The Characteristic Polynomial

1. Coefficients of the characteristic polynomial

Consider the eigenvalue problem for an \( n \times n \) matrix \( A \),

\[
A \vec{v} = \lambda \vec{v}, \quad \vec{v} \neq 0.
\]  
(1)

The solution to this problem consists of identifying all possible values of \( \lambda \) (called the eigenvalues), and the corresponding non-zero vectors \( \vec{v} \) (called the eigenvectors) that satisfy eq. (1). Consider the \( n \times n \) identity matrix \( I \). Noting that \( I \vec{v} = \vec{v} \), one can rewrite eq. (1) as

\[
(A - \lambda I) \vec{v} = 0.
\]  
(2)

This is a set of \( n \) homogeneous equations. If \( A - \lambda I \) is an invertible matrix, then one can simply multiply both sides of eq. (2) by \((A - \lambda I)^{-1}\) to conclude that \( \vec{v} = 0 \) is the unique solution. By definition, the zero vector is not an eigenvector. Thus, in order to find non-trivial solutions to eq. (2), one must demand that \( A - \lambda I \) is not invertible, or equivalently,

\[
p(\lambda) \equiv \det(A - \lambda I) = 0.
\]  
(3)

Eq. (3) is called the characteristic equation. Evaluating the determinant yields an \( n \)th order polynomial in \( \lambda \), called the characteristic polynomial, which we have denoted above by \( p(\lambda) \).

The determinant in eq. (3) can be evaluated by the usual methods. It takes the form,

\[
p(\lambda) = \det(A - \lambda I) = \begin{vmatrix}
    a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\
    a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda
\end{vmatrix}
\]  

\[= (-1)^n \left[ \lambda^n + c_1 \lambda^{n-1} + c_2 \lambda^{n-2} + \cdots + c_{n-1} \lambda + c_n \right],\]  
(4)

where \( A = [a_{ij}] \). The coefficients \( c_i \) are to be computed by evaluating the determinant. Note that we have identified the coefficient of \( \lambda^n \) to be \((-1)^n\). This arises from one term in the determinant that is given by the product of the diagonal elements. It is easy to show that this is the only possible source of the \( \lambda^n \) term in the characteristic polynomial. It is then convenient to factor out the \((-1)^n\) before defining the coefficients \( c_i \).

Two of the coefficients will be derived in problem 1 of Problem Set 1,

\[
c_1 = -\text{Tr} A, \quad c_n = (-1)^n \det A.
\]  
(5)
It then follows that the general form for the characteristic polynomial is:

\[ p(\lambda) = \det(A - \lambda I) = (-1)^n \left[ \lambda^n - \lambda^{n-1} \text{Tr} \ A + c_2 \lambda^{n-2} + \cdots + (-1)^{n-1}c_{n-1}\lambda + (-1)^n \det A \right]. \quad (6) \]

The explicit expressions for \( c_2, c_3, \ldots, c_{n-1} \) are more complicated than those of eq. (5). In the Appendix to these notes, I will provide explicit expressions for these coefficients in terms of traces of powers of \( A \).

By the fundamental theorem of algebra, an \( n \)th order polynomial equation of the form \( p(\lambda) = 0 \) possesses precisely \( n \) roots. Thus, the solution to \( p(\lambda) = 0 \) has \( n \) potentially complex roots, which are denoted by \( \lambda_1, \lambda_2, \ldots, \lambda_n \). These are the eigenvalues of \( A \). If a root is non-degenerate (i.e., only one root has a particular numerical value), then we say that the root has multiplicity one—it is called a \textit{simple root}. If a root is degenerate (i.e., more than one root has a particular numerical value), then we say that the root has multiplicity \( p \), where \( p \) is the number of roots with that same value—such a root is called a \textit{multiple root}. For example, a double root (as its name implies) arises when precisely two of the roots of \( p(\lambda) \) are equal. In the counting of the \( n \) roots of \( p(\lambda) \), multiple roots are counted according to their multiplicity.

One can always factor a polynomial in terms of its roots. Thus, eq. (4) implies that:

\[ p(\lambda) = (-1)^n(\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n), \]

where multiple roots appear according to their multiplicity. Multiplying out the \( n \) factors above yields

\[
p(\lambda) = (-1)^n \left[ \lambda^n - \lambda^{n-1} \sum_{i=1}^{n} \lambda_i + \lambda^{n-2} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \lambda_i \lambda_j + \cdots \\
+ \lambda^{n-k} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_k=1}^{n} \lambda_{i_1} \lambda_{i_2} \cdots \lambda_{i_k} + \cdots + \lambda_1 \lambda_2 \cdots \lambda_n \right]. \quad (7)
\]

Comparing with eq. (6), it immediately follows that:

\[ \text{Tr} \ A = \sum_{i=1}^{n} \lambda_i = \lambda_1 + \lambda_2 + \cdots + \lambda_n, \quad \det A = \lambda_1 \lambda_2 \lambda_3 \cdots \lambda_n. \]

The coefficients \( c_2, c_3, \ldots, c_{n-1} \) are also determined by the eigenvalues. In general,

\[
c_k = (-1)^k \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_k=1}^{n} \lambda_{i_1} \lambda_{i_2} \cdots \lambda_{i_k}, \quad \text{for} \ k = 1, 2, \ldots, n. \quad (8)
\]
2. The Cayley-Hamilton Theorem

**Theorem:** Given an $n \times n$ matrix $A$, the characteristic polynomial is defined by $p(\lambda) = \det(A - \lambda I) = (-1)^n [\lambda^n + c_1 \lambda^{n-1} + c_2 \lambda^{n-2} + \cdots + c_{n-1} \lambda + c_n]$, it follows that

$$p(A) = (-1)^n \left[ A^n + c_1 A^{n-1} + c_2 A^{n-2} + \cdots + c_{n-1} A + c_n I \right] = 0,$$

(9)

where $A^0 \equiv I$ is the $n \times n$ identity matrix and $0$ is the $n \times n$ zero matrix.

**False proof:** The characteristic polynomial is $p(\lambda) = \det(A - \lambda I)$. Setting $\lambda = A$, we get $p(A) = \det(A - A I) = \det(A - A) = \det(0) = 0$. This “proof” does not make any sense. In particular, $p(A)$ is an $n \times n$ matrix, but in this false proof we obtained $p(A) = 0$ where 0 is a number.

**Correct proof:** Recall that the adjugate (sometimes called the classical adjoint) of a matrix $S$, denoted by $\text{adj} \ S$, is the transpose of the matrix of cofactors.\(^2\) The cofactor expansion of the determinant is equivalent to the equation\(^3\)

$$S \text{adj} \ S = I \det S,$$

(10)

In particular, setting $S = A - \lambda I$, it follows that

$$(A - \lambda I) \text{adj}(A - \lambda I) = p(\lambda) I,$$

(11)

where $p(\lambda) = \det(A - \lambda I)$ is the characteristic polynomial. Since $p(\lambda)$ is an $n$th-order polynomial, it follows from eq. (11) that $\text{adj}(A - \lambda I)$ is a matrix polynomial of order $n - 1$. Thus, we can write:

$$\text{adj}(A - \lambda I) = B_0 + B_1 \lambda + B_2 \lambda^2 + \cdots + B_{n-1} \lambda^{n-1},$$

where $B_0, B_1, \ldots, B_{n-1}$ are $n \times n$ matrices (whose explicit forms are not required in these notes). Inserting the above result into eq. (11) and using eq. (4), one obtains:

$$(A - \lambda I)(B_0 + B_1 \lambda + B_2 \lambda^2 + \cdots + B_{n-1} \lambda^{n-1}) = (-1)^n \left[ \lambda^n + c_1 \lambda^{n-1} + \cdots + c_{n-1} \lambda + c_n \right] I.$$

(12)

Eq. (12) is true for any value of $\lambda$. Consequently, the coefficient of $\lambda^k$ on the left-hand side of eq. (12) must equal the coefficient of $\lambda^k$ on the right-hand side of eq. (12), for

---

\(^1\)In the expression for $p(\lambda)$, we interpret $c_n$ to mean $c_n \lambda^0$. Thus, when evaluating $p(A)$, the coefficient $c_n$ multiplies $A^0 \equiv I$.

\(^2\)The cofactor matrix of $S$ is the $n \times n$ matrix $C$, whose matrix elements are given by $C_{ij} = (-1)^{i+j} M_{ij}$, where $M_{ij}$ is the determinant of the matrix obtained from $S$ by deleting the $i$th row and $j$th column of $S$. Then, the the adjugate of $S$ is defined as adj $S = CT$, where $T$ indicates the transpose. That is, $\text{adj}(S)_{ij} = C_{ji} = (-1)^{i+j} M_{ji}$. For further details, see Ref. [2].

\(^3\)Note that if $\det S \neq 0$, then we may divide both sides of eq. (10) by the determinant and identify $S^{-1} = \text{adj} S / \det S$, since the inverse satisfies $SS^{-1} = I$. 

---

3
\( k = 0, 1, 2, \ldots, n \). This yields the following \( n + 1 \) equations:

\[
AB_0 = (-1)^n c_n \mathbf{I},
\]

\[
-B_{k-1} + AB_k = (-1)^n c_{n-k} \mathbf{I}, \quad k = 1, 2, \ldots, n - 1,
\]

\[
-B_{n-1} = (-1)^n \mathbf{I}.
\]

Using eqs. (13)–(15), we can evaluate the matrix polynomial \( p(A) \).

\[
p(A) = (-1)^n \left[ A^n + c_1 A^{n-1} + c_2 A^{n-2} + \cdots + c_{n-1} A + c_n \mathbf{I} \right]
\]

\[
= AB_0 + (-B_0 + B_1 A) A + (-B_1 + B_2) A^2 + \cdots + (-B_{n-2} + B_{n-1} A) A^{n-1} - B_{n-1} A^n
\]

\[
= A(B_0 - B_0) + A^2(B_1 - B_1) + A^3(B_2 - B_2) + \cdots + A^{n-1}(B_{n-2} - B_{n-2}) + A^n(B_{n-1} - B_{n-1})
\]

\[
= 0,
\]

which completes the proof of the Cayley-Hamilton theorem.

A notable feature of the Cayley-Hamilton theorem is that it provides a new way to compute the inverse of an \( n \times n \) matrix \( A \). The inverse \( A^{-1} \) exists if and only if \( c_n = (-1)^n \det A \neq 0 \) [cf. eq. (5)]. Multiplying eq. (9) by \( A^{-1} \) and dividing by \( c_n \) yields,

\[
A^{-1} = \left( \frac{(-1)^{n-1}}{\det A} \right) \left[ A^{n-1} + c_1 A^{n-2} + \cdots + c_{n-2} A + c_{n-1} \mathbf{I} \right], \quad \text{for } \det A \neq 0.
\]

Remarkably, the inverse matrix \( A^{-1} \) can always be expressed as a matrix polynomial of degree \( n - 1 \). Similarly, the Cayley-Hamilton theorem can be used to show that any matrix power, or more generally any function \( f(A) \) that can be expressed as a series (either finite or infinite) of the form,

\[
f(A) = \sum_j a_j A^j,
\]

can be expressed as a matrix polynomial of degree at most \( n - 1 \).

It is instructive to illustrate the Cayley-Hamilton theorem for \( 2 \times 2 \) matrices. In this case,

\[
p(\lambda) = \lambda^2 - \lambda \text{Tr } A + \det A.
\]

Hence, by the Cayley-Hamilton theorem,

\[
p(A) = A^2 - A \text{Tr } A + \mathbf{I} \det A = 0.
\]

Let us take the trace of this equation. Since \( \text{Tr } \mathbf{I} = 2 \) for the \( 2 \times 2 \) identity matrix,

\[
\text{Tr}(A^2) - (\text{Tr } A)^2 + 2 \det A = 0.
\]

It follows that for any \( 2 \times 2 \) matrix,

\[
\det A = \frac{1}{2} \left[ (\text{Tr } A)^2 - \text{Tr}(A^2) \right], \quad \text{and} \quad A^{-1} = \frac{1}{\det A} \left[ \mathbf{I} \text{Tr } A - A \right].
\]

You can easily verify the results of eq. (17) for any \( 2 \times 2 \) matrix.
Appendix: Identifying the coefficients of the characteristic polynomial in terms of traces

The characteristic polynomial of an $n \times n$ matrix $A$ is given by:

$$p(\lambda) = \det(A - \lambda I) = (-1)^n \left[ \lambda^n + c_1 \lambda^{n-1} + c_2 \lambda^{n-2} + \cdots + c_{n-1} \lambda + c_n \right].$$

In Section 1, we identified:

$$c_1 = -\text{Tr} \, A, \quad c_n = (-1)^n \det A. \quad (18)$$

One can also derive expressions for $c_2, c_3, \ldots, c_{n-1}$ in terms of traces of powers of $A$. In this appendix, I will exhibit the relevant results without proofs (which can be found in the references at the end of these notes). Let us introduce the notation:

$$t_k = \text{Tr}(A^k).$$

Then, the following recursive equation can be proven:

$$c_k = -\frac{1}{k} \sum_{i=1}^{k} c_{k-i} t_i, \quad \text{for } 1 \leq k \leq n,$$

where $c_0 \equiv 1$. More explicitly,

$$t_1 + c_1 = 0 \quad \text{and} \quad t_k + c_1 t_{k-1} + \cdots + c_{k-1} t_1 + k c_k = 0, \quad k = 2, 3, \ldots, n. \quad (19)$$

These equations are called the Newton’s identities. A nice proof of these identities can be found in Ref. [3]. The equations exhibited in eq. (19) are recursive, since one can solve for the $c_k$ in terms of the traces $t_1, t_2, \ldots, t_k$ iteratively by starting with $c_1 = -t_1$, and then proceeding step by step by solving the equations with $k = 2, 3, \ldots, n$ in successive order. This recursive procedure yields:

$$c_1 = -t_1,$$

$$c_2 = \frac{1}{2}(t_1^2 - t_2),$$

$$c_3 = -\frac{1}{6}t_3^3 + \frac{1}{2}t_1 t_2 - \frac{1}{3} t_3,$$

$$c_4 = \frac{1}{24} t_4^4 - \frac{1}{4} t_1 t_3^2 + \frac{1}{2} t_1 t_2 + \frac{1}{8} t_2^2 - \frac{1}{4} t_4,$$

and so on. The results above can be summarized by the following equation [4],

$$c_m = -\frac{t_m}{m} + \frac{1}{2!} \sum_{i=1}^{m-1} \sum_{j=1}^{m-1} \frac{t_i t_j}{i+j} - \frac{1}{3!} \sum_{i=1}^{m-2} \sum_{j=1}^{m-2} \sum_{k=1}^{m-2} \frac{t_i t_j t_k}{i+j+k} + \cdots + \frac{(-1)^m t_m^m}{m!}, \quad m = 1, 2, \ldots, n.$$
Note that by using $c_n = (-1)^n \det A$, one obtains a general expression for the determinant in terms of traces of powers of $A$,

\[
\det A = (-1)^n c_n = (-1)^n \left[ -\frac{t_n}{n} + \frac{1}{2!} \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \frac{t_i t_j}{ij} - \frac{1}{3!} \sum_{i=1}^{n-2} \sum_{j=1}^{n-2} \sum_{k=1}^{n-2} \frac{t_i t_j t_k}{ijk} + \cdots + \frac{(-1)^n t^n}{n!} \right],
\]

where $t_k \equiv \text{Tr}(A^k)$. One can verify that:

\[
\begin{align*}
\det A &= -\frac{1}{2} \left[ (\text{Tr} A)^2 - \text{Tr}(A^2) \right], \quad \text{for any } 2 \times 2 \text{ matrix,} \\
\det A &= -\frac{1}{6} \left[ (\text{Tr} A)^3 - 3 \text{Tr} A \text{Tr}(A^2) + 2 \text{Tr}(A^3) \right], \quad \text{for any } 3 \times 3 \text{ matrix,}
\end{align*}
\]

det. The coefficients of the characteristic polynomial, $c_k$, can also be expressed directly in terms of the eigenvalues of $A$, as shown in eq. (8).

**BONUS MATERIAL**

One can derive another closed-form expression for the $c_k$. To see how to do this, let us write out the Newton identities explicitly.

Eq. (19) for $k = 1, 2, \ldots, n$ yields:

\[
\begin{align*}
&c_1 = -t_1, \\
t_1 c_1 + 2c_2 = -t_2, \\
t_2 c_1 + t_1 c_2 + 3c_3 = -t_3, \\
&\vdots \quad \vdots \\
t_{k-1} c_1 + t_{k-2} c_2 + \cdots + t_1 c_{k-1} + k c_k = -t_k, \\
&\vdots \quad \vdots \\
t_{n-1} c_1 + t_{n-2} c_2 + \cdots + t_1 c_{n-1} + n c_n = -t_n.
\end{align*}
\]

Consider the first $k$ equations above (for any value of $k = 1, 2, \ldots, n$). This is a system of linear equations for $c_1, c_2, \ldots, c_k$, which can be written in matrix form:

\[
\begin{pmatrix}
1 & 0 & 0 & \cdots & 0 & 0 \\
t_1 & 2 & 0 & \cdots & 0 & 0 \\
t_2 & t_1 & 3 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
t_{k-2} & t_{k-3} & t_{k-4} & \cdots & k - 1 & 0 \\
t_{k-1} & t_{k-2} & t_{k-3} & \cdots & t_1 & k
\end{pmatrix}
\begin{pmatrix}
c_1 \\
c_2 \\
c_3 \\
\vdots \\
c_{k-1} \\
c_k
\end{pmatrix}
= \begin{pmatrix}
-t_1 \\
-t_2 \\
-t_3 \\
\vdots \\
-t_{k-1} \\
-t_k
\end{pmatrix}.
\]
Applying Cramer’s rule, we can solve for \( c_k \) in terms of \( t_1, t_2, \ldots, t_k \) [5]:

\[
c_k = \begin{vmatrix}
1 & 0 & 0 & \cdots & 0 & -t_1 \\
t_1 & 2 & 0 & \cdots & 0 & -t_2 \\
t_2 & t_1 & 3 & \cdots & 0 & -t_3 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
t_{k-2} & t_{k-3} & t_{k-4} & \cdots & k-1 & -t_{k-1} \\
t_{k-1} & t_{k-2} & t_{k-3} & \cdots & t_1 & -t_k
\end{vmatrix}.
\]

The denominator is the determinant of a lower triangular matrix, which is equal to the product of its diagonal elements. Hence,

\[
c_k = \frac{1}{k!} \begin{vmatrix}
1 & 0 & 0 & \cdots & 0 & -t_1 \\
t_1 & 2 & 0 & \cdots & 0 & -t_2 \\
t_2 & t_1 & 3 & \cdots & 0 & -t_3 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
t_{k-2} & t_{k-3} & t_{k-4} & \cdots & k-1 & -t_{k-1} \\
t_{k-1} & t_{k-2} & t_{k-3} & \cdots & t_1 & -t_k
\end{vmatrix}.
\]

It is convenient to multiply the \( k \)th column by \(-1\), and then move the \( k \)th column over to the first column (which requires a series of \( k-1 \) interchanges of adjacent columns). These operations multiply the determinant by \((-1)^1\) and \((-1)^{k-1}\) respectively, leading to an overall sign change of \((-1)^k\). Hence, our final result is:\(^4\)

\[
c_k = \frac{(-1)^k}{k!} \begin{vmatrix}
t_1 & 1 & 0 & 0 & \cdots & 0 \\
t_2 & t_1 & 2 & 0 & \cdots & 0 \\
t_3 & t_2 & t_1 & 3 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
t_{k-1} & t_{k-2} & t_{k-3} & t_{k-4} & \cdots & k-1 \\
t_k & t_{k-1} & t_{k-2} & t_{k-3} & \cdots & t_1
\end{vmatrix}, \quad k = 1, 2, \ldots, n.
\]

\(^4\)This result is derived in section 4.1 on p. 20 of Ref. [5]. However, the determinantal expression given in Ref. [5] for \( \sigma_k \equiv (-1)^k c_k \) contains a typographical error—the diagonal series of integers, 1, 1, 1, \ldots, 1, appearing just above the main diagonal of \( \sigma_k \) should be replaced by 1, 2, 3, \ldots, \( k-1 \).
We can test this formula by evaluating the first three cases $k = 1, 2, 3$:

\[ c_1 = -t_1, \quad c_2 = \frac{1}{2!} t_1^2 t_2 - t_2, \quad c_3 = \frac{1}{3!} t_1^3 - 3t_1 t_2 t_3, \]

which coincide with the previously stated results. Finally, setting $k = n$ yields the determinant of the $n \times n$ matrix $A$, $\det A = (-1)^n c_n$, in terms of traces of powers of $A$,

\[
\det A = \frac{1}{n!} \begin{vmatrix}
  t_1 & 1 & 0 & 0 & \cdots & 0 \\
  t_2 & t_1 & 2 & 0 & \cdots & 0 \\
  t_3 & t_2 & t_1 & 3 & \cdots & 0 \\
  \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
  t_{n-1} & t_{n-2} & t_{n-3} & t_{n-4} & \cdots & n-1 \\
  t_n & t_{n-1} & t_{n-2} & t_{n-3} & \cdots & t_1
\end{vmatrix},
\]

where $t_k \equiv \text{Tr}(A^k)$. Indeed, one can check that our previous results for the determinants of a $2 \times 2$ matrix and a $3 \times 3$ matrix are recovered.

REFERENCES


