**Vector Differentiation Supplementary Notes**

This notes is a very small part of vector differentiation properties to get you started. You can prove each property easily by yourself by using $2 \times 2$ matrices and $2 \times 1$ vectors.

# 1 Matrix Transpose

Define square matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$,

1. $(\boldsymbol{A}^T)^T = \boldsymbol{A}$
2. $(\boldsymbol{A} + \boldsymbol{B})^T = \boldsymbol{A}^T + \boldsymbol{B}^T$
3. $(c\boldsymbol{A})^T = c\boldsymbol{A}^T$
4. $(\boldsymbol{A}\boldsymbol{B})^T = \boldsymbol{B}^T\boldsymbol{A}^T$
5. $(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C})^T = \boldsymbol{C}^T\boldsymbol{B}^T\boldsymbol{A}^T$
6. $\boldsymbol{A}^T = \boldsymbol{A}$ if $\boldsymbol{A}$ is symmetric

*Note: the above property works not only for square matrices, but any matrix which dimensions are compatible with one another, e.g. for $\boldsymbol{AB}$ where $\boldsymbol{A}$ is an $n \times d$ matrix and $\boldsymbol{B}$ is a $d \times k$ matrix.*

# 2 Basic Rules

1. **Chain Rule**. Given vector-valued functions $f, y$, taking in vector $\boldsymbol{x} \in \mathbb{R}^d$ as an input: $f(y(\boldsymbol{x}))$. Then,
$$\frac{dfy(\boldsymbol{x})}{d\boldsymbol{x}} = \frac{df}{dy} \cdot \frac{dy}{d\boldsymbol{x}}$$
2. **Total Derivatives**. Given vector-valued function $f, y$, taking in vector $\boldsymbol{x} \in \mathbb{R}^d$ as an input: $f(\boldsymbol{x}, y(\boldsymbol{x}))$. Then,
$$\frac{df(\boldsymbol{x}, y(\boldsymbol{x}))}{d\boldsymbol{x}} = \frac{df}{d\boldsymbol{x}} + \frac{df}{dy} \cdot \frac{dy}{d\boldsymbol{x}}$$

# 3 Vector Differentiation

Define a column vector $\boldsymbol{x} = [x_1, x_2, ..., x_d]^T, \boldsymbol{x} \in \mathbb{R}^2$, The derivative of a scalar output function $f(\boldsymbol{x})$ with respect to a **column** vector $\boldsymbol{x}$ can be defined as a **row** vector,

$$\nabla f(\boldsymbol{x}) = \frac{d}{d\boldsymbol{x}} f(\boldsymbol{x}) = \left[ \frac{\partial f}{\partial x_1} \; \frac{\partial f}{\partial x_2} \; ... \; \frac{\partial f}{\partial x_d} \right]$$

or a **column** vector (which is the transpose of the above), depending on whichever format is necessary. The following are some of the common properties,

1. Let a scalar $y$ be defined as an inner product of two vectors $\boldsymbol{b} \in \mathbb{R}^d$, $\boldsymbol{x} \in \mathbb{R}^d$, and $\boldsymbol{b}$ does not depend on $\boldsymbol{x}$,
$$y = \boldsymbol{b}^T\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{y},$$
then,
$$\frac{dy}{d\boldsymbol{x}} = \boldsymbol{b}^T.$$
*Note: it can be also $\boldsymbol{b}$, depending on which format is necessary.*

2. Let a scalar $y$ be defined as an inner product of the two same vectors $\boldsymbol{x} \in \mathbb{R}^d$,

$$y = \boldsymbol{x}^T \boldsymbol{x},$$

then,

$$\frac{dy}{d\boldsymbol{x}} = 2\boldsymbol{x}^T.$$

*Note: similarly, it can be also $2\boldsymbol{x}$, depending on which format is necessary*

3. Let a scalar $y$ be defined as a product of vector $\boldsymbol{b} \in \mathbb{R}^n$, $n \times d$ matrix $\boldsymbol{A}$, and vector $\boldsymbol{x} \in \mathbb{R}^d$, and $\boldsymbol{A}$ does not depend on $\boldsymbol{x}$ or $\boldsymbol{y}$,

$$y = \boldsymbol{b}^T \boldsymbol{A}\boldsymbol{x},$$

then,

$$\frac{dy}{d\boldsymbol{x}} = \boldsymbol{b}^T \boldsymbol{A}, \ \frac{dy}{d\boldsymbol{b}} = \boldsymbol{x}^T \boldsymbol{A}^T.$$

4. Let a scalar $y$ be defined as a product of a vector $\boldsymbol{x} \in \mathbb{R}^d$, $n \times d$ matrix $\boldsymbol{A}$, and another vector $\boldsymbol{x}$, where $\boldsymbol{A}$ does not depend on $\boldsymbol{x}$,

$$y = \boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x},$$

then,

$$\frac{dy}{d\boldsymbol{x}} = \boldsymbol{x}^T(\boldsymbol{A} + \boldsymbol{A}^T).$$

In the case where $\boldsymbol{A}$ is a symmetric matrix,

$$\frac{dy}{d\boldsymbol{x}} = 2\boldsymbol{x}^T \boldsymbol{A}$$

*Note: in the symmetric matrix case, the answer can also be $\frac{dy}{d\boldsymbol{x}} = 2\boldsymbol{A}\boldsymbol{x}$, depending on which format is necessary*

5. Let a vector $\boldsymbol{y} \in \mathbb{R}^m$ be defined as a product between a $n \times d$ matrix $\boldsymbol{A}$ and vector $\boldsymbol{x} \in \mathbb{R}^d$, and $\boldsymbol{A}$ does not depend on $\boldsymbol{x}$,

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x},$$

then,

$$\frac{d\boldsymbol{y}}{d\boldsymbol{x}} = \boldsymbol{A}.$$


Materials referenced from:

1. `https://ccrma.stanford.edu/~dattorro/matrixcalc.pdf`
2. `http://www.atmos.washington.edu/~dennis/MatrixCalculus.pdf`
3. `http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/3274/pdf/imm3274.pdf`