# Vector Calculus Examples

**Jingyang You**
**jingyang.you@anu.edu.au**

## 1 Important Facts

**# Properties**

$\mathbf{a}^T\mathbf{b} = (\mathbf{a}^T\mathbf{b})^T = \mathbf{b}^T\mathbf{a}$ (because $\mathbf{a}^T\mathbf{b}$ is a scalar)

$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$

$(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$

$(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$

$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

$(\mathbf{A}+\mathbf{B})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1}\mathbf{A}^{-1}$ (Note the difference, assuming $\mathbf{A}, \mathbf{B}$ are invertible)

$\mathbf{det}(\mathbf{AB}) = \mathbf{det}(\mathbf{A}) \cdot \mathbf{det}(\mathbf{B})$

$\mathbf{det}(c\mathbf{A}) = c^n \cdot \mathbf{det}(\mathbf{A})$

$\mathbf{det}(\mathbf{A}^T) = \mathbf{det}(\mathbf{A})$

$\mathbf{tr}(\mathbf{A}) = \mathbf{tr}(\mathbf{A}^T)$

$\mathbf{tr}(\mathbf{AB}) = \mathbf{tr}(\mathbf{BA})$ ($\mathbf{A}, \mathbf{B}$ needn't be square but have to be commutable)

$\mathbf{tr}(\mathbf{A}+\mathbf{B}) = \mathbf{tr}(\mathbf{A}) + \mathbf{tr}(\mathbf{B})$

$\mathbf{tr}(\mathbf{aa}^T) = \mathbf{a}^T\mathbf{a}$

**# Ideas**

In most cases, treat matrices as collections of vectors. For example:

$$\mathbf{A} = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_m^T \end{bmatrix} = \begin{bmatrix} a_1' & a_2' & \cdots & a_k' \end{bmatrix} \in \mathbb{R}^{m\times k}$$

where $a_i^T$ are $\mathbb{R}^{1\times k}$ row vectors, $a_j'$ are $\mathbb{R}^m$ column vectors. This comes in handy because

$$\mathbf{Ax} = \begin{bmatrix} a_1^T\mathbf{x} \\ a_2^T\mathbf{x} \\ \vdots \\ a_n^T\mathbf{x} \end{bmatrix}$$

As you may find, the matrix multiplication now can be written in a super compact form.

## 2 Vector Calculus

**Caution: In this course, we'll be using the numerator layout. In the following courses like SML and AML, we'll be using the denominator layout**. You'll see the difference in the next subsection.

## # Four types of problems

In this course, we only deal with **four** types of vector calculus. Each type will be accompanied by an example to demonstrate the correct way to do it.

---

### 1. scalar numerator $f : \mathbb{R}^k \to \mathbb{R}$, vector denominator $\mathbf{x} \in \mathbb{R}^k$

$$\frac{df(\mathbf{x})}{\mathbf{dx}} = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} & \frac{\partial f(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f(\mathbf{x})}{\partial x_k} \end{bmatrix} \in \mathbb{R}^{1\times k}$$

**Example:**

$$\frac{d\mathbf{x}^T\mathbf{x}}{\mathbf{dx}} = \begin{bmatrix} \frac{\partial \sum_{i=1}^k x_i^2}{\partial x_1} & \frac{\partial \sum_{i=1}^k x_i^2}{\partial x_2} & \cdots & \frac{\partial \sum_{i=1}^k x_i^2}{\partial x_k} \end{bmatrix} = \begin{bmatrix} 2x_1 & 2x_2 & \cdots & 2x_k \end{bmatrix} = 2\mathbf{x}^T$$

---

**ONE VERY IMPORTANT NOTE:** In other courses and many literature including matrix cookbook, we use the **column form** to represent the vector in the denominator (**denominator layout**).

$$\frac{df(\mathbf{x})}{\mathbf{dx}} = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_k} \end{bmatrix}$$

This will result in an additional transpose in the result when we are finding the derivative over a **vector**. No transpose will be added if the denominator is a **matrix**.

---

### 2. vector numerator $f : \mathbb{R} \to \mathbb{R}^k/\mathbb{R}^{1\times k}$, scalar denominator $x \in \mathbb{R}$

$$\frac{\mathbf{d}f(x)}{dx} = \begin{bmatrix} \frac{df_1(x)}{dx} \\ \frac{df_2(x)}{dx} \\ \vdots \\ \frac{df_k(x)}{dx} \end{bmatrix} \in \mathbb{R}^k$$

**Example:**

$$\frac{d\begin{bmatrix} \sin x & \cos x \end{bmatrix}}{dx} = \begin{bmatrix} \frac{d\sin x}{dx} \\ \frac{d\cos x}{dx} \end{bmatrix} = \begin{bmatrix} \cos x \\ -\sin x \end{bmatrix}$$

**Note that to use this rule, the vector in the numerator does not have to be in the row form. Both column and row form lead to the same result by definition.**

**Example:**

$$\frac{d\begin{bmatrix} \sin x \\ \cos x \end{bmatrix}}{dx} = \frac{d\begin{bmatrix} \sin x & \cos x \end{bmatrix}}{dx} = \begin{bmatrix} \cos x \\ -\sin x \end{bmatrix}$$

---

**3. vector numerator $f : \mathbb{R}^m \to \mathbb{R}^k/\mathbb{R}^{1\times k}$, vector denominator $\mathbf{x} \in \mathbb{R}^m$**

$$\frac{\mathbf{d}f(\mathbf{x})}{\mathbf{dx}} \underline{\underline{\text{Applying 2.}}} \begin{bmatrix} \frac{df_1(x)}{\mathbf{dx}} \\ \frac{df_2(x)}{\mathbf{dx}} \\ \vdots \\ \frac{df_k(x)}{\mathbf{dx}} \end{bmatrix}$$

$$\underline{\underline{\text{Applying 1.}}} \begin{bmatrix} \frac{df_1(x)}{dx_1} & \frac{df_1(x)}{dx_2} & \cdots & \frac{df_1(x)}{dx_m} \\ \frac{df_2(x)}{dx_1} & \frac{df_2(x)}{dx_2} & \cdots & \frac{df_2(x)}{dx_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{df_k(x)}{dx_1} & \frac{df_k(x)}{dx_2} & \cdots & \frac{df_k(x)}{dx_m} \end{bmatrix} \in \mathbb{R}^{k\times m}$$

**Important Example: Find the derivative of $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$ over $\mathbf{x} \in \mathbb{R}^n$.**
In general, we define four steps of finding the derivative:

**(1) Write the general elementwise derivative, so that there is no remaining vector form in the numerator containing the denominator**

**(2) Separate the numerator, so that each separated term is trivially related to the denominator**

**(3) Calculate the derivative**

**(4) Convert the elementwise derivative back to the vector form**

Now, take a look at this example. We follow the steps to find the derivative.

**Step (1): Write the general elementwise derivative, so that there is no remaining vector form in the numerator containing the denominator.**

The elementwise derivative would be $\frac{\partial \mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial x_t}$. To evaluate the derivative properly, we need to make sure that there is no remaining vector form in the numerator containing the denominator. The current numerator has vector $\mathbf{x}$, which involves $x_t$. So, we need to convert is to a summation of scalars. Different from the example in 1. where the formula $\mathbf{x}^T\mathbf{x}$ can be converted to a summation easily, $\mathbf{x}^T\mathbf{A}\mathbf{x}$ is not that easy. Here I would like to introduce a really useful trick to convert vector forms to summations. The matrix multiplication $\mathbf{A}\mathbf{x}$ can be expressed as

$$\mathbf{A}\mathbf{x} = \begin{bmatrix} a_1^T\mathbf{x} \\ a_2^T\mathbf{x} \\ \vdots \\ a_n^T\mathbf{x} \end{bmatrix}$$

Thus, the term $\mathbf{x}^T\mathbf{A}\mathbf{x}$ is equivalent to

$$\mathbf{x}^T \begin{bmatrix} a_1^T\mathbf{x} \\ a_2^T\mathbf{x} \\ \vdots \\ a_n^T\mathbf{x} \end{bmatrix} = \sum_{i=1}^{n} x_i a_i^T \mathbf{x} \overset{\text{Expand}}{=\!=\!=} \sum_{j=1}^{n}\sum_{i=1}^{n} x_i A_{ij} x_j$$

We can also do this by expanding $\mathbf{x}^T\mathbf{A}$ first. Analogously,

$$\mathbf{x}^T\mathbf{A} = \begin{bmatrix} \mathbf{x}^T a_1' & \mathbf{x}^T a_2' & \cdots & \mathbf{x}^T a_n' \end{bmatrix}$$

Therefore,

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \begin{bmatrix} \mathbf{x}^T a_1' & \mathbf{x}^T a_2' & \cdots & \mathbf{x}^T a_n' \end{bmatrix} \mathbf{x} = \sum_{i=1}^{n} \mathbf{x}^T a_i' x_i = \sum_{i=1}^{n}\sum_{j=1}^{n} x_j A_{ji} x_i$$

which leads to the same result. This is because

$$\sum_{i=1}^{n}\sum_{j=1}^{n} x_j A_{ji} x_i \overset{\text{Swap}}{=\!=\!=} \sum_{j=1}^{n}\sum_{i=1}^{n} x_i A_{ij} x_j$$

**By now we can ensure there is no remaining vector form containing the denominator.**

We proceed to step (2) with the elementwise derivative $\frac{\partial \sum_{j=1}^{n}\sum_{i=1}^{n} x_i x_j A_{ij}}{\partial x_t}$.

**Step (2): Separate the numerator, so that each separated term is trivially related to the denominator**

The numerator,

$$\sum_{j=1}^{n}\sum_{i=1}^{n} x_i x_j A_{ij}$$

is by no means trivially related to the denominator $x_t$. Both $x_i, x_j$ can be or not be $x_t$. So, we need to break down the numerator into several cases, and in each case the relation becomes trivial. Here, $x_i x_j$ can be classified into four cases: (1) $i = t, j \neq t$, (2) $j = t, i \neq t$, (3) $i = t, j = t$ and (4) $i \neq t, j \neq t$. Case 4 can be safely ignored, as none of them is $x_t$, resulting in a derivative of 0. Thus the partial derivative can be written as

$$\frac{\partial \sum_{j=1}^{n}\sum_{i=1}^{n} x_i x_j A_{ij}}{\partial x_t} = \overbrace{\frac{\partial \sum_{j\neq t}^{n} x_t x_j A_{tj}}{\partial x_t}}^{\textbf{Case 1}} + \overbrace{\frac{\partial \sum_{i\neq t}^{n} x_i x_t A_{it}}{\partial x_t}}^{\textbf{Case 2}} + \overbrace{\frac{\partial x_t^2 A_{tt}}{\partial x_t}}^{\textbf{Case 3}}$$

We have now explicitly separated the numerator. This can indeed help as in case one, for example, $x_j$ can never be $x_t$. This ensures that $x_t x_j$ is always a linear function of $x_t$, and its derivative over $x_t$ is always $x_j$.

**Step (3): Calculate the derivative**

As you may find, this step is the easiest.

$$\begin{aligned}\overbrace{\frac{\partial \sum_{j\neq t}^{n} x_t x_j A_{tj}}{\partial x_t}}^{\textbf{Case 1}} + \overbrace{\frac{\partial \sum_{i\neq t}^{n} x_i x_t A_{it}}{\partial x_t}}^{\textbf{Case 2}} + \overbrace{\frac{\partial x_t^2 A_{tt}}{\partial x_t}}^{\textbf{Case 3}} &= \sum_{j\neq t}^{n} x_j A_{tj} + \sum_{i\neq t}^{n} x_i A_{it} + 2x_t A_{tt} \\ &= \left(\sum_{j\neq t}^{n} x_j A_{tj} + x_t A_{tt}\right) + \left(\sum_{i\neq t}^{n} x_i A_{it} + x_t A_{tt}\right) \\ &= \sum_{j=1}^{n} x_j A_{tj} + \sum_{i=1}^{n} x_i A_{it}\end{aligned}$$

**Step (4): Convert back to vector form**

We now have the elementwise derivative

$$\frac{\partial \mathbf{x}^T \mathbf{A}\mathbf{x}}{\partial x_t} = \sum_{j=1}^{n} x_j A_{tj} + \sum_{i=1}^{n} x_i A_{it}$$

We can easily observe that these terms are actually the scalar form of the dot product. We convert them to vector form for a more compact result

$$\frac{\partial \mathbf{x}^T \mathbf{A}\mathbf{x}}{\partial x_t} = a_t^T \mathbf{x} + a_t'^T \mathbf{x}$$

where $a_t^T, a_t'$ are the $t$-th row and column of $\mathbf{A}$. Till now we have written out the vector form for one single entry of $\mathbf{x}$, $x_t$. If we just repeat the above steps for every entry (not actually repeating, just changing the subscript from $t$ to other indices from 1 to $n$), we will get

$$\frac{d\mathbf{x}^T \mathbf{A}\mathbf{x}}{\mathbf{dx}} = \begin{bmatrix} a_1^T \mathbf{x} & a_2^T \mathbf{x} & \cdots & a_n^T \mathbf{x} \end{bmatrix} + \begin{bmatrix} a_1'^T \mathbf{x} & a_2'^T \mathbf{x} & \cdots & a_n'^T \mathbf{x} \end{bmatrix}$$

Anyone remember this from above?

$$\mathbf{A}\mathbf{x} = \begin{bmatrix} a_1^T \mathbf{x} \\ a_2^T \mathbf{x} \\ \vdots \\ a_n^T \mathbf{x} \end{bmatrix}$$

Apply this we get

$$\frac{d\mathbf{x}^T \mathbf{A}\mathbf{x}}{\mathbf{dx}} = (\mathbf{A}\mathbf{x})^T + \left(\mathbf{A}^T \mathbf{x}\right)^T = \mathbf{x}^T (\mathbf{A} + \mathbf{A}^T)$$

So that is our beatiful first order derivative.

**4. scalar numerator $f : \mathbb{R}^{m\times k} \to \mathbb{R}$, matrix denominator $\mathbf{X} \in \mathbb{R}^{m\times k}$**

$$\frac{df(\mathbf{X})}{\mathbf{dX}} = \begin{bmatrix} \frac{df(\mathbf{X})}{dX_{11}} & \frac{df(\mathbf{X})}{dX_{12}} & \cdots & \frac{df(\mathbf{X})}{dX_{1k}} \\ \frac{df(\mathbf{X})}{dX_{21}} & \frac{df(\mathbf{X})}{dX_{22}} & \cdots & \frac{df(\mathbf{X})}{dX_{2k}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{df(\mathbf{X})}{dX_{m1}} & \frac{df(\mathbf{X})}{dX_{m2}} & \cdots & \frac{df(\mathbf{X})}{dX_{mk}} \end{bmatrix}$$

**Example:**

Find the derivative of $\mathbf{tr}(\mathbf{X})$ over $\mathbf{X} \in \mathbb{R}^{n\times n}$.

$$\frac{d\ \mathbf{tr}(\mathbf{X})}{\mathbf{dX}} = \begin{bmatrix} \frac{\partial\ \mathbf{tr}(\mathbf{X})}{\partial X_{11}} & \frac{\partial\ \mathbf{tr}(\mathbf{X})}{\partial X_{12}} & \cdots & \frac{\partial\ \mathbf{tr}(\mathbf{X})}{\partial X_{1n}} \\ \frac{\partial\ \mathbf{tr}(\mathbf{X})}{\partial X_{21}} & \frac{\partial\ \mathbf{tr}(\mathbf{X})}{\partial X_{22}} & \cdots & \frac{\partial\ \mathbf{tr}(\mathbf{X})}{\partial X_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial\ \mathbf{tr}(\mathbf{X})}{\partial X_{n1}} & \frac{\partial\ \mathbf{tr}(\mathbf{X})}{\partial X_{n2}} & \cdots & \frac{\partial\ \mathbf{tr}(\mathbf{X})}{\partial X_{nn}} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial \sum_{k=1}^{n} X_{kk}}{\partial X_{11}} & \frac{\partial \sum_{k=1}^{n} X_{kk}}{\partial X_{12}} & \cdots & \frac{\partial \sum_{k=1}^{n} X_{kk}}{\partial X_{1n}} \\ \frac{\partial \sum_{k=1}^{n} X_{kk}}{\partial X_{21}} & \frac{\partial \sum_{k=1}^{n} X_{kk}}{\partial X_{22}} & \cdots & \frac{\partial \sum_{k=1}^{n} X_{kk}}{\partial X_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \sum_{k=1}^{n} X_{kk}}{\partial X_{n1}} & \frac{\partial \sum_{k=1}^{n} X_{kk}}{\partial X_{n2}} & \cdots & \frac{\partial \sum_{i=1}^{n} X_{kk}}{\partial X_{nn}} \end{bmatrix}$$

**Again, we consider a general case in row $i$, column $j$:**

$$\frac{\partial \sum_{k=1}^{n} X_{kk}}{\partial X_{ij}}$$

Obviously, the numerator only contains the diagonal elements. This means, if $i \neq j$, the derivative is 0, since there is no relevant part in the numerator. Otherwise, if $i = j$, the parital derivative is now:

$$\frac{\partial \sum_{k=1}^{n} X_{kk}}{\partial X_{ij}} = \frac{dX_{ij}}{dX_{ij}} = \frac{dX_{ii}}{dX_{ii}} = \frac{dX_{jj}}{dX_{jj}} = 1$$

This implies every diagonal element in the derivative matrix is 1, and other elements are all 0. This means

$$\frac{d\ \mathbf{tr}(\mathbf{X})}{\mathbf{dX}} = \mathbf{I}$$