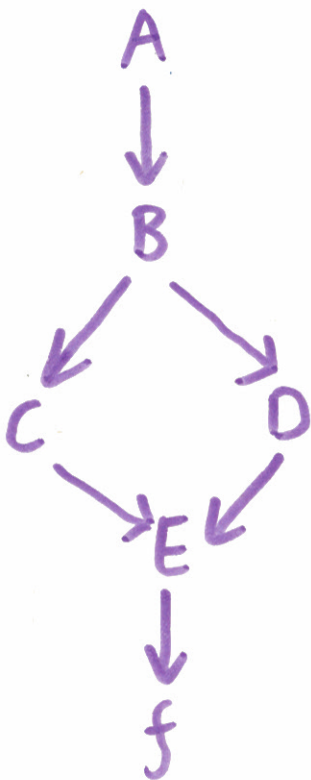


# Differentiating functions



If  $E = CD$  if I have  $M \times M$  matrices  
That costs  $O(M^3)$

# Chain Rule of Differentiation

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

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

Example

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

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

---

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

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  $X$

$$\bar{X}_{ij} = \frac{\partial E}{\partial X_{ij}} \quad \text{~~the~~$$

Start at end. For example output  $f$

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

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

Backpropagate

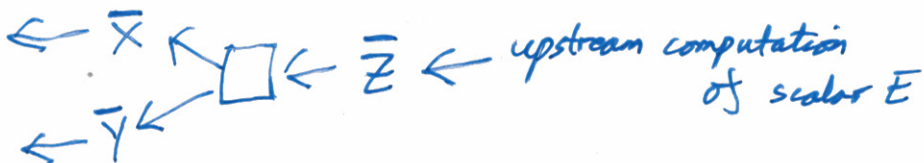
Compute  $\bar{X}$  for parents of quantities have derivative signals for.

# Example Matrix Multiplication



$$Z = XY$$

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



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

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

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

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

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

$$\dots X_{ij} Y_{jn} + \dots$$

$\parallel$   
 $m$

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

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

$$= \sum_m \bar{Z}_{ml} X_{mk}$$

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

# Elementwise Functions

$$\rightarrow \underline{a} \rightarrow \boxed{\text{S}} \rightarrow \underline{h} \rightarrow h_k = g(a_k)$$

$$\leftarrow \underline{\bar{a}} \leftarrow \boxed{\text{ }} \leftarrow \underline{\bar{h}} \leftarrow \begin{array}{l} \text{upstream} \\ \text{of} \end{array} \begin{array}{l} \text{computation} \\ \text{of} \end{array} \text{ scalar } E$$

$$\bar{a}_i = \frac{\partial E}{\partial a_i} = \frac{\partial E}{\partial h_i} \frac{\partial h_i}{\partial a_i} \quad \left[ \text{where } g'(z) = \frac{\partial g(a)}{\partial a} \Big|_{a=z} \right]$$
$$= \bar{h}_i g'(a_i)$$

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

⌞ Elementwise product  
Hadamard " "

Or .\* in Matlab  
(\* is Python)