Univ. of Washington

Lecture : Vector Derivatives

Winter 2021

Lecturer: Dan Calderone

1 Matrix Derivatives

Derivatives are linear maps that convert perturbations in function arguments into perturbations in the function themselves. Consider $x \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$. $f(x)$ is a scalar. The derivative $\frac{\partial f}{\partial x}$ is the row vector

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}$$

such that

$$\Delta f \approx \frac{\partial f}{\partial x} \Delta x = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{bmatrix} \tag{1}$$

where $\Delta f \in \mathbb{R}$ and $\Delta x \in \mathbb{R}^n$ are perturbations in f and x, respectively. Note that if f is linear, ie. $f(x) = b^T x$, then $\frac{\partial f}{\partial x} = b^T$. Note that the perturbation form in (1) can be useful in computing vector derivatives in tricky situations. For example, suppose $f(x) = x^T Q x + b^T x$. In order to compute the derivative, we can perturb each instance of x separately and add up the perturbations. (The ability to perturb each instance of x separately is called the *product rule*.) Then we rearrange the right hand side (RHS) into the form of (1).

$$\Delta f = \Delta x^T Q x + x^T Q \Delta x + b^T \Delta x \tag{2}$$

Noticing that each of the terms in the RHS is a scalar, we can transpose as necessary.

$$\begin{aligned} \Delta f &= (\Delta x^T Q x)^T + x^T Q \Delta x + b^T \Delta x && (3) \\ &= \big(x^T (Q + Q^T) + b^T\big) \Delta x && (4) \\ \Rightarrow \qquad \frac{\partial f}{\partial x} &= x^T (Q + Q^T) + b^T && (5) \end{aligned}$$

Now suppose $f(x)$ is a vector valued function $f : \mathbb{R}^n \to \mathbb{R}^n$. The derivative is now an $m \times n$ matrix

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \tag{6}$$

such that

$$\Delta f = \begin{bmatrix} \Delta f_1 \\ \vdots \\ \Delta f_m \end{bmatrix} \approx \frac{\partial f}{\partial x} \Delta x = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{bmatrix} \tag{7}$$

where $\Delta f \in \mathbb{R}^m$ and $\Delta x \in \mathbb{R}^n$. Note that when $\frac{\partial f}{\partial x}$ is a matrix it is referred to as a *Jacobian*.

Now suppose we have a scalar function $f(x)$ and we want to compute its second derivative. Differentiating once gives

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix} \tag{8}$$

Now treating $\frac{\partial f}{\partial x}$ as a vector valued function, we can compute the second derivative

$$\frac{\partial^2 f}{\partial x^2} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \tag{9}$$

The matrix $\frac{\partial^2 f}{\partial x^2}$ is symmetric since $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$ and is referred to as the *Hessian* of the function $f(x)$. Second derivatives are used to approximate perturbations of first derivatives

$$\Delta \frac{\partial f}{\partial x} \approx \Delta x^T \frac{\partial^2 f}{\partial x^2} = \begin{bmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{bmatrix}^T \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \tag{10}$$

For the quadratic function $f(x) = x^T Q x + b^T x$, we can use the perturbative perspective to compute

$$\Delta \frac{\partial f}{\partial x} = \Delta x^T \frac{\partial^2 f}{\partial x^2} = \Delta x^T (Q + Q^T) \Rightarrow \qquad \frac{\partial^2 f}{\partial x^2} = Q + Q^T \tag{11}$$