

# Discussion Materials

Oct. 11, 2019

## 1 Derivatives with vectors and matrices

All the vectors are assumed to be a column vector.

### 1.1 vector-by-scalar

$$\frac{\partial \mathbf{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} & \frac{\partial y_2}{\partial x} & \cdots & \frac{\partial y_n}{\partial x} \end{bmatrix}$$

$\mathbf{y}$  is a column vector,  $x$  is a scalar, the result is a row vector which has the same shape with  $\mathbf{y}^\top$ .

### 1.2 scalar-by-vector

$$\frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} & \cdots & \frac{\partial y}{\partial x_m} \end{bmatrix}^\top$$

$\mathbf{x}$  is a column vector,  $y$  is a scalar, the result is a column vector which has the same shape with  $\mathbf{x}$ .

### 1.3 vector-by-vector

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_n}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_m} & \frac{\partial y_2}{\partial x_m} & \cdots & \frac{\partial y_n}{\partial x_m} \end{bmatrix}$$

Both  $\mathbf{x}$  and  $\mathbf{y}$  are column vectors of length  $m$ ,  $n$  respectively. The result should be an  $m \times n$  matrix. (Lay out columns by  $\mathbf{x}$ , and rows by  $\mathbf{y}^\top$ .)

### 1.4 matrix-by-scalar

$$\frac{\partial \mathbf{Y}}{\partial x} = \begin{bmatrix} \frac{\partial y_{11}}{\partial x} & \frac{\partial y_{12}}{\partial x} & \cdots & \frac{\partial y_{1n}}{\partial x} \\ \frac{\partial y_{21}}{\partial x} & \frac{\partial y_{22}}{\partial x} & \cdots & \frac{\partial y_{2n}}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_{m1}}{\partial x} & \frac{\partial y_{m2}}{\partial x} & \cdots & \frac{\partial y_{mn}}{\partial x} \end{bmatrix}$$

Here  $\mathbf{Y}$  is a matrix and  $x$  is a scalar, and the result has the same shape of  $\mathbf{Y}$ . Note this is not consistent with 1.1, the result is no longer in the shape of  $\mathbf{Y}^\top$ .

## 1.5 scalar-by-matrix

$$\frac{\partial y}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{12}} & \cdots & \frac{\partial y}{\partial x_{1n}} \\ \frac{\partial y}{\partial x_{21}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{m1}} & \frac{\partial y}{\partial x_{m2}} & \cdots & \frac{\partial y}{\partial x_{mn}} \end{bmatrix}$$

Here  $\mathbf{X}$  is a matrix and  $y$  is a scalar, and the result has the same shape of  $\mathbf{X}$ . This is consistent with 1.2, the result is in the shape of  $\mathbf{X}$ .

## 1.6 Derivative of Vector Product

Suppose both  $\mathbf{u}$  and  $\mathbf{v}$  are column vectors, then the derivative of the dot product will be

$$\frac{\partial \mathbf{u}^\top \mathbf{v}}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}}{\partial \mathbf{x}}^\top \mathbf{v} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}} \mathbf{u}$$

## 1.7 Chain Rule

Suppose both  $\mathbf{f}$  and  $\mathbf{g}$  are vector functions (input and output are both vectors).  $\mathbf{x}$  is also a column vector.

$$\frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{x}))}{\partial \mathbf{x}} = \frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}} \frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{x}))}{\partial \mathbf{g}(\mathbf{x})}$$

Note that this is different from chain rule of scalar functions. The order should be reversed. It's very easy to make mistakes, so try not to use chain rule if not necessary.

# 2 Practice

For the first 3 cases, prove it using  $2 \times 2$  matrix  $\mathbf{A}$  or 2 dimensional vector  $\mathbf{y}$ .

1.  $\frac{\partial \mathbf{Ax}}{\partial \mathbf{x}} = \mathbf{A}^\top$
2.  $\frac{\partial \mathbf{y}^\top \mathbf{x}}{\partial \mathbf{x}} = \mathbf{y}$
3.  $\frac{\partial \mathbf{x}^\top \mathbf{A}}{\partial \mathbf{x}} = \mathbf{A}$
4.  $\frac{\partial \mathbf{x}^\top \mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{x}$
5.  $\frac{\partial \mathbf{x}^\top \mathbf{Ax}}{\partial \mathbf{x}} = \mathbf{Ax} + \mathbf{A}^\top \mathbf{x}$  (using 1.6, set  $\mathbf{u} = \mathbf{A}^\top \mathbf{x}$ ,  $\mathbf{v} = \mathbf{x}$ , and calculate  $\frac{\partial \mathbf{u}^\top \mathbf{v}}{\partial \mathbf{x}}$ )

# 3 Least Squares Problem

## 3.1 Formulation

For over-determined linear systems, the least squares problem is a optimization problem:

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_2^2$$

where  $\mathbf{A} \in R^{n \times d}$ ,  $d < n$  is a skinny matrix with rank  $d$ .  $\|\mathbf{x}\|_2$  is  $l_2$ -norm of vector  $\mathbf{x}$ , which is defined as  $\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$ .

## 3.2 Solution

$$\mathcal{L}(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|_2^2 = (\mathbf{Ax} - \mathbf{b})^\top (\mathbf{Ax} - \mathbf{b})$$

Set  $\frac{\partial \mathcal{L}(\mathbf{x})}{\partial \mathbf{x}} = 0$ , calculate the solution. ( $\mathbf{x} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}$ )

## 4 Minimum Norm Problem

### 4.1 Formulation

For under-determined linear systems, the minimum norm problem is:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_2^2$$

$$\text{s. t. } \mathbf{Ax} - \mathbf{b} = \mathbf{0}$$

where  $\mathbf{A} \in R^{n \times d}$ ,  $d > n$  is a fat matrix with rank  $n$ .

### 4.2 Solution

Using Lagrangian multiplier:

$$\mathcal{L}(\mathbf{x}, \lambda) = \|\mathbf{x}\|_2^2 + \lambda^\top (\mathbf{Ax} - \mathbf{b})$$

Set  $\frac{\partial \mathcal{L}(\mathbf{x}, \lambda)}{\partial \mathbf{x}} = 0$ ,  $\frac{\partial \mathcal{L}(\mathbf{x}, \lambda)}{\partial \lambda} = 0$ , calculate the solution. ( $\mathbf{x} = \mathbf{A}^\top (\mathbf{AA}^\top)^{-1} \mathbf{b}$ )