Differentiability:

$\vec{e}_i^{\,T} = (0,\dots,0,\underset{i}{1},0,\dots)$

$\downarrow$

Recall partial derivatives $\frac{\partial f}{\partial x_i}(\vec{x}^*) = \lim_{h\to 0} \frac{f(\vec{x}^* + h\vec{e}_i) - f(\vec{x}^*)}{h}$

Directional derivative at $\vec{x}^*$: Let $\|\vec{u}\| = 1$.

directional derivative in direction $\vec{u}$:

$$\lim_{h\to 0} \frac{f(\vec{x}^* + h\vec{u}) - f(\vec{x}^*)}{h}$$

Differentiability at x^* is somewhat stronger than existence of the partial / directional derivatives:

Def: If there exists $\vec{p}^{\,T}$ such that

$$\lim_{h\to 0} \left| \frac{f(\vec{x}^* + h\vec{u}) - f(x^*)}{h} - \vec{p}^{\,T}\vec{u} \right| = 0$$

then f is differentiable at x^*.

(and if so: $\vec{p}^{\,T} = \nabla f(\vec{x}^*)$)

In other words:

"The linear approximation is 'good' " ☺

For transformations $\vec{F}$: replace $\vec{p}^{\,T}$ by $\vec{A}$ and $|\cdot|$ by $\|\cdot\|$

There are "ugly" examples that do not behave as "1-variable intuition" suggests.

"Ugly example": $$f(x,y) = \begin{cases} 1 & \text{if } y = x^2 \neq 0 \\ 0 & \text{elsewhere.} \end{cases}$$

At all points where $f(x,y) = 0$, we have
$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0$$ and all directional derivatives $= 0$.

But at $(0,0)$, "the first order approximation could be bad":

$$f(x,y) \approx f(0,0) + \nabla f(0,0)\begin{pmatrix} x \\ y \end{pmatrix} \; ?$$

Put $y = x^2$. It says "$1 \approx 0$" even arbitrarily close to $(0,0)$. Bad!

<u>Fact</u>: If $f \in \underbrace{C'(S)}$, S <u>open</u>

| i.e. the partial 1^{st} derivatives exist
| and are continuous <u>on</u> S

then f differentiable on S.

Differentiation : product & chain rules

Notation :

f, g : real-valued functions

$\vec{F}, \vec{G}$: transformations

$\vec{u}, \vec{v}$: "anything" vector valued : free or intermediate variables

$\vec{p}, \vec{A}$: "parameters" : vector of constants resp matrix of constants

$\frac{\partial \vec{F}}{\partial \vec{x}}$ (note overarrow on both) : the Jacobian

$$\begin{pmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \vdots & \ddots & \\ \frac{\partial F_m}{\partial x_1} & \cdots & \frac{\partial F_m}{\partial x_n} \end{pmatrix} = \begin{pmatrix} \nabla F_1 \\ \vdots \\ \nabla F_m \end{pmatrix}$$

(each row = a gradient)

If $\vec{G} = \vec{G}(\vec{u}, \vec{v})$:

$\frac{\partial \vec{G}}{\partial \vec{u}}$: fix $\vec{v}$ as constant, consider as function of $\vec{u}$, take partial derivatives

---

Note : ∇f = row

Hessian of f = Jacobian of $\overbrace{(\nabla f)^T}^{\text{column, transformation}}$

$$= \begin{pmatrix} \nabla \frac{\partial f}{\partial x_1} \\ \nabla \frac{\partial f}{\partial x_2} \\ \vdots \\ \nabla \frac{\partial f}{\partial x_n} \end{pmatrix}$$

Simplest:

$f(\vec{x}) =$ real constant $\qquad \nabla f(\vec{x}) = \overset{(1\times n)}{0^T}$ if $\vec{x} = (x_1 \dots x_n)^T$

$\vec{F}(\vec{x}) = \vec{q} \in R^m$, constant: $\qquad \frac{\partial \vec{F}}{\partial \vec{x}}(\vec{x}) = \vec{O}_{m\times n}$

$f(\vec{x}) = \vec{p}^T\vec{x} \quad (= \vec{x}^T\vec{p})$, $\vec{p}$ const: $\qquad \nabla f(\vec{x}) = \vec{p}^T$

$\vec{F}(\vec{x}) = \vec{A}\vec{x}$, $\vec{A}$ const. matrix: $\qquad \frac{\partial \vec{F}}{\partial \vec{x}}(\vec{x}) = \vec{A}$.

Chain rules:

$$f(\underset{\substack{\downarrow \\ n\text{-vector}}}{\vec{x}}) = g(\underbrace{\vec{u}(\vec{x})}_{k\text{-vector}}) \qquad \nabla f(\vec{x}) = \underset{1\times k}{\nabla g(\vec{u}(x))} \; \underset{k\times n}{\frac{\partial \vec{u}}{\partial \vec{x}}(\vec{x})}$$

$$\vec{F}(\vec{x}) = \vec{G}(\vec{u}(\vec{x})) \qquad \begin{pmatrix} \nabla F_1(\vec{x}) \\ \vdots \\ \nabla F_n(\vec{x}) \end{pmatrix} = \begin{pmatrix} \nabla G_1(\vec{u}(x)) \\ \vdots \\ \vdots \end{pmatrix} \frac{\partial \vec{u}}{\partial \vec{x}}$$

i.e $\quad \frac{\partial \vec{F}}{\partial \vec{x}} = \frac{\partial \vec{G}}{\partial \vec{u}} \frac{\partial \vec{u}}{\partial \vec{x}}$

$$\vec{F}(\vec{x}) = \vec{G}(\underset{k\text{-vector}}{\vec{u}(\vec{x})}, \underset{l\text{-vector}}{\vec{v}(x)}):$$

$$\frac{\partial F}{\partial \vec{x}} = \underset{m\times k}{\frac{\partial \vec{G}}{\partial \vec{u}}} \; \underset{k\times n}{\frac{\partial \vec{u}}{\partial \vec{x}}} + \underset{m\times l}{\frac{\partial \vec{G}}{\partial \vec{v}}} \; \underset{l\times n}{\frac{\partial \vec{v}}{\partial \vec{x}}}$$

Product rule

Dot: $f(\vec{x}) = \underset{1\times k}{\vec{u}(\vec{x})^T}\ \underset{k\times 1}{\vec{v}(\vec{x})}$

Use the chain rule: $\frac{\partial}{\partial v}\left[\vec{u}^T\vec{v}\right] = \vec{u}^T$

$$\frac{\partial}{\partial u}\left[\vec{u}^T\vec{v}\right] = \vec{v}^T$$

$$\text{so}\quad \boxed{\nabla f = \vec{u}^T\frac{\partial\vec{v}}{\partial\vec{x}} + \vec{v}^T\frac{\partial\vec{u}}{\partial\vec{x}}}\qquad (1\times k)\times(k\times n)$$

- Note here the "switching order".
  Other authors would just let $\frac{\partial\,\vec{\text{row}}}{\partial\vec{x}} =$ "transpose of "our" Jacobian" =

- Example: $\nabla[\underbrace{\vec{x}^T\vec{A}}_{u^T}\underset{\vec{v}}{\vec{x}}] = \vec{x}^T\vec{A}\,\boxed{\vec{I}}$ ← $\vec{v} = \vec{x}$ Jacobian $= \vec{I}$

$$+\ \vec{x}^T\underbrace{\left(\text{Jacobi }\vec{A}^T\vec{x}\right)}_{\vec{A}^T}$$

$$= \vec{x}^T\vec{A} + \vec{x}^T\vec{A}^T$$

(You could also: $(\vec{x}^T+\vec{q})^T\vec{A}(\vec{x}+\vec{q})$

$$= \underbrace{\vec{x}^T\vec{A}\vec{x}}_{\text{function}} + \underbrace{\vec{x}^T\vec{A}\vec{q} + \vec{q}^T\vec{A}\vec{x}}_{\underbrace{\vec{x}^T(\vec{A}+\vec{A}^T)}_{\text{derivative}}\vec{q}} + \underbrace{\vec{q}^T\vec{A}\vec{q}}_{\text{higher order}}$$
)

Scaling: $\vec{F}(\vec{x}) = \overset{\text{number}}{f(\vec{x})}\ \overset{\text{vector}}{\vec{G}(\vec{x})} = \begin{pmatrix} fG_1 \\ \\ fG_n \end{pmatrix}$ $\quad \nabla F_i = \nabla f\ G_i + f\ \nabla G_i$

$$\text{So}\quad \boxed{\frac{\partial\vec{F}}{\partial\vec{x}} = \underset{m\times 1}{\vec{G}}\ \underset{1\times n}{\nabla f} + \underset{\text{scalar}}{f}\ \underset{m\times 1}{\frac{\partial G}{\partial x}}}$$

Example: $f(\vec{x}) = \|\vec{x}\| = \sqrt{\vec{x}^T\vec{x}}$,

Find the gradient vector & Hessian matrix ...

$f = \sqrt{g}$; $g = \vec{x}^T\vec{x}$, $\nabla g(\vec{x}) = 2\vec{x}^T$

$$\nabla f = \frac{1}{2\sqrt{g}} \nabla g = \frac{1}{\sqrt{g}} \vec{x}^T = \underline{\underline{\frac{\vec{x}^T}{\|\vec{x}\|}}}$$

$$\text{Hessian} = \text{Jacobi}\left[\frac{\vec{x}}{\|\vec{x}\|}\right]$$

a scaling $(\vec{x}^T\vec{x})^{-1/2}$ of $\vec{x}$

$\uparrow$ Jacobian $= I$

$$\text{Hessian}[f] = \vec{x}\ \nabla\left[(\vec{x}^T x)^{-1/2}\right] + \frac{1}{\|\vec{x}\|}\vec{I}$$

$$= \frac{1}{\|\vec{x}\|}\vec{I} + \vec{x}\left(\underbrace{-\frac{1}{2}(\vec{x}^T\vec{x})^{-\frac{3}{2}}}_{\text{scaling}} \cdot 2x^T\right)$$

$$= \frac{1}{\|\vec{x}\|}\left(I - \frac{1}{\underset{\|x\|^2}{\vec{x}^T\vec{x}}}\ \underbrace{\vec{x}\vec{x}^T}_{\text{this is }\underline{\text{not}}\ \vec{x}^T\vec{x}\,!}\right)$$

$\vec{x}\vec{x}^T$ is n×n, element (i,j) is $x_i x_j$

Exercise: $f(\vec{x}) = g(\vec{A}\vec{x})$, $\vec{A}$ is m×n.

Find gradient and Hessian.

Example: $\vec{F}$ is $m \times 1$, $\vec{u}$ is $n \times 1$, $n \geq m$,
$\vec{x}$ is $(n-m) \times 1$

$$\vec{F}(\vec{x}, \vec{u}) = \underset{\substack{\uparrow \\ \text{const}}}{\vec{p}}$$

: m eq's.
Can typically determine
m variables: the $\vec{u}$.

Implicit differentiation:

$$\frac{\partial \vec{F}}{\partial \vec{x}} + \frac{\partial \vec{F}}{\partial \vec{u}} \frac{\partial \vec{u}}{\partial \vec{x}} = \vec{0}_{m \times (n-m)}$$

$$\frac{\partial \vec{u}}{\partial \vec{x}} = - \underbrace{\left(\frac{\partial \vec{F}}{\partial \vec{u}} \right)}_{\text{provided this inverse exists.}}^{-1} \frac{\partial \vec{F}}{\partial \vec{x}} \qquad \infty$$

Theory: <u>When</u> can we take $\vec{u} = \vec{u}(\vec{x})$
as a C^1 function?

Answer: that the RHS of ∞ is
well-defined: $\vec{F} \in C^1$ and

$$\left| \frac{\partial \vec{F}}{\partial \vec{u}} \right| \neq 0 .$$

is <u>sufficient</u>.

Special case: Inverse functions

$\vec{y} = \vec{P}(\vec{x})$, both $\vec{y}$ and $\vec{x}$ are n-vectors

Inverting $\Longleftrightarrow$ $\vec{x} = \vec{G}(\vec{y})$

$$\vec{I} = \frac{\partial \vec{F}}{\partial \vec{x}} \frac{\partial \vec{G}}{\partial \vec{y}} \quad \text{yields} \quad \frac{\partial \vec{G}}{\partial \vec{y}} = \left(\frac{\partial \vec{F}}{\partial \vec{x}} \right)^{-1}$$

An inverse exists (locally) as long as this exists.