2B21 MATHEMATICAL METHODS IN PHYSICS & ASTRONOMY

Colin Wilkin

Department of Physics & Astronomy, UCL

****

1 Linear Vector Spaces and Matrices

1.1 Revision of determinants

3×3 determinants were introduced briefly in the first year course and I am going to spend the first hour or so reinforcing this material.

1.1.1 Two-by-Two Determinants

Let us start off with a 2×2 determinant, which is an object with two rows and columns, sandwiched between two vertical lines. It just represents an ordinary scalar quantity. The numerical value of a 2×2 determinant is given by

$$\Delta = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} \,. \tag{1.1}$$

for any set of four numbers a_{ij}. The notation with the indices is conventional and we will see why it is chosen when we discuss matrix multiplication. As a simple example, consider

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

Rule 1

Interchanging rows and columns leaves a determinant unchanged.

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

Rule 2

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

Rule 3

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

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

Rule 4

A determinant vanishes if two rows (or columns) are multiples of each other. For example, if $a_{i2} = \alpha\, a_{i1}$ for $i = 1, 2$, then $\Delta = \alpha\, a_{11}a_{21} - \alpha\, a_{11}a_{21} = 0$.

Rule 5

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

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

Rule 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.

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

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

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.

1.1.2 Three-by-Three Determinants

All the above rules will be valid for a general $N \times N$ determinant. A 3×3 determinant can be expanded by the first row as

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

$$= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}\,. \tag{1.2}$$

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.**

Alternatively, we could expand the determinant by the second column say;

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

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

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

The rule of Sarrus

One simple way of remembering how to expand a 3×3 determinant is through the *rule of Sarrus*. Note that this is **not** valid for a 4×4 or higher order determinant. In this prescription, we write down again the first and second columns at the end of the determinant as:

$$\begin{matrix} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{matrix} \tag{1.3}$$

There are now SIX diagonals that lead from the top row to the bottom. The ones pointing to the right get a plus sign, those to the left a minus sign. Thus $a_{12}a_{23}a_{31}$ is positive, whereas $a_{12}a_{21}a_{33}$ is negative. This agrees with the result given in Eq. (1.2).

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.

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

When determinants are evaluated on a computer, the algorithm generally involves subtracting linear combinations 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.

1.1.3 Higher order determinants

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

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

$$= a_{11} \begin{vmatrix} a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \\ a_{42} & a_{43} & a_{44} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} & a_{24} \\ a_{31} & a_{33} & a_{34} \\ a_{41} & a_{43} & a_{44} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} & a_{24} \\ a_{31} & a_{32} & a_{34} \\ a_{41} & a_{42} & a_{44} \end{vmatrix} - a_{14} \begin{vmatrix} a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{vmatrix} \tag{1.4}$$

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

1.1.4 Solving linear simultaneous equations: CRAMER's rule

Consider the simultaneous equations

$$\begin{aligned} a_{11}\,x_1 + a_{12}\,x_2 + a_{13}\,x_3 &= b_1\,, \\ a_{21}\,x_1 + a_{22}\,x_2 + a_{23}\,x_3 &= b_2\,, \\ a_{31}\,x_1 + a_{32}\,x_2 + a_{33}\,x_3 &= b_3 \end{aligned} \tag{1.5}$$

for the unknown x_i. The solution is

$$x_1 = \begin{vmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{vmatrix} \Bigg/ \Delta\,, \quad x_2 = \begin{vmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{vmatrix} \Bigg/ \Delta\,, \quad x_3 = \begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{vmatrix} \Bigg/ \Delta\,, \tag{1.6}$$

where

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}\,. \tag{1.7}$$

We just replace the appropriate column with the column of numbers from the right hand side. This is called Cramer's rule and gives just the same results as matrix inversion — but rather quicker!

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 $x_1 = 3$.

1.2 Three-dimensional Vectors

You were introduced to vectors in an ordinary (real) 3-dimensional Euclidean space in the first year 1B21 course. We are going to start this course by going over some of the results obtained there and then generalise the definitions and results to **complex** spaces with n-dimensions. This will be of importance for the 2B22 Quantum Mechanics course.

We can define a three-dimensional Euclidean space by introducing three mutually orthogonal basis vectors $\hat{\imath}$, $\hat{\jmath}$ and $\hat{k}$, as was done in 1B21. However, it is awkward to generalise this notation to an arbitrary number of dimensions and so we use instead $\underline{\hat{e}}_1 = \hat{\imath}$, $\underline{\hat{e}}_2 = \hat{\jmath}$, and $\underline{\hat{e}}_3 = \hat{k}$. These basis vectors have unit length,

$$\underline{\hat{e}}_1 \cdot \underline{\hat{e}}_1 = \underline{\hat{e}}_2 \cdot \underline{\hat{e}}_2 = \underline{\hat{e}}_3 \cdot \underline{\hat{e}}_3 = 1\,, \tag{1.8}$$

and they are perpendicular to each other;

$$\underline{\hat{e}}_1 \cdot \underline{\hat{e}}_2 = \underline{\hat{e}}_2 \cdot \underline{\hat{e}}_3 = \underline{\hat{e}}_3 \cdot \underline{\hat{e}}_1 = 0\,. \tag{1.9}$$

These properties may be summarised in one equation as

$$\underline{\hat{e}}_i \cdot \underline{\hat{e}}_j = \delta_{ij}\,, \tag{1.10}$$

where the Kronecker delta δ_{ij} is a very useful shorthand notation for

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{1.11}$$

Any vector $\underline{v}$ in this three-dimensional space may be written down in terms of its components along the $\underline{\hat{e}}_i$. I am switching notation here so that I shall underline vectors, rather than putting an arrow on the top. This brings it into line with the notation for matrices. Thus

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

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

$$v_i = \underline{\hat{e}}_i \cdot \underline{v}\;. \tag{1.12}$$

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

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

$$\underline{u}\,\cdot\,\underline{v} = (u_1\,\underline{\hat{e}}_1 + u_2\,\underline{\hat{e}}_2 + u_3\,\underline{\hat{e}}_3)\cdot(v_1\,\underline{\hat{e}}_1 + v_2\,\underline{\hat{e}}_2 + v_3\,\underline{\hat{e}}_3) = u_1\,v_1 + u_2\,v_2 + u_3\,v_3 = \sum_{i=1}^{3} u_i\,v_i\;. \tag{1.13}$$

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

$$v^2 = \underline{v}\cdot\underline{v} = v_1^2 + v_2^2 + v_3^2\;. \tag{1.14}$$

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

$$v = \mid \underline{v} \mid = \sqrt{v^2} = \sqrt{v_1^2 + v_2^2 + v_3^2}\;. \tag{1.15}$$

A unit vector has length one.

A vector is the zero vector if and only if all its components vanish. Thus

$$\underline{v} = \underline{0} \qquad \Longleftrightarrow \qquad (v_1\,, v_2\,, v_3) = (0,0,0)\;. \tag{1.16}$$

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

$$\underline{\hat{e}}_3 \neq \alpha\,\underline{\hat{e}}_1 + \beta\,\underline{\hat{e}}_2\;, \tag{1.17}$$

where α and β are 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

$$\underline{u} = \begin{pmatrix} 1\\ 2\\ 3 \end{pmatrix} \;:\; \underline{v} = \begin{pmatrix} 4\\ 5\\ 6 \end{pmatrix} \;:\; \underline{w} = \begin{pmatrix} 9\\ 12\\ 15 \end{pmatrix}\;. \tag{1.18}$$

Clearly then

$$\underline{w} = \underline{u} + 2\underline{v}\;. \tag{1.19}$$

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

The three-dimensional space S_3 is defined as one where there are three, **BUT NO MORE**, orthonormal linearly independent vectors $\underline{\hat{e}}_i$. 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 $\underline{v}$ in the component form;

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

Note that the $\underline{\hat{e}}_i$ are not unique. We could for example rotate the system through 45° and use these new axes as basis vectors.

So far we haven't done anything which was not done in 1B21, although the notation is a little bit different with the $\underline{\hat{e}}_i$. We now have to generalise all this to an arbitrary number of dimensions and also let the components become complex.

1.3 Linear Vector Space

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

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

$$\begin{aligned}\underline{a}+\underline{b} &= \underline{c}\in S.\\ \underline{c}=\underline{a}+\underline{b} &= \underline{b}+\underline{a}\quad\text{(Commutative law)}\\ (\underline{a}+\underline{b})+\underline{c} &= \underline{a}+(\underline{b}+\underline{c})\quad\text{(Associative law)}\,.\end{aligned}\tag{1.20}$$

2. Multiplication by a scalar (possibly complex)

$$\begin{aligned}\underline{a}\in S &\Longrightarrow \lambda\,\underline{a}\in S\quad(\lambda\text{ a complex number})\,,\\ \lambda\,(\underline{a}+\underline{b}) &= \lambda\,\underline{a}+\lambda\,\underline{b}\,,\\ \lambda\,(\mu\,\underline{a}) &= (\lambda\,\mu)\,\underline{a}\quad(\mu\text{ another complex number})\,.\end{aligned}\tag{1.21}$$

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

$$\underline{a}+\underline{0}=\underline{a}\tag{1.22}$$

   for all vectors $\underline{a}$.

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

$$\underline{a}+(-\underline{a})=\underline{0}\,.\tag{1.23}$$

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

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

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

   By definition, an n-dimensional complex vector space S_n contains just n linearly independent vectors. Hence any vector $\underline{X}$ can be written as a linear combination

$$\underline{X}=c_1\,\underline{X}_1+c_2\,\underline{X}_2\;\cdots\;c_n\,\underline{X}_n\,.\tag{1.24}$$

6. **Basis vectors and components**
   Any set of n linearly independent vectors can be used as a basis for an n-dimensional vector space, which means that the basis is not unique. Once the basis has been chosen, any vector can be written uniquely as a linear combination

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

   Up to this point we have not assumed that the basis vectors are orthogonal. For certain physical problems it is very convenient to work with basis vectors which are not perpendicular — for example, when dealing with crystals with hexagonal symmetry. However, in this course we are only going to work with basis vectors $\hat{\underline{e}}_i$ which are orthogonal and of unit length.

7. **Definition of scalar product**
   The only difference with the results of the 1B21 course is that we want now to let the coefficients c_i in Eq. (1.24) be complex. Such complex spaces are important for the Quantum Mechanics course.

Suppose that we write a vector $\underline{v}$ in terms of its components v_i along the basis vectors $\underline{\hat{e}}_i$, and similarly for another vector $\underline{u}$. Then the scalar product of these two vectors will be defined by

$$(\underline{u}\,,\underline{v}) = \underline{u}\cdot\underline{v} = u_1^*\,v_1 + u_2^*\,v_2 + \cdots + u_n^*\,v_n\ . \tag{1.25}$$

The only difference to the usual form on the right hand side is that we have introduced complex conjugation on all the components u_i since the vectors have to be allowed to be complex. This is the only essential difference with the straightforward real vectors used in 1B21. To stress this difference though, we sometimes use a different notation on the left hand side and denote the scalar product by $(\underline{u}\,,\,\underline{w})$ rather than $\underline{u}\cdot\underline{w}$.

Note that

$$(\underline{v}\,,\,\underline{u}) = v_1^*\,u_1 + v_2^*\,u_2 + \cdots + v_n^*\,u_n = (\underline{u}\,,\,\underline{w})^*\ . \tag{1.26}$$

Thus, in general, the scalar product is a complex scalar.

8. **Consequences of the definition**

   (a) If $\underline{y} = \alpha\,\underline{u} + \beta\,\underline{v}$ then $(\underline{w}\,,\,\underline{y}) = \alpha\,(\underline{w}\,,\,\underline{u}) + \beta(\underline{w}\,,\,\underline{v})$.

   (b) Putting $\underline{u} = \underline{v}$, we see that

   $$u^2 = (\underline{u}\,,\,\underline{u}) = u_1^*\,u_1 + u_2^*\,u_2 + \cdots + u_n^*\,u_n = \mid u_1\mid^2 + \mid u_2\mid^2 + \cdots + \mid u_n\mid^2\ . \tag{1.27}$$

   This is the generalisation of Pythagoras's theorem for complex numbers. Since the $\mid u_i\mid^2$ are real and cannot be negative, then $u^2 \geq 0$. It therefore makes sense to talk about $u = \sqrt{u^2}$ as the real length of a complex vector. In particular, if $u = 1$, we call $\underline{u}$ a unit vector.

   (c) We say that two vectors are orthogonal if $(\underline{u}\,,\,\underline{v}) = 0$.

   (d) Components of a vector are given by the scalar product $v_i = (\underline{\hat{e}}_i\,,\,\underline{v})$.

Representations

Given a set of basis vectors $\underline{\hat{e}}_i$, any vector $\underline{v}$ in an n-dimensional space can be written uniquely in the form $\underline{v} = \sum_{i=1}^{n} v_i\,\underline{\hat{e}}_i$. The set of numbers v_i, $i = 1,\cdots,n$ (the components) are said to represent the vector $\underline{v}$ 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. This whole business of matrices (and much of the third year Quantum Mechanics) is connected with what happens when we change the basis vectors.

1.4 Linear Transformations

Suppose that we perform some operation on a vector $\underline{v}$ which changes it into another vector in the space S_n. We could, for example, rotate the vector. Let us denote the operation by $\hat{A}$ and, instead of tediously saying that $\hat{A}$ acts on $\underline{v}$, write it symbolically as $\hat{A}\,\underline{v}$. By assumption, therefore, $\underline{u} = \hat{A}\,\underline{v}$ is another vector in the same space S_n. To agree with the notation of the 2B22 Quantum Mechanics course, I shall try to put a hat on all the operators.

In 1B21 you were shown that all the manipulations of vectors were simplified by working with components. To investigate this further, we have first to see how the operation $\hat{A}$ changes the basis vectors $\underline{\hat{e}}_1$, $\underline{\hat{e}}_2, \cdots, \underline{\hat{e}}_n$. For the sake of definiteness, let us look at $\underline{\hat{e}}_1$, which has a 1 in the first position and zeros everywhere else:

$$\underline{\hat{e}}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ : \\ 0 \end{pmatrix} \quad (n \text{ terms in the column}). \tag{1.28}$$

Now look at the result of acting upon $\underline{\hat{e}}_1$ with the operator $\hat{A}$. This gives rise to a vector which we shall denote by $\underline{a}_1$ because it started from $\underline{\hat{e}}_1$. Thus

$$\underline{a}_1 = \hat{A}\,\underline{\hat{e}}_1\,. \tag{1.29}$$

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

$$\underline{a}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \vdots \\ a_{n1} \end{pmatrix}. \tag{1.30}$$

To specify the action of $\hat{A}$ completely, we must say how it acts on <u>all</u> the basis vectors $\underline{\hat{e}}_i$;

$$\underline{a}_i = \hat{A}\,\underline{\hat{e}}_i = \begin{pmatrix} a_{1i} \\ a_{2i} \\ a_{3i} \\ \vdots \\ a_{ni} \end{pmatrix}. \tag{1.31}$$

This means that we have to give the n^2 numbers a_{ji}, $(j = 1, 2, \cdots, n; i = 1, 2, \cdots, n)$.

Instead of writing $\underline{a}_i$ explicitly as a column vector, we can use the basis vectors once again to show that

$$\underline{a}_i = a_{1i}\,\underline{\hat{e}}_1 + a_{2i}\,\underline{\hat{e}}_2 + a_{3i}\,\underline{\hat{e}}_3 + \cdots + a_{ni}\,\underline{\hat{e}}_n = \sum_{j=1}^{n} a_{ji}\,\underline{\hat{e}}_j\,. \tag{1.32}$$

We have here used the fact that $\underline{\hat{e}}_i$ has 1 in the i'th position and 0's everywhere else.

Just as in 1B21, once we know how the basis vectors transform, it is (in principle) easy to evaluate the action of $\hat{A}$ on some vector $\underline{v} = \sum_i v_i\,\underline{\hat{e}}_i$. Then

$$\underline{u} = \hat{A}\,\underline{v} = \sum_i (\hat{A}\,\underline{\hat{e}}_i)\,v_i = \sum_{i,j} a_{ji}\,v_i\,\underline{\hat{e}}_j\,. \tag{1.33}$$

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

$$\underline{u} = \sum_j u_j\,\underline{\hat{e}}_j\,, \tag{1.34}$$

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

$$u_j = \sum_{i=1}^{n} a_{ji}\,v_i\,. \tag{1.35}$$

As we shall see in a few minutes, this is just the law for matrix multiplication. Many of you will have seen it for 2×2 matrices from GCSE. For $n\times n$, the sums are just a bit bigger! Some of you will notice that the basis vectors transform with $\sum_j a_{ji}\,\underline{\hat{e}}_j$, whereas the components involve the other index $\sum_i a_{ji}\,v_i$.

The set of numbers a_{ij} represents the abstract operator $\hat{A}$ in the particular basis that we have chosen; these n^2 numbers determine completely the effect of $\hat{A}$ on any arbitrary vector. We say that the vector undergoes a <u>linear transformation</u>. It is convenient to arrange all these numbers into a square array

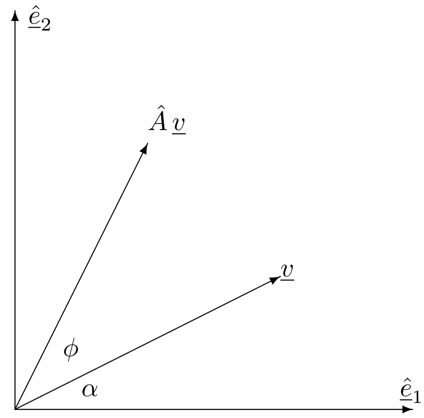
$$\underline{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}, \tag{1.36}$$

and this construct we call a matrix. This one is in fact a square matrix with n rows and n columns.

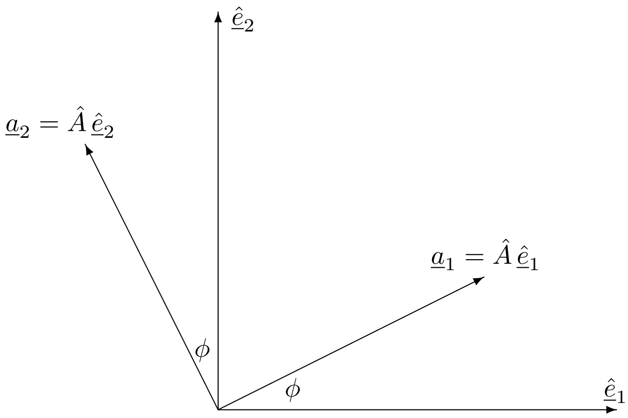
The lecturer in 1B21 insisted that you signify a vector by putting an arrow on the top, underline it, or put a tilde under or over it or write it in bold in order to distinguish it from a scalar. The 2B21 lecturer similarly exhorts the students to write something on the $\underline{A}$ in order to show that it is a matrix. The textbooks tend to use bold face — here we are going just to underline the symbol.

Concrete example # 1

Let $\hat{A}$ be the operator which rotates a vector in two dimensions through an angle ϕ anticlockwise.



We want to find the matrix representation of the operator $\hat{A}$. Do this by looking at what happens to the basis vectors under the rotation.



Using simple trigonometry,

$$\begin{aligned} \underline{a}_1 = \hat{A}\,\hat{\underline{e}}_1 &= \cos\phi\,\hat{\underline{e}}_1 + \sin\phi\,\hat{\underline{e}}_2 \\ &= a_{11}\,\hat{\underline{e}}_1 + a_{21}\,\hat{\underline{e}}_2\,. \end{aligned}$$

Hence $a_{11} = \cos\phi$ and $a_{21} = \sin\phi$.

Similarly,

$$\begin{aligned} \underline{a}_2 = \hat{A}\,\hat{\underline{e}}_2 &= -\sin\phi\,\hat{\underline{e}}_1 + \cos\phi\,\hat{\underline{e}}_2 \\ &= a_{12}\,\hat{\underline{e}}_1 + a_{22}\,\hat{\underline{e}}_2\,, \end{aligned}$$

so that $a_{12} = -\sin\phi$ and $a_{22} = \cos\phi$.

The two-dimensional rotation matrix therefore takes the form

$$\underline{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} . \tag{1.37}$$

We now have to check whether this gives an answer which is consistent with the first picture. Here

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} v\cos\alpha \\ v\sin\alpha \end{pmatrix}$$

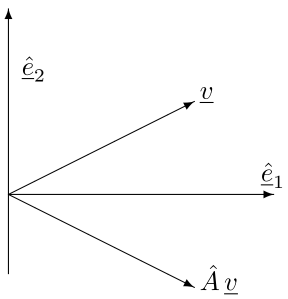
so that

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} v\cos\alpha \\ v\sin\alpha \end{pmatrix} = \begin{pmatrix} v\cos\alpha\cos\phi - v\sin\alpha\sin\phi \\ v\sin\alpha\cos\phi + v\cos\alpha\sin\phi \end{pmatrix} = \begin{pmatrix} v\cos(\alpha+\phi) \\ v\sin(\alpha+\phi) \end{pmatrix} .$$

The latter is exactly what you get from applying trigonometry to the diagram.

Concrete example #2

We want the matrix representation for a reflection in the x-axis.



In this case

$$\begin{aligned} \hat{A}\,\underline{\hat{e}}_1 &= \underline{\hat{e}}_1 \\ \hat{A}\,\underline{\hat{e}}_2 &= -\underline{\hat{e}}_2 \,. \end{aligned}$$

Hence $a_{11} = 1$, $a_{22} = -1$, $a_{21} = a_{12} = 0$ and

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

As a test, see what happens to the vector in the picture:

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

entirely as expected.

1.5 Multiple Transformations; Matrix 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

$$\begin{aligned} \underline{w} &= \hat{B}\,\underline{v} \\ \underline{u} &= \hat{A}\,\underline{w} \,. \\ \underline{u} &= \hat{A}\,\hat{B}\,\underline{v} = \hat{C}\underline{v} \,. \end{aligned} \tag{1.38}$$

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

$$\begin{aligned} w_i &= \sum_j b_{ij}\, v_j \\ u_k &= \sum_i a_{ki}\, w_i \\ &= \sum_{i,j} a_{ki}\, b_{ij} v_j \\ &= \sum_j c_{kj}\, v_j \;. \end{aligned} \tag{1.39}$$

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

$$c_{kj} = \sum_{i=1}^{n} a_{ki}\, b_{ij} \;. \tag{1.40}$$

This is the law for the multiplication of two matrices $\underline{A}$ and $\underline{B}$. The product matrix has the elements c_{kj}. For 2×2 matrices you had the rule at A-level or even at GCSE!

Matrices can be used to represent the action of linear operations, such as reflection and rotation, on vectors. Now that we know how to combine such operations through matrix multiplication, we can build up quite complicated operations. This leads us quite naturally to the study of the properties of matrices in general.

1.6 Properties of Matrices

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

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

A matrix having n rows and m columns is called an $n \times m$ matrix. The above examples are 2×3 and 2×2. A square matrix clearly has $n = m$. The general matrix is written

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots \\ a_{21} & a_{22} & a_{23} & \cdots \\ a_{31} & a_{32} & a_{33} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix} .$$

There is often confusion between the matrices discussed here and the determinants which were introduced in the 1B21 course. The only obvious apparent difference when you look at them is that a matrix is an array surrounded by brackets whereas a determinant has rather got vertical lines. They are, however, very different beasts. The determinant $\mid A \mid$ is a single number (or algebraic expression). A matrix $\underline{A}$ is a whole array of $n \times m$ numbers which represents a transformation.

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

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

Rules

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

2. There exists an $n \times m$ zero-matrix where all the elements are zero.

3. There exists a unit matrix. This is an $n \times n$ square matrix with ones down the diagonal and zeros everywhere else.
$$\underline{I} = \underline{E} = \begin{pmatrix} 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix} .$$
Some books do use $\underline{E}$ for this. In component form
$$I_{ij} = \delta_{ij} \; ,$$
where the Kronecker-delta has been employed.

4. Addition or Subtraction.
The sum of two matrices $\underline{A}$ and $\underline{B}$ can only be defined if they have the same number of n rows and the same number m of columns. If this is the case, then the matrix $\underline{C}$ is also $n \times m$ and has elements
$$c_{ij} = a_{ij} + b_{ij} \; .$$
It follows immediately that $\underline{A}+\underline{B} = \underline{B}+\underline{A}$ (commutative law of addition) and $(\underline{A}+\underline{B})+\underline{C} = \underline{A}+(\underline{B}+\underline{C})$ (associative law).

5. Multiplication by a scalar.
$$\underline{B} = \lambda \underline{A} \quad \Longrightarrow \quad b_{ij} = \lambda a_{ij} \; .$$

6. Matrix multiplication:
$$\underline{C} = \underline{A}\,\underline{B} \quad \Longrightarrow \quad c_{ij} = \sum_{k=1}^{n} a_{ik}\, b_{kj} \; .$$
Note that matrix multiplication can only be defined if the number of columns in $\underline{A}$ is equal to the number of rows in $\underline{B}$. Then if $\underline{A}$ is $m \times n$ and $\underline{B}$ is $n \times p$, then $\underline{C}$ is $m \times p$.

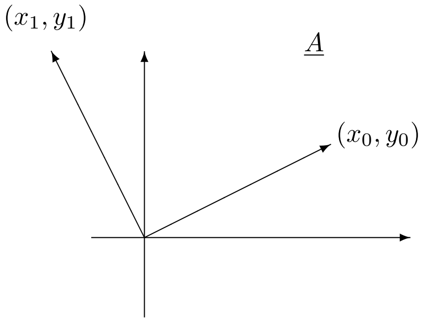
   Note that matrix multiplication is **NOT** commutative; $\underline{A}\,\underline{B} \neq \underline{B}\,\underline{A}$. One of the multiplications might not even be defined! If $\underline{A}$ is $m \times n$ and $\underline{B}$ is $n \times m$, then $\underline{A}\,\underline{B}$ is $m \times m$ and $\underline{B}\,\underline{A}$ is $n \times n$.

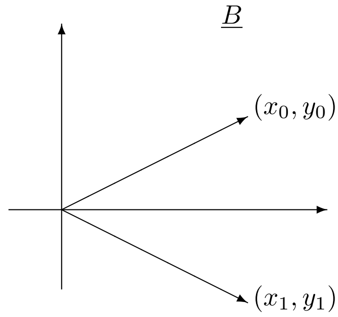
   Matrices do not commute because they are constructed to represent linear operations and, in general, such operations do not commute. It can matter in which order you do certain operations.

   On the other hand,
$$\begin{aligned} \underline{A}\,(\underline{B}\,\underline{C}) &= (\underline{A}\,\underline{B})\,\underline{C} \\ \underline{A}\,(\underline{B}+\underline{C}) &= \underline{A}\,\underline{B} + \underline{A}\,\underline{C} \; . \end{aligned}$$
I shall assume that you are all familiar with the actual multiplication process in practice. If you are not, then you have been warned!

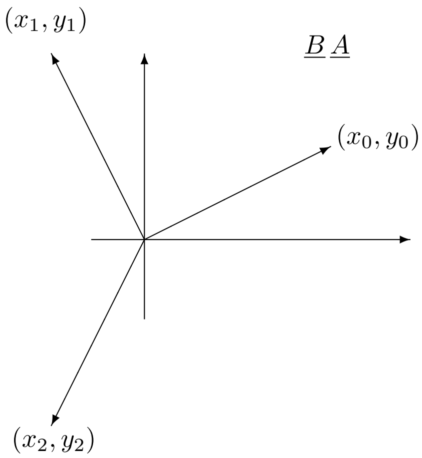
Example #1

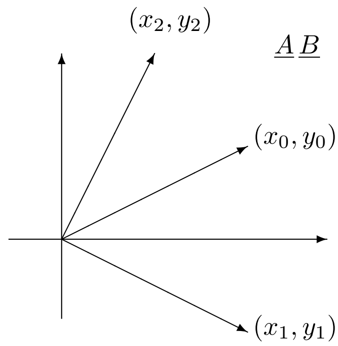
Let $\underline{A}$ represent a rotation of 90° around the z-axis and $\underline{B}$ a reflection in the x-axis.





For the combination $\underline{B}\,\underline{A}$, we first act with $\underline{A}$ and then $\underline{B}$. In the case of $\underline{A}\,\underline{B}$ it is the other way around and this leads to a different result, as shown in the picture.





Clearly the end point (x_2, y_2) is very different in the two cases so that the operations corresponding to $\underline{A}$ and $\underline{B}$ obviously don't commute. We now want to show exactly the same results using matrix manipulation, in order to illustrate the power of matrix multiplication.

We have already constructed 2×2 matrix representing the two-dimensional rotation through angle ϕ.

$$\underline{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{for } \phi = 90^\circ \, .$$

Similarly, for the reflection in the x-axis,

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

Hence

$$\begin{aligned} \underline{A}\,\underline{B} &= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ \underline{B}\,\underline{A} &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \, , \end{aligned}$$

so that in the $\underline{A}\,\underline{B}$ case $x_2 = y_0$ and $y_2 = x_0$. The x and y coordinates are simply interchanged. In the other case both x_2 and y_2 get an extra minus sign. This is exactly what we see in the picture.

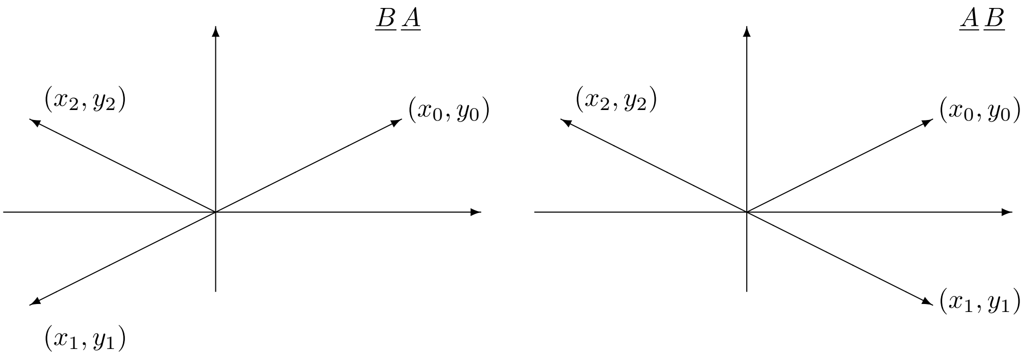
Example #2

It may of course happen that two operators commute, as for example when one represents a rotation through 180° and the other a reflection in the x-axis. Then

$$\underline{A} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \underline{B} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \underline{A}\,\underline{B} = \underline{B}\,\underline{A} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The combined operation describes a reflection in the y-axis.

Geometrically these operations correspond to



You should note that the final point $(x_2, y_2) = (-x_0, y_0)$ is the same in both diagrams, but the intermediate point is not.

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.

$$|\, \underline{A}\,\underline{B} \,| = |\, \underline{A} \,| \times |\, \underline{B} \,| \ . \tag{1.41}$$

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

One consequence of this is that, although $\underline{A}\,\underline{B} \neq \underline{B}\,\underline{A}$, their determinants are equal. In the first example that I gave of matrix multiplication, we see that $|\, \underline{A}\,\underline{B} \,| =$
$|\, \underline{B}\,\underline{A} \,| = -1$. This result for the determinant of products will prove very useful later.

1.7 Special Matrices

Multiplication by the unit matrix

Let $\underline{A}$ be an $n \times n$ matrix and $\underline{I}$ the $n \times n$ unit matrix. Then

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

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

$$\underline{A}\,\underline{I} = \underline{A} \,. \tag{1.42}$$

Similarly

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

and

$$\underline{I}\,\underline{A} = \underline{A}\,. \tag{1.43}$$

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

Diagonal matrices

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

$$\underline{A} = \begin{pmatrix} a_1 & 0 & 0 & \cdots \\ 0 & a_2 & 0 & \cdots \\ 0 & 0 & a_3 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}\ .$$

Thus

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

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

$$(\underline{A}\,\underline{B})_{ij} = \sum_k A_{ik}\,B_{kj} = \sum_k a_i\,\delta_{ik}\,\delta_{kj}\,b_k = (a_i\,b_i)\,\delta_{ij}\ .$$

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

Transposing matrices

The transposed matrix $\underline{A}^T$ is just the original matrix $\underline{A}$ with its rows and columns interchanged. Hence

$$(\underline{A}^T)_{ij} = (\underline{A})_{ji}\ . \tag{1.44}$$

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

Consequences

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

b) If $\underline{A}^T = \underline{A}$, we call $\underline{A}$ symmetric.
If $\underline{A}^T = -\underline{A}$, we call $\underline{A}$ antisymmetric.

c) There is a trick when transposing a product of matrices. To see this, look at $\underline{C} = \underline{A}\,\underline{B}$, which has elements

$$c_{ij} = \sum_k a_{ik}\,b_{kj}\ .$$

Now

$$(\underline{C}^T)_{ji} = c_{ij} = \sum_k a_{ik}\,b_{kj} = \sum_k (\underline{A}^T)_{ki}\,(\underline{B}^T)_{jk} = \sum_k (\underline{B}^T)_{jk}\,(\underline{A}^T)_{ki} = (\underline{B}^T\,\underline{A}^T)_{ji}\ .$$

Hence

$$(\underline{A}\,\underline{B})^T = \underline{B}^T\,\underline{A}^T\ . \tag{1.45}$$

When you transpose a product of matrices, you must reverse the order of the multiplication. This is true no matter how many terms there are;

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

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

d) If $\underline{A}^T\,\underline{A} = \underline{I}$, we say that $\underline{A}$ is an orthogonal matrix. You should check that the two-dimensional rotation matrix

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

is orthogonal. For this, all you really need is $\cos^2\phi + \sin^2\phi = 1$. The matrix $\underline{A}$ rotates the system through an angle ϕ, while the transpose matrix $\underline{A}^T$ rotates it back through an angle $-\phi$. Because of this, orthogonal matrices are of great practical use in different branches of Physics.

Taking the determinant of the defining equation, and using the determinant of a product rule, we find that

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

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

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

so that $|\,\underline{A}\,| = \pm 1$.

e) Suppose $\underline{A}$ and $\underline{B}$ are orthogonal matrices. Then their product $\underline{C} = \underline{A}\,\underline{B}$ is also orthogonal.

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

The physical meaning of this is that, since the matrix for the rotation about the x-axis is orthogonal and so is the rotation about the y-axis, then the matrix for a rotation about the y-axis followed by one about the x-axis is also orthogonal.

Complex conjugation

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

$$(\underline{A}^*)_{ij} = a^*_{ij}\,. \tag{1.46}$$

For example

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

If $\underline{A} = \underline{A}^*$ we say that the matrix is real.

Hermitian conjugation

This is just a combination of complex conjugation and transposition and it is probably more important than either – especially in Quantum Mechanics. It is sometimes called the **Hermitian adjoint** and denoted by a dagger (†).

$$\underline{A}^\dagger = (\underline{A}^T)^* = (\underline{A}^*)^T\,. \tag{1.47}$$

Thus $(\underline{A}^\dagger)^\dagger = \underline{A}$.

For example

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

If $\underline{A}^\dagger = \underline{A}$, we call $\underline{A}$ Hermitian.
If $\underline{A}^\dagger = -\underline{A}$, we call $\underline{A}$ antiHermitian.

Clearly all real symmetric matrices are Hermitian, but there are also other possibilities. For example,

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

is Hermitian.

The rule for Hermitian conjugates of products is exactly the same as for transpositions, so that

$$(\underline{A}\,\underline{B})^\dagger = \underline{B}^\dagger\,\underline{A}^\dagger\,. \tag{1.48}$$

Unitary Matrices

A matrix $\underline{U}$ is unitary if

$$\underline{U}^\dagger\,\underline{U} = \underline{I}\,. \tag{1.49}$$

At the risk of being very repetitive, let me stress that unitary matrices are very important in Quantum Mechanics!

Just as we did for orthogonal matrices, we can find out something about the determinant of $\underline{U}$ by using the determinant of a matrix product rule.

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

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

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

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

1.8 Matrix Inversion

Explicit 2×2 evaluation

We want now to define the inverse of a <u>square</u> matrix $\underline{A}$ and obtain a simple way of evaluating it. The inverse, $\underline{B} = \underline{A}^{-1}$, is defined to be that matrix which, when multiplied by $\underline{A}$, gives the unit matrix;

$$\underline{B}\,\underline{A} = \underline{I}\,.$$

Consider the following concrete example where

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

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

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

and this gives

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

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

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

Rule for 2×2 matrices

We clearly need some automated way of evaluating inverse matrices so that the lecturer can ask questions in the examination paper! Let me try to motivate the result with this particular numerical example. Then we shall generalise and only justify it afterwards. From this example

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

You will notice that inside the bracket, all the coefficients are exchanged across the diagonal between $\underline{A}$ and $\underline{A}^{-1}$. 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

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

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

We have to show that this simple observation is true for the inverse of any 2×2 matrix. Consider a general

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

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

$$(\underline{A}^{-1})^T = \frac{1}{(\alpha\delta - \beta\gamma)} \begin{pmatrix} \delta & -\beta \\ -\gamma & \alpha \end{pmatrix} \quad \text{and} \quad \underline{A}^{-1} = \frac{1}{(\alpha\delta - \beta\gamma)} \begin{pmatrix} \delta & -\gamma \\ -\beta & \alpha \end{pmatrix} .$$

It is left as a simple exercise for the student to verify that the $\underline{A}^{-1}$ defined in this way does indeed satisfy $\underline{A}^{-1}\,\underline{A} = \underline{I}$.

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

Cofactors and minors

In the first lecture I asserted that a 3×3 determinant can be expanded by the first row (Laplace's rule) as

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

The 2×2 sub-determinants are obtained by striking out the rows and columns containing respectively a_{11}, a_{12} and a_{13}. We call these sub-determinants 2×2 **minors** of the determinant Δ.

Define the 2×2 minor obtained by striking out the i'th row and j'th column to be M_{ij}. From the examples given above, it is fairly clear that

$$\Delta = \sum_j a_{ij}\, M_{ij}\, (-1)^{i+j} = \sum_i a_{ij}\, M_{ij}\, (-1)^{i+j} \; . \tag{1.50}$$

The first form corresponds to expanding by row-i, the second to column-j. After summing over j, the answer does not depend upon the value of i, *i.e.* on which row has been used for the expansion.

One trouble about this formula is the irritating $(-1)^{i+j}$ factor which always arises in expanding determinants. One way of sweeping it under the carpet is to define the **cofactor** matrix $\underline{C} = [C_{ij}]$ with this explicit factor contained therein:

$$C_{ij} = (-1)^{i+j}\, M_{ij} \; , \tag{1.51}$$

so that

$$\Delta = \sum_{i \text{ or } j} a_{ij}\, C_{ij} \; . \tag{1.52}$$

This merely puts the minus sign problem somewhere else!

If $\underline{A}$ is a 3×3 matrix, then so is $\underline{C}$. We define the **adjoint** matrix to be the transpose of $\underline{C}$, which means that the indices i and j are switched around:

$$[\underline{A}^{\text{adj}}]_{ij} = C_{ji} \; . \tag{1.53}$$

Theorem

For any square matrix,

$$\underline{A}^{-1} = \underline{A}^{\text{adj}} / \mid A \mid \; . \tag{1.54}$$

This clearly agrees with our experience in the case of a 2×2 matrix. For a 3×3 matrix one can write down the most general form, carry out the operations outlined above, to show explicitly that $\underline{A}^{-1}\,\underline{A} = \underline{I}$. The formula in Eq. (1.54) is valid for any size matrix, but you won't be asked to work out anything bigger than 3×3 in this course. All that I will do now is give you an explicit example to show how to carry out these operations in practice.

Example Find the inverse of

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

The matrix of minors is

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

The cofactor matrix changes a few signs to give

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

The adjoint matrix involves changing rows and columns:

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

Now

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

Hence

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

You can check that this is right by doing the explicit $\underline{A}^{-1}\underline{A}$ multiplication.

Note that if $\mid A \mid = 0$, we say that the determinant is **singular** and $\underline{A}^{\mathrm{adj}}$ does not exist. [It has some infinite elements.]

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

Properties of the inverse matrix

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

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

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

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

Hence

$$(\underline{A}\,\underline{B})^{-1} = \underline{B}^{-1}\,\underline{A}^{-1}\,. \tag{1.55}$$

Just as in transposing products, one reverses the order before inverting each matrix.

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

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

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

g) **Matrix division**

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

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

1.9 Solution of Linear Simultaneous Equations

In the 1B21 course you were shown how to solve simultaneous equations of the 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 the unknown x_i as the ratio of two determinants. The result was proved in the 2×2 case and I want here to give an indication of a more general proof.

The equations 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

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

One can write down the solution immediately by multiplying both sides by $\underline{A}^{-1}$ to leave

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

All that remains is to evaluate the result!

Using the previous expression for the inverse matrix,

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

Assuming for the moment that the determinant does not vanish, this leads to Cramer's rule discussed in the first lecture. $\sum_i (\underline{A}^{\mathrm{adj}})_{ji}\,b_i$ is the determinant obtained by replacing the j'th column of $\underline{A}$ by the column vector $\underline{b}$. There are many special cases of this formula; I draw your attention to only two:

a) Suppose that $\mid A \mid = 0$. In such a case the matrix $\underline{A}$ is singular and we cannot define the inverse matrix. Provided that the equations are mutually consistent, this means that (at least) one of the equations is not linearly independent of the others. We do not have n equations for n unknowns but rather only $n-1$ equations. We must therefore limit our ambitions and 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.

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

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

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 $\mid A \mid = 0$.

1.10 Eigenvalues and Eigenvectors

Let $\underline{A}$ be an $n \times n$ square matrix and $\underline{X}$ an $n \times 1$ column vector such that

$$\underline{A}\,\underline{X} = \lambda\,\underline{X} = \lambda\,\underline{I}\,\underline{X}\,, \tag{1.56}$$

where λ is some scalar number. In such a case we say that λ is an **eigenvalue** of the matrix $\underline{A}$ and that $\underline{X}$ is the corresponding eigenvector. Half of Quantum Mechanics seems to be devoted to searching for eigenvalues!

To attack the problem, rearrange Eq. (1.56) as

$$(\underline{A} - \lambda\,\underline{I})\,\underline{X} = \underline{0}\,. \tag{1.57}$$

This is a set of n homogeneous linear equations which only has an interesting solution provided that

$$| \underline{A} - \lambda \underline{I} | = 0 \, . \tag{1.58}$$

Writing this out explicitly,

$$\begin{vmatrix} (a_{11}-\lambda) & a_{12} & a_{13} & \cdots \\ a_{21} & (a_{22}-\lambda) & a_{23} & \cdots \\ a_{31} & a_{32} & (a_{33}-\lambda) & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{vmatrix} = 0 \, . \tag{1.59}$$

This is an equation for the required eigenvalues λ. It is a polynomial of degree n in λ and hence there must be n solutions. These roots are not necessarily real (even if all the a_{ij} are real) and some of the roots may be equal to others. This polynomial equation is called the characteristic equation of the eigenvalue problem.

Let us label the roots as

$$\lambda_1, \, \lambda_2, \, \cdots , \lambda_n \, .$$

If two of the eigenvalues are equal, then we say that the eigenvalue has a two-fold degeneracy, or that it is doubly-degenerate. Similarly, if there are r equal roots then this corresponds to an r-fold degeneracy.

Suppose that we have solved the characteristic equation to get the eigenvalues λ_i. We then have to solve

$$(\underline{A} - \lambda_i \, \underline{I}) \, \underline{X}_i = \underline{0}$$

to find the corresponding eigenvector $\underline{x}_i$. There are therefore n eigenvectors $\underline{X}_i$ which can be written in terms of components as

$$\underline{X}_i = \begin{pmatrix} x_{1i} \\ x_{2i} \\ : \\ : \\ x_{ni} \end{pmatrix} .$$

Example Find the eigenvalues and eigenvectors of the matrix

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

The characteristic equation is

$$| \underline{A} - \lambda \underline{I} | = \begin{vmatrix} (3-\lambda) & 2 \\ 1 & (4-\lambda) \end{vmatrix} = (3-\lambda)(4-\lambda) - 2 = 0 \, ,$$

giving solutions $\lambda_1 = 5$ and $\lambda_2 = 2$.

In the case of $\lambda_1 = 5$, we have

$$\begin{pmatrix} (3-\lambda) & 2 \\ 1 & (4-\lambda) \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{21} \end{pmatrix} == \begin{pmatrix} -2 & 2 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{21} \end{pmatrix} = \underline{0} \, .$$

This gives the two equations

$$\begin{aligned} -2x_{11} + 2x_{21} &= 0, \\ x_{11} - x_{21} &= 0. \end{aligned}$$

Of course these equations are not linearly independent and so the solution must involve some arbitrary constant c_1;

$$x_{11} = x_{21} = c_1 \, .$$

Similarly, for $\lambda_2 = 2$, we get

$$x_{12} = c_2, \quad x_{22} = -\tfrac{1}{2} c_2 \, .$$

In summary

$$\lambda_1 = 5 \qquad \Longrightarrow \qquad \underline{X}_1 = c_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} ,$$

$$\lambda_2 = 2 \qquad \Longrightarrow \qquad \underline{X}_2 = c_2 \begin{pmatrix} 1 \\ -\frac{1}{2} \end{pmatrix} .$$

The c_i are arbitrary constants but for many purposes it is convenient to choose their sizes such that the $\underline{X}_i$ are unit vectors. You remember that we defined the scalar product of two (possibly complex) vectors through

$$(\underline{a}\,,\underline{b}) = a_1^*\,b_1 + a_2^*\,b_2 + \cdots + a_n^*\,b_n = \underline{a}^\dagger\,\underline{b}\,.$$

In order that the lengths of the eigenvectors be unity, we need

$$\underline{X}_1^\dagger\,\underline{x}_1 = \underline{X}_2^\dagger\,\underline{x}_2 = 1\,.$$

The first of these equations means that

$$(c_1^* \;\; c_1^*) \begin{pmatrix} c_1 \\ c_1 \end{pmatrix} = 2 \mid c_1 \mid^2 = 1\,.$$

The phase of c_1 is really completely arbitrary — this equation only fixes the magnitude of a potentially complex number c_1. Taking it to be real and positive, then $c_1 = 1/\sqrt{2}$.

The second equation results in

$$(c_2^* \quad -\tfrac{1}{2}c_2^*) \begin{pmatrix} c_2 \\ -\frac{1}{2}c_2 \end{pmatrix} = \tfrac{5}{4} \mid c_2 \mid^2 = 1\,,$$

and so $c_2 = 2/\sqrt{5}$.

The final answer is, therefore,

$$\lambda_1 = 5 \qquad \Longrightarrow \qquad \underline{X}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} ,$$

$$\lambda_2 = 2 \qquad \Longrightarrow \qquad \underline{X}_2 = \frac{2}{\sqrt{5}} \begin{pmatrix} 1 \\ -\frac{1}{2} \end{pmatrix} .$$

1.11 Eigenvalues of Unitary Matrices

By definition, any unitary matrix $\underline{U}$ satisfies

$$\underline{U}^\dagger\,\underline{U} = \underline{U}\,\underline{U}^\dagger = \underline{I}\,.$$

To find the eigenvalues, we have to solve the equation

$$\underline{U}\,\underline{X} = \lambda\,\underline{I}\,\underline{X}\,. \tag{1.60}$$

Now take the Hermitian conjugate of Eq. (1.60),

$$\begin{aligned} (\underline{U}\,\underline{X})^\dagger &= (\lambda\,\underline{I}\,\underline{X})^\dagger \\ \underline{X}^\dagger\,\underline{U}^\dagger &= \lambda^*\,\underline{X}^\dagger\,\underline{I}\,. \end{aligned} \tag{1.61}$$

Note here the trick that the Hermitian conjugate interchanges the order in a product.

Multiply the left hand sides of Eqs. (1.60, 1.61) together and also the right hand sides:

$$\underline{X}^\dagger\,\underline{U}^\dagger\,\underline{U}\,\underline{X} = \lambda^*\lambda\,\underline{X}^\dagger\,\underline{X}\,. \tag{1.62}$$

But $\underline{U}^\dagger\,\underline{U} = \underline{I}$, and $\underline{X}^\dagger\,\underline{X} = X^2$. Hence

$$X^2 = \mid \lambda \mid^2\; X^2\,. \tag{1.63}$$

Since $X^2 \neq 0$, we can divide out by this factor to find that $\mid \lambda \mid = 1$, *i.e.* all the eigenvalues are (possibly complex) numbers of unit modulus;

$$\lambda = e^{i\phi} \quad \text{with } \phi \text{ real.} \tag{1.64}$$

1.12 Eigenvalues of Hermitian Matrices

A Hermitian matrix is one for which $\underline{H} = \underline{H}^\dagger$. Consider two eigenvector equations corresponding to different eigenvalues λ_i and λ_j;

$$\begin{aligned} \underline{H}\,\underline{X}_i &= \lambda_i\,\underline{X}_i\,, && (1.65)\\ \underline{H}\,\underline{X}_j &= \lambda_j\,\underline{X}_j\,. && (1.66)\end{aligned}$$

Take the Hermitian conjugate of Eq. (1.65);

$$\begin{aligned} (\underline{H}\,\underline{X}_i)^\dagger &= (\lambda_i\,\underline{X}_i)^\dagger\,,\\ \underline{X}_i^\dagger\,\underline{H}^\dagger = \underline{X}_i^\dagger\,\underline{H} &= \lambda_i^*\,\underline{X}_i^\dagger\,. && (1.67)\end{aligned}$$

Now multiply Eq. (1.67) on the right by $\underline{X}_j$ to get

$$\underline{X}_i^\dagger\,\underline{H}\,\underline{X}_j = \lambda_i^*\,\underline{X}_i^\dagger\,\underline{X}_j\,. \tag{1.68}$$

Go back to Eq.(1.66) and multiply it on the left by $\underline{X}_i^\dagger$;

$$\underline{X}_i^\dagger\,\underline{H}\,\underline{X}_j = \lambda_j\,\underline{X}_i^\dagger\,\underline{X}_j\,. \tag{1.69}$$

The left hand sides of Eqs. (1.68) and (1.69) are identical and so, for all i and j, the right hand sides have to be as well;

$$(\lambda_i^* - \lambda_j)\,\underline{X}_i^\dagger\,\underline{X}_j = 0\,. \tag{1.70}$$

Take first $i = j$ **:**

$$(\lambda_i^* - \lambda_i)\,\underline{X}_i^\dagger\,\underline{X}_i = (\lambda_i^* - \lambda_i)\,X_i^2 = 0\,. \tag{1.71}$$

But since all of the X_i^2 are non-zero, we see that

$$\lambda_i^* - \lambda_i = 0\,, \tag{1.72}$$

which means that all the eigenvalues are **real**.

Now take $i \neq j$:

If the eigenvalues are non-degenerate, *i.e.* $i \neq j \implies \lambda_i \neq \lambda_j$, then

$$\underline{X}_i^\dagger\,\underline{X}_j = 0\,, \tag{1.73}$$

which means that the corresponding eigenvectors are orthogonal.

If two of the eigenvalues are the same, *i.e.* a particular root is doubly degenerate, then the proof fails because one can then have $\lambda_i - \lambda_j = 0$ for $i \neq j$. Nevertheless, it is still possible to choose linear combinations of the corresponding eigenvectors to make all the eigenvectors orthogonal.

Orthogonal basis set

Suppose that we normalise the eigenvectors of a Hermitian matrix as we did in the 2×2 example. Then the $\underline{\hat{X}_i}$ are unit orthogonal vectors, which we can take as basis vectors for this n-dimensional space. As a consequence, any vector can be written as

$$\underline{V} = \sum_i V_i\,\underline{\hat{X}_i}\,.$$

This simple result will be used extensively in one form or another in the second and third year Quantum Mechanics course. The Hamiltonian (Energy) operator is Hermitian and so its eigenfunctions are orthogonal. Any wave function can be expanded in terms of these eigenfunctions.

1.13 Useful Rules for Eigenvalues

If we group all the different $\underline{\hat{X}_i}$ column vectors together in a single $n \times n$ matrix $\underline{X}$, then the eigenvector equation can then be written in the form

$$\underline{A}\,\underline{X} = \underline{X}\,\underline{\Lambda}\,, \tag{1.74}$$

where $\underline{\Lambda}$ is the diagonal matrix of eigenvalues

$$\underline{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} . \tag{1.75}$$

Take the determinant of Eq. (1.74) and use the determinant of a product rule to show that

$$\mid \underline{A} \mid \mid \underline{X} \mid = \mid \underline{\Lambda} \mid \mid \underline{X} \mid \ .$$

Hence

$$\mid \underline{\Lambda} \mid = \mid \underline{A} \mid \ .$$

We can use this to check that we got the right answer for the 2×2 matrix $\begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix}$. This has determinant $\Delta = 10$ which is indeed equal to the product of the eigenvalues 5 and 2. Great!

Another valuable check is through the trace of a matrix. This quantity is defined as being the sum of the diagonal elements;

$$tr\{\underline{A}\} = \sum_i a_{ii}\,. \tag{1.76}$$

For example, the simple 2×2 matrix given above has $tr\{\underline{A}\} = 7$, which is equal to the sum of the eigenvalues 2 and 5. Is this just luck or is it much deeper?

Rewrite Eq. (1.74) by taking $\underline{X}$ over to the other side as an inverse matrix.

$$\underline{A} = \underline{X}\,\underline{\Lambda}\,\underline{X}^{-1}\,.$$

Now take the trace. Writing it out explicitly with indices,

$$tr\{\underline{A}\} = \sum_i a_{ii} = \sum_{i,j,k} (\underline{X})_{ij}\,(\underline{\Lambda})_{jk}\,(\underline{X}^{-1})_{ki}$$

$$= \sum_{i,j,k} (\underline{\Lambda})_{jk}\,(\underline{X}^{-1})_{ki}\,(\underline{X})_{ij} = tr\{\underline{\Lambda}\,\underline{X}^{-1}\,\underline{X}\} = tr\{\underline{\Lambda}\} = \sum_i \lambda_i\,.$$

The trace of a matrix is equal to the sum of the eigenvalues.

A third useful result is that if the original matrix $\underline{A}$ is Hermitian, then $\underline{X}$ is <u>unitary</u> because

$$\underline{X}_i^\dagger\,\underline{X}_j = \delta_{ij}\,.$$

1.14 Real Quadratic Forms

The general real quadratic form may be written as

$$F = \underline{X}^T\,\underline{A}\,\underline{X} = \sum_{i,j} a_{ij}\,x_i\,x_j\,. \tag{1.77}$$

We can simplify the problem a bit by taking the matrix $\underline{A}$ to be symmetric, *i.e.* $a_{ij} = a_{ji}$. The coefficients can then be read off by inspection. For example, if (Boas, p.422),

$$F = x^2 + 6xy - 2y^2 - 2yz + z^2\,,$$

then $a_{11} = 1$ is the coefficient of the x^2 term. Similarly, $a_{12} = a_{21} = 3$ is half the coefficient of the xy term. The coefficient is shared between two equal elements of the matrix.

We now want to rotate the coordinate system

$$\underline{X} = \underline{R}\,\underline{Y} \tag{1.78}$$

such that the quadratic form has no cross terms of the kind y_1y_2. Thus

$$F = \underline{Y}^T\,\underline{R}^T\,\underline{A}\,\underline{R}\,\underline{Y} = \underline{Y}^T\,\underline{D}\,\underline{Y}\,, \tag{1.79}$$

where $\underline{D}$ is a <u>diagonal</u> matrix.

Since we are interested in rotating the axes, the matrix $\underline{R}$ is orthogonal, $\underline{R}^T\,\underline{R} = \underline{I}$. From Eq. (1.79), we see that we have to find an $\underline{R}$ such that

$$\underline{R}^T\,\underline{A}\,\underline{R} = \underline{D}\,. \tag{1.80}$$

But we have, in principle, already solved this problem. Going back to the notes, we see that $\underline{D}$ is the diagonal matrix of eigenvalues $\underline{\Lambda}$, and $\underline{R}$ is the matrix of eigenvectors.

Example

Diagonalise the quadratic form

$$F = 5x^2 - 4xy + 2y^2\,.$$

First write the form in terms of a matrix

$$F = (x\,,\,y)\begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

Next find the eigenvalues, which means solving the quadratic equation

$$\begin{vmatrix} (5-\lambda) & -2 \\ -2 & (2-\lambda) \end{vmatrix} = \lambda^2 - 7\lambda + 6 = 0\,.$$

There are two solutions, $\lambda_1 = 6$ and $\lambda_2 = 1$. [You could check these by showing that the trace of the matrix equals 7 and its determinant equals 6.]

In the case of $\lambda_1 = 6$, the eigenvector equation is

$$\begin{pmatrix} -1 & -2 \\ -2 & -4 \end{pmatrix}\begin{pmatrix} r_{11} \\ r_{21} \end{pmatrix} = 0\,,$$

which gives $r_{11} = -2r_{21}$. In order that the eigenvector be normalised, we get

$$\underline{r}_1 = \frac{1}{\sqrt{5}}\begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

Similarly for $\lambda_1 = 1$, the eigenvector equation is

$$\begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} r_{12} \\ r_{22} \end{pmatrix} = 0\,,$$

which gives $r_{22} = 2r_{12}$. The normalised eigenvector is

$$\underline{r}_2 = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix},$$

and the whole rotation matrix

$$\underline{R} = \begin{pmatrix} -2/\sqrt{5} & 1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{pmatrix}.$$

Thus

$$F = 6x'^2 + y'^2\,,$$

where

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \underline{R}^T \begin{pmatrix} x \\ y \end{pmatrix} ,$$

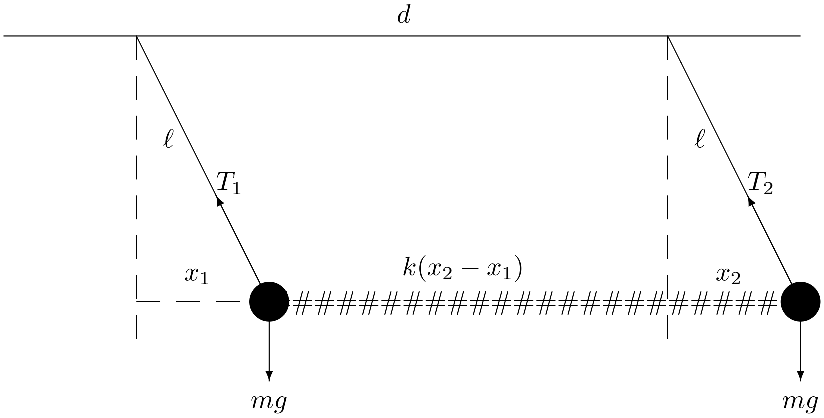
i.e.

$$\begin{aligned} x' &= \frac{1}{\sqrt{5}} (-2x + y) , \\ y' &= \frac{1}{\sqrt{5}} (x + 2y) . \end{aligned}$$

You should check this by putting the expressions for x' and y' into the new expression for F.

1.15 Normal Modes of Oscillation

As a concrete example of the kind of problem to be attacked now, consider two point particles, each of mass m, attached by light inextensible strings of length ℓ to a horizontal beam, the points of suspensions being a distance d apart. Now connect the two masses by a light spring of natural length d and spring constant d. The force pulling the two masses together is $k(x_2 - x_1)$, where x_2 and x_1 are the instantaneous displacements of the masses from equilibrium. The tension T_i in the string produces a restoring horizontal force of mgx_i/ℓ (for small displacements).



The equations of motion of the system are

$$\begin{aligned} m \frac{d^2 x_1}{dt^2} &= -\frac{mg}{\ell} x_1 + k(x_2 - x_1) , \\ m \frac{d^2 x_2}{dt^2} &= -\frac{mg}{\ell} x_2 + k(x_1 - x_2) . \end{aligned}$$

These may be recast into matrix form

$$\frac{d^2 \underline{X}}{dt^2} = \underline{A}\, \underline{X} ,$$

where

$$\underline{A} = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix} = \begin{pmatrix} -g/\ell - k/m & k/m \\ k/m & -g/\ell - k/m \end{pmatrix} .$$

These equations are coupled, in that $\ddot{x}_1$ depends also upon the value of x_2. We now have to find linear combinations of the x_i such that the equations become uncoupled. For this, let

$$\underline{X} = \underline{R}\, \underline{Y} ,$$

where $\underline{R}$ is an orthogonal matrix which does not depend upon time. Hence

$$\underline{R} \frac{d^2 \underline{Y}}{dt^2} = \underline{A}\, \underline{R}\, \underline{Y} .$$

Now multiply on the left by $\underline{R}^T$ and use the fact that $\underline{R}^T\,\underline{R} = \underline{I}$ to obtain

$$\frac{d^2\underline{Y}}{dt^2} = \underline{R}^T\,\underline{A}\,\underline{R}\,\underline{Y}\,.$$

In order that the equations be uncoupled, we need the right-hand side to be a diagonal matrix which, just as for the quadratic form problem, is that of the eigenvalues, $\underline{\Lambda}$:

$$\underline{R}^T\,\underline{A}\,\underline{R} = \underline{\Lambda}\,,$$

where $\underline{R}$ is the matrix of normalised eigenvectors. The new variables y_i satisfy the uncoupled equations

$$\ddot{y} = -\lambda_i\,y\,.$$

The first part of the problem consists of determining the eigenvalues, which are fixed by

$$\left|\begin{array}{cc} -g/\ell - k/m - \lambda & k/m \\ k/m & -g/\ell - k/m - \lambda \end{array}\right| = 0\,.$$

This has the two solutions $\lambda_1 = -g/\ell$ and $\lambda_2 = -g/\ell - 2k/m$. The equations of motion are therefore

$$\begin{aligned} \ddot{y}_1 = &\quad -\omega_1^2\,y_1 &&= -\frac{g}{\ell}\,y_1\,, \\ \ddot{y}_2 = &\quad -\omega_2^2\,y_2 &&= -\left(\frac{g}{\ell} + 2\frac{k}{m}\right)\,y_2\,. \end{aligned}$$

The general solution of these equations is

$$\begin{aligned} y_1 &= \alpha_1 \sin\omega_1 t + \beta_1 \cos\omega_1 t\,, \\ y_2 &= \alpha_2 \sin\omega_2 t + \beta_2 \cos\omega_2 t \end{aligned}$$

To find the relation between the x_i and y_i, we must find the rotation matrix $\underline{R}$, *i.e.* the eigenvectors of $\underline{A}$. For $\lambda_1 = -g/\ell$,

$$\begin{pmatrix} -k/m & k/m \\ k/m & -k/m \end{pmatrix} \begin{pmatrix} r_{11} \\ r_{21} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\,.$$

which gives $r_{11} = r_{21}$ and a normalised eigenvector of $\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$.

For $\lambda_2 = -g/\ell - 2k/m$,

$$\begin{pmatrix} k/m & k/m \\ k/m & k/m \end{pmatrix} \begin{pmatrix} r_{12} \\ r_{22} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\,,$$

which gives $r_{12} = r_{22}$, a normalised eigenvector of $\begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$. The rotation matrix is then

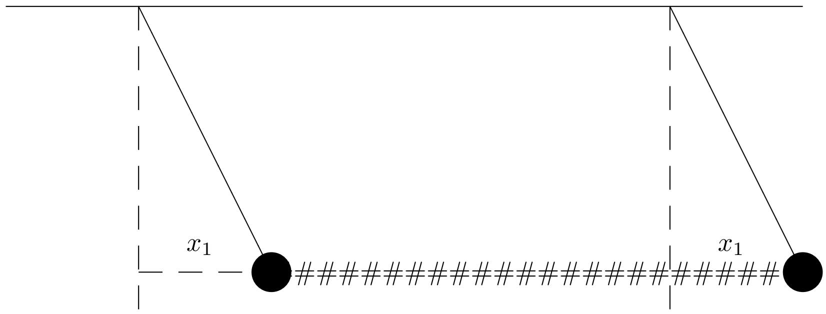
$$\underline{R} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}\,.$$

The old and new coordinates are therefore related by

$$\begin{aligned} x_1 &= \frac{1}{\sqrt{2}}\,(y_1 + y_2) \quad &:& \quad y_1 = \frac{1}{\sqrt{2}}\,(x_1 + x_2)\,, \\ x_2 &= \frac{1}{\sqrt{2}}\,(y_1 - y_2) \quad &:& \quad y_2 = \frac{1}{\sqrt{2}}\,(x_1 - x_2)\,. \end{aligned}$$

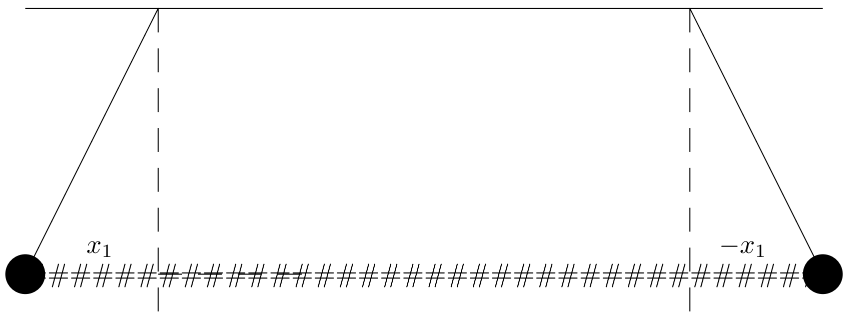
We call one of the uncoupled modes of oscillation a normal mode. Depending upon the boundary conditions, it is possible to excite one of the normal modes independently of the other. It is therefore of interest to look what the two normal modes look like in terms of the x_i.

In normal mode 1, we have $y_2 = 0$, and so $x_1 = x_2 = y_1/\sqrt{2}$.



The two pendulums swing together in phase and of course, since the two pendulums are identical, the spring is neither stretched nor compressed. Effectively the spring doesn't influence this mode at all. It is therefore not surprising that the frequency $\omega_1 = \sqrt{g/\ell}$ is just the same as that for a free pendulum of the same length.

In normal mode 1, we have $y_1 = 0$, and so $x_1 = -x_2 = y_2/\sqrt{2}$.



The two pendulums oscillate out of phase, with the spring being alternately stretched and compressed. Compared to the first normal mode, the restoring forces are here <u>increased</u> because the spring is contributing something. Hence the frequency is higher:

$$\omega_2 = \sqrt{\frac{g}{\ell} + \frac{2k}{m}}\ .$$

In the real world, we have to impose boundary conditions. Suppose at time $t = 0$ we take pendulum 1 to be at rest at the equilibrium position and pendulum 2 to be at rest at displacement $x_2 = a$. What is the subsequent motion? In terms of the y_i variables, at $t = 0$,

$$\begin{aligned} y_1 = \frac{a}{\sqrt{2}} \quad &: \quad y_2 = -\frac{a}{\sqrt{2}}\ , \\ \dot{y}_1 = 0 \quad &: \quad \dot{y}_2 = 0\ . \end{aligned}$$

Hence, at later times, the solutions are

$$\begin{aligned} y_1 &= \frac{a}{\sqrt{2}} \cos\omega_1 t\ , \\ y_2 &= -\frac{a}{\sqrt{2}} \cos\omega_2 t\ . \end{aligned}$$

In terms of the physical variables,

$$\begin{aligned} x_1 &= \frac{a}{2} \left(\cos\omega_1 t - \cos\omega_2 t\right)\ , \\ x_2 &= \frac{a}{2} \left(\cos\omega_2 t + \cos\omega_2 t\right)\ . \end{aligned}$$