§1 回顾变量的微分学

在学习向量和矩阵的求导之前, 最好回顾关于一元函数、多元函数的微分学.

一元函数的微分学 微分算子 $\mathscr{D}$ 是一个线性算子. 这就意味着对于映射 f, g、常数 k, l, 有 $\mathscr{D}(kf+lg) = k\mathscr{D}(f)+l\mathscr{D}(g)$. 此外, 我们还有链式法则, 即 $f(g(x)) = f'_{g(x)}g'_x$.

多元函数的微分学 最重要的是考虑多元函数的复合函数.

§2 向量求导: 两种格式

一句话概括: 向量求导无非就是将多个变量排成了一个比较好的格式. 并且在这个比较好的格式上面求微分. 再继续之前, 我们先来回顾几个重要的技巧:

1. **作为点积的求和**. 对于形如 $\sum_i u_i v_i$ 的式子, 可以使用两个列向量 $\mathbf{u}, \mathbf{v}$ 并且写作点积的形式: 即 $\sum u_i v_i = \mathbf{u}\cdot\mathbf{v} = \mathbf{u}'\mathbf{v}$.
2. **作为矩阵乘法的点积和**. 对于一个矩阵 $\mathbf{X}$, 其中如果第 i 行第 j 列的元素是 $\mathbf{u}_i\cdot\mathbf{v}_j$, 可以定义两个矩阵 $\mathbf{U} = \begin{bmatrix} | & | & & | & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_{n-1} & \mathbf{u}_n \\ | & | & & | & | \end{bmatrix}$、$\mathbf{V} = \begin{bmatrix} | & | & & | & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_{n-1} & \mathbf{v}_n \\ | & | & & | & | \end{bmatrix}$, 使得 $\mathbf{X} = \mathbf{U}'\mathbf{V}$.

2.1. 基本定义.

定义 1. (向量对标量求导) 向量 $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_m \end{bmatrix}'$ 对于标量 x 的求导记作

$$\boxed{\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}x} := \begin{bmatrix} \frac{\mathrm{d}y_1}{\mathrm{d}x} & \frac{\mathrm{d}y_2}{\mathrm{d}x} & \frac{\mathrm{d}y_3}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}y_m}{\mathrm{d}x} \end{bmatrix}'}.$$

注. (1) 这通常称为向量 $\mathbf{y}$ 的切向量. 实际上这里映射 $\mathbf{y}:\mathbb{R}^1\to\mathbb{R}^m$.

(2) 这里我们采用竖着排版的形式. 这样的情形只是一个习惯性操作. 倘若定义 $\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}x} := \begin{bmatrix} \frac{\mathrm{d}y_1}{\mathrm{d}x} & \frac{\mathrm{d}y_2}{\mathrm{d}x} & \frac{\mathrm{d}y_3}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}y_m}{\mathrm{d}x} \end{bmatrix}$, 那么后续的内容除了符号上有更改以外, 没有太多区别.

定义 2. (标量对向量求导) 标量 y 对于向量 $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_m \end{bmatrix}'$ 的**分子形式**记作

$$\boxed{\frac{\partial y}{\partial \mathbf{x}} := \begin{bmatrix} \frac{\mathrm{d}y}{\mathrm{d}x_1} & \frac{\mathrm{d}y}{\mathrm{d}x_2} & \frac{\mathrm{d}y}{\mathrm{d}x_3} & \cdots & \frac{\mathrm{d}y}{\mathrm{d}x_m} \end{bmatrix}}.$$

其**分母形式**可以类似记作

$$\frac{\partial y}{\partial \mathbf{x}} := \begin{bmatrix} \frac{\mathrm{d}y}{\mathrm{d}x_1} & \frac{\mathrm{d}y}{\mathrm{d}x_2} & \frac{\mathrm{d}y}{\mathrm{d}x_3} & \cdots & \frac{\mathrm{d}y}{\mathrm{d}x_m} \end{bmatrix}'.$$

这里**分子形式**的意思是: 生成的向量的**行数**和**分子**的行数相同.

注. (1) 分子形式和分母形式实际上是求导之后对于变量如何排布进行的约定. 如果不加说明, 下列总是使用分子形式.

(2) 实际上这是映射 $\mathbf{y} : \mathbb{R}^n \to \mathbb{R}$. 并且与梯度之间的关系是 $\nabla y = \left(\frac{\partial y}{\partial \mathbf{x}}\right)'$.

定义 3. (向量对向量求导) 向量 $\mathbf{y} = [y_1, y_2, \cdots, y_m]'$ 对向量 $\mathbf{x} = [x_1, x_2, \cdots, x_n]'$ 求导的**分子形式**写作

$$\begin{array}{cc} \boxed{\frac{\partial \mathbf{y}}{\partial \mathbf{x}} := \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}} & m\text{行} \\ n\text{列} & \end{array}$$

其**分母形式**就是分子形式的转置.

这里**分子形式**的意思是: 生成的矩阵的**行数**和**分子**的行数相同.

注. (1) 实际上映射 $\mathbf{y} : \mathbb{R}^m \to \mathbb{R}^n$, 这就是 Jacobian 矩阵.

定义 4. (矩阵对标量求导) 矩阵 $\mathbf{Y} \in \mathbb{R}^{m \times n}$ 对于标量 x 的求导的分子形式记作

$$\boxed{\frac{\partial \mathbf{Y}}{\partial x} = \left(\frac{\partial y_{i,j}}{\partial x}\right)_{i,j}}.$$

分母形式就是分子形式的转置.

这里**分子形式**的意思是: 生成的矩阵的**行数**和**分子**的行数相同.

注. 这时候 $\mathbf{Y} : \mathbb{R} \to \mathbb{R}^{m \times n}$.

定义 5. (标量对矩阵求导) 标量 y 对于矩阵 $\mathbf{X} \in \mathbb{R}^{p \times q}$ 求导的分子形式记作

$$\boxed{\frac{\partial y}{\partial \mathbf{x}} = \left(\frac{\partial y}{\partial x_{j,i}}\right)_{i,j}}.$$

分母形式就是分子形式的转置.

这里**分子形式**的意思是: 生成的矩阵的**行数**和**分母的列数**相同. (因为分子就是一个标量).

注. 实际上这时候映射 $y : \mathbb{R}^{p \times q} \to \mathbb{R}$. 常见的函数为矩阵的迹、矩阵的行列式.

§3 引例: 使用迹 (trace), 化为标量形式

在线性代数中学习过: 某个矩阵 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 的迹 $\mathrm{tr}\mathbf{A} := \sum_{i=1}^{n} A_{ii}$. 现在考虑某个 $f = \mathrm{tr}(g(\mathbf{X}))$, 然后想要求 $\frac{\partial f}{\partial \mathbf{X}}$.

我们把矩阵 $\mathbf{A}$ 的第 i 行第 j 列赋予一个新的记号: $[\mathbf{A}]_{ij}$. 比如我们观察到:

- $[\mathbf{AB}]_{ik} = \sum_j A_{ij} B_{jk}$.
- $[\mathbf{ABC}']_{ij} = \sum_k A_{ik} [\mathbf{BC}']_{kj} = \sum_k A_{ik} \sum_l B_{kl} C_{jl} = \sum_k \sum_l A_{ik} B_{kl} C_{jl}$.

这就使得我们可以把它展开为完全的标量形式了, 进而应用求导公式.

例 1. 求 $f(\mathbf{X}) = \mathrm{tr}(\mathbf{AXB})$, 其中 $\mathbf{A}, \mathbf{B}$ 是常矩阵.

解答. 我们知道

$$\begin{aligned} f &= \mathrm{tr}(\mathbf{AXB}) \\ &= \sum_i [\mathbf{AXB}]_{ii} = \sum_i \sum_j A_{ij} [\mathbf{XB}]_{ji} \\ &= \sum_i \sum_j A_{ij} \sum_k X_{jk} B_{ki} = \sum_i \sum_j \sum_k A_{ij} X_{jk} B_{ki} \end{aligned}$$

固定 j, k 两个下标, 将此式子对于 X_{jk} 求导, 就得到了

$$\begin{aligned} \frac{\partial f}{\partial X_{jk}} &= \sum_i A_{ij} B_{ki} && (\text{下标}k\text{这时候已经固定}) \\ &= [\mathbf{BA}]_{kj} \end{aligned}$$

因为结果肯定与 $\mathbf{X}$ 相同, 那么行列就是 j, k. 因此我们的矩阵求导的形式为 $\frac{\partial \mathrm{tr}(\mathbf{AXB})}{\partial \mathbf{X}} = \mathbf{A}'\mathbf{B}'$.

类似地, 令 $f(\mathbf{X}) = \mathrm{tr}(\mathbf{AX}'\mathbf{B})$,

$$\begin{aligned} f &= \sum_i [\mathbf{AX}'\mathbf{B}]_{ii} = \sum_i \sum_j A_{ij} [\mathbf{X}'\mathbf{B}]_{ji} \\ &= \sum_i \sum_j A_{ij} \sum_k X'_{jk} B_{ki} = \sum_i \sum_j A_{ij} \sum_k X_{kj} B_{ki} \\ &= \sum_i \sum_j \sum_k A_{ij} X_{kj} B_{ki}. \end{aligned}$$

固定 k, j, 对 X_{kj} 求导有: $\frac{\partial f}{\partial X_{jk}} = \sum_i A_{ij} B_{ki} = [\mathbf{BA}]_{kj}$, 由于求导的结果一定与 $\mathbf{X}$ 相同, 那么就是 $\boxed{\dfrac{\partial \mathrm{tr}(\mathbf{AX}'\mathbf{B})}{\partial \mathbf{X}} = \mathbf{BA}.}$

问题在于, 若有多个同时出现的指标该怎么办? 来看下面的内容.

例 2. 考虑 $f = \mathrm{tr}(\mathbf{AXBXC}')$, 求 $\partial f / \partial \mathbf{X}$.$\mathbf{A}, \mathbf{B}, \mathbf{C}$ 为常数矩阵.

解答. 考虑

$$f = \sum_i$$

§4 使用微分形式计算

类似于全微分一样, 矩阵的计算可以用微分形式进行求解. 下面列举常用的微分形式的分子结果.

定理 1. 第一组恒等式:

$$
\begin{aligned}
dy &= \frac{dy}{dx}dx \\
dy &= \frac{dy}{d\mathbf{x}}d\mathbf{x} \\
dy &= \operatorname{tr}(\frac{dy}{d\mathbf{X}}d\mathbf{X}) \\
d\mathbf{y} &= \frac{d\mathbf{y}}{dx}dx \\
d\mathbf{y} &= \frac{d\mathbf{y}}{d\mathbf{x}}d\mathbf{x} \\
d\mathbf{Y} &= \frac{d\mathbf{Y}}{dx}dx
\end{aligned}
$$

定理 2. 第二组恒等式:

$$
\begin{aligned}
d\mathbf{A} &= \mathbf{0} \\
d(a\mathbf{X}) &= a\ d\mathbf{X} \\
d(\mathbf{X}+\mathbf{Y}) &= d\mathbf{X}+d\mathbf{Y} \\
d(\mathbf{X}\mathbf{Y}) &= (d\mathbf{X})\mathbf{Y}+\mathbf{X}(d\mathbf{Y}) \\
d(\mathbf{X_1}\mathbf{X_2}\cdots\mathbf{X_n}) &= (d\mathbf{X_1})\mathbf{X_2}\cdots\mathbf{X_n}+\mathbf{X_1}(d\mathbf{X_2})\cdots\mathbf{X_n}+\cdots+\mathbf{X_1}\mathbf{X_2}\cdots(d\mathbf{X_n}) \\
d(\mathbf{A}\mathbf{X}\mathbf{B}+\mathbf{C}) &= \mathbf{A}(d\mathbf{X})\mathbf{B} \\
d(\mathbf{X}\otimes\mathbf{Y}) &= (d\mathbf{X})\otimes\mathbf{Y}+\mathbf{X}\otimes(d\mathbf{Y}) \\
d(\mathbf{X}\circ\mathbf{Y}) &= (d\mathbf{X})\circ\mathbf{Y}+\mathbf{X}\circ(d\mathbf{Y}) \\
d(\mathbf{X}^\top) &= (d\mathbf{X})^\top \\
d(\mathbf{X}^{-1}) &= -\mathbf{X}^{-1}(d\mathbf{X})\mathbf{X}^{-1} \\
d(\operatorname{tr}(\mathbf{X})) &= \operatorname{tr}(d\mathbf{X}) \\
d(|\mathbf{X}|) &= \operatorname{tr}(adj(\mathbf{X})d\mathbf{X}) = |\mathbf{X}|\operatorname{tr}(\mathbf{X}^{-1}d\mathbf{X}) \\
d(\sigma(\mathbf{X})) &= \sigma'(\mathbf{X})\circ d\mathbf{X} \quad \text{其中 } \sigma \text{ 是逐元素相关的操作}
\end{aligned}
.
$$

定理 3. 第三组恒等式:

$$
\begin{aligned}
d(z(\mathbf{y}(\mathbf{x}))) &= \frac{dz}{d\mathbf{y}}\frac{d\mathbf{y}}{d\mathbf{x}}d\mathbf{x} \\
d(z(y(\mathbf{X}))) &= \operatorname{tr}(\frac{dz}{dy}\frac{dy}{d\mathbf{X}}d\mathbf{X}) \\
d(z(\mathbf{Y}(x))) &= \operatorname{tr}(\frac{dz}{d\mathbf{Y}}\frac{d\mathbf{Y}}{dx}dx) \\
d(\mathbf{z}(\mathbf{y}(\mathbf{x}))) &= \frac{d\mathbf{z}}{d\mathbf{y}}\frac{d\mathbf{y}}{d\mathbf{x}}d\mathbf{x}
\end{aligned}
.
$$

定理 4. 第四组恒等式:

$$
\begin{aligned}
df(x,y,z) &= \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz \\
df(x,\mathbf{y},\mathbf{Z}) &= \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial \mathbf{y}}d\mathbf{y} + \mathrm{tr}(\frac{\partial f}{\partial \mathbf{Z}}d\mathbf{Z}) \\
d\mathbf{f}(x,\mathbf{y}) &= \frac{\partial \mathbf{f}}{\partial x}dx + \frac{\partial \mathbf{f}}{\partial \mathbf{y}}d\mathbf{y} \\
d\mathbf{F}(x,y,z) &= \frac{\partial \mathbf{F}}{\partial x}dx + \frac{\partial \mathbf{F}}{\partial y}dy + \frac{\partial \mathbf{F}}{\partial z}dz
\end{aligned}.
$$

注. Kronecker 乘法是 $\mathbf{A}\otimes\mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix}$.

§5 常见求导结果

5.1. 向量对向量.

1. 常数相关
   (a) 常向量的导数是 0: 若 $\mathbf{a}$ 中不含变量 $\mathbf{x}$, 那么 $\frac{\partial \mathbf{a}}{\partial \mathbf{x}} = \mathbf{0}$.
   (b) 常数进出求导: 若 a 中不含变量 $\mathbf{x}$, $\mathbf{u}=\mathbf{u}(\mathbf{x})$ 那么 $\frac{\partial a\mathbf{u}}{\partial \mathbf{x}} = a\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$.
   (c) 常矩阵进出求导:
      i. 若 $\mathbf{A}\in\mathbb{R}^{n\times m}$ 中不含变量 $\mathbf{x}\in\mathbb{R}^{m\times 1}$, 那么 $\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}$.
      ii. 若 $\mathbf{A}\in\mathbb{R}^{n\times m}$ 中不含变量 $\mathbf{x}\in\mathbb{R}^{m\times 1}$, 那么 $\frac{\partial \mathbf{x}'\mathbf{A}}{\partial \mathbf{x}} = \mathbf{A}'$.
      iii. 若 $\mathbf{A}\in\mathbb{R}^{n\times m}$ 中不含变量 $\mathbf{x}\in\mathbb{R}^{m\times 1}$, 而向量值函数 $\mathbf{u}=\mathbf{u}(\mathbf{x})\in\mathbb{R}^{m\times 1}$, 那么 $\frac{\partial \mathbf{A}\mathbf{u}}{\partial \mathbf{x}} = \mathbf{A}$.
2. 乘法
   (a) 变元在向量: a 中不含 $\mathbf{x}$, 但是 $\mathbf{u}=\mathbf{u}(\mathbf{x})$: $\frac{\partial a\mathbf{u}}{\partial \mathbf{x}} = a\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$.
   (b) 变元在常数: $v=v(\mathbf{x}), \mathbf{a}$ 中不含 $\mathbf{x}$:$\frac{\partial v\mathbf{a}}{\partial \mathbf{x}} = \mathbf{a}\frac{\partial v}{\partial \mathbf{x}}$.
   (c) 两个都有: $v=v(\mathbf{x}), \mathbf{u}=\mathbf{u}(\mathbf{x})$: $\frac{\partial v\mathbf{u}}{\partial \mathbf{x}} = v\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \mathbf{u}\frac{\partial v}{\partial \mathbf{x}}$.
3. 加法: $\mathbf{u}=\mathbf{u}(\mathbf{x}), \mathbf{v}=\mathbf{v}(\mathbf{x})$, 那么 $\frac{\partial(\mathbf{u}+\mathbf{v})}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$.
4. 链式法则
   (a) $\mathbf{u}=\mathbf{u}(\mathbf{x}), \mathbf{g}=\mathbf{g}(\mathbf{u}), \frac{\partial \mathbf{g}}{\partial \mathbf{x}} = \frac{\partial \mathbf{g}}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$.
   (b) $\mathbf{u}=\mathbf{u}(\mathbf{x}), \mathbf{f}=\mathbf{f}(\mathbf{g}), \mathbf{g}=\mathbf{g}(\mathbf{u}), \frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \frac{\partial \mathbf{f}}{\partial \mathbf{g}}\frac{\partial \mathbf{g}}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$.

5.2. 标量对向量.

1. 常数相关: a 中不含 $\mathbf{x}$, $u=u(\mathbf{x})$. $\frac{\partial au}{\partial \mathbf{x}} = a\frac{\partial u}{\partial \mathbf{x}}$.
2. 加法: $u=u(\mathbf{x}), v=v(\mathbf{x})$. $\frac{\partial(u+v)}{\partial \mathbf{x}} = \frac{\partial u}{\partial \mathbf{x}} + \frac{\partial v}{\partial \mathbf{x}}$.
3. 乘法: 如果 $u=u(\mathbf{x}), v=v(\mathbf{x})$, $\frac{\partial uv}{\partial \mathbf{x}} = u\frac{\partial v}{\partial \mathbf{x}} + v\frac{\partial u}{\partial \mathbf{x}}$.
4. 链式法则: $f=f(g), g=g(u), u=u(\mathbf{x}), \frac{\partial f}{\partial \mathbf{x}} = \frac{\partial f}{\partial g}\frac{\partial g}{\partial u}\frac{\partial u}{\partial \mathbf{x}}$.
5. 点积: $\mathbf{u}=\mathbf{u}(\mathbf{x}), \mathbf{v}=\mathbf{v}(\mathbf{x}), \frac{\partial(\mathbf{u}\cdot\mathbf{v})}{\partial \mathbf{x}} = \mathbf{u}'\frac{\partial \mathbf{v}}{\partial \mathbf{x}} + \mathbf{v}'\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$.
6. 二次型: $\mathbf{u}=\mathbf{u}(\mathbf{x}), \mathbf{v}=\mathbf{v}(\mathbf{x}), \mathbf{A}$ 中不含 $\mathbf{x}, \frac{\partial(\mathbf{u}'\mathbf{A}\mathbf{v})}{\partial \mathbf{x}} = \mathbf{u}'\mathbf{A}\frac{\partial \mathbf{v}}{\partial \mathbf{x}} + \mathbf{v}'\mathbf{A}'\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$.

5.3. 向量对标量.

1. 常数相关:
   (a) 常数: a 与 x 无关, $\mathbf{u}=\mathbf{u}(x)$. $\frac{\partial a\mathbf{u}}{\partial x}=a\frac{\partial \mathbf{u}}{\partial x}$.
   (b) 常矩阵: $\mathbf{A}$ 与 x 无关, $\mathbf{u}=\mathbf{u}(x)$. $\frac{\partial \mathbf{A}\mathbf{u}}{\partial x}=\mathbf{A}\frac{\partial \mathbf{u}}{\partial x}$.
2. 转置: 如果 $\mathbf{u}=\mathbf{u}(x)$, 那么 $\frac{\partial \mathbf{u}'}{\partial x}=\left(\frac{\partial \mathbf{u}}{\partial x}\right)'$.
3. 加法: $\mathbf{u}=\mathbf{u}(x),\mathbf{v}=\mathbf{v}(x)$. 那么 $\frac{\partial (\mathbf{u}+\mathbf{v})}{\partial x}=\frac{\partial \mathbf{u}}{\partial x}+\frac{\partial \mathbf{v}}{\partial x}$.
4. 乘法相关
   (a) 向量的外积: $\mathbf{u}=\mathbf{u}(x),\mathbf{v}=\mathbf{v}(x)$. $\frac{\partial (\mathbf{u}'\times\mathbf{v})}{\partial x}=\left(\frac{\partial \mathbf{u}}{\partial x}\right)'\times\mathbf{v}+\mathbf{u}'\times\frac{\partial \mathbf{v}}{\partial x}$.
   (b) 矩阵的外积: $\frac{\partial (\mathbf{U}\times\mathbf{v})}{\partial x}=\frac{\partial \mathbf{U}}{\partial x}\times\mathbf{v}+\mathbf{U}\times\frac{\partial \mathbf{x}}{\partial x}$.
5. 链式法则: $\mathbf{u}=\mathbf{u}(x),\mathbf{f}=\mathbf{f}(\mathbf{g}),\mathbf{g}=\mathbf{g}(\mathbf{u})$. 那么 $\frac{\partial \mathbf{f}}{\partial x}=\frac{\partial \mathbf{f}}{\partial \mathbf{g}}\frac{\partial \mathbf{g}}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial x}$.

5.4. 向量对矩阵.

1. 常数相关: a 不含 $\mathbf{X}$ 相关的变量, $u=u(\mathbf{X})$. 那么 $\frac{\partial au}{\partial \mathbf{X}}=a\frac{\partial u}{\partial \mathbf{x}}$.
2. 加法: 如果 $u=u(\mathbf{X}),v=v(\mathbf{X})$. 那么 $\frac{\partial (u+v)}{\partial \mathbf{X}}=\frac{\partial u}{\partial \mathbf{X}}+\frac{\partial v}{\partial \mathbf{X}}$.
3. 乘法: $u=u(\mathbf{X}),v=v(\mathbf{X})$. $\frac{\partial uv}{\partial \mathbf{X}}=u\frac{\partial v}{\partial \mathbf{X}}+v\frac{\partial u}{\partial \mathbf{X}}$.
4. 链式法则:
   (a) $u=u(\mathbf{X})$, $\frac{\partial f(g(u))}{\partial \mathbf{X}}=\frac{\partial f(g)}{\partial g}\frac{\partial g(u)}{\partial u}\frac{\partial u}{\partial \mathbf{X}}$.
   (b) $\mathbf{U}=\mathbf{U}(\mathbf{X})$, $\frac{\partial g(\mathbf{U})}{\partial X_{ij}}=\operatorname{tr}\left(\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}}\frac{\partial \mathbf{U}}{\partial X_{ij}}\right)$

5.5. 矩阵对标量.

1. 常数相关:
   (a) 标量常数: $\mathbf{U}=\mathbf{U}(x),a$ 不是 x 的函数: $\frac{\partial a\mathbf{U}}{\partial x}=a\frac{\partial \mathbf{U}}{\partial x}$.
   (b) 矩阵: $\mathbf{A},\mathbf{B}$ 不是 x 的函数, $\mathbf{U}=\mathbf{U}(x)$, $\frac{\partial \mathbf{AUB}}{\partial x}=\mathbf{A}\frac{\partial \mathbf{U}}{\partial x}\mathbf{B}$.
2. 加法: $\mathbf{U}=\mathbf{U}(x),\mathbf{V}=\mathbf{V}(x)$, $\frac{\partial (\mathbf{U}+\mathbf{V})}{\partial x}=\frac{\partial \mathbf{U}}{\partial x}+\frac{\partial \mathbf{V}}{\partial x}$.
3. 乘法相关:$\mathbf{U}=\mathbf{U}(x),\mathbf{V}=\mathbf{V}(x)$
   (a) 矩阵乘法: $\frac{\partial (\mathbf{UV})}{\partial x}=\mathbf{U}\frac{\partial \mathbf{V}}{\partial x}+\frac{\partial \mathbf{U}}{\partial x}\mathbf{V}$.
   (b) 张量乘法: $\frac{\partial (\mathbf{U}\otimes\mathbf{V})}{\partial x}=\mathbf{U}\otimes\frac{\partial \mathbf{V}}{\partial x}+\frac{\partial \mathbf{U}}{\partial x}\otimes\mathbf{V}$.
   (c) 逐元素相乘: $\frac{\partial (\mathbf{U}\circ\mathbf{V})}{\partial x}=\mathbf{U}\circ\frac{\partial \mathbf{V}}{\partial x}+\frac{\partial \mathbf{U}}{\partial x}\circ\mathbf{V}$
   (d) 矩阵的逆:
      i. 一元情况: $\frac{\partial \mathbf{U}^{-1}}{\partial x}=-\mathbf{U}^{-1}\frac{\partial \mathbf{U}}{\partial x}\mathbf{U}^{-1}$
      ii. 多元情况: $\mathbf{U}=\mathbf{U}(x,y)$, $\frac{\partial^2 \mathbf{U}^{-1}}{\partial x\partial y}=\mathbf{U}^{-1}\left(\frac{\partial \mathbf{U}}{\partial x}\mathbf{U}^{-1}\frac{\partial \mathbf{U}}{\partial y}-\frac{\partial^2 \mathbf{U}}{\partial x\partial y}+\frac{\partial \mathbf{U}}{\partial y}\mathbf{U}^{-1}\frac{\partial \mathbf{U}}{\partial x}\right)\mathbf{U}^{-1}$
4. 函数的复合: $\mathbf{A}$ 不是 x 的函数, $\mathbf{g}(\mathbf{X})$ 是标量形式的多项式 (可以是无穷级数的形式); $g'(x)$ 是对应的标量求导的结果; $\mathbf{g}'(\mathbf{x})$ 是对应矩阵的变化, 有 $\frac{\partial \mathbf{g}(x\mathbf{A})}{\partial x}=\mathbf{A}\mathbf{g}'(x\mathbf{A})=\mathbf{g}'(x\mathbf{A})\mathbf{A}$.