

# Math 1600 Lecture 32, Section 2, 21 Nov 2014

## Announcements:

Today we finish 4.6 and start Section 5.1. Continue reading **Section 5.1** for next class, and start reading **Section 5.2**. Work through recommended [homework questions](#).

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

## Review: Section 4.6: Markov chains:

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

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

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

**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  $P$ .

## New material:

**Proof of 4.33:** 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.

**Question:** Let  $P = \begin{bmatrix} 1/2 & 1/3 \\ 1/2 & 2/3 \end{bmatrix}$ . **(a)** Find a clever way to figure out the eigenvalues of  $P$  and determine whether it is diagonalizable.

**Solution to (a):** Since  $P$  is stochastic, we know that 1 is one of its eigenvalues. That means that the characteristic polynomial factors as  $(\lambda - 1)(\lambda - a)$  for some number  $a$ . Now  $\det(P) = 1/3 - 1/6 = 1/6$ . On the other hand,  $\det(P)$  is the value of the characteristic polynomial when  $\lambda = 0$ , so we must have  $a = 1/6$ . Therefore the eigenvalues of  $P$  are 1 and  $1/6$ . Since there are two distinct eigenvalues,  $P$  is diagonalizable.

**(b)** Compute  $P^{10}\vec{x}_0$ , where  $\vec{x}_0 = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$ .

One can show that eigenvectors of  $P$  are  $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$  and  $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$  for  $\lambda = 1$  and  $1/6$  respectively. Therefore

$$\begin{bmatrix} 2 & -1 \\ 3 & 1 \end{bmatrix}^{-1} P \begin{bmatrix} 2 & -1 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1/6 \end{bmatrix}$$

So

$$P = \begin{bmatrix} 2 & -1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1/6 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 3 & 1 \end{bmatrix}^{-1}$$

and therefore

$$\begin{aligned}
P^{10}\vec{x}_0 &= \begin{bmatrix} 2 & -1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1^{10} & 0 \\ 0 & (1/6)^{10} \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 3 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} \\
&= \begin{bmatrix} 2 & -1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1/6^{10} \end{bmatrix} \frac{1}{5} \begin{bmatrix} 1 & 1 \\ -3 & 2 \end{bmatrix} \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} \\
&= \frac{1}{5} \begin{bmatrix} 2 + 3/6^{10} & 2 - 2/6^{10} \\ 3 - 3/6^{10} & 3 + 2/6^{10} \end{bmatrix} \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} \\
&= \frac{1}{5} \begin{bmatrix} 2 + \frac{1}{2(6^{10})} \\ 3 - \frac{1}{2(6^{10})} \end{bmatrix}
\end{aligned}$$

## Section 5.1: Orthogonality in $\mathbb{R}^n$ :

Chapter 5 is all about orthogonality. We'll learn about (1) orthogonal projections onto planes and other subspaces; (2) how easy it is to work with an orthogonal basis; (3) how to orthogonally diagonalize matrices; etc.

Recall that vectors  $\vec{u}$  and  $\vec{v}$  are **orthogonal** if  $\vec{u} \cdot \vec{v} = 0$ .

**Definition:** A set  $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$  of vectors in  $\mathbb{R}^n$  is an **orthogonal set** if all pairs of distinct vectors in the set are orthogonal, i.e.  $\vec{v}_i \cdot \vec{v}_j = 0$  whenever  $i \neq j$ .

**Example 5.1:** The following vectors form an orthogonal set:

$$\vec{v}_1 = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{v}_3 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

**Theorem 5.1:** If  $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$  is an orthogonal set of nonzero vectors in  $\mathbb{R}^n$ , then these vectors are linearly independent.

**Proof:**

Suppose that

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

If we dot both sides with  $\vec{v}_1$ , we get that

$$(c_1 \vec{v}_1 + c_2 \vec{v}_2 + \cdots + c_k \vec{v}_k) \cdot \vec{v}_1 = \vec{0} \cdot \vec{v}_1 = 0$$

But the left-hand-side is equal to

$$c_1(\vec{v}_1 \cdot \vec{v}_1) + c_2(\vec{v}_2 \cdot \vec{v}_1) + \cdots + c_k(\vec{v}_k \cdot \vec{v}_1)$$

and all terms are zero except the first. This shows that

$$c_1(\vec{v}_1 \cdot \vec{v}_1) = 0$$

Since  $\vec{v}_1 \neq \vec{0}$ , we must have  $c_1 = 0$ . By dotting with other vectors, we see that all  $c_i$ s are zero.  $\square$

**Definition:** An **orthogonal basis** for a subspace  $W$  of  $\mathbb{R}^n$  is a basis that is also orthogonal.

For example, the vectors in Example 5.1 automatically form a basis for  $\mathbb{R}^3$ . Checking orthogonality is much easier than checking linear independence!

**Fact:** We'll show in Section 5.3 that every subspace *has* an orthogonal basis.

It is very easy to figure out coordinates with respect to an orthogonal basis. Suppose  $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$  is an orthogonal basis and

$$\vec{w} = c_1 \vec{v}_1 + \cdots + c_k \vec{v}_k$$

Then if we dot both sides with  $\vec{v}_1$ , we find

$$\vec{w} \cdot \vec{v}_1 = c_1(\vec{v}_1 \cdot \vec{v}_1)$$

and so

$$c_1 = \frac{\vec{w} \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}$$

**Theorem 5.2:** Let  $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$  be an orthogonal basis for a subspace  $W$  of  $\mathbb{R}^n$ , and let  $\vec{w}$  be any vector in  $W$ . Then the unique scalars  $c_1, c_2, \dots, c_k$  such that

$$\vec{w} = c_1 \vec{v}_1 + \dots + c_k \vec{v}_k$$

are given by the formula

$$c_i = \frac{\vec{w} \cdot \vec{v}_i}{\vec{v}_i \cdot \vec{v}_i}$$

**Example:** Find the coordinates of  $\vec{w} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$  with respect to the basis

$$\vec{v}_1 = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{v}_3 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

**Solution:** The coordinates are

$$c_1 = \frac{\vec{w} \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} = \frac{1}{6}, \quad c_2 = \frac{\vec{w} \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2} = \frac{5}{2}, \quad c_3 = \frac{\vec{w} \cdot \vec{v}_3}{\vec{v}_3 \cdot \vec{v}_3} = \frac{2}{3}$$

So

$$\vec{w} = \frac{1}{6} \vec{v}_1 + \frac{5}{2} \vec{v}_2 + \frac{2}{3} \vec{v}_3$$

Normally you have to solve a system to figure out the coordinates!

**Definition:** An **orthonormal set** is an orthogonal set of *unit* vectors. An **orthonormal basis** for a subspace  $W$  is a basis for  $W$  that is an orthonormal set.

The condition of being orthonormal can be expressed as

$$\vec{v}_i \cdot \vec{v}_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

**Example:** The standard basis  $\{\vec{e}_1, \dots, \vec{e}_n\}$  is an orthonormal basis for  $\mathbb{R}^n$ .

For an orthonormal basis, the formula for the coordinates simplifies:

**Theorem 5.3:** If  $\{\vec{q}_1, \vec{q}_2, \dots, \vec{q}_k\}$  is an orthonormal basis of a subspace  $W$ , and  $\vec{w}$  is in  $W$ , then

$$\vec{w} = (\vec{w} \cdot \vec{q}_1)\vec{q}_1 + \dots + (\vec{w} \cdot \vec{q}_k)\vec{q}_k$$

Note that any orthogonal basis can be converted to an orthonormal basis by dividing each vector by its length. E.g.

$$\vec{q}_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}, \quad \vec{q}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{q}_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

is an orthonormal basis for  $\mathbb{R}^3$ .

**Question:** How many orthonormal bases are there for  $\mathbb{R}^3$ ?

## Orthogonal Matrices

**Definition:** A square matrix  $Q$  with real entries whose columns form an orthonormal set is called an **orthogonal** matrix. (Strange name!)

**Examples:**  $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$  and  $B = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$  and

$$C = [\vec{q}_1 \ \vec{q}_2 \ \vec{q}_3] = \begin{bmatrix} \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{bmatrix}.$$

**Non-example:**  $D = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$  is not orthogonal.

**Note:** In  $\mathbb{R}^2$  and  $\mathbb{R}^3$ , orthogonal matrices correspond exactly to rotations and reflections. This is an important geometric reason to study them. Another reason is that we will see in Section 5.4 that they are related to diagonalization of symmetric matrices.

**Theorems 5.4 and 5.5:**  $Q$  is orthogonal if and only if  $Q^T Q = I$ , i.e. if and only if  $Q$  is invertible and  $Q^{-1} = Q^T$ .

**Proof:**

The  $ij$  entry of  $Q^T Q$  is  $\text{row}_i(Q^T) \cdot \text{col}_j(Q)$ . That is, it equals  $\text{col}_i(Q) \cdot \text{col}_j(Q)$ . So the columns are orthonormal if and only if the  $ij$  entry of  $Q^T Q$  is 1 when  $i = j$  and 0 otherwise, which means that  $Q^T Q = I$ .

Since  $Q$  is square, the last equation is the same as saying the  $Q$  is invertible and  $Q^{-1} = Q^T$ .  $\square$

Note that  $Q^T$  is much easier to calculate than  $Q^{-1}$ !

**Theorem 5.7:** If  $Q$  is orthogonal, then its **rows** form an orthonormal set too.

**Proof:** Since  $Q^T Q = I$ , we must also have  $Q Q^T = I$ . But the last equation says exactly that the rows of  $Q$  are orthonormal.  $\square$

Another way to put it is that  $Q^T$  is also an orthogonal matrix. Look at the examples again.

**Theorem 5.6:** Let  $Q$  be an  $n \times n$  matrix. Then the following statements are equivalent:

- $Q$  is orthogonal.
- $\|Q\vec{x}\| = \|\vec{x}\|$  for every  $\vec{x}$  in  $\mathbb{R}^n$ .
- $Q\vec{x} \cdot Q\vec{y} = \vec{x} \cdot \vec{y}$  for every  $\vec{x}$  and  $\vec{y}$  in  $\mathbb{R}^n$ .

For example,  $T_B$  is rotation, which preserves lengths and angles.

**Partial proof of 5.6:**



(a)  $\implies$  (c): Suppose  $Q$  is orthogonal, so  $Q^T Q = I$ . Then

$$Q\vec{x} \cdot Q\vec{y} = (Q\vec{x})^T (Q\vec{y}) = \vec{x}^T Q^T Q \vec{y} = \vec{x}^T \vec{y} = \vec{x} \cdot \vec{y}$$

(c)  $\implies$  (a): Suppose (c) holds. Take  $\vec{x} = \vec{e}_i$  and  $\vec{y} = \vec{e}_j$ . Then  $Q\vec{x} = Q\vec{e}_i$  is the  $i$ th column of  $Q$  and  $Q\vec{y} = Q\vec{e}_j$  is the  $j$ th column of  $Q$ . (c) says that the dot product of these columns is

$$Q\vec{e}_i \cdot Q\vec{e}_j = \vec{e}_i \cdot \vec{e}_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

So the columns of  $Q$  are orthonormal, which means  $Q$  is orthogonal.

(c)  $\implies$  (b) is clear, by taking  $\vec{x} = \vec{y}$  in (c).

(b)  $\implies$  (c): see text.  $\square$