4. DIAGONALISATION REVISITED

§4.1. Eigenvalues and Eigenvectors

Recall that an **eigenvector v** for a matrix A is a non-zero vector where A**v** = λ**v** for some scalar λ. The scalar λ is called the corresponding **eigenvalue**. The **characteristic polynomial** of A, denoted by $\chi_A(\lambda)$ is $|\lambda I - A|$ and its zeros are precisely the eigenvalues of A.

The **r'th trace** of the $n \times n$ matrix A is the sum of all the $r \times r$ sub-determinants whose diagonal coincides with that of the whole matrix. It is denoted by $\mathrm{tr}_r(A)$ and is the sum of $\binom{n}{r}$ $r \times r$ determinants. Special cases are $\mathrm{tr}_1(A)$, which is just the trace, tr(A) and $\mathrm{tr}_n(A)$ which is |A|.

Then $\chi_A(\lambda) =$
$\lambda^n - \mathrm{tr}(A)\lambda^{n-1} + \mathrm{tr}_2(A)\lambda^{n-2} - \mathrm{tr}_3(A)\lambda^{n-4} + ... + (-1)^n|A|$.
This is by far the easiest way to compute $\chi_A(\lambda)$. Solving $\chi_A(\lambda) = 0$ gives the eigenvalues and, for each one, the non-zero solutions give the corresponding eigenvalues.

If A is an $n \times n$ matrix and S = ($\mathbf{v}_1$, $\mathbf{v}_2$, ..., $\mathbf{v}_n$) is an invertible matrix whose columns are eigenvectors, where $A\mathbf{v}_i = \lambda_i\mathbf{v}_i$, and D = diag($\lambda_1$, λ_2, ..., λ_n) then $A = SDS^{-1}$ and we say that A is **diagonalisable**. In such cases $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ will be a basis of eigenvectors for A.

We can extend these concepts to linear transformations. If $f:U \to V$ is a linear transformation between two vector spaces, over the field F, and v is a non-zero element of U such that $f(v) = \lambda v$ for some $\lambda \in F$ then v is called an **eigenvector** for f and λ is the corresponding **eigenvalue**. This agrees with our previous definition by considering the linear transformation $f(v) = Av$, but it makes sense when there's no obvious matrix in the picture.

Example 1: Let U be the set of differentiable functions of one real variable x and let V be the space of all real functions of one real variable x, both considered as vector spaces over $\mathbb{R}$.

Let $D:U \to V$ be the linear transformation $D(f(x)) = \frac{df(x)}{dx}$.

An eigenvector is a function $f(x)$ such that $\frac{df(x)}{dx} = \lambda f(x)$ for some $\lambda \in \mathbb{R}$. Clearly the eigenvectors are the exponential functions, or more properly, the non-zero multiples of the exponential functions. For any real λ the functions $f(x) = Ce^{\lambda x}$, for any $C \neq 0$, are eigenvectors for D, with the real number λ being the corresponding eigenvalue.

Example 2: Let V be 3-dimensional Euclidean space, over $\mathbb{R}$. Let π be any plane in V that passes through the origin. Then reflection in π is a linear transformation, M. The eigenvalues are ± 1.

For $\lambda = 1$ the eigenvectors are the non-zero vectors in π because these are fixed by the reflection.

For $\lambda = -1$ the eigenvectors are the non-zero vectors that are perpendicular to the plane.

You shouldn't have your thinking about eigenvalues dominated by the equation $|\lambda I - A| = 0$. True, if you're given an $n \times n$ matrix, this is usually the best way to find its eigenvalues. And once you've found the eigenvalues you can then find the eigenvectors. But there are some matrices where it's easier to find the eigenvectors first, and *then* the eigenvalues.

But when you have to find the eigenvalues and eigenvectors of a linear transformation that isn't given by a matrix then $|\lambda I - A|$ won't make sense. How could you use determinants in example 1 or example 2? Of course we can always put in a basis, and represent the linear transformation by a matrix, but this is a long-winded way of going about it. Always think first of the equation $f(v) = \lambda v$.

If you have a linear transformation $f:V \to V$ on a finite-dimensional vector space V, the nicest basis, if such a basis exists, would be a basis of eigenvectors. For, if $\alpha = \{v_1, \ldots, v_n\}$ is a basis of eigenvectors for the $n \times n$ matrix A, with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$ then the matrix of f relative to this basis is simply the diagonal matrix $D = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}$.

But if the linear transformation is $f(v) = Av$ it will be expressed in terms of the standard basis β. To reach the diagonal matrix we simply need a change of basis.

If $S = \left[\frac{\alpha}{\beta}\right]$ then

$$A = \left[\frac{f(\beta)}{\beta}\right] = \left[\frac{\alpha}{\beta}\right]\left[\frac{f(\alpha)}{\alpha}\right]\left[\frac{\alpha}{\beta}\right]^{-1} = SAS^{-1}.$$

Of course, in order to get a basis of eigenvectors, we need to have the full complement of eigenvalues. In other words the characteristic polynomial must split completely into linear factors over the field. This will always be the case if our field is $\mathbb{C}$, the field of complex numbers. But even if the characteristic polynomial splits completely we may still fail to get a basis of eigenvectors.

Example 3: Let A = $\begin{pmatrix} 7 & 2 & -1 \\ -3 & 2 & 4 \\ -2 & -2 & 11 \end{pmatrix}$. The characteristic polynomial is $(\lambda - 10)(\lambda - 5)^2$.

But A – 5I = $\begin{pmatrix} 2 & 2 & -1 \\ -3 & -3 & 4 \\ -2 & -2 & 6 \end{pmatrix} \rightarrow \begin{pmatrix} 2 & 2 & -1 \\ -3 & -3 & 4 \\ 1 & 1 & -3 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 & -3 \\ 2 & 2 & -1 \\ -3 & -3 & 4 \end{pmatrix}$

$\rightarrow \begin{pmatrix} 1 & 1 & -3 \\ 0 & 0 & 5 \\ 0 & 0 & -5 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 & -3 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ so the eigenvectors for λ = 5 are all scalar multiples of $\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$. The dimension of the space of eigenvectors for the eigenvalue 5 is just 1, even though 5 is a double zero of the characteristic polynomial. The space of eigenvectors for λ = 10 also has dimension 1. Hence there's no basis of eigenvectors for $\mathbb{R}^3$. The eigenvectors only span a 2-dimensional subspace.

§4.2. Eigenspaces

If λ is an eigenvalue for the matrix A then the corresponding **eigenspace** is: $\mathbf{E_A(\lambda)} = \{\mathbf{v} \mid A\mathbf{v}) = \lambda v\}$. If λ is an eigenvalue the dimension of $E_A(\lambda)$ must be at least 1. The **total eigenspace** is the space spanned by all the eigenvectors and is denoted by $\mathbf{E_A}$.

We can extend the concept of eigenspaces to linear transformations, but for now we'll concentrate on matrices.

Theorem 1: The $n \times n$ matrix A is diagonalisable if and only if dim $E_A = n$.
Proof: dim $E_A = n$ if and only if there is a basis of eigenvectors. ✋☺

Theorem 2: The total eigenspace is the direct sum of the individual eigenspaces.
Proof: Suppose the distinct eigenvalues of the $n \times n$ matrix A are $\lambda_1, \ldots, \lambda_k$.
For each r let $\mathbf{v}_i \in E_A(\lambda_r)$. Then $A\mathbf{v}_r = \lambda_r\mathbf{v}$.
Suppose $\mathbf{v} = x_1\mathbf{v}_1 + \ldots + x_k\mathbf{v}_k = 0$.
For each r let:
$A_r = (A - \lambda_1 I) \ldots (A - \lambda_{r-1}I)(A - \lambda_{r+1}I) \ldots (A - \lambda_k I)$.
That is, A_r is the product of the factors $A - \lambda_s I$ for all s except for $s = r$. Then:

$$A_r\mathbf{v} = x_1A_r\mathbf{v}_1 + \ldots + x_{r-1}A_r\mathbf{v}_{r-1} + x_rA_r\mathbf{v}_r + \ldots + x_kA_r\mathbf{v}_r$$
$$= 0.$$

Hence $x_rA_r\mathbf{v}_r = 0$ because all the other terms are zero.

$$\text{But } A_r\mathbf{v}_r = (\lambda_r - \lambda_1) \ldots (\lambda_r - \lambda_{r-1}) (\lambda_r - \lambda_{r+1}) \ldots (\lambda_r - \lambda_k)\mathbf{v}_r$$
$$\neq 0.$$

It follows that $x_r = 0$.
Hence $E_A = E_A(\lambda_1) \oplus \ldots \oplus E_A(\lambda_k)$. ✋☺

Example 4: Let $A = \begin{pmatrix} 1 & 8 & -4 \\ 8 & 1 & 4 \\ -4 & 4 & 7 \end{pmatrix}$.

$\chi_A(\lambda) = |\lambda I - A| = (\lambda - 9)^2(\lambda + 9)$ so the eigenvalues are ± 9, with 9 being a repeated eigenvalue.

$E_A(9)$ is the null space of

$$A - 9I = \begin{pmatrix} -8 & 8 & -4 \\ 8 & -8 & 4 \\ -4 & 4 & -2 \end{pmatrix} \to \begin{pmatrix} 2 & -2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Hence $E_A(9) = \left\langle \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix} \right\rangle$.

$E_A(-9)$ is the null space of

$$A + 9I = \begin{pmatrix} 10 & 8 & -4 \\ 8 & 10 & 4 \\ -4 & 4 & 16 \end{pmatrix} \to \begin{pmatrix} 1 & -1 & -4 \\ 5 & 4 & -2 \\ 4 & 5 & 2 \end{pmatrix}$$

$$\to \begin{pmatrix} 1 & -1 & -4 \\ 0 & 9 & 18 \\ 0 & 9 & 18 \end{pmatrix} \to \begin{pmatrix} 1 & -1 & -4 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Hence $E_A(-9) = \left\langle \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix} \right\rangle$. Since $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix}$ are linearly independent. The quickest way to show this in this case is to observe that $\begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix}$ is orthogonal to the other two and so they are linearly independent and hence

$E_A = E_A(9) \oplus E_A(-9)$. Hence E_A has dimension 3 and we have a basis of eigenvectors. It follows that A is diagonalisable.

§4.3. Types of Diagonalisable Matrices

A square matrix is diagonalisable if and only if it has a basis of eigenvectors, or equivalently if every vector is a sum of eigenvectors. In many cases it is possible to conclude that the matrix is diagonalisable because of other properties.

Theorem 3: An $n \times n$ matrix is diagonalisable if it has n distinct eigenvalues.

Proof: If the eigenvalues are $\lambda_1, \ldots, \lambda_n$ then:

$$E_A = E_A(\lambda_1) \oplus \ldots \oplus E_A(\lambda_n)$$

and so $\dim(E_A) = n$. ✋☺

A **projection matrix** is one that satisfies the equation:

$$A^2 = A.$$

Example 5: Suppose that A is a 3×3 real matrix satisfying $A^2 = A$. Then it will be the matrix of a projection onto a plane in $\mathbb{R}^3$. Projecting a point twice will be the same as projection it once.

The eigenvalues of Aare the zeros of $\lambda^2 - \lambda$,

that is, 0 and 1.

$E_A(1)$ is this plane (points on the plane are fixed by the projection).

$E_A(0)$ is the line perpendicular to that plane.

Theorem 4: Projection matrices are diagonalisable.
Proof: Let A be a projection matrix.
Any vector **v** can be expressed as:

$$\mathbf{v} = (\mathbf{v} - A\mathbf{v}) + A\mathbf{v}.$$

Clearly $\mathbf{v} - A\mathbf{v} \in E_A(0)$ and $A\mathbf{v} \in E_A(1)$, so the eigenvectors span the whole space. ✍☺

A matrix A has **finite order** *m* if $A^m = I$.

Theorem 5: Matrices over $\mathbb{C}$ of finite order are diagonalisable.
Proof: Let A be an $n \times n$ matrix over $\mathbb{C}$ such that $A^m = I$.
Let $\theta = e^{2\pi i/m}$ and let **v** be any vector in $\mathbb{C}^n$.
Then $1 + \theta^r + \theta^{2r} + \ldots + \theta^{r(m-1)}$ is the sum of a geometric progression and so, if $\theta^r \neq 1$, this is:

$$\frac{\theta^{rm} - 1}{\theta^r - 1} = \frac{(\theta^m)^r - 1}{\theta^r - 1} = 0.$$

Then $m\mathbf{v}$

$$\left.\begin{array}{l}
= (\mathbf{v} + \theta A\mathbf{v} + \theta^2 A^2\mathbf{v} + \ldots + \theta^{m-1}A^{m-1}\mathbf{v}) \\
+ (\mathbf{v} + \theta^2 A\mathbf{v} + \theta^4 A^2\mathbf{v} + \ldots + \theta^{2(m-1)}A^{m-1}\mathbf{v}) \\
+ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
+ (\mathbf{v} + \theta^{m-1}A\mathbf{v} + \theta^{2(m-1)}A^2\mathbf{v} + \ldots + \theta^{(m-1)^2}A^{m-1}\mathbf{v}) \\
+ (\mathbf{v} + A\mathbf{v} + A^2\mathbf{v} + \ldots + A^{m-1}\mathbf{v}).
\end{array}\right\}$$

This is because the sum of each column, except the first, is zero. The sum of the *r*'th column is

$(1 + \theta^r + \theta^{2r} + \ldots + \theta^{r(m-1)})A^r\mathbf{v} = 0$ when $r > 1$.

Now $A(\mathbf{v} + \theta^r A\mathbf{v} + \theta^{2r}A^2\mathbf{v} + \ldots + \theta^{(m-1)r}A^{m-1}\mathbf{v})$
$= A\mathbf{v} + \theta^r A^2\mathbf{v} + \ldots + \theta^{(m-2)r}A^{m-1}\mathbf{v} + \theta^{(m-1)r}A^m\mathbf{v}$
$= A\mathbf{v} + \theta^r A^2\mathbf{v} + \ldots + \theta^{(m-2)r}A^{m-1}\mathbf{v} + \theta^{-r}\mathbf{v}$
$= \theta^{-r}\mathbf{v} + A\mathbf{v} + \theta^r A^2\mathbf{v} + \ldots + \theta^{(m-2)r}A^{m-1}\mathbf{v}$
$= \theta^{-r}(\mathbf{v} + \theta^r A\mathbf{v} + \theta^{2r}A^2\mathbf{v} + \ldots + \theta^{(m-1)r}A^{m-1}\mathbf{v})$.

Hence each term in the expression for **v** is an eigenvector, and so every vector is a sum of eigenvectors. ✋☺

Example 6: If $A^4 = I$ show that A is diagonalisable.
Solution: Every vector can be expresses as
$\mathbf{v} = \frac{1}{4}\,[(\mathbf{v} - A\mathbf{v}) + (\mathbf{v} - iA\mathbf{v}) + (\mathbf{v} + A\mathbf{v}) + (\mathbf{v} + iA\mathbf{v})]$
$\in \quad E_A(1) \; + \; E_A(i) \; + \; E_A(-1) \; + \; E_A(-i)$.

The following is an alternative proof of the above theorem. It's interesting that this proof uses some elementary calculus, including the Fundamental Theorem of Calculus.

Lemma: Let $\theta = e^{2\pi i/k}$.

Then $\frac{1}{x-1} + \frac{1}{x-\theta} + \frac{1}{x-\theta^2} + \ldots + \frac{1}{x-\theta^{n-1}} = \frac{nx^{n-1}}{x^n - 1}$.

Proof: The integral of the LHS is
$\log(x - 1) + \log(x - \theta) + \log(x - \theta^2) + \ldots + \log(x - \theta^{n-1})$
$= \log[(x - 1)(x - \theta)(x - \theta^2) \ldots (x - \theta^{n-1})]$

$= \log(x^n - 1)$ because of the above factorisation.

Differentiating we get $\dfrac{nx^{n-1}}{x^n - 1}$.

By the Fundamental Theorem of Calculus the derivative of the integral of the LHS is the LHS itself, and so the result follows.

For $r = 0, 1, 2, \ldots, n-1$ define:

$$e_r(x) = \frac{x^n - 1}{x - \theta^r}.$$

This has degree $n - 1$ and is the product of the terms $x - \theta^t$ for all t *except* for $t = r$.

Alternative proof of Theorem 5:

Suppose A is an $n \times n$ matrix over $\mathbb{C}$ where $A^m = I$.

The eigenvalues of A are the m-th roots of unity:

$$1, \theta, \theta^2, \ldots, \theta^{m-1} \text{ where } \theta = e^{2\pi i/m}.$$

From the second lemma we get:

$$e_0(A) + e_1(A) + e_2(A) + \ldots + e_{n-1}(A) = nA^{n-1}.$$

Clearly A is invertible and so:

$$nI = A^{1-n}[e_0(A) + e_1(A) + e_2(A) + \ldots + e_{n-1}(A)].$$

Let $\mathbf{v}$ be any column vector with n components.

Then $(A - \theta^r I)(A^{1-n} e_r(A)\mathbf{v})$

$= A^{1-n}(A - I)(A - \theta I)(A - \theta^2 I) \ldots (A - \theta^{n-1} I)\mathbf{v}$

$= A^{1-n}(A^n - 1)\mathbf{v}$

$= 0.$

It follows that, for all r, $A^{1-n}e_r(A)\mathbf{v}$ is an eigenvector for the eigenvalue θ^r, or it is zero.

Now $n\mathbf{v} = A^{1-n}e_0(A)\mathbf{v} + A^{1-n}e_1(A)\mathbf{v} + \ldots + A^{1-n}e_{n-1}(A)\mathbf{v}$ and so **v** is a linear combination of eigenvectors for A. The eigenvectors thus span the whole space and so A is diagonalizable. ✋☺

Example 7: Suppose $A^3 = I$.
The eigenvalues are 1, ω and ω^2.

$$e_0(x) = (x - \omega)(x - \omega^2),$$
$$e_1(x) = (x - 1)(x - \omega^2) \text{ and}$$
$$e_2(x) = (x - 1)(x - \omega).$$

So $3A^2 = (A - \omega I)(A - \omega^2 I) + (A - I)(A - \omega^2 I) + (A - I)(A - \omega I)$ which means that every vector **v** can be expressed as

$$\mathbf{v} = \frac{1}{3}A(A - \omega I)(A - \omega^2 I)\mathbf{v} + \frac{1}{3}A(A - I)(A - \omega^2 I)\mathbf{v} + \frac{1}{3}(A - I)(A - \omega I)\mathbf{v}.$$

A **cyclic matrix** is one of the form

$$\begin{pmatrix} a_1 & a_2 & \ldots & a_n \\ a_n & a_1 & \ldots & a_{n-1} \\ \ldots & \ldots & \ldots & \ldots \\ a_2 & a_3 & \ldots & a_1 \end{pmatrix}$$

As you move down the rows, the row moves to the right, and wraps around to the beginning.

Example 8: $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \\ 3 & 4 & 1 & 2 \\ 2 & 3 & 4 & 1 \end{pmatrix}$ is a cyclic matrix.

Theorem 6: Cyclic matrices are diagonalisable.

Proof: Let A = $\begin{pmatrix} a_1 & a_2 & \dots & a_n \\ a_n & a_1 & \dots & a_{n-1} \\ \dots & \dots & \dots & \dots \\ a_2 & a_3 & \dots & a_1 \end{pmatrix}$ and let $\theta = e^{2\pi i/n}$.

Then for all r,

$$A\begin{pmatrix} 1 \\ \theta^r \\ \theta^{2r} \\ \dots \\ \theta^{r(n-1)} \end{pmatrix} = (a_1 + a_2\theta^r + \dots + a_n\theta^{r(n-1)})\begin{pmatrix} 1 \\ \theta^r \\ \theta^{2r} \\ \dots \\ \theta^{r(n-1)} \end{pmatrix}.$$

It follows that each $\begin{pmatrix} 1 \\ \theta^r \\ \theta^{2r} \\ \dots \\ \theta^{r(n-1)} \end{pmatrix}$ is an eigenvector for A.

The determinant whose r'th column is $\begin{pmatrix} 1 \\ \theta^r \\ \theta^{2r} \\ \dots \\ \theta^{r(n-1)} \end{pmatrix}$ is a Vandermonde determinant $V(1, \theta, \theta^2, \dots, \theta^{n-1})$.

Since 1, θ, θ^2, …, θ^{n-1} are distinct this Vandermonde determinant is non-zero and so its columns are linearly independent. Hence there's a basis of eigenvectors. ☟☺

§4.4. Normal Matrices

Let A be a square matrix over ℂ. We define A^* to be the 'conjugate-transpose' of A. That is, if $A = (a_{ij})$ then

$A^* = (\overline{a_{ji}})$.

Example 9: If $A = \begin{pmatrix} 1+i & 2-i \\ 4+3i & 6-5i \end{pmatrix}$ find A^*.

Solution: $A^* = \begin{pmatrix} 1-i & 2+i \\ 4-3i & 6+5i \end{pmatrix}^T = \begin{pmatrix} 1-i & 4-3i \\ 2+i & 6+5i \end{pmatrix}$.

We call A^* the **adjoint** of A. But beware. There's another 'adjoint' that's sometimes used in finding inverses. That we denoted by adj(A).

Theorem 7:

(1) $(A + B)^* = A^* + B^*$;

(2) $(\lambda A)^* = \bar{\lambda} A^*$;

(3) $A^{**} = A$;

(4) $(AB)^* = B^* A^*$.

Proof: (1) – (3) are obvious.

(4) Let $A = (a_{ij})$ and $B = (b_{ij})$.

Then $(AB)^* = (\overline{AB})^T = (\bar{A}\,\bar{B})^T = \bar{B}^T\,\bar{A}^T = B^*A^*$. ☟☺

If we use the standard inner product on $\mathbb{C}^n$, where $\langle u \mid v\rangle = u^* v$ then we have the following fundamental property of adjoint.

Theorem 8: $\langle Au \mid v\rangle = \langle u \mid A^*v\rangle$ for all u, v.
Proof: $\langle Au \mid v\rangle = (Au)^*v = u^*A^*v = \langle u \mid A^*v\rangle$.

A **normal matrix** is a square matrix A, over $\mathbb{C}$, such that $AA^* = A^* A$. That is, it commutes with its conjugate transpose. There are several special types of normal matrix.

A Hermitian matrix is one where $A^* = A$. This includes real symmetric matrices. Clearly Hermitian matrices are normal since A commutes with itself.

A **unitary** matrix is one whose conjugate-transpose is its inverse, that is, A is unitary if $A^* = A^{-1}$, or in other words, if $AA^* = I$. Unitary matrices are normal since every invertible matrix commutes with its inverse. A real unitary matrix is called an **orthogonal matrix**. Here $A^{-1} = A^T$.

The columns of a complex matrix are orthonormal (mutually orthogonal and each of unit length) if and only if the matrix is unitary (similarly for rows). Hence if a matrix A has an orthonormal basis of eigenvectors it is diagonalisable. The corresponding eigenmatrix S is unitary and so $A = SDS^*$.

A matrix A is defined to be **skew-Hermitian** if A* = –A. Clearly skew-Hermitian matrices are normal. A real skew-Hermitian matrix is called skew-symmetric. It has the property that $A^T = -A$.

We can summarise the definitions of these special types of normal matrix as follows. We also provide 2 × 2 examples of each. The first example is of a normal matrix that doesn't fit into any of these special categories.

NAME	DEFINITION	EXAMPLE
Normal matrix	AA* = A*A	$\begin{pmatrix} 1+i & 0 \\ 0 & i \end{pmatrix}$
Hermitian matrix	A* = A	$\begin{pmatrix} 3 & 1-2i \\ 1+2i & -1 \end{pmatrix}$
Skew-Hermitian matrix	A* = –A	$\begin{pmatrix} 3i & 1-2i \\ -1-2i & i \end{pmatrix}$
Unitary matrix	$A^* = A^{-1}$	$\begin{pmatrix} \frac{1+i}{2} & \frac{1}{\sqrt{2}} \\ \frac{1-i}{2} & \frac{i}{\sqrt{2}} \end{pmatrix}$

For a real matrix we use different names.

NAME	DEFINITION	EXAMPLE
Real symmetric matrix	$A^T = A$	$\begin{pmatrix} 3 & -2 \\ -2 & -1 \end{pmatrix}$
Real skew-symmetric matrix	$A^T = -A$	$\begin{pmatrix} 0 & -2 \\ 2 & 0 \end{pmatrix}$
Orthogonal matrix	$A^T = A^{-1}$	$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$

Theorem 9: A square matrix over $\mathbb{C}$ is a normal matrix if and only if there exists a unitary matrix U and a diagonal matrix D such that $A = UDU^*$.
Proof: Suppose $A = UDU^*$. Then $A^* = UD^*U^*$.
So $AA^* = UDU^*UD^*U^* = U(DD^*)U^*$.
But D, D* are diagonal matrices so $DD^* = D^*D$.
Hence $AA^* = U(D^*D)U^* = (UD^*U^*)(UDU^*) = A^*A$. ✋☺

We now prove a number of properties of normal matrices.
Theorem 10: Suppose that A is a normal matrix. Then

$E_A(\lambda) \leq E_{A^*}(\overline{\lambda})$.
Proof: Let $\mathbf{v} \in E_A(\lambda)$.
Then $\langle A^*\mathbf{v} - \overline{\lambda}\mathbf{v} \mid A^*\mathbf{v} - \overline{\lambda}\mathbf{v}\rangle = \langle A^*\mathbf{v} \mid A^*\mathbf{v}\rangle - \langle \lambda\mathbf{v} \mid A^*\mathbf{v}\rangle - \langle A^*\mathbf{v} \mid \overline{\lambda}\mathbf{v}\rangle + \langle \overline{\lambda}\mathbf{v} \mid \overline{\lambda}\mathbf{v}\rangle$

$= \langle \mathbf{v} \mid AA^*\mathbf{v}\rangle - \bar{\lambda}\langle \mathbf{v} \mid A^*\mathbf{v}\rangle - \lambda\langle A^*\mathbf{v} \mid \mathbf{v}\rangle + \bar{\lambda}\lambda\langle \mathbf{v} \mid \mathbf{v}\rangle$

$= \langle \mathbf{v} \mid A^*A\mathbf{v}\rangle - \bar{\lambda}\langle A\mathbf{v} \mid \mathbf{v}\rangle - \lambda\langle \mathbf{v} \mid A\mathbf{v}\rangle + \bar{\lambda}\lambda\langle \mathbf{v} \mid \mathbf{v}\rangle$

$= \langle A\mathbf{v} \mid A\mathbf{v}\rangle - \bar{\lambda}\langle \lambda\mathbf{v} \mid \mathbf{v}\rangle - \lambda\langle \mathbf{v} \mid \lambda\mathbf{v}\rangle + \bar{\lambda}\lambda\langle \mathbf{v} \mid \mathbf{v}\rangle$

$= \langle \lambda\mathbf{v} \mid \lambda\mathbf{v}\rangle - \bar{\lambda}\lambda\langle \mathbf{v} \mid \mathbf{v}\rangle - \bar{\lambda}\lambda\langle \mathbf{v} \mid \mathbf{v}\rangle + \bar{\lambda}\lambda\langle \mathbf{v} \mid \mathbf{v}\rangle$

$= \bar{\lambda}\lambda\langle \mathbf{v} \mid \mathbf{v}\rangle - \bar{\lambda}\lambda\langle \mathbf{v} \mid \mathbf{v}\rangle$

$= 0.$

Hence $A^*\mathbf{v} - \lambda\mathbf{v} = 0$ and so $\mathbf{v} \in E_{A^*}(\bar{\lambda})$. ✋☺

Theorem 11: Suppose that A is a normal matrix. Then eigenvectors corresponding to different eigenvalues are orthogonal.

Proof: Let $\mathbf{v} \in E_A(\lambda)$ and $\mathbf{w} \in E_A(\mu)$ where $\lambda \neq \mu$.

Then $\lambda\langle \mathbf{v} \mid \mathbf{w}\rangle = \langle \lambda\mathbf{v} \mid \mathbf{w}\rangle$

$= \langle A\mathbf{v} \mid \mathbf{w}\rangle$

$= \langle \mathbf{v} \mid A^*\mathbf{w}\rangle$

$= \langle \mathbf{v} \mid \bar{\mu}\mathbf{w}\rangle = \mu\langle \mathbf{v} \mid \mathbf{w}\rangle.$

Since $\lambda \neq \mu$ then $\langle \mathbf{v} \mid \mathbf{w}\rangle = 0$. ✋☺

Theorem 12: Suppose that A is a normal matrix. Then E_A has an orthonormal basis of eigenvectors.

Proof: Let the distinct eigenvectors of A be $\lambda_1, \ldots, \lambda_k$.

Then $E_A = E_A(\lambda_1) + \ldots + E_A(\lambda_k)$.

For each i choose an orthonormal basis for $E_A(\lambda_i)$ and take the union of these bases. By Theorem 11 this basis will be an orthonormal basis of E_A. ✋☺

Theorem 13: Suppose that A is a normal matrix and let $\mathbf{v} \in E_A^{\perp}$. Then $A\mathbf{v} \in E_A^{\perp}$.
In other words, $E_A^{\perp}$ is invariant under multiplication by A.
Proof: Let $\mathbf{v} \in E_A^{\perp}$ and let $\mathbf{w} \in E_A$.
Then $\mathbf{w} = \mathbf{v}_1 + \ldots + \mathbf{v}_k$ where each $\mathbf{v}_i \in E_A(\lambda_i)$.
Hence $A^*\,\mathbf{w} = A^*\,\mathbf{v}_1 + \ldots + A^*\,\mathbf{v}_k$

$$= \bar{\lambda}_1\mathbf{v}_1 + \ldots + \bar{\lambda}_k v_k \text{ by Theorem 10.}$$
$$\in E_A.$$

Hence $\langle \mathbf{v} \mid A^*\mathbf{w}\rangle = 0$ and so $\langle A\mathbf{v} \mid \mathbf{w}\rangle = 0$.
Since this holds for all $\mathbf{w} \in E_A$, $A\mathbf{v} \in E_A^{\perp}$. ✋☺

Theorem 14: Suppose that A is a normal matrix. Then the map f: $E_A^{\perp} \to E_A^{\perp}$ defined by $f(\mathbf{v}) = A\mathbf{v}$ is a linear transformation
Proof: It is the restriction of $\mathbf{v} \to A\mathbf{v}$ to $E_A^{\perp}$.
The only thing that isn't obvious is the fact that $f(\mathbf{v}) \in E_A^{\perp}$ for all $\mathbf{v} \in E_A^{\perp}$, and that was proved in Theorem 13. ✋☺

Theorem 15: Suppose that A is a normal matrix. Then $E_A^{\perp} = 0$.
Proof: Suppose $\dim(E_A^{\perp}) \geq 1$.

Then f has at least one eigenvalue λ and a corresponding eigenvector **v**.
Clearly $\mathbf{v} \in E_A$, a contradiction. ✋☺

Theorem 16: Suppose that A is a normal matrix. Then A is unitarily diagonalisable.
Proof: If A is an $n \times n$ matrix then we've shown that $E_A = \mathbb{C}^n$ and hence $\mathbb{C}^n$ has an orthonormal basis of eigenvectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$.
If $U = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ is the matrix whose columns are the $\mathbf{v}_i$ then U is unitary and $U^{-1}AU = D$ for some diagonal matrix D. Since U is unitary we can also write U^{-1} as U^*. ✋☺

EXERCISES FOR CHAPTER 4

Exercise 1: Which of the following matrices are diagonalisable over ℂ? Give reasons.

(a) $\begin{pmatrix} 4 & 11 \\ -2 & -5 \end{pmatrix}$; (b) $\begin{pmatrix} 1 & -3 \\ 1 & 1 \end{pmatrix}$; (c) $\begin{pmatrix} 17 & -20 & -11 \\ 7 & -6 & -6 \\ 14 & -18 & -9 \end{pmatrix}$; (d) $\begin{pmatrix} 3 & 14 & -7 \\ 0 & 3 & 0 \\ 0 & 14 & -4 \end{pmatrix}$;

(e) $\begin{pmatrix} 3 & 7 & 2 & 9 \\ 9 & 3 & 7 & 2 \\ 2 & 9 & 3 & 7 \\ 7 & 2 & 9 & 3 \end{pmatrix}$; (f) $\begin{pmatrix} 5 & 6 & 1 & -2 \\ 6 & 2 & 4 & 2 \\ 1 & 4 & 0 & 8 \\ -2 & 2 & 8 & -1 \end{pmatrix}$.

Exercise 2: Let A have the form $\begin{pmatrix} 2 & -1 & 3 & 1 & 0 \\ 1 & * & * & * & * \\ -1 & * & * & * & * \\ 1 & * & * & * & * \\ -1 & * & * & * & * \end{pmatrix}$.

Suppose that the eigenvalues of A are either 0 or 1. Prove that A is not diagonalisable.

Exercise 3: Suppose A is a matrix such that $A^4 + A^3 + A^2 + A + I = 0$. Show that A is diagonalisable.

Exercise 4: What is wrong with the following 'solution' to the above exercise?
"The zeros of the polynomial $\lambda^4 + \lambda^3 + \lambda^2 + \lambda + 1$ are $e^{2\pi i/5}$, $e^{4\pi i/5}$, $e^{6\pi i/5}$, $e^{8\pi i/5}$, which are all distinct. Therefore if $A^4 + A^3 + A^2 + A + I = 0$ then its eigenvalues are distinct and hence A is diagonalisable."

Exercise 5: Suppose $A^3 = A$. Show that for all $\mathbf{v}$, $A^2\mathbf{v} - \mathbf{v}$, $A^2\mathbf{v} + A\mathbf{v}$ and $A^2\mathbf{v} - A\mathbf{v}$ are all eigenvectors for A. Show that $\mathbf{v} \in \langle A^2\mathbf{v} - \mathbf{v}, A^2\mathbf{v} + A\mathbf{v}, A^2\mathbf{v} - A\mathbf{v}\rangle$. Hence show that A is diagonalisable.

Exercise 6: Let $A = \frac{1}{9}\begin{pmatrix} 11 & -16 & 8 \\ -16 & 11 & 8 \\ 8 & 8 & 23 \end{pmatrix}$.

(a) Show that $\begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix}$ is an eigenvector for A.

(b) Find the eigenvalues of A.

(c) Extend $\left\{\begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix}\right\}$ to an orthogonal basis of eigenvectors for A.

(d) Write down an orthonormal basis of eigenvectors for A.

(e) Find an orthogonal matrix Q and a diagonal matrix D such that $A = QDQ^T$

Exercise 7: Consider the following real matrices:

$$A = \begin{pmatrix} 4 & 7 & -1 & 0 \\ 7 & 1 & 2 & 8 \\ -1 & 2 & 2 & 5 \\ 0 & 8 & 5 & 3 \end{pmatrix}; \; B = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{pmatrix}; \; C = \begin{pmatrix} 3 & 5 \\ -5 & 4 \end{pmatrix}$$

$$; \; D = \frac{1}{5}\begin{pmatrix} 3 & 4 \\ 4 & -3 \end{pmatrix};$$

$$E = \begin{pmatrix} 0 & 5 & -8 \\ -5 & 0 & 1 \\ 8 & -1 & 0 \end{pmatrix}.$$

For each of these determine which of the four adjectives REAL-SYMMETRIC, SKEW-SYMMETRIC and ORTHOGONAL applies.

Exercise 8: Consider the following non-real matrices:

$$A = \begin{pmatrix} 0 & 1-i & 3+5i \\ -1-i & 0 & 4-i \\ -3+5i & -4-i & 0 \end{pmatrix}; \quad B = \begin{pmatrix} 3 & 3+4i \\ 4-3i & 3 \end{pmatrix}; \quad C = \begin{pmatrix} 5 & 1-i & 3+5i \\ 1+i & 2 & -1+2i \\ 3-5i & -1-2i & 1 \end{pmatrix};$$

$$D = \frac{1}{2}\begin{pmatrix} 1-i & 1+i \\ 1-i & -1-i \end{pmatrix}; \quad E = \begin{pmatrix} 3 & 1-2i \\ 1+2i & -1 \end{pmatrix}.$$

For each of these determine which of the four adjectives NORMAL, HERMITIAN, SKEW-HERMITIAN and UNITARY applies.

SOLUTIONS FOR CHAPTER 4

Exercise 1: (a) trace = –1, determinant = 2 so $\chi(\lambda) = \lambda^2 + \lambda + 2$.
Since this has distinct zeros the matrix has distinct eigenvalues and so is diagonalisable.

(b) trace = 2, determinant = 4 so $\chi(\lambda) = \lambda^2 - 2\lambda + 4 = (\lambda - 2)^2$. If the matrix was diagonalisable it would have to be $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$, which it is not.

(c) Let A be the matrix. Then tr(A) = 2, $\mathrm{tr}_2(A) = -15$, |A| = – 36 so

$$\chi_A(\lambda) = \lambda^3 - 2\lambda^2 - 15\lambda + 36$$
$$= (\lambda - 3)^2(\lambda + 4).$$

$$A - 3I = \begin{pmatrix} 14 & -20 & -11 \\ 7 & -9 & -6 \\ 14 & -18 & -12 \end{pmatrix} \rightarrow \begin{pmatrix} 7 & -9 & -6 \\ 14 & -20 & -11 \\ 14 & -18 & -12 \end{pmatrix} \rightarrow \begin{pmatrix} 7 & -9 & -6 \\ 0 & -2 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Since there is only one row of zeros, $\dim E_A(3) = 1$. Clearly $\dim E_A(-4) = 1$ so $\dim E_A = 2$.
Hence A is not diagonalisable. (The eigenvectors don't span $\mathbb{R}^3$.)

(d) The matrix is cyclic and so it is diagonalisable.

(e) The matrix is real symmetric and so is diagonalisable.

Exercise 2: If A was diagonalisable it would have to satisfy the equation $A^2 = A$. But the 1-1 component of A^2 is 1, not 3.

Exercise 3: $A^5 - I = (A - I)(A^4 + A^3 + A^2 + A + I) = 0$, so A is a matrix of finite order and hence is diagonalisable.

Exercise 4: The flaw is that although the fifth roots of unity are distinct, the matrix A might have one of them repeated.

Exercise 5:
$$\begin{aligned} A(A^2\mathbf{v} - \mathbf{v}) &= A^3\mathbf{v} - A\mathbf{v} \\ &= A\mathbf{v} - A\mathbf{v} \\ &= 0, \text{ so } A^2\mathbf{v} - \mathbf{v} \in E_A(0). \end{aligned}$$

$$\begin{aligned} A(A^2\mathbf{v} + A\mathbf{v}) &= A^3\mathbf{v} + A^2\mathbf{v} \\ &= A\mathbf{v} + A^2\mathbf{v} = 0, \text{ so } A^2\mathbf{v} + A\mathbf{v} \in E_A(1). \end{aligned}$$

$$\begin{aligned} A(A^2\mathbf{v} - A\mathbf{v}) &= A^3\mathbf{v} - A^2\mathbf{v} \\ &= A\mathbf{v} - A^2\mathbf{v} \\ &= 0, \text{ so } A^2\mathbf{v} - A\mathbf{v} \in E_A(-1). \end{aligned}$$

$\mathbf{v} = -(A^2\mathbf{v} - \mathbf{v}) - ½(A^2\mathbf{v} + A\mathbf{v}) - ½(A^2\mathbf{v} - A\mathbf{v})$.
Since every vector is a linear combination of eigenvectors A is diagonalisable.

Exercise 6:

(a) $\frac{1}{9}\begin{pmatrix} 11 & -16 & 8 \\ -16 & 11 & 8 \\ 8 & 8 & 23 \end{pmatrix}\begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} -3 \\ 6 \\ 6 \end{pmatrix} = 3\begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix}$ so $\begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix}$ is an eigenvector for A.

(b) tr(A) = 5; $tr_2(A) = 3$; |A| = –9 so $\chi_A(\lambda) = \lambda^3 - 5\lambda^2 + 3\lambda + 9$.
We know from (a) that $\lambda = 3$ is an eigenvalue. We use this fact to factorise $\chi_A(\lambda)$ as $(\lambda - 3)^2(\lambda + 1)$. Hence the eigenvalues are 3, 3, –1.

(c) $A - 3I = \frac{1}{9}\left[\begin{pmatrix} 11 & -16 & 8 \\ -16 & 11 & 8 \\ 8 & 8 & 23 \end{pmatrix} - \begin{pmatrix} 27 & 0 & 0 \\ 0 & 27 & 0 \\ 0 & 0 & 27 \end{pmatrix}\right] = \frac{1}{9}\left[\begin{pmatrix} -16 & -16 & 8 \\ -16 & -16 & 8 \\ 8 & 8 & -4 \end{pmatrix}\right] \to \begin{pmatrix} 2 & 2 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

Let z = 2h, y = 2k. Then $x = h - 2k$ and so a vector in $E_A(3)$ has the form $\begin{pmatrix} h-2k \\ 2k \\ 2h \end{pmatrix}$.

We want such a vector to be orthogonal to $\begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix}$ so we require $2k - h + 4k + 4h = 0$.
This gives $6k + 3h = 0$, that is, $h = -2k$.

Our vector is now $\begin{pmatrix} -4k \\ 2k \\ -4k \end{pmatrix}$. Choosing k = ½ gives the vector $\begin{pmatrix} -2 \\ 1 \\ -2 \end{pmatrix}$. So $\left\{\begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ -2 \end{pmatrix}\right\}$ is an orthogonal basis for $E_A(3)$.

$$A + I = \frac{1}{9}\left[\begin{pmatrix} 11 & -16 & 8 \\ -16 & 11 & 8 \\ 8 & 8 & 23 \end{pmatrix} + \begin{pmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{pmatrix}\right]$$

$$= \frac{1}{9}\left[\begin{pmatrix} 20 & -16 & 8 \\ -16 & 20 & 8 \\ 8 & 8 & 32 \end{pmatrix}\right] \rightarrow \begin{pmatrix} 1 & 1 & 4 \\ 20 & -16 & 8 \\ -16 & 20 & 8 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 & 4 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

So $\begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix}$ is an eigenvector for $\lambda = -1$.

As expected, it's orthogonal to the vectors in $E_A(3)$.

Hence $\left\{ \begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix} \right\}$ is an orthogonal basis of eigenvectors for A.

(d) Each of these vectors has length 3, so $\left\{ \frac{1}{3}\begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix}, \frac{1}{3}\begin{pmatrix} -2 \\ 1 \\ -2 \end{pmatrix}, \frac{1}{3}\begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix} \right\}$ is an orthonormal basis of eigenvectors.

(e) Let Q = $\frac{1}{3}\begin{pmatrix} -1 & -2 & -2 \\ 2 & 1 & 2 \\ 2 & -2 & 1 \end{pmatrix}$ and D = $\begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -1 \end{pmatrix}$.

Then AQ = QD and so $A = QDQ^{-1} = QDQ^T$.

Exercise 7:

$$A^T = A;\ BB^T = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} =$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix};\ C^T = \begin{pmatrix} 3 & -5 \\ 5 & 4 \end{pmatrix};$$

$$DD^T = D^2 = \frac{1}{25}\begin{pmatrix} 3 & 4 \\ 4 & -3 \end{pmatrix}\begin{pmatrix} 3 & 4 \\ 4 & -3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix};\ E^T =$$

$$\begin{pmatrix} 0 & -5 & 8 \\ 5 & 0 & -1 \\ -8 & 1 & 0 \end{pmatrix} = -E.$$

	A	**B**	**C**	**D**	**E**
real-symmetric	√			√	
skew-symmetric					√
orthogonal		√		√	

Exercise 8: $A^* = A$; $BB^* = B^* B = \begin{pmatrix} 34 & 21+21i \\ 21-21i & 34 \end{pmatrix}$; $C^* = C$;

$$DD^* = \frac{1}{4}\begin{pmatrix} 1-i & 1+i \\ 1-i & -1-i \end{pmatrix}\begin{pmatrix} 1+i & 1+i \\ 1-i & -1+i \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

	A	B	C	D
normal	√	√	√	√
Hermitian			√	
skew-Hermitian	√			
unitary				√