Matrix Calculus

Sihan Wei

June 2025

Overview

1. This is not a formal course on matrix calculus, but we'll cover everything you need for this class.
2. You should be comfortable with single-variable calculus and basic linear algebra.
3. Recommended resources:
    - The Matrix Cookbook
    - Wikipedia: Matrix Calculus

Notation

In this note:

- We use **bold lowercase** letters (e.g., $\mathbf{x}, \mathbf{y}, \boldsymbol{\delta}$) to represent **vectors**.
- We use **bold uppercase** letters (e.g., $\mathbf{A}, \mathbf{X}, \boldsymbol{\Delta}$) to represent **matrices**.
- Scalars are written in regular font (e.g., x, y, λ).
- $\nabla f(\mathbf{x})$ denotes the **gradient** of a scalar function w.r.t. vector $\mathbf{x}$.
- $\nabla^2 f(\mathbf{x})$ denotes the **Hessian matrix**.
- $\mathrm{tr}(\cdot)$ denotes the **trace** of a matrix.

Types of Derivatives

We consider six types of derivatives involving scalars, vectors, and matrices:

Output \Input	**Scalar**	**Vector**	**Matrix**
Scalar	$\frac{\partial y}{\partial x}$	$\frac{\partial \mathbf{y}}{\partial x}$	$\frac{\partial \mathbf{Y}}{\partial x}$
Vector	$\frac{\partial y}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$	
Matrix	$\frac{\partial y}{\partial \mathbf{X}}$		

- Scalar with respect to scalar/vector/matrix
- Vector with respect to scalar/vector
- Matrix with respect to scalar

Derivative of a Scalar Function

1. With respect to a scalar (dy/dx): standard single-variable calculus.
2. With respect to a vector ($dy/d\mathbf{x}$): For example, $y = \|\mathbf{x}\|$. If $\mathbf{x} \in \mathbb{R}^n$, then

$$\frac{dy}{d\mathbf{x}} = \begin{bmatrix} \frac{dy}{dx_1} \\ \vdots \\ \frac{dy}{dx_n} \end{bmatrix} \in \mathbb{R}^{n \times 1}.$$

3. With respect to a matrix ($dy/d\mathbf{X}$): For example, $y = \|\mathbf{X}\|_F$, the Frobenius norm. If $\mathbf{X} \in \mathbb{R}^{m \times n}$, then

$$\frac{dy}{d\mathbf{X}} = \begin{bmatrix} \frac{dy}{dX_{11}} & \cdots & \frac{dy}{dX_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{dy}{dX_{m1}} & \cdots & \frac{dy}{dX_{mn}} \end{bmatrix} \in \mathbb{R}^{m \times n}.$$

Derivative of a Vector Function

1. With respect to a scalar ($d\mathbf{y}/dx$): For example, $\mathbf{y} = x\mathbf{v}$ for scalar x and constant vector $\mathbf{v} \in \mathbb{R}^n$. Then

$$\frac{d\mathbf{y}}{dx} = \begin{bmatrix} \frac{dy_1}{dx} & \cdots & \frac{dy_n}{dx} \end{bmatrix} \in \mathbb{R}^{1 \times n}.$$

2. With respect to a vector ($d\mathbf{y}/d\mathbf{x}$): For example, $\mathbf{y} = \mathbf{A}\mathbf{x}$ with constant $\mathbf{A}$. $\mathbf{y} \in \mathbb{R}^n$ and $\mathbf{x} \in \mathbb{R}^m$. Then

$$\begin{aligned} \frac{d\mathbf{y}}{d\mathbf{x}} &= \begin{bmatrix} \nabla y_1(\mathbf{x}) & \nabla y_2(\mathbf{x}) & \cdots & \nabla y_n(\mathbf{x}) \end{bmatrix} \\ &= \begin{bmatrix} \frac{dy_1}{dx_1} & \cdots & \frac{dy_n}{dx_1} \\ \vdots & \ddots & \vdots \\ \frac{dy_1}{dx_m} & \cdots & \frac{dy_n}{dx_m} \end{bmatrix} \in \mathbb{R}^{m \times n}. \end{aligned}$$

Practice

1. Compute $\frac{d}{d\mathbf{x}}\|\mathbf{x}\|_2^2$
2. Compute $\frac{d}{d\mathbf{x}}\mathbf{A}\mathbf{x}$

Practice: Solutions

1. $\frac{d}{d\mathbf{x}}\|\mathbf{x}\|_2^2 = \frac{d}{d\mathbf{x}}(\mathbf{x}^\top \mathbf{x}) = 2\mathbf{x}$
2. $\frac{d}{d\mathbf{x}}(\mathbf{A}\mathbf{x}) = \mathbf{A}$

Chain Rule (Multivariate)

- For scalar functions: If $h(x) = f(g(x))$, then
$$\frac{dh}{dx} = \frac{df}{dg} \cdot \frac{dg}{dx}$$
- For multivariate case: If $h(x) = f(g_1(x), g_2(x))$, then:
$$\frac{dh}{dx} = \frac{\partial f}{\partial g_1} \cdot \frac{dg_1}{dx} + \frac{\partial f}{\partial g_2} \cdot \frac{dg_2}{dx}$$
- **Matrix multiplication is not commutative**: That is, in general, $AB \neq BA$. $\Rightarrow$ **Order matters** when applying the chain rule. Swapping terms can lead to incorrect results or a dimension mismatch.
- **Shape consistency is your best sanity check**: Always check the dimensions of each part in the chain rule. If the final shape doesn't match the derivative you're taking, you've likely reversed or misapplied the chain rule.
- Tip: Write down shapes explicitly when unsure!

Chain Rule: Practice

- Compute $\nabla \|\mathbf{x}\|$
- Compute $\nabla \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2$

Chain Rule: Solutions

- $\frac{d}{d\mathbf{x}}\|\mathbf{x}\|_2 = \frac{\mathbf{x}}{\|\mathbf{x}\|_2}$ (undefined at $\mathbf{x} = 0$)
- $\frac{d}{d\mathbf{w}}\|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 = 2\mathbf{X}^\top(\mathbf{X}\mathbf{w} - \mathbf{y})$

Taylor Expansion

- In 1D:

$$f(x+\delta) = f(x) + f'(x)\delta + \frac{1}{2}f''(x)\delta^2 + o(|\delta|^2)$$

- In $\mathbb{R}^n$:

$$f(\mathbf{x}+\boldsymbol{\delta}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^\top \boldsymbol{\delta} + \frac{1}{2}\boldsymbol{\delta}^\top \nabla^2 f(\mathbf{x})\boldsymbol{\delta} + o(\|\boldsymbol{\delta}\|_2^2)$$

- For matrix-valued input:

$$f(\mathbf{X}+\boldsymbol{\Delta}) = f(\mathbf{X}) + \operatorname{tr}\left(\nabla f(\mathbf{X})^\top \boldsymbol{\Delta}\right) + \frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Delta}^\top \nabla^2 f(\mathbf{X})\boldsymbol{\Delta}\right) + o(\|\boldsymbol{\Delta}\|_F^2)$$

Practice: Quadratic Forms

Consider the quadratic form $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x}$.

1. Use Taylor expansion to compute $\frac{d}{d\mathbf{x}}(\mathbf{x}^\top \mathbf{A}\mathbf{x})$, assuming $\mathbf{A}$ is constant.
2. What happens when $\mathbf{A}$ is symmetric, i.e., $\mathbf{A}^\top = \mathbf{A}$?

Quadratic Form: Solutions

Let $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x}$

1. Using the Taylor expansion or product rule:

$$\frac{d}{d\mathbf{x}}(\mathbf{x}^\top \mathbf{A}\mathbf{x}) = \mathbf{A}^\top \mathbf{x} + \mathbf{A}\mathbf{x}$$

2. If $\mathbf{A}$ is symmetric, then:

$$\frac{d}{d\mathbf{x}}(\mathbf{x}^\top \mathbf{A}\mathbf{x}) = 2\mathbf{A}\mathbf{x}$$