矩阵代数与应用

上海大学计算机工程与科学学院

2024年1月

第七章 矩阵微分与梯度分析

- 矩阵微分
  - ◆Jacobian矩阵与梯度分析
  - ◆一阶实矩阵微分与Jacobian矩阵辨识
- 梯度分析初步
  - ◆实变函数无约束优化的梯度分析
  - ◆平滑凸优化的一阶算法
  - ◆约束凸优化算法

矩阵微分

量与量之间的对应关系(自变量←→ 应变量):

{标量, 向量, 矩阵} ←→ {标量, 向量, 矩阵}

（1） 实函数**矩阵**对**标量**变元的导数 （矩阵，元素都是标量变元的函数）

（2） 实矩阵**函数**对**矩阵**变元的导数 （标量函数，自变量是矩阵）

（3） 实函数**矩阵**对**矩阵**变元的导数

（4） 梯度矩阵、Jacobian矩阵与Hessian矩阵

（5） 实值标量函数的矩阵**微分**及计算

注： 此部分对应参考教材的7.1,7.2

自变量 因变量
标量→标量
标量→向量
标量→矩阵
向量→标量
向量→向量
向量→矩阵
矩阵→标量
矩阵→向量
矩阵→矩阵

实值函数的分类

函数类型\变量类型	标量变元$\boldsymbol{x} \in \mathbb{R}$	向量变元$\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{m}}$	矩阵变元$\boldsymbol{X} \in \mathbb{R}^{\boldsymbol{m \times n}}$
标量函数$\mathrm{f} \in \mathbb{R}$	$\mathrm{f(x)}$ $\mathrm{f}: \mathbb{R} \rightarrow \mathbb{R}$	$\mathrm{f(x)}$ $\mathrm{f}: \mathbb{R}^m \rightarrow \mathbb{R}$	$\cancel{\mathrm{f(x)}}$ $\cancel{\mathrm{f}: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}}$
向量函数$\mathrm{f} \in \mathbb{R}$	$\mathrm{f(x)}$ $\mathrm{f}: \mathbb{R} \rightarrow \mathbb{R}^{\mathrm{p}}$	$\mathrm{f(x)}$ $\mathrm{f}: \mathbb{R}^{\mathrm{m}} \rightarrow \mathbb{R}^p$	$\mathrm{f(x)}$ $\mathrm{f}: \mathbb{R}^{\mathrm{m \times n}} \rightarrow \mathbb{R}^p$
矩阵函数$\mathrm{F} \in \mathbb{R}^{p \times q}$	$\cancel{\mathrm{F(x)}}$ $\cancel{\mathrm{F}: \mathbb{R} \rightarrow \mathbb{R}^{p \times q}}$	$\mathrm{F(x)}$ $\mathrm{F}: \mathbb{R}^m \rightarrow \mathbb{R}^{p \times q}$	$\cancel{\mathrm{F(x)}}$ $\cancel{\mathrm{F}: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{p \times q}}$

复值函数的分类

函数类型\变量类型	标量变元$\mathrm{z}, z^* \in \mathbb{C}$	向量变元$\mathrm{z}, z^* \in \mathbb{C}^m$	矩阵变元$\mathrm{Z}, Z^* \in \mathbb{C}^{m \times n}$
标量函数$\mathrm{f} \in \mathbb{C}$	$\mathrm{f}(\mathrm{z}, z^*)$ $\mathrm{f}: \mathbb{C} \times \mathbb{C} \rightarrow \mathbb{C}$	$\mathrm{f}(\mathrm{z}, z^*)$ $\mathrm{f}: \mathbb{C}^{\mathrm{m}} \times \mathbb{C}^{\mathrm{m}} \rightarrow \mathbb{C}$	$\mathrm{f}(\mathrm{z}, z^*)$ $\mathrm{f}: \mathbb{C}^{m \times n} \times \mathbb{C}^{m \times n} \rightarrow \mathbb{C}$
向量函数$\mathrm{f} \in \mathbb{C}^p$	$\mathrm{f}(\mathrm{z}, z^*)$ $\mathrm{f}: \mathbb{C} \times \mathbb{C} \rightarrow \mathbb{C}^p$	$\mathrm{f}(\mathrm{z}, z^*)$ $\mathrm{f}: \mathbb{C}^{\mathrm{m}} \times \mathbb{C}^{\mathrm{m}} \rightarrow \mathbb{C}^p$	$\mathrm{f}(\mathrm{z}, z^*)$ $\mathrm{f}: \mathbb{C}^{m \times n} \times \mathbb{C}^{m \times n} \rightarrow \mathbb{C}^p$
矩阵函数$\mathrm{F} \in \mathbb{C}^{p \times q}$	$\mathrm{F}(\mathrm{z}, z^*)$ $\mathrm{F}: \mathbb{C} \times \mathbb{C} \rightarrow \mathbb{C}^{p \times q}$	$\mathrm{F}(\mathrm{z}, z^*)$ $\mathrm{F}: \mathbb{C}^{\mathrm{m}} \times \mathbb{C}^{\mathrm{m}} \rightarrow \mathbb{C}^{p \times q}$	$\mathrm{F}(\mathrm{z}, z^*)$ $\mathrm{F}: \mathbb{C}^{m \times n} \times \mathbb{C}^{m \times n} \rightarrow \mathbb{C}^{p \times q}$

(1)实函数矩阵对标量变元的导数

一对多！

$$A'(t) = \frac{\mathrm{d}A(t)}{\mathrm{d}t} = \left(\frac{\mathrm{d}a_{ij}(t)}{\mathrm{d}t}\right)_{m\times n}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(A(t) \pm B(t)\big) = \frac{\mathrm{d}}{\mathrm{d}t}A(t) \pm \frac{\mathrm{d}}{\mathrm{d}t}B(t)$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(A(t)B(t)\big) = \frac{\mathrm{d}}{\mathrm{d}t}A(t)\cdot B(t) + A(t)\frac{\mathrm{d}}{\mathrm{d}t}B(t)$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(a(t)A(t)\big) = \frac{\mathrm{d}a(t)}{\mathrm{d}t}A(t) + a(t)\frac{d}{\mathrm{d}t}A(t)$$

自变量是标量，应变量是由实函数组成的矩阵

$$\frac{\mathrm{d}}{\mathrm{d}t}e^{tA} = Ae^{tA} = e^{tA}A$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\cos(tA) =- Asin(tA) =-\sin(tA)\,A$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\sin(tA) = Acos(tA) =\cos(tA)\,A$$

若A不是方阵，如何定义exp(tA)?

◆ 高阶导数

$$\frac{\mathrm{d}^k A(t)}{\mathrm{d}t^k} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}^{k-1}A(t)}{\mathrm{d}t^{k-1}}\right) = \left(\frac{\mathrm{d}^k a_{ij}(t)}{\mathrm{d}t^{\,k}}\right)_{m\times n}$$

◆ 积分

$$\int_a^b A(t)\mathrm{d}t = \left(\int_a^b a_{ij}(t)\mathrm{d}t\right)_{m\times n}$$

例1 . MATLAB函数expm()和exp()

```
>> A=randn(3,3)
A =
   0.5377   0.8622  -0.4336
   1.8339   0.3188   0.3426
  -2.2588  -1.3077   3.5784

>> exp(A)
ans =
   1.7120   2.3683   0.6482
   6.2582   1.3754   1.4086
   0.1045   0.2704  35.8161

>> expm(A)
ans =
   7.0769   4.0651  -4.8645
   1.9420   1.6985   1.6878
 -38.6739 -23.3133  41.3345
```

例2 . MATLAB函数logm()和log()

```
>> A=randn(3,3)
A =
   1.4090  -1.2075   0.4889
   1.4172   0.7172   1.0347
   0.6715   1.6302   0.7269
>> B=expm(A)
B =
   1.9349  -2.5701  -0.1837
   5.1812   1.6601   3.0920
   5.0943   2.3454   4.3886
>> C=logm(B)
C =
   1.4090  -1.2075   0.4889
   1.4172   0.7172   1.0347
   0.6715   1.6302   0.7269
>> A-C
ans =
  1.0e-14 *
  -0.1554   0.1110  -0.0278
  -0.0444   0.0222  -0.1776
   0.1110  -0.0666   0.0222
```

例3 . 用MATLAB函数logm()计算log(tA)的导数

```
rng default;
A=rand(3,3);
disp(A);
t=3;
dt=0.001;
dA=logm((t+dt)*A)-logm(t*A);
B=dA/dt;
disp(B);
```

注：MATLAB中的exp()和log()函数是按元素计算的. 请注意expm()和exp()的区别.

```
0.8147    0.9134    0.2785
0.9058    0.6324    0.5469
0.1270    0.0975    0.9575

0.3333 + 0.0000i   0.0000 - 0.0000i   0.0000 + 0.0000i
0.0000 - 0.0000i   0.3333 + 0.0000i  -0.0000 - 0.0000i
0.0000 + 0.0000i   0.0000 - 0.0000i   0.3333 + 0.0000i
```

$$\frac{\mathrm{d}}{\mathrm{d}t}\log(tA) = I/t \quad (?)$$

对数函数泰勒级数展开式(公式)

$$\frac{\mathrm{d}}{\mathrm{d}t}\log(tA) = \mathrm{I}/t \quad (?)$$

$$\ln(x)=\sum_{n=1}^{+\infty}(-1)^{n-1}\frac{(x-1)^n}{n}$$

$$\frac{\mathbf{d}}{\mathbf{d}t}\log(tA)=\frac{\mathbf{d}}{\mathbf{d}t}\sum_{n=1}^{+\infty}(-1)^{n-1}\frac{(tA-I)^n}{n}=\sum_{n=1}^{+\infty}(-1)^{n-1}\frac{\mathbf{d}}{\mathbf{d}t}\frac{(tA-I)^n}{n}$$

$$=\sum_{n=1}^{+\infty}(-1)^{n-1}A(tA-I)^{n-1}=(1/t)\sum_{n=1}^{+\infty}(-1)^{n-1}tA(tA-I)^{n-1}$$

只需证明$\sum_{n=1}^{+\infty}(-1)^{n-1}tA(tA-I)^{n-1}=I$即可

令$f=\sum_{n=1}^{+\infty}(-1)^{n-1}(tA-I)^{n-1}$,则显然$f=I-(tA-I)f$，因而

$tAf=I$,即$\sum_{n=1}^{+\infty}(-1)^{n-1}tA(tA-I)^{n-1}=I$成立.

(注：级数必须收敛才行，此推导是形式推导）

应用：矩阵微分方程的解

定理1： 满足初始条件 $\mathbf{x}(t)|_{t=t_0} = \mathbf{x}(t_0)$ 的一阶线性常系数齐次微分方程组

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = \mathbf{A}\mathbf{x}(t)$$

有且仅有唯一解 $\mathbf{x}(t) = \mathrm{e}^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0)$

定理2： 一阶线性常系数非齐次微分方程组

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = \mathbf{A}\mathbf{x}(t) + \mathbf{b}(t)$$

的通解为

$$\mathbf{x}(t) = \mathrm{e}^{tA}\left(\mathbf{c} + \int_{t_0}^{t} \mathrm{e}^{-sA}\,\mathbf{b}(s)\mathrm{d}s\right)$$

其中c为任意常数向量.

定理3：n阶常系数齐次线性微分方程

$$\begin{cases} x^{(n)}(t)+a_1x^{(n-1)}(t)+a_2x^{(n-2)}(t)+\cdots+a_nx(t)=0 \\ x^{(i)}(t)|_{t=t_0}=x^{(i)}(t_0), \quad i=0,1,\cdots,n-1 \end{cases}$$

的解为

$$x(t)=\begin{bmatrix}1 & 0 & \cdots & 0\end{bmatrix}\mathrm{e}^{(t-t_0)A}\begin{bmatrix} x(t_0) \\ x'(t_0) \\ \vdots \\ x^{(n-1)}(t_0)\end{bmatrix}$$

其中

$$A=\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix}$$

注：令$(y_1,y_2,\ldots,y_n)=(x^{(0)},x^{(1)},\ldots,x^{(n-1)})$可推导求解过程.

定理4： n阶常系数非齐次线性微分方程

$$x^{(n)}(t)+a_1x^{(n-1)}(t)+a_2x^{(n-2)}(t)+\cdots+a_nx(t)=f(t)$$

的通解为

$$x(t)=[10\cdots0]\left(\mathrm{e}^{tA}c+\int_{t_0}^{t}\mathrm{e}^{A(t-s)}bf(s)\mathrm{d}s\right)$$

其中c为任意常数向量；$b=[0\quad 0\quad \cdots\quad 1]^{\mathrm{T}}$；而$A$同定理3.

多对一！

(2)实矩阵函数对矩阵变元的导数

$$\frac{\mathrm{d}f(\mathbf{X})}{\mathrm{d}\mathbf{X}}=\left(\frac{\partial f(\mathbf{X})}{\partial x_{ij}}\right)_{m\times n}=\begin{bmatrix}\frac{\partial f(\mathbf{X})}{\partial x_{11}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{1n}}\\ \vdots & \ddots & \vdots\\ \frac{\partial f(\mathbf{X})}{\partial x_{m1}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{mn}}\end{bmatrix}_{m\times n}$$

◆ 列向量偏导和行向量偏导($\mathbf{x}\in\mathbb{R}^{m\times 1}$)

$$\frac{\mathrm{d}f(\mathbf{x})}{\mathrm{d}\mathbf{x}}=\left[\frac{\partial f(\mathbf{x})}{\partial x_1},\cdots,\frac{\partial f(\mathbf{x})}{\partial x_m}\right]^{\mathrm{T}}$$

$$\frac{\mathrm{d}f(\mathbf{x})}{\mathrm{d}\mathbf{x}^{\mathrm{T}}}=\left[\frac{\partial f(\mathbf{x})}{\partial x_1},\cdots,\frac{\partial f(\mathbf{x})}{\partial x_m}\right]$$

$$\partial X=\begin{pmatrix}\partial X_1\\ \vdots\\ \partial X_m\end{pmatrix}\quad \partial_X=\begin{pmatrix}\partial_{X_1}\\ \vdots\\ \partial_{X_m}\end{pmatrix}$$

$$\partial_{\mathrm{x}}=\frac{\partial}{\partial x}$$

自变量是矩阵，应变量是实函数

标量函数对矩阵/向量变元导数的性质

1） 若$\mathbf{X}\in\mathbb{R}^{m\times n}$且$f(\mathbf{X})=c$为常数，则$\frac{\mathrm{d}c}{\mathrm{d}X}=O_{m\times n}$．（常数）

2） 若c_1,c_2为实常数，则

$$\frac{\mathrm{d}(c_1 f(\mathbf{X})+c_2 g(\mathbf{X}))}{\mathrm{d}\mathbf{X}}=c_1\frac{\mathrm{d}f(\mathbf{X})}{\mathrm{d}\mathbf{X}}+c_2\frac{\mathrm{d}g(\mathbf{X})}{\mathrm{d}\mathbf{X}}.\text{（线性法则）}$$

3） $\frac{\mathrm{d}f(\mathbf{X})g(\mathbf{X})}{\mathrm{d}\mathbf{X}}=\frac{\mathrm{d}f(\mathbf{X})}{\mathrm{d}\mathbf{X}}g(\mathbf{X})+f(\mathbf{X})\frac{\mathrm{d}g(\mathbf{X})}{\mathrm{d}\mathbf{X}}$.（乘法法则）

4） 若$g(\mathbf{X})\neq 0$，则

$$\frac{\mathrm{d}f(\mathbf{X})/g(\mathbf{X})}{\mathrm{d}\mathbf{X}}=\frac{1}{g^2(\mathbf{X})}\left[\frac{\mathrm{d}f(\mathbf{X})}{\mathrm{d}\mathbf{X}}g(\mathbf{X})-f(\mathbf{X})\frac{\mathrm{d}g(\mathbf{X})}{\mathrm{d}\mathbf{X}}\right].\text{（商法则）}$$

(*) $\frac{\mathrm{d}g(f(X))}{\mathrm{d}X}=\frac{\mathrm{d}g(f(X))}{\mathrm{d}f(X)}\frac{\mathrm{d}f(X)}{\mathrm{d}X}$.(链式法则)

(**) 变元独立性基本假设！！！

5） $\frac{\mathrm{d}\mathbf{a}^{\mathrm{T}}\mathbf{X}\mathbf{b}}{\mathrm{d}\mathbf{X}}=\mathbf{a}\mathbf{b}^{\mathrm{T}}\quad(\mathbf{X}\in\mathbb{R}^{m\times n})$

$$\frac{\mathrm{d}f(\mathbf{x})}{\mathrm{d}\mathbf{x}}=\left[\frac{\partial f(\mathbf{x})}{\partial x_1},\cdots,\frac{\partial f(\mathbf{x})}{\partial x_m}\right]^{\mathrm{T}}$$

$$\frac{\mathrm{d}f(\mathbf{x})}{\mathrm{d}\mathbf{x}^{\mathrm{T}}}=\left[\frac{\partial f(\mathbf{x})}{\partial x_1},\cdots,\frac{\partial f(\mathbf{x})}{\partial x_m}\right]$$

例4 函数$\boldsymbol{f(X)} = \mathbf{a}^{\mathrm{T}}\mathbf{Xb}$的导数

```
a=rand(1,5)';
b=rand(1,5)';
X=rand(5,5);
delta=0.01;
Z=zeros(5,5);
for ii=1:5
   for jj=1:5
      XX=X;
XX(ii,jj)=XX(ii,jj)+delta;
      f=a'*XX*b-a'*X*b;
      Z(ii,jj)=f/delta;
   end
end
disp(Z);
disp(a*b');
```

0.2942	0.1653	0.2111	0.5179	0.3979
0.0890	0.0500	0.0638	0.1566	0.1203
0.2850	0.1601	0.2045	0.5016	0.3854
0.0852	0.0479	0.0611	0.1500	0.1153
0.3252	0.1827	0.2333	0.5725	0.4398

0.2942	0.1653	0.2111	0.5179	0.3979
0.0890	0.0500	0.0638	0.1566	0.1203
0.2850	0.1601	0.2045	0.5016	0.3854
0.0852	0.0479	0.0611	0.1500	0.1153
0.3252	0.1827	0.2333	0.5725	0.4398

6） 若$\mathbf{X} \in \mathbb{R}^{n\times n}$非奇异，则

$$\frac{\mathrm{d}\mathbf{a}^{\mathrm{T}}\mathbf{X}^{-1}\mathbf{b}}{\mathrm{d}\mathbf{X}} = -\mathbf{X}^{-\mathrm{T}}\mathbf{a}\mathbf{b}^{\mathrm{T}}\mathbf{X}^{-\mathrm{T}} \qquad \mathbf{X}^{-\mathrm{T}} = (\mathbf{X}^{-1})^{\mathrm{T}}$$

7） $$\frac{\mathrm{d}a^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{b}}{\mathrm{d}\mathbf{X}} = \mathbf{X}(\mathbf{a}\mathbf{b}^{\mathrm{T}} + \mathbf{b}\mathbf{a}^{\mathrm{T}})$$

8） $$\frac{\mathrm{d}\mathbf{a}^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{b}}{\mathrm{d}\mathbf{X}} = (\mathbf{a}\mathbf{b}^{\mathrm{T}} + \mathbf{b}\mathbf{a}^{\mathrm{T}})\mathbf{X}$$

$$\frac{\mathrm{d}f(\mathbf{x})}{\mathrm{d}\mathbf{x}} = \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \cdots, \frac{\partial f(\mathbf{x})}{\partial x_m}\right]^{\mathrm{T}}$$

$$\frac{\mathrm{d}f(\mathbf{x})}{\mathrm{d}\mathbf{x}^{\mathrm{T}}} = \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \cdots, \frac{\partial f(\mathbf{x})}{\partial x_m}\right]$$

9） $$\frac{\mathrm{d}\exp(\mathbf{a}^{\mathrm{T}}\mathbf{X}\mathbf{b})}{\mathrm{d}\mathbf{X}} = \mathbf{a}\mathbf{b}^{\mathrm{T}}\exp(\mathbf{a}^{\mathrm{T}}\mathbf{X}\mathbf{b})$$

10） $$\frac{\mathrm{d}\mathbf{a}^{\mathrm{T}}\mathbf{x}}{\mathrm{d}\mathbf{x}} = \frac{\mathrm{d}\mathbf{x}^{\mathrm{T}}\mathbf{a}}{\mathrm{d}\mathbf{x}} = \mathbf{a}$$

11） $$\frac{\mathrm{d}\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{b}}{\mathrm{d}\mathbf{x}} = \mathbf{A}\mathbf{b}, \qquad \frac{\mathrm{d}\mathbf{b}^{\mathrm{T}}\mathbf{A}\mathbf{x}}{\mathrm{d}\mathbf{x}} = \mathbf{A}^{\mathrm{T}}\mathbf{b}$$

12） $$\frac{\mathrm{d}\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}{\mathrm{d}\mathbf{x}} = (\mathbf{A} + \mathbf{A}^{\mathrm{T}})\mathbf{x}$$

$\frac{\mathbf{d}\mathbf{a}^{\mathbf{T}}\mathbf{X}^{-1}\mathbf{b}}{\mathbf{d}\mathbf{X}} = -\mathbf{X}^{-\mathbf{T}}\mathbf{a}\mathbf{b}^{\mathbf{T}}\mathbf{X}^{-\mathbf{T}}$的推导

$$(X+\mathrm{d}X)^{-1}(X+\mathrm{d}X)=I$$

$$\Rightarrow (X+\mathrm{d}X)^{-1}X+(X+\mathrm{d}X)^{-1}\mathrm{d}X=I$$

$$\Rightarrow (X+\mathrm{d}X)^{-1}+(X+\mathrm{d}X)^{-1}(\mathrm{d}X)X^{-1}=X^{-1}$$

$$\Rightarrow (X+\mathrm{d}X)^{-1}-X^{-1}=-(X+\mathrm{d}X)^{-1}(\mathrm{d}X)X^{-1}$$

$$\Rightarrow a^T(X+\mathrm{d}X)^{-1}b-a^TX^{-1}b=-a^T(X+\mathrm{d}X)^{-1}(\mathrm{d}X)X^{-1}b$$

$$\Rightarrow \frac{a^T(X+\mathrm{d}X)^{-1}b-a^TX^{-1}b}{\mathrm{d}X}=\frac{-a^T(X+\mathrm{d}X)^{-1}(\mathrm{d}X)X^{-1}b}{\mathrm{d}X}$$

$$\because \frac{-a^T(X+\mathrm{d}X)^{-1}(\mathrm{d}X)X^{-1}b}{\mathrm{d}X}=\frac{-a^TX^{-1}(\mathrm{d}X)X^{-1}b+o(\|\mathrm{d}X\|)^2}{\mathrm{d}X}$$

(分子二阶微分可以略去)

$$\therefore \frac{a^T(X+\mathrm{d}X)^{-1}b-a^TX^{-1}b}{\mathrm{d}X}=\frac{-a^TX^{-1}(\mathrm{d}X)X^{-1}b}{\mathrm{d}X}$$

$$=-(a^TX^{-1})^T(X^{-1}b)^T$$

$$=-(X^{-1})^Tab^T(X^{-1})^T$$

例5 $\frac{\mathrm{d}\mathbf{a}^{\mathrm{T}}\mathbf{X}^{-1}\mathbf{b}}{\mathrm{d}\mathbf{X}}=-\mathbf{X}^{-\mathrm{T}}\mathbf{a}\mathbf{b}^{\mathrm{T}}\mathbf{X}^{-\mathrm{T}}$的MATLAB程序验证

```
a=rand(1,5)';
b=rand(1,5)';
X=rand(5,5);
X1=inv(X);
delta=0.001;
Z=zeros(5,5);
for ii=1:5
   for jj=1:5
      XX=X;
      XX(ii,jj)=XX(ii,jj)+delta;
      X2=inv(XX);
      f=a'*X2*b-a'*X1*b;
      Z(ii,jj)=f/delta;
   end
end
disp(Z);
disp(-X1'*a*b'*X1');
disp(Z+X1'*a*b'*X1');
```

3.5398	-6.7683	2.9917	-1.6593	-0.7051
2.1314	-4.0853	1.8036	-1.0002	-0.4259
-8.3936	16.2730	-7.0900	3.9679	1.6876
-1.5598	2.9905	-1.3161	0.7315	0.3121
9.5195	-18.0554	8.0547	-4.4498	-1.8892

3.5355	-6.7791	2.9879	-1.6603	-0.7064
2.1317	-4.0874	1.8015	-1.0011	-0.4259
-8.4308	16.1656	-7.1250	3.9592	1.6845
-1.5591	2.9894	-1.3176	0.7322	0.3115
9.4840	-18.1850	8.0150	-4.4538	-1.8949

0.0043	0.0108	0.0038	0.0011	0.0013
-0.0004	0.0021	0.0021	0.0009	-0.0000
0.0372	**0.1074**	0.0350	0.0087	0.0032
-0.0008	0.0011	0.0015	-0.0006	0.0006
0.0355	**0.1296**	0.0397	0.0040	0.0057

✓矩阵迹的微分

13) $\frac{\mathrm{dtr}(\mathbf{X})}{\mathrm{d}\mathbf{X}} = \mathbf{I}$

14) $\frac{\mathrm{dtr}(\mathbf{X}^{-1})}{\mathrm{d}\mathbf{X}} = -(\mathbf{X}^{-2})^{\mathrm{T}}$

15) $\frac{\mathrm{dtr}(\mathbf{AX})}{\mathrm{d}\mathbf{X}} = \frac{\mathrm{dtr}(\mathbf{XA})}{\mathrm{d}\mathbf{X}} = \mathbf{A}^{\mathrm{T}}$

16)

$$\frac{\mathrm{dtr}(\mathbf{AX}^{\mathrm{T}})}{\mathrm{d}\mathbf{X}} = \frac{\mathrm{dtr}(\mathbf{X}^{\mathrm{T}}\mathbf{A})}{\mathrm{d}\mathbf{X}} = \mathbf{A},$$

$$\frac{\mathrm{dtr}(\mathbf{ax}^{\mathrm{T}})}{\mathrm{d}\mathbf{x}} = \frac{\mathrm{dtr}(\mathbf{xa}^{\mathrm{T}})}{\mathrm{d}\mathbf{x}} = \mathbf{a}.$$

17) $\frac{\mathrm{dtr}(\mathbf{XX}^{\mathrm{T}})}{\mathrm{d}\mathbf{X}} = \frac{\mathrm{dtr}(\mathbf{X}^{\mathrm{T}}\mathbf{X})}{\mathrm{d}\mathbf{X}} = 2\mathbf{X}$

18) $\frac{\mathrm{dtr}(\mathbf{AX}^{-1})}{\mathrm{d}\mathbf{X}} = -(\mathbf{X}^{-1}\mathbf{AX}^{-1})^{\mathrm{T}}$

19) $\frac{\mathrm{dtr}(\mathbf{X}^{\mathrm{T}}\mathbf{AX})}{\mathrm{d}\mathbf{X}} = (\mathbf{A}+\mathbf{A}^{\mathrm{T}})\mathbf{X}$

$$\frac{\mathrm{d}f(\mathbf{x})}{\mathrm{d}\mathbf{x}} = \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \cdots, \frac{\partial f(\mathbf{x})}{\partial x_m}\right]^{\mathrm{T}}$$

$$\frac{\mathrm{d}f(\mathbf{x})}{\mathrm{d}\mathbf{x}^{\mathrm{T}}} = \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \cdots, \frac{\partial f(\mathbf{x})}{\partial x_m}\right]$$

$$\frac{\mathrm{d}f(\mathbf{X})}{\mathrm{d}\mathbf{X}} = \left(\frac{\partial f(\mathbf{X})}{\partial x_{ij}}\right)_{m\times n} = \begin{bmatrix} \frac{\partial f(\mathbf{X})}{\partial x_{11}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial x_{m1}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{mn}} \end{bmatrix}_{m\times n}$$

$\frac{\mathrm{dtr}(\mathbf{X}^{-1})}{\mathrm{d}\mathbf{X}} = -\mathbf{X}^{-2\mathrm{T}}$的证明

注意到

$$(X+\mathrm{d}X)^{-1} - X^{-1} = -(X+\mathrm{d}X)^{-1}(\mathrm{d}X)X^{-1}$$

所以

$$\mathrm{d}trace(X^{-1}) = trace((X+\mathrm{d}X)^{-1}) - trace(X^{-1})$$
$$= trace((X+\mathrm{d}X)^{-1} - X^{-1})$$
$$= trace(-(X+\mathrm{d}X)^{-1}(\mathrm{d}X)X^{-1})$$
$$= trace(-(X)^{-1}(\mathrm{d}X)X^{-1})$$
$$\therefore \mathrm{d}trace(X^{-1})/\mathrm{d}X = -(X^{-2})^{T}$$

$$\frac{\mathrm{d}f(\mathbf{x})}{\mathrm{d}\mathbf{x}} = \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \cdots, \frac{\partial f(\mathbf{x})}{\partial x_m}\right]^{\mathrm{T}}$$
$$\frac{\mathrm{d}f(\mathbf{x})}{\mathrm{d}\mathbf{x}^{\mathrm{T}}} = \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \cdots, \frac{\partial f(\mathbf{x})}{\partial x_m}\right]$$

$$\frac{\mathrm{d}f(\mathbf{X})}{\mathrm{d}\mathbf{X}} = \left(\frac{\partial f(\mathbf{X})}{\partial x_{ij}}\right)_{m\times n} = \begin{bmatrix} \frac{\partial f(\mathbf{X})}{\partial x_{11}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{1\mathrm{n}}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial x_{m1}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{mn}} \end{bmatrix}_{m\times n}$$

$\frac{\mathbf{d}\,trace(AXB)}{\mathbf{d}X}$的公式推导和验证有三种技术路线：

（1）基于矩阵代数运算化简该公式，利用已知公式的结果，如

$trace(AXB) = trace(BAX) = trace(CX), (C = BA)$,利用公式

$\frac{\mathbf{d}\,trace(CX)}{\mathbf{d}X} = C^T$,即可.

(2)基于函数的矩阵元素表达式进行推导

（这个是近乎万能的方法,参见“张量代数”）

$$trace(AXB) = trace((a_{ik})(x_{kl})(b_{lj})) = trace(\sum_{k=1}^{n}\sum_{l=1}^{n} a_{ik}x_{kl}b_{lj})$$

$$= \sum_{w=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} a_{wk}x_{kl}b_{lw}$$

$$\left(\frac{\Delta trace(AXB)}{\Delta x_{ij}}\right) = \left(\frac{\sum_{w=1}^{n} a_{wi}\Delta x_{ij}b_{jw}}{\Delta x_{ij}}\right) = \left(\sum_{w=1}^{n} a_{wi}b_{jw}\right) = \left(\sum_{w=1}^{n} b_{jw}a_{wi}\right) = A^T B^T$$

（3）编写程序进行数值验证和辨识

例6 $\frac{\mathrm{d}\,\mathrm{tr}(\mathbf{X}^{-1})}{\mathrm{d}\mathbf{X}} = -\mathbf{X}^{-2T}$的MATLAB验证

```
a=rand(1,5)';
b=rand(1,5)';
X=rand(5,5);
X1=inv(X);
delta=0.001;
Z=zeros(5,5);
for ii=1:5
    for jj=1:5
        XX=X;
        XX(ii,jj)=XX(ii,jj)+delta;
        X2=inv(XX);
        f=trace(X2)-trace(X1);
        Z(ii,jj)=f/delta;
    end
end
disp(Z);
disp(-(X1*X1)');
disp(Z+(X1*X1)');
```

```
-1.0027   0.2716   1.8452   0.0499  -1.5638
-0.2608  -0.5639  -2.5756   0.5131   3.1805
-0.6624   3.1740   1.0724  -1.2269  -2.7109
-0.1048   2.5211   3.7267  -3.3684  -3.5344
 1.8196  -4.5757  -2.4091   3.0176   2.2801

-1.0025   0.2714   1.8455   0.0499  -1.5655
-0.2606  -0.5648  -2.5784   0.5129   3.1769
-0.6632   3.1705   1.0725  -1.2271  -2.7108
-0.1048   2.5177   3.7229  -3.3747  -3.5370
 1.8197  -4.5826  -2.4079   3.0146   2.2809

-0.0001   0.0003  -0.0003  -0.0000   0.0017
-0.0002   0.0008   0.0027   0.0002   0.0037
 0.0008   0.0035  -0.0001   0.0002  -0.0001
 0.0000   0.0034   0.0038   0.0063   0.0026
-0.0001   0.0069  -0.0011   0.0030  -0.0008
```

✓矩阵行列式的微分

$$\frac{\mathrm{d}f(\mathbf{x})}{\mathrm{d}\mathbf{x}} = \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \cdots, \frac{\partial f(\mathbf{x})}{\partial x_m}\right]^{\mathrm{T}}$$
$$\frac{\mathrm{d}f(\mathbf{x})}{\mathrm{d}\mathbf{x}^{\mathrm{T}}} = \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \cdots, \frac{\partial f(\mathbf{x})}{\partial x_m}\right]$$

20) $$\frac{\mathrm{d}|\mathbf{X}|}{\mathrm{d}X} = |\mathbf{X}|\mathbf{X}^{-\mathrm{T}}$$

21) $$\frac{\mathrm{d}|\mathbf{X}^{-1}|}{\mathrm{d}\mathbf{X}} = -|\mathbf{X}|^{-1}\mathbf{X}^{-\mathrm{T}}$$

22) $$\frac{\mathrm{d}\log|\mathbf{X}|}{\mathrm{d}\mathbf{X}} = \frac{1}{|\mathbf{X}|}\frac{\mathrm{d}|\mathbf{X}|}{\mathrm{d}\mathbf{X}}$$

23) $$\frac{\mathrm{d}|\mathbf{X}^{\mathrm{T}}\mathbf{X}|}{\mathrm{d}\mathbf{X}} = 2|\mathbf{X}^{\mathrm{T}}\mathbf{X}|\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} \quad (\mathrm{rank}(\mathbf{X}) = n)$$
$$\frac{\mathrm{d}|\mathbf{X}\mathbf{X}^{\mathrm{T}}|}{\mathrm{d}\mathbf{X}} = 2|\mathbf{X}\mathbf{X}^{\mathrm{T}}|(\mathbf{X}\mathbf{X}^{\mathrm{T}})^{-1}\mathbf{X} \quad (\mathrm{rank}(\mathbf{X}) = m)$$

24) $$\frac{\mathrm{d}|\mathbf{X}^{\mathrm{T}}\mathbf{A}\mathbf{X}|}{\mathrm{d}\mathbf{X}} = |\mathbf{X}^{\mathrm{T}}\mathbf{A}\mathbf{X}| \times \left[\mathbf{A}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{A}\mathbf{X})^{-1} + \mathbf{A}^{\mathrm{T}}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{A}\mathbf{X})^{-\mathrm{T}}\right]$$

25) $$\frac{\mathrm{d}|\mathbf{X}\mathbf{A}\mathbf{X}^{\mathrm{T}}|}{\mathrm{d}\mathbf{X}} = |\mathbf{X}\mathbf{A}\mathbf{X}^{\mathrm{T}}| \times \left[(\mathbf{X}\mathbf{A}\mathbf{X}^{\mathrm{T}})^{-\mathrm{T}}\mathbf{X}\mathbf{A}^{\mathrm{T}} + (\mathbf{X}\mathbf{A}\mathbf{X}^{\mathrm{T}})^{-1}\mathbf{X}\mathbf{A}\right]$$

公式20)的推导：

$\Delta \det(A) = \det(A + \Delta A) - \det(A)$

令$I(i, j)$为只有第i行第j列元素不为零的、非零元素等于1的矩阵，

$\Delta A = \Delta a_{ij} I(i, j)$,则

$\det(A + \Delta A) - \det(A) = \det(A + \Delta a_{ij} I(i, j)) - \det(A) = \Delta a_{ij} A_{ij}$

其中, A_{ij}是A的第i行第j列的代数余子式,也就是$\Delta \det(A) / \Delta a_{ij} = A_{ij}$,

因为$A^{-1} = (A_{ji}) / \det(A)$(这是方阵逆的定义),所以

$$\frac{\mathbf{d} \det(A)}{\mathbf{d} A} = (\Delta \det(A) / \Delta a_{ij}) = \det(A)(A^{-1})^T$$

(3)实函数矩阵对矩阵变元的导数

多对多！

若$\mathbf{F}(\mathbf{X}) \in \mathbb{R}^{p\times q}$是以矩阵$\mathbf{X} \in \mathbb{R}^{m\times n}$为变元的实值函数矩阵，则

$$\frac{\mathrm{d}\mathbf{F}(\mathbf{X})}{\mathrm{d}\mathbf{X}} = \left(\frac{\partial \mathbf{F}(\mathbf{X})}{\partial x_{ij}}\right)_{m\times n} = \begin{bmatrix} \frac{\partial \mathbf{F}(\mathbf{X})}{\partial x_{11}} & \cdots & \frac{\partial \mathbf{F}(\mathbf{X})}{\partial x_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{F}(\mathbf{X})}{\partial x_{m1}} & \cdots & \frac{\partial \mathbf{F}(\mathbf{X})}{\partial x_{mn}} \end{bmatrix}_{\mathrm{mp}\times nq}$$

多对一！

其中

$$\frac{\partial \mathbf{F}(\mathbf{X})}{\partial x_{ij}} = \begin{bmatrix} \frac{\partial f_{11}(\mathbf{X})}{\partial x_{ij}} & \cdots & \frac{\partial f_{1q}(\mathbf{X})}{\partial x_{ij}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{p1}(\mathbf{X})}{\partial x_{ij}} & \cdots & \frac{\partial f_{pq}(\mathbf{X})}{\partial x_{ij}} \end{bmatrix}_{p\times q} \quad \begin{matrix} i = 1,2,\cdots,m \\ j = 1,2,\cdots,n \end{matrix}$$

一对多！

自变量是矩阵，应变量是由实函数组成的矩阵

例

$$\text{已知}X=\begin{pmatrix}x_{11} & x_{12}\\ x_{21} & x_{22}\end{pmatrix},\text{求}\frac{\mathrm{d}X}{\mathrm{d}X},\frac{\mathrm{d}(X^TX)}{\mathrm{d}X}$$

$$\text{解}:\frac{\mathrm{d}X}{\mathrm{d}X}=\left(\frac{\partial X}{\partial x_{ij}}\right)_{2\times2(\text{块})}=\begin{pmatrix}1&0&0&1\\0&0&0&0\\0&0&0&0\\1&0&0&1\end{pmatrix}_{4\times4(\text{元素})}$$

$$\frac{\mathrm{d}(X^TX)}{\mathrm{d}X}=\frac{\mathrm{d}(X^T)X+X^T\cdot\mathrm{d}X}{\mathrm{d}X}=\frac{\mathrm{d}(X^T)X}{\mathrm{d}X}+\frac{X^T\mathrm{d}X}{\mathrm{d}X}$$ **（千万别约分！）**

$$=\left(\frac{\partial(X^T)X}{\partial x_{ij}}\right)_{2\times2(\text{块})}+\left(\frac{X^T\partial X}{\partial x_{ij}}\right)_{2\times2(\text{块})}$$

$$=\begin{pmatrix}\begin{pmatrix}1&0\\0&0\end{pmatrix}X & \begin{pmatrix}0&0\\1&0\end{pmatrix}X\\ \begin{pmatrix}0&1\\0&0\end{pmatrix}X & \begin{pmatrix}0&0\\0&1\end{pmatrix}X\end{pmatrix}_{4\times4(\text{元素})}+\begin{pmatrix}X^T\begin{pmatrix}1&0\\0&0\end{pmatrix} & X^T\begin{pmatrix}0&1\\0&0\end{pmatrix}\\ X^T\begin{pmatrix}0&0\\1&0\end{pmatrix} & X^T\begin{pmatrix}0&0\\0&1\end{pmatrix}\end{pmatrix}_{4\times4(\text{元素})}=\begin{pmatrix}2x_{11}&x_{12}&0&x_{11}\\x_{12}&0&x_{11}&2x_{12}\\2x_{21}&x_{22}&0&x_{21}\\x_{22}&0&x_{21}&2x_{22}\end{pmatrix}$$

例

已知$X=\begin{pmatrix}x_1 & x_2 & x_3 & x_4\end{pmatrix}$，求$\dfrac{\mathrm{d}X}{\mathrm{d}X},\dfrac{\mathrm{d}X}{\mathrm{d}X^T},\dfrac{\mathrm{d}(XX^T)}{\mathrm{d}X}$

解：$\dfrac{\mathrm{d}X}{\mathrm{d}X}=\left(\dfrac{\partial X}{\partial x_i}\right)_{1\times4(\text{块})}=\begin{pmatrix}1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1\end{pmatrix}_{1\times16(\text{元素})}$

$$\frac{\mathrm{d}X}{\mathrm{d}X^T}=\left(\frac{\partial X}{\partial x_i}\right)_{4\times1(\text{块})}=\begin{pmatrix}1 & 0 & 0 & 0\\0 & 1 & 0 & 0\\0 & 0 & 1 & 0\\0 & 0 & 0 & 1\end{pmatrix}_{4\times4(\text{元素})}$$

$$\frac{\mathrm{d}(XX^T)}{\mathrm{d}X}=\frac{(\mathrm{d}X)X^T+X\mathrm{d}X^T}{\mathrm{d}X}=\frac{\mathrm{d}(X)X^T}{\mathrm{d}X}+\frac{X\mathrm{d}X^T}{\mathrm{d}X}$$

（千万不要约分！）

$$=\left(\frac{\partial(X)X^T}{\partial x_i}\right)_{1\times4(\text{块})}+\left(\frac{X\partial X^T}{\partial x_i}\right)_{1\times4(\text{块})}$$

$$=\begin{pmatrix}x_1, & x_2, & x_3, & x_4\end{pmatrix}_{1\times4(\text{元素})}+\begin{pmatrix}x_1, & x_2, & x_3, & x_4\end{pmatrix}_{1\times4(\text{元素})}=2\begin{pmatrix}x_1, & x_2, & x_3, & x_4\end{pmatrix}$$

(4)梯度矩阵、Jacobian矩阵与Hessian矩阵

定义：梯度向量和梯度矩阵

$$\nabla_{\mathrm{x}} f(\mathbf{x}) \stackrel{\text{def}}{=} \frac{\mathrm{d} f(\mathbf{x})}{\mathrm{d}\mathbf{x}} = \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \cdots, \frac{\partial f(\mathbf{x})}{\partial x_m}\right]^{\mathrm{T}}$$

$$\nabla_{\mathrm{X}} f(\mathbf{X}) \stackrel{\text{def}}{=} \frac{\mathrm{d} f(\mathbf{X})}{\mathrm{d}\mathbf{X}} = \begin{bmatrix} \frac{\partial f(\mathbf{X})}{\partial x_{11}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial x_{m1}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{mn}} \end{bmatrix}_{\mathbf{m}\times n}$$

$$\mathrm{vec}(\nabla_{\mathrm{X}} f(\mathbf{X})) = \frac{\mathrm{d} f(\mathbf{X})}{\mathrm{d}\,\mathrm{vec}(\mathbf{X})}$$

$$= \left[\frac{\partial f(\mathbf{X})}{\partial x_{11}}, \cdots, \frac{\partial f(\mathbf{X})}{\partial x_{m1}}, \cdots, \frac{\partial f(\mathbf{X})}{\partial x_{1n}}, \cdots, \frac{\partial f(\mathbf{X})}{\partial x_{mn}}\right]^{\mathrm{T}}$$

函数梯度方向取反所得向量(矩阵)-$\nabla_{\mathrm{x}} f(\mathrm{x})$称为函数在点X处的**梯度流.**

函数梯度方向取反所得向量(矩阵)-$\nabla_{\mathrm{x}} f(\mathrm{x})$称为函数在点X处的**梯度流.**

矩阵的向量化

vec(A):按列堆栈

rvec(A):按行堆栈

$\mathrm{unvec}_{m,n}$(X):按列矩阵化

$\mathrm{unrvec}_{m,n}$(X):按行矩阵化

$$rvec(A) = \left(vec(A^T)\right)^T$$

$$K_{mn} vec(A) = vec(AT)$$

交换矩阵:

$$K_{mn} = \sum_{j=1}^{n} (e_j^T \otimes I_m \otimes e_j)$$

Hadamard积(对应元素乘积):*

Kronecker积: $\otimes$

内积: <X,Y>

$$f(X + \Delta X) = f(X) + < \nabla_{\mathbf{X}} f(\mathbf{X}), \Delta X > + o(||\Delta X||)$$

令$\alpha > 0$, $\Delta X = -\alpha \nabla_{\mathrm{x}} f(\mathrm{x})$，得

$$f(X + \Delta X) = f(X) - \alpha||\nabla_{\mathbf{X}} f(\mathbf{X})|| + o(||\Delta X||)$$

在机器学习领域，α也称为学习率.

```
f=@(x) x+sin(x);
df=@(x) 1+cos(x);
x=1;
alpha=1.95;
for ii=1:100
    dx=-alpha*df(x);
    x=x+dx;
end
disp(x);
disp(f(x));
disp(df(x));
```

alpha=1.95;
-3.1378
-3.1416
7.0159e-06

alpha=0.05;
-2.7065
-3.1280
0.0932

定义：协梯度向量和Jacobian矩阵（协梯度矩阵）：

$$\mathrm{D}_{\mathbf{x}} f(\mathbf{x}) \stackrel{\text{def}}{=} \frac{\mathrm{d}f(\mathbf{x})}{\mathrm{d}\mathbf{x}^{\mathrm{T}}} = \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \cdots, \frac{\partial f(\mathbf{x})}{\partial x_m}\right]$$

$$D_X f(\mathbf{X}) \stackrel{\text{def}}{=} \frac{\mathrm{d}f(\mathbf{X})}{\mathrm{d}\mathbf{X}^{\mathrm{T}}} = \begin{bmatrix} \frac{\partial f(\mathbf{X})}{\partial x_{11}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{m1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial x_{1n}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{mn}} \end{bmatrix}_{n\times m}$$

Covariant form of the gradient vector

Cogradient vector

命题1：给定是指标量函数$f(\mathbf{X})$,其中$\mathbf{X} \in \mathbb{R}^{m\times n}$，则

$$\mathrm{rvec}(\mathrm{D}_X f(\mathbf{X})) = \mathrm{D}_{\mathrm{vec}(X)} f(\mathbf{X})$$

或

$$\mathrm{D}_{\mathbf{X}} f(\mathbf{X}) = \mathrm{unrvec}\left(\mathrm{D}_{\mathrm{vec}(\mathbf{X})} f(\mathbf{X})\right)$$

行向量化
列向量化

命题2： $\nabla_{\mathbf{X}} f(\mathbf{X}) = \mathrm{D}_{\mathbf{X}}^{\mathrm{T}} \mathbf{f}(\mathbf{X})$

命题3： 给定$f(\mathbf{X})$,其中$\mathbf{X} \in \mathbb{R}^{m\times n}$.若已求出$\mathrm{D}_{\mathrm{vec}(x)} f(\mathbf{X})$,则

$$\nabla_X f(\mathbf{X}) = \ \mathrm{unvec}\left(\mathrm{D}_{\mathrm{vec}(\mathbf{X})}^{\mathrm{T}} f(\mathbf{X})\right) \qquad (1)$$

$+$	+	$-$	-
$\pm$	\pm	$\mp$	\mp
$\cdot$	\cdot	$\div$	\div
$\times$	\times	$\setminus$	\setminus
$\cup$	\cup	$\cap$	\cap
$\sqcup$	\sqcup	$\sqcap$	\sqcap
$\vee$	\vee , \lor	$\wedge$	\wedge , \land
$\oplus$	\oplus	$\ominus$	\ominus
$\odot$	\odot	$\oslash$	\oslash
$\otimes$	\otimes	$\bigcirc$	\bigcirc
$\bigtriangleup$	\bigtriangleup	$\bigtriangledown$	\bigtriangledown
$\lhd$	\lhd [a]	$\rhd$	\rhd [a]
$\unlhd$	\unlhd [a]	$\unrhd$	\unrhd [a]

换言之，若

$$\mathrm{D}_{\mathrm{vec}(\mathbf{X})} f(\mathbf{X}) = [d_1, d_2, \cdots, d_{mn}]$$

则

$$[\nabla_{\mathbf{X}} f(\mathbf{X})]_{i,j} = d_{i+(j-1)n} \quad \begin{cases} i = 1, \cdots, m \\ j = 1, \cdots n \end{cases} \qquad (2)$$

定义：多元向量值函数的Jacobian矩阵或协梯度矩阵：

$$D_{\mathbf{x}} f(\mathbf{x}) = \frac{\mathrm{d}\mathbf{f}(\mathbf{x})}{\mathrm{d}\mathbf{x}^{\mathrm{T}}} = \begin{bmatrix} \frac{\mathrm{d}f_1(\mathbf{x})}{\mathrm{d}\mathbf{x}^{\mathrm{T}}} \\ \vdots \\ \frac{\mathrm{d}f_p(\mathbf{x})}{\mathrm{d}\mathbf{x}^{\mathrm{T}}} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_p(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_p(\mathbf{x})}{\partial x_m} \end{bmatrix}_{p\times m}$$

◆ 实值矩阵函数情形(应变量是矩阵)

$$\mathbf{F}(\mathbf{X}) = [f_{kl}]_{k=1,l=1}^{p,q} \in \mathbb{R}^{p\times q} \quad (\text{其中}\mathbf{X} \in \mathbb{R}^{m\times n})$$

$$\mathbf{f}(\mathbf{X}) \overset{\text{def}}{=} \mathrm{vec}(\mathbf{F}(\mathbf{X})) \in \mathbb{R}^{pq\times 1}$$

$$= \left[f_{11}(\mathbf{X}), \cdots, f_{p1}(\mathbf{X}), \cdots, f_{1q}(\mathbf{X}), \cdots, f_{pq}(\mathbf{X})\right]^{\mathrm{T}}$$

列优先！

矩阵函数$\mathbf{F}(\mathbf{X})$的<u>行向量偏导</u>

$$\mathrm{D}_{\mathrm{vec}(\mathbf{X})}\mathbf{F}(\mathbf{X}) \overset{\mathrm{def}}{=} \frac{\mathrm{d}\mathbf{f}(\mathbf{X})}{\mathrm{dvec}^{\mathrm{T}}(\mathbf{X})} = \frac{\mathrm{dvec}\big(\mathbf{F}(\mathbf{X})\big)}{\mathrm{dvec}^{\mathrm{T}}(\mathbf{X})} \in \mathbb{R}^{pq\times mn}$$

$$= \left[\frac{\mathrm{d}f_{11}}{\mathrm{dvec}^{\mathrm{T}}(\mathbf{x})}, \cdots, \frac{\mathrm{d}f_{p1}}{\mathrm{dvec}^{\mathrm{T}}(\mathbf{x})}, \cdots, \frac{\mathrm{d}f_{1q}}{\mathrm{dvec}^{\mathrm{T}}(\mathbf{x})}, \cdots, \frac{\mathrm{d}f_{pq}}{\mathrm{dvec}^{\mathrm{T}}(\mathbf{x})}\right]^{\mathrm{T}}$$

$$= \begin{bmatrix} \frac{\partial f_{11}}{\partial x_{11}} & \cdots & \frac{\partial f_{11}}{\partial x_{m1}} & \cdots & \frac{\partial f_{11}}{\partial x_{1n}} & \cdots & \frac{\partial f_{11}}{\partial x_{mn}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_{p1}}{\partial x_{11}} & \cdots & \frac{\partial f_{p1}}{\partial x_{m1}} & \cdots & \frac{\partial f_{p1}}{\partial x_{1n}} & \cdots & \frac{\partial f_{p1}}{\partial x_{mn}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_{1q}}{\partial x_{11}} & \cdots & \frac{\partial f_{1q}}{\partial x_{m1}} & \cdots & \frac{\partial f_{1q}}{\partial x_{1n}} & \cdots & \frac{\partial f_{1q}}{\partial x_{mn}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_{pq}}{\partial x_{11}} & \cdots & \frac{\partial f_{pq}}{\partial x_{m1}} & \cdots & \frac{\partial f_{pq}}{\partial x_{1n}} & \cdots & \frac{\partial f_{pq}}{\partial x_{mn}} \end{bmatrix}_{pq\times mn}$$

定义：标量函数的**Hessian矩阵(<u>自变量是向量</u>)**

$$\frac{\mathrm{d}^2 f(\mathbf{x})}{\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}^{\mathrm{T}}} = \frac{\mathrm{d}}{\mathrm{d}\mathbf{x}^{\mathrm{T}}}\left[\frac{\mathrm{d}f(\mathbf{x})}{\mathrm{d}\mathbf{x}}\right] = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_m} \\ \vdots & & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_m \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_m \partial x_m} \end{bmatrix}_{m\times m}$$

◆ 记$\nabla_{\mathbf{x}}^2 f(\mathbf{x}) = \mathrm{D}_{\mathbf{x}}(\nabla_{\mathbf{x}} f(\mathbf{x})) = \nabla_{\mathbf{x}^{\mathrm{T}}}(\nabla_{\mathbf{x}} f(\mathbf{x}))$

◆ 实值**标量函数**$f(\mathbf{X})$的Hessian矩阵（<u>**自变量:向量,应变量:标量**</u>）

$$\frac{\mathrm{d}^2 f(\mathbf{X})}{\mathrm{d}\mathbf{X}\mathrm{d}\mathbf{X}^{\mathrm{T}}} = \frac{\mathrm{d}}{\mathrm{d}\mathbf{X}^{\mathrm{T}}}\left[\frac{\mathrm{d}f(\mathbf{X})}{\mathrm{d}\mathbf{X}}\right]$$

对称矩阵！

或记作$\nabla_{X}^2 f(\mathbf{X}) = \mathrm{D}_{\mathbf{X}}(\nabla_{\mathbf{X}} f(\mathbf{X})) = \nabla_{\mathbf{X}^{T}}(\nabla_{\mathbf{X}} f(\mathbf{X}))$

注意字母的小写和大写！

(5)实值标量函数的矩阵微分及计算

A.实值函数的矩阵微分 $\boldsymbol{f}(\boldsymbol{X}+\boldsymbol{dX})=\boldsymbol{f}(\boldsymbol{X})+\mathbf{d}\boldsymbol{f}(\boldsymbol{X})$

◆ 全微分

$$\mathrm{d}f(\mathbf{x})=\frac{\partial f(\mathbf{x})}{\partial x_1}\mathrm{d}x_1+\cdots+\frac{\partial f(\mathbf{x})}{\partial x_m}\mathrm{d}x_m$$

$$=\left[\frac{\partial f(\mathbf{x})}{\partial x_1},\cdots,\frac{\partial f(\mathbf{x})}{\partial x_m}\right]\begin{bmatrix}\mathrm{d}x_1\\ \vdots\\ \mathrm{d}x_m\end{bmatrix}$$

$$=\frac{\mathrm{d}f(\mathbf{x})}{\mathrm{d}\mathbf{x}^{\mathrm{T}}}\mathrm{d}\mathbf{x}$$

函数的微分矩阵: $\mathbf{d}\boldsymbol{f}(\boldsymbol{X})$

$$
\begin{aligned}
\mathrm{d}f(\mathbf{X}) &= \frac{\partial f(\mathbf{X})}{\partial \mathbf{x}_1}\mathrm{d}\mathbf{x}_1 + \cdots + \frac{\partial f(\mathbf{X})}{\partial \mathbf{x}_n}\mathrm{d}\mathbf{x}_n \\
&= \left[\frac{\partial f(\mathbf{X})}{\partial x_{11}}, \cdots, \frac{\partial f(\mathbf{X})}{\partial x_{m1}}\right]\begin{bmatrix} \mathrm{d}x_{11} \\ \vdots \\ \mathrm{d}x_{m1} \end{bmatrix} + \dots + \left[\frac{\partial f(\mathbf{X})}{\partial x_{1n}}, \cdots, \frac{\partial f(\mathbf{X})}{\partial x_{mn}}\right]\begin{bmatrix} \mathrm{d}x_{1n} \\ \vdots \\ \mathrm{d}x_{mn} \end{bmatrix} \\
&= \left[\frac{\partial f(\mathbf{X})}{\partial x_{11}}, \cdots, \frac{\partial f(\mathbf{X})}{\partial x_{m1}}, \cdots, \frac{\partial f(\mathbf{X})}{\partial x_{1n}}, \cdots, \frac{\partial f(\mathbf{X})}{\partial x_{mn}}\right]\begin{bmatrix} \mathrm{d}x_{11} \\ \vdots \\ \mathrm{d}x_{m1} \\ \vdots \\ \mathrm{d}x_{1n} \\ \vdots \\ \mathrm{d}x_{mn} \end{bmatrix} \\
&= \mathrm{rvec}(\mathbf{A})\mathrm{vec}(\mathrm{d}\mathbf{X})
\end{aligned}
$$

其中

$$\mathbf{A} = \mathrm{D}_{\mathbf{x}} f(\mathbf{X}) = \frac{\mathrm{d}f(\mathbf{X})}{\mathrm{d}\mathbf{X}^{\mathrm{T}}} = \begin{bmatrix} \frac{\partial f(\mathbf{X})}{\partial x_{11}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{m1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial x_{1n}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{mn}} \end{bmatrix}$$

且

$$\mathrm{d}\mathbf{X} = \begin{bmatrix} \mathrm{d}x_{11} & \cdots & \mathrm{d}x_{1n} \\ \vdots & \ddots & \vdots \\ \mathrm{d}x_{m1} & \cdots & \mathrm{d}x_{mn} \end{bmatrix}$$

进一步有

$$\mathrm{d}f(\mathbf{X}) = \left(\mathrm{vec}(\mathbf{A}^{\mathrm{T}})\right)^{\mathrm{T}} \mathrm{vec}(\mathrm{d}\mathbf{X})$$

即

$$\boldsymbol{df(X) = tr(AdX)}$$

用此可以推导出很多公式！

$$\boldsymbol{f(X+dX)=f(X)+\mathrm{d}f(X)=f(X)+tr(AdX)}$$

命题4： 一阶偏导矩阵A是唯一确定的.即，若存在A_1和A_2满足

$$\mathrm{d}f(\mathbf{X})=tr(\mathbf{A}_i\mathrm{d}\mathbf{X}),\quad i=1,2$$

则$\mathbf{A}_1=\mathbf{A}_2$.

命题5： 若实值标量函数$f(\mathbf{X})$在$\mathbf{X}$处可微分，则　　　辨识定理

$$\mathrm{d}f(\mathbf{X})=\mathrm{tr}(\mathbf{A}\mathrm{d}\mathbf{X})\quad\Leftrightarrow\quad\nabla_{\mathbf{x}}f(\mathbf{X})=\frac{\mathrm{d}f(\mathbf{X})}{\mathrm{d}\mathbf{X}}=\mathbf{A}^{\mathrm{T}}$$

辨识：满足某些条件，必定是它！

B.实矩阵微分计算

<1> 一般规则

1） $\mathrm{d}(\mathbf{X}^{\mathrm{T}}) = (\mathrm{d}\mathbf{X})^{\mathrm{T}}$

2） $\mathrm{d}(\alpha\mathbf{X} + \beta\mathbf{Y}) = \alpha\mathrm{d}\mathbf{X} + \beta\mathrm{d}\mathbf{Y}$

例1.考虑标量函数$\mathrm{tr}(\mathbf{U})$的微分，得

$$\mathrm{d}(\mathrm{tr}\mathbf{U}) = \mathrm{d}\left(\sum_{i=1}^{n} u_{ii}\right) = \sum_{i=1}^{n} \mathrm{d}\, u_{ii} = \mathrm{tr}(\mathrm{d}\mathbf{U})$$

即有$\mathrm{d}(\mathrm{tr}\mathbf{U}) = \mathrm{tr}(\mathrm{d}\mathbf{U})$.

<2>实矩阵微分的常用计算公式

1) $\mathrm{d}A = 0$

2) $\mathrm{d}(\alpha X) = \alpha \mathrm{d}\mathrm{X}$

3) $\mathrm{d}(X^{\mathrm{T}}) = (\mathrm{d}X)^{\mathrm{T}})$

4) $\mathrm{d}(\mathbf{U} \pm \mathbf{V}) = \mathrm{d}\mathbf{U} \pm \mathrm{d}\mathbf{V}$

5) $\mathrm{d}(\mathbf{AXB}) = \mathbf{A}(\mathrm{d}\mathbf{X})\mathbf{B}$

6)

$$\mathrm{d}(\mathbf{UV}) = (\mathrm{d}\mathbf{U})\mathbf{V} + \mathbf{U}(\mathrm{d}\mathbf{V})$$

$$\mathrm{d}(\mathbf{UVW}) = (\mathrm{d}\mathbf{U})\mathbf{VW} + \mathbf{U}(\mathrm{d}\mathbf{V})\mathbf{W} + \mathbf{UV}(\mathrm{d}\mathbf{W})$$

特别地，若A为常数矩阵，则

$$\mathrm{d}(\mathbf{XAX}^{\mathrm{T}}) = (\mathrm{d}\mathbf{X})\mathbf{AX}^{\mathrm{T}} + \mathbf{XA}(\mathrm{d}\mathbf{X})^{\mathrm{T}}$$

和

$$\mathrm{d}(\mathbf{X}^{\mathrm{T}}\mathbf{AX}) = (\mathrm{d}\mathbf{X})^{\mathrm{T}}\mathbf{AX} + \mathbf{X}^{\mathrm{T}}\mathbf{A}\mathrm{d}\mathbf{X}$$

7) $\mathrm{d}(\mathbf{U}\otimes\mathbf{V})=(\mathrm{d}\mathbf{U})\otimes\mathbf{V}+\mathbf{U}\otimes\mathrm{d}\mathbf{V}$

8) $\mathrm{d}(\mathbf{U}\odot\mathbf{V})=(\mathrm{d}\mathbf{U})\odot\mathbf{V}+\mathbf{U}\odot\mathrm{d}\mathbf{V}$

9) $\mathrm{d}(\mathrm{vec}(\mathbf{X}))=\mathrm{vec}(\mathrm{d}\mathbf{X})$

10) $\mathrm{dlog}\mathbf{X}=\mathbf{X}^{-1}\mathrm{d}\mathbf{X},\quad \mathrm{dlog}(\mathbf{F}(\mathbf{X}))=(\mathbf{F}(\mathbf{X}))^{-1}\mathrm{d}(\mathbf{F}(\mathbf{X}))$

11) $\mathrm{d}|\mathbf{X}|=|\mathbf{X}|\mathrm{tr}(\mathbf{X}^{-1}\mathrm{d}\mathbf{X}),\quad \mathrm{d}|\mathbf{F}(\mathbf{X})|=|\mathbf{U}|\mathrm{tr}(\mathbf{U}^{-1}\mathrm{d}\mathbf{X})$

12) $\mathrm{d}(\mathrm{tr}(\mathbf{X}))=\mathrm{tr}(\mathrm{d}\mathbf{X}),\quad \mathrm{d}(\mathrm{tr}(\mathbf{F}(\mathbf{X})))=\mathrm{tr}(\mathrm{d}(\mathbf{F}(\mathbf{X})))$

13) $\mathrm{d}(\mathbf{X}^{-1})=-\mathbf{X}^{-1}(\mathrm{d}\mathbf{X})\mathbf{X}^{-1}$

14) $$\mathrm{d}\left(\mathbf{X}^{\dagger}\right)=-\mathbf{X}^{\dagger}(\mathrm{d}\mathbf{X})\mathbf{X}^{\dagger}+\mathbf{X}^{\dagger}\left(\mathbf{X}^{\dagger}\right)^{\mathrm{T}}(\mathrm{d}\mathbf{X}^{\mathrm{T}})\left(\mathbf{I}-\mathbf{X}\mathbf{X}^{\dagger}\right)+\left(\mathbf{I}-\mathbf{X}^{\dagger}\mathbf{X}\right)(\mathrm{d}\mathbf{X}^{\mathrm{T}})\left(\mathbf{X}^{\dagger}\right)^{\mathrm{T}}\mathbf{X}^{\dagger}$$

$$\mathrm{d}\left(\mathbf{X}^{\dagger}\mathbf{X}\right)=\mathbf{X}^{\dagger}(\mathrm{d}\mathbf{X})\left(\mathbf{I}-\mathbf{X}^{\dagger}\mathbf{X}\right)+\left(\mathbf{X}^{\dagger}(\mathrm{d}\mathbf{X})\left(\mathbf{I}-\mathbf{X}^{\dagger}\mathbf{X}\right)\right)^{\mathrm{T}}$$

$$\mathrm{d}\left(\mathbf{X}\mathbf{X}^{\dagger}\right)=\left(\mathbf{I}-\mathbf{X}\mathbf{X}^{\dagger}\right)(\mathrm{d}\mathbf{X})\mathbf{X}^{\dagger}+\left(\left(\mathbf{I}-\mathbf{X}\mathbf{X}^{\dagger}\right)(\mathrm{d}\mathbf{X})\mathbf{X}^{\dagger}\right)^{\mathrm{T}}$$

提示：可尝试用SVD分解来证明

C.利用矩阵微分计算梯度矩阵

$$\mathrm{d}f(\mathbf{X}) = \mathrm{tr}(\mathbf{A}\mathrm{d}\mathbf{X}) \quad \Leftrightarrow \quad \nabla_{\mathbf{X}} f(\mathbf{X}) = \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{A}^{\mathrm{T}}$$

- 一般标量函数的梯度矩阵

$$\frac{\mathrm{d}a^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{b}}{\mathrm{d}\mathbf{X}} = \mathbf{X}(\mathbf{a}\mathbf{b}^{\mathrm{T}} + \mathbf{b}\mathbf{a}^{\mathrm{T}})$$

$\mathbf{d}a^TX^TXb = \mathbf{d}tr(a^TX^TXb) = tr(\mathbf{d}a^TX^TXb) = tr(a^T\mathbf{d}(X^TX)b)$

$= tr(ba^T\mathbf{d}(X^TX)) = tr(ba^T\mathbf{d}(X^T)X)) + tr(ba^TX^T\mathbf{d}X))$

$= tr(Xba^T\mathbf{d}(X^T))) + tr(ba^TX^T\mathbf{d}X)) = tr(\mathbf{d}Xab^TX^T) + tr(ba^TX^T\mathbf{d}X))$

$= tr((ab^TX^T + ba^TX^T)\mathbf{d}X)$

- 迹函数的梯度矩阵

$$\frac{\mathrm{dtr}(\mathbf{X})}{\mathrm{d}\mathbf{X}} = \mathbf{I}$$

- 行列式的梯度矩阵

$$\frac{\mathrm{d}|\mathbf{X}|}{\mathrm{d}X} = |\mathbf{X}|\mathbf{X}^{-\mathrm{T}}, \qquad \mathbf{d|X|=tr(|X|X^{-1}dX)}$$

D.二阶实微分矩阵与实Hessian矩阵

令x, $\mathbf{x}$,$\mathbf{X}$分别代表函数的实标量变元、m×1实向量变元和$m\times n$实矩阵变元，而$f(\cdot)$，$\mathbf{f}(\cdot)$，$\mathbf{F}(\cdot)$分别表示实标量函数、$p\times 1$
实向量函数和$p\times q$实矩阵函数.

表3 二阶辨识表

实函数	二阶实微分矩阵	实Hessian矩阵H	H的维数
$f(x)$ $f(\mathbf{x})$ $f(\mathbf{X})$	$\mathrm{d}^2[f(x)]=\beta(\mathrm{d}x)^2$ $\mathrm{d}^2[f(x)]=(\mathrm{d}x)^{\mathrm{T}}B\mathrm{d}x$ $\mathrm{d}^2[f(X)]=(\mathrm{dvec}(X))^{\mathrm{T}}B\mathrm{d}(\mathrm{vec}(X))$	$\mathbf{H}[f(x)]=\beta$ $\mathbf{H}[f(\mathbf{x})]=\frac{1}{2}(\mathbf{B}+\mathbf{B}^{\mathrm{T}})$ $\mathbf{H}[f(\mathbf{X})]=\frac{1}{2}(\mathbf{B}+\mathbf{B}^{\mathrm{T}})$	1×1 $m\times m$ $mn\times mn$
$\mathbf{f}(x)$ $\mathbf{f}(\mathbf{x})$ $\mathbf{f}(\mathbf{X})$	$\mathrm{d}^2[f(x)]=b(\mathrm{d}x)^2$ $\mathrm{d}^2[f(x)]=(I_m\otimes\mathrm{d}x)^{\mathrm{T}}B\mathrm{d}x$ $\mathrm{d}^2[f(X)]=(I_m\otimes\mathrm{dvec}(X))^{\mathrm{T}}B\mathrm{d}(\mathrm{vec}(X))$	$\mathbf{H}[\mathbf{f}(x)]=\mathbf{b}$ $\mathbf{H}[\mathbf{f}(\mathbf{x})]=\frac{1}{2}\left[\mathbf{B}+(\mathbf{B}')_v\right]$ $\mathbf{H}[\mathbf{f}(\mathbf{X})]=\frac{1}{2}\left[\mathbf{B}+(\mathbf{B}')_v\right]$	$p\times 1$ $pm\times m$ $pmn\times mn$
$\mathbf{F}(x)$ $\mathbf{F}(\mathbf{x})$ $\mathbf{F}(\mathbf{X})$	$\mathrm{d}^2[\mathbf{F}(x)]=\mathbf{B}(\mathrm{d}x)^2$ $\mathrm{d}^2[\mathrm{vec}(\mathbf{F})]=\left(\mathbf{I}_{mp}\otimes\mathrm{d}\mathbf{x}\right)^{\mathrm{T}}\mathbf{B}\mathrm{d}\mathbf{x}$ $\mathrm{d}^2[\mathrm{vec}(\mathbf{F})]$ $=\left(\mathbf{I}_{mp}\otimes\mathrm{dvec}(\mathbf{X})\right)^{\mathrm{T}}\mathbf{B}\mathrm{d}(\mathrm{vec}(\mathbf{X}))$	$\mathbf{H}[\mathbf{F}(x)]=\mathrm{vec}(\mathbf{B})$ $\mathbf{H}[\mathbf{F}(\mathbf{x})]=\frac{1}{2}\left[\mathbf{B}+(\mathbf{B}')_v\right]$ $\mathbf{H}[\mathbf{F}'(\mathbf{x})]=\frac{1}{2}\left[\mathbf{B}+(\mathbf{B}')_v\right]$	$pq\times 1$ $pmq\times m$ $pmqn\times mn$

表中，对于实向量函数$\boldsymbol{f}$，

$$\mathbf{B}=\begin{bmatrix}\mathbf{B}_1\\ \mathbf{B}_2\\ \vdots\\ \mathbf{B}_p\end{bmatrix},\quad (\mathbf{B}')_v=\begin{bmatrix}\mathbf{B}_1^{\mathrm{T}}\\ \mathbf{B}_2^{\mathrm{T}}\\ \vdots\\ \mathbf{B}_p^{\mathrm{T}}\end{bmatrix}$$

而对于实矩阵函数$\mathbf{F}$，

$$\mathbf{B}=\begin{bmatrix}\mathbf{B}_{11}\\ \vdots\\ \mathbf{B}_{p1}\\ \vdots\\ \mathbf{B}_{1q}\\ \vdots\\ \mathbf{B}_{pq}\end{bmatrix}\quad (\mathbf{B}')_v=\begin{bmatrix}\mathbf{B}_{11}^{\mathrm{T}}\\ \vdots\\ \mathbf{B}_{p1}^{\mathrm{T}}\\ \vdots\\ \mathbf{B}_{1q}^{\mathrm{T}}\\ \vdots\\ \mathbf{B}_{pq}^{\mathrm{T}}\end{bmatrix}$$

定理：令$f(\mathbf{X})$是矩阵$\mathbf{X} \in \mathbb{R}^{m\times n}$的实值标量函数,并可二次微分,则

$$\mathrm{d}^2 f(\mathbf{X}) = \mathrm{tr}(\mathbf{B}(\mathrm{d}\mathbf{X})^{\mathrm{T}}\mathbf{C}\mathrm{d}\mathbf{X}) \Leftrightarrow \mathbf{H}(f(\mathbf{X})) = \frac{1}{2}(\mathbf{B}^{\mathrm{T}} \otimes \mathbf{C} + \mathbf{B} \otimes \mathbf{C}^{\mathrm{T}})$$

$$\mathrm{d}^2 f(\mathbf{X}) = \mathrm{tr}(\mathbf{B}(\mathrm{d}\mathbf{X})\mathbf{C}\mathrm{d}\mathbf{X}) \Leftrightarrow \mathbf{H}(f(\mathbf{X})) = \frac{1}{2}\mathbf{K}_{nm}(\mathbf{B}^{\mathrm{T}} \otimes \mathbf{C} + \mathbf{C}^{\mathrm{T}} \otimes \mathbf{B})$$

式中$\mathbf{K}_{nm}$为交换矩阵.

$$K_{mn} = \sum_{j=1}^{n}(e_j^T \otimes I_m \otimes e_j)$$

$\mathbf{d^2X}$, $\boldsymbol{d}(\mathbf{dX})$两者有区别!

$$\frac{\mathrm{d}\mathbf{a}^{\mathrm{T}}\mathbf{X}\mathbf{b}}{\mathrm{d}\mathbf{X}} = \mathbf{a}\mathbf{b}^{\mathrm{T}} \quad (\mathbf{X} \in \mathbb{R}^{m\times n})$$

$$f(X) = a^T X b$$

$$\mathbf{d}f(X) = \mathbf{d}(a^T X = b) = \mathbf{d}tr(a^T Xb) = tr(\mathbf{d}(a^T Xb)) = tr(a^T \mathbf{d}Xb) = tr(ba^T \mathbf{d}X)$$

$$\mathrm{D}_X f(X) = ba^T, \nabla_X f(X) = ab^T$$

$$\mathbf{d}^2 f(X) = \mathbf{d}(tr(ba^T \mathbf{d}X)) = tr(\mathbf{d}(ba^T \mathbf{d}X)) = tr(\mathbf{d}(ba^T)\mathbf{d}X)) + tr(ba^T \mathbf{d}(\mathbf{d}X))) = 0$$

$$H(f(X)) = 0$$

$$f(X) = a^T X^2 b$$

$$\mathbf{d}f(X) = tr(a^T \mathbf{d}(X^2)b) = tr(ba^T \mathbf{d}X^2) = tr(ba^T((\mathbf{d}X)X + X\mathbf{d}X))$$

$$= tr(Xba^T \mathbf{d}X) + tr(ba^T X\mathbf{d}X)) = tr((Xba^T + ba^T X)\mathbf{d}X)$$

$$\mathrm{D}_X f(X) = Xba^T + ba^T X, \nabla_X f(X) = ab^T X^T + X^T ab^T$$

$$\mathbf{d}^2 f(X) = \mathbf{d}((Xba^T + ba^T X)\mathbf{d}X) = tr(\mathbf{d}(Xba^T + ba^T X)\mathbf{d}X))$$

$$= tr((\mathbf{d}Xba^T + ba^T \mathbf{d}X)\mathbf{d}X)) = tr(\mathbf{d}Xba^T \mathbf{d}X) + tr(ba^T \mathbf{d}X\mathbf{d}X)$$

$$H(f(X)) = \frac{1}{2} K_{nn}(I_n \otimes ba^T + ab^T \otimes I_n + ab^T \otimes I_n + I_n \otimes ba^T) = K_{nn}(I_n \otimes ba^T + ab^T \otimes I_n)$$

7.3 梯度与无约束最优化

无约束极小化： $\min\limits_{x\in\mathbb{R}} f(x) \quad \left\{= \max\limits_{x\in\mathbb{R}}[-f(x)]\right\}$

$f(x)$称为**目标函数(objective function)**:当仅用于极小化问题时，又称为**代价函数(cost function)**.

主要思想是：松弛和逼近

松弛：找一个序列$\{X_k\}_{k=0}^{\infty}$,使得$\{f(X_k)\}_{k=0}^{\infty}$是松弛序列，即$f(X_k) \geq f(X_{k+1})$.

逼近：找一个简单函数代替目标函数，该简单函数与目标函数有相同的极值性质.

7.3.1 单变量函数的平稳点和极值点

$f(x)$的自变量x的邻域： B(x; r)={Y:|Y-x|<r}

$f(x)$的极值点x*： $f(x^*) \le f(x)$, x*, x∈ $B(x^*, r)$

$f(x)$的平稳点x*： $\partial f(x^*)$ =0

平稳点只是极值点的候选点.

假设函数在x处二阶可微，则有Taylor展式成立：

$$f(x+\Delta x) = f(x) + f'(x)\Delta x + \frac{1}{2} f''(x)(\Delta x)^2 + o(\Delta x^3)$$

平稳点(stationary point)，极小值点，严格极小值点，极大值点，严格极大值点，鞍点(saddle point)

7.3.2 向量的函数的平稳点和极值点

$f(X)$的自变量X的邻域：B(X; r)={Y: ||Y-X||<r}

$f(X)$的极值点X*： $f(X^*) \leq f(X)$，X*, X∈ $B(X^*, r)$

$f(X)$的平稳点X*： $\partial f(X^*)$ =0

平稳点只是极值点的候选点.

假设函数在X处二阶可微，则有Taylor展示成立：

$$f(X+\Delta X)=f(X)+\frac{\partial f(X)}{\partial X^T}\Delta X+\frac{1}{2}(\Delta X)^T\frac{\partial^2 f(X)}{\partial X \partial X^T}\Delta X+o(\|\Delta X\|^3)$$

平稳点(stationary point)，极小值点，严格极小值点，极大值点，严格极大值点，鞍点(saddle point)

7.3.3 矩阵的函数的平稳点和极值点

$f(X)$的自变量X的邻域：B(X; r)={Y: ||Y-X||<r}

$f(X)$的极值点X*： $f(X^*) \leq f(X)$, X*, X∈ $B(X^*, r)$

$f(X)$的平稳点X*： $\partial f(X^*)$ =0

平稳点只是极值点的候选点.

假设函数在矩阵X处二阶可微，则有Taylor展示成立：

$$f(X+\Delta X)=f(X)+\frac{\partial f(X)}{\partial vec(X)^T}vec(\Delta X)$$

$$+\frac{1}{2}(vec(\Delta X))^T\frac{\partial^2 f(X)}{\partial vec(X)\partial vec(X)^T}vec(\Delta X)+o(\|\Delta X\|^3)$$

平稳点(stationary point)，极小值点，严格极小值点，极大值点，严格极大值点，鞍点(saddle point)

表 7.3.1 实变函数的平稳点和极值点的条件

实变函数	$f(x):\mathbb{R}\to\mathbb{R}$	$f(\boldsymbol{x}):\mathbb{R}^n\to\mathbb{R}$	$f(\boldsymbol{X}):\mathbb{R}^{m\times n}\to\mathbb{R}$
平稳点	$\left.\frac{\partial f(x)}{\partial x}\right\|_{x=c}=0$	$\left.\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}\right\|_{\boldsymbol{x}=\boldsymbol{c}}=\boldsymbol{0}$	$\left.\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}}\right\|_{\boldsymbol{X}=\boldsymbol{C}}=\boldsymbol{O}_{m\times n}$
局部极小点	$\left.\frac{\partial^2 f(x)}{\partial x\partial x}\right\|_{x=c}\geqslant 0$	$\left.\frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x}\partial \boldsymbol{x}^{\mathrm{T}}}\right\|_{\boldsymbol{x}=\boldsymbol{c}}\succeq 0$	$\left.\frac{\partial^2 f(\boldsymbol{X})}{\partial \mathrm{vec}(\boldsymbol{X})\partial(\mathrm{vec}\,\boldsymbol{X})^{\mathrm{T}}}\right\|_{\boldsymbol{X}=\boldsymbol{C}}\succeq 0$
严格局部极小点	$\left.\frac{\partial^2 f(x)}{\partial x\partial x}\right\|_{x=c}>0$	$\left.\frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x}\partial \boldsymbol{x}^{\mathrm{T}}}\right\|_{\boldsymbol{x}=\boldsymbol{c}}\succ 0$	$\left.\frac{\partial^2 f(\boldsymbol{X})}{\partial \mathrm{vec}(\boldsymbol{X})\partial(\mathrm{vec}\,\boldsymbol{X})^{\mathrm{T}}}\right\|_{\boldsymbol{X}=\boldsymbol{C}}\succ 0$
局部极大点	$\left.\frac{\partial^2 f(x)}{\partial x\partial x}\right\|_{x=c}\leqslant 0$	$\left.\frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x}\partial \boldsymbol{x}^{\mathrm{T}}}\right\|_{\boldsymbol{x}=\boldsymbol{c}}\preceq 0$	$\left.\frac{\partial^2 f(\boldsymbol{X})}{\partial \mathrm{vec}(\boldsymbol{X})\partial(\mathrm{vec}\,\boldsymbol{X})^{\mathrm{T}}}\right\|_{\boldsymbol{X}=\boldsymbol{C}}\preceq 0$
严格局部极大点	$\left.\frac{\partial^2 f(x)}{\partial x\partial x}\right\|_{x=c}<0$	$\left.\frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x}\partial \boldsymbol{x}^{\mathrm{T}}}\right\|_{\boldsymbol{x}=\boldsymbol{c}}\prec 0$	$\left.\frac{\partial^2 f(\boldsymbol{X})}{\partial \mathrm{vec}(\boldsymbol{X})\partial(\mathrm{vec}\,\boldsymbol{X})^{\mathrm{T}}}\right\|_{\boldsymbol{X}=\boldsymbol{C}}\prec 0$
鞍　点	$\left.\frac{\partial^2 f(x)}{\partial x\partial x}\right\|_{x=c}$ 不定	$\left.\frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x}\partial \boldsymbol{x}^{\mathrm{T}}}\right\|_{\boldsymbol{x}=\boldsymbol{c}}$ 不定	$\left.\frac{\partial^2 f(\boldsymbol{X})}{\partial \mathrm{vec}(\boldsymbol{X})\partial(\mathrm{vec}\,\boldsymbol{X})^{\mathrm{T}}}\right\|_{\boldsymbol{X}=\boldsymbol{C}}$ 不定

定义：给定一个Hermitian矩阵$\boldsymbol{H}$，称向量$\boldsymbol{p}$为（相对于矩阵的$\boldsymbol{H}$）

（1）**正曲率方向**，若二次型$\boldsymbol{p}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{p} > 0$；

（2）**零曲率方向**，若二次型$\boldsymbol{p}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{p} = 0$；

（3）**负曲率方向**，若二次型$\boldsymbol{p}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{p} < 0$

定义：当矩阵$\boldsymbol{H}$是非线性函数$f(\mathrm{X})$的Hessian矩阵时，称

（1）$\boldsymbol{p}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{p}$为函数$f$沿着方向$\boldsymbol{p}$的曲率（curvature）；

（2）满足$\boldsymbol{p}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{p} > 0$的向量$\boldsymbol{p}$为$f$的正曲率方向；

（3）满足$\boldsymbol{p}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{p} < 0$的向量$\boldsymbol{p}$为$f$的负曲率方向.

◆ **曲率方向也就是函数的最大变化率方向**

定理：令$f(\boldsymbol{w})$是复变量$\boldsymbol{w}$的实值函数. 通过将$\boldsymbol{w}$和$\boldsymbol{w}^*$视为独立变元，实目标函数$f(\boldsymbol{w})$的曲率方向由共轭梯度向量$\nabla_{w^*} f(\boldsymbol{w})$给出.

7.3.4 实变函数的梯度分析

向量变元情形下目标函数$\boldsymbol{f}(\mathrm{X})$的局部极小点条件:

（1）必要条件： 若$\mathrm{X}_{\mathbf{0}}$是$f(\mathrm{X})$的局部极小点，则该函数在点$\mathrm{X}_{\mathbf{0}}$的共轭梯度为零向量，并且共轭梯度的梯度即Hessian矩阵半定，即

$$\left.\frac{\partial f(\mathrm{X})}{\partial \mathrm{X}^{T}}\right|_{X=X_0} = \mathbf{0}, \quad \left.\frac{\partial^2 f(\boldsymbol{X})}{\partial X \partial \boldsymbol{X}^{\mathrm{T}}}\right|_{X=X_0} \geq 0$$

（2）充分条件： 若函数$f(\boldsymbol{X})$在$\boldsymbol{X}_{\mathbf{0}}$的共轭梯度为零向量，并且Hessian矩阵正定，即

$$\left.\frac{\partial f(\mathrm{X})}{\partial \mathrm{X}^{T}}\right|_{X=X_0} = \mathbf{0}, \quad \left.\frac{\partial^2 f(\boldsymbol{X})}{\partial X \partial \boldsymbol{X}^{\mathrm{T}}}\right|_{X=X_0} > 0$$

则$\boldsymbol{X}_{\mathbf{0}}$是函数$f(\boldsymbol{X})$的严格局部极小点.

具有等式约束的优化问题

$$\min_{s.t.\ A\mathbf{x}=\mathbf{b}} f(\mathbf{x}) \qquad \text{其中} \mathbf{A} \in \mathbb{R}^{m\times n} (m \le n)$$

Lagrange乘子法

$$J(\mathbf{x}) = f(\mathbf{x}) + \boldsymbol{\lambda}^{\mathrm{T}}(\mathbf{b} - \mathbf{A}\mathbf{x})$$

梯度分析：

（1）计算目标函数的梯度向量（矩阵）.

（2）计算目标函数的Hessian矩阵，并分析其正定性.

梯度分析与最优化：

（1）设计合适的目标函数.

（2）令梯度矩阵等于零矩阵，得到优化问题的平稳点.

（3）利用负梯度，得到梯度算法.

（4）利用Hessian矩阵的正定性，分析梯度算法的收敛性能（平稳点是否为局部或全局极小点）.

7.4 平滑凸优化的一阶算法

凸集和凸函数

$$f(\alpha X+(1-\alpha)Y)\le \alpha f(X)+(1-\alpha)f(Y)$$

严格凸函数

$$f(\alpha X+(1-\alpha)Y)< \alpha f(X)+(1-\alpha)f(Y)$$

凸函数辨识的充分必要条件

$$f(Y)\ge f(X)+<\nabla_X f(X),Y-X> \qquad \text{一阶充要条件}$$

$$H_X f(X)=\frac{\partial^2 f(X)}{\partial X\partial X^T}\succ\succ 0 \qquad \text{二阶充要条件}$$

符号 $\succ\succ$ 表示非负定

梯度下降法

$$f(X+\Delta X)=f(X)+\frac{\partial f(X)}{\partial X^T}\Delta X+\frac{1}{2}(\Delta X)^T\frac{\partial^2 f(X)}{\partial X\partial X^T}\Delta X+o(\|\Delta X\|^3)$$

令 $H_X f(X)=\frac{\partial^2 f(X)}{\partial X\partial X^T}\approx\frac{1}{t}I$,略去高阶无穷小,则该函数的一个逼近为

$$f(X+\Delta X)=f(X)+\frac{\partial f(X)}{\partial X^T}\Delta X+\frac{1}{2t}(\Delta X)^T\Delta X$$

对其求梯度，得解

$$\Delta X=-t\frac{\partial f(X)}{\partial X}$$

$$\mathrm{d}(f(X+\Delta X))=\mathrm{d}\ \mathrm{tr}(f(X)+\frac{\partial f(X)}{\partial X^T}\Delta X+\frac{1}{2t}(\Delta X)^T\Delta X)$$

$$=\ \mathrm{tr}\left(\mathrm{d}\left(f(X)+\frac{\partial f(X)}{\partial X^T}\Delta X+\frac{1}{2t}(\Delta X)^T\Delta X\right)\right)$$

$$=\mathrm{tr}\left(\frac{\partial f(X)}{\partial X^T}\mathrm{d}(\Delta X)+\frac{1}{2t}\mathrm{d}((\Delta X)^T\Delta X)\right)=\mathrm{tr}\left((\frac{\partial f(X)}{\partial X^T}+\frac{1}{t}(\Delta X)^T)\mathrm{d}(\Delta X)\right)$$

所以 $\frac{\partial f(X)}{\partial X^T}+\frac{1}{t}(\Delta X)^T$ 是该函数关于 ΔX 的协梯度矩阵.根据极值点的必要性，有

$$\frac{\partial f(X)}{\partial X^T}+\frac{1}{t}(\Delta X)^T=0\Rightarrow\Delta X=-t\frac{\partial f(X)}{\partial X}$$

Newton法

$$f(X+\Delta X)=f(X)+\frac{\partial f(X)}{\partial X^T}\Delta X+\frac{1}{2}(\Delta X)^T\frac{\partial^2 f(X)}{\partial X\partial X^T}\Delta X+o(\|\Delta X\|^3)$$

略去高阶无穷小,则该函数的一个逼近为

$$f(X+\Delta X)=f(X)+\frac{\partial f(X)}{\partial X^T}\Delta X+\frac{1}{2}(\Delta X)^T\frac{\partial^2 f(X)}{\partial X\partial X^T}\Delta X$$

此表达式的自变量是ΔX

对于自变量的增量，它也是极值．对增量求梯度，得

$$\Delta X=-\left(\frac{\partial^2 f(X)}{\partial X\partial X^T}\right)^{-1}\frac{\partial f(X)}{\partial X}$$

$$\mathrm{d}(f(X+\Delta X))=\mathrm{d}\,\mathrm{tr}(f(X)+\frac{\partial f(X)}{\partial X^T}\Delta X+\frac{1}{2}(\Delta X)^T\frac{\partial^2 f(X)}{\partial X\partial X^T}\Delta X)$$

$$=\mathrm{tr}\left(\mathrm{d}\left(f(X)+\frac{\partial f(X)}{\partial X^T}\Delta X+\frac{1}{2}(\Delta X)^T\frac{\partial^2 f(X)}{\partial X\partial X^T}\Delta X\right)\right)$$

$$=\mathrm{tr}\left(\frac{\partial f(X)}{\partial X^T}\mathrm{d}(\Delta X)+\mathrm{d}((\Delta X)^T\frac{\partial^2 f(X)}{\partial X\partial X^T}\Delta X)\right)=\mathrm{tr}\left(\left(\frac{\partial f(X)}{\partial X^T}+(\Delta X)^T\frac{\partial^2 f(X)}{\partial X\partial X^T}\right)\mathrm{d}(\Delta X)\right)$$

所以$\frac{\partial f(X)}{\partial X^T}+(\Delta X)^T\frac{\partial^2 f(X)}{\partial X\partial X^T}$是该函数关于$\Delta X$的协梯度矩阵.根据极值点的必要性,极值点为

$$\frac{\partial f(X)}{\partial X^T}+(\Delta X)^T\frac{\partial^2 f(X)}{\partial X\partial X^T}=0\Rightarrow\Delta X=-\left(\frac{\partial^2 f(X)}{\partial X\partial X^T}\right)^{-1}\frac{\partial f(X)}{\partial X}$$

算法

（1）令

$$\Delta X_k = -H^{-1}(f(X_k)\nabla_X f(X_k) = -\left(\frac{\partial^2 f(X_k)}{\partial X \partial X^T}\right)^{-1} \frac{\partial f(X_k)}{\partial X}$$ （牛顿法）

或

$$\Delta X_k = -\nabla_X f(X_k) = -\frac{\partial f(X_k)}{\partial X}$$ （最速下降法）

（2）选择步长

$$\mu_k > 0$$

（3）更新

$$X_{k+1} = X_k + \mu_k \Delta X_k$$

例

$$X_{Tik} = \arg\min_{X}(\| AX-b \|_2^2 + \lambda \| X \|_2^2).$$

$$X_{Tik} = (A^H A + \lambda I)^{-1} A^H b.$$

$$\mathrm{d}(\| AX-b \|_2^2 + \lambda \| X \|_2^2) = \mathrm{d}(\mathrm{tr}(\| AX-b \|_2^2 + \lambda \| X \|_2^2))$$

$$= tr(\mathrm{d}(AX-b)^T (AX-b) + \lambda \mathrm{d}(X^T X)))$$

$$= tr(\mathrm{d}X^T A^T (AX-b) + (AX-b)^T A\mathrm{d}X + \lambda \mathrm{d}(X^T X)))$$

$$= tr((AX-b)^T A\mathrm{d}X) + tr((AX-b)^T A\mathrm{d}X) + 2\lambda tr(X^T \mathrm{d}X)$$

因而

$$\nabla_X f = \left(2(AX-b)^T A + 2\lambda X^T\right)^T$$

函数梯度矩阵辨识定理:

$$\mathrm{d}f(\mathbf{X}) = \mathrm{tr}(\mathbf{A}\mathrm{d}\mathbf{X}) \quad \Leftrightarrow \quad \nabla_{\mathbf{X}} f(\mathbf{X}) = \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{A}^{\mathrm{T}}$$

$$\mathrm{d}(\| AX-b \|_2^2+\lambda \| X \|_2^2) = \mathrm{d}(\mathrm{tr}(\| AX-b \|_2^2+\lambda \| X \|_2^2))$$
$$= tr(\mathrm{d}(AX-b)^T(AX-b)+\lambda\mathrm{d}(X^TX)))$$
$$= tr(\mathrm{d}X^TA^T(AX-b)+(AX-b)^TA\mathrm{d}X+\lambda\mathrm{d}(X^TX)))$$
$$= tr((AX-b)^TA\mathrm{d}X)+tr((AX-b)^TA\mathrm{d}X)+2\lambda tr(X^T\mathrm{d}X)$$

因而

$$\nabla_X f = \left(2(AX-b)^TA+2\lambda X^T\right)^T$$

所以

$$\mathrm{d}^2(\| AX-b \|_2^2+\lambda \| X \|_2^2) = \mathrm{d}tr\left(\left(2(AX-b)^TA+2\lambda X^T\right)dX\right)$$
$$= tr\left(\mathrm{d}\left(2(AX-b)^TA+2\lambda X^T\right)dX\right) = tr\left(\mathrm{d}\left(2(X^TA^T-b^T)A+2\lambda X^T\right)dX\right)$$
$$= tr\left(2\mathrm{d}X^TA^TAdX+2\lambda\mathrm{d}X^TdX\right) == tr\left(\mathrm{d}X^T(2A^TA+2\lambda I)dX\right)$$
$$\Rightarrow H(\| AX-b \|_2^2+\lambda \| X \|_2^2) = 2A^TA+2\lambda I$$

$B=(1)_{1\times 1}$

Hessian矩阵辨识定理之一：

$$\mathrm{d}^2 f(\mathbf{X}) = \mathrm{tr}(\mathbf{B}(\mathrm{d}\mathbf{X})^{\mathrm{T}}\mathbf{C}\mathrm{d}\mathbf{X}) \Leftrightarrow \mathbf{H}(f(\mathbf{X})) = \frac{1}{2}(\mathbf{B}^{\mathrm{T}} \otimes \mathbf{C} + \mathbf{B} \otimes \mathbf{C}^{\mathrm{T}})$$

最速下降法

```
A=rand(4,4);
b=rand(4,1);
X=zeros(4,1);
lamda=.2;
Xt=(A'*A+lamda*eye(4,4))\A'*b;
disp([0,Xt']);
Df=zeros(4,4);
jj=10;kk=1;
for ii=1:200
   df=2*(A*X-b)'*A+2*lamda*X';
   Df=-df';
   X=X+lamda*Df;
   if jj==ii
      disp([ii,X']);
      kk=kk+1;
       jj=kk*40;
   end
end
disp([ii,X']);
```

lamda=0.2 （迭代一百次，与理论值就相同了）

0	**-0.3382**	**0.3822**	**0.2542**	**0.5825**
10.0000	-0.2971	0.3736	0.1639	0.4810
80.0000	-0.3382	0.3822	0.2542	0.5826
120.0000	-0.3382	0.3822	0.2542	0.5825
160.0000	-0.3382	0.3822	0.2542	0.5825
200.0000	-0.3382	0.3822	0.2542	0.5825
200.0000	**-0.3382**	**0.3822**	**0.2542**	**0.5825**

lamda的值不能太大，太大就发散.

lamda1=0.1 （迭代两百次，与理论值就相同了）

0	-0.1373	-0.4025	-0.5467	1.2368
10.0000	0.1009	-0.0211	-0.1409	0.5440
80.0000	-0.1230	-0.3595	-0.5334	1.1917
120.0000	-0.1346	-0.3910	-0.5447	1.2268
160.0000	-0.1367	-0.3995	-0.5463	1.2345
200.0000	-0.1372	-0.4017	-0.5466	1.2363
200.0000	-0.1372	-0.4017	-0.5466	1.2363

lamda参数混淆了！！！

```
A=rand(4,4);
b=rand(4,1);
X=zeros(4,1);
lamda=.2;
Xt=(A'*A+lamda*eye(4,4))\A'*b;
disp([0,Xt']);

lamda1=0.15;
Df=zeros(4,4);
jj=10;kk=1;
for ii=1:200
    df=2*(A*X-b)'*A+2*lamda*X';
    Df=-df';
    X=X+lamda1*Df;
    if jj==ii
        disp([ii,X']);
        kk=kk+1;
         jj=kk*40;
    end
end
disp([ii,X']);
```

lamda1=0.15 （迭代一百次，与理论值就相同了）

0	**0.2878**	**0.3976**	**0.3428**	**0.1711**
10.0000	**0.3260**	**0.3407**	**0.2934**	**0.1992**
80.0000	**0.2882**	**0.3976**	**0.3421**	**0.1712**
120.0000	**0.2879**	**0.3976**	**0.3427**	**0.1711**
160.0000	**0.2878**	**0.3976**	**0.3428**	**0.1711**
200.0000	**0.2878**	**0.3976**	**0.3428**	**0.1711**
200.0000	**0.2878**	**0.3976**	**0.3428**	**0.1711**

lamda1的值不能太大，太大就发散.

lamda1=0.1 （迭代两百次，与理论值就相同了）

0	**0.1264**	**-0.0274**	**0.0519**	**0.0388**
10.0000	0.0909	0.0032	0.0519	0.0436
80.0000	0.1262	-0.0273	0.0523	0.0385
120.0000	0.1264	-0.0274	0.0520	0.0387
160.0000	0.1264	-0.0274	0.0519	0.0387
200.0000	0.1264	-0.0274	0.0519	0.0388
200.0000	**0.1264**	**-0.0274**	**0.0519**	**0.0388**

牛顿法（收敛速度快）

```
A=rand(4,4);
b=rand(4,1);
X=zeros(4,1);
lamda=.2;
%Tikhonov解
Xt=(A'*A+lamda*eye(4,4))\A'*b;
disp([0,Xt']);
lamda1=1.2;%学习率
Df=zeros(4,4);
jj=5;kk=1;
for ii=1:200
   df=2*(A*X-b)'*A+2*lamda*X';%协梯度
   Df=2*(A'*A+lamda*eye(4,4));%H矩阵
   X=X-lamda1*Df\df';
   if jj==ii
      disp([ii,X']);
      kk=kk+1;
       jj=kk*5;
   end
end
disp([ii,X']);
```

lamda=.2，lamda1=1.2

0	-0.0838	0.2688	0.0570	0.1199
5	**-0.0838**	**0.2687**	**0.0570**	**0.1198**
10	-0.0838	0.2688	0.0570	0.1199
15	-0.0838	0.2688	0.0570	0.1199
20	-0.0838	0.2688	0.0570	0.1199
25	-0.0838	0.2688	0.0570	0.1199

5步收敛到一个比较精确的值！

牛顿法比最速下降法收敛速度快.

lamda=10.2，lamda1=1.2

0	-0.0838	0.2688	0.0570	0.1199
5	**-0.0838**	**0.2687**	**0.0570**	**0.1198**
10	-0.0838	0.2688	0.0570	0.1199
15	-0.0838	0.2688	0.0570	0.1199
20	-0.0838	0.2688	0.0570	0.1199
25	-0.0838	0.2688	0.0570	0.1199

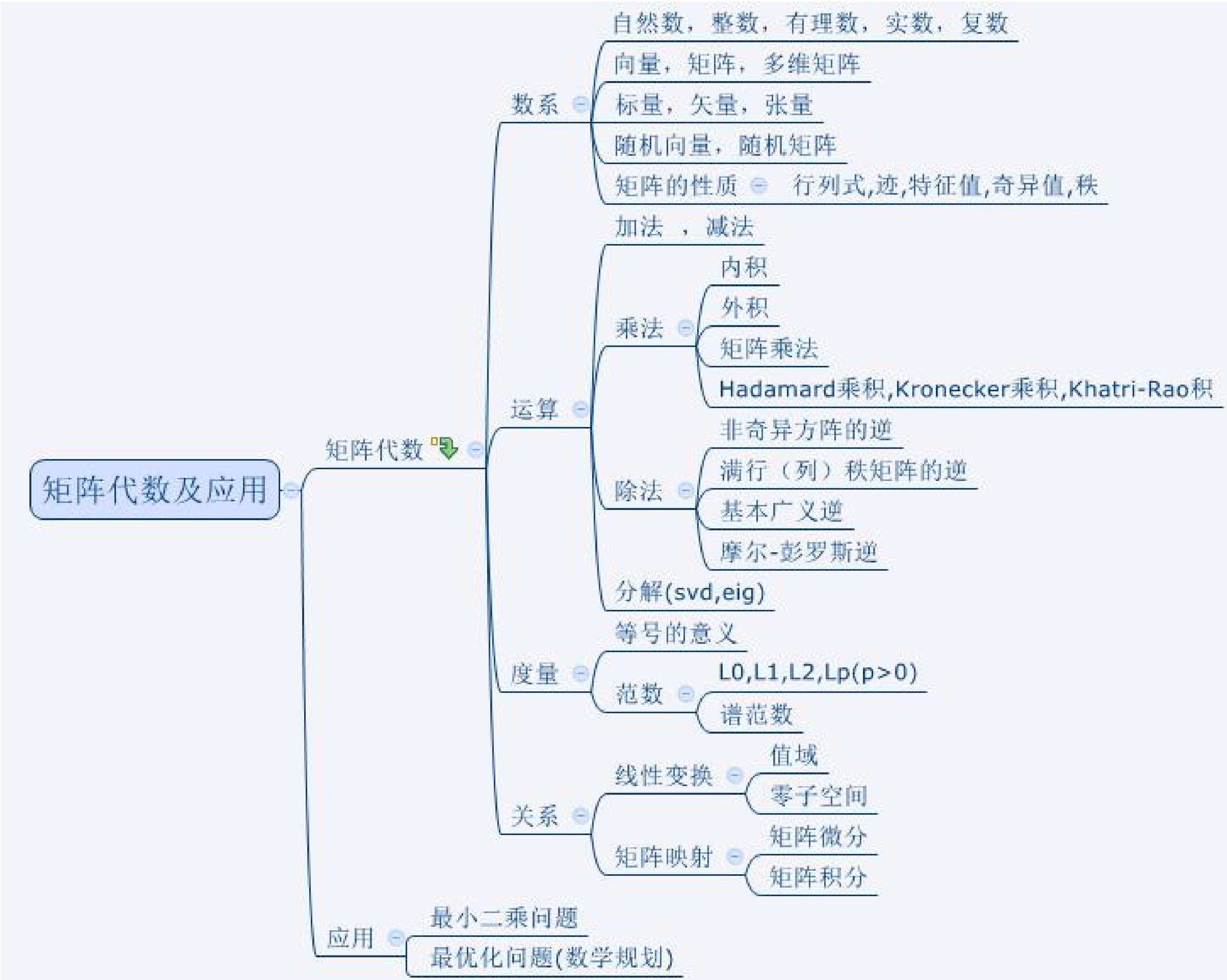
也是5步收敛到一个比较精确的值！

作业

7.9,7.12

实验 (选做)

9.编程验证7.11的结果（参见PPT第31页）


矩阵代数及应用
矩阵代数
数系
自然数，整数，有理数，实数，复数
向量，矩阵，多维矩阵
标量，矢量，张量
随机向量，随机矩阵
矩阵的性质
行列式,迹,特征值,奇异值,秩
运算
加法 ，减法
乘法
内积
外积
矩阵乘法
Hadamard乘积,Kronecker乘积,Khatri-Rao积
除法
非奇异方阵的逆
满行（列）秩矩阵的逆
基本广义逆
摩尔-彭罗斯逆
分解(svd,eig)
度量
等号的意义
范数
L0,L1,L2,Lp(p>0)
谱范数
关系
线性变换
值域
零子空间
矩阵映射
矩阵微分
矩阵积分
应用
最小二乘问题
最优化问题(数学规划)

矩阵代数纲要

谢谢！

gtding@shu.edu.cn

66135773，计算机大楼921

上海大学计算机工程与科学学院

2024年1月