

Vectors, Matrices and Determinants

Simon C Benjamin

Chapter 2 cont.: Vector Fields and Calculus

2.5 Motivation

So far we have understood that individual vectors can be useful to describe physical phenomena and effects such as displacement, velocity, force, etc. Now we will consider the idea of a vector *function* which means a vector that varies according to some parameter(s) that we feed in. For example, a velocity whose direction and strength depends on time, $\mathbf{v}(t)$, or a force like gravity whose direction and strength depends on position in space, $\mathbf{F}(x, y, z)$. Extending the notion of vectors in this way is extremely powerful for physics, materials science and engineering.

2.6 Scalar fields and Vector fields

The idea of a *scalar function* will be familiar to the reader: a simple example is $s(t) = -\frac{1}{2}gt^2$, which is the distance fallen by an object in time t . Another example would be

$$F(\mathbf{r}) = \frac{G M_e m}{|\mathbf{r}|^2}$$

which is the *magnitude* of the gravitational force on a point mass m above the surface of the Earth (mass M_e) when the displacement vector from the centre of the Earth to the point is $\mathbf{r}$. Note that for this second example we ‘fed in’ a vector but ‘got back’ a scalar – thus it is still a *scalar function*. Also note that this equation is only true when $|\mathbf{r}| > R_e$ i.e. we are above the Earth’s surface because otherwise we don’t get the full effect of Earth’s mass. In these notes we’ll always assume $|\mathbf{r}| > R_e$ when using Earth’s gravity as an example.

Our $F(\mathbf{r})$ is also a good example of a *scalar field* because it gives us a scalar for every point in space – here, 3D space is the field over which the function acts.

Now consider if we wanted not just the magnitude of the force acting on our point mass m but also the direction, i.e. we want to know the whole force vector. Well we know that the force acts in a direction towards the centre of the Earth, i.e. back along the direction of $\mathbf{r}$. Therefore we can write

$$\mathbf{F}(\mathbf{r}) = \frac{G M_e m}{|\mathbf{r}|^2}(-\hat{\mathbf{r}}) = -\frac{G M_e m \mathbf{r}}{|\mathbf{r}|^3} = \frac{G M_e m}{|\mathbf{r}|^3}(x\mathbf{i} + y\mathbf{j} + z\mathbf{k}) \quad (2.1)$$

3: Introducing Matrices

Now this object $\mathbf{F}(\mathbf{r})$ is a vector function, because it is a vector that is a function of something. We can also call it a vector field because it gives us a vector for every location in a field (here, 3D space).

It's often useful to draw a vector field. There are (at least) two common ways of doing this. They are shown in Figure 2.1.

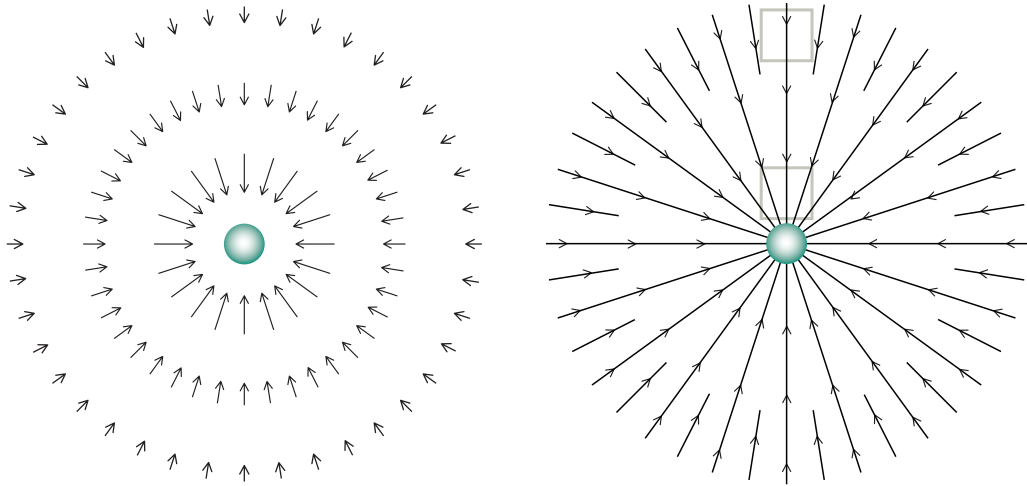


Figure 2.1: Earth's gravity is well described by a vector field. Two common ways of drawing a vector field are shown here. On the left, we simply draw the usual direction/magnitude vector arrows at various points we choose. On the right, a method that can be simpler to draw – now the density of lines in the figure *roughly* indicates the strength of the field (compare the line density in the two grey boxes).

Gravity is a natural choice to think about, because we are used to referring to 'gravitational field', but it is worth noting that vector fields don't need to refer to forces. For example, consider a solid object rotating around a fixed axis. Every point in the object is moving with a velocity that depends on the distance from the centre, $\mathbf{r}$. But the magnitude of the velocity increases as we go further from the centre, and its direction is perpendicular to $\mathbf{r}$. How can we write the vector field $\mathbf{v}(\mathbf{r})$ that gives us the velocity of any point?

[Space for answer]

- Magnitude of the velocity increases *linearly* according to $|\mathbf{r}|$.
- Direction of the velocity can be obtained as $\hat{\mathbf{p}} \times \hat{\mathbf{r}}$, where $\mathbf{p}$ is a vector along the axis.
- We'll need a constant to say how fast the object is rotating. Let's say that a point that is 'unit distance' from the axis has magnitude of velocity

(speed) s . Then we get

$$\mathbf{v}(\mathbf{r}) = s |\mathbf{r}| \hat{\mathbf{p}} \times \hat{\mathbf{r}} = s \hat{\mathbf{p}} \times \mathbf{r}$$

2.7 Notation for vector fields

Typically a vector field may be written with the same notation as a vector (e.g. **bold** in a typeset document, underlined in hand-written notes). Commonly $\mathbf{E}$ is used for an electric field, and $\mathbf{B}$ for a magnetic field. It may be that notation such as $\mathbf{E}(x, y, z)$ or equivalently $\mathbf{E}(\mathbf{r})$ is used to stress that this is a vector field and therefore the vector varies with position in space; here $\mathbf{r}$ would be understood to mean the position vector $\mathbf{r} = x \mathbf{i} + y \mathbf{j} + z \mathbf{k}$.

2.8 Addition of vector fields

Mathematically, we can add two vector fields just as we may add two vectors; and just as with simple vectors, whether or not this makes sense physically depends on what the vectors represent. For the electric field, for example, it is physically correct to simply add the fields that result from, say, two different electrostatic charges in order to find the field that would exist when both charges are present. It is quite instructive to do this for a negative and a positive charge at two distinct locations; a tool like Mathematica makes it easy to visualise the result.

2.9 Differentiating a vector function

We may want to differentiate a vector function, e.g. we might want to talk about how fast the gravitational force is changing as we move our test mass about in space, or how fast a vector that depends on time is evolving.

3: Introducing Matrices

For ordinary scalar functions, we say that the derivative can be thought of as the limit as we move an infinitesimal distance,

$$\frac{d}{dx}f(x) = \lim_{\delta x \rightarrow 0} \frac{f(x + \delta x) - f(x)}{\delta x}$$

and we say that a function is *differentiable* if this limit is well-defined. There should be no discontinuity in $f(x)$ at x , so that the function varies smoothly.

We have exactly the same thing for vector functions. Suppose that we have a vector that is a function of time, e.g.

$$\mathbf{d}(t) = u t \mathbf{i} + \left(-\frac{a t^2}{2}\right) \mathbf{j}$$

where a and u are constants, while $\mathbf{i}$ and $\mathbf{j}$ are the usual Cartesian unit vectors (what could this vector function be?) then

$$\frac{d}{dt}\mathbf{d}(t) = \lim_{\delta t \rightarrow 0} \frac{\mathbf{d}(t + \delta t) - \mathbf{d}(t)}{\delta t} = u \mathbf{i} + (-a t) \mathbf{j}$$

The vector function is *differentiable* because it varies smoothly with its parameter, time in this case. Similarly we could consider a vector function that depends on multiple parameters, e.g. a vector field that depends on the point in space, specified by x , y , z or instead by $\mathbf{r}$. If it varies smoothly as we change the parameter(s) we feed in (the function's arguments) then it is differentiable.

2.10 Vector field as a gradient of a Scalar field

Many simple vector fields can thought of as derivative of a scalar field. This sounds like a complex idea, but it's quite simple. Consider again the gravitational field around the Earth. One *scalar* field that we could consider, is the potential energy of an object due to gravity. The higher we lift an object, the more its potential energy changes. The total potential energy that a test mass m has because of Earth's gravity is

$$P(\mathbf{r}) = -\frac{G M_e m}{|\mathbf{r}|} \text{ where, as usual, we assume } |\mathbf{r}| > R_e \quad (2.2)$$

The minus sign is because we have to provide energy to move the mass away from the centre, and at infinite distance it's natural to say the energy is zero. We can compare this energy equation (2.2) to our earlier vector field equation for gravity, Eqn. (2.1). There must be a relationship, since

we know from Newton that *work done*=*force*×*distance*. In fact the force field Eqn. (2.1) must be related to the energy equation Eqn. (2.2) by taking a gradient.

Why is that? It's quite intuitive really: if the gravitational force is strong in some direction, then if we move our mass in that direction we will make a large change in its potential energy – because if we're moving against the force we'll have to do a lot of work, or if we're going with the force then we can make the mass do a lot of work. Said the other way around, if our potential energy changes a lot as we move in a certain direction, there must be a strong force in that direction.

We can capture the relationship by finding the gradient of our scalar field in the x , y and z directions and setting these three gradients as the strength of the force field in the same three directions. So,

$$\mathbf{F}(\mathbf{r}) = - \left(\frac{\partial}{\partial x} P(\mathbf{r}) \right) \mathbf{i} - \left(\frac{\partial}{\partial y} P(\mathbf{r}) \right) \mathbf{j} - \left(\frac{\partial}{\partial z} P(\mathbf{r}) \right) \mathbf{k} \quad (2.3)$$

Think about our choice of sign: If the potential energy increases with x (positive gradient) then the gravitational force must be in the opposite direction (we are doing work to move the mass). Let's work it through.

$$P(\mathbf{r}) = P(x\mathbf{i} + y\mathbf{j} + z\mathbf{k}) = \frac{-G M_e m}{\sqrt{x^2 + y^2 + z^2}}$$

$$\begin{aligned} \frac{\partial}{\partial x} P(\mathbf{r}) &= \frac{\partial}{\partial x} \frac{-G M_e m}{(x^2 + y^2 + z^2)^{\frac{1}{2}}} \\ &= \frac{-G M_e m}{(x^2 + y^2 + z^2)^{\frac{3}{2}}} \frac{-2x}{2} \\ &= \frac{G M_e m}{|\mathbf{r}|^3} x \end{aligned}$$

similarly

$$\frac{\partial}{\partial y} P(\mathbf{r}) = \frac{G M_e m}{|\mathbf{r}|^3} y \quad \text{and} \quad \frac{\partial}{\partial z} P(\mathbf{r}) = \frac{G M_e m}{|\mathbf{r}|^3} z$$

so using Eqn. (2.3) we find

$$\mathbf{F}(\mathbf{r}) = \left(\frac{-G M_e m}{|\mathbf{r}|^3} \right) (x\mathbf{i} + y\mathbf{j} + z\mathbf{k}) = - \frac{G M_e m \mathbf{r}}{|\mathbf{r}|^3}$$

2.11 The Grad operator

We've seen that we can derive the force due to gravity as the gradient of the potential energy *scalar field*. In fact there is often a relationship between the scalar and the vector fields, and it is useful to define a symbol which means 'take the gradient of this scalar field to make a vector field'.

$$\text{grad } S \equiv \nabla S \equiv \frac{\partial S}{\partial x} \mathbf{i} + \frac{\partial S}{\partial y} \mathbf{j} + \frac{\partial S}{\partial z} \mathbf{k}$$

The symbol ∇ is called 'nabla', but the expression ∇S is often pronounced 'grad S'. We can now write Eqn. (2.3) more neatly as $\mathbf{F}(\mathbf{r}) = -\nabla P$.

Note 1: The gravitational and electrostatic fields can be written as the gradient of a scalar field, but many vector fields cannot – one example is the 'velocity field' we thought about for the spinning disk. [How can we see it's impossible to write it as a grad?](#)

Note 2: If our scalar field S has only two variables, say x and y but no z , then that's fine – in this case we'll just get a two-dimensional vector field that has no component in the $\mathbf{k}$ direction.

2.12 Divergence and the Div operator

Look at Figure 2.2, which shows the gravitational field due to a point mass (left) and the electric field due to two charges (right). Clearly the points corresponding to the mass and the charges are special: Nearby, all the vector field lines point toward the point (or away, for the positive charge). This phenomenon is called *divergence* and can be measured as follows.

We define the divergence, or 'Div operator' as follows. Here vector field $\mathbf{F}$ is a function of position in space, i.e. $\mathbf{F}(\mathbf{r})$, but we just write $\mathbf{F}$ for simplicity

$$\text{div } \mathbf{F} \equiv \nabla \cdot \mathbf{F} \equiv \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}$$

where F_x , F_y and F_z are just the Cartesian components of $\mathbf{F}$,

$$\mathbf{F} = F_x \mathbf{i} + F_y \mathbf{j} + F_z \mathbf{k}$$

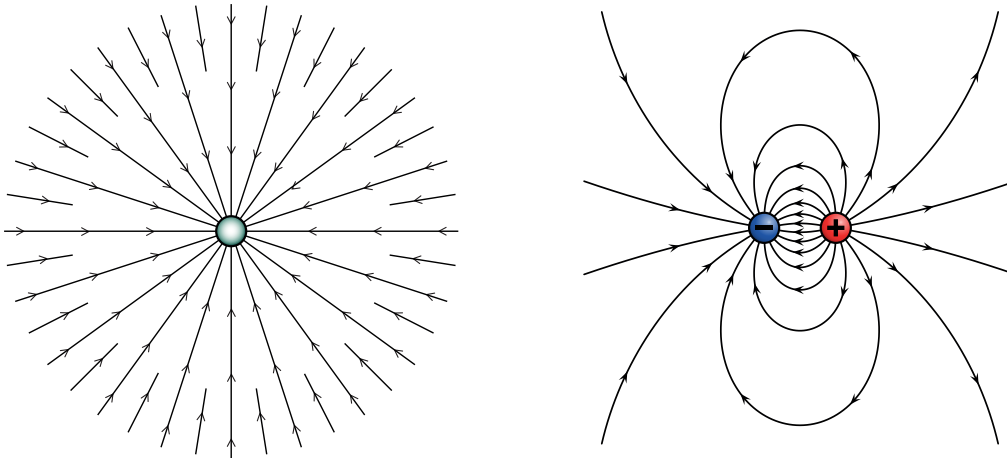


Figure 2.2: The gravitational field around a point mass (left), and the electric field around two points charges, one of either sign (right). In both cases, the Div operator will produce zero everywhere except the particles' locations.

Notice that the symbol for the Div operator, $\nabla \cdot$ looks a bit like a dot product. This is because one way to remember it is as a 'kind of' dot product between a pseudo-vector formed from the differentials $\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}$ and a second vector which is just $\mathbf{F}$.

In the exercises, there is a question in which you prove that for a gravitational field as shown in Fig. 2.2 the divergence is *zero* everywhere except where $\mathbf{r} = \mathbf{0}$. Generally, the divergence of a gravitational field corresponds to the density of mass at that point (and so, is zero whether there is no mass). The equivalent thing is true of the electric field: the divergence at some point in space corresponds to the charge density there – and this is in fact one of Maxwell's laws.

2.13 The Curl operator

The Grad operator turns a scalar field into a vector field, and the Div operator does the reverse: it turns a vector field into a scalar field (which may be zero). There is a third operator that we'll define here for completeness: The Curl transforms one vector field into another.

$$\text{curl } \mathbf{F} \equiv \nabla \times \mathbf{F} \equiv \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} \right) \mathbf{i} + \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x} \right) \mathbf{j} + \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right) \mathbf{k}$$

where F_x , F_y and F_z are just the Cartesian components of $\mathbf{F}$,

$$\mathbf{F} = F_x \mathbf{i} + F_y \mathbf{j} + F_z \mathbf{k}$$

We won't discuss the curl operator further here, except to note that it would actually give a zero vector field if applied to either of the fields in Fig. 2.2. The curl will give a non-zero vector field if the original field has closed loops in it, as magnetic fields do.

2.14 Integrating along a path in a Vector field

We have considered e.g. the force $\mathbf{F}$ exerted by someone pushing on a carriage that runs on rails. If the carriage moves a distance given by vector $\mathbf{D}$ then the work done will be simply $\mathbf{F} \cdot \mathbf{D}$. Now we want to extend this to think about a **vector field**. Let's go back again to the example of gravity.

Suppose we now consider the case that some test mass (for now we'll assume a mass of 1 Kg to save writing m_t everywhere) is moved through a gravitational field from a point with position vector $\mathbf{r}_0$ to $\mathbf{r}_1$, i.e. it experiences a displacement $\mathbf{D} = \mathbf{r}_1 - \mathbf{r}_0$.

How much energy would we use (or indeed, gain) if we make this happen? If we were in a *uniform* force field, i.e. one whose direction and strength were the same everywhere $\mathbf{F}(\mathbf{r}) = \mathbf{F}_U$, then the answer would be simply

$$(\text{force}) \cdot (\text{displacement vector}), \text{ i.e. } \mathbf{F}_U \cdot \mathbf{D}$$

which is essentially the same as for our person-pushing-train scenario. But what is interesting here is that we are moving the test mass through a gravitational field with varying strength. To describe what happens we can break up the path we take into N little steps labelled $\delta \mathbf{l}_i$ then

$$\mathbf{D} = \sum_{i=1}^{i=N} \delta \mathbf{l}_i \text{ so that we would have the total work } W_{tot} = \sum_{i=1}^{i=N} \mathbf{F}(\mathbf{r}) \cdot \delta \mathbf{l}_i.$$

[Sketch here]

In the limit of our little steps becoming infinitesimal we can rewrite the sum as an integral:

$$\mathbf{D} = \int_C d\mathbf{l} \text{ so that we would have the total work } W_{tot} = \int_C \mathbf{F}(\mathbf{r}) \cdot d\mathbf{l}$$

where it is conventional to use the subscript C on the integral to indicate that we are integrating along some specific path that we have in mind (sometimes S is written instead). This is therefore called a **path integral**, and also sometimes called a ‘line integral’.

The idea is that each step is a tiny displacement, so small that the vector field is effectively constant throughout the displacement. This works provided that the strength and direction of the vector field is *continuous* meaning that it doesn’t suddenly jump. For physical fields like gravity and magnetic or electrostatic force, this will always be the case.

To actually do the integral we need to specify the path (which amounts to the same thing as listing all the little $\delta\mathbf{l}_i$ displacements). We could describe our progress along the line using a scalar quantity s that goes from zero to one. Suppose that we go in a straight line. Then as we move along the path our position vector $\mathbf{r}$ is given by

$$\mathbf{r}(s) = \mathbf{r}_0 + s\mathbf{D}, \text{ remembering that } \mathbf{D} \equiv (\mathbf{r}_1 - \mathbf{r}_0) \text{ and } s \text{ goes from } 0 \text{ to } 1.$$

What about our small quantities $\delta\mathbf{l}$ (and $d\mathbf{l}$, the limit of becoming very small)? We should be able to get them by considering a small change in the position vector, so

$$\delta\mathbf{l} = \mathbf{r}(s + \delta s) - \mathbf{r}(s) = \mathbf{D}\delta s, \text{ and taking the infinitesimal limit } d\mathbf{l} = \mathbf{D}ds.$$

[Sketch]

3: Introducing Matrices

Therefore our integral will be

$$\begin{aligned} W_{tot} &= \int_C \mathbf{F}_g(\mathbf{r}(s)) \cdot d\mathbf{l} = \int_{s=0}^{s=1} \mathbf{F}_g(\mathbf{r}_0 + s \mathbf{D}) \cdot \mathbf{D} ds \\ &= -GM_e \int_{s=0}^{s=1} \frac{(\mathbf{r}_0 + s \mathbf{D}) \cdot \mathbf{D} ds}{|\mathbf{r}_0 + s \mathbf{D}|^3} \end{aligned} \quad (2.4)$$

Let's make things easy on ourselves and evaluate this by assuming that we are moving our test mass directly away from the centre of the earth, so that $\mathbf{D}$ and $\mathbf{r}_0$ are in the same direction. Then $\mathbf{r}_0 = r_0 \hat{\mathbf{D}}$ so

$$\begin{aligned} W_{tot} &= -GM_e \int_{s=0}^{s=1} \frac{(r_0 + sD) \hat{\mathbf{D}} \cdot \mathbf{D} ds}{|(r_0 + sD) \hat{\mathbf{D}}|^3} \quad \text{where } D = |\mathbf{D}| \\ &= -GM_e \int_{s=0}^{s=1} \frac{(r_0 + sD) D ds}{(r_0 + sD)^3} \\ &= -GM_e \int_{s=0}^{s=1} \frac{D ds}{(r_0 + sD)^2} \\ &= -GM_e \int_{R=r_0}^{R=r_1} \frac{dR}{R^2} \quad \text{using the change of variable } R = r_0 + sD. \end{aligned}$$

This last integral is of course just what you would write down if asked “what is the change in energy if a mass of 1 Kg rises from height r_0 to new height r_1 ”. The integral comes out as $GM_e(1/r_1 - 1/r_0)$.

Aside: can we reconcile this with the rule that the energy change in lifting a mass by a vertical height h is just mgh ?

2.15 Path independence

Would it have mattered if we took a different path between $\mathbf{r}_0$ and $\mathbf{r}_1$, when we worked out Eqn. (2.4)? In fact it would make no difference – except to make the derivation harder! It makes intuitive sense that you gain the same energy however you move a mass from one point to another (ignoring things like air resistance and all that, of course). But actually it's an example of a powerful general rule.

The integral along some path in a vector field, $\mathbf{F}(\mathbf{r})$, only depends on the end points and *not* the specific path taken, if the vector field can be obtained from a scalar field using grad. So

If $\mathbf{F}(\mathbf{r}) = \nabla S(\mathbf{r})$ then $\int_C \mathbf{F} d\mathbf{l}$ is path independent.

Advanced note: In fact another equivalent condition is, ‘if $\nabla \times \mathbf{F} = \mathbf{0}$ then the path integral is path-independent’ because in fact if $\mathbf{F}(\mathbf{r}) = \nabla S(\mathbf{r})$ then it must be the case that $\nabla \times \mathbf{F} = \mathbf{0}$ and vice versa.

2.16 Integrating over a surface

Sometimes we want to integrate a vector field over a surface – this sounds complicated but is generally quite straightforward. Let’s consider Gauss’s law, which states

$$\int_S \mathbf{E} \cdot d\mathbf{a} = \frac{1}{\epsilon_0} \int \rho dV.$$

Gauss’s law comes into play when we are interested in some region of space that contains electric charge; the integral on the left refers to the surface of that region, and the integral on the right refers to the volume of the region. Note: some authors would write this as

$$\iint_S \mathbf{E} \cdot d\mathbf{a} = \frac{1}{\epsilon_0} \iiint \rho dV.$$

where the extra integral signs are to help communicate that the one on the left is a surface (2D) and the one on the right is a volume (3D). Another convention sometimes seen is to put a circle on the integral since we are thinking about a *closed* surface.

$$\oiint_S \mathbf{E} \cdot d\mathbf{a} = \frac{1}{\epsilon_0} \iiint \rho dV.$$

[Sketch]

3: Introducing Matrices

The right hand side here is simple: the quantity ρ is the density of electric charge in the region – it might be a constant or it might vary, so we could have $\rho(\mathbf{r})$. Since the total charge in some tiny volume dV is ρdV we see that the integral is simply the *total charge* within this entire region.

The left hand side is more interesting. We must integrate by combining the tiny surface area fragments $d\mathbf{a}$ – but we notice that the $\mathbf{a}$ here is bold, denoting a vector, and there is a dot product between this vector and the electric vector field $\mathbf{E}$.

To understand what we are supposed to do here, think of any closed surface as being made of tiny flat sections, which together create the curved surface. How can we describe a little flat piece of surface (or ‘surface element’) with using a vector? Well we choose the normal to that element as the direction of the vector, and the amount of area corresponds to the magnitude of the vector.

[Sketch]

By far the easiest surface to integrate over will be the sphere centred at the origin. Then the direction of each surface element $d\mathbf{a}$ will simply be $\hat{\mathbf{r}}$ and we only have to figure out the size of the surface element. If we use spherical polar coordinates, then an elementary ‘piece’ of the surface at radius r is given by $r^2 \sin(\theta) d\theta d\psi$ where θ is the polar angle, and ψ is the azimuthal angle. So then,

$$d\mathbf{a} = (r^2 \sin(\theta) d\theta d\psi) \hat{\mathbf{r}}$$

Now if we are fortunate enough that the vector field $\mathbf{E}$ is also spherically symmetric, so that the direction at any point is directly out (or in) along $\hat{\mathbf{r}}$, then we’ll simply find that

$$\mathbf{E} \cdot d\mathbf{a} = |\mathbf{E}| (r^2 \sin(\theta) d\theta d\psi) \quad (2.5)$$

and we can then proceed to do the integral in the normal way, i.e. completing the $\int_0^\pi d\theta$ and $\int_0^{2\pi} d\psi$ integrals in turn.

3: Introducing Matrices

Here we'll introduce the idea of a matrix, how to manipulate matrices, and their use for representing transformations.

3.17 What is a matrix?

In everyday language, a matrix is an enclosure, a structure that contains or confines other entities. In mathematics, a matrix is an ordered array of numbers. We will be concerned with matrices that can be written as a rectangular array of numbers, and we will characterise the size of the array by the number of rows and columns it has.

3.17.1 Notation

We will write our matrices with the numbers inside normal, curved brackets (some texts use square brackets $[]$ instead). Each specific number within the matrix is referred to as an element. When we wish to write the matrix as a single symbol, we will use a bold capital letter in these typeset notes, .e.g. **A**. However, in handwritten work it is common to use a double underlining, as in A.

$$\mathbf{A} = \begin{pmatrix} 2 & -3 & 4 \\ 11 & 0 & \frac{7}{5} \end{pmatrix}$$

The matrix shown here has two rows and three columns; it is a 2×3 matrix. If we want to talk about a specific element, we often use subscripts to identify the row and column index of that element. So,

$$A_{2,3} = \text{"element at row} = 2, \text{ column} = 3\text{"} = \frac{7}{5}.$$

Similarly, we often write the indices as variables in order to describe sums, manipulations and so on. So, the sum of the elements in the first column is

$$\sum_{i=1}^2 A_{i,1} = 2 + 11 = 13.$$

3.17.2 Row, Column and Square Matrices

Three particular matrix shapes have special names: A **row** matrix is a matrix with only one row, while a **column** matrix has only one column. Finally a

square matrix, as one would guess, is a matrix with the same number of rows as columns.

$$\mathbf{B} = \begin{pmatrix} 2 & -3 & 4 \end{pmatrix} \quad \mathbf{C} = \begin{pmatrix} 7 \\ 2 \\ -3 \end{pmatrix} \quad \mathbf{D} = \begin{pmatrix} 9 & 4 & 7 \\ 2 & 1 & 0 \\ 3 & -3 & 1 \end{pmatrix}$$

As we will see presently, we can think of row or column matrices as ways of writing the simple vectors that we've been thinking about earlier.

3.17.3 Equating and adding matrices

For two matrices to be equal, they must have the same row/column dimensions, and every element must match:

$$\mathbf{A} = \mathbf{B} \quad \Rightarrow \quad A_{i,j} = B_{i,j} \quad \forall i, j.$$

Matrix **addition** is the simplest operation on matrices: it is defined in the obvious way, i.e. if $\mathbf{C} = \mathbf{A} + \mathbf{B}$ then each element of $\mathbf{C}$ is the sum of the two corresponding elements in $\mathbf{A}$ and $\mathbf{B}$:

$$\mathbf{C} = \mathbf{A} + \mathbf{B} \quad \Rightarrow \quad C_{i,j} = A_{i,j} + B_{i,j}$$

For this to make sense, it is necessary that $\mathbf{A}$ and $\mathbf{B}$ have the same row/column dimensions. From the definition we see at once, that the addition operation is commutative: $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$.

Example

3.17.4 Scalar Multiplication

We can multiply a matrix by a simple number (i.e. a scalar) just by multiplying **every** element by that number:

$$\mathbf{B} = k\mathbf{A} \quad \Rightarrow \quad B_{i,j} = kA_{i,j}$$

3.17.5 Diagonal

When we speak of the diagonal of a matrix, we mean the set of elements going from top-left to bottom-right, i.e. the elements which have their row index and column index equal: $A_{i,i}$.

If we make the statement that a given matrix “is diagonal”, we mean that only these diagonal elements can be non-zero. Normally the term is used only for square matrices. Example:

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

Note: one or more of the diagonal elements might **also** be zero.

3.17.6 Identity Matrix

If a square matrix of any size is diagonal and the elements of the diagonal are all unity, then it is called an identity matrix. Presently we will see why it deserves this name.

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

3.17.7 Transposition

The transpose of a matrix $\mathbf{A}$, is written as $\mathbf{A}^T$ and is defined as the matrix obtained by exchanging the row/column index of every element in the original. Therefore, the transpose will have as many columns as the original matrix has rows, and vice versa. Also, the diagonal elements won’t change (since their row index is the same as their column index). Really the transpose operation is like a reflection of the matrix about its diagonal.

$$\mathbf{A} = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{pmatrix} \quad \mathbf{B} = \mathbf{A}^T = \begin{pmatrix} a_{1,1} & a_{2,1} \\ a_{1,2} & a_{2,2} \\ a_{1,3} & a_{2,3} \end{pmatrix} \quad B_{i,j} = A_{j,i}$$

3.18 Matrix Multiplication

The multiplication of two matrices $\mathbf{A}$ and $\mathbf{B}$ is a more complex operation. It is computationally intensive, involving many different multiplications.

The operation is only possible if the number of **columns** in $\mathbf{A}$ is equal to the number of **rows** in $\mathbf{B}$. To find the elements of $\mathbf{C} = \mathbf{AB}$ we then proceed as follows:

To find $C_{1,1}$ we focus on row number 1 of $\mathbf{A}$ and column number 1 of $\mathbf{B}$. We multiply the first element of the row, by the first element in the column. Similarly we multiply the second element of the row, by the second element of the column, and so on. We then add up all these products: this is the value of $C_{1,1}$.

To find the element of $\mathbf{C}$ at row i , column j , that is to say $C_{i,j}$, we focus on only row number i of $\mathbf{A}$ and column number j of $\mathbf{B}$. We then do the same trick of multiplying the pairs and summing.

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} r & s & t \\ u & v & w \\ x & y & z \end{pmatrix} = \begin{pmatrix} \color{red}{ar+bu+cx} & \square & \square \\ \square & \square & \square \\ \square & \square & \square \end{pmatrix} \quad \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} r & s & t \\ u & v & w \\ x & y & z \end{pmatrix} = \begin{pmatrix} \square & \square & \square \\ \square & \square & \color{red}{dt+ew+fz} \\ \square & \square & \square \end{pmatrix}$$

Figure 3.3: Multiplication.

Formally, we can write the definition of multiplication as follows:

$$C_{i,j} = \sum_k A_{i,k} B_{k,j}$$

Example

Is it now clear why the identity matrix $\mathbf{I}$ is special? It's because $\mathbf{AI} = \mathbf{A}$ for any $\mathbf{A}$. Note: we can easily find an example to prove that matrix multiplication is not commutative, i.e. that in general

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

However, it is associative, meaning $\mathbf{A(BC)} = (\mathbf{AB})\mathbf{C}$, and “distributive over addition”, so $\mathbf{A(B + C)} = \mathbf{AB + AC}$ and similarly $(\mathbf{A + B})\mathbf{C} = \mathbf{AC + BC}$.

3.19 Matrices for representing equations

One way to see why our definition of matrix multiplication is useful, is to consider the use of matrices to represent simultaneous equations. Suppose that we are given the following:

$$\begin{aligned} x + 5y &= -1 \\ 2x + 6y &= 10 \end{aligned}$$

Using matrix multiplication we can write this in a compact form:

$$\begin{pmatrix} 1 & 5 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -1 \\ 10 \end{pmatrix}$$

Now suppose that we want to change variables, to a and b , where $x = 5a - 7b$ and $y = -a + b$. In other words, we are now saying that

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

If we want to write the equations purely in terms of a and b , we can now do so just by substituting our x, y column:

$$\begin{pmatrix} 1 & 5 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 5 & -7 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} -1 \\ 10 \end{pmatrix}$$

and since multiplication is associative, we are free to multiply the first two:

$$\begin{pmatrix} 1 \times 5 + 5 \times -1 & 1 \times -7 + 5 \times 1 \\ 2 \times 5 + 6 \times -1 & 2 \times -7 + 6 \times 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} -1 \\ 10 \end{pmatrix}$$

$$\begin{pmatrix} 0 & -2 \\ 4 & -8 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} -1 \\ 10 \end{pmatrix}.$$

So, we can make the change of variables (and it just so happens that the new variables make one of the equations very simple, since it is saying $-2b = 10$). Of course, really we would like to use matrices to **solve** the equations in a systematic way! This is something we will learn to do later in the course.

3.20 Matrices as deformations of a solid

One very powerful and general use of matrices is for representing transformations. To see what this means, we will begin by writing our familiar $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$ unit vectors as column matrices.

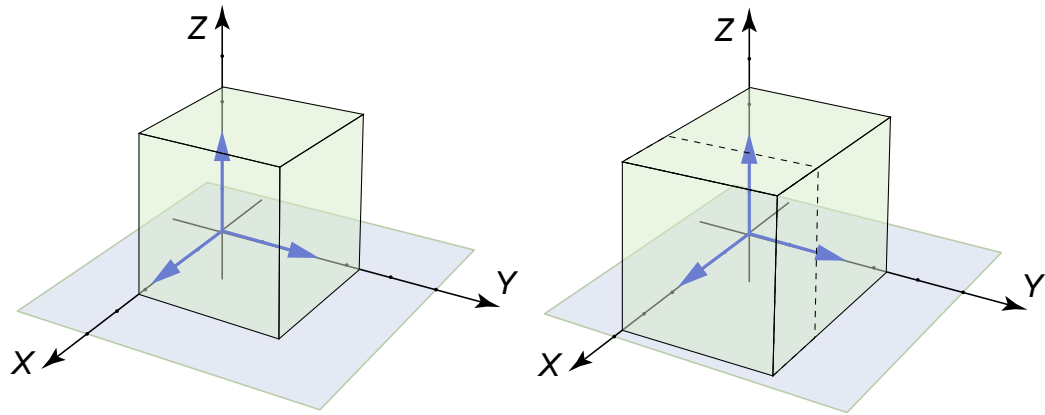
$$\hat{\mathbf{i}} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \hat{\mathbf{j}} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \hat{\mathbf{k}} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Now if we act on any one of these vectors with a 3×3 matrix $\mathbf{M}$, we will get a new column vector of length 3. So, for example,

$$\mathbf{M} \hat{\mathbf{i}} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} = a\hat{\mathbf{i}} + b\hat{\mathbf{j}} + c\hat{\mathbf{k}}$$

Thus we can think of the matrix $\mathbf{M}$ as a transformation that changes $\hat{\mathbf{i}}$ into $a\hat{\mathbf{i}} + b\hat{\mathbf{j}} + c\hat{\mathbf{k}}$. In general it will also transform $\hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$ (but with different constants rather than a, b, c). Now if we imagine an object such as a cube, with position vectors defining each corner, then $\mathbf{M}$ will transform all those vectors — and hence transform the object itself.

3.20.1 Uniaxial strain



Suppose that some object is subjected to a force so that it is uniformly **stretched** in a particular direction. If the stretch is along the x -axis, then the position vector of any atom in the object will transform like this:

$$x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}} \rightarrow x(1 + \epsilon)\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}}.$$

In other words, the x -component (i.e. the component in the direction of $\hat{\mathbf{i}}$) has scaled by $(1 + \epsilon)$, however the other orthogonal components are unchanged. In terms of the three Cartesian unit vectors, we can write

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 + \epsilon \\ 0 \\ 0 \end{pmatrix} \quad \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

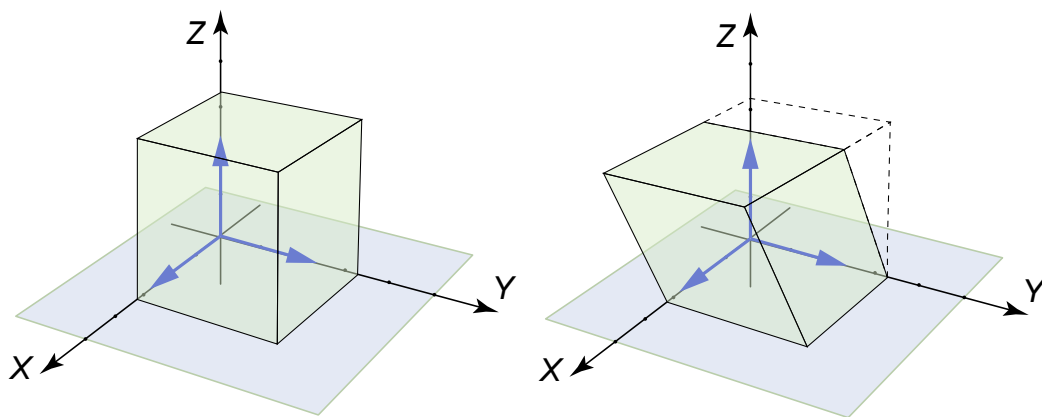
It is easy to see that the matrix $\mathbf{M}$ that will cause this transformation is simply,

$$\mathbf{M} = \begin{pmatrix} 1 + \epsilon & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Note that the matrix is constructed, column-by-column, from the transformed versions of $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$.

3.20.2 Simple shear

A shear is a slightly more complicated transformation. The object is deformed as shown in the figure. We might imagine atomic planes sliding past one another, in response to some force applied to the x - y planes and in the direction of x . Suppose we look at the deformed cube from a viewpoint along the positive y -axis, so seeing things in the x - z plane. We'd notice that the the previously-square face of the cube is now a parallelogram. So the deformation can be said to be **'in' the x - z plane, in the x direction**.



Notice that as before, the position vector of a point inside the volume (or a corner) will pick up an additional component in the $\hat{\mathbf{i}}$ direction, BUT

3: Introducing Matrices

unlike the previous case of simple strain, here this additional $\hat{\mathbf{i}}$ component will now scale according to component of $\hat{\mathbf{k}}$ in the original vector.

$$x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}} \rightarrow (x + \epsilon z)\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}}.$$

Fortunately this remains straightforward to capture in the form of a transformation matrix $\mathbf{M}$. We simply write

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} \epsilon \\ 0 \\ 1 \end{pmatrix}$$

It is easy to see that the matrix $\mathbf{M}$ that will cause this transformation is simply,

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

Thinking about how a cube would be deformed, we see that the sides would ‘lean’ an angle θ where

$$\tan(\theta) = \epsilon.$$

We can similarly work out the shear transformations for other choices for the plane to which the force is applied, and the direction of the force.

Examples

3.20.3 Uniform dilation

Suppose that an object increases in size uniformly (e.g. if expanding freely due to heat). Then we would write

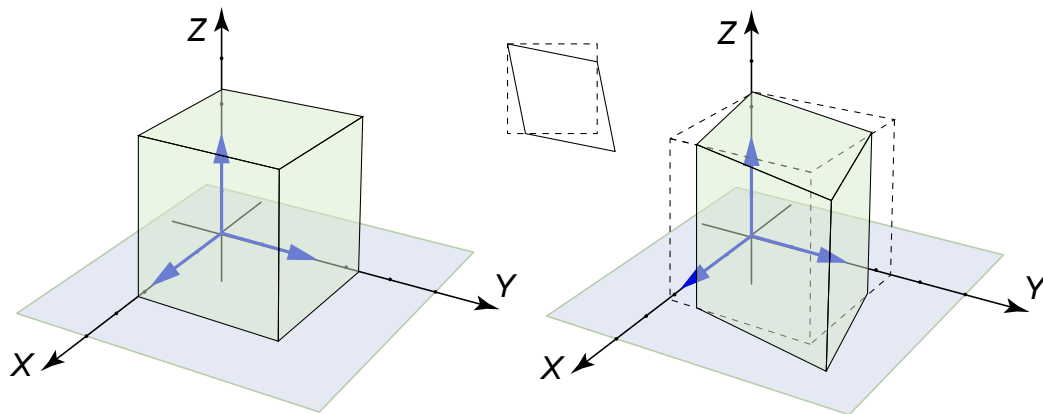
$$x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}} \rightarrow x(1 + \epsilon)\hat{\mathbf{i}} + y(1 + \epsilon)\hat{\mathbf{j}} + z(1 + \epsilon)\hat{\mathbf{k}}.$$

Leading us to

$$\mathbf{M} = \begin{pmatrix} 1 + \epsilon & 0 & 0 \\ 0 & 1 + \epsilon & 0 \\ 0 & 0 & 1 + \epsilon \end{pmatrix}.$$

3.20.4 Pure shear

The diagram shows the effect of a pure shear, where the direction of the extension is $\hat{\mathbf{i}} + \hat{\mathbf{j}}$ and the deformation is said to be ‘**in the x - y plane**’ (we would look at the deformed cube from a viewpoint along the z -axis to see exactly what’s happening).



We see that position vectors will be transformed as

$$x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}} \rightarrow (x + \epsilon y)\hat{\mathbf{i}} + (y + \epsilon x)\hat{\mathbf{j}} + z\hat{\mathbf{k}}.$$

So we will write

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ \epsilon \\ 0 \end{pmatrix} \quad \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} \epsilon \\ 1 \\ 0 \end{pmatrix} \quad \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Thus our transformation matrix $\mathbf{M}$ is

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

3.20.5 Combining deformations

Because matrix multiplication is not commutative (i.e. generally $\mathbf{AB} \neq \mathbf{BA}$) we will find that a strain deformation followed by a shear will yield a different object, versus applying these in reverse order.

Example

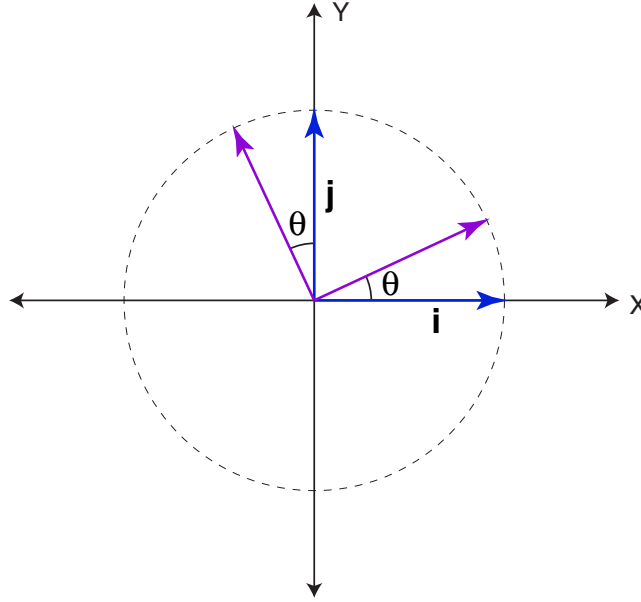
3.21 Rotation matrices

Not all transformations need to deform an object. Instead we can leave object's shape unchanged and simply rotate it.

3.21.1 z -axis rotation

Suppose that we want to describe a rotation of an object about the z -axis. Then the z -coordinate of any point, i.e. the component of $\hat{\mathbf{k}}$ in the posi-

tion vector, will not change. We can focus on the x - y plane to see what is happening.



Imagine looking down towards the x - y plane, from a standpoint along the positive z axis. From the figure, we see how the Cartesian unit vectors are transformed.

$$\hat{\mathbf{i}} \rightarrow \cos(\theta)\hat{\mathbf{i}} + \sin(\theta)\hat{\mathbf{j}} \quad \hat{\mathbf{j}} \rightarrow \cos(\theta)\hat{\mathbf{j}} - \sin(\theta)\hat{\mathbf{i}}.$$

Thus, a vector in the plane is transformed as

$$x\hat{\mathbf{i}} + y\hat{\mathbf{j}} \rightarrow (\cos(\theta)x - \sin(\theta)y)\hat{\mathbf{i}} + (\cos(\theta)y + \sin(\theta)x)\hat{\mathbf{j}}.$$

In our column notation we write

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \\ 0 \end{pmatrix} \quad \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \\ 0 \end{pmatrix} \quad \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

thus our transformation matrix $\mathbf{M}$ is

$$\mathbf{M} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

This rotation is in the anticlockwise direction, when viewed from the standpoint of the positive z axis. If we want a negative rotation, then of course we

3: Introducing Matrices

need only change the sign of the rotation angle, $\theta \rightarrow -\theta$. Then notice that $\cos(-\theta) = \cos(\theta)$ but $\sin(-\theta) = -\sin(\theta)$, so we have

$$\mathbf{P} = \begin{pmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We should find that the matrix product of these two, in either order, will just give the Identity matrix (why?).

$$\begin{aligned} \mathbf{PM} &= \begin{pmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \cos^2(\theta) + \sin^2(\theta) & 0 & 0 \\ 0 & \sin^2(\theta) + \cos^2(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} = \mathbf{I} \end{aligned}$$

Similarly, we can obtain the matrices that represent rotations about the x or the y axes. Note that if we consider two rotations about different axes, they will not commute in general: the order matters.

Examples