 2/8/2023

Matrix Calculus
Motivation: Scientific Machine Learning
Automatic Differentiation

Theoretical Linear Algebra Notions:
Vector Spaces + Linear Operators

Vector Space: V
Given $v_1, v_2 \in V$ $v_1 + v_2 \in V$ vector addition
given $c \in \mathbb{R}$, $v \in V$ $cv \in V$ scalar multiplication
other axioms (see wikipedia)

1) Examples: $\mathbb{R}^n = \left\{ x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right\}$
In particular $\mathbb{R}$
2) Polynomials of degree at most n in 1 variable
3) $m \times n$ matrices $\mathbb{R}^{m,n}$
4) $n_1 \times n_2 \times \dots \times n_k$ arrays $\mathbb{R}^{n_1, n_2, \dots, n_k}$
5) Continuous functions: e.g. $(\sin + \cos)(x) = \sin x + \cos x$

Linear Operators (Linear Transformation)
L maps V to W
$L[c_1 v_1 + c_2 v_2] = c_1 L[v_1] + c_2 L[v_2]$ $\forall c_1, c_2 \in \mathbb{R}$ $\forall v_1, v_2 \in V$
"Square Bracket"
1) $V = \mathbb{R}^n$, $W = \mathbb{R}^m$, A is any $m \times n$ matrix

$$L[v] = Av$$

Linear Algebra: All linear transformations can be written as a matrix.
Wrong Impression: Doesn't mean they should!

2) $\mathbb{R}^n$ to $\mathbb{R}$

all linear transform can be written as row vectors

$$L[v] = w^T v = w \cdot v = \sum w_i v_i$$

3) <u>Example</u>

$A \in \mathbb{R}^{m,n}$

$L[A]_{ij}$ = Average of ~~(north(A_{ij}) + east~~ A_{ij}'s four neighbors ($A_{i\pm1}$, $A_{j\pm1}$) (0 if off edge)

* Can be written as $mn \times mn$ matrix with entries 0 & $1/4$.
* Usually best not to.

4) V = diff functions W = continuous functions

$$L(f) = f'$$

5) $V = \mathbb{R}^{n,n}$ $W = \mathbb{R}^{n,n}$, ~~A~~ $A \in \mathbb{R}^{n,n}$ fixed

$$L[X] = AX + XA \quad \text{for all } X \in \mathbb{R}^{n,n}$$

Can be written as $n^2 \times n^2$ matrix if you flatten the input & output

6) $V = W = \mathbb{R}$

$f(x) = ax + b$ is a linear function but not a linear operator

~~Ex~~ all line operators have the form $f(x) = ax$

2 2/8/2023

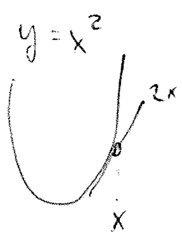
(one acid test if $L[0] \neq 0$ it's not a linear operator)

Scalar calculus

$$f(x+\delta x) - f(x) = f'(x)\,\delta x + o(\delta x)^2 \qquad \delta x \text{ finite perturb}$$

$$df = f'(x)\,dx \qquad dx \text{ infinitesimal}$$

(or informally a small number)



at every x there is a slope
at every x there is a linearization:

$$d(\text{square}) \approx (2x)\,dx$$

$$(3.0001^2 - 3^2) \approx 6\,(.0001)$$

Multiply by 6

Vector to Scalar $\mathbb{R}^n \to \mathbb{R}$

$$df = \nabla f \cdot dx = (\nabla f)^T dx$$

$$(\nabla f)_i = \frac{\partial f}{\partial x_i}$$

e.g. $f(x) = x^T x = \sum x_i^2$

$$\frac{\partial f}{\partial x_i} = 2x_i \qquad \nabla f = 2x$$

$$df = 2x^T dx$$

$$\nabla f = 2x$$

$$df = f'(x)\,dx \qquad f'(x) = 2x^T$$

$$\cancel{f'(x)[dx] = 2x^T dx}$$

$$\cancel{f'(x) = 2x}$$

Lin operator: $f'(x)[\delta x] = 2x^T(\delta x)$

Vector to Vector

$$f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} \qquad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$\begin{pmatrix} df_1 \\ df_2 \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{pmatrix} \begin{pmatrix} dx_1 \\ dx_2 \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} dx_1 + \frac{\partial f_1}{\partial x_2} dx_2 \\ \frac{\partial f_2}{\partial x_1} dx_1 + \frac{\partial f_2}{\partial x_2} dx_2 \end{pmatrix}$$

2×2 Jacobian

In general $f(x)$ take $\mathbb{R}^n$ to $\mathbb{R}^m$

$$df = \underbrace{f'(x)}_{\substack{m \times n \\ \text{matrix}}} \, dx$$

Note! $m = 1$, $f'(x) = 1 \times n = (\nabla f)^T$

(3)

2/8/2023

Rules

$f(x) = g(x) \pm h(x)$ $\qquad df = dg \pm dh$

$f' = g' + h'$

Product $\qquad df = dg\,h + g\,dh$

$f'(x)[dx] = g'(x)[dx]\,h(x) + g(x)\,h'(x)[dx]$

$(\text{not } (g'h + gh')dx)$

Chain Rule

$f(x) = g(h(x))$

$df = f'(x)[dx] = \cancel{g'(}h'($

$g'(h(x))[h'(x)[dx]]$

$f'(x) = \underbrace{g'(h(x)) \quad h'(x)}_{\text{composition}}$

$$d(A^2) = A\,dA + dA\,A$$

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

$$d(A^3) = A^2\,dA + A\,dA\,A + dA\,A^2$$