LECTURE 7

Orthogonal diagonalisation of symmetric matrices

Last time, we saw that, if H is Hermitian, any of its eigenvectors corresponding to different eigenvalues must be orthogonal.

We also stated a version of the spectral theorem, which said that Hermitian matrices are always diagonalisable. Remember: an $n \times n$ matrix is diagonalisable if it has n linearly-independent eigenvectors, which allows us to construct the $n \times n$ matrix S that satisfies:

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

If an $n \times n$ Hermitian matrix has n different eigenvalues, diagonalisability follows straightforwardly. There will be n *orthogonal* eigenvectors corresponding to the n different eigenvalues, so clearly these eigenvectors are linearly independent.

If you're worried that an eigenvector might not necessarily exist for each eigenvalue, look back to Lecture 6 from last Term. We saw that if

$$(\mathsf{A} - \lambda \mathrm{I})\mathbf{v} = \mathbf{0}$$

and $\det(\mathsf{A} - \lambda \mathrm{I}) = 0$, then non-trivial solutions *must* exist for $\mathbf{v}$.

However, the argument breaks down if any of the eigenvalues come from repeated roots of the auxiliary equation. Then, we won't have n different eigenvalues. What happens in this case?

For simplicity, we will explore this issue for real, symmetric matrices (which, as we saw last time, are a special case of Hermitian matrices). Analogous arguments can be made for complex Hermitian matrices, but we will leave the result for a simple footnote at the end.

Repeated roots of the characteristic equation

To see what can happen when roots of the characteristic equation are repeated, let's examine the two cases below:

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

We'll find the eigenvalues and eigenvectors of each.

For A, the characteristic equation is

$$(1 - \lambda)^2 - 1 \times 0 = 0$$

which simplifies to $(1 - \lambda)^2 = 0$. Hence $\lambda = 1$ is a repeated root. We shall say that the eigenvalue $\lambda_1 = 1$ has an *algebraic multiplicity* of 2, because it comes from a double root of the characteristic polynomial. More generally, we say that any eigenvalue corresponding to a repeated root is *degenerate*.

Turning to the eigenvectors, we need

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

which gives $y = 0$. Hence, the eigenvectors all take the form

$$\begin{pmatrix} x \\ 0 \end{pmatrix}.$$

Geometrically, they all point along a single line. We say that the 'eigenspace' (basically the collection of all eigenvectors) of this eigenvalue is one-dimensional, and that the eigenvalue has a *geometric multiplicity* of 1. As usual, we can pick a convenient eigenvector such as

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The eigenspace of an eigenvalue is more formally defined as the set of all possible eigenvectors with that eigenvalue, *plus the zero vector* (which is not counted as an eigenvector).

Now, let's examine B. Its characteristic equation is

$$(1-\lambda)^2 - 0 \times 0 = 0$$

which also has a double solution of $\lambda = 1$. So $\lambda_1 = 1$ is an eigenvalue of B with algebraic multiplicity 2.

On the other hand, when we come to the eigenvectors, we now need to solve

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Any values of x and y are allowed, so the eigenvector can be any vector of the form

$$\begin{pmatrix} x \\ y \end{pmatrix}.$$

with arbitrary x and y. The eigenspace is therefore two-dimensional, because the vector can point anywhere in the plane, and hence the eigenvalue of 1 has geometric multiplicity 2.

To summarise, we have found that:

Matrix	**Eigenvalue**	**Alg. mult.**	**Geom. mult.**
$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$	$\lambda = 1$	2	1
$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\lambda = 1$	2	2

So we see that the geometric multiplicity does not necessarily equal the algebraic multiplicity. Matrix A cannot be diagonalised: we cannot construct two linearly independent eigenvectors for S when they can only be taken from a one-dimensional eigenspace.

It turns out, though, that the *geometric multiplicity of an eigenvalue always equals its algebraic multiplicity* for a real symmetric matrix. That is, if the eigenvalue comes from an repeated root of the characteristic polynomial with multiplicity n, its eigenspace will be n-dimensional. The proof of this is a bit tricky, so I have relegated it to an appendix for those who are interested.

Indeed, it is true for any Hermitian matrix.

This is why symmetric matrices are diagonalisable. An $n \times n$ symmetric matrix will have a characteristic polynomial with n real roots, if we count repeated roots according to the number of times they appear. Eigenvectors with different eigenvalues will be orthogonal and therefore, automatically linearly independent. Eigenvectors corresponding to a repeated root of the characteristic polynomial (say of multiplicity m) will not necessarily be orthogonal or even linearly independent, but the eigenspace will be m-dimensional and so we will be able to *choose* m linearly-dependent vectors. In fact, we will be able to choose m *orthonormal vectors* within that eigenspace. (This is known as constructing an *orthonormal basis* for the eigenspace.) We'll see later why using orthonormal vectors is particularly nice.

The existence of n roots is guaranteed by the fundamental theorem of algebra. In this case, all n roots must be real, because all the eigenvalues of a symmetric matrix are real (as proved last time).

Anyway, enough words. Let's see how this works for a more interesting example.

Example 1. Find a matrix that diagonalises

$$\mathsf{A} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

I'll leave it up to you to show that the characteristic equation is

$$(\lambda - 1)^2(\lambda - 4) = 0.$$

Hence the eigenvalues are 4 and 1. Notice that $\lambda = 1$ is a double root, so it has an algebraic multiplicity of 2.

Now, we look for the eigenvectors. For $\lambda = 4$, it is the same as we've done many times before. Substituting the eigenvalue into the eigenvector equation gives

$$\begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

From here, it is easy to show that $x = y = z$ and hence an eigenvector is

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

For $\lambda = 1$, the equation we need to solve is

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

This gives us

$$x + y + z = 0$$

three times. (Notice that *two* of the equations have become redundant.) We now have a single constraint to satisfy, i.e. $z = -x - y$, so any eigenvector of the form

$$\begin{pmatrix} x \\ y \\ -x - y \end{pmatrix}$$

is an eigenvector with eigenvalue 1.

Geometrically, the eigenvector can point anywhere in the plane $x + y + z = 0$. Therefore, the geometric multiplicity of λ_2 is 2. We should not be surprised, because we said that the algebraic multiplicity was 2 earlier, and we stated that the two are always the same for symmetric matrices.

To diagonalise A, we need to find three linearly-independent eigenvectors. Therefore, let's choose

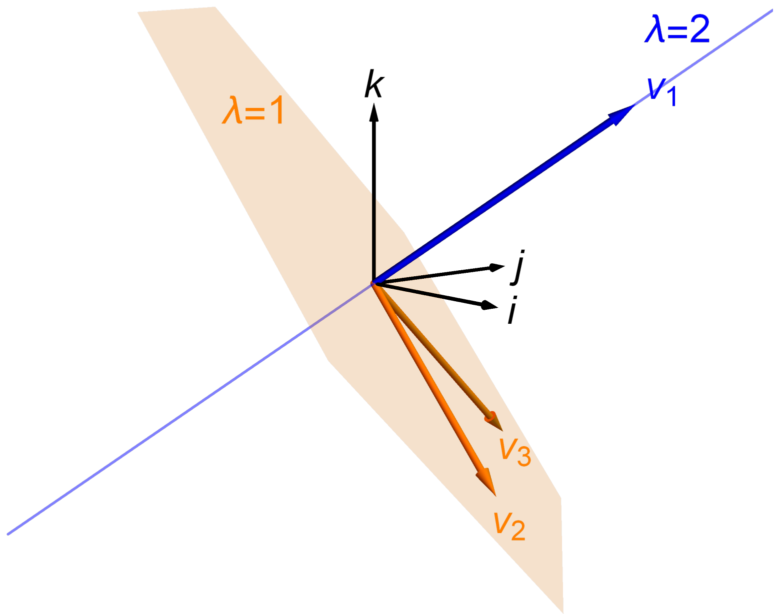
$$\mathbf{v}_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \quad \text{and} \quad \mathbf{v}_3 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

from the possibilities for $\lambda = 1$. These are clearly linearly independent. They are also orthogonal to (and therefore linearly independent of) $\mathbf{v}_1$; I'll say more on that in a bit. Hence, a matrix that diagonalises A is

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

You may wish to check that $\mathsf{S}^{-1}\mathsf{A}\mathsf{S} = \mathsf{D}$.

On the next page, I've drawn a diagram to show what's really going on in the preceding example.



The thin blue line represents the eigenspace of $\lambda = 4$. Any vector that points along this line (i.e. is a multiple of $\mathbf{v}_1$) is an eigenvector with eigenvalue 4. We saw why that was in Lecture 4: any multiple of an eigenvector is also an eigenvector with the same eigenvalue.

The two-dimensional eigenspace of $\lambda = 1$ is shown as the orange plane. It's the plane $x + y + z = 0$ (look back to the previous example to see where this equation comes from). Any vector in this plane is an eigenvector with eigenvalue 1. We chose the two non-parallel vectors $\mathbf{v}_2$ and $\mathbf{v}_3$, when answering the question in the previous example.

Notice that any linear combination of $\mathbf{v}_2$ and $\mathbf{v}_3$ will also lie in the orange plane, so it will also be an eigenvector with eigenvalue 1. I will leave the proof as an exercise.

This is the generalisation of taking any multiple of $\mathbf{v}_1$ to form another eigenvector with eigenvalue 4.

Notice that the blue line is perpendicular to the orange plane. This is guaranteed here, because the matrix A is symmetric. All eigenvectors with different eigenvalues must be orthogonal. Therefore, the whole eigenspaces have to be orthogonal.

This illustrates why all symmetric matrices are diagonalizable. Eigenvectors with different eigenvalues are necessarily linearly independent. Furthermore, we can always choose enough linearly independent eigenvectors within any given eigenspace, to provide the right number of eigenvectors overall to form the matrix S.

Orthogonal diagonalization

From the picture above, it's clear that we could choose two eigenvectors in the $\lambda = 1$ plane that are mutually orthogonal. They would, of course, automatically be orthogonal to the $\lambda = 2$ eigenvector. We could then normalise all three eigenvectors to give a set of three *orthonormal eigenvectors*.

There is an advantage to doing this. If the eigenvectors are chosen to be orthonormal, then the columns of S will be orthonormal. Therefore, S will be an orthogonal matrix, which means that

$$\mathsf{S}^{-1} = \mathsf{S}^{\mathsf{T}}.$$

Then, A can be diagonalized using

$$\boxed{\mathsf{D} = \mathsf{S}^{\mathsf{T}}\mathsf{A}\mathsf{S} \qquad (7.1)}$$

instead of having to deal with the matrix inverse in $\mathsf{S}^{-1}\mathsf{A}\mathsf{S}$.

In other words, a symmetric matrix is not only guaranteed to be diagonalisable; it can always be diagonalised by an *orthogonal transformation* of the form (7.1). We just need to make sure that the eigenvectors we use to construct S are orthonormal.

When all the eigenvalues are different, it's very easy to ensure that the eigenvectors are orthonormal, because they are already guaranteed to be mutually orthogonal. So we just need to normalise them. For example, in Lecture 4 we saw that

$$\mathsf{A} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

has the following eigenvectors and eigenvalues:

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \qquad \lambda_1 = -1$$

$$\mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \qquad \lambda_2 = 1$$

$$\mathbf{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \qquad \lambda_3 = 2.$$

Note the orthogonality. So we can just normalise them as follows

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$$

$$\mathbf{v}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

$$\mathbf{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

and then use them to form an orthogonal matrix S as

$$\mathsf{S} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

I'll leave it to you to confirm that (7.1) is satisfied using this S and its transpose.

When eigenvalues are repeated, things are trickier. In Example 1, the eigenvectors we chose for $\lambda = 1$ weren't orthogonal. So we need some way of choosing orthogonal eigenvectors for degenerate eigenvalues.

There are various tricks. One is called *Gram-Schmidt orthogonalisation*, which works for eigenvectors of any sized matrix. It is beautiful, but I don't have time to cover it. Given that you are most likely to encounter degenerate eigenvalues in 3×3 matrices, I'll show you the dirtiest trick of them all.

Example 2. Find an orthogonal matrix that diagonalises A in Example 1

The issue is ensuring orthogonality of the eigenvectors. We know that $\mathbf{v}_1$ and $\mathbf{v}_2$ must already be orthogonal, but how do we find the third, orthogonal eigenvector that we need?

Answer: use the cross product! Let's take $\mathbf{v}_1 \times \mathbf{v}_2$:

$$\mathbf{v}_1 \times \mathbf{v}_2 = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 1 & 1 \\ 1 & 0 & -1 \end{vmatrix}$$

You can show that this gives a vector

$$\begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}$$

It is necessarily an eigenvector with eigenvalue 1, because it must be perpendicular to $\mathbf{v}_1$ by the definition of the cross product. (Look at the diagram a couple of pages back.) And now we've ensured that it's also perpendicular to $\mathbf{v}_2$.

Hence, by normalising all three vectors, we obtain

$$\mathsf{S} = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{pmatrix}.$$

Not a pretty answer, but it works.

Geometrical picture of orthogonal diagonalisation

I'll end the lecture with the 'classic' example of orthogonal diagonalisation. It's motivated by the following question: sketch the graph of

$$5x^2 + 8xy + 5y^2 = 1.$$

To answer the question, we note first that

$$5x^2 + 8xy + 5y^2 = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{v}^{\mathsf{T}}\mathsf{A}\mathbf{v}.$$

If this is not clear, multiply it out yourself to check. Notice that there are infinitely-many choices of A that would have worked; we've chosen the one case where it is symmetric.

Our problem then becomes: sketch the graph of

$$\mathbf{v}^{\mathsf{T}}\mathsf{A}\mathbf{v} = 1. \tag{7.2}$$

The eigenvalues of A are 1 and 9. The eigenvectors can be found as usual; upon normalising them and putting them into the S matrix, we obtain

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

Notice that this has turned out to be the matrix for a clockwise rotation by $\pi/4$.

Now, let

$$\mathbf{v} = \mathsf{S}\mathbf{u}. \tag{7.3}$$

Then,

$$\begin{aligned} \mathbf{v}^{\mathsf{T}}\mathsf{A}\mathbf{v} &= (\mathsf{S}\mathbf{u})^{\mathsf{T}}\mathsf{A}(\mathsf{S}\mathbf{u}) \\ &= \mathbf{u}^{\mathsf{T}}\mathsf{S}^{\mathsf{T}}\mathsf{A}\mathsf{S}\mathbf{u} \\ &= \mathbf{u}^{\mathsf{T}}\mathsf{D}\mathbf{u}. \end{aligned}$$

So, from (7.2), our problem becomes: sketch the graph of

$$\mathbf{u}^{\mathsf{T}}\mathsf{D}\mathbf{u} = 1. \tag{7.4}$$

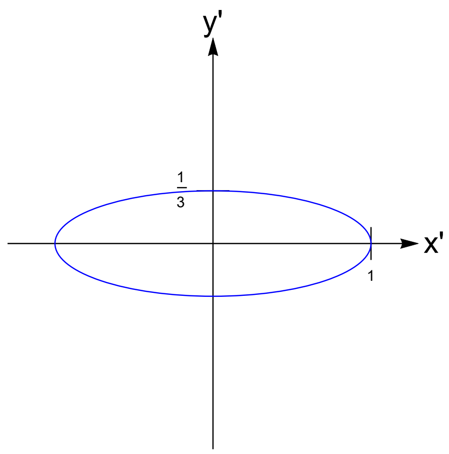
Let's denote the components of $\mathbf{u}$ by x' and y'. Then,

$$\begin{aligned} \mathbf{u}^{\mathsf{T}}\mathsf{D}\mathbf{u} &= \begin{pmatrix} x' & y' \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} \\ &= (x')^2 + 9(y')^2. \end{aligned}$$

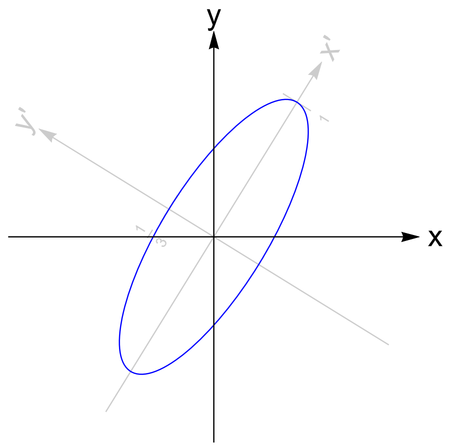
Hence, (7.4) becomes

$$(x')^2 + 9(y')^2$$

which is clearly an ellipse.



But hang on... we were asked to sketch the graph in the xy plane, not our $x'y'$ plane. Switching between the two is very easy, though. From (7.3), we see that the x and y axes are just the x' and y' axes rotated clockwise by $\pi/4$. Hence, in the xy plane, the curve is



This illustrates the beauty of orthogonal diagonalisation. As we've seen, orthogonal transformations just rotate (and possibly) reflect the axes, so orthogonal diagonalisation is all about looking at a problem in a rotated set of axes that makes it simpler.

The more general case of Hermitian matrices

We've focussed on symmetric matrices for simplicity. Analogous arguments can be made for the more general case of a Hermitian matrix. Instead of being diagonalised by an orthogonal matrix, they turn out to be diagonalised by a *unitary matrix* that satisfies

$$\mathsf{U}^\dagger = \mathsf{U}^{-1}.$$

'Diagonalizing the Hamiltonian' is a phrase you'll hear a lot in quantum mechanics next year.