1、求右 Kronecker 积并验证 tr(A⊗B)=tr(A)tr(B)

**2、右$Kronecker$积：**$A \otimes B = [a_1 B, \ldots, a_n B]$。**左$Kronecker$积：**$B \otimes A = [A \otimes B]_{left} = [Ab_1, \ldots, Ab_q]$。**广义$Kronecker$积：**$\{A_i\}_N \otimes B = [A_1 b^1, \ldots, A_N b^N]^T$。

2、

**1、**$A^+$：**若$A$满列秩**，则$A^+ = L = (A^H A)^{-1} A^H$，**左伪逆矩阵。若$A$满行秩**，则$A^+ = R = A^H (AA^H)^{-1}$，**右伪逆矩阵。**

因为$Range(A) = Range(AA^+A) \subset Range(AA^+) \subset Range(A)$，所以$Range(AA^+) = Range(A)$，同时由于$AA^+AA^+ = AA^+$是幂等矩阵，所以$AA^+$是**$A$列空间的投影**。同理，$Range(A^H) = Range(A^H(A^+)^H A^H) \subset Range(A^H(A^+)^H) \subset Range(A^H)$，又因为$Range(A^H(A^+)^H) = Range(A^+A)$，所以$Range(A^+A) = Range(A^H)$，同时由于$A^+AA^+A = A^+A$是幂等矩阵，所以$A^+A$是$A^H$列空间的投影，即**$A$行空间的投影。**

3、

**3、奇异值分解：**$A = U\Sigma V^T$，**左奇异向量：**$(AA^T)u_i = \lambda_i u_i$、**右奇异向量：**$(A^TA)v_i = \lambda_i v_i$、**奇异值矩阵：**$\sigma_i = \frac{Av_i}{u_i}$。其中$U^TU = I, V^TV = I$。

**几何意义：**$V$表示原始域的标准正交基，$U$表示经过$A$变换后的辅助域的标准正交基。矩阵$A$的作用是将一个向量$X$先在$V$这组正交基向量的空间作旋转，然后对每个方向进行一定的缩放（缩放因子就是各个特征值），最后在$U$这组正交基向量的空间再次作旋转。

4、

$$[\frac{\partial tr(XB)}{\partial X^T}]_{ij} = \frac{\partial}{\partial x_{ji}} \sum_{p=1}^m \sum_{l=1}^n x_{pl} b_{lp} = \sum_{p=1}^m \sum_{l=1}^n b_{lp} \frac{\partial x_{pl}}{\partial x_{ji}} = b_{ij}，D_X tr(XB) = B。$$

5、

**②$KL$散度：**$D_{KL}(X||AS) = \sum_{i=1}^N \sum_{j=1}^M (x_{ij} ln \frac{x_{ij}}{[AS]_{ij}} - x_{ij} + [AS]_{ij})$的最小化问题。梯度为$\frac{\partial D_{KL}(X||AS)}{\partial a_{ik}} = \sum_{j=1}^M (-\frac{S_{kj}x_{ij}}{[AS]_{ij}} + S_{kj})$，$\frac{\partial D_{KL}(X||AS)}{\partial S_{kj}} = \sum_{i=1}^N (-\frac{a_{ik}x_{ij}}{[AS]_{ij}} + a_{ik})$，从而其梯度下降为$a_{ik} \leftarrow a_{ik} - \mu_{ik} \sum_{j=1}^M (-\frac{S_{kj}x_{ij}}{[AS]_{ij}} + S_{kj})$，$S_{kj} \leftarrow S_{kj} - \eta_{kj} \sum_{i=1}^N (-\frac{a_{ik}x_{ij}}{[AS]_{ij}} + a_{ik})$。令$\mu_{ik} = \frac{a_{ik}}{\sum_{j=1}^M S_{kj}}$，$\eta_{kj} = \frac{S_{kj}}{\sum_{i=1}^N a_{ik}}$，则改写为乘法运算$a_{ik} = a_{ik} \frac{\sum_{j=1}^M S_{kj}x_{ij}/[AS]_{ij}}{\sum_{j=1}^M S_{kj}}$，$S_{kj} \leftarrow S_{kj} \frac{\sum_{i=1}^N a_{ik}x_{ij}/[AS]_{ij}}{\sum_{i=1}^N a_{ik}}$。

**3、 $L_1$范数最小化的梯度分析**

一个实变元t∈R的**符号函数**定义为 $\mathrm{sgn}(t) = \begin{cases} +1, & t > 0 \\ 0, & t = 0 \\ -1, & t < 0 \end{cases}$

绝对值函数f(t)=|t|的次微分为 $\partial f(t) = \frac{\partial |t|}{\partial t} = \begin{cases} \{+1\}, & t > 0 \\ [-1,+1], & t = 0 \\ \{-1\}, & t < 0 \end{cases}$

于是，对$L_1$正则化最小二乘问题

$$\min_{\mathbf{x}} J(\lambda, \mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1 。$$

目标函数的梯度为

$$\nabla_{\mathbf{x}} J(\boldsymbol{\lambda}, \mathbf{x}) = \frac{\partial J(\boldsymbol{\lambda}, \mathbf{x})}{\partial \mathbf{x}} = -\mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x}) + \boldsymbol{\lambda}\nabla_{\mathbf{x}}\|\mathbf{x}\|_1 = -\mathbf{c} + \boldsymbol{\lambda}\nabla_{\mathbf{x}}\|\mathbf{x}\|_1 ,$$

其中，称$\mathbf{c}=\mathbf{A}^T(\mathbf{y}-\mathbf{A}\mathbf{x})$为**残差相关性**(residual correlation)向量；$\nabla_{\mathbf{x}}\|\mathbf{x}\|_1$为$L_1$范数$\|\mathbf{x}\|_1$的梯度向量，其第$i$分量为

$$\nabla_{x_i}\|\mathbf{x}\|_1 = \frac{\partial\|\mathbf{x}\|_1}{\partial x_i} = \begin{cases} \{+1\}, & x_i > 0 \\ [-1,+1], & x_i = 0 \\ \{-1\}, & x_i < 0 \end{cases} \quad (i = 1,2,\ldots,n)$$

从而，$L_1$正则化最小二乘问题的**平稳条件**为

$$\mathbf{c} = \lambda\nabla_{\mathbf{x}}\|\mathbf{x}\|_1 ,$$

或者为

$$c_i = \begin{cases} \{+\lambda\}, & x_i > 0 \\ [-\lambda,+\lambda], & x_i = 0 \\ \{-\lambda\}, & x_i < 0 \end{cases} \quad (i = 1,2,\ldots,n)$$

由于$L_1$范数最小化是一个凸函数优化问题，所以以上**平稳点条件也是**$L_1$范数最小化的最优解的**充分必要条件**。

6、

7、

**8、实对称矩阵的特征值是实数**，证明：设实对称矩阵$A$的复数特征值为$\lambda$，对应$n$维非零向量$X$，使得$AX=\lambda X$成立，设$\bar{\lambda}$为$\lambda$的共轭复数，因为$A$是实对称矩阵，所以$A=\bar{A}=A^T$，则$A\bar{X}=\bar{A}\bar{X}=\overline{AX}=\overline{\lambda X}=\bar{\lambda}\bar{X}$，同时左乘$\bar{X}^T AX=\bar{X}^T\lambda X=\lambda\bar{X}^T X$，又因为$\bar{X}^T AX=\bar{X}^T A^T X=(A\bar{X})^T X=(\bar{\lambda}\bar{X})^T X=\bar{\lambda}\bar{X}^T X$，即$(\lambda-\bar{\lambda})\bar{X}^T X=(\lambda-\bar{\lambda})[\bar{X},X]=0$，又因为非零向量$X$内积$>0$，即$\lambda=\bar{\lambda}$

**9、正定矩阵特征值大于0**，证明：设正定矩阵$A$有一个非正的特征值$k$，对应有非零特征向量$x$，则$Ax=kx$，但$x^H Ax=kx^H x<=0$,这与$A$正定矛盾。

8、

**10、L矩阵是半正定矩阵**，证明：对任意向量$f$，有

$$f^T Lf=f^T(D-W)f=f^T Df-f^T Wf=\sum_{i=1}^n d_i f_i^2-\sum_{i,j=1}^n w_{ij}f_i f_j=\frac{1}{2}(\sum_{i=1}^n d_i f_i^2-2\sum_{i,j=1}^n w_{ij}f_i f_j+\sum_{j=1}^n d_j f_j^2)$$

$$=\frac{1}{2}[\sum_{i=1}^n(\sum_{j=1}^n w_{ij})f_i^2-2\sum_{i,j=1}^n w_{ij}f_i f_j+\sum_{j=1}^n(\sum_{i=1}^n w_{ji})f_j^2]=\frac{1}{2}[\sum_{i=1}^n(\sum_{j=1}^n w_{ij})f_i^2-2\sum_{i,j=1}^n w_{ij}f_i f_j+\sum_{j=1}^n(\sum_{i=1}^n w_{ij})f_j^2]$$

$$=\frac{1}{2}(\sum_{i,j=1}^n w_{ij}f_i^2-2\sum_{i,j=1}^n w_{ij}f_i f_j+\sum_{i,j=1}^n w_{ij}f_j^2)=\frac{1}{2}\sum_{i,j=1}^n w_{ij}(f_i^2-2f_i f_j+f_j^2)=\frac{1}{2}\sum_{i,j=1}^n w_{ij}(f_i-f_j)^2$$，因为矩阵$W$非负，所以$f^T Lf\geqslant 0$

9、

**例2:** 令m×n矩阵A的奇异值分解为$\mathbf{U\Lambda}V^T$，则以下极小化问题

$$\min_{C\in R^{n\times n}}\|\mathbf{C}\|_*+\frac{\tau}{2}\|\mathbf{A}-\mathbf{AC}\|_F^2$$

最优解为$\hat{\mathbf{C}}=\mathbf{V}_1(\mathbf{I}-\frac{1}{\tau}\mathbf{\Lambda}_1^{-2})\mathbf{V}_1^T$，其中$\mathbf{U}=[\mathbf{U}_1,\mathbf{U}_2]$，$\mathbf{\Lambda}=diag(\mathbf{\Lambda}_1,\mathbf{\Lambda}_2)$和$\mathbf{V}=[\mathbf{V}_1,\mathbf{V}_2]$分别为左奇异矩阵$\mathbf{U}$、对角矩阵$\mathbf{\Lambda}$及右奇异矩阵$\mathbf{V}$对应于指标集 $\mathbf{I}_1=\{i:\lambda_i>1/\sqrt{\tau}\}$ 和 $\mathbf{I}_2=\{i:\lambda_i\le 1/\sqrt{\tau}\}$划分。且最优值为

$$f(\hat{\mathbf{C}})=\sum_{i\in\mathbf{I}_1}(1-\frac{1}{2\tau}\lambda_i^{-2})+\frac{\tau}{2}\sum_{i\in\mathbf{I}_2}\lambda_i^2\text{。}$$

**证明:** 为了说明$\hat{\mathbf{C}}$是极小解，由于目标函数为$\mathbf{C}$的凸函数，故只要说明目标函数在$\hat{\mathbf{C}}$的次梯度包含$\mathbf{0}$，即

$$0\in\partial\|\hat{\mathbf{C}}\|_*-\tau\mathbf{A}^T\mathbf{A}(\mathbf{I}-\hat{\mathbf{C}})\text{。}$$

由于矩阵$\mathbf{C}$的核范数的次梯度

$$\partial\|\mathbf{C}\|_*=\{\mathbf{U_C V_C^T}+\mathbf{W}:\ \mathbf{W}\in R^{n\times n},\mathbf{U_C^T W}=\mathbf{0},\mathbf{WV_C}=\mathbf{0},\|\mathbf{W}\|_{spec}\le 1\},$$

其中$\mathbf{U_C}$和$\mathbf{V_C}$分别为矩阵$\mathbf{C}$于截尾左、右奇异值矩阵。

令 $\mathbf{U_C}=\mathbf{V_C}=\mathbf{V}_1$，则

$$\mathbf{I}-\hat{\mathbf{C}}=(\mathbf{V}_1\mathbf{V}_1^T+\mathbf{V}_2\mathbf{V}_2^T)-\mathbf{V}_1(\mathbf{I}-\frac{1}{\tau}\mathbf{\Lambda}_1^{-2})\mathbf{V}_1^T=\underline{\frac{1}{\tau}\mathbf{V}_1\mathbf{\Lambda}_1^{-2}\mathbf{V}_1^T+\mathbf{V}_2\mathbf{V}_2^T}\quad\text{和}\quad\mathbf{A}^T\mathbf{A}=\mathbf{V\Lambda}^2\mathbf{V}^T=\underline{\mathbf{V}_1\mathbf{\Lambda}_1^2\mathbf{V}_1^T+\mathbf{V}_2\mathbf{\Lambda}_2^2\mathbf{V}_2^T},$$

从而

$$\mathbf{V}_1\mathbf{V}_1^T+\mathbf{W}-\tau\mathbf{A}^T\mathbf{A}(\mathbf{I}-\hat{\mathbf{C}})=\mathbf{V}_1\mathbf{V}_1^T+\mathbf{W}-\tau(\frac{1}{\tau}\mathbf{V}_1\mathbf{V}_1^T+\mathbf{V}_2\mathbf{\Lambda}_2^2\mathbf{V}_2^T)=\mathbf{W}-\tau\mathbf{V}_2\mathbf{\Lambda}_2^2\mathbf{V}_2^T\text{。}$$

于是，令$\mathbf{W}=\tau\mathbf{V}_2\mathbf{\Lambda}_2^2\mathbf{V}_2^T$，则 $\mathbf{V}_1\mathbf{V}_1^T+\mathbf{W}-\tau\mathbf{A}^T\mathbf{A}(\mathbf{I}-\hat{\mathbf{C}})=\mathbf{0}$，且$\mathbf{V}_1^T\mathbf{W}=\mathbf{0}$，$\mathbf{WV}_1=\mathbf{0}$ 和 $\|\mathbf{W}\|_{spec}=\|\tau\mathbf{\Lambda}_2^2\|_{spec}\le 1$。

此表明目标函数在$\hat{\mathbf{C}}$的次梯度包含$\mathbf{0}$。又因为目标函数为凸函数，所以$\hat{\mathbf{C}}$是优化问题的极小解。

最优目标函数值为

$$f(\hat{\mathbf{C}})=\|\hat{\mathbf{C}}\|_*+\frac{\tau}{2}\|\mathbf{A}-\mathbf{A}\hat{\mathbf{C}}\|_F^2=\left\|\mathbf{I}-\frac{1}{\tau}\mathbf{\Lambda}_1^{-2}\right\|_*+\frac{\tau}{2}\left\|\frac{1}{\tau}\mathbf{U}_1\mathbf{\Lambda}_1^{-1}\mathbf{V}_1^T+\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{V}_2^T\right\|_F^2$$

$$=\sum_{i\in\mathbf{I}_1}(1-\frac{1}{\tau}\lambda_i^{-2})+\frac{\tau}{2}(\sum_{i\in\mathbf{I}_1}\frac{1}{\tau^2}\lambda_i^{-2}+\sum_{i\in\mathbf{I}_2}\lambda_i^2)=\sum_{i\in\mathbf{I}_1}(1-\frac{1}{2\tau}\lambda_i^{-2})+\frac{\tau}{2}\sum_{i\in\mathbf{I}_2}\lambda_i^2\text{。}$$

不可微函数极小点的一阶条件：
点$\mathbf{x}^*$是凸函数$f(\mathbf{x})$的一个极小点
<u>当且仅当</u>$f(\mathbf{x})$在$\mathbf{x}^*$次可微的，
且$0\in\partial f(\mathbf{x}^*)$。