MA 105 : Calculus

D1 - Lecture 16

Sandip Singh

Department of Mathematics

Autumn 2023, IIT Bombay, Mumbai

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 $\frac{\partial f}{\partial x}(x, y)$ and $\frac{\partial f}{\partial y}(x, y)$ 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. $\square$

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 at b and c, respectively, while x is varied, we get the partial derivative of the function f with respect to x at the point (a, b, c) as

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

In a similar way, we can define the partial derivatives

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

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

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

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.