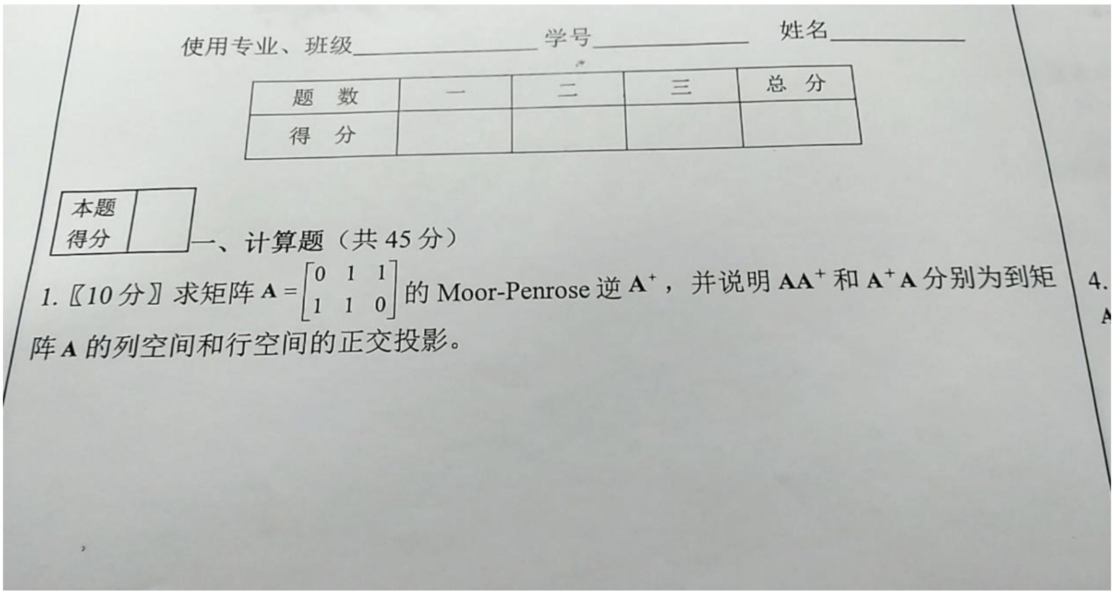
第一题课后习题 278



使用专业、班级________ 学号________ 姓名________

题 数	一	二	三	总 分
得 分				

本题得分	

一、计算题（共 45 分）

1.〖10 分〗求矩阵 $\mathbf{A}=\begin{bmatrix}0&1&1\\1&1&0\end{bmatrix}$ 的 Moor-Penrose 逆 $\mathbf{A}^{+}$，并说明 $\mathbf{A}\mathbf{A}^{+}$ 和 $\mathbf{A}^{+}\mathbf{A}$ 分别为到矩阵 $\mathbf{A}$ 的列空间和行空间的正交投影。

第二题 ppt

二、广义Rayleigh商

(1) 广义Rayleigh商的定义与性质

设矩阵$\mathbf{A}$和$\mathbf{B}$均为n×n维Hermitian矩阵，且$\mathbf{B}$是正定的。矩阵对$(\mathbf{A},\mathbf{B})$的广义Rayleigh商R($\mathbf{x}$)定义为：

$$R(\mathbf{x})=\frac{\mathbf{x}^H\mathbf{A}\mathbf{x}}{\mathbf{x}^H\mathbf{B}\mathbf{x}}, \tag{5.4}$$

其中，$\mathbf{x}$是待选择的向量，目的是使广义Rayleigh商最大化或最小化。

为了求解广义Rayleigh商R($\mathbf{x}$)，定义一个新向量 $\tilde{\mathbf{x}}=B^{1/2}\mathbf{x}$，代入(5.4)得

$$R(\tilde{\mathbf{x}})=\frac{\tilde{\mathbf{x}}^H(\mathbf{B}^{-1/2})^H\mathbf{A}(\mathbf{B}^{-1/2})\tilde{\mathbf{x}}}{\tilde{\mathbf{x}}^H\tilde{\mathbf{x}}},$$

其中$B^{1/2}$表示矩阵B的平方根。这样，矩阵对$(\mathbf{A},\mathbf{B})$的广义Rayleigh商**等价为**矩阵$(\mathbf{B}^{-1/2})^H\mathbf{A}(\mathbf{B}^{-1/2})$的标准Rayleigh商，从而广义Rayleigh商的最小值为该矩阵的最小特征，最大值为该矩阵的最大特征值。

第三题

0 分】求矩阵 $\mathbf{A} = \begin{bmatrix}3 & 3\\ -4 & 4\end{bmatrix}$ 的奇异值分解，并说明此分解的几何意义。

$\begin{bmatrix}-4 & 4\end{bmatrix}$ 的奇异值分解，并说明此分解的几何意义。

解. $AA^T = \begin{pmatrix}3 & 3\\ -4 & 4\end{pmatrix}\begin{pmatrix}3 & -4\\ 3 & 4\end{pmatrix} = \begin{pmatrix}18 & 0\\ 0 & 32\end{pmatrix}$

特征值 $\lambda_1 = 18$ $\lambda_2 = 32$ 对应的特征向量分别为 $\begin{pmatrix}1\\0\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix}$

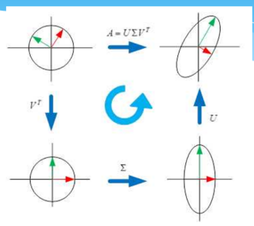
A^TA 为 $\begin{pmatrix}3 & -4\\ 3 & 4\end{pmatrix}\begin{pmatrix}3 & 3\\ -4 & 4\end{pmatrix} = \begin{pmatrix}25 & -7\\ -7 & 25\end{pmatrix}$

特征值为 $\lambda_1 = 32$ $\lambda_2 = 18$ 对应的特征向量分别为 $\begin{pmatrix}-0.7071 & 0.7071\\ 0.7071 & 0.7071\end{pmatrix}$

$\therefore$ A的奇异值分解为 $\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}18 & \\ & 32\end{pmatrix}\begin{pmatrix}-0.7071 & 0.7071\\ 0.7071 & 0.7071\end{pmatrix}$

(7)矩阵奇异值分解的几何与物理意义

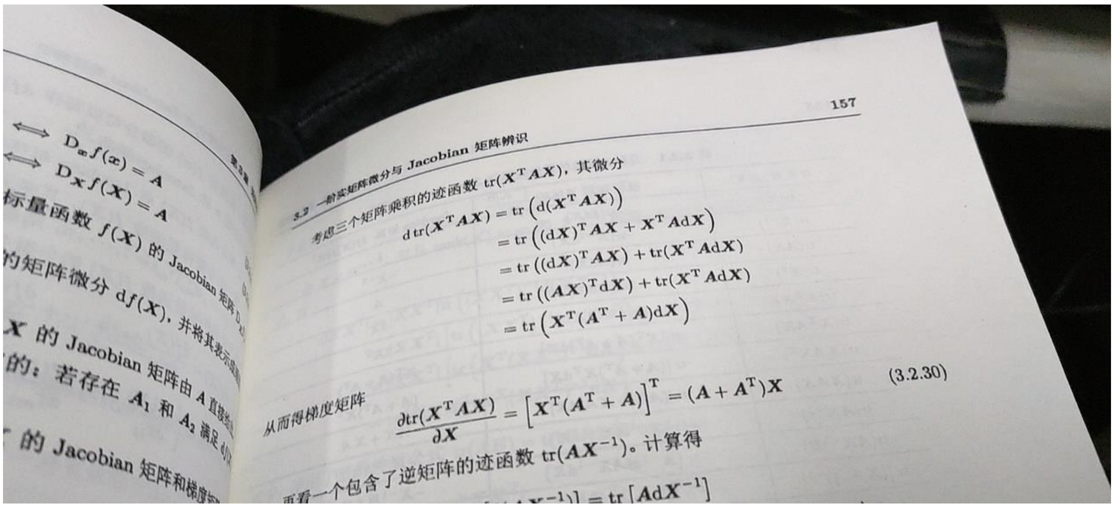
由矩阵 **A** 的奇异值分解 $\mathbf{A}=\mathbf{U}\Sigma\mathbf{V}^T$ 可知，**A** 可以分解成三个矩阵相乘的形式，其中 **V** 表示了原始域的标准正交基，**U** 表示经过 **A** 变换后的辅助域的标准正交基，Σ 表示了 **V** 中的向量与 **U** 中相对应向量之间的关系。



由上图发现，线性变换 **Ax** 可以分解为**旋转**、**缩放**、**旋转**这三种基本线性变换，即矩阵 **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**（缩放因子就是各个奇异值）；最后在 **U** 这组正交基向量的空间再次作**旋转**。

因此，奇异值分解是将一个矩阵原本混合在一起的三种作用效果分解出来。

第四题 157



157

3.2 一阶实矩阵微分与 Jacobian 矩阵辨识

$\Longleftrightarrow \quad \mathrm{D}_{\boldsymbol{x}} f(\boldsymbol{x}) = \boldsymbol{A}$

$\Longleftrightarrow \quad \mathrm{D}_{\boldsymbol{X}} f(\boldsymbol{X}) = \boldsymbol{A}$

标量函数 $f(\boldsymbol{X})$ 的 Jacobian 矩阵

的矩阵微分 $\mathrm{d}f(\boldsymbol{X})$，并将其表示成

$\boldsymbol{X}$ 的 Jacobian 矩阵由 $\boldsymbol{A}$ 直接给出

的：若存在 $\boldsymbol{A}_1$ 和 $\boldsymbol{A}_2$ 满足

的 Jacobian 矩阵和梯度矩阵

考虑三个矩阵乘积的迹函数 $\mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})$，其微分

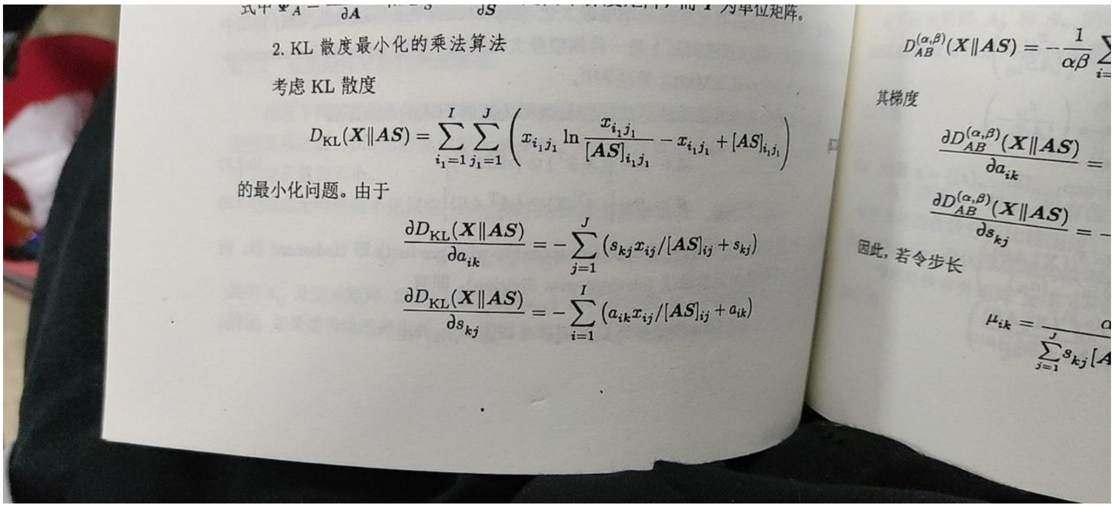
$$
\begin{aligned}
\mathrm{d}\,\mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}) &= \mathrm{tr}\left(\mathrm{d}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})\right) \\
&= \mathrm{tr}\left((\mathrm{d}\boldsymbol{X})^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\mathrm{d}\boldsymbol{X}\right) \\
&= \mathrm{tr}\left((\mathrm{d}\boldsymbol{X})^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}\right) + \mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\mathrm{d}\boldsymbol{X}) \\
&= \mathrm{tr}\left((\boldsymbol{A}\boldsymbol{X})^{\mathrm{T}}\mathrm{d}\boldsymbol{X}\right) + \mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\mathrm{d}\boldsymbol{X}) \\
&= \mathrm{tr}\left(\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{A})\mathrm{d}\boldsymbol{X}\right)
\end{aligned}
$$

从而得梯度矩阵

$$
\frac{\partial \mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})}{\partial \boldsymbol{X}} = \left[\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{A})\right]^{\mathrm{T}} = (\boldsymbol{A} + \boldsymbol{A}^{\mathrm{T}})\boldsymbol{X} \tag{3.2.30}
$$

再看一个包含了逆矩阵的迹函数 $\mathrm{tr}(\boldsymbol{A}\boldsymbol{X}^{-1})$。计算得

第五题 366



2. KL 散度最小化的乘法算法

考虑 KL 散度

$$
D_{\mathrm{KL}}(\boldsymbol{X}\|\boldsymbol{A}\boldsymbol{S}) = \sum_{i_1=1}^{I}\sum_{j_1=1}^{J}\left(x_{i_1j_1}\ln\frac{x_{i_1j_1}}{[\boldsymbol{A}\boldsymbol{S}]_{i_1j_1}} - x_{i_1j_1} + [\boldsymbol{A}\boldsymbol{S}]_{i_1j_1}\right)
$$

的最小化问题。由于

$$
\frac{\partial D_{\mathrm{KL}}(\boldsymbol{X}\|\boldsymbol{A}\boldsymbol{S})}{\partial a_{ik}} = -\sum_{j=1}^{J}\left(s_{kj}x_{ij}/[\boldsymbol{A}\boldsymbol{S}]_{ij} + s_{kj}\right)
$$

$$
\frac{\partial D_{\mathrm{KL}}(\boldsymbol{X}\|\boldsymbol{A}\boldsymbol{S})}{\partial s_{kj}} = -\sum_{i=1}^{I}\left(a_{ik}x_{ij}/[\boldsymbol{A}\boldsymbol{S}]_{ij} + a_{ik}\right)
$$

其梯度

因此，若令步长

所以梯度下降算法为

$$a_{ik} \leftarrow a_{ik} - \mu_{ik} \times \left[-\sum_{j=1}^{J} \left(s_{kj} x_{ij} / [\boldsymbol{AS}]_{ij} + s_{kj} \right) \right]$$

$$s_{kj} \leftarrow s_{kj} - \eta_{kj} \times \left[-\sum_{i=1}^{I} \left(a_{ik} x_{ij} / [\boldsymbol{AS}]_{ij} + a_{ik} \right) \right]$$

若令

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{J} s_{kj}}, \quad \eta_{kj} = \frac{1}{\sum_{i=1}^{I} a_{ik}}$$

则梯度下降算法可以改写为乘法算法[308]

$$a_{ik} \leftarrow a_{ik} \frac{\sum_{j=1}^{J} s_{kj} x_{ik} / [\boldsymbol{AS}]_{ik}}{\sum_{j=1}^{J} s_{kj}} \tag{6.7.15}$$

$$s_{kj} \leftarrow s_{kj} \frac{\sum_{i=1}^{I} a_{ik} x_{ik} / [\boldsymbol{AS}]_{ik}}{\sum_{i=1}^{I} a_{ik}} \tag{6.7.16}$$

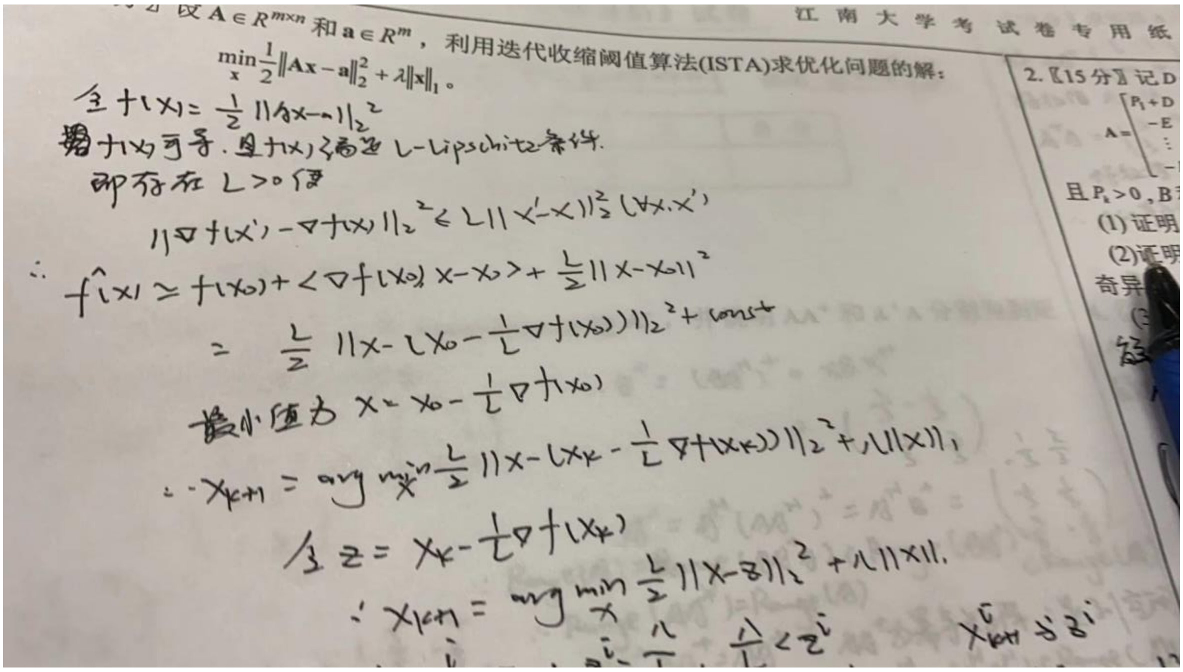
其矩阵形式为

$$\boldsymbol{A} \leftarrow \left[\boldsymbol{A} \oslash \left[\boldsymbol{1}_I \otimes (\boldsymbol{S}\boldsymbol{1}_K)^{\mathrm{T}} \right] \right] * \left[[\boldsymbol{X} \oslash (\boldsymbol{AS})]\, \boldsymbol{S}^{\mathrm{T}} \right] \tag{6.7.17}$$

$$\boldsymbol{S} \leftarrow \left[\boldsymbol{S} \oslash \left[(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{1}_I) \otimes \boldsymbol{1}_K \right] \right] * \left[\boldsymbol{A}^{\mathrm{T}} [\boldsymbol{X} \oslash (\boldsymbol{AS})] \right] \tag{6.7.18}$$

式中 $\boldsymbol{1}_I$ 是全部元素为 1 的 I 维列向量。

第六题



江南大学考试卷专用纸

设 $A\in R^{m\times n}$ 和 $a\in R^{m}$，利用迭代收缩阈值算法(ISTA)求优化问题的解：

$$\min_{x}\frac{1}{2}\|Ax-a\|_2^2+\lambda\|x\|_1。$$

令 $f(x)=\frac{1}{2}\|Ax-a\|_2^2$

易得 $f(x)$ 可导，且 $f(x)$ 满足 L-Lipschitz 条件，

即存在 $L>0$ 使

$$\|\nabla f(x')-\nabla f(x)\|_2^2\le L\|x'-x\|_2^2\ (\forall x,x')$$

$$\therefore\ \hat{f}(x)\simeq f(x_0)+\langle\nabla f(x_0),x-x_0\rangle+\frac{L}{2}\|x-x_0\|^2$$

$$=\frac{L}{2}\|x-(x_0-\frac{1}{L}\nabla f(x_0))\|_2^2+\text{const}$$

最小值为 $x=x_0-\frac{1}{L}\nabla f(x_0)$

$$\therefore\ x_{k+1}=\arg\min_{x}\frac{L}{2}\|x-(x_k-\frac{1}{L}\nabla f(x_k))\|_2^2+\lambda\|x\|_1$$

令 $z=x_k-\frac{1}{L}\nabla f(x_k)$

$$\therefore\ x_{k+1}=\arg\min_{x}\frac{L}{2}\|x-z\|_2^2+\lambda\|x\|_1$$

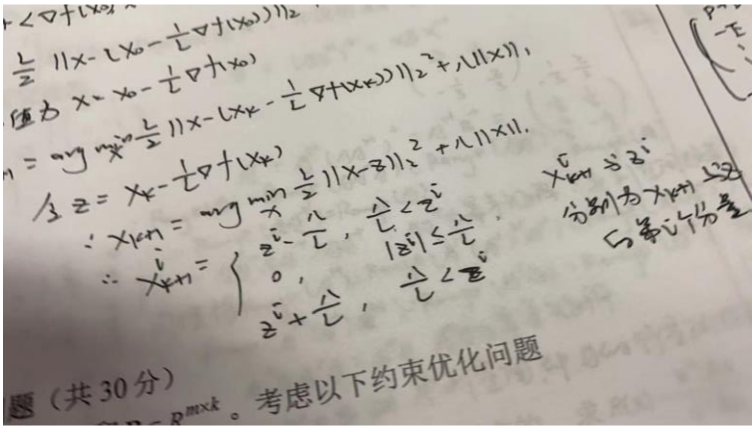
2.【15分】记D

且 $P_k>0$，B

(1) 证明

(2) 证明

奇异



$$=\frac{L}{2}\|x-(x_0-\frac{1}{L}\nabla f(x_0))\|_2^2$$

最小值为 $x=x_0-\frac{1}{L}\nabla f(x_0)$

$$x_{k+1}=\arg\min_{x}\frac{L}{2}\|x-(x_k-\frac{1}{L}\nabla f(x_k))\|_2^2+\lambda\|x\|_1$$

令 $z=x_k-\frac{1}{L}\nabla f(x_k)$

$$\therefore\ x_{k+1}=\arg\min_{x}\frac{L}{2}\|x-z\|_2^2+\lambda\|x\|_1$$

x_{k+1}^{i} 与 z^{i} 分别为 x_{k+1} 与 z 第 i 个分量

$$\therefore\ x_{k+1}^{i}=\begin{cases} z^{i}-\frac{\lambda}{L}, & \frac{\lambda}{L}<z^{i}\\ 0, & |z^{i}|\le\frac{\lambda}{L}\\ z^{i}+\frac{\lambda}{L}, & \frac{\lambda}{L}<z^{i}\end{cases}$$

题（共30分）

$B\in R^{m\times k}$。考虑以下约束优化问题

第七题

四、奇异值分解的一个重要应用(Sparse Principal Component Analysis, SPCA, 2012)

令 $\mathbf{M}_{n\times p}$和$\mathbf{N}_{n\times k}$两个矩阵。考虑约束极小化问题

$$\min_{\mathbf{A}\in R^{p\times k}} \|\mathbf{M}-\mathbf{N}\mathbf{A}^T\|_F^2$$
$$s.t.\quad \mathbf{A}^T\mathbf{A}=\mathbf{I}_{k\times k}。$$

假设$\mathbf{M}^T\mathbf{N}$ 的截尾SVD为 $\mathbf{U}\mathbf{D}\mathbf{V}^T$,其中$\mathbf{D}=diag(d_1,d_2,\cdots,d_r),d_i>0$,$r=\text{rank}\ (\mathbf{M}^T\mathbf{N})$，则优化问题的最优解为

$$\hat{\mathbf{A}}=\mathbf{U}\mathbf{V}^T。$$

证明： 由于$\mathbf{A}^T\mathbf{A}=\mathbf{I}$，所以

$$\|\mathbf{M}-\mathbf{N}\mathbf{A}^T\|_F^2=tr((\mathbf{M}-\mathbf{N}\mathbf{A}^T)^T(\mathbf{M}-\mathbf{N}\mathbf{A}^T))=\underline{tr(\mathbf{M}^T\mathbf{M})}-2tr(\mathbf{M}^T\mathbf{N}\mathbf{A}^T)+\underline{tr(\mathbf{N}^T\mathbf{N})}。$$

于是，极小化$\|\mathbf{M}-\mathbf{N}\mathbf{A}^T\|_F^2$只需极大化$tr(\mathbf{M}^T\mathbf{N}\mathbf{A}^T)$即可。利用$\mathbf{M}^T\mathbf{N}$的SVD有

$$tr(\mathbf{M}^T\mathbf{N}\mathbf{A}^T)=tr(\mathbf{U}\mathbf{D}\mathbf{V}^T\mathbf{A}^T)=tr(\mathbf{U}\mathbf{D}\mathbf{A}^{*T})=tr(\mathbf{A}^{*T}\mathbf{U}\mathbf{D}),$$

其中$\mathbf{A}^*=\mathbf{A}\mathbf{V}$。由于$\mathbf{V}$是$k\times r$列正交的且$\mathbf{A}^T\mathbf{A}=\mathbf{I}$，所以 $\mathbf{A}^{*T}\mathbf{A}^*=\mathbf{I}_{r\times r}$，即$\mathbf{A}^*$的列正交。

既然 $\mathbf{D}$为正对角元的对角矩阵，所以 $tr(\mathbf{A}^{*T}\mathbf{U}\mathbf{D})$ 为 $\mathbf{A}^{*T}\mathbf{U}$ 的对角元的非负线性组合(系数为奇异值)，从而当 $\mathbf{A}^{*T}\mathbf{U}$的对角元均为极大时，$tr(\mathbf{A}^{*T}\mathbf{U}\mathbf{D})$才达到极大。而 $\mathbf{A}^{*T}\mathbf{U}$ 的对角元分别为

$$a_1^{*T}u_1,a_2^{*T}u_2,\cdots,a_r^{*T}u_r,$$

其中 a_j^* 和 u_j分别为$\mathbf{A}^*$和$\mathbf{U}$的第j列，j=1,2,…,r。

令 $\mathbf{M}_{n\times p}$和$\mathbf{N}_{n\times k}$两个矩阵。考虑约束极小化问题

$$\min_{\mathbf{A}\in R^{p\times k}} \|\mathbf{M}-\mathbf{N}\mathbf{A}^T\|_F^2$$
$$s.t.\quad \mathbf{A}^T\mathbf{A}=\mathbf{I}_{k\times k}。$$

假设$\mathbf{M}^T\mathbf{N}$ 的截尾SVD为 $\mathbf{U}\mathbf{D}\mathbf{V}^T$，其中 $\mathbf{D}=diag(d_1,d_2,\cdots,d_r),d_i>0$,$r=\text{rank}\ (\mathbf{M}^T\mathbf{N})$，则优化问题的最优解为

$$\hat{\mathbf{A}}=\mathbf{U}\mathbf{V}^T。$$

$$\boxed{\mathbf{A}^{*T}\mathbf{A}^*=\mathbf{I}_{r\times r}\quad \mathbf{U}^T\mathbf{U}=\mathbf{I}_{r\times r}}$$

由Cauchy-Schwartz不等式得

$$a_j^{*T}u_j\leqslant\|a_j^*\|\times\|u_j\|=1\times1=1，当且仅当时\ a_j^*=u_j\ 时等号成立。$$

此表明，当 $a_j^*=u_j$时 $a_j^{*T}u_j$达到极大值1，即当$\mathbf{A}^*=\mathbf{U}$时 $\mathbf{A}^{*T}\mathbf{U}$的对角元均达到极大值1。

于是，由$\mathbf{A}^*=\mathbf{A}\mathbf{V}$得优化问题的极小解即 $tr(\mathbf{M}^T\mathbf{N}\mathbf{A}^T)$的极大解为

$$\hat{\mathbf{A}}=\mathbf{A}^*\mathbf{V}^T=\mathbf{U}\mathbf{V}^T。$$

$$\boxed{tr(\mathbf{M}^T\mathbf{N}\mathbf{A}^T)=tr(\mathbf{U}\mathbf{D}\mathbf{V}^T\mathbf{A}^T)=tr(\mathbf{U}\mathbf{D}\mathbf{A}^{*T})=tr(\mathbf{A}^{*T}\mathbf{U}\mathbf{D})}$$

$$\boxed{\mathbf{A}^{*T}\mathbf{U}\ 的对角元分别为\ a_1^{*T}u_1,a_2^{*T}u_2,\cdots,a_r^{*T}u_r}$$

第八题

记不清了