# Explanatory Notes for 6.390

Shaunticlair Ruiz (Current TA)

Fall 2022

### 7.X.14 Derivative: matrix/scalar

Now, we have our general form for creating derivatives.

We'll get our derivative of the form

$$\frac{\partial(\text{Matrix})}{\partial(\text{Scalar})} = \frac{\partial M}{\partial s} \tag{1}$$

We have a matrix $M$ in the shape $(r \times k)$ and a scalar $s$. Our **input** is a **scalar**, and our **output** is a **matrix**.

$$M = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1r} \\ m_{21} & m_{22} & \cdots & m_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ m_{k1} & m_{k2} & \cdots & m_{kr} \end{bmatrix} \tag{2}$$

This may seem concerning: before, we divided **inputs** across **rows**, and **outputs** across **columns**. But in this case, we have **no** input axes, and **two** output axes.

Well, let's try to make this work anyway.

What did we do before, when we didn't know how to handle a **new** derivative? We compared it to **old** versions: we built our vector/vector case using the vector/scalar case and the scalar/vector case.

We did this by **compressing** one of our *vectors* into a *scalar* temporarily: this works, because we want to treat each of these objects the **same way**.

We don't know how to work with Matrix/Scalar, but what's the **closest** thing we do know? **Vector/Scalar**.

How do we accomplish that? As we saw above, a matrix is a **vector** of **vectors**. We could turn it into a **vector** of **scalars**.

**Concept 1**

A **matrix** can be thought of as a **column vector** of **row vectors** (or vice versa).

So, we can use our earlier technique and convert the **row vectors** into **scalars**.

We'll replace the **row vectors** in our matrix with **scalars**.

$$M = \begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_k \end{bmatrix} \tag{3}$$

Now, we can pretend our matrix is a vector! We've got a derivative for that:

$$\frac{\partial M}{\partial s} = \begin{bmatrix} \frac{\partial M_1}{\partial s} & \frac{\partial M_2}{\partial s} & \cdots & \frac{\partial M_r}{\partial s} \end{bmatrix} \tag{4}$$

Aha - we have the same form that we did for our vector/vector derivative! Each derivative is a column vector. Let's expand it out:

$$\frac{\partial M}{\partial s} = \left. \begin{bmatrix} \overbrace{\begin{bmatrix} \frac{\partial m_{11}}{\partial s} \\ \frac{\partial m_{12}}{\partial s} \\ \vdots \\ \frac{\partial m_{1r}}{\partial s} \end{bmatrix}, \begin{bmatrix} \frac{\partial m_{21}}{\partial s} \\ \frac{\partial m_{22}}{\partial s} \\ \vdots \\ \frac{\partial m_{2r}}{\partial s} \end{bmatrix}, \cdots \begin{bmatrix} \frac{\partial m_{k1}}{\partial s} \\ \frac{\partial m_{k2}}{\partial s} \\ \vdots \\ \frac{\partial m_{kr}}{\partial s} \end{bmatrix}}^{\text{Column } j \text{ matches } m_{j?}} \end{bmatrix} \right\} \text{Row } i \text{ matches } m_{?i} \tag{5}$$

**Definition 2**

If $M$ is a matrix in the shape $(r \times k)$ and $s$ is a scalar,

Then we define the **matrix derivative** $\partial M / \partial s$ as the $(k \times r)$ matrix:

$$\frac{\partial M}{\partial s} = \left. \overbrace{\begin{bmatrix} \frac{\partial m_{11}}{\partial s} & \frac{\partial m_{21}}{\partial s} & \cdots & \frac{\partial m_{k1}}{\partial s} \\ \frac{\partial m_{12}}{\partial s} & \frac{\partial m_{22}}{\partial s} & \cdots & \frac{\partial m_{k2}}{\partial s} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial m_{1r}}{\partial s} & \frac{\partial m_{2r}}{\partial s} & \cdots & \frac{\partial m_{kr}}{\partial s} \end{bmatrix}}^{\text{Column } j \text{ matches } m_{j?}} \right\} \text{Row } i \text{ matches } m_{?i}$$

This matrix has the transpose of the shape of $M$.

### 7.X.15 Derivative: scalar/matrix

We'll get our derivative of the form

$$\frac{\partial(\text{Scalar})}{\partial(\text{Matrix})} = \frac{\partial s}{\partial M} \tag{6}$$

We have a matrix $M$ in the shape $(r \times k)$ and a scalar $s$. Our **input** is a **matrix**, and our **output** is a **scalar**.

Let's do what we did last time: break it into **row vectors**.

$$M = \begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_k \end{bmatrix} \tag{7}$$

The gradient for this "vector" gives us a **column vector**:

$$\frac{\partial s}{\partial M} = \begin{bmatrix} \frac{\partial s}{\partial M_1} \\ \frac{\partial s}{\partial M_2} \\ \vdots \\ \frac{\partial s}{\partial M_k} \end{bmatrix} \tag{8}$$

This time, each derivative is a **row vector**. Let's **expand**:

$$\frac{\partial s}{\partial M} = \begin{bmatrix} \begin{bmatrix} \frac{\partial s}{\partial m_{11}} & \frac{\partial s}{\partial m_{12}} & \cdots & \frac{\partial s}{\partial m_{1r}} \end{bmatrix} \\ \begin{bmatrix} \frac{\partial s}{\partial m_{21}} & \frac{\partial s}{\partial m_{22}} & \cdots & \frac{\partial s}{\partial m_{2r}} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \frac{\partial s}{\partial m_{k1}} & \frac{\partial s}{\partial m_{k2}} & \cdots & \frac{\partial s}{\partial m_{kr}} \end{bmatrix} \end{bmatrix} \tag{9}$$

**Definition 3**

If $M$ is a matrix in the shape $(r \times k)$ and $s$ is a scalar,

Then we define the **matrix derivative** $\partial s/\partial M$ as the $(r \times k)$ matrix:

$$\frac{\partial s}{\partial M} = \begin{array}{cl} \text{Column } j \text{ matches } m_{?j} & \\ \begin{bmatrix} \frac{\partial s}{\partial m_{11}} & \frac{\partial s}{\partial m_{12}} & \cdots & \frac{\partial s}{\partial m_{1r}} \\ \frac{\partial s}{\partial m_{21}} & \frac{\partial s}{\partial m_{22}} & \cdots & \frac{\partial s}{\partial m_{2r}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial s}{\partial m_{k1}} & \frac{\partial s}{\partial m_{k2}} & \cdots & \frac{\partial s}{\partial m_{kr}} \end{bmatrix} & \text{Row } i \text{ matches } m_{i?} \end{array}$$

This matrix has the same shape as $M$.