

For the next few weeks, we focus on differentiation of multivariate functions (in a different order from the textbook):

$\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  (i.e. **vector-valued** functions)

- This week: differentiating a multivariate function (§12.2, 12.3 first two pages, 12.4 first two pages, 12.6 first two pages, fifth and sixth pages)
- Week 8: the chain rule, for differentiating compositions (§12.5, and the matrix version in §12.6)

$f : \mathbb{R}^n \rightarrow \mathbb{R}$  (i.e. **scalar-valued** functions)

- Week 9: direction of greatest increase, tangent planes, Taylor polynomials (§12.7, 12.9 first four pages)
- Week 10: classifying critical points (§13.1, the subsection “Classifying Critical Points” until Example 7; the rest is in Week 11)
- Week 11: finding maxima and minima (§13.1-13.5 8E, §13.1-13.4 7E)

Notation: we call the  $m$  “outputs” of  $\mathbf{f}$  by  $f_1, f_2, \dots, f_m$ , these are the **coordinate functions**. e.g.  $\mathbf{f} : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  is denoted  $\mathbf{f}(x, y) = (f_1(x, y), f_2(x, y), f_3(x, y))$ . We will often analyze  $\mathbf{f}$  by analysing its coordinate functions separately.

For one-variable functions, the derivative is the limit of a difference quotient:

$$\frac{df}{dx} = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

To discuss the differentiability of multivariate functions, we must first define the limit of a multivariate function  $\lim_{(x,y) \rightarrow (a,b)} g(x,y)$ .

Unlike the limits of 2-variable Riemann sums that we saw in multiple integration, the limit of a 2-variable function **cannot** be calculated by taking 1D limits separately in the  $x$  and  $y$  directions. It requires a more careful analysis.

On the next page we give an informal definition of a limit for functions  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , but we will concentrate on when the domain is  $\mathbb{R}^2$  and the codomain is  $\mathbb{R}$ .

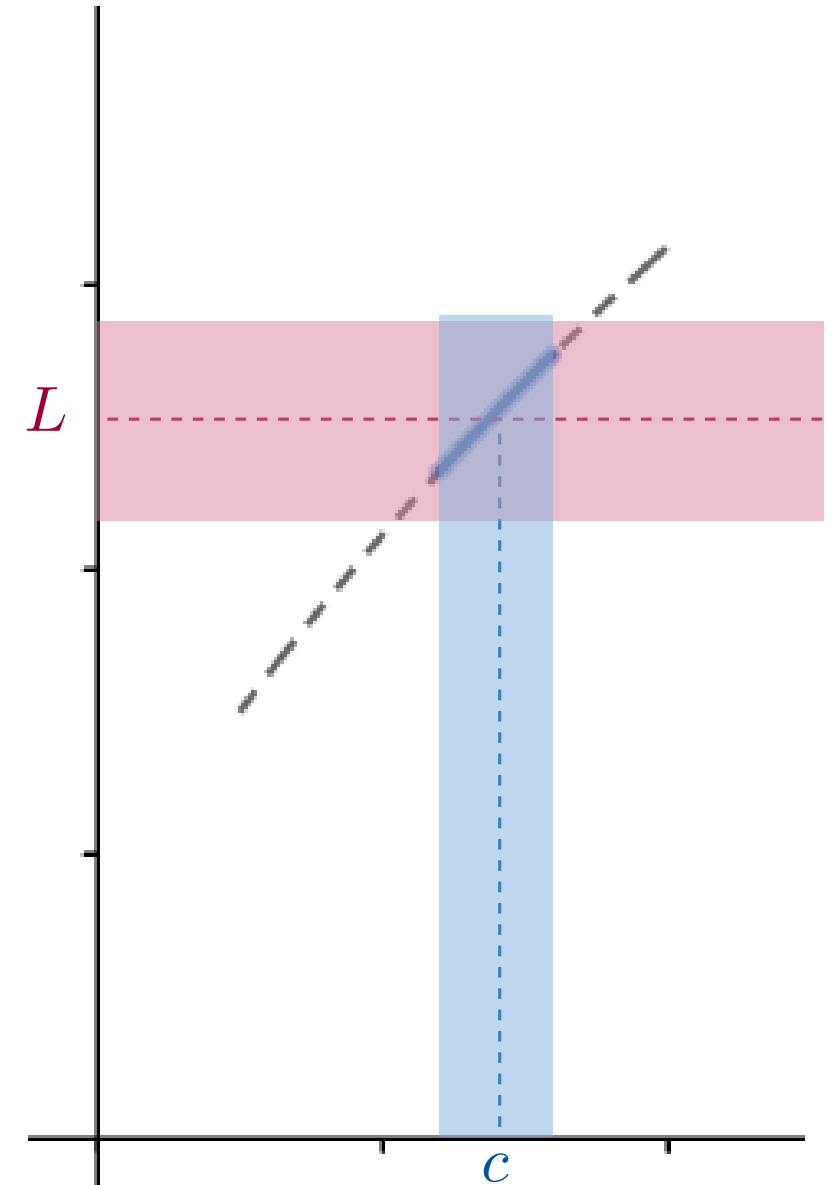
We will first discuss ways to show that a limit does not exist (p5-9), and then ways to evaluate a limit that does exist (p10-13).

## §12.2: Limits and Continuity

Remember the informal definition of a single-variable limit:

**Definition:** Given a function  $f(x)$  defined near a point  $c$ , the statement  $\lim_{x \rightarrow c} f(x) = L$  means: we can ensure that  $f(x)$  is as close as we want to  $L$  by taking  $x$  close enough (but not equal) to  $c$ .

In other words: given any small interval around  $L$  (the height of the red rectangle), we can find a small interval around  $c$  (the width of the blue rectangle) so the values of  $f(x)$  when  $x$  is in this small “blue” interval all lie in the “red” interval around  $L$  (i.e. the part of the graph of  $f$  in the blue rectangle is also in the red rectangle).

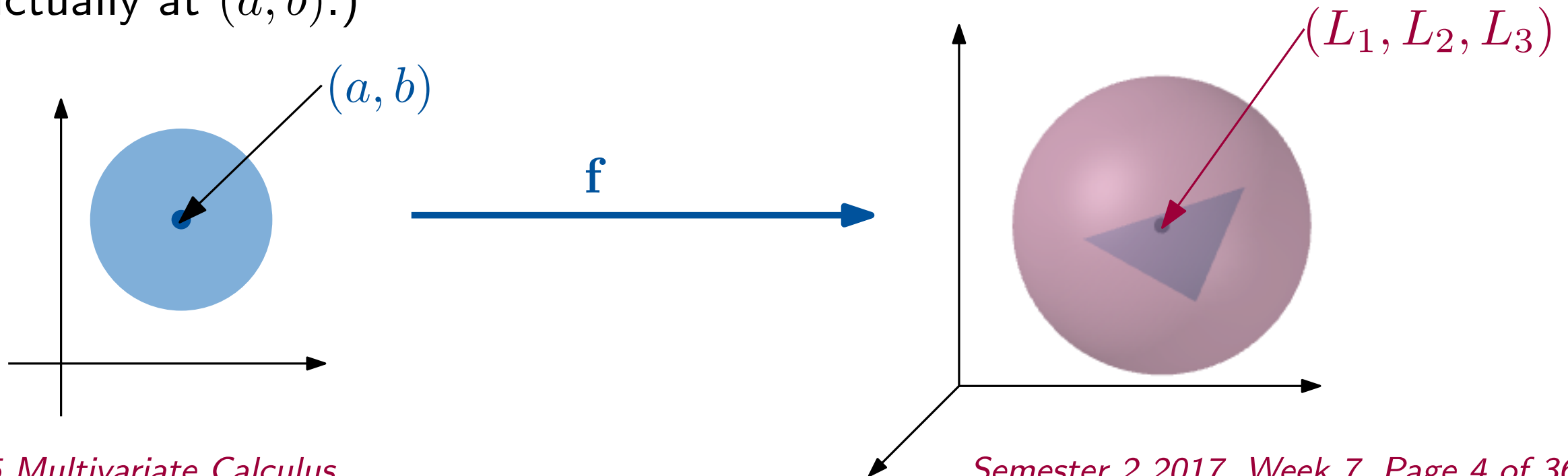


A limit of a multivariate function is the same idea, using balls and spheres instead of intervals:

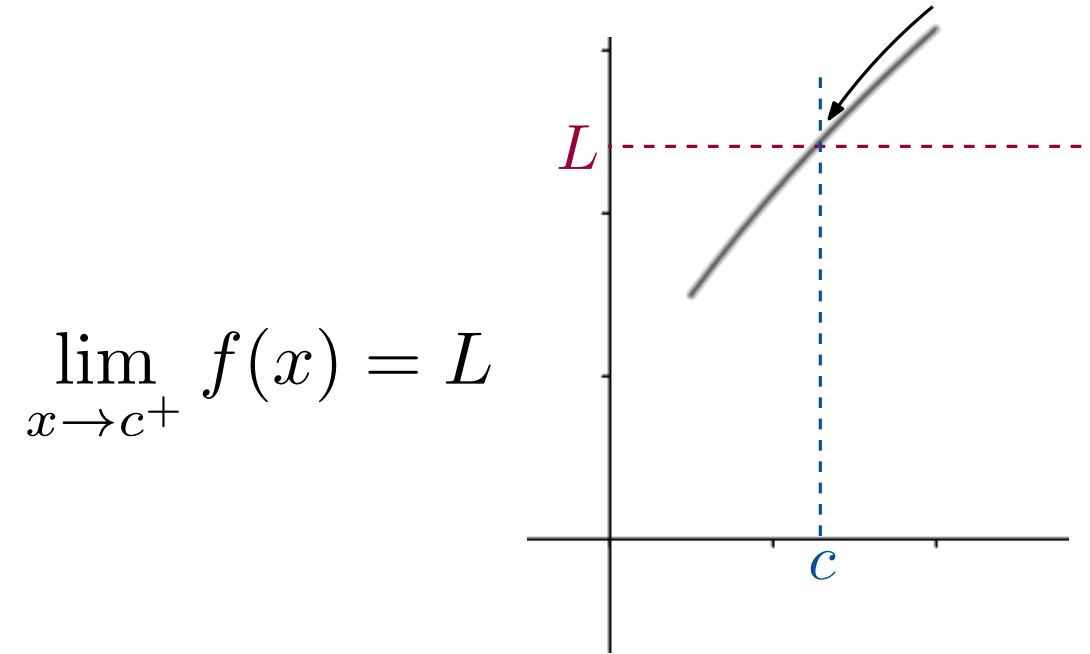
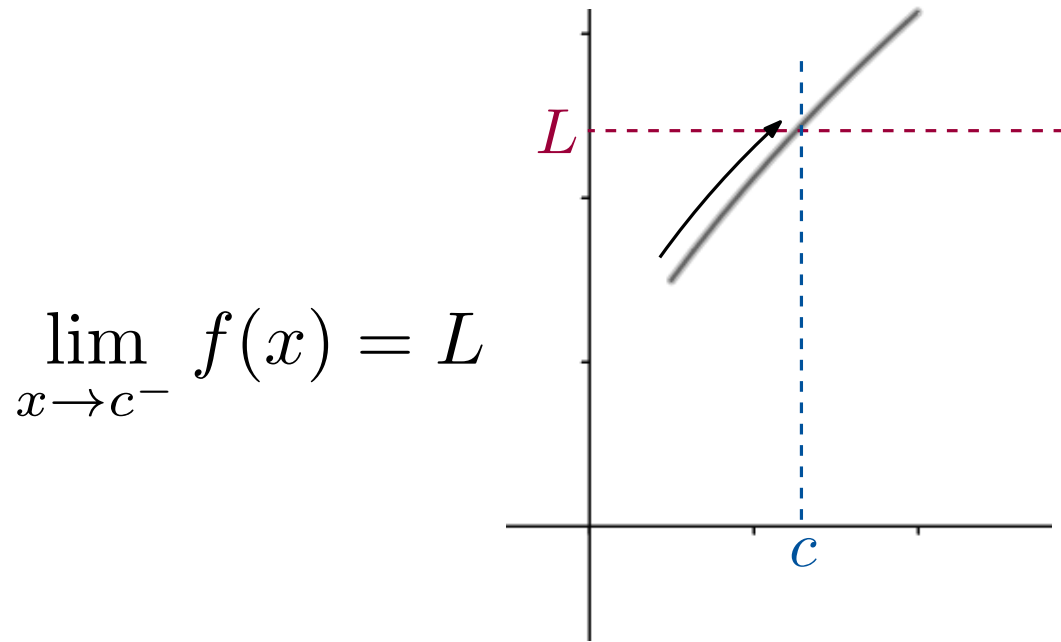
**Example:** Given a 2-variable function  $\mathbf{f} : \mathcal{D} \rightarrow \mathbb{R}^3$  whose domain  $\mathcal{D}$  contains points close to  $(a, b)$ , the statement  $\lim_{(x,y) \rightarrow (a,b)} \mathbf{f}(x, y) = (L_1, L_2, L_3)$  means:

given a small sphere around  $(L_1, L_2, L_3)$ , we can find a small disk around  $(a, b)$  such that the image of this disk under  $\mathbf{f}$  is entirely contained in the sphere.

(Strictly speaking,  $\mathbf{f}(a, b)$  does not need to be in the sphere for the limit statement to hold, because a limit is about how a function behaves **around**  $(a, b)$  but not actually at  $(a, b)$ .)



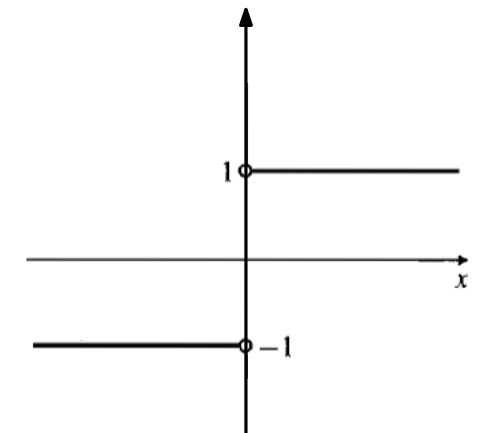
Another informal way to think about single-variable limits is: how does  $f(x)$  behave as  $x$  moves towards  $c$ ? This is the **one-sided limit**:



It is a theorem that  $\lim_{x \rightarrow c} f(x)$  exists if and only if the two one-sided limits exist and are equal - i.e.  $f(x)$  “goes towards” the same number no matter how we move towards  $c$ .

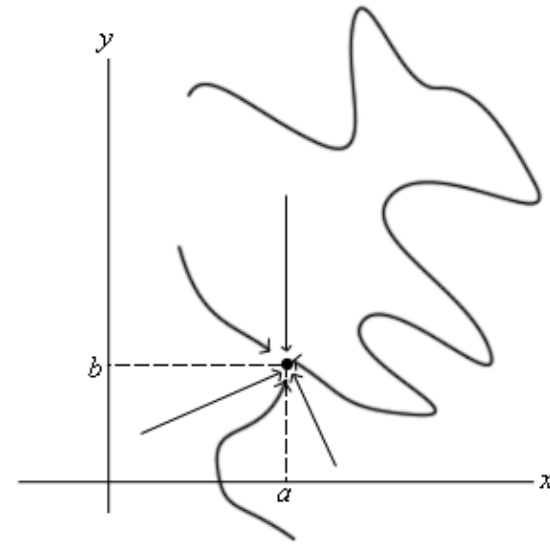
**Example:**  $\lim_{x \rightarrow 0} \frac{x}{|x|}$  does not exist because  $\lim_{x \rightarrow 0^-} \frac{x}{|x|} = -1$ ,

$\lim_{x \rightarrow 0^+} \frac{x}{|x|} = 1$ , and these limits are not equal.



The same is true for multivariate limits, but there are now many more ways for  $(x, y)$  to approach  $(a, b)$ .

Each way of approach can be formalised as a **path**, i.e. a function  $t \mapsto (x(t), y(t))$  such that  $x(c) = a, y(c) = b$ . (Imagine drawing one of the paths in the diagram, and recording the position of your pen at time  $t$ . Write  $c$  for the time that your pen reaches the point of interest  $(a, b)$ .) We then study  $f$  by considering the values that  $f$  takes along the path, i.e. by considering the composition  $f(x(t), y(t))$ .



**Theorem: Multivariate Limits and Paths:** Let  $f : \mathcal{D} \rightarrow \mathbb{R}$  be a two-variable function, and suppose  $\mathcal{D}$  contains points arbitrarily close to  $(a, b)$ . We have

$\lim_{(x,y) \rightarrow (a,b)} f(x, y) = L$  if and only if, for **all paths**  $t \mapsto (x(t), y(t))$  such that  $x(c) = a, y(c) = b$ , the limits  $\lim_{t \rightarrow c} f(x(t), y(t))$  **all exist and are equal to  $L$** .

Because the existence of the 2D limit is equivalent to the existence of 1D limits along **infinitely many** paths, it is not practical to use this theorem to prove the existence of a 2D limit. However, the theorem is useful for showing a 2D limit

**HK, doesn't exist:** simply find two paths along which the limits are different.

**Example:** Show that the limit  $\lim_{(x,y) \rightarrow (-1,2)} \frac{x^2 - 1}{4x^2 - y^2}$  does not exist.

**Example:** Show that the limit  $\lim_{(x,y) \rightarrow (0,0)} \frac{x^2 y}{x^3 + y^3}$  does not exist.



**Example:** Show that the limit  $\lim_{(x,y) \rightarrow (0,0)} \frac{x^4}{x^2y + y^2}$  does not exist.

Now we give some strategies for showing that a 2-variable limit does exist. This will use the concept of continuity, which has the same definition as the 1D case.

**Definition:** An  $n$ -variable function  $\mathbf{f} : \mathcal{D} \rightarrow \mathbb{R}^m$  is *continuous* at a point  $(a_1, \dots, a_n)$  in the domain  $\mathcal{D}$  if

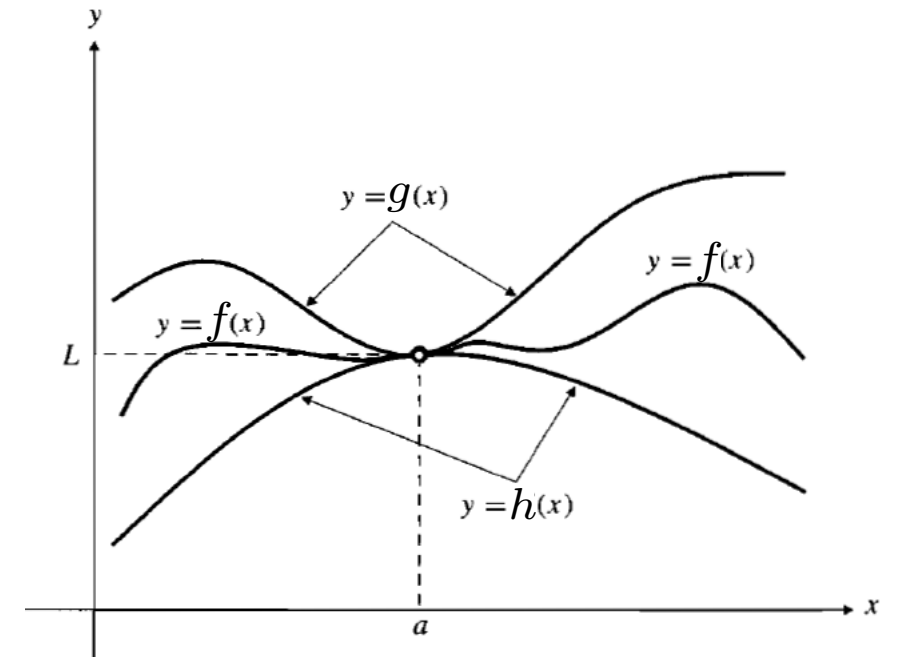
$$\lim_{(x_1, \dots, x_n) \rightarrow (a_1, \dots, a_n)} \mathbf{f}(x_1, \dots, x_n) = \mathbf{f}(a_1, \dots, a_n).$$

As in the 1D case, elementary functions (i.e. sums, products and compositions of  $x^n$ ,  $e^x$ ,  $\ln x$ ,  $\sin x$ ,  $\cos x$ ) are continuous. So the following example is easy:

**Example:** Evaluate the limit  $\lim_{(x,y) \rightarrow (-1,2)} \frac{x^2 - 2}{y^2 - 1}$ , or prove that it does not exist.

In more complicated examples, our main tool for evaluating limits is the squeeze theorem. The multivariate squeeze theorem is a very simple extension of the 1D statement. (The diagram is in 1D, but we can easily imagine a 2D version.)

**Squeeze Theorem:** Suppose there are functions  $g(x, y)$  and  $h(x, y)$  such that, for all points  $(x, y)$  in the domain of  $f$  that are near  $(a, b)$ , we have the inequality  $h(x, y) \leq f(x, y) \leq g(x, y)$ . Suppose also that  $\lim_{(x,y) \rightarrow (a,b)} g(x, y) = \lim_{(x,y) \rightarrow (a,b)} h(x, y) = L$ . Then  $\lim_{(x,y) \rightarrow (a,b)} f(x, y)$  exists and is also equal to  $L$ .



We will often choose our squeezing functions  $g(x, y)$  and  $h(x, y)$  to be elementary functions, so their limits are easy to calculate.

In the special case where we want to show  $\lim_{(x,y) \rightarrow (a,b)} f(x, y) = 0$ , it is enough to find one squeezing function  $g(x, y)$  such that  $\lim_{(x,y) \rightarrow (a,b)} g(x, y) = 0$  and

**HKB**  $-g(x, y) \leq f(x, y) \leq g(x, y)$  - this inequality is equivalent to  $|f(x, y)| \leq g(x, y)$ .

Here's a simple 2-dimensional version of the standard 1D squeeze theorem example (see Homework 3 final question).

**Example:** Evaluate  $\lim_{(x,y) \rightarrow (0,0)} xy^2 \sin\left(\frac{1}{y}\right)$ , or prove that the limit does not exist.

Most applications of the Squeeze Theorem in 2D don't involve trigonometric functions, and look more like this next example:

**Example:** Evaluate the limit  $\lim_{(x,y) \rightarrow (0,0)} \frac{yx^2}{x^2 + y^2}$ , or prove that it does not exist.

Summary on analysing the limit  $\lim_{(x,y) \rightarrow (a,b)} f(x,y)$ :

**Step 1** If  $f$  is a continuous function that is defined at  $(a,b)$ , then the limit is  $f(a,b)$  (see p10).

**Step 2** Evaluate the limit of  $f$  along straight line paths:  $\lim_{x \rightarrow a} f(x,b)$ ,  $\lim_{y \rightarrow b} f(a,y)$ ,  $\lim_{x \rightarrow 0} f(x, mx)$  (if  $(a,b) = (0,0)$ ). If you find two different limits, then  $f$  does not have a limit at  $(a,b)$  (see p8).

**Step 3** Evaluate the limit of  $f$  along paths of the form  $y = x^n$  or  $(x,y) = (t^i, t^j)$  (if  $(a,b) = (0,0)$ ). If these give different limits from the paths in Step 2, then  $f$  does not have a limit at  $(a,b)$  (see p9).

**Step 4** Try to prove that  $\lim_{(x,y) \rightarrow (a,b)} f(x,y)$  is the value of the limits found in Steps 2 and 3, by using the Squeeze Theorem (see p13).

(Note that the squeeze theorem only applies to scalar-valued functions, because there is no concept of  $<$  or  $>$  in  $\mathbb{R}^m$  for  $m > 1$ . But we can apply the squeeze theorem to the coordinate functions of a vector-valued function -  $\mathbf{f}$  has a limit if and only if each coordinate function  $f_i$  has a limit.)

## §12.3-4: Partial Derivatives

Remember that the derivative of a single-variable function is the limit of a difference quotient, measuring the rate of change of  $f$  as we change the input variable  $x$ :

$$f'(x) = \frac{df}{dx} = \frac{d}{dx} f(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

## §12.3-4: Partial Derivatives

Remember that the derivative of a single-variable function is the limit of a difference quotient, measuring the rate of change of  $f$  as we change the input variable  $x$ :

$$f'(x) = \frac{df}{dx} = \frac{d}{dx} f(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

The derivative of a 2-variable function will be a more complicated concept because there are many different ways to change the input variables  $(x, y)$ . We start with the simplest way, where we fix one variable and change the other:

**Definition:** The *first-order partial derivatives* of the function  $f(x, y)$ , with respect to the variables  $x$  and  $y$  respectively, are given by:

$$f_x(x, y) = \frac{\partial f}{\partial x} = \frac{\partial}{\partial x} f(x, y) = \lim_{h \rightarrow 0} \frac{f(x+h, y) - f(x, y)}{h}$$

$$f_y(x, y) = \frac{\partial f}{\partial y} = \frac{\partial}{\partial y} f(x, y) = \lim_{k \rightarrow 0} \frac{f(x, y+k) - f(x, y)}{k}$$

These are **1D limits**, not the 2D limits of the previous section.



**Definition:** The *first-order partial derivatives* of the function  $f(x, y)$ , with respect to the variables  $x$  and  $y$  respectively, are given by:

$$f_x(x, y) = \frac{\partial f}{\partial x} = \frac{\partial}{\partial x} f(x, y) = \lim_{h \rightarrow 0} \frac{f(x + h, y) - f(x, y)}{h}$$

$$f_y(x, y) = \frac{\partial f}{\partial y} = \frac{\partial}{\partial y} f(x, y) = \lim_{k \rightarrow 0} \frac{f(x, y + k) - f(x, y)}{k}$$

It is clear how to define first-order partial derivatives for an  $n$ -variable function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  - there will be  $n$  of them, one for each variable (which will change that variable and keep the other  $n - 1$  variables fixed).

And for a vector-valued function  $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , we can take partial derivatives of each coordinate function, so there will be  $nm$  partial derivatives altogether (see p32).

(The textbook writes  $f_1(x, y)$  to mean “differentiate with respect to the first variable”, i.e. what we are calling  $f_x(x, y)$ . But this causes problems when  $\mathbf{f}$  is vector-valued, because  $f_i$  is also the  $i$ th coordinate function of  $\mathbf{f}$ .)

If  $f$  is an elementary function, then we can use our single-variable differentiation rules to calculate  $\frac{\partial f}{\partial x}$ , by treating  $y$  as a constant (and similarly for  $\frac{\partial f}{\partial y}$ ).

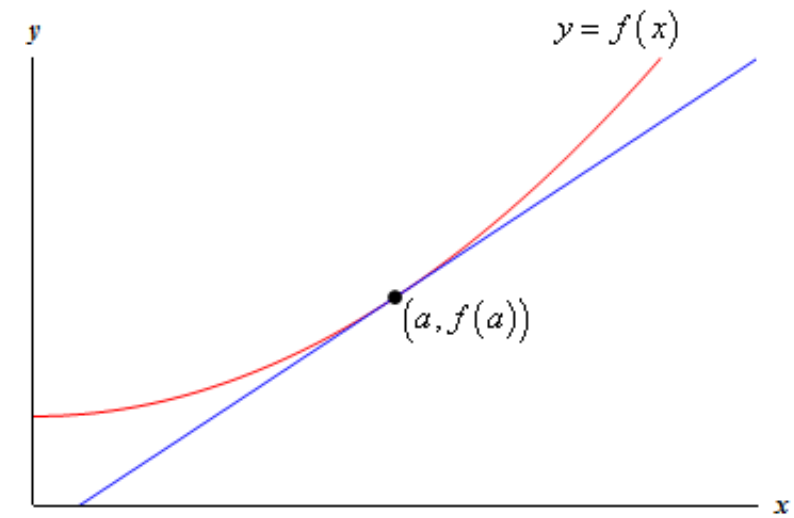
**Example:** Find the first-order partial derivatives of  $f(x, y) = \frac{xy}{x+1}$  at  $(1, 2)$ .

When  $f$  is defined by different formulae around  $(a, b)$ , we need to use the limit definition to calculate the partial derivatives at  $(a, b)$ .

**Example:** Let  $f(x, y) = \begin{cases} \frac{y^3}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}$ . Find  $f_y(0, 0)$ .

## The geometric meaning of partial derivatives

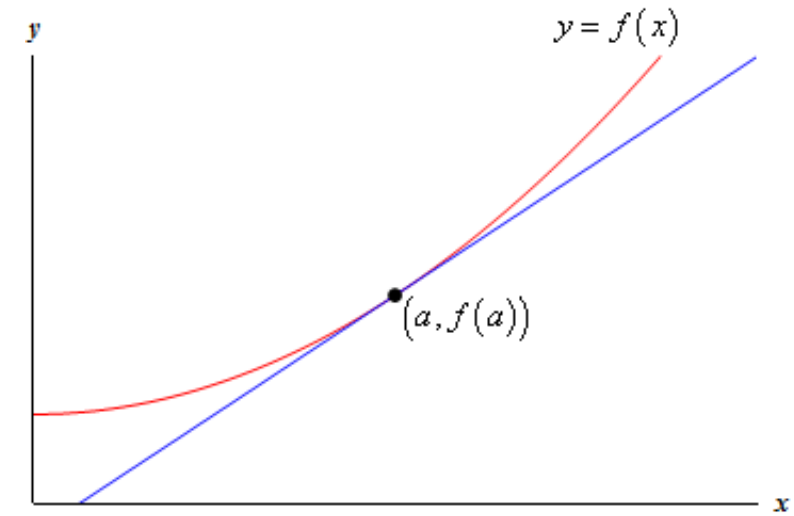
Recall that the derivative  $f'(a)$  of a single-variable function  $f$  is the slope of the tangent line to the graph of  $f$  at the point  $(a, f(a))$ .



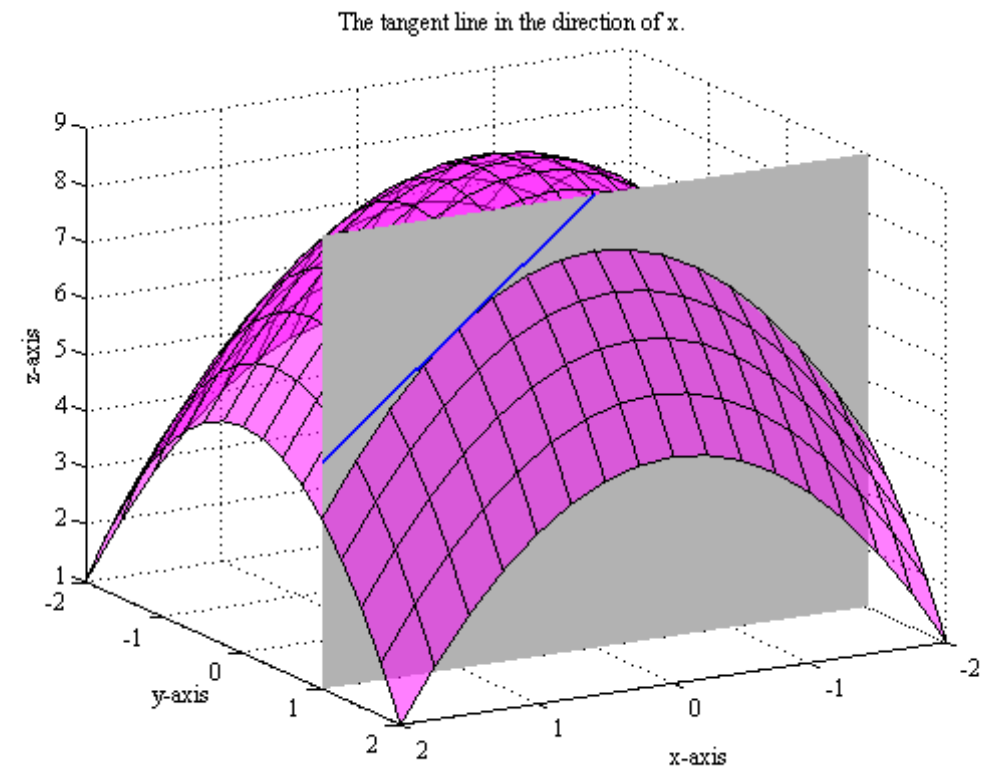
(pictures from Paul's Online Math Notes and WikiHow)

## The geometric meaning of partial derivatives

Recall that the derivative  $f'(a)$  of a single-variable function  $f$  is the slope of the tangent line to the graph of  $f$  at the point  $(a, f(a))$ .



The partial derivatives  $f_x(a, b)$  and  $f_y(a, b)$  are also the slopes of tangent lines to certain curves: these curves are the graphs of  $f(x, b)$  and  $f(a, y)$ , which are the intersections of the graph of  $f$  with the planes  $y = b$  and  $x = a$  respectively.



(pictures from Paul's Online Math Notes and WikiHow)

## Higher order partial derivatives

Remember that the second derivative of a single-variable function comes from taking the derivative twice:  $f''(x) = \frac{d^2 f}{dx^2} = \frac{d^2}{dx^2} f(x) = \frac{d}{dx} \left( \frac{df}{dx} \right)$ .

## Higher order partial derivatives

Remember that the second derivative of a single-variable function comes from taking the derivative twice:  $f''(x) = \frac{d^2 f}{dx^2} = \frac{d^2}{dx^2} f(x) = \frac{d}{dx} \left( \frac{df}{dx} \right)$ .

Similarly, we can take the derivative twice of a 2-variable function  $f(x, y)$  - now there are many combinations depending on which variable we differentiate with respect to, as they can be different variables for the first time and the second time:

**Definition:** The *second-order partial derivatives* of the function  $f(x, y)$  are:

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

It is clear how to define third-order partial derivatives, by repeated partial differentiation.

**Example:** Find the second-order partial derivatives of  $f(x, y) = \frac{xy}{x+1}$  at  $(1, 2)$ .



In the previous example, we found that  $\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y}$ . This is a general fact for well-behaved functions:

**Theorem 1: Equality of Mixed Partial Derivatives:** (Clairaut's Theorem)

Suppose  $p, q$  are two  $k$ th-order partial derivatives of  $f$  obtained by differentiating with respect to the same set of  $k$  variables (possibly with repetition) but in different orders. Suppose also that  $p, q$  are continuous at  $(a, b)$ , and all  $k - 1$ th-order partial derivatives of  $f$  are continuous around  $(a, b)$ . Then  $p(a, b) = q(a, b)$ .

**Example:** Clairaut's Theorem says that, if  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  has third-order partial derivatives that are continuous everywhere, then there are only four different third-order partial derivatives:

$$f_{xxx};$$

$$f_{xxxy} = f_{xyxx} = f_{yxxx};$$

$$f_{xyyy} = f_{yyxy} = f_{yyyx};$$

$$f_{yyyy}.$$

The proof of Clairaut's Theorem uses the 1D mean value theorem separately in the  $x$  and  $y$  directions - see the textbook.

## §12.6: Linear Approximation and Differentiability

Remember that a single-variable function  $f$  is said to be differentiable at a point  $a$  if the derivative  $f'(a)$  exists.

It is not a good idea to say that a multivariate function is differentiable if all its partial derivatives exist, since there are discontinuous functions that have partial derivatives. One example is  $f(x, y) = \begin{cases} \frac{2xy}{x^2+y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0). \end{cases}$  On ex sheet

#13 Q1, you showed that  $f$  does not have a limit at  $(0, 0)$ , so  $f$  is not continuous, but because  $f$  is always 0 on the  $x$  and  $y$  axes, its partial derivatives do exist and are 0. (You will prove this carefully on Homework 4.)

The existence of partial derivatives means that  $f$  is well-behaved in the  $x$  and  $y$  directions, but  $f$  can still be horrible in other directions. A good definition of differentiability at  $(a, b)$  must be a statement about all points around  $(a, b)$ .

We will say that  $f$  is differentiable if it is **locally well-approximated by a linear function**.

Let's first understand what this means for a single-variable function  $f : \mathbb{R} \rightarrow \mathbb{R}$ . Remember that the **linearisation** of  $f$  at  $a$  is

$$L(x) = f(a) + f'(a)(x - a).$$

This linear function is important because we can use it to approximate  $f(x)$  when  $x$  is near  $a$ , i.e. when  $x = a + h$  and  $h$  is small. This approximation is good because the error satisfies

$$\begin{aligned} f(x) - L(x) &= f(x) - f(a) - f'(a)(x - a) \\ f(a + h) - L(a + h) &= f(a + h) - f(a) - f'(a)h \\ \frac{f(a + h) - L(a + h)}{h} &= \frac{f(a + h) - f(a)}{h} - f'(a) \\ \lim_{h \rightarrow 0} \frac{f(a + h) - L(a + h)}{h} &= \lim_{h \rightarrow 0} \frac{f(a + h) - f(a)}{h} - f'(a) = 0 \end{aligned}$$

In other words, the error is small compared to  $h$ , the distance from  $a$  to  $x$ .

Remember again that the linearisation of  $f : \mathbb{R} \rightarrow \mathbb{R}$  at  $a$  is  $L(x) = f(a) + f'(a)(x - a)$ .  
A straightforward multivariate generalisation is:

**Definition:** The *linearisation* of a function  $f(x, y)$  at  $(a, b)$  is

$$L(x, y) = f(a, b) + \underset{\substack{\text{rate of change of } f \\ \text{with respect to } x}}{\underset{\uparrow}{f_x(a, b)}} \underset{\substack{\text{change} \\ \text{in } x}}{\underset{\uparrow}{(x - a)}} + \underset{\substack{\text{rate of change of } f \\ \text{with respect to } y}}{\underset{\uparrow}{f_y(a, b)}} \underset{\substack{\text{change} \\ \text{in } y}}{\underset{\uparrow}{(y - b)}}.$$

Remember again that the linearisation of  $f : \mathbb{R} \rightarrow \mathbb{R}$  at  $a$  is  $L(x) = f(a) + f'(a)(x - a)$ .  
A straightforward multivariate generalisation is:

**Definition:** The *linearisation* of a function  $f(x, y)$  at  $(a, b)$  is

$$L(x, y) = f(a, b) + \underbrace{f_x(a, b)}_{\substack{\text{rate of change of } f \\ \text{with respect to } x}} \underbrace{(x - a)}_{\substack{\text{change} \\ \text{in } x}} + \underbrace{f_y(a, b)}_{\substack{\text{rate of change of } f \\ \text{with respect to } y}} \underbrace{(y - b)}_{\substack{\text{change} \\ \text{in } y}}.$$

And then  $f$  is differentiable if the error from using  $L(x, y)$  to approximate  $f(x, y)$  is small compared to the distance from  $(x, y)$  to  $(a, b)$ :

**Definition:** A function  $f(x, y)$  is *differentiable* at  $(a, b)$  if

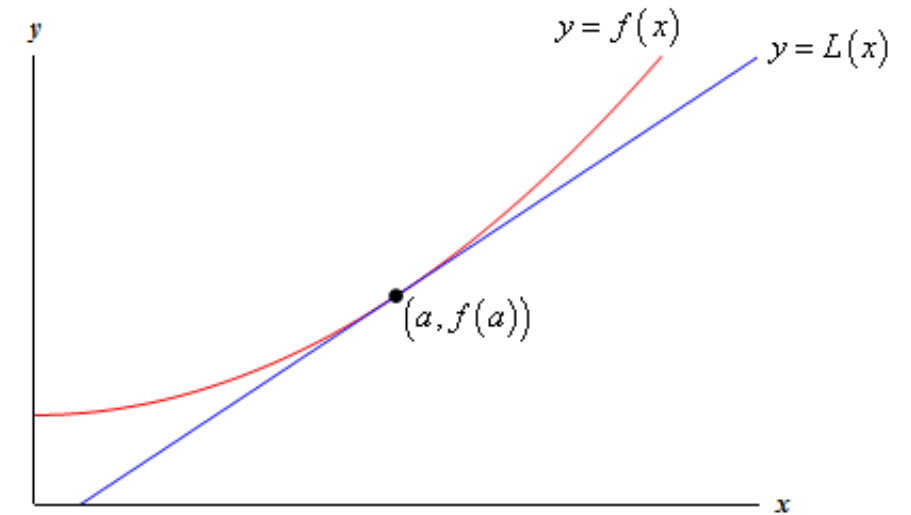
$$\lim_{(h, k) \rightarrow (0, 0)} \frac{f(a + h, b + k) - f(a, b) - f_x(a, b)h - f_y(a, b)k}{\sqrt{h^2 + k^2}} = 0.$$

It is clear how to generalise these definitions for  $n$ -variable scalar-valued functions. They also make sense for vector-valued functions, using addition and scalar-multiplication in  $\mathbb{R}^m$  (see p35).

Before continuing with the theory of differentiability, let us make sure we understand the linearisation and its applications:

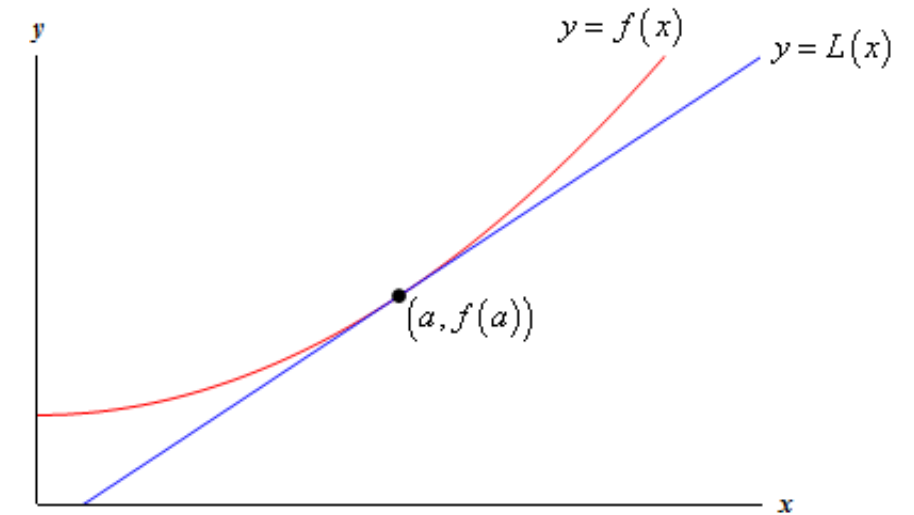
**Example:** Calculate the linearisation of  $f(x, y) = x^2y$  at  $(1, 2)$ , and use it to estimate  $f(1.1, 1.8)$ .

To motivate the second main application of linearisations, recall that the graph of a single-variable linearisation  $y = f(a) + f'(a)(x - a)$  is the tangent line to the graph of  $f$  at  $(a, f(a))$ .



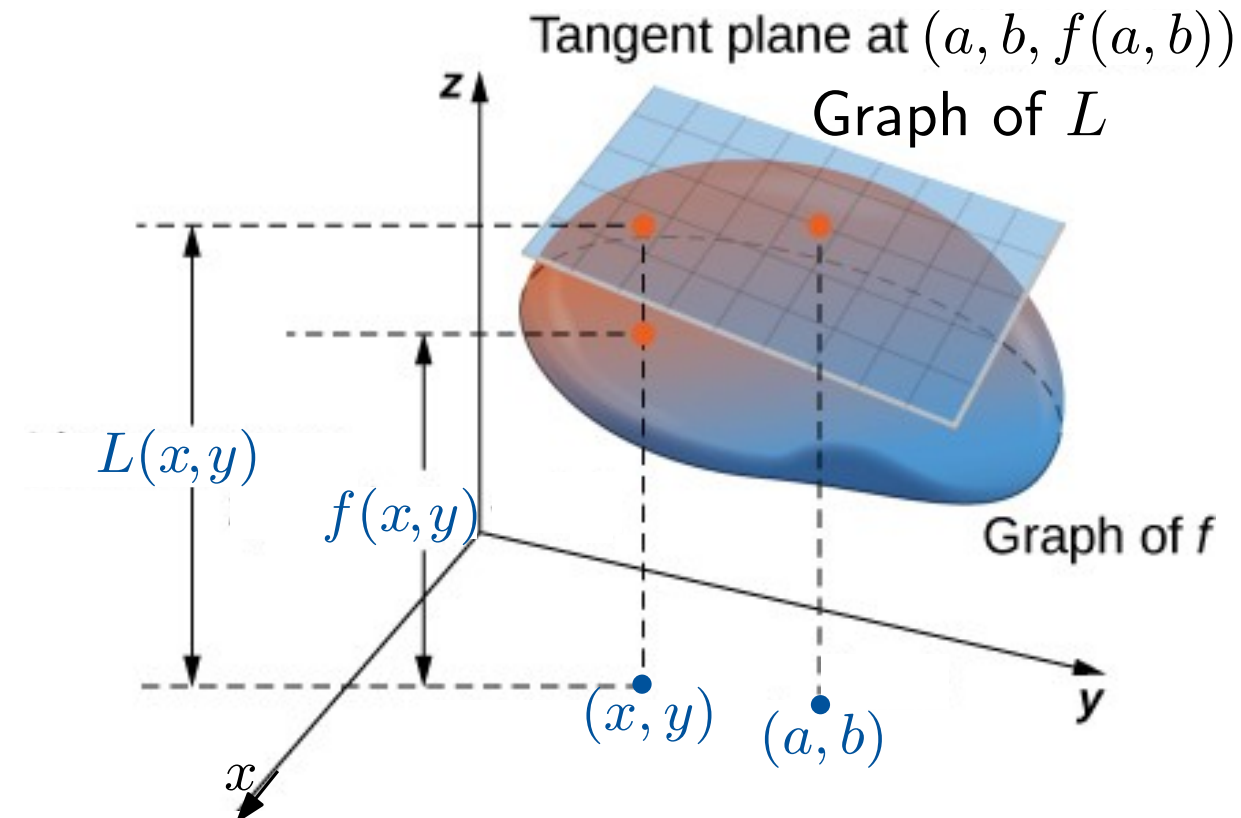
(pictures from Paul's Online Math Notes and archive.cnx.org)

To motivate the second main application of linearisations, recall that the graph of a single-variable linearisation  $y = f(a) + f'(a)(x - a)$  is the tangent line to the graph of  $f$  at  $(a, f(a))$ .



Consider the linearisation of a 2-variable function  $f(x, y)$  at  $(a, b)$ . Its graph is  $z = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)$ , and this is the tangent plane to the graph of  $f$  at  $(a, b, f(a, b))$ .

In Week 9 (§12.7) we will see a more general method that computes tangent planes to any surface, not just to a graph.



(pictures from Paul's Online Math Notes and archive.cnx.org)



Back to differentiability: remember that

**Definition:** A function  $f(x, y)$  is *differentiable* at  $(a, b)$  if

$$\lim_{(h,k) \rightarrow (0,0)} \frac{f(a+h, b+k) - f(a, b) - f_x(a, b)h - f_y(a, b)k}{\sqrt{h^2 + k^2}} = 0.$$

We will rarely need to use this definition to check if functions are differentiable, thanks to the following theorem:

**Theorem 4: Continuous Partial Derivatives guarantee Differentiability:**

Given a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , if all its partial derivatives  $\frac{\partial f}{\partial x_i}$  are continuous around  $(a, b)$ , then  $f$  is differentiable at  $(a, b)$ .

The main idea of the proof (p30) is to write  $f(a+h, b+k) - f(a, b)$  in the definition of differentiable in terms of partial derivatives, using a multivariate mean value theorem. On the next page we state precisely the 2D mean value theorem, you can imagine the analogous statement when the domain is  $\mathbb{R}^n$ .

We are using the 1D MVT phrased as follows: if  $g$  is differentiable on  $[a, a + h]$ , then there is a point between  $a$  and  $a + h$ , i.e. a point  $a + \theta h$  for  $\theta$  between 0 and 1, such that  $f(a + h) - f(a) = hf'(a + \theta h)$ .

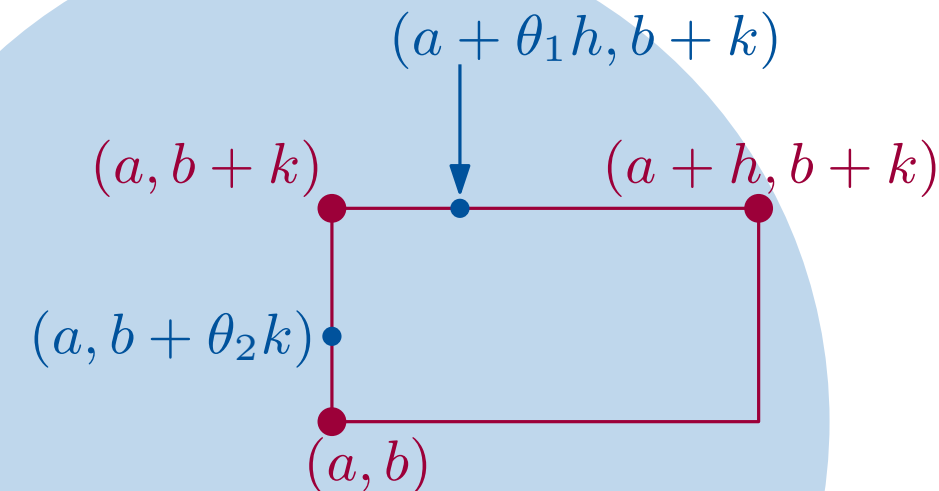
**Theorem 3: Mean Value Theorem (MVT):** Suppose  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  has continuous partial derivatives  $f_x$  and  $f_y$  on a small disk around  $(a, b)$ . If  $h, k$  are small enough that  $(a + h, b + k)$  are in this disk, then there exist numbers  $\theta_1, \theta_2$  between 0 and 1 such that

$$f(a + h, b + k) - f(a, b) = hf_x(a + \theta_1 h, b + k) + kf_y(a, b + \theta_2 k).$$

**Proof:**

$$\begin{aligned} f(a + h, b + k) - f(a, b) &= \underbrace{f(a + h, b + k) - f(a, b + k)}_{\substack{\text{1D MVT in} \\ x\text{-direction}}} + \underbrace{f(a, b + k) - f(a, b)}_{\substack{\text{1D MVT in} \\ y\text{-direction}}} \\ &= hf_x(a + \theta_1 h, b + k) + kf_y(a, b + \theta_2 k). \end{aligned}$$

We are using the 1D MVT phrased as follows: if  $g$  is differentiable on  $[a, a + h]$ , then there is a point between  $a$  and  $a + h$ , i.e. a point  $a + \theta h$  for  $\theta$  between 0 and 1, such that  $f(a + h) - f(a) = hf'(a + \theta h)$ .



**Proof:** (of Theorem 4, sketch):

Recall that MVT says there are numbers  $\theta_1, \theta_2$  between 0 and 1 with  $f(a + h, b + k) - f(a, b) = hf_x(a + \theta_1 h, b + k) + kf_y(a, b + \theta_2 k)$ .

We now use this to show that, if  $f(x, y)$  has continuous partial derivatives, then  $f$  is differentiable. So we need to show that

$$\frac{f(a + h, b + k) - f(a, b) - f_x(a, b)h - f_y(a, b)k}{\sqrt{h^2 + k^2}} \rightarrow 0 \text{ as } (h, k) \rightarrow (0, 0).$$

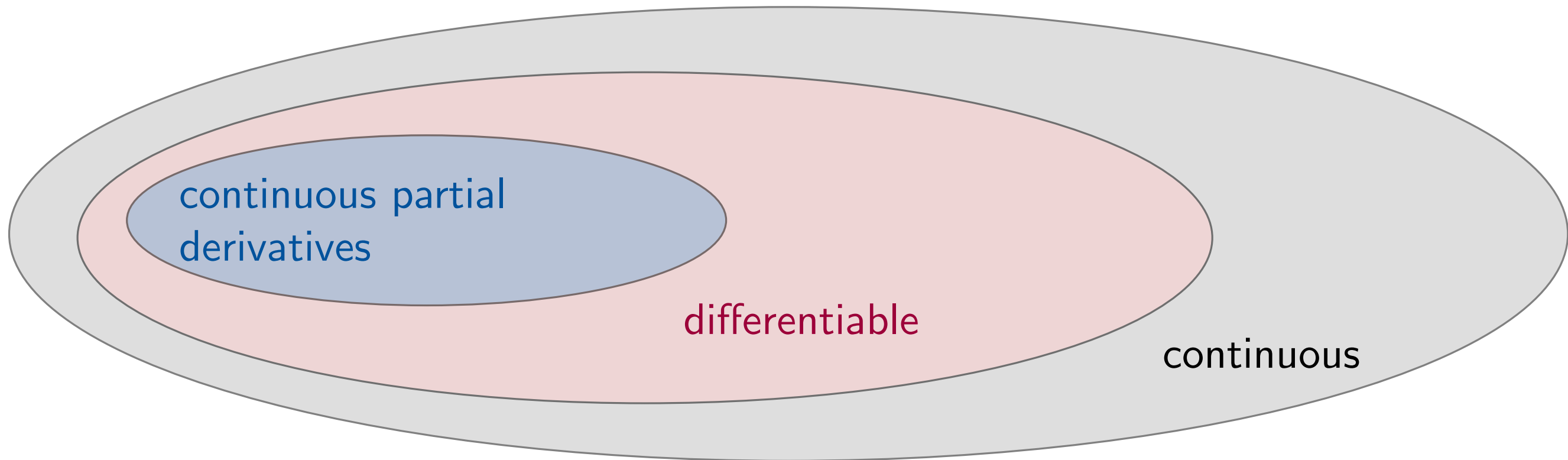
Using MVT to replace the first two terms in the numerator:

$$\begin{aligned} & \frac{hf_x(a + \theta_1 h, b + k) + kf_y(a, b + \theta_2 k) - f_x(a, b)h - f_y(a, b)k}{\sqrt{h^2 + k^2}} \\ &= \underbrace{\frac{h}{\sqrt{h^2 + k^2}}}_{\text{is finite.}} \underbrace{(f_x(a + \theta_1 h, b + k) - f_x(a, b))}_{\text{goes to 0 because } f_x \text{ is continuous}} + \underbrace{\frac{k}{\sqrt{h^2 + k^2}}}_{\text{is finite.}} \underbrace{(f_y(a, b + \theta_2 k) - f_y(a, b))}_{\text{goes to 0 because } f_y \text{ is continuous}} \end{aligned}$$

There is one more important result about differentiability - for simplicity we state it below for 2-variable functions, but it holds for any  $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ :

**Theorem: Differentiable Functions are Continuous:** If  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  is differentiable at  $(a, b)$ , then it is continuous at  $(a, b)$ .

So the hierarchy of functions is as follows:



**Theorem: Differentiable Functions are Continuous:** If  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  is differentiable at  $(a, b)$ , then it is continuous at  $(a, b)$ .

**Proof:** (sketch, same as the 1D proof):

We show that  $\lim_{(h,k) \rightarrow (0,0)} f(a+h, b+k) - f(a, b) = 0$ .

$$\begin{aligned}
 & f(a+h, b+k) - f(a, b) \\
 = & f(a+h, b+k) - f(a, b) - f_x(a, b)h - f_y(a, b)k + f_x(a, b)h + f_y(a, b)k \\
 = & \underbrace{\sqrt{h^2 + k^2}}_{\substack{\text{goes to 0} \\ \text{because it is} \\ \text{a continuous} \\ \text{function of} \\ (h, k)}} \underbrace{\frac{f(a+h, b+k) - f(a, b) - f_x(a, b)h - f_y(a, b)k}{\sqrt{h^2 + k^2}}}_{\substack{\text{goes to 0 because } f \text{ is differentiable}}} + \underbrace{f_x(a, b)h + f_y(a, b)k}_{\substack{\text{goes to 0 because} \\ \text{it is a continuous} \\ \text{function of } (h, k)}}
 \end{aligned}$$

## Derivatives of vector-valued functions

Consider a function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ . We can compute the linearisation of its coordinate functions separately:

$$L_1(x, y) = f_1(a, b) + \left. \frac{\partial f_1}{\partial x} \right|_{(a,b)} (x - a) + \left. \frac{\partial f_1}{\partial y} \right|_{(a,b)} (y - b)$$

$$L_2(x, y) = f_2(a, b) + \left. \frac{\partial f_2}{\partial x} \right|_{(a,b)} (x - a) + \left. \frac{\partial f_2}{\partial y} \right|_{(a,b)} (y - b)$$

$$L_3(x, y) = f_3(a, b) + \left. \frac{\partial f_3}{\partial x} \right|_{(a,b)} (x - a) + \left. \frac{\partial f_3}{\partial y} \right|_{(a,b)} (y - b)$$

## Derivatives of vector-valued functions

Consider a function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ . We can compute the linearisation of its coordinate functions separately:

$$L_1(x, y) = f_1(a, b) + \left. \frac{\partial f_1}{\partial x} \right|_{(a,b)} (x - a) + \left. \frac{\partial f_1}{\partial y} \right|_{(a,b)} (y - b)$$

$$L_2(x, y) = f_2(a, b) + \left. \frac{\partial f_2}{\partial x} \right|_{(a,b)} (x - a) + \left. \frac{\partial f_2}{\partial y} \right|_{(a,b)} (y - b)$$

$$L_3(x, y) = f_3(a, b) + \left. \frac{\partial f_3}{\partial x} \right|_{(a,b)} (x - a) + \left. \frac{\partial f_3}{\partial y} \right|_{(a,b)} (y - b)$$



## Derivatives of vector-valued functions

Consider a function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ . We can compute the linearisation of its coordinate functions separately:

$$L_1(x, y) = f_1(a, b) + \left. \frac{\partial f_1}{\partial x} \right|_{(a,b)} (x - a) + \left. \frac{\partial f_1}{\partial y} \right|_{(a,b)} (y - b)$$

$$L_2(x, y) = f_2(a, b) + \left. \frac{\partial f_2}{\partial x} \right|_{(a,b)} (x - a) + \left. \frac{\partial f_2}{\partial y} \right|_{(a,b)} (y - b)$$

$$L_3(x, y) = f_3(a, b) + \left. \frac{\partial f_3}{\partial x} \right|_{(a,b)} (x - a) + \left. \frac{\partial f_3}{\partial y} \right|_{(a,b)} (y - b)$$

This is matrix multiplication

$$\begin{pmatrix} \left. \frac{\partial f_1}{\partial x} \right|_{(a,b)} & \left. \frac{\partial f_1}{\partial y} \right|_{(a,b)} \\ \left. \frac{\partial f_2}{\partial x} \right|_{(a,b)} & \left. \frac{\partial f_2}{\partial y} \right|_{(a,b)} \\ \left. \frac{\partial f_3}{\partial x} \right|_{(a,b)} & \left. \frac{\partial f_3}{\partial y} \right|_{(a,b)} \end{pmatrix} \begin{pmatrix} x - a \\ y - b \end{pmatrix}$$

So one way to organise the  $mn$  partial derivatives of a function  $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , that makes sense with linear algebra, is with an  $m \times n$  matrix:

**Definition:** The *Jacobian matrix*  $D\mathbf{f}(\mathbf{x})$  of a function  $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is the  $m \times n$  matrix with  $\frac{\partial f_i}{\partial x_j}$  in row  $i$  and column  $j$ :

$$D\mathbf{f}(\mathbf{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}.$$

As observed on the previous page, we can write the linearisation of a vector-valued function using the Jacobian matrix:

$$\mathbf{L}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + \underbrace{D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a})}_{\text{matrix-vector multiplication}}.$$

matrix-vector multiplication

**Example:** Calculate the Jacobian matrix of  $\mathbf{f}(x, y) = \left( \frac{xy}{x+1}, x^2y, x \right)$  at  $(1, 2)$ , and use it to estimate  $\mathbf{f}(1.1, 2.3)$ .

## Non-examinable: the derivative as a linear transformation

Recall that the Jacobian matrix is

$$D\mathbf{f}(\mathbf{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix},$$

and the linearisation is:

$$\mathbf{L}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a}).$$

The linear transformation represented by the Jacobian matrix is called *the derivative* of  $\mathbf{f}$ . It allows a definition of differentiability without reference to coordinates:  $\mathbf{f}$  is differentiable at  $\mathbf{a}$  if there is a linear transformation  $D\mathbf{f}(\mathbf{a})$  such that

$$\lim_{\mathbf{x} \rightarrow \mathbf{a}} \frac{\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a})}{|\mathbf{x} - \mathbf{a}|} = \mathbf{0}.$$