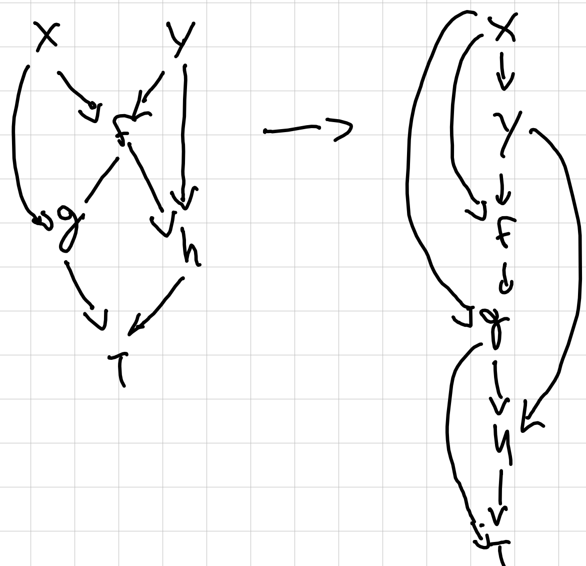
Take a computational graph and transform it to a graph in topological order. Example



$$T(x,y) = T[g(x,f(x,y)),h(f(x,y),y)]$$

Now it is easier to traverse the tree (going from x to T or from T to x).

So, taking the derivative (eg $\frac{dT}{dY}$) is simple, since you can just look one node at a time.

To compute the derivatives, we traversing the graph in reverse topological order, and accumulate derivatives. Add possible derivatives, start at zero.

$$\frac{\partial T}{\partial V} = D_V = 0 \quad \forall V \in [T, h, g, f, x, y]$$

Then, we set $\underline{D_T = 1} = \frac{\partial T}{\partial T}$

Then, we going through the input nodes of T, and we start accumulating derivatives

In this example, we accumulate the derivatives wrt g and h:

$$D_g = \frac{\partial T}{\partial g} \times \underbrace{D_T}_{\partial T/\partial T}, \; D_h = \frac{\partial T}{\partial h} \times D_T$$

Then, we go to the next node in reverse topological order, h. Here we look at its inputs and accumulate

$$D_f = \frac{\partial h}{\partial f} \times D_h = \frac{\partial h}{\partial f} \frac{\partial T}{\partial h}$$

$$D_y = \frac{\partial h}{\partial y} D_h = \frac{\partial h}{\partial y} \frac{\partial T}{\partial h}$$

$\Bigg\{$ $\frac{\partial h}{\partial y}$ is "local derivative" so, $D_y \neq \partial T/\partial y$. We need to get other contributions

After h, we go to g, and compute

$$D_x = \frac{\partial g}{\partial x} \times D_g$$

$$D_f = D_f + \frac{\partial g}{\partial f} \times D_g = \frac{\partial h}{\partial f}\frac{\partial T}{\partial h} + \frac{\partial g}{\partial f}\frac{\partial T}{\partial g} = \frac{\partial T}{\partial f}!$$

Notice that we encounter a node as an input to another node for multiple times, we update its D_v, by adding the previous and new contributions. The reason is simple; if a node is an input in multiple nodes, then the function is of the form.

$$T(h(f, \ldots), g(f, \ldots), \ldots)$$

So, in order to correctly capture the (partial) derivative wrt f, we do (schematically)

$$\frac{\partial T}{\partial f} = \frac{\partial T}{\partial h}\frac{\partial h}{\partial f} + \frac{\partial T}{\partial g}\frac{\partial g}{\partial f}$$

So, we add all the contributions coming from the nodes h and g.

After g, we go to f and do

$$D_x = D_x + \frac{\partial f}{\partial x} D_f = \frac{\partial g}{\partial x}\frac{\partial T}{\partial g} + \frac{\partial f}{\partial x}\frac{\partial T}{\partial f} = \frac{\partial T}{\partial x}!$$

$$D_y = D_y + \frac{\partial f}{\partial y} D_f = \frac{\partial h}{\partial y}\frac{\partial T}{\partial h} + \frac{\partial f}{\partial y}\frac{\partial T}{\partial f} = \frac{\partial T}{\partial y}!$$