作业

（1）用矩阵符号推演证明 $B=S^{-1}AS$

$$\mathbf{V}=\mathbf{U}S, L(\mathbf{U})=\mathbf{U}A, L(\mathbf{V})=\mathbf{V}B$$
$$\Rightarrow \mathbf{V}B=L(\mathbf{V})=L(\mathbf{U}S)=L(\mathbf{U})S=\mathbf{U}AS$$
$$\Rightarrow \mathbf{U}SB=\mathbf{U}AS$$
$$\Rightarrow SB=AS$$
$$\Rightarrow B=S^{-1}AS$$

注：

1）黑体字母表示基向量从左到右排列的一个代数表示，此处不能理解成一个矩阵。

2）矢量、线性变换是线性空间里的研究对象，它们的存在不依赖于基的存在。

线性空间：R^n(维数为 n), R^m(维数为 m)

矢量（向量）：$X\in R^n, Y\in R^m$, n和m都是大于零的整数

线性变换：$\begin{cases} L: \ R^n \to R^m, \\ L(\alpha_1 X_1+\alpha_2 X_2)=\alpha_1 L(X_1)+\alpha_2 L(X_2), \\ X_1, X_2\in R^n, \alpha_1, \alpha_2\in R\text{(注：复数域类同)} \end{cases}$

3）但是，要研究它们，就必须给线性空间建立坐标系，也就是选取一组最大线性无关组做为基。继而，所有矢量都可以用这组基向量线性表示，该表示的系数按基向量的排列顺序排列，就得到了该矢量的代数表示：向量。线性变换也如此，用它们对基向量进行变换，变换后的系数按矩阵方式排列，就得到了它的代数表示：矩阵。当有了代数表示后，矢量和线性变换之间的数量关系及运算，就全部落实到它们的代数表示上了，此时，矢量和线性变换这种抽象的东西，就具体化了。只有基向量选定后，矢量和线性变换的表示才能落实下来。一般我们忽略了基的问题，是因为默认选择的基是标准正交基(此处特指：空间里的坐标轴方向上的单位向量们)。

- 空间上的基向量是选取的，不唯一
- 矢量在不同基下的表示也是不一样的
- 同一矢量在不同基下的代数表示可以互相转换（计算）得到，称为坐标变换
- 对于线性变换，以上说法也正确

设$\mathbf{U}=(\mathbf{u}_1,\mathbf{u}_2,\cdots,\mathbf{u}_n), \widetilde{\mathbf{U}}=(\widetilde{\mathbf{u}}_1,\widetilde{\mathbf{u}}_2,\cdots,\widetilde{\mathbf{u}}_n)$是选定的两组$R^n$上的基

$\mathbf{V}=(\mathbf{v}_1,\mathbf{v}_2,\cdots,\mathbf{v}_m), \widetilde{\mathbf{V}}=(\widetilde{\mathbf{v}}_1,\widetilde{\mathbf{v}}_2,\cdots,\widetilde{\mathbf{v}}_m)$是选定的两组$R^m$上的基

则矢量 **X** 和 **Y** 在 R^n 和 R^m 上有如下表示（用黑体表示矢量或基向量，正体表示向量或矩阵）

$$\mathbf{X}=x_1\mathbf{u}_1+x_2\mathbf{u}_2+\cdots+x_n\mathbf{u}_n=\sum_{i=1}^{n}x_i\mathbf{u}_i=(\mathbf{u}_1,\mathbf{u}_1,\cdots,\mathbf{u}_1)\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}=\mathbf{U}X$$

$$\mathbf{X}=\widetilde{x}_1\widetilde{\mathbf{u}}_1+\widetilde{x}_2\widetilde{\mathbf{u}}_2+\cdots+\widetilde{x}_n\widetilde{\mathbf{u}}_n=\sum_{i=1}^{n}\widetilde{x}_i\widetilde{\mathbf{u}}_i=(\widetilde{\mathbf{u}}_1,\widetilde{\mathbf{u}}_1,\cdots,\widetilde{\mathbf{u}}_1)\begin{pmatrix}\widetilde{x}_1\\\widetilde{x}_2\\\vdots\\\widetilde{x}_n\end{pmatrix}=\widetilde{\mathbf{U}}\widetilde{X}$$

$$\mathbf{Y} = y_1\mathbf{v}_1 + y_2\mathbf{v}_2 + \cdots + y_m\mathbf{v}_m = \sum_{i=1}^{m} x_i\mathbf{v}_i = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m)\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \mathbf{V}Y$$

$$\mathbf{Y} = \tilde{y}_1\tilde{\mathbf{v}}_1 + \tilde{y}_2\tilde{\mathbf{v}}_2 + \cdots + \tilde{y}_m\tilde{\mathbf{v}}_m = \sum_{i=1}^{m} \tilde{y}_i\tilde{\mathbf{v}}_i = (\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \cdots, \tilde{\mathbf{v}}_m)\begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_m \end{pmatrix} = \tilde{\mathbf{V}}\tilde{Y}$$

注：左边是繁写，右边是简写，是数学记号的问题。数学记号的使用，是为了方便。
大家不要纠缠矢量和向量的学术意义上的区别，不同人与不同人的看法不一样。为方便起见，为了概念的细微辨识，本文所说矢量，是指一种物理量。向量是矢量的一种数学表示形式。明白这个区别就行。
对于线性变换，必有

$$L(\mathbf{u}_i) = a_{i1}\mathbf{v}_1 + a_{i2}\mathbf{v}_2 + \cdots + a_{im}\mathbf{v}_m = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m)(a_{i1}, a_{i2}, \cdots, a_{im})^T, i = 1,2,\cdots,n$$

也就是

$$L((\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n) = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m)\begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \cdots & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{pmatrix}$$

简记为

$L(\mathbf{U}) = \mathbf{V}A$（注：此处的 A 与本题解答的红色部分里的 A 没有关系，此处只是这里使用的一个记号，其含义不解释了，自明。写此公式的时候，没有看红色部分里的那个公式，已经写好了，不想再换其他符号了）
这个 A 称为线性变换在基 **U** 和 **V** 下的表示矩阵。如果自变量空间和应变量空间各选两组基，则一共有四种表示矩阵（A, A_1, A_2, A_3）。

$L(\mathbf{U}) = \mathbf{V}A$

$L(\mathbf{U}) = \tilde{\mathbf{V}}A_1$

$L(\tilde{\mathbf{U}}) = \mathbf{V}A_2$

$L(\tilde{\mathbf{U}}) = \tilde{\mathbf{V}}A_3$

它们都表示同一线性变换 L
如果明白了以上内容，关于基向量的事情，就没有障碍了。

（2）用矩阵符号推演证明如下两个公式

令I为单位矩阵，则：

$$eig(\mathrm{I} + \mathrm{cA}) = \mathrm{I} + \mathrm{c} \cdot \mathrm{eig}(\mathrm{A})$$

$$\mathrm{eig}(\mathrm{A} - \mathrm{cI}) = \mathrm{eig}(\mathrm{A}) - \mathrm{c}$$

解：
证明：

①记$\lambda = eig(\mathrm{A})$,
$\because \exists \mathbf{x}$, 使 $\mathrm{A}\,\mathbf{x} = \lambda\mathbf{x}$,
$\therefore (c\mathrm{A}+\mathrm{I})\mathbf{x} = c\lambda\mathbf{x} + \mathbf{x} = (1+c\,\lambda)\mathbf{x}$,
$\therefore 1+c\,\lambda$是$c\mathrm{A}+\mathrm{I}$的特征值,
得证。
②记$\lambda = eig(\mathrm{A})$,
$\because \exists \mathbf{x}$, 使 $\mathrm{A}\,\mathbf{x} = \lambda\mathbf{x}$,
$\therefore (\mathrm{A}-c\,\mathrm{I})\mathbf{x} = \lambda\mathbf{x} - c\mathbf{x} = (\lambda - c)\mathbf{x}$,
$\therefore \lambda - c$是$\mathrm{A}-c\mathrm{I}$的特征值,
得证。

(3)书本习题

1.1

解：根据图得到：

$$\begin{cases} x1 + x5 = 800; \\ x1 + x4 = 400 + x2; \\ x3 + 600 = x2; \\ 1600 + x3 = 400 + x7; \\ x7 = x4 + x6; \\ x5 + x6 = 1000; \end{cases}$$

整理可得：

$$\begin{cases} x1 + x5 = 800; \\ x1 - x2 + x4 = 400; \\ x2 - x3 = 600; \\ x3 - x7 = -1200; \\ x4 + x6 - x7 = 0; \\ x5 + x6 = 1000; \end{cases}$$

有 7 个未知数，但只有 6 个式子，所以矩阵的秩<7，即有无穷多个解。
写出增广矩阵：

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 800 \\ 1 & -1 & 0 & 1 & 0 & 0 & 0 & 400 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 600 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & -1200 \\ 0 & 0 & 0 & 1 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1000 \end{bmatrix}$$

解得：

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} 800 \\ 400 \\ -200 \\ 0 \\ 0 \\ 1000 \\ 1000 \end{bmatrix} + k_1 \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 1 \\ -1 \\ 0 \end{bmatrix} + k_2 \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

网络流问题。把 x6 和 X7 作为控制参数，也就是令 k1=1000-X6, k2=x7-1000，则网络运行状态就完全确定。

1.2

解：

$$\begin{cases} 3I_2 + 5I_3 = 4; \\ 5I_1 + 3I_2 = 2; \\ I_1 + I_3 = I_2; \end{cases}$$

写出增广矩阵：

$$\begin{bmatrix} 0 & 3 & 5 & 4 \\ 5 & 3 & 0 & 2 \\ 1 & -1 & 1 & 0 \end{bmatrix}$$

解得：

$$\begin{cases} I_1 = 0.0727(A) \\ I_2 = 0.5455(A) \\ I_3 = 0.4727(A) \end{cases}$$

1.3

解：

(1): 因为 $rank\left[B, AB, A^2B\right] = rank \begin{bmatrix} 0 & 1 & 0 \\ 1 & -0.9 & 0.81 \\ 1 & 0.5 & 0.25 \end{bmatrix} = 3$

所以矩阵对(A, B)是可控的。

(2): 因为 $rank\left[B, AB, A^2B\right] = rank \begin{bmatrix} 1 & 0.5 & 0.19 \\ 1 & 0.7 & 0.45 \\ 0 & 0 & 0 \end{bmatrix} = 2$

所以矩阵对(A, B)是不可控的。

1.7

解:

证明如下:

$$\because x_1 \neq x_2, y_1 \neq y_2$$

$$\therefore \text{两点式方程可表示为}: \frac{y-y_1}{x-x_1} = \frac{y_2-y_1}{x_2-x_1}$$

$$\text{整理后可得}: (y_2-y_1)x-(x_2-x_1)\mathrm{y} = y_2x_1-x_2y_1$$

$$\therefore \text{写出上式的等效形式为}: \begin{vmatrix} 1 & x & y \\ 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \end{vmatrix} = 0$$

得证。

1.8

解:

证明如下:

将三个点的坐标代入到平面方程中，可以得到:

$$\begin{cases} ax_1+by_1+cz_1=-1 \\ ax_2+by_2+cz_2=-1 \\ ax_3+by_3+cz_3=-1 \end{cases}$$

利用 Gramer 法则，可以得到:

$$a=\frac{\begin{vmatrix} -1 & y_1 & z_1 \\ -1 & y_2 & z_2 \\ -1 & y_3 & z_3 \end{vmatrix}}{\begin{vmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{vmatrix}}, \mathrm{b}=\frac{\begin{vmatrix} x_1 & -1 & z_1 \\ x_2 & -1 & z_2 \\ x_3 & -1 & z_3 \end{vmatrix}}{\begin{vmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{vmatrix}}, c=\frac{\begin{vmatrix} x_1 & y_1 & -1 \\ x_2 & y_2 & -1 \\ x_3 & y_3 & -1 \end{vmatrix}}{\begin{vmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{vmatrix}}$$

所以可以将平面方程表示为:

$$\begin{vmatrix} -1 & y_1 & z_1 \\ -1 & y_2 & z_2 \\ -1 & y_3 & z_3 \end{vmatrix}x+\begin{vmatrix} x_1 & -1 & z_1 \\ x_2 & -1 & z_2 \\ x_3 & -1 & z_3 \end{vmatrix}y+\begin{vmatrix} x_1 & y_1 & -1 \\ x_2 & y_2 & -1 \\ x_3 & y_3 & -1 \end{vmatrix}z=\begin{vmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{vmatrix}$$

将此式化简，为:

$$\begin{vmatrix} 1 & x & y & z \\ 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ 1 & x_3 & y_3 & z_3 \end{vmatrix}=0$$

得证。

1.10

解：

若向量两两求内积均为 0，那么这就是一组正交向量；

可求得：$\begin{bmatrix}1\\0\\1\\2\end{bmatrix}^T \bullet \begin{bmatrix}-1\\1\\1\\0\end{bmatrix}=0;\begin{bmatrix}-1\\1\\1\\0\end{bmatrix}^T \bullet \begin{bmatrix}-1\\-2\\1\\0\end{bmatrix}=0;\begin{bmatrix}1\\0\\1\\2\end{bmatrix}^T \bullet \begin{bmatrix}-1\\-2\\1\\0\end{bmatrix}=0$

所以这是一组正交向量。

1.11

解：

证明如下：

$uv^Tu=(v^Tu)u \Rightarrow (v^Tu)$是$uv^T$的特征值，$u$是对应的特征向量

$uv^TX=\lambda X \Rightarrow u(v^TX)=(v^TX)u=\lambda X \Rightarrow X=\alpha u, \lambda=(v^Tu)$

则必定能找到与u正交的一组两两正交的单位向量$u_2,\cdots,u_n$

令u_1是u的单位化向量,$U=(u_1,u_2,\cdots,u_n)$,则

$U^{-1}uv^TU=diag(\lambda,0,\cdots,0), U^{-1}(I+uv^T)U=diag(1+\lambda,1,\cdots,1)$,

所以$\det(I+uv^T)=1+\lambda$.

2.11

$$Q(X)=X^{\mathrm{T}}\begin{pmatrix}1&0&0&0\\0&-1&0&0\\0&0&-1&0\\0&0&0&-1\end{pmatrix}X$$

$$Q(AY)=Y^{\mathrm{T}}A^T\begin{pmatrix}1&0&0&0\\0&-1&0&0\\0&0&-1&0\\0&0&0&-\end{pmatrix}AY$$

$$A^T\begin{pmatrix}1&0&0&0\\0&-1&0&0\\0&0&-1&0\\0&0&0&-1\end{pmatrix}A=\begin{pmatrix}1&0&0&0\\0&-1&0&0\\0&0&-1&0\\0&0&0&-1\end{pmatrix}$$

这是洛伦兹矩阵的定义

显然，两个洛伦兹矩阵的乘积还是洛伦兹矩阵。

在洛伦兹矩阵的变换下，本题所述二次型的数学表达形式不变!!! (这是洛伦兹

变换的特点)

2.12

证:

1）

$$\because M=\lambda P+(1-\lambda)Q\rightarrow\sum_{i=1}^{n}m_{ij}=\lambda\sum_{i=1}^{n}p_{ij}+(1-\lambda)\sum_{i=1}^{n}q_{ij}=\lambda+(1-\lambda)=1$$

$\therefore M=\lambda P+(1-\lambda)Q$是Markov矩阵

2）

$$\because 令M=PQ\rightarrow\sum_{i=1}^{n}m_{ij}=\sum_{i=1}^{n}\sum_{k=1}^{n}p_{ik}q_{kj}=\sum_{k=1}^{n}\sum_{i=1}^{n}p_{ik}q_{kj}=\sum_{k=1}^{n}q_{kj}=1$$

$\therefore PQ$也为Markov矩阵

2.13

证:

$\because M是Markov矩阵$

$\therefore \sum_{i=1}^{n}m_{ij}=1$

$\therefore \sum_{i=1}^{n}y_i=\sum_{j=1}^{n}\sum_{i=1}^{n}m_{ij}x_j=\sum_{j=1}^{n}x_j$

$\because x$是概率向量

$\therefore \sum_{j=1}^{n}x_j=1$

$\therefore \sum_{i=1}^{n}y_i=1且y_i\geq 0(i=1,2,\cdots \mathrm{n})$

$\therefore$ 可证y=Mx是概率向量

3.6

```
>> A=[-2,4,3;0,0,0;-1,5,2]
A =
    -2     4     3
     0     0     0
    -1     5     2
>> [U,D]=eig(A)
U =
   -0.9487   -0.7071   -0.7548
         0         0    0.1078
   -0.3162   -0.7071   -0.6470
D =
    -1     0     0
     0     1     0
     0     0     0
>> U*D/U
ans =
   -2.0000    4.0000    3.0000
         0         0         0
   -1.0000    5.0000    2.0000
```

所以，解为:

```
>> U*(D-2*D)/U
ans =
    2.0000   -4.0000   -3.0000
         0         0         0
    1.0000   -5.0000   -2.0000
```

3.7

```
>> A=[1,1,2;-1,2,1;0,1,3]
A =
     1     1     2
    -1     2     1
     0     1     3
>> [U,D]=eig(A)
U =
    0.4082   -0.5774    0.7071
    0.8165    0.5774    0.0000
   -0.4082   -0.5774    0.7071
D =
    1.0000         0         0
         0    2.0000         0
         0         0    3.0000
```

4.3

A =

149	-50	-154
537	180	546
-27	9	-25

>> [U,S,V]=svd(A)

U =

0.0228	0.9994	0.0247
-0.9988	0.0217	0.0434
0.0428	-0.0257	0.9988

S =

787.6038	0	0
0	219.4293	0
0	0	**16.3725**

V =

-0.6782	0.7349	-0.0004
-0.2292	-0.2110	0.9502
-0.6982	-0.6445	-0.3115

4.7

正定矩阵的特征值是正的，能对角化，有 n 个相互正交的特征向量。

4.12

$$A = U\Sigma V, A^T = V^T\Sigma^T U^T$$

$$\begin{pmatrix} 0 & A^T \\ A & 0 \end{pmatrix} X = \begin{pmatrix} 0 & A^T \\ A & 0 \end{pmatrix}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \lambda \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

$$A^T X_2 = \lambda X_1, AX_1 = \lambda X_2$$

$$\Rightarrow AA^T X_2 = \lambda A X_1, A^T A X_1 = \lambda A^T X_2$$

$$\Rightarrow AA^T X_2 = \lambda^2 X_2, A^T A X_1 = \lambda^2 X_1$$

$\Rightarrow X_1 \leftarrow V, X_2 \leftarrow U$（注：$\leftarrow$ 表示取U和V中的列向量)

$$\because AA^T = U\Sigma V(U\Sigma V)^T = U\Sigma VV^T\Sigma^T U^T = U\Sigma\Sigma^T U^T$$

$\therefore AA^T U = U\Sigma\Sigma^T U^T U = U\Sigma\Sigma^T$ (也就是奇异值分别乘上U的各列)

对于$A^T A, V$也有相同讨论

6.4

解 ：满秩分解 $A^+ = G^H (F^H A G^H)^{-1} F^H.$

$$\begin{pmatrix}1\\5\\7\end{pmatrix} = \begin{pmatrix}1\\5\\7\end{pmatrix}(1)$$

$$\begin{pmatrix}1\\5\\7\end{pmatrix}^+ = (1)\left(\begin{pmatrix}1 & 5 & 7\end{pmatrix}\begin{pmatrix}1\\5\\7\end{pmatrix}(1)\right)^{-1}\begin{pmatrix}1 & 5 & 7\end{pmatrix}$$
$$= \frac{1}{1^2+5^2+7^2}\begin{pmatrix}1 & 5 & 7\end{pmatrix}$$
$$= \frac{1}{75}\begin{pmatrix}1 & 5 & 7\end{pmatrix}$$
$$= \begin{pmatrix}0.133 & 0.0667 & 0.0933\end{pmatrix}$$

利用彭罗斯逆的性质 $A† = (A^H A)† A^H$，也可以。

Matlab:

```
>> pinv([1 5 7]')

ans =

    0.0133    0.0667    0.0933

>> pinv([1 5 7]')*75

ans =

    1.0000    5.0000    7.0000
```

6.5

按定义证明

$(1) AA^T(A(A^TA)^{-2}A^T)AA^T = AA^T$

$(2)(A(A^TA)^{-2}A^T)AA^T(A(A^TA)^{-2}A^T) = A(A^TA)^{-2}A^T$

$(3)(AA^T(A(A^TA)^{-2}A^T))^T = (A(A^TA)^{-2}A^T)^T AA^T = A(A^TA)^{-2}A^TAA^T = A(A^TA)^{-1}A^T$

$AA^T(A(A^TA)^{-2}A^T) = A(A^TA)^{-1}A^T$

(4)略

6.7

证明： $A = U\Sigma V, A^2 = U\Sigma^2 U^H = V\Sigma^2 V^H, A^{+2} = U^H \Sigma^{+2} U = V^H \Sigma^{+2} V$

$M = V^H \Sigma^+ U^H, M^2 = V^H \Sigma^{+2} V$

6.12

方程是相容的，有解，根据推论 1，和定理 1 的证明，可得证明步骤。

6.13

$$A\mathbf{x} + \varepsilon = \mathbf{b}$$
$$Q(x) = (A\mathbf{x} - \mathbf{b})^T W (A\mathbf{x} - \mathbf{b})$$
$$\mathbf{d}Q(x) = tr(\mathbf{d}(A\mathbf{x} - \mathbf{b})^T W (A\mathbf{x} - \mathbf{b}))$$
$$= tr(\mathbf{dx}^T A^T W (A\mathbf{x} - \mathbf{b})) + tr((A\mathbf{x} - \mathbf{b})^T W A \mathbf{dx})$$
$$= 2tr((A\mathbf{x} - \mathbf{b})^T W A \mathbf{dx})$$
$$\Rightarrow \nabla_{\mathbf{x}} Q = ((A\mathbf{x} - \mathbf{b})^T W A)^T$$
$$\Rightarrow ((A\mathbf{x} - \mathbf{b})^T W A)^T = 0$$
$$\Rightarrow A^T W^T (A\mathbf{x} - \mathbf{b}) = \mathbf{0}$$
$$\Rightarrow A^T W^T A\mathbf{x} - A^T W^T \mathbf{b} = \mathbf{0}$$
$$\Rightarrow A^T W^T A\mathbf{x} = A^T W^T \mathbf{b}$$
$$\mathbf{x} = (A^T W^T A)^+ A^T W^T \mathbf{b} (\text{极小范数最小二乘解})$$

$$\|\varepsilon\|_2^2 = \varepsilon^{\mathrm{T}} \varepsilon = \sqrt{\varepsilon^{\mathrm{T}} \varepsilon \varepsilon^{\mathrm{T}} \varepsilon}$$
$$\|\varepsilon\|_2^4 = \varepsilon^{\mathrm{T}} \varepsilon \varepsilon^{\mathrm{T}} \varepsilon (\text{极值属性一样})$$
$$\varepsilon^{\mathrm{T}} \varepsilon \varepsilon^{\mathrm{T}} \varepsilon = \varepsilon^{\mathrm{T}} (\varepsilon \varepsilon^{\mathrm{T}} - W + W) \varepsilon = \varepsilon^{\mathrm{T}} (\varepsilon \varepsilon^{\mathrm{T}} - W) \varepsilon + \varepsilon^{\mathrm{T}} W \varepsilon$$
$$= (A\mathbf{x} - b)^{\mathbf{T}} (\varepsilon \varepsilon^{\mathrm{T}} - W)(A\mathbf{x} - \mathbf{b}) + Q(\mathbf{x})$$
当把$\mathbf{x}$看成已经求得的解时，
$$\mathbf{E}((A\mathbf{x} - b)^{\mathbf{T}} (\varepsilon \varepsilon^{\mathrm{T}} - W)(A\mathbf{x} - \mathbf{b}) + Q(\mathbf{x}))$$
$$= (A\mathbf{x} - b)^{\mathbf{T}} (\mathbf{E} \varepsilon \varepsilon^{\mathrm{T}} - W)(A\mathbf{x} - \mathbf{b}) + Q(\mathbf{x})$$
$$= (A\mathbf{x} - b)^{\mathbf{T}} (R - W)(A\mathbf{x} - \mathbf{b}) + Q(\mathbf{x})$$
当$W = R$时$, \mathbf{E}\|\varepsilon\|_2^4 = \mathbf{E}\varepsilon^{\mathrm{T}} W \varepsilon$

7.9

$$\mathbf{d}f(\mathbf{x}) = \mathbf{d}tr(f(\mathbf{x})) = \mathbf{d}tr(a^T \mathbf{x}) = tr(a^T \mathbf{dx}))$$
$$\mathbf{d}^2 f(\mathbf{x}) = \mathbf{d}tr(a^T \mathbf{dx}) = tr(a^T \mathbf{d}(\mathbf{dx})) = 0 \Rightarrow H = 0$$
$$\mathbf{d}f(\mathbf{x}) = \mathbf{d}tr(f(\mathbf{x})) = \mathbf{d}tr(\mathbf{x}^T A \mathbf{x}) = tr(\mathbf{x}^T (A + A^T) \mathbf{dx}))$$
$$\mathbf{d}^2 f(\mathbf{x}) = \mathbf{d}tr(\mathbf{x}^T (A + A^T) \mathbf{dx})) = tr(\mathbf{dx}^T (A + A^T) \mathbf{dx})) \Rightarrow H = A + A^T$$

7.12

$$\mathbf{d}tr(AX^{-1}B) = \mathbf{d}tr(tr(AX^{-1}B)) = \mathbf{d}tr(AX^{-1}B) = tr(BA\mathbf{d}X^{-1})$$
$$= tr(-BAX^{-1}(\mathbf{d}X)X^{-1}) = tr(-X^{-1}BAX^{-1}(\mathbf{d}X))$$
$$\Rightarrow \nabla_X tr(AX^{-1}B) = -X^{-1}BAX^{-1}$$
$$\mathbf{d}tr(AX^T BXC) = \mathbf{d}tr(CAX^T BX) = tr(CA\mathbf{d}X^T BX) + tr(CAX^T B\mathbf{d}X)$$
$$= tr(BXCA\mathbf{d}X^T) + tr(CAX^T B\mathbf{d}X)$$
$$= tr((BXCA)^T \mathbf{d}X) + tr(CAX^T B\mathbf{d}X)$$
$$\Rightarrow \nabla_X tr(AX^T BXC) = BXCA)^T + CAX^T B$$