Matrix Calculus

§3.1 Kronecker Products and Jacobians

Q: Let $f(x) = \|x\|_2$ from $\mathbb{R}^n \to \mathbb{R}$

$= \sqrt{x_1^2 + \cdots + x_n^2}$. Ask for $\nabla_x f$.

Let $r = \|x\|$, $r^2 = x^T x$, take differential.

$2r\,dr = 2x^T dx$, $dr = \frac{x^T}{r} dx$,

$\nabla_x r = \frac{x}{r}$.

Recall. Jacobian is the ~~graph~~ mat that $[x_1 \; x_2 \; x_3]$

$$\begin{bmatrix} \frac{\partial x_1}{\partial x_1} & \frac{x_2}{\partial x_2} & \frac{x_3}{\partial x_3} \\ \vdots & \vdots & \\ \frac{x_3}{\partial x_1} & \frac{x_3}{\partial x_2} & \frac{x_3}{\partial x_3} \end{bmatrix}.$$

Never construct Jacobian, use linear transformation instead.

1. Kronecker Products

Example. $\begin{pmatrix} a \\ b \end{pmatrix} \otimes \begin{pmatrix} p & r \\ q & s \end{pmatrix} = \begin{pmatrix} \boxed{\begin{matrix} ap & aq \\ ar & as \end{matrix}} \\ \boxed{\begin{matrix} bp & bq \\ br & bs \end{matrix}} \end{pmatrix}$

$$\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} \otimes \begin{pmatrix} \triangledown & \heartsuit \\ @ & \heartsuit \end{pmatrix} = \begin{pmatrix} a\boxed{\begin{matrix} \triangledown & \heartsuit \\ @ & \heartsuit \end{matrix}} & b\square & c\square \\ d\square & e\square & f\square \end{pmatrix}.$$

$$I_2 \otimes \begin{bmatrix} p & q \\ r & s \end{bmatrix} = \begin{bmatrix} \begin{matrix} p & q \\ r & s \end{matrix} & 0 \\ 0 & \begin{matrix} p & q \\ r & s \end{matrix} \end{bmatrix}$$

$$\begin{bmatrix} p & r \\ q & s \end{bmatrix} \otimes I_2 = \begin{bmatrix} p & 0 & q & 0 \\ 0 & p & 0 & q \\ r & 0 & s & 0 \\ 0 & r & 0 & s \end{bmatrix}$$

- Jacobian is a function $I_2 \otimes X + X^T \otimes I_2$.
- key kronecker identify

$$(A \otimes B)^T \cdot \mathrm{vec}(C) = \mathrm{vec}(BCA^T)$$

• Useful identities

a) $(A \otimes B)^T = A^T \otimes B^T$

b) $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$

c) $\det(A \otimes B) = \det(A)^m \det(B)^n$, $A \in \mathbb{R}^{mn}$, $B \in \mathbb{R}^{m \times m}$

d) $\operatorname{trace}(A \otimes B) = \operatorname{trace} A \operatorname{trace} B$

e) $A \otimes B$ is orthogonal if A, B are orthognal.

f) $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$

g) If $Au = \lambda u$, $Bv = \mu v$, if $X = vu^T$,
$BXA^T = \lambda\mu X$, also $AX^TB^T = \lambda\mu X^T$,
$A \otimes B$ and $B \otimes A$ have same eigenvals.
and transposed eigenvecs.

Eg. $d(X^2) = XdX + dX\,X$
$= XdXI + IdX\,X$

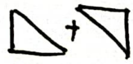
$\operatorname{vec}(dX^2) = (I \otimes X + X^T \otimes I)\operatorname{vec}(dX)$. ⊃ by key identity.

↓ $I \otimes X + X^T \otimes I$ is a linear op. i.e.

$dX^2 = (I \otimes X + X^T \otimes I)[dX]$ $\qquad (A \otimes B)[M] \equiv BMA^T$

Eg. $d(X^3) = X^2dX + XdXX + dXX^2$.
$$= \underbrace{(I \otimes X^2 + X^T \otimes X + (X^T)^2 \otimes I)}_{\text{operator}} \cdot [dX]$$

2. Example. LU decomposition, $\qquad A = LU$

no other than $\mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$ functions. n^2  $\to n^2$

prod rule: $d(LU) = LdU + dL\,U$.
$= (I \otimes L + U^T \otimes I)[dL, dU]$

$dL = \begin{pmatrix} 0 & 0 \\ dL_{11} & 0 \end{pmatrix}$ $\quad dU = \begin{pmatrix} dU_{11} & dU_{12} \\ 0 & dU_{22} \end{pmatrix}$.

$\operatorname{vec} = \begin{pmatrix} dl_{11} \\ du_{11} \\ du_{12} \\ \vdots \end{pmatrix}$.

3. Example. Traceless symmetric eigenproblem.

$$S=\begin{bmatrix} p & s \\ +s & -p \end{bmatrix} \qquad \Lambda=\begin{bmatrix} r & 0 \\ 0 & -r \end{bmatrix}. \qquad Q=\begin{bmatrix} \cos(\frac{1}{2}\theta) & -\sin(\frac{1}{2}\theta) \\ \sin(\frac{1}{2}\theta) & \cos(\frac{1}{2}\theta) \end{bmatrix}$$

$$Q\Lambda Q^T=\begin{bmatrix} r\cos\theta & r\sin\theta \\ r\sin\theta & -r\cos\theta \end{bmatrix}.$$

Catesian → Polar

4. Example. full 2x2 symmetric eigenproblem

$$\begin{pmatrix} p & s \\ s & r \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}^T$$

$\lambda_1, \lambda_2, \theta \rightarrow p, r, s.$ characteristic, polimomial, etc.

$\leftarrow$ matrix multiplication.

The det of $\lambda_1 - \lambda_2$ will be the scale factor.