LECTURE 6

Date of Lecture: February 3, 2020

The symbol ☡ is for flagging a cautionary comment or a tricky argument. It occurs in the margins and is Knuth's version of Bourbaki's "dangerous bend symbol".

An n-tuple $(x_1, \dots, x_n)$ of symbols (x_i not necessarily real or complex numbers) will also be written as a column vector when convenient. Thus

$$(x_1, \dots, x_n) = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

A map $\boldsymbol{f}$ from a set S to a product set $T_1 \times \cdots \times T_n$ will often be written as an n-tuple $\boldsymbol{f} = (f_1, \dots, f_n)$, with f_i a map from S to T_i, and hence, by the above convention, as a column vector

$$\boldsymbol{f} = \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}.$$

(See Remark 2.2.2 of Lecture 5.)

The default norm on Euclidean spaces of the form $\mathbf{R}^n$ is the Euclidean norm $\| \, \|_2$ and we will simply denote it as $\| \, \|$.

☡ Note that $(x_1, \dots, x_n) \neq [x_1 \dots x_n]$. Each side is the transpose of the other.

1. Derivatives

1.1. Partial derivatives.

Many of you might have been introduced to partial derivatives already in an intuitive way. For example if

$$f(x, y, z) = x^3yz + z\sin(x^2 + y^2)$$

then

$$\frac{\partial f}{\partial x}(x, y, z) = 3x^2yz + 2xz\cos(x^2 + y^2),$$
$$\frac{\partial f}{\partial y}(x, y, z) = x^3z + 2yz\cos(x^2 + y^2),$$

and

$$\frac{\partial f}{\partial z}(x, y, z) = x^3y + \sin(x^2 + y^2).$$

The idea is of course that when computing, say, $\frac{\partial f}{\partial x}$, we differentiate with respect to x and pretend y and z are constants. Similar considerations are at work when computing other partial derivatives. More precisely, it is clear that the right definition for the partial derivative of the above f with respect to x is the limit $\frac{\partial f}{\partial x} = \lim_{h \to 0} \frac{f(x+h,y,z) - f(x,y,z)}{h}$. Analogous definitions can be given for other partial derivatives. A little thought shows that if $\boldsymbol{p} = (x, y, z)$ and $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, and $\boldsymbol{e}_3$ are the standard basis vectors for $\mathbf{R}^3$ (e.g. $\boldsymbol{e}_1 = (1, 0, 0)$), then $\frac{\partial f}{\partial x} = \lim_{h \to 0} \frac{f(\boldsymbol{p}+h\boldsymbol{e}_1)-f(\boldsymbol{p})}{h}$, $\frac{\partial f}{\partial y} = \lim_{h \to 0} \frac{f(\boldsymbol{p}+h\boldsymbol{e}_2)-f(\boldsymbol{p})}{h}$, and $\frac{\partial f}{\partial z} = \lim_{h \to 0} \frac{f(\boldsymbol{p}+h\boldsymbol{e}_3)-f(\boldsymbol{p})}{h}$.

Here is the formal definition.

Definition 1.1.1. Let $\boldsymbol{f} \colon U \to \mathbf{R}^m$ be a map where U is an open subset of $\mathbf{R}^n$ and let $\boldsymbol{a}$ be a point in U. For $i \in \{1, \dots, n\}$, the i^{th} partial derivative of $\boldsymbol{f}$ is said to exist at $\boldsymbol{a}$ if the limit

$$\lim_{h \to 0} \frac{\boldsymbol{f}(\boldsymbol{a} + h\boldsymbol{e}_i) - \boldsymbol{f}(\boldsymbol{a})}{h}.$$

exists. In such a case the i^{th} partial derivative of $\boldsymbol{f}$ at $\boldsymbol{a}$ is defined to be the above limit. The i^{th} partial derivative of $\boldsymbol{f}$ at $\boldsymbol{a}$, when it exists, is denoted by any of the following symbols:

$$\frac{\partial \boldsymbol{f}}{\partial x_i}(\boldsymbol{a}), \quad \frac{\partial \boldsymbol{f}}{\partial x_i}\Big|_{\boldsymbol{x}=\boldsymbol{a}}, \quad \text{and} \quad D_i\boldsymbol{f}(\boldsymbol{a}).$$

If $\boldsymbol{f} = (f_1, \dots, f_m)$, clearly $\frac{\partial \boldsymbol{f}}{\partial x_i}(\boldsymbol{a}) = \big(\frac{\partial f_1}{\partial x_i}(\boldsymbol{a}), \dots, \frac{\partial f_m}{\partial x_i}(\boldsymbol{a})\big)$.

1.2. **Properties of the derivative.** Let U be open in $\mathbf{R}^n$, $\boldsymbol{f} \colon U \to \mathbf{R}^m$ a map, and $\boldsymbol{a}$ a point in U. Recall that in Definition 2.2.1 of Lecture 5 we said that $\boldsymbol{f}$ is differentiable at $\boldsymbol{a}$ if there exists $T \in L(\mathbf{R}^n, \mathbf{R}^m)$ such that

$$\lim_{\boldsymbol{h} \to \boldsymbol{0}} \frac{\|\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{h}) - \boldsymbol{f}(\boldsymbol{a}) - T\boldsymbol{h}\|}{\|\boldsymbol{h}\|} = 0.$$

If this happens, we call T the derivative of $\boldsymbol{f}$ at $\boldsymbol{a}$. If $\boldsymbol{f}$ is differentiable at $\boldsymbol{a}$, we write $\boldsymbol{f}'(\boldsymbol{a})$ for its derivative at $\boldsymbol{a}$.

The following lemma shows that the derivative is well defined when it exists.

Lemma 1.2.1. *Let U be an open subset of $\mathbf{R}^n$, $\boldsymbol{a}$ a point in U, and $\boldsymbol{f} \colon U \to \mathbf{R}^m$ a map. Let $A, B \in L(\mathbf{R}^n, \mathbf{R}^m)$ be such that*

$$\lim_{\boldsymbol{h} \to \boldsymbol{0}} \frac{\|\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{h}) - \boldsymbol{f}(\boldsymbol{a}) - A\boldsymbol{h}\|}{\|\boldsymbol{h}\|} = \lim_{\boldsymbol{h} \to \boldsymbol{0}} \frac{\|\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{h}) - \boldsymbol{f}(\boldsymbol{a}) - B\boldsymbol{h}\|}{\|\boldsymbol{h}\|} = 0.$$

Then $A = B$.

Proof. Let $C = B - A$. Then

$$\begin{aligned} \frac{\|C\boldsymbol{h}\|}{\|\boldsymbol{h}\|} &= \frac{\|(\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{h}) - \boldsymbol{f}(\boldsymbol{a}) - A\boldsymbol{h}) - (\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{h}) - \boldsymbol{f}(\boldsymbol{a}) - B\boldsymbol{h})\|}{\|\boldsymbol{h}\|} \\ &\leq \frac{\|\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{h}) - \boldsymbol{f}(\boldsymbol{a}) - A\boldsymbol{h}\|}{\|\boldsymbol{h}\|} + \frac{\|\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{h}) - \boldsymbol{f}(\boldsymbol{a}) - B\boldsymbol{h}\|}{\|\boldsymbol{h}\|}. \end{aligned}$$

It follows that

$$\lim_{\boldsymbol{h} \to \boldsymbol{0}} \frac{\|C\boldsymbol{h}\|}{\|\boldsymbol{h}\|} = 0.$$

Fix a non-zero vector $\boldsymbol{v} \in \mathbf{R}^n$ and let $\boldsymbol{h}_t = t\boldsymbol{v}$ for $t \in \mathbf{R}$. Then the above equality gives

$$\lim_{t \to 0} \frac{\|C\boldsymbol{h}_t\|}{\|\boldsymbol{h}_t\|} = 0.$$

Now $\frac{\|C\boldsymbol{h_t}\|}{\|\boldsymbol{h_t}\|} = \frac{\|C(t\boldsymbol{v})\|}{\|t\boldsymbol{v}\|} = \frac{\|C\boldsymbol{v}\|}{\|\boldsymbol{v}\|}$ for all $t \neq 0$. Thus the above limit gives $\frac{\|C\boldsymbol{v}\|}{\|\boldsymbol{v}\|} = 0$, i.e. $C\boldsymbol{v} = \mathbf{0}$. Since $\boldsymbol{v}$ was an arbitrary non-zero element of $\mathbf{R}^n$, this means $C = 0$, i.e. $A = B$. □

Theorem 1.2.2. *Let U be an open subset of $\mathbf{R}^n$, $\boldsymbol{a}$ a point in U, and $\boldsymbol{f}\colon U \to \mathbf{R}^m$ a map. Suppose $\boldsymbol{f}$ is differentiable at $\boldsymbol{a}$. Then $\boldsymbol{f}$ is continuous at $\boldsymbol{a}$.*

Proof. Let $T = \boldsymbol{f}'(\boldsymbol{a})$. Then (as T is continuous, being a linear transformation),

$$\|\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{h}) - \boldsymbol{f}(\boldsymbol{a})\| \leq \|\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{h}) - \boldsymbol{f}(\boldsymbol{a}) - T\boldsymbol{h}\| + \|T\boldsymbol{h}\| \longrightarrow 0 \quad (\text{as } \boldsymbol{h} \longrightarrow \mathbf{0}).$$

This proves that $\boldsymbol{f}$ is continuous at $\boldsymbol{a}$. □

1.2.3. Let $\boldsymbol{f}\colon U \to \mathbf{R}^m$ be a map, where U is open in $\mathbf{R}^n$. The map $\boldsymbol{f}$ is said to be differentiable on U if $\boldsymbol{f}$ is differentiable at every point of U. In this case we have a map $\boldsymbol{f}'\colon U \to L(\mathbf{R}^n, \mathbf{R}^m)$ given by $\boldsymbol{x} \mapsto \boldsymbol{f}'(\boldsymbol{x})$. The map $\boldsymbol{f}'$ is also denoted $D\boldsymbol{f}$.

Similarly, if $D_i\boldsymbol{f}(\boldsymbol{x})$ exists for every $\boldsymbol{x} \in U$, we say the i^{th} partial derivative of $\boldsymbol{f}$ exists on U. In this case we regard the i^{th} partial derivative of $\boldsymbol{f}$ as an $\mathbf{R}^m$-valued function on U and denote the function as $D_i\boldsymbol{f}$, or as $\frac{\partial \boldsymbol{f}}{\partial x_i}$.

1.2.4. Using the standard bases on $\mathbf{R}^n$ and $\mathbf{R}^m$ we have a canonical vector space isomorphism $L(\mathbf{R}^n, \mathbf{R}^m) \xrightarrow{\sim} M_{\mathbf{R}}(m, n)$ where $M_{\mathbf{R}}(m, n)$ is the space of $m \times n$ matrices. The latter is isomorphic to $\mathbf{R}^{mn}$. In any case, via these isomorphisms $L(\mathbf{R}^n, \mathbf{R}^m)$ acquires a norm, and since all norms on a finite dimensional vector space are equivalent, if $D\boldsymbol{f}$ exists on U, then the notion of the continuity as well as the differentiability of $D\boldsymbol{f}$ at a point on U or on all of U makes sense, independent of the norm imposed on $L(\mathbf{R}^n, \mathbf{R}^n)$. Moreover if the derivative of $D\boldsymbol{f}$ exists at a point in U then the value of the derivative at that point is also independent of the norm imposed. The same discussion applies to each partial derivative $D_i\boldsymbol{f}$.

2. The matrix of $\boldsymbol{f}'(\boldsymbol{a})$

Throughout $\boldsymbol{f}$ is a function on an open set U in $\mathbf{R}^n$ taking values in $\mathbf{R}^m$.

2.1. **Matrix representations.** Suppose $A\colon \mathbf{R}^n \to \mathbf{R}^m$ is a linear transformation and $[A]$ its matrix representation with respect to the standard bases on $\mathbf{R}^n$ and $\mathbf{R}^m$. Let $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ be the standard basis of $\mathbf{R}^n$. This means

$$\boldsymbol{e}_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \longleftarrow \quad i^{\text{th}} \text{ place} \quad .$$

Then it is well known that

$$A\boldsymbol{e}_j = j^{\text{th}} \text{ column of } [A], \qquad (j \in \{1, \ldots, n\}). \tag{2.1.1}$$

2.1.2. Now suppose $T = \boldsymbol{f}'(\boldsymbol{a})$ for some point $\boldsymbol{a} \in U$. Fix $j \in \{1, \dots, n\}$. We have

$$\begin{aligned}\lim_{h\to 0}\left\|\frac{\boldsymbol{f}(\boldsymbol{a}+h\boldsymbol{e}_j)-\boldsymbol{f}(\boldsymbol{a})}{h} - T\boldsymbol{e}_j\right\| &= \lim_{h\to 0}\frac{\|\boldsymbol{f}(\boldsymbol{a}+h\boldsymbol{e}_j)-\boldsymbol{f}(\boldsymbol{a})-T(h\boldsymbol{e}_j)\|}{|h|}\\ &= \lim_{h\to 0}\frac{\|\boldsymbol{f}(\boldsymbol{a}+h\boldsymbol{e}_j)-\boldsymbol{f}(\boldsymbol{a})-T(h\boldsymbol{e}_j)\|}{\|h\boldsymbol{e}_j\|}\\ &= 0\end{aligned}$$

since $T = \boldsymbol{f}'(\boldsymbol{a})$. Thus $\lim_{h\to 0}\frac{\boldsymbol{f}(\boldsymbol{a}+h\boldsymbol{e}_j)-\boldsymbol{f}(\boldsymbol{a})}{h}$ exists and

$$\lim_{h\to 0}\frac{\boldsymbol{f}(\boldsymbol{a}+h\boldsymbol{e}_j)-\boldsymbol{f}(\boldsymbol{a})}{h} = T\boldsymbol{e}_j. \tag{2.1.2.1}$$

In other words $D_j\boldsymbol{f}(\boldsymbol{a})$ exists and

$$T\boldsymbol{e}_j = D_j\boldsymbol{f}(\boldsymbol{a}). \tag{2.1.2.2}$$

We therefore have the following result:

Theorem 2.1.3. *Let U be an open subset of $\mathbf{R}^n$, $\boldsymbol{a}$ a point in U, and $\boldsymbol{f}\colon U \to \mathbf{R}^m$ a map. Let $\boldsymbol{f} = (f_1, \dots, f_m)$. If $\boldsymbol{f}$ is differentiable at $\boldsymbol{a}$ then the partial derivatives $D_j\boldsymbol{f}(\boldsymbol{a})$ exist for $1 \le j \le n$ and the matrix representation of $\boldsymbol{f}'(\boldsymbol{a})$ in the standard bases of $\mathbf{R}^n$ and $\mathbf{R}^m$ is*

$$J\boldsymbol{f}(\boldsymbol{a}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\boldsymbol{a}) & \dots & \frac{\partial f_1}{\partial x_n}(\boldsymbol{a}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\boldsymbol{a}) & \dots & \frac{\partial f_m}{\partial x_n}(\boldsymbol{a}) \end{bmatrix}$$

Proof. This is an immediate consequence of (2.1.1), (2.1.2.1), and (2.1.2.2). □

For the corollary below, it may be a good idea to revisit the discussion in §§§1.2.4.

Corollary 2.1.4. *If $\boldsymbol{f}$ is differentiable on U then $D\boldsymbol{f}\colon U \to L(\mathbf{R}^n, \mathbf{R}^m)$ is continuous if and only if the partial derivatives $D_i\boldsymbol{f}\colon U \to \mathbf{R}^m$ are continuous for $1 \le i \le n$.*

Proof. This is immediate from the theorem, the discussion in §§§1.2.4 and the fact that a function $\boldsymbol{g}\colon \mathbf{R}^l \to \mathbf{R}^k$ is continuous if and only each g_i, $1 \le i \le k$, is continuous, where $\boldsymbol{g} = (g_1, \dots, g_k)$. □

Definition 2.1.5. If $\boldsymbol{f}$ satisfies the hypotheses of Theorem 2.1.3, the matrix $J\boldsymbol{f}(\boldsymbol{a})$ displayed in the theorem is called the *Jacobian matrix of $\boldsymbol{f}$ at $\boldsymbol{a}$*. If $\boldsymbol{f}$ is differentiable on all of U, then the matrix of functions

$$J\boldsymbol{f} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

is called the *Jacobian matrix* of $\boldsymbol{f}$.

Remark 2.1.6. If we blur the distinction between $L(\mathbf{R}^n, \mathbf{R}^m)$ and $M_{\mathbf{R}}(m, n)$ using the standard bases on $\mathbf{R}^n$ and $\mathbf{R}^m$, then $D\boldsymbol{f} = J\boldsymbol{f}$. In many textbooks $J\boldsymbol{f}$ is called the derivative of $\boldsymbol{f}$. We too will often do away with the distinction between $L(\mathbf{R}^n, \mathbf{R}^m)$ and $M_{\mathbf{R}}(m, n)$.

Theorem 2.1.7. (The Chain Rule in Several Variables) *Let* $U \subset \mathbf{R}^n$, $V \subset \mathbf{R}^m$ *be open subsets,* $\boldsymbol{f}\colon U \to \mathbf{R}^m$ *and* $\boldsymbol{g}\colon V \to \mathbf{R}^k$ *maps with* $\boldsymbol{f}(U) \subset V$, $\boldsymbol{a}$ *a point in* U *and* $\boldsymbol{b} = \boldsymbol{f}(\boldsymbol{a})$. *If* $\boldsymbol{f}$ *is differentiable at* $\boldsymbol{a}$, *and* $\boldsymbol{g}$ *at* $\boldsymbol{b}$, *then* $\boldsymbol{g} \circ \boldsymbol{f}$ *is differentiable at* $\boldsymbol{a}$ *and*

$$(\boldsymbol{g} \circ \boldsymbol{f})'(\boldsymbol{a}) = \boldsymbol{g}'(\boldsymbol{b})\boldsymbol{f}'(\boldsymbol{a}).$$

Proof. Let $A = \boldsymbol{f}'(\boldsymbol{a})$ and $B = \boldsymbol{g}'(\boldsymbol{b})$. For $\boldsymbol{h}$ such that $\boldsymbol{a} + \boldsymbol{h} \in U$ and $\boldsymbol{k}$ such that $\boldsymbol{b} + \boldsymbol{k} \in V$ set

$$\boldsymbol{\varepsilon}(\boldsymbol{h}) = \boldsymbol{f}(\boldsymbol{a} + \boldsymbol{h}) - \boldsymbol{f}(\boldsymbol{a}) - A\boldsymbol{h} \quad \text{and} \quad \boldsymbol{\eta}(\boldsymbol{k}) = \boldsymbol{g}(\boldsymbol{b} + \boldsymbol{k}) - \boldsymbol{g}(\boldsymbol{b}) - B\boldsymbol{k}.$$

Then

$$(*) \qquad \lim_{\boldsymbol{h} \to \boldsymbol{0}} \frac{\|\boldsymbol{\varepsilon}(\boldsymbol{h})\|}{\|\boldsymbol{h}\|} = \lim_{\boldsymbol{k} \to \boldsymbol{0}} \frac{\|\boldsymbol{\eta}(\boldsymbol{k})\|}{\|\boldsymbol{k}\|} = 0.$$

Let $\zeta\colon (V - \boldsymbol{b}) \smallsetminus \{\boldsymbol{0}\} \to [0, \infty)$ be the map $\boldsymbol{k} \mapsto \|\boldsymbol{\eta}(\boldsymbol{k})\|/\|\boldsymbol{k}\|$. By $(*)$, $\lim_{\boldsymbol{k} \to \boldsymbol{0}} \zeta(\boldsymbol{k}) = 0$, and so we can extend ζ to a continuous function on $V - \boldsymbol{b}$ by setting $\zeta(\boldsymbol{0}) = 0$.

Let W be a ball centered at $\boldsymbol{0} \in \mathbf{R}^n$ such that $W + \boldsymbol{a} \subset U$. Let $\boldsymbol{h} \in W \smallsetminus \{\boldsymbol{0}\}$ and set

$$\boldsymbol{k} = \boldsymbol{f}(\boldsymbol{a} + \boldsymbol{h}) - \boldsymbol{f}(\boldsymbol{a}) = \boldsymbol{f}(\boldsymbol{a} + \boldsymbol{h}) - \boldsymbol{b}.$$

Note that

$$(**) \qquad A\boldsymbol{h} = \boldsymbol{k} - \boldsymbol{\varepsilon}(\boldsymbol{h}).$$

With $\|A\|$ and $\|B\|$ denoting operator norms on A and B respectively (see (2.1.1) of Lecture 5), we have the following chain of inequalities.

$$\begin{aligned}
\frac{\|\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{h})) - \boldsymbol{g}(\boldsymbol{f}(\boldsymbol{a})) - BA\boldsymbol{h}\|}{\|\boldsymbol{h}\|} &= \frac{\|\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{h})) - \boldsymbol{g}(\boldsymbol{f}(\boldsymbol{a})) - B(\boldsymbol{k} - \boldsymbol{\varepsilon}(\boldsymbol{h}))\|}{\|\boldsymbol{h}\|} \quad \text{(via } (**)\text{)} \\
&= \frac{\|\boldsymbol{g}(\boldsymbol{b}+\boldsymbol{k}) - \boldsymbol{g}(\boldsymbol{b}) - B(\boldsymbol{k}) + B\boldsymbol{\varepsilon}(\boldsymbol{h}))\|}{\|\boldsymbol{h}\|} \\
&\le \frac{\|\boldsymbol{g}(\boldsymbol{b}+\boldsymbol{k}) - \boldsymbol{g}(\boldsymbol{b}) - B(\boldsymbol{k})\|}{\|\boldsymbol{h}\|} + \frac{\|B\boldsymbol{\varepsilon}(\boldsymbol{h}))\|}{\|\boldsymbol{h}\|} \\
&= \frac{\|\boldsymbol{\eta}(\boldsymbol{k})\|}{\|\boldsymbol{h}\|} + \frac{\|B\boldsymbol{\varepsilon}(\boldsymbol{h}))\|}{\|\boldsymbol{h}\|} \\
&= \zeta(\boldsymbol{k})\frac{\|\boldsymbol{k}\|}{\|\boldsymbol{h}\|} + \frac{\|B\boldsymbol{\varepsilon}(\boldsymbol{h}))\|}{\|\boldsymbol{h}\|} \\
&= \zeta(\boldsymbol{k})\frac{\|A\boldsymbol{h} + \boldsymbol{\varepsilon}(\boldsymbol{h})\|}{\|\boldsymbol{h}\|} + \frac{\|B\boldsymbol{\varepsilon}(\boldsymbol{h}))\|}{\|\boldsymbol{h}\|} \quad \text{(via } (**)\text{)} \\
&\le \zeta(\boldsymbol{k})\frac{\|A\|\|\boldsymbol{h}\| + \|\boldsymbol{\varepsilon}(\boldsymbol{h})\|}{\|\boldsymbol{h}\|} + \frac{\|B\|\|\boldsymbol{\varepsilon}(\boldsymbol{h}))\|}{\|\boldsymbol{h}\|} \\
&= \zeta(\boldsymbol{k})\left\{\|A\| + \frac{\|\boldsymbol{\varepsilon}(\boldsymbol{h})\|}{\|\boldsymbol{h}\|}\right\} + \|B\|\frac{\|\boldsymbol{\varepsilon}(\boldsymbol{h}))\|}{\|\boldsymbol{h}\|}.
\end{aligned}$$

By the continuity of $\boldsymbol{f}$ at $\boldsymbol{a}$ we see that $\boldsymbol{k} \to \boldsymbol{0}$ as $\boldsymbol{h} \to \boldsymbol{0}$. By the continuity of ζ we have $\zeta(\boldsymbol{k}) \to 0$ as $\boldsymbol{k} \to \boldsymbol{0}$. Hence by $(*)$, the expression on the last line of the above chain of inequalities tends to 0 as $\boldsymbol{h} \to \boldsymbol{0}$. This proves the theorem. □

3. The class $\mathscr{C}^k$

3.1. **Higher order derivatives.** From the discussion §§§1.2.4 we know that if $\boldsymbol{f}\colon U \to \mathbf{R}^m$ is differentiable on U, where U, as usual, is open in $\mathbf{R}^n$, then one can talk about $D\boldsymbol{f}$ being differentiable at a point $\boldsymbol{a}$, and write $D^2\boldsymbol{f}(\boldsymbol{a})$ for $D\circ D\boldsymbol{f}(\boldsymbol{a}) = D(D\boldsymbol{f})(\boldsymbol{a})$. In view of Theorem 2.1.3, if $D^2\boldsymbol{f}(\boldsymbol{a})$ exists then $D_i \circ D_j\boldsymbol{f}(\boldsymbol{a})$ exists for all $1 \le i, j \le n$. One can of course take this further to any order. Partial derivatives (when they exist) of the form $D_{i_1}\circ D_{i_2}\circ \ldots D_{i_k}(\boldsymbol{f})$ (respectively $D_{i_1}\circ D_{i_2}\circ \ldots D_{i_k}(\boldsymbol{f})(\boldsymbol{a})$) are called *partial derivatives of order* k (respectively *partial derivatives of order* k *at* $\boldsymbol{a}$). Often they are just called k^{th} *order partials*, or k^{th} *order partial derivatives* (the "at $\boldsymbol{a}$" being added when needed). One often writes $D_{i_1\ldots i_k}$ or $\frac{\partial^k}{\partial x_{i_1}\ldots\partial x_{i_k}}$ for $D_{i_1}\circ\ldots\circ D_{i_k}$. Thus we have

$$D_{i_1}\circ\ldots\circ D_{i_k} = D_{i_1\ldots i_k} = \frac{\partial^k}{\partial x_{i_1}\ldots\partial x_{i_k}}.$$

At a point $\boldsymbol{a}$ the following notations are also common (when the left most iterated partial derivative exists at $\boldsymbol{a}$)

$$D_{i_1}\circ\ldots\circ D_{i_k}\boldsymbol{f}(\boldsymbol{a}) = D_{i_1\ldots i_k}\boldsymbol{f}(\boldsymbol{a}) = \frac{\partial^k\boldsymbol{f}}{\partial x_{i_1}\ldots\partial x_{i_k}}(\boldsymbol{a}) = \left.\frac{\partial^k\boldsymbol{f}}{\partial x_{i_1}\ldots\partial x_{i_k}}\right|_{\boldsymbol{x}=\boldsymbol{a}}.$$

Definition 3.1.1. Let U be an open subset of $\mathbf{R}^n$, and $\boldsymbol{f}\colon U\to\mathbf{R}^m$ a map and $k\ge 1$. We say $\boldsymbol{f}$ is k*-times differentiable on* U if $D^k\boldsymbol{f}$ exists on U. We say $\boldsymbol{f}\in\mathscr{C}^k$, or sometimes $\boldsymbol{f}\in\mathscr{C}^k(U)$ if $\boldsymbol{f}$ is k-times differentiable and $D^k\boldsymbol{f}$ is continuous on U.

Further obvious terminology is as follows. With $\boldsymbol{f}$ as above and $\boldsymbol{a}\in U$, $\boldsymbol{f}$ is 1*-times differentiable at* $\boldsymbol{a}$ if it is differentiable at $\boldsymbol{a}$. For $k>1$, we say $\boldsymbol{f}$ is k*-times differentiable at* $\boldsymbol{a}$ if it is $k-1$ times differentiable in an open neighbourhood of $\boldsymbol{a}$ and $D^{k-1}\boldsymbol{f}$ is differentiable at $\boldsymbol{a}$.

About these notes. These course notes are a reasonably faithful record of the lectures given at the Chennai Mathematical Institute (CMI) in the January-April semester of 2019-20. The course is Analysis II, a core course for first year undergraduates at CMI. For more material related to this course, visit `https://www.cmi.ac.in/~pramath/teaching.html#ANA2`. If you have comments on these notes, or on related course material, please send an email to pramath@cmi.ac.in.

References

[R] W. Rudin, *Principles of Mathematical Analysis*, (Third Edition), McGraw-Hill, New Delhi, 1976.