28 April 2020

Equations That Have Changed the World

Professor Chris Budd

Introduction

Mathematics is full of equations. An equation condenses a huge amount of information into a single statement and solving the equation is like reading a great mystery novel to find who-dunnit at the end. It is the job of the mathematician both to formulate equations and to unpack them to discover the information that they contain. An often-asked question is whether mathematics is an invention or a discovery. In a very real sense, it is both. Formulating an equation is a process of invention and then solving it is a process of discovery. As an example, finding out the properties of the solutions of the equation

$$u_t = \Delta u + \lambda f(u)$$

has kept me busy in a process of discovery for most of my life as a professional mathematician. (For the record this is the equation which describes how things catch light and burn.)

Some equations are more famous than others. Perhaps the most famous of all is Pythagoras' theorem

$$a^2 + b^2 = c^2$$

which relates the sides a,b and c of a right angled triangle, and which has been known to be true for over 3000 years. The closely related equation, and also very famous equation,

$$a^n + b^n = c^n$$

is the basis of Fermat's Last Theorem. It has only been shown recently by Wiles and Taylor that this equation has no solutions with a,b and c integers, if n is greater than 2.

In a poll of the all time greatest equations nearly every mathematician would cite Euler's identity

$$e^{i\pi} + 1 = 0$$

which is in turn a special case of his incredibly useful and general formula

$$e^{i\theta} = \cos(\theta) + i\sin(\theta).$$

Equations can also describe the universe. An excellent example is Newton's law of gravitation, published in 1690.

$$\mathbf{F}_{12} = \frac{G\ m_1\ m_2\ (\mathbf{x}_1 - \mathbf{x}_2)}{\|\mathbf{x}_1 - \mathbf{x}_2\|^3}.$$

This equation describes the attractive force between two massive bodies, gives and excellent description of the whole of the Solar system. On an even larger scale, the Einstein Field Equations describe the evolution of the whole of the universe.

I have decided to describe in this talk not to talk about the equations above, which are covered in many other books, but about own my *five* favourite equations. Like all lists of five favourite things, whether it is music, art or films, this is a necessarily subjective list. Others have tried similar exercises. Ian Stewart, himself a very distinguished former Gresham Professor, wrote the book [1] of his 17 best equations. I am pleased to see some overlap with my list, but there are also some significant differences. The equations that made it onto my list are those which I meet every day in my work, and which, on a computer, are the equations which many others become under discretisation. They are also equations which take us on a tour of the whole of mathematics. This distinguishes them from those on another excellent list in the book by Graeme Farmelo [2], which celebrates the great equations of physics. Again, I will argue that many of these great equations are fundamentally identical from a mathematical perspective. The great advantage of this approach is that by working out how to solve one equation, we can then how to solve many others.

So, my list is those five equations which, in my opinion have changed the world both by their practical utility and also how they have changed mathematics in the search for how to solve them.

Equation one: The linear equation

$$A\,x = b$$

This simple looking equation has arguably changed the world in more ways than any other equation ever written down. It may not be as glamorous as Schrodingers equation, or the Einstein field equations, or as hard to solve as Fermat's last theorem proved to be, but its impact on our lives in everything that we do is almost limitless. In this equation x is an unknown, b is known, and A is a (known) operator which maps x to b. We often call b the *right hand side* of this equation. This equation links a known to an unknown. So, by solving it we are able to find out things that we never knew before. This is the true essence of the way that maths helps us to unravel the secrets of the universe.

In this equation the operator A can be many things, a number, a square matrix, a rectangular matrix, or even a differential operation such as

$$\frac{d^2x}{dt^2} \quad \text{or even} \quad \frac{\partial^2 x}{\partial y^2} + \frac{\partial^2 x}{\partial z^2}$$

The linear equation arises when we try to predict the weather, do an MRI or CAT scan in medical imaging, design an aircraft, train a machine to learn, control a chemical engineering plant or even go shopping. Basically, pretty well everything. However, on a computer all of the various operators and applications lead to the same equation to solve, in which A is a rectangular matrix operator.

The simplest example of this equation was known to the Greeks and occurs when A, x and b are simply numbers.

It follows then that

$$x = b/A.$$

This equation has a unique solution provided that the number A is not equal to zero. If b and A are integers then x is a rational number. This distinguishes it from numbers such as the square root of 2, which cannot be expressed in this form.

A key feature of this equation is that it is *linear*.

This means that if x is the solution when the right hand side is b, then lambda x is the solution if the right hand side is lambda b. Also if x_1 is the solution for a right hand side b_1, and x_2 the solution when the right hand side is b_2, then x_1 + x_2 is the solution when the right hand side is b_1 + b_2.

The most interesting case occurs when A is a *matrix* and x and b are vectors. As an example, we might want to go shopping. Let's suppose that we want to buy x apples and y bananas. Apples cost £2 each and bananas £3 and we have a budget of £17. At the same time, we need to make up our vitamin C content, and each apple has 4 units of vitamin C, and each banana has 3 units. The total vitamin C content to reach is then 25 units. The question is then, what values of x and y should we take?

The equations that we need to satisfy are then:

$$2x + 3y = 17 \qquad \text{and} \qquad 4x + 3y = 25$$

This system can be expressed in the form A x = b as a matrix equation which looks like:

$$\begin{pmatrix} 2 & 3 \\ 4 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 17 \\ 25 \end{pmatrix}.$$

This is much harder to solve than the simpler equation earlier. In fact to solve it we need to construct the inverse matrix A^{-1} so that

$$x = A^{-1} b$$

For our problem this looks like

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -3/6 & 3/6 \\ 4/6 & -2/6 \end{pmatrix} \begin{pmatrix} 17 \\ 25 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$$

The difficulty of solving a matrix equation rapidly increases with the number of unknowns. For the problem above we have two unknowns, and the problem is easy to solve. However, if a problem has N unknowns then the difficulty is increases like N^3.

The reason this matters, is that the equation A x = b has far more applications than just sorting out our shopping. These vary from medical imaging to weather forecasting and from designing aircraft to building a bridge. In fact, it is the central equation for any problem when we have a (large number) of unknowns which we need to determine from a set of knowns and a relation between them. If the relation is a linear operator, then it leads directly to the equation. In contrast if it is nonlinear then the same equations arise repeatedly in a sequence of iterations to find a sequence of approximations approaching to the correct solution.

Just as importantly it can either arise in its own right, as in the problem we have described, or as a discretisation of a partial differential equation.

In these sorts of problems the number of unknowns N can easily be a billion or higher. Such problems can take weeks to solve!

I will demonstrate this though two very different examples.

The first is the *Poisson equation*

$$-\nabla^2 u = f$$

Here u(x,y,z) is a function, as is f(x,y,z), and $\nabla^2 u$ is the differential operator

$$\nabla^2 u = u_{xx} + u_{yy} + u_{zz}$$

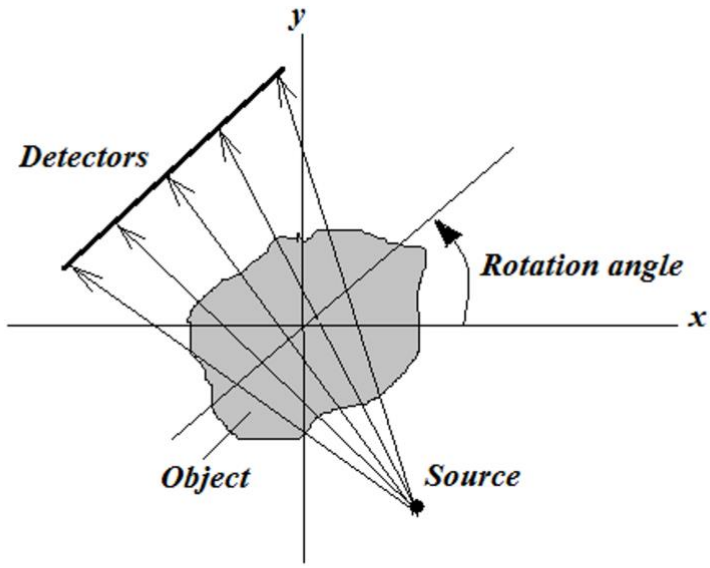
If u is a voltage and f is a charge then this is the fundamental equation of electrostatics, if u is a potential and f is a mass then it is the fundamental equation of (Newtonian) gravity, if u is a pressure and f is a velocity then it is the Pressure Poisson equation which lies at the heart of meteorology. In all cases we need to solve it as accurately as possible. To do this we introduce a mesh, and approximate the solution by just considering its values on the mesh. The *differential* operator on the function above then becomes a *discrete* operator on the values on the mesh, and such a discrete operator can be represented as a square matrix. For the operator u_{xx} on a mesh with 6 points this matrix takes the following form.

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

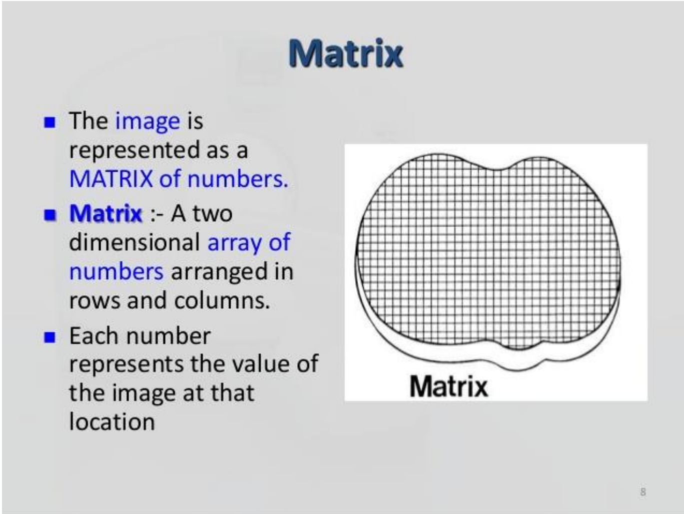
You will notice that most of the entries in this square matrix are zero. This is called a *sparse matrix* and many of the matrices arising in applications take this form. Special techniques have been developed to exploit this structure to allow such matrices to be inverted much more rapidly.

A major application of this approach comes in my favourite subject of weather forecasting. The causes of our weather are many and various, but the main drivers of our weather are the air pressure and the air temperature, coupled to the rotation of the earth. These in turn lead to movements of the air, which we see as wind, and of moisture, which we see as rain, snow, clouds or fog. The motion of the air satisfies the Navier-Stokes equations, which we will meet later in this talk and are nonlinear. However, a crucial part of solving these (what most weather forecasters consider to be the hardest part of the problem) is to determine the pressure. This is done by solving the Pressure Poisson equation, exactly the sort of equation as described above.

A quite different application arises in medical imaging. One powerful technique for looking inside the body is to use computerised axial tomography, or CAT for short. In this technique M X-rays are shone though the body, and the attenuation of each is measured. This gives a vector of measurements b = (b_1, b_2, b_3, b_4, … , b_M).



The body is then divided up into a mesh as illustrated below



In each square in the mesh the density of the body is x and the denser the body is, the more it absorbs X-rays. The relation of the density to the absorption is well understood and is a linear operator A just as before. Thus, by looking at how much the X-rays are absorbed over the body it is, in principle, possible to work out the density.

A crucial difference, between this problem and the shopping problem is that the number of unknowns N, which is the number of mesh cells, is generally very much larger than the total number, M, of the X-rays. The resulting M x N matrix is rectangular. Here is an example of a typical such problem in which N = 5 and M = 2.

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}.$$

The difference between this and the shopping problem is that it is unlikely that it has a unique solution. In fact, there may be many different solutions to the same problem. As an example, we could take N = 2 and M = 1 and have the simple problem.

$$x_1 + x_2 = 1$$

This equation has an *infinite* number of solutions including (1, 0), (0, 1) and (1/2, 1/2). This is a big problem if you are a doctor trying to interpret an image as there may be many possible images created by the computer, all of which fit the data. This is typical of an inverse problem. However, if you were a doctor then you would exercise your clinical judgement in filtering out which solution looks most appropriate. A rather similar approach is used in solving an inverse problem in which we use extra information about the solution to work out what is the best one. As an example for the above problem we could ask for the smallest solution (x1,x2) which satisfies the equation, where smallest is defined by the solution which minimises

$$x_1^2 + x_2^2$$

In this case we do have a unique answer, which is given by (x_1, x_2) = (1/2, 1/2).

For the more general case we can build in similar type of *a-priori information* where we say in advance what we think the solution will look like. This information is then combined with the data in a systematic way (using Bayesian analysis) to give an answer. The resulting method is then used widely in medical imaging. This gives images such as this one

I have spent several pages describing the equation A x = b without ever really saying how we go about actually solving it.

There are essentially two ways to do this. In a *direct method* we aim to find an expression for the inverse of the matrix A. One of the best ways to do this called Gaussian Elimination [3]. This method was originally described in 179AD in the Chinese mathematical text Chapter Eight: *Rectangular Arrays* of *The Nine Chapters on the Mathematical Art*. It was then rediscovered by Newton in 1670 and perfected in 1810 by Gauss in a manner suitable for hand calculations. Gaussian elimination is still used in various forms today and is probably the best method for solving a problem with under N = 100,000 unknowns. However, for much larger problems, such as those in meteorology, it is too slow and memory intensive. Indeed, the time it takes to do a calculation is proportional to N^3 which rapidly increase as N increases. In this case a different approach is needed. The most powerful of these is to use iterative methods. In such a method an initial guess is made about the solution. This guess is then successively improved until the solution is obtained. One of the nice features about this method for a problem such as medical imaging, is that we can stop this process at any point if we think that the solution looks good enough. This cannot be done in a direct method. As a result, iterative methods are now very widely used. The current best of these as a compromise between speed and simplicity of use is the amazing *Conjugate gradient method* [4]. Every student of science should be taught about this method which is the method of choice to solve Ax = b, and even works well for ill-posed inverse problems.

Equation two: The matrix eigenvalue equation

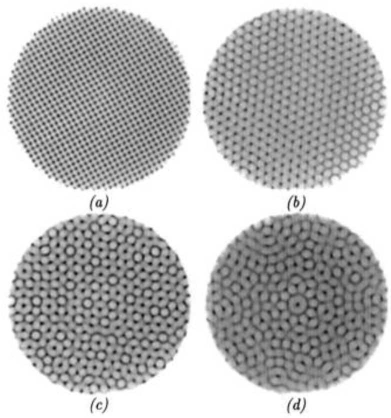
$$A\,\mathbf{x} = \lambda\,\mathbf{x}$$

For our second equation we will look at what is usually called the *matrix eigenvalue problem*. In this problem A is a matrix or a differential operator as before, and lambda and x are both unknowns. This equation can claim to be at the heart of all of modern technology, physics, chemistry, and engineering.

I will illustrate this equation by looking at our behaviour in a shower. Showers will feature later in this talk in another context. Imagine that you are in the shower and start singing. In particular sing your way through a scale. Some of the notes in the scale will sound much louder than others. These are called the *resonant frequencies* of the shower. They are closely linked to the above equation. In particular, if A is the Laplacian differential operator, and omega is a resonant frequency then omega satisfies the (Helmholtz) equation

$$\nabla^2 \mathbf{x} = -\omega^2 \mathbf{x}$$

Which is simply a version of the matrix eigenvalue equation with A= nabla^2 and lambda = -omega^2. In fact, if you replace A by the discrete approximation we looked at in the previous section, then these two equations become identical. The values of omega which satisfy this equation are the resonant vibrational frequencies of the shower, and the vectors x tell you the shapes of the shower as it is vibrating. These can be very complex, and the complexity is all built into the equation above. As an example, here are some of the vibrational modes of the very similar problem of water in a circular dish.



The same equation allows us to find the resonant modes of many other systems, which may be suspension bridges, cars, aeroplanes and football stadia. The latter is vital, as if a crowd is singing, and dancing to, the team song of the home team, and this contains a resonant frequency of the stadium, then the crowd may shake it to bits.

Let's now have a look at a simple example of this equation with the matrix

$$A = \begin{pmatrix} 2.5 & 0.5 \\ 0.5 & 2.5 \end{pmatrix}$$

Then a careful calculation shows that

$$A \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad A \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

so, the possible values of lambda are lambda = 3 and lambda = 2.

The possible values of lambda are called the *eigenvalues* of A, and the possible vectors x are called the *eigenvectors*.

More generally if A is an n*n matrix then the matrix eigenvalue equation has n solutions. The eigenvalues then satisfy the equation

$$\det(A - \lambda I) = 0$$

where det(A) is the determinant of the matrix A.

For a general 2*2 matrix of the form

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

then the determinant condition leads to the quadratic equation

$$\lambda^2 - (a+d)\lambda + ad - bc = 0$$

This is one of the many reasons that the quadratic equation is so important. What is particularly interesting is that if the original problem is to do with vibrational modes, and solutions of the quadratic equation are real numbers then the vibrations simply die away. Whereas if they are complex numbers then this leads to solutions of the original problem which oscillate.

In the case of an n*n matrix we get a polynomial of degree n which in general can't be solved exactly. It is in fact very hard to solve the matrix eigenvalue equation even on a computer. The best algorithm to do it, the QR algorithm [5], is really quite slow and hard to use.

However, it is well worth solving the matrix eigenvalue equation, because its applications are so diverse.

As well as giving the resonant frequencies of suspension bridges, the same equation allows us to work out the vibrational frequencies of molecules, which is essential for *chemistry*. Further chemistry can be done by solving Schrodinger's equation, which takes the form

$$i\hbar \frac{\partial \psi}{\partial t} = \left[\frac{-\hbar^2}{2m} \nabla^2 + V(\mathbf{r}, t) \right] \psi$$

This equation lies at the heart of quantum mechanics. By solving it we can find the wave function describing the state of a quantum-mechanical system. This means everything we might want to know about. If this lecture was a list of the great equations of physics, then this equation would be nearly top of the list. However, as soon as we discretise it on a computer it becomes the matrix eigenvalue equation, and this makes it into this section.

In a similar manner the matrix eigenvalue equation, via the Helmholtz equation we looked at earlier, lies at the heart of any technology that makes use of *waves*. Such technology includes all of radio, TV, radar, mobile phones, WiFi as well as acoustics.

However, perhaps my favourite application comes from Google. If ever I am asked to justify studying mathematics, I point out that Google relies heavily on mathematical ideas, as indeed does the whole of the Internet.

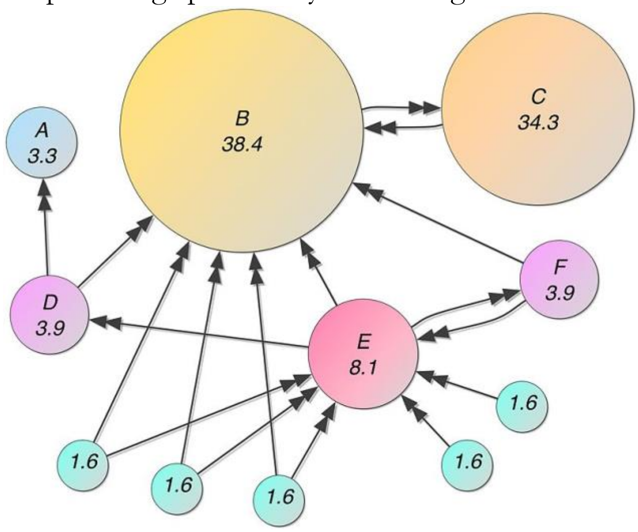
The original algorithm behind Google, called the PageRank algorithm, in fact involves solving exactly the matrix eigenvalue problem. The PageRank algorithm is used to search website pages, and appropriately enough, was invented by Larry Page, who was one of the founders of Google. Legend had it that he devised the algorithm as an undergraduate after attending a maths course in Stanford about eigenvalues. Suppose that we want to search for a topic such as Gresham College. The idea behind the algorithm, is to assign all the webpages which mention Gresham a rank. The rule for associating a rank, are that

- Every page P_i mentioning Gresham gets a well-defined rank R_i.
- Each page P_i will be pointed to by m other pages. These will have ranks R_ij, and each will respectively point to N_i1, N_i2, …, N_im other pages.
- The rank R_i of page P_i satisfies the condition:

$$R_i = \frac{(1-d)}{N} + d\left(\frac{R_{i1}}{N_{i1}} + \frac{R_{i12}}{N_{i2}} + \ldots + \frac{R_{im}}{N_{im}}\right).$$

Here d is the *damping factor* eg. d = 0.85 and N is the total number of pages.

The formula uses a model of a *random surfer* who reaches their target site after several clicks, then switches to a random page. The PageRank value of a page reflects the chance that the random surfer will land on that page by clicking on a link. Here is an example of a set of pages which satisfy this property with d = 0.85, which is taken from [6]. The numbers show the percentage probability for landing on each of the 11 pages.



In this figure you can see that page B has the highest rank and is thus the most likely to be found.

The equation for the PageRank R involves R on both sides and to find it we must solve a linear equation of the form of Equation 1. However, this is very hard to do because the size of this system is absolutely huge. Google has to search through over one billion websites to find the rank. Worse still Google has to do this calculation very fast indeed, usually in under a second, if it is to be useful to anyone searching the Internet. However, there is a very neat way to do the calculation by turning it into Equation 2. In fact, it can be shown that the vector of all R values satisfies the matrix eigenvalue problem with eigenvalue one, given by

M R = R

Here M is called the *augmented adjacency matrix*, expressing the way that the internet is interconnected. In fact, R is an eigenvector of M with an *eigenvalue of 1*. This makes it easy to find R quickly. It is done iteratively by starting with a guess R^0 and then performing the iteration

$$R^{n+1} = MR^n$$

This procedure converges rapidly to the correct answer. It is the basis of the Google algorithm which gives the right answer in a fraction of a second.

So, whenever you are using Google you are solving a huge matrix eigenvalue problem, very very fast. Who says that maths isn't useful!

Equation 3: The Fourier Transform and its inverse.

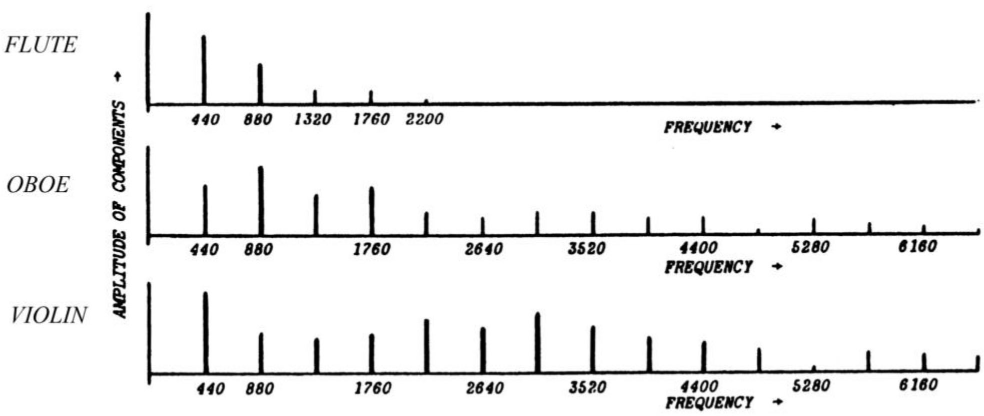
$$\hat{f}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega t} f(t)\, dt,$$

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t} \hat{f}(\omega)\, d\omega$$

My next equation comes as a pair. Suppose that you listen to an orchestra, can you work out from the sound what musical instruments are playing. Or if you hear the sound of a violin, can you work out how to reproduce the sound. On a seemingly different topic, if you have a blurred photograph, can you recreate the original unblurred picture. All of these are questions which can be answered by using the Fourier transform, which is the first of the equations above, and the inverse Fourier transform, which is the second. The Fourier Transform lies at the heart of the telecommunications industry. It is also of vital importance in most area of mathematics (including probability and statistics) as well as in physics, chemistry, medical imaging, and many more besides.

The Fourier transform (FT) expresses (or decomposes) a function f(t) as a combination of waves with frequency omega and amplitude fhat(omega). One way to think about this is to imagine an orchestra with different instruments each playing notes of different frequencies. The first of the equations tells us how loudly each instrument is being played, and the second shows how the different instruments combine to give the total sound that we hear. In fact, we can do better than that. In the orchestra we will have several instruments such a violin, a trumpet, and oboe and a piano. Each can play the same note, say A, but this note will sound very different on each instrument. The reason for this is that the instrument makes not just the note A, but all of the harmonics of that note as well. The amplitude of these harmonics is different for each these instruments, and this is why they sound different. Now, by using the Fourier transform, if we feed in a recording f(t) of the instrument into the first formula, then the function fhat(omega) tells us the amplitude of the different harmonics. This is very useful. If we know the amplitude of the different harmonics of an instrument, then we can recreate the sound of it. This is how synthesisers work.

Below we can see the harmonics of the notes from a flute, and oboe and a violin calculated using the Fourier Transform. You can see from this that the flute has relatively few harmonics which explains its very pure tone. In contrast the large number of harmonics for the violin leads to the richness of its tone.



The Fourier transform is a method for turning one function f(t) into another fhat(omega). The inverse Fourier transform turns fhat(omega) back into f(t).

If a function f(t) is a simple sinusoid function

$$f(t) = \sin(\omega_0 t)$$

then the Fourier Transform is a single harmonic (to be precise it is a pair of delta functions centred on omega = omega_0 and omega = -omega_0). This is why it is so effective in sorting out the harmonics of the flute, violin etc.

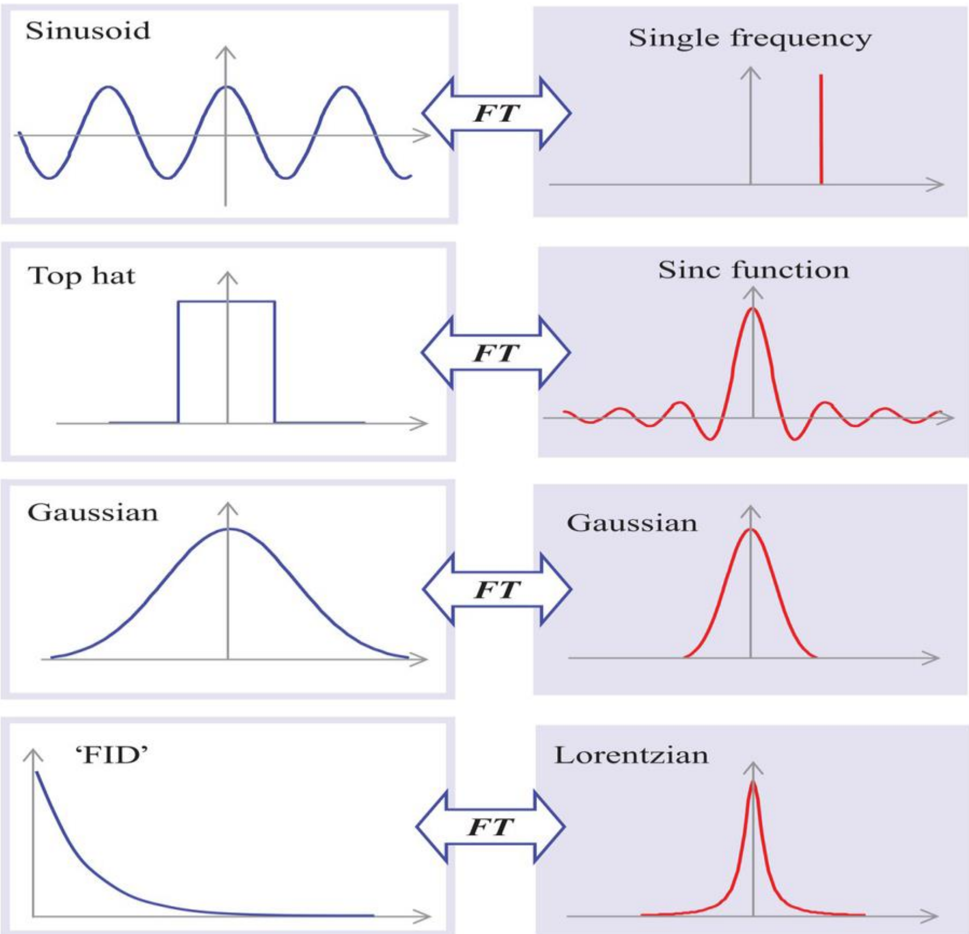
However, it can be applied to many more functions. An example is given by taking f(t) equal to the top-hat function.

$$f(t) = 0 \quad \text{if} \quad |t| > 1, \qquad f(t) = 1 \quad \text{if} \quad |t| < 1.$$

An example of this function would come from going into a darkened room, switching on the light for a short time, and then switching it off again. The function f(t) is then the illumination of the room. The Fourier Transform of this is the function

$$\hat{f}(\omega) = \frac{\sin(\omega)}{\omega}$$

Which is also called sinc(omega). Some more examples of functions and their Fourier Transforms are given below



If the sinc function is familiar (especially if you are a physicist) it is because it is the diffraction pattern of light through a slit. This is not a coincidence. Light is made up of waves, and the Fourier Transform pays a very important role in both optics and acoustics in calculating many different types of diffraction pattern. For an essentially similar reason it is also a central tool in quantum mechanics.

The Fourier Transform also plays a very important role in signal and image processing. Imagine that you are trying to transmit information through a communications channel (for example radio channel or a telephone wire). The information inevitably gets distorted. Similarly, if we take a photograph of a subject, then the photograph can get blurred. The process behind this blurring is well understood. If f(t) is the signal going through the channel, and g(t) is a blurring function, which is a property of the channel, then the output h(t) of the channel is given by

$$h(t) = \int_{-\infty}^{\infty} f(t-\tau)\, g(\tau)\, d\tau.$$

The function h(t) is often called the *convolution* of f(t) with g(t) and is written as

$$h = f * g$$

The convolution integral is hard to compute directly, and to work it out naively would take a large amount of computing time. However, one of my favourite formulae in the whole of mathematics gives a simple expression for h in terms of the Fourier Transform and is given by

$$\hat{h} = \hat{f}\ \hat{g}.$$

This result is called the *convolution theorem* and its central importance to modern telecommunications cannot be overstated. To work out h, we find the Fourier transforms of f and g, multiply them and apply the inverse Fourier transform. Perhaps more importantly, if we receive h(t) and we want to find out f(t) then we divide the Fourier Transform of h by that of g and take the inverse to find f. If the function h represents the intensity levels in a blurred photograph, then f will be the intensity levels of the unblurred original. In practice we have to be a bit more careful, as h(t) will also be distorted by noise. However, this basic idea lies at the heart of many of the algorithms used to remove distortion from a signal. As such it is of extreme importance in dealing with distorted TV, radio and mobile phone signals. The human brain also does something similar when it processes the sounds that it receives from the ear. The modern technique of Magnetic Resonance Imaging (MRI) makes extensive use of the Fourier transform to make it work.

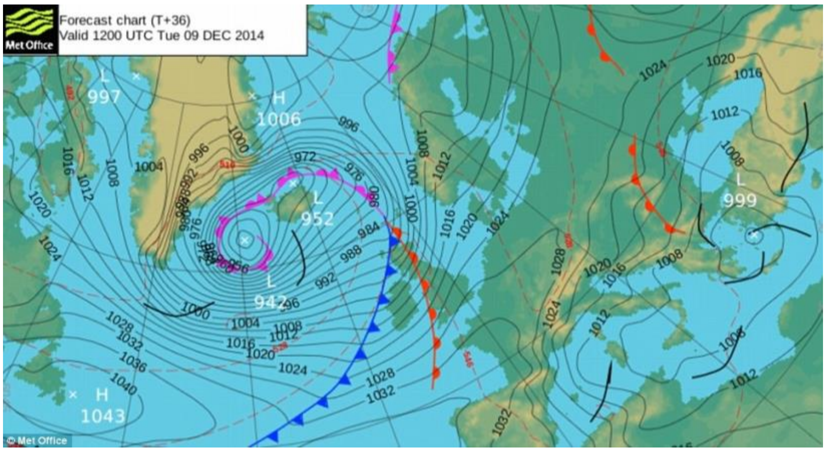
Before After

I hope that you can see that the Fourier Transform lies at the heart of a huge amount of modern technology. However, none of this would be possible if it were not possible to compute the Fourier Transform quickly. Fortunately, the Fourier Transform can be calculated rapidly by using Fast Fourier Transform or FFT. This algorithm was reinvented and implemented on a digital computer by Cooley and Tukey in 1965 [7], although it had been known since pioneering work by Gauss in the early 19_{th} Century. The FFT transformed the telecommunications industry and led directly to the modern digital revolution. It is a good example (as is the Page Rank algorithm) of a piece of mathematics this led to a whole new technology. Shortly after the introduction of the FFT, the Beatles released the single 'We can Work it Out', but maybe that was just a coincidence.

Equation 4: The Navier-Stokes equations and a touch of chaos.

You may not have heard of the Navier-Stokes equations, but you encounter them every day. These are the equations that describe the weather, and in an extended form, our climate!

We all know that the weather is important to our lives, that it can sometimes be predictable and other times very unpredictable. It seems strange that all of the different types of behaviour we associate with the weather can be described by a single set of equations, but we believe that this is the case, and they make a fitting subject for our fourth great equation.



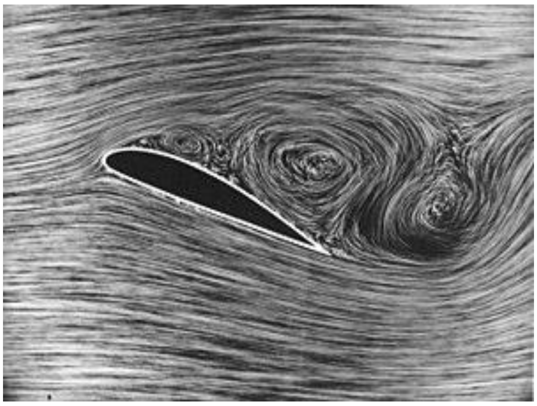
We start by thinking about exactly what we mean by *weather*. Weather is the combination of the motion of the atmosphere, coupled to the motion of the oceans, the transport of moisture within the atmosphere, all coupled to changes in the pressure and temperature of the air. It is possible to write down (partial differential) equations which describe all of this mathematically. In their totality these equations are rather complicated, however at the guts of them are the equations that describe the underlying motion of the air on its own. These equations are the same for air as they are for water or any other fluid and they were derived in the 19th century by the two mathematicians Navier and Stokes, hence their name the *Navier-Stokes equations*. (Although it is fair to say that they are a natural extension of Newton's laws of motion, and a still useful approximation to them, the Euler Equations, was derived by Euler in 1757.)

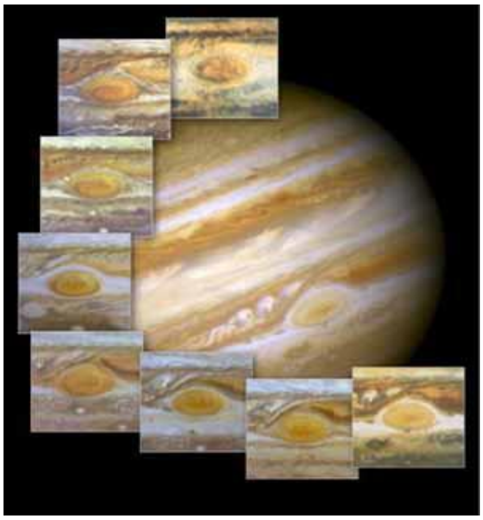
Imagine that you are looking at a point in the atmosphere. At this point the air will have a *velocity* **u**, a density rho, and a *pressure* P. It is then affected by the rotation f of the Earth and its gravity g. These are all related together by the *Navier-Stokes equations* which describe how changes to the velocity in *time* are related to changes in the velocity and the pressure in *space*. Brace yourselves...here they come!

$$\frac{D\mathbf{u}}{Dt} + 2\mathbf{f} \times \mathbf{u} = -\frac{1}{\rho}\nabla P + \frac{1}{Re}\nabla^2 \mathbf{u} - g\mathbf{k},$$

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

Although these equations look rather brutal, by the time you have done a university course in mathematics, physics or engineering they will become old friends. The term *Re* in the equations is called the *Reynolds Number*. It is low if the fluid is very sticky (like treacle) and high if it is hardly sticky at all, like air or water. What is really remarkable is that the same equations describe the motion of the water in a cup of coffee, the evolution of a hurricane, the behaviour of the ocean currents, the flow of air around an aircraft wind, the flow of coolant in a power station, the future climate of the Earth, the motion of blood around the body and even the atmosphere of Jupiter - all of this physics is coded into just one set of equations.

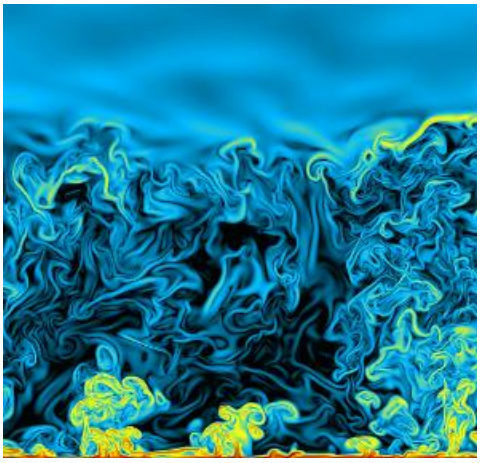




Unfortunately, there is some bad news to come. The first piece of bad news is that the Navier-Stokes equations are very, very hard to solve. We only know of a few exact solutions (that is, solutions which we can write down using a formula), usually for problems which are of little or no physical interest.

A lot of work has been done on finding *approximate solutions* which work for certain important physical situations, such as the flow of water in a pipe. The procedures for finding such solutions dominate the subject called *fluid mechanics*, which you may meet in a university course in applied mathematics, physics or engineering (especially aeronautical engineering). Fortunately, it is also possible to write computer programmes which can find numerical solutions to these equations. Indeed, there is a huge industry called *computational fluid dynamics* devoted to this task. It is computer programs of this type which are used by the meteorological office to help predict the weather. They are also used in the design of aircraft and cars, the study of blood flow, the design of power stations, the analysis of the effects of pollution, the study of the insides of a star, calculations of climate change and, in a notable success story, in the design of *Thrust 2*, the first supersonic racing car. (This calculation was done by the computational fluid dynamics team at the University of Swansea).

The next piece of bad news is that even in the best of circumstances these programs can take even the fastest computer an enormous time to run, and these computers can often only solve relatively simple problems. They are also quite unable to deal with the phenomenon called *turbulence*. Turbulence is the complex behaviour that fluids show on small length scales. (An understanding of turbulence is one of the great problems in physics for the new millennium.) You can see or feel turbulence every time that you look at a cloud, examine the motion of the water in a waterfall or stick your head out of a car window. An example of a turbulent flow is shown below. No computer programme on earth can simulate this behaviour exactly, and at the present all that we have are rough approximations. It is a sobering thought that these approximations have to be used every time a simulation is made of a safety-critical situation (such as the effects of a fire or of a coolant leak in a nuclear power station). The approximations are not bad (giving errors of around 20%), but this situation is hardly satisfactory.



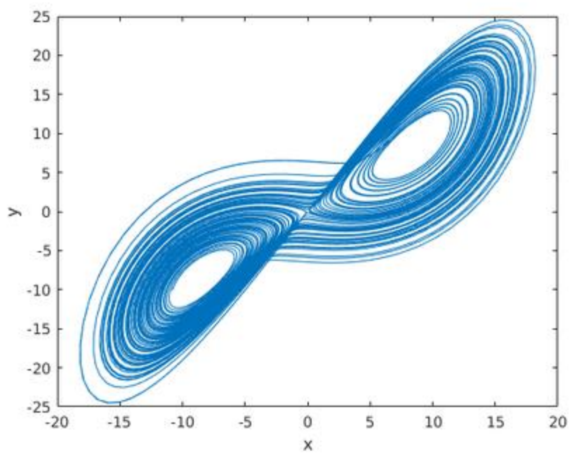
Turbulence is closely related to the (possibly simpler) subject of *chaos* which I covered in an earlier lecture. A system is chaotic if it is described by (simple) mathematical equations, but it has behaviour which is complex and apparently unpredictable. The reason that the weather is very hard to predict is that the Navier-Stokes equations seem to have chaotic solutions. This makes it very hard to predict the weather accurately more than two weeks into the future. This phenomenon was first recognised by Lorenz in the 1960s. Lorenz was studying a simplification of the Navier-Stokes equations, now called the Lorenz equations. This system of three *ordinary* differential equations is much simpler than the Navier-Stokes equations, but is still a valid approximation to an atmospheric flow with convection. These equations are given by:

$$\frac{dx}{dt} = \sigma(y - x),$$

$$\frac{dy}{dt} = x(\rho - z) - y,$$

$$\frac{dz}{dt} = xy - \beta z.$$

In the 1960's electronic computers had just become available, and Lorentz was able to find (approximate) solutions of the above equations using these. He was in for a shock! Rather than behaving regularly, the solutions were very erratic. The term chaotic was given to them shortly after. If you plot the (x,y) graph of the solutions you can see the complexity, in a butterfly shaped curve, which is now called a strange attractor. Lorentz was delighted. At last he understood why the weather was so complex.



It is interesting to note that the Lorentz equations are in many ways very similar in form to the SIR equations currently being used to predict the behaviour of the COVID-19 epidemic [8]. Here S is the stock of the susceptible population, I is the stock of the infected population, R is the stock of the recovered population, and N is the sum of these.

$$
\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta IS}{N}, \\
\frac{dI}{dt} &= \frac{\beta IS}{N} - \gamma I, \\
\frac{dR}{dt} &= \gamma I.
\end{aligned}
$$

Both models comprise three ordinary differential equations with quadratic nonlinearities. However, they have very different solutions. Most importantly the SIR model can, and is, used to give long term predictions about the epidemic. In contrast the Lorentz equations predict that the weather is unpredictable.

All of the above issues are very important to the way that we use the Navier-Stokes equations to help us to understand the physical world around us, but to mathematicians they pale into insignificance when compared to what is the subject of the sixth Millennium Prize Problem of the Clay Institute. (See [9] for a description of all of the others.) This problem was posed in 2000 and there is a $1000 000 reward if you can solve it. The problem is not whether we can solve the Navier-Stokes equations (either exactly or using a computer), but *whether they have any solutions at all.*

You may feel that this is an unimportant question - after all it is obvious - isn't it? - that the equations *must* have a solution. After all we have lots of weather. However, there are plenty of examples in mathematics of equations which don't have solutions. For example, before the invention of negative numbers, the equation $x+1=0$ had no solution. The Greeks thought that all numbers could be expressed in terms of fractions (rational numbers) and had a very deep shock when they discovered that the equation $x^2=2$ did not have a solution, which could be expressed as a fraction. Similarly, if you only knew about real numbers then it would not be possible to solve the equation $x^2=-1$. It is quite possible that a situation could occur in which a possible solution of the Navier-Stokes equations starts by being completely physical, but quickly becomes infinite and fails to represent anything corresponding to the physical situation that the equations are trying to describe.

The situation in 2020 is that no one has yet managed to show that the solutions of the Navier-Stokes equations correspond to real physical solutions for all time. Conversely, no one has found a "solution" of the Navier-Stokes equations that becomes infinite and loses its physical meaning. If such a solution were to be found, would it really be nonsense, or might it give us some insight into the problem of turbulence (the latter being the view of the author)? We don't know! What we do know is that for the cases that we can compute, the Navier-Stokes equations do seem to give a very accurate description of the motion of fluids and that they also seem to be *uniquely hard.* If we make a small change to the equations, then we can answer all of the questions but return to the physically motivated equations and nothing is certain. Mathematicians seem evenly divided as to whether solutions exist or not, and the question of the existence of solutions of the Navier-Stokes equations seems likely to stay unresolved for a long time.

The shower equation

My last choice of an equation is often called the *(hot) shower equation*, and is given by

$$\frac{dx}{dt} = -\lambda\, x(t-\tau).$$

Here tau is called a delay. I include this equation not only because of its intrinsic mathematical interest, but also because it helps us to control many things, including, possibly, the spread of the current COVID-19 pandemic. Other applications of this equation are in climate dynamics, lasers, El Nino, agriculture, population dynamics, control theory in general and (of course) the behaviour of showers. It deserves to be better known.

Imagine that you are in a shower. It feels too cold, so you turn the heating knob up. Nothing happens immediately because it takes water time to get around the pipes. So, you end up turning up the up the heat of the water higher and higher. Eventually the water gets around the pipes and into the shower and then onto you. It is then far too hot, ouch! So, you immediately turn the water down. But by the time it gets to you it is now too cold, brrrr! So, you turn it up again. And so, the cycle of turning the shower up and down continues, and it seems impossible to get it right. I'm sure all of us have been in this situation. If not, go and try it for your self.

The shower equation describes this situation. If x(t) is the temperature of the shower as we feel it, the equation says that the change in this temperature (due to our turning the control knob, is a consequence of what happened a time tau ago, where tau is the time that it takes for the water to get around the plumbing. The parameter lambda tells us how quickly the shower responds to changes in the knob setting. The shower equation, or similar, arises in many problems where we are trying to control the output of a system with a delay in it. A very good example is the use of methods by the government to try to control the current COVID-19 emergency.

We can solve this equation by posing a possible solution of the form

$$x(t) = e^{\alpha t}$$

If we substitute this into the shower equation, we get

$$\alpha e^{\alpha t} = -\lambda e^{\alpha(t-\tau)} = -\lambda e^{\alpha t} e^{-\alpha \tau}$$

So that alpha satisfies the lovely (transcendental) equation

$$\alpha = -\lambda \, e^{-\alpha \tau}.$$

Or, if $z = -\alpha \tau$ and $\gamma = \lambda \tau$

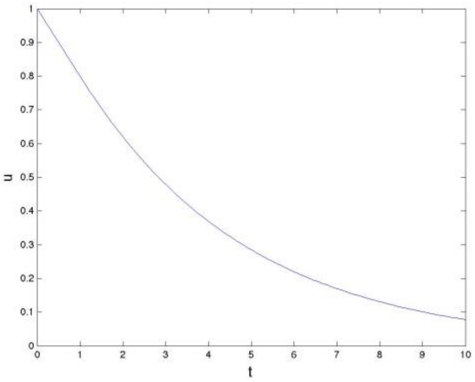
$$z = \gamma e^{z}$$

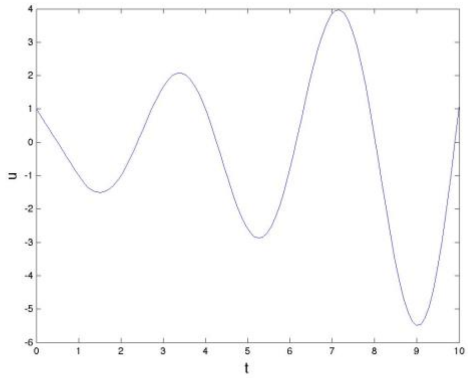
The solutions of this equation depend upon gamma. If

$$\gamma < 1/e = 0.3679..$$

then this equation has two real solutions, both of which lead to solutions of the shower equation which decay to zero. This means that this equation is stable, and we can control the shower. However, is gamma > 0.3679 then the solutions oscillate in the manner we alluded to in the introduction to the equation. As gamma increases beyond pi/2 = 1.5708… then the oscillations get larger, and the shower gets more and more uncontrollable. In contrast, if there was no delay then as gamma increases the solutions would tend more and more rapidly to zero and the shower would be very easy to control. So, we see that a bit of a delay fundamentally changes the nature of the way that the shower can be controlled.

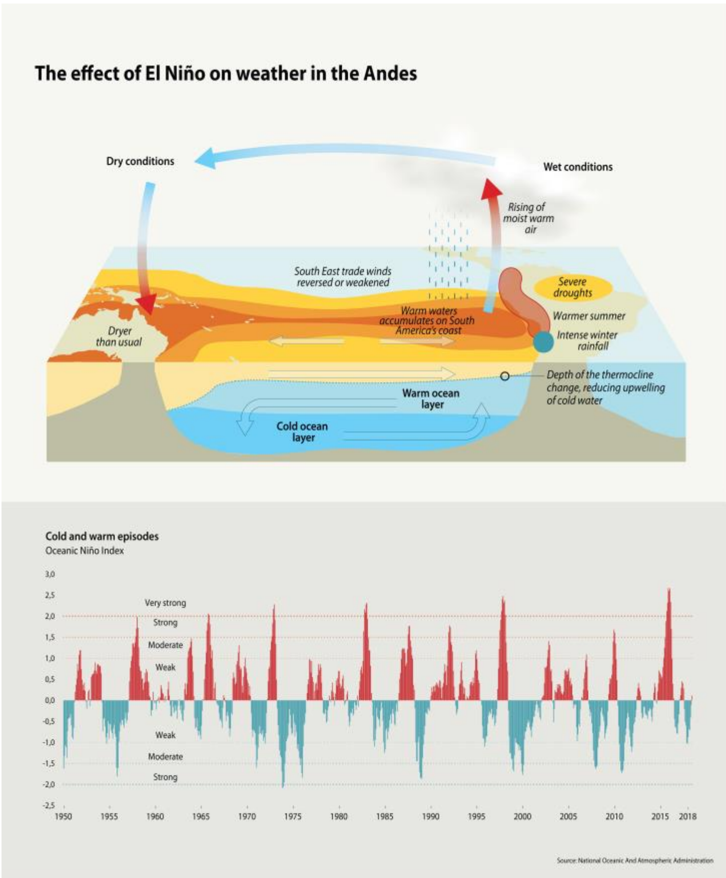
The two graphs below illustrate this. These are calculated by solving the shower equation on a computer and assuming that there is a past history for which $x(t) = 1$ if $t < 0$. On the left is the resulting stable evolution of the solutions of the shower equation when gamma = 0.2 and on the right the unstable oscillations that we see in the solutions when gamma = 2.





Now, showers are of course an important part of our lives (well certainly of mine) and controlling them is important. However, the applications of the shower equation extend far beyond showers.

One of the most important of these is in the study of the dynamics of the climate. The reason for this is that many climate phenomena take time to have an effect. For example, if we change the amount of Carbon Dioxide that we put into the atmosphere now, then we will have to wait some time before we see the actual effects of this on the Earth's temperature. This makes it both hard to determine the effects of Carbon Dioxide reduction and can potentially lead to uncontrolled oscillations. Another example comes in the changes to the temperature of the South Pacific in the El Nino Southern Oscillation (ENSO). This is an irregular change in the temperature with an approximate four-year interval between warming events. El Nino has a major effect on the economy not only in the region but of the whole world. If we could predict it better, then this would help the communities in the Pacific to prepare for it.



The ENSO is caused by an interaction between the ocean currents and the atmosphere, which changes the temperature of sea. It can be modelled by a equation very similar to the shower equation. In this case the delay is caused by the time that it takes for the ocean currents to travel from the West coast of South America to the East coast of Asia and then return (see the above figure.) The change in the temperature of the oceans is then delayed by this amount leading to a similar equation to the shower equation. This leads to the periodicity that we see. In fact there are extra nonlinear terms, which lead to the irregularity that we see above, which is an example of chaotic dynamics superimposed on a periodic oscillation.

The shower equation is, unfortunately, very relevant to our current emergency due to the COVID-19 virus. The virus has an incubation time of between 5 days and two weeks during which no symptoms are visible. Thus, any intervention that is made by Government to control the virus will have to wait for this period before it has any effect. This leads directly to versions of the shower equation, the so-called SIR equation with latency and control, that are used to help understand, and control, the epidemic. A critical component in this analysis is the time lag from controls via policy intervention, to observed response, as well as the uncertainty in model, parameters, and states of the system. I said earlier that the SIR equations (unlike the Lorentz equations) have predictable solutions [8]. However, like the ENSO system, once delay is built into them, things become much more uncertain, and this is the problem that we are facing at them moment.

At time of writing it remains to see how controllable the (health and economic) system is as a result.

References

[1] I. Stewart, (2012), *17 equations that changed the world,* Profile Books.

[2] G. Farmelo, (2002), *It Must Be Beautiful: Great Equations Of Modern Science,* Granta Books.

[3] Wikipedia article on Gaussian Elimination https://en.wikipedia.org/wiki/Gaussian_elimination

[4] Wikipedia article on the Conjugate Gradient method
https://en.wikipedia.org/wiki/Conjugate_gradient_method

[5] Wikipedia article on the QR algorithm
https://en.wikipedia.org/wiki/QR_algorithm

[6] Wikipedia article on the PageRank algorithm
https://en.wikipedia.org/wiki/PageRank

[7] J. Cooley and J. Tukey, (1965), *An algorithm for the machine calculation of complex Fourier series,* Math. Comput. **19** (90), pp 297–301

[8] A series of articles on the COVID-19 epidemic in the Plus Maths magazine
https://plus.maths.org/content/tags/covid-19

[9] K. Devlin, (2005), *The Millennium Problems,* Granta Books.