Today

1. optimization
2. differentiation

Allen-Cahn

$$
\begin{aligned}
u_t - 0.0001u_{xx} + 5u^3 - 5u &= 0 \\
u(0,x) &= x^2\cos(\pi x) \\
u(t,-1) &= u(t,1) \\
u_x(t,-1) &= u_x(t,1)
\end{aligned}
$$

Let $f(x,t) = u_t - 0.0001\,u_{xx} + 5u^3 - 5u$

let L = objective or Loss

$$= \frac{1}{N}\sum_{i=1}^{N} \left| f(x^i, t^i) \right|$$

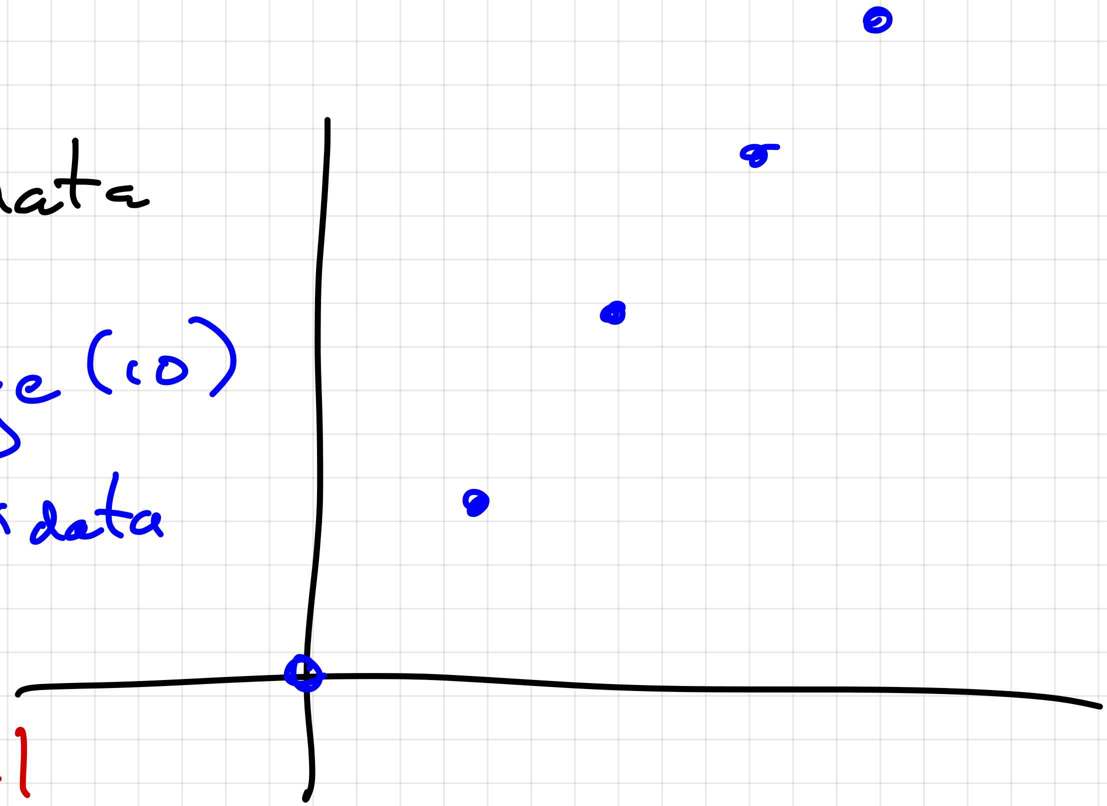
$$x,\ t \to \boxed{\text{model}} \to u \to u_t,\ u_x,\ u_{xx} \to L$$

Demo

Have some data

x_{data} = arange(10)

y_{data} = 2 * xdata



Goal: find model

$$y = wx + b$$

Use an Error:

let $\hat{w}$ be a guess at w

$\hat{b}$

let $\hat{y} = \hat{w}x + \hat{b}$

$$E = \frac{1}{N}\sum(\hat{y} - y_{data})^2$$

In this case

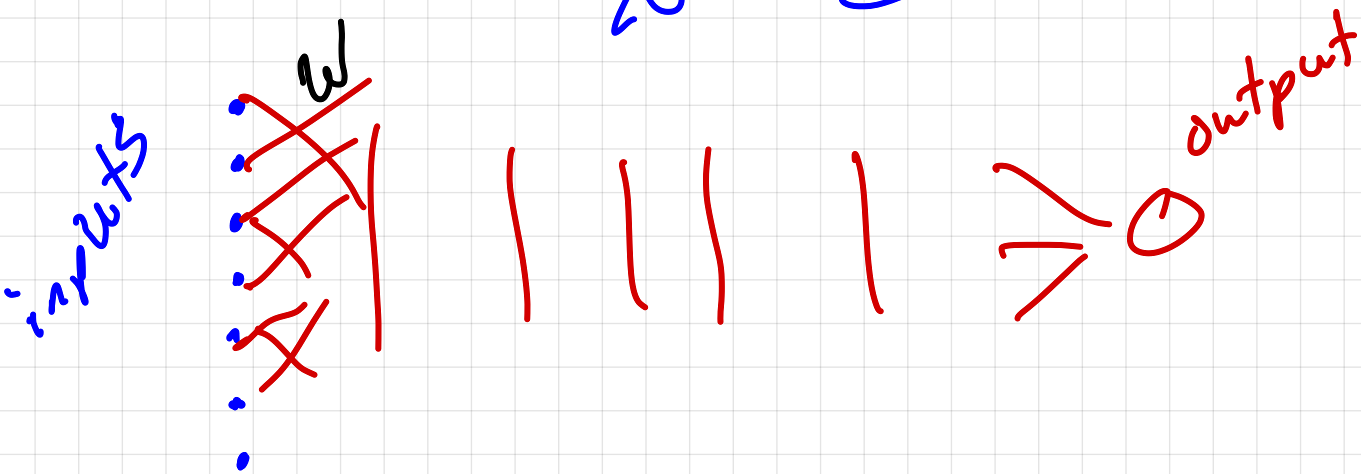
data

$$L = \frac{1}{N} \sum_{n=1}^{N} \left(\hat{y}_n - y_n \right)^2$$

$$= \frac{1}{N} \sum_{n=1}^{N} \left(w x_n - y_n \right)^2$$

$$\frac{\partial L}{\partial w} = \frac{1}{N} \sum_{n=1}^{N} \frac{\partial}{\partial w} \left(w x_n - y_n \right)^2$$

$$= \frac{1}{N} \sum 2 \left(w x_n - y_n \right) x_n$$

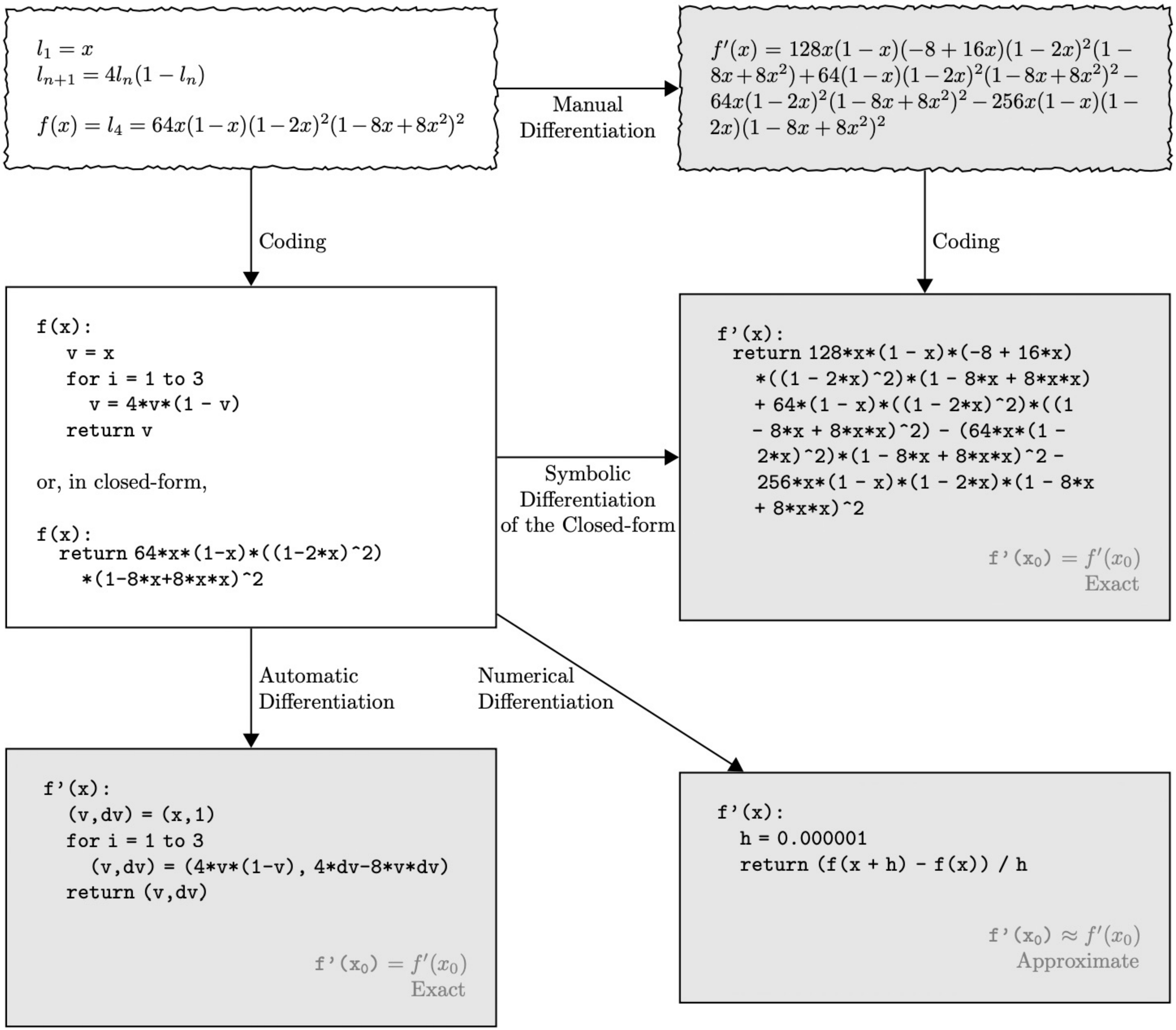


Many ways

1. symbolic diff : $x^2 \to 2x$

2. numerical diff : $\dfrac{u(x+h) - u(x-h)}{2h}$

3. automatic diff :



l1 = x
ln+1 = 4ln(1 − ln)
f(x) = l4 = 64x(1−x)(1−2x)^2(1−8x+8x^2)^2
Manual Differentiation
f′(x) = 128x(1 − x)(−8 + 16x)(1 − 2x)^2(1 − 8x+8x^2)+64(1−x)(1−2x)^2(1−8x+8x^2)^2 − 64x(1−2x)^2(1−8x+8x^2)^2 − 256x(1−x)(1 − 2x)(1 − 8x + 8x^2)^2
Coding
Coding
f(x):
v = x
for i = 1 to 3
v = 4*v*(1 - v)
return v
or, in closed-form,
f(x):
return 64*x*(1-x)*((1-2*x)^2)
*(1-8*x+8*x*x)^2
Symbolic Differentiation of the Closed-form
f'(x):
return 128*x*(1 - x)*(-8 + 16*x)
*((1 - 2*x)^2)*(1 - 8*x + 8*x*x)
+ 64*(1 - x)*((1 - 2*x)^2)*((1
- 8*x + 8*x*x)^2) - (64*x*(1 -
2*x)^2)*(1 - 8*x + 8*x*x)^2 -
256*x*(1 - x)*(1 - 2*x)*(1 - 8*x
+ 8*x*x)^2
f'(x0) = f′(x0)
Exact
Automatic Differentiation
Numerical Differentiation
f'(x):
(v,dv) = (x,1)
for i = 1 to 3
(v,dv) = (4*v*(1-v), 4*dv-8*v*dv)
return (v,dv)
f'(x0) = f′(x0)
Exact
f'(x):
h = 0.000001
return (f(x + h) - f(x)) / h
f'(x0) ≈ f′(x0)
Approximate

Back to the chain rule:

$$y = f(g(h(\underline{x})))$$

$w_0 = x$

$w_1 = h(w_0)$

$w_2 = g(w_1)$

$w_3 = f(w_2) = y$

$$\frac{dy}{dx} = \frac{dy}{dg}\frac{dg}{dx}$$

$$= \frac{dy}{dg} \cdot \frac{dg}{dh} \cdot \frac{dh}{dx}$$

$$= \frac{dy}{dw_2} \cdot \frac{dw_2}{dw_1} \frac{dw_1}{dx}$$

$$y = e^{\sin(x^2)}$$

let $w_0 = x$

$w_1 = w_0^2$

$w_2 = \sin(w_1)$

$w_3 = e^{w_2}$

$$\frac{dy}{dx} = e^{w_2} \cos(w_1)\, 2w_0$$

$$\frac{dy}{dx} = \frac{dy}{dw_2} \cdot \frac{dw_2}{dw_1} \cdot \frac{dw_1}{dx}$$

3 2 1

Forward accumulation: inside-out

1 2 3

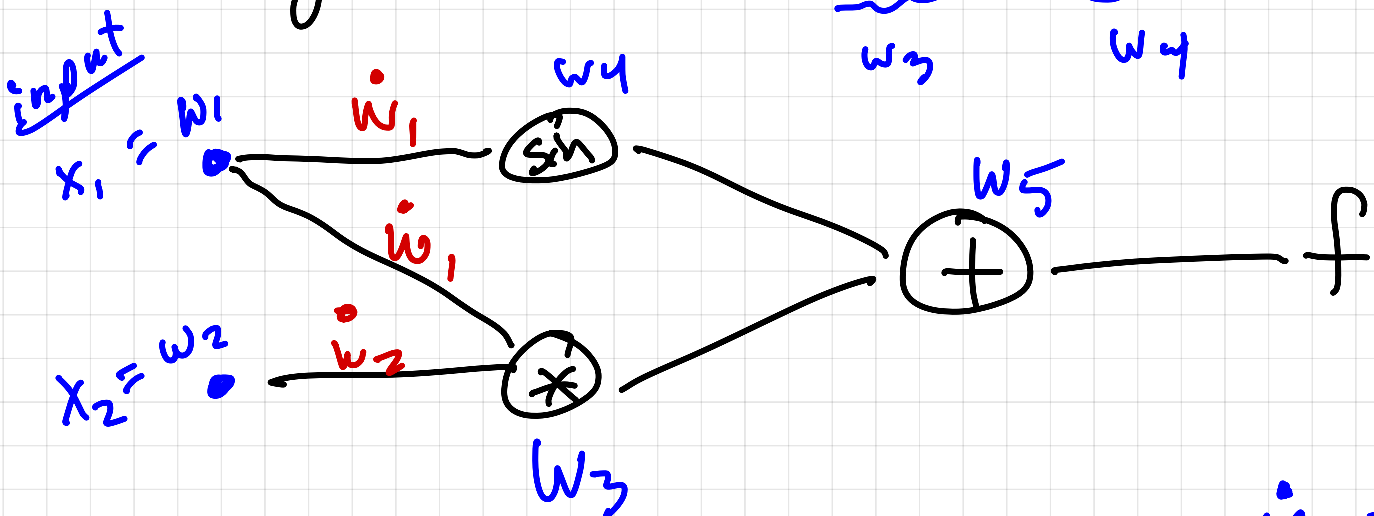
Reverse accum. outside in

Example

$$y = f(x_1, x_2) = \underbrace{x_1 x_2}_{w_3} + \underbrace{\sin(x_1)}_{w_4}$$

$$\dot{w}_i = \sum_j \frac{\partial w_i}{\partial w_j} \dot{w}_j$$

$\dot{j}$ are input



$$w_1 = x_1$$

$$w_2 = x_2$$

$$w_3 = w_1 * w_2$$

$$w_4 = \sin(w_1)$$

$$w_5 = w_3 + w_4$$

$$\dot{w}_1 = 1$$

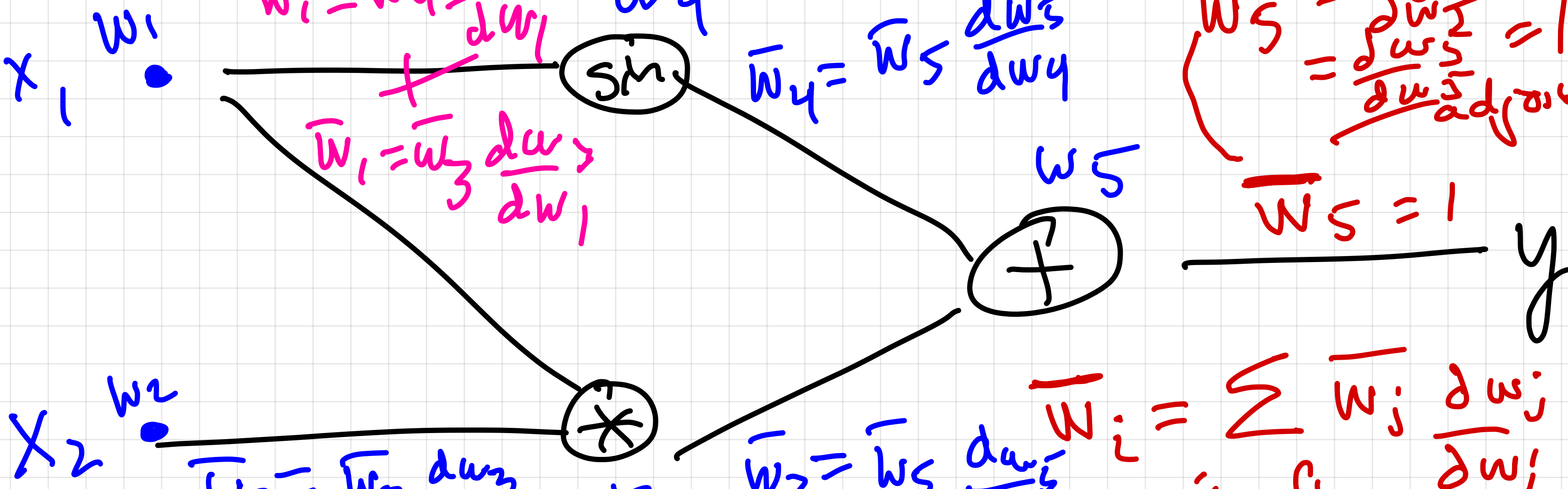
$$\dot{w}_2 = 0$$

$$\dot{w}_3 = \frac{\partial w_3}{\partial w_1} \dot{w}_1 + \frac{\partial w_3}{\partial w_2} \cdot \dot{w}_2$$

$$= w_2 \dot{w}_1 + w_1 \dot{w}_2$$

$$\dot{w}_4 = \cos w_1 \cdot \dot{w}_1$$

$$\dot{w}_5 = 1 \cdot \dot{w}_3 + 1 \cdot \dot{w}_4$$



$\bar{w}_1 = \bar{w}_4 \frac{dw_4}{dw_1}$

$\bar{w}_1 = \bar{w}_3 \frac{dw_3}{dw_1}$

$\bar{w}_4 = \bar{w}_5 \frac{dw_5}{dw_4}$

$\bar{w}_2 = \bar{w}_3 \frac{dw_3}{dw_2}$

$\bar{w}_3 = \bar{w}_5 \frac{dw_5}{dw_3}$

$\bar{w}_5 = \frac{\partial y}{\partial w_5} = 1$

$= \frac{\partial w_5}{\partial w_5}$ adjoint

$\bar{w}_5 = 1$

$\bar{w}_i = \sum_{j,\text{ after}} \bar{w}_j \frac{\partial w_j}{\partial w_i}$

$w_1 = x_1$

$w_2 = x_2$

$w_3 = w_1 * w_2$

$w_4 = \sin(w_1)$

$w_5 = w_3 + w_4$

fwd prop

reverse accum

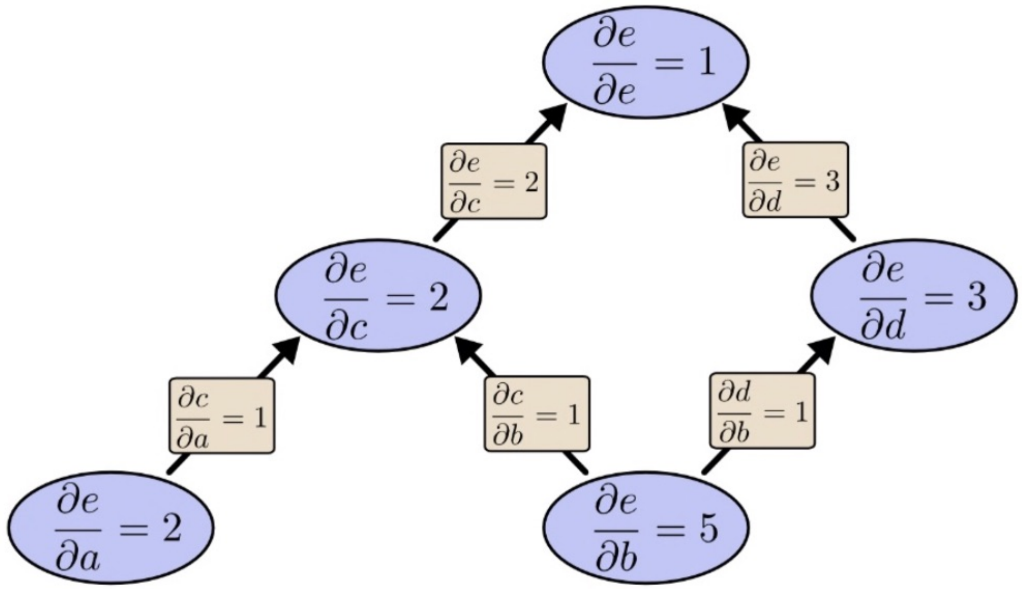
$\bar{w}_1 = \bar{w}_4 \cos(w_1) + \bar{w}_3 w_2$
$= \cos(w_1) + w_2$

$\bar{w}_2 = \bar{w}_3 \cdot w_1 = w_1$

$\bar{w}_3 = \bar{w}_5 \cdot 1 = 1 \cdot 1 = 1$

$\bar{w}_4 = \bar{w}_5 \cdot 1 = 1 \cdot 1 = 1$

$\bar{w}_5 = 1$



When I say that reverse-mode differentiation gives us the derivative of e with respect to every node, I really do mean *every node*. We get both $\frac{\partial e}{\partial a}$ and $\frac{\partial e}{\partial b}$, the derivatives of e with respect to both inputs. Forward-mode differentiation gave us the derivative of our output with respect to a single input, but reverse-mode differentiation gives us all of them.

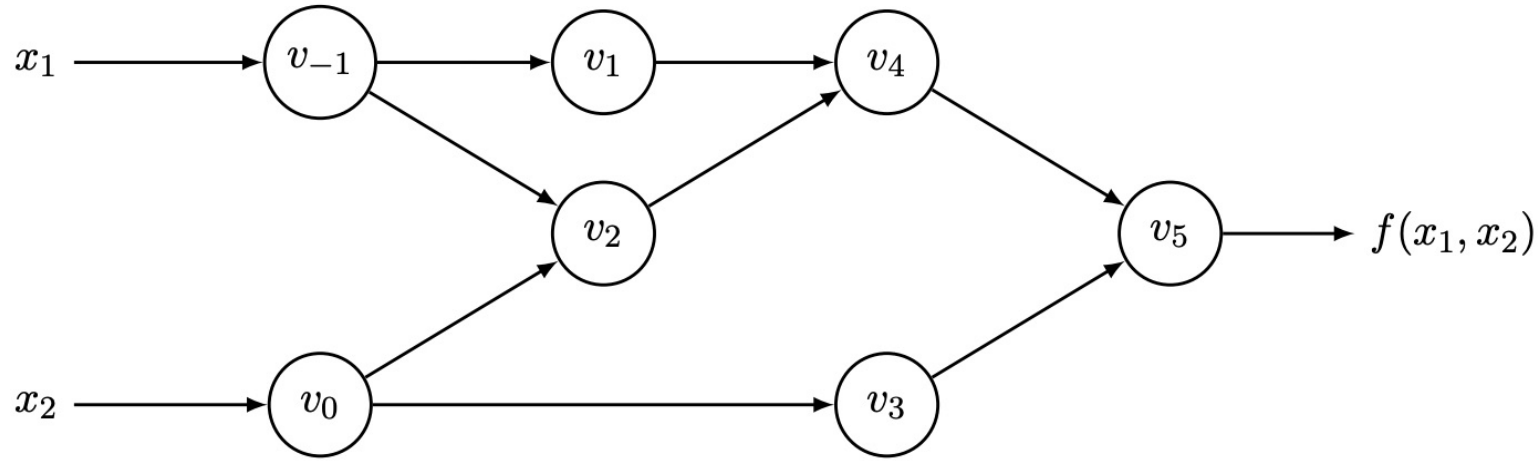
For this graph, that's only a factor of two speed up, but imagine a function with a million inputs and one output. Forward-mode differentiation would require us to go through the graph a million times to get the derivatives. Reverse-mode differentiation can get them all in one fell swoop! A speed up of a factor of a million is pretty nice!

Try it

$$f(x_1, x_2) = \ln(x_1) + x_1 x_2 - \sin(x_2)$$

1. Draw the compute graph
2. Write forward accum.
3. " reverse "

$$\begin{matrix} & 2 & \\ 1 & + & = 6 \\ & 3 & \end{matrix}$$



$v_{-1} = 2$ for example

$v_0 = 5$

$v_1 = \ln v_{-1}$

$v_2 = v_{-1} * v_0$

$v_3 = \sin v_0$

$v_4 = v_1 + v_2$

$v_5 = v_4 - v_3$

$\parallel$

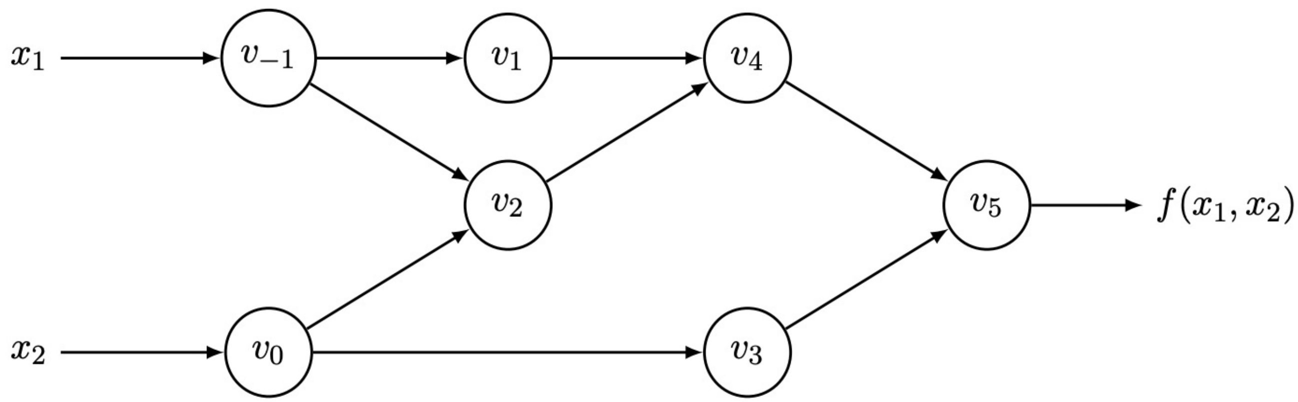
y

Forward Primal Trace

v_{-1}	$= x_1$	$= 2$
v_0	$= x_2$	$= 5$
v_1	$= \ln v_{-1}$	$= \ln 2$
v_2	$= v_{-1} \times v_0$	$= 2 \times 5$
v_3	$= \sin v_0$	$= \sin 5$
v_4	$= v_1 + v_2$	$= 0.693 + 10$
v_5	$= v_4 - v_3$	$= 10.693 + 0.959$
y	$= v_5$	$= 11.652$

Forward Tangent (Derivative) Trace

$\dot{v}_{-1}$	$= \dot{x}_1$	$= 1$
$\dot{v}_0$	$= \dot{x}_2$	$= 0$
$\dot{v}_1$	$= \dot{v}_{-1}/v_{-1}$	$= 1/2$
$\dot{v}_2$	$= \dot{v}_{-1} \times v_0 + \dot{v}_0 \times v_{-1}$	$= 1 \times 5 + 0 \times 2$
$\dot{v}_3$	$= \dot{v}_0 \times \cos v_0$	$= 0 \times \cos 5$
$\dot{v}_4$	$= \dot{v}_1 + \dot{v}_2$	$= 0.5 + 5$
$\dot{v}_5$	$= \dot{v}_4 - \dot{v}_3$	$= 5.5 - 0$
$\boldsymbol{\dot{y}}$	$\boldsymbol{= \dot{v}_5}$	$\boldsymbol{= 5.5}$



Forward Primal Trace

$$
\begin{aligned}
v_{-1} &= x_1 && = 2 \\
v_0 &= x_2 && = 5 \\
\hline
v_1 &= \ln v_{-1} && = \ln 2 \\
v_2 &= v_{-1} \times v_0 && = 2 \times 5 \\
v_3 &= \sin v_0 && = \sin 5 \\
v_4 &= v_1 + v_2 && = 0.693 + 10 \\
v_5 &= v_4 - v_3 && = 10.693 + 0.959 \\
\hline
y &= v_5 && = 11.652
\end{aligned}
$$

Reverse Adjoint (Derivative) Trace

$$
\begin{aligned}
&\boldsymbol{\bar{x}_1 = \bar{v}_{-1}} && && \boldsymbol{= 5.5} \\
&\boldsymbol{\bar{x}_2 = \bar{v}_0} && && \boldsymbol{= 1.716} \\
\hline
&\bar{v}_{-1} = \bar{v}_{-1} + \bar{v}_1 \frac{\partial v_1}{\partial v_{-1}} && = \bar{v}_{-1} + \bar{v}_1 / v_{-1} && = 5.5 \\
&\bar{v}_0 = \bar{v}_0 + \bar{v}_2 \frac{\partial v_2}{\partial v_0} && = \bar{v}_0 + \bar{v}_2 \times v_{-1} && = 1.716 \\
&\bar{v}_{-1} = \bar{v}_2 \frac{\partial v_2}{\partial v_{-1}} && = \bar{v}_2 \times v_0 && = 5 \\
&\bar{v}_0 = \bar{v}_3 \frac{\partial v_3}{\partial v_0} && = \bar{v}_3 \times \cos v_0 && = -0.284 \\
&\bar{v}_2 = \bar{v}_4 \frac{\partial v_4}{\partial v_2} && = \bar{v}_4 \times 1 && = 1 \\
&\bar{v}_1 = \bar{v}_4 \frac{\partial v_4}{\partial v_1} && = \bar{v}_4 \times 1 && = 1 \\
&\bar{v}_3 = \bar{v}_5 \frac{\partial v_5}{\partial v_3} && = \bar{v}_5 \times (-1) && = -1 \\
&\bar{v}_4 = \bar{v}_5 \frac{\partial v_5}{\partial v_4} && = \bar{v}_5 \times 1 && = 1 \\
\hline
&\bar{v}_5 = \bar{y} && = 1 &&
\end{aligned}
$$