Hence $J\bar{h}(\bar{a}) = J(l \circ \bar{k})(\bar{a}) = [Jl(\bar{k}(\bar{a}))][J\bar{k}(\bar{a})]$

$$= [g_1(\bar{a}) \cdots g_m(\bar{a})\ f_1(\bar{a}) \cdots f_m(\bar{a})] \begin{bmatrix} J\bar{f}(\bar{a}) \\ J\bar{g}(\bar{a}) \end{bmatrix}$$

$$[D\bar{h}(\bar{a})]\bar{v} = [\bar{g}(\bar{a})\ \bar{f}(\bar{a})]\begin{bmatrix} J\bar{f}(\bar{a}) \\ J\bar{g}(\bar{a}) \end{bmatrix}\bar{v} = [\bar{g}(\bar{a})\ \bar{f}(\bar{a})]\begin{bmatrix} [J\bar{f}(\bar{a})]\bar{v} \\ [J\bar{g}(\bar{a})]\bar{v} \end{bmatrix}$$

$$= \bar{g}(\bar{a}) \cdot [J\bar{f}(\bar{a})]\bar{v} + \bar{f}(\bar{a}) \cdot [J\bar{g}(\bar{a})]\bar{v}$$

$$= D\bar{f}(\bar{a})\bar{v} \cdot \bar{g}(\bar{a}) + \bar{f}(\bar{a}) \cdot D\bar{g}(\bar{a})\bar{v}$$

10/19/2016 >

proof of Chain rule:

We're assuming $\bar{r}(\bar{h}) := \bar{g}(\bar{a}+\bar{h}) - \bar{g}(\bar{a}) - [D\bar{g}(\bar{a})]\bar{h}$ has $\lim_{\bar{h}\to\bar{0}} \dfrac{\bar{r}(\bar{h})}{|\bar{h}|} = \bar{0}$

and $\bar{s}(\bar{k}) := \bar{f}(\bar{g}(\bar{a})+\bar{k}) - \bar{f}(\bar{g}(\bar{a})) - [D\bar{f}(\bar{g}(\bar{a}))]\bar{k}$ has $\lim_{\bar{k}\to\bar{0}} \dfrac{\bar{s}(\bar{k})}{|\bar{k}|} = \bar{0}$

So we analyze

$$\bar{f}(\bar{g}(\bar{a}+\bar{h})) \overset{\text{defn of } \bar{r}(\bar{h})}{=} \bar{f}\big(\bar{g}(\bar{a}) + \underbrace{[D\bar{g}(\bar{a})]\bar{h} + \bar{r}(\bar{h})}_{\text{call this } \bar{k} :=}\big)$$

$$= \bar{f}(\bar{g}(\bar{a}) + \bar{k})$$

$$\overset{\text{defn of } \bar{s}(\bar{k})}{=} \bar{f}(\bar{g}(\bar{a})) + [D\bar{f}(\bar{g}(\bar{a}))]\bar{k} + \bar{s}(\bar{k})$$

$$= \bar{f}(\bar{g}(\bar{a})) + \underbrace{[D\bar{f}(\bar{g}(\bar{a}))][D\bar{g}(\bar{a})]\bar{h}}_{\substack{\text{what we wanted} \\ \text{to approximate } \bar{f}(\bar{g}(\bar{a}+\bar{h})) - \bar{f}(\bar{g}(\bar{a}))!}} + \underbrace{[D\bar{f}(\bar{g}(\bar{a}))]\bar{r}(\bar{h}) + \bar{s}(\bar{k})}_{\text{error term?}}$$

Thus we need to show the error term has

$$\lim_{\bar{h}\to\bar{0}} \frac{[D\bar{f}(\bar{g}(\bar{a}))]\bar{r}(\bar{h}) + \bar{s}(\bar{k})}{|\bar{h}|} = \bar{0}$$

Easy to see $\lim_{\bar{h}\to\bar{0}} \dfrac{[D\bar{f}(\bar{g}(\bar{a}))]\bar{r}(\bar{h})}{|\bar{h}|} = \bar{0}$ since $\left| [D\bar{f}(\bar{g}(\bar{a}))] \dfrac{\bar{r}(\bar{h})}{|\bar{h}|} \right| \le \underbrace{\overbrace{|D\bar{f}(\bar{g}(\bar{a}))|}^{\text{matrix norm}}}_{\text{fixed!}} \underbrace{\dfrac{|\bar{r}(\bar{h})|}{|\bar{h}|}}_{\to 0 \text{ as } \bar{h}\to 0}$

(59) So it remains to show $\lim_{\bar{h}\to\bar{0}} \frac{\bar{s}(\bar{k})}{|\bar{h}|} \overset{?}{=} \bar{0}$ when $\bar{k} = [D\bar{g}(\bar{a})\bar{h}] + \bar{r}(\bar{h})$

$$\parallel$$

$$\lim_{\bar{h}\to\bar{0}} \frac{\bar{s}\left([D\bar{g}(\bar{a})]\bar{h} + \bar{r}(\bar{h}) \right)}{|\bar{h}|}$$

an ϵ that we can choose.

Since $\lim_{\bar{h}\to\bar{0}} \frac{\bar{r}(\bar{h})}{\bar{h}} = \bar{0}$, $\exists\, \delta_1 > 0$ with $\frac{|\bar{r}(\bar{h})|}{|\bar{h}|} \le 1$ for $|\bar{h}| < \delta_1$

i.e. $|\bar{r}(\bar{h}) \le |\bar{h}|$

and thus also
$$\begin{aligned} \left| [D\bar{g}(\bar{a})]\bar{h} + \bar{r}(\bar{h}) \right| &\le \left| [D\bar{g}(\bar{a})]\bar{h} \right| + |\bar{r}(\bar{h})| \\ &\le \left|[D\bar{g}(\bar{a})]\right| |\bar{h}| + |\bar{h}| \\ &= \left(1 + |[D\bar{g}(\bar{a})]|\right) |\bar{h}| \end{aligned}$$

Since $\lim_{\bar{k}\to\bar{0}} \frac{\bar{s}(\bar{k})}{|\bar{k}|} = \bar{0}$,
given $\epsilon > 0$, $\exists\, \delta_2$ such that $\frac{|\bar{s}(\bar{k})|}{|\bar{k}|} < \epsilon$ for $|\bar{k}| < \delta_2$.

Hence for $\delta < \min\left(\delta_1, \frac{\delta_2}{1 + |[D\bar{g}(\bar{a})]|}\right)$, if $|\bar{h}| < \delta$ then one has
$$\begin{aligned} \left| [D\bar{g}(\bar{a})]\bar{h} + \bar{r}(\bar{h}) \right| &\le \left(1 + |[D\bar{g}(\bar{a})]|\right) \frac{\delta_2}{1 + |[D\bar{g}(\bar{a})]|} \\ &= \delta_2 \end{aligned}$$

and thus
$$\frac{\left|\bar{s}\left([D\bar{g}(\bar{a})]\bar{h} + \bar{r}(\bar{h})\right)\right|}{\left| [D\bar{g}(\bar{a})]\bar{h} + \bar{r}(\bar{h}) \right|} < \epsilon$$

i.e. $\left|\bar{s}\left([D\bar{g}(\bar{a})]\bar{h} + \bar{r}(\bar{h})\right)\right| < \epsilon \left| [D\bar{g}(\bar{a})]\bar{h} + \bar{r}(\bar{h}) \right| \le \epsilon\left(1 + |[D\bar{g}(\bar{a})]|\right)|\bar{h}|$

Dividing left & right by $|\bar{h}|$ gives $\frac{s(\bar{k})}{|\bar{h}|} = \frac{\left|\bar{s}\left([D\bar{g}(\bar{a})]\bar{h} + \bar{r}(\bar{h})\right)\right|}{|\bar{h}|} < \epsilon$ ∎

(61)

§1.9 M.V.T. and differentiability criteria

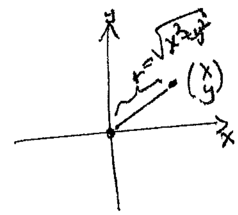
Sadly, sometimes $\bar{f}$ can be continuous at $\bar{x}=\bar{a}$,
have all partial derivatives existing at $\bar{x}=\bar{a}$,
(so Jf exists as a matrix)
but not be differentiable there!

EXAMPLE 1.9.3:

$$f:\mathbb{R}^2 \to \mathbb{R}$$

$$f\begin{pmatrix}x\\y\end{pmatrix} = \begin{cases} \dfrac{x^2y}{x^2+y^2} & \text{if } \begin{pmatrix}x\\y\end{pmatrix} \neq \begin{pmatrix}0\\0\end{pmatrix} \\ 0 & \text{if } \begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix} \end{cases}$$

a removable discontinuity at $\begin{pmatrix}0\\0\end{pmatrix}$ $\longrightarrow$



is actually continuous & differentiable on $\mathbb{R}^2-\{\bar{0}\}$
and continuous at $\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$

since $\left|\dfrac{x^2y}{x^2+y^2}\right| = \dfrac{|x|^2|y|}{|x^2+y^2|} \leq \dfrac{r^3}{r^2}=r$ where $r:=\sqrt{x^2+y^2}$,

since $|x|\leq r$, $|y|\leq r$

and $r\to 0$ as $\begin{pmatrix}x\\y\end{pmatrix}\to\begin{pmatrix}0\\0\end{pmatrix}$

with partial derivatives at $\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$ existing:

$$\left.\frac{\partial f}{\partial x}\right|_{\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}} = \lim_{h\to 0}\frac{f\left(\begin{pmatrix}0\\0\end{pmatrix}+\begin{pmatrix}h\\0\end{pmatrix}\right)-f\begin{pmatrix}0\\0\end{pmatrix}}{h} = \lim_{h\to 0}\frac{0/h^2}{h}=0$$

$$\left.\frac{\partial f}{\partial y}\right|_{\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}} = \lim_{h\to 0}\frac{f\left(\begin{pmatrix}0\\0\end{pmatrix}+\begin{pmatrix}0\\h\end{pmatrix}\right)-f\begin{pmatrix}0\\0\end{pmatrix}}{h} = \lim_{h\to 0}\frac{0/h^2}{h}=0$$

$$\Rightarrow \left[Jf\begin{pmatrix}0\\0\end{pmatrix}\right]=\begin{bmatrix}0 & 0\end{bmatrix}$$

However in the direction $\bar{v}=\begin{bmatrix}1\\1\end{bmatrix}$, the dir. derivative has a different value (i.e. not 0):

$$\lim_{h\to 0}\frac{f\left(\begin{pmatrix}0\\0\end{pmatrix}+h\begin{pmatrix}1\\1\end{pmatrix}\right)-f\begin{pmatrix}0\\0\end{pmatrix}}{h} = \lim_{h\to 0}\frac{h^3/(h^2+h^2)}{h} = \lim_{h\to 0}\frac{1}{2}=\frac{1}{2} \quad \begin{pmatrix}!\\0\end{pmatrix}$$

This disagrees with $\left[Jf\begin{pmatrix}0\\0\end{pmatrix}\right]\bar{v} = \begin{bmatrix}0 & 0\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}=0$,

so f is not differentiable at $\begin{pmatrix}0\\0\end{pmatrix}$.

What's the problem?
The partial derivatives aren't continuous in some neighborhood U of $\begin{pmatrix}0\\0\end{pmatrix}$:

$$\left.\begin{aligned}\frac{\partial f}{\partial x} &= \frac{(x^2+y^2)2xy - x^2y(2x)}{(x^2+y^2)^2} = \frac{2xy^3}{(x^2+y^2)^2}\\ \frac{\partial f}{\partial y} &= \frac{(x^2+y^2)x^2 - x^2y(2y)}{(x^2+y^2)^2} = \frac{x^4-x^2y^2}{(x^2+y^2)^2}\end{aligned}\right\} \text{if } \begin{pmatrix}x\\y\end{pmatrix}\neq\begin{pmatrix}0\\0\end{pmatrix}$$

$\leftarrow$ Can check they have different limits as one approaches $\begin{pmatrix}0\\0\end{pmatrix}$ on different lines.

We need to avoid this pathology.

<u>DEF'N 1.9.6</u>: A function $\bar{f}: \underset{\mathbb{R}^n}{\overset{\text{open}}{\mathcal{U}}} \longrightarrow \mathbb{R}^m$ is called <u>continuously</u> <u>differentiable on $\mathcal{U}$</u> if all its partial derivatives $\frac{\partial f_i}{\partial x_j}$ $\begin{matrix} i=1,\dots,m \\ j=1,\dots,n \end{matrix}$ <u>exist</u> on $\mathcal{U}$, and are <u>continuous on $\mathcal{U}$</u>.

(<u>NOTATION</u>: $\bar{f}$ is C^1 on $\mathcal{U}$)

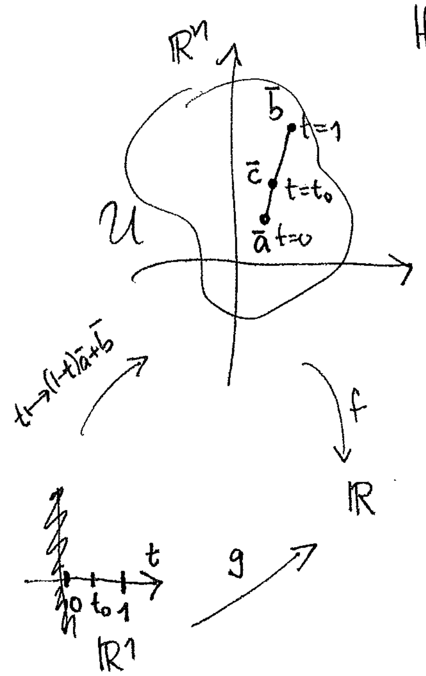
<u>THM 1.9.8</u>: If $\bar{f}$ is C^1 on $\mathcal{U}$, then it is <u>differentiable at every $\bar{a} \in \mathcal{U}$</u> (and $D\bar{f}(\bar{a})$ has matrix $J\bar{f}(\bar{a}) = \left[\frac{\partial f_i}{\partial x_j}(\bar{a}) \right]_{\substack{i=1,\dots,m \\ j=1,\dots,n}}$, of course)

10/21/2016 >

<u>proof</u>: We are going to need the <u>multivariate</u> <u>version</u> of M.V.T. here (and again later):

<u>THM 1.9.1</u> (multivariate MVT)

If $f: \underset{\mathbb{R}^n}{\overset{\text{open}}{\mathcal{U}}} \longrightarrow \mathbb{R}$ contains a line segment $[\bar{a}, \bar{b}]$ and f is differentiable on $\mathcal{U}$, then $\exists$ some $\bar{c} \in [\bar{a}, \bar{b}]$ with $[Df(\bar{c})](\bar{b} - \bar{a}) = f(\bar{b}) - f(\bar{a})$.



In particular (COR. 1.9.2), if $|[Df(\bar{c})]| \le M \quad \forall \bar{c} \in [a, b]$ then $|f(\bar{b}) - f(\bar{a})| \le M |\bar{b} - \bar{a}|$

<u>proof of multivar. MVT</u>:

Parametrize $[\bar{a}, \bar{b}] = \{(1-t)\bar{a} + t\bar{b} : 0 \le t \le 1\}$ and apply usual MVT to $g: [0,1] \longrightarrow \mathbb{R}$ (continuous on $[0,1]$ - why? differentiable on $(0,1)$ - why?)

$$g(t) = f((1-t)\bar{a} + t\bar{b}) \quad \left(= \begin{cases} f(\bar{a}) & \text{if } t=0 \\ f(\bar{b}) & \text{if } t=1 \end{cases} \right)$$

to give $t_0 \in (0,1)$ with $g'(t_0) = \frac{g(1) - g(0)}{1-0} = f(\bar{b}) - f(\bar{a})$)

$$\| \quad \lim_{h \to 0} \frac{g(t_0 + h) - g(t_0)}{h}$$

name [redacted] $\bar{c} := (1-t_0)\bar{a} + t_0\bar{b}$

$$= \lim_{h \to 0} \frac{f(\bar{c} + h(\bar{b} - \bar{a})) - f(\bar{c})}{h}$$

= dir. deriv. of f at $\bar{c}$ in dir. $\bar{b} - \bar{a}$

$$= [Df(\bar{c})](\bar{b} - \bar{a}) \quad \blacksquare$$