Matrix Calculus

Lecture by Farrukh Rahman, transcribed by Matthew Caseres

Scalar-valued functions of a single variable

Use the chain rule to take the derivative of the function below.

$$\begin{aligned}
&f : \mathbb{R} \to \mathbb{R}, \quad x \in \mathbb{R} \\
&y = f(x) = x^2 \ln\left(x^2\right) \\
&y' = \tfrac{df}{dx} = 2x\left(1 + \ln\left(x^2\right)\right)
\end{aligned}$$

Vector-valued functions of a single variable

A function can take a single variable and output a vector.

$$\begin{aligned}
&f : \mathbb{R} \to \mathbb{R}^3 \\
&f(x) = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} = \begin{bmatrix} 3x^2 + x \\ e^x \\ \ln(x) \end{bmatrix} \\
&\tfrac{df}{dx} = \begin{bmatrix} \frac{df_1}{dx} \\ \frac{\partial f_2}{dx} \\ \frac{df_3}{dx} \end{bmatrix} = \begin{bmatrix} 6x + 1 \\ e^x \\ \frac{1}{x} \end{bmatrix} \in \mathbb{R}^{3 \times 1}
\end{aligned}$$

Matrix valued functions of a single variable

Differentiate each function with respect to the variable.

$$\begin{aligned}
&f(x) = \begin{bmatrix} f_1 & f_2 \\ f_3 & f_4 \end{bmatrix} = \begin{bmatrix} \cos(x) & \sin(x) \\ e^x & \tanh(x) \end{bmatrix} \\
&\tfrac{df}{dx} = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_2}{\partial x} \\ \frac{\partial f_3}{\partial x} & \frac{\partial f_4}{\partial x} \end{bmatrix} = \begin{bmatrix} -\sin(x) & \cos(x) \\ e^x & 1 - \tanh^2(x) \end{bmatrix}
\end{aligned}$$

Scalar-valued functions of a multiple variables

We take the derivative with respect to the transposed elements of x.

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$y = x_1 x_2 x_3$$

$$\frac{dy}{dx} = \begin{bmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} & \frac{\partial y}{\partial x_3} \end{bmatrix} = \begin{bmatrix} x_2 x_3 & x_1 x_3 & x_1 x_2 \end{bmatrix}$$

This is in numerator layout.

Scalar-valued functions of matrices

We take the derivative with respect to the transposed elements of x. This is a generalization of the previous case.

$$
\begin{aligned}
&f : \mathbb{R}^{2\times 3} \to \mathbb{R} \\
&x = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix} \\
&f(x) = \sum_i \sum_j x_{ij} \\
&\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{21}} \\ \frac{\partial f}{\partial x_{12}} & \frac{\partial f}{\partial x_{22}} \\ \frac{\partial f}{\partial x_{13}} & \frac{\partial f}{\partial x_{23}} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 2}
\end{aligned}
$$

Vector-valued functions of multiple variables

We differentiate each function in the vector with respect to the vector x. Each component function f_i is treated as a scalar-valued function when differentiated.

$$
\begin{aligned}
&f : R^3 \to R^2 \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\
&f(\vec{x}) = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_1 x_2 + x_1 x_3 \\ x_1^2 + 2x_3 \end{bmatrix} \\
&\frac{df(\vec{x})}{d\vec{x}} = \begin{bmatrix} \dfrac{df_1}{dx} \\ \dfrac{df_2}{dx} \end{bmatrix} = \begin{bmatrix} \frac{\delta f_1}{\delta x_1} & \frac{\delta f_1}{\partial x_2} & \frac{\delta f_1}{\delta x_3} \\ \frac{\delta f_2}{\partial x_1} & \frac{\delta f_2}{\delta x_2} & \frac{\partial f_2}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 1 + x_2 + x_3 & x_1 & x_1 \\ 2x_1 & 0 & 2 \end{bmatrix}
\end{aligned}
$$

Another example (identity function)

$$
\begin{aligned}
x &= \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\
f(x) &= \begin{bmatrix} f_1 \\ f_2 \\ f_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\
\frac{\partial f}{\partial x} &= \begin{bmatrix} \frac{\partial f_1}{\partial x} \\ \frac{\partial f_2}{\partial x} \\ \frac{\partial f_3}{\partial x} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
\end{aligned}
$$

Vector-valued functions of a matrix

We differentiate each function in the vector with respect to the matrix x.

$$
f : \mathbb{R}^{2\times 3} \to \mathbb{R}^4
$$

$$
x = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix}
$$

$$f(x) = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{bmatrix} = \begin{bmatrix} \sum_i \sum_j x_{ij} \\ x_{11}x_{22}x_{23} \\ x_{12}x_{21} \\ x_{21}x_{22}x_{23} \end{bmatrix}$$

$$\frac{df}{dx} = \begin{bmatrix} \frac{df_1}{dx} \\ \frac{df_2}{dx} \\ \frac{df_3}{dx} \\ \frac{df_4}{dx} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \\ \begin{bmatrix} x_{22}x_{23} & 0 \\ 0 & x_{11}x_{23} \\ 0 & x_{11}x_{22} \end{bmatrix} \\ \begin{bmatrix} 0 & x_{12} \\ x_{21} & 0 \\ 0 & 0 \end{bmatrix} \\ \begin{bmatrix} 0 & x_{22}x_{23} \\ 0 & x_{21}x_{23} \\ 0 & x_{21}x_{22} \end{bmatrix} \end{bmatrix} \in \mathbb{R}^{4\times3\times2}$$

According to Wikipedia numerator layout is "lay out according to $\mathbf{y}$ and $\mathbf{x}^T$" and that denominator layour is "according to $\mathbf{y}^T$ and $\mathbf{x}$". No examples here are in denominator format, and the slides seem to follow this rule as well. This is why everything is in the shape of y and then nested inside it is in the shape of x^T.

Matrix-valued functions of multiple variables

Much the same as for vector-valued functions. Regardless the dimension of x the answer will be

$$f(x) = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix}$$

$$\frac{df}{dx} = \begin{bmatrix} \frac{df_{11}}{dx} & \frac{df_{12}}{dx} \\ \frac{df_{21}}{dx} & \frac{df_{22}}{dx} \end{bmatrix}$$

So we have a $\mathbf{2} \times \mathbf{2} \times \left(\textbf{the dimensions from } \frac{df_{ij}}{dx} \textbf{ which are same as } x^T\right)$.