To prepare for the more challenging topics that we will meet later in the course, we need to start off with some introductory material. In this section we will introduce two new co-ordinate systems (cylindrical and spherical co-ordinates), and briefly review the meaning of partial differentiation and the chain rule for partial derivatives. By the end of this section, you should be able to convert Cartesian co-ordinates into cylindrical and spherical co-ordinates and convert partial differential equations presented in Cartesian co-ordinates to partial differential equations in alternative co-ordinate systems. For further reading and examples, take a look at the book Calculus by James Stewart (a recommended text for your first year calculus courses). **You should also review any notes you have from first year on polar co-ordinates and partial differentiation**.

1.1. Co-ordinate Systems

In three space dimensions, a point P may be represented in the Cartesian system by the co-ordinates (x, y, z) or as a vector of the form $x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$, where

$$\mathbf{i} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad \mathbf{j} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad \mathbf{k} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

are the standard basis vectors in $\mathbf{R}^3$.

Cartesian co-ordinates are convenient to describe objects such as rectangles and bricks, or objects composed of different sized bricks. For example, we understand that the set of points

$$\{(x, y, z) \,|\, 0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq z \leq 1\}$$

describes the unit cube (a cube with sides of length one, with vertices at $(0, 0, 0)$, $(0, 1, 0)$ etc). To describe objects and surfaces with curvature, such as circles, spheres, cylinders, cones etc., it is often simpler to work in alternative co-ordinate systems.

Revision: polar co-ordinates.

You already know that polar co-ordinates (r, θ) are useful for describing flat, circular geometries in **two** space dimensions. The Cartesian co-ordinates (x, y) of a point P tell us the distances to travel from the origin along the x and y axes to locate P. The co-ordinates (r, θ) give us the length of the line that connects P to the origin, and the angle between that line and the positive x-axis (measured in an anti-clockwise direction from the positive x-axis). Note that $r \geq 0$ and $0 \leq \theta < 2\pi$. Using basic trigonometry, if we know the polar co-ordinates (r, θ), then the Cartesian co-ordinates are given by

$$x = r\cos\theta, \qquad y = r\sin\theta.$$

Similarly, if we know the Cartesian co-ordinates, then

$$r = \sqrt{x^2 + y^2}, \qquad \theta = \tan^{-1}(y/x).$$

In three dimensions, there are two other useful co-ordinate systems: cylindrical and spherical co-ordinates. [**See figures in the separate handout**].

[1] Dr. C.E. Powell, School of Mathematics, University of Manchester.

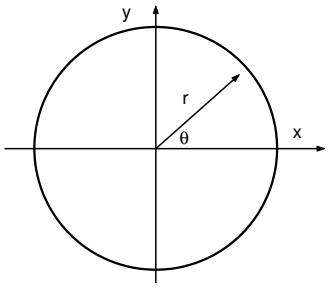


Figure 1: In $\mathbb{R}^2$, a point can be represented by Cartesian co-ordinates (x, y) or polar co-ordinates (r, θ).

Cylindrical co-ordinates

These are a natural extension of polar co-ordinates to three dimensions. Any point P in $\mathbb{R}^3$ can be represented by

$$(r, \theta, z)$$

where[2] (r, θ) are the standard polar co-ordinates of the point that is the projection of P onto the x-y plane **[see figure]**, and z is the height (above or below the x-y plane) of P. Viewed another way, this says that any point P in three dimensions can be thought of as lying on the surface of a cylinder of some radius r, of infinite height, with the z-axis running through the centre. The co-ordinates θ and z tell us where on the surface of the cylinder the point is located.

With the aid of simple trigonometry **[see figure]**, if we know the cylindrical co-ordinates of a point P then the Cartesian co-ordinates are given by:

$$x = r\cos\theta, \qquad y = r\sin\theta, \qquad z = z.$$

Similarly, if we know the Cartesian co-ordinates, the cylindrical co-ordinates are given by

$$r = \sqrt{x^2 + y^2}, \qquad \theta = \tan^{-1}(y/x), \qquad z = z.$$

The co-ordinate z is the same in both systems. As with polar co-ordinates, we must be careful when applying $\tan^{-1}$. Recall that tan is a periodic function, so there are infinitely many values θ such that $\tan(\theta)$ is equal to a given value of y/x. The value of θ we want must lie in the range $[0, 2\pi)$.

Example. Find the cylindrical co-ordinates of the point $(x, y, z) = (3, -3, -1)$.

Answer: Clearly, $z = -1$ and $r = \sqrt{9+9} = \sqrt{18}$. Direct calculation (using a calculator) gives:

$$\theta = \tan^{-1}(-3/3) = \tan^{-1}(-1) = -\pi/4.$$

This is outside the desired range. In fact, we have

$$\tan^{-1}(-1) = -\pi/4 + n\pi,$$

for any $n \in \mathbb{N}$. Here, $\theta = -\pi/4 + 2\pi = 7\pi/4$ (or 315 degrees). Sketch a diagram to convince yourself that this is the correct value! There are also plenty more examples to try on Exercise Sheet 1.

[2]Note that some textbooks use different letters in place of r and θ for cylindrical co-ordinates. We will use the notation that is consistent with what we already know for polar co-ordinates.

Spherical co-ordinates

These co-ordinates are useful for geometries with radial symmetry (e.g., spheres). Any point P can be thought of as lying on the surface of a sphere, of some radius ρ, centred at the origin [**see figure**]. If (x, y, z) are the Cartesian co-ordinates of P then the distance from P to the origin is

$$\rho = \sqrt{x^2 + y^2 + z^2}.$$

This is the radius of the sphere upon whose surface P lies. We need to supply two other co-ordinates (angles) to uniquely specify the location of P. The spherical co-ordinates are (ρ, θ, ϕ), where

- ρ is the distance from P to the origin.
- θ is the same angle used in the polar and cylindrical systems. That is, if we project P onto the (x, y) plane [**see figure**], θ is the angle measured anti-clockwise from the positive x-axis to the line joining the projected point to the origin.
- ϕ is the angle between the positive z-axis and the line connecting P to the origin.

Note that $\rho > 0$ (if $\rho = 0$ then we are at the origin), $0 \leq \theta < 2\pi$, and $0 \leq \phi \leq \pi$. You might like to think about the angles ϕ and θ as longitude and latitude.

With some basic trigonometry [**see figure**], we find

$$r = \rho \sin \phi$$

and so

$$x = \rho \sin \phi \cos \theta, \qquad y = \rho \sin \phi \sin \theta.$$

In addition

$$z = \rho \cos \phi.$$

Alternatively, if we know the Cartesian co-ordinates (x, y, z) then,

$$\rho = \sqrt{x^2 + y^2 + z^2}, \qquad \theta = \tan^{-1}(y/x).$$

Clearly $\tan \phi = r/z$ and so the angle ϕ can be calculated by

$$\phi = \tan^{-1}\left(\frac{\sqrt{x^2 + y^2}}{z}\right).$$

Example. Find the spherical co-ordinates of the point $(x, y, z) = (3, -3, -1)$.

Answer: We have $\rho = \sqrt{9 + 9 + 1} = \sqrt{19}$ and we already know $\theta = 7\pi/4$ (or 315 degrees, from the previous example). Finally, $\phi = \tan^{-1}\left(\sqrt{18}/-1\right) = -1.3393 + n\pi$. The correct value is achieved with $n = 1$, giving $\phi = 1.8023$ radians (to 4 d.p.) or $\phi = 103.26$ degrees (to two d.p). Now try the other examples on Exercise Sheet 1!

You should now be able to describe standard surfaces like planes, cones, spheres etc in Cartesian, cylindrical and/or spherical co-ordinates. Sometimes, working in a particular co-ordinate system vastly simplifies the equation for the surface.

Example. What surface is represented by the equation $\rho = 2$? How would the same surface be represented in Cartesian co-ordinates?

Answer: $\rho = 2$ is the equation for a sphere of radius 2, centred at the origin. More precisely, $\rho = 2$ means the set of points (ρ, θ, ϕ) for which $\rho = 2$ is fixed and θ and ϕ take any of their permitted values. In Cartesian co-ordinates, the same surface is represented by the set of points (x, y, z) such that

$$\sqrt{x^2 + y^2 + z^2} = 2.$$

As you can see, the equation is far simpler in spherical co-ordinates.

1.2. Partial Differentiation

A 'differential equation' is an equation containing an unknown function, say u, and some of its derivatives. If u is a function of a single variable x, then the derivatives are

$$\frac{du}{dx}, \quad \frac{d^2u}{dx^2}, \quad \frac{d^3u}{dx^3}, \quad \ldots.$$

A differential equation associated with a function of **single** variable is called an **ordinary differential equation** or **ODE**. For example[3]

$$\frac{d^2u}{dx^2} + \lambda u = 0.$$

The first derivative $\frac{du}{dx}$ represents the rate of change of u with respect to x. A more formal definition is: the slope of the tangent line to the curve $y = u(x)$ at a given point x. For example, consider the simple function $u(x) = x^2$. We all know the derivative is $2x$. Using the formal definition,

$$\frac{du}{dx} = \lim_{h\to 0} \frac{u(x+h) - u(x)}{h} = \lim_{h\to 0} \frac{(x+h)^2 - x^2}{h} = \lim_{h\to 0} 2x + h = 2x.$$

If u is a function of **more than one** variable, e.g., $u(x, y)$, then there are other derivatives to consider. We can calculate the rate of change of u with respect to either x or y. The first **partial derivatives** are denoted

$$\frac{\partial u}{\partial x}, \qquad \frac{\partial u}{\partial y},$$

or u_x, u_y, for short. There are three second partial derivatives, including

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

which is usually abbreviated to u_{xy}, etc. Partial derivatives are rates of change with respect to one variable, whilst the other variables are held fixed.

Example. Consider the function

$$u(x, y) = \left(x^2 + 3y^2\right) \exp\left(-x^2 - y^2\right).$$

[3]You studied how to solve second-order ODEs with constant coefficients in the first year and it will be very important in this course that you remember how to do this. If you have forgotten, you should revise it now.

To find the partial derivative u_x, recall that we treat y like a constant and differentiate as if u were only a function of x as follows

$$\begin{aligned}\frac{\partial u}{\partial x} &= \left(x^2+3y^2\right)(-2x)\exp\left(-x^2-y^2\right)+\exp\left(-x^2-y^2\right)(2x)\\ &= \left(2x-2x^3-6xy^2\right)\exp\left(-x^2-y^2\right)\\ &= 2x\left(1-x^2-3y^2\right)\exp\left(-x^2-y^2\right).\end{aligned}$$

This is the rate of change of u with respect to x if y is kept fixed. For example, if $y=0$, then

$$\frac{\partial u}{\partial x}\,|_{y=0} = 2x\left(1-x^2\right)\exp\left(-x^2\right).$$

Similarly, $\frac{\partial u}{\partial y}$ is found by treating x as a constant, and differentiating with respect to y.

Note that if $u=u(x,y)$ is a function of two variables then fixing a value for y amounts to taking a cross-section through the surface $z=u(x,y)$. The partial derivative u_x actually gives the slope of the tangent to the curve that is that cross-section. We can plot the function $u(x,y)$ from the above example easily in MATLAB, by fixing values for x and y and then interpreting $z=u(x,y)$ as the output of the function. Try the following MATLAB commands.

```
>> x=linspace(-4,4,64); y=linspace(-4,4,64); [xx,yy]=meshgrid(x,y);
>> zz=(xx.^2+3.*yy.^2).*exp(-xx.^2-yy.^2);
>> surf(xx,yy,zz); axis square; shg
```

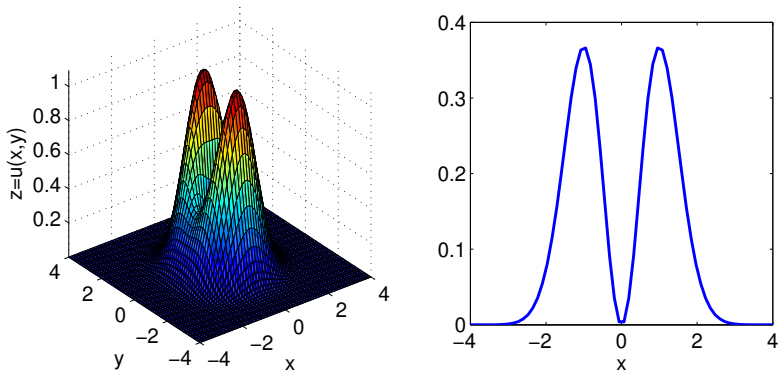


Figure 2: Left: MATLAB `surf` plot of the function $u(x,y)=\left(x^2+3y^2\right)\exp\left(-x^2-y^2\right)$. Right: cross-section through $z=u(x,y)$ at $y=0$.

If we set $y=0$ then $u(x,0)=x^2\exp(-x^2)$ is a cross-section through the surface $z=u(x,y)$ (this is shown in the right plot in Figure 2). In the example above, we found $\frac{\partial u}{\partial x}\,|_{y=0}$, which is the slope of the tangent line to this cross-section.

Formally, the first partial derivatives of a function u of two variables x and y are given by

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

Example. Use the above definition to find the first partial derivatives of the function $u(x,y) = x^2 + y^2$.

Answer: Only the calculation for u_x is shown here. u_y is left as an exercise. We have

$$\frac{\partial u}{\partial x} = \lim_{h\to 0} \frac{(x+h)^2 + y^2 - \left(x^2+y^2\right)}{h} = \lim_{h\to 0} \frac{2xh + h^2}{h} = 2x.$$

1.3. The Chain Rule for Partial Derivatives.

For functions of a single variable, we have the well-known chain rule. Suppose u is a function of t and t is a function of x. Then u is a function of x (because it depends on t) and

$$\frac{du}{dx} = \frac{du}{dt}\frac{dt}{dx}.$$

Example. Suppose $u = \sin^2(t)$ and $x = \sin t$. Use the chain rule to find $\frac{du}{dx}$.

Answer: Differentiating u with respect to t and x with respect to t (which is natural since u is given as a function of t and x is given as a function of t) gives

$$\frac{du}{dt} = 2\sin t\cos t, \qquad \frac{dx}{dt} = \cos t.$$

Now, since[4] $\frac{dt}{dx} = \left(\frac{dx}{dt}\right)^{-1}$, the chain rule gives

$$\frac{du}{dx} = 2\sin t = 2x.$$

(Of course, we could also have made the substitution $x = \sin t$ immediately to give $u = x^2$ and then differentiated with respect to x straightforwardly).

A similar rule holds for functions of more than one variable. Here, we'll focus on functions of two variables. Suppose $u = u(s,t)$ and $s = s(x,y)$, and $t = (x,y)$. Then, u is a function of x and y through its dependence on s and t. The chain rule gives:

$$\begin{aligned}
\frac{\partial u}{\partial x} &= \frac{\partial u}{\partial t}\frac{\partial t}{\partial x} + \frac{\partial u}{\partial s}\frac{\partial s}{\partial x}, \\
\frac{\partial u}{\partial y} &= \frac{\partial u}{\partial t}\frac{\partial t}{\partial y} + \frac{\partial u}{\partial s}\frac{\partial s}{\partial y}.
\end{aligned}$$

Notice that there are two terms in each expression. This is because the function u depends on x and y in two distinct ways: via s and via t.

Example. Suppose $u = t\sin s$ and $s = x^2 + y^2$, $t = 2x + 4y$. Use the chain rule to find $\frac{\partial u}{\partial x}$ and $\frac{\partial u}{\partial y}$.

Answer: Applying the chain rule gives;

$$\begin{aligned}
\frac{\partial u}{\partial x} &= 2\sin s + 2xt\cos s = 2\sin\left(x^2+y^2\right) + 2x\left(2x+4y\right) \\
\frac{\partial u}{\partial y} &= 4\sin s + 2yt\cos s = 4\sin\left(x^2+y^2\right) + 2y\left(2x+4y\right)\cos\left(x^2+y^2\right).
\end{aligned}$$

[4] for functions of **one** variable only!

Note that for functions of more than one variable, it is very important to realise that $\frac{\partial x}{\partial t} \neq \left(\frac{\partial t}{\partial x}\right)^{-1}$ (in general). For instance, suppose x and t are both functions of two variables, say x depends on t and s, and t depends on x and y. Then, $\frac{\partial x}{\partial t}$ is the rate of change of x with respect to t when s is fixed but $\frac{\partial t}{\partial x}$ is the rate of change of t with respect to x when y is fixed.

We can use the chain rule to convert derivatives in one co-ordinate system to partial derivatives in another co-ordinate system.

Example. Suppose that a calculation or a given equation involves the partial derivatives $\frac{\partial u}{\partial x}$ and $\frac{\partial u}{\partial y}$ where x and y are understood to be Cartesian co-ordinates. Now suppose that it is preferable to think of u as a function of polar co-ordinates. Convert the derivatives $\frac{\partial u}{\partial x}$ and $\frac{\partial u}{\partial y}$ into derivatives with respect to r and θ.

Answer: We know that r and θ are both functions of x and y. That is,

$$r = \sqrt{x^2 + y^2}, \qquad \theta = \tan^{-1}(y/x).$$

When viewed as a function of polar co-ordinates, u depends on x via its dependence on r and θ. Hence, the chain rule gives:

$$\frac{\partial u}{\partial x} = \frac{\partial u}{\partial r}\frac{\partial r}{\partial x} + \frac{\partial u}{\partial \theta}\frac{\partial \theta}{\partial x}.$$

Similarly,

$$\frac{\partial u}{\partial y} = \frac{\partial u}{\partial r}\frac{\partial r}{\partial y} + \frac{\partial u}{\partial \theta}\frac{\partial \theta}{\partial y}.$$

Differentiating the expressions for r and θ with respect to x and y gives (**exercise**)

$$\frac{\partial u}{\partial x} = \cos\theta\frac{\partial u}{\partial r} - \frac{\sin\theta}{r}\frac{\partial u}{\partial \theta},$$

and similarly,

$$\frac{\partial u}{\partial y} = \sin\theta\frac{\partial u}{\partial r} + \frac{\cos\theta}{r}\frac{\partial u}{\partial \theta}.$$

Laplace's equation is a classical differential equation that we will study later in the course. **[See the handout on Classical PDEs]**. In two dimensions, this is written as

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

To convert this equation into say, polar co-ordinates, we need to convert the second derivatives with respect to x and y into derivatives with respect to r and θ. Given the above expressions for the first derivatives, you can now attempt this (see the last question on Exercise Sheet 1).

1.4. Solutions to PDEs

Solutions to ODEs usually have arbitrary constants in them. For example, consider the first-order ODE,

$$\frac{du}{dx} = -\lambda x.$$

To solve this, we can integrate both sides with respect to x to obtain

$$u(x) = \int -\lambda x \, dx = -\frac{\lambda x^2}{2} + C,$$

where C is any constant (note that this is an indefinite integral). In addition, consider

$$\frac{dv}{dx} = -\lambda v.$$

The solution is $v(x) = Ce^{-\lambda x}$, where C is any constant. In both cases, **infinitely** many solutions exist. To obtain a **unique** solution, extra conditions are needed. For example, if we know that $u(0) = 1$ in the first example, then

$$-\lambda(0) + C = 1 \Rightarrow C = 1,$$

and so the unique solution is

$$u(x) = -\frac{\lambda x^2}{2} + 1.$$

Solutions to **partial differential equations (PDEs)** contain arbitrary functions in place of arbitrary constants. For example, consider the simple PDE

$$\frac{\partial u}{\partial x} = 0$$

where we assume that u is a function of x and y. Integrating both sides with respect to x gives

$$u(x, y) = \int 0 \, dx = C(y),$$

where $C(y)$ is an arbitrary function of y (this acts like a constant when differentiating with respect to x). Check - any function of y (only) that you can think of satisfies the above partial differential equation. To obtain a unique solution, we need extra conditions such as $u(x, 0) = f(x)$ or $u(x, 1) = g(x)$ where $f(x), g(x)$ are specified functions.

Example. Find the unique solution $u(x, t)$ to the PDE $\frac{\partial u}{\partial x} = t + 1$ such that $u(0, t) = f(t)$.

Answer: Integrating with respect to x gives

$$u(x, t) = xt + x + C(t),$$

where $C(t)$ is an arbitrary function of t. Since $u(0, t) = f(t)$, we know $C(t) = f(t)$ (a specific function) so the unique solution is

$$u(x, t) = x(t + 1) + f(t).$$

In section 4, we will study a method called Separation of Variables for finding exact solutions to a certain class of partial differential equations (PDEs). To do this, it will be necessary to express a given function of **one** variable $f(x)$ as a series

$$f(x) = \sum_{n=1}^{\infty} c_n \phi_n(x), \qquad 0 < x < L, \tag{1}$$

called a **Fourier series**. In a Fourier series, the coefficients c_n are expressed as integrals and the functions ϕ_n are **orthogonal**. You met orthogonal vectors in the first year Linear Algebra course. You should revise your lecture notes on orthogonality from that course, and additionally, read the **handout on Orthogonal vectors**.

2.1. Orthogonal functions

It is impossible to expand **all** functions $f(x)$ and we certainly shouldn't write $=$ in (1) until we know whether the series converges. However, we can expand a certain class of functions, known as **piecewise continuous** functions.

Definition: Piecewise continuous (pwc). Given a function $f(x)$ on $[a, b]$, $f(x)$ is **piecewise continuous (pwc)** on $[a, b]$ if there exists a finite number of points x_n, with

$$a = x_0 < x_1 < \ldots < x_N = b$$

such that:

1. $f(x)$ is continuous on each open subinterval (x_{n-1}, x_n).
2. $f(x)$ has a finite limit at each end of each open interval.

In summary: $f(x)$ is allowed to have breaks/jumps as long as there are only a finite number of these, and $f(x)$ does not blow up (go to infinity) anywhere in $[a, b]$.

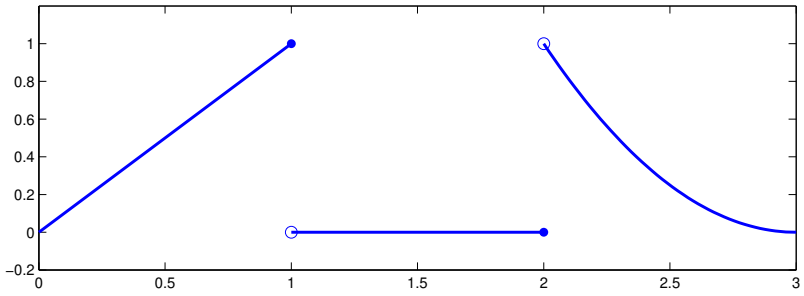


Figure 1: An example of a piecewise continuous function $f(x)$.

[1] Dr. C.E. Powell, School of Mathematics, University of Manchester.

Example. Consider the function $f(x)$ shown in Figure 1. $f(x)$ is pwc on $[0,3]$ (but not continuous) since $f(x)$ is continuous on each of the open subintervals $(0,1),(1,2),(2,3)$ and the limits of $f(x)$ at the end points of these open subintervals exist (are finite). In particular, $f(0^+)=0, f(1^-)=1$, $f(1^+)=1, f(2^-)=0$, and $f(2^+)=1, f(3^-)=0$.

Now, for pwc functions, we can define orthogonality. Compare the following definitions for functions with those you already know for vectors.

Definition: Inner-product. Given two functions $f(x), g(x)$ that are integrable on $[a,b]$, their **inner-product** is

$$(f,g) := \int_a^b f(x)g(x)\,dx.$$

Definition: Norm. The **norm** of an integrable function $f(x)$ on $[a,b]$ is

$$\|f\| := \sqrt{(f,f)} = \left(\int_a^b f(x)^2\,dx\right)^{1/2}.$$

Definition: Orthogonal. Two pwc functions $f(x)$ and $g(x)$ on $[a,b]$ are **orthogonal** when

$$(f,g) = \int_a^b f(x)g(x)\,dx = 0$$

and **orthonormal** when, in addition to being orthogonal, we have $\|f\| = 1 = \|g\|$.

2.2. Generalised Fourier Series

Suppose $f(x)$ is pwc on $[a,b]$ and let $\{\phi_n(x)\}_{n=1}^{\infty}$ be a specified set of orthogonal pwc functions on $[a,b]$. Does it make sense to write:

$$f(x) = \sum_{n=1}^{\infty} c_n\phi_n(x), \qquad 0<x<L\,?$$

We can't write $=$ until we have checked if the sum **converges**. If it converges, we can work out the coefficients easily. Taking the inner-product of both sides with any of the functions $\phi_i(x)$ gives

$$\int_a^b f(x)\phi_i(x)\,dx \;=\; \int_a^b \left(\sum_{n=1}^{\infty} c_n\phi_n(x)\right)\phi_i(x)\,dx = \sum_{n=1}^{\infty} c_n \int_a^b \phi_n(x)\phi_i(x)\,dx.$$

Since $(\phi_n,\phi_i)=0$ unless $n=i$, only one of the integrals in the sum is non-zero. Hence,

$$\int_a^b f(x)\phi_i(x)\,dx \;=\; c_i\int_a^b \phi_i(x)\phi_i(x)\,dx$$

and solving for c_i gives

$$c_i = \frac{\int_a^b f(x)\phi_i(x)\,dx}{\int_a^b \phi_i(x)^2\,dx} = \frac{(f,\phi_i)}{(\phi_i,\phi_i)}.$$

The series becomes

$$f(x) = \sum_{n=1}^{\infty} \frac{(f,\phi_n)}{(\phi_n,\phi_n)}\,\phi_n(x).$$

This is a **generalised Fourier series** and the coefficients c_n are called **Fourier coefficients**.

2.3. Fourier Series

Usually, when we talk about 'Fourier Series', we make a particular choice for the set of orthogonal functions $\{\phi_n(x)\}_{n=1}^{\infty}$.

Consider the integral

$$\int_{-L}^{L} \sin\left(\frac{n\pi x}{L}\right) \cos\left(\frac{m\pi x}{L}\right)\, dx, \quad n, m = 1, 2, \ldots.$$

Simplifying the integrand gives

$$\begin{aligned} & \frac{1}{2}\int_{-L}^{L} \sin\left(\frac{(n+m)\pi x}{L}\right) + \sin\left(\frac{(n-m)\pi x}{L}\right)\, dx \\ = \; & \frac{1}{2}\left[-\frac{L}{(n+m)\pi}\cos\left(\frac{(n+m)\pi x}{L}\right) - \frac{L}{(n-m)\pi}\cos\left(\frac{(n-m)\pi x}{L}\right)\right]_{-L}^{L} = 0. \end{aligned}$$

This tells us that the set of functions $\{\sin(\pi x/L), \sin(2\pi x/L), \ldots\}$ and the set $\{\cos(\pi x/L), \cos(2\pi x/L), \ldots\}$ are mutually orthogonal on the interval $[-L, L]$.

We can also show (see Exercise Sheet 2) that

$$\int_{-L}^{L} \cos\left(\frac{n\pi x}{L}\right) \cos\left(\frac{m\pi x}{L}\right) = \begin{cases} 0 & n \neq m \\ L & n = m \neq 0 \\ 2L & n = m = 0 \end{cases}$$

and

$$\int_{-L}^{L} \sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{m\pi x}{L}\right) = \begin{cases} 0 & n \neq m \\ L & n = m \neq 0 \end{cases}.$$

With a little more work (see Exercise Sheet 2), it can be shown that the set

$$\{1, \sin(\pi x/L), \cos(\pi x/L), \sin(2\pi x/L), \cos(2\pi x/L), \ldots\}$$

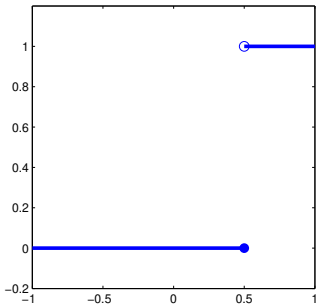
is an **orthogonal** set on $[-L, L]$ (we assume $L \in \mathbb{R}$ with $L > 0$). This is the set we will use for $\{\phi_n(x)\}_{n=1}^{\infty}$.

Definition: Fourier Series. The Fourier series of a function $f(x)$ that is pwc on $[-L, L]$ is

$$a_0 + \sum_{n=1}^{\infty} a_n \cos\left(\frac{n\pi x}{L}\right) + \sum_{n=1}^{\infty} b_n \sin\left(\frac{n\pi x}{L}\right)$$

with Fourier coefficients

$$\begin{aligned} a_0 &= \frac{1}{2L}\int_{-L}^{L} f(x)\, dx, \\ a_n &= \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right)\, dx, \quad n = 1, 2, \ldots, \\ b_n &= \frac{1}{L}\int_{-L}^{L} f(x)\sin\left(\frac{n\pi x}{L}\right)\, dx, \quad n = 1, 2, \ldots. \end{aligned}$$



Example. Compute the Fourier Series of the following pwc function

$$f(x) = \begin{cases} 0 & -1 \le x \le 1/2 \\ 1 & 1/2 < x \le 1 \end{cases}.$$

Direct integration gives:

$$a_0 = \frac{1}{2}\int_{-1}^{1} f(x)\,dx = \frac{1}{2}\int_{1/2}^{1} 1\,dx = \frac{1}{4}$$

(note that $L = 1$ here). Similarly,

$$a_n = \frac{1}{1}\int_{-1}^{1} f(x)\cos(n\pi x)\;dx = \int_{1/2}^{1} \cos(n\pi x)\;dx = -\frac{1}{n\pi}\sin\left(\frac{n\pi}{2}\right),$$

and

$$\begin{aligned} b_n &= \frac{1}{1}\int_{-1}^{1} f(x)\sin(n\pi x)\;dx = \int_{1/2}^{1} \sin(n\pi x)\;dx \\ &= \left[-\frac{1}{n\pi}\cos(n\pi x)\right]_{1/2}^{1} = -\frac{1}{n\pi}\cos(n\pi) + \frac{1}{n\pi}\cos\left(\frac{n\pi}{2}\right) \\ &= \frac{1}{n\pi}\left(\cos\left(\frac{n\pi}{2}\right) - (-1)^n\right). \end{aligned}$$

Putting all this together, the Fourier series associated with $f(x)$ is

$$\frac{1}{4} + \sum_{n=1}^{\infty}\left(-\frac{1}{n\pi}\sin\left(\frac{n\pi}{2}\right)\right)\cos(n\pi x) + \sum_{n=1}^{\infty}\frac{1}{n\pi}\left(\cos\left(\frac{n\pi}{2}\right) - (-1)^n\right)\sin(n\pi x).$$

Notice that we did not say $f(x)$ is equal to the Fourier series. A natural question now is: to what function does the Fourier series converge? Does it converge to $f(x)$ (the function in the figure above)? This is investigated on a separate **Handout on Fourier Series**, which you should now read.

The numerical investigation in MATLAB described in the handout reveals that the Fourier series is not equal to $f(x)$ everywhere. We observe that:

1. the Fourier series agrees with $f(x)$ on $[-1, 1]$ (i.e., converges to $f(x)$) **except** at three points: $x = -1$, $x = 0.5$ and $x = 1$. At those points, the Fourier series converges to the value $1/2$.

2. if we plot the Fourier series on $(-\infty, \infty)$ we just see copies of the series on $[-1, 1]$, shifted. The Fourier series is a periodic function.

Definition: Periodic function. A function $f(x)$ is periodic with period T if, for all x, $f(x+T) = f(x)$.

Examples: $\sin(3x), \sin(4\pi x), \tan(x), \ldots$ are periodic functions.

Given a function $f(x)$ defined on a fixed interval, we can always make a 'periodic version' of it, by extending it to the whole real number line.

Definition: Periodic extension. Let $f(x)$ be defined on $[-L, L]$. The periodic extension $\tilde{f}(x)$ of $f(x)$ is defined by

$$\tilde{f}(x) = \begin{cases} f(x) & -L \le x < L \\ \tilde{f}(x - 2L) & x \ge L \\ \tilde{f}(x + 2L) & x < -L \end{cases} .$$

It is easier to draw a picture. Basically, we take $f(x)$ on $[-L, L]$ and copy it on the adjacent intervals of length $2L$, taking care not to give $f(x)$ duplicate values at the end points of any of these subintervals. Of course, we cannot sketch $\tilde{f}(x)$ on the whole of $(-\infty, \infty)$, but it is best to include at least the intervals to the right and left of the principle interval: $[-3L, -L], [-L, L]$ and $[L, 3L]$.

Example. Sketch the periodic extension of the function

$$f(x) = \begin{cases} 0 & -1 \le x \le 1/2 \\ 1 & 1/2 < x \le 1 \end{cases} .$$

Using the definition, we copy the definition of $f(x)$ on $[-1, 1)$ (not including the right end point) and then on the interval $[1, 3)$ we copy the definition of $f(x)$ on $[-1, 1)$. Similarly, on the interval $[-3, -1)$ we copy the definition of $f(x)$ on $[-1, 1)$.

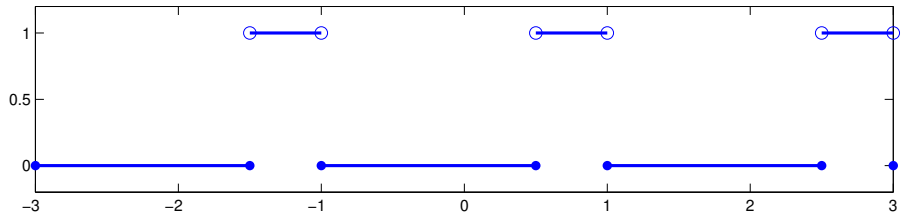


Figure 2: Periodic extension $\tilde{f}(x)$ of the function $f(x)$ defined on the fixed interval $[-1, 1]$.

Looking at the handout, we see that the Fourier series associated with $f(x)$ 'looks like' the periodic extension $\tilde{f}(x)$, except at the points where $\tilde{f}(x)$ jumps. This is not a coincidence. Fourier's Theorem explains the connection between the Fourier series of $f(x)$ and the periodic extension of $f(x)$. However, the Theorem is valid only for functions that are **piecewise smooth** (a stricter condition than pwc).

Definition: Piecewise smooth (pws). If $f(x)$ and $\frac{df}{dx}$ are piecewise continuous (pwc) on some partition of $[a, b]$ then $f(x)$ is piecewise smooth on $[a, b]$.

Fourier's Theorem. Let $g(x)$ be piecewise smooth on the interval $[-L, L]$ and periodic, with period $2L$. The Fourier series

$$a_0 + \sum_{n=1}^{\infty} a_n \cos\left(\frac{n\pi x}{L}\right) + \sum_{n=1}^{\infty} b_n \sin\left(\frac{n\pi x}{L}\right)$$

associated with $g(x)$ converges to

$$\frac{1}{2}\left(g(x^+) + g(x^-)\right)$$

at every $x \in (-\infty, \infty)$.

Notice that the theorem can only be applied to pws periodic functions. If we are given a pws function $f(x)$ on a fixed interval $[-L, L]$ then $\tilde{f}(x)$ is both pws and periodic. The periodic extension $\tilde{f}(x)$ agrees with $f(x)$ on $[-L, L)$ (the interval we care about). So, to learn about the Fourier series associated with $f(x)$, we can apply the theorem with $g(x) = \tilde{f}(x)$ (treat $f(x)$ as if it were a periodic function). Now, if $\tilde{f}(x)$ is a function that is continuous at x then

$$\frac{1}{2}\left(\tilde{f}(x^+) + \tilde{f}(x^-)\right) = \tilde{f}(x)$$

but at points where $\tilde{f}(x)$ jumps, the Fourier series converges to the **average** of $\tilde{f}(x^+)$ and $\tilde{f}(x^-)$. This explains why the Fourier series investigated on the handout converged to $1/2$ at $x = -1, 1/2$ and 1.

The Theorem is useful because we can use it to sketch a Fourier series of a given pws function $f(x)$ on $[-L, L]$ without having to compute the Fourier coefficients.

1. First, sketch the periodic extension of $f(x)$
2. At points of discontinuity, mark the average value
3. Pull out the piece of the graph that corresponds to the interval $[-L, L]$ of interest

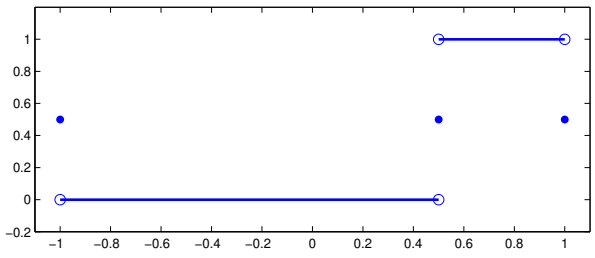


Figure 3: Fourier series of the function $f(x)$ on the interval $[-1, 1]$.

Example. Sketch the Fourier series associated with the function

$$f(x) = \begin{cases} 0 & -1 \le x \le 1/2 \\ 1 & 1/2 < x \le 1 \end{cases}.$$

Following the above steps, we obtain the following picture shown in Figure 3.

2.4 Fourier Sine and Cosine Series

Fourier series of odd and even function have special forms.

Definition: Odd function. A function $f(x)$ is odd if $f(-x) = -f(x)$ for all $x \in \mathbb{R}$.

Examples: x, x^{17}, $\sin(x)$ are odd functions.

Definition: Even function. A function $f(x)$ is even if $f(-x) = f(x)$ for all $x \in \mathbb{R}$.

Examples: x^2, x^{100}, $\cos(x)$ are even functions.

Consider the integral of an odd or an even function over a symmetric interval of the form $[-L, L]$. In general,

$$f(x)\,\text{odd} \Rightarrow \int_{-L}^{L} f(x)\,dx = 0$$

and

$$f(x)\,\text{even} \Rightarrow \int_{-L}^{L} f(x)\,dx = 2\int_{0}^{L} f(x)\,dx.$$

Now, suppose $g(x)$ is odd. The associated Fourier coefficients are:

$$\begin{aligned}
a_0 &= \frac{1}{2L}\int_{-L}^{L} g(x)\,dx = 0 \\
a_n &= \frac{1}{L}\int_{-L}^{L} \underbrace{g(x)}_{odd}\underbrace{\cos\left(\frac{n\pi x}{L}\right)}_{even}\,dx = 0 \\
b_n &= \frac{1}{L}\int_{-L}^{L} \underbrace{g(x)}_{odd}\underbrace{\sin\left(\frac{n\pi x}{L}\right)}_{odd}\,dx = \frac{2}{L}\int_{0}^{L} g(x)\sin\left(\frac{n\pi x}{L}\right)\,dx.
\end{aligned}$$

Above, we have used the fact that the product of an odd and an even function is odd, while the product of two odd functions is even (can you prove this?). Since $a_0 = a_n = 0$, the cosine terms and the constant term in the Fourier series drop out, leaving only the sine terms. The Fourier series of an odd function is a sine series. Using the above calculation for b_n, the Fourier series has the form

$$\sum_{n=1}^{\infty}\left(\frac{2}{L}\int_{0}^{L} g(x)\sin\left(\frac{n\pi x}{L}\right)\,dx\right)\sin\left(\frac{n\pi x}{L}\right).$$

Similarly, the Fourier series of an even function is called a cosine series, since all the coefficients in front of the sine terms are zero (see Exercise Sheet 3).

Now, consider the following question: If a function $f(x)$ is **not** odd, can we still represent it as a Fourier sine series? We will need to be able to do this when we apply the method of Separation of Variables, later on. Suppose $f(x)$ is pws on the half interval $[0, L]$. We can extend $f(x)$ to an odd function on the interval $[-L, L]$ (create an 'odd version' of it) in such a way that the extended function agrees with $f(x)$ on the original interval $[0, L]$. The Fourier series associated with this extended odd function will be a sine series. Similarly, any function $f(x)$ defined on $[0, L]$ can be extended to an even function on $[-L, L]$ and the associated Fourier series will be a cosine series.

Definition: Odd extension. Given $f(x)$ on $[0, L]$, the odd extension is

$$f_{\text{odd}}(x) = \begin{cases} f(x) & 0 \le x \le L \\ -f(-x) & -L \le x < 0 \end{cases}.$$

Definition: Even extension. Given $f(x)$ on $[0, L]$, the even extension is

$$f_{\text{even}}(x) = \begin{cases} f(x) & 0 \le x \le L \\ f(-x) & -L \le x < 0 \end{cases}.$$

Examples. Figure 4 shows two functions defined on $[0, 1]$ and their odd and even extensions on $[-1, 1]$.

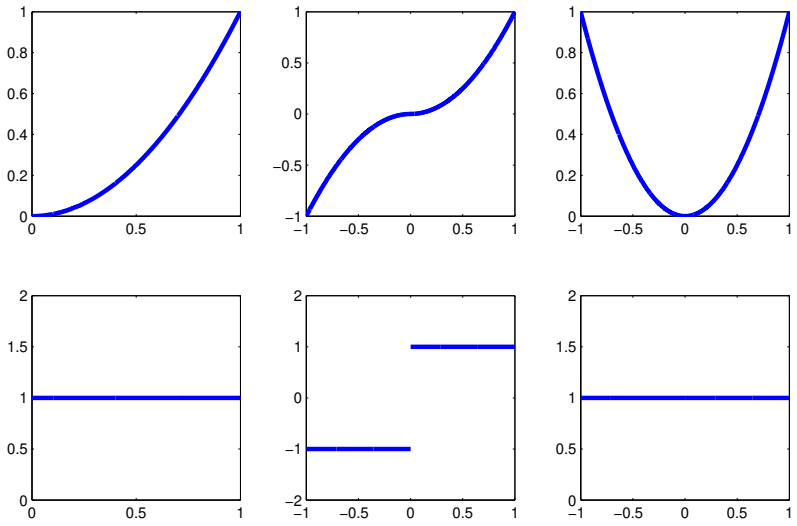


Figure 4: Left: $f(x)$ defined for $x \in [0, 1]$. Middle: odd extension $f_{\text{odd}}(x)$ on $[-1, 1]$. Right: even extension $f_{\text{even}}(x)$ on $[-1, 1]$.

Once again, using Fourier's Theorem, we can sketch a Fourier sine or cosine series, without integrating to find the coefficients.

Example. Find the Fourier sine series associated with the function $f(x) = 1$ on $[0, 1]$.

First, we sketch the Fourier sine series by applying the following steps.

1. Sketch $f(x)$ on $[0, 1]$ (see the bottom left plot in Figure 4).
2. Sketch $f_{\text{odd}}(x)$ on $[-1, 1]$ (see the bottom middle plot in Figure 4).
3. Sketch the periodic extension $\tilde{f}_{\text{odd}}(x)$ on $(-\infty, \infty)$
4. At points of discontinuity, mark the average value (i.e., apply Fourier's Theorem).
5. Extract the piece of the graph that corresponds to the interval $[0, 1]$ of interest.

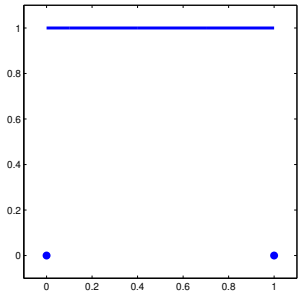


Figure 5: Fourier sine series associated with the function $f(x) = 1$ on $[0, 1]$.

This gives the graph in Figure 5. Notice that the Fourier sine series agrees with $f(x)$ in the interval $(0, 1)$ but converges to zero at the end points. To work out the sine series explicitly, we need to compute the coefficients of the Fourier series associated with the odd extension

$$f_{\text{odd}}(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ -1 & -1 \leq x < 0 \end{cases}.$$

Since this is an odd function, we know $a_0 = 0 = a_n$ and

$$b_n = \frac{2}{1}\int_0^1 f_{\text{odd}}(x)\sin(n\pi x)\, dx = 2\int_0^1 \sin(n\pi x)\, dx = \frac{2}{n\pi}\left(1 - (-1)^n\right).$$

Hence, the Fourier sine series is

$$\sum_{n=1}^{\infty} \frac{2}{n\pi}\left(1 - (-1)^n\right)\sin(n\pi x).$$

You should compare this with the Fourier cosine series of the same function (see Exercise Sheet 3).

In the next section, we will study a method called Separation of Variables for finding exact solutions to a certain class of partial differential equations (PDEs). In this section, we introduce div, grad notation, for writing PDEs in a compact way, and discuss some of the properties of PDEs. PDEs are classified into different groups, according to their properties. By now, you should have read the **handout on Classical PDEs**, which contains a list of all the PDEs you will meet in this course.

Definition: PDE, solution. A partial differential equation or *PDE* is an equation containing partial derivatives of an unknown function (say, u). A *solution* is a function which, when substituted for u, satisfies the PDE.

Example. The equation

$$\frac{\partial u}{\partial x} + \frac{\partial^2 u}{\partial y^2} = 0$$

is a PDE. The function $u = 6x - 3y^2$ is a solution.

A PDE (without extra conditions) will not have a unique solution. In fact, there will usually be infinitely many solutions. Conditions such as $u(x, 0) = g(x)$ (where $u = u(x, t)$ and t represents time) are called *initial conditions*. Conditions such as $u(0, t) = f(t)$ (where x represents space) are called *boundary conditions*. Unique solutions are only possible if extra conditions are supplied.

3.1. Grad, div notation

The symbol ∇ ('grad' or gradient) is a differential operator. When we apply it to a function, it produces the vector of first partial derivatives (in Cartesian coordinates) of that function.

- In 1d, with $u = u(x)$, we have $\nabla u = \frac{\partial u}{\partial x}$ (a one component vector).
- In 2d, with $u = u(x, y)$, we have $\nabla u = \left(\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}\right)$.
- In 3d, with $u = u(x, y, z)$, we have $\nabla u = \left(\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial u}{\partial z}\right)$.

Notice that ∇ acts on *scalar* functions and produces a *vector*. The symbol $\nabla\cdot$ ('div' or divergence) also represents a differential operator and is used on vector functions $\mathbf{F} = (F_1, F_2, F_3)$ as follows:

$$\nabla \cdot \mathbf{F} = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) \cdot (F_1, F_2, F_3) = \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z}.$$

We can use the two operators together in the following way.

$$\nabla \cdot (\nabla u) = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) \cdot \left(\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial u}{\partial z}\right) = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}.$$

The symbol ∇^2 is called the *Laplacian* operator. It is short-hand for $\nabla \cdot \nabla$. Hence, we apply it as

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

When working with PDEs, we use this notation as much as possible.

[1] Dr. C.E. Powell, School of Mathematics, University of Manchester.

3.2. Classification of PDEs

To know how to solve PDEs, we first need to classify them. Not all solution methods are suitable for all types of PDEs. We use the following words to classify PDEs:

- order
- linear or nonlinear
- homogeneous or non-homogeneous
- elliptic, parabolic or hyperbolic (for second order PDEs).

Definition: Order. The *order* of a PDE is the order of the highest derivative appearing in the equation.

Example. The PDE

$$\frac{\partial^3 u}{\partial x^2 \partial y} + \frac{\partial^2 u}{\partial x \partial y} = x$$

has order three (or 'is third order').

We can write any PDE in operator notation

$$\mathcal{L}u = f,$$

where u is the solution, $\mathcal{L}$ is a differential operator (that acts on u) and f is a collection of terms involving only the independent variables (and not u).

Example. The PDE

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} - 2x = 0$$

can be written as $\mathcal{L}u = f$ with

$$\mathcal{L} := \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}, \qquad f = 2x.$$

Definition: Linear. The operator $\mathcal{L}$ is *linear* if, for any two functions u_1, u_2 that are linearly independent, and any $c \in \mathbb{R}$,

1. $\mathcal{L}(u_1 + u_2) + \mathcal{L}u_1 + \mathcal{L}u_2$,
2. $\mathcal{L}(cu_1) = c\mathcal{L}u_1$.

If $\mathcal{L}$ is linear, then the PDE $\mathcal{L}u = f$ is *linear*. Otherwise, it is *nonlinear*.

Example. The PDE

$$\frac{\partial u}{\partial x} + \sin(u) = 0$$

is nonlinear because

$$\mathcal{L}(u_1 + u_2) = \frac{\partial (u_1 + u_2)}{\partial x} + \sin(u_1 + u_2) \neq \frac{\partial u_1}{\partial x} + \sin(u_1) + \frac{\partial u_2}{\partial x} + \sin(u_2) = \mathcal{L}u_1 + \mathcal{L}u_2.$$

However, the PDE $\frac{\partial u}{\partial x} + \sin(x) = 0$ is linear!

Basically, linear PDEs can have no terms in them in which the solution or its derivatives have powers greater than one, no products of derivatives, logs, trigonometric functions etc ...

Definition: Homogeneous. A PDE of the form $\mathcal{L}u = 0$ is homogeneous (i.e., there are no terms that do not involve the unknown u or its derivatives).

Example. The PDE $u_{tt} = u_{xx}$ can be written as $u_{tt} - u_{xx} = 0$, or $\mathcal{L}u = 0$ where

$$\mathcal{L} = \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2}.$$

Hence, it is homogeneous.

Definition: Elliptic, parabolic, hyperbolic. A second order PDE for a function $u(x,t)$ (i.e., a function of two variables) can be written:

$$Au_{xx} + Bu_{xt} + Cu_{tt} + Du_x + Eu_t + Fu + G = 0,$$

where A–F are functions of x and t only for linear problems and $G = 0$ for homogeneous problems. We say the PDE is

- elliptic if $B^2 - 4AC < 0$.
- parabolic if $B^2 - 4AC = 0$.
- hyperbolic if $B^2 - 4AC > 0$.

(Notice that this test involves only the coefficients of the second derivative terms).

Example. Laplace's equation in two dimensions for $u = u(x,y)$ is

$$\nabla^2 u = 0, \qquad \text{or} \qquad u_{xx} + u_{yy} = 0.$$

We have $A = 1, B = 0$ and $C = 1$ so $B^2 - 4AC = -4$. Since $-4 < 0$, Laplace's equation is elliptic.

On the **handout on Classical PDEs** you will also find examples of parabolic and hyperbolic PDEs. See also Exercise Sheet 4.

PDEs that are linear and homogeneous have a very important property. Let u_1 and u_2 be two linearly independent solutions to $\mathcal{L}u = 0$ and suppose that $\mathcal{L}$ is linear. Let $c_1, c_2 \in \mathbb{R}$ and define $v = c_1u_1 + c_2u_2$. Then, using the properties of $\mathcal{L}$ and the fact that u_1 and u_2 are solutions gives

$$\mathcal{L}v = \mathcal{L}(c_1u_1 + c_2u_2) = \mathcal{L}(c_1u_1) + \mathcal{L}(c_2u_2) = c_1\mathcal{L}(u_1) + c_2\mathcal{L}(u_2) = 0 + 0 = 0.$$

Hence, v is also a solution. Since c_1 and c_2 are arbitrary, there are infinitely many solutions. We can do a similar exercise if we start with three linearly independent solutions, then four, and five .... This leads us to the **Principle of Superposition**. This says that if we can find a set of linearly independent solutions of a linear and homogeneous PDE, then any linear combination of them is also a solution.

Theorem. Principle of Superposition. Let $\{u_1, u_2, \dots,\}$ be a set of infinitely many linearly independent solutions to $\mathcal{L}u = 0$ where $\mathcal{L}$ is linear. Then, for any $c_i \in \mathbb{R}$,

$$u = \sum_{i=1}^{\infty} c_i u_i$$

is also a solution.

The above result also says that linear and homogeneous PDEs may have solutions that are in the form of an infinite series. This should remind you of our earlier discussion of Fourier Series and indeed, we will link back to Fourier Series later.

Second-order, linear, homogeneous PDEs can be solved by the method of Separation of Variables. Below, we briefly discuss three model problems that belong to this class. One is elliptic, one is parabolic and one is hyperbolic. The heat equation and the wave equation were also discussed on the **handout on Classical PDEs**.

3.3. The one-dimensional heat equation

Consider a one-dimensional metal wire of length L. Let u denote the temperature in the wire. We assume $u = u(x,t)$ since the temperature will depend on time and also on the position along the wire. If there are no heat sources or sinks along the wire, then heat energy moves via the process of diffusion only. The mathematical model of this physical law is given by

$$\frac{\partial u}{\partial t} = K\frac{\partial^2 u}{\partial x^2},$$

or $u_t = K\nabla^2 u$, where K is the thermal conductivity coefficient. To obtain a unique solution, we need to specify boundary conditions and one initial condition. These will depend on the physical situation we want to model. Possibilites for **boundary conditions** include:

- $u(0,t) = 0$ and $u(L,0) = 0$ (the temperature is zero at the boundary).
- $u(0,t) = 10$ and $u(L,0) = 0$ (the temperature is ten at one end of the wire, but zero at the other).
- $u_x(0,t) = 0$ and $u_x(L,t) = 0$ (the flow of heat at the ends of the wire is zero, or, the boundaries are 'insulated').

For the **initial condition**, we need to specify an initial temperature at $t = 0$. That is, $u(x,0) = f(x)$, for some $f(x)$ that also agrees with the chosen boundary conditions.

Example: Heat Equation. Consider the heat equation

$$u_t = u_{xx}, \qquad 0 < x < 1,\ t > 0,$$

with $u(0,t) = 0$ and $u(1,t) = 0$ and initial condition

$$u(x,0) = \sin(\pi x) + \sin(3\pi x).$$

The solution is

$$u(x,t) = e^{-t}\sin(\pi x) + e^{-9t}\sin(3\pi x).$$

This is plotted for various values of t in Figure 1. Can you interpret what is happening physically? (Think of the different graphs as representing frames in a movie).

3.3. The one-dimensional wave equation

Consider a string, which when laid flat, has length L. Let $u(x,t)$ be the displacement of the string from the horizontal axis. Then, u satisfies the PDE

$$u_{tt} = c^2 u_{xx},$$

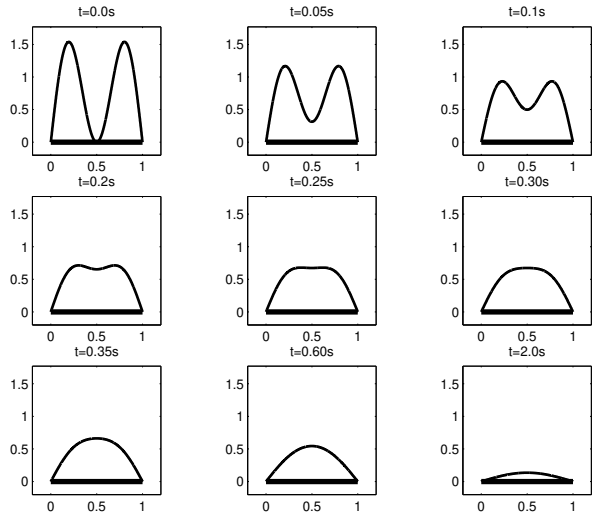


Figure 1: Snapshots of the solution $u(x,t)$ to the heat equation example at different values of t (time).

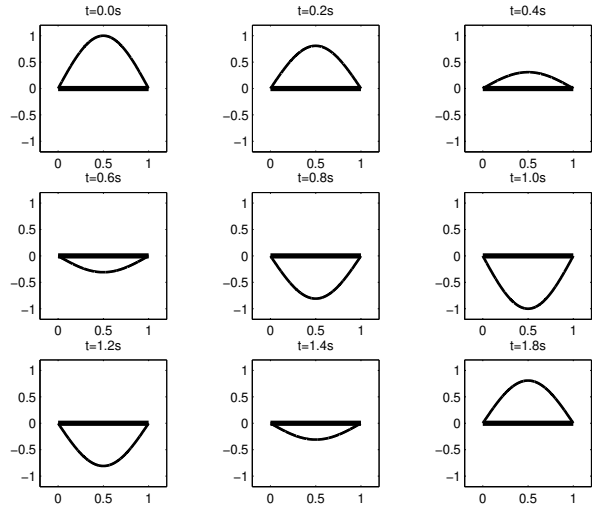


Figure 2: Snapshots of the solution $u(x,t)$ to the wave equation example at different values of t (time).

where c^2 is the wave speed (related to the properties of the string). If the string is fixed to the horizontal axis at both ends (at $x = 0$ and $x = L$) then there is no displacement and the boundary conditions are simply

$$u(0,t) = 0, \quad u(L,t) = 0.$$

We also need **two** initial conditions. For example,

$$u(x,0) = f(x), \qquad u_t(x,0) = g(x),$$

for given functions $f(x)$ and $g(x)$.

Example: Wave Equation. Consider the wave equation

$$u_{tt} = u_{xx}, \qquad 0 < x < 1,\ t > 0,$$

with $u(0,t) = 0$ and $u(1,t) = 0$ and initial conditions

$$u(x,0) = \sin(\pi x) \quad \text{and} \quad \frac{\partial u(x,0)}{\partial t} = 0.$$

The solution is

$$u(x,t) = \sin(\pi x)\cos(\pi t).$$

This is plotted for various values of t in Figure 2. Can you interpret what is happening physically?

3.4. Laplace's equation in two dimensions

Consider the PDE $\nabla^2 u = 0$ or $u_{xx} + u_{yy} = 0$. Notice that this is the heat eqation in two dimensions with $u_t = 0$. Laplace's equation can therefore be used to model steady-state temperatures. Suppose then, that we want to model the temperature u in a rectangular plate represented by the region $[0, L_x] \times [0, L_y]$. Here, the temperature is a function of the two space coordinates x and y but does not depend on time, t. The lengths L_x and L_y give the dimensions of the plate. To find a unique steady-state temperature, we need to supplement the PDE with four boundary conditions (one for each side of the plate), such as

$$u(x,0) = f_1(x), \quad u(x,L_y) = f_2(x), \quad u(0,y) = g_1(y), \quad u(L_x,y) = g_2(x).$$

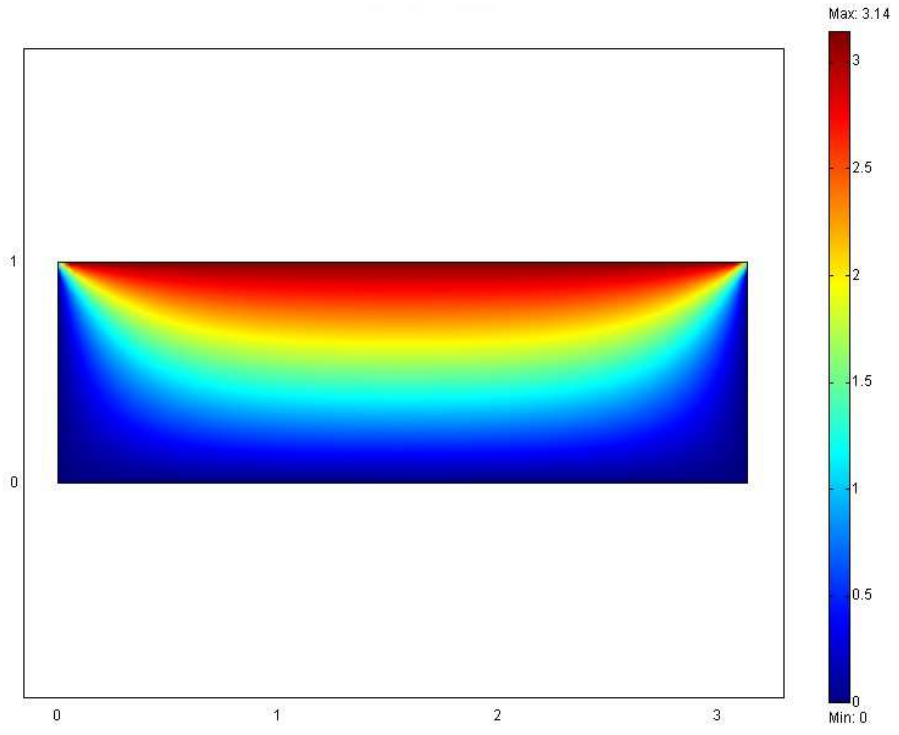


Figure 3: MATLAB surf plot of the solution $u(x,t)$ to Laplace's equation.

Example: Laplace's Equation. Consider the PDE

$$u_{xx} + u_{yy} = 0, \qquad 0 < x < 3,\, 0 < y < 1,$$

with boundary conditions

$$u(x, 0) = 0, \quad u(x, 3) = \pi, \quad u(0, y) = 0, \quad u(1, y) = 0.$$

A MATLAB surf plot of the solution is shown in Figure 3. In this case, there is no neat expression we can write down for $u(x, y)$. The solution is an infinite series. We will study how to find it in the next section.

Few PDEs have *exact* solutions. Even if exact solutions do exist, they are often too expensive to find or to use in practical situations. In real life applications (for example, weather forecasting, aerospace engineering), scientists use computational methods to *approximate* the solutions. We will look at one such family of methods, called finite difference methods, in the next section. Separation of Variables can only be used to find the exact solution to second-order, linear, homogeneous PDEs (like the heat equation, the wave equation and Laplace's equation) with linear, homogeneous boundary conditions. The method has several steps and is best learned by solving specific examples (see below). The general ideas are also summarised on the **handout on Separation of Variables**.

4.1. The one-dimensional heat equation

Consider the following problem. Find $u(x,t)$ satisfying

$$\frac{\partial u}{\partial t} = K\frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < L, \quad t > 0,$$

together with the boundary conditions

$$u(0,t) = 0, \qquad u(L,t) = 0,$$

and the initial condition

$$u(x,0) = u_0(x).$$

The first thing to note is that we are interested in non-zero (non-trivial) solutions. We look for 'separated solutions' of the form

$$u(x,t) = X(x)T(t)$$

where X is a function of x only and T is a function of t only. Substituting this expression into the PDE gives:

$$\frac{\partial (XT)}{\partial t} = K\frac{\partial^2 (XT)}{\partial x^2}$$

and since X does not depend on t and T does not depend on x, this becomes

$$X\frac{dT}{dt} = KT\frac{d^2X}{dx^2}.$$

Notice that we no longer have partial derivatives. The next step is to re-arrange the equation so that everything that depends on x is on one side, and everything that depends on t is on the other. We have

$$\frac{1}{KT}\frac{dT}{dt} = \frac{1}{X}\frac{d^2X}{dx^2}.$$

Note that is is ok to divide by X and T. Since we want non-zero solutions, $u = XT$ is not the zero function, so X and T are not the zero function. We have a choice about which side of the equation to put the constant K. Here, we'll put it on the t side but it doesn't really matter.

Now, notice that the left-hand side of the above equation is a **function of t only** and the right-hand side is a **function of x only**. Since x and t are independent variables, the two sides of the equation

[1] Dr. C.E. Powell, School of Mathematics, University of Manchester.

can only be equal if they are both equal to a constant. We don't know this constant (yet) but we'll call it λ (the separation constant). Hence

$$\frac{T'}{KT} = \frac{X''}{X} = \lambda.$$

(We are now using dash notation as short-hand for the derivatives). Hence,

$$T' - \lambda KT = 0 \quad \text{and} \quad X'' - \lambda X = 0.$$

The PDE has been transformed into two decoupled ODEs![2]

If we knew λ, then we would know the general form of the solutions $X(x)$ and $T(t)$ to our two ODEs. Unfortunately, we don't know λ so we have to consider three different cases: $\lambda = 0$, $\lambda > 0$ and $\lambda < 0$. Any (or all) of these might be possible. We check them all. First, consider the PDE for X. We can only find solutions if we have boundary conditions for X. We know that $u = XT$ satisfies the original boundary conditions, so

$$u(0,t) = 0 \Rightarrow X(0)T(t) = 0 \Rightarrow X(0) = 0$$

(since $T(t)$ is not the zero function). Similarly,

$$u(L,t) = 0 \Rightarrow X(L)T(t) = 0 \Rightarrow X(L) = 0.$$

We now have two boundary conditions for X and we need to solve

$$X''(x) - \lambda X(x) = 0, \quad 0 < x < 1, \quad \text{such that} \quad X(0) = 0 = X(L).$$

The above equation is called an *eigenvalue problem*. To solve it, we need to find both $X(x)$ and λ. Non-zero solutions $X(x)$ are called *eigenfunctions* and the corresponding values λ are called *eigenvalues*. We test possible values for λ to see whether for those values, non-zero solutions $X(x)$ exist.

$\lambda = 0$ If $\lambda = 0$ then the ODE becomes $X''(x) = 0$ and the general solution is $X(x) = Ax + B$. The boundary condition $X(0) = 0$ tells us that $B = 0$ and the condition $X(L) = 0$ gives $AL = 0$. Since $L \neq 0$, we have $A = 0$. Since $A = B = 0$, $X(x) = 0$. Hence, **there are no non-zero solutions** when $\lambda = 0$.

$\lambda > 0$ If $\lambda > 0$ then we can write $\lambda = \omega^2$ for some $\omega \in \mathbb{R}$ with $\omega \neq 0$. The general solution to the ODE in this case[3] is

$$X(x) = Ae^{\omega x} + Be^{-\omega x}.$$

Imposing the boundary conditions $X(0) = 0$ and $X(L) = 0$ once again gives $A = B = 0$ so **there are no non-zero solutions** when $\lambda > 0$.

$\lambda < 0$ If $\lambda < 0$ then we can write $\lambda = -\omega^2$ for some $\omega \in \mathbb{R}$ with $\omega \neq 0$. The general solution in this case is

$$X(x) = A\cos(\omega x) + B\sin(\omega x).$$

The condition $X(0) = 0$ gives $A = 0$. The condition $X(L) = 0$ then gives $B\sin(\omega L) = 0$. If $B = 0$ then we only have the zero solution again. Non-zero solutions can exist when $\sin(\omega L) = 0$. In

[2] In first year calculus courses you studied methods for solving ODEs. It is important now that you remember these methods. Revise your notes if you have forgotten.

[3] Perhaps you remember that the characteristic or auxiliary equation is $m^2 - \lambda = 0$. To determine the general form of the solution, we examine the roots of this equation....

other words, when $\omega L = n\pi$ for any $n \in \mathbb{Z} \setminus \{0\}$. We don't include zero because we know that $\omega L \neq 0$. The eigenvalues are therefore

$$\lambda = -\omega^2 = -\left(\frac{n\pi}{L}\right)^2, \quad n = 1, 2, \ldots$$

Note that we don't need to include negative values of n now because we only want to record all the distinct eigenvalues. To emphasis that there are infinitely many eigenvalues, one for each value of n, we can also write

$$\lambda_n = -\left(\frac{n\pi}{L}\right)^2, \quad n = 1, 2, \ldots$$

The corresponding non-zero solution X_n associated with λ_n is

$$X_n(x) = B_n \sin\left(\frac{n\pi x}{L}\right), \quad n = 1, 2, \ldots$$

Here B_n is any non-zero constant. However, when we talk about 'eigenfunctions' we usually just set these constants to one and write

$$X_n(x) = \sin\left(\frac{n\pi x}{L}\right), \quad n = 1, 2, \ldots$$

So far, we have solved the ODE for $X(x)$ and found the corresponding values of λ but we have not yet solved the second ODE for $T(t)$. The good news is that we don't need to test λ again. We've already found the values λ_n that give non-zero solutions $X_n(x)$. We have one solution $T_n(x)$ for each λ_n satisfying

$$T_n'(t) - \lambda_n K T_n(t) = 0.$$

This is a first order ODE with general solution[4]

$$T_n(t) = C_n e^{K\lambda t}, \quad n = 1, 2, \ldots$$

(where C_n is an arbitrary constant). Putting everything together, we have found infinitely many separated solutions to the heat equation of the form

$$u_n(x,t) = X_n(x)T_n(t) = A_n e^{-K(n\pi/L)^2 t} \sin\left(\frac{n\pi x}{L}\right), \quad n = 1, 2, \ldots,$$

for arbitrary constants A_n. (Note that when we multiply two arbitrary constants B_n and C_n, we just get another arbitrary constant).

So, have we solved the problem? Not quite. Each function $u_n(x,t)$ satisfies the PDE and also the boundary conditions but we have not yet used the initial condition. In general, none of the functions $u_n(x,t)$ will satisfy the initial condition $u_n(x,0) = u_0(x)$ (unless $u_0(x)$ is a very special function). To find a solution that does satisfy the initial condition, the final step is to apply the Principle of Superposition. Since the PDE and the boundary conditions are linear and homogeneous, and the functions $u_n(x,t)$ are linearly independent [5], we know that any linear combination is also a solution. Consider

$$u(x,t) = \sum_{n=1}^{\infty} A_n e^{-K(n\pi/L)^2 t} \sin\left(\frac{n\pi x}{L}\right). \tag{1}$$

[4] You know this from first year! Recall the method of integrating factors, if it is not obvious.

[5] We haven't actually proved this but the eigenfunctions are always linearly independent.

This is also a solution to the heat equation and it satisfies zero boundary conditions. We can make it satisfy the initial condition if we can find coefficients A_n so that

$$u(x,0) = \sum_{n=1}^{\infty} A_n \sin\left(\frac{n\pi x}{L}\right) = u_0(x).$$

In other words, if the function $u_0(x)$ can be written as a Fourier Sine Series, with Fourier coefficients A_n! Using the orthogonality property of the eigenfunctions (i.e., the sine functions), we know that the Fourier coefficients are

$$A_n = \frac{\int_0^L u_0(x)\sin(n\pi/L)\,dx}{\int_0^L \sin^2(n\pi/L)\,dx} = \frac{2}{L}\int_0^L u_0(x)\sin(n\pi/L)\,dx. \tag{2}$$

Hence, the final solution is the infinite series $u(x,t)$ given in (1) with coefficients defined in (2).

Summary. It is a good idea to remember the **steps** of the method. Do not try and memorize the answer to the above problem! The boundary conditions are key. They determine the eigenvalues λ_n and the eigenfunctions $X_n(x)$. Changing just one boundary conditions leads to a completely different solution. Try question 4 on Exercise Sheet 5 and compare your answer with the above calculation.

4.2. The one-dimensional wave equation

Consider the following problem. Find $u(x,t)$ satisfying

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < L, \quad t > 0,$$

such that

$$\frac{\partial u(0,t)}{\partial x} = 0, \qquad \frac{\partial u(L,t)}{\partial x} = 0,$$

and

$$u(x,0) = u_0(x), \qquad \frac{\partial u(x,0)}{\partial t} = u_1(x).$$

We can apply the method of Separation of Variables exactly as we did above for the heat equation but there are two key differences.

- two initial conditions need to be imposed after we find all the separated solutions
- we obtain a second order ODE for $T(t)$

You are asked to find solutions to the wave equation in questions 1 and 2 on Exercise Sheet 6, so full details are not given here. To start off, we set $u(x,t) = X(x)T(t)$ and substitute this into the PDE to obtain

$$XT'' = c^2 TX''$$

or,

$$\frac{T''}{c^2 T} = \frac{X''}{X} = \lambda,$$

where λ is the separation constant. We obtain two second order ODEs:

$$X'' - \lambda X = 0, \qquad T'' - c^2\lambda T = 0.$$

The ODE for $X(x)$ is the same as before, but now the boundary conditions are different. The first task is to use the boundary conditions for u to obtain boundary conditions for X and then find the eigenvalues λ and the corresponding eigenfunctions (non-zero solutions) X. Once we know λ, we can then also solve the ODE for $T(t)$.

4.3. Laplace's equation in two dimensions

Consider the following problem. Find $u(x, y)$ satisfying

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0, \qquad 0 < x < a, \quad 0 < y < b.$$

Can we use Separation of Variables? Following the same steps in the examples above, we can start by substituting $u(x, y) = X(x)Y(y)$ into the PDE to give $YX'' + XY'' = 0$. Rearranging gives

$$\frac{X''}{X} = -\frac{Y''}{Y} = \lambda,$$

where λ is the separation constant. We obtain two ODEs

$$X'' - \lambda X = 0, \qquad Y'' + \lambda Y = 0.$$

Note the difference in the sign! To make any progress, we'll need boundary conditions for both $X(x)$ and $Y(y)$.

Earlier, we said that 'Separation of Variables can only be used for problems with homogeneous (i.e. zero) boundary conditions' but if we apply zero boundary conditions everywhere on $[0, a] \times [0, b]$ then the only solution to Laplace's equation is $u(x, y) = 0$. In fact, what we really need is the property that when we add up all the separated solutions, the series solution should still satisfy the boundary conditions. If the boundary conditions are homogeneous (i.e., zero), then adding up infinitely many functions that are zero on the boundary still produces something that is zero on the boundary. If we add up infinitely many separated solutions that are equal to say, one on the boundary, then the sum of the solutions is not equal to one on the boundary. In fact, the sum blows up. For Laplace's equation on a rectangle, we can have zero boundary conditions on three sides and a non-zero condition on the fourth side. We treat the non-zero condition like an 'initial condition' in the sense that we only impose it only on the series solution, not on the individual separated solutions.

Consider the boundary conditions

$$u(0, y) = 0, \quad u(a, y) = 0, \quad u(x, 0) = 0, \quad u(x, b) = f(x).$$

Substituting $u(x, y) = X(x)Y(y)$ in the first three of these conditions gives $X(0) = 0, X(a) = 0$ and $Y(0) = 0$. Since we have a full set of boundary conditions for the ODE $X'' - \lambda X = 0$, we solve that one first. We've solved this ODE before (in the heat equation example) but it doesn't hurt to review it again here. We test the possible cases for λ as follows.

$\lambda = 0$ If $\lambda = 0$ then the ODE becomes $X''(x) = 0$ and the general solution is $X(x) = Ax + B$. The boundary conditions $X(0) = 0$ and $X(a) = 0$ tell us that **there are no non-zero solutions** in this case.

$\lambda > 0$ If $\lambda > 0$ then we can write $\lambda = \omega^2$ for some $\omega \in \mathbb{R}$ with $\omega \neq 0$. The general solution to the ODE in this case is

$$X(x) = Ae^{\omega x} + Be^{-\omega x}.$$

Imposing the boundary conditions $X(0) = 0$ and $X(a) = 0$ reveals that there are also **no non-zero solutions**.

$\lambda < 0$ If $\lambda < 0$ then we can write $\lambda = -\omega^2$ for some $\omega \in \mathbb{R}$ with $\omega \neq 0$. The general solution in this case is

$$X(x) = A\cos(\omega x) + B\sin(\omega x).$$

The condition $X(0) = 0$ gives $A = 0$. The condition $X(a) = 0$ then gives $B\sin(\omega a) = 0$. If $B = 0$ then we only have the zero solution again. Non-zero solutions exist when $\sin(\omega a) = 0$. In other words, when $\omega a = n\pi$ for any $n \in \mathbb{Z} \setminus \{0\}$ (since $\omega L \neq 0$). The eigenvalues are therefore

$$\lambda_n = -\omega^2 = -\left(\frac{n\pi}{a}\right)^2, \quad n = 1, 2, \ldots$$

The non-zero solution X_n associated with λ_n is

$$X_n(x) = B_n \sin\left(\frac{n\pi x}{a}\right), \quad n = 1, 2, \ldots$$

where B_n is an arbitrary constant.

Next, we need to solve

$$Y_n''(y) + \lambda_n Y_n(y) = 0, \quad \text{such that} \quad Y_n(0) = 0,$$

where we know that $\lambda_n < 0$. In this case the general form of the solution is

$$Y_n(y) = A_n e^{\omega_n y} + B_n e^{-\omega_n y}.$$

The boundary condition $Y_n(0) = 0$ tells us that $A_n + B_n = 0$ or $B_n = -A_n$. Hence,

$$Y_n(y) = A_n\left(e^{\omega_n y} - e^{-\omega_n y}\right) = 2A_n \sinh(\omega_n y) = 2A_n \sinh\left(\frac{n\pi y}{a}\right).$$

We have found infinitely many separated solutions

$$u_n(x, y) = X_n(x)Y_n(y) = A_n \sin\left(\frac{n\pi x}{a}\right)\sinh\left(\frac{n\pi y}{a}\right), n = 1, 2, \ldots.$$

(The product of two arbitrary constants is just an arbitrary constant). None of these individual solutions satisfies the fourth boundary condition. Applying the Principle of Superposition,

$$u(x, y) = \sum_{n=1}^{\infty} A_n \sin\left(\frac{n\pi x}{a}\right)\sinh\left(\frac{n\pi y}{a}\right)$$

is also a solution to the the PDE, and satisfies the three zero boundary conditions. To impose the final non-zero boundary condition, we need to find coefficients A_n such that

$$f(x) = u(x, b) = \sum_{n=1}^{\infty} A_n \sin\left(\frac{n\pi x}{a}\right)\sinh\left(\frac{n\pi b}{a}\right).$$

Using the orthogonality property of the sine functions (on the interval $[0, a]$), we have

$$A_n \sinh\left(\frac{n\pi b}{a}\right) = \frac{\int_0^a f(x)\sin\left(\frac{n\pi x}{a}\right)\,dx}{\int_0^a \sin^2\left(\frac{n\pi x}{a}\right)\,dx}.$$

Calculating the integral on the denominator and rearranging gives

$$A_n = \left(\frac{2}{a}\right)\frac{1}{\sinh\left(\frac{n\pi b}{a}\right)}\int_0^a f(x)\sin\left(\frac{n\pi x}{a}\right)\,dx.$$

4.4. The wave equation on a disk

Consider a circular drum (i.e., a thin circular membrane, fixed to a rigid circular frame) of radius a. We can solve the wave equation to model vibrations in the membrane, given its initial shape, for example at $t = 0$. Recall, the wave equation is

$$\frac{\partial^2 u}{\partial t^2} = \nabla^2 u$$

(where we have set $c^2 = 1$ for simplicity). In two space dimensions, this is

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}.$$

On circular (two-dimensional) geometries it is best to work in polar coordinates. So, using the chain rule for partial derivatives, we can transform this into

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}.$$

To simplify things a bit, let's assume

- the vibrations are radially symmetric (do not depend on θ)
- the interesting vibrations correspond to negative values of the separation constant only (the case $\lambda < 0$ or $\lambda = -\omega^2$).

The first assumption means that u is a function of t and r only and we need to solve

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r}, \qquad 0 < r < a, \quad t > 0.$$

This reduced problem is one-dimensional in space. (Starting at the centre of the drum and drawing a straight line to the boundary, we've effectively taken a one-dimensional slice). We'll need two boundary conditions, one at $r = 0$ and one at $r = a$. If the membrane is fixed to the frame, then the vibration is zero there. The correct boundary condition is

$$u(a, t) = 0.$$

At $r = 0$ we will impose the condition

$$| u(0, t) | < \infty.$$

This simply says that the size of the vibrations at the centre of the drum are finite.

To apply separation of variables, we substitute $u(r, t) = R(r)T(t)$ into the PDE to obtain two ODEs

$$r^2 R''(r) + rR'(r) - \lambda r^2 R(r) = 0,$$

and

$$T''(t) - \lambda T(t) = 0.$$

Substituting into the boundary conditions gives

$$R(a) = 0, \qquad | R(0) | < \infty.$$

Since we have two boundary conditions for R, we can solve the ODE for $R(r)$ first, using the assumption that $\lambda = -\omega^2$ for some $\omega \in \mathbb{R}$.

The above ODE for $R(r)$ is perhaps one that you have not met before. It is called Bessel's equation of order zero. The general form of the solution is

$$R(r) = AJ_0(\omega r) + BY_0(\omega r),$$

where A, B are arbitrary constants. The functions Y_0 and J_0 are special functions called Bessel functions. More details about these functions, and how to investigate them in MATLAB are available on the **handout on Bessel functions**. A key point is that $J_0(0)$ is finite but Y_0 blows up at the origin. So, to satisfy the boundary condition at $r = 0$, we must have $B = 0$. The boundary conditions at $r = a$ then gives

$$0 = R(r) = AJ_0(\omega a).$$

If A is zero, then we only have $R(r) = 0$. Non-zero solutions exist when $J_0(\omega a) = 0$. That is, when ωa is a zero of the function J_0. In fact, J_0 has infinitely many zeros (just like sine and cosine). Let α_n denote a zero of J_0, for $n = 1, 2, \ldots$ then we have

$$\omega_n = \frac{\alpha_n}{a}, \qquad n = 1, 2, \ldots.$$

The eigenvalues are

$$\lambda_n = -\left(\frac{\alpha_n}{a}\right)^2, \qquad n = 1, 2, \ldots,$$

and the eigenfunctions are

$$R_n(r) = A_n J_0(\omega_n r/a), \qquad n = 1, 2, \ldots.$$

Next, we need to solve for $T_n(t)$ and combine with $R_n(r)$ to find the separated solutions $u_n(r, t) = R_n(r)T_n(t)$.

Finite difference methods are schemes for approximating the solution to an ODE or PDE, at a discrete set of grid points. We begin by studying schemes for the reaction-diffusion equation and the convection-diffusion equation in one dimension (both ODEs) and then discuss finite differences for the one-dimensional heat equation (a PDE).

5.1. The reaction-diffusion equation

Consider the following problem. Find $u(x)$ satisfying

$$-\frac{d^2u(x)}{dx^2} + r(x)\,u(x) = f(x), \qquad 0 < x < 1,$$

together with the boundary conditions

$$u(0) = 0, \qquad u(1) = 0.$$

The coefficient $r(x)$ is called the reaction coefficient. It controls the strength of the reaction and hence, the size of the reaction term compared to the diffusion term. We will assume that $r(x)$ and $f(x)$ are real-valued functions and a unique solution exists. When working with approximation schemes, it is useful to find examples with exact solutions, to investigate the accuracy of the approximation.

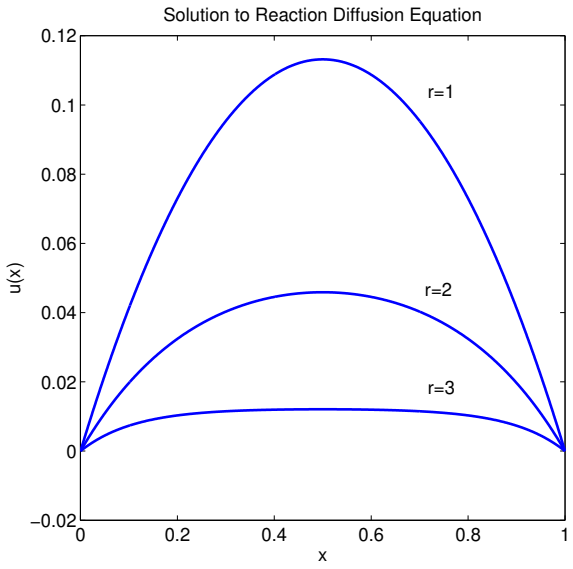


Figure 1: The exact solution to the reaction-diffusion equation when $r = 1, 2, 3$ and $f = 1$.

Example. Suppose $r(x) = 0$ and $f(x) = 1$. Clearly, the exact solution is $u(x) = \frac{1}{2}(x - x^2)$.

Example. Suppose $r(x) = \omega^2$ and $f(x) = 1$. The exact solution (see Figure 1) is

$$u(x) = \frac{1}{\omega^2} - \frac{e^{\omega x} + e^{\omega(1-x)}}{\omega^2(1 + e^{\omega})}.$$

[1] Dr. C.E. Powell, School of Mathematics, University of Manchester.

To approximate $u(x)$ we apply the following steps.

1. Set up a grid of uniformly spaced points on the interval $[0,1]$ (the interval on the x-axis on which we want to solve the ODE). If we divide $[0,1]$ into N sub-intervals of width $h = 1/N$, there are $N+1$ grids points x_j with

$$x_0 = 0, \quad x_1 = h, \quad x_2 = 2h, \quad \ldots, \quad x_N = Nh = 1.$$

2. Write down the differential equation at each grid point. We have

$$-\frac{d^2u(x_j)}{dx^2} + r(x_j)\,u(x_j) = f(x_j), \qquad j = 1, 2, \ldots, N-1,$$

   and the boundary conditions give

$$u(x_0) = 0, \qquad u(x_N) = 0.$$

3. Approximate the derivatives by **finite difference formulae**.

4. Solve the resulting set of $N-1$ algebraic equations for approximate values $U_j \approx u(x_j)$.

There are lots of possibilities in step 3. We will consider only a 'centred' finite difference scheme.

Definition: Centred Difference. The centred difference δu of $u(x)$ at $x = x_j$ is defined as

$$\delta u(x_j) = u\left(x_j + \frac{h}{2}\right) - u\left(x_j - \frac{h}{2}\right).$$

Applying the above definition twice (exercise) gives the second centred difference

$$\delta^2 u(x_j) = \delta\left(\delta u(x_j)\right) = u(x_{j+1}) - 2u(x_j) + u(x_{j-1}).$$

Notice that if we divide by the distance between the x-values where u is evaluated, then we can use the centred difference $\delta u(x_j)$ to approximate the first derivative. That is

$$\frac{1}{h}\delta u(x_j) = \frac{u\left(x_j + \frac{h}{2}\right) - u\left(x_j - \frac{h}{2}\right)}{h} \approx \frac{du(x_j)}{dx}.$$

Since the second derivative is the first derivative of the first derivative, we also have

$$\frac{1}{h^2}\delta^2 u(x_j) \approx \frac{d^2u(x_j)}{dx^2}.$$

If we make this approximation in the equations written in step 2, then the solution is no longer $u(x_j)$ but an approximate value $U_j \approx u(x_j)$. We have

$$-\frac{U_{j+1} - 2U_j + U_{j-1}}{h^2} + r_j\,U_j = f_j, \qquad j = 1, 2, \ldots N-1,$$

where we have also used the short-hand notation $r_j = r(x_j), f_j = f(x_j)$. At the boundaries, there is no need to make an approximation, because we already know the solution there so

$$U_0 = u(x_0) = 0, \qquad U_N = u(x_N) = 0.$$

Rearranging the finite difference equations gives

$$-\frac{1}{h^2}U_{j-1} + \left(\frac{2}{h^2} + r_j\right)U_j - \frac{1}{h^2}U_{j+1} = f_j, \qquad j = 1, 2, \ldots, N-1.$$

This is the centred finite difference scheme for the reaction-diffusion equation. We can write the $N-1$ equations for the approximations U_j as a matrix problem

$$\begin{pmatrix} b_1 & c & & & & \\ a & b_2 & c & & & \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & \ddots & c \\ & & & & a & b_{N-1} \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ \vdots \\ \vdots \\ U_{N-1} \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ \vdots \\ \vdots \\ f_{N-1} \end{pmatrix},$$

where

$$a = -\frac{1}{h^2}, \qquad b_j = \frac{2}{h^2} + r_j, \quad c = -\frac{1}{h^2}.$$

So, we have to solve a tridiagonal system of equations to obtain approximate values $U_1, \ldots, U_{N-1}$. See the **handout on Centred Finite Differences for the Reaction-Diffusion Equation** for some examples and details about how to solve this system in MATLAB.

The above centred finite difference equations were derived from the approximation for the second derivative

$$\frac{1}{h^2}\delta^2 u(x_j) \approx \frac{d^2u(x_j)}{dx^2}.$$

Is this a good approximation? First, we consider the following Taylor series expansions about x_j

$$u(x_{j+1}) = u(x_j + h) = u(x_j) + hu'(x_j) + \frac{h^2}{2}u''(x_j) + \frac{h^3}{6}u'''(x_j) + \frac{h^4}{24}u^{(4)}(x_j) + \ldots$$

and

$$u(x_{j-1}) = u(x_j - h) = u(x_j) - hu'(x_j) + \frac{h^2}{2}u''(x_j) - \frac{h^3}{6}u'''(x_j) + \frac{h^4}{24}u(4)(x_j) - \ldots.$$

Substituting these into the approximation for the second derivative gives

$$\begin{aligned} \frac{1}{h^2}\delta^2 u(x_j) &= \frac{u(x_j + h) - 2u(x_j) + u(x_j - h)}{h^2} \\ &= \frac{d^2u(x_j)}{dx^2} + \frac{h^2}{12}\frac{d^4u(x_j)}{dx^4} + \text{ terms with higher powers of } h. \end{aligned}$$

Hence, the error in approximating the true second derivative at x_j is

$$\frac{1}{h^2}\delta^2 u(x_j) - \frac{d^2u(x_j)}{dx^2} = \frac{h^2}{12}\frac{d^4u(x_j)}{dx^4} + \text{ terms with higher powers of } h.$$

What about the error in approximating the solution $u(x_j)$ by U_j?

Definition: Global error. The global error in $u(x_j)$ at $x = x_j$ is

$$e_j = u(x_j) - U_j.$$

Definition: Convergence. A finite difference scheme converges if

$$\max_{0\leq j\leq N} \mid e_j \mid \rightarrow 0, \qquad \text{as} \qquad N \rightarrow \infty.$$

The experiments on the **handout on Centred Finite Differences for the Reaction-Diffusion Equation** show that the centred scheme does converge for the test problems considered. In general, if we don't know the exact solution, then we can't study the global error. We need a way to the quantify error in our approximation that doesn't require the exact solution.

Definition: Truncation error. The truncation error T_j at x_j is the remainder when the exact solution $u(x_j)$ is substituted into the finite difference scheme.

The centred finite difference scheme applied to the reaction-diffusion equation gives

$$-\frac{U_{j+1} - 2U_j + U_{j-1}}{h^2} + r_j\, U_j = f_j, \qquad j = 1, 2, \ldots N-1,$$

so substituting $u(x_j)$ for the approximate value U_j and finding the remainder (when we subtract the right-hand side from the left-hand side) gives

$$T_j = -\left(\frac{u(x_{j+1}) - 2u(x_j) + u(x_{j-1})}{h^2}\right) + r_j u(x_j) - f_j, \qquad j = 1, 2, \ldots N-1.$$

Now, we also know that $u(x_j)$ satisfies the differential equation so

$$0 = -\frac{d^2u(x_j)}{dx^2} + r_j u(x_j) - f_j, \qquad j = 1, 2, \ldots N-1.$$

Subtracting these gives

$$T_j = \frac{d^2u(x_j)}{dx^2} - \frac{1}{h^2}\left(u(x_{i+1}) - 2u(x_j) + u(x_j - 1)\right).$$

Hence, the truncation error T_j is the error in approximating the second derivative by the centred finite difference. Using the Taylor series calculation above, we have

$$T_j = -\frac{h^2}{12}\frac{d^4u(x_j)}{dx^4} + \text{ terms with higher powers of } h.$$

Definition: Order. The order of a finite difference method is the lowest power of h occurring in T_j.

The centred scheme for the reaction-diffusion equation is second-order.

5.2. The convection-diffusion equation

Consider the following problem. Find $u(x)$ satisfying

$$-\frac{d^2u(x)}{dx^2} + w\,\frac{du(x)}{dx} = f(x), \qquad 0 < x < 1,$$

together with the boundary conditions

$$u(0) = \alpha, \qquad u(1) = \beta.$$

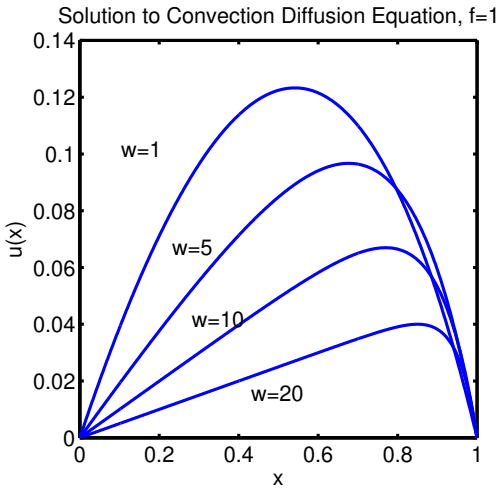


Figure 2: The exact solution to the convection-diffusion equation when $f = 1$ and we have zero boundary conditions.

The coefficient w is called the convection coefficient and we will assume this is a constant. The exact solution in the case of zero boundary conditions is shown in Figure 2.

Notice that there are two derivatives in this equation. We need finite difference approximations for both the second derivative and the first derivative. On Exercise Sheet 7, you are asked to investigate two different schemes. Details of numerical experiments performed in MATLAB are given on the **handout on finite differences for the convection-diffusion problem**.

Centred finite differences

The centred method uses the same centred approximation to the second derivative that we used for the reaction-diffusion equation. That is,

$$\frac{d^2u(x_j)}{dx^2} \approx \frac{1}{h^2}\delta^2 u(x_j), \qquad j = 1, 2, \ldots, N-1.$$

In addition, we use a centred approximation for the first derivative given by

$$\frac{du(x_j)}{dx} \approx \frac{u(x_j + h) - u(x_j - h)}{2h} = \frac{u(x_{j+1}) - u(x_{j-1})}{2h}.$$

Combining these approximations, the resulting finite difference equations are

$$-\left(\frac{U_{j+1} - 2U_j + U_{j-1}}{h^2}\right) + w\left(\frac{U_{j+1} - U_{j-1}}{2h}\right) = f_j, \qquad j = 1, 2, \ldots, N-1,$$

and, using the boundary conditions, we have $U_0 = \alpha$, $U_N = \beta$. Again, we have to solve a system of $N-1$ algebraic equations to find $U_1, \ldots, U_{N-1}$. The numerical experiments performed on the handout reveal that there is a problem with this scheme. There are non-physical oscillations in the approximation. The scheme is **unstable**. The scheme is stable only when

$$\frac{|w|h}{2} \leq 1.$$

(You can test this out yourself by doing the numerical experiments on the handout and varying w and h). This means that when w is large, h has to be very small to get a stable solution. The smaller h is, the more equations we have to solve and the more expensive the method becomes.

Upwind finite differences

Alternatively, we can keep the centred approximation to the second derivative, as before, but use an 'improved' approximation for the first derivative. The so-called upwind approximation is

$$\frac{du(x_j)}{dx} \approx \frac{u(x_j) - u(x_j - h)}{h} = \frac{u(x_j) - u(x_{j-1})}{h}.$$

The resulting finite difference scheme is stable. There are no restrictions no h. However, for the same value of h, the method is less accurate (see Exercise Sheet 7 and the experiments on the **handout on finite differences for the convection-diffusion problem**).

5.3. The one-dimensional heat equation

Consider the following problem. Find $u(x,t)$ satisfying the one-dimensional heat equation,

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < 1, \quad t > 0,$$

together with the boundary conditions

$$u(0,t) = 0, \qquad u(1,t) = 0,$$

and the initial condition

$$u(x,0) = u_0(x).$$

We already know that the exact solution can be expressed as an infinite series, using the method of separation of variables. However, an infinite series is not useful for practical purposes. For instance, if we want to know the value of u at some particular value of x and t, then it is not possible to use the infinite series. (We can never add up infinitely many quantities). We could truncate the series after a few terms and use the truncated series as an approximation. Alternatively, we can use finite difference schemes.

We follow the same steps as for ODEs but to start off, we need a grid in space and time (a space-time grid). The easiest thing is to think about space in the horizontal direction and time t running in the vertical direction. Since $0 < x < 1$, we may use N intervals of length $h = 1/N$ in the x-direction, and label the x-values as

$$x_0 = 0, \quad x_1 = h, \quad \ldots, \quad x_j = jh, \quad x_N = 1.$$

In the t-direction, we'll use intervals of size k, with t-values

$$t_0 = 0, \quad t_1 = k, \quad \ldots, \quad t_m = mh, \quad \ldots.$$

Note that there is no upper limit on the time variable t. We just keep taking steps until we reach the value of t where we want to approximate the solution. We now have two discretisation parameters: h (the spacing in the x-direction) and k (the spacing in the t-direction).

Let U_j^m denote our approximation to $u(x_j, t_m)$. Starting from the known initial values

$$U_j^0 = u(x_j, t_0) = u(x_j, 0) = u_0(x_j), \qquad j = 0, 1, \ldots, N,$$

we compute approximations sequentially at each time step. We briefly look at two finite difference schemes: the explicit method and the implicit method.

Explicit scheme

The heat equation has two derivatives and we need to approximate both of them, at the grid points (x_j, t_m). For the 'explicit' scheme, we replace the time derivative with a 'first forward difference' and the space derivative with a 'second centred difference'. That is

$$\frac{\partial^2 u(x_j,t_m)}{\partial x^2} \approx \frac{u(x_{j+1},t_m) - 2u(x_j,t_m) + u(x_{j-1},t_m)}{h^2}$$

and

$$\frac{\partial u(x_j,t_m)}{\partial t} \approx \frac{u(x_j,t_{m+1}) - u(x_j,t_m)}{k}.$$

Notice that to approximate the partial derivative with respect to t, we keep the value of x constant and only take a difference over distinct t values, and similarly for the partial derivative with respect to x.

If we write down the heat equation at the point (x_j, t_m) and replace the derivatives with the above approximations, and then replace the exact values $u(x_j, t_m)$ with approximate ones, we obtain the finite difference equations

$$\frac{U_j^{m+1} - U_j^m}{k} = \frac{U_{j+1}^m - 2U_j^m + U_{j-1}^m}{h^2}, \qquad j = 1, \ldots, N-1, \quad m = 0, 1, \ldots$$

or, writing $\nu = \frac{k}{h^2}$, we can rearrange to give

$$U_{j+1}^m = U_j^m + \nu\left(U_{j+1}^m - 2U_j^m + U_{j-1}^m\right), \qquad j = 1, \ldots, N-1, \quad m = 0, 1, \ldots,$$

with

$$U_0^m = 0, \quad U_N^m = 0, \qquad m = 0, 1, 2, \ldots,$$

(from the boundary conditions) and

$$U_j^0 = u_0(x_j), \qquad j = 0, 1, \ldots, N,$$

(from the initial condition).

The above scheme is called 'explicit' because new values at time level $m+1$ can be computed from the old values at time level m directly. We start from the initial values U_j^0, $j = 0, 1, \ldots, N$ (which are given) and then compute the values U_j^1 at time $t_1 = k$, then at $t_2 = k$, etc. In matrix notation, at step $m+1$, we have to compute

$$\begin{pmatrix} U_1^{m+1} \\ U_2^{m+1} \\ \vdots \\ \vdots \\ \vdots \\ U_{N-1}^{m+1} \end{pmatrix} = \begin{bmatrix} 1-2\nu & \nu & 0 & 0 & \ldots & 0 \\ \nu & 1-2\nu & \nu & 0 & \ldots & 0 \\ 0 & \nu & 1-2\nu & \nu & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \nu & 1-2\nu & \nu \\ 0 & 0 & 0 & 0 & \nu & 1-2\nu \end{bmatrix} \begin{pmatrix} U_1^m \\ U_2^m \\ \vdots \\ \vdots \\ \vdots \\ U_{N-1}^m \end{pmatrix}, \quad m = 0, 1, 2, \ldots$$

At step $m+1$, we know the vector on the right. To compute the vector on the left, all we have to do is a matrix-vector product. No linear system of equations needs to be solved. More details about this computation, and how to do experiments in MATLAB are given on the **handout on the Explicit Finite Difference Scheme for the Heat Equation**.

Example On the handout, results of two experiments are reported. The initial condition is chosen as

$$u_0(x) = \begin{cases} 2x & 0 \le x \le 1/2 \\ 2-2x & 1/2 \le x \le 1 \end{cases}.$$

The explicit finite difference scheme is applied in two different ways. First, we set $h = 1/20$ (so, choose $N = 20$ intervals in space) and $k = 0.0012$. In this case, we have $\nu = k/h^2 = 0.48$. The finite difference approximation seems to be a good one, at each time step (see the pictures on the handout). Now, if we repeat the experiment but choose $k = 0.0013$ (and keep $h = 1/20$) then the approximation has non-physical oscillations and the approximation is terrible! As time progresses, these oscillations only get worse. In this second case, we have $\nu = k/h^2 = 0.52$.

The message from this experiment is that we need to choose h and k very carefully. These parameters have to be compatible. It can be shown that the explicit scheme is **unstable** whenever $\nu > 1/2$. See Exercise Sheet 8 for a further investigation of this issue. To analyse the explicit method properly, we'll need to investigate the error. Again, we must distinguish between global error and truncation error (which can be analysed, even if we don't know the exact solution).

Definition: Global error. The global error in $u(x_j, t_m)$ at the point (x_j, t_m) is

$$e_j^m = u(x_j, t_m) - U_j^m.$$

Definition: Truncation error. The truncation error T_j^m at a point (x_j, t_m) is the remainder when the exact solution is substituted into the approximation scheme.

So, for the explicit scheme,

$$T_j^m = \frac{u(x_j, t_{m+1}) - u(x_j, t_m)}{k} - \left(\frac{u(x_j, t_m) - 2u(x_j, t_m) + u(x_{j-1}, t_m)}{h^2} \right).$$

If we expand the terms $u(x_j, t_{m+1}), u(x_{j+1}, t_m)$ and $u(x_{j-1}, t_m)$ about (x_j, t_m) using Taylor series (in 2 variables now), we find that:

$$T_j^m = \frac{k}{2}\frac{\partial^2 u(x_j, t_m)}{\partial t^2} - \frac{h^2}{12}\frac{\partial^4 u(x_j, t_m)}{\partial x^4} + \text{ terms with higher powers of } h \text{ and } k.$$

So, the scheme is first order in time and second order in space. See Exercise Sheet 8.

Implicit scheme

The restriction that $\nu < 1/2$ for the explicit scheme means that we need $k \le \frac{1}{2}h^2$. This poses a restriction on the time step and the method becomes expensive for small h. We can improve stability, however, by changing the approximation used for the time derivative.

The 'implicit' finite difference scheme for the heat equation uses the same centred approximation for the space derivative as before

$$\frac{\partial^2 u(x_j, t_m)}{\partial x^2} \approx \frac{u(x_{j+1}, t_m) - 2u(x_j, t_m) + u(x_{j-1}, t_m)}{h^2}$$

and a 'backward' difference approximation for the time derivative

$$\frac{\partial u(x_j, t_m)}{\partial t} \approx \frac{u(x_j, t_m) - u(x_j, t_{m-1})}{k}.$$

If we write down the heat equation at the point (x_j, t_m) and replace the derivatives with the above approximations, and then replace the exact values $u(x_j, t_m)$ with approximate ones, we obtain the finite difference equations

$$\frac{U_j^m - U_j^{m-1}}{k} = \frac{U_{j+1}^m - 2U_j^m + U_{j-1}^m}{h^2}, \qquad j = 1, \ldots, N-1, \quad m = 1, 2, \ldots$$

or, writing $\nu = \frac{k}{h^2}$, and relabelling $m-1$ as m we can rearrange to give

$$U_{j+1}^m = -\nu U_{j-1}^{m+1} + (1+2\nu)\, U_j^{m+1} - \nu U_{j+1}^{m+1}, \qquad j = 1, \ldots, N-1, \quad m = 0, 1, \ldots,$$

with

$$U_0^m = 0, \quad U_N^m = 0, \qquad m = 0, 1, 2, \ldots,$$

(from the boundary conditions) and

$$U_j^0 = u_0(x_j), \qquad j = 0, 1, \ldots, N,$$

(from the initial condition).

This time, we do need to solve a linear system of equations to compute the approximation at the next time level. At step $m+1$ we have to solve:

$$\begin{bmatrix} 1+2\nu & -\nu & 0 & 0 & \ldots & 0 \\ -\nu & 1+2\nu & -\nu & 0 & \ldots & 0 \\ 0 & -\nu & 1+2\nu & -\nu & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & -\nu & 1+2\nu & -\nu \\ 0 & 0 & 0 & 0 & -\nu & 1+2\nu \end{bmatrix} \begin{pmatrix} U_1^{m+1} \\ U_2^{m+1} \\ \vdots \\ \vdots \\ \vdots \\ U_{N-1}^{m+1} \end{pmatrix} = \begin{pmatrix} U_1^m \\ U_2^m \\ \vdots \\ \vdots \\ \vdots \\ U_{N-1}^m \end{pmatrix}, \quad m = 0, 1, 2, \ldots$$

The implicit scheme converges for all values of $\nu > 0$. There is no restriction on how we choose k and h (for stability). On the **handout on the Implicit Finite Difference Scheme for the Heat Equation**, the same experiment that was conducted with the explicit scheme is repeated. No oscillations are now observed when we choose $\nu = 0.52$.

You met vectors in the first year. Vector calculus is essentially calculus on vectors. We will need to differentiate vectors and perform integrals involving vectors. In particular, we will look at two fundamental results called The Divergence Theorem and Stokes Theorem. Now would be an ideal time to revise any notes you have on vectors. Some of the basic facts that you should already know about vectors, and operations on vectors that involve derivatives (divergence $\nabla\cdot$, gradient ∇, curl $\nabla\times$) are summarised on the **handout on Div, Grad and Curl**. Exercise Sheet 9 also contains questions that are intended as revision.

6.1. Introduction

A vector field $\mathbf{F}$ in three dimensions is a rule which tells us how to associate a vector with each point (x, y, z). See the **handout on Div, Grad and Curl**. For example, the velocity of a fluid is a vector field. In general,

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

where F_x, F_y and F_z are functions of x, y, z. The handout shows the following two-dimensional examples (with $F_z = 0$).

$$\mathbf{F}_1 = x\mathbf{i} + y\mathbf{j}, \qquad \mathbf{F}_2 = \frac{-y\mathbf{i} + x\mathbf{j}}{\sqrt{x^2 + y^2}}.$$

The vector field

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

is called the radial direction vector. We will meet it several times. Note that $\mathbf{F}_1$ is the two-dimensional version.

Divergence and curl are two important mathematical operations on vector fields. Recall,

$$\nabla\cdot\mathbf{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}.$$

The divergence gives a measure of 'net mass flow' If $\nabla\cdot\mathbf{F} = 0$ then no mass is created or destroyed. A simple calculation reveals that $\nabla\cdot\mathbf{F}_2 = 0$ (check!). If we focus on a small region in the x-y plane then we can see that mass is conserved by observing that the arrows pointing into that region are matched by arrows of the same length pointing out of that region.

The curl $\nabla\times\mathbf{F}$ gives a measure of twisting or 'curling' of a vector field. A simple calculation reveals that $\nabla\times\mathbf{F}_2 = 0$ (check!). The curl essentially tells us how a particle released into the flow field rotates. Recall, the curl is calculated via

$$\begin{aligned}
\nabla\times\mathbf{F} &= \left(\frac{\partial}{\partial x}\mathbf{i}, \frac{\partial}{\partial y}\mathbf{j}, \frac{\partial}{\partial z}\mathbf{k}\right)\times(F_x\mathbf{i} + F_y\mathbf{j} + F_z\mathbf{k}) \\
&= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ F_x & F_y & F_z \end{vmatrix} \\
&= \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}\right)\mathbf{i} - \left(\frac{\partial F_z}{\partial x} - \frac{\partial F_x}{\partial z}\right)\mathbf{j} + \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}\right)\mathbf{k}.
\end{aligned}$$

Example. Consider $\mathbf{F} = x^2y\mathbf{i} + xz\mathbf{j} + xyz\mathbf{k}$ then

$$\nabla \times \mathbf{F} = (xz - x)\mathbf{i} - yz\mathbf{j} + (z - x^2)\mathbf{k}.$$

There are two fundamental identities involving the operators div, grad and curl. For all scalar functions $f = f(x, y, z)$,

$$\nabla \times (\nabla f) = 0.$$

The curl of a gradient is always zero. In addition, for all vector fields $\mathbf{F}$,

$$\nabla \cdot (\nabla \times \mathbf{F}) = 0.$$

The divergence of the curl of a vector is always zero. You are asked to prove these identities on Exercise Sheet 9.

6.2. Volume integrals (of scalar functions)

Recall, the double integral

$$\int\int_D f(x, y)dA,$$

where D is a two-dimensional region in the x-y plane and $dA = dxdy$ (the area element in Cartesian coordinates) represents the volume between the surface $z = f(x, y)$ and the region D.

If D is a rectangle, then we have constant limits of integration. We can then easily swap the order of integration without worrying about the limits. For example, if $D = [0, 1] \times [0, 2]$ and $f(x, y) = x^2 + y^2$ then

$$\int\int_D f(x, y)\, dA = \int_{x=0}^{1}\int_{y=0}^{2} x^2 + y^2\, dydx = \int_{y=0}^{2}\int_{x=0}^{1} x^2 + y^2\, dxdy.$$

(You have been doing this since A-level). If D is a region with a more complicated shape, then we usually don't have constant limits of integration. Limits for the inner integral must be expressed as functions of the outer variables.

Example Compute the integral

$$\int\int_D 1 - x - y\, dA,$$

where D is the right-angled triangle with vertices $(0, 0)$, $(1, 0)$ and $(0, 1)$. If we choose to perform the y integral first, then

$$\int\int_D 1 - x - y\, dA = \int_{x=0}^{1}\int_{y=0}^{1-x} 1 - x - y\, dydx.$$

Alternatively, if we perform the x integral first, then

$$\int\int_D 1 - x - y\, dA = \int_{y=0}^{1}\int_{x=0}^{1-y} 1 - x - y\, dxdy.$$

In both cases, you should find that the answer is $1/6$. The important point is that we cannot just swap the order of integration and keep the same limits.

Now, a volume integral (of a scalar function f) over a three-dimensional volume V is denoted

$$\int\int\int_V f\, dV,$$

where dV is the so-called volume element. We use three integral signs here to emphasis that the integral is over a three-dimensional volume, but it is also ok to only write one integral sign. In Cartesian coordinates, you already know that the volume element is

$$dV = dxdydz.$$

Remember that an integral is really defined as a limit. We replace the integral with a sum over pieces of V. On each piece we evaluate the function f and multiply by the volume of the piece. We then take the limit of the sum as the number of pieces tends to infinity. In Cartesian coordinates, a natural way to break up a volume V is into small bricks. If the brick has lengths dx, dy and dz in each of the x, y and z coordinate directions then the volume of the piece is $dxdydz$. This is the 'volume element'.

If the volume V is a simple brick, then we have constant limits of integration. So, if $V = [a, b] \times [c, d] \times [e, f]$ then

$$\int\int\int_V f(x, y, z)\, dV = \int_{z=e}^{f}\int_{y=c}^{d}\int_{x=a}^{b} f(x, y, z)\, dxdydz.$$

In this case, we can swap the order of integration easily and we don't need to worry about changing the limits. As with double integrals, if V is more complicated then we need to pay attention to the limits of integration. Drawing a picture of V usually helps determine the correct limits!

Example Evaluate

$$\int\int\int_V 1 + xy\, dV$$

where V is the tetrahedron with vertices $(0, 0, 0)$, $(0, 0, 1)$, $(0, 1, 0)$ and $(1, 0, 0)$. Lets perform the z integral first, followed by the y integral and finally the x integral. If we perform the z integral first, then we need to provide limits for z as a function of the outer variables y and x. For a fixed x and a fixed y, the z coordinate is bounded by the face of the tetrahedron in the x-y plane (where $z = 0$) and the face of the tetrahedron that coincides with the plane $z = 1 - x - y$. Hence,

$$\int\int\int_V 1 + xy\, dV = \int_x\int_y \left(\int_{z=0}^{1-x-y} 1 + xy\, dz\right) dy\, dx.$$

The limits for y should then be expressed as functions of the outer variable x. We have

$$\int\int\int_V 1 + xy\, dV = \int_x\int_{y=0}^{1-x}\int_{z=0}^{1-x-y} 1 + xy\, dz\, dy\, dx.$$

Finally, the variable x varies from 0 to 1 so

$$\int\int\int_V 1 + xy\, dV = \int_{x=0}^{1}\int_{y=0}^{1-x}\int_{z=0}^{1-x-y} 1 + xy\, dz\, dy\, dx = 7/40.$$

If V is not a brick but a more complicated shape that is not easy to describe in the Cartesian co-ordinate system, then it may be easier to work in an alternative coordinate system. If we do this, however, care must be taken to convert the volume element dV in the proper way. For instance, in cylindrical coordinates, dV does **not** mean $drd\theta dz$.

You already learned how to do double integrals in polar coordinates (revise your first year notes if you have forgotten). Recall,

$$\int\int_A f\, dA = \int_x\int_y f(x, y)\, dx\, dy = \int_r\int_\theta f(x(r, \theta), y(r, \theta))\, r\, dr\, d\theta.$$

The area element here in polar coordinates is $dA = rdrd\theta$. In Cartesian coordinates, we have $dA = dxdy$. Where does the extra factor of r come from? Recall that r is the determinant of the Jacobian matrix

$$\begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix}.$$

It accounts for the change in area when we map a small rectangle with area $dxdy$ into polar co-ordinates. The mapped rectangle is not a rectangle.

In three dimensions, we know that the volume element for **Cartesian coordinates** is

$$dV = dxdydz.$$

In **cylindrical coordinates** we have

$$dV = rdrd\theta dz$$

and in **spherical coordinates** we have

$$dV = \rho^2 \sin\phi \, d\rho d\phi \, d\theta.$$

You should learn these formulae. To work out the last two explicitly, you can write down the 3×3 Jacobian matrix and find its determinant, exactly as you did for polar co-ordinates.

Let's test out the claim that $dV = \rho^2 \sin\phi \, d\rho d\phi \, d\theta$ for spherical co-ordinates. Note that

$$\int\int\int_V 1 dV$$

gives the volume of V. If we get the volume element dV and the limits of integration right, then integrating one over a sphere should give the volume of the sphere (which you already know how to compute).

Example Find the volume of a sphere of radius two. Integrating one over V where V is the sphere centred at the origin with radius two gives

$$\begin{aligned}
\int\int\int_V 1 \, dV &= \int\int\int_V 1 \, \rho^2 \sin\phi \, d\rho \, d\theta \, d\phi \\
&= \int_{\phi=0}^{\pi} \int_{\theta=0}^{2\pi} \int_{\rho=0}^{2} \rho^2 \sin\phi \, d\rho \, d\theta \, d\phi \\
&= \int_{\phi=0}^{\pi} \int_{\theta=0}^{2\pi} \left(\frac{8 \sin\phi}{3} \right) d\theta \, d\phi \\
&= \int_{\phi=0}^{\pi} \frac{16\pi \sin\phi}{3} \, d\phi \\
&= -\frac{16\pi \cos(\pi)}{3} + \frac{16\pi \cos(0)}{3} \\
&= \frac{16\pi}{3} + \frac{16\pi}{3} = \frac{32\pi}{3}.
\end{aligned}$$

Now, of course, we also know that the standard formula for the volume of a sphere is $\frac{4}{3}\pi\rho^3$ where ρ is the radius. Since the radius is two, the volume is $\frac{4}{3} \times \pi \times 8 = \frac{32}{3}\pi$, which matches the above calculation.

There is an important theorem that connects: the divergence $\nabla \cdot \mathbf{F}$ of a vector field, a closed volume V, and its surface S. The theorem is quite technical. We present it first, and then investigate it.

Theorem: The Divergence Theorem. Let V be a bounded, closed region in space with piecewise smooth boundary S. Let $\hat{\mathbf{n}}$ be the unit normal vector to S, pointing **outward**. Then, if $\mathbf{F}$ is a differentiable vector field,

$$\int\int\int_V \nabla \cdot \mathbf{F}\, dV = \int\int_S \mathbf{F} \cdot \hat{\mathbf{n}}\, dS.$$

In essence, this result says that the total divergence of a vector field in a bounded region V in space is equal to the net flow (or 'flux') across the boundary of the surface in the normal direction. Note that $\nabla \cdot \mathbf{F}$ is a scalar function so the integral on the left-hand side is a standard volume integral (which you know how to evaluate). Note that $\mathbf{F} \cdot \hat{\mathbf{n}}$ is the dot product of two vectors and this is also a scalar function. So, to evaluate the right-hand side of the equation, we first need to know how to find the normal vector to the surface S, and then how to evaluate surface integrals of scalar functions.

Unit normal vectors to surfaces

For some surfaces, it is easy to determine the unit normal vector.

Example Consider the unit cube. That is, the cube whose edges all have length one, and one of the vertices is the origin $(0, 0, 0)$. The cube has six faces, which are portions of the surfaces $z = 0$, $z = 1$ (the top and bottom faces), $x = 0$, $x = 1$, $y = 0$ and $y = 1$ (the side faces). The cube is a close volume. The unit vector that points out of the cube at each of the six faces is aligned with one of the x, y and z coordinate axes. On the top surface, the vector must point straight up (and have length one), so $\hat{\mathbf{n}} = \mathbf{k}$. On the bottom surface, the vector must point straight down (and have length one), so $\hat{\mathbf{n}} = -\mathbf{k}$. Similarly, on the four side faces, we have $\hat{\mathbf{n}} = -\mathbf{i}$, $\hat{\mathbf{n}} = \mathbf{i}$, $\hat{\mathbf{n}} = -\mathbf{j}$, and $\hat{\mathbf{n}} = \mathbf{j}$.

For curved surfaces it is more complicated to determine $\hat{\mathbf{n}}$. For surfaces of the form $z = f(x, y)$, we have the following general formula.

Let S be a portion of a surface of the form $z = f(x, y)$. A unit normal vector to S is

$$\hat{\mathbf{n}} = \frac{-\frac{\partial f}{\partial x}\mathbf{i} - \frac{\partial f}{\partial y}\mathbf{j} + \mathbf{k}}{\sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}}.$$

Note that this vector is simply 'a' normal vector to the surface. If we reverse the sign, then it is still pointing in a direction normal to the surface. For closed surfaces, if we want the outward pointing normal, we must pay attention to the sign.

Example Let S be the surface of the unit sphere. On S, we have $x^2 + y^2 + z^2 = 1$. So, the surface can be expressed as

$$z = \pm\sqrt{1 - x^2 - y^2} = f(x, y).$$

The positive square root corresponds to points on the upper hemisphere and the negative square root corresponds to the lower hemisphere. Differentiating gives

$$\frac{\partial f}{\partial x} = -\frac{x}{z}, \quad \frac{\partial f}{\partial y} = -\frac{y}{z},$$

(for both signs) and so

$$\sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2} = \sqrt{1 + \frac{x^2}{z^2} + \frac{y^2}{z^2}} = \frac{1}{z}\sqrt{z^2 + x^2 + y^2}.$$

Using the above formula, we have

$$\hat{\mathbf{n}} = \frac{x\mathbf{i} + y\mathbf{j} + z\mathbf{k}}{\sqrt{z^2 + x^2 + y^2}}.$$

We know that on S, $x^2 + y^2 + z^2 = 1$. Hence, a unit normal vector to S is

$$\hat{\mathbf{n}} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}.$$

This is, of course, the radial position vector. It points in the outward normal direction at all points on the surface of the sphere.

6.3. Surface integrals

Let S be a portion of a surface $z = f(x, y)$. Note that here S could be open or closed. The integral

$$\int\int_S 1\, dS$$

is the surface area of S and

$$\int\int_S G(x, y, z)\, dS$$

is the surface integral of G. If the surface is of the form $z = f(x, y)$ then we can convert the surface integral into a standard double integral over a flat region D in the x-y plane by performing a change of variable. To do this, we first need to relate a small area δS on the (possibly curved) surface S to its projection δD (or 'shadow') onto the x-y plane.

It can be shown that the relationship between δS and δD is

$$\delta S = \frac{\delta D}{\hat{\mathbf{n}} \cdot \mathbf{k}}$$

where $\hat{\mathbf{n}}$ is the unit normal vector to S on δS (and obviously $\mathbf{k}$ is the unit normal vector to δD, pointing up). So,

$$\int\int_S G(x, y, z)\, dS = \int\int_S G(x, y, z(x, y)) \frac{1}{\hat{\mathbf{n}} \cdot \mathbf{k}}\, dxdy.$$

Using the formula for $\hat{\mathbf{n}}$, we have

$$\hat{\mathbf{n}} \cdot \mathbf{k} = \frac{1}{\sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}}$$

and so,

$$\int\int_S G(x, y, z)\, dS = \int\int_S G(x, y, z(x, y)) \sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}\, dxdy.$$

Basically, when converting to a standard integral in the x-y plane, we just need to remember the square root factor to account for the change in curvature. Note that if S is itself flat, and parallel to the x-y plane, then $\hat{\mathbf{n}} = \mathbf{k}$ and $\hat{\mathbf{n}} \cdot \mathbf{k} = \mathbf{k} \cdot \mathbf{k} = 1$, so there is no extra factor.

Example Evaluate the surface integral $\int\int_S z^2\, dS$, where S is the open surface corresponding the portion of the surface of the unit sphere that lies in the first octant ($x \geq 0$, $y \geq 0$, $z \geq 0$). **[Note - it would help to draw a diagram of S here].**

S is a portion of the surface

$$z = +\sqrt{1 - x^2 - y^2} = f(x, y).$$

We have already seen that

$$\frac{\partial f}{\partial x} = -\frac{x}{z}, \quad \frac{\partial f}{\partial y} = -\frac{y}{z}$$

and so

$$\sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2} = \sqrt{1 + \frac{x^2}{z^2} + \frac{y^2}{z^2}} = \frac{1}{z}\sqrt{z^2 + x^2 + y^2} = \frac{1}{z}$$

on S. Hence,

$$\int\int_S z^2 \, dS = \int\int_D z^2 \frac{1}{z} \, dx \, dy = \int\int_D z \, dx \, dy = \int\int_D \sqrt{1 - x^2 - y^2} \, dx \, dy.$$

This is now just a standard double integral but we need to work out the limits of integration. D is the projection of S onto the x-y plane. This is a quarter circle bounded by the lines $x = 0$, $y = 0$ and $x^2 + y^2 = 1$. It would be easier to evaluate this integral in polar co-ordinates. So, changing variables once again,

$$\int\int_S z^2 \, dS = \int_{\theta=0}^{\pi/2} \int_{r=0}^{1} \sqrt{1 - r^2} \, r dr d\theta.$$

(Don't forget the extra factor of r in the area element when we change variables). The integral with respect to r can be done by substitution or by just identifying a function whose derivative is the given integrand

$$\int\int_S z^2 \, dS = \int_{\theta=0}^{\pi/2} \left[-\frac{\left(1 - r^2\right)^{3/2}}{3} \right]_0^1 d\theta = \int_{\theta=0}^{\pi/2} \frac{1}{3} \, d\theta = \frac{1}{3} \times \frac{\pi}{2} = \frac{\pi}{6}.$$

You will find more surface integral examples on Exercise Sheet 10. Note that we can perform integrals over surfaces that are open (as in the above example) or closed (for example, the surface of a whole sphere). In the Divergence Theorem, however, we have a surface integral of a function of the form

$$G(x, y, z) = \mathbf{F} \cdot \hat{\mathbf{n}},$$

where the surface S is always **closed**.

6.4. The Divergence Theorem

We now have all the ingredients we need to evaluate both integrals in the Divergence Theorem.

Example Show that the Divergence Theorem

$$\int\int\int_V \nabla \cdot \mathbf{F} \, dV = \int\int_S \mathbf{F} \cdot \hat{\mathbf{n}} \, dS$$

is satisfied for the vector field $\mathbf{F} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ when V is the unit sphere.

First, we evaluate the left-hand side. Clearly, $\nabla \cdot \mathbf{F} = 3$. Hence,

$$\int\int\int_V \nabla \cdot \mathbf{F} dV = 3 \int\int\int_V 1 \, dS,$$

which is three times the volume of the sphere. This is $3\,(4\pi/3) = 4\pi$.

For the integral on the right, we know that $\hat{\mathbf{n}} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ (the radial direction vector). So,

$$\mathbf{F} \cdot \hat{\mathbf{n}} = x^2 + y^2 + z^2 = 1$$

on S. Hence,

$$\int\int_S \mathbf{F} \cdot \hat{\mathbf{n}}\, dS = \int\int_S 1\, dS,$$

which is just the surface area of the sphere. Using the standard formula for surface area, this is $4\pi(1)^2 = 4\pi$. Both integrals match.

Example Verify that the Divergence Theorem holds when

$$\mathbf{F} = (y - x)\mathbf{i} + (y - z)\mathbf{j} + (x - y)\mathbf{k},$$

and V is the unit cube.

First we note that $\nabla \cdot \mathbf{F} = -1 + 1 + 0 = 0$. Hence $\int\int\int_V \nabla \cdot \mathbf{F}\, dV = 0$. It is not easy to describe the surface S of the unit cube, or the normal vector to S, with a single equation. In this case, it is best to split S into 6 distinct parts (the 6 faces of the cube). That is,

$$\int\int_S \mathbf{F} \cdot \hat{\mathbf{n}}\, dS = \sum_{i=1}^{6} \int\int_{S_i} \mathbf{F} \cdot \hat{\mathbf{n}}_i\, dS_i,$$

where S_i is the ith face, and $\hat{\mathbf{n}}_i$ is the unit normal vector to that face (pointing out of the cube). We evaluate these surface integrals one at a time. For example, suppose that S_1 is the top face. Then, $z = 1$ on S_1 and $\hat{\mathbf{n}}_1 = \mathbf{k}$ so $\mathbf{F} \cdot \hat{\mathbf{n}}_1 = x - y$. Now, since S_1 is flat,

$$\int\int_{S_1} \mathbf{F} \cdot \hat{\mathbf{n}}_1\, dS_1 = \int_{y=0}^{1}\int_{x=0}^{1} x - y\, dxdy = 0.$$

Similarly for the other faces .... See Exercise Sheet 10.

6.5. Line integrals and Stokes Theorem

You studied line integrals (for lines in the x-y plane) in first year calculus. If you have forgotten, please revise your notes now. See also the **handout: Revision on line integrals in the plane**.

Consider the integral of a scalar function f along a line segment c in three space dimensions, where c starts at point a and finishes at point b. We write

$$\int_c f(x, y, z)\, ds.$$

Here, s is the 'arc length parameter'. We can use this coordinate to parametrise any line. At point a, $s = 0$ and at point b, the value of s is the length of the line segment joining a and b. If the coordinates x, y and z of any point on the line can be expressed easily in terms of s, then we just have a standard one dimensional integral

$$\int_{s=0}^{\text{length of line}} f(x(s), y(s), z(s))\, ds.$$

Now, let $\mathbf{F}$ be a vector field and consider

$$\int_c \mathbf{F} \cdot \hat{\mathbf{t}}\, ds,$$

where $\hat{\mathbf{t}}$ is the unit tangent vector to the curve c. This integral gives the work done by the vector field to move an object along c. If we know how to find $\hat{\mathbf{t}}$ then, given c, this is just a standard line integral, as above.

Draw any line segment (part of a curve) and mark two points, with arc length parameter s and $s + \delta s$. Let (x, y, z) be the standard Cartesian coordinates of the first point, and $(x + \delta x, y + \delta y, z + \delta z)$ be the coordinates of the second. Now draw a **straight** line joining these two points. This straight line has the vector equation

$$\delta x\mathbf{i} + \delta y\mathbf{j} + \delta z\mathbf{k} = (x(s + \delta s) - x(s))\,\mathbf{i} + (y(s + \delta s) - y(s))\,\mathbf{j} + (z(s + \delta s) - z(s))\,\mathbf{k}.$$

If we divide by δs and take the limit as $\delta s \to 0$ then we obtain the vector

$$\frac{dx}{ds}\mathbf{i} + \frac{dy}{ds}\mathbf{j} + \frac{dz}{ds}\mathbf{k}.$$

This is, by definition, the unit tangent vector to the curve at (x, y, z). Hence

$$\hat{\mathbf{t}} = \frac{dx}{ds}\mathbf{i} + \frac{dy}{ds}\mathbf{j} + \frac{dz}{ds}\mathbf{k}.$$

Using this expression for $\hat{\mathbf{t}}$, we can also write

$$\int_c \mathbf{F} \cdot \hat{\mathbf{t}}\, ds = \int_c \mathbf{F} \cdot \left(\frac{dx}{ds}\mathbf{i} + \frac{dy}{ds}\mathbf{j} + \frac{dz}{ds}\mathbf{k} \right) ds = \int_c \mathbf{F} \cdot \mathbf{dr},$$

where

$$\mathbf{dr} = dx\mathbf{i} + dy\mathbf{j} + dz\mathbf{k}.$$

Some textbooks use the notation $\int_c \mathbf{F} \cdot \mathbf{dr}$ but we will stick to $\int_c \mathbf{F} \cdot \hat{\mathbf{t}}\, ds$. Recall also that when c is a closed loop we write $\oint_c$ and not $\int_c$.

Example Let $\mathbf{F} = y\mathbf{i} - x\mathbf{j}$ and let c denote the closed curve shown in Figure 1. (Note that this curve lies in the x-y plane). Evaluate

$$\oint_c \mathbf{F} \cdot \hat{\mathbf{t}}\, ds.$$

Using the expression for $\hat{\mathbf{t}}$ (with $z = 0$) gives

$$\oint_c \mathbf{F} \cdot \hat{\mathbf{t}}\, ds = \oint_c (y\mathbf{i} - x\mathbf{j}) \cdot \left(\frac{dx}{ds}\mathbf{i} + \frac{dy}{ds}\mathbf{j} \right) ds = \oint_c \left(y\frac{dx}{ds} - x\frac{dy}{ds} \right) ds.$$

We break the curve up into three parts, and on each straight line segment, express x and y as functions of s. That is,

$$\oint_c \left(y\frac{dx}{ds} - x\frac{dy}{ds} \right) ds = \sum_{i=1}^{3} \oint_{c_i} \left(y(s)\frac{dx(s)}{ds} - x(s)\frac{dy(s)}{ds} \right) ds.$$

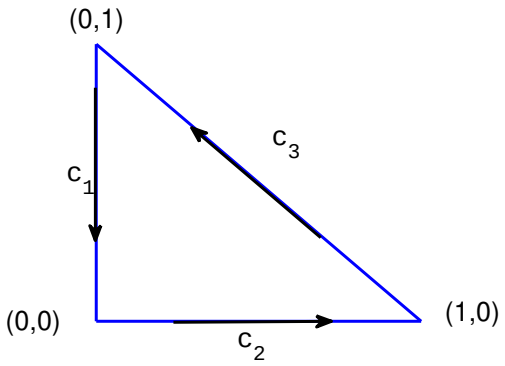


Figure 1: A closed loop corresponding to the boundary of a triangle. The arrows give the orientation of the path taken.

On c_1 (the line joining $(0,1)$ to $(0,0)$), we have $x = 0$ and $y = 1 - s$. (Check: the length of the line is 1. When $y = 1$, $s = 0$ and when $y = 0$, $s = 1$). So,

$$\oint_{c_1} \left(y(s)\frac{dx(s)}{ds} - x(s)\frac{dy(s)}{ds} \right) ds = \int_{s=0}^{1} ((1-s)0 - 0)\, ds = 0.$$

On c_2, we have $y = 0$ and $x = s$, so

$$\oint_{c_2} \left(y(s)\frac{dx(s)}{ds} - x(s)\frac{dy(s)}{ds} \right) ds = \int_{s=0}^{1} (0 - s0)\, ds = 0.$$

Finally, on c_3, we have $x = 1 - s/\sqrt{2}$, and $y = s/\sqrt{2}$ so

$$\oint_{c_3} \left(y(s)\frac{dx(s)}{ds} - x(s)\frac{dy(s)}{ds} \right) ds = \int_{s=0}^{1} \left(s/\sqrt{2}\right)\left(-1/\sqrt{2}\right) - \left(1 - s/\sqrt{2}\right)\left(1/\sqrt{2}\right) ds = -1.$$

Now, we have an important theorem that connects: the curl of a vector field, an open surface, and the closed curve c that 'spans' the surface. For a flat surface S in two dimensions, the closed curve c is just the boundary of S (as in the triangle example above). In three dimensions, an open surface may have curvature. Consider, for example, the upper half of the surface of the unit sphere. This surface is spanned by the closed loop corresponding to the set of points $x^2 + y^2 = 1$.

Theorem: Stokes Theorem. Let c be a closed curve and let S be an open surface spanned by c. Let $\hat{\mathbf{n}}$ be the unit normal vector to S (oriented with respect to the right-hand rule). Then, if $\mathbf{F}$ is a differentiable vector field,

$$\oint_c \mathbf{F} \cdot \hat{\mathbf{t}}\, ds = \int\int_S (\nabla \times \mathbf{F}) \cdot \hat{\mathbf{n}}\, dS.$$

Note that it is important here that the vectors $\hat{\mathbf{n}}$ and $\hat{\mathbf{t}}$ are oriented correctly with respect to one another. In words, the theorem says that the integral of $\mathbf{F} \cdot \hat{\mathbf{t}}$ around a closed loop c is equal to the integral of the normal component of the curl of $\mathbf{F}$ across the surface S spanned by c.

Example Let us return to the above example with $\mathbf{F} = y\mathbf{i} - x\mathbf{j}$, where S is a right-angled triangle and c is the boundary. We have already computed

$$\oint_c \mathbf{F} \cdot \hat{\mathbf{t}}\, ds,$$

where the tangent vector points in the direction shown in Figure 1. Let $\hat{\mathbf{n}}$ be the unit normal vector to S pointing up (in the positive z direction). That is, $\hat{\mathbf{n}} = \mathbf{k}$. Then, since

$$\nabla \times \mathbf{F} = -2\mathbf{k},$$

we have

$$\int\int_S (\nabla \times \mathbf{F}) \cdot \hat{\mathbf{n}}\, dS = -2 \times \int\int_S 1\, dS,$$

which is minus two times the area of the triangle. Hence,

$$\int\int_S (\nabla \times \mathbf{F}) \cdot \hat{\mathbf{n}}\, dS = -2 \times \frac{1}{2} = -1.$$

This integral matches the line integral computed above. We see that Stokes Theorem is indeed satisfied.

When S is a flat surface, lying in the x-y plane, as in the above example, the computation is relatively straight forward. Exercise Sheet 10 has more challenging examples with curved surfaces. In the next example S is still a triangle but is not lying in the x-y plane.

Example Verify Stokes Theorem when $\mathbf{F} = \left(y + y^2\right)\mathbf{k}$ and c is the boundary of the triangle S with vertices $(0,0,1)$, $(1,0,0)$, $(0,1,0)$.

First we compute

$$\nabla \times \mathbf{F} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ 0 & 0 & y + y^2 \end{vmatrix} = (1 + 2y)\mathbf{i}.$$

The surface here is a portion of the plane $x + y + z = 1$ or equivalently, a portion of the surface $z = f(x, y)$ where $f(x, y) = 1 - x - y$. Using the general formula for finding a unit vector that is normal to this surface gives

$$\hat{\mathbf{n}} = \frac{-(-1)\mathbf{i} - (-1)\mathbf{j} + \mathbf{k}}{\sqrt{3}} = \frac{1}{\sqrt{3}}\left(\mathbf{i} + \mathbf{j} + \mathbf{k}\right).$$

The integral on the left-hand side of the theorem is

$$\int\int_S (\nabla \times \mathbf{F}) \cdot \hat{\mathbf{n}}\, dS = \frac{1}{\sqrt{3}} \int\int_S 1 + 2y\, dS.$$

We can convert this into a standard double integral in the x-y plane as follows.

$$\frac{1}{\sqrt{3}} \int\int_S 1 + 2y\, dS = \frac{1}{\sqrt{3}} \int_y \int_x 1 + 2y\, \sqrt{3}\, dxdy.$$

We need the correct limits for x and y. The projection of S onto the x-y plane is a triangle bounded by the lines $x = 0$, $y = 0$ and $x + y = 1$ so

$$\frac{1}{\sqrt{3}} \int\int_S 1 + 2y\, dS = \int_{y=0}^{1} \int_{x=0}^{1-y} 1 + 2y\ dxdy = \frac{5}{6}.$$

Next, we compute the line integral on the right-hand side. Note that the normal vector chosen points in the upward direction. The tangent vector $\hat{\mathbf{t}}$ to the boundary of the triangle should be oriented accordingly. (We must travel in an anti-clockwise direction). First, note that

$$\mathbf{F} \cdot \hat{\mathbf{t}} = \left(y + y^2\right) \mathbf{k} \cdot \left(\frac{dx}{ds}\mathbf{i} + \frac{dy}{ds}\mathbf{j} + \frac{dz}{ds}\mathbf{k} \right) = \left(y + y^2\right) \frac{dz}{ds}$$

and so

$$\oint_c \mathbf{F} \cdot \hat{\mathbf{t}}\, ds = \oint_c \left(y + y^2\right) \frac{dz}{ds}\, ds.$$

We break the boundary up into three parts, making sure to travel in the right direction. Let c_1 denote the line segment from $(0,0,1)$ to $(1,0,0)$, let c_2 denote the line segment from and $(1,0,0)$ to $(0,1,0)$ and let c_3 denote the line segment from and $(0,1,0)$ to $(0,0,1)$.

On c_1, $s = 0$ at $(0,0,1)$ and $s = \sqrt{2}$ at $(1,0,0)$. Along this line, $y = 0$ so

$$\oint_{c_1} \left(y + y^2\right) \frac{dz}{ds}\, ds = \oint_{s=0}^{\sqrt{2}} 0 \, \frac{dz(s)}{ds}\, ds = 0.$$

On c_2, $s = 0$ at $(1,0,0)$ and $s = \sqrt{2}$ at $(0,1,0)$. Along this line, $z = 0$ so

$$\oint_{c_1} \left(y + y^2\right) \frac{dz}{ds}\, ds = \oint_{s=0}^{\sqrt{2}} (y + y^2)\, 0 \, ds = 0.$$

Finally, on c_3, $s = 0$ at $(0,1,0)$ and $s = \sqrt{2}$ at $(0,0,1)$. Along this line, $y = 1 - s/\sqrt{2}$ and $z = s/\sqrt{2}$ so

$$\begin{aligned} \oint_{c_3} \left(y + y^2\right) \frac{dz}{ds}\, ds &= \oint_{s=0}^{\sqrt{2}} \left[\left(1 - s/\sqrt{2}\right) + \left(1 - s/\sqrt{2}\right)^2 \right] \left(1/\sqrt{2}\right)\, ds \\ &= \oint_{s=0}^{\sqrt{2}} 2 - \frac{3s}{\sqrt{2}} + \frac{s^2}{2}\, ds = \frac{5}{6}. \end{aligned}$$

We see that both integrals match.