Calculus, trigonometry, vectors and geometry

January 20, 2017

Overview

These notes review the minimal mathematics you absolutely will need:

- Algebra, exponentials, logarithms and trigonometry
- Calculus: basic integrals and derivatives, chain and product rule, integration by parts and change of variables
- Vectors: basic vector algebra
- Linear equations
- Quadratic equations
- Geometry: a few basic formulae

These will be required throughout the course. *Learn this material now if any of it seems new. There is a possibility that partial derivatives may be new to you if you are taking multivariable differential calculus as a co-requisite; still, you should read the relevant sections below.*

Algebra

In order to succeed in a subject that involves a lot of calculus, algebra needs to be second nature. Besides the basic rules of associativitiy, commutativity and distributivity you need to know how to manipulate fractions and exponentiation. In particular, you need to know how to add fractions and divide by fractions

$$\frac{1}{a} + \frac{1}{b} = \frac{a+b}{ab}.$$

$$\frac{1}{\frac{1}{a}} = a$$

Do not expect to get much sympathy if you write something like $1/a + 1/b = 1/(a+b)$.

Exponentiation is another area prone to errors. You need to remember the following rules

$$a^b \times a^c = a^{b+c}$$
$$\left(a^b\right)^c = a^{bc}$$
$$a^{-b} = \frac{1}{a^b}$$

Again, do not write things like $1/a^b = a^{1/b}$ or $(a^b)^c = a^{b^c}$.

You should also know how to do binomial expansions,

$$(a+b)^n = \sum_{r=0}^{n} \frac{n!}{r!(n-r)!} a^{n-r} b^r.$$

Note 1 *If you are in doubt whether some equality holds or not, you can often check by trying. For instance, if you had written*

$$1/a + 1/b = 1/(a+b)$$

then substituting $a = b = 1$ *would give* $1/1 + 1/1 = 2$ *on the left, and* $1/(1+1) = 1/2$ *on the right.*

Similarly, if you had written $1/a^b = a^{1/b}$*, trying* $a = 2$*,* $b = 1$ *would give* $1/(2^1) = 1/2$ *on the left,* $2^{1/1} = 2$ *on the right, or if you had put* $(a^b)^c = a^{(b^c)}$*, then* $a = 2$*,* $b = 1$*,* $c = 2$ *would give* $2^2 = 4$ *on the left,* $2^{\left(1^2\right)} = 2^1 = 2$ *on the right.*

You also need to know how to use the rules above in order to solve equations. This is not the place for an exhaustive list — hopefully this is all high school material — but for instance you need to know that if $y = x^a$ then $x = y^{1/a}$.

Exponentials and logarithms

If you can remember the rules above for manipulating exponentials, the dealing with the exponential function $\exp(x)$ should be relatively straightforward:

$$\exp(x) = e^x$$
$$\exp(x)\exp(y) = \exp(x+y)$$
$$\frac{\exp(x)}{\exp(y)} = \exp(x-y)$$
$$[\exp(x)]^a = \exp(ax)$$

Do not write things like $\exp(x) + \exp(y) = \exp(x+y)$.

In principle, if you can remember how to exponentiate, then the rules for manipulating logarithms follow. In this course, the notation 'log' will be reserved for the logarithm with base e. Logarithms with base 10 are really an accident of using a base 10 number system, which in turn probably has a lot to do with humans having 10 fingers. Base 10 logarithms do not naturally appear in physics. The rules you need to remember are

$$\begin{aligned} \log(ab) &= \log(a) + \log(b) \\ b\log(a) &= \log(a^b) \\ -\log(a) &= \log(1/a), \end{aligned}$$

where the last rule is really the second with $b = -1$.

You may also want to use logarithms to solve equations. For instance, if you are given a and c so that

$$a^b = c$$

and need to find b, you can write use the fact that $a = exp(\log(a))$ to write

$$a^b = e^{log(a)b} = c$$

and therefore that

$$\log(a)b = \log(c), \qquad b = \frac{\log(c)}{\log(a)}$$

Trigonoemetry

You should know the basic definitions of $\sin(\theta)$, $\cos(\theta)$ and $\tan(\theta)$ for a right-angled triangle (figure 1). You should also know that

$$\begin{aligned} \tan(\theta) &= \frac{\sin(\theta)}{\cos(\theta)} \\ \sin^2(\theta) + \cos^2(\theta) &= 1 \\ \sin(\alpha+\beta) &= \sin(\alpha)\cos(\beta) + \sin(\beta)\cos(\alpha) \\ \cos(\alpha+\beta) &= \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta). \\ \sin(\alpha)\cos(\beta) &= \frac{1}{2}\left[\sin(\alpha+\beta) + \sin(\alpha-\beta)\right] \\ \sin(\alpha)\sin(\beta) &= \frac{1}{2}\left[\cos(\alpha-\beta) - \cos(\alpha+\beta)\right] \\ \cos(\alpha)\cos(\beta) &= \frac{1}{2}\left[\cos(\alpha+\beta) + \cos(\alpha-\beta)\right] \end{aligned}$$

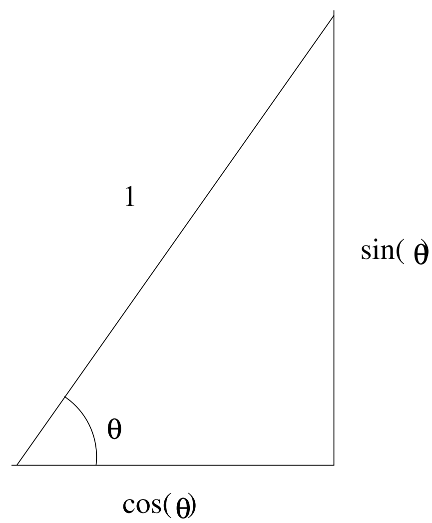

Figure 1: Basic definitions of $\sin(\theta)$ and $\cos(\theta)$.

Calculus

To calculate derivatives and integrals in this course, all you need to know in principle is the table below plus a number of rules that allow you to apply the table below to more complicated functions.

Function $f(x)$	Derivative $f'(x)$	Integral $\int f(x)$
x^n	nx^{n-1}	$x^{n+1}/(n+1)+C$ if $n \neq -1$, $\log(x)+C$ if $n=-1$
$\sin(x)$	$\cos(x)$	$-\cos(x)+C$
$\cos(x)$	$-\sin(x)$	$\sin(x)+C$
$\exp(x)$	$\exp(x)$	$\exp(x)+C$
$\log(x)$	$1/x$	$x\log(x)-x+C$

Rules: Differentiation

The rules you need to know are the chain rule and the product rule. The chain rule is for composite functions,

$$\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x).$$

Example 1 *Let $F(x) = (ax+b)^n$, where a and b are constants. Let $f(y) = y^n$, $g(x) = ax+b$. Then $f'(y) = ny^{n-1}$, $g'(x) = a$, and so*

$$F'(x) = f'(g(x))g'(x) = n(ax+b)^{n-1} \times a = an(ax+b)^{n-1}.$$

The chain rule applies when functions are multiplied with each other,

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x)g(x) = f'(x)g(x) + f(x)g'(x).$$

Example 2 *Let* $F(x) = x\exp(x)$*. Let* $f(x) = x$*,* $g(x) = \exp(x)$*.*

$$F'(x) = f'(x)g(x) + f(x)g'(x) = \exp(x) + x\exp(x) = (x+1)\exp(x).$$

In addition, the following rule is worth knowing: if the function $y = f(x)$ can be inverted as $x = g(y)$, then

$$\frac{\mathrm{d}y}{\mathrm{d}x} = 1/\frac{\mathrm{d}x}{\mathrm{d}y}.$$

Example 3 *Let* $y = f(x) = \exp(x)$*. Then* $x = g(y) = \log(y)$*. We have*

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \exp(x)$$

and

$$\frac{\mathrm{d}x}{\mathrm{d}y} = 1/y = 1/\exp(x).$$

Rules: Integration

When integrating, you are basically trying to find an anti-derivative, i.e.

$$\int f(x)\,\mathrm{d}x = F(x) + C$$

if $F'(x) = f(x)$. This also works the other way round, i.e., if $F(x) = \int f(x)\,\mathrm{d}x$ then $F'(x) = f(x)$.

To apply limits, use

$$\int_a^b f(x)\,\mathrm{d}x = [F(x) + C]_{x=a}^{x=b} = [F(b) + C] - [F(a) + C] = F(b) - F(a).$$

The constant of integration disappears when taking a definite integral. This is also clear from the idea of a Riemann sum taken between definite limits: this has a definite value, so there can be no arbitrary constant of integration that gets added. Also, note that the integration variable x is definite integrals is just a *dummy variable*: we could also write it as y, x', x_1 etc. That is,

$$\int_a^b f(x)\,\mathrm{d}x \equiv \int_a^b f(y)\,\mathrm{d}y \equiv \int_a^b f(x_1)\,\mathrm{d}x_1.$$

This is important when reading certain formulae: in definite integrals, the variable of integration has no special meaning. The only restriction is that it *cannot* be the same as one of the limits. In other words, it is legitimate to write

$$\sin(x) = \int_0^x \cos(y) r dy$$

but *not*

$$\sin(x) = \int_0^x \cos(x)\,\mathrm{d}x :$$

x cannot simultaenously be fixed (x on the left-hand side and the limit) and summed over (integral).

The equivalent of the product rule for integration is a change of variable for integrals. If $y = y(x)$ is *invertible*,[1] then

$$\int f(y)\,\mathrm{d}y = \int f(y(x))\frac{\mathrm{d}y}{\mathrm{d}x}\,\mathrm{d}x.$$

If there are limits, then these are transformed, too

$$\int_a^b f(y)\,\mathrm{d}y = \int_{y^{-1}(a)}^{y^{-1}(b)} f(y(x))\frac{\mathrm{d}y}{\mathrm{d}x}\,\mathrm{d}x,$$

where $y^{-1}(\cdot)$ is the inverse of the function $y(\cdot)$. The change of variables rules can be used in 'both ways':

Example 4 *Consider the integral* $\int \tan(x)\,\mathrm{d}x = \int -\frac{1}{\cos(x)}[-\sin(x)]\,\mathrm{d}x$. *If we let* $y(x) = \cos(x)$, *we can recognize*

$$\int -\frac{1}{\cos(x)}[-\sin(x)]\,\mathrm{d}x = \int -\frac{1}{y(x)}\frac{\mathrm{d}y}{\mathrm{d}x}\,\mathrm{d}x = \int -\frac{1}{y}\,\mathrm{d}y = -\log(y)+C = -\log(\cos(x))+C.$$

Example 5 *Consider the integral* $\int \frac{\mathrm{d}y}{\sqrt{1-y^2}}$. *Let* $y = \sin(x)$, $y'(x) = \cos(x)$ *so*

$$\int \frac{\mathrm{d}y}{\sqrt{1-y^2}} = \int \frac{1}{\sqrt{1-\sin^2(x)}}\cos(x)\,\mathrm{d}x = \int \frac{1}{\cos(x)}\cos(x)\,\mathrm{d}x = x+C = \sin^{-1}(y)+C.$$

Example 6 *Consider the integral* $\int (ax+b)^n\,\mathrm{d}x$. *Let* $u = ax+b$, *from which* $x = (u-b)/a$ *so that* $\mathrm{d}x/\mathrm{d}u = 1/a$ *and*

$$\int (ax+b)^n\,\mathrm{d}x = \int u^n\frac{1}{a}\,\mathrm{d}u = \frac{u^{n+1}}{a(n+1)} + C = \frac{(ax+b)^n}{a(n+1)} + C$$

The equivalent of the product rule for integration is integration by parts:

$$\int f(x)g'(x)\,\mathrm{d}x = f(x)g(x) - \int f'(x)g(x)\,\mathrm{d}x$$

In general, when confronted with the integral of a product, choose f to be a function that becomes *simpler* when differentiated.

[1] Meaning that for each y there is exactly one value of $x = x(y)$

Example 7 *Consider* $\int x \exp(x)\,\mathrm{d}x$. *Let* $f(x) = x$, $g'(x) = \exp(x)$. *Then* $f'(x) = 1$, $g(x) = \exp(x)$. *So*

$$\int f(x)g'(x)\,\mathrm{d}x = x\exp(x) - \int 1 \cdot \exp(x)\,\mathrm{d}x = (x-1)\exp(x) + C.$$

Example 8 *Consider the integral* $\int \log(x)\,\mathrm{d}x$. *Let* $f(x) = \log(x)$, $g'(x) = 1$. *Then* $f'(x) = 1/x$, $g(x) = x$.

$$\int \log(x)\,\mathrm{d}x = x\log(x) - \int (1/x)\cdot x\,\mathrm{d}x = x\log(x) - x + C.$$

Note that you can also check whether you got an indefinite integral right by differentiating it. You should get the integrand back.

Partial derivatives

Partial derivatives help to extend the idea of a derivative to functions of more than one variable. Particulary, the partial derivative of a function $h(x, y)$ with respect to one of its arguments, say x, is essentially the ordinary derivative with respect to x, taken with y held constant:

$$\frac{\partial h}{\partial x} = \lim_{\delta x \to 0} \frac{h(x + \delta x, y) - h(x, y)}{\delta x}.$$

In other word, it is the derivative with respect to x where y is treated like any other constant in the functional form of h.

Example 9 *Let* $h(x, y) = x^2 y^3$. *Then*

$$\frac{\partial h}{\partial x} = 2y^3 x.$$

Similarly,

$$\frac{\partial h}{\partial y} = 3x^2 y^2.$$

Chain and product rules also apply to partial derivatives. For instance, if $F(x, y, z) = f(g(x, y, z))$, then

$$\frac{\partial F}{\partial x} = \frac{\mathrm{d}f}{\mathrm{d}g}\frac{\partial g}{\partial x}.$$

Or, as a slightly more complicated example, if $f(x, y, z) = F(r(x, y, z), \theta(x, y, z), \psi(x, y, z))$, then

$$\frac{\partial f}{\partial x} = \frac{\partial F}{\partial r}\frac{\partial r}{\partial x} + \frac{\partial F}{\partial \theta}\frac{\partial \theta}{\partial x} + \frac{\partial F}{\partial \psi}\frac{\partial \psi}{\partial x},$$

where f, F, r, θ and ψ can be arbitrary functions.

Also, for sufficiently smooth functions of more than one variable, higher order derivatives commute, e.g.,

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x},$$

where

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right), \qquad \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right).$$

Rules: partial differentiation

The product rule is unchanged. For instance, if f and g are functions of (x, y), then

$$\frac{\partial (fg)}{\partial x} = \frac{\partial f}{\partial x} g + f \frac{\partial g}{\partial x}$$

The chain rule is a bit more complicated. Let $f = f(x, y)$ and $x = x(s, t)$, $y = y(s, t)$. Then

$$\frac{\partial f}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s}$$

where $\partial f/\partial s$ of course means the derivative taken at constant t, while $\partial f/\partial x$ means the derivative at constant y and $\partial f/\partial y$ means the derivative at constant x.

Taylor series

A sufficiently 'smooth' function can be approximated near a point $x = x_0$ if its derivatives at that point are known. For instance, if we write

$$\delta x = x - x_0,$$

then the function $f(x)$ can be approximated as[2]

$$f(x) \approx f(x_0) + f'(x_0)\delta x + f''(x_0)\frac{(\delta x)^2}{2!} + f'''(x_0)\frac{(\delta x)^3}{3!} + \ldots$$

for small enough δx.

Note 2 *Taylor series can be justified by noting that*

$$f(x) - f(x_0) = \int_{x_0}^{x} f'(x_1)\,\mathrm{d}x_1.$$

[2] There are some technicalities here, which won't concern us in this course, surrounding the convergence of this series as more and more terms are included.

Beware that x_1 here is just a dummy variable*: it doesn't matter what we use as the integration variable. But*

$$\int_{x_0}^{x} f'(x_1)\,\mathrm{d}x_1 \int_{x_0}^{x} f'(x_0)\,\mathrm{d}x_1 + \int_{x_0}^{x} f'(x_1) - f'(x_0)\,\mathrm{d}x_1$$
$$f'(x_0)(x - x_0) + \int_{x_0}^{x} f'(x_1) - f'(x_0)\,\mathrm{d}x_1$$

This gives

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \int_{x_0}^{x} f'(x_1) - f'(x_0)\,\mathrm{d}x_1, \qquad (1)$$

The right-hand side gives the first two terms of the Taylor series, plus a correction (the integral) which is small if $f'(x)$ does not change much over the interval from x_0 to x (so that $f'(x_1) \approx f'(x_0)$ in the interval).

To get more terms in the Taylor series, we can apply the same approach again to $f'(x_1) - f'(x_0)$,

$$f'(x_1) - f'(x_0) = f''(x_0)(x_1 - x_0) + \int_{x_0}^{x_1} f''(x_2) - f''(x_0)\,\mathrm{d}x_2, \qquad (2)$$

where x_2 is another dummy variable. Substituting for $f'(x_1) - f'(x_0)$ in (1) gives

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \int_{x_0}^{x} f''(x_0)(x_1 - x_0)\,\mathrm{d}x_1 + \int_{x_0}^{x} \left[\int_{x_0}^{x_1} f''(x_2) - f''(x_0)\,\mathrm{d}x_2\right] \mathrm{d}x_1.$$

We can work out the first integral on the right-hand side exactly

$$\int_{x_0}^{x} f''(x_0)(x_1 - x_0)\,\mathrm{d}x_1 = f''(x_0)\left[(x_1 - x_0)^2/2\right]_{x_0}^{x} = f''(x_0)\frac{(x - x_0)^2}{2},$$

and substituting into (1) gives us the first three terms of the Taylor series plus another correction term:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + f''(x_0)\frac{(x - x_0)^2}{2} + \int_{x_0}^{x} \left[\int_{x_0}^{x_1} f''(x_2) - f''(x_0)\,\mathrm{d}x_2\right] \mathrm{d}x_1.$$

We can continue like this to get any number of terms in the Taylor series: for instance, we can rewrite $f''(x_2) - f''(x_0)$ in the same way we re-wrote $f'(x_1) - f'(x_0)$ in (2) to get the next term in the Taylor series.

Example 10 *Let $f(x) = \cos(x)$ and $x_0 = 0$. Then $f(x_0) = \cos(0) = 1$, $f'(x_0) = -\sin(0) = 0$, $f''(x_0) = -\cos(0) = -1$, $f'''(x_0) = \sin(0) = 0$ etc, and $\delta x = x - x_0 = x$, so*

$$f(x) \approx 1 - x^2/2 + \ldots$$

This also works for partial derivatives, e.g. if $\delta x = x - x_0$, $\delta y = y - y_0$, and the notation

$$\left.\frac{\partial f}{\partial x}\right|_{x=x_0, y=y_0}$$

means $\partial f/\partial x$ evaluated at the point (x_0, y_0), then

$$\begin{aligned} f(x,y) \;\approx\; & f(x_0,y_0) + \left.\frac{\partial f}{\partial x}\right|_{x=x_0,y=y_0} \delta x + \left.\frac{\partial f}{\partial y}\right|_{x=x_0,y=y_0} \delta y \\ & + \left.\frac{\partial^2 f}{\partial x^2}\right|_{x=x_0,y=y_0} \frac{(\delta x)^2}{2!} + 2\left.\frac{\partial^2 f}{\partial x \partial y}\right|_{x=x_0,y=y_0} \frac{\delta x \delta y}{2!} + \left.\frac{\partial^2 f}{\partial y^2}\right|_{x=x_0,y=y_0} \frac{(\delta y)^2}{2!} + \dots \end{aligned}$$

Vectors

Vectors are quantities that have ‘magnitude and direction’. This is a fairly vague statement. In a more concrete way, think of a vector as having components in some coordinate system. For instance

$$\mathbf{a} = 2\mathbf{i} + 3\mathbf{j} - 4\mathbf{k}$$

has component 2 along the x-axis, component 3 along the y-axis, and component -4 along the z-axis. Here **i**, **j** and **k** are the unit vectors along the x-, y- and z-axes, respectively, and we use bold face to denote vectors (you may also have seen vectors written as letters with arrows over them).

Another way of writing the vector **a** would be in column form,

$$\mathbf{a} = \begin{pmatrix} 2 \\ 3 \\ -4 \end{pmatrix}.$$

or in row form,

$$\mathbf{a} = (2, 3, -4).$$

If you are familiar with this last form, then all you need to remember is that the unit vector **i**, **j** and **k** are then simply

$$\mathbf{i} = (1,0,0), \qquad \mathbf{j} = (0,1,0), \qquad \mathbf{k} = (0,0,1),$$

so that writing $\mathbf{a} = 2\mathbf{i} + 3\mathbf{j} - 4\mathbf{k}$ means the same as

$$\mathbf{a} = 2\times(1,0,0)+3\times(0,1,0)-4\times(0,0,1) = (2,0,0)+(0,3,0)+(0,0,-4) = (2,3,-4).$$

Addition and multiplication by scalars

Given two vectors $\mathbf{a} = a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k}$ and $\mathbf{b} = b_x\mathbf{i} + b_y\mathbf{j} + b_z\mathbf{k}$, we have addition and subtraction defined through

$$\mathbf{a} + \mathbf{b} = (a_x + b_x)\mathbf{i} + (a_y + b_y)\mathbf{j} + (a_z + b_z)\mathbf{k}.$$

Example 11 *Let* $\mathbf{a} = 2\mathbf{i} + 3\mathbf{j} - 4\mathbf{k}$ *as above, and* $\mathbf{b} = 3\mathbf{i} + \mathbf{j} + 7\mathbf{k}$. *Then*

$$\mathbf{a} + \mathbf{b} = 5\mathbf{i} + 4\mathbf{j} + 3\mathbf{k}.$$

Similarly, if $\mathbf{a} = a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k}$ and λ is a scalar (i.e., just a number) then their product $\lambda\mathbf{a}$ is defined as

$$\lambda\mathbf{a} = \lambda a_x\mathbf{i} + \lambda a_y\mathbf{j} + \lambda a_z\mathbf{k}.$$

Example 12 *Let* $\mathbf{a} = 2\mathbf{i} + 3\mathbf{j} - 4\mathbf{k}$ *as above, and* $\lambda = -3$. *Then*

$$\lambda\mathbf{a} = -6\mathbf{i} - 9\mathbf{j} + 12\mathbf{k}.$$

Vector addition and multiplication by a scalar are commutative and distributive in the usual way:

$$\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a},$$
$$\lambda(\mathbf{a} + \mathbf{b}) = \lambda\mathbf{a} + \lambda\mathbf{b}.$$

These results follow from the definition of addition and scalar multiplication above.

Magnitudes and vector products

The magnitude of a vector $\mathbf{a} = a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k}$ can be thought of as the 'length' of the vector, and is defined as the root of the sum of the squares of its components,

$$|\mathbf{a}| = \sqrt{a_x^2 + a_y^2 + a_z^2}.$$

This definition can be linked to the idea of the length of a vector as a line through Pythagoras' Theorem.

Example 13 *Let* $\mathbf{a} = 2\mathbf{i} + 3\mathbf{j} - 4\mathbf{k}$ *as above. Then*

$$|\mathbf{a}| = \sqrt{2^2 + 3^2 + (-4)^2} = \sqrt{29}.$$

The scalar product of two vectors $\mathbf{a} = a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k}$ and $\mathbf{b} = b_x\mathbf{i} + b_y\mathbf{j} + b_z\mathbf{k}$ is defined as

$$\mathbf{a} \cdot \mathbf{b} = a_xb_x + a_yb_y + a_zb_z.$$

This is also equal to the product of the magnitudes of the two vectors times the cosine of the angle between them,

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos(\theta),$$

where θ is the angle between the vectors. Clearly, we have

$$\mathbf{a} \cdot \mathbf{a} = |\mathbf{a}|^2$$

from the definition of the scalar product.

Example 14 *Let* $\mathbf{a} = \mathbf{i}$ *and* $\mathbf{b} = \mathbf{j}$*. Clearly these two vectors are at right angles to each other: one is parallel to the* x*-axis, the other is parallel to the* y*-axis. And sure enough, we have* $a_x = 1$, $a_y = 0$, $a_z = 0$, *while* $b_x = 0$, $b_y = 1$, $b_z = 0$, *and so*

$$\mathbf{a} \cdot \mathbf{b} = a_x b_x + a_y b_y + a_z b_z = 0.$$

The scalar product satisfies the usual rules of commutativity and distributivity,

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$$
$$\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c}.$$

Again these rules follow from the definition of the vector product above.

Scalar products can be used to figure out the components of a vector in the direction of the coordinate axes. For instance, if $\mathbf{a} = a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k}$, then

$$a_x = \mathbf{a} \cdot \mathbf{i}, \qquad a_y = \mathbf{a} \cdot \mathbf{j}, \qquad a_z = \mathbf{a} \cdot \mathbf{k}.$$

In other words, we can write

$$\mathbf{a} = (\mathbf{a} \cdot \mathbf{i})\mathbf{i} + (\mathbf{a} \cdot \mathbf{j})\mathbf{j} + (\mathbf{a} \cdot \mathbf{k})\mathbf{k}.$$

This is also sometimes referred to as decomposing $\mathbf{a}$ into its x-, y- and z-components.

The vector product of two vectors $\mathbf{a} = a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k}$ and $\mathbf{b} = b_x\mathbf{i} + b_y\mathbf{j} + b_z\mathbf{k}$ is written as

$$\mathbf{a} \times \mathbf{b}.$$

This is defined as a vector whose direction is perpendicular to both, $\mathbf{a}$ and $\mathbf{b}$, and whose magnitude is the product of the magnitudes of the two vectors times the sine of the angle between them, so that

$$|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}||\mathbf{b}| \sin(\theta),$$

while (on account of the direction of $\mathbf{a} \times \mathbf{b}$), we have

$$(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{a} = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{b} = 0.$$

In terms of components, the vector product can be written as

$$\mathbf{a} \times \mathbf{b} = (a_x b_y - b_x a_y)\mathbf{k} + (a_y b_z - b_y a_z)\mathbf{i} + (a_z b_x - b_z a_x)\mathbf{j}.$$

The vector product is distributive, but not commutative:

$$\begin{aligned} \mathbf{a} \times \mathbf{b} &= -\mathbf{b} \times \mathbf{a} \\ \mathbf{a} \times (\mathbf{b} + \mathbf{c}) &= \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}. \end{aligned} \tag{3}$$

Example 15 *If you cannot memorize the formula for the vector product in terms of the components of the vectors given above, you can always work it by recognizing the following vector products of the unit vectors* $\mathbf{i}$*,* $\mathbf{j}$ *and* $\mathbf{k}$*,*

$$\mathbf{i} \times \mathbf{j} = \mathbf{k}, \qquad \mathbf{j} \times \mathbf{k} = \mathbf{i}, \qquad \mathbf{k} \times \mathbf{i} = \mathbf{j}.$$

and using the commutativity and distributivity rules above:

$$\begin{aligned} \mathbf{a} \times \mathbf{b} &= (a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k}) \times (b_x\mathbf{i} + b_y\mathbf{j} + b_z\mathbf{k}) \\ &= a_x\mathbf{i} \times b_x\mathbf{i} + a_x\mathbf{i} \times b_y\mathbf{j} + a_x\mathbf{i} \times b_z\mathbf{k} \\ &\quad + a_y\mathbf{j} \times b_x\mathbf{i} + a_y\mathbf{j} \times b_y\mathbf{j} + a_y\mathbf{j} \times b_z\mathbf{k} \\ &\quad + a_z\mathbf{k} \times b_x\mathbf{i} + a_z\mathbf{k} \times b_y\mathbf{j} + a_z\mathbf{k} \times b_z\mathbf{k} \\ &= a_x b_y\mathbf{k} + a_x b_z(-\mathbf{j}) \\ &\quad + a_y b_x(-\mathbf{k}) + a_y b_z\mathbf{i} \\ &\quad + a_z b_x\mathbf{j} + a_z b_y(-\mathbf{i}) \\ &= (a_x b_y - b_x a_y)\mathbf{k} + (a_y b_z - b_y a_z)\mathbf{i} + (a_z b_x - b_z a_x)\mathbf{j}. \end{aligned}$$

You may also have seen other ways of doing this, for instance writing the vector product in the 'determinant' form

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix}.$$

Use whichever method works best for you.

Linear systems of equations

A system of linear equations takes the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ &\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m \end{aligned}$$

where all the coefficients a_{11}, a_{12}, $\dots a_{mn}$ are known, as are the coefficients b_1, b_2, $\dots b_m$. The variables x_1, x_2 etc are the unknowns. An example would be

$$x_1 + 2x_2 + 3x_3 = 0 \tag{5a}$$
$$x_1 - 2x_2 = 1 \tag{5b}$$

Often, you will see x, y and z used instead of x_1, x_2, x_3, in which case (5) might be

$$x + 2y + 3z = 0 \tag{6a}$$
$$x - 2y = 1 \tag{6b}$$

This works well if you have three or fewer unknowns, but not so well when you have more.

To solve a set of linear equations uniquely (meaning, so that there is a single answer for x_1, x_2 etc), you generally need to have as many equations as you have unknowns. In terms of (4), that would mean $m = n$. Consequently (5) does not have a unique solution: any combination of x_1 and x_2 satisfiying $x_1 - 2x_2 = 1$ will be a solution, provided we also put $x_3 = -(x_1 + 2x_2)/3$. To get a unique solution, you have to introduce another equation.

The most straightforward (but also laborious) way of solving linear equations is by *Gaussian elimination*. Start with one variable (say x_n) and use the last equation to express x_n in terms of the x_1, x_2 etc, here

$$x_n = (b_m - a_{m1}x_1 - a_{m2}x_2 - \dots - a_{m(n-1)}x_{n-1}))/a_{mn}.$$

The substitute this into the remaining equations, so you get

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}(b_m - a_{m1}x_1 - a_{m2}x_2 - \dots - a_{m(n-1)}x_{n-1})/a_{mn} &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}(b_1 - a_{11}x_1 - a_{12}x_2 - \dots - a_{m(n-1)}x_{n-1})/a_{mn} &= b_2 \\
&\dots \\
a_{(m-1)1}x_1 + a_{(m-1)2}x_2 + \dots + a_{(m-1)n}(b_m - a_{m1}x_1 - a_{m2}x_2 - \dots - a_{m(n-1)}x_{n-1})/a_{mn} &= b_{m-1}
\end{aligned}$$

Now you can collect coefficients of x_1, x_2 etc,

$$\begin{aligned}
\left(a_{11} - \frac{a_{1n}a_{m1}}{a_{mn}}\right)x_1 + \dots + \left(a_{1(n-1)} - \frac{a_{1n}a_{m(n-1)}}{a_{mn}}\right)x_{n-1} &= b_1 - \frac{a_{1n}b_m}{a_{mn}} \\
&\dots \\
\left(a_{(m-1)1} - \frac{a_{(m-1)n}a_{m1}}{a_{mn}}\right)x_1 + \dots + \left(a_{(m-1)(n-1)} - \frac{a_{(m-1)n}a_{m(n-1)}}{a_{mn}}\right)x_{n-1} &= b_{m-1} - \frac{a_{(m-1)n}b_1}{a_{mn}}
\end{aligned}$$

This takes the form of a smaller system of linear equations with $m-1$ equations and $n-1$ unknowns,

$$\begin{aligned}
a'_{11}x_1 + a'_{12}x_2 + \dots + a'_{1(n-1)}x_{n-1} &= b_{1-1} \\
&\dots \\
a'_{(m-1)1}x_1 + a'_{(m-1)2}x_2 + \dots + a'_{(m-1)(n-1)}x_{n-1} &= b'_{m-1}
\end{aligned}$$

if we define

$$a'_{11} = a_{11} - a_{1n}a_{m1}/a_{mn}, \qquad a_{12'} = a_{12} - a_{1n}a_{m2}/a_{mn}, \qquad \ldots \qquad b'_1 = b_1 - a_{1n}b_m/a_{mn} \qquad \text{etc.}$$

We can then use the same method again and again. For the case of n equations in n unknowns, we hope to arrive at one equation in one unknown (say x_1), from which we can then find the other unknowns x_2, x_3 by substituting back in.

This all looks very complicated, but if you are working with numerical coefficients rather than with general coefficeints denoted by letters, things get simpler.

Example 16 *Take as an example* (6)*. As discussed, to get a unique answer you need to add another equation. Let this be*

$$2x + y + 3z = 0. \tag{6c}$$

First, eliminate z. You can do this either by writing $z = (-2x - y)/3$ from (6c)*, or by subtracting* (6c) *and from* (6a) *to get*

$$(x + 2y + 3z) - (2x + y + 3z) = 0 - 0$$

which becomes

$$-x + y = 0$$

this in combination with (6b) *gives a system of two equations in two unknowns:*

$$-x + y = 0$$
$$x - 2y = 1$$

In this case, it is easier to eliminate x than y: simply add the two equations to get

$$-y = 1$$

or

$$y = -1.$$

Then, working backwards, we also had

$$-x + y = 0$$

so

$$x = y = -1.$$

Also, we had $2x + y + 3z = 0$, or

$$z = -(2x + y)/3 = -1.$$

Note 3 *Having as many equations as unknowns is not actually a guarantee of having a unique solution, or any solution at all. Take for instance* (6) *but with* (6c) *replaced by*

$$2x + 3z = 0. \tag{6d}$$

Again eliminate z by subtracting (6d) *from* (6a)*:*

$$(x + 2y + 3z) - (2x + 3z) = 0 - 0,$$

which becomes

$$2y - x = 0. \tag{6e}$$

However, in addition to (6e)*, we also have* (6b)*,*

$$x - 2y = 1.$$

Adding the two to eliminate y, we find

$$(2y - x) + (x - 2y) = 0 + 1,$$

which becomes

$$0 = 1$$

because eliminating y eliminates x at the same time. As a result, there cannot be a solutoin.

A different situation would occur if, instead of (6d)*, we had*

$$2x + 3z = 1. \tag{6f}$$

Following the same procedure of subtracting (6a) *from* (6f) *to eliminate z gives*

$$(x + 2y + 3z) - (2x + 3z) = 0 - 1,$$

or

$$-x + 2y = -1.$$

This is however the same as (6b) *with both sides multiplied by -1. Therefore we do not have a unique solution; any combination of x and y satisfying $x - 2y = 1$ solves the problem, with $z = (1 - 2x)/3$.*

Note 4 *Linear algebra is a much more systematic way of doing what we have done for a few examples above, and more. Instead of writing the linear system* (4) *in the form above, we would instead write*

$$\mathbf{Ax} = \mathbf{b}$$

where $\mathbf{A}$ *is a* matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{n1} \\ a_{21} & a_{22} & \dots & a_{n2} \\ & & \dots & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

and $\mathbf{x}$ *and* $\mathbf{b}$ *are* column vectors

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{pmatrix}$$

and the multiplication of a column vector by a matrix gives the i*th component of* $\mathbf{b}$*,* b_i*, as*

$$b_i = \sum_{j=1}^{n} A_{ij} x_j.$$

Linear algebra then tells you, among other things, about conditions under which $\mathbf{A}$ *has an inverse* $\mathbf{A}^{-1}$ *such that*

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

and how to compute that inverse.

Quadratic equations

A polynomial is a function of the form

$$f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n.$$

In fact, you can think of a Taylor series as trying to approximate a more complicated function by a polynomial. In general, you cannot solve a *polynomial equation*

$$a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n = 0$$

analytically.

However, when the polynomial is quadratic ($n = 2$), then you can. The idea is that you want to write a general quadratic equation

$$ax^2 + bx + c = 0$$

in the form

$$(x + k)^2 = d.$$

This suggests that

$$x + k = \sqrt{d}$$

However, you have to remember that $(-\sqrt{d})^2 = (\sqrt{d})^2 = d$, so there is a second possiblity,

$$-(x+k) = \sqrt{d}.$$

or

$$x + k = -\sqrt{d}.$$

In short, we write this as

$$x + k = \pm\sqrt{d},$$

so that

$$x = -k \pm \sqrt{d}.$$

How do we get a general quadratic equation ax^2+bx+c into the form $(x+k)^2 = d$? First, get rid of the coefficient of x^2 by dividing by a:

$$x^2 + \frac{b}{a}x + \frac{c}{a}. \tag{7}$$

The trick now is to remember that

$$(x+k)^2 = x^2 + 2kx + k^2,$$

so if we define k as $k = b/(2a)$, then (7) can be written as

$$x^2 + 2kx + \frac{c}{a} = 0$$

To get it into the form $(x+k)^2 = d$, we can recognize that we still need a term $+k^2$ on the left. However, we cannot make it appear out of thin air, so we add k^2 to both, left and right,

$$x^2 + 2kx + k^2 + \frac{c}{a} = k^2,$$

or

$$(x+k)^2 + \frac{c}{a} = k^2.$$

This is called *completing the square*. Rearranging,

$$(x+k)^2 = k^2 - \frac{c}{a}.$$

This means we should define $d = k^2 - c/a$.

As

$$x = -k \pm \sqrt{d},$$

this means, with $k = b/(2a)$ and $d = k^2 - c/a = b^2/4a^2 - c/a = (b^2 - 4ac)/(4a^2)$, we get

$$x = -\frac{b}{2a} \pm \sqrt{\frac{b^2 - 4ac}{4a^2}},$$

or, written more succinctly,

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

You can try to memorize this formula. Much better, understand the method and you can always re-derive the formula.

Note 5 *There are cases where you cannot get a solution that consists of ordinary real numbers. The formula above involves the square root of $b^2 - 4ac$; however, $b^2 - 4ac$ could easily be negative. An easy way of seeing this is to try to solve the quadratic equation $x^2 + 1 = 0$, which is already in the form $(x + k)^2 = d$, with $k = 0$, $d = -11$. Imaginary numbers were invented to handle situations like this.*

Note 6 *You should also know what it means to factorize a polynomial. That means to write it in the form*

$$f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n = a_n(x - x_1)(x - x_2)(x - x_3)\ldots(x - x_n).$$

The constants x_1, x_2 etc. are such that $f(x_1) = f(x_2) = f(x_3) = \ldots = f(x_n) = 0$ (because putting $x = x_1$, $x = x_2$ etc in the product above will ensure that one of the factors in the product is zero). As a result, factorizing a polynomial is equivalent to finding its roots.

Example 17 *Take*

$$2x^2 + 8x + 6 = 0$$

Instead of simply applying the formula above, we can complete the square. First, get rid of the coefficient of x^2 by dividing by 2:

$$x^2 + 4x + 3 = 0.$$

To complete the square, look at the linear term, in this case $4x$. We want this to take the form $2kx$, which in this case requires $k = 2$. We also need a term $k^2 = 2^2$ on the left, so we add it to both sides:

$$x^2 + 2 \times 2 \times x + 2^2 + 3 = 2^2.$$

Collect the first three terms into one 'perfect square' $(x + 2)^2$ on the left, and put $2^2 = 4$ on the right,

$$(x + 2)^2 + 3 = 4.$$

Now take 3 to the right,

$$(x + 2)^2 = 1.$$

Taking the square root and allowing a plus or a minus sign gives

$$x + 2 = \pm 1$$

so

$$x = -1 - 2 \qquad or \qquad x = 1 - 2$$

meaning

$$x = -3 \qquad or \qquad x = -1.$$

Sometimes you can also 'spot' an answer without going through the whole procedure. For instance, you might spot that expanding $(x+3)(x+1)$ *gives*

$$\begin{aligned}(x+3)(x+1) &= x^2 + x + 3x + 3 \\ &= x^2 + 4x + 3\end{aligned}$$

so $x^2 + 4x + 3 = 0$ *can be written as*

$$(x+3)(x+1) = 0.$$

In order for a product of two factors to be zero, one of the factors must be zero, so either $x + 3 = 0$ *or* $x + 1 = 0$*, giving* $x = -3$ *or* -1*.*

Geometry

You should know a few basic results about volumes and areas of standard shapes like circles, triangles, cylinders, spheres, prisms and tetrahedra. Some of these are given in the tables below

	circumference	area
Triangle	Sum of side lengths	$\frac{1}{2}$ base $\times$ height
circle	$2\pi r$	πr^2

Table 1: Circumferences and areas of two-dimensional shapes

	surface area	volume
Cylinder	$2\pi rh$ (curved part only)	$\pi r^2 h$
Sphere	$4\pi r^2$	$\frac{4}{3}\pi r^3$
Tetrahedron	sum of surface triangles	$\frac{1}{3}$ base $\times$ height
Prism	sum of surfaces	base $\times$ height

Table 2: Surface areas and volumes of some common three-dimensional shapes

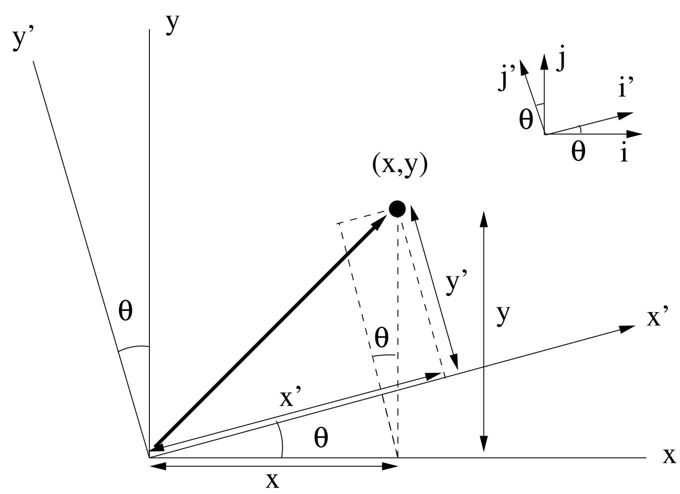

Figure 2: A rotated coordinate system (x', y') with unit vectors $\mathbf{i}'$ and $\mathbf{j}'$.

More on vectors: the meaning of 'magnitude and direction'

The material in this section is not required for the course but is useful.

For our purposes, knowing that a physically (or geometrically) defined vector 'has a magnitude and a direction', and using the rules above, suffices. There is however a deeper meaning to the idea that a vector has a magnitude and a direction: namely, that these should in some way not depend on the choice of coordinate system. A vector that connects point A to point B in one Cartesian coordinate system should still connect A to B in, say, a Cartesian coordinate system that has been rotated with respect to the original one. Of course, the direction as 'seen' in the new coordinate system will be different from that 'seen' in the original coordinate system, but the vector still connects A to B, and its length remains the same.

To get a simple understanding of this, consider a vector $x\mathbf{i}+y\mathbf{j}$ in a two-dimensional Cartesian coordinate system with coordinates (x, y). This vector can be thought of as connecting the origin to the point with coordinates (x, y). Now rotate the coordinate axis through an angle θ anticlockwise to give new coordinate axes with coordinates (x', y'). Clearly, we have from figure 2

$$x' = x\cos(\theta) + y\sin(\theta), \qquad y' = -\sin(\theta)x + \cos(\theta)y.$$

Similarly, we can define unit vectors parallel to the x'- and y' axis as $\mathbf{i}'$ and $\mathbf{j}'$, respectively (figure 2). These can be related to $\mathbf{i}$ and $\mathbf{j}$ by looking at the x- and y-components of $\mathbf{i}'$ and $\mathbf{j}'$. From the discussion above, we have

$$\mathbf{i}' = (\mathbf{i}' \cdot \mathbf{i})\mathbf{i} + (\mathbf{i}' \cdot \mathbf{j})\mathbf{j},$$
$$\mathbf{j}' = (\mathbf{j}' \cdot \mathbf{i})\mathbf{i} + (\mathbf{j}' \cdot \mathbf{j})\mathbf{j},$$

or equally

$$\mathbf{i} = (\mathbf{i}\cdot\mathbf{i}')\mathbf{i}' + (\mathbf{i}\cdot\mathbf{j}')\mathbf{j}',$$
$$\mathbf{j} = (\mathbf{j}\cdot\mathbf{i}')\mathbf{i}' + (\mathbf{j}\cdot\mathbf{j}')\mathbf{j}',$$

But we also know that the dot products are the products of the magnitudes of the vectors times the cosines of the angles between the different vectors. As all the vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{i}'$ and $\mathbf{j}'$ are unit vectors, this means that the dot products are just the cosines of the angles between the vectors, and from figure 2,

$$\begin{aligned}
\mathbf{i}'\cdot\mathbf{i} &= \cos(\theta),\\
\mathbf{i}'\cdot\mathbf{j} &= \cos(\pi/2-\theta) = \sin(\theta),\\
\mathbf{j}'\cdot\mathbf{i} &= \cos(\pi/2+\theta) = -\sin(\theta),\\
\mathbf{j}'\cdot\mathbf{j} &= \cos(\theta).
\end{aligned}$$

Therefore, if we have a vector defined by $\mathbf{a} = a_x\mathbf{i} + a_y\mathbf{j}$, then

$$\begin{aligned}
\mathbf{a} = a_x\mathbf{i} + a_y\mathbf{j}&\\
&= a_x\left[(\mathbf{i}\cdot\mathbf{i}')\mathbf{i}' + (\mathbf{i}\cdot\mathbf{j}')\mathbf{j}'\right] + a_y\left[(\mathbf{j}\cdot\mathbf{i}')\mathbf{i}' + (\mathbf{j}\cdot\mathbf{j}')\mathbf{j}'\right]\\
&= \left[a_x\mathbf{i}\cdot\mathbf{i}' + a_y\mathbf{j}\cdot\mathbf{i}'\right]\mathbf{i}' + \left[a_x\mathbf{i}\cdot\mathbf{j}' + a_y\mathbf{j}\cdot\mathbf{j}'\right]\mathbf{j}'\\
&= \left[a_x\cos(\theta) + a_y\sin(\theta)\right]\mathbf{i}' + \left[-a_x\sin(\theta) + a_y\cos(\theta)\right]\mathbf{j}'
\end{aligned}$$

We can write this as $\mathbf{a} = a'_x\mathbf{i}' + a'_y\mathbf{j}'$ if we define

$$a'_x = a_x\cos(\theta) + a_y\sin(\theta), \qquad a'_y = -a_x\sin(\theta) + a_y\cos(\theta).$$

Clearly, the x' and y' components of the vector in the new coordinate system are related in a very specific way to the x and y coordinates in the old coordinate system — in fact, the relationship between (a'_x, a'_y) and (a_x, a_y) is analogous to the relationship between (x', y') and (x, y) above, as was to be expected as we could express the point (x, y) as a position vector $x\mathbf{i}+y\mathbf{j}$. The relationship can be written in the matrix form

$$\begin{pmatrix} a'_x \\ a'_y \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}\begin{pmatrix} a_x \\ a_y \end{pmatrix}, \tag{8}$$

or symbolically,

$$\mathbf{a}' = \mathbf{R}\mathbf{a}, \tag{9}$$

where $\mathbf{a}' = (a'_x, a'_y)^{\mathrm{T}}$, $\mathbf{a} = (a_x, a_y)^{\mathrm{T}}$ (note that the superscript T denotes the transpose of a vector or matrix) and

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

It is important that the length of the vector — which we have *defined* as $\sqrt{a_x^2 + a_y^2}$ — should not depend on the coordinate system. But

$$\begin{aligned}
|\mathbf{a}|^2 &= a_x'^2 + a_y'^2 \\
&= (a_x \cos(\theta) + a_y \sin(\theta))^2 + (-a_x \sin(\theta) + a_y \cos(\theta))^2 \\
&= a_x^2 \cos^2(\theta) + 2a_x a_y \cos(\theta)\sin(\theta) + a_y^2 \sin^2(\theta) + a_x^2 \sin^2(\theta) - 2a_x a_y \sin(\theta)\cos(\theta) + a_y^2 \cos^2(\theta) \\
&= a_x^2 \left[\cos^2(\theta) + \sin^2(\theta)\right] + a_y^2 \left[\cos^2(\theta) + \sin^2(\theta)\right] \\
&= a_x^2 + a_y^2.
\end{aligned}$$

As required, the formula for computing the length of a vector does not depend on the particular Cartesian coordinate system chosen. $|\mathbf{a}|$ is said to be *invariant* under a change of coordinate system. This makes $|\mathbf{a}| = \sqrt{a_x^2 + a_y^2 + a_z^2}$ a somewhat special combination of the components a_x, a_y and a_z: For instance, if we had defined a function f as $f(\mathbf{a}) = 2a_x^2 + a_y^2$, we would have in terms of the components a_x', a_y' that

$$f(\mathbf{a}) = a_x'^2 + a_y'^2 + (a_x' \cos(\theta) - a_y'(\sin(\theta))^2,$$

which is not of the form $2a_x'^2 + a_y'^2$. Functions that depend on the choice of coordinate system like $f(\mathbf{a})$ are problematic in physical theories: physics does not depend on the way of measuring position chosen.

There is a link with the matrix formulation (9) here: in standard matrix notation,

$$|\mathbf{a}'|^2 = \mathbf{a}'^{\mathrm{T}}\mathbf{a}'$$

where

$$\mathbf{a}' = \begin{pmatrix} a_x' \\ a_y' \end{pmatrix}.$$

But we have $\mathbf{a}' = \mathbf{R}\mathbf{a}$, and hence

$$\begin{aligned}
|\mathbf{a}'|^2 &= (\mathbf{R}\mathbf{a})^{\mathrm{T}}\,\mathbf{R}\mathbf{a} \\
&= \mathbf{a}^{\mathrm{T}}\mathbf{R}^{\mathrm{T}}\mathbf{R}\mathbf{a}
\end{aligned}$$

But

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

where $\mathbf{I}$ is the identity matrix. This property of the rotation matrix $\mathbf{R}$ (i.e., that $\mathbf{R}^{\mathrm{T}}\mathbf{R} = \mathbf{I}$) is usually called orthonormality. Using this, we find

$$|\mathbf{a}'|^2 = \mathbf{a}^{\mathrm{T}}\mathbf{I}\mathbf{a} = \mathbf{a}^{\mathrm{T}}\mathbf{a} = |\mathbf{a}|^2;$$

in other words, the invariance of the length of the vector arises because the transformation from the xy to the $x'y'$ coordinate system involves an orthonormal matrix $\mathbf{R}$.

Similarly, we expect the dot product of two vectors to be invariant: neither the magnitude of the vectors nor the angle between them should depend on the coordinate system. This is also easy to show in the same way as above. In matrix notation,

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{a}^{\mathrm{T}}\mathbf{b},$$

treating $\mathbf{a}$ and $\mathbf{b}$ as column vectors. If $\mathbf{a}' = \mathrm{R}\mathbf{a}$ and $\mathbf{b}' = \mathrm{R}\mathbf{b}$, then

$$\mathbf{a}'^{\mathrm{T}}\mathbf{b}' = (\mathbf{Ra})^{\mathrm{T}}\,\mathbf{Rb} = \mathbf{a}^{\mathrm{T}}\mathbf{R}^{\mathrm{T}}\mathbf{Rb} = \mathbf{a}^{\mathrm{T}}\mathbf{b}$$

as required.

The idea of some quantities behaving in a particular way under transformations to different coordinate systems and others remaining unchanged can be taken a lot further, but the example above hopefully gives you an example for what is involved.