8. Matrix Algebra

08/30/22

Homework: Problems **6.1, 6.6, 7.7, 7.22**, and **7.25** are due on **Tuesday, September 6**.

We start by defining matrices.

Matrix. An $m \times n$ *matrix* is a rectangular array $\mathbf{A}$ of $m \times n$ *elements* arranged in m rows and n columns.

For our purposes, the *elements* will be real or complex numbers or functions taking real or complex values, although more generality is allowed. The elements could themselves even be matrices.[1]

A generic element of the $m \times n$ matrix $\mathbf{A}$ can be written a_{ij} for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. The element a_{ij} is in row i, column j. We can write the matrix $\mathbf{A}$ as

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{pmatrix}$$

where I've highlighted the indices indicating row 3 in red, and column 2 in light blue.

We may sometimes write the matrix as $\mathbf{A} = \left[a_{ij}\right]$. This is not to be confused with notations such as $(a+b)_{ij}$, used to describe the ij element of $(\mathbf{A} + \mathbf{B})$.

Matrices $\mathbf{A}$ and $\mathbf{B}$ are *equal* if they are both the same size and shape ($m \times n$) and the matrices have identical entries. That is, $a_{ij} = b_{ij}$ for every $i = 1, \ldots, m$ and $j = 1, \ldots, n$.

[1] There is also no problem requiring elements be in some arbitrary field $\mathbb{F}$, or be functions with values in $\mathbb{F}$.

8.1 Matrix Addition

Matrix Addition. We can add two $m \times n$ matrices together by adding the corresponding elements. Thus $(a + b)_{ij} = a_{ij} + b_{ij}$.

Matrices of different sizes and shapes cannot be added together. The formulas above would sometimes make no sense in that case. We will say that matrices are *conformable for addition* if they are the same size and shape, that is, both have m rows and n columns.

It follows that

$$\begin{aligned}
\mathbf{A} + \mathbf{B} &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix} \\
&= \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{pmatrix}
\end{aligned}$$

Addition is both associative and commutative for matrices because it is associative and commutative for numbers. Thus

$$\begin{aligned}
(\mathbf{A} + \mathbf{B}) + \mathbf{C} &= \mathbf{A} + (\mathbf{B} + \mathbf{C}) && \text{addition associates} \\
\mathbf{A} + \mathbf{B} &= \mathbf{B} + \mathbf{A} && \text{addition commutes}
\end{aligned}$$

8.2 Demonstration that Matrix Addition Commutes

For the commutative case,

$$\begin{aligned}
\mathbf{B}+\mathbf{A} &= \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \\
&= \begin{pmatrix} b_{11}+a_{11} & b_{12}+a_{12} & \cdots & b_{1n}+a_{1n} \\ b_{21}+a_{21} & b_{22}+a_{22} & \cdots & b_{2n}+a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1}+a_{m1} & b_{m2}+a_{m2} & \cdots & b_{mn}+a_{mn} \end{pmatrix} \\
&= \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{pmatrix} \\
&= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix} \\
&= \mathbf{A}+\mathbf{B}
\end{aligned}$$

8.3 Multiplying a Matrix by a Scalar

We can also multiply a matrix by a *scalar* α, meaning a number α.

Scalar Multiplication. The matrix $\alpha\mathbf{A}$ is defined by $(\alpha a)_{ij} = \alpha a_{ij}$. We always write the scalar on the left. The product $\mathbf{A}\alpha$ is undefined.

This means

$$\alpha\mathbf{A} = \begin{pmatrix} \alpha a_{11} & \alpha a_{12} & \cdots & \alpha a_{1n} \\ \alpha a_{21} & \alpha a_{22} & \cdots & \alpha a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha a_{m1} & \alpha a_{m2} & \cdots & \alpha a_{mn} \end{pmatrix}$$

Scalar multiplication obeys two additive distributive laws. One distributes scalar multiplication over matrix addition, the other distributes scalar addition when multiplying scalars by the same matrix.

$$\alpha(\mathbf{A} + \mathbf{B}) = \alpha\mathbf{A} + \alpha\mathbf{B} \qquad \text{scalar distributive law I}$$
$$(\alpha + \beta)\mathbf{A} = \alpha\mathbf{A} + \beta\mathbf{A} \qquad \text{scalar distributive law II}$$

8.4 The Zero Matrix

One special matrix is the zero matrix.

Zero Matrix. The $m \times n$ *zero matrix,* $\mathbf{0}$, is defined by $a_{ij} = 0$ for every $i = 1, \ldots, m$ and $j = 1, \ldots, n$. Every element is zero.

Not surprisingly, adding the zero matrix to any matrix gives that matrix as the sum. It is an *additive identity*. It is easy to show that there can be only one additive identity. The zero matrix of size $m \times n$ can also be obtained by multiplying any matrix of size $m \times n$ by zero: $0\mathbf{A} = \mathbf{0}$.

$$\begin{aligned} \mathbf{0} + \mathbf{A} = \mathbf{A} + \mathbf{0} = \mathbf{A} &\qquad \text{additive identity} \\ 0\mathbf{A} = \mathbf{0} &\qquad \text{0 and additive identity} \end{aligned}$$

8.5 Additive Inverse

Each matrix $\boldsymbol{A}$ has an additive inverse, a matrix we can add to $\boldsymbol{A}$ to obtain the zero matrix.

Additive Inverse. Each matrix $\boldsymbol{A}$ has a unique *additive inverse* $-\boldsymbol{A}$, which can be obtained by multiplying $\boldsymbol{A}$ by (-1).

We know $-\boldsymbol{A}$ is an additive inverse because

$$\begin{aligned}
-\boldsymbol{A} + \boldsymbol{A} &= \boldsymbol{A} - \boldsymbol{A} \\
&= (1)\boldsymbol{A} + (-1)\boldsymbol{A} \\
&= (1-1)\boldsymbol{A} \\
&= 0\boldsymbol{A} \\
&= \boldsymbol{0}.
\end{aligned}$$

We have

$$\begin{aligned}
\boldsymbol{A} + (-\boldsymbol{A}) = (-\boldsymbol{A}) + \boldsymbol{A} &= \boldsymbol{0} && \text{additive inverse} \\
(-1)\boldsymbol{A} &= -\boldsymbol{A} && -1 \text{ and additive inverse}
\end{aligned}$$

8.6 Matrix Addition and Scalar Multiplication

This gives us the following properties involving matrix addition and scalar multiplication. These and other properties listed later only apply to conformable matrices.

$(\mathbf{A}+\mathbf{B})+\mathbf{C}=\mathbf{A}+(\mathbf{B}+\mathbf{C})$	addition associates
$\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$	addition commutes
$0\mathbf{A}=\mathbf{0}$	0 and additive identity
$(-1)\mathbf{A}=-\mathbf{A}$	-1 and additive inverse
$\mathbf{0}+\mathbf{A}=\mathbf{A}+\mathbf{0}=\mathbf{A}$	additive identity
$\mathbf{A}+(-\mathbf{A})=(-\mathbf{A})+\mathbf{A}=\mathbf{0}$	additive inverse
$\alpha(\mathbf{A}+\mathbf{B})=\alpha\mathbf{A}+\alpha\mathbf{B}$	scalar distributive law I
$(\alpha+\beta)\mathbf{A}=\alpha\mathbf{A}+\beta\mathbf{A}$	scalar distributive law II

There are no real surprises when it comes to matrix addition and multiplication of a matrix by a number (*scalar multiplication*). However, we have yet to consider matrix multiplication.

8.7 Complex Matrices

There's no reason to restrict ourselves to real numbers when defining matrices. We could use any field. The only fields that we will use are the real numbers $\mathbb{R}$ and complex numbers $\mathbb{C}$.

Recall that the complex numbers can be written $z = a + bi$ where $a, b \in \mathbb{R}$ and the imaginary unit i is defined as the square root of -1, $i = \sqrt{-1}$. We arbitrarily call this the positive square root of -1. Then $-i = -\sqrt{-1}$.

The real numbers are contained in the complex numbers. Just set $b = 0$. The purely imaginary numbers are also there, with $a = 0$.

One key fact about the complex numbers is that every solution to any complex polynomial equation in one variable is a complex number. In fact, the Fundamental Theorem of Algebra states that every complex polynomial $p(z)$ of degree n in z can be factored as

$$p(z) = \alpha(z - \lambda_1) \cdots (z - \lambda_n)$$

where α and each $\lambda_i \in \mathbb{C}$.[2]

This is not true if we restrict ourselves to real numbers. The equation $x^2 + 1 = 0$ has no real solutions and cannot be factored using real numbers. However, there are two complex solutions: i and $-i$. The complex factorization is

$$x^2 + 1 = (x - i)(x + i).$$

[2] As I write this in August 2020, Wikipedia points out that the theorem was named when algebra was primarily about solving polynomial equations. They comment that "Additionally, it is not fundamental for modern algebra; its name was given at a time when algebra was synonymous with theory of equations." This is true in the sense that the Fundamental Theorem of Algebra is not as fundamental to algebra as it used to be. Nonetheless, huge chunks of modern algebra are still focused on the theory of equations, it's just that the theory has reached an incredible level of abstraction. One only has to look at Wiles's proof of Fermat's Last Theorem to see this.

8.8 Aside on Fields

Field. A *field* is a set $\mathbb{F}$ with two binary operations, denoted $+$ (addition) and $\cdot$ (multiplication). They obey the following axioms

1. Addition and multiplication are both *associative*. For all $a, b, c \in \mathbb{F}$, $(a + b) + c = a + (b + c)$ and $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.
2. Addition and multiplication are both *commutative*. For all $a, b, c \in \mathbb{F}$, $a + b = b + a$ and $a \cdot b = b \cdot a$.
3. Addition and multiplication both have *identities*, denoted 0 and 1, respectively. For all $a \in \mathbb{F}$, $a + 0 = a$ and $a \cdot 1 = a$.
4. Every element of $\mathbb{F}$ has an *additive inverse*. For all $a \in \mathbb{F}$, there is $-a \in \mathbb{F}$ with $a + (-a) = 0$.
5. Every non-zero element of $\mathbb{F}$ has a *multiplicative inverse*. For $a \in \mathbb{F}$, $a \neq 0$, there is $a^{-1} \in \mathbb{F}$ with $a \cdot a^{-1} = 1$. The notation $1/a$ is also used for a^{-1}.
6. Multiplication *distributes* over addition. For all $a, b, c \in \mathbb{F}$, $a \cdot (b + c) = a \cdot b + a \cdot c$.

One can easily show that the additive and multiplicative identities are unique. E.g., if a and b are both multiplicative identities, $a = b \cdot a = a \cdot b = b$.

8.9 Examples of Fields

Examples of fields include the rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$, and the complex numbers $\mathbb{C}$, and the set of rational complex numbers.

If p is a prime number, and a and b are integers, we say $a = b \mod p$ if $a-b$ is divisible by p. This defines a set of p equivalence classes of integers, which we call $0, 1, \ldots, (p-1)$. This set, the set of integers mod p can be denoted $\mathbb{Z}/(p)$ or $\mathbb{Z}/p\mathbb{Z}$ (the natural notation $\mathbb{Z}_p$ is sometimes used, but is also used for the p-adic integers). The set $\mathbb{Z}/(p)$ is a field when p is prime.

If n is not prime, we can still consider the integers mod n, but they do not form a field. If $n > 0$ is not prime we can factor $n = km$ where $k, m > 1$. Then while $k, m \neq 0 \mod n$, $km = 0 \mod n$. This means that $\mathbb{Z}/(n)$ is not a field.

We can also form fields by taking the rational numbers and adding multiples of square roots. E.g., the numbers that can be written $a + b\sqrt{2}$ where $a, b \in \mathbb{Q}$ form a field. The key point is that $(\sqrt{2})^{-1} = \frac{1}{2}\sqrt{2}$ is such a number.

8.10 Transpose of a Matrix

A matrix can be transformed by transposing it.

Transpose of a Matrix. Given an $m \times n$ matrix $\mathbf{A}$, its *transpose*, $\mathbf{A}^{\mathsf{T}}$ is the $n \times m$ matrix defined by $a^{\mathsf{T}}_{ij} = a_{ji}$.

In other words, we interchange rows and columns to transpose the matrix. Basically, we are flipping it along its *main diagonal*, consisting of the elements a_{ii}.

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}^{\mathsf{T}} = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}.$$

It is easy to verify that the transpose is compatible with both addition and scalar multiplication.

$$(\mathbf{A} + \mathbf{B})^{\mathsf{T}} = \mathbf{A}^{\mathsf{T}} + \mathbf{B}^{\mathsf{T}}$$
$$(\alpha\mathbf{A})^{\mathsf{T}} = \alpha\mathbf{A}^{\mathsf{T}}$$

It is also easy to see that the transpose of a transpose is the original matrix.

$$(\mathbf{A}^{\mathsf{T}})^{\mathsf{T}} = \mathbf{A}.$$

8.11 Symmetric and Skew-Symmetric Matrices

If a matrix is square, it may be its own transpose, $\mathbf{A} = \mathbf{A}^\mathsf{T}$, meaning $a_{ij} = a_{ji}$ for all i and j. Such matrices are called *symmetric*. A related concept is skew-symmetry. A matrix $\mathbf{A}$ is *skew-symmetric* if $\mathbf{A}^\mathsf{T} = -\mathbf{A}$.

The matrix $\mathbf{A}$ is symmetric and $\mathbf{B}$ is skew-symmetric.

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix}.$$

As with $\mathbf{B}$, the main diagonal of any skew-symmetric matrix is zero since $a_{ii} = -a_{ii}$.

Any square matrix can be decomposed into a sum of a symmetric matrix and a skew-symmetric matrix.

Theorem 8.11.1. *If $\mathbf{A}$ is a square matrix, then $\mathbf{B} = \frac{1}{2}(\mathbf{A}+\mathbf{A}^\mathsf{T})$ is symmetric, $\mathbf{C} = \frac{1}{2}(\mathbf{A} - \mathbf{A}^\mathsf{T})$ is skew-symmetric. Moreover, $\mathbf{A} = \mathbf{B} + \mathbf{C}$.*

Proof. Taking the transposes of $\mathbf{B}$ and $\mathbf{C}$ shows they are symmetric and skew-symmetric, respectively. Simple addition shows $\mathbf{B} + \mathbf{C} = \mathbf{A}$. ∎

8.12 Hermitian Conjugate of a Matrix

A related concept that only affects complex matrices is the Hermitian conjugate. The *complex conjugate* of $z = a + bi$ where $a, b \in \mathbb{R}$ is $\bar{z} = a - bi$. One nice property of the conjugate is that $z\bar{z} = \bar{z}z = a^2 + b^2 = |z|^2$.

Hermitian Conjugate. The *Hermitian conjugate* $\mathbf{A}^*$ of a matrix $\mathbf{A}$ is the complex conjugate of $\mathbf{A}^\mathsf{T}$. Thus $a^*_{ij} = \bar{a}_{ji}$.

It is easy to see how the Hermitian conjugate is compatible with addition and, with a twist, scalar multiplication.

$$(\mathbf{A} + \mathbf{B})^* = \mathbf{A}^* + \mathbf{B}^*$$
$$(\alpha\mathbf{A})^* = \bar{\alpha}\mathbf{A}^*$$

As with transposes, it is easy to see that the Hermitian conjugate of a Hermetian conjugate is the original matrix.

$$(\mathbf{A}^*)^* = \mathbf{A}.$$

8.13 Hermitian and Anti-Hermitian Matrices

A matrix is *Hermitian* if $\mathbf{A}^* = \mathbf{A}$ and *skew-Hermitian* or *anti-Hermitian* if $\mathbf{A}^* = -\mathbf{A}$. In particular, Hermitian matrices obey $a_{ii} = \bar{a}_{ii}$, implying that the main diagonal is real; anti-Hermitian matrices obey $a_{ii} = -\bar{a}_{ii}$, yielding a purely imaginary diagonal.

Hermitian matrices are a complex matrix analog of real numbers and skew-Hermitian matrices are an analog of purely imaginary numbers, even though neither need be purely real or imaginary, except on the main diagonal. For example, the matrix

$$\mathbf{A} = \begin{pmatrix} +i & -1 \\ +1 & -i \end{pmatrix}$$

is skew-Hermitian as

$$\mathbf{A}^* = \begin{pmatrix} -i & +1 \\ -1 & +i \end{pmatrix} = -\mathbf{A}.$$

Any square matrix can be decomposed into a sum of a Hermitian matrix and a anti-Hermitian matrix.

Theorem 8.13.1. *If $\mathbf{A}$ is a square matrix, then $\mathbf{B} = \frac{1}{2}(\mathbf{A}+\mathbf{A}^*)$ is Hermitian, $\mathbf{C} = \frac{1}{2}(\mathbf{A} - \mathbf{A}^*)$ is anti-Hermitian. Moreover $\mathbf{A} = \mathbf{B} + \mathbf{C}$.*

Proof. Taking the Hermitian conjugates of $\mathbf{B}$ and $\mathbf{C}$ shows they are Hermitian and anti-Hermitian, respectively. Simple addition shows $\mathbf{B} + \mathbf{C} = \mathbf{A}$. ∎

8.14 Matrix Multiplication I

If the sizes are right, matrices can be multiplied. The size condition is that the number of columns in the first matrix must be equal to the number of rows in the second. We can multiply an $m \times n$ matrix by an $n \times k$ matrix to obtain an $m \times k$ matrix. Multiplying them in the opposite order is only possible is $k = m$.

Multiplication gives us a second type of conformability. Two matrices $\mathbf{A}$ and $\mathbf{B}$ are *conformable for multiplication* if the number of columns in $\mathbf{A}$ and number of rows in $\mathbf{B}$ are the same.

Matrix Multiplication. When matrices are conformable for multiplication, the *matrix product* $\mathbf{A} \times \mathbf{B}$ (sometimes written $\mathbf{AB}$) is defined as follows. Then

$$(a \times b)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$$

or using the summation notation

$$(a \times b)_{ij} = \sum_{h=1}^{n} a_{ih}b_{hj}$$

for all $i = 1, \ldots, m$ and $j = 1, \ldots, k$.

The matrix product easily relates to the transpose. When we take the transpose of a matrix product, we get the product of the transposes in reverse order: $(\mathbf{AB})^{\mathsf{T}} = \mathbf{B}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}$. The same thing happens with Hermitian conjugates: $(\mathbf{AB})^* = \mathbf{B}^*\mathbf{A}^*$.

8.15 Matrix Multiplication II

Let $\mathbf{A}$ and $\mathbf{B}$ be $m \times n$ and $n \times k$ matrices, respectively. Such matrices are conformable for multiplication. To get a better feel for how matrix multiplication works, we examine the ij element of $\mathbf{A} \times \mathbf{B}$, $(a \times b)_{ij}$. To compute it, we use row i of $\mathbf{A}$ and column j of $\mathbf{B}$. Then

$$(a \times b)_{ij} = \begin{pmatrix} a_{i1} & a_{i2} & a_{i3} & \cdots a_{ij} \end{pmatrix} \times \begin{pmatrix} b_{1j} \\ b_{2j} \\ b_{3j} \\ \vdots \\ b_{kj} \end{pmatrix}$$

We match up the corresponding terms of the two vectors (row and column), multiply them together, and add to get the result. Thus

$$(a \times b)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \cdots + a_{ij}b_{kj}.$$

We do this for every row i of $\mathbf{A}$ and column j of $\mathbf{B}$, until we've covered all possibilities. If you're familiar with the dot product, you can think of this as the dot product of the vectors $(a_{i.})$ and $(b_{.j})$.

8.16 Matrix Multiplication: Examples

We consider some examples of multiplying matrices of different sizes and shapes together.

$$\begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix} \times \begin{pmatrix} 10 & 11 & 13 \\ 7 & 5 & 1 \end{pmatrix} = \begin{pmatrix} 24 & 21 & 15 \\ 31 & 26 & 16 \end{pmatrix}.$$

These two matrices cannot be multiplied in the opposite order because you can't multiply a 2×3 matrix by a 2×2 matrix. In general, we cannot assume that matrix multiplication commutes.

What if the multiplication makes sense? Consider the 1×3 matrix

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}.$$

Then

$$\boldsymbol{A} \times \boldsymbol{A}^{\mathsf{T}} = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \times \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = (1 + 4 + 9) = (14).$$

If we take the product in the opposite order, we get something entirely different.

$$\boldsymbol{A}^{\mathsf{T}} \times \boldsymbol{A} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \times \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix}.$$

Here both products are defined, but the matrices still fail to commute. The size differences between the products make it impossible for the matrices to commute.

8.17 Matrix Multiplication: Square Examples

When matrices are *square* (same number of rows and columns) and the same size, it makes sense to multiply them in either order. Now both products are square and have the same size. That still doesn't guarantee that they commute! Thus

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \times \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 7 & 8 & 9 \\ 4 & 5 & 6 \end{pmatrix} \tag{8.17.1}$$

but

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 3 & 2 \\ 4 & 6 & 5 \\ 7 & 9 & 8 \end{pmatrix} \tag{8.17.2}$$

The products are not the same. These two matrices do not commute under multiplication (of course, addition is still commutative).

Interestingly enough, in equation (8.17.1), **pre**-multiplication by the matrix consisting of 0's and 1's switches the second and third rows of the $1, 2, 3, \ldots$ matrix. The pre-multiplication has carried out an elementary row operation.

Even more interestingly, in equation (8.17.2), **post**-multiplication by the same matrix switches the columns of the $1, 2, 3, \ldots$ matrix.

8.18 The Identity Matrix: Definition

To define the identity matrix, we first define the Kronecker delta, which is a function from index pairs ij to the real numbers.[3]

Kronecker delta. The *Kronecker delta*, δ_{ij} is defined by

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

The Kronecker delta will be useful for defining matrices, beginning with the identity matrix.

Identity Matrix. We define the $n \times n$ *identity matrix* I_n by $i_{ij} = \delta_{ij}$. In other words,

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

[3] Leopold Kronecker (1823–1891) used the Kronecker delta in his lectures. Although it's an extremely useful construction, it didn't find it's way into to the literature until Luther Pfahler Eisenhard used it in his 1926 book *Riemannian Geometry*, Princeton University Press, Princeton, NJ.

8.19 The Identity Matrix: Properties

The identity matrix has the property that for any $m \times n$ matrix $\mathbf{A}$,

$$\mathbf{I}_m \times \mathbf{A} = \mathbf{A} \text{ and } \mathbf{A} \times \mathbf{I}_n = \mathbf{A} \qquad \text{multiplicative identities.}$$

In other words, $\mathbf{I}_m$ is a (left) multiplicative identity, and $\mathbf{I}_n$ is a (right) multiplicative identity.

Theorem 8.19.1. *Suppose* $\mathbf{A}$ *is an* $m \times n$ *matrix. Then* $\mathbf{I}_m \times \mathbf{A} = \mathbf{A} = \mathbf{A} \times \mathbf{I}_n$.

Proof. We consider the first equality. Call the product $\mathbf{B}$. Then

$$b_{ij} = \sum_{k=1}^{m} \delta_{ik} a_{kj} = a_{ij}$$

as all the terms with $i \neq k$ are zero due to the Kronecker delta. So we have $\mathbf{B} = \mathbf{A}$.

A similar argument establishes the other equality. ∎

If $\mathbf{A}$ is a square matrix, $n = m$ and $\mathbf{I} = \mathbf{I}_m$ commutes with $\mathbf{A}$. In fact, any scalar multiple of $\mathbf{I}$ commutes with $\mathbf{A}$. There may be other matrices that commute with $\mathbf{A}$.

8.20 Matrix Multiplication: Basic Properties

So what can we say about matrix multiplication? For conformable matrices, the following identities hold:

$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \times \mathbf{B}) \times \mathbf{C}$	multiplication associates
$\alpha(\mathbf{A} \times \mathbf{B}) = (\alpha\mathbf{A}) \times \mathbf{B} = \mathbf{A} \times (\alpha\mathbf{B})$	scalar associative law
$(\mathbf{A} + \mathbf{B}) \times \mathbf{C} = \mathbf{A} \times \mathbf{C} + \mathbf{B} \times \mathbf{C}$	matrix distributive law I
$\mathbf{A} \times (\mathbf{B} + \mathbf{C}) = \mathbf{A} \times \mathbf{B} + \mathbf{A} \times \mathbf{C}$	matrix distributive law II
$\mathbf{A} = \mathbf{I}_m \times \mathbf{A} = \mathbf{A} \times \mathbf{I}_n$	identity matrices

You'll notice the absence of a commutative law for matrix multiplication. Matrix multiplication usually does not commute. Suppose $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $n \times k$. Then $\mathbf{A} \times \mathbf{B}$ exists, but $\mathbf{B} \times \mathbf{A}$ will only exist if $k = m$. In the latter case, $\mathbf{A} \times \mathbf{B}$ is $m \times m$ and $\mathbf{B} \times \mathbf{A}$ is $n \times n$. These two products can only be the same if $n = m$. In other words, we can only think about matrix multiplication commuting when both matrices are both square and the same size. The following examples illustrate this.

8.21 Linear Systems and Matrices

We can use matrix multiplication to write any linear system as a matrix product. Recall our original linear system.

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ a_{31}x_1 + a_{32}x_2 + \cdots + a_{3n}x_n &= b_3 \\ \vdots \qquad\qquad\qquad & \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{6.7.2}$$

We let $\mathbf{x}$ denote the $n \times 1$ column vector of variables and $\mathbf{b}$ the column vector of constant terms. Thus

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

Let $\mathbf{A}$ be the $m \times n$ coefficient matrix with ij element a_{ij}. Then the system (6.7.2) can be written

$$\mathbf{A}\mathbf{x} = \mathbf{b}.$$

To make the formula work, we **have** to write the vectors of variables and constant terms as column vectors, not row vectors.

As you may guess, matrix algebra will be useful in solving these systems.

8.22 Vectors

Vectors. A vector in $\mathbb{R}^n$ is an n-tuple of real numbers. There are two types: row vectors and column vectors. The first can be written as $1 \times n$ matrices, while the column vectors are written as $n \times 1$ matrices. We denote vectors by lowercase bold letters such as $\mathbf{a}$ or $\mathbf{x}$. Vectors are not numbers. They are n-tuples of numbers.

Vectors can be rows or columns. E.g.,

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix} \qquad \text{or} \qquad \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

8.23 Matrix Inverses

A matrix that has an multiplicative inverse is called invertible.

Invertible Matrix. If $\mathbf{A}$ is a square matrix, it has an *inverse* if there is a matrix $\mathbf{A}^{-1}$ with $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$. A matrix with an inverse is called *invertible*.

Invertible matrices are always non-singular.

Theorem 8.23.1. *Suppose an* $n \times n$ *matrix* $\mathbf{A}$ *is invertible. Then* rank $\mathbf{A} = n$ *and* $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ *is the unique solution to the system* $\mathbf{A}\mathbf{x} = \mathbf{b}$. *Moreover,* $\mathbf{A}$ *is non-singular.*

Proof. If $\mathbf{A}$ is invertible, then $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ solves the system for every $\mathbf{b}$ by the rules of matrix algebra. By Corollary 7.30.1, rank $\mathbf{A}$ is the number of rows, n. The number of columns is also n, so $\mathbf{A}$ is non-singular by Corollary 7.31.2. ∎

Non-singular Matrices. It is not hard to show that any non-singular matrix is invertible. We do this later, immediately prior to the statement of Theorem 8.39.1.

8.24 Invertibility and Transposition

We can also relate the transpose and inverse of a matrix. The inverse of the transpose is the transpose of the inverse and the inverse of the Hermitian conjugate is the Hermitian conjugate of the inverse.

Theorem 8.24.1. *Suppose* $\mathbf{A}$ *is invertible. Then* $(\mathbf{A}^{-1})^{\mathsf{T}} = (\mathbf{A}^{\mathsf{T}})^{-1}$ *and* $(\mathbf{A}^{-1})^{*} = (\mathbf{A}^{*})^{-1}$.

Proof. We know $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I} = \mathbf{A}^{-1}\mathbf{A}$. Take the transpose to obtain $(\mathbf{A}^{-1})^{\mathsf{T}}\mathbf{A}^{\mathsf{T}} = \mathbf{I} = \mathbf{A}^{\mathsf{T}}(\mathbf{A}^{-1})^{\mathsf{T}}$, showing that the transpose of $\mathbf{A}^{-1}$ is the inverse of $\mathbf{A}^{\mathsf{T}}$. The same argument applies to the Hermitian conjugate. ∎

8.25 Left and Right Inverses

If $\mathbf{A}$ is an $m \times n$ matrix with $m \neq n$, it is still possible to find either a left inverse or a right inverse. A *left inverse* is an $n \times m$ matrix $\mathbf{B}$ with $\mathbf{BA} = \mathbf{I}_n$ and a *right inverse* is an $n \times m$ matrix $\mathbf{C}$ with $\mathbf{AC} = \mathbf{I}_m$. If $\mathbf{A}$ has both left and right inverses, it must be square and the inverses must be identical as shown in Theorem 8.27.1.

For example,

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 4 \end{pmatrix}$$

has right inverse

$$\mathbf{B} = \begin{pmatrix} 2 & 1 \\ 1 & -2 \\ -1 & 1 \end{pmatrix}$$

because

$$\mathbf{A} \times \mathbf{B} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 4 \end{pmatrix} \times \begin{pmatrix} 2 & 1 \\ 1 & -2 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

However,

$$\mathbf{B} \times \mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & -2 \\ -1 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 4 \end{pmatrix} = \begin{pmatrix} 3 & 6 & 10 \\ -1 & -2 & -5 \\ 0 & 0 & 1 \end{pmatrix}.$$

In fact, $\mathbf{A}$ has no left inverse.

An example of a matrix with a left inverse but not a right inverse is $\mathbf{A}^T$, which has left inverse $\mathbf{B}^T$.

8.26 One-sided Inverses and Linear Systems

The one-sided inverses are connected to the properties of the linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$.

Theorem 8.26.1.

1. *If $\mathbf{A}$ has a left inverse, then there is at most one solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$ and* rank $\mathbf{A}$ *is equal to the number of columns.*
2. *If $\mathbf{A}$ has a right inverse, then there is a solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$ and* rank $\mathbf{A}$ *is equal to the number of rows.*

Proof. (1) Suppose $\mathbf{A}$ has a left inverse $\mathbf{B}$. Suppose it has two solutions, $\mathbf{x}$ and $\mathbf{x}'$. Then

$$\mathbf{A}\mathbf{x} = \mathbf{b} \text{ and } \mathbf{A}\mathbf{x}' = \mathbf{b}.$$

Apply the left inverse to both equations, yielding

$$\mathbf{x} = \mathbf{B}\mathbf{A}\mathbf{x} = \mathbf{B}\mathbf{b} \text{ and } \mathbf{x}' = \mathbf{B}\mathbf{A}\mathbf{x}' = \mathbf{B}\mathbf{b}.$$

Combining them, we see that $\mathbf{x} = \mathbf{x}'$. There is at most one solution. By Corollary 7.31.2, rank $\mathbf{A}$ is the number of columns of $\mathbf{A}$.

(2) Suppose $\mathbf{A}$ has a right inverse $\mathbf{C}$. Then $\mathbf{A}(\mathbf{C}\mathbf{b}) = (\mathbf{A}\mathbf{C})\mathbf{b} = \mathbf{b}$, so $\mathbf{x} = \mathbf{C}\mathbf{b}$ is a solution to the system. Since this system always has a solution, rank $\mathbf{A}$ is the number of rows of $\mathbf{A}$ by Corollary 7.30.1. ∎

8.27 Two-Sided Inverses

If a matrix has both left and right inverses, they must be the same and the matrix must be invertible.

Theorem 8.27.1. *If an $m \times n$ matrix $\mathbf{A}$ has both a left inverse $\mathbf{B}$ and a right inverse $\mathbf{C}$, then $\mathbf{B} = \mathbf{C}$ and $m = n$. Furthermore, $\mathbf{B} = \mathbf{C}$ is the inverse of $\mathbf{A}$.*

Proof. Suppose $\mathbf{B}$ is a left inverse and $\mathbf{C}$ a right inverse. Then $\mathbf{BA} = \mathbf{I}$. It follows that

$$\begin{aligned}(\mathbf{BA})\mathbf{C} &= \mathbf{IC}\\ \mathbf{B}(\mathbf{AC}) &= \mathbf{C}\\ \mathbf{BI} &= \mathbf{C}\\ \mathbf{B} &= \mathbf{C},\end{aligned}$$

showing that the two inverses must be identical.

Theorem 8.26.1 tells us that rank $\mathbf{A}$ = #cols = #rows. This means that $\mathbf{A}$ is non-singular.

Finally, since $\mathbf{A}$ is $m \times m$ and $\mathbf{BA} = \mathbf{I}_m = \mathbf{AB}$, which implies $\mathbf{B} = \mathbf{C} = \mathbf{A}^{-1}$, the inverse of $\mathbf{A}$. ∎

8.28 Inverse of a Scalar Product

If an $n \times n$ matrix is invertible and $\alpha \neq 0$, the scalar product $\alpha\mathbf{A}$ is also invertible.

Theorem 8.28.1. *Suppose $\mathbf{A}$ is an invertible $n \times n$ matrix and $\alpha \neq 0$. Then $\alpha\mathbf{A}$ is also invertible with $(\alpha\mathbf{A})^{-1} = \alpha^{-1}\mathbf{A}^{-1}$.*

Proof.

$$\begin{aligned}(\alpha\mathbf{A}) \times \alpha^{-1}\mathbf{A}^{-1} &= \mathbf{A} \times \mathbf{A}^{-1} \\ &= \mathbf{I} \\ &= \mathbf{A}^{-1} \times \mathbf{A} \\ &= \alpha^{-1}\mathbf{A}^{-1} \times (\alpha\mathbf{A})\end{aligned}$$

establishing the result. ∎

8.29 Inverse of a Matrix Product

If two $n \times n$ matrices are invertible, their matrix product is also invertible.

Theorem 8.29.1. *Suppose* $\mathbf{A}$ *and* $\mathbf{B}$ *are invertible* $n \times n$ *matrices. Then* $\mathbf{AB}$ *is also invertible with* $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

Proof. The proof is simple. Both

$$\begin{aligned}(\mathbf{B}^{-1}\mathbf{A}^{-1}) \times (\mathbf{AB}) &= \big((\mathbf{B}^{-1}\mathbf{A}^{-1})\mathbf{A}\big)\mathbf{B} \\ &= \big(\mathbf{B}^{-1}(\mathbf{A}^{-1}\mathbf{A})\big)\mathbf{B} \\ &= (\mathbf{B}^{-1}\mathbf{I})\mathbf{B} \\ &= \mathbf{B}^{-1}\mathbf{B} \\ &= \mathbf{I}\end{aligned}$$

and

$$\begin{aligned}(\mathbf{AB}) \times (\mathbf{B}^{-1}\mathbf{A}^{-1}) &= \big((\mathbf{AB})\mathbf{B}^{-1}\big)\mathbf{A}^{-1} \\ &= \big(\mathbf{A}(\mathbf{BB}^{-1})\big)\mathbf{A}^{-1} \\ &= \mathbf{AA}^{-1} \\ &= \mathbf{I},\end{aligned}$$

establishing the result. ∎

8.30 Inverses of Diagonal Matrices

A square matrix is called a *diagonal matrix* if the only non-zero elements are those on the *main diagonal*, elements of the form a_{ii}. We will denote a diagonal matrix with $(\lambda_1, \ldots, \lambda_n)$ on the diagonal by $\operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Thus

$$\operatorname{diag}(\lambda_1, \ldots, \lambda_n) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

It is easily verified that diagonal matrices with no zeros on the diagonal can be inverted. In fact

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}^{-1} = \begin{pmatrix} 1/\lambda_1 & 0 & \cdots & 0 \\ 0 & 1/\lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & 1/\lambda_n \end{pmatrix}$$

Whenever $\lambda_1 \cdots \lambda_n \neq 0$.

8.31 All Diagonal Matrices Commute

It is also the case that any two diagonal matrices of the same size commute. In fact, their product is also a diagonal matrix with the product of the diagonal elements on the diagonal.

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \begin{pmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \mu_n \end{pmatrix} = \begin{pmatrix} \lambda_1\mu_1 & 0 & \cdots & 0 \\ 0 & \lambda_2\mu_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_n\mu_n \end{pmatrix}$$
$$= \begin{pmatrix} \mu_1\lambda_1 & 0 & \cdots & 0 \\ 0 & \mu_2\lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \mu_n\lambda_n \end{pmatrix} = \begin{pmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \mu_n \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

Or more concisely,

$$\begin{aligned} \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \times \operatorname{diag}(\mu_1, \ldots, \mu_n) \\ &= \operatorname{diag}(\lambda_1\mu_1, \ldots, \lambda_n\mu_n) \\ &= \operatorname{diag}(\mu_1\lambda_1, \ldots, \mu_n\lambda_n) \\ &= \operatorname{diag}(\mu_1, \ldots, \mu_n) \times \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \end{aligned}$$

8.32 Intro to Elementary Row Matrices

There are two classes of elementary matrices—elementary row matrices and elementary column matrices. Pre-multiplying a matrix $\boldsymbol{A}$ by an elementary row matrix carries out the corresponding elementary row operation. Post-multiplying by an elementary column matrix carries out the corresponding elementary column operation.

We saw this earlier. Recall equation (8.17.1).

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \times \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 7 & 8 & 9 \\ 4 & 5 & 6 \end{pmatrix} \tag{8.17.1}$$

Pre-multiplying by the matrix

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

switched the second and third rows—an elementary row operation. In equation (8.17.2), post-multiplying by the same matrix switched the second and third columns, an elementary column operation.

8.33 Elementary Row Matrices I: Switching Rows

More generally, suppose we form the matrix $\mathbf{E}_{ij}$ by taking the identity matrix, and switching the i^{th} and j^{th} rows (this is the same as switching the i^{th} and j^{th} columns). Then pre-multiplying any matrix $\boldsymbol{A}$ by this matrix will switch $\boldsymbol{A}$'s i^{th} and j^{th} rows.

The matrix $\mathbf{E}_{ij}$ is the $m \times m$ matrix with elements

$$e_{hk} = \begin{cases} 0 & \text{when } hk = ii \text{ or } hk = jj \\ 1 & \text{when } hk = ij \text{ or } hk = ji \\ \delta_{hk} & \text{otherwise.} \end{cases}$$

We can now calculate the product for any $m \times n$ matrix $\boldsymbol{A}$. Let $c_{k\ell}$ denote the elements of $\mathbf{E}_{ij} \times \boldsymbol{A}$.

$$c_{k\ell} = \sum_{h=1}^{m} e_{kh} a_{h\ell} = \begin{cases} a_{k\ell} & \text{when } k \neq i, j \\ a_{j\ell} & \text{when } k = i \\ a_{i\ell} & \text{when } k = j. \end{cases}$$

To see this, first suppose $k \neq i, j$. Then $e_{kh} = \delta_{kh}$ and $c_{k\ell} = \sum_h \delta_{kh} a_{h\ell} = a_{k\ell}$. If instead, $k = i$, we have $c_{i\ell} = \sum_h e_{ih} a_{h\ell} = a_{j\ell}$. Finally, if $k = j$, the sum is $c_{j\ell} = \sum_h e_{jh} a_{h\ell} = a_{i\ell}$.

In other words, the i^{th} and j^{th} rows of $\boldsymbol{A}$ have been switched.

8.34 Elementary Row Matrices II: Multiplying a Row

The other two types of elementary row operations have their own elementary matrices. All are formed by applying the desired row operation to the identity matrix.

To multiply row i by $r \neq 0$, we define the matrix $\mathbf{E}_i(r)$ by

$$e_{hk} = \begin{cases} r\delta_{ik} & \text{when } h = i \\ \delta_{hk} & \text{when } h \neq i. \end{cases}$$

For example,

$$\mathbf{E}_2(3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

As before, we use the notation $[c_{k\ell}]$ for the product. If $k \neq i$, $c_{k\ell} = \sum_h e_{kh}a_{h\ell} = \sum_h \delta_{kh}a_h\ell = a_{k\ell}$ and if $k = i$, $c_{i\ell} = \sum_h e_{ih}a_{h\ell} = \sum_h r\delta_{ih}a_{h\ell} = ra_{i\ell}$.

Only the i^{th} row (column) is changed, and it is multiplied by r.

8.35 Elementary Row Matrices III

The third type of elementary row operation adds a multiple of one row to another. To add r times row i to row j, we define the matrix $\mathbf{E}_{ij}(r)$ by

$$e_{hk} = \begin{cases} e_{jj} = 1 & \\ e_{ji} = r & \\ e_{jk} = 0 & \text{when } k \neq i, j \\ \delta_{hk} & \text{when } h \neq j. \end{cases}$$

The only change from the identity occurs in row j, where there is an r in column i instead of a 0.

Here's an example, adding r times row 2 to row 3:

$$\mathbf{E}_{23}(r) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & r & 1 \end{pmatrix}.$$

Showing that this works is left as an exercise to the reader.

8.36 Elementary Row Matrices: Invertibility 09/01/22

The elementary row matrices are all invertible, and the inverses are also elementary row matrices. For $r \neq 0$, we have

$$\begin{aligned} E_{ij}^{-1} &= E_{ij}, \\ E_i(r)^{-1} &= E_i(1/r), \text{ and} \\ E_{ij}(r)^{-1} &= E_{ij}(-r). \end{aligned}$$

8.37 Elementary Column Matrices

So what about the elementary column operations. In equation (8.17.2), we post-multiplied by $\mathbf{E}_{23}$ and swapped the 2^{nd} and 3^{rd} columns.

Let $\mathbf{A}$ be a square matrix. Because $\mathbf{E}_{ij}$ is symmetric,

$$(\mathbf{A}\mathbf{E}_{ij})^{\mathsf{T}} = \mathbf{E}_{ij}\mathbf{A}^{\mathsf{T}} \qquad \text{or} \qquad \mathbf{A}\mathbf{E}_{ij} = (\mathbf{E}_{ij}\mathbf{A}^{\mathsf{T}})^{\mathsf{T}}.$$

Rows i and j of $\mathbf{A}^{\mathsf{T}}$ are swapped in the transpose of $\mathbf{A}\mathbf{E}_{ij}$, so the columns i and j of $\mathbf{A}$ must be swapped in $\mathbf{A}\mathbf{E}_{ij}$ itself. This means that post-multiplying by $\mathbf{E}_{ij}$ swaps columns i and j.

The same argument applies to $\mathbf{E}_i(r)$, showing that post-multiplying by $\mathbf{E}_i(r)$ multiplies column i by r.

As for $\mathbf{E}_{ij}(r)$, this argument shows that post-multiplying by $\mathbf{E}_{ij}(r)^{\mathsf{T}} \neq \mathbf{E}_{ij}(r)$ adds r times column i to column j. In other words, the family of elementary column matrices consists of matrices of the form $\mathbf{E}_{ij}$, $\mathbf{E}_i(r)$, and $\mathbf{E}_{ij}(r)^{\mathsf{T}}$.

8.38 Matrix Squares and Square Roots

The matrices $\mathbf{E}_{ij}$ have a particularly interesting property. If we square them, we get the identity matrix.

$$\mathbf{E}_{ij}^2 = \mathbf{I}.$$

We can think of the matrices $\mathbf{E}_{ij}$ as square roots of the identity matrix. They are not the only non-trivial square roots. The matrix

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

is also a square root of the identity. Matrices are quite different from real numbers in this respect as 1 has only two square roots.

We also don't need imaginary numbers to find square roots of $-\mathbf{I}$. One example is skew-symmetric.

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \times \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = -\mathbf{I}.$$

8.39 Row Operations and Inversion

Suppose that $\mathbf{A}$ is a non-singular matrix. Such a matrix can be row-reduced to the identity matrix (Lemma 7.33.1). That means that there are elementary matrices $\mathbf{E}_1, \ldots, \mathbf{E}_k$ with $(\mathbf{E}_k \cdots \mathbf{E}_2\mathbf{E}_1)\mathbf{A} = \mathbf{I}$. Then the inverse of $\mathbf{A}$ can be expressed as a product of elementary matrices $\mathbf{E}_k \cdots \mathbf{E}_1$. Since each $\mathbf{E}_i$ is invertible, so is their product, which is $\mathbf{A}^{-1}$.

Combined with Theorem 8.23.1, we have proven that non-singularity and invertibility are the same.

Theorem 8.39.1. *An $n \times n$ matrix $\mathbf{A}$ is non-singular if and only if it is invertible.*

This also gives us a method for finding the inverse. Consider the matrix

$$\big(\mathbf{A} \mid \mathbf{I}\big).$$

We row-reduce this by pre-multiplying by $\mathbf{A}^{-1} = \mathbf{E}_k \cdots \mathbf{E}_1$. What we get is

$$\big(\mathbf{A}^{-1}\mathbf{A} \mid \mathbf{A}^{-1}\big) = \big(\mathbf{I} \mid \mathbf{A}^{-1}\big).$$

In other words, by row-reducing

$$\big(\mathbf{A} \mid \mathbf{I}\big),$$

we obtain the inverse of $\mathbf{A}$ in the right-hand portion of the row-reduced matrix.

8.40 Invertible Matrices: Product of Elementary Matrices

It follows that any invertible matrix can be written as the product of elementary matrices.

Theorem 8.40.1. *Let A be an $n \times n$ invertible matrix. Then there are elementary matrices $F_1, \ldots, F_k$ with $A = F_1 F_2 \cdots F_k$.*

Proof. Using the notation above, we have $A^{-1} = E_k \cdots E_1$. Take the inverse. Since the inverse of any of the elementary matrix is also an elementary matrix, we may set $F_i = E_i^{-1}$. ∎

8.41 Input-Output Systems Revisited

Earlier, we examined input-output systems. Suppose we have an input-output model without labor and that the input coefficient matrix is $n \times n$. Let $\mathbf{c}$ be the desired consumption vector. Given outputs $\mathbf{x}$, the required input is $\mathbf{A}\mathbf{x}$. For this to work, we must have $\mathbf{c} + \mathbf{A}\mathbf{x} = \mathbf{x}$. In other words, $\mathbf{c}$ must solve $\mathbf{c} = (\mathbf{I} - \mathbf{A})\mathbf{x}$ or $\mathbf{x} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{c}$. The inputs must be non-negative for this to be *feasible*. When does that happen?

To make things commensurate, we will measure inputs and outputs by their dollar values, and write the input coefficient matrix so that it shows the dollar cost of inputs for one dollar's worth of output. We will expect that the dollar value of output exceeds the dollar value of input (firms can make profits).

That means that a_{ij} is the cost of i used in the production of one dollar's worth of j. Then

$$\sum_{i=1}^{n} a_{ij} = \text{ cost to produce \$1 worth of } j$$

and that the positive profit condition is

$$\sum_{i=1}^{n} a_{ij} < 1 \text{ for every } j. \tag{8.41.3}$$

8.42 Theorem on Input-Output Systems

A matrix obeying equation (8.41.3), such as $\mathbf{I} - \mathbf{A}$, is said to have a *dominant diagonal*. The diagonal elements of $\mathbf{I} - \mathbf{A}$ are $1 - a_{jj}$ while the off-diagonal elements are $-a_{ij}$. It follows that the absolute sum of the off-diagonal elements of column j is $\sum_{i \neq j} a_{ij}$. By equation (8.41.3), this is less than absolute value of the diagonal element in column j, $1 - a_{jj}$. Matrices with dominant diagonals have some nice properties.

Theorem 8.42.1. *If each $a_{ij} \geq 0$ and for every j, $\sum_{i=1}^{n} a_{ij} < 1$, then $(\mathbf{I} - \mathbf{A})^{-1}$ exists and each entry is non-negative.*

Proof. We will not do this in class. The proof is in section 8.5 of Simon and Blume. ∎

There is a corollary, which provides an answer to the question of when inputs are non-negative.

Corollary 8.42.2. *Under the conditions of Theorem 8.42.1, for all non-negative $\mathbf{c}$, $x = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{c}$ is non-negative.*

Proof. Since each element of $(\mathbf{I} - \mathbf{A})^{-1}$ is non-negative, the matrix product shows that each x_i is the sum of non-negative numbers. ∎

The corollary tells us that any non-negative consumption vector is feasible in this input-output model under the positive profit condition: $\sum_i a_{ij} < 1$ for every j.

8.43 Summary of Matrix Algebra

For conformable matrices:

$(\mathbf{A}+\mathbf{B})+\mathbf{C}=\mathbf{A}+(\mathbf{B}+\mathbf{C})$	addition associates
$\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$	addition commutes
$0\mathbf{A}=\mathbf{0}$	0 and additive identity
$(-1)\mathbf{A}=-\mathbf{A}$	-1 and additive inverse
$\mathbf{0}+\mathbf{A}=\mathbf{A}+\mathbf{0}=\mathbf{A}$	additive identity
$\mathbf{A}+(-\mathbf{A})=(-\mathbf{A})+\mathbf{A}=\mathbf{0}$	additive inverse
$\mathbf{A}\times(\mathbf{B}\times\mathbf{C})=(\mathbf{A}\times\mathbf{B})\times\mathbf{C}$	multiplication associates
$\alpha(\mathbf{A}\times\mathbf{B})=(\alpha\mathbf{A})\times\mathbf{B}=\mathbf{A}\times(\alpha\mathbf{B})$	scalars and matrix multiplication
$\alpha(\mathbf{A}+\mathbf{B})=\alpha\mathbf{A}+\alpha\mathbf{B}$	scalar distributive law I
$(\alpha+\beta)\mathbf{A}=\alpha\mathbf{A}+\beta\mathbf{A}$	scalar distributive law II
$\alpha(\mathbf{A}\times\mathbf{B})=(\alpha\mathbf{A})\times\mathbf{B}=\mathbf{A}\times(\alpha\mathbf{B})$	scalar associative law
$(\mathbf{A}+\mathbf{B})\times\mathbf{C}=\mathbf{A}\times\mathbf{C}+\mathbf{B}\times\mathbf{C}$	matrix distributive law I
$\mathbf{A}\times(\mathbf{B}+\mathbf{C})=\mathbf{A}\times\mathbf{B}+\mathbf{A}\times\mathbf{C}$	matrix distributive law II
$\mathbf{I}_m\times\mathbf{A}=\mathbf{A}$ and $\mathbf{A}\times\mathbf{I}_n=\mathbf{A}$	multiplicative identities, $\mathbf{A}$ is $m\times n$
$\mathbf{A}\times\mathbf{A}^{-1}=\mathbf{A}^{-1}\times\mathbf{A}=\mathbf{I}$	multiplicative inverse
$(\alpha\mathbf{A})^{-1}=\alpha^{-1}\mathbf{A}^{-1}$	inverse of scalar multiple
$(\mathbf{AB})^{-1}=\mathbf{B}^{-1}\mathbf{A}^{-1}$	inverse of matrix product
$(\mathbf{A}+\mathbf{B})^\mathsf{T}=\mathbf{A}^\mathsf{T}+\mathbf{B}^\mathsf{T}$	transpose of sum
$(\alpha\mathbf{A})^\mathsf{T}=\alpha\mathbf{A}^\mathsf{T}$	transpose of scalar multiple
$(\mathbf{AB})^\mathsf{T}=\mathbf{B}^\mathsf{T}\mathbf{A}^\mathsf{T}$	transpose of matrix product
$(\mathbf{A}+\mathbf{B})^*=\mathbf{A}^*+\mathbf{B}^*$	conjugate of sum
$(\alpha\mathbf{A})^*=\bar{\alpha}\mathbf{A}^*$	conjugate of scalar multiple
$(\mathbf{AB})^*=\mathbf{B}^*\mathbf{A}^*$	conjugate of matrix product

9. Determinants

Determinants are functions defined on the set of square matrices. They take values in the same field used to define the matrices (real, complex, rational numbers, etc.)

Let $\mathbf{A}$ be an $n \times n$ matrix. We will inductively define the *determinant of* $\mathbf{A}$, $\det \mathbf{A}$. If $n = 1$,

$$\mathbf{A} = (a_{11}) \qquad \text{and} \qquad \det \mathbf{A} = a_{11}.$$

If we have defined determinants for matrices up to size $n-1$, we define the determinant for an $n \times n$ matrix $\mathbf{A}$ in terms of those lesser determinants by

$$\det \mathbf{A} = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n} = \sum_{i=1}^{n} a_{ij}C_{ij}$$

where the ij-*cofactor* of a_{ij} is

$$C_{ij} = (-1)^{i+j} \det \mathbf{A}_{ij} = (-1)^{i+j} M_{ij}.$$

The matrix $\mathbf{A}_{ij}$ is the $(n-1) \times (n-1)$ submatrix of $\mathbf{A}$ formed by removing row i and column j. The number $M_{ij} = \det \mathbf{A}_{ij}$ is referred to as the ij-*minor*.

To see how submatrices work, if

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \text{ then } \mathbf{A}_{23} = \begin{pmatrix} 1 & 2 \\ 7 & 8 \end{pmatrix}.$$

9.1 Determinants and Cofactors

Another notation for the determinant is to replace the matrix parentheses or brackets by vertical bars:

$$\det \boldsymbol{A} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}.$$

We state one result without proof.

Determinant Fact. The determinant can be calculated by expanding by cofactors along any row or any column. But you must use the same row or column for the entire calculation.

9.2 Determinants of Diagonal Matrices

It's easy to calculate the determinant of a diagonal matrix directly from the definition.

Theorem 9.2.1. *Let* $\mathbf{D}_n = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ *be an* $n \times n$ *diagonal matrix. Its determinant is* $\det \mathbf{D}_n = \lambda_1 \lambda_2 \cdots \lambda_n$.

Proof. We prove this by induction on the size of the matrix. It is true for $n = 1$, as

$$\det \mathbf{D}_1 = \det(a_{11}) = a_{11} = \lambda_1.$$

Now suppose it is true for $n \times n$ matrices. Then we expand an $(n+1) \times (n+1)$ matrix $\mathbf{D}_{n+1}$ along its top row:

$$\begin{aligned}
\det D_{n+1} &= \lambda_1 C_{11} + 0C_{12} + \cdots + 0C_{1,n+1} \\
&= \lambda_1 C_{11} \\
&= \lambda_1 (-1)^{1+1} \det \boldsymbol{A}_{11} \\
&= \lambda_1 \big(\det \operatorname{diag} \lambda_2, \ldots, \lambda_{n+1} \big) \\
&= \lambda_1 \lambda_2 \cdots \lambda_{n+1}.
\end{aligned}$$

The third line uses the definition of C_{11}, and the induction hypothesis was used in line four. This shows that the result is true for $(n+1)$ if it is true for n. Since we already showed it was true for $n = 1$, follows that it is true for every $n = 1, 2, \ldots$. ∎

9.3 Determinants Do Not Add

Although it may happen in some special cases, determinants generally do not add. That is, the usual case is that $\det \mathbf{A} + \det \mathbf{B} \neq \det(\mathbf{A} + \mathbf{B})$.

For example

$$0 = \begin{vmatrix} 1 & 0 \\ 0 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 \\ 0 & 1 \end{vmatrix} \neq \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1.$$

There are some cases where they do add. Here's one where both sides are zero.

$$0 = \begin{vmatrix} 1 & 0 \\ 0 & 0 \end{vmatrix} + \begin{vmatrix} 1 & 0 \\ 0 & 0 \end{vmatrix} = \begin{vmatrix} 2 & 0 \\ 0 & 0 \end{vmatrix} = 0.$$

Examples where they do add and both sides are of the equation are zero are much easier to create than those where both sides are not zero.

Since we have a formula for the determinant of diagonal matrices, we can investigate that case a little more closely

When matrices are diagonal, with $\mathbf{A} = \operatorname{diag}(a_1, \ldots, a_n)$ and $\mathbf{B} = \operatorname{diag}(b_1, \ldots, b_n)$, the condition for additivity of the determinant is

$$\prod_{i=1}^{n} a_i + \prod_{i=1}^{n} b_i = \prod_{i=1}^{n} (a + b)_i$$

This usually fails even in the 2×2 case, where it requires $a_1 b_2 + a_2 b_1 = 0$. A non-trivial example that works is

$$\begin{vmatrix} -2 & 0 \\ 0 & 2 \end{vmatrix} + \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = -3 = \begin{vmatrix} -1 & 0 \\ 0 & 3 \end{vmatrix}$$

9.4 Triangular Matrices

A matrix $\mathbf{A}$ is an *upper triangular matrix* if $a_{ij} = 0$ whenever $i > j$. An upper triangular matrix looks like this.

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} & a_{1n} \\ 0 & a_{22} & \cdots & a_{2,n-1} & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{n-1,n-1} & a_{n-1,n} \\ 0 & 0 & \cdots & 0 & a_{nn} \end{pmatrix}.$$

All the elements below the main diagonal are zero in an upper triangular matrix. A lower triangular matrix is the opposite, everything above the main diagonal is zero, so $\mathbf{A}$ is a *lower triangular matrix* if $a_{ij} = 0$ whenever $i < j$. A lower triangular matrix looks like this.

$$\mathbf{A} = \begin{pmatrix} a_{11} & 0 & \cdots & 0 & 0 \\ a_{21} & a_{22} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n-1,1} & a_{n-1,2} & \cdots & a_{n-1,n-1} & 0 \\ a_{n1} & a_{n2} & \cdots & a_{n,n-1} & a_{nn} \end{pmatrix}.$$

Theorem 9.4.1. *If an $n \times n$ matrix $\mathbf{A}$ is either an upper or lower triangular matrix, then* $\det \mathbf{A} = a_{11}a_{22}\cdots a_{nn}$.

Proof. For a lower triangular matrix, we repeat the proof of Theorem 9.2.1. The upper triangular case is a similar induction, except we expand along the first column rather rather than the first row. ∎

9.5 2×2 Determinants

You might be thinking this is easy after computing determinants for diagonal and triangular matrices. That's because we're starting with the easy ones.

The determinant of a 2×2 matrix is still easy, but includes something besides the diagonal terms. Suppose $\mathbf{A}$ is a 2×2 matrix. Then

$$\det \mathbf{A} = a_{11}C_{11} + a_{12}C_{12}.$$

Now $\mathbf{A}_{11} = (a_{22})$ and $\mathbf{A}_{12} = (a_{21})$. Using the formula for size one determinants, we find $\det \mathbf{A}_{11} = a_{22}$ and $\det \mathbf{A}_{12} = a_{21}$. The cofactors are then $C_{11} = a_{22}$ and $C_{12} = -a_{21}$. Then we have

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

9.6 More on 2×2 Determinants

One way to remember this is the following: We multiply the numbers on the main diagonal (NW to SE) and subtract the product of the numbers on the anti-diagonal (SW to NE).

For example,

$$\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = 1(4) - 2(3) = -2,$$

and

$$\begin{vmatrix} 15 & 3 \\ 7 & 12 \end{vmatrix} = 15(12) - 3(7) = 180 - 21 = 159.$$

Determinants can also be zero.

$$\begin{vmatrix} 1 & 2 \\ 2 & 4 \end{vmatrix} = 4 - 4 = 0.$$

When we get to the 3×3 case, we'll start to see the general pattern. But first, we establish some general results with the current definition.

9.7 Alternating Functions

If we interchange any two rows of a matrix, it flips the sign of the determinant. This can only happen if $n \geq 2$. This result is important because it tells us how one of the elementary row (column) operations affects the determinant.

Before proving this, we introduce some more terminology.

Alternating. A function $f(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is *alternating* if whenever we interchange two of the $\mathbf{x}_i$, f is multiplied by (-1), flipping the sign.

We can regard the determinant as a function of the rows (or columns) of an $n \times n$ matrix. As such, Theorem 9.8.1 tells us that determinants are alternating, both with respect to row interchange and column interchange.

9.8 Determinants are Alternating I

Theorem 9.8.1. *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix with* $n \geq 2$. *Form* $\mathbf{B}$ *from* $\mathbf{A}$ *by interchanging any two rows or columns of* $\mathbf{A}$. *Then* $\det \mathbf{B} = -\det \mathbf{A}$.

Proof. We prove this by induction, starting when $n = 2$, When $n = 2$, we use the formula for determinants of size two.

$$\det \mathbf{B} = \begin{vmatrix} a_{21} & a_{22} \\ a_{11} & a_{12} \end{vmatrix} = a_{21}a_{12} - a_{11}a_{22} = -\det \mathbf{A}.$$

When $n \geq 3$, we have a little more room. We will interchange rows h and k of $\mathbf{A}$. We use induction on the size of the matrix. Suppose the result is true for matrices of size n with $n \geq 2$.

Proof Continues ...

9.9 Determinants are Alternating II

Remainder of Proof. For the induction step, we consider matrices of size $n + 1 \geq 3$ and expand the determinant along a row that is not part of the interchange. That is, we use row $i \neq h, k$. Then

$$\begin{aligned}
\det \mathbf{B} &= a_{i1}C'_{i1} + a_{i2}C'_{i2} + \cdots + a_{i,n+1}C'_{i,n+1} \\
&= \sum_{j=1}^{n+1} a_{ij}C'_{ij} \\
&= -\sum_{j=1}^{n+1} a_{ij}C_{ij} \\
&= -\det \mathbf{A}.
\end{aligned}$$

Here C'_{ij} are the cofactors in $\mathbf{B}$. The induction hypothesis is used to get from the second to third row as $C'_{ij} = (-1)^{i+j} \det \mathbf{B}_{ij} = -(-1)^{i+j} \det \mathbf{A}_{ij} = -C_{ij}$. This because the two rows h and k are in each submatrix $\mathbf{A}_{ij}$. They are interchanged in each $\mathbf{A}_{ij}$ to get each of the $\mathbf{B}_{ij}$, reversing the sign by the induction hypothesis. Then we put the determinant back together in the last line to finish the induction step. It follows that the result is true for $n = 2, 3, \ldots$.

The column case is the same, but expanded along an uninvolved column. ∎

9.10 Determinants with Repetition

When a row or column is repeated, the determinant is zero. We already did the main part of the work for this in Theorem 9.8.1.

Theorem 9.10.1. *Suppose* $\mathbf{A}$ *is an* $n \times n$ *matrix with* $n \geq 2$. *If either a row or column is repeated, then* $\det \mathbf{A} = 0$.

Proof. Let i and j be the repeated rows. If we interchange rows i and j, we still have matrix $\mathbf{A}$. But by Theorem 9.8.1, $\det \mathbf{A} = -\det \mathbf{A}$. Then $2 \det \mathbf{A} = 0$, so $\det \mathbf{A} = 0$.

9.11 Determinants of Size 3

Now that we have the determinants of size two under control, we can proceed to size three.

$$\det \boldsymbol{A} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13}.$$

We now use the formula for the size two determinants to find

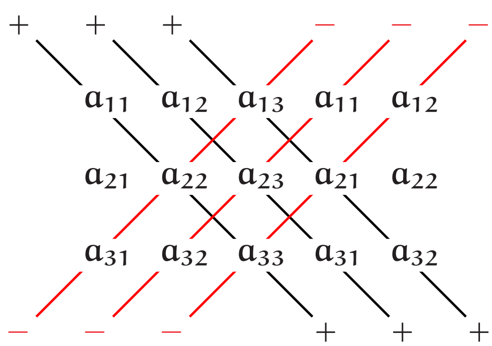
$$\begin{aligned}\det \boldsymbol{A} &= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{31}a_{23}) \\ &\quad + a_{13}(a_{21}a_{32} - a_{31}a_{22}) \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} \\ &\quad + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}\end{aligned}$$

9.12 Computing 3×3 Determinants: Sarrus Rule

A way to remember 3×3 determinants is to use the *Sarrus Rule*.[1] Start by repeating the first two columns, yielding a 3×5 array.

$$\begin{array}{ccccc} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{array}$$

Then attach plus signs to the first 3 diagonals and minus signs to the first three anti-diagonals.



Compute the determinant by multiplying the coefficients along each of the black lines marked with a plus and adding the three products, then multiply the coefficients along each of the red lines marked with a minus and subtract the three products. The result is the determinant.

[1] Pierre Frédéric Sarrus (1798–1861) was a French mathematican who wrote several treatises. He's perhaps best known for the Sarrus Rule, used for computing 3×3 determinants and for Sarrus numbers. Although not mathematics, he also developed the first mechanical linkage that could transform rotary motion into perfect linear motion.

9.13 Another View of Determinants

One way to think about the 3×3 determinant

$$
\begin{aligned}
\det \mathbf{A} = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} \\
- a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} \\
+ a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}
\end{aligned}
$$

is to notice that each of the six terms is composed of elements from every row and every column. The first product, $a_{11}a_{22}a_{33}$ uses row 1 and column 1, then row 2 and column 2, and finally row 3 and column 3. The second product, $a_{11}a_{23}a_{32}$ again uses row 1 and column 1, then row 2 and column 3, and finally row 3 and column 2. Each row is used once, each column is used once.

As for the signs, the plus sign is applied when the column numbers are in the same order as the row numbers such as '123' and '123'. The minus appears when there is a reversal such as '123' and '132'. The same thing happens in the next pair where '123' is matched with '213', which gets a negative sign. while '123' is matched with '231' with a plus sign. In case there two switches of adjacent elements, to '213' and then to '231' with the two minus signs canceling. Finally, in the third pair '123' goes with '312' where two switches, to '132' and then '312' resulting in a plus sign. The last term takes '123' to '321' (one more switch) and so a minus sign.

9.14 Paths Through the Matrix

What's happening here is we are taking all possible paths from the top to the bottom of the matrix (or left side to the right side) where each product takes elements from each row and column. For each product, we assign the sign based on whether we have an even number of interchanges in the indices (positive), or an odd number (negative).

This also works on the 2×2 determinant. Then the rows are '12'. The positive sign is applied when the columns go in the same order, '12'. The negative is applies when the columns are in the opposite order, '21'. The determinant is then $a_{11}a_{22} - a_{12}a_{21}$.

We generalize this to every size n by writing it in terms of permutations.

9.15 Determinants via Permutations

We can define determinants using permutations of the indices.

Permutation. We say a function $\sigma\colon \{1,\dots,n\} \to \{1,\dots,n\}$ is a *permutation* if σ takes each value $\{1,\dots,n\}$ exactly once.

In other words, '$\sigma(1)\dots\sigma(n)$' is a rearrangement of '$12\dots n$'. The sign of a permutation, $\operatorname{sgn}\sigma$, is $+1$ when an even number of interchanges of adjacent elements of '$12\dots n$' yield '$\sigma(1)\dots\sigma(n)$'. The sign is -1 when an odd number of interchanges is involved. Let P_n denote the set of permutations of '$12\dots n$'. There are $n!$ permutations of '$12\dots n$'.[2]

[2] The order of the numbers matters for permutations. '12' and '21' are different permutations. If we only care about which numbers are involved, but not their order, we refer to *combinations*. Thus '124' and '241' represent the same combination of numbers.

9.16 Determinant Formula using Permutations

We can now write the determinant as

$$\det \boldsymbol{A} = \sum_{\sigma \in P_n} (\operatorname{sgn} \sigma) \left(\prod_{i=1}^{n} a_{i\sigma(i)} \right).$$

This is what we just described on the previous page. Each element in $\prod_{i=1}^{n} a_{i\sigma(i)}$ is from a different row ($i = 1, \ldots n$) and a different column ($\sigma(1), \ldots, \sigma(n)$). The sign of of each product is determined by the number of interchanges in the permutation $\sigma(i)$.

When $n = 3$ there are 6 permutations to consider: '123', '132', '312', '321', '231', and '213'. Since each is created by a single interchange from the previous permutation, the signs alternate. For a 3×3 matrix $\boldsymbol{A}$ the formula yields

$$\begin{aligned} \det \boldsymbol{A} = & a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{13}a_{21}a_{32} \\ & - a_{13}a_{22}a_{31} + a_{12}a_{23}a_{31} - a_{12}a_{21}a_{33} \end{aligned}$$

which is the previously calculated value.

13. Functions of Several Variables

Before continuing with determinants, it will be helpful to upgrade the language we use to describe mappings.[1] We need to be able to discuss functions in more detail.

13.1 The Language of Functions

Suppose X and Y are sets. A *function* from X to Y is a rule that assigns an element of Y to every element of X. We write $f\colon X \to Y$ or

$$X \xrightarrow{f} Y$$

to indicate f is a function from X to Y. Here X is referred to as the *domain of* f, $X = \operatorname{dom} f$, and Y is the *target space of* f.

Given $\mathbf{x} \in X$, $f(\mathbf{x})$ denotes the element of Y that f assigns to $\mathbf{x}$. We sometimes write $\mathbf{x} \mapsto f(\mathbf{x})$ to indicate that f assigns $f(\mathbf{x})$ to $\mathbf{x}$. Every element of X is mapped to exactly one element of Y. However, multiple elements of X can map to the same element of Y.[2]

Given a set X, its *power set* is $\mathcal{P}(X) = \{\text{subsets of } X\}$. We can also use a function defined on X on subsets of X If $A \subset X$, the *image* of A under f is

$$f(A) = \big\{f(\mathbf{x}) : \mathbf{x} \in A\big\}.$$

Here we regard the function as a map between power sets, $f\colon \mathcal{P}(X) \to \mathcal{P}(Y)$. Of course $f(A) \subset Y$. The image of the domain X is referred to as the *range*, $\operatorname{ran} f = f(X)$. The range is contained in the target space, $f(X) \subset Y$, but need not coincide with it.

[1] See also Chapter 13 of Simon and Blume.

[2] Function-like mappings that map to non-empty sets of values are called *correspondences*. We will not study them at this time.

13.2 Surjective (Onto) Functions

If the range does coincide with the target space, we say that the function f is *onto* or *surjective*.[3]

Onto = Surjective. We say that a function f is *onto* or *surjective* if the range of f is the entire target space, (i.e., $f(X) \equiv \operatorname{ran} f = Y$).

► **Example 13.2.1: Two Functions.** For example, suppose $f\colon \mathbb{R} \to \mathbb{R}$ is defined by $f(x) = x^2 + 2$. If $A = [0, 2]$, $f(A) = [2, 6]$. The range of f is the interval $[2, +\infty)$. As this is smaller than the target space $\mathbb{R}$, f is not onto.

The function $g\colon \mathbb{R} \to \mathbb{R}$ defined by $g(x) = x^3$ has $\operatorname{ran} g = \mathbb{R}$ because if $y \in \mathbb{R}$, $y = g\big(y^{1/3}\big)$ and $y^{1/3} \in \operatorname{dom} g$. ◄

[3] Why two terms? One-to-one and onto are older English terms. Injective, surjective, and bijective were popularized by the (mostly French) Nicolas Bouraki group as well as some prominent American mathematicians in the 1950's. The prominent algebraist Saunders Mac Lane seems to have had a particular dislike for one-to-one and onto.

13.3 Injective (One-to-One) Functions

One-to-One = Injective. A function f is *one-to-one* or *injective* if $f(\mathbf{x}) = f(\mathbf{x}')$ implies that $\mathbf{x} = \mathbf{x}'$.

▶ **Example 13.3.1: Matrix Functions.** When $\mathbf{A}$ is an $m \times n$ matrix, the function $f\colon \mathbb{R}^n \to \mathbb{R}^m$ defined by $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$ is onto $\mathbb{R}^m$ if and only if $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b}$ in $\mathbb{R}^n$. It follows that $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$ is onto $\mathbb{R}^m$ if and only if $\operatorname{rank}\mathbf{A} = m$ by Corollary 7.30.1.

Now f is one-to-one if and only if $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x}'$ implies $\mathbf{x} = \mathbf{x}'$. That is, if and only if $\mathbf{A}(\mathbf{x} - \mathbf{x}') = \mathbf{0}$ implies $\mathbf{x} - \mathbf{x}' = \mathbf{0}$. The function f will be one-to-one if and only if $\mathbf{A}\mathbf{x} = \mathbf{0}$ has only one solution, $\mathbf{x} = \mathbf{0}$. Corollary 7.29.1 tells us that f is one-to-one if and only if $n = \operatorname{rank}\mathbf{A}$.

It follows that $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$ is both one-to-one and onto if and only if $\operatorname{rank}\mathbf{A} = n = m$, which is equivalent to $\mathbf{A}$ being invertible. ◀

13.4 Bijective Functions

Bijective. If f is both one-to-one and onto (injective and surjective), we call it *bijective*.

Theorem 13.4.1. *If* $f\colon X \to Y$ *is bijective, for each* $y \in Y$*, there is a unique* $x(y) \in X$ *with* $f\big(x(y)\big) = y$.

Proof. Since f is onto, there is an $x(y) \in X$ that is mapped back to y. That is, with $f\big(x(y)\big) = y$. Since f is one-to-one, that $x(y)$ is unique. ∎

We call the function $y \mapsto x(y)$ the *inverse of* f and denote it by f^{-1}. Thus $f^{-1}\colon Y \to X$ and $f\big(f^{-1}(y)\big) = y$. Also, $f^{-1}\big(f(x)\big) = x$, since $f(x)$ is the unique element of Y that is the image of the point $\mathbf{x} \in X$.

13.5 Linear Functions

Linear functions are those that preserve the two basic linear operations: vector addition and scalar multiplication.

Linear Transformation. Let $f\colon \mathbb{R}^n \to \mathbb{R}^m$. The function $f\colon \mathbb{R}^n \to \mathbb{R}$ is a *linear function* or *linear transformation* if for every $\alpha \in \mathbb{R}$ and every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

1. $f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$ and
2. $f(\alpha\mathbf{x}) = \alpha f(\mathbf{x})$.

Setting $\mathbf{x} = \mathbf{y} = \mathbf{0}$, condition (1) implies $f(\mathbf{0}) = 0$ for any linear function.

The two criteria for linearity can be combined as

$$f(\alpha\mathbf{x} + \mathbf{y}) = \alpha f(\mathbf{x}) + f(\mathbf{y}) \tag{13.5.1}$$

for all scalars α and vectors $\mathbf{x}$ and $\mathbf{y}$. It's pretty obvious that the two linearity conditions imply equation (13.5.1).

To see that equation (13.5.1) implies both conditions, set $\alpha = 1$, which implies $f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$. This implies $f(\mathbf{0}) = \mathbf{0}$ as above. Finally, setting $\mathbf{y} = \mathbf{0}$ yields the second condition $f(\alpha\mathbf{x}) = \alpha f(\mathbf{x})$.

13.6 Any Matrix Defines a Linear Function

The transformation $T_{\mathbf{A}}$. Given an $m \times n$ matrix $\mathbf{A}$, define the function $T_{\mathbf{A}} \colon \mathbb{R}^n \to \mathbb{R}^m$ by $T_{\mathbf{A}}(\mathbf{x}) = \mathbf{A}\mathbf{x}$.

Theorem 13.6.1. *Let $\mathbf{A}$ be an $m \times n$ matrix. The function $T_{\mathbf{A}} \colon \mathbb{R}^n \to \mathbb{R}^m$ defined by $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$ is a linear function.*

Proof. That this is a linear function follows from the rules of matrix algebra. Let $\alpha \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then

$$
\begin{aligned}
T_{\mathbf{A}}(\alpha\mathbf{x} + \mathbf{y}) &= \mathbf{A}(\alpha\mathbf{x} + \mathbf{y}) \\
&= \mathbf{A}(\alpha\mathbf{x}) + \mathbf{A}\mathbf{y} = \alpha(\mathbf{A}\mathbf{x}) + \mathbf{A}\mathbf{y} \\
&= \alpha T_{\mathbf{A}}(\mathbf{x}) + T_{\mathbf{A}}(\mathbf{y})
\end{aligned}
$$

showing that $T_{\mathbf{A}}$ is linear. ∎

We will see in Theorem 10.6.1 that all linear transformations $T \colon \mathbb{R}^n \to \mathbb{R}^m$ can be written $T = T_{\mathbf{A}}$ for some $m \times n$ matrix $\mathbf{A}$.

13.7 Multilinear Functions

We can write $\mathbb{R}^{nk} = \overbrace{\mathbb{R}^n \times \cdots \mathbb{R}^n}^{\text{k times}}$. We write elements of $\mathbb{R}^{nk}$ in the form $(\mathbf{x}_1, \dots, \mathbf{x}_k)$ with each $\mathbf{x}_i \in \mathbb{R}^n$.

Multilinearity. A function $\overbrace{\mathbb{R}^n \times \cdots \mathbb{R}^n}^{\text{k times}} \to \mathbb{R}$ is called k-*linear* if it is separately linear on each of the k copies of $\mathbb{R}^n$. The term *multilinear* is used in the generic case, and *bilinear* is used when f is 2-linear.

A variety of multilinear objects are generically referred to as *tensors*. A k-multilinear function is a k-tensor. You'll notice this is a little vague. Being limited to $\mathbb{R}^n$ is also rather constraining, but reasonable for an introduction to tensors.[4]

[4] So far as I know, a definition of what exactly a tensor is or is not that is both precise and general not does not exist. One influential attempts was that of H. Whitney (1938) Tensor Products of Abelian Groups, *Duke Math. J.* **4**, 495–528.

Hassler Whitney (1907–1989) was an American mathematician known for his work on manifolds, particularly embedding of manifolds, characteristic classes, and geometric integration, including extensions of Stokes' Theorem. His unusual first name was his maternal grandmother's maiden name. California's Mount Whitney is named for his uncle, Josiah Whitney.

His work on graph theory in the 1930's formed the basis for the later computer proof of the Four Color Theorem. He worked on extension of smooth functions to larger domains. He also showed that sufficiently smooth n-manifolds as defined intrinsically by Veblen and Whitehead (see section 15.22) could be embedded as submanifolds of $\mathbb{R}^{2n+1}$. He also developed the extrinsic definition of manifolds, defining them as subsets of some $\mathbb{R}^n$. The 1954 Nash Embedding Theorem (yes, that John Nash) reduced the smoothness required.

13.8 Bilinear Forms

We can use any $n \times n$ matrix $\mathbf{A} = [a_{ij}]$ to define a bilinear function from $\mathbb{R}^{2n}$ to $\mathbb{R}$. Such functions are called *bilinear forms* or *quadratic forms*.[5] Set

$$f(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i y_j = \mathbf{x}^T \mathbf{A} \mathbf{y}.$$

Then f is bilinear. We'll show that it is linear in the second coordinate using matrix notation.

$$\begin{aligned} f(\mathbf{x}, \alpha\mathbf{y} + \mathbf{z}) &= \mathbf{x}^T \mathbf{A}(\alpha\mathbf{y} + \mathbf{z}) \\ &= \mathbf{x}^T \mathbf{A}(\alpha\mathbf{y}) + \mathbf{x}^T \mathbf{A}\mathbf{z} \\ &= \alpha(\mathbf{x}^T \mathbf{A}\mathbf{y}) + \mathbf{x}^T \mathbf{A}\mathbf{z} \\ &= \alpha f(\mathbf{x}, \mathbf{y}) + f(\mathbf{x}, \mathbf{z}) \end{aligned}$$

The case of the first coordinate is similar.

[5] See section 13.3 of Simon and Blume for the basic definition. We will study them more in Chapter 16.

13.9 Bilinear Forms Using Coordinates 09/06/22

Homework: Problems **8.3, 8.18, 8.29, 9.8**, and **9.13** are due on **Tuesday, September 13**.

a This can also be shown by explicitly using the coordinates of the vectors and elements of the matrix. To do so, we introduce the shorthand that

$$\sum_{ij=1}^{n} \text{ means } \sum_{i=1}^{n}\sum_{j=1}^{n}, \quad \sum_{ijk=1}^{n} \text{ means } \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n},$$

and similarly for larger sets of indices.

This allows us to write bilinear forms as

$$f(\mathbf{x}, \mathbf{y}) = \sum_{ij=1}^{n} a_{ij} x_i y_j.$$

13.10 Tensors I

Just as we can define 2-linear functions, a 4-linear function A can be defined by[6]

$$A(\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}) = \sum_{hijk=1}^{n} a_{hijk} w_h x_i y_j z_k.$$

The function A is linear in each of its four coordinates, in other words, the 4-dimensional array $[a_{hijk}]$ defines a *tensor*, more specifically, a 4-tensor.

We show this for the first coordinate.

$$\begin{aligned} A(\mathbf{w} + \alpha \mathbf{w}', \mathbf{x}, \mathbf{y}, \mathbf{z}) &= \sum_{hijk=1}^{n} a_{hijk}(w_h + \alpha w'_h) x_i y_j z_k \\ &= \sum_{hijk=1}^{n} a_{hiij} w_h x_i y_j z_k + \alpha \sum_{hijk=1}^{n} a_{hiij} w'_h x_i y_j z_k \\ &= \alpha A(\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}) + \alpha A(\mathbf{w}', \mathbf{x}, \mathbf{y}, \mathbf{z}) \end{aligned}$$

This shows that A is linear in the first coordinate. The argument for linearity in the other coordinates is similar.

[6] At this point, we are well beyond Simon and Blume. Since this is related to the previous material, I leave it under the Chapter 13 heading.

13.11 Tensors II

More generally, anything of the form

$$A(\mathbf{x}_1, \ldots, \mathbf{x}_k) = \sum_{j_1 \cdots j_k = 1}^{n} a_{j_1 \cdots j_k} x_{1j_1} \cdots x_{kj_k}$$

is k-linear, giving us a tensor $A = \left[a_{j_1 \cdots j_k}\right]$.

This type of method, involving summation over coordinates is similar to Ricci and Levi-Civita's absolute differential calculus, developed for use in differential geometry.[7] Albert Einstein made it the language of general relativity. However, the notation is **simplified here** by focusing on tensors that are functions solely of ordinary (contravariant) vectors.[8] Einstein paid particular attention to the geometry, which he allowed to vary by location. Our problems are simpler because any geometry they involve is Euclidean.

Modern approaches to tensors emphasize coordinate-free methods. Although this can make many things easier, it can also make understanding more difficult due to the higher level of abstraction.

---

[7] The Ricci calculus or absolute differential calculus consists of rules for using index notation and manipulating it for tensor fields on a differentiable manifold (which can be $\mathbb{R}^m$). It was developed by Ricci and Levi-Civita in the late 19^{th} century.
Gregorio Ricci-Curbastro (1853–1925) was an Italian mathematician best known for inventing tensor calculus. He also worked on the theory of the real numbers.
The Italian mathematician Tullio Levi-Civita (1873–1941) is best known for his work on tensor calculus and its applications to the theory of general relativity. He also worked on teh three-body problem, analytical mechanics, and hydrodynamics.

[8] Albert Einstein (1879–1955) was one of the greatest physicists. In a single year he published pathbreaking papers on the photo-electric effect, Brownian motion, and special relativity. Any one of those would have put him in the top rank of physicists. Over the next decade he working on incorporating gravity and acceleration (with identical effects by the Principle of Equivalence) into special relativity, which transformed it into general relativity. Other important work of his included the theory of heat capacity and Bose-Einstein statistics describing the behavior of particles with integral spin.

9. Determinants Again

We return to determinants after our brief look at general properties of functions. In fact, we are now doing a little of both, because determinants are a special type of function. One of their properties is that they are multilinear.

9.15 Determinants are Row and Column n-functions

Our real reason for being interested in multilinear functions at this moment is that the determinant is multilinear. So what is the determinant a function of? We can treat the determinant as a function of either the rows or the columns of $\mathbf{A}$. For the row case, let $\mathbf{a}_i$ be the i^{th} row of $\mathbf{A}$. Then we can write

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{pmatrix},$$

which lets us think of the determinant as a function of the rows, a n-function $f_n(\mathbf{a}_1, \dots, \mathbf{a}_n) = \det \mathbf{A}$.

Similarly, let $\mathbf{a}^j$ be the j^{th} column of $\mathbf{A}$, to make it a n-function by using the columns.

$$\mathbf{A} = \left(\mathbf{a}^1 \;\middle|\; \mathbf{a}^2 \;\middle|\; \cdots \;\middle|\; \mathbf{a}^n \right).$$

9.16 Multilinearity of Determinants

We now show that determinants are multilinear.

Theorem 9.16.1. *Let $\mathbf{A}$ be an $n \times n$ matrix, and $f_n(\mathbf{a}_1, \dots, \mathbf{a}_n) = \det \mathbf{A}$, where the $\mathbf{a}_i$ are the rows of $\mathbf{A}$. Then f_n is n-linear in $(\mathbf{a}_1, \dots, \mathbf{a}_n)$ for $n \geq 1$. If we regard the determinant as a function of the columns of $\mathbf{A}$, $\mathbf{a}^i$, it is n-linear in the columns.*

Proof. Replace row i by $\mathbf{a}_i + \alpha\mathbf{a}'_i$ for any scalar α and vector $\mathbf{a}'_i \in \mathbb{R}^n$. We now expand the determinant f_n on the row of interest, row i.

$$
\begin{aligned}
f_n(\mathbf{a}_1, \dots, \mathbf{a}_i + \alpha\mathbf{a}'_i, \dots, \mathbf{a}_n)& \\
&= (a_{i1} + \alpha a'_{i1})C_{i1} + \cdots + (a_{in} + \alpha a'_{in})C_{in} \\
&= (a_{i1}C_{i1} + \cdots + a_{in}C_{in}) + \alpha(a'_{i1}C_{i1} + \cdots + a'_{in}C_{in}) \\
&= f_n(\mathbf{a}_1, \dots, \mathbf{a}_n) + \alpha f_n(\mathbf{a}'_1, \dots, \mathbf{a}'_n)
\end{aligned}
$$

where $\mathbf{a}'_j = \mathbf{a}_j$ for $j \neq i$. This shows that f_n is linear separately in each $\mathbf{a}_i$, and so is n-linear.

The proof in terms of columns is basically the same, but expands along the column of interest. ∎

The multilinearity of the determinant means that we know how determinants behave under the third elementary row operation, adding a non-zero multiple of row i to row j. Using multilinearity, we can write the determinant as $\det(\mathbf{a}_1, \dots, \mathbf{a}_n) + \alpha \det(\mathbf{a}'_1, \dots, \mathbf{a}'_n)$ where row i is repeated in the latter. Since det is alternating, the second term is zero and we are left with $\det \mathbf{A}$. In other words, the third elementary row operation leaves the determinant unchanged, $\det \mathbf{E}_{ij}(r)\mathbf{A} = \det \mathbf{A}$.

9.17 Determinants: Yet Another Definition

We are now ready for the third definition of determinants.

The *determinant* on the set of $n \times n$ matrices is defined as the function from the set of $n \times n$ matrices to $\mathbb{R}$ that:

1. Is an alternating function of the rows of the matrices
2. Is n-linear in the rows of the matrices
3. Maps the identity matrix to 1.

It's possible to show there is only one such function. The normalization, so that $D_n(\mathbf{I}_n) = 1$ is required for this.

9.18 The Determinant Theorem

The Determinant Theorem. *Let D_n be a function from the set of $n \times n$ matrices to $\mathbb{R}$. Given a matrix $\mathbf{A}$, we regard $D_n(\mathbf{A})$ as a function of the rows of $\mathbf{A}$, which we denote $\mathbf{a}_1, \ldots, \mathbf{a}_n$. Such a function is the determinant of $\mathbf{A}$ if and only if*

1. D_n *is an alternating function.*
2. D_n *is n-linear.*
3. $D_n(\mathbf{I}_n) = 1$.

Proof. Only if case ($\Rightarrow$): We have proven this in pieces already. There are three relevant theorems using the cofactor definition. Theorem 9.8.1 showed that the determinant as defined by cofactor expansion, is an alternating function. Theorem 9.16.1 showed that it is also multilinear, and Theorem 9.2.1 implies that $\det \mathbf{I} = 1$.

It's not hard to verify these properties also hold for the permutation definition. As a result, conditions (1)–(3) hold for both the co-factor and permutation definitions of the determinant. When we show that the solution to (1)–(3) is unique, we will know that all three definitions yield the same determinant.

If case ($\Leftarrow$): This will follow over the next two pages

Fact. The Determinant Theorem remains true if we require it to be alternating and multilinear in terms of columns rather than rows.

9.19 The Determinant Theorem, II

If (i). To prove the if portion of the determinant theorem, we will examine the effects of the elementary row operations on any alternating multilinear function f_n of the rows of a matrix of size n.

Because f_n is alternating, interchanging rows flips the sign of f_n. Because of multilinearity, multiplying a row by a scalar multiplies f_n by that same scalar.

The third row operation is adding a non-zero multiple of one row to another. Suppose we replace $\mathbf{a}_i$ by $r\mathbf{a}_i + \mathbf{a}_j$. By multilinearity,

$$\begin{aligned} &f_n(\mathbf{a}_1, \dots, r\mathbf{a}_i + \mathbf{a}_j, \dots, \mathbf{a}_n) \\ &\qquad = f_n(\mathbf{a}_1, \dots, \mathbf{a}_n) + rf_n(\mathbf{a}_1, \dots, \mathbf{a}_i, \dots, \mathbf{a}_n) \end{aligned}$$

with the $r\mathbf{a}_i + \mathbf{a}_j$ term in the j^{th} row. The second term in the last line has $\mathbf{a}_i$ in both row i and row j. Because f_n is alternating, that term is zero. Then

$$f_n(\mathbf{a}_1, \dots, \mathbf{a}_i + r\mathbf{a}_j, \dots, \mathbf{a}_n) = f_n(\mathbf{a}_1, \dots, \mathbf{a}_n).$$

This means that the third elementary row operation leaves f_n unchanged.

9.20 The Determinant Theorem, III

If (ii). We can now provide a recipe for finding f_n by row reduction. Let $\mathbf{R}$ be a reduced row-echelon form of an $n \times n$ matrix $\mathbf{A}$. Let m be the number of row interchanges in the reduction and $r_1, \dots, r_k$ be the scalar multiples of rows in the reduction. Then

$$f_n(\mathbf{A}) = (-1)^m \left(\prod_{h=1}^{k} r_h \right) f_n(\mathbf{R}).$$

There two possibilities for $\mathbf{R}$. If $\mathbf{A}$ is non-singular, then $\mathbf{R} = \mathbf{I}_n$ by Corollary 7.32.1 and Lemma 7.33.1. It follows that

$$\begin{aligned} f_n(\mathbf{A}) &= (-1)^m \left(\prod_{h=1}^{k} r_h \right) f_n(\mathbf{I}_n) \\ &= (-1)^m \left(\prod_{h=1}^{k} r_h \right) \end{aligned}$$

since $f_n(\mathbf{I}_n) = 1$.

Otherwise, $\operatorname{rank} \mathbf{A} < n$ and $\mathbf{R}$ will have a zero row. If we multiply that row by zero, $\mathbf{R}$ doesn't change, but f_n is multiplied by zero. This concludes the proof of the if portion of the Determinant Theorem, and so the proof of the whole theorem. ∎

Incidentally, this also shows that the determinant as defined by cofactor expansion is the same as in the Determinant Theorem. Since it is easy to show that the permutation definition is also alternating, multilinear, and takes the value 1 on identity matrices, the permutation method also defines the same determinant.

9.21 Determinants of Transposes

Further consideration of the Determinant Theorem shows that it yields the same determinant if we use columns instead of rows. If we apply the column version to $\mathbf{A}^\mathsf{T}$, that is the same as the row version applied to $\mathbf{A}$, so $\det \mathbf{A} = \det \mathbf{A}^\mathsf{T}$.

The determinant of a transposed matrix is the same as the determinant of the original matrix. We state this as a theorem.

Theorem 9.21.1. *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix. Then* $\det \mathbf{A} = \det \mathbf{A}^\mathsf{T}$.

Proof. See above. ∎

9.22 Determinants of Products

Although we will not prove it, the determinant of the product of two (or more) matrices is the product of the determinants.

Theorem 9.22.1. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be* $n \times n$ *matrices. Then* $\det(\mathbf{AB}) = (\det \mathbf{A})(\det \mathbf{B})$.

Proof. See the proof of Theorem 26.4 in Simon and Blume. ∎

9.23 Determinants of Inverses

We can now show the following theorem on the determinant of inverses.

Theorem 9.23.1. *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix. Then* $\mathbf{A}$ *is invertible if and only if* $\det \mathbf{A} \neq 0$*. In that case,* $\det \mathbf{A}^{-1} = 1/\det \mathbf{A}$*. Moreover,* $\mathbf{A}$ *is invertible if and only if* $\det \mathbf{A} \neq 0$*.*

Proof. Now suppose $\mathbf{A}$ is invertible. Then $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$. It follows that

$$\begin{aligned}(\det \mathbf{A})(\det \mathbf{A}^{-1}) &= \det(\mathbf{A}\mathbf{A}^{-1}) \\ &= \det \mathbf{I} \\ &= 1.\end{aligned}$$

This means that $\det \mathbf{A}^{-1} = 1/\det \mathbf{A}$ and that $\det \mathbf{A} \neq 0$. In the course of proving the Determinant Theorem, we had shown that $\det \mathbf{A}$ is non-zero if and only rank $\mathbf{A} = n$. It follows that $\mathbf{A}$ is invertible if and only if $\det \mathbf{A} \neq 0$. ∎

9.24 The Adjoint Matrix

Adjoint Matrix. Let $\mathbf{A}$ be an $n \times n$ matrix. Define the *adjoint* of $\mathbf{A}$, adj $\mathbf{A}$ by $(\operatorname{adj} \mathbf{A})_{ij} = C_{ji}$ where C_{ji} is the ji-cofactor of (j, i).

Notice the implicit transposition in the definition of adjoint. The ji-cofactor is used for the ij entry.

Theorem 9.24.1. *Let $\mathbf{A}$ be an invertible $n \times n$ matrix. Then*

$$\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \operatorname{adj} \mathbf{A}.$$

Proof. It is enough to show that $\mathbf{A} \times \operatorname{adj} \mathbf{A} = (\det \mathbf{A})\mathbf{I}_n$. Now

$$\mathbf{A} \times \operatorname{adj} \mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} C_{11} & \cdots & C_{n1} \\ \vdots & \ddots & \vdots \\ C_{1n} & \cdots & C_{nn} \end{pmatrix} = \big(\det \mathbf{A}\big)\mathbf{I}_n \tag{9.24.1}$$

We look at the ij element of the product, which is

$$a_{i1}C_{j1} + a_{i2}C_{j2} + \cdots + a_{in}C_{jn}. \tag{9.24.2}$$

If $i = j$, equation (9.24.2) becomes

$$a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} = \det \mathbf{A}.$$

It is the expansion of the determinant of $\mathbf{A}$ along row i.

What if $i \neq j$? In that case, we have expanded along row j as far as the cofactors are concerned, but have used row i for the a_{i1}. Since $i \neq j$, we are computing the determinant of a matrix where row i occurs both in row i and row j. The result of course is 0. Equation (9.24.1) follows, and so does the theorem. ∎

9.25 Cramer's Rule

A closely related result is *Cramer's Rule*. We prove it by recalling that $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ and using the adjoint formula for the inverse.[1]

Cramer's Rule. *Let $\mathbf{A}$ be an invertible $n \times n$ matrix. Then the equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a unique solution when $\mathbf{x}$ and $\mathbf{b}$ are $n \times 1$ vectors. The solution is*

$$x_i = \frac{\det \mathbf{B}_i}{\det \mathbf{A}}$$

where $\mathbf{B}_i$ is the matrix obtained from $\mathbf{A}$ by replacing the i^{th} column of $\mathbf{A}$ by $\mathbf{b}$.

Proof. Now $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} = (\det \mathbf{A})^{-1}(\operatorname{adj} \mathbf{A})\mathbf{b}$. Since the ij-element of $\operatorname{adj} \mathbf{A}$ is C_{ji}, it follows that the i^{th} element of $\mathbf{x}$ is

$$x_i = \frac{1}{\det \mathbf{A}} \sum_{j=1}^{n} b_j C_{ji}.$$

The sum on the right-hand side is the expansion of $\det \mathbf{B}_i$ along its i^{th}column, $\mathbf{b}$, in which case we multiply b_j by C_{ji}. But then $x_i = \det \mathbf{B}_i / \det \mathbf{A}$, proving Cramer's Theorem. ∎

October 30, 2022

[1] The Genevan mathematecian Gabriel Cramer (1704–1752) is best-known for Cramer's Rule, and for his work on algebraic curves. In particular, he showed that an n^{th} degree curve in general position is determined by $n(n+3)/2$ points of the curve.