Course 6 March 30, 2021

6. Chain Rule

$$(f\circ g)' = (f'\circ g)\cdot g' \qquad f'(x)\in\mathbb{R} \qquad df(x)\in L(\ldots)$$

6.1. Theorem Let $A\subseteq\mathbb{R}^n$, $B\subseteq\mathbb{R}^m$, $a\in \operatorname{int} A$, $g: A\to B$ s.t. g is differentiable at a and $g(a)\in\operatorname{int} B$, $f: B\to\mathbb{R}^p$ s.t. f is differentiable at $g(a)$. Then $f\circ g: A\to\mathbb{R}^p$ is differentiable at a and

$$d(f\circ g)(a) = df(g(a))\circ dg(a)$$

$$J(f\circ g)(a) = J(f)(g(a))\cdot J(g)(a)$$

Proof.

$$\begin{tikzcd} A\subseteq\mathbb{R}^n \arrow[r, "g"] \arrow[dr, "f\circ g"'] & B\subseteq\mathbb{R}^m \arrow[d, "f"] \\ & \mathbb{R}^p \end{tikzcd} \qquad \begin{tikzcd} \mathbb{R}^n \arrow[r, "dg(a)"] \arrow[dr, dashed, "d(f\circ g)(a)"'] & \mathbb{R}^m \arrow[d, "df(g(a))"] \\ & \mathbb{R}^p \end{tikzcd}$$

Let $b := g(a)$, $\varphi := dg(a)$, $\psi := df(g(a))$.

Since g is diff. at a $\Rightarrow$ $\exists\, \omega: A\to\mathbb{R}^m$ s.t.

(1) $\lim\limits_{x\to a}\omega(x) = 0_m$

(2) $\forall x\in A$: $\quad g(x) = g(a) + \varphi(x-a) + \|x-a\|\cdot\omega(x)$

Since f is diff. at $b=g(a)$ $\Rightarrow$ $\exists\ \sigma : B \to \mathbb{R}^p$ s.t.

(3) $\lim\limits_{u\to b} \sigma(u) = 0_p$

(4) $\forall u\in B$: $f(u) = f(b) + \psi(u-b) + \|u-b\| \cdot \sigma(u)$

Without restricting the generality we may assume that $\sigma(b) = 0_p$, i.e. σ is continuous at b.

Letting $u = g(x)$ in (4) $\Rightarrow$

$\forall x\in A$: $f(g(x)) = f(b) + \psi(g(x)-b) + \| g(x)-b\| \cdot \sigma(g(x))$

$$(f\circ g)(x) = (f\circ g)(a) + \psi(\underbrace{g(x)-g(a)}_{\text{by } (2)}) + \|g(x)-g(a)\|\, \sigma(g(x))$$

$$(f\circ g)(x) = (f\circ g)(a) + \psi\big(\varphi(x-a) + \|x-a\|\, \omega(x)\big) + \|g(x)-g(a)\|\, \sigma(g(x))$$

$\forall x \in A$: $(f\circ g)(x) = (f\circ g)(a) + (\psi\circ\varphi)(x-a) + \|x-a\|\, \psi(\omega(x)) + \|g(x)-g(a)\|\, \sigma(g(x))$

Consider $\rho : A \to \mathbb{R}^p$ $\quad \rho(x) = \psi(\omega(x)) + \dfrac{\|g(x)-g(a)\|}{\|x-a\|}\, \sigma(g(x))$ if $x \neq a$

$\rho(a) = 0_p$

$\Rightarrow$ $\forall x \in A$: $(f\circ g)(x) = (f\circ g)(a) + (\psi\circ\varphi)(x-a) + \|x-a\|\, \rho(x)$

We will prove that $\lim\limits_{x\to a} \rho(x) \overset{*}{=} 0_p$ $\Big\}\Rightarrow$

$\Rightarrow$ $f\circ g$ is diff. at and $d(f\circ g)(a) = \psi\circ\varphi = df(g(a))\circ dg(a)$

$$\forall x \in A \setminus \{a\}: \quad 0 \le \|\rho(x)\| \le \|\varphi(\omega(x))\| + \frac{\|g(x)-g(a)\|}{\|x-a\|} \cdot \|\sigma(g(x))\| \overset{(2)}{=}$$

$$= \|\varphi(\omega(x))\| + \frac{\|\varphi(x-a) + \|x-a\|\,\omega(x)\|}{\|x-a\|} \cdot \|\sigma(g(x))\|$$

$$\le \|\varphi\| \cdot \|\omega(x)\| + \frac{\|\varphi(x-a)\| + \|x-a\| \cdot \|\omega(x)\|}{\|x-a\|} \cdot \|\sigma(g(x))\|$$

$$\le \|\varphi\| \cdot \|\omega(x)\| + \frac{\|\varphi\| \cdot \cancel{\|x-a\|} + \cancel{\|x-a\|} \cdot \|\omega(x)\|}{\cancel{\|x-a\|}} \cdot \|\sigma(g(x))\|$$

$$\forall x \in A: \quad \underbrace{0}_{\to 0} \le \underbrace{\|\rho(x)\|}_{\to 0} \le \underbrace{\|\varphi\| \cdot \underbrace{\|\omega(x)\|}_{\xrightarrow{x\to a}\, 0} + (\|\varphi\| + \underbrace{\|\omega(x)\|}_{\to 0}) \cdot \underbrace{\|\sigma(g(x))\|}_{\xrightarrow{x\to a}\, 0}}_{\xrightarrow{x\to a}\, 0}$$

g diff at $\Rightarrow$ g is cont. at a; σ is cont. at $b = g(a)$ $\Big\}$ $\Rightarrow$ $\sigma \circ g$ is cont. at $\Rightarrow$

$$\Rightarrow \lim_{x\to a} \sigma(g(x)) = \sigma(g(a)) = \sigma(b) = 0_p$$

Hence $\lim_{x \to a} \|\rho(x)\| = 0 \Rightarrow \lim_{x\to a} \rho(x) = 0_p$

We have $J(f\circ g)(a) = [d(f\circ g)(a)] = [df(g(a)) \circ dg(a)]$

$$= [df(g(a))] \cdot [dg(a)] = J(f)(g(a)) \cdot J(g)(a).$$

<u>Remark</u> Let $A \subseteq \mathbb{R}^n$, $B \subseteq \mathbb{R}^m$, $a \in \operatorname{int} A$

$g = g(x_1, \dots, x_n) : A \to B$ s.t. g is diff. at a, $g(a) \in \operatorname{int} B$ $\quad \dfrac{\partial g}{\partial x_j}$

$f = f(u_1, \dots, u_m) : B \to \mathbb{R}$ $\quad f \circ g : A \to \mathbb{R}$

$g = (g_1, \dots, g_m)$

$$\Rightarrow \underbrace{J(f \circ g)(a)}_{1 \times n} = \underbrace{J(f)(g(a))}_{1 \times m} \cdot \underbrace{J(g)(a)}_{m \times n}$$

$$\left(\frac{\partial (f \circ g)}{\partial x_1}(a) \quad \frac{\partial (f \circ g)}{\partial x_2}(a) \quad \cdots \quad \frac{\partial (f \circ g)}{\partial x_n}(a) \right) =$$

$$= \left(\frac{\partial f}{\partial u_1}(g(a)) \quad \frac{\partial f}{\partial u_2}(g(a)) \quad \cdots \quad \frac{\partial f}{\partial u_m}(g(a)) \right) \cdot \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(a) & \frac{\partial g_1}{\partial x_2}(a) & \cdots & \frac{\partial g_1}{\partial x_n}(a) \\ \frac{\partial g_2}{\partial x_1}(a) & \frac{\partial g_2}{\partial x_2}(a) & \cdots & \frac{\partial g_2}{\partial x_n}(a) \\ \cdots & \cdots & & \\ \frac{\partial g_m}{\partial x_1}(a) & \frac{\partial g_m}{\partial x_2}(a) & \cdots & \frac{\partial g_m}{\partial x_n}(a) \end{pmatrix}$$

$$\Rightarrow \frac{\partial (f \circ g)}{\partial x_j}(a) = \frac{\partial f}{\partial u_1}(g(a)) \cdot \frac{\partial g_1}{\partial x_j}(a) + \frac{\partial f}{\partial u_2}(g(a)) \cdot \frac{\partial g_2}{\partial x_j}(a) + \cdots + \frac{\partial f}{\partial u_m}(g(a)) \cdot \frac{\partial g_m}{\partial x_j}(a)$$

$j = \overline{1, n}$

$$\frac{\partial (f \circ g)}{\partial x_j}(a) = \sum_{i=1}^{m} \frac{\partial f}{\partial u_i}(g(a)) \cdot \frac{\partial g_i}{\partial x_j}(a) \qquad j = \overline{1, n}$$

$$\frac{\partial(f\circ g)}{\partial x_j} = \sum_{i=1}^{m} \frac{\partial f}{\partial u_i} \cdot \frac{\partial g_i}{\partial x_j} \qquad j=\overline{1,n}$$

Some special cases

1) Let $g : \mathbb{R} \to \mathbb{R}^3$ $\qquad g(t) = (x(t), y(t), z(t))$

$f = f(x,y,z) : \mathbb{R}^3 \to \mathbb{R}$

$F = f\circ g : \mathbb{R} \to \mathbb{R}$

$$\frac{dF}{dt}(t) = \frac{\partial f}{\partial x}(x(t), y(t), z(t)) \cdot \frac{dx}{dt}(t) + \frac{\partial f}{\partial y}(x(t), y(t), z(t)) \cdot \frac{dy}{dt}(t) + \frac{\partial f}{\partial z}(x(t), y(t), z(t)) \cdot \frac{dz}{dt}(t)$$

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

2) Let $g : \mathbb{R}^3 \to \mathbb{R}^2$ $\qquad g(x,y,z) = (u(x,y,z), v(x,y,z))$

$f = f(u,v) : \mathbb{R}^2 \to \mathbb{R}$

$F = f\circ g : \mathbb{R}^3 \to \mathbb{R}$

$$\frac{\partial F}{\partial x}(x,y,z) = \frac{\partial f}{\partial u}(u(x,y,z), v(x,y,z)) \cdot \frac{\partial u}{\partial x}(x,y,z) + \frac{\partial f}{\partial v}(u(x,y,z), v(x,y,z)) \cdot \frac{\partial v}{\partial x}(x,y,z)$$

$$\frac{\partial F}{\partial y}(x,y,z) = \frac{\partial f}{\partial u}(u(x,y,z), v(x,y,z)) \cdot \frac{\partial u}{\partial y}(x,y,z) + \frac{\partial f}{\partial v}(\text{——} \;''\; \text{——}) \cdot \frac{\partial v}{\partial y}(x,y,z)$$

$$\frac{\partial F}{\partial z}(x,y,z) = \frac{\partial f}{\partial u}(\text{——} \;''\; \text{——}) \cdot \frac{\partial u}{\partial z}(x,y,z) + \frac{\partial f}{\partial v}(\text{——} \;''\; \text{——}) \cdot \frac{\partial v}{\partial z}(x,y,z)$$

$$\boxed{\begin{aligned} \frac{\partial F}{\partial x} &= \frac{\partial f}{\partial u} \cdot \frac{\partial u}{\partial x} + \frac{\partial f}{\partial v} \cdot \frac{\partial v}{\partial x} \\ \frac{\partial F}{\partial y} &= \\ \frac{\partial F}{\partial z} &= \end{aligned}}$$

7. Mean value theorems

7.1. Definition. Let $A \subseteq \mathbb{R}^n$, $f : A \to \mathbb{R}$. A point $a \in A$ is called a local minimum (resp. local maximum) for f if $\exists V \in \mathcal{V}(a)$ s.t. $\forall x \in V \cap A$ one has

$(*)$ $\quad f(a) \le f(x) \quad$ (resp. $f(a) \ge f(x)$)

If $(*)$ holds for all $x \in A$ $\Rightarrow$ a is called a global minimum (resp. global maximum) for f

Local minima and local maxima = local extrema

Global — " — global — " — = global extrema

7.2. Theorem (P. Fermat) Let $A \subseteq \mathbb{R}^n$, $f : A \to \mathbb{R}$, a be a point satisfying the following conditions: (i) $a \in \operatorname{int} A$

(ii) a is a local extremum for f

(iii) f is partially differentiable at a

Then $\nabla f(a) = 0_n$ i.e. $\frac{\partial f}{\partial x_j}(a) = 0 \quad \forall j = \overline{1,n}$.

Proof Assume that a is a local minimum for f

$a \in \operatorname{int} A \Rightarrow \exists r > 0$ s.t. $[a_1 - r, a_1 + r] \times \dots \times [a_n - r, a_n + r] \subseteq A$

$a = (a_1, \dots, a_n)$

Let $j \in \{1, \dots, n\}$ and let $g : [a_j - r, a_j + r] \to \mathbb{R}$

$$g(x) := f(a_1, \dots, a_{j-1}, x, a_{j+1}, \dots, a_n)$$

a is a local minimum for f $\Rightarrow$ a_j is a local minimum for g $\Big\}$ Fermat thm to g $\Rightarrow$

f is partially diff. at a $\Rightarrow$ g is diff. at a_j and

$$g'(a_j) = \frac{\partial f}{\partial x_j}(a)$$

$\Rightarrow g'(a_j) = 0$

$$\Downarrow$$

$$\frac{\partial f}{\partial x_j}(a) = 0.$$

<u>Remarks</u> 1° If $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ is partially diff at some point $a \in \operatorname{int} A$ and $\nabla f(a) = 0_n$, then a is called a critical (or stationary) point of f

a is a local extremum of f $\Rightarrow$ a is a critical point of f

$$\not\Leftarrow$$

<u>Counterexample</u> $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) = x^2 - y^2$

$\nabla f(x,y) = (2x, -2y)$

$\nabla f(x,y) = (0,0) \iff (x,y) = (0,0)$

$\Rightarrow (0,0)$ is a critical point of f

$f(x,0) = x^2 > 0 = f(0,0) \quad \forall x \in \mathbb{R}\setminus\{0\} \Rightarrow (0,0)$ cannot be a local maximum for f

$f(0,y) = -y^2 < 0 = f(0,0) \quad \forall y \in \mathbb{R}\setminus\{0\} \Rightarrow (0,0)$ —— " —— minimum for f

$\Rightarrow (0,0)$ is not a local extremum for f

$z = f(x,y) = x^2 - y^2 \rightarrow$ hyperbolic paraboloid

$\rightarrow$ the critical points of a function that are not local extrema are called <u>saddle points</u>

2) Let $A \subseteq \mathbb{R}^n$ be a compact set s.t. $\operatorname{int} A \neq \emptyset$

$f : A \to \mathbb{R}$ continuous on A and diff on $\operatorname{int} A$

By the Weierstrass Thm $\Rightarrow f$ is bounded and attains its bounds on A

$$m := \min_{x \in A} f(x) \qquad M := \max_{x \in A} f(x)$$

Let $C := \{x \in \operatorname{int} A \mid \nabla f(x) = 0_n\}$

Assume C = finite

$$m_1 := \min_{x \in C} f(x) \qquad M_1 = \max_{x \in C} f(x)$$

We may define $m_2 := \min_{x \in \operatorname{bd} A} f(x) \qquad M_2 := \max_{x \in \operatorname{bd} A} f(x)$

$\operatorname{bd} A = (\operatorname{cl} A) \cap \operatorname{cl}(\mathbb{R}^n \setminus A)$

$\Rightarrow \operatorname{bd} A \subseteq \operatorname{cl} A = A$ – bounded

$\downarrow$

closed and bounded

$\Downarrow$

$\operatorname{bd} A$ is compact

$$m = \min\{m_1, m_2\} \qquad M = \max\{M_1, M_2\}$$

7.3. **Lemma** Let A be an open convex subset of $\mathbb{R}^n$, let $f : A \to \mathbb{R}^m$ be a differentiable function on A, let $a, b \in A$, and let $F : [0,1] \to \mathbb{R}^m$

$$F(t) = f\big((1-t)a + tb\big)$$

Then F is diff. on $[0,1]$ and

$$F'(t) = df((1-t)a + tb)(b-a) \qquad \forall t \in [0,1]$$

Proof. Fix $t_0 \in [0,1]$. Let $(1-t_0)a + t_0 b =: \bar{a}$, $v := b-a$

$$\lim_{t \to t_0} \frac{1}{t-t_0}\left[F(t) - F(t_0)\right] \underset{s=t-t_0}{=} \lim_{s \to 0} \frac{1}{s}\left[F(t_0+s) - F(t_0)\right]$$

$$= \lim_{s \to 0} \frac{1}{s}\Big[f\big((1-t_0-s)a + (t_0+s)b\big) - f\big((1-t_0)a + t_0 b\big)\Big]$$

$$= \lim_{s \to 0} \frac{1}{s}\Big[f(\bar{a} + sv) - f(\bar{a})\Big] = f'(\bar{a}; v) = df(\bar{a})(v)$$

$\Rightarrow$ F is differentiable at t_0 and $F'(t_0) = df\big((1-t_0)a + t_0 b\big)(b-a)$.

7.4. Theorem (MVT for scalar functions of several variables) Let A be an open convex subset of $\mathbb{R}^n$, $f: A \to \mathbb{R}$ be a diff. function on A. Then $\forall\, a, b \in A$ $\exists\, \xi \in (0,1)$ s.t. letting $c := (1-\xi)a + \xi b$ we have

$$f(b) - f(a) = df(c)(b-a)$$

Proof Let $a, b \in A$, and let $F: [0,1] \to \mathbb{R}$, $F(t) := f((1-t)a + tb)$

L7.3 $\Rightarrow$ F is diff. on $[0,1]$ $\xRightarrow{\text{Lagrange MVT}}$ $\exists\, \xi \in (0,1)$ s.t. $F(1) - F(0) = F'(\xi)$

But $F(1) = f(b)$, $F(0) = f(a)$, $F'(\xi) \overset{\text{L7.3}}{=} df((1-\xi)a + \xi b)(b-a)$ $\Bigg\} \Rightarrow$

$\Rightarrow$ $f(b) - f(a) = df(c)(b-a)$, where $c := (1-\xi)a + \xi b$.

7.5. Theorem (MVT for vector functions of several variables). Let A be an open convex subset of $\mathbb{R}^n$, $f: A \to \mathbb{R}^m$ be a differentiable function on A. Then $\forall\, a, b \in A$ $\exists\, \xi \in (0,1)$ s.t. letting $c := (1-\xi)a + \xi b$ we have

$$\|f(b) - f(a)\| \leq \|df(c)\| \cdot \|b-a\|.$$

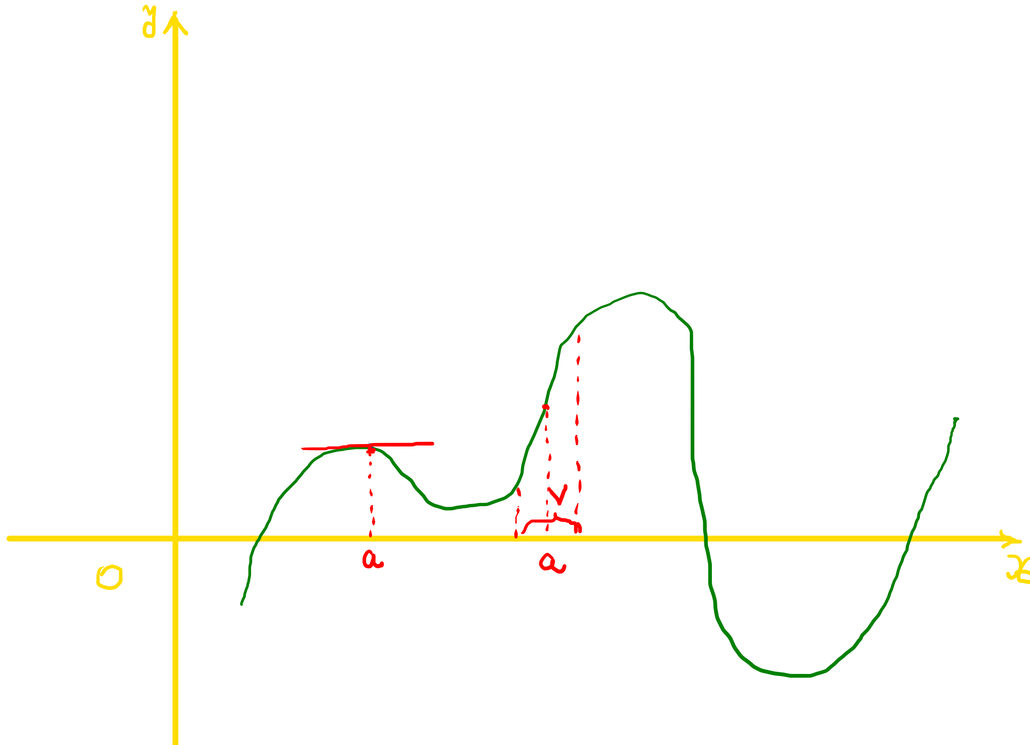
Proof. Let $a, b \in A$, and let $F: [0,1] \to \mathbb{R}^m$, $F(t) := f((1-t)a + tb)$

L7.3 $\Rightarrow$ F is diff on $[0,1]$ $\xRightarrow[\text{functions of a real variable}]{\text{MVT for vector}}$ $\exists\, \xi \in (0,1)$ s.t. $\|F(1) - F(0)\| \leq \|F'(\xi)\|$

But $F(1) = f(b)$, $F(0) = f(a)$, $F'(\xi) = df((1-\xi)a + \xi b)(b-a)$ $\Bigg\} \Rightarrow$

$\Rightarrow$ $\|f(b) - f(a)\| \leq \|df(c)(b-a)\| \leq \|df(c)\| \cdot \|b-a\|$, where $c := (1-\xi)a + \xi b$.

8. The inverse function theorem (The local inversion theorem)



$f : A \subseteq \mathbb{R} \to \mathbb{R}$ continuously diff on int A

$a \in \text{int} A$ s.t. $f'(a) \neq 0$

$\Rightarrow$ $f'(a) > 0$ or $f'(a) < 0$

Since f' is continuous at a $\Rightarrow$

$\Rightarrow \exists V = (a-r, a+r) \in \mathcal{V}(a)$

s.t. $f'(x) > 0 \quad \forall x \in V$

$\Rightarrow f|_V$ is increasing

$\Rightarrow f|_V$ is injective

$\Rightarrow \tilde{f} : V \to f(V)$

$\tilde{f}(x) := f(x)$ is bijective

8.1. Definition. Let $A \subseteq \mathbb{R}^n$ be an open set. A function $f : A \to \mathbb{R}^m$ is said to be of class C^1 on A if is partially diff on A and all its partial derivatives $\frac{\partial f}{\partial x_j} : A \to \mathbb{R}^m$ are continuous on A, $\forall j = \overline{1,n}$

8.2. Definition Let $A, B \subseteq \mathbb{R}^n$ be open sets. A function $f : A \to B$ is called a diffeomorphism if f is bijective, differ. on A, with $f^{-1} : B \to A$ diff. on B.

f is called a C^1-diffeomorphism if f is bijective, of class C^1 on A, with f^{-1} of class C^1 on B