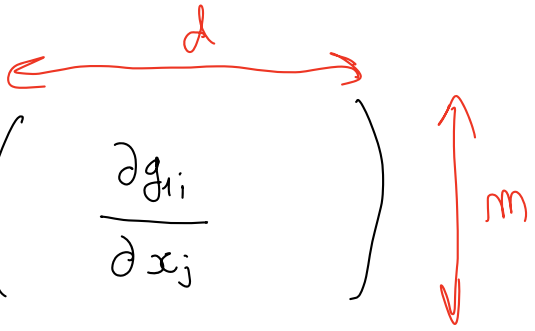
$$f(x) = g_3(g_2(g_1(x)))$$

$x \in \mathbb{R}^d$ $\quad g_1 : \mathbb{R}^d \to \mathbb{R}^m$ $\quad g_2 : \mathbb{R}^m \to \mathbb{R}^n$ $\quad g_3 : \mathbb{R}^n \to \mathbb{R}$

$$\nabla f = \begin{pmatrix} \partial f/\partial x_1 \\ \vdots \\ \partial f/\partial x_d \end{pmatrix}$$

$$J_1 = Jac(g_1) = \begin{pmatrix} \dfrac{\partial g_{1i}}{\partial x_j} \end{pmatrix}$$



$$\nabla f = (J_3 \; J_2 \; J_1)^T$$

$J_3 (J_2 J_1)$

$(J_3 J_2) J_1$

$\underbrace{J_3}_{\in \mathbb{R}^{1\times n}} \quad \underbrace{J_2 \; J_1}_{\in \mathbb{R}^{n\times m} \quad \in \mathbb{R}^{m\times d}}$

1) $(J_2 J_1 \in \mathbb{R}^{n\times d}$
$\Rightarrow$ ndm multiplications to compute it

2) $\underbrace{J_3}_{\in \mathbb{R}^{1\times n}} \underbrace{(J_2 J_1)}_{\in \mathbb{R}^{n\times d}} \Rightarrow nd$ multiplications

Cost is $ndm + nd$

1) $J_3 J_2 \Rightarrow nm$ multiplications

2) $(J_3 J_2) J_1 \Rightarrow md$ multip. cost is $nm + md$

Assumptions: $d \geq m \geq n \geq 1$

Ccl: $ndm + nd \geq nm + md$

So $J_3 (J_2 J_1)$ is more costly

and we prefer $(J_3 J_2) J_1$

$x \rightarrow g_1 \rightarrow g_2 \rightarrow g_3$