9. Differential Calculus

Definition 9.1

Let V and W be two Banach spaces, $D \subseteq V$ be open in V and let $a \in D$. A function $f : D \longrightarrow W$ is called differentiable at a if there exists a continuous linear transformation $T \in \mathcal{L}(V,W)$ such that

$$(*) \qquad \lim_{x \to a} \frac{f(x) - f(a) - T(x-a)}{\|x-a\|} = 0$$

Remark 9.2

In the hypothesis of the above definition, for a function $f : D \longrightarrow W$ the following are equivalent:

(i) f is differentiable at a;

(ii) there exists $T \in \mathcal{L}(V,W)$ s.t. $\displaystyle\lim_{x \to a} \frac{\| f(x) - f(a) - T(x-a)\|}{\|x-a\|} = 0$;

(iii) there exist $T \in \mathcal{L}(V,W)$ and $r : D \to W$ s.t.

(a) $f(x) = f(a) + T(x-a) + r(x)$ for all $x \in D$, and

(b) $\displaystyle\lim_{x \to a} \frac{r(x)}{\|x-a\|} = 0$;

(iv) there exist $T \in \mathcal{L}(V,W)$ and $r' : D \to W$ s.t.

(a) $f(x) = f(a) + T(x-a) + \|x-a\| \cdot r'(x)$ for all $x \in D$, and

(b) $\displaystyle\lim_{x \to a} r'(x) = 0$.

(v) there exist $T \in \mathcal{L}(V,W)$ and a map $\varphi : U_0 \longrightarrow W$, where $U_0 \subseteq V$ is a neighborhood of 0 s.t.

(a) $f(a+h) = f(a) + T(h) + \varphi(h)$ for all $h \in U_0$, and

(b) $$\lim_{h \to 0} \frac{\varphi(h)}{\|h\|} = 0;$$

(vi) there exist $T \in \mathcal{L}(V,W)$ and a map $\psi : U_0 \to W$, where U_0 is a neighborhood of 0 in V s.t.

(a) $f(a+h) = f(a) + T(h) + \|h\| \cdot \psi(h)$ for all $h \in U_0$, and

(b) $\lim_{h \to 0} \psi(h) = 0$.

Proof

Exercise. □

Proposition 8.3

If f is differentiable at a, then the continuous linear transformation satisfying $(*)$ is uniquely determined.

Proof

Assume that $S \in \mathcal{L}(V,W)$ is another function having property $(*)$. It follows that

$$\lim_{x \to a} \frac{T(x-a) - S(x-a)}{\|x-a\|} = \lim_{x \to a} \frac{\big(f(x) - f(a) - T(x-a)\big) - \big(f(x) - f(a) - S(x-a)\big)}{\|x-a\|}$$

$$= \lim_{x \to a} \frac{f(x) - f(a) - T(x-a)}{\|x-a\|} - \lim_{x \to a} \frac{f(x) - f(a) - S(x-a)}{\|x-a\|}$$

$$= 0.$$

We shall prove that $T(y) = S(y)$ for all $y \in V$. If $y = 0$, then $T(0) = S(0) = 0$. Assume now that $y \neq 0$. Since $a \in D$ and D is open in V, there exists $r > 0$ s.t. $U_r(a) \subseteq D$. Then by setting $\delta := \frac{r}{\|y\|}$ we get that for all $t \in]0, \delta[$,

$\|(a+ty) - a\| = \|ty\| = t \cdot \|y\| < \delta \cdot \|y\| = r$, so $(a+ty) \in U_r(a) \subseteq D$.

Since $\lim_{t\to 0}(a+ty)=a$, we get that

$$0=\lim_{t\to 0}\frac{T((a+ty)-a)-S((a+ty)-a)}{\|(a+ty)-a\|}=\lim_{t\to 0}\frac{T(ty)-S(ty)}{t\cdot\|y\|}=$$

$$=\lim_{t\to 0}\frac{t(T(y)-S(y))}{t\cdot\|y\|}=\lim_{t\to 0}\frac{T(y)-S(y)}{\|y\|}=\frac{T(y)-S(y)}{\|y\|}.$$

Hence, $T(y)=S(y)$. $\square$

Definition 8.4

The uniquely determined continuous linear transformation T of the Definition 8.1 is called the derivative of f at a and is denoted by $f'(a)$ or $Df(a)$.

Definition 8.5

We say that $f: D\to W$ is differentiable if f is differentiable at every point $a\in D$. In that case, the derivative f' is a map

$$f': D\longrightarrow \mathcal{L}(V,W),\qquad a\longmapsto f'(a).$$

from D into the Banach space $(\mathcal{L}(V,W),\|\cdot\|)$, where $\|\cdot\|$ is the operator norm.

We say that f is continuously differentiable or that f is of class C^1 if f' is continuous.

Remark 9.6

If V and W are finite dimensional normed spaces (and, hence, Banach spaces by Theorem 6.43), then any linear transformation $T: V \to W$ is continuous (see Proposition 6.46).

Moreover, since any two norms on a finite dimensional normed space are equivalent, it follows that the property of a function to be differentiable does not depend on V and W.

Remark 9.7

In the case $V = W = \mathbb{K}$ we get the usual definition of differentiability in $\mathbb{K}$ from Analysis I.

Proof

Use the fact that a function $T: \mathbb{K} \to \mathbb{K}$ is linear iff there exists $L \in \mathbb{K}$ s.t. $T(x) = Lx$ for all $x \in \mathbb{K}$.

$\square$

Proposition 9.8

Let V and W be Banach spaces, $D \subseteq V$ be open in V, $a \in D$ and $f: D \to W$ be differentiable at a. Then there exists a constant $c > 0$ s.t. for all x from a neighborhood of a

$$\|f(x) - f(a)\| \leq c\|x - a\|.$$

In particular, f is continuous at a.

Proof

Let us use Remark 9.2 (iv). Then there exists $r: D \to W$ s.t.

$$f(x) = f(a) + f'(a)(x-a) + \|x-a\| \, r(x) \text{ for all } x \in D \text{ and } \lim_{x\to a} r(x) = 0.$$

We get that for all $x \in D$.

$$\|f(x) - f(a)\| = \| f'(a)\cdot(x-a) + \|x-a\| \cdot r(x)\| \le \|f'(a)\cdot(x-a)\| + $$
$$+ \|x-a\| \cdot \|r(x)\| \le \left(\|f'(a)\| + \|r(x)\|\right) \|x-a\|.$$

Since $\lim_{x\to a} r(x) = 0$, there exists $\delta > 0$ s.t.

$$(\forall x \in D)\left(0 < \|x-a\| < \delta \Rightarrow \|r(x)\| < 1\right).$$

Let $U := U_\delta(a)$. Then for all $x \in U$,

$$\|f(x) - f(a)\| \le (\|f'(a)\| + 1)\cdot\|x-a\| = c\|x-a\|,$$

where $c = \|f'(a)\| + 1$.

In particular, this implies that

$$\lim_{x\to a} \|f(x) - f(a)\| \le \lim_{x\to a} c\cdot\|x-a\| = 0,$$

so $\lim_{x\to a} f(x) = f(a)$. Thus, f is continuous at a. □

Operations with differentiable functions

In this section V, W, U are Banach spaces.

Proposition 8.8

Let D be open in V, $a \in D$ and $f, g : D \to W$. Assume that f and g are differentiable at a. Then

(i) $f+g$ is differentiable at a, and

$$(f+g)'(a) = f'(a) + g'(a).$$

(ii) for all $c \in \mathbb{R}$, $c \cdot f$ is differentiable at a, and

$$(c \cdot f)'(a) = c \cdot f'(a).$$

<u>Proof</u>

Since f, g are differentiable at a, there are $r_1, r_2 : D \to W$

s.t.

$$f(x) = f(a) + f'(a)(x-a) + r_1(x), \quad \lim_{x \to a} \frac{r_1(x)}{\|x-a\|} = 0$$

$$g(x) = g(a) + g'(a)(x-a) + r_2(x), \quad \lim_{x \to a} \frac{r_2(x)}{\|x-a\|} = 0.$$

(i) Let $r := r_1 + r_2$. Then

$$(f+g)(x) = (f+g)(a) + (f'(a) + g'(a))(x-a) + r(x), \text{ and}$$

$$\lim_{x \to a} \frac{r(x)}{\|x-a\|} = \lim_{x \to a} \frac{r_1(x)}{\|x-a\|} + \lim_{x \to a} \frac{r_2(x)}{\|x-a\|} = 0.$$

Thus, $f+g$ is differentiable at a and $(f+g)'(a) = f'(a) + g'(a)$.

(ii) Let $r := c r_1$. Then

$$(cf)(x) = cf(a) + (cf'(a))(x-a) + r(x), \text{ and}$$

$$\lim_{x \to a} \frac{r(x)}{\|x-a\|} = c \cdot \lim_{x \to a} \frac{r_1(x)}{\|x-a\|} = 0.$$

<u>Proposition 8.10</u>

Let D be open in V and E be open in W. Let $f : D \to E$ and $g : E \to U$ be two functions. Assume that f is differentiable at $a \in D$ and g is differentiable at $f(a)$. Then $g \circ f$ is differentiable at a, and

$$(g \circ f)'(a) = g'(f(a)) \circ f'(a).$$

Before giving the proof, we make explicit the meaning of the above formula. Note that $f'(a): V \to W$ is a continuous linear transformation, and $g'(f(a)): W \to U$ is also a continuous linear transformation. These transformations can be composed to get a continuous linear transformation from V to U.

Proof

For brevity, we set $b = f(a)$, $T_1 = f'(a)$ and $T_2 = g'(f(a)) = g'(b)$. Hence, we have to prove that $g \circ f$ is differentiable at a, and $(g \circ f)'(a) = T_2 \circ T_1$.

We use Remark 8.2 (iii). Since f is differentiable at a and g is differentiable at b, there exist $r_1 : D \to W$ and $r_2 : E \to U$ s.t.

$$f(x) = f(a) + T_1(x-a) + r_1(x) \text{ for all } x \in D, \text{ and } \lim_{x \to a} \frac{r_1(x)}{\|x-a\|} = 0$$

$$g(y) = g(b) + T_2(y-b) + r_2(y) \text{ for all } y \in E, \text{ and } \lim_{y \to b} \frac{r_2(y)}{\|y-b\|} = 0.$$

Let $r : D \to U$, $r(x) = (g \circ f)(x) - (g \circ f)(a) - (T_2 \circ T_1)(a)$.

Thus, $(g \circ f)(x) = (g \circ f)(a) + (T_2 \circ T_1)(a) + r(x)$ for all $x \in D$. It remains to prove that $\lim_{x \to a} \frac{r(x)}{\|x-a\|} = 0$.

We have that for all $x \in D$,

$$r(x) = g(f(x)) - g(b) - T_2(T_1(a)) = \big(g(b) + T_2(f(x) - b) + r_2(f(x))\big) -$$
$$- g(b) - T_2(T_1(a)) = T_2(f(x) - b) - T_2(T_1(a)) + r_2(f(x)) =$$
$$= T_2(f(x) - b - T_1(a)) + r_2(f(x)) = T_2(r_1(x)) + r_2(f(x)).$$

It follows that

$$\frac{\|r(x)\|}{\|x-a\|} = \frac{\|T_2(r_1(x)) + r_2(f(x))\|}{\|x-a\|} \le \frac{\|T_2(r_1(x))\|}{\|x-a\|} + \frac{\|r_2(f(x))\|}{\|x-a\|}$$

$$\le \|T_2\| \cdot \frac{\|r_1(x)\|}{\|x-a\|} + \frac{\|r_2(f(x))\|}{\|x-a\|} = \|T_2\| \cdot \frac{\|r_1(x)\|}{\|x-a\|} +$$

$$+ \frac{\|r_2(f(x))\|}{\|f(x)-f(a)\|} \cdot \frac{\|f(x)-f(a)\|}{\|x-a\|}$$

$$\le \|T_2\| \cdot \frac{\|r_1(x)\|}{\|x-a\|} + c \cdot \frac{\|r_2(f(x))\|}{\|f(x)-f(a)\|}$$ for all x from a neighborhood of a, by Prop. 9.8.

Since $\lim\limits_{x\to a}\left(\|T_2\| \cdot \frac{\|r_1(x)\|}{\|x-a\|} + c \cdot \frac{\|r_2(f(x))\|}{\|f(x)-f(a)\|}\right) = \|T_2\| \cdot 0 +$

$$+ c \cdot \lim_{x\to a} \frac{\|r_2(f(x))\|}{\|f(x)-f(a)\|} \overset{y:=f(x)}{=\!=} c \cdot \lim_{y\to b} \frac{\|r_2(y)\|}{\|y-b\|} = c \cdot 0 = 0,$$

it follows that $\lim\limits_{x\to a} \frac{\|r(x)\|}{\|x-a\|} = 0$, so $\lim\limits_{x\to a} \frac{r(x)}{\|x-a\|} = 0$. $\square$

<u>Proposition 9.11</u>

Let D be open in V and E be open in W. Let $f: D \to E$ be a bijective function and $a \in D$. If f is differentiable at a with invertible derivative $f'(a)$ and if the inverse mapping $f^{-1}: E \to D$ is continuous at $b = f(a)$, then f^{-1} is differentiable at b with derivative

$$(f^{-1})'(b) = (f'(a))^{-1} = (f'(f^{-1}(b)))^{-1}.$$

Proof

Since f is differentiable at a, by Remark 8.2 (iv), there exists $r: D \to W$ s.t.

(1) $f(x) = f(a) + f'(a)(x-a) + \|x-a\| \cdot r(x)$ for all $x \in D$, and

$\lim_{x \to a} r(x) = 0$.

By taking $x = f^{-1}(y)$ in (1), we get that

(2) $y = b + f'(a)(f^{-1}(y) - f^{-1}(b)) + \|f^{-1}(y) - f^{-1}(b)\| \cdot r(f^{-1}(y))$

for all $y \in E$.

For brevity, let us set $T := (f'(a))^{-1}$. By applying T to both members of (2), it follows that

$T(y-b) = f^{-1}(y) - f^{-1}(b) + T\left(\|f^{-1}(y) - f^{-1}(b)\| \cdot r(f^{-1}(y))\right)$,

so

$f^{-1}(y) = f^{-1}(b) + T(y-b) + \|f^{-1}(y) - f^{-1}(b)\| \cdot (-(T \circ r \circ f^{-1})(y))$

for all $y \in E$.

Let $r': E \to W$, $r'(y) = \dfrac{\|f^{-1}(y) - f^{-1}(b)\|}{\|y-b\|} \cdot (-T \circ r \circ f^{-1})(y)$ if $y \neq b$

$r'(b) = 0$

Then for all $y \in E$,

$f^{-1}(y) = f^{-1}(b) + T(y-b) + \|y-b\| \cdot r'(y)$.

Hence, it remains to prove that $\lim_{y \to b} r'(y) = 0$.

First, let us remark that, since f^{-1} is continuous at $b=f(a)$, we have that $\lim_{y\to b} f^{-1}(y) = f^{-1}(b) = a$. It follows that

$\lim_{y\to b} r(f^{-1}(y)) = \lim_{x\to a} r(x) = 0$, so $\lim_{y\to b} (-T\circ r\circ f^{-1})(y) =$

$= -T(0) = 0.$

Now, we shall prove that there exists $M>0$ s.t.

$$\|f^{-1}(y) - f^{-1}(b)\| \leq M\cdot \|y-b\| \quad \text{for all } y \text{ from a neighborhood of } b.$$

It follows then that $\lim_{y\to b} r'(y) = 0$.

We have that for all $y\in E$,

$$\|f^{-1}(y)-f^{-1}(b)\| \leq \|T(y-b)\| + \|f^{-1}(y)-f^{-1}(b)\|\cdot\|-(T\circ r\circ f^{-1})(y)\|$$

$$\leq \|T\|\cdot\|y-b\| + \|f^{-1}(y)-f^{-1}(b)\|\cdot\|-(T\circ r\circ f^{-1})(y)\|.$$

Since $\lim_{y\to b}\left(-(T\circ r\circ f^{-1})(y)\right)=0$, we get that there exists a neighborhood G of b s.t.

$$\|-(T\circ r\circ f^{-1})(y)\| \leq \frac{1}{2} \quad \text{for all } y\in G\setminus\{b\}.$$

Hence, for all $y\in G\setminus\{b\}$,

$$\|f^{-1}(y)-f^{-1}(b)\| \leq \|T\|\cdot\|y-b\| + \frac{1}{2}\cdot\|f^{-1}(y)-f^{-1}(b)\|,$$

so

(3) $\|f^{-1}(y)-f^{-1}(b)\| \leq 2\cdot\|T\|\cdot\|y-b\|.$

Inequality (3) is also true for $y=b$.

$\square$

Proposition 9.12

Let D be open in V, $a \in D$ and $f, g : D \to \mathbb{R}$. Assume that f, g are differentiable at a. Then

(i) $f \cdot g : D \to \mathbb{R}$, $(f \cdot g)(x) = f(x) \cdot g(x)$ is differentiable at a, and

$$(f \cdot g)'(a) = f(a) \cdot g'(a) + g(a) \cdot f'(a).$$

(ii) if $f(a) \neq 0$, then $\frac{1}{f}$ is differentiable at a, and

$$\left(\frac{1}{f}\right)'(a) = -\frac{1}{f(a)^2} f'(a)$$

Proof

(i) Let $T := f(a) \cdot g'(a) + g(a) \cdot f'(a)$. Then vor all $v \in V$, we have that

$$T(v) = f(a) \cdot g'(a)(v) + g(a) \cdot f'(a)(v).$$

We have that

$$\lim_{x \to a} \frac{(f \cdot g)(x) - (f \cdot g)(a) - T(x-a)}{\|x-a\|} =$$

$$= \lim_{x \to a} \frac{f(x) g(x) - f(a) g(a) - f(a) g'(a)(x-a) - g(a) f'(a)(x-a)}{\|x-a\|} =$$

$$= \lim_{x \to a} \frac{(f(x) - f(a) - f'(a)(x-a)) g(a) + f(x)(g(x) - g(a) - g'(a)(x-a))}{\|x-a\|} +$$

$$+ \frac{(f(x) - f(a)) g'(a)(x-a)}{\|x-a\|}$$

$$= \lim_{x \to a} \frac{f(x) - f(a) - f'(a)(x-a)}{\|x-a\|} \cdot g(a) + \lim_{x \to a} f(x) \cdot \frac{g(x) - g(a) - g'(a)(x-a)}{\|x-a\|}$$

$$+ \lim_{x \to a} \frac{f(x) - f(a)}{\|x-a\|} \cdot g'(a)(x-a) = 0 \cdot g(a) + f(a) \cdot 0 + 0 = 0,$$

since f is continuous at a, so $\lim_{x \to a} f(x) = f(a)$, and

$\frac{\|f(x) - f(a)\|}{\|x-a\|} \leq c$ for x in some neighborhood of a (see Prop. 8.8)

$\lim_{x \to a} f'(a)(x-a) = f'(a)(0) = 0.$

(ii) Assume that $f(a) \neq 0$. Since f is continuous at a, there exists

a neighborhood G of a s.t. $f(x) \neq 0$ for all $x \in G \subseteq D$.

Then $\frac{1}{f} : G \longrightarrow W$, $x \mapsto \frac{1}{f(x)}$ is well defined.

Consider the differentiable real function $g: \mathbb{R} \setminus \{0\} \to \mathbb{R}$ defined

by $g(y) = \frac{1}{y}$. Then $\frac{1}{f} = g \circ f$.

We can apply the Chain Rule to get that $\frac{1}{f}$ is differentiable at

a with derivative

$$\left(\frac{1}{f}\right)'(a) = g'(f(a)) \cdot f'(a) = -\frac{1}{f(a)^2} f'(a).$$

$\square$

Differentiability in $\mathbb{R}^n$

In the sequel, we consider $V = \mathbb{R}^n$, $W = \mathbb{R}^m$, where $m, n \in \mathbb{N}$ and the norms are the Euclidean norms, if not otherwise stated.

A function $f: D \to \mathbb{R}^m$, where $D \subseteq \mathbb{R}^n$, can be written in the form

$$f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix}$$ for all $x \in \mathbb{R}^n$, where $f_j : D \to \mathbb{R}$, $j = 1, \dots, m$ are the components of f. We shall write $f = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}$.

Example 8.14

Let $f: \mathbb{R}^3 \to \mathbb{R}^2$, $f(x,y,z) = \begin{pmatrix} x^2 - y^2 \\ 2xz + 1 \end{pmatrix}$. Then $f = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}$, where

$f_1, f_2 : \mathbb{R}^3 \to \mathbb{R}$, $f_1(x,y,z) = x^2 - y^2$, $f_2(x,y,z) = 2xz + 1$.

Proposition 8.15

Let D be open in $\mathbb{R}^n$, $a \in D$ and $f: D \to \mathbb{R}^m$. Then

$$f \text{ is differentiable at } a \iff f_1, \dots, f_m \text{ are differentiable at } a.$$

In this case, the derivative $f'(a) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is given by

$$f'(a) = \begin{pmatrix} f_1'(a) \\ \vdots \\ f_m'(a) \end{pmatrix}.$$

That is, $(f'(a))_j = f_j'(a)$ for all $j = 1, \dots, m$.

Proof

We have that

f is differentiable at a with derivative $f'(a) \iff$

$$\iff \lim_{x \to a} \frac{f(x) - f(a) - f'(a)(x-a)}{\|x-a\|} = 0$$

$$\iff \lim_{x \to a} \left(\frac{f_1(x) - f_1(a) - (f'(a))_1(x-a)}{\|x-a\|} \cdots \frac{f_m(x) - f_m(a) - (f'(a))_m(x-a)}{\|x-a\|} \right)^T = 0$$

$$\iff \text{for all } j = 1, \dots, m, \quad \lim_{x \to a} \frac{f_j(x) - f_j(a) - (f'(a))_j(x-a)}{\|x-a\|} = 0$$

$\iff$ for all $j = 1, \dots, m$, f_j is differentiable at a and $f_j'(a) = (f'(a))_j$.

The last equivalence is true iff $(f'(a))_j \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$. It is easy to prove that if $T: \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, $T = \begin{pmatrix} T_1 \\ \vdots \\ T_m \end{pmatrix}$, then T_j is a linear transformation for all $j = 1, \dots, m$. $\square$

Remark 8.16

As a consequence of the above proposition, we get that for a path $\gamma: I \to \mathbb{R}^n$, where I is an open interval in $\mathbb{R}$,

γ is differentiable $\iff$ γ is differentiable in the sense of Definition 8.18.

If $f: D \subseteq \mathbb{R}^n \to \mathbb{R}^m$, where D is open, is differentiable at $a \in D$, then the derivative $f'(a)$ is a linear transformation $\mathbb{R}^n \to \mathbb{R}^m$. Then $f'(a)$ can be represented by a $m \times n$-matrix and sometimes $f'(a)$ is identified with this matrix. If $A = (a_{ij})_{\substack{1 \le i \le m \\ 1 \le j \le n}}$ is the associated matrix and $h = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} \in \mathbb{R}^n$, then

$$f'(a)(h) = Ah = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}.$$

If $f: D \subseteq \mathbb{R}^n \to \mathbb{R}$ is real-valued, then $f'(a)$ is a linear transformation $\mathbb{R}^n \to \mathbb{R}$, so it is represented by a $1\times n$-matrix, that is a row vector. The transpose $[f'(a)]^T$ of this $1\times n$-matrix is a $n\times 1$-matrix, a column vector.

Definition 8.17

Let $f: D \subseteq \mathbb{R}^n \to \mathbb{R}$, where D is open in $\mathbb{R}^n$, and assume that f is differentiable at a. The gradient of f at a, denoted by $\operatorname{grad} f(a)$ or $\nabla f(a)$, is defined as

$$\operatorname{grad} f(a) := [f'(a)]^T.$$

Directional and partial derivatives

In the sequel, $D \subseteq \mathbb{R}^n$ is open, $a \in D$ and $f: D \to \mathbb{R}^m$.

Remark 8.18

For any vector $v \in \mathbb{R}^n$, there exists a neighborhood $U_0 \subseteq \mathbb{R}$ of 0 s.t. $a+tv \in D$ for all $t \in U_0$.

Proof

Since $a \in D$ and D is open, there exists $r>0$ s.t. $U_r(a) \subseteq D$. Let $v \in \mathbb{R}^n$. If $v=0$, we can take any neighborhood of 0, since $a+t0 = a \in D$. Assume that $v \neq 0$, so $\|v\| \neq 0$. Then by setting $\delta := \frac{r}{\|v\|}$ we get that for all $t \in U_0 :=]-\delta, \delta[$,

$$\|(a+tv)-a\| = \|tv\| = |t| \cdot \|v\| < \delta \cdot \|v\| = r,$$

so $a+tv \in U_r(a) \subseteq D$. $\square$

Definition 9.19

Let $v \in \mathbb{R}^n$, $v \neq 0$. The directional derivative of f at a in the direction v is defined as

$$D_v f(a) = \lim_{t \to 0} \frac{f(a+tv) - f(a)}{t},$$

provided the limit exists.

Proposition 9.20

Let $f = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}$ and $v \in \mathbb{R}^n \setminus \{0\}$. Then

$$D_v f(a) \text{ exists} \iff D_v f_j(a) \text{ exists for all } j = 1, \dots, m.$$

In this case, $D_v f(a) = \begin{pmatrix} D_v f_1(a) \\ \vdots \\ D_v f_m(a) \end{pmatrix}$.

Proof

Similar to the proof of Proposition 9.15. □

Proposition 9.21

Assume that f is differentiable at a. Then for any $v \in \mathbb{R}^n$, $v \neq 0$, the directional derivative $D_v f(a)$ exists and

$$D_v f(a) = f'(a)(v).$$

Proof

Let $v \in \mathbb{R}^n$, $v \neq 0$. Since f is differentiable at a, by Remark 9.2 (vi), there exists a map $\psi : U_0 \to \mathbb{R}^m$, where is a neighborhood of 0 in $\mathbb{R}^n$,

s.t.

(1) $\quad f(a+h) = f(a) + f'(a)(h) + \|h\| \cdot \psi(h) \quad$ for all $h \in U_0$

and $\lim\limits_{h \to 0} \psi(h) = 0$.

As in Remark 1.18, we can find a neighborhood V_0 of 0 in $\mathbb{R}$ s.t. $tv \in U_0$ for all $t \in V_0$. Then for $t \in V_0$ we can take $h := tv$ in (1) to get that

$$f(a+tv) = f(a) + f'(a)(tv) + \|tv\| \cdot \psi(tv), \text{ so}$$

$$f(a+tv) - f(a) = t \cdot f'(a)(v) + |t| \cdot \|v\| \cdot \psi(tv).$$

We get that

(2) $$\frac{f(a+tv) - f(a)}{t} = f'(a)(v) + \frac{|t|}{t} \cdot \|v\| \cdot \psi(tv) \quad \text{for all } t \in V_0 \setminus \{0\}.$$

Since $\lim\limits_{t \to 0} tv = 0$, we have that $\lim\limits_{t \to 0} \psi(tv) = 0$ by (1). Moreover, $\left| \frac{|t|}{|t} \cdot \|v\| \right| = \|v\|$, so we can conclude that $\lim\limits_{t \to 0} \left(\frac{|t|}{t} \cdot \|v\| \cdot \psi(tv) \right) = 0$.

Thus, by (2), $\lim\limits_{t \to 0} \frac{f(a+tv) - f(a)}{t} = f'(a)(v)$. That is, the directional derivative $D_v f(a)$ exists and $D_v f(a) = f'(a)(v)$.

$\square$

Thus, differentiability at a implies the existence of all directional derivatives at a. We shall see later that the converse is not true. There exist functions having all directional derivatives at a point without being differentiable at that point.

Proposition 8.21 can be used to compute $f'(a)$ when we know that f is differentiable at a. If $v_1, \dots, v_n$ is a basis of $\mathbb{R}^n$, then every vector $h \in \mathbb{R}^n$ can be uniquely represented as a linear combination $h = \sum_{i=1}^{n} \alpha_i v_i$, with $\alpha_i \in \mathbb{R}$. Since $f'(a)$ is linear, we get that

$$f'(a)(h) = f'(a)\left(\sum_{i=1}^{n} \alpha_i v_i\right) = \sum_{i=1}^{n} \alpha_i\, f'(a)(v_i) = \sum_{i=1}^{n} \alpha_i D_{v_i} f(a).$$

Therefore $f'(a)$ is known if the directional derivatives $D_{v_i} f(a)$ for the basis vectors are known.

Let us consider the standard basis $e_1, \dots, e_n \in \mathbb{R}^n$, where

$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow i\text{th position} \qquad \text{for all } i = 1, \dots, n.$$

<u>Definition 8.22</u>

Let $i = 1, \dots, n$. The <u>i-th partial derivative of f at a</u> or <u>the partial derivative of f at a with respect to the i-th variable</u> is by definition the directional derivative $D_{e_i} f(a)$ at a in the direction e_i, provided it exists.

We use the following notations for the i-th partial derivative of f at a:

$$\frac{\partial f}{\partial x_i}(a), \quad \partial_i f(a), \; D_i f(a).$$

If $\frac{\partial f}{\partial x_i}(a)$ exists, we say also that f is <u>partial differentiable at a with respect to the i-th variable</u>. We say that f is <u>partial differentiable at a</u> if $\frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_n}(a)$ exists.

If $\frac{\partial f}{\partial x_i}(a)$ exists for all $a \in D$ we get a **function** $\frac{\partial f}{\partial x_i} : D \to \mathbb{R}^m$ whose value at any $a \in D$ is $\frac{\partial f}{\partial x_i}(a)$; $\frac{\partial f}{\partial x_i}$ is the <u>i-th partial derivative.</u>

<u>Remark 1.23</u>

Let $i = 1, \dots, n$ and $a = (a_1, \dots, a_n) \in D$.

(i) $$\frac{\partial f}{\partial x_i}(a) = \lim_{x_i \to a_i} \frac{f(a_1, \dots, a_{i-1}, x_i, a_{i+1}, \dots, a_n) - f(a)}{x_i - a_i}.$$

(ii) If $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ is real-valued, then
$$\frac{\partial f}{\partial x_i}(a) = \lim_{x_i \to a_i} \frac{\varphi_i(x_i) - \varphi_i(a_i)}{x_i - a_i} = \varphi_i'(a_i),$$
where $\varphi_i : \mathbb{R} \to \mathbb{R}$, $\varphi_i(t) = f(a_1, \dots, a_{i-1}, t, a_{i+1}, \dots, a_n)$.

<u>Proof</u>

(i) $$\frac{\partial f}{\partial x_i}(a) = D_{e_i} f(a) = \lim_{t \to 0} \frac{f(a + t e_i) - f(a)}{t} =$$
$$= \lim_{t \to 0} \frac{f(a_1, \dots, a_{i-1}, a_i + t, a_{i+1}, \dots, a_n) - f(a)}{t}$$
$$\underset{x_i := a_i + t}{=} \lim_{x_i \to a_i} \frac{f(a_1, \dots, a_{i-1}, x_i, a_{i+1}, \dots, a_n) - f(a)}{x_i - a_i}$$

(ii) Apply (i). $\square$

<u>Remark 1.24</u>

Let $f = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}$. Then for all $i = 1, \dots, n$,
$$\frac{\partial f}{\partial x_i}(a) \text{ exists} \iff \frac{\partial f_1}{\partial x_i}(a), \dots, \frac{\partial f_m}{\partial x_i}(a) \text{ exist.}$$

In this case, $\dfrac{\partial f}{\partial x_i}(a) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_i}(a) \\ \vdots \\ \dfrac{\partial f_m}{\partial x_i}(a) \end{pmatrix}$.

<u>Proof</u>

It is a consequence of Proposition 8.20. □

Let $D \subseteq \mathbb{R}^n$ be open, $f : D \to \mathbb{R}^m$, $f = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}$ and $a \in D$. In order to compute a partial derivative of f at a, we compute the partial derivatives at a of the components $f_1, \dots, f_m$. Thus, we reduce the problem to real-valued functions.

If $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ is real-valued, then we compute the partial derivative $\dfrac{\partial f}{\partial x_i}(a)$ as follows: we fix all the variables except for the i-th, putting them equal to $a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n$ and we consider the real function $t \longmapsto f(a_1, \dots, a_{i-1}, t, a_{i+1}, \dots, a_n)$, $t \in \mathbb{R}$; now we differentiate this function at a as in Analysis I and the result is $\dfrac{\partial f}{\partial x_i}(a)$.

<u>Definition 8.25</u>

Assume that $f : D \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is partial differentiable at $a \in D$.
Then the matrix of partial derivatives

$$\begin{pmatrix} \dfrac{\partial f_1}{\partial x_1}(a) & \dfrac{\partial f_1}{\partial x_2}(a) & \dots & \dfrac{\partial f_1}{\partial x_n}(a) \\ \dfrac{\partial f_2}{\partial x_1}(a) & \dfrac{\partial f_2}{\partial x_2}(a) & \dots & \dfrac{\partial f_2}{\partial x_n}(a) \\ \dots & \dots & \dots & \dots \\ \dfrac{\partial f_m}{\partial x_1}(a) & \dfrac{\partial f_m}{\partial x_2}(a) & \dots & \dfrac{\partial f_m}{\partial x_n}(a) \end{pmatrix}$$

is called the Jacobi matrix of f at a and is denoted by $\mathcal{J}f(a)$.
If $m=n$, then the matrix $\mathcal{J}f(a)$ is square and its determinant $\det \mathcal{J}f(a)$ is called the Jacobian determinant or the Jacobian of f at a.
It is frequently denoted $\dfrac{\partial(f_1,\dots,f_n)}{\partial(x_1,\dots,x_n)}$ or $\dfrac{D(f_1,\dots,f_n)}{D(x_1,\dots,x_n)}$.

Remark 9.26

If $f:\mathbb{R}^3\to\mathbb{R}$, one also uses the notation $\dfrac{\partial f}{\partial x}, \dfrac{\partial f}{\partial y}, \dfrac{\partial f}{\partial z}$ instead of $\dfrac{\partial f}{\partial x_1}, \dfrac{\partial f}{\partial x_2}, \dfrac{\partial f}{\partial x_3}$.

Example 9.27

(i) Let $f:\mathbb{R}^2\to\mathbb{R}$, $f(x_1,x_2)=x_1^2x_2+x_2^3$.
To find $\dfrac{\partial f}{\partial x_1}(x_1,x_2)$ we hold x_2 constant and differentiate only with respect to x_1; this yields

$$\frac{\partial f}{\partial x_1}(x_1,x_2)=\frac{d(x_1^2x_2+x_2^3)}{dx_1}=2x_1x_2.$$

Similarly, to find $\dfrac{\partial f}{\partial x_2}(x_1,x_2)$ we hold x_1 constant and differentiate only with respect to x_2:

$$\frac{\partial f}{\partial x_2}(x_1,x_2)=\frac{d(x_1^2x_2+x_2^3)}{dx_2}=x_1^2+3x_2^2.$$

Thus, f is partial differentiable on $\mathbb{R}^2$ and for all $a=(a_1,a_2)\in\mathbb{R}^2$,

$$\frac{\partial f}{\partial x_1}(a)=2a_1a_2,\quad \frac{\partial f}{\partial x_2}(a)=a_1^2+3a_2^2,\quad \mathcal{J}f(a)=\left(\frac{\partial f}{\partial x_1}(a)\ \ \frac{\partial f}{\partial x_2}(a)\right)$$
$$=\left(2a_1a_2\ \ a_1^2+3a_2^2\right).$$

(ii) Let $f : \mathbb{R}^2 \to \mathbb{R}^2$, $f(x_1, x_2) = \begin{pmatrix} x_1^2 - x_2^2 \\ 2x_1x_2 \end{pmatrix}$.

The components of f are $f_1, f_2 : \mathbb{R}^2 \to \mathbb{R}$, $f_1(x_1, x_2) = x_1^2 - x_2^2$, $f_2(x_1, x_2) = 2x_1x_2$.

Then f is partial differentiable on $\mathbb{R}^2$ and for all $a = (a_1, a_2) \in \mathbb{R}^2$ we get that

$$J_f(a_1, a_2) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a_1, a_2) & \frac{\partial f_1}{\partial x_2}(a_1, a_2) \\ \frac{\partial f_2}{\partial x_1}(a_1, a_2) & \frac{\partial f_2}{\partial x_2}(a_1, a_2) \end{pmatrix} =$$

$$= \begin{pmatrix} 2a_1 & -2a_2 \\ 2a_2 & 2a_1 \end{pmatrix}.$$

(iii) Let $f : \mathbb{R}^2 \to \mathbb{R}^3$, $f(x, y) = \begin{pmatrix} x^2 - 2xy \\ x^2 + y^3 \\ \sin x \end{pmatrix}$

The components of f are $f_1, f_2, f_3 : \mathbb{R}^3 \to \mathbb{R}$,

$f_1(x, y) = x^2 - 2xy$, $f_2(x, y) = x^2 + y^3$, $f_3(x, y) = \sin x$.

For all $(x, y) \in \mathbb{R}^2$, f is partial differentiable at (x, y) and

$$J_f(x, y) = \begin{pmatrix} \frac{\partial f_1}{\partial x}(x, y) & \frac{\partial f_1}{\partial y}(x, y) \\ \frac{\partial f_2}{\partial x}(x, y) & \frac{\partial f_2}{\partial y}(x, y) \\ \frac{\partial f_3}{\partial x}(x, y) & \frac{\partial f_3}{\partial y}(x, y) \end{pmatrix} = \begin{pmatrix} 2x - 2y & -2x \\ 2x & 3y^2 \\ \cos x & 0 \end{pmatrix}$$

In particular, $J_f(0, 1) = \begin{pmatrix} -2 & 0 \\ 0 & 3 \\ 1 & 0 \end{pmatrix}$.

(iv) Let $f: \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) = x^{\frac{1}{3}} y^{\frac{1}{3}} = \sqrt[3]{x} \cdot \sqrt[3]{y}$.

Then for all $(x,y) \in \mathbb{R}^2 \setminus \{(0,0)\}$,

$$Jf(x,y) = \left(\frac{\partial f}{\partial x}(x,y) \quad \frac{\partial f}{\partial y}(x,y) \right) = \left(\frac{1}{3} \cdot y^{\frac{1}{3}} \cdot x^{-\frac{2}{3}}, \ \frac{1}{3} x^{\frac{1}{3}} y^{-\frac{2}{3}} \right)$$

$$= \left(\frac{1}{3} \frac{\sqrt[3]{y}}{\sqrt[3]{x^2}} \quad \frac{1}{3} \frac{\sqrt[3]{x}}{\sqrt[3]{y^2}} \right).$$

To obtain the partial derivative at $(0,0)$ we can not simply substitute $(x,y) = (0,0)$. In this case, we use the definition.

$$\frac{\partial f}{\partial x}(0,0) = \lim_{x \to 0} \frac{f(x,0) - f(0,0)}{x - 0} = \lim_{x \to 0} \frac{0-0}{x} = 0.$$

$$\frac{\partial f}{\partial y}(0,0) = \lim_{y \to 0} \frac{f(0,y) - f(0,0)}{y - 0} = 0.$$

Thus, $Jf(0,0) = (0 \quad 0)$.

<u>Proposition 8.28</u>

Let $D \subseteq \mathbb{R}^n$ be open, $a \in D$ and $f: D \to \mathbb{R}^m$. If f is differentiable at a, then f is partial differentiable at a and for all $h \in \mathbb{R}^n$

$$f'(a)(h) = Jf(a) \cdot h.$$

Thus, the matrix associated to $f'(a)$ is the Jacobi matrix of f at a.

<u>Proof</u>

By Proposition 8.21, f is partial differentiable at a and for all $i = 1, \dots, n$

$$\frac{\partial f}{\partial x_i}(a) = D_{e_i} f(a) = f'(a)(e_i).$$

Let $h = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} \in \mathbb{R}^n$. Then $h = \sum_{i=1}^{n} h_i e_i$, so

$$f'(a)(h) = f'(a)\left(\sum_{i=1}^{n} h_i e_i\right) = \sum_{i=1}^{n} h_i f'(a)(e_i) = \sum_{i=1}^{n} h_i \frac{\partial f}{\partial x_i}(a)$$

$$= \sum_{i=1}^{n} h_i \cdot \begin{pmatrix} \frac{\partial f_1}{\partial x_i}(a) \\ \vdots \\ \frac{\partial f_m}{\partial x_i}(a) \end{pmatrix}, \text{ by Remark 9.24}$$

$$= \begin{pmatrix} \sum_{i=1}^{n} h_i \cdot \frac{\partial f_1}{\partial x_i}(a) \\ \vdots \\ \sum_{i=1}^{n} h_i \cdot \frac{\partial f_m}{\partial x_i}(a) \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ & & \\ \frac{\partial f_m}{\partial x_1}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a) \end{pmatrix} \cdot \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$$

$$= J_f(a) \cdot h.$$

Remark 9.29

It is possible that all partial derivatives exist at a without f being differentiable at a. Then the Jacobi matrix can be formed, but it does not represent the derivative $f'(a)$, which does not exist.

Remark 9.30

To check whether f is differentiable at a, one first verifies that f is partial differentiable at a. This is a necessary condition for the existence of $f'(a)$, by Proposition 9.28. Then one forms the Jacobi matrix $J_f(a)$ and considers the associated linear transformation $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, $T(h) = J_f(a)h$. Finally, one tests if $\lim_{x \to a} \frac{f(x) - f(a) - T(x-a)}{\|x-a\|} = 0$ holds. If this holds, then f is differentiable at a with derivative $f'(a) = T$.

The following example shows that a function can have all directional derivatives at a without being differentiable at a.

Example 8.31

Let $f: \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$f(x_1,x_2) = \begin{cases} 0 & \text{for } (x_1,x_2) = (0,0) \\ \dfrac{|x_1| \cdot x_2}{\sqrt{x_1^2+x_2^2}} & \text{for } (x_1,x_2) \neq (0,0). \end{cases}$$

Then f has all directional derivatives at $(0,0)$, but f is not differentiable at $(0,0)$.

Proof

Let $v = (v_1,v_2) \in \mathbb{R}^2 \setminus \{(0,0)\}$. Then

$$D_v f(0,0) = \lim_{t\to 0} \frac{f((0,0)+tv) - f(0,0)}{t} = \lim_{t\to 0} \frac{f(tv) - f(0,0)}{t} = \lim_{t\to 0} \frac{f(tv)}{t}$$

$$= \lim_{t\to 0} \frac{1}{t} \cdot \frac{|tv_1| \cdot t v_2}{\sqrt{t^2 v_1^2 + t^2 v_2^2}} = \lim_{t\to 0} \frac{|t||v_1| \cdot v_2}{|t| \cdot \sqrt{v_1^2+v_2^2}} =$$

$$= \lim_{t\to 0} \frac{|v_1| \cdot v_2}{\sqrt{v_1^2+v_2^2}} = \frac{|v_1| \cdot v_2}{\sqrt{v_1^2+v_2^2}}.$$

It follows that $\frac{\partial f}{\partial x_1}(0,0) = D_{e_1} f(0,0) = \frac{|1| \cdot 0}{\sqrt{1+0}} = 0$ and $\frac{\partial f}{\partial x_2}(0,0) = D_{e_2} f(0,0) = 0$. Thus, $J_f(0,0) = (0 \;\; 0)$.

Assume now that f is differentiable at $(0,0)$. By Proposition 8.28, we must have for all $v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \in \mathbb{R}^2$:

$$f'(0,0)(v) = J_f(0,0) \cdot v = (0 \;\; 0) \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0.$$

Thus, $D_v f(0,0) = f'(0,0)(v) = 0$. Yet, the preceding computations yield for $v=(1,1)$ that $D_v f(0,0) = \frac{1}{\sqrt{2}}$.

We have got a contradiction. Thus, f is not differentiable at $(0,0)$. □

<u>Proposition 9.32</u>

Let D be open in $\mathbb{R}^n$, $a \in D$ and $f: D \to \mathbb{R}^m$. Assume that f is partial differentiable on D and that the partial derivatives $\frac{\partial f}{\partial x_1}, \dots \frac{\partial f}{\partial x_n}$ are continuous at a.

Then f is differentiable at a and $f'(a)$ has as associate matrix the Jacobi matrix $J_f(a)$ of f at a.

<u>Proof</u>

By Propositions 9.15 and Remark 9.24, it follows that it is enough to consider the case $m=1$. We have to prove that $f: D \to \mathbb{R}$ is differentiable at a and that for all $h = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} \in \mathbb{R}^n$,

$$f'(a)(h) = J_f(a)\cdot h = \left(\frac{\partial f}{\partial x_1}(a) \ \dots \ \frac{\partial f}{\partial x_n}(a)\right)\begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} =$$

$$= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a)\cdot h_i .$$

Let $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ be defined by $T(h) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a) h_i$.

Let $\varepsilon > 0$. Since for any $i = 1, \dots, n$, the function $\frac{\partial f}{\partial x_i}: D \to \mathbb{R}$ is continuous at a, there exists $r_i > 0$ s.t. $U_{r_i}(a) \subseteq D$ and

$$(\forall x \in D)\left(x \in U_{r_i}(a) \Rightarrow \left|\frac{\partial f}{\partial x_i}(x) - \frac{\partial f}{\partial x_i}(a)\right| < \frac{\varepsilon}{n}\right).$$

Let $r := \min_{i=1,..,n} r_i$. Then $U_r(a) \subseteq U_{r_i}(a) \subseteq D$ and

(1) $(\forall i=1,..,n)(\forall x \in D)\left(x \in U_r(a) \Rightarrow \left|\frac{\partial f}{\partial x_i}(x) - \frac{\partial f}{\partial x_i}(a)\right| < \frac{\varepsilon}{n}\right)$.

Let $x \in U_r(a)$. Then

$$f(x) - f(a) - T(x-a) = f(x_1, x_2, \dots, x_n) - f(a_1, a_2, \dots, a_n) - \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a)(x_i - a_i)$$

$$= \left[f(x_1, x_2, \dots, x_n) - f(a_1, x_2, \dots, x_n) - \frac{\partial f}{\partial x_1}(a)(x_1 - a_1)\right] +$$

$$+ \left[f(a_1, x_2, \dots, x_n) - f(a_1, a_2, x_3, \dots, x_n) - \frac{\partial f}{\partial x_2}(a)(x_1 - a_1)\right] +$$

$$+ \dots\dots$$

$$+ \left[f(a_1, a_2, \dots, a_{n-1}, x_n) - f(a_1, a_2, \dots, a_n) - \frac{\partial f}{\partial x_n}(a)(x_n - a_n)\right].$$

For any $i = 1, \dots, n$ let us define the real function

$$\varphi_i : [a_i, x_i] \to \mathbb{R}, \quad \varphi_i(t) = f(a_1, \dots, a_{i-1}, t, x_{i+1}, \dots, x_n) - \frac{\partial f}{\partial x_i}(a) \cdot t,$$

where we have assumed without loss of generality that $a_i \le x_i$; the same reasoning applies when $x_i \le a_i$.

Then for any $t \in [a_i, x_i]$ we have that $(a_1, \dots, a_{i-1}, t, x_{i+1}, \dots, x_n) \in D$ since $\|(a_1, \dots, a_{i-1}, t, x_{i+1}, \dots, x_n) - (a_1, \dots, a_n)\| = \|(0, \dots, 0, t, x_{i+1} - a_{i+1}, \dots, x_n - a_n)\| = \sqrt{(t-a_i)^2 + \sum_{k=i+1}^{n}(x_k - a_k)^2} \le \sqrt{(x_i - a_i)^2 + \sum_{k=i+1}^{n}(x_k - a_k)^2} \le$

$\le \|x - a\| < r$, so $(a_1, \dots, a_{i-1}, t, x_{i+1}, \dots, x_n) \in U_r(a) \subseteq D$.

It follows that $\varphi_i'(t) = \lim_{y \to t} \frac{\varphi_i(y) - \varphi_i(t)}{y - t}$

$$= \lim_{y \to t} \frac{f(a_1, \dots, a_{i-1}, y, x_{i+1}, \dots, x_n) - f(a_1, \dots, a_{i-1}, t, x_{i+1}, \dots, x_n)}{y - t}$$

$$-\frac{\partial f}{\partial x_i}(a) = \frac{\partial f}{\partial x_i}(a_1,\dots,a_{i-1},t,x_{i+1},\dots,x_n) - \frac{\partial f}{\partial x_i}(a)$$

Hence, φ_i is differentiable on $[a_i,x_i]$ so we can apply the Mean Value Theorem from Analysis I to get the existence of $\xi_i \in]a_i,x_i[$ s.t.

$$\varphi_i(x_i) - \varphi_i(a) = \varphi_i'(\xi_i)(x_i-a_i).$$

That is,

$$f(a_1,\dots,a_{i-1},x_i,x_{i+1},\dots,x_n) - f(a_1,\dots,a_{i-1},a_i,x_{i+1},\dots,x_n) - \frac{\partial f}{\partial x_i}(a)(x_i-a_i) =$$

$$= \left(\frac{\partial f}{\partial x_i}(a_1,\dots,a_{i-1},\xi_i,x_{i+1},\dots,x_n) - \frac{\partial f}{\partial x_i}(a)\right)\cdot(x_i-a_i).$$

Hence,

$$|f(x)-f(a)-T(x-a)| \le \sum_{i=1}^n \left| f(a_1,\dots,a_{i-1},x_i,x_{i+1},\dots,x_n) - f(a_1,\dots,a_{i-1},a_i,x_{i+1},\dots,x_n) - \frac{\partial f}{\partial x_i}(a)(x_i-a_i)\right|$$

$$= \sum_{i=1}^n |\varphi_i(x_i)-\varphi_i(a_i)| = \sum_{i=1}^n \left|\frac{\partial f}{\partial x_i}(a_1,\dots,a_{i-1},\xi_i,x_{i+1},\dots,x_n) - \frac{\partial f}{\partial x_i}(a)\right|\cdot|x_i-a_i|$$

$$\overset{(1)}{<} \frac{\varepsilon}{n}\cdot\sum_{i=1}^n |x_i-a_i| \le \frac{\varepsilon}{n}\cdot\sum_{i=1}^n \|x-a\| =$$

$$= \frac{\varepsilon}{n}\cdot n\cdot\|x-a\| = \varepsilon\cdot\|x-a\|.$$

Thus, we have got that

$$(\forall x\in D)\left(x\in U_r(a) \Rightarrow \frac{|f(x)-f(a)-T(x-a)|}{\|x-a\|} < \varepsilon\right)$$

This shows that $\lim_{x\to a} \frac{|f(x)-f(a)-T(x-a)|}{\|x-a\|} = 0$, that is f is differentiable and $f'(a)=T$.

$\square$