exercice 1 Soient $g : \mathbb{R}^3 \to \mathbb{R}^3$ et $f : \mathbb{R}^3 \to \mathbb{R}^3$ deux fonctions. Supposons que

$$g : \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \to \begin{pmatrix} 3x_1 + -2x_2 + x_3 \\ -x_1 + 4x_3 \\ -3x_1 + x_2 - x_3 \end{pmatrix} \qquad f : \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \to \begin{pmatrix} e^{x_1+x_2} \\ e^{x_2+x_3} \\ e^{x_3+x_1} \end{pmatrix}$$

Calculer les gradients de $g \circ f$ et $f \circ g$.

Démonstration. Commençons par calculer les gradients de f et g qui sont toutes deux des fonctions de $\mathbb{R}^3$ dans $\mathbb{R}^3$.

$$\nabla g : \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \to \begin{pmatrix} 3 & -2 & 1 \\ -1 & 0 & 4 \\ -3 & 1 & -1 \end{pmatrix}$$

$$\nabla f : \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \to \begin{pmatrix} e^{x_1+x_2} & e^{x_1+x_2} & 0 \\ 0 & e^{x_2+x_3} & e^{x_2+x_3} \\ e^{x_3+x_1} & 0 & e^{x_3+x_1} \end{pmatrix}$$

On applique alors la formule de la chain rule : $\nabla(g \circ f) = (\nabla(g)(f) \times \nabla f$.

$$\nabla(g \circ f) : \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \to \begin{pmatrix} 3e^{x_1+x_2} + e^{x_3+x_1} & 3e^{x_1+x_2} - 2e^{x_2+x_3} & e^{x_3+x_1} - 2e^{x_2+x_3} \\ 4e^{x_3+x_1} - e^{x_1+x_2} & -e^{x_1+x_2} & 4e^{x_3+x_1} \\ -3e^{x_1+x_2} - e^{x_3+x_1} & -3e^{x_1+x_2} + e^{x_2+x_3} & e^{x_2+x_3} - e^{x_3+x_1} \end{pmatrix}$$

Le deuxième gradient (ou jacobienne) ne tient pas sur une ligne... donc on va l'écrire par colonne

$$\text{colonne}_1 \nabla(f \circ g) : \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \to \begin{pmatrix} 3e^{3x_1+-2x_2+x_3+-x_1+4x_3} - e^{3x_1+-2x_2+x_3+-x_1+4x_3} \\ -e^{-x_1+4x_3+-3x_1+x_2-x_3} - 3e^{-x_1+4x_3+-3x_1+x_2-x_3} \\ 3e^{-3x_1+x_2-x_3+3x_1+-2x_2+x_3} - 3e^{-3x_1+x_2-x_3+3x_1+-2x_2+x_3} \end{pmatrix}$$

$$\text{colonne}_2 \nabla(f \circ g) : \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \to \begin{pmatrix} -2e^{3x_1+-2x_2+x_3+-x_1+4x_3} \\ e^{-x_1+4x_3+-3x_1+x_2-x_3} \\ -2e^{-3x_1+x_2-x_3x_3+3x_1+-2x_2+x_3} + e^{-3x_1+x_2-x_3+3x_1+-2x_2+x_3} \end{pmatrix}$$

$$\text{colonne}_3 \nabla(f \circ g) : \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \to \begin{pmatrix} e^{3x_1+-2x_2+x_3+-x_1+4x_3} + 4e^{3x_1+-2x_2+x_3+-x_1+4x_3} \\ 4e^{-x_1+4x_3+-3x_1+x_2-x_3} - e^{-x_1+4x_3+-3x_1+x_2-x_3} \\ e^{-3x_1+x_2-x_3+3x_1+-2x_2+x_3} - e^{-3x_1+x_2-x_3+3x_1+-2x_2+x_3} \end{pmatrix}$$

□

exercice 2 (**regression logistique**) Soit F un perceptron appris par cross-entropy (on verra ça en cours d'IA ou durant les PST).

$$f : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to \begin{pmatrix} w_{1,1}x_1 + w_{1,2}x_2 \\ w_{2,1}x_1 + w_{2,2}x_2 \end{pmatrix} \qquad \text{softmax} : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to \begin{pmatrix} \frac{e^{x_1}}{e^{x_1}+e^{x_2}} \\ \frac{e^{x_1}}{e^{x_1}+e^{x_2}} \end{pmatrix} \qquad \text{loss} : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to -y_1 \log(x_1) - y_2 \log(x_2)$$

Calculer le gradient de $F = \text{loss} \circ \text{softmax} \circ f$ dans le cas $y_1 = 1$ et $y_2 = 0$.

Démonstration. Calculons les gradients de chaque étape de F.

$$\nabla f : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to \begin{pmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \end{pmatrix}$$

$$\nabla \text{softmax} : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to \begin{pmatrix} \frac{e^{x_1}(e^{x_1}+e^{x_2})-e^{x_1 2}}{(e^{x_1}+e^{x_2})^2} & \frac{-e^{x_1}e^{x_2}}{(e^{x_1}+e^{x_2})^2} \\ \frac{-e^{x_1}e^{x_2}}{(e^{x_1}+e^{x_2})^2} & \frac{e^{x_2}(e^{x_1}+e^{x_2})-e^{x_2 2}}{(e^{x_1}+e^{x_2})^2} \end{pmatrix} = \begin{pmatrix} \frac{e^{x_1}e^{x_2}}{(e^{x_1}+e^{x_2})^2} & \frac{-e^{x_1}e^{x_2}}{(e^{x_1}+e^{x_2})^2} \\ \frac{-e^{x_1}e^{x_2}}{(e^{x_1}+e^{x_2})^2} & \frac{e^{x_1}e^{x_2}}{(e^{x_1}+e^{x_2})^2} \end{pmatrix}$$

$$\nabla \text{loss} : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to \begin{pmatrix} \frac{1}{x_1} & 0 \end{pmatrix}$$

Nous allons utiliser la chain rule deux fois : d'avord pour calculer $\nabla(\text{loss} \circ \text{softmax})$

$$\nabla(\text{loss} \circ \text{softmax}) : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to \begin{pmatrix} \frac{e^{x_1}+e^{x_2}}{e^{x_1}} & 0 \end{pmatrix} \begin{pmatrix} \frac{e^{x_1}e^{x_2}}{(e^{x_1}+e^{x_2})^2} & \frac{-e^{x_1}e^{x_2}}{(e^{x_1}+e^{x_2})^2} \\ \frac{-e^{x_1}e^{x_2}}{(e^{x_1}+e^{x_2})^2} & \frac{e^{x_1}e^{x_2}}{(e^{x_1}+e^{x_2})^2} \end{pmatrix} = \begin{pmatrix} \frac{e^{x_2}}{e^{x_1}+e^{x_2}} & \frac{-e^{x_2}}{e^{x_1}+e^{x_2}} \end{pmatrix}$$

La loss (corss-entropy) simplifie joliment les choses... Calculons maintenant le gradient de F

$$\nabla F = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to \nabla(\text{loss} \circ \text{softmax})(f(x)) \times \nabla f(x) = \begin{pmatrix} \frac{e^{w_{2,1}x_1+w_{2,2}x_2}}{e^{w_{1,1}x_1+w_{1,2}x_2}+e^{w_{2,1}x_1+w_{2,2}x_2}} & \frac{-e^{w_{2,1}x_1+w_{2,2}x_2}}{e^{w_{1,1}x_1+w_{1,2}x_2}+e^{w_{2,1}x_1+w_{2,2}x_2}} \end{pmatrix} W$$

où W est la matrice des poids W égale au gradient de f calculé précédemment. □