Abstract Algebra II

George Voutsadakis[1]

[1] Mathematics and Computer Science
Lake Superior State University

LSSU Math 342

1 Modules Over Principal Ideal Domains

- Preview
- The Basic Theory
- The Rational Canonical Form
- The Jordan Canonical Form

Subsection 1

Preview

The Structure Theorem and its Applications

- We will prove a Structure Theorem for finitely generated modules over Principal Ideal Domains.
- This theorem is an example of the ideal structure of the ring (which is particularly simple for P.I.D.s) being reflected in the structure of its modules.
    - If we apply this result in the case where the P.I.D. is the ring of integers $\mathbb{Z}$, then we obtain a proof of the Fundamental Theorem of Finitely Generated Abelian Groups.
    - If we apply this structure theorem in the case where the P.I.D. is the ring $F[x]$ of polynomials in x with coefficients in a field F, we obtain the basic results on:
        - the Rational Canonical Form of a matrix;
        - the Jordan Canonical Form of a matrix.

Application on Finitely Generated Abelian Groups

- We saw that any finitely generated abelian group is isomorphic to the direct sum of cyclic abelian groups, either $\mathbb{Z}$ or $\mathbb{Z}/n\mathbb{Z}$, for some positive integer $n \neq 0$.
- An abelian group is the same thing as a $\mathbb{Z}$-module.

  Since the ideals of $\mathbb{Z}$ are precisely the trivial ideal (0) and the principal ideals $(n) = n\mathbb{Z}$ generated by positive integers n, the Fundamental Theorem of Finitely Generated Abelian Groups in the language of modules says that:
  - Any finitely generated $\mathbb{Z}$-module is the direct sum of modules of the form $\mathbb{Z}/I$, where I is an ideal of $\mathbb{Z}$ (these are the cyclic $\mathbb{Z}$-modules);
  - The direct sum is unique when written in a particular form.
- Note the correspondence between the ideal structure of $\mathbb{Z}$ and the structure of its (finitely generated) modules, the finitely generated abelian groups.

Application on Vector Spaces and Linear Transformations

- The Fundamental Theorem of Finitely Generated Modules over a P.I.D. states that the same result holds when the Principal Ideal Domain $\mathbb{Z}$ is replaced by any P.I.D.
- In particular, we have seen that a module over the ring $F[x]$ of polynomials in x with coefficients in the field F is the same thing as a vector space V together with a fixed linear transformation T of V (where the element x acts on V by the linear transformation T).
- The Fundamental Theorem in this case will say that:
    - Such a vector space is the direct sum of modules of the form $F[x]/I$, where I is an ideal of $F[x]$, hence is either the trivial ideal (0) or a principal ideal $(f(x))$ generated by some nonzero polynomial $f(x)$ (these are the cyclic $F[x]$-modules);
    - The direct sum is unique when written in a particular form.
- Translated back into the language of vector spaces and linear transformations, it yields information on the linear transformation T.

Jordan and Rational Canonical Forms of a Matrix

- For example, suppose V is a vector space of dimension n over F and we choose a basis for V.
- Giving a linear transformation T of V to itself is the same thing as giving an $n \times n$ matrix A, with coefficients in F (choosing a different basis for V gives a different matrix B for T which is similar to A, i.e., is of the form $P^{-1}AP$, for some invertible matrix P which defines the change of basis).
- We will see that the Fundamental Theorem in this situation implies (under the assumption that the field F contains all the "eigenvalues" for the given linear transformation T) that:

  > There is a basis for V so that the associated matrix for T is as close to being a diagonal matrix as possible and so has a particularly simple form.

  This is the Jordan canonical form.
- The rational canonical form is another simple form for the matrix for T (that does not require the eigenvalues for T to be elements of F).

Example

- Let $V = \mathbb{Q}^3 = \{(x, y, z) : x, y, z \in \mathbb{Q}\}$ be the 3-dimensional vector space of ordered 3-tuples with entries from the field $F = \mathbb{Q}$ of rational numbers.

  Suppose T is the linear transformation

  $$T(x, y, z) = (9x + 4y + 5z, -4x - 3z, -6x - 4y - 2z), x, y, z \in \mathbb{Q}.$$

  If we take the standard basis $e_1 = (1, 0, 0), e_2 = (0, 1, 0), e_3 = (0, 0, 1)$ for V then the matrix A representing this linear transformation is

  $$A = \begin{pmatrix} 9 & 4 & 5 \\ -4 & 0 & -3 \\ -6 & -4 & -2 \end{pmatrix}.$$

Example (Cont'd)

- We will see that the Jordan canonical form for this matrix A is the much simpler matrix $B = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$ obtained by taking instead the basis $f_1 = (2, -1, -2), f_2 = (1, 0, -1), f_3 = (3, -2, -2)$ for V, since in this case

$$\begin{array}{rcl} T(f_1) & = & T(2, -1, -2) = (4, -2, -4) = 2f_1 + 0f_2 + 0f_3; \\ T(f_2) & = & T(1, 0, -1) = (4, -1, -4) = 1f_1 + 2f_2 + 0f_3; \\ T(f_3) & = & T(3, -2, -2) = (9, -6, -6) = 0f_1 + 0f_2 + 3f_3. \end{array}$$

So the columns of the matrix representing T with respect to this basis are $(2, 0, 0), (1, 2, 0)$ and $(0, 0, 3)$. Hence, T has matrix B with respect to this basis. In particular A is similar to the simpler matrix B. This linear transformation T cannot be diagonalized (i.e., there is no choice of basis for V for which the corresponding matrix is a diagonal matrix). So B is as close to a diagonal matrix for T as is possible.

Subsection 2

The Basic Theory

Noetherian Modules and Noetherian Rings

- Let R be a ring and let M be a left R-module.

Definition (Noetherian Modules and Noetherian Rings)

(1) The left R-module M is said to be a **Noetherian R-module** or to satisfy the **ascending chain condition on submodules** (or **A.C.C. on submodules**) if there are no infinite increasing chains of submodules, i.e., whenever $M_1 \subseteq M_2 \subseteq M_3 \subseteq \cdots$ is an increasing chain of submodules of M, then there is a positive integer m, such that, for all $k \geq m$, $M_k = M_m$ (so the chain becomes stationary at stage m: $M_m = M_{m+1} = M_{m+2} = \cdots$).

(2) The ring R is said to be **Noetherian** if it is Noetherian as a left module over itself, i.e., if there are no infinite increasing chains of left ideals in R.

- One can formulate analogous notions of A.C.C. on right and on two-sided ideals in a (possibly noncommutative) ring R.

  For noncommutative rings these properties need not be related.

Characterizing Noetherian Modules

Theorem

Let R be a ring and let M be a left R-module. Then the following are equivalent:

(1) M is a Noetherian R-module.

(2) Every nonempty set of submodules of M contains a maximal element under inclusion.

(3) Every submodule of M is finitely generated.

(1) implies (2): Assume M is Noetherian and let Σ be any nonempty collection of submodules of M. Choose any $M_1 \in \Sigma$. If M_1 is a maximal element of Σ, (2) holds. If M_1 is not maximal, there is some $M_2 \in \Sigma$, such that $M_1 \subsetneq M_2$. If M_2 is maximal in Σ, (2) holds. Otherwise, there is an $M_3 \in \Sigma$, properly containing M_2. Proceeding in this way, if (2) fails, we can produce by the Axiom of Choice an infinite strictly increasing chain of elements of Σ, contrary to (1).

Characterizing Noetherian Modules (Cont'd)

(2) implies (3): Assume (2) holds and let N be any submodule of M. Let Σ be the collection of all finitely generated submodules of N. Since $\{0\} \in \Sigma$, this collection is nonempty. By (2), Σ contains a maximal element N'. If $N' \neq N$, let $x \in N - N'$. Since $N' \in \Sigma$, the submodule N' is finitely generated by assumption. Hence, the submodule generated by N' and x is also finitely generated. This contradicts the maximality of N'. So $N = N'$ is finitely generated.

(3) implies (1): Assume (3) holds and let $M_1 \subseteq M_2 \subseteq M_3 \subseteq \cdots$ be a chain of submodules of M. Let $N = \bigcup_{i=1}^{\infty} M_i$. Note that N is a submodule. By (3), N is finitely generated by, say, $a_1, a_2, \ldots, a_n$. Since $a_i \in N$, for all i, each a_i lies in one of the submodules in the chain, say M_{j_i}. Let $m = \max\{j_1, j_2, \ldots, j_n\}$. Then $a_i \in M_m$, for all i. So the module they generate is contained in M_m, i.e., $N \subseteq M_m$. This implies $M_m = N = M_k$, for all $k \geq m$, which proves (1).

P.I.D.s are Noetherian Rings

Corollary

If R is a P.I.D., then every nonempty set of ideals of R has a maximal element and R is a Noetherian ring.

- The P.I.D. R satisfies condition (3) in the theorem with $M = R$.
- Recall that even if M itself is a finitely generated R-module, submodules of M need not be finitely generated.
  So the condition that M be a Noetherian R-module is in general stronger than the condition that M be a finitely generated R-module.

Linear Dependence

Proposition

Let R be an integral domain and let M be a free R-module of rank $n < \infty$. Then any $n+1$ elements of M are R-linearly dependent, i.e., for any $y_1, y_2, \ldots, y_{n+1} \in M$, there are elements $r_1, r_2, \ldots, r_{n+1} \in R$, not all zero, such that $r_1 y_1 + r_2 y_2 + \cdots + r_{n+1} y_{n+1} = 0$.

- The quickest way of proving this is to:
  - Embed R in its quotient field F (since R is an integral domain);
  - Observe that since $M \cong R \oplus R \oplus \cdots \oplus R$ (n times) we obtain $M \subseteq F \oplus F \oplus \cdots \oplus F$;
  - The latter is an n-dimensional vector space over F;
  - So any $n+1$ elements of M are F-linearly dependent;
  - By clearing the denominators of the scalars (by multiplying through by the product of all the denominators, for example), we obtain an R-linear dependence relation among the $n+1$ elements of M.

Linear Dependence (Alternative Proof)

- Alternatively, let $e_1, \ldots, e_n$ be a basis of the free R-module M and let $y_1, \ldots, y_{n+1}$ be any $n+1$ elements of M. For $1 \leq i \leq n+1$, write

  $$y_i = a_{1i}e_1 + a_{2i}e_2 + \cdots + a_{ni}e_n$$

  in terms of the basis $e_1, e_2, \ldots, e_n$. Let A be the $(n+1) \times (n+1)$ matrix whose:

  - i, j entry is $a_{ij}, 1 \leq i \leq n, 1 \leq j \leq n+1$;
  - last row is zero.

  Certainly $\det A = 0$. Since R is an integral domain, the columns of A are R-linearly dependent. Any dependence relation on the columns of A gives a dependence relation on the y_i's.

Torsion Submodules

- If R is any integral domain and M is any R-module recall that

  $$\mathrm{Tor}(M) = \{x \in M : rx = 0, \text{ for some nonzero } r \in R\}$$

  is a submodule of M (called **the torsion submodule of** M).
- If N is any submodule of $\mathrm{Tor}(M)$, N is called a **torsion submodule of** M.
- So the torsion submodule of M is the union of all torsion submodules of M, i.e., is the maximal torsion submodule of M.
- If $\mathrm{Tor}(M) = 0$, the module M is said to be **torsion free**.

Annihilator Ideals

- For any submodule N of M, the **annihilator of** N is the ideal of R defined by
$$\mathrm{Ann}(N) = \{r \in R : rn = 0, \text{for all } n \in N\}.$$
- Note that if N is not a torsion submodule of M then $\mathrm{Ann}(N) = (0)$.
- It is easy to see that if N, L are submodules of M with $N \subseteq L$, then $\mathrm{Ann}(L) \subseteq \mathrm{Ann}(N)$.
- If R is a P.I.D. and $N \subseteq L \subseteq M$, with $\mathrm{Ann}(N) = (a)$ and $\mathrm{Ann}(L) = (b)$, then $a \mid b$.

  In particular, the annihilator of any element x of M divides the annihilator of M (this is implied by Lagrange's Theorem when $R = \mathbb{Z}$).

The Rank of an R-Module

Definition (Rank of a Module)

For any integral domain R the **rank** of an R-module M is the maximum number of R-linearly independent elements of M.

- The preceding proposition states that for a free R-module M over an integral domain the rank of a submodule is bounded by the rank of M.
- This notion of rank agrees with previous uses of the same term:
  - If the ring $R = F$ is a field, then the rank of an R-module M is the dimension of M as a vector space over F;

    Moreover, any maximal set of F-linearly independent elements is a basis for M.
  - For a general integral domain, however, an R-module M of rank n need not have a "basis", i.e., need not be a free R-module even if M is torsion free.

    So some care is necessary with the notion of rank, particularly with respect to the torsion elements of M.

Free Modules of Finite Rank over P.I.D.s

- If N is a submodule of a free module of finite rank over a P.I.D., then N is again a free module of finite rank; Furthermore it is possible to choose generators for the two modules related in a simple way.

Theorem

Let R be a Principal Ideal Domain, let M be a free R-module of finite rank n and let N be a submodule of M. Then:

(1) N is free of rank $m, m \leq n$;

(2) There exists a basis $y_1, y_2, \ldots, y_n$ of M so that $a_1 y_1, a_2 y_2, \ldots, a_m y_m$ is a basis of N, where $a_1, a_2, \ldots, a_m$ are nonzero elements of R with the divisibility relations $a_1 \mid a_2 \mid \cdots \mid a_m$.

- The theorem is trivial for $N = \{0\}$. Assume $N \neq \{0\}$. For each R-module homomorphism φ of M into R, the image $\varphi(N)$ of N is a submodule of R, i.e., an ideal in R. Since R is a P.I.D., this ideal must be principal, say $\varphi(N) = (a_\varphi)$, for some $a_\varphi \in R$.

Free Modules of Finite Rank over P.I.D.s (Cont'd)

- Let $\Sigma = \{(a_\varphi) : \varphi \in \mathrm{Hom}_R(M, R)\}$ be the collection of the principal ideals in R obtained in this way from the R-module homomorphisms of M into R. The collection Σ is certainly nonempty since taking φ to be the trivial homomorphism shows that $(0) \in \Sigma$. By a preceding corollary, Σ has at least one maximal element i.e., there is at least one homomorphism ν of M to R so that the principal ideal $\nu(N) = (a_\nu)$ is not properly contained in any other element of Σ. Let $a_1 = a_\nu$, for this maximal element. Let $y \in N$ be an element mapping to the generator a_1 under the homomorphism $\nu : \nu(y) = a_1$.

  Claim: The element a_1 is nonzero.

  Let $x_1, x_2, \ldots, x_n$ be any basis of the free module M. Let $\pi_i \in \mathrm{Hom}_R(M, R)$ be the natural projection homomorphism onto the i-th coordinate with respect to this basis. Since $N \neq \{0\}$, there exists an i such that $\pi_i(N) \neq 0$. This shows that Σ contains more than just the trivial ideal (0). Since (a_1) is a maximal element of Σ it follows that $a_1 \neq 0$.

Free Modules of Finite Rank over P.I.D.s (Cont'd)

- Claim: Element a_1 divides $\varphi(y)$, for every $\varphi \in \text{Hom}_R(M, R)$.

  Let d be a generator for the principal ideal generated by a_1 and $\varphi(y)$. Then d is a divisor of both a_1 and $\varphi(y)$ in R and $d = r_1 a_1 + r_2 \varphi(y)$, for some $r_1, r_2 \in R$. Consider the homomorphism $\psi = r_1 \nu + r_2 \varphi$ from M to R. Then $\psi(y) = (r_1 \nu + r_2 \varphi)(y) = r_1 a_1 + r_2 \varphi(y) = d$. So $d \in \psi(N)$. Hence, $(d) \subseteq \psi(N)$. But d is a divisor of a_1. So we also have $(a_1) \subseteq (d)$. Then $(a_1) \subseteq (d) \subseteq \psi(N)$. By the maximality of (a_1), we must have equality: $(a_1) = (d) = \psi(N)$. In particular, $(a_1) = (d)$ shows that $a_1 \mid \varphi(y)$ since d divides $\varphi(y)$.
- If we apply this to the projection homomorphisms π_i, we see that a_1 divides $\pi_i(y)$, for all i. Write $\pi_i(y) = a_1 b_i$, for some $b_i \in R$, $1 \leq i \leq n$. Define $y_1 = \sum_{i=1}^n b_i x_i$. Note that $a_1 y_1 = y$. But $a_1 = \nu(y) = \nu(a_1 y_1) = a_1 \nu(y_1)$ and a_1 is a nonzero element of the integral domain R. This shows $\nu(y_1) = 1$.

Free Modules of Finite Rank over P.I.D.s (Cont'd)

- Claim: The element y_1 can be taken as one element in a basis for M and $a_1 y_1$ can be taken as one element in a basis for N, i.e., we have:
  (a) $M = Ry_1 \oplus \ker\nu$;
  (b) $N = Ra_1y_1 \oplus (N \cap \ker\nu)$.

(a) Let x be arbitrary in M and write $x = \nu(x)y_1 + (x - \nu(x)y_1)$. We have $\nu(x - \nu(x)y_1) = \nu(x) - \nu(x)\nu(y_1) = \nu(x) - \nu(x) \cdot 1 = 0$. Hence, $x - \nu(x)y_1$ is in the kernel of ν. So x can be written as the sum of an element in Ry_1 and an element in $\ker\nu$. Hence, $M = Ry_1 + \ker\nu$. To see that the sum is direct, suppose ry_1 is also an element in the kernel of ν. Then $0 = \nu(ry_1) = r\nu(y_1) = r$. This shows that $ry_1 = 0$.

(b) Observe that $\nu(x')$ is divisible by a_1, for every $x' \in N$, by the definition of a_1 as a generator for $\nu(N)$. Write $\nu(x') = ba_1$ where $b \in R$, Then the decomposition in (a) is $x' = \nu(x')y_1 + (x' - \nu(x')y_1)$ $= ba_1y_1 + (x' - ba_1y_1)$, where the second summand is in $\ker\nu$ and in N. So $N = Ra_1y_1 + (N \cap \ker\nu)$. The fact that the sum in (b) is direct is a special case of the directness of the sum in (a).

Free Modules of Finite Rank over P.I.D.s: Part 1

- We now prove Part 1 by induction on the rank, m, of N.

  If $m = 0$, then N is a torsion module. Hence $N = 0$, since a free module is torsion free. So Part 1 holds trivially.

  Assume then that $m > 0$. Since the sum in (b) above is direct, we see easily that $N \cap \ker\nu$ has rank $m - 1$. By induction, $N \cap \ker\nu$ is then a free R-module of rank $m - 1$. Again by the directness of the sum in (b) we see that adjoining $a_1 y_1$ to any basis of $N \cap \ker\nu$ gives a basis of N. So N is also free (of rank m). This proves Part 1.

Free Modules of Finite Rank over P.I.D.s: Part 2

- We prove Part 2 by induction on n, the rank of M.
  Applying Part 1 to the submodule $\ker\nu$ shows that this submodule is free. Since the sum in (a) is direct it is free of rank $n-1$. By the induction assumption applied to the module $\ker\nu$ (which plays the role of M) and its submodule $\ker\nu\cap N$ (which plays the role of N):

  > There is a basis $y_2, y_3, \ldots, y_n$ of $\ker\nu$, such that $a_2y_2, a_3y_3, \ldots, a_my_m$ is a basis of $N\cap\ker\nu$, for some elements $a_2, a_3, \ldots, a_m$ of R, with $a_2 \mid a_3 \mid \cdots \mid a_m$.

  Since the sums (a) and (b) are direct, $y_1, y_2, \ldots, y_n$ is a basis of M and $a_1y_1, a_2y_2, \ldots, a_my_m$ is a basis of N. It now remains to show that a_1 divides a_2.

  Define a homomorphism φ from M to R by $\varphi(y_1)=\varphi(y_2)=1$ and $\varphi(y_i)=0$, for all $i>2$, on the basis for M. Then, we have $a_1=\varphi(a_1y_1)$. So $a_1\in\varphi(N)$. Hence, $(a_1)\subseteq\varphi(N)$. By the maximality of (a_1) in Σ, it follows that $(a_1)=\varphi(N)$. Since $a_2=\varphi(a_2y_2)\in\varphi(N)$, we then have $a_2\in(a_1)$, i.e., $a_1\mid a_2$.

Cyclic R-Modules

- Recall that the left R-module C is a **cyclic** R-module (for any ring R, not necessarily commutative nor with 1) if there is an element $x \in C$, such that $C = Rx$.
- We can then define an R-module homomorphism
$$\pi : R \to C$$
by $\pi(r) = rx$, which will be surjective by the assumption $C = Rx$.
- The First Isomorphism Theorem gives an isomorphism of (left) R-modules
$$R/\mathrm{ker}\pi \cong C.$$
- If R is a P.I.D., $\mathrm{ker}\pi$ is a principal ideal, (a). So the cyclic R-modules C are of the form $R/(a)$ where $(a) = \mathrm{Ann}(C)$.
- The cyclic modules are the simplest modules (since they require only one generator).

Fundamental Theorem, Existence: Invariant Factor Form

Theorem 5. (Fundamental Theorem, Existence: Invariant Factor Form)

Let R be a P.I.D. and let M be a finitely generated R-module.

(1) Then M is isomorphic to the direct sum of finitely many cyclic modules. More precisely,

$$M \cong R^r \oplus R/(a_1) \oplus R/(a_2) \oplus \cdots \oplus R/(a_m),$$

for some integer $r \geq 0$ and nonzero elements $a_1, a_2, \ldots, a_m$ of R which are not units in R and which satisfy the divisibility relations $a_1 \mid a_2 \mid \cdots \mid a_m$.

(2) M is torsion free if and only if M is free.

(3) In the decomposition in (1),

$$\text{Tor}(M) \cong R/(a_1) \oplus R/(a_2) \oplus \cdots \oplus R/(a_m).$$

In particular M is a torsion module if and only if $r = 0$ and in this case the annihilator of M is the ideal (a_m).

Proof of Fundamental Theorem

- The module M can be generated by a finite set of elements by assumption. Let $x_1, x_2, \ldots, x_n$ be a set of generators of M of minimal cardinality. Let R^n be the free R-module of rank n with basis $b_1, b_2, \ldots, b_n$. Define the homomorphism $\pi : R^n \to M$ by setting $\pi(b_i) = x_i$, for all i. π is automatically surjective since $x_1, \ldots, x_n$ generate M. By the First Isomorphism Theorem for modules we have $R^n/\text{ker}\pi \cong M$. Now, by the theorem, applied to R^n and the submodule $\text{ker}\pi$ we can choose another basis $y_1, y_2, \ldots, y_n$ of R^n so that $a_1 y_1, a_2 y_2, \ldots, a_m y_m$ is a basis of $\text{ker}\pi$ for some elements $a_1, a_2, \ldots, a_m$ of R with $a_1 \mid a_2 \mid \cdots \mid a_m$. This implies

$$\begin{array}{rcl} M & \cong & R^n/\text{ker}\pi \\ & = & (Ry_1 \oplus Ry_2 \oplus \cdots \oplus Ry_n)/(Ra_1 y_1 \oplus Ra_2 y_2 \oplus \cdots \oplus Ra_m y_m). \end{array}$$

Proof of Fundamental Theorem (Cont'd)

- To identify the quotient on the right hand side we use the natural surjective R-module homomorphism

$$Ry_1 \oplus Ry_2 \oplus \cdots \oplus Ry_n \to R/(a_1) \oplus R/(a_2) \oplus \cdots \oplus r/(a_m) \oplus R^{n-m}$$

that maps $(\alpha_1 y_1, \ldots, \alpha_n y_n)$ to $(\alpha_1 \mod (a_1), \ldots, \alpha_m \mod (a_m)$, $\alpha_{m+1}, \ldots, \alpha_n)$. The kernel of this map is the set of elements where a_i divides α_i, $i = 1, 2, \ldots, m$, i.e., $Ra_1y_1 \oplus Ra_2y_2 \oplus \cdots \oplus Ra_my_m$. Hence we obtain $M \cong R/(a_1) \oplus R/(a_2) \oplus \cdots \oplus R/(a_m) \oplus R^{n-m}$. If a is a unit in R, then $R/(a) = 0$. So in this direct sum we may remove any of the initial a_i which are units. This gives (1) (with $r = n - m$).

$R/(a)$ is a torsion R-module, for any nonzero element a of R. Hence, (1) immediately implies M is a torsion free module if and only if $M \cong R^r$. This is (2).

The annihilator of $R/(a)$ is the ideal (a). Hence, Part (3) is immediate from the definitions.

Idea for Uniqueness and Betti Number of M

- We will prove the uniqueness of the decomposition, namely that if we have
$$M \cong R^{r'} \oplus R/(b_1) \oplus R/(b_2) \oplus \cdots \oplus R/(b_{m'}),$$
for some integer $r' \geq 0$ and nonzero elements $b_1, b_2, \ldots, b_{m'}$ of R which are not units with $b_1 \mid b_2 \mid \cdots \mid b_{m'}$, then $r = r'$, $m = m'$ and $(a_i) = (b_i)$ (so $a_i = b_i$ up to units) for all i.
The divisibility condition $a_1 \mid a_2 \mid \cdots \mid a_m$ gives the uniqueness.

Definition (Betti Number and Invariant Factors)

The integer r in the theorem is called the **free rank** or the **Betti number** of M and the elements $a_1, a_2, \ldots, a_m \in R$ (defined up to multiplication by units in R) are called the **invariant factors** of M.

- Until we have proved that the invariant factors of M are unique, we refer to *a set of invariant factors* for M (and similarly for the free rank), by which we mean any elements giving a decomposition for M as in Part (1) of the theorem.

Fundamental Theorem, Existence: Elementary Divisors

Theorem (Fundamental Theorem, Existence: Elementary Divisor Form)

Let R be a P.I.D. and let M be a finitely generated R-module. Then M is the direct sum of a finite number of cyclic modules whose annihilators are either (0) or generated by powers of primes in R, i.e.,

$$M \cong R^r \oplus R/(p_1^{\alpha_1}) \oplus R/(p_2^{\alpha_2}) \oplus \cdots \oplus R/(p_t^{\alpha_t}),$$

where $r \geq 0$ is an integer and $p_1^{\alpha_1}, \ldots, p_t^{\alpha_t}$ are positive powers of (not necessarily distinct) primes in R.

- Suppose a is a nonzero element of the Principal Ideal Domain R. R is also a Unique Factorization Domain. So we can write

  $$a = up_1^{\alpha_1} p_2^{\alpha_2} \cdots a_s^{\alpha_s},$$

  where the p_i are distinct primes in R and u is a unit. This factorization is unique up to units. So the ideals $(p_i^{\alpha_i})$, $i = 1, \ldots, s$, are uniquely defined.

Fundamental Theorem: Elementary Divisors (Cont'd)

- For $i \neq j$, we have $(p_i^{\alpha_i}) + (p_j^{\alpha_j}) = R$, since the sum of these two ideals is generated by a greatest common divisor, which is 1 for distinct primes p_i, p_j. Put another way, the ideals $(p_i^{\alpha_i})$, $i = 1, \ldots, s$, are comaximal in pairs.

  The intersection of all these ideals is the ideal (a), since a is the least common multiple of $p_1^{\alpha_1}, p_2^{\alpha_2}, \ldots, p_s^{\alpha_s}$. Then the Chinese Remainder Theorem shows that

  $$R/(a) \cong R/(p_1^{\alpha_1}) \oplus R/(p_2^{\alpha_2}) \oplus \cdots \oplus R/(p_s^{\alpha_s})$$

  as rings and also as R-modules.

  Applying this to the modules in the theorem, allows us to write each of the direct summands $R/(a_i)$ for the invariant factor a_i of M as a direct sum of cyclic modules whose annihilators are the prime power divisors of a_i.

Elementary Divisors of M

- We proved the Elementary Divisor Form by using the prime power factors of the invariant factors for M.
- We will see that the decomposition of M into a direct sum of cyclic modules whose annihilators are (0) or prime powers is unique:

  The integer r and the ideals $(p_1^{\alpha_1}), \ldots, (p_t^{\alpha_t})$ are uniquely defined for M.

Definition (Elementary Divisors)

Let R be a P.I.D. and let M be a finitely generated R-module. The prime powers $p_1^{\alpha_1}, \ldots, p_t^{\alpha_t}$ (defined up to multiplication by units in R) are called the **elementary divisors** of M.

The Primary Decomposition Theorem

- Suppose M is a finitely generated torsion module over the Principal Ideal Domain R.
- If for the distinct primes $p_1, p_2, \ldots, p_n$ in the decomposition, we group together all the cyclic factors corresponding to the same p_i, M can be written as $M = N_1 \oplus N_2 \oplus \cdots \oplus N_n$, where N_i consists of all the elements of M which are annihilated by some power of the prime p_i.

Theorem (The Primary Decomposition Theorem)

Let R be a P.I.D. and let M be a nonzero torsion R-module (not necessarily finitely generated) with nonzero annihilator a. Suppose the factorization of a into distinct prime powers in R is $a = up_1^{\alpha_1} p_2^{\alpha_2} \cdots p_n^{\alpha_n}$ and let $N_i = \{x \in M : p_i^{\alpha_i} x = 0\}$, $1 \leq i \leq n$. Then N_i is a submodule of M with annihilator $p_i^{\alpha_i}$ and is the submodule of M of all elements annihilated by some power of p_i. We have $M = N_1 \oplus N_2 \oplus \cdots \oplus N_n$. If M is finitely generated, then each N_i is the direct sum of finitely many cyclic modules whose annihilators are divisors of $p_i^{\alpha_i}$.

The Proof of the Primary Decomposition Theorem

- We have proved the results in the case where M is finitely generated over R.

  In the general case it is clear that N_i is a submodule of M with annihilator dividing $p_i^{\alpha_i}$. Since R is a P.I.D., the ideals $(p_i^{\alpha_i})$ and $(p_j^{\alpha_j})$ are comaximal for $i \neq j$. So the direct sum decomposition of M can be proved easily by modifying the argument in the proof of the Chinese Remainder Theorem to apply it to modules. Using this decomposition we see that the annihilator of N_i is precisely $p_i^{\alpha_i}$.

Definition (Primary Components)

The submodule N_i in the previous theorem is called the p_i-**primary component** of M.

- Notice that with this terminology the elementary divisors of a finitely generated module M are just the invariant factors of the primary components of Tor(M).

Isomorphism Lemma

- Note that if M is any module over a commutative ring R and a is an element of R, then $aM = \{am : m \in M\}$ is a submodule of M.
- Recall also that in a Principal Ideal Domain R, the nonzero prime ideals are maximal.

  Hence, the quotient of R by a nonzero prime ideal is a field.

Lemma

Let R be a P.I.D. and let p be a prime in R. Let F denote the field $R/(p)$.

(1) Let $M = R^r$. Then $M/pM \cong F^r$.

(2) Let $M = R/(a)$ where a is a nonzero element of R. Then
$$M/pM \cong \begin{cases} F, & \text{if } p \text{ divides } a \text{ in } R \\ 0, & \text{if } p \text{ does not divide } a \text{ in } R \end{cases}$$

(3) Let $M = R/(a_1) \oplus R/(a_2) \oplus \cdots \oplus R/(a_k)$, where each a_i is divisible by p. Then $M/pM \cong F^k$.

Proof of the Lemma

(1) There is a natural map from R^r to $(R/(p))^r$ defined by mapping $(\alpha_1, \ldots, \alpha_r)$ to $(\alpha_1 \mod (p), \ldots, \alpha_r \mod (p))$. This is clearly a surjective R-module homomorphism. Its kernel consists of the r-tuples all of whose coordinates are divisible by p, i.e., pR^r. So $R^r/pR^r \cong (R/(p))^r$.

(2) This follows from the Isomorphism Theorems: Note first that $p(R/(a))$ is the image of the ideal (p) in the quotient $R/(a)$. Hence, it is $(p)+(a)/(a)$. The ideal $(p)+(a)$ is generated by a greatest common divisor of p and a. Hence, it is (p), if p divides a, and it is $R=(1)$, otherwise. Thus, $pM=(p)/(a)$, if p divides a and $pM=R/(a)=M$, otherwise. If p divides a, then $M/pM=(R/(a))/((p)/(a)) \cong R/(p)$. If p does not divide a, then $M/pM=M/M=0$.

(3) This follows from (2) as in the proof of Part (1) of the Invariant Factor Form of the Fundamental Theorem, Existence.

Fundamental Theorem, Uniqueness

Theorem (Fundamental Theorem, Uniqueness)

Let R be a P.I.D.

(1) Two finitely generated R-modules M_1 and M_2 are isomorphic if and only if they have the same free rank and the same list of invariant factors.

(2) Two finitely generated R-modules M_1 and M_2 are isomorphic if and only if they have the same free rank and the same list of elementary divisors.

- If M_1 and M_2 have the same free rank and list of invariant factors or the same free rank and list of elementary divisors, then they are clearly isomorphic.

  Suppose that M_1 and M_2 are isomorphic. Any isomorphism between M_1 and M_2 maps the torsion in M_1 to the torsion in M_2. So we must have $\text{Tor}(M_1) \cong \text{Tor}(M_2)$.

Fundamental Theorem, Uniqueness (Cont'd)

- Then, if r_1 is the free rank of M_1 and r_2 is the free rank of M_2,

$$R^{r_1} \cong M_1/\text{Tor}(M_1) \cong M_2/\text{Tor}(M_2) \cong R^{r_2}.$$

Let p be any nonzero prime in R. Then from $R^{r_1} \cong R^{r_2}$ we obtain $R^{r_1}/pR^{r_1} \cong R^{r_2}/pR^{r_2}$. By the previous lemma, this implies $F^{r_1} \cong F^{r_2}$, where F is the field R/pR. Hence, we have an isomorphism of an r_1-dimensional vector space over F with an r_2-dimensional vector space over F. It follows that $r_1 = r_2$. Thus, M_1 and M_2 have the same free rank.

Fundamental Theorem: List of Elementary Divisors

- We are reduced to showing that M_1 and M_2 have the same lists of invariant factors and elementary divisors.

  To do this we need only work with the isomorphic torsion modules $\mathrm{Tor}(M_1)$ and $\mathrm{Tor}(M_2)$. So, we may assume that both M_1 and M_2 are torsion R-modules.
- We first show they have the same elementary divisors.

  It suffices to show that for any fixed prime p the elementary divisors which are a power of p are the same for both M_1 and M_2.

  If $M_1 \cong M_2$, then the p-primary submodule of M_1 ($=$ the direct sum of the cyclic factors whose elementary divisors are powers of p) is isomorphic to the p-primary submodule of M_2, since these are the submodules of elements which are annihilated by some power of p. We are therefore reduced to the case of proving that if two modules M_1 and M_2 which have annihilator a power of p are isomorphic then they have the same elementary divisors.

Fundamental Theorem: List of Elementary Divisors (II)

- We proceed by induction on the power of p in the annihilator of M_1 (which is the same as the annihilator of M_2 since M_1 and M_2 are isomorphic).

  If this power is 0, then both M_1 and M_2 are 0 and we are done.

  Otherwise, M_1 (and M_2) have nontrivial elementary divisors. Suppose the elementary divisors of M_1 are given by $\underbrace{p, p, \ldots, p}_{m \text{ times}}$, p^{α_1}, p^{α_2}, ..., p^{α_s}, where $2 \leq \alpha_1 \leq \alpha_2 \leq \cdots \leq a_s$, i.e., M_1 is the direct sum of cyclic modules with generators $x_1, x_2, \ldots, x_m, x_{m+1}, \ldots, x_{m+s}$ say, whose annihilators are $(p), (p), \ldots, (p), (p^{\alpha_1}), \ldots, (p^{\alpha_s})$, respectively. Then the submodule pM_1 has elementary divisors $p^{\alpha_1 - 1}$, $p^{\alpha_2 - 1}$, ..., $p^{\alpha_s - 1}$ since pM_1 is the direct sum of the cyclic modules with generators $px_1, px_2, \ldots, px_m, px_{m+1}, \ldots, px_{m+s}$ whose annihilators are $(1), (1), \ldots, (1), (p^{\alpha_1 - 1}), \ldots, (p^{\alpha_s - 1})$, respectively.

Fundamental Theorem: List of Elementary Divisors (III)

- Similarly, if the elementary divisors of M_2 are given by $\underbrace{p, p, \ldots, p}_{n \text{ times}}$, $p^{\beta_1}, p^{\beta_2}, \ldots, p^{\beta_t}$, where $2 \leq \beta_1 \leq \beta_2 \leq \cdots \leq \beta_t$, then pM_2 has elementary divisors $p^{\beta_t - 1}, p^{\beta_2 - 1}, \ldots, p^{\beta_t - 1}$.

  Since $M_1 \cong M_2$, also $pM_1 \cong pM_2$. Moreover, the power of p in the annihilator of pM_1 is one less than the power of p in the annihilator of M_1. By induction, the elementary divisors for pM_1 are the same as the elementary divisors for pM_2. I.e., $s = t$ and $\alpha_i - 1 = \beta_i - 1$, for $i = 1, 2, \ldots, s$. Hence, $\alpha_i = \beta_i$, for $i = 1, 2, \ldots, s$. Finally, since also $M_1/pM_1 \cong M_2/pM_2$, we get, by the lemma, that $F^{m+s} \cong F^{n+t}$. This shows that $m + s = n + t$. Since $s = t$, we get $m = n$. This proves that the set of elementary divisors for M_1 is the same as the set of elementary divisors for M_2.

Fundamental Theorem: List of Invariant Factors

- We now show that M_1 and M_2 must have the same invariant factors.

  Suppose $a_1 \mid a_2 \mid \cdots \mid a_m$ are invariant factors for M_1. We obtain a set of elementary divisors for M_1 by taking the prime power factors of these elements. Note that then the divisibility relations on the invariant factors imply that:

  - a_m is the product of the largest of the prime powers among these elementary divisors;
  - a_{m-1} is the product of the largest prime powers among these elementary divisors once the factors for a_m have been removed;

    $\vdots$

  If $b_1 \mid b_2 \mid \cdots \mid b_n$ are invariant factors for M_2, then we similarly obtain a set of elementary divisors for M_2 by taking the prime power factors of these elements. But we showed above that the elementary divisors for M_1 and M_2 are the same. It follows that the same is true of the invariant factors.

Elementary Divisors and Invariant Factors

Corollary

Let R be a P.I.D. and let M be a finitely generated R-module.

(1) The elementary divisors of M are the prime power factors of the invariant factors of M.

(2) The largest invariant factor of M is the product of the largest of the distinct prime powers among the elementary divisors of M; the next largest invariant factor is the product of the largest of the distinct prime powers among the remaining elementary divisors of M; and so on.

- The procedure in (1) gives a set of elementary divisors.

  Since the elementary divisors for M are unique by the theorem, it follows that the procedure in (1) gives *the set of elementary divisors*.

  Similarly for (2).

Subsection 3

The Rational Canonical Form

Setup for Applying the Classification Theorems

- We now apply the results on finitely generated modules in the special case where the P.I.D. is the ring $F[x]$ of polynomials in x with coefficients in a field F.
- Let V be a finite dimensional vector space over F of dimension n and let T be a fixed linear transformation of V (i.e., from V to itself).
- We can consider V as an $F[x]$-module where the element x acts on V as the linear transformation T.
- So any polynomial in x acts on V as the same polynomial in T.
- Since V has finite dimension over F by assumption, it is by definition finitely generated as an F-module, hence certainly finitely generated as an $F[x]$-module.

  So the classification theorems apply.

Decompositions and Canonical Forms

- Any nonzero free $F[x]$-module (being isomorphic to a direct sum of copies of $F[x]$) is an infinite dimensional vector space over F.
- So if V has finite dimension over F, then it must in fact be a torsion $F[x]$-module (i.e., its free rank is 0).
- It follows from the Fundamental Theorem that then V is isomorphic as an $F[x]$-module to the direct sum of cyclic, torsion $F[x]$-modules.
- This decomposition of V allows us to choose a basis for V with respect to which the matrix representation for the linear transformation T is in a specific simple form:
  - When we use the invariant factor decomposition of V, we obtain the rational canonical form for the matrix for T.
  - When we use the elementary divisor decomposition (and when F contains all the eigenvalues of T) we obtain the Jordan canonical form.
- The uniqueness portion of the Fundamental Theorem ensures that the rational and Jordan canonical forms are unique (which is why they are referred to as canonical).

Eigenvalues, Eigenvectors and Eigenspaces

Definition (Eigenvalue, Eigenvector, Eigenspace)

(1) An element λ of F is called an **eigenvalue** of the linear transformation T if there is a nonzero vector $v \in V$, such that $T(v) = \lambda v$. In this situation v is called an **eigenvector** of T **with corresponding eigenvalue** λ.

(2) If A is an $n \times n$ matrix with coefficients in F, an element λ is called an **eigenvalue** of A **with corresponding eigenvector** v if v is a nonzero $n \times 1$ column vector such that $Av = \lambda v$.

(3) If λ is an eigenvalue of the linear transformation T, the set $\{v \in V : T(v) = \lambda v\}$ is called the **eigenspace** of T **corresponding to the eigenvalue** λ. Similarly, if λ is an eigenvalue of the $n \times n$ matrix A, the set of $n \times 1$ matrices v with $Av = \lambda v$ is called the **eigenspace** of A **corresponding to the eigenvalue** λ.

Eigenvalues of Transformations and of Matrices

- If we fix a basis $\mathcal{B}$ of V, then any linear transformation T of V has an associated $n \times n$ matrix A.
- Conversely, if A is any $n \times n$ matrix, then the map T defined by $T(v) = Av$, for $v \in V$, where the v on the right is the $n \times 1$ vector consisting of the coordinates of v with respect to the fixed basis $\mathcal{B}$ of V, is a linear transformation of V.
- Then v is an eigenvector of T with corresponding eigenvalue λ if and only if the coordinate vector of v with respect to $\mathcal{B}$ is an eigenvector of A with eigenvalue λ.

  In other words, the eigenvalues for the linear transformation T are the same as the eigenvalues for the matrix A of T with respect to any fixed basis for V.

Determinants and Eigenvalues

Definition (Determinant of a Linear Transformation)

The **determinant** of a linear transformation from V to V is the determinant of any matrix representing the linear transformation (not dependent on the choice of the basis used).

Proposition

The following are equivalent:

(1) A is an eigenvalue of T;

(2) $\lambda I - T$ is a singular linear transformation of V;

(3) $\det(\lambda I - T) = 0$.

- Since λ is an eigenvalue of T with corresponding eigenvector v if and only if v is a nonzero vector in the kernel of $\lambda I - T$, it follows that (1) and (2) are equivalent.

  (2) and (3) are equivalent by our results on determinants.

Characteristic Polynomials and Eigenvalues

Definition (Characteristic Polynomial)

Let x be an indeterminate over F. The polynomial $\det(xI - T)$ is called the **characteristic polynomial** of T and will be denoted $c_T(x)$.
If A is an $n \times n$ matrix with coefficients in F, $\det(xI - A)$ is called the **characteristic polynomial** of A and will be denoted $c_A(x)$.

- It is easy to see by expanding the determinant that the characteristic polynomial of either T or A is a monic polynomial of degree $n = \dim V$.
- The preceding proposition says that the set of eigenvalues of T (or A) is precisely the set of roots of the characteristic polynomial of T (of A, respectively).

  In particular, T has at most n distinct eigenvalues.

The Generator of the Annihilator of V in $F[x]$

- Consider V as a module over $F[x]$ via the linear transformation T.

  We argued it is a torsion $F[x]$-module.
- The annihilator $\mathrm{Ann}(V)$ of V in $F[x]$ is an ideal of $F[x]$.

  But $F[x]$ is a P.I.D.

  So $\mathrm{Ann}(V)$ is generated by a unique monic polynomial $m(x) \in F[x]$:
  - $m(x)$ is the unique monic polynomial of minimal degree annihilating V (i.e., $m(T)$ is the 0 linear transformation);
  - If $f(x) \in F[x]$ is any polynomial annihilating V, $m(x)$ divides $f(x)$.

The Minimal Polynomial

- The ring of all $n \times n$ matrices over F is isomorphic to the collection of all linear transformations of V to itself.
- So for any $n \times n$ matrix A over F, there is similarly a unique monic polynomial of minimal degree with $m(A)$ the zero matrix.

Minimal Polynomial

The unique monic polynomial which generates the ideal $\text{Ann}(V)$ in $F[x]$ is called the **minimal polynomial** of T and will be denoted $m_T(x)$.
The unique monic polynomial of smallest degree which when evaluated at the matrix A is the zero matrix is called the **minimal polynomial** of A and will be denoted $m_A(x)$.

Minimal Polynomial and Invariant Factors

- By the Invariant Factor Form Theorem, we have an isomorphism

$$V \cong F[x]/(a_1(x)) \oplus F[x]/(a_2(x)) \oplus \cdots \oplus F[x]/(a_m(x))$$

of $F[x]$-modules where $a_1(x), a_2(x), \ldots, a_m(x)$ are polynomials in $F[x]$ of degree at least one satisfying $a_1(x) \mid a_2(x) \mid \cdots \mid a_m(x)$.

These invariant factors $a_i(x)$ are only determined up to a unit in $F[x]$. But since the units of $F[x]$ are precisely the nonzero elements of F (i.e., the nonzero constant polynomials), we may make these polynomials unique by stipulating that they be monic.

Since the annihilator of V is the ideal $(a_m(x))$, we obtain:

Proposition

The minimal polynomial $m_T(x)$ is the largest invariant factor of V. All the invariant factors of V divide $m_T(x)$.

Action of Multiplication by x

- We now choose a basis for each of the direct summands for V in the decomposition for which the matrix for T is quite simple.
  The linear transformation T acting on V is the element x acting by multiplication on each of the factors of the direct sum.
  We have seen that the elements $1, \overline{x}, \overline{x}^2, \ldots, \overline{x}^{k-1}$ give a basis for the vector space $F[x]/(a(x))$, where
  $$a(x) = x^k + b_{k-1}x^{k-1} + \cdots + b_1 x + b_0$$
  is any monic polynomial in $F[x]$ and $\overline{x} = x \mod (a(x))$.
  With respect to this basis the linear transformation of multiplication by x acts in a simple manner:
  $$\begin{array}{l} 1 \mapsto \overline{x}; \quad \overline{x} \mapsto \overline{x}^2; \quad \overline{x}^2 \mapsto \overline{x}^3; \ldots, \overline{x}^{k-2} \mapsto \overline{x}^{k-1}; \\ \overline{x}^{k-1} \mapsto \overline{v}^k = -b_0 - b_1\overline{x} - \cdots - b_{k-1}\overline{x}^{k-1}. \end{array}$$
  The last equality holds because $\overline{x}^k + b_{k-1}\overline{x}^{k-1} + \cdots + b_1\overline{x} + b_0 = 0$, since $a(\overline{x}) = 0$ in $F[x]/(a(x))$.

The Companion Matrix

- In matrix form with respect to the basis $1, \overline{x}, \overline{x}^2, \ldots, \overline{x}^{k-1}$, multiplication by x takes the form:

$$\begin{pmatrix} 0 & 0 & \cdots & \cdots & \cdots & -b_0 \\ 1 & 0 & \cdots & \cdots & \cdots & -b_1 \\ 0 & 1 & \cdots & \cdots & \cdots & -b_2 \\ 0 & 0 & \ddots & & & \vdots \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & \cdots & 1 & -b_{k-1} \end{pmatrix}.$$

Definition (Companion Matrix)

Let $a(x) = x^k + b_{k-1}x^{k-1} + \cdots + b_1 x + b_0$ be any monic polynomial in $F[x]$. The **companion matrix** of $a(x)$ is the $k \times k$ matrix with 1's down the first subdiagonal, $-b_0, -b_1, \ldots, -b_{k-1}$ down the last column and zeros elsewhere. The companion matrix of $a(x)$ will be denoted by $\mathcal{C}_{a(x)}$.

The Matrix of T

- Consider $V \cong F[x]/(a_1(x)) \oplus F[x]/(a_2(x)) \oplus \cdots \oplus F[x]/(a_m(x))$ and let $\mathcal{B}_i$ be the elements of V corresponding to the basis chosen above for the cyclic factor $F[x]/(a_i(x))$ under the isomorphism.
- Then by definition the linear transformation T acts on $\mathcal{B}_i$ by the companion matrix for $a_i(x)$, since we have seen that this is how multiplication by x acts.
- The union $\mathcal{B}$ of the $\mathcal{B}_i$'s gives a basis for V since the sum is direct.
- With respect to this basis the linear transformation T has as matrix the direct sum of the companion matrices for the invariant factors, i.e., $\begin{pmatrix} \mathcal{C}_{a_1(x)} & & & \\ & \mathcal{C}_{a_2(x)} & & \\ & & \ddots & \\ & & & \mathcal{C}_{a_n(x)} \end{pmatrix}$.
- This matrix is uniquely determined from the invariant factors of the $F[x]$-module V and the list of invariant factors uniquely determines the module V up to isomorphism as an $F[x]$-module.

The Rational Canonical Form

Definition (Rational Canonical Form)

(1) A matrix is said to be in **rational canonical form** if it is the direct sum of companion matrices for monic polynomials $a_1(x), \ldots, a_m(x)$ of degree at least one with $a_1(x) \mid a_2(x) \mid \cdots \mid a_m(x)$. The polynomials $a_i(x)$ are called the **invariant factors** of the matrix. Such a matrix is also said to be a **block diagonal matrix** with blocks the companion matrices for the $a_i(x)$.

(2) A **rational canonical form** for a linear transformation T is a matrix representing T which is in rational canonical form.

Uniqueness of the Rational Canonical Form

- The process we used to determine the matrix of T from the direct sum decomposition is reversible.

  Suppose $b_1(x), b_2(x), \ldots, b_t(x)$ are monic polynomials in $F[x]$ of degree at least one such that $b_i(x) \mid b_{i+1}(x)$, for all i.

  Suppose for some basis $\mathcal{E}$ of V, that the matrix of T with respect to the basis $\mathcal{E}$ is the direct sum of the companion matrices of the $b_i(x)$.

  Then V must be a direct sum of T-stable subspaces D_i, one for each $b_i(x)$, in such a way that the matrix of T on each D_i is the companion matrix of $b_i(x)$.

  Let $\mathcal{E}_i$ be the corresponding (ordered) basis of D_i (so $\mathcal{E}$ is the union of the $\mathcal{E}_i$) and let e_i be the first basis element in $\mathcal{E}_i$.

  Then it is easy to see that D_i is a cyclic $F[x]$-module with generator e_i and that the annihilator of D_i is $b_i(x)$.

The Uniqueness Theorem

- Thus, the torsion $F[x]$-module V decomposes into a direct sum of cyclic $F[x]$-modules in two ways satisfying the theorem conditions. Since the invariant factors are unique, $a_i(x)$ and $b_i(x)$ must differ by a unit factor in $F[x]$. Since they are monic, $a_i(x) = b_i(x)$, for all i.

Theorem (Rational Canonical Form for Linear Transformations)

Let V be a finite dimensional vector space over the field F and let T be a linear transformation of V.

(1) There is a basis for V with respect to which the matrix for T is in rational canonical form, i.e., is a block diagonal matrix whose diagonal blocks are the companion matrices for monic polynomials $a_1(x), a_2(x), \ldots, a_m(x)$ of degree at least one with $a_1(x) \mid a_2(x) \mid \cdots \mid a_m(x)$.

(2) The rational canonical form for T is unique.

- The word **rational** indicates that this canonical form is calculated entirely within the field F and exists for any linear transformation T.

Similarity in the Language of $F[x]$-Modules

Theorem

Let S and T be linear transformations of V. Then the following are equivalent:

(1) S and T are similar linear transformations;

(2) the $F[x]$-modules obtained from V via S and via T are isomorphic $F[x]$-modules;

(3) S and T have the same rational canonical form.

(1) implies (2): Assume there is a non singular linear transformation U, such that $S = UTU^{-1}$. The vector space isomorphism $U : V \to V$ is also an $F[x]$-module homomorphism, where x acts on the first V via T and on the second via S. For example

$$U(xv) = U(Tv) = UT(v) = SU(v) = x(Uv).$$

Hence this is an $F[x]$-module isomorphism of the two modules in (2).

Similarity in the Language of $F[x]$-Modules (Cont'd)

(2) implies (3): Assume (2) holds and denote by:

- V_1 the vector space V made into an $F[x]$-module via S;
- V_2 the space V made into an $F[x]$-module via T.

Since $V_1 \cong V_2$ as $F[x]$-modules they have the same list of invariant factors. Thus S and T have a common rational canonical form.

(3) implies (1): Assume (3) holds. S and T have the same matrix representation with respect to some choice of (possibly different) bases of V by assumption. Hence, they are, up to a change of basis, the same linear transformation of V. Hence, are similar.

Rational Canonical Form for Matrices

- Let A be any $n \times n$ matrix with entries from F. Let V be an n-dimensional vector space over F.
  - We can then define a linear transformation T on V by choosing a basis for V and setting $T(v) = Av$, where v on the right hand side means the $n \times 1$ column vector of coordinates of v with respect to our chosen basis.
  - Clearly, the matrix for this T with respect to this basis is A.

Theorem (Rational Canonical Form for Matrices)

Let A be an $n \times n$ matrix over the field F.

(1) The matrix A is similar to a matrix in rational canonical form, i.e., there is an invertible $n \times n$ matrix P over F, such that $P^{-1}AP$ is a block diagonal matrix whose diagonal blocks are the companion matrices for monic polynomials $a_1(x), a_2(x), \ldots, a_m(x)$ of degree at least one with $a_1(x) \mid a_2(x) \mid \cdots \mid a_m(x)$.

(2) The rational canonical form for A is unique.

Invariant Factors of Matrices and Similarity

Definition

The **invariant factors** of an $n \times n$ matrix over a field F are the invariant factors of its rational canonical form.

Theorem

Let A and B be $n \times n$ matrices over the field F. Then A and B are similar and only if A and B have the same rational canonical form.

- If A is a matrix with entries from a field F and F is a subfield of a larger field K, then we may also consider A as a matrix over K.

Corollary (Rational Canonical Form and Field)

Let A and B be two $n \times n$ matrices over a field F and suppose F is a subfield of the field K.

(1) The rational canonical form of A is the same whether it is computed over K or over F. The minimal and characteristic polynomials and the invariant factors of A are the same whether A is considered as a matrix over F or as a matrix over K.

Similarity and Field

Corollary (Cont'd)

Let A and B be two $n \times n$ matrices over a field F and suppose F is a subfield of the field K.

(2) The matrices A and B are similar over K if and only if they are similar over F, i.e., there exists an invertible $n \times n$ matrix P with entries from K, such that $B = P^{-1}AP$ if and only if there exists an (in general different) invertible $n \times n$ matrix Q with entries from F, such that $B = Q^{-1}AQ$.

(1) Let M be the rational canonical form of A when computed over the smaller field F. Since M satisfies the conditions in the definition of the rational canonical form over K, the uniqueness of the rational canonical form implies that M is also the rational canonical form of A over K. Hence the invariant factors of A are the same whether A is viewed over F or over K.

The Proof (Cont'd)

- In particular, since the minimal polynomial is the largest invariant factor of A, it also does not depend on the field over which A is viewed.

  It is clear from the determinant definition of the characteristic polynomial of A that this polynomial depends only on the entries of A.

(2) If A and B are similar over the smaller field F they are clearly similar over K.

Conversely, if A and B are similar over K, they have the same rational canonical form over K. Then, they have the same rational canonical form over F, hence are similar over F.

- The corollary asserts, among other things, that:
  - The rational canonical form for an $n \times n$ matrix A is an $n \times n$ matrix with entries in the smallest field containing the entries of A;
  - This canonical form is the same matrix even if we allow conjugation of A by nonsingular matrices whose entries come from larger fields.

Characteristic Polynomial and Invariant Factors

- A connection between the characteristic polynomial of a matrix (or of a linear transformation) and its invariant factors is given by

Lemma

Let $a(x) \in F[x]$ be any monic polynomial.

(1) The characteristic polynomial of the companion matrix of $a(x)$ is $a(x)$.

(2) If M is the block diagonal matrix $M = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_k \end{pmatrix}$, given by the direct sum of matrices $A_1, A_2, \ldots, A_k$, then the characteristic polynomial of M is the product of the characteristic polynomials of $A_1, A_2, \ldots, A_k$.

- The proof is omitted.

The Cayley-Hamilton Theorem

Proposition

Let A be an $n \times n$ matrix over the field F.

(1) The characteristic polynomial of A is the product of all the invariant factors of A.

(2) (**The Cayley-Hamilton Theorem**) The minimal polynomial of A divides the characteristic polynomial of A.

(3) The characteristic polynomial of A divides some power of the minimal polynomial of A. In particular these polynomials have the same roots, not counting multiplicities.

- The same statements are true if the matrix A is replaced by a linear transformation T of an n-dimensional vector space over F.

Proof of the Cayley-Hamilton Theorem

(1) Let B be the rational canonical form of A. By the previous lemma, the block diagonal form of B shows that the characteristic polynomial of B is the product of the characteristic polynomials of the companion matrices of the invariant factors of A. The characteristic polynomial of the companion matrix $\mathcal{C}_{a(x)}$ for $a(x)$ is just $a(x)$. Thus, the characteristic polynomial for B is the product of the invariant factors of A. Since A and B are similar, they have the same characteristic polynomial.

(2) Immediate from (1) since the minimal polynomial for A is the largest invariant factor of A.

(3) The fact that all the invariant factors divide the largest one immediately implies (3).

- Part (2) says that:
  - The matrix A satisfies its own characteristic polynomial: $c_A(A) = 0$.
  - Thus, the degree of the minimal polynomial for A is at most n.

Invariant Factor Decomposition Algorithm: Step (1)

- Let V be an $F[x]$-module with vector space basis $[e_1, e_2, \ldots, e_n]$.

  Let T be the linear transformation of V to itself defined by x.

  Let $A = (a_{ij})$ be the $n \times n$ matrix associated to T and this choice of basis for V, i.e.,

$$T(e_j) = xe_j = \sum_{i=1}^{n} a_{ij} e_i.$$

(1) Use the following three elementary row and column operations to diagonalize the matrix $xI - A$ over $F[x]$, keeping track of the row operations used:

(a) Interchange two rows or columns (denoted $R_i \leftrightarrow R_j$ and $C_i \leftrightarrow C_j$);

(b) Add a multiple (in $F[x]$) of one row or column to another (denoted $R_i \leftarrow R_i + p(x)R_j$ and $C_i \leftarrow C_i + p(x)C_j$);

(c) Multiply any row or column by a unit in $F[x]$, i.e., by a nonzero element in F (which will be denoted by uR_i and uC_i).

Invariant Factor Decomposition Algorithm: Step (2)

(2) Beginning with the $F[x]$-module generators $[e_1, e_2, \ldots, e_n]$, for each row operation used in Step (1), change the set of generators by the following rules:

   (a) If the i-th row is interchanged with the j-th row, then interchange the i-th and j-th generators;

   (b) If $p(x)$ times the j-th row is added to the i-th row, then subtract $p(x)$ times the i-th generator from the j-th generator;

   (c) If the i-th row is multiplied by the unit $u \in F$, then divide the i-th generator by u.

Invariant Factor Decomposition Algorithm: Step (3)

(3) When $xI - A$ has been diagonalized, the generators $[e_1, e_2, \ldots, e_n]$ for V will be in the form of $F[x]$-linear combinations of $e_1, e_2, \ldots, e_n$. Use $xe_j = T(e_j) = \sum_{i=1}^{n} a_{ij} e_i$ to write these elements as F-linear combinations of $e_1, e_2, \ldots, e_n$.

When $xI - A$ has been diagonalized,

- the first $n - m$ of these linear combinations are 0 (providing a useful numerical check on the computations);
- the remaining m linear combinations are nonzero, i.e., the generators for V are in the form $[0, \ldots, 0, f_1, \ldots, f_m]$.

The elements $f_1, \ldots, f_m$ are a set of $F[x]$-module generators for the cyclic factors in the invariant factor decomposition of V (with annihilators $(a_1(x)), \ldots, (a_m(x))$, respectively):

$$\begin{array}{l} V = F[x]f_1 \oplus F[x]f_2 \oplus \cdots \oplus F[x]f_m, \\ F[x]f_i \cong F[x]/(a_i(x)), \quad i = 1, 2, \ldots, m, \end{array}$$

giving the Invariant Factor Decomposition of the $F[x]$-module V.

Invariant Factor Decomposition Algorithm: Steps (4),(5)

(4) The corresponding vector space basis for each cyclic factor of V is then given by the elements

$$f_i, Tf_i, T^2 f_i, \ldots, T^{\deg a_i(x)-1} f_i.$$

(5) Write the k-th element of the vector space basis computed in Step (4) in terms of the original vector space basis $[e_1, e_2, \ldots, e_n]$;
Use the coordinates for the k-th column of an $n \times n$ matrix P.
Then $P^{-1}AP$ is in rational canonical form (with diagonal blocks the companion matrices for the $a_i(x)$).
This is the matrix for the linear transformation T with respect to the vector space basis in Step (4).

Converting a Matrix to Rational Canonical Form: Step (1)

- Let A be an $n \times n$ matrix with entries in the field F.
  1. Use the following three elementary row and column operations to diagonalize the matrix $xI - A$ over $F[x]$, keeping track of the row operations used:
     (a) Interchange two rows or columns (denoted $R_i \leftrightarrow R_j$ and $C_i \leftrightarrow C_j$);
     (b) Add a multiple (in $F[x]$) of one row or column to another (denoted $R_i \leftarrow R_i + p(x)R_j$ and $C_i \leftarrow C_i + p(x)C_j$);
     (c) Multiply any row or column by a unit in $F[x]$, i.e., by a nonzero element in F (denoted uR_i and uC_i).

  Define $d_1, \ldots, d_m$ to be the degrees of the monic nonconstant polynomials $a_1(x), \ldots, a_m(x)$ appearing on the diagonal.

Converting a Matrix to Rational Canonical Form: Step (2)

(2) Beginning with the $n \times n$ identity matrix P', for each row operation used in (1), change the matrix P' by the following rules:

  (a) If $R_i \leftrightarrow R_j$, then interchange the i-th and j-th columns of P' (i.e., $C_i \leftrightarrow C_j$ for P');

  (b) If $R_i \leftarrow R_i + p(x)R_j$, then subtract the product of the matrix $p(A)$ times the i-th column of P' from the j-th column of P' (i.e., $C_j \leftarrow C_j - p(A)C_i$ for P');

  (c) If uR_i, then divide the elements of the i-th column of P' by u (i.e., $u^{-1}C_i$ for P').

Converting a Matrix to Rational Canonical Form: Step (3)

(3) When $xI - A$ has been diagonalized, the first $n - m$ columns of the matrix P' are 0 (providing a useful numerical check on the computations) and the remaining m columns of P' are nonzero.
For each $i = 1, 2, \ldots, m$, multiply the i-th nonzero column of P' successively by $A^0 = I, A^1, A^2, \ldots, A^{d_i - 1}$, where d_i is the integer in (1); Use the resulting column vectors (in this order) as the next d_i columns of an $n \times n$ matrix P.
Then $P^{-1}AP$ is in rational canonical form (whose diagonal blocks are the companion matrices for the polynomials $a_1(x), \ldots a_m(x)$ in (1)).

Linear Transformations/Matrices vs Finite Abelian Groups

- In the theory of canonical forms for linear transformations/matrices:
  - The characteristic polynomial plays the role of the order of a finite abelian group;
  - The minimal polynomial plays the role of the exponent.

  They are the same invariants, one for modules over the Principal Ideal Domain $\mathbb{Z}$ and the other for modules over the Principal Ideal Domain $F[x]$.

Problems on Matrices and on Finite Abelian Groups

- The characteristic polynomial / order and the minimal polynomial / exponent dualities suggest solving problems for linear transformations / matrices directly analogous to those we considered for finite abelian groups:
  - (A) Determine the rational canonical form of a given matrix (Decompose a finite abelian group as a direct product of cyclic groups);
  - (B) Determine whether two given matrices are similar (Determine whether two given finite abelian groups are isomorphic);
  - (C) Determine all similarity classes of matrices over F with a given characteristic polynomial (Determine all abelian groups of a given order);
  - (D) Determine all similarity classes of $n \times n$ matrices over F with a given minimal polynomial (Determine all abelian groups of rank at most n of a given exponent).

Example

- We find the rational canonical forms of the following matrices over $\mathbb{Q}$ and determine if they are similar:
  $A = \begin{pmatrix} 2 & -2 & 14 \\ 0 & 3 & -7 \\ 0 & 0 & 2 \end{pmatrix}, B = \begin{pmatrix} 0 & -4 & 85 \\ 1 & 4 & -30 \\ 0 & 0 & 3 \end{pmatrix}, C = \begin{pmatrix} 2 & 2 & 1 \\ 0 & 2 & -1 \\ 0 & 0 & 3 \end{pmatrix}.$
  A direct computation shows that all three of these matrices have the same characteristic polynomial:
  $$c_A(x) = c_B(x) = c_C(x) = (x-2)^2(x-3).$$
  Since the minimal and characteristic polynomials have the same roots, the only possibilities for the minimal polynomials are $(x-2)(x-3)$ or $(x-2)^2(x-3)$. We quickly find that:
  - $(A-2I)(A-3I) = 0$,
  - $(B-2I)(B-3I) \neq 0$ (the $1,1$-entry is nonzero) and
  - $(C-2I)(C-3I) \neq 0$ (the $1,2$-entry is nonzero).

  So $m_A(x) = (x-2)(x-3)$, $m_B(x) = m_C(x) = (x-2)^2(x-3)$.

Example (Cont'd)

- Hence:
  - There are no additional invariant factors for B and C;
  - Since the invariant factors for A divide the minimal polynomial and have product the characteristic polynomial, A has for invariant factors the polynomials $x-2$, $(x-2)(x-3)=x^2-5x+6$.

  We conclude that B and C are similar and neither is similar to A.

  Note that $(x-2)^2(x-3)=x^3-7x^2+16x-12$.

  Hence, the rational canonical forms are:

$$\left(\begin{array}{rrr} 2 & 0 & 0 \\ 0 & 0 & -6 \\ 0 & 1 & 5 \end{array}\right), \quad \left(\begin{array}{rrr} 0 & 0 & 12 \\ 1 & 0 & -16 \\ 0 & 1 & 7 \end{array}\right), \quad \left(\begin{array}{rrr} 0 & 0 & 12 \\ 1 & 0 & -16 \\ 0 & 1 & 7 \end{array}\right).$$

Invariant Factor Decomposition

- We use row and column operations (in $\mathbb{Q}[x]$) to reduce the matrix $xI - A = \begin{pmatrix} x-2 & 2 & -14 \\ 0 & x-3 & 7 \\ 0 & 0 & x-2 \end{pmatrix}$ to diagonal form.

$$\begin{pmatrix} x-2 & 2 & -14 \\ 0 & x-3 & 7 \\ 0 & 0 & x-2 \end{pmatrix} \xrightarrow[R_1 \leftarrow R_1 + R_2]{} \begin{pmatrix} x-2 & x-1 & -7 \\ 0 & x-3 & 7 \\ 0 & 0 & x-2 \end{pmatrix} \xrightarrow[C_1 \leftarrow C_1 - C_2]{}$$

$$\begin{pmatrix} -1 & x-1 & -7 \\ -x+3 & x-3 & 7 \\ 0 & 0 & x-2 \end{pmatrix} \xrightarrow[-R_1]{} \begin{pmatrix} 1 & -x+1 & 7 \\ -x+3 & x-3 & 7 \\ 0 & 0 & x-2 \end{pmatrix} \xrightarrow[R_2 \leftarrow R_2 + (x-3)R_1]{}$$

$$\begin{pmatrix} 1 & -x+1 & 7 \\ 0 & -x^2+5x-6 & 7(x-2) \\ 0 & 0 & x-2 \end{pmatrix} \xrightarrow[C_2 \leftarrow C_2 + (x-1)C_1]{}$$

$$\begin{pmatrix} 1 & 0 & 7 \\ 0 & -x^2+5x-6 & 7(x-2) \\ 0 & 0 & x-2 \end{pmatrix} \xrightarrow[C_3 \leftarrow C_3 - 7C_1]{} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -x^2+5x-6 & 7(x-2) \\ 0 & 0 & x-2 \end{pmatrix} \xrightarrow[-C_2]{}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & x^2-5x+6 & 7(x-2) \\ 0 & 0 & x-2 \end{pmatrix} \xrightarrow[R_2 \leftarrow R_2 - 7R3]{} \begin{pmatrix} 1 & 0 & 0 \\ 0 & x^2-5x+6 & 0 \\ 0 & 0 & x-2 \end{pmatrix} \xrightarrow[R_2 \leftrightarrow R3,\ C_2 \leftrightarrow C_3]{}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & x-2 & 0 \\ 0 & 0 & x^2-5x+6 \end{pmatrix}$$

Invariant Factor Decomposition (Cont'd)

- This determines the invariant factors $x-2, x^2-5x+6$ for this matrix.

  Let now V be a 3-dimensional vector space over $\mathbb{Q}$ with basis e_1, e_2, e_3. Let T be the corresponding linear transformation (which defines the action of x on V), i.e., $xe_1 = T(e_1) = 2e_1$, $xe_2 = T(e_2) = -2e_1 + 3e_2$, $xe_3 = T(e_3) = 14e_1 - 7e_2 + 2e_3$.

  The row operations used in the reduction above were $R_1 \leftarrow R_1 + R_2, -R_1, R_2 \leftarrow R_2 + (x-3)R_1, R_2 \leftarrow R_2 - 7R_3, R_2 \leftrightarrow R_3$.

  Starting with the basis $[e_1, e_2, e_3]$ for V and changing it according to the rules given in the text, we obtain

$$\begin{array}{rcl}[e_1, e_2, e_3] & \rightarrow & [e_1, e_2 - e_1, e_3] \rightarrow [-e_1, e_2 - e_1, e_3] \\ & \rightarrow & [-e_1 - (x-3)(e_2 - e_1), e_2 - e_1, e_3] \\ & \rightarrow & [-e_1 - (x-3)(e_2 - e_1), e_2 - e_1, e_3 + 7(e_2 - e_1)] \\ & \rightarrow & [-e_1 - (x-3)(e_2 - e_1), e_3 + 7(e_2 - e_1), e_2 - e_1].\end{array}$$

Invariant Factor Decomposition (Cont'd)

- Using the formulas above for the action of x, we see that these last elements are the elements $[0, -7e_1 + 7e_2 + e_3, -e_1 + e_2]$ of V corresponding to the elements $1, x-2$ and $x^2 - 5x + 6$ in the diagonalized form of $xI - A$, respectively.

  The elements $f_1 = -7e_1 + 7e_2 + e_3$ and $f_2 = -e_1 + e_2$ are therefore $\mathbb{Q}[x]$-module generators for the two cyclic factors of V in its invariant factor decomposition as a $\mathbb{Q}[x]$-module.

  The corresponding $\mathbb{Q}$-vector space bases for these two factors are then f_1 and f_2, $xf_2 = Tf_2$, i.e., $-7e_1 + 7e_2 + e_3$ and $-e_1 + e_2$, $T(-e_1 + e_2) = -4e_1 + 3e_2$.

  Then the matrix $P = \begin{pmatrix} -7 & -1 & -4 \\ 7 & 1 & 3 \\ 1 & 0 & 0 \end{pmatrix}$ conjugates A into its rational canonical form: $P^{-1}AP = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & -6 \\ 0 & 1 & 5 \end{pmatrix}$.

Converting A Directly to Rational Canonical Form

- We use the row operations involved in the diagonalization of $xI - A$ to determine the matrix P' of the algorithm above:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow[C_2 \leftarrow C_2 - C_1]{} \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow[-C_1]{} \begin{pmatrix} -1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow[C_1 \leftarrow C_1 - (A-3I)C_2]{}$$

$$\begin{pmatrix} 0 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow[C_3 \leftarrow C_3 + 7C_2]{} \begin{pmatrix} 0 & -1 & -7 \\ 0 & 1 & 7 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow[C_2 \leftrightarrow C_3]{} \begin{pmatrix} 0 & -7 & -1 \\ 0 & 7 & 1 \\ 0 & 1 & 0 \end{pmatrix} = P'$$

  Here we have $d_1 = 1$ and $d_2 = 2$, corresponding to the second and third nonzero columns of P', respectively. The columns of P are therefore given by
  $$\begin{pmatrix} -7 \\ 7 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, A\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -4 \\ 3 \\ 0 \end{pmatrix},$$
  respectively, which again gives the matrix P above.

A 4×4 Matrix

- Consider the matrix $D = \begin{pmatrix} 1 & 2 & -4 & 4 \\ 2 & -1 & 4 & -8 \\ 1 & 0 & 1 & -2 \\ 0 & 1 & -2 & 3 \end{pmatrix}$. We compute its characteristic polynomial:

$$\begin{vmatrix} x-1 & -2 & 4 & -4 \\ -2 & x+1 & -4 & 8 \\ -1 & 0 & x-1 & 2 \\ 0 & -1 & 2 & x-3 \end{vmatrix} = \begin{vmatrix} x-1 & -2 & 4 & -4 \\ 0 & x+1 & -2(x+1) & 4 \\ -1 & 0 & x-1 & 2 \\ 0 & -1 & 2 & x-3 \end{vmatrix} =$$

$$(x-1)\begin{vmatrix} x+1 & -2(x+1) & 4 \\ 0 & x-1 & 2 \\ -1 & 2 & x-3 \end{vmatrix} - \begin{vmatrix} -2 & 4 & -4 \\ x+1 & -2(x+1) & 4 \\ -1 & 2 & x-3 \end{vmatrix} =$$

$$(x-1)\begin{vmatrix} x+1 & 0 & 4 \\ 0 & x-1 & 2 \\ -1 & 0 & x-3 \end{vmatrix} - \begin{vmatrix} 0 & 0 & -2(x-1) \\ x+1 & -2(x+1) & 4 \\ -1 & 2 & x-3 \end{vmatrix} =$$

$$(x-1)^2\begin{vmatrix} x+1 & 4 \\ -1 & x-3 \end{vmatrix} + 2(x-1)\begin{vmatrix} x+1 & -2(x+1) \\ -1 & 2 \end{vmatrix} =$$

$$(x-1)^2(x^2-2x-3+4) + 2(x-1)(2x+2-2x-2) =$$

$$(x-1)^2(x-1)^2 + 2(x-1) \cdot 0 = (x-1)^4.$$

A 4 × 4 Matrix

- Consider the matrix $D = \begin{pmatrix} 1 & 2 & -4 & 4 \\ 2 & -1 & 4 & -8 \\ 1 & 0 & 1 & -2 \\ 0 & 1 & -2 & 3 \end{pmatrix}$. We computed the characteristic polynomial of D, $(x-1)^4$. The possible minimal polynomials are then $x-1, (x-1)^2, (x-1)^3$ and $(x-1)^4$. We have

$$\begin{aligned} D - I &\neq 0; \\ (D-I)^2 &= \begin{pmatrix} 0 & 2 & -4 & 4 \\ 2 & -2 & 4 & -8 \\ 1 & 0 & 0 & -2 \\ 0 & 1 & -2 & 2 \end{pmatrix} \begin{pmatrix} 0 & 2 & -4 & 4 \\ 2 & -2 & 4 & -8 \\ 1 & 0 & 0 & -2 \\ 0 & 1 & -2 & 2 \end{pmatrix} = 0. \end{aligned}$$

So the minimal polynomial for D is $(x-1)^2$.
There are then two possible sets of invariant factors:

$$x-1, x-1, (x-1)^2 \quad \text{and} \quad (x-1)^2, (x-1)^2.$$

A 4×4 Matrix (Cont'd)

- To determine the invariant factors for D we apply the procedure of the previous example to the 4×4 matrix
$$xI - D = \begin{pmatrix} x-1 & -2 & 4 & -4 \\ -2 & x+1 & -4 & 8 \\ -1 & 0 & x-1 & 2 \\ 0 & -1 & 2 & x-3 \end{pmatrix}.$$

$$\begin{pmatrix} x-1 & -2 & 4 & -4 \\ -2 & x+1 & -4 & 8 \\ -1 & 0 & x-1 & 2 \\ 0 & -1 & 2 & x-3 \end{pmatrix} \xrightarrow{R_1 \leftrightarrow R_3} \begin{pmatrix} -1 & 0 & x-1 & 2 \\ -2 & x+1 & -4 & 8 \\ x-1 & -2 & 4 & -4 \\ 0 & -1 & 2 & x-3 \end{pmatrix}$$

$$\xrightarrow{-R_1} \begin{pmatrix} 1 & 0 & -x-+1 & -2 \\ -2 & x+1 & -4 & 8 \\ x-1 & -2 & 4 & -4 \\ 0 & -1 & 2 & x-3 \end{pmatrix} \xrightarrow{R_2 \leftarrow R_2+2R_1} \begin{pmatrix} 1 & 0 & -x-+1 & -2 \\ 0 & x+1 & -2x-2 & 4 \\ x-1 & -2 & 4 & -4 \\ 0 & -1 & 2 & x-3 \end{pmatrix}$$

$$\xrightarrow{R_3 \leftarrow R_3-(x-1)R_1} \begin{pmatrix} 1 & 0 & -x-+1 & -2 \\ 0 & x+1 & -2x-2 & 4 \\ 0 & -2 & x^2-2x+5 & 2x-6 \\ 0 & -1 & 2 & x-3 \end{pmatrix} \xrightarrow{C_3 \leftarrow C_3+(x-1)C_1} \begin{pmatrix} 1 & 0 & 0 & -2 \\ 0 & x+1 & -2x-2 & 4 \\ 0 & -2 & x^2-2x+5 & 2x-6 \\ 0 & -1 & 2 & x-3 \end{pmatrix}$$

$$\xrightarrow{C_4 \leftarrow C_4+2C_1} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & x+1 & -2x-2 & 4 \\ 0 & -2 & x^2-2x+5 & 2x-6 \\ 0 & -1 & 2 & x-3 \end{pmatrix} \xrightarrow{R_2 \leftrightarrow R_4} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 2 & x-3 \\ 0 & -2 & x^2-2x+5 & 2x-6 \\ 0 & x+1 & -2x-2 & 4 \end{pmatrix}$$

$$\xrightarrow{-R_2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & -x+3 \\ 0 & -2 & x^2-2x+5 & 2x-6 \\ 0 & x+1 & -2x-2 & 4 \end{pmatrix}$$

A 4 × 4 Matrix (Cont'd)

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & -x+3 \\ 0 & -2 & x^2-2x+5 & 2x-6 \\ 0 & x+1 & -2x-2 & 4 \end{pmatrix} \xrightarrow{R_3 \leftarrow R_3+2R_2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & -x+3 \\ 0 & 0 & (x-1)^2 & 0 \\ 0 & x+1 & -2x-2 & 4 \end{pmatrix}$$

$$\xrightarrow{R_4 \leftarrow R_4-(x+1)R_2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & -x+3 \\ 0 & 0 & (x-1)^2 & 0 \\ 0 & 0 & 0 & (x-1)^2 \end{pmatrix} \xrightarrow{C_3 \leftarrow C_3+2C_2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -x+3 \\ 0 & 0 & (x-1)^2 & 0 \\ 0 & 0 & 0 & (x-1)^2 \end{pmatrix}$$

$$\xrightarrow{C_4 \leftarrow C_4+(x-3)C_2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & (x-1)^2 & 0 \\ 0 & 0 & 0 & (x-1)^2 \end{pmatrix}.$$

Thus, the invariant factors for D are $(x-1)^2$, $(x-1)^2$.

Moreover, a series of elementary row and column operations diagonalizing $xI-D$ is $R_1 \leftrightarrow R_3$, $-R_1$, $R_2 \leftarrow R_2+2R_1$, $R_3 \leftarrow R_3-(x-1)R_1$, $C_3 \leftarrow C_3+(x-1)C_1$, $C_4 \leftarrow C_4+2C_1$, $R_2 \leftrightarrow R_4$, $-R_2$, $R_3 \leftarrow R_3+2R_2$, $R_4 \leftarrow R_4-(x+1)R_2$, $C_3 \leftarrow C_3+2C_2$, $C_4 \leftarrow C_4+(x-3)C_2$.

4×4 Invariant Factor Decomposition

- In the reduction, the following row operations were used:

$$\begin{array}{llll} R_1 \leftrightarrow R_3, & -R_1, & R_2 \leftarrow R_2 + 2R_1, & R_3 \leftarrow R_3 - (x-1)R_1, \\ R_2 \leftrightarrow R_4, & -R_2, & R_3 \leftarrow R_3 + 2R_2, & R_4 \leftarrow R_4 - (x+1)R_2. \end{array}$$

Translate to column operations:

$$\begin{array}{llll} C_1 \leftrightarrow C_3, & -C_1 & C_1 \leftarrow C_1 - C_2, & C_1 \leftarrow C_1 + (x-1)C_3, \\ C_2 \leftrightarrow C_4, & -C_2, & C_2 \leftarrow C_2 - 2C_3, & C_2 \leftarrow C_2 + (x+1)C_4. \end{array}$$

Use on a fixed basis for V, say $[e_1, e_2, e_3, e_4]$:

$$\begin{array}{l} [e_1, e_2, e_3, e_4] \to [e_3, e_2, e_1, e_4] \to [-e_3, e_2, e_1, e_4] \\ \to [-2e_2 - e_3, e_2, e_1, e_4] \to [(x-1)e_1 - 2e_2 - e_3, e_2, e_1, e_4] \\ \to [(x-1)e_1 - 2e_2 - e_3, e_4, e_1, e_2] \\ \to [(x-1)e_1 - 2e_2 - e_3, -e_4, e_1, e_2] \\ \to [(x-1)e_1 - 2e_2 - e_3, -2e_1 - e_4, e_1, e_2] \\ \to [(x-1)e_1 - 2e_2 - e_3, -2e_1 + (x+1)e_2 - e_4, e_1, e_2]. \end{array}$$

4×4 Invariant Factor Decomposition (Cont'd)

- If e_1, e_2, e_3, e_4 is a basis for V, then we found that the generators of V corresponding to the factors are

$$(x-1)e_1 - 2e_2 - e_3 = 0, -2e_1 + (x+1)e_2 - e_4 = 0, e_1, e_2.$$

Hence a vector space basis for the two direct factors in the invariant decomposition of V in this case is given by e_1, Te_1 and e_2, Te_2, where T is the linear transformation defined by D, i.e., $e_1, e_1 + 2e_2 + e_3$ and $e_2, 2e_1 - e_2 + e_4$.

So $P = \begin{pmatrix} 1 & 1 & 0 & 2 \\ 0 & 2 & 1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ and $P^{-1}DP = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 2 \end{pmatrix}$.

4×4 Rational Canonical Form

- We determine the matrix P' of the algorithm from the row operations used in the diagonalization of $xI - D$:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{C_1 \leftrightarrow C_3} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{-C_1} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{C_1 \leftarrow C_1 - 2C_2}$$
$$\begin{pmatrix} 0 & 0 & 1 & 0 \\ -2 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{C_1 \leftarrow C_1 + (D-I)C_3} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{C_2 \leftrightarrow C_4} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \xrightarrow{-C_2}$$
$$\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \xrightarrow{C_2 \leftarrow C_2 - 2C_3} \begin{pmatrix} 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \xrightarrow{C_2 \leftarrow C_2 + (D+I)C_4} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} =$$

P'. Here we have $d_1 = 2$ and $d_2 = 2$, corresponding to the third and fourth nonzero columns of P'. The columns of P are therefore given by

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, D \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, D \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ 0 \\ 1 \end{pmatrix}$$

respectively, which again gives the matrix P above.

Determining All Similarity Classes

- We determine all similarity classes of matrices A with entries from $\mathbb{Q}$ with characteristic polynomial $(x^4-1)(x^2-1)$.

  First note that any matrix with a degree 6 characteristic polynomial must be a 6×6 matrix.

  The polynomial $(x^4-1)(x^2-1)$ factors into irreducibles in $\mathbb{Q}[x]$ as

  $$(x-1)^2(x+1)^2(x^2+1).$$

  Since the minimal polynomial $m_A(x)$ for A has the same roots as $c_A(x)$, it follows that $(x-1)(x+1)(x^2+1)$ divides $m_A(x)$.

Determining All Similarity Classes (Invariant Factors)

- Suppose $a_1(x), \ldots, a_m(x)$ are the invariant factors of some A.

  Then we must have:
  - $a_m(x) = m_A(x)$;
  - $a_i(x) \mid a_{i+1}(x)$ (in particular, all the invariant factors divide $m_A(x)$);
  - $a_1(x)a_2(x)\cdots a_m(x) = (x^4-1)(x^2-1)$.

  One easily sees that the only permissible lists under these constraints are:

  (a) $(x-1)(x+1), (x-1)(x+1)(x^2+1)$;
  (b) $x-1, (x-1)(x+1)^2(x^2+1)$;
  (c) $x+1, (x-1)^2(x+1)(x^2+1)$;
  (d) $(x-1)^2(x+1)^2(x^2+1)$.

  We can now write out the corresponding direct sums of companion matrices to obtain representatives of the 4 similarity classes.

Finding All Similarity Classes

- We find all similarity classes of 3×3 matrices A with entries from $\mathbb{Q}$ satisfying $A^6 = I$. Its minimal polynomial divides $x^6 - 1$.

  In $\mathbb{Q}[x]$, the complete factorization is

  $$x^6 - 1 = (x-1)(x+1)(x^2 - x + 1)(x^2 + x + 1).$$

  Conversely, if B is any 3×3 matrix whose minimal polynomial divides $x^6 - 1$, then $B^6 = I$.

  The minimal polynomial for B has degree at most 3.

  Thus, the possibilities for the minimal polynomial are:

  (a) $x - 1$;

  (b) $x + 1$;

  (c) $x^2 - x + 1$;

  (d) $x^2 + x + 1$;

  (e) $(x-1)(x+1)$;

  (f) $(x-1)(x^2 - x + 1)$;

  (g) $(x-1)(x^2 + x + 1)$;

  (h) $(x+1)(x^2 - x + 1)$;

  (i) $(x+1)(x^2 + x + 1)$.

Finding All Similarity Classes (Cont'd)

- Under the constraints of the rational canonical form these give rise to the following permissible lists of invariant factors:

  (i) $x-1, x-1, x-1$;

  (ii) $x+1, x+1, x+1$;

  (iii) $x-1, (x-1)(x+1)$;

  (iv) $x+1, (x-1)(x+1)$;

  (v) $(x-1)(x^2-x+1)$;

  (vi) $(x-1)(x^2+x+1)$;

  (vii) $(x+1)(x^2-x+1)$;

  (viii) $(x+1)(x^2+x+1)$.

  Note that it is impossible to have a suitable set of invariant factors if the minimal polynomial is x^2+x+1 or x^2-x+1. One can now write out the corresponding rational canonical forms; for example, (i) is I, (ii) is $-I$ and (iii) is $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$.

  Another way of phrasing this is that any 3×3 matrix with entries from $\mathbb{Q}$ whose order divides 6 is similar to one of these 8 matrices.

Subsection 4

The Jordan Canonical Form

Assumption on Eigenvalues

- The elementary divisors of a module are the prime power divisors of its invariant factors.
- For the $F[x]$-module V, the invariant factors were monic polynomials $a_1(x), a_2(x), \ldots, a_m(x)$ of degree ≥ 1, $a_1(x) \mid a_2(x) \mid \cdots \mid a_m(x)$.
- So the associated elementary divisors are the powers of the irreducible polynomial factors of these polynomials.
- To obtain the simplest possible elementary divisors we shall assume that the polynomials $a_1(x), a_2(x), \ldots, a_m(x)$ factor completely into linear factors, i.e., that the elementary divisors of V are powers $(x - \lambda)^k$ of linear polynomials. Since the product of the elementary divisors is the characteristic polynomial, this is equivalent to the assumption that the field F contains all the eigenvalues of T.
- Then V is the direct sum of finitely many cyclic $F[x]$-modules of the form $F[x]/(x - \lambda)^k$, where $\lambda \in F$ is one of the eigenvalues of T, corresponding to the elementary divisors of V.

Choosing a Special Basis

- We now choose a vector space basis for each of the direct summands corresponding to the elementary divisors of V for which the corresponding matrix for T is particularly simple.
- By definition of the $F[x]$-module structure the linear transformation T acting on V is the element x acting by multiplication on each of the direct summands $F[x]/(x-\lambda)^k$.
- Consider the elements $(\overline{x}-\lambda)^{k-1},(\overline{x}-\lambda)^{k-2},\ldots,\overline{x}-\lambda,1$ in the quotient $F[x]/(x-\lambda)^k$.
- Expanding each of these polynomials in $\overline{x}$, we see that the matrix relating these elements to the F-basis $\overline{x}^{k-1},\overline{x}^{k-2},\ldots,\overline{x},1$ of $F[x]/(x-\lambda)^k$ is upper triangular with 1's along the diagonal.
- Since this is an invertible matrix (having determinant 1), it follows that the elements above are an F-basis for $F[x]/(x-\lambda)^k$.

Jordan Block of Size k With Eigenvalue λ

- With respect to the basis $\overline{x}^{k-1}, \overline{x}^{k-2}, \ldots, \overline{x}, 1$ of $F[x]/(x-\lambda)^k$, the linear transformation of multiplication by x acts in a particularly simple manner.

  Note that $x = \lambda + (x-\lambda)$ and that $(\overline{x}-\lambda)^k = 0$ in the quotient.

  So we have:

$$\begin{array}{lcl}
(\overline{x}-\lambda)^{k-1} & \mapsto & \lambda \cdot (\overline{x}-\lambda)^{k-1} + (\overline{x}-\lambda)^k = \lambda(\overline{x}-\lambda)^{k-1}; \\
(\overline{x}-\lambda)^{k-2} & \mapsto & \lambda \cdot (\overline{x}-\lambda)^{k-2} + (\overline{x}-\lambda)^{k-1}; \\
 & \vdots & \\
\overline{x}-\lambda & \mapsto & \lambda \cdot (\overline{x}-\lambda) + (\overline{x}-\lambda)^2; \\
1 & \mapsto & \lambda \cdot 1 + (\overline{x}-\lambda).
\end{array}$$

Jordan Block of Size k With Eigenvalue λ (Cont'd)

- Thus, the matrix for multiplication by x is

$$\begin{pmatrix} \lambda & 1 & & & \\ & \lambda & \ddots & & \\ & & \ddots & 1 & \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix}$$

  where the blank entries are all zero.

Definition (Elementary Jordan Matrix or Jordan Block)

The $k \times k$ matrix with λ along the main diagonal and 1 along the first superdiagonal, depicted above, is called the $k \times k$ **elementary Jordan matrix with eigenvalue** λ or the **Jordan block of size** k **with eigenvalue** λ.

Jordan Canonical Form

- Applying this to each of the cyclic factors of V in its elementary divisor decomposition, we obtain a vector space basis for V with respect to which the linear transformation T has as matrix the direct sum of the Jordan blocks corresponding to the elementary divisors of V.
- I.e., it is block diagonal with Jordan blocks: $\begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_t \end{pmatrix}$.
- The matrix is uniquely determined up to permutation of the blocks along the diagonal by the elementary divisors of the $F[x]$-module V.
- Conversely, the list of elementary divisors uniquely determines the module V up to $F[x]$-module isomorphism.

Definition (Jordan Canonical Form)

(1) A matrix is said to be in **Jordan canonical form** if it is a block diagonal matrix with Jordan blocks along the diagonal.

(2) A **Jordan canonical form** for a linear transformation T is a matrix representing T which is in Jordan canonical form.

Jordan Canonical Form for Linear Transformations

- We have proved that any linear transformation T has a Jordan canonical form.
- It follows from the uniqueness of the elementary divisors that the Jordan canonical form is unique up to a permutation of the Jordan blocks along the diagonal.

Theorem (Jordan Canonical Form for Linear Transformations)

Let V be a finite dimensional vector space over the field F and let T be a linear transformation of V. Assume F contains all the eigenvalues of T.

(1) There is a basis for V with respect to which the matrix for T is in Jordan canonical form, i.e., is a block diagonal matrix whose diagonal blocks are the Jordan blocks for the elementary divisors of V.

(2) The Jordan canonical form for T is unique up to a permutation of the Jordan blocks along the diagonal.

Jordan Canonical Form for Matrices

Theorem (Jordan Canonical Form for Matrices)

Let A be an $n \times n$ matrix over the field F and assume F contains all the eigenvalues of A.

(1) The matrix A is similar to a matrix in Jordan canonical form, i.e., there is an invertible $n \times n$ matrix P over F, such that $P^{-1}AP$ is a block diagonal matrix whose diagonal blocks are the Jordan blocks for the elementary divisors of A.

(2) The Jordan canonical form for A is unique up to a permutation of the Jordan blocks along the diagonal.

- The Jordan canonical form for a matrix A is as close to being a diagonal matrix as possible.

Corollary

(1) If a matrix A is similar to a diagonal matrix D, then D is the Jordan canonical form of A.

(2) Two diagonal matrices are similar if and only if their diagonal entries are the same up to a permutation.

Diagonalizability Criterion

Corollary

If A is an $n \times n$ matrix with entries from F and F contains all the eigenvalues of A, then A is similar to a diagonal matrix over F if and only if the minimal polynomial of A has no repeated roots.

- Suppose A is similar to a diagonal matrix. The minimal polynomial of a diagonal matrix has no repeated roots (its roots are precisely the distinct elements along the diagonal). But similar matrices have the same minimal polynomial. It follows that the minimal polynomial for A has no repeated roots.

Diagonalizability Criterion (Sufficiency)

- Conversely, suppose the minimal polynomial for A has no repeated roots and let B be the Jordan canonical form of A. The matrix B is a block diagonal matrix with elementary Jordan matrices down the diagonal. The minimal polynomial for B is the least common multiple of the minimal polynomials of the Jordan blocks. It is easy to see directly that a Jordan block of size k with eigenvalue λ has minimal polynomial $(x - \lambda)^k$ (note that this is immediate from the fact that each elementary Jordan matrix gives the action on a cyclic $F[x]$-submodule whose annihilator is $(x - \lambda)^k$). Since A and B have the same minimal polynomial, the least common multiple of the $(x - \lambda)^k$ cannot have any repeated roots. It follows that k must be 1. Hence, each Jordan block must be of size one and B is a diagonal matrix.

From Invariant Factors to Elementary Divisors

- Assume that the field F contains all the eigenvalues of T (or A) so both the rational and Jordan canonical forms exist over F.
- The process of passing from one form to the other is exactly the same as for finite abelian groups.
- From invariant factors to elementary divisors:
  - Write each invariant factor as a product of distinct linear factors to powers;
  - The resulting set of powers of linear polynomials is the set of elementary divisors.

Example: If the invariant factors of T are

$$(x-1)(x-3)^3, (x-1)(x-2)(x-3)^3, (x-1)(x-2)^2(x-3)^3,$$

then the elementary divisors are

$$(x-1), (x-3)^3, (x-1), (x-2), (x-3)^3, (x-1), (x-2)^2, (x-3)^3.$$

From Elementary Divisors to Invariant Factors

- The largest invariant factor is the product of the largest of the distinct prime powers among the elementary divisors.
- The next largest invariant factor is the product of the largest of the distinct prime powers among the remaining elementary divisors, and so on.
- From elementary divisors to invariant factors:
  - Arrange the elementary divisors into n separate lists, one for each eigenvalue.

    In each list arrange the polynomials in increasing (i.e., nondecreasing) degree.
  - Arrange for all n lists to have the same length by appending an appropriate number of the constant polynomial 1.
  - Form the i-th invariant factor by taking the product of the i-th polynomial in each of these lists.

From Elementary Divisors to Invariant Factors: Example

- If the elementary divisors of T are

  $$(x-1)^3, (x+4), (x+4)^2, (x-5)^2, (x-1)^5, (x-1)^3, (x-5)^3, (x-1)^4, (x+4)^3.$$

  The intermediate lists are:

  $$\begin{array}{lllll} (1) & (x-1)^3 & (x-1)^3 & (x-1)^4 & (x-1)^5 \\ (2) & 1 & x+4 & (x+4)^2 & (x+4)^3 \\ (3) & 1 & 1 & (x-5)^2 & (x-5)^3 \end{array}$$

  So the list of invariant factors is:

  $$(x-1)^3, (x-1)^3(x+4), (x-1)^4(x+4)^2(x-5)^2, (x-1)^5(x+4)^3(x-5)^3.$$

Elementary Divisor Decomposition Algorithm (1)-(4)

(1-3) The first three steps in the algorithm are those from the Invariant Factor Decomposition Algorithm.

(4) For each invariant factor $a(x)$ computed for A, write

$$a(x) = (x-\lambda_1)^{\alpha_1}(x-\lambda_2)^{\alpha_2}\cdots(x-\lambda_s)^{\alpha_s},$$

where $\lambda_1, \ldots, \lambda_s \in F$ are distinct.
Let $f \in V$ be the $F[x]$-module generator for the cyclic factor corresponding to the invariant factor $a(x)$ computed in (3).
Then the elements

$$\frac{a(x)}{(x-\lambda_1)^{\alpha_1}}f, \frac{a(x)}{(x-\lambda_2)^{\alpha_2}}f, \ldots, \frac{a(x)}{(x-\lambda_s)^{\alpha_s}}f$$

(note that the $\frac{a(x)}{(x-\lambda_i)^{\alpha_i}} \in F[x]$ are polynomials) are $F[x]$-module generators for the cyclic factors of V corresponding to the elementary divisors

$$(x-\lambda_1)^{\alpha_1}, (x-\lambda_2)^{\alpha_2}, \ldots, (x-\lambda_s)^{\alpha_s}.$$

Elementary Divisor Decomposition Algorithm (5)-(6)

(5) If $g_i = \frac{a(x)}{(x-\lambda_i)^{\alpha_i}} f$ is the $F[x]$-module generator for the cyclic factor of V corresponding to the elementary divisor $(x-\lambda_i)^{\alpha_i}$, then the corresponding vector space basis for this cyclic factor of V is given by the elements

$$(T-\lambda_i)^{\alpha_i-1} g_i, (T-\lambda_i)^{\alpha_i-2} g_i, \ldots (T-\lambda_i) g_i, g_i.$$

(6) Write the k-th element of the vector space basis computed in (5) in terms of the original vector space basis $[e_1, e_2, \ldots, e_n]$ for V and use the coordinates for the k-th column of an $n \times n$ matrix P.

Then $P^{-1}AP$ is in Jordan canonical form (with Jordan blocks appearing in the order used in (5) for the cyclic factors of V).

Converting a Matrix to Jordan Canonical Form

(1-2) The first two steps are those from the algorithm for Converting an $n \times n$ matrix to Rational Canonical Form.

(3) When $xI - A$ has been diagonalized the first $n - m$ columns of the matrix P' are 0 and the remaining m columns of P' are nonzero. For each successive $i = 1, 2, \ldots, m$:

(a) Factor the i-th nonconstant diagonal element (which is of degree d_i):

$$a(x) = (x - \lambda_1)^{\alpha_1}(x - \lambda_2)^{\alpha_2} \cdots (x - \lambda_s)^{\alpha_s},$$

where $\lambda_1, \ldots, \lambda_s \in F$ are distinct (here $a(x) = a_i(x)$ is the i-th nonconstant diagonal element and s depends on i).

(b) Multiply the i-th nonzero column of P' successively by the d_i matrices:

$$\begin{gathered}(A - \lambda_1 I)^{\alpha_1 - 1}(A - \lambda_2 I)^{\alpha_2} \cdots (A - \lambda_s I)^{\alpha_s}\\ (A - \lambda_1 I)^{\alpha_1 - 2}(A - \lambda_2 I)^{\alpha_2} \cdots (A - \lambda_s I)^{\alpha_s}\\ \vdots\\ (A - \lambda_1 I)^{0}(A - \lambda_2 I)^{\alpha_2} \cdots (A - \lambda_s I)^{\alpha_s}\end{gathered}$$

Converting a Matrix to Jordan Canonical Form (Cont'd)

(3) (b)

$$\begin{gathered}
(A-\lambda_1 I)^{\alpha_1}(A-\lambda_2 I)^{\alpha_2-1}\cdots(A-\lambda_s I)^{\alpha_s}\\
(A-\lambda_1 I)^{\alpha_1}(A-\lambda_2 I)^{\alpha_2-2}\cdots(A-\lambda_s I)^{\alpha_s}\\
\vdots\\
(A-\lambda_1 I)^{\alpha_1}(A-\lambda_2 I)^{0}\cdots(A-\lambda_s I)^{\alpha_s}\\
\vdots(A-\lambda_1 I)^{\alpha_1}(A-\lambda_2 I)^{\alpha_2}\cdots(A-\lambda_s I)^{\alpha_s-1}\\
(A-\lambda_1 I)^{\alpha_1}(A-\lambda_2 I)^{\alpha_2}\cdots(A-\lambda_s I)^{\alpha_s-2}\\
\vdots\\
(A-\lambda_1 I)^{\alpha_1}(A-\lambda_2 I)^{\alpha_2}\cdots(A-\lambda_s I)^{0}
\end{gathered}$$

(c) Use the column vectors resulting from (b) (in that order) as the next d_i columns of an $n \times n$ matrix P.

Then $P^{-1}AP$ is in Jordan canonical form (whose Jordan blocks correspond to the ordering of the factors in (a)).

Example 1

- Let
$$A=\begin{pmatrix} 2 & -2 & 14 \\ 0 & 3 & -7 \\ 0 & 0 & 2 \end{pmatrix}, B=\begin{pmatrix} 0 & -4 & 85 \\ 1 & 4 & -30 \\ 0 & 0 & 3 \end{pmatrix}, C=\begin{pmatrix} 2 & 2 & 1 \\ 0 & 2 & -1 \\ 0 & 0 & 3 \end{pmatrix}$$
and let $F=\mathbb{Q}$.

  Note that $\mathbb{Q}$ contains all the eigenvalues for these matrices.

  We have already determined the invariant factors of these matrices:
$$A: x-2,(x-2)(x-3); \quad B:(x-2)^2(x-3); \quad C:(x-2)^2(x-3).$$
  The elementary divisors of A are $x-2, x-2$ and $x-3$; those of B and C are $(x-2)^2$ and $x-3$. So the respective Jordan canonical forms are: $\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$.

Example 2

- For the matrix $A = \begin{pmatrix} 2 & -2 & 14 \\ 0 & 3 & -7 \\ 0 & 0 & 2 \end{pmatrix}$ we determined that
$$f_1 = -7e_1 + 7e_2 + e_3 \quad \text{and} \quad f_2 = -e_1 + e_2$$
were $\mathbb{Q}[x]$-module generators for the two cyclic factors of V in its invariant factor decomposition, corresponding to the invariant factors $x-2$ and $(x-2)(x-3)$, respectively.
We apply the first algorithm to find the $\mathbb{Q}[x]$-module generators for the three cyclic factors of V in its elementary divisor decomposition, corresponding to the elementary divisors $x-2, x-2$, and $x-3$:
$$\begin{array}{rcl} f_1 & = & -7e_1 + 7e_2 + e_3; \\ (x-3)f_2 & = & (x-3)(-e_1+e_2) = -xe_1 + xe_2 + 3e_1 - 3e_2 \\ & = & -2e_1 - 2e_1 + 3e_2 + 3e_1 - 3e_2 = -e_1; \\ (x-2)f_2 & = & (x-2)(-e_1+e_2) = -xe_1 + xe_2 + 2e_1 - 2e_2 \\ & = & -2e_1 - 2e_1 + 3e_2 + 2e_1 - 2e_2 = -2e_1 + e_2. \end{array}$$

Example 2 (Cont'd)

- Then the matrix $P = \begin{pmatrix} -7 & -1 & -2 \\ 7 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$ conjugates A into its Jordan canonical form: $P^{-1}AP = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$. The columns of this matrix can also be obtained following the second algorithm above, using the nonzero columns of the matrix P' computed previously:

$$\begin{aligned}
(A-2I)^0 \begin{pmatrix} -7 \\ 7 \\ 1 \end{pmatrix} &= \begin{pmatrix} -7 \\ 7 \\ 1 \end{pmatrix}; \\
(A-2I)^0(A-3I)^1 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} &= \begin{pmatrix} -1 & -2 & 14 \\ 0 & 0 & -7 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}; \\
(A-2I)^1(A-3I)^0 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} &= \begin{pmatrix} 0 & -2 & -14 \\ 0 & 1 & -7 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}.
\end{aligned}$$

Example 3

- For the 4×4 matrix $D = \begin{pmatrix} 1 & 2 & -4 & 4 \\ 2 & -1 & 4 & -8 \\ 1 & 0 & 1 & -2 \\ 0 & 1 & -2 & 3 \end{pmatrix}$ the invariant factors were $(x-1)^2, (x-1)^2$, with corresponding $\mathbb{Q}[x]$-module generators $f_1 = e_1$ and $f_2 = e_2$, respectively.

  Hence, the elementary divisors are $(x-1)^2$, $(x-1)^2$.

  The corresponding vector space bases for these two factors are given by:

$$\begin{aligned}
(T-1)f_1 &= \begin{pmatrix} 1 & 2 & -4 & 4 \\ 2 & -1 & 4 & -8 \\ 1 & 0 & 1 & -2 \\ 0 & 1 & -2 & 3 \end{pmatrix} e_1 - e_1 = 2e_2 + e_3; \\
e_1; & \\
(T-1)f_2 &= \begin{pmatrix} 1 & 2 & -4 & 4 \\ 2 & -1 & 4 & -8 \\ 1 & 0 & 1 & -2 \\ 0 & 1 & -2 & 3 \end{pmatrix} e_2 - e_2 = 2e_1 - 2e_2 + e_4; \\
e_2. &
\end{aligned}$$

Example 3 (Cont'd)

- Thus, the matrix $P = \begin{pmatrix} 0 & 1 & 2 & 0 \\ 2 & 0 & -2 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$ conjugates D into its Jordan canonical form: $P^{-1}DP = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$.
- The columns of this matrix can also be obtained following the second algorithm above, using the nonzero columns of the matrix $P' = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ computed previously.

Example 3 (Cont'd)

- We calculate

$$\begin{aligned}
(D-I)^1\begin{pmatrix}1\\0\\0\\0\end{pmatrix} &= \begin{pmatrix}0&2&-4&-4\\2&-2&4&-8\\1&0&0&-2\\0&1&-2&2\end{pmatrix}\begin{pmatrix}1\\0\\0\\0\end{pmatrix}=\begin{pmatrix}0\\2\\1\\0\end{pmatrix};\\
(D-I)^0\begin{pmatrix}1\\0\\0\\0\end{pmatrix} &= \begin{pmatrix}1\\0\\0\\0\end{pmatrix};\\
(D-I)^1\begin{pmatrix}0\\1\\0\\0\end{pmatrix} &= \begin{pmatrix}0&2&-4&-4\\2&-2&4&-8\\1&0&0&-2\\0&1&-2&2\end{pmatrix}\begin{pmatrix}0\\1\\0\\0\end{pmatrix}=\begin{pmatrix}2\\-2\\0\\1\end{pmatrix};\\
(D-I)^0\begin{pmatrix}0\\1\\0\\0\end{pmatrix} &= \begin{pmatrix}0\\1\\0\\0\end{pmatrix}.
\end{aligned}$$

Example 4

- The set of similarity classes of 6×6 matrices with entries from $\mathbb{C}$ with characteristic polynomial $(x^4 - 1)(x^2 - 1)$ consists of the 4 classes represented by the rational canonical forms in the preceding set of examples (there are no additional lists of invariant factors over $\mathbb{C}$).

  Their Jordan canonical forms cannot all be written over $\mathbb{Q}$, however.

  Example: Suppose the invariant factors are

  $$(x-1)(x+1) \quad \text{and} \quad (x-1)(x+1)(x^2+1)$$

  Then the elementary divisors are:

  $$x-1, x+1, x-1, x+1, x-i, x+i,$$

  where i is a square root of -1 in $\mathbb{C}$.

  So the Jordan form for this matrix is a diagonal matrix with diagonal entries $1, 1, -1, -1, i, -i$.

Example 5

- The set of similarity classes of 3×3 matrices A over $\mathbb{C}$ satisfying $A^6 = I$ is considerably larger than that over $\mathbb{Q}$.

  If A is any such matrix, $m_A(x) \mid x^6 - 1$.

  Since the latter polynomial has no repeated roots in $\mathbb{C}$, the minimal polynomial of A has no repeated roots.

  The Jordan canonical form of A is a diagonal matrix.

  This diagonal matrix has the same minimal polynomial.

  Hence, its 6-th power is also the identity.

  So each diagonal entry is a 6-th root of unity:

  - For each list $\zeta_1, \zeta_2, \zeta_3$ of 6-th roots of unity we obtain a Jordan canonical form;
  - Two such forms are the same (i.e., give rise to similar matrices) if and only if the lists are permuted versions of each other.

  There are, up to similarity, $6 + 6 \cdot 5 + \frac{6 \cdot 5 \cdot 4}{6} = 56$ classes of such A's.