Multiplying through by A, and then substituting from (8.6.4) for A^n on the right, we obtain

$$(8.6.5)\quad A^{n+1} = \frac{a_{n-1}a_0}{a_n^2} I_n + \left(\frac{a_{n-1}a_1}{a_n^2} - \frac{a_0}{a_n}\right) A + \cdots + \left(\frac{a_{n-1}^2}{a_n^2} - \frac{a_{n-2}}{a_n}\right) A^{n-1}.$$

By continuing this process, we can express any positive integral power of A as a linear combination of $I, A, \ldots, A^{n-1}$.

If A^{-1} exists, by multiplying (8.6.3) by A^{-1} and then solving for A^{-1}, we obtain

$$(8.6.6)\quad A^{-1} = -\frac{a_1}{a_0} I - \frac{a_2}{a_0} A - \cdots - \frac{a_n}{a_0} A^{n-1}.$$

Multiplying again by A^{-1} and then substituting for A^{-1} on the right from (8.6.6), we obtain

$$(8.6.7)\quad A^{-2} = \left(\frac{a_1^2}{a_0^2} - \frac{a_2}{a_0}\right) I_n + \left(\frac{a_1a_2}{a_0^2} - \frac{a_3}{a_0}\right) A + \cdots + \left(\frac{a_1a_{n-1}}{a_0^2} - \frac{a_n}{a_0}\right) A^{n-2} + \left(\frac{a_1a_n}{a_0^2}\right) A^{n-1}.$$

Thus all negative integral powers of A may also be expressed as linear combinations of $I_n, A, \ldots, A^{n-1}$ when A^{-1} exists.

These procedures are used in deriving formulas for automatic computation.

Another possible application of the Cayley-Hamilton theorem is to the evaluation of $\varphi(\lambda)$ itself. In fact, dividing $\varphi(A)$ by $a_n = (-1)^n$, we obtain an equation

$$A^n + \alpha_1 A^{n-1} + \cdots + \alpha_{n-1}A + \alpha_n I_n = 0,$$

from which

$$(8.6.8)\quad (A^nX) + \alpha_1(A^{n-1}X) + \cdots + \alpha_{n-1}(AX) + \alpha_n X = 0,$$

where X is an arbitrarily chosen fixed vector. This matrix equation is equivalent to n scalar equations in the n unknowns $\alpha_1, \alpha_2, \ldots, \alpha_n$, which may be solved by any appropriate procedure provided the coefficient matrix is nonsingular. *This method of determining $\varphi(\lambda)$ often involves very tedious computations.*

8.7 The Minimum Polynomial of a Matrix

There are, of course, polynomial functions of a square matrix A, other than its characteristic function $\varphi(A)$, which reduce to zero. In fact, if $f(\lambda) = \varphi(\lambda)g(\lambda)$ where $g(\lambda)$ is any polynomial in λ, then $f(A) = 0$ also.

Among all not identically zero polynomials $p(\lambda)$ such that $p(A) = 0$, there must exist some of lowest degree. If this degree is μ, then $0 < \mu \leqslant n$.

Let each such polynomial of degree μ be divided by the coefficient of λ^μ. Then the result must be the same in every case, for if different polynomials were obtained, the difference of two of these, say $p_1(\lambda)$ and $p_2(\lambda)$, would be free of λ^μ,

$$p_1(\lambda) - p_2(\lambda) = \alpha_{\mu-1}\lambda^{\mu-1} + \alpha_{\mu-2}\lambda^{\mu-2} + \cdots + \alpha_1\lambda + \alpha_0.$$

Substitution of A would now yield

$$p_1(A) - p_2(A) \equiv \alpha_{\mu-1}A^{\mu-1} + \alpha_{\mu-2}A^{\mu-2} + \cdots + \alpha_1 A + \alpha_0 I_n = 0,$$

so that we would have an equation of degree less than μ satisfied by A. The contradiction proves our claim. This unique polynomial of lowest degree μ which vanishes at A is called the **minimum polynomial** of A and is denoted by $m(\lambda)$. We summarize in

Theorem 8.7.1: *For a given square matrix A, there exists a unique polynomial $m(\lambda)$ of lowest degree μ, in which the coefficient of λ^μ is equal to unity, such that $m(A) = 0$.*

To illustrate, consider the scalar matrix

$$A = \begin{bmatrix} \alpha & 0 \\ 0 & \alpha \end{bmatrix}.$$

Here $\varphi(\lambda) = (\alpha - \lambda)^2$, but the minimum polynomial is just $\lambda - \alpha$ since $A - \alpha I = 0$ here.

We prove now

Theorem 8.7.2: *Every polynomial $p(\lambda)$ such that $p(A) = 0$ is exactly divisible by $m(\lambda)$.*

To prove this, let the quotient when $p(\lambda)$ is divided by $m(\lambda)$ be $q(\lambda)$ and let the remainder, which is of degree less than μ, be $r(\lambda)$. Then we have

$$p(\lambda) \equiv m(\lambda)\, q(\lambda) + r(\lambda).$$

Substitution of A now yields

$$r(A) = 0.$$

Since $r(\lambda)$ is of degree less than μ, this implies $m(\lambda)$ is not a minimum polynomial unless $r(\lambda) \equiv 0$. Thus

$$p(\lambda) \equiv m(\lambda)\, q(\lambda),$$

and the theorem is proved.

A particular case of this result is that the minimum polynomial is a divisor of the characteristic polynomial $\varphi(\lambda)$. The relation between $\varphi(\lambda)$ and $m(\lambda)$ is more closely defined in the next two theorems.

Theorem 8.7.3: *Every linear factor $\lambda - \lambda_1$ of $\varphi(\lambda)$ is also a factor of $m(\lambda)$.*

We know in fact that dividing $m(\lambda)$ by $\lambda - \lambda_1$ yields an identity

$$m(\lambda) \equiv (\lambda - \lambda_1)\, s(\lambda) + r,$$

where r is a constant. Substituting A for λ, we have

$$m(A) = (A - \lambda_1 I_n)\, s(A) + rI_n = 0,$$

so that, if $r \neq 0$, we have

$$(A - \lambda_1 I_n)\left(\frac{-s(A)}{r}\right) = I_n.$$

This means $A - \lambda_1 I_n$ has an inverse whereas, since λ_1 is a characteristic root, it is singular. The contradiction shows that r must be zero, so that

$$m(\lambda) \equiv (\lambda - \lambda_1)\, s(\lambda),$$

as was to be proved.

An immediate consequence is

Theorem 8.7.4: *If the characteristic roots of A are distinct,*

$$\varphi(\lambda) = (-1)^n\, m(\lambda).$$

Indeed, in this case $\varphi(\lambda) = (-1)^n \prod_{j=1}^{n} (\lambda - \lambda_j)$ and $m(\lambda) = \prod_{j=1}^{n} (\lambda - \lambda_j)$, by the preceding theorem.

One might be tempted to conclude at this point that the minimum polynomial is simply the product of the distinct factors of $\varphi(\lambda)$. That this is not the case may be shown by an example. Let

$$A = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 1 & \cdots & 1 \\ 0 & 0 & 1 & \cdots & 1 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}_n$$

so that $\varphi(\lambda) \equiv (-1)^n(\lambda - 1)^n$.

Then $A - I \neq 0$ so that $\lambda - 1$ is not the minimum polynomial. However, it is easily verified that $(A - I)^n = 0$ while no lesser power of $A - I$ is zero. Thus the minimum polynomial is $(\lambda - 1)^n$ in this case.

An equation $A^p + \alpha_{p-1}A^{p-1} + \cdots + \alpha_1 A + \alpha_0 I = 0$, which is satisfied by a fixed matrix A, says that the matrices $I, A, A^2, \ldots, A^p$ are linearly dependent. Hence the minimum polynomial of A may be found by determining the smallest positive integer p for which such a dependence exists. Since $\varphi(A) = 0$, we have $p \leqslant n$. For small values of n, in the case of matrices with integer entries, the computation is sometimes practicable by the following scheme.

We begin by computing $A^2, A^3, \ldots, A^n$. We then form a row matrix B_i for each A^i (B_0 for I) such that

$$B_i = [A_1{}^i, A_2{}^i, \ldots, A_n{}^i],$$

where $A_j{}^i$ is the jth row of A^i. Thus each row B_i has n^2 elements. Now we form an $(n + 1) \times n^2$ matrix

$$B = \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_n \end{bmatrix}.$$

The problem is to find the smallest integer p such that the first p rows of B are dependent. To record conveniently the row operations required to determine this, we append an identity matrix of order $n + 1$ to B, thus obtaining an $(n + 1) \times (n^2 + n + 1)$ array $[B, I_{n+1}]$. We perform on the identity matrix I_{n+1} the same row operations as on B. However, the dependence of the rows of B is all that is at issue. For example, if

$$A = \begin{bmatrix} -1 & 3 & 0 & 3 \\ -1 & 3 & 1 & 3 \\ 1 & -1 & -2 & 0 \\ 1 & -1 & -1 & -1 \end{bmatrix},$$

we obtain this 5×21 matrix:

$$\begin{array}{r} I: \\ A: \\ A^2: \\ A^3: \\ A^4: \end{array} \left[\begin{array}{rrrrrrrrr} 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 3 & 0 & 3 & -1 & 3 & 1 & 3 & 1 \\ 1 & 3 & 0 & 3 & 2 & 2 & -2 & 3 & -2 \\ -1 & 9 & 0 & 9 & -3 & 11 & 3 & 9 & 3 \\ 1 & 15 & 0 & 15 & 4 & 12 & -4 & 15 & -4 \end{array}\right.$$

$$\left.\begin{array}{rrrrrrr:rrrrr} 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ -1 & -2 & 0 & 1 & -1 & -1 & -1 & 0 & 1 & 0 & 0 & 0 \\ 2 & 3 & 0 & -2 & 2 & 2 & 1 & 0 & 0 & 1 & 0 & 0 \\ -3 & -4 & 0 & 3 & -3 & -3 & -1 & 0 & 0 & 0 & 1 & 0 \\ 4 & 5 & 0 & -4 & 4 & 4 & 1 & 0 & 0 & 0 & 0 & 1 \end{array}\right].$$

Now any column among the first n^2, which is linearly dependent on preceding columns, may be deleted without altering the dependence of the rows in any way. This permits the successive deletion of columns 3, 4, 7, 8, 9, 10, 12, 13, 14, 15, 16 in this example, so that only this matrix needs to be considered:

$$\left[\begin{array}{rrrrr|rrrrr} 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 3 & -1 & 3 & -2 & 0 & 1 & 0 & 0 & 0 \\ 1 & 3 & 2 & 2 & 3 & 0 & 0 & 1 & 0 & 0 \\ -1 & 9 & -3 & 11 & -4 & 0 & 0 & 0 & 1 & 0 \\ 1 & 15 & 4 & 12 & 5 & 0 & 0 & 0 & 0 & 1 \end{array}\right].$$

By sweep-out, *using row operations only*, we now obtain:

$$\left[\begin{array}{rrrrr|rrrrr} 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 3 & -1 & 4 & -1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 3 & -3 & 3 & -2 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2 & -3 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -2 & -3 & 0 & 1 \end{array}\right].$$

The 5×5 matrix on the right, being in fact a product of matrices of elementary transformations, records what has been done to the abbreviated rows representing I, A, A^2, A^3, A^4 respectively. Indeed, if we interpret I, A, A^2, A^3, A^4 to mean *rows*, the product

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ -2 & -1 & 1 & 0 & 0 \\ -2 & -3 & 0 & 1 & 0 \\ 0 & -2 & -3 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} I \\ A \\ A^2 \\ A^3 \\ A^4 \end{bmatrix}$$

represents exactly what we must do to the original array in order to reduce rows four and five to rows of zeros. From the fourth row, we have, then,

$$-2I - 3A + 0 \cdot A^2 + A^3 + 0 \cdot A^4 = 0,$$

that is,

$$\text{(8.7.1)} \qquad A^3 - 3A - 2I = 0,$$

so that the minimal equation is

$$\lambda^3 - 3\lambda - 2 = 0,$$

with roots $-1, -1, 2$.

Notice that the fifth row yields

$$A^4 - 3A^2 - 2A = 0,$$

which is an immediate consequence of the equation (8.7.1) yielded by row four.

Although this method of finding the minimum polynomial is perfectly general, the computational difficulties increase rapidly with n, as is the case with most matrix calculations. This explains why effective procedures for use with digital computers are so essential.

8.8 Characteristic Roots of a Polynomial Function of a Matrix A

The following theorem, for which Exercises 9 and 15, Section 8.3 have been preparation, is also useful in applications:

Theorem 8.8.1: *If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the characteristic roots, distinct or not, of a matrix A of order n, and if $g(A)$ is any polynomial function of A, then the characteristic roots of $g(A)$ are $g(\lambda_1), g(\lambda_2), \ldots, g(\lambda_n)$.*

We know that

$$\det (A - \lambda I_n) \equiv (-1)^n(\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n),$$

and we wish to prove that, for any polynomial function $g(A)$,

$$\det (g(A) - \lambda I_n) \equiv (-1)^n(\lambda - g(\lambda_1))(\lambda - g(\lambda_2)) \cdots (\lambda - g(\lambda_n)).$$

Now suppose that $g(x)$ is of degree r in x and that, for a fixed value of λ, the roots of $g(x) - \lambda = 0$ are $x_1, x_2, \ldots, x_r$. Then we have

$$g(x) - \lambda \equiv \alpha(x - x_1)(x - x_2) \cdots (x - x_r),$$

where α is the coefficient of x^r in $g(x)$. Hence

$$g(A) - \lambda I_n \equiv \alpha(A - x_1 I_n)(A - x_2 I_n) \cdots (A - x_r I_n),$$

so that, if $\varphi(\lambda)$ is the characteristic polynomial of A,

$$\begin{aligned} \det (g(A) - \lambda I_n) &= \alpha^n \det (A - x_1 I_n) \det (A - x_2 I_n) \cdots \det (A - x_r I_n) \\ &= \alpha^n \varphi(x_1)\, \varphi(x_2) \cdots \varphi(x_r) \\ &= \alpha^n(-1)^n(x_1 - \lambda_1)(x_1 - \lambda_2) \cdots (x_1 - \lambda_n) \\ &\quad \times (-1)^n(x_2 - \lambda_1)(x_2 - \lambda_2) \cdots (x_2 - \lambda_n) \\ &\quad \vdots \\ &\quad \times (-1)^n(x_r - \lambda_1)(x_r - \lambda_2) \cdots (x_r - \lambda_n). \end{aligned}$$

Fortunately, the answer is yes. To see why, let V be the subspace of $\underline{C}^{k^2}$ spanned by I, A, ..., A^{k-1} and suppose $\sum_{n=0}^{\infty} \alpha_n A^n$ converges to f(A). The Cayley-Hamilton theorem implies that the partial sums, $\sum_{n=0}^{p} \alpha_n A^n$ all lie in V, therefore their limit, f(A), must also lie in V, and hence $f(A) = \beta_0 I + \beta_1 A + \cdots + \beta_{k-1} A^{k-1}$ for some scalars $\beta_i \in \underline{C}$. If we define $r(\tau) = \sum_{i=0}^{k-1} \beta_i \tau^i$, then according to the Lagrange-Sylvester theorem we have

$$r(A) = P\left(\bigoplus_{i=1}^{m} \sum_{j=0}^{n_i-1} \frac{r^{(j)}(\lambda_i)}{j!} U^j_{n_i}\right)P^{-1}$$

and

$$f(A) = P\left(\bigoplus_{i=1}^{m} \frac{f^{(j)}(\lambda_i)}{j!} U^j_{n_i}\right)P^{-1}$$

Comparing entries in $P^{-1}r(A)P$ with those in $P^{-1}f(A)P$, we have $r^{(j)}(\lambda_i) = f^{(j)}(\lambda_i)$ for each $1 \le i \le m$ and each $0 \le j < n_i$. Now $\ell(\lambda) = \max\{n_i : \lambda = \lambda_i\}$, so

$$r^{(j)}(\lambda) = f^{(j)}(\lambda) \qquad (17)$$

for each eigenvalue λ of A and each $0 \le j < \ell(\lambda)$ for this polynomial $r(\tau)$. We've already seen that any solution r to Eq. (17) has r(A) = f(A), so this method will always work.

17. THE MINIMAL POLYNOMIAL REVIEWED

The polynomial $p(\tau) = \sum_{i=0}^{m} \beta_i \tau^i$ is a *minimal polynomial* of the $k \times k$ matrix A iff

1. p(A) = 0
2. If $s(\tau)$ is a nonzero polynomial of degree m' and s(A) = 0, then $m' \ge m$
3. $\beta_m = 1$

In other words, $p(\tau)$ is a monic polynomial of least degree which has A for a root.

How do we know there is a minimal polynomial for A? How do we find one? How many are there?

Every Matrix Has a Minimal Polynomial

The finite sequence of matrices I, A, A^2, ..., A^{k^2} is linearly dependent (as a sequence of $k^2 + 1$ members of the k^2-dimensional space of all $k \times k$ matrices). Let m be the smallest index t such that I, A, ..., A^t is linearly dependent. Therefore there exist γ_i, not all zero, such that $\sum_{i=0}^{m} \gamma_i A^i = 0$. In particular, $\gamma_m \neq o$ because of the minimality of m. Therefore

$$A^m + \sum_{i=0}^{m-1} \frac{\gamma_i}{\gamma_m} A_i = 0$$

If we define the polynomial p by

$$p(\tau) = \tau^m + \sum_{i=0}^{m-1} \frac{\gamma_i}{\gamma_m} \tau^i \quad \text{for all } \tau$$

then (1) p(A) = 0, (2) no polynomial of degree less than p's has A for a root, and (3) p is monic; that is, p is a minimal polynomial for A.

How to Find a Minimal Polynomial for A

The proof of existence of a minimal polynomial suggests an algorithm for finding one:

1. Try to solve $A = \lambda_0 I$ for λ_0. If there is no solution,
2. Try to solve $A^2 = \lambda_0 I + \lambda_1 A$ for λ_0, λ_1. If there is no solution,

⋮

j. Try to solve $A^j = \sum_{i=0}^{j-1} \lambda_i A^i$ for $\lambda_0, \ldots, \lambda_{j-1}$. If there is no solution,

⋮

(This procedure must stop at some step as seen in Sec. 17.A)

⋮

m. $A^m = \sum_{i=0}^{m-1} \lambda_i A^i$;

so

$$p(\tau) = \tau^m - \sum_{i=0}^{m-1} \lambda_i \tau^i$$

Example: Suppose A is as in Example 4 of Sec. 16.

1. Try to solve $A = \lambda I$ for λ. There is no solution (otherwise, e.g., $-2 = \lambda$ and $2 = \lambda$).
2. Try to solve $A^2 = \lambda_0 I + \lambda_1 A$. If there is a solution then comparing the first rows of both sides:

   $[-6, 4, -4, 10] = [\lambda_0 - 2\lambda_1, 2\lambda_1, -2\lambda_1, 4\lambda_1]$

   and we'd have $4 = 2\lambda_1$ and $10 = 4\lambda_1$; so there is no solution.
3. Try to solve $A^3 = \lambda_0 I + \lambda_1 A + \lambda_2 A^2$. If there is a solution, then comparing first rows of both sides:

   $[-12, 6, -6, 20] = [\lambda_0 - 2\lambda_1 - 6\lambda_2, 2\lambda_1 + 4\lambda_2, 4\lambda_1 + 10\lambda_2]$

   and we'd have $\lambda_2 = 4$, $\lambda_1 = -5$, and $\lambda_0 = 2$. Is there a solution? Yes, because $2I - 5A + 4A^2 = A^3$.

Therefore $\tau^3 - 4\tau^2 + 5\tau - 2$ is a minimal polynomial for A.

Every Matrix Has at Most One Minimal Polynomial

Suppose p' and p are minimal polynomials for A. Dividing p' by p, we find polynomials q (the quotient polynomial) and r (the remainder polynomial) such that

$$p'(\tau) = p(\tau)q(\tau) + r(\tau) \quad \text{for all } \tau$$

Therefore, $r(A) = p'(A) - p(A)q(A) = 0$ by part (1) of the definition of minimal polynomial. Therefore part (2) of that definition ensures that r is the zero polynomial because the remainder's degree is always smaller than the divisor's. Consequently, $p'(\tau) = p(\tau)q(\tau)$. Part (2) implies that p' and p have the same degree, therefore $q(\tau)$ is a constant polynomial [$q(\tau) = \alpha$ for all τ]. Part (3) implies that $\alpha = 1$; so $p'(\tau) = p(\tau)$ for all τ, and hence every matrix has at most one minimal polynomial.

Exercises

82. Show that the minimal polynomial of A is a divisor of every polynomial having A for a root, i.e., if m_A is the minimal polynomial of A and f is any polynomial such that $f(A) = 0$ then there is some polynomial q such that

$$f(\tau) = m_A(\tau)q(\tau) \quad \text{for all } \tau.$$

83. Show that λ is an eigenvalue of A iff $m_A(\lambda) = 0$.