§4.1. Eigenvalues and Eigenvectors

If \( A \) is a square matrix and \( \mathbf{v} \) is a non-zero vector such that \( A\mathbf{v} = \lambda \mathbf{v} \) we say that \( \mathbf{v} \) is an **eigenvector** of \( A \) and \( \lambda \) is the corresponding **eigenvalue**. Writing the equation as \( A\mathbf{v} - \lambda \mathbf{v} = 0 \), or \((\lambda I - A)\mathbf{v} = 0\) we see that if there is a non-zero solution for \( \mathbf{v} \) then the matrix \( \lambda I - A \) must be non-invertible and so \( |\lambda I - A| = 0 \).

If \( A \) is an \( n \times n \) matrix then \( |\lambda I - A| \) will be a polynomial of degree \( n \). We call this the **characteristic polynomial** of \( A \) and denote it by \( \chi_A(\lambda) \). So the eigenvalues of \( A \) are precisely the zeros of the characteristic polynomial.

Finding the characteristic polynomial and solving it is the normal way to find eigenvalues. And once we find the eigenvectors we can find the corresponding eigenvalues. But sometimes we find the eigenvectors first. Consider the following example.

**Example 1:** Find the eigenvalues and eigenvectors of \( A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \\ 3 & 4 & 1 & 2 \\ 4 & 1 & 2 & 3 \end{pmatrix} \).

**Solution:** It would take a lot of work to follow the traditional path of computing the characteristic polynomial first. It turns out to be \( \lambda^4 - 12\lambda^3 + 20\lambda^2 + 16\lambda - 160 \) and so there would remain the difficult job of finding its zeros.

But note that the sum of every row is the same. This means that \( \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \) is an eigenvector, for

\[
\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \\ 3 & 4 & 1 & 2 \\ 4 & 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = 10 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}
\]

is just another way of saying that every row adds to 10. So in this case we’ve found the eigenvector first, and then the eigenvalues. What about the other eigenvalues?

Note that each row is the same as the one above but rotated one place to the right, with the component that falls off the right-hand end going down to the left-hand end. This pattern can be encapsulated by the equation

\[
\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \\ 3 & 4 & 1 & 2 \\ 4 & 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ k \\ k^2 \\ k^3 \end{pmatrix} = (1 + 2k + 3k^2 + 4k^3) \begin{pmatrix} 1 \\ k \\ k^2 \\ k^3 \end{pmatrix}
\]

where \( k^4 = 1 \).

So for every 4’th root of 1, \( \begin{pmatrix} 1 \\ k \\ k^2 \\ k^3 \end{pmatrix} \) is an eigenvector and \( 1 + 2k + 3k^2 + 4k^3 \) is the corresponding eigenvalue. Putting \( k = 1, i, -1 \) and \(-i \) we conclude that the eigenvalues of \( A \) are 10, \(-2 \pm i \) and \(-2 \).

So \( \chi_A(\lambda) = (\lambda - 10)(\lambda + 2)(\lambda + 2 + 2i)(\lambda + 2 - 2i) = \lambda^4 - 12\lambda^3 + 20\lambda^2 + 16\lambda - 160 \).

So in this special case we found the eigenvectors first, then the eigenvalues and lastly the characteristic polynomial. This is the reverse of what we would normally do.
Theorem 1: The eigenvalues of a square matrix \( A \) are the zeros of its characteristic polynomial.

**Proof:** \( \lambda \) is an eigenvalue of \( A \) if and only if \( A\mathbf{v} = \lambda \mathbf{v} \) has a non-zero solution for \( \mathbf{v} \). This holds if and only if \( |\lambda I - A| = 0 \) which means that \( \lambda \) is a zero of the characteristic polynomial.

Theorem 2: The trace of a square matrix is the sum of its eigenvalues and the determinant is the product of its eigenvalues. (Here multiple zeros are counted according to their multiplicities.)

**Proof:** Let \( A = (a_{ij}) \) be a square matrix and let its characteristic polynomial be

\[
\lambda^n + c_{n-1}\lambda^{n-1} + \ldots + c_1\lambda + c_0.
\]

Then the sum of the eigenvalues is \(-c_{n-1}\). But the only term of degree \( n - 1 \) in \( |\lambda I - A| \) comes from \((\lambda - a_{11})(\lambda - a_{22}) \ldots (\lambda - a_{nn})\) and this has coefficient \(-\{a_{11} + a_{22} + \ldots + a_{nn}\} = -\text{tr}(A)\).

So \( c_{n-1} = -\text{tr}(A) \) and hence the sum of the eigenvalues is \( \text{tr}(A) \).

The product of the eigenvalues is \((-1)c_0\). Now the constant term is the value of the characteristic polynomial when \( \lambda = 0 = |A| = (-1)^n|A| \). (Remember that taking out the factor of \(-1\) from each row changes the sign of the determinant.)

Two \( n \times n \) matrices \( A, B \) are said to be similar if \( B = S^{-1}AS \) for some invertible matrix \( S \).

In ordinary algebra, the algebra of a field, the equation \( B = S^{-1}AS \) would give us \( B = A \) by cancelling the \( S^{-1} \) and the \( S \). But such “remote” cancelling, permissible in a field, is not permitted for a non-commutative system such as the system of \( n \times n \) matrices. We can only cancel if the element and its inverse are adjacent. In a field it doesn’t matter if they’re not because we can rearrange terms until they are. For matrices we just can’t do that because matrices in general don’t commute.

Similarity is clearly an equivalence relation in that:

1. every matrix is similar to itself – just take \( S = I \);
2. if \( A \) is similar to \( B \) then \( B \) is similar to \( A \);
3. if \( A \) is similar to \( B \) and \( B \) is similar to \( C \) then \( A \) is similar to \( C \).

Similar matrices are indeed similar in the non-technical sense in that they have many common properties.

Theorem 3: Similar matrices have the same characteristic polynomial.

**Proof:** Suppose that \( B = S^{-1}AS \) for some invertible matrix \( S \). The characteristic polynomial for \( B \) is \(|\lambda I - B| = |\lambda I - S^{-1}AS| = |S^{-1}(\lambda I - A)S| = |S|^{-1}|\lambda I - A||S| = |\lambda I - A|\).

**Corollary:** Similar matrices have the same determinant and the same trace.

§4.2. Calculating the Characteristic Polynomial

If \( k \geq 1 \), the \( k \)-th trace, \( \text{tr}_k(A) \), of a square matrix \( A \) is the sum of all \( k \times k \) sub-determinants got from \( A \) by deleting corresponding rows and columns (so that the diagonal of each sub-determinant coincides with the diagonal of \( A \). We define \( \text{tr}_0(A) = 1 \).

**Theorem 4:** Suppose \( A \) is an \( n \times n \) matrix and \( D = D(\lambda_1, \ldots, \lambda_n) = \begin{vmatrix} \lambda_1 - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda_2 - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda_n - a_{nn} \end{vmatrix} \) is a polynomial in \( n \) commuting variables \( \lambda_1, \ldots, \lambda_n \). For every subset \( S = \{s_1, \ldots, s_k\} \) of \( \{1, 2, \ldots, n\} \) with \( s_1 < s_2 < \ldots < s_k \) the coefficient of \( \lambda_{s_1}\ldots\lambda_{s_k} \) in \( D \) is \((-1)^{n-k}\) times the sub-determinant got by deleting all rows and columns except those corresponding to the elements of \( S \).

**Proof:** We prove this by induction on \( k \).

If \( k = 0 \), \( S = \emptyset \) and the required coefficient is the constant term of \( D \) which is \(|-A| = (-1)^n|A|\).
Suppose $k \geq 1$. Expanding $D$ along row $s_1$ the only term in $\lambda_{s_1} \ldots \lambda_{s_k}$ arises from
$(\lambda_{s_1} - a_{s_1})$ times the determinant $D'$ got from $D$ by deleting row $s_1$ and column $s_1$. This is the
coefficient of $\lambda_{s_2} \ldots \lambda_{s_k}$ in $D'$ which, by induction, is $(-1)^{(n-1)-(k-1)}$ times the sub-determinant $D''$ got
from $D'$ by deleting the rows and columns corresponding
$s_2, \ldots, s_k$. But this is $(-1)^{n-k}$ times the sub-determinant got from $D$ by deleting the rows and
columns corresponding to the elements of $S$.

**Theorem 5:** The characteristic polynomial of the $n \times n$ matrix $A$
$$
\chi_A(\lambda) = \lambda^n - \text{tr}(A)\lambda^{n-1} + \text{tr}_2(A)\lambda^{n-2} + \ldots + (-1)^{n-k} \text{tr}_k(A)\lambda^k + \ldots + (-1)^n |A|.
$$

**Proof:** With $D(\lambda_1, \ldots, \lambda_n)$ defined above $\chi_A(A) = D(\lambda, \ldots, \lambda)$.

**Example 2:** Find the characteristic polynomial of $A = \begin{pmatrix} 1 & 2 & 4 \\ 3 & 1 & 5 \\ 6 & 2 & 7 \end{pmatrix}$.

**Solution:** $\text{tr}_1(A) = 1 + 1 + 7 = 9$, $\text{tr}_2(A) = \begin{vmatrix} 1 & 2 \\ 3 & 1 \end{vmatrix} + \begin{vmatrix} 1 & 5 \\ 6 & 2 \end{vmatrix} = -25$,
$\text{tr}_3(A) = 124 = 15$ and so the characteristic polynomial is $\lambda^3 - 9\lambda^2 - 25\lambda - 15$.

§4.3. Diagonalisable Matrices

The simplest matrices to work with are the diagonal matrices. Not only do you add by
adding corresponding components, matrix multiplication for diagonal matrices simplifies to just
multiplying corresponding components. Moreover the eigenvalues are simply the diagonal
components and so the determinant of a diagonal matrix is simply the product of the diagonal
components.

If a matrix is similar to a diagonal matrix the components of that diagonal matrix are the
eigenvalues, both of the diagonal matrix and the original matrix. (Remember that similar matrices
have the same eigenvalues.)

A matrix is defined to be **diagonalisable** if it is similar to a diagonal matrix. One advantage
of having a diagonalisable matrix is that its powers are easy to compute. For if $A = S^{-1}DS$
then
$$
A^n = (S^{-1}DS)^n = S^{-1}DSS^{-1}DS \ldots S^{-1}DS^2S = S^{-1}D^nS,
$$
and if $D = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}$
then $D^n = \begin{pmatrix} \lambda_1^n & 0 & \ldots & 0 \\ 0 & \lambda_2^n & \ldots & 0 \\ \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \lambda_n^n \end{pmatrix}$.

Are all matrices diagonalisable? No, but they generally are. An example of a non-
diagonalisable matrix is $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ which has a double eigenvalue of $1$. If $A$ is similar to a
diagonal matrix $D$ then $D$ must be the diagonal matrix with the same eigenvalues. In other words
we must have $D = I$. But if $A = S^{-1}DS$ then $A = S^{-1}S = I$, which is clearly not so.

Here is a partial list of diagonalisable matrices. It doesn’t cover every diagonalisable matrix, but it
will be clear that to be diagonalisable is the rule, not the exception.

(1) $n \times n$ matrices with $n$ distinct eigenvalues;
(2) matrices $A$ such that $A^m = I$ for some positive integer $m$;
(3) real symmetric matrices.

If $A$, $S$ are an $n \times n$ matrices and the columns of $S$ are eigenvectors for $A$ then $S$ is called an **eigenmatrix** of $A$.
Theorem 5: If \( E = (v_1, v_2, \ldots, v_n) \) is an eigenmatrix for \( A \) then \( AE = ED \)

where \( D = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix} \) and \( \lambda_i \) is the eigenvalue corresponding to the eigenvector \( v_i \).

**Proof:**

\[
AE = A(v_1, v_2, \ldots, v_n) = (Av_1, Av_2, \ldots, Av_n) = (\lambda_1 v_1, \lambda_2 v_2, \ldots, \lambda_n v_n) \]

\[
= (v_1, v_2, \ldots, v_n) \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix} = ED.
\]

**Corollary:** A matrix is diagonalisable if and only if it has an invertible eigenmatrix.

**Proof:** Suppose \( A \) has an invertible eigenmatrix \( E \). Then \( AE = ED \) and so \( A = EDE^{-1} \).

Let \( S = E^{-1} \). Then \( A = S^{-1}DS \).

Conversely suppose that \( A \) is diagonalisable. Then \( A = S^{-1}DS \) for some diagonal matrix \( D \) and invertible matrix \( S \). Let \( E = S^{-1} \). Then \( A = EDE^{-1} \) and so \( AE = ED \). Let \( D = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix} \) and \( E = (v_1, v_2, \ldots, v_n) \). Hence \( AE = (Av_1, Av_2, \ldots, Av_n) = (\lambda_1 v_1, \lambda_2 v_2, \ldots, \lambda_n v_n) \) and so each \( v_i \) is an eigenvector for \( A \).

§4.4. Non-Negative Matrices

The **spectrum** of a matrix \( A \) is the set of eigenvalues. It is denoted by \( \sigma(A) \). The **spectral radius** of \( A \) is the largest modulus of the eigenvalues of \( A \).

A real matrix \( A = (a_{ij}) \) is called **non-negative** if each \( a_{ij} \geq 0 \) and **positive** if each \( a_{ij} > 0 \). We write these properties as \( A \geq 0 \) and \( A > 0 \) respectively.

**Theorem 6 (Perron):** If \( A > 0 \) and \( r \) is the spectral radius of \( A \) \( r \in \sigma(A) \) and \( |z| < r \) for all other \( z \in \sigma(A) \).

**Corollary:** If \( A^k > 0 \) for some \( k \) and \( r \) is the spectral radius of \( A \) then \( r \in \sigma(A) \) and \( |z| < r \) for all other \( z \in \sigma(A) \).

**Proof:** Let \( \sigma(A) = \{\lambda_1, \lambda_2, \ldots, \lambda_n\} \) with \( |\lambda_i| \geq |\lambda_{i+1}| \) for \( r = 1, 2, \ldots, n - 1 \). Then the spectral radius of \( A \) is \( |\lambda_1| \). Then \( \sigma(A^k) = \{\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k\} \) and the spectral radius of \( A^k \) is \( |\lambda_1|^k \). Since \( A^k > 0 \), \( |\lambda_1|^k \in \sigma(A^k) \) and \( |\lambda_2|^k < |\lambda_1|^k \). Hence \( |\lambda_1| \in \sigma(A) \) and \( |\lambda_2| < |\lambda_1| \).