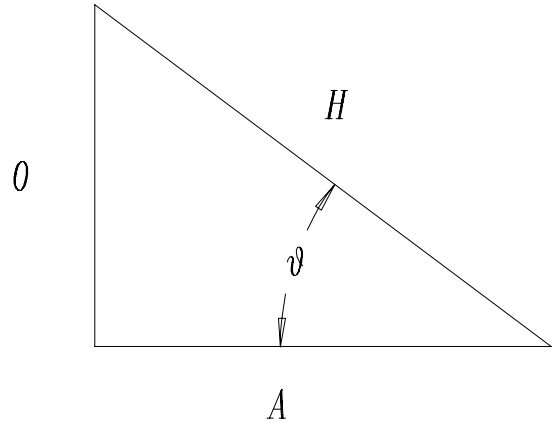
Mathematical review

This chapter is intended as a brief review of key ideas and techniques of integral and differential calculus, vector analysis in Cartesian 3-dimensional space, and the mathematical theory of probability.

1. Review of Trigonometry

The parts of trigonometry we need are based on the Law of Pythagoras and algebra. Consider a right triangle, as shown below.



The Pythagorean Theorem states "the square of the hypotenuse of a right triangle equals the sum of the squares of the sides". In symbols,

$$H^2 = A^2 + O^2$$

This geometrical fact leads to the famous trigonometric identities. We define the following functions of the angle θ shown in the Figure:

$$\sin\theta = \frac{O}{H} \qquad \cos\theta = \frac{A}{H}$$

$$\tan\theta = \frac{O}{A} \qquad \cot\theta = \frac{A}{O}$$

These trigonometric functions are **pure numbers**, because they are the ratio of two lengths. That is, they have no units associated with them. If we divide the Pythagorean Theorem on both sides by H^2, we find

$$\frac{H^2}{H^2} = 1 = \frac{A^2}{H^2} + \frac{O^2}{H^2}$$

$$= \left(\frac{A}{H}\right)^2 + \left(\frac{O}{H}\right)^2 = \cos^2\theta + \sin^2\theta$$

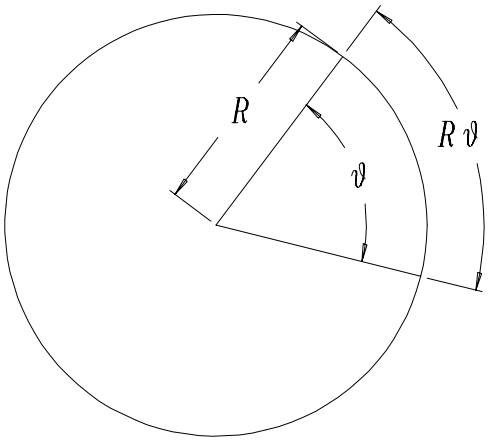
That is, for any angle,

$$\cos^2\theta + \sin^2\theta = 1 .$$

Next we see that, by definition,

$$\cot\theta = \frac{1}{\tan\theta} .$$

Next we recall the different ways of writing angles. It is common to divide a complete circle into 360 degrees. Each degree is further subdivided into 60 *minutes of arc*, and each minute into 60 *seconds of arc*. That is, a degree has 3600 seconds of arc. The other common way of writing angles is based on circles. If a circle has radius *R*, then the arc subtended between two radial lines, as shown below,



has length $R\theta$ by definition. So a complete circle corresponds to an angle of 2π *radians*.

Trigonometry also gives us formulas for adding angles:

$$\sin(\theta+\varphi) \equiv \sin\theta\cos\varphi + \cos\theta\sin\varphi$$

$$\cos(\theta+\varphi) \equiv \cos\theta\cos\varphi - \sin\theta\sin\varphi$$

From these we can derive all the standard formulas—for example, let $\varphi = \theta$: then

$$\sin 2\theta \equiv 2\sin\theta\cos\theta$$

$$\cos 2\theta \equiv \cos^2\theta - \sin^2\theta$$

$$= 2\cos^2\theta - 1$$

$$= 1 - 2\sin^2\theta\ .$$

The latter two relations lead to *half-angle* formulas. Let $2\theta = \varphi$; then $\theta = \frac{\varphi}{2}$ and we find:

$$\cos\left(\frac{\varphi}{2}\right) = \pm\left(\frac{1+\cos\varphi}{2}\right)^{1/2}$$

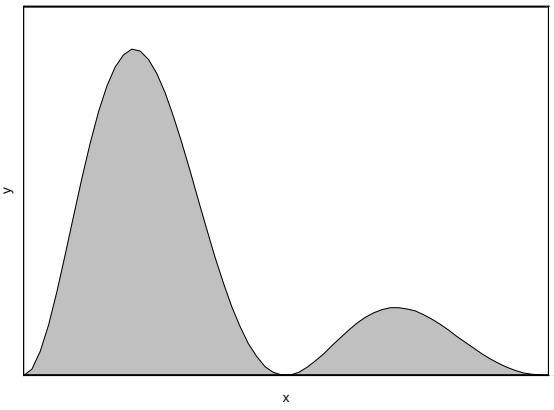
$$\sin\left(\frac{\varphi}{2}\right) = \pm\left(\frac{1-\cos\varphi}{2}\right)^{1/2}\ .$$

2. Areas and Volumes

Shape	Formula
circle	$Area = \pi R^2$
sphere	$Area = 4\pi R^2$
sphere	$Vol = \frac{4\pi}{3} R^2$
pyramid (or cone)	$Vol = \frac{1}{3} A \times H$, where A=base area, H=height

3. Integration

The definite integral $\int_a^b f(x)\,dx$ is the area between the graph of the function and the x-axis as shown below:



We estimate the integral by breaking up the area into narrow rectangles of width w that approximate the height of the curve at that point and then adding the areas of the rectangles[1]. For rectangles of non-zero width the method gives an approximation. If we calculate with rectangles that consistently protrude above the curve (assume for simplicity the curve lies above the x-axis), and with rectangles that consistently lie below the curve, we capture the exact area between two approximations. We say that we have *bounded* the integral above and below. In mathematical language,

$$w \sum_{n=0}^{(b-a)/w} \min\Big[f(a+nw), f(a+nw+w)\Big]$$

$$\le \int_a^b f(x)\,dx$$

$$\le w \sum_{n=0}^{(b-a)/w} \max\Big[f(a+nw), f(a+nw+w)\Big]\ .$$

1. If a rectangle lies below the horizontal axis, its area is considered to be negative.

It is easy to see that each rectangle in the upper bound is about $w|f'(x)|$ too high[2] on average, hence overestimates the area by about $\frac{1}{2}w^2|f'(x)|$. There are $(b–a)/w$ such rectangles, so if $|f'(x)|$ remains finite over the interval $[a, b]$ the total discrepancy will be smaller than

$$\frac{1}{2}w\,(b-a)\,\max_{a\le x\le b}\,|f'(x)|\ .$$

Similarly, the lower bound will be low by about the same amount. This means that if we halve w (by taking twice as many points), the accuracy of the approximation will double. The mathematical definition of $\int_a^b f(x)\,dx$ is the number we get by taking the limit as the width w of the rectangles becomes arbitrarily small. We know that such a limit exists because the actual area has been captured between lower and upper bounds that shrink together as we take more points.

As a historical/cultural note, the notation for an integral involves the integral sign $\int$ itself and a factor dx. The dx reminds us first, which variable we are integrating with respect to; and second, that we are summing up rectangles of height $f(x)$ and width Δx. The integral sign $\int$ is actually an elongated letter S, standing for "sum", to remind us that integration consists of adding the areas of a large number of rectangles.

4. Differentiation

The derivative of a function is defined by a limiting process. Assuming it exists, we call the derivative of a function the limit

$$f'(x) \overset{df}{=} \lim_{\Delta x\,\to\,0}\left[\frac{f(x+\Delta x)\ -\ f(x)}{\Delta x}\right].$$

The notation $f'(x)$ is Isaac Newton's[3]; Leibniz, the co-inventor of calculus, used the notation $\frac{df}{dx}$ to remind us of the underlying limiting process.

It is easy to use the definition of a derivative to derive rules for differentiating certain kinds of functions—for example the *product rule*

$$\frac{d}{dx}\Big[f(x)\,g(x)\Big] = f(x)\,\frac{dg(x)}{dx} + \frac{df(x)}{dx}\,g(x)\ ,$$

and the *chain rule*

$$\frac{d}{dx}f\Big(g(x)\Big) = f'\Big(g(x)\Big)\,g'(x)\ .$$

Here are some common derivatives that the student should know without having to look up

Table of first derivatives

function	derivative
x^α	$\alpha\,x^{\alpha-1}$
$\sin x$	$\cos x$
$\cos x$	$-\sin x$
e^x	e^x
a^x	$a^x \log a$
$\log x$	x^{-1}

2. $f'(x)$ is the slope of the line tangent to the curve at the point x. It is called the ***first derivative*** of $f(x)$. See the definition in §4 below.
3. Newton is one of two mathematicians credited with the invention of the calculus—the other is Gottfried Leibniz. Since the calculus was invented (independently) by them in the late 17th Century, it is today more than 300 years old and therefore sufficiently venerable to count as a *classic* discipline, like Latin, Greek or Western literature.

Table of first derivatives

function	derivative
$\tan x$	$\sec^2 x$
$\sin^{-1} x$	$\left(1-x^2\right)^{-1/2}$
$\cos^{-1}(x)$	$-\left(1-x^2\right)^{-1/2}$
$\tan^{-1}(x)$	$\left(1+x^2\right)^{-1/2}$

5. The fundamental theorem of calculus

Suppose we think of $\int_a^b f(x)\,dx$ as a function—call it $F(b)$—of the upper limit, b. What would happen if we compared the area $F(b)$ with the area $F(b+\Delta b)$? We see that the difference between the two is (for small Δb) is

$$\Delta F(b) = F(b+\Delta b) - F(b)$$
$$\approx f(b)\,\Delta b + O\left((\Delta b)^2\right).$$

so that

$$\frac{dF(b)}{db} = \lim_{\Delta b\to 0}\frac{1}{\Delta b}\left(\int_a^{b+\Delta b} dx - \int_a^b f(x)\,dx\right)$$
$$\to f(b)\,.$$

This is a fancy way to say that integration and differentiation are *inverse operations*, in the same sense as multiplication and division, or addition and subtraction. It is usual to write the primitive integral, or *primitive* of a function as

$$F(x) \overset{df}{=} \int^x f(x')\,dx'\,.$$

(Here we have written the variable of integration as x' to distinguish it from the upper limit—there is a good deal of sloppyness about this in mathematical literature, the upper limit and variable of integration often being denoted by the same letter.)

6. Functions of several variables

In physical applications we must often deal with quantities that depend on several independent variables. For example, the temperature can differ from one place to the next, as well as varying in time[4].

Consider a function $q(x, t)$ depending on both position x (in one space dimension, for simplicity) and time t. We might be interested in how it varies if—say—time increases by a small amount while we hold the position constant; or conversely we might want to know how it varies from one position to an adjacent one at a given moment of time. This introduces the notion of *partial derivatives*, defined by

$$\frac{\partial q}{\partial t} \overset{df}{=} \lim_{\Delta t\to 0}\left[\frac{q(x,\,t+\Delta t) - q(x,\,t)}{\Delta t}\right]$$

and

$$\frac{\partial q}{\partial x} \overset{df}{=} \lim_{\Delta x\to 0}\left[\frac{q(x+\Delta x,\,t) - q(x,\,t)}{\Delta x}\right].$$

That is, partial derivatives are computed just like ordinary derivatives except we pretend the other variable (that we are not differentiating with respect to) is just a constant; the curly d (∂) reminds us of this. An example might be

$$q(x,\,t) = 37.2x^{\alpha}\cos t$$

for which the partial derivatives are

4. Anyone who has ever added hot (or cold!) water while taking a bath can verify this from personal experience.

$$\frac{\partial q}{\partial t} = -37.2\, x^{\alpha} \sin t$$

$$\frac{\partial q}{\partial x} = 37.2\, \alpha\, x^{\alpha - 1} \cos t\,.$$

A somewhat more complex example is

$$q(x, t) = e^{-3x^2/t}$$

for which

$$\frac{\partial q}{\partial t} = \frac{3x^2}{t^2}\, e^{-3x^2/t}$$

$$\frac{\partial q}{\partial x} = \frac{-6x}{t}\, e^{-3x^2/t}\,.$$

A question that arises frequently when considering functions of several variables is "differentiating under the integral sign". That is, suppose we define a function

$$Q(t) = \int_a^b q(x, t)\, dx\,.$$

Is it legitimate to compute the derivative with respect to t by interchanging the order of differentiation and integration? Or, put another way, can we naively write

$$\frac{d}{dt} Q(t) = \lim_{\Delta t \to 0} \left[\frac{Q(t + \Delta t) - Q(t)}{\Delta t} \right]$$

$$= \int_a^b \lim_{\Delta t \to 0} \left[\frac{q(x, t + \Delta t) - q(x, t)}{\Delta t} \right] dx$$

$$= \int_a^b \frac{\partial q(x, t)}{\partial t}\, dx$$

with any possibility of getting the same result both ways? The answer to this rhetorical question is "Yes, if the function $q(x, t)$ is sufficiently well behaved."

Mathematicians derive unholy glee from finding horrible functions that serve as counter-examples to any given proposition. Because differentiating under an integral sign requires interchanging the order of performing two limiting processes, it is virtually guaranteed that there exist some functions for which such interchange will yield false results. However we do not permit such functions in the branch of mathematics used to describe physical processes. That is, for the purposes of this course we may assume that any function we shall be interested in has the necessary properties to guarantee that any operation involving the interchange of limiting processes will be independent of the order.

Another case of this sort involves higher partial derivatives. It is clear what we mean by

$$\frac{\partial^2 q(x, t)}{\partial t^2} \text{ or } \frac{\partial^2 q(x, t)}{\partial x^2}\,.$$

But what does the *mixed* partial derivative

$$\frac{\partial^2 q(x, t)}{\partial t\, \partial x}$$

mean, and is it the same as the mixed partial derivative

$$\frac{\partial^2 q(x, t)}{\partial x\, \partial t}\,?$$

The process needed to get the first mixed derivative is simply to calculate the function

$$p(x, t) = \frac{\partial q(x, t)}{\partial x}$$

and then to calculate $\dfrac{\partial p}{\partial t}$. To get the second we differentiate in the opposite order. Although there certainly exist functions for which the orders cannot be interchanged, any function we shall encounter in physical applications will permit interchanging the order.

7. Ordinary differential equations

We are sometimes confronted by equations of the form

$$\frac{dx}{dt} = f(x, t) .$$

Such an equation is called an ordinary differential equation (ODE) of first order[5].

The simplest such equation has the form

$$\frac{dx}{dt} = f(t) ;$$

where $f(t)$ is a known function of t. That is, the unknown function does not appear on the right side. Such equations can be solved by integrating both sides with respect to the independent variable, t:

$$x(t) - x(a) = \int_a^t \frac{dx}{ds}\, ds = \int_a^t f(s)\, ds .$$

The equation

$$\frac{dx}{dt} = f(x, t)$$

can sometimes be solved with an *integrating factor*: if we can find a function $g(x, t)$ which, when it multiplies both sides

$$g(x, t)\frac{dx}{dt} = g(x, t)\, f(x, t)$$

converts both sides to perfect derivatives (of other functions) then we have reduced the problem to one of performing some integrals.

The *linear* ODE of first order can always be solved by an integrating factor. Writing it in the form

$$\frac{dx}{dt} + x\, f(t) = g(t)$$

where $f(t)$ and $g(t)$ are known functions, we see that the factor

$$F(t) = \exp\left[\int^t f(s)\, ds\right]$$

turns the left side into a perfect derivative and the right side into a known function:

$$F(t)\left(\frac{dx}{dt} + x\, f(t)\right) \equiv \frac{d}{dt}\Big(x\, F(t)\Big) = F(t)\, g(t) .$$

The solution therefore has the form

$$x(t)\, F(t) = \text{constant} + \int^t F(s)\, g(s)\, ds .$$

Since $F(t)$ is never zero we can always solve the above for $x(t)$, adjusting the constant of integration to match an initial value.

Finally, some equations are *separable*: they have the form

$$\frac{dx}{dt} = A(x)\, B(t)$$

where A and B are known functions, so we can rewrite them as

$$\frac{dx}{A(x)} = B(t)\, dt$$

and integrate both sides (with respect to their respective variables). Here is an example that arises in the description of an object with air resistance moving vertically in a gravitational field:

$$v\frac{dv}{dy} = -g \pm \Gamma v^2 .$$

The sign depends on whether the object is rising (–) or falling (+). Hence

$$\frac{v\, dv}{g \mp \Gamma v^2} = -dy$$

5. What makes it "first order" is the fact that the highest derivative of the unknown function that appears in it is the first derivative.

or

$$y = y_0 - \frac{1}{2}\int_{v_0^2}^{v^2} \frac{du}{g \mp \Gamma u}$$

where we have changed the variable of integration to $u = v^2$, and where y_0 is a constant of the motion with an obvious meaning. The remaining integral will be recognized from the table of derivatives as the function

$$\int^{u} \frac{du}{g \mp \Gamma u} = \frac{1}{\Gamma}\ln\left(g \mp \Gamma u\right),$$

so that the full solution becomes

$$y = y_0 - \frac{1}{2\Gamma}\ln\left(\frac{g \mp \Gamma v^2}{g \mp \Gamma v_0^2}\right).$$

We now look at some simple second-order differential equations. The first thing to know is that in general they are very hard. For example, the equation[6]

$$\ddot{x} + p(t)\,\dot{x} + q(t)\,x = f(t)$$

is *linear* (the function $x(t)$ and its derivatives appear only to first degree) but its solutions can be complicated functions. A considerable literature—deriving mainly from the 19th and early 20th Centuries—discusses specific examples of this equation that arise in physics problems. This literature is summarized in many good books on mathematical methods of physics[7].

Here we shall only consider three types of second order ODEs. The first has constant coefficients (although the driving function $f(t)$, sometimes known as the inhomogeneous term, need not be a constant). Suppose it is zero, and p and q are constants. Then the function

$$x(t) = A\,e^{\lambda t}$$

solves the equation, as long as λ is a root of the *secular equation*

$$\lambda^2 + p\lambda + q = 0\,.$$

For example, the equation for the viscous-damped harmonic oscillator has the form[8]

$$\ddot{x} + \gamma\dot{x} + \omega^2 x = 0\,;$$

its secular equation is

$$\lambda^2 + \gamma\lambda + \omega^2 = 0$$

with solutions (here $i^2 = -1$)

$$\lambda = -\frac{\gamma}{2} \pm i\Omega = -\frac{\gamma}{2} \pm i\left(\omega^2 - \frac{\gamma^2}{4}\right)^{1/2}.$$

This is less horrific than it looks—it just means the solutions have the form

$$x(t) = \left[A\cos\Omega t + B\sin\Omega t\right] e^{-\gamma t/2}$$

where A and B must be adjusted to match the initial conditions.

The second type of second-order equation has the form

$$\ddot{x} = f(x)\,.$$

This equation can always be integrated once by means of the integrating factor $\dot{x}$:

$$\dot{x}\,\ddot{x} = \dot{x}\,f(x)\,,$$

or

6. Here we have adopted the "dot" notation of Sir Isaac Newton. A single dot over a variable denotes its first derivative, two dots the second derivative, *etc*.
7. For example, J. Mathews and R.L. Walker, *Mathematical Methods of Physics* (W.A. Benjamin, Inc., New York, 19xx).
8. We shall be using this equation a lot to describe the motion of limbs, Brownian motion of charged particles in a magnetic field, or the behavior of sensory cilia in a fluid.

$$\frac{d}{dt}\left(\frac{\dot{x}^2}{2}\right) = \frac{d}{dt}\int^{x(t)} f(s)\,ds\,.$$

Integrating both sides we have

$$\frac{\dot{x}^2}{2} = \int^{x(t)} f(s)\,ds + \text{constant}.$$

This equation is separable, but the result may not be integrable in closed form. For example, if we eschew the usual linear approximation for the simple pendulum we obtain

$$\ddot{\theta} + \omega^2 \sin\theta = 0$$

giving a first ntegral

$$\frac{1}{2}\dot{\theta}^2 + \omega^2 \cos\theta = E\,;$$

we can solve for $\dot{\theta}(t)$ and separate the resulting equation:

$$\frac{d\theta}{dt} = \pm\left[2E - 2\omega^2\cos\theta\right]^{1/2}$$

or

$$\int^{\theta} du\left[E - \omega^2\cos u\right]^{-1/2} = \pm\, t\sqrt{2}\,.$$

The integral defines something called an elliptic function, which is rather more complicated than we need to deal with here.

Finally, suppose the second order equation is really a first order one in disguise. For example our old friend, the object falling freely, with air resistance,

$$\ddot{y} = -g \mp \Gamma\dot{y}^2$$

has this form if we change variables. Let

$$v = \dot{y}\,;$$

then

$$\frac{dv}{dt} = -g \mp \Gamma v^2$$

which is separable and can be integrated once. However, the integral involves the arctangent function, which can be rather ugly to integrate further. Hence we employed a trick: the chain rule of differentiation lets us write

$$\frac{dv}{dt} \equiv \frac{dv}{dy}\cdot\frac{dy}{dt} \equiv v\frac{dv}{dy}$$

which is how we eliminated the time variable in favor of the (vertical) distance traveled.

8. Vectors

We now consider vectors in 3-dimensional Cartesian coordinate spaces. Normally we choose three orthogonal directions and label the coordinate axes x, y, z . The common labelling is *right-handed* as shown below (a left-handed coordinate system would be one in which the positive and negative directions along any one axis were interchanged).

A vector is an ordered set of three numbers representing a point in the coordinate system in terms of distances from the origin taken along the axes. Thus the vector

$$\vec{r} \overset{df}{=} (x, y, z) \equiv x\,\hat{\mathbf{x}} + y\,\hat{\mathbf{y}} + z\,\hat{\mathbf{z}}$$

represents a point a distance x along the x-axis, a distance y along the y-axis, *etc.*

We can multiply a vector by a constant:

$$\lambda\vec{r} \overset{df}{=} (\lambda x, \lambda y, \lambda z)\,;$$

thus the preceding definition introduces an alternative notation, in which the vector is represented as the sum of three vectors pointing along the three axes. The special vector $\hat{\mathbf{x}}$ is a unit vector, that is, one whose length is 1; and similarly for $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$.

We can add two vectors: if

$$\vec{r} \equiv x\,\hat{\mathbf{x}} + y\,\hat{\mathbf{y}} + z\,\hat{\mathbf{z}}$$

and

$$\vec{s} \equiv u\,\hat{\mathbf{x}} + v\,\hat{\mathbf{y}} + w\,\hat{\mathbf{z}},$$

then

$$\vec{r} + \vec{s} \equiv (x+u)\,\hat{\mathbf{x}} + (y+v)\,\hat{\mathbf{y}} + (z+w)\,\hat{\mathbf{z}}\ .$$

We can combine two vectors to make a scalar:

$$\vec{r}\cdot\vec{s} \equiv xu + yv + zw\,.$$

This is called the *scalar product*, or "dot product" of two vectors. Its geometrical meaning is the cosine of the angle between the two vectors multiplied by the product of their lengths:

$$\vec{r}\cdot\vec{s} = |\vec{r}|\ |\vec{s}|\ \cos\theta\,;$$

from this it is clear that the dot product of a vector with itself is just the square of its length:

$$\vec{r}\cdot\vec{r} = |\vec{r}|^2 \equiv x^2 + y^2 + z^2\,.$$

Finally, in three dimensions it is possible to define a vector multiplication of two vectors:

$$\vec{r}\times\vec{s} \equiv (yw - zv)\,\hat{\mathbf{x}} + (zu - xw)\,\hat{\mathbf{y}} + (xv - yu)\,\hat{\mathbf{z}}\,.$$

We can remember the precise definition of the *vector product* (sometimes called the "cross-product") using either of two mnemonic devices: first, we can write the vector product as a determinant which can be evaluated using Cramer's rule:

$$\vec{r}\times\vec{s} = \det\begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ x & y & z \\ u & v & w \end{vmatrix} = -\vec{s}\times\vec{r}\,.$$

Alternatively, note that the unit vectors have the vector products

$$\hat{\mathbf{x}}\times\hat{\mathbf{y}} \equiv \hat{\mathbf{z}}$$

$$\hat{\mathbf{y}}\times\hat{\mathbf{z}} \equiv \hat{\mathbf{x}}$$

$$\hat{\mathbf{z}}\times\hat{\mathbf{x}} \equiv \hat{\mathbf{y}}$$

(also note that the vector product of any vector with itself is zero because of the antisymmetry).

As we can see from the rules for unit vectors, the cross product of two vectors is a vector that is perpendicular to the plane formed by the two vectors being multiplied. The length of the product vector is given by

$$|\vec{r}\times\vec{s}\,| = |\vec{r}|\ |\vec{s}|\ \sin\theta$$

where θ is the angle between them.

The vector product is needed to express compactly the force on a moving charged particle in an external magnetic field (the Lorentz force):

$$\vec{F} = \frac{Q}{c}\,\vec{v}\times\vec{B}\,,$$

and thence the force on a current-carrying wire in a magnetic field (the principle of the electric motor).

9. Vector analysis

The last topic in this primer summarizes certain operations involving partial differentiation and multiple integration, that arise frequently in physical applications.

Suppose we want the difference between a function of position at a point $\vec{r} = (x, y, z)$ and a neighboring point, which we shall call

$$\vec{r} + d\vec{r} = (x + dx, y + dy, z + dz)\,.$$

That is, we want to calculate

$$d\Phi \overset{df}{=} \Phi(\vec{r} + d\vec{r}) - \Phi(\vec{r})$$

$$= \Phi(x + dx, y + dy, z + dz) - \Phi(x, y, z)\,.$$

Since each of the small differences dx, dy, dz is independent of the other two, we see that to first order in the differences,

$$d\Phi = \frac{\partial\Phi}{\partial x}\,dx + \frac{\partial\Phi}{\partial y}\,dy + \frac{\partial\Phi}{\partial z}\,dz\,.$$

This looks like a scalar product of the two vectors

$$d\vec{r} = (dx, dy, dz) = dx\,\hat{\mathbf{x}} + dy\,\hat{\mathbf{y}} + dz\,\hat{\mathbf{z}}$$

and

$$\nabla\Phi = \frac{\partial\Phi}{\partial x}\hat{\mathbf{x}} + \frac{\partial\Phi}{\partial y}\hat{\mathbf{y}} + \frac{\partial\Phi}{\partial z}\hat{z} .$$

The latter is called the *gradient* of Φ.

Thus we write the change in Φ compactly as

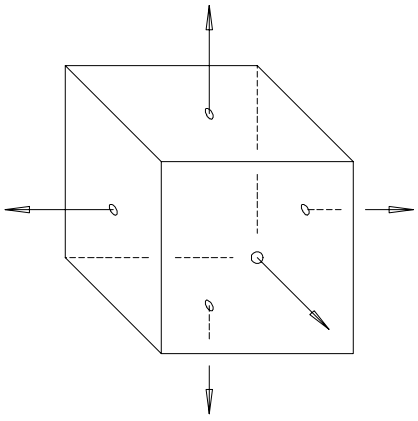
$$d\Phi = \nabla\Phi \cdot d\vec{r} .$$

Another type of partial derivative that arises frequently is the *divergence* of a vector. That is, suppose we have a vector, each of whose components is a function of position. (Such a vector is called a *vector field*.) An example might be an electric field or a magnetic field—either can be expected to vary from point to point both in magnitude and direction. A vector field that is particularly interesting in this course is the *flux* of a moving fluid. If the fluid has a number-density $n(x, y, z, t)$ and a local velocity vector $\vec{v}$ its flux vector (or current density) is defined to be

$$\vec{j} \overset{df}{=} n\,\vec{v} .$$

This has the physical meaning of the number of particles flowing across a unit area perpendicular to the flux vector, per unit time.

Suppose we calculate the net outflow from an infinitesimal volume such as that shown below. By convention we erect unit vectors



$$\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_6$$

perpendicular to each face of the parallelopiped, pointing outward from their respective faces; the net outflow is then

$$-\frac{dN}{dt} = \sum_{k=1}^{6} \vec{j}(k) \cdot \hat{\mathbf{u}}_k\, dS_k$$

where the area of the k'th face is dS_k and $\vec{j}(k)$ is the value of $\vec{j}$ at that face. If we work out the details we find that the net outflow is proportional to something called the *divergence* of $\vec{j}$:

$$\operatorname{div}\vec{j} \equiv \nabla\cdot\vec{j} \overset{df}{=} \frac{\partial}{\partial x}j_x + \frac{\partial}{\partial y}j_y + \frac{\partial}{\partial z}j_z .$$

The reason it is often written in the form $\nabla\cdot\vec{j}$ is mnemonic: if we think of the symbol ∇ as a vector in its own right,

$$\nabla = \frac{\partial}{\partial x}\hat{\mathbf{x}} + \frac{\partial}{\partial y}\hat{\mathbf{y}} + \frac{\partial}{\partial z}\hat{z} ,$$

and forget for the moment that its "components" are operations rather than numbers, then the "dot product" of that vector with a vector field is just the sum of the "products" of the corresponding components of the two vectors. At this point we recall that the components of ∇ are operators of differentiation, so when they "multiply" the corresponding components of the vector field, we must understand this as an instruction to carry out the differentiation.

An operation that arises so often it is given its own name is to take the divergence of a vector field that is itself the gradient of a scalar function: for example, in electrostatics we write

$$\nabla\cdot\vec{E} = 4\pi\rho \quad \text{(one of Maxwell's equations)}$$

where

$$\vec{E} = -\nabla\Phi .$$

Combining these we have the Poisson equation (also called the inhomogeneous Laplace equation)

$$\text{div}\,(\nabla\Phi) \equiv \nabla\cdot(\nabla\Phi) = -4\pi\rho\,.$$

Working out the details we see that

$$\nabla\cdot(\nabla\Phi) = \frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2} + \frac{\partial^2\Phi}{\partial z^2}$$

so we define the *Laplacian operator*,

$$\nabla^2 \equiv \nabla\cdot\nabla \overset{df}{=} \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\,.$$

It can be enormously convenient to treat mathematical operations as abstract entities that can be separated from the things they are operating on and added, multiplied and so forth according to appropriate rules. This approach was introduced (along with the vector notation) by the brilliantly eccentric English physicist Oliver Heaviside (1850-1925). It is now so universally used that few recall the origin of these ideas.

To see how valuable the operational calculus can be, we now derive Green's theorem[9] in just a couple of lines. In solving certain partial differential equations arising in electromagnetic theory, acoustics or quantum mechanics, we construct a function of the form

$$G = \Phi\nabla^2\Psi - \Psi\nabla^2\Phi\,.$$

Now it is easy to show (basically by the product rule of differentiation) that

$$G \equiv \nabla\cdot\Big(\Phi\nabla\Psi - \Psi\nabla\Phi\Big).$$

Thus if we integrate G over some volume, then by Gauss's theorem we can convert the integral to one over a surface surrounding that volume. In other words,

$$\iiint_V G\,dV = \iiint_V \nabla\cdot\Big(\Phi\nabla\Psi - \Psi\nabla\Phi\Big)dV$$

$$= \iint_S \Big(\Phi\nabla\Psi - \Psi\nabla\Phi\Big)\cdot d\vec{S}\,.$$

Using Green's theorem it is possible to specify the solutions of certain partial differential equations, throughout the volume of interest, knowing only the values of the function (and of another, called the Green's function) on S, and the components of their gradients normal to S. Showing how this is done would, however, take us too far afield.

Before leaving the subject of vector analysis, it is worth stressing certain identities. First,

$$\nabla\times(\nabla\Phi) = \textbf{curl}\,(\textbf{grad}\,\Phi) \equiv 0$$

$$\nabla\cdot(\nabla\times\vec{A}) = \textbf{div}\cdot(\textbf{curl}\,\vec{A}) \equiv 0\,.$$

And we sometimes need the identity

$$\nabla\times(\nabla\times\vec{A}) = \textbf{curl}\,(\textbf{curl}\,\vec{A})$$

$$\equiv \nabla(\nabla\cdot\vec{A}) - \nabla^2\vec{A}\,.$$

10. Basic ideas of probability theory

Probability theory is so fundamental in physics it is all but taken for granted. The basic notion is a numerical representation of the likelihood of a chance event, which we call a probability. For concreteness let us imagine a cubical die (that is, one of a pair of dice) whose six faces are marked with 1–6 spots. If we throw the die there are six possible distinct outcomes: it can land on one of its faces, with the face opposite up. Assuming, by

9. This was a major piece of work in its time and required pages of tedious calculation to prove, working with individual vector components.

symmetry, that there is no reason for one outcome to be more likely than another, we assign *a priori* probabilities $\frac{1}{6}$ to each outcome.

We should draw attention, despite their obviousness, to two aspects of this assignment: first, a probability is a number p in the range $0 \le p \le 1$. Second, the events we are describing must be distinct, or mutually exclusive: if 3 dots are showing, then 5 dots cannot be simultaneously showing.

Let us consider more complicated situations: suppose A and B are two possible events, and suppose we perform an experiment with n equally likely outcomes:

$n_1 =$ outcomes with A but not B
$n_2 =$ outcomes with B but not A
$n_3 =$ outcomes with both A and B
$n_4 =$ outcomes with neither A nor B

Since this exhausts all the possibilities we have

$$n_1 + n_2 + n_3 + n_4 = n\,.$$

The probability of A is

$$p(A) = \frac{n_1 + n_3}{n}$$

and that of B is

$$p(B) = \frac{n_2 + n_3}{n}\,.$$

But there are more complex possibilities. For example, the probability of A or B (or both—this is "or" in the logical sense) is

$$p(A + B) = \frac{n_1 + n_2 + n_3}{n}$$

and the probability of both occurring (the *joint* probability) is

$$p(AB) = \frac{n_3}{n}\,.$$

Last we define *conditional* probabilities: for example, the probability that A occurs, given that we know B has occured, is

$$p(A\,|\,B) = \frac{n_3}{n_2 + n_3}$$

(that is, in the set of all outcomes with B occurring, what fraction have A occurring also). Similarly,

$$p(B\,|\,A) = \frac{n_3}{n_1 + n_3}\,.$$

From these examples we can deduce some general rules:

$$p(AB) = p(B)\,p(A\,|\,B) \equiv p(A)\,p(B\,|\,A)$$

and

$$p(A + B) = p(A) + p(B) - p(AB)\,.$$

An example of the latter is the probability that if you draw two cards from each of two well-shuffled decks, at least one will be a king. The probability of drawing a king is assumed to be the same as that of drawing any other card, namely $\frac{1}{13}$, and since the events are independent (that is, $p(B\,|\,A) \equiv p(B)$) we have

$$p(\text{ at least 1 king }) = \frac{1}{13} + \frac{1}{13} - \frac{1}{169}\,.$$

It is worth noting that if two events are mutually exclusive,

$$p(AB) = 0\,.$$

Suppose we have a third event, C. Since

$$p(B)\,p(A\,|\,B) = p(A)\,p(B\,|\,A)$$

it must be true that

$$p(B\,|\,A) = \frac{p(B)}{p(A)}\,p(A\,|\,B)$$

and for event C,

$$p(C\,|\,A) = \frac{p(C)}{p(A)}\,p(A\,|\,C)\,.$$

If we divide the first relation by the second, $p(A)$ cancels out and we have *Bayes's theorem*

$$\frac{p(B|A)}{p(C|A)} = \frac{p(B)}{p(C)}\,\frac{p(A|B)}{p(A|C)}\,.$$

Here is an example of Bayes's theorem: suppose we have drawers A, B and C containing two gold, one gold and one silver, and two silver coins, respectively. If the drawers are unlabelled and we take a coin from one (without looking inside), what is the probability that we the other coin is gold? That is we want to know the probability that, having found a gold coin (call this event D), the drawer we chose at random is drawer A. In symbols we want to calculate $p(A|D)$. Now we know that $p(C|D) = 0$ (because then we would definitely have pisked a silver coin). We also know that since we had to have picked one of the drawers,

$$p(A|D) + p(B|D) + p(C|D) = 1\,.$$

Bayes's theorem tells us that

$$\frac{p(A|D)}{p(B|D)} = \frac{p(A)}{p(B)}\,\frac{p(D|A)}{p(D|B)} = \frac{⅓}{⅓}\,\frac{1}{½} = 2\,;$$

hence we conclude

$$p(A|D) = ⅔\,, \qquad p(B|D) = ⅓\,.$$

This problem, incidentally, is equivalent to one that the columnist Marilyn Vos Savant (the person with the world's highest IQ) got right and many mathematics professors got wrong.

Bayes's theorem can give insight into the correctness and power of a theory, given certain observations.

Thus consider the probability $p(M|GR)$ that General Relativity is correct, given that it predicts the (observed) residual perihelion precession of Mercury's orbit. General Relativity gives a unique answer for this value, so the probability of observing it would in theory be a δ-function. However, observational error smears the probability distribution into a sharply peaked curve centered at the optimum value. What can we compare the theory to? Suppose we consider alternate physics based on some non-uniform mass distribution (that has not yet been observed) that leads to a term proportional to $\frac{1}{r^3}$ in the gravitational potential of the Sun. The coefficient of this extra term is not well-determined since the only restriction is that it must be too small to influence the orbits of the other planets (since these are not seen to precess). Thus we must assign a much broader probability distribution to the alternate theory, $p(M|A)$. Bayes's theorem says that

$$\frac{P(GR|M)}{p(A|M)} = \frac{p(GR)}{p(A)}\,\frac{p(M|GR)}{p(M|A)}\,;$$

we do not know the a priori probabilities, $p(GR)$, $p(A)$ that either theory is correct. But it is reasonable to assume in the absence of evidence to the contrary that they are roughly equal. In this case, the experimental result combined with the sharp prediction of General Relativity favors Einstein's theory over plausible alternatives by something like 40 to 1.

Now we discuss some well-known probability distributions. Among discrete distributions the most common are the binary and Poisson distributions. They can be derived by writing down equations that describe the processes, then solving the equations.

Consider an experiment involving two outcomes—for example heads or tails, with probabilities p and q such that $p + q = 1$. Define the probability to have n heads after a series of N tosses of the coin as $P_n(N)$. Now it is easy to see that

$$P_n(N) = p\,P_{n-1}(N-1) + q\,P_n(N-1)$$

or in words, the probability that after $N-1$ tosses there were $n-1$ heads, times the probability that the next toss produced heads, plus the probability that there were already n heads and one threw a tail. This is a difference equation in two indices, and difficult to solve without a trick: the idea is to multiply P_n by a variable s^n and sum over n to

produce a new function called the *generating function*

$$G_N(s) = \sum_{n=0}^{N} P_n(N)\, s^n .$$

Note that $G_N(1) = 1$.

It is easy to see that the generating function satisfies the equation

$$G_N(s) = \Big(p\, s + q\Big) G_{N-1}(s)$$

whose solution is

$$G_N(s) = G_N(1)\, \Big(p\, s + q\Big)^N \equiv \Big(p\, s + q\Big)^N$$

from which we can recover the probabilities by expanding the binomial:

$$G_N(s) = \sum_{n=0}^{N} \binom{N}{n} p^n\, q^{N-n}\, s^n$$

or

$$P_n(N) = \binom{N}{n} p^n\, q^{N-n} .$$

This is called the *binomial distribution*. Just for fun let us calculate the expected number of heads in N tosses: this is just

$$\langle n \rangle_N = \sum_{n=0}^{N} n \binom{N}{n} p^n\, q^{N-n} \equiv \left. \frac{dG_N(s)}{ds} \right|_{s=1} .$$

The last identity gives us an easy way to evaluate the sum: we get

$$\langle n \rangle_N = \left. \frac{d}{ds} \Big(p\, s + q\Big)^N \right|_{s=1} = pN .$$

We could also ask for the *variance* of this number, defined by

$$\mathrm{Var}\ (n)_N = \langle\, (n - \langle n \rangle)^2\, \rangle_N$$

$$= \sum_{n=0}^{N} (n - pN)^2 \binom{N}{n} p^n\, q^{N-n} .$$

Evaluating this sum (for example, by the previous trick) we find

$$\mathrm{Var}\ (n)_N = pqN .$$

Thus the statistical uncertainty or standard deviation (square root of the variance) is

$$\sigma = \sqrt{p\, q\, N} \ .$$

Next we consider a physical process like a telephone exchange or a Geiger counter. Events (telephone calls, radioactive decays) come randomly spaced in time, but at some average rate α so that in a time t we may expect to have counted $\alpha\, t$ such events. Now if $P_n(t)$ is the probability at time t of having observed n events, we may say that

$$P_n(t + dt) = P_{n-1}(t)\, \alpha\, dt + P_n(t) \Big(1 - \alpha\, dt\Big).$$

This leads to the differential-difference equation

$$\frac{d}{dt} P_n(t) = \alpha \Big(P_{n-1}(t) - P_n(t)\Big).$$

Defining the appropriate generating function by analogy with the preceding example, we find the differential equation

$$\frac{d}{dt} G(s, t) = \alpha\, (s - 1)\, G(s, t)$$

or

$$G(s, t) = e^{-\alpha t}\, e^{\alpha t s} \equiv \sum_{n=0}^{\infty} e^{-\alpha t} \frac{(\alpha\, t)^n}{n!}\, s^n .$$

That is, if we define $\lambda = \alpha t$ as the expected number of hits, the probability of n hits is

$$P_n = e^{-\lambda} \frac{\lambda^n}{n!} .$$

This is called the *Poisson* distribution.

It is easily seen that the probabilities sum to 1 for positive λ and that

$$\langle n \rangle = \lambda$$

$$\mathrm{Var}(n) = \lambda .$$

That is, the statistical uncertainty in a counting experiment is always

$$\sigma = \left[\langle n \rangle\right]^{1/2}.$$

Poisson distributions apply to political polling as well as to the theory of statistical fluctuations in a gas. In a gas of mean number-density $\bar{n}$ a volume ΔV contains

$$N = \bar{n}\,\Delta V \pm \left(\bar{n}\,\Delta V\right)^{1/2}$$

(that is, the average number $\pm\sigma$). If we make the volume too small, therefore, the relative size of the fluctuations becomes comparable to the average number, hence there is no meaning to the averaging process. Such effects show up in microcircuits as noise (because charge comes in electron-size units). In the days of vacuum tubes the limiting factor in amplification was the "shot effect" noise (resembling lead shot or BBs dropped on the floor) because at low currents only a few electrons could be expected to arrive in any time interval.