8. Matrix Algebra

9/1/20

We start by defining matrices.

Matrix. An $m \times n$ *matrix* is a rectangular array $\mathbf{A}$ of mn *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.[1]

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

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

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

Matrices $\mathbf{A}$ and $\mathbf{B}$ are *equal* if they are both $m \times n$ and have identical entries, $a_{ij} = b_{ij}$ for every $i = 1, \dots, m$ and $j = 1, \dots, n$.

[1] For example, there is 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 cannot be added together. We will say that matrices are *conformable for addition* if they have the same number of rows and 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.

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

8.2 Multiplying a Matrix by a Scalar

We can also multiply a matrix by a number α.

Scalar Multiplication. The matrix $\alpha\mathbf{A}$ is defined by $(\alpha a)_{ij} = \alpha a_{ij}$.

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 is defined only on the left, not the right. Scalar multiplication obeys two additive distributive laws. One distributes scalar multiplication over matrix addition, the other distributes scalar addition when multiplying scalars by a matrix.

$$\begin{aligned} \alpha(\mathbf{A} + \mathbf{B}) &= \alpha\mathbf{A} + \alpha\mathbf{B} && \text{scalar distributive law I} \\ (\alpha + \beta)\mathbf{A} &= \alpha\mathbf{A} + \beta\mathbf{A} && \text{scalar distributive law II} \end{aligned}$$

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

Not surprisingly, adding the zero matrix to any matrix has no effect. It is easy to show that there can be only one additive identity. The zero matrix can also be obtained by multiplying any matrix by zero: $0\mathbf{A} = \mathbf{0}$.

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

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

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

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

We have

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

8.3 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
$0\mathbf{A} = \mathbf{0}$	scalars and additive identity
$(-1)\mathbf{A} = -\mathbf{A}$	scalars 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.

8.4 Complex Matrices

There's no reason to restrict ourselves to real numbers when defining matrices. We could use any number field. The only one that we will use are the 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, $i = \sqrt{-1}$. 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 each $\lambda_i \in \mathbb{C}$.[2]

This is not true of real numbers. The equation $x^2 + 1 = 0$ has no real solutions. However, there are two complex solutions: i and $-i$. Accordingly, $x^2+1 = (x-i)(x+i)$.

[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.

8.5 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.

$$\begin{aligned} (\mathbf{A} + \mathbf{B})^\mathsf{T} &= \mathbf{A}^\mathsf{T} + \mathbf{B}^\mathsf{T} \\ (\alpha\mathbf{A})^\mathsf{T} &= \alpha\mathbf{A}^\mathsf{T} \end{aligned}$$

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 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 8.5.1. *If $\mathbf{A}$ is a square matrix, $\mathbf{B} = (\mathbf{A} + \mathbf{A}^\mathsf{T})/2$ is symmetric, $\mathbf{C} = (\mathbf{A} - \mathbf{A}^\mathsf{T})/2$ is skew-symmetric, and $\mathbf{A} = \mathbf{B} + \mathbf{C}$.*

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

8.6 Hermitian Conjugate of a Matrix

A related concept that only effects 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.

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

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 the complex matrix analog of real numbers and skew-Hermitian matrices are the 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 8.6.1. *If $\mathbf{A}$ is a square matrix, $\mathbf{B} = (\mathbf{A} + \mathbf{A}^*)/2$ is Hermitian, $\mathbf{C} = (\mathbf{A} - \mathbf{A}^*)/2$ is anti-Hermitian, and $\mathbf{A} = \mathbf{B} + \mathbf{C}$.*

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

8.7 Matrix Multiplication

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 $k = n$.

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, the *matrix product* $\mathbf{A} \times \mathbf{B}$ (also 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: $(\mathbf{AB})^* = \mathbf{B}^*\mathbf{A}^*$.

8.8 Matrix Multiplication: Basic Properties

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

$$
\begin{array}{rl}
\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \times \mathbf{B}) \times \mathbf{C} & \text{multiplication associates} \\
\alpha(\mathbf{A} \times \mathbf{B}) = (\alpha\mathbf{A}) \times \mathbf{B} = \mathbf{A} \times (\alpha\mathbf{B}) & \text{scalar associative law} \\
(\mathbf{A} + \mathbf{B}) \times \mathbf{C} = \mathbf{A} \times \mathbf{C} + \mathbf{B} \times \mathbf{C} & \text{matrix distributive law I} \\
\mathbf{A} \times (\mathbf{B} + \mathbf{C}) = \mathbf{A} \times \mathbf{B} + \mathbf{A} \times \mathbf{C} & \text{matrix distributive law II}
\end{array}
$$

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.

8.9 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

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

Then

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

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

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

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.

8.10 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

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \times \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 7 & 8 & 9 \\ 4 & 5 & 6 \end{pmatrix} \tag{8.10.1}$$

but

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 3 & 2 \\ 4 & 6 & 5 \\ 7 & 9 & 8 \end{pmatrix} \tag{8.10.2}$$

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

Interestingly enough, in equation (8.10.1), **pre**-multiplication by the $0, 1$ matrix 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 (8.10.2), **post**-multiplication by the same matrix switches the columns of the $1, 2, 3, \ldots$ matrix.

8.11 Linear Systems and Matrices

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

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
a_{31}x_1 + a_{32}x_2 + \cdots + a_{3n}x_n &= b_3 \\
\vdots \qquad\qquad\qquad & \quad \vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m
\end{aligned}
\tag{6.5.2}
$$

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

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

Let $\mathbf{A}$ be the $m \times n$ coefficient matrix with ij element a_{ij}. Then the system (6.5.2) can be written

$$
\mathbf{Ax} = \mathbf{b}.
$$

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.

8.12 The Identity Matrix

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,

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

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 8.12.1. *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. $\square$

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 $\mathbf{A}$.

8.13 Inverse Matrices

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 non-singular.

Theorem 8.13.1. *Suppose an $n \times n$ matrix $\mathbf{A}$ is invertible. Then* $\operatorname{rank} \mathbf{A} = n$ *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 7.22.1, $\operatorname{rank} \mathbf{A}$ is the number of rows, n. The number of columns is also n, so $\mathbf{A}$ is non-singular by Corollary 7.23.2. □

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 8.23.1.

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 conjugate is the conjugate of the inverse.

Theorem 8.13.2. *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. □

8.14 Left and Right Inverses

If 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 $\mathbf{B}A = \mathbf{I}_n$ and a *right inverse* is an $n \times m$ matrix $\mathbf{C}$ with $A\mathbf{C} = \mathbf{I}_m$. Notice that if A has both left and right inverses, it must be square and the inverses must be identical as shown in Theorem 8.16.1.

For example,

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

has right inverse

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

because

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

However,

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

In fact, A has no left inverse.

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

8.15 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 8.15.1.

1. *If $\mathbf{A}$ has a left inverse, then there is at most one solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$ and* $\operatorname{rank}\mathbf{A}$ *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* $\operatorname{rank}\mathbf{A}$ *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

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

Combining them, we see that $\mathbf{x} = \mathbf{x}'$. There is at most one solution. By Corollary 7.23.2, $\operatorname{rank}\mathbf{A}$ 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, $\operatorname{rank}\mathbf{A}$ is the number of rows of $\mathbf{A}$ by Corollary 7.22.1. □

8.16 Two-Sided Inverses

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

Theorem 8.16.1. *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 8.15.1 tells us that $\operatorname{rank} \mathbf{A} = \#\text{cols} = \#\text{rows}$. 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 $\mathbf{A}$. □

8.17 Matrix Inverses and Scalar Products

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

Theorem 8.17.1. *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. □

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

Theorem 8.17.2. *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. □

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

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,

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

8.19 Elementary Row Matrices I

There are two classes of elementary matrices—elementary row matrices and elementary column matrices. Pre-multiplying a matrix $\mathbf{A}$ 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 (8.10.1).

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \times \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 7 & 8 & 9 \\ 4 & 5 & 6 \end{pmatrix} \tag{8.10.1}$$

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 (8.10.2), post-multiplying by the same matrix switched the second and third columns, an elementary column operation.

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 $\mathbf{A}$ by this matrix will switch $\mathbf{A}$'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 $\mathbf{A}$. Let $c_{k\ell}$ denote the elements of $\mathbf{E}_{ij} \times \mathbf{A}$.

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

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

8.20 Elementary Row Matrices II

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

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

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

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.

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

$$\begin{aligned} \mathbf{E}_{ij}^{-1} &= \mathbf{E}_{ij}, \\ \mathbf{E}_i(r)^{-1} &= \mathbf{E}_i(1/r), \text{ and} \\ \mathbf{E}_{ij}(r)^{-1} &= \mathbf{E}_{ij}(-r). \end{aligned}$$

Two examples:

$$\mathbf{E}_2(3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } \mathbf{E}_{32}(r) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & r \\ 0 & 0 & 1 \end{pmatrix}.$$

8.21 Some Matrix 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 $\mathbf{E}_{ij}$ as a square root 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.

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

8.22 Elementary Column Matrices

Two types of the elementary matrices, $\mathbf{E}_{ij}$ and $\mathbf{E}_i(r)$ are symmetric. The other type, $\mathbf{E}_{ij}(r)$ is not symmetric.

There are elementary column operations corresponding to the elementary row operations. There are three of them: Interchanging two columns, multiplying a column by a non-zero scalar, and adding a non-zero multiple of one column to another.

The matrices that carry out the first two operations when **post**-multiplied are $\mathbf{E}_{ij}$ and $\mathbf{E}_i(r)$. However, the third elementary column operation requires a different matrix, $\mathbf{E}_{ij}(r)^{\mathbf{T}}$.

Where does the transpose come in? The symmetry of the other two types of matrices means for those elementary column operations, we use the same matrices as the elementary row operations.

In fact, if we apply an elementary row operation to $\mathbf{A}$, it means that we have applied an elementary column operation to $\mathbf{A}^{\mathbf{T}}$. If $\mathbf{E}$ is the elementary row matrix that does this, the transformed matrix is $\mathbf{E} \times \mathbf{A}$, and when we put it in column form by transposing we obtain $(\mathbf{E} \times \mathbf{A})^{\mathbf{T}} = \mathbf{A}^{\mathbf{T}} \times \mathbf{E}^{\mathbf{T}}$. Any elementary column operation can be obtained by post-multiplying by the transpose of the corresponding elementary row matrix.

8.23 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 7.25.1). That means that there are elementary matrices $\mathbf{E}_1, \dots, \mathbf{E}_k$ 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 identity 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 8.13.1, we have proven that non-singularity and invertibility are the same.

Theorem 8.23.1. *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

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

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.

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

Theorem 8.23.2. *Let $\mathbf{A}$ be an $n \times n$ invertible matrix. Then there are elementary matrices $\mathbf{F}_1, \dots, \mathbf{F}_k$ with $\mathbf{A} = \mathbf{F}_1\mathbf{F}_2 \cdots \mathbf{F}_k$.*

Proof. Using the notation above, we have $\mathbf{A}^{-1} = \mathbf{E}_k \cdots \mathbf{E}_1$. Take the inverse. Since the inverse of any of the elementary matrics is also an elementary matrix, we may set $\mathbf{F}_i = \mathbf{E}_i^{-1}$. $\square$

8.24 Input-Output Systems

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 x, the required input is $\mathbf{A}x$. For this to work, we must have $\mathbf{c} + \mathbf{A}x = x$. In other words, $\mathbf{c}$ must solve $\mathbf{c} = (\mathbf{I} - \mathbf{A})x$ or $x = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{c}$. 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 are making profits).

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

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

and that the positive profit condition is

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

We have the following result.

Theorem 8.24.1. *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 8.24.2. *Under the conditions of Theorem 8.24.1, 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.

8.25 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
$0\mathbf{A} = \mathbf{0}$	scalars and additive identity
$(-1)\mathbf{A} = -\mathbf{A}$	scalars 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
$\alpha(\mathbf{A} + \mathbf{B}) = \alpha\mathbf{A} + \alpha\mathbf{B}$	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 defined only for square matrices.[1] Let $\mathbf{A}$ be an $n \times n$ matrix. We will inductively define the *determinant of* $\mathbf{A}$, $\det \mathbf{A}$. If $n = 1$, $\mathbf{A} = (a_{11})$ and $\det \mathbf{A} = a_{11}$. If we have defined determinants up to size $n - 1$, we define the determinant for an $n \times n$ matrix $\mathbf{A}$ 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

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

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

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

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

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.

[1] This chapter draws on Chapters 9 and 26

9.1 Determinants of Diagonal Matrices

9/3/20

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

Theorem 9.1.1. *Let* $\mathbf{D}_n = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ *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. Then we expand along the top row:

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

where the last line follows from the induction hypothesis. 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 $n = 1, 2, \ldots$. □

9.2 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

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

The left-hand side is $0 + 0$ while the right-hand side is 1.

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

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

Examples where they do add and both sides 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, $\mathbf{A} = \operatorname{diag}(a_1, \ldots, a_n)$ and $\mathbf{B} = \operatorname{diag}(b_1, \ldots, b_n)$, the condition for additivity of the determinant is

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

Even in the 2×2 case, this requires $a_1 b_2 + a_2 b_1 = 0$. For larger matrices, the conditions for additivity of the determinant become more stringent.

9.3 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 9.3.1. *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 9.1.1. The upper triangular case is a similar induction, except we expand along the first column rather rather than the first row. □

9.4 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 $\boldsymbol{A}$ is a 2×2 matrix. Then

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

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

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

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.

9.5 Interchanging Rows

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.

Theorem 9.5.1. *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 $n \geq 2$.

For the induction step, we consider matrices of size $n + 1 \geq 3$ and expand the determinant along 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 finsh 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. □

Alternating. A function $f(x_1, \ldots, x_n)$ is *alternating* if whenever we interchange two of the x_i, f is multiplied by (-1), flipping the sign.

Theorem 9.5.1 tells us that determinants are alternating, both with respect to row interchange and column interchange.

9.6 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 9.5.1.

Theorem 9.6.1. *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 9.5.1, $\det \mathbf{A} = -\det \mathbf{A}$. Then $2 \det \mathbf{A} = 0$, so $\det \mathbf{A} = 0$.

9.7 Determinants of Size 3

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

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

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

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

A way to remember 3×3 determinants is to repeat the first two columns, and attach a plus sign to the first 3 diagonals, and a minus sign to the first three anti-diagonals.

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

9.8 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.

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. 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.

9.9 Determinants via Permutation

A second commonly used definition of determinants uses permutations of the indices.

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

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

We can now write the determinant as

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

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 $\mathbf{A}$ the formula yields

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

which is the previously calculated value.

9.10 The Language of Functions

Before continuing with determinants, it will be helpful to upgrade our terminology.[2] We need to be able to discuss 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 $f\colon X \to Y$ 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 $\mathbf{x}$. We can also use functions on sets. If $A \subset X$, the *image* of A under f is $f(A) = \{f(\mathbf{x}) : \mathbf{x} \in A\}$. Of course $f(A) \subset Y$. The image of the domain X is referred to as the range, $\operatorname{ran} f = f(X)$.

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

► **Example 9.10.1: 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, 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 $y^{1/3} \in \operatorname{dom} g$. ◄

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 9.10.2: 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 7.22.1.

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 7.21.2 tells us that f is one-to-one if and only if $n = \operatorname{rank} \mathbf{A}$. ◄

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

Theorem 9.10.3. *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 $x \mapsto x(y)$ 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 x is the unique element of Y that is the image of a point in X.

[2] See also section 13.1 of Simon and Blume.

9.11 Linear and Multilinear Functions

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

$$f(\alpha\mathbf{x} + \mathbf{y}) = \alpha f(\mathbf{x}) + f(\mathbf{y}) \tag{9.11.1}$$

for all scalars α and vectors $\mathbf{x}$ and $\mathbf{y}$. It's pretty obvious that the two linearity conditions imply equation (9.11.1).

To see that equation (9.11.1) 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 $f(\alpha\mathbf{x}) = \alpha f(\mathbf{x})$.

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

Theorem 9.11.1. *Let $\mathbf{A}$ be an $m \times n$ matrix. The function $T_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^m$ 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 10.5.1 that all linear transformations $T\colon \mathbb{R}^n \to \mathbb{R}^m$ have can be written $T = T_{\mathbf{A}}$ for some $n \times n$ matrix $\mathbf{A}$.

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

Multilinearity. A function $\mathbb{R}^{nk} \to \mathbb{R}$ is k-*linear* if it is separately linear in each coordinate k. 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.

9.12 Determinants are Multilinear

Our real reason for being interested so 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, $f_n(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \det \mathbf{A}$.

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

Theorem 9.12.1. *Let $\mathbf{A}$ be an $n \times n$ matrix, and $f_n(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \det \mathbf{A}$, where the $\mathbf{a}_i$ are the rows (columns) of $\mathbf{A}$. Then f_n is n-linear in $(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ for $n \geq 1$.*

Proof. Replace row i by $\mathbf{a}_i + \alpha \mathbf{a}'_i$ for any scalar α and vector $\mathbf{a}'_i \in \mathbb{R}^{n+1}$. We now expand the determinant f_n on the row of interest, row i.

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

showing 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 one row to another. In particular, $\det \mathbf{E}_{ij}(r) = 1$.

9.13 Bilinear Forms

We can use a 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*.[3] Set

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

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

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

The case of the first coordinate is similar.

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

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

and similarly for larger sets of indices.

We write

$$\begin{aligned} f(\mathbf{x}, \alpha\mathbf{y} + \mathbf{z}) &= \sum_{ij=1}^{n} a_{ij} x_i (\alpha y_j + z_j) \\ &= \sum_{ij=1}^{n} a_{ij} x_i (\alpha y_j) + \sum_{ij=1}^{n} a_{ij} x_i z_j \\ &= \alpha \sum_{ij=1}^{n} a_{ij} x_i y_j + \sum_{ij=1}^{n} a_{ij} x_i z_j \\ &= \alpha f(\mathbf{x}, \mathbf{y}) + f(\mathbf{x}, \mathbf{z}) \end{aligned}$$

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

9.14 Tensors

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

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

The 4-dimensional array $\left[a_{hijk}\right]$ is an example of a *tensor*, more specifically, a 4-tensor. We can define a k-tensor by the k-linear mapping

$$A(x_1, \ldots, x_k) = \sum_{j_1 \cdots j_k = 1}^{n} a_{j_1 \cdots j_k} x_{1j_1} \cdots x_{kj_k}$$

where each $x_i = (x_{i1}, \ldots, x_{in}) \in \mathbb{R}^n$ gives us a tensor on $\mathbb{R}^{nk}$ described by the k-dimensional array $\left[a_{j_1 \cdots j_k}\right]$.

More generally, anything of the form

$$A(x_1, \ldots, x_k) = \sum_{j_1 \cdots j_k = 1}^{n} a_{j_1 \cdots j_k} x_{1j_1} \cdots x_{kj_k}$$

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 and made famous by Albert Einstein. However, the notation is **simplified here** by focusing on tensors that are functions solely of ordinary (contravariant) vectors.

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

9.15 Determinants: Yet Another Definition

The third definition of determinant consists of the three conditions stated in the following 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 as a function defined on the rows of $\mathbf{A}$, $\mathbf{a}_1, \ldots, \mathbf{a}_n$. Such a function is the determinant if and only if*

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

Proof. Only if case: We have proven this in pieces already. There are three relevant theorems using the cofactor definition. Theorem 9.5.1 showed that the determinant as defined by cofactor expansion, is an alternating function. Theorem 9.12.1 showed that it is also multilinear, and Theorem 9.1.1 implies that $\det \mathbf{I} = 1$.

It's not terribly hard to verify these properties also hold for the permutation definition.

If case: 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.

9.16 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 a_i by $a_i + ra_j$. By multilinearity,

$$f_n(a_1, \dots, a_i + ra_j, \dots, a_n) = f_n(a_1, \dots, a_n) + rf_n(a_1, \dots, a_j, \dots, a_n)$$

The latter term has a_j in both row i and row j. Since f_n is alternating, interchanging those rows flips the sign. But it also leaves $f_n(a_1, \dots, a_j, \dots, a_n)$ unchanged. The only way that can happen is if it is zero. Thus

$$f_n(a_1, \dots, a_i + ra_j, \dots, a_n) = f_n(a_1, \dots, a_n).$$

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

9.17 The Determinant Theorem, III

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

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

There two possibilities for $\mathbf{R}$. If $\boldsymbol{A}$ is non-singular, then $\mathbf{R} = \mathbf{I}_n$ by Corollary 7.24.1 and Lemma 7.25.1. It follows that

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

in which case $f_n(\mathbf{I}_n)$ uniquely determines f_n.

Otherwise, rank $\boldsymbol{A} < n$ and $\mathbf{R}$ will have a zero row. If we multiple that row by zero, $\mathbf{R}$ doesn't change, but f_n has to take the value 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.

9.18 Determinants, Transposes, and Inverses

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

Proof. See above. □

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

Theorem 9.18.2. *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. □

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

Theorem 9.18.3. *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 $\mathbf{AA}^{-1} = \mathbf{I}$. It follows that

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

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 $\operatorname{rank} \mathbf{A} = n$. It follows that $\mathbf{A}$ is invertible if and only if $\det \mathbf{A} \neq 0$. □

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

Theorem 9.19.1. *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

$$\mathbf{A} \times \operatorname{adj}\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} C_{11} & \cdots & C_{n1} \\ \vdots & \ddots & \vdots \\ C_{1n} & \cdots & C_{nn} \end{pmatrix} = \left(\det \mathbf{A}\right)\mathbf{I}_n \tag{9.19.2}$$

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

$$a_{i1}C_{j1} + a_{i2}C_{j2} + \cdots + a_{in}C_{jn}. \tag{9.19.3}$$

If $i = j$, equation (9.19.3) becomes

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

The point is that it is the expansion of the determinant $\mathbf{A}$ 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 (9.19.2) follows, and so does the theorem. □

9.20 Cramer's Rule

A closely related result is *Cramer's Rule*, which we state without proof. To prove it, recall that $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ and use the adjoint formula for the inverse.

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 $\mathbf{b}$.

September 28, 2020