$$\begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = f(\theta) = f(\theta_1, \theta_2, \theta_3, \theta_4)$$

$$= \begin{bmatrix} \theta_1 & \theta_2 \\ \theta_3 & \theta_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$\llcorner$ —— fixed (dataset

$$\begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \\ a_5 & a_6 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = g(a_1, a_2, a_3 \ldots a_6)$$

Jacobian $\dfrac{\partial h}{\partial \theta} = \begin{bmatrix} \dfrac{\partial h_1}{\partial \theta_1} & \dfrac{\partial h_1}{\partial \theta_2} & \dfrac{\partial h_1}{\partial \theta_3} & \dfrac{\partial h_1}{\partial \theta_4} \\[2mm] \dfrac{\partial h_2}{\partial \theta_1} & \dfrac{\partial h_2}{\partial \theta_2} & \dfrac{\partial h_2}{\partial \theta_3} & \dfrac{\partial h_2}{\partial \theta_4} \end{bmatrix}$
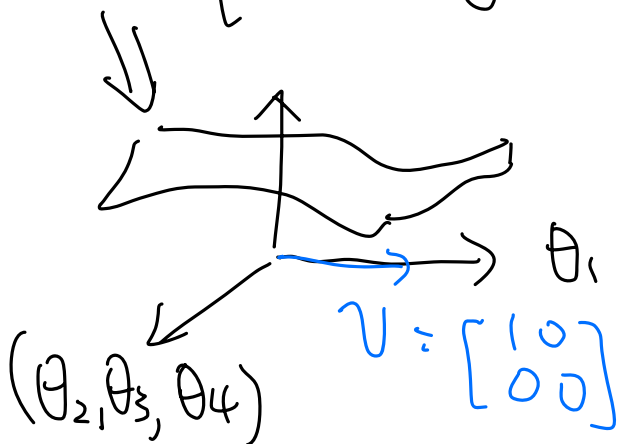
$\dfrac{\partial h}{\partial \theta} v \overset{R^4 \text{ in } \theta \text{ space}}{\downarrow} = \begin{bmatrix} v_1 & v_2 \\ v_3 & v_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

(by def $\lim\limits_{\varepsilon} \dfrac{f(\theta + \varepsilon v) - f(\theta)}{\varepsilon}$)

When $v = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ $\Rightarrow$ $\dfrac{\partial h}{\partial \theta} v = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$$= \begin{bmatrix} x_1 \\ 0 \end{bmatrix} = \begin{bmatrix} \dfrac{\partial h_1}{\partial \theta_1} \\[2mm] \dfrac{\partial h_2}{\partial \theta_1} \end{bmatrix}$$



$(\theta_2, \theta_3, \theta_4)$ $\theta_1$

$v = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$

Jacobian $\dfrac{\partial q}{\partial h} = \begin{bmatrix} \dfrac{\partial q_1}{\partial h_1} & \dfrac{\partial q_1}{\partial h_2} \\[10pt] \dfrac{\partial q_2}{\partial h_1} & \dfrac{\partial q_2}{\partial h_2} \\[10pt] \dfrac{\partial q_3}{\partial h_1} & \dfrac{\partial q_3}{\partial h_3} \end{bmatrix}$

$$\frac{\partial q}{\partial h} v_h = \lim_{\varepsilon} \frac{\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \\ a_5 & a_6 \end{bmatrix} \begin{bmatrix} h_1 + \varepsilon v_{h_1} \\ h_2 + \varepsilon v_{h_2} \end{bmatrix} - \begin{bmatrix} a_1 & a_2 \\ \cdots \\ \cdots \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}}{\varepsilon}$$

$$= \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \\ a_5 & a_6 \end{bmatrix} \begin{bmatrix} v_{h_1} \\ v_{h_2} \end{bmatrix}$$

When you use the previous $\begin{bmatrix} \dfrac{\partial h_1}{\partial \theta_1} \\[10pt] \dfrac{\partial h_1}{\partial \theta_2} \end{bmatrix}$,

we are essentially computing

$$
\begin{bmatrix}
\dfrac{\partial q_1}{\partial h_1} & \dfrac{\partial q_1}{\partial h_2} \\[6pt]
\dfrac{\partial q_2}{\partial h_1} & \dfrac{\partial q_2}{\partial h_2} \\[6pt]
\dfrac{\partial q_3}{\partial h_1} & \dfrac{\partial q_3}{\partial h_3}
\end{bmatrix}
\begin{bmatrix}
\dfrac{\partial h_1}{\partial \theta_1} \\[10pt]
\dfrac{\partial h_2}{\partial \theta_1}
\end{bmatrix}
$$

$$
=
\begin{bmatrix}
\dfrac{\partial q_1}{\partial h_1}\cdot\dfrac{\partial h_1}{\partial \theta_1} & \dfrac{\partial q_1}{\partial h_2}\dfrac{\partial h_2}{\partial \theta_1} \\[10pt]
- & \sim & - \\[6pt]
- & \sim & \smile & \sim
\end{bmatrix}
$$

via evaluating:

$$
=
\begin{bmatrix}
a_1 & a_2 \\
a_3 & a_4 \\
a_5 & a_6
\end{bmatrix}
\begin{bmatrix}
x_1 \\
0
\end{bmatrix}
$$

( push forwards )

Contrast to reverse mode,

When you already have:

$$\begin{bmatrix} \frac{\partial L}{\partial h_1} \\ \frac{\partial L}{\partial h_2} \end{bmatrix}, \text{ to obtain } \begin{bmatrix} \frac{\partial L}{\partial \theta_1} & \frac{\partial L}{\partial \theta_2} \\ \frac{\partial L}{\partial \theta_3} & \frac{\partial L}{\partial \theta_4} \end{bmatrix}$$

by chain rule:

$$\frac{\partial L}{\partial \theta_1} = \frac{\partial L}{\partial h_1} \frac{\partial h_1}{\partial \theta_1} + \frac{\partial L}{\partial h_2} \frac{\partial h_2}{\partial \theta_1}$$

$$\frac{\partial L}{\partial \theta_2} = \frac{\partial L}{\partial h_1} \frac{\partial h_1}{\partial \theta_2} + \frac{\partial L}{\partial h_2} \frac{\partial h_2}{\partial \theta_2}$$

$$\frac{\partial L}{\partial \theta_3} = \frac{\partial L}{\partial h_1} \frac{\partial h_1}{\partial \theta_3} + \frac{\partial L}{\partial h_2} \frac{\partial h_2}{\partial \theta_3}$$

$$\frac{\partial L}{\partial \theta_4} = \quad - \quad - \quad - \quad -$$

$$\Rightarrow \frac{\partial L}{\partial \theta_1} = \frac{\partial L}{\partial h_1} x_1 + \frac{\partial L}{\partial h_2} \cdot 0$$

$$\frac{\partial L}{\partial \theta_2} = \frac{\partial L}{\partial h_1} x_2 + \frac{\partial L}{\partial h_2} \times 0$$

$$\frac{\partial L}{\partial \theta_3} = \frac{\partial L}{\partial h_1} \cdot 0 + \frac{\partial L}{\partial h_2} \times x_1$$

$$\frac{\partial L}{\partial \theta_4} = \frac{\partial L}{\partial h_1} \times 0 + \frac{\partial L}{\partial h_2} \times x_2$$

$$\Rightarrow \begin{bmatrix} \frac{\partial L}{\partial \theta_1} & \frac{\partial L}{\partial \theta_2} \\ \frac{\partial L}{\partial \theta_3} & \frac{\partial L}{\partial \theta_4} \end{bmatrix} = \begin{bmatrix} \frac{\partial L}{\partial h_1} \\ \frac{\partial L}{\partial h_2} \end{bmatrix} \begin{bmatrix} x_1 & x_2 \end{bmatrix}$$

Connection to vector-jacobian product

$$\frac{\partial L}{\partial \theta} = \underbrace{\frac{\partial L}{\partial h}}_{vector} \cdot \boxed{\frac{\partial h}{\partial \theta}} \quad \longrightarrow \text{ no need to materialize}$$

$$\underset{Jacobian}{}$$

via evaluating

$$= \begin{bmatrix} \frac{\partial L}{\partial h_1} \\ \frac{\partial L}{\partial h_2} \end{bmatrix} \begin{bmatrix} x_1 & x_2 \end{bmatrix}$$