LECTURE 6

Hermitian matrices

Last time, we saw how to diagonalise a matrix by a similarity transformation. This has applications throughout chemistry, including in quantum mechanics, group theory, chemical kinetics and (as we saw in the second example last time), constructing the normal modes of vibration of a molecule.

I mentioned that there was a caveat: it doesn't always work. Let's see why.

When things go wrong: defective matrices

Suppose we wanted to solve the following system of differential equations:

$$\begin{aligned}\frac{\mathrm{d}y_1}{\mathrm{d}t} &= y_1 + y_2 \\ \frac{\mathrm{d}y_2}{\mathrm{d}t} &= y_2.\end{aligned}$$

In matrix form, this becomes

$$\mathbf{y}' = \mathsf{A}\mathbf{y} \tag{6.1}$$

with

$$\mathsf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

As we saw in Lecture 4, this matrix has one distinct eigenvector:

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \text{with } \lambda_1 = 1.$$

We therefore don't have enough eigenvectors to build a 2×2 matrix S, so the procedure in the previous section doesn't work.

"Aha!", you might say. Let's just use a multiple of $\mathbf{v}_1$ as the second column of S, because we know that this would also be an eigenvector. Unfortunately this won't work either. The matrix S needs to be invertible to write

$$\mathsf{S}^{-1}\mathsf{A}\mathsf{S} = \mathsf{D}. \tag{6.2}$$

If the columns of S are linearly dependent, then S will necessarily have a determinant of zero and not be invertible.

So, we see that some matrices cannot be diagonalised. They don't have enough linearly dependent eigenvectors. We call these matrices *defective*. On the other hand, matrices that can be diagonalised are called *diagonalisable*.

In general, there is no simple way to tell whether a matrix is diagonalisable or defective. One just has to calculate the eigenvalues and eigenvectors, and then count whether there are enough *linearly-independent* eigenvectors to form S.

However, there is a beautiful theorem of linear algebra, known as the *spectral theorem*, that guarantees diagonalisability for certain matrices. I will state a special case of the result:

> All Hermitian matrices are diagonalisable.

The next section will explain what we mean by a Hermitian matrix. We'll see that we've actually encountered many already.

Hermitian matrices

Up to now, we've looked only at matrices involving real numbers. Now it is time to generalise to complex numbers. There are two reasons for this:

1. The matrices of quantum mechanics are complex.
2. It is actually easier to prove certain results if we consider complex matrices.

So, we will now allow matrices to have complex numbers as elements. This requires us to introduce some new definitions.

Given a complex matrix A, we construct its *complex conjugate* A^* by taking the complex conjugate of every element. For example:

$$\begin{pmatrix} \mathrm{i} & -\mathrm{i} \\ 0 & 3 \end{pmatrix}^* = \begin{pmatrix} -\mathrm{i} & \mathrm{i} \\ 0 & 3 \end{pmatrix}.$$

As explained earlier, the adjoint matrix is not the same as the adjugate matrix that arose when calculating the matrix inverse. Some people use the word 'adjoint' to refer to both.

We then define the *adjoint* of a complex matrix by taking the complex conjugate and the transpose. A 'dagger' symbol is used to represent this:

$$\mathsf{A}^\dagger = \left(\mathsf{A}^*\right)^\mathsf{T}.$$

i.e.

$$\begin{aligned} \begin{pmatrix} \mathrm{i} & -\mathrm{i} \\ 0 & 3 \end{pmatrix}^\dagger &= \begin{pmatrix} -\mathrm{i} & \mathrm{i} \\ 0 & 3 \end{pmatrix}^\mathsf{T} \\ &= \begin{pmatrix} -\mathrm{i} & 0 \\ \mathrm{i} & 3 \end{pmatrix}. \end{aligned}$$

You can convince yourself that it doesn't matter which way round one takes the complex conjugate and the transpose: the resulting adjoint matrix is the same either way. Using what we know about complex conjugates and transposes, it follows that

$$\begin{aligned} (\mathsf{AB})^\dagger &= \left(\mathsf{A}^*\mathsf{B}^*\right)^\mathsf{T} \\ &= \left(\mathsf{B}^*\right)^\mathsf{T}\left(\mathsf{A}^*\right)^\mathsf{T} \\ &= \mathsf{B}^\dagger\mathsf{A}^\dagger. \end{aligned}$$

Hence, using the associativity of matrix multiplication, it follows that

$$(\mathsf{AB} \cdots \mathsf{Z})^\dagger = \mathsf{Z}^\dagger \cdots \mathsf{B}^\dagger \mathsf{A}^\dagger, \tag{6.3}$$

just as this follows for the transpose of a product of real matrices. As we shall see, the adjoint will be the natural generalisation of the transpose when dealing with complex matrices.

We define a *Hermitian* matrix to be one that is equal to its adjoint:

$$\boxed{\mathsf{H} = \mathsf{H}^\dagger \qquad \text{Hermitian matrix} \qquad (6.4)}$$

For example, the following matrices are Hermitian. You should check them both for yourself, to make sure you fully understand the definition.

$$\begin{pmatrix} 1 & \mathrm{i} \\ -\mathrm{i} & -1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 1+\mathrm{i} \\ 0 & 2 & \mathrm{i} \\ 1-\mathrm{i} & -\mathrm{i} & 3 \end{pmatrix}$$

Complex vectors

We saw a few lectures ago that vectors are really just a special case of matrices. By introducing complex numbers into matrices, we have therefore implicitly made complex vectors a possibility too.

Let's make things explicit. We shall say that column vectors and row vectors can now have complex elements, e.g.

$$\begin{pmatrix} 0 \\ \mathrm{i} \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} -\mathrm{i} & 2\mathrm{i} & -3\mathrm{i} \end{pmatrix}.$$

It is impossible to visualise these as arrows in n-dimensional space. But they can still be treated algebraically. It seems reasonable to define addition and subtraction of complex vectors in the same as for real vectors. But what about multiplication?

Last Term, we defined the scalar product of two real vectors in terms of their lengths and the angle between them. We can't do this for complex vectors, because the concepts of length and angle don't make sense if we can't draw the vectors as arrows. So, we'll switch to the other form of the dot product.

For real (three-dimensional) vectors, we saw that

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + a_3 b_3.$$

This has the important property that

$$|\mathbf{a}|^2 = \mathbf{a} \cdot \mathbf{a} = |a_1|^2 + |a_2|^2 + |a_3|^2 \geq 0.$$

Specifically, $\mathbf{a} \cdot \mathbf{a} = 0$ only if $\mathbf{a} = \mathbf{0}$. On the other hand, if

$$\mathbf{a} \cdot \mathbf{b} = 0$$

for two different vectors, we say that the vectors are orthogonal.

With all this in mind, we *define* the scalar product of two complex vectors to be

Notice that complex vectors don't satisfy $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$ in general.

$$\mathbf{u} \cdot \mathbf{v} = u_1^* v_1 + u_2^* v_2 + \cdots + u_n^* v_n.$$

The complex conjugates are included so that

$$\begin{aligned} \mathbf{u} \cdot \mathbf{u} &= u_1^* u_1 + u_2^* u_2 + \cdots + u_n^* u_n \\ &= |u_1|^2 + |u_2|^2 + \cdots + |u_n|^2 \\ &\geq 0, \end{aligned}$$

with

$$\mathbf{u} \cdot \mathbf{u} = 0 \qquad \text{only if} \qquad \mathbf{u} = \mathbf{0}. \tag{6.5}$$

as for the real case. This means we can define the 'length' of a complex vector to be

$$|\mathbf{u}|^2 = \mathbf{u} \cdot \mathbf{u},$$

In other words, if the complex conjugates were not included in the definition of the scalar product, we could find non-zero vectors that would be 'orthogonal' to themselves. The definition of orthogonality wouldn't be very useful in that case.

and we can say that that two complex vectors are orthogonal if

$$\mathbf{u} \cdot \mathbf{v} = 0, \tag{6.6}$$

Finally, we saw a couple of lectures ago that

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{a}^\mathsf{T} \mathbf{b}.$$

when **a** and **b** are column vectors. It is easy to show that the generalisation to complex vectors is

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^\dagger \mathbf{v}.$$

We'll need this in the next section.

Notice that the new definitions for complex vectors reduce exactly to our normal definitions when the vectors are real. Complex conjugates can be dropped, and the adjoint therefore reduces to the transpose. So what we prove now for the general, complex case will also apply when everything is real.

Eigenvalues of Hermitian matrices

This is the first of two proofs for today. We're going to show that the eigenvalues of Hermitian matrices must be real, even though the matrices themselves can be complex.

Suppose **v** is an eigenvector of the Hermitian matrix H:

$$\mathsf{H}\mathbf{v} = \lambda \mathbf{v}.$$

Then, we can pre-multiply both sides of the equation by $\mathbf{v}^\dagger$ to get

$$\mathbf{v}^\dagger \mathsf{H} \mathbf{v} = \lambda \mathbf{v}^\dagger \mathbf{v}. \tag{6.7}$$

Now, we'll take the adjoint of both sides:

$$\left(\mathbf{v}^\dagger \mathsf{H} \mathbf{v}\right)^\dagger = \left(\lambda \mathbf{v}^\dagger \mathbf{v}\right)^\dagger.$$

Using (6.3), this turns into

$$\mathbf{v}^\dagger \mathsf{H}^\dagger \mathbf{v} = \mathbf{v}^\dagger \mathbf{v} \lambda^\dagger.$$

λ is just a scalar, so it is unaffected by the transpose part of the adjoint. That means that $\lambda^\dagger = \lambda^*$, and we can bring it to the front of the right-hand side to get

$$\mathbf{v}^\dagger \mathsf{H}^\dagger \mathbf{v} = \lambda^* \mathbf{v}^\dagger \mathbf{v}.$$

Finally, using that H is Hermitian, (6.4), the LHS reduces to

$$\mathbf{v}^\dagger \mathsf{H} \mathbf{v} = \lambda^* \mathbf{v}^\dagger \mathbf{v}. \tag{6.8}$$

Notice that (6.8) has exactly the same LHS as (6.7). If we take one away from the other, then

$$0 = (\lambda - \lambda^*) \mathbf{v}^\dagger \mathbf{v}. \tag{6.9}$$

But we saw earlier that

$$\begin{aligned} \mathbf{v}^\dagger \mathbf{v} &= \mathbf{v} \cdot \mathbf{v} \\ &= 0 \quad \text{only if} \quad \mathbf{v} = \mathbf{0}. \end{aligned}$$

Therefore, given that $\mathbf{0}$ is not considered to be an eigenvector, it must be that

$$\mathbf{v}^\dagger \mathbf{v} \neq 0$$

for any eigenvector $\mathbf{v}$. Hence, from (6.9), we need $(\lambda - \lambda^*) = 0$ in order to satisfy the equation. Therefore

$$\lambda = \lambda^*$$

which proves that λ must be real for any eigenvector.

Example 1. Find the eigenvalues of

$$\mathsf{H} = \begin{pmatrix} 1 & \mathrm{i} \\ -\mathrm{i} & 1 \end{pmatrix}$$

As usual, we set the characteristic equation $\det(\mathsf{H} - \lambda \mathrm{I}) = 0$ to find the eigenvalues. This leads to

$$(1 - \lambda)^2 + \mathrm{i}^2 = 0$$

which gives

$$(1 - \lambda)^2 = 1$$

and hence $\lambda = 0$ or $\lambda = 2$. Both eigenvalues are real, even though H has complex entries, because H is Hermitian.

Eigenvectors of a Hermitian matrix

The second proof builds on the first. We will show that eigenvectors of a Hermitian matrix, *that correspond to different eigenvalues*, must be orthogonal. The method of the proof follows a similar structure to that above.

The eigenvectors of a Hermitian matrix $\mathbf{H}$ satisfy

$$\mathsf{H} \mathbf{v}_i = \lambda_i \mathbf{v}_i.$$

This means that we can introduce a second eigenvector and write

$$\mathbf{v}_j^\dagger \mathsf{H}\mathbf{v}_i = \lambda_i \mathbf{v}_j^\dagger \mathbf{v}_i. \tag{6.10}$$

Alternatively, taking the adjoint of both sides, and using $\mathsf{H} = \mathsf{H}^\dagger$, gives

$$\mathbf{v}_i^\dagger \mathsf{H}\mathbf{v}_j = \lambda_i^* \mathbf{v}_i^\dagger \mathbf{v}_j.$$

However, since we showed that all eigenvalues are real, this simplifies to

$$\mathbf{v}_i^\dagger \mathsf{H}\mathbf{v}_j = \lambda_i \mathbf{v}_i^\dagger \mathbf{v}_j. \tag{6.11}$$

The trick now is to subtract one equation from the other, as before. To make things cancel out nicely, we will first swap the labels i and j in (6.11) to get

If swapping the labels looks dodgy to you, the alternative way to get to (6.12) is to start again but from

$$\mathsf{H}\mathbf{v}_j = \lambda_j \mathbf{v}_j,$$

and do the steps leading to (6.11) after premultiplying by $\mathbf{v}_i^\dagger$.

$$\mathbf{v}_j^\dagger \mathsf{H}\mathbf{v}_i = \lambda_j \mathbf{v}_j^\dagger \mathbf{v}_i. \tag{6.12}$$

Taking this from (6.10) gives

$$0 = (\lambda_i - \lambda_j)\mathbf{v}_j^\dagger \mathbf{v}_i \tag{6.13}$$

If the eigenvectors have different eigenvalues, i.e. $\lambda_i \neq \lambda_j$, then it must be the case that

$$\mathbf{v}_j^\dagger \mathbf{v}_i = 0.$$

Hence $\mathbf{v}_j \cdot \mathbf{v}_i = 0$, i.e. the eigenvectors are orthogonal.

Notice that the proof says nothing about eigenvectors with the same eigenvalue. Equation (6.13) is satisfied when $\lambda_i = \lambda_j$, regardless of whether the eigenvectors are orthogonal or not. The consequences of this will rear their collective head in the next lecture.

Example 2. Find the eigenvectors of H in Example 1, and show they are orthogonal.

For $\lambda = 0$, we want to solve

$$\begin{pmatrix} 1 & \mathrm{i} \\ -\mathrm{i} & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{0}.$$

Hence

$$x + \mathrm{i}y = 0 \tag{6.14}$$
$$-\mathrm{i}x + y = 0 \tag{6.15}$$

Notice, as usual, that these equations are linearly-dependent. The second is just $-\mathrm{i}$ times the first. From the second, we see that $y = \mathrm{i}x$ and hence an eigenvector is

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ \mathrm{i} \end{pmatrix}.$$

For $\lambda = 2$, we want to solve

$$\begin{pmatrix} -1 & \mathrm{i} \\ -\mathrm{i} & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{0}.$$

This leads to the eigenvector

$$\mathbf{v}_2 = \begin{pmatrix} \mathrm{i} \\ 1 \end{pmatrix}.$$

To show that they are orthogonal, we use the definition in (6.6):

$$\begin{aligned} \mathbf{v}_1 \cdot \mathbf{v}_2 &= \mathbf{v}_1^\dagger \mathbf{v}_2 \\ &= \begin{pmatrix} 1 & -\mathrm{i} \end{pmatrix} \begin{pmatrix} \mathrm{i} \\ 1 \end{pmatrix} \\ &= 1 \times \mathrm{i} + (-\mathrm{i}) \times 1 \\ &= 0. \end{aligned}$$

Notice that it was important to take the complex conjugate as part of the adjoint operation on $\mathbf{v}_1$. We would not have obtained zero had we used the normal definition of the dot product for real vectors.

Real symmetric matrices

The following matrix is clearly Hermitian: taking the complex conjugate makes no difference, and neither does taking the transpose.

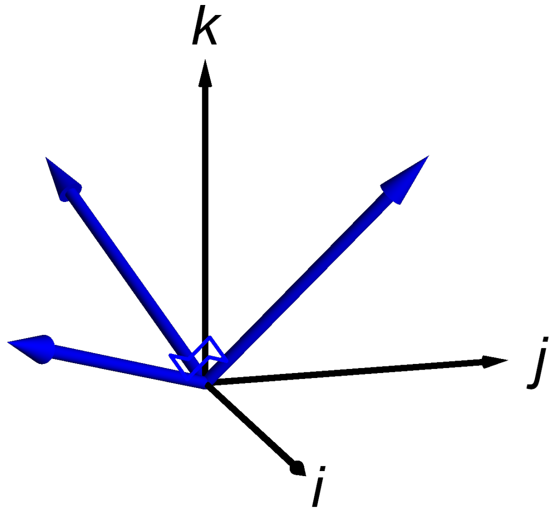
$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{pmatrix}$$

It is an example of a *real symmetric* matrix. It's 'real' because all its elements are real numbers. It's 'symmetric' because it satisfies the following condition:

$$\mathsf{A} = \mathsf{A}^\mathsf{T} \qquad \text{Symmetric matrix}$$

Real symmetric matrices are a special case of Hermitian matrices, so their eigenvalues are guaranteed to be real, and (for different eigenvalues) their eigenvectors are guaranteed to be orthogonal.

In this case, the matrix has a ghastly set of eigenvectors and eigenvalues. The eigenvalues are $\lambda_1 \simeq 11.34$, $\lambda_2 \simeq -0.52$ and $\lambda_3 \simeq 0.17$. Since all the eigenvalues are different, all the eigenvectors must be orthogonal. We can confirm that in this case by plotting the eigenvectors: they look like the figure below.


k
j
i