31.05.22

Auto-diff

→ Types of diff

→ Hard-Coded — Complex / Closed form

→ Numeric — inaccurate / slow

→ Symbolic — expression swell / closed form

→ Automatic

→ Used as far as 2050, only ~2018 in ML.

→ Back-propagation

-> Confusion of AD

-> Provides numeric exp using symbolic rules

Numeric Differentiation (finite diff)

For a function $f: \mathbb{R}^N \to \mathbb{R}$

$$\frac{\partial f(\bar{x})}{\partial x_i} \approx \frac{f(\bar{x} + h\bar{e}_i) - f(\bar{x})}{h}$$

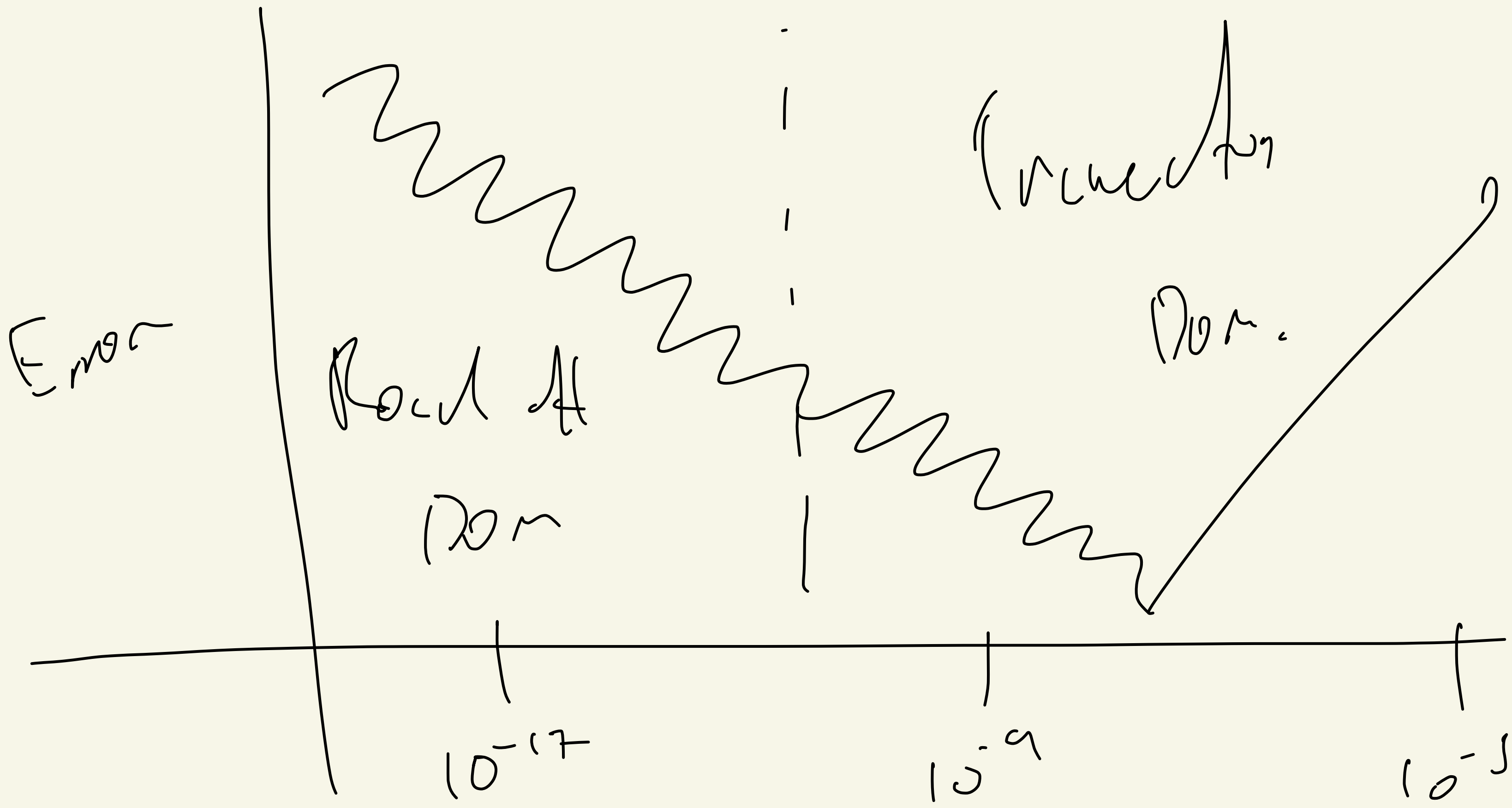
-> $O(n)$ for n x_i

-> Parameter h

-> Truncation error

-> h is not 0

-> Round-off error



Center differences

$$\frac{\partial f(\bar{x})}{\partial x_i} = \frac{f(\bar{x} + h\bar{e}_i) - f(\bar{x} - h\bar{e}_i)}{2h} + O(h^2)$$

B.f: $f: \mathbb{R}^n \to \mathbb{R}^m \Rightarrow 2mn$ computations $O(mn)$

↳ Does not solve either error fully.

→ Main problem

→ Truncation a bit

→ mostly $\mathcal{O}(mn)$ scaling

⇒ NN's are resilient to float point error.

↳ Gupta et. al. 2015

Symbolic Differentiation

$$\partial_x (f(x) + g(x)) = \partial_x f(x) + \partial_x g(x)$$

$$\partial_x (f(x) \cdot g(x)) = \partial_x f(x) g(x) + f(x) \partial_x g(x)$$

⇒ Good for understanding tangent space.

⇒ Solving for minima

→ expression swell

$$l_{n+1} = 4 l_n (1 - l_n)$$

n	l_n	$\partial_x l_n$
1	x	1
2	$4x(1-x)$	$4(1-x) - 4x$
3	$16x(1-x)(1-2x)^2$	$16(1-x)(1-2x)^2 - 16x(1-2x)^2 - 64x(1-x)(1-2x)$
4	$64x(1-x)(1-2x)^2(1-8x+8x^2)^2$	~4 lines

⇒ Not feasible if all we want is the value

Automatic Differentiation.

→ Non-standard interpretation of code

→ All numeric computations are a composition of elementary operations for which derivatives are known.

→ Typically

→ binary arithmetic

→ sign switch

→ transcendental functions

→ exp

→ log

→ Trig funcs

→ Wengert List (1964) or Trace

Notation

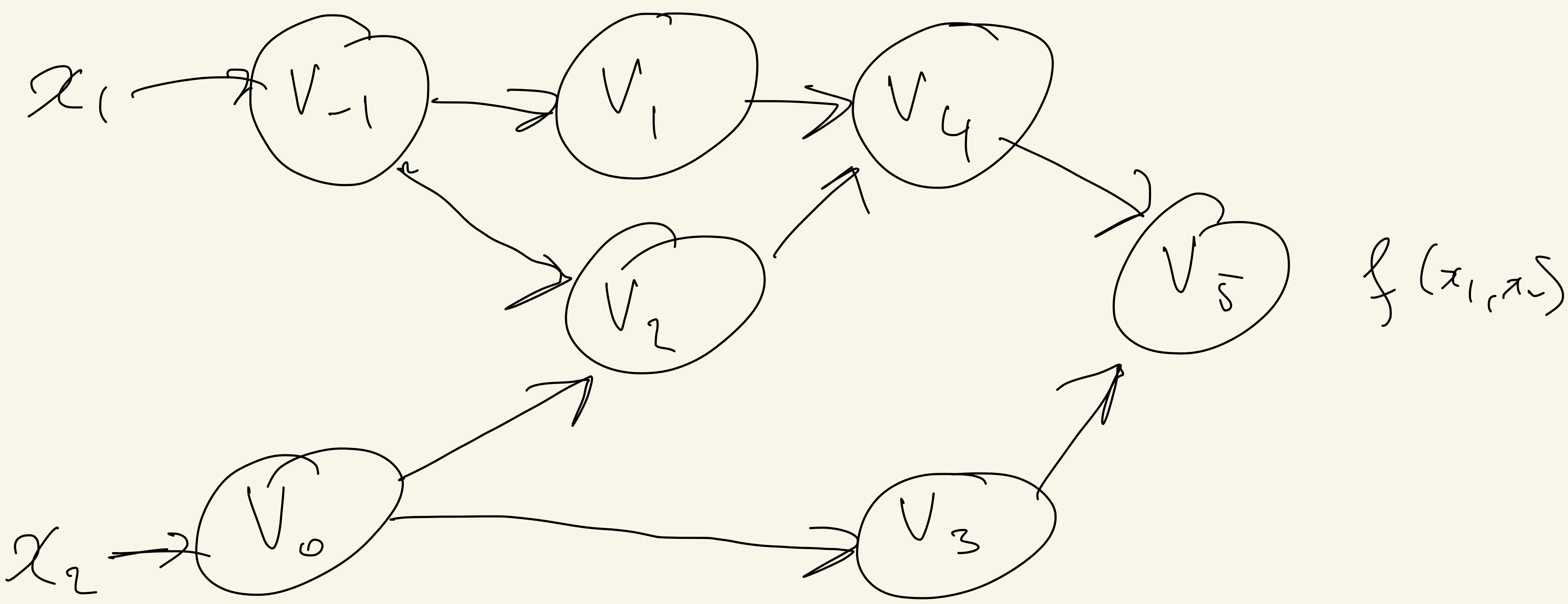
$f: \mathbb{R}^n \to \mathbb{R}^m$

input vars: v_{i-n}, $i = 1, \dots, n$

intermediate vars v_i, $i = 1, \dots, l$

output $y_{m-i} = v_{l-i}$, $i = m-1, \dots, 0$

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



Forward Mode (tangent linear)

→ Generate a tangent trace

$$\dot{v}_i = \frac{\partial v_i}{\partial x_1} \quad \forall\, v_i$$

Set $\dot{x}_i = 1$ and $\dot{x}_j = 0 \;\; \forall j \neq i$

Primal Trace	Tangent Trace
$v_{-1} = x_1 = 2$	$\dot{v}_{-1} = \dot{x}_1 = 1$
$v_0 = x_2 = 5$	$\dot{v}_0 = \dot{x}_2 = 0$
$v_1 = \ln v_{-1} = \ln 2$	$\dot{v}_1 = \dot{v}_{-1}/v_{-1} = 1/2$
$v_2 = v_{-1} \cdot v_0 = 2 \cdot 5$	$\dot{v}_2 = \dot{v}_{-1} \cdot v_0 + \dot{v}_0 \cdot v_{-1} = 1 \cdot 5 + 0 \cdot 2$
$v_3 = \sin v_0 = \sin 5$	$\dot{v}_3 = \dot{v}_0 \cdot \cos v_0 = 0 \cdot \cos 5$
$v_4 = v_1 + v_2 = 0.693 + 10$	$\dot{v}_4 = \dot{v}_1 + \dot{v}_2 = 0.5 + 5$
$v_5 = v_4 - v_3 = 10.693 + 0.959$	$\dot{v}_5 = \dot{v}_4 - \dot{v}_3 = 5.5 - 0$
$y = v_5 = 11.652$	$\dot{y} = \dot{v}_5 = 5.5$

-> Efficient for

$$\boxed{f : \mathbb{R}^n \to \mathbb{R}^m \ : \ n \ll m}$$

↳ Scals lik $\mathcal{O}(n)$

Dual Numbers

-> Truncated taylor series

$$v + \dot{v}\varepsilon$$

↳ nilpotent number : $\varepsilon^2 = 0$

! $\varepsilon \neq 0$

$$(v + \dot{v}\varepsilon)(u + \dot{u}\varepsilon) = \underbrace{uv + u\dot{v}\varepsilon + v\dot{u}\varepsilon}_{} + \cancel{\dot{u}\dot{v}\varepsilon^2}^{=0}$$

↳ Product Rule

→ Set up a regime s.t

$$f(v + \dot{v}\varepsilon) = f(v) + f'(v)\dot{v}\varepsilon$$

→ Gives chain rule

$$f(g(v + \dot{v}\varepsilon)) = f(g(v) + g'(v)\dot{v}\varepsilon)$$

$$= f(g(v)) + f'(g(v))g'(v)\dot{v}\varepsilon$$

⇒ Dual numbers carry tangent values with the primal.

Reverse Mode (Cotangent linear)

→ Complement each intermediate with adjoint

$$\bar{v}_i = \frac{\partial y_j}{\partial v_i}$$

→ We are interested in:

$$\frac{\partial y}{\partial v_0} = \frac{\partial y}{\partial v_2}\frac{\partial v_2}{\partial v_0} + \frac{\partial y}{\partial v_3}\frac{\partial v_3}{\partial v_0}$$

$$\bar{v}_0 = \bar{v}_2 \frac{\partial v_2}{\partial v_0} + \bar{v}_3 \frac{\partial v_3}{\partial v_0}$$

Two step

$$\bar{v}_0 = \bar{v}_3 \frac{\partial v_3}{v_0} \quad ; \quad \bar{v}_0 = \bar{v}_0 + \bar{v}_2 \frac{\partial v_2}{\partial v_0}$$

Slowly

$$u_1 = f(v)$$

$$u_2 = g(v)$$

$$t = h(u_1, u_2)$$

it can be shown

$$\partial_v t = \sum_i \frac{dt}{du_i} \frac{du_i}{dv}$$

example

$$z = x_1 x_2 + \sin(x_1)$$

Forward

$$w_1 = x_1$$

$$w_2 = x_2$$

$$w_3 = w_1 \cdot w_2$$

$$w_4 = \sin(w_1)$$

$$w_5 = w_3 + w_4$$

$$z = w_5$$

Reverse

$$\frac{dz}{dz} = 1$$

$$\Rightarrow w_5 = z$$

$$\frac{dz}{dw_5} = 1$$

and $\frac{\partial w_5}{\partial w_3} = 1 = \frac{\partial w_5}{\partial w_4} = 1$

$$\frac{dz}{dw_3} = \overbrace{\frac{dz}{dw_5}}^{\text{one part}} \cdot \frac{dw_5}{dw_3} = 1 \times 1 = 1$$

$$\frac{dz}{dw_4} = \frac{dz}{dw_5} \cdot \frac{dw_5}{dw_4} = 1 \times 1 = 1$$

$w_3 = w_2 \cdot w_1 \quad \therefore \quad \frac{dw_3}{dw_2} = w_1$

$$\frac{dz}{dw_2} = \frac{dz}{dw_3} \frac{dw_3}{dw_2} = 1 \times w_1 = w_1$$

$$\frac{dz}{dw_2} = w_1 = 2$$

$$w_2 \propto w_3, w_4$$

$$\frac{dw_3}{dw_1} = w_2 \qquad \frac{dw_4}{dw_1} = \cos(w_1)$$

$$\frac{dz}{dw_1} = \frac{dz}{dw_3}\frac{dw_3}{dw_1} + \frac{dz}{dw_4}\frac{dw_4}{dw_1}$$

$$= w_2 + \cos(w_1)$$

$$= 2.58$$

$$\therefore \boxed{\frac{dz}{dx_1} = 2.58, \quad \frac{dz}{dx_2} = 2}$$

Scaling

$$f : \mathbb{R}^n \to \mathbb{R}^m$$

$$\left.\begin{array}{ll} \text{forward} & t \propto n \cdot c \cdot \text{ops}(f) \\ \\ \text{rev} & t \propto m \cdot c \cdot \text{ops}(f) \end{array}\right\} \begin{array}{l} c \overset{!}{<} 6 \\ \\ c \sim [2,3] \end{array}$$

$$\left.\begin{array}{ll} \text{for} & n > m \\ & \hookrightarrow \text{rev} \\ \\ \text{for} & n \ll m \\ & \hookrightarrow \text{form.} \end{array}\right\} \text{also depends on what you need.}$$