矩阵求导

Chen Hu

2021 年 3 月 28 日

目录

1 函数与标量，向量和矩阵 3
1.1 function 为标量 3
1.1.1 input 为标量 3
1.1.2 input 为向量 4
1.1.3 input 为矩阵 4
1.2 function 为向量 4
1.2.1 标量变元 4
1.2.2 向量变元 4
1.2.3 矩阵变元 5
1.3 function 为矩阵 5
1.3.1 标量变元 5
1.3.2 向量变元 5
1.3.3 矩阵变元 5

2 矩阵求导本质 6

3 矩阵求导结果的布局 6
3.1 直观上看 6
3.2 向量变元的实值标量函数 $f(\boldsymbol{x}), \boldsymbol{x} = [x_1, \ldots, x_n]^T$ 7
3.2.1 行向量偏导形式 (又称行偏导向量形式) 7

3.2.2 梯度向量形式 (又称列向量偏导形式) 7

3.3 矩阵变元的实值标量函数 $f(\boldsymbol{X}), \boldsymbol{X}_{m\times n} = (x_{ij})_{i=1,j=1}^{m,n}$ 7

3.3.1 $vec(\boldsymbol{X})$ 7

3.3.2 行向量偏导形式 (又称行偏导向量形式) 7

3.3.3 Jacobian 矩阵形式 7

3.3.4 梯度向量形式 (又称列向量偏导形式) 8

3.3.5 梯度矩阵形式 8

4 矩阵变元的实矩阵函数 $\boldsymbol{F}(\boldsymbol{X}), \boldsymbol{X}_{m\times n} = (x_{ij})_{i=1,j=1}^{m,n}, \boldsymbol{F}_{p\times q} = (f_{ij})_{i=1,j=1}^{p,q}$ 9

4.1 Jacobian 矩阵形式 9

4.2 梯度矩阵形式 10

5 向量变元的实值标量函数 10

5.1 函数形式 10

5.2 四个法则 11

5.2.1 常数求导 11

5.2.2 线性法则 11

5.2.3 乘积法则 11

5.2.4 商法则 11

5.3 几个公式 11

5.3.1 公式 1 11

5.3.2 公式 2 12

5.3.3 公式 3 12

5.3.4 公式 4 12

6 矩阵变元的实值标量函数 12

6.1 函数形式 12

6.2 四个法则 13

6.2.1 常数求导 13

6.2.2 线性法则 13

6.2.3 乘积法则 13

6.2.4 商法则 13

6.3 几个公式 13
6.3.1 公式 1 13
6.3.2 公式 2 13
6.3.3 公式 3 14
6.3.4 公式 4 14

7 矩阵的迹 **14**
7.1 定义 14
7.2 性质 14

8 微分与全微分 **15**

9 矩阵的微分 **15**
9.1 向量变元的实值标量函数 15
9.2 矩阵变元的实值标量函数 16
9.3 矩阵变元的实矩阵函数 17

1 函数与标量，向量和矩阵

考虑一个函数 $function(input)$, 针对 function 的类型，input 类型，可以将这个函数分为九种不同的种类

1.1 function 为标量

function 是一个实值标量函数，用细体小写字母 f 表示

1.1.1 input 为标量

1. 称 function 的变元是标量，用细体小写字母 x 表示

2. 示例

$$f(x) = x + 2 \tag{1}$$

1.1.2 input 为向量

1. 称 function 的变元为向量，用粗体小写字母 $\boldsymbol{x}$ 表示
2. 示例：设 $\boldsymbol{x} = [x_1, x_2, x_3]^T$

$$f(\boldsymbol{x}) = a_1x_1^2 + a_2x_2^2 + a_3x_3^2 + a_4x_1x_2 \tag{2}$$

1.1.3 input 为矩阵

1. 称 function 的变元为矩阵，用粗体的大写字母 $\boldsymbol{X}$ 表示
2. 示例：设 $\boldsymbol{X}_{3\times 2} = (x_{ij})_{i=1,j=1}^{3,2}$

$$f(\boldsymbol{X}) = a_1x_{11}^2 + a_2x_{12}^2 + a_3x_{21}^2 + a_4x_{22}^2 + a_5x_{31}^2 + a_6x_{32}^2 \tag{3}$$

1.2 function 为向量

function 是一个实向量函数，用粗体小写字母 f 表示

1.2.1 标量变元

$$\boldsymbol{f}_{3\times 1}(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ f_3(x) \end{bmatrix} = \begin{bmatrix} x+1 \\ 2x+1 \\ 3x^2+1 \end{bmatrix} \tag{4}$$

1.2.2 向量变元

1. 令 $\boldsymbol{x} = [x_1, x_2, x_3]^T$
2. 示例

$$\boldsymbol{F}_{3\times 1}(\boldsymbol{x}) = \begin{bmatrix} f_1(\boldsymbol{x}) \\ f_2(\boldsymbol{x}) \\ f_3(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} x_1 + x_2 + x_3 \\ x_1^2 + 2x_2 + x_3 \\ x_1x_2 + x_2 + 2x_3 \end{bmatrix} \tag{5}$$

1.2.3 矩阵变元

1. 令 $\boldsymbol{X} = (x_{ij})_{i=1,j=1}^{3,2}$
2. 示例

$$
\begin{aligned}
\boldsymbol{F}_{3\times 1}(\boldsymbol{X}) &= \begin{bmatrix} f_1(\boldsymbol{X}) \\ f_2(\boldsymbol{X}) \\ f_3(\boldsymbol{X}) \end{bmatrix} \\
&= \begin{bmatrix} x_{11}+x_{12}+x_{21}+x_{22}+x_{31}+x_{32} \\ 3x_{11}+x_{12}+x_{21}+x_{22}+x_{31}+x_{32} \\ 5x_{11}+x_{12}+x_{21}+x_{22}+x_{31}+x_{32} \end{bmatrix}
\end{aligned} \tag{6}
$$

1.3 function 为矩阵

1.3.1 标量变元

$$
\boldsymbol{F}_{3\times 2}(x) = \begin{bmatrix} f_{11}(x) & f_{12}(x) \\ f_{21}(x) & f_{22}(x) \\ f_{32}(x) & f_{32}(x) \end{bmatrix} = \begin{bmatrix} x+1 & 2x+2 \\ x^2+1 & 2x^2+1 \\ x^3+1 & 2x^3+1 \end{bmatrix} \tag{7}
$$

1.3.2 向量变元

1. 令 $\boldsymbol{x} = [x_1, x_2, x_3]^T$
2. 示例

$$
\boldsymbol{F}_{3\times 2}(\boldsymbol{x}) = \begin{bmatrix} f_{11}(\boldsymbol{x}) & f_{12}(\boldsymbol{x}) \\ f_{21}(\boldsymbol{x}) & f_{22}(\boldsymbol{x}) \\ f_{32}(\boldsymbol{x}) & f_{32}(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} 2x_1+x_2+x_3 & 2x_1+2x_2+x_3 \\ 2x_1+2x_2+x_3 & x_1+2x_2+x_3 \\ 2x_1+x_2+2x_3 & x_1+2x_2+2x_3 \end{bmatrix} \tag{8}
$$

1.3.3 矩阵变元

1. 令 $\boldsymbol{X} = (x_{ij})_{i=1,j=1}^{3,2}$

2. 示例

$$\boldsymbol{F}_{3\times 2}(\boldsymbol{X}) = \begin{bmatrix} f_{11}(\boldsymbol{X}) & f_{12}(\boldsymbol{X}) \\ f_{21}(\boldsymbol{X}) & f_{22}(\boldsymbol{X}) \\ f_{32}(\boldsymbol{X}) & f_{32}(\boldsymbol{X}) \end{bmatrix}$$
$$= \begin{bmatrix} x_{11}+x_{12}+x_{21}+x_{22}+x_{31}+x_{32} & 2x_{11}+x_{12}+x_{21}+x_{22}+x_{31}+x_{32} \\ 3x_{11}+x_{12}+x_{21}+x_{22}+x_{31}+x_{32} & 4x_{11}+x_{12}+x_{21}+x_{22}+x_{31}+x_{32} \\ 5x_{11}+x_{12}+x_{21}+x_{22}+x_{31}+x_{32} & 6x_{11}+x_{12}+x_{21}+x_{22}+x_{31}+x_{32} \end{bmatrix} \tag{9}$$

function 是一个实矩阵函数，用粗体大写字母 F 表示

2 矩阵求导本质

矩阵求导本质就是 function 对每个 f 分别对变元中每个元素逐个求偏导，只不过也写成了向量，矩阵形式而已

3 矩阵求导结果的布局

3.1 直观上看

1. 分子布局：分子是列向量形式，分母是行向量形式

$$\frac{\partial \boldsymbol{f}_{2\times 1}(\boldsymbol{x})}{\partial \boldsymbol{x}^T_{3\times 1}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} \end{bmatrix} \tag{10}$$

2. 分母布局：分母是列向量形式，分子是行向量形式

$$\frac{\partial \boldsymbol{f}^T_{2\times 1}(\boldsymbol{x})}{\partial \boldsymbol{x}_{3\times 1}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} \\ \frac{\partial f_1}{\partial x_3} & \frac{\partial f_2}{\partial x_3} \end{bmatrix} \tag{11}$$

3.2 向量变元的实值标量函数 $f(\boldsymbol{x}), \boldsymbol{x} = [x_1, \ldots, x_n]^T$

3.2.1 行向量偏导形式 (又称行偏导向量形式)

$$D_x f(\boldsymbol{x}) = \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}^T} = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right] \tag{12}$$

3.2.2 梯度向量形式 (又称列向量偏导形式)

$$\nabla_x f(\boldsymbol{x}) = \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right]^T \tag{13}$$

3.3 矩阵变元的实值标量函数 $f(\boldsymbol{X}), \boldsymbol{X}_{m\times n} = (x_{ij})_{i=1,j=1}^{m,n}$

3.3.1 $vec(\boldsymbol{X})$

1. 矩阵 $\boldsymbol{X}$ 按列堆栈来向量化
2. 示例

$$vec(\boldsymbol{X}) = [x_{11}, x_{21}, \ldots, x_{m1}, x_{12}, x_{22}, \ldots, x_{m2}, \ldots, x_{1n}, x_{2n}, \ldots, x_{mn}]^T \tag{14}$$

3.3.2 行向量偏导形式 (又称行偏导向量形式)

1. 先把矩阵变元 $\boldsymbol{X}$ 按 $vec(\boldsymbol{X})$ 向量化，再对该向量变元应用等式 (12)
2. 示例

$$\begin{aligned} D_{vec(\boldsymbol{X})} f(\boldsymbol{X}) &= \frac{\partial f(\boldsymbol{X})}{\partial vec^T(\boldsymbol{X})} \\ &= \left[\frac{\partial f}{\partial x_{11}}, \frac{\partial f}{\partial x_{21}}, \ldots, \frac{\partial f}{\partial x_{m1}}, \frac{\partial f}{\partial x_{12}}, \frac{\partial f}{\partial x_{22}}, \ldots, \frac{\partial f}{\partial x_{m2}}, \frac{\partial f}{\partial x_{1n}}, \frac{\partial f}{\partial x_{2n}}, \ldots, \frac{\partial f}{\partial x_{mn}}\right] \end{aligned} \tag{15}$$

3.3.3 Jacobian 矩阵形式

1. 先把矩阵变元 $\boldsymbol{X}$ 进行转置，再对转置后的每个元素逐个求偏导，结果布局和转置布局一样

2. 示例

$$
\begin{aligned}
D_{\boldsymbol{X}} f(\boldsymbol{X}) &= \frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}_{m\times n}^{T}} \\
&= \begin{bmatrix}
\frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{21}} & \cdots & \frac{\partial f}{\partial x_{m1}} \\
\frac{\partial f}{\partial x_{12}} & \frac{\partial f}{\partial x_{22}} & \cdots & \frac{\partial f}{\partial x_{m2}} \\
\vdots & \vdots & \vdots & \vdots \\
\frac{\partial f}{\partial x_{1n}} & \frac{\partial f}{\partial x_{2n}} & \cdots & \frac{\partial f}{\partial x_{mn}}
\end{bmatrix}_{n\times m}
\end{aligned}
\tag{16}
$$

3.3.4 梯度向量形式 (又称列向量偏导形式)

1. 先把原矩阵变元 $\boldsymbol{X}$ 按 vec 向量化，转变为向量变元，再对该变元使用等式 (13)

2. 示例：

$$
\begin{aligned}
\nabla_{vec(\boldsymbol{X})} f(\boldsymbol{X}) &= \frac{\partial f(\boldsymbol{X})}{\partial vec(\boldsymbol{X})} \\
&= \left[\frac{\partial f}{\partial x_{11}}, \frac{\partial f}{\partial x_{21}}, \ldots, \frac{\partial f}{\partial x_{m1}}, \frac{\partial f}{\partial x_{12}}, \frac{\partial f}{\partial x_{22}}, \ldots, \frac{\partial f}{\partial x_{m2}}, \frac{\partial f}{\partial x_{1n}}, \frac{\partial f}{\partial x_{2n}}, \ldots, \frac{\partial f}{\partial x_{mn}}\right]^{T}
\end{aligned}
\tag{17}
$$

3.3.5 梯度矩阵形式

1. 直接对原矩阵变元 $\boldsymbol{X}$ 每个位置元素逐个求偏导，结果布局和原矩阵布局一致

2. 示例

$$
\begin{aligned}
\nabla_{\boldsymbol{X}} f(\boldsymbol{X}) &= \frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}_{m\times n}^{T}} \\
&= \begin{bmatrix}
\frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{12}} & \cdots & \frac{\partial f}{\partial x_{1n}} \\
\frac{\partial f}{\partial x_{21}} & \frac{\partial f}{\partial x_{22}} & \cdots & \frac{\partial f}{\partial x_{2n}} \\
\vdots & \vdots & \vdots & \vdots \\
\frac{\partial f}{\partial x_{m1}} & \frac{\partial f}{\partial x_{m2}} & \cdots & \frac{\partial f}{\partial x_{mn}}
\end{bmatrix}_{m\times n}
\end{aligned}
\tag{18}
$$

4 矩阵变元的实矩阵函数 $\boldsymbol{F}(\boldsymbol{X}), \boldsymbol{X}_{m\times n} = (x_{ij})^{m,n}_{i=1,j=1}, \boldsymbol{F}_{p\times q} = (f_{ij})^{p,q}_{i=1,j=1}$

4.1 Jacobian 矩阵形式

1. 先把矩阵变元 $\boldsymbol{X}$ 按 vec 向量化，转换为向量变元

$$vec(\boldsymbol{X}) = [x_{11}, x_{21}, \ldots, x_{m1}, x_{12}, x_{22}, \ldots, x_{m2}, \ldots, x_{1n}, x_{2n}, \ldots, x_{mn}]^T \tag{19}$$

2. 将实矩阵函数 $\boldsymbol{F}$ 按 vec 向量化，转换为实向量函数

$$vec(F(\boldsymbol{X})) = [f_{11}(\boldsymbol{X}), f_{21}(\boldsymbol{X}), \ldots, f_{m1}(\boldsymbol{X}), f_{12}(\boldsymbol{X}), f_{22}(\boldsymbol{X}), \ldots, f_{m2}(\boldsymbol{X}), \ldots, f_{1n}(\boldsymbol{X}), f_{2n}(\boldsymbol{X}), \ldots \tag{20}$$

3. 写出布局为 $pq \times mn$ 的矩阵

$$\begin{aligned} D_{\boldsymbol{X}}\boldsymbol{F}(\boldsymbol{X}) &= \frac{\partial vec_{pg\times 1}(\boldsymbol{F}(\boldsymbol{X}))}{\partial vec^T_{mn\times 1}\boldsymbol{X}} \\ &= \begin{bmatrix} \frac{\partial f_{11}}{\partial x_{11}} & \frac{\partial f_{11}}{\partial x_{21}} & \cdots & \frac{\partial f_{11}}{\partial x_{m1}} & \frac{\partial f_{11}}{\partial x_{12}} & \frac{\partial f_{11}}{\partial x_{22}} & \cdots & \frac{\partial f_{11}}{\partial x_{m2}} & \cdots & \frac{\partial f_{11}}{\partial x_{1n}} & \frac{\partial f_{11}}{\partial x_{2n}} & \cdots & \frac{\partial f_{11}}{\partial x_{mn}} \\ \frac{\partial f_{21}}{\partial x_{11}} & \frac{\partial f_{21}}{\partial x_{21}} & \cdots & \frac{\partial f_{21}}{\partial x_{m1}} & \frac{\partial f_{21}}{\partial x_{12}} & \frac{\partial f_{21}}{\partial x_{22}} & \cdots & \frac{\partial f_{21}}{\partial x_{m2}} & \cdots & \frac{\partial f_{21}}{\partial x_{1n}} & \frac{\partial f_{21}}{\partial x_{2n}} & \cdots & \frac{\partial f_{21}}{\partial x_{mn}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_{p1}}{\partial x_{11}} & \frac{\partial f_{p1}}{\partial x_{21}} & \cdots & \frac{\partial f_{p1}}{\partial x_{m1}} & \frac{\partial f_{p1}}{\partial x_{12}} & \frac{\partial f_{p1}}{\partial x_{22}} & \cdots & \frac{\partial f_{p1}}{\partial x_{m2}} & \cdots & \frac{\partial f_{p1}}{\partial x_{1n}} & \frac{\partial f_{p1}}{\partial x_{2n}} & \cdots & \frac{\partial f_{p1}}{\partial x_{mn}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_{p2}}{\partial x_{11}} & \frac{\partial f_{p2}}{\partial x_{21}} & \cdots & \frac{\partial f_{p2}}{\partial x_{m1}} & \frac{\partial f_{p2}}{\partial x_{12}} & \frac{\partial f_{p2}}{\partial x_{22}} & \cdots & \frac{\partial f_{p2}}{\partial x_{m2}} & \cdots & \frac{\partial f_{p2}}{\partial x_{1n}} & \frac{\partial f_{p2}}{\partial x_{2n}} & \cdots & \frac{\partial f_{p2}}{\partial x_{mn}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_{1q}}{\partial x_{11}} & \frac{\partial f_{1q}}{\partial x_{21}} & \cdots & \frac{\partial f_{1q}}{\partial x_{m1}} & \frac{\partial f_{1q}}{\partial x_{12}} & \frac{\partial f_{1q}}{\partial x_{22}} & \cdots & \frac{\partial f_{1q}}{\partial x_{m2}} & \cdots & \frac{\partial f_{1q}}{\partial x_{1n}} & \frac{\partial f_{1q}}{\partial x_{2n}} & \cdots & \frac{\partial f_{1q}}{\partial x_{mn}} \\ \frac{\partial f_{2q}}{\partial x_{11}} & \frac{\partial f_{2q}}{\partial x_{21}} & \cdots & \frac{\partial f_{2q}}{\partial x_{m1}} & \frac{\partial f_{2q}}{\partial x_{12}} & \frac{\partial f_{2q}}{\partial x_{22}} & \cdots & \frac{\partial f_{2q}}{\partial x_{m2}} & \cdots & \frac{\partial f_{2q}}{\partial x_{1n}} & \frac{\partial f_{2q}}{\partial x_{2n}} & \cdots & \frac{\partial f_{2q}}{\partial x_{mn}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_{pq}}{\partial x_{11}} & \frac{\partial f_{pq}}{\partial x_{21}} & \cdots & \frac{\partial f_{pq}}{\partial x_{m1}} & \frac{\partial f_{pq}}{\partial x_{12}} & \frac{\partial f_{pq}}{\partial x_{22}} & \cdots & \frac{\partial f_{pq}}{\partial x_{m2}} & \cdots & \frac{\partial f_{pq}}{\partial x_{1n}} & \frac{\partial f_{pq}}{\partial x_{2n}} & \cdots & \frac{\partial f_{pq}}{\partial x_{mn}} \end{bmatrix}_{pq\times} \end{aligned} \tag{21}$$

4.2 梯度矩阵形式

$$
\begin{aligned}
\nabla_{\boldsymbol{X}}\boldsymbol{F}(\boldsymbol{X}) &= \frac{\partial vec_{pg\times 1}^T(\boldsymbol{F}(\boldsymbol{X}))}{\partial vec_{mn\times 1}\boldsymbol{X}} \\
&= \begin{bmatrix}
\frac{\partial f_{11}}{\partial x_{11}} & \frac{\partial f_{21}}{\partial x_{11}} & \cdots & \frac{\partial f_{p1}}{\partial x_{11}} & \frac{\partial f_{12}}{\partial x_{11}} & \frac{\partial f_{22}}{\partial x_{11}} & \cdots & \frac{\partial f_{p2}}{\partial x_{11}} & \cdots & \frac{\partial f_{1q}}{\partial x_{11}} & \cdots & \frac{\partial f_{pq}}{\partial x_{11}} \\
\frac{\partial f_{11}}{\partial x_{21}} & \frac{\partial f_{21}}{\partial x_{21}} & \cdots & \frac{\partial f_{p1}}{\partial x_{21}} & \frac{\partial f_{12}}{\partial x_{21}} & \frac{\partial f_{22}}{\partial x_{21}} & \cdots & \frac{\partial f_{p2}}{\partial x_{21}} & \cdots & \frac{\partial f_{1q}}{\partial x_{21}} & \cdots & \frac{\partial f_{pq}}{\partial x_{21}} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\frac{\partial f_{11}}{\partial x_{m1}} & \frac{\partial f_{21}}{\partial x_{m1}} & \cdots & \frac{\partial f_{p1}}{\partial x_{m1}} & \frac{\partial f_{12}}{\partial x_{m1}} & \frac{\partial f_{22}}{\partial x_{m1}} & \cdots & \frac{\partial f_{p2}}{\partial x_{m1}} & \cdots & \frac{\partial f_{1q}}{\partial x_{m1}} & \cdots & \frac{\partial f_{pq}}{\partial x_{m1}} \\
\frac{\partial f_{11}}{\partial x_{12}} & \frac{\partial f_{21}}{\partial x_{12}} & \cdots & \frac{\partial f_{p1}}{\partial x_{12}} & \frac{\partial f_{12}}{\partial x_{12}} & \frac{\partial f_{22}}{\partial x_{12}} & \cdots & \frac{\partial f_{p2}}{\partial x_{12}} & \cdots & \frac{\partial f_{1q}}{\partial x_{12}} & \cdots & \frac{\partial f_{pq}}{\partial x_{12}} \\
\frac{\partial f_{11}}{\partial x_{22}} & \frac{\partial f_{21}}{\partial x_{22}} & \cdots & \frac{\partial f_{p1}}{\partial x_{22}} & \frac{\partial f_{12}}{\partial x_{22}} & \frac{\partial f_{22}}{\partial x_{22}} & \cdots & \frac{\partial f_{p2}}{\partial x_{22}} & \cdots & \frac{\partial f_{1q}}{\partial x_{22}} & \cdots & \frac{\partial f_{pq}}{\partial x_{22}} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\frac{\partial f_{11}}{\partial x_{m2}} & \frac{\partial f_{21}}{\partial x_{m2}} & \cdots & \frac{\partial f_{p1}}{\partial x_{m2}} & \frac{\partial f_{12}}{\partial x_{m2}} & \frac{\partial f_{22}}{\partial x_{m2}} & \cdots & \frac{\partial f_{p2}}{\partial x_{m2}} & \cdots & \frac{\partial f_{1q}}{\partial x_{m2}} & \cdots & \frac{\partial f_{pq}}{\partial x_{m2}} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\frac{\partial f_{11}}{\partial x_{1n}} & \frac{\partial f_{21}}{\partial x_{1n}} & \cdots & \frac{\partial f_{p1}}{\partial x_{1n}} & \frac{\partial f_{12}}{\partial x_{1n}} & \frac{\partial f_{22}}{\partial x_{1n}} & \cdots & \frac{\partial f_{p2}}{\partial x_{1n}} & \cdots & \frac{\partial f_{1q}}{\partial x_{1n}} & \cdots & \frac{\partial f_{pq}}{\partial x_{1n}} \\
\frac{\partial f_{11}}{\partial x_{2n}} & \frac{\partial f_{21}}{\partial x_{2n}} & \cdots & \frac{\partial f_{p1}}{\partial x_{2n}} & \frac{\partial f_{12}}{\partial x_{2n}} & \frac{\partial f_{22}}{\partial x_{2n}} & \cdots & \frac{\partial f_{p2}}{\partial x_{2n}} & \cdots & \frac{\partial f_{1q}}{\partial x_{2n}} & \cdots & \frac{\partial f_{pq}}{\partial x_{2n}} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\frac{\partial f_{11}}{\partial x_{mn}} & \frac{\partial f_{21}}{\partial x_{mn}} & \cdots & \frac{\partial f_{p1}}{\partial x_{mn}} & \frac{\partial f_{12}}{\partial x_{mn}} & \frac{\partial f_{22}}{\partial x_{mn}} & \cdots & \frac{\partial f_{p2}}{\partial x_{mn}} & \cdots & \frac{\partial f_{1q}}{\partial x_{mn}} & \cdots & \frac{\partial f_{pq}}{\partial x_{mn}}
\end{bmatrix}_{mn\times pq}
\end{aligned} \tag{22}
$$

5 向量变元的实值标量函数

5.1 函数形式

1. 函数形式 $f(\boldsymbol{x}), \boldsymbol{x} = [x_1, x_2, \ldots, x_n]^T$
2. 使用梯度形式，有

$$
\begin{aligned}
\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) &= \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} \\
&= \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right]^T
\end{aligned} \tag{23}
$$

5.2 四个法则

5.2.1 常数求导

与一元函数求导相同，结果为零向量

$$\frac{\partial c}{\partial \boldsymbol{x}} = \boldsymbol{0}_{n\times 1} \tag{24}$$

其中，c 为常数

5.2.2 线性法则

与一元函数求导法则相同：相加再求导等于求导再相加，常数提外面

$$\frac{\partial [c_1 f(\boldsymbol{x}) + c_2 g(\boldsymbol{x})]}{\partial \boldsymbol{x}} = c_1 \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} + c_2 \frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}} \tag{25}$$

其中，c_1, c_2 为常数

5.2.3 乘积法则

与一元函数求导乘积法则相同，前导后不导加前不导后导

$$\frac{\partial [f(\boldsymbol{x}) g(\boldsymbol{x})]}{\partial \boldsymbol{x}} = \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} g(\boldsymbol{x}) + f(\boldsymbol{x}) \frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}} \tag{26}$$

5.2.4 商法则

与一元函数求导商法则相同，上导下不导减上不导下导除以下的平方

$$\frac{\partial \left[\frac{f(\boldsymbol{x})}{g(\boldsymbol{x})}\right]}{\partial \boldsymbol{x}} = \frac{1}{g^2(\boldsymbol{x})} \left[\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} g(\boldsymbol{x}) - f(\boldsymbol{x}) \frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}}\right] \tag{27}$$

5.3 几个公式

5.3.1 公式 1

$$\frac{\partial (\boldsymbol{x}^T \boldsymbol{a})}{\partial \boldsymbol{x}} = \frac{\partial (\boldsymbol{a}^T \boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{a} \tag{28}$$

其中，$\boldsymbol{a}$ 为常数向量，$\boldsymbol{a} = (a_1, a_2, \ldots, a_n)^T$

5.3.2 公式 2

$$\frac{\partial(\boldsymbol{x}^T\boldsymbol{x})}{\partial\boldsymbol{x}} = 2\boldsymbol{x} \tag{29}$$

5.3.3 公式 3

$$\frac{\partial(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{a})}{\partial\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}^T\boldsymbol{x} \tag{30}$$

其中，$\boldsymbol{A}_{n\times n}$ 为常数矩阵

5.3.4 公式 4

$$\frac{\partial(\boldsymbol{a}^T\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{b})}{\partial\boldsymbol{x}} = \boldsymbol{a}\boldsymbol{b}^T\boldsymbol{x} + \boldsymbol{b}\boldsymbol{a}^T\boldsymbol{x} \tag{31}$$

其中, $\boldsymbol{a}, \boldsymbol{b}$ 为常数常量

6 矩阵变元的实值标量函数

6.1 函数形式

1. 函数形式

$$f(\boldsymbol{X}), \boldsymbol{X}_{m\times n} = (x_{ij})_{i=1,j=1}^{m,n} \tag{32}$$

2. 使用梯度矩阵形式

$$\begin{aligned}\nabla_{\boldsymbol{X}} f(\boldsymbol{X}) &= \frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}_{m\times n}} \\ &= \begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{12}} & \cdots & \frac{\partial f}{\partial x_{1n}} \\ \frac{\partial f}{\partial x_{21}} & \frac{\partial f}{\partial x_{22}} & \cdots & \frac{\partial f}{\partial x_{2n}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f}{\partial x_{m1}} & \frac{\partial f}{\partial x_{m2}} & \cdots & \frac{\partial f}{\partial x_{mn}} \end{bmatrix}_{m\times n}\end{aligned} \tag{33}$$

6.2 四个法则

6.2.1 常数求导

$$\frac{\partial c}{\partial \boldsymbol{X}} = \mathbf{0}_{m\times n} \tag{34}$$

6.2.2 线性法则

$$\frac{\partial[c_1 f(\boldsymbol{X}) + c_2 g(\boldsymbol{X})]}{\partial \boldsymbol{X}} = c_1\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}} + c_2\frac{\partial g(\boldsymbol{X})}{\partial \boldsymbol{X}} \tag{35}$$

6.2.3 乘积法则

$$\frac{\partial[f(\boldsymbol{X}) g(\boldsymbol{X})]}{\partial \boldsymbol{X}} = \frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}} g(\boldsymbol{X}) + f(\boldsymbol{X})\frac{\partial g(\boldsymbol{X})}{\partial \boldsymbol{x}} \tag{36}$$

6.2.4 商法则

$$\frac{\partial\left[\frac{f(\boldsymbol{X})}{g(\boldsymbol{X})}\right]}{\partial \boldsymbol{X}} = \frac{1}{g^2(\boldsymbol{X})}\left[\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}} g(\boldsymbol{X}) - f(\boldsymbol{X})\frac{\partial g(\boldsymbol{X})}{\partial \boldsymbol{X}}\right] \tag{37}$$

6.3 几个公式

6.3.1 公式 1

$$\frac{\partial(\boldsymbol{a}^T\boldsymbol{X}\boldsymbol{b})}{\partial \boldsymbol{X}} = \boldsymbol{a}\boldsymbol{b}^T) \tag{38}$$

其中，$\boldsymbol{a}_{m\times 1}, \boldsymbol{b}_{n\times 1}$ 为常数向量，$\boldsymbol{a} = (a_1, a_2, \ldots, a_m)^T, \boldsymbol{b} = (b_1, b_2, \ldots, b_n)^T$

6.3.2 公式 2

$$\frac{\partial(\boldsymbol{a}^T\boldsymbol{X}^T\boldsymbol{b})}{\partial \boldsymbol{X}} = \boldsymbol{b}\boldsymbol{a}^T) \tag{39}$$

其中，$\boldsymbol{a}_{n\times 1}, \boldsymbol{b}_{m\times 1}$ 为常数向量，$\boldsymbol{a} = (a_1, a_2, \ldots, a_n)^T, \boldsymbol{b} = (b_1, b_2, \ldots, b^T_)$

6.3.3 公式 3

$$\frac{\partial(\boldsymbol{a}^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{b})}{\partial\boldsymbol{X}} = \boldsymbol{a}\boldsymbol{b}^T\boldsymbol{X} + \boldsymbol{b}\boldsymbol{a}^T\boldsymbol{X} \tag{40}$$

其中，$\boldsymbol{a}_{m\times1}, \boldsymbol{b}_{m\times1}$ 为常数向量，$\boldsymbol{a} = (a_1, a_2, \ldots, a_m)^T, \boldsymbol{b} = (b_1, b_2, \ldots, b_m)^T$

6.3.4 公式 4

$$\frac{\partial(\boldsymbol{a}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{b})}{\partial\boldsymbol{X}} = \boldsymbol{X}\boldsymbol{b}\boldsymbol{a}^T + \boldsymbol{X}\boldsymbol{a}\boldsymbol{b}^T \tag{41}$$

其中，$\boldsymbol{a}_{n\times1}, \boldsymbol{b}_{n\times1}$ 为常数向量，$\boldsymbol{a} = (a_1, a_2, \ldots, a_n)^T, \boldsymbol{b} = (b_1, b_2, \ldots, b_n)^T$

7 矩阵的迹

7.1 定义

$n \times n$ 的方阵 $\boldsymbol{A}_{n\times n}$ 的主对角线元素之和称为矩阵 $\boldsymbol{A}$ 的迹 (trace), 记为 $tr(\boldsymbol{A})$

7.2 性质

1. 标量的迹：一个标量 x 可以看作 1×1 的矩阵，它的迹就是它自己

2. 线性法则: 相加再求迹等于求迹再相加, 标量提外面

$$tr(c_1\boldsymbol{A} + c_2\boldsymbol{B}) = c_1 tr(\boldsymbol{A}) + c_2 tr(\boldsymbol{B}) \tag{42}$$

3. 转置：转置矩阵的迹等于原矩阵的迹

4. 乘积的迹的本质对于两个阶数都是 $m \times n$ 的矩阵 $\boldsymbol{A}_{m\times n}, \boldsymbol{B}_{m\times n}$, 其中一个矩阵诚意另一个矩阵的转置的迹，本质上是 $\boldsymbol{A}_{m\times n}, \boldsymbol{B}_{m\times n}$ 两个矩

阵对应位置的元素相乘并相加，可以理解为向量点积在矩阵上的推广

$$tr(\boldsymbol{A}\boldsymbol{B}^T) = \begin{aligned} & a_{11}b_{11} + a_{12}b_{12} + \ldots + a_{1n}b_{1n} \\ & + a_{21}b_{21} + a_{22}b_{22} + \ldots + a_{2n}b_{2n} \\ & + \ldots \\ & + a_{m1}b_{m1} + a_{m2}b_{m2} + \ldots + a_{mn}b_{mn} \end{aligned} \tag{43}$$

5. 交换律：矩阵乘积位置互换，迹不变
6. 更多矩阵的交换律：

$$tr(\boldsymbol{ABC}) = tr(\boldsymbol{CAB}) = tr(\boldsymbol{BCA}) \tag{44}$$

7. 熟练使用

$$tr(\boldsymbol{A}\boldsymbol{B}^T) = tr(\boldsymbol{B}\boldsymbol{A}^T) = tr(\boldsymbol{A}^T\boldsymbol{B}) = tr(\boldsymbol{B}\boldsymbol{A}^T) \tag{45}$$

8 微分与全微分

9 矩阵的微分

9.1 向量变元的实值标量函数

1. 函数形式

$$f(\boldsymbol{x}), \boldsymbol{x} = [x_1, x_2, \ldots, x_n]^T \tag{46}$$

2. 全微分

$$\begin{aligned} df(\boldsymbol{x}) &= \frac{\partial f}{\partial x_1}dx_1 + \frac{\partial f}{\partial x_2}dx_2 + \ldots + + \frac{\partial f}{\partial x_n}dx_n \\ &= \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right) \begin{bmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{bmatrix} \end{aligned} \tag{47}$$

3. 因为结果是标量，也可以写为迹的形式

$$\begin{aligned} df(\boldsymbol{x}) &= \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}\right) \begin{bmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{bmatrix} \\ &= tr(\left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}\right) \begin{bmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{bmatrix}) \end{aligned} \tag{48}$$

9.2 矩阵变元的实值标量函数

1. 函数形式

$$f(\boldsymbol{X}), \boldsymbol{X}_{m\times n} = (x_{ij})_{i=1,j=1}^{m,n} \tag{49}$$

2. 全微分

$$\begin{aligned} df(\boldsymbol{X}) = &\frac{\partial f}{\partial x_{11}} dx_{11} + \frac{\partial f}{\partial x_{12}} dx_{12} + \dots + + \frac{\partial f}{\partial x_{1n}} dx_{1n} \\ &+ \frac{\partial f}{\partial x_{21}} dx_{21} + \frac{\partial f}{\partial x_{22}} dx_{22} + \dots + + \frac{\partial f}{\partial x_{2n}} dx_{2n} \\ &+ \dots \\ &+ \frac{\partial f}{\partial x_{m1}} dx_{m1} + \frac{\partial f}{\partial x_{m2}} dx_{m2} + \dots + + \frac{\partial f}{\partial x_{mn}} dx_{mn} \end{aligned} \tag{50}$$

3. 上式其实就是矩阵 $\left(\frac{\partial f}{\partial x_{ij}}\right)_{i=1,j=1}^{m,n}$ 与矩阵 $(dx_{ij})_{i=1,j=1}^{m,n}$ 对应位置的元素

相乘并相加，从等式 (43) 可以看出，上式可以写为两个矩阵相乘的迹

$$
\begin{aligned}
df(\boldsymbol{X}) &= \frac{\partial f}{\partial x_{11}}dx_{11} + \frac{\partial f}{\partial x_{12}}dx_{12} + \ldots + + \frac{\partial f}{\partial x_{1n}}dx_{1n} \\
&+ \frac{\partial f}{\partial x_{21}}dx_{21} + \frac{\partial f}{\partial x_{22}}dx_{22} + \ldots + + \frac{\partial f}{\partial x_{2n}}dx_{2n} \\
&+ \ldots \\
&+ \frac{\partial f}{\partial x_{m1}}dx_{m1} + \frac{\partial f}{\partial x_{m2}}dx_{m2} + \ldots + + \frac{\partial f}{\partial x_{mn}}dx_{mn} \\
&= tr(\begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{21}} & \ldots & \frac{\partial f}{\partial x_{m1}} \\ \frac{\partial f}{\partial x_{12}} & \frac{\partial f}{\partial x_{22}} & \ldots & \frac{\partial f}{\partial x_{m2}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f}{\partial x_{1n}} & \frac{\partial f}{\partial x_{2n}} & \ldots & \frac{\partial f}{\partial x_{mn}} \end{bmatrix}_{n\times m} \begin{bmatrix} dx_{11} & dx_{12} & \ldots & dx_{1n} \\ dx_{21} & dx_{22} & \ldots & dx_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ dx_{m1} & dx_{m2} & \ldots & dx_{mn} \end{bmatrix}_{m\times n})
\end{aligned}
\tag{51}
$$

9.3 矩阵变元的实矩阵函数

1. 函数形式

$$\boldsymbol{F}(\boldsymbol{X}), \boldsymbol{F}_{p\times q} = (f_{ij})_{i=1,j=1}^{p,q}, \boldsymbol{X}_{m\times n} = (x_{ij})_{i=1,j=1}^{m,n} \tag{52}$$

2. 全微分: 设 $f_{ij}(\boldsymbol{X})$ 可微

$$d\boldsymbol{F}_{p\times q}(\boldsymbol{X}) = \begin{bmatrix} df_{11}(\boldsymbol{X}) & f_{12}(\boldsymbol{X}) & \ldots & f_{1q}(\boldsymbol{X}) \\ df_{21}(\boldsymbol{X}) & f_{22}(\boldsymbol{X}) & \ldots & f_{2q}(\boldsymbol{X}) \\ \vdots & \vdots & \vdots & \vdots \\ df_{p1}(\boldsymbol{X}) & f_{p2}(\boldsymbol{X}) & \ldots & f_{pq}(\boldsymbol{X}) \end{bmatrix}_{p\times q} \tag{53}$$