

Matrices
MA1S1

Tristan McLoughlin

November 9, 2014

Anton & Rorres: Ch 1.3-1.8

Basic matrix notation

We have studied matrices as a tool for solving systems of linear equations but now we want to study them in their own right. Let us start by recalling some definitions:

A **matrix** is a rectangular array or table of numbers. We called the individual numbers **entries** of the matrix and referred to them by their row and column numbers. The rows are numbered $1, 2, \dots$ from the top and the columns are numbered $1, 2, \dots$ from left to right. So we use what you might think of as a

(row, column)

coordinate system for the entries of a matrix.

In the example

$$\begin{bmatrix} 1 & 1 & 2 & 5 \\ 1 & 11 & 13 & -2 \\ 2 & 1 & 3 & 4 \end{bmatrix}$$

13 is the $(2, 3)$ entry, the entry in row 2 and column 3.

The matrix above is called a 3×4 matrix because it has 3 rows and 4 columns.

Basic matrix notation

We can talk about matrices of all different sizes such as

$$\begin{array}{cccc} \begin{bmatrix} 4 & 5 \\ 7 & 11 \end{bmatrix} & \begin{bmatrix} 4 \\ 7 \end{bmatrix} & \begin{bmatrix} 4 & 7 \end{bmatrix} & \begin{bmatrix} 4 & 5 \\ 7 & 11 \\ 13 & 13 \end{bmatrix} \\ 2 \times 2 & 2 \times 1 & 1 \times 2 & 3 \times 2 \end{array}$$

and in general we can have $m \times n$ matrices for any $m \geq 1$ and $n \geq 1$.

Matrices with just one row are called *row matrices*, e.g.

$$[x_1, x_2, \dots, x_n]$$

While we use the term *column matrix* for a matrix with just one column.
Here is an $n \times 1$ (column) matrix

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

We consider two matrices to be the ‘same’ matrix only if they are absolutely identical. They have to have the same shape (same number of rows and same number of columns) and they have to have the same numbers in the same positions.

Thus, all the following are different matrices

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \neq \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} \neq \begin{bmatrix} 2 & 1 & 0 \\ 3 & 4 & 0 \end{bmatrix} \neq \begin{bmatrix} 2 & 1 \\ 3 & 4 \\ 0 & 0 \end{bmatrix}$$

Double subscripts

When we want to discuss a matrix without listing the numbers in it, that is when we want to discuss a matrix that is not yet specified or an unknown matrix we use a notation with double subscripts

$$\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$$

This is a 2×2 matrix where the $(1, 1)$ entry is x_{11} , the $(1, 2)$ entry is x_{12} and so on.

Carrying this idea further, when we want to discuss an $m \times n$ matrix X and refer to its entries we write

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

So the (i, j) entry of X is called x_{ij} .

Sometimes we want to write something like this but we don't want to take up space for the whole picture and we write an abbreviated version like

$$X = [x_{ij}]_{1 \leq i \leq m, 1 \leq j \leq n}$$

To repeat what we said about when matrices are equal using this kind of notation, suppose we have two $m \times n$ matrices

$$X = [x_{ij}]_{1 \leq i \leq m, 1 \leq j \leq n} \text{ and } Y = [y_{ij}]_{1 \leq i \leq m, 1 \leq j \leq n}$$

Then $X = Y$ means the mn scalar equations $x_{ij} = y_{ij}$ must all hold (for each (i, j) with $1 \leq i \leq m, 1 \leq j \leq n$). And if an $m \times n$ matrix equals an $r \times s$ matrix, we have to have $m = r$ (same number of rows), $n = s$ (same number of columns) **and** all the entries equal.

Arithmetic with matrices

In much the same way as we did with n -tuples we now define addition of matrices. We only allow addition of matrices that are of the same size. Two matrices of different sizes cannot be added.

To add two matrices we simply add each corresponding component. For example

$$\begin{bmatrix} 2 & 1 \\ 3 & -4 \\ 0 & 7 \end{bmatrix} + \begin{bmatrix} 6 & -2 \\ 15 & 12 \\ -9 & 21 \end{bmatrix} = \begin{bmatrix} 2+6 & 1+(-2) \\ 3+15 & -4+12 \\ 0+(-9) & 7+21 \end{bmatrix} = \begin{bmatrix} 8 & -1 \\ 18 & 8 \\ -9 & 28 \end{bmatrix}$$

More generally, if we take two $m \times n$ matrices

$$X = [x_{ij}]_{1 \leq i \leq m, 1 \leq j \leq n} \text{ and } Y = [y_{ij}]_{1 \leq i \leq m, 1 \leq j \leq n}$$

then we define

$$X + Y = [x_{ij} + y_{ij}]_{1 \leq i \leq m, 1 \leq j \leq n}$$

(the $m \times n$ matrix with the $(1, 1)$ entry being the sum of the $(1, 1)$ entries of X and Y , the $(1, 2)$ entry the sum of the $(1, 2)$ entries of X and Y , and so on).

To define the scalar multiple kX , for a number k and a matrix X we just multiply every entry of X by k .

For example

$$8 \begin{bmatrix} 2 & 1 \\ 3 & -4 \\ 0 & 7 \end{bmatrix} = \begin{bmatrix} 8(2) & 8(1) \\ 8(3) & 8(-4) \\ 8(0) & 8(7) \end{bmatrix} = \begin{bmatrix} 16 & 8 \\ 24 & -32 \\ 0 & 56 \end{bmatrix}$$

The general definition is:

$$X = [x_{ij}]_{1 \leq i \leq m, 1 \leq j \leq n}$$

is any $m \times n$ matrix and k is any real number then kX is another $m \times n$ matrix. Specifically

$$kX = [kx_{ij}]_{1 \leq i \leq m, 1 \leq j \leq n}$$

We see that if we multiply by $k = 0$ we get a matrix where all the entries are 0.

The zero matrix

The $m \times n$ matrix where every entry is 0 is called the $m \times n$ zero matrix. Thus we have zero matrices of every possible size.

If X is a matrix then we can say

$$X + \mathbf{0} = X$$

if $\mathbf{0}$ means the zero matrix of the same size as X . If we wanted to make the notation less ambiguous, we could write something like $\mathbf{0}_{m,n}$ for the $m \times n$ zero matrix. Then things we can note are that if X is any $m \times n$ matrix then

$$X + \mathbf{0}_{m,n} = X, \quad 0X = \mathbf{0}_{m,n}$$

We will not usually go to the lengths of indicating the size of the zero matrix we mean in this way. We will write the zero matrix as $\mathbf{0}$ and try to make it clear what size matrices we are dealing with from the context.

Matrix transformations

Let us temporarily return to our study of linear equations. Consider the equations:

$$\begin{aligned}y_1 &= x_1 + 2x_2 \\ y_2 &= \quad + 3x_2\end{aligned}$$

We can think of these equations as a matrix equation

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 + 2x_2 \\ 3x_2 \end{bmatrix}$$

This can be viewed as a rule for transforming the vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ into the vector } \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} .$$

The transformation is described by the matrix

$$\mathbf{M} = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$$

Matrix transformations

If we write this as $\mathbf{y} = \mathbf{M}\mathbf{x}$ this tells us how to multiply matrices:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 + 2x_2 \\ 3x_2 \end{bmatrix}$$

Now we will generalise this to multiplying arbitrary matrices.

Matrix multiplication

This is a rather new thing, compared to the ideas we have discussed up to now. As we saw certain matrices can be multiplied and their product is another matrix.

If X is an $m \times n$ matrix and Y is an $n \times p$ matrix then the product XY will make sense and it will be an $m \times p$ matrix.

For example, then

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & -2 \\ 2 & -1 & 3 & 1 \\ 4 & 2 & 6 & 4 \end{bmatrix}$$

is going to make sense. It is the product of

$$2 \times 3 \text{ by } 3 \times 4$$

and the result is going to be 2×4 . (We have to have the same number of columns in the left matrix as rows in the right matrix. The outer numbers, the ones left after ‘cancelling’ the same number that occurs in the middle, give the size of the product matrix.)

As an example of a product that **is not defined** and will not make sense

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 7 & 8 \\ 9 & 10 \end{bmatrix} \quad 2 \times \textcolor{blue}{3} \text{ by } \textcolor{red}{2} \times 2$$

Back to the example that will make sense, what we have explained so far is the shape of the product

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & -2 \\ 2 & -1 & 3 & 1 \\ 4 & 2 & 6 & 4 \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} & z_{13} & z_{14} \\ z_{21} & z_{22} & z_{23} & z_{24} \end{bmatrix}$$

and we still have to explain how to calculate the z_{ij} , the entries in the product.

We'll concentrate on one example to try and show the idea. Say we look at the entry z_{23} , the $(2,3)$ entry in the product. What we do is take row 2 of the left matrix 'times' column 3 of the right matrix

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & -2 \\ 2 & -1 & 3 & 1 \\ 4 & 2 & 6 & 4 \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} & z_{13} & z_{14} \\ z_{21} & z_{22} & z_{23} & z_{24} \end{bmatrix}$$

The way we multiply the row $[4 \ 5 \ 6]$ times the column

$$\begin{bmatrix} 1 \\ 3 \\ 6 \end{bmatrix}$$

is a very much reminiscent of a dot product

$$(4)(1) + (5)(3) + (6)(6) = z_{23}$$

In other words $z_{23} = 55$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & -2 \\ 2 & -1 & 3 & 1 \\ 4 & 2 & 6 & 4 \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} & z_{13} & z_{14} \\ z_{21} & z_{22} & 55 & z_{24} \end{bmatrix}$$

If we calculate all the other entries in the same sort of way (row i on the left times column j on the right gives z_{ij}) we get

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & -2 \\ 2 & -1 & 3 & 1 \\ 4 & 2 & 6 & 4 \end{bmatrix} = \begin{bmatrix} 17 & 4 & 25 & 12 \\ 38 & 7 & 55 & 21 \end{bmatrix}$$

As we discussed the only way to get used to the way to multiply matrices is to do some practice. Consider the two matrices:

$$A = \begin{bmatrix} 4 & 1 & 12 \\ 0 & 1 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 5 & 4 \\ 1 & 1 \\ -2 & 3 \end{bmatrix}$$

As these are 2×3 and a 3×2 matrix they can be multiplied to give 2×2 .

So we have:

$$AB = \begin{bmatrix} 4(5) + 1(1) + 12(-2) & 4(4) + 1(1) + 12(3) \\ 0(5) + 1(1) + 4(-2) & 0(4) + 1(1) + 4(3) \end{bmatrix} = \begin{bmatrix} -3 & 53 \\ -7 & 13 \end{bmatrix}$$

It is possible to explain in a succinct formula what the rule is for calculating the entries of the product matrix.

In the equation

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2p} \\ \vdots & & \ddots & \vdots \\ z_{m1} & z_{m2} & \cdots & z_{mp} \end{bmatrix}$$

the (i, k) entry z_{ik} of the product is got by taking the dot product of the i^{th}

row $[x_{i1} \ x_{i2} \ \dots \ x_{in}]$ of the first matrix times the k^{th} column $\begin{bmatrix} y_{1k} \\ y_{2k} \\ \vdots \\ y_{nk} \end{bmatrix}$ of the

second.

In short

$$z_{ik} = x_{i1}y_{1k} + x_{i2}y_{2k} + \cdots + x_{in}y_{nk}$$

or with the Sigma notation for sums, we can rewrite this as

$$\sum_{j=1}^n x_{ij}y_{jk} = z_{ik} \quad (\text{for } 1 \leq i \leq m, 1 \leq k \leq p).$$

Hence for $X = [x_{ij}]_{1 \leq i \leq m, 1 \leq j \leq n}$ and $Y = [y_{ij}]_{1 \leq i \leq n, 1 \leq j \leq p}$ we have

$$XY = Z$$

where

$$Z = \left[\sum_{j=1}^n x_{ij}y_{jk} \right]_{1 \leq i \leq m, 1 \leq k \leq p}.$$

In previous weeks we considered the system of m linear equations in n unknowns:

$$\begin{array}{cccccc} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ \vdots & & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & b_m \end{array}$$

If we introduce the $m \times n$ coefficient matrix

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

and the column vectors (or $n \times 1$ and $m \times 1$ matrices)

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Then we can write the equations as

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

This in fact gives us another way to think about matrix multiplication as we can write

$$\begin{aligned} A\mathbf{x} &= \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix} \\ &= x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} \end{aligned}$$

Which can be stated as:

We can express the product $A\mathbf{x}$ as a linear combination of the column vectors of A with coefficients given by the entries of \mathbf{x} .

In fact the same picture can be used when multiplying general matrices.

We can phrase the earlier construction in an *equivalent* fashion by saying the in the product of matrices A and B , the **first** column vector of AB is the linear combination of column vectors of A with coefficients from the **first** row of B . The **second** column vector is given by the column vectors of A with coefficients from the **second** row of B . And so on.

For example:

$$\begin{bmatrix} 1 & 2 & 4 \\ 2 & 6 & 0 \end{bmatrix} \begin{bmatrix} 4 & 1 & 4 & 3 \\ 0 & -1 & 3 & 1 \\ 2 & 7 & 5 & 2 \end{bmatrix} = \begin{bmatrix} 12 & 27 & 30 & 13 \\ 8 & -4 & 26 & 12 \end{bmatrix}$$

Where

$$\begin{bmatrix} 12 \\ 8 \end{bmatrix} = 4 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 0 \begin{bmatrix} 2 \\ 6 \end{bmatrix} + 2 \begin{bmatrix} 4 \\ 0 \end{bmatrix}$$

and

$$\begin{bmatrix} 27 \\ -4 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 2 \\ 6 \end{bmatrix} + 7 \begin{bmatrix} 4 \\ 0 \end{bmatrix}$$

and so on.

Matrix multiplication

Recall how we multiply two matrices together: Given two matrices A and B ,

$$A = \begin{bmatrix} 2 & 2 & 8 \\ 5 & 4 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 5 \\ 1 & -1 \\ 2 & 1 \end{bmatrix}$$

we can multiply them to get a 2×2 matrix as follows:

$$\begin{aligned} AB &= \begin{bmatrix} 2 & 2 & 8 \\ 5 & 4 & 3 \end{bmatrix} \begin{bmatrix} 3 & 5 \\ 1 & -1 \\ 2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2(3) + 2(1) + 8(2) & 2(5) + 2(-1) + 8(1) \\ 5(3) + 4(1) + 3(2) & 5(5) + 4(-1) + 3(1) \end{bmatrix} \end{aligned}$$

Matrix multiplication

Recall how we multiply two matrices together: Given two matrices A and B ,

$$A = \begin{bmatrix} 2 & 2 & 8 \\ 5 & 4 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 5 \\ 1 & -1 \\ 2 & 1 \end{bmatrix}$$

we can multiply them to get a 2×2 matrix as follows:

$$\begin{aligned} AB &= \begin{bmatrix} 2 & 2 & 8 \\ 5 & 4 & 3 \end{bmatrix} \begin{bmatrix} 3 & 5 \\ 1 & -1 \\ 2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2(3) + 2(1) + 8(2) & 2(5) + 2(-1) + 8(1) \\ 5(3) + 4(1) + 3(2) & 5(5) + 4(-1) + 3(1) \end{bmatrix} \end{aligned}$$

Matrix multiplication

Recall how we multiply two matrices together: Given two matrices A and B ,

$$A = \begin{bmatrix} 2 & 2 & 8 \\ 5 & 4 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 5 \\ 1 & -1 \\ 2 & 1 \end{bmatrix}$$

we can multiply them to get a 2×2 matrix as follows:

$$\begin{aligned} AB &= \begin{bmatrix} 2 & 2 & 8 \\ 5 & 4 & 3 \end{bmatrix} \begin{bmatrix} 3 & 5 \\ 1 & -1 \\ 2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2(3) + 2(1) + 8(2) & 2(5) + 2(-1) + 8(1) \\ 5(3) + 4(1) + 3(2) & 5(5) + 4(-1) + 3(1) \end{bmatrix} \end{aligned}$$

Matrix multiplication

Recall how we multiply two matrices together: Given two matrices A and B ,

$$A = \begin{bmatrix} 2 & 2 & 8 \\ 5 & 4 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 5 \\ 1 & -1 \\ 2 & 1 \end{bmatrix}$$

we can multiply them to get a 2×2 matrix as follows:

$$\begin{aligned} AB &= \begin{bmatrix} 2 & 2 & 8 \\ 5 & 4 & 3 \end{bmatrix} \begin{bmatrix} 3 & 5 \\ 1 & -1 \\ 2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2(3) + 2(1) + 8(2) & 2(5) + 2(-1) + 8(1) \\ 5(3) + 4(1) + 3(2) & 5(5) + 4(-1) + 3(1) \end{bmatrix} \end{aligned}$$

Matrix multiplication

Recall how we multiply two matrices together: Given two matrices A and B ,

$$A = \begin{bmatrix} 2 & 2 & 8 \\ 5 & 4 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 5 \\ 1 & -1 \\ 2 & 1 \end{bmatrix}$$

we can multiply them to get a 2×2 matrix as follows:

$$\begin{aligned} AB &= \begin{bmatrix} 2 & 2 & 8 \\ 5 & 4 & 3 \end{bmatrix} \begin{bmatrix} 3 & 5 \\ 1 & -1 \\ 2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2(3) + 2(1) + 8(2) & 2(5) + 2(-1) + 8(1) \\ 5(3) + 4(1) + 3(2) & 5(5) + 4(-1) + 3(1) \end{bmatrix} \\ &= \begin{bmatrix} 24 & 16 \\ 25 & 24 \end{bmatrix} \end{aligned}$$

Properties of matrix multiplication

Matrix multiplication has the properties that you would expect of any multiplication and the standard rules of algebra work out as long as you keep the order of the products intact:

- (i) If A and B are both $m \times n$ matrices and C is $n \times p$, then

$$(A + B)C = AC + BC$$

(This is the right distributive law for multiplication.)

- (ii) If A is an $m \times n$ matrices and B and C are both $n \times p$ then

$$A(B + C) = AB + AC .$$

(This is the left distributive law for multiplication.)

- (iii) If k is a scalar, and A is an $m \times n$ matrices while C is $n \times p$ then

$$(kA)C = k(AC) = A(kC) .$$

- (iv) If A is an $m \times n$ matrices and B is $n \times p$ and C is $p \times q$, then the two ways of calculating ABC work out the same:

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

(This is known as the associative law for multiplication.)

While most of these properties are fairly obvious let us try to prove the first one: From the definition of matrix multiplication and addition the entries of $(A + B)C$ are given by

$$[(A + B)C]_{ij} = \sum_{k=1}^n (a_{ik} + b_{ik})c_{kj}$$

but because of the distributivity of normal multiplication

$$\begin{aligned} [(A + B)C]_{ij} &= \sum_{k=1}^n (a_{ik} + b_{ik})c_{kj} \\ &= \sum_{k=1}^n (a_{ik}c_{kj} + b_{ik}c_{kj}) \\ &= \sum_{k=1}^n a_{ik}c_{kj} + \sum_{k=1}^n b_{ik}c_{kj} \\ &= [AC]_{ij} + [BC]_{ij} . \end{aligned}$$

Where on the right we have the entries of the matrix $AC + BC$.

We won't prove the property of associativity instead we will give an example:

We saw before that $AB \neq BA$ in general for matrices.

In a bit more detail the situation is as follows:

(a) BA does not have to make sense if AB makes sense.

We saw before that $AB \neq BA$ in general for matrices.

In a bit more detail the situation is as follows:

(a) BA does not have to make sense if AB makes sense.

For example: As we saw, if A is a 3×4 matrix and B is 3×3 , then AB does not make sense. But BA is a product of a 3×3 times a 3×4 — so it makes sense.

We saw before that $AB \neq BA$ in general for matrices.

In a bit more detail the situation is as follows:

- (a) BA does not have to make sense if AB makes sense.
- (b) It can be that AB and BA both make sense but they are different sizes.

We saw before that $AB \neq BA$ in general for matrices.

In a bit more detail the situation is as follows:

- (a) BA does not have to make sense if AB makes sense.
- (b) It can be that AB and BA both make sense but they are different sizes.
For example: If A is a 2×3 matrix and B is a 3×2 matrix, then AB is 2×2 while BA is 3×3 . As they are different sizes AB and BA are certainly not equal.

We saw before that $AB \neq BA$ in general for matrices.

In a bit more detail the situation is as follows:

- (a) BA does not have to make sense if AB makes sense.
- (b) It can be that AB and BA both make sense but they are different sizes.
- (c) The more tricky case is the case where the matrices A and B are *square matrices* of the same size.

We saw before that $AB \neq BA$ in general for matrices.

In a bit more detail the situation is as follows:

- (a) BA does not have to make sense if AB makes sense.
- (b) It can be that AB and BA both make sense but they are different sizes.
- (c) The more tricky case is the case where the matrices A and B are *square matrices* of the same size.

A square matrix is an $n \times n$ matrix for some n . Notice that the product of two $n \times n$ matrices is another $n \times n$ matrix. Still, it is usually not the case that $AB = BA$ when A and B are $n \times n$.

There are some perhaps non-intuitive consequences of the properties of matrix multiplication. For example for real numbers:

- If $ab = ac$ and $a \neq 0$ then $b = c$.
- If $ab = 0$ then at least one of a or b is 0.

Neither of these is true for matrices.

Consider

Identity Matrices

There are some special square matrices which deserve a special name. We've already seen the zero matrix (which makes sense for any size — can be $m \times n$ and need not be square).

Another special matrix is the $n \times n$ **identity matrix** which we denote by I_n . So

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and in general I_n is the $n \times n$ matrix with 1 in all the 'diagonal' entries and zeroes off the diagonal.

By the **diagonal entries** of an $n \times n$ matrix we mean the (i, i) entries for $i = 1, 2, \dots, n$. We do not talk of the diagonal for rectangular matrices.

Identity Matrices

The reason for the name is that the identity matrix is a multiplicative identity.

That is $I_m A = A$ and $A = A I_n$ for any $m \times n$ matrix A . For example

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 5 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 3 & 4 \end{bmatrix}$$

and

$$\begin{bmatrix} 2 & 5 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 3 & 4 \end{bmatrix} .$$

Identity matrices show up naturally in the study of the reduced row echelon form of square (i.e. $n \times n$) matrices.

If R is the reduced row echelon form of an $n \times n$ matrix then either R has one or more rows of zeros or R is the $n \times n$ identity matrix I_n .

Inverse matrices — basic ideas

Definition: If A is an $n \times n$ matrix, then another $n \times n$ matrix C is called the *inverse matrix* for A if it satisfies

$$AC = I_n \text{ and } CA = I_n.$$

We write A^{-1} for the inverse matrix C (if there is one).

The idea for this definition is that the identity matrix is analogous to the number 1, in the sense that $1k = k1 = k$ for every real number k while $AI_n = I_nA = A$ for every $n \times n$ matrix A .

Then the key thing about the reciprocal of a nonzero number k is that the product

$$\left(\frac{1}{k}\right)k = 1$$

We insist that the inverse should work on both sides but we will see a theorem that says that if A and C are $n \times n$ matrices and $AC = I_n$, then automatically $CA = I_n$ must also hold.

Inverse matrices — basic ideas

When does there exist an inverse for a given matrix A ?

It is not enough that A should be nonzero. One way to see this is to look at a system of n linear equations in n unknowns written in matrix form.

If the $n \times n$ matrix A has an inverse matrix C then we can multiply both sides of the equation $A\mathbf{x} = \mathbf{b}$ by C from the left to get

$$\begin{aligned} A\mathbf{x} &= \mathbf{b} \\ \Rightarrow C(A\mathbf{x}) &= C\mathbf{b} \\ \Rightarrow (CA)\mathbf{x} &= C\mathbf{b} \\ \Rightarrow I_n\mathbf{x} &= C\mathbf{b} \\ \Rightarrow \mathbf{x} &= C\mathbf{b} \end{aligned}$$

So we find that the system of n equation in n unknowns given by $A\mathbf{x} = \mathbf{b}$, for any right hand side \mathbf{b} , will just have the **one** solution $\mathbf{x} = C\mathbf{b}$.

Invertible Matrices

For a system of n equation in n unknowns given by

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

where A is an invertible matrix with inverse C and for any right hand side \mathbf{b} , will just have the **one** solution $\mathbf{x} = C\mathbf{b}$. We have seen for many systems of linear equations there are infinite families of solutions or sometimes no solutions.

This amounts to a significant restriction on A and hence not all matrices have inverses. We are led to the definition:

Definition. An $n \times n$ matrix A is called **invertible** if there is an $n \times n$ inverse matrix for A .

We now consider how to find the inverse of a given matrix A . The method will work quite efficiently for large matrices as well as for small ones.

Invertible Matrices - an example

To make things more concrete, we'll think about a specific example

$$A = \begin{bmatrix} 2 & 3 \\ 2 & 5 \end{bmatrix}$$

How can we find

$$C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

so that $AC = I_2$.

Writing out that equation we want C to satisfy we get

$$AC = \begin{bmatrix} 2 & 3 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2$$

If you think of how matrix multiplication works, this amounts to two different equations for the columns of C

$$\begin{bmatrix} 2 & 3 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} c_{11} \\ c_{21} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 & 3 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} c_{12} \\ c_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

According to the reasoning we used above to get to equation $A\mathbf{x} = \mathbf{b}$, each of these represents a system of 2 linear equations in 2 unknowns:

$$2c_{11} + 3c_{21} = 1$$

$$2c_{11} + 5c_{21} = 0$$

and

$$2c_{12} + 3c_{22} = 0$$

$$2c_{12} + 5c_{22} = 1$$

We know how to solve these!

To solve them we can use **Gauss-Jordan elimination** (or Gaussian elimination) twice, once for the augmented matrix for the first system of equations,

$$\left[\begin{array}{cc|c} 2 & 3 & 1 \\ 2 & 5 & 0 \end{array} \right]$$

and again for the second system

$$\left[\begin{array}{cc|c} 2 & 3 & 0 \\ 2 & 5 & 1 \end{array} \right]$$

If we were to write out the steps for the Gauss-Jordan eliminations, we'd find that we were repeating the exact same steps the second time as the first time.

There is a trick to solve at once two systems of linear equations, where the coefficients of the unknowns are the same in both, but the right hand sides are different.

The trick is to write both columns after the dotted line, like this

$$\left[\begin{array}{cc|cc} 2 & 3 & 1 & 0 \\ 2 & 5 & 0 & 1 \end{array} \right]$$

We row reduce this matrix

$$\begin{aligned} \left[\begin{array}{cc|cc} 1 & \frac{3}{2} & \frac{1}{2} & 0 \\ 2 & 5 & 0 & 1 \end{array} \right] & \text{Old Row 1} \times \frac{1}{2} \\ \left[\begin{array}{cc|cc} 1 & \frac{3}{2} & \frac{1}{2} & 0 \\ 0 & 2 & -1 & 1 \end{array} \right] & \text{Old Row 2} - 2 \times \text{Old Row 1} \\ \left[\begin{array}{cc|cc} 1 & \frac{3}{2} & \frac{1}{2} & 0 \\ 0 & 1 & -\frac{1}{2} & \frac{1}{2} \end{array} \right] & \text{Old Row 2} \times \frac{1}{2} \end{aligned}$$

this is now in row echelon form now. We now perform the final steps of Gauss-Jordan

$$\left[\begin{array}{cc|cc} 1 & 0 & \frac{5}{4} & -\frac{3}{4} \\ 0 & 1 & -\frac{1}{2} & \frac{1}{2} \end{array} \right] \text{Old Row 1} - \frac{3}{2} \times \text{Old Row 2}$$

This is in reduced row echelon form so Gauss-Jordan is finished.

The **first** column after the dotted line gives the solution to the **first** system, the one for the **first** column of C . The **second** column after the dotted line relates to the **second** system, the one for the **second** column of C .

That means we have

$$\begin{bmatrix} c_{11} \\ c_{21} \end{bmatrix} = \begin{bmatrix} \frac{5}{4} \\ -\frac{1}{2} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} c_{12} \\ c_{22} \end{bmatrix} = \begin{bmatrix} -\frac{3}{4} \\ \frac{1}{2} \end{bmatrix}$$

So we find that the matrix C has to be

$$C = \begin{bmatrix} \frac{5}{4} & -\frac{3}{4} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

We can multiply out and check that it is indeed true that $AC = I_2$ (which has to be the case unless we made a mistake) and that $CA = I_2$ (a fact which has to be true automatically as we will see).

$$\begin{aligned} AC &= \begin{bmatrix} 2 & 3 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} \frac{5}{4} & -\frac{3}{4} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 2\left(\frac{5}{4}\right) + 3\left(-\frac{1}{2}\right) & 2\left(-\frac{3}{4}\right) + 3\left(\frac{1}{2}\right) \\ 2\left(\frac{5}{4}\right) + 5\left(-\frac{1}{2}\right) & 2\left(-\frac{3}{4}\right) + 5\left(\frac{1}{2}\right) \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} CA &= \begin{bmatrix} \frac{5}{4} & -\frac{3}{4} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 2 & 5 \end{bmatrix} = \begin{bmatrix} \left(\frac{5}{4}\right)(2) + \left(-\frac{3}{4}\right)(2) & \left(\frac{5}{4}\right)(3) + \left(-\frac{3}{4}\right)(5) \\ \left(-\frac{1}{2}\right)(2) + \left(\frac{1}{2}\right)(2) & \left(-\frac{1}{2}\right)(3) + \left(\frac{1}{2}\right)(5) \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

Which is exactly the required property of the inverse matrix!

Invertible 2 Matrices

Let's consider the most general 2×2 matrix:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

and use the above method. In this case the augmented matrix is

$$\begin{aligned} & \left[\begin{array}{cc|cc} 1 & \frac{b}{a} & : & \frac{1}{a} & 0 \\ c & d & : & 0 & 1 \end{array} \right] && \text{Old Row 1} \times \frac{1}{a} \\ \Rightarrow & \left[\begin{array}{cc|cc} 1 & \frac{b}{a} & : & \frac{1}{a} & 0 \\ 0 & d - c\frac{b}{a} & : & -c\frac{1}{a} & 1 \end{array} \right] && \text{Old Row 2} - c \times \text{Old Row 1} \\ \Rightarrow & \left[\begin{array}{cc|cc} 1 & \frac{b}{a} & : & \frac{1}{a} & 0 \\ 0 & 1 & : & \frac{-c}{ad-cb} & \frac{a}{ad-cb} \end{array} \right] && \frac{1}{d - c\frac{b}{a}} \times \text{Old Row 2} \end{aligned}$$

This ends **Gaussian elimination**. Now we do the remaining steps of Gauss-Jordan.

Removing the entry above the leading one in row two:

$$\begin{aligned} & \left[\begin{array}{cc|c} 1 & 0 & \frac{1}{a} \left(1 - \frac{bc}{ad-cb} \right) \\ 0 & 1 & \frac{-c}{ad-cb} \end{array} \right] \quad \text{Old Row 1} - \frac{b}{a} \times \text{Old Row 2} \\ &= \left[\begin{array}{cc|c} 1 & 0 & \frac{d}{ad-cb} \\ 0 & 1 & \frac{-c}{ad-cb} \end{array} \right] \end{aligned}$$

Thus we see that the inverse of an arbitrary 2×2 matrix, A , is

$$A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Here we can see the determinant of the 2×2 matrix appear:

$$\det(A) = ad - bc$$

(note the down-right, plus sign, down-left minus sign rule works).

This general approach works for larger matrices too. If we start with an $n \times n$ matrix

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

and we look for an $n \times n$ matrix

$$C = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}$$

where $AC = I_n$, we want

$$AC = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I_n$$

This means that the columns of C have to satisfy systems of n linear equations in n unknowns of the form

$$A(j^{\text{th}} \text{ column of } C) = j^{\text{th}} \text{ column of } I_n$$

We can solve all of these n systems of equations together because they have the same matrix A of coefficients for the unknowns.

We do this by writing an **augmented matrix** where there are n columns after the dotted line. The columns to the right of the dotted line, the right hand sides of the various systems we want to solve to find the columns of C , are going to be the columns of the $n \times n$ identity matrix.

Method for finding the inverse A^{-1} of an $n \times n$ matrix A .

Use **Gauss-Jordan elimination** to row reduce the augmented matrix

$$[A | I_n] = \left[\begin{array}{cccc|cccc} a_{11} & a_{12} & \cdots & a_{1n} & 1 & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & a_{2n} & 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & 0 & 0 & \cdots & 1 \end{array} \right]$$

We should end up with a reduced row echelon form that looks like

$$\left[\begin{array}{cccc|cccc} 1 & 0 & \cdots & 0 & c_{11} & c_{12} & \cdots & c_{1n} \\ 0 & 1 & \cdots & 0 & c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & c_{n1} & c_{n2} & \cdots & c_{nn} \end{array} \right]$$

or in summary $[I_n | A^{-1}]$.

If we don't end up with a matrix of the form $[I_n | C]$ it means that there is no inverse for A .

Elementary matrices

We now make a link between elementary row operations and matrix multiplication.

Recall now the 3 types of elementary row operations:

- (i) multiply all the numbers in some row by a nonzero factor (and leave every other row unchanged)
- (ii) replace any chosen row by the difference between it and a multiple of some other row.
- (iii) Exchange the positions of some pair of rows in the matrix.

Definition: An $n \times n$ *elementary matrix* E is the result of applying a single elementary row operation to the $n \times n$ identity matrix I_n .

Examples. We use $n = 3$ in these examples. Recall

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- (i) Row operation: Multiply row 2 by -5 . Corresponding elementary matrix

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- (ii) Row operation: Add 4 times row 1 to row 3 (same as subtracting (-4) times row 1 from row 3). Corresponding elementary matrix

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix}$$

- (iii) Row operation: swap rows 2 and 3.

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Row operations and Elementary Matrices

The idea is that if A is an $m \times n$ matrix, then doing one single row operation on A is equivalent to multiplying A on the left by an elementary matrix E (to get EA), and E should be the $m \times m$ elementary matrix for that same row operation.

Row operations and Elementary Matrices

Examples. We use the following A to illustrate this idea,

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix}$$

- (1) Row operation: Add (-5) times row 1 to row 2.
The corresponding E (let's call it E_1):

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and so $E_1 A$ is

$$E_1 A = \begin{bmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -4 & -8 & -12 \\ 9 & 10 & 11 & 12 \end{bmatrix}$$

(Same as doing the row operation to A .)

- (2) Row operation: Suppose in addition we also want to add (-9) times row 1 to row 3. In the context of multiplying by elementary matrices, we need a different elementary matrix for the second step, let's call it E_2 :

$$E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -9 & 0 & 1 \end{bmatrix}$$

What we want in order to do first one and then the next row operation is

$$\begin{aligned} E_2 E_1 A &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -9 & 0 & 1 \end{bmatrix} E_1 A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -9 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -4 & -8 & -12 \\ 9 & 10 & 11 & 12 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -4 & -8 & -12 \\ 0 & -8 & -16 & -24 \end{bmatrix} \end{aligned}$$

where E_1 is the elementary matrix we used first.

So the first row operation changes A to E_1A , and then the second changes that to E_2E_1A .

If we do a whole sequence of several row operations (as we would do if we followed the Gaussian elimination recipe further) we can say that the end result after k row operations is that we get

$$E_k E_{k-1} \dots E_3 E_2 E_1 A$$

where E_i is the elementary matrix for the i^{th} row operation we did.

Elementary matrices are invertible

We heard before all elementary row operations are reversible by another elementary row operation. It follows that every elementary matrix E has an inverse that is another elementary matrix.

Elementary matrices are invertible

For example, take E to be the 3×3 elementary matrix corresponding to the row operation “add (-5) times row 1 to row 2”. So

$$E = \begin{bmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Then the reverse row operation is “add 5 times row 1 to row 2”, and the elementary matrix for that is

$$\tilde{E} = \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Thinking in terms of row operations, or just by multiplying out the matrices we see that the result of applying second row operation to E :

$$\tilde{E}E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3$$

and $E\tilde{E} = I_3$ also.

Theory about invertible matrices

Proposition

If A is an invertible $n \times n$ matrix, then its inverse A^{-1} is also invertible and

$$(A^{-1})^{-1} = A$$

Proof.

What we know about the inverse A^{-1} (from its definition) is that

$$AA^{-1} = I_n \text{ and } A^{-1}A = I_n$$

In words, the inverse of A is a matrix with the property that A times the inverse and the inverse times A are both the identity matrix I_n .

But looking at the two equations again and focussing on A^{-1} rather than on A , we see that there is a matrix which when multiplied by A^{-1} on the right or the left gives I_n . And that matrix is A . So A^{-1} has an inverse and the inverse is A . □

This is an example of a mathematical proof of a theorem. Theorems are the mathematical analogue of the laws of science (the second law of thermodynamics, Boyle's law, Newton's Laws and so on), but there is a difference.

In Science, a law is a way of summarising observations made in experiments or a conjecture based on some principle. The law should then be checked with further experiments and if it checks out, it becomes accepted as a fact.

Such laws have to have a precise statement for them to work. Roughly they say that given a certain situation, some particular effect or result will happen. Sometimes the "certain situation" may be somewhat idealised. So one may interpret the law as saying that the effect or result should be very close to the observed effect or result if the situation is almost exactly valid.

Sometimes we don't even know what approximations or idealisations we are making in stating our physical laws and when we do new experiments in different arenas we discover our laws are only approximations and don't in fact hold as stated in general.

In mathematics we expect our results to be exactly true as long as our assumptions hold, furthermore our laws, once proven, will always be true as stated.

Theorem

Products of invertible matrices are invertible, and the inverse of the product is the product of the inverses taken in the reverse order.

In more mathematical language, if A and B are two invertible $n \times n$ matrices, then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.

Proof.

Start with any two invertible $n \times n$ matrices A and B , and look at

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AI_nA^{-1} = AA^{-1} = I_n$$

And look also at

$$(B^{-1}A^{-1})(AB) = B^{-1}(B^{-1}B)A = B^{-1}I_nB = B^{-1}B = I_n$$

This shows that $B^{-1}A^{-1}$ is the inverse of AB (because multiplying AB by $B^{-1}A^{-1}$ on the left or the right gives I_n). So it shows that $(AB)^{-1}$ exists, or in other words that AB is invertible, as well as showing the formula for $(AB)^{-1}$. □

Theorem (equivalent ways to see that a matrix is invertible)

Let A be an $n \times n$ (square) matrix.

The following are equivalent statements about A , meaning that if any one of them is true, then the others have to be true as well. (And if one is not true, the others must all be not true.)

- (a) *A is invertible (has an inverse)*
- (b) *the equation $A\mathbf{x} = \mathbf{0}$ (where \mathbf{x} is an unknown $n \times 1$ column matrix, $\mathbf{0}$ is the $n \times 1$ zero column) has only the solution $\mathbf{x} = \mathbf{0}$*
- (c) *the reduced row echelon form for A is I_n*
- (d) *A can be written as a product of elementary matrices*

In principle there are a lot of things to prove in the theorem.

Starting with any one of the 4 items, assuming that that statement is valid for a given $n \times n$ matrix A , we should provide a line of logical reasoning why all the other items have to be also true about that same A . We don't do this by picking examples of matrices A , but by arguing about a matrix where we don't specifically know any of the entries.

But we then have 4 times 3 little proofs to give, 12 proofs in all. So it would be long even if each individual proof is very easy.

There is a trick to reduce the number of proofs from 12 to only 4. We prove a cyclical number of steps

$$\begin{array}{ccc} (a) & \Rightarrow & (b) \\ \uparrow & & \downarrow \\ (d) & \Leftarrow & (c) \end{array}$$

The idea then is to prove 4 things only

- (a) \Rightarrow (b) In this step we assume only that statement (a) is true about A , and then we show that (b) must also be true.
- (b) \Rightarrow (c) In this step we assume only that statement (b) is true about A , and then we show that (c) must also be true.
- (c) \Rightarrow (d) Similarly we assume (c) and show (d) must follow.
- (d) \Rightarrow (a) In the last step we assume (d) and show that (a) must follow.

When we have done this we will be able to deduce all the statements from any one of the 4.

Starting with (say) the knowledge that (c) is a true statement the third step above shows that (d) must be true. Then the next step tells us (a) must be true and the first step then says (b) must be true. In other words, starting at any point around the ring

$$\begin{array}{ccc} (a) & \Rightarrow & (b) \\ \uparrow & & \downarrow \\ (d) & \Leftarrow & (c) \end{array}$$

(or at any corner of the square) we can work around to all the others.

Link 1

- (a) A is invertible (has an inverse)
- (b) the equation $A\mathbf{x} = \mathbf{0}$ (where \mathbf{x} is an unknown $n \times 1$ column matrix, $\mathbf{0}$ is the $n \times 1$ zero column) has only the solution $\mathbf{x} = \mathbf{0}$

Proof: (a) \Rightarrow (b).

Assume A is invertible and A^{-1} is its inverse.

Consider the equation $A\mathbf{x} = \mathbf{0}$ where \mathbf{x} is some $n \times 1$ matrix and $\mathbf{0}$ is the $n \times 1$ zero matrix. Multiply both sides by A^{-1} on the left to get

$$\begin{aligned}A^{-1}A\mathbf{x} &= A^{-1}\mathbf{0} \\I_n\mathbf{x} &= \mathbf{0} \\ \mathbf{x} &= \mathbf{0}\end{aligned}$$

Therefore $\mathbf{x} = \mathbf{0}$ is the only possible solution of $A\mathbf{x} = \mathbf{0}$.



Link 2

- (b) the equation $A\mathbf{x} = \mathbf{0}$ (where \mathbf{x} is an unknown $n \times 1$ column matrix, $\mathbf{0}$ is the $n \times 1$ zero column) has only the solution $\mathbf{x} = \mathbf{0}$
- (c) the reduced row echelon for A is I_n

Proof: (b) \Rightarrow (c).

Assume now that $\mathbf{x} = \mathbf{0}$ is the only possible solution of $A\mathbf{x} = \mathbf{0}$.

That means that when we solve $A\mathbf{x} = \mathbf{0}$ by using Gauss-Jordan elimination on the augmented matrix

$$\left[\begin{array}{c|c} & \begin{matrix} 0 \\ 0 \\ \vdots \\ 0 \end{matrix} \end{array} \right] = [A \quad | \quad \mathbf{0}]$$

we can't end with free variables.

We know that we must end up with a reduced row echelon form that has as many leading ones as there are unknowns. Since we are dealing with n equations in n unknowns, that means A row reduces to I_n . □

Link 3

- (c) the reduced row echelon form for A is I_n
- (d) A can be written as a product of elementary matrices

Proof: (c) \Rightarrow (d).

Suppose now that A row reduces to I_n .

Write down an elementary matrix for each row operation we need to row-reduce A to I_n . Say they are E_1, E_2, \dots, E_k . Recall that all elementary matrices have inverses. So we must have

$$\begin{aligned}E_k E_{k-1} \dots E_2 E_1 A &= I_n \\E_k^{-1} E_k E_{k-1} \dots E_2 E_1 A &= E_k^{-1} I_n \\I_n E_{k-1} \dots E_2 E_1 A &= E_k^{-1} \\E_{k-1} \dots E_2 E_1 A &= E_k^{-1} \\E_{k-1}^{-1} E_{k-1} \dots E_2 E_1 A &= E_{k-1}^{-1} E_k^{-1} \\E_{k-2} \dots E_2 E_1 A &= E_{k-1}^{-1} E_k^{-1}\end{aligned}$$

So, when we keep going in this way, we end up with

$$A = E_1^{-1} E_2^{-1} \dots E_{k-1}^{-1} E_k^{-1}$$

So we have (d) because inverses of elementary matrices are again elementary matrices.



Link 4

- (d) A can be written as a product of elementary matrices
- (a) A is invertible (has an inverse)

Proof: (d) \Rightarrow (a).

If A is a product of elementary matrices,

$$A = E_k E_{k-1} \dots E_2 E_1$$

then

$$\begin{aligned} E_1^{-1} E_2^{-1} \dots E_{k-1}^{-1} E_k^{-1} A &= E_1^{-1} E_2^{-1} \dots E_{k-1}^{-1} E_k^{-1} E_k E_{k-1} \dots E_2 E_1 \\ &= E_1^{-1} E_2^{-1} \dots E_{k-1}^{-1} I_n E_{k-1} \dots E_2 E_1 \\ &= E_1^{-1} E_2^{-1} \dots E_{k-2}^{-1} I_n E_{k-2} \dots E_2 E_1 \\ &= I_n \end{aligned}$$

we can use the earlier facts that the inverse of the product is the product of the inverses in the reverse order to show that A is invertible and its inverse is

$$A^{-1} = E_1^{-1} E_2^{-1} \dots E_{k-1}^{-1} E_k^{-1}$$

So we get (a).



Summary

Hence we have shown every link in our proof

$$\begin{array}{ccc} (a) & \Rightarrow & (b) \\ \Uparrow & & \Downarrow \\ (d) & \Leftarrow & (c) \end{array}$$

as required by the theorem.

Theorem

If A and B are two $n \times n$ matrices and if $AB = I_n$, then $BA = I_n$.

Proof.

The idea is to apply the previous theorem to the matrix B rather than to A . Consider the equation $B\mathbf{x} = \mathbf{0}$ (where \mathbf{x} and $\mathbf{0}$ are $n \times 1$). Multiply that equation by A on the left to get

$$\begin{aligned}AB\mathbf{x} &= B\mathbf{0} \\ I_n\mathbf{x} &= \mathbf{0} \\ \mathbf{x} &= \mathbf{0}\end{aligned}$$

So $\mathbf{x} = \mathbf{0}$ is the only possible solution of $B\mathbf{x} = \mathbf{0}$.

That means B satisfies condition (b) of the previous Theorem. Thus by the theorem, B^{-1} exists. Multiply the equation $AB = I_n$ by B^{-1} on the right to get

$$\begin{aligned}ABB^{-1} &= I_nB^{-1} \\ AI_n &= B^{-1} \\ A &= B^{-1}\end{aligned}$$

So, we get

$$BA = BB^{-1} = I_n.$$



Special matrices

There are matrices that have a special form that makes calculations with them much easier than the same calculations are as a rule.

Diagonal matrices

For square matrices (that is $n \times n$ for some n) $A = (a_{ij})_{i,j=1}^n$ we say that A is a *diagonal matrix* if $a_{ij} = 0$ whenever $i \neq j$. Thus in the first few cases $n = 2, 3, 4$ diagonal matrices look like

$$\begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix}$$

$$\begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix}$$

$$\begin{bmatrix} a_{11} & 0 & 0 & 0 \\ 0 & a_{22} & 0 & 0 \\ 0 & 0 & a_{33} & 0 \\ 0 & 0 & 0 & a_{44} \end{bmatrix}$$

Examples with numbers

$$\begin{bmatrix} 4 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 13 \end{bmatrix}, \quad \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

(These are 3×3 examples.)

Diagonal matrices are easy to multiply

$$\begin{bmatrix} 4 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 \\ 0 & 12 & 0 \\ 0 & 0 & 4 \end{bmatrix} = \begin{bmatrix} -4 & 0 & 0 \\ 0 & 60 & 0 \\ 0 & 0 & 24 \end{bmatrix}$$

$$\begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} \begin{bmatrix} b_{11} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & b_{33} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & 0 & 0 \\ 0 & a_{22}b_{22} & 0 \\ 0 & 0 & a_{33}b_{33} \end{bmatrix}$$

The idea is that all that needs to be done is to multiply the corresponding diagonal entries to get the diagonal entries of the product (which is again diagonal).

Based on this we can rather easily figure out how to get the inverse of a diagonal matrix. For example if

$$A = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 6 \end{bmatrix}$$

then

$$A^{-1} = \begin{bmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{5} & 0 \\ 0 & 0 & \frac{1}{6} \end{bmatrix}$$

because if we multiply these two diagonal matrices we get the identity.

We could also figure out A^{-1} the usual way, by row-reducing $[A \mid I_3]$. The calculation is actually quite easy. Starting with

$$[A \mid I_3] = \left[\begin{array}{ccc|ccc} 4 & 0 & 0 & 1 & 0 & 0 \\ 0 & 5 & 0 & 0 & 1 & 0 \\ 0 & 0 & 6 & 0 & 0 & 1 \end{array} \right]$$

we just need to divide each row by something to get to

$$[A \mid I_3] = \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 1 & 0 & 0 & \frac{1}{5} & 0 \\ 0 & 0 & 1 & 0 & 0 & \frac{1}{6} \end{array} \right]$$

Special Matrices

Upper triangular matrices

This is the name given to square matrices where all the non-zero entries are on or above the diagonal.

A 4×4 example is

$$A = \begin{bmatrix} 4 & -3 & 5 & 6 \\ 0 & 3 & 7 & -9 \\ 0 & 0 & 0 & 6 \\ 0 & 0 & 0 & -11 \end{bmatrix}$$

Another way to express it is that all the entries that are definitely below the diagonal have to be 0. Some of those on or above the diagonal can be zero also. They can all be zero and then we would have the zero matrix, which would be technically upper triangular. All diagonal matrices are also counted as upper triangular.

The precise statement then is that an $n \times n$ matrix

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

is *upper triangular* when

$$a_{ij} = 0 \text{ whenever } i > j.$$

It is fairly easy to see that if A and B are two $n \times n$ upper triangular matrices, then

the sum $A + B$ and the product AB

are both upper triangular.

Example

Let us consider the upper triangular matrices:

$$A = \begin{bmatrix} 3 & 4 & 5 & 6 \\ 0 & 7 & 8 & 9 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 & -2 & 5 \\ 0 & 4 & -1 & 5 \\ 0 & 0 & 8 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Then the sum is

$$A + B = \begin{bmatrix} 4 & 6 & 3 & 11 \\ 0 & 11 & 7 & 14 \\ 0 & 0 & 9 & 4 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

while the product is

$$AB = ?$$

Example

Let us consider the upper triangular matrices:

$$A = \begin{bmatrix} 3 & 4 & 5 & 6 \\ 0 & 7 & 8 & 9 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 & -2 & 5 \\ 0 & 4 & -1 & 5 \\ 0 & 0 & 8 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Then the sum is

$$A + B = \begin{bmatrix} 4 & 6 & 3 & 11 \\ 0 & 11 & 7 & 14 \\ 0 & 0 & 9 & 4 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

while the product is

$$AB = \begin{bmatrix} 3 & 22 & 30 & 51 \\ 0 & 28 & 57 & 60 \\ 0 & 0 & 8 & 4 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

Also inverting upper triangular matrices is relatively painless because the Gaussian elimination parts of the process are almost automatic.

As an example, we look at the (upper triangular)

$$A = \begin{bmatrix} 3 & 4 & 5 & 6 \\ 0 & 7 & 8 & 9 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

We should row reduce

$$\left[\begin{array}{cccc|cccc} 3 & 4 & 5 & 6 & : & 1 & 0 & 0 & 0 \\ 0 & 7 & 8 & 9 & : & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 & : & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 & : & 0 & 0 & 0 & 1 \end{array} \right]$$

and the first few steps are to divide row 1 by 3, row 2 by 7 and row 4 by 3, to get

$$\left[\begin{array}{cccc|cccc} 1 & \frac{4}{3} & \frac{5}{3} & 2 & : & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 1 & \frac{8}{7} & \frac{9}{7} & : & 0 & \frac{1}{7} & 0 & 0 \\ 0 & 0 & 1 & 2 & : & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & : & 0 & 0 & 0 & \frac{1}{3} \end{array} \right]$$

This is then already in row echelon form and to get the inverse we need to get to reduced row echelon form (starting by clearing out above the last leading 1, then working back up). The end result should be

$$\left[\begin{array}{cccc|ccc} 1 & 0 & 0 & 0 & : & \frac{1}{3} & -\frac{4}{21} & -\frac{1}{7} & 0 \\ 0 & 1 & 0 & 0 & : & 0 & \frac{1}{7} & -\frac{8}{7} & \frac{1}{3} \\ 0 & 0 & 1 & 0 & : & 0 & 0 & 1 & -\frac{2}{3} \\ 0 & 0 & 0 & 1 & : & 0 & 0 & 0 & \frac{1}{3} \end{array} \right]$$

It is quite easy to see that an upper triangular matrix is invertible exactly when the diagonal entries are all nonzero. Another way to express this same thing is that the product of the diagonal entries should be nonzero.

It is also easy enough to see from the way the above calculation of the inverse worked out that the inverse of an upper triangular matrix will be again upper triangular.

Strictly upper triangular matrices

These are matrices which are upper triangular and also have all zeros on the diagonal. This can also be expressed by saying that there should be zeros on and below the diagonal.

The precise statement then is that an $n \times n$ matrix

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

is *strictly upper triangular* when

$$a_{ij} = 0 \text{ whenever } i \geq j.$$

Example

An example is

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix}$$

This matrix is certainly *not* invertible. To be invertible we need *each* diagonal entry to be nonzero while his matrix is at the other extreme — all diagonal entries are 0.

For this matrix

$$A^2 = AA = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and

$$A^3 = AA^2 = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \mathbf{0}$$

In fact this is not specific to the example. Every strictly upper triangular matrix

$$A = \begin{bmatrix} 0 & a_{12} & a_{13} \\ 0 & 0 & a_{23} \\ 0 & 0 & 0 \end{bmatrix}$$

has

$$A^2 = \begin{bmatrix} 0 & 0 & a_{12}a_{23} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ and } A^3 = \mathbf{0}.$$

In general an $n \times n$ strictly upper triangular matrix A has $A^n = \mathbf{0}$. This is an example of nilpotent matrix.

Definition

A square matrix A is called *nilpotent* if some power of A is the zero matrix.

This shows again a significant difference between ordinary multiplication of numbers and matrix multiplication. It is not true that $AB = \mathbf{0}$ means that A or B has to be $\mathbf{0}$.

The question of which matrices have an inverse is also more complicated than it is for numbers. Every nonzero number has a reciprocal, but there are many nonzero matrices that fail to have an inverse.

Given we've discussed upper triangular matrices (and strictly upper triangular) we might also discuss lower triangular matrices. In fact we could repeat most of the same arguments for them, with small modifications, but there is an operation to flip one to the other

Transposes

The **transpose** of a matrix is what you get by writing the rows as columns. More precisely, we can take the transpose of any $m \times n$ matrix A ,

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

by writing the entries of the first row $a_{11}, a_{12}, \dots, a_{1n}$ down the first column of the transpose, the entries $a_{21}, a_{22}, \dots, a_{2n}$ of the second row down the second column, etc. We get a new $n \times m$ matrix, which we denote A^t

$$A^t = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & & \ddots & \\ a_{1n} & a_{2n} & \cdots & a_{nm} \end{bmatrix}$$

Another way to describe it is that the (i, j) entry of the transpose is a_{ji} = the (j, i) entry of the original matrix.

Examples are

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}, \quad A^t = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{bmatrix}$$

$$A = \begin{bmatrix} 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix}, \quad A^t = \begin{bmatrix} 4 & 7 & 10 \\ 5 & 8 & 11 \\ 6 & 9 & 12 \end{bmatrix}$$

Yet another way to describe it is that it is the matrix got by reflecting the original matrix in the “diagonal” line, or the line where $i = j$ (row number = column number).

So we see that if we start with an upper triangular

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix}$$

then the transpose

$$A^t = \begin{bmatrix} a_{11} & 0 & 0 \\ a_{12} & a_{22} & 0 \\ a_{13} & a_{23} & a_{33} \end{bmatrix}$$

is *lower triangular* (has all nonzero entries on or below the diagonal).

Facts about transposes

- (i) $A^{tt} = A$ (transpose twice gives back the original matrix)
- (ii) $(A + B)^t = A^t + B^t$ (if A and B are matrices of the same size).
- (iii) $(kA)^t = kA^t$ for A a matrix and k a scalar.
- (iv) $(AB)^t = B^t A^t$ (the transpose of a product is the product of the transposes taken in the reverse order — provided the product AB makes sense).

So if A is $m \times n$ and B is $n \times p$, then $(AB)^t = B^t A^t$. Note that B^t is $p \times n$ and A^t is $n \times m$ so that $B^t A^t$ makes sense and is a $p \times m$ matrix, the same size as $(AB)^t$.

The proof requires a bit of notation and organisation so we won't do it in detail. Here is what we would need to do just for the 2×2 case.

Take any two 2×2 matrices, which we write out as

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

Then

$$A^t = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}, \quad B^t = \begin{bmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{bmatrix}$$

and we can find

$$\begin{aligned} AB &= \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}, \\ (AB)^t &= \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{21}b_{11} + a_{22}b_{21} \\ a_{11}b_{12} + a_{12}b_{22} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix} \end{aligned}$$

while

$$B^t A^t = \begin{bmatrix} b_{11}a_{11} + b_{21}a_{12} & b_{11}a_{21} + b_{21}a_{22} \\ b_{12}a_{11} + b_{22}a_{12} & b_{12}a_{21} + b_{22}a_{22} \end{bmatrix} = (AB)^t$$

A final property of transposes is:

- (v) If A is an invertible square matrix then A^t is also invertible and $(A^t)^{-1} = (A^{-1})^t$ (the inverse of the transpose is the same as the transpose of the inverse).

This is easy to see: Let A be an invertible $n \times n$ matrix. We know from the definition of A^{-1} that

$$AA^{-1} = I_n \text{ and } A^{-1}A = I_n$$

Take transposes of both equations to get

$$(A^{-1})^t A^t = I_n^t = I_n \text{ and } A^t (A^{-1})^t = I_n^t = I_n$$

Therefore we see that A^t has an inverse and that the inverse matrix is $(A^{-1})^t$.

Lower triangular matrices

We can use the transpose to transfer what we know about upper triangular matrices to lower triangular ones. Let us take 3×3 matrices as an example, though what we say will work similarly for $n \times n$.

If

$$A = \begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

is lower triangular, then its transpose

$$A^t = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ 0 & a_{22} & a_{32} \\ 0 & 0 & a_{33} \end{bmatrix}$$

is upper triangular. So we know that A^t has an inverse exactly when the product of its diagonal entries

$$a_{11}a_{22}a_{33} \neq 0$$

But that is the same as the product of the diagonal entries of A . So lower triangular matrices have an inverse exactly when the product of the diagonal entries is nonzero.

- ① Another thing we know is that $(A^t)^{-1}$ is again upper triangular. So $((A^t)^{-1})^t = (A^{-1})^{tt} = A^{-1}$ is lower triangular. Thus the inverse of a lower triangular matrix is again lower triangular (if it exists).
- ② Using $(AB)^t = B^t A^t$ we can show that the product of lower triangular matrices is again lower triangular. As $B^t A^t$ is a product of upper triangulars it is upper triangular and then $AB = ((AB)^t)^t = (B^t A^t)^t$ is the transpose of an upper triangular and so AB is lower triangular.
- ③ Finally, we could use transposes to show that strictly *lower triangular matrices* have to be nilpotent (some power of them is the zero matrix).

Or we could figure these out by working them out in more or less the same way as we did for the strictly upper triangular case.

Symmetric matrices

A matrix A is called *symmetric* if $A^t = A$.

Symmetric matrices must be square as the transpose of an $m \times n$ matrix is $n \times m$. So if $m \neq n$, then A^t and A are not even the same size — and so they could not be equal.

One way to say what ‘symmetric’ means for a square matrix is to say that the numbers in positions symmetrical around the diagonal are equal.

Examples are

$$\begin{bmatrix} -3 & 1 & -1 \\ 1 & 14 & 33 \\ -1 & 33 & 12 \end{bmatrix}, \quad \begin{bmatrix} 5 & 1 & 2 & -3 \\ 1 & 11 & 25 & 0 \\ 2 & 25 & -41 & 6 \\ -3 & 0 & 6 & -15 \end{bmatrix}$$

Trace of a matrix

The *trace* of a matrix is a number that is quite easy to compute and which partly characterises the matrix. It is the sum of the diagonal entries.

So

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

has

$$\text{trace}(A) = 1 + 4 = 5$$

and

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ -7 & -8 & -6 \end{bmatrix}$$

has

$$\text{trace}(A) = 1 + 5 + (-6) = 0$$

For 2×2

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \Rightarrow \text{trace}(A) = a_{11} + a_{22}$$

and for 3×3 ,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \Rightarrow \text{trace}(A) = a_{11} + a_{22} + a_{33}$$

Properties of the trace

- (i) $\text{trace}(A + B) = \text{trace}(A) + \text{trace}(B)$ (if A and B are both $n \times n$)
- (ii) $\text{trace}(kA) = k \text{trace}(A)$ (if k is a scalar and A is a square matrix)
- (iii) $\text{trace}(A^t) = \text{trace}(A)$ (if A is any square matrix)
- (iv) $\text{trace}(AB) = \text{trace}(BA)$ for A and B square matrices of the same size (or even for A $m \times n$ and B an $n \times m$ matrix).

The last property is the only one that is at all hard to check out. The others are pretty easy to see.

To check the last one, we should write out the entries of A and B , work out the diagonal entries of AB and the sum of them. Then work out the sum of the diagonal entries of BA and their sum. Rearranging we should see we get the same answer.

Let us look at the simple example of 2×2 matrices.

In the 2×2 we would take

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

(without saying what the entries are specifically) and look at

$$AB = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & * \\ * & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}$$

$$\Rightarrow \text{trace}(AB) = a_{11}b_{11} + a_{12}b_{21} + a_{21}b_{12} + a_{22}b_{22}$$

(where the asterisks mean something goes there but we don't have to figure out what goes in those places).

$$BA = \begin{bmatrix} b_{11}a_{11} + b_{12}a_{21} & * \\ * & b_{21}a_{12} + b_{22}a_{22} \end{bmatrix}$$

$$\Rightarrow \text{trace}(BA) = b_{11}a_{11} + b_{12}a_{21} + b_{21}a_{12} + b_{22}a_{22}$$

Hence we see that

$$\text{trace}(AB) = \text{trace}(BA) .$$

The idea is that now we know this is always going to be true for 2×2 matrices A and B . To check it for all the sizes is not really that much more difficult but it requires a bit of notation to be able to keep track of what is going on.

Application of matrices to graphs

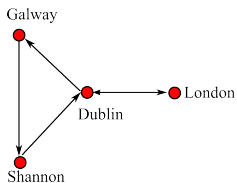
Graph theory is a subject that is somehow abstract, but at the same time rather close to applications.

Fairly simple minded examples would be an intercity rail network (nodes would be stations and the edges would correspond to the existence of a direct line from one station to another), or an airline route network, or an ancestral tree graph, or a telecommunications network.

Mathematically a graph is something that has vertices (also known as nodes) and edges (also called paths) joining some of the nodes to some others.

For our situation, we consider something more like an airline network (joining different airports by direct flights), but we will take account of the fact that some airports might not be connected by flights that go direct in both directions.

Here is a route network for a start-up airline that has two routes it flies. One goes Dublin \rightarrow London and London \rightarrow Dublin, while another route makes a round trip Dublin \rightarrow Galway \rightarrow Shannon \rightarrow Dublin.



The arrows on the edges mean that this is an example of a *directed graph*. Here we allow one-way edges and bi-directional edges between nodes (or vertices) of the graph, which we draw by indicating arrows.

To get the *vertex matrix* for a graph like this, we first number or order the vertices, for instance

| | |
|---------|---|
| Dublin | 1 |
| London | 2 |
| Galway | 3 |
| Shannon | 4 |

and then we make a matrix, a 4×4 matrix in this case since there are 4 vertices, according to the following rules. The entries of the matrix are either 0 or 1. All diagonal entries are 0. The (i, j) entry is 1 if there is a direct edge from vertex i to vertex j (in that direction).

That is we form the matrix M with entries:

$$m_{ij} = \begin{cases} 1, & P_i \rightarrow P_j \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

So in our example graph, the vertex matrix is

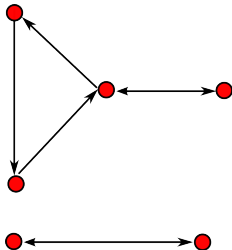
$$M = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

For instance the first row is 0 1 1 0 because there is a direct link $1 \rightarrow 2$ and $1 \rightarrow 3$, but no direct link $1 \rightarrow 4$ (Dublin to Galway is not directly linked).

In more precise terms: A **directed graph** is a finite set of elements, $\{P_1, P_2, \dots, P_n\}$ together with a collection of ordered pairs (P_i, P_j) of distinct elements of this set with no ordered pair being repeated.

The elements of this set are called **vertices**, and the ordered pairs are called **directed edges**. We also use the notation $P_i \rightarrow P_j$ to indicate a directed edge in a graph.

In general a graph can have separate components.



In science very complicated graphs and networks commonly appear: e.g. **metabolic pathways**

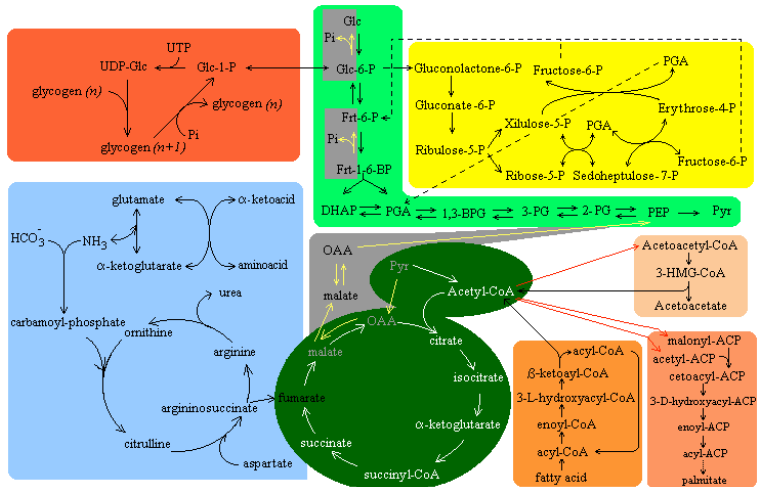


Image from <http://www2.ufp.pt/~pedros/bq/integration.htm>

In science very complicated graphs and networks commonly appear: e.g. metabolic pathways

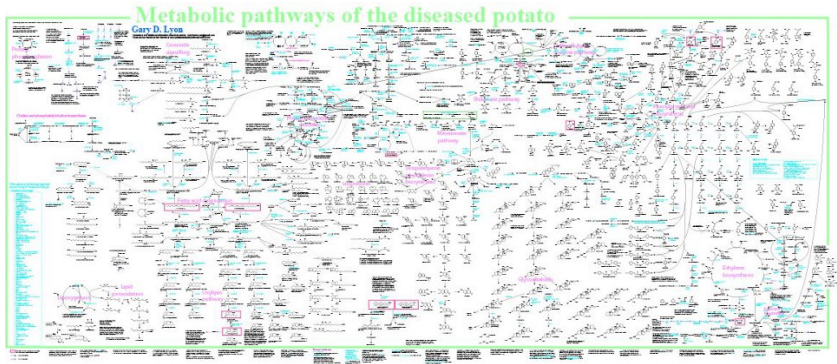


Image from <http://potatometabolicpathways.webs.com>

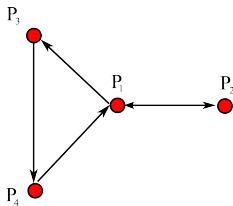
Having a matrix description of graphs is often very useful. Here is one result that makes a connection to matrix multiplication.

Theorem

If M is the vertex matrix of a directed graph, then the entries of M^2 give the numbers of 2-step (or 2-hop) connections.

More precisely, the (i, j) entry of M^2 gives the number of ways to go from vertex i to vertex j with exactly 2 steps (or exactly one intermediate vertex). Similarly M^3 gives the number of 3-step connections, and so on for higher powers of M .

Consider again the graph:



with the vertex matrix

$$M = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

In our example

$$M^2 = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

The diagonal 1's in the matrix correspond to the fact that there is a round trip Dublin \rightarrow London \rightarrow Dublin (or $1 \rightarrow 2 \rightarrow 1$) and also London \rightarrow Dublin \rightarrow London. The 1 in the top right corresponds to the connection Dublin \rightarrow Galway \rightarrow Shannon.

If we add $M + M^2$ we get nonzero entries in every place where there is a connection in 1 or 2 steps

$$M + M^2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

and the zeros off the diagonal there correspond to the connections that can't be made in either a direct connection or a 2-hop connection (which is Gaway \rightarrow London and London \rightarrow Shannon in our example). Although we don't see it here the numbers in the matrix M^2 can be bigger than 1 if there are two routes available using 2-hops.

Clique

Definition A subset of a directed graph is called a **clique** if it satisfies the following three conditions:

- ① The subset contains at least three vertices.
- ② For each pair of vertices P_i and P_j in the subset, both $P_i \rightarrow P_j$ and $P_j \rightarrow P_i$ are true.
- ③ The subset is as large as possible; that is, it is not possible to add another vertex to the subset and still satisfy condition (ii).

A clique is in some sense the largest possible subgroup of vertices that are maximally connected to each other.

For simple graphs we can find cliques by inspection but for more complicated diagrams it is useful to have a tool.

For a graph we can form the matrix S with the entries:

$$s_{ij} = \begin{cases} 1, & P_i \leftrightarrow P_j \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

This is similar to the vertex matrix but only has non-zero entries where both vertices are connected. That is $s_{ij} = 1$ if $m_{ij} = m_{ji} = 1$ but $s_{ij} = 0$ if $m_{ij} = 0$ and/or $m_{ji} = 0$.

We note that this matrix is automatically symmetric as

$$S = S^t .$$

We can then identify cliques by using the fact that if $s_{ij}^{(3)}$ are the entries of the matrix S^3 , then a vertex P_i belongs to some clique if and only if $s_{ii}^{(3)} \neq 0$.

Consider the vertex matrix

$$M = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix} \quad (3)$$

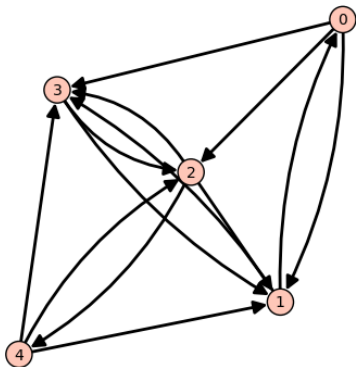
we can find the graph corresponding to the with a bit of work or by using sage ...

Graphs and Matrices

```
M = Matrix([[0, 1, 1, 1, 0],[1, 0, 0, 1, 0],[0, 1, 0, 1, 1],[0, 1, 1, 1, 0]]); M
```

```
[0 1 1 1 0]
[1 0 0 1 0]
[0 1 0 1 1]
[0 1 1 0 0]
[0 1 1 1 0]
```

```
G=DiGraph(M)
P = G.plot()
P.show( figsize=5)
```



Does this matrix have any cliques? The S matrix is

$$S = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \quad (4)$$

and so (can calculate using for example SAGE)

$$S^3 = \begin{pmatrix} 0 & 2 & 1 & 0 & 0 \\ 2 & 0 & 0 & 3 & 1 \\ 1 & 0 & 0 & 3 & 2 \\ 0 & 3 & 3 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 \end{pmatrix} \quad (5)$$

and as there are no non-zero diagonal elements there are no cliques.