Math 34600 L (43597) - Lectures 01

Ethan Akin
Office: Marshak 325
Email: eakin@ccny.cuny.edu

Spring, 2024

Contents

Introduction

- The book is W. Keith Nicholson *Linear Algebra with Applications* . A pdf version of the book is free and you can order an inexpensive hardcopy.
- Keep up with the homework. I will be collecting it every Thursday.
- Ask questions.
- The course information sheet with the term's homework assignments is posted on my site:

  $http : /math.sci.ccny.cuny.edu//peoplename = EthanAkin$

- I will be posting there a pdf of the slides I am using here. The first set is already up. I will post the others as we get to them.
- The class will meet from 9:30am to 10:45 on Tuesdays and Thursdays in NAC 6/121. If due to inclement weather, I am unable to come in, or if we are switched to online mode, we will meet at the scheduled time using Blackboard Collaborate Ultra.
- Office: MR (Marshak) 325A

  Office Hours: Tuesday 11:00am-1:00pm, Thursday 11:00am-12:00.

  Email: eakin@ccny.cuny.edu

Number Facts

First recall the names of some properties of real numbers. For addition we have the

$$\text{Commutative Law:} \quad a + b = b + a,$$

$$\text{Associative Law:} \quad a + (b + c) = (a + b) + c.$$

We use these without thinking about them. We add up a bunch of numbers in any order we want.

For example, we may group the positives together and then the negatives and them compute the difference.

It is the same for multiplication.

On the other hand, when we combine multiplication and addition, we have the

$$\text{Distributive Law:} \quad a(b + c) = ab + ac.$$

Hidden here is the Order of Operations Rule of Notation we requires that we perform the multiplications before the additions. Thus, we need not write $(ab) + (ac)$.

This is important enough that its use in different directions have different names: From left to right it is “clearing parentheses” or “multiplying out”, while from right to left it is “common monomial factoring” or just “factoring”.

This is all familiar, but we should pause to look back at the commutative law for multiplication: $3 \cdot 5 = 5 \cdot 3$.

But remember multiplication is repeated addition. So

three 5's is $5 \ + \ 5 \ + \ 5$.

five 3's is $3 \ + \ 3 \ + \ 3 \ + \ 3 \ + \ 3$.

If the fact that these are equal doesn't surprise you, think about repeated multiplication

$5 \times 5 \times 5$ is not equal to $3 \times 3 \times 3 \times 3 \times 3$.

Why does it work for addition?

Look at this rectangle of 15 x's.

x x x x x
x x x x x
x x x x x

We can think of this as five columns of three x's each or as three rows of five x's each.

This RECTANGLE PICTURE where we think of the same rectangle of numbers either as a collection of rows or as a collection of columns is going to be essential throughout the course.

Furthermore, we will repeatedly use the fact that you get the same thing if you first add up the rows or if you first add up the columns.

That is, $5x + 5x + 5x = 3x + 3x + 3x + 3x + 3x$. Both are $15x$.

But what if you multiply? You don't get $3^5 = 5^3$. What you get is $x^5 \cdot x^5 \cdot x^5 = x^3 \cdot x^3 \cdot x^3 \cdot x^3 \cdot x^3$. Both are x^{15}. This is the second law of exponents: $(x^n)^m = (x^m)^n = x^{n \cdot m}$.

Vectors from the Past

When you met vectors in Calculus or Physics they were defined as objects with magnitude and direction (as opposed to a scalar with magnitude alone). They were represented as arrows pointing in the appropriate direction and with length representing the magnitude.

Addition of vectors is described using the 'parallelogram law' which is justified because this is how velocities or forces combine. Think of the Force Table Experiment in Physics.

Multiplying by a positive scalar preserves the direction, changing only the length by a suitable factor. Multiplying by (-1) reverses the direction, preserving the length.

Computation using the geometric definitions can be tedious, but when coordinates are introduced, the computations become easy. For addition we just add the corresponding coordinates and for multiplication by a scalar c we just multiply each coordinate by c.

This pairing of geometric meaning with coordinate computation is crucial for the initial applications of vectors.

In the case of the dot product, the original definition is given by coordinates, but the applications again depend on the geometric meaning: $\mathbf{a} \cdot \mathbf{b} \;=\; |\mathbf{a}||\mathbf{b}| \cos(\theta)$ where θ is the angle between the two vectors.

We will move into the theory of Linear Algebra in two steps. First, we will focus on just the coordinate approach, looking at different arrays of real numbers. Later we will redefine what a vector is when we consider abstract vector spaces.

To illustrate this move, we will extend the definition of dot product between vectors in $\mathbb{R}^3$ to vectors in $\mathbb{R}^n$ where n can be any positive integer (think $n = 17$). So $\mathbf{a} \in \mathbb{R}^n$ is a list of length n consisting of real numbers. We write

$$\mathbf{a} \;=\; (a_1, a_2, \ldots, a_n).$$

We define the *dot product*

$$\mathbf{a} \cdot \mathbf{b} \;=\; a_1 b_1 + a_2 b_2 + \cdots + a_n b_n.$$

That is, just as in the $\mathbb{R}^3$ case we multiply the corresponding coordinates and then sum the products.

Notice that $\mathbf{a} \cdot \mathbf{a} = a_1^2 + a_2^2 + \cdots + a_n^2$ which is positive unless $\mathbf{a}$ is the zero vector

From the commutative law of multiplication for real numbers we get $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$. That is, we can reverse the order in the dot product.

If we multiply $\mathbf{b}$ by a scalar c, we can pull c out of the dot product. That is, $\mathbf{a} \cdot (c\mathbf{b}) = c(\mathbf{a} \cdot \mathbf{b})$ because the factor of c which occurs in each term of the sum can be "factored" out of the sum. Notice that the notation hides the occurrence of two different kinds of multiplication: In $c\mathbf{b}$ the scalar c multiplies times the vector $\mathbf{b}$, while $c(\mathbf{a} \cdot \mathbf{b})$ is the product of two real numbers.

The important property $\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c}$ is shown by using the following array:

$$
\begin{array}{ccccc|cc}
a_1 b_1 & + & a_2 b_2 & + & \ldots & = & \mathbf{a} \cdot \mathbf{b} \\
a_1 c_1 & + & a_2 c_2 & + & \ldots & = & \mathbf{a} \cdot \mathbf{c} \\
\hline\hline
a_1(b_1 + c_1) & + & a_2(b_2 + c_2) & + & \ldots & = & \mathbf{a} \cdot (\mathbf{b} + \mathbf{c})
\end{array}
$$

Notice this is an example of adding different ways in the RECTANGLE PICTURE.

Matrices: Linear Operations

In addition to horizontal (and vertical) lists of numbers, we will use rectangular arrays of numbers. Such a rectangular arrangement is called a *matrix* (plural *matrices*). These are initially motivated by their application to systems of linear equations.

$$\begin{array}{rcrcrcr} 3x & & & - & 2z & = & 7 \\ x & - & y & + & z & = & 5 \\ & & y & - & z & = & -2 \end{array}$$

In order to describe the system all that we need is the location and value of each of the various coefficients,

$$\begin{pmatrix} 3 & 0 & -2 \\ 1 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix}$$

is called the *coefficient matrix* of the system. When we adjoin the numbers on the other side of the equations, we get the *augmented matrix* of the system.

$$\left(\begin{array}{ccc|c} 3 & 0 & -2 & 7 \\ 1 & -1 & 1 & 5 \\ 0 & 1 & -1 & -2 \end{array}\right)$$

We will first consider matrices as objects in themselves and a bit later see how they are used to solve systems of equations.

For positive integers m and n an $m \times n$ matrix A is a rectangular array of numbers with m rows and n columns.

We use the notation $A = [a_{ij}]$ with the row index i varying from 1 to m and the column index j varying from 1 to n.

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

We will sometimes put in a comma for clarity so that $a_{3,5}$ is the $(3,5)$ entry: the number in the 3^{rd} row and the 5^{th} column.

We call $m \times n$ the *size* or *shape* of an $m \times n$ matrix A.

An important special case is when $m = n$. We call this a *square matrix* of size $n \times n$.

When $m = 1$ so there is only one row we call A a *row vector* of *length* n and when $n = 1$ we call it a *column vector* of *length* m.

When we look at an $m \times n$ matrix we can see it as a stack of m row vectors each of length n, or, alternatively, as a list of n column vectors each of length m. This is just the RECTANGLE PICTURE described above.

The linear operations of addition and multiplication by scalars works just as it did for vectors in $\mathbb{R}^3$.

We can only add two matrices of the same shape and when A and B are both $m \times n$ matrices, we add by summing the corresponding entries for each position. That is, if $A = [a_{ij}]$ and $B = [b_{ij}]$, then $A + B = [a_{ij} + b_{ij}]$.

If k is a scalar, i.e. a real number, (or sometimes a complex number), then for kA we just multiply each entry by k. That is, if $A = [a_{ij}]$, then $kA = [ka_{ij}]$.

We then have that $A + B$ and kA also have the same shape, namely $m \times n$.

For each shape $m \times n$, there is an $m \times n$ *zero matrix* 0 with all entries equal to 0. It acts like the number zero does under addition. That is, $A + 0 = 0 + A = A$. Furthermore, the zero matrix equals $0A$ for any $m \times n$ matrix A.

The matrix $-A = (-1)A$ *cancels* A. That is, $A + (-A) = (-A) + A = 0$ and we write $A - B$ for $A + (-B)$.

The obvious rules for this matrix arithmetic are given as Theorem 2.1.1 on page 40 of WKN.

$$\begin{aligned} A+B &= B+A, & A+(B+C) &= (A+B)+C, \\ k(pA) &= (kp)A & 1A &= A \\ k(A+B) &= kA+kB & (k+p)A &= kA+pA. \end{aligned} \tag{1.1}$$

Each of these is checked by looking at each entry separately. We apply these rules without thinking about them just as we do with arithmetic of numbers.

If $A_1, \ldots, A_p$ is a list of $m \times n$ matrices and $k_1, \ldots, k_p$ is a list of scalars, then $k_1 A_1 \ + \ k_2 A_2 \ + \ \ldots \ + \ k_p A_p$ is the *linear combination* of $A_1, \ldots, A_p$ with coefficients $k_1, \ldots, k_p$. Here, just as with numbers, we are using the order of operations rule which says do the scalar multiplications first and then add up.

The transpose of the $m \times n$ matrix A is the $n \times m$ matrix A^T obtained by interchanging the rows and columns, so that if $A = [a_{ij}]$ then $A^T = [a_{ji}]$. So $(A^T)^T = A$. Again by looking at each entry separately we see that

$$(A + B)^T = A^T + B^T \qquad (kA)^T = k(A^T). \tag{1.2}$$

As the book points out we can think of obtaining A^T by flipping A across the *main diagonal*, the entries $a_{11}, a_{22}, \ldots$.

A and A^T have the same shape only when $m = n$ and so when A is a square matrix.

A square matrix A is called *symmetric* when $A = A^T$. For any square matrix A, the sum $A + A^T$ is symmetric (prove it!)

An important example of a symmetric is a *diagonal matrix* with zero entries off the main diagonal.

Let us look at the Exercises: 2.1/2bfg, 3bdfh, 4b, 11, 12, 13, 14d, pages 44-45.

Sigma Notation for Sums

We pause to review the Sigma notation for sums which you have seen for Riemann sums and for infinite series.

If $\mathbf{a} = (a_i) \in \mathbb{R}^n$, then $\sum_{i=1}^{n} a_i$ means $a_1 + a_2 + \ldots a_n$.

The distributive law, clearing parentheses, implies

$$k \cdot \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} k \cdot a_i. \tag{1.3}$$

Also,

$$\sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i = \sum_{i=1}^{n} (a_i + b_i) \tag{1.4}$$

when $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$.

This is really another example of the RECTANGLE PICTURE:

$$\begin{array}{cccccc|cc} a_1 & + & a_2 & + & \dots & a_n & = & \sum_i a_i \\ b_1 & + & b_2 & + & \dots & b_n & = & \sum_i b_i \\ \hline\hline (a_1+b_1) & + & (a_2+b_2) & + & \dots & (a_n+b_n) & = & \sum_i (a_i+b_i) \end{array}$$

The dot product can be written as $\mathbf{a} \cdot \mathbf{b} \ = \ \sum_{i=1}^{n} \ a_i \cdot b_i$.

Now suppose that $A = (a_{ij})$ is an $m \times n$ matrix. We can write $S = \sum_{i,j} a_{ij}$ for the sum of all of the entries.

But now we are going to use the RECTANGLE PICTURE again.

If we fix the row index i, then $\sum_{j=1}^{n} a_{ij}$ just adds up row i. If we do this for each i and then add up the row sums, we get S.

Instead we can compute each column sum $\sum_{i=1}^{m} a_{ij}$ and then add up the column sums to get S. That is,

$$\sum_i \Big(\sum_j a_{ij}\Big) = \sum_{i,j} a_{ij} = \sum_j \Big(\sum_i a_{ij}\Big). \qquad (1.5)$$

You have seen this in a more advanced context.

In calculus 3, to compute the integral of $f(x, y)$ over a rectangle $[a, b] \times [c, d]$, you can use either of the iterated integrals:

$$\int_{x=a}^{x=b} \left(\int_{y=c}^{y=d} f(x, y)\, dy \right) dx \quad \text{or} \quad \int_{y=c}^{y=d} \left(\int_{x=a}^{x=b} f(x, y)\, dx \right) dy.$$

Matrix Times Vector

Unlike addition and scalar multiplication of matrices, matrix multiplication is not at all intuitive. In particular, it is not what you would first think of. You do not take two $m \times n$ matrices and multiply the corresponding entries. Just as with fraction addition where you don't just add the tops and add the bottoms, there is a more complicated procedure which is justified by its usefulness.

We follow the book by first looking at multiplication of a matrix times a column vector.

First, remember that when we write $A = [a_{ij}]$ for an $m \times n$ matrix, the i is the row index, varying from 1 to m as we move down the stack of rows and j is the column index, varying from 1 to n as we move right along the list of columns.

The number labeled a_{ij} is the (i, j) entry, the number in the i^{th} row and the j^{th} column.

We multiply the $m \times n$ matrix $A = [a_{ij}]$ times the column vector $X = [x_j]$ of length n. A column vector of length n is the same thing as an $n \times 1$ matrix. We omit the column index, writing x_j for x_{j1} since there is only one column.

When we multiply A times X we obtain a column vector B of length m. So $B = [b_i]$. The *Multiplication Formula* is given by:

$$AX = B, \quad \text{with} \quad b_i = \sum_{j=1}^{n} a_{ij}x_j = a_{i1}x_1 + a_{i2}x_2 \cdots + a_{in}x_n. \tag{1.6}$$

However, it is the picture rather than the formula which reveals what is going on and why this weird definition is important.

There are two pictorial representations of the same thing.

Think of X as a column of n unknowns. Row i of A has length n. We take the dot product with X and set it equal b_i. This gives us a system of m linear equations in n unknowns with coefficient matrix A and augmented matrix $A|B$. Thus, we can write the system of m equations using a single matrix equation $AX \ = \ B$. See illustration Matrix Multiplication 1a.

Instead, think of X as a column of n coefficients, and regard A as a list of n columns $A = [C_1 \ C_2 \ldots \ C_n]$ each of length m. Then $B = AX$ writes the column B as a linear combination of the columns of A.

$$x_1 C_1 + x_2 C_2 + \cdots + x_n C_n \ = \ B.$$

See illustration Matrix Multiplication 1b.

From the Multiplication Formula or from the properties of the dot product we obtain the properties of multiplication for a matrix times a vector. This is Theorem 2.2.2 of the book. Assume that A, B are $m \times n$ matrices, X, Y are $n \times 1$ column vectors and k is a scalar.

$$
\begin{gathered}
(kA)X \;=\; A(kX) \;=\; k(AX). \\
A(X+Y) \;=\; AX \;+\; AY, \qquad (A+B)X \;=\; AX \;+\; BX.
\end{gathered} \tag{1.7}
$$

These follow because, from Equation (1.3)

$$\sum_j (ka_{ij})x_j \;=\; \sum_j a_{ij}(kx_j) \;=\; k\sum_j a_{ij}x_j.$$

While (1.4) implies

$$\sum_j a_{ij}(x_j + y_j) \;=\; \sum_j (a_{ij}x_j + a_{ij}y_j) \;=\; (\sum_j a_{ij}x_j) + (\sum_j a_{ij}y_j).$$

and also

$$\sum_j (a_{ij} + b_{ij})x_j \;=\; \sum_j (a_{ij}x_j + b_{ij}x_j) \;=\; (\sum_j a_{ij}x_j) + (\sum_j b_{ij}x_j).$$

Multiplication of Matrices

If $A = [a_{ij}]$ is an $m \times n$ matrix and $B = [b_{jk}]$ is an $n \times p$ matrix, then we the product $AB = C = [c_{ik}]$ is an $m \times p$ matrix. Notice we use the same index j as the column index for A and the row index for B because in both cases j ranges from 1 to n.

It is given by the Multiplication Formula

$$AB = C, \qquad \text{with} \quad c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}. \tag{1.8}$$

Again it is the picture which reveals the meaning of the formula.

We cut A into rows. There are m rows, each of length n. We cut B into columns. There are p columns, each of length n.

The ik entry, c_{ik} of $C = A \cdot B$ is the dot product of Row_i of A with Col_k of B. This makes sense because each of these has length n. See illustration Matrix Multiplication 2.

In actually doing matrix multiplication, write B above and to the right of A so that the product AB is below B and to the right of A.

Of the properties for multiplication, most important is the Commutative Law because it is usually not true.

If A is $m \times n$, and B is $n \times p$, then AB is $m \times p$. If $m \neq p$, then BA is not even defined.

If $m = p$, then AB is $m \times m$ and BA is $n \times n$. Both are square matrices but if $n \neq m$, they are of different size and so certainly not equal.

Finally, if $n = m = p$, then A, B, AB and BA are all square matrices of the same size but still it is usually not true that $AB = BA$.

For example, if A has a row of zeroes, then the corresponding row in the product AB consists of zeroes.

Similarly, if B has a column of zeroes, the corresponding column in the product AB consists of zeroes.

To summarize, in matrix multiplication the order of the factors is very important and $AB \; = \; BA$ is usually NOT TRUE. For example, with X an $n \times 1$ column vector, AX is an $m \times 1$ column vector, but XA is undefined. If Y is a $1 \times m$ row vector, then YA is a $1 \times n$ row vector.

We can use second picture for matrix multiplication, but it is usually less important when B is not a single column vector. See illustration Matrix Multiplication 2b.

For the Associative Law for matrix multiplication we will use the sum formula and the RECTANGLE PICTURE.

Start with matrices A, B and C. Suppose $A = (a_{ij})$ is $m \times n$, $B = (b_{jk})$ is $n \times p$ and $C = (c_{k\ell})$ is $p \times q$.

AB is $m \times p$ with $(AB)_{ik} \;=\; \sum_j \; a_{ij} b_{jk}$.

BC is $n \times q$ with $(BC)_{j\ell} \;=\; \sum_k \; b_{jk} c_{k\ell}$.

$(AB)C$ and $A(BC)$ are both $m \times q$ and we have:

$$((AB)C)_{i\ell} = \sum_k (AB)_{ik} c_{k\ell} =$$

$$\sum_k (\sum_j a_{ij} b_{jk}) c_{k\ell} = \sum_k (\sum_j a_{ij} b_{jk} c_{k\ell}).$$

$$(A(BC)_{i\ell} = \sum_j a_{ij} (BC)_{j\ell} =$$

$$\sum_j a_{ij} (\sum_k b_{jk} c_{k\ell}) = \sum_j (\sum_k a_{ij} b_{jk} c_{k\ell}).$$

The end results are equal from the RECTANGLE PICTURE or, to be precise, from equation (1.5). So

$$(AB)C = A(BC).$$

The other properties of multiplication follow directly from the properties of multiplying a matrix times a column vector.

$$(kA)B \;=\; A(kB) \;=\; k(AB).$$
$$A(B+C) \;=\; AB \;+\; AC, \qquad (A+D)B \;=\; AB \;+\; DB. \tag{1.9}$$

Here A and D are $m \times n$ while B and C are $n \times p$.

Also we have

$$(AB)^T \;=\; B^T A^T. \tag{1.10}$$

Notice the reverse of order. B^T is $p \times n$ and A^T is $n \times m$.

Let us look at Exercises 2.3/1ghi, 5ac, 6c, 7c, pages 76-77.

Block Multiplication

With A $m \times n$ and B $n \times p$ matrices so that AB is an $m \times p$ matrix, partitioning the matrices in various ways can lead to associated multiplication formulae.

$$\left(\frac{A_1}{A_2}\right)\left(B_1 \quad | \quad B_2\right) = \begin{pmatrix} A_1B_1 & A_1B_2 \\ A_2B_1 & A_21B_2 \end{pmatrix}.$$

Here A_1 is $u \times n$,A_2 is $(m-u) \times n$, B_1 is $n \times v$, and B_2 is $n \times (p-v)$ for some u, v.

$$\left(A_1 \quad | \quad A_2\right)\left(\frac{B_1}{B_2}\right) = \left(A_1B_1 + A_2B_2\right).$$

Here A_1 is $m \times u$, A_2 is $m \times (n-u)$, B_1 is $u \times p$ and B_2 is $(n-u) \times p$. Notice that this time the partition of the columns of A must use the same size divisions as does the partition of the rows of B.

As is indicated in the book, the most important application of Block Multiplication concerns $n \times n$ square matrices A and B in block triangular form with A_1 and B_1 square matrices of the same size:

$$\begin{pmatrix} A_1 & X \\ 0 & A_2 \end{pmatrix} \begin{pmatrix} B_1 & Y \\ 0 & B_2 \end{pmatrix} = \begin{pmatrix} A_1 B_1 & A_1 Y + X B_2 \\ 0 & A_2 B_2 \end{pmatrix}.$$

Identity and Inverse Matrices

We have seen that for each shape $m \times n$ there is a zero matrix 0_{mn}, with every entry equal to 0. It acts as the number zero does with respect to addition. That is,
$A + 0_{mn} = 0_{mn} + A = A$ for any $m \times n$ matrix A.

There are analogues for the number 1 as well. For every n there is an $n \times n$ square matrix I_n. However, not all the entries are 1. Instead, the main diagonal entries, the $(1, 1), (2, 2), \ldots$ entries equal 1 and the off-diagonal entries are 0. So, for example,

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

If A is an $m \times n$ matrix then $I_m A = A = A I_n$. This is easily checked by looking at the pictures for multiplication.

Alternatively, we can use the multiplication formula. The *ij* entry of the identity matrix I is the so-called *Kronecker delta*

$$\delta_{ij} \;=\; \begin{cases} 1 \text{ if } i = j, \\ 0 \text{ if } i \neq j. \end{cases}$$

So $(AI)_{ik} \;=\; \sum_j \; a_{ij}\delta_{jk}$. But $\delta_{jk} = 0$ except when $j = k$ and $\delta_{kk} = 1$ So $(AI)_{ik} \;=\; a_{ik}$. That is, $AI = A$.

Similarly, $IA = A$.

We have seen that any $m \times n$ matrix A has an additive inverse $-A$ which cancels it under addition. That is, $A + (-A) = 0$.

Only a nonzero number has a reciprocal which cancels it under multiplication ("You can't divide by zero."). In the same way an $n \times n$ matrix A may have an inverse A^{-1} which cancels it under multiplication, but not all do.

The inverse A^{-1} of an $n \times n$ matrix A is the matrix which satisfies $AA^{-1} = A^{-1}A = I_n$. Since A cancels A^{-1} we have $(A^{-1})^{-1} \;=\; A$.

We speak of <u>the</u> inverse because A has at most one inverse. If Q is another matrix which cancels A, then

$$Q = QI_n = Q(AA^{-1}) = (QA)A^{-1} = I_nA^{-1} = A^{-1}.$$

As you would expect, the square zero matrices do not have inverses, but many nonzero square matrices are also *singular* (meaning they don't have an inverse).

If A has a row of zeroes, then it cannot happen that $AB = I$ for any matrix B (Why not?). Similarly if A has a column of zeroes then it cannot happen that $BA = I$.

Thus, if A has a row or column of zeroes then it does not have an inverse.

In the positive direction we have $(AB)^{-1} = B^{-1}A^{-1}$.

This simple looking equation requires some interpretation. It says: If $n \times n$ matrices A and B each have inverses, then the product AB has an inverse and its inverse is the product of the inverse of A and B in the reversed order.

We prove this by showing that $B^{-1}A^{-1}$ does what the inverse of AB is supposed to do, namely to cancel AB:

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$$

and similarly in the other direction. Notice it won't work if we used $A^{-1}B^{-1}$ instead. The product $ABA^{-1}B^{-1}$ is often not equal to I because you can't usually move the B past the A^{-1} to cancel things out. In general, for invertible $n \times n$ matrices $A_1, A_2, \ldots, A_k$ we have

$$(A_1A_2 \ldots A_k)^{-1} = A_k^{-1}A_{k-1}^{-1} \ldots A_1^{-1}. \tag{1.11}$$

Suppose that the $n \times n$ matrix A is in block triangular form with $A = \begin{pmatrix} A_1 & X \\ 0 & A_2 \end{pmatrix}$ with A_1 $u \times u$, A_2 $(n-u) \times (n-u)$ and X $u \times (n-u)$. If A_1 and A_2 are invertible, then A is invertible with

$$A^{-1} = \begin{pmatrix} A_1^{-1} & -A_1^{-1} X A_2^{-1} \\ 0 & A_2^{-1} \end{pmatrix}$$

which you check by Block Multiplication. Conversely, as in Example 2.4.11, if $P = \begin{pmatrix} C & V \\ W & D \end{pmatrix}$, then

$$I_n = A\ P = \begin{pmatrix} A_1 C + XW & A_1 V + XD \\ A_2 W & A_2 D \end{pmatrix}$$

So $A_2 W = 0, A_1 C + XW = I_u, A_2 D = I_{n-u}$. We use a fact proved later that we need only check cancellation on one side. It follows that A_2 is invertible and so $W = A_2^{-1} A_2 W = 0$. Hence, $A_1 C = I_u$ and so A_1 is invertible as well.

In the 2×2 case there is a simple way of getting the inverse if it exists. For a 2×2 matrix A define the *adjugate* by :

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \Longrightarrow \quad A' = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{1.12}$$

That is, interchange the diagonal elements and reverse the sign of the off-diagonal elements.

Let $det = ad - bc$, the determinant of A. Check that $AA' = A'A = detI$. So if $det \neq 0$, then $A^{-1} = \frac{1}{det}A'$.

If $det = 0$, then $AA' = 0$ and so if A were invertible, $A' = A^{-1}AA' = 0$ and so all the entries of A would be zero. Thus, A is not in fact invertible.

All of this will generalize, but in a not so simple way, to the $n \times n$ case.

Let us look at Exercises 2.4/2b, 26d, pages 90, 93.

Echelon Form

Remember that to describe a system of linear equations, we use the *augmented matrix* $(A|B)$ where A is the *coefficient matrix* and B is the column vector of constants on the right side of the equal signs.

In order to solve a system of equations we will manipulate the augmented matrix to get an augmented matrix of a system which is equivalent - that is, with the same solutions - but which is easy to solve.

We first describe the form which is our goal. Then we will describe how to get from the original system to one with the special form.

A matrix is in *Echelon Form* when it satisfies the following conditions:

(i) Any zero rows are at the bottom.

(ii) The first nonzero entry from the left in any nonzero row is a 1, called a *leading* 1.

(iii) Each leading 1 is to the right of the leading 1's in the rows above it.

The matrix is in Reduced Echelon Form when it satisfies, in addition:

(iv) Each leading 1 is the only nonzero entry in its column.

Now suppose the augmented matrix is in Echelon Form.

Consistency: If there is a leading 1 in the last, constants, column then the system is *inconsistent*. That is the is no solution. Observe that the equation $0x_1 + 0x_2 + \ldots 0x_n \; = \; 1$ has no solution.

If the system is consistent, then the leading 1's all occur in columns associated with variables. These are called the leading 1 variables or the *pivot variables*. The remaining variables, if any, are called *free variables*.

To solve the system we work up from the bottom. We use each nonzero row to solve for the pivot variable associated with its leading 1. Each free variable is a parameter, an arbitrary constant, which may take on any value.

For example, suppose that the row is $(0, 0, 1, 7, -1, 2| - 3)$. The pivot variable is x_3 and we have $x_3 = -3 - 7x_4 + x_5 - 2x_6$. Because we are working up from the bottom, we have already solved for x_4, x_5 and x_6. Each of these is a free variable parameter, or else a pivot variable for which we have a solution and we substitute it into this equation.

If x_6 and x_4 were free variables, then we would let $x_6 = r$ and $x_4 = s$. If we had found $x_5 = 11 - 2r$, then we substitute to get

$$x_3 = -3 - 7s + (11 - 2r) - 2r = 8 - 7s - 4r.$$

$$\left(\begin{array}{ccccccc|c} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 2 & 1 & -4 & -3 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 7 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

is in Echelon Form with solution, working from the bottom.

$$x_7 = r.$$

$$x_6 = 5 - 2r.$$

$$x_5 = s.$$

$$x_4 = 7 - x_5 - x_6 - x_7 =$$

$$7 - s - (5 - 2r) - r = 2 - s + r.$$

$$x_3 = -3 - 2x_5 - x_6 + 4x_7 =$$

$$-3 - 2s - (5 - 2r) + 4r = -8 - 2s + 8r.$$

$$x_2 = 0.$$

$$x_1 = t.$$

From Reduced Echelon Form, it is easier to solve:

$$\left(\begin{array}{ccccc|c} 1 & 3 & 0 & -2 & 0 & 3 \\ 0 & 0 & 1 & 1 & 0 & -5 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{array}\right)$$

$$\begin{aligned} x_5 &= -1. \\ x_4 &= r. \\ x_3 &= -5 - r. \\ x_2 &= s. \\ x_1 &= 3 - 3s + 2r. \end{aligned}$$

Let us look at Exercises 1.2/3bd, page 17.

We now introduce the operations that we use to transform a matrix into Echelon Form.

The Elementary Row Operations are of three types.

Type I: Interchange two rows:
$R_i \leftrightarrow R_j$.

Type II: Multiply a row by a nonzero scalar:
$R_i \rightarrow kR_i$.

Type III: To a row add a multiple of a copy of another row:
$R_i \rightarrow R_i + kR_j$.

Each operation preserves the shape of the matrix and, when applied to an augmented matrix, each yields an augmented matrix for a system with the same solutions.

Notice that each type is reversible by an operation of the same type.

Type I: $R_i \leftrightarrow R_j$ is its own inverse.

Type II: $R_i \rightarrow kR_i$ is reversed by $R_i \rightarrow (1/k)R_i$ (which is why we required $k \neq 0$).

Type III: $R_i \rightarrow R_i + kR_j$ is reversed by $R_i \rightarrow R_i - kR_j$.

For two $m \times n$ matrices A, B we call B row equivalent to A if we can get from A to B by a sequence of elementary row operations. We write this as $A \sim B$. We allow the sequence to be empty (no operations at all) which shows that $A \sim A$ for any matrix A. It is clear that $A \sim B$ and $B \sim C$ implies $A \sim C$. Finally, because the operations are reversible, we see that $A \sim B$ implies $B \sim A$. This says that row equivalence is an *equivalence relation*.

We can transform any matrix A to a row equivalent matrix in Echelon Form and from there to Reduced Echelon Form by using the *Gaussian Algorithm*.

Step 1: If the matrix consists entirely of zeroes, stop. The matrix is in Row Echelon form.

Step 2: Find the first column from the left with a nonzero entry k. If it is not in the first row use a Type I operation to move it to the top.

Step 3: Use the Type II operation $R_1 \to (1/k)R_1$ to convert the k to a leading 1.

Step 4: Use Type III operations $R_i \to R_i - a_i R_1$ to obtain 0's in each row below the leading 1. Here a_i is the entry of R_i in the column of the leading one.

Step 5: Repeat Steps 1 - 4 on the matrix consisting of the rows below R_1.

The process stops at a Step 1 or after a Step 4 when there are no more rows. The matrix is now in Echelon Form.

Notice that we are moving down and to the right as we transform to Echelon Form.

Once the matrix is in Echelon Form we may further transform to get Reduced Echelon Form. Notice that already below each leading 1 there are only 0's. Now we move up and to the left to eliminate the nonzero entries above the leading 1's

Step 6: For the leading 1 in R_i, use Type III operations $R_j \to R_j - a_j R_i$ to obtain 0's above the leading 1 in R_i. Here a_j is the entry of R_j in the column of the R_i leading 1.

We repeat, moving upwards until above every leading 1 there are only 0's. The matrix is now in Reduced Echelon Form. Notice that going to Reduced Echelon Form does not change the number or position of the leading 1's.

While there may be several matrices in Echelon Form which are row equivalent to A, there is only one in Reduced Echelon Form. This is Theorem 2.5.4 proved in the book on page 99, but we will omit the proof.

We are not restricted to the steps given in the algorithm. Often we obtain the leading 1's in other ways to avoid fractions. For example, consider

$$A = \begin{pmatrix} 2 & 3 \\ 1 & 5 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix} \quad C = \begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix}$$

For A we use $R_1 \leftrightarrow R_2$ to get the leading 1. For B we would use $R_2 \to R_2 - R_1$ to get a leading 1 in the second row which we would then switch to the top using $R_1 \leftrightarrow R_2$. For C we can use $R_2 \to 2R_2$ and then $R_2 \to R_2 - R_1$ to get a leading 1 in the second row which again we would switch to the top using Type I operation.

Rank of a Matrix

The *rank* of a matrix A is the number of leading 1's in a matrix in Echelon Form which is row equivalent to A. That is, put A in Echelon Form and count the leading ones. Notice that for an $m \times n$ matrix the rank r is at most the minimum of m and n because there is at most one leading 1 in any row or column.

An augmented matrix represents an inconsistent system when there is a leading 1 in its last column when it is put in Echelon Form. Thus, the system is consistent when the rank of the augmented matrix equals the rank of the coefficient matrix.

For a consistent system with coefficient matrix $m \times n$ and rank r, the number of equations is m and the number of variables is n. So the number of pivot variables is $r \leq n$ and the number of free variables, i.e. the number of parameters, is $n - r$.

Now assume that A is an $m \times m$ matrix which is row equivalent to Q in Reduced Echelon Form. The following is the first version of the so-called *Fredholm Alternative*.

Fredholm Alternative (version 1): For A an $m \times m$ matrix row equivalent to Q in Reduced Echelon Form, Either:

(i) $r = RankA \; = \; m$ and so $Q = I_m$ because there is a leading 1 in every row and every column.

or

(ii) $r = RankA \; < \; m$ and so Q has a row of zeroes. In fact, it has $m - r$ rows of zeroes.

Homogeneous Systems

A system of equations 0 solutions when the system is inconsistent, 1 solution when the system is consistent and $n = r$ and infinitely many solutions when the system is consistent and $n > r$.

A system is called *homogeneous* when the column of constants is 0. In matrix form a homogeneous system is $AX = 0$.

If X_1, X_2 are solutions of $AX = B$, then the difference $Z = X_1 - X_2$ is a solution of the homogeneous system. If X_p is a particular solution of the system $AX = B$, then all solutions are of the form $X_p + Z$ where Z is a solution of the homogeneous system.

A homogeneous system always has the *trivial solution* $Z = 0$. This is the only solution when $n = r$ and there are infinitely many solutions when $n > r$.

Elementary Matrices

The fundamental property of row operations is the following. The every row operation, labeled *RowOP* we have

$$RowOp(AB) \;=\; RowOp(A)B. \tag{2.1}$$

For any row operation, the associated elementary $m \times m$ matrix is defined to be $E = RowOP(I_m)$. That is, apply the row operation to the identity matrix.

For any $m \times n$ matrix A we have:

$$RowOp(A) \;=\; RowOp(I_m A) \;=\; RowOp(I_m)A \;=\; E\,A. \tag{2.2}$$

That is, doing a row operation to A is the same as multiplying on the left by the associated elementary matrix.

Recall that every row operation $RowOP$ has a reverse row operation which we will label $\overline{RowOP}$. Let E and $\overline{E}$ be the associated elementary matrices.

$$I_m \;=\; \overline{RowOP}(RowOp(I_m)) \;=\; \overline{RowOP}(E) \;=\; \overline{E}\; E$$

and similarly $I_m \;=\; E\;\overline{E}$. That is $\overline{E}$ cancels E and so $\overline{E} = E^{-1}$.

Thus, every elementary matrix is invertible and its inverse is the elementary matrix associated with the reverse row operation.

If A is an $m \times n$ matrix and Q is the matrix in Reduced Echelon Form which is row equivalent to A, then there is a sequence of row operations which takes us from A to Q.

$$\begin{aligned} Q &= RowOp_k(RowOp_{k-1}(\dots RowOp_2(RowOp_1(A))\dots)) \\ &= E_k \ E_{k-1} \dots E_2 \ E_1 \ A. \end{aligned} \tag{2.3}$$

where E_1, E_2 etc are the elementary matrices associated with $RowOp_1, RowOp_2, \dots$.

Because the product of invertible matrices is invertible, $U = E_k \ E_{k-1} \dots E_2 \ E_1$ is invertible with inverse $U^{-1} = E_1^{-1} \ E_2^{-1} \dots E_{k-1}^{-1} \ E_k^{-1}$. and we have

$$Q = U \ A, \qquad U^{-1} \ Q = A. \tag{2.4}$$

Fredholm Alternative (version 2): For A an $m \times m$ matrix row equivalent to Q in Reduced Echelon Form, so that $Q = UA$ and $U^{-1}Q = A$ with U a product of elementary matrices. Either:

(i) $r = RankA \; = \; m$, $Q = I_m$ and so $U^{-1} = A$ is a product of elementary matrices and so is invertible with inverse U. For each $m \times 1$ column vector B the system $AX = B$ has the unique solution $X = A^{-1}B$, because $AX = B$ implies $X = A^{-1}AX = A^{-1}B$ and, conversely, $X = A^{-1}B$ implies $AX = AA^{-1}B = B$. For B any $m \times p$ matrix, $AB = 0$ implies $B = A^{-1}0 = 0$.

or else

(ii) $r = RankA < m$, Q has $m - r$ rows of zeroes, and so A is not invertible since $Q = UA$ is not invertible. There exist $m \times 1$ column vectors B such that the system $AX = B$ is inconsistent and so has no solution. If $\mathbf{e}_m = \begin{pmatrix} 0 \\ 0 \\ \dots \\ 1 \end{pmatrix}$ then $\left(Q \mid \mathbf{e}_m\right)$ is the augmented matrix of an inconsistent system and so $\left(A \mid U^{-1}\mathbf{e}_m\right)$ is the augmented matrix of an inconsistent system.

On the other hand, if $AX = B$ is consistent, for example if $B = 0$ and the system is homogeneous, then there are infinitely many solutions with $m - r$ parameters. If X is a nontrivial solution of the homogeneous system $AX = 0$ and B consists of m copies of the column X, then B is a nonzero $m \times m$ matrix such that $AB = 0$.

We have seen that the product of invertible matrices is invertible. Conversely, if a product of square matrices is invertible then each factor is invertible.

Theorem 2.01: If A, B are $m \times m$ matrices, then $AB = C$ is invertible if and only if both A and B are invertible.

Proof: We have already seen that if A and B are invertible, then $(AB)^{-1} = B^{-1}A^{-1}$.

Now suppose that $C = AB$ is invertible. From what we have just seen, it follows that UC is invertible for any invertible matrix U. If A is not invertible, then there exists U a product of elementary matrices, and so invertible, such that UA has a row of zeroes. Then $UC = UAB$ has a row of zeroes, but this is impossible because UC is invertible.

It follows that A is invertible. Hence, $B = A^{-1}AB = A^{-1}C$ is invertible as well.

□

Thus, we have (technically by induction)for $m \times m$ matrices $A_1, A_2, \ldots, A_k$

$$\text{The product } A_1 A_2 \ldots A_k \text{ is invertible} \iff A_1, A_2, \ldots, A_k \text{ are all invertible.}$$

We used earlier the fact that for square matrices it suffices to check cancelation on one side.

Corollary 2.02: If A, B are $m \times m$ matrices with $AB = I_m$, then A and B are invertible with $A^{-1} = B$.

Proof: From the Theorem it follows that A and B are invertible because I_m is invertible.

It follows that $A^{-1} = A^{-1} I_m = A^{-1} AB = B$.
□

A square matrix is invertible if and only if its transpose is invertible. In fact,

$$(A^T)^{-1} = (A^{-1})^T. \tag{2.5}$$

We have to show that $(A^{-1})^T$ cancels A^T. Remember that $(AB)^T = B^T A^T$. Therefore

$$I_m \;=\; I_m^T \;=\; (A^{-1}A)^T \;=\; A^T(A^{-1})^T.$$

For an $m \times m$ matrix A, we can use the Gaussian Algorithm to compute the inverse of A if it exists.

Begin with the $m \times 2m$ matrix $(A \mid I_m)$ with A on the left and the identity on the right. We do row operations to put this matrix in Reduced Echelon Form.

Case 1: If A is invertible, then when we reach Reduced Echelon Form we have $Q = I_m = UA$ on the left side. We are doing the same row operations on the right side and so we end with $UI_m = U$ on the right. That is, when we reach the identity on the left, we have the inverse of A on the right.

Case 2: If A is not invertible, then at some point we see a row of zeroes on the left side. We may then stop, recognizing that A is row equivalent to a matrix with a row of zeroes and so we see that A is not invertible.

Let us look at Exercises 2.4/2k, 3b, page 91.

Linear Transformations

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is a function which satisfies:

T1: $T(X + Y) = T(X) + T(Y)$. That is, you get the same result whether you first add and then apply T or if you apply T to each term and then add.

T2: $T(cX) = cT(X)$. That is, you can 'pull out' constants.

It follows that applying T commutes with forming linear combinations:

$$T(c_1X_1+c_2X_2+\cdots+c_kX_k) = c_1T(X_1)+c_2T(X_2)+\cdots+c_kT(X_k). \tag{2.6}$$

If A is an $m \times n$ matrix, then $T_A : \mathbb{R}^n \to \mathbb{R}^m$ is defined by $T_A(X) = AX$. The properties of matrix multiplication imply that T_A is linear. In fact, every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is of this form.

To see this we define the *standard basis* of $\mathbb{R}^n$ to be the list $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ of columns of the identity matrix I_n. Later we will discuss the idea of a basis in general. For $X \in \mathbb{R}^n$ we have:

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ \cdot \\ \cdot \\ 0 \end{pmatrix} + \cdots + x_n \begin{pmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ 1 \end{pmatrix}.$$

That is, $X \;=\; x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n$.

Applying the linear map T we have
$T(X) \;=\; x_1 T(\mathbf{e}_1) + x_2 T(\mathbf{e}_2) + \cdots + x_n T(\mathbf{e}_n)$.

This means that if $A = [T(\mathbf{e}_1)\, T(\mathbf{e}_2) \ldots T(\mathbf{e}_n)]$, the matrix with columns $T(\mathbf{e}_1), T(\mathbf{e}_2), \ldots T(\mathbf{e}_n)$ then
$T(X) = AX = T_A(X)$.

Notice that each column $T(\mathbf{e}_i)$ lies in $\mathbb{R}^m$ and so A is an $m \times n$ matrix.

If $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}^p \to \mathbb{R}^n$ are linear maps, then the *composition*$T \circ S : \mathbb{R}^p \to \mathbb{R}^m$ is defined by $T \circ S(X) = T(S(X))$. That is, $T \circ S$ of X is T of S of X. This is the notion of composed map to which the Chain Rule is applied in Calculus.

The composition of linear maps is linear. This is easy to check directly, but it follows from the fact that composition corresponds to matrix multiplication.

For A an $m \times n$ and B an $n \times p$ matrix

$$T_A \circ T_B \;=\; T_{AB}. \tag{2.7}$$

Because

$$T_A \circ T_B(X) \;=\; A(BX) \;=\; (AB)X \;=\; T_{AB}(X).$$

For example, the counterclockwise rotation R_θ through the angle θ sends $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ to $\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$ and sends $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ to $\begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$. Thus,

$$R_\theta \;=\; T_A \quad \text{with} \quad A \;=\; \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

As shown in Example 2.5.7 on page 111, we can obtain the addition formulae for sine and cosine by observing $R_\theta \circ R_\phi \;=\; R_{\theta+\phi}$ and multiplying the corresponding matrices.

For I_n we have $T_{I_n} = I : \mathbb{R}^n \to \mathbb{R}^n$, the identity map with $I(X) \ = \ X$ for all X in $\mathbb{R}^n$.

It then follows that if A is an invertible $n \times n$ matrix, then $T_{A^{-1}} : \mathbb{R}^n \to \mathbb{R}^n$ is the inverse function for T_A. That is, $T_{A^{-1}} \circ T_A = I = T_A \circ T_{A^{-1}}$. So

$$(T_A)^{-1} \ = \ T_{A^{-1}}.$$

Determinants

You should already be familiar with 2×2 determinants and even, via the cross product in Calculus, with 3×3 determinants. In general, the determinant is a function which associates to every $n \times n$ square matrix A a number $det(A)$.

Before describing the properties which characterize the determinant, we first mention two important properties which are not true.

$$det(A + B) \;=\; det(A) + det(B) \quad \text{NOT TRUE.}$$

$$det(kA) \;=\; kdet(A) \quad \text{NOT TRUE.}$$

That is, the function *det* is not a linear transformation. Now for the properties which are true.

DET 1: For the $n \times n$ identity matrix, $det(I_n) = 1$.

We will later see that $det(AB) = det(A)det(B)$. With $A = B = I_n$ this implies that $det(I_n) = det(I_n)det(I_n)$.

Now the solutions of the equation $x = xx = x^2$, i.e. $x(x-1) = x^2 - x = 0$ are 0 and 1.

If the determinant of I_n were equal to 0, then for every $n \times n$ matrix A, $A = AI_n$ would imply $det(A) = det(A)det(I_n) = 0$. That is, every matrix would have determinant 0.

We are left with the condition that $det(I_n) = 1$.

DET 2: The determinant is a multi-linear function of the rows.

We mentioned that the determinant is not a linear function. When we looked at the dot product we saw that if you fix one vector, and were allowed to vary the other vector, then the dot product was a linear function of the variable vector. That is, the dot product was linear in each variable separately. We call such a function of two variables *bi-linear*. Similarly, if we fix all but one row of a matrix and allow the remaining row to vary, the determinant is a linear function of the remaining, variable, row.

To see what this means we will use the notation $R_i(A)$ to mean the i^{th} row of the matrix A.

DET 2a: If A and B are $n \times n$ matrices with $R_i(B) = cR_i(A)$ and $R_j(B) = R_j(A)$ for all $j \neq i$, then $det(B) = cdet(A)$.

That is, we can 'pull out' of the determinant a common factor c from any row.

Corollary 3.01: If A has a row of zeroes, then $det(A) = 0$.

Proof: The number 0 is a common factor of a row of zeroes and so we can pull it out of the determinant.
□

Corollary 3.02: $det(cA) \; = \; c^n det(A)$.

Proof: We can pull a common factor of c from each of the n rows. □

DET 2b: If A, B and C are $n \times n$ matrices with $R_i(C) = R_i(A) + R_i(B)$ and $R_j(C) = R_j(B) = R_j(A)$ for all $j \neq i$, then $det(C) = det(A) + det(B)$.

For example, if

$$A = \begin{pmatrix} a_1 & a_2 & a_3 \\ x_1 & x_2 & x_3 \\ b_1 & b_2 & b_3 \end{pmatrix}, B = \begin{pmatrix} a_1 & a_2 & a_3 \\ y_1 & y_2 & y_3 \\ b_1 & b_2 & b_3 \end{pmatrix}$$

and

$$C = \begin{pmatrix} a_1 & a_2 & a_3 \\ x_1 + y_1 & x_2 + y_2 & x_3 + y_3 \\ b_1 & b_2 & b_3 \end{pmatrix}$$

then $det(C) = det(A) + det(B)$.

DET 3: Interchanging two rows multiplies the determinant by (-1). That is, if B is obtained from A by interchanging two rows and leaving the rest fixed, then $det(B) = -det(A)$.

Corollary 3.03: If A has two identical rows, i.e. $R_i(A) = R_j(A)$ for some $i \neq j$, then $det(A) = 0$.

Proof: The matrix B obtained by interchanging the two rows is the same as A. From DET 3, we have $det(B) = -det(A)$. But since $B = A$, $det(B) = det(A)$. That is, $-det(A) = det(A)$. The only number x such that $-x = x$ is $x = 0$.
□

We now show that from DET 2b and this *Identical Row Property*, we can verify DET3.

We will look at several matrices with all the rows except the i and j rows fixed. We will write $\begin{pmatrix} R \\ S \end{pmatrix}$ for the matrix with R in the i^{th} row, S in the j^{th} row.

$$0 \;=\; det \begin{pmatrix} R_i(A) + R_j(A) \\ R_i(A) + R_j(A) \end{pmatrix} \;=$$

$$det \begin{pmatrix} R_i(A) \\ R_i(A) \end{pmatrix} + det \begin{pmatrix} R_j(A) \\ R_j(A) \end{pmatrix} + det \begin{pmatrix} R_i(A) \\ R_j(A) \end{pmatrix} + det \begin{pmatrix} R_j(A) \\ R_i(A) \end{pmatrix}$$

Therefore, $det \begin{pmatrix} R_i(A) \\ R_j(A) \end{pmatrix} \;=\; -det \begin{pmatrix} R_j(A) \\ R_i(A) \end{pmatrix}$.

These three properties characterize the determinant. The remaining properties are derived from these.

DET 4: For each elementary row operation there is a nonzero multiplier μ so that for any $n \times n$ matrix A $det(RowOp(A)) \ = \ \mu det(A)$. In particular, if $E = RowOP(I_n)$ is the associated elementary matrix, then $det(E) = \mu$.

DET 4a: For a Type I row operation $R_i \leftrightarrow R_j$, $\mu = -1$ and so $det(RowOp(A)) \ = \ -det(A)$.

Proof: This is a restatement of DET 3.

DET 4b: For a Type II row operation $R_i \to c\ R_i$ with $c \neq 0$, $\mu = c$ and so $det(RowOp(A)) \ = \ c\ det(A)$.

Proof: This is a restatement of DET 2a.

DET 4c: For a Type III row operation $R_i \to R_i + k\ R_j$, $\mu = 1$ and so $det(RowOp(A)) \ = \ det(A)$.

Proof: Use the notation of the proof of the derivation of DET 3 and apply DET 2a and DET 2b.

$$det(RowOp(A)) \ = \ det \begin{pmatrix} R_i(A) + kR_j(A) \\ R_j(A) \end{pmatrix} \ =$$

$$det \begin{pmatrix} R_i(A) \\ R_j(A) \end{pmatrix} + k\ det \begin{pmatrix} R_j(A) \\ R_j(A) \end{pmatrix}.$$

The second term $k\ det \begin{pmatrix} R_j(A) \\ R_j(A) \end{pmatrix} \ = \ 0$ by the Identical Rows Property. So

$$det(RowOp(A)) \ = \ det \begin{pmatrix} R_i(A) \\ R_j(A) \end{pmatrix} \ = \ det(A).$$

Applying DET 4 repeatedly we get

DET 4d: If A is an $n \times n$ matrix with $B = RowOp_k(RowOp_{k-1}(\dots(RowOp_1(A))\dots))$, then $det(B) = \mu_k \mu_{k-1} \dots \mu_1 det(A)$ where $\mu_1, \mu_2, \dots$ are the multipliers of $RowOp_1, ROwOp_2, \dots$.

Corollary 3.04: The $n \times n$ matrix A is invertible if and only if $det(A) \neq 0$.

Proof: Choose the row operations so that B is in Reduced Echelon Form. Let $\mu = \mu_k \mu_{k-1} \dots \mu_1 \neq 0$.

If A is not invertible, and so the rank is less than n, then B has a row of zeroes and so $det(B) = 0$. Since $\mu \neq 0$, $det(A) = 0$.

If A is invertible, then $B = I_n$ and so $det(B) = 1$. Hence, $det(A) = 1/\mu$.
□

In particular, this gives an efficient way of computing the determinant. Use the Gaussian Algorithm to put A in Echelon Form, keeping track of the multiplier for each operation.

If you hit a row of zeroes along the way, A is noninvertible and $det(A) = 0$.

Otherwise, when you reach Echelon Form, the determinant is the reciprocal of the product of the multipliers.

You need not go all the way to Reduced Echelon Form because the final steps are all Type III operations with multiplier equal to 1.

Let us look at Exercises 3.1/1e, 5bd, page 153.

Fredholm Alternative (final version): For A an $m \times m$ matrix row equivalent to Q in Reduced Echelon Form, so that $Q = UA$ and $U^{-1}Q = A$ with U a product of elementary matrices. Either:

(i)[A is *Nonsingular*]

- $r = RankA \;=\; m$,
- $Q = I_m$ and so $U^{-1} = A$ is a product of elementary matrices and is invertible with inverse U.
- For each $m \times 1$ column vector B the system $AX = B$ has the unique solution $X = A^{-1}B$.
- For B any $m \times p$ matrix, $AB = 0$ implies $B = A^{-1}0 = 0$.
- The determinant $det(A)$ is nonzero.

or else

(ii)[A is *Singular*]

- $r = RankA \; < \; m$,
- Q has $m - r$ rows of zeroes, and A is not invertible.
- There exist $m \times 1$ column vectors B such that the system $AX = B$ is inconsistent and so has no solution.
- If $AX = B$ is consistent, for example if $B = 0$ and the system is homogeneous, then there are infinitely many solutions with $m - r$ parameters.
- There exists B a nonzero $m \times m$ matrix such that $AB = 0$.
- The determinant $det(A)$ equals zero.

DET 5: If A and B are $n \times n$ matrices, then $det(AB) = det(A)det(B)$.

Proof: We have seen that AB is invertible if and only if both A and B are invertible. So AB is noninvertible, and so $det(AB) = 0$ if and only if either A or B is noninvertible and so $det(A)det(B) = 0$.

If A is invertible, then I_n is row equivalent to A and so we can write

$$A \; = \; E_k E_{k-1} \dots E_1 = RowOp_k(RowOp_{k-1}(\dots(RowOp_1(I_n))\dots))$$

and similarly

$$B \; = E'_\ell E'_{\ell-1} \dots E'_1 = RowOp'_\ell(RowOp'_{\ell-1}(\dots(RowOp'_1(I_n))\dots))$$

for some other list of row operations and associated elementary matrices. So then:

$$AB = E_k E_{k-1} \dots E_1 E'_\ell E'_{\ell-1} \dots E'_1 =$$
$$RowOp_k(\dots(RowOp_1(RowOp'_\ell(\dots(RowOp'_1(I_n))\dots)))\dots).$$

By DET 4d, $det(AB)$ is the product of the multipliers $\mu_k \dots \mu_1 \mu'_\ell \dots \mu'_1$, while $det(A) = \mu_k \dots \mu_1$ and $det(B) = \mu'_\ell \dots \mu'_1$

Thus, $det(AB) = det(A)det(B)$.
□

DET 6: If A is an $n \times n$ matrix, then $det(A^T) = det(A)$.

Proof: A is invertible if and only if A^T is invertible and so $det(A) = 0$ implies $det(A^T) = 0$.

If E is the elementary matrix associated with the Type I operation $R_i \leftrightarrow R_j$, then $E_{ij} = E_{ji} = 1$ and the only other nonzero entries are $E_{kk} = 1$ for $k \neq i, j$. Hence, $E^T = E$.

If E is the elementary matrix associated with the Type II operation $R_i \to c\ R_i$, then $E_{ii} = c$ and the only other nonzero entries are $E_{kk} = 1$ for $k \neq i$. Hence, $E^T = E$.

If E is the elementary matrix associated with the Type III operation $R_i \to R_i + c\ R_j$, then $E_{ij} = c$ and and the only other nonzero entries are $E_{kk} = 1$ for all k. So E^T is the elementary matrix associated with the Type III operation $R_j \to R_j + c\ R_i$.

So for any elementary matrix E, the transpose E^T is an elementary matrix associated with the same type and with the same multiplier.

Now assume that A is invertible and so is a product of elementary matrices $A = E_k \dots E_1$ with determinant $\mu_k \dots \mu_1$, the product of the multipliers.

The transpose of a product is the product of the transposes with the order reversed. So $A^T = E_1^T \dots E_k^T$ with determinant $\mu_1 \dots \mu_k = det(A)$.
□

Corollary 3.05: If A has a column of zeroes or two identical columns, then $det(A) = 0$.

Proof: A^T has a row of zeroes or two identical rows and so $det(A) = det(A^T) = 0$.

DET 7: If $A = \begin{pmatrix} A_1 & X \\ 0 & A_2 \end{pmatrix}$ with A_1 a $u \times u$ matrix, A_2 a $v \times v$ matrix and $n = u + v$ (that is, A is in Block Triangular Form), then $det(A) = det(A_1)det(A_2)$.

Proof (Sketch): We saw before that A is invertible if and only if A_1 and A_2 are invertible. So $det(A) = 0$ if and only if $det(A_1) = 0$ or $det(A_2) = 0$ and so if and only if $det(A_1)det(A_2) = 0$. So we may look just at the case with A_1 and A_2 invertible.

By Block multiplication, $A = \begin{pmatrix} I_u & 0 \\ 0 & A_2 \end{pmatrix} \begin{pmatrix} A_1 & X \\ 0 & I_v \end{pmatrix}$.

So $det(A) \; = \; det(\begin{pmatrix} I_u & 0 \\ 0 & A_2 \end{pmatrix})det(\begin{pmatrix} A_1 & X \\ 0 & I_v \end{pmatrix})$.

Use Type III Row Operations to knock out the non-zero entries of X. So we obtain $det(\begin{pmatrix} A_1 & X \\ 0 & I_v \end{pmatrix}) \; = \; det(\begin{pmatrix} A_1 & 0 \\ 0 & I_v \end{pmatrix})$

There are row operations which reduce A_1 to I_u. These same row operations, with the same multipliers, reduce $\begin{pmatrix} A_1 & 0 \\ 0 & I_v \end{pmatrix}$ to $\begin{pmatrix} I_u & 0 \\ 0 & I_v \end{pmatrix} = I$ and so $det(\begin{pmatrix} A_1 & 0 \\ 0 & I_v \end{pmatrix}) = det(A_1)$.

There are row operations which reduce A_2 to I_v. Using row operations with the same multipliers, we reduce $\begin{pmatrix} I_u & 0 \\ 0 & A_2 \end{pmatrix}$ to $\begin{pmatrix} I_u & 0 \\ 0 & I_v \end{pmatrix} = I$. So $det \begin{pmatrix} I_u & 0 \\ 0 & A_2 \end{pmatrix}) = det(A_2)$ as well.

$\square$

Although we have used it implicitly several times, we will later need to use the *Principle of Induction* explicitly.

A proposition $P(n)$ which depends on a positive integer n is true for all n when

- $P(1)$ is true.
- $P(n)$ can be proved when $P(k)$ is assumed to be true for all $k < n$ (Inductive Hypothesis).

That is, having checked the Initial Step, $P(1)$, we are allowed to assume what we are trying to prove but only for the earlier levels.

Rather than being a circular argument, an inductive argument spirals upward.

Corollary 3.06: If A is an upper triangular matrix, so that

$a_{ij} = 0$ when $i > j$, e.g. $A = \begin{pmatrix} a_{11} & x & x & x \\ 0 & a_{22} & x & x \\ 0 & 0 & a_{33} & x \\ . & . & . & . \\ . & . & . & . \end{pmatrix}$, then

$det(A) = a_{11}a_{22}a_{33}\ldots.$

Proof: We use induction on the size n of the matrix. The result is obvious for a 1×1 matrix.

In Block form $A = \begin{pmatrix} a_{11} & X \\ 0 & A_2 \end{pmatrix}$ with A_2 an $(n-1) \times (n-1)$ upper triangular matrix.

By DET 7 $det(A) = a_{11}det(A_1)$ and by the Inductive Hypothesis $det(A_1) = a_{22}a_{33}\ldots.$
We have proved the result for an $n \times n$ matrix, using the result for an $(n-1) \times (n-1)$ matrix. So the result is true for all n by induction.

Cofactor Expansion

For an $n \times n$ matrix A, the *ij minor* $m_{ij}(A)$, or just M_{ij} when A is understood, is the determinant of the $(n-1) \times (n-1)$ matrix A_{ij} obtained by deleting the i^{th} row and the j^{th} column of A. The minors m_{jj} are called the *principal minors* of A.

The *ij cofactor*, or *signed minor*, is $c_{ij}(A) = (-1)^{i+j} m_{ij}(A)$. Thus the signs alternate as we move along a row or column:

$$\begin{pmatrix} + & - & + & - & \cdot & \cdot \\ - & + & - & + & \cdot & \cdot \\ + & - & + & - & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

In particular, a $+$ is always attached to a principal minor.

Notice that $m_{ij}(A^T) = m_{ji}(A)$ and so $c_{ij}(A^T) = c_{ji}(A)$.

DET 8: If A is an $n \times n$ matrix and $i, j \leq n$, then we can compute the determinant of A by expanding along row i or along column j:

$$det(A) = \sum_{k=1}^{n} a_{ik} c_{ik} = a_{i1} c_{i1} + a_{i2} c_{i2} + \cdots + a_{in} c_{in}$$

$$= \sum_{k=1}^{n} a_{kj} c_{kj} = a_{1j} c_{1j} + a_{2j} c_{2j} + \cdots + a_{nj} c_{nj}.$$

Proof: Expanding along a row for A is the same as expanding along a column of A^T. Since $det(A) = det(A^T)$ it will be enough to check that if $det'(A)$ is defined by expanding along a fixed column j, then det' satisfies the conditions DET 1, DET 2 and DET 3 as these conditions characterize the determinant. We are assuming that the $(n-1) \times (n-1)$ determinant satisfies these conditions.

We will just sketch the arguments.

DET 1: If $A = I_n$ then $a_{kj} = 0$ except when $k = j$ and $a_{jj} = 1$.

Observe that $A_{jj} = I_{n-1}$ and so principal minor $m_{jj} = det(I_{n-1}) = 1$.

Because it is a principal minor its sign is positive.

Hence, $det'(I_n) = a_{jj}c_{jj} = 1$.

DET 2a: Suppose $R_\ell(B) = cR_\ell(A)$ and $R_k(B) = R_k(A)$ for all $k \neq \ell$.

It follows that $A_{\ell j} = B_{\ell j}$ which implies $c_{\ell j}(A) = c_{\ell j}(B)$ with $b_{\ell j} = c\ a_{\ell j}$.

If $k \neq \ell$, then $b_{kj} = a_{kj}$, but one row of B_{kj} is c times the corresponding row of A_{kj}. This implies that $c_{kj}(B) = c\ c_{kj}(A)$.

So a factor of c pulls out of each term of the column expansion for B.

DET 2b: Similarly, suppose $R_\ell(C) = R_\ell(A) + R_\ell(B)$ and $R_k(C) = R_k(A) = R_k(B)$ for all $k \neq \ell$.

It follows that $C_{\ell j} = A_{\ell j} = B_{\ell j}$ which implies $c_{\ell j}(C) = c_{\ell j}(A) = c_{\ell j}(B)$ with $c_{\ell j} = \ a_{\ell j} + b_{\ell j}$.

If $k \neq \ell$, then $c_{kj} = a_{kj} = b_{kj}$, but one row of C_{kj} is the sum of the corresponding rows of A_{kj} and B_{kj}. This implies that $c_{kj}(C) = \ c_{kj}(A) + c_{kj}(B)$.

So the column expansion for C splits into the sum of the column expansion of A and that of B.

DET 3 is actually the trickiest property to check. It is actually sufficient to assume that the two rows being switched are adjacent. To show this it suffices to show that if $R_i(A) = R_{i+1}(A)$, then $det'(A) = 0$.

If $k \neq i$ or $i + 1$, then two rows of A_{kj} are equal and so $c_{kj} = 0$.

Since row i equals row $i + 1$, we have $a_{ij} = a_{(i+1)j}$ and $A_{ij} = A_{(i+1)j}$ which implies that for the minors $m_{ij} = m_{(i+1)j}$ On the other hand, the signs $(-1)^{i+j}$ and $(-1)^{i+1+j}$ are opposite. This means that $c_{ij} = -c_{(i+1)j}$.

The nonzero terms in the column expansion are $a_{ij}c_{ij}$ and $a_{(i+1)j}c_{(i+1)j}$ and these sum to 0.
□

Corollary 3.07: If you expand using the entries of a row with the cofactors of a different row, or the entries of a column with the cofactors of a different column, then you get 0. That is, if $i_1 \neq i_2$ or $j_1 \neq j_2$, then

$$0 = \sum_{k=1}^{n} a_{i_1 k} c_{i_2 k} = a_{i_1 1} c_{i_2 1} + a_{i_1 2} c_{i_2 2} + \cdots + a_{i_1 n} c_{i_2 n}$$

$$= \sum_{k=1}^{n} a_{k j_1} c_{k j_2} = a_{1 j_1} c_{1 j_2} + a_{2 j_1} c_{2 j_2} + \cdots + a_{n j_1} c_{n j_2}.$$

Proof: Consider the row case. Notice that the sum never uses the entries in row i_2 since they are deleted in the computation of $c_{i_2 k}$ for every k.

Define B so that $R_{i_2}(B) = R_{i_1}(A)$ and $R_k(B) = R_k(A)$ for all $k \neq i_2$. Thus, $R_{i_1}(B) = R_{i_1}(A) = R_{i_2}(B)$. Because it has two identical rows, $det(B) = 0$.

Therefore $b_{i_2 k} = a_{i_1 k}$ and $c_{i_2 k}(A) = c_{i_2 k}(B)$ for every k. So we have:

$$\sum_{k=1}^{n} a_{i_1 k} c_{i_2 k}(A) \; = \; \sum_{k=1}^{n} b_{i_2 k} c_{i_2 k}(B) \; = \; det(B) \; = \; 0.$$

The column sum result is proved the same way, or else follows by using the transpose.
□

Now for an $n \times n$ matrix A define the *adjugate* $adj(A)$ to be the transpose of the cofactor matrix so that
$adj(A)_{ij} \;=\; c_{ji}(A)$.

We have already seen this in the 2×2 case. If $A \;=\; \begin{pmatrix} a & b \\ c & d \end{pmatrix}$,
then $adj(A) \;=\; \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$. We saw in that case that
$A\ adj(A) = det(A)I_2$. The analogous result is true in general.

DET 9: If A is an $n \times n$ matrix, then
$A\ adj(A) \;=\; det(A)I_n \;=\; adj(A)\ A$. in particular, if
$det(A) \neq 0$, then $A^{-1} = \frac{1}{det(A)} adj(A)$.

Proof: The ij entry of $A\ adj(A)$ is the dot product of the i^{th} row of A with the j^{th} column of $adj(A)$ which is the j^{th} row of the cofactor matrix. So the result follows from DET 8 and its Corollary.
□

When $n > 3$ the cofactor method is usually an inefficient way to compute the determinant as compared with the method which uses the Gaussian algorithm. However, it is useful in various constructions. For example:

Cramer's Rule: For A a nonsingular $n \times n$ matrix and B an $n \times 1$ column vector, let

$$A_i = \begin{pmatrix} C_1(A) & C_2(A) & \cdot & \cdot & C_{i-1}(A) & B & C_{i+1}(A) & \cdot & \cdot & C_n(A) \end{pmatrix}.$$

That is, replace the i^{th} column of A by the column B.

The unique solution $X = A^{-1}B$ of the system $AX = B$ is given by

$$x_i \; = \; det(A_i) \div det(A).$$

Proof: $X = A^{-1}B = \frac{1}{det(A)} adj(A)B$.

So $det(A)x_i$ is the dot product of B with the i^{th} row of $adj(A)$ which is the i^{th} column of the cofactor matrix of A.

The i^{th} column of the cofactor matrix of A is the same as the i^{th} column of the cofactor matrix of A_i. On the other hand, B is the i^{th} column of A_i.

By DET 8, the dot product of the i^{th} column of the cofactor matrix of A_i with the i^{th} column of A_i is the determinant of A_i computed by expanding along the i^{th} column.

That is, $det(A)x_i = det(A_i)$.
$\square$

Let us look at Exercises 3.2/ 4, 10f, 27, pages 167, 168.