Linear Algebra Notes

Paul Ellis, Manhattanville College

September 19, 2017

Contents

Preface

These are my personal lecture notes for teaching linear algebra, and are based on W.W.L. Chen's wonderful notes, which are available free online here:

https://rutherglen.science.mq.edu.au/wchen/lnlafolder/lnla.html

I have taught this course (using this text) during Fall semester of 2011, 2012, 2014, and 2016, so I thought maybe I should write down some of my observations.

Each chapter begins with my notes about if and how I deviate significantly from Chen, as well as a list of the exercises I generally assign. The numbered sections are material that significantly differs from or is not contained in Chen. The appendix is wholly new, and is a retelling of a classic example of László Babai.

-Paul

Chapter 1

Systems of Linear Equations

I follow Chen's Chapter 1, mostly. I give examples and homework from Sections 1.8, 1.9, 1.10, but I don't put them on the exams. I skip the application in Section 1.11 entirely.

Exercises in Chapter 1: 1, 3, 4, 6b, 7, 9, 11, 13, 18, 19e

Challenge problem: Prove that each of the three Row Operations do not change the set of solutions to a system of linear equations.

Chapter 2

Introduction to Matrix Algebra

I follow the general outine of Chen's Chapter 2. However, I skip the applications in Sections 2.10, 2.11, 2.13 entirely.

Exercises in Chapter 2: 1, 3, 4, 5, 6d, 6e*, 7ab, 8ab, 9a (do it as stated, but also with the matrix $d = (20000, 30000)$), 15, 16, 17, 18ac

2.1 Proof of Associativity

I make a point to give an example of and prove the associative property of matrix multiplication. There's a few reasons for this. First, some students think it is obvious, and don't realize until doing an example or two that it is a thing that actually requires proof. Next, a number of proofs in this chapter rely on this property (and to a lesser extent, distributivity), and I like to emphasize how much we can build on this one property. These later proofs will sometimes look absolutely trivial, and so I like to pause in the middle of them and say somthing like "but see, we are using associativity here, and remember that we did take the time to prove that non-trivial result!" Finally, it is a good exercise in playing with the definition of matrix multiplication. And I cannot recall where at the moment, but I'm sure there are similar calulations in the middle of later proofs. Oh, and of course, there's the fact that the element in row i and column j in a product of adjacency matrices is the number of paths through a network from node i to node j, which is kind of visible in this expansion:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}b_{11}c_{11} + a_{11}b_{12}c_{21} + a_{12}b_{21}c_{11} + a_{12}b_{22}c_{21} & a_{11}b_{11}c_{12} + a_{11}b_{12}c_{22} + a_{12}b_{21}c_{12} + a_{12}b_{22}c_{22} \\ a_{21}b_{11}c_{11} + a_{21}b_{12}c_{21} + a_{22}b_{21}c_{11} + a_{22}b_{22}c_{21} & a_{21}b_{11}c_{12} + a_{21}b_{12}c_{22} + a_{22}b_{21}c_{12} + a_{22}b_{22}c_{22} \end{pmatrix}$$

2.2 Computational Miscellany

Here are some useful computations that follow immediately from the definition of matrix multiplication, but nevertheless are not obvious on first glance, and are used without comment at least once somewhere in this course.

*skippable

Theorem 2.2.1. *Suppose A is and $m \times n$ matrix, and that $v \in \mathbb{R}^n$ is a column vector. In particular, suppose that the rows of A are $r_1, r_2, \ldots, r_m$ Then*

$$Av = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{pmatrix} v = \begin{pmatrix} r_1 \cdot v \\ r_2 \cdot v \\ \vdots \\ r_m \cdot v \end{pmatrix}$$

Example 2.2.2. To visualize this, notice: $\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} ax + by + cz \\ dx + ey + fz \\ gx + hy + iz \end{pmatrix}$

Theorem 2.2.3. *Suppose A is an $m \times n$ matrix, and B is an $n \times p$ matrix. Supose the columns of B are $c_1, c_2, \ldots, c_p$. Then*

$$AB = A \begin{pmatrix} c_1 & c_2 & \ldots & c_p \end{pmatrix} = \begin{pmatrix} Ac_1 & Ac_2 & \ldots & Ac_p \end{pmatrix}$$

Example 2.2.4. To visualize this, notice:

$$\begin{aligned}
\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} j & k & l \\ m & n & o \\ p & q & r \end{pmatrix} &= \begin{pmatrix} aj + bm + cp & ak + bn + cq & al + bo + cr \\ dj + em + fp & dk + en + fq & dl + eo + fr \\ gj + hm + ip & gk + hn + iq & gl + ho + ir \end{pmatrix} \\
&= \left(\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} j \\ m \\ p \end{pmatrix} \quad \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} k \\ n \\ q \end{pmatrix} \quad \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} l \\ o \\ r \end{pmatrix} \right) \\
&= \left(A \begin{pmatrix} j \\ m \\ p \end{pmatrix} \quad A \begin{pmatrix} k \\ n \\ q \end{pmatrix} \quad A \begin{pmatrix} l \\ o \\ r \end{pmatrix} \right)
\end{aligned}$$

2.3 The inverse of a matrix

From now on, unless otherwise specified, every matrix in this course is presumed to be a square matrix.

We now investigate the (multiplicative) *inverse* of a matrix. Before we discuss this, think about what the multiplicative inverse of a real number is. For example, why do we say that 2 and $\frac{1}{2}$ are multiplicative inverses? It's precisely that $2 \cdot \frac{1}{2} = 1$. And what is special about the number 1? It is the multiplicative identity, which means that $1 \cdot x = x$ and $x \cdot 1 = x$ for any real number x. As it happens, there is a multiplicative identity for matrices:

Definition 2.3.1. Suppose $A = (a_{ij})$ is an $n \times n$ matrix. Then the entries a_{ii} together make up the *diagonal* of A.

For example, in the matrix $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$, the entries 1, 5, 9 make up the diagonal. Similarly, we would say that 2, 3, 6 are above the diagonal and that 4, 7, 8 are below the diagonal.

Definition 2.3.2. We define I_n to be the $n \times n$ matrix with 1 for each diagonal entry, and 0 for every other entry. Formally, we say $I_n = (a_{ij})$, where

$$(a_{ij}) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Some examples:

$$I_1 = \begin{pmatrix} 1 \end{pmatrix} \qquad I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad I_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Lemma 2.3.3. *If A is any $n \times n$ matrix, then $AI_n = A = I_nA$. Thus we call I_n the (multiplicative)* identitiy matrix*.*

Rather than proving this formaly, simply check it for a generic 3×3 matrix. It will then become obvious. Finally, we can define:

Definition 2.3.4. If A and B are both $n \times n$ matrices so that $AB = I_n$ and $BA = I_n$, then we say that A and B are *inverses* of each other. In this case, we write $B = A^{-1}$ and $A = B^{-1}$

In Section 2.6, we will prove that $AB = I_n \iff BA = I_n$, and so this definition is actually redundant. For now, it is the one we will use.

Does A^{-1} always exist? NO! For example, let $A = \begin{pmatrix} 0 & 0 \\ 2 & 3 \end{pmatrix}$. Then if $B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is any other 2×2 matrix, we see that

$$AB = \begin{pmatrix} 0 & 0 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 2a + 3c & 2b + 3d \end{pmatrix} \neq \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

This is quite different than the situation for real numbers. The only real number without an inverse is 0

Definition 2.3.5. If A^{-1} exists, we say that A is *invertible*.

Lemma 2.3.6. *If A^{-1} exists, then it is unique.*

Proof. Suppose B and C are both inverses of A. Then

$$B = BI = B(AC) = (BA)C = IC = C$$

□

Lemma 2.3.7. *If A^{-1} and B^{-1} both exist, then $(AB)^{-1}$ exists, and, in fact, $(AB)^{-1} = B^{-1}A^{-1}$*

Proof. $(AB)\left(B^{-1}A^{-1}\right) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$. Thus $B^{-1}A^{-1}$ is the inverse of AB □

Lemma 2.3.8. $\left(A^{-1}\right)^{-1} = A$

Proof. $\left(A^{-1}\right)^{-1}$ and A^{-1} are inverses of each other. Also A and A^{-1} are inverses of each other. Hence A and $\left(A^{-1}\right)^{-1}$ are both inverses of A^{-1}. So by Lemma 2.3.6, $\left(A^{-1}\right)^{-1} = A$ □

2.4 How do we find the inverse of a matrix?

Question 2.4.1. *If we have a matrix A, how do we find A^{-1}?*

As it happens, our method will use elementary row operations! As an illustration, let's start with the 2×2 case. Suppose $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $\begin{pmatrix} e & f \\ g & h \end{pmatrix}$ are inverses of each other, and let's see if we can find a way to solve for the second matrix in terms of the first one.* Each of the following statements is equivalent to the next one:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} ae+bg \\ ce+dg \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ AND } \begin{pmatrix} af+bh \\ cf+dh \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} e \\ g \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ AND } \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} f \\ h \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

*A really nice exercise (which I totally do assign) is to find and prove an explicit formula for the 2×2 case. However, the point here is to find a method that works for larger matrices, as well.

$$\begin{array}{cc} e & g \end{array} \qquad\qquad\qquad \begin{array}{cc} f & h \end{array}$$

$$\left(\begin{array}{cc|c} a & b & 1 \\ c & d & 0 \end{array}\right) \text{ AND } \left(\begin{array}{cc|c} a & b & 0 \\ c & d & 1 \end{array}\right)$$

Now suppose we solve these systems of linear equations by using elementary row operations to put them into reduced row echelon form. Then we will have:

$$\left(\begin{array}{cc|c} 1 & 0 & e \\ 0 & 1 & g \end{array}\right) \text{ AND } \left(\begin{array}{cc|c} 1 & 0 & f \\ 0 & 1 & h \end{array}\right)$$

But we will have used the same sequence of elementary row operations in each case. Thus we can do them simultaneously. Therefore

$$\left(\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right) \text{ row reduces to } \left(\begin{array}{cc|cc} 1 & 0 & e & f \\ 0 & 1 & g & h \end{array}\right)$$

And this gives us a method to find an inverse for the matrix A! To summarize:

- Start with $\left(\begin{array}{c|c} A & I \end{array}\right)$
- Use elementary row operations to obtain $\left(\begin{array}{c|c} I & B \end{array}\right)$
- Then B is actually A^{-1}

Example 2.4.2. Let's find the inverse of $\begin{pmatrix} 3 & 4 \\ 5 & 6 \end{pmatrix}$...

...answer is $\begin{pmatrix} -3 & 2 \\ \frac{5}{2} & -\frac{3}{2} \end{pmatrix}$

Example 2.4.3. Let's find the inverse of $\begin{pmatrix} 1 & 0 & 5 \\ 1 & 1 & 0 \\ 3 & 2 & 6 \end{pmatrix}$...

...answer is $\begin{pmatrix} 6 & 10 & -5 \\ -6 & -9 & 5 \\ -1 & -2 & 1 \end{pmatrix}$

Example 2.4.4. Let's find the inverse of $\begin{pmatrix} 1 & 3 & 0 & -1 \\ 0 & 1 & -2 & -1 \\ -2 & -6 & 3 & 2 \\ 3 & 5 & 8 & -3 \end{pmatrix}$....

...answer is $\begin{pmatrix} -\frac{9}{2} & -1 & -2 & \frac{1}{2} \\ \frac{25}{12} & 0 & \frac{2}{3} & -\frac{1}{4} \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ \frac{3}{4} & -1 & 0 & -\frac{1}{4} \end{pmatrix}$

Don't forget to check the answer in each example!

2.5 When is a Matrix Invertible? Consider Elementary Matrices.

We have now seen how to find the inverse of a matrix, but it also makes sense to ask, more generally,

Question 2.5.1. *Is a given matrix invertible?*

On the most basic level, it might be helpful to know the answer to this before we go about trying to calculate the inverse of a large matrix. However, the question of invertibility itself is will turn out to be quite interesting. Throughout this course, we will find a surprising number of ways of answering this question. And so invertibility will connect many seemingly unrelated concepts.

In any case, the key to our first answer of the invertibility question will be elementary matrices:

Definition 2.5.2. An $n \times n$ matrix is an *elementary matrix* if it it the result of performing *one* elementary row operation on the identity matrix I_n.

For example if we apply the operation $R_1 + 5R_2 \to R_1$ to I_3, we obtain the matrix $\begin{pmatrix} 1 & 5 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

and if we apply the operation $R_1 \leftrightarrow R_2$ to I_3, we obtain the matrix $\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

and if we apply the operation $87R_3 \to R_3$ to the matrix I_3, we obtain the matrix $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 87 \end{pmatrix}$

So what makes these matrices special? Well, let's see what happens when we multiply another matrix on the left by one of these:

$$\begin{pmatrix} 1 & 5 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11}+5a_{21} & a_{12}+5a_{22} & a_{13}+5a_{23} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 87 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 87a_{31} & 87a_{32} & 87a_{33} \end{pmatrix}$$

It's the same as applying the elementary row operation that defined the matrix in the first place!

Lemma 2.5.3. *Suppose the $n \times n$ elemtary matrix E is defined by applying the elentary row operation R to the martix I_n. Then if A is any other $n \times n$ matrix, then the matrix obtained by applying R to A is just EA.*

Proof. Exercise. Or just look at the above example. □

So elementary matrices give us a link between matrix multiplication and EROs. If we work a little, we can really exploit this link.

First, note that each elementary row operation has an inverse operation which 'undoes' it: The inverse of $(R_i + kR_j) \to R_i$ is $(R_i - kR_j) \to R_i$, the inverse of $R_i \leftrightarrow R_j$ is $R_i \leftrightarrow R_j$, and the inverse of $kR_i \to R_i$ is $\frac{1}{k}R_i \to R_i$.

Lemma 2.5.4. *Each elmentary matrix is invertible. In particular, if the elementary matrix E is defined by the elementary row operation R, then E^{-1} is defined by inverse of R.*

Proof. Exercise. Or just examine the inverses of the matrices in the above example. □

Next, let's think a bit about how two different row operations are composed together. Suppose the matrix B is formed by applying the elementary row operation R_1 to the matrix A, and then appling the operation R_2 to the result. Suppose E_1 and E_2 are the corresponding elementary matrices. Then $B = E_2(E_1 A)$. By the associativity of matrix multiplication, we will simply write $B = E_2 E_1 A$. But this also means that to get from B to A, we must first undo R_2, and then undo R_1, so $A = E_1^{-1}(E_2^{-1}B) = E_1^{-1}E_2^{-1}B$.

Putting all this together, we get the following definition:

Definition 2.5.5. Two matrices A, B are called *row equivalent* if B can be formed from A via a (finite) sequence of row operations. This is the same as saying: there is a (finite) sequence of elementary matrices $E_1, E_2, \ldots, E_n$ so that

$$B = E_n \ldots E_3 E_2 E_1 A$$

which is equivalent to $E_1^{-1} E_2^{-1} E_3^{-1} \ldots E_n^{-1} B = A$

So what does this all have to do with finding the inverse of a matrix? Well, we saw already that if the matrix A is invertible, then we start with $\left(A \mid I \right)$ And perform elementary row operations until we get $\left(I \mid A^{-1} \right)$. In other words, if A is row equivalent to I, then it has an inverse. It turns out that the converse (and much more) is true:

Theorem 2.5.6. *Suppose A is an $n \times n$ matrix. Then the folowing are equivalent.*

(a) A is invertible

(b) The matrix equation $Ax = \mathcal{O}$ has only the trivial solution $x = \mathcal{O}$

(c) A and I_n are row equivalent

(d) for any $n \times 1$ matrix b, the matrix equation $Ax = b$ has a solution

(e) for any $n \times 1$ matrix b, the matrix equation $Ax = b$ has the unique solution $x = A^{-1}b$

Before we prove this, you may think that (d) and (e) are redundant. Certainly (e) is a stronger statement than (d), and so if invertible matrices satisfy the stronger statement (e), then they also satisfy the weaker statement (d). However, each of these implications goes both ways. So if we are instead trying to prove that a given matrix is invertible, it will be useful that we only have to prove the weaker statement (d), and not the stonger statement (e).

Proof. We will prove $(a) \implies (b) \implies (c) \implies (a)$ and $(a) \implies (e) \implies (d) \implies (a)$. Note that this consists of 6 sub-proofs.
$\underline{(a) \implies (b)}$: Assume A^{-1} exists. Let x_0 be a solution to $Ax = \mathcal{O}$. Then

$$x_0 = Ix_0 = (A^{-1}A)x_0 = A^{-1}(Ax_0) = A^{-1}\mathcal{O} = \mathcal{O}$$

Thus $x_0 = \mathcal{O}$ is the only solution to $Ax = \mathcal{O}$.
$\underline{(b) \implies (c)}$: Assume $Ax = \mathcal{O}$ has only the trivial solution $x = \mathcal{O}$. Now think of this in terms of systems of linear equations! It means that when we apply row reductions and backsolving to this system of linear equations:

$$\left(\begin{array}{c|c} \text{A} & \begin{matrix} 0 \\ 0 \\ \vdots \\ 0 \end{matrix} \end{array}\right)$$

that we should get the unique solution $x_1 = 0, x_2 = 0, \ldots, x_n = 0$. But this means that the RREF is

$$\left(\begin{array}{cccc|c} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{array}\right)$$

And in terms of matrices, we've just said that A is row equivalent to I_n!

$\underline{(c) \implies (a)}$: See the paragraph preceeding this theorem.

$\underline{(a) \implies (e)}$: Assume A^{-1} exists, and let b be any $n \times 1$ matrix. Let $x_0 = A^{-1}b$. Then

$$Ax_0 = A(A^{-1}b) = (AA^{-1})b = I_n b = b$$

So why is this the only solution? Well, let y be another solution. That is, suppose $Ay = b$. Then

$$y = Iy = (A^{-1}A)y = A^{-1}(Ay) = A^{-1}b = x_0$$

$\underline{(e) \implies (d)}$: Obvious.

$\underline{(d) \implies (a)}$: Assume that for every $n \times 1$ matrix b, that $Ax = b$ has a solution. Then, for each $1 \le i \le n$, let b_i be the $n \times 1$ matrix whose entries are all 0, except that the ith entry is 1.*. In other words, b_i is the ith column of I_n. Next for each $1 \le i \le n$, let x_i be a solution to $Ax = b_i$. Then†

$$\begin{aligned} A\begin{pmatrix} x_1 & x_2 & \ldots & x_n \end{pmatrix} &= \begin{pmatrix} Ax_1 & Ax_2 & \ldots & Ax_n \end{pmatrix} \\ &= \begin{pmatrix} b_1 & b_2 & \ldots & b_n \end{pmatrix} \\ &= I_n \end{aligned}$$

Thus $\begin{pmatrix} x_1 & x_2 & \ldots & x_n \end{pmatrix}$ is the inverse of A. □

*So, for example, if $n = 3$, then $b_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $b_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, and $b_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

†See Theorem 2.2.3 above to exaplin the first equals sign

2.6 Sidedness of matrix inverses

Finally, we end this bit with a note about 'one-sided' inverses. In the definition of matrix inverse, we said that A and B are inverses of each other if both $AB = I$ *and* $BA = I$. But is this redundant? If one is true, does that make the other true? The reason we ask this kind of question is that if we want to prove a given matrix has an inverse, it will be easier to just prove one of the two statements, instead of having to prove both.* In fact, they *are* equivalent, but it will require some work to prove this.

Lemma 2.6.1. *Let A be an $n \times n$ matrix. Suppose it has a 'right inverse,' that is, that there is a matrix C so that $AC = I$. Then A also has a 'left inverse,' that is, there is a matrix B so that $BA = I$.*

Proof. Let $AC = I$. Assume toward a contradiction that A does **not** have a left inverse. Next let $E_1, E_2, \ldots, E_k$ be the sequence of elementary matrices which reduces A to reduced row echelon form, so that $E_k \ldots E_2 E_1 A$ is in row reduced echelon form. The only way for a square matrix to be in RREF is if it is the identity matrix, or if it has an all-zero row. Now $E_k \ldots E_2 E_1 A$ is not the identity martix, else $E_k \ldots E_2 E_1$ would be a left inverse for A. So we conclude that $E_k \ldots E_2 E_1 A$ has an all-zero row.

Next, note that if the ith row of a matrix P is all-zeros, and if Q is any other matrix, then the ith row of PQ is also all-zeros. Therefore $(E_k \ldots E_2 E_1 A)C$ has an all-zero row. But we also have

$$(E_k \ldots E_2 E_1 A)C = E_k \ldots E_2 E_1 (AC) = E_k \ldots E_2 E_1 I = E_k \ldots E_2 E_1$$

so $E_k \ldots E_2 E_1$ has an all-zero row. But recall that elementary matrices all have inverses, and so then

$$(E_k \ldots E_2 E_1)(E_1^{-1} E_2^{-1} \ldots E_k^{-1}) = E_k(\ldots(E_2(E_1 E_1^{-1})E_2^{-1})\ldots)E_k^{-1} = I_n$$

has an all zero row, which is a contradiction. □

We can use this to prove that a matrix can only have one right inverse:

Corollary 2.6.2. *Let A be an $n \times n$ matrix. Suppose C and C' are both right inverses for A. That is, assume $AC = I$ and $AC' = I$. Then $C = C'$*

Please note that this kind of 'cancellation property' does not work in general! Can you think of an example where $AC = AC'$ but $C \neq C'$?

*This is the kind of persnikety thing that never comes up when dealing with real numbers, since, in that realm, multiplication is commutative, and so $ab = 1$ and $ba = 1$ mean exactly the same thing. But never forget: matrix multiplication is **not** commutative!

Proof. Assume $AC = I$ and $AC' = I$. Then by the previous lemma, there is a matrix B so that $BA = I$. Then we get

$$C = IC = (BA)C = B(AC) = BI = B(AC') = (BA)C' = IC' = C'$$

□

Exercise 2.6.3. *Let A be an $n \times n$ matrix. Suppose it has a 'left inverse,' that is, that there is a matrix B so that $BA = I$. Then A also has a 'right inverse,' that is, there is a matrix C so that $AC = I$.*

Corollary 2.6.4. *(Also an exercise) Let A be an $n \times n$ matrix. Suppose B and B' are both left inverses for A. That is, assume $BA = I$ and $B'A = I$. Then $B = B'$*

So if a matrix has an inverse 'on one side,' then it has one 'on the other side,' too. And, finally, the 'two' inverses must be the same:

Lemma 2.6.5. *Suppose A has both a left inverse and a right inverse. Then the two inverses are the same. That is, suppose $BA = I$ and $AC = I$. Then $B = C$.*

Proof. Let $BA = I$ and $AC = I$. Then

$$B = BI = B(AC) = (BA)C = IC = C$$

□

Notice that this proof (as well as the proof of the above corollary) relies on the *associativity* of matrix multiplication.

Conclusion: If a matrix A has an inverse on either side, then that inverse works on both sides of A, and in fact, it is the the only inverse for A. So there is no ambiguity in just calling that matrix A^{-1}.

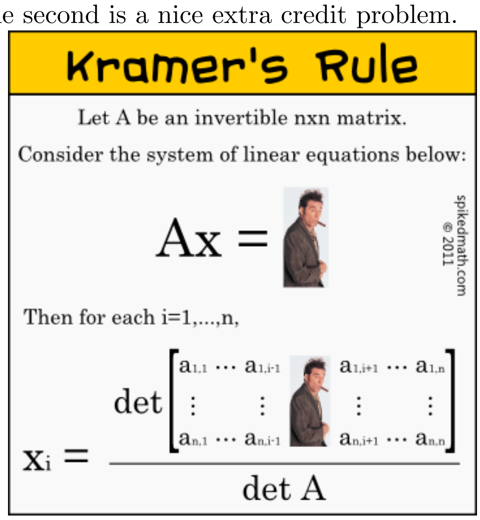
Chapter 3

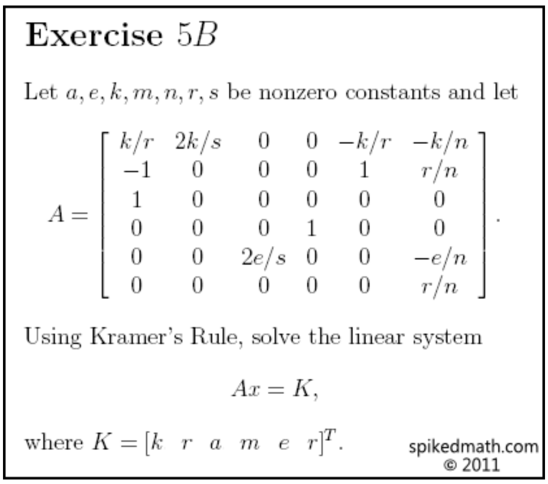
Determinants

My only comment here is that I do not usually give the permutation definition of determinant. I simply assert that the cofactor definition does not depend on the choice of row or column, though we do explore this fact with an example. ☹

Exercises in Chapter 3: 2pq, 3, 4, 6, 7a*, 8, 9, 10

These graphics are fun. The first a correct statement of Cramer's Rule. The second is a nice extra credit problem.





*skippable

Chapter 4

Vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$

This chapter is mostly a review of precalculus. I skim 4.6, and skip 4.7 entirely. For 4.6, for example, I may just include

- Equation of a plane
- An example of finding the equation of a plane containg three given points
- Use the same example to show the method preceeding Example 4.6.5 in Chen
- State the formula for the distance froma point to a plane, and leave the proof as an exercise. (This is analagous to the 2 dimensional case, anyway)

I certainly skip all the stuff with parametric forms. Even though, in hindsight, this does connect to the parametric forms from Chapter 1, it's a bit of a distraction here. This will be treated better as Proposition 6G.

<u>Exercises in Chapter 4:</u> 1, 4 (for the last part, find all unit vectors perpendicular to both u and v), 5, 6 (do this without resorting to coordinates), 9 (For 9a, compute directly. For 9b, the student must use 9a to deduce it.), 10bdefg

Chapter 5

Vector spaces

Exercises in Chapter 5: 4, 1, 2, 3, 6, 5*

*skippable

Chapter 6

Row Space, Column Space, Null Space: Rank and Nullity

Exercises in Chapter 6: 1 (but also find a basis for the null space in each case)

Chapter 7

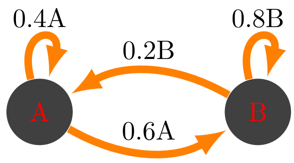
Eigenvalues and Eigenvectors

These two sections are new material for the end of this chapter.

Exercises in Chapter 7: 1, 2, 3, 4, 5, 6, 7, and prove the Cayley-Hamilton Theorem for 2x2 matrices.

7.1 An application of diagonalization to Markov Chains

Suppose there are two brands A and B, and let A_n and B_n represent the market share of each brand at time n. Supose that at each time increment, that 60% of A switches to B and 20% of B switches to A.



In other words, if a consumer currently buys brand A, then during the next time interval, they have a 40% chance of staying with brand A, and a 60% chance of sweitching to brand B. Similarly, if a consumer currently buys brand B, then during the next time interval, they have a 80% chance of staying with brand B, and a 20% chance of switching to brand A. Thus

$$\begin{pmatrix} A_{n+1} \\ B_{n+1} \end{pmatrix} = \begin{pmatrix} 0.4A_n + 0.2B_n \\ 0.6A_n + 0.8B_n \end{pmatrix} = \begin{pmatrix} 0.4 & 0.2 \\ 0.6 & 0.8 \end{pmatrix} \begin{pmatrix} A_n \\ B_n \end{pmatrix}$$

and so we get the pattern

$$\begin{pmatrix} A_{n+k} \\ B_{n+k} \end{pmatrix} = \begin{pmatrix} 0.4 & 0.2 \\ 0.6 & 0.8 \end{pmatrix}^k \begin{pmatrix} A_n \\ B_n \end{pmatrix}$$

So to see what happens in this situation 'in the long run', we let $E = \begin{pmatrix} 0.4 & 0.2 \\ 0.6 & 0.8 \end{pmatrix}$ and evaluate E^k for very large values of k. Now E has eigenvalues $\lambda = 1$ and $\lambda = 0.2$ with corresponding eigenvectors $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$.* So that

*Check my work!

gives us:

$$\begin{pmatrix}1 & 0\\ 0 & 0.2\end{pmatrix} = \begin{pmatrix}1 & 1\\ 3 & -1\end{pmatrix}^{-1}\begin{pmatrix}0.4 & 0.2\\ 0.6 & 0.8\end{pmatrix}\begin{pmatrix}1 & 1\\ 3 & -1\end{pmatrix}$$

$$\begin{pmatrix}1 & 0\\ 0 & 0.2\end{pmatrix} = \begin{pmatrix}\frac{1}{4} & \frac{1}{4}\\ -\frac{3}{4} & -\frac{1}{4}\end{pmatrix}\begin{pmatrix}0.4 & 0.2\\ 0.6 & 0.8\end{pmatrix}\begin{pmatrix}1 & 1\\ 3 & -1\end{pmatrix}$$

$$\text{So } E = \begin{pmatrix}0.4 & 0.2\\ 0.6 & 0.8\end{pmatrix} = \begin{pmatrix}1 & 1\\ 3 & -1\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 0.2\end{pmatrix}\begin{pmatrix}\frac{1}{4} & \frac{1}{4}\\ -\frac{3}{4} & -\frac{1}{4}\end{pmatrix}$$

$$\text{So } E^k = \begin{pmatrix}0.4 & 0.2\\ 0.6 & 0.8\end{pmatrix} = \begin{pmatrix}1 & 1\\ 3 & -1\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 0.2\end{pmatrix}^k\begin{pmatrix}\frac{1}{4} & \frac{1}{4}\\ -\frac{3}{4} & -\frac{1}{4}\end{pmatrix} = \begin{pmatrix}1 & 1\\ 3 & -1\end{pmatrix}\begin{pmatrix}1^k & 0\\ 0 & 0.2^k\end{pmatrix}\begin{pmatrix}\frac{1}{4} & \frac{1}{4}\\ -\frac{3}{4} & -\frac{1}{4}\end{pmatrix}$$

So as $k \to \infty$, E^k approaches

$$E^\infty = \begin{pmatrix}1 & 1\\ 3 & -1\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}\frac{1}{4} & \frac{1}{4}\\ -\frac{3}{4} & -\frac{1}{4}\end{pmatrix} = \begin{pmatrix}1 & 1\\ 3 & -1\end{pmatrix}\begin{pmatrix}\frac{1}{4} & \frac{1}{4}\\ 0 & 0\end{pmatrix} = \begin{pmatrix}\frac{1}{4} & \frac{1}{4}\\ \frac{3}{4} & \frac{3}{4}\end{pmatrix}$$

Finally, this means that

$$\begin{pmatrix}A_\infty\\ B_\infty\end{pmatrix} = E^\infty\begin{pmatrix}A_n\\ B_n\end{pmatrix} = \begin{pmatrix}\frac{1}{4} & \frac{1}{4}\\ \frac{3}{4} & \frac{3}{4}\end{pmatrix}\begin{pmatrix}A_n\\ B_n\end{pmatrix} = \begin{pmatrix}\frac{1}{4}A_n + \frac{1}{4}B_n\\ \frac{3}{4}A_n + \frac{3}{4}B_n\end{pmatrix} = \begin{pmatrix}\frac{1}{4}(A_n + B_n)\\ \frac{3}{4}(A_n + B_n)\end{pmatrix}$$

So 'in the long run', brand A ends up with $\frac{1}{4}$ of the total market share, while brand B ends up with $\frac{3}{4}$ of the total market share.

Finally, we can see that this limit state is also an equilibrium state for this system. Suppose that brand A currently has x market share, and that brand B currently has $3x$ of the market share. Then in the next interation:

$$\begin{pmatrix}0.4 & 0.2\\ 0.6 & 0.8\end{pmatrix}\begin{pmatrix}x\\ 3x\end{pmatrix} = \begin{pmatrix}x\\ 3x\end{pmatrix}$$

....we see that each brand has the same amount of market share as before.

7.2 Cayley-Hamilton Theorem

In this section, we work to analyze the characteristic polynomial of a matrix. Recall that if A is a matrix, then the *characteristic equation of* A is $\det(A - xI) = 0$, and that the *characteristic polynomial of* A is $\det(A - xI)$. Before, we used the variable λ to emphasize that this equation is used to find eigenvalues, but here we use the variable x to emphasize that this object is a function. In fact, it is true to its name:

Lemma 7.2.1. *If A is an $n \times n$ matrix, then the characteristic polynomial of A is a polynomial of degree n.*

Proof. Exercise □

So for the rest of this section, we let A be an $n \times n$ matrix, and let its characteristic polynomial be

$$a_n x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \ldots + a_3x^3 + a_2x^2 + a_1x + a_0$$

where $a_n, a_{n-1}, a_{n-2}, \ldots, a_3, a_2, a_1, a_0$ are real numbers.

Lemma 7.2.2. *Let $a_nx^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \ldots + a_3x^3 + a_2x^2 + a_1x + a_0$ be the characteristic polynomial of the matrix A. Then $a_0 = \det(A)$*

Proof. If we let $x = 0$ in the definition of the characteristic polynomial, then we get the equation:

$$\det(A - 0I) = a_n 0^n + a_{n-1}0^{n-1} + a_{n-2}0^{n-2} + \ldots + a_3 0^3 + a_2 0^2 + a_1 0 + a_0$$

which simplifies to $\det(A) = a_0$. □

Corollary 7.2.3. *Let $a_nx^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \ldots + a_3x^3 + a_2x^2 + a_1x + a_0$ be the characteristic polynomial of the matrix A. Then*

$$a_0 = 0 \iff \det(A) = 0 \iff A \textit{ is not invertible}$$

The main result of this section is the following theorem, which says that if we 'plug in' the matrix A for x in its own characteristic polynomial, then the result is 0. This may *seem* obvious, since $\det(A - AI) = \det(\mathcal{O}) = 0$. However, this is a false proof, as the variable x stands for a scalar, not a matrix! And we know very well that matrix multiplication does *not* work the same as scalar multiplication. Even then, note that the determinant is a scalar, while a polynomial funcion of a matrix is again a matrix. So really, this is quite surprising:

Theorem 7.2.4 (Cayley-Hamilton)**.** *Let $a_nx^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \ldots + a_3x^3 + a_2x^2 + a_1x + a_0$ be the characteristic polynomial of the matrix A. Then*

$$a_nA^n + a_{n-1}A^{n-1} + a_{n-2}A^{n-2} + \ldots + a_3A^3 + a_2A^2 + a_1A + a_0I = \mathcal{O}$$

Before we prove this theorem, let's see why it is useful. First, we get a new formula for A^{-1}:

Corollary 7.2.5. *Let* $a_nx^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \ldots + a_3x^3 + a_2x^2 + a_1x + a_0$ *be the characteristic polynomial of the matrix* A*. Suppose also that* A *is invertible. Then*

$$A^{-1} = -\frac{1}{a_0}\left(a_nA^{n-1} + a_{n-1}A^{n-2} + a_{n-2}A^{n-3} + \ldots + a_3A^2 + a_2A + a_1I\right)$$

Proof. By Cayley-Hamilton,

$$a_nA^n + a_{n-1}A^{n-1} + a_{n-2}A^{n-2} + \ldots + a_3A^3 + a_2A^2 + a_1A + a_0I = \mathcal{O}$$

$$a_nA^n + a_{n-1}A^{n-1} + a_{n-2}A^{n-2} + \ldots + a_3A^3 + a_2A^2 + a_1A = -a_0I$$

$$\left(a_nA^{n-1} + a_{n-1}A^{n-2} + a_{n-2}A^{n-3} + \ldots + a_3A^2 + a_2A + a_1I\right)A = -a_0I$$

Now since A is invertible, $a_0 \neq 0$, so we get:

$$-\frac{1}{a_0}\left(a_nA^{n-1} + a_{n-1}A^{n-2} + a_{n-2}A^{n-3} + \ldots + a_3A^2 + a_2A + a_1I\right)A = I$$

which means that $-\frac{1}{a_0}\left(a_nA^{n-1} + a_{n-1}A^{n-2} + a_{n-2}A^{n-3} + \ldots + a_3A^2 + a_2A + a_1I\right)$ is the inverse of A. □

Next we can use Cayley-Hamilton to calculate powers of A as linear combinations of $I, A, A^2, A^3, \ldots, A^{n-1}$:

Example 7.2.6. Let $A = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}$ Then $\det(A - xI) = x^2 - 2x + 1$. So by Cayley-Hamilton, $A^2 - 2A + I = \mathcal{O}$, and so $A^2 = 2A - I$. But then if we multiply both sides by A, substitute, and simplify, we get a formula for A^3:

$$A^3 = (2A - I)A = 2A^2 - A = 2(2A - I) - A = 3A - 2I$$

again, multiplying both sides by A, we get a formula for A^4:

$$A^4 = (3A - 2I)A = 3A^2 - 2A = 3(2A - I) - 2A = 4A - 3I$$

and so on...

Finally, let's prove the main theorem:

proof of Cayley-Hamilton. Again, let $\det(A - xI) = a_nx^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \ldots + a_3x^3 + a_2x^2 + a_1x + a_0$. Consider the adjoint $\text{adj}(A - xI)$, which is the transpose of the cofactor matrix of $A - xI$. Thus each entry of $\text{adj}(A - xI)$ is a polynomial of degree at most $n - 1$. So $\text{adj}(A - xI)$ can be expressed as

$$\text{adj}(A - xI) = x^{n-1}B_{n-1} + x^{n-2}B_{n-2} + \ldots + xB_1 + B_0$$

where $B_{n-1}, B_{n-2}, \ldots, B_1, B_0$ are $n \times n$ matrices with scalar entries. *

Then

$$(A - xI)\frac{1}{\det(A - xI)}\operatorname{adj}(A - xI) = I$$

$$(A - xI)\operatorname{adj}(A - xI) = \det(A - xI)I$$

$$(A - xI)(x^{n-1}B_{n-1} + x^{n-2}B_{n-2} + \ldots + xB_1 + B_0) = (a_nx^n + a_{n-1}x^{n-1} + \ldots + a_1x + a_0)I$$

$$(x^{n-1}AB_{n-1} + x^{n-2}AB_{n-2} + \ldots + xAB_1 + AB_0) + (x^nB_{n-1} + x^{n-1}B_{n-2} + \ldots + x^2B_1 + xB_0) =$$
$$= a_nx^nI + a_{n-1}x^{n-1}I + \ldots + a_1xI + a_0I$$

Then matching the coefficients in like powers of x in this equation, we get the following system of equations:

$$\begin{aligned}
-B_{n-1} &= a_nI \\
-B_{n-2} + AB_{n-1} &= a_{n-1}I \\
-B_{n-3} + AB_{n-2} &= a_{n-2}I \\
&\vdots \\
-B_1 + AB_2 &= a_2I \\
-B_0 + AB_1 &= a_1I \\
AB_0 &= a_0I
\end{aligned}$$

next, multiply each side of each equation *on the left* by different powers of A:

$$\begin{aligned}
A^n(-B_{n-1}) &= A^n(a_nI) \\
A^{n-1}(-B_{n-2} + AB_{n-1}) &= A^{n-1}(a_{n-1}I) \\
A^{n-2}(-B_{n-3} + AB_{n-2}) &= A^{n-2}(a_{n-2}I) \\
&\vdots \\
A^2(-B_1 + AB_2) &= A^2(a_2I) \\
A(-B_0 + AB_1) &= A(a_1I) \\
AB_0 &= a_0I
\end{aligned}$$

*For example, if $\operatorname{adj}(A - xI) = \begin{pmatrix} x^2 & x^2 + x + 1 \\ 2x - 1 & 0 \end{pmatrix}$, then it also can be expressed as $x^2\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} + x\begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$

which simplifies to:

$$\begin{aligned}
-A^n B_{n-1} &= a_n A^n \\
-A^{n-1} B_{n-2} + A^n B_{n-1} &= a_{n-1} A^{n-1} \\
-A^{n-2} B_{n-3} + A^{n-1} B_{n-2} &= a_{n-2} A^{n-2} \\
&\vdots \\
-A^2 B_1 + A^3 B_2 &= a_2 A^2 \\
-A B_0 + A^2 B_1 &= a_1 A \\
A B_0 &= a_0 I
\end{aligned}$$

Finally, if we add all these equations together, the sum on the left telescopes to $\mathcal{O}$, and we obtain

$$\mathcal{O} = a_n A^n + a_{n-1} A^{n-1} + a_{n-2} A^{n-2} + \ldots + a_3 A^3 + a_2 A^2 + a_1 A + a_0 I$$

□

Chapter 8

Linear Transformations

Up to and including Section 8.7, follow Chen's notes. The following section replaces 8.8 through 8.10. The section after that is new material.

I save some time in class by putting the pdf of Chen on the screen to go over Examples 8.8.1 and 8.8.2.

Exercises in Chapter 8: 1a, 2, 3a*, 4b*c*de, 5, 6, 8, 15ac, 16*, 19, 21, 22bcd

8.1 Similar Matrices

Question 8.1.1. *What happens when we represent the same linear operator with respect to different bases?*

Generally, we can define the matrix of a linear transformation from one vector space to another, with respect to some fixed bases:

Definition 8.1.2. Suppose $T : V \to W$ is a linear transformation, where V and W are finite-dimensional. Suppose further that $\mathcal{B} = \{v_1, \ldots, v_n\}$ is a basis for V and that $\mathcal{C} = \{w_1, \ldots, w_m\}$ is a basis for W. Then we define the matrix

$$A = (\ [T(v_1)]_{\mathcal{C}}\ [T(v_2)]_{\mathcal{C}}\ \ldots\ [T(v_n)]_{\mathcal{C}})$$

to be *the matrix for the linear transformation T with respect to $\mathcal{B}$ and $\mathcal{C}$*

The columns of A tell us, in terms of $\mathcal{C}$, what T does to the elements of $\mathcal{B}$. Furthermore, multiplication by A is the same as applying T (with respect to these bases):

Exercise 8.1.3. *For any vector $v \in V$, $A[v]_{\mathcal{B}} = [T(v)]_{\mathcal{C}}$*

Exercise 8.1.4. *If T is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ and if the bases are the standard bases, then A is just the standard matrix for T.*

This construction becomes quite interesting if, instead of looking at different transformations, we instead look at a particular linear operator and see what happens to the matrix A when we change the basis in question.

Definition 8.1.5. If, as above, $V = W$ and $\mathcal{B} = \mathcal{C}$, the matrix A is called *the matrix for the linear operator T with respect to the basis $\mathcal{B}$.*

*Skippable

(Now look at Examples 8.8.1 and 8.8.2 in Chen)

In Example 8.8.2, we see that $A = \begin{pmatrix} 1 & -1 \\ 1 & 4 \end{pmatrix}$ represents T with respect to the basis $\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$, and that $A' = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$ represents T with respect to the standard basis $\mathcal{E} = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$.

So how are A and A' related?

More generally, suppose that $T : V \to V$ is a linear operator and that $\mathcal{B} = \{u_1, \ldots, u_n\}$ and $\mathcal{B}' = \{v_1, \ldots, v_n\}$ are both bases for V. Let A be the matrix for T with respect to $\mathcal{B}$. Let A' be the matrix for T with respect to $\mathcal{B}'$.

Now let $v, w \in V$ so that $T(v) = w$. Then $A[v]_{\mathcal{B}} = [w]_{\mathcal{B}}$ and $A'[v]_{\mathcal{B}'} = [w]_{\mathcal{B}'}$. Next let

$$\begin{aligned} P &= (\ [T(v_1)]_{\mathcal{B}}\ [T(v_2)]_{\mathcal{B}}\ \ldots\ [T(v_n)]_{\mathcal{B}}) \\ P^{-1} &= (\ [T(v_1)]_{\mathcal{B}'}\ [T(v_2)]_{\mathcal{B}'}\ \ldots\ [T(v_n)]_{\mathcal{B}'}) \end{aligned}$$

be the transition matrices from $\mathcal{B}'$ to $\mathcal{B}$ and $\mathcal{B}$ to $\mathcal{B}'$, respectively. Putting these al together, we get $P[v]_{\mathcal{B}'} = [v]_{\mathcal{B}}$ and $P[w]_{\mathcal{B}'} = [w]_{\mathcal{B}}$. So $[w]_{\mathcal{B}'} = P^{-1}[w]_{\mathcal{B}}$. Therefore,

$$\begin{aligned} A'[v]_{\mathcal{B}'} &= [w]_{\mathcal{B}'} \\ &= P^{-1}[w]_{\mathcal{B}} \\ &= P^{-1}A[v]_{\mathcal{B}} \\ &= P^{-1}AP[v]_{\mathcal{B}'} \end{aligned}$$

But since the choice of v was arbitrary, it must be the case that $A' = P^{-1}AP$.

Example 8.1.6. We can check that this is true for Example 8.8.2. In fact, $P = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$ and $P^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ are the transition matrices from $\mathcal{E}$ to $\mathcal{B}$ and $\mathcal{B}$ to $\mathcal{E}$, respectively. Then

$$P^{-1}AP = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} = A'$$

It turns out that this purely algebraic relation is itself quite interesting. Thus it has a name:

Definition 8.1.7. If A and A' are square matrices satisfying $P^{-1}AP = A'$, for some invertible matrix P, then we say that A and A' are *similar*.

So we have just proved:

Theorem 8.1.8. *If A and A' represent the same linear operator, but with respect to different bases, then they are similar.*

The converse is also true. Suppose $P^{-1}AP = A'$, and let T be the operator for which A is the *standard* matrix. It is easy to show that A' is then the matrix for T with respect to the basis given by the columns of P. (See again the above example.)

In fact, we have seen this simlarity relation before. Another way to say that *a matrix is diagonalizable* is that *it is similar to a diagonal matrix.* So let's explore this concept a bit more.

Lemma 8.1.9. *Similarity is an equivalence relation.*

Therefore, we can just say "A and A' are similar to each other", rather than having to specify "A is similar to A'" or "A' is similar to A" each time.

Proof. Reflexivity: $A = I^{-1}AI$

Symmetry: Suppose A is similar to A'. That is, $P^{-1}AP = A'$, for some invertible matrix P. Then multiplying both sides by P on the left and P^{-1} on the right gives:

$$\begin{aligned} P^{-1}AP &= A' \\ P(P^{-1}AP)P^{-1} &= P(A')P^{-1} \\ A &= (P^{-1})^{-1}(A')P^{-1} \end{aligned}$$

Then A' is similar to A

Transitivity: Suppose A is similar to B and that B is similar to C. That is, $P^{-1}AP = B$ and $Q^{-1}BQ = C$, for some invertible matrices P, Q. Then

$$C = Q^{-1}BQ = Q^{-1}P^{-1}APQ = (PQ)^{-1}A(PQ)$$

So A is similar to C □

In group theory, this 'similarity' relation is called 'conjugacy'. Notice that the above proof used no facts about matrices, just associativity, the existence of an identity, and that $Q^{-1}P^{-1} = (PQ)^{-1}$. That said, when two matrices are similar, they really do have a lot in common:

Theorem 8.1.10. *Suppose A and A' are similar matrices. Then*

(a) $\text{Rank}(A) = \text{Rank}(A')$

(b) $\det(A) = \det(A')$

(c) $\chi(A) = \chi(A')$

(d) A and A' have the exact same eigenvalues.

(e) $\operatorname{tr}(A) = \operatorname{tr}(A')$

$\chi(A)$ denotes the characteristic polynomial of A, and $\operatorname{tr}(A)$ is the *trace* of A, which is the sum of the diagonal entries of A. Literally, $\operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii}$.

Proof. Suppose $P^{-1}AP = A'$

(b)

$$\begin{aligned}\det(A') &= \det(P^{-1}AP) \\ &= \det(P^{-1})\det(A)\det(P) \\ &= \frac{1}{\det(P)}\det(A)\det(P) \\ &= \det(A)\end{aligned}$$

(c)

$$\begin{aligned}\chi(A') &= \chi(P^{-1}AP) \\ &= \det(P^{-1}AP - \lambda I) \\ &= \det(P^{-1}AP - \lambda P^{-1}IP) \\ &= \det(P^{-1}AP - P^{-1}\lambda IP) \\ &= \det\left(P^{-1}(AP - \lambda IP)\right) \\ &= \det\left(P^{-1}(A - \lambda I)P\right) \\ &= \det(P^{-1})\det(A - \lambda I)\det(P) \\ &= \frac{1}{\det(P)}\det(A - \lambda I)\det(P) \\ &= \det(A - \lambda I) \\ &= \chi(A)\end{aligned}$$

(a) First, as an exercise, prove that, for any two matrices A, B,

$$\operatorname{Rank}(AB) \leq \operatorname{Rank}(A) \text{ and } \operatorname{Rank}(AB) \leq \operatorname{Rank}(B).$$

Then $\operatorname{Rank}(A') = \operatorname{Rank}(P^{-1}AP) \leq \operatorname{Rank}(A)$. Similarly, $\operatorname{Rank}(A) = \operatorname{Rank}(PA'P^{-1}) \leq \operatorname{Rank}(A')$. Thus $\operatorname{Rank}(A) = \operatorname{Rank}(A')$.

(d) First, note that for any two matrices A, B,

$$\operatorname{tr}(AB) = \sum_{i=1}^{n}(AB)_{ii} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ji} = \sum_{j=1}^{n}\sum_{i=1}^{n} b_{ji}a_{ij} = \sum_{j=1}^{n}(BA)_{jj} = \operatorname{tr}(BA)^{*}$$

*We can visualize how this works by looking at 2×2 matrices:

$$\operatorname{tr}\left[\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} e & f \\ g & h \end{pmatrix}\right] = \operatorname{tr}\begin{pmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{pmatrix} = ae+bg+cf+dh$$

$$\operatorname{tr}\left[\begin{pmatrix} e & f \\ g & h \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right] = \operatorname{tr}\begin{pmatrix} ea+fc & eb+fd \\ ga+hc & gb+hd \end{pmatrix} = ae+fc+gb+hd$$

Then $\mathrm{tr}(A') = \mathrm{tr}([P^{-1}A]P) = \mathrm{tr}(P[P^{-1}A]) = \mathrm{tr}(A)$. □

We can even say that A and A' have the same eigenvectors:

Theorem 8.1.11. *Suppose A and A' represent the same linear operator $T : V \to V$ with respect to bases $\mathcal{B}$ and $\mathcal{B}'$. Let $v \in V$ be a vector. Then $[v]_{\mathcal{B}}$ is an eigenvector of A with corresponding eigenvalue λ if and only if $[v]_{\mathcal{B}'}$ is an eigenvector of A' with corresponding eigenvalue λ.*

Proof. Assume $[v]_{\mathcal{B}}$ is an eigenvector of A with corresponding eigenvalue λ. Let P be the transition matrix from $\mathcal{B}'$ to $\mathcal{B}$. As above, we have $A' = P^{-1}AP$. Thus

$$A'[v]_{\mathcal{B}'} = P^{-1}AP[v]_{\mathcal{B}'} = P^{-1}A[v]_{\mathcal{B}} = P^{-1}\lambda[v]_{\mathcal{B}} = \lambda P^{-1}[v]_{\mathcal{B}} = \lambda[v]_{\mathcal{B}'}$$

So $[v]_{\mathcal{B}'}$ is an eigenvector of A' with corresponding eigenvalue λ. □

To summarize: the geometric fact of two matrices representing the same linear operator is equivalent to the algebraic fact of the two matrices being similar. Furthermore, when two matrices do represent the same linear operator (i.e., are similar) then they also have the same rank, determinant, characterisitc polynomial, eigenvalues, and trace. They also have the same eigenvectors, but with respect to different bases.

This equivalence has some surprising applications. We will explore one of them, namely, that we can prove a fact about eigenspaces that we probably suspected (but couldn't prove) back in Chapter 7.

8.2 Multiplicity of eigenvalues

Definition 8.2.1. Suppose λ_0 is an eigenvalue of the matrix A.

- The *geometric multiplicity* of λ_0 is the dimension of the corresponding eigenspace.
- The *algebraic mulitplicity* of λ_0 is the multiplicity of the root λ_0 in $\chi(A)$. In other words, the number of factors of $(x - \lambda_0)$ in $\chi(A)$.

Example 8.2.2. Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Then $\chi(A) = (x-1)^2$, so the eigenvalue $\lambda = 1$ has algebraic multiplicity 2. On the other hand, a basis for the eigenspace corresponding to this eigenvalue is $\{\begin{pmatrix} 1 & 0 \end{pmatrix}\}$. Thus the eigenvalue $\lambda = 1$ has algebraic multiplicity 1.

If we compare this with all the examples from Chapter 7, we see that there is a pattern here. But now we can prove it:

Theorem 8.2.3. *If λ_0 is an eigenvalue of the matrix A, then*

$$(\textit{geometric multiplicity of } \lambda_0) \le (\textit{algebraic multiplicity of } \lambda_0)$$

Recall that an $n \times n$ matrix is diagonalizable if and only if we can find n linearly independent eigenvectors. Therefore, the above theorem implies that A is diagonalizable if and only if:

- We can factor $\chi(A)$ into n linear factors; and
- The geometric multiplicity of each eigenvalue is equal to its algebraic multiplicity.

Proof of theorem. Let A be an $n \times n$ matrix, and let λ_0 be an eigenvalue of A. Suppose λ_0 has geometric multiplicity of k, for some $k \le n$. Then let $\{v_1, v_2, \dots v_k\}$ be a basis for this eigenspace. Now extend $\{v_1, v_2, \dots v_k\}$ to a basis for $\mathbb{R}^n$:

$$\{v_1, v_2, \dots v_k, v_{k+1}, \dots, v_n\}$$

and let $\{e_1, e_2, \dots, e_n\}$ be the standard basis for $\mathbb{R}^n$. Finally, let T be the linear transformation whose *standard* matrix is A.

Let A' be the matrix for T wirth respect to the basis $\{v_1, v_2, \dots v_k, v_{k+1}, \dots, v_n\}$. Note that if $1 \le i \le k$, then $T(v_i) = Av_i = \lambda_0 v_i$. Therefore the ith column of A' is $\lambda_0 e_i$, that is, the column with λ_0 in the$i$th entry and zeroes everywhere else. Thus A' looks like the following, where the upper left quadrant is a $k \times k$ matrix.

$$\left(\begin{array}{ccc|c} \lambda_0 & & 0 & \\ & \ddots & & \text{stuff} \\ 0 & & \lambda_0 & \\ \hline & 0 & & \text{stuff} \end{array}\right)$$

Therefore,

$$\begin{aligned} \chi(A') &= \det(A' - xI) \\ &= \det \left(\begin{array}{ccc|c} (\lambda_0 - x) & & 0 & \\ & \ddots & & \text{stuff} \\ 0 & & (\lambda_0 - x) & \\ \hline & 0 & & \text{stuff} \end{array}\right) \\ &= (\lambda_0 - x)^k (\text{something}) \end{aligned}$$

But since A and A' are matrix representations of the linear operator T with respect to different bases, they are similar. Thus

$$\chi(A) = \chi(A') = (\lambda_0 - x)^k(\text{something}).$$

In other words, this polynomial has *at least* k factors of $(\lambda_0 - x)$. So

$$(\text{the algebraic multiplicity of } \lambda_0 \text{ in } A) \geq k = (\text{the geometric multiplicity of } \lambda_0 \text{ in } A)$$

$\square$

Chapter 9

Real Inner Product Spaces

How much I get into the Gram-Schidt process, as well as Chapter 10, is time dependent. One semester, I was only able to prove GS, and then skim the examples from the book by putting the pdf on the screen. On the other hand, during another semester, we fully finished Chapter 9, covered the good bits of Chapter 10, and then had a student presentation of Chapter 12!

In either case, I leave the entire last class for my Oddtown presentation, which is the appendix in these notes.

Exercises in Chapter 9: 1c, 3b, 5b, 6, 7, 10, 11, 12, 13, 14*, 15, 16, 17 (for #15, find an orthogonal basis for spanu1,u2, u3 instead.)

Exercises in Chapter 10: 2, 3, 4, 5, 6, 9e

*skippable

Appendix A

Oddtown

This section is simply a restatement of the first section of the book Linear Algebra Methods in Combinatorics by Babai and Frankl.

We will describe two towns, Eventown and Oddtown, where the inhabitants like to form clubs. In each city, they have the strange law that no two clubs can have the exact same list of members. Mathematically, this means that each club is a different subset of the town. Now, let's suppose that Eventown and Oddtown both have 40 inhabitants. With just this rule, that means that there are $2^40 = 1,099,511,627,776$ possible clubs, as this is the number of subsets of a set with 40 elements.

In response to this, the city council in Eventown passed two laws:

(E1) Each club must have an even number of members.

(E2) Any two clubs must share an even number of members.

Certaintly Eventown can no longer have 2^{40} clubs. But how many could it have? One way to get a lot of clubs is for the inhabitants to form 20 married couples, and then make sure that if one inhabitant joins a club, then their spouse does as well. This construction gives $2^{20} = 1,048,576$ possible clubs, as this is the numbner of ways to choose subsets of the set of married couples. That's still a lot of clubs! It turns out that this construction is optimal, though the proof is a bit much for this class:

Theorem (Berlekamp 1969, Graver 1975)**.** *If Eventown has n inhabitants, then there are at most $2^{\frac{n}{2}}$ clubs in Eventown.*

Well the city council in Oddtown sees that Eventown still has over a million clubs, so they decide to pass a slightly different set of laws:

(O1) Each club must have an **odd** number of members.

(O2) Any two clubs must share an even number of members.

So how many clubs could Oddtown have? One simple construction is to make 40 differnt clubs, each with one member. But this is not very many clubs, is it? Well, it turns out that this is optimal!

Theorem (Berlekamp 1969)**.** *If Oddtown has n inhabitants, then there are at most n clubs in Oddtown.*

This is very surprising, given the fact that Eventown can have over a million clubs! But how can we prove this? We first need to discuss:

Linear Algebra with fields other than $\mathbb{R}$

Everything we have done this semester has been done with scalars that are real numbers. However, most[*] of what we have seen could have been proven using any other field besides $\mathbb{R}$. So what is a field?

Definition. A set F together with binary operations $+, \cdot$ and constants $0, 1$ forms a **field** if

- $+, \cdot$ are both associative and commutative
- $0, 1$ are the additive and multiplicative identities, respectively
- Every element of F has an additive inverse
- Every element of F (except for 0) has a multiplicative inverse
- Distributivity: $a \cdot (b + c) = a \cdot b + a \cdot c$ and $(b + c) \cdot a = b \cdot a + c \cdot a$
- Closure properties: If a, b are elements of F, then $a + b$ and $a \cdot b$ are elements of F

So $\mathbb{R}$ forms a field, as do $\mathbb{C}$[†] and $\mathbb{Q}$. Note that $\mathbb{Z}$ is not a field, as most elements do not have multiplicative inverses. Also $\mathbb{N}$ is not a field, as most elements do not have multiplicative or additive inverses. However, in that latter two examples, all other properties of a field are satisfied.

For Oddtown, what we are really interested in is finite fields. The easiest finite fields to understand are those that come from modular arithemtic. If we draw our the addition tables for $\mathbb{F}_2$, $\mathbb{F}_3$, and $\mathbb{F}_4$[‡], we can see that $\mathbb{F}_2$ and $\mathbb{F}_3$ are fields, but that $\mathbb{F}_4$ is not[§]. It turns out that $\mathbb{F}_n$ is a field precisely when n is prime, but that is a discussion for another time. For now, we focus on $\mathbb{F}_2$.

---

[*] A good exercise would be to see which theorems actually used the fact that the scalars were real numbers...

[†] If we define $\mathbb{C}$ to be the set of numbers of the form $a+bi$, where a, b are real numbers, then the one property of fields that is not obvious is that multiplicative inverses exist. However $(a + bi)^{-1} = \frac{a - bi}{a^2 + b^2} = \frac{a}{a^2 + b^2} + \frac{-b}{a^2 + b^2} i$

[‡] Here's where my LaTeX patience ran out...

[§] Here, the key is again to check for the existence of multiplicative inverses, but the distributive property probably warrants an exercise, as well.

Now back to Oddtown...

To prove the Oddtown theorem, we encode each club as a vector in the vector space of dimension 40 over the field $\mathbb{F}_2$. In other words, each club is a list of 40 zeros and ones. Each coordinate represents an inhabitant of Oddtown. A given club will then have a 1 in that coordinate if that inhabitant is a member of the club, and a 0 if they are not. For example the club containing inhabitants 2,4,7 would correspond to the vector

$$(0,1,0,1,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)$$

Thus we are working in the vector space $(\mathbb{F}_2)^{40}$. As before, this makes a 40 dimensional vector space, with the usual dot product.*

What does this dot product look like? Suppose $u, v \in (\mathbb{F}_2)^{40}$. Then

$$\begin{aligned} u \cdot v &= u_1 v_1 + u_2 v_2 + \ldots + u_{40} v_{40} \\ &= \begin{cases} 0 & \text{if } u \text{ and } v \text{ share an even number of ones} \\ 1 & \text{if } u \text{ and } v \text{ share an odd number of ones} \end{cases} \end{aligned}$$

For example

$$\begin{aligned} (0,\mathbf{1},1,0,\mathbf{1}) \cdot (1,\mathbf{1},0,0,\mathbf{1}) &= 0 \cdot 1 + 1 \cdot 1 + 1 \cdot 0 + 0 \cdot 0 + 1 \cdot 1 \\ &= 0 + \mathbf{1} + 0 + 0 + \mathbf{1} \\ &= 0 \end{aligned}$$

Claim. *The club vectors of Oddtown form a linearly independent set*

Thus there are at most 40 clubs, since the vector space has dimension 40!

Proof of claim. Suppose $v_1, v_2, \ldots, v_R$ are club vectors and that $c_1, c_2, \ldots, c_R \in \mathbb{F}_2$ are scalars such that

$$c_1 v_1 + c_2 v_2 + \ldots + c_R v_R = \mathcal{O}$$

Let $1 \leq i \leq n$. We want to show that $c_i = 0$. We take the dot product of both sides of the above equation with v_i:

$$\begin{aligned} (c_1 v_1 + c_2 v_2 + \ldots + c_i v_i + \ldots + c_R v_R) \cdot v_i &= \mathcal{O} \cdot v_i \\ c_1(v_1 \cdot v_i) + c_2(v_2 \cdot v_i) + \ldots + c_i(v_i \cdot v_i) + \ldots + c_R(v_R \cdot v_i) &= 0 \\ c_1(0) + c_2(0) + \ldots + c_i(1) + \ldots + c_R(0) &= 0 \\ 0 + 0 + \ldots + c_i + \ldots + 0 &= 0 \\ c_i &= 0 \end{aligned}$$

□

*Be careful, though, as this is not an inner product. $\|(1,1)\| = \sqrt{(1,1)\cdot(1,1)} = \sqrt{1 \cdot 1 + 1 \cdot 1} = \sqrt{1+1} = 0$. So this is not positive definite.