First we show that, $\nabla_A tr(AB) = B^\top$ for two matrices A and B. By definition,

$$tr(AB) = \sum_{j=1}^{n}\sum_{i=1}^{m} a_{ji}b_{ij}$$
$$\implies \frac{\partial}{\partial a_{ij}} tr(AB) = b_{ji}$$
$$\implies \nabla_A tr(AB) = B^\top$$

Also, $tr(AB) = tr(BA)$

Now, lets assume $tr(ABA^\top C) = tr(XY^\top C) = \mathbf{H}(X,Y)$ where $X = f(A) = AB$ and $Y = A$.

Using chain rule:

$$\begin{aligned}\nabla_A \mathbf{H}(X,Y) &= \nabla_X \mathbf{H}(X,Y)\nabla_A X + \nabla_Y \mathbf{H}(X,Y)\nabla_A Y\\ &= \nabla_X \mathbf{H}(X,Y)\nabla_A AB + \nabla_Y \mathbf{H}(X,Y)\nabla_A A\\ &= \nabla_X \mathbf{H}(X,Y)B^\top + \nabla_Y \mathbf{H}(X,Y)\end{aligned}$$

The proof for the highlighted text is mentioned below

Now,

$$\begin{aligned}\nabla_X \mathbf{H}(X,Y) &= \nabla_X tr(XY^\top C)\\ &= (Y^\top C)^\top\\ &= C^\top Y\\ &= C^\top A\end{aligned}$$

And,

$$\begin{aligned}\nabla_Y \mathbf{H}(X,Y) &= \nabla_Y tr(XY^\top C)\\ &= \nabla_Y tr(Y^\top CX)\\ &= (\nabla_Y^\top tr(Y^\top CX))^\top\\ &= CX\\ &= CAB\end{aligned}$$

Thus, $\nabla_A tr(ABA^\top C) = C^\top AB^\top + CAB$

Matrix-by-matrix derivative formula

Tanmoy Chakraborty

August 22, 2024

Given:
- A is an $m \times n$ matrix.
- B is an $n \times p$ matrix.

We want to compute the derivative of $C = AB$ with respect to A, where C is an $m \times p$ matrix. We are interested in how each element of C changes with respect to each element of A.

Step-by-Step Derivation

Let $C = AB$, where:

$$C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$$

i.e., the (i, j)-th element of C is the dot product of the i-th row of A and the j-th column of B.

Now, we compute the derivative of C_{ij} with respect to A_{pq}. There are two cases:

1. **When** $p = i$: In this case, A_{ik} is directly involved in the sum, so the derivative will be non-zero when $q = k$.

2. **When** $p \neq i$: The derivative will be zero because the row of A corresponding to p does not influence the row corresponding to i.

The derivative of C_{ij} with respect to A_{pq} is:

$$\frac{\partial C_{ij}}{\partial A_{pq}} = B_{qj} \delta_{ip}$$

where δ_{ip} is the *Kronecker delta*, which is 1 if $i = p$, and 0 otherwise.

Derivative of the Matrix Product AB

Now we can express the full derivative of the matrix product AB with respect to A:

$$\frac{\partial(AB)}{\partial A} = B^T$$

This derivative can be interpreted as follows: - The derivative of the matrix product $C = AB$ with respect to A results in the matrix B^T, the transpose of B.

Tensor Formulation

In tensor form, the derivative of the matrix $C = AB$ (i.e., treating the derivative of each element of C with respect to each element of A as a separate partial derivative) is a third-order tensor, where:

$$\frac{\partial(C_{ij})}{\partial A_{pq}} = B_{qj}\delta_{ip}$$

This means that for every element of C, the derivative with respect to A depends only on the corresponding elements of B, and is zero whenever the row index p does not match the row index i of C.

Matrix Notation Summary

To summarize the key points:

$$\frac{\partial(AB)}{\partial A} = B^T$$

This is a fundamental result in matrix calculus, often used in optimization problems involving matrix products.