Foundations of Mathematical Physics: Vectors, Tensors and Fields

2008 – 2009

John Peacock

www.roe.ac.uk/japwww/teaching/vtf.html

Textbooks The standard recommended text for this course (and later years) is Riley, Hobson & Bence *Mathematical Methods for Physics and Engineering* (Cambridge). A slightly more sophisticated approach, which can often be clearer once you know what you are doing, is taken by Arfken & Weber *Mathematical Methods for Physicists* (Academic Press).

Lecture 1: The meaning of vectors

Because we inhabit a world with more than one spatial dimension, physical phenomena frequently require us to distinguish between

Scalar : a quantity specified by a single number;

Vector : a quantity specified by a number (magnitude) and a **direction**;

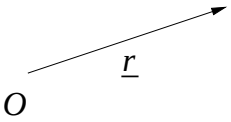
e.g. speed is a scalar, velocity is a vector. Vector algebra is an essential physics tool for describing vector quantities in a compact fashion. Modern notation is not that old: it was invented in the 1880s by Gibbs and by Heaviside. Earlier physicists from Newton to Maxwell had to work much harder to solve their problems.

Notation: Textbooks often denote vectors by boldface: $\boldsymbol{A}$, or occasionally the arrow notation: $\vec{A}$. But for writing vectors, the easiest notation is the underline: $\underline{A}$. Denote a vector by $\underline{A}$ and its magnitude by $|\underline{A}|$ or A. *Always* underline a vector to distinguish it from its magnitude. A unit vector is often denoted by a hat $\hat{\underline{A}} = \underline{A} \,/\, A$ and represents a direction.

The main intention of this course is to develop skill in using vector methods to solve problems in physics. As such, it deliberately repeats some material that has been seen before (in MM3 and AM3). The approach will be relatively informal; but this is no excuse for lack of rigour. It is important to be able to derive the key results in the subject.

1.1 Geometrical view: position vectors

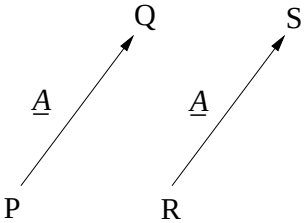
A vector is fundamentally a geometrical object, as can be seen by starting with the most basic example, the **position vector**. This is drawn as a line between an origin and a given point, with an arrow showing the direction. It is often convenient to picture this vector in a concrete way, as a thin rod carrying a physical arrowhead.



The position vector of a point relative to an origin O is normally written $\underline{r}$, which has length r (the radius of the point from the origin) and points along the unit vector $\hat{\underline{r}}$.

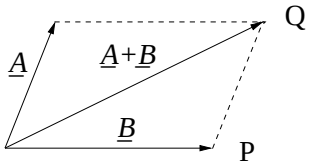
Formally speaking, this 'directed line segment' is merely a **representation** of the more abstract idea of a vector, and different kinds of vectors can be represented by a position vector: *e.g.* for a velocity vector we would draw a position vector pointing in the same direction as the velocity, and set the length proportional to the speed. This geometrical viewpoint suffices to demonstrate some of the basic properties of vectors:

Independence of origin



Vectors are unchanged by being **transported**: as drawn, both displacements from P to Q and from R to S represent the same vector. In effect, both are position vectors, but with P and R treated as the origin: the choice of origin is arbitrary.

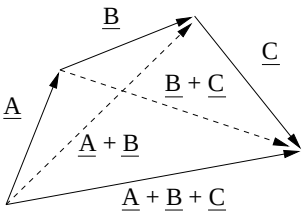
Addition of vectors: parallelogram law



$\underline{A} + \underline{B} = \underline{B} + \underline{A}$ (**commutative**).

From this, we see that the vector $\underline{A}$ that points from P to Q is just the position vector of the last point minus that of the first: we write this as $\underline{PQ} = \underline{OQ} - \underline{OP} = \underline{r}_Q - \underline{r}_P$. We prove this by treating $\underline{A}$ and $\underline{A} + \underline{B}$ as the two position vectors in the above diagram.

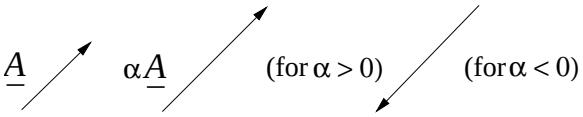
This generalises to any number of vectors: the **resultant** is obtained by adding the vectors nose to tail. This lets us prove that vector addition is **associative**:



A geometrical demonstration that $(\underline{A} + \underline{B}) + \underline{C} = \underline{A} + (\underline{B} + \underline{C})$.

Multiplication by scalars

A vector may be multiplied by a scalar to give a new vector *e.g.*



Also

$$\begin{aligned} |\alpha\underline{A}| &= |\alpha||\underline{A}| \\ \alpha(\underline{A} + \underline{B}) &= \alpha\underline{A} + \alpha\underline{B} \qquad \text{(distributive)} \\ (\alpha + \beta)\underline{A} &= \alpha\underline{A} + \beta\underline{A} \qquad \text{(distributive)} \\ \alpha(\beta\underline{A}) &= (\alpha\beta)\underline{A} \qquad \text{(associative).} \end{aligned}$$

In summary, as far as addition of vectors is concerned, or of multiplication by scalars, the power of vector notation is just that you treat vectors as if they were just a number (a 'directed number'). The important exception of multiplication of vectors will be dealt with shortly. In the meantime, there are already some common mistakes to avoid:

1. You can add vectors, but you can't add vectors and scalars.
2. Check that all quantities in a vector equation are of the same type: e.g. any equation *vector* = *scalar* is clearly wrong. (The only exception to this is if we lazily write *vector* = 0 when we mean $\underline{0}$.)
3. Never try to divide by a vector – there is no such operation.

1.2 Coordinate geometry

Although the geometrical view of vectors is fundamental, in practice it is often easier to convert vectors to a set of numbers: this is the approach to geometry pioneered by Descartes in 1637 (hence Cartesian coordinates). Now, a position vector is represented by either a row or column of numbers (**row vector** or **column vector**):

$$\underline{r} = (x, y, z) \quad \text{or} \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

assuming three dimensions for now. These numbers are the **components** of the vector. When dealing with matrices, we will normally assume the column vector to be the primary form – but otherwise it is most convenient to use row vectors.

It should be clear that this xyz triplet is just a representation of the vector. But we will commonly talk as if (x, y, z) *is* the vector itself. The coordinate representation makes it easy to prove all the results considered above: to add two vectors, we just have to add the coordinates. For example, associativity of vector addition then follows just because addition of numbers is associative.

Basis vectors

A more explicit way of writing a Cartesian vector is to introduce basis vectors denoted by either $\underline{i}$, $\underline{j}$ and $\underline{k}$ or $\underline{e}_x$, $\underline{e}_y$ and $\underline{e}_z$ which point along the x, y and z-axes. These basis vectors are **orthonormal**: *i.e.* they are all unit vectors that are mutually perpendicular. The $\underline{e}_z$ vector is related to $\underline{e}_x$ and $\underline{e}_y$ by the r.h. screw rule.

The key idea of basis vectors is that any vector can be written as a **linear superposition** of different multiples of the basis vectors. If the components of a vector $\underline{A}$ are A_x, A_y, A_z, then we write

$$\underline{A} \;=\; A_x\,\underline{i} + A_y\,\underline{j} + A_z\,\underline{k} \quad \text{or} \quad \underline{A} = A_x\,\underline{e}_x + A_y\,\underline{e}_y + A_z\,\underline{e}_z\ .$$

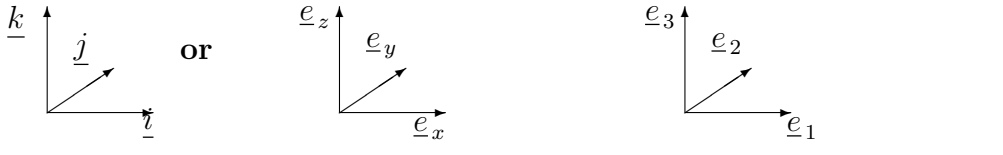
In row-vector notation, the basis vectors themselves are just

$$\underline{i} = \underline{e}_x = (1,0,0) \quad \underline{j} = \underline{e}_y = (0,1,0) \quad \underline{k} = \underline{e}_z = (0,0,1)$$

1.3 Suffix or Index notation

A more systematic labelling of basis vectors is by $\underline{e}_1$, $\underline{e}_2$ and $\underline{e}_3$. i.e. instead of $\underline{i}$ we write $\underline{e}_1$, instead of $\underline{j}$ we write $\underline{e}_2$, instead of $\underline{k}$ we write $\underline{e}_3$. This scheme is known as the **suffix** notation. Its great advantages are that it generalises easily to any number of dimensions and greatly simplifies manipulations and the verification of various identities (see later in the course).

Old Notation **New Notation**



$\underline{r} = x\underline{i} + y\underline{j} + z\underline{k}$ $\underline{r} = x\underline{e}_x + y\underline{e}_y + z\underline{e}_z$ $\underline{r} = x_1\underline{e}_1 + x_2\underline{e}_2 + x_3\underline{e}_3$

Thus any vector $\underline{A}$ in N dimensions can be written in this new notation as

$$\underline{A} = A_1\,\underline{e}_1 + A_2\,\underline{e}_2 + A_3\,\underline{e}_3 + \cdots = \sum_{i=1}^{N} A_i\,\underline{e}_i\ .$$

Free and dummy indices

We have written the components of $\underline{A}$ as A_i. This can be a common source of confusion: if we had written A_j instead, does that make any difference? In words, A_i means "the i^{th} component of $\underline{A}$". Thus, i is a **free index**: it means some integer that we have not yet specified, and indeed we might as well have called it j. The only important thing is to be consistent: if vectors $\underline{A}$ and $\underline{B}$ are equal, then all their individual components are equal, so we can write $A_i = B_i$ and get a relation that is implicitly true for any value i without having to write them all out. The relation $A_j = B_j$ expresses exactly the same result and is just a good a choice. But $A_i = B_j$ would be meaningless, because the indices don't balance. Where the value of an index is summed over, as in $\underline{A} = \sum_{i=1}^{N} A_i\,\underline{e}_i$, all possible values of i are used, and it is called a **dummy index**. Again, we can happily replace i by j or whatever, provided it is done consistently:

$$\boxed{\underline{A} = \sum_{i=1}^{3} A_i\,\underline{e}_i = \sum_{j=1}^{3} A_j\,\underline{e}_j.}$$

Example: epicycles

To illustrate the idea of shift of origin, consider the position vectors of two planets 1 & 2 (Earth and Mars, say) on circular orbits: $\underline{r}_1 = r_1(\cos\omega_1 t, \sin\omega_1 t)$, $\underline{r}_2 = r_2(\cos\omega_2 t, \sin\omega_2 t)$. The position vector of Mars as seen from Earth is

$$\underline{r}_{21} = \underline{r}_2 - \underline{r}_1 = (r_2\cos\omega_2 t - r_1\cos\omega_1 t,\ r_2\sin\omega_2 t - r_1\sin\omega_1 t).$$

This describes an *epicycle*: Mars moves on a small circle of radius r_1 whose centre is carried round in a circle of radius r_2. Although epicycles have a bad press, notice that this is an exact description of Mars's motion, using the same number of free parameters as the heliocentric view. Fortunately, planetary orbits are not circles, otherwise the debate over whether the Sun or the Earth made the better origin might have continued much longer.

1.4 Vector physics: independence of basis

Although the coordinate approach is convenient and practical, expressing vectors as components with respect to a basis is against the spirit of the power of vectors as a tool for physics – which is that *physical laws relating vectors must be true independent of the coordinate system being used*. Consider the case of **vector dynamics**:

$$\underline{F} = m\underline{a} = m\,\frac{d}{dt}\,\underline{v} = m\,\frac{d^2}{dt^2}\,\underline{r}.$$

In one compact statement, this equation says that $F = ma$ is obeyed separately by all the components of F and a. The simplest case is where one of the basis vectors points in the

direction of $\underline{F}$, in which case there is only one scalar equation to consider. But the vector equation is true whatever basis we use. We will return to this point later when we consider how the components of vectors alter when the basis used to describe them is changed.

Example: centripetal force

As an example of this point in action, consider again circular motion in 2D: $\underline{r} = r(\cos\omega t, \sin\omega t)$. What force is needed to produce this motion? We get the acceleration by differentiating twice w.r.t. t:

$$\underline{F} = m\underline{a} = m\,\frac{d^2}{dt^2}\underline{r} = m\,r(-\omega^2\sin\omega t, -\omega^2\cos\omega t) = -m\omega^2\,\underline{r}.$$

Although we have used an explicit set of components as an intermediate step, the final result just says that the required force is $m\omega^2 r$, directed radially inwards.

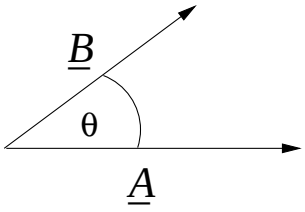
Lecture 2: Multiplying vectors

So far, vector notation is completely pain-free: we just treat vectors as if they were numbers and the algebra of addition or subtraction is identical. What about multiplication? What could $\underline{A}\ \underline{B}$ mean? To see this, we have to think geometrically, and there are two aspects that resemble multiplication: the projection of one vector onto another, and the area of the parallelogram formed by two vectors.

2.1 Scalar or dot product

The scalar product (also known as the dot product) between two vectors is defined as

$$\boxed{(\underline{A}\cdot\underline{B}) \equiv AB\cos\theta, \text{ where } \theta \text{ is the angle between } \underline{A} \text{ and } \underline{B}}$$

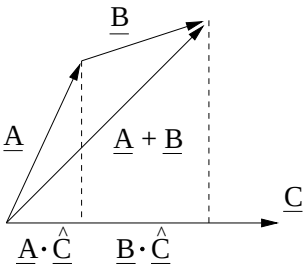


$(\underline{A}\cdot\underline{B})$ is a scalar — *i.e.* a single number. By definition, the scalar product is **commutative**: $(\underline{A}\cdot\underline{B}) = (\underline{B}\cdot\underline{A})$.

.

The geometrical significance of the scalar product is that it **projects** one vector onto another: $\underline{A}\cdot\hat{\underline{B}}$ is the component of A in the direction of the unit vector $\hat{\underline{B}}$, and its magnitude is $A\cos\theta$. This viewpoint makes it easy to prove that the scalar product is

Distributive over addition $\boxed{(\underline{A}+\underline{B})\cdot\underline{C} = \underline{A}\cdot\underline{C} + \underline{B}\cdot\underline{C}}$.



The components of $\underline{A}$ and $\underline{B}$ along $\underline{C}$ clearly add to make the components of $\underline{A}+\underline{B}$ along $\underline{C}$.

Scalar product in terms of components

You know very well that the scalar product is worked out in practice using the components of the vector:

$$\boxed{\underline{A}\cdot\underline{B}=\sum_{i=1}^{3}A_i\,B_i\;;}$$

Let's prove that these two definitions are identical; this requires the distributive property of the scalar product. If $\underline{A}=\sum_i A_i\underline{e}_i$, then

$$\underline{A}\cdot\underline{e}_1 \;\;=\;\; (A_1\underline{e}_1+A_2\underline{e}_2+A_3\underline{e}_3)\cdot\underline{e}_1=A_1,$$

so the orthonormality of the basis vectors picks out the projection of $\underline{A}$ in the direction of $\underline{e}_1$, and similarly for the components A_2 and A_3. In general we may write

$$\boxed{\underline{A}\cdot\underline{e}_i=\underline{e}_i\cdot\underline{A}\equiv A_i \quad \text{or sometimes} \quad (\underline{A})_i.}$$

If we now write $\underline{B}=\sum_i B_i\underline{e}_i$, then $\underline{A}\cdot\underline{B}$ amounts to picking out in turn A_i and multiplying it by B_i. So we recover the standard formula for the scalar product based on (i) distributivity; (ii) orthonormality of the basis.

Summary of properties of scalar product

(i) $\underline{A}\cdot\underline{B}=\underline{B}\cdot\underline{A}\;\;;\;\;\underline{A}\cdot(\underline{B}+\underline{C})=\underline{A}\cdot\underline{B}+\underline{A}\cdot\underline{C}$

(ii) $\hat{\underline{n}}\cdot\underline{A}$ = the scalar projection of $\underline{A}$ onto $\hat{\underline{n}}$, where $\hat{\underline{n}}$ is a unit vector

(iii) $(\hat{\underline{n}}\cdot\underline{A})\,\hat{\underline{n}}$ = the vector projection of $\underline{A}$ onto $\hat{\underline{n}}$

(*iv*) $\boxed{\underline{A}\cdot\underline{A}=|\underline{A}|^2}$ which defines the magnitude of a vector. For a unit vector $\hat{\underline{A}}\cdot\hat{\underline{A}}=1$

Example: parallel and perpendicular components

A vector may be resolved with respect to some direction $\hat{\underline{n}}$ into a parallel component $\underline{A}_{\parallel}=(\hat{\underline{n}}\cdot\underline{A})\hat{\underline{n}}$. There must therefore be a perpendicular component $\underline{A}_{\perp}=\underline{A}-\underline{A}_{\parallel}$. If this reasoning makes sense, we should find that $\underline{A}_{\parallel}$ and $\underline{A}_{\perp}$ are at right angles. To prove this, evaluate

$$\underline{A}_{\perp}\cdot\hat{\underline{n}}=(\underline{A}-(\hat{\underline{n}}\cdot\underline{A})\hat{\underline{n}})\cdot\hat{\underline{n}}=\underline{A}\cdot\hat{\underline{n}}-\hat{\underline{n}}\cdot\underline{A}=0$$

(because $\hat{\underline{n}}\cdot\hat{\underline{n}}=1$).

Example: the cosine rule

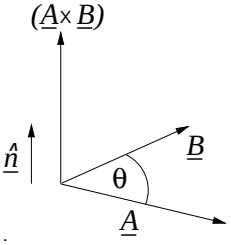
Consider two position vectors $\underline{A}$ and $\underline{B}$. They define a triangle, whose third side is the vector $\underline{C} = \underline{A} - \underline{B}$. $C^2 = (\underline{A} - \underline{B})^2 = (\underline{A} - \underline{B}) \cdot (\underline{A} - \underline{B}) = A^2 + B^2 - 2\underline{A} \cdot \underline{B}$. Hence we have a simple derivation of $C^2 = A^2 + B^2 - 2AB\cos\theta$, where θ is the angle between $\underline{A}$ and $\underline{B}$.

2.2 The vector or 'cross' product

The vector product represents the fact that two vectors define a parallelogram. This geometrical object has an area, but also an orientation in space – which can be represented by a vector.

$$(\underline{A} \times \underline{B}) \equiv AB\sin\theta\, \hat{\underline{n}} \text{ , where } \hat{\underline{n}} \text{ in the 'right-hand screw direction'}$$

i.e. $\hat{\underline{n}}$ is a unit vector normal to the plane of $\underline{A}$ and $\underline{B}$, in the direction of a right-handed screw for rotation of $\underline{A}$ to $\underline{B}$ (through $< \pi$ radians).

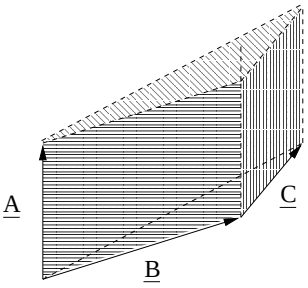


$(\underline{A} \times \underline{B})$ is a vector — i.e. it has a direction and a length.

It is important to note that the idea of the vector product is unique to three dimensions. In 2D, the area defined by two vectors is just a scalar: there is no choice about the orientation. In N dimensions, it turns out that $N(N-1)/2$ numbers are needed to specify the size and orientation of an element of area. So representing this by a vector is only possible for $N = 3$. The idea of such an 'oriented area' always exists, of course, and the general name for it is the **wedge product**, denoted by $\underline{A} \wedge \underline{B}$. You can feel free to use this notation instead of $\underline{A} \times \underline{B}$, since they are the same thing in 3D.

The vector product shares a critical property with the scalar product, but unfortunately one that is not as simple to prove; it is

Distributive over addition $\underline{A} \times (\underline{B} + \underline{C}) = \underline{A} \times \underline{B} + \underline{A} \times \underline{C}$.



It is easy enough to see that $\underline{A} \times (\underline{B} + \underline{C}) = \underline{A} \times \underline{B} + \underline{A} \times \underline{C}$ if all three vectors lie in the same plane. The various parallelograms of interest (shaded in this figure) differ by the triangles at top and bottom, which are clearly of identical shape.

When the three vectors are not in the same plane, the proof is more involved. What we shall do for now is to *assume* that the result is true in general, and see where it takes us. We can then work backwards at the end.

2.3 The vector product in terms of components

Because of the distributive property, when we write $\underline{A} \times \underline{B}$ in terms of components, the expression comes down to a sum of products of the basis vectors with each other:

$$\underline{A} \times \underline{B} = \left(\sum_{i=1}^{3} A_i \, \underline{e}_i \right) \times \left(\sum_{j=1}^{3} B_j \, \underline{e}_j \right) = \sum_{i=1}^{3} \sum_{j=1}^{3} A_i \, B_j \, (\underline{e}_i \times \underline{e}_j) \; .$$

Almost all the cross products of basis vectors vanish. The only one we need is the one that defines the z axis:

$$\boxed{\underline{e}_1 \times \underline{e}_2 = \underline{e}_3,}$$

and cyclic permutations of this. If the order is reversed, so is the sign: $\underline{e}_2 \times \underline{e}_1 = -\underline{e}_3$. In this way, we get

$$\underline{A} \times \underline{B} = \underline{e}_1(A_2B_3 - A_3B_2) + \underline{e}_2(A_3B_1 - A_1B_3) + \underline{e}_3(A_1B_2 - A_2B_1)$$

from which we deduce that

$$(\underline{A} \times \underline{B})_1 \;=\; (A_2B_3 - A_3B_2) \text{ , etc.}$$

So finally, we have recovered the familiar expression in which the vector product is written as a determinant:

$$\boxed{\underline{A} \times \underline{B} = \begin{vmatrix} \underline{e}_1 & \underline{e}_2 & \underline{e}_3 \\ A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \end{vmatrix} .}$$

If we were to take this as the *definition* of the vector product, it is easy to see that the distributive property is obeyed. But now, how do we know that the geometrical properties of $\underline{A} \times \underline{B}$ are satisfied? One way of closing this loop is to derive the determinant expression in another way. If $\underline{A} \times \underline{B} = \underline{C}$, then $\underline{C}$ must be perpendicular to both vectors: $\underline{A} \cdot \underline{C} = \underline{B} \cdot \underline{C} = 0$. With some effort, these simultaneous equations can be solved to find the components of $\underline{C}$ (within some arbitrary scaling factor, since there are only two equations for three components). Or we can start by assuming the determinant and show that $\underline{A} \times \underline{B}$ is perpendicular to $\underline{A}$ and $\underline{B}$ and has magnitude $AB \sin\theta$. Again, this is quite a bit of algebra. For the present, we content ourselves with checking a simple test case, where we choose $\underline{A} = (A, 0, 0)$ along the x axis, and $\underline{B} = (B_1, B_2, 0)$ in the xy plane. This gives $\underline{A} \times \underline{B} = (0, 0, AB_2)$, which points in the z direction as required. From the scalar product, $\cos\theta = AB_1/AB = B_1/B$ (where $B = \sqrt{B_1^2 + B_2^2}$), so $\sin\theta = \sqrt{1 - \cos^2\theta} = \sqrt{1 - B_1^2/B^2} = B_2/B$. Hence $AB_2 = AB\sin\theta$, as required.

Summary of properties of vector product

(i) $\underline{A} \times \underline{B} = -\underline{B} \times \underline{A}$

(ii) $\underline{A} \times \underline{B} = 0$ if $\underline{A}, \underline{B}$ are parallel

(iii) $\underline{A} \times (\underline{B} + \underline{C}) = \underline{A} \times \underline{B} + \underline{A} \times \underline{C}$

(iv) $\underline{A} \times (\alpha \underline{B}) = \alpha \underline{A} \times \underline{B}$

Lecture 3: More vector multiplication and physical applications

3.1 The scalar triple product

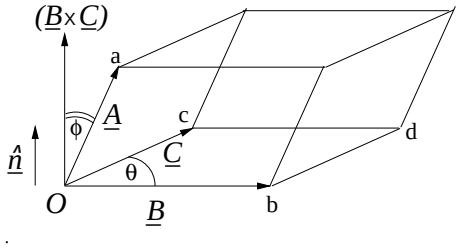
By introducing a third vector, we extend the geometrical idea of an area to the volume of the parallelipiped. The scalar triple product is defined as follows

$$(\underline{A}, \underline{B}, \underline{C}) \equiv \underline{A} \cdot (\underline{B} \times \underline{C})$$

Notes

(i) If $\underline{A}$, $\underline{B}$ and $\underline{C}$ are three concurrent edges of a parallelepiped, the volume is $(\underline{A}, \underline{B}, \underline{C})$.

To see this, note that:



$$\begin{aligned}
\text{area of the base} &= \text{area of parallelogram } Obdc \\
&= B\,C\,\sin\theta = |\underline{B} \times \underline{C}| \\
\text{height} &= A\cos\phi = \hat{\underline{n}} \cdot \underline{A} \\
\text{volume} &= \text{area of base} \times \text{height} \\
&= B\,C\,\sin\theta\,\hat{\underline{n}} \cdot \underline{A} \\
&= \underline{A} \cdot (\underline{B} \times \underline{C})
\end{aligned}$$

(ii) If we choose $\underline{C}, \underline{A}$ to define the base then a similar calculation gives volume $= \underline{B} \cdot (\underline{C} \times \underline{A})$
We deduce the following symmetry/antisymmetry properties:

$$(\underline{A}, \underline{B}, \underline{C}) = (\underline{B}, \underline{C}, \underline{A}) = (\underline{C}, \underline{A}, \underline{B}) = -(\underline{A}, \underline{C}, \underline{B}) = -(\underline{B}, \underline{A}, \underline{C}) = -(\underline{C}, \underline{B}, \underline{A})$$

(iii) If $\underline{A}, \underline{B}$ and $\underline{C}$ are **coplanar** (*i.e.* all three vectors lie in the same plane) then $V = (\underline{A}, \underline{B}, \underline{C}) = 0$, and vice-versa.

It is easy to write down an expression for the scalar triple product in terms of components:

$$\begin{aligned}
\underline{A} \cdot (\underline{B} \times \underline{C}) &= \sum_{i=1}^{3} A_i (\underline{B} \times \underline{C})_i \\
&= A_1(B_2C_3 - C_2B_3) - A_2(B_1C_3 - C_1B_3) + A_3(B_1C_2 - C_1B_2) \\
&= \begin{vmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{vmatrix}.
\end{aligned}$$

The symmetry properties of the scalar triple product may be deduced from this by noting that interchanging two rows (or columns) changes the value by a factor -1.

3.2 The vector triple product

There are *several* ways of combining 3 vectors to form a new vector.
e.g. $\underline{A} \times (\underline{B} \times \underline{C})$; $(\underline{A} \times \underline{B}) \times \underline{C}$, etc. Note carefully that *brackets are important*, since

$$\underline{A} \times (\underline{B} \times \underline{C}) \neq (\underline{A} \times \underline{B}) \times \underline{C} \, .$$

Expressions involving two (or more) vector products can be simplified by using the identity

$$\boxed{\underline{A} \times (\underline{B} \times \underline{C}) = \underline{B}(\underline{A} \cdot \underline{C}) - \underline{C}(\underline{A} \cdot \underline{B})} \, .$$

This is a result you must memorise (say "back cab" and picture a 'black cab' taxi reversing). If you worry that you may have misremembered the bracket order, remember that $(\underline{A} \times \underline{B}) \times \underline{C}$ would have to be orthogonal to $\underline{C}$ and hence made up from $\underline{A}$ and $\underline{B}$.

To prove this (or at least make it plausible), we can exploit the freedom to choose a basis, and take coordinates such that $\underline{C} = (0, 0, C)$ points along the z axis. In this case,

$$\underline{B} \times \underline{C} = \begin{vmatrix} \underline{e}_1 & \underline{e}_2 & \underline{e}_3 \\ B_1 & B_2 & B_3 \\ 0 & 0 & C \end{vmatrix} = (CB_2, -CB_1, 0).$$

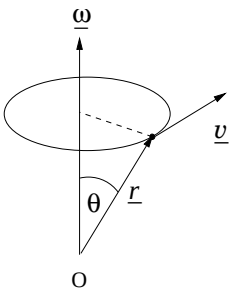
$$\underline{A} \times (\underline{B} + \underline{C}) = \begin{vmatrix} \underline{e}_1 & \underline{e}_2 & \underline{e}_3 \\ A_1 & A_2 & A_3 \\ CB_2 & -CB_1 & 0 \end{vmatrix} = (CA_3B_1, CA_3B_2, -[CA_1B_1 + CA_2B_1]).$$

Finally, we write the result as a relation between vectors, in which case it becomes independent of coordinates, in the same way as we deduced the centripetal force earlier.

3.3 The vector product and angular momentum

(i) Angular velocity

Consider a point in a rigid body rotating with **angular velocity** $\underline{\omega}$: $|\underline{\omega}|$ is the angular speed of rotation measured in radians per second and $\underline{\hat{\omega}}$ lies along the axis of rotation. Let the position vector of the point with respect to an origin O on the axis of rotation be $\underline{r}$.



You should convince yourself that $\underline{v} = \underline{\omega} \times \underline{r}$ by checking that this gives the right direction for $\underline{v}$; that it is perpendicular to the plane of $\underline{\omega}$ and $\underline{r}$; that the magnitude $|\underline{v}| = \omega r \sin\theta = \omega \times$ radius of circle in which the point is travelling

(ii) Angular momentum

Now consider the **angular momentum** of the particle defined by $\underline{L} = \underline{r} \times (m\underline{v})$ where m is the mass of the particle.

Using the above expression for $\underline{v}$ we obtain

$$\underline{L} = m\underline{r} \times (\underline{\omega} \times \underline{r}) = m\left[\underline{\omega} r^2 - \underline{r}(\underline{r} \cdot \underline{\omega})\right]$$

where we have used the identity for the vector triple product. Note that $\underline{L} = 0$ if $\underline{\omega}$ and $\underline{r}$ are parallel. Note also that only if $\underline{r}$ is perpendicular to $\underline{\omega}$ do we obtain $\underline{L} = m\underline{\omega}r^2$, which means that only then are $\underline{L}$ and $\underline{\omega}$ in the same direction.

(iii) Torque

This last result sounds peculiar, but makes a good example of how physical laws are independent of coordinates. The **torque** or **moment** of a force about the origin $\underline{G} = \underline{r} \times \underline{F}$ where $\underline{r}$ is the position vector of the point where the force is acting and $\underline{F}$ is the force vector through that point. Torque causes angular momentum to change:

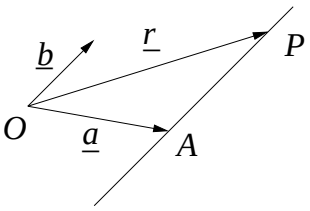
$$\frac{d}{dt}\underline{L} = \dot{\underline{r}} \times m\underline{v} + \underline{r} \times m\underline{a} = 0 + \underline{r} \times \underline{F} = \underline{G}.$$

If the origin is in the centre of the circle, then the centripetal force generates zero torque – but otherwise G is nonzero. This means that $\underline{L}$ has to change with time. Here, we have assumed that the vector product obeys the product rule under differentiation. This will be easy to prove a little later.

3.4 Equations of lines and planes

The Equation of a Line

Suppose that P lies on a line which passes through a point A which has a position vector $\underline{a}$ with respect to an origin O. Let P have position vector $\underline{r}$ relative to O and let $\underline{b}$ be a vector through the origin in a direction parallel to the line.



We may write

$$\underline{r} = \underline{a} + \lambda\underline{b}$$

which is the **parametric equation of the line** *i.e.* as we vary the parameter λ from $-\infty$ to ∞, $\underline{r}$ describes all points on the line.

Rearranging and using $\underline{b} \times \underline{b} = 0$, we can also write this as

$$(\underline{r} - \underline{a}) \times \underline{b} \;=\; 0$$

or

$$\boxed{\underline{r} \times \underline{b} = \underline{c}}$$

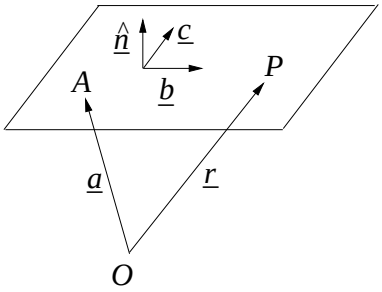
where $\underline{c} = \underline{a} \times \underline{b}$ is normal to the plane containing the line and origin.

Notes

(i) $\underline{r} \times \underline{b} = \underline{c}$ is an **implicit equation** for a line

(ii) $\underline{r} \times \underline{b} = 0$ is the equation of a line through the origin.

The Equation of a Plane



$\underline{r}$ is the position vector of an arbitrary point P on the plane
$\underline{a}$ is the position vector of a fixed point A in the plane
$\underline{b}$ and $\underline{c}$ are parallel to the plane but non-collinear: $\underline{b} \times \underline{c} \neq 0$.

We can express the vector $\underline{AP}$ in terms of $\underline{b}$ and $\underline{c}$, so that:

$$\underline{r} \;=\; \underline{a} + \underline{AP} \;=\; \underline{a} + \lambda \underline{b} + \mu \underline{c}$$

for some λ and μ. This is the **parametric equation of the plane**.

We define the unit normal to the plane

$$\hat{\underline{n}} \;=\; \frac{\underline{b} \times \underline{c}}{|\underline{b} \times \underline{c}|} \,.$$

Since $\underline{b} \cdot \hat{\underline{n}} = \underline{c} \cdot \hat{\underline{n}} = 0$, we have the implicit equation

$$(\underline{r} - \underline{a}) \cdot \hat{\underline{n}} \;=\; 0 \,.$$

Alternatively, we can write this as

$$\boxed{\underline{r} \cdot \hat{\underline{n}} \;=\; p} \,,$$

where $p = \underline{a} \cdot \hat{\underline{n}}$ is the perpendicular distance of the plane from the origin. **Note**: $\underline{r} \cdot \underline{a} \;=\; 0$ is the equation for a plane through the origin (with unit normal $\underline{a}/|\underline{a}|$).

This is a very important equation which you must be able to recognise. In algebraic terms, it means that $ax + by + cz + d = 0$ is the equation of a plane.

Example: Is the following set of equations consistent?

$$\underline{r} \times \underline{b} = \underline{c}$$
$$\underline{r} = \underline{a} \times \underline{c}$$

Geometrical interpretation: the first equation is the (implicit) equation for a line whereas the second equation is the (explicit) equation for a point. Thus the question is whether the point is on the line. If we insert the 2nd into the into the l.h.s. of the first we find

$$\underline{r} \times \underline{b} = (\underline{a} \times \underline{c}) \times \underline{b} = -\underline{b} \times (\underline{a} \times \underline{c}) = -\underline{a}\,(\underline{b} \cdot \underline{c}) + \underline{c}\,(\underline{a} \cdot \underline{b})$$

Now $\underline{b} \cdot \underline{c} = \underline{b} \cdot (\underline{r} \times \underline{b}) = 0$ thus the previous equation becomes

$$\underline{r} \times \underline{b} = \underline{c}\,(\underline{a} \cdot \underline{b})$$

so that, on comparing with the first given equation, we require

$$\underline{a} \cdot \underline{b} = 1$$

for the equations to be consistent.

Lecture 4: More on suffix notation

So far, we have been revising material that should have been relatively familiar. Now it is time to introduce some more powerful tools – whose idea is to make vector calculations quicker and easier to perform.

To revise the issues we face, consider the vector equation

$$\underline{A} = (\underline{B} \cdot \underline{C})\, \underline{D},$$

which must hold for each component separately:

$$A_i = (\underline{B} \cdot \underline{C})\, D_i \quad i = 1, 2, 3.$$

The free index i occurs **once and only once** in each term of the equation. Every term in the equation must be of the same kind *i.e.* have the same free indices. In order to write the scalar product that appears in the second term in suffix notation, we must avoid using i as a dummy index, since as we have already used it as a free index. It is better to write

$$A_i = \left(\sum_{k=1}^{3} B_k C_k \right) D_i$$

since this can then be rewritten without the brackets as

$$A_i = \sum_{k=1}^{3} B_k C_k D_i.$$

4.1 The Kronecker delta symbol δ_{ij}

We define the symbol δ_{ij} (pronounced "delta i j"), where i and j can take on the values 1 to 3, such that

$$\boxed{\begin{array}{rcl} \delta_{ij} & = & 1 \text{ if } i = j \\ & = & 0 \text{ if } i \neq j \end{array}}$$

i.e. $\delta_{11} = \delta_{22} = \delta_{33} = 1$ and $\delta_{12} = \delta_{13} = \delta_{23} = \cdots = 0$.

The equations satisfied by the **orthonormal basis vectors** $\underline{e}_i$ can all now be written as

$$\boxed{\underline{e}_i \cdot \underline{e}_j \;=\; \delta_{ij}} \;.$$

e.g. $\quad \underline{e}_1 \cdot \underline{e}_2 \;=\; \delta_{12} = 0 \quad ; \quad \underline{e}_1 \cdot \underline{e}_1 \;=\; \delta_{11} = 1$

Notes

(i) Since there are two free indices i and j, $\underline{e}_i \cdot \underline{e}_j \;=\; \delta_{ij}$ is equivalent to 9 equations

(ii) $\delta_{ij} \;=\; \delta_{ji}$ [*i.e.* δ_{ij} is *symmetric* in its indices.]

(iii) $\sum_{i=1}^{3} \delta_{ii} \;=\; 3 \quad (= \delta_{11} + \delta_{22} + \delta_{33})$

(iv) $\sum_{j=1}^{3} A_j \delta_{jk} \;= A_1\delta_{1k} + A_2\delta_{2k} + A_3\delta_{3k}$

Remember that k is a free index. Thus if $k = 1$ then only the first term on the rhs contributes and rhs $= A_1$, similarly if $k = 2$ then rhs $= A_2$ and if $k = 2$ then rhs $= A_3$. Thus we conclude that

$$\boxed{\sum_{j=1}^{3} A_j \delta_{jk} \;= A_k}$$

In other words, the Kronecker delta picks out the kth term in the sum over j. This is in particular true for the multiplication of two Kronecker deltas:

$\sum_{j=1}^{3} \delta_{ij}\delta_{jk} \;= \delta_{i1}\delta_{1k} + \delta_{i2}\delta_{2k} + \delta_{i3}\delta_{3k} \;= \delta_{ik}$

Generalising the reasoning in (iv) implies the so-called **sifting property**:

$$\boxed{\sum_{j=1}^{3} (\text{anything })_j \delta_{jk} \;= (\text{anything })_k}$$

where $(\text{anything})_j$ denotes any expression that has a free index j. In effect, δ_{ij} is a kind of mathematical virus, whose main love in life it to replace the index you first thought of with one of its own. Once you get used to this behaviour, it's a very powerful trick.

Examples of the use of this symbol are:

1.
$$\begin{aligned} \underline{A} \cdot \underline{e}_j \;&=\; \left(\sum_{i=1}^{3} A_i \underline{e}_i\right) \cdot \underline{e}_j \;=\; \sum_{i=1}^{3} A_i \,(\underline{e}_i \cdot \underline{e}_j) \\ &=\; \sum_{i=1}^{3} A_i \delta_{ij} \;=\; A_j, \quad \text{since terms with } i \neq j \text{ vanish.} \end{aligned}$$

2.
$$\begin{aligned} \underline{A} \cdot \underline{B} \;&=\; \left(\sum_{i=1}^{3} A_i \underline{e}_i\right) \cdot \left(\sum_{j=1}^{3} B_j \underline{e}_j\right) \\ &=\; \sum_{i=1}^{3}\sum_{j=1}^{3} A_i B_j \,(\underline{e}_i \cdot \underline{e}_j) \;=\; \sum_{i=1}^{3}\sum_{j=1}^{3} A_i B_j \delta_{ij} \\ &=\; \sum_{i=1}^{3} A_i B_i \quad \text{or} \quad \sum_{j=1}^{3} A_j B_j. \end{aligned}$$

Matrix representation of δ_{ij}

We may label the elements of a (3×3) matrix $\underline{\underline{M}}$ as M_{ij},

$$\underline{\underline{M}} \;=\; \begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{pmatrix}.$$

Note the 'double-underline' convention that we shall use to denote matrices and distinguish them from vectors and scalars. Textbooks normally use a different convention and denote a matrix in bold, $\mathbf{M}$, but this is not practical for writing matrices by hand.

Thus we see that if we write δ_{ij} as a matrix we find that it is the identity matrix $\underline{\underline{I}}$.

$$\delta_{ij} \;=\; \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

4.2 Einstein summation convention

As you will have noticed, the novelty of writing out summations in full soon wears thin. A way to avoid this tedium is to adopt the Einstein summation convention; by adhering strictly to the following rules the summation signs are suppressed.

Rules

(i) Omit summation signs

(ii) If a suffix appears twice, a summation is implied *e.g.* $A_iB_i = A_1B_1 + A_2B_2 + A_3B_3$
Here i is a *dummy* index.

(iii) If a suffix appears only once it can take any value *e.g.* $A_i = B_i$ holds for $i = 1, 2, 3$
Here i is a *free* index. Note that there may be more than one free index. **Always** check that the free indices match on both sides of an equation *e.g.* $A_j = B_i$ is WRONG.

(iv) A given suffix **must not** appear more than twice in any term of an expression. Again, **always** check that there are no multiple indices *e.g.* $A_iB_iC_i$ is WRONG.

Examples

$\underline{A} = A_i\underline{e}_i$ *here i is a dummy index.*

$\underline{A} \cdot \underline{e}_j = A_i\underline{e}_i \cdot \underline{e}_j = A_i\delta_{ij} = A_j$ *here i is a dummy index but j is a free index.*

$\underline{A} \cdot \underline{B} = (A_i\underline{e}_i) \cdot (B_j\underline{e}_j) = A_iB_j\delta_{ij} = A_jB_j$ *here i,j are dummy indices.*

$(\underline{A} \cdot \underline{B})(\underline{A} \cdot \underline{C}) = A_iB_iA_jC_j$ *again i,j are dummy indices.*

Armed with the summation convention one can rewrite many of the equations of the previous lecture without summation signs *e.g.* the sifting property of δ_{ij} now becomes

$$\boxed{(\text{anything})_j\delta_{jk} \;= (\text{anything})_k}$$

so that, for example, $\delta_{ij}\delta_{jk} = \delta_{ik}$

From now on, except where indicated, the summation convention will be assumed. You should make sure that you are completely at ease with it.

4.3 Levi-Civita symbol ϵ_{ijk}

We saw in the last lecture how δ_{ij} could be used to greatly simplify the writing out of the orthonormality condition on basis vectors.

We seek to make a similar simplification for the vector products of basis vectors *i.e.* we seek a simple, uniform way of writing the equations

$$\underline{e}_1 \times \underline{e}_2 = \underline{e}_3 \quad \underline{e}_2 \times \underline{e}_1 = -\underline{e}_3 \quad \underline{e}_1 \times \underline{e}_1 = 0 \quad \textit{etc.}$$

To do so we define the Levi-Civita symbol ϵ_{ijk} (pronounced 'epsilon i j k'), where i, j and k can take on the values 1 to 3, such that

$$\boxed{\begin{aligned} \epsilon_{ijk} &= +1 \text{ if } ijk \text{ is an } \textit{even} \text{ permutation of 123 ;} \\ &= -1 \text{ if } ijk \text{ is an } \textit{odd} \text{ permutation of 123 ;} \\ &= \ \ 0 \text{ otherwise (} \textit{i.e.} \text{ 2 or more indices are the same) .} \end{aligned}}$$

An *even* permutation consists of an *even* number of transpositions.
An *odd* permutations consists of an *odd* number of transpositions.

For example,
$$\begin{aligned} \epsilon_{123} &= +1\,; \\ \epsilon_{213} &= -1 \ \{ \text{ since } (123) \to (213) \text{ under } \textit{one} \text{ transposition } [1 \leftrightarrow 2]\}\,; \\ \epsilon_{312} &= +1 \ \{(123) \to (132) \to (312); \text{ 2 transpositions; } [2 \leftrightarrow 3][1 \leftrightarrow 3]\}\,; \\ \epsilon_{113} &= 0\,; \quad \epsilon_{111} = 0\,; \text{ etc.} \end{aligned}$$

$$\boxed{\epsilon_{123} = \epsilon_{231} = \epsilon_{312} = +1 \ ; \quad \epsilon_{213} = \epsilon_{321} = \epsilon_{132} = -1 \ ; \quad \text{all others} = 0 \ .}$$

4.4 Vector product

The equations satisfied by the vector products of the (right-handed) orthonormal basis vectors $\underline{e}_i$ can now be written uniformly as

$$\boxed{\underline{e}_i \times \underline{e}_j \ = \ \epsilon_{ijk}\, \underline{e}_k} \quad (i,j = 1,2,3) \ .$$

For example,

$$\underline{e}_1 \times \underline{e}_2 \ = \ \epsilon_{121}\, \underline{e}_1 + \epsilon_{122}\, \underline{e}_2 + \epsilon_{123}\, \underline{e}_3 \ = \ \underline{e}_3 \quad ; \quad \underline{e}_1 \times \underline{e}_1 \ = \ \epsilon_{111}\, \underline{e}_1 + \epsilon_{112}\, \underline{e}_2 + \epsilon_{113}\, \underline{e}_3 \ = \ 0$$

Also,
$$\underline{A} \times \underline{B} = A_i \ B_j \ \underline{e}_i \times \underline{e}_j = \epsilon_{ijk}\, A_i B_j\, \underline{e}_k$$

This gives the very important relation for the components of the vector product:

$$\boxed{(\underline{A} \times \underline{B})_k \ = \ \epsilon_{ijk}\, A_i B_j}$$

It also allows us to see directly that the definition of $\underline{A} \times \underline{B}$ in terms of a determinant is correct: the general expression for a determinant of a 3×3 matrix $\underline{\underline{M}}$ is

$$\det(\underline{\underline{M}}) = |\underline{\underline{M}}| = \epsilon_{ijk} M_{1i} M_{2j} M_{3k},$$

with similar relations for matrices of different size (for a 2×2 we need $\epsilon_{12} = +1$, $\epsilon_{21} = -1$).

Example: differentiation of $\underline{A} \times \underline{B}$

Now we can prove the result we assumed earlier in differentiating $\underline{A} \times \underline{B}$:

$$\frac{d}{dt}\,\underline{A} \times \underline{B} = \frac{d}{dt}\,(\epsilon_{ijk}\, A_i B_j\, \underline{e}_k),$$

but ϵ_{ijk} and $\underline{e}_k$ are not functions of time, so we can use the ordinary product rule on the numbers A_i and B_j:

$$\frac{d}{dt}\,\underline{A} \times \underline{B} = (\epsilon_{ijk}\, \dot{A}_i B_j\, \underline{e}_k) + (\epsilon_{ijk}\, A_i \dot{B}_j\, \underline{e}_k) = \dot{\underline{A}} \times \underline{B} + \underline{A} \times \dot{\underline{B}}.$$

Lecture 5: Coordinate transformations and change of basis

5.1 Linear transformation of basis

Suppose $\{\underline{e}_i\}$ and $\{\underline{e}_i{}'\}$ are two different orthonormal bases. The new basis vectors must have components in the old basis, so clearly $\underline{e}_1{}'$ can be written as a linear combination of the vectors $\underline{e}_1, \underline{e}_2, \underline{e}_3$:

$$\underline{e}_1{}' = \lambda_{11}\underline{e}_1 + \lambda_{12}\underline{e}_2 + \lambda_{13}\underline{e}_3$$

with similar expressions for $\underline{e}_2{}'$ and $\underline{e}_3{}'$. In summary,

$$\boxed{\underline{e}_i{}' = \lambda_{ij}\,\underline{e}_j}$$

(assuming summation convention) where λ_{ij} $(i = 1, 2, 3$ and $j = 1, 2, 3)$ are the 9 numbers relating the basis vectors $\underline{e}_1{}'$, $\underline{e}_2{}'$ and $\underline{e}_3{}'$ to the basis vectors $\underline{e}_1$, $\underline{e}_2$ and $\underline{e}_3$.

Notes

(*i*) λ_{ij} are nine numbers defining the change of basis or 'linear transformation'. They are sometimes known as 'direction cosines'.

(i) Since $\underline{e}_i{}'$ are orthonormal

$$\underline{e}_i{}' \cdot \underline{e}_j{}' \;=\; \delta_{ij}\ .$$

Now the l.h.s. of this equation may be written as

$$(\lambda_{ik}\,\underline{e}_k) \cdot (\lambda_{j\ell}\,\underline{e}_\ell) = \lambda_{ik}\lambda_{j\ell}\,(\underline{e}_k \cdot \underline{e}_\ell) = \lambda_{ik}\lambda_{j\ell}\delta_{k\ell} = \lambda_{ik}\,\lambda_{jk}$$

(in the final step we have used the sifting property of $\delta_{k\ell}$) and we deduce

$$\boxed{\lambda_{ik}\lambda_{jk} \;=\; \delta_{ij}}$$

(ii) In order to determine λ_{ij} from the two bases consider

$$\underline{e}_i{}' \cdot \underline{e}_j \;=\; (\lambda_{ik}\,\underline{e}_k) \cdot \underline{e}_j = \lambda_{ik}\,\delta_{kj} = \lambda_{ij}\,.$$

Thus

$$\boxed{\underline{e}_i{}' \cdot \underline{e}_j = \lambda_{ij}}$$

5.2 Inverse relations

Consider expressing the unprimed basis in terms of the primed basis and suppose that

$$\underline{e}_i = \mu_{ij}\,\underline{e}_j{}'.$$

Then $\lambda_{ki} \;=\; \underline{e}_k{}' \cdot \underline{e}_i \;= \mu_{ij}\,(\underline{e}_k{}' \cdot \underline{e}_j{}') \;= \mu_{ij}\,\delta_{kj} = \mu_{ik}$ so that

$$\boxed{\mu_{ij} = \lambda_{ji}}$$

Note that $\underline{e}_i \cdot \underline{e}_j \;=\; \delta_{ij} \;= \lambda_{ki}\,(\underline{e}_k{}' \cdot \underline{e}_j) \;= \lambda_{ki}\lambda_{kj}$ and so we obtain a second relation

$$\boxed{\lambda_{ki}\lambda_{kj} \;=\; \delta_{ij}}\,.$$

5.3 The transformation matrix

We may label the elements of a 3×3 matrix $\underline{\underline{M}}$ as M_{ij}, where i labels the row and j labels the column in which M_{ij} appears:

$$\underline{\underline{M}} \;=\; \begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{pmatrix}.$$

The summation convention can be used to describe matrix multiplication. The ij component of a product of two 3×3 matrices M, N is given by

$$\boxed{(MN)_{ij} = M_{i1}\,N_{1j} + M_{i2}\,N_{2j} + M_{i3}\,N_{3j} = M_{ik}\,N_{kj}}$$

Likewise, recalling the definition of the transpose of a matrix $(M^T)_{ij} = M_{ji}$

$$(M^T N)_{ij} = (M^T)_{ik} N_{kj} = M_{ki} N_{kj}$$

We can thus arrange the numbers λ_{ij} as elements of a square matrix, denoted by λ and known as the **transformation matrix**:

$$\underline{\underline{\lambda}} \;=\; \begin{pmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & \lambda_{22} & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & \lambda_{33} \end{pmatrix}$$

We denote the matrix transpose by $\underline{\underline{\lambda}}^T$ and define it by $(\lambda^T)_{ij} = \lambda_{ji}$ so we see that $\underline{\underline{\mu}} = \underline{\underline{\lambda}}^T$ is the transformation matrix for the inverse transformation.

We also note that δ_{ij} may be thought of as elements of a 3×3 unit matrix:

$$\begin{pmatrix} \delta_{11} & \delta_{12} & \delta_{13} \\ \delta_{21} & \delta_{22} & \delta_{33} \\ \delta_{31} & \delta_{32} & \delta_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \underline{\underline{I}}.$$

i.e. the matrix representation of the Kronecker delta symbol is the **identity matrix** $\underline{\underline{I}}$.

The inverse relation $\lambda_{ik}\lambda_{jk} = \lambda_{ki}\lambda_{kj} = \delta_{ij}$ can be written in matrix notation as

$$\boxed{\underline{\underline{\lambda}}\,\underline{\underline{\lambda}}^T = \underline{\underline{\lambda}}^T\underline{\underline{\lambda}} = \underline{\underline{I}}} \quad , \quad i.e. \quad \boxed{\underline{\underline{\lambda}}^{-1} = \underline{\underline{\lambda}}^T} \; .$$

This is the definition of an **orthogonal matrix** and the transformation (from the $\underline{e}_i$ basis to the $\underline{e}_i{}'$ basis) is called an **orthogonal transformation**.

Now from the properties of determinants, $|\underline{\underline{\lambda}}\,\underline{\underline{\lambda}}^T| = |\underline{\underline{I}}| = 1 = |\lambda|\,|\lambda^T|$ and $|\lambda^T| = |\lambda|$, we have that $|\lambda|^2 = 1$ hence

$$|\lambda| \;=\; \pm 1 \; .$$

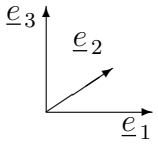
If $\lvert\lambda\rvert = +1$ the orthogonal transformation is said to be 'proper'
If $\lvert\lambda\rvert = -1$ the orthogonal transformation is said to be 'improper'

Handedness of basis

An improper transformation has an unusual effect. In the usual Cartesian basis that we have considered up to now, the basis vectors $\underline{e}_1$, $\underline{e}_2$, and $\underline{e}_3$ form a *right-handed* basis, that is, $\underline{e}_1 \times \underline{e}_2 = \underline{e}_3$, $\underline{e}_2 \times \underline{e}_3 = \underline{e}_1$ and $\underline{e}_3 \times \underline{e}_1 = \underline{e}_2$.

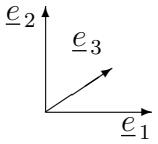
However, we could choose $\underline{e}_1 \times \underline{e}_2 = -\underline{e}_3$, and so on, in which case the basis is said to be *left-handed*.

right handed



$$\underline{e}_3 = \underline{e}_1 \times \underline{e}_2$$
$$\underline{e}_1 = \underline{e}_2 \times \underline{e}_3$$
$$\underline{e}_2 = \underline{e}_3 \times \underline{e}_1$$
$$\boxed{(\underline{e}_1, \underline{e}_2, \underline{e}_3) = 1}$$

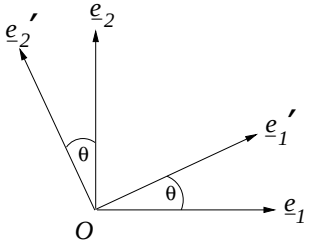
left handed



$$\underline{e}_3 = \underline{e}_2 \times \underline{e}_1$$
$$\underline{e}_1 = \underline{e}_3 \times \underline{e}_2$$
$$\underline{e}_2 = \underline{e}_1 \times \underline{e}_3$$
$$\boxed{(\underline{e}_1, \underline{e}_2, \underline{e}_3) = -1}$$

5.4 Examples of orthogonal transformations

Rotation about the $\underline{e}_3$ axis. We have $\underline{e}_3{}' = \underline{e}_3$ and thus for a rotation through θ,



$$
\begin{aligned}
\underline{e}_3{}' \cdot \underline{e}_1 &= \underline{e}_1{}' \cdot \underline{e}_3 = \underline{e}_3{}' \cdot \underline{e}_2 = \underline{e}_2{}' \cdot \underline{e}_3 = 0\,, \quad \underline{e}_3{}' \cdot \underline{e}_3 = 1 \\
\underline{e}_1{}' \cdot \underline{e}_1 &= \cos\theta \\
\underline{e}_1{}' \cdot \underline{e}_2 &= \cos(\pi/2 - \theta) = \sin\theta \\
\underline{e}_2{}' \cdot \underline{e}_2 &= \cos\theta \\
\underline{e}_2{}' \cdot \underline{e}_1 &= \cos(\pi/2 + \theta) = -\sin\theta
\end{aligned}
$$

Thus

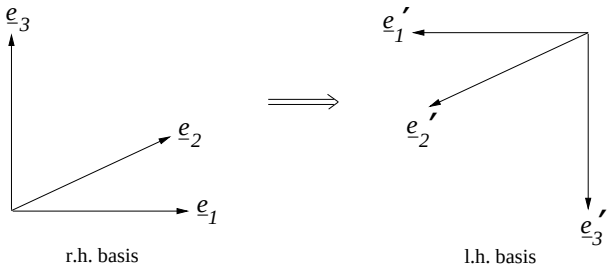
$$
\underline{\underline{\lambda}} = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.
$$

It is easy to check that $\underline{\underline{\lambda}}\,\underline{\underline{\lambda}}^T = \underline{\underline{I}}$. Since $|\lambda| = \cos^2\theta + \sin^2\theta = 1$, this is a *proper* transformation. Note that rotations cannot change the handedness of the basis vectors.

Inversion or Parity transformation. This is defined such that $\underline{e}_i{}' = -\underline{e}_i$.

$$
\text{i.e. } \lambda_{ij} = -\delta_{ij} \qquad \text{or} \qquad \lambda = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} = -\underline{\underline{I}}\,.
$$

Clearly $\underline{\underline{\lambda}}\,\underline{\underline{\lambda}}^T = \underline{\underline{I}}$. Since $|\lambda| = -1$, this is an *improper* transformation. Note that the handedness of the basis is reversed: $\underline{e}_1{}' \times \underline{e}_2{}' = -\underline{e}_3{}'$



Reflection. Consider reflection of the axes in $\underline{e}_2$–$\underline{e}_3$ plane so that $\underline{e}_1{}' = -\underline{e}_1$, $\underline{e}_2{}' = \underline{e}_2$ and $\underline{e}_3{}' = \underline{e}_3$. The transformation matrix is

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

Since $|\lambda| = -1$, this is an *improper* transformation. Again the handedness of the basis changes.

5.5 Products of transformations

Consider a transformation $\underline{\underline{\lambda}}$ to the basis $\{\underline{e}_i{}'\}$ followed by a transformation $\underline{\underline{\rho}}$ to another basis $\{\underline{e}_i{}''\}$

$$
\underline{e}_i \overset{\lambda}{\Longrightarrow} \underline{e}_i{}' \overset{\rho}{\Longrightarrow} \underline{e}_i{}''
$$

Clearly there must be an orthogonal transformation $\underline{e}_i \overset{\xi}{\Longrightarrow} \underline{e}_i{}''$

Now

$$
\underline{e}_i{}'' = \rho_{ij}\underline{e}_j{}' = \rho_{ij}\lambda_{jk}\underline{e}_k = (\rho\lambda)_{ik}\underline{e}_k \quad \text{so} \quad \boxed{\underline{\underline{\xi}} = \underline{\underline{\rho}}\,\underline{\underline{\lambda}}}
$$

Note the order of the product: the matrix corresponding to the first change of basis stands to the right of that for the second change of basis. In general, transformations do not commute so that $\underline{\underline{\rho}}\,\underline{\underline{\lambda}} \neq \underline{\underline{\lambda}}\,\underline{\underline{\rho}}$. Only the inversion and identity transformations commute with all transformations.

Example: Rotation of θ about $\underline{e}_3$ then reflection in yz plane

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -\cos\theta & -\sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

whereas, if we reverse the order,

$$\begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -\cos\theta & \sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

5.6 Improper transformations

We may write any improper transformation $\underline{\underline{\xi}}$ (for which $|\xi| = -1$) as $\underline{\underline{\xi}} = \left(-\underline{\underline{I}}\right) \underline{\underline{\lambda}}$ where $\underline{\underline{\lambda}} = -\underline{\underline{\xi}}$ and $|\lambda| = +1$ Thus an improper transformation can always be expressed as a proper transformation followed by an inversion.

e.g. consider $\underline{\underline{\xi}}$ for a reflection in the $1-3$ plane which may be written as

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

Identifying $\underline{\underline{\lambda}}$ from $\underline{\underline{\xi}} = \left(-\underline{\underline{I}}\right) \underline{\underline{\lambda}}$ we see that λ is a rotation of π about $\underline{e}_2$.

Lecture 6: Transformation properties of vectors and scalars

6.1 Transformation of vector components

Let $\underline{A}$ be any vector, with components A_i in the basis $\{\underline{e}_i\}$ and A'_i in the basis $\{\underline{e}_i{}'\}$ *i.e.*

$$\boxed{\underline{A} = A_i\,\underline{e}_i = A'_i\,\underline{e}_i{}'}\ .$$

The components are related as follows, taking care with dummy indices:

$$A'_i = \underline{A} \cdot \underline{e}_i{}' = (A_j \underline{e}_j) \cdot \underline{e}_i{}' = (\underline{e}_i{}' \cdot \underline{e}_j) A_j = \lambda_{ij} A_j$$

$$\boxed{A'_i = \lambda_{ij} A_j}$$

So the new components are related to the old ones by the same matrix transformation as applies to the basis vectors. The inverse transformation works in a similar way:

$$A_i \;=\; \underline{A} \cdot \underline{e}_i \;=\; (A'_k \underline{e}_k{}') \cdot \underline{e}_i = \lambda_{ki} A'_k \;= (\lambda^T)_{ik} A'_k .$$

Note carefully that we do *not* put a prime on the vector itself – there is only one vector, $\underline{A}$, in the above discussion.

However, the *components* of this vector are different in different bases, and so are denoted by A_i in the basis $\{\underline{e}_i\}$, A'_i in the basis $\{\underline{e}_i{}'\}$, etc.

In matrix form we can write these relations as

$$\boxed{\begin{pmatrix} A'_1 \\ A'_2 \\ A'_3 \end{pmatrix} = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & \lambda_{22} & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & \lambda_{33} \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} = \underline{\underline{\lambda}} \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix}}$$

Example: Consider a rotation of the axes about $\underline{e}_3$

$$\begin{pmatrix} A'_1 \\ A'_2 \\ A'_3 \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} = \begin{pmatrix} \cos\theta\, A_1 + \sin\theta\, A_2 \\ \cos\theta\, A_2 - \sin\theta\, A_1 \\ A_3 \end{pmatrix}$$

A direct check of this using trigonometric considerations is significantly harder!

6.2 The Transformation of the scalar product

Let $\underline{A}$ and $\underline{B}$ be vectors with components A_i and B_i in the basis $\{\underline{e}_i\}$ and components A'_i and B'_i in the basis $\{\underline{e}_i{}'\}$. In the basis $\{\underline{e}_i\}$, the scalar product, denoted by $(\underline{A} \cdot \underline{B})$, is

$$(\underline{A} \cdot \underline{B}) \;= A_i B_i \,.$$

In the basis $\{\underline{e}_i{}'\}$, we denote the scalar product by $(\underline{A} \cdot \underline{B})'$, and we have

$$\begin{aligned} (\underline{A} \cdot \underline{B})' \;&=\; A'_i B'_i \;= \lambda_{ij} A_j \lambda_{ik} B_k \;= \delta_{jk} A_j B_k \\ &=\; A_j B_j = (\underline{A} \cdot \underline{B}) \,. \end{aligned}$$

Thus the scalar product is the same evaluated in any basis. This is of course expected from the geometrical definition of scalar product which is independent of basis. We say that the scalar product is *invariant* under a change of basis.

Summary We have now obtained an algebraic definition of scalar and vector quantities. Under the orthogonal transformation from the basis $\{\underline{e}_i\}$ to the basis $\{\underline{e}_i{}'\}$, defined by the transformation matrix $\lambda : \underline{e}_i{}' = \lambda_{ij}\, \underline{e}_j$, we have that

- A **scalar** is a single number ϕ which is invariant:

$$\boxed{\phi' \;=\; \phi}\;.$$

Of course, not all scalar quantities in physics are expressible as the scalar product of two vectors *e.g.* mass, temperature.

- A **vector** is an 'ordered triple' of numbers A_i which transforms to A'_i :

$$\boxed{A'_i \;= \lambda_{ij} A_j}\;.$$

6.3 Transformation of the vector product

Improper transformations have an interesting effect on the vector product. Consider the case of inversion, so that $\underline{e}_i{}' = -\underline{e}_i$, and all the signs of vector components are flipped: $A'_i = -A_i$ etc.

If we now calculate the vector product $\underline{C} = \underline{A} \times \underline{B}$ in the new basis using the determinant formula, we obtain

$$\begin{vmatrix} \underline{e}_1{}' & \underline{e}_2{}' & \underline{e}_3{}' \\ A'_1 & A'_2 & A'_3 \\ B'_1 & B'_2 & B'_3 \end{vmatrix} = (-)^3 \begin{vmatrix} \underline{e}_1 & \underline{e}_2 & \underline{e}_3 \\ A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \end{vmatrix} = - \begin{vmatrix} \underline{e}_1 & \underline{e}_2 & \underline{e}_3 \\ A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \end{vmatrix}$$

which is $-\underline{C}$ as calculated in the original basis!

This shouldn't be surprising, since we defined the vector product direction in the first place assuming a right-hand screw rule, and the determinant expression was derived assuming that the basis was also right-handed. If we use this same formula in a left-handed basis, the direction of the vector product produced by the formula is reversed compared to our original definition.

The real solution to this is to stick to the geometrical definition, in which case the vector product is always a well-defined entity (the notion of a right-handed screw is not changed just because we use a left-handed basis).

But if we regard the simple determinant form as the *definition* of the vector product (which is frequently what is assumed), then we have to live with the possibility of $\underline{A} \times \underline{B}$ flipping sign. It is then an example of what is known as a 'pseudovector' or an 'axial vector' which has the transformation law

$$\boxed{C'_i = (\det \lambda)\lambda_{ij} C_j}$$

so that it behaves just like a vector under proper transformations, for which $\det \lambda = +1$, but picks up an extra minus sign under improper transformations, for which $\det \lambda = -1$.

6.4 Summary of story so far

We take the opportunity to summarise some key points of what we have done so far. N.B. this is NOT a list of everything you need to know.

Key points from geometrical approach

You should recognise on sight that

$$\begin{aligned} \underline{r} \times \underline{b} = \underline{c} \quad & \text{is a line } (\underline{r} \text{ lies on a line}) \\ \underline{r} \cdot \underline{a} = d \quad & \text{is a plane } (\underline{r} \text{ lies in a plane}) \end{aligned}$$

Useful properties of scalar and vector products to remember

$$\begin{aligned} \underline{a} \cdot \underline{b} = 0 \quad &\Leftrightarrow \quad \text{vectors orthogonal} \\ \underline{a} \times \underline{b} = 0 \quad &\Leftrightarrow \quad \text{vectors collinear} \\ \underline{a} \cdot (\underline{b} \times \underline{c}) = 0 \quad &\Leftrightarrow \quad \text{vectors co-planar or linearly dependent} \\ \underline{a} \times (\underline{b} \times \underline{c}) \quad &= \quad \underline{b}(\underline{a} \cdot \underline{c}) - \underline{c}(\underline{a} \cdot \underline{b}) \end{aligned}$$

Key points of suffix notation

We label orthonormal basis vectors $\underline{e}_1, \underline{e}_2, \underline{e}_3$ and write the expansion of a vector $\underline{A}$ as

$$\underline{A} = \sum_{i=1}^{3} A_i \underline{e}_i$$

The Kronecker delta δ_{ij} can be used to express the orthonormality of the basis

$$\underline{e}_i \cdot \underline{e}_j = \delta_{ij}$$

The Kronecker delta has a very useful sifting property

$$\sum_j [\cdots]_j \delta_{jk} = [\cdots]_k$$

$(\underline{e}_1, \underline{e}_2, \underline{e}_3) = \pm 1$ determines whether the basis is right- or left-handed

Key points of summation convention

Using the summation convention we have for example

$$\underline{A} = A_i \underline{e}_i$$

and the sifting property of δ_{ij} becomes

$$[\cdots]_j \delta_{jk} = [\cdots]_k$$

We introduce ϵ_{ijk} to enable us to write the vector products of basis vectors in a r.h. basis in a uniform way

$$\underline{e}_i \times \underline{e}_j = \epsilon_{ijk} \underline{e}_k$$

.

The vector products and scalar triple products in a r.h. basis are

$$\underline{A} \times \underline{B} = \begin{vmatrix} \underline{e}_1 & \underline{e}_2 & \underline{e}_3 \\ A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \end{vmatrix} \quad \text{or equivalently} \quad (\underline{A} \times \underline{B})_i \;=\; \epsilon_{ijk} A_j B_k$$

$$\underline{A} \cdot (\underline{B} \times \underline{C}) = \begin{vmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{vmatrix} \quad \text{or equivalently} \quad \underline{A} \cdot (\underline{B} \times \underline{C}) \;=\; \epsilon_{ijk} A_i B_j C_k$$

Key points of change of basis

The new basis is written in terms of the old through

$$\underline{e}_i{}' = \lambda_{ij}\underline{e}_j \quad \text{where } \lambda_{ij} \text{ are elements of a } 3\times 3 \text{ transformation matrix } \underline{\underline{\lambda}}$$

$\underline{\underline{\lambda}}$ is an orthogonal matrix, the defining property of which is $\underline{\underline{\lambda}}^{-1} = \underline{\underline{\lambda}}^T$ and this can be written as

$$\underline{\underline{\lambda}}\,\underline{\underline{\lambda}}^T = \underline{\underline{I}} \quad \text{or} \quad \lambda_{ik}\lambda_{jk} = \delta_{ij}$$

$|\lambda| = \pm 1$ decides whether the transformation is proper or improper i.e. whether the handedness of the basis is changed.

Key points of algebraic approach

A **scalar** is defined as a number that is invariant under an orthogonal transformation

A **vector** is defined as an object $\underline{A}$ represented in a basis by numbers A_i which transform to A'_i through

$$A'_i = \lambda_{ij}A_j.$$

or in matrix form

$$\begin{pmatrix} A'_1 \\ A'_2 \\ A'_3 \end{pmatrix} = \underline{\underline{\lambda}} \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix}$$

Lecture 7: Vector spaces

The whole apparatus of vector algebra has some general aspects that apply in other contexts – but where the physical interpretation is very different. To stress this point, mathematicians like to develop the abstract idea of a vector space. We now give an informal summary of the main elements of this structure, and illustrate it with a couple of key applications.

7.1 Vector space definition

The following definition of a vector space captures all the main elements we have been using so far. This list is not for examination, but it is interesting to see how many things need to be specified, including items that might have been thought rather obvious. A vector space V is a collection of items (vectors) with an addition operation, and the ability to multiply by scalars.

1. Closure under addition: $\underline{A} + \underline{B} \in V$ if $\underline{A}, \underline{B} \in V$.
2. Closure under multiplication: $\alpha\underline{A} \in V$ if $\underline{A} \in V$.
3. Neutral element for addition: $\underline{A} + \underline{0} = \underline{A}$.
4. Neutral element for multiplication: $1\underline{A} = \underline{A}$.

5. Inverse element for addition: $\underline{A} + (-\underline{A}) = \underline{0}$.
6. Commutivity: $\underline{A} + \underline{B} = \underline{B} + \underline{A}$.
7. Associativity: $\underline{A} + (\underline{B} + \underline{C}) = (\underline{A} + \underline{B}) + V_C$.
8. Consistency: $\alpha(\beta\underline{A}) = \alpha\beta(\underline{A})$.
9. Distributivity type 1: $\alpha(\underline{A} + \underline{B}) = \alpha\underline{A} + \alpha\underline{B}$.
10. Distributivity type 2: $(\alpha + \beta)\underline{A} = \alpha\underline{A} + \beta\underline{A}$.

7.2 Linear independence and dimensionality

This impressively long formal list doesn't add anything to what we know already, but one more detailed aspect is worth emphasising, and this is the idea of how to specify the **dimensionality** of the vector space.

Consider two vectors $\underline{A}$ and $\underline{B}$ that satisfy the equation $\alpha\underline{A} + \beta\underline{B} = 0$. If this is satisfied for *non-zero* α and β then we can solve the equation to find $\underline{B} = -\frac{\alpha}{\beta}\underline{A}$. Clearly $\underline{A}$ and $\underline{B}$ are **collinear** (either parallel or anti-parallel), and then $\underline{A}$ and $\underline{B}$ are said to be **linearly dependent**. If this equation can be satisfied *only* for $\alpha = \beta = 0$, then $\underline{A}$ and $\underline{B}$ are **linearly independent**, and no λ can be found such that $\underline{B} = \lambda\underline{A}$).

Similarly, in 3 dimensions three vectors are linearly dependent if we can find non-trivial α, β, γ (i.e. not all zero) such that

$$\boxed{\alpha\underline{A} + \beta\underline{B} + \gamma\underline{C} = 0},$$

otherwise $\underline{A}, \underline{B}, \underline{C}$ are linearly independent (no one is a linear combination of the other two).

The geometrical interpretation of linear dependence in three dimensions is that

$$\boxed{\text{three linearly dependent vectors} \Leftrightarrow \text{three coplanar vectors}}$$

To see this note that if $\alpha\underline{A} + \beta\underline{B} + \gamma\underline{C} = 0$ then

$$\begin{array}{ll} \alpha \neq 0 & \alpha\underline{A} \cdot (\underline{B} \times \underline{C}) = 0 \Rightarrow \underline{A}, \underline{B}, \underline{C} \quad \text{are coplanar} \\ \alpha = 0 & \text{then } \underline{B} \text{ is collinear with } \underline{C} \text{ and } \underline{A}, \underline{B}, \underline{C} \text{ are coplanar} \end{array}$$

The dimensionality of a space is then defined in these terms as follows: *For a space of dimension n one can find at most n linearly independent vectors.*

7.3 Skew basis

A basis in which the basis vectors are *orthogonal* and *normalised* (of unit length) is called an **orthonormal** basis. But as defined above, basis vectors need satisfy neither of these

criteria. Even if the basis vectors are unit (which is a simple matter of definition), they need not be orthogonal – in which case we have a **skew basis**. How do we define components in such a case? It turns out that there are two different ways of achieving this.

We can express the vector as a linear superposition:

$$\underline{V} = \sum_i V_i^{(1)} \underline{e}_i,$$

where $\underline{e}_i$ are the basis vectors. But we are used to extracting components by taking the dot product, so we might equally well want to define a second kind of component by

$$V_i^{(2)} = \underline{V} \cdot \underline{e}_i.$$

These numbers are not the same, as we see by inserting the first definition into the second:

$$V_i^{(2)} = \left(\sum_j V_j^{(1)} \underline{e}_j \right) \cdot \underline{e}_i.$$

This cannot be simplified further if we lack the usual orthonormal basis $\underline{e}_i \cdot \underline{e}_j = \delta_{ij}$, in which case a given type-2 component is a mixture of all the different type-1 components.

For (non-examinable) interest, the two types are named respectively **contravariant** and **covariant** components. It isn't possible to say that one of these definitions is better than another: we sometimes want to use both types of component, as with the modulus-squared of a vector:

$$V^2 = \underline{V} \cdot \underline{V} = \underline{V} \cdot \left(\sum_i V_i^{(1)} \underline{e}_i \right) = \sum_i V_i^{(1)} V_i^{(2)}.$$

This looks like the usual rule for the dot product, but both kinds of components are needed.

7.4 A generalization: Complete sets of functions

A less familiar example of a vector space arises in the case of functions. Consider a Taylor series about the origin (a Maclaurin series), whereby some function $f(x)$ (note that x is not a vector) is expressed as a power series:

$$f(x) = a + bx + cx^2 + dx^3 + \cdots$$

This looks a bit similar to the expansion of a vector

$$\underline{A} = A_1 \underline{e}_1 + A_2 \underline{e}_2 + A_3 \underline{e}_3 + \cdots,$$

with the powers of x playing the role of the basis vectors. Is there anything real to this resemblance? There certainly is in practice: just as, when two vectors are equal, we say all the components separately must be equal, so we commonly equate coefficients of different powers of x in expansions of equations like $f(x) = g(x)$. To prove that, if two vectors are equal, all their components are separately equal, we would use the existence of an orthonormal basis. Write $\underline{A} = \underline{B}$ in components:

$$A_i \underline{e}_i = B_j \underline{e}_j$$

(using the summation convention) and take the dot product with $\underline{e}_k$. All terms in the sums vanish via orthogonality except the k ones, so we get $A_k = B_k$.

Inner product of functions

The corresponding idea when dealing with functions is to define orthonormality in terms of integration over some domain of size L. If we have a set of functions of x, $f_i(x)$, where the i label distinguishes different functions, then the f_i are mutually orthonormal if

$$\boxed{\langle f_i \mid f_j \rangle \equiv \int_{-L/2}^{+L/2} f_i^*(x) f_j(x)\, dx = \delta_{ij}.}$$

This integral over a product of the f's defines the **inner product**, which is the general version of the scalar product. Note the complex conjugate: this is irrelevant for real functions, but as you know there are examples (quantum-mechanical wavefunctions) where things are not necessarily real.

Simple powers of x don't satisfy orthonormality, but they can be recast so that they do. Start with a constant for f_0; take $f_1(x) \propto x$, which is clearly orthogonal to f_0; f_2 must be a polynomial of one order higher:

$$\begin{aligned} f_0(x) &= a \\ f_1(x) &= bx \\ f_2(x) &= c + dx + ex^2 \end{aligned}$$

The three constants in f_2 arise from orthonormality. Consider the orthogonality relations:

$$\int_{-L/2}^{+L/2} a(c + dx + ex^2)\, dx = \int_{-L/2}^{+L/2} bx(c + dx + ex^2)\, dx = 0.$$

From the second of these, it is easy to see that $d = 0$; the first relation gives $cL + eL^3/12 = 0$, so $f_2(x) \propto 1 - 12x^2/L^2$. The coefficients of proportionality, if needed, come from normalization via the integral of the square of the f's: $a = 1/\sqrt{L}$ etc. This process can be extended indefinitely, generating a set of polynomial functions called the **Legendre polynomials**.

Completeness

The Legendre polynomials are one example of a **complete set** of functions, by which we mean that any function $g(x)$ can be represented as a linear superposition of orthonormal **basis functions**:

$$g(x) = \sum_i a_i\, f_i(x).$$

The analogy with vectors should be clear: the coefficients a_i can be written as a row or column vector, and any coefficient of interest can be projected out using the inner product:

$$a_i = \langle f_i \mid g \rangle.$$

The important step here is, of course, how you know that a given set of functions is complete; this is something that will have to be left for later years.

Fourier series

The most important complete set in physics is where the basis functions are oscillatory:

$$f_n(x) \propto \exp(ik_n x)$$

(it should be clear that the i that means $\sqrt{-1}$ is not the same as the index i in k_i, but we have written k_n just to be safe). In order to have these waves orthogonal over the interval L, we must have a whole number of wavelengths spanning the interval: $\lambda_n = L/n$, where n is an integer. Since the **wavenumber** k is $2\pi/\lambda$, that gives

$$k_n = n\,\frac{2\pi}{L}$$

as the allowed wavenumbers. As well as allowing orthonormality over the range $-L/2$ to $+L/2$, this condition means that we are also restricting ourselves to functions that are periodic, repeating on the scale L.

The use of the complex exponential may seem surprising, especially since we have been expanding real functions so far. There are two attitudes to this. One is to remember that $\exp i\theta = \cos\theta + i\sin\theta$, so we are really saying real functions can be written using sin and cos waves:

$$f(x) = \sum_n a_n \sin(k_n x) + \sum_n b_n \cos(k_n x).$$

Alternatively, in terms of the complex expansion, we want any imaginary terms to disappear. So, if we write

$$f(x) = \sum_n c_n \exp(ik_n x),$$

then $f(x)$ is guaranteed to be real if the coefficients c_n are Hermitian: the coefficient c associated with wavenumber k are the complex conjugate of those associated with $-k$.

Fourier analysis will be very important in subsequent years.

Lecture 8: Tensors

Physical relations between vectors

The simplest physical laws are expressed in terms of scalar quantities that are independent of our choice of basis *e.g.* the gas law $P = nkT$ relating three scalar quantities (pressure, number density and temperature), which will in general all vary with position.

At the next level of complexity are laws relating vector quantities, such as $\underline{F} = m\underline{a}$:

(i) These laws take the form *vector* = *scalar* × *vector*

(ii) They relate two vectors in the *same* direction

If we consider Newton's Law, for instance, then in a particular Cartesian basis $\{\underline{e}_i\}$, $\underline{a}$ is represented by its components $\{a_i\}$ and $\underline{F}$ by its components $\{F_i\}$ and we can write

$$F_i \;=\; m\,a_i$$

In another such basis $\{\underline{e}_i{}'\}$

$$F_i' \;=\; m\,a_i'$$

where the set of numbers, $\{a_i'\}$, is in general different from the set $\{a_i\}$. Likewise, the set $\{F_i'\}$ differs from the set $\{F_i\}$, but of course

$$a_i' \;=\; \lambda_{ij} a_j \quad \text{and} \quad F_i' \;=\; \lambda_{ij} F_j$$

Thus we can think of $\underline{F} = m\underline{a}$ as representing an infinite set of relations between measured components in various bases. Because all vectors transform the same way under orthogonal transformations, the relations have the *same form* in all bases. We say that Newton's Law, expressed in component form, is **form invariant** or **covariant**.

This is why our proof of the 'BAC-CAB' rule using a special coordinate system wasn't a cheat. Rather, it uses the principle of **manifest covariance**: if we write down a candidate relation between two vectors $\underline{A} = \underline{B}$, we only need to show that it holds for one basis. The case of changing basis involves a transformation matrix, which acts in the same way on each side of the equation; so if two vectors are equal in one basis, they are always equal.

8.1 Examples of more complicated laws

Physical laws often relate two vectors, but in general these may point in different directions. We then have the case where there is a linear relation between the various components of the vectors, and there are many physical examples of this.

Ohm's law in an anisotropic medium

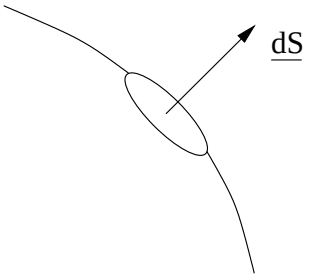
The vector form of Ohm's Law says that an applied electric field $\underline{E}$ produces a current density $\underline{J}$ (current per unit area) in the same direction: $\underline{J} = \sigma \underline{E}$, where σ is the conductivity (to see that this is the familiar $V = RI$, consider a tube of cross-sectional area A and length L: $I = JA$, $V = EL$ and $R = L/(\sigma A)$). This only holds for conducting media that are isotropic, i.e. the same in all directions. This is certainly not the case in crystalline media, where the regular lattice will favour conduction in some directions more than in others.

The most general relation between $\underline{J}$ and $\underline{E}$ which is linear and is such that $\underline{J}$ vanishes when $\underline{E}$ vanishes is of the form

$$J_i = G_{ij} E_j$$

where G_{ij} are the components of the *conductivity tensor* in the chosen basis, and characterise the conduction properties when $\underline{J}$ and $\underline{E}$ are measured in that basis. Thus we need nine numbers, G_{ij}, to characterise the conductivity of an anisotropic medium.

Stress tensor

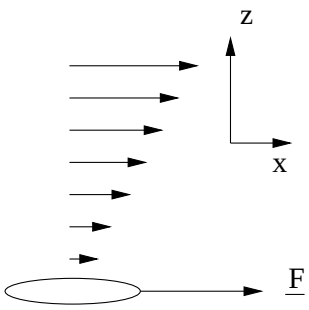


Consider a surface acted on by the pressure of a fluid. The force on an area element $\underline{dS}$ is $\underline{F} = -P\underline{dS}$ for isotropic pressure (a minus sign because the force acts *into* the surface).

In general, we have

$$F_i = s_{ij}\, dS_j,$$

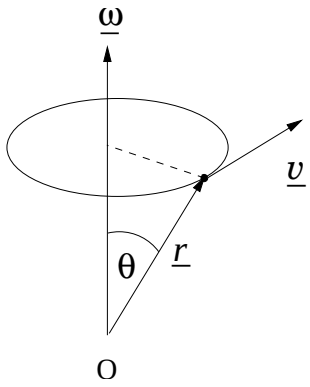
where s_{ij} are the components of the **stress tensor**. Thus, where we deal only with pressure, $s_{ij} = -P\delta_{ij}$ and the stress tensor is diagonal.



The most important example of anisotropic stress is in a viscous shear flow. Suppose a fluid moves in the x direction, but v_x changes in the z direction. The force per unit area acting in the x direction on a surface in the xy plane is $\eta\, dv_x/dz$, where η is the coefficient of viscosity. In this case, the only non-zero component of the stress tensor is

$$s_{13} = \eta\, dv_x/dz.$$

Rotation and the inertia tensor



Consider a particle of mass m at a point $\underline{r}$ in a rigid body rotating with **angular velocity** $\underline{\omega}$. Recall that $\underline{v} = \underline{\omega} \times \underline{r}$. You were asked to check that this gives the right direction for $\underline{v}$; that it is perpendicular to the plane of $\underline{\omega}$ and $\underline{r}$; that the magnitude $|\underline{v}| = \omega r \sin\theta =$ $\omega\times$ radius of circle in which the point is travelling

Now consider the **angular momentum** of the particle about the origin O, defined by $\underline{L} = \underline{r} \times \underline{p} = \underline{r} \times (m\underline{v})$ where m is the mass of the particle. Using the identity for the vector triple product, we obtain

$$\underline{L} = m\underline{r} \times (\underline{\omega} \times \underline{r}) = m\left[\underline{\omega}(\underline{r}\cdot\underline{r}) - \underline{r}(\underline{r}\cdot\underline{\omega})\right].$$

Note that only if $\underline{r}$ is perpendicular to $\underline{\omega}$ do we obtain $\underline{L} = mr^2\underline{\omega}$, which means that only then are $\underline{L}$ and $\underline{\omega}$ in the same direction.

Taking components of this equation in an orthonormal basis $\{\underline{e}_i\}$, we find that

$$\begin{aligned} L_i &= m\left[\omega_i(\underline{r}\cdot\underline{r}) - x_i(\underline{r}\cdot\underline{\omega})\right] \\ &= m\left[r^2\omega_i - x_i x_j\omega_j\right] \quad \text{noting that} \quad \underline{r}\cdot\underline{\omega} = x_j\omega_j \\ &= m\left[r^2\delta_{ij} - x_i x_j\right]\omega_j \quad \text{using} \quad \omega_i = \delta_{ij}\omega_j \end{aligned}$$

Thus

$$\boxed{L_i = I_{ij}(O)\,\omega_j \qquad \text{where} \qquad I_{ij}(O) = m\left[r^2\,\delta_{ij} - x_i x_j\right]}$$

$I_{ij}(O)$ are the components of the **inertia tensor**, relative to O, in the $\underline{e}_i$ basis.

Note that there is a potential confusion here, since $\underline{\underline{I}}$ is often used to mean the identity matrix. Normally it is clear from context what is intended, but you can always use the alternative notation $\underline{\underline{\delta}}$ for the identity, since we have seen that its components are δ_{ij}.

8.2 Rank of tensors

The set of nine numbers, T_{ij}, representing a tensor of the above sort can be written as a 3×3 array, which are the components of a matrix:

$$\underline{\underline{T}} = \begin{pmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{pmatrix}$$

Because it has 2 indices, this would be called a **second rank tensor**.

Scalars and vectors are called tensors of rank zero and one respectively, where *rank = no. of indices in a Cartesian basis*. We can also define tensors of rank greater than two. Our friend ϵ_{ijk} is a tensor of rank three, whereas δ_{ij} is another tensor of rank two.

8.3 Transformation properties of tensors

Suppose we consider an orthogonal transformation of basis. Simply changing basis cannot alter the form of a physical law, which must be valid in any basis. Therefore, if our relation reads $A_i = T_{ij}B_j$, we must have

$$A'_i = T'_{ij}B'_j \quad \text{where} \quad A'_i = \lambda_{ij}A_j \quad \text{and} \quad B'_j = \lambda_{jk}B_k$$

Thus we deduce that

$$\lambda_{ij}A_j = \lambda_{ij}T_{jk}B_k = T'_{ij}\lambda_{jk}B_k$$

which we can rewrite as

$$(T'_{ij}\lambda_{jk} - \lambda_{ij}T_{jk})B_k = 0$$

This must be true for arbitrary vector $\underline{B}$ and hence $T'_{ij}\lambda_{jk} = \lambda_{ij}T_{jk}$. In matrix notation, this is just $\underline{\underline{T}}'\,\underline{\underline{\lambda}} = \underline{\underline{\lambda}}\,\underline{\underline{T}}$. If we multiply on the right by $\underline{\underline{\lambda}}^T = \underline{\underline{\lambda}}^{-1}$, this gives the general law for transforming the components of a second rank tensor:

$$\boxed{\underline{\underline{T}}' = \underline{\underline{\lambda}}\,\underline{\underline{T}}\,\underline{\underline{\lambda}}^T}$$

In terms of components, this is written as

$$T'_{ij} = \lambda_{ik}\,T_{k\ell}\,(\lambda^T)_{\ell j} = \lambda_{ik}\lambda_{j\ell}\,T_{k\ell}.$$

Note that we get one instance of λ for each tensor index, according to the rule that applies for vectors. This applies for tensors of any rank.

Notes

(i) It is not quite right to say that a second rank tensor is a matrix: the tensor is the fundamental object and is *represented in a given basis* by a matrix.

(ii) Nevertheless, it is reasonable to say informally that a tensor is a **physical matrix**: a matrix that occurs in a physics equation relating two vectors, which must hold in any basis.

(iii) But not all matrices are tensors: *e.g.* the transformation matrix λ is not a tensor but nine numbers defining the transformation between *two different bases*.

(iv) Most tensors change components when the basis is changed, but some do not: δ_{ij} and ϵ_{ijk} are **isotropic tensors** whose components are always the same (easy to prove for δ_{ij}, which is just the identity matrix).

Lecture 9: More on tensors

9.1 Invariants

Trace of a tensor: the trace of a tensor is defined as the sum of the diagonal elements T_{ii}. Consider the trace of the matrix representing the tensor in the transformed basis

$$\begin{aligned} T'_{ii} &= \lambda_{ir}\lambda_{is}T_{rs} \\ &= \delta_{rs}T_{rs} = T_{rr} \end{aligned}$$

Thus the trace is the same, evaluated in any basis and is a scalar invariant.

Determinant: it can be shown that the determinant is also an invariant.

Symmetry of a tensor: if the matrix T_{ij} representing the tensor is *symmetric* then

$$T_{ij} = T_{ji}$$

Under a change of basis

$$\begin{aligned} T'_{ij} &= \lambda_{ir}\lambda_{js}T_{rs} \\ &= \lambda_{ir}\lambda_{js}T_{sr} \quad \text{using symmetry} \\ &= \lambda_{is}\lambda_{jr}T_{rs} \quad \text{relabelling} \\ &= T'_{ji} \end{aligned}$$

Therefore a symmetric tensor remains symmetric under a change of basis. Similarly (exercise) an antisymmetric tensor $T_{ij} = -T_{ji}$ remains antisymmetric.

In fact one can decompose an arbitrary second rank tensor T_{ij} into a symmetric part S_{ij} and an antisymmetric part A_{ij} through

$$\boxed{S_{ij} = \frac{1}{2}\left[T_{ij} + T_{ji}\right] \quad A_{ij} = \frac{1}{2}\left[T_{ij} - T_{ji}\right]}$$

9.2 The inertia tensor

We saw earlier that for a single particle of mass m, located at position $\underline{r}$ with respect to an origin O on the axis of rotation of a rigid body

$$L_i = I_{ij}(O)\,\omega_j \qquad \text{where} \qquad I_{ij}(O) = m\left\{r^2\,\delta_{ij} - x_i x_j\right\}$$

where $I_{ij}(O)$ are the components of the inertia tensor, relative to O, in the basis $\{\underline{e}_i\}$.

For a **collection of N particles** of mass m^α at $\underline{r}^\alpha$, where $\alpha = 1 \ldots N$,

$$\boxed{I_{ij}(O) = \sum_{\alpha=1}^{N} m^\alpha \left\{(\underline{r}^\alpha \cdot \underline{r}^\alpha)\,\delta_{ij} - x_i^\alpha x_j^\alpha\right\}} \tag{1}$$

For a **continuous body**, the sums become integrals, giving

$$\boxed{I_{ij}(O) \;=\; \int_V \rho(\underline{r})\left\{(\underline{r}\cdot\underline{r})\,\delta_{ij} - x_i x_j\right\}\,dV}\;.$$

Here, $\rho(\underline{r})$ is the density at position $\underline{r}$. $\rho(\underline{r})\,dV$ is the mass of the volume element dV at $\underline{r}$. For laminae (flat objects) and solid bodies, these are 2- and 3-dimensional integrals respectively.

If the basis is *fixed relative to the body*, the $I_{ij}(O)$ are **constants** in time.

Consider the **diagonal** term

$$\begin{aligned} I_{11}(O) &= \sum_\alpha m^\alpha \left\{(\underline{r}^\alpha\cdot\underline{r}^\alpha) - (x_1^\alpha)^2\right\} \\ &= \sum_\alpha m^\alpha \left\{(x_2^\alpha)^2 + (x_3^\alpha)^2\right\} \\ &= \sum_\alpha m^\alpha\,(r_\perp^\alpha)^2\,, \end{aligned}$$

where $r_\perp^\alpha$ is the **perpendicular distance** of m^α from the $\underline{e}_1$ axis through O.

This term is called the **moment of inertia** about the $\underline{e}_1$ axis. It is simply the mass of each particle in the body, multiplied by the square of its distance from the $\underline{e}_1$ axis, summed over all of the particles. Similarly the other diagonal terms are moments of inertia.

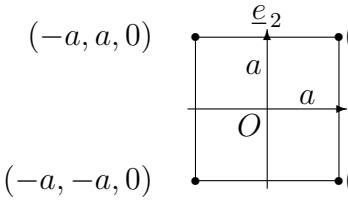
The **off-diagonal** terms are called the **products of inertia**, having the form, for example

$$I_{12}(O) \;=\; -\sum_\alpha m^\alpha\, x_1^\alpha\, x_2^\alpha\,.$$

Example

Consider 4 masses m at the vertices of a square of side $2a$.

(i) O **at centre of the square.**



For $m^{(1)} = m$ at $(a,a,0)$, $\underline{r}^{(1)} = a\underline{e}_1 + a\underline{e}_2$, so $\underline{r}^{(1)}\cdot\underline{r}^{(1)} = 2a^2$, $x_1^{(1)} = a$, $x_2^{(1)} = a$ and $x_3^{(1)} = 0$

$$\underline{\underline{I}}(O) \;=\; m\left\{2a^2\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix} - a^2\begin{pmatrix}1&1&0\\1&1&0\\0&0&0\end{pmatrix}\right\} \;=\; ma^2\begin{pmatrix}1&-1&0\\-1&1&0\\0&0&2\end{pmatrix}\,.$$

For $m^{(2)} = m$ at $(a,-a,0)$, $\underline{r}^{(2)} = a\underline{e}_1 - a\underline{e}_2$, so $\underline{r}^{(2)}\cdot\underline{r}^{(2)} = 2a^2$, $x_1^{(2)} = a$ and $x_2^{(2)} = -a$

$$\underline{\underline{I}}(O) \;=\; m\left\{2a^2\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix} - a^2\begin{pmatrix}1&-1&0\\-1&1&0\\0&0&0\end{pmatrix}\right\} \;=\; ma^2\begin{pmatrix}1&1&0\\1&1&0\\0&0&2\end{pmatrix}\,.$$

For $m^{(3)} = m$ at $(-a,-a,0)$, $\underline{r}^{(3)} = -a\underline{e}_1 - a\underline{e}_2$, so $\underline{r}^{(3)}\cdot\underline{r}^{(3)} = 2a^2$, $x_1^{(3)} = -a$ and $x_2^{(3)} = -a$

$$\underline{\underline{I}}(O) \;=\; m\left\{2a^2\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix} - a^2\begin{pmatrix}1&1&0\\1&1&0\\0&0&0\end{pmatrix}\right\} \;=\; ma^2\begin{pmatrix}1&-1&0\\-1&1&0\\0&0&2\end{pmatrix}\,.$$

For $m^{(4)} = m$ at $(-a, a, 0)$, $\underline{r}^{(4)} = -a\underline{e}_1 + a\underline{e}_2$, so $\underline{r}^{(4)} \cdot \underline{r}^{(4)} = 2a^2$, $x_1^{(4)} = -a$ and $x_2^{(4)} = a$

$$\underline{\underline{I}}(O) \;=\; m\left\{2a^2\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix} - a^2\begin{pmatrix}1&-1&0\\-1&1&0\\0&0&0\end{pmatrix}\right\} \;=\; ma^2\begin{pmatrix}1&1&0\\1&1&0\\0&0&2\end{pmatrix}.$$

Adding up the four contributions gives the inertia tensor for all 4 particles as

$$\boxed{\underline{\underline{I}}(O) \;=\; 4ma^2\begin{pmatrix}1&0&0\\0&1&0\\0&0&2\end{pmatrix}.}$$

Note that the final inertia tensor is diagonal and in this basis the products of inertia are all zero (of course, there are other bases where the tensor is not diagonal). This implies the basis vectors are eigenvectors of the inertia tensor. For example, if $\underline{\omega} = \omega(0,0,1)$ then $\underline{L}(O) = 8m\omega a^2(0,0,1)$.

In general $\underline{L}(O)$ is not parallel to $\underline{\omega}$. For example, if $\underline{\omega} = \omega(0,1,1)$ then $\underline{L}(O) = 4m\omega a^2(0,1,2)$. Note that the inertia tensors for the individual masses are not diagonal.

9.3 The parallel axes theorem

If G is the **centre of mass** of the body its position vector $\underline{R}$ is given by

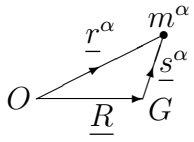
$$\underline{R} \;= \sum_\alpha m^\alpha \underline{r}^\alpha \Big/ M,$$

where $\underline{r}^\alpha$ are the position vectors relative to O and $M = \sum_\alpha m^\alpha$, is the **total mass** of the system.

The parallel axes theorem states that

$$\boxed{I_{ij}(O) \;= I_{ij}(G) + M\left\{(\underline{R}\cdot\underline{R})\,\delta_{ij} - R_iR_j\right\}},$$

Proof: Let $\underline{s}^\alpha$ be the position of m^α **with respect to** G, then

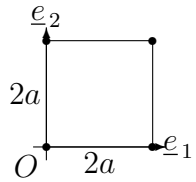


$$\begin{aligned}
I_{ij}(G) &= \sum_\alpha m^\alpha\left\{(\underline{s}^\alpha\cdot\underline{s}^\alpha)\,\delta_{ij} - s_i^\alpha s_j^\alpha\right\};\\
I_{ij}(O) &= \sum_\alpha m^\alpha\left\{(\underline{r}^\alpha\cdot\underline{r}^\alpha)\,\delta_{ij} - x_i^\alpha x_j^\alpha\right\}\\
&= \sum_\alpha m^\alpha\left\{(\underline{R}+\underline{s}^\alpha)^2\,\delta_{ij} - (\underline{R}+\underline{s}^\alpha)_i(\underline{R}+\underline{s}^\alpha)_j\right\}\\
&= M\left\{R^2\delta_{ij} - R_iR_j\right\} + \sum_\alpha m^\alpha\left\{(\underline{s}^\alpha\cdot\underline{s}^\alpha)\,\delta_{ij} - s_i^\alpha s_j^\alpha\right\}\\
&\qquad +2\,\delta_{ij}\,\underline{R}\cdot\sum_\alpha m^\alpha\underline{s}^\alpha - R_i\sum_\alpha m^\alpha s_j^\alpha - R_j\sum_\alpha m^\alpha s_i^\alpha\\
&= M\left\{R^2\delta_{ij} - R_iR_j\right\} + I_{ij}(G)
\end{aligned}$$

the cross terms vanishing since

$$\sum_\alpha m^\alpha s_i^\alpha \;=\; \sum_\alpha m^\alpha \left(r_i^\alpha - R_i\right) \;=\; 0\,.$$

Example: use of the parallel axes theorem



Consider the same arrangement of masses as before but with O at one corner of the square *i.e.* a (massless) lamina of side $2a$, with masses m at each corner and the origin O at the bottom, left so that the masses are at $(0,0,0)$, $(2a,0,0)$, $(0,2a,0)$ and $(2a,2a,0)$

We have $M = 4m$ and

$$\begin{aligned}\underline{OG} \;=\; \underline{R} \;&=\; \frac{1}{4m}\left\{m(0,0,0) + m(2a,0,0) + m(0,2a,0) + m(2a,2a,0)\right\}\\ &=\; (a,a,0)\end{aligned}$$

and so G is at the centre of the square and $R^2 = 2a^2$. We can now use the parallel axis theorem to relate the inertia tensor of the previous example to that of the present

$$\underline{\underline{I}}(O) - \underline{\underline{I}}(G) \;=\; 4m\left\{2a^2\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix} - a^2\begin{pmatrix}1&1&0\\1&1&0\\0&0&0\end{pmatrix}\right\} = \;4ma^2\begin{pmatrix}1&-1&0\\-1&1&0\\0&0&2\end{pmatrix}.$$

From the previous example,

$$\underline{\underline{I}}(G) \;=\; 4ma^2\begin{pmatrix}1&0&0\\0&1&0\\0&0&2\end{pmatrix} \quad \text{and hence}$$

$$\underline{\underline{I}}(O) \;=\; 4ma^2\begin{pmatrix}1+1&0-1&0\\0-1&1+1&0\\0&0&2+2\end{pmatrix} = \;4ma^2\begin{pmatrix}2&-1&0\\-1&2&0\\0&0&4\end{pmatrix}$$

Lecture 10: Eigenvectors of real, symmetric tensors

If $\underline{\underline{T}}$ is a (2nd-rank) tensor an eigenvector $\underline{n}$ of $\underline{\underline{T}}$ obeys (in any basis)

$$\boxed{\underline{\underline{T}}\,\underline{n} = t\,\underline{n}}\,.$$

The tensor acts on the **eigenvector** to produce a vector in the *same* direction, but changed in length by a factor t (the **eigenvalue**).

10.1 Construction of the eigenvectors

The matrix equation to solve is simply rearranged to one with a zero rhs: $\left(\underline{\underline{T}} - t\,\underline{\underline{I}}\right)\,\underline{n} = 0$. You should know the standard result that such a matrix equation has a nontrivial solution ($\underline{n}$ nonzero) if and only if

$$\boxed{\det\left(\underline{\underline{T}} - t\,\underline{\underline{I}}\right) \equiv 0}\ .$$

i.e.

$$\begin{vmatrix} T_{11}-t & T_{12} & T_{13} \\ T_{21} & T_{22}-t & T_{23} \\ T_{31} & T_{32} & T_{33}-t \end{vmatrix} = 0\ .$$

This equation, known as the **characteristic** or **secular** equation, is a **cubic** in t, giving 3 real solutions $t^{(1)}$, $t^{(2)}$ and $t^{(3)}$ and corresponding eigenvectors $\underline{n}^{(1)}$, $\underline{n}^{(2)}$ and $\underline{n}^{(3)}$.

Example:

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

The characteristic equation reads

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

Thus
$$(1-t)\{t(t-1)-1\} - \{(1-t)-0\} = 0$$

and so
$$(1-t)\{t^2-t-2\} = (1-t)(t-2)(t+1) = 0.$$

Thus the solutions are $t=1$, $t=2$ and $t=-1$.

We now find the eigenvector for each of these eigenvalues, writing $\underline{n} = (n_1, n_2, n_3)$:

$$\begin{aligned} (1-t)\,n_1 &+ n_2 &&= 0 \\ n_1 &- t\,n_2 &+ n_3 &= 0 \\ &n_2 &+ (1-t)\,n_3 &= 0. \end{aligned}$$

For $t = t^{(1)} = 1$, this is

$$\left.\begin{aligned} n_2 &= 0 \\ n_1 - n_2 + n_3 &= 0 \\ n_2 &= 0 \end{aligned}\right\} \Longrightarrow n_2 = 0\,; n_3 = -n_1\,.$$

Note that we only get two equations: we could never expect the components to be determined completely, since any multiple of $\underline{n}$ will also be an eigenvector. Thus $n_1 : n_2 : n_3 = 1 : 0 : -1$ and a *unit* vector in the direction of $\underline{n}^{(1)}$ is

$$\underline{\hat{n}}^{(1)} = \frac{1}{\sqrt{2}}(1,0,-1)\ .$$

For $t = t^{(2)} = 2$, we have

$$\left.\begin{array}{rcrcrcl} -n_1 & + & n_2 & & & = & 0 \\ n_1 & - & 2n_2 & + & n_3 & = & 0 \\ & & n_2 & - & n_3 & = & 0 \end{array}\right\} \Longrightarrow n_2 \;=\; n_3 \;=\; n_1 \,.$$

Thus $n_1 : n_2 : n_3 \;=\; 1:1:1$ and a *unit* vector in the direction of $\underline{n}^{(2)}$ is

$$\underline{\hat{n}}^{(2)} \;=\; \frac{1}{\sqrt{3}}(1,1,1) \,.$$

For $t = t^{(3)} = -1$, a similar calculation (exercise) gives

$$\underline{\hat{n}}^{(3)} \;=\; \frac{1}{\sqrt{6}}(1,-2,1) \,.$$

Notes:

(1) We can equally well replace $\underline{n}$ by $-\underline{n}$ in any case.

(2) $\underline{\hat{n}}^{(1)} \cdot \underline{\hat{n}}^{(2)} = \underline{\hat{n}}^{(1)} \cdot \underline{\hat{n}}^{(3)} = \underline{\hat{n}}^{(2)} \cdot \underline{\hat{n}}^{(3)} = 0$, so the eigenvectors are mutually orthogonal.

10.2 Important theorem and proof

Theorem: If T_{ij} is *real* and *symmetric*, its eigenvalues are *real*. The eigenvectors corresponding to *distinct* eigenvalues are *orthogonal*.

Proof: Let $\underline{A}$ and $\underline{B}$ be eigenvectors, with eigenvalues a and b respectively, then

$$\begin{aligned} T_{ij}A_j &= a\,A_i \\ T_{ij}B_j &= b\,B_i \end{aligned}$$

We multiply the first equation by B_i^*, and sum over i, giving

$$T_{ij}A_jB_i^* \;=\; a\,A_iB_i^*$$

We now take the complex conjugate of the second equation, multiply by A_i and sum over i, to give

$$T_{ij}^*B_j^*A_i \;=\; b^*\,B_i^*A_i$$

Since T_{ij} is *real* and *symmetric*, $T_{ij}^* = T_{ji}$, and so we have created $T_{ij}A_jB_i^*$ twice. Subtracting the two right-hand sides gives

$$\boxed{(a-b^*)\,A_iB_i^* \;=\; 0} \,.$$

Case 1: If we choose $\underline{B} = \underline{A}$, $A_iA_i^* \;=\; \sum_{i=1}^{3}|A_i|^2 \;> 0$, so

$$\boxed{a \;=\; a^*}\;.$$

Thus, we have shown that the eigenvalues are real.

Since a is real and T_{ij} are real, real A_i can be found.

Case 2: If we choose $\underline{B} \neq \underline{A}$, and *assume* for now that $a \neq b$, then $(a - b^*)$ is non-zero, implying

$$\boxed{\underline{A} \cdot \underline{B} \;=\; 0}\;.$$

Thus the eigenvectors are orthogonal if the eigenvalues are distinct.

10.3 Degenerate eigenvalues

This neat proof won't always work. If the characteristic equation for the eigenvalue, t, takes the form

$$(t^{(1)} - t)(t^{(2)} - t)^2 = 0,$$

there is a repeated root and we have a *doubly degenerate* eigenvalue $t^{(2)}$.

Claim: In the case of a real, symmetric tensor we can nevertheless always find TWO mutually orthogonal solutions for $\underline{n}^{(2)}$ (which are both orthogonal to $\underline{n}^{(1)}$).

Example

$$\underline{\underline{T}} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \;\Rightarrow\; \left|\underline{\underline{T}} - t\underline{\underline{I}}\right| = \begin{vmatrix} -t & 1 & 1 \\ 1 & -t & 1 \\ 1 & 1 & -t \end{vmatrix} = 0 \Rightarrow \boxed{t = 2 \text{ and } t = -1 \text{ (twice)}}\;.$$

For $t = t^{(1)} = 2$ with eigenvector $\underline{n}^{(1)}$

$$\left.\begin{array}{rcrcrcl} -2n_1 & + & n_2 & + & n_3 & = & 0 \\ n_1 & - & 2n_2 & + & n_3 & = & 0 \\ n_1 & + & n_2 & - & 2n_3 & = & 0 \end{array}\right\} \Longrightarrow \left\{\begin{array}{rcl} n_2 & = & n_3 \;=\; n_1 \\ \hat{\underline{n}}^{(1)} & = & \frac{1}{\sqrt{3}}(1,1,1)\;. \end{array}\right.$$

For $t = t^{(2)} = -1$ with eigenvector $\underline{n}^{(2)}$

$$n_1^{(2)} + n_2^{(2)} + n_3^{(2)} = 0$$

is the only independent equation. This can be written as $\underline{n}^{(1)} \cdot \underline{n}^{(2)} = 0$ which is the equation for a plane normal to $\underline{n}^{(1)}$. Thus any vector orthogonal to $\underline{n}^{(1)}$ is an eigenvector with eigenvalue -1. It is clearly possible to choose an infinite number of different pairs of orthogonal vectors that are restricted to this plane, and thus orthogonal also to $\underline{n}^{(1)}$.

If the characteristic equation is of form

$$(t^{(1)} - t)^3 = 0$$

then we have a triply degenerate eigenvalue $t^{(1)}$. In fact, this only occurs if the tensor is equal to $t^{(1)}\delta_{ij}$ which means it is 'isotropic' and any direction is an eigenvector with eigenvalue $t^{(1)}$.

10.4 Diagonalisation of a real, symmetric tensor

In the basis $\{\underline{e}_i\}$ the tensor T_{ij} is, in general, non-diagonal. i.e. T_{ij} is non-zero for $i \neq j$. But we now show that it is always possible to find a basis in which the tensor becomes diagonal. Moreover, this basis is such that the we use the normalised eigenvectors (the **principal axes**) as the basis vectors.

It is relatively easy to see that this works, by considering the action of $\underline{\underline{T}}$ on a vector $\underline{V}$:

$$\underline{\underline{T}}\,\underline{V} = \underline{\underline{T}} \sum_i V_i\,\underline{e}_i = \sum_i V_i\,\underline{\underline{T}}\,\underline{e}_i.$$

If the basis vectors are eigenvectors with eigenvalues $t^{(i)}$, then

$$\underline{\underline{T}}\,\underline{V} = \sum_i V_i\,t^{(i)}\,\underline{e}_i,$$

so the effect of $\underline{\underline{T}}$ is to multiply the components of the vector by the eigenvalues. Clearly, this can only happen if

$$\underline{\underline{T}} = \begin{pmatrix} t^{(1)} & 0 & 0 \\ 0 & t^{(2)} & 0 \\ 0 & 0 & t^{(3)} \end{pmatrix}.$$

Let's now prove this more carefully. We want to convert to the basis where $\underline{e}_i{}' = \underline{n}^{(i)}$, the normalised eigenvectors of $\underline{\underline{T}}$. Choose the elements of the transformation matrix to be

$$\lambda_{ij} \;=\; \underline{e}_i{}' \cdot \underline{e}_j \;=\; \underline{n}^{(i)} \cdot \underline{e}_j \;=\; n_j^{(i)}\;;$$

i.e. the **rows** of λ are the components of the **normalised eigenvectors** of T.

In the basis $\{\underline{e}_i{}'\}$

$$T'_{ij} = (\lambda T \lambda^T)_{ij}$$

Now since the **columns** of λ^T are the **normalised eigenvectors** of T we see that

$$\underline{\underline{T}}\,\underline{\underline{\lambda}}^T = \begin{pmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{pmatrix} \begin{pmatrix} n_1^{(1)} & n_1^{(2)} & n_1^{(3)} \\ n_2^{(1)} & n_2^{(2)} & n_2^{(3)} \\ n_3^{(1)} & n_3^{(2)} & n_3^{(3)} \end{pmatrix} = \begin{pmatrix} t^{(1)}n_1^{(1)} & t^{(2)}n_1^{(2)} & t^{(3)}n_1^{(3)} \\ t^{(1)}n_2^{(1)} & t^{(2)}n_2^{(2)} & t^{(3)}n_2^{(3)} \\ t^{(1)}n_3^{(1)} & t^{(2)}n_3^{(2)} & t^{(3)}n_3^{(3)} \end{pmatrix}$$

$$\underline{\underline{\lambda}}\,\underline{\underline{T}}\,\underline{\underline{\lambda}}^T \;=\; \begin{pmatrix} n_1^{(1)} & n_2^{(1)} & n_3^{(1)} \\ n_1^{(2)} & n_2^{(2)} & n_3^{(2)} \\ n_1^{(3)} & n_2^{(3)} & n_3^{(3)} \end{pmatrix} \begin{pmatrix} t^{(1)}n_1^{(1)} & t^{(2)}n_1^{(2)} & t^{(3)}n_1^{(3)} \\ t^{(1)}n_2^{(1)} & t^{(2)}n_2^{(2)} & t^{(3)}n_2^{(3)} \\ t^{(1)}n_3^{(1)} & t^{(2)}n_3^{(2)} & t^{(3)}n_3^{(3)} \end{pmatrix} = \begin{pmatrix} t^{(1)} & 0 & 0 \\ 0 & t^{(2)} & 0 \\ 0 & 0 & t^{(3)} \end{pmatrix}$$

from the orthonormality of the $\underline{n}^{(i)}$ (rows of λ; columns of λ^T).

Thus, with respect to a basis defined by the eigenvectors or principal axes of the tensor, the tensor has *diagonal form*. [i.e. $\underline{\underline{T}}' = \mathrm{diag}\{t^{(1)}, t^{(2)}, t^{(3)}\}$.] The diagonal basis is often referred to as the '**principal axes basis**'.

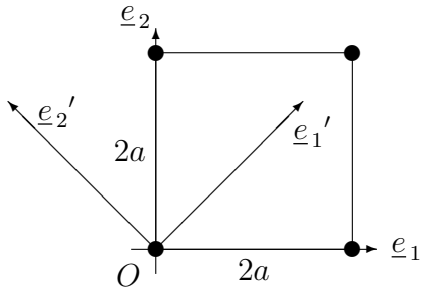
Note: In the diagonal basis the trace of a tensor is the sum of the eigenvalues; the determinant of the tensor is the product of the eigenvalues. Since the trace and determinant are

invariants this means that in any basis the trace and determinant are the sum and products of the eigenvalues respectively.

Example: Diagonalisation of the inertia tensor Consider the inertia tensor studied earlier: four masses arranged in a square with the origin at the left hand corner

$$\underline{\underline{I}}(O) \;=\; 4ma^2 \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

It is easy to check (exercise) that the eigenvectors (or principal axes of inertia) are $(\underline{e}_1 + \underline{e}_2)$ (eigenvalue $4ma^2$), $(\underline{e}_1 - \underline{e}_2)$ (eigenvalue $12ma^2$) and $\underline{e}_3$ (eigenvalue $16ma^2$).



Defining the $\underline{e}_i{}'$ basis as normalised eigenvectors: $\underline{e}_1{}' = \frac{1}{\sqrt{2}}(\underline{e}_1 + \underline{e}_2)$; $\underline{e}_2{}' = \frac{1}{\sqrt{2}}(-\underline{e}_1 + \underline{e}_2)$; $\underline{e}_3{}' = \underline{e}_3$, one obtains

$$\underline{\underline{\lambda}} \;=\; \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (\text{ a rotation of } \pi/4 \text{ about } \underline{e}_3 \text{ axis})$$

and the inertia tensor in the basis $\{\underline{e}_i{}'\}$ has components $I'_{ij}(O) = \left(\lambda\, I(O)\lambda^T\right)_{ij}$ so that

$$\begin{aligned} \underline{\underline{I}}'(O) \;&=\; 4ma^2 \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &=\; 4ma^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{pmatrix} . \end{aligned}$$

We see that the tensor is diagonal with diagonal elements which are the eigenvalues (principal moments of inertia).

Remark: Diagonalisability is a very special and useful property of real, symmetric tensors. It is a property also shared by the more general class of Hermitian operators which you will meet in quantum mechanics in third year. A general tensor does not share the property. For example a real non-symmetric tensor cannot be diagonalised.

Lecture 11: Fields

In physics we often have to consider properties that vary in some region of space e.g. temperature of a body. To do this we require the concept of fields.

If to each point $\underline{r}$ in some region of ordinary 3D space there corresponds a **scalar** $\phi(x_1, x_2, x_3)$, then $\phi(\underline{r})$ is a **scalar field**.

Examples: temperature distribution in a body $T(\underline{r})$, pressure in the atmosphere $P(\underline{r})$, electric charge density or mass density $\rho(\underline{r})$, electrostatic potential $\phi(\underline{r})$.

Similarly a **vector field** assigns a vector $\underline{V}(x_1, x_2, x_3)$ to each point $\underline{r}$ of some region.

Examples: velocity in a fluid $\underline{v}(\underline{r})$, electric current density $\underline{J}(\underline{r})$, electric field $\underline{E}(\underline{r})$, magnetic field $\underline{B}(\underline{r})$

A vector field in 2D can be represented graphically, at a carefully selected set of points $\underline{r}$, by an arrow whose length and direction is proportional to $\underline{V}(\underline{r})$ *e.g.* wind velocity on a weather forecast chart.

11.1 Level surfaces of a scalar field

If $\phi(\underline{r})$ is a non-constant scalar field, then the equation $\phi(\underline{r}) = c$ where c is a constant, defines a **level surface** (or equipotential) of the field. Level surfaces do not intersect (otherwise ϕ would be multi-valued at the point of intersection).

Familiar examples in two dimensions, where they are level curves rather than level surfaces, are the contours of constant height on a geographical map, $h(x_1, x_2) = c$. Also isobars on a weather map are level curves of pressure $P(x_1, x_2) = c$.

Examples in three dimensions:

(i) Suppose that

$$\phi(\underline{r}) = x_1^2 + x_2^2 + x_3^2 = x^2 + y^2 + z^2$$

The level surface $\phi(\underline{r}) = c$ is a sphere of radius $\sqrt{c}$ centred on the origin. As c is varied, we obtain a family of level surfaces which are concentric spheres.

(ii) Electrostatic potential due to a point charge q situated at the point $\underline{a}$ is

$$\phi(\underline{r}) = \frac{q}{4\pi\epsilon_0} \frac{1}{|\underline{r} - \underline{a}|}$$

The level surfaces are concentric spheres centred on the point $\underline{a}$.

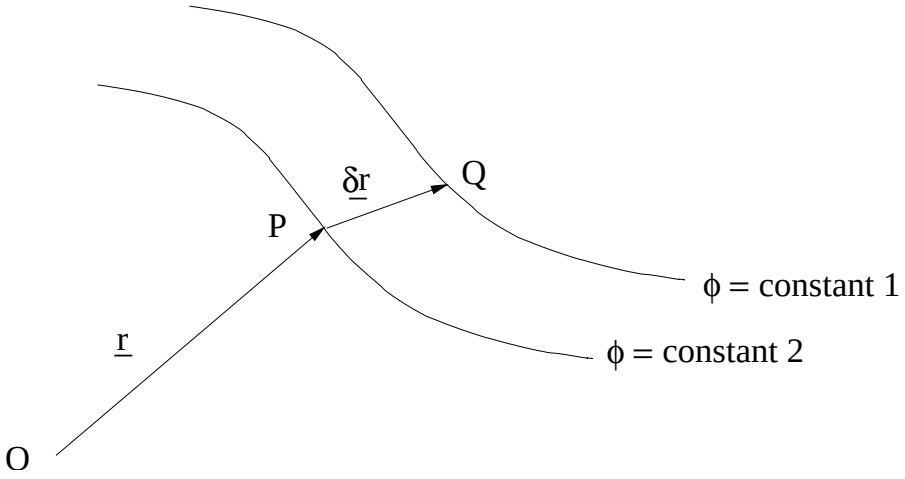
(iii) Let $\phi(\underline{r}) = \underline{k} \cdot \underline{r}$. The level surfaces are planes $\underline{k} \cdot \underline{r} =$ constant with normal $\underline{k}$.

(iv) Let $\phi(\underline{r}) = \exp(i\underline{k} \cdot \underline{r})$. Note that this a complex scalar field. Since $\underline{k} \cdot \underline{r} =$ constant is the equation for a plane, the level surfaces are planes.

11.2 Gradient of a scalar field

How does a scalar field change as we change position? As an example think of a 2D contour map of the height $h = h(x, y)$ of a hill. If we are on the hill and move in the $x - y$ plane then the change in height will depend on the direction in which we move. In particular, there will be a direction in which the height increases most steeply ('straight up the hill') We now introduce a formalism to describe how a scalar field $\phi(\underline{r})$ changes as a function of $\underline{r}$.

Let $\phi(\underline{r})$ be a scalar field. Consider 2 nearby points: P (position vector $\underline{r}$) and Q (position vector $\underline{r} + \underline{\delta r}$). Assume P and Q lie on *different* level surfaces as shown:



Now use a Taylor series for a function of 3 variables to evaluate the change in ϕ as we move from P to Q

$$\begin{aligned}
\delta\phi &\equiv \phi(\underline{r} + \underline{\delta r}) - \phi(\underline{r}) \\
&= \phi(x_1 + \delta x_1, x_2 + \delta x_2, x_3 + \delta x_3) - \phi(x_1, x_2, x_3) \\
&= \frac{\partial\phi(\underline{r})}{\partial x_1}\,\delta x_1 + \frac{\partial\phi(\underline{r})}{\partial x_2}\,\delta x_2 + \frac{\partial\phi(\underline{r})}{\partial x_3}\,\delta x_3 + O(\,\delta x_i^2\,).
\end{aligned}$$

We have of course assumed that all the partial derivatives exist. Neglecting terms of order $(\,\delta x_i^2\,)$ we can write

$$\boxed{\delta\phi = \underline{\nabla}\,\phi(\underline{r}) \cdot \underline{\delta r}}$$

where the 3 quantities

$$\left(\underline{\nabla}\,\phi(\underline{r})\right)_i = \frac{\partial\phi(\underline{r})}{\partial x_i}$$

form the Cartesian components of a **vector field**. We write

$$\boxed{\underline{\nabla}\,\phi(\underline{r}) \equiv \underline{e}_i\,\frac{\partial\phi(\underline{r})}{\partial x_i} = \underline{e}_1\,\frac{\partial\phi(\underline{r})}{\partial x_1} + \underline{e}_2\,\frac{\partial\phi(\underline{r})}{\partial x_2} + \underline{e}_3\,\frac{\partial\phi(\underline{r})}{\partial x_3}}$$

or in the old 'x, y, z' notation (where $x_1 = x$, $x_2 = y$ and $x_3 = z$)

$$\underline{\nabla}\,\phi(\underline{r}) = \underline{e}_1\,\frac{\partial\phi(\underline{r})}{\partial x} + \underline{e}_2\,\frac{\partial\phi(\underline{r})}{\partial y} + \underline{e}_3\,\frac{\partial\phi(\underline{r})}{\partial z}$$

The **vector field** $\underline{\nabla}\,\phi(\underline{r})$, pronounced 'grad phi', is called the **gradient** of $\phi(\underline{r})$.

11.3 The operator 'del'

We can think of the **vector operator** $\underline{\nabla}$ (pronounced 'del') acting on the **scalar field** $\phi(\underline{r})$ to produce the **vector field** $\underline{\nabla}\,\phi(\underline{r})$.

In Cartesians:

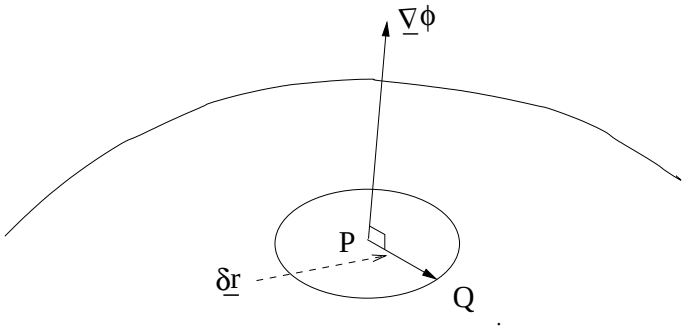
$$\boxed{\underline{\nabla} = \underline{e}_i \frac{\partial}{\partial x_i} = \underline{e}_1 \frac{\partial}{\partial x_1} + \underline{e}_2 \frac{\partial}{\partial x_2} + \underline{e}_3 \frac{\partial}{\partial x_3}}$$

We call $\underline{\nabla}$ an 'operator' since it operates on something to its *right*. It is a vector operator since it has vector transformation properties.

11.4 Interpretation of the gradient

In deriving the expression for $\delta\phi$ above, we assumed that the points P and Q lie on *different* level surfaces. Now consider the situation where P and Q are nearby points on the *same* level surface. In that case $\delta\phi = 0$ and so

$$\delta\phi = \underline{\nabla}\phi(\underline{r}) \cdot \underline{\delta r} = 0$$



The infinitesimal vector $\underline{\delta r}$ lies in the level surface at $\underline{r}$, and the above equation holds for all such $\underline{\delta r}$, hence

$$\boxed{\underline{\nabla}\phi(\underline{r}) \text{ is } \textbf{normal to the level surface at } \underline{r}.}$$

11.5 Directional derivative

Now consider the change, $\delta\phi$, produced in ϕ by moving distance δs in some direction say $\hat{\underline{s}}$.

Then $\underline{\delta r} = \hat{\underline{s}}\,\delta s$ and

$$\delta\phi = \underline{\nabla}\phi(\underline{r}) \cdot \underline{\delta r} = (\underline{\nabla}\phi(\underline{r}) \cdot \hat{\underline{s}})\,\delta s$$

As $\delta s \to 0$, the rate of change of ϕ as we move in the direction of $\hat{\underline{s}}$ is

$$\frac{d\phi(\underline{r})}{ds} = \hat{\underline{s}} \cdot \underline{\nabla}\phi(\underline{r}) = |\underline{\nabla}\phi(\underline{r})| \cos\theta \tag{2}$$

where θ is the angle between $\hat{\underline{s}}$ and the normal to the level surface at $\underline{r}$.

$\hat{\underline{s}} \cdot \underline{\nabla}\phi(\underline{r})$ is the **directional derivative** of the scalar field ϕ in the direction of $\hat{\underline{s}}$.

Note that the directional derivative has its *maximum* value when $\underline{s}$ is parallel to $\underline{\nabla}\phi(\underline{r})$, and is *zero* when $\underline{s}$ lies in the level surface. Therefore

$\underline{\nabla}\phi$ points in the direction of the **maximum** rate of increase in ϕ

Also recall that this direction is normal to the level surface. For a familiar example think of the contour lines on a map. The steepest direction is perpendicular to the contour lines.

Example: calculate the gradient of $\phi = r^2 = x^2 + y^2 + z^2$

$$\begin{aligned} \underline{\nabla}\phi(\underline{r}) &= \left(\underline{e}_1 \frac{\partial}{\partial x} + \underline{e}_2 \frac{\partial}{\partial y} + \underline{e}_3 \frac{\partial}{\partial z} \right) (x^2 + y^2 + z^2) \\ &= 2x\,\underline{e}_1 + 2y\,\underline{e}_2 + 2z\,\underline{e}_3 = 2\underline{r} \end{aligned}$$

Example: Find the directional derivative of $\phi = xy(x+z)$ at point $(1, 2, -1)$ in the $(\underline{e}_1 + \underline{e}_2)/\sqrt{2}$ direction.

$$\underline{\nabla}\phi = (2xy + yz)\underline{e}_1 + x(x+z)\underline{e}_2 + xy\underline{e}_3 = 2\underline{e}_1 + 2\underline{e}_3$$

at $(1, 2, -1)$. Thus at this point

$$\frac{1}{\sqrt{2}} (\underline{e}_1 + \underline{e}_2) \cdot \underline{\nabla}\phi = \sqrt{2}$$

Physical example: Let $T(\underline{r})$ be the temperature of the atmosphere at the point $\underline{r}$. An object flies through the atmosphere with velocity $\underline{v}$. Obtain an expression for the rate of change of temperature experienced by the object.

As the object moves from $\underline{r}$ to $\underline{r} + \underline{\delta r}$ in time δt, it sees a change in temperature

$$\delta T(\underline{r}) = \underline{\nabla} T(\underline{r}) \cdot \underline{\delta r}.$$

For a small time interval, $\delta T \simeq (dT/dt)\delta t$ and $\underline{\delta r} \simeq \underline{v}\,\delta t$, so dividing by δt gives

$$\frac{dT(\underline{r})}{dt} = \underline{v} \cdot \underline{\nabla} T(\underline{r})$$

Lecture 12: More on differentiation of fields

12.1 Maxima and minima

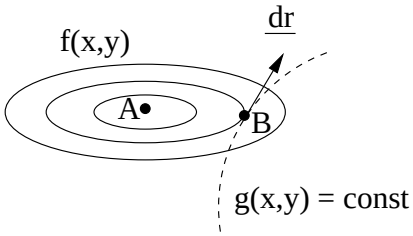
From this reasoning, it is easy to see the criterion that has to be satisfied at a maximum or minimum of a field $f(\underline{r})$ (or a stationary point):

$$\underline{\nabla} f = 0.$$

A more interesting case is a *conditional* extremum: find a stationary point of f subject to the condition that some other function $g(\underline{r})$ is constant. In effect, we need to see how f varies as we move along a level line of the function g. If $\underline{dr}$ lies in that level line, then we have

$$\underline{\nabla} f \cdot \underline{dr} = \underline{\nabla} g \cdot \underline{dr} = 0.$$

But if $\underline{dr}$ points in a different direction, then $\underline{\nabla} f \cdot \underline{dr}$ would be non-zero in general: this is the difference between conditional and unconditional extrema.



Consider the function f, which has a maximum at point A. If we follow the dotted level line, on which the function g is a constant, the value of f constrained in this way reaches a maximum at point B. Here, the directional derivative of f is zero along a vector $\underline{dr}$ that is tangent to the level line: *i.e.* $\underline{\nabla} f \cdot \underline{dr} = 0$.

However, there is a very neat way of converting the problem into an unconditional one. Just write

$$\underline{\nabla}(f + \lambda g) = 0,$$

where the constant λ is called a **Lagrange multiplier**. We want to choose it so that $\underline{\nabla}(f + \lambda g) \cdot \underline{dr} = 0$ for *any* $\underline{dr}$. Clearly it is satisfied for the initial case where $\underline{dr}$ lies in the level line of g, in which case $\underline{dr}$ is perpendicular to $\underline{\nabla} g$ – and also to $\underline{\nabla} f$. To make a general vector, we have to add a component in the direction of $\underline{\nabla} g$ – but the effect of moving in this direction will be zero if we choose

$$\lambda = -(\underline{\nabla} f \cdot \underline{\nabla} g) \,/\, |\underline{\nabla} g|^2,$$

evaluated at the desired solution. Since we don't know this in advance, λ is called an **undetermined multiplier**.

Example If we don't know λ, what use is it? The answer is we find it out at the end. Consider $f = r^2$ and find a stationary point subject to the condition $g = x + y = 1$. We have $\underline{\nabla}(x^2 + y^2 + z^2 + \lambda[x + y]) = 0$; in components, this is $(2x + \lambda, 2y + \lambda, 2z) = 0$, so we learn $z = 0$ and $x = y = -\lambda/2$. Now, since $x + y = 1$, this requires $\lambda = -1$, and so the required stationary point is $(1/2, 1/2, 0)$.

12.2 Examples on gradient

Now we do some exercises on the gradient using suffix notation. As usual suffix notation is most convenient for proving more complicated identities.

1. Let $\phi(\underline{r}) = r^2 = x_1^2 + x_2^2 + x_3^2$, then

$$\underline{\nabla}\phi(\underline{r}) = \left(\underline{e}_1 \frac{\partial}{\partial x_1} + \underline{e}_2 \frac{\partial}{\partial x_2} + \underline{e}_3 \frac{\partial}{\partial x_3}\right)(x_1^2 + x_2^2 + x_3^2) = 2x_1\,\underline{e}_1 + 2x_2\,\underline{e}_2 + 2x_3\,\underline{e}_3 = 2\underline{r}$$

In suffix notation

$$\underline{\nabla}\phi(\underline{r}) = \underline{\nabla}\, r^2 = \left(\underline{e}_i \frac{\partial}{\partial x_i}\right)(x_j x_j) = \underline{e}_i\,(\delta_{ij}x_j + x_j\delta_{ij}) = \underline{e}_i\, 2x_i = 2\underline{r}$$

In the above we have used the important property of partial derivatives

$$\boxed{\frac{\partial x_i}{\partial x_j} = \delta_{ij}}$$

The level surfaces of r^2 are spheres centred on the origin, and the gradient of r^2 at $\underline{r}$ points radially outward with magnitude $2r$.

2. Let $\phi = \underline{a} \cdot \underline{r}$ where $\underline{a}$ is a *constant* vector.

$$\underline{\nabla}\,(\underline{a} \cdot \underline{r}) = \left(\underline{e}_i \frac{\partial}{\partial x_i}\right)(a_j x_j) = \underline{e}_i\, a_j \delta_{ij} = \underline{a}$$

This is not surprising, since the level surfaces $\underline{a} \cdot \underline{r} = c$ are planes orthogonal to $\underline{a}$.

12.3 Identities for gradients

If $\phi(\underline{r})$ and $\psi(\underline{r})$ are real scalar fields, then:

1. **Distributive law**

$$\boxed{\underline{\nabla}\,\big(\phi(\underline{r}) + \psi(\underline{r})\big) \;=\; \underline{\nabla}\,\phi(\underline{r}) + \underline{\nabla}\,\psi(\underline{r})}$$

Proof:

$$\underline{\nabla}\,\big(\phi(\underline{r}) \,+\, \psi(\underline{r})\big) \;=\; \underline{e}_i \frac{\partial}{\partial x_i}\,\big(\phi(\underline{r}) \,+\, \psi(\underline{r})\big) \;=\; \underline{\nabla}\,\phi(\underline{r}) \,+\, \underline{\nabla}\,\psi(\underline{r})$$

2. **Product rule**

$$\boxed{\underline{\nabla}\,\big(\phi(\underline{r})\,\psi(\underline{r})\big) \;=\; \psi(\underline{r})\;\underline{\nabla}\,\phi(\underline{r}) + \phi(\underline{r})\;\underline{\nabla}\,\psi(\underline{r})}$$

Proof:

$$\begin{aligned}
\underline{\nabla}\,\big(\phi(\underline{r})\,\psi(\underline{r})\big) \;&=\; \underline{e}_i \frac{\partial}{\partial x_i}\,\big(\phi(\underline{r})\,\psi(\underline{r})\big) \\
&=\; \underline{e}_i \left(\psi(\underline{r})\,\frac{\partial\phi(\underline{r})}{\partial x_i} \;+\; \phi(\underline{r})\,\frac{\partial\psi(\underline{r})}{\partial x_i}\right) \\
&=\; \psi(\underline{r})\;\underline{\nabla}\,\phi(\underline{r}) \;+\; \phi(\underline{r})\;\underline{\nabla}\,\psi(\underline{r})
\end{aligned}$$

3. **Chain rule:** If $F(\phi(\underline{r}))$ is a scalar field, then

$$\boxed{\underline{\nabla}\, F(\phi(\underline{r})) \;=\; \frac{\partial F(\phi)}{\partial \phi}\; \underline{\nabla}\, \phi(\underline{r})}$$

Proof:

$$\underline{\nabla}\, F(\phi(\underline{r})) \;=\; \underline{e}_i \frac{\partial}{\partial x_i}\, F(\phi(\underline{r})) \;=\; \underline{e}_i \frac{\partial F(\phi)}{\partial \phi}\; \frac{\partial \phi(\underline{r})}{\partial x_i} \;=\; \frac{\partial F(\phi)}{\partial \phi}\; \underline{\nabla}\, \phi(\underline{r})$$

Example of Chain Rule: If $\phi(\underline{r}) \;=\; r$ and $F(\phi(\underline{r})) = \phi(\underline{r})^n = r^n$, then

$$\underline{\nabla}\, (r^n) \;=\; (n\, r^{n-1})\, \hat{\underline{r}} \;=\; (n\, r^{n-2})\, \underline{r}.$$

12.4 More on vector operators

We have seen how $\underline{\nabla}$ acts on a scalar field to produce a vector field. We can make products of the vector operator $\underline{\nabla}$ with other vector quantities to produce new operators and fields in the same way as we could make scalar and vector products of two vectors.

For example, recall that the directional derivative of ϕ in direction $\hat{\underline{s}}$ was given by $\hat{\underline{s}} \cdot \underline{\nabla}\phi$. Generally, we can interpret $\underline{A} \cdot \underline{\nabla}$ as a **scalar operator**:

$$\underline{A} \cdot \underline{\nabla} = A_i \frac{\partial}{\partial x_i}$$

i.e. $\underline{A} \cdot \underline{\nabla}$ acts on a scalar field to its *right* to produce another scalar field

$$(\underline{A} \cdot \underline{\nabla})\, \phi(\underline{r}) \;=\; A_i \frac{\partial \phi(\underline{r})}{\partial x_i} \;=\; A_1 \frac{\partial \phi(\underline{r})}{\partial x_1} \;+\; A_2 \frac{\partial \phi(\underline{r})}{\partial x_2} \;+\; A_3 \frac{\partial \phi(\underline{r})}{\partial x_3}$$

Actually we can also act with this operator on a vector field to get another vector field.

$$\begin{aligned}(\underline{A} \cdot \underline{\nabla})\, \underline{V}(\underline{r}) \;&=\; A_i \frac{\partial}{\partial x_i} \underline{V}(\underline{r}) \;=\; A_i \frac{\partial}{\partial x_i} \left(V_j(\underline{r})\, \underline{e}_j\right) \\ &=\; \underline{e}_1\, (\underline{A} \cdot \underline{\nabla})\, V_1(\underline{r}) \;+\; \underline{e}_2\, (\underline{A} \cdot \underline{\nabla})\, V_2(\underline{r}) \;+\; \underline{e}_3\, (\underline{A} \cdot \underline{\nabla})\, V_3(\underline{r})\end{aligned}$$

The alternative expression $\underline{A} \cdot \left(\underline{\nabla}\, \underline{V}(\underline{r})\right)$ is *undefined* because $\underline{\nabla}\, \underline{V}(\underline{r})$ doesn't make sense.

N.B. Great care is required with the order in products since, in general, products involving operators are not commutative. For example

$$\underline{\nabla} \cdot \underline{A} \;\neq\; \underline{A} \cdot \underline{\nabla}$$

$\underline{A} \cdot \underline{\nabla}$ is a scalar differential operator whereas $\underline{\nabla} \cdot \underline{A} = \partial A_i / \partial x_i$ gives a scalar field called the **divergence** of $\underline{A}$.

We now combine the vector operator $\underline{\nabla}$ ('del') with a vector field to define two new operations 'div' and 'curl'. Then we define the Laplacian.

12.5 The Laplacian operator ∇^2

We may take the *divergence* of the *gradient* of a scalar field $\phi(\underline{r})$

$$\underline{\nabla} \cdot (\underline{\nabla}\,\phi(\underline{r})) \;=\; \frac{\partial}{\partial x_i}\,\frac{\partial}{\partial x_i}\,\phi(\underline{r}) \;\equiv\; \nabla^2\phi(\underline{r})$$

∇^2 is the **Laplacian operator**, pronounced 'del-squared'. In Cartesian coordinates

$$\nabla^2 \;=\; \frac{\partial}{\partial x_i}\,\frac{\partial}{\partial x_i}$$

More explicitly

$$\boxed{\nabla^2\,\phi(\underline{r}) \;=\; \frac{\partial^2\phi}{\partial x_1^2} + \frac{\partial^2\phi}{\partial x_2^2} + \frac{\partial^2\phi}{\partial x_3^2} \quad \text{or} \quad \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial x^2}}$$

Example

$$\nabla^2\, r^2 = \frac{\partial}{\partial x_i}\,\frac{\partial}{\partial x_i} x_j x_j = \frac{\partial}{\partial x_i}(2x_i) = 2\delta_{ii} = 6\ .$$

In Cartesian coordinates, the effect of the Laplacian on a vector field $\underline{A}$ is *defined* to be

$$\nabla^2\underline{A}(\underline{r}) \;=\; \frac{\partial}{\partial x_i}\,\frac{\partial}{\partial x_i}\,\underline{A}(\underline{r}) \;=\; \frac{\partial^2}{\partial x_1^2}\,\underline{A}(\underline{r}) \,+\, \frac{\partial^2}{\partial x_2^2}\,\underline{A}(\underline{r}) \,+\, \frac{\partial^2}{\partial x_3^2}\,\underline{A}(\underline{r})$$

The Laplacian acts on a vector field to produce another vector field.

12.6 Divergence

We define the **divergence** of a vector field $\underline{A}$ (pronounced 'div A') as

$$\boxed{\text{div}\,\underline{A}(\underline{r}) \;\equiv\; \underline{\nabla} \cdot \underline{A}(\underline{r})}$$

In Cartesian coordinates

$$\boxed{\begin{aligned} \underline{\nabla} \cdot \underline{A}(\underline{r}) = \frac{\partial}{\partial x_i}\,A_i(\underline{r}) \;&=\; \frac{\partial A_1(\underline{r})}{\partial x_1} + \frac{\partial A_2(\underline{r})}{\partial x_2} + \frac{\partial A_3(\underline{r})}{\partial x_3} \\ \text{or} \qquad &\frac{\partial A_x(\underline{r})}{\partial x} + \frac{\partial A_y(\underline{r})}{\partial y} + \frac{\partial A_z(\underline{r})}{\partial z} \quad \text{in } x, y, z \text{ notation} \end{aligned}}$$

Example: $\underline{A}(\underline{r}) = \underline{r} \quad \Rightarrow \quad \boxed{\underline{\nabla} \cdot \underline{r} \;=\; 3}$ a *very* useful & important result

$$\underline{\nabla} \cdot \underline{r} = \frac{\partial x_1}{\partial x_1} + \frac{\partial x_2}{\partial x_2} + \frac{\partial x_3}{\partial x_3} = 1 + 1 + 1 = 3$$

In suffix notation

$$\underline{\nabla} \cdot \underline{r} = \frac{\partial x_i}{\partial x_i} = \delta_{ii} = 3.$$

Lecture 13: Curl and its meaning

13.1 Curl

We define the curl of a vector field, curl $\underline{A}$, as the *vector* field

$$\boxed{\text{curl}\,\underline{A}(\underline{r}) \;\equiv\; \underline{\nabla} \times \underline{A}(\underline{r})}$$

In Cartesian coordinates, this means that the ith component of $\underline{\nabla} \times \underline{A}$ is

$$\boxed{\left(\underline{\nabla} \times \underline{A}\right)_i \;=\; \epsilon_{ijk}\,\frac{\partial}{\partial x_j}\,A_k}$$

More explicitly, we can use a determinant form (cf. the expression of the vector product)

$$\boxed{\underline{\nabla} \times \underline{A} \;=\; \begin{vmatrix} \underline{e}_1 & \underline{e}_2 & \underline{e}_3 \\ \frac{\partial}{\partial x_1} & \frac{\partial}{\partial x_2} & \frac{\partial}{\partial x_3} \\ A_1 & A_2 & A_3 \end{vmatrix} \quad \text{or} \quad \begin{vmatrix} \underline{e}_x & \underline{e}_y & \underline{e}_z \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ A_x & A_y & A_z \end{vmatrix} .}$$

Example: $\underline{A}(\underline{r}) = \underline{r} \quad \Rightarrow \quad \boxed{\underline{\nabla} \times \underline{r} \;=\; 0}$ another *very* useful & important result

$$\begin{aligned} \underline{\nabla} \times \underline{r} \;&=\; \underline{e}_i\,\epsilon_{ijk}\,\frac{\partial}{\partial x_j}\,x_k \\ &=\; \underline{e}_i\,\epsilon_{ijk}\,\delta_{jk} = \underline{e}_i\,\epsilon_{ijj} = 0 \end{aligned}$$

or, using the determinant formula, $\underline{\nabla} \times \underline{r} \;=\; \begin{vmatrix} \underline{e}_1 & \underline{e}_2 & \underline{e}_3 \\ \frac{\partial}{\partial x_1} & \frac{\partial}{\partial x_2} & \frac{\partial}{\partial x_3} \\ x_1 & x_2 & x_3 \end{vmatrix} \;\equiv\; 0$

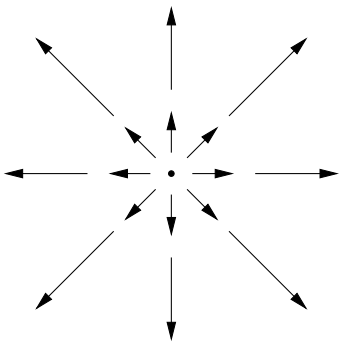
Example: Compute the curl of $\underline{V} = x^2 y \underline{e}_1 + y^2 x \underline{e}_2 + xyz\underline{e}_3$:

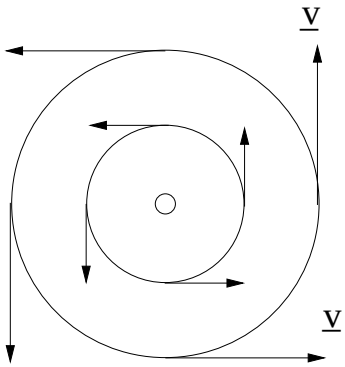
$$\underline{\nabla} \times \underline{V} \;=\; \begin{vmatrix} \underline{e}_1 & \underline{e}_2 & \underline{e}_3 \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ x^2 y & y^2 x & xyz \end{vmatrix} = \underline{e}_1(xz - 0) - \underline{e}_2(yz - 0) + \underline{e}_3(y^2 - x^2)$$

13.2 Physical interpretation of 'div' and 'curl'

Full interpretations of the divergence and curl of a vector field are best left until after we have studied the Divergence Theorem and Stokes' Theorem respectively. However, we can gain some intuitive understanding by looking at simple examples where div and/or curl vanish.

First consider the radial field $\underline{A} = \underline{r}$; $\underline{\nabla} \cdot \underline{A} = 3$; $\underline{\nabla} \times \underline{A} = 0$. We sketch the vector field $\underline{A}(\underline{r})$ by drawing at selected points vectors of the appropriate direction and magnitude. These give the tangents of 'flow lines'. Roughly speaking, in this example the divergence is positive because bigger arrows come out of a point than go in. So the field 'diverges'. (Once the concept of flux of a vector field is understood this will make more sense.)





Now consider the field $\underline{v} = \underline{\omega} \times \underline{r}$ where $\underline{\omega}$ is a constant vector. One can think of $\underline{v}$ as the velocity of a point in a rigid rotating body. We sketch a cross-section of the field $\underline{v}$ with $\underline{\omega}$ chosen to point out of the page. We can calculate $\underline{\nabla} \times \underline{v}$ as follows:

$$\begin{aligned}
\underline{\nabla} \times \left(\underline{\omega} \times \underline{r}\right) &= \underline{e}_i \, \epsilon_{ijk} \, \frac{\partial}{\partial x_j} \, \left(\underline{\omega} \times \underline{r}\right)_k \quad = \underline{e}_i \, \epsilon_{ijk} \, \frac{\partial}{\partial x_j} \, \epsilon_{k\ell m} \, \omega_\ell \, x_m \\
&= \underline{e}_i \, \left(\delta_{i\ell} \, \delta_{jm} - \delta_{im} \, \delta_{j\ell}\right) \, \omega_\ell \, \delta_{jm} \qquad \left(\text{since } \frac{\partial \omega_\ell}{\partial x_j} = 0\right) \\
&= \underline{e}_i \, \left(\omega_i \, \delta_{jj} - \delta_{ij} \, \omega_j\right) \; = \; \underline{e}_i \, 2 \, \omega_i \; = \; 2\underline{\omega}
\end{aligned}$$

Thus we obtain yet another *very* useful & important result:

$$\boxed{\underline{\nabla} \times \left(\underline{\omega} \times \underline{r}\right) = 2\underline{\omega}}$$

To understand intuitively the non-zero curl imagine that the flow lines are those of a rotating fluid with a small ball centred on a flow line of the field. The centre of the ball will follow the flow line. However the effect of the neighbouring flow lines is to make the ball rotate. Therefore the field has non-zero 'curl' and the axis of rotation gives the direction of the curl.

For this rotation-like field, the divergence is zero. To prove this, write the components of $\underline{\nabla} \cdot (\underline{\omega} \times \underline{r})$ as

$$\frac{\partial}{\partial x_i} \, \epsilon_{ijk} \omega_j r_k = \epsilon_{ijk} \omega_j \delta_{ik} = \epsilon_{jii} \omega_j = 0,$$

where we have differentiated r_k to get δ_{ik} and used the fact that $\epsilon_{ijk} = 0$ if two indices are equal. So our two examples are complementary: one has zero curl, the other has zero divergence.

Terminology:

1. If $\underline{\nabla} \cdot \underline{A}(\underline{r}) = 0$ in some region R, $\underline{A}$ is said to be **solenoidal** in R.

2. If $\underline{\nabla} \times \underline{A}(\underline{r}) = 0$ in some region R, $\underline{A}$ is said to be **irrotational** in R.

13.3 Vector operator identities

There are many identities involving div, grad, and curl. It is not necessary to know *all* of these, but you are advised to be able to produce from memory expressions for $\underline{\nabla} r$, $\underline{\nabla} \cdot \underline{r}$, $\underline{\nabla} \times \underline{r}$, $\underline{\nabla}\phi(r)$, $\underline{\nabla}(\underline{a} \cdot \underline{r})$, $\underline{\nabla} \times (\underline{a} \times \underline{r})$, $\underline{\nabla}(fg)$, and first four identities given below. You should be *familiar* with the rest and to be able to *derive* and *use* them when necessary.

Most importantly you should be at ease with div, grad and curl. This only comes through practice and deriving the various identities gives you just that. In these derivations the advantages of suffix notation, the summation convention and ϵ_{ijk} will become apparent.

In what follows, $\phi(\underline{r})$ is a scalar field; $\underline{A}(\underline{r})$ and $\underline{B}(\underline{r})$ are vector fields.

13.3.1 Distributive laws

1. $\underline{\nabla} \cdot (\underline{A} + \underline{B}) = \underline{\nabla} \cdot \underline{A} + \underline{\nabla} \cdot \underline{B}$
2. $\underline{\nabla} \times (\underline{A} + \underline{B}) = \underline{\nabla} \times \underline{A} + \underline{\nabla} \times \underline{B}$

The proofs of these are straightforward using suffix or 'x y z' notation and follow from the fact that div and curl are linear operations.

13.3.2 Product laws

The results of taking the div or curl of **products** of vector and scalar fields are predictable but need a little care:

3. $\underline{\nabla} \cdot (\phi\, \underline{A}) = \phi\, \underline{\nabla} \cdot \underline{A} + \underline{A} \cdot \underline{\nabla}\phi$
4. $\underline{\nabla} \times (\phi\, \underline{A}) = \phi\, (\underline{\nabla} \times \underline{A}) + (\underline{\nabla}\phi) \times \underline{A} = \phi\, (\underline{\nabla} \times \underline{A}) - \underline{A} \times \underline{\nabla}\phi$

Proof of (4):

$$\begin{aligned} \underline{\nabla} \times (\phi\, \underline{A}) &= \underline{e}_i\, \epsilon_{ijk}\, \frac{\partial}{\partial x_j}\, (\phi\, A_k) \\ &= \underline{e}_i\, \epsilon_{ijk} \left(\phi \left(\frac{\partial A_k}{\partial x_j} \right) + \left(\frac{\partial \phi}{\partial x_j} \right) A_k \right) \\ &= \phi\, (\underline{\nabla} \times \underline{A}) + (\underline{\nabla}\phi) \times \underline{A}. \end{aligned}$$

13.3.3 Products of two vector fields

Things start getting complicated:

5. $\underline{\nabla}\, (\underline{A} \cdot \underline{B}) = (\underline{A} \cdot \underline{\nabla})\, \underline{B} + (\underline{B} \cdot \underline{\nabla}) \underline{A} + \underline{A} \times (\underline{\nabla} \times \underline{B}) + \underline{B} \times (\underline{\nabla} \times \underline{A})$
6. $\underline{\nabla} \cdot (\underline{A} \times \underline{B}) = \underline{B} \cdot (\underline{\nabla} \times \underline{A}) - \underline{A} \cdot (\underline{\nabla} \times \underline{B})$

7. $\underline{\nabla} \times (\underline{A} \times \underline{B}) \;=\; \underline{A}\,(\underline{\nabla} \cdot \underline{B}) \;-\; \underline{B}\,(\underline{\nabla} \cdot \underline{A}) \;+\; (\underline{B} \cdot \underline{\nabla})\,\underline{A} \;-\; (\underline{A} \cdot \underline{\nabla})\,\underline{B}$

The trickiest of these is the one involving the triple vector product. Remembering 'BAC–CAB', we might be tempted to write $\underline{\nabla} \times (\underline{A} \times \underline{B}) = \underline{A}\,(\underline{\nabla} \cdot \underline{B}) \;-\; \underline{B}\,(\underline{\nabla} \cdot \underline{A})$; where do the extra terms come from? To see this, use the general technique for all such manipulations, which is to *write things in components, keeping terms in order.* So the 'BAC–CAB' rule actually says

$$[\underline{A} \times (\underline{B} \times \underline{C})]_i = A_j B_i C_j - A_j B_j C_i.$$

Thus, when $A_i = \partial/\partial x_i$, the derivative operates to the right and makes two terms from differentiating a product.

13.3.4 Identities involving 2 gradients

8. $\underline{\nabla} \times (\underline{\nabla}\phi) \;=\; 0$ — curl grad ϕ is always zero.

9. $\underline{\nabla} \cdot (\underline{\nabla} \times \underline{A}) \;=\; 0$ — div curl $\underline{A}$ is always zero.

10. $\underline{\nabla} \times (\underline{\nabla} \times \underline{A}) \;=\; \underline{\nabla}(\underline{\nabla} \cdot \underline{A}) \;-\; \nabla^2 \underline{A}$

Proofs are easily obtained in Cartesian coordinates using suffix notation:

Proof of (8)
$$\underline{\nabla} \times (\underline{\nabla}\,\phi) \;=\; \underline{e}_i\,\epsilon_{ijk}\,\frac{\partial}{\partial x_j}\,(\underline{\nabla}\,\phi)_k \;=\; \underline{e}_i\,\epsilon_{ijk}\,\frac{\partial}{\partial x_j}\,\frac{\partial}{\partial x_k}\,\phi.$$

Now, partial derivatives commute, and so we will get the same contribution $(\partial/\partial x_j)\,(\partial/\partial x_k)$ twice, with the order of j and k reversed in ϵ_{ijk}. This changes the sign of ϵ_{ijk}, and so the two terms exactly cancel each other.

Proof of (10)
$$[\underline{A} \times (\underline{B} \times \underline{C})]_i = A_j B_i C_j - A_j B_j C_i.$$

So now if the first two terms are derivatives, there is no product rule to apply. Moreover, A_j and B_i will commute, since partial derivative commute. This immediately lets us prove the result.

Finally, when a scalar field ϕ depends only on the magnitude of the position vector $r = |\underline{r}|$, we have

$$\nabla^2\,\phi(r) \;=\; \phi''(r) \;+\; \frac{2\phi'(r)}{r}$$

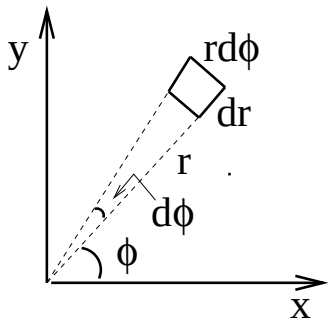
where the prime denotes differentiation with respect to r. Proof of this relation is left to the tutorial.

Lecture 14: Integrals over Fields

14.1 Polar co-ordinate systems

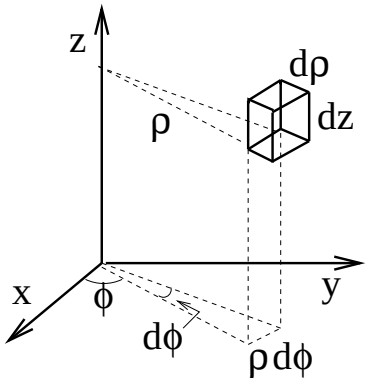
Before commencing with integral vector calculus we review here polar co-ordinate systems. Here dV indicates a volume element and dA an area element. Note that different conventions, *e.g.* for the angles ϕ and θ, are sometimes used.

Plane polar co-ordinates



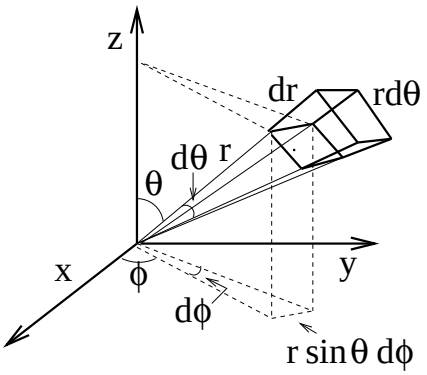
$$x = r\cos\phi$$
$$y = r\sin\phi$$
$$dA = r\,dr\,d\phi$$

Cylindrical polar co-ordinates



$$x = \rho\cos\phi$$
$$y = \rho\sin\phi$$
$$z = z$$
$$dV = \rho\,d\rho\,d\phi\,dz$$

Spherical polar co-ordinates



$$x = r\sin\theta\cos\phi$$
$$y = r\sin\theta\sin\phi$$
$$z = r\cos\theta$$
$$dV = r^2\sin\theta\,dr\,d\theta\,d\phi$$

14.2 Scalar and vector integration and line integrals

You should already be familiar with integration in $\mathbb{R}^1, \mathbb{R}^2, \mathbb{R}^3$. Here we review integration of a scalar field with an example.

Consider a hemisphere of radius a centred on the $\underline{e}_3$ axis and with bottom face at $z = 0$. If the mass density (a scalar field) is $\rho(r) = \sigma/r$ where σ is a constant, then what is the total mass?

It is most convenient to use spherical polars, so that

$$M = \int_{\text{hemisphere}} \rho(\underline{r})dV = \int_0^a r^2\rho(r)dr \int_0^{\pi/2} \sin\theta d\theta \int_0^{2\pi} d\phi = 2\pi\sigma \int_0^a r dr = \pi\sigma a^2$$

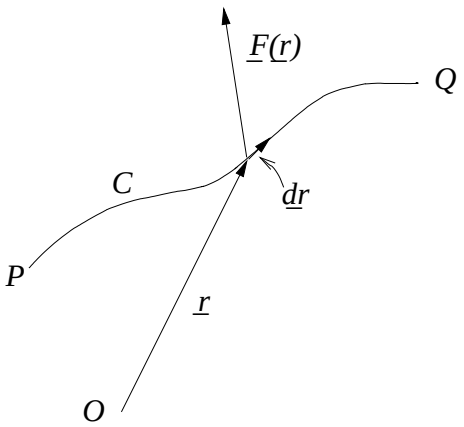
Now consider the centre of mass vector

$$M\underline{R} = \int_V \underline{r}\rho(\underline{r})dV$$

This is our first example of integrating a vector field (here $\underline{r}\rho(\underline{r})$). To do so simply integrate each component using $\underline{r} = r\sin\theta\cos\phi\underline{e}_1 + r\sin\theta\sin\phi\underline{e}_2 + r\ \cos\theta\underline{e}_3$

$$\begin{aligned}
MX &= \int_0^a r^3\rho(r)dr\int_0^{\pi/2}\sin^2\theta d\theta\int_0^{2\pi}\cos\phi\ d\phi = 0 \quad \text{since } \phi \text{ integral gives 0}\\
MY &= \int_0^a r^3\rho(r)dr\int_0^{\pi/2}\sin^2\theta d\theta\int_0^{2\pi}\sin\phi\ d\phi = 0 \quad \text{since } \phi \text{ integral gives 0}\\
MZ &= \int_0^a r^3\rho(r)dr\int_0^{\pi/2}\sin\theta\cos\theta d\theta\int_0^{2\pi}d\phi = 2\pi\sigma\int_0^a r^2dr\int_0^{\pi/2}\frac{\sin 2\theta}{2}d\theta\\
&= \frac{2\pi\sigma a^3}{3}\left[\frac{-\cos 2\theta}{4}\right]_0^{\pi/2} = \frac{\pi\sigma a^3}{3} \qquad \Rightarrow \underline{R} = \frac{a}{3}\underline{e}_3
\end{aligned}$$

14.3 Line integrals



As an example, consider a particle constrained to move on a wire. Only the component of the force along the wire does any work. Therefore the work done in moving the particle from $\underline{r}$ to $\underline{r} + d\underline{r}$ is

$$dW = \underline{F}\cdot d\underline{r}\ .$$

The total work done in moving particle along a wire which follows some curve C between two points P, Q is

$$W_C = \int_P^Q dW = \int_C \underline{F}(\underline{r})\cdot d\underline{r}\ .$$

This is a line integral along the curve C.

More generally let $\underline{A}(\underline{r})$ be a vector field defined in the region R, and let C be a curve in R joining two points P and Q. $\underline{r}$ is the position vector at some point on the curve; $\underline{dr}$ is an infinitesimal vector *along* the curve at $\underline{r}$.

The magnitude of $\underline{dr}$ is the infinitesimal **arc length**: $ds = \sqrt{\underline{dr}\cdot\underline{dr}}$.

We define $\hat{\underline{t}}$ to be the **unit vector** tangent to the curve at $\underline{r}$ (points in the direction of $\underline{dr}$)

$$\hat{\underline{t}} = \frac{d\underline{r}}{ds}$$

Note that, in general, $\displaystyle\int_C \underline{A}\cdot\underline{dr}$ **depends on the path** joining P and Q.

In Cartesian coordinates, we have

$$\int_C \underline{A}\cdot\underline{dr} = \int_C A_i dx_i = \int_C (A_1dx_1 + A_2dx_2 + A_3dx_3)$$

14.4 Parametric representation of a line integral

Often a curve in 3D can be parameterised by a single parameter e.g. if the curve were the trajectory of a particle then time would be the parameter. Sometimes the parameter of a line integral is chosen to be the arc-length s along the curve C.

Generally for parameterisation by λ (varying from λ_P to λ_Q)

$$x_i \;=\; x_i(\lambda), \quad \text{with } \lambda_P \le \lambda \le \lambda_Q$$

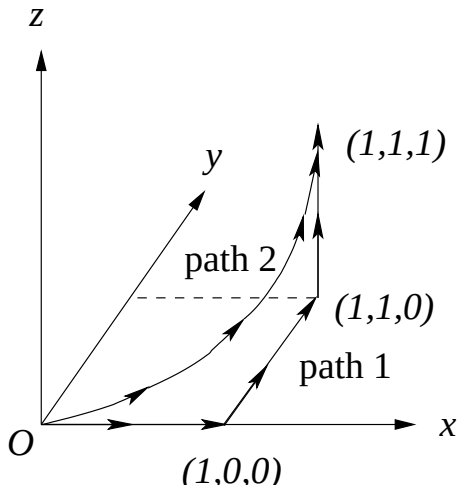
then

$$\int_C \underline{A}\cdot\underline{dr} \;=\; \int_{\lambda_P}^{\lambda_Q}\left(\underline{A}\cdot\frac{\underline{dr}}{d\lambda}\right)d\lambda = \int_{\lambda_P}^{\lambda_Q}\left(A_1\,\frac{dx_1}{d\lambda} \,+\, A_2\,\frac{dx_2}{d\lambda} \,+\, A_3\,\frac{dx_3}{d\lambda}\right)\,d\lambda$$

If necessary, the curve C may be subdivided into sections, each with a different parameterisation (piecewise smooth curve).

Example: $\underline{A} = (3x^2+6y)\,\underline{e}_1 - 14yz\underline{e}_2 \,+ 20xz^2\underline{e}_3$. Evaluate $\displaystyle\int_C \underline{A}\cdot\underline{dr}$ between the points with Cartesian coordinates $(0,0,0)$ and $(1,1,1)$, along the paths C:

1. $(0,0,0)\to(1,0,0)\to(1,1,0)\to(1,1,1)$ (straight lines).
2. $x=\lambda$, $y=\lambda^2$ $z=\lambda^3$; from $\lambda=0$ to $\lambda=1$.



1. • Along the line from $(0,0,0)$ to $(1,0,0)$, we have $y=z=0$, so $dy=dz=0$, hence $\underline{dr}=\underline{e}_1\,dx$ and $\underline{A}=3x^2\,\underline{e}_1$, (here the parameter is x):

$$\int_{(0,0,0)}^{(1,0,0)}\underline{A}\cdot\underline{dr} \;=\; \int_{x=0}^{x=1}3x^2\,dx \;=\; \left[x^3\right]_0^1 \;=\; 1$$

• Along the line from $(1,0,0)$ to $(1,1,0)$, we have $x=1$, $dx=0$, $z=dz=0$, so $\underline{dr}=\underline{e}_2\,dy$ (here the parameter is y) and

$$\underline{A} \;=\; \left(3x^2+6y\right)\big|_{x=1}\,\underline{e}_1 \;=\; (3+6y)\,\underline{e}_1.$$

$$\int_{(1,0,0)}^{(1,1,0)}\underline{A}\cdot\underline{dr} \;=\; \int_{y=0}^{y=1}(3+6y)\;\underline{e}_1\cdot\underline{e}_2\,dy \;=\; 0.$$

- Along the line from $(1,1,0)$ to $(1,1,1)$, we have $x = y = 1$, $dx = dy = 0$, and hence $\underline{dr} = \underline{e}_3\, dz$ and $\underline{A} = 9\,\underline{e}_1 - 14z\,\underline{e}_2 + 20z^2\,\underline{e}_3$, therefore

$$\int_{(1,1,0)}^{(1,1,1)} \underline{A}\cdot\underline{dr} \;=\; \int_{z=0}^{z=1} 20z^2\,dz \;=\; \left[\frac{20}{3}z^3\right]_0^1 \;=\; \frac{20}{3}$$

Adding up the 3 contributions we get

$$\int_C \underline{A}\cdot\underline{dr} \;=\; 1 \;+\; 0 \;+\; \frac{20}{3} \;=\; \frac{23}{3} \qquad \text{along path (1)}$$

2. To integrate $\underline{A} = (3x^2+6y)\,\underline{e}_1 - 14yz\underline{e}_2 + 20xz^2\underline{e}_3$ along path (2) (where the parameter is λ), we write

$$\begin{aligned} \underline{r} &= \lambda\,\underline{e}_1 + \lambda^2\,\underline{e}_2 + \lambda^3\,\underline{e}_3 \\ \frac{d\underline{r}}{d\lambda} &= \underline{e}_1 \;+\; 2\lambda\,\underline{e}_2 \;+\; 3\lambda^2\,\underline{e}_3 \\ \underline{A} &= \left(3\lambda^2 + 6\lambda^2\right)\,\underline{e}_1 \;-\; 14\lambda^5\,\underline{e}_2 \;+\; 20\lambda^7\,\underline{e}_3 \quad \text{so that} \end{aligned}$$

$$\int_C \left(\underline{A}\cdot\frac{d\underline{r}}{d\lambda}\right) d\lambda = \int_{\lambda=0}^{\lambda=1} \left(9\lambda^2 \;-\; 28\lambda^6 \;+\; 60\lambda^9\right)\,d\lambda = \left[3\lambda^3 \;-\; 4\lambda^7 + 6\lambda^{10}\right]_0^1 \;=\; 5$$

$$\text{Hence} \qquad \int_C \underline{A}\cdot\underline{dr} \;=\; 5 \qquad \text{along path (2)}$$

In this case, the integral of $\underline{A}$ from $(0,0,0)$ to $(1,1,1)$ depends on the path taken.

The line integral $\displaystyle\int_C \underline{A}\cdot\underline{dr}$ is a **scalar** quantity. Another **scalar** line integral is $\displaystyle\int_C f\,ds$ where $f(\underline{r})$ is a scalar field and ds is the infinitesimal arc-length introduced earlier.

Line integrals around a **simple** (doesn't intersect itself) **closed** curve C are denoted by $\displaystyle\oint_C$

$$\text{e.g.} \qquad \oint_C \underline{A}\cdot\underline{dr} \qquad \equiv \text{the } \textbf{circulation} \text{ of } \underline{A} \text{ around } C$$

We can also define **vector** line integrals e.g.

1. $\displaystyle\int_C \underline{A}\,ds \;=\; \underline{e}_i \int_C A_i\,ds$ in Cartesian coordinates.
2. $\displaystyle\int_C \underline{A}\times\underline{dr} \;=\; \underline{e}_i\,\epsilon_{ijk}\int_C A_j\,dx_k$ in Cartesians.

Example : Consider a current of magnitude I flowing along a wire following a closed path C. The magnetic force on an element $\underline{dr}$ of the wire is $I\underline{dr}\times\underline{B}$ where $\underline{B}$ is the magnetic

field at $\underline{r}$. Let $\underline{B}(\underline{r}) = x\,\underline{e}_1 + y\,\underline{e}_2$. Evaluate $\oint_C \underline{B} \times \underline{dr}$ for a circular current loop of radius a in the $x-y$ plane, centred on the origin.

$$\begin{aligned} \underline{B} &= a\cos\phi\,\underline{e}_1 + a\sin\phi\,\underline{e}_2 \\ \underline{dr} &= (-a\sin\phi\,\underline{e}_1 + a\cos\phi\,\underline{e}_2)\,d\phi \end{aligned}$$

Hence $$\oint_C \underline{B} \times \underline{dr} = \int_0^{2\pi} \left(a^2\cos^2\phi + a^2\sin^2\phi\right) \underline{e}_3\,d\phi = \underline{e}_3\,a^2 \int_0^{2\pi} d\phi = 2\pi a^2\,\underline{e}_3$$

Lecture 15: The scalar potential

Consider again the work done by a force. If the force is *conservative*, i.e. total energy is conserved, then the work done is equal to minus the change in potential energy

$$dV = -dW = -\underline{F}\cdot\underline{dr} = -F_i dx_i$$

Now we can also write dV as

$$dV = \frac{\partial V}{\partial x_i} dx_i = (\underline{\nabla} V)_i dx_i$$

Therefore we can identify $\boxed{\underline{F} = -\underline{\nabla} V}$

Thus the force is minus the gradient of the (scalar) potential. The minus sign is conventional and chosen so that potential energy decreases as the force does work.

In this example we knew that a potential existed (we postulated conservation of energy). More generally we would like to know under what conditions can a vector field $\underline{A}(\underline{r})$ be written as the gradient of a scalar field ϕ, *i.e.* when does $\underline{A}(\underline{r}) = (\pm)\,\underline{\nabla}\,\phi(\underline{r})$ hold?

Aside: A **simply connected region**, R, is one for which every closed curve in R can be shrunk continuously to a point while remaining entirely in R. The inside of a sphere is simply connected while the region between two concentric cylinders is **not** simply connected: it is doubly connected. For this course we shall be concerned with simply connected regions

15.1 Theorems on scalar potentials

For a vector field $\underline{A}(\underline{r})$ defined in a simply connected region R, the following three statements are equivalent, *i.e.* **any one implies the other two**:

1. $\underline{A}(\underline{r})$ can be written as the **gradient** of a **scalar potential** $\phi(\underline{r})$

$$\boxed{\underline{A}(\underline{r}) = \underline{\nabla}\,\phi(\underline{r}) \quad \text{with} \quad \phi(\underline{r}) = \int_{\underline{r}_0}^{\underline{r}} \underline{A}(\underline{r}')\cdot\underline{dr}'}$$

   where $\underline{r}_0$ is some arbitrary fixed point in R.

2. (a) $\oint_C \underline{A}(\underline{r}')\cdot\underline{dr}' = 0$, where C is any **closed** curve in R

(b) $\phi(\underline{r}) \equiv \int_{\underline{r}_0}^{\underline{r}} \underline{A}(\underline{r}') \cdot \underline{dr}'$ does not depend on the path between $\underline{r}_0$ and $\underline{r}$.

3. $\underline{\nabla} \times \underline{A}(\underline{r}) = 0$ for all points $\underline{r} \in R$

Proof that (2) implies (1)
Consider two neighbouring points $\underline{r}$ and $\underline{r} + \underline{dr}$, *define* the potential as an integral that is independent of path:

$$\phi(\underline{r}) = \int_{\underline{r}_0}^{\underline{r}} \underline{A}(\underline{r}') \cdot \underline{dr}'.$$

The starting point, $\underline{r}_0$ is arbitrary, so the potential can always have an arbitrary constant added to it. Now, the change in ϕ corresponding to a change in $\underline{r}$ is

$$d\phi(\underline{r}) = \underline{A}(\underline{r}) \cdot \underline{dr}.$$

But, by Taylor's theorem, we also have

$$d\phi(\underline{r}) = \frac{\partial \phi(\underline{r})}{\partial x_i}\, dx_i = \underline{\nabla}\, \phi(\underline{r}) \cdot \underline{dr}$$

Comparing the two different equations for $d\phi(\underline{r})$, which hold for all $\underline{dr}$, we deduce

$$\underline{A}(\underline{r}) = \underline{\nabla}\, \phi(\underline{r})$$

Thus we have shown that **path independence** implies the existence of a scalar potential ϕ for the vector field $\underline{A}$. (Also path independence implies 2(a)).

Proof that (1) implies (3) (the easy bit)

$$\underline{A} = \underline{\nabla}\, \phi \quad \Rightarrow \quad \underline{\nabla} \times \underline{A} = \underline{\nabla} \times \left(\underline{\nabla}\, \phi\right) \equiv 0$$

because curl (grad ϕ) is identically zero (ie it is zero for *any* scalar field ϕ).

Proof that (3) implies (2) (the hard bit)
We defer the proof until we have met Stokes' theorem.

Terminology: A vector field is

- **irrotational** if $\underline{\nabla} \times \underline{A}(\underline{r}) = 0$.
- **conservative** if $\underline{A}(\underline{r}) = \underline{\nabla}\phi$.
- For simply connected regions we have shown irrotational and conservative are synonymous. But note that for a multiply connected region this is not the case.

15.2 Finding scalar potentials

We have shown that the scalar potential $\phi(\underline{r})$ for a *conservative* vector field $\underline{A}(\underline{r})$ can be constructed from a line integral which is *independent* of the path of integration between the endpoints. Therefore, a convenient way of evaluating such integrals is to integrate along a **straight line** between the points $\underline{r}_0$ and $\underline{r}$. Choosing $\underline{r}_0 = 0$, we can write this integral in parametric form as follows:

$$\underline{r}' = \lambda\, \underline{r} \quad \text{where} \quad \{0 \le \lambda \le 1\} \quad \text{so} \quad \underline{dr}' = d\lambda\, \underline{r} \quad \text{and therefore}$$

$$\phi(\underline{r}) \;=\; \int_{\lambda=0}^{\lambda=1} \underline{A}(\lambda\,\underline{r})\cdot(d\lambda\,\underline{r})$$

Example: Let $\underline{A}(\underline{r}) \;=\; 2\,(\underline{a}\cdot\underline{r})\,\underline{r} \;+\; r^2\,\underline{a}$ where $\underline{a}$ is a constant vector.
It is straightforward to show that $\underline{\nabla}\times\underline{A}=0$. Thus

$$\begin{aligned}
\phi(\underline{r}) \;&=\; \int_0^{\underline{r}} \underline{A}(\underline{r}')\cdot\underline{dr}' \;=\; \int_0^1 \underline{A}(\lambda\,\underline{r})\cdot(d\lambda\,\underline{r}) \\
&=\; \int_0^1 \Big[2\,(\underline{a}\cdot\lambda\,\underline{r})\,\lambda\,\underline{r} \;+\; \lambda^2 r^2\,\underline{a}\Big]\cdot(d\lambda\,\underline{r}) \\
&=\; \Big[2\,(\underline{a}\cdot\underline{r})\,\underline{r}\cdot\underline{r} \;+\; r^2\,(\underline{a}\cdot\underline{r})\Big]\int_0^1 \lambda^2\,d\lambda \\
&=\; r^2\,(\underline{a}\cdot\underline{r})
\end{aligned}$$

Sometimes it is possible to see the answer without constructing it:

$$\underline{A}(\underline{r}) \;=\; 2\,(\underline{a}\cdot\underline{r})\,\underline{r} \;+\; r^2\,\underline{a} \;=\; (\underline{a}\cdot\underline{r})\underline{\nabla}r^2 \;+\; r^2\,\underline{\nabla}(\underline{a}\cdot\underline{r}) \;=\; \underline{\nabla}\Big((\underline{a}\cdot\underline{r})\,r^2 \;+\; \text{const}\Big)$$

in agreement with what we had before if we choose const $=\;0$. While this method is not as systematic as Method 1, it can be quicker if you spot the trick.

15.3 Conservative forces: conservation of energy

Let us now see how the name *conservative field* arises. Consider a vector field $\underline{F}(\underline{r})$ corresponding to the only force acting on some test particle of mass m. We will show that for a conservative force (where we can write $\underline{F}=-\underline{\nabla}V$) the total energy is **constant** in time.

Proof: The particle moves under the influence of Newton's Second Law:

$$m\ddot{\underline{r}} = \underline{F}(\underline{r}).$$

Consider a small displacement $\underline{dr}$ along the path taking time dt. Then

$$m\ddot{\underline{r}}\cdot\underline{dr} = \underline{F}(\underline{r})\cdot\underline{dr} = -\underline{\nabla}V(\underline{r})\cdot\underline{dr}.$$

Integrating this expression along the path from $\underline{r}_A$ at time $t=t_A$ to $\underline{r}_B$ at time $t=t_B$ yields

$$m\int_{\underline{r}_A}^{\underline{r}_B} \ddot{\underline{r}}\cdot\underline{dr} = -\int_{\underline{r}_A}^{\underline{r}_B} \underline{\nabla}V(\underline{r})\cdot\underline{dr}.$$

We can simplify the left-hand side of this equation to obtain

$$m\int_{\underline{r}_A}^{\underline{r}_B} \ddot{\underline{r}}\cdot\underline{dr} = m\int_{t_A}^{t_B} \ddot{\underline{r}}\cdot\dot{\underline{r}}\,dt = m\int_{t_A}^{t_B} \tfrac{1}{2}\,\tfrac{d}{dt}\,\dot{\underline{r}}^2 dt = \tfrac{1}{2}m[v_B^2 - v_A^2],$$

where v_A and v_B are the magnitudes of the velocities at points A and B respectively.

The right-hand side simply gives

$$-\int_{\underline{r}_A}^{\underline{r}_B} \underline{\nabla} V(\underline{r}) \cdot \underline{dr} = -\int_{\underline{r}_A}^{\underline{r}_B} dV = V_A - V_B$$

where V_A and V_B are the values of the potential V at $\underline{r}_A$ and $\underline{r}_B$, respectively. Therefore

$$\frac{1}{2}mv_A^2 + V_A = \frac{1}{2}mv_B^2 + V_B$$

and the total energy $\quad E = \frac{1}{2}mv^2 + V \quad$ is **conserved**, i.e. *constant in time*.

15.4 Physical examples of conservative forces

Newtonian Gravity and the *electrostatic force* are both conservative. *Frictional forces* are not conservative; energy is dissipated and work is done in traversing a closed path. In general, time-dependent forces are not conservative.

The foundation of Newtonian Gravity is **Newton's Law of Gravitation**. The force $\underline{F}$ on a particle of mass m_1 at $\underline{r}$ due to a particle of mass m at the origin is given by

$$\underline{F} = -\frac{G\,m\,m_1}{r^2}\,\hat{\underline{r}},$$

where $G \simeq 6.673 \times 10^{-11}\,\mathrm{N\,m^2\,kg^2}$ is Newton's Gravitational Constant.

The **gravitational field** $\underline{G}(\underline{r})$ (due to the mass at the origin) is formally defined as

$$\underline{G}(\underline{r}) = \lim_{m_1 \to 0} \frac{\underline{F}(\underline{r})}{m_1}.$$

so that the gravitational field due to the test mass m_1 can be ignored. The **gravitational potential** can be obtained by spotting the direct integration for $\underline{G} = -\underline{\nabla}\phi$

$$\phi = -\frac{Gm}{r}\ .$$

Alternatively, to calculate by a line integral choose $\underline{r}_0 = \infty$ then

$$\begin{aligned} \phi(\underline{r}) &= -\int_{\infty}^{\underline{r}} \underline{G}(\underline{r}') \cdot \underline{dr}' = -\int_{\infty}^{1} \underline{G}(\lambda\underline{r}) \cdot d\lambda\underline{r} \\ &= \int_{\infty}^{1} \frac{Gm\,(\hat{\underline{r}} \cdot \underline{r})}{r^2}\frac{d\lambda}{\lambda^2} = -\frac{Gm}{r} \end{aligned}$$

NB In this example the vector field $\underline{G}$ is singular at the origin $\underline{r} = 0$. This implies we have to exclude the origin and it is not possible to obtain the scalar potential at $\underline{r}$ by integration along a path from the origin. Instead we integrate from infinity, which in turn means that the gravitational potential at infinity is zero.

NB Since $\underline{F} = m_1\underline{G} = -\underline{\nabla}(m_1\phi)$ the potential energy of the mass m_1 is $V = m_1\phi$. The distinction (a convention) between potential and potential energy is a common source of confusion.

Electrostatics: Coulomb's Law states that the force $\underline{F}$ on a particle of charge q_1 at $\underline{r}$ in the electric field $\underline{E}$ due to a particle of charge q at the origin is given by

$$\underline{F} = q_1 \underline{E} = \frac{q_1\, q}{4\pi\epsilon_0\, r^2}\, \underline{\hat{r}}$$

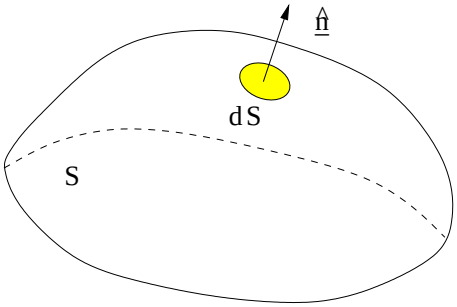
where $\epsilon_0 = 8.854\,187\,817\cdots \times 10^{-12}\, C^2 N^{-1} m^{-2}$ is the **Permittivity of Free Space** and the 4π is conventional. More strictly,

$$\underline{E}(\underline{r}) = \lim_{q_1 \to 0} \frac{\underline{F}(\underline{r})}{q_1}.$$

The **electrostatic potential** is taken as $\phi = 1/(4\pi\epsilon_0 r)$ (obtained by integrating $E = -\underline{\nabla}\phi$ from infinity to $\underline{r}$) and the potential energy of a charge q_1 in the electric field is $V = q_1\phi$.

Note that mathematically electrostatics and gravitation are very similar, the only real difference being that gravity between two masses is always attractive, whereas like charges repel.

Lecture 16: Surface integrals



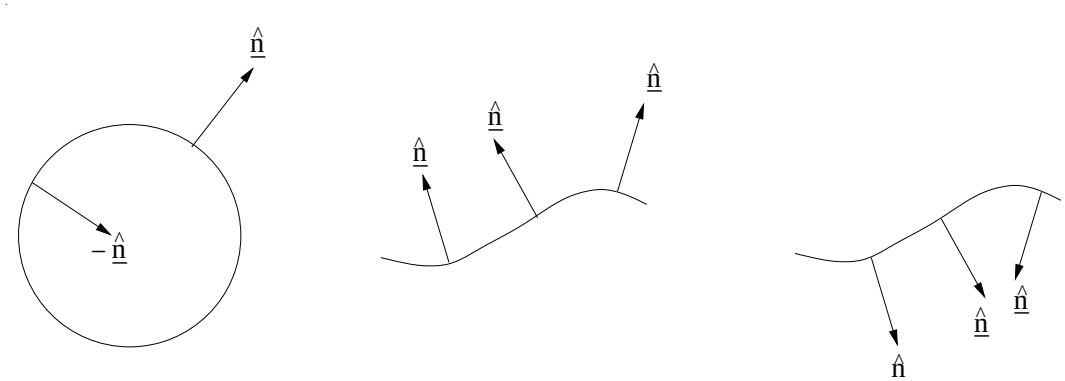
Let S be a two-sided surface in ordinary three-dimensional space as shown. If an infinitesimal element of surface with (scalar) area dS has unit normal $\underline{\hat{n}}$, then the infinitesimal **vector element of area** is *defined* by

$$\boxed{\underline{dS} \;=\; \underline{\hat{n}}\, dS}$$

Example: if S lies in the (x, y) plane, then $\underline{dS} = \underline{e}_3\, dx\, dy$ in Cartesian coordinates.

Physical interpretation: $\underline{dS} \cdot \underline{\hat{a}}$ gives the projected (scalar) element of area onto the plane with unit normal $\underline{\hat{a}}$.

For **closed** surfaces (eg, a sphere) we *choose* $\underline{\hat{n}}$ to be the **outward normal**. For **open** surfaces, the sense of $\underline{\hat{n}}$ is arbitrary — except that it is chosen in the same sense for all elements of the surface.



If $\underline{A}(\underline{r})$ is a vector field defined on S, we define the (normal) **surface integral**

$$\int_S \underline{A} \cdot \underline{dS} \;=\; \int_S \left(\underline{A} \cdot \hat{\underline{n}}\right)\, dS \;=\; \lim_{\substack{m \to \infty \\ \delta S \to 0}} \sum_{i=1}^{m} \left(\underline{A}(\underline{r}^{\,i}) \cdot \hat{\underline{n}}^{\,i}\right)\, \delta S^i$$

where we have formed the Riemann sum by dividing the surface S into m small areas, the ith area having vector area $\underline{\delta S}^i$. Clearly, the quantity $\underline{A}(\underline{r}^{\,i}) \cdot \hat{\underline{n}}^{\,i}$ is the component of $\underline{A}$ *normal* to the surface at the point $\underline{r}^{\,i}$

Note that the integral over S is really a double integral, since it is an integral over a 2D surface. Sometimes the integral over a *closed* surface is denoted by $\oint_S \underline{A} \cdot \underline{dS}$.

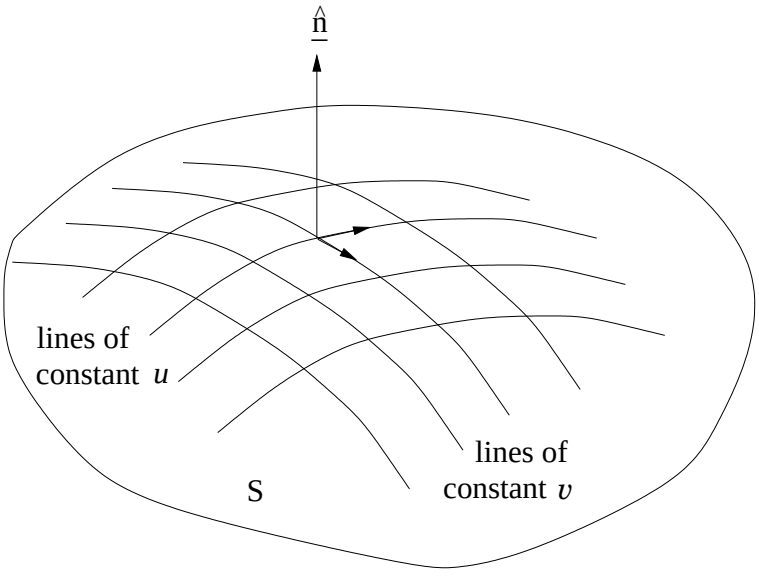
16.1 Parametric form of the surface integral

Often, we will need to carry out surface integrals explicitly, and we need a procedure for turning them into double integrals. Suppose the points on a surface S are defined by **two** real parameters u and v:-

$\underline{r} \;=\; \underline{r}(u,v) \;=\; (x(u,v),\, y(u,v),\, z(u,v))$ then

- the lines $\underline{r}(u,v)$ for fixed u, variable v, and
- the lines $\underline{r}(u,v)$ for fixed v, variable u

are **parametric lines** and form a **grid** on the surface S as shown. In other words, u and v form a coordinate system on the surface – although usually not a Cartesian one.



If we change u and v by du and dv respectively, then $\underline{r}$ changes by $\underline{dr}$:-

$$\underline{dr} \;=\; \frac{\partial \underline{r}}{\partial u}\, du \;+\; \frac{\partial \underline{r}}{\partial v}\, dv,$$

so that there are two linearly independent vectors generated by varying either u or v. The vector element of area, $\underline{dS}$, generated by these two vectors has magnitude equal to the area of

the infinitesimal parallelogram shown in the figure, and points perpendicular to the surface:

$$\underline{dS} \;=\; \left(\frac{\partial \underline{r}}{\partial u}\,du\right) \times \left(\frac{\partial \underline{r}}{\partial v}\,dv\right) = \left(\frac{\partial \underline{r}}{\partial u} \times \frac{\partial \underline{r}}{\partial v}\right)\,du\,dv$$

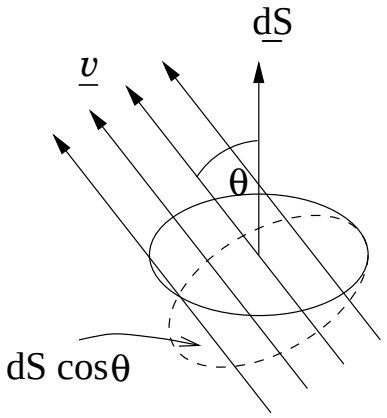
$$\boxed{\underline{dS} = \left(\frac{\partial \underline{r}}{\partial u} \times \frac{\partial \underline{r}}{\partial v}\right)\,du\,dv}$$

Finally, our integral is parameterised as

$$\boxed{\int_S \underline{A}\cdot\underline{dS} \;=\; \int_v\int_u \underline{A}\cdot\left(\frac{\partial \underline{r}}{\partial u} \times \frac{\partial \underline{r}}{\partial v}\right)\,du\,dv}$$

We will give an explicit example of this below, where the coordinates are the spherical polar angles.

16.2 The concept of flux



Let $\underline{v}(\underline{r})$ be the velocity at a point $\underline{r}$ in a moving fluid. In a small region, where $\underline{v}$ is approximately constant, the **volume** of fluid crossing the element of vector area $\underline{dS} = \underline{\hat{n}}\,dS$ in time dt is

$$\left(|\underline{v}|\;dt\right)\;(dS\,\cos\theta) \;=\; \left(\underline{v}\cdot\underline{dS}\right)\;dt$$

since the area *normal* to the direction of flow is $\underline{\hat{v}}\cdot\underline{dS} = dS\,\cos\theta$.

Therefore

$$\underline{v}\cdot\underline{dS} \;=\; \textbf{volume per unit time of fluid crossing } \underline{dS}$$

$$\text{hence}\quad \int_S \underline{v}\cdot\underline{dS} \;=\; \textbf{volume per unit time of fluid crossing a finite surface } S$$

More generally, for a vector field $\underline{A}(\underline{r})$:

$$\boxed{\text{The surface integral } \int_S \underline{A}\cdot\underline{dS} \text{ is called the } \textbf{flux} \text{ of } \underline{A} \text{ through the surface } S.}$$

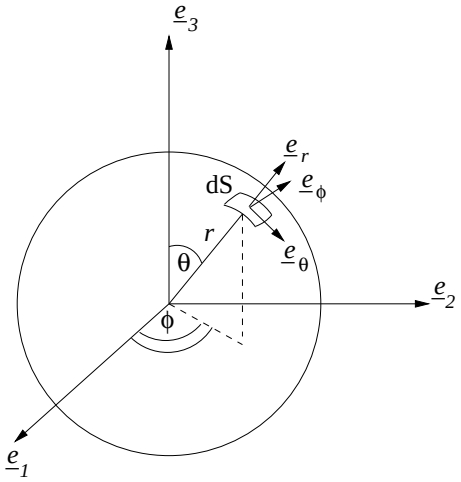
The concept of flux is useful in many different contexts e.g. flux of molecules in an gas; electromagnetic flux etc

Example: Let S be the surface of sphere $x^2 + y^2 + z^2 = a^2$. Evaluate the total flux of the vector field $\underline{A} = \underline{\hat{r}}/r^2$ out of the sphere.

This is easy, since $\underline{A}$ and $\underline{dS}$ are parallel, so $\underline{A} \cdot \underline{dS} = dS/r^2$. Therefore, we want the total area of the surface of the sphere, divided by r^2, giving 4π. Let's now prove this much more pedantically, by constructing an explicit expression for $\underline{dS}$ and carrying out the integral.

An arbitrary point $\underline{r}$ on S may be parameterised by spherical polar co-ordinates θ and ϕ

$$\begin{aligned} \underline{r} &= a\sin\theta\cos\phi\,\underline{e}_1 + a\sin\theta\sin\phi\,\underline{e}_2 + a\cos\theta\,\underline{e}_3 \qquad \{0 \le \theta \le \pi,\ 0 \le \phi \le 2\pi\} \\ \text{so} \quad \frac{\partial \underline{r}}{\partial\theta} &= a\cos\theta\cos\phi\,\underline{e}_1 + a\cos\theta\sin\phi\,\underline{e}_2 - a\sin\theta\,\underline{e}_3 \\ \text{and} \quad \frac{\partial \underline{r}}{\partial\phi} &= -a\sin\theta\sin\phi\,\underline{e}_1 + a\sin\theta\cos\phi\,\underline{e}_2 + 0\,\underline{e}_3 \end{aligned}$$



Therefore

$$\begin{aligned} \frac{\partial \underline{r}}{\partial\theta} \times \frac{\partial \underline{r}}{\partial\phi} &= \begin{vmatrix} \underline{e}_1 & \underline{e}_2 & \underline{e}_3 \\ a\cos\theta\cos\phi & a\cos\theta\sin\phi & -a\sin\theta \\ -a\sin\theta\sin\phi & +a\sin\theta\cos\phi & 0 \end{vmatrix} \\ &= a^2\sin^2\theta\cos\phi\,\underline{e}_1 + a^2\sin^2\theta\sin\phi\,\underline{e}_2 + a^2\sin\theta\cos\theta\left[\cos^2\phi + \sin^2\phi\right]\underline{e}_3 \\ &= a^2\sin\theta\,(\sin\theta\cos\phi\,\underline{e}_1 + \sin\theta\sin\phi\,\underline{e}_2 + \cos\theta\,\underline{e}_3) \\ &= a^2\sin\theta\,\underline{\hat{r}} \\ \underline{\hat{n}} &= \underline{\hat{r}} \\ \underline{dS} &= \frac{\partial \underline{r}}{\partial\theta} \times \frac{\partial \underline{r}}{\partial\phi}\,d\theta\,d\phi = a^2\sin\theta d\theta\,d\phi\,\underline{\hat{r}} \end{aligned}$$

On the surface S, $r = a$ and the vector field $\underline{A}(\underline{r}) = \underline{\hat{r}}/a^2$. Thus the flux of $\underline{A}$ is

$$\int_S \underline{A} \cdot \underline{dS} = \int_0^{\pi} \sin\theta d\theta \int_0^{2\pi} d\phi = 4\pi$$

Spherical basis: The normalised vectors (shown in the figure)

$$\underline{e}_\theta = \frac{\partial \underline{r}}{\partial\theta} \Big/ \left|\frac{\partial \underline{r}}{\partial\theta}\right| \quad ; \quad \underline{e}_\phi = \frac{\partial \underline{r}}{\partial\phi} \Big/ \left|\frac{\partial \underline{r}}{\partial\phi}\right| \quad ; \quad \underline{e}_r = \underline{\hat{r}}$$

form an orthonormal set. This is the basis for spherical polar co-ordinates and is an example of a non-Cartesian basis since the $\underline{e}_\theta, \underline{e}_\phi, \underline{e}_r$ depend on position $\underline{r}$.

16.3 Other surface integrals

If $f(\underline{r})$ is a scalar field, a scalar surface integral is of the form

$$\int_S f \, dS,$$

of which the **surface area**, $\displaystyle\int_S dS$, of the surface S is a special case.

We may also define vector surface integrals

$$\int_S f \, \underline{dS} \qquad \int_S \underline{A} \, dS \qquad \int_S \underline{A} \times \underline{dS}$$

Each of these is a double integral, and is evaluated in a similar fashion to the scalar integrals, the result being a vector in each case.

The **vector area** of a surface is defined as $\underline{S} = \displaystyle\int_S \underline{dS}$. For a closed surface this is always zero.

Example: the vector area of an (open) hemisphere of radius a is found using spherical polars to be

$$\underline{S} = \int_S \underline{dS} = \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi/2} a^2 \sin\theta \underline{e}_r \, d\theta \, d\phi \, .$$

Using $\underline{e}_r = \sin\theta\cos\phi \, \underline{e}_1 + \sin\theta \sin\phi \, \underline{e}_2 + \cos\theta \, \underline{e}_3$ we obtain

$$\begin{aligned} \underline{S} &= \underline{e}_1 \, a^2 \int_0^{\pi/2} \sin^2\theta d\theta \int_0^{2\pi} \cos\phi d\phi \; + \; \underline{e}_2 \, a^2 \int_0^{\pi/2} \sin^2\theta d\theta \int_0^{2\pi} \sin\phi d\phi \\ &\quad + \underline{e}_3 \, a^2 \int_0^{\pi/2} \sin\theta\cos\theta d\theta \int_0^{2\pi} d\phi \\ &= 0 + 0 + \underline{e}_3 \pi a^2 \end{aligned}$$

The vector surface of the full sphere is zero since the contributions from upper and lower hemispheres cancel; also the vector area of a *closed* hemisphere is zero since the vector area of the bottom face is $-\underline{e}_3 \pi a^2$.

Lecture 17: Volume integrals and the divergence theorem

17.1 Parametric form of volume integrals

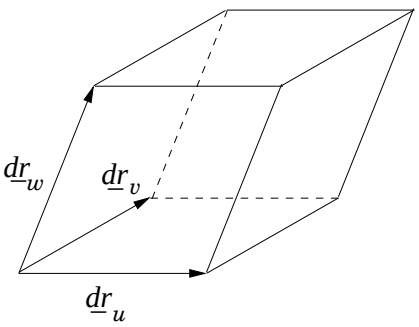
Here we discuss the *parametric* form of volume integrals. Suppose we can write $\underline{r}$ in terms of three real parameters u, v and w, so that $\underline{r} = \underline{r}(u, v, w)$. If we make a small change in each of these parameters, then $\underline{r}$ changes by

$$\underline{dr} \; = \; \frac{\partial \underline{r}}{\partial u} \, du \; + \; \frac{\partial \underline{r}}{\partial v} \, dv \; + \; \frac{\partial \underline{r}}{\partial w} \, dw$$

Along the curves $\{v = \text{constant},\ w = \text{constant}\}$, we have $dv = 0$ and $dw = 0$, so $\underline{dr}$ is simply

$$\underline{dr}_u \ = \ \frac{\partial \underline{r}}{\partial u}\, du$$

with $\underline{dr}_v$ and $\underline{dr}_w$ having analogous definitions.



The vectors $\underline{dr}_u$, $\underline{dr}_v$ and $\underline{dr}_w$ form the sides of an infinitesimal parallelepiped of volume

$$dV \ = \ \left|\underline{dr}_u \cdot \underline{dr}_v \times \underline{dr}_w\right|$$

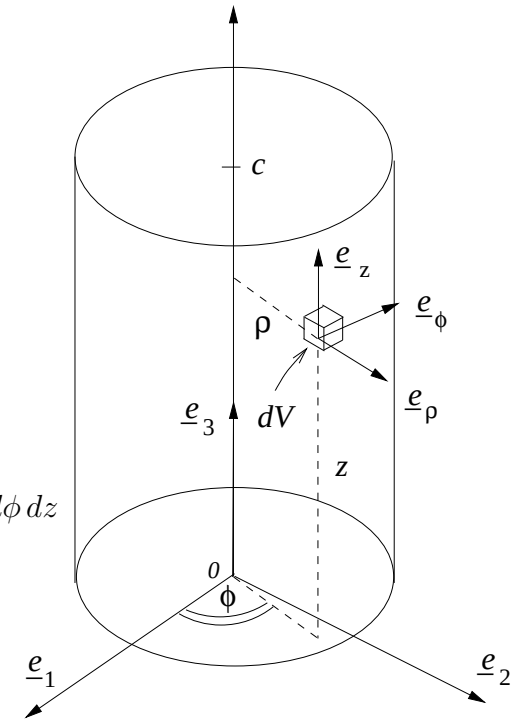
$$\boxed{dV = \left|\frac{\partial \underline{r}}{\partial u} \cdot \frac{\partial \underline{r}}{\partial v} \times \frac{\partial \underline{r}}{\partial w}\right| du\, dv\, dw}$$

Example: Consider a circular cylinder of radius a, height c. We can parameterise $\underline{r}$ using cylindrical polar coordinates. Within the cylinder, we have

$$\underline{r} = \rho\cos\phi\,\underline{e}_1 \ + \ \rho\sin\phi\,\underline{e}_2 \ + \ z\underline{e}_3 \quad \{0 \le \rho \le a,\ 0 \le \phi \le 2\pi,\ 0 \le z \le c\}$$

Thus
$$\frac{\partial \underline{r}}{\partial \rho} \ = \ \cos\phi\,\underline{e}_1 \ + \ \sin\phi\,\underline{e}_2$$
$$\frac{\partial \underline{r}}{\partial \phi} \ = \ -\rho\,\sin\phi\,\underline{e}_1 \ + \ \rho\,\cos\phi\,\underline{e}_2$$
$$\frac{\partial \underline{r}}{\partial z} \ = \ \underline{e}_3$$

and so
$$dV \ = \ \left|\frac{\partial \underline{r}}{\partial \rho} \cdot \frac{\partial \underline{r}}{\partial \phi} \times \frac{\partial \underline{r}}{\partial z}\right| d\rho\, d\phi\, dz \ = \ \rho\, d\rho\, d\phi\, dz$$



The **volume** of the cylinder is

$$\int_V dV \ = \ \int_{z=0}^{z=c}\int_{\phi=0}^{\phi=2\pi}\int_{\rho=0}^{\rho=a} \rho\, d\rho\, d\phi\, dz \ = \ \pi\, a^2 c.$$

Cylindrical basis: the normalised vectors (shown on the figure) form a non-Cartesian basis where

$$\underline{e}_\rho = \frac{\partial \underline{r}}{\partial \rho} \Big/ \left|\frac{\partial \underline{r}}{\partial \rho}\right| \quad ; \quad \underline{e}_\phi = \frac{\partial \underline{r}}{\partial \phi} \Big/ \left|\frac{\partial \underline{r}}{\partial \phi}\right| \quad ; \quad \underline{e}_z = \frac{\partial \underline{r}}{\partial z} \Big/ \left|\frac{\partial \underline{r}}{\partial z}\right|$$

Exercise: For **Spherical Polars** $\underline{r} = r\sin\theta\cos\phi\,\underline{e}_1 + r\sin\theta\sin\phi\,\underline{e}_2 + r\cos\theta\,\underline{e}_3$ show that

$$dV \ = \ \left|\frac{\partial \underline{r}}{\partial r} \cdot \frac{\partial \underline{r}}{\partial \theta} \times \frac{\partial \underline{r}}{\partial \phi}\right| dr\, d\theta\, d\phi \ = \ r^2 \sin\theta\, dr\, d\theta\, d\phi$$

17.2 Integral definition of divergence

If $\underline{A}$ is a vector field in the region R, and P is a point in R, then the divergence of $\underline{A}$ at P may be **defined** by

$$\boxed{\operatorname{div} \underline{A} \;=\; \lim_{\mathrm{V}\to 0} \frac{1}{\mathrm{V}} \int_{\mathrm{S}} \underline{A}\cdot\underline{dS}}$$

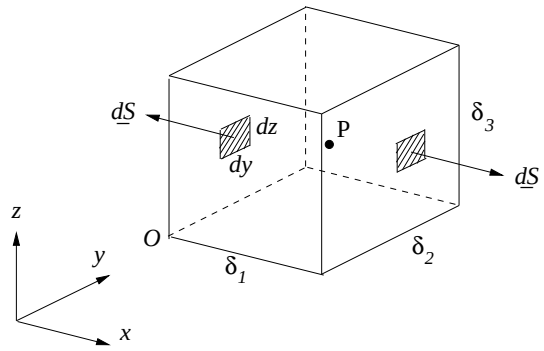
where S is a **closed** surface in R which encloses the volume V. The limit must be taken so that the point P is within V.

This definition of $\operatorname{div} \underline{A}$ is **basis independent**.

We now prove that our original definition of *div* is recovered in Cartesian co-ordinates

Let P be a point with Cartesian coordinates (x_0, y_0, z_0) situated at the *centre* of a small rectangular block of size $\delta_1 \times \delta_2 \times \delta_3$, so its volume is $\delta V = \delta_1\, \delta_2\, \delta_3$.

- On the **front** face of the block, orthogonal to the x axis at $x = x_0 + \delta_1/2$ we have *outward* normal $\underline{\hat{n}} = \underline{e}_1$ and so $\underline{dS} = \underline{e}_1\, dy\, dz$
- On the **back** face of the block orthogonal to the x axis at $x = x_0 - \delta_1/2$ we have *outward* normal $\underline{\hat{n}} = -\underline{e}_1$ and so $\underline{dS} = -\underline{e}_1\, dy\, dz$



Hence $\underline{A} \cdot \underline{dS} = \pm\, A_1\, dy\, dz$ on these two faces. Let us denote the two surfaces orthogonal to the $\underline{e}_1$ axis by S_1.

The contribution of these two surfaces to the integral $\displaystyle\int_S \underline{A} \cdot \underline{dS}$ is given by

$$\begin{aligned}
\int_{S_1} \underline{A} \cdot \underline{dS} \;&=\; \int_z\int_y \Big\{ A_1(x_0 + \delta_1/2, y, z) \;-\; A_1(x_0 - \delta_1/2, y, z) \Big\}\, dy\, dz \\
&=\; \int_z\int_y \Big\{ \Big[A_1(x_0, y, z) \;+\; \frac{\delta_1}{2}\, \frac{\partial A_1(x_0, y, z)}{\partial x} \;+\; O(\delta_1^2) \Big] \\
&\qquad\qquad - \Big[A_1(x_0, y, z) \;-\; \frac{\delta_1}{2}\, \frac{\partial A_1(x_0, y, z)}{\partial x} \;+\; O(\delta_1^2) \Big] \Big\}\, dy\, dz \\
&=\; \int_z\int_y \delta_1\, \frac{\partial A_1(x_0, y, z)}{\partial x}\, dy\, dz
\end{aligned}$$

where we have dropped terms of $O(\delta_1^2)$ in the Taylor expansion of A_1 about (x_0, y, z).

So

$$\frac{1}{\delta V} \int_{S_1} \underline{A} \cdot \underline{dS} = \frac{1}{\delta_2\, \delta_3} \int_z\int_y \frac{\partial A_1(x_0, y, z)}{\partial x}\, dy\, dz$$

As we take the limit $\delta_1, \delta_2, \delta_3 \to 0$ the integral tends to $\frac{\partial A_1(x_0,y_0,z_0)}{\partial x}\,\delta_2\,\delta_3$ and we obtain

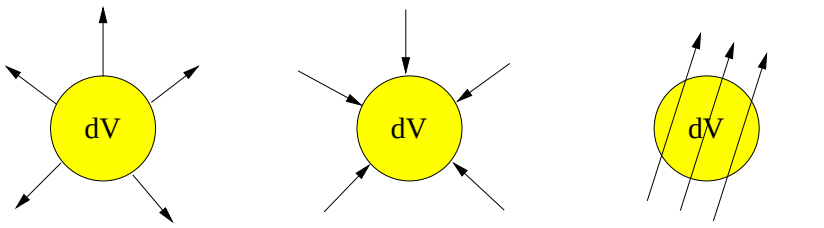
$$\lim_{\delta V \to 0} \frac{1}{\delta V} \int_{S_1} \underline{A} \cdot \underline{dS} \;=\; \frac{\partial A_1(x_0, y_0, z_0)}{\partial x}$$

With similar contributions from the other 4 faces, we find

$$\text{div}\,\underline{A} \;=\; \frac{\partial A_1}{\partial x} + \frac{\partial A_2}{\partial y} + \frac{\partial A_3}{\partial z} \;=\; \underline{\nabla} \cdot \underline{A}$$

in agreement with our original definition in Cartesian co-ordinates.

Note that the integral definition gives an intuitive understanding of the divergence in terms of net flux leaving a small volume around a point $\underline{r}$. **In pictures:** for a small volume dV



div $\underline{A}$ > 0 div $\underline{A}$ < 0 div $\underline{A}$ = 0
(flux in = flux out)

17.3 The divergence theorem (Gauss's theorem)

If $\underline{A}$ is a vector field in a volume V, and S is the closed surface bounding V, then

$$\boxed{\int_V \underline{\nabla} \cdot \underline{A}\, dV \;=\; \int_S \underline{A} \cdot \underline{dS}}$$

Proof : We derive the divergence theorem by making use of the integral definition of div $\underline{A}$

$$\text{div}\,\underline{A} \;=\; \lim_{\text{V}\to 0} \frac{1}{\text{V}} \int_{\text{S}} \underline{A} \cdot \underline{dS}.$$

Since this **definition** of div $\underline{A}$ is valid for volumes of arbitrary shape, we can build a smooth surface S from a large number, N, of blocks of volume ΔV^i and surface ΔS^i. We have

$$\text{div}\,\underline{A}(\underline{r}^i) \;=\; \frac{1}{\Delta V^i} \int_{\Delta S^i} \underline{A} \cdot \underline{dS} \;+\; (\epsilon^i)$$

where $\epsilon^i \to 0$ as $\Delta V^i \to 0$. Now multiply both sides by ΔV^i and sum over all i

$$\sum_{i=1}^{N} \text{div}\,\underline{A}(\underline{r}^i)\,\Delta V^i \;=\; \sum_{i=1}^{N} \int_{\Delta S^i} \underline{A} \cdot \underline{dS} \;+\; \sum_{i=1}^{N} \epsilon^i\,\Delta V^i$$

On rhs the contributions from surface elements *interior* to S cancel. This is because where two blocks touch, the outward normals are in *opposite* directions, implying that the contributions to the respective integrals cancel.

Taking the limit $N \to \infty$ we have, as claimed,

$$\int_V \underline{\nabla} \cdot \underline{A}\, dV \;=\; \int_S \underline{A} \cdot \underline{dS}\,.$$

Lecture 18: Continuity equation and curl

18.1 Examples of the divergence theorem

Volume of a body:

Consider the volume of a body:

$$V = \int_V dV$$

Recalling that $\underline{\nabla} \cdot \underline{r} = 3$ we can write

$$V = \frac{1}{3} \int_V \underline{\nabla} \cdot \underline{r}\, dV$$

which using the divergence theorem becomes

$$V = \frac{1}{3} \int_S \underline{r} \cdot \underline{dS}$$

Example: Consider the hemisphere $x^2 + y^2 + z^2 \leq a^2$ centred on $\underline{e}_3$ with bottom face at $z = 0$. Recalling that the divergence theorem holds for a *closed* surface, the above equation for the volume of the hemisphere tells us

$$V = \frac{1}{3} \left[\int_{hemisphere} \underline{r} \cdot \underline{dS} + \int_{bottom} \underline{r} \cdot \underline{dS} \right] .$$

On the bottom face $\underline{dS} = -\underline{e}_3\, dS$ so that $\underline{r} \cdot \underline{dS} = -z\, dS = 0$ since $z = 0$. Hence the only contribution comes from the (open) surface of the hemisphere and we see that

$$V = \frac{1}{3} \int_{hemisphere} \underline{r} \cdot \underline{dS} .$$

We can evaluate this by using spherical polars for the surface integral. As was derived above, for a hemisphere of radius a

$$\underline{dS} = a^2 \sin\theta\, d\theta\, d\phi\, \underline{e}_r .$$

On the hemisphere $\underline{r} \cdot \underline{dS} = a^3 \sin\theta\, d\theta\, d\phi$ so that

$$\int_S \underline{r} \cdot \underline{dS} = a^3 \int_0^{\pi/2} \sin\theta\, d\theta \int_0^{2\pi} d\phi = 2\pi a^3$$

giving the anticipated result

$$V = \frac{2\pi a^3}{3} .$$

18.2 Continuity equation

Consider a fluid with density field $\rho(\underline{r})$ and velocity field $\underline{v}(\underline{r})$. We have seen previously that the volume flux (volume per unit time) flowing across a surface is given by $\int_S \underline{v} \cdot \underline{dS}$. The corresponding mass flux (mass per unit time) is given by

$$\int_S \rho \underline{v} \cdot \underline{dS} \equiv \int_S \underline{J} \cdot \underline{dS}$$

where $\underline{J} = \rho \underline{v}$ is called the *mass current*.

Now consider a volume V bounded by the *closed* surface S containing no sources or sinks of fluid. Conservation of mass means that the outward mass flux through the surface S must be equal to the rate of decrease of mass contained in the volume V.

$$\int_S \underline{J} \cdot \underline{dS} = -\frac{\partial M}{\partial t} \ .$$

The mass in V may be written as $M = \int_V \rho \, dV$. Therefore we have

$$\frac{\partial}{\partial t} \int_V \rho \, dV + \int_S \underline{J} \cdot \underline{dS} = 0 \ .$$

We now use the divergence theorem to rewrite the second term as a volume integral and we obtain

$$\int_V \left[\frac{\partial \rho}{\partial t} + \underline{\nabla} \cdot \underline{J} \right] dV = 0$$

Now since this holds for arbitrary V we must have that

$$\boxed{\frac{\partial \rho}{\partial t} + \underline{\nabla} \cdot \underline{J} = 0 \ .}$$

This equation, known as the **continuity equation**, appears in many different contexts since it holds for any *conserved* quantity. Here we considered mass density ρ and mass current $\underline{J}$ of a fluid; but equally it could have been number density of molecules in a gas and current of molecules; electric charge density and electric current vector; thermal energy density and heat current vector; or even more abstract conserved quantities such as probability density.

18.3 Sources and sinks

Static case: Consider *time independent* behaviour where $\partial \rho / \partial t = 0$. The continuity equation tells us that for the density to be constant in time we must have $\underline{\nabla} \cdot \underline{J} = 0$ so that flux into a point equals flux out.

However if we have a **source** or a **sink** of the field, the divergence is not zero at that point. In general the quantity

$$\frac{1}{V} \int_S \underline{A} \cdot \underline{dS}$$

tells us whether there are sources or sinks of the vector field $\underline{A}$ within V: if V contains

- a **source**, then $\int_S \underline{A} \cdot \underline{dS} = \int_V \underline{\nabla} \cdot \underline{A} \, dV > 0$
- a **sink**, then $\int_S \underline{A} \cdot \underline{dS} = \int_V \underline{\nabla} \cdot \underline{A} \, dV < 0$

If S contains neither sources nor sinks, then $\int_S \underline{A} \cdot \underline{dS} = 0$.

As an example consider **electrostatics**. You will have learned that electric field lines are conserved and can only start and stop at charges. A positive charge is a source of electric

field (i.e. creates a positive flux) and a negative charge is a sink (i.e. absorbs flux or creates a negative flux).

The electric field due to a charge q at the origin is

$$\underline{E} = \frac{q}{4\pi\epsilon_0 r^2}\,\hat{\underline{r}}.$$

It is easy to verify that $\underline{\nabla}\cdot\underline{E} = 0$ except at the origin where the field is singular.

The flux integral for this type of field across a sphere (of any radius) around the origin was evaluated previously and we find the flux out of the sphere as:

$$\int_S \underline{E}\cdot\underline{dS} = \frac{q}{\epsilon_0}$$

Now since $\underline{\nabla}\cdot\underline{E} = 0$ away from the origin the results holds for *any surface enclosing the origin.* Moreover if we have several charges enclosed by S then

$$\int_S \underline{E}\cdot\underline{dS} = \sum_i \frac{q_i}{\epsilon_0}.$$

This recovers *Gauss' Law* of electrostatics.

We can go further and consider a *charge density* of $\rho(\underline{r})$ per unit volume. Then

$$\int_S \underline{E}\cdot\underline{dS} = \int_V \frac{\rho(\underline{r})}{\epsilon_0} dV\ .$$

We can rewrite the lhs using the divergence theorem

$$\int_V \underline{\nabla}\cdot\underline{E}\, dV = \int_V \frac{\rho(\underline{r})}{\epsilon_0} dV\ .$$

Since this must hold for arbitrary V we see

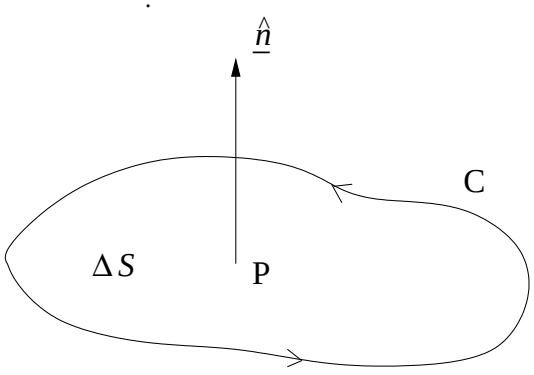
$$\boxed{\underline{\nabla}\cdot\underline{E} = \frac{\rho(\underline{r})}{\epsilon_0}}$$

which holds for all $\underline{r}$ and is one of Maxwell's equations of Electromagnetism.

18.4 Two definitions of curl

18.4.1 Line integral definition of curl

Let ΔS be a small planar surface containing the point P, bounded by a **closed** curve C, with unit normal $\hat{\underline{n}}$ and (scalar) area ΔS. Let $\underline{A}$ be a vector field defined on ΔS.



The component of $\underline{\nabla}\times\underline{A}$ parallel to $\hat{\underline{n}}$ is defined to be

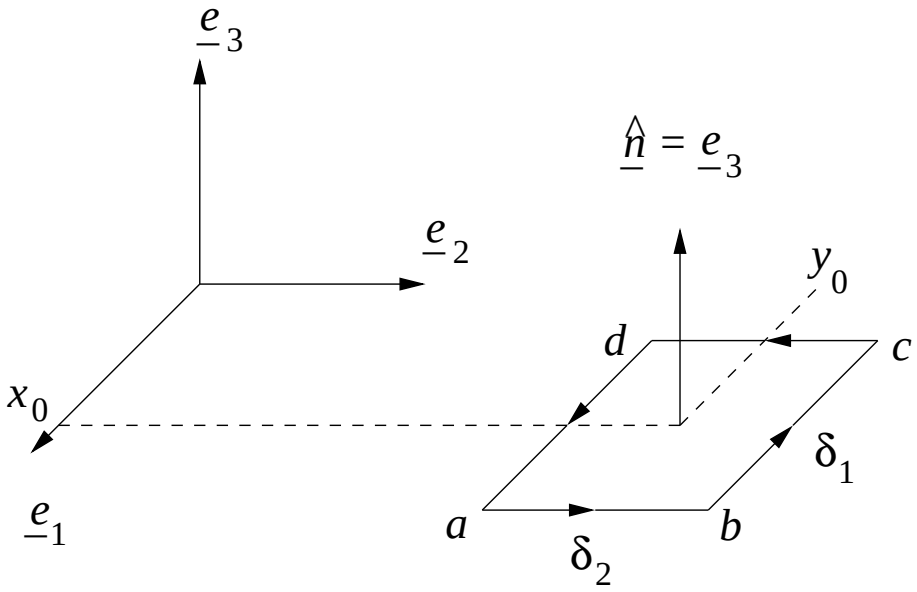
$$\boxed{\underline{\hat{n}}\cdot\left(\underline{\nabla}\times\underline{A}\right) \;=\; \lim_{\Delta S\to 0}\;\frac{1}{\Delta S}\;\oint_C \underline{A}\cdot\underline{dr}}$$

NB: the integral around C is taken in the right-hand sense with respect to the normal $\underline{\hat{n}}$ to the surface – as in the figure above.

This definition of curl is **independent of the choice of basis**. The usual Cartesian form for **curl** $\underline{A}$ can be recovered from this general definition by considering small rectangles in the $(\underline{e}_1-\underline{e}_2)$, $(\underline{e}_2-\underline{e}_3)$ and $(\underline{e}_3-\underline{e}_1)$ planes respectively, but you are not required to prove this.

18.4.2 Cartesian form of curl

Let P be a point with Cartesian coordinates (x_0, y_0, z_0) situated at the *centre* of a small rectangle $C = abcd$ of size $\delta_1 \times \delta_2$, area $\Delta S = \delta_1\,\delta_2$, in the $(\underline{e}_1-\underline{e}_2)$ plane.



The line integral around C is given by the sum of four terms

$$\oint_C \underline{A}\cdot\underline{dr} \;=\; \int_a^b \underline{A}\cdot\underline{dr} \;+\; \int_b^c \underline{A}\cdot\underline{dr} \;+\; \int_c^d \underline{A}\cdot\underline{dr} \;+\; \int_d^a \underline{A}\cdot\underline{dr}$$

Since $\underline{r} = x\underline{e}_1 + y\underline{e}_2 + z\underline{e}_3$, we have $\underline{dr} = \underline{e}_1\,dx$ along $d \to a$ and $c \to b$, and $\underline{dr} = \underline{e}_2\,dy$ along $a \to b$ and $d \to c$. Therefore

$$\oint_C \underline{A}\cdot\underline{dr} \;=\; \int_a^b A_2\,dy \;-\; \int_c^b A_1\,dx \;-\; \int_d^c A_2\,dy \;+\; \int_d^a A_1\,dx$$

For small δ_1 & δ_2, we can Taylor expand the integrands, viz

$$\begin{aligned}
\int_d^a A_1\,dx &= \int_d^a A_1(x, y_0 - \delta_2/2, z_0)\,dx \\
&= \int_{x_0-\delta_1/2}^{x_0+\delta_1/2} \left[A_1(x, y_0, z_0) - \frac{\delta_2}{2}\,\frac{\partial A_1(x, y_0, z_0)}{\partial y} + O(\delta_2^2)\right] dx \\
\int_c^b A_1\,dx &= \int_c^b A_1(x, y_0 + \delta_2/2, z_0)\,dx \\
&= \int_{x_0-\delta_1/2}^{x_0+\delta_1/2} \left[A_1(x, y_0, z_0) + \frac{\delta_2}{2}\,\frac{\partial A_1(x, y_0, z_0)}{\partial y} + O(\delta_2^2)\right] dx
\end{aligned}$$

so

$$\begin{aligned}
\frac{1}{\Delta S}\left[\int_d^a \underline{A}\cdot\underline{dr} + \int_b^c \underline{A}\cdot\underline{dr}\right] &= \frac{1}{\delta_1\,\delta_2}\left[\int_d^a A_1\,dx - \int_c^b A_1\,dx\right] \\
&= \frac{1}{\delta_1\delta_2}\int_{x_0-\delta_1/2}^{x_0+\delta_1/2}\left[-\delta_2\frac{\partial A_1(x, y_0, z_0)}{\partial y} + O(\delta_2^2)\right] dx \\
&\to -\frac{\partial A_1(x_0, y_0, z_0)}{\partial y} \quad \text{as} \quad \delta_1,\,\delta_2 \to 0
\end{aligned}$$

A similar analysis of the line integrals along $a \to b$ and $c \to d$ gives

$$\frac{1}{\Delta S}\left[\int_a^b \underline{A}\cdot\underline{dr} + \int_c^d \underline{A}\cdot\underline{dr}\right] \to \frac{\partial A_2(x_0, y_0, z_0)}{\partial x} \quad \text{as} \quad \delta_1,\,\delta_2 \to 0$$

Adding the results gives for our line integral definition of curl yields

$$\underline{e}_3\cdot\left(\underline{\nabla}\times\underline{A}\right) = \left(\underline{\nabla}\times\underline{A}\right)_3 = \left[\frac{\partial A_2}{\partial x} - \frac{\partial A_1}{\partial y}\right]\Bigg|_{(x_0, y_0, z_0)}$$

in agreement with our original definition in Cartesian coordinates.

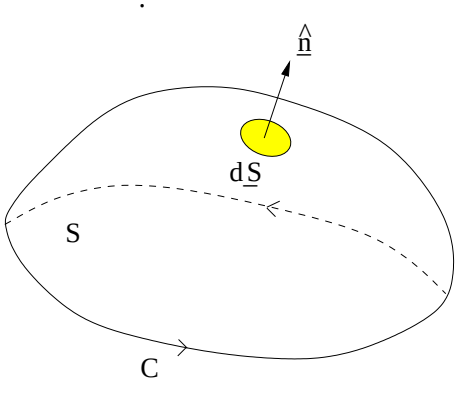
The other components of curl $\underline{A}$ can be obtained from similar rectangles in the $(\underline{e}_2 - \underline{e}_3)$ and $(\underline{e}_1 - \underline{e}_3)$ planes, respectively.

Lecture 19: Stokes' theorem

If S is an **open** surface, bounded by a simple **closed** curve C, and $\underline{A}$ is a vector field defined on S, then

$$\boxed{\oint_C \underline{A}\cdot\underline{dr} = \int_S \left(\underline{\nabla}\times\underline{A}\right)\cdot\underline{dS}}$$

where C is traversed in a right-hand sense about $\underline{dS}$. (As usual $\underline{dS} = \underline{\hat{n}}dS$ and $\underline{\hat{n}}$ is the unit normal to S).



Proof:

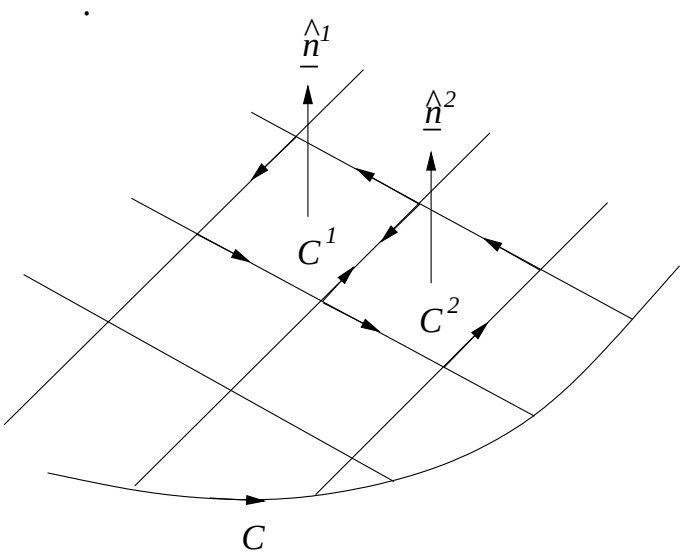
Divide the surface area S into N adjacent small surfaces as indicated in the diagram. Let $\underline{\Delta S^i} = \Delta S^i \, \underline{\hat{n}}^i$ be the vector element of area at $\underline{r}^i$. Using the integral definition of curl,

$$\underline{\hat{n}} \cdot \big(\text{curl } \underline{A}\big) \;=\; \underline{\hat{n}} \cdot \big(\underline{\nabla} \times \underline{A}\big) \;=\; \lim_{\Delta S \to 0} \; \frac{1}{\Delta S} \oint_C \underline{A} \cdot \underline{dr}$$

we multiply by ΔS^i and sum over all i to get

$$\sum_{i=1}^{N} \; \big(\underline{\nabla} \times \underline{A}(\underline{r}^i)\big) \cdot \underline{\hat{n}}^i \; \Delta S^i \;=\; \sum_{i=1}^{N} \; \oint_{C^i} \underline{A} \cdot \underline{dr} \; + \; \sum_{i=1}^{N} \; \epsilon^i \, \Delta S^i$$

where C^i is the curve enclosing the area ΔS^i, and the quantity $\epsilon^i \to 0$ as $\Delta S^i \to 0$.



Since each small closed curve C^i is traversed in the same sense, then, from the diagram, all contributions to $\sum_{i=1}^{N} \; \oint_{C^i} \underline{A} \cdot \underline{dr}$ **cancel**, except on those curves where part of C^i lies on the curve C. For example, the line integrals along the common sections of the two small closed curves C^1 and C^2 **cancel exactly**. Therefore

$$\sum_{i=1}^{N} \; \oint_{C^i} \underline{A} \cdot \underline{dr} \;=\; \oint_C \underline{A} \cdot \underline{dr}$$

Hence

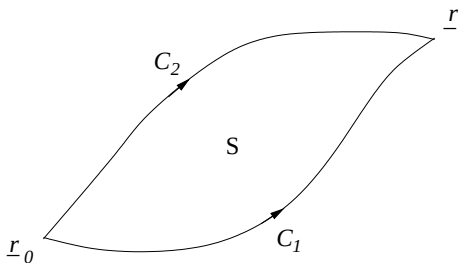
$$\oint_C \underline{A} \cdot \underline{dr} \;=\; \int_S \big(\underline{\nabla} \times \underline{A}\big) \cdot \underline{dS} \;=\; \int_S \underline{\hat{n}} \cdot \big(\underline{\nabla} \times \underline{A}\big) \; dS$$

19.1 Applications of Stokes' theorem

We have seen that if a vector field is irrotational (curl vanishes) then a line integral is independent of path. We can now prove this statement using Stokes' theorem.

Proof:

Let $\underline{\nabla} \times \underline{A}(\underline{r}) = 0$ in R, and consider the **difference** of two line integrals from the point $\underline{r}_0$ to the point $\underline{r}$ along the two curves C_1 and C_2 as shown:



$$\int_{C_1} \underline{A}(\underline{r}') \cdot \underline{dr}' - \int_{C_2} \underline{A}(\underline{r}') \cdot \underline{dr}'$$

We use $\underline{r}'$ as integration variable to distinguish it from the **limits** of integration $\underline{r}_0$ and $\underline{r}$.

We can rewrite this as the integral around the **closed** curve $C = C_1 - C_2$:

$$\begin{aligned} \int_{C_1} \underline{A}(\underline{r}') \cdot \underline{dr}' - \int_{C_2} \underline{A}(\underline{r}') \cdot \underline{dr}' &= \oint_C \underline{A}(\underline{r}') \cdot \underline{dr}' \\ &= \int_S \underline{\nabla} \times \underline{A} \cdot \underline{dS} = 0 \end{aligned}$$

In the above, we have used Stokes' theorem to write the *line* integral of $\underline{A}$ around the closed curve $C = C_1 - C_2$, as the *surface* integral of $\underline{\nabla} \times \underline{A}$ over an open surface S bounded by C. This integral is zero because $\underline{\nabla} \times \underline{A} = 0$ everywhere in R. Hence

$$\underline{\nabla} \times \underline{A}(\underline{r}) = 0 \quad \Rightarrow \quad \oint_C \underline{A}(\underline{r}') \cdot \underline{dr}' = 0$$

for *any* closed curve C in R as claimed.

Clearly, the converse is also true ı.e. if the line integral between two points is path independent then the line integral around any closed curve (connecting the two points) is zero. Therefore

$$0 = \oint_C \underline{A}(\underline{r}') \cdot \underline{dr}' = \int_S \underline{\nabla} \times \underline{A} \cdot \underline{dS}$$

where we have used Stokes' theorem and since this holds for any S the field must be irrotational.

Ampère's Law

In Physics 2 you will have met the integral form of Ampère's law, which describes the magnetic field $\underline{B}$ produced by a steady current $\underline{J}$:

$$\oint_C \underline{B} \cdot \underline{dr} = \mu_0 \int_S \underline{J} \cdot \underline{dS}$$

where the closed curve C bounds the surface S *i.e.* the rhs is the current flux across S. We can rewrite the lhs using Stokes' theorem to obtain

$$\int_S (\underline{\nabla} \times \underline{B}) \cdot \underline{dS} = \mu_0 \int_S \underline{J} \cdot \underline{dS} \, .$$

Since this holds for any surface S we must have

$$\underline{\nabla} \times \underline{B} - \mu_0 \underline{J} = 0$$

which is the differential form of Ampère's law and is one of Maxwell's equations (see next year).

Planar Areas

Consider a planar surface in the $\underline{e}_1-\underline{e}_2$ plane and the vector field

$$\underline{A} = \frac{1}{2}\left[-y\underline{e}_1 + x\underline{e}_2\right] \ .$$

We find $\underline{\nabla} \times \underline{A} = \underline{e}_3$. Since a vector element of area normal to a planar surface in the $\underline{e}_1-\underline{e}_2$ plane is $\underline{dS} = dS\,\underline{e}_3$ we can obtain the area in the following way

$$\int_S \underline{\nabla} \times \underline{A} \cdot \underline{dS} = \int_S \underline{e}_3 \cdot \underline{dS} = \int_S dS = S$$

Now we can use Stokes' theorem to find

$$\begin{aligned} S &= \oint_C \underline{A} \cdot \underline{dr} = \frac{1}{2}\oint_C (-y\underline{e}_1 + x\underline{e}_2) \cdot (\underline{e}_1 dx + \underline{e}_2 dy) \\ &= \frac{1}{2}\oint_C (x\,dy - y\,dx) \end{aligned}$$

where C is the closed curve bounding the surface.

e.g. To find the area inside the curve

$$x^{2/3} + y^{2/3} = 1$$

use the substitution $x = \cos^3\phi$, $y = \sin^3\phi$, $0 \le \phi \le 2\pi$ then

$$\frac{dx}{d\phi} = -3\cos^2\phi\,\sin\phi \quad ; \quad \frac{dy}{d\phi} = 3\sin^2\phi\,\cos\phi$$

and we obtain

$$\begin{aligned} S &= \frac{1}{2}\oint_C \left(x\,\frac{dy}{d\phi} - y\,\frac{dx}{d\phi}\right) d\phi \\ &= \frac{1}{2}\int_0^{2\pi} \left(3\cos^4\phi\sin^2\phi + 3\sin^4\phi\cos^2\phi\right) d\phi \\ &= \frac{3}{2}\int_0^{2\pi} \sin^2\phi\cos^2\phi\ d\phi = \frac{3}{8}\int_0^{2\pi} \sin^2 2\phi\, d\phi = \frac{3\pi}{8} \end{aligned}$$

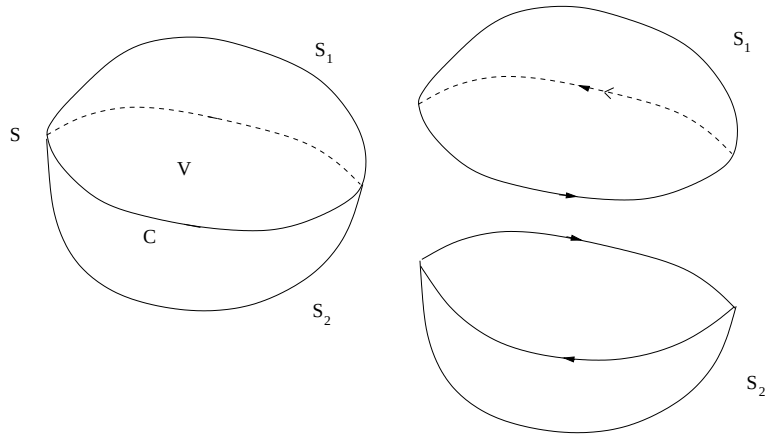
19.2 Example on joint use of divergence and Stokes' theorems

Example: show that $\underline{\nabla} \cdot \underline{\nabla} \times \underline{A} \equiv 0$ independent of co-ordinate system:

Let S be a closed surface, enclosing a volume V. Applying the divergence theorem to $\underline{\nabla} \times \underline{A}$, we obtain

$$\int_V \underline{\nabla} \cdot \left(\underline{\nabla} \times \underline{A}\right)\ dV \ = \ \int_S \left(\underline{\nabla} \times \underline{A}\right) \cdot \underline{dS}$$

Now divide S into two surfaces S_1 and S_2 with a **common** boundary C as shown below



Now use Stokes' theorem to write

$$\int_S \left(\underline{\nabla} \times \underline{A}\right) \cdot \underline{dS} \;=\; \int_{S_1} \left(\underline{\nabla} \times \underline{A}\right) \cdot \underline{dS} + \int_{S_2} \left(\underline{\nabla} \times \underline{A}\right) \cdot \underline{dS} \;=\; \oint_C \underline{A} \cdot \underline{dr} - \oint_C \underline{A} \cdot \underline{dr} \;=\; 0$$

where the second line integral appears with a minus sign because it is traversed in the **opposite** direction. (Recall that Stokes' theorem applies to curves traversed in the right hand sense with respect to the outward normal of the surface.)

Since this result holds for arbitrary volumes, we must have

$$\underline{\nabla} \cdot \underline{\nabla} \times \underline{A} \equiv 0$$