Maths for Computer Science
Calculus

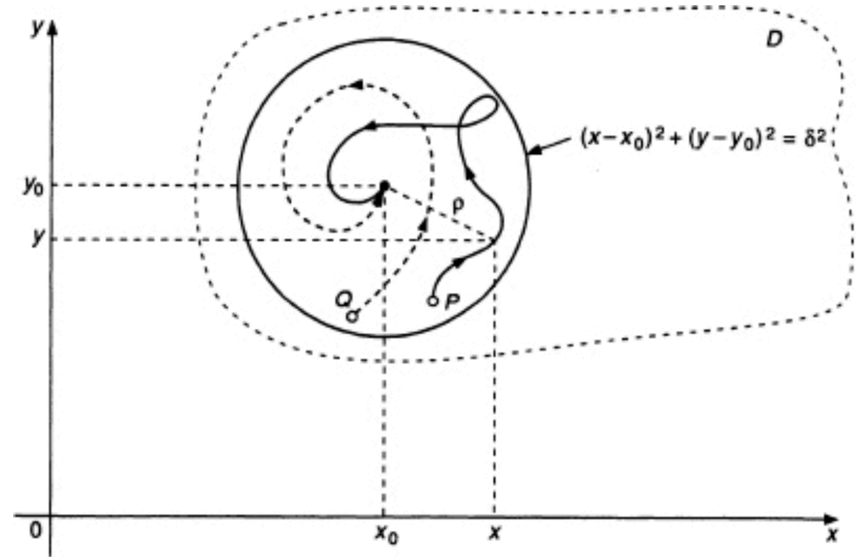
Prof. Magnus Bordewich

Durham University

Parameterised curves

Limits of multivariate functions

Recall $\lim_{(x,y)\to(x_0,y_0)} f(x,y) = L$ exists if and only if L is independent of the path taken (provided $\rho \to 0$).

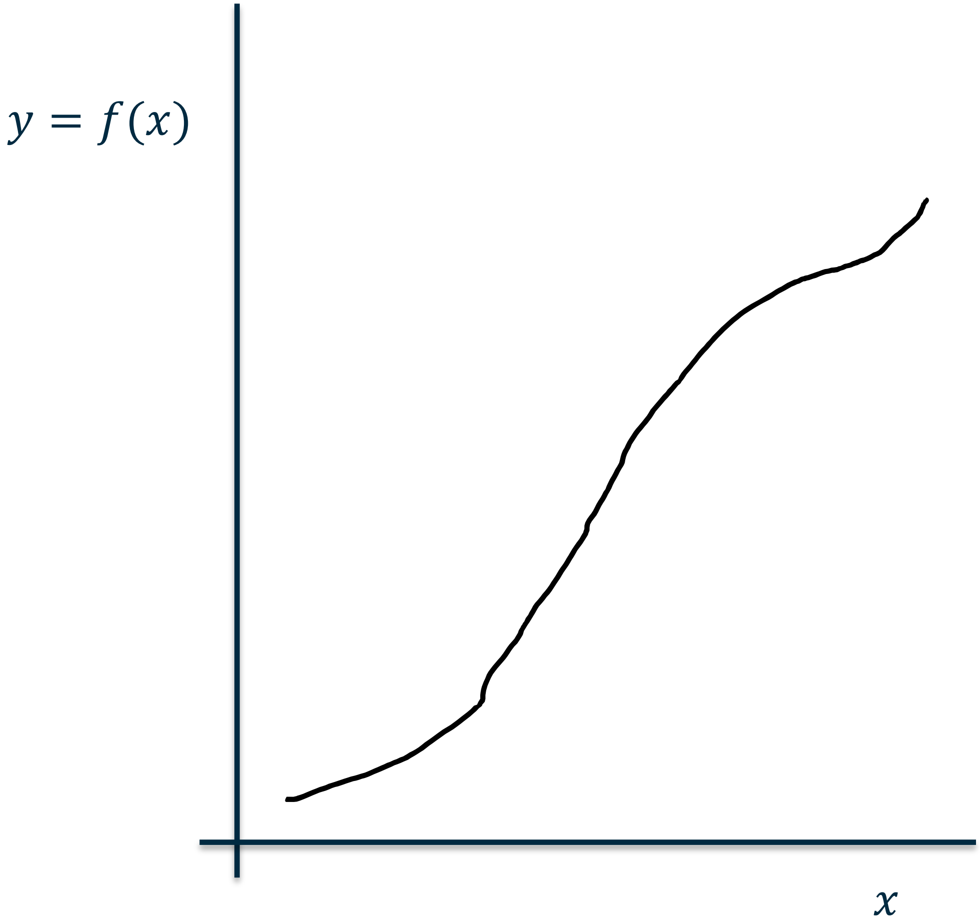


How do we deal with the path?

Durham University

Image: *Mathematics for Engineers and Scientists, Alan Jeffrey*

Graphs

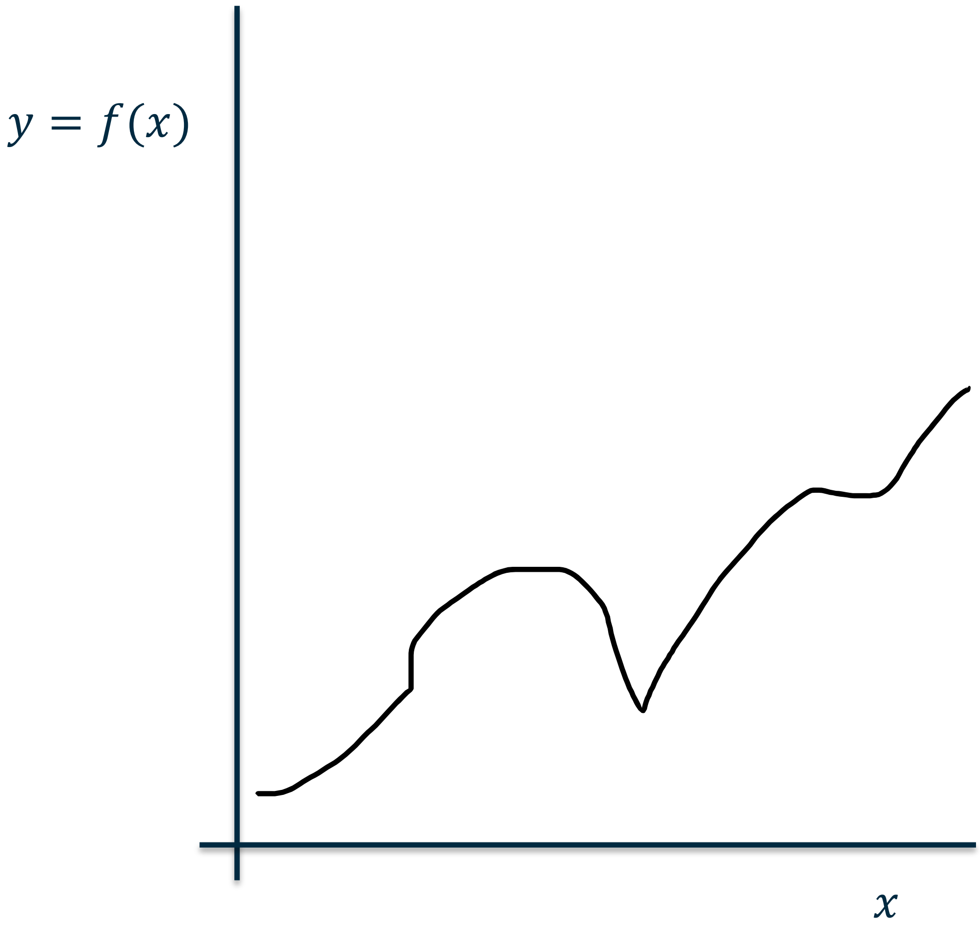


Does this function have an inverse?

The function f has an inverse because it is one-to-one and onto.

If it is continuous then it must be **strictly monotonic**.

Graphs

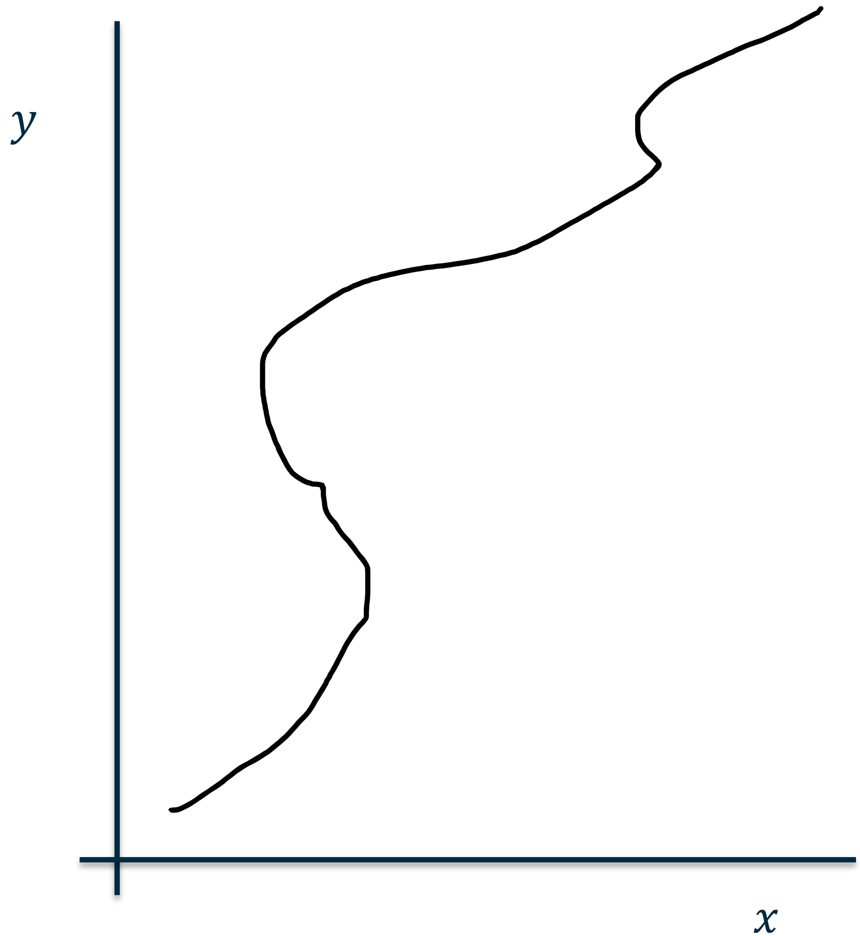


Does this function have an inverse?

The function f has no inverse because it is not one-to-one.

It is not **monotonic**.

Graphs

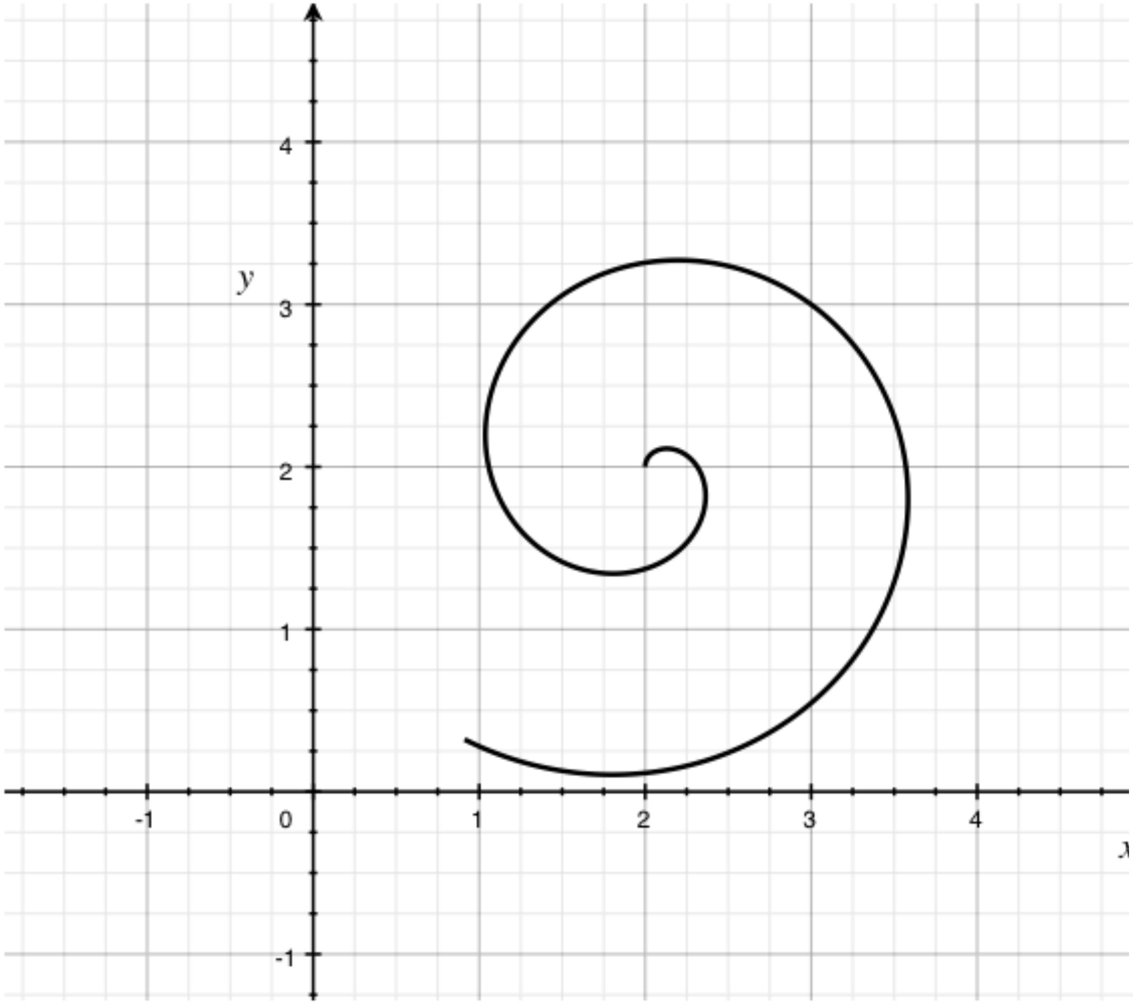


Does this function have an inverse?

This line is not a function.
It does not map each value of x
To a unique value of y.

Durham University

Parameterised curves



What if we want to represent complex curves such as this spiral?

We **parameterise** the curve: rather than explicitly assuming y is a function of x, we instead take x and y to be given by functions of some new parameter t.

Here I took

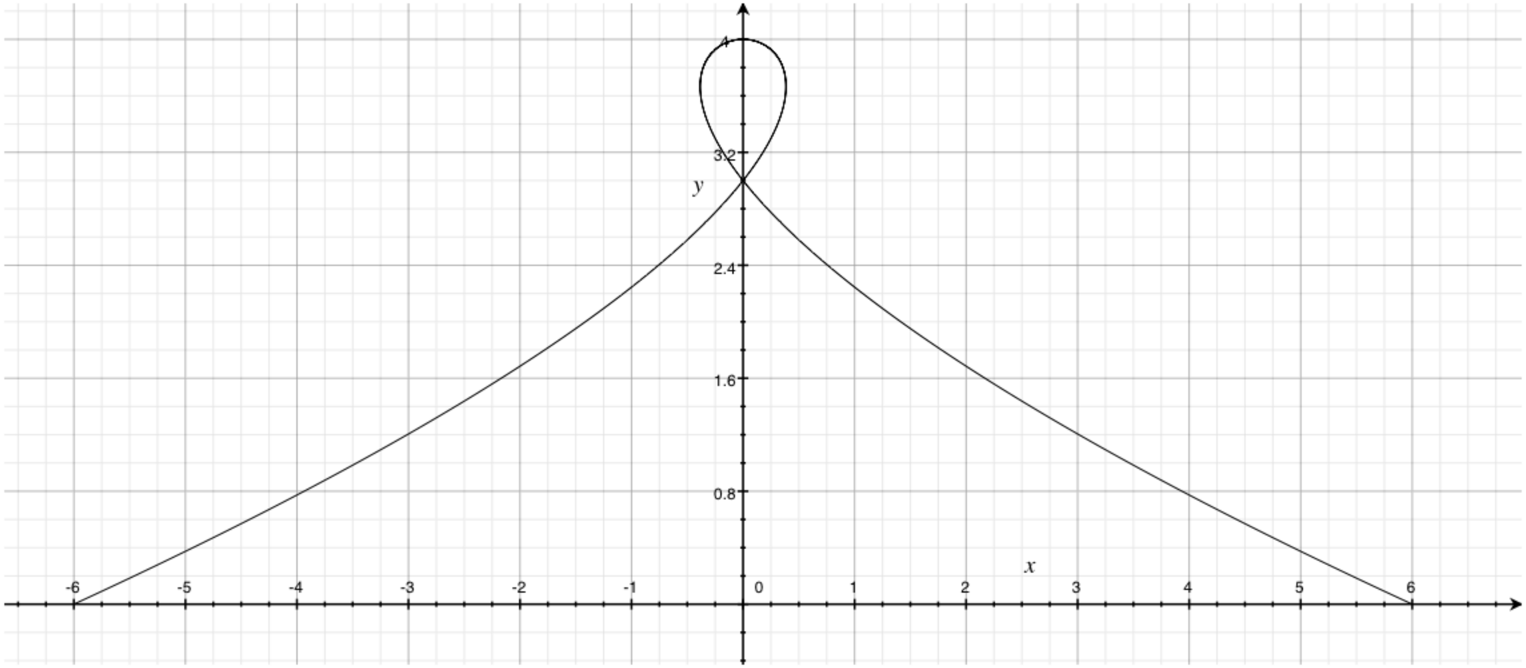
$x = r(t) = 2 + \frac{t}{5}\cos t$ and

$y = s(t) = 2 + \frac{t}{5}\sin t$

for $t \in [0,10]$.

Parameterised curves

Here I took
$x = r(t) = t^3 - t$ and
$y = s(t) = 4 - t^2$ for $t \in [-2,2]$.
This curve has **direction.**

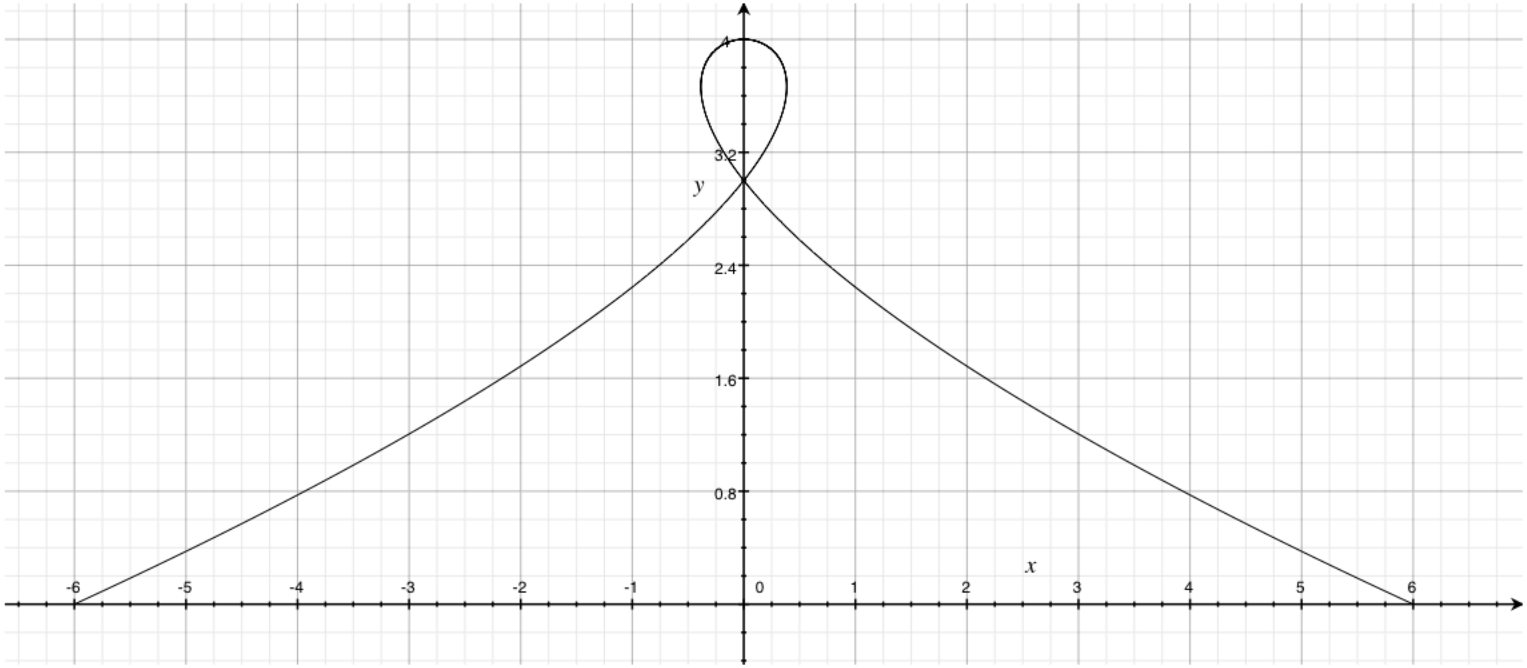


Parameterised curves

A different parameterization can give the same curve:

- In the opposite direction: $x = r(t) = -t^3 + t,\ y = s(t) = 4 - t^2$
- At a different 'speed' $x = r(t) = -8t^3 + 2t,\ y = s(t) = 4 - 4t^2$

When we come to differentiate such curves, the parameterization will matter!



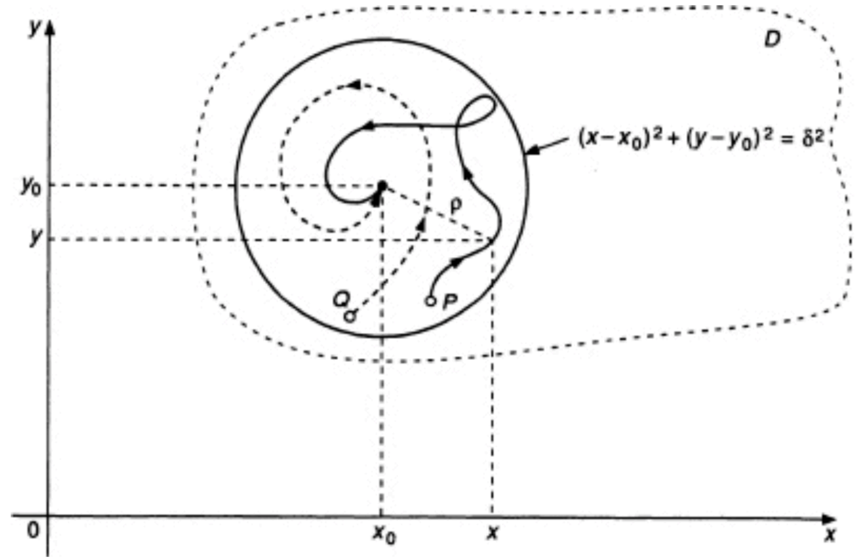
Continuity and limits

A curve $(x, y) = (r(t), s(t))$ has a limit (r, s) at a point $t = a$ if and only if $r(t)$ and $s(t)$ have limits r and s, respectively, at $t = a$.

A curve $(x, y) = (r(t), s(t))$ is continuous at a point $t = a$ if and only if $r(t)$ and $s(t)$ are continuous at $t = a$.

Limits of multivariate functions

Recall $\lim_{(x,y)\to(x_0,y_0)} f(x,y) = L$ exists if and only if L is independent of the path taken (provided $\rho \to 0$).



So if the path is defined by a curve $(x, y) = (x(t), y(t))$ then we are Interested in how f behaves as a function of t.

We can use the chain rule:

As t changes a little bit, x changes at a rate $\frac{dx}{dt}$ and y changes at a rate $\frac{dy}{dt}$.

For each change in x, f changes at a rate $\frac{\partial f}{\partial x}$, and for y, f changes at a rate $\frac{\partial f}{\partial y}$.

The combined change in f is $\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$.

Durham University

Image: *Mathematics for Engineers and Scientists, Alan Jeffrey*

Jacobians

Differentiability of multivariate functions

For $f(x_1, x_2, \dots, x_n)\colon \mathbb{R}^n \to \mathbb{R}$.

$$\frac{\partial f}{\partial x_i}(\boldsymbol{x}) = \lim_{h \to 0} \frac{f(\boldsymbol{x} + h.\boldsymbol{e_i}) - f(\boldsymbol{x})}{h}.$$

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}\right)^T$$

∇f is a vector pointing in the direction of the greatest rate of increase of f and having magnitude the rate of increase of f in that direction.

Reminder: for multivariate functions, partial derivatives does not imply continuity or differentiability.

f is **differentiable** at $\boldsymbol{x_0}$ if for all unit vectors $\boldsymbol{v}$ we have

$$\nabla_v f(\boldsymbol{x_0}) = \boldsymbol{v}.\nabla f$$

Theorem: f is **differentiable** at $\boldsymbol{x_0}$ if all its partial derivatives exist **and are continuous** at $\boldsymbol{x_0}$.

Example

Consider $f(x,y) = \begin{cases} 1 & xy = 0 \\ 0 & xy \neq 0 \end{cases}$

This has partial derivatives at (0,0), e.g.

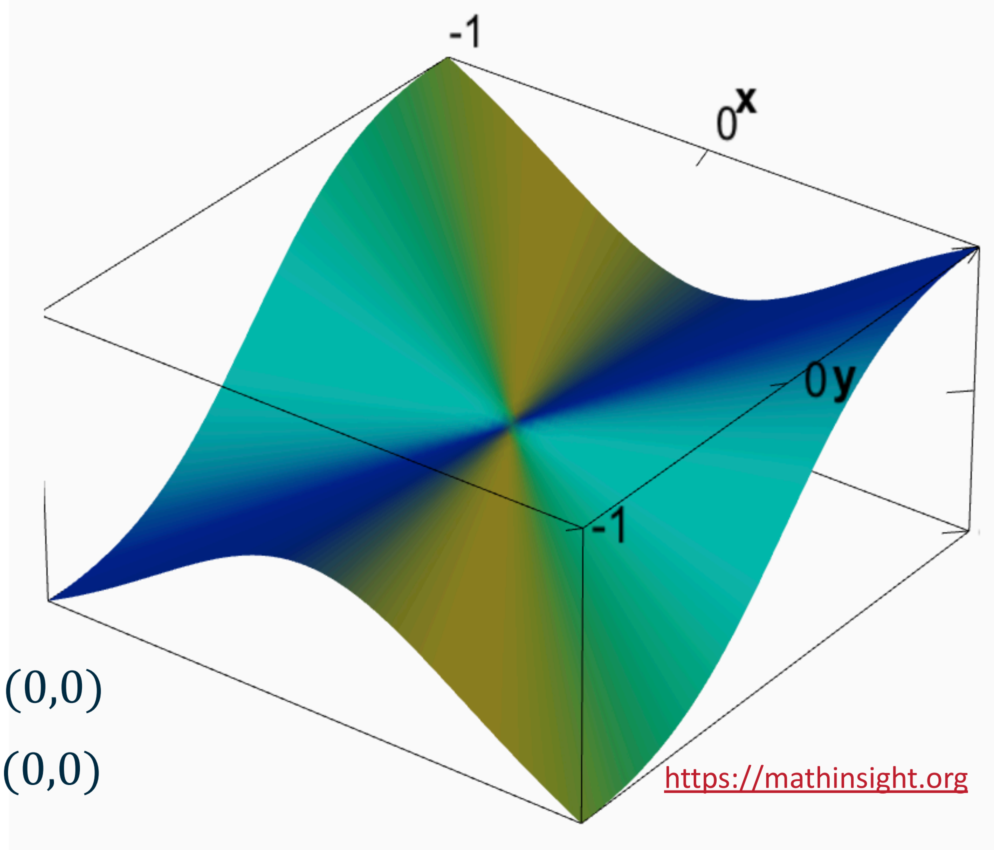
$$\frac{\partial f}{\partial y}(0,0) = \lim_{y \to 0} \frac{f(0,y) - f(0,0)}{y - 0} = \lim_{y \to 0} \frac{1 - 1}{y} = 0$$

But just away from $(0,0)$ the partial derivatives do not exist and therefore are not continuous at $(0,0)$.

For any $x > 0$:

$$\frac{\partial f}{\partial y}(x,0) = \lim_{y \to 0} \frac{f(x,y) - f(x,0)}{y - 0} = \lim_{y \to 0} \frac{0 - 1}{y} \to -\infty$$

Example



Consider $f(x,y) = \begin{cases} \frac{x^2y}{x^2+y^2} & (x,y) \neq (0{,}0) \\ 0 & (x,y) = (0{,}0) \end{cases}$

Now f is continuous at all points.

It has partial derivatives:

$$\frac{\partial f}{\partial x}(x,y) = \frac{2xy(x^2+y^2) - x^2y(2x)}{(x^2+y^2)^2} = \frac{2x^3y + 2xy^3 - 2x^3y}{(x^2+y^2)^2}$$

$$\frac{\partial f}{\partial y}(x,y) = \frac{x^2(x^2+y^2) - x^2y(2y)}{(x^2+y^2)^2} = \frac{x^4 + x^2y^2 - 2x^2y^2}{(x^2+y^2)^2}$$

Even at $(0{,}0)$ these exist and are both 0.

But $\frac{\partial f}{\partial y}(x,0)$ is 1 when $x \neq 0$, but 0 when $x = 0$, so not continuous.

Vector valued functions

So far we have considered a vector input and real valued output.

E.g. $f: \mathbb{R}^n \to \mathbb{R}$ given by $f(x_1, x_2, \ldots, x_n) = \sum_j \frac{x_j}{n}$ might give the average brightness of a digital camera sensor.

What if we instead wanted to adjust the brightness?

We could map each pixel i with a function $f_i(x_1, x_2, \ldots, x_n) = x_i - \sum_j \frac{x_j}{n} + 128$ to make the average brightness 128.

But rather than writing out separate functions for every pixel, we can think of the output as a vector $\left(f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_n(\boldsymbol{x})\right)$.

Now we have a vector valued function $\boldsymbol{f}: \mathbb{R}^n \to \mathbb{R}^n$.

The derivative of the i^{th} component of the output with respect to the j^{th} input is:

$$\frac{\partial f_i}{\partial x_j}.$$

Jacobian matrix

We have a function $\boldsymbol{f}: \mathbb{R}^n \to \mathbb{R}^m$ where $\boldsymbol{f} = (f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \dots, f_m(\boldsymbol{x}))$.

Set $\boldsymbol{J}_{ij} = \dfrac{\partial f_i}{\partial x_j}$.

The resulting **matrix of partial derivatives** is called the **Jacobian matrix**:

$$\mathbf{J} = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{pmatrix}$$

So the i^{th} row of the Jacobian matrix of $\boldsymbol{f}$ is the gradient ∇f_i (transposed).

And the j^{th} column is the derivative of $\boldsymbol{f}$ w.r.t. x_j: $\dfrac{\partial \boldsymbol{f}}{\partial x_j}$.

Durham University

Notation warning

Everyone agrees on most mathematical concepts – but they do not always write them the same way!

This often leads to confusion and you will just have to get used to it, and check what the notation is in a given text or book.

This is why even advanced books contain a glossary and summary of notation!

For a function $\boldsymbol{f}: \mathbb{R}^n \to \mathbb{R}^m$ the matrix of derivatives may be denoted:

$$\boldsymbol{J}, \boldsymbol{J}_{\boldsymbol{f}}, D\boldsymbol{f}, \nabla\boldsymbol{f}$$

Or even simply $\frac{\partial(f_1,f_2,\ldots,f_m)}{\partial(x_1,x_2,\ldots,x_n)}$.

Occasionally the Jacobian is defined as the transpose of this matrix.

Example

$$\boldsymbol{f}: \mathbb{R}^{\mathbf{3}} \to \mathbb{R}^{\mathbf{2}} = (2x_1 + x_2 - x_3, x_1 + 3x_2 - 4x_3)$$

$$\boldsymbol{g}: \mathbb{R}^{\mathbf{2}} \to \mathbb{R}^{\mathbf{2}} = (\sin(x_1 + x_2), \cos(x_1 - x_2))$$

$$\boldsymbol{J_f} = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 3 & -4 \end{pmatrix} \qquad \boldsymbol{J_g} = \begin{pmatrix} \cos(x_1 + x_2) & \cos(x_1 + x_2) \\ -\sin(x_1 - x_2) & \sin(x_1 - x_2) \end{pmatrix}$$

$$\boldsymbol{J_g}\left(\frac{3\pi}{4}, \frac{\pi}{4}\right) = \begin{pmatrix} \cos(\pi) & \cos(\pi) \\ -\sin(\pi/2) & \sin(\pi/2) \end{pmatrix} = \begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix}$$

What about $\boldsymbol{g} \circ \boldsymbol{f}: \mathbb{R}^{\mathbf{3}} \to \mathbb{R}^{\mathbf{2}}$?

$$\boldsymbol{g} \circ \boldsymbol{f} = (\sin(3x_1 + 4x_2 - 5x_3), \cos(x_1 - 2x_2 + 3x_3))$$

Durham University

Example

What about $\boldsymbol{g} \circ \boldsymbol{f} \colon \mathbb{R}^3 \to \mathbb{R}^2$?

$$\boldsymbol{g} \circ \boldsymbol{f} = (\sin(3x_1 + 4x_2 - 5x_3), \cos(x_1 - 2x_2 + 3x_3))$$

$$\boldsymbol{J}_{\boldsymbol{g}\circ\boldsymbol{f}} = \begin{pmatrix} 3\cos(3x_1 + 4x_2 - 5x_3) & 4\cos(3x_1 + 4x_2 - 5x_3) & -5\cos(3x_1 + 4x_2 - 5x_3) \\ -\sin(x_1 - 2x_2 + 3x_3) & 2\sin(x_1 - 2x_2 + 3x_3) & -3\sin(x_1 - 2x_2 + 3x_3) \end{pmatrix}$$

$$\boldsymbol{J}_{\boldsymbol{g}\circ\boldsymbol{f}}\left(\frac{8\pi}{20}, -\frac{\pi}{20}, 0\right) = \begin{pmatrix} 3\cos(\pi) & 4\cos(\pi) & -5\cos(\pi) \\ -\sin(\pi/2) & 2\sin(\pi/2) & -3\sin(\pi/2) \end{pmatrix} = \begin{pmatrix} -3 & -4 & 5 \\ -1 & 2 & -3 \end{pmatrix}$$

But notice: $\begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix} . \begin{pmatrix} 2 & 1 & -1 \\ 1 & 3 & -4 \end{pmatrix} = \begin{pmatrix} -3 & -4 & 5 \\ -1 & 2 & -3 \end{pmatrix}$

What is this magic?!

Example

It is the chain rule!

$$\frac{\partial g \circ f_i}{\partial x_j} = \frac{\partial g_i}{\partial f_1}\frac{\partial f_1}{\partial x_j} + \frac{\partial g_i}{\partial f_2}\frac{\partial f_2}{\partial x_j} = \boldsymbol{\nabla}_{\boldsymbol{g}_{\boldsymbol{i}}} \frac{\partial \boldsymbol{f}}{\partial x_j}$$

So

$$\boldsymbol{J}_{\boldsymbol{g}\circ\boldsymbol{f}} = \boldsymbol{J}_{\boldsymbol{g}}\,\boldsymbol{J}_{\boldsymbol{f}}$$

Linear approximations

For a univariate function the derivative gives details of the tangent line:

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

So

$$f(x+h) \approx f(x) + hf'(x)$$

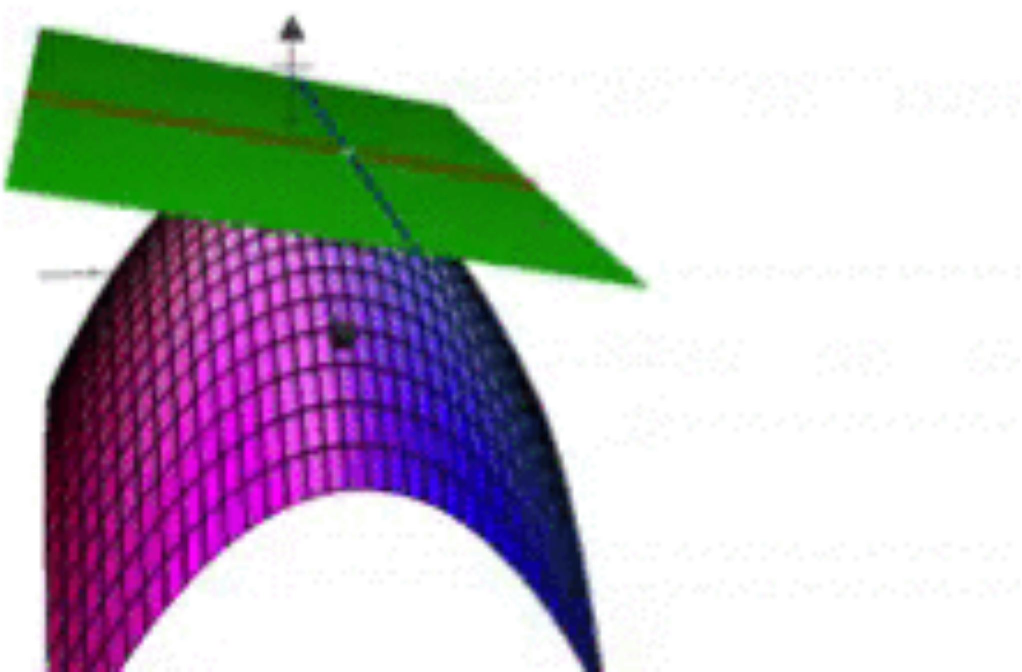
Linear approximations

For a surface we have seen that

$$\nabla_v f(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0 + h.v_x, y_0 + h.v_y) - f(x_0, y_0)}{h}$$

So

$$f(x_0 + h.v_x, y_0 + h.v_y) = f(x_0, y_0) + h\nabla_v f(x_0, y_0)$$



Durham University

Linear approximations

For a multivariate function : $\mathbb{R}^{\boldsymbol{n}} \to \mathbb{R}^{\boldsymbol{m}}$ the Jacobian maxtrix gives the best linear approximation to $\boldsymbol{f}$ at $\boldsymbol{x}$:

$$\boldsymbol{f}(\boldsymbol{x} + h\boldsymbol{v}) = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{x})(h\boldsymbol{v})$$

Note: $\boldsymbol{x}, \boldsymbol{v}, h\boldsymbol{v}$ are n-dimensional vectors,

$\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{x})$ is an m by n matrix,

$\boldsymbol{f}(\boldsymbol{x}), \boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{x})(h\boldsymbol{v})$ are m-dimensional vectors.