Mathematics For Quant Finance

Introduction to Mathematical Methods

Riaz Ahmad*

This is a revision course designed to act as a mathematics refresher. The volume of work covered is significantly large so the emphasis is on working through the notes and problem sheets. The four topics covered in detail are

- Calculus
- Linear Algebra
- Probability
- Differential Equations

*r.ahmad@7city.com

1 Introduction to Calculus

1.1 Basic Terminology

Mathematics is not actually that complex. Historically books on the subject tend to be written in a fairly incomprehensible manner by mathematicians who want to look clever. In addition the basic notation used can put many off. Once familiar with a particular style of terminology, another text or lecturer uses a different style of nomenclature. Quant Finance is no exception. We start by presenting some standard mathematical shorthand:

Natural Numbers $\mathbb{N} = \{0, 1, 2, 3, ...\}$

Integers $(\pm\mathbb{N})\,\mathbb{Z} = \{0,\ \pm 1,\ \pm 2,\ \pm 3,\ .....\}$

Rationals $\mathbb{Q} = \left\{\frac{1}{2},\ 0.76,\ 2.25,\ 0.3333333....\right\}$

Irrationals $\left\{\sqrt{2},\ 0.01001000100001...,\ \pi,\ e\right\}$

Reals $\mathbb{R}$ all the above

Complex numbers $\mathbb{C} = \left\{x + iy : i = \sqrt{-1}\right\}$

$\exists$	there exists	$\longrightarrow$	which gives
$\forall$	for all	s.t	such that
$\therefore$	therefore	$!x$	a unique x
$\because$	because	iff	if and only if

$\equiv$	equivalent
$\sim$	similar
$\in$	an element of

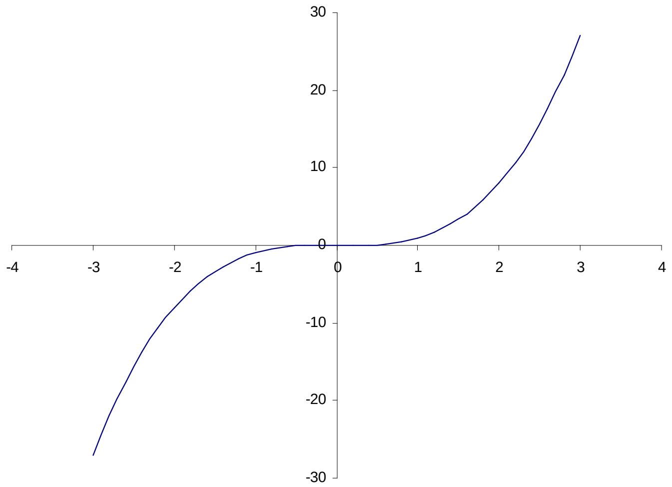
1.2 Functions

A *function* $f(x)$ of a single variable x is a rule that assigns each element of a set X (written $x \in X$) to exactly one element y of a set Y $(y \in Y)$. A function is denoted by the form $y = f(x)$ or:

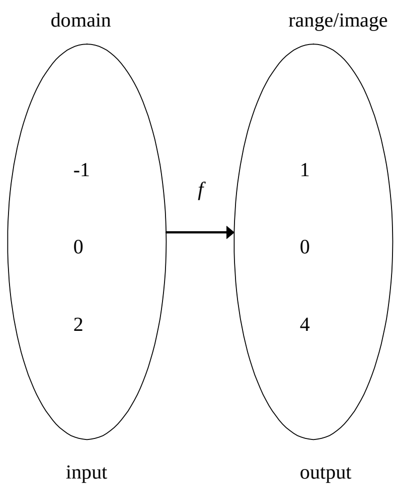
$$x \mapsto f(x).$$

We can also write $f : X \longrightarrow Y$

For example,if $f(x) = x^3$, then $f(-1) = -1^3 = -1$



We often write $y = f(x)$ where y is the *dependent variable* and x is the *independent variable.*



The set X is called the *domain* of f and the set Y is called the *image* (or *range*), written $\mathbf{Dom}\ f$ and **Im** f, in turn.

For a given value of x there should be at most one value of y. The definition of a function requires this property.

The example above is a $\mathbf{1} - \mathbf{1}$ function.

$f(x) = x^2$ is a *many to one function.*

$f(x) = \sqrt{x}$ is a *one to many mapping.* Note we do not use the term function here.

The *inverse function* $f^{-1}(x)$ is defined so that

$$f\left(f^{-1}(x)\right) = x \text{ and } f^{-1}(f(x)) = x.$$

Thus $\sqrt{x}$ and x^2 are inverse mappings (for $x \geq 0$).

Further Notation:

$(a, b) = a < x < b$ *open interval*

$[a, b] = a \leq x \leq b$ *closed interval*

$(a, b] = a < x \leq b$ semi-*open/closed interval*

$[a, b) = a \leq x < b$ semi-*open/closed interval*

Example 1: What is the inverse of $y = 2x^2 - 1$.

i.e. we want y^{-1}. This can be done by rearranging the function as

$$x = \sqrt{\frac{y+1}{2}}$$

therefore $y^{-1}(x) = \sqrt{\frac{x+1}{2}}$.

Check:

$$yy^{-1}(x) = 2\left(\sqrt{\frac{x+1}{2}}\right)^2 - 1 = x = y^{-1}y(x)$$

Example 2: Consider $f(x) = 1/x$, therefore $f^{-1}(x) = 1/x$

$$\mathbf{Dom} f = (-\infty, 0) \cup (0, \infty) \text{ or } \mathbb{R} - \{0\}$$

Returning to the earlier example

$$y = 2x^2 - 1$$

clearly $\mathbf{Dom} f = \mathbb{R}$ (clearly)

and for

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

to exist we require the term inside the square root sign to be non-negative, i.e. $\frac{x+1}{2} \geq 0 \implies x \geq -1$, therefore $\mathbf{Dom} f = \{[-1, \infty)\}$.

An *even function* is one which has the property

$$f(-x) = f(x)$$

e.g. $f(x) = x^2$.

The function we plotted earlier $f(x) = x^3$ is an example of an *odd function* because

$$f(-x) = -f(x).$$

1.2.1 Polynomials

These are functions which involve powers of x,

$$\begin{aligned} f(x) &= a_0 + a_1 x + a_2 x^2 + \ldots.. \\ &\quad .. + a_{n-1} x^{n-1} + a_n x^n. \end{aligned}$$

The highest power is called the *degree* of the polynomial - so $f(x)$ is an n^{th} degree polynomial. We can express this more compactly as

$$f(x) = \sum_{k=0}^{n} a_k x^k$$

where the coefficients of x are constants.

$k = 1, 2$ gives a linear and quadratic in turn. The most general form of quadratic equation is

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

which can be solved for x using the formula

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

There are three cases to consider:

(1) $b^2 - 4ac > 0 \longrightarrow x_1 \neq x_2 \in \mathbb{R}$: 2 distinct real roots

(2) $b^2 - 4ac = 0 \longrightarrow x = x_1 = x_2 = -\dfrac{b}{2a} \in \mathbb{R}$: one two fold root

(3) $b^2 - 4ac < 0 \longrightarrow x_1 \neq x_2 \in \mathbb{C}$ - Complex conjugate pair

1.2.2 Explicit/Implicit Representation

When we express a function as $y = f(x)$, then we can obtain y corresponding to a (known) value of x. We say y is an *explicit* function. All known terms are on the right hand side (rhs) and unknown on the left hand side (lhs). For example

$$y = 2x^2 + 4x - 16 = 0$$

Occasionally we may write a function in an *implicit* form $f(x, y) = 0$, although in general there is no guarantee that for each x there is a unique y.

A trivial example is

$$y - x^2 = 0,$$

which in its current form is implicit. Simple rearranging gives $y = x^2$ which is explicit.

A more complex example is

$$4y^4 - 2y^2x^2 - yx^2 + x^2 + 3 = 0.$$

So we see all known and unknown variables are bundled together. An implicit form which does not give rise to a function is

$$y^2 + x^2 - 16 = 0.$$

This can be written as

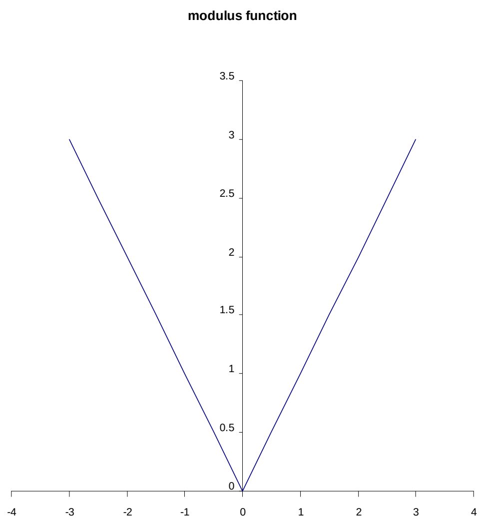
$$y = \sqrt{16 - x^2}.$$

and e.g. for $x = 0$ we can have either $y = 4$ or $y = -4$, i.e. one to many.

1.2.3 The Modulus Function

Sometimes we wish to obtain the absolute value of a number, i.e. positive part. For example the absolute value of -3.9 is 3.9. In maths there is a function which gives us the absolute value of a variable x called the *modulus function*, written $|x|$ and defined as

$$y = |x| = \begin{cases} x & x \geq 0 \\ -x & x < 0 \end{cases}$$



1.2.4 The exponential and log functions

The *logarithm* (or simply log) was introduced to solve equations of the form

$$a^p = N$$

and we say p is log of N to base a. That is we take logs of both sides ($\log_a$)

$$\log_a a^p = \log_a N$$

which gives

$$p = \log_a N.$$

By definition $\log_a a = 1$ (important).

We will often need the exponential function e^x and the (natural) logarithm $\log_e x$ or $(\ln x)$. Here

$$e = 2.718281828\ldots.$$

which is the approximation to

$$\left(1 + \frac{1}{n}\right)^n$$

when n is very large. Similarly the exponential function can be approximated from

$$\left(1+\frac{x}{n}\right)^n$$

$\ln x$ and e^x are mutual inverses:

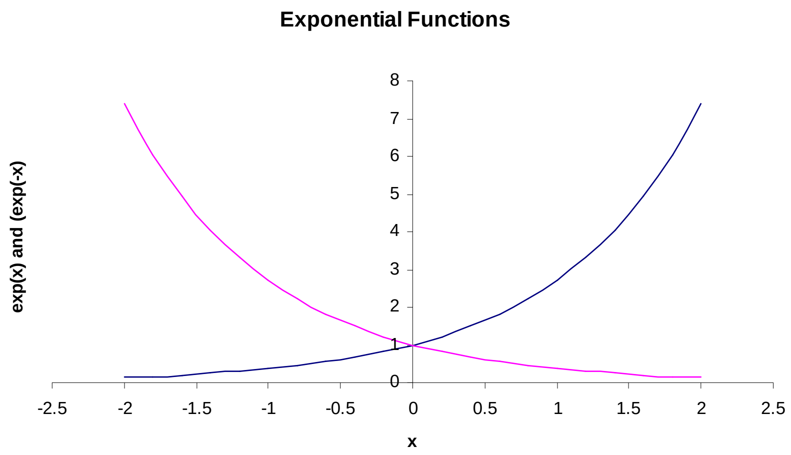
$$\log\left(e^x\right) = e^{\log x} = x.$$

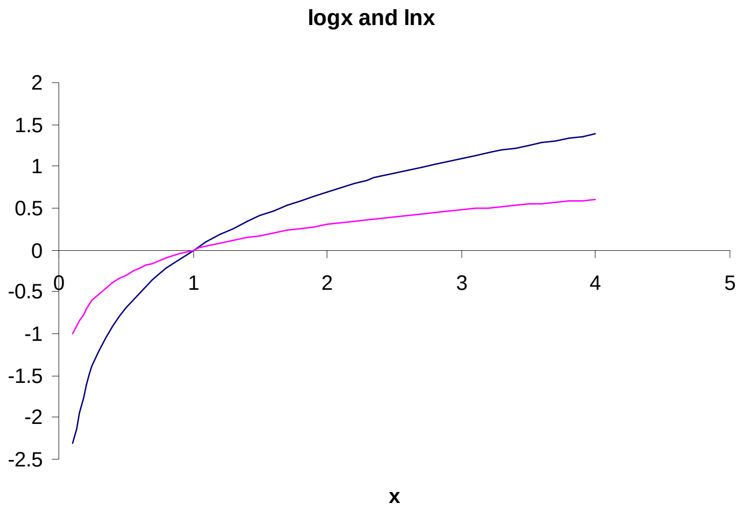
Also

$$\frac{1}{e^x} = e^{-x}.$$

Here we have used the property $(x^a)^b = x^{ab}$, which allowed us to write $\frac{1}{e^x} = (e^x)^{-1} = e^{-x}$.

Their graphs look like this:





We see that $\log_{10} x$ grows faster than natural logarithm.

Note that e^x is always strictly positive. It tends to zero as x becomes very large and negative, and to infinity as x becomes large and positive. To get an idea of how quickly e^x grows, note the approximation $e^5 \approx 150$.

Later we will also see $e^{-x^2/2}$, which is particularly useful in probability. This function decays particularly rapidly as $|x|$ increases. The term $-x^2/2$ is called the *exponent*, i.e. the "bit" above the exponential. When there is a large exponent (many terms) then it is common to write

$$\exp(.......)$$

Note:

$$e^x e^y = e^{x+y}, \qquad e^0 = 1$$

(recall $x^a.x^b = x^{a+b}$) and

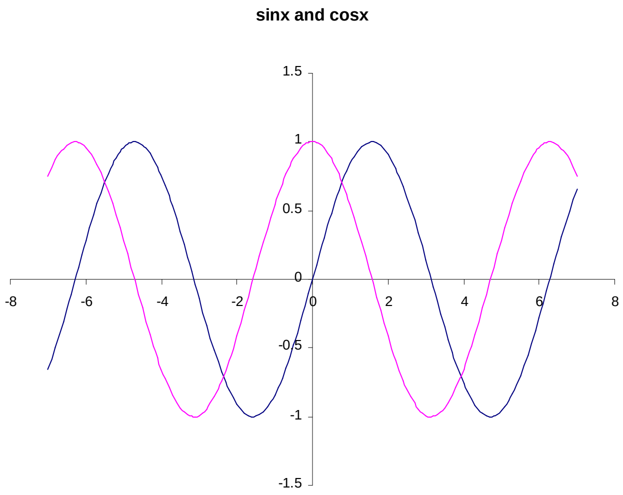
$$\log(xy) = \log x + \log y, \quad \log(1/x) = -\log x, \;\; \log 1 = 0.$$

$$\log\left(\frac{x}{y}\right) = \log x - \log y.$$

$$\begin{aligned} \mathbf{Dom}\,(e^x) &= \mathbb{R} \\ \mathrm{Im}\,(e^x) &= (0,\infty) \end{aligned}$$

$$\begin{aligned} \mathbf{Dom}\,(\ln x) &= (0,\infty) \\ \mathrm{Im}\,(\ln x) &= \mathbb{R} \end{aligned}$$

1.2.5 Trigonometric Functions



$\sin x$ is an *odd* function, i.e. $\sin(-x) = -\sin x$.

It is *periodic* with period 2π: $\sin(x + 2\pi) = \sin x$. This means that after every 360° it repeats itself.

$\sin x = 0 \iff x = n\pi \ \forall n \in \mathbb{Z}$

$\mathbf{Dom} = \mathbb{R}$ and $\mathrm{Im} = [-1, 1]$

$\cos x$ is an *even* function, i.e. $\cos(-x) = \cos x$.

It is *periodic* with period 2π: $\cos(x + 2\pi) = \cos x$.

$\cos x = 0 \iff x = (2n+1)\frac{\pi}{2}\ \forall n \in \mathbb{Z}$

$\mathbf{Dom} = \mathbb{R}$ and $\mathrm{Im} = [-1, 1]$

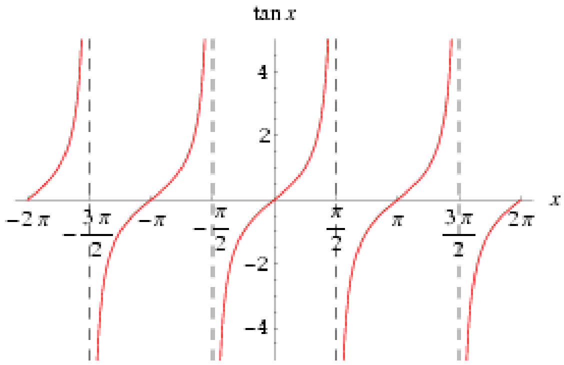
Trigonometric Identities:

$$\begin{aligned}
\cos^2 x + \sin^2 x &= 1 \\
\sin(x \pm y) &= \sin x \cos y \pm \cos x \sin y \\
\cos(x \pm y) &= \cos x \cos y \mp \sin x \sin y \\
\sin\left(x + \frac{\pi}{2}\right) &= \cos x \\
\cos\left(\frac{\pi}{2} - x\right) &= \sin x
\end{aligned}$$

$$\tan x = \frac{\sin x}{\cos x}$$

This is an odd function: $\tan(-x) = \tan x$

Periodic: $\tan(x + \pi) = \tan x$



$$\mathbf{Dom} = \{x : \cos x \neq 0\} = \left\{x : x \neq (2n+1)\,\frac{\pi}{2};\; n \in \mathbb{Z}\right\} = \mathbb{R} - \left\{(2n+1)\,\frac{\pi}{2};\; n \in \mathbb{Z}\right\}$$

The inverse trigonometric functions are defined by

$$\sec x = \frac{1}{\cos x}; \quad \csc x = \frac{1}{\sin x}; \quad \cot x = \frac{1}{\tan x}$$

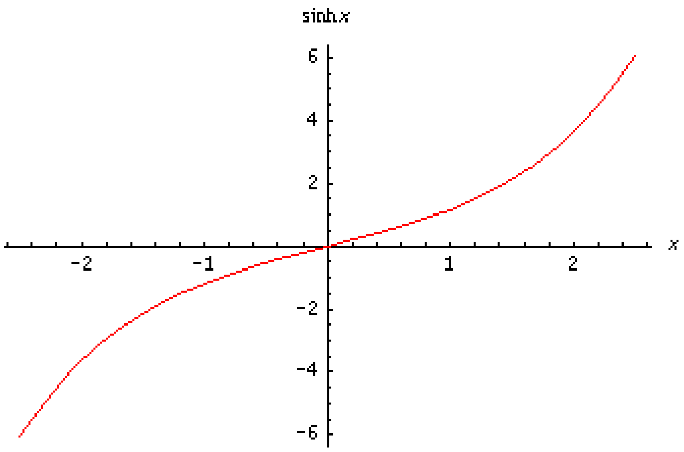
1.2.6 Hyperbolic Functions

$$\sinh x = \frac{1}{2}\left(e^{x} - e^{-x}\right)$$

Odd function: $\sinh(-x) = -\sinh x$

$\mathbf{Dom} = \mathbb{R}$

$\mathrm{Im} = \mathbb{R}$

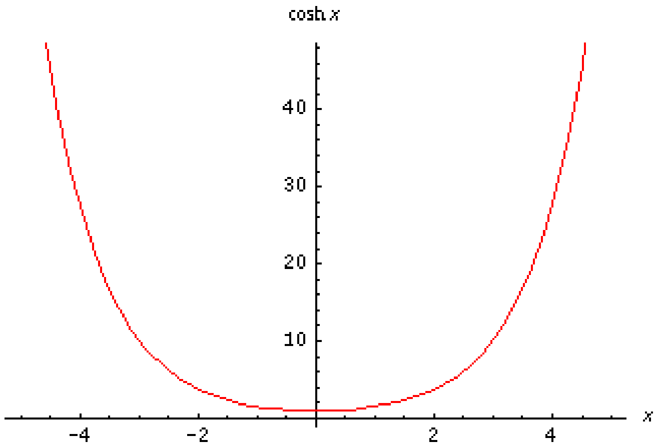


$$\cosh x = \frac{1}{2}\left(e^{x} + e^{-x}\right)$$

Even function: $\cosh(-x) = \cosh x$

Dom $=\mathbb{R}$

$\mathsf{Im} = [1, \infty)$



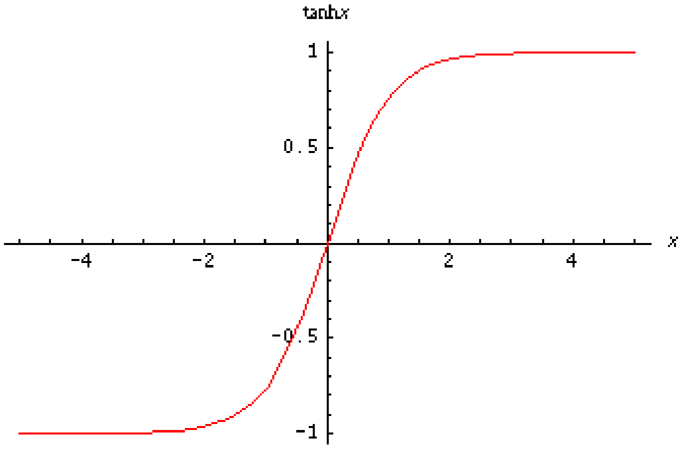
Identities:

$$\begin{aligned}
\cosh^2 x - \sinh^2 x &= 1 \\
\sinh(x+y) &= \sinh x \cosh y + \cosh x \sinh y \\
\cosh(x+y) &= \cosh x \cosh y + \sinh x \sinh y
\end{aligned}$$

$$\tanh x = \frac{\sinh x}{\cosh x}$$

Dom $=\mathbb{R}$

$\mathsf{Im} = (-1, 1)$



Inverse Hyperbolic Functions

$y = \sinh^{-1} x \longrightarrow x = \sinh y = \frac{\exp y - \exp(-y)}{2};$

$2x = \exp y - \exp(-y)$

multiply both sides by $\exp y$ to obtain $2xe^y = e^{2y} - 1$ which can be written as

$$(e^y)^2 - 2x(e^y) - 1 = 0.$$

This gives us a quadratic in e^y therefore

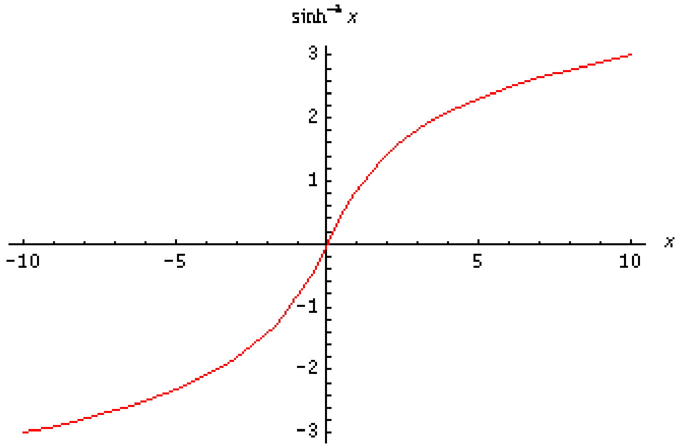
$$e^y = \frac{2x \pm \sqrt{4x^2 + 4}}{2} = x \pm \sqrt{x^2 + 1}$$

Now $\sqrt{x^2 + 1} > x \Longrightarrow x - \sqrt{x^2 + 1} < 0$ and we know that $e^y > 0$ therefore we have $e^y = x + \sqrt{x^2 + 1}$. Hence taking logs of both sides gives us

$$\sinh^{-1} x = \ln \left| x + \sqrt{x^2 + 1} \right|$$

$$\mathbf{Dom}\left(\sinh^{-1}\mathbf{x}\right) = \mathbb{R}$$

$$\mathrm{Im}\left(\sinh^{-1} x\right) = \mathbb{R}$$



Similarly $y = \cosh^{-1} x \longrightarrow x = \cosh y = \frac{\exp y + \exp(-y)}{2}$;

$2x = \exp y + \exp(-y)$ and again multiply both sides by $\exp y$ to obtain

$$(e^y)^2 - 2x(e^y) + 1 = 0.$$

and

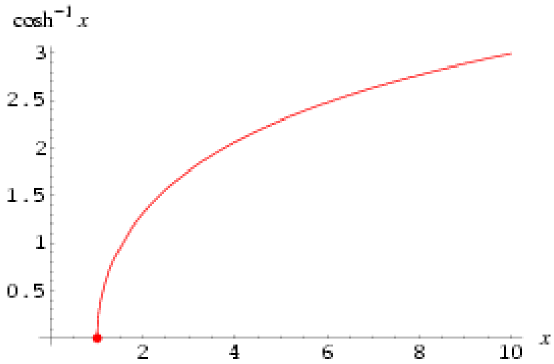
$$e^y = x + \sqrt{x^2 - 1}$$

We take the positive root (not both) to ensure this is a function.

$$\cosh^{-1} x = \ln \left| x + \sqrt{x^2 - 1} \right|$$

$$\mathbf{Dom}\left(\cosh^{-1}\mathbf{x}\right) = [1, \infty)$$

$$\mathrm{Im}\left(\cosh^{-1} x\right) = [0, \infty)$$



We finish off by obtaining an expression for $\tanh^{-1} x$. Put $y = \tanh^{-1} x \longrightarrow$

$$x = \tanh y = \frac{\exp y - \exp(-y)}{\exp y + \exp(-y)};$$

$$x \exp y + x \exp(-y) = \exp y - \exp(-y)$$

and as before multiply through by e^y

$$\begin{aligned} x \exp 2y + x &= \exp 2y - 1 \\ \exp 2y\,(1-x) &= 1 + x \longrightarrow \exp 2y = \frac{1+x}{1-x} \end{aligned}$$

taking logs gives

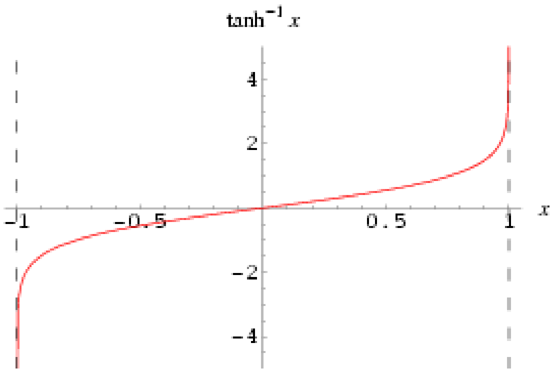
$$2y = \ln\left|\frac{1+x}{1-x}\right|$$

hence

$$\tanh^{-1} x = \tfrac{1}{2}\ln\left|\frac{1+x}{1-x}\right|$$

$\mathbf{Dom}\left(\tanh^{-1}\mathbf{x}\right) = (-1, 1)$

$\operatorname{Im}\left(\tanh^{-1} x\right) = \mathbb{R}$



1.3 Limits

Choose a point x_0 and function $f(x)$. Suppose we are interested in this function near the point $x = x_0$. The function need not be defined at $x = x_0$. We write $f(x) \longrightarrow l$ as $x \longrightarrow x_0$, "if $f(x)$ gets closer and closer to l as x gets close to x_0". Mathematically we write this as

$$\lim_{x \longrightarrow x_0} f(x) \longrightarrow l,$$

if $\exists$ a number l such that

- Whenever x is close to x_0

- $f(x)$ is close to l.

Let us have a look at a few basic examples and corresponding "tricks" to evaluate them

Example 1:

$$\lim_{x \longrightarrow 0} \left(x^2 + 2x + 3\right) \longrightarrow 0 + 0 + 3 \longrightarrow 3;$$

Example 2:

$$\lim_{x \longrightarrow \infty} e^{-x} \longrightarrow 0; \quad \lim_{x \longrightarrow \infty} e^{x} \longrightarrow \infty; \quad \lim_{x \longrightarrow 0} e^{x} \longrightarrow e^{0} = 1.$$

Example 3:

$$\lim_{x \longrightarrow \infty} \frac{x^2 + 2x + 2}{3x^2 + 4} \quad = \quad \lim_{x \longrightarrow \infty} \frac{\frac{x^2}{x^2} + \frac{2x}{x^2} + \frac{2}{x^2}}{\frac{3x^2}{x^2} + \frac{4}{x^2}} =$$

$$\lim_{x \longrightarrow \infty} \frac{1 + \frac{2}{x} + \frac{2}{x^2}}{3 + \frac{4}{x^2}} \quad \longrightarrow \quad \frac{1}{3}.$$

Example 4:

$$\lim_{x \longrightarrow 3} \frac{x^2 - 9}{x - 3} = \lim_{x \longrightarrow 3} \frac{(x+3)(x-3)}{(x-3)} = \lim_{x \longrightarrow 3} (x+3) \longrightarrow 6$$

The limit only exists if

$$\begin{aligned} f(x) &\longrightarrow l \text{ as } x \to x_0^- \\ f(x) &\longrightarrow l \text{ as } x \to x_0^+ \end{aligned}$$

More Examples:

$$\lim_{x \longrightarrow 0} \sin x \longrightarrow 0$$

$$\lim_{x \longrightarrow 0} \frac{\sin x}{x} \longrightarrow 1$$

$$\lim_{x \longrightarrow 0} |x| \longrightarrow 0$$

What about $\lim_{x \to 0} \frac{|x|}{x}$?

$$
\begin{aligned}
\lim_{x \to 0^+} \frac{|x|}{x} &= 1 \\
\lim_{x \to 0^-} \frac{|x|}{x} &= -1
\end{aligned}
$$

therefore $\frac{|x|}{x}$ does not tend to a limit as $x \to 0$.

1.4 Continuity

A function $f(x)$ is **continuous** at x_0 if

$$\lim_{x \to x_0} f(x) = f(x_0).$$

That is, 'we can draw its graph without taking the pen off the paper'.

1.5 Differentiation

How fast does a function $f(x)$ change with x? The **gradient** or **derivative** of $f(x)$,written

$$f'(x) \text{ or } \frac{df}{dx}$$

is defined for each x as

$$f'(x) = \lim_{\delta x \to 0} \frac{f(x + \delta x) - f(x)}{\delta x}$$

assuming the limit exists (it may not). Differentiability implies continuity (but converse does not always hold).

The earlier form of the derivative given is also called a *forward derivative.* Other possible definitions of the derivative are

$$\begin{aligned} f'(x) &= \lim_{h \to 0} \frac{1}{h}\left(f(x) - f(x-h)\right) \quad \textit{backward} \\ f'(x) &= \lim_{h \to 0} \frac{1}{2h}\left(f(x+h) - f(x-h)\right) \quad \textit{centred} \end{aligned}$$

Examples:

Differentiating x^2 from first principles:

$$\begin{aligned} f(x) &= x^2 \\ f(x+h) &= (x+h)^2 = x^2 + 2xh + h^2 \\ \frac{f(x+h) - f(x)}{h} &= \frac{2hx + h^2}{h} \\ &= 2x + h \\ &\longrightarrow 2x \quad \text{as} \quad h \to 0; \end{aligned}$$

$$\frac{d}{dx}x^n = nx^{n-1};$$

$$\frac{d}{dx}e^x = e^x; \quad \frac{d}{dx}e^{ax} = ae^{ax};$$

$$\begin{aligned}
\frac{d}{dx}\log x &= \frac{1}{x} \\
\frac{d}{dx}\cos x &= -\sin x \\
\frac{d}{dx}\sin x &= \cos x \\
\frac{d}{dx}\tan x &= \sec^2 x
\end{aligned}$$

and so on. Take these as defined (standard results).

The inverse trigonometric functions are defined by

$$\sec x = \frac{1}{\cos x}; \quad \csc x = \frac{1}{\sin x}; \quad \cot x = \frac{1}{\tan x}$$

Examples:

$$\begin{aligned} f(x) &= x^5 \to f'(x) = 5x^4 \\ g(x) &= e^{3x} \to g'(x) = 3e^{3x} = 3g(x) \end{aligned}$$

1.5.1 Rules For Differentiation

Linearity

If λ and μ are constants and $y = \lambda f(x) + \mu g(x)$ then

$$\frac{dy}{dx} = \frac{d}{dx}(\lambda f(x) + \mu g(x)) = \lambda f'(x) + \mu g'(x).$$

Thus if $y = 3x^2 - 6e^{-2x}$ then

$$dy/dx = 6x + 12e^{-2x}.$$

1.5.2 Product Rule

If $y = f(x)\,g(x)$ then

$$\frac{dy}{dx} = f'(x)\,g(x) + f(x)\,g'(x)\,.$$

Thus if $y = x^3 e^{3x}$ then

$$dy/dx = 3x^2 e^{3x} + x^3 \left(3e^{3x}\right) = 3x^2 (1 + x)\, e^{3x}.$$

We can derive this rule as follows. Put $h(x) = f(x)\,g(x)$. If $\frac{dh}{dx}$ exists, we can define

$$\begin{aligned} h'(x) &= \lim_{\delta x \to 0} \frac{h(x + \delta x) - h(x)}{\delta x} \\ &= \lim_{\delta x \to 0} \frac{f(x + \delta x)\,g(x + \delta x) - f(x)\,g(x)}{\delta x} \end{aligned}$$

To evaluate the limit, we perform a small trick, subtract and add $f(x + \delta x)\,g(x)$ from the numerator

$$\frac{f(x+\delta x)g(x+\delta x)-f(x+\delta x)g(x)+f(x+\delta x)g(x)-f(x)g(x)}{\delta x}$$

which can be written as

$$\lim_{\delta x\to 0}\left(f\left(x+\delta x\right).\frac{g(x+\delta x)-g(x)}{\delta x}+g\left(x\right).\frac{f(x+\delta x)-f(x)}{\delta x}\right)$$

so

$$\begin{aligned} h'\left(x\right) &= \lim_{\delta x\to 0} f\left(x+\delta x\right).\lim_{\delta x\to 0}\frac{g\left(x+\delta x\right)-g\left(x\right)}{\delta x} \\ &\quad +\lim_{\delta x\to 0} g\left(x\right).\lim_{\delta x\to 0}\frac{f\left(x+\delta x\right)-f\left(x\right)}{\delta x} \\ &= f\left(x\right)\lim_{\delta x\to 0}\frac{g\left(x+\delta x\right)-g\left(x\right)}{\delta x}+ \\ &\quad g\left(x\right).\lim_{\delta x\to 0}\frac{f\left(x+\delta x\right)-f\left(x\right)}{\delta x} \\ &= f\left(x\right)g'\left(x\right)+g\left(x\right)f'\left(x\right) \end{aligned}$$

1.5.3 Function of a Function Rule

Differentiation is often a matter of breaking a complicated problem up into simpler components.

The function of a function rule is one of the main ways of doing this. If

$y = f(g(x))$ then

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

Thus if $y = e^{4x^2}$ then

$$dy/dx = e^{4x^2} 4.2x = 8xe^{4x^2}.$$

So differentiate the whole function, then multiply by the derivative of the "inside" $(g(x))$.

Another way to think of this is in terms of the **chain rule.**

Write $y = f(g(x))$ as

$$y = f(u), \quad u = g(x).$$

Then

$$\begin{aligned} \frac{dy}{dx} &= \frac{d}{dx} f(u) = \frac{du}{dx}\frac{d}{du} f(u) = g'(x) f'(u) \\ &= g'(x) f'(g(x)). \end{aligned}$$

Symbolically, we write this as

$$\frac{dy}{dx} = \frac{du}{dx}\frac{dy}{du}$$

provided u is a function of x alone.

Thus for $y = e^{4x^2}$, write $u = 4x^2$, $y = e^u$. Then

$$\frac{dy}{dx} = \frac{du}{dx}\frac{dy}{du} = 8xe^{4x^2}.$$

Further examples:

$y = \sin x^3$

$$\begin{aligned} y &= \sin u, \text{ where } u = x^3 \\ y' &= \cos u.3x^2 \longrightarrow y' = 3x^2 \cos x^3 \end{aligned}$$

$y = \tan^2 x$: this is how we write $(\tan x)^2$ so put

$$\begin{aligned} y &= u^2 \text{ where } u = \tan x \\ y' &= 2u.\sec^2 x \longrightarrow y' = 2\tan x \sec^2 x \end{aligned}$$

$y = \ln \sin x$. Put $u = \sin x \longrightarrow y = \ln u$

$$\frac{dy}{du} = \frac{1}{u}, \quad \frac{du}{dx} = \cos x$$

hence $y' = \cot x$.

1.5.4 Quotient Rule

If $y = \dfrac{f(x)}{g(x)}$ then

$$\frac{dy}{dx} = \frac{g(x)\, f'(x) - f(x)\, g'(x)}{(g(x))^2}.$$

Thus if $y = e^{3x}/x^2$,

$$\frac{dy}{dx} = \frac{x^2 3e^{3x} - 2xe^{3x}}{x^4} = \frac{3x - 2}{x^3}e^{3x}.$$

This is a combination of the product rule and the function of a function (or chain) rule. It is very simple to derive:

Staring with $y = \dfrac{f(x)}{g(x)}$ and writing as $y = f(x)(g(x))^{-1}$ we apply the product rule

$$\frac{dy}{dx} = \frac{df}{dx}(g(x))^{-1} + f(x)\frac{d}{dx}(g(x))^{-1}$$

Now use the chain rule on $(g(x))^{-1}$; i.e. write $u = g(x)$ so

$$\begin{aligned}\frac{d}{dx}(g(x))^{-1} &= \frac{du}{dx}\frac{d}{du}u^{-1} = g'(x)\left(-u^{-2}\right) \\ &= -\frac{g'(x)}{g(x)^2}.\end{aligned}$$

Then

$$\frac{dy}{dx} = \frac{1}{g(x)}\frac{df}{dx} - f(x)\frac{g'(x)}{g(x)^2} = \frac{f'(x)}{g(x)} - \frac{f(x)\,g'(x)}{g(x)^2}.$$

To simplify we note that the common denominator is $g(x)^2$ hence

$$\frac{dy}{dx} = \frac{g(x)\, f'(x) - f(x)\, g'(x)}{g(x)^2}.$$

Examples:

$$\begin{aligned} \frac{d}{dx}(xe^x) &= x\frac{d}{dx}(e^x) + e^x \frac{d}{dx}(x) \\ &= xe^x + e^x = e^x (x+1)\,; \end{aligned}$$

$$\begin{aligned} \frac{d}{dx}(e^x/x) &= \frac{x\,(e^x)' - e^x\,(x)'}{(x)^2} = \frac{xe^x - e^x}{x^2} \\ &= \frac{e^x}{x^2}(x-1)\,; \end{aligned}$$

$$\begin{aligned} \frac{d}{dx}\left(e^{-x^2}\right) &= \frac{d}{dx}(e^u) \quad \text{where} \quad u = -x^2 \therefore du = -2x dx \\ &= (-2x)\, e^{-x^2}. \end{aligned}$$

1.5.5 Implicit Differentiation

Consider the function

$$y = a^x$$

where a is a constant. If we take natural log of both sides

$$\ln y = x \ln a$$

and now differentiate both sides by applying the chain rule to the left hand side

$$\begin{aligned} \frac{1}{y}\frac{dy}{dx} &= \ln a \\ \frac{dy}{dx} &= y \ln a \end{aligned}$$

and replace y by a^x to give

$$\frac{dy}{dx} = a^x \ln a.$$

This is an example of *implicit differentiation.*

We could have obtained the same solution by initially writing a^x as a combination of a **log** and **exp**

$$\begin{aligned} y &= \exp(\ln a^x) = \exp(x \ln a) \\ y' &= \frac{d}{dx}\left(e^{x \ln a}\right) = e^{x \ln a}\frac{d}{dx}(x \ln a) \\ &= a^x \ln a. \end{aligned}$$

Consider the earlier implicit function given by

$$4y^4 - 2y^2x^2 - yx^2 + x^2 + 3 = 0.$$

The resulting derivative will also be an implicit function. Differentiating gives

$$16y^3y' - 2\left(2yy'x^2 + 2y^2x\right) - \left(y'x^2 + 2xy\right) = -2x$$

$$\begin{aligned} \left(16y^3 - 4yx^2 - x^2\right)y' &= -2x + 4y^2x + 2xy \\ y' &= \frac{-2x + 4y^2x + 2xy}{16y^3 - 4yx^2 - x^2} \end{aligned}$$

1.5.6 Alternative Proof of the Product Rule

The proof of this rule can be fairly rigorous. However we can present a fairly simple working using the log of a function to obtain the proof of the product rule: Start with

$$y = f(x)\, g(x)$$

and now take log of both sides

$$\log y = \log f(x) + \log g(x)$$

differentiating implicity gives

$$\begin{aligned} \frac{1}{y}\frac{dy}{dx} &= \frac{1}{f}\frac{df}{dx} + \frac{1}{g}\frac{dg}{dx} \\ &= \frac{gdf + fdg}{fgdx} \end{aligned}$$

taking $y = f(x)\, g(x)$ across gives

$$\begin{aligned} \frac{dy}{dx} &= \left(\frac{gdf + fdg}{fgdx}\right) fg \\ &= \frac{gdf + fdg}{dx} \\ &= g\frac{df}{dx} + f\frac{dg}{dx}. \end{aligned}$$

1.5.7 Higher Derivatives

These are defined recursively;

$$\begin{aligned} f''(x) &= \frac{d^2 f}{dx^2} = \frac{d}{dx}\left(\frac{df}{dx}\right) \\ f'''(x) &= \frac{d^{\,3} f}{dx^3} = \frac{d}{dx}\left(\frac{d^2 f}{dx^2}\right) \end{aligned}$$

and so on. For example:

$$\begin{aligned} f(x) &= 4x^3 \\ f^{\,\prime}(x) &= 12x^2 \longrightarrow f''(x) = 24x \\ f'''(x) &= 24 \longrightarrow f^{(iv)}(x) = 0. \end{aligned}$$

so for any n^{th} degree polynomial

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots\dots + a_1 x + a_0$$

we have $f^{(n+1)}(x) = 0$.

Consider another example

$$\begin{aligned} f(x) &= e^x \\ f'(x) &= e^x \longrightarrow f''(x) = e^x \\ &\vdots \\ f^{(n)}(x) &= e^x = f(x) \end{aligned}$$

$$\begin{aligned} f(x) &= \log x \\ f'(x) &= 1/x \\ f''(x) &= -1/x^2 \\ f'''(x) &= 2/x^3. \end{aligned}$$

Warning

Not all functions are differentiable everywhere. For example, $1/x$ has the derivative $-1/x^2$ but only for $x \neq 0$.

Easy way is to "look for a hole", e.g. $f(x) = \dfrac{1}{x-2}$ does not exist at $x = 2$.

$x = 2$ is called a *singularity* for this function. We say $f(x)$ is *singular* at the point $x = 2$.

1.5.8 Further Limits

This will be an application of differentiation. Consider the limiting case

$$\lim_{x \to a} \frac{f(x)}{g(x)} \equiv \frac{0}{0}$$

This is called an *indeterminate form.* Then *L' Hospitals rule* states

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)} = \; .... \; = \lim_{x \to a} \frac{f^{(r)}(x)}{g^{(r)}(x)}$$

for r such that we have the indeterminate form 0/0. If for $r+1$ we have

$$\lim_{x \to a} \frac{f^{(r+1)}(x)}{g^{(r+1)}(x)} \to A$$

where A is not of the form 0/0 then

$$\lim_{x \to a} \frac{f(x)}{g(x)} \equiv \lim_{x \to a} \frac{f^{(r+1)}(x)}{g^{(r+1)}(x)}.$$

Note: Very important to verify quotient has this indeterminate form before using L'Hospitals rule. Else we end up with an incorrect solution. We can also use this rule for the form $\dfrac{\infty}{\infty}$.

Examples:

1.

$$\lim_{x\to 0}\frac{\cos x + 2x - 1}{3x} \equiv \frac{0}{0}$$

So differentiate both numerator and denominator $\longrightarrow$

$$\lim_{x\to 0}\frac{\frac{d}{dx}(\cos x + 2x - 1)}{\frac{d}{dx}(3x)} = \lim_{x\to 0}\frac{-\sin x + 2}{3} \neq \frac{0}{0} \to \frac{2}{3}$$

2. $\lim_{x\to 0}\frac{e^x + e^{-x} - 2}{1 - \cos 2x}$; quotient has form $0/0$. By L' Hospital's rule we have $\lim_{x\to 0}\frac{e^x - e^{-x}}{2\sin 2x}$, which has indeterminate form $0/0$ again for 2nd time, so we apply L' Hospital's rule again

$$\lim_{x\to 0}\frac{e^x + e^{-x}}{4\cos 2x} = \frac{1}{2}.$$

3. $\lim_{x\to\infty}\frac{x^2}{\ln x} \equiv \frac{\infty}{\infty} \Rightarrow$ use L'Hospital , so $\lim_{x\to\infty}\frac{2x}{1/x} \to \infty$

4. $\lim\limits_{x\to\infty} \dfrac{e^{3x}}{\ln x} \equiv \dfrac{\infty}{\infty} \Rightarrow \lim\limits_{x\to\infty} 3xe^{3x} \to \infty$

5. $\lim\limits_{x\to\infty} x^2e^{-3x} \equiv 0.\infty$, so we convert to form ∞/∞ by writing $\lim\limits_{x\to\infty} \dfrac{x^2}{e^{3x}}$, and now use L'Hospital (differentiate twice), which gives $\lim\limits_{x\to\infty} \dfrac{2}{9e^{3x}} \to 0$

6. $\lim\limits_{x\to 0} \dfrac{\sin x}{x} \equiv \lim\limits_{x\to 0} \cos x \approx 1$

What is example 6. saying?

When x is very close to 0 then $\sin x \approx x$. That is $\sin x$ can be approximated with the function x for small values.

1.6 Taylor Series

Many functions are so complicated that it is not easy to see what they look like. If we only want to know what a function looks like *locally* , we can approximate it by simpler functions: polynomials. The crudest approximation is by a constant: if $f(x)$ is continuous at x_0,

$$f(x) \approx f(x_0)$$

for x near x_0.

Before we consider this in a more formal manner we start by looking at a simple motivating example:

Consider $f(x) = e^x$.

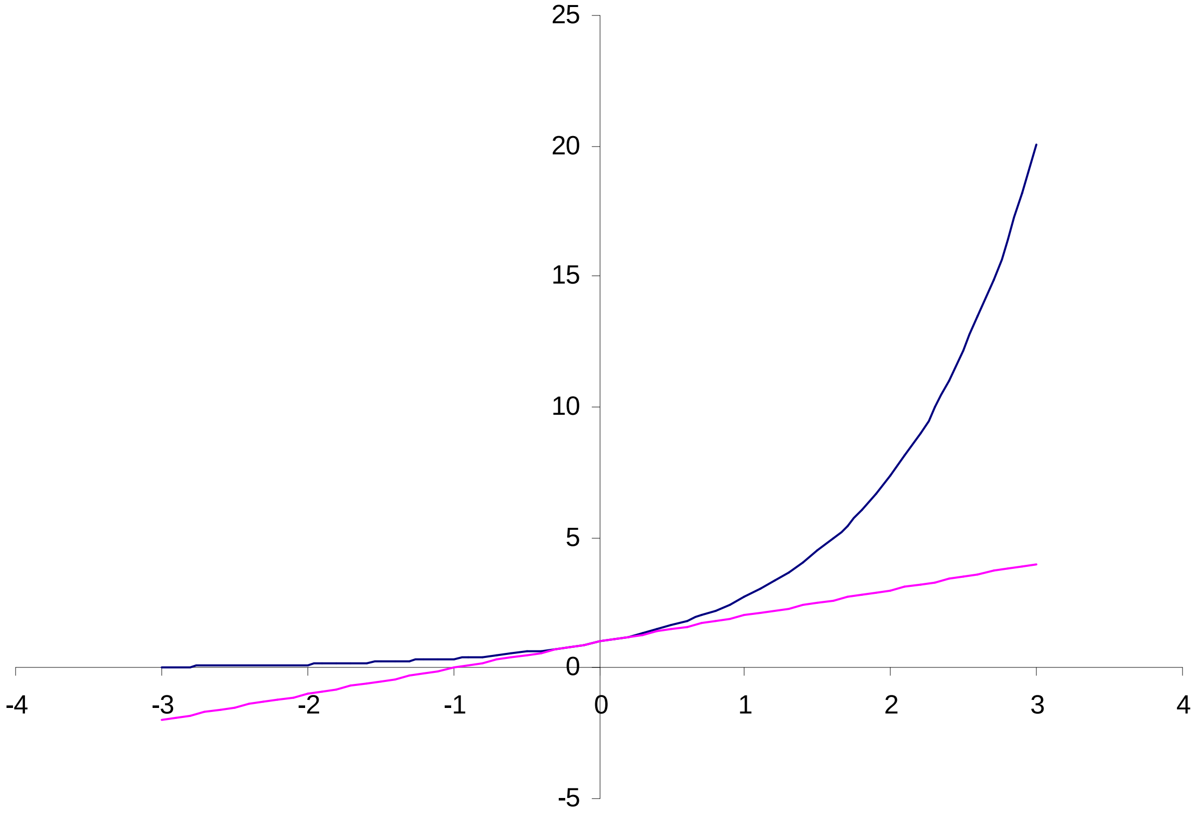
Suppose we wish to approximate this function for very small values of x (i.e. $x \longrightarrow 0$). We know at $x = 0$, $\frac{df}{dx} = 1$. So this is the gradient at $x = 0$. We can find the equation of the line that passes through a point (x_0, y_0) using

$$y - y_0 = m(x - x_0).$$

Here $m = \frac{df}{dx} = 1$, $x_0 = 0$, $y_0 = 1$, so $y = 1 + x$, is a polynomial. What information have we ascertained from this?

If $x \longrightarrow 0$ then the point $(x, 1 + x)$ on the tangent is close to the point (x, e^x) on the graph $f(x)$ and hence

$$e^x \approx 1 + x$$


25
20
15
10
5
0
-5
-4
-3
-2
-1
0
1
2
3
4

Suppose now that we are not that close to 0. We look for a second degree polynomial (i.e. quadratic)

$$g(x) = ax^2 + bx + c \longrightarrow g' = 2ax + b \longrightarrow g'' = 2a$$

If we want this parabola $g(x)$ to have

(i) same y intercept as f :

$$g(0) = f(0) \Longrightarrow c = 1$$

(ii) same tangent as f

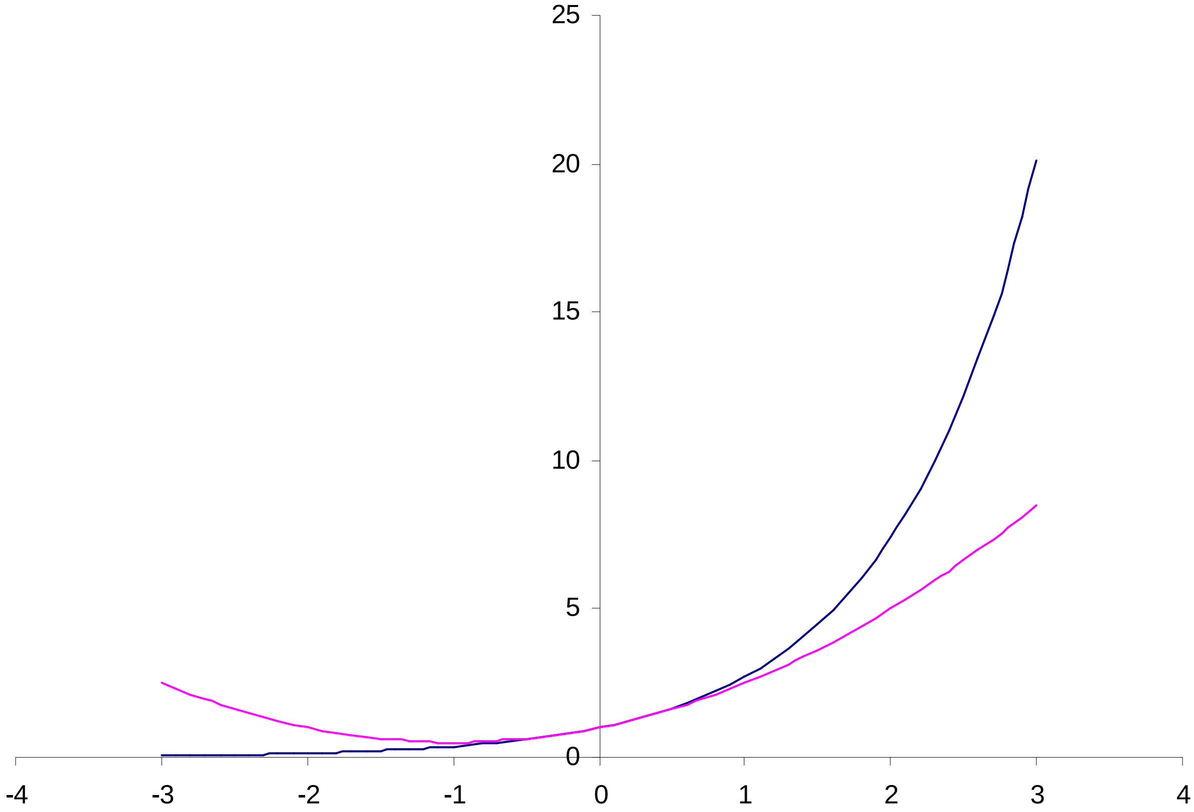
$$g'(0) = f'(0) \Longrightarrow b = 1$$

(iii) same curvature as f

$$g''(0) = f''(0) \Longrightarrow 2a = 1$$

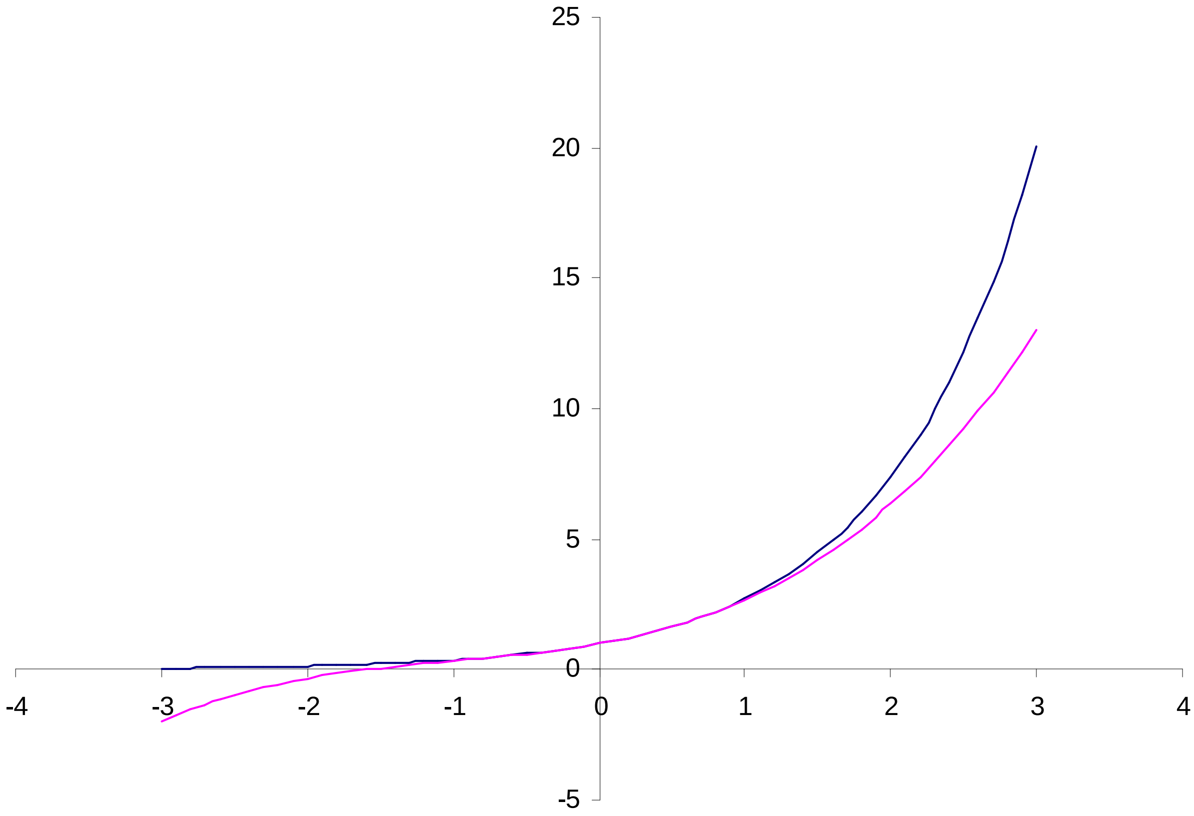
This gives

$$e^x \approx g(x) = \frac{1}{2}x^2 + x + 1$$


25
20
15
10
5
0
-4
-3
-2
-1
0
1
2
3
4

Moving further away we would look at a third order polynomial $h(x)$ which gives

$$e^x \approx h(x) = \frac{1}{3!}x^3 + \frac{1}{2!}x^2 + x + 1$$



and so on.

Better is to approximate by the tangent at x_0. This makes the approximation *and* its derivative agree with the function:

$$f(x) \approx f(x_0) + (x - x_0) f'(x_0).$$

Better still is by the best fit parabola (quadratic), which makes the first two derivatives agree:

$$f(x) \approx f(x_0) + (x - x_0) f'(x_0) + \frac{1}{2}(x - x_0)^2 f''(x_0).$$

This process can be continued indefinitely as long as f can be differentiated often enough.

The n^{th} term is

$$\frac{1}{n!}f^{(n)}(x_0)(x-x_0)^n,$$

where $f^{(n)}$ means the n^{th} derivative of f and $n! = n.(n-1)\ldots 2.1$ is the factorial.

$x_0 = 0$ is the special case, called *Maclaurin Series.*

Examples:

Expanding about the origin $x_0 = 0$,

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots + \frac{x^n}{n!}$$

Near 0, the logarithm looks like

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots + (-1)^n \frac{x^{n+1}}{(n+1)!}$$

How can we obtain this? Put $f(x) = \log(1+x)$, then $f(0) = 0$

$$\begin{array}{cc}
f'(x) = \frac{1}{1+x} & f'(0) = 1 \\
f''(x) = -\frac{1}{(1+x)^2} & f''(0) = -1 \\
f'''(x) = \frac{2}{(1+x)^3} & f'''(0) = 2 \\
f^{(4)}(x) = -\frac{6}{(1+x)^4} & f^{(4)}(0) = -6
\end{array}$$

Thus

$$\begin{aligned}
f(x) &= \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n \\
&= 0 + \frac{1}{1!}x + \frac{(-1)}{2!}x^2 + \frac{1}{3!}.2x^3 + \frac{(-6)}{4!}x^4 + \ldots.. \\
&= x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots
\end{aligned}$$

Taylor's theorem, in general, is this : If $f\ (x)$ and its first n derivatives exist (and are continuous) on some interval containing the point x_0 then

$$
\begin{aligned}
f(x) \;=\; & f(x_0) + \tfrac{1}{1!} f'(x_0)(x - x_0) + \\
& \tfrac{1}{2!} f''(x_0)(x - x_0)^2 + \ldots \\
& + \tfrac{1}{(n-1)!} f^{(n-1)}(x_0)(x - x_0)^{n-1} + R_n(x)
\end{aligned}
$$

where $R_n(x) = (1/n!)\, f^{\ (n)}(\xi)(x - x_0)^n$, ξ is some (usually unknown) number between x_0 and x and $f^{(n)}$ is the n^{th} derivative of f.

We can expand about any point $x = a$, and shift this point to the origin, i.e. $x - x_0 \equiv 0$ and we express in powers of $(x - x_0)^n$.

So for $f(x) = \sin x$ about $x = \pi/4$ we will have

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}\left(\frac{\pi}{4}\right)}{n!} (x - \pi/4)^n$$

where $f^{(n)}\left(\frac{\pi}{4}\right)$ is the n^{th} derivative of $\sin x$ at $x_0 = \pi/4$.

As another example suppose we wish to expand $\log(1+x)$ about $x_0 = 2$, i.e. $x - 2 = 0$ then

$$f(x) = \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(2)(x-2)^n$$

where $f^{(n)}(2)$ is the n^{th} derivative of $\log(1+x)$ evaluated at the point $x = 2$.

Note that $\log(1+x)$ does not exist for $x = -1$.

1.6.1 The Binomial Expansion

The *Binomial Expansion* is the Taylor expansion of $(1+x)^n$ where n is a positive integer. It reads:

$$(1+x)^n = 1+nx+\frac{n(n-1)}{2!}x^2+\frac{n(n-1)(n-2)}{3!}x^3+\dots .$$

We can extend this to expressions of the form

$$(1+ax)^n = 1+n(ax)+\frac{n(n-1)}{2!}(ax)^2+\frac{n(n-1)(n-2)}{3!}(ax)^3+\dots$$

$$(p+ax)^n = \left[p\left(1+\frac{a}{p}x\right)\right]^n = p^n\left[1+n\left(\frac{a}{p}x\right)+\dots\dots\right]$$

The binomial coefficients are found in Pascal's triangle:

$$
\begin{array}{ccccccccccccl}
 & & & & & 1 & & & & & & & \text{(n=0)} \ (1+x)^0 \\
 & & & & 1 & & 1 & & & & & & \text{(n=1)} \ (1+x)^1 \\
 & & & 1 & & 2 & & 1 & & & & & \text{(n=2)} \ (1+x)^2 \\
 & & 1 & & 3 & & 3 & & 1 & & & & \text{(n=3)} \ (1+x)^3 \\
 & 1 & & 4 & & 6 & & 4 & & 1 & & & \text{(n=4)} \ (1+x)^4 \\
1 & & 5 & & 10 & & 10 & & 5 & & 1 & & \text{(n=5)} \ (1+x)^5
\end{array}
$$

and so on ...

As an example consider:

$(1+x)^3 \quad n = 3 \Rightarrow 1 \quad 3 \quad 3 \quad 1 \quad \therefore (1+x)^3 = 1 + 3x + 3x^2 + x^3$

$(1+x)^5 \quad n = 5 \rightarrow (1+x)^5 = 1 + 5x + 10x^2 + 10x^3 + 5x^4 + x^5$.

If n is not an integer the theorem still holds but the coefficients are no longer integers. For example,

$$(1+x)^{-1} = 1 - x + x^2 - x^3 + \dots \, .$$

and

$$(1+x)^{1/2} = 1 + \frac{1}{2}x + \left(\frac{1}{2}\right)\left(-\frac{1}{2}\right)\frac{x^2}{2!} \dots \, .$$

Example: We looked at $\lim\limits_{x\to 0} \frac{\sin x}{x} \to 1$ (by L'Hospital).
We can also do this using Taylor series:

$$\begin{aligned}
\lim_{x\to 0} \frac{\sin x}{x} &\sim \lim_{x\to 0} \frac{x - x^3/3! + x^5/5! + \ldots.}{x} \\
&\sim \lim_{x\to 0} \left(1 - x^2/3! + x^4/5! + \ldots.\right) \\
&\to 1.
\end{aligned}$$

1.7 Integration

1.7.1 The Indefinite Integral

The indefinite integral of $f(x)$,

$$\int f(x)\,dx,$$

is any function $F(x)$ whose derivative equals $f(x)$. Thus if

$$F(x) = \int f(x)\,dx \qquad \text{then} \qquad \frac{dF}{dx}(x) = f(x).$$

Since the derivative of a constant, C, is zero $(dC/dx = 0)$, the indefinite integral of $f(x)$ is only determined up to an arbitrary constant;

if $\dfrac{dF}{dx} = f(x)$ then

$$\frac{d}{dx}\left(F(x) + C\right) = \frac{dF}{dx}(x) + \frac{dC}{dx} = \frac{dF}{dx}(x) = f(x).$$

Thus we must always include an arbitrary constant of integration in an indefinite integral.

Simple examples are

$$\begin{aligned}
\int x^n dx &= \frac{1}{n+1}x^{n+1} + C && (n \neq -1), \\
\int \frac{dx}{x} &= \log(x) + C, \\
\int e^{ax} dx &= \frac{1}{a}e^{ax} + C && (a \neq 0), \\
\int \cos ax dx &= \frac{1}{a}\sin ax + C \\
\int \sin ax dx &= -\frac{1}{a}\cos ax + C
\end{aligned}$$

Linearity

Integration is linear:

$$\int (\alpha f(x) + \beta g(x))\, dx = \alpha \int f(x)\, dx + \beta \int g(x)\, dx$$

for constants A and B. Thus, for example

$$\begin{aligned} \int \left(Ax^2 + Bx^3\right) dx &= A \int x^2 dx + B \int x^3 dx \\ &= \frac{A}{3}x^3 + \frac{B}{4}x^4 + C, \end{aligned}$$

$$\int (3e^x + 2/x)\, dx = 3 \int e^x dx + 2 \int \frac{dx}{x} = 3e^x + 2\log(x) + C,$$

and so forth.

1.7.2 The Definite Integral

The **definite integral,**

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

is the area under the graph of $f(x)$, between $x = a$ and $x = b$, with positive values of $f(x)$ giving positive area and negative values of $f(x)$ contributing negative area. It can be computed if the indefinite integral is known. For example

$$\int_1^3 x^3 dx = \left[\frac{1}{4}x^4\right]_1^3 = \frac{1}{4}\left(3^4 - 1^4\right) = 20,$$

$$\int_{-1}^1 e^x dx = [e^x]_{-1}^1 = e - 1/e.$$

Note that the definite integral is also linear in the sense that

$$\int_a^b (Af(x) + Bg(x))\, dx = A\int_a^b f(x)\, dx + B\int_a^b g(x)\, dx.$$

Note also that a definite integral

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

does not depend on the variable of integration, x in the above, it only depends on the function f and the limits of integration (a and b in this case); the area under a curve does not depend on what we choose to call the horizontal axis.

So

$$\int_a^b f(x)\, dx = \int_a^b f(y)\, dy = \int_a^b f(z)\, dz.$$

We should never confuse the variable of integration with the limits of integration; a definite integral of the form

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

is at best potentially confusing and at worst meaningless.

Consider the following

$$\mathrm{erf}\,(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-s^2} ds,$$

$$\mathrm{erf}\,c\,(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-s^2} ds,$$

$$N\,(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-s^2/2} ds.$$

Note that, by definition, if $a < b < c$ then

$$\int_a^c f(x)\,dx = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx.$$

By convention (which is not unreasonable if we think of a definite integral in terms of an area)

$$\int_c^a f(x)\,dx = -\int_a^c f(x)\,dx.$$

With this convention we find that for any a, b and c

$$\int_a^c f(x)\,dx = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx.$$

1.7.3 Integration by Substitution

This involves the change of variable and used to evaluate integrals of the form

$$\int g\left(f\left(x\right)\right)f'\left(x\right)dx,$$

and can be evaluated by writing $z = f\left(x\right)$ so that $dz/dx = f'\left(x\right)$ or $dz = f'\left(x\right)dx$. Then the integral becomes

$$\int g\left(z\right)dz.$$

For example:

$$\begin{aligned}
\int \frac{x}{1+x^2}dx &= \frac{1}{2}\int \frac{dz}{z} \\
&= \frac{1}{2}\log\left(z\right)+C = \frac{1}{2}\log\left(1+x^2\right)+C \\
&= \log\left(\sqrt{1+x^2}\right)+C
\end{aligned}$$

if we put $z = 1 + x^2$ so $dz = 2xdx$.

Similarly:

$$\begin{aligned}\int xe^{-x^2}dx &= -\frac{1}{2}\int e^z dz \\ &= -\frac{1}{2}e^z + C = -\frac{1}{2}e^{-x^2} + C\end{aligned}$$

this time with $z = -x^2$ so $dz = -2xdx$;

$$\begin{aligned}\int \frac{1}{x}\log(x)\,dx &= \int z\ dz = \frac{1}{2}z^2 + C \\ &= \frac{1}{2}(\log(x))^2 + C\end{aligned}$$

with $z = \log(x)$ so $dz = dx/x$ and

$$\begin{aligned}\int e^{x+e^x}dx &= \int e^x e^{e^x}dx = \int e^z dz \\ &= e^z + C = e^{e^x} + C\end{aligned}$$

with $z = e^x$ so $dz = e^x dx$.

The method can be used for definite integrals too. In this case it is usually more convenient to change the limits of integration at the same time as changing the variable; this is not strictly necessary, but it can save a lot of time.

For example, consider

$$\int_1^2 e^{x^2} 2x dx.$$

Write $z = x^2$, so $dz = 2xdx$. Now consider the limits of integration; when $x = 2$, $z = x^2 = 4$ and when $x = 1$, $z = x^2 = 1$. Thus

$$\begin{aligned} \int_{x=1}^{x=2} e^{x^2} 2x dx &= \int_{z=1}^{z=4} e^z dz \\ &= [e^z]_{z=1}^{z=4} = e^4 - e^1. \end{aligned}$$

Further examples: consider

$$\int_{x=1}^{x=2} \frac{2x dx}{1+x^2}.$$

In this case we could write $z = 1 + x^2$, so $dz = 2x dx$ and $x = 1$ corresponds to $z = 2$, $x = 2$ corresponds to $z = 5$, and

$$\begin{aligned} \int_{x=1}^{x=2} \frac{2x}{1+x^2} dx &= \int_{z=2}^{z=5} \frac{dz}{z} \\ &= [\ln(z)]_{z=2}^{z=5} = \log(5) - \ln(2) \\ &= \ln(5/2) \end{aligned}$$

We can solve the same problem without change of limit, i.e.

$$\left\{ \ln \left| 1 + x^2 \right| \right\}_{x=1}^{x=2} \longrightarrow \ln 5 - \ln 2 = \ln 5/2.$$

Or consider

$$\int_{x=1}^{x=e} 2\frac{\log(x)}{x}dx$$

in which case we should choose $z = \log(x)$ so $dz = dx/x$ and $x = 1$ gives $z = 0$, $x = e$ gives $z = 1$ and so

$$\int_{x=1}^{x=e} 2\frac{\log(x)}{x}dx = \int_{z=0}^{z=1} 2z\,dz = \left[z^2\right]_{z=0}^{z=1} = 1.$$

When we make a substitution like $z = f(x)$ we are implicitly assuming that $dz/dx = f'(x)$ is neither infinite nor zero. It is important to remember this implicit assumption.

Consider the integral

$$\int_{-1}^{1} x^2 dx = \frac{1}{3}\left[x^3\right]_{x=-1}^{x=1} = \frac{1}{3}(1-(-1)) = \frac{2}{3}.$$

Now put $z = x^2$ so $dz = 2x dx$ or $dz = 2\sqrt{z}\, dx$ and when $x = -1$, $z = x^2 = 1$ and when $x = 1$, $z = x^2 = 1$, so

$$\int_{x=-1}^{x=1} x^2 dx = \frac{1}{2}\int_{z=1}^{z=1} \frac{dz}{\sqrt{z}} = 0$$

as the area under the curve $1/\sqrt{z}$ between $z = 1$ and $z = 1$ is obviously zero.

It is clear that $x^2 > 0$ except at $x = 0$ and therefore that

$$\int_{-1}^{1} x^2 dx = \frac{2}{3}$$

must be the correct answer. The substitution $z = x^2$ gave

$$\int_{x=-1}^{x=1} x^2 dx = \frac{1}{2}\int_{z=1}^{z=1} \frac{dz}{\sqrt{z}} = 0$$

which is obviously wrong. So why did the substitution fail?

It failed because $f'(x) = dz/dx = 2x$ changed signs between $x = -1$ and $x = 1$. In particular, $dz/dx = 0$ at $x = 0$, the function $z = x^2$ is not invertible for $-1 \leq x \leq 1$.

Moral: when making a substitution make sure that $dz/dx \neq 0$.

Earlier we saw the definition of the **CDF** for the Normal Distribution

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-s^2/2} ds$$

If $x \longrightarrow \infty$ then we know (by the fact that the area under a PDF has to sum to unity) that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-s^2/2} ds = 1.$$

This can be used to obtain an important result. First we make the substitution $x = s/\sqrt{2}$ to give $dx = ds/\sqrt{2}$, hence the integral becomes

$$\sqrt{2} \int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{2\pi}$$

and hence we obtain

$$\begin{aligned} \int_{-\infty}^{\infty} e^{-x^2} dx &= \sqrt{\pi} \Longrightarrow \\ \int_{0}^{\infty} e^{-x^2} dx &= \frac{\sqrt{\pi}}{2}. \end{aligned}$$

1.7.4 Integration by Parts

This is based on the product rule. In usual notation, if $y = u(x)\,v(x)$ then

$$\frac{dy}{dx} = \frac{du}{dx}v + u\frac{dv}{dx}$$

so that

$$\frac{du}{dx}v = \frac{dy}{dx} - u\frac{dv}{dx}$$

and hence integrating

$$\int \frac{du}{dx}vdx = \int \frac{dy}{dx}dx - \int u\frac{dv}{dx}dx = y(x) - \int u\frac{dv}{dx}dx + C$$

or

$$\int \frac{du}{dx}vdx = u(x)\,v(x) - \int u(x)\,\frac{dv}{dx}dx + C$$

i.e.

$$\int u'v dx = uv - \int uv' dx + C$$

This is useful, for instance, if $v(x)$ is a polynomial and $u(x)$ is an exponential.

How can we use this formula? Consider the example

$$\int xe^x dx$$

Put

$$\begin{array}{cc} v = x & u' = e^x \\ v' = 1 & u = e^x \end{array}$$

hence

$$\begin{aligned} \int xe^x dx &= uv - \int u \frac{dv}{dx} dx \\ &= xe^x - \int e^x .1 dx = e^x (x-1) + C \end{aligned}$$

The formula we are using is the same as

$$\int v du = uv - \int u dv + C$$

Now using the same example $\int xe^x dx$

$$\begin{array}{cc} v = x & du = e^x dx \\ dv = dx & u = e^x \end{array}$$

and

$$\begin{aligned}\int v du &= uv - \int u dv = xe^x - \int e^x dx \\ &= e^x (x-1) + C\end{aligned}$$

Another example

$$\int \underbrace{x^2}_{v(x)} \underbrace{e^{2x}}_{u'} dx = \underbrace{\frac{1}{2} x^2 e^{2x}}_{uv} - \int \underbrace{x e^{2x}}_{uv'} dx + C$$

and using integration by parts again

$$\int x e^{2x} dx = \frac{1}{2} x e^{2x} - \frac{1}{2} \int e^{2x} dx = \frac{1}{4} (2x-1) e^{2x} + D$$

so

$$\int x^2 e^{2x} dx = \frac{1}{4} \left(2x^2 - 2x + 1\right) e^{2x} + E.$$

Important Example:

$$\int e^x \cos x dx$$

so set $I = \int e^x \cos x dx$. Now put

$$\begin{array}{cc} v = e^x & u' = \cos x \\ v' = e^x & u = \sin x \end{array}$$

which gives

$$I = e^x \sin x - \int e^x \sin x dx$$

need to obtain $\int e^x \sin x dx$ for a second time by parts so put

$$\begin{array}{cc} v = e^x & u' = \sin x \\ v' = e^x & u = -\cos x \end{array}$$

and we have

$$\int e^x \sin x dx = -e^x \cos x + \underbrace{\int e^x \cos x dx}_{I}$$

so putting together with the earlier integral

$$\begin{aligned} I &= e^x \sin x - (-e^x \cos x + I) \\ 2I &= e^x (\sin x + \cos x) \end{aligned}$$

hence

$$\int e^x \cos x dx = \frac{e^x}{2} (\sin x + \cos x) + C$$

1.7.5 Other Results

$$\int \frac{f'(x)}{f(x)}dx = \ln|f(x)| + C$$

e.g.

$$\int \frac{3}{1+3x}dx = \ln|1+3x| + C$$

$$\int \frac{1}{2+7x}dx = \frac{1}{7}\int \frac{7}{2+7x}dx = \frac{1}{7}\ln|2+7x| + C$$

This allows us to state a standard result

$$\int \frac{1}{a+bx}dx = \frac{1}{b}\ln|a+bx| + C$$

How can we re-do the earlier example

$$\int \frac{x}{1+x^2}dx,$$

which was initially treated by substitution? We note that we can write this integral as

$$\begin{aligned}\frac{1}{2}\int \frac{2x}{1+x^2}dx &= \frac{1}{2}\ln\left|1+x^2\right| + C \\ &= \ln\sqrt{\left|1+x^2\right|} + C\end{aligned}$$

1.7.6 Partial Fractions

Consider a fraction where both numerator and denominator are polynomial functions, i.e.

$$h\ (x) = \frac{f(x)}{g(x)} \equiv \frac{\sum\limits_{n=0}^{N} a_n x^n}{\sum\limits_{n=0}^{M} b_n x^n}$$

where deg $f(x) <$ deg $g(x)$, i.e. $N < M$. Then $h\ (x)$ is called a *partial fraction.* Suppose

$$\frac{c}{(x+a)(x+b)} \equiv \frac{A}{(x+a)} + \frac{B}{(x+b)}$$

then writing

$$c = A(x+b) + B(x+a)$$

and solving for A and B allows us to obtain partial fractions.

The simplest way to achieve this is by setting $x = -b$ to obtain the value of B, then putting $x = -a$ yields A.

Example: $\dfrac{1}{(x-2)(x+3)}$. Now write

$$\frac{1}{(x-2)(x+3)} \equiv \frac{A}{x-2} + \frac{B}{x+3}$$

which becomes

$$1 = A(x+3) + B(x-2)$$

Setting $x = -3 \rightarrow B = -1/5$; $\quad x = 2 \rightarrow A = 1/5$.
So

$$\frac{1}{(x-2)(x+3)} \equiv \frac{1}{5(x-2)} - \frac{1}{5(x+3)}.$$

There is another quicker and simpler method to obtain partial fractions, called the "*cover-up*" rule. As an example consider

$$\frac{x}{(x-2)(x+3)} \equiv \frac{A}{x-2} + \frac{B}{x+3}.$$

Firstly, look at the term $\dfrac{A}{x-2}$. The denominator vanishes for $x = 2$, so take the expression on the LHS and "cover-up" $(x-2)$. Now evaluate the remaining expression, i.e. $\dfrac{x}{(x+3)}$ for $x = 2$, which gives 2/5. So $A = 2/5$.

Now repeat this, by noting that $\dfrac{B}{x+3}$ does not exist at $x = -3$. So cover up $(x+3)$ on the LHS and evaluate $\dfrac{x}{(x-2)}$ for $x = -3$, which gives $B = 3/5$.

Any rational expression $\dfrac{f(x)}{g(x)}$ (with degree of $f(x)$ $<$ degree of $g(x)$) such as above can be written

$$\frac{f(x)}{g(x)} \equiv F_1 + F_2 + \ \ldots\ldots\ldots \ + F_k$$

where each F_i has form

$$\frac{A}{(px+q)^m} \quad \text{or} \quad \frac{Cx+D}{\left(ax^2+bx+c\right)^n}$$

where $\dfrac{A}{(px+q)^m}$ is written as

$$\frac{A_1}{(px+q)} + \frac{A_2}{(px+q)^2} + \ \ldots\ldots \ + \frac{A}{(px+q)^m}$$

and $\dfrac{Cx+D}{\left(ax^2+bx+c\right)^n}$ becomes

$$\frac{C_1x+D_1}{ax^2+bx+c} + \ \ldots\ldots + \frac{C_nx+D_n}{\left(ax^2+bx+c\right)^n}$$

Examples:

$$\frac{3x-2}{(4x-3)(2x+5)^3} \equiv \frac{A}{4x-3} + \frac{B}{2x+5} + \frac{C}{(2x+5)^2} + \frac{D}{(2x+5)^3}$$

$$\frac{4x^2+13x-9}{x(x+3)(x-1)} \equiv \frac{A}{x} + \frac{B}{x+3} + \frac{C}{(x-1)}$$

$$\frac{3x^3-18x^2+29x-4}{(x+1)(x-2)^3} \equiv \frac{A}{x+1} + \frac{B}{x-2} + \frac{C}{(x-2)^2} + \frac{D}{(x-2)^3}$$

$$\frac{5x^2-x+2}{\left(x^2+2x+4\right)^2(x-1)} \equiv \frac{Ax+B}{x^2+2x+4} + \frac{Cx+D}{\left(x^2+2x+4\right)^2} + \frac{E}{x-1}$$

$$\frac{x^2-x-21}{\left(x^2+4\right)^2(2x-1)} \equiv \frac{Ax+B}{x^2+4} + \frac{Cx+D}{\left(x^2+4\right)^2} + \frac{E}{2x-1}$$

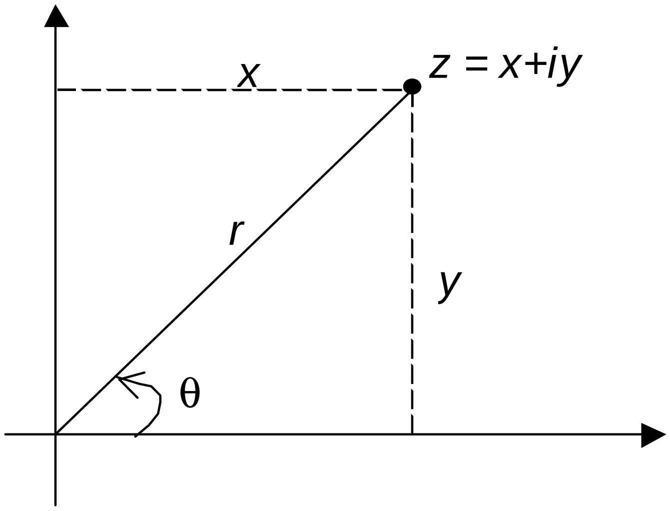
1.8 Complex Numbers

A complex number z is defined by $z = x + iy$ where $x,\ y \in \mathbb{R}$ and $i = \sqrt{-1}$. It follows that $i^2 = -1$.

A complex number z may also be expressed in polar co-ordinate form as

$$z = r\,(\cos\theta + i\sin\theta)$$

where r is always positive and θ counter-clockwise from Ox. So $x = r\cos\theta$, $y = r\sin\theta$



So

$$x = r\cos\theta,\ \ y = r\sin\theta;\ \ \ r = +\sqrt{x^2 + y^2},\ \ \theta = \arctan\frac{y}{x}$$

We call the $x-$ axis the real line and the $y-$ axis the imaginary line.

The set of all complex numbers is denoted $\mathbb{C}$, and for any complex number z we write $z \in \mathbb{C}$. We can think of $\mathbb{R} \subset \mathbb{C}$.

1.8.1 Arithmetic

Given any two complex numbers $z_1 = a + ib, \quad z_2 = c + id$ the following definitions hold:

Addition & Subtraction $\quad z_1 \pm z_2 = (a \pm c) + i\,(b \pm d)$

Multiplication $\quad z_1 \times z_2 = (ac - bd) + i\,(ad + bc)$

Division $\quad \dfrac{z_1}{z_2} = \dfrac{a + ib}{c + id} = \dfrac{(ac + bd) + i\,(bc - ad)}{c^2 + d^{\,2}}$

here we have simply multiplied by $\dfrac{c - id}{c - id}$ and note that $(c + id)\,(c - id) = c^2 + d^2$

Examples

$z_1 = 1 + 2i, \quad z_2 = 3 - i$

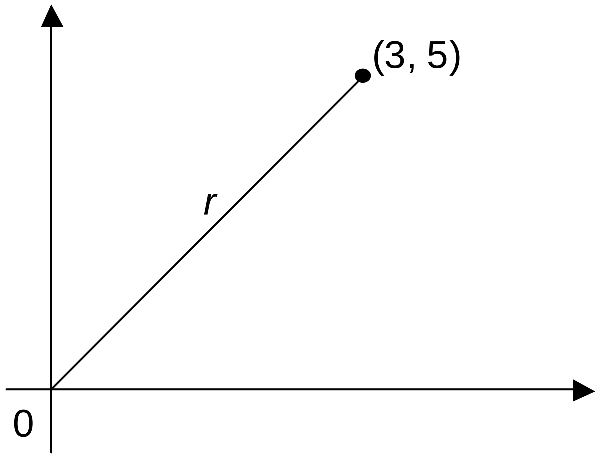
$z_1 + z_2 = (1 + 3) + i\,(2 - 1) = 4 + i$; $\quad z_1 + z_2 = (1 - 3) + i\,(2 - (-1)) = -2 + 3i$

$$z_1 \times z_2 = (1.3 - 2. - 1) + i\,(1. - 1 + 2.3) = 5 + 5i$$

$$\frac{z_1}{z_2} = \frac{1+2i}{3-i}.\frac{3+i}{3+i} = \frac{1+7i}{10}$$

1.8.2 Modulus and Argument

Given $z = x + iy$, the *modulus* of z denoted $|z|$ is defined $|z| = r = +\sqrt{x^2 + y^2}$ (as given earlier). So we are using Pythagoras to calculate the length of point joining the origin to the point $z\,(x,\ y)$. As an example, consider the complex number $z = 3 + 5i$, which is represented in the $x - y$ plane (first quadrant) by the point $(3,\ 5)$. The modulus can be calculated from $|3 + 5i| = \sqrt{3^2 + 5^2}$, to give $r = \sqrt{34}$.



1.8.3 Complex Conjugate

We define *complex conjugate* of z by $\overline{z}$ where

$$\overline{z} = x - iy.$$

$\overline{z}$ is the reflection of z in the real line. So for example if $z = 1 - 2i$, then $\overline{z} = 1 + 2i$

1. $\overline{(\overline{z})} = z$

2. $\overline{(z_1 + z_2)} = \overline{z_1} + \overline{z}_2$

3. $\overline{(z_1 z_2)} = \overline{z}_1 \overline{z}_2$

4. $z+\overline{z} \;= 2x = 2\,\mathsf{Re}\,z \qquad \Rightarrow \mathsf{Re}\,z = \dfrac{z+\overline{z}}{2}$

5. $z-\overline{z} \;= 2iy = 2i\,\mathsf{Im}\,z \qquad \Rightarrow \mathsf{Im}\,z = \dfrac{z-\overline{z}}{2i}$

6. $z.\overline{z} \;= (x+iy)\,(x-iy) = |z|^2$

7. $|\overline{z}|^2 = \overline{z}\overline{(\overline{z})} = \overline{z}z = |z|^2 \qquad \Rightarrow |\overline{z}| = |z|$

8. $\dfrac{z_1}{z_2} = \dfrac{z_1}{z_2}.\dfrac{\overline{z}_2}{\overline{z}_2} = \dfrac{z_1\overline{z}_2}{|\overline{z}_2|^2}$

9. $|z_1 z_2|^2 = |z_1|^2\,|z_2|^2$

1.8.4 Polar Form:

We return to the polar form representation of complex numbers. We now introduce a new notation. If $z \in \mathbb{C}$, then

$$z = r\,(\cos\theta + i\sin\theta) = re^{i\theta}.$$

Hence

$$e^{i\theta} = \cos\theta + i\sin\theta,$$

which is a special relationship called *Eulers Identity.* Knowing $\sin\theta$ is an odd function gives $e^{-i\theta} = \cos\theta - i\sin\theta$. Referring to the earlier figure, we have:

$$|z| = r, \quad \arg z = \theta$$

If

$$z_1 = r_1 e^{i\theta_1} \quad \text{and} \quad z_2 = r_2 e^{i\theta_2}$$

then

$$\begin{aligned} z_1 z_2 &= r_1 r_2 e^{i(\theta_1+\theta_2)} \Rightarrow |z_1 z_2| = r_1 r_2 = |z_1|\,|z_2| \\ \arg(z_1 z_2) &= \theta_1 + \theta_2 = \arg(z_1) + \arg(z_2). \end{aligned}$$

If $z_2 \neq 0$ then

$$\frac{z_1}{z_2} = \frac{r_1 e^{i\theta_1}}{r_2 e^{i\theta_2}} = \frac{r_1}{r_2} e^{i(\theta_1-\theta_2)}$$

and hence

$$\begin{aligned} \left|\frac{z_1}{z_2}\right| &= \frac{|z_1|}{|z_2|} = \frac{r_1}{r_2} \\ \arg\left(\frac{z_1}{z_2}\right) &= \theta_1 - \theta_2 = \arg(z_1) - \arg(z_2) \end{aligned}$$

Eulers Formula: Let θ be any angle, then

$$\exp(i\theta) = \cos\theta + i\sin\theta.$$

We can prove this by considering the Taylor series for $\exp(x)$, $\sin x$, $\cos x$

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots\ldots\ldots\ldots + \frac{x^n}{n!} \qquad \text{(a)}$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \ldots\ldots\ldots\ldots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} \quad \text{(b)}$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \ldots\ldots\ldots\ldots + (-1)^n \frac{x^{2n}}{(2n)!} \quad \text{(c)}$$

Replacing x by the purely imaginary quantity $i\theta$ in (a), we obtain

$$\begin{aligned} e^{i\theta} &= 1 + i\theta + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \ldots\ldots\ldots\ldots + \frac{(i\theta)^n}{n!} \\ &= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \ldots\ldots\ldots\ldots\right) + \\ &\quad\; i\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \ldots\ldots\ldots\right) \\ &= \cos\theta + i\sin\theta \end{aligned}$$

Note: When $\theta = \pi$ then $\exp i\pi = -\mathbf{1}$ and $\theta = \pi/2$ gives $\exp(i\pi/2) = i$.

1.9 Functions of Several Variables

A function can depend on more than one variable. For example, the value of an option depends on the underlying asset price S (for 'spot' or 'share') and time t. We can write its value as $V(S, t)$.

The value also depends on other parameters such as the exercise price E, interest rate r and so on. Although we could write $V(S, t, E, r, ...)$, it is usually clearer to leave these other variables out.

Depending on the application, the independent variables may be x and t for space and time, or two space variables x and y, or S and t for price and time, and so on.

1.9.1 Partial Derivatives

Consider a function $z = f(x, y)$, which can be thought of as a surface in x, y, z space. We can think of x and y as positions on a two dimensional grid (or as spacial variables) and z as the height of a surface above the (x, y) grid.

How do we differentiate a function $f(x, y)$ of *two* variables? What if there are more independent variables?

The **partial derivative** of $f(x, y)$ with respect to x is written

$$\frac{\partial f}{\partial x}$$

(note ∂ and not d). It is the $x-$ derivative of f *with y held fixed:*

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

The other partial derivative, $\partial f/\partial y$, is defined similarly but now x is held fixed:

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

$$\frac{\partial f}{\partial x} \qquad \text{and} \qquad \frac{\partial f}{\partial y}$$

are sometimes written as f_x and f_y.

Examples

If

$$f(x, t) = x + t^2 + xe^{-t^2}$$

then

$$\frac{\partial f}{\partial x} = f_x = 1 + 0 + 1 \cdot e^{-t^2}$$

$$\frac{\partial f}{\partial t} = f_t = 0 + 2t + x \cdot (-2t)\, e^{-t^2}.$$

The convention is, treat the other variable like a constant.

Let $z = x^3 y^2 + \sin xy$ then

$$z_x = 3x^2y^2 + y\cos xy, \quad z_y = 2x^3y + x\cos xy$$

1.9.2 Higher Derivatives

Like ordinary derivatives, these are defined recursively:

$$\begin{aligned}
\frac{\partial^2 f}{\partial x^2} &= f_{xx} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right), \\
\frac{\partial^2 f}{\partial x \partial y} &= f_{xy} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right), \\
\frac{\partial^2 f}{\partial y \partial x} &= f_{yx} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right),
\end{aligned}$$

and

$$\frac{\partial^2 f}{\partial y^2} = f_{yy} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right).$$

If f is well-behaved, the 'mixed' partial derivatives are equal:

$$f_{xy} = f_{yx}.$$

i.e. the second order derivatives exist and are continuous. In the previous example we have

$$\begin{aligned} z_{xx} &= 6xy^2 - y^2 \sin xy, \quad z_{yy} = 2x^3 - x^2 \sin xy, \\ z_{xy} &= 6x^2y + \cos xy - xy \sin xy = z_{yx}. \end{aligned}$$

Examples:

With $f\ (x,t) = x + t^2 + xe^{-t^2}$ as above,

$$f_x = 1 + e^{-t^2}$$

so

$$f_{xx} = 0; \quad f_{xt} = -2te^{-t^2}$$

Also

$$f_t = 2t - 2xte^{-t^2}$$

so

$$f_{tx} = -2te^{-t^2}; \quad f_{tt} = 2 - 2xe^{-t^2} + 4xt^2e^{-t^2}$$

Note that $f_{xt} = f_{tx}$.

1.9.3 The Chain Rule I

Suppose that $x = x(s)$ and $y = y(s)$ and $F(s) = f\ (x(s), y(s))$. Then

$$\frac{dF}{ds}(s) = \frac{dx}{ds}(s)\frac{\partial f}{\partial x}(x(s), y(s)) + \frac{dy}{ds}(s)\frac{\partial f}{\partial y}(x(s), y(s))$$

Thus if $f(x,y) = x^2 + y^2$ and $x(s) = \cos(s)$, $y(s) = \sin(s)$ we find that $F(s) = f(x(s), y(s))$ has derivative

$$\frac{dF}{ds} = -\sin(s)\cdot 2\cos(s) + \cos(s)\cdot 2\sin(s) = 0$$

which is what it should be, since $F(s) = \cos^2(s) + \sin^2(s) = 1$,

i.e. a constant.

Example: Calculate $\dfrac{dz}{dt}$ at $t = \pi/2$ where

$$z = \exp\left(xy^2\right) \qquad x = t\cos t, \quad y = t\sin t.$$

Chain rule gives

$$\begin{aligned} \frac{dz}{dt} &= \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} \\ &= y^2 \exp\left(xy^2\right)(-t\sin t + \cos t) + \\ &\quad 2xy\exp\left(xy^2\right)(\sin t + t\cos t)\,. \end{aligned}$$

At $t = \pi/2$ $x = 0$, $y = \pi/2 \Rightarrow \left.\dfrac{dz}{dt}\right|_{t=\pi/2} = -\dfrac{\pi^3}{8}.$

1.9.4 The Chain Rule II

Suppose that $x = x(u, v)$, $y = y(u, v)$ and that $F(u, v) = f\ (x(u, v),\ y(u, v))$. Then

$$\frac{\partial F}{\partial u} = \frac{\partial x}{\partial u}\frac{\partial f}{\partial x} + \frac{\partial y}{\partial u}\frac{\partial f}{\partial y} \quad \text{and} \quad \frac{\partial F}{\partial v} = \frac{\partial x}{\partial v}\frac{\partial f}{\partial x} + \frac{\partial y}{\partial v}\frac{\partial f}{\partial y}.$$

This is sometimes written as

$$\frac{\partial}{\partial u} = \frac{\partial x}{\partial u}\frac{\partial}{\partial x} + \frac{\partial y}{\partial u}\frac{\partial}{\partial y}, \qquad \frac{\partial}{\partial v} = \frac{\partial x}{\partial v}\frac{\partial}{\partial x} + \frac{\partial y}{\partial v}\frac{\partial}{\partial y}.$$

so is essentially a differential operator.

Example:

$$T = x^3 - xy + y^3 \quad \text{where} \quad x = r\cos\theta, \quad y = r\sin\theta$$

$$\begin{aligned}
\frac{\partial T}{\partial r} &= \frac{\partial T}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial T}{\partial y}\frac{\partial y}{\partial r} = \cos\theta\left(3x^2 - y\right) + \sin\theta\left(3y^2 - x\right) \\
&= \cos\theta\left(3r^2\cos^2\theta - r\sin\theta\right) + \\
&\quad \sin\theta\left(3r^2\sin^2\theta - r\cos\theta\right) \\
&= 3r^2\left(\cos^3\theta + \sin^3\theta\right) - 2r\cos\theta\sin\theta \\
&= 3r^2\left(\cos^3\theta + \sin^3\theta\right) - r\sin 2\theta.
\end{aligned}$$

$$\begin{aligned}
\frac{\partial T}{\partial \theta} &= \frac{\partial T}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial T}{\partial y}\frac{\partial y}{\partial \theta} \\
&= -r\sin\theta\left(3x^2 - y\right) + r\cos\theta\left(3y^2 - x\right) \\
&= -r\sin\theta\left(3r^2\cos^2\theta - r\sin\theta\right) + \\
&\quad r\cos\theta\left(3r^2\sin^2\theta - r\cos\theta\right) \\
&= 3r^3\cos\theta\sin\theta\,(\sin\theta - \cos\theta) + \\
&\quad r^2\left(\sin^2\theta - \cos^2\theta\right). \\
&= r^2(\sin\theta - \cos\theta)(3r\cos\theta\sin\theta + \sin\theta + \cos\theta)
\end{aligned}$$

1.9.5 Extensions

If $x = x(u, v, w)$, $y = y(u, v, w)$ and $F(u, v, w) = f(x(u, v, w),\ y(u, v, w))$ then

$$\begin{aligned}
\frac{\partial F}{\partial u} &= \frac{\partial x}{\partial u}\frac{\partial f}{\partial x} + \frac{\partial y}{\partial u}\frac{\partial f}{\partial y} \\
\frac{\partial F}{\partial v} &= \frac{\partial x}{\partial v}\frac{\partial f}{\partial x} + \frac{\partial y}{\partial v}\frac{\partial f}{\partial y} \\
\frac{\partial F}{\partial w} &= \frac{\partial x}{\partial w}\frac{\partial f}{\partial x} + \frac{\partial y}{\partial w}\frac{\partial f}{\partial y}
\end{aligned}$$

If $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$ and $F(u, v) = f\ (x(u, v),\ y(u, v),\ z(u, v))$ then

$$\begin{aligned}
\frac{\partial F}{\partial u} &= \frac{\partial x}{\partial u}\frac{\partial f}{\partial x} + \frac{\partial y}{\partial u}\frac{\partial f}{\partial y} + \frac{\partial z}{\partial u}\frac{\partial f}{\partial z} \\
\frac{\partial F}{\partial v} &= \frac{\partial x}{\partial v}\frac{\partial f}{\partial x} + \frac{\partial y}{\partial v}\frac{\partial f}{\partial y} + \frac{\partial z}{\partial v}\frac{\partial f}{\partial z}
\end{aligned}$$

So we can generalise this to obtain a chain rule for

$$F(x_1, x_2, \dots\dots x_m) =$$
$$f\left(\begin{array}{c} X_1(x_1, x_2, \dots\dots x_m),\ X_2(x_1, x_2, \dots\dots x_m),\ \dots\dots \\ X_n(x_1, x_2, \dots\dots x_m) \end{array}\right)$$

by the result

$$\begin{aligned}
\frac{\partial F}{\partial x_1} &= \frac{\partial X_1}{\partial x_1}\frac{\partial f}{\partial X_1} + \frac{\partial X_2}{\partial x_1}\frac{\partial f}{\partial X_2} + \dots\dots\dots + \frac{\partial X_n}{\partial x_1}\frac{\partial f}{\partial X_n} \\
\frac{\partial F}{\partial x_2} &= \frac{\partial X_1}{\partial x_2}\frac{\partial f}{\partial X_1} + \frac{\partial X_2}{\partial x_2}\frac{\partial f}{\partial X_2} + \dots\dots\dots + \frac{\partial X_n}{\partial x_2}\frac{\partial f}{\partial X_n} \\
&\vdots \\
&\vdots \\
\frac{\partial F}{\partial x_m} &= \frac{\partial X_1}{\partial x_m}\frac{\partial f}{\partial X_1} + \frac{\partial X_2}{\partial x_m}\frac{\partial f}{\partial X_2} + \dots\dots\dots + \frac{\partial X_n}{\partial x_m}\frac{\partial f}{\partial X_n}.
\end{aligned}$$

Naturally this can be written in a more compact (and pedantic) way

$$\frac{\partial F}{\partial x_i} = \sum_{j=1}^{n} \frac{\partial X_j}{\partial x_i}\frac{\partial f}{\partial X_j}, \quad i = 1, \dots, m$$

1.9.6 Taylor for two Variables

Assuming that a function $f(x,t)$ is differentiable enough, near $x = x_0$, $t = t_0$,

$$\begin{aligned} f(x,t) &= f(x_0,t_0) + (x - x_0) f_x(x_0,t_0) + \\ & (t - t_0) f_t(x_0,t_0) \\ & + \frac{1}{2}\left[\begin{array}{c} (x - x_0)^2 f_{xx}(x_0,t_0) \\ +2(x - x_0)(t - t_0) f_{xt}(x_0,t_0) \\ +(t - t_0)^2 f_{tt}(x_0,t_0) \end{array}\right] + \ \ldots. \end{aligned}$$

That is,

$$\begin{aligned} f\ (x,t) &= \text{constant} + \ \text{linear} + \text{quadratic} \\ & + \ldots. \end{aligned}$$

The error in truncating this series after the second order terms tends to zero faster than the included terms. This result is particularly important for Itô's lemma in Stochastic Calculus.

Suppose a function $f = f(x, y)$ and both x, y change by a small amount, so $x \longrightarrow x + \delta x$ and $y \longrightarrow y + \delta y$, then we can examine the change in f using a two dimensional form of Taylor

$$\begin{aligned} f(x + \delta x, y + \delta y) \quad = \quad & f(x, y) + f_x \delta x + f_y \delta y + \\ & \frac{1}{2} f_{xx} \delta x^2 + \frac{1}{2} f_{yy} \delta y^2 + \\ & f_{xy} \delta x \delta y + O\left(\delta x^3, \delta y^3\right). \end{aligned}$$

By taking $f(x, y)$ to the lhs, writing

$$df = f(x + \delta x, y + \delta y) - f(x, y)$$

and considering only linear terms, i.e.

$$df = \frac{\partial f}{\partial x} \delta x + \frac{\partial f}{\partial y} \delta y$$

we obtain a formula for the *differential* or *total change* in f.

1.10 Special Functions

1.10.1 The Gamma Function

The Gamma Function $\Gamma(x)$ is defined as

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt \quad (x > 0)$$

Note $\int_0^\infty e^{-t} dt = 1$

Integration by parts gives us $\int_0^\infty e^{-t} t^x dt = \Gamma(x+1) =$

$$\begin{aligned} x\int_0^\infty e^{-t} t^{x-1} dt &= x(x-1)\int_0^\infty e^{-t} t^{x-2} dt \quad (1) \\ &= \ldots\ldots.. = x! \end{aligned}$$

Important results:

$$\begin{aligned} \Gamma(n+1) &= n! \ (n \geq 0) \\ \Gamma(1) &= 1 \end{aligned}$$

and also from (1)

$$\Gamma(x+1) = x\Gamma(x).$$

If we make the substitution $t = u^2$ in $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt$ we obtain

$$\Gamma(x) = 2\int_0^\infty e^{-u^2} u^{2x-1} du$$

and put $x = 1/2$ so that

$$\Gamma\left(\tfrac{1}{2}\right) = 2\int_0^\infty e^{-u^2} du$$

and we know from the error function that $\int_0^\infty e^{-u^2} du = \sqrt{\pi}/2$, hence

$$\Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi}.$$

Examples:

1. $\Gamma(4) = 3! = 6;\qquad \dfrac{\Gamma(4)}{\Gamma(5)} = \dfrac{3!}{4!} = \dfrac{1}{4};$

2. $\Gamma\left(\frac{5}{2}\right) -$ use $\Gamma(x+1) = x\Gamma(x)$ with $x = 3/2$

$$\begin{aligned}\Gamma\left(\frac{5}{2}\right) &= \frac{3}{2}\Gamma\left(\frac{3}{2}\right) = \frac{3}{2}\left(\frac{1}{2}\Gamma\left(\frac{1}{2}\right)\right) = \frac{3}{2}\cdot\frac{1}{2}\cdot\sqrt{\pi} \\ &= \frac{3}{4}\sqrt{\pi}\end{aligned}$$

3. $\Gamma\left(-\frac{3}{2}\right) -$ now use $\Gamma(x) = \dfrac{\Gamma(x+1)}{x}$

$$\Gamma\left(-\frac{3}{2}\right) = \frac{\Gamma\left(-\frac{3}{2}+1\right)}{-3/2} = -\frac{2}{3}\Gamma\left(-\frac{1}{2}\right)$$

$$= -\tfrac{2}{3}\left[\frac{\Gamma\left(\frac{1}{2}\right)}{-1/2}\right] = -\tfrac{2}{3}\cdot -2\cdot\sqrt{\pi} = \tfrac{4}{3}\sqrt{\pi}$$

2 Introduction to Linear Algebra

2.1 Properties of Vectors

We consider real $n-$dimensional vectors belonging to the set $\mathbb{R}^n$. An $n-$tuple

$$\underline{v} = (v_1,\ v_2,\ ....,\ v_n) \in \mathbb{R}^n$$

is a vector of dimension n. The elements v_i $(i = 1, ..., n)$ are called components of $\underline{v}$.

Any pair $\underline{u},\ \underline{v} \in \mathbb{R}^n$ are equal iff

1. the corresponding components u_i's and v_i's are equal

2. dimensions of both vectors are the same and we write $\underline{u} = \underline{v}$.

Examples:

$\underline{u}_1 = (1,\ 0)\,,\ \underline{u}_2 = \left(1,\ e,\ \sqrt{3},\ 6\right),\ \underline{u}_3 = (3,\ 4)\,,\ \underline{u}_4 = (\pi,\ \ln 3,\ 2,\ 1)$

1. $\underline{u}_1$, $\underline{u}_3 \in \mathbb{R}^2$ and $\underline{u}_2$, $\underline{u}_4 \in \mathbb{R}^4$

2. $(x+y,\ x-z,\ 2z-1) = (3,\ -2,\ 5)$ .For equality to hold corresponding components are equal, so

$$\left.\begin{array}{c} x+y=3 \\ x-z=-2 \\ 2z-1=5 \end{array}\right\} \Rightarrow x=1;\ y=2;\ z=3$$

2.1.1 Vector Arithmetic

Let $\underline{u}$, $\underline{v} \in \mathbb{R}^n$. Then *vector addition* is defined as

$$\underline{u} + \underline{v} = (u_1 + v_1,\ u_2 + v_2,\ \ldots\ldots\ldots.,\ u_n + v_n)$$

If $k \in \mathbb{R}$ is any scalar then

$$k\underline{u} = (ku_1,\ ku_2,\ \ldots\ldots\ldots.,\ ku_n)$$

Note: vector addition only holds if the dimensions of each are identical.

Examples:

$\underline{u} = (3,\ 1,\ -2,\ 0)\,,\ \underline{v} = (5,\ -5,\ 1,\ 2)\,,\ \underline{w} = (0,\ -5,\ 3,\ 1)$

1. $\underline{u}+\underline{v} = (3+5,\ 1-5,\ -2+1,\ 0+2) = (8,\ -4,\ -1,\ 2)$

2. $2\underline{w} = (2.0,\ 2.(-5)\,,\ 2.3,\ 2.1) = (0,\ -10,\ 6,\ 2)$

3. $\underline{u} + \underline{v} - 2\underline{w} = (8,\ -4,\ -1,\ 2) - (0,\ -10,\ 6,\ 2) = (8,\ 6,\ -7,\ 0)$

$\underline{1} \in \mathbb{R}^n$ is given by $(1,\ 1, \ldots.., \ 1)$.

Similarly $\underline{0} = (0,\ 0,\ \ldots..,\ 0)$ is the *zero vector*.

Vectors can also be multiplied together using the *dot product* . If $\underline{u},\ \underline{v} \in \mathbb{R}^n$ then the dot product denoted by $\underline{u}.\underline{v}$ is

$$\underline{u}.\underline{v} = u_1 v_1 + u_2 v_2 + \ldots\ldots\ldots\ldots + u_n v_n \in \mathbb{R}$$

which is clearly a scalar quantity. The operation is commutative , i.e.

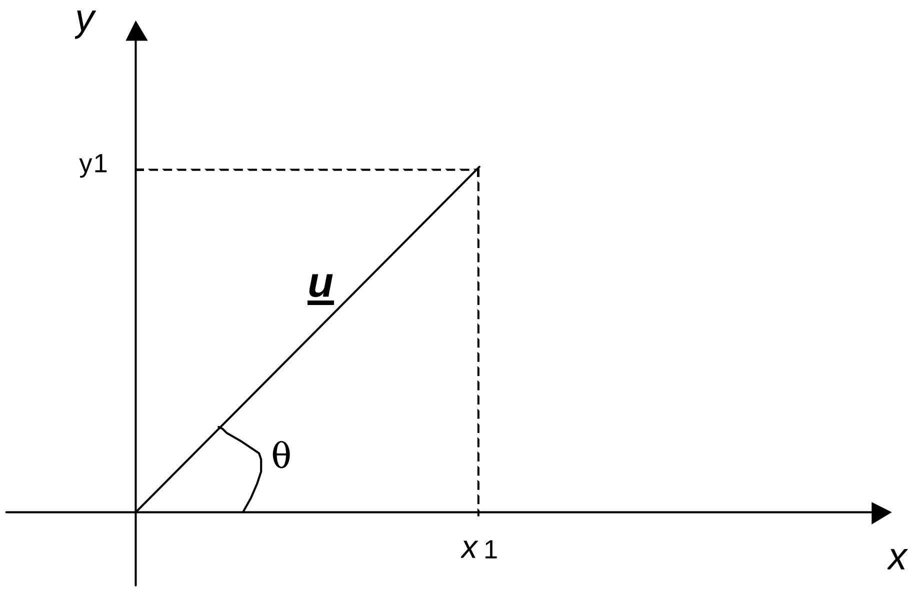
$$\underline{u}.\underline{v} = \underline{v}.\underline{u}$$

If a pair of vectors have a scalar product which is zero, they are said to be *orthogonal.*

Geometrically this means that the two vectors are perpendicular to each other.

2.1.2 Concept of Length in $\mathbb{R}^n$

Recall in 2-D $\underline{u} = (x_1, y_1)$



The length or *magnitude* of $\underline{u}$, written $|u|$ is given by Pythagoras

$$|u| = \sqrt{(x_1)^2 + (y_1)^2}$$

and the angle θ the vector makes with the horizontal is

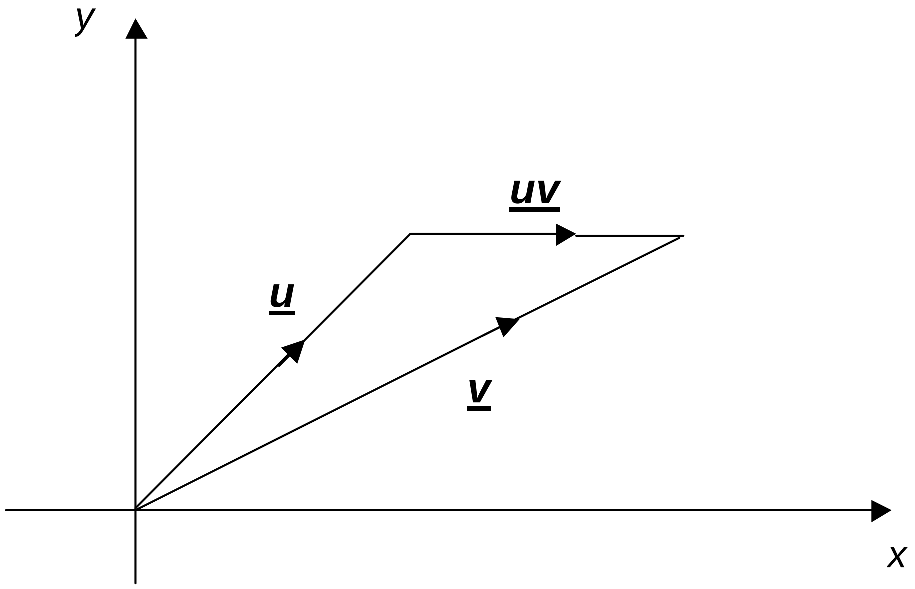
$$\theta = \text{arctan}\,\frac{y_1}{x_1}.$$

Any vector $\underline{u}$ can be expressed as

$$\underline{u} = |u|\,\underline{\widehat{u}}$$

where $\underline{\widehat{u}}$ is the *unit vector* because $|\underline{\widehat{u}}| = 1$.

Given any two vectors $\underline{u}$, $\underline{v} \in \mathbb{R}^2$, we can calculate the distance between them



$$\begin{aligned} |\underline{v} - \underline{u}| &= |(v_1\ ,\ v_2) - (u_1\ ,\ u_2)| \\ &= \sqrt{(v_1 - u_1)^2 + (v_2 - u_2)^2} \end{aligned}$$

In 3D (or $\mathbb{R}^3$) a vector $\underline{v} = (x_1,\ y_1,\ z_1)$ has length/magnitude

$$|v| = \sqrt{(x_1)^2 + (y_1)^2 + (z_1)^2}.$$

To extend this to $\mathbb{R}^n$, is similar.

Consider $\underline{v} = (v_1,\ v_2,\ \ldots\ldots\ldots,\ v_n) \in \mathbb{R}^n$. The length of $\underline{v}$ is called the *norm* and denoted $\|\underline{v}\|$, where

$$\|\underline{v}\| = \sqrt{(v_1)^2 + (v_2)^2 + \ldots\ldots\ldots + (v_n)^2}$$

If $\underline{u},\ \underline{v} \in \mathbb{R}^n$ then the distance between $\underline{u}$ and $\underline{v}$ is can be obtained in a similar fashion

$$\|\underline{v} - \underline{u}\| = \sqrt{(v_1 - u_1)^2 + (v_2 - u_2)^2 + \ldots\ldots\ldots + (v_n - u_n)^2}$$

We mentioned earlier that two vectors $\underline{u}$ and $\underline{v}$ in two dimension are orthogonal if $\underline{u}.\underline{v} = 0$.

The idea comes from the definition

$$\underline{u}.\underline{v} = |u|\ .\ |v| \cos\theta.$$

Re-arranging gives the angle between the two vectors. Note when $\theta = \pi/2$ $\underline{u}.\underline{v} = 0$.

If $\underline{u}, \ \underline{v} \in \mathbb{R}^n$ we write

$$\underline{u}.\underline{v} = |\,|u|\,|.|\,|v|\,|\cos\theta$$

Examples: Consider the following vectors

$$\begin{aligned} \underline{u} &= (2,\ -1,\ 0,\ -3)\ ,\ \ \underline{v} = (1,\ -1,\ -1,\ 3)\ , \\ \underline{w} &= (1,\ 3,\ -2,\ 2) \end{aligned}$$

$$\|\underline{u}\| = \sqrt{(2)^2 + (-1)^2 + (0)^2 + (-3)^2} = \sqrt{14}$$

Distance between $\underline{v}$ & $\underline{w} = \|\underline{w} - \underline{v}\| =$

$$\sqrt{(1-1)^2 + (3-(-1))^2 + (-2-(-1))^2 + (2-3)^2}$$
$$= \ 3\sqrt{2}$$

The angle between $\underline{u}$ & $\underline{w}$ can be obtained from

$$\cos\theta = \frac{\underline{u}.\underline{v}}{||u||\,||v||}.$$

Hence

$$\begin{aligned}\cos\theta &= \frac{(2,\ -1,\ 0,\ -3).(1,\ -1,\ -1,\ 3)}{2\sqrt{3}\sqrt{14}} = -\sqrt{\frac{3}{14}} \rightarrow \\ \theta &= \cos^{-1}\left(-\sqrt{\frac{3}{14}}\right)\end{aligned}$$

2.2 Matrices

A *matrix* is a rectangular array $A = \left(a_{i\ j}\right)$ for $i = 1, ..., m$; $j = 1, ..., n$ written

$$A = \begin{pmatrix} a_{11} & a_{12} & .. & .. & .. & a_{1n} \\ a_{21} & a_{22} & .. & .. & .. & a_{2n} \\ \vdots & & & & & \vdots \\ \vdots & & & & & \vdots \\ \vdots & .. & .. & .. & .. & \vdots \\ a_{m1} & a_{m2} & .. & .. & .. & a_{mn} \end{pmatrix}$$

and is an $(m \times n)$ matrix, i.e. m rows and n columns.

If $m = n$ the matrix is called *square*. The product mn gives the number of elements in the matrix.

A vector which we have already seen is simply a case of a $(m \times 1)$ matrix, i.e.

$$\underline{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ \vdots \\ x_m \end{pmatrix}$$

2.2.1 Matrix Arithmetic

Let $A,\ B \in\ {}^m\mathbb{R}^n$

$A + B =$

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & .. & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & .. & .. & b_{1n} \\ b_{21} & b_{22} & .. & .. & b_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & .. & .. & .. & \vdots \\ b_{m1} & b_{m2} & .. & .. & b_{mn} \end{pmatrix}$$

and the corresponding elements are added to give

$$\begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + +b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & .. & \cdots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{pmatrix} = B{+}A$$

Matrices can only added if they are of the same form.

Examples:

$$
\begin{aligned}
A &= \begin{pmatrix} 1 & -1 & 2 \\ 0 & 3 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 4 & 0 & -3 \\ -1 & -2 & 3 \end{pmatrix}, \\
C &= \begin{pmatrix} 2 & -3 & 1 \\ 5 & -1 & 2 \\ -1 & 0 & 3 \end{pmatrix}, \quad D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}
\end{aligned}
$$

$$
A+B = \begin{pmatrix} 5 & -1 & -1 \\ -1 & 1 & 7 \end{pmatrix}; \quad C+D = \begin{pmatrix} 3 & -3 & 1 \\ 5 & 0 & 2 \\ -1 & 0 & 4 \end{pmatrix}
$$

We cannot perform any other combination of addition as A and B are (2×3) and C and D are (3×3).

2.2.2 Matrix Multiplication

To multiply two square matrices $\mathbf{A}$ and $\mathbf{B}$, so that $\mathbf{C} = \mathbf{AB}$, the elements of $\mathbf{C}$ are found from the recipe

$$C_{ij} = \sum_{k=1}^{N} A_{ik} B_{kj}.$$

That is, the i th row of $\mathbf{A}$ is dotted with the j th column of $\mathbf{B}$. For example,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{pmatrix}.$$

Note that in general $\mathbf{AB} \neq \mathbf{BA}$. The general rule for multiplication is

$$A_{pn}\ B_{nm} \to C_{pm}$$

Example:

$$
\begin{aligned}
& \begin{pmatrix} 2 & 1 & 0 \\ 2 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 3 \\ 1 & 2 \end{pmatrix} \\
= \; & \begin{pmatrix} 2.1 + 1.0 + 0.1 & 2.2 + 1.3 + 0.2 \\ 2.1 + 0.0 + 2.1 & 2.2 + 0.3 + 2.2 \end{pmatrix} \\
= \; & \begin{pmatrix} 2 & 7 \\ 4 & 8 \end{pmatrix}
\end{aligned}
$$

2.2.3 Transpose

The **transpose** of a matrix with entries A_{ij} is the matrix with entries A_{ji}; the entries are 'reflected' across the leading diagonal, i.e. rows become columns. The transpose of $\mathbf{A}$ is written $\mathbf{A}^{\mathsf{T}}$. If $\mathbf{A} = \mathbf{A}^{\mathsf{T}}$ then $\mathbf{A}$ is **symmetric**. For example, of the matrices

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}, \qquad \mathbf{C} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix},$$

we have $\mathbf{B} = \mathbf{A}^T$ and $\mathbf{C} = \mathbf{C}^T$. Note that for any matrix $\mathbf{A}$ and B

(i) $(\mathbf{A} + B)^T = \mathbf{A}^T + B^T$

(ii) $\left(\mathbf{A}^T\right)^T = \mathbf{A}$

(iii) $(kA)^T = kA^T$, k is a scalar

(iv) $(AB)^T = B^T A^T$

Example:

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \\ 2 & 2 \end{pmatrix} \rightarrow A^T = \begin{pmatrix} 2 & 1 & 2 \\ 1 & 2 & 2 \end{pmatrix}$$

2.2.4 Matrix Representation of Linear Equations

We begin by considering a two-by-two set of equations for the unknowns x and y :

$$\begin{aligned} ax + by &= p \\ cx + dy &= q \end{aligned}$$

The solution is easily found. To get x, multiply the first equation by d, the second by b, and subtract to eliminate y :

$$(ad - bc)\,x = dp - bq.$$

Then find y :

$$(ad - bc)\,y = aq - cp.$$

This works and gives a unique solution *as long as* $ad - bc \neq 0$.

If $ad - bc = 0$, the situation is more complicated: there may be no solution at all, or there may be many.

Examples:

Here is a system with a unique solution:

$$\begin{aligned} x - y &= 0 \\ x + y &= 2 \end{aligned}$$

The solution is $x = y = 1$.

Now try

$$\begin{aligned} x - y &= 0 \\ 2x - 2y &= 2 \end{aligned}$$

Obviously there is no solution: from the first equation $x = y$, and putting this into the second gives $0 = 2$. Here $ad - bc = 1\,(-2) - (1-)\,2 = 0$.

Also note what is being said:

$$\left.\begin{array}{c} x = y \\ x = 1 + y \end{array}\right\} \text{ Impossible.}$$

Lastly try

$$\begin{aligned} x - y &= 1 \\ 2x - 2y &= 2. \end{aligned}$$

The second equation is twice the first so gives no new information. Any x and y satisfying the first equation satisfy the second. This system has many solutions.

Note: If we have one equation for two unknowns the system is undetermined and has many solutions. If we have *three* equations for two unknowns, it is over-determined and in general has no solutions at all.

Then the general (2×2) system is written

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} p \\ q \end{pmatrix}$$

or

$$\mathbf{A}\underline{\mathbf{x}} = \underline{\mathbf{p}}.$$

The equations can be solved if the matrix $\mathbf{A}$ is **invertible.** This is the same as saying that its **determinant**

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

is not zero.

These concepts generalise to systems of N equations in N unknowns. Now the matrix $\mathbf{A}$ is $N \times N$ and the vectors $\mathbf{x}$ and $\mathbf{p}$ have N entries.

Here are two special forms for $\mathbf{A}$. One is the identity matrix, which has its own inverse

$$\mathsf{I} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & \\ 0 & 0 & 1 & \dots & \vdots \\ \vdots & & \ddots & & 0 \\ 0 & & \dots & 0 & 1 \end{pmatrix}.$$

and for any $\mathbf{x}$, $\mathsf{I}\mathbf{x} = \mathbf{x}$. The other is the **tridiagonal form.** This is common in finite difference numerical schemes.

$$A = \begin{pmatrix} * & * & 0 & \cdots & \cdots & 0 \\ * & \ddots & \ddots & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & \ddots & * \\ 0 & \cdots & \cdots & 0 & * & * \end{pmatrix}$$

There is a main diagonal, and one above called the *super diagonal* and one below called the *sub-diagonal.*

A *symmetric matrix* A has the property

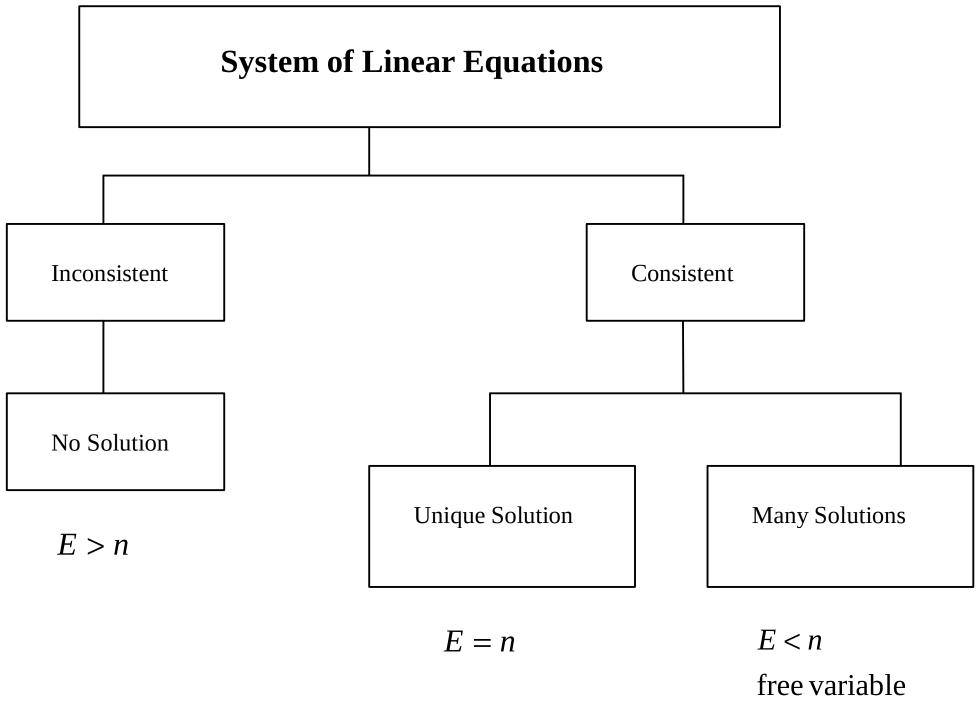
$$\begin{aligned} \mathbf{A}^{\mathsf{T}} &= A \\ \left(a_{ij}\right) &= \left(a_{ji}\right) \end{aligned}$$

so the leading diagonal acts as a "mirror". Clearly a symmetric matrix is square. For example

$$\begin{pmatrix} \alpha & a & b & c \\ a & \beta & d & e \\ b & d & \gamma & f \\ c & e & f & \delta \end{pmatrix}$$

is symmetric.

To conclude:



where $E =$ number of equations and $n =$ unknowns.

The theory and numerical analysis of linear systems accounts for quite a large branch of mathematics.

2.3 Using Matrix Notation For Solving Linear Systems

The usual notation for systems of linear equations is that of matrices and vectors. Consider the system

$$\begin{aligned} ax + by + cz &= p \\ dx + ey + fz &= q \\ gx + hy + iz &= r \end{aligned} \qquad (*)$$

for the unknown variables x, y, z. We gather the unknowns x, y and z and the given p, q and r into vectors:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix}, \qquad \begin{pmatrix} p \\ q \\ r \end{pmatrix}$$

and put the coefficients into a matrix

$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}.$$

A is called the *coefficient matrix* of the linear system $(*)$ and the special matrix formed by

$$\left(\begin{array}{ccc|c} a & b & c & p \\ d & e & f & q \\ g & h & i & r \end{array}\right)$$

is called the *augmented matrix.*

Now consider a general linear system consisting of m equations in n unknowns which can be written in augmented form as

$$\left(\begin{array}{cccccc|c} a_{11} & a_{12} & .. & .. & .. & a_{1n} & b_1 \\ a_{21} & a_{22} & .. & .. & .. & a_{2n} & b_2 \\ \vdots & & & & & \vdots & \vdots \\ \vdots & & & & & \vdots & \vdots \\ \vdots & .. & .. & .. & .. & \vdots & \vdots \\ a_{m1} & a_{m2} & - & - & - & a_{mn} & b_m \end{array}\right).$$

We can perform a series of row operations on this matrix and reduce it to a simplified matrix of the form

$$\left(\begin{array}{cccccc|c} a_{11} & a_{12} & .. & .. & .. & a_{1n} & b_1 \\ 0 & a_{22} & .. & .. & .. & a_{2n} & b_2 \\ 0 & 0 & & & & \vdots & \vdots \\ 0 & 0 & 0 & & & \vdots & \vdots \\ \vdots & .. & .. & .. & .. & \vdots & \vdots \\ 0 & 0 & - & - & 0 & a_{mn} & b_m \end{array}\right).$$

Such a matrix is said to be of *echelon form* if the number of zeros preceding the first nonzero entry of each row increases row by row.

A matrix A is said to be *row equivalent* to a matrix B, written $A \sim B$ if B can be obtained from A from a finite sequence of operations called *elementary row operations* of the form:

[ER$_1$]: Interchange the i th and j th rows: $R_i \leftrightarrow R_j$

[ER$_2$]:Replace the i th row by itself multiplied by a nonzero constant k : $R_i \rightarrow kR_i$

[ER$_3$]:Replace the i th row by itself plus k times the j th row: $R_i \rightarrow R_i + kR_j$

These have no affect on the solution of the of the linear system which gives the augmented matrix.

Examples:

Solve the following linear systems

1.

$$\left.\begin{array}{r} 2x + y - 2z = 10 \\ 3x + 2y + 2z = 1 \\ 5x + 4y + 3z = 4 \end{array}\right\} \equiv A\underline{x} = \underline{b} \ \text{ with}$$

$$A = \begin{pmatrix} 2 & 1 & -2 \\ 3 & 2 & 2 \\ 5 & 4 & 3 \end{pmatrix} \ \text{ and } \ \underline{b} = \begin{pmatrix} 10 \\ 1 \\ 4 \end{pmatrix}$$

The augmented matrix for this system is

$$\left(\begin{array}{ccc|c} 2 & 1 & -2 & 10 \\ 3 & 2 & 2 & 1 \\ 5 & 4 & 3 & 4 \end{array}\right) \underset{R_3 \to 2R_3 - 5R_1}{\overset{R_2 \to 2R_2 - 3R_1}{\sim}} \left(\begin{array}{ccc|c} 2 & 1 & -2 & 10 \\ 0 & 1 & 10 & -28 \\ 0 & 3 & 16 & -42 \end{array}\right)$$

$$\underset{R_1 \to R_1 - R_2}{\overset{R_3 \to R_3 - 3R_2}{\sim}} \left(\begin{array}{ccc|c} 2 & 0 & -12 & 38 \\ 0 & 1 & 10 & -28 \\ 0 & 0 & -14 & 42 \end{array}\right)$$

$$\begin{aligned} -14z &= 42 \to z = -3 \\ y + 10z &= -28 \to y = -28 + 30 = 2 \\ x - 6z &= 19 \to x = 19 - 18 = 1 \end{aligned}$$

Therefore solution is unique with

$$\underline{x} = \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}$$

2.

$$\left.\begin{array}{c} x+2y-3z=6 \\ 2x-y+4z=2 \\ 4x+3y-2z=14 \end{array}\right\}$$

$$\left(\begin{array}{ccc|c} 1 & 2 & -3 & 6 \\ 2 & -1 & 4 & 2 \\ 4 & 3 & -2 & 14 \end{array}\right) \underset{R_3 \to R_3 - 4R_1}{\overset{R_2 \to R_2 - 2R_1}{\sim}}$$
$$\left(\begin{array}{ccc|c} 1 & 2 & -3 & 6 \\ 0 & -5 & 10 & -10 \\ 0 & -5 & 10 & -10 \end{array}\right) \underset{R_2 \to 0.5R_2}{\overset{R_3 \to R_3 - R_2}{\sim}}$$
$$\left(\begin{array}{ccc|c} 1 & 2 & -3 & 6 \\ 0 & 1 & -2 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Number of equations is less than number of unknowns.

$y-2z=2$ so $z=a$ is a free variable$\Rightarrow$ $y=2(1+a)$

$x+2y-3z=6 \rightarrow x=6-2y+3z=2-a$

$\Rightarrow x=2-a; \quad y=2(1+a); \quad z=a$

Therefore there are many solutions

$$\underline{x} = \begin{pmatrix} 2 - a \\ 2\,(1 + a) \\ a \end{pmatrix}$$

3.

$$\left.\begin{array}{c} x + 2y - 3z = -1 \\ 3x - y + 2z = 7 \\ 5x + 3y - 4z = 2 \end{array}\right\}$$

$$\left(\begin{array}{ccc|c} 1 & 2 & -3 & -1 \\ 3 & -1 & 2 & 7 \\ 5 & 3 & -4 & 2 \end{array}\right) \underset{R_3 \to R_3 - 5R_1}{\overset{R_2 \to R_2 - 3R_1}{\sim}}$$

$$\left(\begin{array}{ccc|c} 1 & 2 & -3 & -1 \\ 0 & -7 & 11 & 10 \\ 0 & -7 & 11 & 7 \end{array}\right) \overset{R_3 \to R_3 - R_2}{\sim}$$

$$\left(\begin{array}{ccc|c} 1 & 2 & -3 & -1 \\ 0 & -7 & 11 & 10 \\ 0 & 0 & 0 & -3 \end{array}\right)$$

The last line reads $0 = -3$. Also middle iteration shows that the second and third equations are inconsistent.

Hence no solution exists.

2.4 Matrix Inverse

The **inverse** of a matrix $\mathbf{A}$, written $\mathbf{A}^{-1}$, satisfies

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}.$$

It may not always exist, but if it does, the solution of the system

$$\mathbf{A}\mathbf{x} = \mathbf{p}$$

is

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{p}.$$

The inverse of the matrix for the special case of a 2×2 matrix

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

is

$$\frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

provided that $ad - bc \neq 0$.

The inverse of any $n \times n$ matrix A is defined as

$$A^{-1} = \frac{1}{|A|} \text{adj } A$$

where adj $A = \left[(-1)^{\mathsf{i+j}} \left| M_{\mathsf{ij}} \right| \right]^{\mathsf{T}}$ is the adjoint, i.e. we form the matrix of A's cofactors and transpose it.

M_{ij} is the square sub-matrix obtained by "covering the i^{th} row and j^{th} column", and its determinant is called the **Minor** of the element a_{ij}. The term $A_{ij} = (-1)^{i+j} \left| M_{ij} \right|$ is then called the **cofactor** of a_{ij}.

Consider the following example with

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 3 \end{pmatrix}$$

So the determinant is given by $|A| =$

$$(-1)^{1+1} A_{11} |M_{11}| + (-1)^{1+2} A_{12} |M_{12}| + (-1)^{1+3} A_{13} |M_{13}|$$

$$\begin{aligned} &= 1 \begin{vmatrix} 2 & 1 \\ 1 & 3 \end{vmatrix} - 1 \begin{vmatrix} 1 & 1 \\ 0 & 3 \end{vmatrix} + 0 \begin{vmatrix} 1 & 2 \\ 0 & 1 \end{vmatrix} \\ &= (2 \times 3 - 1 \times 1) - (1 \times 3 - 1 \times 0) + 0 = 5 - 3 \\ &= 2 \end{aligned}$$

Here we have expanded about the 1^{st} row - we can do this about any row. If we expand about the 2^{nd} row - we should still get $|A| = 2$.

We now calculate the adjoint:

$$(-1)^{1+1}\,M_{11} = +\begin{vmatrix} 2 & 1 \\ 1 & 3 \end{vmatrix} \quad (-1)^{1+2}\,M_{12} = -\begin{vmatrix} 1 & 1 \\ 0 & 3 \end{vmatrix}$$

$$(-1)^{1+3}\,M_{13} = +\begin{vmatrix} 1 & 2 \\ 0 & 1 \end{vmatrix}$$

$$(-1)^{2+1}\,M_{21} = -\begin{vmatrix} 1 & 0 \\ 1 & 3 \end{vmatrix} \quad (-1)^{2+2}\,M_{22} = +\begin{vmatrix} 1 & 0 \\ 0 & 3 \end{vmatrix}$$

$$(-1)^{2+3}\,M_{23} = -\begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix}$$

$$(-1)^{3+1}\,M_{31} = +\begin{vmatrix} 1 & 0 \\ 2 & 1 \end{vmatrix} \quad (-1)^{3+2}\,M_{32} = -\begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix}$$

$$(-1)^{3+3}\,M_{33} = +\begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix}$$

$$\text{adj } A = \begin{pmatrix} 5 & -3 & 1 \\ -3 & 3 & -1 \\ 1 & -1 & 1 \end{pmatrix}^{\mathsf{T}}$$

We can now write the inverse of A (which is symmetric)

$$A^{-1} = \frac{1}{2}\begin{pmatrix} 5 & -3 & 1 \\ -3 & 3 & -1 \\ 1 & -1 & 1 \end{pmatrix}$$

Elementary row operations (as mentioned above) can be used to simplify a determinant, as increased numbers of zero entries present, requires less calculation. There are two important points, however. Suppose the value of the determinant is $|A|$, then:

[ER$_1$]: $R_i \leftrightarrow R_j \Rightarrow |A| \rightarrow -|A|$

[ER$_2$]: $R_i \rightarrow kR_i \Rightarrow |A| \rightarrow k\,|A|$

2.5 Orthogonal Matrices

A matrix $\mathbf{P}$ is **orthogonal** if

$$\mathbf{P}\mathbf{P}^{\mathsf{T}} = \mathbf{P}^{\mathsf{T}}\mathbf{P} = \mathsf{I}.$$

This means that the rows and columns of $\mathbf{P}$ are orthogonal and have unit length. It also means that

$$\mathbf{P}^{-1} = \mathbf{P}^{\mathsf{T}}.$$

In two dimensions, orthogonal matrices have the form

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \qquad \text{or} \qquad \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$$

for some angle θ and they correspond to rotations or reflections.

So rows and columns being orthogonal means row i·row $j = 0$, i.e. they are perpendicular to each other.

$$\begin{aligned}(\cos\theta,\ \sin\theta)\cdot(-\sin\theta,\ \cos\theta) &= \\ -\cos\theta\sin\theta+\sin\theta\cos\theta &= 0 \\ (\cos\theta,\ \sin\theta)\cdot(\sin\theta,\ -\cos\theta) &= \\ \cos\theta\sin\theta-\sin\theta\cos\theta &= 0\end{aligned}$$

$$\underline{v} = (\cos\theta,\ -\sin\theta)^T \to |\underline{v}| = \cos^2\theta + (-\sin\theta)^2 = 1$$

Finally, if $P = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$ then

$$P^{-1} = \frac{1}{\underbrace{\cos^2\theta - \left(-\sin^2\theta\right)}_{=1}} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = P^T.$$

2.6 Eigenvalues and Eigenvectors

If $\mathbf{A}$ is a square matrix, $\underline{\mathbf{v}}$ is an **eigenvector** of $\mathbf{A}$ with **eigenvalue** λ if

$$\mathbf{A}\underline{\mathbf{v}} = \boldsymbol{\lambda}\underline{\mathbf{v}} \qquad \text{or} \qquad (\mathbf{A} - \boldsymbol{\lambda}\mathsf{I})\,\underline{\mathbf{v}} = \mathbf{0}.$$

An $N \times N$ matrix has exactly N eigenvalues, not all necessarily real or distinct; they are the roots of the *characteristic equation*

$$\det(\mathbf{A} - \boldsymbol{\lambda}\mathsf{I}) = 0.$$

and each solution has a corresponding eigenvector $\underline{\mathbf{v}}$. $\mathbf{A} - \boldsymbol{\lambda}\mathsf{I}$ is the *characteristic polynomial.*

The eigenvectors are in some sense special directions for the matrix $\mathbf{A}$. In complete generality this is a vast topic. Many Boundary-Value

Problems can be reduced to eigenvalue problems.

We will just look at real symmetric matrices for which $\mathbf{A} = \mathbf{A}^\mathsf{T}$. For these matrices

- The eigenvalues are real;

- The eigenvectors corresponding to distinct eigenvalues are orthogonal;

- The matrix can be **diagonalised:** that is, there is an orthogonal matrix $\mathbf{P}$ such that

$$A = PDP^\mathsf{T} \qquad \text{or} \qquad P^\mathsf{T}AP = D$$

where $\mathbf{D}$ is **diagonal**, that is only the entries on the leading diagonal are nonzero, and these are equal to the eigenvalues of $\mathbf{A}$.

$$D = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \lambda_n \end{pmatrix}$$

Example:

$$A = \begin{pmatrix} 3 & 3 & 3 \\ 3 & -1 & 1 \\ 3 & 1 & -1 \end{pmatrix}$$

then

$$\begin{aligned} \det(A - \boldsymbol{\lambda}\mathrm{I}) &= \begin{vmatrix} 3-\lambda & 3 & 3 \\ 3 & -1-\lambda & 1 \\ 3 & 1 & -1-\lambda \end{vmatrix} \\ &= -\lambda^3 + \lambda^2 + 24\lambda + 36 = 0 \\ &= (\lambda+3)(\lambda+2)(\lambda-6) \end{aligned}$$

so that the eigenvalues, i.e. the roots of this equation, are $\lambda_1 = -3$, $\lambda_2 = -2$ and $\lambda_3 = 6$.

Eigenvectors are now obtained from

$$\begin{pmatrix} 3-\lambda_i & 3 & 3 \\ 3 & -1-\lambda_i & 1 \\ 3 & 1 & -1-\lambda_i \end{pmatrix} \underline{\mathbf{v}}_i = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad i = 1,\ 2,\ 3$$

$$\lambda_1 = -3: \qquad \begin{pmatrix} 6 & 3 & 3 \\ 3 & 2 & 1 \\ 3 & 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Upon row reduction we have $\left(\begin{array}{ccc|c} 2 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \rightarrow y = z$, so put $z = a$ and $2x = -y - z \rightarrow x = -\alpha \therefore \underline{\mathbf{v}}_1 = \alpha \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}$

Similarly

$$\lambda_2 = -2: \ \underline{\mathbf{v}}_2 = \beta \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \qquad \lambda_3 = 6: \ \underline{\mathbf{v}}_3 = \gamma \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$$

If we take $\alpha = \beta = \gamma = 1$ the corresponding eigenvectors are

$$\underline{\mathbf{v}}_1 = \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}, \ \underline{\mathbf{v}}_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \ \underline{\mathbf{v}}_3 = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$$

Now normalise these, i.e. $|\underline{\mathbf{v}}| = 1$. Use $\widehat{\underline{\mathbf{v}}} = \underline{\mathbf{v}} / |\underline{\mathbf{v}}|$ for normalised eigenvectors

$$\widehat{\underline{\mathbf{v}}}_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}, \quad \widehat{\underline{\mathbf{v}}}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \quad \widehat{\underline{\mathbf{v}}}_3 = \frac{1}{\sqrt{6}} \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$$

Hence

$$\mathbf{P} = \begin{pmatrix} \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{pmatrix} \to \mathbf{P}^{\mathsf{T}} = \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix}$$

so that

$$
\begin{aligned}
\mathbf{P}^{\mathsf{T}}\mathbf{A}\mathbf{P} &= \begin{pmatrix} -3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 6 \end{pmatrix} \\
&= D.
\end{aligned}
$$

2.7 Criteria for invertibility

A system of linear equations is uniquely solvable if and only if the matrix $\mathbf{A}$ is invertible. This in turn is true if any of the following is:

1. If and only if the determinant is nonzero;

2. If and only if all the eigenvalues are nonzero;

3. If (but not only if) it is **strictly diagonally dominant.**

In practise it takes far too long to work out the determinant. The second criterion is often useful though, and there are quite quick methods for working out the eigenvalues. The third method is explained on the next page.

(Note: there are many other criteria for invertibility.)

A matrix $\mathbf{A}$ with entries A_{ij} is strictly diagonally dominant if

$$|A_{ii}| > \sum_{j \neq i} \left| A_{ij} \right| .$$

That is, the diagonal element in each row is bigger in modulus that the sum of the moduli of the off-diagonal elements in that row.

Examples:

$$\begin{pmatrix} 2 & 0 & 1 \\ 1 & 4 & 2 \\ 1 & 3 & 6 \end{pmatrix} \quad \text{is s.d.d. and so invertible;}$$

$$\begin{pmatrix} 1 & 0 & 2 \\ 2 & 5 & 1 \\ 3 & 2 & 13 \end{pmatrix} \quad \text{is not s.d.d. but still invertible;}$$

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{is neither s.d.d. \ nor invertible.}$$

3 Introduction to Probability

3.1 Preliminaries

A set Ω of all possible outcomes of some given experiment is called the *sample space*.

A particular outcome $\omega \in \Omega$ is called a *sample point*, or *sample path* for a stochastic process.

An *event* Ψ is a set of outcomes, i.e. $\Psi \subset \Omega$.

Example 1

Experiment: A dice is rolled and the number appearing on top is observed. The sample space consists of the 6 possible numbers:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

If the number 4 appears then $\omega = 4$ is a sample point, clearly $4 \in \Omega$.

Let Ψ_1, Ψ_2, $\Psi_3 =$ events that an even, odd, prime number occurs respectively.

So

$$\Psi_1 = \{2,\ 4,\ 6\}\,,\ \Psi_2 = \{1,\ 3,\ 5\}\,,\ \ \Psi_3 = \{2,\ 3,\ 5\}$$

$\Psi_1 \cup \Psi_3 = \{\ 2,\ 3,\ 4,\ 5,\ 6\} -$ event that an even or prime number occurs.

$\Psi_2 \cap \Psi_3 = \{3,\ 5\} -$ event that odd and prime number occurs.

$\Psi_3^c = \{1,\ 4,\ 6\} -$ event that prime number does not occur (complement of event).

$\boxed{\text{Example 2}}$ Experiment:

Toss a coin twice and observe the sequence of heads (H) and tails (T) that appears. Sample space

$$\Omega = \{\text{HH, TT, HT, TH}\}$$

Let Ψ_1 be event that at least one head appears, and Ψ_2 be event that both tosses are the same:

$$\Psi_1 = \{\text{HH, HT, TH}\}\,, \;\; \Psi_2 = \{\text{HH, TT}\}$$

$\Psi_1 \cap \Psi_2 = \{\text{HH}\}$

Events are subsets of Ω, but not all subsets of Ω are events.

3.1.1 Random Variables

Outcomes of experiments are not always numbers, e.g. 2 heads appearing; picking an ace from a deck of cards. We need some way of assigning real numbers to each random event. Random variables assign numbers to events.

Thus a *random variable* (RV) X is a function which maps from the sample space Ω to the set of real numbers

$$X : \omega \in \Omega \rightarrow \mathbb{R},$$

i.e. it associates a number $X(\omega)$ with each outcome ω.

Consider the example of tossing a coin and suppose we are paid £1 for each head and we lose £1 each time a tail appears. We know that $\mathbb{P}(\mathrm{H}) = \mathbb{P}(\mathrm{T}) = \frac{1}{2}$. So now we can assign the following outcomes

$$\begin{aligned} \mathbb{P}(1) &= \frac{1}{2} \\ \mathbb{P}(-1) &= \frac{1}{2} \end{aligned}$$

Mathematically, if our random variable is X, then

$$X = \begin{cases} +1 & \text{if H} \\ -1 & \text{if T} \end{cases}$$

or using the notation above $X : \omega \in \{\text{H,T}\} \to \{-1, 1\}$.

The probability that the RV takes on each possible value is called the *probability distribution.*

If X is a RV then

$$\mathbb{P}(X = a) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = a\})$$

is the probability that a occurs (or X maps onto a).

$P(a \leq X \leq b) =$ probability that X lies in the interval $[a,\ b] =$

$$\mathbb{P}(\{\omega \in \Omega : a \leq X(\omega) \leq b\})$$

$$X : \underset{\text{Domain}}{\Omega} \longrightarrow \underset{\text{Range (finite)}}{\mathbb{R}}$$

$X(\Omega) = \{x_1, ..., x_n\} = \{x_i\}_{1 \leq i \leq n}$

$\mathbb{P}[x_i] = \mathbb{P}[X = x_i] = f\ \ (x_i)\ \forall\ i.$

So the earlier coin tossing example gives

$$\mathbb{P}(X = 1) = \frac{1}{2};\ \ \mathbb{P}(X = -1) = \frac{1}{2}$$

$f(x_i)$ is the probability distribution of X.

This is called a *discrete probability distribution.*

x_i	x_1	x_2	....	x_n
$f(x_i)$	$f(x_1)$	$f(x_2)$	....	$f(x_n)$

There are two properties of the distribution $f\ \ (x_i)$

(i) $f(x_i) \geq 0\ \ \forall\ i \in [1,\ n]$

(ii) $\sum\limits_{i=1}^{n} f(x_i) = 1$, i.e. sum of all probabilities is one.

3.1.2 Mean/Expectation

The *mean* μ measures the centre (average) of the distribution

$$\begin{aligned} \mu &= \mathbb{E}[X] = \sum_{i=1}^{n} x_i \ f\ (x_i) \\ &= x_1 f(x_1) + x_2 f(x_2) + \ldots.. + x_n f(x_n) \end{aligned}$$

which is equal to the weighted average of all possible values of X together with associated probabilities.

This is also called the *first moment.*

Example:

x_i	2	3	8
$f(x_i)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

$$\begin{aligned} \mu &= \mathbb{E}[X] = \sum_{i=1}^{3} x_i f(x_i) = 2\left(\frac{1}{4}\right) + 3\left(\frac{1}{2}\right) + 8\left(\frac{1}{4}\right) \\ &= 4 \end{aligned}$$

3.1.3 Variance/Standard Deviation

This measures the spread (dispersion) of X about the mean.

Variance $\mathbb{V}[X] =$

$$\mathbb{E}\left[(X-\mu)^2\right] = \mathbb{E}\left[X^2\right] - \mu^2 = \sum_{i=1}^{n} x_i^2 f(x_i) - \mu^2 = \sigma^2$$

$\mathbb{E}\left[(X-\mu)^2\right]$ is also called the *second moment about the mean.*

From the previous example we have $\mu = 4$, therefore

$$\begin{aligned}\mathbb{V}[X] &= \left(2^2\left(\frac{1}{4}\right) + 3^2\left(\frac{1}{2}\right) + 8^2\left(\frac{1}{4}\right)\right) - 16 \\ &= 5.5 = \sigma^2 \rightarrow \sigma = 2.34\end{aligned}$$

3.1.4 Rules for Manipulating Expectations

Suppose $X,\ Y$ are random variables and $\alpha,\ \beta,\ \lambda \in \mathbb{R}$ are constant scalar quantities. Then

- $\mathbb{E}\left[\lambda X\right] = \lambda \mathbb{E}\left[X\right]$
- $\mathbb{E}\left[X + Y\right] = \mathbb{E}\left[X\right] + \mathbb{E}\left[Y\right],$ (linearity)
- $\mathbb{V}\left[\alpha X + \beta\right] = \alpha^2 \mathbb{V}\left[X\right]$
- $\mathbb{E}\left[XY\right] = \mathbb{E}\left[X\right]\mathbb{E}\left[Y\right],$
- $\mathbb{V}\left[X + Y\right] = \mathbb{V}\left[X\right] + \mathbb{V}\left[Y\right]$

The last two are provided $X,\ Y$ are independent.

3.1.5 Continuous Random Variables

As the number of discrete events becomes very large, individual probabilities $f\ (x_i) \to 0$. Now look at the continuous case.

Instead of $f\ (x_i)$ we now have $p(x)$ which is a continuous distribution called as *probability density function, PDF.*

$$P(a \leq X \leq b) = \int_a^b p(x)\,dx$$

The *cumulative distribution function* $F(x)$ of a RV X is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} p(x)\,dx$$

$F(x)$ is related to the PDF by

$$p(x) = \frac{dF}{dx}$$

(fundamental theorem of calculus) provided $F(x)$ is differentiable. However unlike $F(x)$, $p(x)$ may have singularities (and may be unbounded).

3.1.6 Special Expectations:

Given any PDF $p(x)$ of X.

Mean $\mu = \mathbb{E}[X] = \int_{\mathbb{R}} xp(x)\,dx$.

Variance $\sigma^2 = \mathbb{V}[X] = \mathbb{E}\left[(X-\mu)^2\right] = \int_{\mathbb{R}} x^2 p(x)\,dx - \mu^2$

(2^{nd} moment about the mean).

The n^{th} moment about zero is defined as

$$\begin{aligned}\mu_n &= \mathbb{E}[X^n] \\ &= \int_{\mathbb{R}} x^n p(x)\,dx.\end{aligned}$$

In general, for any function h

$$\mathbb{E}\left[h\left(X\right)\right] = \int_{\mathbb{R}} h\left(x\right) p\left(x\right) dx.$$

where X is a RV following the distribution given by $p\left(x\right)$.

Moments about the mean are given by

$$\mathbb{E}\left[\left(X-\mu\right)^n\right];\ n = 2,\ 3,\ \ldots$$

The special case $n = 2$ gives the variance σ^2.

3.1.7 Skewness and Kurtosis

Having looked at the variance as being the second moment about the mean, we now discuss two further moments centred about μ, that provide further important information about the probability distribution.

Skewness is a measure of the asymmetry of a distribution (i.e. lack of symmetry) about its mean. A distribution that is identical to the left and right about a centre point is symmetric.

The third central moment, i.e. third moment about the mean scaled with σ^3

$$\frac{\mathbb{E}\left[(X-\mu)^3\right]}{\sigma^3}$$

is called the *skew* and is a measure of the skewness (a non-symmetric distribution is called *skewed*).

Any distribution which is symmetric about the mean has a skew of zero.

Negative values for the skewness indicate data that are skewed left and positive values for the skewness indicate data that are skewed right.

By skewed left, we mean that the left tail is long relative to the right tail. Similarly, skewed right means that the right tail is long relative to the left tail.

The fourth centred moment scaled by the variance, called the *kurtosis* is defined

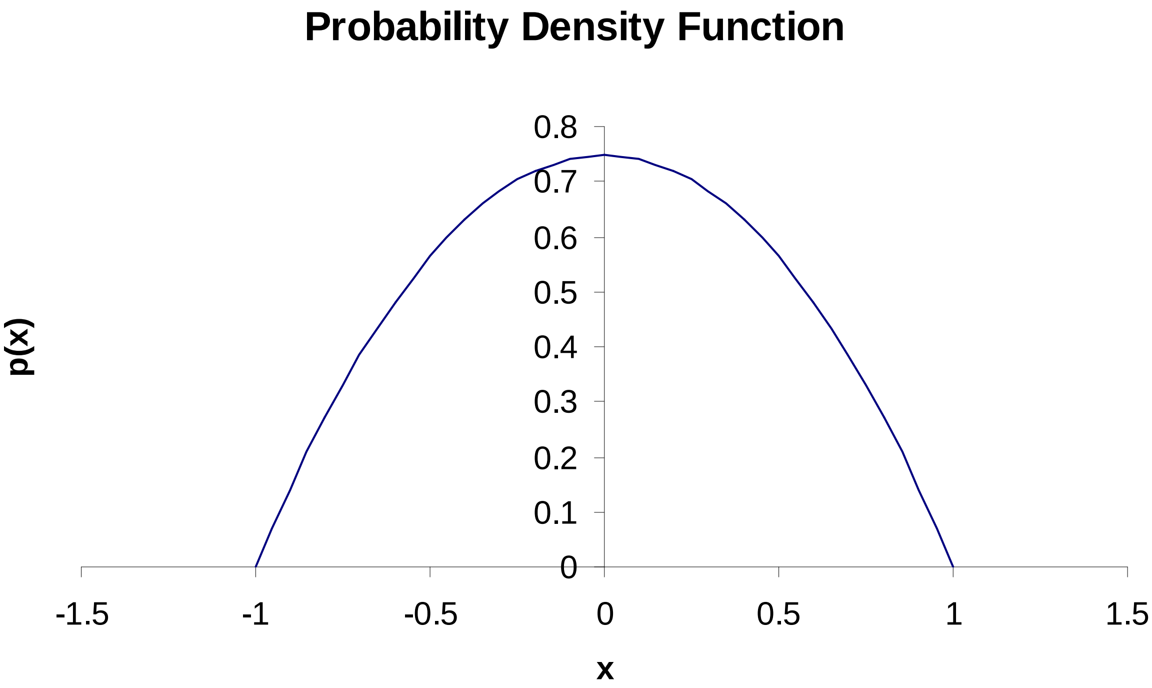
$$\frac{\mathbb{E}\left[(X-\mu)^4\right]}{\sigma^4}.$$

This is a measure of how much of the distribution is out in the tails at large negative and positive values of X.

Example:

Consider a continuous PDF

$$p(x) = \begin{cases} k\left(1 - x^2\right) & |x| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$



i) Calculate k :

We know

$$\int_{-\infty}^{\infty} p\left(x\right) dx = 1 \therefore k \int_{-1}^{1} \left(1 - x^{2}\right) dx = 1$$

$$k \left(x - \tfrac{1}{3} x^{3}\right)\Big|_{-1}^{1} \rightarrow k = \tfrac{3}{4}$$

ii) $\mathbb{E}\left[X\right] = \int_{\mathbb{R}} x p\left(x\right) dx = \frac{3}{4} \int_{-1}^{1} \left(x - x^{3}\right) dx$

If $f\left(x\right)$ is an odd function, i.e. $f\left(-x\right) = -f\left(x\right)$ then $\int_{-a}^{a} f\left(x\right) dx = 0 \quad \therefore \quad \mu = \mathbf{E}\left[X\right] = 0.$

iii) $\mathbb{V}[X] = \int_{\mathbb{R}} x^2 p(x)\,dx - \mu^2 = \int_{\mathbb{R}} x^2 p(x)\,dx$

$$= \tfrac{3}{4}\int_{-1}^{1} \left(x^2 - x^4\right) dx.$$

If $f\ (x)$ is an even function, i.e. $f\ (-x) = f\ (x)$ then $\int_{-a}^{a} f\ (x)\,dx = 2\int_{0}^{a} f\ (x)\,dx \quad \therefore$

$$\begin{aligned}\mathbb{V}[X] &= \tfrac{3}{2}\int_{0}^{1} \left(x^2 - x^4\right) dx = \tfrac{3}{2}\left(\tfrac{1}{3}x^3 - \tfrac{1}{5}x^5\right)\Big|_0^1 \\ &= \tfrac{1}{5} = \sigma^2 \longrightarrow \text{ standard deviation } \sigma \approx 0.45\end{aligned}$$

iv) Calculate the probability that a random variable X which follows this distribution, lies in the interval $\left(-\frac{1}{3},\ \frac{1}{2}\right)$. So

$$\mathbb{P}\left(-\tfrac{1}{3} \le X \le \tfrac{1}{2}\right) = \int_{-1/3}^{1/2} p(x)\,dx$$

$$= \tfrac{3}{4}\int_{-1/3}^{1/2} \left(1 - x^2\right) dx \approx 0.58$$

3.2 Normal Distribution

The *normal* (or *Gaussian*) distribution $N\left(\mu,\ \sigma^2\right)$ with mean and standard deviation μ and σ^2 in turn is defined in terms of its density function

$$p\left(x\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\left(x-\mu\right)^2}{2\sigma^2}\right).$$

For the special case $\mu = 0$ and $\sigma = 1$ it is called the *standard normal* distribution $N\left(0,\ 1\right)$.

This is also verified by making the substitution

$$\phi = \frac{x-\mu}{\sigma}$$

in $p\left(x\right)$ which gives

$$\Phi\left(\phi\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\phi^2\right)$$

and clearly has zero mean and unit variance:

$$\mathbb{E}\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma}\mathbb{E}\left[X-\mu\right] = 0,$$

$$\mathbb{V}\left[\frac{X-\mu}{\sigma}\right] = \mathbb{V}\left[\frac{X}{\sigma} - \frac{\mu}{\sigma}\right]$$

Now $\mathbb{V}\left[\alpha X + \beta\right] = \alpha^2 \mathbb{V}\left[X\right]$ (standard result), hence

$$\frac{1}{\sigma^2}\mathbb{V}\left[X\right] = \frac{1}{\sigma^2}.\sigma^2 = 1$$

Its cumulative distribution function is

$$F\left(x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}\phi^2} d\phi = P\left(-\infty \leq X \leq x\right).$$

The skewness of $N\left(0,\ 1\right)$ is zero and its kurtosis is 3.

3.3 Moment Generating Function

The *moment generating function* of X, denoted $M_X(\theta)$ is given by

$$M_X(\theta) = \mathbb{E}\left[e^{\theta x}\right] = \int_{\mathbb{R}} e^{\theta x} p(x)\, dx$$

provided the expectation exists. We can expand as a power series to obtain

$$M_X(\theta) = \sum_{n=0}^{\infty} \frac{\theta^n \mathbb{E}(X^n)}{n!}$$

so the n^{th} moment is the coefficient of $\theta^n/n!$, or the n^{th} derivative evaluated at zero.

How do we arrive at this result?

We use the Taylor series expansion for the exponential function: $\int_{\mathbb{R}} e^{\theta x} p(x)\, dx =$

$$\int_{\mathbb{R}} \left(1 + \theta x + \frac{(\theta x)^2}{2!} + \frac{(\theta x)^3}{3!} + \ldots\ldots\right) p(x)\, dx$$

$$
\begin{aligned}
&= \underbrace{\int_{\mathbb{R}} p(x)\,dx}_{1} + \theta \underbrace{\int_{\mathbb{R}} x p(x)\,dx}_{\mathbb{E}(X)} + \frac{\theta^2}{2!} \underbrace{\int_{\mathbb{R}} x^2 p(x)\,dx}_{\mathbb{E}(X^2)} + \\
&\quad \frac{\theta^3}{3!} \underbrace{\int_{\mathbb{R}} x^3 p(x)\,dx}_{\mathbb{E}(X^3)} + \ldots. \\
&= 1 + \theta \mathbb{E}(X) + \frac{\theta^2}{2!} \mathbb{E}\left(X^2\right) + \frac{\theta^3}{3!} \mathbb{E}\left(X^3\right) + \ldots. \\
&= \sum_{n=0}^{\infty} \frac{\theta^n \mathbb{E}(X^n)}{n!}.
\end{aligned}
$$

Calculating Moments

The k^{th} moment m_k of the random variable X can now be obtained by differentiating, i.e.

$$\begin{aligned} m_k &= M_X^{(k)}(0)\,; \quad k = 0, 1, 2, \ldots \\ M_X^{(k)}(0) &= \left.\frac{d^k}{d\theta^k} M_X(\theta)\right|_{\theta=0} \end{aligned}$$

A useful result in finance is the MGF for the normal distribution. If $X \sim N\left(\mu, \sigma^2\right)$, then we can construct a standard normal $\phi \sim N(0,1)$ by setting $\phi = \dfrac{X-\mu}{\sigma} \Longrightarrow X = \mu + \sigma\phi$.

The MGF is

$$\begin{aligned} M_\theta(X) &= \mathbb{E}\left[e^{\theta x}\right] = \mathbb{E}\left[e^{\theta(\mu+\phi\sigma)}\right] \\ &= e^{\theta\mu}\mathbb{E}\left[e^{\theta\sigma\phi}\right] \end{aligned}$$

So the MGF of X is therefore equal to the MGF of ϕ but with θ replaced by $\theta\sigma$.This is much nicer than trying to calculate the MGF of $X \sim N\left(\mu, \sigma^2\right)$.

$$\begin{aligned}
\mathbb{E}\left[e^{\theta\phi}\right] &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\theta x} e^{-x^2/2} dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\theta x - x^2/2} dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(x^2-2\theta x+\theta^2-\theta^2\right)} dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\theta)^2+\frac{1}{2}\theta^2} dx \\
&= e^{\frac{1}{2}\theta^2}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\theta)^2} dx
\end{aligned}$$

Now do a change of variable - put $u = x - \theta$

$$\begin{aligned}
\mathbb{E}\left[e^{\theta\phi}\right] &= e^{\frac{1}{2}\theta^2}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2} du \\
&= e^{\frac{1}{2}\theta^2}
\end{aligned}$$

Thus

$$\begin{aligned}
M_\theta(X) &= e^{\theta\mu}\mathbb{E}\left[e^{\theta\sigma\phi}\right] \\
&= e^{\theta\mu+\frac{1}{2}\theta^2\sigma^2}
\end{aligned}$$

To get the simpler formula for a standard normal distribution put $\mu = 0,\ \sigma = 1$ to get $M_\theta(X) = e^{\frac{1}{2}\theta^2}$.

We can now obtain the first four moments for a standard normal

$$
\begin{aligned}
m_1 &= \left.\frac{d}{d\theta}e^{\frac{1}{2}\theta^2}\right|_{\theta=0} \\
&= \left.\theta e^{\frac{1}{2}\theta^2}\right|_{\theta=0} = 0
\end{aligned}
$$

$$
\begin{aligned}
m_2 &= \left.\frac{d^2}{d\theta^2}e^{\frac{1}{2}\theta^2}\right|_{\theta=0} \\
&= \left.\left(\theta^2+1\right)e^{\frac{1}{2}\theta^2}\right|_{\theta=0} = 1
\end{aligned}
$$

$$
\begin{aligned}
m_3 &= \left.\frac{d^3}{d\theta^3}e^{\frac{1}{2}\theta^2}\right|_{\theta=0} \\
&= \left.\left(\theta^3+3\theta\right)e^{\frac{1}{2}\theta^2}\right|_{\theta=0} = 0
\end{aligned}
$$

$$
\begin{aligned}
m_4 &= \left.\frac{d^4}{d\theta^4}e^{\frac{1}{2}\theta^2}\right|_{\theta=0} \\
&= \left.\left(\theta^4+6\theta^2+3\right)e^{\frac{1}{2}\theta^2}\right|_{\theta=0} = 3
\end{aligned}
$$

The latter two are particularly useful in calculating the skew and kurtosis.

3.4 Correlation

The covariance is useful in studying the statsitical dependence between two random. If $X,\ Y$ are RV's, then their covariance is defined as:

$$
\begin{aligned}
\mathrm{Cov}\,(X,\ Y) &= \mathbb{E}\left[\left(X - \underbrace{\mathbb{E}\,(X)}_{=\mu_x}\right)\left(Y - \underbrace{\mathbb{E}\,(Y)}_{=\mu_y}\right)\right] \\
&= \mathbb{E}\,[XY] - \mu_x \mu_y
\end{aligned}
$$

which we denote as σ_{XY}. **Note:**

$$
\mathrm{Cov}\,(X,\ X) = \mathbb{E}\left[(X - \mu_x)^2\right] = \sigma^2.
$$

$X,\ Y$ are *correlated* if

$$
\mathbb{E}\left[(X - \mu_x)\left(Y - \mu_y\right)\right] \neq 0.
$$

We can then define an important dimensionless quantity (used in finance) called the *correlation coefficient* and denoted as $\rho_{XY}\,(X,\ Y)$ where

$$
\rho_{XY} = \frac{\mathrm{Cov}\,(X,\ Y)}{\sigma_x \sigma_y}.
$$

The correlation can be thought of as a normalised covariance, as $|\rho_{XY}| \leq 1$, for which the following conditions are properties:

i. $\rho(X,\ Y) = \rho(Y,\ X)$

ii. $\rho(X,\ \pm X) = \pm 1$

iii. $-1 \leq \rho \leq 1$

$\rho_{XY} = -1 \Rightarrow$ perfect negative correlation

$\rho_{XY} = 1 \Rightarrow$perfect correlation

$\rho_{XY} = 0 \Rightarrow X,\ Y$ uncorrelated

Why is the correlation coefficient bounded by ± 1? Justification of this requires a result called the *Cauchy-Schwartz inequality*. This is a theorem which most students encounter for the first time in linear algebra (although we have not discussed this). Let's start off with the version for random variables (RVs) X and Y, then the Cauchy-Schwartz inequality is

$$[\mathbb{E}[XY]]^2 \leq \mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right].$$

We know that the covariance of X, Y is

$$\sigma_{XY} = \mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right]$$

If we put

$$\begin{aligned} \mathbb{V}[X] &= \sigma_X^2 = \mathbb{E}\left[(X-\mu_X)^2\right] \\ \mathbb{V}[Y] &= \sigma_Y^2 = \mathbb{E}\left[(Y-\mu_Y)^2\right]. \end{aligned}$$

From Cauchy-Schwartz we have

$$(\mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right])^2 \leq \mathbb{E}\left[(X-\mu_X)^2\right]\mathbb{E}\left[(Y-\mu_Y)^2\right]$$

or we can write

$$\sigma_{XY}^2 \leq \sigma_X^2\sigma_Y^2$$

Divide through by $\sigma_X^2\sigma_Y^2$

$$\frac{\sigma_{XY}^2}{\sigma_X^2\sigma_Y^2} \leq 1$$

and we know that the left hand side above is ρ_{XY}^2, hence

$$\rho_{XY}^2 = \frac{\sigma_{XY}^2}{\sigma_X^2\sigma_Y^2} \leq 1$$

and since ρ_{XY} is a real number, this implies $|\rho_{XY}| \leq 1$ which is the same as

$$-1 \leq \rho_{XY} \leq +1.$$

4 Differential Equations

4.1 Introduction

2 Types of Differential Equation (D.E)

(i) Ordinary Differential Equation (O.D.E)

Equation involving (ordinary) derivatives

$$x,\ y,\ \frac{dy}{dx},\ \frac{d^{\,2}y}{dx^2},\ \ldots\ldots\ldots,\ \frac{d^{\,n}y}{dx^n} \qquad \text{(some fixed } n\text{)}$$

y is some unknown function of x together with its derivatives, i.e.

$$F\left(x,\ y,\ y',\ y'',\ \ldots\ldots,\ y^{\,(n)}\right) = 0 \qquad (1)$$

Note $y^4 \neq y^{(4)}$

Also if $y = y\,(t)$, where t is time, then we often write

$$\dot{y} = \frac{dy}{dt}, \quad \ddot{y} = \frac{d^{\,2}y}{dt^2}, \quad \ldots\ldots, \quad \ddot{\ddot{y}} = \frac{d^{\,4}y}{dt^4}$$

(ii) Partial Differential Equation (PDE)

Involve partial derivatives, i.e. unknown function dependent on two or more variables,

e.g.

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

More complicated to solve - better for modelling real-life situations, e.g. finance, engineering & science.

Order of the highest derivative is the **order of the DE**

An ode is of **degree** r if $\dfrac{d\,^n y}{dx^n}$ (where n is the order of the derivative) appears with power r

$\left(r \epsilon \mathbb{Z}^+\right)$ – the definition of n and r is distinct. Assume that any ode has the property that each

$\dfrac{d^\ell y}{dx^\ell}$ appears in the form $\left(\dfrac{d^\ell y}{dx^\ell}\right)^r \to \left(\dfrac{d^n y}{dx^n}\right)^r$ order n and degree r.

Examples:

	DE	order	degree
(1)	$y' = 3y$	1	1
(2)	$(y')^3 + 4\sin y = x^3$	1	3
(3)	$\left(y^{(4)}\right)^2 + x^2\left(y^{(2)}\right)^5 + (y')^6 + y = 0$	4	2
(4)	$y'' = \sqrt{y' + y + x}$	2	2
(5)	$y'' + x\,(y')^3 - xy = 0$	2	1

Note - example (4) above can be written as $(y'')^2 = y' + y + x$

We will consider ODE's of degree one, and of the form

$$a_n(x)\frac{d^n y}{dx^n}+a_{n-1}(x)\frac{d^{n-1}y}{dx^{n-1}}+....+a_1(x)\frac{dy}{dx}+a_0(x)\,y=g(x)$$

$$\equiv \sum_{i=0}^{n} a_i(x)\,y^{(i)}(x)=g(x) \quad \text{(more pedantic)}$$

Note: $y^{(0)}(x)$ - zeroth derivative, i.e. $y(x)$.

This is a <u>Linear</u> ODE of order n, i.e. $r = 1$ $\forall$ (for all) terms. Linear also because $a_i(x)$ not a function of $y^{(i)}(x)$ - else equation is Non-linear.

Examples:

	DE	Nature of DE
(1)	$2xy'' + x^2y' - (\sin x)\, y = x^2$	Linear
(2)	$yy'' + xy' + y = 2$	$a_2 = y \Rightarrow$ Non-Linear
(3)	$y'' + \sqrt{y'} + y = x^2$	Non-Linear $\because (y')^{\frac{1}{2}}$
(4)	$\dfrac{d^4y}{dx^4} + y^4 = 0$	Non-Linear - y^4

Our aim is to solve our ODE either explicitly or by finding the most general $y\,(x)$ satisfying it or implicitly by finding the function y implicitly in terms of x, via the most general function g s.t $g\,(x, y) = 0$.

Suppose that y is given in terms of x and n arbitrary constants of integration c_1, c_2, ..., c_n.

So $\widetilde{g}\,(x,\ c_1,\ c_2, \ldots\ldots., c_n) = 0$. Differentiating $\widetilde{g}$, n times to get $(n+1)$ equations involving

$$c_1,\ c_2, \ldots\ldots., c_n,\ x,\ y,\ y',\ y'',\ \ldots\ldots,\ y^{\,(n)}.$$

Eliminating $c_1,\ c_2, \ldots\ldots., c_n$ we get an ODE

$$\widetilde{f}\ \left(x,\ y,\ y',\ y'',\ \ldots\ldots,\ y^{\ (n)}\right) = 0$$

Examples:

(1) $y = x^3 + ce^{-3x}$ (so 1 constant c)

$\Rightarrow \dfrac{dy}{dx} = 3x^2 - 3ce^{-3x}$, so eliminate c by taking $3y + y' = 3x^3 + 3x^2$

i.e.

$$-3x^2\,(x+1) + 3y + y' = 0$$

(2) $y = c_1e^{-x} + c_2e^{2x}$ (2 constant's so differentiate twice)

$y' = -c_1e^{-x} + 2c_2e^{2x} \Rightarrow y'' = c_1e^{-x} + 4c_2e^{2x}$

Now

$$\left.\begin{array}{ll} y + y' = 3c_2e^{2x} & \text{(a)} \\ y' + y'' = 6c_2e^{2x} & \text{(b)} \end{array}\right\}$$

and 2(a)=(b) $\therefore$ 2 $(\ y + y') = y + y'' \rightarrow$

$$y'' - 2y' - y = 0.$$

Conversely it can be shown (under suitable conditions) that the general solution of an n$^{\text{th}}$ order ode will involve n arbitrary constants. If we specify values (i.e. boundary values) of

$$y,\ y',\ \ldots\ldots\ldots\ldots,\ y^{(n)}$$

for values of x, then the constants involved may be determined.

A solution $y = y(x)$ of (0.1) is a function that produces zero upon substitution into the lhs of (1).

Example:

$y'' - 3y' + 2y = 0$ is a 2nd order equation and $y = e^x$ is a solution.

$y = y' = y'' = e^x$ - substituting in equation gives $e^x - 3e^x + 2e^x = 0$. So we can verify that a function is the solution of a DE simply by substitution.

Exercise:

(1) Is $y(x) = c_1 \sin 2x + c_2 \cos 2x$ (c_1, c_2 arbitrary constants) a solution of $y'' + 4y = 0$

(2) Determine whether $y = x^2 - 1$ is a solution of $\left(\dfrac{dy}{dx}\right)^4 + y^2 = -1$

4.1.1 Initial & Boundary Value Problems

A DE together with conditions, an unknown function $y(x)$ and its derivatives, all given at the same value of independent variable x is called an **Initial Value Problem** (IVP).

e.g. $y'' + 2y' = e^x$; $y(\pi) = 1$, $y'(\pi) = 2$ is an IVP because both conditions are given at the same value $x = \pi$.

A **Boundary Value Problem** (BVP) is a DE together with conditions given at different values of x, i.e. $y'' + 2y' = e^x$; $y(0) = 1$, $y(1) = 1$.

Here conditions are defined at different values $x = 0$ and $x = 1$.

A solution to an IVP or BVP is a function $y(x)$ that both solves the DE and satisfies all given initial or boundary conditions.

Exercise: Determine whether any of the following functions

(a) $y_1 = \sin 2x$ (b) $y_2 = x$ (c) $y_3 = \frac{1}{2}\sin 2x$ is a solution of the IVP

$$y'' + 4y = 0; \quad y(0) = 0, \quad y'(0) = 1$$

4.2 First Order Ordinary Differential Equations

Standard form for a first order DE (in the unknown function $y(x)$) is

$$y' = f(x, y) \tag{2}$$

so given a 1^{st} order ode

$$F\left(x, y, y'\right) = 0$$

can often be rearranged in the form (2), e.g.

$$xy' + 2xy - y = 0 \Rightarrow \quad y' = \frac{y - 2x}{x}$$

4.2.1 One Variable Missing

This is the simplest case

y missing:

$$y' = f(x) \qquad \text{solution is } y = \int f(x)dx$$

x missing:

$$y' = f(y) \qquad \text{solution is } x = \int \frac{1}{f(y)}dy$$

Example:

$$y' = \cos^2 y \ , \quad y = \frac{\pi}{4} \text{ when } x = 2$$

$$\Rightarrow x = \int \frac{1}{\cos^2 y}dy = \int \sec^2 y \ dy \Rightarrow x = \tan y + c \ ,$$

c is a constant of integration.

This is the general solution. To obtain a particular solution use

$$y\,(2) = \frac{\pi}{4} \rightarrow 2 = \tan\frac{\pi}{4} + c \Rightarrow c = 1$$

so rearranging gives

$$y = \arctan\,(x - 1)$$

4.2.2 Variable Separable

$$y' = g(x)\,h(y) \tag{3}$$

So $f(x,y) = g(x)\,h(y)$ where g and h are functions of x only and y only in turn. So

$$\frac{dy}{dx} = g(x)\,h(y) \rightarrow \int \frac{dy}{h(y)} = \int g(x)\,dx + c$$

c – arbitrary constant

Examples:

1. $\dfrac{dy}{dx} = \dfrac{x^2+2}{y}$

$$\int y\;dy = \int \left(x^2 + 2\right) dx \rightarrow \frac{y^2}{2} = \frac{x^3}{3} + 2x + c$$

2. $\dfrac{dy}{dx} = y \ln x$ subject to $y = 1$ at $x = e$ ($y(e) = 1$)

$$\int \frac{dy}{y} = \int \ln x \; dx \quad \text{Recall: } \int \ln x \; dx = x(\ln x - 1)$$

$$\ln y = x(\ln x - 1) + c \rightarrow y = A \exp(x \ln x - x)$$

$A -$ arb. constant

now putting $x = e$, $y = 1$ gives $A = 1$. So solution becomes

$$y = \exp(\ln x^x) \exp(-x) \rightarrow y = \frac{x^x}{e^x} \Rightarrow y = \left(\frac{x}{e}\right)^x$$

4.2.3 Linear Equations

These are equations of the form

$$y' + P(x)\,y = Q(x) \tag{4}$$

which are similar to (3), but the presence of $Q(x)$ renders this no longer separable. We look for a function $R(x)$, called an **Integrating Factor** (I.F) so that

$$R(x)y' + R(x)P(x)\,y = \frac{d}{dx}(R(x)y) \tag{5}$$

So upon multiplying the lhs of (4), it becomes a derivative of $R(x)y$, i.e.

$$Ry' + RPy = Ry' + R'y$$

from (4).

This gives $RPy = R'y \Rightarrow R(x)P(x) = \dfrac{dR}{dx}$, which is a DE for R which is separable, hence

$$\int \frac{dR}{R} = \int P dx + c \rightarrow \ln R = \int P dx + c$$

So $R(x) = K \exp\left(\int P\ dx\right)$, hence there exists a function $R(x)$ with the required property.

Multiply (4) through by $R(x)$

$$\underbrace{R(x)\left(y' + P(x)y\right)}_{=\frac{d}{dx}(R(x)y)} = R(x)Q(x)$$

$$\frac{d}{dx}(Ry) = R(x)Q(x) \to R(x)y = \int R(x)Q(x)dx + B$$

B – arb. constant.

We also know the form of $R(x) \to$

$$yK\exp\left(\int P\ dx\right) = \int K\exp\left(\int P\ dx\right)Q(x)dx + B$$

divide through by K to give

$$y\,\mathrm{exp}\left(\int P\ dx\right) = \int \mathrm{exp}\left(\int P\ dx\right) Q(x)dx + \text{ constant.}$$

So we can take $K = 1$ in the expression for $R(x)$.

To solve $y' + P(x)\,y = Q(x)$ calculate $\boxed{R(x) = \mathrm{exp}\left(\int P\ dx\right)}$, which is the I.F

Examples:

1. Solve $xy' - y = x^3$

This is currently not in standard form. However, dividing through by x gives

$$y' - \frac{1}{x}y = x^2$$

Now comparing with (4) gives $P(x) \equiv -\dfrac{1}{x}$ & $Q(x) \equiv x^2$, therefore

I.F $R(x) = \exp\left(\int -\frac{1}{x}\, dx\right) = \exp\left(-\ln x\right) = \frac{1}{x}$.

Multiply DE by $\frac{1}{x} \to$

$$
\begin{aligned}
\frac{1}{x}\left(y' - \frac{1}{x}y\right) &= x \Rightarrow \frac{d}{dx}\left(\frac{y}{x}\right) = x \to \int d\left(x^{-1}y\right) \\
&= \int x dx + c
\end{aligned}
$$

$$\Rightarrow \frac{y}{x} = \frac{x^2}{2} + c \therefore \text{GS is } y = \frac{x^3}{2} + cx$$

2. Obtain the general solution of $(1 + ye^x)\dfrac{dx}{dy} = e^x$

$$\frac{dy}{dx} = (1 + ye^x)\, e^{-x} = e^{-x} + y \Rightarrow$$

$$\frac{dy}{dx} - y = e^{-x}$$

Which is a linear equation, with $P = -1$; $Q = e^{-x}$

$$\text{I.F } R(y) = \exp\left(\int -dx\right) = e^{-x}$$

so multiplying DE by I.F

$$\begin{aligned} e^{-x}\left(y' - y\right) &= e^{-2x} \rightarrow \frac{d}{dx}\left(ye^{-x}\right) = e^{-2x} \Rightarrow \\ \int d\left(ye^{-x}\right) &= \int e^{-2x} dx \end{aligned}$$

$$ye^{-x} = -\frac{1}{2}e^{-2x} + c$$

$$\therefore y = ce^{x} - \frac{1}{2}e^{-x} \text{ is the GS}$$

4.3 Second Order ODE's

Typical second order ODE (degree 1) is

$$y'' = f\left(x, y, y'\right)$$

solution involves two arbitrary constants.

4.3.1 Simplest Cases

A y', y missing, so $\boxed{y'' = f\ (x)}$

Integrate wrt x (twice): $y = \int \left(\int f\ (x)\ dx\right) dx$

Example: $y'' = 4x$

GS $y = \displaystyle\int \left(\int 4x\ dx\right) dx = \int \left[2x^2 + C\right] dx = \frac{2x^3}{3} + Cx + D$

B y missing, so $\boxed{y'' = f\ (y',\ x)}$

Put $P = y' \rightarrow y'' = \dfrac{dP}{dx} = f\ (P, x)$, i.e. $P' = f\ (P, x)$ - first order ode

Solve once $\rightarrow P(x)$

Solve again $\rightarrow y(x)$

Example: Solve $x\dfrac{d^{\,2}y}{dx^2} + 2\dfrac{dy}{dx} = x^3$

Note: $\boxed{\text{A}}$ is a special case of $\boxed{\text{B}}$

C y' and x missing, so $\boxed{y'' = f\ (y)}$

Put $p = y'$, then

$$\begin{aligned} \frac{d^{\,2}y}{dx^2} &= \frac{dp}{dx} = \frac{dp}{dy}\frac{dy}{dx} = p\frac{dp}{dy} \\ &= f\ (y) \end{aligned}$$

So solve 1st order ode

$$p\frac{dp}{dy} = f\left(y\right)$$

which is separable, so

$$\int p\ dp = \int\ f\ \ (\ y)\ \ dy \longrightarrow$$

$$\frac{1}{2}p^2 = \int f\left(y\right) dy + \text{const.}$$

Example: Solve $y^3 y'' = 4$

$\Rightarrow y'' = \dfrac{4}{y^3}$. Put $p = y' \longrightarrow \dfrac{d^2y}{dx^2} = p\dfrac{dp}{dy} = \dfrac{4}{y^3}$

$\therefore\ \int p\ dp\ =\ \displaystyle\int \frac{4}{y^3}\ dy\ \Rightarrow\ p^2\ =\ -\frac{4}{y^2} + D\quad \therefore\quad p\ =$ $\dfrac{\pm\sqrt{Dy^2 - 4}}{y}$, so from our definition of p,

$$\frac{dy}{dx} = \frac{\pm\sqrt{Dy^2 - 4}}{y} \Rightarrow \int dx = \int \frac{\pm\ y}{\sqrt{Dy^2 - 4}} dy$$

Integrate rhs by substitution (i.e. $u = Dy^2 - 4$) to give

$$x = \frac{\pm\sqrt{Dy^2 - 4}}{D} + E \rightarrow \left[D\left(x - E\right)^2\right] = Dy^2 - 4$$

$$\therefore \text{GS is } Dy^2 - D^2\left(x - E\right)^2 = 4$$

D x missing: $\quad y'' = f\ \ (y',\ y)$

Put $\ P = y'$, so $\dfrac{d^2y}{dx^2} = P\dfrac{dP}{dy} = f\ \ (P, y)$ - 1st order ODE

4.3.2 Linear ODE's of Order at least 2

General n$^{\text{th}}$ order linear ode is of form:

$a_n(x)\,y^{(n)}+a_{n-1}(x)\,y^{(n-1)}+.......+a_1(x)\,y'+a_0(x)\,y = g(x)$

Use symbolic notation:

$$D \equiv \frac{d}{dx}\ ;\ \ D^r \equiv \frac{d^{\ r}}{dx^{\ r}} \quad \text{so } D^r y \equiv \frac{d^{\ r} y}{dx^{\ r}}$$

$\therefore\ \ a_r D^r \equiv a_r(x)\dfrac{d^{\ r}}{dx^{\ r}}$ so

$$a_r D^r y = a_r(x)\frac{d^{\ r} y}{dx^{\ r}}$$

Now introduce

$$L = a_n D^n + a_{n-1} D^{n-1} + a_{n-2} D^{n-2} + \dots\dots\dots\dots + a_1 D + a_0$$

so we can write a linear ode in the form

$$L\ y = g$$

$L-$ Linear Differential Operator of order n and its definition will be used throughout.

If $g(x) = 0\ \forall x$, then $L\ y = 0$ is said to be **HOMOGENEOUS**.

$L\ y = 0$ is said to be the homogeneous part of $L\ y = g$.

L is a linear operator because as is trivially verified:

(1) $L\ (y_1 + y_2) = L\ (y_1) + L(y_2)$

(2) $L\ (cy) = cL(y) \qquad c \in \mathbb{R}$

GS of $Ly = g$ is given by

$$\boxed{y = y_c + y_p}$$

where y_c- Complimentary Function & y_p- Particular Integral (or Particular Solution)

$$\left.\begin{array}{l} y_c \text{ is solution of } \ Ly = 0 \\ y_p \text{ is solution of } \ Ly = g \end{array}\right\} \therefore \ \text{GS } y = y_c + y_p$$

Look at homogeneous case $Ly = 0$. Put Ⓢ $=$ all solutions of $Ly = 0$. Then Ⓢ forms a vector space of dimension n. Functions $y_1(x)$, ..., $y_n(x)$ are LINEARLY DEPENDENT if $\exists\ \lambda_1$, ..., $\lambda_n \in \mathbb{R}$ (not all zero) s.t

$$\lambda_1 y_1(x) + \lambda_2 y_2(x) \ + ... + \ \lambda_n y_n(x) = 0$$

Otherwise y_i's $(i = 1,\ \ldots..,\ n)$ are said to be LINEARLY INDEPENDENT (Lin. Indep.) $\Rightarrow$

whenever

$$\lambda_1 y_1(x) + \lambda_2 y_2(x) + \ldots\ldots\ldots. + \lambda_n y_n(x) = 0 \quad \forall x$$

then $\lambda_1 = \lambda_2 = \ \ldots\ldots... = \ \lambda_n = 0.$

FACT:

(1) $L-$ n^{th} order linear operator, then $\exists\ n$ Lin. Indep. solutions $y_1,\ \ldots..,\ y_n$ of $Ly = 0$ s.t GS of $Ly = 0$ is given by

$$y = \lambda_1 y_1 + \lambda_2 y_2 + \ldots\ldots\ldots. + \lambda_n y_n \qquad \underset{1 \le i \le n}{\lambda_i \in \mathbb{R}}\ .$$

(2) Any n Lin. Indep. solutions of $Ly = 0$ have this property.

To solve $Ly = 0$ we need only find by "hook or by crook" n Lin. Indep. solutions.

4.3.3 Linear ODE's with Constant Coefficients

Consider Homogeneous case: $Ly = 0$.

All basic features appear for the case $n = 2$, so we analyse this.

$$L\ y = a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = 0 \qquad a,\ b,\ c \in \mathbb{R}$$

Try a solution of the form $\qquad y = \exp(\lambda x)$

$$L\left(e^{\lambda x}\right) = \left(aD^2 + bD + c\right) e^{\lambda x}$$

hence $a\lambda^2 + b\lambda + c = 0$ and so λ is a root of the quadratic equation

$a\lambda^2 + b\lambda + c = 0$ **AUXILLIARY EQUATION (A.E)**

There are three cases to consider:

(1) $b^2 - 4ac > 0$

So $\lambda_1 \neq \lambda_2 \in \mathbb{R}$, so GS is

$$\boxed{y = c_1 \exp(\lambda_1 x) + c_2 \exp(\lambda_2 x)}$$

c_1, c_2 $-$ arb. const.

(2) $b^2 - 4ac = 0$

So $\lambda = \lambda_1 = \lambda_2 = -\dfrac{b}{2a}$

Clearly $e^{\lambda x}$ is a solution of $L\,y = 0$ - but theory tells us there exist two solutions for a 2nd

order ode. So now try $y = x \exp(\lambda x)$

$$\begin{aligned} L\left(xe^{\lambda x}\right) &= \left(aD^2 + bD + c\right)\left(xe^{\lambda x}\right) \\ &= \underbrace{\left(a\lambda^2 + b\lambda + c\right)}_{=0}\left(xe^{\lambda x}\right) + \underbrace{(2a\lambda + b)}_{=0}\left(e^{\lambda x}\right) \\ &= 0 \end{aligned}$$

This gives a 2nd solution $\therefore$ GS is $y = c_1 \exp(\lambda x) + c_2 x \exp(\lambda x)$, hence

$$\boxed{y = (c_1 + c_2 x)\exp(\lambda x)}$$

(3) $b^2 - 4ac < 0$

So $\lambda_1 \neq \lambda_2 \in \mathbb{C}$ - Complex conjugate pair $\lambda = p \pm iq$ where

$$p = -\frac{b}{2a}, \qquad q = \frac{1}{2a}\sqrt{\left|b^2 - 4ac\right|} \quad (\neq 0)$$

Hence

$$\begin{aligned} y &= c_1 \exp(p + iq)\,x + c_2 \exp(p - iq)\,x \\ &= c_1 e^{px} e^{iqx} + c_2 e^{px} e^{-iqx} = e^{px}\left(c_1 e^{iqx} + c_2 e^{-iqx}\right) \end{aligned}$$

Eulers identity gives $\exp(\pm i\theta) = \cos\theta \pm i \sin\theta$

Simplifying (using Euler) then gives the GS

$$y(x) = e^{px}(A\cos qx + B\sin qx)$$

Examples:

(1) $y'' - 3y' - 4y = 0$

Put $y = e^{\lambda x}$ to obtain A.E

A.E: $\lambda^2 - 3\lambda - 4 = 0 \rightarrow (\lambda - 4)(\lambda + 1) = 0 \Rightarrow$ $\lambda = 4$ & -1 - 2 distinct $\mathbb{R}$ roots

$$\text{GS} \quad y(x) = Ae^{4x} + Be^{-x}$$

(2) $y'' - 8y' + 16y = 0$

A.E $\lambda^2 - 8\lambda + 16 = 0 \to (\lambda - 4)^2 = 0 \Rightarrow \lambda = 4$
, 4 (2 fold root)

'go up one', i.e. instead of $y = e^{\lambda x}$, take $y = xe^{\lambda x}$

$$\text{GS} \quad y(x) = (C + Dx)\, e^{4x}$$

(3) $y'' - 3y' + 4y = 0$

A.E $\lambda^2 - 3\lambda + 4 = 0 \to \lambda = \dfrac{3 \pm \sqrt{9 - 16}}{2} = \dfrac{3 \pm i\sqrt{7}}{2}$

$\lambda_1 = \dfrac{3 + i\sqrt{7}}{2}, \ \ \lambda_2 = \dfrac{3 - i\sqrt{7}}{2} \ \equiv p \pm iq$

$\left(p = \dfrac{3}{2}, \ \ q = \dfrac{\sqrt{7}}{2}\right)$

$$y = e^{\frac{3}{2}x}\left(a\cos\frac{\sqrt{7}}{2}x + b\sin\frac{\sqrt{7}}{2}x\right)$$

4.4 General n^{th} Order Equation

Consider

$$L\,y = a_n y^{(n)} + a_{n-1} y^{(n-1)} + \dots\dots\dots + a_1 y' + a_0 y = 0$$

then

$$\begin{aligned} L &\equiv D^n + \widehat{a}_{n-1} D^{n-1} + \widehat{a}_{n-2} D^{n-2} + \dots\dots\dots \\ &\quad + \widehat{a}_1 D + \widehat{a}_0 \\ \widehat{a}_i &\in \mathbb{R}\ (0 \le i \le n-1) \end{aligned}$$

$\left(\text{we have divided through by } a_n, \text{ i.e. } \widehat{a}_i = \dfrac{a_i}{a_n}\right)$ so $L\,y =$ 0

A.E becomes $\boxed{\lambda^n + \widehat{a}_{n-1}\lambda^{n-1} + \dots\dots\dots\dots + \widehat{a}_1\lambda + \widehat{a}_0 = 0}$

$\boxed{\text{Case 1}}$ (Basic)

n distinct roots λ_1, ..., λ_n then $e^{\lambda_1 x}$, $e^{\lambda_2 x}$, ..., $e^{\lambda_n x}$ are n Lin. Indep. solutions giving a GS

$$y = \beta_1 e^{\lambda_1 x} + \beta_2 e^{\lambda_2 x} + ... + \beta_n e^{\lambda_n x}$$

$\beta_i -$ arb.

$\boxed{\text{Case 2}}$

If λ is a real $r-$ fold root of the A.E then $e^{\lambda x}$, $xe^{\lambda x}$, $x^2 e^{\lambda x}$,.........., $x^{r-1} e^{\lambda x}$ are r Lin. Indep. solutions of $Ly = 0$, i.e.

$$y = e^{\lambda x}\left(\alpha_1 + \alpha_2 x + \alpha_3 x^2 .... + \alpha_r \, x^{r-1}\right)$$

$\alpha_i -$ arb.

$\boxed{\text{Case 3}}$

If $\lambda = p + iq$ is a r - fold root of the A.E then so is $p - iq$

$$\left.\begin{array}{c} e^{px}\cos qx,\ \ xe^{px}\cos qx,\ \ldots\ldots\ldots\ldots, x^{r-1}e^{px}\cos qx \\ e^{px}\sin qx,\ xe^{px}\sin qx,\ \ldots\ldots\ldots\ldots, x^{r-1}e^{px}\sin qx \end{array}\right\}$$

$$\longrightarrow \quad 2r \text{ Lin. Indep. solutions of } L\ y = 0$$

$$\begin{aligned} \text{GS}\ \ y \ &= \ e^{px}\left(c_1 + c_2x + c_3x^2 + \ldots\ldots\ldots\ldots\right)\cos qx + \\ &\quad\ e^{px}\left(C_1 + C_2x + C_3x^2 + \ldots\ldots\ldots\ldots\right)\sin qx \end{aligned}$$

$\boxed{\text{Examples:}}$ Find the GS of each ODE

(1) $y^{(4)} - 5y'' + 6y = 0$

A.E: $\lambda^4 - 5\lambda^2 + 6 = 0 \rightarrow \left(\lambda^2 - 2\right)\left(\lambda^2 - 3\right) = 0$

So $\lambda = \pm\sqrt{2}$, $\lambda = \pm\sqrt{3}$ - four distinct roots

$\therefore$ GS $y = Ae^{\sqrt{2}x} + Be^{-\sqrt{2}x} + Ce^{\sqrt{3}x} + De^{-\sqrt{3}x}$ (Case 1)

(2) $\dfrac{d^{\,6}y}{dx^6} - 5\dfrac{d^{\,4}y}{dx^4} = 0$

A.E: $\lambda^6 - 5\lambda^4 = 0$ roots: 0, 0, 0, 0, $\pm\sqrt{5}$

GS $y = Ae^{\sqrt{5}x} + Be^{-\sqrt{5}x} + \left(C + Dx + Ex^2 + Fx^3\right)$ ($\because$ exp(0) $= 1$)

(3) $\dfrac{d^{\,4}y}{dx^4} + 2\dfrac{d^{\,2}y}{dx^2} + y = 0$

A.E: $\lambda^4 + 2\lambda^2 + 1 = \left(\lambda^2 + 1\right)^2 = 0 \qquad \lambda = \pm i$ is a 2 fold root.

Example of Case (3)

$$y = A\cos x + Bx\cos x + C\sin x + Dx\sin x$$

4.5 Non-Homogeneous Case - Method of Undetermined Coefficients

$$\text{GS} \quad y = \text{C.F} + \text{P.I}$$

C.F comes from the roots of the A.E

There are three methods for finding P.I

(a) "Guesswork" - which we are interested in

(b) Annihilator

(c) D-operator Method

(a) Guesswork Method

If the rhs of the ode $g(x)$ is of a certain type, we can guess the form of P.S. We then try it out and determine the numerical coefficients.

The method will work when $g(x)$ has the following forms

i. Polynomial in x $\quad g(x) = p_0 + p_1 x + p_2 x^2 + \ldots\ldots\ldots + p_m x^m$.

ii. An exponential $\quad g(x) = Ce^{kx}$ $\quad$ (Provided k is <u>not</u> a root of A.E).

iii. Trigonometric terms, $g(x)$ has the form $\sin ax$, $\cos ax$ (Provided ia is <u>not</u> a root of A.E).

iv. $g(x)$ is a combination of i. , ii. , iii. provided $g(x)$ does not contain part of the C.F (in which case use other methods).

Examples:

(1) $y'' + 3y' + 2y = x^2$

GS $y = \text{C.F} + \text{P.I} = y_c + y_p$

C.F: A.E gives

$\lambda^2 + 3\lambda + 2 = 0 \Rightarrow \lambda = -1,\ -2 \therefore y_c = ae^{-x} + be^{-2x}$

P.I Now $g(x) = x^2$,

so try $y_p = p_0 + p_1 x + p_2 x^2 \quad \rightarrow y_p' = p_1 + 2p_2 x \quad \rightarrow y_p'' = 2p_2$

Now substitute these in to the DE, ie

$2p_2 + 3(p_1 + 2p_2 x) + 2\left(p_0 + p_1 x + p_2 x^2\right) = x^2$ and equate coefficients of x^n

$O\left(x^2\right):\qquad 2p_2 = 1 \Rightarrow p_2 = \frac{1}{2}$

$O(x)$: $\qquad 6p_2 + 2p_1 = 0 \Rightarrow p_1 = -\frac{3}{2}$

$O\left(x^0\right)$: $\qquad 2p_2 + 3p_1 + 2p_0 = 0 \Rightarrow p_0 = \frac{7}{4}$

$$\therefore \text{GS} \quad y = ae^{-x} + be^{-2x} + \frac{7}{4} - \frac{3}{2}x + \frac{1}{2}x^2$$

(2) $y'' + 3y' + 2y = 3e^{5x}$

The homogeneous part is the same as in (1), so $y_C = Ae^{-x} + Be^{-2x}$. For the non-homog. part we note that $g(x)$ has the form e^{kx}, so try $y_p = Ce^{5x}$, and $k = 5$ is not a solution of the A.E.

Substituting y_p into the DE gives

$$C\left(5^2 + 15 + 2\right)e^{5x} = 3e^{5x} \;\rightarrow C = \frac{1}{14}$$

$$\therefore y = Ae^{-x} + Be^{-2x} + \frac{1}{14}e^{5x}$$

(3) $y'' - 5y' - 6y = \cos 3x$

A.E: $\lambda^2 - \lambda - 6 = 0 \Rightarrow \lambda = -1, 6 \Rightarrow y_c = \alpha e^{-x} + \beta e^{6x}$

Guided by the rhs, i.e. $g(x)$ is a trigonometric term, we try

$$y_p = A\cos 3x + B\sin 3x$$

$\rightarrow {y_p}' = -3A\sin 3x + 3B\cos 3x \rightarrow {y_p}'' = -9A\cos 3x - 9B\sin 3x$ and substitute into DE.

Collecting coefficients of $\sin 3x$ and $\cos 3x$ gives:

$O(\cos 3x):\qquad -9A - 15B - 6A = 1$

$O(\sin 3x):\qquad -9B + 15A - 6B = 0$

$A = -\frac{1}{30} = B \rightarrow y_p = -\frac{1}{30}(\cos 3x + \sin 3x)$, so general solution becomes

$$y = \alpha e^{-x} + \beta e^{6x} - \frac{1}{30}(\cos 3x + \sin 3x)$$

4.5.1 Failure Case

Consider the DE $y'' - 5y' + 6y = e^{2x}$, which has a CF given by $y(x) = \alpha e^{2x} + \beta e^{3x}$. To find a

PS, if we try $y_p = Ae^{2x}$, we have upon substitution

$$Ae^{2x}[4 - 10 + 6] = e^{2x}$$

so when $k \ (= 2)$ is also a solution of the C.F , then the trial solution $y_p = Ae^{kx}$ fails, so we must seek the existence of an alternative solution.

The methods given should be used in such cases.

Statement

a) $Ly = y'' + ay' + b = \alpha e^{kx}$ - trial function is normally $y_p = Ce^{kx}$.

If k is a root of the A.E then $L\left(Ce^{kx}\right) = 0$ so this substitution does not work. In this case, we try $y_p = Cxe^{kx}$ - so 'go one up'.

This works provided k is not a repeated root of the A.E, if so try $y_p = Cx^2e^{kx}$, and so forth ....

b) $Ly = g$ where $g(x)$ has the form $(\alpha \sin mx + \beta \cos mx)\, e^{px}$ try

$$y_p = (c_1 \sin mx + c_2 \cos mx)\, e^{px}$$

provided $p + im$ is not a root of the A.E. If $p + im$ is a root then 'go one up' so try

$$y_p = (c_1 \sin mx + c_2 \cos mx)\, xe^{px} \text{ , etc.}$$

c) Finally, if $g(x) = g_1(x) + g_2(x) + g_3(x)$ where

$$\begin{aligned} g_1(x) &= \sum_{k=0}^{m} g_k x^k, \quad g_2(x) = Ce^{kx}, \\ g_3(x) &= (\alpha \sin mx + \beta \cos mx)\, e^{px} \end{aligned}$$

Then try

$$\boxed{y_p = \overline{y}_p(x) + \widetilde{y}_p(x) + \widehat{y}_p(x)}$$

where

$$\begin{aligned} \overline{y}_p(x) &= p_0 + p_1 x + p_2 x^2 + \ldots\ldots\ldots + p_m x^m \\ \widetilde{y}_p(x) &= Ce^{kx} \\ \widehat{y}_p(x) &= (c_1 \sin mx + c_2 \cos mx)\, e^{px} \end{aligned}$$

4.6 Linear ODE's with Variable Coefficients - Euler Equation

In the previous sections we have looked at various second order DE's with constant coefficients. We now introduce a 2$^{\text{nd}}$ order equation in which the coefficients are variable in x. An equation of the form

$$L\,y = ax^2\frac{d^2y}{dx^2} + \beta x\frac{dy}{dx} + cy = g(x)$$

is called a Cauchy-Euler equation. Note the relationship between the coefficient and corresponding derivative term, ie $a_n(x) = ax^n$ and $\dfrac{d^{\,n}y}{dx^n}$, i.e. both power and order of derivative are n.

The equation is still linear. To solve the homogeneous part, we look for a solution of the form

$$y = x^{\lambda}$$

So $y' = \lambda x^{\lambda-1} \rightarrow y'' = \lambda(\lambda-1)x^{\lambda-2}$, which upon substitution yields the quadratic, A.E.

$$a\lambda^2 + b\lambda + c = 0$$

[where $b = (\beta - a)$] which can be solved in the usual way - there are 3 cases to consider, depending upon the nature of $b^2 - 4ac$.

$\boxed{\text{Case 1:}}$ $b^2 - 4ac > 0 \to \lambda_1,\ \lambda_2 \in \mathbb{R}$ - 2 real distinct roots

$$\text{GS } y = Ax^{\lambda_1} + Bx^{\lambda_2}$$

$\boxed{\text{Case 2:}}$ $b^2 - 4ac = 0 \to \lambda = \lambda_1 = \lambda_2 \in \mathbb{R}$ - 1 real (double fold) root

$$\text{GS } y = x^{\lambda}\left(A + B\ln x\right)$$

$\boxed{\text{Case 3:}}$ $b^2 - 4ac < 0 \to \lambda = \alpha \pm i\beta \in \mathbb{C}$ - pair of complex conjugate roots

$$\text{GS } y = x^{\alpha}\left(A\cos\left(\beta\ln x\right) + B\sin\left(\beta\ln x\right)\right)$$

$\boxed{\text{Example 1}}$ Solve $x^2y'' - 2xy' - 4y = 0$

Put $y = x^\lambda \Rightarrow y' = \lambda x^{\lambda-1} \Rightarrow y'' = \lambda(\lambda-1)x^{\lambda-2}$ and substitute in DE to obtain (upon simplification) the A.E. $\quad \lambda^2 - 3\lambda - 4 = 0 \rightarrow (\lambda-4)(\lambda+1) = 0$

$\Rightarrow \lambda = 4$ & -1 : 2 distinct $\mathbb{R}$ roots. So GS is

$$y(x) = Ax^4 + Bx^{-1}$$

$\boxed{\text{Example 2}}$ Solve $\quad x^2y'' - 7xy' + 16y = 0$

So assume $y = x^\lambda$

A.E $\quad \lambda^2 - 8\lambda + 16 = 0 \Rightarrow \lambda = 4\ ,\ 4$ (2 fold root)

'go up one', i.e. instead of $y = x^\lambda$, take $y = x^\lambda \ln x$ to give

$$y(x) = x^4(A + B\ln x)$$

$\boxed{\text{Example 3}}$ Solve $\qquad x^2y'' - 3xy' + 13y = 0$

Assume existence of solution of the form $y = x^\lambda$

A.E becomes $\qquad \lambda^2 - 4\lambda + 13 = 0 \rightarrow \lambda = \dfrac{4 \pm \sqrt{16 - 52}}{2} = \dfrac{4 \pm 6i}{2}$

$\lambda_1 = 2+3i,\ \lambda_2 = 2-3i\ \equiv \alpha \pm i\beta \qquad (\alpha = 2,\ \ \beta = 3)$

$$y = x^2 \left(A \cos\left(3 \ln x\right) + B \sin\left(3 \ln x\right)\right)$$

4.6.1 Reduction to constant coefficient

The Euler equation considered above can be reduced to the constant coefficient problem discussed earlier by use of a suitable transform. To illustrate this simple technique we use a specific example.

Solve $\quad x^2y'' - xy' + y = \ln x$

Use the substitution $x = e^t$ i.e. $t = \ln x$. We now rewrite the the equation in terms of the variable t, so require new expressions for the derivatives (chain rule):

$$\frac{dy}{dx} = \frac{dy}{dt}\frac{dt}{dx} = \frac{1}{x}\frac{dy}{dt}$$

$$\begin{aligned}\frac{d^{\,2}y}{dx^2} &= \frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{d}{dx}\left(\frac{1}{x}\frac{dy}{dt}\right) = \frac{1}{x}\frac{d}{dx}\frac{dy}{dt} - \frac{1}{x^2}\frac{dy}{dt} \\ &= \frac{1}{x}\frac{dt}{dx}\frac{d}{dt}\frac{dy}{dt} - \frac{1}{x^2}\frac{dy}{dt} = \frac{1}{x^2}\frac{d^2y}{dt^2} - \frac{1}{x^2}\frac{dy}{dt}\end{aligned}$$

$\therefore$ the Euler equation becomes

$$x^2\left(\frac{1}{x^2}\frac{d^2y}{dt^2}-\frac{1}{x^2}\frac{dy}{dt}\right)-x\left(\frac{1}{x}\frac{dy}{dt}\right)+y=t \qquad \rightarrow$$

$$y''(t)-2y'(t)+y=t$$

The solution of the homogeneous part , ie C.F. is $y_c = e^t(A+Bt)$.

The particular solution (P.S.) is obtained by using $y_p = p_0 + p_1 t$ to give $y_p = 2+t$

The GS of this equation becomes

$$y(t)=e^t(A+Bt)+2+t$$

which is a function of t . The original problem was $y = y(x)$, so we use our transformation $t=\ln x$ to get the GS

$$y=x(A+B\ln x)+2+\ln x.$$

4.7 Partial Differential Equations

4.7.1 Introduction

The formation (and solution) of PDE's forms the basis of a large number of mathematical models used to study physical situations arising in science, engineering and medicine.

More recently their use has extended to the modelling of problems in finance and economics.

We now look at the second type of DE, i.e. PDE's. These have partial derivatives instead of ordinary derivatives.

One of the underlying equations in finance, the Black-Scholes equation for the price of an option $V(S,t)$ is an example of a linear PDE

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r - D) S \frac{\partial V}{\partial S} - rV = 0$$

providing σ, D, r are not functions of V or any of its derivatives.

4.7.2 Similarity Reduction

Model equation is

$$\frac{\partial p}{\partial t} = c^2 \frac{\partial^2 p}{\partial y^2}$$

for the unknown function $p = p(y,t)$ which is a probability density function.

We assume a solution of the following form exists:

$$p(y,t) = t^{\alpha} f\left(\frac{y}{t^{\beta}}\right)$$

where α, β are constants to be determined. So put

$$\xi = \frac{y}{t^{\beta}}$$

which allows us to obtain the following derivatives

$$\frac{\partial \xi}{\partial y} = \frac{1}{t^{\beta}}; \quad \frac{\partial \xi}{\partial t} = -\beta y t^{-\beta-1}$$

we can now say

$$p(y,t) = t^{\alpha} f(\xi)$$

therefore

$$\frac{\partial p}{\partial y} = \frac{\partial p}{\partial \xi}\frac{\partial \xi}{\partial y} = t^{\alpha} f'(\xi) \cdot \frac{1}{t^{\beta}} = t^{\alpha-\beta} f'(\xi)$$

$$\begin{aligned} \frac{\partial^2 p}{\partial y^2} &= \frac{\partial}{\partial y}\left(\frac{\partial p}{\partial y}\right) = \frac{\partial}{\partial y}\left(t^{\alpha-\beta} f'(\xi)\right) \\ &= \frac{\partial \xi}{\partial y}\frac{\partial}{\partial \xi}\left(t^{\alpha-\beta} f'(\xi)\right) \\ &= t^{\alpha-\beta}\frac{1}{t^{\beta}}\frac{\partial}{\partial \xi} f'(\xi) = t^{\alpha-2\beta} f''(\xi) \end{aligned}$$

$$\frac{\partial p}{\partial t} = t^{\alpha}\frac{\partial}{\partial t} f(\xi) + \alpha t^{\alpha-1} f(\xi)$$

we can use the chain rule to write

$$\frac{\partial}{\partial t} f(\xi) = \frac{\partial f}{\partial \xi} \cdot \frac{\partial \xi}{\partial t} = -\beta y t^{-\beta-1} f'(\xi)$$

so we have

$$\frac{\partial p}{\partial t} = \alpha t^{\alpha-1} f(\xi) - \beta y t^{\alpha-\beta-1} f'(\xi)$$

and then substituting these expressions in to the pde gives

$$\alpha t^{\alpha-1} f(\xi) - \beta y t^{\alpha-\beta-1} f'(\xi) = c^2 t^{\alpha-2\beta} f''.$$

We know from ξ that

$$y = t^{\beta} \xi$$

hence the equation above becomes

$$\alpha t^{\alpha-1} f(\xi) - \beta \xi t^{\alpha-1} f'(\xi) = c^2 t^{\alpha-2\beta} f''.$$

For the similarity solution to exist we require the equation to be independent of t, i.e. $\alpha - 1 = \alpha - 2\beta \Longrightarrow \beta = 1/2$, therefore

$$\alpha f - \frac{1}{2} \xi f' = c^2 f''$$

thus we have so far

$$p = t^{\alpha} f\left(\frac{y}{\sqrt{t}}\right)$$

which gives us a whole family of solutions dependent upon the choice of α.

We know that p represents a pdf, hence

$$\int_{\mathbb{R}} p\left(y,t\right)dy = 1 = \int_{\mathbb{R}} t^{\alpha} f\left(\frac{y}{\sqrt{t}}\right)dy$$

change of variables $u = y/\sqrt{t} \longrightarrow du = dy/\sqrt{t}$ so the integral becomes

$$t^{\alpha+1/2}\int_{-\infty}^{\infty} f\left(u\right)du = 1$$

which we need to normalise independent of time t. This is only possible if $\alpha = -1/2$.

So the D.E becomes

$$-\frac{1}{2}\left(f + \xi f'\right) = c^2 f''.$$

We have an exact derivative on the lhs, i.e. $\dfrac{d}{d\xi}\left(\xi f\right) = f + \xi f'$, hence

$$-\frac{1}{2}\frac{d}{d\xi}\left(\xi f\right) = c^2 f''$$

and we can integrate once to get

$$-\frac{1}{2}(\xi f) = c^2 f' + K.$$

We set $K = 0$ in order to get the correct solution, i.e.

$$-\frac{1}{2}(\xi f) = c^2 f'$$

which can be solved as a simple first order variable separable equation:

$$f(\xi) = A \exp\left(-\tfrac{1}{4c^2}\xi^2\right)$$

and returning to

$$p(y,t) = t^{-1/2} f(\xi)$$

becomes

$$p(y,t) = \frac{A}{\sqrt{t}} \exp\left(-\frac{y^2}{4tc^2}\right).$$

This is a pdf for a variable y that is normally distributed with mean zero and standard deviation $c\sqrt{2t}$, which we ascertained by the following comparison:

$$-\frac{1}{2}\frac{y^2}{2tc^2} : -\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}$$

i.e. $\mu \equiv 0$ and $\sigma^2 \equiv 2tc^2$.