Problem 1

(a)$\forall A, B \in \mathbb{R}^{n\times n}$，以及满足$||\vec{x}|| = 1$的$\vec{x}$，我们有

$$\begin{aligned} ||(A+B)\vec{x}|| &= ||A\vec{x} + B\vec{x}|| \\ &\leq ||A\vec{x}|| + ||B\vec{x}|| \\ &\leq ||A|| + ||B|| \end{aligned}$$

其中最后一个不等号是由$||A||$的定义。

所以

$$||A+B|| = \max\{||(A+B)\vec{x}|| : ||\vec{x}|| = 1\} \leq ||A|| + ||B||$$

(b)利用等价定义：

$$||A|| = \max_{\vec{x}\in\mathbb{R}^n\setminus\{0\}} \frac{||A\vec{x}||}{||\vec{x}||}$$

那么显然有

$$\begin{aligned} \frac{||A\vec{x}||}{||\vec{x}||} &\leq ||A|| \\ ||A\vec{x}|| &\leq ||A||.\,||\vec{x}|| \end{aligned}$$

任取$\vec{x} \in \mathbb{R}^n$，我们有

$$\begin{aligned} ||AB\vec{x}|| &= ||A(B\vec{x})|| \\ &\leq ||A||.\,||B\vec{x}|| \\ &\leq ||A||.\,||B||.\,||\vec{x}|| \end{aligned}$$

所以

$$\frac{||AB\vec{x}||}{||\vec{x}||} \leq ||A||.\,||B||$$

对左边取最大值可得：

$$||AB|| \leq ||A||.\,||B||$$

(c)原不等式等价于

$$||A^k|| \geq |\lambda|^k$$

对A的特征值λ_i，取对应的特征向量$\vec{x}_i \in \mathbb{R}^n$，我们有

$$\begin{aligned}||A^k\overrightarrow{x_i}|| &= ||A^{k-1}\lambda_i\overrightarrow{x_i}|| \\ &= |\lambda_i|||A^{k-1}\overrightarrow{x_i}|| \\ &= \dots \\ &= |\lambda_i|^k.||\overrightarrow{x_i}||\end{aligned}$$

所以

$$\frac{||A^k\overrightarrow{x_i}||}{||\vec{x}||} = |\lambda_i|^k$$

因此对任意特征值λ_i，我们有

$$||A^k|| = \max_{\vec{x}\in\mathbb{R}^n\setminus\{0\}} \frac{||A^k\vec{x}||}{||\vec{x}||} \geq |\lambda_i|^k$$

(d)我们假设$||\vec{x}||_1 = 1$，即

$$\sum_{i=1}^{n}|x_i| = 1$$

计算$||A\vec{x}||_1$可得:

$$\begin{aligned}||A\vec{x}||_1 &= \sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}||x_j| \\ &= \sum_{j=1}^{n}\sum_{i=1}^{n}|a_{ij}||x_j| \\ &\leq \sum_{j=1}^{n}(\max_{1\leq j\leq n}\sum_{i=1}^{n}|a_{ij}|)|x_j| \\ &= (\max_{1\leq j\leq n}\sum_{i=1}^{n}|a_{ij}|)\sum_{j=1}^{n}|x_j| \\ &= \max_{1\leq j\leq n}\sum_{i=1}^{n}|a_{ij}|\end{aligned}$$

所以我们有

$$||A||_1 = \max_{1\leq j\leq n}\sum_{i=1}^{n}|a_{ij}|$$

补充题:

由(c)可得，对任意特征值λ_i，我们有

$$\lim_{k\to\infty}||A^k||^{\frac{1}{k}} \geq |\lambda_i|$$

所以

$$\lim_{k\to\infty}||A^k||^{\frac{1}{k}}\geq \max\{|\lambda_i|\}=\rho(A)$$

接下来证明另一个方向的不等式，证明参考维基百科

$$\lim_{k\to\infty}A^k=0\Leftrightarrow \rho(A)<1$$

注意到对于任意矩阵A，假设其特征值为

$$\lambda_1,\ldots,\lambda_n$$

那么$\frac{A}{k}$的特征为

$$\frac{\lambda_1}{k},\ldots,\frac{\lambda_n}{k}$$

所以$\forall\epsilon>0$，构造如下矩阵

$$A_+=\frac{1}{\rho(A)+\epsilon}A$$

由之前叙述可得

$$\rho(A_+)<1$$

所以

$$\lim_{k\to\infty}A_+^k=0$$

所以由极限的定义可得，存在N_+，当$k\geq N_+$时，我们有

$$||A_+^k||=\frac{||A^k||}{(\rho(A)+\epsilon)^k}<1$$

从而

$$||A^k||^{\frac{1}{k}}\leq\rho(A)+\epsilon$$

结合之前的结果可得

$$\rho(A)\leq||A^k||^{\frac{1}{k}}\leq\rho(A)+\epsilon$$

令$\epsilon\to 0$，我们有

$$\rho(A)=||A^k||^{\frac{1}{k}}$$

Problem 2

(a)对于固定的k，记

$$E_{c,l}=I+c\vec{e}_l\vec{e}_k^T$$

我们计算$E_{c_1,s}E_{c_2,s}$

$$\begin{aligned}E_{c_1,s}E_{c_2,t} &= (I+c_1\vec{e}_s\vec{e}_k^T)(I+c_2\vec{e}_t\vec{e}_k^T)\\ &= I+(c_1\vec{e}_s+c_2\vec{e}_t)\vec{e}_k^T+c_1c_2\vec{e}_s\vec{e}_k^T\vec{e}_t\vec{e}_k^T\\ &= I+(c_1\vec{e}_s+c_2\vec{e}_t)\vec{e}_k^T+c_1c_2\vec{e}_s(\vec{e}_k^T\vec{e}_t)\vec{e}_k^T\\ &= I+(c_1\vec{e}_s+c_2\vec{e}_t)\vec{e}_k^T\end{aligned}$$

注意到forward substitution等价于左乘矩阵$E_{c,l}, l > k$，结合上述事实，我们有

$$\begin{aligned}M_k &= \prod_{i=k+1}^{n} E_{c_i,i}\\ &= I+(\sum_{i=k+1}^{n} c_i\vec{e}_i)\vec{e}_k^T\end{aligned}$$

记

$$\vec{m}_k = -(\sum_{i=k+1}^{n} c_i\vec{e}_i)$$

第一部分结论得证。接着验证

$$L_kM_k = I$$

事实上，我们有

$$\begin{aligned}L_kM_k &= (I+\vec{m}_k\vec{e}_k^T)(I-\vec{m}_k\vec{e}_k^T)\\ &= I+\vec{m}_k\vec{e}_k^T-\vec{m}_k\vec{e}_k^T+\vec{m}_k(\vec{e}_k^T\vec{m}_k)\vec{e}_k^T\\ &= I+\vec{m}_k(\vec{e}_k^T\vec{m}_k)\vec{e}_k^T\end{aligned}$$

注意到$\vec{e}_k$的第k个元素为1，$\vec{m}_k$第k个元素为0，所以

$$\vec{m}_k(\vec{e}_k^T\vec{m}_k)\vec{e}_k^T = \vec{0}$$

因此

$$L_kM_k = I$$

(b)

$$\begin{aligned}L_kP^{(ij)} &= (I+\vec{m}_k\vec{e}_k^T)P^{(ij)}\\ &= P^{(ij)}+\vec{m}_k\left(\vec{e}_k^TP^{(ij)}\right)\end{aligned}$$

因为$\vec{e}_k^TP^{(ij)}$的作用是交换$\vec{e}_k^T$的第i,j列，而$\vec{e}_k^T$的第i,j列均为0，$k \neq i,j$，所以

$$\vec{e}_k^TP^{(ij)} = \vec{e}_k^T$$

注意到我们显然有

$$P^{(ij)}P^{(ij)} = I$$

所以

$$
\begin{aligned}
L_k P^{(ij)} &= P^{(ij)} + \vec{m}_k \left(\vec{e}_k^T P^{(ij)}\right) \\
&= P^{(ij)} + \vec{m}_k \vec{e}_k^T \\
&= P^{(ij)} + P^{(ij)} P^{(ij)} \overrightarrow{m_k} \vec{e}_k^T \\
&= P^{(ij)} (I + P^{(ij)} \overrightarrow{m_k} \vec{e}_k^T)
\end{aligned}
$$

(c)令

$$
\begin{aligned}
G(k) &= P_{k+1} L_{k+1} \ldots P_{n-1} L_{n-1} \\
G'(k) &= P_{k+1} \ldots P_{n-1} L_{k+1}^p \ldots L_{n-1}^p
\end{aligned}
$$

我们的目标是证明

$$
G(0) = G'(0)
$$

所以关于k做数学归纳法即可。当$k = n-2$时，

$$
\begin{aligned}
G(n-2) &= P_{n-1} L_{n-1} = P_{n-1}(I + \vec{m}_{n-1} \vec{e}_{n-1}^T) \\
G'(n-2) &= P_{n-1} L_{n-1}^p = P_{n-1}(I + \vec{m}_{n-1} \vec{e}_{n-1}^T)
\end{aligned}
$$

所以

$$
G(n-2) = G'(n-2)
$$

因此$k = n-2$时结论成立。假设$k = s$时结论，现在证明$k = s-1$时结论也成立，此时有

$$
\begin{aligned}
G(s) &= P_{s+1} L_{s+1} \ldots P_{n-1} L_{n-1} \\
&= G'(s) \\
&= P_{s+1} \ldots P_{n-1} L_{s+1}^p \ldots L_{n-1}^p
\end{aligned}
$$

由定义，我们有

$$
\begin{aligned}
G(s-1) &= P_s L_s G(s) \\
&= P_s L_s P_{s+1} \ldots P_{n-1} L_{s+1}^p \ldots L_{n-1}^p
\end{aligned}
$$

利用(b)计算$P_s L_s P_{s+1} \ldots P_{n-1}$可得

$$
\begin{aligned}
P_s L_s P_{s+1} \ldots P_{n-1} &= P_s (I + \vec{m}_s \vec{e}_s) P_{s+1} \ldots P_{n-1} \\
&= P_s P_{s+1} (I + P_{s+1} \vec{m}_s \vec{e}_s) P_{s+2} \ldots P_{n-1}
\end{aligned}
$$

注意到$P_{s+1}\vec{m}_s$的特点依然为$i \le s$的元素为0，所以仍然可以使用(b)的性质，即

$$
(I + P_{s+1} \vec{m}_s \vec{e}_s) P_{s+2} = P_{s+2} (I + P_{s+2} P_{s+1} \vec{m}_s \vec{e}_s)
$$

所以

$$
\begin{aligned}
P_s L_s P_{s+1} \dots P_{n-1} &= P_s P_{s+1} (I + P_{s+1} \vec{m}_s \vec{e}_s) P_{s+2} \dots P_{n-1} \\
&= P_s P_{s+1} P_{s+2} (I + P_{s+2} P_{s+1} \vec{m}_s \vec{e}_s) P_{s+3} \dots P_{n-1} \\
&\dots \\
&= P_s \dots P_{n-1} (I + P_{n-1} \dots P_{s+1} \vec{m}_s \vec{e}_s) \\
&= P_s \dots P_{n-1} L_s^p
\end{aligned}
$$

从而

$$
\begin{aligned}
G(s-1) &= P_s L_s P_{s+1} \dots P_{n-1} L_{s+1}^p \dots L_{n-1}^p \\
&= P_s \dots P_{n-1} L_s^p L_{s+1}^p \dots L_{n-1}^p \\
&= G'(s-1)
\end{aligned}
$$

因此$k = s-1$时结论也成立。

(d)首先考虑

$$S_k \triangleq \vec{m}_k \vec{e}_k^T$$

由$\vec{m}_k, \vec{e}_k$的性质可得，只有当$i \geq k+1$且$j = k$时，$(S_k)_{ij} \neq 0$，所以当$i < j$时，我们必然有

$$(S_k)_{ij} = 0$$

所以S_k是下三角阵，注意到作用$P_{n-1} \dots . P_{k+1}$只会交换位于$i \geq k+1$且$j = k$位置上的元素的相对位置，所以我们依然有$P_{n-1} \dots . P_{k+1} S_k$是下三角阵，所以

$$L_k^p = I + P_{n-1} \dots . P_{k+1} S_k$$

是下三角阵，因此

$$L_1^p \dots L_{n-1}^p$$

是下三角阵。

Problem 3

(a)令$e_1, \dots, e_n$为$\mathbb{R}^n$上标准正交积，那么

$$\vec{x} = \sum_{i=1}^{n} x_i e_i, \vec{y} = \sum_{j=1}^{n} y_j e_j$$

所以

$$
\begin{aligned}
\langle \vec{x}, \vec{y} \rangle &= \langle \sum_{i=1}^{n} x_i e_i, \sum_{j=1}^{n} y_j e_j \rangle \\
&= \sum_{i=1}^{n} \langle x_i e_i, \sum_{j=1}^{n} y_j e_j \rangle & \text{由性质}3 \\
&= \sum_{i=1}^{n} x_i \langle e_i, \sum_{j=1}^{n} y_j e_j \rangle & \text{由性质}2 \\
&= \sum_{i=1}^{n} x_i \langle \sum_{j=1}^{n} y_j e_j, e_i \rangle & \text{由性质}1 \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i y_j \langle e_j, e_i \rangle & \text{由性质}2 \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i y_j \langle e_i, e_j \rangle & \text{由性质}1
\end{aligned}
$$

记

$$A_{ij} = \langle e_i, e_j \rangle$$

那么

$$\langle \vec{x}, \vec{y} \rangle = \vec{x}^T A \vec{y}$$

注意到

$$\langle \vec{x}, \vec{x} \rangle = \vec{x}^T A \vec{x} \geq 0$$

当且仅当$\vec{x} = \vec{0}$时等号成立，所以A是正定矩阵，因此存在对称矩阵M，使得

$$A = M^T M = M^2$$

所以

$$
\begin{aligned}
\langle \vec{x}, \vec{y} \rangle &= \vec{x}^T A \vec{y} \\
&= \vec{x}^T M^T M \vec{y} \\
&= (M\vec{y})^T (M\vec{y})
\end{aligned}
$$

(b)利用(a)，不难得到如下等价定义：

$$d(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})^T A (\vec{x} - \vec{y})}$$

其中A是正定矩阵。

(c)记

$$X=\begin{bmatrix} —(\vec{x}^{(1)})^T— \\ —(\vec{x}^{(2)})^T— \\ \vdots \\ —(\vec{x}^{(n)})^T— \end{bmatrix} \in \mathbb{R}^{n\times d}, Y=\begin{bmatrix} —(\vec{y}^{(1)})^T— \\ —(\vec{y}^{(2)})^T— \\ \vdots \\ —(\vec{y}^{(n)})^T— \end{bmatrix} \in \mathbb{R}^{n\times d}$$

我们可以将问题描述如下，找到矩阵A，使得

$$\sum_{i=1}^{n} d^2(\vec{x}_i,\vec{y}_i)=\sum_{i=1}^{n}(\vec{x}_i-\vec{y}_i)^T A(\vec{x}_i-\vec{y}_i)$$

达到最小值。对上述函数化简可得

$$\begin{aligned}\sum_{i=1}^{n} d^2(\vec{x}_i,\vec{y}_i)&=\sum_{i=1}^{n}(\vec{x}_i-\vec{y}_i)^T A(\vec{x}_i-\vec{y}_i)\\&=\sum_{i=1}^{n}(\vec{x}_i-\vec{y}_i)^T M^T M(\vec{x}_i-\vec{y}_i)\\&=\sum_{i=1}^{n}(M(\vec{x}_i-\vec{y}_i))^T M(\vec{x}_i-\vec{y}_i)\\&=\|M(X-Y)^T\|_2^2\end{aligned}$$

由于我们要比较距离的接近程度，所以这里增加如下约束条件：

$$\|X-Y\|_2^2=\|(X-Y)^T\|_2^2=1$$

所以可以将原问题转化为熟悉的形式。

Problem 4

备注，题目中虽然没有讲，但我推断这里默认Tikhonov regularization为

$$\Gamma=I$$

否则有些内容无法讨论。

(a)加上Tikhonov regularization，我们的目标是最小化

$$\begin{aligned}J(\vec{a})&=\sum_{i=1}^{k}\left(||y^{(i)}-\vec{a}.\vec{x}^{(i)}||_2^2\right)+||\Gamma\vec{a}||_2^2\\&=\sum_{i=1}^{k}\left(||y^{(i)}-\vec{a}(\vec{x}^{(i)})^T||_2^2\right)+\vec{a}^T\Gamma^T\Gamma\vec{a}\end{aligned}$$

定义

$$X=\begin{bmatrix}-(\vec{x}^{(1)})^T-\\-(\vec{x}^{(2)})^T-\\\vdots\\-(\vec{x}^{(k)})^T-\end{bmatrix},\vec{y}=\begin{bmatrix}y^{(1)}\\y^{(2)}\\\vdots\\y^{(k)}\end{bmatrix}$$

那么上述问题可以转化为最小化

$$\begin{aligned}J(\vec{a})&=||X\vec{a}-\vec{y}||_2^2+\vec{a}^T\Gamma^T\Gamma\vec{a}\\&=(X\vec{a}-\vec{y})^T(X\vec{a}-\vec{y})+\vec{a}^T\Gamma^T\Gamma\vec{a}\\&=\vec{a}^TX^TX\vec{a}-2\vec{a}^TX^T\vec{y}+\vec{y}^T\vec{y}+\vec{a}^T\Gamma^T\Gamma\vec{a}\end{aligned}$$

关于$\vec{a}$求梯度可得

$$\nabla_{\vec{a}}J(\vec{a})=2X^TX\vec{a}-2X^T\vec{y}+2\Gamma^T\Gamma\vec{a}$$

令上式为0可得正规方程。

$$(X^TX+\Gamma^T\Gamma)\vec{a}=X^T\vec{y}$$

(b)将$\vec{a}$分解为

$$\vec{a}=\vec{a}_{\perp}+\vec{a}_{||}$$

其中

$$\begin{gathered}\vec{a}_{||}\in\text{span}\{\vec{x}^{(1)},\ldots,\vec{x}^{(k)}\}\\(\vec{x}^{(i)})^T.\vec{a}_{\perp}=0,\forall 1\le i\le k\end{gathered}$$

所以

$$\begin{gathered}X\vec{a}=X(\vec{a}_{\perp}+\vec{a}_{||})=X\vec{a}_{||}\\||X\vec{a}-\vec{y}||_2^2=||X\vec{a}_{||}-\vec{y}||_2^2\end{gathered}$$

不难发现我们有

$$\begin{aligned}||\vec{a}||^2&=\vec{a}^T\vec{a}\\&=\vec{a}_{\perp}^T\vec{a}_{\perp}+\vec{a}_{||}^T\vec{a}_{||}\\&=||\vec{a}_{\perp}||^2+||\vec{a}_{||}||^2\end{aligned}$$

利用$\Gamma=I$的事实，所以

$$\begin{aligned}\vec{a}^T\Gamma^T\Gamma\vec{a}&=\vec{a}^T\vec{a}\\&\ge||\vec{a}_{||}||^2\end{aligned}$$

从而

$$
\begin{aligned}
J(\vec{a}) &= ||X\vec{a} - \vec{y}||_2^2 + \vec{a}^T\Gamma^T\Gamma\vec{a} \\
&\geq ||X\vec{a}_{\parallel} - \vec{y}||_2^2 + ||\vec{a}_{\parallel}||^2
\end{aligned}
$$

当且仅当

$$\vec{a} = \vec{a}_{\perp}$$

等号成立。所以我们只需要考虑

$$\vec{a} \in \mathrm{span}\{\vec{x}^{(1)}, \ldots, \vec{x}^{(k)}\}$$

的情形。

(c)记

$$\vec{c} = (c_1, \ldots, c_k)^T$$

那么

$$\vec{a} = X^T\vec{c}$$

带入正规方程可得

$$
\begin{aligned}
(X^TX + \Gamma^T\Gamma)X^T\vec{c} &= X^T\vec{y} \\
(X^TX + I)X^T\vec{c} &= X^T\vec{y} \\
X^T(XX^T + I)\vec{c} &= X^T\vec{y} \\
XX^T(XX^T + I)\vec{c} &= XX^T\vec{y} \\
(XX^T + I)\vec{c} &= \vec{y} \\
\vec{c} &= (XX^T + I)^{-1}\vec{y}
\end{aligned}
$$

注意这里我们有

$$XX^T + I \in \mathbb{R}^{k\times k}$$

(d)注意到

$$(XX^T)_{i,j} = (\vec{x}^{(i)})^T\vec{x}^{(j)}$$

如果使用特征转换，不妨设

$$
Z = \begin{bmatrix}
—(\vec{z}^{(1)})^T— \\
—(\vec{z}^{(2)})^T— \\
\vdots \\
—(\vec{z}^{(k)})^T—
\end{bmatrix}
$$

那么

$$
\begin{gathered}
(ZZ^T)_{i,j} = (\phi(\vec{x}^{(i)}))^T\phi(\vec{x}^{(j)}) = K(\vec{x}^{(i)}, \vec{x}^{(j)}) \\
ZZ^T = K
\end{gathered}
$$

所以

$$\vec{c} = (K+I)^{-1}\vec{y}, \vec{a} = Z^T\vec{a} = Z^T(K+I)^{-1}\vec{y}$$

$$\vec{a}.\,\phi(\vec{x}) = \vec{a}^T\phi(\vec{x}) = \vec{y}^T(K+I)^{-1}Z\phi(\vec{x}) = \vec{y}^T(K+I)^{-1}\begin{bmatrix} K(x_1,x) \\ K(x_2,x) \\ \vdots \\ K(x_k,x) \end{bmatrix}$$

补充题，直接利用泰勒展开即可：

$$\begin{aligned} e^{-\pi(x-y)^2} &= e^{-\pi x^2}e^{-\pi y^2}e^{2\pi xy} \\ &= e^{-\pi x^2}e^{-\pi y^2}(\sum_{n=0}^{\infty}\frac{(2\pi xy)^n}{n!}) \\ &= e^{-\pi x^2}e^{-\pi y^2}(\sum_{n=0}^{\infty}\sqrt{\frac{(2\pi)^n}{n!}}x^n.\sqrt{\frac{(2\pi)^n}{n!}}y^n) \end{aligned}$$

所以

$$\phi(x) = e^{-\pi x^2}(1,\ldots,\sqrt{\frac{(2\pi)^n}{n!}}x^n,\ldots)$$