5.IV Jordan Form

This section uses material from three optional subsections: Direct Sum, Determinants Exist, and Other Formulas for the Determinant.

The chapter on linear maps shows that every $h\colon V \to W$ can be represented by a partial-identity matrix with respect to some bases $B \subset V$ and $D \subset W$. This chapter revisits this issue in the special case that the map is a linear transformation $t\colon V \to V$. Of course, the general result still applies but with the codomain and domain equal we naturally ask about having the two bases also be equal. That is, we want a canonical form to represent transformations as $\text{Rep}_{B,B}(t)$.

After a brief review section, we began by noting that a block partial identity form matrix is not always obtainable in this B, B case. We therefore considered the natural generalization, diagonal matrices, and showed that if its eigenvalues are distinct then a map or matrix can be diagonalized. But we also gave an example of a matrix that cannot be diagonalized and in the section prior to this one we developed that example. We showed that a linear map is nilpotent — if we take higher and higher powers of the map or matrix then we eventually get the zero map or matrix — if and only if there is a basis on which it acts via disjoint strings. That led to a canonical form for nilpotent matrices.

Now, this section concludes the chapter. We will show that the two cases we've studied are exhaustive in that for any linear transformation there is a basis such that the matrix representation $\text{Rep}_{B,B}(t)$ is the sum of a diagonal matrix and a nilpotent matrix in its canonical form.

5.IV.1 Polynomials of Maps and Matrices

Recall that the set of square matrices is a vector space under entry-by-entry addition and scalar multiplication and that this space $\mathcal{M}_{n\times n}$ has dimension n^2. Thus, for any $n\times n$ matrix T the n^2+1-member set $\{I, T, T^2, \dots, T^{n^2}\}$ is linearly dependent and so there are scalars $c_0, \dots, c_{n^2}$ such that $c_{n^2}T^{n^2} + \cdots + c_1T + c_0I$ is the zero matrix.

1.1 Remark This observation is small but important. It says that every transformation exhibits a generalized nilpotency: the powers of a square matrix cannot climb forever without a "repeat".

1.2 Example Rotation of plane vectors $\pi/6$ radians counterclockwise is represented with respect to the standard basis by

$$T = \begin{pmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{pmatrix}$$

and verifying that $0T^4 + 0T^3 + 1T^2 - 2T - 1I$ equals the zero matrix is easy.

1.3 Definition For any polynomial $f(x) = c_n x^n + \cdots + c_1 x + c_0$, where t is a linear transformation then $f(t)$ is the transformation $c_n t^n + \cdots + c_1 t + c_0(\mathrm{id})$ on the same space and where T is a square matrix then $f(T)$ is the matrix $c_n T^n + \cdots + c_1 T + c_0 I$.

1.4 Remark If, for instance, $f(x) = x - 3$, then most authors write in the identity matrix: $f(T) = T - 3I$. But most authors don't write in the identity map: $f(t) = t - 3$. In this book we shall also observe this convention.

Of course, if $T = \mathrm{Rep}_{B,B}(t)$ then $f(T) = \mathrm{Rep}_{B,B}(f(t))$, which follows from the relationships $T^j = \mathrm{Rep}_{B,B}(t^j)$, and $cT = \mathrm{Rep}_{B,B}(ct)$, and $T_1 + T_2 = \mathrm{Rep}_{B,B}(t_1 + t_2)$.

As Example 1.2 shows, there may be polynomials of degree smaller than n^2 that zero the map or matrix.

1.5 Definition The *minimal polynomial* $m(x)$ of a transformation t or a square matrix T is the polynomial of least degree and with leading coefficient 1 such that $m(t)$ is the zero map or $m(T)$ is the zero matrix.

A minimal polynomial always exists by the observation opening this subsection. A minimal polynomial is unique by the 'with leading coefficient 1' clause. This is because if there are two polynomials $m(x)$ and $\hat{m}(x)$ that are both of the minimal degree to make the map or matrix zero (and thus are of equal degree), and both have leading 1's, then their difference $m(x) - \hat{m}(x)$ has a smaller degree than either and still sends the map or matrix to zero. Thus $m(x) - \hat{m}(x)$ is the zero polynomial and the two are equal. (The leading coefficient requirement also prevents a minimal polynomial from being the zero polynomial.)

1.6 Example We can see that $m(x) = x^2 - 2x - 1$ is minimal for the matrix of Example 1.2 by computing the powers of T up to the power $n^2 = 4$.

$$T^2 = \begin{pmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{pmatrix} \quad T^3 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad T^4 = \begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix}$$

Next, put $c_4T^4 + c_3T^3 + c_2T^2 + c_1T + c_0I$ equal to the zero matrix

$$\begin{aligned}
-(1/2)c_4 \phantom{{}-c_3} + (1/2)c_2 + (\sqrt{3}/2)c_1 + c_0 &= 0 \\
-(\sqrt{3}/2)c_4 - c_3 - (\sqrt{3}/2)c_2 - (1/2)c_1 \phantom{{}+c_0} &= 0 \\
(\sqrt{3}/2)c_4 + c_3 + (\sqrt{3}/2)c_2 + (1/2)c_1 \phantom{{}+c_0} &= 0 \\
-(1/2)c_4 \phantom{{}-c_3} + (1/2)c_2 + (\sqrt{3}/2)c_1 + c_0 &= 0
\end{aligned}$$

and use Gauss' method.

$$\begin{aligned}
c_4 \phantom{{}+c_3} - c_2 - \sqrt{3}c_1 - 2c_0 &= 0 \\
c_3 + \sqrt{3}c_2 + 2c_1 + \sqrt{3}c_0 &= 0
\end{aligned}$$

Setting c_4, c_3, and c_2 to zero forces c_1 and c_0 to also come out as zero. To get a leading one, the most we can do is to set c_4 and c_3 to zero. Thus the minimal polynomial is quadratic.

Using the method of that example to find the minimal polynomial of a 3×3 matrix would mean doing Gaussian reduction on a system with nine equations in ten unknowns. We shall develop an alternative. To begin, note that we can break a polynomial of a map or a matrix into its components.

1.7 Lemma Suppose that the polynomial $f(x) = c_n x^n + \cdots + c_1 x + c_0$ factors as $k(x-\lambda_1)^{q_1}\cdots(x-\lambda_\ell)^{q_\ell}$. If t is a linear transformation then these two are equal maps.

$$c_n t^n + \cdots + c_1 t + c_0 = k\cdot(t-\lambda_1)^{q_1}\circ\cdots\circ(t-\lambda_\ell)^{q_\ell}$$

Consequently, if T is a square matrix then $f(T)$ and $k\cdot(T-\lambda_1 I)^{q_1}\cdots(T-\lambda_\ell I)^{q_\ell}$ are equal matrices.

Proof. This argument is by induction on the degree of the polynomial. The cases where the polynomial is of degree 0 and 1 are clear. The full induction argument is Exercise 1.7 but the degree two case gives its sense.

A quadratic polynomial factors into two linear terms $f(x) = k(x-\lambda_1)\cdot(x-\lambda_2) = k(x^2 + (\lambda_1+\lambda_2)x + \lambda_1\lambda_2)$ (the roots λ_1 and λ_2 might be equal). We can check that substituting t for x in the factored and unfactored versions gives the same map.

$$\begin{aligned}\bigl(k\cdot(t-\lambda_1)\circ(t-\lambda_2)\bigr)\,(\vec{v}) &= \bigl(k\cdot(t-\lambda_1)\bigr)\,(t(\vec{v})-\lambda_2\vec{v})\\ &= k\cdot\bigl(t(t(\vec{v}))-t(\lambda_2\vec{v})-\lambda_1 t(\vec{v})-\lambda_1\lambda_2\vec{v}\bigr)\\ &= k\cdot\bigl(t\circ t\,(\vec{v})-(\lambda_1+\lambda_2)t(\vec{v})+\lambda_1\lambda_2\vec{v}\bigr)\\ &= k\cdot(t^2-(\lambda_1+\lambda_2)t+\lambda_1\lambda_2)\,(\vec{v})\end{aligned}$$

The third equality holds because the scalar λ_2 comes out of the second term, as t is linear. QED

In particular, if a minimial polynomial $m(x)$ for a transformation t factors as $m(x) = (x-\lambda_1)^{q_1}\cdots(x-\lambda_\ell)^{q_\ell}$ then $m(t) = (t-\lambda_1)^{q_1}\circ\cdots\circ(t-\lambda_\ell)^{q_\ell}$ is the zero map. Since $m(t)$ sends every vector to zero, at least one of the maps $t-\lambda_i$ sends some nonzero vectors to zero. So, too, in the matrix case — if m is minimal for T then $m(T) = (T-\lambda_1 I)^{q_1}\cdots(T-\lambda_\ell I)^{q_\ell}$ is the zero matrix and at least one of the matrices $T-\lambda_i I$ sends some nonzero vectors to zero. Rewording both cases: at least some of the λ_i are eigenvalues. (See Exercise 29.)

Recall how we have earlier found eigenvalues. We have looked for λ such that $T\vec{v} = \lambda\vec{v}$ by considering the equation $\vec{0} = T\vec{v} - x\vec{v} = (T-xI)\vec{v}$ and computing the determinant of the matrix $T-xI$. That determinant is a polynomial in x, the characteristic polynomial, whose roots are the eigenvalues. The major result of this subsection, the next result, is that there is a connection between this characteristic polynomial and the minimal polynomial. This results expands on the prior paragraph's insight that some roots of the minimal polynomial are eigenvalues by asserting that every root of the minimal polynomial is an eigenvalue and further that every eigenvalue is a root of the minimal polynomial (this is because it says '$1\leq q_i$' and not just '$0\leq q_i$').

1.8 Theorem (Cayley-Hamilton) If the characteristic polynomial of a transformation or square matrix factors into

$$k \cdot (x - \lambda_1)^{p_1} (x - \lambda_2)^{p_2} \cdots (x - \lambda_\ell)^{p_\ell}$$

then its minimal polynomial factors into

$$(x - \lambda_1)^{q_1} (x - \lambda_2)^{q_2} \cdots (x - \lambda_\ell)^{q_\ell}$$

where $1 \leq q_i \leq p_i$ for each i between 1 and ℓ.

The proof takes up the next three lemmas. Although they are stated only in matrix terms, they apply equally well to maps. We give the matrix version only because it is convenient for the first proof.

The first result is the key — some authors call it the Cayley-Hamilton Theorem and call Theorem 1.8 above a corollary. For the proof, observe that a matrix of polynomials can be thought of as a polynomial with matrix coefficients.

$$\begin{pmatrix} 2x^2 + 3x - 1 & x^2 + 2 \\ 3x^2 + 4x + 1 & 4x^2 + x + 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix} x^2 + \begin{pmatrix} 3 & 0 \\ 4 & 1 \end{pmatrix} x + \begin{pmatrix} -1 & 2 \\ 1 & 1 \end{pmatrix}$$

1.9 Lemma If T is a square matrix with characteristic polynomial $c(x)$ then $c(T)$ is the zero matrix.

Proof. Let C be $T - xI$, the matrix whose determinant is the characteristic polynomial $c(x) = c_n x^n + \cdots + c_1 x + c_0$.

$$C = \begin{pmatrix} t_{1,1} - x & t_{1,2} & \ldots & \\ t_{2,1} & t_{2,2} - x & & \\ \vdots & & \ddots & \\ & & & t_{n,n} - x \end{pmatrix}$$

Recall that the product of the adjoint of a matrix with the matrix itself is the determinant of that matrix times the identity.

$$c(x) \cdot I = \operatorname{adj}(C)C = \operatorname{adj}(C)(T - xI) = \operatorname{adj}(C)T - \operatorname{adj}(C) \cdot x \qquad (*)$$

The entries of $\operatorname{adj}(C)$ are polynomials, each of degree at most $n - 1$ since the minors of a matrix drop a row and column. Rewrite it, as suggested above, as $\operatorname{adj}(C) = C_{n-1}x^{n-1} + \cdots + C_1 x + C_0$ where each C_i is a matrix of scalars. The left and right ends of equation $(*)$ above give this.

$$\begin{aligned} c_n I x^n + c_{n-1} I x^{n-1} + \cdots + c_1 I x + c_0 I = {} & (C_{n-1}T)x^{n-1} + \cdots + (C_1 T)x + C_0 T \\ & - C_{n-1}x^n - C_{n-2}x^{n-1} - \cdots - C_0 x \end{aligned}$$

Equate the coefficients of x^n, the coefficients of x^{n-1}, etc.

$$\begin{aligned} c_n I &= -C_{n-1} \\ c_{n-1} I &= -C_{n-2} + C_{n-1} T \\ &\vdots \\ c_1 I &= -C_0 + C_1 T \\ c_0 I &= C_0 T \end{aligned}$$

Multiply (from the right) both sides of the first equation by T^n, both sides of the second equation by T^{n-1}, etc. Add. The result on the left is $c_n T^n + c_{n-1} T^{n-1} + \cdots + c_0 I$, and the result on the right is the zero matrix. QED

We sometimes refer to that lemma by saying that a matrix or map *satisfies* its characteristic polynomial.

1.10 Lemma Where $f(x)$ is a polynomial, if $f(T)$ is the zero matrix then $f(x)$ is divisible by the minimal polynomial of T. That is, any polynomial satisfied by T is divisable by T's minimal polynomial.

Proof. Let $m(x)$ be minimal for T. The Division Theorem for Polynomials gives $f(x) = q(x)m(x) + r(x)$ where the degree of r is strictly less than the degree of m. Plugging T in shows that $r(T)$ is the zero matrix, because T satisfies both f and m. That contradicts the minimality of m unless r is the zero polynomial. QED

Combining the prior two lemmas gives that the minimal polynomial divides the characteristic polynomial. Thus, any root of the minimal polynomial is also a root of the characteristic polynomial. That is, so far we have that if $m(x) = (x - \lambda_1)^{q_1} \ldots (x - \lambda_i)^{q_i}$ then $c(x)$ must has the form $(x - \lambda_1)^{p_1} \ldots (x - \lambda_i)^{p_i} (x - \lambda_{i+1})^{p_{i+1}} \ldots (x - \lambda_\ell)^{p_\ell}$ where each q_j is less than or equal to p_j. The proof of the Cayley-Hamilton Theorem is finished by showing that in fact the characteristic polynomial has no extra roots λ_{i+1}, etc.

1.11 Lemma Each linear factor of the characteristic polynomial of a square matrix is also a linear factor of the minimal polynomial.

Proof. Let T be a square matrix with minimal polynomial $m(x)$ and assume that $x - \lambda$ is a factor of the characteristic polynomial of T, that is, assume that λ is an eigenvalue of T. We must show that $x - \lambda$ is a factor of m, that is, that $m(\lambda) = 0$.

In general, where λ is associated with the eigenvector $\vec{v}$, for any polynomial function $f(x)$, application of the matrix $f(T)$ to $\vec{v}$ equals the result of multiplying $\vec{v}$ by the scalar $f(\lambda)$. (For instance, if T has eigenvalue λ associated with the eigenvector $\vec{v}$ and $f(x) = x^2 + 2x + 3$ then $(T^2 + 2T + 3)(\vec{v}) = T^2(\vec{v}) + 2T(\vec{v}) + 3\vec{v} = \lambda^2 \cdot \vec{v} + 2\lambda \cdot \vec{v} + 3 \cdot \vec{v} = (\lambda^2 + 2\lambda + 3) \cdot \vec{v}$.) Now, as $m(T)$ is the zero matrix, $\vec{0} = m(T)(\vec{v}) = m(\lambda) \cdot \vec{v}$ and therefore $m(\lambda) = 0$. QED

1.12 Example We can use the Cayley-Hamilton Theorem to help find the minimal polynomial of this matrix.

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

First, its characteristic polynomial $c(x) = (x-1)(x-2)^3$ can be found with the usual determinant. Now, the Cayley-Hamilton Theorem says that T's minimal polynomial is either $(x-1)(x-2)$ or $(x-1)(x-2)^2$ or $(x-1)(x-2)^3$. We can decide among the choices just by computing:

$$(T - 1I)(T - 2I) = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 2 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and

$$(T - 1I)(T - 2I)^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and so $m(x) = (x-1)(x-2)^2$.

Exercises

✓ **1.13** What are the possible minimal polynomials if a matrix has the given characteristic polynomial?
(a) $8 \cdot (x-3)^4$ **(b)** $(1/3) \cdot (x+1)^3(x-4)$ **(c)** $-1 \cdot (x-2)^2(x-5)^2$
(d) $5 \cdot (x+3)^2(x-1)(x-2)^2$
What is the degree of each possibility?

✓ **1.14** Find the minimal polynomial of each matrix.

(a) $\begin{pmatrix} 3 & 0 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 4 \end{pmatrix}$ **(b)** $\begin{pmatrix} 3 & 0 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}$ **(c)** $\begin{pmatrix} 3 & 0 & 0 \\ 1 & 3 & 0 \\ 0 & 1 & 3 \end{pmatrix}$ **(d)** $\begin{pmatrix} 2 & 0 & 1 \\ 0 & 6 & 2 \\ 0 & 0 & 2 \end{pmatrix}$

(e) $\begin{pmatrix} 2 & 2 & 1 \\ 0 & 6 & 2 \\ 0 & 0 & 2 \end{pmatrix}$ **(f)** $\begin{pmatrix} -1 & 4 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & -4 & -1 & 0 & 0 \\ 3 & -9 & -4 & 2 & -1 \\ 1 & 5 & 4 & 1 & 4 \end{pmatrix}$

1.15 Find the minimal polynomial of this matrix.

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

✓ **1.16** What is the minimal polynomial of the differentiation operator d/dx on $\mathcal{P}_n$?