For the next three weeks, we only look at scalar-valued functions $f : \mathbb{R}^n \to \mathbb{R}$.
Our overall aim would be to find the maximum and minimum of these functions.
This week we look at two ideas which are useful to this goal:

- The gradient vector (pp2-8, §12.7 in the textbook)
- Taylor polynomials (pp13-23, §12.9 in the textbook)

In passing, we will also discuss rates of change of f in any direction, and tangent planes and normal lines to surfaces (p9-12).

§12.7: Gradients and Directional Derivatives

Recall that the partial derivatives f_x, f_y of a 2-variable function measure the rate of change when we fix one variable and change the other, i.e. the rate of change in the x or y direction, which in vector notation is the $\mathbf{i}$ or $\mathbf{j}$ direction.

$$f_x(a,b) = \lim_{h\to 0} \frac{f(a+h,b) - f(a,b)}{h}; \quad f_y(a,b) = \lim_{h\to 0} \frac{f(a,b+h) - f(a,b)}{h}$$

What about the rate of change in other directions, e.g. the $2\mathbf{i}+\mathbf{j}$ direction? Equivalently, what is the rate of change of f when x increases twice as fast as y?

§12.7: Gradients and Directional Derivatives

Recall that the partial derivatives f_x, f_y of a 2-variable function measure the rate of change when we fix one variable and change the other, i.e. the rate of change in the x or y direction, which in vector notation is the $\mathbf{i}$ or $\mathbf{j}$ direction.

$$f_x(a,b) = \lim_{h\to 0} \frac{f(a+h,b) - f(a,b)}{h}; \quad f_y(a,b) = \lim_{h\to 0} \frac{f(a,b+h) - f(a,b)}{h}$$

What about the rate of change in other directions, e.g. the $2\mathbf{i}+\mathbf{j}$ direction? Equivalently, what is the rate of change of f when x increases twice as fast as y?

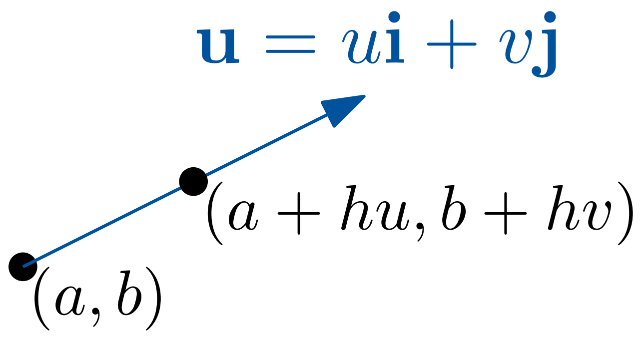
Because we are interested in the direction of change of the input, and not the length of the change vector, we should use a unit vector, i.e. work with $\frac{2}{\sqrt{5}}\mathbf{i} + \frac{1}{\sqrt{5}}\mathbf{j}$.

Definition: Let $\mathbf{u} = u\mathbf{i} + v\mathbf{j}$ be a unit vector. The *directional derivative of $f(x,y)$ at (a,b) in the direction of $\mathbf{u}$* is

$$D_{\mathbf{u}} f(a,b) = \lim_{h\to 0^+} \frac{f(a+hu, b+hv) - f(a,b)}{h}.$$

Definition: Let $\mathbf{u} = u\mathbf{i} + v\mathbf{j}$ be a unit vector. The *directional derivative of $f(x, y)$ at (a, b) in the direction of $\mathbf{u}$* is

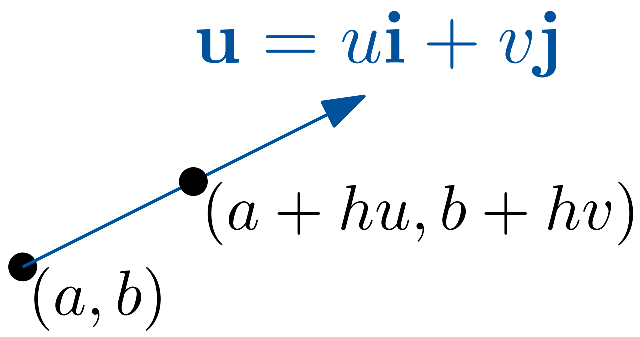
$$D_{\mathbf{u}} f(a, b) = \lim_{h \to 0^+} \frac{f(a + hu, b + hv) - f(a, b)}{h}.$$



This is the rate of change of f as you move from (a, b) in the direction $\mathbf{u}$.

Definition: Let $\mathbf{u} = u\mathbf{i} + v\mathbf{j}$ be a unit vector. The *directional derivative of $f(x, y)$ at (a, b) in the direction of $\mathbf{u}$* is

$$D_{\mathbf{u}} f(a, b) = \lim_{h \to 0^+} \frac{f(a + hu, b + hv) - f(a, b)}{h}.$$



This is the rate of change of f as you move from (a, b) in the direction $\mathbf{u}$.

Observe that, if f is differentiable, then the right hand side in the above definition is $\left.\frac{d}{dh} f(x, y)\right|_{h=0}$, where $x(h) = a + hu$ and $y(h) = b + hv$.
We can calculate this derivative using the multivariate chain rule:

$$\left.\frac{d}{dh} f(x, y)\right|_{h=0} = \left.\left(\frac{\partial f}{\partial x}\frac{dx}{dh} + \frac{\partial f}{\partial y}\frac{dy}{dh}\right)\right|_{h=0} = \left.\frac{\partial f}{\partial x}\right|_{(x,y)=(a,b)} u + \left.\frac{\partial f}{\partial y}\right|_{(x,y)=(a,b)} v.$$

So we can easily calculate the directional derivatives of a differentiable function, using its partial derivatives. This formula is usually expressed in terms of the dot product of the unit vector $\mathbf{u}$ and a vector that contains the partial derivatives.

Definition: Given a function $f(x,y)$ with partial derivatives at (a,b), the *gradient vector of f at (a,b)* is

$$\mathbf{grad} f(a,b) = \nabla f(a,b) = f_x(a,b)\mathbf{i} + f_y(a,b)\mathbf{j}.$$

Similarly, the gradient of an n-variable function at $(a_1,\dots,a_n)$ is a vector in $\mathbb{R}^n$.

What we showed on the previous page is:

Theorem 7: Calculating directional derivatives using the gradient: If $f:\mathbb{R}^n \to \mathbb{R}$ is differentiable at $(a_1,\dots,a_n)$ and $\mathbf{u}$ is a unit vector, then directional derivative of f at $(a_1,\dots,a_n)$ in the direction of $\mathbf{u}$ is

$$D_{\mathbf{u}} f(a_1,\dots,a_n) = \mathbf{u} \bullet \nabla f(a_1,\dots,a_n).$$

Example: Find the rate of change of $f(x, y) = x^2 - y^2$ at $(3, -1)$ in the direction $2\mathbf{i} + \mathbf{j}$. Is f increasing or decreasing in this direction?

The following example explains why it is useful to put the partial derivatives into a gradient vector.

Example: Let $f(x, y) = x^2 + y^2$.

a. Draw the level curves of f.

b. Draw on the same diagram $\nabla f(1, 1)$ and $\nabla f(-1, 1)$.

c. By considering the value of f at points close to $(1, 1)$, estimate the direction at $(1, 1)$ in which f increases most quickly.

We record below the observations from the previous example. These properties hold for (scalar-valued) functions of any number of variables.

Theorem: Geometric properties of the gradient vector:

a. At (a, b), the function $f(x, y)$ increases most rapidly in the direction of $\nabla f(a, b)$. The maximum rate of increase is $|\nabla f(a, b)|$.
b. At (a, b), the function $f(x, y)$ decreases most rapidly in the direction of $-\nabla f(a, b)$. The maximum rate of decrease is $|\nabla f(a, b)|$.
c. $\nabla f(a, b)$ is perpendicular to the level set of f at (a, b).

We record below the observations from the previous example. These properties hold for (scalar-valued) functions of any number of variables.

Theorem: Geometric properties of the gradient vector:

a. At (a, b), the function $f(x, y)$ increases most rapidly in the direction of $\nabla f(a, b)$. The maximum rate of increase is $|\nabla f(a, b)|$.
b. At (a, b), the function $f(x, y)$ decreases most rapidly in the direction of $-\nabla f(a, b)$. The maximum rate of decrease is $|\nabla f(a, b)|$.
c. $\nabla f(a, b)$ is perpendicular to the level set of f at (a, b).

Proof: (of a,b) For a unit vector $\mathbf{u}$, the rate of change of f at (a, b) in the direction of $\mathbf{u}$ is $D_{\mathbf{u}} f(a, b) = \mathbf{u} \bullet \nabla f(a, b)$. By a property of the dot product, this is $|\mathbf{u}|\,|\nabla f(a, b)| \cos\theta$, where θ is the angle between $\mathbf{u}$ and $\nabla f(a, b)$. So the rate of change of f is maximised when $\cos\theta$ is maximised - i.e. when $\cos\theta = 1$, i.e. $\theta = 0$, i.e. when $\mathbf{u}$ is in the same direction as $\nabla f(a, b)$. Similarly, the rate of change of f is minimised (i.e. most negative) when $\cos\theta = -1$, i.e. when $\mathbf{u}$ is in the opposite direction to $\nabla f(a, b)$.

Theorem: Geometric properties of the gradient vector:

c. $\nabla f(a,b)$ is perpendicular to the level set of f at (a,b).

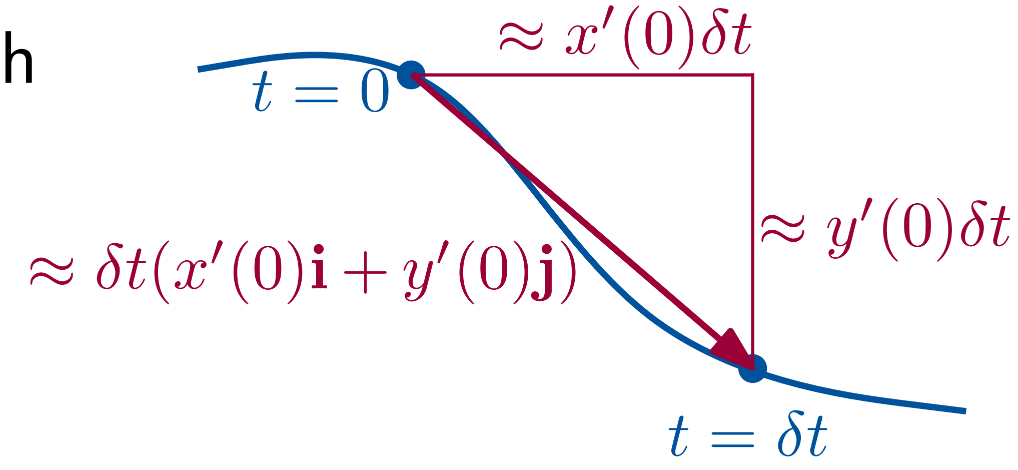
Proof: (of c, sketch) Suppose $(x(t), y(t))$ is a parametrisation of the level set of f that passes through (a,b) and $(a,b) = (x(0), y(0))$.

Because f does not change along the level set: $\displaystyle \frac{d}{dt} f(x(t), y(t)) = 0$

By the multivariate chain rule: $\displaystyle \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = 0$

In particular, when $t = 0$: $\nabla f(a,b) \bullet (x'(0)\mathbf{i} + y'(0)\mathbf{j}) = 0$

So $\nabla f(a,b)$ is perpendicular to $x'(0)\mathbf{i} + y'(0)\mathbf{j}$, which from the picture is tangent to the level curve of f.



(For higher dimensions, apply this argument to all curves $(x_1(t), \dots, x_n(t))$ on the level set of f, to deduce that $\nabla(a_1, \dots, a_n)$ must be perpendicular to all directions tangent to the level set.)

Application of the gradient vector: finding tangent planes and normal lines

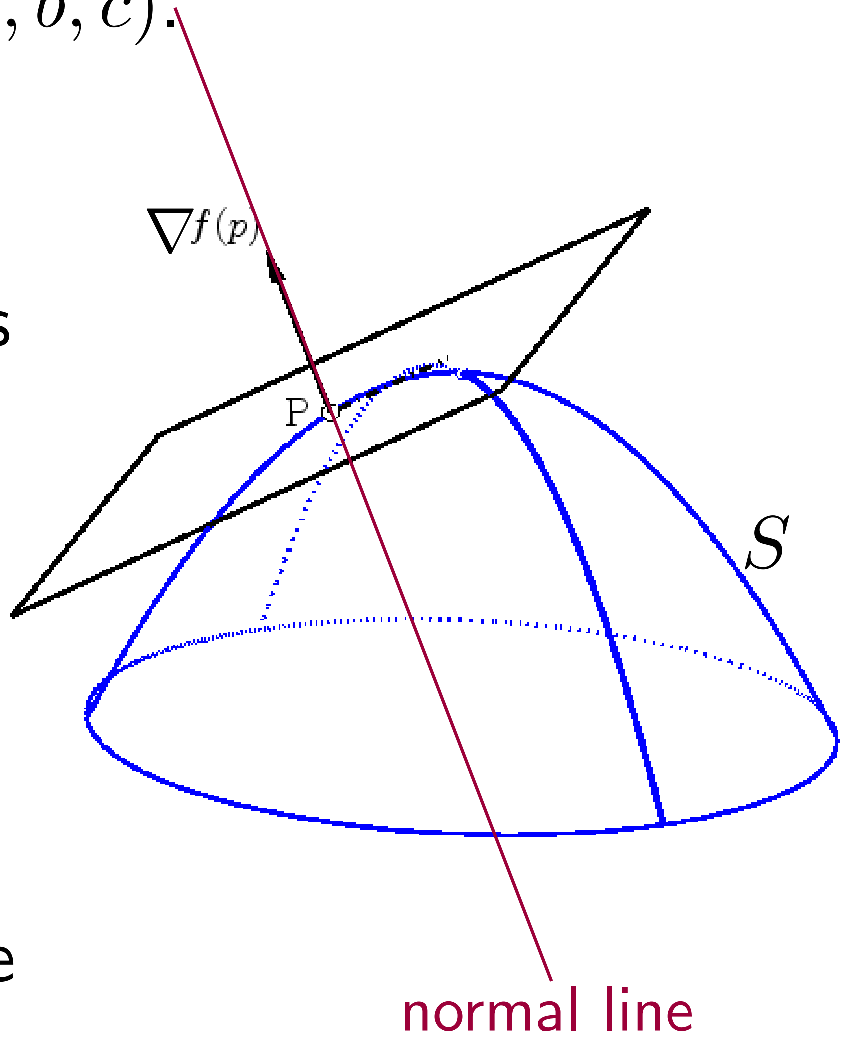
The geometric properties of the gradient on the previous page also apply to 3-variable functions $f(x, y, z)$. In particular:

c. $\nabla f(a, b, c)$ is perpendicular to the level set of f at (a, b, c).

The level set of a 3-variable function is a surface in $\mathbb{R}^3$, let's call it S. So

- the line through (a, b, c) in the direction $\nabla f(a, b, c)$ is the *normal line* to S at (a, b, c), meaning it intersects S perpendicularly at (a, b, c);
- the plane through (a, b, c) with normal $\nabla f(a, b, c)$ is the tangent plane to S at (a, b, c).

One reason to be interested in the normal line: given a point Q, what is the point P on S that is closest to Q? The line segment $\overrightarrow{QP}$ must be perpendicular to S - so we look for a point P where the normal line goes through Q.



(picture from Mathematics Online)

Because we can express any surface defined by an equation as the level set of a function (see week 2 p14), we can use this technique to find the normal line and tangent plane to any surface (see also ex. sheet #17 Q2).

Example: Find an equation for tangent plane to the surface $2x + 2\ln(2y) = 9 - z^2$ at the point $(x, y, z) = (4, \frac{1}{2}, 1)$.

We can use this technique to find tangent planes to graphs:

Example: Find an equation in standard form for the tangent plane to the graph of $f(x, y) = 3ye^{-x}$ when $x = 0$ and $y = 2$.

Now we repeat the previous example for a general function $f(x, y)$, to show how the gradient method of finding tangent planes includes the formula for the tangent plane to a graph (i.e. that it is the graph of the linearisation, see week 8 p26):

The graph of a 2-variable function $f(x, y)$ is $z = f(x, y)$. Call this surface S.
S is the level surface of a different 3-variable function $F(x, y, z) = z - f(x, y)$.
So the tangent plane to S at $(a, b, f(a, b))$ has normal vector

$$\nabla F(a,b,f(a,b)) = \left(\frac{\partial}{\partial x}(z - f(x,y))\mathbf{i} + \frac{\partial}{\partial y}(z - f(x,y))\mathbf{j} + \frac{\partial}{\partial z}(z - f(x,y))\mathbf{k} \right)\Bigg|_{(a,b,f(a,b))}$$

$$= \left(-\frac{\partial f}{\partial x}\mathbf{i} - \frac{\partial f}{\partial y}\mathbf{j} + 1\mathbf{k} \right)\Bigg|_{(a,b,f(a,b))} = -\left.\frac{\partial f}{\partial x}\right|_{(a,b)} \mathbf{i} - \left.\frac{\partial f}{\partial y}\right|_{(a,b)} \mathbf{j} + 1\mathbf{k}.$$

So the equation of the tangent plane is

$$-\left.\frac{\partial f}{\partial x}\right|_{(a,b)} (x-a) - \left.\frac{\partial f}{\partial y}\right|_{(a,b)} (y-b) + 1(z - f(a,b)) = 0, \text{ which rearranges}$$

to $z = f(a,b) + \left.\frac{\partial f}{\partial x}\right|_{(a,b)}(x-a) + \left.\frac{\partial f}{\partial y}\right|_{(a,b)}(y-b)$, the graph of the linearisation.

§12.9: Taylor Polynomials

Given a differentiable single-variable function f, its linearisation at a is a linear function that approximates f near a:

$$f(a+h) \approx L(a+h) = f(a) + f'(a)h.$$

To obtain a better approximation, we can use the *nth order Taylor polynomial* of f about a: (note $P_1 = L$)

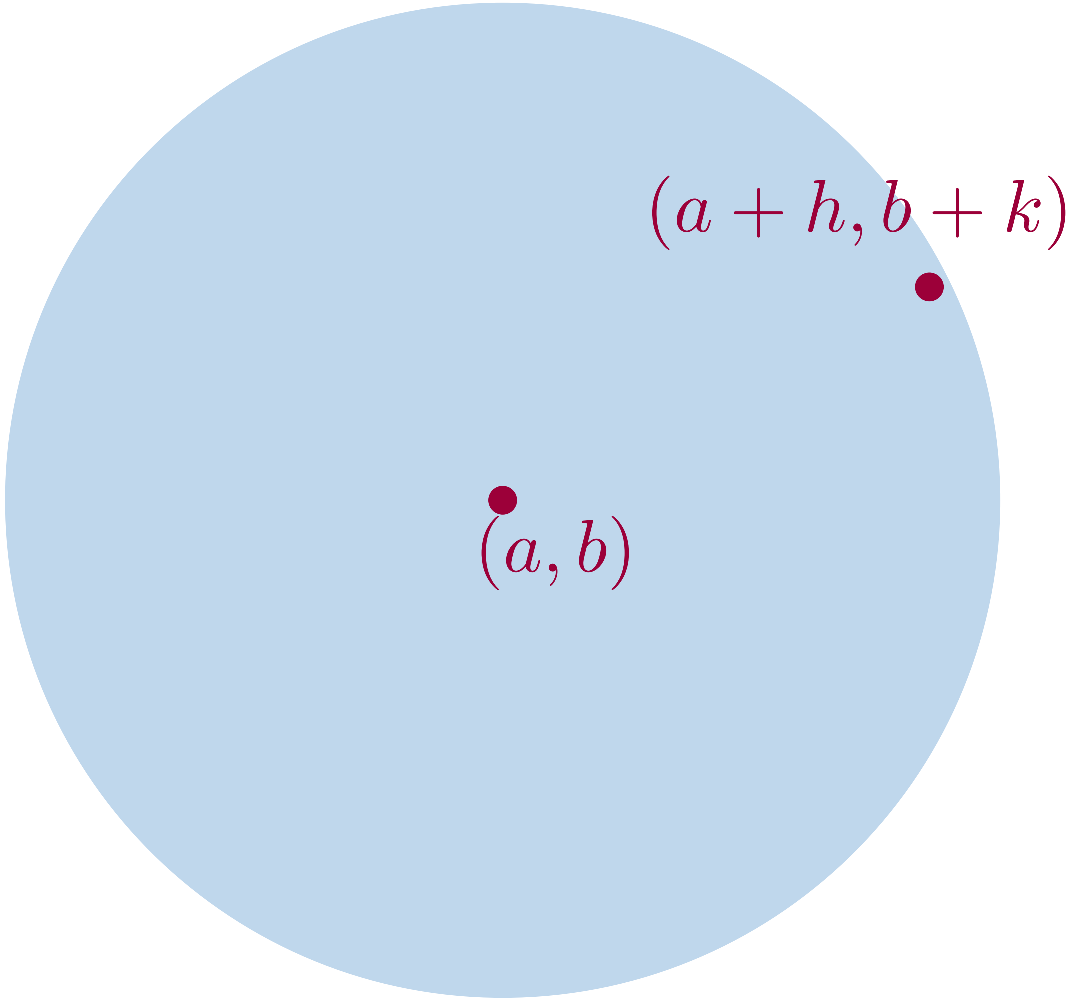
$$f(a+h) \approx P_n(a+h) = f(a) + f'(a)h + \frac{f''(a)}{2!}h^2 + \cdots + \frac{f^{(n)}(a)}{n!}h^n.$$

Example: $(a=0)$

$$e^x \approx \underset{\displaystyle e^0}{\underset{\|}{1}} + \underset{\frac{d}{dx}e^x\big|_{x=0}}{\underset{\|}{1}}\,x + \underset{\frac{d^2}{dx^2}e^x\big|_{x=0}}{\underset{\|}{1}}\,\frac{x^2}{2!} + \cdots + \underset{\frac{d^n}{dx^n}e^x\big|_{x=0}}{\underset{\|}{1}}\,\frac{x^n}{n!}$$

Similarly, for multivariate functions, we can obtain a better approximation than the linearisation by using a degree n polynomial. For example, the third order Taylor polynomial of a 2-variable function f about (a, b) will have the form:

$$f(a+h, b+k) \approx ? + (?h + ?k) + (?h^2 + ?hk + ?k^2) + (?h^3 + ?h^2k + ?hk^2 + ?k^3)$$



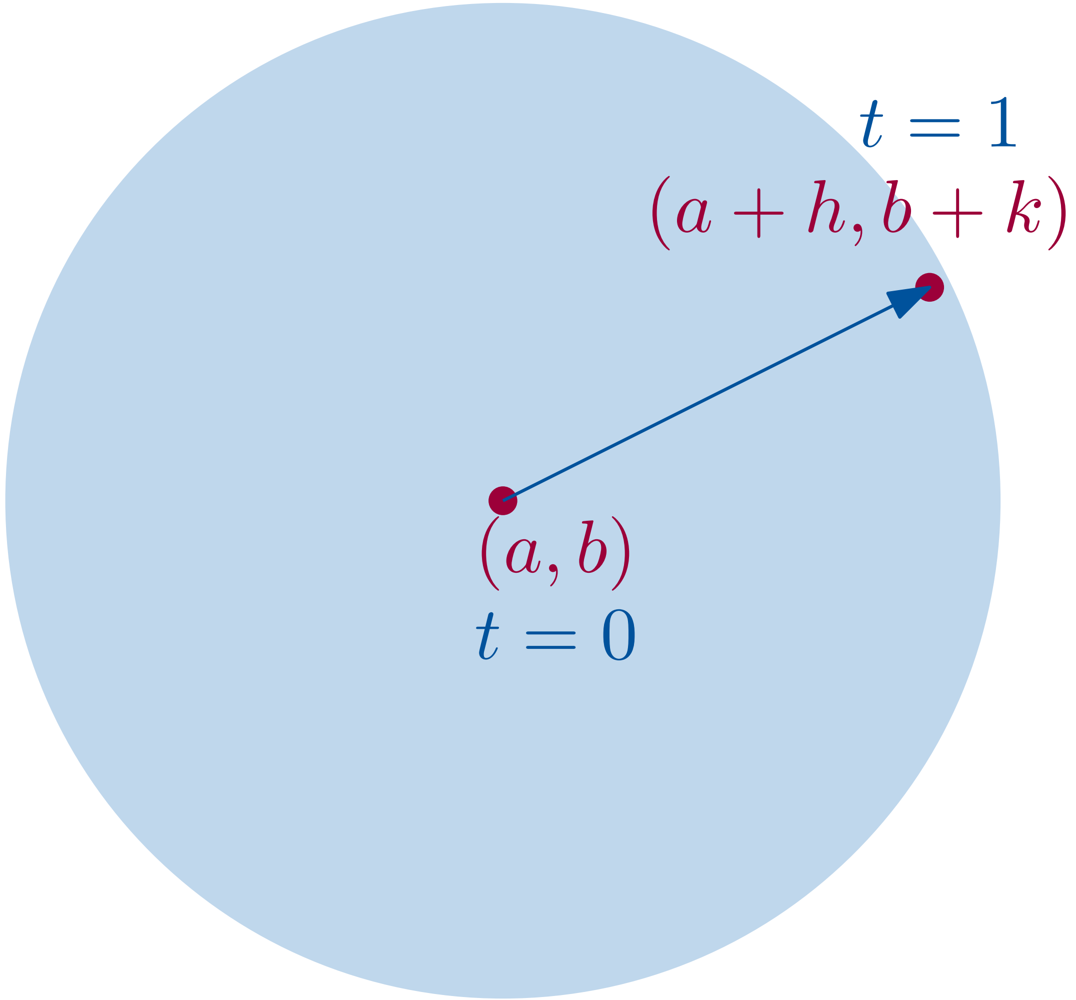
Similarly, for multivariate functions, we can obtain a better approximation than the linearisation by using a degree n polynomial. For example, the third order Taylor polynomial of a 2-variable function f about (a, b) will have the form:

$$f(a+h, b+k) \approx ?+(?h+?k)+(?h^2+?hk+?k^2)+(?h^3+?h^2k+?hk^2+?k^3)$$

To derive such a Taylor polynomial, let's simplify the problem to a 1D problem: fix a point $(a+h, b+k)$ consider f only on the path between (a, b) and $(a+h, b+k)$.

More specifically, let $x(t)=a+th$, $y(t)=b+tk$ (for fixed h, k) and let $F(t)$ be the composition $F(t)=f(x(t), y(t))=f(a+th, b+tk)$.

We will find the 1D Taylor polynomial for $F(t)$ about 0, then substitute in $t=1$.



Recall: $x = a + th$, $y = b + tk$, $F(t) = f(x(t), y(t)) = f(a + th, b + tk)$.

The nth-order Taylor polynomial of $F(t)$ about $t = 0$ is

$$P_n(t) = F(0) + F'(0)\,t + \frac{F''(0)}{2!}\,t^2 + \cdots + \frac{F^{(n)}(0)}{n!}t^n.$$

Recall: $x = a + th$, $y = b + tk$, $F(t) = f(x(t), y(t)) = f(a + th, b + tk)$.

The nth-order Taylor polynomial of $F(t)$ about $t = 0$ is

$$P_n(t) = \boxed{F(0)} + \boxed{F'(0)}t + \frac{\boxed{F''(0)}}{2!}t^2 + \cdots + \frac{F^{(n)}(0)}{n!}t^n.$$

$F(0) = f(a, b)$.

Recall: $x = a + th$, $y = b + tk$, $F(t) = f(x(t), y(t)) = f(a + th, b + tk)$.

The nth-order Taylor polynomial of $F(t)$ about $t = 0$ is

$$P_n(t) = \boxed{F(0)} + \boxed{F'(0)}\,t + \frac{\boxed{F''(0)}}{2!}t^2 + \cdots + \frac{F^{(n)}(0)}{n!}t^n.$$

$F(0) = f(a, b)$.

Using multivariate chain rule:

$$\begin{aligned} F'(t) &= \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} \\ &= f_x h + f_y k \end{aligned}$$

$$F'(0) = f_x(a, b)h + f_y(a, b)k$$

This agrees with the linearisation.

Recall: $x = a + th$, $y = b + tk$, $F(t) = f(x(t), y(t)) = f(a + th, b + tk)$.

The nth-order Taylor polynomial of $F(t)$ about $t = 0$ is

$$P_n(t) = \boxed{F(0)} + \boxed{F'(0)}\,t + \frac{\boxed{F''(0)}}{2!}t^2 + \cdots + \frac{F^{(n)}(0)}{n!}t^n.$$

$F(0) = f(a, b)$.

Using multivariate chain rule:

$$F'(t) = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$
$$= f_x h + f_y k$$
$$F'(0) = f_x(a, b)h + f_y(a, b)k$$

This agrees with the linearisation.

Using multivariate chain rule (in the second line):

$$F''(t) = \frac{d}{dt}F'(t) = \frac{d}{dt}(f_x h + f_y k) = h\frac{df_x}{dt} + k\frac{df_y}{dt}$$
$$= h\left(\frac{\partial f_x}{\partial x}\frac{dx}{dt} + \frac{\partial f_x}{\partial y}\frac{dy}{dt}\right) + k\left(\frac{\partial f_y}{\partial x}\frac{dx}{dt} + \frac{\partial f_y}{\partial y}\frac{dy}{dt}\right)$$
$$= h\left(\frac{\partial f_x}{\partial x}h + \frac{\partial f_x}{\partial y}k\right) + k\left(\frac{\partial f_y}{\partial x}h + \frac{\partial f_y}{\partial y}k\right)$$
$$= h(f_{xx}h + f_{xy}k) + k(f_{yx}h + f_{yy}k)$$
$$F''(0) = f_{xx}(a, b)h^2 + 2f_{xy}(a, b)hk + f_{yy}(a, b)k^2$$

(using $f_{xy} = f_{yx}$ in the last line)

Recall: $x = a + th$, $y = b + tk$, $F(t) = f(x(t), y(t)) = f(a + th, b + tk)$.

The nth-order Taylor polynomial of $F(t)$ about $t = 0$ is

$$P_n(t) = F(0) + F'(0)\,t + \frac{F''(0)}{2!}\,t^2 + \cdots + \frac{F^{(n)}(0)}{n!}t^n.$$

$$= f(a,b) + (f_x(a,b)h + f_y(a,b)k)\,t + \frac{f_{xx}(a,b)h^2 + 2f_{xy}(a,b)hk + f_{yy}(a,b)k^2}{2!}t^2 + \dots$$

Recall: $x = a + th$, $y = b + tk$, $F(t) = f(x(t), y(t)) = f(a + th, b + tk)$.

The nth-order Taylor polynomial of $F(t)$ about $t = 0$ is

$$P_n(t) = F(0) + F'(0)\,t + \frac{F''(0)}{2!}t^2 + \cdots + \frac{F^{(n)}(0)}{n!}t^n.$$

$$= f(a,b) + (f_x(a,b)h + f_y(a,b)k)\,t + \frac{\boxed{f_{xx}(a,b)h^2 + 2f_{xy}(a,b)hk + f_{yy}(a,b)k^2}}{2!}t^2 + \ldots$$

$$F''(t) = \frac{d}{dt}(f_x h + f_y k)$$
$$= f_{xx}h^2 + f_{xy}hk + f_{yx}kh + f_{yy}k^2$$
$$F''(0) = f_{xx}(a,b)h^2 + 2f_{xy}(a,b)hk + f_{yy}(a,b)k^2$$

Notice the pattern in our calculation of $F''(0)$: each differentiation creates two sets of terms, one set where we differentiate with respect to x and multiply by h, and one set where we differentiate with respect to y and multiply by k.

So we expect $F'''(0) = ?f_{xxx}(a,b)h^3 + ?f_{xxy}(a,b)h^2k + ?f_{xyy}(a,b)hk^2 + ?f_{yyy}(a,b)k^3$. (Because of equality of mixed partial derivatives, these four are the only different third-order partial derivatives, see week 7 p22.)

In the calculation of $F''(0)$, the term $f_{xy}(a,b)hk$ has coefficient 2 because there are "two orders" to differentiate with respect to x and y each once: x first then y, or y first then x.

So the coefficient of the f_{xxy} term in $F'''(0)$ is the number of orders to differentiate with respect to x twice and to y once. There are three such ways: $f_{xxy}, f_{xyx}, f_{yxx}$. By the same argument, the coeffficient of the f_{xyy} term is also 3, and the coefficients of the f_{xxx} and f_{yyy} terms are both 1.

Hence $F'''(0) = f_{xxx}(a,b)h^3 + 3f_{xxy}(a,b)h^2k + 3f_{xyy}(a,b)hk^2 + f_{yyy}(a,b)k^3$.

In the calculation of $F''(0)$, the term $f_{xy}(a,b)hk$ has coefficient 2 because there are "two orders" to differentiate with respect to x and y each once: x first then y, or y first then x.

So the coefficient of the f_{xxy} term in $F'''(0)$ is the number of orders to differentiate with respect to x twice and to y once. There are three such ways: $f_{xxy}, f_{xyx}, f_{yxx}$. By the same argument, the coeffficient of the f_{xyy} term is also 3, and the coefficients of the f_{xxx} and f_{yyy} terms are both 1.

Hence $F'''(0) = f_{xxx}(a,b)h^3 + 3f_{xxy}(a,b)h^2k + 3f_{xyy}(a,b)hk^2 + f_{yyy}(a,b)k^3$.

For larger n,

$$F^{(n)}(0) = \underbrace{\frac{\partial^n f}{\partial x^n}(a,b)h^n}_{j=n} + \cdots + \underbrace{\boxed{\frac{n!}{j!(n-j)!}}}_{\text{number of ways to choose } j \text{ objects from } n \text{ objects}} \frac{\partial^n f}{\partial x^j y^{n-j}}(a,b)h^j k^{n-j} + \cdots + \underbrace{\frac{\partial^n f}{\partial y^n}(a,b)k^n}_{j=0}.$$

Putting it all together:
$x = a + th$, $y = b + tk$, $F(t) = f(x(t), y(t)) = f(a + th, b + tk)$.

The nth-order Taylor polynomial of $F(t)$ about $t = 0$ is

$$P_n(t) = F(0) + F'(0)\,t + \frac{F''(0)}{2!}\,t^2 + \cdots + \frac{F^{(n)}(0)}{n!}t^n.$$

So the *nth-order Taylor polynomial of $f(x, y)$ about $(x, y) = (a, b)$* is

$$P_n(1) = F(0) + F'(0) + \frac{F''(0)}{2!} + \cdots + \frac{F^{(n)}(0)}{n!}.$$

$$= f(a,b) + (f_x(a,b)h + f_y(a,b)k) + \frac{f_{xx}(a,b)h^2 + 2f_{xy}(a,b)hk + f_{yy}(a,b)k^2}{2!} + \ldots$$

$$+ \frac{1}{n!}\frac{\partial^n f}{\partial x^n}(a,b)h^n + \cdots + \frac{1}{j!(n-j)!}\frac{\partial^n f}{\partial x^j y^{n-j}}(a,b)h^j k^{n-j} + \cdots + \frac{1}{n!}\frac{\partial^n f}{\partial y^n}(a,b)k^n.$$

Example: Find the second-order Taylor polynomial of $f(x,y) = \dfrac{\sin x}{y}$ about $(x,y) = (0,1)$.

If we want a high order Taylor polynomial, it is often faster to multiply and/or substitute into the Taylor polynomials of the following important 1D functions (if you don't remember them exactly, you can always do some differentiation to double-check):

$$e^t = 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \dots$$

$$\sin t = t - \frac{t^3}{3!} + \frac{t^5}{5!} - \frac{t^7}{7!} + \dots$$

$$\cos t = 1 - \frac{t^2}{2!} + \frac{t^4}{4!} - \frac{t^6}{6!} + \dots$$

$$\ln(1+t) = t - \frac{t^2}{2} + \frac{t^3}{3} - \dots$$

$$\frac{1}{1+t} = 1 - t + t^2 - t^3 + \dots$$

Example: (compare p19) Find the fourth-order Taylor polynomial of $f(x, y) = \dfrac{\sin x}{y}$ about $(x, y) = (0, 1)$.

Example: Find the third-order Taylor polynomial of $\ln(2 + x + 2y^2)$ about $(0, 0)$.

When expanding a Taylor polynomial about (a, b), we need to consider powers of $x - a$ and $y - b$. Hence we need to do some algebraic manipulation such that the 't' that we substitute for in the single-variable Taylor series is in terms of $x - a$ and $y - b$.

Example: (compare ex. sheet #18 q2) Find the third-order Taylor polynomial of $e^x\sqrt{y}$ about $(2, 1)$.