MA 105 : Calculus

D1 and D4 - Upcoming Lectures

Sandip Singh

Department of Mathematics, IIT Bombay

September 11, 2023

Partial derivatives

As before, U will denote a subset of $\mathbb{R}^2$. Given a function $f : U \to \mathbb{R}$, we can fix one of the variables and view the function f as a function of the other variable alone. We can then take the derivative of this one variable function.

To make things precise, fix x_2.

Definition: The partial derivative of $f : U \to \mathbb{R}$ with respect to x_1 at the point (a, b) is defined by

$$\frac{\partial f}{\partial x_1}(a, b) := \lim_{x_1 \to a} \frac{f((x_1, b)) - f((a, b))}{x_1 - a} = \lim_{t \to 0} \frac{f((a + t, b)) - f((a, b))}{t}.$$

Similarly, one can define the partial derivative with respect to x_2.

In this case the variable x_1 is fixed and f is regarded only as a function of x_2:

$$\frac{\partial f}{\partial x_2}(a, b) := \lim_{x_2 \to b} \frac{f((a, x_2)) - f((a, b))}{x_2 - b} = \lim_{t \to 0} \frac{f((a, b + t)) - f((a, b))}{t}.$$

Directional Derivatives

The partial derivatives are special cases of the directional derivative. Let $v = (v_1, v_2)$ be a unit vector. Then v specifies a direction in $\mathbb{R}^2$.

Definition: The directional derivative of f in the direction v at a point $x = (x_1, x_2)$ is defined as

$$\nabla_v f(x) = \lim_{t \to 0} \frac{f(x + tv) - f(x)}{t} = \lim_{t \to 0} \frac{f((x_1 + tv_1, x_2 + tv_2)) - f((x_1, x_2))}{t}.$$

It measures the rate of change of the function f at x along the path $x + tv$.

Observe that if we take $v = (1, 0)$ in the above definition, we obtain $\partial f / \partial x_1$, while $v = (0, 1)$ yields $\partial f / \partial x_2$.

Consider the function

$$f(x_1, x_2) = \begin{cases} 1 & \text{if } x_1 = 0 \text{ or if } x_2 = 0 \\ 0 & \text{otherwise.} \end{cases}$$

It should be clear to you that since this function is constant along the two axes,

$$\frac{\partial f}{\partial x_1}(0,0) = 0 \quad \text{and} \quad \frac{\partial f}{\partial x_2}(0,0) = 0.$$

On the other hand, $f(x_1, x_2)$ is not continuous at the origin. (why?)

Thus, a function may have both partial derivatives (and, in fact, any directional derivative - see the next slide) but still not be continuous.

This suggests that for a function of two variables, just requiring that both partial derivatives exist is not a good or useful definition of "differentiability" (why?).

In the section on iterated limits, we studied the following function from Exercise 5.5:

$$f(x,y) = \frac{x^2y^2}{x^2y^2 + (x-y)^2} \quad \text{for} \quad (x,y) \neq (0,0).$$

Let us further set $f(0,0) = 0$. You can check that every directional derivative of f at $(0,0)$ exists and is equal to 0, except along $y = x$ (that is, along the unit vector $v = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$) when the directional derivative is not defined.

However, we have already seen that the function is not continuous at the origin since we have shown that $\lim_{(x,y)\to(0,0)} f(x,y)$ does not exist. For an example with directional derivatives in all directions see Exercise 5.3(i).

Conclusion: All directional derivatives may exist at a point even if the function is discontinuous at that point.

Let us go back and examine the notion of differentiability for a function $f(x)$ of one variable.

Suppose f is differentiable at the point x_0. When is a line passing through the point $(x_0, f(x_0))$ on the curve $y = f(x)$ is the tangent line to the curve?

Recall that the equation of a line passing through the point $(x_0, f(x_0))$ and having slope m is

$$y = f(x_0) + m(x - x_0).$$

If we consider the difference $f(x) - f(x_0) - m(x - x_0)$ and write $h = (x - x_0)$, we see that the difference can be rewritten as

$$f(x_0 + h) - f(x_0) - m \cdot h.$$

The above line is the tangent line to the curve $y = f(x)$ at the point $(x_0, f(x_0))$ on the curve if $m = f'(x_0)$, which is equivalent of saying that

$$f(x_0 + h) - f(x_0) - m \cdot h = o(h) = \varepsilon_1(h)h$$

where $\varepsilon_1(h)$ is a function of h that goes to 0 as h goes to 0, and in this case the function $o(h) = \varepsilon_1(h)|h|$ is a function of h "that goes to zero faster than h" (that is, $\lim_{h \to 0} \frac{o(h)}{|h|} = 0$).

The preceding idea generalises to two (or more) dimensions.

Let $f(x, y)$ be a function which has both partial derivatives. In the two variable case we need to look at the difference of $z = f(x, y)$ (defining a surface in $\mathbb{R}^3$) and $z = g(x, y)$ (defining a plane in $\mathbb{R}^3$).

Let us first recall how to find the equation of a plane passing through the point $P = (x_0, y_0, z_0)$.

It is the graph of the function

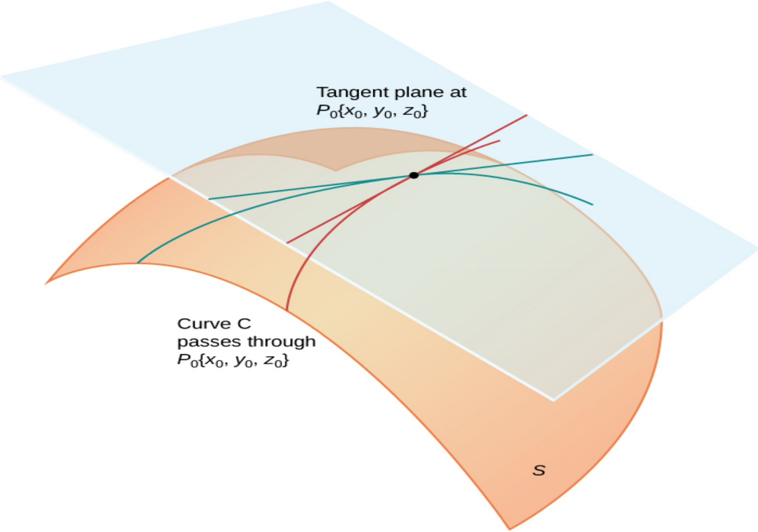
$$z = g(x, y) = z_0 + a(x - x_0) + b(y - y_0).$$

Let us determine the tangent plane to the surface $z = f(x, y)$ passing through a point $P = (x_0, y_0, z_0)$ *on the surface*. In other words, we have to determine the constants a and b so that the above plane becomes the tangent plane to the surface $z = f(x, y)$ at the point $P = (x_0, y_0, z_0)$.

If we fix the y variable as $y = y_0$ and treat $f(x, y)$ only as a function of x, we get a curve on the surface $z = f(x, y)$.

Similarly, if we treat $g(x, y)$ as function only of x (by fixing $y = y_0$), we obtain a line on the plane $z = g(x, y)$.

The tangent plane in a picture



https://openstax.org/books/calculus-volume-3/pages/4-4-tangent-planes-and-linear-approximations

The tangent to the curve passing through (x_0, y_0, z_0) must be the same as the line passing through (x_0, y_0, z_0), and, in any event, their slopes (which are given by the derivatives of the curve $z = f(x, y_0)$ at $x = x_0$ and the line $z = g(x, y_0)$ at $x = x_0$, resp.) must be the same.

Since the above derivatives with respect to x are same as the partial derivatives with respect to x, we get

$$\frac{dz}{dx}(x_0) = \frac{\partial f}{\partial x}(x_0, y_0) = \frac{\partial g}{\partial x}(x_0, y_0) = a.$$

Arguing in exactly the same way but fixing the x variable as $x = x_0$ and varying the y variable, we obtain

$$\frac{\partial f}{\partial y}(x_0, y_0) = \frac{\partial g}{\partial y}(x_0, y_0) = b.$$

Hence, the equation of the tangent plane to the surface $z = f(x, y)$ at the point (x_0, y_0) is (remember that $z_0 = f(x_0, y_0)$)

$$z = f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0).$$

Differentiability for functions of two variables

We now define differentiability for functions of two variables by imitating the one variable definition, but using the "$o(h)$" version.

We let $(x, y) = (x_0, y_0) + (h, k) = (x_0 + h, y_0 + k)$.

Definition A function $f : U \to \mathbb{R}$ is said to be differentiable at a point (x_0, y_0) if $\frac{\partial f}{\partial x}(x_0, y_0)$ and $\frac{\partial f}{\partial y}(x_0, y_0)$ exist and

$$\lim_{(h,k)\to(0,0)} \frac{\left|f(x_0 + h, y_0 + k) - f(x_0, y_0) - \frac{\partial f}{\partial x}(x_0, y_0)h - \frac{\partial f}{\partial y}(x_0, y_0)k\right|}{\|(h, k)\|} = 0.$$

We could rewrite this as

$$\begin{aligned} f((x_0, y_0) + (h, k)) - f(x_0, y_0) - \frac{\partial f}{\partial x}(x_0, y_0)h - \frac{\partial f}{\partial y}(x_0, y_0)k \\ = \varepsilon(h, k)\|(h, k)\| \end{aligned}$$

where $\varepsilon(h, k)$ is a function that goes to 0 as $(h, k) \to (0, 0)$.

This form of differentiability now looks exactly like the one variable version case (put $o(h, k) = \varepsilon(h, k)\|(h, k)\|$). Can you guess the derivative $f'(x_0, y_0)$ of the function $f(x, y)$ at (x_0, y_0)?

The derivative as a linear map

We can rewrite the differentiability criterion once more as follows.

We define the 1×2 matrix

$$Df(x_0, y_0) = \begin{pmatrix} \frac{\partial f}{\partial x}(x_0, y_0) & \frac{\partial f}{\partial y}(x_0, y_0) \end{pmatrix}.$$

A 1×2 matrix can be multiplied by a column vector (which is a 2×1 matrix) to give a real number. In particular:

$$\begin{pmatrix} \frac{\partial f}{\partial x}(x_0, y_0) & \frac{\partial f}{\partial y}(x_0, y_0) \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} = \frac{\partial f}{\partial x}(x_0, y_0)h + \frac{\partial f}{\partial y}(x_0, y_0)k$$

that is,

$$Df(x_0, y_0) \begin{pmatrix} h \\ k \end{pmatrix} = \frac{\partial f}{\partial x}(x_0, y_0)h + \frac{\partial f}{\partial y}(x_0, y_0)k.$$

The definition of differentiability can thus be reformulated using matrix notation.

Definition 2: The function $f(x, y)$ is said to be differentiable at a point (x_0, y_0) if there exists a matrix denoted $Df(x_0, y_0)$ with the property that

$$f((x_0, y_0)+(h, k))-f(x_0, y_0)-Df(x_0, y_0)\begin{pmatrix} h \\ k \end{pmatrix} = o(h, k) = \varepsilon(h, k)\|(h, k)\|$$

for some function $\varepsilon(h, k)$ that goes to 0 as (h, k) goes to $(0, 0)$.

Viewing the derivative as a matrix allows us to view it as a linear map from $\mathbb{R}^2 \to \mathbb{R}$. Given a 1×2 matrix A and two column vectors v and w, we see that

$$A \cdot (v + w) = A \cdot v + A \cdot w \quad \text{and} \quad A \cdot (\lambda v) = \lambda(A \cdot v),$$

for any real number λ.

As we have seen before, functions satisfying the above two properties are called linear functions or linear maps. Thus, the map $v \to A \cdot v$ gives a linear map from $\mathbb{R}^2$ to $\mathbb{R}$.

A condition for differentiability

Exercise: Show that a function $f(x, y)$ is differentiable in the sense of Definition 1 if and only if it is differentiable in the sense of Definition 2 with

$$Df(x_0, y_0) = \begin{pmatrix} \frac{\partial f}{\partial x}(x_0, y_0) & \frac{\partial f}{\partial y}(x_0, y_0) \end{pmatrix}.$$

The matrix $Df(x_0, y_0)$ is called the Derivative matrix of the function $f(x, y)$ at the point (x_0, y_0).

Theorem 1: Let $f : U \to \mathbb{R}$. If the partial derivatives $\frac{\partial f}{\partial x}(x_0, y_0)$ and $\frac{\partial f}{\partial y}(x_0, y_0)$ exist and are continuous in a neighbourhood of a point (x_0, y_0) (that is, in a region of the plane of the form $\{(x, y) \mid \|(x, y) - (x_0, y_0)\| < r\}$ for some $r > 0$), then f is differentiable at (x_0, y_0).

Remark: We omit the proof of this theorem. However, we note that a function whose partial derivatives exist and are continuous is said to be of class $\mathcal{C}^1$. The theorem says that every $\mathcal{C}^1$ function is differentiable.

Differentiability $\Rightarrow$ continuity

Theorem: Let U be a subset of $\mathbb{R}^2$ and $(x_0, y_0) \in U$. If $f : U \longrightarrow \mathbb{R}$ is differentiable at (x_0, y_0), then f is continuous at (x_0, y_0).

Proof: Note that for $\epsilon = 1, \exists \delta_1 > 0$ such that

$|f((x_0, y_0) + (h, k)) - f(x_0, y_0)| = \left| \varepsilon(h, k)\|(h, k)\| + Df(x_0, y_0) \begin{pmatrix} h \\ k \end{pmatrix} \right|$

$\leq |\varepsilon(h, k)| \cdot \|(h, k)\| + \|Df(x_0, y_0)\| \cdot \|(h, k)\|$ (property of dot product)

$< (1 + K)\|(h, k)\|$ (since $\varepsilon(h, k)$ tends to 0 as (h, k) tends to $(0, 0)$)

whenever $||(h, k)|| < \delta_1$ (where $K = \|Df(x_0, y_0)\|$).

Therefore, for a given $\epsilon > 0$, if we take $\delta = \min\{\delta_1, \frac{\epsilon}{1+K}\}$, then

$$|f((x_0, y_0) + (h, k)) - f(x_0, y_0)| < \epsilon$$

whenever $||(h, k)|| < \delta$. Hence the differentiable function f is continuous. $\square$

The Gradient

When viewed as a row vector rather than as a matrix, the derivative matrix is called the gradient and is denoted $\nabla f(x_0, y_0)$. Thus

$$\nabla f(x_0, y_0) = \left(\frac{\partial f}{\partial x}(x_0, y_0), \frac{\partial f}{\partial y}(x_0, y_0) \right).$$

In terms of the coordinate vectors $\mathbf{i}$ and $\mathbf{j}$, the gradient can be written as

$$\nabla f(x_0, y_0) = \frac{\partial f}{\partial x}(x_0, y_0)\mathbf{i} + \frac{\partial f}{\partial y}(x_0, y_0)\mathbf{j}.$$

As we will see later in this lecture, the gradient is related to the directional derivative in the direction v:

$$\nabla_v f = \nabla f \cdot v.$$

Three variables

For the next few slides, we will assume that $f : U \to \mathbb{R}$ is a function of three variables, that is, U is a subset of $\mathbb{R}^3$. In this case, if we denote the variables by x, y and z, we get three partial derivatives as follows: we hold two of the variables constant and vary the third. For instance, if y and z are kept fixed while x is varied, we get the partial derivative with respect to x at the point (a, b, c):

$$\frac{\partial f}{\partial x}(a, b, c) = \lim_{x \to a} \frac{f(x, b, c) - f(a, b, c)}{x - a} = \lim_{t \to 0} \frac{f(a + t, b, c) - f(a, b, c)}{t}.$$

In a similar way, we can define the partial derivatives

$$\frac{\partial f}{\partial y}(a, b, c) \quad \text{and} \quad \frac{\partial f}{\partial z}(a, b, c).$$

Once we have the three partial derivatives we can once again define the gradient of f:

$$\nabla f(a, b, c) = \left(\frac{\partial f}{\partial x}(a, b, c), \frac{\partial f}{\partial y}(a, b, c), \frac{\partial f}{\partial z}(a, b, c) \right).$$

Differentiability in three variables

Exercise 1: Formulate a definition of differentiability for a function of three variables.

Exercise 2: Formulate the analogue of Theorem 1 (of this section) for a function of three variables.

We can also define differentiability for functions from $\mathbb{R}^m$ to $\mathbb{R}^n$ where m and n are any positive integers. We will do this in detail in this course when m and n have the values 1 and 2 and 3.

Finally, the rules for the partial derivatives of sums, differences, products and quotients of functions $f, g : U \to \mathbb{R}$, ($U \subset \mathbb{R}^m$, $m = 2, 3$) are exactly analogous to those for the derivative of functions of one variable.

The derivative of vector-valued functions

We now define the derivative of a function $f : U \to \mathbb{R}^n$, where U is a subset of $\mathbb{R}^m$.

Recall that we can write $f = (f_1, f_2, \ldots, f_n)$ where $f_j = \pi_j \circ f : U \to \mathbb{R}$ and $\pi_j : \mathbb{R}^n \to \mathbb{R}$ is the projection on the j-th coordinate defined as $(y_1, y_2, \ldots, y_n) \mapsto y_j$.

The function f is said to be differentiable at a point x if there exists an $n \times m$ matrix $Df(x)$ such that

$$\lim_{h \to (0,0,\ldots,0)} \frac{\|f(x+h) - f(x) - Df(x)(h)\|}{\|h\|} = 0$$

where $x = (x_1, x_2, \ldots, x_m)$, $h = (h_1, h_2, \ldots, h_m)$ are vectors in $\mathbb{R}^m$ and $Df(x)(h) = Df(x) \cdot h$ is a vector in $\mathbb{R}^n$ (we are considering h here as a column vector, that is, a matrix of order $m \times 1$).

The matrix $Df(x)$ is usually called the total derivative of f. It is also referred as the Jacobian matrix. What are its entries?

From our experience in the 1×2 case we might guess (correctly!) that the entries will be the partial derivatives.

Here is the total derivative or the derivative matrix written out fully:

$$Df(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \cdots & \frac{\partial f_1}{\partial x_m}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) & \cdots & \frac{\partial f_2}{\partial x_m}(x) \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ \frac{\partial f_n}{\partial x_1}(x) & \frac{\partial f_n}{\partial x_2}(x) & \cdots & \frac{\partial f_n}{\partial x_m}(x) \end{bmatrix}.$$

In the 3×1 case (that is, when $m = 1$, $n = 3$ and $f = (f_1, f_2, f_3) : U(\subseteq \mathbb{R}) \to \mathbb{R}^3$) we get

$$f'(t) = Df(t) = \begin{bmatrix} f_1'(t) \\ f_2'(t) \\ f_3'(t) \end{bmatrix}.$$

As before, the derivative may be viewed as a linear map, this time from $\mathbb{R}^m$ to $\mathbb{R}^n$ (or, in the case just above, from $\mathbb{R}$ to $\mathbb{R}^3$).

Norm of a matrix

Let

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

be an $n \times m$ matrix with entries in $\mathbb{R}$. One can define the norm of the matrix A as

$$\|A\| = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2}.$$

Just by using the fact that

$$|a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{im}x_m| \leq \sqrt{\sum_{j=1}^{m} a_{ij}^2}\sqrt{\sum_{j=1}^{m} x_j^2}$$

one can show easily that

$$\|A(x)\| \leq \|A\| \cdot \|x\|$$

for $x = (x_1, x_2, \ldots, x_m) \in \mathbb{R}^m$.

An Exercise and a Remark

Exercise: Following the proof of the continuity of differentiable scalar fields (and by using the property of the norm of a matrix), show that the differentiable vector-valued functions are also continuous.

Note that the scalar (real) valued functions of multi-variables are also known as scalar fields and vector-valued functions as vector fields.

Remark: Theorem 1 holds in this greater generality - a function from $\mathbb{R}^m$ to $\mathbb{R}^n$ is differentiable at a point x_0 if all the partial derivatives $\frac{\partial f_i}{\partial x_j}$ $1 \leq i \leq n$, $1 \leq j \leq m$, are continuous in a neighborhood of x_0 (define a neighborhood of x_0 in $\mathbb{R}^m$!).

Rules for the total derivative

Rule 1: Just like in the one variable case, if f and g are differentiable

$$D(f+g)(x) = Df(x) + Dg(x)$$

and

$$D(cf)(x) = cDf(x),\ \forall c \in \mathbb{R}.$$

Rule 2: Somewhat harder, but only because the notation gets more cumbersome, is the Chain rule:

$$D(f \circ g)(x) = Df(g(x)) \circ Dg(x),$$

where $\circ$ on the right hand side denotes the matrix multiplication.

We will prove the chain rule in a special case when

$$g(t) = (x(t), y(t)) : I(\subseteq \mathbb{R}) \to \mathbb{R}^2 \text{ and } f : \mathbb{R}^2 \to \mathbb{R}.$$

The Chain Rule

We now study the situation where we have composition of functions. We assume that $x, y : I \to \mathbb{R}$ are differentiable functions from some interval (open or closed) to $\mathbb{R}$. Thus the pair $(x(t), y(t))$ defines a function from I to $\mathbb{R}^2$. Suppose we have a function $f : \mathbb{R}^2 \to \mathbb{R}$ which is differentiable. We would like to study the derivative of the composite function $z(t) = f(x(t), y(t))$ from I to $\mathbb{R}$.

Theorem 2: With notation as above

$$\frac{dz}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

For a function $w(t) = f(x(t), y(t), z(t))$ in three variables, the chain rule takes the form

$$\frac{dw}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt}.$$

Notation: $f_x = \frac{\partial f}{\partial x}, \quad f_y = \frac{\partial f}{\partial y}, \quad f_z = \frac{\partial f}{\partial z}.$

The proof of the chain rule in two variable case

Since $x(t)$, $y(t)$ are differentiable,

$$\frac{x(t+h)-x(t)-hx'(t)}{h} = \varepsilon_1(h) \Rightarrow x(t+h) = x(t) + h[x'(t)+\varepsilon_1(h)],$$

$$\frac{y(t+h)-y(t)-hy'(t)}{h} = \varepsilon_2(h) \Rightarrow y(t+h) = y(t) + h[y'(t)+\varepsilon_2(h)]$$

where $\varepsilon_1(h)$ and $\varepsilon_2(h)$ are functions of h that go to zero as h goes to zero. Hence

$$f(x(t+h), y(t+h)) = f(x(t)+h[x'(t)+\varepsilon_1(h)], y(t)+h[y'(t)+\varepsilon_2(h)]). \quad (1)$$

Since the function $f(x,y)$ is differentiable,

$$\begin{aligned}
& f\left(x(t)+h[x'(t)+\varepsilon_1(h)], y(t)+h[y'(t)+\varepsilon_2(h)]\right) \\
&= f(x(t),y(t)) + Df \cdot \begin{bmatrix} h[x'(t)+\varepsilon_1(h)] \\ h[y'(t)+\varepsilon_2(h)] \end{bmatrix} \\
&+ \|(h[x'(t)+\varepsilon_1(h)], h[y'(t)+\varepsilon_2(h)])\| \cdot \varepsilon(h[x'(t)+\varepsilon_1(h)], h[y'(t)+\varepsilon_2(h)]) \\
&= f(x(t),y(t)) + Df \cdot \begin{bmatrix} h[x'(t)+\varepsilon_1(h)] \\ h[y'(t)+\varepsilon_2(h)] \end{bmatrix} + h \cdot \epsilon_3(h) \\
&= f(x(t),y(t)) + f_x \cdot h[x'(t)+\varepsilon_1(h)] + f_y \cdot h[y'(t)+\varepsilon_2(h)] + h \cdot \varepsilon_3(h)
\end{aligned}$$

(since $Df = [f_x \;\; f_y]$)

$$= f(x(t), y(t)) + f_x \cdot h \cdot x'(t) + f_y \cdot h \cdot y'(t) \qquad (2)$$
$$+h[f_x \cdot \varepsilon_1(h) + f_y \cdot \varepsilon_2(h) + \varepsilon_3(h)]$$

where $\varepsilon_3(h)$ is also a function of h which goes to zero as h goes to zero (and $\varepsilon(h_1, h_2)$ goes to zero as (h_1, h_2) go zero).

Now (1) and (2) imply that

$$\lim_{h \to 0} \frac{f(x(t+h), y(t+h)) - f(x(t), y(t)) - f_x x'(t)h - f_y y'(t)h}{h}$$
$$= \lim_{h \to 0} [f_x \cdot \varepsilon_1(h) + f_y \cdot \varepsilon_2(h) + \varepsilon_3(h)] = 0$$

and hence

$$\lim_{h \to 0} \frac{f(x(t+h), y(t+h)) - f(x(t), y(t))}{h} = f_x x'(t) + f_y y'(t)$$

that is,

$$\frac{d}{dt} f(x(t), y(t)) = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}. \qquad \square$$

An application to tangents of curves

A simple example to verify the chain rule: Let $z = f(x, y) = xy$, $x(t) = t^3$ and $y(t) = t^2$. Then $z(t) = t^5$, so $z'(t) = 5t^4$.

On the other hand, using the chain rule we get

$$z'(t) = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = y \cdot 3t^2 + x \cdot 2t = 3t^4 + 2t^4 = 5t^4.$$

Example: A continuous mapping $c : I \to \mathbb{R}^n$ on an interval $I \subseteq \mathbb{R}$ is called a path or curve in $\mathbb{R}^n$, $(n = 2, 3)$. The function $c(t)$ will be given by a tuple of functions form.

Let us consider a curve $c(t)$ in $\mathbb{R}^3$. Each point on the curve will be given by a triple of coordinates which will depend on t, that is, the curve can be described by a triple of functions $(g(t), h(t), k(t))$.

We can write

$$c(t) = g(t)\mathbf{i} + h(t)\mathbf{j} + k(t)\mathbf{k}, \quad \text{and if } c'(t_0) = g'(t_0)\mathbf{i} + h'(t_0)\mathbf{j} + k'(t_0)\mathbf{k}$$

exists and is nonzero, it represents the tangent vector to the curve $c(t)$ at the point $c(t_0)$.

For an example, consider the curve $c(t) = (t, \sqrt{1-t^2})$ in $\mathbb{R}^2$ and defined on the interval $[-1, 1]$. Observe that the curve $c(t)$ represents the **upper unit semicircle** centered at the origin.

You can verify easily that whenever $c'(t_0)$ exists and is nonzero, the tangent line to the circle $c(t)$ at the point $c(t_0)$ is the line that passes through the point $c(t_0)$ and is parallel to the tangent vector $c'(t_0)$. $\square$

So far our example has nothing to do with the chain rule. Suppose $z = f(x, y)$ is a surface, and our curve given by $c(t) = (g(t), h(t), f(g(t), h(t))$ lies on the surface $z = f(x, y)$.

Let us compute the tangent vector to the curve at $c(t_0) = (x_0, y_0, z_0)$. It is given by

$$c'(t_0) = g'(t_0)\mathbf{i} + h'(t_0)\mathbf{j} + k'(t_0)\mathbf{k},$$

where $k(t) = f(g(t), h(t))$. Using the chain rule we see that

$$k'(t_0) = \frac{\partial f}{\partial x}(x_0, y_0)g'(t_0) + \frac{\partial f}{\partial y}(x_0, y_0)h'(t_0).$$

We can further show that this tangent vector lies on the tangent plane to the surface $z = f(x, y)$.

Indeed, we have already seen that the tangent plane has the equation

$$z = f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0).$$

A normal vector to this plane is given by

$$\left(-\frac{\partial f}{\partial x}(x_0, y_0), -\frac{\partial f}{\partial y}(x_0, y_0), 1\right).$$

Thus, to verify that the tangent vector

$$c'(t_0) = \left(g'(t_0), h'(t_0), \frac{\partial f}{\partial x}(x_0, y_0)g'(t_0) + \frac{\partial f}{\partial y}(x_0, y_0)h'(t_0)\right)$$

at the point $c(t_0) = (x_0, y_0, z_0)$ on the curve $c(t)$ lies on the plane, we need only check that its dot product with the normal vector is 0. But this is now clear. $\square$

The Gradient of scalar fields

When viewed as a row vector rather than as a matrix, the derivative matrix of $f : U \to \mathbb{R}$ at a point $(a_1, a_2, \ldots, a_m) \in U \subseteq \mathbb{R}^m$ is called the gradient of f at $(a_1, a_2, \ldots, a_m)$ and is denoted as $\nabla f(a_1, a_2, \ldots, a_m)$. Thus

$$\nabla f(a_1, a_2, \ldots, a_m) = \left(\frac{\partial f}{\partial x_1}(a_1, a_2, \ldots, a_m), \ldots, \frac{\partial f}{\partial x_m}(a_1, a_2, \ldots, a_m) \right).$$

For the case $m = 2$, in terms of the coordinate vectors $\mathbf{i}$ and $\mathbf{j}$ the gradient can be written as

$$\nabla f(x_0, y_0) = \frac{\partial f}{\partial x}(x_0, y_0)\mathbf{i} + \frac{\partial f}{\partial y}(x_0, y_0)\mathbf{j}.$$

As we will see in the next lecture, the gradient is related to the directional derivative in the direction v:

$$\nabla_v f = \nabla f \cdot v.$$

Another application of the chain rule: Directional derivatives

Let $U \subset \mathbb{R}^3$ and let $f : U \to \mathbb{R}$ be differentiable. We want to relate the directional derivative to the gradient.

We consider the (differentiable) curve $c(t) = (x_0, y_0, z_0) + tv$, where $v = (v_1, v_2, v_3)$ is a **unit vector**. We can rewrite $c(t)$ as $c(t) = (x_0 + tv_1, y_0 + tv_2, z_0 + tv_3)$. We apply the chain rule to compute the derivative of the function $f(c(t))$ (observe that $\frac{d}{dt}f(c(t))$ at $t = 0$ is same as the directional derivative of f at (x_0, y_0, z_0) in the direction of the vector $v = (v_1, v_2, v_3)$):

$$\frac{d}{dt}f(c(0)) = \frac{\partial f}{\partial x}(x_0, y_0, z_0)v_1 + \frac{\partial f}{\partial y}(x_0, y_0, z_0)v_2 + \frac{\partial f}{\partial z}(x_0, y_0, z_0)v_3$$

and this can be rewritten as

$$\nabla_v f(x_0, y_0, z_0) = \frac{d}{dt}f(c(0)) = \nabla f(x_0, y_0, z_0) \cdot v.$$

Of course, the same argument works when $U \subset \mathbb{R}^2$ and f is a function of two variables.

The Chain Rule and Gradients

The preceding argument is a special case of a more general fact.

Let $c(t)$ be a (differentiable) curve in $\mathbb{R}^3$. Then, by writing $c(t) = (x(t), y(t), z(t))$ and then using the chain rule for the derivative of $f(c(t))$ we obtain that

$$\frac{d}{dt} f(c(t)) = \nabla f(c(t)) \cdot c'(t).$$

Going back to the directional derivative, we can ask ourselves the following question. In what direction is f changing fastest at a given point (x_0, y_0, z_0)? In other words, in which direction does the directional derivative attain its largest value?

Using what we have just learnt, we are looking for a unit vector $v = (v_1, v_2, v_3)$ such that

$$\nabla f(x_0, y_0, z_0) \cdot v$$

is as large as possible (we assume that $\nabla f(x_0, y_0, z_0) \neq (0, 0, 0)$).

We rewrite the preceding dot product as

$$\nabla f(x_0, y_0, z_0) \cdot v = \|\nabla f(x_0, y_0, z_0)\| \|v\| \cos \theta$$

where θ is the angle between v and $\nabla f(x_0, y_0, z_0)$. Since v is a unit vector, this gives

$$\nabla f(x_0, y_0, z_0) \cdot v = \|\nabla f(x_0, y_0, z_0)\| \cos \theta.$$

The maximum value on the right hand side is obviously attained when $\theta = 0$, that is, when v points in the direction of ∇f. In other words the function is increasing fastest in the direction v given by ∇f. Thus the unit vector that we seek is

$$v = \frac{\nabla f(x_0, y_0, z_0)}{\|\nabla f(x_0, y_0, z_0)\|}.$$

Surfaces defined implicitly

So far we have only been considering surfaces of the form $z = f(x, y)$, where f was a function on a subset of $\mathbb{R}^2$. We now consider a more general type of surface S defined implicitly:

$$S = \{(x, y, z) \mid f(x, y, z) = b\}$$

where b is a constant. Most surfaces we have come across are usually described in this form: for instance, the sphere which is given by $x^2 + y^2 + z^2 = r^2$ or the right circular cone $x^2 + y^2 - z^2 = 0$. Let us try to understand what a tangent plane is more precisely.

If S is a surface, a tangent plane to S at a point $s \in S$ (if it exists) is a plane that contains the tangent lines at s to all curves passing through s and lying on S.

If $c(t)$ is a curve on the surface S given by $f(x, y, z) = b$, we see that $f(c(t)) = b$, and hence

$$\frac{d}{dt} f(c(t)) = 0.$$

On the other hand, by the chain rule,

$$0 = \frac{d}{dt} f(c(t)) = \nabla f(c(t)) \cdot c'(t).$$

Thus, if $s_0 = c(t_0) = (x_0, y_0, z_0)$ is a point on the surface, we see that

$$\nabla f(c(t_0)) \cdot c'(t_0) = 0,$$

for every curve $c(t)$ on the surface S passing through $s_0 = (x_0, y_0, z_0)$.

Hence, if $\nabla f(x_0, y_0, z_0) \neq 0$, then $\nabla f(x_0, y_0, z_0)$ is perpendicular to the tangent plane of S at (x_0, y_0, z_0). How to determine a vector perpendicular to the tangent plane at the point (x_0, y_0, z_0) on the surface S given by $z = f(x, y)$? It is determined by $\nabla(z - f(x, y))$ at the point (x_0, y_0, z_0).

The equation of the tangent plane

Since we know that the gradient of f is normal to the level surface $S = \{(x, y, z) \in \mathbb{R}^3 \mid f(x, y, z) = c\}$ (provided the gradient is nonzero), it allows us to write down the equation of the tangent plane of S at the point $s_0 = (x_0, y_0, z_0)$. The equation of this plane is

$$f_x(x_0, y_0, z_0)(x - x_0) + f_y(x_0, y_0, z_0)(y - y_0) + f_z(x_0, y_0, z_0)(z - z_0) = 0.$$

For the curve $f(x, y) = c$, by considering y as a function of x (implicitly) we obtain (by differentiating $f(x, y)$ with respect to x and using the chain rule) that

$$f_x(x_0, y_0) + f_y(x_0, y_0)\frac{dy}{dx} = 0 \Rightarrow \frac{dy}{dx} = -\frac{f_x(x_0, y_0)}{f_y(x_0, y_0)}$$

and hence the equation of the tangent line to the curve $f(x, y) = c$ passing through (x_0, y_0) is

$$y - y_0 = -\frac{f_x(x_0, y_0)}{f_y(x_0, y_0)}(x - x_0),$$

that is, $f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) = 0.$

Gravitational force as gradient of the potential energy

Let $\mathbf{r}$ denote the position vector

$$x\mathbf{i} + y\mathbf{j} + z\mathbf{k},$$

of a point $P = (x, y, z)$ in $\mathbb{R}^3$. Instead of writing $\|\mathbf{r}\|$, it is customary to write r. This notation is very useful.

For instance, the Newton's Law of Gravitation can be expressed as

$$\mathbf{F} = -\frac{GMm}{r^3} \cdot \mathbf{r}$$

where the mass M is assumed to be at the origin, $\mathbf{r}$ denotes the position vector of the mass m, G is a constant and $\mathbf{F}$ denotes the gravitational force between the two (point) masses.

A simple computation shows that

$$\nabla\left(\frac{1}{r}\right) = -\frac{\mathbf{r}}{r^3}.$$

Thus the gravitational force at any point can be expressed as the gradient of a function called the potential function/energy.

Keeping our previous discussion in mind, we know that if

$$V = -\frac{GMm}{r}, \quad \mathbf{F} = -\frac{GMm}{r^3}\cdot\mathbf{r} = -\nabla V.$$

What are the level surfaces of V? Clearly, r must be a constant on these level sets, so the level surfaces are spheres.

Since $\mathbf{F} = -\nabla V$, we see that the gravitational force $\mathbf{F}$ is orthogonal to the sphere and points towards the origin. $\square$

Review - problems involving the gradient

Exercise 1: Find the points on the hyperboloid $x^2 - y^2 + 2z^2 = 1$ where the normal line is parallel to the line that joins the points $(3, -1, 0)$ and $(5, 3, 6)$.

Solution: The hyperboloid is an implicitly defined surface. A normal vector at a point (x_0, y_0, z_0) on the hyperboloid is given by the gradient of the function $x^2 - y^2 + 2z^2$ at (x_0, y_0, z_0):

$$\nabla f(x_0, y_0, z_0) = (2x_0, -2y_0, 4z_0).$$

We require this vector to be parallel to the line joining the points $(3, -1, 0)$ and $(5, 3, 6)$. This line lies in the same direction as the vector $(5 - 3, 3 + 1, 6 - 0) = (2, 4, 6)$. Thus we need only solve the equations

$$(2x_0, -2y_0, 4z_0) = \lambda(2, 4, 6),$$

for some $\lambda \in \mathbb{R}$ such that the point (x_0, y_0, z_0) lies on the hyperboloid. By solving the above equations, we find that $x_0 = \lambda$, $y_0 = -2\lambda$ and $z_0 = (3/2)\lambda$. Substituting x_0, y_0, z_0 in the equation of the hyperboloid yields $\lambda = \pm\sqrt{2/3}$. $\square$

Problems involving the gradient, continued

Exercise 2: Find the directions in which the directional derivative of $f(x, y) = x^2 + \sin xy$ at the point $(1, 0)$ has the value 1.

Solution: We compute ∇f first:

$$\nabla f(x, y) = (2x + y \cos xy, x \cos xy),$$

so at $(1, 0)$ we get, $\nabla f(1, 0) = (2, 1)$.

To find the directional derivative in the direction $v = (v_1, v_2)$ (where v is a unit vector), we simply take the dot product with the gradient:

$$\nabla_v f(1, 0) = \nabla f(1, 0) \cdot v = (2, 1) \cdot (v_1, v_2) = 2v_1 + v_2.$$

This will have value "1" when $2v_1 + v_2 = 1$, subject to $v_1^2 + v_2^2 = 1$, which yields $v_1 = 0, v_2 = 1$ or $v_1 = 4/5, v_2 = -3/5$. $\square$

Review of the gradient

Exercise 3: Find $D_u F(2,2,1)$ where D_u denotes the directional derivative of the function $F(x,y,z) = 3x - 5y + 2z$ and u is the unit vector in the direction of the outward normal to the sphere $x^2 + y^2 + z^2 = 9$ at the point $(2,2,1)$.

Solution: The unit outward normal to the sphere $g(x,y,z) = x^2 + y^2 + z^2 = 9$ at $(2,2,1)$ is given by

$$u = \frac{\nabla g(2,2,1)}{\|\nabla g(2,2,1)\|}.$$

We see that $\nabla g(2,2,1) = (4,4,2)$ so the corresponding unit vector u is $\frac{1}{3}(2,2,1)$.

To get the directional derivative we simply take the dot product of $\nabla F(2,2,1) = (3,-5,2)$ with $u = \frac{1}{3}(2,2,1)$:

$$D_u F(2,2,1) = (3,-5,2) \cdot \frac{1}{3}(2,2,1) = -\frac{2}{3}.$$

$\square$