线性代数(2023-2024)第十三次作业参考答案

1 复习知识点

- 牢记Gram–Schmidt算法，熟练掌握如何用Gram–Schmidt算法将一个内积空间里的一组基底转化为一组标准正交基底。作为重要应用，熟练掌握使用Gram–Schmidt算法得到对称矩阵的各个特征空间的标准正交基底，从而将对称矩阵正交对角化。
- 二次型的定义，orthogonal change of variables的定义，Theorem 7.17 (The principal axes theorem)的内容，以及熟练掌握如何用orthogonal change of variables化简一个给定的二次型，参考讲义196页-197页，198页-199页的两个例子。
- 二次型的positive definite, negative definite, indefinite与其对应的对称矩阵特征值的关系，如何判断一个对称矩阵是否是positive defintie。
- 熟练掌握QR分解与奇异值分解(SVD)的具体计算过程。

2 习题部分

Problem A(6 Points)

令V为一个内积空间，$\langle\cdot,\cdot\rangle$为$V$上的内积。假设$S=\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n\}$是$V$的一组基底。对$S$使用Gram–Schmidt算法，依次得到

$$\boldsymbol{v}_1=\boldsymbol{u}_1/\|\boldsymbol{u}_1\|;$$
$$\boldsymbol{v}_2=\frac{\boldsymbol{u}_2-\langle\boldsymbol{u}_2,\boldsymbol{v}_1\rangle\boldsymbol{v}_1}{\|\boldsymbol{u}_2-\langle\boldsymbol{u}_2,\boldsymbol{v}_1\rangle\|};$$
$$\boldsymbol{v}_j=\frac{\boldsymbol{u}_j-\langle\boldsymbol{u}_j,\boldsymbol{v}_1\rangle\boldsymbol{v}_1-\ldots-\langle\boldsymbol{u}_j,\boldsymbol{v}_{j-1}\rangle\boldsymbol{v}_{j-1}}{\|\boldsymbol{u}_j-\langle\boldsymbol{u}_j,\boldsymbol{v}_1\rangle\boldsymbol{v}_1-\ldots-\langle\boldsymbol{u}_j,\boldsymbol{v}_{j-1}\rangle\boldsymbol{v}_{j-1}\|},j=1,\ldots,n.$$

用尽量严格的数学语言证明，对任何$j=1,\ldots,n$，都有

$$\text{span}\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_j\}=\text{span}\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_j\}.$$

答案：我们对$j=1,\ldots,n$使用数学归纳法。

对于$j=1$，由于$\boldsymbol{v}_1=\boldsymbol{u}_1/\|\boldsymbol{u}_1\|;$，显然有span$\{\boldsymbol{v}_1\}=$ span$\{\boldsymbol{u}_1\}$。

现在假设对某个$1\le j<n$，我们有span$\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_j\}=$ span$\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_j\}$。我们考虑$j+1$的情况。首先注意，令$\boldsymbol{v}'_{j+1}=\boldsymbol{u}_{j+1}-(\langle\boldsymbol{u}_{j+1},\boldsymbol{v}_1\rangle\boldsymbol{v}_1+\ldots+\langle\boldsymbol{u}_{j+1},\boldsymbol{v}_j\rangle\boldsymbol{v}_j)$，那么$\boldsymbol{v}_{j+1}=\frac{\boldsymbol{v}'_{j+1}}{\|\boldsymbol{v}'_{j+1}\|}$。因此不难看出span$\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_j,\boldsymbol{v}_{j+1}\}=$ span$\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_j,\boldsymbol{v}'_{j+1}\}$。因此我们只需证明

$$\text{span}\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_j,\boldsymbol{u}_{j+1}\}=\text{span}\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_j,\boldsymbol{v}'_{j+1}\}.$$

令$W=\text{span}\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_j\}$。由归纳假设，我们有$W=\text{span}\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_j\}=\text{span}\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_j\}$。那么$\boldsymbol{v}'_{j+1}=\boldsymbol{u}_{j+1}-(\langle\boldsymbol{u}_{j+1},\boldsymbol{v}_1\rangle\boldsymbol{v}_1+\ldots+\langle\boldsymbol{u}_{j+1},\boldsymbol{v}_j\rangle\boldsymbol{v}_j)=\boldsymbol{u}_{j+1}-\text{proj}_W(\boldsymbol{u}_{j+1})$，特别地，我们有

$$\text{proj}_W(\boldsymbol{u}_{j+1})\in W=\text{span}\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_j\},$$

也即存在实数$d_1,\ldots,d_j$使得$\text{proj}_W(\boldsymbol{u}_{j+1})=d_1\boldsymbol{u}_1+\ldots+d_j\boldsymbol{u}_j$。因此，任何关于$\boldsymbol{v}_1,\ldots,\boldsymbol{v}_j,\boldsymbol{v}'_{j+1}$的线性组合

$$c_1\boldsymbol{v}_1+\ldots+c_j\boldsymbol{v}_j+c_{j+1}\boldsymbol{v}'_{j+1}$$

都可以表示为

$$c_1\boldsymbol{v}_1+\ldots+c_j\boldsymbol{v}_j+c_{j+1}(\boldsymbol{u}_{j+1}-(d_1\boldsymbol{u}_1+\ldots+d_j\boldsymbol{u}_j))=c_1\boldsymbol{v}_1+\ldots+c_j\boldsymbol{v}_j-c_{j+1}(d_1\boldsymbol{u}_1+\ldots+d_j\boldsymbol{u}_j)+c_{j+1}\boldsymbol{u}_{j+1}.$$

再次利用归纳假设，$c_1\boldsymbol{v}_1+\ldots+c_j\boldsymbol{v}_j\in W=\text{span}\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_j\}=\text{span}\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_j\}$，因此存在另一组系数$k_1,\ldots,k_j$使得

$$c_1\boldsymbol{v}_1+\ldots+c_j\boldsymbol{v}_j=k_1\boldsymbol{u}_1+\ldots+k_j\boldsymbol{u}_j,$$

因此等式

$$c_1\boldsymbol{v}_1+\ldots+c_j\boldsymbol{v}_j+c_{j+1}\boldsymbol{v}'_{j+1}=c_1\boldsymbol{v}_1+\ldots+c_j\boldsymbol{v}_j-c_{j+1}(d_1\boldsymbol{u}_1+\ldots+d_j\boldsymbol{u}_j)+c_{j+1}\boldsymbol{u}_{j+1}.$$

又可以表示为

$$c_1\boldsymbol{v}_1+\ldots+c_j\boldsymbol{v}_j+c_{j+1}\boldsymbol{v}'_{j+1}=k_1\boldsymbol{u}_1+\ldots+k_j\boldsymbol{u}_j-c_{j+1}(d_1\boldsymbol{u}_1+\ldots+d_j\boldsymbol{u}_j)+c_{j+1}\boldsymbol{u}_{j+1}.$$

显然，这意味着$c_1\boldsymbol{v}_1+\ldots+c_j\boldsymbol{v}_j+c_{j+1}\boldsymbol{v}'_{j+1}\in\text{span}\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_{j+1}\}$，也即我们得到$\text{span}\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_j,\boldsymbol{v}'_{j+1}\}\subset\text{span}\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_{j+1}\}$。
同理可证，对于任何关于$\boldsymbol{u}_1,\ldots,\boldsymbol{u}_{j+1}$的线性组合

$$c_1\boldsymbol{u}_1+\ldots+c_j\boldsymbol{u}_j+c_{j+1}\boldsymbol{u}_{j+1},$$

由于$\boldsymbol{u}_{j+1}=\boldsymbol{v}'_{j+1}+\text{proj}_W(\boldsymbol{u}_{j+1})$，且$W=\text{span}\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_j\}=\text{span}\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_j\}$故而有$\text{proj}_W(\boldsymbol{u}_{j+1})\in\text{span}\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_j\}$，以及$c_1\boldsymbol{u}_1+\ldots+c_j\boldsymbol{u}_j\in W=\text{span}\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_j\}$，我们有

$$c_1\boldsymbol{u}_1+\ldots+c_j\boldsymbol{u}_j+c_{j+1}\boldsymbol{u}_{j+1}=\underbrace{(c_1\boldsymbol{u}_1+\ldots+c_j\boldsymbol{u}_j+c_{j+1}\text{proj}_W(\boldsymbol{u}_{j+1}))}_{\in\text{span}\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_j\}}+c_{j+1}\boldsymbol{v}'_{j+1}$$

即$c_1\boldsymbol{u}_1+\ldots+c_j\boldsymbol{u}_j+c_{j+1}\boldsymbol{u}_{j+1}\in\text{span}\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_j,\boldsymbol{v}'_{j+1}\}$，因此

$$\text{span}\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}'_{j+1}\}\subset\text{span}\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_j,\boldsymbol{v}'_{j+1}\}$$

综上，得到

$$\text{span}\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_j,\boldsymbol{u}_{j+1}\}=\text{span}\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_j,\boldsymbol{v}'_{j+1}\},$$

因此该结论对任何$j = 1, \ldots, n$成立。

Problem B(6 Points), 2022-2023年线性代数期末考试题

Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ be m vectors in $\mathbb{R}^n$. Denote by "$\cdot$" the Euclidean inner product on $\mathbb{R}^n$. Prove that $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ are linearly dependent if and only if the determinant

$$\begin{vmatrix} \boldsymbol{u}_1 \cdot \boldsymbol{u}_1 & \boldsymbol{u}_1 \cdot \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_1 \cdot \boldsymbol{u}_m \\ \boldsymbol{u}_2 \cdot \boldsymbol{u}_1 & \boldsymbol{u}_2 \cdot \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_2 \cdot \boldsymbol{u}_m \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{u}_m \cdot \boldsymbol{u}_1 & \boldsymbol{u}_m \cdot \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_m \cdot \boldsymbol{u}_m \end{vmatrix} = 0.$$

答案： 令$A = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_m \end{bmatrix} \in M_{n\times m}$，即$A$的列向量依次为$\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$，那么

$$A^\top A = \begin{bmatrix} \boldsymbol{u}_1 \cdot \boldsymbol{u}_1 & \boldsymbol{u}_1 \cdot \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_1 \cdot \boldsymbol{u}_m \\ \boldsymbol{u}_2 \cdot \boldsymbol{u}_1 & \boldsymbol{u}_2 \cdot \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_2 \cdot \boldsymbol{u}_m \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{u}_m \cdot \boldsymbol{u}_1 & \boldsymbol{u}_m \cdot \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_m \cdot \boldsymbol{u}_m \end{bmatrix} \in M_{m\times m}.$$

由**讲义Theorem 4.38**可知，$\text{rank}(A^\top A) = \text{rank}(A)$，因此我们有

$$\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\} \text{ 线性相关} \iff \text{rank}(A) < m \iff \text{rank}(A^\top A) < m \iff \det(A^\top A) = 0.$$

Problem C(6 Points)

已知$\boldsymbol{v} = (1, 1, 1)$是二次型$Q_A(\boldsymbol{x}) = 2x_1^2 + x_2^2 + ax_3^2 + 2x_1x_2 + 2bx_1x_3 + 2x_2x_3$对应的对称矩阵$A$的一个特征向量，其中$a, b$为待定实数。

1. (2 points) 求a与b的值，并求出对阵矩阵A。

   答案： 设对称矩阵$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$。注意$a_{12} = a_{21}, a_{13} = a_{31}, a_{23} = a_{32}$。

   此时有

   $$Q_A(\boldsymbol{x}) = a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + 2a_{12}x_1x_2 + 2a_{13}x_1x_3 + 2a_{23}x_2x_3.$$

   比较$Q_A(\boldsymbol{x}) = 2x_1^2 + x_2^2 + ax_3^2 + 2x_1x_2 + 2bx_1x_3 + 2x_2x_3$的系数，可得

   $$a_{11} = 2, a_{22} = 1, a_{33} = a, a_{12} = a_{21} = 1, a_{13} = a_{31} = b, a_{23} = a_{32} = 1,$$

即$A=\begin{bmatrix}2&1&b\\1&1&1\\b&1&a\end{bmatrix}$。已知$\boldsymbol{v}=(1,1,1)$是$A$的一个特征向量，因此存在$\lambda\in\mathbb{R}$使得

$$A\boldsymbol{v}=\lambda\boldsymbol{v}\Rightarrow\begin{bmatrix}3+b\\3\\a+b+1\end{bmatrix}=\begin{bmatrix}\lambda\\\lambda\\\lambda\end{bmatrix},$$

由此可得$\lambda=3,a=2,b=0$。因此$A=\begin{bmatrix}2&1&0\\1&1&1\\0&1&2\end{bmatrix}$。

2. (3 points) 利用orthogonal change of variables $\boldsymbol{y}=P^\top\boldsymbol{x}$ 求出二次型关于$\boldsymbol{y}=(y_1,y_2,y_3)$的表达式。

   答案： 通过解A的特征方程

   $$p_A(\lambda)=\det(\lambda I_3-A)=0$$

   可解得A的三个特征值为$\lambda_1=0,\lambda_2=2,\lambda_3=3$，因此$A$的正交对角化$D=P^\top AP$里出现的对角矩阵为

   $$D=\begin{bmatrix}0&&\\&2&\\&&3\end{bmatrix}.$$

   因此二次型关于$\boldsymbol{y}=P^\top\boldsymbol{x}$的表达式为$0y_1^2+2y_2^2+3y_3^2=2y_2^2+3y_3^2$。

3. (1 point) 判断对称矩阵A是否是positive definite。

   答案： 由**讲义Theorem 7.19**可知，A是positvie definite当且仅当它的所有特征值都大于零。我们已经知道A有一个为0的特征值，故A并非positive definite。实际上由于A的所有特征值都是非负的，所以A是positive semidefinite.

Problem D(6 Points), 2022-2023年线性代数期末延期考试题

Let $A=\begin{bmatrix}1&1&1\\1&0&1\\-1&-1&1\end{bmatrix}$. Find a decomposition $A=QRP$ such that: P and Q are orthogonal matrices, and

$$R=\begin{bmatrix}a&0&d\\0&b&e\\0&0&c\end{bmatrix}$$

is an upper triangular matrix with $a, b, c, d, e \in \mathbb{R}$.

答案： 该题的解答思路如下：由于rank$(A) = 3$以及$A = QRP$的分解中要求Q为正交矩阵，R为上三角矩阵，我们首先想到的就是对A本身进行QR分解。那么，令$A = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \boldsymbol{u}_3 \end{bmatrix}$，即$\boldsymbol{u}_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$，$\boldsymbol{u}_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$，$\boldsymbol{u}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$，我们利用Gram–Schmidt算法将其转化为一组标准正交基底$\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$，即

$$\boldsymbol{v}_1 = \frac{1}{\|\boldsymbol{u}_1\|}\boldsymbol{u}_1, \boldsymbol{v}_2 = \frac{\boldsymbol{u}_2 - (\boldsymbol{u}_2 \cdot \boldsymbol{v}_1)\boldsymbol{v}_1}{\|\boldsymbol{u}_2 - (\boldsymbol{u}_2 \cdot \boldsymbol{v}_1)\boldsymbol{v}_1\|}, \boldsymbol{v}_3 = \frac{\boldsymbol{u}_3 - (\boldsymbol{u}_3 \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 - (\boldsymbol{u}_3 \cdot \boldsymbol{v}_2)\boldsymbol{v}_2}{\|\boldsymbol{u}_3 - (\boldsymbol{u}_3 \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 - (\boldsymbol{u}_3 \cdot \boldsymbol{v}_2)\boldsymbol{v}_2\|}.$$

假设$A = QR$，那么$Q = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 \end{bmatrix}$，$R = \begin{bmatrix} \boldsymbol{u}_1 \cdot \boldsymbol{v}_1 & \boldsymbol{u}_2 \cdot \boldsymbol{v}_1 & \boldsymbol{u}_3 \cdot \boldsymbol{v}_1 \\ \boldsymbol{u}_1 \cdot \boldsymbol{v}_2 & \boldsymbol{u}_2 \cdot \boldsymbol{v}_2 & \boldsymbol{u}_3 \cdot \boldsymbol{v}_2 \\ \boldsymbol{u}_1 \cdot \boldsymbol{v}_3 & \boldsymbol{u}_2 \cdot \boldsymbol{v}_3 & \boldsymbol{u}_3 \cdot \boldsymbol{v}_3 \end{bmatrix}$。然而，此时我们注意到，$R$第一行第二列上的项为$\boldsymbol{u}_2 \cdot \boldsymbol{v}_1 = \frac{1}{\|\boldsymbol{u}_1\|}\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \neq 0$，与题目所要求的$R = \begin{bmatrix} a & 0 & d \\ 0 & b & e \\ 0 & 0 & c \end{bmatrix}$的形式不符合。因此我们不能直接对$A$进行QR分解。

经过观察可知，$\boldsymbol{u}_2 \cdot \boldsymbol{u}_3 = 0$，因此，以上QR分解过程暗示我们，我们需要将$A$的第一列与第三列互换得到矩阵

$$\tilde{A} = \begin{bmatrix} \boldsymbol{u}_3 & \boldsymbol{u}_2 & \boldsymbol{u}_1 \end{bmatrix},$$

再对$\tilde{A}$进行QR分解，则可以得到如题目要求的上三角矩阵R。所以我们现在先考虑如何从A得到$\tilde{A}$。显然，要交换A的第一列与第三列，我们可以先交换$A^\top$的第一行与第三行，再对交换过后的矩阵进行一次转置即可；而将$A^\top$的第一行与第三行进行交换，只需对$A^\top$左乘一个初等矩阵

$$P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix},$$

因此

$$\tilde{A} = \begin{bmatrix} \boldsymbol{u}_3 & \boldsymbol{u}_2 & \boldsymbol{u}_1 \end{bmatrix} = (PA^\top)^\top = (A^\top)^\top P^\top = AP^\top.$$

显然，$P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$是一个**正交矩阵**。

现在，我们对$\tilde{A}=\begin{bmatrix}\boldsymbol{w}_1 & \boldsymbol{w}_2 & \boldsymbol{w}_3\end{bmatrix}=\begin{bmatrix}1 & 1 & 1\\ 1 & 0 & 1\\ 1 & -1 & -1\end{bmatrix}$进行QR分解，可求得

$$\boldsymbol{q}_1=\boldsymbol{w}_1/\|\boldsymbol{w}_1\|=\begin{bmatrix}1/\sqrt{3}\\ 1/\sqrt{3}\\ 1/\sqrt{3}\end{bmatrix},\boldsymbol{q}_2=\frac{\boldsymbol{w}_2-(\boldsymbol{w}_2\cdot\boldsymbol{q}_1)\boldsymbol{q}_1}{\|\boldsymbol{w}_2-(\boldsymbol{w}_2\cdot\boldsymbol{q}_1)\boldsymbol{q}_1\|}=\begin{bmatrix}1/\sqrt{2}\\ 0\\ -1/\sqrt{2}\end{bmatrix}$$

(注意，实际上我们有$\boldsymbol{w}_2\cdot\boldsymbol{q}_1=0$),

$$\boldsymbol{q}_3=\frac{\boldsymbol{w}_3-(\boldsymbol{w}_3\cdot\boldsymbol{q}_1)\boldsymbol{q}_1-(\boldsymbol{w}_3\cdot\boldsymbol{q}_2)\boldsymbol{q}_2}{\|\boldsymbol{w}_3-(\boldsymbol{w}_3\cdot\boldsymbol{q}_1)\boldsymbol{q}_1-(\boldsymbol{w}_3\cdot\boldsymbol{q}_2)\boldsymbol{q}_2\|}=\begin{bmatrix}-1/\sqrt{6}\\ 2/\sqrt{6}\\ -1/\sqrt{6}\end{bmatrix},$$

因此$Q=\begin{bmatrix}1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6}\\ 1/\sqrt{3} & 0 & 2/\sqrt{6}\\ 1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6}\end{bmatrix}$，同时

$$R=\begin{bmatrix}\boldsymbol{w}_1\cdot\boldsymbol{q}_1 & \boldsymbol{w}_2\cdot\boldsymbol{q}_1 & \boldsymbol{w}_3\cdot\boldsymbol{q}_1\\ \boldsymbol{w}_1\cdot\boldsymbol{q}_2 & \boldsymbol{w}_2\cdot\boldsymbol{q}_2 & \boldsymbol{w}_3\cdot\boldsymbol{q}_2\\ \boldsymbol{w}_1\cdot\boldsymbol{q}_3 & \boldsymbol{w}_2\cdot\boldsymbol{q}_3 & \boldsymbol{w}_3\cdot\boldsymbol{q}_3\end{bmatrix}=\begin{bmatrix}\sqrt{3} & 0 & 1/\sqrt{3}\\ 0 & \sqrt{2} & \sqrt{2}\\ 0 & 0 & 2/\sqrt{6}\end{bmatrix}.$$

由以上，我们得到

$$AP^\top=\tilde{A}=QR,$$

两边同时右乘P，利用$P^\top P=I$，可得

$$A=QRP,$$

即为所求分解。

Problem E(6 Points), 最小二乘法

通过这道题目，我们简单了解一个线性代数的具体应用：最小二乘法(Least Squares)与最佳逼近(Best Approximation)。

这个应用的背景如下：假设$A\in M_{m\times n}$与$\boldsymbol{b}\in\mathbb{R}^m$给定，但方程组$A\boldsymbol{x}=\boldsymbol{b}$无解。我们试图找到一个向量$\boldsymbol{x}\in\mathbb{R}^n$使得$A\boldsymbol{x}$与$\boldsymbol{b}$的距离尽可能的小，即：

$$\|\boldsymbol{b}-A\boldsymbol{y}\|\geq\|\boldsymbol{b}-A\boldsymbol{x}\|>0$$

对任何$\boldsymbol{y}\in\mathbb{R}^n$都成立。满足这个条件的向量$\boldsymbol{x}\in\mathbb{R}^n$被称为最小二乘法(Least Squares)的解，$\boldsymbol{b}-A\boldsymbol{x}$被称为最小二乘法误差向量(least squares error vector)，$\|\boldsymbol{b}-A\boldsymbol{x}\|$被称为最小二乘法误差(least squares error)。

现在我们讨论如何求出这个最小二乘法的解$\boldsymbol{x}$。这里我们需要用到下面的这个结论。

- (1 point) 令V为一个内积空间，$\langle\cdot,\cdot\rangle$为$V$上的一个内积，$W$是$V$的一个有限维子空间，$\boldsymbol{b}\in V$。证明对于任何$\boldsymbol{w}\in W$，都有

$$\|\boldsymbol{b}-\mathrm{proj}_W(\boldsymbol{b})\|\leq\|\boldsymbol{b}-\boldsymbol{w}\|,$$

且等号只有当$\boldsymbol{w}=\mathrm{proj}_W(\boldsymbol{b})$时才能取到。

提示： 对于任何$\boldsymbol{w}\in W$，我们有

$$\boldsymbol{b}-\boldsymbol{w}=(\boldsymbol{b}-\mathrm{proj}_W(\boldsymbol{b}))+(\mathrm{proj}_W(\boldsymbol{b})-\boldsymbol{w}).$$

由于$\boldsymbol{b}$可以表示为$\boldsymbol{b}=\mathrm{proj}_W(\boldsymbol{b})+\mathrm{proj}_{W^\perp}(\boldsymbol{b})$(讲义Theorem 7.15)，我们又有$\boldsymbol{b}-\mathrm{proj}_W(\boldsymbol{b})=\mathrm{proj}_{W^\perp}(\boldsymbol{b})$，因此$\boldsymbol{b}-\mathrm{proj}_W(\boldsymbol{b})$实际上与$\mathrm{proj}_W(\boldsymbol{b})-\boldsymbol{w}\in W$正交。因此利用勾股定理可得

$$\|\boldsymbol{b}-\boldsymbol{w}\|^2=\|\boldsymbol{b}-\mathrm{proj}_W(\boldsymbol{b})\|^2+\|\mathrm{proj}_W(\boldsymbol{b})-\boldsymbol{w}\|^2.$$

利用这一等式可完成证明。

答案： 由提示可得，

$$\|\boldsymbol{b}-\boldsymbol{w}\|^2=\|\boldsymbol{b}-\mathrm{proj}_W(\boldsymbol{b})\|^2+\|\mathrm{proj}_W(\boldsymbol{b})-\boldsymbol{w}\|^2.$$

显然，由于$\|\mathrm{proj}_W(\boldsymbol{b})-\boldsymbol{w}\|^2\geq 0$，以上等式告诉我们

$$\|\boldsymbol{b}-\boldsymbol{w}\|^2=\|\boldsymbol{b}-\mathrm{proj}_W(\boldsymbol{b})\|^2+\|\mathrm{proj}_W(\boldsymbol{b})-\boldsymbol{w}\|^2\geq\|\boldsymbol{b}-\mathrm{proj}_W(\boldsymbol{b})\|^2$$

恒成立。如果$\boldsymbol{w}\neq\mathrm{proj}_W(\boldsymbol{b})$，那么必然有$\|\mathrm{proj}_W(\boldsymbol{b})-\boldsymbol{w}\|^2>0$，此时有$\|\boldsymbol{b}-\boldsymbol{w}\|^2>\|\boldsymbol{b}-\mathrm{proj}_W(\boldsymbol{b})\|^2$。因此可得，等号只有当$\boldsymbol{w}=\mathrm{proj}_W(\boldsymbol{b})$时才能取到。

现在令$V=\mathbb{R}^m$为欧式内积空间。那么以上结果告诉我们

- (1 point) 证明：$\boldsymbol{x}\in\mathbb{R}^n$是最小二乘法的解当且仅当$A\boldsymbol{x}=\mathrm{proj}_W(\boldsymbol{b})$，这里$W$是$A$的列空间。

答案： $\boldsymbol{x}\in\mathbb{R}^n$是最小二乘法的解当且仅当$\|\boldsymbol{b}-A\boldsymbol{y}\|\geq\|\boldsymbol{b}-A\boldsymbol{x}\|$对任何$\boldsymbol{y}\in\mathbb{R}^n$都成立。显然，由于

$$\{A\boldsymbol{y}:\boldsymbol{y}\in\mathbb{R}^n\}=\mathrm{RAN}(A)=\mathrm{Col}(A)=W$$

是A的列空间，见**讲义Theorem 4.51**，$\|\boldsymbol{b}-A\boldsymbol{y}\|\geq\|\boldsymbol{b}-A\boldsymbol{x}\|$对任何$\boldsymbol{y}\in\mathbb{R}^n$都成立等价于要求

$$\|\boldsymbol{b}-A\boldsymbol{x}\|\leq\|\boldsymbol{b}-\boldsymbol{w}\|$$

对任何$\boldsymbol{w} \in W = \mathrm{Col}(A)$成立。因此，由上一问的结论可知，这等价于要求$A\boldsymbol{x} = \mathrm{proj}_W(\boldsymbol{b})$。因此证毕。

现在，为了求解$A\boldsymbol{x} = \mathrm{proj}_W(\boldsymbol{b})$($W$是$A$的列空间)，我们首先将其写为

$$\boldsymbol{b} - A\boldsymbol{x} = \boldsymbol{b} - \mathrm{proj}_W(\boldsymbol{b}).$$

等式两边同乘以$A^\top$，可得

$$A^\top(\boldsymbol{b} - A\boldsymbol{x}) = A^\top(\boldsymbol{b} - \mathrm{proj}_W(\boldsymbol{b})) = A^\top(\mathrm{proj}_{W^\perp}(\boldsymbol{b})).$$

- (1 point) 证明： $A^\top(\mathrm{proj}_{W^\perp}(\boldsymbol{b})) = \mathbf{0}$。

答案： 由**英文教材Theorem 4.8.7**可知，$\mathrm{Null}(A^\top) = W^\perp = \mathrm{Col}(A)^\perp$，因此利用$\mathrm{proj}_{W^\perp}(\boldsymbol{b}) \in W^\perp$可得$\mathrm{proj}_{W^\perp}(\boldsymbol{b}) \in \mathrm{Null}(A^\top)$，从而有$A^\top(\mathrm{proj}_{W^\perp}(\boldsymbol{b})) = \mathbf{0}$。

因此，由以上结果，我们实际上得到，$\boldsymbol{x} \in \mathbb{R}^n$是最小二乘法的解当且仅当$A^\top(\boldsymbol{b} - A\boldsymbol{x}) = A^\top(\mathrm{proj}_{W^\perp}(\boldsymbol{b})) = \mathbf{0}$，也即当且仅当$\boldsymbol{x} \in \mathbb{R}^n$满足方程

$$(A^\top A)\boldsymbol{x} = A^\top \boldsymbol{b}.$$

以上方程也被称为normal system。

- (3 points) 令$A = \begin{bmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{bmatrix}$，$\boldsymbol{b} = \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix}$。证明$A\boldsymbol{x} = \boldsymbol{b}$无解，并利用normal system$(A^\top A)\boldsymbol{x} = A^\top \boldsymbol{b}$.求出最小二乘法的解$\boldsymbol{x}$，接着求出最小二乘法的误差$\|\boldsymbol{b} - A\boldsymbol{x}\|$。

答案： 通过计算$A\boldsymbol{x} = \boldsymbol{b}$的增广矩阵的阶梯型，容易看出该方程组无解。

要求最小二乘法的解$\boldsymbol{x}$，需要求解normal system

$$(A^\top A)\boldsymbol{x} = A^\top \boldsymbol{b}.$$

经计算可得

$$A^\top A = \begin{bmatrix} 14 & -3 \\ -3 & 21 \end{bmatrix}, \quad A^\top \boldsymbol{b} = \begin{bmatrix} 1 \\ 10 \end{bmatrix},$$

由此可得$(A^\top A)\boldsymbol{x} = A^\top \boldsymbol{b}$的解为

$$\boldsymbol{x} = \begin{bmatrix} \frac{17}{95} \\ \frac{143}{285} \end{bmatrix}.$$

此时有$\boldsymbol{b}-A\boldsymbol{x}=\begin{bmatrix}\frac{1232}{285}\\ -\frac{154}{285}\\ \frac{4}{3}\end{bmatrix}$，从而得到

$$\|\boldsymbol{b}-A\boldsymbol{x}\|\approx 4.556.$$

Bonus: 不计入分数，但强烈建议亲手计算

- 计算$A=\begin{bmatrix}1&0&0\\1&1&0\\1&1&1\end{bmatrix}$的QR分解。

  答案： 见讲义203–204页。

- 计算$A=\begin{bmatrix}1&1\\0&1\\1&0\end{bmatrix}$的奇异值分解(SVD)。

  答案： 见讲义207–208页。

Deadline: 22:00, January 14.

作业提交截止时间：1月14日晚上22：00。