Vector Calculus

University of Cambridge Part IA Mathematical Tripos

David Tong

Department of Applied Mathematics and Theoretical Physics,
Centre for Mathematical Sciences,
Wilberforce Road,
Cambridge, CB3 OBA, UK

http://www.damtp.cam.ac.uk/user/tong/vc.html
d.tong@damtp.cam.ac.uk

Recommended Books and Resources

There are many good books on vector calculus that will get you up to speed on the basic ideas, illustrated with an abundance of examples.

- H.M Schey, "*Div, Grad, Curl, and all That*"
- Jerrold Marsden and Anthony Tromba, "*Vector Calculus*"

Schey develops vector calculus hand in hand with electromagnetism, using Maxwell's equations as a vehicle to build intuition for differential operators and integrals. Marsden and Tromba is a meatier book but the extra weight is because it goes slower, not further. Neither of these books cover much (if any) material that goes beyond what we do in lectures. In large part this is because the point of vector calculus is to give us tools that we can apply elsewhere and the next steps involve turning to other courses.

- Baxandall and Liebeck, "*Vector Calculus*"

This book does things differently from us, taking a more rigorous and careful path through the subject. For the most part, this involves being more careful from the off about what spaces different objects live in. All of this will be treated in later courses, but if you're someone who likes all their i's dotted, ϵ's small, and $\hbar$'s uncrossed, then this is an excellent place to look.

Contents

Acknowledgements

These lecture notes are far from novel. Large chunks of them have been copied wholesale from the excellent lecture notes of Ben Allanach and Jonathan Evans who previously taught this course. I've also benefitted from the detailed notes of Stephen Cowley. My thanks to Jonathan Evans, Julia Gog and Maria Gutierrez for helpful discussions and suggestions on the material in the course.

I am supported by a Royal Society Wolfson Merit award and by a Simons Investigator award.

0 Introduction

The development of calculus was a watershed moment in the history of mathematics. In its simplest form, we start with a function

$$f : \mathbb{R} \to \mathbb{R}$$

Provided that the function is continuous and smooth, we can do some interesting things. We can differentiate. And integrate. It's hard to overstate the importance of these operations. It's no coincidence that the discovery of calculus went hand in hand with the beginnings of modern science. It is, among other things, how we describe change.

The purpose of this course is to generalise the concepts of differentiation and integration to functions, or maps, of the form

$$f : \mathbb{R}^m \to \mathbb{R}^n \tag{0.1}$$

with m and n positive integers. Our goal is simply to understand the different ways in which we can differentiate and integrate such functions. Because points in $\mathbb{R}^m$ and $\mathbb{R}^n$ can be viewed as vectors, this subject is called *vector calculus*. It also goes by the name of *multivariable calculus*.

The motivation for extending calculus to maps of the kind (0.1) is manifold. First, given the remarkable depth and utility of ordinary calculus, it seems silly not to explore such an obvious generalisation. As we will see, the effort is not wasted. There are several beautiful mathematical theorems awaiting us, not least a number of important generalisations of the fundamental theorem of calculus to these vector spaces. These ideas provide the foundation for many subsequent developments in mathematics, most notably in geometry. They also underlie every law of physics.

Examples of Maps

To highlight some of the possible applications, here are a few examples of maps (0.1) that we will explore in greater detail as the course progresses. Of particular interest are maps

$$f : \mathbb{R} \to \mathbb{R}^n \tag{0.2}$$

These define *curves* in $\mathbb{R}^n$. A geometer might want to understand how these curves twist and turn in the higher dimensional space or, for $n = 3$, how the curve ties itself in knots. For a physicist, maps of this type are particularly important because they describe the trajectory of a particle. Here the codomain $\mathbb{R}^n$ is identified as physical space, an interpretation that is easiest to sell when $n = 3$ or, for a particle restricted to move on a plane, $n = 2$.

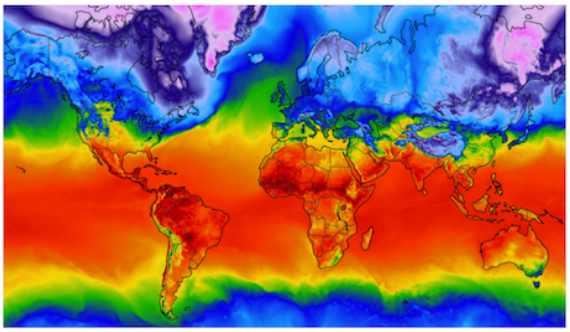
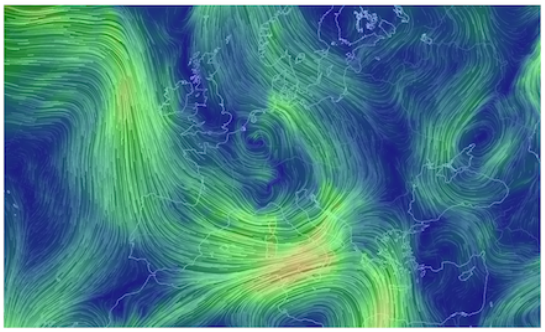



Figure 1. On the left, the temperature on the surface of the Earth is an example of a map from $\mathbb{R}^2 \to \mathbb{R}$, also known as a scalar field. On the right, the wind on the surface of the Earth blows more or less horizontally and so can be viewed as a map from $\mathbb{R}^2 \to \mathbb{R}^2$, also known as a vector field. (To avoid being co-opted by the flat Earth movement, I should mention that, strictly speaking, each of these is a map from $\mathbf{S^2}$ rather than $\mathbb{R}^2$.)

Before we go on, it will be useful to introduce some notation. We'll parameterise $\mathbb{R}$ by the variable t. Meanwhile, we denote points in $\mathbb{R}^n$ as $\mathbf{x}$. A curve (0.2) in $\mathbb{R}^n$ is then written as

$$f : t \to \mathbf{x}(t)$$

Here $\mathbf{x}(t)$ is the image of the map. But, in many situations below, we'll drop the f and just refer to $\mathbf{x}(t)$ as the map. For a physicist, the parameter t is usually viewed as time. In this case, repeated differentiation of the map with respect to t gives us first velocity, and then acceleration.

Going one step further, we could consider maps $f : \mathbb{R}^2 \to \mathbb{R}^n$ as defining a surface in $\mathbb{R}^n$. Again, a geometer might be interested in the curvature of this surface and this, it turns out requires an understanding of how to differentiate the maps. There are then obvious generalisations to higher dimensional surfaces living in higher dimensional spaces.

From the physics perspective, in the map (0.2) that defines a curve the codomain $\mathbb{R}^n$ is viewed as physical space. A conceptually different set of functions arise when we think of the domain $\mathbb{R}^m$ as physical space. For example, we could consider maps of the kind

$$f : \mathbb{R}^3 \to \mathbb{R}$$

where $\mathbb{R}^3$ is viewed as physical space. Physicists refer to this as a *scalar field*. (Mathematicians refer to it as a map from $\mathbb{R}^3$ to $\mathbb{R}$.) A familiar example of such a map is temperature: there exists a temperature at every point in this room and that gives a map $T(\mathbf{x})$. This is shown in Figure 1. A more fundamental, and ultimately more interesting, example of a scalar field is the Higgs field in the Standard Model of particle physics.

As one final example, consider maps of the form

$$f : \mathbb{R}^3 \to \mathbb{R}^3$$

where, again, the domain $\mathbb{R}^3$ is identified with physical space. Physicists call these *vector fields*. (By now, you can guess what mathematicians call them.) In fundamental physics, two important examples are provided by the electric field $\mathbf{E}(\mathbf{x})$ and magnetic field $\mathbf{B}(\mathbf{x})$, first postulated by Michael Faraday: each describes a three-dimensional vector associated to each point in space.

1 Curves

In this section, we consider maps of the form

$$f : \mathbb{R} \to \mathbb{R}^n$$

A map of this kind is called a *parameterised curve*, with $\mathbb{R}$ the parameter and the curve the image of the map in $\mathbb{R}^n$. In what follows, we will denote the curve as C.

Whenever we do explicit calculations, we need to introduce some coordinates. The obvious ones are Cartesian coordinates, in which the vector $\mathbf{x} \in \mathbb{R}^n$ is written as

$$\mathbf{x} = (x^1, \ldots, x^n) = x^i \mathbf{e}_i$$

where, in the second expression, we're using summation convention and explicitly summing over $i = 1, \ldots, n$. Here $\{\mathbf{e}_i\}$ is a choice of orthonormal basis vectors, satisfying $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$. For $\mathbb{R}^n = \mathbb{R}^3$, we'll also write these as $\{\mathbf{e}_i\} = \{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$. (The notation $\{\mathbf{e}_i\} = \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ is also standard, although we won't adopt it in these lectures.)

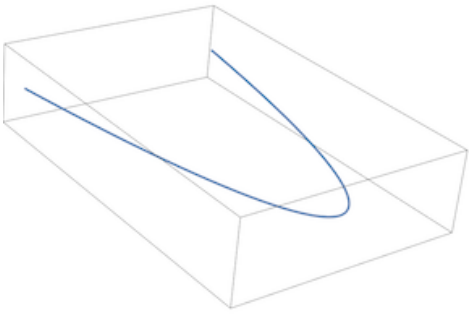
The image of the function can then be written as $\mathbf{x}(t)$. In physics, we might think of this as the trajectory of a particle evolving in time t. Here, we'll mostly just view the curve as an abstract mathematical map, with t nothing more than a parameter labelling positions along the curve. In fact, one of themes of this section is that, for many calculations, the choice of parameter t is irrelevant.

Examples

Here are two simple examples. Consider first the map $\mathbb{R} \to \mathbb{R}^3$ that takes the form

$$\mathbf{x}(t) = (at, bt^2, 0)$$

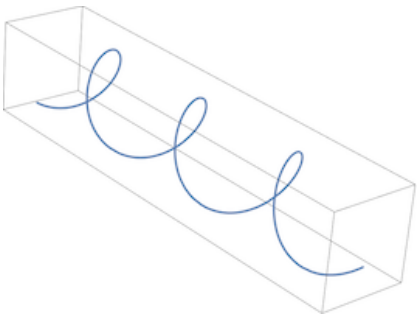
The image of the map is the parabola $a^2 y = bx^2$, lying in the plane $z = 0$, and is shown on the right.



This looks very similar to what you would draw if asked to plot the graph $y = bx^2/a^2$, with the additional requirement of $z = 0$ prompting the artistic flourish that results in a curve suspended in 3d. Obviously, the curve $\mathbf{x}(t)$ and the functions $y = bx^2/a^2$ (with $z = 0$) are related, but they're not quite the same thing. The function $y = bx^2/a^2$ is usually thought of as a map $\mathbb{R} \to \mathbb{R}$ and in plotting a graph you include both the domain and codomain. In contrast, on the right we've plotted only the image of the curve $\mathbf{x}(t)$ in $\mathbb{R}^3$; the picture loses all information about the domain coordinate t.

Here is a second example that illustrates the same point. Consider

$$\mathbf{x}(t) = (\cos t, \sin t, t) \tag{1.1}$$



The resulting curve is a helix, shown to the right. Like any other curve, the choice of parameterisation is not unique. We could, for example, consider the different map

$$\mathbf{x}(t) = (\cos \lambda t, \sin \lambda t, \lambda t)$$

This gives the same helix as (1.1) for any choice of $\lambda \in \mathbb{R}$ as long as $\lambda \neq 0$. In some contexts this matters. If, for example, t is time, and $\mathbf{x}(t)$ is the trajectory of a rollercoaster then the fate of the contents of your stomach depends delicately on the value of λ. However, there will be some properties of the curve that are independent of the choice of parameterisation and, in this example, independent of λ. It is these properties that will be our primary interest in this section.

Before we go on, a pedantic mathematical caveat. It may be that the domain of the curve is not all of $\mathbb{R}$. For example, we could have the map $\mathbb{R} \to \mathbb{R}^2$ given by $\mathbf{x}(t) = (t, \sqrt{1-t^2})$. This makes sense only for the interval $t \in [-1, +1]$ and you should proceed accordingly.

1.1 Differentiating the Curve

The vector function $\mathbf{x}(t)$ is *differentiable* at t if, as $\delta t \to 0$, we can write

$$\mathbf{x}(t + \delta t) - \mathbf{x}(t) = \dot{\mathbf{x}}(t)\, \delta t + \mathcal{O}(\delta t^2) \tag{1.2}$$

You should think of this expression as defining the derivative $\dot{\mathbf{x}}(t)$. If the derivative $\dot{\mathbf{x}}$ exists everywhere then the curve is said to be *smooth*. This means that it is continuous and, as the name suggests, not egregiously jagged.

There are some notational issues to unpick in this expression. First, $\mathcal{O}(\delta t^2)$ includes all terms that scale as δt^2 or smaller as $\delta t \to 0$. This "big-O" notation is commonly used in physics and applied mathematics. In pure maths you will also see the "little o" notation $o(\delta t)$ which means "strictly smaller than δt" as $\delta t \to 0$. Roughly speaking $o(\delta t)$ is the same thing as $\mathcal{O}(\delta t^2)$. (In other courses you may encounter situations where this speaking is too rough to be accurate, but it will suffice for our needs.) We'll stick with big-O notation throughout these lectures.

We've denoted the derivative in (1.2) with a dot, $\dot{\mathbf{x}}(t)$. This was Newton's original notation for the derivative and, 350 years later, comes with some sociological baggage. In physics, a dot is nearly always used to denote differentiation with respect to time, so the velocity of a particle is $\dot{\mathbf{x}}$ and the acceleration is $\ddot{\mathbf{x}}$. Meanwhile a prime, like $f'(x)$, is usually used to denote differentiation with respect to space. This is deeply ingrained in the psyche of physicists, so much so that I get a little shudder if I see something like $x'(t)$, even though it's perfectly obvious that it means dx/dt. Mathematicians, meanwhile, seem to have no such cultural hang-ups on this issue. (They reserve their cultural hang-ups for a 1000 other issues.)

We write the left-hand side of (1.2) as

$$\delta\mathbf{x}(t) = \mathbf{x}(t+\delta t) - \mathbf{x}(t)$$

The derivative is then the vector

$$\frac{d\mathbf{x}}{dt} = \dot{\mathbf{x}}(t) = \lim_{\delta t\to 0}\frac{\delta\mathbf{x}}{\delta t}$$

Here the familiar notation $d\mathbf{x}/dt$ for the derivative is due to Leibniz and works if we're differentiating with respect to time, space, or anything else. We'll also sometimes use the slightly sloppy notation and write

$$d\mathbf{x} = \dot{\mathbf{x}}\, dt$$

which, at least for now, really just means the same thing as (1.2) except we've dropped the $\mathcal{O}(\delta t^2)$ terms.

It's not difficult to differentiate vectors and, at least in Cartesian coordinates with the basis vectors $\mathbf{e}_i$, we can just do it component by component

$$\mathbf{x}(t) = x^i(t)\mathbf{e}_i \quad \Rightarrow \quad \dot{\mathbf{x}}(t) = \dot{x}^i(t)\mathbf{e}_i$$

The same is true if we work in any other choice of basis vectors $\{\mathbf{e}_i\}$ provided that these vectors themselves are independent of t. (In the lectures on Dynamics and Relativity we encounter an example where the basis vectors *do* depend on time and you have to be more careful. This arises in Section 6 on "Non-Inertial Frames".)

More generally, given a function $f(t)$ and two vector functions $\mathbf{g}(t)$ and $\mathbf{h}(t)$, it's simple to check that the following Leibniz identities hold

$$\begin{aligned}\frac{d}{dt}(f\mathbf{g}) &= \frac{df}{dt}\mathbf{g} + f\frac{d\mathbf{g}}{dt}\\ \frac{d}{dt}(\mathbf{g}\cdot\mathbf{h}) &= \frac{d\mathbf{g}}{dt}\cdot\mathbf{h} + \mathbf{g}\cdot\frac{d\mathbf{h}}{dt}\end{aligned}$$

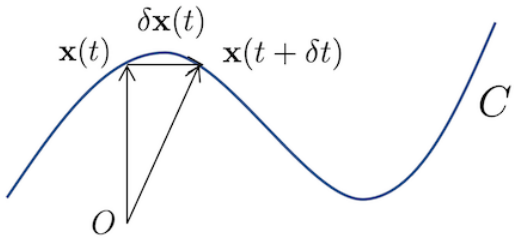


Figure 2. The derivative is the tangent vector to the curve.

Moreover, if $\mathbf{g}(t)$ and $\mathbf{h}(t)$ are vectors in $\mathbb{R}^3$, we also have the cross-product identity

$$\frac{d}{dt}(\mathbf{g}\times\mathbf{h}) = \frac{d\mathbf{g}}{dt}\times\mathbf{h} + \mathbf{g}\times\frac{d\mathbf{h}}{dt}$$

As usual, we have to be careful with the ordering of terms in the cross product because for example, $d\mathbf{g}/dt\times\mathbf{h} = -\mathbf{h}\times d\mathbf{g}/dt$.

1.1.1 Tangent Vectors

There is a nice geometric meaning to the derivative $\dot{\mathbf{x}}(t)$ of a parameterised curve C: it gives the tangent to the curve and is called, quite reasonably, the *tangent vector*. This is shown in Figure 2.

The direction of the tangent vector $\dot{\mathbf{x}}(t)$ is geometrical (at least up to a sign): it depends only on the curve C itself, and not on the choice of parameterisation. In contrast, the magnitude of the tangent vector $|\dot{\mathbf{x}}(t)|$ does depend on the parameterisation. This is obvious mathematically, since we're differentiating with respect to t, and also physically where $\dot{\mathbf{x}}$ is identified with the velocity of a particle.

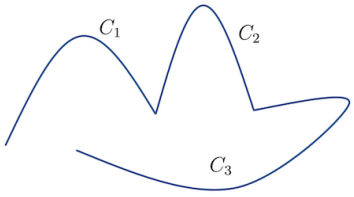
Sometimes, you may find yourself with an unwise choice of parameterisation in which the derivative vector $\dot{\mathbf{x}}$ vanishes at some point. For example, consider the curve in $\mathbb{R}^2$ given by

$$\mathbf{x}(t) = (t^3, t^3)$$

The curve C is just the straight line $x = y$. The tangent vector $\dot{x} = 3t^2(1,1)$ which clearly points along the line $x = y$ but with magnitude $3\sqrt{2}t^2$ and so vanishes at $t = 0$. Clearly this is not a property of C itself, but of our choice of parameterisation. We get the same curve C from the map $\mathbf{x}(t) = (t, t)$ but now the tangent vector is everywhere non-vanishing.

A parameterisation is called *regular* if $\dot{\mathbf{x}}(t) \neq 0$ for any t. In what follows, we will assume that we are dealing with regular parameterisations except, perhaps, at isolated points. This means that we can divide the curve into segments, each of which is regular.

As a slightly technical aside, we will sometimes have cause to consider curves that are *piecewise smooth* curves of the form $C = C_1 + C_2 + \ldots$, where the end of one curve lines up with the beginning of the next, as shown on the right. In this case, a tangent vector exists everywhere except at the cusps where two curves meet.



1.1.2 The Arc Length

We can use the tangent vectors to compute the length of the curve. From Figure 2, we see that the distance between two nearby points is

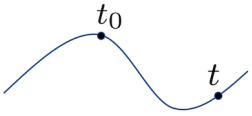
$$\delta s = |\delta \mathbf{x}| + \mathcal{O}(|\delta \mathbf{x}|^2) = |\dot{\mathbf{x}}\, \delta t| + \mathcal{O}(\delta t^2)$$

We then have

$$\frac{ds}{dt} = \pm \left| \frac{d\mathbf{x}}{dt} \right| = \pm |\dot{\mathbf{x}}| \tag{1.3}$$

where we get the plus sign for distances measured in the direction of increasing t, and the minus sign in the direction of decreasing t.

If we pick some starting point t_0 on the curve, then the distance along the curve to any point $t > t_0$ is given by



$$s = \int_{t_0}^{t} dt' \; |\dot{\mathbf{x}}(t')|$$

This distance is called the *arc length*, s. Because $|\dot{\mathbf{x}}| > 0$, this is a positive and strictly increasing function as we move away in the direction $t > t_0$. It is a negative, and strictly decreasing function in the direction $t < t_0$.

Although the tangent vector $\dot{\mathbf{x}}$ depends on the choice of parameterisation, the arc length s does not. We can pick a different parameterisation of the curve $\tau(t)$, which we will take to be an invertible and smooth function. We will also assume that $d\tau/dt > 0$

so that they both measure "increasing time" in the same direction. The chain rule tells us that

$$\frac{d\mathbf{x}}{dt} = \frac{d\mathbf{x}}{d\tau}\frac{d\tau}{dt} \tag{1.4}$$

We can then compute the arc length using the τ parameterisation: it is

$$s = \int_{t_0}^{t} dt' \ |\dot{\mathbf{x}}(t')| = \int_{\tau_0}^{\tau} d\tau' \ \frac{dt'}{d\tau'}\left|\frac{d\mathbf{x}}{d\tau'}\frac{d\tau'}{dt'}\right| = \int_{\tau_0}^{\tau} d\tau' \ \left|\frac{d\mathbf{x}}{d\tau'}\right| \tag{1.5}$$

In the second equality, we find the contribution from the chain rule (1.4) together with a factor from the measure that comes from integrating over $d\tau$ instead of dt. These then cancel in the third equality. The upshot is that we can compute the arc length using any parameterisation that we wish. Or, said differently, the arc length is independent of the choice of parameterisation of the curve.

We can now turn this on its head. All parameterisations of the curve give the same arc length. But this means that the arc length itself is, in many ways, the only natural parameterisation of the curve. We can then think of $\mathbf{x}(s)$ with the corresponding tangent vector $d\mathbf{x}/ds$. From (1.3), we see that this choice of the tangent vector always has unit length: $|d\mathbf{x}/ds| = 1$.

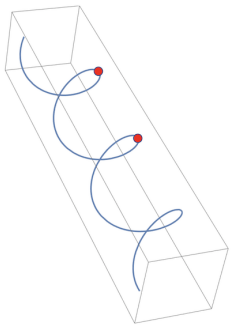
As an aside: these kind of issues raise their head in the physics of special relativity where time means different things for people moving at different speeds. This means that there is no universally agreed "absolute time" and so different people will parameterise the trajectory of a particle $\mathbf{x}(t)$ in different ways. There's no right or wrong way, but it's annoying if someone does it differently to you. (Admittedly, this is only likely to happen if they are travelling at an appreciable fraction of the speed of light relative to you.) Happily there is something that everyone can agree on, which is the special relativistic version of arc length. It's known as *proper time*. You can read more about this in the lectures on Dynamics and Relativity.

An Example

To illustrate these ideas, let's return to our helix example of (1.1). We had $\mathbf{x}(t) = (\cos t, \sin t, t)$ and so $\dot{\mathbf{x}}(t) = (-\sin t, \cos t, 1)$. Our defining equation (1.3) then becomes (taking the positive sign)

$$\frac{ds}{dt} = |\dot{\mathbf{x}}| = \sqrt{2}$$

If we take $t_0 = 0$, then the arc length measured from the point $\mathbf{x} = (1, 0, 0)$ is $s = \sqrt{2}t$. In particular, after time $t = 2\pi$ we've made a full rotation and sit at $\mathbf{x} = (1, 0, 2\pi)$. These two points are shown as red dots in the figure. Obviously the direct route between the two has distance 2π. Our analysis above shows that the distance along the helix is $s = \sqrt{8}\pi$.



1.1.3 Curvature and Torsion

There is a little bit of simple geometry associated to these ideas. Given a curve C, parameterised by its arc length s, we have already seen that the tangent vector

$$\mathbf{t} = \frac{d\mathbf{x}}{ds}$$

has unit length, $|\mathbf{t}| = 1$. (Note: don't confuse the bold faced tangent vector $\mathbf{t}$ with our earlier parameterisation t: they're different objects!) We can also consider the "acceleration" of the curve with respect to the arc length, $d^2\mathbf{x}/ds^2$. The magnitude of this "acceleration" is called the *curvature*

$$\kappa(s) = \left|\frac{d^2\mathbf{x}}{ds^2}\right| \tag{1.6}$$

To build some intuition, we can calculate the curvature of a circle of radius R. If we start with a simple parameterisation $\mathbf{x}(t) = (R\cos t, R\sin t)$ then you can check using (1.3) that the arc length is $s = Rt$. We then pick the new parameterisation $\mathbf{x}(s) = (R\cos(s/R), R\sin(s/R))$. We then find that a circle of radius R has constant curvature

$$\kappa = \frac{1}{R}$$

Note, in particular, that as $R \to \infty$, the circle becomes a straight line which has vanishing curvature.

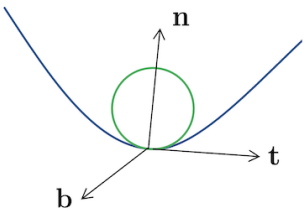
There is also a unit vector associated to this "acceleration", defined as

$$\mathbf{n} = \frac{1}{\kappa}\frac{d^2\mathbf{x}}{ds^2} = \frac{1}{\kappa}\frac{d\mathbf{t}}{ds}$$

This is known as the *principal normal*. Note that the factor of $1/\kappa$ ensures that $|\mathbf{n}| = 1$.

Importantly, if $\kappa \neq 0$ then $\mathbf{n}$ is perpendicular to the tangent vector $\mathbf{t}$. This follows from the fact that $\mathbf{t} \cdot \mathbf{t} = 1$ and so $d/ds(\mathbf{t} \cdot \mathbf{t}) = 2\kappa \mathbf{n} \cdot \mathbf{t} = 0$. This means that $\mathbf{t}$ and $\mathbf{n}$ define a plane, associated to every point in the curve. This is known as the *osculating plane*.

For any point s on the curve, there is an associated osculating plane. Now draw a circle in that plane that touches the curve at the point s, whose curvature matches $\kappa(s)$. This is called the *osculating circle* and is shown in green in the figure. This is the circle that just kisses the curve at s



Next we can ask: how does the osculating plane vary as we move along the curve? This is simplest to discuss if we restrict to curves in $\mathbf{R}^3$. In this case, we have the cross product at our disposal and we can define the unit normal to the osculating plane as

$$\mathbf{b} = \mathbf{t} \times \mathbf{n}$$

This is known as the *binormal*, to distinguish it from the normal $\mathbf{n}$. The three vectors $\mathbf{t}$, $\mathbf{n}$ and $\mathbf{b}$ define an orthonormal basis for $\mathbb{R}^3$ at each point s along the curve (at least as long as $\kappa(s) \neq 0$.) This basis twists and turns along the curve.

Note that $|\mathbf{b}| = 1$ which, using the same argument as for $\mathbf{t}$ above, tells us that $\mathbf{b} \cdot d\mathbf{b}/ds = 0$. In addition, we have $\mathbf{t} \cdot \mathbf{b} = 0$ which, after differentiating, tells us that

$$0 = \frac{d\mathbf{t}}{ds} \cdot \mathbf{b} + \mathbf{t} \cdot \frac{d\mathbf{b}}{ds} = \kappa \mathbf{n} \cdot \mathbf{b} + \mathbf{t} \cdot \frac{d\mathbf{b}}{ds}$$

But, by definition, $\mathbf{n} \cdot \mathbf{b} = 0$. So we learn that $\mathbf{t} \cdot d\mathbf{b}/ds = 0$. In other words, $d\mathbf{b}/ds$ is orthogonal to both $\mathbf{b}$ and to $\mathbf{t}$. Which means that it must be parallel to $\mathbf{n}$. We define the *torsion* $\tau(s)$ as a measure of how the binormal changes

$$\frac{d\mathbf{b}}{ds} = -\tau(s)\mathbf{n} \tag{1.7}$$

From the definition, you can see that the torsion is a measure of $\dddot{\mathbf{x}}$. The minus sign means that if the top of the green circle in the figure tilts towards us, then $\tau > 0$; if it tilts away from us then $\tau < 0$. Heuristically, the curvature captures how much the curve fails to be a straight line, while the torsion captures how much the curve fails to be planar.

The Frenet-Serret Equations

There is a closed set of formulae describing curvature and torsion. These are the *Frenet-Serret* equations,

$$\frac{d\mathbf{t}}{ds} = \kappa\mathbf{n} \tag{1.8}$$

$$\frac{d\mathbf{b}}{ds} = -\tau\mathbf{n} \tag{1.9}$$

$$\frac{d\mathbf{n}}{ds} = \tau\mathbf{b} - \kappa\mathbf{t} \tag{1.10}$$

The first of these (1.8) is simply the definition of the normal $\mathbf{n}$.

That leaves us with (1.10). We'll again start with the definition $\mathbf{b} = \mathbf{t} \times \mathbf{n}$, and this time take the cross product with $\mathbf{t}$. The triple product formula then gives us

$$\mathbf{b} \times \mathbf{t} = (\mathbf{t} \times \mathbf{n}) \times \mathbf{t} = (\mathbf{t} \cdot \mathbf{t})\,\mathbf{n} - (\mathbf{n} \cdot \mathbf{t})\,\mathbf{t} = \mathbf{n}$$

Now taking the derivative with respect to s, using (1.8) and (1.9) and noting that $\mathbf{b} = \mathbf{t} \times \mathbf{n}$ and $\mathbf{t} = \mathbf{n} \times \mathbf{b}$ then gives us (1.10).

It's useful to rewrite the first two equations (1.8) and (1.9) using $\mathbf{n} = \mathbf{b} \times \mathbf{t}$ so that we have

$$\frac{d\mathbf{t}}{ds} = \kappa(\mathbf{b} \times \mathbf{t}) \quad \text{and} \quad \frac{d\mathbf{b}}{ds} = -\tau(\mathbf{b} \times \mathbf{t})$$

This is six first order equations for six unknowns, $\mathbf{b}(s)$ and $\mathbf{t}(s)$. If we are given $\kappa(s)$ and $\tau(s)$, together with initial conditions $\mathbf{b}(0)$ and $\mathbf{t}(0)$, then we can solve for $\mathbf{b}(s)$ and $\mathbf{t}(s)$ and can subsequently solve for the curve $\mathbf{x}(s)$. The way to think about this is that the curvature and torsion $\kappa(s)$ and $\tau(s)$ specify the curve, up to translation and orientation.

1.2 Line Integrals

Given a curve C in $\mathbb{R}^n$ and some function defined over $\mathbb{R}^n$, we may well wish to integrate the function along the curve. There are different stories to tell for scalar and vector fields and we deal with each in turn.

1.2.1 Scalar Fields

A scalar field is a map

$$\phi : \mathbb{R}^n \to \mathbb{R}$$

With coordinates $\mathbf{x}$ on $\mathbb{R}^n$, we'll denote this scalar field as $\phi(\mathbf{x})$.

Given a parameterised curve C in $\mathbb{R}^n$, which we denote as $\mathbf{x}(t)$, it might be tempting to put these together to get the function $\phi(\mathbf{x}(t))$ which is a composite map $\mathbb{R} \to \mathbb{R}$. We could then just integrate over t in the usual way.

However, there's a catch. The result that you get will depend both on the function ϕ, the curve C, *and* the choice of parameterisation of the curve. There's nothing wrong this per se, but it's not what we want here. For many purposes, it turns out to be more useful to have a definition of the integral that depends only on the function ϕ and the curve C, but gives the same answer for any choice of parameterisation of the curve.

One way to achieve this is to work with the arc length s which, as we've seen, is the natural parameterisation along the curve. We can integrate from point $\mathbf{a}$ to point $\mathbf{b}$, with $\mathbf{x}(s_a) = \mathbf{a}$ and $\mathbf{x}(s_b) = \mathbf{b}$ and $s_a < s_b$, by defining the *line integral*

$$\int_C \phi \, ds = \int_{s_a}^{s_b} \phi(\mathbf{x}(s)) \, ds$$

where the right-hand side is now viewed as a usual one-dimensional integral.

This line integral is, by convention, defined so that $\int_C ds$ gives the length of the curve C and, in particular, is always positive. In other words, there's no directional information in this integral: it doesn't matter what way you move along the curve.

Suppose that we're given a parameterised curve C in terms of some other parameter $\mathbf{x}(t)$, with $\mathbf{x}(t_a) = \mathbf{a}$ and $\mathbf{x}(t_b) = \mathbf{b}$. The usual change of variables tells us that

$$\int_C \phi \, ds = \int_{t_a}^{t_b} \phi(\mathbf{x}(t)) \frac{ds}{dt} \, dt$$

We can then use (1.3). If $t_b > t_a$ then we have $ds/dt = +|\dot{\mathbf{x}}|$ and

$$\int_C \phi \, ds = \int_{t_a}^{t_b} \phi(\mathbf{x}(t)) \, |\dot{\mathbf{x}}(t)| \, dt \tag{1.11}$$

Meanwhile, if $t_b < t_a$ then we have $ds/dt = -|\dot{\mathbf{x}}|$ and

$$\int_C \phi \, ds = \int_{t_b}^{t_a} \phi(\mathbf{x}(t)) \, |\dot{\mathbf{x}}(t)| \, dt$$

We see that the line integral comes with the length of the tangent vector $|\dot{\mathbf{x}}|$ in the integrand. This is what ensures that the line integral is actually independent of the choice of parameterisation: the argument is the same as the one we used in (1.5) to

show that the arc length is invariant under reparameterisations: upon a change of variables, the single derivative d/dt in $\dot{\mathbf{x}}$ cancels the Jacobian from the integral $\int dt$. Furthermore, the minus signs work out so that you're always integrating from a smaller value of t to a larger one, again ensuring that $\int_C ds$ is positive and so can be interpreted as the length of the curve.

1.2.2 Vector Fields

Vector fields are maps of the form

$$\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^n$$

So that at each point $\mathbf{x} \in \mathbb{R}^n$ we have a vector-valued object $\mathbf{F}(\mathbf{x})$. We would like to understand how to integrate a vector field along a curve C.

There are two ways to do this. We could work component-wise, treating each component like the scalar field example above. After doing the integration, this would leave us with a vector.

However it turns out that, in many circumstances, it's more useful to integrate the vector field so that the integral gives us just a number. We do this integrating the component of the vector field that lies tangent to the curve. Usually, this is what is meant by the *line integral* of a vector field.

In more detail, suppose that our curve C has a parameterisation $\mathbf{x}(t)$ and we wish to integrate from t_a to t_b, with $\mathbf{x}(t_a) = \mathbf{a}$ and $\mathbf{x}(t_b) = \mathbf{b}$. The line integral of a vector field $\mathbf{F}$ along C is defined to be

$$\int_C \mathbf{F}(\mathbf{x}) \cdot d\mathbf{x} = \int_{t_a}^{t_b} \mathbf{F}(\mathbf{x}(t)) \cdot \dot{\mathbf{x}}(t)\ dt \tag{1.12}$$

Once again, this doesn't depend on the choice of parameterisation t. This is manifest in the expression on the left where the parameterisation isn't mentioned. The right-hand side is invariant for the same reason as (1.11).

This time, however, there's a slightly different story to tell about minus signs. We should think of each curve C as coming with an *orientation*, which is the direction along the curve. Equivalently, it can be thought of as the direction of the tangent vector $\dot{\mathbf{x}}$. In the example above, the orientation of the curve is from $\mathbf{a}$ to $\mathbf{b}$. This then determines the limits of the integral, from t_a to t_b, since $\mathbf{x}(t_a) = \mathbf{a}$ and $\mathbf{x}(t_b) = \mathbf{b}$. Note that the limits are always this way round, regardless of whether our parameterisation has $t_a < t_b$ or whether $t_b > t_a$: the orientation determines the limits, not the parameterisation.

In summary, the line integral for a scalar field $\int_C \phi\, ds$ is independent of the orientation and, if ϕ is positive, the integral will also be positive. In contrast, the integral of the vector field $\int_C \mathbf{F}\cdot\dot{\mathbf{x}}\, dt$ depends on the orientation. Flip the orientation of the curve, and the integral will change sign.

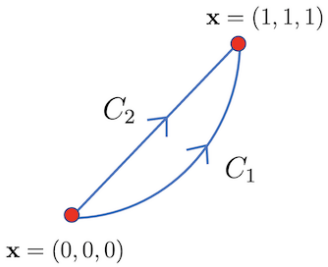
An Example

As a slightly baroque example, consider the vector field in $\mathbb{R}^3$,

$$\mathbf{F}(\mathbf{x}) = (xe^y, z^2, xy)$$

To evaluate the line integral, we also need to specify the curve C along which we perform the integral. We'll consider two options, both of which evolve from $\mathbf{x}(t=0) = (0,0,0)$ to $\mathbf{x}(t=1) = (1,1,1)$. Our first curve is

$$C_1: \qquad \mathbf{x}(t) = (t, t^2, t^3)$$



This is shown in the figure. Evaluated on C_1, we have $\mathbf{F}(\mathbf{x}(t)) = (te^{t^2}, t^6, t^3)$. Meanwhile $\dot{\mathbf{x}} = (1, 2t, 3t^2)$ so we have

$$\begin{aligned}\int_{C_1} \mathbf{F}\cdot d\mathbf{x} &= \int_0^1 dt\ \mathbf{F}\cdot\dot{\mathbf{x}} \\ &= \int_0^1 dt\ \left(te^{t^2} + 2t^7 + 3t^5\right) = \frac{1}{4}\left(1+2e\right)\end{aligned}$$

Our second curve is simply the straight line

$$C_2: \qquad \mathbf{x}(t) = (t, t, t)$$

Evaluated on this curve, we have $\mathbf{F}(\mathbf{x}(t)) = (te^t, t^2, t^2)$. Now the tangent vector is $\dot{x} = (1,1,1)$ and the integral is

$$\int_{C_2} \mathbf{F}\cdot d\mathbf{x} = \int_0^1 dt\ \mathbf{F}\cdot\dot{\mathbf{x}} = \int_0^1 dt\ \left(te^t + 2t^2\right) = \frac{5}{3} \tag{1.13}$$

(The first of these integrals is done by an integration by parts.)

The main lesson to take from this is the obvious one: the answers are different. The result of a line integral generally depends on both the thing you're integrating $\mathbf{F}$ and the choice of curve C.

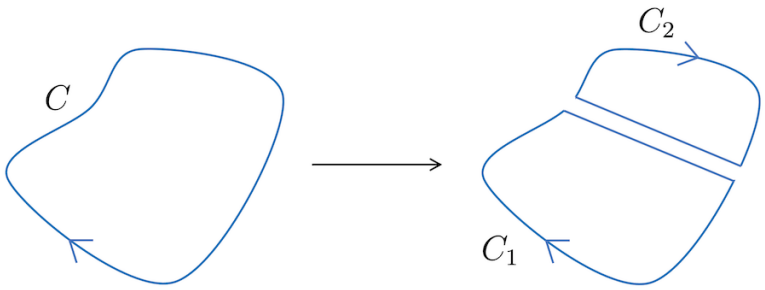


Figure 3. Decomposing a curve by introducing new segments with opposite orientations.

More Curves, More Integrals

We'll see plenty more examples of line integrals, both in this course and in later ones. Here are some comments to set the scene.

First, there will be occasions when we want to perform a line integral around a *closed curve* C, meaning that the starting and end points are the same, $\mathbf{a} = \mathbf{b}$. For such curves, we introduce new notation and write the line integral as

$$\oint_C \mathbf{F} \cdot d\mathbf{x}$$

with the little circle on the integral sign there to remind us that we're integrating around a loop. This quantity is called the *circulation* of $\mathbf{F}$ around C. The name comes from Fluid Mechanics where we might view $\mathbf{F}$ as the velocity field of a fluid, and the circulation quantifies the swirling motion of the fluid.

In other occasions, we may find ourselves in a situation in which the curve C decomposes into a number of piecewise smooth curves C_i, joined up at their end points. We write $C = C_1 + C_2 + \ldots$, and the line integral is

$$\int_C \mathbf{F} \cdot d\mathbf{x} = \int_{C_1} \mathbf{F} \cdot d\mathbf{x} + \int_{C_2} \mathbf{F} \cdot d\mathbf{x} + \ldots$$

It is also useful to think of the curve $-C$ as the same as the curve C but with the opposite orientation. This means that we have the expression

$$\int_{-C} \mathbf{F}(\mathbf{x}) \cdot d\mathbf{x} = -\int_C \mathbf{F}(\mathbf{x}) \cdot d\mathbf{x}$$

For example, we could return to our previous baroque example and consider the closed curve $C = C_1 - C_2$. This curve starts at $\mathbf{x} = (0, 0, 0)$, travels along C_1 to $\mathbf{x} = (1, 1, 1)$

and then returns back along C_2 in the opposite direction to the arrow. From our previous answers, we have

$$\oint_C \mathbf{F} \cdot d\mathbf{x} = \int_{C_1} \mathbf{F} \cdot d\mathbf{x} - \int_{C_2} \mathbf{F} \cdot d\mathbf{x} = \frac{1}{4}(1+2e) - \frac{5}{3}$$

There are lots of games that we can play like this. For example, it's sometimes useful to take a smooth closed curve C and decompose it into two piecewise smooth segments,. An example is shown in Figure 3, where we've introduced two new segments, which should be viewed as infinitesimally close to each other. These two new segments have opposite orientation and so cancel out in any integral. In this way, we can think of the original curve as $C = C_1 + C_2$. We'll see other examples of these kinds of manipulations as we progress.

1.3 Conservative Fields

Here's an interesting question. In general the line integral of a vector field depends on the path taken. But is this ever not the case? In other words, are there some vector fields $\mathbf{F}$ for which the line integral depends only on the end points and not on the route you choose to go between them?

Such a vector field $\mathbf{F}$ would obey

$$\int_{C_1} \mathbf{F} \cdot d\mathbf{x} = \int_{C_2} \mathbf{F} \cdot d\mathbf{x}$$

for any C_1 and C_2 that share the same end points $\mathbf{a}$ and $\mathbf{b}$ and the same orientation. Equivalently, we could consider the closed curve $C = C_1 - C_2$ and write this as

$$\oint_C \mathbf{F} \cdot d\mathbf{x} = 0$$

for all closed curves C. To answer this question about vector fields, we first need to introduce a new concept for scalar fields.

1.3.1 The Gradient

Let's return to the scalar field

$$\phi : \mathbb{R}^n \to \mathbb{R}$$

We want to ask: how can we differentiate such a function?

With Cartesian coordinates $\mathbf{x} = (x^1, \ldots, x^n)$ on $\mathbb{R}^n$, the scalar field is a function $\phi(x^1, \ldots, x^n)$. Given such a function of several variables, we can always take *partial derivatives*, which means that we differentiate with respect to one variable while keeping all others fixed. For example,

$$\frac{\partial \phi}{\partial x^1} = \lim_{\epsilon \to 0} \frac{\phi(x^1 + \epsilon, x^2, \ldots, x^n) - \phi(x^1, x^2, \ldots, x^n)}{\epsilon} \tag{1.14}$$

If all n partial derivatives exist then the function is said to be *differentiable*.

The partial derivatives offer n different ways to differentiate our scalar field. We will sometimes write this as

$$\partial_i \phi = \frac{\partial \phi}{\partial x^i} \tag{1.15}$$

where the ∂_i can be useful shorthand when doing long calculations. While the notation of the partial derivative tells us what's changing it's just as important to remember what's kept fixed. If, at times, there's any ambiguity this is sometimes highlighted by writing

$$\left(\frac{\partial \phi}{\partial x^1} \right)_{x^2, \ldots, x^n}$$

where the subscripts tell us what remains unchanged as we vary x^1. We won't use this notation in these lectures since it should be obvious what variables are being held fixed.

The n different partial derivatives can be packaged together into a vector field. To do this, we introduce the orthonormal basis of vectors $\{\mathbf{e}_i\}$ associated to the coordinates x^i. The *gradient* of a scalar field is then a vector field, defined as

$$\nabla \phi = \frac{\partial \phi}{\partial x^i} \, \mathbf{e}_i \tag{1.16}$$

where we're using the summation convention in which we implicitly sum over the repeated $i = 1, \ldots, n$ index.

Because $\nabla \phi$ is a vector field, it may be more notationally consistent to write it in bold font as $\boldsymbol{\nabla} \phi$. However, I'll stick with $\nabla \phi$. There's no ambiguity here because the symbol ∇ only ever means the gradient, never anything else, and so is *always* a vector. It's one of the few symbols in mathematics and physics whose notational meaning is fixed.

For scalar fields $\phi(x, y, z)$ in $\mathbb{R}^3$, the gradient is

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

where we've written the orthonormal basis as $\{\mathbf{e}_i\} = \{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$.

There's a useful way to view the vector field $\nabla\phi$. To see this, note that if we want to know how the function ϕ changes in a given direction $\hat{\mathbf{n}}$, with $|\hat{\mathbf{n}}| = 1$, then we just need to take the inner product $\hat{\mathbf{n}} \cdot \nabla\phi$. This is known as the *directional derivative* and sometimes denoted $D_{\mathbf{n}}\phi = \hat{\mathbf{n}} \cdot \nabla\phi$. Obviously the directional derivative is maximal at any point $\mathbf{x}$ when $\hat{\mathbf{n}}$ lies parallel to $\nabla\phi(\mathbf{x})$. But this is telling us something important: at each point in space, the vector $\nabla\phi(\mathbf{x})$ is pointing in the direction in which $\phi(\mathbf{x})$ changes most quickly.

1.3.2 Back to Conservative Fields

First a definition. A vector field $\mathbf{F}$ is called *conservative* if it can be written as

$$\mathbf{F} = \nabla\phi$$

for some scalar field ϕ which, in this context, is referred to as a *potential*. (The odd name "conservative" derives from the conservation of energy in Newtonian mechanics we will see the connection to this below.) Finally, we can answer the question that we introduced at the beginning of this section: when is a line integral independent of the path?

Claim: The line integral around any closed curve vanishes if and only if $\mathbf{F}$ is conservative.

Proof: Consider a conservative vector field of the form $\mathbf{F} = \nabla\phi$. We'll integrate this along a curve C that interpolates from point $\mathbf{a}$ to point $\mathbf{b}$, with parameterisation $\mathbf{x}(t)$. We have

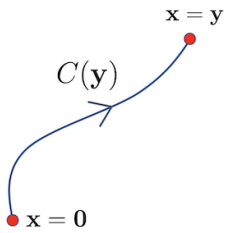
$$\int_C \mathbf{F} \cdot d\mathbf{x} = \int_C \nabla\phi \cdot d\mathbf{x} = \int_{t_a}^{t_b} \frac{\partial\phi}{\partial x^i}\frac{dx^i}{dt}\,dt = \int_{t_a}^{t_b} \frac{d}{dt}\phi(\mathbf{x}(t))\,dt$$

where the last equality follows from the chain rule. But now we have the integral of a total derivative, so

$$\int_C \mathbf{F} \cdot d\mathbf{x} = \Big[\phi(\mathbf{x}(t))\Big]_{t_a}^{t_b} = \phi(\mathbf{b}) - \phi(\mathbf{a})$$

which depends only on the end points as promised.

Conversely, given the vector field $\mathbf{F}$ whose integral vanishes when taken around any closed curve, it is always possible to construct a potential ϕ. We first choose a value of ϕ at the origin. There's no unique choice here, reflecting the fact that the potential ϕ is only defined up to an overall constant. We can take $\phi(\mathbf{0}) = 0$. Then, at any other point $\mathbf{y}$, we define $\mathbf{x} = \mathbf{y}$



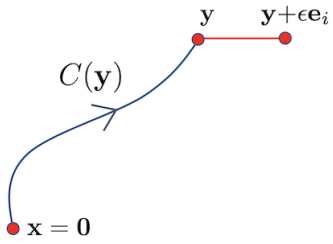
$$\phi(\mathbf{y}) = \int_{C(\mathbf{y})} \mathbf{F} \cdot d\mathbf{x}$$

where $C(\mathbf{y})$ is a curve that starts at the origin and ends at the point $\mathbf{y}$ as shown in the figure above. Importantly, by assumption $\oint \mathbf{F} \cdot d\mathbf{x} = 0$, so it doesn't matter which curve C we take: they all give the same answer.

It remains only to show that $\nabla\phi = \mathbf{F}$. This is straightforward. Reverting to our original definition of the partial derivative (1.14), we have

$$\frac{\partial\phi}{\partial x^i}(\mathbf{y}) = \lim_{\epsilon\to 0} \frac{1}{\epsilon}\left[\int_{C(\mathbf{y}+\epsilon\mathbf{e}_i)} \mathbf{F}\cdot d\mathbf{x} - \int_{C(\mathbf{y})} \mathbf{F}\cdot d\mathbf{x}\right]$$

The first integral goes along $C(\mathbf{y})$, and then continues along the red line shown in the figure to the right. Meanwhile, the second integral goes back along $C(\mathbf{y})$. The upshot is that the difference between them involves only the integral along the red line



$$\frac{\partial\phi}{\partial x^i}(\mathbf{y}) = \lim_{\epsilon\to 0} \frac{1}{\epsilon} \int_{\text{red line}} \mathbf{F}\cdot d\mathbf{x}$$

The red line is taken to be the straight line in the x^i direction. This means that the line integral projects onto the F_i component of the vector $\mathbf{F}$. Since we're integrating this over a small segment of length ϵ, the integral gives $\int_{\text{red line}} \mathbf{F}\cdot d\mathbf{x} \approx F_i\epsilon$ and, after taking the limit $\epsilon \to 0$, we have

$$\frac{\partial\phi}{\partial x^i}(\mathbf{y}) = F_i(\mathbf{y})$$

This is our desired result $\nabla\phi = \mathbf{F}$. □

It's clear that the result above is closely related to the fundamental theorem of calculus: the line integral of a conservative vector field is the analog of the integral of a total derivative and so is given by the end points. We'll meet more analogies along the same lines as we proceed.

Given a vector field $\mathbf{F}$, how can we tell if there's a corresponding potential so that we can write $\mathbf{F} = \nabla\phi$? There's one straightforward way to check: for a conservative vector field, the components $\mathbf{F} = F_i\mathbf{e}_i$ are given by

$$F_i = \frac{\partial\phi}{\partial x^i}$$

Differentiating again, we have

$$\frac{\partial F_i}{\partial x^j} = \frac{\partial^2\phi}{\partial x^i x^j} = \frac{\partial F_j}{\partial x^i} \tag{1.17}$$

where the second equality follows from the fact that the order of partial derivatives doesn't matter (at least for suitably well behaved functions). This means that a necessary condition for $\mathbf{F}$ to be conservative if that $\partial_i F_j = \partial_j F_i$. Later in these lectures we will see that (at least locally) this is actually a sufficient condition.

An Example

Consider the (totally made up) vector field

$$\mathbf{F} = (3x^2y\sin z, x^3\sin z, x^3y\cos z)$$

Is this conservative? We have $\partial_1 F_2 = 3x^2\sin z = \partial_2 F_1$ and $\partial_1 F_3 = 3x^2y\cos z = \partial_3 F_1$ and, finally, $\partial_2 F_3 = x^3\cos z = \partial_3 F_2$. So it passes the derivative test. Indeed, it's not then hard to check that

$$\mathbf{F} = \nabla\phi \quad \text{with} \quad \phi = x^3y\sin z$$

Knowing this makes it trivial to evaluate the line integral $\int_C \mathbf{F}\cdot d\mathbf{x}$ along any curve C since it is given by $\phi(\mathbf{b}) - \phi(\mathbf{a})$ where $\mathbf{a}$ and $\mathbf{b}$ are the end points of C.

Exact Differentials

There is a slightly different and more abstract way of phrasing the idea of a conservative vector field. First, given a function $\phi(\mathbf{x})$ on $\mathbb{R}^n$, the *differential* is defined to be

$$d\phi = \frac{\partial\phi}{\partial x^i}\,dx^i = \nabla\phi\cdot d\mathbf{x}$$

It's a slightly formal object, obviously closely related to the derivative. The differential is itself a function of $\mathbf{x}$ and captures how much the function ϕ changes as we move in any direction.

Next, consider a vector field $\mathbf{F}(\mathbf{x})$ on $\mathbb{R}^n$. We can take the inner product with an infinitesimal vector to get the object $\mathbf{F} \cdot d\mathbf{x}$. In fancy maths language, this is called a *differential form*. (Strictly it's an object known as a differential one-form) It's best to think of $\mathbf{F} \cdot d\mathbf{x}$ as something that we should integrate along a curve.

A differential form is said to be *exact* if it can be written as

$$\mathbf{F} \cdot d\mathbf{x} = d\phi$$

for some function ϕ. This is just a rewriting of our earlier idea: a differential is exact if and only if the vector field is conservative. In this case, it takes the form $\mathbf{F} = \nabla\phi$ and so the associated differential is

$$\mathbf{F} \cdot d\mathbf{x} = \frac{\partial \phi}{\partial x^i}\, dx^i = d\phi$$

where the last equality follows from the chain rule.

1.3.3 An Application: Work and Potential Energy

There's a useful application of these ideas in Newtonian mechanics. The trajectory $\mathbf{x}(t)$ of a particle is governed by Newton's second law which reads

$$m\ddot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$$

where, in this context, $\mathbf{F}(\mathbf{x})$ can be thought of as a force field. An important concept in Newtonian mechanics is the kinetic energy of a particle, $K = \frac{1}{2}m\dot{\mathbf{x}}^2$. (This is more often denoted as T in theoretical physics.) As the particle's position changes in time, the kinetic energy changes as

$$K(t_2) - K(t_1) = \int_{t_1}^{t_2} \frac{dK}{dt}\, dt = \int_{t_1}^{t_2} m\dot{\mathbf{x}} \cdot \ddot{\mathbf{x}}\, dt = \int_{t_1}^{t_2} \dot{\mathbf{x}} \cdot \mathbf{F}\, dt = \int_C \mathbf{F} \cdot d\mathbf{x}$$

The line integral of the force $\mathbf{F}$ along the trajectory C of the particle is called the *work done*.

Something special happens for conservative forces. These can be written as

$$\mathbf{F} = -\nabla V \tag{1.18}$$

for some choice of V. (Note: the minus sign is just convention.) From the result above, for a conservative force the work done depends only on the end points, not on the path taken. We then have

$$K(t_2) - K(t_1) = \int_C \mathbf{F} \cdot d\mathbf{x} = -V(t_2) + V(t_1) \quad \Rightarrow \quad K(t) + V(t) = \text{constant}$$

We learn that a conservative force, one that can be written as (1.18), has a conserved energy $E = K + V$. Indeed, it's this conservation of energy that lends it's name to the more general idea of a "conservative" vector field. We'll have use of these ideas in the lectures on Dynamics and Relativity.

1.3.4 A Subtlety

Here's a curious example. Consider the vector field on $\mathbb{R}^2$ given by

$$\mathbf{F} = \left(-\frac{y}{x^2+y^2}, \frac{x}{x^2+y^2} \right)$$

Is this conservative? If we run our check (1.17), we find

$$\frac{\partial F_x}{\partial y} = \frac{\partial F_y}{\partial x} = \frac{y^2 - x^2}{(x^2+y^2)^2}$$

which suggests that this is, indeed, a conservative field. Indeed, you can quickly check that

$$\mathbf{F} = \nabla\phi \quad \text{with} \quad \phi(x,y) = \tan^{-1}\left(\frac{y}{x}\right)$$

(To see this, write $\tan\phi = y/x$ and recall that $\partial(\tan\phi)/\partial x = (\cos\phi)^{-2}\partial\phi/\partial x = (1 + \tan^2\phi)\partial\phi/\partial x$ with a similar expression when you differentiate with respect to y. A little algebra will then convince you that the above is true.)

Let's now integrate $\mathbf{F}$ along a closed curve C that is a circle of radius R surrounding the origin. We take $\mathbf{x}(t) = (R\cos t, R\sin t)$ with $0 \leq t < 2\pi$ and the line integral is

$$\oint_C \mathbf{F}\cdot d\mathbf{x} = \int_0^{2\pi} \mathbf{F}\cdot\frac{d\mathbf{x}}{dt}\,dt = \int_0^{2\pi}\left(-\frac{\sin t}{R}\cdot(-R\sin t) + \frac{\cos t}{R}\cdot R\cos t\right)dt = 2\pi$$

Well, that's annoying! We've just proven that the integral of any conservative vector field around a close curve C necessarily vanishes, and yet one of our first examples seems to show otherwise! What's going on?

The deal is that $\phi(x,y)$ is *not* a well behaved function on $\mathbb{R}^2$. In particular, it's not continuous along the y-axis: as $x \to 0$ the function ϕ approaches either $+\pi/2$ or $-\pi/2$ depending on whether y/x is positive or negative. Implicit in our previous proof was the requirement that we have a continuous function ϕ, well defined everywhere on $\mathbb{R}^2$. Strictly speaking, a conservative field should have $\mathbf{F} = \nabla\phi$ with ϕ continuous.

Relatedly, $\mathbf{F}$ itself isn't defined everywhere on $\mathbb{R}^2$ because it is singular at the origin. Strictly speaking, $\mathbf{F}$ is only defined on the plane $\mathbb{R}^2$ with the point at the origin removed. We write this as $\mathbb{R}^2 - \{0,0\}$,

We learn that we should be careful. The line integral of a conservative vector field around a closed curve C is only vanishing if the vector field is well defined everywhere inside C.

Usually pathological examples like this are of interest only to the most self-loathing of pure mathematicians. But not in this case. The subtlety that we've seen above later blossoms into some of the most interesting ideas in both mathematics and physics where it underlies key aspects in the study of topology. In the above example, the space $\mathbb{R}^2 - \{0, 0\}$ has a different topology from $\mathbb{R}^2$ because in the latter case all loops are contractible, while in the former case there are non-contractible loops that circle the origin. It turns out that one can characterise the topology of a space by studying the kinds of functions that live on it. In particular, the functions that satisfy the check (1.17) but cannot be written as $\mathbf{F} = \nabla\phi$ with a continuous ϕ play a particularly important role, as they encode a lot of information about the topology of the underlying space.

2 Surfaces (and Volumes)

The main purpose of this chapter is to understand how to generalise the idea of an integral. Rather than integrating over a line, we will instead look at how to integrate over a 2d surface. We'll then see how to generalise to the integration over a 3d volume or, more generally, an n-dimensional space.

2.1 Multiple Integrals

We'll start by explaining what it means to integrate over a region in $\mathbb{R}^2$ or over a region in $\mathbb{R}^3$. The former are called area, or surface, integrals; the latter volume integrals. By the time we've understood volume integrals, the extension to $\mathbb{R}^n$ will be obvious.

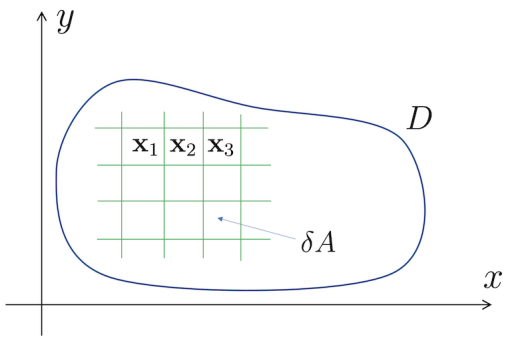
2.1.1 Area Integrals

Consider a region $D \subset \mathbb{R}^2$. Given a scalar function $\phi : \mathbb{R}^2 \to \mathbb{R}$, we want to find a way to integrate ϕ over D. We write this as

$$\int_D \phi(\mathbf{x})\, dA \tag{2.1}$$

You should think of the area element dA as representing an infinitesimally small area, with the $\int$ sign telling us that we're summing over many such small areas, in much the same was as $\int dx$ should be thought of as summing over infinitesimally small line elements dx. The area element is also written as $dA = dx\, dy$.

The rough idea underlying the integral is straightforward. First, we find a way to tesselate D with some simple shape, say a rectangle or other polygon. Each shape has common area δA. Admittedly, there might be some difficulty in making this work around the edge, but we'll ignore this for now. Then we might approximate the integral as

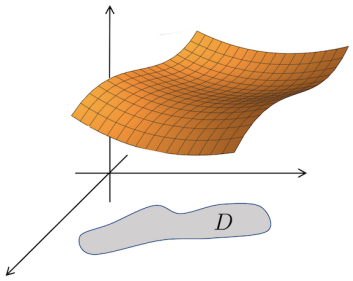


$$\int_D \phi(\mathbf{x})\, dA \;\approx\; \sum_n \phi(\mathbf{x}_n)\, \delta A$$

where $\mathbf{x}_n$ is a point in the middle of each shape. We can then consider making δA smaller and smaller, so that we tesselate the region D with finer and finer shapes. Intuitively, we might expect that as $\delta A \to 0$, the sum converges on an answer and, moreover, this answer is independent on any choices that we made along the way, such

as what shape we use and how we deal with the edges. When the limit exists – as it will for any sensible choice of function ϕ and region D – then it converges to the integral. If the function in the integrand is simply $\phi = 1$ then the integral (2.1) calculates the area of the region D.

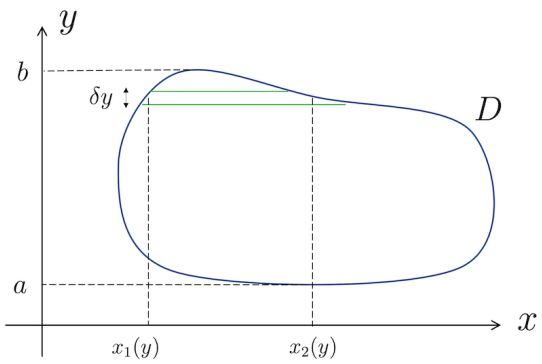
Just as an ordinary, one-dimensional integral can be viewed as the area under a curve, so too can an area integral be viewed as the volume under a function. This interpretation follows simply by plotting $z = \phi(x, y)$ in 3d as shown to the right.



Evaluating Area Integrals

In practice, we evaluate area integrals (or, indeed, higher dimensional integrals) by reducing them to multiple ordinary integrals.

There are a number of different ways to do this, and some may be more convenient than others, although all will give the same answer. For example, we could parcel our region D into narrow horizontal strips of width δy like so:

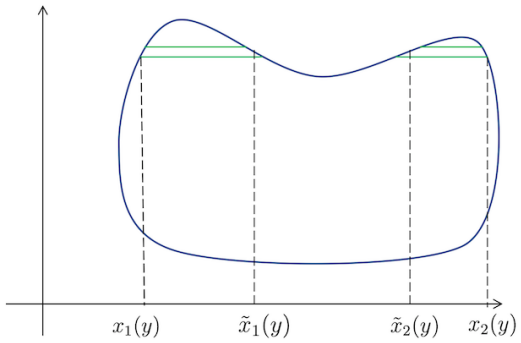


For each value of y, we then do the x integral between the two limits $x_1(y)$ and $x_2(y)$. We then subsequently sum over all such strips by doing the y integral between the two outer limits of the shape which we call a and b. The net result is

$$\int_D \phi(x, y)\, dA = \int_a^b dy \int_{x_1(y)}^{x_2(y)} dx\ \phi(x, y) \tag{2.2}$$

In this approach, the information about the shape D appears in the limits of the integral $x_1(y)$ and $x_2(y)$ which trace the outline of D as y changes.

If your shape is suitably annoying, then one or more of the integrals may have to be decomposed into disjoint sets. An example is shown on the right. In this case, for some values of y we need two further functions $\tilde{x}_1(y)$ and $\tilde{x}_2(y)$ to trace the outline of D and the integral in (2.2) is defined to be



$$\int_{x_1(y)}^{x_2(y)} dx = \int_{x_1(y)}^{\tilde{x}_1(y)} dx + \int_{\tilde{x}_2(y)}^{x_2(y)} dx$$

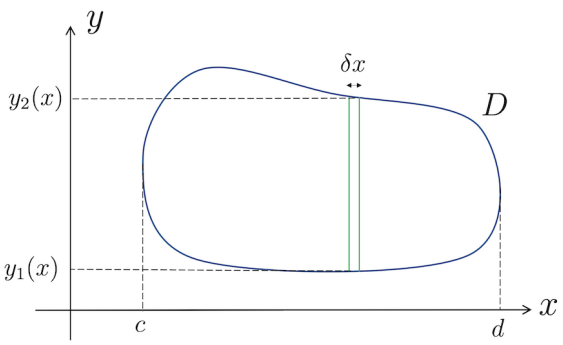
with the obvious generalisation if more disjoint intervals are needed.

We should pause at this point to make a comment on notation. You may be used to writing integrals as $\int$(integrand) dx, with the thing you're integrating sandwiched between the $\int$ sign and the dx. Indeed, that's the convention that we've been using up until now. But, as you progress through mathematics, there is a time to dump this notation and we have now reached that time. When performing multiple integrals, it becomes annoying to remember where you should place all those dx's and dy's, not least because they're not conveying any further information. So we instead write integrals as $\int dx$ (integrand), with the dx placed next to the integral sign. There's nothing deep in this. It's just a different convention, albeit one that holds your hand a little less. Think of it like that time you took the training wheels off your bike.

Our new notation does, however, retain the idea of ordering. You should work from right to left, first performing the $\int dx$ integration in (2.2) to get a function of y, and subsequently performing the $\int dy$ integration.

Note also that the number of $\int$ signs is not conserved in (2.2). On the left, $\int dA$ is an area integral and so requires us to do two normal integrals which are then written explicitly on the right. Shortly we will meet volume integrals and denote them as $\int dV$. Some texts prefer a convention in which there is a conservation of integral signs and so write area integrals as $\int\int dA$ and volume integrals as $\int\int\int dV$. The authors of these texts aren't string theorists and have never had to perform an integral in ten dimensions. Here we refuse to adopt this notation on the grounds that it looks silly.

There is a different way to do the integral (2.2). We could just as well divide our formula D into vertical strips of width δx, so that it looks like this:



For each value of x, we do the y integral between the two limits $y_1(x)$ and $y_2(x)$. As before, these functions trace the shape of the region D. We then subsequently sum over all strips by doing the x integral between the two outer limits of the shape which we now call c and d. Now the result is

$$\int_D \phi(x,y)\, dA = \int_c^d dx \int_{y_1(x)}^{y_2(x)} dy\ \phi(x,y) \tag{2.3}$$

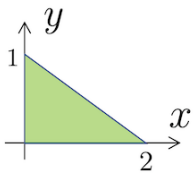
There are other ways to divide up the region D, some of which we will meet below when we discuss different coordinate choices. *Fubini's theorem*, proven in 1907, states that, for suitably well behaved functions $\phi(x,y)$ and regions D, all different ways of decomposing the integral agree. We won't prove this theorem here but it guarantees that the result that you get from doing the integrals in (2.2) coincides with the result from (2.3).

An Example

As a simple example, consider the function

$$\phi(x,y) = x^2 y$$

integrated over the triangle D shown in the figure.



We'll do the area integral in two different ways. If we first do the $\int dx$ integration, as in (2.2), then we have

$$\int_D \phi\, dA = \int_0^1 dy \int_0^{2-2y} dx\ x^2 y = \int_0^1 dy\ y \left[\frac{x^3}{3}\right]_0^{2-2y} = \frac{8}{3}\int_0^1 dy\ y(1-y)^3 = \frac{2}{15}$$

Meanwhile, doing the $\int dy$ integration first, as in (2.3), we have

$$\int_D \phi\, dA = \int_0^2 dx \int_0^{1-x/2} dy\ x^2 y = \int_0^2 dx\ x^2 \left[\frac{y^2}{2}\right]_0^{1-x/2} = \frac{1}{2}\int_0^2 dx\ x^2 \left(1-\frac{x}{2}\right)^2 = \frac{2}{15}$$

The two calculations give the same answer as advertised.

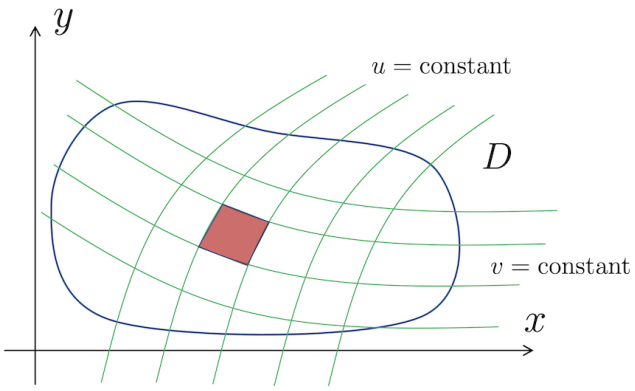


Figure 4. A change of coordinates from (x, y) to (u, v).

2.1.2 Changing Coordinates

Our discussion above was very much rooted in Cartesian coordinates. What if we choose to work with a different set of coordinates on $\mathbb{R}^2$?

Consider a change of variables $(x, y) \to (u, v)$. To be a good change of coordinates, the map should be smooth and invertible and we will assume that this is the case. The region D can then equally well be parameterised by coordinates (u, v). An example is shown in Figure 4, with lines of constant u and constant v plotted in green. We want to know how to do the area integral in the (u, v) coordinates.

Claim: The area integral can be written as

$$\int_D dx\, dy\ \phi(x, y) = \int_{D'} du\, dv\ |J(u, v)|\, \phi(u, v) \tag{2.4}$$

The region D in the (x, y) plane is mapped into a different region D' in the (u, v) plane. Here $\phi(u, v)$ is slightly sloppy shorthand: it means the function $\phi(x(u, v), y(u, v))$. The additional term $J(u, v)$ is called the *Jacobian* and is given by the determinant

$$J(u, v) = \begin{vmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{vmatrix}$$

The Jacobian is an important enough object that it also gets its own notation and is sometimes written as

$$J = \frac{\partial(x, y)}{\partial(u, v)}$$

Proof(ish): Here is a sketch of the proof to give you some intuition for why this is the right thing to do. We evaluate the integral by summing over small areas δA, formed by lines of constant u and v as shown by the red shaded region in Figure 4. The sides of this small region have length δu and δv respectively, but what is its area? It's not simply $\delta u\, \delta v$ because the sides aren't at necessarily right angles. Instead, the small shaded region is approximately a parallelogram.

We think of the original coordinates as functions of the new, so $x = x(u, v)$ and $y = y(u, v)$. If we make vary u and v slightly, then the change in the original x and y coordinates is

$$\delta x = \frac{\partial x}{\partial u}\delta u + \frac{\partial x}{\partial v}\delta v + \ldots \quad \text{and} \quad \delta y = \frac{\partial y}{\partial u}\delta u + \frac{\partial y}{\partial v}\delta v + \ldots$$

where the $+\ldots$ hide second order terms $\mathcal{O}(\delta u^2)$, $\mathcal{O}(\delta v^2)$ and $\mathcal{O}(\delta u\, \delta v)$. This means that we have

$$\begin{pmatrix} \delta x \\ \delta y \end{pmatrix} = \begin{pmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{pmatrix} \begin{pmatrix} \delta u \\ \delta v \end{pmatrix}$$

The small parallelogram is then spanned by the two vectors $\mathbf{a} = (\frac{\partial x}{\partial u}, \frac{\partial y}{\partial u})\delta u$ and $\mathbf{b} = (\frac{\partial x}{\partial v}, \frac{\partial y}{\partial v})\delta v$. Recall that the area of a parallelogram is $|\mathbf{a} \times \mathbf{b}|$, so we have

$$\delta A = \left|\frac{\partial(x, y)}{\partial(u, v)}\right| \delta u\, \delta v = |J|\, \delta u\, \delta v$$

which is the promised result □

An Example: 2d Polar Coordinates

There is one particular choice of coordinates that vies with Cartesian coordinates in their usefulness. This is plane polar coordinates, defined by

$$x = \rho \cos\phi \quad \text{and} \quad y = \rho \sin\phi$$

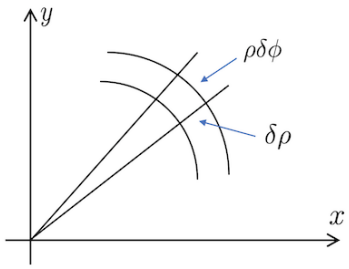
where the radial coordinate $\rho \geq 0$ and the angular coordinate takes values in $\phi \in [0, 2\pi)$. (Note: we used $\phi(x, y)$ to describe a general scalar field earlier in this section. This shouldn't be confused with the coordinate ϕ that we've introduced here.) We can easily compute the Jacobian to find

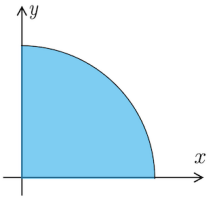
$$J = \frac{\partial(x, y)}{\partial(\rho, \theta)} = \begin{vmatrix} \cos\phi & -\rho\sin\phi \\ \sin\phi & \rho\cos\phi \end{vmatrix} = \rho$$

So we learn that the area element is given by

$$dA = \rho\, d\rho d\phi$$

There is also a simple graphical explanation of this result: it follows by looking at the area of the rounded square shape in the figure to the right (again, ignoring terms second order in $\delta\phi$ and $\delta\rho$).





Let's now use this to do an integral. Let D be a the wedge in the (x, y) plane defined by $x \geq 0$, $y \geq 0$ and $x^2 + y^2 \leq R^2$. This is shown to the left. In polar coordinates, this region is given by

$$0 \leq \rho \leq R \quad \text{and} \quad 0 \leq \phi \leq \frac{\pi}{2}$$

We'll integrate the function $f = e^{-(x^2+y^2)/2} = e^{-\rho^2/2}$ over the region D. In polar coordinates, we have

$$\int_D f\, dA = \int_0^{\pi/2} d\phi \int_0^R d\rho\ \rho e^{-\rho^2/2}$$

where the extra power of ρ in the integrand comes from the Jacobian. The $\int d\phi$ integral just gives us $\pi/2$, while the $\int d\rho$ integral is easily done. We have

$$\int_D f\, dA = \frac{\pi}{2}\Big[- e^{-\rho^2/2}\Big]_0^R = \frac{\pi}{2}\left(1 - e^{-R^2/2}\right)$$

As a final application, consider taking the limit $R \to \infty$, so that we're integrating over the quadrant $x, y \geq 0$. Clearly the answer is $\int_D f\, dA = \pi/2$. Back in Cartesian coordinates, this calculation becomes

$$\int_D f\, dA = \int_0^\infty dx \int_0^\infty dy\ e^{-(x^2+y^2)/2} = \left(\int_0^\infty dx\ e^{-x^2/2}\right)\left(\int_0^\infty dy\ e^{-y^2/2}\right)$$

Comparing to our previous result, we find the well-known expression for a Gaussian integral

$$\int_0^\infty dx\ e^{-x^2/2} = \sqrt{\frac{\pi}{2}}$$

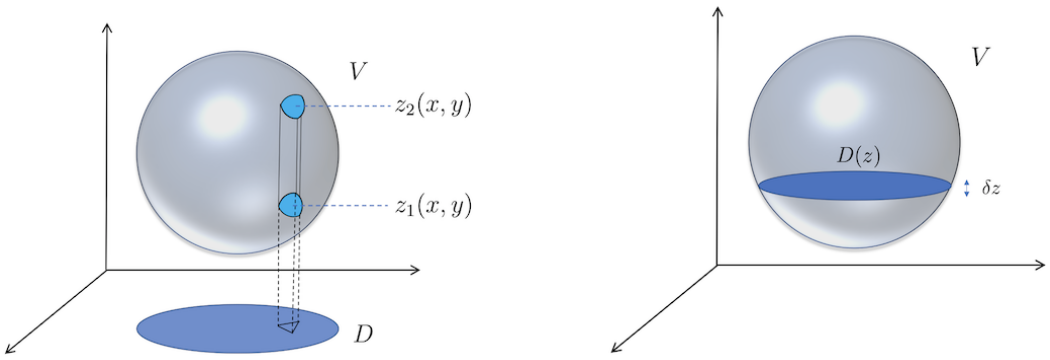


Figure 5. Two different ways to do a volume integral. On the left: perform the $\int dz$ integral first; on the right, perform the $\int_{D(z)} dA$ area integral first.

2.1.3 Volume Integrals

Most of this chapter will be devoted to discussing surfaces, but this is as good a place as any to introduce volume integrals because they are a straightforward generalisation of area integrals.

The basic idea should by now be familiar. The integration of a scalar function $\phi : \mathbb{R}^3 \to \mathbb{R}$ over a three-dimensional region V can be approximated by dividing the region into many small 3d pieces, each with volume δV and located at some position $\mathbf{x}_n$. You then find a way to take the limit

$$\int_V \phi(\mathbf{x})\, dV \;= \lim_{\delta V \to 0} \sum_n \phi(\mathbf{x}_n)\, \delta V$$

In practice, we evaluate volume integrals in the same way as we evaluate area integrals: by performing successive integrations. If we use Cartesian coordinates (x, y, z) we have a number of ways to proceed. For example, we could choose to first do the $\int dz$ integral, subsequently leaving us with an area integral over the (x, y) plane.

$$\int_V \phi(x,y,z)\, dV = \int dA \int_{z_1(x,y)}^{z_2(x,y)} dz\ \phi(x,y,z)$$

This approach is shown on the left-hand side of Figure 5. Alternatively, we could first do an area integral over some sliver of the region V and subsequently integrate over all slivers. This is illustrated on the right-hand side of Figure 5 and results in an integral of the form

$$\int_V \phi(x,y,z)\, dV = \int dz \int_{D(z)} dxdy\ \phi(x,y,z)$$

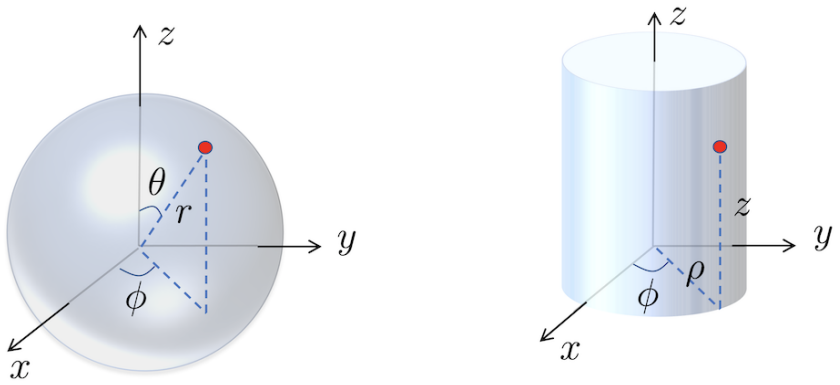


Figure 6. Spherical polar coordinates on the left, and cylindrical polar coordinates on the right.

As before, for suitably nice functions ϕ and regions V, the order of integration is unimportant.

There are many reasons to do a volume integral. You might, for example, want to know the volume of some object, in which case you just integrate the function $\phi = 1$. Alternatively, it's common to integrate a density of something, which means stuff per unit volume. Integrating the density over the region V tells you the amount of stuff in V. Examples of stuff that we will meet in other courses include mass, electric charge and probability.

2.1.4 Spherical Polar and Cylindrical Polar Coordinates

If your region V is some blocky shape, then Cartesian coordinates are probably the right way forward. However, for many applications it is more convenient to use a different choice of coordinates.

Given an invertible, smooth transformation $(x, y, z) \to (u, v, w)$ then the volume elements are mapped to

$$dV = dx\, dy\, dz = |J|\, du\, dv\, dw$$

with the Jacobian given by

$$J = \frac{\partial(x, y, z)}{\partial(u, v, w)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{vmatrix}$$

The sketch of the proof is identical to the 2d case: the volume of the appropriate parallelepiped is $\delta V = |J|\delta u \delta v \delta w$.

Two sets of coordinates are particularly useful. The first is *spherical polar coordinates*, related to Cartesian coordinates by the map

$$\begin{aligned} x &= r\sin\theta\,\cos\phi \\ y &= r\sin\theta\,\sin\phi \\ z &= r\cos\theta \end{aligned} \tag{2.5}$$

The range of the coordinates is $r \in [0,\infty)$, $\theta \in [0,\pi]$ and $\phi \in [0, 2\pi)$. The Jacobian is

$$\frac{\partial(x,y,z)}{\partial(u,v,w)} = r^2\sin\theta \quad \Rightarrow \quad dV = r^2\sin\theta\, dr\, d\theta\, d\phi \tag{2.6}$$

The second is *cylindrical polar coordinates*, which coincides with plane polar coordinates in the (x, y) plane, leaving z untouched

$$\begin{aligned} x &= \rho\cos\phi \\ y &= \rho\sin\phi \\ z &= z \end{aligned} \tag{2.7}$$

with $\rho \in [0,\infty)$ and $\phi \in [0, 2\pi)$ and, of course, $z \in (-\infty, +\infty)$. (Later in the course, we will sometimes denote the radial coordinate in cylindrical polar coordinates as r instead of ρ.) This time the Jacobian is

$$\frac{\partial(x,y,z)}{\partial(u,v,w)} = \rho \quad \Rightarrow \quad dV = \rho\, d\rho\, d\phi\, dz$$

We can do some dimensional analysis to check that these results make sense. In spherical polars we have one coordinate, r, with dimensions of length and two dimensionless angular coordinates. Correspondingly, the Jacobian has dimension length^2 to ensure that dV has the dimension of volume. In cylindrical polars, we have two coordinates with dimension of length, ρ and z, and just a single angular coordinate. This is the reason that the Jacobian now has dimension of length rather than length^2.

Example 1: The Volume of a Sphere

Consider a spherically symmetric function $f(r)$. We can integrate it over a ball of radius R using spherical polar coordinates, with $dV = r^2 \sin\theta \, dr d\theta d\phi$ to get

$$\begin{aligned}\int_V f \, dV &= \int_0^R dr \int_0^\pi d\theta \int_0^{2\pi} d\phi \; r^2 f(r) \sin\theta \\ &= 2\pi \Big[-\cos\theta \Big]_0^\pi \int_0^R dr \; r^2 f(r) \\ &= 4\pi \int_0^R dr \; r^2 f(r)\end{aligned}$$

In particular, if we take $f(r) = 1$ then we get the volume of a sphere $\text{Vol} = 4\pi R^3/3$.

Example 2: A Cylinder Cut Out of a Sphere

Next consider a more convoluted example: we want the volume of a sphere of radius R, with a cylinder of radius $s < R$ removed from the middle. The region V is then $x^2 + y^2 + z^2 \leq R^2$, together with $x^2 + y^2 \geq s^2$. Note that we don't just subtract the volume of a cylinder from the that of a sphere because the top of the cylinder isn't flat: it stops where it intersects the sphere.

In cylindrical coordinates, the region V spans $s \leq \rho \leq R$ and $-\sqrt{R^2 - \rho^2} \leq z \leq \sqrt{R^2 - \rho^2}$. And, of course, $0 \leq \phi < 2\pi$. We have $dV = \rho \, d\rho dz d\phi$ and

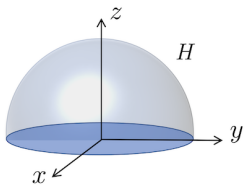
$$\text{Vol} = \int_V dV = \int_0^{2\pi} d\phi \int_s^R d\rho \; \rho \int_{-\sqrt{R^2-\rho^2}}^{\sqrt{R^2-\rho^2}} dz = 4\pi \int_s^R d\rho \; \rho \sqrt{R^2 - \rho^2}$$

It is now straightforward to do the integral to find the volume

$$\text{Vol} = \frac{4\pi}{3}(R^2 - s^2)^{3/2}$$

Example 3: Electric Charge On a Hemisphere

Consider a density of electric charge that increases linearly in the z-direction, with $f(z) = f_0 z/R$, in a hemisphere H of radius R, with $z \geq 0$ and f_0 a constant. What is the total charge in H?



In spherical polar coordinates, the coordinates for the hemisphere H are $0 \leq r \leq R$ and $0 \leq \phi < 2\pi$ and, finally, $0 \leq \theta \leq \pi/2$, which restricts us to the hemisphere with

$z \geq 0$. We integrate the function $f = f_0 r\cos\theta/R$ over H with $dV = r^2\sin\theta\, dr d\theta d\phi$ to find

$$\int_H f\, dV = \frac{f_0}{R}\int_0^{2\pi} d\phi \int_0^{\pi/2} d\theta \int_0^R dr\ r^2\sin\theta\, r\cos\theta = \frac{2\pi f_0}{R}\left[\frac{r^4}{4}\right]_0^R \left[\frac{1}{2}\sin^2\theta\right]_0^{\pi/2} = \frac{1}{4}\pi R^3 f_0$$

As a quick check on our answer, note that f_0 is the charge density so the dimensions of the final answer are correct: the total charge is equal to the charge density times a volume.

Vector Valued Integrals

We can also integrate vector valued fields $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$ over a volume V. There's nothing subtle here: we just do the integral component by component and the final answer is also a vector.

A common example arises when we compute the centre of mass. Let $\rho(\mathbf{x})$ be the density of an object. (Note that this isn't a great choice of notation if we're working in cylindrical polar coordinates.) The total mass is

$$M = \int_V \rho(\mathbf{x})\, dV$$

and the *centre of mass* is

$$\mathbf{X} = \frac{1}{M}\int_V \mathbf{x}\rho(\mathbf{x})\, dV$$

For example, consider again the solid hemisphere H from the previous example, covering $0 \leq r \leq R$ and $z \geq 0$. We'll take this object to have constant density ρ. The total mass is

$$M = \int_H \rho\, dV = \frac{2\pi}{3}\rho R^3$$

Writing $\mathbf{X} = (X, Y, Z)$ for the centre of mass, we need to compute the three components individually. We have

$$\begin{aligned} X = \frac{\rho}{M}\int_H x\, dV &= \frac{\rho}{M}\int_0^{2\pi} d\phi \int_0^R dr \int_0^{\pi/2} d\theta\ xr^2\sin\theta \\ &= \frac{\rho}{M}\int_0^{2\pi} d\phi \int_0^R dr \int_0^{\pi/2} d\theta\ r^3\sin^2\theta\cos\phi = 0 \end{aligned}$$

where the integral $\int d\phi \cos\phi = 0$. A similar calculation shows that $Y = 0$. Indeed, the fact that the centre of mass lies at $(X, Y) = (0, 0)$ follows on symmetry grounds. We're left only computing the centre of mass in the z-direction. This is

$$Z = \frac{\rho}{M}\int_H z\, dV = \frac{\rho}{M}\int_0^{2\pi} d\phi \int_0^R dr \int_0^{\pi/2} d\theta\ r^3 \cos\theta \sin\theta = \frac{3R}{8}$$

We learn that the centre of mass sits at $\mathbf{X} = (0, 0, 3R/8)$.

Generalisation to $\mathbb{R}^n$

Finally, it is straightforward to generalise multiple integrals to $\mathbb{R}^n$. If we make a smooth, invertible change of coordinates from Cartesian $x^1, \ldots, x^n$ to some other coordinates $u^1, \ldots, u^n$ then the integral over some n-dimensional region M is

$$\int_M f(x^i)\, dx^1\ \ldots dx^n = \int_{M'} f(x(u^i))\, |J|\, du^1 \ldots du^n$$

where the Jacobian

$$J = \frac{\partial(x^1, \ldots, x^n)}{\partial(u^1, \ldots, u^n)} = \det\left(\frac{\partial x^i}{\partial u^a}\right)$$

is the obvious generalisation of our previous results.

2.2 Surface Integrals

Our next task is to understand how to integrate over a surface that doesn't lie flat in $\mathbb{R}^2$, but is instead curved in some way in $\mathbb{R}^3$. We will start by looking at how we define such surfaces in the first place.

2.2.1 Surfaces

There are (at least) two different ways to describe a surface in $\mathbb{R}^3$.

- A surface can be viewed as the level set of a function,

  $$F(x, y, z) = 0$$

  This is one condition on three variables, so results in a two-dimensional surface in $\mathbb{R}^3$. (In general, a single constraint like this results in an $(n-1)$-dimensional space in $\mathbb{R}^n$. Alternatively we say that the space has *codimension* one.)

- We can consider a *parameterised surface*, defined by the map

  $$\mathbf{x} : \mathbb{R}^2 \to \mathbb{R}^3$$

  This is the extension of the parameterised curve that we discussed in Section 1. This now defines a dimension two surface in any space $\mathbb{R}^3$.

At each point on the surface, we can define a *normal vector*, $\mathbf{n}$ which points perpendicularly away from the surface. When the surface is defined as the level set of function $F(\mathbf{x}) = 0$, the normal vector lies in the direction

$$\mathbf{n} \sim \nabla F$$

To see this, note that $\mathbf{m} \cdot \nabla F$ describes the rate of change of F in the direction $\mathbf{m}$. If $\mathbf{m}$ lies tangent to the surface then we have, by definition, $\mathbf{m} \cdot \nabla F = 0$. Conversely, the normal to the surface $\mathbf{n}$ lies in the direction in which the function F changes most quickly, and this is ∇F.

It's traditional to normalise the normal vector, so we usually define

$$\mathbf{n} = \pm \frac{1}{|\nabla F|} \nabla F$$

where we'll say more about the choice of minus sign below.

Meanwhile, for the parameterised surface $\mathbf{x}(u, v) \in \mathbb{R}^3$, we can construct two tangent vectors to the surface, namely

$$\frac{\partial \mathbf{x}}{\partial u} \quad \text{and} \quad \frac{\partial \mathbf{x}}{\partial v}$$

where each partial derivative is taken holding the other coordinate fixed. Each of these lies within the surface, so the normal direction is

$$\mathbf{n} \sim \frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v}$$

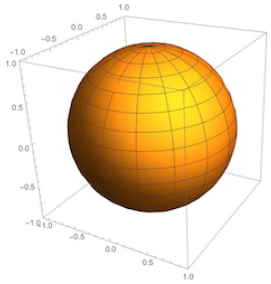
If $\mathbf{n} \neq 0$ anywhere on the surface then the parameterisation is said to be *regular*. Note that, although a parameterised surface can be defined in any $\mathbb{R}^n$, the normal direction is only unique in $\mathbb{R}^3$ where we have the cross product at our disposal.

Examples

Here are a number of examples using the definition involving a level set. A sphere of radius R is defined by

$$F(x, y, z) = x^2 + y^2 + z^2 - R^2 = 0$$

the normal direction is given by $\nabla F = 2(x, y, z)$ and points radially outwards.



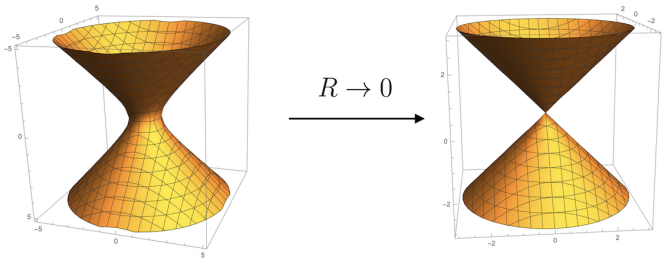


Figure 7. When the hyperboloid degenerates to a cone, there is no well defined normal at the origin.

A hyperboloid is defined by

$$F(x, y, z) = x^2 + y^2 - z^2 - R^2 = 0$$

with normal direction given by $\nabla F = 2(x, y, -z)$. Note that for both the sphere and hyperboloid, the normal vector is nowhere vanishing because the origin $\mathbf{x} = 0$ doesn't lie on the surface. However, if we take the limit $R \to 0$ then the hyperboloid degenerates to two cones, meeting at the origin. In this case, $\nabla F = 0$ at the origin, reflecting the fact there is no unique direction away from the surface at this point. This is shown in Figure 7

2.2.2 Surfaces with Boundaries

A surface S can have a boundary. This is a piecewise smooth closed curve. If there are several boundaries, then this curve should be thought of as having several disconnected pieces.

For example, we could define the surfaces above now restricted to the region $z \geq 0$. In this case both the sphere and hyperboloid are truncated and their boundary is the circle $x^2 + y^2 = R^2$ in the $z = 0$ plane.

The boundary of a surface S is denoted ∂S with ∂ the standard notation to denote the boundary of any object. For example, later in the lectures we will denote the boundary of a 3d volume V as ∂V. You might reasonably wonder why we use the partial derivative symbol ∂ to denote the boundary of something. There are some deep and beautiful reasons behind this that will only become apparent in later courses. But there is also a simple, intuitive reason. Consider a collection of 3d objects, all the same shape but each bigger than the last. We'll denote these volumes as V_r. Then, roughly

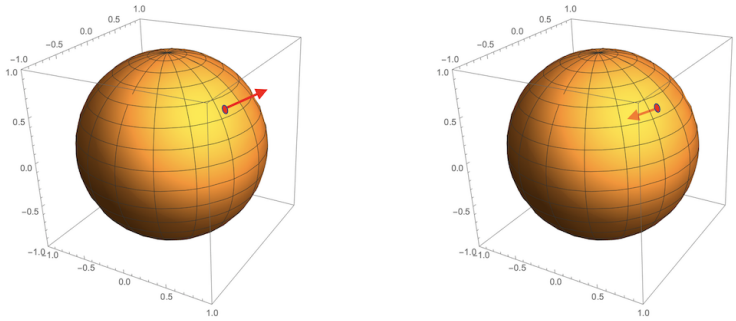


Figure 8. Two orientations of a sphere, with the unit normal pointing outwards or inwards.

speaking, you can view the boundary surface as

$$\partial V_r = \lim_{\epsilon \to 0} \frac{1}{\epsilon}\Big(V_{r+\epsilon} \setminus V_r\Big)$$

where $\setminus$ means that you remove the 3d object V_r from inside the slightly larger object $V_{r+\epsilon}$. This, of course, looks very much like the formula for a derivative.

This "derivative equals boundary" idea also shows up when we calculate volumes, areas and lengths. For example, a disc of radius r has area πr^2. The length of the boundary is $\frac{d}{dr}(\pi r^2) = 2\pi r$. This relation continues to higher dimensional balls and spheres.

There is something important lurking in the idea of a boundary. The boundary is necessarily a *closed* curve C, meaning that it has no end points. Another way of saying this is that a closed curve C itself has no boundary, or $\partial C = 0$. We see that if a curve arises as the boundary of a surface, then the curve itself has no boundary. This is captured in the slogan "the boundary of a boundary vanishes" or, in equation form, $\partial^2 S = 0$. It is a general and powerful principle that extends to higher dimensional objects where $\partial^2(\text{anything}) = 0$. The idea that the boundary of a boundary vanishes is usually expressed simply as $\partial^2 = 0$.

A couple of quick definitions. A surface is said to be *bounded* if it doesn't stretch off to infinity. More precisely, a bounded surface can be contained within some solid sphere of fixed radius. A surface that does stretch off to infinity is said to be *unbounded*. Obviously, the sphere is a bounded surface, while the hyperboloid is unbounded.

Finally, a bounded surface with no boundary is said to be *closed*.

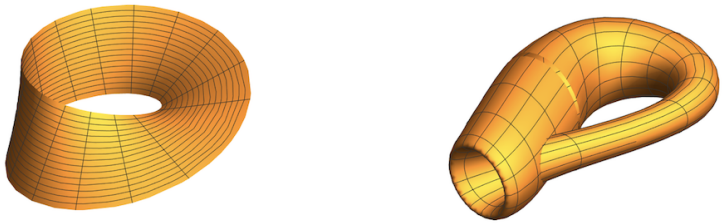


Figure 9. Two unorientable surfaces: the Möbius strip on the left, and the Klein bottle on the right.

2.2.3 Orientability

As long as the normal vector $\mathbf{n} \neq 0$, we can always normalise it so that it has unit length. But in general, there is no canonical way to fix the sign. This is a matter of convention and determines what we mean by "outside the surface" and what we mean by "inside".

A surface is said to be *orientable* if there is a consistent choice of unit normal $\mathbf{n}$ which varies smoothly over the surface. The sphere and hyperboloid above are both orientable, with the two choices of an orientation for the sphere shown in Figure 8. Throughout these lectures we will work only with orientable surfaces. For such surfaces, a choice of sign fixes the unit normal everywhere and is said to determine the *orientation* of the surface.

We note in passing that unorientable surfaces exist. You can easily make one of these in the comfort of your own home. Take a strip of paper and glue the two ends together. You've got two different ways to glue them as shown by the arrows below:

If you glue by aligning the arrows shown on the left, then you're just left with a boring strip of paper. But if you glue with the arrows aligned on the right, then you end up with something new and exciting: an unorientable surface, known as a *Möbius strip*. You can see one that I made earlier on the left of Figure 9. If you pick a normal vector and evolve it smoothly around the strip then you'll find that it comes back pointing in the other direction. Relatedly, the Möbius strip has a single boundary, rather than two.

We can also make closed, unorientable surfaces using a similar construction. This time we glue together two edges, like so

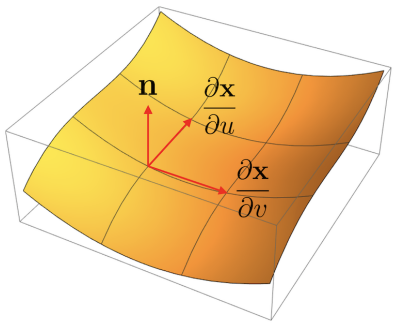
Again, you have various choices. If you glue with the arrows aligned as shown on the left, then you'll end up with a torus which is very much orientable. If you glue with the arrows aligned as shown on the right then you get an unorientable surface called a *Klein bottle*. It's a little tricky to draw embedded in 3d space (and, indeed, tricky to make with paper and glue) as it appears to intersect itself, but an attempt is shown on the right of Figure 9.

2.2.4 Scalar Fields

We're now in a position to start integrating objects over surfaces. For this, we work with parameterised surfaces $\mathbf{x}(u, v)$.

Sit at some point (u, v) on the surface, and move in both directions by some small amount δu and δv. This defines an approximate parallelogram on the surface, as shown in the figure. The area of this parallelogram is

$$\delta S = \left| \frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} \right| \delta u \, \delta v$$



where, as usual, we've dropped higher order terms. This is called the *scalar area*. (We'll see the need for the adjective "scalar" below when we introduce a variant known as the vector area.)

Now we're in a position to define the surface integral of a scalar field $\phi(\mathbf{x})$. Given a parameterised surface S, the surface integral is given by

$$\int_S \phi(\mathbf{x}) \, dS = \int_D du dv \, \left| \frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} \right| \phi(\mathbf{x}(u, v)) \tag{2.8}$$

where D is the appropriate region in the (u, v) plane. This is now the same kind of area integral that we learned to do in Section 2.1.

The area integral of a scalar field does not depend on the orientation of the surface. It doesn't matter what you choose as the inside of the surface S and what you choose as the outside, the integral of a scalar field over S always gives the same answer. In particular, if we integrate $\phi = 1$ over a surface S then we get the area of that surface, and this is always positive. This is entirely analogous to the line integral of a scalar field that we met in Section 1.2 that was independent of the orientation of the curve.

Reparameterisation Invariance

Importantly, the surface integral (2.8) is independent of the choice of parameterisation of the surface. To see this, suppose that we replace our original parameterisation $\mathbf{x}(u, v)$ with an alternative parameterisation $\mathbf{x}(\tilde{u}, \tilde{v})$, both of which are assumed to be regular. We then have

$$\frac{\partial \mathbf{x}}{\partial u} = \frac{\partial \mathbf{x}}{\partial \tilde{u}}\frac{\partial \tilde{u}}{\partial u} + \frac{\partial \mathbf{x}}{\partial \tilde{v}}\frac{\partial \tilde{v}}{\partial u} \quad \text{and} \quad \frac{\partial \mathbf{x}}{\partial v} = \frac{\partial \mathbf{x}}{\partial \tilde{u}}\frac{\partial \tilde{u}}{\partial v} + \frac{\partial \mathbf{x}}{\partial \tilde{v}}\frac{\partial \tilde{v}}{\partial v}$$

Taking the cross-product, we have

$$\frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} = \frac{\partial(\tilde{u}, \tilde{v})}{\partial(u, v)} \frac{\partial \mathbf{x}}{\partial \tilde{u}} \times \frac{\partial \mathbf{x}}{\partial \tilde{v}}$$

This means that the scalar area element can equally well be written as

$$dS = \left| \frac{\partial \mathbf{x}}{\partial \tilde{u}} \times \frac{\partial \mathbf{x}}{\partial \tilde{v}} \right| d\tilde{u}\, d\tilde{v}$$

where we've used the result (2.4) which, in the current context, is $d\tilde{u}\, d\tilde{v} = \frac{\partial(\tilde{u},\tilde{v})}{\partial(u,v)}\, du\, dv$.

The essence of this calculation is the same as we saw for line integrals: the two derivatives $\partial/\partial u$ and $\partial/\partial v$ in the integrand cancel the Jacobian factor under a change of variables. The upshot is that we can write the surface integral (2.8) using any parameterisation that we wish: the answer will be the same.

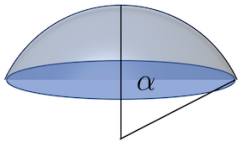
An Example

Consider a sphere of radius R. Let S be the subregion that sits at an angle $\theta \leq \alpha$ from the vertical. This is the grey region shown in the figure. We want to compute the area of this cap.

We start by constructing a parameterisation of the sphere. This is straightforward if we use the spherical polar angles θ and ϕ defined in (2.5) as parameters. We have

$$\mathbf{x}(\theta, \phi) = R(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta) := R\,\mathbf{e}_r$$

Here $\mathbf{e}_r$ is the unit vector that points radially outwards. (We will also use the notation $\mathbf{e}_r = \hat{\mathbf{r}}$ later in these lectures.) We can then easily calculate



$$\frac{\partial \mathbf{x}}{\partial \theta} = R(\cos\theta\cos\phi, \cos\theta\sin\phi, -\sin\theta) := R\,\mathbf{e}_\theta$$
$$\frac{\partial \mathbf{x}}{\partial \phi} = R(-\sin\theta\sin\phi, \sin\theta\cos\phi, 0) := R\sin\theta\,\mathbf{e}_\phi$$

Here, by construction, $\mathbf{e}_\theta$ and $\mathbf{e}_\phi$ are unit vectors pointing in the direction of increasing θ and ϕ respectively. We'll have more to say about the triplet of vectors $\mathbf{e}_r$, $\mathbf{e}_\theta$ and $\mathbf{e}_\phi$ in Section 3.3. For now, we can compute

$$\frac{\partial \mathbf{x}}{\partial \theta} \times \frac{\partial \mathbf{x}}{\partial \phi} = R^2 \sin\theta\,\mathbf{e}_r$$

From this, we have the scalar area element

$$dS = R^2 \sin\theta\, d\theta\, d\phi \tag{2.9}$$

We've seen a result very similar to this before. The volume element in spherical polar coordinates (2.6) is $dV = r^2\sin\theta\, dr\, d\theta\, d\phi$. Our area element over a sphere simply comes from setting $r = R$ and ignoring the dr piece of the volume element.

It is now straightforward to compute the area. We have

$$A = \int_0^{2\pi} d\phi \int_0^{\alpha} d\theta\ R^2 \sin\theta = 2\pi R^2(1 - \cos\alpha)$$

Note that if we set $\alpha = \pi$ then we get the area of a full sphere: $A = 4\pi R^2$.

2.2.5 Vector Fields and Flux

Now we turn to vector fields. There is a particularly interesting and useful way to integrate a vector field $\mathbf{F}(\mathbf{x})$ over a surface S so that we end up with a number. We do this by taking the inner product of the vector field with the normal to the surface, $\mathbf{n}$, so that

$$\int_S \mathbf{F}(\mathbf{x})\cdot\mathbf{n}\, dS = \int_D du dv\ \left(\frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v}\right) \cdot \mathbf{F}(\mathbf{x}(u,v)) \tag{2.10}$$

This is called the *flux* of $\mathbf{F}$ through S.

The definition of the flux is independent of our choice of parameterisation: the argument is identical to the one we saw above for a scalar field.

It's convenient to introduce some new notation. The *vector area element* is defined as

$$d\mathbf{S} = \mathbf{n}\, dS = \left(\frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} \right) du\, dv$$

This has magnitude dS and points in the normal direction $\mathbf{n}$.

The flux of a vector field depends on the orientation of the surface S. This can be seen in the presence of the normal vector in (2.10). In the parameterised surface $\mathbf{x}(u, v)$, the choice of orientation can be traced to the parameterisation (u, v) and, in particular, the order in which they appear in the cross product. Changing the orientation of the surface flips the sign of the flux.

The physical importance of the flux can be seen by thinking about a fluid. Let $\mathbf{F}(\mathbf{x})$ be the *velocity field* of a fluid. (Usually we would denote this as $\mathbf{u}(\mathbf{x})$ or $\mathbf{v}(\mathbf{x})$, but we've already used u and v as the parameters of the surface so we'll adopt the non-standard name $\mathbf{F}$ for the velocity to avoid confusion.) In a small time δt, the amount of fluid flowing through a small surface element δS is given by

$$\text{Fluid Flow} = \mathbf{F}\, \delta t \cdot \mathbf{n}\, \delta S$$

where the dot product ensures that we don't include the component of fluid that flows parallel to the surface. Integrating over the whole surface, we see that the flux of fluid

$$\text{Flux} = \int_S \mathbf{F} \cdot d\mathbf{S}$$

is the amount of fluid crossing S *per unit time*. In other words, the flux is the rate of fluid flow.

We also talk of "flux" in other contexts, where there's no underlying flow. For example, in our course on Electromagnetism, we will spend some time computing the flux of the electric field through various surfaces, $\int_S \mathbf{E} \cdot d\mathbf{S}$.

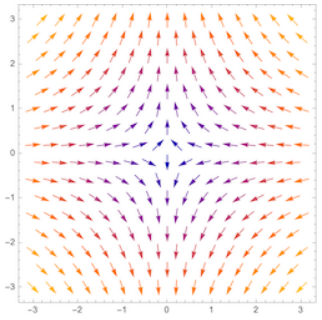
An Example

Consider the vector field

$$\mathbf{F} = (-x, 0, z)$$

This is plotted in the $y =$ constant plane in the figure.

We want to integrate this vector field over the hemispherical cap, subtended by the angle α that we used as an example in Section 2.2.4. This is the region of a sphere of radius R, spanned by the polar coordinates



$$0 \leq \theta \leq \alpha \quad \text{and} \quad 0 \leq \phi < 2\pi$$

We know from our previous work that

$$d\mathbf{S} = R^2 \sin\theta\, \mathbf{e}_r\, d\theta\, d\phi \quad \text{with} \quad \mathbf{e}_r = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$$

In particular, we have

$$\mathbf{F}\cdot\mathbf{e}_r = -x\sin\theta\cos\phi + z\cos\theta = R(-\sin^2\theta\cos^2\phi + \cos^2\theta)$$

The flux through the hemispherical cap is then

$$\int \mathbf{F}\cdot d\mathbf{S} = \int_0^\alpha d\theta \int_0^{2\pi} d\phi\ R^3\sin\theta\left[-\sin^2\theta\cos^2\phi + \cos^2\theta\right]$$

We use $\int_0^{2\pi} d\phi\ \cos^2\phi = \pi$ to get

$$\begin{aligned}\int \mathbf{F}\cdot d\mathbf{S} &= \pi R^3 \int_0^\alpha d\theta\ \sin\theta\left[-\sin^2\theta + 2\cos^2\theta\right] \\ &= \pi R^3\Big[\cos\theta\sin^2\theta\Big]_0^\alpha = \pi R^3\cos\alpha\sin^2\alpha \end{aligned} \tag{2.11}$$

2.2.6 A Sniff of the Gauss-Bonnet Theorem

The methods described in this section have many interesting applications to geometry. Here we sketch two important ideas. We prove neither.

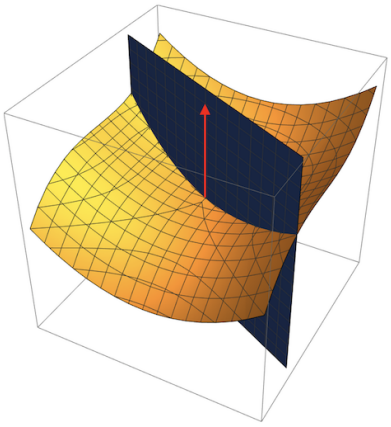
Consider a surface S and pick a point with normal $\mathbf{n}$. We can construct a plane containing $\mathbf{n}$, as shown in the figure. The intersection of the original surface and the plane describes a curve C that lies in S. Associated to this curve is a curvature κ, defined in (1.6).

Now, we rotate the plane about $\mathbf{n}$. As we do so, the curve C changes and so too does the curvature. Of particular interest are the maximum and minimum curvatures

$$\kappa_{\min} \leq \kappa \leq \kappa_{\max}$$

These are referred to as *principal curvatures*. The *Gaussian curvature* of the surface S at our chosen point is then defined to be

$$K = \kappa_{\min}\kappa_{\max}$$



As defined, the curvature K would appear to have as much to do with the embedding of the surface in $\mathbb{R}^3$ as the surface itself. The *theorem egregium* (or "remarkable" theorem) due to Gauss, is the statement that this is misleading: the curvature K is a property of the surface alone, irrespective of any choice of embedding. We say that K is *intrinsic* to the surface.

The idea that curved surfaces have a life of their own, independent of their embedding, is an important one. It generalises to higher dimensional spaces, known as *manifolds*, which are the subject of differential geometry. In physics, curved space (or, more precisely, curved spacetime) provides the framework for our understanding of gravity. Both Riemannian geometry and its application to gravity will be covered in lectures on General Relativity.

The Gaussian curvature K has a number of interesting properties. Here's one. Consider a *geodesic triangle* drawn on the surface as shown in Figure 10. This means that we connect three points with *geodesics*, which are lines of the shortest distance as measured using the arc length (1.5). Let θ_1, θ_2 and θ_3 be the interior angles of the triangle, defined by the inner product of tangent vectors of the geodesic curves. Then it turns out that

$$\theta_1 + \theta_2 + \theta_3 = \pi + \int_D K\, dS \tag{2.12}$$

where D is the interior region of the triangle. If the triangle is drawn on flat $\mathbb{R}^2$, then $K = 0$ and this theorem reduces to the well known statement that the angles of a triangle add up to π.

We can check this formula for the simple case of a triangle drawn on a sphere. If the sphere has radius R then the geodesics are great circles and, as we saw in Section 1.1, they all have curvature $\kappa = 1/R$. Correspondingly, the Gaussian curvature for a sphere is $K = 1/R^2$. A geodesic triangle is shown in the figure to the below: it has two right-angles $\pi/2$ sitting at the equator, and an angle α at the top.

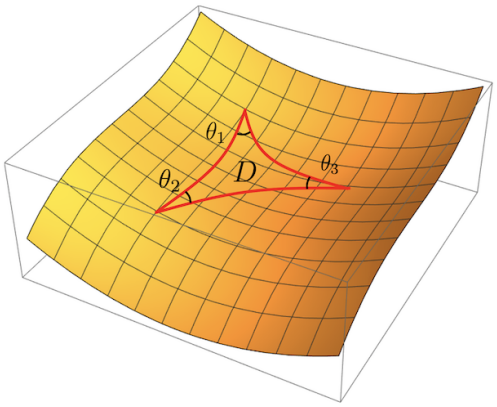
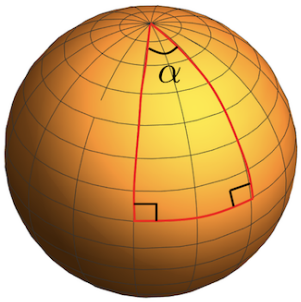


Figure 10. A geodesic triangle inscribed on a surface.

The area of the region inside the triangle is $A = \alpha R^2$ (so that $A = 2\pi R^2$ when $\alpha = 2\pi$ which is the area of the upper hemisphere.). We then have



$$\int_D K\, dS = \frac{A}{R^2} = \alpha$$

which agrees with the result (2.12).

Here's another beautiful application of the Gaussian curvature. Consider a closed surface S. Any such surface can be characterised by the number of holes that it has. This number of holes is known as the *genus*. Three examples are given in Figure 11: a sphere with $g = 0$, a torus with $g = 1$ and some kind of baked-good with genus $g = 3$. It turns out that if you integrate the Gaussian curvature over the entire surface then you get

$$\int_S K\, dS = 4\pi(1 - g) \tag{2.13}$$

This result is all kinds of wonderful. The genus g tells us about the topology of the surface. It's a number that only makes sense when you stand back and look at the object as a whole. In contrast, the Gaussian curvature is a locally defined object: at any given point it depends only on the neighbourhood of that point. But this result tells us that integrating something local can result in something global.

The round sphere provides a particularly simple example of this result. As we've seen above, the Gaussian curvature is $K = 1/R^2$ which, when integrated over the

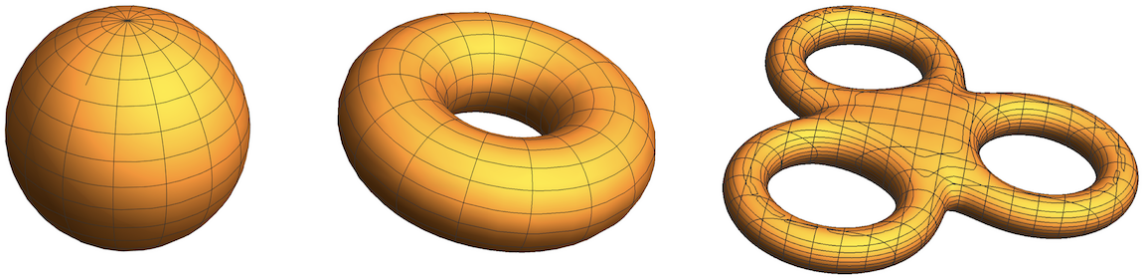


Figure 11. Three closed surfaces with different topologies. The sphere has genus $g = 0$, the torus has genus $g = 1$ and the surface on the right has $g = 3$.

whole sphere, does indeed give 4π as befits a surface of genus $g = 0$. However, this simple calculation hides the magic of the formula (2.13). Suppose that we start to deform the sphere. We might choose to pull it out in some places, push it inwards in others. We could try to mould some likeness of our face in some part of it. Everything that we do changes the local Gaussian curvature. It will increase in some parts and decrease in others. But the formula (2.13) tells us that this must, at the end of the day, cancel out. As long as we don't tear the surface, so its topology remains that of a sphere, the integral of K will always give 4π.

The results (2.12) and (2.13) are two sides of the wondrous Gauss-Bonnet theorem. A proof of this theorem will have to wait for later courses. (You can find a somewhat unconventional proof using methods from physics in the lectures on Supersymmetric Quantum Mechanics. This proof also works for a more powerful generalisation to higher dimensional spaces, known as the Chern-Gauss-Bonnet theorem.)

3 Grad, Div and Curl

In this section we're going to further develop the ways in which we can differentiate. We'll be particularly interested in how we can differentiate scalar and vector fields. Our definitions will be straightforward but, at least for the time being, we won't be able to offer the full intuition behind these ideas. Perhaps ironically, the full meaning of how to differentiate will become clear only in Section 4 where we also learn the corresponding different ways to integrate.

3.1 The Gradient

We've already seen how to differentiate a scalar field $\phi : \mathbb{R}^n \to \mathbb{R}$. Given Cartesian coordinates x^i with $i = 1, \ldots, n$ on $\mathbb{R}^n$, the gradient of ϕ is defined as

$$\nabla\phi = \frac{\partial\phi}{\partial x^i}\,\mathbf{e}_i \tag{3.1}$$

Note that differentiating a scalar field leaves us with a vector field.

The definition above relies on a choice of Cartesian coordinates. Later in this section, we'll find expressions for the gradient in different coordinate systems. But there is also a definition of the gradient that does not rely on any coordinate choice at all. This starts by considering a point $\mathbf{x} \in \mathbb{R}^n$. We don't, yet, think of $\mathbf{x}$ as defined by a string of n numbers: that comes only with a choice of coordinates. Instead, it should be viewed as an abstract point in $\mathbb{R}^n$.

The first principles, coordinate-free definition of the gradient $\nabla\phi$ simply compares the value of ϕ at some point $\mathbf{x}$ to the value at some neighbouring point $\mathbf{x} + \mathbf{h}$ with $h = |\mathbf{h}| \ll 1$. For a differentiable function ϕ, we can write

$$\phi(\mathbf{x} + \mathbf{h}) = \phi(\mathbf{x}) + \mathbf{h} \cdot \nabla\phi + \mathcal{O}(h^2) \tag{3.2}$$

where this should be thought of as the definition of the gradient $\nabla\phi$. Note that it's similar in spirit to our definition of the tangent to a curve $\dot{\mathbf{x}}$ given in (1.2). If we pick a choice of coordinates, with $\mathbf{x} = (x^1, \ldots, x^n)$, then we can take $\mathbf{h} = \epsilon\,\mathbf{e}_i$ with $\epsilon \ll 1$. The definition (3.2) then coincides with (3.1),

An Example

Consider the function on $\mathbb{R}^3$,

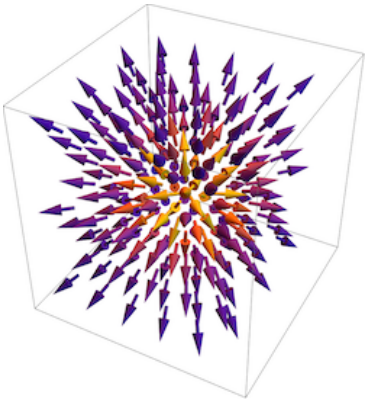
$$\phi(x, y, z) = -\frac{1}{\sqrt{x^2 + y^2 + z^2}} = -\frac{1}{r}$$

where $r^2 = x^2 + y^2 + z^2$ is the distance from the origin. We have

$$\frac{\partial \phi}{\partial x} = \frac{x}{(x^2+y^2+z^2)^{3/2}} = \frac{x}{r^3}$$

and similar for the others. The gradient is then given by

$$\nabla \phi = \frac{x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}}{r^3} = \frac{\hat{\mathbf{r}}}{r^2}$$



where, in the final expression, we've introduced the unit vector $\hat{\mathbf{r}}$ which points out radially outwards in each direction, like the spikes on a hedgehog as shown in the figure. The vector field $\nabla\phi$ points radially, decreasing as $1/r^2$. Vector fields of this kind are important in electromagnetism where they describe the electric field $\mathbf{E}(\mathbf{x})$ arising from a charged particle.

An Application: Following a Curve

Suppose that we're given a curve in $\mathbb{R}^n$, defined by the map $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$, together with a scalar field $\phi : \mathbb{R}^n \to \mathbb{R}$. Then we can combine these into the composite map $\phi(\mathbf{x}(t)) : \mathbb{R} \to \mathbb{R}$. This is simply the value of the scalar field evaluated on the curve. We can then differentiate this map along the curve using the higher dimensional version of the chain rule.

$$\frac{d\phi(\mathbf{x}(t))}{dt} = \frac{\partial \phi}{\partial x^i}\frac{dx^i}{dt}$$

This has a nice, compact expression in terms of the gradient,

$$\frac{d\phi(\mathbf{x}(t))}{dt} = \nabla\phi \cdot \frac{d\mathbf{x}}{dt}$$

This tells us how the function $\phi(\mathbf{x})$ changes as we move along the curve.

3.2 Div and Curl

At this stage we take an interesting and bold mathematical step. We view ∇ as an object in its own right. It is called the *gradient operator*.

$$\nabla = \mathbf{e}_i \frac{\partial}{\partial x^i} \tag{3.3}$$

This is both a vector and an operator. The fact that ∇ is an operator means that it's just waiting for a function to come along (from the right) and be differentiated.

The gradient operator ∇ sometimes goes by the names *nabla* or *del*, although usually only when explaining to students in a first course on vector calculus that ∇ sometimes goes by the names *nabla* or *del*. (Admittedly, the latex command for ∇ is \nabla which helps keep the name alive.)

With ∇ divorced from the scalar field on which it originally acted, we can now think creatively about how it may act on other fields. As we've seen, a vector field is defined to be a map

$$\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^n$$

Given two vectors, we all have a natural urge to dot them together. This gives a derivative acting on vector fields known as the *divergence*

$$\nabla \cdot \mathbf{F} = \left(\mathbf{e}_i \frac{\partial}{\partial x^i}\right) \cdot (\mathbf{e}_j F_j) = \frac{\partial F_i}{\partial x^i}$$

where we've used the orthonormality $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$. Note that the gradient of a scalar field gave a vector field. Now the divergence of a vector field gives a scalar field.

The divergence isn't the only way to differentiate a vector field. If we're in $\mathbb{R}^n$, a vector field has N components and we could differentiate each of these in one of N different directions. This means that there are N^2 different meanings to the "derivative of a vector field". But the divergence turns out to be the combination that is most useful.

Both the gradient and divergence operations can be applied to fields in $\mathbb{R}^n$. In contrast, our final operation holds only for vector fields that map

$$\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$$

In this case, we can take the cross product. This gives a derivative of a vector field known as the *curl*,

$$\nabla \times \mathbf{F} = \left(\mathbf{e}_i \frac{\partial}{\partial x^i}\right) \times (\mathbf{e}_j F_j) = \epsilon_{ijk} \frac{\partial F_j}{\partial x^i} \mathbf{e}_k$$

Or, written out in its full glory,

$$\nabla \times \mathbf{F} = \left(\frac{\partial F_3}{\partial x^2} - \frac{\partial F_2}{\partial x^3}, \frac{\partial F_1}{\partial x^3} - \frac{\partial F_3}{\partial x^1}, \frac{\partial F_2}{\partial x^1} - \frac{\partial F_1}{\partial x^2}\right) \tag{3.4}$$

The curl of a vector field is, again, a vector field. It can also be written as the determinant

$$\nabla \times \mathbf{F} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ \frac{\partial}{\partial x^1} & \frac{\partial}{\partial x^2} & \frac{\partial}{\partial x^3} \\ F_1 & F_2 & F_3 \end{vmatrix}$$

As we proceed through these lectures, we'll build intuition for the meaning of these two derivatives. We will see, in particular, that the divergence $\nabla \cdot \mathbf{F}$ measures the net flow of the vector field $\mathbf{F}$ into, or out of, any given point. Meanwhile, the curl $\nabla \times \mathbf{F}$ measures the rotation of the vector field. A full understanding of this will come only in Section 4 when we learn to undo the differentiation through integration. For now we will content ourselves with some simple examples.

Simple Examples

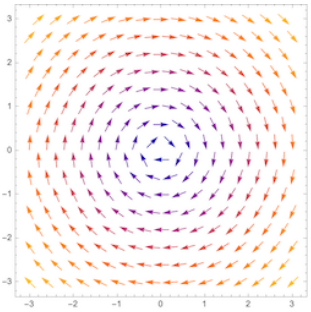
Consider the vector field

$$\mathbf{F}(\mathbf{x}) = (x^2, 0, 0)$$

Clearly this flows in a straight line, with increasing strength. It has $\nabla \cdot \mathbf{F} = 2x$, reflecting the fact that the vector field gets stronger as x increases. It also has $\nabla \times \mathbf{F} = 0$.

Next, consider the vector field

$$\mathbf{F}(\mathbf{x}) = (y, -x, 0)$$



This swirls, as shown in the figure on the right. We have $\nabla \cdot \mathbf{F} = 0$ and $\nabla \times \mathbf{F} = (0, 0, -2)$. The curl points in the $\hat{\mathbf{z}}$ direction, perpendicular to the plane of the swirling.

Finally, we can consider the hedgehog-like radial vector field that we met previously,

$$\mathbf{F} = \frac{\hat{\mathbf{r}}}{r^2} = \frac{1}{(x^2 + y^2 + z^2)^{3/2}} (x, y, z) \tag{3.5}$$

You can check that this obeys $\nabla \cdot \mathbf{F} = 0$ and $\nabla \times \mathbf{F} = 0$. Or, to be more precise, it obeys these equations *almost* everywhere. Clearly something fishy is going on at the origin $r = 0$. In fact, we will later see that we can make this less fishy: a correct statement is

$$\nabla \cdot \mathbf{F} = 4\pi \delta^3(\mathbf{x})$$

where $\delta^3(\mathbf{x})$ is the higher-dimensional version of the Dirac delta function. We'll understand this result better in Section 5 where we will wield the Gauss divergence theorem.

When evaluating the derivatives of radial fields, like the hedgehog (3.5), it's best to work with the radial distance r, given by $r^2 = x^i x^i$. Taking the derivative then gives $2r\partial r/\partial x^i = 2x^i$ and we have $\partial r/\partial x^i = x^i/r$. You can then check that, for any integer p,

$$\nabla r^p = \mathbf{e}_i \frac{\partial (r^p)}{\partial x^i} = p r^{p-1} \hat{\mathbf{r}}$$

Meanwhile, the vector $\mathbf{x} = x_i \mathbf{e}_i$ can equally well be written as $\mathbf{x} = \mathbf{r} = r\hat{\mathbf{r}}$ which highlights that it points outwards in the radial direction. We have

$$\nabla \cdot \mathbf{r} = \frac{\partial x^i}{\partial x^i} = \delta_{ii} = n$$

where the n arises because we're summing over all $i = 1, \ldots, n$. (Obviously, if we're working in $\mathbb{R}^3$ then $n = 3$.) We can also take the curl

$$\nabla \times \mathbf{r} = \epsilon_{ijk} \frac{\partial x^j}{\partial x^i} \mathbf{e}_k = 0$$

which, of course, as always holds only in $\mathbb{R}^3$.

3.2.1 Some Basic Properties

There are a number of straightforward properties obeyed by grad, div and curl. First, each of these is a *linear* differential operator, meaning that

$$\begin{aligned} \nabla(\alpha\phi + \psi) &= \alpha\nabla\phi + \nabla\psi \\ \nabla \cdot (\alpha\mathbf{F} + \mathbf{G}) &= \alpha\nabla \cdot \mathbf{F} + \nabla \cdot \mathbf{G} \\ \nabla \times (\alpha\mathbf{F} + \mathbf{G}) &= \alpha\nabla \times \mathbf{F} + \nabla \times \mathbf{G} \end{aligned}$$

for any scalar fields ϕ and ψ, vector fields $\mathbf{F}$ and $\mathbf{G}$, and any constant α.

Next, each of them has a Leibniz property, which means that they obey a generalisation of the product rule. These are

$$\begin{aligned} \nabla(\phi\psi) &= \phi\nabla\psi + \psi\nabla\phi \\ \nabla \cdot (\phi\mathbf{F}) &= (\nabla\phi) \cdot \mathbf{F} + \phi(\nabla \cdot \mathbf{F}) \\ \nabla \times (\phi\mathbf{F}) &= (\nabla\phi) \times \mathbf{F} + \phi(\nabla \times \mathbf{F}) \end{aligned}$$

In the last of these, you need to be careful about the placing and ordering of ∇, just like you need to be careful about the ordering of any other vector when dealing with the cross product. The proof of any of these is simply an exercise in plugging in the

component definition of the operator and using the product rule. For example, we can prove the second equality thus:

$$\nabla \cdot (\phi \mathbf{F}) = \frac{\partial(\phi F_i)}{\partial x^i} = \frac{\partial \phi}{\partial x^i} F_i + \phi \frac{\partial F_i}{\partial x^i} = (\nabla \phi) \cdot \mathbf{F} + \phi(\nabla \cdot \mathbf{F})$$

There are also a handful of further Leibnizian properties involving two vector fields. The first of these is straightforward to state:

$$\nabla \cdot (\mathbf{F} \times \mathbf{G}) = (\nabla \times \mathbf{F}) \cdot \mathbf{G} - \mathbf{F} \cdot (\nabla \times \mathbf{G})$$

This is simplest to prove using index notation. Alternatively, it follows from the usual scalar triple product formula for three vectors. To state the other properties, we need one further small abstraction. Given a vector field $\mathbf{F}$ and the gradient operator ∇, we can construct further differential operators. These are

$$\mathbf{F} \cdot \nabla = F_i \frac{\partial}{\partial x^i} \quad \text{and} \quad \mathbf{F} \times \nabla = \mathbf{e}_k \epsilon_{ijk} F_i \frac{\partial}{\partial x^j}$$

Note that the vector field $\mathbf{F}$ sits on the left, so isn't acted upon by the partial derivative. Instead, each of these objects is itself a differential operator, just waiting for something to come along so that it can differentiate it. In particular, these constructions appear in two further identities

$$\begin{aligned} \nabla(\mathbf{F} \cdot \mathbf{G}) &= \mathbf{F} \times (\nabla \times \mathbf{G}) + \mathbf{G} \times (\nabla \times \mathbf{F}) + (\mathbf{F} \cdot \nabla)\mathbf{G} + (\mathbf{G} \cdot \nabla)\mathbf{F} \\ \nabla \times (\mathbf{F} \times \mathbf{G}) &= (\nabla \cdot \mathbf{G})\mathbf{F} - (\nabla \cdot \mathbf{F})\mathbf{G} + (\mathbf{G} \cdot \nabla)\mathbf{F} - (\mathbf{F} \cdot \nabla)\mathbf{G} \end{aligned}$$

Again, these are not difficult to prove: they follow from expanding out the left-hand side in components.

3.2.2 Conservative is Irrotational

Recall that a conservative vector field $\mathbf{F}$ is one that can be written as

$$\mathbf{F} = \nabla \phi$$

for some scalar field ϕ. We also say that $\mathbf{F}$ is *irrotational* if $\nabla \times \mathbf{F} = 0$. There is a beautiful theorem that says these two concepts are actually equivalent:

Theorem: For fields defined *everywhere* on $\mathbb{R}^3$, conservative is the same as irrotational.

$$\nabla \times \mathbf{F} = 0 \quad \Longleftrightarrow \quad \mathbf{F} = \nabla \phi$$

Half Proof: It is trivial to prove this in one direction, Suppose that $\mathbf{F} = \nabla\phi$, so that $F_i = \partial_i\phi$. Then

$$\nabla \times \mathbf{F} = \epsilon_{ijk}\partial_i F_j \mathbf{e}_k = \epsilon_{ijk}\partial_i\partial_j\phi\, \mathbf{e}_k = 0$$

which vanishes because the ϵ_{ijk} symbol means that we're anti-symmetrising over ij, but the partial derivatives $\partial_i\partial_j$ are symmetric, so the terms like $\partial_1\partial_2 - \partial_2\partial_1$ cancel.

It is less obvious that the converse statement holds, i.e. that irrotational implies conservative. We'll show this only in Section 4.4 where it appears as a corollary of Stokes' theorem. □

Recall that in Section 1.3 we showed that the line integral of a conservative field was independent of the path taken. Putting this together with the result above, we have the following, equivalent statements:

$$\nabla \times \mathbf{F} = 0 \quad \Longleftrightarrow \quad \mathbf{F} = \nabla\phi \quad \Longleftrightarrow \quad \oint_C \mathbf{F} \cdot d\mathbf{x} = 0$$

where we've yet to see the proof of the first $\Longrightarrow$. In fact, we will complete this step through Stokes' theorem which shows that the statement on the far-left is equivalent to the statement on the far-right.

3.2.3 Solenoidal Fields

Here is another definition. A vector field $\mathbf{F}$ is called *divergence free* or *solenoidal* if $\nabla \cdot \mathbf{F} = 0$. (The latter name comes from electromagnetism, where a magnetic field $\mathbf{B}$ is most easily generated by a tube with a bunch of wires wrapped around it known as a "solenoid" and has the property $\nabla \cdot \mathbf{B} = 0$.)

There is a nice theorem about divergence free fields that is a counterpart to the one above:

Theorem: Any divergence free field can be written as the curl of something else,

$$\nabla \cdot \mathbf{F} = 0 \quad \Longleftrightarrow \quad \mathbf{F} = \nabla \times \mathbf{A}$$

again, provided that $\mathbf{F}$ is defined everywhere on $\mathbb{R}^3$. Note that $\mathbf{A}$ is not unique. In particular, if you find one $\mathbf{A}$ that does the job then any other $\mathbf{A} + \nabla\phi$ will work equally as well. In later courses, we will see that this theorem and the previous one both get subsumed into a single theorem known as the *Poincaré lemma*.

Proof: It's again straightforward to show this one way. If $\mathbf{F} = \nabla \times \mathbf{A}$, then $F_i = \epsilon_{ijk}\partial_j A_k$ and so

$$\nabla \cdot \mathbf{F} = \partial_i(\epsilon_{ijk}\partial_j A_k) = 0$$

which again vanishes for the symmetry reasons.

This time, we will prove the converse statement by explicitly exhibiting a vector potential $\mathbf{A}$ such that $\mathbf{F} = \nabla \times \mathbf{A}$. We pick some arbitrary point $\mathbf{x}_0 = (x_0, y_0, z_0)$ and then construct the following vector field

$$\mathbf{A}(\mathbf{x}) = \left(\int_{z_0}^{z} F_y(x, y, z')\, dz' \ , \ \int_{x_0}^{x} F_z(x', y, z_0)\, dx' - \int_{z_0}^{z} F_x(x, y, z')\, dz' \ , \ 0 \right) \tag{3.6}$$

Since $A_z = 0$, the definition of the curl (3.4) becomes

$$\nabla \times \mathbf{A} = \left(-\frac{\partial A_y}{\partial z} \ , \ \frac{\partial A_x}{\partial z} \ , \ \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \right)$$

Using the ansatz (3.6), we find that the first two components of $\nabla \times \mathbf{A}$ immediately give what we want

$$(\nabla \times \mathbf{A})_x = F_x(x, y, z) \quad \text{and} \quad (\nabla \times \mathbf{A})_y = F_y(x, y, z)$$

both of which follow from the fundamental theorem of calculus. Meanwhile, we still have a little work ahead of us for the final component

$$(\nabla \times \mathbf{A})_z = F_z(x, y, z_0) - \int_{z_0}^{z} \frac{\partial F_x}{\partial x}(x, y, z')\, dz' - \int_{z_0}^{z} \frac{\partial F_y}{\partial y}(x, y, z')\, dz'$$

At this point we use the fact that $\mathbf{F}$ is solenoidal, so $\nabla \cdot \mathbf{F} = 0$ and so $\partial F_z/\partial z' = -(\partial F_x/\partial x + \partial F_y/\partial y)$. We then have

$$(\nabla \times \mathbf{A})_z = F_z(x, y, z_0) + \int_{z_0}^{z} \frac{\partial F_z}{\partial z'}(x, y, z')\, dz' = F_z(x, y, z)$$

This is the result we want. □

Note that both theorems above come with a caveat: the fields must be defined everywhere on $\mathbb{R}^3$. This is important as counterexamples exist that do not satisfy this requirement, similar to the one that we met in a previous context in Section 1.3.4. These counterexamples will take on a life of their own in future courses where they provide the foundations to think about topology, both in mathematics and physics.

We've seen two related results above. A vector field $\mathbf{F} = \nabla\phi$ obeys $\nabla \times \mathbf{F} = 0$ and a vector field $\mathbf{F} = \nabla \times \mathbf{A}$ obeys $\nabla \cdot \mathbf{F} = 0$. In fact, it can be shown that the most general vector field on $\mathbb{R}^3$ can be decomposed a

$$\mathbf{F} = \nabla\phi + \nabla \times \mathbf{A}$$

for some ϕ and $\mathbf{A}$. This is known as the *Helmholtz decomposition*. We won't prove this statement here, although it follows from the result above if you can show that, for any $\mathbf{F}$, there always exists a potential ϕ such that $\mathbf{F} - \nabla\phi$ is solenoidal. (This ultimately follows from properties of the Laplace equation that we describe in section 5.2.)

3.2.4 The Laplacian

The *Laplacian* is a second order differential operator defined by

$$\nabla^2 = \nabla \cdot \nabla = \frac{\partial^2}{\partial x^i \partial x^i}$$

For example, in 3d the Laplacian takes the form

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

This is a scalar differential operator meaning that, when acting on a scalar field ϕ, it gives back another scalar field $\nabla^2\phi$. Similarly, it acts component by component on a vector field $\mathbf{F}$, giving back another vector field $\nabla^2\mathbf{F}$. If we use the vector triple product formula, we find

$$\nabla \times (\nabla \times \mathbf{F}) = \nabla(\nabla \cdot \mathbf{F}) - \nabla^2\mathbf{F}$$

which we can rearrange to give an alternative expression for the Laplacian acting on the components of a vector field

$$\nabla^2\mathbf{F} = \nabla(\nabla \cdot \mathbf{F}) - \nabla \times (\nabla \times \mathbf{F})$$

We'll devote Section 5 to solving various equations involving the Laplacian.

3.2.5 Some Vector Calculus Equations in Physics

I mentioned in the introduction that all laws of physics are written in the language of vector calculus (or, in the case of general relativity, a version of vector calculus extended to curved spaces, known as differential geometry). Here, for example, are the four equations of electromagnetism, known collectively as the *Maxwell equations*

$$\begin{aligned} \nabla \cdot \mathbf{E} &= \frac{\rho}{\epsilon_0} \quad , \quad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \\ \nabla \cdot \mathbf{B} &= 0 \quad , \quad \nabla \times \mathbf{B} = \mu_0 \left(\mathbf{J} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right) \end{aligned} \tag{3.7}$$

Here $\mathbf{E}$ and $\mathbf{B}$ are the electric and magnetic fields, while $\rho(\mathbf{x})$ is a scalar field that describes the distribution of electric charge in space and $\mathbf{J}(\mathbf{x})$ is a vector field that describes the distribution of electric currents. The equations also include two constants of nature, ϵ_0 and μ_0 which describe the strengths of the electric and magnetic forces respectively.

This simple set of equations describes everything we know about electricity, magnetism and light. Extracting this information requires the tools that we will develop in the rest of these lectures. Along the way, we will sometimes turn to the Maxwell equations to illustrate new ideas.

You'll find the Laplacian sitting in many other equations of physics. For example, the Schrödinger equation describing a quantum particle is written using the Laplacian. A particularly important equation, that crops up in many places, is the *heat equation*,

$$\frac{\partial T}{\partial t} = D\nabla^2 T$$

This tells us, for example, how temperature $T(\mathbf{x}, t)$ evolves over time. Here D is called the *diffusion constant*. This same equation also governs the spread of many other substances when there is some random element in the process, such as the constant bombardment from other atoms. For example, the smell of that guy who didn't shower before coming to lectures spreads through the room in manner described by the heat equation.

3.3 Orthogonal Curvilinear Coordinates

The definition of all our differential operators relied heavily on using Cartesian coordinates. The purpose of this section is simply to ask what these objects look like in different coordinate systems. As usual, the spherical polar and cylindrical polar coordinates in $\mathbb{R}^3$ will be of particular interest to us.

In general, we can describe a point $\mathbf{x}$ in $\mathbb{R}^3$ using some coordinates u, v, w, so $\mathbf{x} = \mathbf{x}(u, v, w)$. Changing either of these coordinates, leaving the others fixed, results in a change in $\mathbf{x}$. We have

$$d\mathbf{x} = \frac{\partial \mathbf{x}}{\partial u}\, du + \frac{\partial \mathbf{x}}{\partial v}\, dv + \frac{\partial \mathbf{x}}{\partial w}\, dw \tag{3.8}$$

Here $\partial\mathbf{x}/\partial u$ is the tangent vector to the lines defined by $v, w =$ constant, with similar statements for the others. A given set of coordinates provides a good parameterisation of some region provided that

$$\frac{\partial \mathbf{x}}{\partial u} \cdot \left(\frac{\partial \mathbf{x}}{\partial v} \times \frac{\partial \mathbf{x}}{\partial w} \right) \neq 0$$

The coordinate (u, v, w) are said to be *orthogonal curvilinear* if the three tangent vectors are mutually orthogonal. Here the slightly odd name "curvilinear" reflects the fact that these tangent vectors are typically not constant, but instead depend on position. We'll see examples shortly.

For orthogonal curvilinear coordinates, we can always define orthonormal tangent vectors simply by normalising them. We write

$$\frac{\partial \mathbf{x}}{\partial u} = h_u \mathbf{e}_u \quad , \quad \frac{\partial \mathbf{x}}{\partial v} = h_v \mathbf{e}_v \quad , \quad \frac{\partial \mathbf{x}}{\partial w} = h_w \mathbf{e}_w$$

where we've introduced *scale factors* $h_u, h_v, h_w > 0$ and $\mathbf{e}_u$, $\mathbf{e}_v$ and $\mathbf{e}_w$ form a right-handed orthonormal basis so that $\mathbf{e}_u \times \mathbf{e}_v = \mathbf{e}_w$. This can always be achieved simply by ordering the coordinates appropriately. Our original equation (3.8) can now be written as

$$d\mathbf{x} = h_u \mathbf{e}_u du + h_v \mathbf{e}_v dv + h_w \mathbf{e}_w dw \tag{3.9}$$

Squaring this, we have

$$d\mathbf{x}^2 = h_u^2\, du^2 + h_v^2\, dv^2 + h_w^2\, dw^2$$

from which it's clear that h_u, h_v and h_w are scale factors that tell us the change in length as we change each of the coordinates.

Throughout this section, we'll illustrate everything with three coordinate systems.

Cartesian Coordinates

First, Cartesian coordinates are easy:

$$\mathbf{x} = (x, y, z) \quad \Longrightarrow \quad h_x = h_y = h_z = 1 \quad \text{and} \quad \mathbf{e}_x = \hat{\mathbf{x}},\ \mathbf{e}_y = \hat{\mathbf{y}},\ \mathbf{e}_z = \hat{\mathbf{z}}$$

Cylindrical Polar Coordinates

Next, cylindrical polar coordinates are defined by (see also (2.7))

$$\mathbf{x} = (\rho \cos\phi, \rho \sin\phi, z)$$

with $\rho \geq 0$ and $\phi \in [0, 2\pi)$ and $z \in \mathbb{R}$. Inverting,

$$\rho = \sqrt{x^2 + y^2} \quad \text{and} \quad \tan\phi = \frac{y}{x}$$

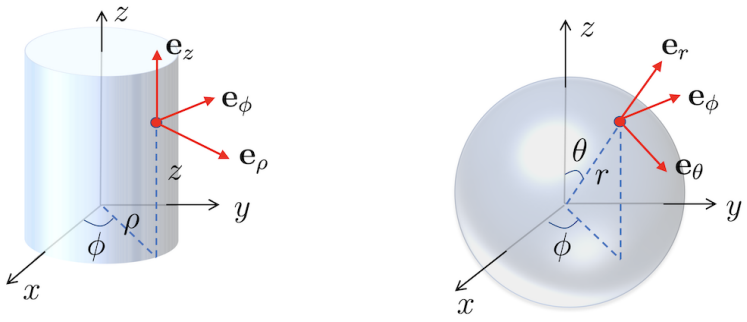


Figure 12. Cylindrical polar coordinates, on the left, and spherical polar coordinates, on the right.

It's straightforward to calculate

$$\begin{aligned}\mathbf{e}_\rho &= \hat{\boldsymbol{\rho}} = (\cos\phi, \sin\phi, 0)\\ \mathbf{e}_\phi &= \hat{\boldsymbol{\phi}} = (-\sin\phi, \cos\phi, 0)\\ \mathbf{e}_z &= \hat{\mathbf{z}}\end{aligned}$$

with

$$h_\rho = h_z = 1 \quad \text{and} \quad h_\phi = \rho$$

The three orthonormal vectors are shown on the left-hand side of Figure 12 in red. Note, in particular, that the vectors depend on ϕ and rotate as you change the point at which they're evaluated.

Spherical Polar Coordinates

Spherical polar coordinates are defined by (see also (2.5).)

$$\mathbf{x} = (r\sin\theta\cos\phi, r\sin\theta\sin\phi, r\cos\theta)$$

with $r \geq 0$ and $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi)$. Inverting,

$$r = \sqrt{x^2 + y^2 + z^2} \quad , \quad \tan\theta = \frac{\sqrt{x^2+y^2}}{z} \quad , \quad \tan\phi = \frac{y}{x}$$

Again, we can easily calculate the basis vectors

$$\begin{aligned}\mathbf{e}_r &= \hat{\mathbf{r}} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)\\ \mathbf{e}_\theta &= \hat{\boldsymbol{\theta}} = (\cos\theta\cos\phi, \cos\theta\sin\phi, -\sin\theta)\\ \mathbf{e}_\phi &= \hat{\boldsymbol{\phi}} = (-\sin\phi, \cos\phi, 0)\end{aligned}$$

These are shown in the right-hand side of Figure 12 in red. This time, the scaling factors are

$$h_r = 1 \quad , \quad h_\theta = r \quad , \quad h_\phi = r\sin\theta$$

We'll now see how various vector operators appear when written in polar coordinates.

3.3.1 Grad

The gradient operator is straightforward. If we shift the position from $\mathbf{x}$ to $\mathbf{x} + \delta\mathbf{x}$, then a scalar field $f(\mathbf{x})$ changes by

$$df = \nabla f \cdot d\mathbf{x} \tag{3.10}$$

This definition can now be used in any coordinate system. In a general coordinate system we have

$$df = \frac{\partial f}{\partial u}du + \frac{\partial f}{\partial v}dv + \frac{\partial f}{\partial w}dw = \nabla f \cdot (h_u \mathbf{e}_u du + h_v \mathbf{e}_v dv + h_w \mathbf{e}_w dw)$$

Using the orthonormality of the basis elements vectors, and comparing the terms on the left and right, this then gives us the gradient operator

$$\nabla f = \frac{1}{h_u}\frac{\partial f}{\partial u}\mathbf{e}_u + \frac{1}{h_v}\frac{\partial f}{\partial v}\mathbf{e}_v + \frac{1}{h_w}\frac{\partial f}{\partial w}\mathbf{e}_w \tag{3.11}$$

In cylindrical polar coordinates, the gradient of a function $f(\rho, \phi, z)$ is

$$\nabla f = \frac{\partial f}{\partial \rho}\hat{\boldsymbol{\rho}} + \frac{1}{\rho}\frac{\partial f}{\partial \phi}\hat{\boldsymbol{\phi}} + \frac{\partial f}{\partial z}\hat{\mathbf{z}}$$

In spherical polar coordinates, the gradient of a function $f(r, \theta, \phi)$ is

$$\nabla f = \frac{\partial f}{\partial r}\hat{\mathbf{r}} + \frac{1}{r}\frac{\partial f}{\partial \theta}\hat{\boldsymbol{\theta}} + \frac{1}{r\sin\theta}\frac{\partial f}{\partial \phi}\hat{\boldsymbol{\phi}}$$

Note, in particular, that when we differentiate with respect to an angle there is always a compensating 1/length prefactor to make sure that the dimensions are right.

3.3.2 Div and Curl

To construct the div and curl in a general coordinate system, we first extract the vector differential operator

$$\nabla = \frac{1}{h_u}\mathbf{e}_u\frac{\partial}{\partial u} + \frac{1}{h_v}\mathbf{e}_v\frac{\partial}{\partial v} + \frac{1}{h_w}\mathbf{e}_w\frac{\partial}{\partial w} \tag{3.12}$$

where, importantly, we've placed the vectors to the left of the differentials because, as we've seen, the basic vectors now typically depend on the coordinates. If we act on a function f with this operator, we recover the gradient (3.11). But now we have this abstract operator, we can also take it to act on a vector field $\mathbf{F}(u,v,w)$. We can expand the vector field as

$$\mathbf{F}(u,v,w) = F_u\mathbf{e}_u + F_v\mathbf{e}_v + F_w\mathbf{e}_w$$

Each of the components depends on the coordinates u, v and w. But so too, in general, do the basis vectors $\{\mathbf{e}_u, \mathbf{e}_v, \mathbf{e}_w\}$. This means that when the derivatives in the differential operator (3.12) hit $\mathbf{F}$, they also act on both the components and the basis vectors.

Given an explicit expression for the basis vectors, it's not hard to see what happens when they are differentiated. For example, in cylindrical polar coordinates we find

$$\nabla\cdot\mathbf{F} = \frac{1}{\rho}\frac{\partial(\rho F_\rho)}{\partial\rho} + \frac{1}{\rho}\frac{\partial F_\phi}{\partial\phi} + \frac{\partial F_z}{\partial z}$$

and

$$\nabla\times\mathbf{F} = \left(\frac{1}{\rho}\frac{\partial F_z}{\partial\phi} - \frac{\partial F_\phi}{\partial z}\right)\hat{\boldsymbol{\rho}} + \left(\frac{\partial F_\rho}{\partial z} - \frac{\partial F_z}{\partial\rho}\right)\hat{\boldsymbol{\phi}} + \frac{1}{\rho}\left(\frac{\partial(\rho F_\phi)}{\partial\rho} - \frac{\partial F_\rho}{\partial\phi}\right)\hat{\mathbf{z}}$$

There is a question on Examples Sheet 2 that asks you to explicitly verify this. Meanwhile, in spherical polar coordinates, we have

$$\nabla\cdot\mathbf{F} = \frac{1}{r^2}\frac{\partial(r^2F_r)}{\partial r} + \frac{1}{r\sin\theta}\frac{\partial(\sin\theta F_\theta)}{\partial\theta} + \frac{1}{r\sin\theta}\frac{\partial F_\phi}{\partial\phi}$$

and

$$\begin{aligned}\nabla\times\mathbf{F} = {}& \frac{1}{r\sin\theta}\left(\frac{\partial(\sin\theta F_\phi)}{\partial\theta} - \frac{\partial F_\theta}{\partial\phi}\right)\hat{\mathbf{r}} \\ & + \frac{1}{r}\left(\frac{1}{\sin\theta}\frac{\partial F_r}{\partial\phi} - \frac{\partial(rF_\phi)}{\partial r}\right)\hat{\boldsymbol{\theta}} \\ & + \frac{1}{r}\left(\frac{\partial(rF_\theta)}{\partial r} - \frac{\partial F_r}{\partial\theta}\right)\hat{\boldsymbol{\phi}}\end{aligned}$$

For completeness, we also give the general results

Claim: Given a vector field $\mathbf{F}(u,v,w)$ in a general orthogonal, curvilinear coordinate system, the divergence is given by

$$\nabla\cdot\mathbf{F} = \frac{1}{h_uh_vh_w}\left(\frac{\partial}{\partial u}(h_vh_wF_u) + \frac{\partial}{\partial v}(h_uh_wF_v) + \frac{\partial}{\partial w}(h_uh_vF_w)\right) \tag{3.13}$$

and the curl is given by the determinant

$$\nabla \times \mathbf{F} = \frac{1}{h_u h_v h_w} \begin{vmatrix} h_u \mathbf{e}_u & h_v \mathbf{e}_v & h_w \mathbf{e}_w \\ \frac{\partial}{\partial u} & \frac{\partial}{\partial v} & \frac{\partial}{\partial w} \\ h_u F_u & h_v F_v & h_w F_w \end{vmatrix}$$

where the derivatives on the second line should now be thought of as acting on the third line only, but not the first. This means that, in components, we have

$$\nabla \times \mathbf{F} = \frac{1}{h_v h_w} \left(\frac{\partial}{\partial v}(h_w F_w) - \frac{\partial}{\partial w}(h_v F_v) \right) \mathbf{e}_u + \text{two similar terms}$$

Proof: Not now. Later. It turns out to be a little easier when we have some integral technology in hand. For this reason, we'll revisit this in Section 4.4.4.

3.3.3 The Laplacian

Finally, we have the Laplacian. From (3.11) and (3.13), this takes the general form

$$\nabla^2 f = \nabla \cdot \nabla f = \frac{1}{h_u h_v h_w} \left[\frac{\partial}{\partial u} \left(\frac{h_v h_w}{h_u} \frac{\partial f}{\partial u} \right) + \frac{\partial}{\partial v} \left(\frac{h_u h_w}{h_v} \frac{\partial f}{\partial v} \right) + \frac{\partial}{\partial w} \left(\frac{h_u h_v}{h_w} \frac{\partial f}{\partial w} \right) \right]$$

Obviously in Cartesian coordinates, the Laplacian is

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}$$

In cylindrical polar coordinates it takes the form

$$\nabla^2 f = \frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\rho \frac{\partial f}{\partial \rho} \right) + \frac{1}{\rho^2} \frac{\partial^2 f}{\partial \phi^2} + \frac{\partial^2 f}{\partial z^2} \tag{3.14}$$

and in spherical polar coordinates

$$\nabla^2 f = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial f}{\partial r} \right) + \frac{1}{r^2 \sin\theta} \frac{\partial}{\partial \theta} \left(\sin\theta \frac{\partial f}{\partial \theta} \right) + \frac{1}{r^2 \sin^2\theta} \frac{\partial^2 f}{\partial \phi^2} \tag{3.15}$$

The most canonical of canonical physics textbooks is J.D. Jackson's "Classical Electrodynamics". I don't know of any theoretical physicist who doesn't have a copy on their shelf. It's an impressive book but I'm pretty sure that, for many, the main selling point is that it has these expressions for div, grad and curl in cylindrical and polar coordinates printed on the inside cover. You can also find these results collated on the last pages of these lecture notes. We'll return to the Laplacian in different coordinate systems in Section 5.2 where we'll explore the solutions to equations like $\nabla^2 f = 0$.

4 The Integral Theorems

The fundamental theorem of calculus states that integration is the inverse of the differentiation, in the sense that

$$\int_a^b dx\ \frac{df}{dx} = f(b) - f(a)$$

In this section, we describe a number of generalisations of this result to higher dimensional integrals. Along the way, we will also gain some intuition for the meaning of the various vector derivative operators.

4.1 The Divergence Theorem

The *divergence theorem*, also known as *Gauss' theorem*, states that, for a smooth vector field $\mathbf{F}(\mathbf{x})$ over $\mathbb{R}^3$,

$$\int_V \nabla \cdot \mathbf{F}\, dV = \int_S \mathbf{F} \cdot d\mathbf{S} \tag{4.1}$$

where V is a bounded region whose boundary $\partial V = S$ is a piecewise smooth closed surface. The integral on the right-hand side is taken with the normal $\mathbf{n}$ pointing outward.

The Meaning of the Divergence

We'll prove the divergence theorem shortly. But first, let's make good on our promise to build some intuition for the divergence. To this end, integrate $\nabla \cdot \mathbf{F}$ over some region of volume V centred at the point $\mathbf{x}$. If the region is small enough, then $\nabla \cdot \mathbf{F}$ will be roughly constant, and so

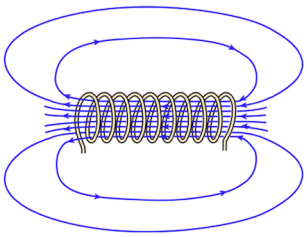
$$\int_V \nabla \cdot \mathbf{F}\, dV \approx V\, \nabla \cdot \mathbf{F}(\mathbf{x})$$

and this becomes exact as the region shrinks to zero size. The divergence theorem then provides a coordinate independent definition of the divergence

$$\nabla \cdot \mathbf{F} = \lim_{V \to 0} \frac{1}{V} \int_S \mathbf{F} \cdot d\mathbf{S} \tag{4.2}$$

This is the result that we advertised in Section 3: the right way to think about the divergence of a vector field is as the net flow into, or out of, a region. If $\nabla \cdot \mathbf{F} > 0$ at some point $\mathbf{x}$, then there is a net flow out of that point; if $\nabla \cdot \mathbf{F} < 0$ at some point $\mathbf{x}$ then there is a net flow inwards.

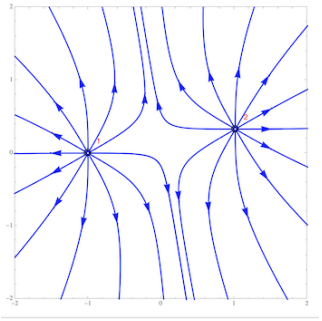
We can illustrate this by looking at a couple of the Maxwell equations (3.7). The magnetic field $\mathbf{B}$ is solenoidal, obeying



$$\nabla \cdot \mathbf{B} = 0$$

This means that the magnetic vector field can't pile up anywhere: at any given point in space, there is as much magnetic field coming in as there is going out. This leads us to draw the magnetic field as continuous, never ending streamlines. For example, the magnetic field lines for solenoid, a long coil of wire carrying a current, is shown in the figure (taken from the website hyperphysics).

Meanwhile, electric field $\mathbf{E}$ obeys

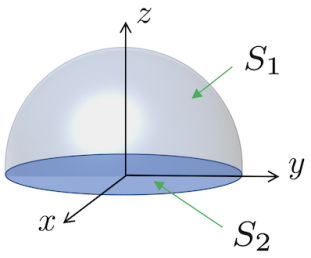


$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}$$

where $\rho(\mathbf{x})$ is the electric charge density. In any region of space where there's no electric charge, so $\rho(\mathbf{x}) = 0$, the electric field lines act just like the magnetic field and can't pile up anywhere. However, the presence of electric charge changes this, and causes the field lines to pile up or disappear. In other words, the electric charge acts as a source or a sink for electric field lines. The electric field lines arising from two pointlike, positive charges which act as sources, are shown in the figure.

Example

Before proving the theorem, we first give an example. Take the volume V to be the solid hemispherical ball, defined as $x^2 + y^2 + z^2 \leq R^2$ and $z \geq 0$. Then boundary of V then has two pieces



$$\partial V = S_1 + S_2$$

where S_1 is the hemisphere and S_2 the disc in the $z = 0$ plane. We'll integrate the vector field

$$\mathbf{F} = (0, 0, z + R)$$

The $+R$ doesn't contribute in the volume integral since we have $\nabla \cdot \mathbf{F} = 1$. Then

$$\int_V \nabla \cdot \mathbf{F}\, dV = \int_V dV = \frac{2}{3}\pi R^3 \tag{4.3}$$

which is the volume of the hemispherical ball. For the surface integral, we work with S_1 and S_2 separately. On the hemisphere S_1, the unit normal vector is $\mathbf{n} = \frac{1}{R}(x, y, z)$ and so

$$\mathbf{F} \cdot \mathbf{n} = \frac{z(z+R)}{R} = R\cos\theta(\cos\theta + 1)$$

where we've used polar coordinates $z = R\cos\theta$. The integral is then

$$\begin{aligned}\int_{S_1} \mathbf{F} \cdot d\mathbf{S} &= \int_0^{2\pi} d\phi \int_0^{\pi/2} d\theta \ (R^2 \sin\theta)\, R\cos\theta(\cos\theta + 1) \\ &= 2\pi R^3 \left[-\frac{1}{3}\cos^3\theta - \frac{1}{2}\cos^2\theta \right]_0^{\pi/2} \\ &= = 2\pi R^3 \left(\frac{1}{3} + \frac{1}{2} \right) = \frac{5\pi R^3}{3} \end{aligned} \tag{4.4}$$

where the $R^2 \sin\theta$ factor in the first line is the Jacobian that we previously saw in (2.9). Meanwhile, for the integral over the disc S_2, we have the normal vector $\mathbf{n} = (0, 0, -1)$, and so (remembering that the disc sits at $z = 0$),

$$\mathbf{F} \cdot \mathbf{n} = -R \quad \Rightarrow \quad \int_{S_2} \mathbf{F} \cdot d\mathbf{S} = (-R) \times \pi R^2$$

with πR^2 the area of the disc. Adding these together, we have

$$\int_{S_1+S_2} \mathbf{F} \cdot d\mathbf{S} = \frac{2}{3}\pi R^3$$

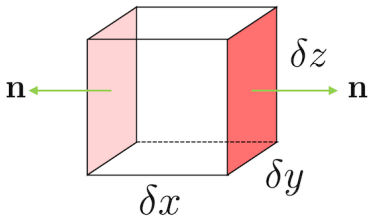
which reproduces the volume integral as promised.

It's worth tracking what became of the $+R$ term in the vector field $\mathbf{F}$. Obviously it didn't contribute to the volume integral. For the surface integral over S_1, it gave the $+1/2$ term in the penultimate expression in (4.4). This was then cancelled by the surface integral over S_2, which only received a contribution from the $+R$ term. We see that this constant vector field when in the top surface, and out the bottom surface, giving no contribution to the overall surface integral. This is how we get agreement with the volume integral which, due to the derivative, is oblivious to any constant (or, indeed, divergent free) components of $\mathbf{F}$.

4.1.1 A Proof of the Divergence Theorem

We start by giving an informal sketch of the basic idea underlying the divergence theorem. We'll then proceed with a more rigorous proof.

To get some intuition for the divergence theorem, take the volume V and divide it up into a bunch of small cubes. A given cube $V_{\mathbf{x}}$ has one corner of the cube sitting at $\mathbf{x} = (x, y, z)$ and sides of lengths δx, δy and δz.



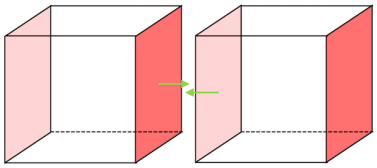
For a small enough cube, we can think of $\mathbf{F} \cdot \mathbf{n}$ as being approximately constant on any given side. To start, we look at the flux of $\mathbf{F}$ through the two sides that lie in the (y, z) plane is given by

$$[F_x(x + \delta x, y, z) - F_x(x, y, z)]\, \delta y\, \delta z \approx \frac{\partial F_x}{\partial x} \delta x\, \delta y\, \delta z \tag{4.5}$$

where the minus sign comes because the flux is calculated using the outward pointing normal and the right-hand side comes from Taylor expanding $F_x(x + \delta x, y, z)$. We get similar expressions for the integrals over the sides that lie in the (x, y) plane and in the (x, z) plane. Summing over six sides, the total flux through the surface of this tiny cube is then

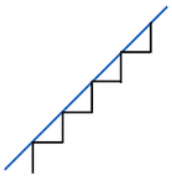
$$\int_{\text{tiny}} \mathbf{F} \cdot d\mathbf{S} = \left(\frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} \right) \delta x\, \delta y\, \delta z = \nabla \cdot \mathbf{F}\, \delta x\, \delta y\, \delta z$$

But now we've tiled our volume V with a whole bunch of these cubes, we can apply the formulae above to each of them. On the right-hand side, we add up the value of $\nabla \cdot \mathbf{F}$ in each cube. This, of course, is the volume integral that we're after. On the left-hand side, something more interesting happens. Now we get a term like the left-hand side of (4.5) for each box, and we sum over all boxes. But this means that all contributions from interior faces cancel out because the outward normal of one box is in the opposite direction to the outward normal from the other box. The upshot is that any interior contribution to the flux vanishes, and we are left only with the contribution from the boundary $S = \partial V$. This then gives us the claimed result



$$\int_S \mathbf{F} \cdot d\mathbf{S} = \int_V \nabla \cdot \mathbf{F}\, dV$$

The derivation above is simple and intuitive, but it might leave you a little nervous. The essence of the divergence theorem is to relate a bulk integral to a boundary integral. But it's not obvious that the boundary can be well approximated by stacking cubes together. To give an analogy, if you try to approximate a 45° line by a series of horizontal and vertical lines, as shown on the right, then the total length of the steps is always going to be $\sqrt{2}$ larger than the length of the horizontal line, no matter how fine you make them. You might worry that these kind of issues afflict the proof above. For that reason, we now give a more careful derivation of the divergence theorem.



Before we proceed, first note that, suitably interpreted, the divergence theorem holds in arbitrary dimension $\mathbb{R}^n$, where a "surface" now means a codimension one subspace. In particular, the divergence theorem holds in $\mathbb{R}^2$, where a surface is a curve. This result, which is interesting in its own right, will serve as a warm-up exercise to proving the general divergence theorem.

The 2d Divergence Theorem: Let $\mathbf{F}$ be a vector field in $\mathbb{R}^2$. Then

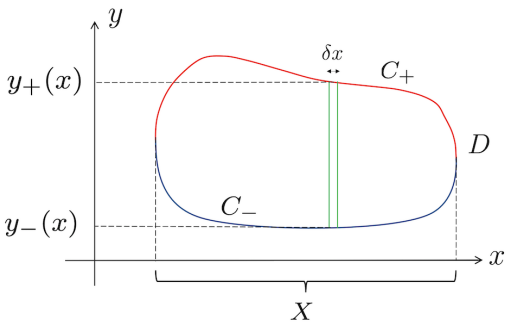
$$\int_D \nabla \cdot \mathbf{F}\, dA = \int_C \mathbf{F} \cdot \mathbf{n}\, ds \tag{4.6}$$

where D is a region in $\mathbb{R}^2$, bounded by the closed curve C and $\mathbf{n}$ is the outward normal to C.

Proof of the 2d Divergence Theorem: For simplicity, we'll assume that $\mathbf{F} = F(x,y)\,\hat{\mathbf{y}}$. The proof that we're about to give also works if $\mathbf{F}$ points solely in the $\hat{\mathbf{x}}$ direction, but a general $\mathbf{F}$ is just a linear sum of the two.

We then have

$$\int_D \nabla \cdot \mathbf{F}\, dA = \int_X dx \int_{y_-(x)}^{y_+(x)} dy\, \frac{\partial F}{\partial y}$$

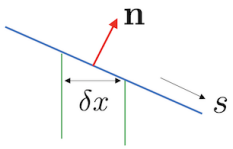


where, as the notation shows, we've chosen to do the area integral by first integrating over y, and then over x. We'll assume, for now, that the region D is convex, as shown in the figure, so that each $\int dy$ is over just a single interval with limits $y_\pm(x)$. These limits trace out an upper curve C_+, shown

in red in the figure, and a lower curve C_- shown in blue. We then have

$$\int_D \nabla \cdot \mathbf{F}\, dA = \int_X dx\, \Big(F(x, y_+(x)) - F(x, y_-(x))\Big)$$

We've succeeded in converting the area integral into an ordinary integral, but it's not quite of the line integral form that we need. The next part of the proof is to massage the integral over $\int dx$ into a line integral over $\int ds$. This is easily achieved if we look at the zoomed-in figure to the right. Along the upper curve C_+, the normal $\mathbf{n}$ points upwards and makes an angle $\cos\theta = \hat{\mathbf{y}} \cdot \mathbf{n}$ with the vertical. Moving a small distance δs along the curve is equivalent to moving



$$\delta x = \cos\theta\, \delta s = \hat{\mathbf{y}} \cdot \mathbf{n}\, \delta s \quad \text{along } C_+$$

Along the lower curve, C_-, the normal $\mathbf{n}$ points downwards and so $\hat{\mathbf{y}} \cdot \mathbf{n}$ is negative. We then have

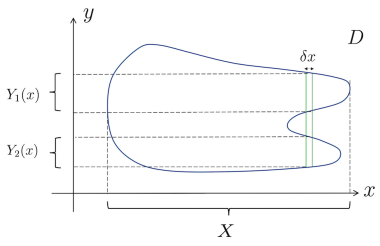
$$\delta x = -\hat{\mathbf{y}} \cdot \mathbf{n}\, \delta s \quad \text{along } C_-$$

The upshot is that we can write the area integral as

$$\begin{aligned}\int_D \nabla \cdot \mathbf{F}\, dA &= \int_X ds\, \Big(\mathbf{n} \cdot \mathbf{F}(x, y_+(x)) + \mathbf{n} \cdot \mathbf{F}(x, y_-(x))\Big) \\ &= \int_{C_+} \mathbf{F} \cdot \mathbf{n}\, ds + \int_{C_-} \mathbf{F} \cdot \mathbf{n}\, ds \\ &= \int_C \mathbf{F} \cdot \mathbf{n}\, ds\end{aligned}$$

with $C = C_+ + C_- = \partial D$ the boundary of the region.

We're left with one small loophole to close: if the region D is not convex, then the range of the integral $\int dy$ may be over two or more disconnected intervals, as shown in the figure. In this case, the boundary curve decomposes into more pieces, but the basic strategy still holds. □



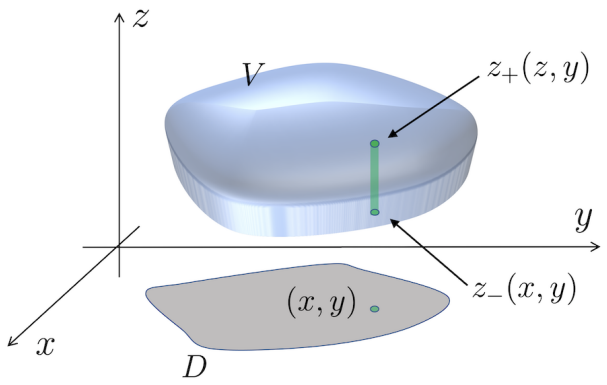


Figure 13. Performing the $\int dz$ integral for the proof of the 3d divergence theorem.

Proof of the 3d Divergence Theorem

The proof of the 3d (or, indeed, higher dimensional) divergence theorem follows using the same strategy. If we focus on $\mathbf{F} = F(x,y,z)\,\hat{\mathbf{z}}$ we have

$$\begin{aligned}\int_V \nabla\cdot\mathbf{F}\,dV &= \int_D dA \int_{z_-(x,y)}^{z_+(x,y)} dz\;\frac{\partial F}{\partial z}\\ &= \int_D dA\,\Big(F(x,y,z_+(x,y)) - F(x,y,z_-(x,y))\Big)\end{aligned}$$

where the limits of the integral $z_\pm(x,y)$ are the upper and lower surfaces of the volume V. The area integral over D is an integral in the (x,y) plane, while to prove Gauss' theorem we need to convert this into a surface integral over $S = \partial V$. This step of the argument is the same as before: at any given point, the different between $dA = dxdy$ and dS is the angle $\cos\theta = \mathbf{n}\cdot\hat{\mathbf{z}}$ (up to a sign). This then gives the promised result (4.1). □

The Divergence Theorem for Scalar Fields

There is a straightforward extension of the divergence theorem for scalar fields ϕ:

Claim: For $S = \partial V$, we have

$$\int_V \nabla\phi\,dV = \int_S \phi\,d\mathbf{S}$$

Proof: Consider the divergence theorem (4.1) with $\mathbf{F} = \phi\mathbf{a}$ where $\mathbf{a}$ is a constant vector. We have

$$\int_V \nabla\cdot(\phi\mathbf{a})dV = \int_S (\phi\mathbf{a})\cdot d\mathbf{S} \quad\Rightarrow\quad \mathbf{a}\cdot\left(\int_V \nabla\phi\,dV - \int_S \phi\,d\mathbf{S}\right) = 0$$

This is true for any constant vector $\mathbf{a}$, and so the expression in the brackets must itself vanish. □

4.1.2 Carl Friedrich Gauss (1777-1855)

Gauss is regarded by many as the greatest mathematician of all time. He made seminal contributions to number theory, algebra, geometry, and physics.

Gauss was born to working class parents in what is now Lower Saxony, Germany. In 1795 he went to study at the university of Göttingen and remained there for the next 60 years.

There are remarkably few stories about Gauss that do not, at the end of the day, boil down to the observation that he was just really good at maths. There is even a website that has collected well over 100 retellings of how Gauss performed the sum $\sum_1^{100} n$ when still a foetus. (You can find an interesting dissection of this story here.)

4.2 An Application: Conservation Laws

Of the many important applications of the divergence theorem, one stands out. In many situations, we have the concept of a conservation law: some quantity that doesn't change over time. There are conservation laws in fundamental physics, including energy, momentum, angular momentum and electric charge and several more that emerge when we look to more sophisticated theories. There are also approximate conservation laws at play when we model more complicated systems. For example, if you're interested in how the population distribution of some species evolves over time then it might well serve you to ignore birth rates and traffic accidents and consider the total number of animals to be fixed.

In all these cases, the quantity is conserved. But we can say something stronger than that: it is conserved *locally*. For example, an electric charge sitting in the palm of your hand can't disappear and turn up on Jupiter. That would satisfy a "global" conservation of charge, but that's not the way the universe works. If the electric charge disappears from your hand, then most likely it has fallen off and is now sitting on the floor. Or, said more precisely, it must have moved to a nearby region of space.

The divergence theorem provides the technology to describe local conservation laws of this type. First, we introduce the *density* $\rho(\mathbf{x}, t)$ of the conserved object. For the purposes of this discussion, we will take this to be the density of electric charge, although it could equally well be the density of any of the other conserved quantities

described above. The total electric charge in some region V is then given by the integral

$$Q = \int_V \rho \, dV$$

The conservation of charge is captured by the following statement: there exists a vector field $\mathbf{J}(\mathbf{x}, t)$ such that

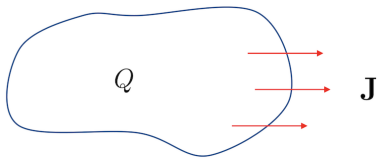
$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0$$

This is known as the *continuity* equation and $\mathbf{J}$ is called the *current density*.

The continuity equation doesn't tell us that the density ρ can't change in time; that would be overly prohibitive. But it does tell us that ρ must change only in a certain way. This ensures that the change in the charge Q in a fixed region V is given by

$$\frac{dQ}{dt} = \int_V \frac{\partial \rho}{\partial t} \, dV = -\int_V \nabla \cdot \mathbf{J} \, dV = -\int_S \mathbf{J} \cdot d\mathbf{S}$$

where the second equality follows from the continuity equation and the third from the divergence theorem at some fixed time t. We learn that the charge inside a region can only change if there is a current flowing through the surface of that region. This is how the conservation of charge is enforced locally.



The intuition behind this idea is straightforward. If you want to keep tabs on the number of people in a nightclub, you don't continuously count them. Instead you measure the number of people entering and leaving through the door.

If the current is known to vanish outside some region, so $\mathbf{J}(\mathbf{x}) = 0$ for $|\mathbf{x}| > R$, then the total charge contained inside that region must be unchanging. Often, in such situations, we ask only that $\mathbf{J}(\mathbf{x}, t) \to 0$ suitably quickly as $|\mathbf{x}| \to \infty$, in which case the total charge is unchanging

$$Q_{\text{total}} = \int_{\mathbb{R}^3} \rho \, dV \quad \text{and} \quad \frac{dQ_{\text{total}}}{dt} = 0$$

In later courses, we'll see many examples of the continuity equation. The example of electric charge discussed above will be covered in the lectures on Electromagnetism, where the flux of $\mathbf{J}$ through a surface S is

$$I = \int_S \mathbf{J} \cdot d\mathbf{S}$$

and is what we usually call the electric current.

We will also see the same equation in the lectures on Quantum Mechanics where $\rho(\mathbf{x})$ has the interpretation of the probability density for a particle to be at some point $\mathbf{x}$ and $Q = \int_V \rho\, dV$ is the probability that the particle sits in some region V. Obviously, in this example we must have $Q_{\text{total}} = 1$ which is the statement that particle definitely sits somewhere.

Finally, the continuity equation also plays an important role in Fluid Mechanics where the mass of the fluid is conserved. In that case, $\rho(\mathbf{x}, t)$ is the density of the fluid and the current is $\mathbf{J} = \rho\mathbf{u}$ where $\mathbf{u}(\mathbf{x}, t)$ is the *velocity field*. The continuity equation then reads

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0$$

In this case the flux is the mass of fluid that passes through a surface S in time t.

In many circumstances, liquids can be modelled as *incompressible*, meaning that $\rho(\mathbf{x}, t)$ is a constant in both space and time. In these circumstances, we have $\dot{\rho} = \nabla\rho = 0$ and the continuity equation tells us that the velocity field is necessarily solenoidal:

$$\nabla \cdot \mathbf{u} = 0 \tag{4.7}$$

This makes sense: for a solenoidal vector field, the flow into any region must be accompanied by an equal outgoing flow, telling us that the fluid can't pile up anywhere, as expected for an incompressible fluid. The statement that fluids are incompressible is a fairly good approximation until we come to think about sound, which arises because of changes in the density which propagate as waves.

4.2.1 Conservation and Diffusion

There is a close connection between conserved quantities and the idea of diffusion. We'll illustrate this with the idea of energy conservation. The story takes a slightly different form depending on the context, but here we'll think of the energy contained in a hot gas. First, since energy is conserved there is necessarily a corresponding continuity equation

$$\frac{\partial \mathcal{E}}{\partial t} + \nabla \cdot \mathbf{J} = 0 \tag{4.8}$$

where $\mathcal{E}(\mathbf{x}, t)$ is the energy density of the gas, and $\mathbf{J}$ is the *heat current* which tells us how energy is transported from one region of space to another.

At this point we need to invoke a couple of physical principles. First, the energy density in a gas is proportional to the temperature of the gas,

$$\mathcal{E}(\mathbf{x}, t) = cT(\mathbf{x}, t) \tag{4.9}$$

where c_V is the specific heat capacity. Next comes a key step: in hot systems, where everything is jiggling around randomly, the heat flow is due to temperature differences between different parts of the system. The relation between the two is captured by the equation

$$\mathbf{J} = -\kappa \nabla T \tag{4.10}$$

where κ is called the *thermal conductivity* and the minus sign ensures that heat flows from hot to cold. This relation is known as *Fick's law*. Neither (4.9) nor (4.10) are fundamental equations of physics and both can be derived from first principles by thinking about the motion of the underlying atoms. (This will be described in the lectures on Statistical Physics and, for Fick's law, the lectures on Kinetic Theory.)

Combining the continuity equation (4.8) with the definition of temperature (4.9) and Fick's law (4.10), we find the heat equation

$$\frac{\partial T}{\partial t} = D\nabla^2 T$$

where the diffusion constant is given by $D = \kappa/c$. This tells us how the temperature of a system evolves. As we mentioned previously, the same heat equation describes the diffusive motion of any conserved quantity.

4.2.2 Another Application: Predator-Prey Systems

We'll see more applications of the divergence theorem in Section 5, mainly in the context of the gravitational and electrostatic forces. However, the uses of the theorem are many and varied and stretch far beyond applications to the laws of physics. Here we give an example in the world of ecology which is modelled mathematically by differential equations. As we'll see, the use of ∇ here is somewhat novel because we're not differentiating with respect to space but with respect to some more abstract variables.

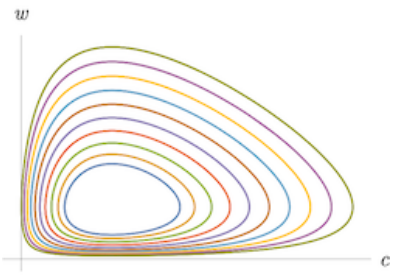
First some background. Predator-prey systems describe the interaction between two species. We will take our predators to be wolves. (Because they're cool.) We will denote the population of wolves at a given time t as $w(t)$. The wolves prey upon something cute and furry. We will denote the population of this cute, furry thing as $c(t)$.

We want to write down a system of differential equations to describe the interaction between wolves and cute furry things. The simplest equations were first written down by Lotka and Volterra and (after some rescaling) take the form

$$\frac{dw}{dt} = w(-\alpha + c)$$
$$\frac{dc}{dt} = c(\beta - w)$$

with $\alpha, \beta > 0$ are some constants. There is a clear meaning to the different terms in these equations. Without food, the wolves die out. That is what the $-\alpha w$ term in the first equation is telling us which, if $c = 0$, will cause the wolf population to decay exponentially quickly. In contrast, without wolves the cute furry things eat grass and prosper. That's what the $+\beta c$ term in the second equation is telling us which, if $w = 0$, ensures that the population of cute furry things grows exponentially. The second term in each equation, $\pm wc$, tells us what happens when the wolves and cute furry things meet. The $\pm$ sign means that it's good news for one, less good for the other.

The Lotka-Volterra equations are straightforward to solve. There is a fixed point at $c = \alpha$ and $w = \beta$ at which the two populations are in equilibrium. Away from this, we find periodic orbits as the two populations wax and wane. To see this, we think of $w = w(c)$ and write the pair of equations as



$$\frac{dw}{dc} = \frac{w(c-\alpha)}{c(\beta - w)}$$

This equation is separable and we have

$$\int \frac{\beta - w}{w}\, dw = \int \frac{c-\alpha}{c}\, dc \quad \Rightarrow \quad \beta \log \omega - \omega + \alpha \log c - c = \text{constant}$$

These orbits are plotted in the (c, w) plane, also known as the phase plane, for different constants in the figure.

So much for the Lotka-Volterra equations. Let's now look at something more complicated. Suppose that there is some intra-species competition: a little wolfy bickering that sometimes gets out of hand, and some cute, furry in-fighting. We can model this by adding extra terms to the original equations:

$$\frac{dw}{dt} = w(-\alpha + c - \mu w)$$
$$\frac{dc}{dt} = c(\beta - w - \nu c) \tag{4.11}$$

where the two new constants are also positive, $\mu, \nu > 0$. Both new terms come with minus signs, which is appropriate because fighting is bad.

What do we do now? There is still a fixed point, now given by $(1+\mu\nu)w = \beta - \nu\alpha$ and $(1+\mu\nu)c = \alpha + \mu\beta$. But what happens away from this fixed point? Do the periodic orbits that we saw earlier persist? Or does something different happen?

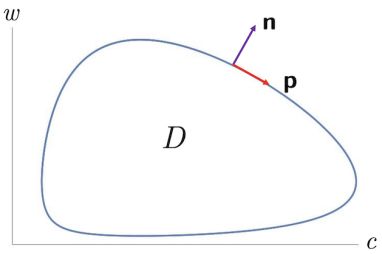
Sadly, we can't just solve the differential equation like we did before because it's no longer separable. Instead, we're going to need a more creative method to understand what's going on. This is where the divergence theorem comes in. We will use it to show that, provided $\mu \neq 0$ or $\nu \neq 0$, the periodic orbits of the Lotka-Volterra equation no longer exist.

We first change notation a little. We write the pair of predator-prey equations (4.11) in vector form

$$\frac{d\mathbf{a}}{dt} = \mathbf{p} \quad \text{with} \quad \mathbf{a} = \begin{pmatrix} w \\ c \end{pmatrix} \quad \text{and} \quad \mathbf{p} = \begin{pmatrix} w(-\alpha + c - \mu w) \\ c(\beta - w - \nu c) \end{pmatrix}$$

Any solution to these equations traces out a path $\mathbf{a}(t)$ in the animal phase plane. The re-writing above makes it clear that $\mathbf{p}$ is the tangent to this path. The question that we wish to answer is: does this path close? In other words, is there a periodic orbit?

It turns out that there are no periodic orbits. To show this, we will suppose that periodic orbits exist and then argue by contradiction. The normal $\mathbf{n}$ to the path $\mathbf{a}(t)$ obeys $\mathbf{n} \cdot \mathbf{p} = 0$, as shown in the figure. This means that if we integrate any function $b(w, c)$ around the periodic orbit we have



$$\oint b(w,c)\, \mathbf{p} \cdot \mathbf{n}\, dt = 0$$

By the 2d divergence theorem, this in turn means that the following integral over the area enclosed by the periodic orbit must also vanish:

$$\int_D \nabla \cdot [b(w,c)\mathbf{p}]\; dA = 0$$

where, in this context, the gradient operator is $\nabla = (\partial/\partial w, \partial/\partial c)$. At this juncture, the trick is to find a cunning choice of function $b(w, c)$. The one that works for us is $b = 1/wc$. This is because we have

$$\nabla \cdot \frac{\mathbf{p}}{wc} = -\frac{\mu}{c} - \frac{\nu}{w}$$

Both of these terms are strictly negative. (For this it is important to remember that populations w and c are strictly positive!) But if $\nabla \cdot (\mathbf{p}/wc)$ is always negative then there's no way to integrate it over a region and get zero. Something has gone wrong. And what's gone wrong was our original assumption of closed orbits. We learn that the nice periodic solutions of the Lotka-Volterra equations are spoiled by any intra-species competition. We're left just with the fixed point which is now stable. All of which is telling us that a little in-fighting may not be so bad after all. It keeps things stable.

The general version of the story above goes by the name of the Bendixson-Dulac theorem and is a powerful tool in the study of dynamical systems.

4.3 Green's Theorem in the Plane

Let $P(x, y)$ and $Q(x, y)$ be smooth functions on $\mathbb{R}^2$. Then

$$\int_A \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA = \oint_C P dx + Q dy \tag{4.12}$$

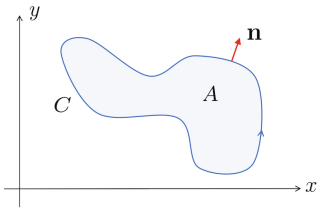
where A is a bounded region in the plane and $C = \partial A$ is a piecewise smooth, non-intersecting closed curve which is traversed anti-clockwise.

Proof: Green's theorem is equivalent to the 2d divergence theorem (4.6). Let $\mathbf{F} = (Q, -P)$ be a vector field in $\mathbb{R}^2$. We then have

$$\int_A \nabla \cdot \mathbf{F} \, dA = \int_A \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA \tag{4.13}$$

If $\mathbf{x}(s) = (x(s), y(s))$ is the parameterised curve C, then the tangent vector is $\mathbf{t}(s) = (x'(s), y'(s))$ and the normal vector $\mathbf{n} = (y'(s), -x'(s))$ obeys $\mathbf{n} \cdot \mathbf{t} = 0$.

You'll need to do a little sketch to convince yourself that, as shown on the right, $\mathbf{n}$ is the outward pointing normal provided that the arc length s increases in the anti-clockwise direction. We then have



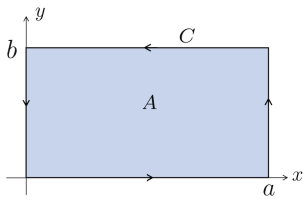
$$\mathbf{F} \cdot \mathbf{n} = Q \frac{dy}{ds} + P \frac{dx}{ds}$$

and so the integral around C is

$$\int_C \mathbf{F} \cdot \mathbf{n} \, ds = \int_C P dx + Q dy \tag{4.14}$$

The 2d divergence theorem is the statement that the left-hand sides of (4.13) and (4.14) are equal; Green's theorem in the plane is the statement that the right-hand sides are equal. □

Applied to a rectangular region, Green's theorem in the plane reduces to the fundamental theorem of calculus. We take the rectangular region to be $0 \le x \le a$ and $0 \le y \le b$. Then



$$\begin{aligned}\int_A -\frac{\partial P}{\partial y}\,dA &= -\int_0^a dx \int_0^b dy\ \frac{\partial P}{\partial y} \\ &= \int_0^a dx\Big(-P(x,b) + P(x,0)\Big) = \int_C P\,dx\end{aligned}$$

where only the horizontal segments contribute, and the minus signs are such that C is traversed anti-clockwise. Meanwhile, we also have

$$\begin{aligned}\int_A \frac{\partial Q}{\partial x}\,dA &= \int_0^b dy \int_0^a dx\ \frac{\partial Q}{\partial x} \\ &= \int_0^b dy\Big(Q(a,y) - Q(0,y)\Big) = \int_C Q\,dy\end{aligned}$$

where, this time, only the vertical segments contribute.

Green's theorem also holds if the area A has a number of disconnected components, as shown in Figure 14. In this case, the integral should be done in an anti-clockwise direction around the exterior boundary, and in a clockwise direction on any interior boundary. The quickest way to see this is to do the integration around a continuous boundary, as shown in the right-hand figure, with an infinitesimal gap. The two contributions across the gap then cancel.

An Example

Let $P = x^2y$ and $Q = xy^2$. We'll take A to be the region bounded by the parabola $y^2 = 4ax$ and the line $x = a$, both with $-2a \le y \le 2a$. Then Green's theorem in the plane tells us that

$$\int_A (y^2 - x^2)\,dA = \int_C x^2y\,dx + xy^2\,dy$$

But this was a problem on the examples sheet, where you found that both give the answer $\frac{104}{105}a^4$.

4.3.1 George Green (1793-1841)

George Green was born in Nottingham, England, the son of a miller. If you were born to a family of millers in the 18$^{\text{th}}$ century, they didn't send you to a careers officer at school to see what you want to be when you grow up. You'd be lucky just to be sent to school. Green got lucky. He attended school for an entire year before joining his father baking and running the mill.

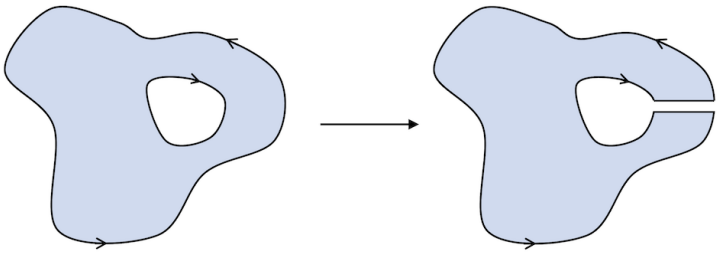


Figure 14. Don't mind the gap. Green's theorem for an area with disconnected boundaries.

It is not known where Green learned his mathematics. The Nottingham subscription library held some volumes, but not enough to provide Green with the background that he clearly gained. Yet, from his mill, Green produced some of the most striking mathematics of his time, including the development of potential theory and, most importantly, the formalism of Green's functions that you will meet in Section 5, as well as in later courses. Much of this was contained in a self-published pamphlet, from 1828, entitled "An Essay on the Application of Mathematical Analysis to the Theories of Electricity and Magnetism". 51 copies were printed.

Green's reputation spread and, at the age of 40, with no formal education, and certainly no Latin or Greek, Green the miller came to Cambridge as a mathematics undergraduate, clothes covered in flour and pretending it was chalk. (University motto: nurturing imposter syndrome since 1209.) With hindsight, this may not have been the best move. Green did well in his exams, but his published papers did not reach the revolutionary heights of his work in the mill. He got a fellowship at Caius, developed a taste for port, then gout, and died before he reached his 50th birthday.

There are parallels between Green's story and that of Ramanujan who came to Cambridge several decades later. To lose one self-taught genius might be regarded as a misfortune. To lose two begins to look like carelessness.

4.4 Stokes' Theorem

Stokes' theorem is an extension of Green's theorem, but where the surface is no longer restricted to lie in a plane.

Let S be a smooth surface in $\mathbb{R}^3$ with boundary $C = \partial S$ a piecewise smooth curve. *Stokes' theorem* states that, for any smooth vector field $\mathbf{F}(\mathbf{x})$, we have

$$\int_S \nabla \times \mathbf{F} \cdot d\mathbf{S} = \int_C \mathbf{F} \cdot d\mathbf{x}$$

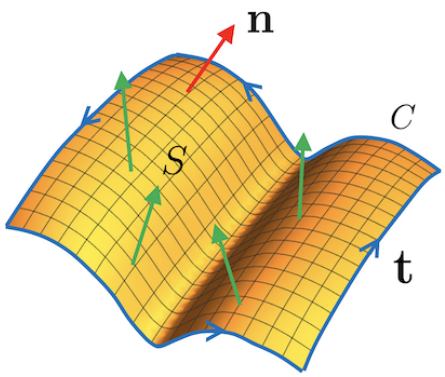


Figure 15. The surface S and bounding curve C for Stokes' theorem. The normal to the surface is shown (at one point) by the red arrow. The theorem invites us to compute the flux of a vector field $\mathbf{F}$, shown by the green arrows, through the surface, and compare it to the line integral around the boundary.

The orientations of $\mathbf{S}$ and $\mathbf{C}$ should be *compatible*. The former is determined by the choice of normal vector $\mathbf{n}$ to S; the latter by the choice of tangent vector $\mathbf{t}$ to C. The two are said to be compatible if $\mathbf{t} \times \mathbf{n}$ points out of S. In practice, this means that if you orient the open surface so that $\mathbf{n}$ points towards you, then the orientation of C is anti-clockwise. The general set-up is shown in Figure 15.

Note that there will typically be many surfaces S that share the same boundary C. By Stokes' theorem, the integral of $\nabla \times \mathbf{F}$ over S must give the same answer for all such surfaces. The theorem also holds if the boundary ∂S consists of a number of disconnected components, again with their orientation determined by that of S.

We'll give a proof of Stokes' theorem shortly. But first we put it to some use.

The Meaning of the Curl

Stokes' theorem gives us some new intuition for the curl of a vector field. If we integrate $\nabla \times \mathbf{F}$ over a small enough surface such that $\nabla \times \mathbf{F}$ is approximately constant, then we will have

$$\int_S \nabla \times \mathbf{F} \cdot d\mathbf{S} \approx A\, \mathbf{n} \cdot (\nabla \times \mathbf{F})$$

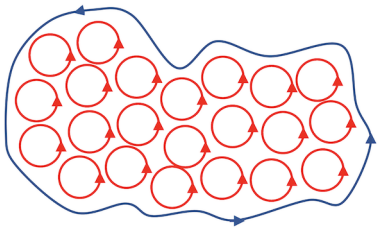
where A is the area and $\mathbf{n}$ the normal of the surface. Taking the limit in which this area shrinks to zero, Stokes' theorem then tell us that

$$\mathbf{n} \cdot (\nabla \times \mathbf{F}) = \lim_{A \to 0} \frac{1}{A} \int_C \mathbf{F} \cdot d\mathbf{x} \tag{4.15}$$

In other words, at any given point, the value of $\nabla \times \mathbf{F}$ in the direction $\mathbf{n}$ tells us about the circulation of $\mathbf{F}$ in the plane normal to $\mathbf{n}$

A useful benchmark comes from considering the vector field $\mathbf{u} = \boldsymbol{\omega} \times \mathbf{x}$, which describes a rigid rotation with angular velocity $\boldsymbol{\omega}$. (See, for example, the lectures on Dynamics and Relativity.) In that case, we have $\nabla \times \mathbf{u} = 2\boldsymbol{\omega}$, so twice the angular velocity.

Turning this on its head, we can get some intuition for Stokes' theorem itself. The curl of the vector field tells us about the local circulation of $\mathbf{F}$. When you integrate this circulation over some surface S, most of it cancels out because the circulation going one way is always cancelled by a neighbouring circulation going the other, as shown in the figure. The only thing that's left when you integrate over the whole surface is the circulation around the edge.



A Corollary: Irrotational Implies Conservative

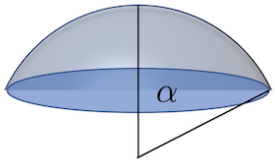
Before we prove Stokes' theorem, we can use it to tie off a thread that we previously left hanging. Recall that in Section 3.2, we proved that $\mathbf{F} = \nabla\phi \implies \nabla \times \mathbf{F} = 0$, but we didn't then have the tools to prove the converse. Now we do. It follows straightforwardly from Stokes' theorem because an irrotational vector field, obeying $\nabla \times \mathbf{F} = 0$, necessarily has

$$\oint_C \mathbf{F} \cdot d\mathbf{x} = 0$$

around any closed curve C. But we showed in Section 1.2 that any such conservative field can be written as $\mathbf{F} = \nabla\phi$ for some potential ϕ.

An Example

Let S be the cap of a sphere of radius R that is covered by the angle $0 \leq \theta \leq \alpha$, as shown in the figure. We'll take



$$\mathbf{F} = (0, xz, 0) \quad \Rightarrow \quad \nabla \times \mathbf{F} = (-x, 0, z) \quad (4.16)$$

This is the example that we discussed in Section 2.2.5, where we computed (see (2.11))

$$\int_S \nabla \times \mathbf{F} \cdot d\mathbf{S} = \pi R^3 \cos\alpha \sin^2\alpha \tag{4.17}$$

That leaves us with the line integral around the rim. This curve C is parameterised by the angle ϕ and is given by

$$\mathbf{x}(\phi) = R(\sin\alpha\cos\phi, \sin\alpha\sin\phi, \cos\alpha) \quad \Rightarrow \quad d\mathbf{x} = R(-\sin\alpha\sin\phi, \sin\alpha\cos\phi, 0)\, d\phi$$

We then have

$$\int_C \mathbf{F} \cdot d\mathbf{x} = \int_0^{2\pi} d\phi\ Rxz\sin\alpha\cos\phi = R^3\sin^2\alpha\cos\alpha \int_0^{2\pi} d\phi\ \cos^2\phi = \pi R^3 \sin^2\alpha\cos\alpha$$

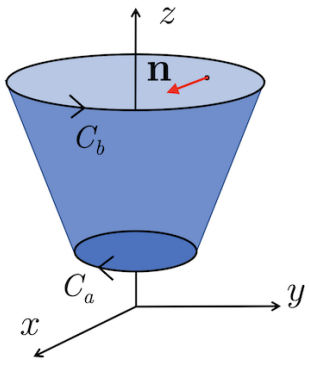
in agreement with the surface integral (4.17).

Another Example

As a second example, consider the conical surface S defined by $z^2 = x^2 + y^2$ with $0 < a \leq z \leq b$. This surface is parameterised, in cylindrical polar coordinates, by

$$\mathbf{x}(\rho, \phi) = (\rho\cos\phi, \rho\sin\phi, \rho) \tag{4.18}$$

with $a \leq \rho \leq b$ and $0 \leq \phi < 2\pi$. We can compute two tangent vectors



$$\frac{\partial \mathbf{x}}{\partial \rho} = (\cos\phi, \sin\phi, 1) \quad \text{and} \quad \frac{\partial \mathbf{x}}{\partial \phi} = \rho(-\sin\phi, \cos\phi, 0)$$

and take their cross product to get the normal

$$\mathbf{n} = \frac{\partial \mathbf{x}}{\partial \rho} \times \frac{\partial \mathbf{x}}{\partial \phi} = (-\rho\cos\phi, -\rho\sin\phi, \rho)$$

This points inwards, as shown in the figure. The associated vector area element is

$$d\mathbf{S} = (-\cos\phi, -\sin\phi, 1)\rho d\rho\, d\phi$$

We'll integrate the same vector field (4.16) over this surface. We have

$$\nabla \times \mathbf{F} \cdot d\mathbf{S} = (x\cos\phi + z)\rho\, d\rho\, d\phi = \rho^2(\cos^2\phi + 1)d\rho\, d\phi$$

where we've substituted in the parametric expressions for x and z from (4.18). The integral is then

$$\int_S \nabla \times \mathbf{F} \cdot d\mathbf{S} = \int_a^b d\rho \int_0^{2\pi} d\phi\ \rho^2(1+\cos^2\phi) = \pi(b^3 - a^3) \tag{4.19}$$

Now the surface has two boundaries, and we must integrate over both of them. We write $\partial S = C_b - C_a$ where C_b has radius b and C_a radius a. Note the minus sign, reflecting the fact that the orientation of the two circles is opposite.

For a circle of radius R, we have $\mathbf{x}(\phi) = R(\cos\phi, \sin\phi, 1)$, and so $d\mathbf{x} = R(-\sin\phi, \cos\phi, 0)\, d\phi$ and

$$\int_{C_R} \mathbf{F} \cdot d\mathbf{x} = \int_0^{2\pi} d\phi\ R^3\cos^2\phi = \pi R^3$$

Remembering that the orientation of C_a in the opposite direction, we reproduce the surface integral (4.19).

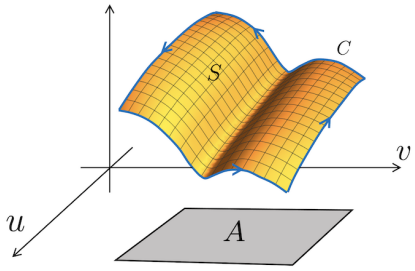
4.4.1 A Proof of Stokes' Theorem

It's clear that Stokes' theorem is a version of Green's theorem in the plane, but viewed through 3d glasses. Indeed, it's trivial to show that the latter follows from the former. Consider the vector field $\mathbf{F} = (P, Q, 0)$ in $\mathbb{R}^3$ and a surface S that lies flat in the $z = 0$ plane. The normal to this surface is $\mathbf{n} = \hat{\mathbf{z}}$, and we have

$$\int_S \nabla \times \mathbf{F} \cdot d\mathbf{S} = \int_S \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dS$$

But Stokes' theorem then tells us that this can also be written as

$$\int_C \mathbf{F} \cdot d\mathbf{x} = \int_C P dx + Q dy$$



However, with a little more work we can also show that the converse is true. In other words, we can lift Green's theorem out of the plane to find Stokes' theorem.

Consider a parameterised surface S defined by $\mathbf{x}(u,v)$ and denote the associated area in the (u,v) plane as A. We parameterise the boundary $C = \partial S$ as $\mathbf{x}(u(t),v(t))$ and the corresponding boundary ∂A as $(u(t),v(t))$. The key idea is to use Green's theorem in the (u,v) plane for the area A and then uplift this to prove Stokes theorem for the surface S.

We start by looking at the integral around the boundary. It is

$$\int_C \mathbf{F}\cdot d\mathbf{x} = \int_C \mathbf{F}\cdot\left(\frac{\partial\mathbf{x}}{\partial u}\,du + \frac{\partial\mathbf{x}}{\partial v}\,dv\right) = \int_{\partial A} F_u\,du + F_v\,dv$$

where $F_u = \mathbf{F}\cdot\partial\mathbf{x}/\partial u$ and $F_v = \mathbf{F}\cdot\partial\mathbf{x}/\partial v$. Now we're in a position to invoke Green's theorem, in the form

$$\int_{\partial A} F_u\,du + F_v\,dv = \int_A \left(\frac{\partial F_v}{\partial u} - \frac{\partial F_u}{\partial v}\right) dA$$

Now our task is clear. We should look at the partial derivatives on the right hand side. We just need to be careful about what thing depends on what thing:

$$\frac{\partial F_v}{\partial u} = \frac{\partial}{\partial u}\left(\mathbf{F}\cdot\frac{\partial\mathbf{x}}{\partial v}\right) = \frac{\partial}{\partial u}\left(F_i\frac{\partial x^i}{\partial v}\right) = \left(\frac{\partial F_i}{\partial x^j}\frac{\partial x^j}{\partial u}\right)\frac{\partial x^i}{\partial v} + F_i\,\frac{\partial^2 x^i}{\partial u\partial v}$$

Meanwhile, we have

$$\frac{\partial F_u}{\partial v} = \frac{\partial}{\partial v}\left(\mathbf{F}\cdot\frac{\partial\mathbf{x}}{\partial u}\right) = \frac{\partial}{\partial v}\left(F_i\frac{\partial x^i}{\partial u}\right) = \left(\frac{\partial F_i}{\partial x^j}\frac{\partial x^j}{\partial v}\right)\frac{\partial x^i}{\partial u} + F_i\,\frac{\partial^2 x^i}{\partial v\partial u}$$

Subtracting the second expression from the first, the second derivative terms cancel, leaving us with

$$\frac{\partial F_v}{\partial u} - \frac{\partial F_u}{\partial v} = \frac{\partial x^j}{\partial u}\frac{\partial x^i}{\partial v}\left(\frac{\partial F_i}{\partial x^j} - \frac{\partial F_j}{\partial x^i}\right) = (\delta_{jk}\delta_{il} - \delta_{jl}\delta_{ik})\,\frac{\partial x^k}{\partial u}\frac{\partial x^l}{\partial v}\frac{\partial F_i}{\partial x^j}$$

At this point we wield everyone's favourite index notation identity

$$\epsilon_{jip}\epsilon_{pkl} = \delta_{jk}\delta_{il} - \delta_{jl}\delta_{ik}$$

We then have

$$\frac{\partial F_v}{\partial u} - \frac{\partial F_u}{\partial v} = \epsilon_{jip}\epsilon_{pkl}\frac{\partial x^k}{\partial u}\frac{\partial x^l}{\partial v}\frac{\partial F_i}{\partial x^j} = (\nabla\times\mathbf{F})\cdot\left(\frac{\partial\mathbf{x}}{\partial u}\times\frac{\partial\mathbf{x}}{\partial v}\right)$$

8. If X, Y, Z be functions of the rectangular co-ordinates x, y, z, dS an element of any limited surface, l, m, n the cosines of the inclinations of the normal at dS to the axes, ds an element of the bounding line, shew that

$$\iint\left\{l\left(\frac{dZ}{dy}-\frac{dY}{dx}\right)+m\left(\frac{dX}{dz}-\frac{dZ}{dx}\right)+n\left(\frac{dY}{dx}-\frac{dX}{dy}\right)\right\}dS$$
$$=\int\left(X\frac{dx}{ds}+Y\frac{dy}{ds}+Z\frac{dz}{ds}\right)ds,$$

the differential coefficients of X, Y, Z being partial, and the single integral being taken all round the perimeter of the surface.

Figure 16. You may now turn the page... the original version of Stokes' theorem, set as an exam question.

Now we're done. Following through the chain of identities above, we have

$$\begin{aligned}\int_C \mathbf{F}\cdot d\mathbf{x} &= \int_A \left(\frac{\partial F_v}{\partial u}-\frac{\partial F_u}{\partial v}\right)\,dudv\\ &= \int_A (\nabla\times\mathbf{F})\cdot\left(\frac{\partial \mathbf{x}}{\partial u}\times\frac{\partial \mathbf{x}}{\partial v}\right)\,dudv\\ &= \int_S (\nabla\times\mathbf{F})\cdot d\mathbf{S}\end{aligned}$$

This is Stokes' theorem. □

4.4.2 George Gabriel Stokes (1819-1903)

Stokes was born in County Sligo, Ireland, but moved to Cambridge shortly after his 19th birthday and remained there for the next 66 years, much of it as Lucasian professor. He contributed widely to different area of mathematics and physics, with the Navier-Stokes equation, describing fluid flow, a particular highlight.

What we now call Stokes' theorem was communicated to Stokes by his friend William Thomson, better known by his later name Lord Kelvin. The theorem first appeared in print in 1854 as part of the Smith's prize examination competition, a second set of exams aimed at those students who felt the Tripos wasn't brutal enough.

If you're in Cambridge and looking for a tranquil place away from the tourists to sit, drink coffee, and ponder the wider universe, then you could do worse than the Mill Road cemetery, large parts of which are overgrown, derelict, and beautiful. Stokes is buried there, as is Cayley, although both gravestones were destroyed long ago. You can find Stokes' resting place nestled between the graves of his wife and daughter[1].

4.4.3 An Application: Magnetic Fields

Consider an infinitely long wire carrying a current. What is the magnetic field that is produced? We can answer this by turning to the Maxwell equations (3.7). For time independent situations, like this, one of the equations reads

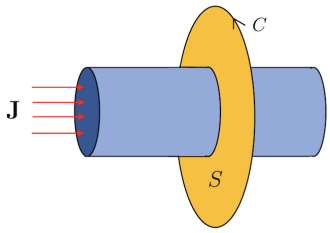
$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} \tag{4.20}$$

where $\mathbf{J}$ is the current density and μ_0 is a constant of nature that determines the strength of the magnetic field and has some pretentious name that I can never remember. Another of the Maxwell equations reads $\nabla \cdot \mathbf{B} = 0$ and in most situations we should solve this in conjunction with (4.20) but here it will turn out, somewhat fortuitously, that if we just find the obvious solution to (4.20) then it solves $\nabla \cdot \mathbf{B} = 0$ automatically.

The equation (4.20) provides a simple opportunity to use Stokes' theorem. We integrate both sides over a surface S that cuts through the wire, as shown in the figure to the right. We then have

$$\int_S \nabla \times \mathbf{B} \cdot d\mathbf{S} = \int_C \mathbf{B} \cdot d\mathbf{x} = \mu_0 \int_S \mathbf{J} \cdot d\mathbf{S} = \mu_0 I$$

where the integral of the current density gives I, the total current through the wire. This equation tells us that there must be a circulation of the magnetic field around the wire. In particular, there must be a component of $\mathbf{B}$ that lies tangent to any curve C that bounds a surface S.



Let's suppose that the wire lies in the z-direction. (Rotate your head or your screen if you

[1] A long, tree lined avenue runs north off Mill Road. At the end, turn right to enter the cemetery. There is a gravel path immediately off to your left, which you should ignore, but take the first mud track that runs parallel to it. Just after the gravestone bearing the name "Frederick Cooper" you will find the Stokes' family plot.

don't like the z direction to be horizontal.) Then if S is a disc of radius ρ, then the boundary $C = \partial S$ is paramterised by the curve

$$\mathbf{x} = \rho(\cos\phi, \sin\phi, 0) \quad \Longrightarrow \quad \mathbf{t} = \frac{\partial \mathbf{x}}{\partial \phi} = \rho(-\sin\phi, \cos\phi, 0)$$

We'll make the obvious guess that $\mathbf{B}$ lies in the same direction as $\mathbf{t}$ and work with the ansatz

$$\mathbf{B}(\mathbf{x}) = b(\rho)(-\sin\phi, \cos\phi, 0)$$

Then $\mathbf{B} \cdot \mathbf{t} = \rho b(\rho)$. Provided that ρ is bigger than the radius of the wire, Maxwell's equation tells us that

$$\mu_0 I = \int_C \mathbf{B} \cdot d\mathbf{x} = \int_0^{2\pi} d\phi \ \rho b(\rho) \quad \Longrightarrow \quad \mathbf{B}(\mathbf{x}) = \frac{\mu_0 I}{2\pi\rho}(-\sin\phi, \cos\phi, 0)$$

You can check that this answer also satisfies the other Maxwell equation $\nabla \cdot \mathbf{B} = 0$. We learn that the magnetic field circulates around the wire, and drops off as $1/\rho$ with ρ the distance from the wire.

4.4.4 Changing Coordinates Revisited

Back in Section 3.3, we wrote down the expressions for the divergence and curl in a general orthonormal curvilinear coordinate system. Now we can offer a proof using the integral theorems above.

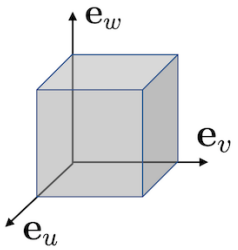
Claim: The divergence of a vector field $\mathbf{F}(u, v, w)$ in a general orthogonal, curvilinear coordinate system is given by

$$\nabla \cdot \mathbf{F} = \frac{1}{h_u h_v h_w}\left(\frac{\partial}{\partial u}(h_v h_w F_u) + \frac{\partial}{\partial v}(h_u h_w F_v) + \frac{\partial}{\partial w}(h_u h_v F_w)\right) \tag{4.21}$$

Proof: We sketch a proof that works with the integral definition of the divergence (4.2),

$$\nabla \cdot \mathbf{F} = \lim_{V \to 0} \frac{1}{V} \int_S \mathbf{F} \cdot d\mathbf{S}$$

We can take the volume V to consist of a small cuboid at point (u, v, w) with sides parallel to the basis vectors $\mathbf{e}_u$, $\mathbf{e}_v$ and $\mathbf{e}_w$. The volume of the cube is $h_u h_v h_w \delta u \, \delta v \, \delta w$. Meanwhile, the area of, say, the



upper face in the figure is roughly $h_u h_v \delta u\, \delta v$. Since h_u and h_v may depend on the coordinates, this could differ from the area of the lower face, albeit only by a small amount δw. Then, assuming that $\mathbf{F}$ is roughly constant on each face, we have

$$\begin{aligned}\int_S \mathbf{F}\cdot d\mathbf{S} &\approx \Big[h_u h_v F_w(u,v,w+\delta w) - h_u h_v F_w(u,v,w)\Big]\delta u\,\delta v + \text{two more terms}\\ &\approx \frac{\partial}{\partial w}(h_u h_v F_w)\delta u\,\delta v\,\delta w + \text{two more terms}\end{aligned}$$

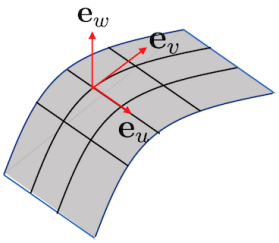
Dividing through by the volume then gives us the advertised expression for $\nabla\cdot\mathbf{F}$. □

Claim: The curl of a vector field $\mathbf{F}(u,v,w)$ in a general orthogonal, curvilinear coordinate system is given by

$$\begin{aligned}\nabla\times\mathbf{F} &= \frac{1}{h_u h_v h_w}\begin{vmatrix} h_u\mathbf{e}_u & h_v\mathbf{e}_v & h_w\mathbf{e}_w \\ \frac{\partial}{\partial u} & \frac{\partial}{\partial v} & \frac{\partial}{\partial w} \\ h_u F_u & h_v F_v & h_w F_w\end{vmatrix}\\ &= \frac{1}{h_v h_w}\left(\frac{\partial}{\partial v}(h_w F_w) - \frac{\partial}{\partial w}(h_v F_v)\right)\mathbf{e}_u + \text{two similar terms}\end{aligned}$$

Proof: This time we use the integral definition of curl (4.15)

$$\mathbf{n}\cdot(\nabla\times\mathbf{F}) = \lim_{A\to 0}\frac{1}{A}\int_C \mathbf{F}\cdot d\mathbf{x}$$



We'll take a surface S with normal $\mathbf{n} = \mathbf{e}_w$ and integrate over a small region, bounded by one of the squares in the figure on the right. The area of the square $h_u h_v \delta u\,\delta v$ while the length of each side is $h_u\delta u$ and $h_v\delta v$. Assuming that the square is small enough so that $\mathbf{F}$ is roughly constant along any given side, we have

$$\begin{aligned}\int_C \mathbf{F}\cdot d\mathbf{x} &\approx h_u F_u(u,v)\delta u + h_v F_v(u+\delta u,v)\delta v - h_u F_u(u,v+\delta v)\delta u - h_v F_v(u,v)\delta v\\ &\approx \Big[\frac{\partial}{\partial u}(h_v F_v) - \frac{\partial}{\partial v}(h_u F_u)\Big]\delta u\,\delta v\end{aligned}$$

Dividing by the area, this gives

$$\mathbf{e}_w\cdot\nabla\times\mathbf{F} = \frac{1}{h_u h_v}\Big[\frac{\partial}{\partial u}(h_v F_v) - \frac{\partial}{\partial v}(h_u F_u)\Big]$$

which is one of the three promised terms in the expression for $\nabla\times\mathbf{F}$. □

5 The Poisson and Laplace Equations

Until now, our focus has been very much on understanding how to differentiate and integrate functions of various types. But, with this under our belts, we can now take the next step and explore various differential equations that are written in the language of vector calculus. Our goal in this section is to find solutions to the Poisson equation and the related Laplace equation. This we will do in Section 5.2. But first we will explain why these equations underly two of the most important forces in the universe.

5.1 Gravity and Electrostatics

The first two fundamental forces to be discovered are also the simplest to describe mathematically. Newton's law of gravity states that two masses, m and M, separated by a distance r will experience a force

$$\mathbf{F}(r) = -\frac{GMm}{r^2}\hat{\mathbf{r}} \tag{5.1}$$

with G Newton's constant, a fundamental constant of nature that determines the strength of the gravitational force. Meanwhile, Coulomb's law states that two electric charges, q and Q, separated by a distance r will experience a force

$$\mathbf{F}(r) = \frac{Qq}{4\pi\epsilon_0 r^2}\hat{\mathbf{r}} \tag{5.2}$$

with the electric constant ϵ_0 a fundamental constant of nature that determines the inverse strength of the electrostatic force. The extra factor 4π reflects the fact that in the century between the Newton and Coulomb people had figured out where factors of 4π should sit in equations.

Most likely it will not have escaped your attention that these two equations are essentially the same. The only real difference is that overall minus sign which tells us that two masses always attract while two like charges repel. The question that we would like to ask is: why are the forces so similar?

Certainly it's not true that there is a deep connection between gravity and the electrostatic force, at least not one that we've uncovered to date. In particular, when masses and charges start to move, both the forces described above are replaced by something different and more complicated – general relativity in the case of gravity, the full Maxwell equations (3.7) in the case of the Coulomb force – and the equations of these theories are very different from each other. Yet, when we restrict to the simple, static set-up, the forces take the same form.

The reason for this is twofold. First, both forces are described by fields. Second, space has three dimensions. The purpose of this section is to explain this in more detail. And, for this, we need the tools of vector calculus.

5.1.1 Gauss' Law

Each of the force equations (5.1) and (5.2) contains some property that characterises the force: mass for gravity and electric charge for the electrostatic force. For our purposes, it will be useful to focus on one of the particles that carries mass m and charge q. We call this a *test particle*, meaning that we'll look at how this particle is buffeted by various forces but won't, in turn, consider its effect on any other particle. Physically, this is appropriate if $m \ll M$ and $q \ll Q$. Then it is useful to write the equation in a way that separates the properties of the test particle from the other. The force experienced by the test particle is

$$\mathbf{F}(\mathbf{x}) = m\mathbf{g}(\mathbf{x}) + q\mathbf{E}(\mathbf{x})$$

where $\mathbf{g}(\mathbf{x})$ is the gravitational and $\mathbf{E}(\mathbf{x})$ is the electric field. Clearly Newton's law is telling us that a particle of mass M sets up a gravitational field

$$\mathbf{g}(\mathbf{x}) = -\frac{GM}{r^2}\hat{\mathbf{r}} \tag{5.3}$$

while a particle with electric charge Q sets up an electric field

$$\mathbf{E}(\mathbf{x}) = \frac{Q}{4\pi\epsilon_0 r^2}\hat{\mathbf{r}} \tag{5.4}$$

So far this is just a trivial rewriting of the force laws. However, we will now reframe these force laws in the language of vector calculus. Instead of postulating the $1/r^2$ force laws (5.3) and (5.4), we will replace them by two properties of the fields from which everything else follows. Here we specify the first property; the second will be explained in Section 5.1.2.

The first property is that if you integrate the relevant field over a closed surface, then it captures the amount of "stuff" inside this surface. For the gravitational field, this stuff is mass

$$\int_S \mathbf{g} \cdot d\mathbf{S} = -4\pi GM \tag{5.5}$$

while for the electric field it is charge

$$\int_S \mathbf{E} \cdot d\mathbf{S} = \frac{Q}{\epsilon_0} \tag{5.6}$$

Again, the difference in minus sign signals the important attractive/repulsive difference between the two forces. In contrast, the factors of $4\pi G$ and $1/\epsilon_0$ are simply convention for how we characterise the strength of the fields. These two equations are known as *Gauss' law*. Or, more precisely, "Gauss' law in integrated form". We'll see the other form below.

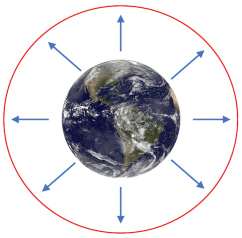
Examples

For concreteness, let's focus on the gravitational field. We will take a sphere of radius R and total mass M. We will require that the density of the sphere is spherically symmetric, but not necessarily constant. The spherical symmetry of the problem then ensures that the gravitational field itself is spherically symmetric, with $\mathbf{g}(\mathbf{x}) = g(r)\hat{\mathbf{r}}$. If we then integrate the gravitational field over any spherical surface S of radius $r > R$, we have

$$\int_S \mathbf{g} \cdot d\mathbf{S} = \int_S g(r) dS = 4\pi r^2 g(r)$$

where we recognise $4\pi r^2$ as the area of the sphere. From Gauss' law (5.5) we then have

$$\mathbf{g}(r) = -\frac{GM}{r^2}\hat{\mathbf{r}} \tag{5.7}$$



This reproduces Newton's force law (5.1). Note, however, that we've extended Newton's law beyond the original remit of point particles: the gravitational field (5.7) holds for any spherically symmetric distribution of mass, provided that we're outside this mass. For example, it tells us that the gravitational field of the Earth (at least assuming spherical symmetry) is indistinguishable from the gravitational field of a point-like particle with the same mass, sitting at the origin. This way of solving for the vector field is known as the *Gauss flux method*.

Another rather cute consequence of this is that, at least for spherically symmetric mass distributions, you don't feel the mass outside you. According to Gauss' law, the gravitational field at any point is determined only by what lies inside a sphere of a given radius. So if, for example, you were able to hollow out the centre of a planet (unlikely, admittedly) then anyone living there would feel no gravitational force from the mass that surrounds them.

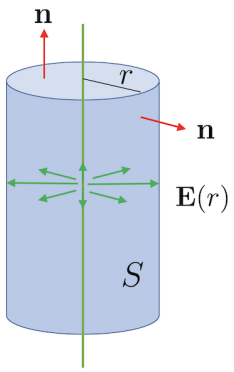
For our second example, we turn to the electric field. Consider an infinite line of charge, with charge per unit length σ. This situation is crying out for cylindrical polar coordinates. Until now, we've always called the radial direction in cylindrical polar coordinates ρ but, for reasons that will become clear shortly, for this example alone we will instead call the radial direction r as shown in the figure. The symmetry of the problem shows that the electric field is radial so takes the form $\mathbf{E}(r) = E(r)\hat{\mathbf{r}}$. Integrating over cylinder S of radius r and length L we have

$$\int_S \mathbf{E} \cdot d\mathbf{S} = 2\pi r L E(r)$$

where there is no contribution from the end caps because $\mathbf{n} \cdot \mathbf{E} = 0$ there, with $\mathbf{n}$ the normal vector. The total charge inside this surface is $Q = \sigma L$. From Gauss' law (5.6), we then have the electric field

$$\mathbf{E}(r) = \frac{\sigma}{2\pi\epsilon_0 r}\hat{\mathbf{r}}$$

Note that the $1/r$ behaviour arises because the symmetry of the problem ensures that the electric field lies in a plane. Said differently, the electric field from an infinite charged line is the same as we would get from a point particle in a flatland world of two dimensions.



More generally, if space were $\mathbb{R}^n$, then the Gauss' law equations (5.5) and (5.6) would still be the correct description of the gravitational and electric fields. Repeating the calculations above would then tell us that a point charge gives rise to an electric field

$$\mathbf{E}(r) = \frac{1}{A_{n-1}\epsilon_0 r^{n-1}}\hat{\mathbf{r}}$$

where $A_n r^n$ is the "surface area" of an n-dimensional sphere S^n of radius r. (For what it's worth, the prefactor is $A_{n-1} = 2\pi^{n/2}/\Gamma(n/2)$ where $\Gamma(x)$ is the gamma function which coincides with the factorial function $\Gamma(x) = (x-1)!$ when x is integer.) For the rest of this section, we'll keep our feet firmly in $\mathbb{R}^3$.

Gauss' Law Again

There's a useful way to rewrite the Gauss' law equations (5.5) and (5.6). For the gravitational field, we introduce the *density*, or mass per unit volume, $\rho(\mathbf{x})$. Invoking the divergence theorem then, for any volume V bounded by S, we have

$$\int_V \nabla \cdot \mathbf{g}\, dV = \int_S \mathbf{g} \cdot d\mathbf{S} = -4\pi GM = -4\pi G \int_V \rho(\mathbf{x})\, dV$$

But, rearranging, we have

$$\int_V \Big(\nabla \cdot \mathbf{g} + 4\pi G\rho(\mathbf{x})\Big) dV = 0$$

for any volume V. This can only hold if the integrand itself vanishes, so we must have

$$\nabla \cdot \mathbf{g} = -4\pi G\rho(\mathbf{x}) \tag{5.8}$$

This is also known as *Gauss' law* for the gravitational field, now in differential form. The equivalence with the earlier integrated form (5.5) follows, as above, from the divergence theorem.

We can apply the same manipulations to the electric field. This time we introduce the *charge density* $\rho_e(\mathbf{x})$. We then get Gauss' law in the form

$$\nabla \cdot \mathbf{E} = \frac{\rho_e(\mathbf{x})}{\epsilon_0} \tag{5.9}$$

This is the first of the Maxwell equations (3.7). (In our earlier expression, we denoted the charge density as $\rho(\mathbf{x})$. Here we've added the subscript ρ_e to distinguish it from mass density.) The manipulations that we've described above show that Gauss' law is a grown-up version of the Coulomb force law (5.2).

5.1.2 Potentials

In our examples above, we used symmetry arguments to figure out the direction in which the gravitational and electric fields are pointing. But in many situations we don't have that luxury. In that case, we need to invoke the second important property of these vector fields: they are both conservative.

Recall that, by now, we have a number of different ways to talk about conservative vector fields. Such fields are necessarily irrotational $\nabla \times \mathbf{g} = \nabla \times \mathbf{E} = 0$. Furthermore, their integral vanishes when integrated around any closed curve C,

$$\oint_C \mathbf{g} \cdot d\mathbf{x} = \oint_C \mathbf{E} \cdot d\mathbf{x} = 0$$

You can check that both of these hold for the examples, such as the $1/r^2$ field, that we discussed above (as long as the path C avoids the singular point at the origin).

Here the key property of a conservative vector field is that it can be written in terms of an underlying scalar field,

$$\mathbf{g} = -\nabla\Phi \quad \text{and} \quad \mathbf{E} = -\nabla\phi \tag{5.10}$$

where $\Phi(\mathbf{x})$ is the gravitational potential and $\phi(\mathbf{x})$ the electrostatic potential. Note the additional minus signs in these definitions. We saw in the discussion around (1.18) that the existence of such potentials ensures that test particles experiencing these forces have a conserved energy:

$$\text{energy} = \frac{1}{2}m\dot{\mathbf{x}}^2 + m\Phi(\mathbf{x}) + q\phi(\mathbf{x})$$

Combining the differential form of the Gauss' law (5.8) and (5.9) with the existence of the potentials (5.10), we find that the gravitational and electric fields are determined, in general, by solutions to the following equations

$$\nabla^2\Phi = 4\pi G\rho(\mathbf{x}) \quad \text{and} \quad \nabla^2\phi = -\frac{\rho_e(\mathbf{x})}{\epsilon_0}$$

Equations of this type are known as the *Poisson equation*. In the special case where the "source" $\rho(\mathbf{x})$ on the right-hand side vanishes, this reduces to the *Laplace equation*, for example

$$\nabla^2\Phi = 0$$

These two equations are commonplace in mathematics and physics. Here we have derived them in the context of gravity and electrostatics, but their applications spread much further.

To give just one further example, in Fluid Mechanics the motion of the fluid is described by a velocity field $\mathbf{u}(\mathbf{x})$. If the flow is irrotational, then $\nabla \times \mathbf{u} = 0$ and the velocity can be described by a potential function $\mathbf{u} = \nabla\phi$. If, in addition, the fluid is incompressible then $\nabla \cdot \mathbf{u} = 0$ and we once again find ourselves solving the Laplace equation $\nabla^2\phi = 0$.

5.2 The Poisson and Laplace Equations

In the rest of this section we will develop some methods to solve the Poisson equation. We change notation and call the potential $\psi(\mathbf{x})$ (to avoid confusion with the polar angle ϕ). We are then looking for solutions to

$$\nabla^2\psi(\mathbf{x}) = -\rho(\mathbf{x})$$

The goal is to solve for $\psi(\mathbf{x})$ given a "source" $\rho(\mathbf{x})$. As we will see, the domain in which $\psi(\mathbf{x})$ lives, together with associated boundary conditions, also plays an important role in determining $\psi(\mathbf{x})$.

The Laplace equation $\nabla^2\psi = 0$ is linear. This means that if $\psi_1(\mathbf{x})$ is a solution and $\psi_2(\mathbf{x})$ is a solution, then so too is $\psi_1(\mathbf{x}) + \psi_2(\mathbf{x})$. Any solution to the Laplace equation acts as a complementary solution to the Poisson equation. This should then be accompanied by a particular solution for a given source $\rho(\mathbf{x})$ on the right-hand side.

5.2.1 Isotropic Solutions

Both the Laplace and Poisson equations are partial differential equation. Life is generally much easier if we're asked to solve ordinary differential equations rather than partial differential equations. For the Poisson equation, this is what we get if we have some kind of symmetry, typically one aligned to some polar coordinates.

For example, if we have spherical symmetry then we can look for solutions of the form $\psi(\mathbf{x}) = \psi(r)$. Using the form of the Laplacian (3.15), Laplace equation becomes

$$\begin{aligned} \nabla^2\psi = 0 \quad &\Rightarrow \quad \frac{d^2\psi}{dr^2} + \frac{2}{r}\frac{d\psi}{dr} = \frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\psi}{dr}\right) = 0 \\ &\Rightarrow \quad \psi(r) = \frac{A}{r} + B \end{aligned} \tag{5.11}$$

for some constants A and B. Clearly the A/r solution diverges as $r \to 0$ so we should be cautious in claiming that this solves the Laplace equation at $r = 0$. (We will shortly see that it doesn't, but it does solve a related Poisson equation.) Note that the solution A/r is relevant in gravity or in electrostatics, where $\psi(r)$ has the interpretation as the potential for a point charge.

Meanwhile, in cylindrical polar coordinates we will also denote the radial direction as r to avoid confusion with the source ρ in the Poisson equation. The Laplace equation becomes

$$\begin{aligned} \nabla^2\psi = 0 \quad &\Rightarrow \quad \frac{d^2\psi}{dr^2} + \frac{1}{r}\frac{d\psi}{dr} = \frac{1}{r}\frac{d}{dr}\left(r\frac{d\psi}{dr}\right) = 0 \\ &\Rightarrow \quad \psi(r) = A\log r + B \end{aligned} \tag{5.12}$$

This again diverges at $r = 0$, this time corresponding to the entire z axis.

Note that if we ignore the z direction, as we have above, then cylindrical polar coordinates are the same thing as 2d polar coordinates, and the log form is the rotationally invariant solution to the Laplace equation in $\mathbb{R}^2$. In general, in $\mathbb{R}^n$, the non-constant solution to the Laplace equation is $1/r^{n-2}$. The low dimensions of $\mathbb{R}^2$ and $\mathbb{R}$ are special because the solution grows asymptotically as $r \to \infty$, while for $\mathbb{R}^n$ with $n \geq 3$, the rotationally invariant solution to the Laplace equation decays to a constant asymptotically.

If $\psi(r)$ is a solution to the Laplace equation, then so too is any derivative of $\psi(r)$. For example, if we take the spherically symmetric solution $\psi(r) = 1/r$, then we can construct a new solution

$$\psi_{\rm dipole}(\mathbf{x}) = \mathbf{d} \cdot \nabla \left(\frac{1}{r}\right) = -\frac{\mathbf{d} \cdot \mathbf{x}}{r^3}$$

for any constant vector $\mathbf{d}$ and, again, with $r \neq 0$. This kind of solution in important in electrostatics where it arises as the large distance solution for a *dipole*, two equal and opposite charges at a fixed distance apart.

Discontinuities and Boundary Conditions

In many situations, we must specify some further data when solving the Poisson equations. Typically this is some kind of boundary condition and, in some circumstances, a requirement of continuity and smoothness on the solution.

This can be illustrated with a simple example. Suppose that we are looking for a spherically symmetric solution to:

$$\nabla^2 \psi = \begin{cases} -\rho_0 & r \leq R \\ 0 & r > R \end{cases}$$

with ρ_0 constant. We will further ask that $\psi(r = 0)$ is non-singular, that $\psi(r) \to 0$ as $r \to \infty$, and that $\psi(\mathbf{x})$ and $\psi'(\mathbf{x})$ are continuous. We will now see that all of these conditions give us a unique solution.

First look inside $r \leq R$. As we mentioned above, a solution to the Poisson equation can be found by adding a complementary solution and a particular solution. Since we're looking for a spherically symmetric particular solution, we can restrict our ansatz to $\psi(r) = r^p$ for some p. It's simple to check that $\nabla^2 r^p = p(p+1)r^{p-2}$. This then gives us the general solution

$$\psi(r) = \frac{A}{r} + B - \frac{1}{6}\rho_0 r^2 \qquad r \leq R$$

But now we can start killing some terms by invoking the boundary conditions. In particular, the requirement that $\psi(r)$ is non-singular at $r = 0$ tells us that we must have $A = 0$. Meanwhile, outside $r > R$ the most general solution is

$$\psi(r) = \frac{C}{r} + D$$

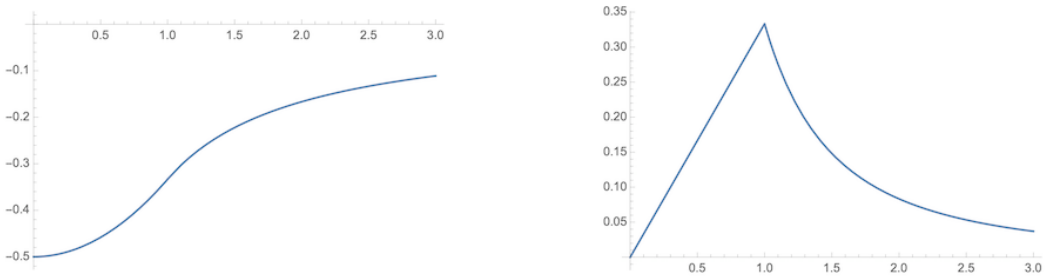


Figure 17. The plot of $\Phi = -4\pi G\psi$ on the left, with the radius $R = 1$ the cross over point. This is more apparent in the gravitational field $g = -\Phi'$ shown on the right.

Now we must have $D = 0$ if $\psi(r) \to 0$ as $r \to \infty$. To finish, we must patch these two solutions at $r = R$, invoking continuity

$$\psi(r = R) = B - \frac{1}{6}\rho_0 R^2 = \frac{C}{R}$$

and smoothness

$$\psi'(r = R) = -\frac{1}{3}\rho_0 R = -\frac{C}{R^2}$$

These determine our last two unknown constants, B and C. Putting this together, we have a unique solution

$$\psi(r) = \begin{cases} \frac{1}{6}\rho_0(3R^2 - r^2) & r \leq R \\ \frac{1}{3}\rho_0 R^3/r & r > R \end{cases}$$

This example has application for the gravitational potential $\Phi = -4\pi G\psi$ of a planet of radius R and density ρ_0. The plot of Φ is shown on the left of Figure 17; the plot of the gravitational field $g = -d\Phi/dr$ is on the right, where we see a linear increase inside the planet, before we get to the more familiar $1/r^2$ fall-off.

5.2.2 Some General Results

So far our solutions to the Poisson equation take place in $\mathbb{R}^3$. (Or, more precisely, $\mathbb{R}^3 - \{0, 0\}$ for the $1/r$ solution (5.11) and $\mathbb{R}^3 - \mathbb{R}$ for the $\log r$ solution (5.12).) In general, we may want to solve the Poisson or Laplace equations $\nabla^2\psi = -\rho$ in some bounded region V. In that case, we must specify boundary conditions on ∂V.

There are two common boundary conditions:

- Dirichlet condition: We fix $\psi(\mathbf{x}) = f(\mathbf{x})$ for some specific $f(\mathbf{x})$ on ∂V.

- Neumann condition: We fix $\mathbf{n}\cdot\nabla\psi(\mathbf{x}) = g(\mathbf{x})$ for some specific $g(\mathbf{x})$ on ∂V, where $\mathbf{n}$ is the outwardly pointing normal of ∂V.

The Neumann boundary condition is sometimes specified using the slightly peculiar notation $\partial\psi/\partial\mathbf{n} := \mathbf{n}\cdot\nabla\psi$. Or even, sometimes, $\partial\psi/\partial n$. We have the following statement of uniqueness:

Claim: Consider the Poisson equation on a bounded region V, with either Dirichlet or Neumann boundary conditions specified on each boundary ∂V. If a solution exists, then it is unique. (In the case of Neumann boundary conditions everywhere, the solution is only unique up to a constant.)

Proof: Let $\psi_1(\mathbf{x})$ and $\psi_2(\mathbf{x})$ both satisfy the Poisson equation with the specified boundary conditions. Then $\psi(\mathbf{x}) = \psi_1 - \psi_2$ obeys $\nabla^2\psi = 0$ and either $\psi = 0$ or $\mathbf{n}\cdot\nabla\psi = 0$ on ∂V. Then consider

$$\int_V \nabla\cdot(\psi\nabla\psi)\,dV = \int_V \left(\nabla\psi\cdot\nabla\psi + \psi\nabla^2\psi\right)\,dV = \int_V |\nabla\psi|^2 dV$$

But by the divergence theorem, we have

$$\int_V \nabla\cdot(\psi\nabla\psi)\,dV = \int_{\partial V} \psi\nabla\psi\cdot d\mathbf{S} = \int_{\partial V} \psi(\mathbf{n}\cdot\nabla\psi)\,dS = 0$$

where either Dirichlet or Neumann boundary conditions set the boundary term to zero. Because $|\nabla\psi|^2 \geq 0$, the integral can only vanish is $\nabla\psi = 0$ everywhere in V, so ψ must be constant. If Dirichlet boundary conditions are imposed anywhere, then that constant must be zero. □

This result means that if we can find any solution – say an isotropic solution, or perhaps a separable solution of the form $\psi(\mathbf{x}) = \Phi(r)Y(\theta)$ – then this must be the unique solution. By considering the limit of large spheres, it is also possible to extend the proof to solutions on $\mathbb{R}^3$, with the boundary condition $\psi(\mathbf{x}) \to 0$ suitably quickly as $r \to \infty$.

Note, however, that this doesn't necessarily tell us that a solution exists. For example, suppose that we wish to solve the Poisson equation $\nabla^2\psi = \rho(\mathbf{x})$ with a fixed Neumann boundary condition $\mathbf{n}\cdot\nabla\psi = g(\mathbf{x})$ on ∂V. Then there can only be a solution provided that there is a particular relationship between ρ and g,

$$\int_V \nabla^2\psi\,dV = \int_{\partial V} \nabla\psi\cdot d\mathbf{S} \quad\Longleftrightarrow\quad \int_V \rho\,dV = \int_S g\,dS$$

In other situations, there may well be other requirements.

If the region V has several boundaries, it's quite possible to specify a different type of boundary condition on each, and the uniqueness statement still holds. This kind of problem arises in electromagnetism where you solve for the electric field in the presence of a bunch of "conductors" (for now, conductors just means a chunk of metal). The electric field vanishes inside a conductor because, if it didn't, then the electric charges inside would move around until they created a counterbalancing field. So any attempt to solve for the electric field outside the conductors must take this into account by imposing certain boundary conditions on the surface of the conductor. It turns out that both Dirichlet and Neumann boundary conditions are important here. If the conductor is "grounded", meaning that it is attached to some huge reservoir of charge like the Earth, then it sits at some fixed potential, typically $\psi = 0$. This is a Dirichlet boundary condition. In contrast, if the conductor is isolated and carries some non-vanishing charge then it will act as a source of electric field, but this field is always emitted perpendicular to the boundary. This, then, specifies $\mathbf{n} \cdot \mathbf{E} = -\mathbf{n} \cdot \nabla\psi$, giving Neumann boundary conditions. You can learn more about this in the lectures on Electromagnetism.

Green's Identities

The proof of the uniqueness theorem used a trick known as Green's (first) identity, namely

$$\int_V \phi\nabla^2\psi \, dV = -\int_V \nabla\phi \cdot \nabla\psi \, dV + \int_S \phi\nabla\psi \cdot d\mathbf{S}$$

This is essentially a 3d version of integration by parts and it follows simply by applying the divergence theorem to $\phi\nabla\psi$. We used it in the above proof with $\phi = \psi$, but the more general form given above is sometimes useful, as is a related formula that follows simply by anti-symmetrisation,

$$\int_V \left(\phi\nabla^2\psi - \psi\nabla^2\phi\right) dV = \int_S \left(\phi\nabla\psi - \psi\nabla\phi\right) \cdot d\mathbf{S}$$

This is known as Green's second identity.

Harmonic Functions

Solutions to the Laplace equation

$$\nabla^2\psi = 0$$

arise in many places in mathematics and physics. These solutions are so special that they get their own name: they are called *harmonic functions*. Here are two properties

of these functions

Claim: Suppose that ψ is harmonic in a region V that includes the solid sphere with boundary $S_R : |\mathbf{x} - \mathbf{a}| = R$. Then the value of ψ at $\mathbf{a}$, the centre of the sphere, is given by $\psi(\mathbf{a}) = \bar{\psi}(R)$ where

$$\bar{\psi}(R) = \frac{1}{4\pi R^2} \int_{S_R} \psi(\mathbf{x})\, dS$$

is the average of ψ over S_R. This is known as the *mean value property.*

Proof: In spherical polar coordinates centred on $\mathbf{a}$, the area element is $dS = r^2 \sin\theta d\theta\, d\phi$, so

$$\bar{\psi}(r) = \frac{1}{4\pi} \int d\phi \int d\theta\ \sin\theta\, \psi(r, \theta, \phi)$$

and

$$\begin{aligned} \frac{d\bar{\psi}(R)}{dr} &= \frac{1}{4\pi} \int d\phi \int d\theta\ \sin\theta\, \frac{\partial \psi(R)}{\partial r} = \frac{1}{4\pi R^2} \int_{S_R} \frac{\partial \psi(R)}{\partial r} dS \\ &= \frac{1}{4\pi R^2} \int_{S_R} \nabla\psi \cdot d\mathbf{S} = \int_{\text{Ball}} \nabla^2 \psi\, dV = 0 \end{aligned}$$

But clearly $\bar{\psi}(R) \to \psi(\mathbf{a})$ as $R \to 0$ so we must have $\bar{\psi}(R) = \psi(\mathbf{a})$ for all R. □

Claim: A harmonic function can have neither a maximum nor minimum in the interior of a region V. Any maximum of minimum must lie on the boundary ∂V.

Proof: If ψ has a local maximum at $\mathbf{a}$ in V then there exists an ϵ such that $\psi(\mathbf{x}) < \psi(\mathbf{a})$ for all $|\mathbf{x} - \mathbf{a}| < \epsilon$. But, we know that $\bar{\psi}(R) = \psi(\mathbf{a})$ and this contradicts the assumption for any $0 < R < \epsilon$. □

This is consistent with our standard analysis of maxima and minima. Usually we would compute the eigenvalues λ_i of the Hessian $\partial^2\psi/\partial x^i \partial x^j$. For a harmonic function $\nabla^2\psi = \partial^2\psi/\partial x^i \partial x^i = 0$. Since the trace of the Hessian vanishes, we must have eigenvalues of opposite sign since $\sum_i \lambda_i = 0$. Hence, any stationary point must be a saddle. Note that this standard analysis is inconclusive when $\lambda_i = 0$, but the argument using the mean value property closes this loophole.

5.2.3 Integral Solutions

There is a particularly nice way to write down an expression for the general solution to the Poisson equation in $\mathbb{R}^3$, with

$$\nabla^2\psi = -\rho(\mathbf{x}) \qquad (\text{  })$$

at least for a localised source $\rho(\mathbf{x})$ that drops off suitably fast, so $\rho(\mathbf{x}) \to 0$ as $r \to \infty$.

To this end, let's look back to what is, perhaps, our simplest "solution",

$$\psi(\mathbf{x}) = \frac{\lambda}{4\pi r} \tag{5.13}$$

for some constant λ. The question we want to ask is: what equation does this actually solve?! We've seen in (5.11) that it solves the Laplace equation $\nabla^2\psi = 0$ when $r \neq 0$. But clearly something's going on at $r = 0$ because the function diverges there. In the language of physics, we would say that there is a point particle sitting at $r = 0$, carrying some mass or charge, giving rise to this potential. What is the correct mathematical way of capturing this?

To see that there must be something going on at $r = 0$, let's replay the kind of Gauss flux games that we met in Section 5.1. We integrate $\nabla^2\psi$, with ψ given by (5.13), over a volume V which we take to be a spherical region of radius R, to find

$$\int_V \nabla^2\psi \, dV = \int_S \nabla\psi \cdot d\mathbf{S} = -\lambda$$

Comparing to (), we see that the function (5.13) must solve the Poisson equation with a source and this source must obey

$$\int_V \rho(\mathbf{x}) \, dV = \lambda$$

This makes sense physically, since $\int \rho dV$ is the total mass, or total charge, which does indeed determine the overall scaling λ of the potential. But what mathematical function obeys $\rho(\mathbf{x}) = 0$ for all $\mathbf{x} \neq 0$ yet, when integrated over all space, gives a non-vanishing constant λ?

The answer is that $\rho(\mathbf{x})$ must be proportional to the 3d *Dirac delta function*,

$$\rho(\mathbf{x}) = \lambda\, \delta^3(\mathbf{x})$$

The Dirac delta function should be thought of as an infinitely narrow spike, located at the origin. It has the properties

$$\delta^3(\mathbf{x}) = 0 \quad \text{for} \quad \mathbf{x} \neq 0$$

and, when integrated against any function $f(\mathbf{x})$ over any volume V that includes the origin, it gives

$$\int_V f(\mathbf{x})\, \delta^3(\mathbf{x})\, dV = f(\mathbf{x} = \mathbf{0})$$

The superscript in $\delta^3(\mathbf{x})$ is there to remind us that the delta function should be integrated over a 3-dimensional volume before it yields something finite. In particular, when integrated against a constant function, we get a measure of the height of the spike,

$$\int_V \delta^3(\mathbf{x})\, dV = 1$$

The Dirac delta function is an example of a generalised function, also known as a distribution. And it is exactly what we need to source the solution $\psi \sim 1/r$. We learn that the function (5.13) is not a solution to the Laplace equation, but rather a solution to the Poisson equation with a delta function source

$$\nabla^2 \psi = -\lambda\, \delta^3(\mathbf{x}) \quad \Rightarrow \quad \psi(\mathbf{x}) = \frac{\lambda}{4\pi r} \tag{5.14}$$

With this important idea in hand, we can now do something quite spectacular: we can use it to write down an expression for a solution to the general Poisson equation.

Claim: The Poisson equation () has the integral solution

$$\psi(\mathbf{x}) = \frac{1}{4\pi} \int_{V'} \frac{\rho(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}\, dV' \tag{5.15}$$

where the integral is over a region V' parameterised by $\mathbf{x}'$.

Proof: First, some simple intuition behind this formula. A point particle at $\mathbf{x}'$ gives rise to a potential of the form $\psi(\mathbf{x}) = \rho(\mathbf{x}')/4\pi|\mathbf{x}-\mathbf{x}'|$, which is just our solution (5.14), translated from the origin to point $\mathbf{x}'$. The integral solution (5.15) then just takes advantage of the linear nature of the Poisson equation and sums a whole bunch of these solutions.

The technology of the delta function allows us to make this precise. We can evaluate

$$\nabla^2\psi = \frac{1}{4\pi}\int_{V'} \rho(\mathbf{x}')\,\nabla^2\left(\frac{1}{|\mathbf{x}-\mathbf{x}'|}\right)\,dV'$$

where you have to remember that ∇^2 differentiates $\mathbf{x}$ and cares nothing for $\mathbf{x}'$. We then have the result

$$\nabla^2\,\frac{1}{|\mathbf{x}-\mathbf{x}'|} = -4\pi\delta^3(\mathbf{x}-\mathbf{x}')$$

which is just a repeat of (5.14), but with the location of the source translated from the origin to the new point $\mathbf{x}'$. Using this, we can continue our proof

$$\nabla^2\psi = -\int_{V'} \rho(\mathbf{x}')\,\delta^3(\mathbf{x}-\mathbf{x}')\,dV' = -\rho(\mathbf{x})$$

which is what we wanted to show. □

The technique of first solving an equation with a delta function source and subsequently integrating to find the general solution is known as the *Green's function* approach. It is a powerful method to solve differential equations and we will meet it again in many further courses.

6 Tensors

A famously annoying definition of a tensor is:

> A tensor is something whose components transform like a tensor

This becomes even more annoying when you appreciate that this is, in fact, one of the better definitions of a tensor. The purpose of this section is to explain why this definition is not as dumb as it sounds and to give some insight into what it means to be a tensor.

Very roughly speaking, tensors are generalisations of objects like vectors and matrices. In index notation, a vector has a single index while a matrix has two indices. A tensor is an object with any number of indices, something like $T_{ij\ldots k}$.

However, this simplistic description hides the most important property of a tensor. Vectors, matrices and, more generally, tensors are more than just a list of numbers. Instead, those numbers should be thought of as a useful way of characterising the underlying object and, because of this, inherit some properties of that underlying object. As we will see, the key property is how the list of numbers transform under a change of basis.

We will start by explaining this in more detail, firstly with vectors and then building up to the definition of a tensor. Initially we will keep the discussion restricted to some (admittedly rather dry) mathematical formalism. Then, in Section 6.2 we will describe some physical examples.

6.1 What it Takes to Make a Tensor

Not any list of n numbers constitutes a vector in $\mathbb{R}^n$. Or, said more precisely, not any list of n numbers constitutes the components of a vector in $\mathbb{R}^n$. For example, if you write down the heights of the first three people you met this morning, that doesn't make a vector in $\mathbb{R}^3$. Instead, a vector comes with certain responsibilities. In particular, the components describe an underlying object which should be independent of the choice of basis. As we now explain, that means that the components should transform in the right way under rotations.

We consider a point $\mathbf{x} \in \mathbb{R}^n$. If we wish to attach some coordinates to this point, we first need to introduce a set of basis vectors $\{\mathbf{e}_i\}$ with $i = 1, \ldots, n$. We will take these to be orthonormal, meaning that $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$. Any vector can then be expressed as

$$\mathbf{x} = x_i \mathbf{e}_i \tag{6.1}$$

Usually we conflate the components $x_i = (x_1, \ldots, x_n)$ with the "vector". But, for our purposes, we should remember that these are just a useful way of representing the more abstract object $\mathbf{x}$. In particular, we're entirely at liberty to take a different set of basis vectors,

$$\mathbf{e}'_i = R_{ij}\mathbf{e}_j$$

If we ask that $\mathbf{e}'_i$ are also orthonormal, so $\mathbf{e}'_i \cdot \mathbf{e}'_j = \delta_{ij}$, then we have

$$\mathbf{e}'_i \cdot \mathbf{e}'_j = R_{ik}R_{jl}\,\mathbf{e}_k \cdot \mathbf{e}_l = R_{ik}R_{jk} = \delta_{ij}$$

or, in matrix notation,

$$RR^T = \mathbb{1}$$

Matrices of this kind are said to be *orthogonal*. We write $R \in O(n)$. Taking the determinant, we have $\det R = \pm 1$. Those matrices with $\det R = +1$ correspond to rotations and are said to be *special orthogonal*. We write $R \in SO(n)$. In $\mathbb{R}^3$, a rotation $R \in SO(3)$ takes a right-handed orthonormal basis into another right-handed orthonormal basis. Those matrices with $\det R = -1$ correspond to a rotation together with a reflection and take a right-handed basis to a left-handed basis.

Under a change of basis, the vector $\mathbf{x}$ itself doesn't change. But its components do. We have

$$\mathbf{x} = x_i\mathbf{e}_i = x'_i\mathbf{e}'_i = x'_iR_{ij}\mathbf{e}_j$$

So the components transform under the same rotation matrix R,

$$x_j = R_{ij}x'_i \quad \Rightarrow \quad x'_i = R_{ij}x_j \tag{6.2}$$

A *tensor* T is a generalisation of these ideas to an object with more indices. Just as the vector $\mathbf{x}$ has an identity independent of any choice of basis, so too does the tensor T. But when measured with respect to a chosen basis $\{\mathbf{e}_i\}$, a *tensor of rank* p has components $T_{i_1\ldots i_p}$. When we change the basis using (6.1), the tensor transforms as

$$T'_{i_1\ldots i_p} = R_{i_1j_1}\ldots R_{i_pj_p}T_{j_1\ldots j_p} \tag{6.3}$$

This is known as the *tensor transformation rule*. A tensor of rank p is sometimes referred to simply as a p-tensor.

The simplest examples of tensors are very familiar. A tensor of rank 0 is just a number, or scalar, T. Here there's no requirement because a number doesn't change if you do a rotation: $T' = T$. So any single number can be said to be a tensor, although it isn't a particularly helpful designation.

A tensor of rank 1 is a vector. Here, however, it's important that the components of the vector transform as $T'_j = R_{ij}T_j$. If they don't transform in this way, then you don't have a tensor on your hands. You just have a bunch of numbers.

A tensor of rank 2 is a matrix that transforms as $T'_{ij} = R_{ik}R_{jl}T_{kl}$. Again, the transformation property is key. Just because you have an array of numbers A_{ij}, arranged in an $n \times n$ grid, doesn't mean that you have a 2-tensor. You have to check the transformation property holds. Otherwise, as with a vector, the array of numbers isn't a tensor; it's just a bunch of numbers.

What's a Tensor and What's Not?

It's worth elaborating on the definition of a tensor. For example, suppose that someone hands you a matrix, say

$$T_{ij} = \begin{pmatrix} 3 & 8 & 0 \\ 5 & -4 & 3 \\ 1 & 1 & 3 \end{pmatrix}$$

and asks you: "is this a tensor?". It's natural to answer yes. After all, it's written as T_{ij} which is the name we've given to a tensor. And it looks for all the world like a matrix. So is it a tensor? The answer is: we don't know. We haven't been given enough information[2]. As we've stressed several times, a tensor isn't just a bunch of numbers arranged in some pattern. This sometimes goes by the name of an *array* of numbers. Instead, we only know that a given array of numbers is a tensor if it transforms as (6.3). That means that we need to firstly know what basis the array of numbers above has been measured in. And then we need to know what the array looks like when measured in other bases. Only then do we have enough information to say whether this is a tensor or not. It's a tensor only if transforms as (6.3): this transformation law is the definition of a tensor.

[2]When I gave these lectures on 6th March 2023, I used the example of $\binom{7}{0}$ which, at least on that day, most certainly wasn't a tensor. It was a football score. A glorious, wonderful, humiliating, shameful football score.

Here's another example. In a given basis, the position of a point is given by x_i. We write this as the components of a vector

$$x_i = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

This is a tensor. Indeed, our starting point is that the components of this simple vector transforms in the tensorial way (6.2). This is just the statement that the components of this vector transform in the familiar way under rotation.

Suppose that you now square each of these elements and decide to write them as a column vector. We'll give it a fancy name Λ_i, complete with that hanging i index,

$$\Lambda_i = \begin{pmatrix} x^2 \\ y^2 \\ z^2 \end{pmatrix}$$

That i index makes this look for all the world like it's a tensor. But it's not. We know that after a rotation, $x_i \to x'_i = R_{ij}x_j$. This means that if we do a rotation and then measure the components of the array Λ'_i we get

$$\Lambda'_i = \begin{pmatrix} (R_{11}x + R_{12}y + R_{13}z)^2 \\ (R_{21}x + R_{22}y + R_{33}z)^2 \\ (R_{31}x + R_{32}y + R_{33}z)^2 \end{pmatrix}$$

But that's most definitely not how a tensor transforms! It's not the rule (6.3) that we wanted. The upshot is that Λ_i is not a tensor and it was a little bit naughty to write it as Λ_i because it suggests that it has some property that it doesn't.

Relatedly, this explains something that you may have wondered about in school. Suppose that you're given two vectors. You know that you can take an inner product to get a scalar, or you can take the cross-product to get another vector. But what stops you from doing something much simpler, just multiplying the component of one vector with the corresponding component of another vector to get a third vector. It seems like such an obvious thing to do. But it's a bad thing to do, precisely because the thing you end up with is not a tensor. It does not transform in the way (6.2), which is how components of a vector should transform.

There is a similar story for matrices. If you have two matrices, then there's a ridiculously complicated way to multiply them, multiplying rows with columns. Why don't we just do something much simpler and multiply entries together component by component? You've probably guessed the answer by now. If we started with genuine matrices, meaning that they transform (6.3), then the object that you get if you do proper matrix multiplication will also transform as (6.3), but the simpler, stupid way to multiplying will not.

Why are we making such a big deal about this? What is so special about things that transform nicely as (6.3) under rotations? Well, there are several answers to this, depending on taste. At the most basic level, if you're a physicist, then you might genuinely want to know how something looks in different, rotated frames of reference.

Moreover, once you realise that there's a preferred way for things to transform — the tensor way (6.3) – this brings some extra power to the calculations, a little like dimensional analysis. Suppose that you have an equation of the form "left-hand side" = "right-hand side". If the thing on the left is a tensor then the thing on the right better also be a tensor. And sometimes there's not many tensors available, which limits your options for what the thing on the right can actually be. We'll see an example of this in Section 6.1.3 when we'll use tensors to make some scary looking integrals a little more palatable.

The discussion above is very much from a physics perspective. But what about a pure maths perspective? This gives a more formal, but arguably cleaner, definition of a tensor. We'll explain this imminently in Section 6.1.1.

We'll meet a number of tensors as we proceed. But there is a one that is special: this is the rank 2 tensor δ_{ij} or, equivalently, the unit matrix. Importantly, it has the same 0 and 1 entries in any basis because, under the transformation (6.3), it becomes

$$\delta'_{ij} = R_{ik}R_{jl}\delta_{kl} = \delta_{ij}$$

We will devote Section 6.1.3 to "invariant tensors" which, like δ_{ij}, take the same form in any basis.

6.1.1 Tensors as Maps

There is something a little strange about the definition of a tensor given above. We first pick a set of coordinates, and the transformation law (6.3) then requires that the tensor transforms nicely so that, ultimately, nothing depends on these coordinates. But, if that's the case, surely there should be a definition of a tensor that doesn't rely on coordinates at all!

There is. A tensor T of rank p is a multi-linear map that takes p vectors, $\mathbf{a}, \mathbf{b}, \ldots, \mathbf{c}$ and spits out a number in $\mathbb{R}$,

$$T(\mathbf{a}, \mathbf{b}, \ldots, \mathbf{c}) = T_{i_1 i_2 \ldots i_p} a_{i_1} b_{i_2} \ldots c_{i_p} \tag{6.4}$$

Here "multi-linear" means that T is linear in each of the entries $\mathbf{a}, \mathbf{b}, \ldots, \mathbf{c}$ individually. By evaluating T on a all possible vectors $\mathbf{a}, \mathbf{b}, \ldots, \mathbf{c}$, we get the components $T_{i_1 i_2 \ldots i_p}$. The transformation rule (6.3) is simply the statement that the map T is independent of the choice of basis, and we can equally well write

$$\begin{aligned} T(\mathbf{a}, \mathbf{b}, \ldots, \mathbf{c}) &= T'_{i_1 i_2 \ldots i_p} a'_{i_1} b'_{i_2} \ldots c'_{i_p} \\ &= (R_{i_1 j_1} R_{i_2 j_2} \ldots R_{i_p j_p} T_{j_1 j_2 \ldots j_p})(R_{i_1 k_1} a_{k_1})(R_{i_2 k_2} b_{k_2}) \ldots (R_{i_p k_p} c_{k_p}) \\ &= T_{j_1 j_2 \ldots j_p} a_{j_1} b_{j_2} \ldots c_{j_p} \end{aligned}$$

which follows because $R^T R = \mathbb{1}$ or, in components, $R_{ij} R_{ik} = \delta_{jk}$. The key is that this formula takes the same form in any basis.

Tensors as Maps Between Vectors

Rather than thinking of a tensor as a map from many vectors to $\mathbb{R}$, you can equivalently think of it as a map from some lower-rank tensor to another. For example, in (6.4), if you don't fill in the first entry, then a rank p tensor can equally well be viewed as taking $(p-1)$ vectors and spitting out a single vector

$$a_i = T_{i\, j_1 \ldots j_{p-1}} b_{j_1} \ldots c_{j_{p-1}}$$

This is the way that tensors typically arise in physics or applied mathematics, where the most common example is simply a rank 2 tensor, defined as a map from one vector to another

$$\mathbf{u} = T\mathbf{v} \quad \Rightarrow \quad u_i = T_{ij} v_j$$

Until now, we've simply called T a matrix but for the equation $\mathbf{u} = T\mathbf{v}$ to make sense, T must transform as a tensor (6.3). This is inherited from the transformation rules of the vectors, $u'_i = R_{ij} u_j$ and $v'_i = R_{ij} v_j$, giving

$$u'_i = T'_{ij} v'_j \quad \text{with} \quad T'_{ij} = R_{ik} R_{jl} T_{kl}$$

Written as a matrix equation, this is $T' = RTR^T$.

6.1.2 Tensor Operations

Given a bunch of tensors, there are some manipulations that leave you with another tensor. Here we describe these operations.

- We can *add* and *subtract* tensors of the same rank, so if S and T are both tensors of rank p then so too is $S + T$. We can also multiply a tensor by a constant α and it remains a tensor.

- If S is a tensor of rank p and T a tensor of rank q, then the *tensor product* $S \otimes T$ is a tensor of rank $p + q$, defined by

$$(S \otimes T)_{i_1 \ldots i_p j_1 \ldots j_q} = S_{i_1 \ldots i_p} T_{j_1 \ldots j_q}$$

  You can check that the components of $(S \otimes T)$ do indeed satisfy the transformation rule (6.3). In particular, if we have p different vectors $\mathbf{a}$, $\mathbf{b}$, ..., $\mathbf{c}$ then we can construct a tensor

$$T = \mathbf{a} \otimes \mathbf{b} \otimes \ldots \otimes \mathbf{c} \quad \text{with} \quad T_{i_1 \ldots i_p} = a_{i_1} b_{i_2} \ldots c_{i_p}$$

- Given a tensor T of rank p, we can construct a new tensor S of rank $(p-2)$ by *contracting* on two indices using δ_{ij},

$$S_{k_1 \ldots k_{p-2}} = \delta_{ij} T_{ijk_1 \ldots k_{p-2}}$$

  For a rank 2 tensor, the contraction is what we call the trace, $\mathrm{Tr}\, T = T_{ii}$. It's a valid tensor operation because the end result is a scalar that does not transform under rotations

$$T'_{ii} = R_{ij} R_{ik} T_{jk} = \delta_{jk} T_{jk} = T_{jj}$$

  The same derivation shows that higher rank tensors can also be contracted, with the additional indices unaffected by the contraction.

  Combining a contraction with a tensor product gives a way to contract two different tensors together. For example, given a p-tensor P and q-tensor Q, we can form a $p + q - 2$ tensor by contracting, say, the first index on each to get $P_{ik_1 \ldots k_{p-1}} Q_{i l_1 \ldots l_{q-1}}$. This may sound abstract, but it's very much something you've seen before: given a pair of 1-tensors $\mathbf{a}$ and $\mathbf{b}$, also known as vectors, we can combine them to get a 0-tensor, also known as a number

$$\mathbf{a} \cdot \mathbf{b} = a_i b_i$$

  This, of course, is just the inner-product. It is a useful operation precisely because the 0-tensor on the right-hand side is, like all 0-tensors, independent of the choice of basis that we choose to express the vectors.

The Quotient Rule

In practice, it's not hard to recognise a tensor when you see one. In any setting, they're usually just objects with a bunch of i and j indices, each of which clearly transforms as a vector. If in doubt, you can just check explicitly how the thing transforms. (There are cases where this check is needed. In later courses, you'll meet an object called the Levi-Civita connection Γ^i_{jk} which looks for all the world like a tensor but turns out, on closer inspection, to be something more subtle.)

There is a more formal way to say this. Let $T_{i_1\ldots i_{p+q}}$ be a bunch of numbers that you think might comprise a tensor of rank $p+q$ in some coordinate basis. If $T_{i_1\ldots i_{p+q}}$ are indeed the components of a tensor then you can feed it a rank q tensor $u_{j_1\ldots j_q}$ and it will spit back a rank p tensor

$$v_{i_1\ldots i_p} = T_{i_i\ldots i_p j_1\ldots j_q} u_{j_1\ldots j_q} \tag{6.5}$$

There is a converse to this statement. If for every tensor $u_{j_1\ldots j_q}$, the output $v_{i_1\ldots i_p}$ defined in (6.5) is a tensor, then $T_{i_i\ldots i_p j_1\ldots j_q}$ are the components of a tensor. This is called the *quotient rule*.

It is straightforward, if a little fiddly, to prove the quotient rule. It's sufficient to restrict attention to tensors u formed from the tensor product of vectors $u_{j_1\ldots j_q} = c_{j_1}\ldots d_{j_q}$. Then, by assumption, $v_{i_1\ldots i_p} = T_{i_i\ldots i_p j_1\ldots j_q} u_{j_1\ldots j_q}$ is a tensor. If we then contract with p further vectors $\mathbf{a},\ldots,\mathbf{b}$ then $v_{i_1\ldots i_p} a_{i_1}\ldots b_{i_p} = T_{i_1\ldots i_p j_1\ldots j_q} a_{i_1}\ldots b_{i_p} c_{j_1}\ldots d_{j_q}$ is necessarily a scalar. This is then enough to ensure the correct transformation rule (6.3) for the components $T_{i_1\ldots i_p j_1\ldots j_q}$.

Symmetry and Anti-Symmetry

The symmetrisation properties of tensors are worthy of comment. A tensor that obeys

$$T_{ijp\ldots q} = \pm T_{jip\ldots q}$$

is said to be *symmetric* (for $+$) or *anti-symmetric* (for $-$) in the indices i and j. If a tensor is (anti)-symmetric in one coordinate system then it is (anti)-symmetric in any coordinate system

$$T'_{ijp\ldots q} = R_{ik}R_{jl}R_{pr}\ldots R_{qs}T_{klr\ldots s} = \pm R_{ik}R_{jl}R_{pr}\ldots R_{qs}T_{lkr\ldots s} = \pm T'_{jip\ldots q}$$

A tensor that is (anti)-symmetric in all pairs of indices is said to be *totally (anti)-symmetric*. Note that for tensors in $\mathbb{R}^n$, there are no anti-symmetric tensors of rank $p > n$ because at least one of the indices must take the same value and so the tensor necessarily vanishes. A totally anti-symmetric tensor of rank p in $\mathbb{R}^n$ has $\binom{n}{p}$ independent components.

Let's now restrict our attention to $\mathbb{R}^3$. A tensor of rank 2 is our new fancy name for a 3×3 matrix T_{ij}. In general, it has 9 independent components. We can always decompose it into the symmetric and anti-symmetric pieces

$$S_{ij} = \frac{1}{2}(T_{ij} + T_{ji}) \quad \text{and} \quad A_{ij} = \frac{1}{2}(T_{ij} - T_{ji})$$

which have 6 and 3 independent components respectively. Our discussion above shows that S and A are each, themselves, tensors. In fact, the symmetric piece can be decomposed further,

$$S_{ij} = P_{ij} + \frac{Q}{3}\delta_{ij}$$

where $Q = S_{ii}$ is the trace of S and carries a single degree of freedom, while P_{ij} is the traceless part of S and carries 5. The importance of this decomposition is that A, P and Q are individually tensors. In contrast, if you were to take, say, the upper-left-hand component of the original matrix T_{ij} then that doesn't form a tensor.

In $\mathbb{R}^3$, we can also rewrite an anti-symmetric matrix in terms of a vector,

$$A_{ij} = \epsilon_{ijk}B_k \quad \Longleftrightarrow \quad B_k = \frac{1}{2}\epsilon_{ijk}A_{ij}$$

The upshot is that in any 3×3 matrix can be decomposed as

$$T_{ij} = P_{ij} + \epsilon_{ijk}B_k + \frac{1}{3}\delta_{ij}Q \tag{6.6}$$

where $P_{ii} = 0$.

6.1.3 Invariant Tensors

There are two important invariant tensors in $\mathbb{R}^n$.

- We've met the first already: it is the rank 2 tensor δ_{ij}. As we noted previously, this is invariant because

  $$\delta'_{ij} = R_{ik}R_{jl}\delta_{kl} = \delta_{ij}$$

  Note that δ_{ij} is invariant under any $R \in O(n)$.

- The rank n totally anti-symmetric tensor $\epsilon_{i_1 \ldots i_n}$. This is defined by $\epsilon_{12\ldots n} = +1$. If you swap any two indices you get a minus sign. In particular, if any two indices are repeated, the epsilon symbol vanishes. This is invariant because

  $$\epsilon'_{i_1 \ldots i_n} = R_{i_1 j_1} \ldots R_{i_n j_n}\epsilon_{j_1 \ldots j_n} = \det R\, \epsilon_{i_1 \ldots i_n} = \epsilon_{i_1 \ldots i_n}$$

Note that the epsilon symbol is only invariant under $R \in SO(n)$ but it is not invariant under $R \in O(n)$ with $\det R = -1$. It picks up a minus sign under reflections. The invariance of ϵ_{ijk} in $\mathbb{R}^3$ is the reason why the cross-product $(\mathbf{a} \times \mathbf{b})_i = \epsilon_{ijk} a_j b_k$ is itself a vector. Or, said differently, why the triple product $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \epsilon_{ijk} a_i b_j c_k$ is independent of the choice of basis.

In general, a tensor is said to be invariant under a given rotation R if

$$T'_{i_1 \ldots i_n} = R_{i_1 j_1} \ldots R_{i_n j_n} T_{j_1 \ldots j_n} = T_{i_1 \ldots i_n}$$

A tensor that is invariant under all rotations R is said to be *isotropic*. Obviously all tensors of rank 0 are isotropic. What about higher rank tensors?

Claim: The only non-zero isotropic tensors in $\mathbb{R}^3$ of rank $p = 1, 2$ or 3 are $T_{ij} = \alpha \delta_{ij}$ and $T_{ijk} = \beta \epsilon_{ijk}$ with α and β constant. In particular, there are no isotropic tensors of rank 1 (essentially because a vector always points in a preferred direction).

Proof: The idea is simply to look at how tensors transform under a bunch of specific rotations by π or $\pi/2$ about certain axes.

For example, consider a tensor of rank 1, so that

$$T'_i = R_{ij} T_j \quad \text{with} \quad R_{ij} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & +1 \end{pmatrix} \tag{6.7}$$

Requiring $T'_i = T_i$ gives $T_1 = T_2 = 0$. Clearly a similar argument, using a different R, also gives $T_3 = 0$.

For a tensor of rank 2, consider the transformation

$$T'_{ij} = \tilde{R}_{ik} \tilde{R}_{jl} T_{kl} \quad \text{with} \quad \tilde{R}_{ij} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & +1 \end{pmatrix} \tag{6.8}$$

which is a rotation by $\pi/2$ about the z-axis. The rotation gives $T'_{13} = T_{23}$ and $T'_{23} = -T_{13}$ so if $T'_{ij} = T_{ij}$, we must have $T_{13} = T_{23} = 0$. Meanwhile $T'_{11} = T_{22}$. Similar arguments tell us that all off-diagonal elements must vanish and all diagonal elements must be equal: $T_{11} = T_{22} = T_{33} = \alpha$ for some α. Hence $T_{ij} = \alpha \delta_{ij}$.

Finally, for a rank 3 tensor we have

$$T'_{ijk} = R_{il}R_{jp}R_{kq}T_{lpq}$$

If we pick R given in (6.7), then we find $T'_{133} = -T_{133}$ and $T'_{111} = -T_{111}$. Similar arguments show that an isotropic tensor must have $T_{ijk} = 0$ unless i, j and k are all distinct. Meanwhile, if we pick $R = \tilde{R}$ given in (6.8), then we get $T'_{123} = -T_{213}$. We end up with the result we wanted: T_{ijk} is isotropic if and only if $T_{ijk} = \beta\epsilon_{ijk}$ for some constant β. □

Although we won't prove it here, all other isotropic tensors can be formed from δ_{ij} and ϵ_{ijk}. For example, the only isotropic 4-tensor in $\mathbb{R}^3$ is

$$T_{ijkl} = \alpha\delta_{ij}\delta_{kl} + \beta\delta_{ik}\delta_{jl} + \gamma\delta_{il}\delta_{jk}$$

with α, β and γ constants. You could try to cook up something involving ϵ_{ijk} but it doesn't give anything new. In particular, $\epsilon_{ijk}\epsilon_{ilp} = \delta_{jl}\delta_{kp} - \delta_{jp}\delta_{kl}$.

There is also an analogous result in $\mathbb{R}^n$: all isotropic tensors can be constructed from the symmetric 2-tensor δ_{ij} and the totally anti-symmetric n-tensor $\epsilon_{i_1\ldots i_n}$.

Invariant Integrals

It is sometimes possible to use invariance properties to immediately write down the index structure of an integral, without doing the hard work of evaluating everything term by term. Suppose that we have some integral of the form

$$T_{ij\ldots k} = \int_V f(r)x_i x_j \ldots x_k \, dV$$

with $r = |\mathbf{x}|$. Then under a rotation, we have

$$T'_{ij\ldots k} = R_{ip}R_{jq}\ldots R_{kr}T_{pq\ldots r} = \int_V f(r)x'_i x'_j \ldots x'_k \, dV$$

with, as usual, $x'_i = R_{ij}x_j$. But if we now change the integration variables to x', both $r = |\mathbf{x}| = |\mathbf{x}'|$ and $dV = dV'$ are invariant. (The latter because the Jacobian is $\det R = 1$). If the domain of integration is also rotationally invariant, so $V = V'$, then the final result must itself be an invariant tensor, $T'_{ij\ldots k} = T_{ij\ldots k}$.

Here are some examples. First, suppose that we have a 3d integral over the interior of a sphere of radius R, given by

$$T_i = \int_V \rho(r) x_i \, dV \tag{6.9}$$

This must be equal to some invariant 1-tensor (i.e. a vector), but there are no such objects. In other words, we can say immediately that $T_i = 0$. You can check this straightforwardly by doing the integral in, say, spherical polar coordinates.

Things change if we look at an integral with two hanging indices,

$$T_{ij} = \int_V \rho(r) x_i x_j \, dV \tag{6.10}$$

(In Section 6.2, we will find integrals of this form arising when we compute the inertia tensor of a sphere.) By the argument above T_{ij} must be an isotropic tensor and hence proportional to δ_{ij},

$$T_{ij} = \int_V \rho(r) x_i x_j \, dV = \alpha \delta_{ij}$$

for some α. If we take the trace, we get

$$\int_V \rho(r) r^2 \, dV = 3\alpha$$

Hence,

$$T_{ij} = \frac{1}{3}\delta_{ij} \int_V \rho(r) r^2 \, dV = \frac{4\pi}{3}\delta_{ij} \int_0^R dr \; \rho(r) r^4 \tag{6.11}$$

For example, if $\rho(r) = \rho_0$ is constant, then $T_{ij} = \frac{4}{15}\pi\rho_0 R^5 \delta_{ij}$.

Here's a slightly more complicated example (taken from the calculation of Stokes flow around a sphere in Fluid Mechanics). Consider the surface integral over a sphere of radius R,

$$\tilde{T}_k = a_j \int_{\mathbf{S}^2} dS_i \; \frac{x_i x_j x_k}{r^5}$$

This time we have a vector $\mathbf{a}$ in the game, so it must be the case that $\tilde{T}_k = \beta a_k$ for some constant β. One way to compute β is to strip off the vector $\mathbf{a}$ and instead look at

$$\tilde{T}_{jk} = \int_{\mathbf{S}^2} dS_i \; \frac{x_i x_j x_k}{r^5} = \beta \delta_{jk}$$

which now should be proportional to the invariant tensor δ_{jk} as shown, with the same coefficient β since $\tilde{T}_k = T_{jk}a_j = \beta a_k$. At this point, we again take the trace over the j and k indices to get

$$\int_{\mathbf{S}^2} dS_i \; \frac{x_i x_j x_j}{r^5} = 3\beta$$

But this integral is given by

$$\int_{\mathbf{S}^2} dS_i \; \frac{x_i}{r^3} = \int_{\mathbf{S}^2} d\mathbf{S} \cdot \frac{\mathbf{n}}{r^2} = 4\pi$$

and so we get $\beta = 4\pi/3$.

6.1.4 Tensor Fields

A tensor field over $\mathbb{R}^3$ is the assignment of a tensor $T_{i\ldots k}(\mathbf{x})$ to every point $\mathbf{x} \in \mathbb{R}^3$. This is the generalisation of a vector field

$$\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$$

to a map of the kind

$$T : \mathbb{R}^3 \to \mathbb{R}^m$$

with m the number of components of the tensor. So, for example, a map that assigns a symmetric, traceless rank 2 tensor $P_{ij}(\mathbf{x})$ to every point has $m = 5$.

The tensor field $T_{i\ldots k}(\mathbf{x})$ is sometimes denoted as $T_{i\ldots k}(x^l)$ which is supposed to show that the field depends on all coordinates $x^1, \ldots, x^3$. It's not great notation because the indices as subscripts are supposed to take some definite values, while the index l in the argument is supposed to denote the whole set of indices. It's especially bad notation when combined with the summation convention and we won't adopt it here.

There is one very famous example of a tensor field. Einstein's theory of general relativity is described by a rank 2 tensor at every point in space. This is called the *metric*. The dynamics of this rank 2 tensor field describe gravity. (I've brushed something rather important under the rug here. Einstein's theory is a rank 2 tensor in *spacetime*, not just in space. Which means that the rank 2 tensor is a 4×4 matrix, rather than a 3×3 matrix.)

Before we move on, it's worth pausing to mention a slightly subtle point. Not all maps $\mathbb{R}^3 \to \mathbb{R}^3$ qualify as "vector fields". The point $\mathbf{x}$ in the codomain $\mathbb{R}^3$ is a vector and so its components transform in the appropriate way under rotation. To be a vector field, the components of the map must transform under the *same* rotation. Similar comments hold for a tensor field.

To illustrate this, the electric field $\mathbf{E}(\mathbf{x})$ is an example of a vector field. If you rotate in space, and so change $\mathbf{x}$, then the direction $\mathbf{E}$ also changes: the rotation acts on both the argument $\mathbf{x}$ and the function itself $\mathbf{E}$.

In contrast, there are maps $\mathbb{R}^3 \to \mathbb{R}^3$ where, although the domain and codomain have the same dimension, vectors in them transform under different rotations. For example, in particle physics there exists an object called a *quark field* which, for our (admittedly, slightly dumbed down) purposes, can be thought of as a map $\mathbb{R}^3 \to \mathbb{R}^3$. This is a quantum field whose ripples are the particles that we call quarks, but these details can be safely ignored for the next couple of years of your life. We will write this field as $q_a(\mathbf{x})$ where the $a = 1, 2, 3$ label is the "colour" of the quark. If we rotate in space, then $\mathbf{x}$ changes but the colour of the quark does not. There is then an independent rotation that acts on the codomain and rotates the colour, but leaves the point in space unchanged. Because the rotations that act on the domain and codomain are unrelated, the quark field is usually not referred to as a vector field.

Taking Derivatives

Given a tensor field, we can always construct higher rank tensors by taking derivatives. In fact, we've already seen a prominent example of this earlier in these lectures. There, we started with a scalar field $\phi(\mathbf{x})$ and differentiated to get the gradient $\nabla\phi$. This means that we start with a rank 0 tensor and differentiate to get a rank 1 tensor.

Strictly speaking, we didn't previously prove that $\nabla\phi$ is a vector field. But it's straightforward to do so. As we've seen above, we need to show that it transforms correctly under rotations. Any vector $\mathbf{v}$ can be decomposed in two different ways,

$$\mathbf{v} = v^i \mathbf{e}_i = v'^i \mathbf{e}'_i$$

where $\{\mathbf{e}_i\}$ and $\{\mathbf{e}'_i\}$ are two orthonormal bases, each obeying $\mathbf{e}_i \cdot \mathbf{e}_j = \mathbf{e}'_i \cdot \mathbf{e}'_j = \delta_{ij}$, and v^i and v'^i are the two different coordinates for $\mathbf{v}$. If we expand $\mathbf{x}$ in this way

$$\mathbf{x} = x_i \mathbf{e}_i = x'_i \mathbf{e}'_i \quad \Longrightarrow \quad x_i = (\mathbf{e}_i \cdot \mathbf{e}'_j) x'_j \quad \Longrightarrow \quad \frac{\partial x^i}{\partial x'_j} = \mathbf{e}_i \cdot \mathbf{e}'_j$$

Here $\mathbf{e}_i \cdot \mathbf{e}'_j$ is the rotation matrix that takes us from one basis to the other. Meanwhile, we can always expand one set of basis vectors in terms of the other,

$$\mathbf{e}_i = (\mathbf{e}_i \cdot \mathbf{e}'_j)\mathbf{e}'_j = \frac{\partial x^i}{\partial x'^j}\mathbf{e}'_j$$

This tells us that we could equally as well write the gradient as

$$\nabla\phi = \frac{\partial\phi}{\partial x^i}\mathbf{e}_i = \frac{\partial\phi}{\partial x^i}\frac{\partial x^i}{\partial x'^j}\mathbf{e}'_j = \frac{\partial\phi}{\partial x'^j}\mathbf{e}'_j$$

This is the expected result: if you work in a different primed basis, then you have the same definition of $\nabla\phi$, but just with primes on both $\mathbf{e}'_i$ and $\partial/\partial x'^i$. This means that the components $\partial_i\phi$ transform correctly under a rotation, so $\nabla\phi$ is indeed a vector.

We can extend the result above to any, suitably smooth, tensor field $T(\mathbf{x})$ of rank p. We can differentiate this any number of times to get a new tensor field of rank, say, $p+q$,

$$X_{i_1\ldots i_q j_1\ldots j_p} = \frac{\partial}{\partial x_{i_1}}\cdots\frac{\partial}{\partial x_{i_q}} T_{j_1\ldots j_p}(\mathbf{x}) \tag{6.12}$$

To verify that this is indeed a tensor, we need to check how it changes under a rotation. In a new basis, we have $x'_i = R_{ij}x_j$ (where $R_{ij} = \mathbf{e}'_i\cdot\mathbf{e}_j$ in the notation above) and so

$$\frac{\partial x'_i}{\partial x_j} = R_{ij} \quad\Longrightarrow\quad \frac{\partial}{\partial x'_i} = \frac{\partial x_j}{\partial x'_i}\frac{\partial}{\partial x_j} = R_{ij}\frac{\partial}{\partial x_j}$$

which is the result we need for X in (6.12) to qualify as a tensor field.

We can implement any of the tensorial manipulations that we met previously for tensor fields. For example, if we start with a vector field $\mathbf{F}(\mathbf{x})$, we can form a rank 2 tensor field

$$T_{ij}(\mathbf{x}) = \frac{\partial F_i}{\partial x_j}$$

But we saw in (6.6) that any rank 2 tensor field can be decomposed into various pieces. There is an anti-symmetric piece

$$A_{ij}(\mathbf{x}) = \epsilon_{ijk}B_k(\mathbf{x}) \quad\text{with}\quad B_k = \frac{1}{2}\epsilon_{ijk}\frac{\partial F_i}{\partial x_j} = -\frac{1}{2}\left(\nabla\times\mathbf{F}\right)_k$$

and a trace piece

$$Q = \frac{\partial F_i}{\partial x_i} = \nabla\cdot\mathbf{F}$$

and, finally, a symmetric, traceless piece

$$P_{ij}(\mathbf{x}) = \frac{1}{2}\left(\frac{\partial F_i}{\partial x_j} + \frac{\partial F_j}{\partial x_i}\right) - \frac{1}{3}\nabla\cdot\mathbf{F}$$

Obviously, the first two of these are familiar tensors (in this case a scalar and vector) from earlier sections.

6.2 Physical Examples

Our discussion above was rooted firmly in mathematics. There are many places in physics where tensors appear. Here we give a handful of examples.

6.2.1 Electric Fields in Matter

Apply an electric field $\mathbf{E}$ to a lump of stuff. A number of things can happen.

If the lump of stuff is an insulator then the material will become *polarised*. This means that the positive electric charge will be pushed in one direction, the negative in another until the lump of stuff acts like a dipole. (This is described in some detail in Section 7 of the lectures on Electromagnetism.) One might think that the resulting polarisation vector $\mathbf{P}$ points in the same direction as the electric field $\mathbf{E}$, but that's too simplistic. For many lumps of stuff, the underlying crystal structure allows the electric charges to shift more freely in some directions than others. The upshot is that the relation between polarisation $\mathbf{P}$ and applied electric field $\mathbf{E}$ is given by

$$\mathbf{P} = \alpha\mathbf{E}$$

where α is a matrix known as the *polarisation tensor*. In a given basis, it has components α_{ij}.

There is a similar story if the lump of stuff is a conductor. This time an applied electric field gives rise to a current density $\mathbf{J}$. Again, the current is not necessarily parallel to the electric field. The relationship between them is now

$$\mathbf{J} = \sigma\mathbf{E}$$

This is known as *Ohm's law*. In general σ is a 3×3 matrix known as the *conductivity tensor*; in a given basis, it has components σ_{ij}.

What can we say about σ when the material is isotropic, meaning that it looks the same in all directions? In this case, no direction is any different from any other. With no preferred direction, the conductivity tensor must be proportional to an invariant tensor, so that it looks the same in all coordinate systems. What are our options?

For 3d materials, the only option is $\sigma_{ij} = \sigma\delta_{ij}$, which ensures that the current does indeed run parallel to the electric field. In this case σ is just referred to as the *conductivity*.

However, suppose that we're dealing with a thin wafer of material in which both the current and electric field are restricted to lie in a plane. This changes the story because now we're dealing with vectors in $\mathbb{R}^2$ rather than $\mathbb{R}^3$ and $\mathbb{R}^2$ is special because there are two invariant 2-tensors in this dimension: δ_{ij} and ϵ_{ij}. This means that the most general conductivity tensor for an isotropic 2d material takes the form

$$\sigma_{ij} = \sigma_{xx}\delta_{ij} + \sigma_{xy}\epsilon_{ij} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ -\sigma_{xy} & \sigma_{xx} \end{pmatrix}$$

Here σ_{xx} is called the *longitudinal conductivity* while σ_{xy} is called the *Hall conductivity*. If $\sigma_{xy} \neq 0$ then an electric field in the x-direction induces a current in the y-direction.

As an aside, it turn out that the seemingly mundane question of understanding σ_{xy} in real materials is closely tied to some of the most interesting breakthroughs in mathematics in recent decades! This is the subject of the Quantum Hall Effect.

6.2.2 The Inertia Tensor

Another simple example of a tensor arises in Newtonian mechanics. A *rigid body* rotating about the origin can be modelled by some number of masses m_a at positions $\mathbf{x}_a$, all moving with velocity $\dot{\mathbf{x}}_a = \boldsymbol{\omega} \times \mathbf{x}_a$. Here $\boldsymbol{\omega}$ is known as the angular velocity. The angular velocity $\boldsymbol{\omega}$ is related to the angular momentum $\mathbf{L}$ by

$$\mathbf{L} = I\boldsymbol{\omega} \tag{6.13}$$

with I the *inertia tensor*. The angular momentum does not necessarily lie parallel to the angular velocity and, correspondingly, I is in general a matrix, rather than a single number. In fact, we can easily derive an expression for the inertia tensor. The angular momentum is

$$\mathbf{L} = \sum_a m_a \mathbf{x}_a \times \dot{\mathbf{x}}_a = \sum_a m_a \mathbf{x}_a \times (\boldsymbol{\omega} \times \mathbf{x}_a) = \sum_a m_a \Big(|\mathbf{x}_a|^2 \boldsymbol{\omega} - (\mathbf{x}_a \cdot \boldsymbol{\omega})\mathbf{x}_a \Big)$$

In components, $L_i = I_{ij}\omega_j$, where

$$I_{ij} = \sum_a m_a \Big(|\mathbf{x}_a|^2 \delta_{ij} - (\mathbf{x}_a)_i (\mathbf{x}_a)_j \Big)$$

For a continuous object with density $\rho(\mathbf{x})$, we can replace the sum with a volume integral

$$I_{ij} = \int_V \rho(\mathbf{x}) \Big(|\mathbf{x}|^2 \delta_{ij} - x_i x_j \Big)\, dV \tag{6.14}$$

So, for example, $I_{33} = \int \rho(x_1^2 + x_2^2)\, dV$ and $I_{12} = -\int \rho x_1 x_2\, dV$.

An Example: A Sphere

For a ball of radius R and density $\rho(r)$, the inertia tensor is

$$I_{ij} = \int_V \rho(r)(r^2\delta_{ij} - x_i x_j)\, dV$$

The second of these terms is the integral (6.10) that we simplified in Section 6.1.3 using isotropy arguments. Using (6.11), we have

$$I_{ij} = \frac{2}{3}\delta_{ij}\int_V \rho(r) r^2\, dV = \frac{8\pi}{3}\delta_{ij}\int_0^R dr\ \rho(r) r^4$$

For example, if $\rho(r) = \rho_0$ is constant, then $I_{ij} = \frac{8}{15}\pi\rho_0 R^5\delta_{ij} = \frac{2}{5}MR^2\delta_{ij}$ where M is the mass of the sphere.

Another Example: A Cylinder

The sphere is rather special because the inertia tensor is proportional to δ_{ij}. That's not the case more generally. Consider, for example, a solid 3d cylinder of radius a and height $2L$, with uniform density ρ. The mass is $M = 2\pi a^2 L\rho$. We align the cylinder with the z-axis and work in cylindrical polar coordinates $x = r\cos\phi$ and $y = r\sin\phi$. The components of the inertia tensor are then

$$I_{33} = \int_V \rho(x^2+y^2)\, dV = \rho\int_0^{2\pi} d\phi \int_0^a dr \int_{-L}^{+L} dz\ r\, r^2 = \rho\pi L a^4$$
$$I_{11} = \int_V \rho(y^2+z^2)\, dV = \rho\int_0^{2\pi} d\phi \int_0^a dr \int_{-L}^{+L} dz\ r(r^2\sin^2\phi + z^2) = \rho\pi a^2 L\left(\frac{a^2}{2} + \frac{2L^2}{3}\right)$$

By symmetry, $I_{22} = I_{11}$. For the off-diagonal elements, we have

$$I_{13} = -\int_V \rho x_1 x_3\, dV = -\rho\int_0^{2\pi} d\phi \int_0^a dr \int_{-L}^{L} dz\ r^2 z\cos\phi = 0$$

where the integral vanishes due to the ϕ integration. Similarly, $I_{12} = I_{13} = 0$. We find that the inertia tensor for the cylinder is

$$I = \text{diag}\left(M\left(\frac{a^2}{4} + \frac{L^2}{3}\right)\ ,\ M\left(\frac{a^2}{4} + \frac{L^2}{3}\right)\ ,\ \frac{1}{2}Ma^2\right) \tag{6.15}$$

Note that the inertia tensor is diagonal in our chosen coordinates.

The Eigenvectors of the Inertia Tensor

The inertia tensor I defined in (6.14) has a special property: it is symmetric

$$I_{ij} = I_{ji}$$

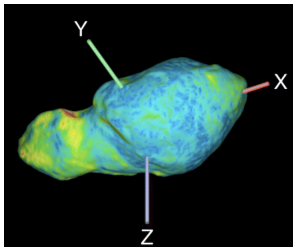
Any symmetric matrix I can always be diagonalised by an appropriate rotation. This means that there exists an $R \in SO(n)$ such that

$$I' = RIR^T = \text{diag}(I_1, I_2, I_3)$$

Another way of saying this is that any symmetric rank 2 tensor has a basis of orthonormal eigenvectors $\{\mathbf{e}_i\}$, with I_i the corresponding eigenvalues.

In the case of the inertia tensor, the eigenvectors $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$ are called the *principal axes* of the solid. It means that any object, no matter how complicated, has its own preferred set of orthonormal axes embedded within it. If the object has some symmetry, then the principal axes will always be aligned with this symmetry. This, for example, was the case for the cylinder that we computed above where aligning the cylinder with the z-axis automatically gave us a diagonal inertia tensor (6.15).

In general, it will be less obvious where the principal axes lie. For example, the figure on the right shows the asteroid Toutatis, which is notable for its lumpy shape. The principal axes are shown embedded in the asteroid.



From (6.13), the angular momentum $\mathbf{L}$ is aligned with the angular velocity $\boldsymbol{\omega}$ only if a body spins about one of its principal axes. It turns out that, in this case, nice things happen and the body spins smoothly. However, if $\mathbf{L}$ and $\boldsymbol{\omega}$ are misaligned, the body exhibits more complicated tumbling, wobbling motion as it spins. You can learn all about this in the lectures on Classical Dynamics. (For what it's worth, Toutatis does not spin about a principal axes.)

6.2.3 Higher Rank Tensors

You might reasonably complain that, after all that work defining tensors, the examples that we've given here are nothing more exotic than matrices, mapping one vector to another. And you would be right. However, as we get to more sophisticated theories of physics, tensors of higher rank do make an appearance. Here we don't give full details, but just say a few words to give you a flavour of things to come.

Perhaps the simplest example arises in the theory of elastic materials. These materials can be subjected to *strain*, which describes the displacement of the material at each point, and *stress*, which describes the forces acting on the material at each point. But each of these is itself a 2-tensor (strictly a tensor field). The *strain tensor* e_{ij} is a symmetric tensor that describes the way the displacement in the x^i direction varies in the x^j. The *stress tensor* σ_{ij} describes the component of the force F_i across a plane normal to x^j. These two tensors are related by

$$\sigma_{ij} = C_{ijkl}e_{kl}$$

This is the grown up version of Hooke's law. In general an elastic material is characterised by the *elasticity tensor*, also known as the *stiffness tensor*, C_{ijkl}.

Higher rank tensors also appear prominently in more advanced descriptions of geometry. In higher dimensions, the simple Gaussian curvature that we met in Section 2 isn't enough to capture all the interesting ways in which spaces can curve in different directions. Instead, it is replaced by a 4-tensor R_{ijkl} known as the Riemann curvature. In the context of physics, this 4-tensor describes the bending of space and time and is needed for the grown-up version of Newton's law of gravity.

6.3 A Unification of Integration Theorems

In this final section, we turn back to matters of mathematics. The three integral theorems that we met in Section 4 are obviously closely related. To end these lectures, we show how they can be presented in a unified framework. This requires us to introduce some novel and slightly formal ideas. These go quite a bit beyond what is usually covered in an introductory course on vector calculus, but we will meet these objects again in later courses on Differential Geometry and General Relativity. View this section as a taste of things to come.

6.3.1 Integrating in Higher Dimensions

Our unified framework will give us integral theorems in any dimension $\mathbb{R}^n$. If you look back at Section 4, you'll notice that the divergence theorem already holds in any $\mathbb{R}^n$. Meanwhile, Stokes' theorem is restricted to surfaces in $\mathbb{R}^3$ for the very simple reason that the cross-product is only defined in $\mathbb{R}^3$. This suggests that before we can extend our integral theorems to higher dimensions, we should first ask a more basic question: how do we extend the cross product to higher dimensions?

The introduction of tensors gives us a way to do this. Given two vectors $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{R}^3$, the cross-product is

$$(\mathbf{a} \times \mathbf{b})_i = \epsilon_{ijk} a_j b_k$$

From this perspective, the reason that the cross product can only be employed in $\mathbb{R}^3$ is because it's only there that the ϵ_{ijk} symbol has three entries. If, in contrast, we're in $\mathbb{R}^4$ then we have ϵ_{ijkl} and so if we feed it two vectors $\mathbf{a}$ and $\mathbf{b}$, then we find ourselves with a tensor of rank 2, $T_{ij} = \epsilon_{ijkl} a_k b_l$.

The tensors that we get from an epsilon symbol are always special, in the sense that they are totally anti-symmetric. The anti-symmetry condition doesn't impose any extra constraint on a 0-tensor ϕ or a 1-tensor a_i as these are just scalar fields and vector fields respectively. It only kicks in when we get to tensors of rank 2 or higher.

With this in mind, we can revisit the cross product. We can define the cross product in any dimension $\mathbb{R}^n$: it is a map that eats two vectors $\mathbf{a}$ and $\mathbf{b}$ and spits back an anti-symmetric $(n-2)$-tensor

$$(\mathbf{a} \times \mathbf{b})_{i_1 \ldots i_{n-2}} = \epsilon_{i_1 \ldots i_n} a_{i_{n-1}} b_{i_n}$$

The only thing that's special about $\mathbb{R}^3$ is that we get back another vector, rather than a higher dimensional tensor.

There is also a slightly different role played by the epsilon symbol $\epsilon_{i_1,\ldots,i_n}$: it provides a map from anti-symmetric p-tensors to anti-symmetric $(n-p)$-tensors, simply by contracting indices,

$$\epsilon : \; T_{i_1 \ldots i_p} \;\mapsto\; \frac{1}{(n-p)!} \epsilon_{i_1 \ldots i_n} T_{i_{n-p+1} \ldots i_n} \tag{6.16}$$

This map goes by the fancy name of the *Hodge dual*. (Actually, it's an entirely trivial version of the Hodge dual. The proper Hodge dual is a generalisation of this idea to curved spaces.)

Our next step is to think about what this has to do with integration. Recall that earlier in these lectures we found two natural ways to integrate vector fields in $\mathbb{R}^3$. The first is along a line

$$\int_C \mathbf{F} \cdot d\mathbf{x} \tag{6.17}$$

which captures the component vector field *tangent* to the line. We can perform this procedure in any dimension $\mathbb{R}^n$. The second operation is to integrate a vector field over a surface

$$\int_S \mathbf{F} \cdot d\mathbf{S} \tag{6.18}$$

where $d\mathbf{S}$ points in the direction normal to the surface. This integration captures the component of the vector field *normal* to the surface and only makes sense in $\mathbb{R}^3$. This is because it's only in $\mathbb{R}^3$ that a two-dimensional surface has a unique normal. More operationally, this normal, which is buried in the definition of $d\mathbf{S}$, requires us to use the cross product. For a parameterised surface $\mathbf{x}(u, v)$, the vector area element is

$$d\mathbf{S} = \frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v}\, du\, dv$$

or, in components,

$$dS_i = \epsilon_{ijk} \frac{\partial x^j}{\partial u} \frac{\partial x^k}{\partial v}\, du\, dv$$

Now comes a mathematical sleight of hand. Rather than thinking of (6.18) as the integral of a vector field projected normal to the surface, instead think of it as the integral of an anti-symmetric 2-tensor $F_{ij} = \epsilon_{ijk} F_k$ integrated *tangent* to the surface. We then have

$$\int_S \mathbf{F} \cdot d\mathbf{S} = \int_S F_{ij}\, dS_{ij} \quad \text{with} \quad dS_{ij} = \frac{1}{2}\left(\frac{\partial x^j}{\partial u} \frac{\partial x^k}{\partial v} - \frac{\partial x^j}{\partial v} \frac{\partial x^k}{\partial u} \right) du\, dv \tag{6.19}$$

This is the same equation as before, just with the epsilon symbol viewed as part of the integrand F_{ij} rather than as part of the measure dS_i. Note that we've retained the anti-symmetry of the area element dS_{ij} that was inherent in our original cross product definition of $d\mathbf{S}$. Strictly speaking this isn't necessary because we're contracting with anti-symmetric indices in F_{ij}, but it turns out that it's best to think of both objects F_{ij} and dS_{ij} as individually anti-symmetric.

This new perspective suggests a way to generalise to higher dimensions. In the line integral (6.17) we're integrating a vector field over a line. In the surface integral (6.19), we're really integrating an anti-symmetric 2-tensor over a surface. The key idea is that one can integrate a totally anti-symmetric p-tensor over a p-dimensional subspace.

Specifically, given an anti-symmetric p-tensor, the generalisation of the line integral (6.17) is the integration over a p-dimensional subspace,

$$\int_M T_{i_1\ldots i_p}\, dS_{i_1\ldots i_p} \tag{6.20}$$

where $\dim(M) = p$. Here $dS_{i_1\ldots i_p}$ is a higher dimensional version of the "area element" defined in (6.19).

Alternatively, the higher dimensional version of the surface integral (6.18) involves first mapping the p-tensor to an $(n-p)$-tensor using the Hodge dual. This can subsequently be integrated over an $(n-p)$-dimensional subspace

$$\int_{\tilde{M}} T_{i_1\ldots i_p}\epsilon_{i_1\ldots i_p j_1\ldots j_{n-p}}\, d\tilde{S}_{j_1\ldots j_{n-p}} \tag{6.21}$$

with $\dim(\tilde{M}) = n - p$.

In fact, we've already met an integral of the form (6.21) elsewhere in these lectures, since this is what we're implicitly doing when we integrate a scalar field over a volume. In this case the "area element" is just $dS_{i_1\ldots i_n} = \frac{1}{n!}\epsilon_{i_1\ldots i_n}\, dV$ and the two epsilon symbols just multiply to a constant.. When actually computing a volume integral, this extra machinery is more of a distraction than a help.. But if we want to know how to think about things more generally then it's extremely useful.

6.3.2 Differentiating Anti-Symmetric Tensors

We've now learned how to integrate anti-symmetric tensors. Our next step is to learn how to differentiate them. We've already noted in (6.12) that we can differentiate a p tensor once to get a tensor of rank $p+1$, but in general differentiating loses the anti-symmetry property. As we now explain, there is a way to restore it so that when we differentiate a totally anti-symmetric p tensor, we end up with a totally anti-symmetric $(p+1)$-tensor.

For a scalar field, things are trivial. We can construct a vector field $\nabla\phi$ and this is automatically "anti-symmetric" because there's nothing to anti-symmetrise.

If we're given a vector field F_i, we can differentiate and then anti-symmetrise by hand. I will introduce a new symbol for "differentiation and anti-symmetrisation" and write

$$(\mathcal{D}F)_{ij} := \frac{1}{2}\left(\frac{\partial F_i}{\partial x^j} - \frac{\partial F_j}{\partial x^i}\right)$$

where the anti-symmetry is manifest on the right-hand side. I should confess that the notation $\mathcal{D}F$ is not at all standard. In subsequent courses, this object is usually viewed as something called a "differential form" and written simply as dF but the notation dF is loaded with all sorts of other connotations which are best ignored at this stage. Hence the made-up notation $\mathcal{D}F$.

In $\mathbb{R}^3$, this anti-symmetric differentiation is equivalent to the curl using the Hodge map (6.16),

$$(\nabla \times \mathbf{F})_i = \epsilon_{ijk}(\mathcal{D}F)_{jk}$$

But now we can extend this definition to any anti-symmetric p-tensor. We can always differentiate and anti-symmetrise to get a $(p+1)$-tensor defined by

$$(\mathcal{D}T)_{i_1\ldots i_{p+1}} = \frac{1}{p+1}\left(\frac{\partial T_{i_1\ldots i_p}}{\partial x^{i_{p+1}}} + p \text{ further terms}\right)$$

where the further terms involve replacing the derivative $\partial/\partial x^{i_{p+1}}$ with one of the other coordinates $\partial/\partial x^j$ so that the whole shebang is fully anti-symmetric.

Note that, with this definition of $\mathcal{D}$, if we differentiate twice then we take a p-tensor to a $(p+2)$-tensor. But this $(p+2)$-tensor always vanishes!

$$(\mathcal{D}\mathcal{D}T)_{i_1\ldots i_{p+2}} = 0$$

for any tensor T. This is because we'll have two derivatives contracted with an epsilon and is the higher dimensional generalisation of the statements that $\nabla \times \nabla\phi = 0$ or $\nabla \cdot (\nabla \times \mathbf{F}) = 0$.

As an aside: this is actually the second time in these lectures that we've seen something vanish when you act twice, although you'd be forgiven for failing to notice the connection. Here our new anti-symmetric derivative obeys $\mathcal{D}^2(\text{anything}) = 0$. But we previously saw that the "boundary of a boundary" is always zero. This means that if a higher dimensional space (really a manifold) M has boundary ∂M then $\partial(\partial M) = 0$. Conceptually, these two ideas are very different but one can't help but be struck by the similarity of the equations $\mathcal{D}^2(\text{anything}) = 0$ and $\partial^2(\text{anything}) = 0$, even though the "anything"'s are very different objects in the two formulae. It turns out that this similarity is pointing at a deep connection between the topology of spaces and the kinds of tensors that one can put on these spaces. In fancy maths words, this is the link between homology and cohomology.

Finally, we can now state the general integration theorem. Given an anti-symmetric p-tensor T, then

$$\int_M (\mathcal{D}T)_{i_1\ldots i_{p+1}}\, dS_{i_1\ldots i_{p+1}} = \int_{\partial M} T_{i_1\ldots i_p}\, dS_{i_1\ldots i_p} \tag{6.22}$$

Here $\dim(M) = p+1$ and, therefore the boundary has $\dim(\partial M) = p$. Note that we don't use a different letter to distinguish the integration measure over these various spaces: everything is simply dS and you have to look closer at the indices to see what kind of space you're integrating over.

The equation (6.22) is a unification of all integration theorems. It contains the fundamental theorem of calculus (when $p = 0$), the divergence theorem (when $p = n-1$) and Stokes' theorem (when $p = 1$ and $\mathbb{R}^n = \mathbb{R}^3$). Geometers refer to this generalised theorem simply as *Stokes' theorem* since that is the original result that it resembles most. The proof is simply a higher dimensional version of the proofs that we sketched previously.

There is, to put it mildly, quite a lot that I'm sweeping under the rug in the discussion above. In particular, the full Stokes' theorem does not hold only in $\mathbb{R}^n$ but in a general curved space known as a manifold. In that context, one has to be a lot more careful about what kind of tensors we're dealing with and, as I mentioned above, Stokes' theorem should be written using a kind of anti-symmetric tensor known as a *differential form*. None of this really matters when working in flat space, but the differences become crucial when thinking about curved spaces. If you want to learn more, these topics will be covered in glorious detail in later courses on Differential Geometry or, for physicists, General Relativity.

What You Really Need

Here are expressions for div, grad, curl and the Laplacian in various coordinate systems.

Cartesian: $\mathbf{x} = (x, y, z)$

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

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

$$\nabla \times \mathbf{F} = \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}\right)\hat{\mathbf{x}} + \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}\right)\hat{\mathbf{y}} + \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}\right)\hat{\mathbf{z}}$$

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}$$

Cylindrical Polars: $\mathbf{x} = (\rho\cos\phi, \rho\sin\phi, z)$

$$\nabla f = \frac{\partial f}{\partial \rho}\hat{\boldsymbol{\rho}} + \frac{1}{\rho}\frac{\partial f}{\partial \phi}\hat{\boldsymbol{\phi}} + \frac{\partial f}{\partial z}\hat{\mathbf{z}}$$

$$\nabla \cdot \mathbf{F} = \frac{1}{\rho}\frac{\partial(\rho F_\rho)}{\partial \rho} + \frac{1}{\rho}\frac{\partial F_\phi}{\partial \phi} + \frac{\partial F_z}{\partial z}$$

$$\nabla \times \mathbf{F} = \left(\frac{1}{\rho}\frac{\partial F_z}{\partial \phi} - \frac{\partial F_\phi}{\partial z}\right)\hat{\boldsymbol{\rho}} + \left(\frac{\partial F_\rho}{\partial z} - \frac{\partial F_z}{\partial \rho}\right)\hat{\boldsymbol{\phi}} + \frac{1}{\rho}\left(\frac{\partial(\rho F_\phi)}{\partial \rho} - \frac{\partial F_\rho}{\partial \phi}\right)\hat{\mathbf{z}}$$

$$\nabla^2 f = \frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho\frac{\partial f}{\partial \rho}\right) + \frac{1}{\rho^2}\frac{\partial^2 f}{\partial \phi^2} + \frac{\partial^2 f}{\partial z^2}$$

Spherical Polars: $\mathbf{x} = (r\sin\theta\cos\phi, r\sin\theta\sin\phi, r\cos\theta)$

$$\nabla f = \frac{\partial f}{\partial r}\hat{\mathbf{r}} + \frac{1}{r}\frac{\partial f}{\partial \theta}\hat{\boldsymbol{\theta}} + \frac{1}{r\sin\theta}\frac{\partial f}{\partial \phi}\hat{\boldsymbol{\phi}}$$

$$\nabla\cdot\mathbf{F} = \frac{1}{r^2}\frac{\partial (r^2 F_r)}{\partial r} + \frac{1}{r\sin\theta}\frac{\partial(\sin\theta F_\theta)}{\partial\theta} + \frac{1}{r\sin\theta}\frac{\partial F_\phi}{\partial\phi}$$

$$\nabla\times\mathbf{F} = \frac{1}{r\sin\theta}\left(\frac{\partial(\sin\theta F_\phi)}{\partial\theta} - \frac{\partial F_\theta}{\partial\phi}\right)\hat{\mathbf{r}} + \frac{1}{r}\left(\frac{1}{\sin\theta}\frac{\partial F_r}{\partial\phi} - \frac{\partial(rF_\phi)}{\partial r}\right)\hat{\boldsymbol{\theta}} + \frac{1}{r}\left(\frac{\partial(rF_\theta)}{\partial r} - \frac{\partial F_r}{\partial\theta}\right)\hat{\boldsymbol{\phi}}$$

$$\nabla^2 f = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial f}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial f}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 f}{\partial\phi^2}$$