Differentiating functions

$$\underline{x} \to \underline{h}^{(1)} \to \underline{h}^{(2)} \to \cdots \to \underline{f} \to E$$

$$\begin{tikzcd}
 & A \arrow[d] & \\
 & B \arrow[dl] \arrow[dr] & \\
C \arrow[dr] & & D \arrow[dl] \\
 & E \arrow[d] & \\
 & f &
\end{tikzcd}$$

If $E = CD$ if I have $M \times M$ matrices

That costs $O(M^3)$

Chain Rule of Differentiation

$$f(x,y) = x^2 y$$

$$\frac{\partial f}{\partial y} = x^2 \quad , \quad \frac{\partial f}{\partial x} = 2xy$$

Example

$$\left.\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \end{aligned}\right] \rightarrow f(\theta, r) = r^3 \cos^2\theta \sin\theta$$

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

$$\frac{\partial f}{\partial A_{ij}} = \sum_{k,l,m,n,o,p,q,r} \frac{\partial B_{kl}}{\partial A_{ij}} \left(\frac{\partial D_{mn}}{\partial B_{kl}} \frac{\partial E_{qr}}{\partial D_{mn}} + \underbrace{\frac{\partial C_{op}}{\partial B_{kl}}}_{} \frac{\partial E_{qr}}{\partial C_{op}} \right) \frac{\partial f}{\partial E_{qr}}$$

4-dimensional array

could be $M \times M \times M \times M$, $O(M^4)$

Reverse-Mode Differentiation or Backpropagation

Notation: (Giles 2008)

If we're computing a scalar, eg. an error E

If we have some intermediate term X

We want $\bar{X}$ same size as X

$$\bar{X}_{ij} = \frac{\partial E}{\partial X_{ij}}$$

Start at end. For example output f

$$E = (f-y)^2 \qquad \frac{\partial E}{\partial f} = 2(f-y)$$

$$\bar{f} = 2(f-y)$$

Backpropagate

Compute $\bar{X}$ for parents of quantities have derivative signals for.

Example Matrix Multiplication

$$Z = XY$$

$$Z_{mn} = \sum_p X_{mp} Y_{pn}$$

upstream computation of scalar E

$$\bar{Z}_{mn} = \frac{\partial E}{\partial Z_{mn}} \quad \text{want} \quad \frac{\partial E}{\partial X_{ij}} = \bar{X}_{ij}, \quad \frac{\partial E}{\partial Y_{kl}} = \bar{Y}_{kl}$$

Giles (2008) describes neat linear algebra to find $\bar{X}, \bar{Y}$. Here we use brute force:

$$\frac{\partial E}{\partial X_{ij}} = \sum_{mn} \underbrace{\frac{\partial E}{\partial Z_{mn}}}_{\bar{Z}_{mn}} \underbrace{\frac{\partial Z_{mn}}{\partial X_{ij}}}_{\delta_{mi} Y_{jn}} = \sum_n \bar{Z}_{in} Y_{jn}$$

$$\boxed{\bar{X} = \bar{Z} Y^T}$$

$$\rightarrow \quad Z_{mn} = X_{m2} Y_{2n} + X_{m5} Y_{5n} + \dots$$

$$\dots \underset{\substack{\| \\ m}}{X_{ij}} Y_{jn} + \dots$$

$$\delta_{mi} = \begin{cases} 1 & m = i \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial E}{\partial Y_{kl}} = \sum_{m,n} \underbrace{\frac{\partial E}{\partial Z_{mn}}}_{\bar{Z}_{mn}} \underbrace{\frac{\partial Z_{mn}}{\partial Y_{kl}}}_{X_{mk}\, \delta_{nl}}$$

$$= \sum_{m} \bar{Z}_{ml} X_{mk}$$

$$\boxed{\bar{Y} = X^T \bar{Z}}$$

Elementwise Functions

$$\to \underline{a} \to \boxed{} \to \underline{h} \to \qquad h_k = g(a_k)$$

$$\leftarrow \bar{\underline{a}} \leftarrow \boxed{} \leftarrow \bar{\underline{h}} \leftarrow$$ upstream computation of scalar E

$$\bar{a}_i = \frac{\partial E}{\partial a_i} = \frac{\partial E}{\partial h_i}\frac{\partial h_i}{\partial a_i}$$

$$= \bar{h}_i \, g'(a_i)$$

[where $g'(z) = \left.\frac{\partial g(a)}{\partial a}\right|_{a=z}$

$$\bar{\underline{a}} = \bar{\underline{h}} \odot g'(\underline{a})$$

Elementwise product
Hadamard " "

Or .* in Matlab
(* in Python)