

Math 1600B Lecture 31, Section 2, 24 March 2014

Announcements:

Course evaluations at start.

Today we finish 4.4 and cover the Markov chains part of Section 4.6. Not covering Section 4.5, or rest of 4.6 (which contains many interesting applications!) **Read Section 5.1** for next class.

Tutorials: Quiz 9 covers Section 4.4 only.

Office hour: Mon 1:30-2:30 and Wed 10:30-11:15, MC103B.

Help Centers: Monday-Friday 2:30-6:30 in MC 106.

Questions to discuss with your neighbour: What does it mean for A and B to be similar? What properties do similar matrices have in common? What does it mean for A to be diagonalizable? How do we tell if it is, and how do we diagonalize A ?

Review of Section 4.4:

Definition: Let A and B be $n \times n$ matrices. We say that A is **similar** to B ($A \sim B$) if there is an invertible matrix P such that $P^{-1}AP = B$.

Theorem 4.22: Let A and B be similar matrices. Then A and B have the same determinant, rank, characteristic polynomial and eigenvalues.

Definition: A is **diagonalizable** if it is similar to some diagonal matrix.

If A is similar to a diagonal matrix D , then D must have the eigenvalues of A on the diagonal. But how to find P ?

Theorem 4.23: Let A be an $n \times n$ matrix. If P is an $n \times n$ matrix whose columns are linearly independent eigenvectors of A , then $P^{-1}AP$ is a diagonal matrix D with the corresponding eigenvalues of A on the diagonal.

On the other hand, if P is any invertible matrix such that $P^{-1}AP$ is diagonal, then the columns of P are linearly independent eigenvectors of A . It follows that A is diagonalizable if and only if it has n linearly independent eigenvectors.

This theorem is one of the main reasons we want to be able to find eigenvectors of a matrix. Moreover, the more eigenvectors the better, so this motivates allowing complex eigenvectors.

Theorem 4.24: If $\lambda_1, \dots, \lambda_k$ are distinct eigenvalues of A and, for each i , \mathcal{B}_i is a basis for the eigenspace E_{λ_i} , then the union of the \mathcal{B}_i 's is a linearly independent set.

Combining Theorems 4.23 and 4.24 gives the following important consequence:

Theorem: An $n \times n$ matrix is diagonalizable if and only if the sum of the geometric multiplicities of the eigenvalues is n .

In particular:

Theorem 4.25: If A is an $n \times n$ matrix with n distinct eigenvalues, then A is diagonalizable.

So it is important to understand the geometric multiplicities better. Here is a helpful result:

Lemma 4.26: If λ_1 is an eigenvalue of an $n \times n$ matrix A , then
geometric multiplicity of $\lambda_1 \leq$ algebraic multiplicity of λ_1

It follows that the only way for the geometric multiplicities to add to n is if they are equal to the algebraic multiplicities and the algebraic multiplicities add to n :

Theorem 4.27 (The Diagonalization Theorem): Let A be an $n \times n$ matrix with distinct eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$. Let their geometric multiplicities be g_1, g_2, \dots, g_k and their algebraic multiplicities be a_1, a_2, \dots, a_k . Then the following are equivalent:

- a. A is diagonalizable.
- b. $g_1 + \cdots + g_k = n$.
- c. $g_i = a_i$ for each i and $a_1 + \cdots + a_k = n$.

Note: This is stated incorrectly in the text. The red part must be added unless you are working over \mathbb{C} , in which case it is automatic that $a_1 + \cdots + a_k = n$. With the way I have stated it, it is correct over \mathbb{R} or over \mathbb{C} .

Summary of diagonalization: Given an $n \times n$ matrix A , we would like to determine whether A is diagonalizable, and if it is, find the invertible matrix P and the diagonal matrix D such that $P^{-1}AP = D$. The result may depend upon whether you are working over \mathbb{R} or \mathbb{C} .

Steps:

1. Compute the characteristic polynomial $\det(A - \lambda I)$ of A .
2. Find the roots of the characteristic polynomial and their algebraic multiplicities by factoring.
3. If the algebraic multiplicities don't add up to n , then A is not diagonalizable, and you can stop. (If you are working over \mathbb{C} , this can't happen.)
4. For each eigenvalue λ , compute the dimension of the eigenspace E_λ . This is the geometric multiplicity of λ , and if it is less than the algebraic multiplicity, then A is not diagonalizable, and you can stop.
5. Compute a basis for the eigenspace E_λ .
6. If for each eigenvalue the geometric multiplicity equals the algebraic multiplicity, then you take the n eigenvectors you found and put them in the columns of a matrix P . Put the eigenvalues in the same order on the diagonal of a matrix D .
7. **Check** that $AP = PD$.

Note that step 4 only requires you to find the row echelon form of $A - \lambda I$, as the number of free variables here is the geometric multiplicity. In step 5, you solve the system.

New material:

Powers:

Suppose $P^{-1}AP = D$, where D is diagonal. Then $A = PDP^{-1}$. We can use this to compute powers of A . For example,

$$\begin{aligned} A^5 &= (PDP^{-1})(PDP^{-1})(PDP^{-1})(PDP^{-1})(PDP^{-1}) \\ &= PD^5P^{-1} \end{aligned}$$

and D^5 is easy to compute since D is diagonal: you just raise the diagonal entries to the fifth power.

More generally, $A^k = PD^kP^{-1}$. This is clearly an efficient way to compute powers! Note that we need to know P , not just D , to do this. Also note that even if A is real, it would work to diagonalize A over \mathbb{C} . The answer would be real, but the intermediate calculations would be complex.

See Example 4.29 for a sample calculation. We'll illustrate this result with an example from Markov Chains.

Review of Markov chains:

A **Markov chain** has a finite set of states $1, 2, \dots, n$ and there is an $n \times n$ matrix P (called the **transition matrix**) with the property that the ij entry P_{ij} is the probability that you transition from state j to state i in one time step.

Since you must transition to some state, $P_{1j} + \dots + P_{nj} = 1$. That is, the entries in each column sum to 1. Moreover, each entry $P_{ij} \geq 0$. Such a P is called a **stochastic matrix**.

We can represent the current state of the system with a **state vector** $\vec{x} \in \mathbb{R}^n$. The i th entry of \vec{x} may denote the number of people/objects in state i . Or we may divide by the total number, so the i th entry of \vec{x} gives the fraction of people/objects in state i . In this case, \vec{x} has non-negative entries that sum to 1 and is called a **probability vector**.

If \vec{x}_k denotes the state after k time steps, then the state after one more

time step is given by

$$\vec{x}_{k+1} = P\vec{x}_k.$$

It follows that $\vec{x}_k = P^k\vec{x}_0$. Therefore:

The ij entry $(P^k)_{ij}$ of P^k is the probability of going from state j to state i in k steps.

A state \vec{x} such that $P\vec{x} = \vec{x}$ is called a **steady state vector**. This is the same as an eigenvector with eigenvalue 1. In Lecture 22, we proved:

Theorem 4.30: Every stochastic matrix has a steady state vector, i.e. it has $\lambda = 1$ as an eigenvalue.

We proved this using the fact that P and P^T have the same eigenvalues, and then noticing that the vector with all 1's is an eigenvector of P^T with eigenvalue 1.

Example: We studied toothpaste usage, and had transition matrix

$$P = \begin{bmatrix} 0.70 & 0.20 \\ 0.30 & 0.80 \end{bmatrix}.$$

We noticed experimentally that a given starting state tends to the state $\begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix}$ and that

$$\begin{bmatrix} 0.70 & 0.20 \\ 0.30 & 0.80 \end{bmatrix} \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix}.$$

We then found this steady state vector algebraically by solving $(I - P)\vec{x} = \vec{0}$. [It is equivalent to solve $(P - I)\vec{x} = \vec{0}$.]

With our new tools, we can go further now.

Section 4.6: Markov chains:

Let's compute powers of the matrix P above. One can show that P has characteristic polynomial

$$\det(P - \lambda I) = \lambda^2 - 1.5\lambda + 0.5 = (\lambda - 1)(\lambda - 0.5)$$

and so has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 0.5$. The eigenspaces are

$$E_1 = \text{span}\left(\begin{bmatrix} 2 \\ 3 \end{bmatrix}\right) \quad \text{and} \quad E_{0.5} = \text{span}\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right)$$

So if we write $Q = \begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix}$, we have that $Q^{-1}PQ = \begin{bmatrix} 1 & 0 \\ 0 & 0.5 \end{bmatrix} = D$.

Therefore,

$$P^k = QD^kQ^{-1} = \begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} 1^k & 0 \\ 0 & 0.5^k \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix}^{-1}$$

As $k \rightarrow \infty$, $0.5^k \rightarrow 0$, so

$$P^k \rightarrow \begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix}^{-1} = \begin{bmatrix} 0.4 & 0.4 \\ 0.6 & 0.6 \end{bmatrix}$$

It follows that if we start with any state $\vec{x}_0 = \begin{bmatrix} a \\ b \end{bmatrix}$ with $a + b = 1$, we'll find that

$$\vec{x}_k = P^k \vec{x}_0 \rightarrow \begin{bmatrix} 0.4 & 0.4 \\ 0.6 & 0.6 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0.4a + 0.4b \\ 0.6a + 0.6b \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix}$$

This explains why every state tends to the steady state! (It also gives a fast way to compute \vec{x}_k for large k .)

This is a very general phenomenon, which we'll spend the rest of the lecture understanding.

Theorem 4.31: Let P be an $n \times n$ stochastic matrix. Then every eigenvalue λ has $|\lambda| \leq 1$.

If in addition the entries of P are all positive, then all eigenvalues besides $\lambda = 1$ have $|\lambda| < 1$.

The general proof just involves some inequalities, but the notation is confusing. Let's see how the argument goes in the special case of

$$P = \begin{bmatrix} 0.70 & 0.20 \\ 0.30 & 0.80 \end{bmatrix}.$$

The key idea is to study the eigenvalues of P^T , which are the same as those of P . Suppose $\begin{bmatrix} a \\ b \end{bmatrix}$ is an eigenvector of P^T with $0 \leq a \leq b$. Then

$$P^T \begin{bmatrix} a \\ b \end{bmatrix} = \lambda \begin{bmatrix} a \\ b \end{bmatrix} \text{ which means that}$$

$$\begin{bmatrix} 0.7a + 0.3b \\ 0.2a + 0.8b \end{bmatrix} = \lambda \begin{bmatrix} a \\ b \end{bmatrix}$$

The second component gives

$$\lambda b = 0.2a + 0.8b \leq 0.2b + 0.8b = b$$

and so $\lambda \leq 1$. If we allow a and b to be negative or complex, we need to use absolute values, and we can conclude that $|\lambda| \leq 1$.

The other part of the Theorem is similar.

The next theorem helps us understand the long-term behaviour:

Theorem 4.33: Let P be an $n \times n$ stochastic matrix all of whose entries are positive. Then as $k \rightarrow \infty$, $P^k \rightarrow L$, a matrix all of whose columns are equal to the same vector \vec{x} which is a steady state probability vector for P .

Proof: We'll assume P is diagonalizable: $Q^{-1}PQ = D$. So $P^k = QD^kQ^{-1}$. As $k \rightarrow \infty$, D^k approaches a matrix D^* with 1's and 0's on the diagonal (by Theorem 4.31), which means that P^k approaches $L = QD^*Q^{-1}$.

Now that we know that P^k has *some* limit L , we can deduce something about it. Since $\lim_{k \rightarrow \infty} P^k = L$, we have

$$PL = P \lim_{k \rightarrow \infty} P^k = \lim_{k \rightarrow \infty} P^{k+1} = L$$

This means that the columns of L must be steady-state vectors for P . Since the columns of P^k are probability vectors, the same must be true of the columns of L . It's not hard to show that P has a unique steady-state probability vector \vec{x} , so $L = [\vec{x} \ \vec{x} \ \cdots \ \vec{x}]$, as required. \square

Finally, we can deduce that Markov chains tend to their steady states:

Theorem 4.34: Let P be an $n \times n$ stochastic matrix all of whose entries are positive, and let \vec{x}_0 be any initial probability vector. Then as $k \rightarrow \infty$, $\vec{x}_k \rightarrow \vec{x}$, where \vec{x} is the steady state probability vector for P .

Proof: Suppose that \vec{x}_0 has components x_1, x_2, \dots, x_n . Then

$$\begin{aligned} \lim_{k \rightarrow \infty} \vec{x}_k &= \lim_{k \rightarrow \infty} P^k \vec{x}_0 = L \vec{x}_0 \\ &= [\vec{x} \ \vec{x} \ \cdots \ \vec{x}] \vec{x}_0 \\ &= x_1 \vec{x} + x_2 \vec{x} + \cdots + x_n \vec{x} \\ &= (x_1 + x_2 + \cdots + x_n) \vec{x} = \vec{x} \quad \square \end{aligned}$$

This result works both ways: if you compute the eigenvector with eigenvalue 1, that tells you the steady-state vector that other states go to as $k \rightarrow \infty$. But it also means that if you don't know the steady-state vector, you can approximate it by starting with any vector \vec{x}_0 and computing $P^k \vec{x}_0$ for large k !

The latter is what Google does to compute the page rank eigenvector.

Note: A transition matrix P is **regular** if some power P^k of it has positive entries. This means that there is a nonzero probability to get from any starting state to any ending state after some number of steps. The text proves the above results for regular matrices, which is not hard once you know them for matrices with positive entries.