

Explanatory Notes for 6.390

Shauntclair Ruiz (Current TA)

Fall 2022

X.10 Vectors and vectors: vector input, vector output

We'll be combining our two previous derivatives:

$$\frac{\partial(\text{Vector})}{\partial(\text{Vector})} = \frac{\partial \mathbf{w}}{\partial \mathbf{v}} \quad (1)$$

\mathbf{v} and \mathbf{w} are both **vectors**: thus, input and output are both **vectors**.

$$\Delta \mathbf{v} \longrightarrow \boxed{\mathbf{f}} \longrightarrow \Delta \mathbf{w} \quad (2)$$

Written out, we get:

$$\begin{array}{c} \overbrace{\begin{bmatrix} \Delta v_1 \\ \Delta v_2 \\ \vdots \\ \Delta v_m \end{bmatrix}}^{\Delta \mathbf{v}} \longrightarrow \begin{array}{c} \overbrace{\begin{bmatrix} \Delta w_1 \\ \Delta w_2 \\ \vdots \\ \Delta w_n \end{bmatrix}}^{\Delta \mathbf{w}} \end{array} \quad (3)$$

Something pretty complicated! We have m inputs and n outputs. Every input can interact with every output.

So, our derivative needs to have mn different elements. That's a lot!

X.11 The vector/vector derivative

We return to our rule from before. We'll skip the star notation, and jump right to the equation we've gotten for both of our two previous derivatives:

$$\Delta \mathbf{w} = \left(\frac{\partial \mathbf{w}}{\partial \mathbf{v}} \right)^T \Delta \mathbf{v} \quad (4)$$

Hopefully, since we're combining two different derivatives, we should be able to use the same rule here.

With mn different elements, this could get messy very fast. Let's see if we can focus on only **part** of our problem:

$$\begin{bmatrix} \Delta w_1 \\ \Delta w_2 \\ \vdots \\ \Delta w_n \end{bmatrix} = \left(\frac{\partial \mathbf{w}}{\partial \mathbf{v}} \right)^T \begin{bmatrix} \Delta v_1 \\ \Delta v_2 \\ \vdots \\ \Delta v_m \end{bmatrix} \quad (5)$$

One input

We could try focusing on just a single **input** or a single **output**, to simplify things. Let's start with a single v_i .

$$\overbrace{\begin{bmatrix} \Delta w_1 \\ \Delta w_2 \\ \vdots \\ \Delta w_n \end{bmatrix}}^{\Delta w \text{ from } v_i} = \left(\frac{\partial w}{\partial v_i} \right)^T \Delta v_i \quad (6)$$

We now have a simpler case: $\partial \text{Vector} / \partial \text{Scalar}$. We're familiar with this case!

$$\frac{\partial w}{\partial v_i} = \left[\frac{\partial w_1}{\partial v_i}, \frac{\partial w_2}{\partial v_i}, \dots, \frac{\partial w_n}{\partial v_i} \right] \quad (7)$$

We get a vector. What if the **output** is a scalar instead?

One output

$$\Delta w_j = \left(\frac{\partial w_j}{\partial v} \right)^T \begin{bmatrix} \Delta v_1 \\ \Delta v_2 \\ \vdots \\ \Delta v_m \end{bmatrix} \quad (8)$$

We have $\partial \text{Scalar} / \partial \text{Vector}$:

$$\frac{\partial w_j}{\partial v} = \begin{bmatrix} \partial w_j / \partial v_1 \\ \partial w_j / \partial v_2 \\ \vdots \\ \partial w_j / \partial v_m \end{bmatrix} \quad (9)$$

So, our vector-vector derivative is a **generalization** of the two derivatives we did before!

It seems that extending along the **vertical** axis changes our v_i value, while moving along the **horizontal** axis changes our w_j value.

~~~~~

## X.12 General derivative

You might have a hint of what we get: one derivative stretches us along **one** axis, the other along the **second**.

To prove it to ourselves, we can **combine** these concepts. We'll handle solve as if we have one vector, and then **substitute** in the second one.

**Concept 1**

One way to **simplify** our work is to treat **vectors** as **scalars**, and then convert them back into **vectors** after applying some math.

We have to be careful - any operation we apply to the **scalar**, has to match how the **vector** would behave.

This is **equivalent** to if we just focused on one scalar inside our vector, and then stacked all those scalars back into the vector.

This isn't just a cute trick: it relies on an understanding that, at its **basic** level, we're treating **scalars** and **vectors** and **matrices** as the same type of object: a structured array of numbers.

We'll get into "arrays" later.

As always, our goal is to **simplify** our work, so we can handle each piece of it.

- We treat  $\Delta v$  as a scalar so we can get the simplified derivative.

$$\Delta w = \left( \frac{\partial w}{\partial v} \right)^T \Delta v \quad (10)$$

We'll only expand **one** of our vectors, since we know how to manage **one** of them.

$$\begin{bmatrix} \Delta w_1 \\ \Delta w_2 \\ \vdots \\ \Delta w_n \end{bmatrix} = \left( \frac{\partial w}{\partial v} \right)^T \Delta v \quad (11)$$

This time, notice that we **didn't** simplify  $v$  to  $v_i$ . We didn't **remove** the other elements - we still have a full **vector**. But, let's treat it as if it *were* a scalar.

This comes out to:

$$\frac{\partial w}{\partial v} = \overbrace{\left[ \frac{\partial w_1}{\partial v}, \frac{\partial w_2}{\partial v}, \dots, \frac{\partial w_n}{\partial v} \right]}^{\text{Column } j \text{ matches } w_j} \quad (12)$$

- Our "answer" is a row vector. But, each of those derivatives is a **column** vector!

Now that we've taken care of  $\partial w_j$  (one for each column), we can expand our derivatives in terms of  $\partial v_i$ .

First, for  $w_1$ :

$$\frac{\partial \mathbf{w}}{\partial \mathbf{v}} = \left[ \begin{array}{c} \left[ \begin{array}{c} \frac{\partial w_1}{\partial v_1} \\ \frac{\partial w_1}{\partial v_2} \\ \vdots \\ \frac{\partial w_1}{\partial v_m} \end{array} \right] , \quad \frac{\partial w_2}{\partial \mathbf{v}} , \quad \dots \quad \frac{\partial w_n}{\partial \mathbf{v}} \end{array} \right] \quad \left. \begin{array}{l} \text{Column } j \text{ matches } w_j \\ \text{Row } i \text{ matches } v_i \end{array} \right\} \quad (13)$$

And again, for  $w_2$ :

$$\frac{\partial \mathbf{w}}{\partial \mathbf{v}} = \left[ \begin{array}{c} \left[ \begin{array}{c} \frac{\partial w_1}{\partial v_1} \\ \frac{\partial w_1}{\partial v_2} \\ \vdots \\ \frac{\partial w_1}{\partial v_m} \end{array} \right] , \quad \left[ \begin{array}{c} \frac{\partial w_2}{\partial v_1} \\ \frac{\partial w_2}{\partial v_2} \\ \vdots \\ \frac{\partial w_2}{\partial v_m} \end{array} \right] , \quad \dots \quad \frac{\partial w_n}{\partial \mathbf{v}} \end{array} \right] \quad \left. \begin{array}{l} \text{Column } j \text{ matches } w_j \\ \text{Row } i \text{ matches } v_i \end{array} \right\} \quad (14)$$

And again, for  $w_n$ :

$$\frac{\partial \mathbf{w}}{\partial \mathbf{v}} = \left[ \begin{array}{c} \left[ \begin{array}{c} \frac{\partial w_1}{\partial v_1} \\ \frac{\partial w_1}{\partial v_2} \\ \vdots \\ \frac{\partial w_1}{\partial v_m} \end{array} \right] , \quad \left[ \begin{array}{c} \frac{\partial w_2}{\partial v_1} \\ \frac{\partial w_2}{\partial v_2} \\ \vdots \\ \frac{\partial w_2}{\partial v_m} \end{array} \right] , \quad \dots \quad \left[ \begin{array}{c} \frac{\partial w_n}{\partial v_1} \\ \frac{\partial w_n}{\partial v_2} \\ \vdots \\ \frac{\partial w_n}{\partial v_m} \end{array} \right] \end{array} \right] \quad \left. \begin{array}{l} \text{Column } j \text{ matches } w_j \\ \text{Row } i \text{ matches } v_i \end{array} \right\} \quad (15)$$

We have column vectors in our row vector... let's just combine them into a **matrix**.

**Definition 2**

If

- $\mathbf{v}$  is an  $(m \times 1)$  **vector**
- $\mathbf{w}$  is an  $(n \times 1)$  **vector**

Then we define the **vector derivative**  $\partial \mathbf{w} / \partial \mathbf{v}$  as fulfilling:

$$\Delta \mathbf{w} = \left( \frac{\partial \mathbf{w}}{\partial \mathbf{s}} \right)^T \Delta \mathbf{s}$$

Thus, our derivative must be a  $(1 \times n)$  vector

$$\frac{\partial \mathbf{w}}{\partial \mathbf{v}} = \begin{matrix} \text{Column } j \text{ matches } \mathbf{w}_j \\ \left[ \begin{array}{cccc} \frac{\partial w_1}{\partial v_1} & \frac{\partial w_2}{\partial v_1} & \dots & \frac{\partial w_n}{\partial v_1} \\ \frac{\partial w_1}{\partial v_2} & \frac{\partial w_2}{\partial v_2} & \dots & \frac{\partial w_n}{\partial v_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial w_1}{\partial v_m} & \frac{\partial w_2}{\partial v_m} & \dots & \frac{\partial w_n}{\partial v_m} \end{array} \right] \end{matrix} \left. \vphantom{\begin{matrix} \frac{\partial w_1}{\partial v_1} \\ \frac{\partial w_1}{\partial v_2} \\ \vdots \\ \frac{\partial w_1}{\partial v_m} \end{matrix}} \right\} \begin{matrix} \text{Row } i \text{ matches } v_i \end{matrix}$$

This general form can be used for **any** of our matrix derivatives.

So, our matrix can represent any **combination** of two elements! We just assign each **row** to a  $v_i$  component, and each **column** with a  $w_j$  component.

**X.13 More about the vector/vector derivative**Let's show a specific example:  $\mathbf{w}$  is  $(3 \times 1)$ ,  $\mathbf{v}$  is  $(2 \times 1)$ .

$$\frac{\partial \mathbf{w}}{\partial \mathbf{v}} = \left[ \begin{array}{cc|c} \overbrace{\frac{\partial w_1}{\partial v_1}}^{w_1} & \overbrace{\frac{\partial w_2}{\partial v_1}}^{w_2} & \overbrace{\frac{\partial w_3}{\partial v_1}}^{w_3} \\ \frac{\partial w_1}{\partial v_2} & \frac{\partial w_2}{\partial v_2} & \frac{\partial w_3}{\partial v_2} \end{array} \right] \left. \vphantom{\begin{matrix} \frac{\partial w_1}{\partial v_1} \\ \frac{\partial w_1}{\partial v_2} \end{matrix}} \right\} \begin{matrix} v_1 \\ v_2 \end{matrix} \quad (16)$$

Another way to describe the general case:

**Notation 3**

Our matrix  $\partial \mathbf{w} / \partial \mathbf{v}$  is entirely filled with **scalar derivatives**

$$\frac{\partial w_j}{\partial v_i} \quad (17)$$

Where any one **derivative** is stored in

- Row  $i$ 
  - $m$  rows total
- Column  $j$ 
  - $n$  columns total

We can also compress it along either axis (just like how we did to derive this result):

**Notation 4**

Our matrix  $\partial \mathbf{w} / \partial \mathbf{v}$  can be written as

$$\frac{\partial \mathbf{w}}{\partial \mathbf{v}} = \overbrace{\left[ \frac{\partial w_1}{\partial \mathbf{v}}, \frac{\partial w_2}{\partial \mathbf{v}}, \dots, \frac{\partial w_n}{\partial \mathbf{v}} \right]}^{\text{Column } j \text{ matches } w_j}$$

or

$$\frac{\partial \mathbf{w}}{\partial \mathbf{v}} = \left[ \begin{array}{c} \frac{\partial w}{\partial v_1} \\ \frac{\partial w}{\partial v_2} \\ \vdots \\ \frac{\partial w}{\partial v_m} \end{array} \right] \left. \vphantom{\begin{array}{c} \frac{\partial w}{\partial v_1} \\ \frac{\partial w}{\partial v_2} \\ \vdots \\ \frac{\partial w}{\partial v_m} \end{array}} \right\} \text{Row } i \text{ matches } v_i$$

These compressed forms will be useful for deriving our new and final derivatives, **matrix-scalar** pairs.