

Proof of formula for derivatives of a matrix within a Frobenius norm

For matrices \mathbf{W} , \mathbf{A} , \mathbf{B} and \mathbf{C} defined such that the following equation makes sense, we have

$$\frac{\partial \|\mathbf{AWB} + \mathbf{C}\|_2^2}{\partial \mathbf{W}} = 2\mathbf{A}^\top (\mathbf{AWB} + \mathbf{C})\mathbf{B}^\top \quad (1)$$

Proof: let's write $\mathbf{Z} = \mathbf{AWB} + \mathbf{C}$.

$\|\mathbf{AWB} + \mathbf{C}\|_2^2 = \|\mathbf{Z}\|_2^2 = \sum_i \sum_j (Z_{i,j})^2$, and

$$\frac{\partial \|\mathbf{Z}\|_2^2}{\partial \mathbf{W}} = \sum_i \sum_j \frac{\partial (Z_{i,j})^2}{\partial \mathbf{W}} \quad (2)$$

We have $Z_{i,j} = \sum_k A_{i,k}(\mathbf{WB})_{k,j} + C_{i,j}$, or

$$Z_{i,j} = \sum_k A_{i,k} \left(\sum_l W_{k,l} B_{l,j} \right) + C_{i,j} \quad (3)$$

In Equation (2), for any $W_{m,n}$, we have

$$\frac{\partial (Z_{i,j})^2}{\partial W_{m,n}} = 2Z_{i,j} \frac{\partial Z_{i,j}}{\partial W_{m,n}} \quad (4)$$

Using Equation (3),

$$\frac{\partial Z_{i,j}}{\partial W_{m,n}} = A_{i,m} B_{n,j} \quad (5)$$

Combining equations (2), (4) and (5), we have

$$\frac{\partial \|\mathbf{Z}\|_2^2}{\partial W_{m,n}} = \sum_i \sum_j (2Z_{i,j} A_{i,m} B_{n,j}) = 2 \sum_i (\mathbf{A}^\top)_{m,i} \left(\sum_j Z_{i,j} (\mathbf{B}^\top)_{j,n} \right)$$

and finally

$$\frac{\partial \|\mathbf{Z}\|_2^2}{\partial W_{m,n}} = 2(\mathbf{A}^\top \mathbf{Z} \mathbf{B}^\top)_{m,n}$$

Q.E.D.