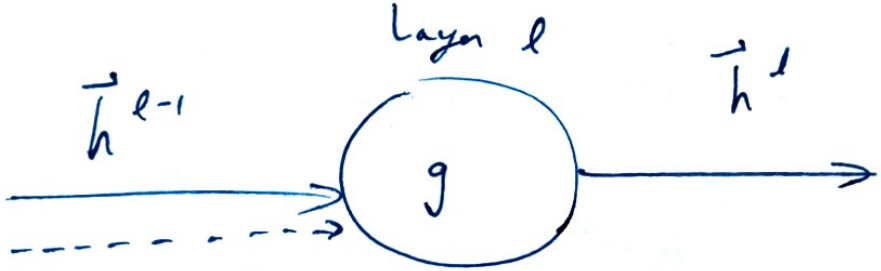
Forward Mode

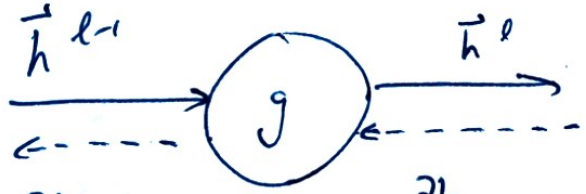


Want $\frac{\partial L}{\partial x}$

$\frac{\partial \vec{h}^{(\ell-1)}}{\partial x}$ — wrt input

$\frac{\partial \vec{h}^{\ell}}{\partial \vec{h}^{\ell-1}} \cdot \frac{\partial \vec{h}^{\ell-1}}{\partial x} = \underbrace{\frac{\partial \vec{h}^{\ell}}{\partial x}}_{\text{output}}$

Reverse Mode



$\frac{\partial L}{\partial \vec{h}^{\ell-1}}$ $\quad$ $\frac{\partial L}{\partial \vec{h}^{\ell}}$

$\frac{\partial L}{\partial \vec{h}^{\ell}} \cdot \frac{\partial \vec{h}^{\ell}}{\partial \vec{h}^{\ell-1}}$

Key Difference:

- Forward propagate derivative output wrt input (At end output is loss)
  - No need for 2 passes, but # passes per input
- Backward propagate derivative of Loss wrt. input
  - Increased storage in proportion to # ops

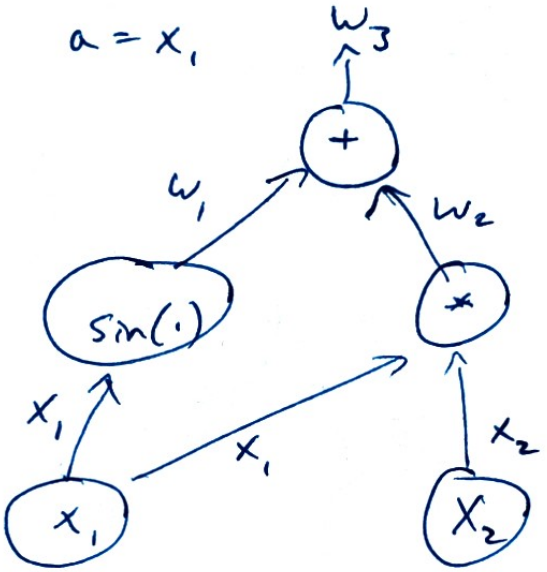
Backward mode preffered for $f: \mathbb{R}^n \to \mathbb{R}^m$ $n \gg m$

Example Forward Mode AD

$$f(x_1, x_2) = x_1 x_2 + \sin(x_1) \qquad \frac{\partial f}{\partial x} = \left[\frac{\partial f}{x_1} \; \frac{\partial f}{x_2}\right]$$

Pick one input: $a = x_1$

Label intermediate



Add - distributor

Max - router

Mul - switcher

Calculate derivative of each variable w.r.t x_1 $(= a)$

$\left(= \frac{\partial x_1}{\partial a}\right)$ $\quad \dot{x}_1 = \frac{\partial x_1}{\partial a} = 1$

$\dot{x}_2 = \frac{\partial x_2}{\partial a} = 0$

$\dot{w}_1 = \frac{\partial w_1}{\partial a} = \frac{\partial w_1}{\partial x_1} \cdot \frac{\partial x_1}{\partial a} = \cos(x_1)\, \dot{x}_1$

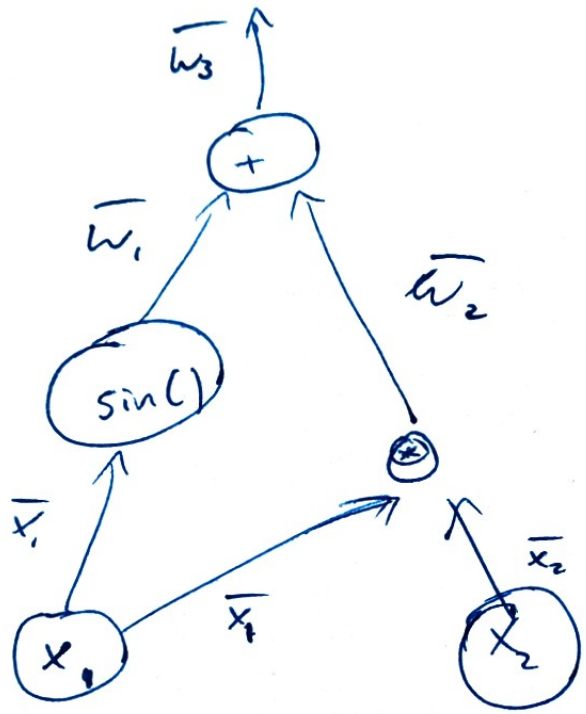
$\dot{w}_2 = \dot{x}_1 x_2 + x_1 \dot{x}_2$

$\dot{w}_3 = \dot{w}_1 + \dot{w}_2$

$$\left[\frac{\partial w_2}{\partial a} = \frac{\partial w_2}{\partial x_1} \cdot \frac{\partial x_1}{\partial x_1} + \frac{\partial w_2}{\partial x_2} \cdot \frac{\partial x_2}{\partial x_1}\right]$$

$$\left[\frac{\partial w_3}{\partial w_1} \cdot \frac{\partial w_1}{\partial x_1} + \frac{\partial w_3}{\partial w_2} \cdot \frac{\partial w_2}{\partial x_1}\right]$$

Reverse Mode AD



$$\bar{w}_3 = \frac{\partial f}{\partial w_3} = 1$$

Notice: Can choose order of evaluation!

$$\bar{w}_1 = \bar{w}_3 \qquad \frac{\partial f}{\partial w_1} = \frac{\partial f}{\partial w_3} \cdot \frac{\partial w_3}{\partial w_1} \cdot 1$$

$$\bar{w}_2 = \bar{w}_3$$

$$\bar{x}_1 = \bar{w}_1 \cdot \cos(x_1) \qquad \frac{\partial f}{\partial w_1} \cdot \frac{\partial w_1}{\partial x_1} \quad \bar{w}_1 \cdot \cos(x_1)$$

$$\bar{x}_1 = \bar{w}_2 \, x_2$$

$$\bar{x}_2 = \bar{w}_2 \, x_1 \qquad \frac{\partial f}{\partial x_2} = \frac{\partial f}{\partial w_2} \cdot \frac{\partial w_2}{\partial x_2} = \bar{w}_2 \cdot x_1$$

Duplicate x_1

Computes gradients for **all** inputs!

Add - distributor

Max - router

Multiply - switcher