Matrix Calculus

§ 1 Derivatives as Linear operators

Revisiting : derivative is <u>linear</u> approximation.

- Linearization : (definition of derivative).

$$\delta f = f(x+\delta x) - f(x) \approx f'(x)\delta x + o(\delta x)$$

small change in the output

small change in the input

- Differential notation.

$$df = f(x+dx) - f(x) = f'(x)dx$$ ← arbitarily small.

differential

Definition of derivative :

$$\boxed{\Delta \text{ output} = (\text{linear operator})\ \Delta \text{ input}}$$

Linear operators. given vectors $v \in$ vector space V, denote as $L[v]$ or Lv

acting on

L : vector v in V → vector Lv out.

satisfying.

$$L[kv_1 + lv_2] = kL[v_1] + lL[v_2].\quad \left(\begin{matrix} v_1, v_2 \in V \\ k, l \in \mathbb{F}\end{matrix}\right).$$

Example. 1) L : multiple by scalar α , $Lv = \alpha v$.

2) What about $Ov = 2v+1$, $v \in \mathbb{R}$. ?

"affine"

No! $O(3v) = 6v+1 \neq 3O(v) = 6v+3$.

3) vector space $V = \{ \text{functions } f(x) : \mathbb{R} \to \mathbb{R} \}$.

$L f(x) = 2f(x)$ ··· linear

$L f(x) = df/dx$ ··· linear if differencable

$L f(x) = \int_0^x f(x)dx$ ··· linear

$L\, f(x) = f(x^2)$: take $L(f(x)+g(x)) = f(x^2)+g(x^2)$

$L(2f(x)) = 2\, f(x^2)$.

RECALL. Scalar function $f(\text{vector input } \vec{x}) \in \mathbb{R}^m$

$$df = f(\underbrace{\vec{x}+d\vec{x}}_{\text{arbitrary change}}) - f(\vec{x}) := \underbrace{f'(x)}_{\text{row vect}}\underbrace{d\vec{x}}_{\text{linear operator on } x} = \text{scalar}$$

$$= \underbrace{(\nabla f)}_{\text{gradient of } f.}^T \cdot d\vec{x}$$

Example. Suppose $f(\vec{x}) = \vec{x}^T \underbrace{A}_{\text{const.}} \vec{x}$, $x \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times m}$.

Normal way. $df = f(\vec{x}+d\vec{x}) - f(\vec{x})$

$$= (\vec{x}+d\vec{x})^T A (\vec{x}+d\vec{x}) - \vec{x} A \vec{x}$$

$\vec{x}$ omitting

$$= x^TAx + dx^TAx + x^TAdx + \underbrace{dx^TA\,dx}_{\text{higher order}}$$
$$- x^TAx$$

$$= dx^TAx + x^TAdx.$$

Since dx^TAx is a <u>scalar</u>, a scalar is always equal to its transpose.

$$= x^TA^Tdx + x^TAdx$$

$$= \underbrace{x^T \underbrace{(A+A^T)}_{\text{symmetric}}}_{f'(x) := \nabla(f)^T} dx.$$

RECALL. Inputs and output in more general vec. spaces

Example. <u>matrix</u> inputs and/or outputs.

$$f(A) = A^{-1} \quad \text{or} \quad A^3 \quad \text{or} \quad U \qquad \boxed{\text{Output matrix}}$$

A: $n \times n$ matrix

U: upper triangular matrix, elim

$$f(A) = \det(A) \quad \text{or} \quad \mathrm{tr}(A) \quad \text{or} \quad \sigma_1(A) \qquad \boxed{\text{Scalar output}}$$

σ_1: largest singular value

Example. $f(A) = A^3$, A is matrix.

product rule: $df \boxed{= dA\,A^2 + A\,dA\,A + A^2\,dA}$

$$= f(A + \underbrace{dA}_{n \times n}) - f(A)$$

$$\Rightarrow \underbrace{f'(A)}_{\text{linear operator}}[dA] = dA\,A^2 + A\,dA\,A + A^2 dA \neq 3A^2 dA \text{ unless } dA \text{ and } A \text{ commute}$$

$\neq$ (any matrix)$\cdot dA$

unless we can cut to col sp. $\longrightarrow$

Example. $f(A) = A^{-1}$

We know $A^{-1}A = I$.

$\Rightarrow d(A^{-1}A) = d(I) = 0$, by product rule, LHS is

$= d(A^{-1})\,A + A^{-1}dA$

That is $d(A^{-1})A = A^{-1}dA$. $\xrightarrow{\text{rmul } A^{-1}}$ $d(A^{-1}) = A^{-1}\,dA\,A^{-1}$

$\Rightarrow df = d(A^{-1})\,dA\,A^{-1} = f'(A)\,dA$.

<u>Live experiment</u>.