DERIVATIVES

5

In our discussion of partial derivatives, we met the function

$$f\begin{pmatrix}x\\y\end{pmatrix} := \begin{cases} \frac{xy^2}{x^2+y^4}, & \begin{pmatrix}x\\y\end{pmatrix} \neq \begin{pmatrix}0\\0\end{pmatrix} \\ 0, & \begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}. \end{cases}$$

This function, while not continuous at $\begin{pmatrix}0\\0\end{pmatrix}$, not only had partial derivatives but directional derivatives in every direction at $\begin{pmatrix}0\\0\end{pmatrix}$:

$$\left(D_{\begin{pmatrix}a\\b\end{pmatrix}} f\right)\begin{pmatrix}0\\0\end{pmatrix} = \lim_{t\to 0} \frac{f\left(\begin{pmatrix}0\\0\end{pmatrix} + t\begin{pmatrix}a\\b\end{pmatrix}\right) - f\begin{pmatrix}0\\0\end{pmatrix}}{t}$$

direction vector

$$= \lim_{t\to 0} \frac{f\begin{pmatrix}ta\\tb\end{pmatrix}}{t} = \lim_{t\to 0} \frac{ab^2 t^2}{a^2 t^2 + b^4 t^4}$$

$$= \begin{cases} b^2/a, & a \neq 0 \\ 0, & a = 0. \end{cases}$$

This is a highly unsatisfactory situation: differentiability is supposed to imply continuity. Moreover, it should imply that a function has a good linear approximation, as with the tangent line in 1-variable calculus — and there is nothing to suggest linear behavior in "b^2/a" at all!

To fix this, we need a notion of differentiability that guarantees continuity, and a notion of derivative that produces a good linear approximation. Let's begin with scalar-valued functions.

§1. The total derivative

Let $S \subset \mathbb{R}^n$ be an open set, $\vec{a} \in S$ be a point, and $f : S \to \mathbb{R}$ a function.

The (total) derivative of f at $\vec{a}$, $Df(\vec{a})$, is a linear transformation

$$T : \mathbb{R}^n \to \mathbb{R}$$

such that as $\vec{h} \to \vec{0}$, the "error"

$$E(\vec{h}) := \frac{f(\vec{a}+\vec{h}) - f(\vec{a}) - T(\vec{h})}{\|\vec{h}\|} \to 0.$$

If such a transformation exists, then f is said to be differentiable at $\vec{a}$.

Theorem: If f is differentiable at $\vec{a}$, then its partials exist there, and $T = Df(\vec{a})$ has matrix given by

$$(*) \quad \left(\frac{\partial f}{\partial x_1}(\vec{a}), \dots, \frac{\partial f}{\partial x_n}(\vec{a}) \right),$$

the transpose of the gradient $\vec{\nabla} f(\vec{a})$.

Moreover, the directional derivatives exist and are given by

$$D_{\vec{v}} f(\vec{a}) = T(\vec{v}) = \vec{\nabla} f(\vec{a}) \cdot \vec{v}.$$

In particular, this means that if it exists, T is unique (determined by $(*)$).

Theorem: If f is differentiable at $\vec{a}$, then its partials exist there, and $T = Df(\vec{a})$ has matrix given by

(*) $$\left(\frac{\partial f}{\partial x_1}(\vec{a}), \dots, \frac{\partial f}{\partial x_n}(\vec{a}) \right),$$

the transpose of the gradient $\vec{\nabla} f(\vec{a})$.

Moreover, the directional derivatives exist and are given by

$$D_{\vec{v}} f(\vec{a}) = T(\vec{v}) = \vec{\nabla} f(\vec{a}) \cdot \vec{v}.$$

In particular, this means that if it exists, T is unique (determined by (*)).

We also see that our function with

$$D_{\begin{pmatrix}1\\1\end{pmatrix}} f\begin{pmatrix}0\\0\end{pmatrix} = \begin{cases} b^2/a \\ 0 \end{cases} \left(\neq \vec{w} \cdot \begin{pmatrix}a\\b\end{pmatrix} \text{ for any } \vec{w} \right)$$

cannot possibly be differentiable at $\vec{0}$.

Proof: $D_{\vec{v}} f(\vec{a}) = \lim_{t \to 0} \dfrac{f(\vec{a}+t\vec{v}) - f(\vec{a})}{t}$

$E(\vec{h})$

$\vec{h} = t\vec{v}$

$$= \lim_{t \to 0} \frac{|t|\|\vec{v}\| \left(\dfrac{f(\vec{a}+\vec{h}) - f(\vec{a}) - T(\vec{h})}{\|\vec{h}\|} \right) + T(t\vec{v})}{t}$$

$$= \lim_{t \to 0} \left(\frac{|t|}{t} \|\vec{v}\| \, E(\vec{h}) + \frac{t\,T(\vec{v})}{t} \right)$$

$E(\vec{h}) \to 0$ since f differentiable

$$= T(\vec{v})$$

$\vec{v} = \sum_{k=1}^{n} v_k \hat{e}_k$ & T linear

$$= \sum_{k=1}^{n} v_k T(\hat{e}_k)$$

$$= \sum_{k=1}^{n} v_k D_{\hat{e}_k} f(\vec{a}) = \sum_{k=1}^{n} v_k \frac{\partial f}{\partial x_k}(\vec{a})$$

matrix given by

$$= \left(\frac{\partial f}{\partial x_1}(\vec{a}), \ldots, \frac{\partial f}{\partial x_n}(\vec{a}) \right) \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

matrix product

$$= \begin{pmatrix} \partial f/\partial x_1 (\vec{a}) \\ \vdots \\ \partial f/\partial x_n (\vec{a}) \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

dot product

$\vec{\nabla} f(\vec{a})$ gradient vector

□

Theorem: If f is differentiable at $\vec{a}$, then its partials exist there, and $T = Df(\vec{a})$ has matrix given by

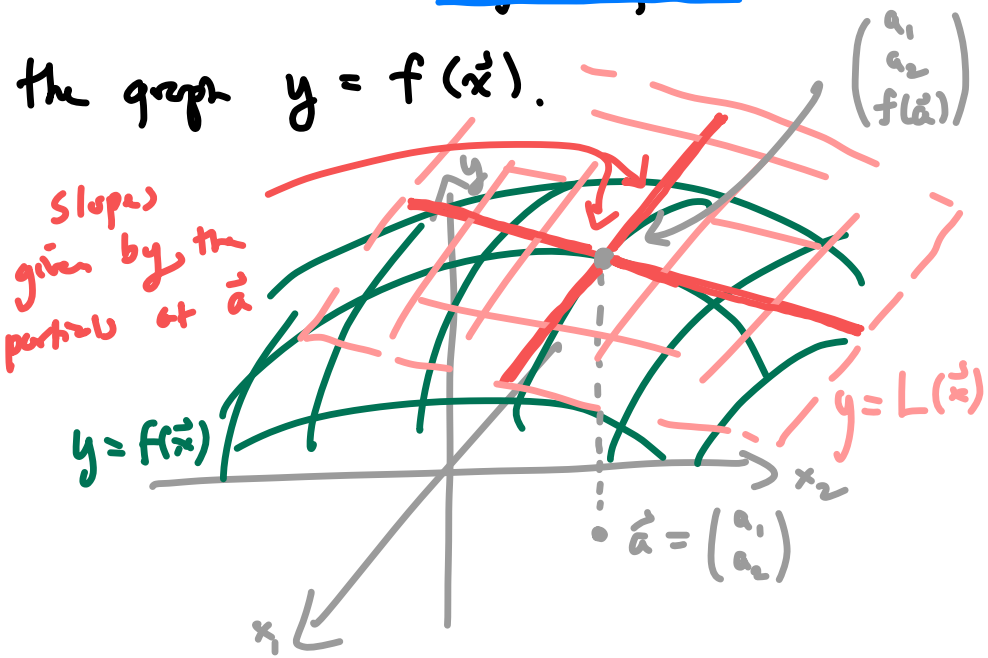
(*) $\left(\dfrac{\partial f}{\partial x_1}(\vec{a}), \ldots, \dfrac{\partial f}{\partial x_n}(\vec{a}) \right)$,

the transpose of the gradient $\vec{\nabla} f(\vec{a})$.

Moreover, the directional derivatives exist and are given by

$$D_{\vec{v}} f(\vec{a}) = T(\vec{v}) = \vec{\nabla} f(\vec{a}) \cdot \vec{v}.$$

The graph of this linear approximation, i.e. $y = L(\vec{x})$ in $\mathbb{R} \times \mathbb{R}^n$, is called the tangent plane to the graph $y = f(\vec{x})$.



Theorem: If f is differentiable at $\vec{a}$, then its partials exist there, and $T = Df(\vec{a})$ has matrix given by

(*) $$\left(\frac{\partial f}{\partial x_1}(\vec{a}), \dots, \frac{\partial f}{\partial x_n}(\vec{a}) \right),$$

the transpose of the gradient $\vec{\nabla} f(\vec{a})$.

Moreover, the directional derivatives exist and are given by

$$D_{\vec{v}} f(\vec{a}) = T(\vec{v}) = \vec{\nabla} f(\vec{a}) \cdot \vec{v}.$$

PROBLEMS ① Find the equation of the tangent plane to $y = f(\vec{x}) = \sqrt{x_1^2 + x_2^2}$ at $\vec{a} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$, and find $D_{\vec{v}} f(\vec{a})$ with $\vec{v} = \begin{pmatrix} -4 \\ 3 \end{pmatrix}$.

② Calculate the directional derivatives (in all directions) of $f\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{cases} \frac{x_1^2 x_2}{x_1^2 + x_2^2}, & \vec{x} \neq \vec{0} \\ 0, & \vec{x} = \vec{0}. \end{cases}$

The linear approximation to f at $\vec{a}$ is given by

$$L(\vec{x}) := f(\vec{a}) + \underbrace{(Df(\vec{a}))(\vec{x} - \vec{a})}_{\text{or } \vec{\nabla} f(\vec{a}) \cdot (\vec{x} - \vec{a})}.$$

It "works" since the definition of the derivative says that, as $\vec{h} \to \vec{0}$,

$$\frac{f(\vec{a} + \vec{h}) - L(\vec{a} + \vec{h})}{\|\vec{h}\|} \left(= E(\vec{h}) \right) \to 0.$$

The take-home from (2) is that one can even have f be continuous at $\vec{0}$ and all its directional derivatives exist at $\vec{0}$, yet fail to be differentiable there. This isn't that surprising, since being well-approximable by a plane means the slopes in various directions have to do more than exist — they have to depend linearly on the direction.

Here is some more evidence that we made the "right" definition:

Theorem: Differentiability at $\vec{a}$ $\Rightarrow$ continuity at $\vec{a}$.

Proof: $0 \leq |f(\vec{a}+\vec{h}) - f(\vec{a})| = |\vec{\nabla} f(\vec{a}) \cdot \vec{h} + \|\vec{h}\| E(\vec{h})|$

$\leq \|\vec{\nabla} f(\vec{a})\| \|\vec{h}\| + \|\vec{h}\| |E(\vec{h})| \to 0$ as $\vec{h} \to \vec{0}$. $\square$

Cauchy-Schwarz + triangle ineq.

While existence of the partials isn't enough to get differentiability (or even continuity!), continuity of the partials is:

Theorem: If $\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}$ exist in a ball about $\vec{a}$, and are continuous at $\vec{a}$, then f is differentiable at $\vec{a}$.

Proof: (WLOG $\vec{a} = \vec{0}$) We need to show

$$(*) \quad \frac{f(\vec{h}) - f(\vec{0}) - \vec{\nabla} f(\vec{0}) \cdot \vec{h}}{\|\vec{h}\|} \to 0 \quad \text{as } \|\vec{h}\| \to 0.$$

Write $\vec{h} = h\vec{u}$, $\vec{u}$ = unit vector, $\vec{v}_k = \sum_{j=1}^{k} u_j \hat{e}_j$

$$\Rightarrow f(h\vec{u}) - f(\vec{0}) = \sum_{k=1}^{n} \{ f(h\vec{v}_k) - f(h\vec{v}_{k-1}) \}$$

($h\vec{v}_{k-1} + hu_k\hat{e}_k$)

MVT

$$= \sum_{k=1}^{n} h u_k \frac{\partial f}{\partial x_k}(\vec{c}_k)$$

($\vec{c}_k$: on segment connecting $h\vec{v}_{k-1}$ to $h\vec{v}_k$)

$$= h \sum_{k=1}^{n} u_k \frac{\partial f}{\partial x_k}(\vec{c}_k) \Rightarrow$$

$(*)$ becomes $\dfrac{h \{ \sum_{k=1}^{n} u_k \left(\frac{\partial f}{\partial x_k}(\vec{c}_k) - \frac{\partial f}{\partial x_k}(\vec{0}) \right) \}}{h \|\vec{u}\|}$, $\|\vec{u}\| = 1$,

which goes to zero by continuity of the $\partial f / \partial x_k$'s — since $h \to 0$ forces $\vec{c}_k \to \vec{0}$. □

All together we arrive at the schematic picture for properties of f:

$C^1 \Rightarrow$ differentiable $\Rightarrow$ (C^0: f continuous)

($\frac{\partial f}{\partial x_i}$ exist + continuous)

$\Downarrow$

$D_{\vec{v}} f$ linear in $\vec{v}$

$\Downarrow$

$D_{\vec{v}} f$ exists $\forall \vec{v}$

$\Downarrow$

$\frac{\partial f}{\partial x_i}$ exist $(\forall i)$

§2. Vector-valued functions

We now turn to the general case.

Consider $\vec{a} \in S \subset \mathbb{R}^n$, $\vec{F}: S \to \mathbb{R}^m$.

Like continuity, differentiability of $\vec{F}$ can be defined to mean that of its component functions $f_1, \dots, f_m$. The components of

$$(*) \quad D_{\vec{v}}\vec{F}(\vec{a}) := \lim_{h \to 0} \frac{\vec{F}(\vec{a}+h\vec{v}) - \vec{F}(\vec{a})}{h}$$

are just the $D_{\vec{v}} f_k(\vec{a})$; while the choice of $T(\vec{v})$ making

$$\vec{E}(\vec{h}) := \frac{\vec{F}(\vec{a}+\vec{h}) - (\vec{F}(\vec{a}) + T(\vec{h}))}{\|\vec{h}\|} \to 0 \quad \text{as } \vec{h} \to \vec{0}$$

agrees with $(*)$ by exactly the same proof as when $m = 1$.

So

$$T(\vec{v}) = D_{\vec{v}}\vec{F}(\vec{a}) = \begin{pmatrix} D_{\vec{v}} f_1(\vec{a}) \\ \vdots \\ D_{\vec{v}} f_m(\vec{a}) \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n \frac{\partial f_1}{\partial x_j}(\vec{a})\, v_j \\ \vdots \\ \sum_{j=1}^n \frac{\partial f_m}{\partial x_j}(\vec{a})\, v_j \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\vec{a}) & \cdots & \frac{\partial f_1}{\partial x_n}(\vec{a}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\vec{a}) & \cdots & \frac{\partial f_m}{\partial x_n}(\vec{a}) \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

$$=: J_{\vec{F}}(\vec{a})\, \vec{v}.$$

Jacobian matrix of $\vec{F}$ at $\vec{a}$

— that is, the linear transformation

$$T: \mathbb{R}^n \to \mathbb{R}^m$$

is represented by the matrix

$$J_F(\vec{a}) = \left[\frac{\partial f_i}{\partial x_j}(\vec{a})\right].$$

It is the linear transformation locally approximating

$$\vec{h} \longmapsto \vec{F}(\vec{a}+\vec{h}) - \vec{F}(\vec{a}).$$

Example: The derivative of a linear function is again the linear function:

$$\vec{F}(\vec{x}) = A\vec{x} = \sum_{j=1}^{n} \vec{a}_j x_j = \begin{pmatrix} \sum_{j=1}^{n} a_{1j} x_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj} x_j \end{pmatrix}$$

($\vec{a}_j$: columns of A)

$$\Downarrow$$

$$f_i(\vec{x}) = \sum_{j=1}^{n} a_{ij} x_j$$

$$\Downarrow$$

$D\vec{F}$ has matrix $\left[\frac{\partial f_i}{\partial x_j}\right] = [a_{ij}] = A$! (at any $\vec{a}$) □

PROBLEM Let $\vec{F}: \mathbb{R}^2 \to \mathbb{R}^2$ be given by $\vec{F}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} xy \\ x^2+y^2 \end{pmatrix}$. Compute $D_{\begin{pmatrix} 3 \\ 4 \end{pmatrix}} \vec{F} \begin{pmatrix} -1 \\ 2 \end{pmatrix}$.

So

$$T(\vec{v}) = D_{\vec{v}}\vec{F}(\vec{a}) = \begin{pmatrix} D_{\vec{v}} f_1(\vec{a}) \\ \vdots \\ D_{\vec{v}} f_m(\vec{a}) \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{n} \frac{\partial f_1}{\partial x_j}(\vec{a})\, v_j \\ \vdots \\ \sum_{j=1}^{n} \frac{\partial f_m}{\partial x_j}(\vec{a})\, v_j \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\vec{a}) & \cdots & \frac{\partial f_1}{\partial x_n}(\vec{a}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\vec{a}) & \cdots & \frac{\partial f_m}{\partial x_n}(\vec{a}) \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

$$=: J_{\vec{F}}(\vec{a})\, \vec{v}.$$

Jacobian matrix of $\vec{F}$ at $\vec{a}$

— that is, the linear transformation

$$D\vec{F}(\vec{a}) := T : \mathbb{R}^n \to \mathbb{R}^m$$

is represented by the matrix

$$J_F(\vec{a}) = \left[\frac{\partial f_i}{\partial x_j}(\vec{a})\right].$$

$D\vec{F}(\vec{a})$ is called the derivative of $\vec{F}$ at $\vec{a}$.

§3. The chain rule

In §3.3 of Shifrin you will find the statements

- $D(\alpha\vec{F}+\beta\vec{G}) = \alpha D\vec{F} + \beta D\vec{G}$

  [linearity of derivatives]

- $D(g\vec{F}) = (Dg)\vec{F} + gD\vec{F}$

i.e. $D(g\vec{F})(\vec{a})\vec{v} = (Dg(\vec{a})\vec{v})\vec{F}(\vec{a}) + g(\vec{a})(D\vec{F}(\vec{a})\vec{v})$

$$D(\vec{F}\cdot\vec{G}) = D\vec{F}\cdot\vec{G} + \vec{F}\cdot D\vec{G}$$

i.e. $D(\vec{F}\cdot\vec{G})(\vec{a})\vec{v} = (D\vec{F}(\vec{a})\vec{v})\cdot G(\vec{a}) + \vec{F}(\vec{a})\cdot(D\vec{G}(\vec{a})\vec{v})$

$$D(\vec{F}\times\vec{G}) = D\vec{F}\times\vec{G} + \vec{F}\times D\vec{G}$$

[Leibniz rules]

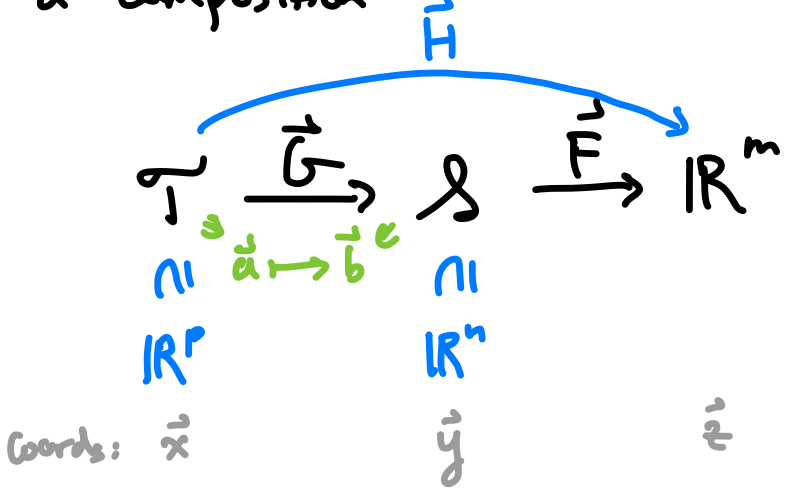
which are tricky to state but easy to prove, and

- $D(\vec{F}\circ\vec{G}) = D\vec{F}\circ D\vec{G}$

i.e. $D(\vec{F}\circ\vec{G})(\vec{a}) = D\vec{F}(\vec{G}(\vec{a}))\circ D\vec{G}(\vec{a})$

which is more interesting — and the one I'll discuss now.

Suppose you are interested in a composition



and want the total derivative

$$D\vec{H}(\vec{a}) : \mathbb{R}^p \to \mathbb{R}^m$$

at $\vec{x} = \vec{a}$.

§3. The chain rule

In §3.3 of Shifrin you will find the statements

- $D(\alpha\vec{F} + \beta\vec{G}) = \alpha D\vec{F} + \beta D\vec{G}$

[linearity of derivatives]

- $D(g\vec{F}) = (Dg)\vec{F} + gD\vec{F}$

i.e. $D(g\vec{F})(\vec{a})\vec{v} = (Dg(\vec{a})\vec{v})\vec{F}(\vec{a}) + g(\vec{a})(D\vec{F}(\vec{a})\vec{v})$

$$D(\vec{F}\cdot\vec{G}) = D\vec{F}\cdot\vec{G} + \vec{F}\cdot D\vec{G}$$

i.e. $D(\vec{F}\cdot\vec{G})(\vec{a})\vec{v} = (D\vec{F}(\vec{a})\vec{v})\cdot\vec{G}(\vec{a}) + \vec{F}(\vec{a})\cdot(D\vec{G}(\vec{a})\vec{v})$

$$D(\vec{F}\times\vec{G}) = D\vec{F}\times\vec{G} + \vec{F}\times D\vec{G}$$

[Leibniz rules]

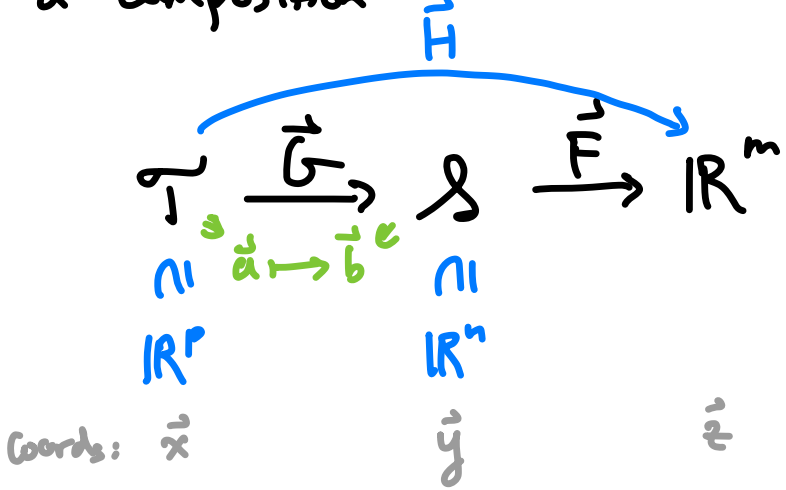
which are tricky to state but easy to prove, and the chain rule

- $D(\vec{F}\circ\vec{G}) = D\vec{F}\circ D\vec{G}$

i.e. $D(\vec{F}\circ\vec{G})(\vec{a}) = D\vec{F}(\vec{G}(\vec{a}))\circ D\vec{G}(\vec{a})$

which is more interesting — and the one I'll discuss now.

Suppose you are interested in a composition



and want the total derivative

$$D\vec{H}(\vec{a}) : \mathbb{R}^p \to \mathbb{R}^m$$

at $\vec{x}=\vec{a}$. If it exists (i.e. $\vec{H}$ is differentiable at $\vec{a}$), this is a linear transformation with the property that

$$\vec{H}(\vec{a}+\vec{u}) - \vec{H}(\vec{a}) = D\vec{H}(\vec{a})\,\vec{u} + \|\vec{u}\|\,\vec{E}_H(\vec{u})$$

where $\vec{E}_H(\vec{u}) \to \vec{0}$ as $\vec{u}\to\vec{0}$. We saw that, moreover, $D\vec{H}(\vec{a})$ is represented by the Jacobian matrix

$$J_H(\vec{a}) = \left[\frac{\partial h_i}{\partial x_j}(\vec{a})\right].$$

Theorem: If $\vec{F}$ is differentiable at $\vec{b}=\vec{F}(\vec{a})$ and $\vec{G}$ is differentiable at $\vec{a}$, then $\vec{H}$ is differentiable at $\vec{a}$, and

$$\begin{cases} D\vec{H}(\vec{a}) = D\vec{F}(\vec{b})\circ D\vec{G}(\vec{a}) & \text{[composition of lin. tr.]} \\ J_H(\vec{a}) = J_F(\vec{b})\cdot J_G(\vec{a}) & \text{[matrix multiplication]} \end{cases}$$

Now

$$\vec{H}(\vec{a}+\vec{u}) - \vec{H}(\vec{a}) = \vec{F}(\vec{G}(\vec{a}+\vec{u})) - \vec{F}(\vec{G}(\vec{a}))$$

(G diff $\Rightarrow$) $\vec{G}(\vec{a}) + D\vec{G}(\vec{a})\vec{u} + \|\vec{u}\|\vec{E}_G(\vec{u})$

($\vec{G}(\vec{a}) = \vec{b}$; call this $\vec{v}$)

$$= \vec{F}(\vec{b}+\vec{v}) - \vec{F}(\vec{b})$$

$$= D\vec{F}(\vec{b})\vec{v} + \|\vec{v}\|\vec{E}_F(\vec{v})$$

$$= D\vec{F}(\vec{b})\big(D\vec{G}(\vec{a})\vec{u}\big) + \|\vec{u}\|\left(D\vec{F}(\vec{b})\vec{E}_G(\vec{u}) + \frac{\|\vec{v}\|}{\|\vec{u}\|}\vec{E}_F(\vec{v})\right)$$

($D\vec{F}(\vec{b})\vec{E}_G(\vec{u}) \to \vec{0}$; $\frac{\|\vec{v}\|}{\|\vec{u}\|}$ bounded by $\bigstar$; $\vec{E}_F(\vec{v}) \to \vec{0}$; as $\vec{u}\to\vec{0}$ ($\Rightarrow \vec{v}\to\vec{0}$))

$$\xrightarrow[\vec{u}\to\vec{0}]{} D\vec{F}(\vec{b})\big(D\vec{G}(\vec{a})\vec{u}\big). \quad \square$$

$$\vec{H}(\vec{a}+\vec{u}) - \vec{H}(\vec{a}) = D\vec{H}(\vec{a})\vec{u} + \|\vec{u}\|\vec{E}_H(\vec{u})$$

where $\vec{E}_H(\vec{u}) \to \vec{0}$ as $\vec{u}\to\vec{0}$. We saw that, moreover, $D\vec{H}(\vec{a})$ is represented by the Jacobian matrix

$$J_H(\vec{a}) = \left[\frac{\partial h_i}{\partial x_j}(\vec{a})\right].$$

Theorem: If $\vec{F}$ is differentiable at $\vec{b} = \vec{F}(\vec{a})$ and $\vec{G}$ is differentiable at $\vec{a}$, then $\vec{H}$ is differentiable at $\vec{a}$, and

$$\begin{cases} D\vec{H}(\vec{a}) = D\vec{F}(\vec{b}) \circ D\vec{G}(\vec{a}) & \text{[composition of lin. tr.]} \\ J_H(\vec{a}) = J_F(\vec{b}) \cdot J_G(\vec{a}) & \text{[matrix multiplication]} \end{cases}$$

Proof: First, $D\vec{G}(\vec{a})\vec{u} = \sum_{k=1}^{n}(\vec{\nabla} g_k(\vec{a})\cdot\vec{u})\hat{e}_k$ ($\vec{\nabla} g_k(\vec{a})$ rows of $D\vec{G}(\vec{a})$)

$\xRightarrow[+\Delta\text{ ineq.}]{\text{C-S}}$ $\|D\vec{G}(\vec{a})\vec{u}\| \le \left(\sum_{k=1}^{n}\|\vec{\nabla} g_k(\vec{a})\|\right)\|\vec{u}\|$

$$\Rightarrow \frac{\|D\vec{G}(\vec{a})\vec{u} + \|\vec{u}\|E_G(\vec{u})\|}{\|\vec{u}\|} \overset{\bigstar}{\le} \sum_{k=1}^{n}\|\vec{\nabla} g_k(\vec{a})\| + E_G(\vec{u})$$

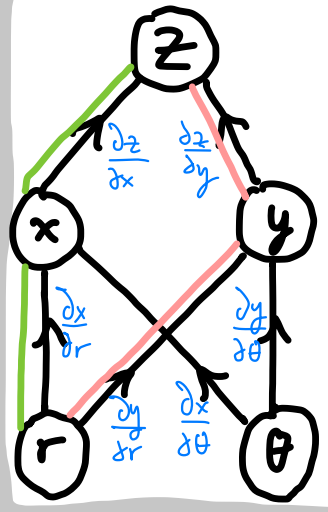
Theorem: If $\vec{F}$ is differentiable at $\vec{b} = \vec{F}(\vec{a})$ and $\vec{G}$ is differentiable at $\vec{a}$, then $\vec{H}$ is differentiable at $\vec{a}$, and

$$\begin{cases} D\vec{H}(\vec{a}) = D\vec{F}(\vec{b}) \circ D\vec{G}(\vec{a}) & \text{[composition of lin. tr.]} \\ J_H(\vec{a}) = J_F(\vec{b}) \cdot J_G(\vec{a}) & \text{[matrix multiplication]} \end{cases}$$

So in matrix form we get

$$\left[\frac{\partial h_i}{\partial x_j}(\vec{a})\right] = \left[\frac{\partial f_i}{\partial y_k}(G(\vec{a}))\right]\left[\frac{\partial g_k}{\partial x_j}(\vec{a})\right],$$

i.e. $\dfrac{\partial h_i}{\partial x_j} = \displaystyle\sum_{k=1}^{n} \frac{\partial f_i}{\partial y_k}\frac{\partial g_k}{\partial x_j}$ for each i, j.



$$\frac{\partial z}{\partial r} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial r}$$

"sum over paths from r to z the products of labels along the path".

Example: Consider the composition

$$\mathbb{R}^2 \xrightarrow{G} \mathbb{R}^2 \xrightarrow{f} \mathbb{R}, \qquad (r,\theta) \mapsto (x,y) \mapsto z, \qquad \tilde{f} = f \circ G$$

where $G(r,\theta) := (r\cos\theta, r\sin\theta)$. Then

- Df means ${}^t\vec{\nabla} f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$,
- $D\tilde{f}$ means ${}^t\vec{\nabla}\tilde{f} = \left(\frac{\partial \tilde{f}}{\partial r}, \frac{\partial \tilde{f}}{\partial \theta}\right)$, and
- $DG = \begin{pmatrix} \partial g_1/\partial r & \partial g_1/\partial \theta \\ \partial g_2/\partial r & \partial g_2/\partial \theta \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$.

The chain rule says

$${}^t\vec{\nabla}\tilde{f} = ({}^t\vec{\nabla} f)\cdot DG$$

$$= \left(\partial f/\partial x, \partial f/\partial y\right)\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

$$= \left(\frac{\partial f}{\partial x}\cos\theta + \frac{\partial f}{\partial y}\sin\theta,\ -\frac{\partial f}{\partial x} r\sin\theta + \frac{\partial f}{\partial y} r\cos\theta\right).$$

A more efficient way to work this out/ memory device is shown at left. □