|m^n$, where h is the number of invariant factors of $XI_n - \alpha$.
Moreover, show that $m = c/\Delta_{n-1}$, where Δ_{n-1} is the H.C.F. of all the $(n-1)$ -th order minors of $XI - \alpha$.
78. Do the Jordan forms $J_n(0)$ have a square root?
79. If α is an invertible complex matrix, then show that α has a square root.
80. Let $f(X) = \prod_{i=1}^r p_i(X)^{n_i}$, where $n_i \geq 1$ and $p_i(X)$ are distinct primes in $k[X]$, k a field, and with $r \geq 1$. Let $P(m)$ denote the number of partitions of $m \in \mathbb{N}$. Prove that the number of matrices upto similarity, with characteristic polynomial $f(X)$, is equal to $\prod_{i=1}^r P(n_i)$.

81. Let A be a ring (possibly non-commutative). Given $a, b \in A$ there is a unit u in A such that $ua = bu$ if and only if there are units f, g in $A[X]$ such that $f(X - a) = (X - b)g$. (Hint: Subtract $(X - b)h(X - a)$ from both sides where h is chosen so that $f - (X - b)h$ has degree zero.