Automatic Differentiation

Question 2: Gradient Through a Matrix Product

Consider $L(Y) \in \mathbb{R}$, for $Y \in \mathbb{R}^{D \times N}$. That is, $L : \mathbb{R}^{D \times N} \to \mathbb{R}$.

Let $Y = H \cdot W$, where $H \in \mathbb{R}^{D \times M}$, $W \in \mathbb{R}^{M \times N}$, and $\cdot$ refers to the matrix product.

Suppose you have already computed $\nabla_Y L$, where $\nabla_Y L \in \mathbb{R}^{D \times N}$.

Prove that,

$$\nabla_H L = \nabla_Y L \cdot W^{\mathrm{T}} \,, \tag{1}$$

and

$$\nabla_W L = H^{\mathrm{T}} \cdot \nabla_Y L \,. \tag{2}$$

Of course, I know you will include statements in your proof that help the reader understand it.

Submit your answers as a PDF file in Kritik. Make sure that nothing in the file can be used to identify you. Remember, the peer-assessment process is anonymous.