Last updated March 19, 2019

UNIVERSITY COLLEGE LONDON

DEPARTMENT OF PHYSICS AND ASTRONOMY

1246 MATHEMATICAL METHODS II

Dr. Frank Krüger and Prof. Serena Viti

Based on the notes from Profs Nikos Konstantinidis, Jon Butterworth, Mike Gillan, Jonathan Tennyson and Sergey Yurchenko.

Please send any comments and corrections to f.kruger@ucl.ac.uk or serena.viti@ucl.ac.uk

These notes are intended as an aid to revision, but they do not contain all the examples and diagrams from the course and are a supplement, not a substitute for your own notes taken in lectures.
They are also still being revised, added to and corrected.

Contents

Chapter 4

Coordinates, Vector Spaces and Linear Transformations

4.1 Introduction

A coordinate system is a choice of labels with which we can describe a physical system. For example, in Cartesian coordinates we choose an origin, plus basis vectors (directions) and scales for three perpendicular axes. The choice itself is arbitrary, and no physical laws should depend upon the choice.

Because of this, it is often good to express physics in *coordinate independent* form. For instance, an expression written as a relationship between vectors is independent of the choice of coordinates. E.g.,

$$\mathbf{F} = m\mathbf{a} \quad \text{and} \quad \mathbf{L} = \mathbf{r} \times \mathbf{p}$$

are both true, whatever axes we choose for the components of the vectors involved. They are also much more compact than writing down the equations component by component.

So one reason for paying attention to transformations between different coordinate systems is to check that **physical laws do not depend upon an arbitrary choice of coordinates**. This is a very powerful principle, and generalisations of it lie behind special relativity as well as advanced areas at the edge of new physics research such as superconductivity, quantum electrodynamics and the standard model of particle physics.

Another reason for paying attention to different coordinate systems and transformations between them is that many physics problems can be made much simpler by an intelligent choice of co-ordinate system (for example, the multi-dimensional integrals considered in the previous section).

Finally, if we develop a general way of discussing coordinate systems, we generalise the concept of a space to potentially include more than three dimensions (useful e.g. in relatvity, or string theory) and allow complex vectors (useful in quantum mechanics).

In the next few lectures we will introduce a general way to discuss and manipulate coordinate systems. We will introduce matrices which are a useful way of representing linear transformations within and between coordinate systems.

First of all let us remind ourselves of the definition of real and complex vector space.

A *complex* vector space S is a slight generalisation of a *real* vector space. It is defined by the property that

$$\forall\ \underline{\mathbf{v}}_1, \underline{\mathbf{v}}_2 \in S \quad \text{and} \quad \forall\ u_1, u_2 \in \mathbb{C} \quad : \quad (u_1\underline{\mathbf{v}}_1 + u_2\underline{\mathbf{v}}_2) \in S\ ,$$

which can be worded as: for all vectors $\underline{\mathbf{v}}_1$ and $\underline{\mathbf{v}}_2$ belonging to the vector space S and for all complex numbers u_1 and u_2, the linear combination $(u_1\underline{\mathbf{v}}_1 + u_2\underline{\mathbf{v}}_2)$ still belongs to S.

In other words, a vector space is defined by the property that by adding together vectors and by multiplying them by scalars one stays within the vector space. The fact that u_1 and u_2 belong to $\mathbb{C}$ qualifies the vector space S as "complex". Had u_1 and u_2 belonged to $\mathbb{R}$, the vector space would have then been "real".

We are now going to start by introducing three dimensional vectors in Euclidian space before generalizing to n-dimensions.

4.2 Three dimensional vectors

Three-dimensional Euclidean space is usually defined by introducing three mutually orthogonal *basis vectors* $\hat{\mathbf{e}}_x$, $\hat{\mathbf{e}}_y$ and $\hat{\mathbf{e}}_z$. However, this notation doesn't generalise to arbitrary number of dimensions, so we use $\underline{\hat{\mathbf{e}}}_1 = \hat{\mathbf{e}}_x$, $\underline{\hat{\mathbf{e}}}_2 = \hat{\mathbf{e}}_y$, and $\underline{\hat{\mathbf{e}}}_3 = \hat{\mathbf{e}}_z$ instead. These basis vectors have unit length,

$$\underline{\hat{\mathbf{e}}}_1 \cdot \underline{\hat{\mathbf{e}}}_1 = \underline{\hat{\mathbf{e}}}_2 \cdot \underline{\hat{\mathbf{e}}}_2 = \underline{\hat{\mathbf{e}}}_3 \cdot \underline{\hat{\mathbf{e}}}_3 = 1, \tag{4.1}$$

and are perpendicular to each other:

$$\underline{\hat{\mathbf{e}}}_1 \cdot \underline{\hat{\mathbf{e}}}_2 = \underline{\hat{\mathbf{e}}}_2 \cdot \underline{\hat{\mathbf{e}}}_3 = \underline{\hat{\mathbf{e}}}_3 \cdot \underline{\hat{\mathbf{e}}}_1 = 0. \tag{4.2}$$

These properties are summarised in one equation as

$$\underline{\hat{\mathbf{e}}}_i \cdot \underline{\hat{\mathbf{e}}}_j = \delta_{ij}, \tag{4.3}$$

where the Kronecker delta δ_{ij} is shorthand for

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{4.4}$$

Any vector $\underline{\mathbf{v}}$ in this three-dimensional space may be written down in terms of its *components* along the $\underline{\hat{\mathbf{e}}}_i$. Thus

$$\underline{\mathbf{v}} = v_1\, \underline{\hat{\mathbf{e}}}_1 + v_2\, \underline{\hat{\mathbf{e}}}_2 + v_3\, \underline{\hat{\mathbf{e}}}_3\ ,$$

where the coefficients v_i may be obtained by taking the scalar product of $\underline{\mathbf{v}}$ with the basis vector $\underline{\hat{\mathbf{e}}}_i$,

$$v_i = \underline{\hat{\mathbf{e}}}_i \cdot \underline{\mathbf{v}}\ . \tag{4.5}$$

(i.e multiply the left hand side and the right hand side by $\underline{\hat{\mathbf{e}}}_i$).

This follows because the $\underline{\hat{\mathbf{e}}}_i$ are perpendicular and have length one.

If we know two vectors $\underline{\mathbf{v}}$ and $\underline{\mathbf{u}}$ in terms of their components, then their scalar product is

$$\underline{\mathbf{u}} \cdot \underline{\mathbf{v}} = (u_1\, \underline{\hat{\mathbf{e}}}_1 + u_2\, \underline{\hat{\mathbf{e}}}_2 + u_3\, \underline{\hat{\mathbf{e}}}_3) \cdot (v_1\, \underline{\hat{\mathbf{e}}}_1 + v_2\, \underline{\hat{\mathbf{e}}}_2 + v_3\, \underline{\hat{\mathbf{e}}}_3) = u_1\, v_1 + u_2\, v_2 + u_3\, v_3 = \sum_{i=1}^{3} u_i\, v_i\,. \tag{4.6}$$

A particularly important case is that of the scalar product of a vector with itself, which gives rise to Pythagoras's theorem

$$v^2 = \underline{\mathbf{v}} \cdot \underline{\mathbf{v}} = v_1^2 + v_2^2 + v_3^2\,. \tag{4.7}$$

The length of a vector $\underline{\mathbf{v}}$ is

$$v = \mid \underline{\mathbf{v}} \mid = \sqrt{v^2} = \sqrt{v_1^2 + v_2^2 + v_3^2}\,. \tag{4.8}$$

If $v = 1$ the vector is called a *unit vector*. A vector is the *zero vector* if and only if all its components vanish. Thus

$$\underline{\mathbf{v}} = \underline{\mathbf{0}} \quad \Longleftrightarrow \quad (v_1\,, v_2\,, v_3) = (0, 0, 0)\,. \tag{4.9}$$

4.2.1 Linear Dependence

A set of vectors $\underline{\mathbf{X}}_1\,,\ \underline{\mathbf{X}}_2\,,\ \cdots\ \underline{\mathbf{X}}_n$ are *linearly dependent* when it is possible to find a set of scalar coefficients c_i (not all zero) such that

$$c_1\, \underline{\mathbf{X}}_1 + c_2\, \underline{\mathbf{X}}_2 + \cdots + c_n\, \underline{\mathbf{X}}_n = \underline{\mathbf{0}}\,.$$

If no such constants c_i exist, then the $\underline{\mathbf{X}}_i$ are *linearly independent*.

The vector $\underline{\mathbf{v}}$ is a *linear combination* of the basis vectors $\underline{\hat{\mathbf{e}}}_i$. Note that the basis vectors themselves are *linearly independent*, because there is no linear combination of the $\underline{\hat{\mathbf{e}}}_i$ which vanishes – unless all the coefficients are zero. Putting it in other words,

$$\underline{\hat{\mathbf{e}}}_3 \neq \alpha\, \underline{\hat{\mathbf{e}}}_1 + \beta\, \underline{\hat{\mathbf{e}}}_2\,, \tag{4.10}$$

where α and β are arbitrary scalars. Clearly, something in the x-direction plus something else in the y-direction cannot give something lying in the z-direction.

On the other hand, for three vectors taken at random, one might well be able to express one of them in terms of the other two. For example, consider the three vectors given in component form by

$$\underline{\mathbf{u}} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \underline{\mathbf{v}} = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} \quad \text{and} \quad \underline{\mathbf{w}} = \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix}. \tag{4.11}$$

Then

$$\underline{\mathbf{w}} = 2\underline{\mathbf{v}} - \underline{\mathbf{u}}, \tag{4.12}$$

or equivalently

$$(-1)\underline{\mathbf{u}} + 2\underline{\mathbf{v}} + (-1)\underline{\mathbf{w}} = \underline{\mathbf{0}}. \tag{4.13}$$

We then say that $\underline{\mathbf{u}}$, $\underline{\mathbf{v}}$ and $\underline{\mathbf{w}}$ are *linearly dependent*. This is an important concept.

Therefore in summary, the three-dimensional space S_3 is defined as one where there are three (but no more) orthonormal linearly independent vectors $\underline{\hat{\mathbf{e}}}_i$. Any vector lying in this three-dimensional space can be written as a linear combination of the basis vectors. All this is really saying is that we can always write $\underline{\mathbf{v}}$ in the component form;

$$\underline{\mathbf{v}} = v_1\,\underline{\hat{\mathbf{e}}}_1 + v_2\,\underline{\hat{\mathbf{e}}}_2 + v_3\,\underline{\hat{\mathbf{e}}}_3\,.$$

Note the $\underline{\hat{\mathbf{e}}}_i$ are not unique. We could, for example, rotate the system through 45° and use these new axes as basis vectors. We will now generalise this to an arbitrary number of dimensions and letting the components become complex. Note that such complex vector spaces are important for Quantum Mechanics.

4.3 *n*-Dimensional Linear Vector Space

4.3.1 Definition

A linear vector space S is a set of abstract quantities $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}, \cdots$, called vectors, which have the following properties:

1. If $\underline{\mathbf{a}} \in S$ and $\underline{\mathbf{b}} \in S$, then

$$\begin{aligned} \underline{\mathbf{a}} + \underline{\mathbf{b}} &= \underline{\mathbf{c}} \in S. \\ \underline{\mathbf{c}} = \underline{\mathbf{a}} + \underline{\mathbf{b}} &= \underline{\mathbf{b}} + \underline{\mathbf{a}} \;\; \text{(Commutative law)} \\ (\underline{\mathbf{a}} + \underline{\mathbf{b}}) + \underline{\mathbf{c}} &= \underline{\mathbf{a}} + (\underline{\mathbf{b}} + \underline{\mathbf{c}}) \;\; \text{(Associative law)}\,. \end{aligned} \tag{4.14}$$

2. Multiplication by a scalar (possibly complex)

$$\begin{aligned} \underline{\mathbf{a}} \in S &\implies \lambda\,\underline{\mathbf{a}} \in S \;\; (\lambda \text{ a complex number})\,, \\ \lambda\,(\underline{\mathbf{a}} + \underline{\mathbf{b}}) &= \lambda\,\underline{\mathbf{a}} + \lambda\,\underline{\mathbf{b}}\,, \\ \lambda\,(\mu\,\underline{\mathbf{a}}) &= (\lambda\,\mu)\,\underline{\mathbf{a}} \;\; (\mu \text{ another complex number})\,. \end{aligned} \tag{4.15}$$

3. There exists a null (zero) vector $\underline{\mathbf{0}} \in S$ such that

$$\underline{\mathbf{a}} + \underline{\mathbf{0}} = \underline{\mathbf{a}} \tag{4.16}$$

for all vectors $\underline{\mathbf{a}}$.

4. For every vector $\underline{\mathbf{a}}$ there exists a unique vector $-\underline{\mathbf{a}}$ such that

$$\underline{\mathbf{a}} + (-\underline{\mathbf{a}}) = \underline{\mathbf{0}}\,. \tag{4.17}$$

Perhaps surprisingly, the solutions $y(x)$ of an n-th order, homogeneous linear ODE,

$$a_n(x)\frac{d^n y}{dx^n} + \ldots + a_1(x)\frac{dy}{dx} + a_0(x)y = 0,$$

satisfy all the above criteria. They therefore form an n-dimensional linear vector space. The vectors are given by functions $y(x)$ that satisfy the ODE. In the case of second-order homogeneous linear ODEs ($n = 2$), if we know two independent solutions $y_1(x)$ and $y_2(x)$, the general solution is given by all linear combinations

$$y(x) = Ay_1(x) + By_2(x).$$

$y_1(x)$ and $y_2(x)$ therefore form a *basis* of this two-dimensional vector space.

4.3.2 Basis vectors and components in n-dimensional space

Any set of n linearly independent vectors $\underline{\mathbf{X}}_1, \underline{\mathbf{X}}_2, \cdots \underline{\mathbf{X}}_n$ can be used as a *basis* for an n-dimensional vector space S_n. This implies that the basis is not unique. Once the basis has been chosen, any vector can be written uniquely as a linear combination

$$\underline{\mathbf{v}} = \sum_{i=1}^{n} v_i \underline{\mathbf{X}}_i$$

of the basis vectors. The set of numbers $v_1, \ldots, v_n$ (the components) are said to *represent* the vector $\underline{\mathbf{v}}$ in that basis. The concept of a vector is more general and abstract than that of the components. The components are somehow man-made. If we rotate the coordinate system then the vector stays in the same direction but the components change.

In this n-dimensional space, unlike in the Euclidian Space, we have not assumed that the basis vectors are unit vectors or orthogonal to one another. For certain physical problems, it is convenient to work with basis vectors which are not perpendicular — e.g. when dealing with crystals with hexagonal symmetry. However, here we will only work with basis vectors $\underline{\hat{\mathbf{e}}}_i$ which are orthogonal and of unit length.

4.3.3 Definition of scalar product in n-dimensional space

Let $\underline{\mathbf{u}} = \sum_{i=1}^{n} u_i \underline{\hat{\mathbf{e}}}_i$ and $\underline{\mathbf{v}} = \sum_{i=1}^{n} v_i \underline{\hat{\mathbf{e}}}_i$ be arbitrary vectors of an n-dimensional vector space S_n with complex components $u_i \in \mathbb{C}$ and $v_i \in \mathbb{C}$. Then the *scalar product* of these two vectors is defined by

$$\underline{\mathbf{u}} \cdot \underline{\mathbf{v}} = u_1^* v_1 + u_2^* v_2 + \cdots + u_n^* v_n. \tag{4.18}$$

The only difference from the usual form is the complex conjugation on all the components u_i, since the vectors have to be allowed to be complex. Note that

$$\underline{\mathbf{v}} \cdot \underline{\mathbf{u}} = v_1^* u_1 + v_2^* u_2 + \cdots + v_n^* u_n = (\underline{\mathbf{u}} \cdot \underline{\mathbf{v}})^*. \tag{4.19}$$

In general, the scalar product is a complex scalar.

Consequences of the definition:

1. $\underline{\mathbf{w}} \cdot (\alpha\, \underline{\mathbf{u}} + \beta\, \underline{\mathbf{v}}) = \alpha\, \underline{\mathbf{w}} \cdot \underline{\mathbf{u}} + \beta \underline{\mathbf{w}} \cdot \underline{\mathbf{v}}$.

2. Putting $\underline{\mathbf{u}} = \underline{\mathbf{v}}$, we see that

$$u^2 = \underline{\mathbf{u}} \cdot \underline{\mathbf{u}} = u_1^*\, u_1 + u_2^*\, u_2 + \cdots + u_n^*\, u_n = |u_1|^2 + |u_2|^2 + \cdots + |u_n|^2\,. \tag{4.20}$$

   Generalisation of Pythagoras's theorem for complex numbers. Since the $|u_i|^2$ are real and cannot be negative, then $u^2 \geq 0$.

3. Two vectors are orthogonal if $\underline{\mathbf{u}} \cdot \underline{\mathbf{v}} = 0$.

4. Components of a vector are given by the scalar product $v_i = \underline{\hat{\mathbf{e}}}_i \cdot \underline{\mathbf{v}}$.

4.4 Matrices and linear transformations

Reminder: what is a set? Any collection of quantities, operators or objects forms a set. Each individual quantity, operator or object is called an element. If the set contains a finite number of elements it is said to be a finite set, otherwise it's called infinite.

4.4.1 Linear Transformations

Very often, e.g. in physics, we need to perform some operation on a vector $\underline{\mathbf{v}}$ which changes it into another vector in the space S_n. For example, rotate the vector. Denote the operation by $\hat{A}$ and, instead of tediously saying that $\hat{A}$ acts on $\underline{\mathbf{v}}$, write it symbolically as $\hat{A}\,\underline{\mathbf{v}}$. By assumption, therefore, $\underline{\mathbf{u}} = \hat{A}\,\underline{\mathbf{v}}$ is another vector in the same space S_n.

How to express the operator $\hat{A}$ in the basis $\underline{\hat{\mathbf{e}}}_1, \underline{\hat{\mathbf{e}}}_2, \cdots, \underline{\hat{\mathbf{e}}}_n$? To investigate this further, see how the operation $\hat{A}$ changes the basis vectors $\underline{\hat{\mathbf{e}}}_1, \underline{\hat{\mathbf{e}}}_2, \cdots, \underline{\hat{\mathbf{e}}}_n$. Let us look at $\underline{\hat{\mathbf{e}}}_1$, which has a 1 in the first position and zeros everywhere else:

$$\underline{\hat{\mathbf{e}}}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (n \text{ terms in the column}). \tag{4.21}$$

The result of $\hat{A}\underline{\hat{\mathbf{e}}}_1$ gives rise to a vector which we shall denote by $\underline{\mathbf{a}}_1$ because it started from $\underline{\hat{\mathbf{e}}}_1$. Thus

$$\underline{\mathbf{a}}_1 = \hat{A}\,\underline{\hat{\mathbf{e}}}_1 \,. \tag{4.22}$$

To write this in terms of components, we must introduce a second index

$$\underline{\mathbf{a}}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \vdots \\ a_{n1} \end{pmatrix} . \tag{4.23}$$

(because the first index now defines the row).

To specify the action of $\hat{A}$ completely, we must define how it acts on *all* the basis vectors $\underline{\hat{\mathbf{e}}}_j$:

$$\underline{\mathbf{a}}_j = \hat{A}\,\underline{\hat{\mathbf{e}}}_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ a_{3j} \\ \vdots \\ a_{nj} \end{pmatrix} . \tag{4.24}$$

This requires n^2 numbers a_{ij} $(i, j = 1, \ldots, n)$. Instead of writing $\underline{\mathbf{a}}_j$ explicitly as a column vector, we can use the basis vectors once again to show that

$$\underline{\mathbf{a}}_j = a_{1j}\,\underline{\hat{\mathbf{e}}}_1 + a_{2j}\,\underline{\hat{\mathbf{e}}}_2 + \ldots + a_{nj}\,\underline{\hat{\mathbf{e}}}_n = \sum_{i=1}^{n} a_{ij}\,\underline{\hat{\mathbf{e}}}_i \,. \tag{4.25}$$

as $\underline{\hat{\mathbf{e}}}_i$ has 1 in the i'th position and 0's everywhere else.

Knowing the basis vectors transformation, it is (in principle) easy to evaluate the action of $\hat{A}$ on some vector $\underline{\mathbf{v}} = \sum_j v_j \, \underline{\hat{\mathbf{e}}}_j$. Then

$$\underline{\mathbf{u}} = \hat{A}\,\underline{\mathbf{v}} = \sum_j v_j(\hat{A}\,\underline{\hat{\mathbf{e}}}_j) = \sum_{ij} a_{ij}\, v_j\, \underline{\hat{\mathbf{e}}}_i \,. \tag{4.26}$$

But, writing $\underline{\mathbf{u}}$ in terms of components as well,

$$\underline{\mathbf{u}} = \sum_i u_i\, \underline{\hat{\mathbf{e}}}_i \,, \tag{4.27}$$

and comparing coefficients of $\underline{\hat{\mathbf{e}}}_i$, we find

$$u_i = \sum_{j=1}^{n} a_{ij}\, v_j \,. \tag{4.28}$$

Note that basis vectors transform with $\sum_i a_{ij}\, \underline{\hat{\mathbf{e}}}_i$, whereas the components involve the other index $\sum_j a_{ij}\, v_j$.

The set of numbers a_{ij} represents the abstract operator $\hat{A}$ in the particular basis chosen; these n^2 numbers determine completely the effect of $\hat{A}$ on any arbitrary vector: the vector undergoes a *linear transformation*. It is convenient to arrange all these numbers into an array

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} , \tag{4.29}$$

called a matrix. This one is in fact a square matrix with n rows and n columns.

Matrices are useful mathematical tools for describing linear transformations in physics which you need to be familiar with.

A matrix is a two-dimensional[1] array of numbers $A = a_{ij}$ where i and j are indices running from 1 to n and m, respectively. We will be using matrices with $n = m = 2, 3$ or 4 (and 1 or course!), but they can be arbitrarily large.

As we signify a vector by underlining it and writing it in bold in order to distinguish it from a scalar, we similarly write $\mathbf{A}$ in bold face in order to show that it is a matrix. You will find other conventions used elsewhere too. It's important to get used to different conventions!

[1]It may (or may not) help to note that the generalisation of vectors and matrices is to tensors. First-rank tensors are vectors, second-rank tensors are matrices, and higher dimensional tensors also exist and are sometimes used in physics, but are beyond the scope of this course.

In general a matrix is a set of elements, which can be either numbers or variables, set out in the form of an array. For example

$$
\underset{\text{(rectangular)}}{\begin{pmatrix} 2 & 6 & 4 \\ -1 & i & 7 \end{pmatrix}} \quad \text{or} \quad \underset{\text{(square)}}{\begin{pmatrix} 0 & -i \\ 3+6i & x^2 \end{pmatrix}}
$$

A matrix having n rows and m columns is called an $n \times m$ matrix. The above examples are 2×3 and 2×2 matrices. A square matrix clearly has $n = m$. Note that matrices are enclosed in brackets.

A vector is a simple matrix which is $n \times 1$ (column vector) or $1 \times n$ (row vector), as in

$$
\begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_n \end{pmatrix} \quad \text{or} \quad (v_1, v_2, v_3, \cdots, v_n) \, .
$$

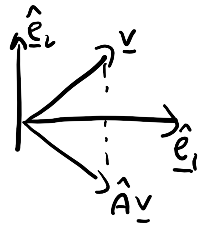
4.4.2 Worked Examples

For the standard cartesian basis $\underline{\hat{\mathbf{e}}}_1 = \hat{\mathbf{e}}_x$ and $\underline{\hat{\mathbf{e}}}_2 = \hat{\mathbf{e}}_y$ of the two-dimensional vector space S_2 determine the matrix representation $\mathbf{A}$ of the operator $\hat{A}$ that

(a) reflects a vector over the x-axis;

(b) rotates a vector through an angle ϕ anti-clockwise.

(a)



Action on basis vectors:

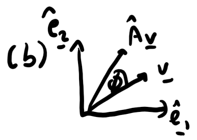
$$\hat{A}\hat{\underline{e}}_1 = \hat{\underline{e}}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$\hat{A}\hat{\underline{e}}_2 = -\hat{\underline{e}}_2 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

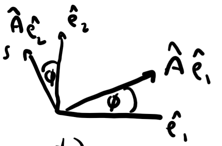
$\therefore \quad \underline{A} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. As a test, apply to the vector $\underline{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ in the above picture:

$$\underline{A}\,\underline{v} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} v_1 \\ -v_2 \end{pmatrix}$$ as expected.

(b)



Acting on basis vectors $\longrightarrow$



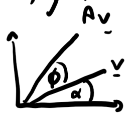
$$\therefore \; a_1 = \hat{A}\hat{\underline{e}}_1 = \begin{pmatrix} \cos\phi \\ \sin\phi \end{pmatrix}$$

$$a_2 = \hat{A}\hat{\underline{e}}_2 = \begin{pmatrix} -\sin\phi \\ \cos\phi \end{pmatrix} \longrightarrow \underline{A} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$$

To check if this matrix indeed represents a rotation by ϕ we write the vector $\underline{V}$ as $\underline{V} = \begin{pmatrix} v\cos\alpha \\ v\sin\alpha \end{pmatrix}$ and act with $\hat{A}$:

$$\underline{A}\underline{v} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}\begin{pmatrix} v\cos\alpha \\ v\sin\alpha \end{pmatrix}$$

$$= \begin{pmatrix} v(\cos\phi\cos\alpha - \sin\phi\sin\alpha) \\ v(\sin\phi\cos\alpha + \cos\phi\sin\alpha) \end{pmatrix} = \begin{pmatrix} v\cos(\phi+\alpha) \\ v\sin(\phi+\alpha) \end{pmatrix}$$



4.4.3 Matrix addition or subtraction

To add (or subtract) two matrices, $\mathbf{M} = \mathbf{A} \pm \mathbf{B}$, one simply adds (subtracts) the corresponding elements,

$$m_{ij} = a_{ij} \pm b_{ij}.$$

So for example, for the two simple matrices,

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 2 & 1 \\ 3 & 1 \end{pmatrix}, \tag{4.30}$$

we have

$$\mathbf{M} = \mathbf{A} + \mathbf{B} = \begin{pmatrix} 3 & 3 \\ 4 & 1 \end{pmatrix}.$$

It follows immediately that $\mathbf{A}+\mathbf{B} = \mathbf{B}+\mathbf{A}$ (commutative law of addition) and $(\mathbf{A}+\mathbf{B})+\mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$ (associative law).

The sum of two matrices **A** and **B** can only be defined if the matrices have the same dimension, e.g. if they are both $m \times n$ matrices. In this case, $\mathbf{M} = \mathbf{A} \pm \mathbf{B}$ is also an $m \times n$ matrix.

4.4.4 Multiple Transformations; Matrix Multiplication

If a matrix **M** is the product of two matrices **A** and **B**, i.e. $\mathbf{M} = \mathbf{AB}$, the elements of **M** are given by:

$$m_{ij} = (\mathbf{AB})_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \equiv a_{ik} b_{kj} \tag{4.31}$$

where we have introduced the "Einstein summation convention" where a repeated index implies a sum over the repeated index. Thus, for two 3×3 matrices:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \qquad \text{and} \qquad \mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}.$$

then their product is:

$$\mathbf{M} = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} & a_{11}b_{13} + a_{12}b_{23} + a_{13}b_{33} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} & a_{21}b_{13} + a_{22}b_{23} + a_{23}b_{33} \\ a_{31}b_{11} + a_{32}b_{21} + a_{33}b_{31} & a_{31}b_{12} + a_{32}b_{22} + a_{33}b_{32} & a_{31}b_{13} + a_{32}b_{23} + a_{33}b_{33} \end{pmatrix}.$$

Again, for the simple matrices in eq. 4.30,

$$\mathbf{M} = \mathbf{AB} = \begin{pmatrix} 8 & 3 \\ 2 & 1 \end{pmatrix}.$$

Note that matrix multiplication is **not in general commutative**, i.e.

$$\mathbf{AB} \neq \mathbf{BA}.$$

For our example,

$$\mathbf{BA} = \begin{pmatrix} 3 & 4 \\ 4 & 6 \end{pmatrix} \neq \mathbf{AB}.$$

but it is associative, i.e.

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$$

for any matrices $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$. Let't prove this (cf. Eq 1.31 in order not to get confused with the indeces):

$$\begin{aligned}
[(\mathbf{AB})\mathbf{C}]_{il} &= \sum_j (\mathbf{AB})_{ij} c_{jl} = \sum_j \sum_k (a_{ik} b_{kj}) c_{jl} \\
&= \sum_j \sum_k a_{ik} (b_{kj} c_{jl}) = \sum_k a_{ik} (\mathbf{BC})_{kl} = [\mathbf{A}(\mathbf{BC})]_{il}.
\end{aligned}$$

Note that matrix multiplication can only be defined if the number of columns in $\mathbf{A}$ is equal to the number of rows in $\mathbf{B}$. Then if $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $n \times p$, then $\mathbf{C}$ is $m \times p$.

To interpret the effect of multiplication, suppose that we know the action of some operator $\hat{A}$ on any vector and also the action of another operator $\hat{B}$. What is the action of the combined operation of $\hat{B}$ followed by $\hat{A}$? Consider

$$\begin{aligned}
\underline{\mathbf{w}} &= \hat{B}\,\underline{\mathbf{v}} \\
\underline{\mathbf{u}} &= \hat{A}\,\underline{\mathbf{w}}\,. \\
\underline{\mathbf{u}} &= \hat{A}\,\hat{B}\,\underline{\mathbf{v}} = \hat{C}\underline{\mathbf{v}}\,.
\end{aligned} \tag{4.32}$$

To find the matrix representation of $\hat{C}$, write the above equations in component form:

$$\begin{aligned}
w_i &= \sum_j b_{ij}\, v_j \\
u_k &= \sum_i a_{ki}\, w_i \\
&= \sum_{i,j} a_{ki}\, b_{ij} v_j \\
&= \sum_j c_{kj}\, v_j\,.
\end{aligned} \tag{4.33}$$

Since this is supposed to hold for any vector $\underline{\mathbf{v}}$, it requires that

$$c_{kj} = \sum_{i=1}^{n} a_{ki}\, b_{ij}\,. \tag{4.34}$$

This is the law for the multiplication of two matrices $\mathbf{A}$ and $\mathbf{B}$. The product matrix has the elements c_{kj}.

Matrices do not commute because they are constructed to represent linear operations and, in general, such operations do not commute. It can matter in which order you do certain operations.

4.5 Determinants

For a square matrix there is a useful number called the determinant of the matrix. For a 1 x 1 matrix the determinant is just the value of the single element; hence we shall show what we mean by a determinant by taking as an example a 2 x 2 matrix. We will then move to higher order matrices.

4.5.1 Two-by-Two Determinants

The determinant is a useful number which is defined for square matrices. For a 2×2 matrix, the determinant is

$$\det A = \begin{vmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} \tag{4.35}$$

This is a second order determinant.

The determinant is an ordinary scalar quantity and can be defined for any set of four numbers a_{ij}. An example:

$$\Delta = \begin{vmatrix} 1 & 3 \\ 4 & 2 \end{vmatrix} = 1 \times 2 - 3 \times 4 = -10\,.$$

4.5.2 Higher order determinants and cofactors

Let's write a 3rd order determinant as:

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

where the element a_{ij} is in row i and column j. To evaluate higher order determinants, we need to evaluate the *cofactor* of each element. The cofactor of the element a_{ij} is defined as

$$c_{ij} = (-1)^{i+j} \det \mathbf{M}_{ij}, \tag{4.36}$$

where $\mathbf{M}_{ij}$ is the matrix remaining when row i and column j have been removed from the matrix $\mathbf{A}$ ($\mathbf{M}_{ij}$ sometimes called Minor). For example, if you remove from the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 4 & 3 \\ 6 & 8 & 9 \\ -2 & 1 & 4 \end{pmatrix}$$

the second row and the third column, the resulting minor matrix is

$$\mathbf{M}_{23} = \begin{pmatrix} 1 & 4 \\ -2 & 1 \end{pmatrix},$$

and the resulting cofactor

$$c_{23} = (-1)^{2+3} \det \mathbf{M}_{23} = - \begin{vmatrix} 1 & 4 \\ -2 & 1 \end{vmatrix} = -(1+8) = -9.$$

The determinant of **A** is obtained by multiplying each element of one row (or one column) by its cofactor and adding the results (see e.g. Boaz Chapter 3 section 3).

To see how this works, a 3×3 determinant can be expanded by the first row as

$$\begin{aligned} \Delta &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}c_{11} + a_{12}c_{12} + a_{13}c_{13} \\ &= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \end{aligned} \tag{4.37}$$

Thus we can express the 3×3 determinant as the sum of three 2×2 ones. **Note particularly the negative sign in front of the second 2×2 determinant**, which comes from the cofactors. Evaluating the 2×2 determinants this gives the final result

$$\begin{aligned} \Delta &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} \end{aligned} \tag{4.38}$$

Alternatively, one can expand the determinant by the second column,

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = -a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{22} \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} - a_{32} \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix},$$

and this gives exactly the same value as before. Pay special attention to the terms which pick up the minus sign. The pattern is:

$$\begin{vmatrix} + & - & + \\ - & + & - \\ + & - & + \end{vmatrix}.$$

A 4×4 determinant can be reduced to four 3×3 determinants as

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix}$$

$$= a_{11} \begin{vmatrix} a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \\ a_{42} & a_{43} & a_{44} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} & a_{24} \\ a_{31} & a_{33} & a_{34} \\ a_{41} & a_{43} & a_{44} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} & a_{24} \\ a_{31} & a_{32} & a_{34} \\ a_{41} & a_{42} & a_{44} \end{vmatrix} - a_{14} \begin{vmatrix} a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{vmatrix} \tag{4.39}$$

Alternatively, one can reduce the size of determinant by taking linear combinations of rows and/or columns (see later). This can be generalised to higher dimensions.

Before we move on let us think for a moment why one would want to calculate a determinant. Determinants are used in many mathematical applications. For example, you can use the determinant of a matrix to solve a system of linear equations, or to calculate the volumes of parallelepipeds. They are also particularly useful in quantum mechanics.

4.5.3 Some properties of determinants

1. If rows are written as columns and columns as rows, the determinant is unchanged.

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

2. A determinant vanishes if one of the rows or columns contains only zeroes.

3. If we multiply a row (or column) by a constant, then the value of the determinant is multiplied by that constant.

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

4. A determinant vanishes if two rows (or columns) are multiples of each other i.e. if $a_{i2} = \alpha\, a_{i1}$ for $i = 1, 2$, then $\Delta = \alpha\, a_{11}a_{21} - \alpha\, a_{11}a_{21} = 0$, as can be seen from this very simple example:

$$\Delta' = \begin{vmatrix} 2 & 4 \\ 6 & 12 \end{vmatrix} = (2 \times 12) - (4 \times 6) = 0$$

5. If we interchange a pair of rows or columns, the determinant changes sign.

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

6. Adding a multiple of one row to another (or a multiple of one column to another) does not change the value of a determinant.

$$\Delta' = \begin{vmatrix} (a_{11} + \alpha\, a_{12}) & a_{12} \\ (a_{21} + \alpha\, a_{22}) & a_{22} \end{vmatrix} = (a_{11} + \alpha\, a_{12})a_{22} - a_{12}(a_{21} + \alpha\, a_{22})$$

$$= [a_{11}a_{22} + \alpha\, a_{12}a_{22} - a_{12}a_{21} - \alpha\, a_{12}a_{22}] = \begin{vmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{vmatrix} + 0 .$$

   This is a very useful rule to help simplify higher order determinants. In our 2×2 example, take 4 times row 1 from row 2 to give

$$\Delta = \begin{vmatrix} 1 & 3 \\ 4 & 2 \end{vmatrix} = \begin{vmatrix} 1 & 3 \\ 0 & -10 \end{vmatrix} = 1 \times (-10) - 3 \times 0 = -10 .$$

   By this trick we have just got one term in the end rather than two.

Examples

1. Evaluate

$$\Delta = \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} .$$

$$\Delta = 1 \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} - 2 \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} + 3 \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} = (45 - 48) - 2\,(36 - 42) + 3\,(32 - 35) = 0 .$$

   The answer is zero because the third row is twice the second minus the first.

2. Evaluate

$$\Delta = \begin{vmatrix} 1 & -3 & -3 \\ 2 & -1 & -11 \\ 3 & 1 & 5 \end{vmatrix} .$$

Add three times column 1 to both columns 2 and 3.

$$\Delta = \begin{vmatrix} 1 & 0 & 0 \\ 2 & 5 & -5 \\ 3 & 10 & 14 \end{vmatrix} = \begin{vmatrix} 5 & -5 \\ 10 & 14 \end{vmatrix} = \begin{vmatrix} 5 & 0 \\ 10 & 24 \end{vmatrix} = 120 .$$

Determinants calculated on computers make often use of subtraction of rows (or columns) such that there is only one element at the top of the first column with zeros everywhere else. This reduces the size of the determinant by one and can be applied systematically. With pencil and paper, this often involves keeping track of fractions. Different books call this technique by different names.

4.5.4 Testing for linear independence using determinants

It follows from Rule 4 in the last section that we can use the determinant to test whether a set of vectors are linearly independent. Consider the vectors in Section 4.2.1:

$$\underline{\mathbf{u}} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} : \ \underline{\mathbf{v}} = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} : \ \underline{\mathbf{w}} = \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix} . \tag{4.40}$$

If we write them in a matrix, and take the determinant, we see that

$$\begin{vmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{vmatrix} = 0. \tag{4.41}$$

If it is not obvious whether a set of vectors are linearly independent, then evaluating this determinant provides a definitive test. If the determinant is zero, they are not linearly independent. If it is non-zero, they are.

4.5.5 Determinant of a Matrix Product

By writing out both sides explicitly, it is straightforward to show that for 2×2 or 3×3 square matrices the determinant of a product of two matrices is equal to the product of the determinants.

$$| \mathbf{A}\,\mathbf{B} | = | \mathbf{A} | \times | \mathbf{B} | \ . \tag{4.42}$$

However, this result is true in general for $n \times n$ square matrices of any size.

Proof

Consider the product of matrices **A** and **B**. Then det(**AB**) is:

$$
\begin{aligned}
&= (ab)_{11}(ab)_{22} - (ab)_{21}(ab)_{12} \\
&= (a_{11}b_{11} + a_{12}b_{21})(a_{21}b_{12} + a_{22}b_{22}) - (a_{21}b_{11} + a_{22}b_{21})(a_{11}b_{12} + a_{12}b_{22}) \\
&= a_{11}a_{22}b_{11}b_{22} + a_{11}a_{21}b_{11}b_{12} + a_{12}a_{21}b_{21}b_{12} + a_{12}a_{22}b_{21}b_{22} - a_{11}a_{21}b_{11}b_{12} - a_{12}a_{21}b_{11}b_{22} - a_{11}a_{22}b_{12}b_{21} - a_{12}a \\
&= a_{11}a_{22}(b_{11}b_{22} - b_{12}b_{21}) - a_{12}a_{21}(b_{11}b_{22} - b_{21}b_{12}) \\
&= (a_{11}a_{22} - a_{12}a_{21})(b_{11}b_{22} - b_{12}b_{21})
\end{aligned}
$$

One consequence of this is that, although $\mathbf{A}\,\mathbf{B} \neq \mathbf{B}\,\mathbf{A}$, their determinants are equal. In the first example that I gave of matrix multiplication, we see that $|\,\mathbf{A}\,\mathbf{B}\,| = |\,\mathbf{B}\,\mathbf{A}\,| = -1$.

4.5.6 Vector product represented as a determinant

One useful application of the determinant is as follows. Consider the determinant

$$
|M| = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix}
$$

Work out the cofactors for the first row. For $\hat{\underline{\mathbf{i}}}$ the cofactor is $a_y b_z - a_z b_y$, for $\hat{\underline{\mathbf{j}}}$ it is $-a_x b_z + a_z b_x$ and for $\hat{\underline{\mathbf{k}}}$ it is $a_x b_y - a_y b_x$. So $|M| = \mathbf{A} \times \mathbf{B}$ where $\mathbf{A} = a_x\hat{\underline{\mathbf{i}}} + a_y\hat{\underline{\mathbf{j}}} + a_z\hat{\underline{\mathbf{k}}}$ and $\mathbf{B} = b_x\hat{\underline{\mathbf{i}}} + b_y\hat{\underline{\mathbf{j}}} + b_z\hat{\underline{\mathbf{k}}}$. This might help you remember how to calculate either vector products or determinants!

4.5.7 Scaling of volume, and the Jacobian

One reason determinants are useful in physics is that if a 2D matrix has $|\mathbf{A}| = a$, then any transformation scales the area by a factor of a. Likewise, for a 3D transformation, volume is scaled by the determinant of the transformation. The notation $detA = |A|$ is suggestive of this; the determinant is in some ways analogous to the modulus of a number or the magnitude of a vector.

As a start, consider the effect of a few simple matrices on the unit square. We can arrange the four column vectors giving the vertices of the unit square in a matrix:

$$
\begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix} \tag{4.43}
$$

and then apply various transformations to them, for example, stretch along the x-axis (all x-coordinates multiplied by two).

$$
\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 2 & 2 \\ 0 & 1 & 0 & 1 \end{pmatrix} \tag{4.44}
$$

The area of the stretched square (now a rectangle) is $2 \times 1 = 2$, and the determinant of the transformation matrix is also 2. So in this case it is clear that the transformed area (2) is equal to the determinant of the transformation matrix (2) multiplied by the original area (1). This is a fairly trivial example of a general rule.

This can be generalised by considering the effect of a linear transformation on the volume (or area) element. Earlier on in the course we looked at how the volume element changes for various different coordinate transformations. We can now write down a general way of finding expressions like this for any coordinate transformation, or equivalently a change of variables in an integration.
Consider two sets of coordinates, (x_1, x_2, x_3) and (x'_1, x'_2, x'_3). Using partial derivatives we can write down the expression for a small change in the prime coordinates, given a small change in the unprimed ones:

$$\begin{aligned} dx'_1 &= \frac{\partial x'_1}{\partial x_1}dx_1 + \frac{\partial x'_1}{\partial x_2}dx_2 + \frac{\partial x'_1}{\partial x_3}dx_3 \\ dx'_2 &= \frac{\partial x'_2}{\partial x_1}dx_1 + \frac{\partial x'_2}{\partial x_2}dx_2 + \frac{\partial x'_2}{\partial x_3}dx_3 \\ dx'_3 &= \frac{\partial x'_3}{\partial x_1}dx_1 + \frac{\partial x'_3}{\partial x_2}dx_2 + \frac{\partial x'_3}{\partial x_3}dx_3 \end{aligned}$$

As a matrix equation this is:

$$\begin{pmatrix} dx'_1 \\ dx'_2 \\ dx'_3 \end{pmatrix} = \begin{pmatrix} \frac{\partial x'_1}{\partial x_1} & \frac{\partial x'_1}{\partial x_2} & \frac{\partial x'_1}{\partial x_3} \\ \frac{\partial x'_2}{\partial x_1} & \frac{\partial x'_2}{\partial x_2} & \frac{\partial x'_2}{\partial x_3} \\ \frac{\partial x'_3}{\partial x_1} & \frac{\partial x'_3}{\partial x_2} & \frac{\partial x'_3}{\partial x_3} \end{pmatrix} \begin{pmatrix} dx_1 \\ dx_2 \\ dx_3 \end{pmatrix}$$

or, more compactly,

$$dr' = Jdr.$$

The determinant of J gives the factor by which the volume element changes when we make the transformation. $|J|$ is called the Jacobian. Evaluating it for the case $(x'_1, x'_2, x'_3) = (x, y, z)$ and $(x_1, x_2, x_3) = (r, \theta, \phi)$ or (r, θ, z) gives the factors we quoted in the Section for spherical and cylindrical coordinates respectively.

4.5.8 Example

Work out the volume element in spherical polar coordinates, given that the volume element in cartesians in $dxdydz$.

Let $x'_1 = x, x'_2 = y, x'_3 = z$ and $x_1 = r, x_2 = \theta, x_3 = \phi$. And we have $x = r\sin\theta\cos\phi, y =$

$r \sin\theta \sin\phi, z = r\cos\theta$. so

$$\begin{aligned} dx &= \frac{\partial x}{\partial r}dr + \frac{\partial x}{\partial \theta}d\theta + \frac{\partial x}{\partial \phi}d\phi \\ dy &= \frac{\partial y}{\partial r}dr + \frac{\partial y}{\partial \theta}d\theta + \frac{\partial y}{\partial \phi}d\phi \\ dz &= \frac{\partial z}{\partial r}dr + \frac{\partial z}{\partial \theta}d\theta + \frac{\partial z}{\partial \phi}d\phi \end{aligned}$$

i.e.

$$\begin{aligned} dx &= \sin\theta\cos\phi dr + r\cos\theta\cos\phi d\theta - r\sin\theta\sin\phi d\phi \\ dy &= \sin\theta\sin\phi dr + r\cos\theta\sin\phi d\theta + r\sin\theta\cos\phi d\phi \\ dz &= \cos\theta dr - r\sin\theta d\theta \end{aligned}$$

So in matrix form we have: and is

$$\begin{pmatrix} dx \\ dy \\ dz \end{pmatrix} = \begin{pmatrix} \sin\theta\cos\phi & r\cos\theta\cos\phi & -r\sin\theta\sin\phi \\ \sin\theta\sin\phi & r\cos\theta\sin\phi & r\sin\theta\cos\phi \\ \cos\theta & -r\sin\theta & 0 \end{pmatrix} \begin{pmatrix} dr \\ d\theta \\ d\phi \end{pmatrix}$$

So the volume element under this transformation scales by

$$\begin{vmatrix} \sin\theta\cos\phi & r\cos\theta\cos\phi & -r\sin\theta\sin\phi \\ \sin\theta\sin\phi & r\cos\theta\sin\phi & r\sin\theta\cos\phi \\ \cos\theta & -r\sin\theta & 0 \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta\cos\phi & r\cos\theta\cos\phi & -r\sin\theta\sin\phi \\ \sin\theta\sin\phi & r\cos\theta\sin\phi & r\sin\theta\cos\phi \end{vmatrix}$$

(swapping two rows twice leaves determinant unchanged). This evaluates to:

$$\cos\theta(r^2(\cos\theta\sin\theta(\cos^2\phi + \sin^2\phi)) + r^2\sin\theta\sin^2\theta(\cos^2\phi + \sin^2\phi)$$

$$\begin{aligned} &= r^2\sin\theta(\cos^2\theta + \sin^2\theta) \\ &= r^2\sin\theta \end{aligned}$$

4.6 Some properties of Matrices

4.6.1 The identity or unit matrix

The multiplicative identity is the matrix which leaves any matrix unchanged when multiplied by it, i.e.

$$\mathbf{IA} = \mathbf{AI} = \mathbf{A}$$

for any square matrix $\mathbf{A}$. Thus, if $\mathbf{A}$ is a $n \times n-$matrix, $\mathbf{I}$ is a matrix with 1 along the diagonal and zero elsewhere;

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

Equivalently, in component notation, let $\mathbf{A}$ be an $n \times n$ matrix and $\mathbf{I}$ the $n \times n$ unit matrix. Then

$$(\mathbf{A}\,\mathbf{I})_{ij} = \sum_k a_{ik}\,\delta_{kj} = a_{ij}\,,$$

since the Kronecker-delta δ_{ij} vanishes unless $i = j$. Thus

$$\mathbf{A}\,\mathbf{I} = \mathbf{A}\,. \tag{4.45}$$

Similarly

$$(\mathbf{I}\,\mathbf{A})_{ij} = \sum_k \delta_{ik}\,a_{kj} = a_{ij}\,,$$

and

$$\mathbf{I}\,\mathbf{A} = \mathbf{A}\,. \tag{4.46}$$

The multiplication on the left or right by $\mathbf{I}$ does not change a matrix $\mathbf{A}$. In particular, the unit matrix $\mathbf{I}$ (or any multiple of it) commutes with any other matrix of the appropriate size.

4.6.2 The transpose of a matrix

To find the transpose of a matrix, switch rows for columns.

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

or for each component:

$$(\mathbf{A}^T)_{ij} = a_{ji}.$$

The transpose of an $n \times m$ matrix is an $m \times n$ matrix.

Transpose of a transpose

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

Symmetric matrices

If $\mathbf{A}^T = \mathbf{A}$, $\mathbf{A}$ is symmetric.
If $\mathbf{A}^T = -\mathbf{A}$, $\mathbf{A}$ is anti-symmetric.

Transposing matrix products

Transposing a product of matrices reverses the order of multiplication.

$$(\mathbf{A}\,\mathbf{B})^T = \mathbf{B}^T\,\mathbf{A}^T\,. \tag{4.47}$$

To prove this, try comparing the ij entries of

$$(\mathbf{A}\,\mathbf{B})^T \tag{4.48}$$

with those of

$$\mathbf{B}^T\,\mathbf{A}^T \tag{4.49}$$

Let's say $\mathbf{C} = \mathbf{AB}$. Then the typical elements of $\mathbf{C}$ are:

$$c_{ij} = \sum_{k=1}^{n} a_{ik}\,b_{kj}\,.$$

By definying element of $\mathbf{C}^T$, c_{ji}, then:

$$c_{ji} = \sum_{k=1}^{n} b_{jk}\,a_{ki}$$

Transposing a product of matrices reverses the order of multiplication. This is true no matter how many terms there are;

$$(\mathbf{A}\,\mathbf{B}\,\mathbf{C})^T = \mathbf{C}^T\,\mathbf{B}^T\,\mathbf{A}^T\,.$$

This rule, which is also true for operators, will be used by the Quantum Mechanics lecturers in the second and third years.

4.6.3 Equal Matrices

Two matrices $\mathbf{A}$ and $\mathbf{B}$ are equal if they have the same number n of rows and m of columns and if all of the corresponding elements are equal.

4.6.4 Multiplication by a scalar

$$\mathbf{B} = \lambda\mathbf{A} \implies b_{ij} = \lambda a_{ij}\,.$$

Orthogonal Matrices

If $\mathbf{A}^T\mathbf{A} = \mathbf{I}$, $\mathbf{A}$ is an **orthogonal** matrix. As an example let's check that the two-dimensional rotation matrix

$$\mathbf{A} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} .$$

is orthogonal. For this, you need $\cos^2\phi + \sin^2\phi = 1$. Matrix $\mathbf{A}$ rotates the system through angle ϕ, while the transpose matrix $\mathbf{A}^T$ rotates it back through angle $-\phi$. Because of this, orthogonal matrices are of great practical use in different branches of Physics. Taking the determinant, and using the determinant of a product rule(Eq. 4.42) gives

$$| \mathbf{A}^T || \mathbf{A} |=| \mathbf{I} |= 1 .$$

But the determinant of a transpose of a matrix is the same as the determinant of the original matrix — it doesn't matter if you switch rows and columns in a determinant. Hence

$$| \mathbf{A} || \mathbf{A} |=| \mathbf{A} |^2 = 1 ,$$

so $| \mathbf{A} |= \pm 1$.

Product of Orthogonal Matrices

Suppose $\mathbf{A}$ and $\mathbf{B}$ are orthogonal matrices. Their product $\mathbf{C} = \mathbf{A}\,\mathbf{B}$ is also orthogonal.

$$\mathbf{C}^T\mathbf{C} = (\mathbf{A}\,\mathbf{B})^T(\mathbf{A}\,\mathbf{B}) = \mathbf{B}^T\mathbf{A}^T\mathbf{A}\,\mathbf{B} = \mathbf{B}^T\mathbf{I}\,\mathbf{B} = \mathbf{B}^T\mathbf{B} = \mathbf{I} .$$

Physical meaning of our example: since the matrix for rotation about the x-axis is orthogonal and so is rotation about the y-axis, then the matrix for rotation about the y-axis followed by one about the x-axis is also orthogonal.

4.6.5 Complex conjugation

To take the complex conjugate of a matrix, just complex-conjugate all its elements:

$$(\mathbf{A}^*)_{ij} = a^*_{ij} . \tag{4.50}$$

For example

$$\mathbf{A} = \begin{pmatrix} -i & 0 \\ 3-i & 6+i \end{pmatrix} \Longrightarrow \mathbf{A}^* = \begin{pmatrix} +i & 0 \\ 3+i & 6-i \end{pmatrix} .$$

If $\mathbf{A} = \mathbf{A}^*$, the matrix is real.

4.6.6 Hermitian conjugation

Hermitian conjugation combines complex conjugation and transposition. It is probably more important than either – especially in Quantum Mechanics. The Hermitian conjugate matrix is sometimes called the **Hermitian adjoint**, and is usually denoted by a dagger (†).

$$\mathbf{A}^\dagger = (\mathbf{A}^T)^* = (\mathbf{A}^*)^T \, . \tag{4.51}$$

Thus $(\mathbf{A}^\dagger)^\dagger = \mathbf{A}$. For example

$$\mathbf{A} = \begin{pmatrix} -i & 0 \\ 3-i & 6+i \end{pmatrix} \Longrightarrow \mathbf{A}^\dagger = \begin{pmatrix} +i & 3+i \\ 0 & 6-i \end{pmatrix} .$$

If $\mathbf{A}^\dagger = \mathbf{A}$, $\mathbf{A}$ is Hermitian.
If $\mathbf{A}^\dagger = -\mathbf{A}$, $\mathbf{A}$ is anti-Hermitian.

All real symmetric matrices are Hermitian, but there are also other possibilities. Eg

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

is Hermitian.

The rule for Hermitian conjugates of products is the same as for transpositions:

$$(\mathbf{A}\,\mathbf{B})^\dagger = \mathbf{B}^\dagger\,\mathbf{A}^\dagger \, . \tag{4.52}$$

4.6.7 Unitary Matrices

Matrix $\mathbf{U}$ is unitary if

$$\mathbf{U}^\dagger\,\mathbf{U} = \mathbf{I} \, . \tag{4.53}$$

Unitary matrices are very important in Quantum Mechanics!

Again consider the determinant product rule, Eq. 4.42.

$$\mid \mathbf{U}^\dagger \mid\mid \mathbf{U} \mid = \mid \mathbf{I} \mid = 1 \, .$$

Changing rows and columns in a determinant does nothing, but Hermitian conjugate also involves complex conjugation. Hence

$$\mid \mathbf{U} \mid^* \mid \mathbf{U} \mid = 1 \, ,$$

and so $\mid \mathbf{U} \mid = e^{i\phi}$, with ϕ being real.

4.6.8 The Trace of a matrix

The trace is the sum of the diagonal elements:

$$Tr(\mathbf{A}) = a_{11} + a_{22} + a_{33} = \Sigma a_{ii} = a_{ii}$$

where I have introduced the Einstein summation convention that **repeated indices in a term imply a sum**. Note that

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

4.7 The multiplicative inverse of a matrix

The inverse of a matrix is defined as $\mathbf{A}^{-1}$ where

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$$

where $\mathbf{I}$ is the identity matrix. Thus inverting a matrix (i.e. finding its identity) is useful for solving matrix equations.

To just quote the result first, the inverse is given by

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\mathbf{C}^T \tag{4.54}$$

where $\mathbf{C}$ is the matrix made up of the cofactors of $\mathbf{A}$ (see Section 4.5.2). $\mathbf{C}^T$ is known as the **adjoint matrix** to $\mathbf{A}$, denoted as $\mathbf{A}^{\mathrm{adj}}$.

Another way of writing this is

$$(\mathbf{A}^{-1})_{ij} = \frac{1}{|\mathbf{A}|}c_{ji}$$

I will now try and motivate this general result using a 2×2 example.

4.7.1 Explicit 2×2 evaluation

Consider

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix} \text{ and } \mathbf{B} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} .$$

We need to determine unknown numbers a, b, c, d from the condition

$$\mathbf{B}\,\mathbf{A} = \begin{pmatrix} a + 4b & 2a + 3b \\ c + 4d & 2c + 3d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} ,$$

[Note that here what we are saying is that the product of the two matrices gives the Identity matrix i.e I will show you that the unknown matrix **B** is the inverse of matrix A]
This gives

$$\begin{aligned} a + 4b &= 1 \qquad a + \tfrac{3}{2}b = 0 , \\ c + 4d &= 0 \qquad c + \tfrac{3}{2}d = \tfrac{1}{2} . \end{aligned}$$

These simultaneous equations have solutions $a = -\frac{3}{5}$, $b = \frac{2}{5}$, $c = \frac{4}{5}$, and $d = -\frac{1}{5}$. In matrix form then

$$\mathbf{B} = \mathbf{A}^{-1} = \tfrac{1}{5}\begin{pmatrix} -3 & 2 \\ 4 & -1 \end{pmatrix} .$$

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix} \text{ and } (\mathbf{A}^{-1})^T = -\tfrac{1}{5}\begin{pmatrix} 3 & -4 \\ -2 & 1 \end{pmatrix} .$$

Notice that inside the bracket, all the coefficients are exchanged across the diagonal between $\mathbf{A}$ and $\mathbf{A}^{-1}$. There are a couple of minus signs, but these are coming in exactly the positions that one gets minus signs when expanding out a 2×2 determinant. The only remaining puzzle is the origin of the factor $-\frac{1}{5}$. Well this is precisely

$$\frac{1}{|\,A\,|} = \frac{1}{(1 \times 3 - 4 \times 2)} = -\frac{1}{5}\,.$$

The determinant $|\,A\,|$ has come in useful after all.

This simple observation is true for the inverse of any 2×2 matrix. Consider

$$\mathbf{A} = \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix}.$$

According to the hand-waving observation above, one would expect

$$\mathbf{A}^{-1} = \frac{1}{(\alpha\delta - \beta\gamma)} \begin{pmatrix} \delta & -\gamma \\ -\beta & \alpha \end{pmatrix}.$$

Verify that the $\mathbf{A}^{-1}$ defined in this way does indeed satisfy $\mathbf{A}^{-1}\,\mathbf{A} = \mathbf{I}$.

IMPORTANT: Do not forget the minus signs and do not forget to transpose the matrix afterward.

This justifies the expression in Eq. 4.54, at least for 2×2 matrices. For a 3×3 matrix one can again write down the most general form, carry out the operations outlined above, and show explicitly that $\mathbf{A}^{-1}\,\mathbf{A} = \mathbf{I}$. Eq. (4.54) is valid for any size matrix, but in this course you won't need to work out anything bigger than 4×4.

Example

Find the inverse of

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

Matrix of minors is

$$\mathbf{M} = \begin{pmatrix} -4 & -2 & -2 \\ 5 & 2 & 3 \\ -8 & -2 & -4 \end{pmatrix}.$$

Cofactor matrix changes a few signs to give

$$\mathbf{C} = \begin{pmatrix} -4 & +2 & -2 \\ -5 & 2 & -3 \\ -8 & 2 & -4 \end{pmatrix}.$$

Adjoint matrix involves changing rows and columns:

$$\mathbf{A}^{\mathrm{adj}} = \begin{pmatrix} -4 & -5 & -8 \\ +2 & 2 & 2 \\ -2 & -3 & -4 \end{pmatrix}.$$

Now

$$| A |= -1 \times (-4) - 2 \times (-2) + 3 \times (-2) = 2 \,.$$

Hence

$$\mathbf{A}^{-1} = \frac{1}{2} \begin{pmatrix} -4 & -5 & -8 \\ 2 & 2 & 2 \\ -2 & -3 & -4 \end{pmatrix} .$$

Can check that this is right by doing the explicit $\mathbf{A}^{-1}\mathbf{A}$ multiplication.

Note that if $| A |= 0$, we say the determinant is singular; $\mathbf{A}^{-1}$ does not exist . [It has some infinite elements.]

There are lots of other ways to do matrix inversion: Gaussian or Gauss-Jordan elimination, as described by Boas. These methods become more important as the size of the matrix goes up.

4.7.2 Properties of the inverse matrix

a) $\mathbf{A}\,\mathbf{A}^{-1} = \mathbf{A}^{-1}\,\mathbf{A} = \mathbf{I}$; a matrix commutes with its inverse.

b) $(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1}$; the operations of inversion and transposition commute.

c) If $\mathbf{C} = \mathbf{A}\,\mathbf{B}$, what is $\mathbf{C}^{-1}$? Consider

$$\mathbf{B}^{-1}\,\mathbf{A}^{-1}\,\mathbf{I} = \mathbf{B}^{-1}\,\mathbf{A}^{-1}\,\mathbf{C}\,\mathbf{C}^{-1} = \mathbf{B}^{-1}\,\mathbf{A}^{-1}\,\mathbf{A}\,\mathbf{B}\,\mathbf{C}^{-1} = \mathbf{B}^{-1}\,\mathbf{B}\,\mathbf{C}^{-1} = \mathbf{C}^{-1} = (\mathbf{A}\,\mathbf{B})^{-1} \,.$$

Hence

$$(\mathbf{A}\,\mathbf{B})^{-1} = \mathbf{B}^{-1}\,\mathbf{A}^{-1} \,. \tag{4.55}$$

reverse the order before inverting each matrix.

d) If $\mathbf{A}$ is orthogonal, *i.e.* $\mathbf{A}^T\,\mathbf{A} = \mathbf{I}$, then $\mathbf{A}^{-1} = \mathbf{A}^T$.

e) If $\mathbf{A}$ is unitary, *i.e.* $\mathbf{A}^\dagger\,\mathbf{A} = \mathbf{I}$, then $\mathbf{A}^{-1} = \mathbf{A}^\dagger$.

f) Using the determinant of a product rule, Eq. 4.42, it follows immediately that $| \mathbf{A}^{-1} |= 1/\, | \mathbf{A} |$.

g) Division of matrices is not really defined, but one can multiply by the inverse matrix. Unfortunately, in general,

$$\mathbf{A}\,\mathbf{B}^{-1} \neq \mathbf{B}^{-1}\,\mathbf{A} \,.$$

4.8 Solution of Linear Simultaneous Equations

We have already seen in fact that simultaneous equations of form

$$\begin{aligned} a_{11}\,x_1 + a_{12}\,x_2 + a_{13}\,x_3 &= b_1\,, \\ a_{21}\,x_1 + a_{22}\,x_2 + a_{23}\,x_3 &= b_2\,, \\ a_{31}\,x_1 + a_{32}\,x_2 + a_{33}\,x_3 &= b_3 \end{aligned}$$

for unknown x_i can be written in matrix form

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix},$$

that is

$$\mathbf{A}\,\underline{\mathbf{x}} = \underline{\mathbf{b}} \;\text{ or }\; \sum_j a_{ij}\,x_j = b_i\,.$$

The solution can immediately be written down then by multiplying both sides by $\mathbf{A}^{-1}$:

$$\underline{\mathbf{x}} = \mathbf{A}^{-1}\,\underline{\mathbf{b}}\,.$$

So this is where the inverse of the matrix is very useful. All that remains is to evaluate the result!

Using the previous expression for the inverse matrix, Eq. 4.54,

$$x_j = \sum_i (\mathbf{A}^{\text{adj}})_{ji}\,b_i / \mid A \mid\,.$$

There are many special cases of this formula; we will now look at two of these special cases and we will then describe a 'short-cut' to doing full matrix inversion.

4.8.1 Vanishing determinant

If $\mid A \mid = 0$. Then matrix **A** is singular the inverse matrix cannot be defined. Provided that the equations are mutually consistent, this means that (at least) one of the equations is not linearly independent of the others. So we do not in fact have n equations for n unknowns but rather only $n-1$ equations. Can only try to solve the equations for $n-1$ of the x_i in terms of the b_i and one of the x_i. It might take some trial and error to find which of the equations to throw away.

4.8.2 Homogeneous equations

If all $b_i = 0$, have to look for a solution of the homogeneous equation

$$\mathbf{A}\,\underline{\mathbf{x}} = \underline{\mathbf{0}}\,.$$

There is, of course, the uninteresting solution where all the $x_i = 0$. Can there be a more interesting solution? The answer is yes, provided that $\mid A \mid = 0$. **Note: this is very important - see 4.10.**

4.8.3 Cramer's rule

If you write it out in full you will see that $\sum_i (\mathbf{A}^{\mathrm{adj}})_{ji}\, b_i$ is the determinant obtained by replacing the j'th column of $\mathbf{A}$ by the column vector $\underline{\mathbf{b}}$. So the solution is

$$x_1 = \begin{vmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{vmatrix} \Big/ \Delta\,, \quad x_2 = \begin{vmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{vmatrix} \Big/ \Delta\,, \quad x_3 = \begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{vmatrix} \Big/ \Delta\,, \tag{4.56}$$

where

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}\,. \tag{4.57}$$

Just replace the appropriate column with the column of numbers from the right hand side. This is called Cramer's rule and can be used as a short cut to doing full matrix inversion.

Example

Use Cramer's rule to solve the following simultaneous equations just for the variable x_1:

$$\begin{aligned} 3x_1 - 2x_2 - x_3 &= 4\,, \\ 2x_1 + x_2 + 2x_3 &= 10\,, \\ x_1 + 3x_2 - 4x_3 &= 5\,. \end{aligned}$$

We can expand the determinant appearing here by the first row as

$$\Delta = \begin{vmatrix} 3 & -2 & -1 \\ 2 & 1 & 2 \\ 1 & 3 & -4 \end{vmatrix} = 3(-4-6) + 2(-8-2) - 1(6-1) = -55\,.$$

Alternatively, adding simultaneously columns 2 and 3 to column 1 gives

$$\Delta = \begin{vmatrix} 0 & -2 & -1 \\ 5 & 1 & 2 \\ 0 & 3 & -4 \end{vmatrix}\,.$$

Expand now by the first column (not forgetting the minus sign)

$$\Delta = -5(8+3) = -55\,.$$

Now by Cramer's rule,

$$\Delta \times x_1 = \begin{vmatrix} 4 & -2 & -1 \\ 10 & 1 & 2 \\ 5 & 3 & -4 \end{vmatrix} = \begin{vmatrix} 4 & -2 & -1 \\ 0 & -5 & 10 \\ 5 & 3 & -4 \end{vmatrix} = \begin{vmatrix} 4 & -2 & -5 \\ 0 & -5 & 0 \\ 5 & 3 & 2 \end{vmatrix} = -5(8+25) = -165\,.$$

Hence $x_1 = 3$.

4.9 Scalar Product using matrices

Finally, we can also write the scalar product of two vectors as a matrix operation.

$$\begin{aligned}
\mathbf{V}\cdot\mathbf{W} &= V_i W_i \\
&= V^T W \\
&= (V_1\, V_2\, V_3)\begin{pmatrix} W_1 \\ W_2 \\ W_3 \end{pmatrix} \\
&= V_1 W_1 + V_2 W_2 + V_3 W_3
\end{aligned}$$

In fact, in general hidden in the middle here is another matrix known as the *metric*, G, of the space in which the vectors are defined;

$$\mathbf{V}\cdot\mathbf{W} = V^T G W = (V_1\, V_2\, V_3)\begin{pmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{pmatrix}\begin{pmatrix} W_1 \\ W_2 \\ W_3 \end{pmatrix}$$

This metric defines how the different coordinates combine to give length elements. For Cartesian coordinates it is equal to the identity matrix, since $ds^2 = dx^2 + dy^2 + dz^2$, so we can ignore it. For cylindrical polars, $ds^2 = dr^2 + r^2 d\theta^2 + dz^2$ and the metric is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

For orthogonal coordinate systems, the metric will always be diagonal (see Section 4.10.2).

4.10 Eigenvalues and eigenvectors of a linear operator

Matrices are very useful in solving many technological problems e.g. those involving coupled oscillations (as we shall see at the end of the course). These problems give rise to equations such as:

$$\mathbf{M}\underline{\mathbf{v}} = \lambda\underline{\mathbf{v}}\,. \tag{4.58}$$

where $\mathbf{M}$ is a matrix, $\underline{\mathbf{v}}$ a column vector and λ a scalar quantity. For non trivial solutions ($\lambda \neq 0$) the values of λ are called the *eigenvalues* of the matrix $\mathbf{M}$ and the corresponding solutions of the given equation are called the *eigenvectors*. [Note: you will also find that the eigenvalues are called *characteristic values* and the eigenvectors *characteristic vectors*].

Eigenvalues can be determined by considering that $\mathbf{M} = \mathbf{IM}$, and hence Eq. (4.58) can be recast as follows:

$$\mathbf{IM}\underline{\mathbf{v}} = \lambda\mathbf{I}\underline{\mathbf{v}} \quad\Rightarrow\quad \mathbf{M}\underline{\mathbf{v}} = \lambda\mathbf{I}\underline{\mathbf{v}} \quad\Rightarrow\quad (\mathbf{M} - \lambda\mathbf{I})\underline{\mathbf{v}} = 0\,,$$

where we have just inserted the identity matrix $\mathbf{I}$ in front of a vector, as can always be done (because remember that $\mathbf{M} = \mathbf{IM}$).

Now, we know from Sect 4.8.2. that a non-zero vector $\underline{\mathbf{v}}$ satisfying the previous equation exists if and only if $\det(\mathbf{M} - \lambda\mathbf{I}) = 0$ (*note that the null vector is a trivial eigenvector of any linear operator: it will be henceforth always disregarded*). We have thus determined the condition for a scalar λ to be eigenvalue of $\mathbf{M}$: λ must be a root of the so-called "characteristic equation", given by

$$\det(\mathbf{M} - \lambda\mathbf{I}) = 0 \,. \tag{4.59}$$

What is the nature of the eigenvalues? Let's expand $\det(\mathbf{M} - \lambda\mathbf{I}) = 0$:

$$\det(\mathbf{M} - \lambda\mathbf{I}) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} - \lambda & \ldots & a_{2n} \\ .. & .. & \ldots & .. \\ a_{n1} & a_{n2} & \ldots & a_{nn} - \lambda \end{vmatrix}$$

where the a_{nn} are the elements of the matrix $\mathbf{M}$ and where the eigenvalues λ are the *roots* of Eq. 4.59 As λ appears only once in each row (column) then the expansion of this determinant is a polynomial of degree n in λ. The polynomial $\det(\mathbf{M} - \lambda\mathbf{I})$ is referred to as the *characteristic polynomial* of $\mathbf{M}$.

More than one linearly independent eigenvector may correspond to the same eigenvalue λ_j because multiplying an eigenvector by a given scalar yields another eigenvector associated to the same eigenvalue. This can be easily proven: assume that $\underline{\mathbf{v}}_j$ is an eigenvector of $\mathbf{M}$ with eigenvalue λ_j (so that $\mathbf{M}\underline{\mathbf{v}}_j = \lambda_j\underline{\mathbf{v}}_j$), and consider the action of $\mathbf{M}$ on the vector $c\underline{\mathbf{v}_\mathbf{j}}$, where $c \in \mathbb{C}$ is any scalar, then:

$$\mathbf{M}(c\underline{\mathbf{v}}_j) = c\mathbf{M}\underline{\mathbf{v}}_j = c\lambda_j\underline{\mathbf{v}}_j = \lambda_j(c\underline{\mathbf{v}}_j)\,,$$

the third equality is true because $\mathbf{M}\underline{\mathbf{v}} = \lambda\underline{\mathbf{v}}$; which proves that $c\underline{\mathbf{v}}_j$ is also an eigenvector with eigenvalue λ_j. Hence, eigenvectors can only be determined up to an arbitrary multiplicative factor.

Let us now specify a systematic, pragmatic recipe to determine eigenvalues and eigenvectors of a linear operator:

1. Write down the characteristic equation (eq 4.59) and solve it to find the eigenvalues λ_j.

2. For each λ, you will have an eigenvector $\underline{\mathbf{v}}$ to give:

$$(\mathbf{M} - \lambda_k\mathbf{I})\underline{\mathbf{v}_\mathbf{k}} = 0 \tag{4.60}$$

where $k = 1,2,.....,n$ (equivalent to the number of eigenvalues).

3. Set one of the elements arbitrarily: to do this, pick one element, say v_1, and make sure the eigensystem does not imply $v_1 = 0$. Them simply set $v_1 = 1$ and solve the eigensystem.

4. Since as we mentioned earlier an eigenvector is indeterminate up to a scalar multiplier, it is convenient to determine *normalised* eigenvectors, *i.e.* eigenvectors of modulus 1. For any eigenvector with real elements v_1, v_2 etc, the corresponding **normalized eigenvector** $\underline{\mathbf{v}}$ is defined as:

$$\underline{\mathbf{v}} = \frac{1}{\|\underline{\mathbf{v}}\|}\begin{pmatrix} v_1 \\ v_2 \\ \cdots \\ v_n \end{pmatrix} \tag{4.61}$$

where $\|\underline{\mathbf{v}}\|$ is the **norm** or Euclidian length of $\underline{\mathbf{v}}$.

5. You then can calculate the normalized eigenvector by remembering that the **norm** or Euclidian length can be evaluated in the folllowing way:

$$\|\underline{\mathbf{v}}\| = \left(v_1^2 + v_2^2 + ... + v_n^2\right)^{1/2} \tag{4.62}$$

Recall now the above session on the scalar product using matrices and see that:

$$\|\underline{\mathbf{v}}\| = \left(\underline{\mathbf{v}}^T \underline{\mathbf{v}}\right)^{1/2} \tag{4.63}$$

Let us now see a concrete example. Consider the 2×2 matrix $\mathbf{A}$:

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

The characteristic equation $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ reads (point **1**)

$$\det\begin{pmatrix} -\lambda & 1 \\ 1 & -\lambda \end{pmatrix} = \lambda^2 - 1 = (\lambda + 1)(\lambda - 1) = 0 .$$

The two solutions of the equation are hence $\lambda_1 = -1$ and $\lambda_2 = 1$. These are the two eigenvalues of $\mathbf{A}$. To find the eigenvector $\underline{\mathbf{v_1}}$ corresponding to λ_1, one has to set

$$\underline{\mathbf{v_1}} = \begin{pmatrix} a \\ b \end{pmatrix}$$

(two rows because the matrix has two rows) and solve the eigensystem $\mathbf{A}\underline{\mathbf{v_1}} = \lambda_1\underline{\mathbf{v_1}}$ for the variables a and b (point **2**):

$$\mathbf{A}\underline{\mathbf{v}}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} b \\ a \end{pmatrix} = \lambda_1\underline{\mathbf{v_1}} = -\underline{\mathbf{v}}_1 = \begin{pmatrix} -a \\ -b \end{pmatrix} .$$

(because λ = -1) Being a vector equation, the previous equality implies two scalar equations: $b = -a$ and $a = -b$. These two equations are identical. The only relevant equation is then

$$b = -a .$$

This equation does not imply $a = 0$. Hence, according to point **3**, we can set $a = 1$ and have

$$\underline{\mathbf{v_1}} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} .$$

Finally, the vector $\underline{\mathbf{v}_1}$ is normalised as per point **4 and 5**:

$$\underline{\mathbf{v}}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} / \sqrt{1^2 + (-1)^2} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} / \sqrt{2}\,.$$

The eigenvector $\underline{\mathbf{v}}_2$ associated to $\lambda_2 = +1$ can be determined as

$$\underline{\mathbf{v}}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} / \sqrt{2}\,.$$

The eigenvalues and eigenvectors of real matrices are not necessarily real: consider for example the matrix $\mathbf{B}$:

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

The characteristic equation $\det(\lambda \mathbf{I} - \mathbf{B}) = 0$ reads (point **1**)

$$\det \begin{pmatrix} \lambda & -1 \\ 1 & \lambda \end{pmatrix} = \lambda^2 + 1 = (\lambda - i)(\lambda + i) = 0\,.$$

The two solutions of the equation are hence $\lambda_1 = -i$ and $\lambda_2 = i$. These are the two eigenvalues of $\mathbf{B}$. To find the non-normalized eigenvector $\underline{\mathbf{v}}_1$ corresponding to λ_1, one has to set

$$\underline{\mathbf{v}}_1 = \begin{pmatrix} a \\ b \end{pmatrix}$$

and solve the eigensystem $\mathbf{B}\underline{\mathbf{v}}_1 = \lambda_1 \underline{\mathbf{v}}_1$ for the variables a and b (point **2**):

$$\mathbf{B}\underline{\mathbf{v}}_1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} b \\ -a \end{pmatrix} = \lambda_1 \underline{\mathbf{v}}_1 = -i\underline{\mathbf{v}}_1 = \begin{pmatrix} -ia \\ -ib \end{pmatrix}.$$

As expected, the vector equation above corresponds to a single independent scalar equation (apparent if one multiplies one of the two rows by i). Setting $a = 1$, as per point **3**, determines $b = -i$, such that

$$\underline{\mathbf{v}}_1 = \begin{pmatrix} 1 \\ -i \end{pmatrix}.$$

The vector $\underline{\mathbf{v}}_1$ is normalised as per point **4**, recalling the rules for the inner product of complex vectors:

$$\underline{\mathbf{v}}_1 = \begin{pmatrix} 1 \\ -i \end{pmatrix} / \sqrt{1^2 + i(-i)} = \begin{pmatrix} 1 \\ -i \end{pmatrix} / \sqrt{2}\,.$$

It can be shown following the same steps that the normalised eigenvector $\underline{\mathbf{v}}_2$ corresponding to $\lambda_2 = +i$ is given by

$$\underline{\mathbf{v}}_2 = \begin{pmatrix} -i \\ 1 \end{pmatrix} / \sqrt{2}\,.$$

4.10.1 Degenerate eigenvalues

As already pointed out, the characteristic equation (eq 4.59), which determines the eigenvalues of the linear operator **M**, is an algebraic equation of order n. If two or more solutions of the equation coincide, the characteristic polynomial is said to have *degenerate* roots. In that case, clearly, there will be less than n eigenvalues.

However, the eigenvalue corresponding to a degenerate root may have a number of corresponding eigenvectors up to its multiplicity (this is not necessarily the case!). Such eigenvectors are determined by the eigensystem, like any eigenvector, but for them the eigensystem might feature more than one redundant equation, so that one may have to arbitrarily set more than one entry of the eigenvector.

This situation is better illustrated with a concrete example. Consider the matrix **B**:

$$\mathbf{B} = \begin{pmatrix} 5 & 1 & 2 \\ 1 & 5 & -2 \\ 2 & -2 & 2 \end{pmatrix} .$$

The characteristic equation $\det(\lambda\mathbf{I} - \mathbf{B}) = 0$ reads

$$\det\begin{pmatrix} \lambda-5 & -1 & -2 \\ -1 & \lambda-5 & 2 \\ -2 & 2 & \lambda-2 \end{pmatrix} = \lambda^3 - 12\lambda^2 + 36\lambda = \lambda(\lambda-6)^2 = 0 ,$$

with solutions 0 and 6. These are the only two eigenvalues of the matrix **B**. However, note that the eigenvalue 6 is a root of the characteristic equation with multiplicity 2 (in that the factor $(\lambda-6)$ to the power 2 occurs in the characteristic polynomial). As anticipated above, the eigenvalue 6 may have up to 2 corresponding linearly independent eigenvectors, in which case it is said to be *degenerate* ('doubly' degenerate, in this instance).

The eigenvector $\underline{\mathbf{v}}_1$ corresponding to $\lambda_1 = 0$ is found by setting

$$\underline{\mathbf{v}}_1 = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

and solving the eigensystem $\mathbf{B}\underline{\mathbf{v}}_1 = \lambda_1\underline{\mathbf{v}}_1$ for the variables a, b and c:

$$\mathbf{B}\underline{\mathbf{v}}_1 = \begin{pmatrix} 5 & 1 & 2 \\ 1 & 5 & -2 \\ 2 & -2 & 2 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 5a+b+2c \\ a+5b-2c \\ 2a-2b+2c \end{pmatrix} = \lambda_1\underline{\mathbf{v}}_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} .$$

Adding up the first and second lines of the previous scalar equation yields

$$a + b = 0 ,$$

such that the third line becomes

$$4a + 2c = 0 .$$

We can then set $c = 1$ to get $a = -1/2$ and $b = 1/2$. Finally, the eigenvector $\underline{\mathbf{v}}_1$ can be normalised to obtain

$$\underline{\mathbf{v}}_1 = \begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix} / \sqrt{6} .$$

Let us now move on to the degenerate eigenvalue $\lambda_2 = 6$. Setting

$$\underline{\mathbf{v}}_2 = \begin{pmatrix} a \\ b \\ c \end{pmatrix} ,$$

the eigensystem $\mathbf{B}\underline{\mathbf{v}}_2 = \lambda_2 \underline{\mathbf{v}}_2$ for the variables a, b and c reads:

$$\mathbf{B}\underline{\mathbf{v}}_1 = \begin{pmatrix} 5 & 1 & 2 \\ 1 & 5 & -2 \\ 2 & -2 & 2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 5a + b + 2c \\ a + 5b - 2c \\ 2a - 2b + 2c \end{pmatrix} = \lambda_2 \underline{\mathbf{v}}_2 = \begin{pmatrix} 6a \\ 6b \\ 6c \end{pmatrix} .$$

The three equations from the vector equality above are all identical and equivalent to

$$a - b - 2c = 0 . \tag{4.64}$$

All vectors satisfying such a relationship are eigenvectors associated to $\lambda_2 = 6$. Clearly, however, we are left with only one condition to determine the three variables a, b and c. Two of them will have hence to be set *arbitrarily*. By setting $a = 1$ and $c = 1$, one gets

$$b = -1 ,$$

such that the first eigenvector corresponding to λ_2 is determined as

$$\underline{\mathbf{v}}_2 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} ,$$

normalised as

$$\underline{\mathbf{v}}_2 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} / \sqrt{3} .$$

Setting $a = 1$ and $c = 0$ instead gives

$$\underline{\mathbf{v}}_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} ,$$

normalised as

$$\underline{\mathbf{v}}_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} / \sqrt{2} .$$

The eigenvectors $\underline{\mathbf{v}}_2$ and $\underline{\mathbf{v}}_3$ are both associated to the same degenerate eigenvalue λ_2. Note that the expressions of $\underline{\mathbf{v}}_2$ and $\underline{\mathbf{v}}_3$ depend on the choice we made for the coefficients when solving the einsystem for λ_2. Different choices could have been made, leading to different eigenvectors $\underline{\mathbf{v}}_2$ and $\underline{\mathbf{v}}_3$. The choice we made was such as to make $\underline{\mathbf{v}}_2$ and $\underline{\mathbf{v}}_3$ orthogonal to each other: this can always be made.

4.10.2 Diagonal matrices

A diagonal matrix is a square matrix with elements only along the diagonal:

$$\mathbf{A} = \begin{pmatrix} a_{11} & 0 & 0 & \cdots \\ 0 & a_{22} & 0 & \cdots \\ 0 & 0 & a_{33} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix} .$$

(We already encountered this). Thus

$$(\mathbf{A})_{ij} = a_i \, \delta_{ij} \, .$$

(recall the Kronecker delta definiton in Section 4.2).

Now consider two diagonal matrices $\mathbf{A}$ and $\mathbf{B}$ of the same size.

$$(\mathbf{A}\,\mathbf{B})_{ij} = \sum_k a_{ik} \, b_{kj} = \sum_k a_i \, \delta_{ik} \, \delta_{kj} \, b_k = (a_i \, b_i) \, \delta_{ij} \, .$$

because if $i = j$ then the only non-zero entries are the matching ones so the product is $a_{ii} b_{ii}$ while if $i \neq j$ the entries in the matrix $\mathbf{A}\,\mathbf{B}$ are zeros. This is what the introduction of the Kronecker delta function is useful for.

Hence $\mathbf{A}\,\mathbf{B}$ is also a diagonal matrix with elements equal to the products of the corresponding individual elements. Note that for diagonal matrices, $\mathbf{A}\,\mathbf{B} = \mathbf{B}\,\mathbf{A}$, so that $\mathbf{A}$ and $\mathbf{B}$ commute.

Many problems using matrices are easier to solve if a transformation can be found which simplifies the structure of the matrix while leaving the eigenvalues unchanged. Diagonalizing a matrix will simplify it. Not all matrices can be diagonalized. Let us first use an example of a diagonalizable matrix and then we can generalize the conditions for diagonalization. Consider the matrix:

$$\mathbf{A} = \begin{pmatrix} 4 & 1 \\ 2 & 3 \end{pmatrix} \tag{4.65}$$

The eigenvalues of this matrix are: $\lambda_1 = 2$, $\lambda_2 = 5$ and the corresponding normalized eigenvectors are:

$$\underline{\mathbf{v}}_1 = \begin{pmatrix} \frac{-1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix}, \underline{\mathbf{v}}_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \tag{4.66}$$

[Calculate them as part of an exercise please!] Consider now the matrix, $\mathbf{B}$, whose columns are formed from the normalized eigenvectors of $\mathbf{A}$. Then

$$\mathbf{B} = \begin{pmatrix} \frac{-1}{\sqrt{5}} & \frac{1}{\sqrt{2}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{2}} \end{pmatrix} \mathbf{B}^{-1} = \begin{pmatrix} \frac{-\sqrt{5}}{3} & \frac{\sqrt{5}}{3} \\ \frac{2\sqrt{2}}{3} & \frac{\sqrt{2}}{3} \end{pmatrix} \tag{4.67}$$

A straightforward calculation now gives:

$$\mathbf{B}^{-1}\mathbf{A}\mathbf{B} = \begin{pmatrix} 2 & 0 \\ 0 & 5 \end{pmatrix} \tag{4.68}$$

This means that the matrix $\mathbf{A}$ is **diagonalizable**.
We now need to find the necessary and sufficient conditions for a matrix to be diagonalizable. From the above example one can infer that such condition is that for a $n\times n$ matrix to be diagonazible it has to have n linearly independent eigenvectors. The diagonalized matrix will have non-zero entries only on the main diagonal.

Let us therefore now generalize the example we used to start with. Assume that a $n \times n$ matrix $\mathbf{M}$ is diagonalisable, with a basis of eigenvectors $\underline{\mathbf{v}}_j$, $j = 1, ..., n$. Let λ_j be the eigenvalue associated to each $\underline{\mathbf{v}}_j$ so that:

$$\mathbf{M}\underline{\mathbf{v}}_j = \lambda_j \underline{\mathbf{v}}_j$$

Let $\mathbf{L}$ be defined as the $n\times n$ matrix with columns equal to the n vectors $\underline{\mathbf{v}}_j$'s. We shall adopt the following notation:

$$\mathbf{L} = \left(\underline{\mathbf{v}}_1, \dots \underline{\mathbf{v}}_j, \dots \underline{\mathbf{v}}_n\right) , \tag{4.69}$$

where each $\underline{\mathbf{v}}_j$ represents the column of entries of the vector $\underline{\mathbf{v}}_j$ in a generic basis. [Note: j is the index, n is the 'maximum' value...]

Since we assumed the matrix is diagonazable then the $\underline{\mathbf{v}}_j$s are linearly independent, and the matrix $\mathbf{L}$ must be invertible: a matrix $\mathbf{L}^{-1}$ exists such that $\mathbf{L}^{-1}\mathbf{L} = \mathbf{I}$. Let us represent $\mathbf{L}^{-1}$ as a matrix of rows:

$$\mathbf{L}^{-1} = \begin{pmatrix} \underline{\mathbf{w}}_1 \\ \vdots \\ \underline{\mathbf{w}}_k \\ \vdots \\ \underline{\mathbf{w}}_n \end{pmatrix} . \tag{4.70}$$

Then $\mathbf{L}^{-1}\mathbf{L} = \mathbf{I}$ can be expressed as

$$\begin{pmatrix} \underline{\mathbf{w}}_1 \\ \vdots \\ \underline{\mathbf{w}}_k \\ \vdots \\ \underline{\mathbf{w}}_n \end{pmatrix} \left(\underline{\mathbf{v}}_1, \dots \underline{\mathbf{v}}_j, \dots \underline{\mathbf{v}}_n\right) = \begin{pmatrix} \underline{\mathbf{w}}_1\underline{\mathbf{v}}_1 & \cdots & \underline{\mathbf{w}}_1\underline{\mathbf{v}}_j & \cdots & \underline{\mathbf{w}}_1\underline{\mathbf{v}}_n \\ \vdots & \ddots & & \iddots & \vdots \\ \underline{\mathbf{w}}_k\underline{\mathbf{v}}_1 & \cdots & \underline{\mathbf{w}}_k\underline{\mathbf{v}}_j & \cdots & \underline{\mathbf{w}}_k\underline{\mathbf{v}}_n \\ \vdots & \iddots & & \ddots & \vdots \\ \underline{\mathbf{w}}_n\underline{\mathbf{v}}_1 & \cdots & \underline{\mathbf{w}}_n\underline{\mathbf{v}}_j & \cdots & \underline{\mathbf{w}}_n\underline{\mathbf{v}}_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

The previous matrix equation can be represented as

$$\underline{\mathbf{w}}_k\underline{\mathbf{v}}_j = \delta_{jk} , \tag{4.71}$$

where the 'Kronecker delta' δ_{jk} is defined by: $\delta_{jk} = 1$ if $j = k$ and $\delta_{jk} = 0$ if $j \neq 0$, as always.

Now, let us apply the ordinary rules of matrix multiplication to evaluate the matrix $\mathbf{D} = \mathbf{L}^{-1}\mathbf{M}\mathbf{L}$, which represents the transformation $\mathbf{M}$ in the new basis of eigenvectors ('eigenbasis'):[2]

[2]Let us recall that, in general, given an invertible matrix $\mathbf{L}$ and a linear operator represented by the matrix $\mathbf{M}$, the transformation $\mathbf{L}^{-1}\mathbf{M}\mathbf{L}$ (known as 'similarity'), represents the linear operator in the new basis given by the columns of $\mathbf{L}$. The fact that the columns of $\mathbf{L}$ form a basis, *i.e.* that they are linearly independent vectors, is equivalent to state that the square matrix $\mathbf{L}$ is invertible.

$$
\begin{aligned}
\mathbf{D} &= \mathbf{L}^{-1}\mathbf{M}\left(\underline{\mathbf{v}}_1, \ldots \underline{\mathbf{v}}_j, \ldots \underline{\mathbf{v}}_n\right) = \mathbf{L}^{-1}\left(\mathbf{M}\underline{\mathbf{v}}_1, \ldots \mathbf{M}\underline{\mathbf{v}}_j, \ldots \mathbf{M}\underline{\mathbf{v}}_n\right) \\
&= \mathbf{L}^{-1}\left(\lambda_1\underline{\mathbf{v}}_1, \ldots \lambda_j\underline{\mathbf{v}}_j, \ldots \lambda_n\underline{\mathbf{v}}_n\right) = \begin{pmatrix} \underline{\mathbf{w}}_1 \\ \vdots \\ \underline{\mathbf{w}}_k \\ \vdots \\ \underline{\mathbf{w}}_n \end{pmatrix} \left(\lambda_1\underline{\mathbf{v}}_1, \ldots \lambda_j\underline{\mathbf{v}}_j, \ldots \lambda_n\underline{\mathbf{v}}_n\right) \\
&= \begin{pmatrix} \lambda_1\underline{\mathbf{w}}_1\underline{\mathbf{v}}_1 & \cdots & \lambda_j\underline{\mathbf{w}}_1\underline{\mathbf{v}}_j & \cdots & \lambda_n\underline{\mathbf{w}}_1\underline{\mathbf{v}}_n \\ \vdots & \ddots & & \iddots & \vdots \\ \lambda_1\underline{\mathbf{w}}_k\underline{\mathbf{v}}_1 & \cdots & \lambda_j\underline{\mathbf{w}}_k\underline{\mathbf{v}}_j & \cdots & \lambda_n\underline{\mathbf{w}}_k\underline{\mathbf{v}}_n \\ \vdots & \iddots & & \ddots & \vdots \\ \lambda_1\underline{\mathbf{w}}_n\underline{\mathbf{v}}_1 & \cdots & \lambda_j\underline{\mathbf{w}}_n\underline{\mathbf{v}}_j & \cdots & \lambda_n\underline{\mathbf{w}}_n\underline{\mathbf{v}}_n \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & \lambda_j & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & \lambda_n \end{pmatrix} .
\end{aligned}
$$

As anticipated above, the expression of the diagonalisable linear transformation $\mathbf{M}$ in the basis given by its eigenvectors *is a diagonal matrix* $\mathbf{D}$*, with the eigenvalues on the main diagonal and zero entries everywhere else.*

In practice, the matrix $\mathbf{L}$ which *diagonalises* the matrix $\mathbf{M}$ is just constructed as the matrix whose column are the eigenvectors of $\mathbf{M}$. For instance, for the matrix $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, the matrix $\mathbf{L}$ is given by

$$\mathbf{L} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} / \sqrt{2} .$$

The inverse $\mathbf{L}^{-1}$ is given by

$$\mathbf{L}^{-1} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} / \sqrt{2} ,$$

and the diagonalisation can be checked directly

$$\mathbf{L}^{-1}\mathbf{A}\mathbf{L} = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} .$$

As illustrated by the following counterexample, *not all square matrices can be diagonalised.* Consider in fact the matrix $\mathbf{J}$:

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

The characteristic equation for $\mathbf{J}$ reads $\lambda^2 = 0$, such that $\lambda_1 = 0$ is the only (potentially doubly degenerate) eigenvalue of $\mathbf{J}$. The generic vector is $\underline{\mathbf{v}}_1 = (a, b)^T$. Note that this is equivalent to

$$\underline{\mathbf{v}}_1 = \begin{pmatrix} a \\ b \end{pmatrix}$$

So the eigensystem for $\lambda_1 = 0$ is

$$\mathbf{J}\underline{\mathbf{v}}_1 = \lambda\underline{\mathbf{v}}_1$$

$$\mathbf{J}\underline{\mathbf{v}}_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

and so

$$\mathbf{J}\underline{\mathbf{v}}_1 = \begin{pmatrix} b \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} .$$

The only condition that can be extracted from the eigensystem is hence $b = 0$. The vector $\underline{\mathbf{v}}_1 = (1, 0)^T$ is in fact an eigenvector of $\mathbf{J}$. However, no other linearly independent eigenvector exists. In particular, the vector $\underline{\mathbf{v}}_2 = (0, 1)^T$, orthogonal to $\underline{\mathbf{v}}_1$ and which would form a basis with the latter, is not an eigenvector of $\mathbf{J}$ as can be directly verified:

$$\mathbf{J}\underline{\mathbf{v}}_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \neq \lambda_2 \underline{\mathbf{v}}_2 .$$

A matrix like $\mathbf{J}$ is referred to as a *Jordan block*. Jordan blocks are never diagonalisable.

4.10.3 Invariants and eigenvalues

It is clear from the expression determined above

$$\mathbf{D} = \mathbf{L}^{-1}\mathbf{M}\mathbf{L} \tag{4.72}$$

(where $\mathbf{D}$ is the diagonal matrix containing the eigenvalues of $\mathbf{M}$ on the main diagonal), that the eigenvalues are invariant under a change of basis of the vector space.

Let us see this with a practical example: consider a change of basis described by the invertible matrix $\mathbf{N}$, such that $\mathbf{M} \to \mathbf{M}' = \mathbf{N}^{-1}\mathbf{M}\mathbf{N}$. By multiplying R.H.S. and L.H.S by $\mathbf{N}$ and $\mathbf{N}^{-1}$ i.e:

$$\mathbf{N}\mathbf{M}'\mathbf{N}^{-1} = \mathbf{N}\mathbf{N}^{-1}\mathbf{M}\mathbf{N}\mathbf{N}^{-1}$$

One has $\mathbf{M} = \mathbf{N}\mathbf{M}'\mathbf{N}^{-1}$, which can be inserted into Eq. (4.72) to obtain

$$\mathbf{D} = \mathbf{L}^{-1}\mathbf{M}\mathbf{L} = \mathbf{L}^{-1}\mathbf{N}\mathbf{M}'\mathbf{N}^{-1}\mathbf{L} = \mathbf{L}'^{-1}\mathbf{M}'\mathbf{L}' \quad \text{with} \quad \mathbf{L}' = \mathbf{N}^{-1}\mathbf{L} .$$

(to get the last passage try 'inserting' $\mathbf{L}'^{-1}$ on both sides of the expression for $\mathbf{L}'$). So, while the matrix of eigenvectors changes from $\mathbf{L}$ into $\mathbf{N}^{-1}\mathbf{L}$, the diagonal matrix of eigenvalues $\mathbf{D}$ stays the same under the change of basis.

The eigenvalues are related to other invariant quantities, which are generally much easier to evaluate. We have in fact encountered two of them already:

- The determinant det($\mathbf{M}$), whose invariance is easily proved.[3]
- The 'trace' $Tr(\mathbf{M})$, which as we recall is defined as the sum of the elements on the main diagonal. The trace equals the sum of all the eigenvalues of a matrix: $Tr(\mathbf{M}) = \sum_{j=1}^{n} \lambda_j$.

For instance, in the case of matrix $\mathbf{A}$ of our previous example, one has $\det(\mathbf{A}) = -1$ and $Tr(\mathbf{A}) = 0$, which are indeed, respectively, the product and the sum of the two eigenvalues -1 and $+1$.

[3]From Binet's formula for the determinant of a product $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$, one has $\det(\mathbf{L}^{-1}\mathbf{M}\mathbf{L}) = \det(\mathbf{L}^{-1})\det(\mathbf{M})\det(\mathbf{L}) = \frac{1}{\det(\mathbf{L})}\det(\mathbf{L})\det(\mathbf{M}) = \det(\mathbf{M})$

4.10.4 An Application of Diagonalization

We can use diagonalization to compute the powers of a matrix, as long as the latter is diagonazable. Let $\mathbf{M}$ be a matrix that can be diagonalised, *i.e.* such that

$$\mathbf{D} = \mathbf{L}^{-1}\mathbf{M}\mathbf{L} \ ,$$

where $\mathbf{D}$ is a diagonal matrix (with the eigenvalues λ_j of $\mathbf{M}$ on the main diagonal and zeros elsewhere) and $\mathbf{L}$ is an invertible matrix (whose columns represent the eigenvectors of $\mathbf{M}$). By the usual 'trick' we can get:

$$\mathbf{M} = \mathbf{L}\mathbf{D}\mathbf{L}^{-1} \ ,$$

Then if we wanted to evaluate $\mathbf{M}^2$, and remembering that the matrix product is associative:

$$\mathbf{M}^2 = (\mathbf{L}\mathbf{D}\mathbf{L}^{-1})^2 = (\mathbf{L}\mathbf{D}\mathbf{L}^{-1})(\mathbf{L}\mathbf{D}\mathbf{L}^{-1}) = \mathbf{L}\mathbf{D}(\mathbf{L}^{-1}\mathbf{L})\mathbf{D}\mathbf{L}^{-1} = \mathbf{L}\mathbf{D}^2\mathbf{L}^{-1}$$

Hence for any power:

$$\mathbf{M}^n = \mathbf{L}\mathbf{D}^n\mathbf{L}^{-1}$$

4.10.5 Summary of special matrices and diagonalization

Before proceeding, let us recall some of the properties of matrices:

- Given a matrix $\mathbf{M}$, the *transpose* of $\mathbf{M}$ is denoted by $\mathbf{M}^T$ and is obtained by writing its rows as the column of $\mathbf{M}$. In terms of matrix entries one has: $\mathbf{M}^T_{jk} = \mathbf{M}_{kj}$ (swap rows and columns).

- Given a matrix $\mathbf{M}$, the *hermitian conjugate* of $\mathbf{M}$ is denoted by $\mathbf{M}^\dagger$ and is obtained by complex conjugation of the transpose $\mathbf{M}^T$: $\mathbf{M}^\dagger = \mathbf{M}^{T*}$. In terms of matrix entries one has: $\mathbf{M}^\dagger_{jk} = \mathbf{M}^*_{kj}$ (swap rows and columns and complex conjugate).

  We can now list some classes of special matrices (we covered most of them already):

- A square matrix $\mathbf{M}$ is said to be *normal* if and only if $\mathbf{M}^\dagger\mathbf{M} = \mathbf{M}\mathbf{M}^\dagger$.

- A square matrix $\mathbf{H}$ is said to be *hermitian* if and only if $\mathbf{H}^\dagger = \mathbf{H}$.

- A square matrix $\mathbf{S}$ is said to be *real and symmetric* if and only if its entries are real and $\mathbf{S}^T = \mathbf{S}$.

- A square matrix $\mathbf{U}$ is said to be *unitary* if and only if $\mathbf{U}^\dagger\mathbf{U} = \mathbf{U}\mathbf{U}^\dagger = \mathbf{I}$.

- A square matrix $\mathbf{O}$ is said to be *orthogonal* if and only if it has real entries and $\mathbf{O}^T\mathbf{O} = \mathbf{O}\mathbf{O}^T = \mathbf{I}$.

  Let us first clarify the dependencies between the sets of matrices defined above:

- Since for real matrices hermitian conjugation is the same as transposition, *orthogonal matrices are unitary*: $\mathbf{O}^\dagger\mathbf{O} = \mathbf{O}^T\mathbf{O} = \mathbf{I}$.

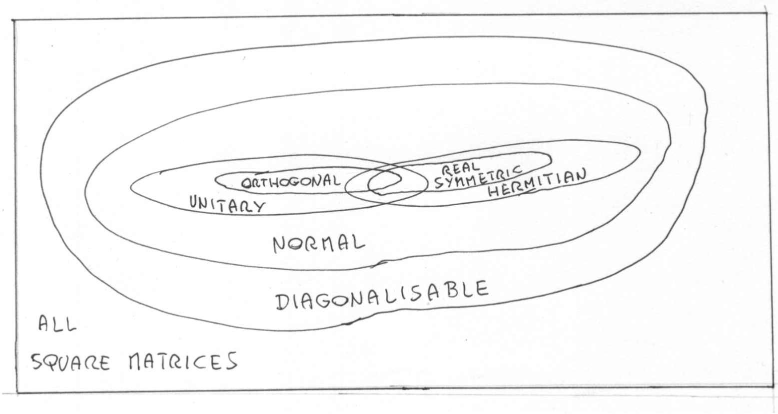


Figure 4.1: Set relationships between classes of special matrices.

- Since for real matrices hermitian conjugation is the same as transposition, *real symmetric matrices are hermitian*: $\mathbf{S}^\dagger = \mathbf{S}^T = \mathbf{S}$.
- *Unitary matrices are normal*: $\mathbf{U}^\dagger\mathbf{U} = \mathbf{I} = \mathbf{U}\mathbf{U}^\dagger$.
- *Hermitian matrices are normal*: $\mathbf{H}^\dagger\mathbf{H} = \mathbf{H}\mathbf{H} = \mathbf{H}\mathbf{H}^\dagger$.
- *Normal matrices are diagonalisable*: unitary, orthogonal, hermitian, and real symmetric matrices are **all** normal and hence diagonalisable

To be *normal*, is, in a sense, the broader sufficient condition known for diagonalisability. Notice that this condition is only sufficient and *not* necessary: there exist diagonalisable matrices which are not normal. In general, given a non-normal matrix, the only way to know whether it is diagonalisable or not is by trying to find its eigenvectors and verifying whether they form a basis of the vector space or not. Notice moreover that, since unitary, orthogonal, hermitian and real symmetric matrices are all normal, and hence diagonalisable, these are other (narrower) sufficient conditions for diagonalisability. The relations between these classes of matrices are depicted in Fig. 4.1. **As an exercsise for you to do at home: take a permutation matrix:**

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

and see how many classes it belongs to.

4.10.6 Eigenvalues and eigenvectors of hermitian matrices

We just saw that any hermitian matrix $\mathbf{H} = \mathbf{H}^\dagger$ is diagonalisable.

Let us then denote by λ_j and λ_k two generic eigenvalues of the hermitian matrix $\mathbf{H}$, with corresponding eigenvectors $\underline{\mathbf{v}}_j$ and $\underline{\mathbf{v}}_k$:

$$\begin{aligned} \mathbf{H}\underline{\mathbf{v}}_j &= \lambda_j \underline{\mathbf{v}}_j \,, && (4.73)\\ \mathbf{H}\underline{\mathbf{v}}_k &= \lambda_k \underline{\mathbf{v}}_k \,. && (4.74)\end{aligned}$$

Let multiply Eq. (4.73) on the left by $\underline{\mathbf{v}}_k^\dagger$ to obtain

$$\underline{\mathbf{v}}_k^\dagger \mathbf{H}\underline{\mathbf{v}}_j = \lambda_j \underline{\mathbf{v}}_k^\dagger \underline{\mathbf{v}}_j \,. \tag{4.75}$$

Because of equation 4.52

$$(\mathbf{A}\,\mathbf{B})^\dagger = \mathbf{B}^\dagger\,\mathbf{A}^\dagger \,.$$

the hermitian conjugate of Eq. (4.74) reads

$$\underline{\mathbf{v}}_k^\dagger \mathbf{H}^\dagger = \underline{\mathbf{v}}_k^\dagger \mathbf{H} = \lambda_k^* \underline{\mathbf{v}}_k^\dagger$$

(where we made use of the property $\mathbf{H} = \mathbf{H}^\dagger$), and can be multiplied on the right by $\underline{\mathbf{v}}_j$ to get

$$\underline{\mathbf{v}}_k^\dagger \mathbf{H}\underline{\mathbf{v}}_j = \lambda_k^* \underline{\mathbf{v}}_k^\dagger \underline{\mathbf{v}}_j \,. \tag{4.76}$$

The left hand sides of Eqs. (4.75) and (4.76) are the same, so that we can equate the right hand sides to obtain

$$(\lambda_j - \lambda_k^*)\underline{\mathbf{v}}_k^\dagger \underline{\mathbf{v}}_j = 0 \,. \tag{4.77}$$

Eq. (4.77) has two consequences:

- If $j = k$, one has $\underline{\mathbf{v}}_k^\dagger \underline{\mathbf{v}}_j = \underline{\mathbf{v}}_j^\dagger \underline{\mathbf{v}}_j = |\underline{\mathbf{v}}_j|^2 \neq 0$ (because null vectors, with all zero entries, are trivial eigenvector of any matrix and are excluded by hypothesis when considering eigenvectors of linear operators), such that one is left with $\lambda_j = \lambda_j^*$: *the eigenvalues of hermitian matrices are always real.*
- If $j \neq k$ and $\lambda_j \neq \lambda_k$, then $\underline{\mathbf{v}}_k^\dagger \underline{\mathbf{v}}_j = 0$: *the eigenvector of hermitian operators associated to different eigenvalues are orthogonal.*

The eigenvectors of a hermitian operator may always be chosen to form an orthonormal basis.

This important result may be rephrased in terms of the transformation that diagonalises $\mathbf{H}$, which we will call $\mathbf{U}$: $\mathbf{U}^{-1}\mathbf{H}\mathbf{U} = \mathbf{D}$. The matrix $\mathbf{U}$ is the matrix with columns equal to the eigenvectors of $\mathbf{H}$: if such eigenvectors are orthonormal, the matrix $\mathbf{U}$ is given by

$$\mathbf{U} = \left(\underline{\mathbf{v}}_1, \ldots \underline{\mathbf{v}}_j, \ldots \underline{\mathbf{v}}_n\right) \,, \quad \text{with} \quad \underline{\mathbf{v}}_j^\dagger \underline{\mathbf{v}}_k = \delta_{jk} \,,$$

Note that $\underline{\mathbf{v}}_j^\dagger \underline{\mathbf{v}}_k = \delta_{jk}$ is the inner product i.e $(\underline{\mathbf{v}}_j, \underline{\mathbf{v}}_k)$

and the inverse can be immediately constructed:

$$\mathbf{U}^{-1} = \begin{pmatrix} \underline{\mathbf{v}}_1^\dagger \\ \vdots \\ \underline{\mathbf{v}}_j^\dagger \\ \vdots \\ \underline{\mathbf{v}}_d^\dagger \end{pmatrix} = \mathbf{U}^\dagger \,.$$

The matrix **U** is hence unitary: *any hermitian matrix is diagonalisable and can be diagonalised by a unitary transformation.*

Also, as we determined above, the eigenvalues of hermitian matrices are real.

Real symmetric matrices being hermitian, their eigenvalues are bound to be real too. Moreover, along the same lines described here, it can be shown that real symmetric matrices can always be diagonalised by orthogonal (that is, essentially, real and unitary) transformations. Orthogonal transformations represent generalised rotations (reducing to the common spatial rotations in dimension 2 and 3): real symmetric matrices can hence always be diagonalised by rotating the basis of the vector space.

4.11 Real quadratic forms

Quadratic forms are polynomials with n variables, all of degree **two**.

Matrices do not only represent linear transformations, but also quadratic combinations of sets of variables. Let $\{x_1, \dots, x_n\}$ and $\{y_1, \dots, y_n\}$be two sets of n variables, assumed for simplicity to be real, and let $\sum_{j,k=1}^{n} q_{jk}x_jy_k$ be any generic quadratic combination of them, with real coefficients q_{jk} (which can of course form a matrix) i.e

$$\sum_{j,k=1}^{n} q_{jk}x_jy_k = q_{11}x_1y_1 + q_{12}x_1y_2 + ... + q_{nn}x_ny_n \tag{4.78}$$

This expression can be written in matrix form:

$$= \underline{\mathbf{x}}^T\mathbf{Q}\underline{\mathbf{y}}$$

setting:

$$\underline{\mathbf{x}} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}, \underline{\mathbf{y}} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}, \mathbf{Q} = \begin{bmatrix} q_{11} & q_{12} & \dots & q_{1n} \\ q_{21} & q_{22} & \dots & q_{2n} \\ . & . & \dots & . \\ q_{n1} & q_{n2} & q_{nn} & \end{bmatrix}$$

Real symmetric matrices hence also represent all the possible *quadratic forms* $Q(\underline{\mathbf{x}}) \in \mathbb{R}$ acting on real vector spaces, which are fed a vector and output a real number. Let us show this via a simple example. Let's consider the form of potential energy:

$$\frac{1}{2}K(x_1 - x_2)^2$$

In matrix form this can be written as:

$$\underline{\mathbf{x}}^T\mathbf{K}\underline{\mathbf{x}}$$

In order to find the elements of the matrix $\mathbf{K}$ note that the main diagonal has the coefficient of each quadratic variable (e.g. q_{11} and q_{22} in this case), while the off diagonal has the coefficient belonging to the product of the two variables. Expand the potential energy to find:

$$\frac{1}{2}K(x_1)^2 + \frac{1}{2}K(x_2)^2 - \frac{1}{2}2x_1x_2$$

The coefficient of the first quadratic variable (x_1) is then $\frac{1}{2}K$ and so it is the coefficient of the second quadratic variable (x_2). Then:

$$q_{12} + q_{21} = -\frac{1}{2}2K$$

and so the off diagonal elements are both $-K/2$.

The applications of quadratic forms are beyond count: for instance, the potential energy of a set of coupled harmonic oscillators (springs) can be represented by a real quadratic form. Notice that because $\mathbf{Q}$ is symmetric, it may always be diagonalised, which is often handy. In the next section, we will solve the dynamics of a system of coupled oscillators: although we will not make explicit use of the potential energy in our solution (simply because we will tackle the equations of motion directly), we will solve the problem by a diagonalisation which is completely analogous to diagonalising the quadratic form for the potential energy of the coupled springs.

4.12 Normal modes of oscillation

To appreciate the usefulness of the algebraic techniques we learned, it is instructive to consider a relevant example where such techniques are applied to solve the dynamics of a classical (non-quantum) system.

Consider three particles of equal mass m joined by springs with elastic constant k and rest length l, and constrained to move on a line (Fig. 4.2). By Newton's and Hook's laws, the equations of motion governing the positions x_1, x_2 and x_3 of the three particles are:

$$
\begin{aligned}
m\ddot{x}_1 &= k(x_2 - x_1 - l)\,, && (4.79)\\
m\ddot{x}_2 &= -k(x_2 - x_1 - l) + k(x_3 - x_2 - l)\,, && (4.80)\\
m\ddot{x}_3 &= -k(x_3 - x_2 - l)\,. && (4.81)
\end{aligned}
$$

These second-order differential equations for the functions $x_1(t)$, $x_2(t)$ and $x_3(t)$ are *coupled*: each solution will hence depend on the other two and is in general not easy to find if the equations are tackled directly in the form given above. However, we will see that the diagonalisation of a matrix will allow us to write down a set of equations equivalent to the above, but much simpler to solve.

Notice that, by defining the vector of positions $\underline{\mathbf{x}} = (x_1, x_2, x_3)^T$ and the vector of derivatives $\underline{\ddot{\mathbf{x}}} = (\ddot{x}_1, \ddot{x}_2, \ddot{x}_3)^T$, the Eqs. (4.79-4.81) can be recast in vector notation as follows:

$$
\underline{\ddot{\mathbf{x}}} = \frac{k}{m} A \underline{\mathbf{x}} + \frac{k}{m} \underline{\mathbf{x}}_0\,, \tag{4.82}
$$

where

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

and

$$
\underline{\mathbf{x}}_0 = \begin{pmatrix} -l \\ 0 \\ l \end{pmatrix}.
$$

This last column vector is made up of all the coefficients independent from the vector $\underline{\mathbf{x}}$ and so first row ('representing' the first equation - eq. 4.80) we have $-l$, second row we have $-l + l = 0$ (see eq. 4.81) and third ro we have l (see independent coefficient in eq. 4.82). The matrix **A** above can be derived by working out the coefficients of each of our variables. It has to be 3×3 because we have 3 equations. Then in the first column we have all the coefficients of x_1, in second column all the coefficioents of x_2 and third column all the coefficients of x_3. The matrix **A** is real and symmetric, and can hence be diagonalised. We leave the diagonalisation procedure to the reader, and just give the result in terms of the eigenvalues λ_1, λ_2 and λ_3 and associated normalised eigenvectors $\underline{\mathbf{v}}_1$, $\underline{\mathbf{v}}_2$ and $\underline{\mathbf{v}}_3$:

$$
\lambda_1 = 0\,, \quad \lambda_2 = -1\,, \quad \lambda_3 = -3\,,
$$

$$
\underline{\mathbf{v}}_1 = \begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix}, \quad \underline{\mathbf{v}}_2 = \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{pmatrix}, \quad \underline{\mathbf{v}}_3 = \begin{pmatrix} 1/\sqrt{6} \\ -2/\sqrt{6} \\ 1/\sqrt{6} \end{pmatrix}.
$$

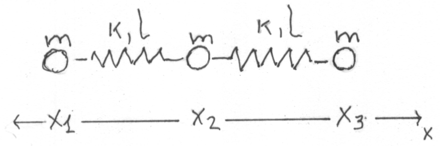


Figure 4.2: Three bodies of mass m coupled by two springs of elastic constant k and rest length l, constrained to move along one dimension (x).

As expected, since the matrix is real and symmetric, the eigenvalues are real and the eigenvectors are real and orthogonal. The transformation $\mathbf{R}$ which diagonalises $\mathbf{A}$ is therefore the orthogonal transformation given by

$$\mathbf{R} = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \end{pmatrix},$$

with $\mathbf{R}^{-1} = \mathbf{R}^T$ (orthogonality). One has then

$$\mathbf{R}^T\mathbf{A}\mathbf{R} = \mathbf{D}\,, \quad \text{with} \quad \mathbf{D} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -3 \end{pmatrix}. \tag{4.83}$$

The previous equation can be multiplied through on the left by $\mathbf{R}$ and on the right by $\mathbf{R}^{\mathrm{T}}$, on both sides, to obtain the equivalent relation

$$\mathbf{A} = \mathbf{R}\mathbf{D}\mathbf{R}^T\,.$$

Let us now insert this expression for $\mathbf{A}$ into Eq. (4.82):

$$\ddot{\underline{\mathbf{x}}} = \frac{k}{m}\mathbf{R}\mathbf{D}\mathbf{R}^T\underline{\mathbf{x}} + \frac{k}{m}\underline{\mathbf{x}}_0\,,$$

which can be multiplied through on the left by $\mathbf{R}^T$ and written as

$$\mathbf{R}^T\ddot{\underline{\mathbf{x}}} = \frac{k}{m}\mathbf{D}\mathbf{R}^T\underline{\mathbf{x}} + \frac{k}{m}\mathbf{R}^T\underline{\mathbf{x}}_0\,. \tag{4.84}$$

Let us now call:

$$\underline{\mathbf{y}} = \mathbf{R}^T\underline{\mathbf{x}}\,, \quad \text{and} \quad \underline{\mathbf{y}}_0 = \mathbf{R}^T\underline{\mathbf{x}}_0 = \begin{pmatrix} 0 \\ -\sqrt{2}l \\ 0 \end{pmatrix} \tag{4.85}$$

(recall the equation for $\underline{\mathbf{x}}_0$), which, if inserted into eq 4.85 yields

$$\ddot{\underline{\mathbf{y}}} = \frac{k}{m}\mathbf{D}\underline{\mathbf{y}} + \frac{k}{m}\underline{\mathbf{y}}_0\,. \tag{4.86}$$

Now, because the matrix **D** is diagonal, the three scalar differential equations contained in (4.86), are much simpler than the original system we set out to solve. In terms of the components y_1, y_2 and y_3 of $\underline{\mathbf{y}}$, one has:

$$\ddot{y}_1 = 0\,, \tag{4.87}$$

$$\ddot{y}_2 = -\frac{k}{m}y_2 + \sqrt{2}\frac{kl}{m}\,, \tag{4.88}$$

$$\ddot{y}_3 = -3\frac{k}{m}y_3\,. \tag{4.89}$$

The threre equations above have been found by looking at explicitly evaluating the $\underline{\mathbf{y}}$ and $\underline{\mathbf{y}}_0$ components in eq. 4.87 i.e for example $\ddot{y}_1 = 0$ because the first column of **D** is all 0s. The first term in the R.H.S of the equation of $\ddot{y}_2$ is determined by multiplying the second column of **D** by k/m and by multiplying the vector $\underline{\mathbf{y}}_0$ by k/m and so on.

Each equation now depends on only one of the three unknown functions: the system of differential equation has been 'decoupled'. The identification of the new variables y_1, y_2 and y_3 which allowed for such a decoupling was made possible by the diagonalisation of the coupling matrix **A**. Notice that, although for simplicity we assumed all the masses, spring constants and rest lengths to be the same, the very same decoupling would have been possible for any system of coupled harmonic oscillators, regardless of such details. The matrix **A** is in fact always symmetric, and hence diagonalisable, for these systems (a consequence, in a sense, of Newton's reaction principle).

The three variables y_1, y_2 and y_3, decoupling the dynamics of the coupled springs, are commonly known as *normal modes of oscillation*. The mode y_1 just represents the centre of mass of the three particles (here the sum of the three positions, as all the masses were assumed equal): in fact, its equation of motion is that of a free particle (the force acting on y_1 is zero), as expected since no external force is acting on the three particles. Modes y_2 and y_3 are instead at times referred to, respectively, as the 'breathing' and the 'Egyptian' mode. Normal modes play a central role in quantum mechanics as well, for instance in the description of quantized electromagnetic fields.