3.8 Diagonalizing Matrices

Last updated: 2020–11–18

A finitely generated module over a principal ideal domain is given by a matrix. By choosing appropriate bases of the source and target, we may classifying these modules.

3.8.1 Definition. An $(m \times n)$-matrix $\mathbf{A} = [a_{i,j}]$ over a commutative ring R is ***diagonal*** if $a_{i,j} = 0$ whenever $i \neq j$.

3.8.2 Definition. Two matrices $\mathbf{A}, \mathbf{B} \in \operatorname{Mat}(m, n, R)$ are ***equivalent*** if there are invertible matrices $\mathbf{C} \in \operatorname{Mat}(m, m, R)$ and $\mathbf{D} \in \operatorname{Mat}(n, n, R)$ such that $\mathbf{A} = \mathbf{C}\,\mathbf{B}\,\mathbf{D}$.

Equivalent matrices represent the same R-module homomorphism under different choices bases for the source and target.

3.8.3 Lemma. *Every matrix over a principal ideal domain is equivalent to a diagonal matrix.*

Proof. Let R be a principal ideal domain. For any elements $a, b \in R$, Theorem 2.8.5 demonstrates that there exists $x, y \in R$ such that $d := \gcd(a, b) = a\,x + b\,y$. Hence, for all $p, q \in R$, we have

$$\begin{bmatrix} a & b \\ p & q \end{bmatrix} \begin{bmatrix} x & -b/d \\ y & a/d \end{bmatrix} = \begin{bmatrix} d & 0 \\ px + qy & (aq - bp)/d \end{bmatrix}$$
$$\begin{bmatrix} x & y \\ -b/d & a/d \end{bmatrix} \begin{bmatrix} a & p \\ b & q \end{bmatrix} = \begin{bmatrix} d & px + qy \\ 0 & (aq - bp)/d \end{bmatrix}$$

and

$$\det \begin{bmatrix} x & -b/d \\ y & a/d \end{bmatrix} = \det \begin{bmatrix} x & y \\ -b/d & a/d \end{bmatrix} = \frac{ax + by}{d} = 1\,.$$

Thus, if a and b are entries in the same row (column) of a matrix $\mathbf{B}$, we can multiply $\mathbf{B}$ on the right (left) by an invertible matrix leaving d in the position occupied by a, leaving 0 in the position occupied by b, and fixing the entries not in the rows or columns of a and b.

Given a matrix $\mathbf{A}$ over the ring R, we multiply on the left and right by invertible matrices to obtain a diagonal matrix. The case $\mathbf{A} = \mathbf{0}$ is trivial, so assume that $\mathbf{A} \neq \mathbf{0}$. Row and column interchanges (left and right multiplications by suitable permutation matrices) bring a nonzero elmement a_1 to the upper left corner. By a sequence of right multiplications by invertible matrices, replace a_1 by a_2, the greatest common divisor of all elements in the first row, leaving zeros in the rest of the first row. Similarly, by left multiplications, replace a_2 by a_3, the greatest common divisor of all elements now in the first column, leaving zeros in the rest of the first column. Continue by replacing a_3 by a_4, the greatest common divisor of all elements now in the first row, and so on. In this manner, we generate a sequence $\langle a_1 \rangle \supseteq \langle a_2 \rangle \supseteq \langle a_3 \rangle \supseteq \cdots$ of ideals. Since R is a unique factorization domain, there exists an index m such that $a_{m+1} = a_m$.

It follows that a_m is the greatest common divisor of the elements in the first row (or column) and the remaining elements in the first column (or row) are zero. Hence, by elementary row or column operations, we can clear both the first row and first column, making all elements zero except for the corner. By induction on the size of the matrix, we can diagonalize $\mathbf{A}$. □

Working over the integers, this form was introduced by Henry Smith (1861).

3.8.4 Definition. An $(m \times n)$-matrix $\mathbf{A} = [a_{i,j}]$ over a commutative ring R is in ***Smith normal form*** if it is diagonal and $a_{i,i}$ divides $a_{i+1,i+1}$ for all $1 \leqslant i < r \leqslant \min(m,n)$ and $a_{j,j} = 0$ for all $r \leqslant j \leqslant \min(m,n)$.

3.8.5 Theorem. *Every matrix over a principal ideal domain is equivalent to a matrix in Smith normal form.*

Proof. By Lemma 3.8.3, it is enough to consider a diagonal matrix. For any elements $a, b \in R$, Theorem 2.8.5 demonstrates that there exists $x, y \in R$ such that $d := \gcd(a,b) = a\,x + b\,y$. It follows that

$$\begin{bmatrix} x & y \\ -b/d & a/d \end{bmatrix}\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}\begin{bmatrix} 1 & -by/d \\ 1 & ax/d \end{bmatrix} = \begin{bmatrix} ax & by \\ -ab/d & ab/d \end{bmatrix}\begin{bmatrix} 1 & -by/d \\ 1 & ax/d \end{bmatrix} = \begin{bmatrix} d & 0 \\ 0 & ab/d \end{bmatrix}.$$

Repeated application of this observation allow one to convert the diagonal matrix into a diagonal matrix with the property that the corner element divides all the remaining elements. By induction on the size of the matrix, we obtain a matrix in Smith normal form. □

3.8.6 Lemma. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be equivalent* $(m \times n)$*-matrix over a principal ideal domain. For all* i*, the ideals generated by the determinants of all* $(i \times i)$*-submatrices are equal.*

Proof. For any $(m \times n)$-matrix $\mathbf{A}$ and any $1 \leqslant i \leqslant \min(m,n)$, let $\Delta_i(\mathbf{A})$ be the ideal generated by the determinants of all $(i \times i)$-submatrices of $\mathbf{A}$. It suffices to show that, for any invertible $(m \times m)$-matrix $\mathbf{C}$, we have $\Delta_i(\mathbf{C}\,\mathbf{A}) = \Delta_i(\mathbf{A})$. The rows of $\mathbf{C}\,\mathbf{A}$ are linear combinations of the rows in $\mathbf{A}$. It follows that the determinants of $(i \times i)$-submatrices of $\mathbf{C}\,\mathbf{A}$ are linear combinations of the determinants of $(i \times i)$-submatrices of $\mathbf{A}$, so $\Delta_i(\mathbf{C}\,\mathbf{A}) \subseteq \Delta_i(\mathbf{A})$. Since $\mathbf{C}$ is invertible, we also have $\Delta_i(\mathbf{A}) = \Delta_i(\mathbf{C}^{-1}\,\mathbf{C}\,\mathbf{A}) \subseteq \Delta_i(\mathbf{C}\,\mathbf{A})$, so $\Delta_i(\mathbf{C}\,\mathbf{A}) = \Delta_i(\mathbf{A})$. □

Each matrix over a principal ideal domain is equivalent to an essentially unique matrix in Smith normal form.

3.8.7 Proposition. *Two* $(m \times n)$*-matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *in Smith normal form over a principal ideal domain are equivalent if and only if there exists a unit* u *such that* $a_{i,i} = c\,b_{i,i}$ *for each* $1 \leqslant i \leqslant \min(m,n)$.

Proof. One verifies that $\Delta_1(\mathbf{A}) = \langle a_{1,1} \rangle$ and $\Delta_i(\mathbf{A})\ \langle a_{i+1,i+1} \rangle = \Delta_{i+1}(\mathbf{A})$ for all $1 \leqslant i \leqslant \min(m,n)$. Hence, the diagonal elements are, up to a unit, determined by the ideals $\Delta_i(\mathbf{A})$. Lemma 3.8.6 implies that two matrices in Smith normal form are equivalent if and only if these ideals are equal. □

3.9 Modules over a Principal Ideal Domain

Last updated: 2020–11–17

Over a Euclidean domain, the Smith normal form of any matrix may be obtained via just elementary row and columns operations. We illustrate this process for a matrix over $\mathbb{F}_{11}[x]$:

$$\begin{bmatrix} 4x^3+4x^2-x+4 & -x^3-x^2+x+1 & 3x^3+3x^2+5 \\ 4x^3-5x+1 & -x^3+3x^2+x-3 & 3x^2-2x^2-4x+3 \\ x^3-x^2+x-1 & -3x^3-x^2+3x+1 & -2x^3+x^2+4x-3 \end{bmatrix}$$

$$\equiv \begin{bmatrix} 4x^3+4x^2-x+4 & -2x+2 & 3x^3+3x^2+5 \\ 4x^3-5x+1 & 3x^2-3x & 3x^2-2x^2-4x+3 \\ x^3-x^2+x-1 & -4x^2-5x-2 & -2x^3+x^2+4x-3 \end{bmatrix} \qquad \mathbf{c}_2 \mapsto \mathbf{c}_2 + 3\,\mathbf{c}_1$$

$$\equiv \begin{bmatrix} 4x^3+4x^2-x+4 & x-1 & 3x^3+3x^2+5 \\ 4x^3-5x+1 & 4x^2-4x & 3x^2-2x^2-4x+3 \\ x^3-x^2+x-1 & 2x^2-3x+1 & -2x^3+x^2+4x-3 \end{bmatrix} \qquad \mathbf{c}_1 \mapsto 5\,\mathbf{c}_2$$

$$\equiv \begin{bmatrix} x-1 & 4x^3+4x^2-x+4 & 3x^3+3x^2+5 \\ 4x^2-4x & 4x^3-5x+1 & 3x^2-2x^2-4x+3 \\ 2x^2-3x+1 & x^3-x^2+x-1 & -2x^3+x^2+4x-3 \end{bmatrix} \qquad \mathbf{c}_2 \leftrightarrow \mathbf{c}_1$$

$$\equiv \begin{bmatrix} x-1 & 0 & 3x^3+3x^2+5 \\ 4x^2-4x & -5x^4-x^3+4x^2+x+1 & 3x^2-2x^2-4x+3 \\ 2x^2-3x+1 & 3x^4-3x^3+5x^2+3x+3 & -2x^3+x^2+4x-3 \end{bmatrix} \qquad \mathbf{c}_1 \mapsto \mathbf{c}_2 + (-4x^2+3x+4)\,\mathbf{c}_1$$

$$\equiv \begin{bmatrix} x-1 & 0 & 0 \\ 4x^2-4x & -5x^4-x^3+4x^2+x+1 & -x^4+2x^3-2x^2-2x+3 \\ 2x^2-3x+1 & 3x^4-3x^3+5x^2+3x+3 & 5x^4-5x^3+4x^2+5x+2 \end{bmatrix} \qquad \mathbf{c}_1 \mapsto \mathbf{c}_2 + (-3x^2+5x+5)\,\mathbf{c}_1$$

$$\equiv \begin{bmatrix} x-1 & 0 & 0 \\ 0 & -5x^4-x^3+4x^2+x+1 & -x^4+2x^3-2x^2-2x+3 \\ 0 & 3x^4-3x^3+5x^2+3x+3 & 5x^4-5x^3+4x^2+5x+2 \end{bmatrix} \qquad \begin{matrix} \mathbf{r}_2 \mapsto \mathbf{r}_2 + (-4x)\,\mathbf{r}_1 \\ \mathbf{r}_3 \mapsto \mathbf{r}_3 + (-2x+1)\,\mathbf{r}_1 \end{matrix}$$

$$\equiv \begin{bmatrix} x-1 & 0 & 0 \\ 0 & -5x^4-x^3+4x^2+x+1 & -5x^2+5 \\ 0 & 3x^4-3x^3+5x^2+3x+3 & 3x^2-3 \end{bmatrix} \qquad \mathbf{c}_2 \mapsto \mathbf{c}_2 + 2\,\mathbf{c}_1$$

$$\equiv \begin{bmatrix} x-1 & 0 & 0 \\ 0 & -5x^4-x^3+4x^2+x+1 & x^2-1 \\ 0 & 3x^4-3x^3+5x^2+3x+3 & -5x^2+5 \end{bmatrix} \qquad \mathbf{c}_3 \mapsto -2\,\mathbf{c}_3$$

$$\equiv \begin{bmatrix} x-1 & 0 & 0 \\ 0 & x^2-1 & -5x^4-x^3+4x^2+x+1 \\ 0 & -5x^2+5 & 3x^4-3x^3+5x^2+3x+3 \end{bmatrix} \qquad \mathbf{c}_2 \leftrightarrow \mathbf{c}_3$$

$$\equiv \begin{bmatrix} x-1 & 0 & 0 \\ 0 & x^2-1 & 0 \\ 0 & 0 & 3x^3+3x^2-3x-3 \end{bmatrix} \qquad \begin{matrix} \mathbf{c}_3 \mapsto \mathbf{c}_3 + (5x^2+x+1)\,\mathbf{c}_2 \\ \mathbf{r}_3 \mapsto \mathbf{r}_3 + 5\,\mathbf{r}_2 \end{matrix}$$

$$\equiv \begin{bmatrix} x-1 & 0 & 0 \\ 0 & (x-1)(x+1) & 0 \\ 0 & 0 & (x-1)(x+1)^2 \end{bmatrix} \qquad \mathbf{c}_3 \mapsto 4\,\mathbf{c}_3$$

3.9.1 Theorem. *For any finitely generated module V over a principal ideal domain R, there exists a unique sequence $I_1 \supseteq I_2 \supseteq \cdots \supseteq I_n$ of proper ideals in R such that $V \cong R/I_1 \oplus R/I_2 \oplus \cdots \oplus R/I_n$.*

Proof of Existence. By Corollary 3.6.4, there exists an exact sequence

$$0 \longrightarrow R^{m_1} \xrightarrow{\varphi} R^{m_0} \longrightarrow V \longrightarrow 0$$

where $m_1 \leqslant m_0$. Theorem 3.8.5 shows that, by choosing suitable bases for the free modules R^{m_1} and R^{m_0}, we may assume that the matrix $\mathbf{M}(\varphi) = [a_{i,j}]$ is in Smith normal form. It follows that

$$V \cong \frac{R}{\langle a_{1,1}\rangle} \oplus \frac{R}{\langle a_{2,2}\rangle} \oplus \cdots \oplus \frac{R}{\langle a_{m_0,m_0}\rangle}$$

where $a_{i,i}$ divides $a_{i+1,i+1}$ for all $1 \leqslant i \leqslant m_1$. If $a_{i,i}$ is a unit, then $R/\langle a_{i,i}\rangle = 0$ is a trivial summand. Let ℓ be the largest integer such that $a_{\ell,\ell}$ is a unit. Setting $n := m_0 - \ell$ and $I_j := \langle a_{\ell+j,\ell+j}\rangle$ for all $1 \leqslant j \leqslant n$, we obtained the desired sequence of ideals in R. □

Outline for a Proof of Uniqueness. Suppose that there exists sequence $J_1 \supseteq J_2 \supseteq \cdots \supseteq J_m$ of ideals such that $V \cong R/J_1 \oplus R/J_2 \oplus \cdots \oplus R/J_m$. We first claim that $m > n$ implies that $J_1 = R$. Setting $S := R/J_1$, we obtain the S-module isomorphisms

$$\bigoplus_{j=1}^{m} \frac{R}{J_j + J_1} \cong \frac{V}{J_1 V} \cong \bigoplus_{i=1}^{n} \frac{R}{I_i + J_1} = S^n .$$

Hence, we can map S^m onto S^n, which confirms that $S = 0$.

We may now assume that $m = n$. It suffices by symmetry to demonstrate that $I_k \subseteq J_k$ for all $1 \leqslant k \leqslant n$. For any $r \in I_k$, we have

$$\bigoplus_{j=1}^{n} \frac{R}{(J_j : r)} \cong rV \cong \bigoplus_{i=k+1}^{n} \frac{R}{(I_i : r)}$$

where $(K : r) = \{a \in R \mid ra \in K\}$. Applying the first paragraph to the R-module rV, we deduce that $(J_1 : r) = (J_2 : r) = \cdots = (J_k : r) = R$, so $x \in J_k$. □

3.9.2 Definition. The ***torsion submodule*** of an R-module V is

$$\tau(V) := \{v \in V \mid rv = 0 \text{ for some } 0 \neq r \in R\} .$$

3.9.3 Corollary. *Let R be a principal ideal domain. For any finitely generated R-module V, there exists a unique nonnegative integer r such that*

$$V \cong \tau(V) \oplus R^r .$$

Proof. Theorem 3.9.1 gives a unique sequence $I_1 \supseteq I_2 \supseteq \cdots \supseteq I_n$ of proper ideals in R such that $V \cong R/I_1 \oplus R/I_2 \oplus \cdots \oplus R/I_n$. Setting k to be the largest integer such that $I_k \neq \langle 0\rangle$, it follows that

$$\tau(V) \cong R/I_1 \oplus R/I_2 \oplus \cdots \oplus R/I_k$$

and $V \cong \tau(V) \oplus R^{n-k}$. □