- Gradient : $\nabla f(x) \in \mathbb{R}^d$
- Jacobienne $\partial F(x) \in \mathbb{R}^{d \times p}$
- Transposée $A \longrightarrow A^T$
  Adjoint

$f : \mathbb{R}^d \longrightarrow \mathbb{R}$

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ | \\ \frac{\partial f}{\partial x_d}(x) \end{pmatrix} \in \mathbb{R}^d$$

$$f(x + \varepsilon\delta) = f(x) + \varepsilon \langle \nabla f(x), \delta \rangle + o(\varepsilon)$$

$F : \mathbb{R}^d \longrightarrow \mathbb{R}^p$

$$\partial F(x) : \mathbb{R}^d \underset{\text{Linéaire}}{\longrightarrow} \mathbb{R}^p \quad \equiv \quad \partial F(x) \in \mathbb{R}^{p \times d}$$

$$F(x) = \begin{bmatrix} F_1(x) \\ \vdots \\ F_p(x) \end{bmatrix} \qquad \partial F(x) \overset{\Delta}{=} \left[\frac{\partial F_i}{\partial x_j}(x) \right]_{\substack{i=1\dots p \\ j=1\dots d}}$$

$$F(x+\varepsilon\delta) = F(x) + \varepsilon \underbrace{\partial F(x)[\delta]}_{\partial F(x) \times \delta} + o(\varepsilon)$$

$$F : \mathbb{R}^d \to \mathbb{R} \qquad p = 1$$

$$\partial F(x) =$$ 

$$\nabla F(x) =$$  $$= \partial F(x)^T$$

$$A = (A_{ij})_{ij} \qquad A^T = (A_{ji})_{ij}$$

$$\langle A[x], y \rangle \stackrel{?}{=} \langle x, A^T[y] \rangle$$

$$" \partial(F \circ G) = \partial F \times \partial G "$$

$$\left[\partial(F \circ G)(x) = \partial F(G(x)) \times \partial G(x) \right]$$

Q: $\nabla f(x)$ calcul

code $f(x)$ $\xrightarrow[\text{"BackProp"}]{\text{Meta}}$ code $\nabla f(x)$

code calcul $f(x)$ K opér°

↳ coût calculer $\nabla f(x)$ en ?? opér°

$$\begin{cases} f : \mathbb{R}^d \to \mathbb{R} \\ \nabla f : \mathbb{R}^d \to \mathbb{R}^d \end{cases}$$

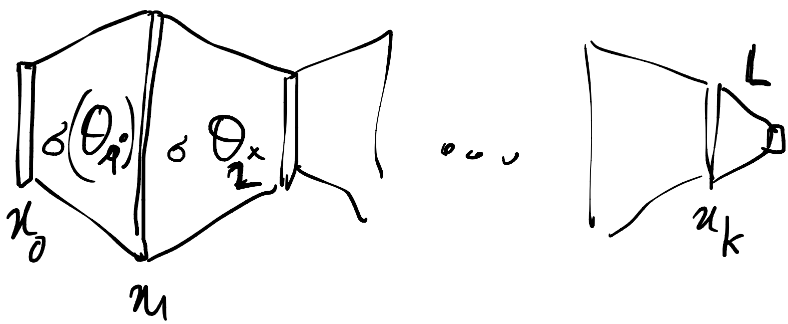
$$Df(x) \approx \left(\frac{-f(x) + f(x+\varepsilon\delta_1)}{\varepsilon}, \dots, \frac{f(x+\varepsilon\delta_d)-f(x)}{\varepsilon} \right)$$

$$(d+1)K$$

Thm: Seppo Linnainmaa 1970
Baur-Strassen 1983
$O(n^{2,7}) \to \omega$

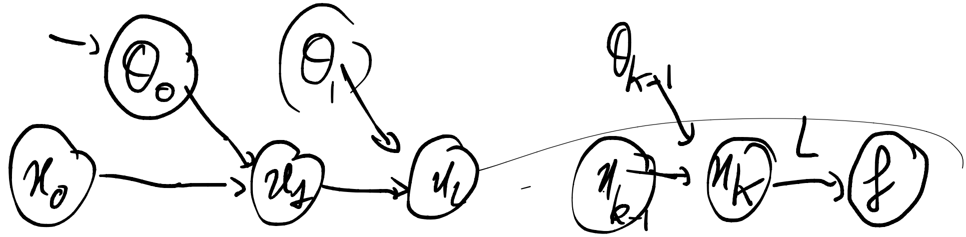
calcula $Df(x)$ 3K optimal

MLP:



$$f(\theta) = L(x_k).$$

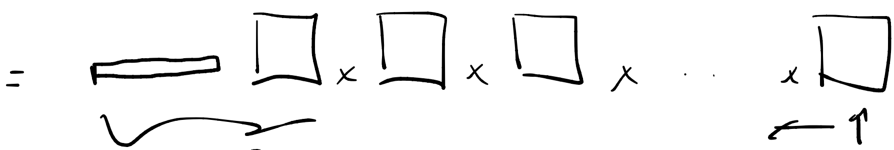
$$x_{k+1} = \sigma(\theta_k @ x_k)$$



<u>Composition</u>: $f(x) = f_k \circ f_{k-1} \cdots \circ f_0(x)$

$$\partial f(x) = \underbrace{\partial f_k(x_k)} \times \partial f_{k-1}(x_{k-1}) \times \cdots \times \partial f_0(x_0)$$

$f(x) \in \mathbb{R}$ $\quad x_k \in \mathbb{R}^d$



d^2 $\quad d^3$

FWD $\Leftarrow$ Kd^3

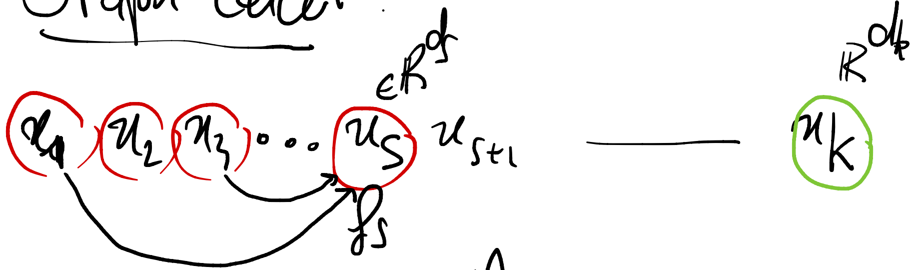
d^2

BWD $\Rightarrow$ Kd^2 /d

[Check Pointing] $\quad$ "RevNet"

[Invertible Neural Net]

Graph calcul :



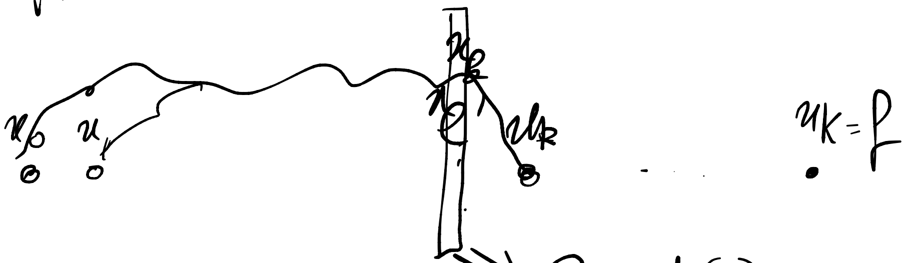
gche $\rightarrow$ dte

$$x_k = f_k(x_{k-1}, x_{k-2} \text{ — } x_1)$$

DAG directed acyclic graph

Tri topo

Diff. auto. FWD :



Parent(k)

$$\frac{\partial x_k}{\partial x_0} \in \mathbb{R}^{d_k \times d_0}$$

$$x_k = f_k\Big((x_l)_{l \in \text{Parent}} \Big)$$

Prop :

$$\frac{\partial x_k}{\partial x_0} \overset{(*)}{=} \sum_{l \in \text{Parent}(k)} \underbrace{\left[\frac{\partial x_k}{\partial x_l}\right]}_{\equiv \frac{\partial f_k(1 \dots)}{\partial x_l}} \times \frac{\partial x_l}{\partial x_0}$$

$x_0 \leftarrow$ init

$x_1 \leftarrow$ init

$\vdots$

$x_s \leftarrow$

$\Longrightarrow$

$$\frac{\partial x_0}{\partial x_0} = Id$$

$$\frac{\partial x_1}{\partial x_0} = 0 \quad \dots \quad = 0$$

for $k = s+1 - K$

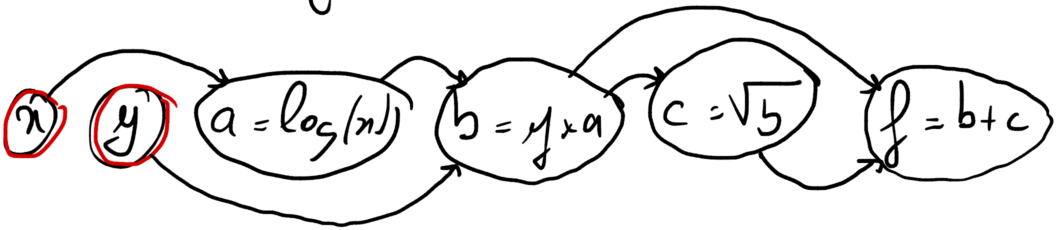
$x_k = f_k(\text{---})$

$\equiv$

for $k = s+1 \dots K$

$$\frac{\partial x_k}{\partial x_0} = \text{chain } (*)$$

$$f(x,y) = y\log(x) + \sqrt{y\log(x)}$$



$$\frac{\partial x}{\partial \cancel{x}\, y} = 1 \longrightarrow 0$$

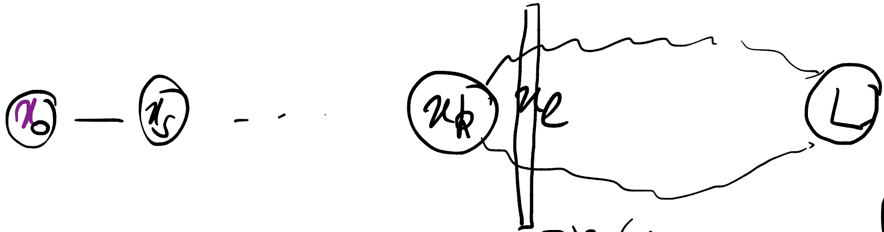
$$\frac{\partial y}{\partial \cancel{x}\, y} = 0 \longrightarrow 1$$

$$\frac{\partial a}{\partial \cancel{x}\, y} = \boxed{\frac{\partial a}{\partial x}} \times \frac{\partial x}{\partial \cancel{x}\, y} = \frac{1}{x} \times \frac{\partial x}{\partial \cancel{x}\, y}$$

$$\frac{\partial b}{\partial \cancel{x}\, y} = \boxed{\frac{\partial b}{\partial a}} \times \frac{\partial a}{\partial \cancel{x}\, y} + \boxed{\frac{\partial b}{\partial y}} \times \frac{\partial y}{\partial \cancel{x}\, y} = y\frac{\partial a}{\partial x} + a\frac{\partial y}{\partial x}$$

$$\frac{\partial c}{\partial \cancel{x}\, y} = \boxed{\frac{\partial c}{\partial b}} \times \frac{\partial b}{\partial \cancel{x}\, y} = \frac{1}{2\sqrt{b}} \times \frac{\partial b}{\partial \cancel{x}\, y}$$

$$\boxed{\frac{\partial f}{\partial x}} = \boxed{\frac{\partial f}{\partial b}} \times \frac{\partial b}{\partial \cancel{x}\, y} + \boxed{\frac{\partial f}{\partial c}} \times \frac{\partial c}{\partial \cancel{x}\, y} = \frac{\partial b}{\partial x} + \frac{\partial c}{\partial x}$$



Fils(k)

$$\frac{\partial x_\ell}{\partial x_k}$$

$$x_\ell = f_\ell(\dots x_k \dots)$$

Prop :

$$\frac{\partial L}{\partial x_k} = \sum_{\ell \in \text{Fils}(k)} \frac{\partial L}{\partial x_\ell} \times \underbrace{\left[\frac{\partial x_\ell}{\partial x_k}\right]}_{\frac{\partial f_\ell}{\partial x_k}}$$

Algo calcul f

x_1
$\leftarrow$ input
x_5

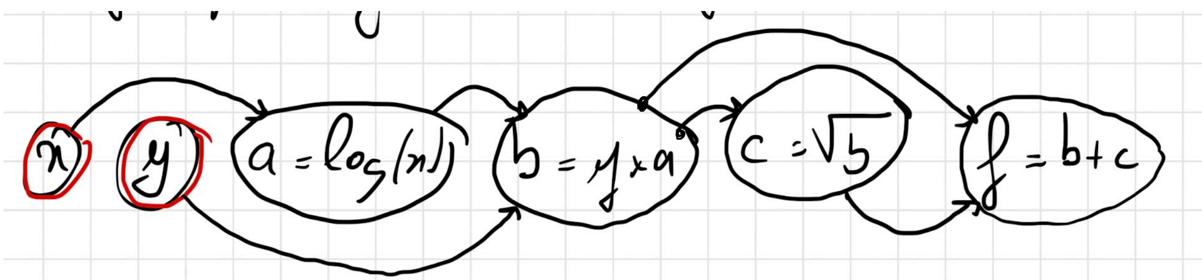
For $k = 5+1 \text{ — } K$

$x_k = f_k(\quad)$

$$\frac{\partial L}{\partial L} = \text{Id}$$

For $k = K \dots 1$

$$\frac{\partial L}{\partial x_k} = \sum \dots$$



$$\frac{\partial f}{\partial f} = 1$$

$$\frac{\partial f}{\partial c} = \frac{\partial f}{\partial f} \times \frac{\partial f}{\partial c} = \frac{\partial f}{\partial f} \times 1$$

$$\frac{\partial f}{\partial b} = \frac{\partial f}{\partial c} \times \frac{\partial c}{\partial b} + \frac{\partial f}{\partial f} \times \frac{\partial f}{\partial b} = \frac{\partial f}{\partial c} \frac{1}{2\sqrt{b}} + \frac{\partial f}{\partial f} \times 1$$

$$\frac{\partial f}{\partial a} = \frac{\partial f}{\partial b} \frac{\partial b}{\partial a} = \frac{\partial f}{\partial b} \times y$$

$$\boxed{\frac{\partial f}{\partial y}} = \frac{\partial f}{\partial b} \times \frac{\partial b}{\partial y} = \frac{\partial f}{\partial b} \times a$$

$$\boxed{\frac{\partial f}{\partial x}} = \frac{\partial f}{\partial a} \times \frac{\partial a}{\partial x} = \frac{\partial f}{\partial a} \times \frac{1}{x}$$

Back Prop en mode grad :

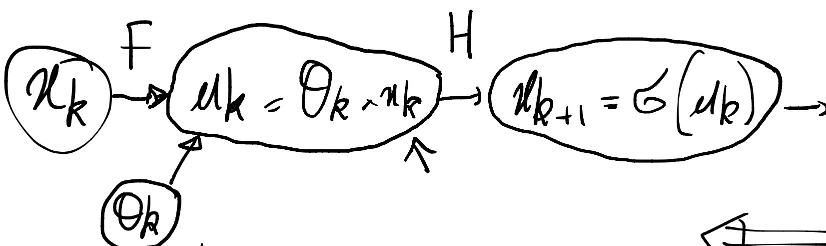
$$\frac{\partial L}{\partial x_k} = \sum_{l \in Fils(k)} \frac{\partial L}{\partial x_l} \times \frac{\partial f_l}{\partial x_k}$$

T

$$(AB)^T = B^T A^T$$

$$\nabla_{x_k} L = \sum_{l \in Fils(k)} \underbrace{\left[\frac{\partial f_l}{\partial x_k}\right]^T (\nabla_{x_l} L)}_{JVP}$$

"Jacobian Vector Product"



$\nabla_{x_{k+1}} L$ = calculé

$$\nabla_{u_k} L = \left[\frac{\partial H(u_k)}{\partial u_k}\right]^T (\nabla_{x_k} L) = \boxed{\sigma'(u) * \nabla_{x_k} L}$$

$$\nabla_{x_k} L = \left[\frac{\partial F}{\partial x_k}(x_k, \theta_k)\right]^T \left(\nabla_{x_k} L\right) = \boxed{\theta^T @ \nabla_{x_k} L}$$

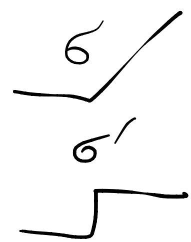
$$+$$

$$\nabla_{\theta_k} L = \left[\frac{\partial F}{\partial \theta_k}(x_k, \theta_k)\right]^T \left(\nabla_{x_k} L\right) = \boxed{\nabla_{x_k} L @ x^T}$$

$$H(u) = \sigma(u) = \left(\sigma(u[i])\right)_i$$

$$\partial H(u) = \mathrm{diag}\left(\sigma'(u[i])\right)_i$$

$$\sigma = \mathrm{ReLU} \qquad \sigma' = \mathbb{1}_{\mathbb{R}^+}$$



$$\partial H(u)^T = \mathrm{diag}\left(\sigma'(u)\right)$$

$$F(x, \theta) = \theta @ x$$

$$\partial_x F(x, \theta)[z] = \theta @ z$$

$$\partial_\theta F(x, \theta)[w] = w @ x$$

$\partial_x F^T$

$$\langle \theta @ z, w \rangle = \langle z, \theta^T @ w \rangle$$

$$\langle w @ x, z \rangle = \langle w, z @ x^T \rangle$$