MA 224 (Real Analysis)

Summary of Lectures - 4

Motivation for the definition of differentiability in $\mathbb{R}^n$: A result due to Frobenius implies that if $n > 2$, then with the usual addition, no multiplication can be defined in $\mathbb{R}^n$ to make $\mathbb{R}^n$ a field. Hence the difference quotient approach to define differentiability/derivative for functions of several real variables won't be meaningful. (Of course, for $n = 2$, we have two notions of differentiability - one will be considered here and the other one is the usual complex derivative which uses the field structure of $\mathbb{R}^2$.)
The alternative approach in case of single real variable is to consider approximation of the graph of a function $f : \mathbb{R} \to \mathbb{R}$ at a point $(x_0, f(x_0)) \in \mathbb{R}^2$ by the tangent line at $(x_0, f(x_0))$. If f is continuous at x_0, then any straight line passing through $(x_0, f(x_0))$ approximates the graph of f at $(x_0, f(x_0))$ in the sense that for each $m \in \mathbb{R}$, $f(x) - f(x_0) - m(x - x_0) \to 0$ as $x \to x_0$. The differentiability of f at x_0 is equivalent to the existence of the unique best possible such approximation (*i.e.* unique $m \in \mathbb{R}$) in the sense that $\lim\limits_{x\to x_0} \frac{|f(x)-f(x_0)-m(x-x_0)|}{|x-x_0|} = 0$. Given any $m \in \mathbb{R}$, the function $x \mapsto mx$ from $\mathbb{R}$ to $\mathbb{R}$ is a linear function (as defined below).

Linear function: A function $L : \mathbb{R}^n \to \mathbb{R}^m$ is called linear if

(a) $L(\mathbf{x} + \mathbf{y}) = L(\mathbf{x}) + L(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, and
(b) $L(\alpha\mathbf{x}) = \alpha L(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$ and for all $\alpha \in \mathbb{R}$.

For example, if $L_1(x, y) = (x - 2y, 0, x + y)$ and $L_2(x, y) = (2x + 1, y)$ for all $(x, y) \in \mathbb{R}^2$, then $L_1 : \mathbb{R}^2 \to \mathbb{R}^3$ is linear whereas $L_2 : \mathbb{R}^2 \to \mathbb{R}^2$ is not linear.

Some facts on linear functions:

(a) If $L : \mathbb{R}^n \to \mathbb{R}^m$ is linear, then $L(\mathbf{0}) = \mathbf{0}$ and there exists $M > 0$ such that $\|L(\mathbf{x})\|_2^{\mathbb{R}^m} \le M\|\mathbf{x}\|_2^{\mathbb{R}^n}$ for all $\mathbf{x} \in \mathbb{R}^n$. Consequently every linear function from $\mathbb{R}^n$ to $\mathbb{R}^m$ is continuous.
(b) A function $L : \mathbb{R}^n \to \mathbb{R}$ is linear iff there exists $\mathbf{a} = (a_1, ..., a_n) \in \mathbb{R}^n$ such that $L(\mathbf{x}) = \langle \mathbf{a}, \mathbf{x} \rangle = \sum\limits_{i=1}^{n} a_i x_i$ for all $\mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^n$.
(c) A function $L : \mathbb{R}^n \to \mathbb{R}^m$ is linear iff there exist linear functions $L_j : \mathbb{R}^n \to \mathbb{R}$ for $j = 1, ..., m$ such that $L(\mathbf{x}) = (L_1(\mathbf{x}), ..., L_m(\mathbf{x}))$ for all $\mathbf{x} \in \mathbb{R}^n$.

Interior point: If $\Omega \subset \mathbb{R}^n$, then a point $\mathbf{x_0} \in \Omega$ is called an interior point of Ω (denoted by $\mathbf{x_0} \in \Omega^0$) if there exists $\delta > 0$ such that $B_\delta(\mathbf{x_0}) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{x_0}\|_2 < \delta\} \subset \Omega$.
For example, if $\Omega = \{(x, y) \in \mathbb{R}^2 : x + y \ge 2\} \subset \mathbb{R}^2$, then $(2, 1) \in \Omega^0$ but $(1, 1) \notin \Omega^0$.

Differentiability and derivative in $\mathbb{R}^n$: A function $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$ is said to be differentiable at $\mathbf{x_0} \in \Omega^0$ if there exists a linear function $L_{\mathbf{x_0}} : \mathbb{R}^n \to \mathbb{R}^m$ such that
$\lim\limits_{\|\mathbf{h}\|_2^{\mathbb{R}^n}\to 0} \frac{\|f(\mathbf{x_0}+\mathbf{h})-f(\mathbf{x_0})-L_{\mathbf{x_0}}(\mathbf{h})\|_2^{\mathbb{R}^m}}{\|\mathbf{h}\|_2^{\mathbb{R}^n}} = 0.$
If f is differentiable at $\mathbf{x_0}$, then $L_{\mathbf{x_0}}$ mentioned above is unique and $L_{\mathbf{x_0}}$ is called the derivative of f at $\mathbf{x_0}$, which we denote by $f'(\mathbf{x_0})$ (or, $Df(\mathbf{x_0})$) $= L_{\mathbf{x_0}}$.
If Ω is a nonempty open subset of $\mathbb{R}^n$, then a function $f : \Omega \to \mathbb{R}^m$ is said to be differentiable if f is differentiable at each $\mathbf{x_0} \in \Omega$.

Remarks: If $f(x) = x^2$ for all $x \in \mathbb{R}$, then according to the usual single-variable calculus notion, $f'(3) = 6$ (a real number) whereas according to the above definition $f'(3)$ is a linear function from $\mathbb{R}$ to $\mathbb{R}$, defined by $f'(3)(x) = 6x$ for all $x \in \mathbb{R}$. Both the interpretations can be seen to be equivalent and the advantage of the second interpretation is that it can be easily generalized (as given in the above definition) to vector-valued functions of several real variables.

Examples:

(a) Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a constant function, so that there exists $\mathbf{y}_0 \in \mathbb{R}^m$ satisfying $f(\mathbf{x}) = \mathbf{y}_0$ for all $\mathbf{x} \in \mathbb{R}^n$. Then f is differentiable and $f'(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^m$ is given by $f'(\mathbf{x})(\mathbf{h}) = \mathbf{0}$ for all $\mathbf{x}, \mathbf{h} \in \mathbb{R}^n$.

(b) Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a linear function. Then f is differentiable and $f'(\mathbf{x}) = f$ for all $\mathbf{x} \in \mathbb{R}^n$.

(c) Let $f(\mathbf{x}) = \|\mathbf{x}\|_2^2$ for all $\mathbf{x} \in \mathbb{R}^n$. Then $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable and $f'(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ is given by $f'(\mathbf{x})(\mathbf{h}) = 2\mathbf{x} \cdot \mathbf{h}$ for all $\mathbf{x}, \mathbf{h} \in \mathbb{R}^n$.

(d) Let $f : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^m$ be a bilinear function, *i.e.* for each $\mathbf{y} \in \mathbb{R}^p$, the function $\mathbf{x} \mapsto f(\mathbf{x}, \mathbf{y})$ from $\mathbb{R}^n$ to $\mathbb{R}^m$ is linear and for each $\mathbf{x} \in \mathbb{R}^n$, the function $\mathbf{y} \mapsto f(\mathbf{x}, \mathbf{y})$ from $\mathbb{R}^p$ to $\mathbb{R}^m$ is linear. Then there exists $M > 0$ such that $\|f(\mathbf{x}, \mathbf{y})\|_2^{\mathbb{R}^m} \le M \|\mathbf{x}\|_2^{\mathbb{R}^n} \|\mathbf{y}\|_2^{\mathbb{R}^p}$ for all $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^p$.
It follows that f is differentiable and $f'(\mathbf{x}, \mathbf{y}) : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^m$ is given by $f'(\mathbf{x}, \mathbf{y})(\mathbf{h}, \mathbf{k}) = f(\mathbf{x}, \mathbf{k}) + f(\mathbf{h}, \mathbf{y})$ for all $\mathbf{x}, \mathbf{h} \in \mathbb{R}^n$ and for all $\mathbf{y}, \mathbf{k} \in \mathbb{R}^p$.

Two questions: Let $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$ and let $\mathbf{x}_0 \in \Omega^0$.

(a) How do we examine the differentiability of f at $\mathbf{x}_0$?

(b) Given that f is differentiable at $\mathbf{x}_0$, how do we find $f'(\mathbf{x}_0)$?

Although the definitions of differentiability and derivative can be used to answer these questions directly (as we have done in the above examples), they are difficult to apply in most cases. Hence we provide alternative answers below.

Partial derivative and directional derivative: Let $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$, let $\mathbf{x}_0 \in \Omega^0$ and let $i \in \{1, ..., n\}$. The partial derivative of f with respect to x_i at $\mathbf{x}_0$ is defined as $\frac{\partial f}{\partial x_i}|_{\mathbf{x}_0} = \lim\limits_{t \to 0} \frac{f(\mathbf{x}_0 + t\mathbf{e}_i) - f(\mathbf{x}_0)}{t}$, provided this limit exists (in $\mathbb{R}$).
If $\mathbf{v} (\neq \mathbf{0}) \in \mathbb{R}^n$, then the directional derivative of f along $\mathbf{v}$ at $\mathbf{x}_0$ is defined as $f'_{\mathbf{v}}(\mathbf{x}_0) = \lim\limits_{t \to 0} \frac{f(\mathbf{x}_0 + t\mathbf{v}) - f(\mathbf{x}_0)}{t}$, provided this limit exists (in $\mathbb{R}$).
We note that partial derivatives are particular cases of directional derivatives.

Remarks: Existence (in $\mathbb{R}$) of all the partial derivatives of f at $\mathbf{x}_0$ does not imply the existence (in $\mathbb{R}$) of other directional derivatives of f at $\mathbf{x}_0$. For example, for the function $f : \mathbb{R}^2 \to \mathbb{R}$, defined by $f(x, y) = \begin{cases} 0 & \text{if } xy = 0, \\ 1 & \text{if } xy \neq 0, \end{cases}$
$f_x(0,0) = f_y(0,0) = 0$ but $f'_{\mathbf{v}}(0,0)$ does not exist (in $\mathbb{R}$) for any $\mathbf{v} = (a, b) \in \mathbb{R}^2$ with $ab \neq 0$.

Answer to question (b): If f is differentiable at $\mathbf{x}_0$ and if $f_j = \pi_j \circ f$ for $j = 1, ..., m$, then each of the partial derivatives $\frac{\partial f_j}{\partial x_i}|_{\mathbf{x}_0}$ $(i = 1, ..., n; j = 1, ..., m)$ exists (in $\mathbb{R}$) and the matrix of the linear function $f'(\mathbf{x}_0) : \mathbb{R}^n \to \mathbb{R}^m$ with respect to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ (called the Jacobian matrix of f at $\mathbf{x}_0$) is given by $[f'(\mathbf{x}_0)] = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}|_{\mathbf{x}_0} & \cdots & \frac{\partial f_1}{\partial x_n}|_{\mathbf{x}_0} \\ \vdots & \vdots & \vdots \\ \frac{\partial f_m}{\partial x_1}|_{\mathbf{x}_0} & \cdots & \frac{\partial f_m}{\partial x_n}|_{\mathbf{x}_0} \end{bmatrix}$.

Answer to question (a): This question is answered with the help of several results listed below.

Result: If f is differentiable at $\mathbf{x}_0$, then f is continuous at $\mathbf{x}_0$.
Hence, if f is not continuous at $\mathbf{x}_0$, then f cannot be differentiable at $\mathbf{x}_0$.
For example, the function $f : \mathbb{R}^2 \to \mathbb{R}$, defined by $f(x, y) = \begin{cases} \frac{xy}{x^2 + y^2} & \text{if } (x, y) \neq (0,0), \\ (0,0) & \text{if } (x, y) = (0,0), \end{cases}$
is not continuous at $(0,0)$ and hence it is not differentiable at $(0,0)$.

Remarks: The converse of the above result is not true, in general. For example, if $f(x, y) = |x| + |y|$ for all $(x, y) \in \mathbb{R}^2$, then $f : \mathbb{R}^2 \to \mathbb{R}$ is continuous at $(0,0)$ but not differentiable at $(0,0)$. In fact, this example also shows that continuity of f at $\mathbf{x}_0$ does not imply even the existence (in $\mathbb{R}$) of partial derivatives of f at $\mathbf{x}_0$.

Result: If $f_j = \pi_j \circ f$ for $j = 1, \ldots, m$, then f is differentiable at $\mathbf{x}_0$ iff for each $j \in \{1, ..., m\}$,

$f_j : \Omega \to \mathbb{R}$ is differentiable at $\mathbf{x}_0$.

In view of this result, from now onwards we may consider the case $m = 1$ only.

Result: If f is differentiable at $\mathbf{x}_0$, then for each $\mathbf{v}(\neq \mathbf{0}) \in \mathbb{R}^n$, the directional derivative $f'_{\mathbf{v}}(\mathbf{x}_0)$ exists (in $\mathbb{R}$) and $f'_{\mathbf{v}}(\mathbf{x}_0) = f'(\mathbf{x}_0)(\mathbf{v})$.
In particular, for each $i \in \{1, ..., n\}$, the partial derivative $\frac{\partial f}{\partial x_i}|_{\mathbf{x}_0}$ exists (in $\mathbb{R}$) and $\frac{\partial f}{\partial x_i}|_{\mathbf{x}_0} = f'(\mathbf{x}_0)(\mathbf{e}_i)$.

Remarks: The above result implies that if at least one of the directional derivatives of f at $\mathbf{x}_0$ does not exist (in $\mathbb{R}$), then f cannot be differentiable at $\mathbf{x}_0$.
However, existence (in $\mathbb{R}$) of all the directional derivatives of f at $\mathbf{x}_0$ does not ensure the differentiability (in fact, even the continuity) of f at $\mathbf{x}_0$. For example, consider $f : \mathbb{R}^2 \to \mathbb{R}$, defined by
$f(x, y) = \begin{cases} \frac{xy^2}{x^2+y^4} & \text{if } (x, y) \neq (0, 0), \\ 0 & \text{if } (x, y) = (0, 0). \end{cases}$
If $\mathbf{v} = (a, b)(\neq (0, 0)) \in \mathbb{R}^2$, then $f'_{\mathbf{v}}(0, 0) = \begin{cases} \frac{b^2}{a} & \text{if } a \neq 0, \\ 0 & \text{if } a = 0, \end{cases}$
but f is not continuous at $(0, 0)$ and hence f is not differentiable at $(0, 0)$.

Result: If for each $i \in \{1, ..., n\}$, $\frac{\partial f}{\partial x_i}$ exists in a neighbourhood of $\mathbf{x}_0$ and is continuous at $\mathbf{x}_0$, then f is differentiable at $\mathbf{x}_0$.
In fact, if $n - 1$ of the partial derivatives $\frac{\partial f}{\partial x_i}$ $(i = 1, ..., n)$ exist in a neighbourhood of $\mathbf{x}_0$ and are continuous at $\mathbf{x}_0$ and the remaining one partial derivative exists at $\mathbf{x}_0$, then also f is differentiable at $\mathbf{x}_0$.

Examples:

(a) The function $f : \mathbb{R}^2 \to \mathbb{R}^3$, defined by $f(x, y) = (x^2 + y^3, x^3 + y^2, 2x^2y^2)$ for all $(x, y) \in \mathbb{R}^2$, is differentiable at $(1, 2)$ and $[f'(1, 2)] = \begin{bmatrix} 2 & 12 \\ 3 & 4 \\ 16 & 8 \end{bmatrix}$.

(b) The function $f : \mathbb{R}^2 \to \mathbb{R}$, defined by $f(x, y) = \begin{cases} x^2 \cos \frac{1}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}$
is differentiable at $(0, 0)$ and $[f'(0, 0)] = [0 \ \ 0]$.
Note that in this example, the first statement of the preceding result is not applicable but the second statement there is applicable in concluding the differentiability of f at $(0, 0)$.

Remarks: f can be differentiable at $\mathbf{x}_0$ even though none of the partial derivatives of f is continuous at $\mathbf{x}_0$. For example, consider $f : \mathbb{R}^2 \to \mathbb{R}$, defined by
$f(x, y) = \begin{cases} (x^2 + y^2) \sin\left(\frac{1}{x^2+y^2}\right) & \text{if } (x, y) \neq (0, 0), \\ 0 & \text{if } (x, y) = (0, 0). \end{cases}$

Result: Let $\frac{\partial f}{\partial x_i}|_{\mathbf{x}_0}$ exist for each $i \in \{1, ..., n\}$. Then f is differentiable at $\mathbf{x}_0$ iff
$\lim\limits_{\|\mathbf{h}\|_2 \to 0} \frac{|f(\mathbf{x}_0+\mathbf{h}) - f(\mathbf{x}_0) - \alpha \cdot \mathbf{h}|}{\|\mathbf{h}\|_2} = 0$, where $\alpha = \left(\frac{\partial f}{\partial x_1}|_{\mathbf{x}_0}, ..., \frac{\partial f}{\partial x_n}|_{\mathbf{x}_0}\right)$.

Examples:

(a) The function $f : \mathbb{R}^2 \to \mathbb{R}$, defined by $f(x, y) = \begin{cases} xy\frac{x^2-y^2}{x^2+y^2} & \text{if } (x, y) \neq (0, 0), \\ 0 & \text{if } (x, y) = (0, 0), \end{cases}$
is differentiable at $(0, 0)$ and $[f'(0, 0)] = [0 \ \ 0]$.

(b) The function $f : \mathbb{R}^2 \to \mathbb{R}$, defined by $f(x, y) = \begin{cases} \frac{xy}{\sqrt{x^2+y^2}} & \text{if } (x, y) \neq (0, 0), \\ 0 & \text{if } (x, y) = (0, 0), \end{cases}$
is not differentiable at $(0, 0)$.

Gradient: If $\frac{\partial f}{\partial x_i}|_{\mathbf{x}_0}$ exists for each $i \in \{1, ..., n\}$, then $\left(\frac{\partial f}{\partial x_1}|_{\mathbf{x}_0}, ..., \frac{\partial f}{\partial x_n}|_{\mathbf{x}_0}\right)$ is called the gradient (vector) of f at $\mathbf{x}_0$ and it is denoted by $\nabla f(\mathbf{x}_0)$ or $\operatorname{grad} f(\mathbf{x}_0)$.
If f is differentiable at $\mathbf{x}_0$ and $\mathbf{u}(\neq \mathbf{0}) \in \mathbb{R}^n$, then $f'_{\mathbf{u}}(\mathbf{x}_0) = \langle \nabla f(\mathbf{x}_0), \mathbf{u} \rangle$ and as a consequence of the (equality condition of) Cauchy-Schwarz inequality, it follows that the function $\mathbf{u} \mapsto f'_{\mathbf{u}}(\mathbf{x}_0)$ attains its maximum value when $\mathbf{u}$ is a positive scalar multiple of $\nabla f(\mathbf{x}_0)$.

Chain rule: Let $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$ and let $g : \Omega' \subset \mathbb{R}^m \to \mathbb{R}^k$ such that $f(\Omega) \subset \Omega'$. If f is differentiable at $\mathbf{x}_0 \in \Omega^0$ and if g is differentiable at $f(\mathbf{x}_0) \in \Omega'^0$, then $g \circ f : \Omega \to \mathbb{R}^k$ is differentiable at $\mathbf{x}_0$ and $(g \circ f)'(\mathbf{x}_0) = g'(f(\mathbf{x}_0)) \circ f'(\mathbf{x}_0)$ (and consequently $[(g \circ f)'(\mathbf{x}_0)] = [g'(f(\mathbf{x}_0))] \cdot [f'(\mathbf{x}_0)]$).

Illustration: Let w be a differentiable function of u and v where each of u and v is a differentiable function of x and y. Then w is a differentiable function of x and y and

$$\frac{\partial w}{\partial x} = \frac{\partial w}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial w}{\partial v}\frac{\partial v}{\partial x},$$
$$\frac{\partial w}{\partial y} = \frac{\partial w}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial w}{\partial v}\frac{\partial v}{\partial y}.$$

Applications:

(a) If $f(\mathbf{x}) = \|\mathbf{x}\|_2$ for all $\mathbf{x} \in \mathbb{R}^n$, then $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at each $\mathbf{x}_0(\neq \mathbf{0}) \in \mathbb{R}^n$ and $f'(\mathbf{x}_0)(\mathbf{h}) = \frac{\mathbf{x}_0 \cdot \mathbf{h}}{\|\mathbf{x}_0\|_2}$ for all $\mathbf{h} \in \mathbb{R}^n$.
(b) Let Ω be an open subset of $\mathbb{R}^n$ and let both $f : \Omega \to \mathbb{R}$ and $g : \Omega \to \mathbb{R}$ be differentiable at $\mathbf{x}_0 \in \Omega$. If $(fg)(\mathbf{x}) = f(\mathbf{x})g(\mathbf{x})$ for all $\mathbf{x} \in \Omega$, then $fg : \Omega \to \mathbb{R}$ is differentiable at $\mathbf{x}_0$ and $(fg)'(\mathbf{x}_0) = g(\mathbf{x}_0)f'(\mathbf{x}_0) + f(\mathbf{x}_0)g'(\mathbf{x}_0)$.

Mean value theorem for real-valued functions: Let Ω be an open subset of $\mathbb{R}^n$ such that $\mathbf{a}, \mathbf{b} \in \Omega$ and $S = \{(1-t)\mathbf{a} + t\mathbf{b} : t \in [0, 1]\} \subset \Omega$. If $f : \Omega \to \mathbb{R}$ is differentiable at each point of S, then there exists $\mathbf{c} \in S$ such that $f(\mathbf{b}) - f(\mathbf{a}) = f'(\mathbf{c})(\mathbf{b} - \mathbf{a})$.

Remarks: A natural generalization of the mean value theorem need not hold if the codomain of the function is $\mathbb{R}^m$, where $m > 1$ (even if the domain is a subset of $\mathbb{R}$). For example, if $f(x) = (\cos x, \sin x)$ for all $x \in \mathbb{R}$, then there cannot exist any $c \in [0, 2\pi]$ such that
$f(2\pi) - f(0) = f'(c)(2\pi)$.

Mean value theorem for vector-valued functions: Let Ω be an open subset of $\mathbb{R}^n$ such that $\mathbf{a}, \mathbf{b} \in \Omega$ and $S = \{(1-t)\mathbf{a} + t\mathbf{b} : t \in [0, 1]\} \subset \Omega$. If $f : \Omega \to \mathbb{R}^m$ is differentiable at each point of S, then for each $\mathbf{y} \in \mathbb{R}^m$, there exists $\mathbf{c} \in S$ such that $\langle f(\mathbf{b}) - f(\mathbf{a}), \mathbf{y} \rangle = \langle f'(\mathbf{c})(\mathbf{b} - \mathbf{a}), \mathbf{y} \rangle$.

Mean value inequality: Let Ω be an open subset of $\mathbb{R}^n$ such that $\mathbf{a}, \mathbf{b} \in \Omega$ and $S = \{(1-t)\mathbf{a} + t\mathbf{b} : t \in [0, 1]\} \subset \Omega$. If $f : \Omega \to \mathbb{R}^m$ is differentiable at each point of S, then there exists $\mathbf{c} \in S$ such that $\|f(\mathbf{b}) - f(\mathbf{a})\|_2 \leq \|f'(\mathbf{c})(\mathbf{b} - \mathbf{a})\|_2$.

Consequence of mean value inequality: Let Ω be an open convex subset of $\mathbb{R}^n$. If $f : \Omega \to \mathbb{R}^m$ is differentiable and if $f'(\mathbf{x}) = \mathbf{0}$ for all $\mathbf{x} \in \Omega$, then f is a constant function.
Instead of assuming Ω to be an open convex subset of $\mathbb{R}^n$, if we assume that Ω is an open connected subset of $\mathbb{R}^n$ (which is a weaker condition), then also the conclusion of the above result remains true.

Continuously differentiable function: Let Ω be an open subset of $\mathbb{R}^n$. A function $f : \Omega \to \mathbb{R}^m$ is said to be continuously differentiable if all the partial derivatives of all the component functions of f are continuous on Ω.

The class of all continuously differentiable functions from Ω to $\mathbb{R}^m$ is denoted by $\mathcal{C}^1(\Omega, \mathbb{R}^m)$.

Inverse function theorem for single variable: Let I be an open interval of $\mathbb{R}$ and let $f \in \mathcal{C}^1(I, \mathbb{R})$ such that $f'(x) \neq 0$ for all $x \in I$. Then

(a) f is one-one,
(b) $f(I)$ is an open interval of $\mathbb{R}$, and
(c) $f^{-1} : f(I) \to I$ is differentiable with $(f^{-1})'(y) = \frac{1}{f'(x)}$ for all $y = f(x)$ with $x \in I$.

Remarks: A direct generalization of the above theorem for several variables is not true. For example, let $f(x, y) = (x \cos y, x \sin y)$ for all $(x, y) \in \Omega = (0, 1) \times (0, a)$, where $a > 2\pi$. Then $f \in \mathcal{C}^1(\Omega, \mathbb{R}^2)$ and $f'(x, y)$ is invertible for all $(x, y) \in \Omega$ but f is not one-one.

Inverse function theorem: Let $f \in \mathcal{C}^1(\Omega, \mathbb{R}^n)$, where Ω is an open subset of $\mathbb{R}^n$. If $\mathbf{x}_0 \in \Omega$ such that $f'(\mathbf{x}_0)$ is invertible (*i.e.* $\det[f'(\mathbf{x}_0)] \neq 0$), then there exists an open subset U of Ω containing $\mathbf{x}_0$ such that

(a) f is one-one on U,
(b) $f(U)$ is an open subset of $\mathbb{R}^n$, and
(c) $f^{-1} : f(U) \to U$ is differentiable with $(f^{-1})'(\mathbf{v}) = (f'(\mathbf{u}))^{-1}$ for all $\mathbf{v} = f(\mathbf{u})$ with $\mathbf{u} \in U$.

Examples:

(a) The set of points of $\mathbb{R}^2$, where the function $f : \mathbb{R}^2 \to \mathbb{R}^2$, defined by $f(x, y) = (x^2 - y^2, xy)$ for all $(x, y) \in \mathbb{R}^2$, is locally invertible, is $\mathbb{R}^2 \setminus \{(0, 0)\}$.
(b) The set of points of $\mathbb{R}^3$, where the function $f : \mathbb{R}^3 \to \mathbb{R}^3$, defined by $f(x, y, z) = (x + y + z, xy + yz + zx, xyz)$ for all $(x, y, z) \in \mathbb{R}^3$, is locally invertible, is $\{(x, y, z) \in \mathbb{R}^3 : x \neq y, y \neq z, z \neq x\}$.

Remarks:

(a) The reason why we have not considered $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$ with $m \neq n$ in the inverse function theorem is provided by the following result:
Let U and V be open subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively and let $f : U \to \mathbb{R}^m$ and $g : V \to \mathbb{R}^n$ be such that $g \circ f = I_U$ and $f \circ g = I_V$. If f is differentiable at $\mathbf{x}_0 \in U$ and g is differentiable at $f(\mathbf{x}_0)$, then $m = n$.
(b) The reason why we need continuously differentiable function instead of just differentiable function in the inverse function theorem is provided by the following example:
The function $f : \mathbb{R} \to \mathbb{R}$, defined by $f(x) = \begin{cases} x + 2x^2 \sin \frac{1}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}$
is not one-one in any neighbourhood of 0 although $f'(0) = 1 \neq 0$.

Implicit function theorem: Let $f \in \mathcal{C}^1(\Omega, \mathbb{R}^m)$, where Ω is an open subset of $\mathbb{R}^n \times \mathbb{R}^m$. If $(\mathbf{a}, \mathbf{b}) \in \Omega$ such that $f(\mathbf{a}, \mathbf{b}) = \mathbf{0}$ and $\det[\frac{\partial f}{\partial \mathbf{y}}(\mathbf{a}, \mathbf{b})] \neq 0$, then there exist an open subset U of $\mathbb{R}^n$ containing $\mathbf{a}$, an open subset W of $\mathbb{R}^n \times \mathbb{R}^m$ containing $(\mathbf{a}, \mathbf{b})$ and a unique continuous function $g : U \to \mathbb{R}^m$ such that $g(\mathbf{a}) = \mathbf{b}$ and $\{(\mathbf{x}, \mathbf{y}) \in W : f(\mathbf{x}, \mathbf{y}) = \mathbf{0}\} = \{(\mathbf{x}, g(\mathbf{x}) : \mathbf{x} \in U\}$. Further, $g \in \mathcal{C}^1(U, \mathbb{R}^m)$ and $[g'(\mathbf{x})] = -[\frac{\partial f}{\partial \mathbf{y}}(\mathbf{x}, g(\mathbf{x}))]^{-1}[\frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}, g(\mathbf{x}))]$ for all $\mathbf{x} \in U$.

Examples:

(a) In a neighbourhood of any point $(x_0, y_0, u_0, v_0) \in \mathbb{R}^4$ which satisfies the equations $x - e^u \cos v = 0$ and $v - e^y \sin x = 0$, there exists a unique solution $(u, v) = \varphi(x, y)$ satisfying $\det[\varphi'(x, y)] = v/x$.
(b) Around the point $(0, 1, 1)$, the equation $xy - z \log y + e^{xz} = 1$ cannot be solved locally as $z = g(x, y)$.
(c) Around the point $(0, 0)$, the equation $x^2 - y^3 = 0$ defines y as a continuous but non-differentiable function of x.

(d) Around the point $(0,0)$, the equation $y^2 - x^4 = 0$ defines y as a (in fact, two) differentiable function of x.

Higher order partial derivatives: Let Ω be an open subset of $\mathbb{R}^n$ and let $f : \Omega \to \mathbb{R}$. The partial derivatives of f may exist on Ω and may have further partial derivatives. The second order partial derivative $\frac{\partial}{\partial x_i}(\frac{\partial f}{\partial x_j})$ is denoted by $\frac{\partial^2 f}{\partial x_i \partial x_j}$ or $f_{x_j x_i}$ if $i \neq j$ and $\frac{\partial^2 f}{\partial x_j^2}$ or $f_{x_j x_j}$ if $i = j$. The analogues of these notations for higher order partial derivatives should be clear.

Example: For the function $f : \mathbb{R}^2 \to \mathbb{R}$, defined by $f(x,y) = \begin{cases} xy\frac{x^2-y^2}{x^2+y^2} & \text{if } (x,y) \neq (0,0), \\ 0 & \text{if } (x,y) = (0,0), \end{cases}$

$f_{xy}(0,0) = -1 \neq 1 = f_{yx}(0,0)$.

Result: Let $f : \Omega \to \mathbb{R}$, where Ω is an open subset of $\mathbb{R}^n$, let $\mathbf{a} \in \Omega$ and let $i, j \in \{1, ..., n\}$. If the partial derivatives $f_{x_i x_j}$ and $f_{x_j x_i}$ exist on Ω and if $f_{x_i x_j}$ and $f_{x_j x_i}$ are continuous at $\mathbf{a}$, then $f_{x_i x_j}(\mathbf{a}) = f_{x_j x_i}(\mathbf{a})$.

Multi-index notation: A multi-index $\alpha = (\alpha_1, ..., \alpha_n)$ is an n-tuple (for some $n \in \mathbb{N}$) of non-negative integers α_i.
For a multi-index $\alpha = (\alpha_1, ..., \alpha_n)$, we define $|\alpha| = \alpha_1 + \cdots + \alpha_n$, $\alpha! = \alpha_1! \cdots \alpha_n!$, $\mathbf{x}^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ (where $\mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^n$) and $\partial^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}$ (where f is a real-valued function of n real variables).
For example, with $n = 3$ and $\mathbf{x} = (x, y, z)$, we have $\mathbf{x}^{(3,1,4)} = x^3 y z^4$ and $\partial^{(0,2,1)} f = \frac{\partial^3 f}{\partial y^2 \partial z}$.
Using multi-indices, the multinomial theorem can be stated as follows: If $\mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^n$ and if $k \in \mathbb{N}$, then $(x_1 + \cdots + x_n)^k = \sum\limits_{|\alpha|=k} \frac{k!}{\alpha!}\mathbf{x}^\alpha$.

Taylor's theorem for several variables: Let $f \in \mathcal{C}^{k+1}(\Omega, \mathbb{R})$, where Ω is an open subset of $\mathbb{R}^n$. If $\mathbf{a}, \mathbf{a} + \mathbf{h} \in \Omega$ and if $S = \{\mathbf{a} + t\mathbf{h} : t \in [0,1]\} \subset \Omega$, then there exists $\theta \in (0,1)$ such that $f(\mathbf{a}+\mathbf{h}) = \sum\limits_{|\alpha| \leq k} \partial^\alpha f(\mathbf{a}) \frac{\mathbf{h}^\alpha}{\alpha!} + \sum\limits_{|\alpha|=k+1} \partial^\alpha f(\mathbf{a} + \theta\mathbf{h}) \frac{\mathbf{h}^\alpha}{\alpha!}$.
In the above equation, the first sum is called the kth order Taylor polynomial of f at $\mathbf{a}$ and the second sum is called the kth order remainder of f at $\mathbf{a}$.

If one wants to find the Taylor polynomial of some order of a function f about a point $\mathbf{a}$, then it may be very tedious to find all the partial derivatives needed in the above formula. The following result helps us to overcome this difficulty.

Result: Let $f \in \mathcal{C}^{k+1}(B_\delta(\mathbf{a}), \mathbb{R})$ for some $\mathbf{a} \in \mathbb{R}^n$ and $\delta > 0$. If $f(\mathbf{a} + \mathbf{h}) = P(\mathbf{h}) + E(\mathbf{h})$ for all $\mathbf{h} \in \mathbb{R}^n$ with $\|\mathbf{h}\|_2 < \delta$, where P is a polynomial of degree $\leq k$ and $\lim\limits_{\|\mathbf{h}\|_2 \to 0} \frac{E(\mathbf{h})}{\|\mathbf{h}\|_2^k} = 0$, then P is the kth order Taylor polynomial of f about $\mathbf{a}$.

Example: The 3rd order Taylor polynomial of $f(x,y) = e^{x^2+y}$ about $(0,0)$ is given by
$P(x,y) = 1 + y + x^2 + \frac{1}{2}y^2 + x^2 y + \frac{1}{6}y^3$.