Math 1600A Lecture 31, Section 2, 22 Nov 2013

Announcements:

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 Monday. Work through recommended homework questions. Some updated!

Tutorials: Quiz 6 is next week. It covers Sections 4.4 and the Markov Chains part of 4.6.
Office hour: Monday, 1:30-2:30, MC103B.
Help Centers: Monday-Friday 2:30-6:30 in MC 106.

Review of Section 4.4

Definition: Let A and B be $n \times n$ matrices. We say that A is **similar** to B if there is an invertible matrix P such that $P^{-1}AP = B$. When this is the case, we write $A \sim B$.

Theorem 4.22: Let A and B be similar matrices. Then:
a. $\det A = \det B$
b. A is invertible iff B is invertible.
c. A and B have the same rank.
d. A and B have the same characteristic polynomial.
e. A and B have the same 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. Then A is diagonalizable if and only if A has n linearly independent eigenvectors.

More precisely, there exist an invertible matrix P and a diagonal matrix D with $P^{-1}AP = D$ if and only if the columns of P are n linearly independent eigenvectors of A and the diagonal entries of D are the corresponding eigenvalues in the same order.

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 $\lambda_1, \ldots, \lambda_k$ 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 in 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

$$\text{geometric multiplicity of } \lambda_1 \leqslant \text{algebraic multiplicity of } \lambda_1$$

We'll prove this in a minute. First, let's look at what it implies:

Let A be an $n \times n$ matrix with distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. Let their geometric multiplicities be $g_1, g_2, \ldots, g_k$ and their algebraic multiplicities be $a_1, a_2, \ldots, a_k$. We know

$$g_i \leqslant a_i \quad \text{for each } i$$

and so

$$g_1 + \cdots + g_k \leqslant a_1 + \cdots + a_k \leqslant n$$

So the only way to have $g_1 + \cdots + g_k = n$ is to have $g_i = a_i$ for each i and $a_1 + \cdots + a_k = n$.

This gives the **main theorem** of the section:

Theorem 4.27 (The Diagonalization Theorem): Let A be an $n \times n$ matrix with distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. Let their geometric multiplicities be $g_1, g_2, \ldots, g_k$ and their algebraic multiplicities be $a_1, a_2, \ldots, a_k$. 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 $\mathbb{C}$.

New material

We still need to prove:

Lemma 4.26: If λ_1 is an eigenvalue of an $n \times n$ matrix A, then

$$\text{geometric multiplicity of } \lambda_1 \leqslant \text{algebraic multiplicity of } \lambda_1$$

Proof (more direct than in text): Suppose that λ_1 is an eigenvalue of A with geometric multiplicity g, and let $\vec{v}_1, \ldots, \vec{v}_g$ be a basis for E_{λ_1}, so

$$A\,\vec{v}_i = \lambda_1\,\vec{v}_i \quad \text{for each } i$$

Let Q be an invertible matrix whose first g columns are $\vec{v}_1, \ldots, \vec{v}_g$:

$$Q = [\ \vec{v}_1 \ \cdots \ \vec{v}_g \quad \text{other vectors} \quad]$$

Since $Q^{-1}Q = I$, we know that $Q^{-1}\,\vec{v}_i = \vec{e}_i$ for $1 \leqslant i \leqslant g$. Also, the first g columns of AQ are $\lambda_1\,\vec{v}_1, \ldots, \lambda_1\,\vec{v}_g$. So the first g columns of $Q^{-1}AQ$ are $\lambda_1\,\vec{e}_1, \ldots, \lambda_1\,\vec{e}_g$. Therefore the matrix $Q^{-1}AQ$ has λ_1 as an eigenvalue with algebraic multiplicity at least g. But $Q^{-1}AQ$ has the same characteristic polynomial as A, so λ_1 must also have algebraic multiplicity at least g for A. $\square$.

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.

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.

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 $1, 2, \ldots, n$ 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, $P_{1j} + \cdots + P_{nj} = 1$. That is, the entries in each column sum to 1. Moreover, each entry $P_{ij} \geqslant 0$. 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 ith entry of $\vec{x}$ may denote the number of people/objects in state i. Or we may divide by the total number, so the ith 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

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

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 $k \to \infty$, $0.5^k \to 0$, so

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

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

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

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:

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

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 $0 \leqslant a \leqslant b$. Then $P^T \begin{bmatrix} a \\ b \end{bmatrix} = \lambda \begin{bmatrix} a \\ b \end{bmatrix}$ 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

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

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.

This 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 $k \to \infty$, $P^k \to L$, 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 $k \to \infty$, 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 $\lim_{k\to\infty} P^k = L$, we have

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

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 $k \to \infty$, $\vec{x}_k \to \vec{x}$, where $\vec{x}$ is the steady state probability vector for P.

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

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

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 $k \to \infty$. 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.