



陈狄

基本无害

429 人赞同了该回答

我天，现有的答案复杂到了让我怀疑自己到底懂不懂的程度。。。

来举个栗子吧： $AB = C \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix} = \begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix}$

其中

$$\begin{cases} c_1 = a_1 b_1 + a_2 b_3 \\ c_2 = a_1 b_2 + a_2 b_4 \\ c_3 = a_3 b_1 + a_4 b_3 \\ c_4 = a_3 b_2 + a_4 b_4 \end{cases}$$

那么

$$\frac{\partial C}{\partial A} \Rightarrow \begin{cases} \partial c_1 / \partial a_1 = b_1 \\ \partial c_1 / \partial a_2 = b_3 \\ \partial c_1 / \partial a_3 = 0 \\ \partial c_1 / \partial a_4 = 0 \\ \partial c_2 / \partial a_1 = b_2 \\ \vdots \\ \partial c_4 / \partial a_4 = b_4 \end{cases}$$

即 C 中每一个元素，对于 A 中每一个元素进行求导。转化成标量的形式就好理解了吧~

至于把以上16个标量求导写成 4×4 的矩阵也好还是16维的向量也好，大多是为了形式（理论）上的美观，或是方便对求导结果的后续使用，亦或是方便编程实现，按需自取，其本质不变。

