ALGEBRA (SPRING TERM 2022-2023)

ZAQ COELHO

Contents

1. Introduction to Matrices

This is a brief example on how matrices are used in our daily lives.

1.1. *Example.* "Meal Planning":

	portions of rice	beans	sausages
meal 1	3	4	2
meal 2	5	3	5

Counting number of proteins and energy:

	units of protein	units of energy
portion of rice	1	5
bean	3	2
sausage	5	7

Hence the total number of proteins and energy contained in each meal is:

	Protein	Energy
meal 1	25	37
meal 2	39	66

One way to view the above example is to introduce variables m_1, m_2 for the meals and keep the variables r, b, s, P, E to denote rice, bean, sausage, protein and energy, respectively. We then have

$$\begin{aligned} m_1 &= 3r + 4b + 2s \\ m_2 &= 5r + 3b + 5s \\ r &= 1P + 5E \\ b &= 3P + 2E \\ s &= 5P + 7E\,. \end{aligned}$$

Date: March 2023.

Therefore we obtain

$$\begin{aligned} m_1 &= 3(1P+5E)+4(3P+2E)+2(5P+7E) = 25P+37E \\ m_2 &= 5(1P+5E)+3(3P+2E)+5(5P+7E) = 39P+66E\,. \end{aligned}$$

We can set out this calculation in a more elegant way:

$$\begin{pmatrix} 3 & 4 & 2 \\ 5 & 3 & 5 \end{pmatrix}\begin{pmatrix} 1 & 5 \\ 3 & 2 \\ 5 & 7 \end{pmatrix} = \begin{pmatrix} 25 & 37 \\ 39 & 66 \end{pmatrix}.$$

The numbers $25, 37, 39, 66$ are computed from the dot products of rows of the first matrix with columns of the second matrix, i.e.,

$$\begin{aligned} 25 &= \begin{pmatrix} 3 & 4 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix}, \quad & 37 &= \begin{pmatrix} 3 & 4 & 2 \end{pmatrix}\begin{pmatrix} 5 \\ 2 \\ 7 \end{pmatrix}, \\ 39 &= \begin{pmatrix} 5 & 3 & 5 \end{pmatrix}\begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix}, \quad & 66 &= \begin{pmatrix} 5 & 3 & 5 \end{pmatrix}\begin{pmatrix} 5 \\ 2 \\ 7 \end{pmatrix}. \end{aligned}$$

The matrix

$$\begin{pmatrix} 25 & 37 \\ 39 & 66 \end{pmatrix}$$

is called the *matrix product*.

1.2. *Example.* Composition of Linear Maps. Consider two maps $g\colon \mathbb{R}^2 \to \mathbb{R}^3$ and $f\colon \mathbb{R}^3 \to \mathbb{R}^2$ given by $g\colon (x,y) \mapsto (a,b,c)$ and $f\colon (a,b,c) \mapsto (\alpha,\beta)$ with

$$\begin{aligned} a &= x+5y \\ b &= 3x+2y \\ c &= 5x+7y \end{aligned}$$

and

$$\begin{aligned} \alpha &= 3a+4b+2c \\ \beta &= 5a+3b+5c\,. \end{aligned}$$

Then we can eliminate a,b,c to express (α,β) in terms of (x,y). This gives the composition $f \circ g\colon (x,y) \mapsto (\alpha,\beta)$ with

$$\begin{aligned} \alpha &= 25x+37y \\ \beta &= 39x+66y\,. \end{aligned}$$

In matrix theory the above example can be expressed as

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 1 & 5 \\ 3 & 2 \\ 5 & 7 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

and

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 3 & 4 & 2 \\ 5 & 3 & 5 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

and hence

$$\begin{aligned}\begin{pmatrix}\alpha\\ \beta\end{pmatrix} &= \begin{pmatrix}3 & 4 & 2\\ 5 & 3 & 5\end{pmatrix}\begin{pmatrix}1 & 5\\ 3 & 2\\ 5 & 7\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}\\ &= \begin{pmatrix}25 & 37\\ 39 & 66\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix} = \begin{pmatrix}25x+37y\\ 39x+66y\end{pmatrix}.\end{aligned}$$

How to think about the matrix product.

Recall the dot product of vectors: if $\mathbf{x} = (x_1, x_2, x_3)$ and $\mathbf{y} = (y_1, y_2, y_3)$ then

$$\mathbf{x}\cdot\mathbf{y} = x_1y_1 + x_2y_2 + x_3y_3 .$$

Then think of

$$\begin{pmatrix}3 & 4 & 2\\ 5 & 3 & 5\end{pmatrix}\begin{pmatrix}1 & 5\\ 3 & 2\\ 5 & 7\end{pmatrix}$$

as

$$\begin{pmatrix}-\mathbf{a}_1-\\ -\mathbf{a}_2-\end{pmatrix}\begin{pmatrix}| & |\\ \mathbf{b}_1 & \mathbf{b}_2\\ | & |\end{pmatrix}$$

where $-\mathbf{a}_1-$ means $\begin{pmatrix}3 & 4 & 2\end{pmatrix}$ and

$$\begin{pmatrix}|\\ \mathbf{b}_1\\ |\end{pmatrix} \quad \text{means} \quad \begin{pmatrix}1\\ 3\\ 5\end{pmatrix} \quad \text{etc.}$$

Then

$$AB = \begin{pmatrix}\mathbf{a}_1\cdot\mathbf{b}_1 & \mathbf{a}_1\cdot\mathbf{b}_2\\ \mathbf{a}_2\cdot\mathbf{b}_1 & \mathbf{a}_2\cdot\mathbf{b}_2\end{pmatrix}.$$

The generalisation is if

$$A = \begin{pmatrix}a_{11} & \cdots & a_{1n}\\ \vdots & \ddots & \vdots\\ a_{m1} & \cdots & a_{mn}\end{pmatrix}$$

is an $m \times n$ matrix and

$$B = \begin{pmatrix}b_{11} & \cdots & b_{1p}\\ \vdots & \ddots & \vdots\\ b_{n1} & \cdots & b_{np}\end{pmatrix}$$

is an $n \times p$ matrix, then the product $C = AB$ is the $m \times p$ matrix

$$C = \begin{pmatrix}c_{11} & \cdots & c_{1p}\\ \vdots & \ddots & \vdots\\ c_{m1} & \cdots & c_{mp}\end{pmatrix}$$

where

$$c_{ik} = \sum_{j=1}^{n} a_{ij}b_{jk}$$

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

1.3. *Example.* We have

$$\begin{pmatrix}1 & -3\\ 2 & 4\end{pmatrix}\begin{pmatrix}-2 & 4\\ 0 & 3\end{pmatrix}=\begin{pmatrix}-2 & -5\\ -4 & 20\end{pmatrix}$$

and

$$\begin{pmatrix}-2 & 4\\ 0 & 3\end{pmatrix}\begin{pmatrix}1 & -3\\ 2 & 4\end{pmatrix}=\begin{pmatrix}6 & 22\\ 6 & 12\end{pmatrix}.$$

Note that the above shows that (square) matrices do not necessarily commute. Consider also

$$\begin{pmatrix}1 & -3 & 2\end{pmatrix}\begin{pmatrix}7\\ 3\\ 6\end{pmatrix}=10$$

and

$$\begin{pmatrix}7\\ 3\\ 6\end{pmatrix}\begin{pmatrix}1 & -3 & 2\end{pmatrix}=\begin{pmatrix}7 & -21 & 14\\ 3 & -9 & 6\\ 6 & -18 & 12\end{pmatrix}.$$

Remark. We can only multiply an $m\times n$ matrix A by a $p\times q$ matrix B to form AB if $n=p$. Similarly, BA is only defined if $q=m$.

In particular, to be able to define both AB and BA we require $q=m$ and $n=p$. In other words, A and B can only be multiplied both ways if A is $m\times n$ and B is $n\times m$.

2. Matrix Algebra

Denote the set of $m\times n$ matrices with elements which are real numbers by $\mathbb{R}^{m\times n}$ (note: everything in this section generalises to matrices with elements being complex numbers, and we write $\mathbb{C}^{m\times n}$).

Then a typical matrix $A\in\mathbb{R}^{m\times n}$ can be written

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

We say a typical element of A is A_{ij} ($i=1,\dots,m$; $j=1,\dots,n$). (Sometimes we write a_{ij}.)

Operations with Matrices and Definitions

Given $A,B\in\mathbb{R}^{m\times n}$, with typical elements A_{ij}, B_{ij}, we define their sum

$$C=A+B$$

by the rule $C_{ij}=A_{ij}+B_{ij}$ for $i=1,\dots,m$; $j=1,\dots,n$, i.e., $(A+B)_{ij}=A_{ij}+B_{ij}$.

2.1. *Example.*

$$\begin{pmatrix}3 & 4\\ 0 & -2\end{pmatrix}+\begin{pmatrix}1 & 7\\ -4 & 9\end{pmatrix}=\begin{pmatrix}4 & 11\\ -4 & 7\end{pmatrix},$$

$$\begin{pmatrix}1 & 2 & 3\end{pmatrix}+\begin{pmatrix}-2 & 1 & 5\end{pmatrix}=\begin{pmatrix}-1 & 3 & 8\end{pmatrix},$$

$$\begin{pmatrix}4\\ -3\\ 7\end{pmatrix}+\begin{pmatrix}1\\ 2\\ 5\end{pmatrix}=\begin{pmatrix}5\\ -1\\ 12\end{pmatrix}.$$

2.2. **Proposition.** *Matrix addition is commutative, i.e.,*

$$A+B=B+A.$$

Proof. We have

$$(A+B)_{ij} = A_{ij} + B_{ij} = B_{ij} + A_{ij} = (B+A)_{ij}$$

for all $i = 1, \ldots, m$; $j = 1, \ldots, n$. $\square$

We define the zero matrix $\mathcal{O}$ as the matrix that have all its elements equal to 0. For instance, in $\mathbb{R}^{2\times 2}$ we have

$$\mathcal{O} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} .$$

So as an example,

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} .$$

In fact, the following is true.

2.3. **Proposition.** *For every matrix $A \in \mathbb{R}^{m\times n}$ we have*

$$\mathcal{O} + A = A + \mathcal{O} = A .$$

Proof.

$$(\mathcal{O} + A)_{ij} = 0 + A_{ij} = A_{ij} = A_{ij} + 0 = (A + \mathcal{O})_{ij} ,$$

for all $i = 1, \ldots, m$; $j = 1, \ldots, n$. $\square$

Let $A \in \mathbb{R}^{m\times n}$ and $r \in \mathbb{R}$. We define the product rA by the rule

$$(rA)_{ij} = rA_{ij} ,$$

for all $i = 1, \ldots, m$; $j = 1, \ldots, n$.

2.4. **Proposition.** *Scalar multiplication satisfy the following properties:*

(i) $(rs)A = r(sA)$,

(ii) $(r+s)A = rA + sA$,

(iii) $r(A+B) = rA + rB$.

for all $A, B \in \mathbb{R}^{m\times n}$ and $r, s \in \mathbb{R}$.

Proof. (i) We have

$$[(rs)A]_{ij} = rsA_{ij} = r[sA]_{ij} .$$

(ii) We compute

$$[(r+s)A]_{ij} = (r+s)A_{ij} = rA_{ij} + sA_{ij} = [rA]_{ij} + [sA]_{ij} .$$

(iii) We have

$$[r(A+B)]_{ij} = r(A+B)_{ij} = [rA]_{ij} + [rB]_{ij} = [rA + rB]_{ij} ,$$

for all $i = 1, \ldots, m$; $j = 1, \ldots, n$. $\square$

2.5. *Example.* For instance,

$$3 \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 3 & 6 \\ 9 & 12 \end{pmatrix} .$$

Note: we define $-A$ to mean the matrix $(-1)A$. So, for instance,

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

Note also that $1A = A$, $\mathcal{O}A = \mathcal{O}$, $r\mathcal{O} = \mathcal{O}$, etc.

Matrix Multiplication

As we have seen earlier, if $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$ then AB is defined by the rule

$$(AB)_{ik} = \sum_{j=1}^{n} A_{ij} B_{jk}$$

for all $i = 1, \dots, m$; $k = 1, \dots, p$.

2.6. **Proposition.** *Matrix multiplication satisfies the following properties:*

(i) *if $A \in \mathbb{R}^{m\times n}$ and $B, C \in \mathbb{R}^{n\times p}$ then*

$$A(B + C) = AB + AC\,,$$

and if $A, B \in \mathbb{R}^{m\times n}$ and $C \in \mathbb{R}^{n\times p}$ then

$$(A + B)C = AC + BC\,;$$

(ii) *if $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$ and $r, s \in \mathbb{R}$ then*

$$(rA)(sB) = rs(AB)\,;$$

(iii) *if $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times p}$ and $C \in \mathbb{R}^{p\times r}$ then*

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

Proof. (i) We have

$$[A(B + C)]_{ik} = \sum_{j=1}^{n} A_{ij}(B_{jk} + C_{jk}) = \sum_{j=1}^{n} A_{ij} B_{jk} + \sum_{j=1}^{n} A_{ij} C_{jk} = [AB]_{ik} + [AC]_{ik}\,,$$

and the second part is similar.
(ii) We compute

$$[(rA)(sB)]_{ik} = \sum_{j=1}^{n} (rA)_{ij}(sB)_{jk} = rs \sum_{j=1}^{n} A_{ij} B_{jk} = rs[AB]_{ik}\,.$$

(iii) Here we have

$$[(AB)C]_{ik} = \sum_{j=1}^{p} (AB)_{ij} C_{jk} = \sum_{j=1}^{p} \left(\sum_{q=1}^{n} A_{iq} B_{qj} \right) C_{jk} = \sum_{q=1}^{n} A_{iq} \left(\sum_{j=1}^{p} B_{qj} C_{jk} \right) = [A(BC)]_{ik}\,,$$

where we used the fact that these were finite sums and it does not matter which order we do the sums. □

2.7. *Example.* We mention again that even if both sides make sense we do not have in general $AB = BA$. For instance,

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

gives

$$AB = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad BA = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Square Matrices

Given two matrices, $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$ we have seen that, to be able to multiply them to form AB, we need $n = p$. So in order to be able to multiply A by itself we need to have $m = n$. Matrices in $\mathbb{R}^{n\times n}$ are called *square matrices*.

2.8. *Example.* We could have

$$\begin{pmatrix} 1 & 0 \\ -4 & 3 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -4 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -16 & 9 \end{pmatrix}.$$

Square matrices have extra properties since we can consider their products as $AA = A^2$, $A^3 = AAA$, etc.

There are some special square matrices that play an important role. A square matrix D in $\mathbb{R}^{n\times n}$ is called a *diagonal matrix*, if $D_{ij} = 0$ whenever $i \neq j$. Hence a diagonal matrix D has the form

$$\begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}.$$

Diagonal matrices all commute and their product is easy to do. For instance,

$$\begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 5 \end{pmatrix} = \begin{pmatrix} 6 & 0 \\ 0 & 20 \end{pmatrix}.$$

In case diagonal elements are all equal to 1, i.e., $d_1 = d_2 = \cdots = d_n = 1$ then we get the $(n \times n)$ *identity matrix* I_n, which has the form

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Note that

$$I_m A = A I_n = A,$$

for any matrix $A \in \mathbb{R}^{m\times n}$.

2.9. **Definition.** Let $A \in \mathbb{R}^{m\times n}$, then we define the *transpose* A^T of A to be the matrix in $\mathbb{R}^{n\times m}$ with

$$A^T_{ij} = A_{ji},$$

for all $i = 1, \ldots, n$; $j = 1, \ldots, m$.

For instance,

$$\begin{pmatrix} 1 & 5 \\ 4 & 2 \end{pmatrix}^T = \begin{pmatrix} 1 & 4 \\ 5 & 2 \end{pmatrix}$$

and

$$\begin{pmatrix} 1 & 3 & 4 \\ -2 & 7 & -5 \end{pmatrix}^T = \begin{pmatrix} 1 & -2 \\ 3 & 7 \\ 4 & -5 \end{pmatrix}.$$

Note that $(A^T)^T = A$.

2.10. **Theorem.** *Given $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$ then*

$$(AB)^T = B^T A^T.$$

Proof. We have

$$(AB)^T_{ik} = (AB)_{ki} = \sum_{j=1}^{n} A_{kj} B_{ji} = \sum_{j=1}^{n} A^T_{jk} B^T_{ij} = \sum_{j=1}^{n} B^T_{ij} A^T_{jk} = (B^T A^T)_{ik},$$

for all $i = 1, \ldots, p$; $j = 1, \ldots, m$. Hence $(AB)^T = B^T A^T$. □

3. General Solution of Systems of Linear Equations

The general problem is how to solve equations of the form

$$\begin{cases} A_{11}x_1 + A_{12}x_2 + \ldots + A_{1n}x_n = b_1 \\ A_{21}x_1 + A_{22}x_2 + \ldots + A_{2n}x_n = b_2 \\ \qquad\qquad\qquad\qquad\quad \vdots \;\; = \vdots \\ A_{m1}x_1 + A_{m2}x_2 + \ldots + A_{mn}x_n = b_m \end{cases}$$

and in matrix notation

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix},$$

where A_{ij}'s and b_i's are given, and the x_j's are unknowns.

Many things may happen. Let us see worked out examples.

(a) There exists a unique solution. For instance,

$$\begin{cases} 5x - 2y &= 4 \\ 3x + y &= 9 \end{cases}$$

has unique solution $x = 2$, $y = 3$.

(b) There exist many solutions. For instance,

$$x + y = 3\,.$$

The general solution is $x = 3 - \lambda$, $y = \lambda$, where $\lambda \in \mathbb{R}$ is a *parameter*.

Another example is

$$\begin{cases} x + y &= 3 \\ 3x + 3y &= 9\,. \end{cases}$$

This is essentially the same as the previous example, since the second equation adds no new information. However, note that this sort of thing can be disguised. Consider for instance,

$$\begin{cases} x + y &= 3 \\ 3x + 3y &= 9 \\ z &= 2\,, \end{cases}$$

with general solution $x = 3 - \lambda$, $y = \lambda$, $z = 2$. This is equivalent to the following system of equations (after row operations):

$$\begin{matrix} r_1 - r_3 \\ r_2 + 3r_3 \\ r_3 + 2r_1 \end{matrix} \begin{cases} x + y - z &= 1 \\ 3x + 3y + 3z &= 15 \\ 2x + 2y + z &= 8\,, \end{cases}$$

The general solution is still $x = 3 - \lambda$, $y = \lambda$, $z = 2$, but it is not as clear as before.

(c) The equations might be inconsistent and there are no solutions. For instance,

$$0x = 3\,,$$

or

$$\begin{cases} x + y &= 4 \\ 2x + 2y &= 9\,. \end{cases}$$

How do we solve this problem in general?

Example Problem

For which values of c are these equations consistent? If or when they are consistent, find the general solution:

$$\begin{cases} x+3y+2z &= 1 \\ 2x+3y+4z &= 5 \\ x+y+2z &= c\,. \end{cases}$$

The strategy is the following (which generalises to all cases, and is called "Gaussian Elimination").

(1) Write the problem is matrix notation:

$$\begin{pmatrix} 1 & 3 & 2 \\ 2 & 3 & 4 \\ 1 & 1 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 5 \\ c \end{pmatrix}.$$

(2) Construct the "Augmented Matrix":

$$\left(\begin{array}{ccc|c} 1 & 3 & 2 & 1 \\ 2 & 3 & 4 & 5 \\ 1 & 1 & 2 & c \end{array}\right)$$

(3) Do "row operations" to get the matrix into *Echelon Form*. In three dimensions the Echelon forms are the following types of matrices.

$$\text{(A)} \begin{pmatrix} \star & \square & \square \\ 0 & \star & \square \\ 0 & 0 & \star \end{pmatrix}, \quad \text{(B)} \begin{pmatrix} \star & \square & \square \\ 0 & \star & \square \\ 0 & 0 & 0 \end{pmatrix}, \quad \text{(C)} \begin{pmatrix} \star & \square & \square \\ 0 & 0 & \star \\ 0 & 0 & 0 \end{pmatrix}, \quad \text{(D)} \begin{pmatrix} \star & \square & \square \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The entry $\star$ is a non-zero entry and they are called *pivots*. The entry $\square$ is any real number, and 0 is the zero entry.

Row Operations

We can perform the following basic operations on rows.

- We can multiply any row by a non-zero constant.
- We can replace any row by its sum with another row.
- We can interchange rows.

We need to do the last above step if the top left corner of the matrix is zero. If that is the case, then swap the top row with another row with a non-zero first element.

Now we apply these principles to our augmented matrix. The first step is to have a set of row operations to achieve a column of zeros beneath the top "1" in the left column. That is:

$$\begin{array}{r} \\ r_2 \to r_2 - 2r_1 \\ r_3 \to r_3 - r_1 \end{array} \left(\begin{array}{ccc|c} 1 & 3 & 2 & 1 \\ 0 & -3 & 0 & 3 \\ 0 & -2 & 0 & c-1 \end{array}\right).$$

Repeat this strategy in the second column to get a zero beneath the entry "-3".

$$\begin{array}{r} \\ \\ r_3 \to r_3 - (2/3)r_2 \end{array} \left(\begin{array}{ccc|c} 1 & 3 & 2 & 1 \\ 0 & -3 & 0 & 3 \\ 0 & 0 & 0 & c-1-(2/3)3 \end{array}\right).$$

The last entry is: $c-1-(2/3)3 = c-3$. This is the Echelon form (B), explained above.

Now we can interpret the bottom line as

$$0x + 0y + 0z \; (= 0) = c - 3\,.$$

So the equations are consistent only if $c = 3$. The next step is to set $c = 3$ and continue to use row operations to get the matrix in *Reduced Echelon Form*.

Reduced Echelon Form

The reduced Echelon forms in three dimensions are the following types of matrices.

$$\text{(A)} \begin{pmatrix} \star & 0 & 0 \\ 0 & \star & 0 \\ 0 & 0 & \star \end{pmatrix}, \quad \text{(B)} \begin{pmatrix} \star & 0 & \square \\ 0 & \star & \square \\ 0 & 0 & 0 \end{pmatrix}, \quad \text{(C)} \begin{pmatrix} \star & \square & 0 \\ 0 & 0 & \star \\ 0 & 0 & 0 \end{pmatrix}.$$

We must have zeros above as well as below the pivots. Returning to our example we have

$$r_1 \to r_1 + r_2 \; \left(\begin{array}{ccc|c} 1 & 0 & 2 & 4 \\ 0 & -3 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{array} \right).$$

Note: In *strict reduced echelon form* it is required that the pivots are equal to 1. So in the above example we would have the strict form:

$$\left(\begin{array}{ccc|c} 1 & 0 & 2 & 4 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{array} \right).$$

We can now translate back into "equation" notation, to get

$$\begin{cases} x + 2z & = 4 \\ \quad -3y & = 3 \\ \qquad 0 & = 0\,. \end{cases}$$

Working from the bottom up, we obtain

$$y = -1 \quad \text{and} \quad x = 4 - 2z\,.$$

Hence the general solution is

$$x = 4 - 2\lambda, \quad y = -1, \quad z = \lambda,$$

where $\lambda \in \mathbb{R}$ is a parameter. This finishes the example problem.

The General Case

In the general case, we can always use row operations to get the matrix A into Echelon form. We start with

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

where $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{b}\mathbb{R}^m$. Here $A \in \mathbb{R}^{m\times n}$. The vector $\mathbf{x}$ are the unknowns and $\mathbf{b}$ is the independent vector, which is given.

We construct the Augmented matrix and perform row operations:

$$(A\,|\,\mathbf{b}) \quad \xrightarrow{\text{row operations}} \quad (C\,|\,\mathbf{d}),$$

where the latter Augmented matrix has the following form:

$$\left(\begin{array}{ccccccc|c} C_{11} & & & & & & & d_1 \\ 0 & \cdots & C_{2\,k_2} & & & & & d_2 \\ \vdots & \vdots & \vdots & & & & & \vdots \\ 0 & \cdots & 0 & \cdots & C_{r\,k_r} & & & d_r \\ 0 & \cdots & \cdots & \cdots & 0 & \cdots & 0 & d_{r+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & d_m \end{array}\right),$$

where $1 = k_1 < k_2 < \ldots < k_r$, and $C_{i\,k_i}$ are the non-zero pivots. The number r is called the *rank* of A (also known as the "row rank" of A).

3.1. **Theorem.** *The rank does not depend on the choices of row operations used to get A into Echelon form.*

Note that $r \le m$. Also note that $r \le n$. In fact, the only way that r could be equal to n is if $k_1 = 1, k_2 = 2, \ldots, k_n = n$ and $m \ge n$.

If $r < m$ (as in the above example) then we require $d_{r+1}, \ldots, d_m$ all to vanish or the equations would be inconsistent.

So assume $r < m$ (and then, $d_{r+1} = \ldots = d_m = 0$), we can proceed further with row operations to get the equations in *Reduced Echelon form*.

$$(C\,|\,\mathbf{d}) \quad \overset{\text{further row operations}}{\longrightarrow} \quad (E\,|\,\mathbf{f}),$$

where the latter has the following form:

$$\left(\begin{array}{ccccccc|c} 1 & & 0 & & 0 & & & f_1 \\ 0 & \cdots & 1 & & 0 & & & f_2 \\ \vdots & \vdots & \vdots & & 0 & & & \vdots \\ 0 & \cdots & 0 & \cdots & 1 & & & f_r \\ 0 & \cdots & \cdots & \cdots & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 0 \end{array}\right).$$

This means that we have columns of zeros above each pivot "1", but the other elements above non-pivots are not necessarily zero.

Translating this back into equations (from bottom up) we get

$$x_{k_r} + \sum_{j=k_r+1}^{n} E_{rj} x_j = f_r \quad \Rightarrow \quad x_{k_r} = f_r - \sum_{j=k_r+1}^{n} E_{rj} x_j \,,$$

$$x_{k_{r-1}} + \sum_{\substack{j=k_{r-1}+1 \\ j\ne k_r}}^{n} E_{r-1,j} x_j = f_{r-1} \quad \Rightarrow \quad x_{k_{r-1}} = f_{r-1} - \sum_{\substack{j=k_{r-1}+1 \\ j\ne k_r}}^{n} E_{r-1,j} x_j \,,$$

and finally,

$$x_1 + \sum_{\substack{j=2 \\ j\notin\{k_2,\ldots,k_r\}}}^{n} E_{1,j} x_j = f_1 \quad \Rightarrow \quad x_1 = f_1 - \sum_{\substack{j=2 \\ j\notin\{k_2,\ldots,k_r\}}}^{n} E_{1,j} x_j \,.$$

So, when $d_{r+1} = \ldots = d_m = 0$ (to avoid inconsistency) we get a solution in which the r quantities $x_1, x_{k_2}, \ldots, x_{k_r}$ are determined in terms of $f_1, \ldots, f_r$ (i.e., in terms of $b_1, \ldots, b_m$, provided they lead to $d_{r+1}, \ldots, d_m$ all zero) and any x_i where $i \in \{1, \ldots, n\} \setminus \{1 = k_1, \ldots, k_r\}$, which can be thought of as $(n - r)$ *parameters*.

Conclusions for the general system of linear equations $A\mathbf{x} = \mathbf{b}$

(1) (Existence Question) For $A\mathbf{x} = \mathbf{b}$ to have a solution for all $\mathbf{b} \in \mathbb{R}^m$ we need $r = m$. Otherwise we get one or more rows of zeros in $(C|\mathbf{d})$ and for some choices of $\mathbf{b}$, the equations are inconsistent. [The consistent choices form an r-dimensional subspace of the m-dimensional set of all $\mathbf{b}$.]

(2) (Uniqueness Question) For $A\mathbf{x} = \mathbf{b}$ to have a unique solution, we need $r = n$. Otherwise, if a solution exists at all, there are many solutions, labelled by the $(n - r)$ parameters. Note that these free parameters span a subspace of $\mathbb{R}^n$ with dimension $n - r$.

(3) In particular, if $r < n$, there is an $(n - r)$ parameter family of solutions to $A\mathbf{x} = \mathbf{0}$. And conversely, if $r = n$ then the only solution to $A\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$, i.e., $x_1 = \ldots = x_n = 0$.

(4) When $r = n$ and the solution is unique, then

$$x_1 = f_1, \cdots, x_n = f_n\,,$$

which means $\mathbf{x} = \mathbf{f}^*$, where, if $m > n$ then $\mathbf{f}^*$ is the n-dimensional column vector whose elements are the same as the first n elements of $\mathbf{f}$, and the last $m - n$ elements being assumed to be all zero.

(5) For the solution to always exist (for all $\mathbf{b}$) and be unique, we need $r = n$ and $r = m$. Hence we need $m = n$ and the matrix A must be a square matrix.

4. The Matrix Inverse

Given $A \in \mathbb{R}^{m\times n}$, when does there exist a matrix $A^{-1} \in \mathbb{R}^{n\times m}$ such that

$$A^{-1}A = I_n \quad \text{and} \quad AA^{-1} = I_m\,?$$

To answer this question we notice that if such an inverse A^{-1} exists, then we can solve $A\mathbf{x} = \mathbf{b}$ for any $\mathbf{b} \in \mathbb{R}^m$, by putting $\mathbf{x} = A^{-1}\mathbf{b}$; and then from the second equation, $A\mathbf{x} = AA^{-1}\mathbf{b} = I_m\mathbf{b} = \mathbf{b}$.

Moreover, the solution would be unique, since if $A\mathbf{x}_1 = \mathbf{b}$ and $A\mathbf{x}_2 = \mathbf{b}$, then $A(\mathbf{x}_2 - \mathbf{x}_1) = \mathbf{0}$. And hence, multiplying on the left by A^{-1} gives $\mathbf{0} = A^{-1}A(\mathbf{x}_2 - \mathbf{x}_1) = \mathbf{x}_2 - \mathbf{x}_1$, which implies $\mathbf{x}_1 = \mathbf{x}_2$.

As a consequence, by our conclusion (5) at the end of last section, in order to have an inverse, the matrix A must be a square matrix $m = n$ with rank $r = m (= n)$.

Let us see an example on how to find the inverse of a rank r square matrix in $\mathbb{R}^{r\times r}$ in practice.

4.1. *Example.* Consider

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

We know that $AA^{-1}\mathbf{b} = \mathbf{b}$ for all $\mathbf{b} \in \mathbb{R}^2$. So $A^{-1}\mathbf{b}$ ($= \mathbf{x}$ say) is the solution to $A\mathbf{x} = \mathbf{b}$. So finding A^{-1} amounts to solving $A\mathbf{x} = \mathbf{b}$ for general $\mathbf{b}$. Hence we do Gaussian elimination on the following Augmented matrix:

$$\left(\begin{array}{cc|c} 1 & 2 & b_1 \\ 3 & 4 & b_2 \end{array}\right) \longrightarrow$$

$$r_2 \to r_2 - 3r_1 \quad \left(\begin{array}{cc|c} 1 & 2 & b_1 \\ 0 & -2 & b_2 - 3b_1 \end{array}\right) \longrightarrow$$

$$r_1 \to r_1 + r_2 \left(\begin{array}{cc|c} 1 & 0 & b_2 - 2b_1 \\ 0 & -2 & b_2 - 3b_1 \end{array}\right) \longrightarrow$$

$$r_2 \to (-1/2)r_2 \left(\begin{array}{cc|c} 1 & 0 & b_2 - 2b_1 \\ 0 & 1 & (3/2)b_1 - (1/2)b_2 \end{array}\right).$$

Hence we have

$$A^{-1}\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} -2b_1 + b_2 \\ (3/2)b_1 - (1/2)b_2 \end{pmatrix},$$

which implies

$$A^{-1} = \begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix}.$$

We can check that we have the right answer because

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$\begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

A neater method is to do “Gaussian elimination” on the extended Augmented matrix:

$$\left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{array}\right) \longrightarrow$$

$$r_2 \to r_2 - 3r_1 \left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & -2 & -3 & 1 \end{array}\right) \longrightarrow$$

$$r_1 \to r_1 + r_2 \left(\begin{array}{cc|cc} 1 & 0 & -2 & 1 \\ 0 & -2 & -3 & 1 \end{array}\right) \longrightarrow$$

$$r_2 \to (-1/2)r_2 \left(\begin{array}{cc|cc} 1 & 0 & -2 & 1 \\ 0 & 1 & 3/2 & -1/2 \end{array}\right),$$

and the matrix on the right hand side of the “bar” is the inverse matrix of A:

$$A^{-1} = \begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix}.$$

The General Case

In the general case of an $n \times n$ matrix A, construct the extended Augmented matrix of $(A|I_n)$. Doing Gaussian elimination on the rows to transform A into the Reduced Echelon form makes it to become $(I_n|A^{-1})$. The reason is that if you have a generic vector $\mathbf{x} \in \mathbb{R}^n$ and an independent vector $\mathbf{b} \in \mathbb{R}^n$ then doing row operations corresponds to the manipulation of the system of n linear equations given by the matrix equation $A\mathbf{x} = \mathbf{b}$. The result after getting into the Reduced Echelon form is $\mathbf{x} = A^{-1}\mathbf{b}$.

5. Determinants

There is a remarkable function on the set of square matrices called the *determinant*. This is a function $\det\colon \mathbb{R}^{n\times n} \to \mathbb{R}$, i.e., $A \mapsto \det(A)$, with the property that A has rank n, if and only if, A is invertible, and if and only if, $\det(A) \neq 0$. The other important property is $\det(AB) = \det(A)\det(B)$ for any pair of square matrices $A, B \in \mathbb{R}^{n\times n}$. So the determinant is a multiplicative function. Also $\det(I_n) = 1$.

The correct way to define the determinant of a square matrix is using the theory of permutations, but here we will define the determinant recursively, by defining the determinant of an $n \times n$ matrix in terms of determinants of $(n-1)\times(n-1)$ matrices. First we consider the special case $n = 2$.

Case $n = 2$: Let

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

Then define the determinant of A by

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

We can prove that

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

and hence A^{-1} exists if and only if $\det(A) \neq 0$. Please check directly that if $\det(A) \neq 0$ then indeed the above expression for A^{-1} satisfies $A^{-1}A = AA^{-1} = I_2$.

We will denote the determinant by

$$\det(A) = \begin{vmatrix} a & b \\ c & d \end{vmatrix}.$$

5.1. *Example.* The determinant of

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

is

$$\det(A) = \begin{vmatrix} 1 & 3 \\ 2 & 4 \end{vmatrix} = 1 \times 4 - 3 \times 2 = -2 \neq 0\,,$$

so the matrix A is invertible. In fact,

$$A^{-1} = \frac{-1}{2} \begin{pmatrix} 4 & -3 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} -2 & 3/2 \\ 1 & -1/2 \end{pmatrix}.$$

Case $n = 3$: Let

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

we define the determinant of A by

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

In order to define the determinant in the general case, we need to define what is a *minor* of a square matrix A.

5.2. **Definition.** Let $A \in \mathbb{R}^{n\times n}$ be a square matrix. We call an $(n-1)\times(n-1)$ matrix A_{ij} a *minor* of A if A_{ij} is obtained from A by removing the row i and column j. So if we have

$$A = \begin{pmatrix} B & | & C \\ - & a_{ij} & - \\ D & | & E \end{pmatrix}$$

then the minor is

$$A_{ij} = \begin{pmatrix} B & C \\ D & E \end{pmatrix}.$$

5.3. **Definition.** Let $A \in \mathbb{R}^{n\times n}$ be a square matrix of size $n \geq 3$. We define the determinant of A by

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}),$$

where A_{ij} is the minor of A obtained from A by removing row i and column j.

The following result can be proved using determinants from the theory of permutations.

5.4. **Theorem.** *The above definition of determinant does not depend on the choice of row i to expand the determinant. We also have*

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}),$$

and this is also independent of the choice of column j to expand the determinant.

5.5. *Example.* Let

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

We can use any row or column to evaluate the determinant, but the simplest is to expand the determinant from column 2, we have

$$\begin{aligned} \det(A) &= \begin{vmatrix} 1 & 0 & 2 \\ -1 & -1 & 3 \\ 4 & 0 & 2 \end{vmatrix} \\ &= (-1)\,0 \begin{vmatrix} -1 & 3 \\ 4 & 2 \end{vmatrix} + (-1) \begin{vmatrix} 1 & 2 \\ 4 & 2 \end{vmatrix} + (-1)\,0 \begin{vmatrix} 1 & 2 \\ -1 & 3 \end{vmatrix} = (-1)(1\times 2 - 2\times 4) = 6. \end{aligned}$$

Also the following result can be proved using the definition of determinant from the theory of permutations.

5.6. **Theorem.** *The following summarise the properties of determinants when performing row operations:*

(i) *If one row of a square matrix A is multiplied by $\lambda \in \mathbb{R}$ then the determinant is multiplied by this scalar, i.e.,* $\det(A) \mapsto \lambda \det(A)$. *[As a consequence, if A is $n\times n$ then* $\det(\lambda A) = \lambda^n \det(A)$.*]*

(ii) *If two rows are interchanged then* $\det(A) \mapsto -\det(A)$.

(iii) *If two rows are equal then* $\det(A) = 0$.

(iv) *If one row is added to another,* $\det(A)$ *is unchanged. In fact, if any multiple of one row is added to another, then* $\det(A)$ *is unchanged.*

Remark: The above result is also true for columns (instead of rows).

The above result shows that we can perform row operations to transform the matrix into Echelon form and keep track of the changes in the determinant. Let A^E be the corresponding Echelon form of the matrix A, then A^E is an upper triangular matrix of the form

$$A^E = \begin{pmatrix} \mu_1 & \square & \cdots & \square \\ 0 & \mu_2 & \square & \vdots \\ \vdots & \vdots & \ddots & \square \\ 0 & 0 & \cdots & \mu_n \end{pmatrix},$$

where $\square$ means any entry (zero or non-zero) and also the diagonal elements μ_i may be zero or non-zero. We have the following result

5.7. **Theorem.** *The determinant of an upper triangular matrix A^E is the product of its diagonal elements, i.e.,*

$$\det(A^E) = \mu_1\mu_2\cdots\mu_n = \prod_{i=1}^{n} \mu_i \,.$$

Proof. The result is true for $n = 2$, since

$$\begin{vmatrix} \mu_1 & \square \\ 0 & \mu_2 \end{vmatrix} = \mu_1\mu_2 \,.$$

Assume the result is true for $n-1$ (with $n-1 \geq 2$) and evaluate the determinant from row n of the matrix A^E to get

$$\det(A^E) = \begin{vmatrix} \mu_1 & \square & \cdots & \square \\ 0 & \mu_2 & \square & \vdots \\ \vdots & \vdots & \ddots & \square \\ 0 & 0 & \cdots & \mu_n \end{vmatrix} = \mu_n \begin{vmatrix} \mu_1 & \square & \cdots & \square \\ 0 & \mu_2 & \square & \vdots \\ \vdots & \vdots & \ddots & \square \\ 0 & 0 & \cdots & \mu_{n-1} \end{vmatrix} = \mu_n \prod_{i=1}^{n-1} \mu_i = \prod_{i=1}^{n} \mu_i \,,$$

where we used the induction step. So the result is true in general. $\square$

Applying the above result we see that $\det(A) \neq 0$ is equivalent to $\det(A^E) \neq 0$, which in turn is equivalent to $\mu_i \neq 0$ for all i. In this case, we conclude that $\det(A) \neq 0$ if and only if the rank of A is n.

5.8. *Example.* Consider the matrix

$$A = \begin{pmatrix} 2 & 3 & 1 \\ -1 & 1 & 4 \\ 1 & 0 & 7 \end{pmatrix}.$$

Using row operations we get

$$\begin{matrix} \\ r_2 \to 2r_2 + r_1 \\ r_3 \to 2r_3 - r_1 \end{matrix} \begin{pmatrix} 2 & 3 & 1 \\ 0 & 5 & 9 \\ 0 & -3 & 13 \end{pmatrix} \quad \text{with } \det = 4\det(A)\,.$$

Then we have

$$\begin{matrix} \\ \\ r_3 \to r_3 + (3/5)r_2 \end{matrix} \begin{pmatrix} 2 & 3 & 1 \\ 0 & 5 & 9 \\ 0 & 0 & 92/5 \end{pmatrix} \quad \text{with } \det = 4\det(A)\,.$$

This implies that

$$4\det(A) = \frac{2 \times 5 \times 92}{5} = 184\,,$$

which gives

$$\det(A) = 46\,.$$

General Definition of Determinants

The general definition of determinants requires the theory of permutations and properties of the symmetric group.

A *permutation* of n objects is a bijection p from the finite set $\{1,2,\ldots,n\}$ to itself. So for each $i \in \{1,2,\ldots,n\}$ we have $p(i) \in \{1,2,\ldots,n\}$.

The inverse function is the inverse permutation, and the identity permutation is the map e given by $e(i) = i$ for all $i \in \{1,2,\ldots,n\}$.

The set of all permutations S_n is called the symmetric group. Note that the cardinality of S_n is $n! = n(n-1)\ldots 2.1$.

Write two copies of the numbers $1,2,\ldots,n$ above each other and draw a straight line segment from i to $p(i)$ for each i. Count the number of crossings of these segments. The permutation p is called even if the number of these crossings is even, and the permutation p is odd if the number of crossings is odd.

For instance, the permutation $1 \mapsto 2, 2 \mapsto 1, 3 \mapsto 3$ has one crossing, so it is odd; whereas the permutation $1 \mapsto 3, 2 \mapsto 1, 3 \mapsto 2$ has two crossings, so it is even. The identity permutation has no crossings, so it is even.

Define a sign function $\operatorname{sgn}(p)$ of the permutation p as $+1$ if the permutation is even, and $\operatorname{sgn}(p) = -1$ if the permutation is odd. Now we can give a general definition of the determinant.

5.9. **Definition.** Let A be an $n \times n$ matrix. The determinant of A is defined by

$$\det(A) = \sum_{p \in S_n} \operatorname{sgn}(p)\, a_{1,p(1)} a_{2,p(2)} \cdots a_{n,p(n)}\,,$$

where $a_{i,j}$ is the ij entry of the matrix A.

All the properties mentioned above can be proved using the above definition and the properties of permutations (i.e., the structure of the symmetric group S_n).

The above definition can also be used in the proof of the following important result.

5.10. **Theorem.** *Let A and B be $n \times n$ matrices, then*

$$\det(AB) = \det(A)\det(B)\,.$$

Cofactors and The Classical Adjoint

The main idea of this subsection is to describe Cramer's rule for the inverse of a square matrix.

5.11. **Definition.** For each pair ij define the *cofactor* c_{ij} of the $n \times n$ matrix A to be

$$c_{ij} = (-1)^{i+j} \det A_{ij}\,,$$

where A_{ij} is the minor of A obtained from A by removing row i and column j. The matrix $C = [c_{ij}]$ is called the *matrix of cofactors* of A.

Using our recursive definition of determinants and the properties of permutations, it can be shown the following couple of results.

5.12. **Theorem.** *For each fixed row i, we have*

$$\det(A) = \sum_{j=1}^{n} a_{ij} c_{ij} \,,$$

and for each fixed column j, we have

$$\det(A) = \sum_{i=1}^{n} a_{ij} c_{ij} \,.$$

5.13. *Example.* For instance,

$$\begin{vmatrix} 1 & 3 & 5 \\ 2 & 0 & 7 \\ -4 & 0 & 2 \end{vmatrix} = -3 \begin{vmatrix} 2 & 7 \\ -4 & 2 \end{vmatrix} = -3(32) = -96 \,.$$

Summing the third column to the second column, does not change the determinant, so we still have

$$\begin{vmatrix} 1 & 8 & 5 \\ 2 & 7 & 7 \\ -4 & 2 & 2 \end{vmatrix} = -96 \,.$$

5.14. **Definition.** Let C be the matrix of cofactors of A, the *adjoint* matrix of A is defined by

$$\operatorname{adj}(A) = C^T \,,$$

i.e., the transpose of the matrix of cofactors.

5.15. **Theorem.** *We have*

$$(AC^T)_{ik} = \sum_{j=1}^{n} a_{ij} (C^T)_{jk} = \sum_{j=1}^{n} a_{ij} c_{kj} = \begin{cases} \det(A) & \text{if } i = k \,, \\ 0 & \text{if } i \neq k \,. \end{cases}$$

We also have

$$(C^T A)_{j\ell} = \sum_{i=1}^{n} (C^T)_{ji} a_{i\ell} = \sum_{i=1}^{n} c_{ij} a_{i\ell} = \begin{cases} \det(A) & \text{if } j = \ell \,, \\ 0 & \text{if } j \neq \ell \,. \end{cases}$$

The above results shows that $A\operatorname{adj}(A) = \operatorname{adj}(A)A = \det(A)I_n$. Hence we have the following result known as *Cramer's Rule* for the inverse of a square matrix.

5.16. **Theorem.** *If A is a square matrix with non-zero determinant and* adj(A) *is the adjoint matrix of A then the inverse matrix of A is given by*

$$A^{-1} = \frac{\operatorname{adj}(A)}{\det(A)} \,.$$

So for instance, we can apply this to a 2×2 matrix

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

to give

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} ,$$

as we obtained earlier.

Geometric Interpretation of the Determinant

Let us start by considering an example.

5.17. *Example.* Take the matrix

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

Consider the multiplication of this matrix on the canonical basis of $\mathbb{R}^2$, $\{\mathbf{i}_1, \mathbf{i}_2\}$, given by

$$\mathbf{i}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{i}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The images of the basis vectors are

$$A\mathbf{i}_1 = \begin{pmatrix} 1 & -1 \\ 2 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad A\mathbf{i}_2 = \begin{pmatrix} 1 & -1 \\ 2 & 2 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}.$$

As a map from $\mathbb{R}^2$ to $\mathbb{R}^2$, the matrix A acts on vectors as

$$A\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x - y \\ 2x + 2y \end{pmatrix}.$$

The matrix A transforms the unit square with sides $\{\mathbf{i}_1, \mathbf{i}_2\}$ to the parallelogram with sides $\{A\mathbf{i}_1, A\mathbf{i}_2\}$. The area of the parallelogram is 4. So the area of the unit square ($= 1$) has been multiplied by the factor 4. This magical number is the modulus of the determinant of the matrix A.

This is in fact a general result. Volumes in higher dimensions are stretched by the modulus of the determinant of the matrix representing the linear transformation.

5.18. **Theorem.** *Let $\{\mathbf{i}_1, \mathbf{i}_2, \ldots, \mathbf{i}_n\}$ be the canonical basis of $\mathbb{R}^n$. Let A be an $n \times n$ matrix. The volume of the cuboid with sides $\{A\mathbf{i}_1, A\mathbf{i}_2, \ldots, A\mathbf{i}_n\}$ is $|\det(A)|$.*

6. Eigenvalues, Eigenvectors, and Diagonalisation

In this section we will see how the computation of eigenvalues and eigenvectors of a square matrix can be used to transform the matrix into a diagonal matrix (after a change of basis vectors).

First we discuss the problem of linear dependence and independence.

6.1. **Definition.** We say that the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ in $\mathbb{R}^n$ are *linear independent* if the only solution to the equation

$$\alpha_1 \mathbf{v}_1 + \cdots + \alpha_m \mathbf{v}_m = \mathbf{0}$$

is $\alpha_1 = \alpha_2 = \ldots = \alpha_m = 0$. If there is a non-zero solution to the above equation, then the vectors are called *linearly dependent*.

In the case the vectors are linearly dependent then there exists $\alpha_i \neq 0$ such that

$$\mathbf{v}_i = \frac{-1}{\alpha_i}(\alpha_1 \mathbf{v}_1 + \cdots + \alpha_{i-1}\mathbf{v}_{i-1} + \alpha_{i+1}\mathbf{v}_{i+1} + \cdots + \alpha_m \mathbf{v}_m).$$

Hence $\mathbf{v}_i$ is a *linear combination* of the other vectors.

Now let A be an $n \times n$ square matrix.

6.2. **Definition.** A column vector $\mathbf{v} \in \mathbb{R}^n$ is said to be an *eigenvector* of A with *eigenvalue* $\lambda \in \mathbb{R}$ (or $\mathbb{C}$) if $\mathbf{v} \neq \mathbf{0}$ and

$$A\mathbf{v} = \lambda\mathbf{v}.$$

We now discuss how to find eigenvectors and eigenvalues of a given matrix A. First consider the following simple example.

6.3. *Example.* Suppose we try to find $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and λ such that

$$\begin{pmatrix} 3 & 2 \\ 2 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} .$$

The first step is to write

$$\begin{pmatrix} 3 & 2 \\ 2 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} ,$$

which is equivalent to

$$\begin{pmatrix} 3-\lambda & 2 \\ 2 & -\lambda \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} . \tag{1}$$

For this to have a non-trivial solution, we require the rank of the matrix to be < 2, which means that the matrix is not invertible, i.e., the determinant is 0. Hence we must have

$$\begin{vmatrix} 3-\lambda & 2 \\ 2 & -\lambda \end{vmatrix} = 0 ,$$

which implies $\lambda^2 - 3\lambda - 4 = 0$, or $(\lambda - 4)(\lambda + 1) = 0$. So the solutions are $\lambda = 4$ or $\lambda = -1$. These are the two possible eigenvalues.

If $\lambda = 4$ the equation (1) becomes

$$\begin{pmatrix} -1 & 2 \\ 2 & -4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} ,$$

which implies

$$x_1 = 2x_2 .$$

So we have

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = c \begin{pmatrix} 2 \\ 1 \end{pmatrix} ,$$

with $c \neq 0$. This is the eigenvector which goes together with (or belongs to) the eigenvalue $\lambda = 4$.

Note that eigenvectors are only determined up to a non-zero multiple, since

$$A\mathbf{v} = \lambda\mathbf{v} \quad \Leftrightarrow \quad A(c\mathbf{v}) = \lambda(c\mathbf{v}) .$$

If $\lambda = -1$ then (1) becomes

$$\begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} ,$$

which implies

$$x_2 = -2x_1 .$$

So the eigenvector which goes with this eigenvalue is

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = c' \begin{pmatrix} -1 \\ 2 \end{pmatrix} ,$$

where $c' \neq 0$.

In general we expect a general matrix in $\mathbb{R}^{n\times n}$ to have n linearly independent eigenvectors (although the corresponding eigenvalues may be complex numbers). In fact, this is true with probability 1 for randomly chosen matrices. However, this is not always true as the next example shows.

6.4. *Example.* Consider the matrix

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} .$$

Then A has a single linearly independent eigenvector associated to the eigenvalue 0. The proof is:

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} ,$$

which implies

$$\begin{aligned} x_2 &= \lambda x_1 \\ 0 &= \lambda x_2 \, . \end{aligned}$$

The second equation means $\lambda = 0$ or $x_2 = 0$. In order to have a non-trivial solution we need $\lambda = 0$, which implies $x_2 = 0$ and x_1 is arbitrary. The conclusion is that there is only one eigenvector $c\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ associated to the eigenvalue $\lambda = 0$.

Now turning back to the general equation $A\mathbf{v} = \lambda \mathbf{v}$, we see that this is equivalent to

$$A\mathbf{v} = \lambda I_n \mathbf{v} \quad \Leftrightarrow \quad (A - \lambda I_n)\mathbf{v} = \mathbf{0} \, .$$

In order for this to have a non-trivial solution $\mathbf{v}$ we need the matrix $(A - \lambda I_n)$ not to be invertible, i.e., its determinant must be zero.

6.5. **Definition.** The *characteristic polynomial* of an $n \times n$ matrix A is defined by

$$p(\lambda) = \det(A - \lambda I_n) \, .$$

The characteristic polynomial is an n-degree polynomial, and its roots (real or complex) are the possible eigenvalues of A.

A sufficient condition for an $n \times n$ matrix A to have n linearly independent eigenvectors if that the characteristic polynomial $p(\lambda)$ have n distinct roots.

Also, a sufficient condition for a matrix $A \in \mathbb{R}^{n\times n}$ to have n linearly independent eigenvectors and for the eigenvalues to be all real is if A is a symmetric matrix (i.e., $A = A^T$).

6.6. **Definition.** Let $A, B \in \mathbb{R}^{n\times n}$, we say that A and B are *similar* if there exists an invertible matrix $M \in \mathbb{R}^{n\times n}$ such that $MB = AM$, i.e., $B = M^{-1}AM$.

6.7. **Theorem** (Diagonalisation)**.** *Suppose a matrix $A \in \mathbb{R}^{n\times n}$ has n linearly independent eigenvectors, i.e.,*

$$A\mathbf{x}_1 = \lambda_1 \mathbf{x}_1, \ldots, A\mathbf{x}_n = \lambda_n \mathbf{x}_n \, .$$

Then A is similar to the diagonal matrix

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} .$$

Proof. Since $\mathbf{x}_1, \dots, \mathbf{x}_n$ are linearly independent and there are n of them, they must span $\mathbb{R}^n$. So any arbitrary vector $\mathbf{x}$ must be a linear combination of these vectors. Hence there exists $c_1, \dots, c_n$ such that

$$\mathbf{x} = c_1\mathbf{x}_1 + \cdots + c_n\mathbf{x}_n\,, \tag{2}$$

and applying A we get

$$A\mathbf{x} = \lambda_1 c_1\mathbf{x}_1 + \cdots + \lambda_n c_n\mathbf{x}_n\,. \tag{3}$$

Writing the coordinates of the eigenvectors as column vectors give

$$\mathbf{x}_1 = \begin{pmatrix} x_1^1 \\ x_1^2 \\ \vdots \\ x_1^n \end{pmatrix}, \dots\,,\ \mathbf{x}_n = \begin{pmatrix} x_n^1 \\ x_n^2 \\ \vdots \\ x_n^n \end{pmatrix}.$$

Now define the matrix M as columns given by these column vectors, i.e.,

$$M = \begin{pmatrix} x_1^1 & x_2^1 & \cdots & x_n^1 \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^n & x_2^n & \cdots & x_n^n \end{pmatrix}.$$

The equation (2) may be rewritten as

$$\mathbf{x} = M\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}, \tag{4}$$

and the equation (3) may be written

$$A\mathbf{x} = M\begin{pmatrix} \lambda_1 c_1 \\ \vdots \\ \lambda_n c_n \end{pmatrix} = M\Lambda\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}, \tag{5}$$

where

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

Moreover, since $\mathbf{x}_1, \dots, \mathbf{x}_n$ are linearly independent, the matrix M has full rank and so is invertible.

Comparing (4) and (5) we see that

$$AM\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = M\Lambda\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}, \quad \text{for all } \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \in \mathbb{R}^n\,.$$

So we must have $AM = M\Lambda$, or $\Lambda = M^{-1}AM$. $\square$

Remark: To see this last step, note that if an $n \times n$ matrix N satisfies $N\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ for all $\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$, then N is the zero matrix. Now apply this to the matrix $N = AM - M\Lambda$.

6.8. *Example.* In our previous example where $A = \begin{pmatrix} 3 & 2 \\ 2 & 0 \end{pmatrix}$, we had the eigenvectors $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ associated to the eigenvalue $\lambda = 4$, and $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$ associated to the eigenvalue $\lambda = -1$. Hence we have

$$M = \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}, \quad M^{-1} = \frac{1}{5}\begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}.$$

Therefore

$$\begin{aligned} M^{-1}AM &= \frac{1}{5}\begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 3 & 2 \\ 2 & 0 \end{pmatrix}\begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix} \\ &= \frac{1}{5}\begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 8 & 1 \\ 4 & -2 \end{pmatrix} \\ &= \frac{1}{5}\begin{pmatrix} 20 & 0 \\ 0 & -5 \end{pmatrix} = \begin{pmatrix} 4 & 0 \\ 0 & -1 \end{pmatrix}. \end{aligned}$$

6.9. *Example.* Consider the following set of vectors:

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}, \quad \mathbf{v}_4 = \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}.$$

We claim that they are linearly independent. We try to find a solution $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ to the equation

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \alpha_3\mathbf{v}_3 + \alpha_4\mathbf{v}_4 = \mathbf{0}$$

and show that the only solution is

$$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0.$$

We have

$$\alpha_1\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \alpha_2\begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} + \alpha_3\begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix} + \alpha_4\begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

So this means

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

where we denote by M the matrix of coefficients.

If M is invertible then the only solution is the trivial solution (all α's equal to 0). And M is invertible if and only if the determinant of M is non-zero.

So we need to compute the determinant

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{vmatrix}$$

Subtract row 1 from rows 2, 3 and 4 to get

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

Developing the determinant from the first column gives

$$\det = 2(-4) - 2(4) = -8 - 8 = -16.$$

Interpreting M as column vectors of eigenvectors of a matrix A and diagonal matrix of eigenvalues D, for instance given by

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

We find A uniquely by $A = MDM^{-1}$.

We get from Excel

$$A = \frac{1}{-16}\begin{pmatrix} -28 & 12 & 4 & -4 \\ 12 & -28 & -4 & 4 \\ 4 & -4 & -28 & 12 \\ -4 & 4 & 12 & -28 \end{pmatrix}$$

Hence we have

$$A\mathbf{v}_1 = \mathbf{v}_1, \quad A\mathbf{v}_2 = \mathbf{v}_2, \quad A\mathbf{v}_3 = 2\mathbf{v}_3, \quad A\mathbf{v}_4 = 3\mathbf{v}_4 .$$

6.10. *Example.* This is a 3×3 example. Find the eigenvalues and eigenvectors of the matrix

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

We seek a non-zero vector $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ and λ such that

$$\begin{pmatrix} 2 & 1 & 0 \\ -1 & 0 & 1 \\ 1 & 3 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

We then write as

$$\begin{pmatrix} 2-\lambda & 1 & 0 \\ -1 & -\lambda & 1 \\ 1 & 3 & 1-\lambda \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \lambda \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

For non-trivial solutions, we require

$$\begin{vmatrix} 2-\lambda & 1 & 0 \\ -1 & -\lambda & 1 \\ 1 & 3 & 1-\lambda \end{vmatrix} = 0 \, .$$

Expand along the top row to get

$$(2-\lambda)\begin{vmatrix} -\lambda & 1 \\ 3 & 1-\lambda \end{vmatrix} - 1\begin{vmatrix} -1 & 1 \\ 1 & 1-\lambda \end{vmatrix} = 0 \, ,$$

which implies

$$(2-\lambda)(\lambda^2-\lambda-3)-(\lambda-2) = 0 \, ,$$

i.e.,

$$(\lambda-2)(\lambda^2-\lambda-2) = 0 \, ,$$

or

$$(\lambda-2)(\lambda-2)(\lambda+1) = 0 \, .$$

So the possible eigenvalues are $\lambda = -1$ or $\lambda = 2$ (with "algebraic multiplicity" 2).

If $\lambda = -1$ we consider the following augmented matrix

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

and after some row operations we arrive at the Echelon form

$$\left(\begin{array}{ccc|c} 1 & 3 & 2 & 0 \\ 0 & 4 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) .$$

This gives $x_3 = c$ (a parameter), and $x_2 = -3c/4$, and $x_3 = c/4$. Hence

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = c\begin{pmatrix} 1/4 \\ -3/4 \\ 1 \end{pmatrix} = c'\begin{pmatrix} 1 \\ -3 \\ 4 \end{pmatrix} .$$

If $\lambda = 2$ we have the augmented matrix

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

and after row operations we obtain the Echelon form

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

This gives $x_2 = 0$ and $x_1 = x_3$. Hence

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = d\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} .$$

Note: We might in another example have got 2-parameters (i.e., two independent eigenvectors with eigenvalue $\lambda = 2$), but in this example we didn't. We say that the "geometric multiplicity" of $\lambda = 2$ here is 1.

However, if we had any matrix A with invertible matrix M such that

$$M^{-1}AM = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix},$$

then we would have geometric multiplicity of the eigenvalue $\lambda = 2$ as 2.

Computing Powers of a Diagonalisable Matrix

If we have $A = M\Lambda M^{-1}$ with M invertible and Λ diagonal, we have

$$A^n = (M\Lambda M^{-1})(M\Lambda M^{-1})\cdots(M\Lambda M^{-1}) = M\Lambda^n M^{-1}\,.$$

As an example we have seen that

$$\begin{pmatrix} 3 & 2 \\ 2 & 0 \end{pmatrix} = M \begin{pmatrix} 4 & 0 \\ 0 & -1 \end{pmatrix} M^{-1}$$

where

$$M = \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}, \quad \text{and} \quad M^{-1} = \frac{1}{5}\begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}.$$

Therefore

$$\begin{aligned} \begin{pmatrix} 3 & 2 \\ 2 & 0 \end{pmatrix}^n &= \frac{1}{5}\begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 4^n & 0 \\ 0 & (-1)^n \end{pmatrix}\begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix} \\ &= \frac{1}{5}\begin{pmatrix} 4^{n+1} + (-1)^n & 2\cdot 4^n - 2\cdot(-1)^n \\ 2\cdot 4^n - 2\cdot(-1)^n & 4^n + 4\cdot(-1)^n \end{pmatrix}. \end{aligned}$$

Application to the Fibonacci sequence

This is an example to find the general term of a sequence defined by a recurrence relation.

6.11. *Example.* Consider the Fibonacci sequence:

$$1 \quad 1 \quad 2 \quad 3 \quad 5 \quad 8 \quad 13 \quad 21 \quad \cdots$$

This is defined as $u_1 = 1$, $u_2 = 1$, and $u_{n+2} = u_{n+1} + u_n$ for all $n \geq 1$. This can be written in matrix form as

$$\begin{pmatrix} u_{n+1} \\ u_{n+2} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} u_n \\ u_{n+1} \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

This implies

$$\begin{pmatrix} u_{n+1} \\ u_{n+2} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}^n \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

We solve this by diagonalising the matrix $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$. The possible eigenvalues are solutions of

$$\begin{vmatrix} -\lambda & 1 \\ 1 & 1-\lambda \end{vmatrix} = 0 \quad \Rightarrow \quad \lambda^2 - \lambda - 1 = 0\,.$$

This gives $\lambda = \dfrac{1 \pm \sqrt{5}}{2}$. Define $\lambda_+ = \dfrac{1+\sqrt{5}}{2}$ and $\lambda_- = \dfrac{1-\sqrt{5}}{2}$. The eigenvector associated to λ_+ is $\begin{pmatrix} 1 \\ \lambda_+ \end{pmatrix}$ and the eigenvector associated to λ_- is $\begin{pmatrix} 1 \\ \lambda_- \end{pmatrix}$.

Carrying on with the diagonalisation we have

$$\Lambda = \begin{pmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{pmatrix} = M^{-1} \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} M$$

where

$$M = \begin{pmatrix} 1 & 1 \\ \lambda_+ & \lambda_- \end{pmatrix} .$$

Now we have

$$M^{-1} = \frac{1}{\lambda_- - \lambda_+} \begin{pmatrix} \lambda_- & -1 \\ -\lambda_+ & 1 \end{pmatrix} = -\frac{1}{\sqrt{5}} \begin{pmatrix} \lambda_- & -1 \\ -\lambda_+ & 1 \end{pmatrix} .$$

Hence

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}^n = -\frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 1 \\ \lambda_+ & \lambda_- \end{pmatrix} \begin{pmatrix} \lambda_+^n & 0 \\ 0 & \lambda_-^n \end{pmatrix} \begin{pmatrix} \lambda_- & -1 \\ -\lambda_+ & 1 \end{pmatrix} .$$

Therefore we conclude that

$$\begin{pmatrix} u_{n+1} \\ u_{n+2} \end{pmatrix} = -\frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 1 \\ \lambda_+ & \lambda_- \end{pmatrix} \begin{pmatrix} \lambda_+^n & 0 \\ 0 & \lambda_-^n \end{pmatrix} \begin{pmatrix} \lambda_- & -1 \\ -\lambda_+ & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} .$$

After some computations and using the fact that $\lambda_+ + \lambda_- = 1$ and reading the first coordinate of the above vector we arrive at

$$u_n = \frac{1}{\sqrt{5}} \left[\left(\frac{1+\sqrt{5}}{2} \right)^n - \left(\frac{1-\sqrt{5}}{2} \right)^n \right] .$$

Application: Solution to systems of coupled linear first-order differential equations.

6.12. *Example.* Find the general solution to

$$\frac{dx}{dt} = 4x + 2y \, , \quad \frac{dy}{dt} = x + 3y \, .$$

Solution: Write in matrix form as

$$\frac{d}{dt} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{6}$$

Try to diagonalise the matrix of coefficients. The possible eigenvalues satisfy

$$\begin{vmatrix} 4-\lambda & 2 \\ 1 & 3-\lambda \end{vmatrix} = 0 \, ,$$

which implies

$$\lambda^2 - 7\lambda + 10 = 0 \quad \Rightarrow \quad (\lambda - 5)(\lambda - 2) = 0 \, .$$

Hence the possible eigenvalues are $\lambda = 5$ or $\lambda = 2$.

If $\lambda = 5$ we have

$$\begin{pmatrix} -1 & 2 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Rightarrow \quad \begin{pmatrix} x \\ y \end{pmatrix} = c \begin{pmatrix} 2 \\ 1 \end{pmatrix} .$$

If $\lambda = 2$ we have

$$\begin{pmatrix} 2 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Rightarrow \quad \begin{pmatrix} x \\ y \end{pmatrix} = d \begin{pmatrix} -1 \\ 1 \end{pmatrix} .$$

So take the matrix

$$M = \begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix} .$$

Although we don't need M^{-1} explicitly, we have

$$M^{-1} = \frac{1}{3}\begin{pmatrix} 1 & 1 \\ -1 & 2 \end{pmatrix}.$$

Then

$$M^{-1}\begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix} M = \begin{pmatrix} 5 & 0 \\ 0 & 2 \end{pmatrix}.$$

Now to solve (6) let

$$\begin{pmatrix} x \\ y \end{pmatrix} = M \begin{pmatrix} x' \\ y' \end{pmatrix},$$

or equivalently

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = M^{-1}\begin{pmatrix} x \\ y \end{pmatrix}.$$

Then

$$\frac{d}{dt}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix} M \begin{pmatrix} x' \\ y' \end{pmatrix}.$$

But

$$M \frac{d}{dt}\begin{pmatrix} x' \\ y' \end{pmatrix} = \frac{d}{dt} M \begin{pmatrix} x' \\ y' \end{pmatrix} = \frac{d}{dt}\begin{pmatrix} x \\ y \end{pmatrix}.$$

Hence we have

$$\frac{d}{dt}\begin{pmatrix} x' \\ y' \end{pmatrix} = M^{-1}\begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix} M \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 5 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} x' \\ y' \end{pmatrix}.$$

This implies that $\frac{d}{dt}x' = 5x'$ with general solution $x' = A\mathrm{e}^{5t}$, and $\frac{d}{dt}y' = 2y'$ with general solution $y' = B\mathrm{e}^{2t}$. Therefore, the general solution for x and y are

$$\begin{pmatrix} x \\ y \end{pmatrix} = M \begin{pmatrix} A\mathrm{e}^{5t} \\ B\mathrm{e}^{2t} \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} A\mathrm{e}^{5t} \\ B\mathrm{e}^{2t} \end{pmatrix} = \begin{pmatrix} 2A\mathrm{e}^{5t} - B\mathrm{e}^{2t} \\ A\mathrm{e}^{5t} + B\mathrm{e}^{2t} \end{pmatrix}.$$

Hence

$$\begin{cases} x(t) = 2A\mathrm{e}^{5t} - B\mathrm{e}^{2t}, \\ y(t) = A\mathrm{e}^{5t} + B\mathrm{e}^{2t}. \end{cases}$$

Example with no real eigenvectors: A real matrix may have no real eigenvectors, so it is not diagonalisable over $\mathbb{R}$, but it may be diagonalisable over the complex numbers $\mathbb{C}$.

6.13. *Example.* Consider the matrix

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Then the possible eigenvalues are solutions of

$$\begin{vmatrix} -\lambda & 1 \\ -1 & -\lambda \end{vmatrix} = 0,$$

which gives

$$\lambda^2 + 1 = 0, \quad \text{or } \lambda = \pm\sqrt{-1} = \pm\mathrm{i}.$$

If $\lambda = \mathrm{i}$ then to find eigenvectors we need to solve

$$\begin{pmatrix} -\mathrm{i} & 1 \\ -1 & -\mathrm{i} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

which implies $-\mathrm{i}x + y = 0$, or $y = \mathrm{i}x$. So there is no non-zero real solution. But there is a complex eigenvector:

$$\begin{pmatrix} x \\ y \end{pmatrix} = c \begin{pmatrix} 1 \\ \mathrm{i} \end{pmatrix}.$$

If $\lambda = -\mathrm{i}$ then to find eigenvectors we need to solve

$$\begin{pmatrix} \mathrm{i} & 1 \\ -1 & \mathrm{i} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

which implies $\mathrm{i}x + y = 0$, or $y = -\mathrm{i}x$. So again there is no non-zero real solution. But there is a complex eigenvector:

$$\begin{pmatrix} x \\ y \end{pmatrix} = d \begin{pmatrix} 1 \\ -\mathrm{i} \end{pmatrix}.$$

In order to diagonalise the matrix A over $\mathbb{C}$ we consider the matrix

$$M = \begin{pmatrix} 1 & 1 \\ \mathrm{i} & -\mathrm{i} \end{pmatrix},$$

which has inverse

$$M^{-1} = \frac{1}{(-2\mathrm{i})} \begin{pmatrix} -\mathrm{i} & -1 \\ -\mathrm{i} & 1 \end{pmatrix}.$$

Now we check that $M^{-1}AM = D$, where D is the diagonal matrix of eigenvalues,

$$D = \begin{pmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{pmatrix}.$$

We have

$$\begin{aligned} M^{-1}AM &= \frac{1}{(-2\mathrm{i})} \begin{pmatrix} -\mathrm{i} & -1 \\ -\mathrm{i} & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ \mathrm{i} & -\mathrm{i} \end{pmatrix} \\ &= \frac{1}{(-2\mathrm{i})} \begin{pmatrix} -\mathrm{i} & -1 \\ -\mathrm{i} & 1 \end{pmatrix} \begin{pmatrix} \mathrm{i} & -\mathrm{i} \\ -1 & -1 \end{pmatrix} \\ &= \frac{1}{(-2\mathrm{i})} \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} = \frac{\mathrm{i}}{2} \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} = \begin{pmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{pmatrix} = D. \end{aligned}$$

The general solution for coupled ordinary differential equations

Suppose A is an $n \times n$ real matrix. We define the norm of A by

$$\|A\| = \max_{\|\mathbf{v}\| \le 1} \|A\mathbf{v}\|,$$

where $\|\mathbf{v}\|$ denotes the Euclidean norm of the vector $\mathbf{v}$ in $\mathbb{R}^n$.

6.14. **Definition.** The exponential of A is an $n \times n$ matrix defined by

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k.$$

It can be shown that

$$\|e^A\| = e^{\|A\|},$$

so the infinite sum of e^A is convergent. Now consider the differential equation

$$\dot{\mathbf{x}} = A\mathbf{x}, \tag{7}$$

where

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} .$$

Suppose $\mathbf{x}(0) = \mathbf{v}$. The general solution of (7) is given by

$$\mathbf{x}(t) = e^{tA}\mathbf{v} .$$

If A is diagonalisable and $A = MDM^{-1}$, where D is the diagonal matrix of eigenvalues of A, and we have

$$\begin{aligned} e^{tA}\mathbf{v} &= \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k \mathbf{v} \\ &= \sum_{k=0}^{\infty} \frac{t^k}{k!} (MDM^{-1})^k \mathbf{v} \\ &= M \sum_{k=0}^{\infty} \frac{t^k}{k!} (D^k) M^{-1} \mathbf{v} , \end{aligned}$$

and this is equal to

$$\mathbf{x}(t) = e^{tA}\mathbf{v} = M \begin{pmatrix} e^{\lambda_1 t} & 0 & \cdots \\ 0 & e^{\lambda_2 t} & \cdots \\ \cdots & \cdots & e^{\lambda_n t} \end{pmatrix} M^{-1}\mathbf{v} .$$

7. Real Symmetric Matrices

In this section we show that real symmetric matrices have real eigenvalues and also eigenvectors associated to different eigenvalues are orthogonal.

First we recall an earlier example, where

$$A = \begin{pmatrix} 3 & 2 \\ 2 & 0 \end{pmatrix} .$$

The eigenvectors were $\mathbf{u} = c\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ associated with eigenvalue $\lambda = 4$, and $\mathbf{v} = c'\begin{pmatrix} -1 \\ 2 \end{pmatrix}$ associated with eigenvalue $\lambda = -1$.

Note: this example happened to be a symmetric matrix. The eigenvectors $\mathbf{u}$ and $\mathbf{v}$ happen to be orthogonal, i.e., their dot product is

$$\mathbf{u}^T\mathbf{v} = \mathbf{v}^T\mathbf{u} = 2(-1) + 1(2) = 0 .$$

This is not a coincidence. In fact we have the following result.

7.1. **Theorem.** *Let $A \in \mathbb{R}^{n\times n}$ with $A^T = A$ (i.e., A is a real and symmetric matrix). Then*

(i) *The eigenvalues of A are real.*
(ii) *Eigenvectors belonging to distinct eigenvalues are orthogonal.*

Proof. To prove this we first note that if $\mathbf{u}$ and $\mathbf{v}$ are column vectors then, if A is symmetric, we have

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

Note also that, if $\mathbf{u}$ has complex elements and we denote the column vector obtained by taking the complex conjugate of all its elements by $\mathbf{u}^*$, then

$$(\mathbf{u}^*)^T\mathbf{u} = \mathbf{u}^T\mathbf{u}^* = |u_1|^2 + \cdots + |u_n|^2 ,$$

and also, if A has real elements,

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

Proof of (i): Now let $A\mathbf{u} = \lambda\mathbf{u}$, with $\mathbf{u} \neq \mathbf{0}$, where λ and the elements of $\mathbf{u}$ are possibly complex. Then

$$A\mathbf{u}^* = \bar{\lambda}\mathbf{u}^* \,,$$

where $\bar{\lambda}$ denotes the complex conjugate of λ. We have

$$\bar{\lambda}(\mathbf{u}^*)^T\mathbf{u} = ((A\mathbf{u})^*)^T\mathbf{u} = (\mathbf{u}^*)^T A\mathbf{u} = (\mathbf{u}^*)^T\lambda\mathbf{u} = \lambda(\mathbf{u}^*)^T\mathbf{u} \,.$$

But $\mathbf{u} \neq \mathbf{0}$, hence $(\mathbf{u}^*)^T\mathbf{u} \neq 0$, which implies $\bar{\lambda} = \lambda$, and λ is real.

Proof of (ii): Suppose $A\mathbf{u} = \lambda\mathbf{u}$ and $A\mathbf{v} = \mu\mathbf{v}$ where $\lambda \neq \mu$, $\mathbf{u} \neq \mathbf{0}$, and $\mathbf{v} \neq \mathbf{0}$. Then

$$(\lambda - \mu)\mathbf{u}^T\mathbf{v} = (\lambda\mathbf{u})^T\mathbf{v} - \mathbf{u}^T(\mu\mathbf{v}) = (A\mathbf{u})^T\mathbf{v} - \mathbf{u}^T A\mathbf{v} = 0 \,.$$

Since $\lambda - \mu \neq 0$, we must have $\mathbf{u}^T\mathbf{v} = 0$. □

Note that, in general, the set $\mathbf{u}$ with $A\mathbf{u} = \lambda\mathbf{u}$ for given A and λ forms a subspace of $\mathbb{R}^n$ which might have more than one dimension. Even in this case, though, it is possible to choose a basis for this subspace which consists of mutually orthogonal vectors (this will be proved in Linear Algebra using Gram-Schmidt orthogonalisation).

Moreover, one can show that, for symmetric matrix, we always have that the set of all eigenvectors spans $\mathbb{R}^n$.

Conclusion: for any symmetric matrix, one can choose (if not already the case) a basis of mutually orthogonal eigenvectors. Furthermore, we can "normalise" each of our $\mathbf{u}$ so that $\mathbf{u}^T\mathbf{u} = 1$. Since if $A\mathbf{u} = \lambda\mathbf{u}$ and $\mathbf{u}^T\mathbf{u} \neq 1$ then simply replace $\mathbf{u}$ by

$$\frac{\mathbf{u}}{\sqrt{\mathbf{u}^T\mathbf{u}}} \,.$$

We then have a set $\mathbf{u}_1, \ldots, \mathbf{u}_n$ of eigenvectors satisfying

$$A\mathbf{u}_i = \lambda_i\mathbf{u}_i \,, \quad i = 1, \ldots, n \,,$$

with some of the λ_i's possibly being the same, which form an *orthonormal basis* for $\mathbb{R}^n$, which means

$$\mathbf{u}_i^T\mathbf{u}_j = \delta_{ij} \,,$$

where δ_{ij} denotes the "*Kronecker delta*" which is a quantity that takes the value 1 when $i = j$ and it is zero if $i \neq j$.

Thanks to the fact that eigenvectors span $\mathbb{R}^n$ we can always diagonalise any symmetric matrix A by taking M to be the matrix

$$\begin{pmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \\ | & | & & | \end{pmatrix} .$$

Then we have $M^{-1}AM = \Lambda$, where Λ is the diagonal matrix of eigenvalues

$$\begin{pmatrix} \lambda_1 & & 0 \\ 0 & \ddots & 0 \\ 0 & & \lambda_n \end{pmatrix} .$$

Now M inherits extra properties. Namely

$$M^T M = \begin{pmatrix} \mathbf{u}_1^T\mathbf{u}_1 & \mathbf{u}_1^T\mathbf{u}_2 & \cdots & \mathbf{u}_1^T\mathbf{u}_n \\ \mathbf{u}_2^T\mathbf{u}_1 & \mathbf{u}_2^T\mathbf{u}_2 & \cdots & \mathbf{u}_2^T\mathbf{u}_n \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{u}_n^T\mathbf{u}_1 & \mathbf{u}_n^T\mathbf{u}_2 & \cdots & \mathbf{u}_n^T\mathbf{u}_n \end{pmatrix} = \begin{pmatrix} 1 & & 0 \\ 0 & \ddots & 0 \\ 0 & & 1 \end{pmatrix} = I_n \, .$$

Note that $M^T M = I$ implies that $\det(M^T)\det(M) = \det(I) = 1$ so that $\det(M) \neq 0$ and then M^{-1} exists. Hence $M^T M M^{-1} = M^{-1}$, which gives $M^T = M^{-1}$.

The conclusion is that to diagonalise a symmetric matrix we don't need to find the inverse of a matrix, we just need to find the transpose. The equation $M^{-1}AM = \Lambda$ becomes $M^T AM = \Lambda$.

7.2. *Example.* Consider again the symmetric matrix

$$A = \begin{pmatrix} 3 & 2 \\ 2 & 0 \end{pmatrix} .$$

Replace $\mathbf{u} = c\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ by $\begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}$, since the norm of $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ is $\sqrt{2^2 + 1^2} = \sqrt{5}$. And replace

$$\mathbf{v} = d\begin{pmatrix} -1 \\ 2 \end{pmatrix} \quad \text{by} \quad \begin{pmatrix} \frac{-1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} .$$

This process is called normalisation of eigenvectors.

Now we have

$$M = \begin{pmatrix} \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix} \quad \text{and} \quad M^T = \begin{pmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{-1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix}$$

Finally,

$$M^T AM = \Lambda = \begin{pmatrix} 4 & 0 \\ 0 & -1 \end{pmatrix} .$$

A square matrix satisfying $M^T M = I$ is called an *orthogonal* matrix. The columns of an orthogonal matrix are mutually orthonormal vectors (they are mutually orthogonal and their norm is 1). It also satisfies $MM^T = I$ and then M^T has mutually orthonormal columns, or the rows of M are mutually orthonormal as well.

Department of Mathematics, University of York