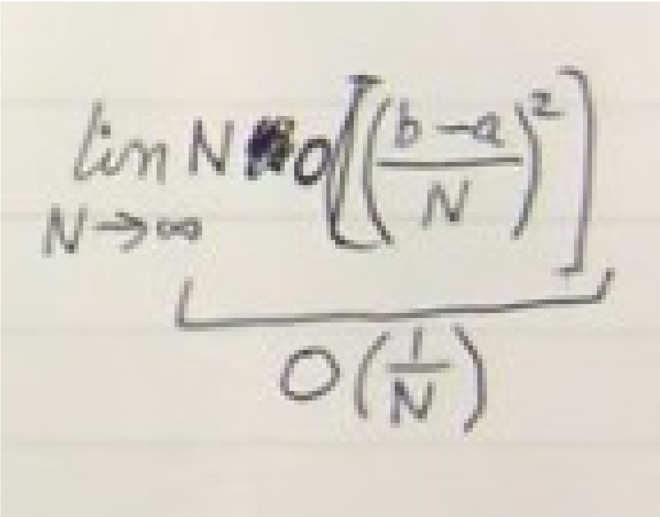
Lectures 1-2:

Basically just a review of A level further maths + Level 6 technical results

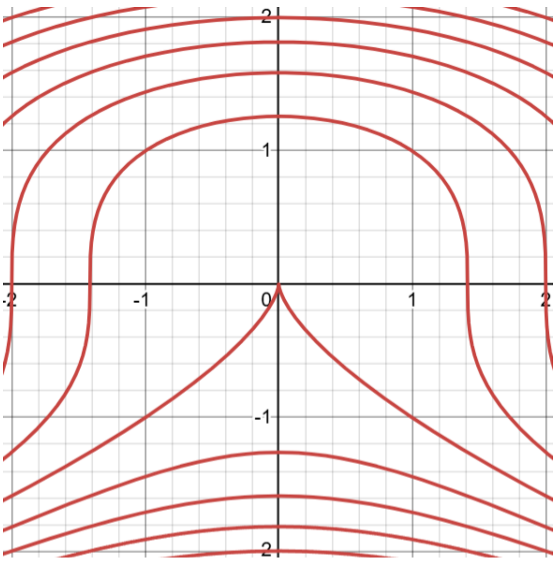
Lecture 3:

The entire lecture was just a review of stuff we've already met at A level, as well as the lecturer doing an unjustified limit that isn't allowed (although the thing he "proved" using this unjustified limit is something we have proved properly in the A level documents). Remember, limits and big O do not commute, even if the lecturer implied they do.



Lecture 4:

We can have functions of multiple variables, like $f: \mathbb{R}^m \to \mathbb{R}^n$. As an example, suppose $z = x^2 + y^3$, then we can sketch this using a contour plot, kind of like elevation maps, where we show lines on the x,y-plane corresponding to when $x^2 + y^3$ is constant. Here is that example:



However, if I imagine this as a 3d graph of a surface with height equal to $x^2 + y^3$, then if I pick a point on this graph and try to find the slope, I have a problem that the slope depends on the direction. Therefore I write $\frac{\partial z}{\partial x}$ for the slope as I move in just the x direction and hold y constant. This is called a partial derivative. In this example, that is 2x, because we differentiate $x^2 + y^3$ and the y^3 vanishes since it is a constant. We put a little thingy in the corner like this to show what's being held constant, as shown below.

$$\left.\frac{\partial z}{\partial x}\right|_y$$

We do need to be careful about showing what is constant in some cases, as for example if f is a function of x, y and z, then $\left.\frac{\partial f}{\partial x}\right|_y$ does not always equal $\left.\frac{\partial f}{\partial x}\right|_z$. For example, in the surface $x^2 + y^3 + z^4 = 1$, then $\left.\frac{\partial f}{\partial x}\right|_y = \frac{d}{dx}(x^2 + y^3 + z^4) = 2x + 4z^3\frac{dz}{dx}$ since y is constant, however $\left.\frac{\partial f}{\partial x}\right|_z = 2x + 3y^2\frac{dy}{dx}$.

Formally, for example, if z is a function of x and y, then

$$\left.\frac{\partial z}{\partial x}\right|_y$$

Is defined as $\lim_{h\to 0}\frac{z(x+h,y)-z(x,y)}{h}$.

Example:

$f(x,y) = x^2 + y^3 + e^{xy^2}$. As a shorthand for the partial derivative with respect to x we often instead write f_x. Since y is treated as constant here, we get that $f_x = 2x + y^2e^{xy^2}$, and $f_y = 3y^2 + 2xye^{xy^2}$. We can compute second partial derivatvies: $f_{xy} = 2ye^{xy^2} + 2xy^3e^{xy^2}$, $f_{xx} = 2 + y^4e^{xy^2}$, and also $f_{yx} = 2ye^{xy^2} + 2xy^3e^{xy^2}$. Notice that $f_{xy} = f_{yx}$ in this case. It turns out this is not a coincidence. We will now prove that this is intuitive, and always true whenever f_{xy} and f_{yx} are continuous, and we will also prove a version of the chain rule for multivariable functions. First, we will need to interpret partial derivatvies as entries of a matrix.

Some precise definitions we need:

Let f be a function from $\mathbb{R}^m$ to $\mathbb{R}^n$.

Then the directional derivative of f at a point a is the slope of f as you move along a vector, which, for a vector u, can be written as $D_uf(a) = \lim_{t\to 0}\frac{f(a+tu)-f(a)}{t}$ whenever this limit exists. This is equal to $\frac{d}{dt}f(a+tu)$ when t=0. When u is a basis vector, like the x axis or the y axis, this is when we get a partial derivative.

We will define the "derivative" of f. Here, h is an m-dimensional vector and f spits out n dimensional vectors. Like in the real number case, we say f is differentiable if there is some A such that

$$f(a+h) = f(a) + Ah + o(h)$$

This means that A is an n*m dimensional matrix as it is a linear map from m dimensions to n dimensions. Intuitively this means we can find an approximation of f near a that resembles a line or a plane or whatever whenever f is differentiable.

We now call A Df. From now on you can think of Df(x) as Df times x in the matrix multiplication sense, since matrix multiplication is really function composition. We have, for a vector u in a fixed direction, $\lim_{t\to 0}\frac{f(a+tu)-f(a)-Df(a)(tu)}{|tu|} = 0$. This is equivalent to saying $\lim_{t\to 0}\frac{f(a+tu)-f(a)-tDf(a)(u)}{|t||u|} = 0$, thus if u is a basis vector, this is equivalent to saying that $Df(a)(u)$ is the partial derivative of f with respect to u.

This is because In this case |u|=1 so we just need $\lim_{t\to 0}\frac{f(a+tu)-f(a)-tDf(a)(u)}{t}=0$ so $Df(a)u=\lim_{t\to 0}\frac{f(a+tu)-f(a)}{t}$ which is clearly the partial derivative. But when we multiply a matrix by a basis vector, we essentially are filtering for a specific column, so the resulting vector, which is the partial derivatives of the components of the output vector of f, is part of the matrix D. This means that D is a matrix pf partial derivatives, which is really nice.

Theorem 1: If all the partial derivatives of f are continuous in a neighbourhood around a then f is differentiable at a in the sense above.

Proof of lemma (screenshots from some other cambridge notes): For each n-dimensional vector h we have the following:

$$f(\mathbf{a}+\mathbf{h})-f(\mathbf{a})=\sum_{j=1}^{n} f(\mathbf{a}+h_1\mathbf{e}_1+\cdots+h_j\mathbf{e}_j)-f(\mathbf{a}+h_1\mathbf{e}_1+\cdots+h_{j-1}\mathbf{e}_{j-1})$$

And we write:

$$\mathbf{h}^{(j)}=h_1\mathbf{e}_1+\cdots+h_j\mathbf{e}_j=(h_1,\cdots,h_j,0,\cdots,0)$$ where the e's are basis vectors (like (1,0,0), etc).

Then we can use the method of differences:

$$f(a+h)-f(a)=\sum_{j=1}^{n}\left(f\left(a+h^{(j)}\right)-f\left(a+h^{(j-1)}\right)\right)$$

$$=\sum_{j=1}^{n}\left(f\left(a+h^{(j-1)}+h_je_j\right)-f\left(a+h^{(j-1)}\right)\right)$$

But now, the mean value theorem from single-variable valvulus allows us to write

$$f(a+h)-f(a)=\sum_{j=1}^{n}h_j\left(\frac{\delta}{\delta e_j}f\left(a+h^{(j-1)}+t_jh_je_j\right)\right)$$

For some t_j between 0 and 1 that depends on j.

$$f(a+h)-f(a)=\sum_{j=1}^{n}h_j\left(\frac{\delta}{\delta e_j}f\left(a+h^{(j-1)}+t_jh_je_j\right)\right)$$

$$=\sum_{j=1}^{n}h_j\frac{\delta}{\delta e_j}f(a)+\sum_{j=1}^{n}h_j\left(\left(\frac{\delta}{\delta e_j}f\left(a+h^{(j-1)}+t_jh_je_j\right)\right)-\left(\frac{\delta}{\delta e_j}f(a)\right)\right)$$

But the partial derivatives are continuous at a so the second term is thus o(h) as $\left(\frac{\delta}{\delta e_j}f\left(a+h^{(j-1)}+t_jh_je_j\right)\right)-\left(\frac{\delta}{\delta e_j}f(a)\right)$ approaches 0 by continuity so when multiplied by h it is o(h). Thus $f_i(a+h)-f_i(a)$, meaning that $Df(a)=\sum_{j=1}^{n}\frac{\delta}{\delta e_j}f(a)$, essentially the matrix of partial derivatives.

Theorem 2: $\frac{\delta^2 f}{\delta x\delta y}=\frac{\delta^2 f}{\delta y\delta x}$ if both of these partial derivatives are continuous at a.

We will prove this. However, this is somewhat obvious, in the sense that, for example, if I move north a bit and measure the change in height, then go back to where I started and move east a bit and to that again, I may measure a slightly different change in height. If I do what I described in the last sentence but with the words “north” and “east” swapped around, the difference of the height differences will be the same, it will always equal the sum of two of the diagonal corner heights minus the height of the other two diagonal corners. Stare at this until it makes sense to you so you have an intuition of what is really going on.

Proof (screenshots from cambridge notes): Let’s assume f is going to $\mathbb{R}^1$, since it is only necessary to show that this is true for each component of any arbitrary f.

Let

$$g_{ij}(t) = f(\mathbf{a} + t\mathbf{e}_i + t\mathbf{e}_j) - f(\mathbf{a} + t\mathbf{e}_i) - f(\mathbf{a} + t\mathbf{e}_j) + f(\mathbf{a}).$$

Then for each fixed t, define $\phi : [0, 1] \to \mathbb{R}$ by

$$\phi(s) = f(\mathbf{a} + st\mathbf{e}_i + t\mathbf{e}_j) - f(\mathbf{a} + st\mathbf{e}_i).$$

Then we get

$$g_{ij}(t) = \phi(1) - \phi(0).$$

We have $g_{ij}(t) = \phi'(s) = t\left(\left(\frac{\delta}{\delta e_i} f\left(a + ste_i + te_j\right)\right) - \left(\frac{\delta}{\delta e_i} f(a + ste_i)\right)\right)$

$= t\left(\frac{\delta}{\delta e_i}\left(f(a + ste_i + te_j) - f(a + ste_i)\right)\right)$ for some s between 0 and 1 because of the single variable mean value theorem. Applying the mean value theorem to $f(a + ste_i + kte_j)$ means there is a k between 0 and 1 such that $g_{ij}(t) = t^2 \frac{\delta^2}{\delta e_j \delta e_i} f(a + ste_i + kte_j)$ and we can do the same for g_{ji}, which is in fact equal to g_{ij} by definition, to get that $g_{ji}(t) = t^2 \frac{\delta^2}{\delta e_i \delta e_j} f(a + \tilde{s}te_i + \tilde{k}te_j)$, so since they are equal we have $t^2 \frac{\delta^2}{\delta e_i \delta e_j} f\left(a + \tilde{s}te_i + \tilde{k}te_j\right) = t^2 \frac{\delta^2}{\delta e_j \delta e_i} f(a + ste_i + kte_j)$. We are interested in the limit of this as t goes to 0, and by continuity, these both converge to $\frac{\delta^2 f}{\delta y \delta x}(a)$, which thus must equal $\frac{\delta^2 f}{\delta x \delta y}(a)$.

Theorem 3: This is the multivariate chain rule. Although we will provide a proof, it is far more important that I am providing an explanation of what the chain rule actually says. We will prove it and then do examples to show why some of the expressions that use the chain rule are the same as the chain rule which we will prove here in the matrix sense. The chain rule says

$$D(g \circ f)(a) = Dg\big(f(a)\big)Df(a)$$

From the definition of the derivatives, we have

$$f(a + h) = f(a) + Df(h) + o(h)$$

and

$$g(f(a) + k) = g(f(a)) + Dg(f(k)) + o(k)$$

Therefore

$g(f(a+h)) = g(f(a) + Df(h) + o(h))$. Let $k = Df(h) + o(h)$, then we have

$$g(f(a+h)) = g(f(a)) + Dg(f(Df(h) + o(h))) + o(Df(h) + o(h))$$

So

$$g\big(f(a+h)\big) = g\big(f(a)\big) + Dg\left(f\big(Df(h)\big)\right) + Dg(f\big(o(h)\big)) + o(Df(h) + o(h))$$

Since derivatives are additive.

Now we will define the operator norm. For a matrix B, let |B| be the largest possible magnitude of a vector of magnitude 1 after being multiplied by B. This is clearly finite, then multiplying everything by the right constants gives that for a vector B, $|Bv| \leq |B||v|$. Therefore $Dg(f\big(o(h)\big))$ is o(h) since D(g∘f) is a matrix. Also, by the triangle inequality, |Ah+o(h)| ≤ |A||h|+|o(h)| ≤ (|A|+1)|h| since an o(h) thing is less than 1 times h when h is sufficiently small by definition. Therefore, since Ah+o(h) is bounded by a constant times h, o(Ah+o(h)) is o(h) as well. The fact that we now have

$$\big(f(a+h)\big) = g\big(f(a)\big) + Dg\left(f\big(Df(h)\big)\right) + o(h)$$

Completes the proof of the chain rule.

Side note: Operator norm notation is annoying because we use absolute value signs to denote both it and the determinant so we have to guess what we want based on context. For both cases, we sometimes use double bars like ||A|| or single bars like |A|.

EXAMPLE:

Let's do an example where we apply the chain rule as given above. Let z be a function of x and y which are both functions of t. Let $z = xy$ where $x = e^t$ and $y = \sin(t)$ Then we can consider a function $f(t) = (x = e^t, y = \sin(t))$ and $g(x,y) = xy$, then $\frac{dz}{dt} = \frac{d\big(g(f(t))\big)}{dt}$. This is a total derivative and equal to the matrix derivative since z(t) is a function from 1 variable to 1 variable.

Since f is a function from 1 variable to 2 variables, the matrix Df in question will be 2x1. The matrix will then look as follows:

$$Df = \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix} = \begin{pmatrix} e^t \\ cos(t) \end{pmatrix}$$

Similarly, Dg will be 1x2 because g is a function from 2 variables to 1 variable.

$$Dg = \begin{pmatrix} \frac{\delta z}{\delta x} & \frac{\delta z}{\delta y} \end{pmatrix} = \begin{pmatrix} y & x \end{pmatrix}$$

$$D(g \circ f) = \begin{pmatrix} \sin(t) & e^t \end{pmatrix}$$

$$D(g \circ f)Df = \begin{pmatrix} \sin(t) & e^t \end{pmatrix} \begin{pmatrix} e^t \\ \cos(t) \end{pmatrix}$$

$$D(g \circ f)Df = \sin(t)\, e^t + \cos(t) e^t$$

And, therefore, since $\frac{dz}{dt} = D(g \circ f)Df = \begin{pmatrix} \frac{\delta z}{\delta x} & \frac{\delta z}{\delta y} \end{pmatrix} \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix}$, so $\frac{dz}{dt} = \frac{\delta z}{\delta x}\frac{dx}{dt} + \frac{\delta z}{\delta y}\frac{dy}{dt}$. This is why that comes from the chain rule.

Notice, however, that the statement $\frac{dz}{dt} = \frac{\delta z}{\delta x}\frac{dx}{dt} + \frac{\delta z}{\delta y}\frac{dy}{dt}$ is intuitively obvious: Since dt is not 0 and only approaching 0, we can write $dz = \frac{\delta z}{\delta x}dx + \frac{\delta z}{\delta y}dy$, which basically says that z changes by the amount z changes as x changes times the amount that x changes plus the amount that z changes as y changes times the amount that y changes. Stare at this until you see why it makes the above formula intuitively obvious. If z is a function of y which is a function of x, then $\frac{dz}{dx} = \frac{\delta z}{\delta x}\frac{dx}{dx} + \frac{\delta z}{\delta y}\frac{dy}{dx} = \frac{\delta z}{\delta x} + \frac{\delta z}{\delta y}\frac{dy}{dx}$.

Notice that the columns correspond to the variable we are differentiating with respect to, and the rows correspond to the variable we are differentiating. Note that the matrices are always compatible because if f is a function from a variables to b variables and g is from c variables to d variables, then f∘g is only valid if b=c, which exactly corresponds with the condition for matrix multiplication to be compatible. So done.

We can also take the chain rule for a function f(x,y) and integrate both sides as follows.

$$\int_{(x_1,y_1)}^{(x_2,y_2)} df = \int_{(x_1,y_1)}^{(x_2,y_2)} \frac{\delta f}{\delta x}dx + \int_{(x_1,y_1)}^{(x_2,y_2)} \frac{\delta f}{\delta y}dy$$

We are taking the path where we move in the x direction until x_2 and then move in the y direction until y_2. Therefore $\int_{(x_1,y_1)}^{(x_2,y_2)} \frac{\delta f}{\delta x}dx + \int_{(x_1,y_1)}^{(x_2,y_2)} \frac{\delta f}{\delta y}dy = \int_{x_1}^{x_2} \frac{\partial f}{\partial x}\Big|_{y=y_1} dx + \int_{y_1}^{y_2} \frac{\partial f}{\partial y}\Big|_{x=x_2} dy$.

$\int_{(x_1,y_1)}^{(x_2,y_2)} df$ just means adding up the small changes in f as you move along the path. Since the sum of these small changes will always be $f(x_2, y_2) - f(x_1, y_1)$, the integral does not depend on the path, and we just wrote it as $\int_{x_1}^{x_2} \frac{\partial f}{\partial x}\Big|_{y=y_1} dx + \int_{y_1}^{y_2} \frac{\partial f}{\partial y}\Big|_{x=x_2} dy$ because this reduces it to two integrals of single variable functions.

Lecture 5:

Here are some examples of using the chain rule.

Lets say we want to work in polar coordinates and write $f(x, y) = f\big(x(r, \theta), y(r, \theta)\big)$ where $x = rcos(\theta)$ and $y = rsin(\theta)$. Then we can apply the chain rule as follows:

$$\frac{\partial f}{\partial \theta}\Big|_r = \frac{\partial f}{\partial x}\Big|_y \frac{\partial x}{\partial \theta}\Big|_r + \frac{\partial f}{\partial y}\Big|_x \frac{\partial y}{\partial \theta}\Big|_r = \frac{\partial f}{\partial y}\Big|_x rcos(\theta) - \frac{\partial f}{\partial x}\Big|_y rsin(\theta)$$

$$\frac{\partial f}{\partial r}\Big|_\theta = \frac{\partial f}{\partial x}\Big|_y \frac{\partial x}{\partial r}\Big|_\theta + \frac{\partial f}{\partial y}\Big|_x \frac{\partial y}{\partial r}\Big|_\theta = \frac{\partial f}{\partial x}\Big|_y cos(\theta) + \frac{\partial f}{\partial y}\Big|_x si\,n(\theta)$$

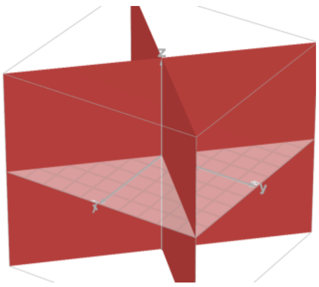
Now let's consider a surface defined by $f(x, y, z) = c$. The chain rule gives $\frac{\partial f}{\partial x}\Big|_z = \frac{\partial f}{\partial x}\frac{\partial x}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial x} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial x}$. However, z is constant so the last term vanishes, and dx/dx is 1, so $\frac{\partial f}{\partial x}\Big|_z = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial x}$. However, on paths where z and f are constant, we have $f_x + f_y \frac{\partial y}{\partial x} = 0$. Rearranging gives $\frac{\partial y}{\partial x}\Big|_z = -\frac{f_x|_{y,z}}{f_y|_{x,z}}$, and similarly for other partial derivatives. In fact, we can easily show from this that the following product, called the cyclical rule, holds:

$$\left.\frac{\partial y}{\partial x}\right|_z \left.\frac{\partial z}{\partial y}\right|_x \left.\frac{\partial x}{\partial z}\right|_y = -1.$$

In the normal two dimensional case, we had $\frac{dy}{dx}\frac{dx}{dy} = 1$. This is still true whenever we are holding all but two variables constant, however, it is in general wrong to assume that for a function f(x,y,z),

$\left.\frac{\partial f}{\partial x}\right|_y \left.\frac{\partial x}{\partial f}\right|_z = 1$, as this is false in general.

Also, here is the function $x^2 - y^2 = 0$.



As you can see, when x=y=0, $\frac{\delta y}{\delta x}$ will not exist as the slope in question could be -1 or 1. This will be reflected when we try to compute it: if $x^2 - y^2 = f = c$, then we use $\left.\frac{\partial y}{\partial x}\right|_z = -\frac{f_x|_{y,z}}{f_y|_{x,z}}$ to get that this is $\left.\frac{\partial y}{\partial x}\right|_z = -\frac{2x}{-2y}$, which is indeed undefined when y=0.

I will also show an example of how to find partial derivatives like $\frac{\delta y}{\delta x}$ on a surface defined parametrically with x, y, and z functions of u and v.

Here is the first case:

$$x = x(u,v), \quad y = y(u,v), \quad z = z(u,v)$$

$$dx = x_u\,du + x_v\,dv$$

$$dy = y_u\,du + y_v\,dv$$

$$dz = z_u\,du + z_v\,dv$$

$$\begin{pmatrix} x_u & x_v \\ y_u & y_v \end{pmatrix}\begin{pmatrix} du \\ dv \end{pmatrix} = \begin{pmatrix} dx \\ dy \end{pmatrix}$$

$$\frac{1}{x_u y_v - x_v y_u}\begin{pmatrix} y_v & -x_v \\ -y_u & x_u \end{pmatrix}\begin{pmatrix} dx \\ dy \end{pmatrix} = \begin{pmatrix} du \\ dv \end{pmatrix}$$

$$dz = z_u\left[\frac{y_v\,dx - x_v\,dy}{x_u y_v - x_v y_u}\right] + z_v\left[\frac{x_u\,dy - y_u\,dx}{x_u y_v - x_v y_u}\right]$$

$$dz = dx\left[\frac{y_v z_u - z_v y_u}{x_u y_v - x_v y_u}\right] + dy\left[\frac{z_v x_u - z_u x_v}{x_u y_v - x_v y_u}\right]$$

$$z_x = \frac{y_v z_u - z_v y_u}{x_u y_v - x_v y_u}$$

Where in the last step I have reverse engineered z_x by equating the coefficients in the chain rule, valid since the step above must be true even when I fix y.

If it is parametric in terms of just t, then we usually cannot do something similar: A curve's trajectory might leave the plane where y is constant so we cannot just fix y. The above proof assumes that the surface is actually differentiable at the point in question when we fix y. For example, the point (1, 0, 0) on the unit sphere is not differentiable with respect to x and with y held constant. This can be seen as the unit sphere can be parametrized as

$x = \sqrt{1 - v^2}\cos(u)\,, y = \sqrt{1 - v^2}\sin(u), z = v.$

Then the denominator of z_x by the above formula is

$-\sqrt{1 - v^2}\sin(u)\frac{-v}{\sqrt{1-v^2}}\sin(u) - \frac{-v}{\sqrt{1-v^2}}\cos(u)\sqrt{1 - v^2}\cos(u) = v(\sin^2(u) + \cos^2(u)) = v$, but since this is at (1, 0, 0), v must be 0 since z=v, so the denominator works out to be 0.

Similarly, if we were to define a curve parametrically in terms of a single variable, then we could achieve that by doing it in terms of u and v but never putting v in any of the equations, but then the derivative of x, y, z with respect to v would be 0, so the denominator would also vanish, consistent with the intuition before that it is not possible.

There is one more topic, and that is differentiating under the integral sign.

If $I(c) = \int_a^b f(c, x)dx$ then $I'(c) = \int_a^b f_c(c, x)dx$. Intuitively this makes sense: We can swap the differentiation and integration order because the sum of the changes is the change of the sums. However, we need to justify this. It is true whenever $f(c, x)$ and $f_c(c, x)$ are continuous everywhere in a closed rectangle with x going from a to b and c in some neighbour its value. If the integral is improper we can just take a limit: If the integral is absolutely convergent then the dominated convergence theorem allows us to take this limit when it exists. We proved DCT in the level 6 technical results document, and now we will show that this differentiation and integration swap is valid.

The following screenshots are from wikipedia.

Now, both of these are true from the fundamental theorem of calculus and the fact that swapping integration bounds changes the sign.

$$\frac{\partial}{\partial b}\left(\int_a^b f(x)\,dx\right) = f(b), \qquad \frac{\partial}{\partial a}\left(\int_a^b f(x)\,dx\right) = -f(a).$$

Let’s define

$$\varphi(\alpha) = \int_a^b f(x, \alpha)\,dx$$

Now, f is continuous on a closed rectangle, and therefore by a theorem in the level 6 technical results document it is uniformly continuous in that rectangle. Thus, there exists a Δα such that

$$|f(x, \alpha + \Delta\alpha) - f(x, \alpha)| < \varepsilon$$

always, for any arbitrary ε.

Also,

$$
\begin{aligned}
\Delta\varphi &= \varphi(\alpha+\Delta\alpha)-\varphi(\alpha) \\
&= \int_a^b f(x,\alpha+\Delta\alpha)\,dx - \int_a^b f(x,\alpha)\,dx \\
&= \int_a^b \left(f(x,\alpha+\Delta\alpha)-f(x,\alpha)\right)\,dx \\
&\leq \varepsilon(b-a).
\end{aligned}
$$

Which implies that ϕ is continuous: The output can be made arbitrary close by making the inputs sufficiently close. Also, by continuity of f_a, there is a Δα with

$$
\left|\frac{f(x,\alpha+\Delta\alpha)-f(x,\alpha)}{\Delta\alpha}-\frac{\partial f}{\partial\alpha}\right|<\varepsilon
$$

As by the mean value theorem the first term of the above screenshot is $f_a(x, a+d)$ with d<Δα, and we pick Δα small enough such that the difference is within ε for any d<Δα. Now, this implies that

$$
\frac{\Delta\varphi}{\Delta\alpha}=\int_a^b \frac{f(x,\alpha+\Delta\alpha)-f(x,\alpha)}{\Delta\alpha}\,dx=\int_a^b \frac{\partial f(x,\alpha)}{\partial\alpha}\,dx+R,
$$

where

$$
|R|<\int_a^b \varepsilon\,dx=\varepsilon(b-a).
$$

The reason for the last term is from the ε bound in the screenshot above. As ε gets smaller, Δα approaches 0, so we have that

$$
\lim_{\Delta\alpha\to 0}\frac{\Delta\varphi}{\Delta\alpha}=\frac{d\varphi}{d\alpha}=\int_a^b \frac{\partial}{\partial\alpha}f(x,\alpha)\,dx
$$

Since R gets small so the two integrals in the screenshot two above that differ by R must approach eachother.

Another theorem says what happens if the integration bounds depend on c:

If $I(c)=\int_{a(c)}^{b(c)} f(x,c)dx$ then $I'(c)=\int_{a(c)}^{b(c)} f_c(x,c)dx+f(b,c)\frac{db}{dc}-f(a,c)\frac{da}{dc}$. Now we need the same continuity conditions as before, and also suppose we have the same continuity conditions in an interval around [a,b] since a and b are changing.

Proof:

Let

$$
\varphi(\alpha)=\int_a^b f(x,\alpha)\,dx,
$$

with a and b depending on α. Then

$$\begin{aligned}\Delta\varphi &= \varphi(\alpha+\Delta\alpha)-\varphi(\alpha)\\ &= \int_{a+\Delta a}^{b+\Delta b} f(x,\alpha+\Delta\alpha)\,dx - \int_a^b f(x,\alpha)\,dx\\ &= \int_{a+\Delta a}^{a} f(x,\alpha+\Delta\alpha)\,dx + \int_a^b f(x,\alpha+\Delta\alpha)\,dx + \int_b^{b+\Delta b} f(x,\alpha+\Delta\alpha)\,dx - \int_a^b f(x,\alpha)\,dx\\ &= -\int_a^{a+\Delta a} f(x,\alpha+\Delta\alpha)\,dx + \int_a^b [f(x,\alpha+\Delta\alpha)-f(x,\alpha)]\,dx + \int_b^{b+\Delta b} f(x,\alpha+\Delta\alpha)\,dx.\end{aligned}$$

In the first and last of the integrals above, we can apply the mean value theorem which essentially says that $\int_a^b f(x)dx = (b-a)f(\xi)$ with $a < \xi < b$ (This is just the standard MVT applied to an antiderivative of f). This gives

$$\Delta\varphi = -\Delta a f(\xi_1,\alpha+\Delta\alpha) + \int_a^b [f(x,\alpha+\Delta\alpha)-f(x,\alpha)]\,dx + \Delta b f(\xi_2,\alpha+\Delta\alpha).$$

Now, for the same mean value theorem argument as In the above proof, dividing everything by Δα and taking a limit does give this for the middle integral:

$$\int_a^b \frac{\partial}{\partial\alpha} f(x,\alpha)\,dx$$

Since $\xi_1 \to a$, and f is continuous, the first term approaches $-\frac{\Delta \mathrm{a}}{\Delta\alpha} f(a,\alpha+\Delta\alpha)$, which approaches $-\frac{\mathrm{da}}{\mathrm{d}\alpha} f(a,\alpha)$, by continuity and the definition of the derivative. Similarly for the last term. So done.

Example:

Suppose we want to evaluate $\int_0^\infty x^n e^{-x} dx$. This can be done with integration by parts and induction, but here is a different method. By a simple substitution,

$\int_0^\infty e^{-\lambda x} dx = \frac{1}{\lambda}$ for $\lambda > 0$. This is continuous with λ in the vicinity of 1 (which is what we will eventually care about), and continuous in x everywhere, and we can just take a limit to infinity.

Differentiating with respect to λ n times gives $\int_0^\infty (-x)^n e^{-\lambda x} dx = (-1)^n \int_0^\infty x^n e^{-\lambda x} dx$. On the other hand, by the power rule, differentiating $\frac{1}{\lambda}$ n times gives $\frac{(n!)(-1)^n}{\lambda^{n+1}}$. Therefore, $(-1)^n \int_0^\infty x^n e^{-\lambda x} dx = \frac{(n!)(-1)^n}{\lambda^{n+1}}$. Taking $\lambda = 1$ and cancelling $n!$ Gives that $\int_0^\infty x^n e^{-x} dx = n!$

Lecture 6:

This lecture was mostly a review from A level of techniques for solving differential equations. However, we do have some definitions:

The **order** of a differential equation is the highest order derivative that appears, so you may hear about "first order" or "second order" differential equations for example.

An **ordinary** differential equation is a differential equation with only an independent (usually x or t) and a dependent variable (usually y).

A **linear** differential equation is a differential equation where all of the terms are y or one of its derivatives multiplied by a function of x (assuming y is the dependent variable and x is the independent variable). It has **constant coefficients** if all these "functions of x" are constant values, and these can be solved using the auxillary equation, as explained in levels 5 and 6.

A **homogenous** differential equation is a differential equation with no terms depending only on x.

The lecture also reviews some numerical methods that we met in further maths (level 5) for approximately solving differential equations.

We also have some obvious power series facts in which the technical details about convergence were justified in previous levels. Suppose $y = \sum_{n=0}^{\infty} a_n x^n$, then inside the radius of convergence,

- $\frac{dy}{dx} = \sum_{n=1}^{\infty} n a_n x^{n-1}$
- $x\frac{dy}{dx} = \sum_{n=1}^{\infty} n a_n x^{n}$
- $xy = \sum_{n=0}^{\infty} a_n x^{n+1} = \sum_{m=1}^{\infty} a_{m-1} x^{m}$

Before we solve differential equations, we need to be careful: We should only solve them on a domain in which they are defined: For example $\frac{dy}{dx} = \frac{y}{x}$ should only be solved on an interval not including x=0 so that everything is defined, otherwise something like y=|x| satisfies the equation everywhere that the stuff is defined. If we add implied constraints like that the solutions have to be infinitely differentiable everywhere then we get better behavior.

Lecture 7:

In simple cases we can find solutions using a series in terms of a_0. For example, if 5y'-3y=0 then 5xy'-3xy=0 so

$$5\sum_{n=1}^{\infty} n a_n x^n - \sum_{m=1}^{\infty} 3a_{m-1} x^m = 0$$

So we can equate coefficients to find $a_1, a_2, a_3, \ldots$ in terms of a_0.

In fact, we can write a recurrence relation: $5na_n = 3a_{n-1}$

$a_n = \frac{3}{5n} a_{n-1}$ for all n so we can apply this repeatedly:

$a_n = \left(\frac{3}{5n}\right)\left(\frac{3}{5(n-1)}\right) a_{n-2} = \left(\frac{3}{5n}\right)\left(\frac{3}{5(n-1)}\right)\left(\frac{3}{5(n-2)}\right) a_{n-3}$ and we can keep going until we get

$$a_n = \frac{\left(\frac{3}{5}\right)^n a_0}{n!}$$

Which means the power series agrees with the taylor series for $a_0 e^{\frac{3x}{5}}$, which we could also derive using the integrating factor.

After this we do more review from A level, such as this cool real world example:

We have 3 radioactive isotopes: A decays to B which decays to C.

At time t, the amount of each isotope is a(t), b(t), c(t).

We have that

$$\frac{da}{dt} = -k_a a$$

Because a decays proportionally to how much A is left.

Therefore $a = a_0 e^{-k_a t}$ which we know from A level.

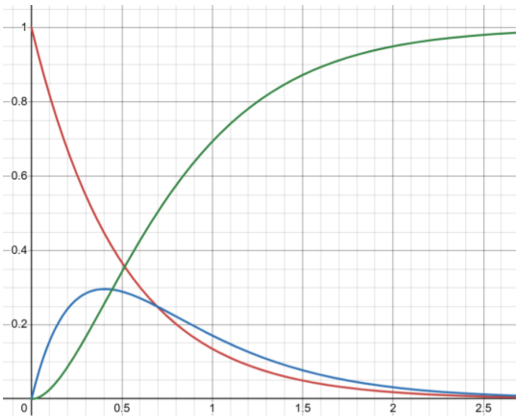
For b, we have to take into account the fact that it is increasing from a's decay and decreasing from its own decay.

$$\frac{db}{dt} = k_a a - k_b b = k_a a_0 e^{-k_a t} - k_b b$$

This is an equation which we can easily solve using an integrating factor, even though the lecturer solves it by guessing for reasons I don't understand.

We can eventually get that if $k_a \neq k_b$ then the unique solution for b satisfying b=0 when t=0 is $-\frac{k_a a_0}{k_b - k_a} e^{-k_b t} + \frac{k_a a_0}{k_b - k_a} e^{-k_a t}$. We can find the solution for c by finding what we would need to add to a and b to get $a_0 = a + b + c$.

Given that this problem came from radioactive decay, it makes sense what the graph of this looks like: a decays exponentially, b starts by increasing but itself decays so it peaks at some point then decays, and c is what is left so it increases over time.



Red = A, Blue = B, Green = C

We can use this idea to find how old something is by dating.

Sometimes boundary/initial conditions may not be about y being a constant at some constant x. For example, for the DE $xy' + (1 - x)y = 1$, the general solution is $y = -\frac{1}{x} + \frac{c}{x} e^x$, and the boundary consition that merely states that y is finite for all x determines that c=1. This is finite by a limit we had to solve by elementary means in the level 4 existence of e proof.

Going back to the radioactivity example in the case $k_a = k_b$, solving it using an integrating factor will eventually give that $b = k_a a_0 t e^{-k_b t}$ is the solution that satisfies that b=0 when t=0.

Lecture 8:

Sometimes equations are separable (meaning you can apply separation of variables from A level, ie it can be written as $\frac{dy}{dx} = \frac{f(x)}{g(y)}$) even if they don't look that way. Example:

$$(x^2y - 3y)\frac{dy}{dx} - 2xy^2 = 4x$$

$$y(x^2 - 3)\frac{dy}{dx} = 2x(2 + y^2)$$

$$\frac{y}{2 + y^2}dy = \frac{2x}{x^2 - 3}dx$$

So we can integrate both sides which gives us solutions in terms of logs, we eventually get

$$y^2 + 2 = C((x^2 - 3)^2)$$

Definition:

An ODE of the form $Q(x,y)\frac{dy}{dx} + P(x,y) = 0$ is exact if there exists an $f(x,y)$ with $\frac{\partial f}{\partial x} = P(x,y) + Q(x,y)\frac{dy}{dx}$ and $\frac{\partial f}{\partial y} = P(x,y)\frac{dx}{dy} + Q(x,y)$. Both are equivalent to $\int df = \int Pdx + \int Qdy$ by previous definitions. If an ODE is exact then the equation is saying df=0 so f=constant is the general solution. We don't know yet how to determine if an equation is exact but we will talk about how to do that and how to apply this.

If an equation is exact, we can use the multivariate chain rule and equate coefficients to get:

$f_x = P(x,y)$ and $f_y = Q(x,y)$.

If P_y and Q_x are continuous, then they are equal to eachother, since $f_{xy} = f_{yx}$ if they are continuous. This is a necessary but not sufficient condition for exactness.

Definition: A domain D is simply connected if it has no holes, ie it is path connected and any closed curve in D can be continuously shrunk to a point in D without leaving D.

Theorem (Poincare lemma): If $P_y = Q_x$ are continuous in a simply connected open domain, then $Pdx + Qdy$ is an exact differential of a single valued function $f(x,y)$. I will give an idea of why it is true then give a proof.

Idea of why the Poincare lemma is true:

We can try to construct a function f by starting somewhere and filling it in based on the tiny changes in x and y. If the domain has holes, it could be that f goes up a spiral staircase and is thus multi-valued (kind of like the complex logarithm looping around the point 0 where it is not defined). But, if there are no holes, then the integral of df around a loop is 0 since it can be continuously deformed into a point where the integral is clearly 0 by simply connectedness. Because the integral of df is 0 f is single-valued at every point. However, we need to demonstrate that the integral does not change as we deform the path – this is not complete and just an intuition.

Proof of Poincare lemma:

Consider a rectangle inside a simply connected open domain. Define (going anticlockwise) its bottom as (a, c) to (b, c), its right as (b, c) to (b, d), its top as (b, d) to (a, d) and its left as (a, d) to (a, c). Let's try to integrate $Pdx + Qdy$ along those sides of the rectangle.

What we end up with is

$$\int_a^b P(x,c)\,dx + \int_c^d Q(b,y)\,dy - \int_a^b P(x,d)\,dx - \int_c^d Q(a,y)\,dy$$

This can be written as

$$\int_a^b [P(x,c) - P(x,d)]\,dx + \int_c^d [Q(b,y) - Q(a,y)]\,dy$$

By the fundamental theorem of calculus, we have the following

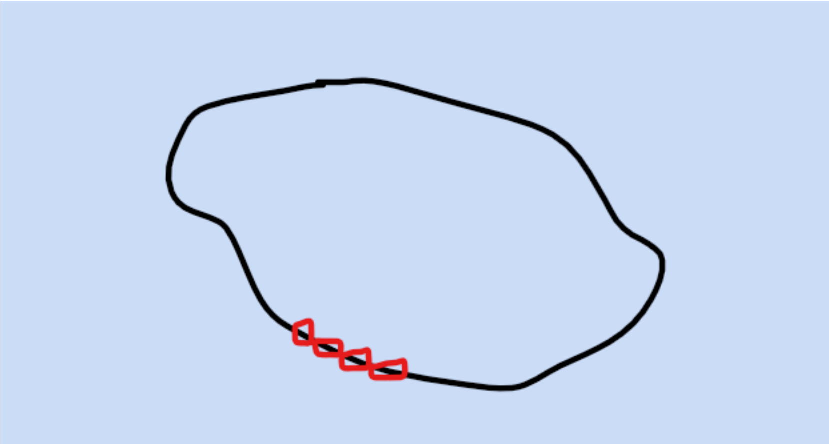
$$P(x,c) - P(x,d) = -\int_c^d P_y(x,y)\,dy, \qquad Q(b,y) - Q(a,y) = \int_a^b Q_x(x,y)\,dx.$$

And therefore we have the following integral, which collapses to zero because $P_y = Q_x$

$$\int_a^b \int_c^d [Q_x(x,y) - P_y(x,y)]\,dy\,dx$$

Now suppose we have a region that has a closed loop as the boundary and can be tiled by a bunch of these rectangles, so a possibly more complicated region but with all right angles and edges aligned to the axes. Then what happens is if we integrate along the boundary of this region, we get 0. The proof is that we can sum the integrals of the rectangles, which are all 0 from earlier, and what happens is all non-boundary edges in the integral vanish since an edge on the right of one rectangle is either part of the boundary or is connected to a left edge of another rectangle, and the integral along those opposite edges cancel since they are going in the opposite direction.

Now we will pick points on our path such that we can take those points to be opposite corners of a bunch of rectangles, and such that these rectangles are all completely inside our SCD which we will call D. This is fine since our path cannot touch the boundary as D is open by assumption. Fix (x_0, y_0) in our domain D. Then it is inside one of our rectangles, and so is any (x, y) in D. Define F to be the integral of $P(x,y)dx + Q(x,y)dy$ from (x_0, y_0) to (x, y) by first moving in a straight line from the starting point to the boundary of its rectangle, along more of the rectangles, until the rectangle of the end point, in a straight line to the end point. Suppose without loss of generality at each rectangle we go horizontally and then vertically. This is well defined since if we do it differently, the difference is 0 by the rectangular-domain property from earlier. We will show that F makes an exact differential as required.



This image shows what I am doing with rectangles: Since the black loop is not on the boundary since the domain is open, I just have to make them sufficiently small.

Lets compute F_x: By continuity of Q_x, we may apply differentiation under the integral sign, which by the way is called Feynman's trick. To find F_x, we will call (x_1, y_1) the first point we go to in our final rectangle, so that $F = \int_{(x_0,y_0)}^{(x_1,y_1)} P(x,y)dx + Q(x,y)dy + \int_{x_1}^{x} P(s,y_0)ds + \int_{y_1}^{y} Q(x,t)dt$ where the first term is on any axis aligned staircase like path.

$F_x = \int_{x_1}^{x} P_x(s,y_0)ds + \int_{y_1}^{y} Q_x(x,t)dt = P(x,y_0) + \int_{y_1}^{y} Q_x(x,t)dt$ by feynman's trick and the fundamental theorem of calculus. Also by FTC,

$P(x,y) - P(x,y_0) = \int_{y_0}^{y} P_y(x,t)dt = \int_{y_0}^{y} Q_x(x,t)dt$ since $P_y = Q_x$. Therefore,

$$F_x = P(x,y_0) + \int_{y_1}^{y} Q_x(x,t)dt + P(x,y) - P(x,y_0) = P(x,y)$$

And we can prove similarly that $F_y = Q(x,y)$.

Thus f is a function whose differential is the exact differential Pdx+Qdy as required. We used simply-connected-ness because when we prove that f is well defined, we use the "rectangle lemma", which has to hold even if the axis aligned path which integrates to 0 encloses the loop, as we asserted that this path integrates to 0, hopefully this makes sense.

Example:

$$6y(y-x)\frac{dy}{dx} + 2x - 3y^2 = 0$$

$$6y(y-x)dy + (2x - 3y^2)dx = 0$$

Now this is exact by the Poincare lemma. This is because $P_y = Q_x = -6y$ and this is defined everywhere and thus on any SCD.

Lets try to solve for P: $f_x = 2x - 3y^2$ so $f = x^2 - 3xy^2 + h(y)$

The point is that h vanishes when we differentiate wrt x so it can be any function of y, kind of like the generalized constant of integration.

For Q: $f_y = 6y^2 - 6xy$ so $f = 2y^3 - 3xy^2 + h(x)$

Thus if we set $h(x) = x^2$ and $h(y) = 2y^3$ we have an exact differential. This is no longer dependent on the Poincare lemma. So the solution is given by

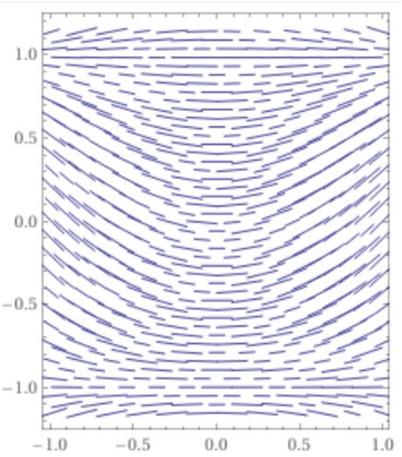
$$2y^3 - 3xy^2 + x^2 = C$$

Sometimes solutions to DEs are impossible to write in closed form, but we can still analyse the behavior of solutions with graphical methods. Each initial condition generates a distinct solution.

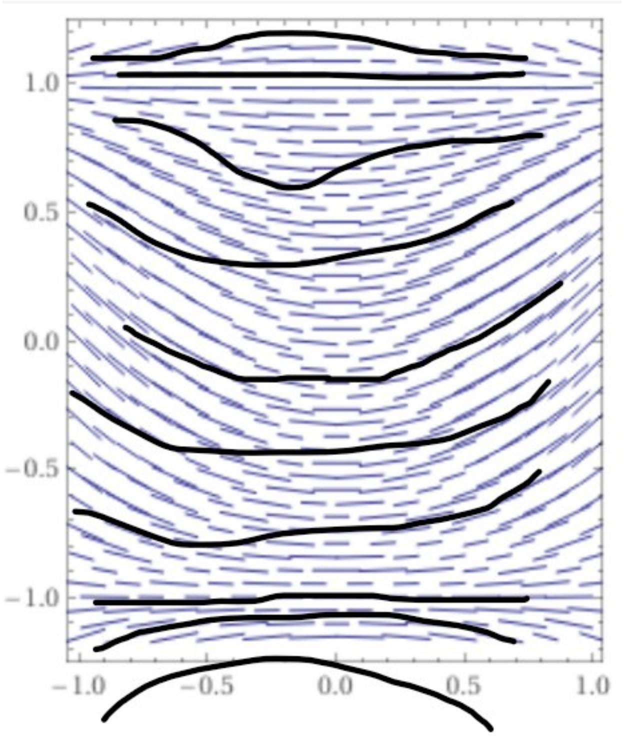
If I have $\frac{dy}{dx} = f(x,y)$ what I can do is draw a derivative vector field and try to sketch the solution curves without solving the equation by following the vector field. Here is what I mean:

For example: $\frac{dy}{dx} = x(1-y^2)$ can be solved, but I will do this example to show the method.

Here is the slope field from wolfram alpha, this is the vector field I mean: We find dy/dx at each point and sketch something like this.



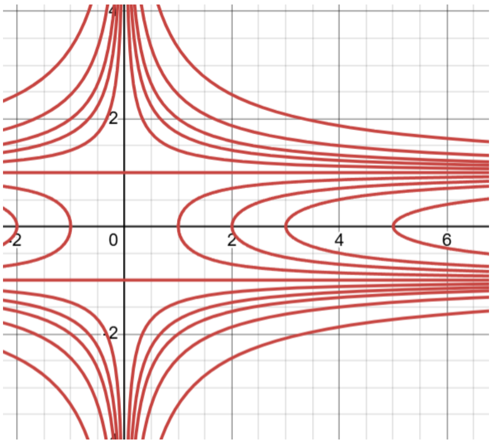
Now here is my attempt to follow the lines to get solution curves. An alternative way is to draw the contours $\frac{dy}{dx} = c$ and try to make a line to connect it appropriately. These contours are called isoclines. In particular, look for when $\frac{dy}{dx} = 0$ and try to solve for y to see if you can identify any constant solutions. In this particular equation we can find this way that $y = \pm 1$ are solutions.



As you can see, the solution y=1 is stable because the solutions around it are "attracted" to it. The solution y=-1 is unstable because the solutions around it are "repelled" from it. Since the derivative is positive if and only if -1<y<1, this implies that all solutions will approach either 1 or negative infinity as x increases, which is why we see this behavior. It's kind of like how if you put a pendulum vertically it may stay stable but if you push it even a little bit it will fall.

Lecture 9:

We will do an example of how we can sketch isoclines for the equation above. We want D (The derivative) to be a constant, and $D = x(1 - y^2)$ so it is constant when $y^2 = 1 - \frac{D}{x}$. Here are what the isoclines look like: We could sketch the graph by trying to fit a curve to be at the right slope at the relevant isoclines.



Definition: A fixed point is a solution y=c. It is stable if when y deviates from c by a sufficiently small amount it converges back to c as x increases. An unstable fixed point is a fixed point that is not stable.

We will now show some methods of determining whether a fixed point is stable. This is called Perturbation analysis. Suppose y=c is a fixed point of $\frac{dy}{dx} = f(x,y)$, then set y=c+ε. Then $\frac{d\varepsilon}{dx} = f(x, c+\varepsilon)$. If f is differentiable at constant x, we can write $\frac{d\varepsilon}{dx} = f(x,c) + \varepsilon f_y(x,c) + o(\varepsilon)$. The first term is zero because of the setup. Therefore if $f_y(x,c)$ is positive, the solution will be unstable since ε will grow, and otherwise the solution is stable since ε will shrink. If $f_y(x,c)$ is zero we need to add more terms to determine stability. The series will look like $f(x,c) + \varepsilon^2 f_{yy}(x,c)$ so we would need to consider the sign of $\varepsilon f_{yy}(x,c)$ to see if epsilon gets bigger or smaller.

Lets do this on the example $\frac{dy}{dx} = x(1-y^2)$. Suppose $c = 1$, then $\frac{d\varepsilon}{dx} = f(x,1) + \varepsilon f_y(x,1) + o(\varepsilon)$

$= \varepsilon(-2\mathrm{xy}) + o(\varepsilon)$. Therefore as x grows, ε grows negatively proportional to itself so it goes to 0. At c=-1 we get that $\frac{d\varepsilon}{dx} = \varepsilon(-2\mathrm{xy}) + \mathrm{o}(\varepsilon)$ which as x grows is now proportional to itself so ε will get larger so the solution is unstable.

Definition: An **autonomous** differential equation is one which does not depend on x (or the independent variable). For first order ones, we can give a formula for the solution but not always a closed form:

If $\frac{dy}{dx} = f(y)$ then $\int dx = \int \frac{1}{f(y)} dy$, so $x + c = \int \frac{1}{f(y)} dy$. However, this is hard to solve in closed form in general.

Example: Consider $\frac{dy}{dx} = y^2$ so $f = y^2$. This has a solution when $y = 0$. Lets see if this is stable or unstable. Lets set $y = \varepsilon$, then $\frac{d\varepsilon}{dx} = \varepsilon\big(\mathrm{f}_y(x,0)\big) + \varepsilon^2\big(f_{yy}(x,0)\big) + o(\varepsilon^2)$. Since f does not depend on x, we can write $\frac{d\varepsilon}{dx} = \varepsilon(0) + \varepsilon^2(2) + o(\varepsilon^2) = \varepsilon\big(2\varepsilon + o(\varepsilon)\big)$. This means the solution is stable if $2\varepsilon + o(\varepsilon)$ is negative, which means it is stable if we perturb y down but unstable if we perturb y up.

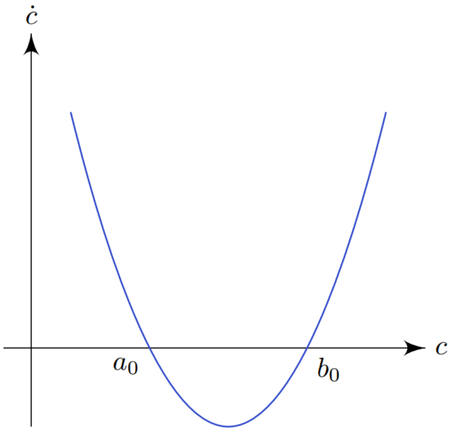
Example (in Chemistry):

Suppose we have a reaction A + B -> C + D where at the start we have a_0, b_0 of A and B and 0 of C and D. Suppose a(t), b(t), c(t) and d(t) are the amount of each chemical we have at any given time. Assume the system is modelled by:

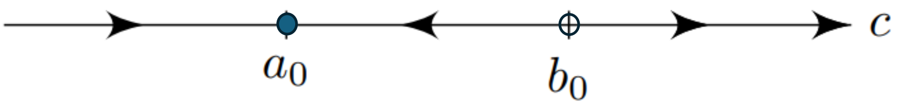
$a(t) + c(t) = a_0$ and $b(t) + c(t) = b_0$ and $c(t) = d(t)$ and $\frac{dc(t)}{dt} = \lambda a(t)b(t)$. These are realistic for some real world reasons but that is irrelevant and I don't really understand it anyway. Rearranging we get $\frac{dc(t)}{dt} = \lambda\big(a_0 - c(t)\big)\big(b_0 - c(t)\big)$. This is an autonomous system and next lecture we will analyze its stability. These ideas could therefore be useful in the context of chemistry.

Lecture 10:

Let's assume that $a_0 < b_0$ without loss of generality. $c = b_0$ and $c = a_0$ are constant solutions. We need to calculate $\frac{df}{dc}$, which is $\lambda(2c - a_0 - b_0)$. At $c = a_0$ this equals $\lambda(a_0 - b_0)$, and at $c = b_0$ this equals $\lambda(b_0 - a_0)$. Recall that the stability depends on the sign of $\frac{df}{dc}$, which means that $c = a_0$ is a stable fixed point and $c = b_0$ is unstable. However, $c = b_0$ is not possible for physical reasons (there would be negative amount of a). For autonomous systems like this, we can quickly see by plotting f against c and looking at its slope at the roots whether each solution is stable. (Screenshots from other cambridge notes)



We can also draw a 1D phase portrait to plot the trajectory of c with time. We can use open and closed circles to represent stable and unstable fixed points.



Example (logistic equation, this is really cool):

Suppose we have a population of size $y(t)$ and a birth rate equal to ay and a death rate equal to $by + cy^2$. Set $a - b = x$, so now we have $\frac{dy}{dt} = xy(1 - \frac{y}{Y})$ where Y is a constant. This is separable and not that hard to solve, but we will not do that – instead we will analyze its behavior.

The fixed points are y=0 and y=Y. The graph of y' against y will be an upside down parabola through 0 and y – Thus the fixed points are 0 and Y. This can be interpreted – If the population is 0 and you add a slight amount to the population it will continue to grow, and it will approach Y.

Now we will do this as a discrete equation (this is another term for a recurrence relation). This is similar to what we did back in that level 4 video about cobweb diagrams.

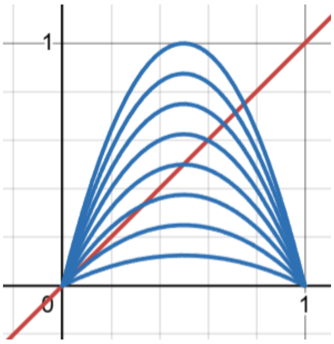
A fixed point of a first order discrete equation (ie an equation of the form $x_{n+1} = f(x_n)$) is defined as a value of x_n with $f(x_n) = x_n$. This means we will stay at x_n forever if we reach it. Recall from level 4 that the point is stable if the derivative of f is strictly between -1 and 1 in the vicinity of the point, and unstable if the derivative of f is strictly greater than 1 or less than -1 in the vicinity of the point.

Note that if the derivative is between 0 and 1 we will move monotonically towards the fixed point, and if it is between 0 and -1 we will oscillate around the fixed point with decreasing magnitude.

Now we will analyze the equation $x_{n+1} = rx_n(1 - x_n)$: This is the discrete logistic equation and it is related to the differential version, but this is where things get really interesting. In fact, my example from level 4 of chaotic behavior was one of these equations with r carefully chosen. We will analyze this in the case that x_n is positive.

Now our $f(x_n) = rx_n(1 - x_n)$. We will sketch this. Since the maximum of $rx_n(1 - x_n)$ turns out to be r/4 by simple calculus, we will only consider r between 0 and 4 so that x stays positive.

Here is x with the family of parabolas:



Solving for fixed ponits gives $x_n = 0, 1 - \frac{1}{r}$. We see that if r<1, we will not have any positive fixed points, and we will have a stable fixed point at 0 by the derivative criterion, so if this is a population then everyone would die.

Again by simple calculus, the derivative of f is 2-r at the non-zero fixed point, and thus when 1<r<3 the fixed point is stable. The fascinating behavior happens when 3<r<4. The recurrence relation behaves chaotically and unpredictably in this case.

Lecture 11:

The lecture is mostly a review of 2nd order ODE's with constant coefficients. We know how to solve these from A level. The lecturer spends the entire lecture giving a much harder derivation than the one in level 6, and therefore this section of the notes will be very short.

Definition: A **differential operator** is something like, for example $(aD^2 + bD + c)$ defined by $(aD^2 + bD + c)y = ay'' + by' + c$. It is linear if it is like a polynomial in D as in the previous example, or equivalently for an operator Ɗ, Ɗ(ax+by)=aƊx+bƊy – It is easy to check that polynomial operators satisfy this relation. What I did implicitly in level 6 was factor this polynomial to derive the solution to 2nd order ODE's with constant coefficients, but I did this without bringing up all this operator jargon.

Definition: Solutions to a DE are linearly independent if they are not linearly dependent, linearly dependent means one of the solutions can be written as a sum of constant multiples of the others.

Lecture 12:

We note that the solution to $ay'' + by' + c$ in the repeated roots case seems different from the solution in the other cases but it can be thought of as the limit of nearby cases.

For example, if the equation is $y'' - 4y' + (4 - \varepsilon^2)y = 0$ then the roots of the characteristic equation are $2 \pm \varepsilon$. We will show another method and then go back to this limiting case.

Theorem: A first order linear homogenous ODE with leading coefficient 1 and the other coefficient continuous has 1 linearly independent solution

Proof: Integrating factor

Suppose we have an equation of the form $y'' + p(x)y' + q(x)y = 0$ and we have a non-zero solution $y_1(x)$. Then we can try a substitution $y(x) = V(x)y_1(x)$. By the product rule, $y' = V'y_1 + Vy_1'$, and $y'' = V''y_1 + 2V'y_1' + Vy_1''$. Therefore, our equation becomes $V''y_1 + 2V'y_1' + Vy_1'' + p(V'y_1 + Vy_1') + q(Vy_1) = 0$. But our assumption was $y_1'' + p(x)y_1' + q(x)y_1 = 0$ since y_1 was a solution. Therefore we can get rid of that and just solve $V''y_1 + 2V'y_1' + p(V'y_1) = 0$. If we let $U = V'$, we have $U'y_1 + U(2y_1' + Py_1) = 0$. The idea is this equation can now be solved for U. We have $\frac{U'}{U} = -\frac{2y_1'}{y_1} - p$ so by integrating both sides, $\ln(U) = -2\ln(y_1) - \int p(x)dx$. Finally, we have that $U = \frac{A}{y_1^2}e^{-\int p(x)dx}$ where A is some constant. We could now integrate U to find V and then get the general solution to the original equation from finding a single solution. This method is called reduction of order.

Note that I almost want to say that from this we prove directly that such an equation has exactly two lineraly independent solutions. The above method gives a proof that this fact holds any time there is a solution that is never zero that we can use as our y_1. I will now prove the full statement, but first we need to introduce a new analysis concept.

Theorem: A second order linear homogenous ODE with leading coefficient 1 and the other coefficients continuous has 2 linearly independent solutions. Let the equation be of the form $y'' + p(x)y + q(x) = 0$ with p and q continuous on some interval.

Proof: Since the equation is linear, we can prove uniqueness of the solution for any $(y_0, y_0') = (a, b)$ by proving that the difference between any such solutions is 0, meaning equivalently if we set $(y_0, y_0') = (0,0)$ then the solution must be identically 0.

Set $\mu(x) = \exp\left(\int_{x_0}^{x} p(t)dt\right)$. Now lets evaluate $(\mu y')' = \mu y'' + \mu' y'$

$= \exp\left(\int_{x_0}^{x} p(t)dt\right) y'' + p(x)\exp\left(\int_{x_0}^{x} p(t)dt\right) y' = -q(x)\mu(x)y(x)$ where the last equality is by the differential equation.

Integrating $(\mu y')' = -\mu q y$ from x_0 to x gives $\mu y' = -\int_{x_0}^{x} \mu(t)q(t)y(t)dt$, where since $y'(x_0) = 0$ we're ok for the lower limit on the left hand side.

Now we have that $y'(s) = -\frac{1}{\mu(s)}\int_{x_0}^{s} \mu(t)q(t)y(t)dt$, so we can integrate again to get:

$$y(x) = -\int_{x_0}^{x} \frac{1}{\mu(s)} \int_{x_0}^{s} \mu(t)q(t)y(t)dt\, ds.$$

Now since $q(t)$ is continuous on $[x_0, x]$ it is bounded there and by extension so is $\mu(t)$ and $\mu(t)^{-1}$ since $\mu(t)$ is never 0 as it is an exponential. Therefore we can use the triangle inequality for integrals and the fact that there is such a bound to get:

$$y(x) \leq C \int_{x_0}^{x} \int_{x_0}^{s} |y(t)| dt\, ds$$

Now define a region $R \coloneqq \{(t, s): x_0 \leq t \leq s \leq x\}$. This is a triangular region. Since we have a non-negative integrand, if it converges then it converges absolutely, so (cf level 6 technical results) we can integrate over the same triangular region in another order.

$$\int_{x_0}^{x} \int_{x_0}^{s} |y(t)| dt\, ds = \int_{t=x_0}^{x} \left(\int_{s=t}^{x} |y(t)| ds \right) dt = \int_{x_0}^{x} |y(t)|(x-t) dt$$

Now we have $|y(x)| \leq C \int_{x_0}^{x} |y(t)|(x-t)dt$. Let $M(x) = \sup_{x_0 \leq t \leq x} |z(t)|$, then

$|y(x)| \leq CM(x) \int_{x_0}^{x} (x-t)dt = CM(x) \frac{(x-x_0)^2}{2}$. Let s be the value such that y attains its maximum absolute value between x_0 and x. We need to justify that we can do this: We will show after this that y is bounded on any interval in order to justify this. Then $M(x) = |y(s)| \leq CM(s) \frac{(s-x_0)^2}{2} \leq CM(x) \frac{(x-x_0)^2}{2}$ since that last function is increasing in x since M is by definition and the square part clearly is. Therefore

$M(x) \leq CM(x) \frac{(x-x_0)^2}{2}$ so $M\left(1 - C\frac{(x-x_0)^2}{2}\right) \leq 0$. Since this holds for any x, it holds for $x = x_0 + h$, but the right hand side is M times a negative thing, so M must be 0, so our function must be 0 on the interval $[x_0, x_0 + h]$. We can repeat this argument $\frac{x-x_0}{h}$ times to get the desired result, and since x was arbitrary this is true everywhere, and its true before x_0 by just flipping everything around. So done.

Now pick $h < \sqrt{\frac{2}{C}}$, meaning $\frac{Ch^2}{2} < 1$. Then for every $t \in [x_0, x_0 + h]$, $\frac{Ch^2}{2} \leq \frac{C(x-x_0)^2}{2}$.

Theorem: Any solution on a closed bounded interval to a differential equation with continuous coefficients $y^{(n)} + p_{n-1}(x)y^{(n-1)} + \cdots + p_1(x)y' + p_0(x)y = 0$ is bounded

Proof: Set $Y(x) = \begin{pmatrix} y(x) \\ y'(x) \\ \vdots \\ y^{(n-1)}(x) \end{pmatrix}$. Then we have $Y'(x) = A(x)Y(x)$ where A is the n*n matrix

$$\begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -p_0 & -p_1 & -p_2 & \dots & -p_{n-1} \end{pmatrix}$$

You can convince yourself that the above equality is true by checking each component.

Now pick an interval J=[A,B]. Then the operator norm (Which we defined in one of the first lectures when we were doing multivariable calculus) is bounded on J since all the entries are bounded. Let M be the supremum of the operator norm of A in J.

We can integrate both sides of $Y'(x) = A(x)Y(x)$ from x_0 to x to get $Y(x) = Y(x_0) + \int_{x_0}^{x} A(t)Y(t)dt$. By the triangle inequality and the operator norm inequality that we derived in the earlier lecture where we introduced operator norms, we must have that $|Y(x)| \le |Y(x_0)| + \int_{x_0}^{x} |A(t)||Y(t)|dt$

$\le |Y(x_0)| + M\int_{x_0}^{x} |Y(t)|dt.$

Now set $v(x) \coloneqq |Y(x_0)| + M\int_{x_0}^{x} |Y(t)|dt$, then we have shown that $|Y(x)| \le v(x)$. $v'(x) = M|y(x)|$ so therefore $v'(x) \le Mv(x)$ since $|Y(x)| \le v(x)$. This certainly looks promising since it seems like v can be bounded by an exponential: Lets make this precise. Let $w(x) = v(x)e^{-Mx}$, then $w'(x) = e^{-Mx}(v'(x) - Mv(x))$ which means w is decreasing by $v'(x) \le Mv(x)$. Therefore since $v(x)e^{-Mx}$ is decreasing, $v(x) \le v(x_0)e^{M(x-x_0)}$. Therefore since $|Y(x)| \le v(x)$,

$|Y(x)| \le v(x_0)e^{M(x-x_0)} = |Y(x_0)|e^{M(x-x_0)}$. Last equality from how v was defined. Therefore on our interval [a,b], Y is bounded by $|Y(x)| \le |Y(x_0)|e^{M(b-a)}$ (since we can do the same argument backwards).

But now, since Y is bounded and its components are y, y', y'', y''', …, none of these components can be larger than the absolute value of Y. In particular y. So y is bounded. So done.

For an n'th order linear ODE, $y^{(n)}(x)$ is determined by all the lower derivatives of y from how we've defined the equations. Differentiating the entire ODE determines all higher derivatives. In fact, if the first n-1 derivatives of y are specified at a point x_0 then we can get a taylor series for y about x_0. At some fixed x, the vector $Y(x) = \begin{pmatrix} y(x) \\ y'(x) \\ y''(x) \\ \vdots \\ y^{(n-1)}(x) \end{pmatrix}$ defines a point in what is called phase space. We will use capital Y to denote this vector and lowercase y to denote a function y(x). As x varies, we trace out a trajectory in phase space.

Example, if $y'' + 4y = 0$, and we have a solution $y_1(x) = \cos(2x)$, then the vector $Y_1(x)$ is $\begin{pmatrix} \cos(2x) \\ -2\sin(2x) \end{pmatrix}$. If $y_2(x) = \sin(2x)$, then $Y_2(x) = \begin{pmatrix} \sin(2x) \\ 2\cos(2x) \end{pmatrix}$.

If we plot y' against y, then our vectors as x varies trace out an ellipse.

Now suppose we have a set of solutions $\{y_i(x)\}$ which are linearly dependent for all x, then it follows that the vectors $\{Y_i(x)\}$ are linearly dependent for all x.

Definition (Wronskian): Suppose we have n solutions to an ODE. Then we define $W(x) = \det\begin{pmatrix} \vdots & \vdots & \dots & \vdots \\ Y_1 & Y_2 & \dots & Y_n \\ \vdots & \vdots & \dots & \vdots \end{pmatrix} = \det\begin{pmatrix} y_1 & y_2 & \cdots & y_n \\ y_1' & y_2' & \cdots & y_n' \\ \vdots & \vdots & \vdots & \vdots \\ y_1^{(n-1)} & y_2^{(n-1)} & \dots & y_n^{(n-1)} \end{pmatrix}$. If the solutions are linearly dependent, W(x)=0 for all x. Otherwise, the solutions are linearly independent. However, it is not necessarily true that if W(x)=0 for all x then the solutions are linearly dependent.

Example: For the $y'' + 4y = 0$ example above, $W(x) = \det\begin{pmatrix} \cos(2x) & \sin(2x) \\ -2\sin(2x) & 2\cos(2x) \end{pmatrix} = 2.$

Lecture 13:

Theorem: Given any 2 solutions of a second order linear ODE y''+p(x)y'+q(x)y=0, if p(x) and q(x) are continuous on an interval I, then either W(x)=0 for all x in I, or W(x) is not 0 for all x in I.

Proof:

$$W = \begin{vmatrix} y_1 & y_2 \\ y_1' & y_2' \end{vmatrix} = y_1 y_2' - y_2 y_1'$$

Therefore $W' = y_1 y_2'' - y_2 y_1''$ by the product rule (Other terms cancel). From the differential equation,

$$W' = -y_1(py_2' + qy_2) + y_2(py_1' + qy_1)$$

$$W' = -py_1y_2' + y_2py_1' = -pW$$

So $W(x) = W(x_0)e^{-\int_{x_0}^{x} p(u)du}$ so if it is not 0 for some x it is not 0 for all x.

Note that sometimes people say "The solutions are linearly independent for some x iff they are linearly independent for all x", but be careful: Linearly independent in this sense literally means the solution vectors evaluated at a point. Linearly independent in the globgal sense means solutions forming a basis. These notes originally had a proof that 2nd order linear ODE's had 2 linearly independent solutions which interchanged these two definitions and was therefore flawed.

We need continuity because W' must be integrable and thus y_2'' and y_1'' being integrable gurantees this. Continuity gurantees that $y_1'' = -p(y_1') - q(y_1)$ is continuous and thus integrable, and for p to be integrable since we integrate it in the above proof, in which the integral should be defined to ensure its exponential is never zero.

Corollary: if p=0, W is constant, and in fact we can find it without solving the ODE.

Example: In Bessel's equation $x^2y'' + xy' + (x^2 - n^2)y = 0$, $W(x) = W(x_0)\exp\left(-\int_{x_0}^{x} \frac{1}{u} du\right) = \frac{C}{x}$ for som C.

Since we know that $y_1 y_2' - y_2 y_1' = W(x_0)e^{-\int_{x_0}^{x} p(u)du}$

So $\frac{d}{dx}\left(\frac{y_2}{y_1}\right) = \frac{1}{y_1^2} W(x_0)e^{-\int_{x_0}^{x} p(u)du}$ by the quotient rule, and this is the same result as from earlier to get a second solution given a first solution to a differential equation.

We can also solve equidimensional equations. These are equations of the form $axy' + by = f(x)$ or $ax^2y'' + bxy' + cy = f(x)$. These equations can be turned into a constant coefficients version using the substitution z=ln(x).

If g(x) is a solution to such an equation with f(x)=0, then consider $y = \frac{dg(kx)}{dx}$. $\frac{dy}{dx} = kg'(kx)$ by the chain rule, so $x\frac{dy}{dx} = (kx)g'(kx)$ and $x^2\frac{d^2y}{dx^2} = (kx)^2g''(kx)$. So $ax^2y'' + bxy' + cy =$ $a(kx)^2g''(ax) + b(ax)g'(ax) + cg(ax) = 0$, so the solutions scale: If g(x) is a solution then so is g(ax).

Example: Suppose we have $ax^2y'' + bxy' + cy = f(x)$ and let z=ln(x). $\frac{dy}{dz} = \frac{dx}{dz}\frac{dy}{dx} = x\frac{dy}{dx}$ since $\frac{dx}{dz} = \exp(z) = x$. $\frac{d^2y}{dz^2} = \frac{d^2x}{dz^2}\frac{dy}{dx} + \frac{dx}{dz}\frac{d^2y}{dx^2}\frac{dx}{dz}$ since $\frac{d}{dz}\left(\frac{dy}{dx}\right) = \frac{dx}{dz}\left(\frac{d}{dx}\left(\frac{dy}{dx}\right)\right)$ by the chain rule.

$= e^z\frac{dy}{dx} + \left(\frac{dx}{dz}\right)^2 y'' = xy' + x^2y''$. Therefore our equation becomes $a\frac{d^2y}{dz^2} + (b-a)\frac{dy}{dz} + cy = f(e^z)$. Therefore we get the characteristic equation $am^2 + (b-a)m + c = 0$, so our complementary function is either of the form $z = Ae^{k_1z} + Be^{k_2z}$ or, if there are repeated roots, $z = (Az + B)e^{kz}$. Reversing the substitution gives $y = Ax^{k_1} + Bx^{k_2}$ or, if repeated roots, $y = (Aln(x) + B)x^k$.

Here is an example of an interesting case: If we have $x^2y'' + xy' + y = 0$, and we do the substitution, although I won't go through the algebra explicitly, it turns out that we get that $y = Ax^i + Bx^{-i}$, and that if we force y real then our solutions are $Acos(\ln(x)) + Bsin(\ln(x))$. As x goes to 0, ln(x) goes to negative infinity, and as this happens, the cos and sin of ln(x) starts to oscillate wildly, as shown in the graph below. I will show a graph of cos(5ln(x)) so you can see the behavior yourself since if I put cos(ln(x)) you can't see the oscillations very well.

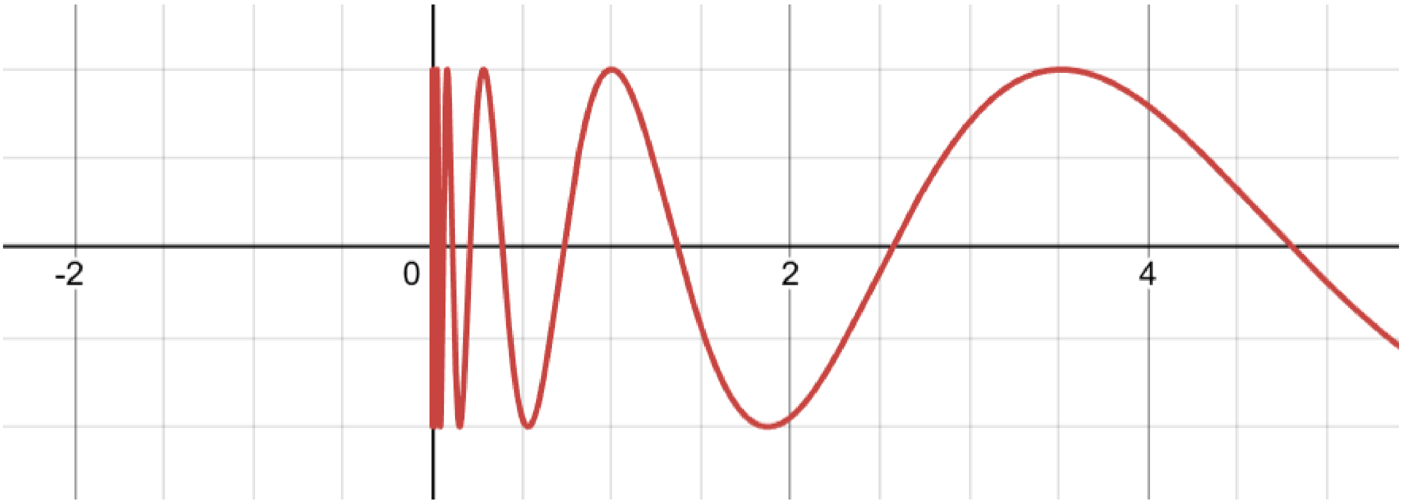


Image: The graph y=cos(5ln(x))

Lecture 14:

Note that we can try particular solutions to a differential equation even when there is a more complicated forcing (right hand side) term. Eg, by using common sense, our guess for the particular solution to y''-5y'+6y=2x+exp(4x) would be Aexp(4x)+Bx+C.

Also note that if the complementary function has repeated roots AND the right hand side shares a term, we can derive by the substitution method that we just have to multiply by x a second time.

For equidimensional equations, if there is a right hand side term of the form x^m, we try a particular integral of the form x^m in the case it does not match a C.F. term. This can be derived from the substitution shown last lecture. If we have a degenerate case with repeated roots or the forcing term matching a CF term, we can try multiplying by factors of $\log(x)$, since we have seen that we get factors of x in the case of the constant coefficient equation, and the substitution $z = \log(x)$ turns an equidimensional equation into such an equation in which we multiply by z in these degenerate cases.

Consider the general equation given by y''+p(x)y'+q(x)y=f(x) in which we have two linearly independent complementary functions y_1, y_2. We can use the solution vectors Y_1, Y_2. Let Y_p be the solution vector for our particular integral and write $Y_p = u(x)Y_1 + v(x)Y_2$. The purpose of this method is to try to solve for u and v as functions of x. Component-wise, we have that $y_p(x) = u(x)y_1(x) + v(x)y_2(x)$, and $y_p'(x) = u(x)y_1'(x) + v(x)y_2'(x)$. We also have that $y_p''(x) = uy_1'' + u'y_1' + vy_2'' + v'y_2'$. Therefore $f(x) = uy_1'' + u'y_1' + vy_2'' + v'y_2' + p(x)(uy_1' + vy_2') + q(x)(uy_1 + vy_2)$ by the original differential equation. Since y_1, y_2 are complementary functions, they satisfy the homogenous equation so we can get rid of these terms and write $u'y_1' + v'y_2' = f(x)$. However, since $u(x)y_1'(x) + v(x)y_2'(x)$ must be the derivative of $u(x)y_1(x) + v(x)y_2(x)$, it means the missing terms are 0. So $u'y_1 + v'y_2 = 0$. We can now see that $\begin{pmatrix} y_1 & y_2 \\ y_1' & y_2' \end{pmatrix}\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} 0 \\ f(x) \end{pmatrix}$. Therefore $\begin{pmatrix} u' \\ v' \end{pmatrix} = \frac{1}{W}\begin{pmatrix} y_2' & -y_2 \\ -y_1' & y_1 \end{pmatrix}\begin{pmatrix} 0 \\ f(x) \end{pmatrix} = \frac{1}{W}\begin{pmatrix} -y_2 f(x) \\ y_1 f(x) \end{pmatrix}$. Now we can integrate these functions to get u and v and thus the particular integral. We get that $y_p = y_2(x)\int \frac{y_1 f(x)}{W} dx - y_1(x)\int \frac{y_2 f(x)}{W} dx$. This is actually the general solution and fixing the constant of integration gives the particular solution – Changing it just changes the solution by multiples of the complementary function.

Lecture 15:

Example: Lets try to find a particular solution to $y'' + 4y = \sin(2x)$. We have two complementary functions: $\sin(2x), \cos(2x)$. We need to find W: We found a few lectures ago that W=-2 if $y_1 = \sin(2x), y_2 = \cos(2x)$. $y_p = \cos(2x)\int \frac{\sin^2(2x)}{-2} dx - \sin(2x)\int \frac{\sin(2x)\cos(2x)}{-2} dx = -\cos(2x)\left(\frac{1}{4}x - \frac{1}{4}\sin(4x)\right) - \frac{1}{4}\sin(2x)\cos(4x)$, dropping the constant since we just need 1 solution. We can use some trig identities and remove some multiples of the complementary function and we will get a particular solution $-\frac{1}{4}x\cos(2x)$.

Lets consider a differential equation that is supposed to represent a physical system but that is irrelevant since this is maths: The equation is $m\ddot{y} + b\dot{y} + ky = f(t)$. We note that if b=0, the complementary function is a sine or cosine wave with frequency $\frac{1}{2\pi}\sqrt{\frac{k}{m}}$. If $b < 2\sqrt{mk}$ (to ensure the characteristic equation has imaginary roots) then what happens is we get a sine wave that decays over time with a lower frequency – We can see this in practice: If $y'' + y = 0$ we get normal sin and cos, and if $y'' + y' + y = 0$ our general solution is $e^{-\frac{x}{2}}\left(A\sin\left(\frac{\sqrt{3}}{2}x\right) + B\cos\left(\frac{\sqrt{3}}{2}x\right)\right)$. We see the frequency decrease and "damping" occur. This is because of the solutions to the characteristic equation: As we increase the middle term the size of the imaginary parts decreases and we gain negative real part. There are irrelevant physical interpretations to this. If we increase b even more, then we just get exponential decay since we have no more sine terms. We note that the disappearence of the sine terms corresponds to the frequency going to zero as $b \to 2\sqrt{mk}$. If b is exactly $2\sqrt{mk}$, we have no more oscillation but the decay is not quite exponential since we have something like xe^{-bx}. We note that in this case and even in the exponential decay case (since we have two different exponential terms), we can have our function grow to start with before decaying. The long term behavior is dominated by the term with slower decay.

But that was just the complementary function. What if we have a forcing term?

We note that since the complementary function always decays in the examples we are considering, the long term behavior goes to a particular integral.

As an example, if the forcing term is a sine function that does not match any complementary function terms, then the behavior goes towards that sine function. If it does match the complementary function, the behavior goes to $xsin(x)$ with some constants added in there. I know I said the physical interpretation is irrelevant – and it is (I lose a brain cell every time this guy uses a physics term in the maths lecture which has happened like 27384293 times today) - but this is what resonant frequency and the breaking a wine glass trick and that kind of thing is about which is kinda cool I guess.

Lecture 16:

Motivation for what we will do: Consider a system that experiences a sudden force from time $T - \varepsilon$ to $T + \varepsilon$, think like striking something with a hammer so it moves suddenly, or something suddenly hitting the ground when it falls. Now think about the limit as $\varepsilon \to 0^+$. The resulting function is something you can recall from the proof of the central limit theorem and chi squared tables: The dirac delta "Function". Essentially a function that is 0 almost everywhere except for 1 point (0) where it is undefined and the integral at that point is 1. Recall that we can interpret this as the limit of a normal distribution with mean 0 and tinier variance, ie as it gets taller and thinner. We used this to get a pdf for a discrete distribution, and now we will use it in this context. We could use any other similar family of functions and take a "limit". But this is should be thought of as a distribution and not a function: it only makes sense when you integrate it.

Note that when this is included in a differential equation it only makes sense when we integrate it and impose all but the two highest derivatives of y to be continuous.

Now consider the equation $m\ddot{y} + b\dot{y} + ky = C\delta(T - t)$: The integral of this on both sides from $T - \varepsilon$ to $T + \varepsilon$ as $\varepsilon \to 0$ is $m\ddot{y} + b\dot{y} + ky = C\delta(T - t)$. Then we get $[m\dot{y} + by]_{T-\varepsilon}^{T+\varepsilon} + k\int_{T-\varepsilon}^{T+\varepsilon} ydx = C$ and we can take the limit as $\varepsilon \to 0$. If we require y to be continuous, then we see that the only term that survives is $[m\dot{y}]_{T-\varepsilon}^{T+\varepsilon}$, and then we see that $\dot{y}$ is discontinuous.

It may happen that if the gradient $\dot{y}$ suddenly changes then the nature of the differential equation makes some oscillation happen.

Properties:

- $\delta(t) = 0$ for all non-zero t.
- $\int_a^b \delta(t)dt = 1$ if a<0, b>0
- $\int_a^b f(t)\delta(t)dt = f(0)$ if f is continuous at 0, a<0, b>0
- $\int_a^b f(t)\delta(t - t_0)dt = \begin{cases} f(t_0): a < t_0 < b \\ 0: t_0 < a, t_0 > b \\ undefined\ if\ t_0 = a, t_0 = b \end{cases}$ provided f is continuous at t_0 since it has to be near $f(t_0)$ near t_0 for the whole limiting idea to work. If f is not continuous the integral is undefined.

In a general case: If $y'' + p(x)y' + q(x)y = \delta(x)$, integrating both sides from $-\varepsilon$ to ε insisting y, p, q continuous gives approximately (by continuity) $[y']_{-\varepsilon}^{\varepsilon} + [p(0)y]_{-\varepsilon}^{\varepsilon} + \int_{-\varepsilon}^{\varepsilon} q(0)ydx = 1$. We see that as earlier, $[y']_{-\varepsilon}^{\varepsilon}$ must be the only term that survives as $\varepsilon \to 0$, thus we see that y' must be discontinuous. The y, p, q continuous condition ensures that $p(x)y, q(x)y$ are bounded near 0 so the other terms indeed vanish. And also if y is not continuous y' will behave like a delta and who knows about y''. So it can be thought of as something "bouncing".

In the higher order case, we must ensure all but the last two derivatives of y we deal with are continuous, then the second to last one will be the discontinuous one, as above.

Example: $y'' - y = 3\delta\left(x - \frac{\pi}{2}\right)$ under conditions $y = 0$ at $x = 0, \pi$.

We need to solve this at either side of $\frac{\pi}{2}$. We can solve $y'' - y = 0$ for $0 < x < \frac{\pi}{2}$, where to satisfy the initial condition we need $y = Asinh(x)$. Similarly, in the other region, we must have $y = Asinh(\pi - x)$. The A's must be equal so that y is continuous at $\frac{\pi}{2}$. Now we must solve for A.

Now integrating both sides of the equation we have that for small ε, we have the approximate condition from integrating both sides that $[y']_{\frac{\pi}{2}-\varepsilon}^{\frac{\pi}{2}+\varepsilon} \approx 3$. Therefore the derivative changes by 3 when we reach $\frac{\pi}{2}$. Therefore we need $-Acosh\left(\frac{\pi}{2}\right) - Acosh\left(\frac{\pi}{2}\right) = 3$, so $A = -\frac{3}{2\cosh\left(\frac{\pi}{2}\right)}$. We can see the bouncy behavior in this diagram below.

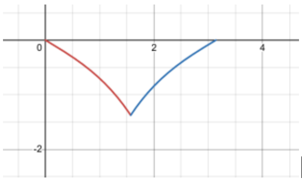


Image: Sketch of the solution above

Now we will define H(x) as the heaviside step function: This is basically an antiderivative of the dirac delta function: More precisely, this is 0 for negative x, 1 for positive x, and undefined at x=0.

Lecture 17:

Heuristically, integrating smoothens a function and differentiating makes a function less smooth. By integrating we can go from the dirac delta function to the heaviside step function and then if we integrate that we get $\frac{1}{2}(x + |x|)$ (dropping constant) which can be called the ramp function, and now our function is continuous everywhere. The idea is integrating things make them be one more time differentiable.

If we have H(x) on the right hand side of a differential equation with continuous coefficients and impose y continuous at x=0 then by integrating the equation and doing a similar argument to last lecture we see the discontinuity has to be in y''.

If we impose that y=0 for x<0 for the ODE y''+py'+qy=H(x) then we have to find a solution to y''+p(x)y'pq(x)y=1 for x>0 on condition y', y=0 when x=0. Or more generally we can solve it in both regions such that y and y' match at x=0 and conditions can be imposed to constrain the solutions.

Recall from A level further maths that we can solve equations like $au_{n+2} + bu_{n+1} + cu_n = f(n)$, and that in Level 6 we mentioned that we can derive the solutions using the same substitution method that we can do to derive the solution to ODE's.

These can actually arise if we try to approximate solutions to differential equations, ie

$$\frac{d^2y}{dx^2}|_{x_n} \approx \frac{y(x_n + h) + y(x_n - h) - 2y(x_n)}{h^2}$$

Example: The fibonacci numbers are 1, 1, 2, 3, 5, 8, 13, ... where each one is the sum of the previous 2. We therefore have the discrete equation $y_{n+2} - y_{n+1} - y_n = 0$ subject to $y_0, y_1 = 1$. The solutions to the characteristic equation are $\frac{1\pm\sqrt{5}}{2}$ so as we can show by substitution (Cf level 5, 6) the general solution is $A\left(\frac{1+\sqrt{5}}{2}\right)^n + B\left(\frac{1-\sqrt{5}}{2}\right)^n$. We can solve for constants to get the values for A and B. Usually, the value $\frac{1+\sqrt{5}}{2}$ is denoted ϕ, or the golden ratio. What ends up happening is we get $y_n = \frac{1}{\sqrt{5}}(\phi^{n+1} - (1-\phi)^{n+1})$. This is now a formula for the fibonacci numbers. A corrolary of this is that the ϕ^{n+1} term dominates so the ratio between fibonacci numbers approaches ϕ in the limit.

In coming lectures we will prove and apply more properties of series solutions to differential equations, which is useful when we cannot find closed form solutions. Meme about series solutions below:

Dexter Chua notes:

We will not prove these results, but merely apply them.

Evil Dexter Chua notes:

We will not apply these results, but merely prove them.

Best of both worlds:

We will prove these results and apply them.

Image: A meme about series solutions.

Lecture 18:

We will now do some definitions that will seem arbitrary but they will make sense once we start doing stuff.

Definition:

An ordinary point is a point where the equation $y'' + p(x)y' + q(x)y = f(x)$ or $y' + p(x)y = f(x)$ has p, q, f analytic meaning it has a valid taylor series in an interval around the point we are considering. We know from level 6 that this implies the existence of valid series solutions.

If we have an equation like $r(x)y'' + p(x)y' + q(x)y = f(x)$, then the point is ordinary provided that when we divide through by r(x) we still have coefficients that are analytic.

If a point x_0 is not an ordinary point it is a singular point. It is a regular singular point if and only if at that point the equation can be written as $(x - x_0)^2 y'' + p(x)(x - x_0)y' + q(x)y = f(x)$ with p, q and f analytic. Otherwise, we have an irregular singular point.

If we have a singular point, we know what to do. We start with the initial conditions and can find our series solution by either successively differentiating the equation or by solving an equation for the next coefficient based on the previous two: I will show an example of this so you get what I mean.

If $y'' + e^x y' + \sin(x)\, y = 0$ and $y = 1 + 2x + ax^2 + \cdots$ and we want to find a, here is what we can do:

$$2a + \cdots + e^x(2 + \cdots) + \sin(x)\,(1 + 2x + \cdots) = 0$$

$$2a + \cdots + \left(1 + x + \frac{x^2}{2} + \cdots\right)(2 + \cdots) + \left(x - \frac{x^3}{6} + \cdots\right)(1 + 2x + \cdots) = 0$$

Solving for the constant coefficient (since we showed we can multiply power series), we get that a must equal -1 since it is 0 on the right hand side. Therefore $y = 1 + 2x - x^2 + \cdots$ valid everywhere since the coefficients are valid everywhere.

Notice the similarity between the equidimensional equations (ie, $ax^2y'' + bxy' + cy = f(x)$)and regular singular points.

Example: Lets look into $(1 - x^2)y'' - 2xy' + 2y = 0$ about x=0 . We can divide through to get $y'' - \frac{2x}{1-x^2}y' + \frac{2}{1-x^2}y = 0$, where each of the coefficients have a valid taylor series for |x|<1, and thus (cf level 6) a series solution about x=0 must be valid for |x|<1. But we have singular points at $x = \pm 1$ which we want to classify. To check if 1 is a regular singular point, we know that therefore $(x-1)^2y'' - \frac{(x-1)^2 2x}{1-x^2}y' + \frac{2(x-1)^2}{1-x^2}y = 0$, and then $(x-1)^2y'' - (x-1)\frac{2x}{1+x}y' + \frac{2(x-1)}{1+x}y = 0$, where we now see that the coefficients are analytic so we have a regular singular point.

Example: $\left(1 + \sqrt{x}\right)y'' - 2xy' + 2y = 0$. We will look at $-\frac{2x}{1+\sqrt{x}}$ and $\frac{2}{1+\sqrt{x}}$. About x=0, we see that the second derivative of $\frac{-2x}{1+\sqrt{x}}$ is not defined at x=0 so we do not have an ordinary point.

We need to now look at $\frac{-2x^2}{1+\sqrt{x}}$ to check if we have a regular singular point, but this does not have a well defined third derivative about x=0. Therefore x=0 is an irregular singular point.

Theorem (For those who are in my year at Cambridge, yes this is "that" theorem, which works for higher order linear equations but we will only prove or apply it for first or second order):

The first part of the theorem says that if $x = x_0$ is an ordinary point of a linear ODE, then there are two linearly independent power series solutions, ie solutions of the form $\sum_{n=0}^{\infty} a_n(x - x_0)^n$ that are convergent in some interval around x_0. We have already proven this in level 6 and have shown in lecture 12 furthermore that it must be that every solution is of this form locally.

The second part of the theorem says that if $x = x_0$ is a regular singular point and the ODE has 0 on the right hand side, ie no forcing term, then there is at least one solution of the form $\sum_{n=0}^{\infty} a_n(x - x_0)^{n+\sigma}$ for some σ (Real or complex: The lecturer said real but I was the one who got him to correct it!) which is valid in an interval around x_0 but not necessarily valid at x_0. This is the same as $(x - x_0)^\sigma \sum_{n=0}^{\infty} a_n(x - x_0)^n$, where σ is real or complex and $a_0 = 0$. This is called a frobenius series. There is no gurantee that we have two linearly independent solutions of this form, but we will come back to this point.

This part is not surprising since the regular singular points were similar to the equidimensional equations so it makes sense that the form of the series solutions is also similar.

We would like to be able to say that if we solve for the coefficients the resulting series we get is actually valid in an interval. Therefore we will prove the second part using a proof almost identical to what we did for the first part in level 6.

Proof of the second part: Lets shift everything over so that $x_0 = 0$. Then we have $x^2y'' + p(x)xy' + q(x)y = 0$. Lets suppose $y = x^\sigma \sum a_n x^n$. Let $p = \sum b_n x^n$, $q = \sum c_n x^n$. Then we can differentiate y inside its radius of convergence (which we will show later is the common radius of congergence of p and q). We can write:

$$y = a_0 x^\sigma + a_1 x^{\sigma+1} + a_2 x^{\sigma+2} + \cdots$$

$$xy' = \sigma a_0 x^\sigma + (\sigma+1) a_1 x^{\sigma+1} + (\sigma+2) a_2 x^{\sigma+2} + \cdots$$

$$x^2 y'' = \sigma(\sigma-1) a_0 x^\sigma + (\sigma+1)\sigma a_1 x^{\sigma+1} + (\sigma+2)(\sigma+1) a_2 x^{\sigma+2} + \cdots$$

Since we need $x^2 y'' + p(x)xy' + q(x)y = 0$, we know that if such a series solution exists, then the x^σ term must satisfy $\sigma(\sigma-1)a_0 x^\sigma + b_0 \sigma a_0 x^\sigma + c_0 a_0 x^\sigma = 0$. But by assumption a_0 is not 0. We now get the indicial equation $\sigma(\sigma-1)a_0 + b_0\sigma + c_0 = 0$. Therefore if such a solution exists, there are two possible values of σ. Note that at an ordinary point if we write it in that equidimensional form then we will have $b_0 = 0 = c_0$ and the roots will be 0 and 1, meaning we will

For now we will consider the larger root if they are real, or any root if they are complex conjugate pairs.

We want to solve for a_n, and we will have to solve an equation by equating the $x^{n+\sigma}$ such as the following:

$a_n[(\sigma+n)(\sigma+n-1) + b_0(\sigma+n) + c_0] + stuff = 0$, which we can solve provided the stuff in the brackets is not zero. This is not a problem since we are picking the larger root of the indicial equation or a complex root – there won't ever be a root bigger by an integer. However, we will return to the case that the roots differ by an integer or are the same, as in the other cases we see that we have two linearly independent series solutions of these forms, but that case is harder. But in this case where the roots are not degenerate, we only need to deal with convergence issues then we will be done – since we have assumed convergence when doing all of this manipulation of the series.

Proposition: Frobenius series actually converge inside the common radius of convergence of the coefficients of the ODE, just like ordinary series solutions at ordinary points.

Proof: This is exactly like the level 6 proof of series solutions for the ordinary case, as we will see. We will again assume we can differentiate power series since once we show that it converges with this algebraic derivative we know that it converges with the true derivative.

Setup: Lets start with an equation of the form $x^2 y'' + xp(x)y' + q(x)y = 0$ where $p(x) = \sum p_n x^n$,

$q(x) = \sum q_n x^n$, and lets pick any root σ of the indicial equation. Then we will look for a solution of the form $y = x^\sigma \sum c_n x^n$ with $c_0 \neq 0$. Let R be the common radius of convergence of p and q, then we want to show that the series $\sum c_n x^n$ converges for every |x|<R. We know already the following:

- $Q(x)y = (\sum_{k=0}^\infty q_k x^k)(\sum_{n=0}^\infty c_n x^{n+\sigma})$
- $xy' = \sum_{n=0}^\infty (n+\sigma) c_n x^{n+\sigma}$
- Therefore, $P(x)xy' = (\sum_{k=0}^\infty p_k x^k)(\sum_{n=0}^\infty (n+\sigma) c_n x^{n+\sigma})$
- $x^2 y'' = \sum_{n=0}^\infty (n+\sigma)(n+\sigma-1) c_n x^{n+\sigma}$

We know from previous work (Level 6 -> Pure maths -> Power series properties for an explanation) that it must be the case that the coefficient of $x^{n+\sigma}$ is as follows:

- $xP(x)y' = (\sum_{n=0}^\infty [\sum_{k=0}^n p_k c_{n-k} (n-k+\sigma)] x^{n+\sigma})$
- $Q(x)y = (\sum_{n=0}^\infty [\sum_{k=0}^n q_k c_{n-k}] x^{n+\sigma})$

Now we can use the differential equation to equate coefficients:

$$(n+\sigma)(n+\sigma-1)c_n + \sum_{p=0}^{n}\left(p_p c_{n-p}(n-p+\sigma) + q_p c_{n-p}\right) = 0$$

Pulling out the p=0 term,

$$(n+\sigma)(n+\sigma-1)c_n + c_n(p_0(n+\sigma)+q_0) + \sum_{p=1}^{n}\left(p_p c_{n-p}(n-p+\sigma) + q_p c_{n-p}\right) = 0$$

$$c_n\left((n+\sigma)(n+\sigma-1) + p_0(n+\sigma)+q_0\right) = -\sum_{p=1}^{n}\left(p_p c_{n-p}(n-p+\sigma) + q_p c_{n-p}\right)$$

Lets define $(n+\sigma)(n+\sigma-1)+p_0(n+\sigma)+q_0$ to be A_n and write

$c_n A_n = -\sum_{p=1}^{n}\left(p_p c_{n-p}(n-p+\sigma)+q_p c_{n-p}\right)$ (We will call this (*))

Here the point is that the n^2 stuff in A_n dominates as n gets large, and the finite values of σ or p_0 or q_0 become negligable. In particular, because of this there exists a constant C such that we have

$|A_n| \geq C(n+1)^2$ if n is large enough. We just need n large enough because these radius of convergence arguments only care about what happens to n when it gets large, and whatever we want can happen in the first hundred terms, or first million terms, or whatever.

Now pick any x with |x|<R and set $A_r^{(1)} := \sum_{p=0}^{\infty}|p_p||x|^p$ and $A_r^{(0)} := \sum_{p=0}^{\infty}|q_p||x|^p$, which are all finite because of the fact that we are inside the common radius of convergence.

Define $M_n := \max_{0\leq k\leq n}|c_k||x|^k$. Take (*) and take absolute values and multiply by $|x|^n$ to get

$$|c_n||A_n||x|^n = \sum_{p=1}^{n}|p_p(n-p+\sigma)+q_p||c_{n-p}||x|^n$$

Now note that $|n-p+\sigma| \leq |n-p| + |\sigma| \leq n+|\sigma|$ by the triangle inequality, so certainly (by the triangle inequality again

$$|c_n||A_n||x|^n \leq (n+|\sigma|)\sum_{p=1}^{n}|p_p||c_{n-p}||x|^n + \sum_{p=1}^{n}|q_p||c_{n-p}||x|^n$$

$$|c_n||A_n||x|^n \leq (n+|\sigma|)\sum_{p=1}^{n}|p_p||c_{n-p}||x|^{n-p}|x|^p + \sum_{p=1}^{n}|q_p||c_{n-p}||x|^{n-p}|x|^p$$

Now we use the definition of M and the A's (noting that the n-p power only goes up to n-1):

$$|c_n||A_n||x|^n \leq (n+|\sigma|)M_{n-1}A_r^{(1)} + M_{n-1}A_r^{(0)}$$

Also, by a previous inequality,

$$C(n+1)^2|c_n||x|^n \leq |c_n||A_n||x|^n \leq (n+|\sigma|)M_{n-1}A_r^{(1)} + M_{n-1}A_r^{(0)}$$

And dividing through gives

$$|c_n||x|^n \leq M_{n-1}\left(\frac{(n+|\sigma|)A_r^{(1)}}{C(n+1)^2} + \frac{M_{n-1}A_r^{(0)}}{C(n+1)^2}\right)$$

For all large enough n, $\frac{n+|m|}{(n+1)^2} \leq \frac{K}{n+1}$ for some constant K. This is because the left hand side times n+1 as n gets large asymptotically approaches 1 and therefore, if say K=2 we eventually get a bound.

Now note that $M_n - M_{n-1} \leq |c_n||x|^n$ since the amount M_n can increase by from M_{n-1} is no more than the new term in our list of terms we are finding a maximum from, since they are all positive.

So, $M_n - M_{n-1} \leq M_{n-1}\left(\frac{(n+|\sigma|)A_r^{(1)}}{C(n+1)^2} + \frac{M_{n-1}A_r^{(0)}}{C(n+1)^2}\right)$ so $M_n \leq M_{n-1}\left(1 + \frac{B}{n+1}\right)$ for some finite constant B.

Now we know that $M_{n+2} \leq M_1 \prod_{k=0}^{n}\left(1 + \frac{B}{k+3}\right) \leq M_1 \prod_{k=0}^{n}\left(1 + \frac{B}{k+2}\right)$. Although these inequalities used assumed n was "large enough", we just have to have multiplication by a constant to deal with the terms that are not large enough, so it will not affect the radius of convergence argument.

Now as before, $\ln(M_{n+2}) \leq \ln(M_1) + \sum_{k=0}^{n} \ln\left(1 + \frac{B}{k+2}\right)$. By the same proof as in level 6 series solutions with S_n, we have that $M_n \leq Cn^a$ for constants c and a, and we again use the inequalities

$$|c_n|^{1/n} \leq \left(\frac{M_n}{|x|^n}\right)^{\frac{1}{n}} \leq \left(\frac{cn^a}{|x|^n}\right)^{\frac{1}{n}} = \frac{c^{1/n}n^{a/n}}{|x|}$$

Again, $c^{1/n}n^{a/n} \to 1$ as n gets large (You can take logs and use L'hopital to demonstrate this). We know from power series properties (Level 6) that the radius of convergence is given by $\frac{1}{\limsup_{n\to\infty} |c_n|^{1/n}}$. But this is at least |x| for any |x|<R, because $\limsup_{n\to\infty} |c_n|^{1/n} \leq \frac{1}{|x|}$ by the inequalities above, so done.

For completeness, I will mention the first order differential equations case and prove the frobenius series converges there:

$xy' + p(x)y = 0$ with $p(x) = \sum_{n=0}^{\infty} p_n x^n$, so $\frac{y'}{y} = -\frac{p(x)}{x} = -\frac{p_0}{x} + (Power\ series)$, so integrating gives $\ln(y) = -p_0 \ln(x) + (Power\ series)$, so $y = Cx^{-p_0}\exp\,(Power\ series)$ where if the power series has radius of convergence R so does the power series of its exponential (since the exponential converges everywhere – yet again see level 6 power series properties). So we do have a frobenius solution.

Now we have to discuss the last case. If we have an irregular singular point, the series solution method may fail completely so what happens is beyond this course. The regular singular point with roots that differ by an integer is something we will deal with next time.

Example: Lets find a series solution of $(1-x^2)y'' - 2xy' + 2y = 0$. For convenience, we multiply through by x^2 to get $(1-x^2)x^2y'' - 2x^3y' + 2x^2y = 0$. We can substitute in the series form and its derivatives to get:

$$(1-x^2)\sum_{n=2}^{\infty} a_n n(n-1)x^n - 2\sum_{n=1}^{\infty} a_n n x^{n+2} + 2\sum_{n=0}^{\infty} a_n x^{n+2} = 0$$

Splitting this further and re-indexing we get

$$\sum_{n=2}^{\infty} a_n n(n-1)x^n - \sum_{n=4}^{\infty} a_{n-2}(n-2)(n-3)x^n - 2\sum_{n=3}^{\infty} a_{n-2}(n-2)x^n + 2\sum_{n=2}^{\infty} a_{n-2}x^n = 0$$

We can now read things off easily and equate coefficients: For $n \geq 2$,

$$a_n n(n-1) - a_{n-2}(n-2)(n-3) - 2a_{n-2}(n-2) + 2a_{n-2} = 0$$

Now lets sipmlify (this is just doing stuff with quadratics)

$$a_n n(n-1) - a_{n-2} n(n-3) = 0$$

We can cancel n since it is non-zero to get

$$a_n(n-1) - a_{n-2}(n-3) = 0$$

Note that a_0 and a_1 are free to vary before we can determine the rest of the coefficients. By how the equation looks, we will want to determine the odd and even terms separately.

By considering odd terms, we see that all odd terms except for 1 must be 0 from the equation above. So we have one solution of the form $y = a_1 x$.

The even terms, we see that $a_n = \left(\frac{n-3}{n-1}\right) a_{n-2} = \left(\frac{n-5}{n-3}\right)\left(\frac{n-3}{n-1}\right) a_{n-4} = \cdots$ But we have the n-3's cancelling and will have similar for the rest of the terms: When we get up to n, we will have $a_n = -\frac{1}{n-1} a_0$. We have a solution of the form $y = a_0\left(1 - x^2 - \frac{x^4}{3} - \frac{x^6}{5} - \cdots\right)$. Recall that $\ln(1 \pm x) = \pm x - \frac{x^2}{2} \pm \frac{x^3}{3} - \frac{x^4}{4} + \cdots$. So we can check that $y = a_0\left(1 - \frac{x}{2}(\ln(1+x) - \ln(1-x))\right) = a_0\left(1 - \frac{x}{2}\ln\left(\frac{1+x}{1-x}\right)\right)$. Now we have two linearly independent solutions, so amazingly we can actually solve the equation in closed form to get $y = a_1 x + a_0\left(1 - \frac{x}{2}\ln\left(\frac{1+x}{1-x}\right)\right)$. We see that near x=1 this becomes undefined and we have logarithmic behavior. It turns out we can get logarithmic behavior at regular singular points and we will see this next lecture – This happens exactly in the roots repeated or differing by an integer case.

Lecture 19:

Example: Consider $4xy'' + 2(1-x^2)y' - xy = 0$. We can multiply by x to get it in "equidimensional" form so that $4x^2y'' + 2(1-x^2)xy' - x^2y = 0$, so we see that x=0 is a regular singular point.

Now lets try a solution $y = x^\sigma \sum_{n=0}^{\infty} a_n x^n$. We get

$\sum_{n=0}^{\infty} a_n x^{n+\sigma}[4(n+\sigma)(n+\sigma-1) + 2(1-x^2)(n+\sigma) - x^2] = 0$

We will look at the x^σ coefficient and try to equate it, which gives the indicial equation

$a_0(4\sigma(\sigma-1) + 2\sigma) = 0$

Since we assume a_0 is not 0, we therefore get that σ is 0 or $\frac{1}{2}$. Because of what we saw last lecture, we see that because these roots do not differ by an integer, there are two linearly independent series solutions that are of this form.

Now lets find the $x^{\sigma+1}$ coefficient: We get:

$a_1[4(\sigma+1)\sigma + 2(\sigma+1)] + a_0[0] = 0$, as those are all the terms that contribute.

Thus $a_1 = 0$.

Now lets think about $x^{n+\sigma}$ for $n \geq 2$. Since the taylor series of the coefficients of our ODE are finite, we can get a recurrence relation that we can use to quickly find our coefficients (although I'm not sure if there is a closed form for them). We will get that

$a_n[4(n+\sigma)(n+\sigma-1)+2(1-x^2)(n+\sigma)-x^2]+a_{n-2}[-2(n-2+\sigma)-1]=0$, as we have to contribute all products of terms that give a multiple $x^{n+\sigma}$.

We can write $2(n+\sigma)(2n+2\sigma-1)a_n = (2n+2\sigma-3)a_{n-2}$. (†)

Now lets plug in the possible values of σ into †

Case 1: $\sigma = 0$

$$a_n = \frac{(2n-3)}{2n(2n-1)}a_{n-2}$$

So we can find that $a_2 = \frac{1}{12}a_0, a_4 = \frac{5}{56}a_2 = \frac{5}{672}a_0$, etc. And all odd terms are 0.

$$y = a_0\left[1+\frac{x^2}{12}+\frac{5x^4}{672}+\cdots\right]$$

And from the recurrence relation we see that the ratio between consecutive terms approaches 0, which means it converges everywhere - this is certainly what we expect because the coefficients of the ODE converges everywhere so the proof last lecture implies this must work!

Case 2: $\sigma = \frac{1}{2}$

$$a_n = \frac{(2n-2)}{2n(2n+1)}a_{n-2} = \frac{n-1}{n(2n+1)}a_{n-2}$$

We will then get, and I won't go through the calculations, but we get that

$$y = a_0x^{\frac{1}{2}}\left[1+\frac{1}{10}x^2+\frac{1}{120}x^4+\cdots\right]$$

And similarly all odd terms are 0.

We found this time two linearly independent frobenius series solutions. As we saw when we derived this last lecture, this happens if the roots do not differ by an integer. If the roots do differ by an integer, we will investigate further.

Lets see first what happens if the root difference is a non-zero integer. In this case, we will get one series solution y_1 with the larger root that looks like $(x-x_0)^\sigma \sum_{n=0}^{\infty} a_n(x-x_0)^n$ since we won't have a problem.

Now we can use reduction of order to try to derive a second solution.

Suppose y_1 satisfies $x^2y_1'' + xp(x)y_1' + q(x)y_1 = 0$, then we will substitute $y = y_1u(x)$. This gives the equation $x^2\big(y_1u(x)\big)'' + xp(x)\big(y_1u(x)\big)' + q(x)y_1u(x) = 0$. Lets expand this out:

$$x^2y_1u'' + 2x^2y_1'u' + x^2y_1''u + xp(x)y_1'u + xp(x)y_1u' + q(x)y_1u = 0$$

But y_1 satisfies the ODE so

$$x^2 y_1 u'' + 2x^2 y_1' u' + xp(x) y_1 u' = 0$$

Now for $x \neq 0$

$$u'' + \frac{2y_1'}{y_1} u' + \frac{p(x)}{x} u' = 0$$

Note that since y_1 is a power series, $\frac{2y_1'}{y_1} = \frac{2\sigma_1}{x} + (Taylor\ series)$ where σ_1 is the other root so it integrates to $2\ln(x) + (Taylor\ series)$. $\frac{p(x)}{x} = \frac{p_0}{x} + (Taylor\ series)$ since p is analytic at x=0. Thus we can integrate and use an integrating factor to get $u' = Cexp\big(-(2\sigma_1 + p_0)\ln(x) + (Taylor\ series)\big)$. Therefore, $u' = x^{-2\sigma_1 - p_0}(Taylor\ series)$ since the exponential of a power series is still a power series valid on the same interval. We can use standard facts about roots of quadratics to get $\sigma_1 + \sigma_2 = 1 - p_0, \sigma_1\sigma_2 = q_0$. We know now that $1 - 2\sigma_1 - p_0 = \sigma_1 + \sigma_2 - 2\sigma_1 = \sigma_2 - \sigma_1$, which is the root difference. Therefore $u' = x^{\sigma_2 - \sigma_1 - 1}(Taylor\ series)$.

We can now integrate u, and here we will assume that the root difference is an integer. We have absolute convergence on the interval we care about and thus can apply the dominated convergence theorem (level 6 technical results) to swap sums and integrals.

We have $u = \sum_{n=0}^{\infty} c_n \int x^{n+\sigma_2-\sigma_1-1} dx$. If $\sigma_2 - \sigma_1$ is sometimes an integer, then since by assumption we know $\sigma_2 \leq \sigma_1$ the power of x in the integral will sometimes be -1 and we will get a log term.

We deduce by reversing the substitution that for the smaller root, we will have

$$y_2 = (x - x_0)^{\sigma} \sum_{n=0}^{\infty} b_n (x - x_0)^n + c y_1 \ln(x - x_0)$$

Where c is a constant that can be determined, but may be very difficult to determine as we have seen from the derivation above.

Note that if we work with the smaller root, we will get a recurrence relation for the coefficients, and the problem happens when we reach the larger root, as we will get $0a_k = something$. If this something is zero, we can pick a_k to be whatever we want and we have a second series solution. Either way, however, we know that a solution of the above form exists with the smaller root and so we have two linearly independent solutions (we will prove soon that this holds for frobenius type solutions as this is not obvious). Therefore there is a log term exactly when this "something" is not zero.

If $\sigma_2 = \sigma_1$ the log term will never have a constant of 0 at the front, but otherwise it might – such as in the case where we try to write an ordinary point in equidimensional form and get roots 0 and 1, but we clearly don't have a logarithmic solution in the ordinary point case, so the constant must be 0. Here is a more direct proof of this:

$y'' + py' + qy = 0 \Rightarrow x^2y'' + x(xp)y' + (x^2q)y = 0$, so the indicial equation is $\sigma(\sigma-1)+0\sigma+0=0$ which has roots 0 and 1.

Example: $x^2y'' - xy = 0$ where x=0 is a regular singular point. The roots of the indicial equation are 0 and 1. We get

$$\sum_{n=0}^{\infty}[a_n(n+\sigma)(n+\sigma-1)x^{n+\sigma} - a_n x^{n+\sigma+1}] = 0$$

For $\sigma = 1$, we can see that by equating coefficients, although I won't go through the details, we get that $a_n n(n+1) = a_{n-1}$. Therefore $a_n = \frac{a_0}{n!(n+1)!}$ so we have a solution $y_1 = a_0 x\left(1 + \frac{x}{2} + \frac{x^2}{12} + \cdots\right)$

If $\sigma = 0$, we cannot solve for coefficients. Therefore $y_2 = \sum_{n=0}^{\infty} b_n x_n + c y_1 \ln(x)$.

We have now shown that we have two linearly independent solutions of "frobenius type" regardless of the case, however this does NOT imply uniqueness because we are not at a point of the form $y'' + py' + qy = 0$ with p and q continuous. But for completeness I will prove that every solution is of this form. But be careful – This is true with a caveat.

Proof: Since this holds on any closed interval that avoids the regular singular point, it holds in general on the open interval (-R, 0) or (0, R) where R is the radius of convergence: For any point in these intervals there are only 2 linearly independent solutions in a closed interval around that point, but the frobenius solutions account for those so there can't be any others. Since the constants in front of the frobenius solutions that determine these solutions are the same in overlapping intervals, they are the same on the whole open interval. The caveat is that this is true on one side of a regular singular point that is not an ordinary point – we can absolutely have a different solution on both sides. unless we force the solution to work in complex numbers. For example, consider the differential equation $x^2y'' - 5xy' + 9y = 0$. The general solution as we would have said before is $x^3\left(A + Bln(x)\right)$, but it turns out that $x^2|x|$ is technially a legitimate third solution in the sense that it satisfies the ODE – only its third derivative doesn't exist but its first two do and really are 0 at x=0 and importantly, it is not of frobenius type – it is an almost-frobenius solution with different constants on each side of 0. The reason this can happen is that the theorem about two solutions assumes continuity when we divide through by the y'' coefficient, which does not happen in any interval containing x=0, and also in the derivation of the solution to equidimensional equations our substitution z=ln(x) only gave us positive number solutions if we want to force real numbers. We could use a substitution like z=ln(-x), but we will have the same problem.

Lecture 20:

We will go back to multivariable functions, see lecture 4 if you need to remind yourself of our earlier work on this.

Now we will assume that functions we are working with have all partial derivatives continuous so that it is differentiable in the lecture 4 sense and we can define gradient vectors or gradient matrices.

Review of a definition from L4 (Directional derivative): If we have a function from $\mathbb{R}^m \to \mathbb{R}$, then the input can be thought of as a vector. The directional derivative is then the rate that the function changes as we move along a certain vector, where the directional derivative with basis vectors is the partial derivatives.

Note that in the $\mathbb{R}^m \to \mathbb{R}$ case, it has a partial derivative matrix D (see lecture 4) and the directional derivative with respect to a vector x is Dx, which in this case is the same as D.x.

We write Df as ∇f, so we call this "D" vector Grad or ∇.

We can by the chain rule write $\mathrm{df} = \mathrm{ds}.\nabla\mathrm{f}$ where ds is the vector $\begin{pmatrix} dx \\ dy \end{pmatrix}$. Now we will assume s that we are differentiating with respect to is a unit vector.

Remark: ∇f is the direction of steepest ascent of f as a point.This is because $s.\nabla\mathrm{f}$ is maximized when s is in the same direction as ∇f by dot product properties.

Rmark: |∇f| is the maximuum slope of f at a point. This is because $|\nabla\mathrm{f}| = |s.\nabla\mathrm{f}\,|$ which is exactly the maximum slope since $\hat{\mathrm{s}}$ is a unit vector parallel to ∇f.

Remark: if s is parallel to contours of f (curves of constant f) then $\frac{df}{ds} = 0$.

Definition: A stationary point is where $\nabla\mathrm{f} = 0$, ie all directional derivatives are 0. In 3D we have local maxima and local minima, or saddle points. I will show image examples:

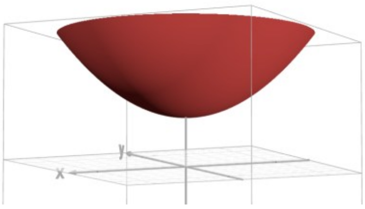


Image: Local minimum, local maxima are the obvious analog. Heuristically, contours are usually elliptical around such a point.

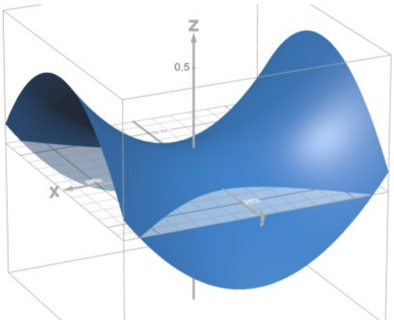


Image: Saddle stationary point. Heuristically, contours are usually hyperola-like around such a point. In a contour plot, contours will cross at a saddle point and look hyperbolic around it. As an example, this is exactly what the contours $x^2 - y^2 = c$ will do.

In 3D, The contours often cross if we are at a stationary point that is not a local minimum or maximum, ie a saddle point. However, this is not always the case.

Now assume further that f is a function which is analytic when we move it along any line. In 3D, or higher dimensions, we want to consider how f varies along the line $x(t) = x_0 + ts$. This is the usual taylor series: $f = f(x_0) + t(f_s) + \frac{1}{2}t^2(f_{ss}) + \cdots = f(x_0) + t(s.\nabla f) + \frac{1}{2}t^2\big(s.\nabla(s.\nabla f)\big) + \cdots$

We can unpack these more:

$$f = f(x_0) + t\left(s_x\frac{\partial f}{\partial x} + s_y\frac{\partial f}{\partial y}\right) + \frac{1}{2}t^2\left(s_x\frac{\partial}{\partial x} + s_y\frac{\partial}{\partial y}\left(s_x\frac{\partial f}{\partial x} + s_y\frac{\partial f}{\partial y}\right)\right) + \cdots$$

$$f = f(x_0) + t\left(s_x\frac{\partial f}{\partial x} + s_y\frac{\partial f}{\partial y}\right) + \frac{1}{2}t^2\left(s_x^2\frac{\partial^2 f}{\partial x^2} + s_y^2\frac{\partial^2 f}{\partial y^2} + 2s_xs_y\frac{\partial^2 f}{\partial x\partial y}\right) + \cdots$$

In the term $s_x^2\frac{\partial^2 f}{\partial x^2}+s_y^2\frac{\partial^2 f}{\partial y^2}+2s_xs_y\frac{\partial^2 f}{\partial x\partial y}$ we can write it as $(s_x \quad s_y)\begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy}\end{pmatrix}\begin{pmatrix} s_x \\ s_y\end{pmatrix}$ where the matrix in the middle is a symmetric matrix called the **Hessian matrix.** We assume all the second partial derivatives are continuous so that we can be allowed to do this. We will call this matrix H.

We can write the thingy above as follows:

$$f = f(x_0) + \left(dx\frac{\partial f}{\partial x} + dy\frac{\partial f}{\partial y}\right) + \frac{1}{2}\left(dx^2\frac{\partial^2 f}{\partial x^2} + dy^2\frac{\partial^2 f}{\partial y^2} + 2dxdy\frac{\partial^2 f}{\partial x\partial y}\right) + \cdots$$

Or we can write it in coordinate independent form where x is vectors:

$$f(x_0 + dx) = f(x_0) + dx.\nabla f + \frac{1}{2}(dx^T H dx) + \cdots$$

Where the derivatives and hessian matrix are evaluated at x_0.

Lecture 21:

The hessian matrix is a symmetric matric so if ∇f is 0 then what happens is the following:

$$f(x_0 + dx) = f(x_0) + \frac{1}{2}(dx'^T D dx') + \cdots$$

Where we diagonalize H (which we always can in this orthogonal way, see vectors and matrices). Then x' is about a perpendicular set of axes that is not necessarily the standard one.

Then we see that if D is all positive or all negative, we have a local extremum, but if D has some positive and some negative entries we have a saddle point.

Definition: A matrix is **positive definite** if all its eigenvalues are >0.

Proposition: An equivalent defintion asserts that $x^T H x > 0$ for all non-zero x.

Proof: $x^T H x = (Px)^T D(Px) = \left(\sqrt{D}Px\right)^T\left(\sqrt{D}Px\right) = \left|\sqrt{D}Px\right|^2 > 0$

Conversely, if $x^T H x > 0$ for all x then so is $(Px)^T D(Px)$. Then if we pick x such that Px is any basis vector, we must get something positive, so D's entries must be all positive.

Where $\sqrt{D}$ is the positive square root of everything in D.

Definition: A matrix is **negative definite** if $x^T H x < 0$ for all non-zero x. It is clear from this definition that this corresponds to a local maximum, and we will see shortly (same reason as above) this corresponds to all eigenvalues being negative.

If x' is the principal axes, then $dx'^T D dx' = \lambda_1 {dx'_1}^2 + \lambda_2 {dx'_2}^2 + \cdots \lambda_n {dx'_n}^2$

We now see that we are have a local maximum if all the eigenvalues are negative and a local minimum if all the eigenvalues are positive, since sufficiently close to the point the higher order terms will be smaller than this second derivative term. If some are 0 or some are negative and others are positive, further analysis is needed.

If a matrix is neither of these it is **indefinite**.

If all eigenvalues are non-zero but mixed signs we are guranteed to have a saddle point.

If some eigenvalues are zero, we need higher terms in the taylor series to classify the stationary points – exactly as in the single variable case.

Example: $f(x, y) = x^2 + y^4$ has a global minimum at $(x, y) = (0,0)$. $\nabla f = (2x, 4y^3)$. And we have that $H = \begin{pmatrix} 2 & 0 \\ 0 & 12y^2 \end{pmatrix}$. At 0, this is equal to $H = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$, so this does not tell us what kind of point it is even though we know that it is a minimum.

Definition: The **signature** of a matrix is the pattern of signs of the ordered subdeterminants of the leading principal minors of H.

Example: For $f(x_1, x_2, \dots, x_n)$, we consider the sign of the determinants

$$|f_{x1x1}|, \begin{vmatrix} f_{x1x1} & f_{x1x2} \\ f_{x2x1} & f_{x2x2} \end{vmatrix}, \begin{vmatrix} f_{x1x1} & f_{x1x2} & f_{x1x3} \\ f_{x2x1} & f_{x2x2} & f_{x2x3} \\ f_{x3x1} & f_{x3x2} & f_{x3x3} \end{vmatrix}, \dots$$

Theorem:

i) A matrix is positive definite if and only if the signature is +, +, +, +,...
ii) A matrix is negative definite if and only if the signature is -, +, -, +,...

Proof:

i) The forward implication is not too bad: For example with vectors like $x = (x_1, x_2, 0, \dots ,0)$, $x^T H x > 0$ always by the hypothesis, but then notice that this is the same as replacing x with (x_1, x_2) and H with its second principal minor. We need to show that if all principal minors have positive determinant then the matrix is positive definite. Starting with the first principal minor, it is positive definite since it is a positive number – this is trivial and there's not much to show here. Note that if H_k (which will denote the k'th principal minor) is positive definite and $Det(H_{k+1}) > 0$ then H_{k+1} must have an even number of negative eigenvalues. Suppose that H_{k+1} has two or more negative eigenvalues with associated eigenvectors u and v with components u_i, v_i. These can be chosen to be orthogonal since they are eigenvectors of a real symmetric matrix with distinct eigenvalues. Consider now $w = v_{k+1}u - u_{k+1}v$ which we will consider to be a row vector which by construction has no k+1 component. It follows that $wH_{k+1} = v_{k+1}uH_{k+1} - u_{k+1}vH_{k+1}$. So,

$wH_{k+1}w^T = (v_{k+1})^2 uH_{k+1}u^T - (v_{k+1})(u_{k+1})uH_{k+1}v^T + (v_{k+1})(u_{k+1})vH_{k+1}u^T + (u_{k+1})^2 vH_{k+1}v^T$ But the middle terms cancel since the two expressions coincide with a sign difference as the matrix vector parts are scalar and transposes of eachother. So,

$$wH_{k+1}w^T = (v_{k+1})^2 uH_{k+1}u^T + (u_{k+1})^2 vH_{k+1}v^T$$

This is less than 0 since u and v are eigenvectors of H_{k+1} with negative eigenvalues. But also, since w has no k+1 component, this is the same as using w_k having the first k components of w and writing $w_k H_k w_k^T$, but this is not negative. This contradiction allows us to conclude by induction that all principal minors are positive definite if they all have positive determinant, and in particular so is the last one. So done.

ii) This is easy once we have part (i) of the theorem. If we take minus a negative definite matrix it becomes positive definite as all the eigenvalues just change sign, and minus a matrix turns the determinant of the k'th minor into the determinant of minus the k'th minor, which by determinant properties multiplies it by $(-1)^k$, converting a - + - + - + ... signature into a + + + + + ... signature. So it is essentially just (i) applied to minus the matrix.

This theorem is very useful as we can now classify stationary points without computing the eigenvalues whenever the determinant of the matrix is non-zero (as this is equivalent to none of the eigenvalues being 0).

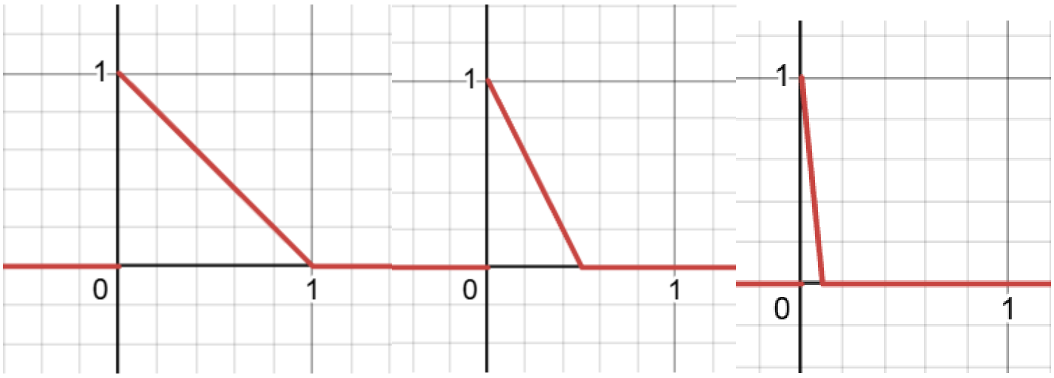
Now suppose f(x,y) has a stationary point at $x_0 = (x_0, y_0)$ and coordinates are alligned with the principal axes/eigenvectors of the hessian matrix. Now assume all the eigenvalues are all non-zero. Consider $x = x_0 + (dx, dy)$, then $f(x) = f(x_0) + \frac{1}{2}dx^2\lambda_1 + \frac{1}{2}dy^2\lambda_2 + o(x^2 + y^2)$. If we assume that we have that the higher order terms are 0, then indeed we will have hyperbolic/elliptical contours near a stationary point.

Now we want to say some things about the behavior of contours f(x,y)=c, in particular that they are actually continuous near a certain point whenever f has continuous second partial derivatives, and that they cross at a saddle point when the eigenvalues are not zero.

Preliminary definition 1: **Uniform convergence** - We know about how pointwise convergence works – a sequence of function converges to a function if it converges at every value. We can say "For each point, there is an ε>0 such that for all x, our sequence of functions f(x) is eventually, after some n, within ε of the limit function". This is different from uniform convergence in the sense that uniform convergence requires n not to depend on x. We need the function to eventually get arbitrarily close at all points at once, not just at each point. An example of a sequence of functions that converges pointwise but fails this is the following:

$$f_n(x) = \begin{cases} 0 < x < \dfrac{1}{n} : 1 - nx \\ Otherwise: 0 \end{cases}$$

This looks like the functions in these images:



Images: Shows examples of this for n=1, 2, 10.

The point is: For any point you choose, say k, after $> \frac{1}{k}$ steps, we will get to 0. So we converge pointwise to 0. It seems like we can pick x=0 and it will converge to 1 but we defined f(0) to be 0. We do not converge uniformly to 0 as we never satisfy that all points get arbitrarily close to 0.

Preliminary definition 2: **Lipschitz continuous** – A function is lipschitz continuous if we always satisfy that $|f(x) < f(y)| < k|x - y|$ for some k. This intuitively means its slope is never greater than k, however it need not be differentiable, it just can't ever be "vertical" or approach being vertical.

Preliminary definition 3: A family of functions F is **Equicontinuous** at a point x_0 if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for all functions f in F,

$$|x_0 - x| < \delta \Rightarrow |f(x_0) - f(x)| < \varepsilon$$

Lemma 1 (Stone weierstrass theorem for real functions on 2D rectangles): For any continuous real function defined on a closed rectangle [a,b]×[c,d] we can define a sequence of Lipschitz continuous functions that converges uniformly to our function.

Proof: Fix ε>0. We will use the known fact (Level 6 technical results) that our function is uniformly continous since it is continuous on a closed interval. So let δ be such that $|x - y| < \delta \Rightarrow |f(x) - f(y)| < \frac{\varepsilon}{2}$(we know this δ exists exactly by this level 6 result). Now what we will do is split our rectangle into rectangles with long side shorter than $\frac{\delta}{\sqrt{2}}$(actually we just need that the longest diagonal is shorter than δ). Now in this rectangle we will define our function g to coincide with f at the corners and then change linearly between the corners. Now for any t in this rectangle with x_0 one of the rectangle's corners and y_0 its opposite corner, $|x_0 - t| < \delta \Rightarrow |f(x_0) - f(t)| < \varepsilon$, and also $|y_0 - t| < \delta \Rightarrow |f(y_0) - f(t)| < \varepsilon$. Therefore what we have is that $f(x_0)$ and $f(y_0)$ are within a band of width 2ε around $f(t)$, so thus $g(t)$ which is between $g(x_0)$ and $g(y_0)$ is in that band, so we must have $|g(t) - f(t)| < \varepsilon$. Therefore if we let $\varepsilon_n \to 0$, we will get uniform convergence, and all of these individual functions are lipschitz continuous because of how we defined them: There slope is bounded by $\sqrt{g_x^2 + g_y^2}$, where these slopes are finite because they are the finite change in the function across the rectangle divided by the non-zero width of the rectangle. So done.

Lemma 2 (Arzela-Ascoli theorem specialized to real functions on 2D closed intervals): Let a sequence of functions f_n be uniformly bounded on a closed bounded interval (For our purposes, we will prove this for 1D or 2D intervals) and equicontinuous, then there exists a subsequence f_{n_k} that converges uniformly to a continuous function f.

Proof: Enumerate the rational points in $\mathbb{R}^2$. To do this, first enumerate the rationals, ie write them out in a list so that we have a bijection between the positive integers and the rational numbers. There are many ways to do this, but one way is to go around like this image below where the x is the numerator and the y is the denominator, but exclude duplicates or 0 denominators.

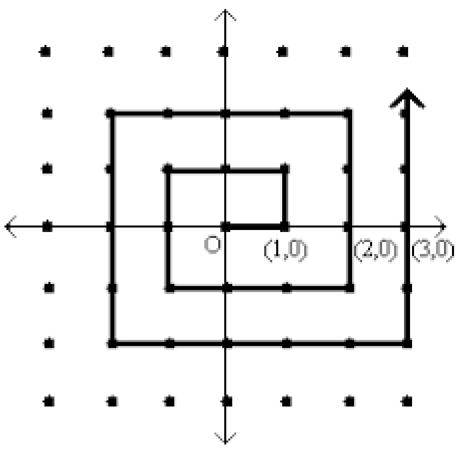


Image: Visual idea of the bijection

Then we know from level 6 that the cartesian product of two listable sets is listable, in fact by a similar diagram to the image above.

Now we filter this enumeration so we only have the rational numbers in our interval. So we have a list of all the rational points in our interval. Call this enumeration $x_1, x_2, \ldots$

Since f_n has a uniform bound M, there is a sequence $f_{n_{1,k}}$ such that $f_{n_{1,k}}(x_1)$ converges pointwise by Bolanzo-Weierstrass (Level 6 technical results). We can find a further subsequence $f_{n_{2,k}}$ of this such that $f_{n_{2,k}}(x_2)$ also converges. We can get an infinite chain of subsequences this way. Now we want to form a sequence of functions f_k defined by $f_k = f_{n_{k,k}}$. By construction, this converges at every rational point. Therefore, given any ε and any rational point x_k, we can find an integer N such that for all n, m>N, we have that $|f_n(x_k) - f_m(x_k)| < \frac{\varepsilon}{3}$. We're making progress.

Since the family F is equicontinuous, there must be an open interval around x_k such that for any s and t in that open interval, $|f(s) - f(t)| < \frac{\varepsilon}{3}$ for all f in our family of functions. Doing this for all x_k gives a covering of our interval using open sets. We will prove shortly the fundamental result that this must admit a finite subcover since the interval is closed, but first I will remark that you can see that if this is true then the theorem about uniform continuity will follow since we can pick delta to be less than the smallest thing in this subcover.

Note that Bolzano-Weierstrass applies on this interval since it is closed: Find a subsequence that converges in the x direction by the 1D version then find a subsequence of that that converges also in the y direction. Assume for a contradiction that we have a countably infinite covering of this interval in open sets $U_1, U_2, U_3, U_4, \ldots$ with no finite sub-cover. Now ennumerate the infinite covering and construct a sequence of points a_n such that a_n is not in the first n-1 open things – possible as those do not cover the entire interval. This has a convergent sub-sequence by bolzano-weierstrass $a_{n_k} \to x$. There is some U_j with x in U_j. But then for all sufficiently large k, a_{n_k} is in U_j since it gets arbitrarily close to x which is in U_j and not on the boundary. But then if $n_k > j$, this contradicts our construction. So we have a finite subcover $U_1, U_2, \ldots, U_j$. There also exists an integer K such that each of these open sets in our subcover contains one of the first K rationals in our list, otherwise our list would be missing every rational in that interval. Finally, for any t in our interval, there are j and k<K such that t and x_k belong to the same interval U_j. For this choice of k, we have, by the triangle ineuqality:

$$|f_n(t) - f_m(t)| \leq |f_n(t) - f_n(x_k)| + |f_m(x_k) - f_n(x_k)| + |f_m(t) - f_m(x_k)|$$

Where we pick n and m to be at least as large as N which is at least large enough such that for all k from 1 to K we have the above inequality $|f_n(x_k) - f_m(x_k)| < \frac{\varepsilon}{3}$, and also from how we defined the U's, $|f_n(t) - f_n(x_k)|, |f_m(t) - f_m(x_k)| < \frac{\varepsilon}{3}$ always. Therefore what we have is that for any t and fixed ε, $|f_n(t) - f_m(t)| \leq \varepsilon$ for m and n large enough. Therefore at each t, we see that the functions value is forced into an arbitrarily small band and thus we have pointwise convergence. We define f to be this pointwise limit, and then $|f_n(t) - f_m(t)| \leq \varepsilon$. Now letting m go to infinity and taking a pointwise limit, $|f_n(t) - f(t)| \leq \varepsilon$, so our sequence of functions converges uniformly. Since ε was arbitrary, we can say that, for example, $|f_n(t) - f(t)| \leq \frac{\varepsilon}{2} < \varepsilon$ so the inequality is strict. However, I need to show that a uniform limit of continuous functions is continuous.

Fix $\varepsilon > 0$. Pick N such that for all n>N, and all t $|f_n(t) - f(t)| < \frac{\varepsilon}{3}$. By continuity, for each x we have that $|x - y| < \delta \Rightarrow |f_N(x) - f_N(y)| < \frac{\varepsilon}{3}$ for some δ. Then, by the triangle inequality,

$$|x - y| < \delta \Rightarrow |f(x) - f(y)| \leq |f(x) - f_N(x)| + |f_N(x) - f_N(y)| + |f(y) - f_N(y)| < \varepsilon.$$

So done.

Lemma 3 (Peano existence theorem): Let $f(x, y)$ be continuous on an open interval D around (x_0, y_0), then the differential equation $y'(x) = f(x, y)$ with initial condition (x_0, y_0) has a solution in a neighbourhood about that point that is not necessarily unique. (It is unique under another mild condition, in fact this is when f is lipschitz continuous, but we do not need that so we will just do the existence theorem for now).

Proof: By replacing y with $y - y_0$ and similarly for x, we can assume that the initial condition is that we must pass through the origin. Since D is open, define a closed rectangle $R := [-x_1, x_1] \times [-y_1, y_1]$ contained in D. On R, the extreme value theorem implies that $\sup_R |f| \leq C < \infty$. Now by Lemma 1, pick a sequence of lipschitz continuous functions f_n converging uniformly to f with $\sup_R |f_k| \leq 2C < \infty$. We define the picard itrations $y_{k,n}: I = [-t_2, t_2] \to \mathbb{R}$ where we set that $t_2 = \min\left(t_1, \frac{y_1}{2C}\right)$, as follows:

$y_{k,0}(t) = 0$ and $y_{k,n+1}(x) = \int_0^x f_k\left(t, y_{k,n}(t)\right) dt$. They are well defined by induction as we have that $\left|y_{k,n+1}(x)\right| \leq \int_0^x \left|f_k\left(t, y_{k,n}(t)\right)\right| dt \leq |x| \sup_R |f_k| \leq t_2 2C \leq y_1$, and thus $(t, y_{k,n}(t))$ is within the domain of f_k. Also, by the triangle inequality for integrals,

$$\left|y_{k,n+1}(x) - y_{k,n}(x)\right| = \int_0^x \left|f_k\left(t, y_{k,n}(t)\right) - f_k\left(t, y_{k,n-1}(t)\right)\right| dt \leq L_k \int_0^x \left|\left(t, y_{k,n}(t)\right) - \left(t, y_{k,n-1}(t)\right)\right| dt$$

Where for each k, an L_k exists by the Lipschitz condition.

Now define $M_{k,n}(x) = \sup_{t \in [0,x]} \left|y_{k,n+1}(t) - y_{k,n}(t)\right| \leq L_k \int_0^x M_{k,n-1}(t) dt$. We also have that $M_{k,0}(x) = \sup_{t \in [0,x]} \left|y_{k,1}(t) - y_{k,0}(t)\right| = \sup_{t \in [0,x]} \left|y_{k,1}(t)\right| \leq \int_0^x |f_k(t, 0)| dt \leq 2C|x|$

Now we will prove by induction what we have the following bound for x in I for which we just proved the base case:

$$M_{k,n}(x) \leq \frac{(2CL_k^n |x|^{n+1})}{(n+1)!}$$

Lets do the induction step. Suppose this is true, then

$$M_{k,n+1}(t) \leq L_k \int_0^x M_{k,n}(t) dt \leq L_k \int_0^x \frac{(2CL_k^n |x|^{n+1})}{(n+1)!} dt \leq 2CL_k^{n+1} \int_0^x \frac{(|x|^{n+1})}{(n+1)!} dt = \frac{(2CL_k^{n+1} |x|^{n+2})}{(n+2)!}$$

As required. Crucially, this tends to 0 as n goes to infinity for all fixed x.

Now for x and x' in I, $\left|y_{k,n+1}(x') - y_{k,n+1}(x)\right| \leq \int_x^{x'} \left|f_k\left(t, y_{k,n}(t)\right)\right| dt \leq 2C|t' - t|$, and thus since this always holds, the family of functions $y_{k,n}$ is equicontinuous: For ε given pick $\delta = \frac{\varepsilon}{2C}$. Therefore by lemma 2, for each k, there is a subsequence y_{k,a_n} converging uniformly to a continuous function y_k.

$$\left|y_{k,a_n}(x) - \int_0^x f_k\left(t, y_{k,a_n}(t)\right) dt\right| = \left|y_{k,a_n}(x) - y_{k,a_n+1}(x)\right| \leq M_{k,a_n}(x) \to 0$$

Thus, for each fixed x, we conclude $y_k(x) = \int_0^x f_k(t, y_k(t))dt$ since the limit of $y_{k,a_n}(x)$ must coincide according to the inequality above, and the integral approaches $\int_0^x f_k(t, y_k(t))dt$ since f is continuous so we can pass the y limit through it and then we are bounded by 2C so we can use dominated convergence (level 6 technical results) to pass the limit through the integral. Since each $y_{k,n}$ has a uniform bound y_1, so does y_k (the limit of those. Now, by the triangle inequality for integrals,

$$|y_k(x) - y_k(x')| \leq \int_{x'}^x |f_k(t, y_k(t))| dt \leq 2C|x - x'|$$

So y_k is equicontinuous so it has a subsequence k_n that converges uniformly to a continuous function y. $y_{k_n}(x) = \int_0^x f_{k_n}\left(t, y_{k_n}(t)\right) dt$, therefore (supposing x is positive since the other way around is the same just with a sign flipped), $\int_0^x f_{k_n}\left(t, y_{k_n}(t)\right) dt \to \int_0^x f(t, y(t))dt$ by the dominated convergence theorem and continuity of f. The limits on each side must coincide, so $y(x) = \int_0^x f(t, y)dt$. By the fundamental theorem of calculus, y is our solution to the differential equation so we are done at last.

Lemma 4 (Implicit function theorem in 2 dimensions): If f is continuously differentiable in a neighbourhood of a point (x_0, y_0) and $f_y(x, y) \neq 0$ in that neighbourhood, then there exists a unique differentiable function g such that $y_0 = g(x_0), f(x, g(x)) = 0$ in a neighbourhood of x_0.

Proof: Lets try to find a g that works. By differentiating the equation $f(x, g(x)) = 0$ we get

$f_x + g'(x)f_y = 0$ so $g'(x) = -\frac{f_x}{f_y}$. Since f is continuously differentiable, f_x and f_y are continuous, and since $f_y \neq 0$, $\frac{f_x}{f_y}$ is continuous, so by Lemma 3 such a function g exists and it is continuous and differentiable.

Also, g(x) actually satisfies this equation:

- Define $h(x) := f(x, g(x))$.
- By the chain rule:

$$h'(x) = f_x(x, g(x)) + f_y(x, g(x))\, g'(x) = f_x - f_y \frac{f_x}{f_y} = 0.$$

- So h is constant. At x_0, $h(x_0) = f(x_0, y_0) = 0$, so $h(x) \equiv 0$ near x_0.

(Image: Screenshot of a proof of the claim above)

Note that for some fixed x, $f(x, y)$ as a function of y is increasing or decreasing (and thus injective) in a neighbourhood since f is continuously differentiable and $f(x_0, y)$ is not flat in y at our point. Therefore we have that if two solutions are the different, $f(x_0, g_1(x)) = f(x_0, g_2(x))$ so by injectivity the two solutions are the same, thus we have uniqueness.

Corollary: Contours around a non-stationary point are continuous. At this point, at least one of the partial derivatives is non-zero, so we can use that one to construct a differentiable local contour as in the theorem above.

Corollary: Contours cross near saddle points where no eigenvalues of the hessian matrix are 0, if the second partials are all continuous.

Proof: Shift everything so the saddle point is at (0, 0, 0) to simplify calculations.

$f(x,y) = \lambda_1 x^2 + \lambda_2 y^2 + o(x^2, y^2)$ where I dropped unnecessary factors of a half, $\lambda_1 > 0, \lambda_2 < 0$. Now scale axes so that $f(x,y) = x^2 - y^2 + o(x^2, y^2)$

Remark: If contours cross then we are at a stationary point since grad is 0 with respect to both of the principal axes at that point. The intersection between $f_x = 0, f_y = 0$ is only the origin in the neighbourhood since those curves have continuous contours near the origin by the implicit function theorem ($f_x = 0$ is a differentiable function x(y) near 0 since $f_{xx} \neq 0$ near 0 by continuity, for example) and their tangent directions are perpendicular so there are no stationary points nearby. Note that we can pick a δ small enough that $|f(x,y) - \frac{1}{2}(x^2 - y^2)| < \frac{1}{4}(x^2 + y^2)$ whenever we are at a distance less than δ.

Region A: $|x| \geq 2|y|$, then $x^2 \geq 4y^2$, so $x^2 - y^2 \geq 3y^2$, so $\frac{1}{2}x^2 - \frac{1}{2}y^2 \geq \frac{1}{2}x^2 - \frac{1}{8}x^2 = \frac{3}{8}x^2$, and

$$|f(x,y) - x^2 + y^2| < \frac{1}{4}(x^2 + y^2) \leq \frac{1}{4}\left(x^2 + \frac{1}{4}x^2\right) = \frac{5}{16}x^2$$

Therefore $f(x,y) \geq \frac{1}{2}(x^2 - y^2) - \left|f(x,y) - \frac{1}{2}(x^2 - y^2)\right| = \frac{1}{16}x^2 > 0.$

Region B: $|y| \geq 2|x|$, symmetric argument gives $f(x,y) \leq \frac{1}{16}y^2 < 0$

Therefore the contours cannot ever lie in these slices, but rather in the two middle wedges. Signs alternate as we move around the circle so they must cross 0 at four points. At any of these points, the grad is non-zero. By the implicit function theorem, if our neighbourhood is small enough that the above contours $f_x = 0, f_y = 0$ lie in those cones, we can find continuously differentiable contours corresponding to $f = 0$ since the derivative that can't be 0 is never 0. So done.

Example: Lets find and classify the stationary points of $f(x,y) = 4x^3 - 12xy + y^2 + 10y + 6$.

$$\nabla f = (12x^2 - 12y, -12x + 2y + 10)$$

Stationary points: $y = x^2$ therefore $-12x + 2x^2 + 10 = 0$ so $x, y = (1,1), (5,25)$.

$$f_{xx} = 24x, f_{xy} = f_{yx} = -12, f_{yy} = 2$$

Therefore the hessian matrix is

$$\begin{pmatrix} 24x & -12 \\ -12 & 2 \end{pmatrix}$$

When x=1, the signature is +-, when x=5 the signature is ++. Therefore x=1 is a saddle point and x=5 is a local minimum. And we will have closed loops as contours near the local minimum and hyperbolic-llike contours near the local maximum: We could find the eigenvectors and stuff in order to sketch these.

Here is an image of the above function as well as a contour plot, which illustrates previous ideas.

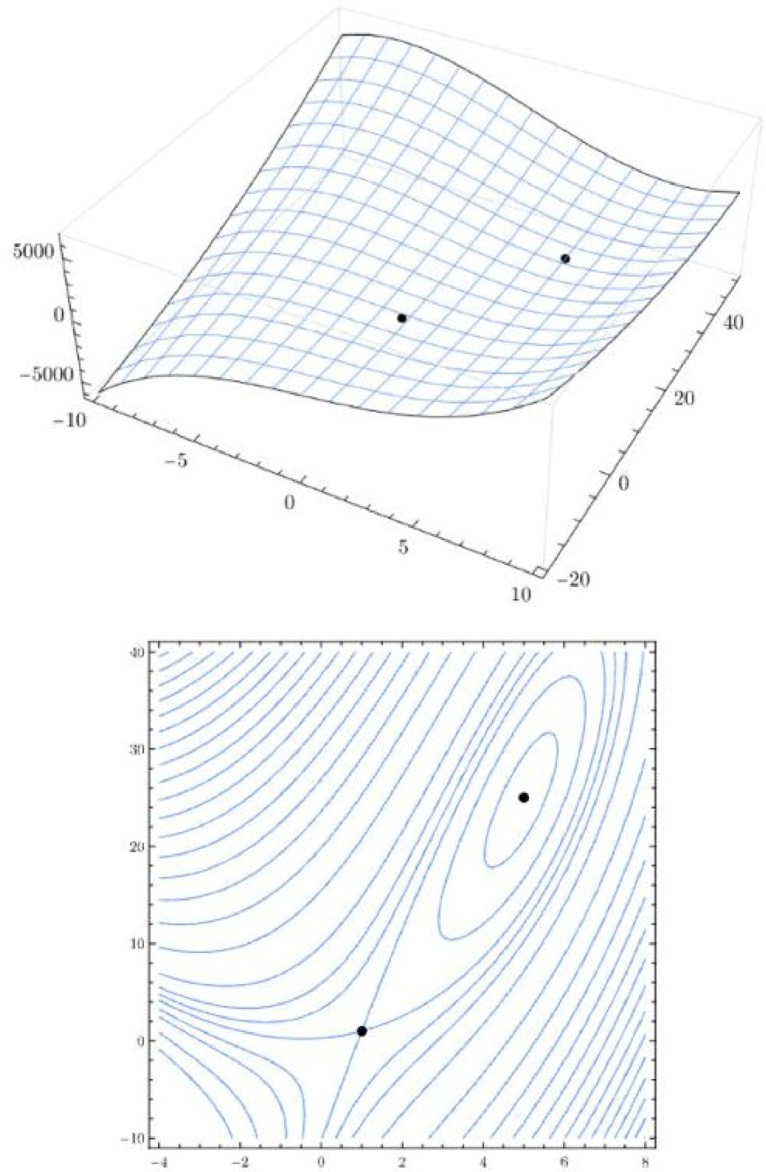


Image: The function and its contour plot with stationary points marked, illustrating above ideas about behavior.

Lecture 22:

Example: Consider 2 dependent variables $y_1(x), y_2(x)$ subject to $y_1' = ay_1 + by_2 + f_1(x)$ and also $y_2' = cy_1 + dy_2 + f_2(x)$. We can write this in matrix form as $\begin{pmatrix} y_1' \\ y_2' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}$. We can also solve this using A level techniques by eliminating a dependent variable and converting it into a second order equation. We will discuss how we can solve it using matrix methods since that scales up easier to higher order cases. We can also start with a second order ODE $y'' + ay' + by = f$ and define $y_1 = y, y_2 = y'$, then we could write this in matrix form as $\begin{pmatrix} y_1' \\ y_2' \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -b & -a \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \begin{pmatrix} 0 \\ f \end{pmatrix}$ to convert it into coupled first order ODEs.

Now we want to be able to solve a matrix equation such as $Y' = MY + F(x)$. We will write the solution to this as $Y_c + Y_p$ (ie, complementary function particular solution form). We will do the boring case where the matrix has constant coefficients.

Lets look for solutions of the form $Y_c = ve^{\lambda x}$ with v a constant vector. Then we want $Mv = {Y_c}' = \lambda v$, so this happens exactly when λ is an eigenvalue of M. Therefore if the eigenvalues are distinct we are guranteed to have n solutions of this form. However, we don't know the general solution using this method. For this course, we just have to be able to find some solutions, and not fully solve for the general solution.

We also will need to find a particular solution by guessing.

As an example, lets find some solutions to the matrix equation $Y' = \begin{pmatrix} -4 & 24 \\ 1 & -2 \end{pmatrix} Y + \begin{pmatrix} 4 \\ 1 \end{pmatrix} e^x$. The eigenvalues of this matrix are 2 and -8. Therefore there are two complementary functions which are $C_1 \begin{pmatrix} 4 \\ 1 \end{pmatrix} e^{2x}$ and $C_2 \begin{pmatrix} -6 \\ 1 \end{pmatrix} e^{-8x}$ (as those are the corresponding eigenvectors). Given the form of the forcing term, we will try a particular solution of the form ue^x. We will get an equation of the form $u = Mu + \begin{pmatrix} 4 \\ 1 \end{pmatrix}$ (cancelling factors of e^x). We can write this as $(I - M)u = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$ where we can find u as the matrix is invertible or else 1 would be an eigenvalue. We find that $u = \begin{pmatrix} -4 \\ -1 \end{pmatrix}$. Therefore we can write that a (not necessarily general, although we can prove it is in fact general if we do this with the other method) family of solutions is $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = C_1 \begin{pmatrix} 4 \\ 1 \end{pmatrix} e^{2x} + C_2 \begin{pmatrix} -6 \\ 1 \end{pmatrix} e^{-8x} + \begin{pmatrix} -4 \\ -1 \end{pmatrix} e^x$. If the matrix is not an eigenvalue further guesses would be needed – we could try multiplying by x, for example, as that often works in situations like this, but I don't know for sure. We would of course get the same answer if we do the thing where we convert it into a second order ODE and solve that.

We can consider phase portraits, similar to what we did in one dimensions earlier on. For autonomous systems of ODEs, we can sketch trajectory in phase space, like a vector field. At non-fixed points there will be one trajectory for each point.

Example: For $Y' = MY$ there is a fixed point when $Y = 0$, or in more generality when $Y \in Ker(M)$. If M has non-repeated eigenvalues and is invertible, we have (general: we can show this using the other method) solution $y = Av_1e^{\lambda_1 x} + Bv_2e^{\lambda_2 x}$. We have a few cases:

Case 1: The eigenvalues are real and have opposite signs. Lets say, for example, that $\lambda_1 > 0 > \lambda_2$. In this case, we can find real eigenvectors (see vectors and matrices lecture 19). Since the trajectory along one eigenvector will go away from the origin and the other one will go towards the origin, here is a sketch of what the phase portrait will look like.

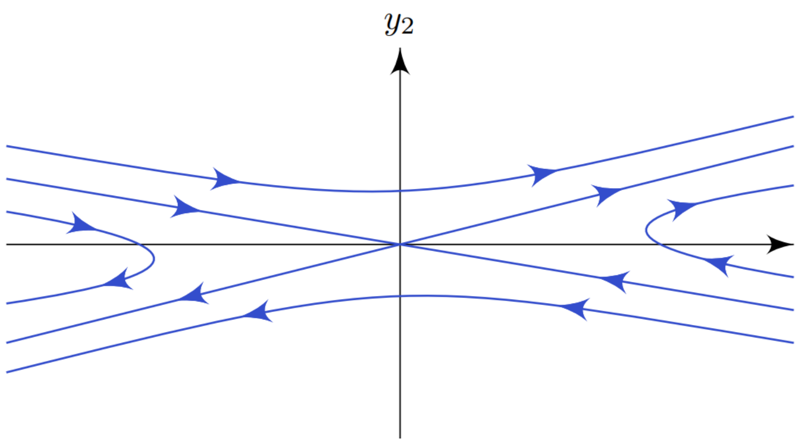


Image: Sketch of the phase portrait.

We call the intersection point a saddle point due to its similarity to actual saddle points.

Case 2: Eigenvalues have same sign and are real and eigenvectors are chosen to be real.

If they are both positive, everything is going away from the origin, like this sketch below. If they are both negative it will look like the sketch but everything will go towards the origin. Therefore each of this cases corresponds to whether the origin is stable or unstable. Note that the larger eigenvalue term will dmoinate at long distances so we will go closer to that one, which we see in the diagram below is v_2. At short distances the opposite happens and the smaller eigenvalue dominates, which is v_1. This explains the behavior that we observe.

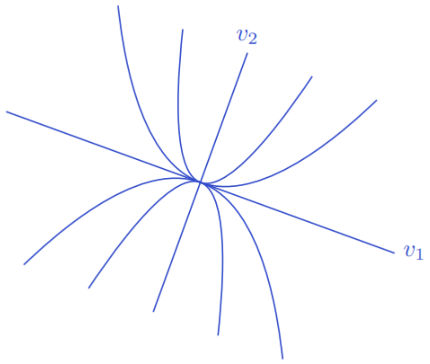


Image: Sketch of the phase portrait

Note that these lines are continuous in all these cases because they are a trajectory of a finite thing and are thus differentiable.

If the eigenvalues are repeated but not 0 all of 2D space will want to go proportional to where it is so we will just have lines going through the origin.

If the roots are not real, then we have sin and cos stuff, which means that we will go in circles, either towards or away from the origin depending on whether the real part is positive or negative, as in the sketch below.

If the eigenvalues are pure imaginary, we will have elliptical paths.

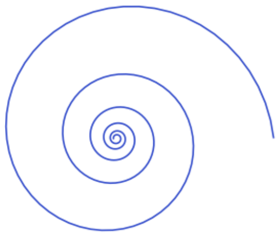


Image: Sketch of the phase portrait (spiral)

Lecture 23:

In the case where our phase portrait traces out an ellipse, the way to determine the trajectory is to find the vector Y' at any point Y in phase space.

We will now do this stuff for autonomous systems of 2 non-linear first order ODE's. A fixed point is a point where $y_1' = y_2' = 0$. If our equation is $y_1' = f(y_1, y_2), y_2' = g(y_1, y_2)$ then we need to solve simaltaneously $f(y_1, y_2) = 0, g(y_1, y_2) = 0$ to find the fixed points. We can investigate the stability of these fixed points and the behavoior around them. We will write

$\left(y_1(x), y_2(x)\right) = (x_0 + \xi(x), y_0 + \eta(x))$ where (x_0, y_0) is a fixed point. If f has a taylor series locally along each direction, we can write that

$\xi'(x) = y_1' = f(x_0 + \xi(x), y_0 + \eta(x)) = f(x_0, y_0) + \xi(x) f_{y_1}(x_0, y_0) + \eta(x) f_{y_2}(x_0, y_0)$. We can write this in matrix form (using $f(x_0, y_0) = 0 = g(x_0, y_0)$) in matrix form as

$$\begin{pmatrix} \xi' \\ \eta' \end{pmatrix} = \begin{pmatrix} f_{y_1} & f_{y_2} \\ g_{y_1} & g_{y_2} \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix}$$

We can then use the eigenvalues to determine the behavior around the fixed point. We will go through an example. Consider the following system of equations which has relevance to the real world:

$$y_1' = 8y_1 - 2y_1^2 - 2y_1y_2$$

$$y_2' = y_1y_2 - y_2$$

The fixxed points are when $8y_1 - 2y_1^2 - 2y_1y_2 = 0, y_1y_2 - y_2 = 0$

$$2y_1(4 - y_1 - y_2) = 0, y_2(y_1 - 1) = 0$$

So $y_1 = 0$ or $y_1 = 4 - y_2$ and $y_2 = 0$ or $y_1 = 1$. By considering the four combinations of possibilites, we get that (0,0), (1,3) and (4,0) are the fixed points.

So our matrix $\begin{pmatrix} f_{y_1} & f_{y_2} \\ g_{y_1} & g_{y_2} \end{pmatrix} = \begin{pmatrix} 8 - 4y_1 - 2y_2 & -2y_1 \\ y_2 & y_1 - 1 \end{pmatrix}$.

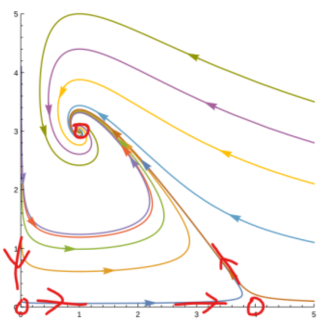
At (0,0), $M = \begin{pmatrix} 8 & 0 \\ 0 & -1 \end{pmatrix}$ so the eigenvalues are 8 and -1 with the axes as eigenvectors. What will happen is we will go towards the fixed point in the y direction and away from it in the x direction.

At (4,0), $M = \begin{pmatrix} -8 & -8 \\ 0 & 3 \end{pmatrix}$ which has eigenvalues -8 and 3. The corresponding eigenvectors are $\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 8 \\ -11 \end{pmatrix}$. The eigenvalues have different signs so we have another saddle point.

At (1,3), $M = \begin{pmatrix} -2 & -2 \\ 3 & 0 \end{pmatrix}$. The eigenvalues are $-1 \pm i\sqrt{5}$. This tells us that close enough to the point, the real part of the eigenvalues will be negative (since we are assuming that the derivatives of f and g are continuous) so the trajectory of the curve will have to be inwards, as that would be the trajectory in the case we force M constant.

At $(\xi, \eta) = (1,0)$ the derivative is $(-2,3)$ so this suggests we spiral anticlockwise inwards.

We can try to put our sketches together we get this:



Which confirms that (1, 3) is the only stable fixed point.

We can study partial (not ordinary) differential equations. These are equations that involve partial derivatives and/or involve multiple variables. This is a massive area of mathematics but we only have time to do a few simple examples.

Consider $f(y,x)$ subject to the equation $\frac{\partial f}{\partial x} - c\frac{\partial f}{\partial y} = 0$. We will do this by asking how f varies along a path $y = A - cx$. Then we have $f(A - cx, x)$, and we can write $\frac{df}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx}$ by the chain rule. Buut then by the PDE, $\frac{df}{dx} = 0$ when $\frac{dy}{dx} = -c$. Therefore f is constant along paths of the form $y = A - cx$. We get that $f(A - cx, x) = g(x_0)$ where g is any differentiable function and $g(x_0) = f(A - cx_0, x_0)$ Rearranging, we get that $f(y,x) = g(y + cx)$ for all a and some differentiable function g. The reason g can be any differentiable function is because we check that $f_x(A) = g_x(y + cx) = cg'(y + cx)$ and that $f_y(A) = g_y(y + cx) = g'(y + cx)$ so the PDE is satisfied. And note that the partial derivatives of f are assumed to exist for this PDE, and this happens exactly if g is differentiable. If we fix one variable here and vary the other variable over time and look at f as a function of that variable, it will move smoothly along the axis at a rate c units per time unit. For this reason, the equation above is often called the **wave equation**. These constant lines are called the **characteristics**.

Lecture 24:

We can take the wave equation and impose an initial condition, such as $f(y,0) = y^2 - 3$. We know that the general solution is $y^2 - 3 = f(y,0) = g(y)$. Therefore we know g. So we have that the general solution is $f(y,x) = (y + cx)^2 - 3$.

We can now add a forcing term and get an equation like

$\frac{\partial f}{\partial x} + 5\frac{\partial f}{\partial y} = e^{-x}$ with initial condition $f(y,0) = e^{-y^2}$ where we need to solve for f. The characteristics are of the form $y = A + 5x$. We can use the chain rule to get that along these paths,

$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial x} + 5\frac{\partial f}{\partial y} = e^{-x}$. Therefore along these paths, $f(x_0 + 5x, x) = g(x_0) - e^{-x}$ But then we know that $f(y,0) = e^{-y^2} = g(y) - 1$ so $g(y) = 1 + e^{-y^2}$ so $f(y,x) = -e^{-(y-5x)^2} - e^{-x}$ as A must be consistent when we set x to 0, and we rearrange to set $x_0 + 5x \coloneqq y$.

Now consider the equation $\frac{\partial^2 f}{dx^2} - c^2\frac{\partial^2 f}{dy^2} = 0$ (second order wave equation). We can factor this and then impose f is twice differentiable in order to let the partial derivatives commute so we can commute these factors to then get that:

$$\left(\frac{\partial}{\partial x} - c\frac{\partial}{\partial y}\right)\left(\frac{\partial}{\partial x} + c\frac{\partial}{\partial y}\right)f = 0$$

Note that both $f(y,x) = g(y + cx)$ and $f(y,x) = h(y - cx)$ with g,h twice continuously differentiable are solutions to this equation because each one goes to 0 by one of the differential operators.

We will now prove that in fact $g(y + cx) + h(y - cx)$ is the general solution. Now let $\xi \coloneqq (y + cx)$ and $\eta \coloneqq (y - cx)$. Now $\frac{\partial}{\partial y}|_x = \frac{\partial \xi}{\partial y}|_x \frac{\partial}{\partial \xi}|_\eta + \frac{\partial \eta}{\partial y}|_x \frac{\partial}{\partial \eta}|_\xi$ by the chain rule.

Similarly $\frac{\partial}{\partial x}|_y = \frac{\partial \xi}{\partial y}|_x \frac{\partial}{\partial \xi}|_\eta + \frac{\partial \xi}{\partial y}|_x \frac{\partial}{\partial \xi}|_\eta$. Because the terms like $\frac{\partial \xi}{\partial y}$ are just c or -c, we could go through the calculations to get $\frac{\partial}{\partial x} - c\frac{\partial}{\partial y} = -2c\frac{\partial}{\partial \eta}, \frac{\partial}{\partial x} + c\frac{\partial}{\partial y} = 2c\frac{\partial}{\partial \xi}$. The wave equation now says that we have that $-4c^2\frac{\partial^2 f}{\partial \xi \partial \eta} = 0$. We now know that $\frac{\partial^2 f}{\partial \xi \partial \eta} = 0$, so $\frac{\partial f}{\partial \xi} = \hat{g}(\xi)$, so $f = h(\eta) + g(\xi)$, remembering from several lectures ago that by undoing a partial derivative we have to add a function of the other variable and not just a constant.

Now we will impose the initial conditions $f(y,0) = \frac{1}{1+y^2}$ and $f_x(y,0) = 0$ and f is twice continuously differentiable. Now we know from what we just did that $f(y,x) = g(y+cx) + h(y-cx)$ so when $g(y) + h(y) = \frac{1}{1+y^2}$ by setting y=0. Also, by the single variable chain rule, $f_x(y,0) = cg'(y) - ch'(y)$, so if we integrate this wrt y we see that $g(y) - h(y) = A$ is a constant. We can write $g(y) = A + h(y)$ so $2g(y) - A = \frac{1}{1+y^2}$, so we can rearrange to get that $g(y) = \frac{1}{2(1+y^2)} + \frac{A}{2}$ and that $h(y) = \frac{1}{2(1+y^2)} - \frac{A}{2}$. Now we can substitute this back into our general solution to get that $f(y,x) = \frac{1}{2}\left[\frac{1}{1+(y+cx)^2} + \frac{1}{1+(y-cx)^2}\right]$ since the A's cancel.

Now lets sketch what this looks like if we fix x and plot f against y. Then when x=0, we just have the simple equation $f(y) = \frac{1}{1+y^2}$ which looks like a sort of bell curve. But then as x changes we will get two bell curves that sort of move over time, but they will be half as tall as the original one, but when x is near 0 they will add together to give a taller bell curve. I will show images of this.

Image: The graphs of this for varying values of x. We see now why it is the wave equation, and we see the waves add together or interfere when x is near 0. We get even more interesting behavior if we had a third independent variable, but that is beyond this course.

In this course we have seen how rich the behavior of these differential equations or systems can be, it's really interesting.