Appendix I

Cayley-Hamilton theorem

I.1 Statement of the theorem

According to the Cayley-Hamilton theorem, every square matrix A satisfies its own characteristic equation (Volume 1, section 9.3.1). Let the characteristic polynomial of A be

$$\chi(\lambda) = \det[\mathsf{A} - \lambda\mathbf{1}]. \tag{I.1}$$

When the determinant is fully expanded and terms in the same power of λ are collected, one obtains

$$\begin{aligned}\chi(\lambda) &= \sum_{j=0}^{n} \chi_j \lambda^j \\ &= \det[\mathsf{A}] + \cdots + (-1)^{n-1}\operatorname{trace}[\mathsf{A}]\lambda^{n-1} + (-1)^n\lambda^n.\end{aligned} \tag{I.2}$$

For example, let

$$\mathsf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 5 & 7 & 11 \\ 13 & 17 & 19 \end{pmatrix}. \tag{I.3}$$

Then

$$\mathsf{A} - \lambda\mathbf{1} = \begin{pmatrix} 1-\lambda & 2 & 3 \\ 5 & 7-\lambda & 11 \\ 13 & 17 & 19-\lambda \end{pmatrix}, \tag{I.4}$$

and

$$\chi(\lambda) = \det[\mathsf{A} - \lambda\mathbf{1}] = 24 + 77\lambda + 27\lambda^2 - \lambda^3. \tag{I.5}$$

It will become clear at the end of Section I.3 why a matrix of integers provides an especially appropriate example.

The Cayley-Hamilton theorem asserts that

$$\chi(\mathbf{A}) = \mathbf{0}, \tag{I.6}$$

where $\mathbf{0}$ is the zero matrix and

$$\begin{aligned}\chi(\mathbf{A}) &= \sum_{j=0}^{n} \chi_j \mathbf{A}^j \\ &= \det[\mathbf{A}]\mathbf{1} + \cdots + (-1)^{n-1}\,\mathrm{trace}\,[\mathbf{A}]\mathbf{A}^{n-1} + (-1)^n \mathbf{A}^n.\end{aligned} \tag{I.7}$$

For the example in Eqs. (I.3–I.5),

$$\chi(\mathbf{A}) = 24 + 77\mathbf{A} + 27\mathbf{A}^2 - \mathbf{A}^3. \tag{I.8}$$

Exercise I.1.2 verifies that $\chi(\mathbf{A}) = \mathbf{0}$ for this example.

Exercises for Section I.1

I.1.1 Verify Eq. (I.5).

I.1.2 Verify that $\chi(\mathbf{A}) = \mathbf{0}$, where $\mathbf{A}$ is defined in Eq. (I.3).

I.2 Elementary proof

Since the Cayley-Hamilton theorem is a fundamental result in linear algebra, it is useful to give two proofs, an elementary one and a more general one. The elementary proof is valid when the eigenvectors of $\mathbf{A}$ span the vector space on which $\mathbf{A}$ acts. The eigenvalue-eigenvector equation can be written in the form

$$\mathbf{AV} = \mathbf{V}\boldsymbol{\Lambda} \tag{I.9}$$

where $\mathbf{V}$ is the matrix of eigenvectors,

$$\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n] \tag{I.10}$$

and $\boldsymbol{\Lambda}$ is a diagonal matrix, the elements of which are the eigenvalues of $\mathbf{A}$ (Volume 1, Eq. (9.217)). Because this proof requires that the eigenvalues of $\mathbf{A}$ must belong to the same number field $\mathbb{F}$ to which the matrix elements of $\mathbf{A}$ belong, $\mathbb{F}$ must be algebraically closed, meaning that the roots of every polynomial equation with coefficients in $\mathbb{F}$ must belong to $\mathbb{F}$. For the purposes of this proof, it is convenient to assume that $\mathbb{F} = \mathbb{C}$, the field of complex numbers.

Because we have assumed that the eigenvectors span the entire vector space $\mathbb{C}^n$, the matrix $\mathbf{V}$ is non-singular. Therefore the inverse matrix $\mathbf{V}^{-1}$ exists. Multiplying both sides of Eq. (I.9) from the right with $\mathbf{V}^{-1}$ results in the equation

$$\mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}. \tag{I.11}$$

Now

$$\mathbf{A}^k = \mathbf{V}\mathbf{\Lambda}^k\mathbf{V}^{-1}, \tag{I.12}$$

from which it follows that

$$\begin{aligned}
\chi(\mathbf{A}) &= \mathbf{V}\chi(\mathbf{\Lambda})\mathbf{V}^{-1} \\
&= \mathbf{V}\begin{pmatrix} \chi(\lambda_1) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \chi(\lambda_n) \end{pmatrix}\mathbf{V}^{-1} \\
&= \mathbf{V}\begin{pmatrix} 0 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & 0 \end{pmatrix}\mathbf{V}^{-1} \\
&= \mathbf{V}\mathbf{0}\mathbf{V}^{-1} \\
&= \mathbf{0},
\end{aligned} \tag{I.13}$$

since each eigenvalue λ_k is a root of the characteristic equation $\chi(\lambda) = 0$. This establishes Eq. (I.6) for the case in which $\mathbf{A}$ is diagonalizable.

Exercises for Section I.2

I.2.1 Verify, without diagonalization or the calculation of eigenvectors, that every 2×2 matrix satisfies the Cayley-Hamilton theorem.

I.2.2 Verify, without diagonalization or the calculation of eigenvectors, that every strictly lower-triangular matrix satisfies the Cayley-Hamilton theorem.

I.2.3 Find the general form for the characteristic equation of the matrix of a proper rotation in $\mathbb{R}^3$. Also show that the eigenvalues are $\lambda = 1,\ e^{\pm i\theta}$, where θ is the angle of rotation.

I.3 General proof

The general proof assumes only scalars that belong to a number field $\mathbb{F}$, and $n \times n$ square matrices. A **matrix of scalars** is one in which every element is a scalar that belongs to $\mathbb{F}$. A **$\boldsymbol{\lambda}$-matrix $\mathbf{B}(\lambda)$** is a matrix, the elements of which are polynomials over $\mathbb{F}$ in an unknown λ. A λ-matrix is equal to the zero matrix, $\mathbf{0}$, if and only if every matrix element is equal to the zero polynomial.

If, in each matrix element of a λ-matrix, one collects terms in powers of λ, the resulting λ-matrix is equal to a sum of matrices of scalars, $\mathbf{B}_j$, each one multiplied by a power of λ:

$$\mathbf{B}(\lambda) = \sum_{j=0}^{l} \lambda^j \mathbf{B}_j \tag{I.14}$$

where $\mathbf{B}_l \neq \mathbf{0}$ unless $\mathbf{B}(\lambda) = \mathbf{0}$.

For example, let

$$\mathbf{B}(\lambda) = \begin{pmatrix} \lambda^2 - 26\lambda - 54 & 2\lambda + 13 & 3\lambda + 1 \\ 5\lambda + 48 & \lambda^2 - 20\lambda - 20 & 11\lambda + 4 \\ 13\lambda - 6 & 17\lambda + 9 & \lambda^2 - 8\lambda - 3 \end{pmatrix}. \tag{I.15}$$

Then

$$\mathbf{B}(\lambda) = \begin{pmatrix} -54 & 13 & 1 \\ 48 & -20 & 4 \\ -6 & 9 & -3 \end{pmatrix} + \lambda \begin{pmatrix} -26 & 2 & 3 \\ 5 & -20 & 11 \\ 13 & 17 & -8 \end{pmatrix} + \lambda^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{I.16}$$

It happens that $\mathbf{B}(\lambda)$ is the matrix of cofactors of the matrix $\mathbf{A} - \lambda\mathbf{1}$ in Eq. (I.4), and that $\mathbf{B}_0$ is the matrix of cofactors of $\mathbf{A}$; more of this later.

To prepare for the proof of the Cayley-Hamilton theorem, it is useful to show that, if a λ-matrix of the form

$$\mathbf{C}(\lambda) = \mathbf{B}(\lambda)(\mathbf{A} - \lambda\mathbf{1}) \tag{I.17}$$

is equal to a matrix of scalars, then $\mathbf{C}(\lambda) = \mathbf{0}$. To see this, expand the right-hand side using Eq. (I.14):

$$\begin{aligned} \mathbf{C}(\lambda) &= \sum_{j=0}^{l} \lambda^j \mathbf{B}_j (\mathbf{A} - \lambda\mathbf{1}) \\ &= -\lambda^{l+1}\mathbf{B}_l + \sum_{j=1}^{l} \lambda^j \left(\mathbf{B}_j\mathbf{A} - \mathbf{B}_{j-1}\right) + \mathbf{B}_0\mathbf{A}. \end{aligned} \tag{I.18}$$

Because $\mathbf{C}(\lambda)$ is equal to a matrix of scalars, the matrix coefficient of the leading term, $\mathbf{B}_l$, must vanish, along with each of the matrix coefficients in the terms $j = l, l-1, \ldots, 1$. But $\mathbf{B}_l = \mathbf{0}$ and $\mathbf{B}_l\mathbf{A} - \mathbf{B}_{l-1} = \mathbf{0}$ imply that $\mathbf{B}_{l-1} = \mathbf{0}$. Continuing the chain of equalities downward in j, one sees that $\mathbf{B}_j = \mathbf{0}$ for $j = l, l-1, \ldots, 0$. It follows that $\mathbf{C}(\lambda) = \mathbf{0}$.

Let $\mathbf{B}(\lambda)$ be the matrix of cofactors of the $n \times n$ matrix $\mathbf{A} - \lambda\mathbf{1}$. Each element of the cofactor matrix $\mathbf{B}(\lambda)$ is obtained from the matrix elements of $\mathbf{A} - \lambda\mathbf{1}$ by evaluating the determinant of a sub-matrix of $\mathbf{A} - \lambda\mathbf{1}$, and is therefore a polynomial of degree $n - 1$ or lower. Therefore $\mathbf{B}(\lambda)$ is a λ-matrix as defined above:

$$\mathbf{B}(\lambda) = \sum_{j=0}^{n-1} \lambda^j \mathbf{B}_j \ . \tag{I.19}$$

By the Laplace expansion of a determinant (Volume 1, Eq. (6.347)),

$$\mathbf{B}(\lambda)(\mathbf{A} - \lambda\mathbf{1}) = \det\left[\mathbf{A} - \lambda\mathbf{1}\right]\mathbf{1} = \chi(\lambda)\mathbf{1}, \tag{I.20}$$

where χ is the characteristic polynomial of the matrix $\mathbf{A}$.

We now evaluate $\chi(\mathbf{A})$ and show that it is equal to the zero matrix. For every $j \in (2:n)$,

$$\mathbf{A}^j - \lambda^j \mathbf{1} = \left(\mathbf{A}^{j-1} + \lambda \mathbf{A}^{j-2} + \cdots + \lambda^{j-2}\mathbf{A} + \lambda^{j-1}\mathbf{1}\right)(\mathbf{A} - \lambda \mathbf{1}). \tag{I.21}$$

Then

$$\begin{aligned}
\chi(\mathbf{A}) - \chi(\lambda)\mathbf{1} &= \sum_{j=0}^{n} \chi_j \left(\mathbf{A}^j - \lambda^j \mathbf{1}\right) \\
&= \sum_{j=2}^{n} \chi_j \left(\mathbf{A}^{j-1} + \lambda \mathbf{A}^{j-2} + \cdots + \lambda^{j-2}\mathbf{A} + \lambda^{j-1}\mathbf{1}\right)(\mathbf{A} - \lambda \mathbf{1}) \\
&\qquad + \chi_1(\mathbf{A} - \lambda \mathbf{1}) \\
&= -\mathbf{G}(\lambda)(\mathbf{A} - \lambda \mathbf{1})
\end{aligned} \tag{I.22}$$

where $\mathbf{G}(\lambda)$ is a λ-matrix. Then

$$\chi(\mathbf{A}) = \chi(\lambda)\mathbf{1} - \mathbf{G}(\lambda)(\mathbf{A} - \lambda \mathbf{1}). \tag{I.23}$$

Eq. (I.20) provides another expression for $\chi(\lambda)\mathbf{1}$. Substituting in Eq. (I.23), one obtains

$$\chi(\mathbf{A}) = \left[\mathbf{B}(\lambda) - \mathbf{G}(\lambda)\right](\mathbf{A} - \lambda \mathbf{1}). \tag{I.24}$$

But $\chi(\mathbf{A})$ is a matrix of scalars, and this equation is of the form of Eq. (I.17). Therefore

$$\chi(\mathbf{A}) = \mathbf{0}. \tag{I.25}$$

After a little study, one realizes that this proof makes no use of division by a scalar. In fact, the Cayley-Hamilton remains true if the number field $\mathbb{F}$ is replaced by a commutative ring R, such as the ring of integers, $\mathbb{Z}$. In this case, the underlying vector space is replaced by an R-module (Volume 1, p. 207). For example, if $R = \mathbb{Z}$, then an underlying space of $n \times 1$ column vectors with elements in a field $\mathbb{F}$ is replaced by a space whose elements are vectors of integers. The only allowable matrices that act on an R-module are matrices whose elements belong to R, as in Eq. (I.3), or are polynomials over R, as in Eq. (I.4). Determinants can still be defined. Each element of the matrix of cofactors is a determinant of a matrix of integers, and therefore is an integer, as in Eq. (I.15). The Laplace expansion of a determinant, Eq. (I.20), is still valid. Since every step of the proof remains valid for R-modules, the conclusion, Eq. (I.25), is also valid when the elements of $\mathbf{A}$ belong to a commutative ring R. The goal of Exercise I.3.2 is to verify the Cayley-Hamilton theorem for the matrix of integers defined in Eq. (I.3).

Exercises for Section I.3

I.3.1 Verify that the matrix in Eq. (I.15) is the matrix of cofactors of the matrix in Eq. (I.4).

I.3.2 Verify Eq. (I.20), using the examples in Eqs. (I.4–I.5) and (I.15).

I.3.3 One of the important practical applications of the Cayley-Hamilton theorem is to reduce the effort required for some matrix computations. Prove that, for a *nonsingular* $n \times n$ matrix A over a field $\mathbb{F}$, the matrix of cofactors of A is given by the expression

$$\mathsf{B}_0 = -\sum_{j=1}^{n} \chi_j \mathsf{A}^{j-1} \tag{I.26}$$

where χ_j is the coefficient of λ^j in the characteristic equation of A. This formula replaces the evaluation of n^2 determinants of $(n-1) \times (n-1)$ matrices with matrix multiplications and additions.

I.3.4 Prove that, if A is a nonsingular matrix over a field $\mathbb{F}$, then

$$\mathsf{A}^{-1} = \frac{1}{\det[\mathsf{A}]} \left((-1)^{n-1} \mathsf{A}^{n-1} + (-1)^{n-2} \operatorname{trace}[\mathsf{A}] \mathsf{A}^{n-2} + \cdots \right). \tag{I.27}$$

I.3.5 Identify every step in the proof of the Cayley-Hamilton theorem presented in Section I.2 that is not valid when the number field $\mathbb{F}$ is replaced by a commutative ring R, and explain why the step is invalid.

I.4 Bibliography

1. Garrett Birkhoff and Saunders Mac Lane, *A Survey of Modern Algebra,* Third Edition, Chapter X, §6 (Macmillan, 1965).

2. Bartel Leendert van der Waerden, *Modern Algebra,* Volume II, Chapter XV, §112 (Frederick Ungar, 1953).