Cal Poly **Brussel**
Algebra

FINITELY GENERATED MODULES OVER A PID

Contents

We prove the fundamental theorem for finitely generated modules over a PID, and apply it to abelian groups and linear transformations. We give the algorithms for putting a given finitely generated abelian group into its canonical form as a direct sum of cyclic groups, for computing the rational canonical and Jordan forms of a linear transformation, and for finding the basis change matrices in each of these cases.

1. **Submodules of Free Modules Over a PID**

Aside from invariance of cardinality, an important property of vector spaces is that they are all free; therefore the subobjects are automatically vector spaces. This is not true for modules over a general commutative ring, or even over an arbitrary integral domain. But it holds over a PID.

Theorem 1.0.1. *Let R be a PID. If M is a free R-module of rank n and $K \subset M$ is a submodule, then K is free of rank $m \leq n$.*

Proof. If $K = 0$, K is free of rank 0, done. Assume $n \geq 1$ and $K \neq 0$. We induct on the rank n.

Base Case. If $n = 1$ then $K \simeq I$ for some nonzero ideal $I \subset R$. Since R is a domain, and R is a PID, K is free of rank 1 ✓.

Inductive Step. Assume $n > 1$ and the result holds for modules of rank $< n$. Set $M = R^n$, and let $M' = R^{n-1} \subset M$ be the first $n-1$ summands, which is the kernel of the map $\pi : M \longrightarrow R$ defined by $\pi(a_1 + \cdots + a_{n-1} + a_n) = a_n$. Then we have a short exact sequence

$$0 \longrightarrow M' \longrightarrow M \xrightarrow{\pi} M'' \longrightarrow 0$$

We “intersect” this sequence with K. Let $K'' = \pi(K) \subset R$, free of rank 0 or 1 by the base case, and let $K' = K \cap M'$, free of rank at most $n-1$ by the induction hypothesis. Then we have

$$0 \longrightarrow K' \longrightarrow K \xrightarrow{\pi} K'' \longrightarrow 0$$

Since K'' is free, $K \simeq K' \oplus K''$ by Hw7#4. If $K'' = (0)$ this shows K is free of rank at most $n-1$, as desired. If $K'' \neq (0)$, $\{k_1, \ldots, k_{m-1}\}$ is a basis for K', and $\{\pi(k_m)\}$ is a basis for K'', then let $\mathsf{B} = \{k_1, \ldots, k_m\}$. B is linearly independent: If $0 = \sum_{i=1}^{m} r_i k_i$ then $\pi(0) = 0 = r_m \pi(k_m)$ for some r_m, hence $r_m = 0$ since $\{\pi(k_m)\}$ is linearly independent. Then $0 = \sum_{i=1}^{m-1} r_i k_i$, hence $r_i = 0$ for all i since the first $m-1$ k_i's are linearly independent. Therefore B is linearly independent. B spans K: If $k \in K$ and $\pi(k) = r_m \pi(k_m)$, then $\pi(k - r_m k_m) = 0$, hence $k - r_m k_m \in K'$, so $k - r_m k_m = \sum_{i=1}^{m-1} r_i k_i$. Therefore $k = \sum_{i=1}^{m} r_i k_i$, which shows B spans K. We conclude B is a basis (of cardinality $m \leq n$), hence K is free of rank $m \leq n$. This completes the inductive step.

We conclude the result is true for all n by induction. □

Remark 1.0.2. The result is false if R is not a PID: If $R = \mathbb{R}[X, Y]$, $M = R$, and $K = (X, Y)$, then K is not free. For example, X and Y span K, but are not linearly independent, since $Y \cdot X - X \cdot Y = 0$. In fact, any two elements $p(X, Y)$, $q(X, Y) \in K$ are dependent, since $qp - pq = 0$. On the other hand, no single element spans K, since X and Y are already not R-multiples of a single element of K: R is a UFD, and X and Y are distinct primes, hence have no common divisor.

Or if $R = \mathbb{Z}/4$, $M = R$, the submodule $R \cdot 2 = \{0, 2\}$ is not free. For 2 is not linearly independent since $2 \cdot 2 = 0$, and there are no other choices for a spanning set!

2. **Smith Normal Form for Matrices Over a PID**

Let R be a commutative ring with 1. Any R-module endomorphism $R^n \longrightarrow R^n$ may be represented by a *matrix* $A \in \mathrm{M}_n(R)$, so that

$$\mathrm{M}_n(R) \simeq \mathrm{End}_R(R^n)$$

As usual, we put $\mathrm{GL}_n(R) = \mathrm{Aut}_R(R^n) = \mathrm{M}_n(R)^\times$.

Definition 2.0.1. We say two $m \times n$ matrices A and A' are *equivalent* if $\exists P \in \mathrm{GL}_m(R), Q \in \mathrm{GL}_n(R)$ such that $A' = PAQ$. This is clearly an equivalence relation. We write $A \sim A'$.

Picture:

$$\begin{tikzcd}
R^n \arrow[r, "A"] & R^m \arrow[d, "P"] \\
R^n \arrow[u, "Q"] \arrow[r, "A'"'] & R^m
\end{tikzcd}$$

If $m = n$ then equivalent matrices are *associates* in the ring $\mathrm{M}_n(R)$, so we can think of this property in $\mathrm{M}_{m\times n}(R)$ as a generalization of the associate property in ring theory.

2.1. **Elementary Row and Column Operations.**

Let e_{ij} denote the ij-th matrix unit in $\mathrm{M}_m(R)$, with a single 1 in the ij-th position. Elementary row operations on an $m \times n$ matrix are left multiplications by the following elements of $\mathrm{GL}_m(R)$

(1) $T_{ij}(b) = I + be_{ij}$ for $i \neq j$ $(R_i \mapsto R_i + bR_j)$
(2) $D_i(u) = I + (u-1)e_{ii}$ for $u \in R^\times$ $(R_i \mapsto uR_i)$
(3) $P_{ij} = I - e_{ii} - e_{jj} + e_{ij} + e_{ji}$ $(R_i \leftrightarrow R_j)$

Elementary column operations on an $m \times n$ matrix are right multiplications by the *transposes* of the corresponding matrices in $\mathrm{GL}_n(R)$: We get $C_i \mapsto C_i + bC_j$, $C_i \mapsto uC_i$, and $C_i \leftrightarrow C_j$, respectively.

Since elementary row and column operations have determinant ± 1 or a unit $u \in R^\times$, and the determinant is multiplicative, equivalent matrices have determinants that differ by units in R.

Definition 2.1.1. Let R be a PID, $A \in \mathrm{M}_{m\times n}(R)$. A *Smith normal form* of A is a diagonal matrix $D = \mathrm{diag}\{s_1, \dots, s_k, 0, \dots, 0\}$ such that $A \sim D$ and $s_i \mid s_j$ if $i < j$.

We first show every matrix has a Smith normal form, and then that the normal form is uniquely determined up to units.

Theorem 2.1.2 (Smith Normal Form)**.** *Suppose R is a PID and $A \in \mathrm{M}_{m\times n}(R)$. Then*

$$A \sim D := \mathrm{diag}\{s_1, \dots, s_k, 0, \dots, 0\} = \mathrm{diag}\{\underbrace{1, \dots, 1}_{k-r}, \underbrace{d_1, \dots, d_r}_{\text{nonunits}}, 0, \dots, 0\}$$

for some $k, r \in \mathbb{N} \cup \{0\}$ and $s_i \in R - \{0\}$ such that $s_i | s_j$ if $i < j$. If R is a Euclidean domain then the matrices $P \in \mathrm{GL}_m(R)$ and $Q \in \mathrm{GL}_n(R)$ such that $D = PAQ$ are products of elementary row and column operations, respectively.

Proof. Assume first that R is a Euclidean domain, which is an integral domain together with a degree function

$$\delta : R - \{0\} \longrightarrow \mathbb{N} \cup \{0\}$$

that satisfies

(1) $\delta(ab) \leq \delta(a) + \delta(b)$ for all $a, b \in R - \{0\}$.
(2) For all $a, b \in R$ with $b \neq 0$ there exist $q, r \in R$ satisfying $a = bq + r$, where either $r = 0$ or $\delta(r) < \delta(b)$.

Algorithm. We may assume $A \neq 0$, else done. If $a_{11} = 0$ then it is easy to show that using elementary row and column operations we may replace it by a nonzero element. Therefore assume $a_{11} \neq 0$.
Claim: If a_{11} does not divide every entry of A, then row/column ops can replace it with an element of smaller degree.
Prove claim: Assume a_{11} doesn't divide some a_{ij}. Either $i = 1$, $j = 1$, or i and j are not 1, and we take these cases in order. If a_{11} does not divide some a_{1j} we have $a_{1j} = a_{11}b_j + b_{11}$, with $\delta(b_{11}) < \delta(a_{11})$, by the Euclidean property, and by committing the column operation $T_{j1}(-b_j)^{\mathrm{t}}$: $C_j \mapsto C_j - b_jC_1$ we replace a_{1j} with b_{11}, and then P^{t}_{1j} replaces a_{11} with b_{11} ✓. Similarly if a_{11} does not divide some a_{i1}, we may replace it with an element of smaller degree using a row operation ✓. If a_{11} divides a_{1j} and a_{i1} but not a_{ij}, then we may replace a_{i1} with zero, which replaces a_{ij} with a_{ij} plus a multiple of a_{1j}, call it $a^{'}_{ij}$. Then using T_{i1} we replace a_{1j} with itself plus $a^{'}_{ij}$, producing an element in row 1 not divisible by a_{11}, and then we lower the degree of a_{11} as before ✓. This proves the claim.

Since the degree function takes values in $\mathbb{N} \cup \{0\}$, and $\delta(b_{11})$ is minimal among all values in $R - \{0\}$ if and only if it is a unit in R, the claim implies we may assume b_{11} divides every element of A. Then we may use row/column ops to zero out every other entry in R_1 and C_1, so that A is equivalent to

$$(*) \qquad \begin{bmatrix} b_{11} & 0 & \cdots & 0 \\ 0 & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ 0 & b_{m2} & \cdots & b_{mn} \end{bmatrix}$$

with $b_{11} \neq 0$ and $b_{11} \mid b_{kl}$. If b_{11} is a unit, we can make $b_{11} = 1$, using an elementary row op.

Continue with the $(m-1) \times (n-1)$ matrix in the southeast corner, noting that always b_{11} will divide everything. By induction, A is equivalent via row/column operations to a diagonal matrix of form $\mathrm{diag}\{s_1, \ldots, s_k, 0, \ldots, 0\}$, with nonzero entries s_i such that $s_i | s_j$ for $i < j$, ✓ . Since row operations are represented by elements of $\mathrm{GL}_m(R)$ and column operations by elements of $\mathrm{GL}_n(R)$, we have $Q \in \mathrm{GL}_n(R)$ and $P \in \mathrm{GL}_m(R)$ such that $D = PAQ$.

If R is not a Euclidean domain we cannot in general diagonalize A with row and column operations alone, and we modify the argument as follows. Instead of δ we use the *length* λ, defined to be the number of primes – with multiplicity – appearing in a (nonzero) prime factorization. Then $\lambda(u) = 0$ if and only if $u \in R^{\times}$, and then u divides everything. We make the analogous claim, that if a_{11} does not divide some a_{ij}, then we may replace it with an element of smaller length. If $a_{11} \nmid a_{1j}$, commit $C_2 \leftrightarrow C_j$ so that $a_{11} \nmid a_{12}$. Let $d = \gcd(a_{11}, a_{12})$, then $\lambda(d) < \lambda(a_{11})$. By

Bezout's Theorem there exist elements x, y such that $a_{11}x + a_{12}y = d$. Note we have used the fact that R is a PID. Put $s = a_{12}d^{-1}$, $t = -a_{11}d^{-1}$, and behold: $\begin{bmatrix} -t & s \\ y & -x \end{bmatrix} \cdot \begin{bmatrix} x & s \\ y & t \end{bmatrix} = I_2$. In particular we have an invertible matrix

$$\begin{bmatrix} x & s \\ y & t \end{bmatrix} \oplus I_{n-2}$$

Right multiplication on A gives a matrix whose first row is $\operatorname{diag}\{d, 0, b_{13}, \ldots, b_{1n}\}$, and $\lambda(d) < \lambda(a_{11})$. Similarly we can lower the length if $a_{11} \nmid a_{i1}$, and the rest of the proof follows the Euclidean case.

Finally, since $s_i \mid s_j$, any units s_i that occur are at the beginning, and left multiplication by $D_i(s_i^{-1})$ converts them to 1's, so that $\operatorname{diag}\{s_1, \ldots, s_q, 0, \ldots, 0\} \sim \operatorname{diag}\{1, \ldots, 1, d_1, \ldots, d_r, 0, \ldots, 0\}$ for nonuits $d_i := s_{k-r+i}$, for some r. □

Example 2.1.3. Let $R = \mathbb{Q}[x]$, a PID. We apply the algorithm to the matrix $\begin{bmatrix} x-2 & -1 \\ 1 & x-3 \end{bmatrix}$:

$$\begin{bmatrix} x-2 & -1 \\ 1 & x-3 \end{bmatrix} \longrightarrow \begin{bmatrix} -1 & x-2 \\ x-3 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & x-2 \\ 3-x & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & x-2 \\ 0 & (x-2)(x-3)+1 \end{bmatrix}$$
$$\longrightarrow \begin{bmatrix} 1 & 0 \\ 0 & x^2-5x+7 \end{bmatrix}$$

The operations are $C_1 \leftrightarrow C_2$, $C_1 \to -C_1$, $R_2 \to R_2 + (x-3)R_1$, and $C_2 \to C_2 + (2-x)C_1$. Thus

$$\begin{bmatrix} 1 & 0 \\ x-3 & 1 \end{bmatrix} \begin{bmatrix} x-2 & -1 \\ 1 & x-3 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & x-2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & x^2-5x+7 \end{bmatrix}$$

2.1.4. **Equivalent Matrices Have the Same Smith Normal Form.** We will show that each equivalence class of equivalent matrices has a uniquely determined representative in Smith normal form, up to units of course.

Definition 2.1.5. Suppose A has degree (m, n), and I and J are subsets of $\{1, \ldots, m\}$ and $\{1, \ldots, n\}$, respectively, such that $m - |I| = n - |J|$.

- The *submatrix* A_{IJ} is the square matrix obtained from A by deleting rows I and columns J.
- The *minor* M_{IJ} of A is $\det(A_{IJ})$.
- The *degree* of a minor M_{IJ} is $\deg(A_{IJ})$.
- The *rank* $\operatorname{rk}(A)$ of $A \in \mathrm{M}_{m\times n}(R)$ is the largest degree of a nonzero minor of A.

Theorem 2.1.6. *Suppose R is a PID, $A \in \mathrm{M}_{m\times n}(R)$, and $\operatorname{rk}(A) = k$. For each $1 \le i \le k$, let $\Delta_i = \Delta_i(A)$ be a gcd of the degree-i minors of A. Suppose*

$$A \sim \operatorname{diag}\{s_1, \ldots, s_k, 0, \ldots, 0\}$$

with $s_i | s_j$ for $i < j$ as in Theorem 2.1.2. Then Δ_i divides Δ_{i+1} for each i, and

$$s_1 \sim \Delta_1,\ s_2 \sim \Delta_2\Delta_1^{-1},\ \cdots, s_k \sim \Delta_k\Delta_{k-1}^{-1}$$

In particular, the s_i are uniquely determined up to units for the equivalence class of A.

Proof. Claim: $A \sim B$ implies $\Delta_i(A) \sim \Delta_i(B)$ $(i \leq k)$ (associates in R). If $P \in \mathrm{GL}_m(R)$ then the hi-entry of PA is $\sum_j p_{hj}a_{ji} \implies R_h(PA) = p_{h1}R_1(A) + p_{h2}R_2(A) + \cdots + p_{hm}R_m(A)$, i.e., the rows of PA are R-linear combinations of the rows of A. The determinant function is alternating and R-multilinear on rows (or columns) $\implies$ degree-i minors of PA are R-linear combinations of the degree-i minors of A. (Try degree-2 minors for $n = 3$.) In a PID, a linear combination of a set of elements is divisible by the gcd of the elements. Therefore $\Delta_i(A)$ divides each degree-i minor of PA, hence $\Delta_i(A)|\Delta_i(PA)$. Similarly if $Q \in \mathrm{GL}_n(R)$ then $\Delta_i(A)|\Delta_i(AQ)$. If $A \sim B$ then $\exists P \in \mathrm{GL}_m(R)$, $Q \in \mathrm{GL}_n(R)$ such that $B = PAQ$, $P^{-1}BQ^{-1} = A$, hence $\Delta_i(A)|\Delta_i(B)|\Delta_i(A)$. Therefore $\Delta_i(A) \sim \Delta_i(B)$ ✓ . Now in particular $A \sim \mathrm{diag}\{s_1, \ldots, s_k, 0, \ldots, 0\}$ by Theorem 2.1.2, and we compute $\Delta_i \sim s_1 \cdots s_i$ for $i : 1 \leq i \leq k$, by inspection. Successive solving for the s_i yields $s_1 \sim \Delta_1$, $s_2 \sim \Delta_2\Delta_1^{-1}, \ldots, s_k \sim \Delta_k\Delta_{k-1}^{-1}$ ✓. The final statement is immediate. □

Remark 2.1.7. This theorem gives an important "determinant shortcut" for computing the s_i, which as we will see are crucial for classifying finitely generated modules over a PID. Note also that if a matrix is similar to a diagonal matrix, the s_i give the diagonalized entries.

Notation. Let R be a PID. Since the Smith normal form of Definition 2.1.1 is uniquely determined (up to units), for any $A \in \mathrm{M}_{m\times n}(R)$ we put

$$\mathrm{SNF}(A) := \mathrm{diag}\{s_1, \ldots, s_k, 0, \ldots, 0\}$$

with nonzero s_i satisfying $s_i|s_j$ for $i < j$.

3. Structure of Finitely Generated Modules Over a PID

Let R be a PID, M a finitely generated R-module, on n generators. By the universal property of free modules, we have a surjective map $\pi : R^n \longrightarrow M$. By Theorem 1.0.1, the kernel K is free of rank $m \leq n$. Since K is finitely generated, we have a surjective R-linear map $L : R^n \longrightarrow K \leq R^n$. Putting these together yields an exact sequence

$$R^n \xrightarrow{L} R^n \xrightarrow{\pi} M \longrightarrow 0$$

where $L(R^n) = K$. We will use this setup to prove a structure theorem about M, by "diagonalizing L" so that M is realized as a coproduct of cyclic R-modules. This amounts to committing basis-change on each of the R^n's. Here's the general notation.

3.1. **Basis Change Notation.** To change bases we need good notation, the better to avoid the madness that comes with confusing "basis change" with "invertible linear transformation". For an invertible matrix can be used in two ways: To write a vector given in one basis in terms of another basis, and to relate two bases.

- Let $\mathbf{e}$, $\mathbf{e}'$ be bases for a free R-module R^n of rank n.
- If $v \in R^n$, write $[v]_{\mathbf{e}}$ for the expression $v = \sum_{i=1}^n v_ie_i = (v_i) \in R^n$.
- We say $[\mathrm{id}]^{\mathbf{e}}_{\mathbf{e}'} : [v]_{\mathbf{e}'} \longmapsto [v]_{\mathbf{e}}$ *writes* $[v]_{\mathbf{e}'}$ *in terms of* $\mathbf{e}$.
- Let $P = [\mathrm{id}]^{\mathbf{e}}_{\mathbf{e}'}$. Then P is in $\mathrm{GL}_n(R)$.
- The j-th column of P is $[e'_j]_{\mathbf{e}}$, i.e., $[e'_j]_{\mathbf{e}} = \sum_{i=1}^n a_{ij}e_i$ (for $P = (a_{ij})$).
- We have $\mathbf{e}' = \mathbf{e}P$, and $P^{\mathrm{t}}\mathbf{e}^{\mathrm{t}} = (\mathbf{e}')^{\mathrm{t}}$, meaning $e'_j = a_{1j}e_1 + \cdots + a_{nj}e_n$.
- We call $P^{-1} = [\mathrm{id}]^{\mathbf{e}'}_{\mathbf{e}} : [v]_{\mathbf{e}} \longmapsto [v]_{\mathbf{e}'}$ the *basis change matrix* (from $\mathbf{e}$ to $\mathbf{e}'$).

- If $L \in \mathrm{End}_R(R^n)$ we write $[L]^{\mathbf{e}}_{\mathbf{e}'} : [v]_{\mathbf{e}'} \to [L(v)]_{\mathbf{e}}$, and $P^{-1}[L]^{\mathbf{e}}_{\mathbf{e}}P = [L]^{\mathbf{e}'}_{\mathbf{e}'}$.

3.2. **Embedding Along the Diagonal.** Suppose $R^n = \coprod_{i=1}^n Re_i$ is the free R-module of rank n, on basis $\mathbf{e} = \{e_1, \dots, e_n\}$. Let $(d_i)e_i \subset Re_i$ be the submodule defined by principal ideals $(d_i) \subset R$. For the quotient we write

$$\frac{Re_i}{(d_i)e_i} = \frac{R}{(d_i)}e_i$$

The R-module homomorphism

$$\begin{aligned} \pi : \coprod_{i=1}^n Re_i &\longrightarrow \coprod_{i=1}^n \frac{R}{(d_i)}e_i \\ e_i &\longmapsto (1 + (d_i))e_i \end{aligned}$$

is surjective, with kernel $K = \{\sum_i m_i e_i \,:\, d_i \mid m_i \;\forall i\} = \coprod_i (d_i)e_i$. By First Isomorphism Theorem,

$$\frac{\coprod_i Re_i}{\coprod_i (d_i)e_i} \simeq \coprod_i \frac{R}{(d_i)}e_i$$

The direct sum expression of K is compatible with R^n in that its summands line up with those of R^n. We will say it is “direct-sum compatible”. Direct sum compatibility of a submodule makes the quotient easy to compute as a direct sum, so “the quotient of the coproducts is the coproduct of the quotients”. Thus if $A : \coprod_i Re_i \longrightarrow \coprod_i Re_i$ is the homomorphism taking e_i to $d_i e_i$, where $d_i \in R$, then $A(\coprod_i Re_i) = \coprod_i (d_i)e_i \leq \coprod_i Re_i$, and the quotient is $\coprod_i R/(d_i)e_i$. We say that A embeds R^n into R^n *along the diagonal*, because it is direct-sum compatible.

In general an R-module homomorphism $A : \coprod_i Re_i \longrightarrow \coprod_i Re_i$ is not along the diagonal, but expresses the image of each e_i as an R-linear combination of the e_i, and the quotient is then not such a simple coproduct. This is our starting point:

3.3. **Presentation Matrix.** Suppose given a surjective map

$$R^n \xrightarrow{\pi} M \longrightarrow 0$$

Let $\mathbf{e} = \{e_1, \dots, e_n\}$ be the standard basis for R^n, and let $K = \ker(\pi) \subset R^n$ the kernel, free on some basis $\underline{y} = \{y_1, \dots, y_m\}$, $m \leq n$. Let $\mathbf{f} = \{f_1, \dots, f_n\}$ be a standard basis for another copy of R^n, and define

$$\begin{aligned} L : R^n &\longrightarrow R^n \\ f_i &\longmapsto \begin{cases} y_i & \text{if } i \leq m \\ 0 & \text{if } i > m \end{cases} \end{aligned}$$

Suppose the expression of each y_j in terms of $\mathbf{e}$ is

$$y_j = \sum_{i=1}^n a_{ij} e_i$$

Set $a_{ij} = 0$ for $j > m$ and let $A = (a_{ij})$, so that $A = [L]^{\mathbf{e}}_{\mathbf{f}} : [R^n]_{\mathbf{f}} \longrightarrow [R^n]_{\mathbf{e}}$.

Definition 3.3.1. The transpose A^{t}, with i-th row $[y_i]_{\mathbf{e}}$, is the *presentation matrix* for M with respect to $\mathbf{e}$ and $\mathbf{f}$.

Remark 3.3.2. The presentation matrix for M depends on the basis. If $\mathbf{e}'$ and $\underline{y}', \mathbf{f}'$ are new bases for R^n and K, R^n, set $P = [\mathrm{id}]^{\mathbf{e}}_{\mathbf{e}'}$ and $Q = [\mathrm{id}]^{\mathbf{f}}_{\mathbf{f}'}$, then $A' = P^{-1}AQ = [L]^{\mathbf{e}'}_{\mathbf{f}'} : [R^n]_{\mathbf{f}'} \to [R^n]_{\mathbf{e}'}$. The new presentation matrix is $A'^{\mathrm{t}} = Q^{\mathrm{t}}A^{\mathrm{t}}P^{-\mathrm{t}}$.

When the embedding $A = [L]^{\mathbf{e}}_{\mathbf{f}}$ is along the diagonal, each f_i is assigned a multiple d_ie_i of e_i, for the same i, as opposed to a more general R-linear combination of the e_i, and by (3.2) the quotient is easy to compute. The fundamental structure theorem below shows us that we can always maneuver into this situation.

Theorem 3.3.3 (Fundamental Structure Theorem)**.** *Suppose R is a PID and M is a finitely generated R-module. Then M is a direct sum of cyclic modules:*

$$M \simeq \coprod_{i=1}^{r} R/(d_i) \oplus R^f$$

for a (possibly empty) set of uniquely determined nonzero nonunits d_i, called invariant factors*, such that $d_i|d_j$ if $i < j$, and a non-negative integer f, called the* rank*.*

Proof. The d_i and f will be shown to be uniquely determined by M in Theorem 3.6.1. Let $\mathbf{x} = x_1, \ldots, x_n$ be a set of generators for M. Let $\mathbf{e} = \{e_i\}$ be a basis for R^n, and define

$$\begin{aligned} \pi : R^n &\longrightarrow M \\ \sum_{i=1}^{n} r_ie_i &\longmapsto \sum_{i=1}^{n} r_ix_i \end{aligned}$$

Let $K = \ker(\pi) \subset R^n$, a free module of rank $m \le n$ by Theorem 1.0.1. Let $\underline{y} = \{y_1, \ldots, y_m\}$ be a basis, let $\mathbf{f} = \{f_1, \ldots, f_n\}$ be the standard basis for another copy of R^n, and define L by

$$\begin{tikzcd} R^n \arrow[r, "L"] & R^n \arrow[r, "\pi"] & M \arrow[r] & 0 \end{tikzcd}$$

where $L(f_i) = y_i$ for $i = 1, \ldots, m$, and $L(f_i) = 0$ for $i > m$. This sequence is exact. Let $A = [L]^{\mathbf{e}}_{\mathbf{f}} : [R^n]_{\mathbf{f}} \longrightarrow [R^n]_{\mathbf{e}}$. so A^{t} is a presentation matrix for M with respect to $\mathbf{e}$ and $\mathbf{f}$.

By Theorem 2.1.2 there exist $P, Q \in \mathrm{GL}_n(R)$ such that

$$P^{-1}AQ = \mathrm{SNF}(A) = \mathrm{diag}\{1, \ldots, 1, d_1, \ldots, d_r, 0, \ldots, 0\}$$

for nonzero nonunits d_i such that $d_i|d_j$ if $i < j$. Assume f zeros. The 1's represent redundancy in the x_i. Since $P, Q \in \mathrm{GL}_n(R)$, we have bases $\mathbf{e}'$, $\mathbf{f}'$ such that $P = [\mathrm{id}]^{\mathbf{e}}_{\mathbf{e}'}$ and $Q = [\mathrm{id}]^{\mathbf{f}}_{\mathbf{f}'}$ then $[L]^{\mathbf{e}'}_{\mathbf{f}'} = P^{-1}AQ : [R^n]_{\mathbf{f}'} \longrightarrow [R^n]_{\mathbf{e}'}$. This matrix takes f_i' to the corresponding multiple of e_i'. Now

$$M \simeq R^n/K \simeq \coprod_{i=1}^{r} R/(d_i) \ \oplus R^f$$

□

Example 3.3.4. Let R be a PID, and let $M = R^2/Rm$, where $m = ae_1 + be_2 \in R^2 = Re_1 \oplus Re_2$. The module Rm is free on basis element m, since R is an integral domain, and we have an exact sequence

$$\begin{tikzcd} Rf_1 \oplus Rf_2 \arrow[r, "L"] & Re_1 \oplus Re_2 \arrow[r, "\pi"] & M \arrow[r] & 0 \end{tikzcd}$$

where

$$A := [L]_{\mathbf{f}}^{\mathbf{e}} = A = \begin{bmatrix} a & 0 \\ b & 0 \end{bmatrix}$$

Let $d = \gcd(a, b) = ar + bs$, then by Theorem 2.1.6 we have $\mathrm{SNF}(A) = \mathrm{diag}\{d, 0\}$, hence

$$M \simeq \frac{R}{(d)} \oplus R$$

We can find the diagonalizing matrix by using elementary row and column operations to put A in Smith normal form. I get $P = \begin{bmatrix} a/d & -s \\ b/d & r \end{bmatrix}$. Check that

$$\mathrm{SNF}(A) = P^{-1}A = \begin{bmatrix} r & s \\ -b/d & a/d \end{bmatrix} \begin{bmatrix} a & 0 \\ b & 0 \end{bmatrix} = \begin{bmatrix} d & 0 \\ 0 & 0 \end{bmatrix}$$

As in the proof, $P^{-1} = [\mathrm{id}]_{\mathbf{e}}^{\mathbf{e}'}$, and $\mathbf{e}' = \mathbf{e}P$ is the new basis with respect to which M is diagonal. Explicitly, $e_1' = (a/d)e_1 + (b/d)e_2$ and $e_2' = -se_1 + re_2$, $P^{-1}A = [L]_{\mathbf{f}}^{\mathbf{e}'}$, and now

$$M = Re_1'/(d)e_1' \oplus Re_2'$$

The summands are generated by $\pi(e_1') = e_1' + (d)e_1'$ and $\pi(e_2') = e_2'$, respectively.

3.4. **Generators for the Diagonalized M.** Given a presentation for a finitely generated module M, we construct in Theorem 3.3.3 a decomposition of M as a coproduct of cyclic modules, by diagonalizing the presentation matrix for M. We can produce the generators of the cyclic direct summands as follows.

Corollary 3.4.1. *Suppose* $A^{\mathrm{t}} \in \mathrm{M}_n(R)$ *is a relations matrix defining* M *as a quotient of* R^n *with respect to* $\mathbf{x} = \{x_1, \ldots, x_n\} \subset M$*, and* $\mathrm{SNF}(A) = P^{-1}AQ$*. Then* $\mathbf{x}' = \mathbf{x}P$ *is the generating set of the cyclic direct summands of* M *in Structure Theorem 3.3.3.*

Proof. The map $\pi : R^n \longrightarrow M$ sends $\mathbf{e}$ to $\mathbf{x}$, hence it sends $\mathbf{e}' = \mathbf{e}P$ to $\mathbf{x}' = \mathbf{x}P$, since π is R-linear. The result follows since the e_i' generate the summands of R^n that form the diagonalized quotient. □

3.5. **Decomposition into** p**-Primary Components.** We prove a second version of Structure Theorem 3.3.3 that replaces the invariant factors with "elementary divisors". Recall that if R is a PID then it is a UFD, and it has prime elements.

Definition 3.5.1. Let R be a PID, and M an R-module.

- The *torsion submodule* of M is the set

$$M_{\mathrm{tor}} = \{m \in M : am = 0 \text{ some } 0 \neq a \in R\}$$

- For each prime $p \in R$, the *p-primary submodule* of M is the set

$$M_p = \{m \in M : p^e m = 0,\ \exists e \in \mathbb{N}\}$$

On the homework we show M_{tor} is a submodule using only that R is an integral domain. If M is finitely generated with invariant factors d_i, then by inspection $M_{\mathrm{tor}} \simeq \coprod_{i=1}^{r} R/(d_i)$. The submodule M_{tor} is defined intrinsically, and so it is a "characteristic submodule", meaning that it is invariant under any automorphism of M.

M_p is a submodule as long as we have primes with which to define it. On the homework we prove that a submodule of a finitely generated module over a PID is finitely generated, using Theorem 1.0.1. Therefore if M is finitely generated then so is M_p, and by Theorem 3.3.3, $M_p \simeq \coprod_{i=1}^{t} R/(p^{e_i})$ for numbers $e_i > 0$ such that $e_i \leq e_{i'}$ if $i < i'$.

Theorem 3.5.2 (Fundamental Structure Theorem$'$). *Suppose R is a PID and M is a finitely generated R-module. Let $\{d_1, \dots, d_r\}$ be the invariant factors of M, let $d_r = \prod_{j=1}^{t_r} p_j^{e_{rj}}$ be a prime factorization of d_r, and let $s_j = \inf\{i : p_j \mid d_i\}$. Then*

$$M \simeq \prod_{j=1}^{t_r} \underbrace{\left(\coprod_{i=s_j}^{r} R/(p_j^{e_{ij}}) \right)}_{M_{p_j}} \oplus R^f$$

for a uniquely determined $f \geq 0$, called the rank, and a (possibly empty) uniquely determined set of prime-powers $p_j^{e_{ij}}$, called elementary divisors*, satisfying $0 < e_{ij}$, $e_{ij} \leq e_{i'j}$ if $i < i'$.*

Proof. We have $M \simeq \coprod_{i=1}^{r} R/(d_i) \oplus R^f$ with $d_1 \mid \cdots \mid d_r$, all nonzero nonunits. Since R is a PID, it is a UFD, so each d_i factors into prime powers in R. Since $d_i \mid d_{i'}$ for $i < i'$, we may order the primes p_j so that those appearing in d_i are $\{p_1, \dots, p_{t_i}\}$, with $t_i \leq t_{i'}$ for $i < i'$, and $d_i \sim \prod_{j=1}^{t_i} p_j^{e_{ij}}$ for unique $e_{ij} > 0$ satisfying $e_{ij} \leq e_{i'j}$ for $i < i'$. By the Chinese Remainder Theorem,

$$R/(d_i) \simeq \prod_{j=1}^{t_i} R/(p_j^{e_{ij}})$$

Now $M \simeq \coprod_{i=1}^{r} \prod_{j=1}^{t_i} R/(p_j^{e_{ij}})$. Consolidating the p_j's leads to the stated expression, with the product of the p_j-primary components forming M_{p_j}. We prove uniqueness in Theorem 3.6.1 below. □

3.6. **Uniqueness of the Invariant Factors.** A different set of generators of M produces a different relations matrix, and it is unclear *a priori* that different relations matrices even have the same size, let alone are equivalent, hence that the d_i's in Structure Theorem 3.3.3 are uniquely determined. The proof that they are is hard, and seems to require the use of primes.

Theorem 3.6.1 (Uniqueness). *Let R be a PID, and let M be a finitely generated R-module. Then the invariant factors d_i of Structure Theorem 3.3.3 and the elementary divisors $p_j^{e_{ij}}$ of Structure Theorem 3.5.2 are uniquely determined up to units, and the rank f in each case is uniquely determined.*

Proof. Let $M \simeq \coprod_{i=1}^{r} R/(d_i) \oplus R^f$ be the decomposition of Theorem 3.3.3. Since $M_{\text{tor}} \simeq \coprod_{i=1}^{r} R/(d_i)$, the quotient M/M_{tor} is isomorphic to R^f, hence f is uniquely determined by the invariance of rank. This reduces the theorem to proving uniqueness for $M = M_{\text{tor}}$.

Assume $M = M_{\text{tor}}$. Let $\{p_j : 1 \leq j \leq t_r\}$ be the set of prime divisors of d_r. By the Structure Theorem 3.3.3, $M_{p_j} \simeq \coprod_{i=s_j}^{r} R/(p_j^{e_{ij}})$ for $e_{ij} \in \mathbb{N}$ satisfying $e_{ij} \leq e_{i'j}$ for $i < i'$. We will show these invariants are uniquely determined. To simplify notation, put $p = p_j$, $e_i = e_{ij}$, and $s = s_j$. We have a filtration $M_p \supset pM_p \supset \cdots \supset p^{e_t} M_p = (0)$, whose factors $M_p^{(k)} := p^k M_p / p^{k+1} M_p$ are $R/(p)$-modules for $k = 0, \dots, e_t - 1$. Since R is a PID and p is prime, (p) is a maximal ideal, so $F = R/(p)$ is a field, hence $M_p^{(k)}$ is an F-vector space, whose dimension $\dim_F M_p^{(k)}$ equals the number of e_i greater than k in the decomposition $M_p = \coprod_{i=1}^{s} R/(p^{e_i})$. This comes from the third isomorphism theorem: $p^k M_p/p^{k+1}M_p$ has as many summands as there are nonzero $p^k R/(p^{e_i})$, ✓. In particular, $\dim_F(M_p/pM_p) = t$, the number of e_i equal to at least 1, $\dim_F(pM_p/p^2M_p)$ is the number of e_i equal to at least 2, and so on. Thus t, e_i, and their multiplicities are uniquely determined by this canonical filtration, as desired.

To finish the proof we observe that in the notation of Theorem 3.5.2, $M_{p_j} \simeq \coprod_{i=s_j}^{r} R/(p_j^{e_{ij}})$, so the exponents e_{ij} of the elementary divisors are uniquely determined, so the elementary divisors $p_j^{e_{ij}}$ are uniquely determined up to units, and since the d_i's are uniquely products of the $p_j^{e_{ij}}$'s up to units, the d_i's are uniquely determined up to units. □

4. Application to Abelian Groups.

Suppose given an abelian group G with a set of relations $K = 0$, which express G as a quotient of $\mathbb{Z}^n$ by the image K of a linear map $A : \mathbb{Z}^n \longrightarrow \mathbb{Z}^n$. As noted above, the map A may or may not be along the diagonal.

4.1. Explicit Examples.

Example 4.1.1. See Example 3.3.4. What is $G = (\mathbb{Z}e_1 \oplus \mathbb{Z}e_2)/\mathbb{Z}(4e_1 + 6e_2)$ as a coproduct of cyclic groups? The quotient is defined by the exact sequence

$$\begin{tikzcd}\mathbb{Z}f_1 \oplus \mathbb{Z}f_2 \arrow[r, "\left[\begin{smallmatrix}4&0\\6&0\end{smallmatrix}\right]"] & \mathbb{Z}e_1 \oplus \mathbb{Z}e_2 \arrow[r, "\pi"] & G \arrow[r] & 0\end{tikzcd}$$

The matrix has a single nonzero invariant factor $\gcd(4,6) = 2$, so we know $G \simeq \mathbb{Z}_2 \oplus \mathbb{Z}$. Explicitly, the row operations $T_{21}(-2)P_{12}T_{21}(-1) = \left[\begin{smallmatrix}-1&1\\3&-2\end{smallmatrix}\right]$ diagonalize this matrix:

$$\mathrm{SNF}(A) = \begin{bmatrix}-1 & 1\\ 3 & -2\end{bmatrix}\begin{bmatrix}4 & 0\\ 6 & 0\end{bmatrix} = \begin{bmatrix}2 & 0\\ 0 & 0\end{bmatrix} \checkmark$$

The transpose of $\mathrm{SNF}(A)$ is another presentation matrix for G, with respect to a different set of generators. We conclude

$$G = \frac{\mathbb{Z}e_1 \oplus \mathbb{Z}e_2}{\mathbb{Z}(4e_1 + 6e_2)} \simeq \mathbb{Z}_2 \oplus \mathbb{Z}$$

Let's find the generators of G that give us this decomposition into cyclic groups. The original generators are $e_1 + K$ and $e_2 + K$, where $K = \mathbb{Z}(4e_1 + 6e_2)$. $P^{-1}A$ is a new presentation matrix with respect to $\mathbf{f}$ and a basis $\mathbf{e}'$ given by $P^{-1} = [\mathrm{id}]_{\mathbf{e}}^{\mathbf{e}'}$. Since $\mathbf{e}' = \mathbf{e}P$,

$$\begin{bmatrix}e_1' & e_2'\end{bmatrix} = \begin{bmatrix}e_1 & e_2\end{bmatrix}\begin{bmatrix}2 & 1\\ 3 & 1\end{bmatrix} = \begin{bmatrix}2e_1 + 3e_2, e_1 + e_2\end{bmatrix}$$

Note that indeed $\mathbb{Z}e_1 \oplus \mathbb{Z}e_2 = \mathbb{Z}(2e_1 + 3e_2) \oplus \mathbb{Z}(e_1 + e_2)$, so $\mathbf{e}'$ is a basis. The new basis gives us a new presentation

$$\begin{tikzcd}\mathbb{Z}f_1 \oplus \mathbb{Z}f_2 \arrow[r, "\left[\begin{smallmatrix}2&0\\0&0\end{smallmatrix}\right]"] & \mathbb{Z}e_1' \oplus \mathbb{Z}e_2' \arrow[r, "\pi"] & G \arrow[r] & 0\end{tikzcd}$$

In these terms, $K = \mathbb{Z}(2e_1')$, $e_1' + K = (2e_1 + 3e_2) + K$ has order 2, and $e_2' + K = (e_1 + e_2) + K$ has infinite order:

$$G = \frac{\mathbb{Z}e_1' \oplus \mathbb{Z}e_2'}{\mathbb{Z}(2e_1')} \simeq \frac{\mathbb{Z}e_1'}{\mathbb{Z}(2e_1')} \oplus \mathbb{Z}e_2' = \mathbb{Z}_2 \oplus \mathbb{Z}$$

Geometrically, G is a cylinder of discrete points, closed under addition, with a single point p_0 of order 2, and a line $\mathbb{Z}(e_1 + e_2)$ winding around the cylinder, combining with p_0 to eventually run through every point. We can replace $e_1 + e_2$ with any element of the coset $e_1 + e_2 + \mathbb{Z}(2e_1 + 3e_2)$; they all have infinite order. This reflects the fact that in the gcd-identity $4r + 6s = 2$, a complete list of choices for r and s is given by the coefficients of e_1 and e_2 for any element of the coset $-e_1 + e_2 + \mathbb{Z}(3e_1 - 2e_2)$, and the top row of the matrix P^{-1} above is then be replaced by $\begin{bmatrix} r & s \end{bmatrix}$, leading to the new generator.

Geometric summary: Let $\ell = \mathbb{Z}(2e_1 + 3e_2)$, the line-subgroup containing $4e_1 + 6e_2$, and let $H = \mathbb{Z}(4e_1 + 6e_2) \le \ell$. Then $\ell/H \simeq \mathbb{Z}_2$, and the quotient is $\ell/H \oplus \ell'$, where $\ell' \simeq \mathbb{Z}$ is some line. We generalize based on this example: Any subgroup $H = \mathbb{Z}(ae_1 + be_2)$ lies on a line $\ell \le \mathbb{Z}e_1 \oplus \mathbb{Z}e_2$, and ℓ/H is a cyclic group of order $d = \gcd(a, b)$, and $(\mathbb{Z}e_1 \oplus \mathbb{Z}e_2)/\mathbb{Z}(ae_1 + be_2)$ will be $\ell/H \oplus \ell'$, isomorphic to $\mathbb{Z}_d \oplus \mathbb{Z}$. Note that the subgroup $\mathbb{Z}$ has d cosets in $\mathbb{Z}_d \oplus \mathbb{Z}$, so on the cylinder there will be d parallel lines.

Example 4.1.2. Determine the direct sum decomposition of the abelian group

$$G = \langle a, b, c : 3a + 9b + 9c = 9a - 3b + 9c = 0 \rangle$$

Since G has three generators and two relations, we may write $G = \mathbb{Z}^3/H$, where $H = \langle y_1, y_2 \rangle$ is the image of the map $L : \coprod_i \mathbb{Z}f_i \longrightarrow \coprod_i \mathbb{Z}e_i$ given by $L(f_1) = y_1 = 3e_1 + 9e_2 + 9e_3$, $L(f_2) = y_2 = 9e_1 - 3e_2 + 9e_3$, and $L(f_3) = 0$. Thus we have an exact sequence

$$\begin{tikzcd} \mathbb{Z}^3 \arrow[r, "L"] & \mathbb{Z}^3 \arrow[r] & G \arrow[r] & 0 \end{tikzcd}$$

and if $A = [L]^{\mathbf{e}}_{\mathbf{f}}$ then the presentation matrix is

$$A^{\mathrm{t}} = \begin{bmatrix} 3 & 9 & 9 \\ 9 & -3 & 9 \\ 0 & 0 & 0 \end{bmatrix}$$

We can read off the invariant factors by Theorem 2.1.6: $d_1 = 3$, $d_2 = \gcd(108, 54, 90)/3 = 6$. Therefore

$$G \simeq \mathbb{Z}_3 \oplus \mathbb{Z}_6 \oplus \mathbb{Z}$$

Explicitly, since $A \sim \operatorname{diag}\{3, 6, 0\}$, we have $P^{-1}AQ = \operatorname{diag}\{3, 6, 0\}$ for some $P, Q \in \mathrm{GL}_3(\mathbb{Z})$. Let $\mathbf{e}' = \mathbf{e}P$ and $\mathbf{f}' = \mathbf{f}Q$ be the new bases defined by P and Q. Then $\operatorname{diag}\{3, 6, 0\}$ embeds $\coprod_i \mathbb{Z}f_i'$ into $\coprod_i \mathbb{Z}e_i'$ along the diagonal. To compute the generators in terms of the a, b, c, we determine P. Commit row reduction on A (not A^{t}) $R_2 \mapsto R_2 - 3R_1$; $R_3 \mapsto R_3 - 3R_1$; $C_2 \mapsto C_2 - 3C_1$; $R_2 \mapsto R_2 - 2R_3$; $R_3 \mapsto R_3 + 3R_2$. Thus $\operatorname{diag}\{3, 6, 0\} = T_{32}(3)T_{23}(-2)T_{31}(-3)T_{21}(-3)AT_{21}(-3)^{\mathrm{t}}$ ✓ . Thus $P^{-1} = T_{32}(3)T_{23}(-2)T_{31}(-3)T_{21}(-3)$, so

$$P = T_{21}(3)T_{31}(3)T_{23}(2)T_{32}(-3) = \begin{bmatrix} 1 & 0 & 0 \\ 3 & -5 & 2 \\ 3 & -3 & 1 \end{bmatrix}$$

Since $\begin{bmatrix} a' & b' & c' \end{bmatrix} = \begin{bmatrix} a & b & c \end{bmatrix} P$ by Corollary 3.4.1, we find $a' = a + 3b + 3c$, $b' = -5b - 3c$, $c' = 2b + c$, and

$$G = \left\langle a', b', c' : |a'| = 3, |b'| = 6 \right\rangle$$

4.2. **Volume of a Lattice Quotient.**

Theorem 4.2.1. *Suppose G is a free abelian group of rank n, and $\imath : H \to G$ is a subgroup. Then G/H is finite if and only if* $\operatorname{rk}(H) = n$*, and if* $\mathbf{x}$ *and* $\underline{y}$ *are bases for G and H, respectively, and $A = [\imath]_{\underline{y}}^{\mathbf{x}}$, then* $|G/H| = \det(A)$.

Proof. The first part is clear: We have an exact sequence

$$\mathbb{Z}^n \xrightarrow{L} \mathbb{Z}^n \xrightarrow{\pi} G/H \longrightarrow 0$$

where π is the projection and $L(\mathbb{Z}^n) = H \leq G = \mathbb{Z}^n$. The quotient is finite if and only if $\operatorname{rk}(L) = n$, which is equivalent to $\operatorname{rk}(H) = n$. Then note that $G/H = \coprod_{i=1}^r Z/(d_i)$, so $|G/H| = \prod_{i=1}^r d_i$, and we just have to compute the d_i. But then $|G/H| = \Delta_r = \det(A)$. Nice result! □

Example 4.2.2. Apply to lattices in $\mathbb{R}^n$. By definition a (full) lattice in $\mathbb{R}^n$ is a free $\mathbb{Z}$-module of rank n. Thus a full lattice is isomorphic to $\mathbb{Z}^n$. If $\mathbb{Z}^n$ is the standard lattice in $\mathbb{R}^n$ and $L : \mathbb{Z}^n \to \Lambda \leq \mathbb{Z}^n$ is a full sublattice, then $|\mathbb{Z}^n/\Lambda| = \det(L)$ computes the volume of the n-tope formed by a basis $\mathbf{e} = \{e_1, \ldots, e_n\}$ for Λ. For example, the 2-tope formed by $\langle 1, 2\rangle$ and $\langle -1, 2\rangle$ has volume 4.

5. **Application to Linear Transformations**

Let k be a field, $k[x]$ the polynomial ring, V an n-dimensional k-vector space, and $T \in \operatorname{End}_k(V)$. We make V into an $k[x]$-module under the rule

$$\begin{aligned} k[x] \times V &\longrightarrow V \\ (f(x), v) &\longmapsto f(T)(v) \end{aligned}$$

Let V_T denote the $k[x]$-module V associated to T. Note that $V_T = V$ as k-vector spaces, but V_T has the additional structure of scalar multiplication in the bigger ring $k[x]$.

Example 5.0.1. Let $V = k^2$ with standard basis $\mathbf{v} = \{v_1, v_2\}$, and let

$$A = [T]_{\mathbf{v}} = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}$$

We compute $x \cdot v_1 = 2v_1 - v_2$, $x^2 \cdot v_1 = 3v_1 - 5v_2$, and $x^3 \cdot v_1 = v_1 - 18v_2$. To apply polynomials in x to elements of V, we just add, for example

$$(x^3 - 2x + 1) \cdot v_1 = -2v_1 - 16v_2$$

The basis $\mathbf{v}$ shows V is finitely generated as a k-module, since any $v \in V$ is a unique k-linear combination of the v_i. This means V_T is finitely generated as a $k[x]$-module as well, since any $v \in V_T$ is, of course, in V, so it is trivially also a $k[x]$-linear combination of the v_i. But $\mathbf{v}$ is not a $k[x]$-basis for V_T. In fact, every $v \in V_T$ is $k[x]$-linearly dependent, because V_T is torsion: Consider the polynomial $p(x) = x^2 - \operatorname{t}(A)x + \det(A) = x^2 - 5x + 7$. This is a nonzero polynomial, but $p(x) \cdot v = p(A) \cdot v = 0$, since

$$p(A) = A^2 - 5A + 7I = \begin{bmatrix} 3 & 5 \\ -5 & 8 \end{bmatrix} - 5\begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix} + \begin{bmatrix} 7 & 0 \\ 0 & 7 \end{bmatrix} = \begin{bmatrix} 3 & 5 \\ -5 & 8 \end{bmatrix} - \begin{bmatrix} 10 & 5 \\ -5 & 15 \end{bmatrix} + \begin{bmatrix} 7 & 0 \\ 0 & 7 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

In fact $p(x)$ is A's characteristic polynomial, defined below.

As noted in Example 5.0.1, since V_T is a finitely generated $k[x]$-module, it is already finitely generated as a k-module. Therefore Structure Theorem 3.3.3 applies in this situation, and we can write V_T as a direct sum of cyclic $k[x]$ modules. Moreover, we observe that the rank f is 0, since V_T has finite k-dimension, hence V_T is a torsion $k[x]$-module, and by Structure Theorem 3.3.3,

$$V_T \simeq \coprod_{i=1}^{r} \frac{k[x]}{(d_i(x))}$$

for uniquely determined nonconstant monic polynomials $d_i(x) \in k[x]$, satisfying $d_i \,|\, d_j$ for $i < j$.

Our goal now is to use this description of V_T to prove things about matrices representing T. For example, recall by Linear Algebra I that T is diagonalizable if and only if V_T has a basis of eigenvectors v_i, and then $V_T = \coprod_{i=1}^{n} k \cdot v_i$. An eigenvector is a $v \in V$ such that $x \cdot v = T(v) = \lambda v$ for some $\lambda \in k$. Therefore T is diagonalizable if and only if we have a $k[x]$-isomorphism $V_T \simeq \coprod_{i=1}^{n} \frac{k[x]}{(x-\lambda_i)}$. By Structure Theorem 3.3.3, there is a $k[x]$-module isomorphism

$$\coprod_{i=1}^{n} \frac{k[x]}{(x-\lambda_i)} \simeq \coprod_{i=1}^{r} \frac{k[x]}{(d_i(x))}$$

We will show that this is equivalent to $d_r(x)$ splitting into distinct linear factors.

5.1. Quotients of k[x].

All ideals of $k[x]$ have form $(f(x))$ for some $f(x) \in k[x]$. We summarize some general observations about the quotient in the following theorem.

Theorem 5.1.1 (Kronecker)**.** *Let k be a field, $k[x]$ the polynomial ring, and $f \in k[x]$ a monic polynomial of degree $n \geq 1$. The quotient ring $k[x]/(f)$ is a k-vector space of dimension n, with k-basis $\{1, \bar{x}, \dots, \bar{x}^{n-1}\}$, where $\bar{x} = x + (f)$. The element $\bar{x}$ is a root of $f(x)$ in $k[x]/(f)$.*

Proof. The quotient $k[x]/(f)$ contains the field k as a subset via the map $a \mapsto a + (f)$. Therefore $k[x]/(f)$ is a k-vector space. Let $\bar{x} = x + (f)$. Then $f(\bar{x}) = 0$ in $k[x]/(f)$, and $\bar{x}^n$ is in the k-span of $\{1, \bar{x}, \dots, \bar{x}^{n-1}\}$: if $f(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + x^n$, then $\bar{x}^n = -(a_0 + a_1\bar{x} + \cdots + a_{n-1}\bar{x}^{n-1})$. It follows that this set spans $k[x]/(f)$ over k. It is linearly independent: $c_0 + c_1\bar{x} + \cdots + c_{n-1}\bar{x}^{n-1} = 0$ implies $f(x) \mid c_0 + c_1 x + \cdots + c_{n-1}x^{n-1}$, and since $k[x]$ is a Euclidean domain, this forces $c_0 + c_1 x + \cdots + c_{n-1}x^{n-1} = 0$, hence $c_i = 0$ for all i. Therefore $\{1, \bar{x}, \dots, \bar{x}^{n-1}\}$ is a k-basis. □

5.2. Polynomials Associated to a Linear Transformation.

Definition 5.2.1. Let V be an n-dimensional k-vector space, $T \in \mathrm{End}_k(V)$, and let $d_1, d_2, \dots, d_r$ be a complete list of the nonconstant monic polynomial invariants of V_T in $k[x]$, with $d_i | d_j$ for $i < j$, as in Structure Theorem 3.3.3. Then

(a) The *minimum polynomial* $m_T(x)$ of T is the invariant $d_r(x)$.
(b) The *characteristic polynomial* $p_T(x)$ of T is the product $\prod_{i=1}^{r} d_i(x)$.

Remark 5.2.2. (i) Since $m_T(x) \cdot V_T = 0$ and $p_T(x) \cdot V_T = 0$, the linear transformations $m_T(T)$ and $p_T(T)$ both equal zero. Once we have shown that $p_T(x)$ is the usual characteristic polynomial, the latter result will become the *Cayley-Hamilton Theorem*: A linear transformation is a root (in $\mathrm{End}_k(V)$) of its characteristic polynomial.

(ii) Since $V_T = \coprod_{i=1}^r k[x]/(d_i)$, and $d_i | d_j$ for $i < j$, $m_T(x)$ is the polynomial $f(x)$ of smallest degree such that $f(T) = 0$. Since $k[x]$ is a PID, the set $\mathrm{Ann}_{k[x]}(V_T)$ is a principal ideal containing $m_T(x)$, and since $m_T(X)$ has smallest degree, it is the generator.

(iii) We know that $k[x]/(d_i)$ is a k-vector space with basis $\{1, x, x^2, \dots, x^{\deg(d_i)-1}\}$. Therefore $\dim_k(k[x]/(d_i)) = \deg(d_i)$. Since $\dim_k(V) = n = \sum_i \deg(d_i)$, the degree of the characteristic polynomial is $\deg(p_T(x)) = \dim_k(V) = n$.

5.3. **Computation of Invariants of a Linear Transformation.** Let V be an n-dimensional k-vector space, $T \in \mathrm{End}_k(V)$. We would like to compute $m_T(x)$ and $p_T(x)$, and all of the invariants in between. To do this we need a presentation matrix for V_T, so we can apply Theorem 2.1.6. Recall we have an exact sequence

$$k[x]^n \xrightarrow{L} k[x]^n \xrightarrow{\pi} V_T \longrightarrow 0$$

where $L \in \mathrm{End}_{k[x]}(k[x]^n) = \mathrm{M}_n(k[x])$. Let $\mathbf{e}$ be the preimage of a basis $\mathbf{v}$ for V, and extend T to $k[x]^n$ by setting $[T]_\mathbf{e} = [T]_\mathbf{v} = (a_{ij}) \in \mathrm{M}_n(k)$. We'll show $L = x - T$, so

$$[L]_\mathbf{e} = xI_n - (a_{ij}) = \begin{bmatrix} x - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & x - a_{22} & \cdots & \vdots \\ \vdots & & \ddots & -a_{n-1\,n} \\ -a_{n1} & \cdots & -a_{nn-1} & x - a_{nn} \end{bmatrix}$$

Note π, T, and multiplication by x are all $k[x]$-linear. Therefore $\pi \circ (x - T)k[x]^n = 0$, so $x - T = 0$ is a relation defining V_T from $k[x]^n$. This means that V_T is described by the generators and relations

$$V_T = \langle e_1, \dots, e_n \ : \ xe_i = T(e_i) \ : \ 1 \le i \le n \rangle$$

where $T(e_i)$ is a k-linear combination of the other e_j. We'll prove the highly plausible fact that these are the *only* relations, meaning that if $\pi(\sum_i f_i(x)e_i) = 0$ for some $f_i(x) \in k[x]$, then $\sum_i f_i(x)e_i$ is contained in the submodule $(x - T)(k[x]^n) \subset k[x]^n$.

Example 5.3.1. Let $T \in \mathrm{End}_k(k^2)$ be given in the standard basis $\mathbf{v}$ by $T(v_1) = 2v_1 - v_2$ and $T(v_2) = v_1 + 3v_2$, as in Example 5.0.1. Let $\mathbf{e}$ be the standard basis for $k[x]^2$, and define $\pi : k[x]^2 \longrightarrow k^2$ by $\pi(e_i) = v_i$. We show that $(x - T)(k[x]^2)$ is contained in $\ker(\pi)$. We compute $(x - T)(e_1) = xe_1 - T(e_1) = xe_1 - (2e_1 - e_2) = (x-2)e_1 + e_2$, and similarly $(x - T)(e_2) = -e_1 + (x-3)e_2$. Since $\pi((x-2)e_1 + e_2) = (T - 2)(v_1) + v_2 = 0$ and $\pi(-e_1 + (x-3)e_2) = -v_1 + (T-3)v_2 = 0$, $(x - T)(k^2) \subset \ker(\pi)$. If these are the only relations then

$$V_T = \langle e_1, e_2 \ : \ xe_1 = 2e_1 - e_2,\ xe_2 = e_1 + 3e_2 \rangle$$

Notice we are just reading off the columns of the (transpose of) the presentation matrix matrix

$$[x - T]_\mathbf{e} = xI - A = \begin{bmatrix} x-2 & -1 \\ 1 & x-3 \end{bmatrix}$$

Theorem 5.3.2. *Let V be an n-dimensional k-vector space, made into a $k[x]$-module V_T via $T \in \mathrm{End}_k(V)$. Extend T from $\mathrm{End}_k(V)$ to $\mathrm{End}_{k[x]}(k[x]^n)$ as above, with $\mathbf{e}$ the standard basis for $k[x]^n$. Then the sequence*

$$0 \longrightarrow k[x]^n \xrightarrow{x-T} k[x]^n \xrightarrow{\pi} V_T \longrightarrow 0$$

is exact, and $\underline{y} = \{y_j = (x - T)e_j : 1 \le j \le n\}$ is a basis for $\ker(\pi) = (x - T)k[x]^n$. In particular $[x - T]_\mathbf{e}^{\mathrm{t}}$ is a presentation matrix for V_T.

Proof. Exactness at V_T means π is surjective, which is how π was defined in the first place. Exactness at the left $k[x]^n$ is because $f = 0$, which forces the rank of the presentation matrix, whatever it is, to be n.

Exactness at the right $k[x]^n$ means $\ker(\pi) = (x-T)k[x]^n = k[x][\underline{y}]$. Since x and T are identified on V_T via π, $k[x][\underline{y}] \subset \ker(\pi)$ as shown above. It remains to show this is the only relation, i.e., that $\underline{y}$ spans $\ker(\pi)$. We claim $k[x]^n = k[x][\underline{y}] + k[\mathbf{e}]$. Since $xe_j = y_j + T(e_j)$, and $T(e_j)$ is a k-linear combination of the e_i, we have $xe_j \in k[x][\underline{y}] + k[\mathbf{e}]$. By induction $x^m e_j \in k[x][\underline{y}] + k[\mathbf{e}]$, for all m, hence $g(x)e_j \in k[x][\underline{y}] + k[\mathbf{e}]$ for any $g(x) \in k[x]$, hence $\sum_i g_i(x)e_i \in k[x][\underline{y}] + k[\mathbf{e}]$ for any $a = \sum_i g_i(x)e_i \in k[x]^n$. This proves the claim.

If $a \in k[x]^n$ is in $\ker(\pi)$, and $a \in \sum_i b_i e_i + k[x][\underline{y}]$, then $\pi(a) = \pi(\sum_i b_i e_i) = \sum_i b_i v_i = 0$, with $b_i \in k$, hence $b_i = 0$ for all i, since $\mathbf{v}$ is a basis. Therefore $a \in k[x][\underline{y}]$, which shows $\ker(\pi) \subset k[x][\underline{y}]$. We conclude $\ker(\pi) = k[x][\underline{y}]$.

Since $\underline{y}$ obviously spans $k[x][\underline{y}]$ over $k[x]$, it remains to show the set $\underline{y}$ is linearly independent. Suppose $\sum_j h_j(x)y_j = 0$, with some $h_j(x) \neq 0$. Since $y_j = xe_j - T(e_j)$, $\sum_{j=1}^n h_j(x)xe_j = \sum_{j=1}^n h_j(x)T(e_j) = \sum_{i,j=1}^n h_j(x)a_{ij}e_i$. Since the e_i are $k[x]$-linearly independent, we conclude $h_i(x)x = \sum_{j=1}^n h_j(x)a_{ij}$. Let r be such that $h_r(x)$ has maximal degree among the $h_i(x)$, then $\deg(h_r(x)x) > \deg(\sum_j h_j(x)a_{ij})$, a contradiction. Therefore $\underline{y}$ is linearly independent, therefore it is a basis for $k[x][\underline{y}]$, hence for $\ker(\pi)$. Since the sequence is exact at $k[x]^n$, it follows that $[x-T]^{\mathrm{t}}_{\mathsf{e}}$ is the presentation matrix for V_T. □

Remark 5.3.3. Note that since V_T is a torsion $k[x]$-module, the rank of $x - T$ is n, hence the determinant of $[x-T]_{\mathsf{e}}$ is nonzero. Theorem 5.3.2 allows us to compute all of the invariants $d_i(x)$, by computing the gcd's of the degree-i minors of $xI - A$, and applying Theorem 2.1.6. In particular:

Corollary 5.3.4. *Let $A \in \mathrm{M}_n(k)$ be a matrix representing T. Then $p_T(x) = \det(xI - A)$.*

Proof. By definition $p_T(x)$ is the product $\prod_{i=1}^r d_i(x)$, where d_i is the i-th invariant of V_T in Structure Theorem 3.3.3. By Theorem 2.1.6, this equals Δ_n, the determinant of any matrix representing $x - T$. □

Example 5.3.5. We continue Example 5.0.1, with $V = k^2$, $\mathbf{e} = \{e_1, e_2\}$ the standard k-basis, and

$$A = [T]_{\mathsf{e}} = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}$$

We have an exact sequence $k[x]^2 \xrightarrow{xI-A} k[x]^2 \xrightarrow{\pi} k^2 \longrightarrow 0$. Applying Theorem 2.1.6 to

$$xI - A = \begin{bmatrix} x-2 & -1 \\ 1 & x-3 \end{bmatrix}$$

we find $\Delta_1 = 1$, and $\Delta_2 = p_A(x)$, so we have one nonzero nonunit $d_1(x) = p_A(x)$. Therefore $m_A(x) = p_A(x) = x^2 - 5x + 7$, and

$$V_T \simeq \frac{k[x]}{(x^2 - 5x + 7)}$$

We know by this that $\mathrm{SNF}(xI - A) = \mathrm{diag}\{1, x^2 - 5x + 7\}$, so for bases $\mathbf{e}'$ and $\mathbf{f}'$ for $k[x]^2$,

$$[xI - T]^{\mathsf{e}'}_{\mathsf{f}'} = \begin{bmatrix} 1 & 0 \\ 0 & x^2 - 5x + 7 \end{bmatrix}$$

is the (transpose of the) presentation matrix of V_T. By Example 2.1.3,

$$\underbrace{\begin{bmatrix} 1 & 0 \\ x-3 & 1 \end{bmatrix}}_{P^{-1}} \underbrace{\begin{bmatrix} x-2 & -1 \\ 1 & x-3 \end{bmatrix}}_{xI-A} \underbrace{\begin{bmatrix} 0 & 1 \\ -1 & x-2 \end{bmatrix}}_{Q} = \begin{bmatrix} 1 & 0 \\ 0 & x^2-5x+7 \end{bmatrix}$$

Thus $P^{-1} = [\mathrm{id}]_{\mathbf{e}}^{\mathbf{e}'}$ and $Q = [\mathrm{id}]_{\mathbf{f}'}^{\mathbf{e}}$. We compute $e_1' = e_1 + (3-x)e_2$ and $e_2' = e_2$, and $f_1' = -e_2$, $f_2' = e_1 + (x-2)e_2$ for the bases of $k[x]^2$ with respect to which we now have an exact sequence

$$0 \longrightarrow [k[x]^2]_{\mathbf{f}'} \xrightarrow{\left[\begin{smallmatrix} 1 & 0 \\ 0 & x^2-5x+7 \end{smallmatrix}\right]} [k[x]^2]_{\mathbf{e}'} \xrightarrow{\pi} k^2 \longrightarrow 0$$

By Corollary 3.4.1, the new set of generators for V_T is $\mathbf{v} = \mathbf{e}P$, so $v_1 = \pi(e_1') = e_1 + (3-A)e_2 = 0$, and $v_2 = \pi(e_2') = e_2$. Only one generator, because the module is cyclic.

5.4. **Similarity.** Let V be an n-dimensional k-vector space. A basis $\mathbf{e}$ for V sets up a 1-1 correspondence between $\mathrm{End}_k(V)$ and $\mathrm{M}_n(k)$, given by $T \longleftrightarrow [T]_{\mathbf{e}}$, under which the j-th column of $[T]_{\mathbf{e}}$ is the n-tuple of coefficients of $T(e_j)$, and $T(e_j)$ is the linear combination of the e_i given by the j-th column vector of $[T]_{\mathbf{e}}$. Thus we have $[T]_{\mathbf{e}}([v]_{\mathbf{e}}) = [T(v)]_{\mathbf{e}}$. We say the matrix $[T]_{\mathbf{e}}$ *represents* T with respect to the basis $\mathbf{e}$.

In the land of linear algebra, linear transformations or matrices are *similar* if they are conjugate under matrix multiplication in a group-theoretic sense, even though $(\mathrm{End}_k(V), \cdot)$ is a multiplicative monoid, not a group. If $S, T \in \mathrm{End}_k(V)$ are conjugate via $M \in \mathrm{Aut}_k(V)$, then writing $k^n = [V]_{\mathbf{e}}$, $A = [S]_{\mathbf{e}}$, $B = [T]_{\mathbf{e}}$, and $P = [M]_{\mathbf{e}}$, we have

$$\begin{tikzcd} V \arrow[r, "S"] \arrow[d, "M"'] & V \arrow[d, "M"] \\ V \arrow[r, "T"'] & V \end{tikzcd} \qquad \begin{tikzcd} k^n \arrow[r, "A"] \arrow[d, "P"'] & k^n \arrow[d, "P"] \\ k^n \arrow[r, "B"'] & k^n \end{tikzcd}$$

The equivalence classes in either case are called *similarity classes.*

Since P is invertible, it can be viewed as a basis change matrix $P = [\mathrm{id}]_{\mathbf{e}}^{\mathbf{f}}$. Then if $A = [T]_{\mathbf{e}}$, and $B = PAP^{-1}$, we have $B = [T]_{\mathbf{f}}$. Thus distinct $A, B \in \mathrm{M}_n(k)$ are similar if and only if they represent the same linear transformation with respect to different bases of V. Conversely, distinct S and T are similar if and only if they are represented by the same matrix with respect to different bases. For if $T = MSM^{-1}$ with $M \in \mathrm{Aut}_k(V)$, then viewing M as $[\mathrm{id}]_{\mathbf{e}}^{\mathbf{f}}$, for some new basis $\mathbf{f}$, we compute

$$[T]_{\mathbf{e}} = [MSM^{-1}]_{\mathbf{e}} = [M]_{\mathbf{e}}[S]_{\mathbf{e}}[M]_{\mathbf{e}}^{-1} = [\mathrm{id}]_{\mathbf{e}}^{\mathbf{f}}[S]_{\mathbf{e}}[\mathrm{id}]_{\mathbf{f}}^{\mathbf{e}} = [S]_{\mathbf{f}}$$

Theorem 5.4.1. *Let k be a field, $A, B \in \mathrm{M}_n(k)$. Let V_A and V_B denote the k-vector space V made into a $k[x]$-module via A and B, respectively. Then the following are equivalent.*

(a) *A and B are similar (in $\mathrm{M}_n(k)$).*
(b) *$V_A \simeq V_B$ as $k[x]$-modules.*
(c) *V_A and V_B have the same (monic) invariant factors.*
(d) *$\mathrm{SNF}(xI - A) = \mathrm{SNF}(xI - B)$.*
(e) *$xI - A$ and $xI - B$ are equivalent (in $\mathrm{M}_n(k[x])$).*

Proof.

(a) $\iff$ (b) : Suppose $B = PAP^{-1}$ for $P \in \mathrm{GL}_n(k)$. Then $P : V_A \to V_B$ is a k-linear transformation, but it is also a $k[x]$-module homomorphism: for since A and B are similar, $B^i = PA^iP^{-1}$, and we have $P(f(x)v) = Pf(A)(v) = f(B)P(v) = f(x)P(v)$ ✓. Since P is invertible, $V_A \simeq V_B$ as $k[x]$-modules, which is (b). Conversely, suppose $V_A \simeq V_B$ as $k[x]$-modules. Since an isomorphism is $k[x]$-linear, it is also k-linear, so it is represented by some $P \in \mathrm{GL}_n(k)$. Then $P(xv) = (PA)(v) = xP(v) = (BP)(v)$ for all v, hence $PA = BP$, so $B = PAP^{-1}$, which is (a).

(b) $\iff$ (c) : By Theorem 3.3.3 (and Theorem 3.6.1) V_A and V_B are each isomorphic to a coproduct of cyclic modules $R/(d_i)$, where the d_i are uniquely determined. Therefore $V_A \simeq V_B$ if and only if they have the same d_i's.

(c) $\iff$ (d) : The d_i's of A and B are *defined* by $\mathrm{SNF}(xI - A)$ and $\mathrm{SNF}(xI - B)$.

(d) $\iff$ (e) : Matrices are equivalent if and only if they have the same invariant factors by Theorem 2.1.6.

□

Remark 5.4.2. If $A \simeq B$ then $p_A(x) = p_B(x)$ by Theorem 5.4.1. The converse is false for $n \geq 2$.

Example 5.4.3. Determine which of the following matrices over $\mathbb{Q}$ are similar.

$$A = \begin{bmatrix} 2 & -2 & 14 \\ 0 & 3 & -7 \\ 0 & 0 & 2 \end{bmatrix} \qquad B = \begin{bmatrix} 0 & -4 & 85 \\ 1 & 4 & -30 \\ 0 & 0 & 3 \end{bmatrix} \qquad C = \begin{bmatrix} 2 & 2 & 1 \\ 0 & 2 & -1 \\ 0 & 0 & 3 \end{bmatrix}$$

By Theorem 5.4.1 it is enough to compute the invariants of the $\mathbb{Q}[x]$-modules V_A, V_B, V_C. Easy to check that $p_A(x) = p_B(x) = p_C(x) = (x-2)^2(x-3)$, so the product of the invariants of V_A, V_B, and V_C are the same. By combinatorics, there are two possibilities: $d_1 = (x-2)$, $d_2 = m(x) = (x-2)(x-3)$, and $d_1 = m(x) = (x-2)^2(x-3)$. It's easy to check that A satisfies $(x-2)(x-3) = 0$, but B and C do not. Therefore

$$V_A \simeq \frac{\mathbb{Q}[x]}{(x-2)} \oplus \frac{\mathbb{Q}[x]}{(x-2)(x-3)} \qquad V_B \simeq V_C \simeq \frac{\mathbb{Q}[x]}{(x-2)^2(x-3)}$$

By Theorem 5.4.1, A is not similar to B, but B is similar to C.

Example 5.4.4 (Trick for Computing $p_T(x)$)**.** The characteristic polynomial is invariant under matrix similarity, so it is associated to a linear transformation T, and can be computed with respect to any matrix representing T. By Corollary 5.3.4, $p_T(x) = \det(xI - A)$ for any matrix $A \in \mathrm{M}_n(k)$ representing T. Here's a shortcut that computes the coefficients of $p_T(x)$ directly from the coefficients of A.

For each subset $I = \{i_1, \dots, i_m\}$ of $\{1, \dots, n\}$, with $m \geq 0$, let M_I be the "diagonal" minor of A obtained by deleting rows I and columns I, and let

$$c_i = \sum_{I : |I| = n-i} M_I$$

the sum of the "diagonal" minors of degree $n - i$. For example $c_{n-1} = \mathrm{t}(A)$, and $c_0 = \det(A)$. Then

$$p_A(x) = x^n - c_{n-1}x^{n-1} + \cdots + (-1)^n c_0$$

Note the c_i are symmetric functions in the coefficients. We can prove it by expanding $\det(xI - A)$ along the first row, and gathering the coefficients of like powers of x.

Example 5.4.5. We can use Theorem 5.4.1 to compute the number of conjugacy classes of the group $\mathrm{GL}_n(\mathbb{F}_q)$, where q is a power of a prime. This group has order

$$|\mathrm{GL}_n(\mathbb{F}_q)| = (q^n - 1)(q^n - q)(q^n - q^2)\cdots(q^n - q^{n-1})$$

To verify this, note first that the number of vectors in $\mathbb{F}_q^n$ is q^n, so there are $q^n - 1$ nonzero vectors. Choose one for the first row, R_1, then it has q multiples, so we are left with $q^n - q$ candidates for the second row. Choose one, R_2. The number of vectors in the linear span $\mathbb{F}_q R_1 + \mathbb{F}_q R_2$ is q^2, so we are left with $q^n - q^2$ choices for the third row. Continuing in this way gives the formula.

The number of conjugacy classes in $\mathrm{GL}_n(\mathbb{F}_q)$ is the number of similarity classes, in this group. Each similarity class is completely determined by its invariant factors $d_1, \ldots, d_r$, with $d_i | d_j$ for $i < j$, and $\sum_i \deg(d_i) = n$. Thus we can count them by choosing a monic characteristic polynomial of degree n, and using combinatorics to categorize possible arrangements of its factors into invariants. Since we are only interested in the invertible elements of $\mathrm{M}_n(\mathbb{F}_q)$, by (5.4.4) we restrict to the characteristic polynomials with nonzero constant term.

For example, $|\mathrm{GL}_2(\mathbb{F}_3)| = 8 \cdot 6 = 48$. There are 9 monic polynomials of degree 2 over $\mathbb{F}_3$, of which 6 have nonzero constant term. If the polynomial is either irreducible or has two distinct linear factors, then its minimum polynomial equals its characteristic polynomial, and there is only one corresponding set of invariants. If the polynomial is a square, then there are two. There are 2 monic squares of degree 2 with nonzero constant term: $(x + 1)^2$ and $(x + 2)^2$. Thus we have $4 + 2 \cdot 2 = 8$ conjugacy classes.

6. Eigenvalues, Vectors, and Spaces

6.1. **Definitions.** Let k be a field, V an n-dimensional k-vector space, $T \in \mathrm{End}_k(V)$, and let V_T denote V made into a $k[x]$-module via T. Eigenvalues, eigenvectors, and eigenspaces are important features of linear transformations, and they play a role in diagonalization and most of the canonical forms. Recall the definitions:

Definition 6.1.1. Let V be an n-dimensional k-vector space, and $T \in \mathrm{End}_k(V)$.

- An *eigenvalue* of T is an element $\lambda \in k$ for which $\ker(T - \lambda)$ is nonzero.
- The *eigenspace* of an eigenvalue λ is the (nonzero) subspace $E_\lambda = \ker(T - \lambda) \subset V$.
- An *eigenvector* of an eigenvalue λ is any element of E_λ.
- An *eigenbasis* is a basis of eigenvectors.
- T is *diagonalizable* if V has an eigenbasis.
- The *algebraic multiplicity* $\mu_a(\lambda)$ is the multiplicity of λ as a root of $p_T(x)$.
- The *geometric multiplicity* $\mu_g(\lambda)$ is $\dim_k E_\lambda$.

6.2. **Main Example.** To examine these definitions in the $k[x]$-module setting, let V equal the coproduct of cyclic $k[x]$-modules $k[x]/(d_i(x))$. For any $\lambda \in k$, multiplication by $x - \lambda$ is a linear transformation on V, i.e., $x - \lambda \in \mathrm{End}_k(V)$, and we have an exact sequence

$$\begin{tikzcd} 0 \arrow[r] & E_\lambda \arrow[r] & \coprod_{i=1}^{r} \dfrac{k[x]}{(d_i(x))} \arrow[r, "x-\lambda"] & \coprod_{i=1}^{r} \dfrac{k[x]}{(d_i(x))} \end{tikzcd}$$

A vector has the form $v = \bar{f}_1(x) + \cdots + \bar{f}_r(x)$, where $\bar{f}_i(x)$ is a coset representative for $f_i(x) + (d_i(x))$.

Multiplication by $x - \lambda$ evidently takes each $k[x]/(d_i(x))$ to itself. We say it *stabilizes* each summand. Since $x - \lambda$ stabilizes the summands, v is in E_λ if and only if $(x - \lambda)\bar{f}_i(x) = 0$ for each i, i.e., each $\bar{f}_i(x)$ is in E_λ. Therefore

$$E_\lambda = \coprod_i E_\lambda \cap \frac{k[x]}{(d_i(x))} \tag{6.2.0.1}$$

Therefore we restrict $x - \lambda$ to a summand, and study the sequence

$$\begin{tikzcd} 0 \arrow[r] & E_\lambda \arrow[r] & \dfrac{k[x]}{(d(x))} \arrow[r, "x-\lambda"] & \dfrac{k[x]}{(d(x))} \end{tikzcd}$$

Let $e = \mu_a(\lambda)$, so $d(x) = (x - \lambda)^e g(x)$, with $x - \lambda \nmid g(x)$. Define $A_\lambda := \ker((x - \lambda)^e)$, so

$$\begin{tikzcd} 0 \arrow[r] & A_\lambda \arrow[r] & \dfrac{k[x]}{(d(x))} \arrow[r, "(x-\lambda)^e"] & \dfrac{k[x]}{(d(x))} \end{tikzcd}$$

is exact. Now we make some observations, all with $V = k[x]/(d(x))$.

(i) λ is an eigenvalue for x if and only if $x - \lambda \mid d(x)$: For by the gcd identity, $x - \lambda$ is a unit in V if and only if $\gcd(x - \lambda, d(x)) = 1$, if and only if multiplication by $x - \lambda$ is an automorphism, and $E_\lambda = (0)$.

(ii) Since $\gcd(x - \lambda, g(x)) = 1$, by Chinese Remainder Theorem we have

$$V = \frac{k[x]}{(x-\lambda)^e} \oplus \frac{k[x]}{(g(x))}$$

and λ is not an eigenvalue for the second summand by (i). Since $x-\lambda$ stabilizes each summand and is an automorphism on the second, E_λ appears already in the restricted exact sequence

$$\begin{tikzcd} 0 \arrow[r] & E_\lambda \arrow[r] & \dfrac{k[x]}{((x-\lambda)^e)} \arrow[r, "x-\lambda"] & \dfrac{k[x]}{((x-\lambda)^e)} \end{tikzcd}$$

(iii) If $e \geq 1$ then $\dim_k E_\lambda = 1$, and E_λ has k-basis $\{(x-\lambda)^{e-1}\}$: For $(x - \lambda)^e \mid (x - \lambda) f(x)$ if and only if $(x - \lambda)^{e-1} \mid f(x)$ by definition of divides, and since we may assume $\deg f(x) \leq e - 1$, we must have $f(x) = c(x - \lambda)^{e-1}$ for some $c \in k$. In particular $\mu_g(\lambda) \leq \mu_a(\lambda)$.

(iv) $\mu_a(\lambda) = \dim_k A_\lambda$: $(x - \lambda)^e$ kills the factor $k[x]/((x - \lambda)^e)$, which has dimension e, and is an automorphism of $k[x]/(g(x))$. Thus we have a (split) exact sequence

$$\begin{tikzcd} 0 \arrow[r] & A_\lambda \arrow[r] & \dfrac{k[x]}{(d(x))} \arrow[r, "(x-\lambda)^e"] & \dfrac{k[x]}{(g(x))} \arrow[r] & 0 \end{tikzcd}$$

(v) x is diagonalizable if and only if $d(x)$ splits into distinct linear factors: Since $\mu_g(\lambda) = 1$ and $\mu_g(\lambda) \leq \mu_a(\lambda)$ by 6.2(iii), and $\mu_a(\lambda)$ is the multiplicity of $x - \lambda$ in $d(x)$ by definition, the only way to get $\sum_\lambda \mu_g(\lambda) = n$ is to have $\mu_a(\lambda) = 1$ and $\sum_\lambda \mu_a(\lambda) = n$, which is the claim.

To summarize:

Theorem 6.2.1. *Let V be a finite-dimensional k-vector space, and let $T \in \operatorname{End}_k(V)$. Then*

(a) *$\lambda \in k$ is an eigenvalue for T if and only if $(x - \lambda) \mid p_T(x)$.*

(b) *$\mu_g(\lambda) = |\{i : x - \lambda \mid d_i(x)\}|$.*

(c) *$\mu_g(\lambda) \leq \mu_a(\lambda)$, and $\sum_\lambda \mu_a(\lambda) \leq \dim_k V$.*

(d) *T is diagonalizable if and only if $m_T(x)$ splits into distinct linear factors.*

Proof. Only (d) is not immediate from the previous discussion. Let V_i be the i-th summand in V's decomposition of Theorem 3.3.3. If T is diagonalizable then V has an eigenbasis, and each $v \in V_i$ can be written $v = \sum_\lambda w_\lambda$ for $w_\lambda \in E_\lambda$. Since $x - \lambda$ stabilizes the summands, we may assume $w_\lambda \in E_\lambda \cap V_i$ by (6.2.0.1). Therefore the eigenvectors of V_i span V_i, hence each V_i has an eigenbasis, and in particular $T|_{V_r}$ is diagonalizable. Consequently $d_r(x) = m_T(x)$ splits into distinct linear factors by (v). Conversely if $m_T(x)$ splits into distinct linear factors then so does each $d_i(x)$, since they all divide $d_r(x)$, and so each V_i has an eigenbasis by (v). The union of these bases is an eigenbasis for V, showing T is diagonalizable. □

Remark 6.2.2. Since $d_r(x) = m_T(x) \mid p_T(x)$, and $p_T(x) = \det(xI - A)$ for any matrix A representing T, $m_T(x)$ is often deducible in small degrees by just substituting A into different candidates for $m_T(x)$ and seeing if the result is zero. On the other hand, $m_T(x)$ is always directly computable by Theorem 2.1.6 as $\Delta_r \Delta_{r-1}^{-1}$, where Δ_i is the gcd of the degree-i minors of $xI - A$.

Example 6.2.3. The matrices

$$A = \begin{bmatrix} 5 & 6 & 0 \\ -3 & -4 & 0 \\ -2 & 0 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} 3 & -1 & 2 \\ -10 & 6 & -14 \\ -6 & 3 & -7 \end{bmatrix}$$

are similar, and diagonalizable. For we compute $p_A(x) = p_B(x) = x^3 - 2x^2 - x + 2$. We factor this by eye into $(x-1)(x+1)(x-2)$. Since the factors are all distinct, this is the minimum polynomial, and so A and B are similar, and diagonalizable.

6.3. **Finding the Basis Change Matrix P for Diagonal Form.** Suppose $T \in \mathrm{End}_k(V)$ and $A = [T]_{\mathbf{e}} \in \mathrm{M}_n(k)$. The procedure is to

(a) Find the eigenvalues, either from a description of T or by explicitly factoring $p_T(x)$.
(b) Find a basis for each $E_\lambda = \ker(\lambda I - A)$ using Gauss-Jordan elimination; let $\mathbf{f} = \{v_1, \dots, v_n\}$ be the eigenbasis for V.
(c) Then $D = P^{-1}AP$ is diagonal for $P = [v_1 \cdots v_n]$. The reason is that $P = [\mathrm{id}]^{\mathbf{e}}_{\mathbf{f}}$.

7. Canonical Forms

Let V be an n-dimensional k-vector space, and let $T \in \mathrm{End}_k(V)$. Each basis $\mathbf{e}$ for V determines a matrix $A = [T]_{\mathbf{e}}$ representing T. A k-basis $\mathbf{e}$ for V that determines a given A probably has nothing to do with the decomposition of the $k[x]$-module V_T in Theorem 3.3.3. But the decomposition itself suggests certain k-bases with respect to which T has a nice form, called a *canonical form*. There are many canonical forms: rational, Jordan, Weyr, Frobenius, and others. Each has its own strengths.

We will study the rational and Jordan canonical forms. Diagonal form is a special case of Jordan canonical form, but we treat it separately because often we just want to show that a matrix or linear transformation is diagonalizable. Many important types of linear transformations, such as those represented by real symmetric matrices, are diagonalizable.

7.1. **Main Example (cont'd).** Suppose $T = x$ and $V = k[x]/(d(x))$.

I. Suppose $d(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 + a_0$. Then the matrix of x with respect to the basis $\mathbf{v} = \{1, x, \ldots, x_{n-1}\}$ for $V = k[x]/(d(x))$ is

$$[x]_{\mathbf{v}} = C(d(x)) \;:=\; \begin{bmatrix} 0 & 0 & \cdots & & 0 & -a_0 \\ 1 & 0 & \cdots & & 0 & -a_1 \\ 0 & 1 & \cdots & & 0 & -a_2 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ & & & & 0 & -a_{n-2} \\ 0 & 0 & \cdots & 0 & 1 & -a_{n-1} \end{bmatrix}$$

called a *companion matrix* of $d_i(x)$.

II. Suppose $d(x) = \prod_{j=1}^{t}(x - \lambda_j)^{e_j}$. By Chinese Remainder Theorem,

$$\frac{k[x]}{(d(x))} = \coprod_{j=1}^{t} \frac{k[x]}{((x - \lambda_j)^{e_j})}$$

Let $V_j = k[x]/((x - \lambda_j)^{e_j})$. The elements $w_i = (x - \lambda_j)^{e_j - i} \in V_j$ define a basis $\mathbf{w}_j = \{w_1, \ldots, w_{e_j}\}$: The w_i are linearly independent since $\{1, x, \ldots, x^{e_j - 1}\}$ are linearly independent; each introduces a higher power of x, up to $x^{e_j - 1}$. The fact that they span is easy. To compute $[x|_{V_j}]_{\mathbf{w}_j}$ we compute $x \cdot w_j$ for each j, for these are the column vectors. Since $(x - \lambda_j)w_1 = (x - \lambda_j)^{e_j} = 0$ in V, $x \cdot w_1 = \lambda_j w_1$. Since $(x - \lambda_j)w_i = w_{i-1}$ for $i : 2 \le i \le e_j$, $x \cdot w_i = w_{i-1} + \lambda_j w_i$. Therefore

$$[x|_{V_j}]_{\mathbf{w}_j} = J_{e_j}(\lambda_j) \;:=\; \begin{bmatrix} \lambda_j & 1 & 0 & \cdots & 0 \\ 0 & \lambda_j & 1 & \ddots & \vdots \\ \vdots & 0 & \lambda_j & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 1 \\ 0 & 0 & \cdots & 0 & \lambda_j \end{bmatrix}$$

$J_{e_j}(\lambda_j)$ is called a *Jordan block*. Since x stabilizes each V_j, we have $x = \coprod_j x|_{V_j}$, and if $\mathbf{w} = \bigcup_j \mathbf{w}_j$,

$$[x]_{\mathbf{w}} = \coprod_{j=1}^{t} [x|_{V_j}]_{\mathbf{w}_j} = \coprod_{j=1}^{t} J_{e_j}(\lambda_j)$$

7.2. **Rational Canonical Form.**

Theorem 7.2.1 (Rational Canonical Form). *Every $T \in \mathrm{End}_k(V)$ has the rational canonical form*

$$[T]_{\mathbf{e}} = \coprod_{i=1}^{r} C(d_i(x))$$

where $C(d_i(x))$ is the companion matrix of 7.1(I). *This form is uniquely determined.*

Proof. We again may reduce to the case $V = k[x]/(d_i(x))$ and $T = x$, where the result follows from 7.1(I). Uniqueness is by Structure Theorem 3.3.3. □

Notation. We write $\mathrm{RCF}(A)$ for the rational canonical form of a matrix $A \in \mathrm{M}_n(k)$.

Example 7.2.2. From Example 6.2.3, the rational canonical forms of

$$A = \begin{bmatrix} 5 & 6 & 0 \\ -3 & -4 & 0 \\ -2 & 0 & 1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 3 & -1 & 2 \\ -10 & 6 & -14 \\ -6 & 3 & -7 \end{bmatrix} \text{ are } \mathrm{RCF}(A) = \mathrm{RCF}(B) = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

7.3. **Finding the Basis Change Matrix P for** $\mathrm{RCF}(A)$**.** To produce a rational canonical form for a given A we just have to compute the $d_i(x)$, from the minors of $xI_n - A$. To produce $B \in \mathrm{GL}_n(k)$ such that $B^{-1}AB$ is in rational canonical form is more involved: Let $\mathbf{v}$ be the standard k-basis of $V = k^n$, and $\mathbf{e}$ the corresponding basis of $k[x]^n$. First find $\mathrm{SNF}(xI - A)$ using row/column operations: we find $P, Q \in \mathrm{GL}_n(k[x])$ such that

$$P^{-1}(xI - A)Q = \mathrm{diag}\{\underbrace{1, \dots, 1}_{t}, d_1, \dots, d_r\}$$

Then $P = [\mathrm{id}]^{\mathbf{e}}_{\mathbf{e}'}$ for some new basis $\mathbf{e}'$, and $\mathbf{v}' = \mathbf{v}P$ is a new set of generators of V, by Corollary 3.4.1, with v'_{t+i} generating the i-th summand, which is isomorphic to $k[x]/(d_i(x))$.

As in Example 5.3.5, $\mathbf{v}'$ may not be a k-basis. Since P is in $\mathrm{GL}_n(k[x])$, the elements of $\mathbf{v}'$ are $k[x]$-linear combinations of the elements of $\mathbf{v}$, where x acts on V as A. In general the resulting v'_i are not linearly independent. In fact, we have $v'_1 = \cdots = v'_t = 0$, where t is the number of 1's. Since $t + r = n$, if $t > 0$ we have $r < n$ generators $\{v'_{t+1}, \dots, v'_{t+r}\}$ of the respective summands of the $k[x]$-module V. Let $m_i = \deg(d_i)$. Each one generates a Kronecker k-basis:

$$\left\{\{v'_{t+1}, xv'_{t+1}, \dots, x^{m_1-1}v'_{t+1}\}, \dots, \{v'_{t+r}, xv'_{t+r}, \dots, x^{m_r-1}v'_{t+r}\}\right\}$$

Since x acts as A, we can compute all of these elements explicitly in terms of $\{v'_1, \dots, v'_r\}$, hence in terms of $\mathbf{v} = \{v_1, \dots, v_n\}$. Let $\mathbf{v}''$ be this new k-basis, and let $B = [\mathrm{id}]^{\mathbf{v}}_{\mathbf{v}''}$. Then $\mathrm{RCF}(A) = B^{-1}AB$.

Example 7.3.1. Let $V = k^2$, with standard basis $\mathbf{v}$, and let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$. Then

$$xI - A = \begin{bmatrix} x-1 & -2 \\ -3 & x-4 \end{bmatrix}$$

and we compute $d_1(x) = 1$ and $d_2(x) = p_A(x) = x^2 - 5x - 2$. By Theorem 7.2.1, we know that there exists a matrix $B \in \mathrm{GL}_2(k)$ such that $B^{-1}AB = \mathrm{RCF}(A) = \begin{bmatrix} 0 & 2 \\ 1 & 5 \end{bmatrix}$. What is B?

First put $xI_2 - A$ into diagonal form. We use the operations $C_1 \leftrightarrow C_2$, $C_1 \mapsto -\frac{1}{2}C_1$, $C_2 \mapsto C_2 - (x-1)C_1$, $R_2 \mapsto R_2 - (-\frac{1}{2}x + 2)R_1$, $R_2 \mapsto 2R_2$ to get

$$\begin{aligned}\begin{bmatrix} 1 & 0 \\ 0 & x^2 - 5x - 2 \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ \frac{1}{2}x - 2 & 1 \end{bmatrix}\begin{bmatrix} x-1 & -2 \\ -3 & x-4 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} -\frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & -x+1 \\ 0 & 1 \end{bmatrix} \\ &= \underbrace{\begin{bmatrix} 1 & 0 \\ x-4 & 2 \end{bmatrix}}_{P^{-1}}\underbrace{\begin{bmatrix} x-1 & -2 \\ -3 & x-4 \end{bmatrix}}_{xI - A}\underbrace{\begin{bmatrix} 0 & 1 \\ -\frac{1}{2} & \frac{1}{2}x - \frac{1}{2} \end{bmatrix}}_{Q}\end{aligned}$$

Thus $P = [\mathrm{id}]^{\mathbf{e}}_{\mathbf{e}'} = \begin{bmatrix} 1 & 0 \\ -\frac{1}{2}x + 2 & \frac{1}{2} \end{bmatrix}$ and $Q = [\mathrm{id}]^{\mathbf{f}}_{\mathbf{f}'} = \begin{bmatrix} 0 & 1 \\ -\frac{1}{2} & \frac{1}{2}x - \frac{1}{2} \end{bmatrix}$. The k-basis $\mathbf{v}$ for V is a set of generators for the $k[x]$-module V. We expect the new set of generators $\mathbf{v}' = \mathbf{v}P$ to generate the summands of V, and in this case there is only one summand. Indeed, since x acts as A, $v_1' = v_1 + (-\frac{1}{2}x + 2)v_2 = 0$ and $v_2' = \frac{1}{2}v_2$. To get the rest of the desired k-basis for V, we apply x to the generator of each summand until we have a k-basis for that summand. In this case, the new k-basis of V is $\mathbf{v}'' = \{v_2', x \cdot v_2'\} = \{\frac{1}{2}v_2, x \cdot \frac{1}{2}v_2\} = \{\frac{1}{2}v_2, v_1 + 2v_2\}$. Thus

$$B = [\mathrm{id}]^{\mathbf{v}}_{\mathbf{v}''} = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & 2 \end{bmatrix}$$

and

$$B^{-1}AB = \begin{bmatrix} -2 & 1 \\ \frac{1}{2} & 0 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ \frac{1}{2} & 2 \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ 1 & 5 \end{bmatrix} \checkmark$$

7.4. **Jordan Canonical Form.** If $p_T(x)$ does not split into linear factors, then T is clearly not diagonalizable. But what if it does split into linear factors. Is it diagonalizable then? The answer is no. The *Jordan canonical form* is the closest we can get. Since we may construct a finite-degree splitting field for over k for $p_T(x)$, this form applies to all matrices over k, as long as we allow coefficients in some finite field extension of k.

Theorem 7.4.1 (Jordan Canonical Form). *Suppose $T \in \mathrm{End}_k(V)$, and $m_T(x)$ splits into linear factors in $k[x]$. Let $d_i(x) = \prod_{j=1}^{t}(x - \lambda_j)^{e_{ij}}$ for $e_{ij} \geq 0$ and distinct λ_j. Then T has the Jordan canonical form*

$$[T]_{\mathbf{w}} = \coprod_{j=1}^{t}\coprod_{i=1}^{r} J_{e_{ij}}(\lambda_j)$$

where $J_{e_{ij}}(\lambda_j)$ is the Jordan block of 7.1(II)*. This form is uniquely determined.*

Proof. We reduce as usual to $V = k[x]/(d_i(x))$ and $T = x$, where the result follows from 7.1(II). Uniqueness follows from the uniqueness of the elementary divisors. □

Notation. We write JCF(A) for the Jordan canonical form of a matrix $A \in \mathrm{M}_n(k)$.

Definition 7.4.2. Let V be an n-dimensional k-vector space, and $T \in \mathrm{End}_k(V)$.

- The *generalized eigenspace* of the eigenvalue λ_j of T is the $(x - \lambda_j)$-primary subspace

$$E^g_{\lambda_j} = \ker((x - \lambda_j)^{e_{rj}}) = \coprod_{i=s_j}^{r} k[x]/(x - \lambda_j)^{e_{ij}}$$

- A *generalized eigenvector* of an eigenvalue λ_j is any element of $E^g_{\lambda_j}$.
- A *generalized eigenbasis* of V is a k-basis that consists of generalized eigenvectors.

Remark 7.4.3. By Theorem 6.2.1(b), $\mu_a(\lambda_j) = \sum_{i=1}^{r} e_{ij}$ is the sum of the degrees of the λ_j-Jordan blocks, and $\mu_g(\lambda_j) = |\{e_{ij} : e_{ij} \neq 0\}|$ is the number of nontrivial λ_j-Jordan blocks.

Example 7.4.4. Find the rational and Jordan canonical forms for

$$A = \begin{bmatrix} 2 & -2 & 14 \\ 0 & 3 & -7 \\ 0 & 0 & 2 \end{bmatrix} \qquad B = \begin{bmatrix} 0 & -4 & 85 \\ 1 & 4 & -30 \\ 0 & 0 & 3 \end{bmatrix}$$

We've seen that $p_A(x) = p_B(x) = x^3 - 7x^2 + 16x - 12 = (x-2)^2(x-3)$, $m_A(x) = (x-2)(x-3)$, and $m_B(x) = p_B(x)$. Therefore

$$V_A \simeq \frac{k[x]}{(x-2)} \oplus \frac{k[x]}{(x^2-5x+6)} \simeq \frac{k[x]}{(x-2)} \oplus \frac{k[x]}{(x-2)} \oplus \frac{k[x]}{(x-3)}$$

$$V_B \simeq \frac{k[x]}{x^3-7x^2+16x-12} \simeq \frac{k[x]}{(x-2)^2} \oplus \frac{k[x]}{(x-3)}$$

Hence

$$\mathrm{RCF}(A) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & -6 \\ 0 & 1 & 5 \end{bmatrix} \qquad \mathrm{RCF}(B) = \begin{bmatrix} 0 & 0 & 12 \\ 1 & 0 & -16 \\ 0 & 1 & 7 \end{bmatrix}$$

$$\mathrm{JCF}(A) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \qquad \mathrm{JCF}(B) = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

Note the algebraic and geometric multiplicites in each case.

Example 7.4.5. Find all possible Jordan canonical forms for $\mathrm{M}_3(\mathbb{C})$, thereby classifying the linear transformations of $\mathbb{C}^3$. Fix $A \in \mathrm{M}_3(\mathbb{C})$. Then $p_A(x) = (x-\lambda_1)(x-\lambda_2)(x-\lambda_3)$ for $\lambda_i \in \mathbb{C}$. Either the λ_i are all distinct, or (WLOG) $\lambda_1 = \lambda_2 \neq \lambda_3$, or all three are equal. We have seen JCF(A) is diagonal in the first case. In the second case there are two possibilities: $m_A(x) = (x-\lambda_1)(x-\lambda_3)$, or $m_A(x) = p_A(x)$. In the third case, there are three: $m_A(x) = (x-\lambda_1)$, $m_A(x) = (x-\lambda_1)^2$, and $m_A(x-\lambda_1)^3$. So here's the list:

$$\begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_1 \end{bmatrix} \begin{bmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_1 & 1 \\ 0 & 0 & \lambda_1 \end{bmatrix} \begin{bmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 1 \\ 0 & 0 & \lambda_1 \end{bmatrix}$$

Note the generalized eigenspaces and algebraic/geometric multiplicities in each case. If an element of $\mathrm{M}_3(\mathbb{C})$ is selected at random, which of these is it most likely to represent?

Example 7.4.6. Suppose

$$V = \frac{k[x]}{(x-\lambda)^2} \oplus \frac{k[x]}{(x-\lambda)^3}$$

Then $p_x(x) = (x-\lambda)^5$, and

$$\mathrm{JCF}(x) = \begin{bmatrix} \lambda & 1 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & \lambda & 1 & 0 \\ 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{bmatrix}$$

The algebraic multiplicity of λ is five, the geometric multiplicity is two.

7.5. **Finding the Basis Change Matrix P for** $\mathrm{JCF}(A)$**.** Suppose $A = [T]_{\mathbf{e}}$ for some k-basis $\mathbf{e}$, and $p_T(x)$ splits into linear factors. Then A has a Jordan canonical form over k, and $\mathrm{JCF}(A) = PAP^{-1}$ for some $P = [\mathrm{id}]_{\mathbf{e}}^{\mathbf{w}} \in \mathrm{GL}_n(k)$. We can write down $\mathrm{JCF}(A)$ if we know the invariants d_i: we get a block $J_{e_{ij}}(\lambda_j)$ for each maximal divisor $(x-\lambda_j)^{e_{ij}}$ of d_i.

Definition 7.5.1. If $V \simeq k[x]/(x-\lambda)^n$, the basis $\mathbf{w}$ for V in 7.1(II), which satisfies $w_{j-1} = (x-\lambda)w_j$, for $j : 2 \le j \le n$, is called a *cycle rooted on (the eigenvector)* w_1. A cycle is generated by any $w_n : (x-\lambda)^n w_n = 0$ and $(x-\lambda)^{n-1} w_n \ne 0$, and then $w_{n-1} = (x-\lambda)w_n, \dots, w_1 = (x-\lambda)w_2$.

The columns P^{-1} are the disjoint union of the cycles rooted in eigenvectors for each Jordan block, as in Definition 7.5.1.

To find a $w_{e_{ij}} \in V$ such that $(x-\lambda_j)^{e_{ij}} w_{e_{ij}} = 0$ but $(x-\lambda_j)^{e_{ij}-1} \ne 0$, use elementary row operations to produce a basis of $\ker((A-\lambda_j I)^{e_{ij}})$, computing $(x-\lambda_j)^{e_{ij}-1}$, and choosing $w_{e_{ij}}$ from the basis by inspection. If there is more than one λ_j-Jordan block then for each successive algorithm we choose a generalized eigenvector for λ_j that is not in the span of the ones we have already found. This is feasible, if tedious, since we can compute a basis of the kernel of $(x-\lambda_j)^{e_{rj}}$.

Example 7.5.2. (a) Let $k = \mathbb{R}$. Find the possible Jordan canonical forms for

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Clearly the answer is

$$\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \quad \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}, \quad \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$$

Let's analyze this further. We compute $p_A(x) = x^2 - (a+d)x + (ad-bc)$. We have distinct roots unless $\Delta = (a+d)^2 - 4(ad-bc) = a^2 - 2ad + 4bc + d^2 = 0$. This defines a hypersurface H in $\mathbb{A}^4(\mathbb{R})$, which has measure zero. By Theorem 6.2.1, $\mathrm{JCF}(A)$ is diagonal if and only if $m_A(x)$ splits into linear factors. Therefore $\mathrm{JCF}(A) = \mathrm{diag}\{\lambda_1, \lambda_2\}$ if $A \notin H$, and if $A \in H$ then $\mathrm{JCF}(A) = \mathrm{diag}\{\lambda, \lambda\}$ if and only if $m_A(x)$ is linear, i.e., $\lambda I - A = 0$, i.e., A is already diagonal.

(b) Let $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. Find $\mathrm{JCF}(A)$ and the basis change matrix P such that $PAP^{-1} = \mathrm{JCF}(A)$. We compute $m_A(x) = p_A(x) = x^2 + 1 = (x-i)(x+i)$, so by Theorem 6.2.1,

$$\mathrm{JCF}(A) = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$$

To compute a basis for the eigenspace of i is to find a basis for

$$\ker(A - iI) = \ker \begin{bmatrix} -i & 1 \\ -1 & -i \end{bmatrix}$$

First put the matrix in reduced row echelon form using row operations. Elementary row operations do not change the kernel, since they are invertible and operate on the left. In this case we apply $T_{21}(i)D_1(-1)P_{12}(A - iI) = \left[\begin{smallmatrix} 1 & i \\ 0 & 0 \end{smallmatrix}\right]$, and conclude the kernel of $A - iI$ of has basis $\{w_1 = \left[\begin{smallmatrix} -i \\ 1 \end{smallmatrix}\right]\}$. Similarly $T_{21}(-i)P_{12}(A + iI) = \left[\begin{smallmatrix} 1 & -i \\ 0 & 0 \end{smallmatrix}\right]$, so the kernel of $A + iI$ has basis $\{w_2 = \left[\begin{smallmatrix} i \\ 1 \end{smallmatrix}\right]\}$. Therefore $\mathbf{w} = \{w_1, w_2\}$ is an eigenbasis, and $P^{-1} = [\mathrm{id}]_{\mathbf{w}}^{\mathbf{e}} = \left[\begin{smallmatrix} -i & i \\ 1 & 1 \end{smallmatrix}\right]$, with inverse $P = \frac{i}{2}\left[\begin{smallmatrix} 1 & -i \\ -1 & -i \end{smallmatrix}\right]$, and compute

$$\mathrm{JCF}(A) = PAP^{-1} = \frac{i}{2}\begin{bmatrix} 1 & -i \\ -1 & -i \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} -i & i \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} \checkmark$$

(c) Do the same for $A = \begin{bmatrix} 2 & -1 \\ 1 & 4 \end{bmatrix}$. We find $p_A(x) = x^2 - 6x + 9 = (x-3)^2$. Since $A \neq 3I$, $m_A(x) = (x-3)^2$, and by this we know A is not diagonalizable and

$$\mathrm{JCF}(A) = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$$

A generalized eigenbasis starts with a nonzero vector w_2 such that $(A-3I)^2 w_2 = 0$ but $(A-3I)w_2 \neq 0$. Since $m_A(x) = (x-3)^2$, any w_2 satisfies the first property. Compute $A - 3I = \left[\begin{smallmatrix} -1 & -1 \\ 1 & 1 \end{smallmatrix}\right]$, and I choose $w_2 = e_1$. To complete the cycle, compute $w_1 = (A-3I)w_2 = \left[\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right]$. Now $\mathbf{w} = \{w_1, w_2\}$ is a generalized eigenbasis, a cycle rooted on the eigenvector w_1. Therefore $P^{-1} = \begin{bmatrix} w_1 & w_2 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 1 & 0 \end{bmatrix}$ and

$$\mathrm{JCF}(A) = PAP^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 2 & -1 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} -1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix} \checkmark$$

Example 7.5.3. Find the Jordan canonical form and basis change matrix for

$$A = [T]_{\mathbf{e}} = \begin{bmatrix} 3 & -1 & -2 \\ 1 & 6 & 1 \\ 1 & 0 & 6 \end{bmatrix}$$

We compute $p_T(x) = (x-5)^3$. Algebraic multiplicity of $\lambda = 5$ is 3. To find $m_T(x)$, compute

$$A - 5I = \begin{bmatrix} -2 & -1 & -2 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \qquad (A-5I)^2 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

Since $(A-5I)^2 \neq 0$, we have $m_T(x) = p_T(x)$. Thus

$$\mathrm{JCF}(A) = J_3(5) = \begin{bmatrix} 5 & 1 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \end{bmatrix}$$

We know $J_3(5) = PAP^{-1}$. What is P? Use the explicit computation of $(A-5I)^2$ to pick w_3 so that $(A-5I)^2 w_3 \neq 0$ by inspection. Then put $w_2 = (A-5I)w_3$, and $w_1 = (A-5I)w_2$:

$$w_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad w_2 = (A-5I)w_3 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \qquad w_1 = (A-5I)w_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

Check that w_1 is an actual eigenvector. Our generalized eigenbasis is

$$\{w_1, w_2, w_3\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$$

Let $P^{-1} = \begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} = [\mathrm{id}]^{\mathbf{e}}_{\mathbf{w}}$, and watch the magic:

$$B = PAP^{-1} = \begin{bmatrix} 0 & 1 & -1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 3 & -1 & -2 \\ 1 & 6 & 1 \\ 1 & 0 & 6 \end{bmatrix}\begin{bmatrix} 1 & -2 & 1 \\ 0 & 1 & 0 \\ -1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 5 & 1 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \end{bmatrix} \quad !!$$