LECTURE 1
LECTURE 2
0. Distinct eigenvalues

I haven't gotten around to stating the following important theorem:

Theorem: A matrix with n distinct eigenvalues is diagonalizable.

Proof (Sketch) Suppose $n = 2$, and let λ_1 and λ_2 be the eigenvalues, $\vec{v}_1, \vec{v}_2$ the eigenvectors. But for all we know, $\vec{v}_1$ and $\vec{v}_2$ are not linearly independent!

Suppose they're not; then we have

$$c_1\vec{v}_1 + c_2\vec{v}_2 = \vec{0}$$

with c_1 and c_2 not both 0.

Multiplying both sides by A, get

$$c_1\lambda_1\vec{v}_1 + c_2\lambda_2\vec{v}_2 = \vec{0}.$$

Multiplying first equation by λ_1 gives

$$c_1\lambda_1\vec{v}_1 + c_2\lambda_1\vec{v}_2 = \vec{0}$$

and subtracting gives

$$c_2(\lambda_2 - \lambda_1)\vec{v}_2 = \vec{0}$$

and from this–since $\lambda_2 \neq \lambda_1$!–we can conclude $c_2 = 0$. It follows similarly that $c_1 = 0$, contradiction.

A similar argument gives the result for any n, but it's not as easy as Strang makes it seem; it requires the fact that the *Vandermonde matrix* is invertible (see Strang, p.98).

Apropos of nothing, I also want to comment:

Fact. A is invertible if and only if 0 is not an eigenvalue of A.

1. Symmetric, Hermitian, unitary matrices

Spectral theorem: A (real) symmetric matrix is diagonalizable.

Strangely enough, the best way to prove this (and I think Strang's proof is very good) is to use *complex* matrices.

Definitions: Recall that the *complex conjugate* of a number $a + bi$ is $a - bi$. Similarly, the complex conjugate of a matrix A is the matrix obtained by replacing each entry with its complex conjugate.

- A^H ("A Hermitian") is the complex conjugate of A^T.

- A is *symmetric* if $A = A^T$.
- A is *Hermitian* if $A = A^H$.
- A is *unitary* if $AA^H = 1$.

Note that "unitary" is the complex analogue of "orthogonal." Indeed, a real unitary matrix is orthogonal.

Note also that $(AB)^H = B^H A^H$.

Give the example of heat diffusion on a circle to suggest the ubiquity of symmetric matrices.

Examples: A typical Hermitian matrix is

$$\begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}.$$

Compute, just for fun, that the eigenvalues are 0 and 2. That they're real numbers, despite the fact that the matrix is complex, is no coincidence! We might want to analyze this before we think about unitary matrices too much.

Definition: Let $\vec{v}$ be a vector with complex entries. We define

$$\|\vec{v}\| = \sqrt{\vec{v}^H \vec{v}}$$

(positive square root.)

Before we prove the spectral theorem, let's prove a theorem that's both stronger and weaker.

Theorem. Let A be an arbitrary matrix. There exists a unitary matrix U such that $U^{-1}AU$ is upper triangular.

We don't have to assume A is symmetric, as in the spectral theorem, but we get a weaker conclusion as a result.

We proceed as follows. We know A has at least one eigenvector $\vec{v}_1$. Let U_1 be a unitary matrix of the form

$$\begin{bmatrix} | & | & | \\ \vec{v}_1 & \dots & \vec{v}_n \\ | & | & | \end{bmatrix}$$

where the latter columns are unspecified. (How do we know there is such a U? The Gram-Schmidt process...) What does $U_1^{-1}AU_1$ look like? Well, what does it do to $\vec{e}_1$? It sends it to

$$U_1^{-1}AU_1\vec{e}_1 = U_1^{-1}A\vec{v}_1 = \lambda_1 U_1^{-1}\vec{v}_1 = \lambda_1\vec{e}_1.$$

So $U_1^{-1}AU_1$ looks like

$$\begin{bmatrix} \lambda_1 & * & * \\ 0 & * & * \\ 0 & 0 & * \end{bmatrix}$$

(Of course, it is only for convenience that we write a 3×3 matrix; the proof works for any n.)

And now we move on down to the "southeast corner" of the matrix, which is a 2×2 matrix A_2. Now A_2 has an eigenvector $\vec{v}_2$. So write U_2 a matrix with a 1 in the upper corner and $\vec{v}_2$ as the second column. This is a little hard for me to type.

What does $U_2^{-1}U_1^{-1}AU_1U_2$ look like?

Well, multiplying by U_2 and U_2^{-1} doesn't affect the first row or column at all; but it turns the second column into

$$\begin{bmatrix} * \\ \lambda_2 \\ 0 \end{bmatrix}$$

just as U_1 turned the first column into

$$\begin{bmatrix} \lambda_2 \\ 0 \\ 0 \end{bmatrix}.$$

Now we just continue this process until we get a triangular matrix similar to A.

I should remark here that to believe the proof (which is well written-up in Strang) you should really work out a 3×3 example on your own. Once you do that, you will really see that the argument is correct. Before you do that, I think it may not be so convincing.

Remark: Our favorite example of non-diagonalizable matrices,

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},$$

is certainly similar to a triangular matrix, because it is itself triangular!

Note that the fact that U was unitary played essentially no role in the above proof. Why did we bother? We bothered because now we're ready to prove the spectral theorem. In fact, we'll prove something even a little better.

Theorem: Every Hermitian matrix is diagonalizable. In particular, every real symmetric matrix is diagonalizable.

Proof. Let A be a Hermitian matrix. By the above theorem, A is "triangularizable"–that is, we can find a unitary matrix U such that

$$U^{-1}AU = T$$

with T upper triangular.

Lemma. $U^{-1}AU$ is Hermitian. **Proof of Lemma.**

$$(U^{-1}AU)^H = U^H A^H (U^{-1})^H = U^{-1}AU.$$

Now we're done! Because if T is upper triangular, then T^H is lower triangular.

Ask: can we write down a triangular Hermitian matrix? Is there such a thing? Take some time to talk about that.

So we have shown that in fact A is similar to a real diagonal matrix, which is to say that A is diagonalizable, and all the eigenvalues of A are real–the very phenomenon we observed for the Hermitian matrix above!

More than that: suppose A is real symmetric. Then the fact that the eigenvalues of A are real means that the eigenvectors of A are also real. So U is a real unitary matrix, so

$$UU^H = UU^T = I;$$

that is, U is an *orthogonal* matrix.

This proves the remarkable fact that *the eigenvectors of a symmetric matrix are mutually orthogonal.*

Theorem. Let A be a Hermitian matrix. Then the eigenvalues of A are real and the eigenvectors are orthogonal (in the Hermitian sense: $v_i^H v_j = 0$.

LECTURE 3

Example: I want to talk about the diffusion of heat around a circular loop of wire. Suppose you have such a loop, and you start by heating one point. How quickly does the heat diffuse over the whole loop? Draw the circle, and choose n nodes where we're going to measure temperature. Then the temperature is governed by the difference equation

$$\vec{x}_{t+1} = A\vec{x}_t$$

where A is a matrix with $a_{ij} = 1/2$ whenever $|i-j| = 1$ or when $|i-j| = n-1$. (This matrix is a pain to type.) For $n = 3$, for instance, the matrix is

$$A_3 = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

Question. Suppose there are 101 nodes. Start with a temperature of 1 at the top node, 0 elsewhere. How long will it take before the temperature is approximately even; say that the temperature is within 0.01 of 1/101 everywhere?

Well, one writes the initial state as

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_{101}\vec{v}_{101}$$

One sees that $\lambda_1 = 1$ is an eigenvalue with eigenvector

$$\vec{v}_1 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}.$$

We know that the eigenvectors are all real numbers (which corresponds with our intiution that there shouldn't be any oscillation in the temperature– it should just progress steadily towards the equilibrium.) Physical intuition also tells us that all eigenvalues are between 1 and -1; otherwise the temperatures would spiral out of control! (It would be nice to prove this fact mathematically, though.)

So write the eigenvalues in order of absolute value:

$$1 = |\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_{101}|.$$

And write our initial state

$$\vec{x}_0 = c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_{101}\vec{v}_{101}.$$

Observation 0: $c_1 = (1/101)$.

Observation 1: $\|c_i\vec{v}_i\| < 1$ for all i. Because $c_i\vec{v}_i$ is nothing other than the orthogonal projection of $\vec{x}_0$ onto the line.

So after a long time,

$$\vec{x}_k = A^k\vec{x}_0 = (1/101)\vec{v}_1 + c_2\lambda_2^k\vec{v}_2 + \ldots + c_{101}\lambda_{101}^k\vec{v}_{101}.$$

So the discrepancy between $\vec{x}_k$ and the uniform temperature distribution is

$$\vec{d} = c_2\lambda_2^k\vec{v}_2 + \ldots + c_{101}\lambda_{101}^k\vec{v}_{101}.$$

How big is this discrepancy? Well, note that

$$\|\vec{d}\| \leq \lambda_2^k + \lambda_3^k + \ldots + \lambda_{101}^k \leq 100\lambda_2^k$$

so in particular, if

$$100\lambda_2^k < 0.01$$

it follows that the length of $\vec{d}$, whence certainly the difference at any point between the temperature and $1/101$, is less than 0.01.

So it all comes down to estimating λ_2. How the heck are we supposed to do that? This kind of problem is *hugely* important. In this simple case, the answer: we know

$$\lambda_2 = \cos \pi/101 = 0.9995.$$

(This can be proved by mathematical induction using the theory of Tchebysheff polynomials.)

And we find that, a priori, 18860 units of time suffice!

To be honest, I'm not a good enough computer programmer to check how good an approximation this is.

Compare with the seven-shuffle theorem.

2. Similar matrices represent the same linear transformation with respect to different bases.

I just want to say a word about this important idea. Let $\mathcal{W}$ be a basis for $\mathbb{R}^n$.

Example: $\vec{w}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \vec{w}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. We can write

$$T\vec{w}_1 = c_{11}\vec{w}_1 + c_{21}\vec{w}_2$$

and

$$T\vec{w}_2 = c_{12}\vec{w}_1 + c_{22}\vec{w}_2$$

In that case, write

$$[T]_{\mathcal{W}} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}.$$

Remark. If $\mathcal{W}$ is the standard basis, then $[T]_{\mathcal{W}}$ is the usual matrix representation of T.

Remark. If $\mathcal{W}$ consists of eigenvectors for T, then $[T]_{\mathcal{W}}$ is diagonal. As in the case above, where T is the transformation usually written as $\begin{bmatrix} 9/8 & 7/8 \\ 7/8 & 9/8 \end{bmatrix}$.

Remark. Let $\mathcal{V}$ and $\mathcal{W}$ be two bases. We can write

$$\vec{w}_1 = s_{11}\vec{v}_1 + s_{21}\vec{v}_2$$

and

$$\vec{w}_2 = s_{12}\vec{v}_1 + s_{22}\vec{v}_2$$

And the corresponding matrix

$$S = \left[\begin{array}{cc} s_{11} & s_{12} \\ s_{21} & s_{22} \end{array}\right]$$

is called (by me) the *connecting matrix* between $\mathcal{V}$ and $\mathcal{W}$. For instance, if $\mathcal{V}$ is the standard basis, then S is the usual matrix

$$\left[\begin{array}{ccc} | & | & | \\ \vec{w}_1 & \dots & \vec{w}_n \\ | & | & | \end{array}\right].$$

Fact. $[T]_{\mathcal{W}} = S^{-1}[T]_{\mathcal{V}}S$.

In particular, the two matrices are similar.

This is a little messy to prove, so I shan't; it's easiest if we take $\mathcal{V}$ to be the standard basis.

Remark: Note that the eigenvalues of a linear transformation do not depend on the basis; that fits well with our knowledge that similar matrices have the same eigenvalues.

3. Jordan canonical form

OK, that was fun. Now I want to talk a *little* about Jordan canonical form.

Remember our GOAL: given A, find a "nice" matrix which is similar to A.

We've already seen that we can find a matrix similar to A which is upper triangular. We can do a little better; we can find a matrix J similar to A which is of Jordan block form; I'll write this on the board, say it makes it not too bad to compute the exponential.