SI231 - Matrix Computations, Fall 2020-21
Homework Set #3

Prof. Yue Qiu and Prof. Ziping Zhao

Name: Tian Yun **Major:** Master in CS

Student No.: 2019232102 **E-mail:** tianyun@shanghaitech.edu.cn

---

Acknowledgements:

1) Deadline: **2020-11-01 23:59:00**

2) Submit your homework at **Gradescope**. Entry Code: **MY3XBJ**. Homework #3 contains two parts, the theoretical part the and the programming part.

3) About the the theoretical part:

   (a) Submit your homework in **Homework 3** in gradescope. Make sure that you have correctly select pages for each problem. If not, you probably will get 0 point.

   (b) Your homework should be uploaded in the **PDF** format, and the naming format of the file is not specified.

   (c) You need to use LATEX in principle.

   (d) Use the given template and give your solution in English. Solution in Chinese is not allowed.

4) About the programming part:

   (a) Submit your codes in **Homework 3 Programming part** in gradescope.

   (b) Detailed requirements see in Problem 2 and Probelm 3.

5) **No late submission is allowed.**

I. UNDERSTANDING PROJECTION

Problem 1. (5 points $\times$ 3)

Suppose that $\mathbf{P} \in \mathbb{R}^{n \times n}$ is a projector onto a subspace $\mathcal{U}$ along its orthogonal complement $\mathcal{U}^\perp$, then it is called the **orthogonal projector** onto $\mathcal{U}$.

1) Prove that an orthogonal projector must be singular if it is not an identity matrix.
2) What is the orthogonal projector onto $\mathcal{U}^\perp$ along the subspace $\mathcal{U}$?
3) Let $\mathcal{U}$ and $\mathcal{W}$ be two subspaces of a vector space $\mathcal{V}$, and denote $\mathbf{P}_\mathcal{U}$ and $\mathbf{P}_\mathcal{W}$ as the corresponding orthogonal projectors, respectively. Prove that $\mathbf{P}_\mathcal{U}\mathbf{P}_\mathcal{W} = 0$ if and only if $\mathcal{U} \perp \mathcal{W}$.

Solution.

1) **proof**: A projection on a vector space V is a linear operator $P : V \mapsto V$ such that $P^2 = P$. $P \in R^{n \times m}$

$\therefore P(I - P) = 0$, and $rank(P(I - P))) = 0$.

$\therefore rank(P) + rank(I - p) - n \le 0$

$\therefore rank(P) + rank(I - p) \le n$

First case : $rank(p) = n$, $\therefore rank(I - p) = 0, \therefore I = p$

Second case: $rank(p) \le n$, $\therefore$ p must be singular.

2) $\mathbf{I} - \mathbf{P}$.

$\mathbf{P} \in \mathbb{R}^{n \times n}$ is a projector onto a subspace $\mathcal{U}$ along its orthogonal complement $\mathcal{U}^\perp$

so for all $x \in \mathcal{U}$, $\mathbf{P}x = x$, for all $x \in \mathcal{U}^\perp$, $\mathbf{P}x = 0$

Now we want to find the orthogonal projector $\mathbf{P}^*$ onto $\mathcal{U}^\perp$ along the subspace $\mathcal{U}$

it means for all $y \in \mathcal{U}^\perp$, $\mathbf{P}^*y = y$, for all $y \in \mathcal{U}$, $\mathbf{P}^*y = 0$

$\mathbf{P}^* = \mathbf{I} - \mathbf{P}$.

because for all $x \in \mathcal{U}^\perp$, $(\mathbf{I} - \mathbf{P})x = x$ and for all $x \in \mathcal{U}$, $(\mathbf{I} - \mathbf{P})x = 0$, satisfied with the analysis above.

3) **proof**:

On the one hand, $\forall x_u \in \mathcal{U}, \exists x \in \mathcal{V}$ s.t. $x_u = \mathbf{P}_\mathcal{U}x$ $\forall x_w \in \mathcal{W}, \exists y \in \mathcal{V}$ s.t. $y_w = \mathbf{P}_\mathcal{W}y$.

if $\mathbf{P}_\mathcal{U}\mathbf{P}_\mathcal{W} = 0$, and $\mathbf{P}_\mathcal{U}x_u = 0$, $\mathbf{P}_\mathcal{W}y_w = 0$ so $x_u^T\mathbf{P}_\mathcal{U}^T\mathbf{P}_\mathcal{W}x_w = 0$

So for any $z_u \in \mathcal{U}$ and because for any $z_w \in \mathcal{W}$ $z_u^T \perp z_w$

$\therefore \mathcal{U} \perp \mathcal{W}$

On the other hand, if $\mathcal{U} \perp \mathcal{W}$ So for any $z_u \in \mathcal{U}$ and for any $z_w \in \mathcal{W}$ $z_u^T z_w = 0$

so for any $z_u \in \mathcal{U}$, we can find a vector $x \in V$ s.t. $\mathbf{P}_\mathcal{U}x = z_u$

and for any $z_v \in \mathcal{V}$ we can find a vector $y \in V$ s.t. $\mathbf{P}_\mathcal{W}y = z_w$

$\because z_u^T z_w = 0$, therefore $(\mathbf{P}_\mathcal{U}x)^T\mathbf{P}_\mathcal{W}y = 0$

So $x^T\mathbf{P}_\mathcal{U}^T\mathbf{P}_\mathcal{W}y = 0$

$\therefore \mathbf{P}_\mathcal{U}^T\mathbf{P}_\mathcal{W} = 0$

II. LEAST SQUARE (LS) PROGRAMMING.

Problem 2. (10 points + 10 points + 5 points)

Write programs to solve the least square problem with specified methods, any programming language is suitable.

$$\mathbf{x} = \arg\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}), \quad f(\mathbf{x}) = ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a matrix representing the predefined data set with m data samples of n dimensions (m=1000, n=210), and $\mathbf{y} \in \mathbb{R}^m$ represents the labels. The data samples are provided in the "data.txt" file, and the labels are provided in the "label.txt" file, you are supposed to load the data before solving the problem.

1) Solve the LS with gradient decent method.

   The gradient descent method for solving problem updates $\mathbf{x}$ as

$$\mathbf{x} = \mathbf{x} - \gamma \cdot \nabla_{\mathbf{x}} f(\mathbf{x}),$$

   where γ is the step size of the gradient decent methods. We suggest that you can set $\gamma = 1e - 5$.

2) Solve the LS by the method of normal equation with Cholesky decomposition and forward/backward substitution.

3) Compare two methods above.

   (a) Basing on the true running results from the program, count the number of "flops"*;

   (b) Compare gradient norm and loss $f(\mathbf{x})$ for results $\mathbf{x} = \mathbf{x_{LS}}$ of above two algorithms.

Notation*: "flop": one flop means one floating point operation, i.e., one addition, subtraction, multiplication, or division of two floating-point numbers, in this problem each floating points operation $+, -, \times, \div, \sqrt{\cdot}$ counts as one "flop".

Hint for gradient decent programming:

1) **Step size selection**: to ensure the convergence of the method, γ is supposed to be selected properly (large step size may accelerate the convergence rate but also may lead to instability, A sufficiently small compensation always ensures that the algorithm converges).

2) **Terminal condition**: the gradient decent is an iteration algorithm that need a terminal condition. In this problem, the algorithm can stop when the gradient of the loss function $f(\mathbf{x})$ at current $\mathbf{x}$ is small enough.

Remarks:

- The solution of the two methods should be printed in files named "sol1.txt" and "sol2.txt" and submitted in gradescope. The format should be same as the input file (210 rows plain text, each rows is a dimension of the final solution).

- Make sure that your codes are executable and are consistent with your solutions.

Solution.

1).You can check in my uploaded file.

2).You can check in my uploaded file.

3). (a)

n = 210 , m = 1000

(1).Gradient descent method:

The stop condition used here is that the change value of cost is less than 1e-6, and a total of k = 85 iterations are performed.

Each iteration needs to be calculated:

1. $X * theta$

2. $X' * (predictions - y)$

3. $theta = theta - alpha * updates$

Assuming that C iterations are required to complete, the time complexity of each iteration is $flop_i = n^2 m + n^2 + mn + 3n$

(2). Normal Equations:

1. $L = chol(X' * X)$;

2. $temp = forwardsub(L', X' * y)$;

3. $theta = backwardsub(L, temp)$;

chol: A total of $O\left(n^2\right)$ elements are calculated, and the calculation of each element contains the multiplication of $O(n)$, so it is $O\left(n^3\right)$

For forward substitution and backward substitution algorithms, since L or L'is a triangular matrix, the first one can be calculated in the order of $y_1, y_2, \cdots$, the required calculation The complexity is $O\left(n^2\right)$, Similarly, the third one can also be calculated like this.

So the total computational complexity is $O\left(n^3 + n^2\right) = O\left(n^3\right)$ flops .The advantage of this algorithm is that when it is necessary to solve different x with relatively different b for the same A, the new part can be completed with $O\left(n^2\right)$ complexity.

(b)

```
>> norm(theta_grad)
ans =
   82.9163565517039
>> norm(theta_normal)
ans =
   82.9165386448065

>> computeCostMulti(X, y, theta_grad)
ans =
   13.3213102758129
>> computeCostMulti(X, y, theta_normal)
ans =
   13.3213070317204
```

theta_grad is the result of gradient descent and theta_norm is the result of norm equation.

norm() fuction will calculate the L2 norm of the vector,and computeCostMulti() function will calculate the loss for $f(x)$

III. UNDERSTANDING THE QR FACTORIZATION

Problem 3 [Understanding the Gram-Schmidt algorithm.]. (5 points + 7 points + 6 points + 7 points)

1) Consider the subspace $\mathcal{S}$ spaned by $\{\mathbf{a}_1, \ldots, \mathbf{a}_4\}$, where

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 3 \\ 4 \\ 5 \\ 6 \end{bmatrix}, \quad \mathbf{a}_4 = \begin{bmatrix} 3 \\ 5 \\ 7 \\ 11 \end{bmatrix}.$$

Use the **classical** Gram-Schimidt algorithm (See Algorithm 1), find a set of orthonormal basis $\{\mathbf{q}_i\}$ for $\mathcal{S}$ by hand (derivation is expected). Do not use decimals in your answers, fraction and n-th roots of numbers are accepted. Verify the orthonormality of the found basis.

Algorithm 1: Classical Gram-Schmidt algorithm

    Input : A collection of linearly independent vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$.

1 **Initilization:** $\widetilde{\mathbf{q}}_1 = \mathbf{a}_1, \mathbf{q}_1 = \widetilde{\mathbf{q}}_1 / \|\widetilde{\mathbf{q}}_1\|_2$

2 **for** $i = 2, \ldots, n$ **do**

3      $\widetilde{\mathbf{q}}_i = \mathbf{a}_i - \sum_{j=1}^{i-1} (\mathbf{q}_j^T \mathbf{a}_i) \mathbf{q}_j$

4      $\mathbf{q}_i = \widetilde{\mathbf{q}}_i / \|\widetilde{\mathbf{q}}_i\|_2$

5 **end**

    Output: $\mathbf{q}_1, \ldots, \mathbf{q}_n$

---

2) Orthogonal projection of vector $\mathbf{a}$ onto a nonzero vector $\mathbf{b}$ is defined as

$$\text{proj}_{\mathbf{b}}(\mathbf{a}) = \frac{\langle \mathbf{a}, \mathbf{b} \rangle}{\langle \mathbf{b}, \mathbf{b} \rangle} \mathbf{b},$$

where $\langle, \rangle$ denotes the inner product of vectors. And for subspace $\mathcal{M}$ with orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$, the orthogonal projector onto subspace $\mathcal{M}$ is given by

$$\mathbf{P} = \mathbf{U}\mathbf{U}^T, \quad \mathbf{U} = [\mathbf{u}_1 | \cdots | \mathbf{u}_k].$$

In the context of **projection of vector** and **projection onto subspace** respectively, can you give another two understandings of the classical Gram-Schmidt algorithm?

3) Consider the subspace $\mathcal{S}$ spaned by $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$,

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ \epsilon \\ \epsilon \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 1 \\ \epsilon \\ 0 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 1 \\ 0 \\ \epsilon \end{bmatrix},$$

where ϵ is a small real number such that $1 + k\epsilon^2 = 1$ ($k \in \mathbb{N}^+$). First complete the pseudo algorithm in Algorithm 3. Then use the **classical** Gram-Schimidt algorithm and the **modified** Gram-Schimidt algorithm respectively, find two sets of basis for $\mathcal{S}$ by hand (derivation is expected). Are the two sets of basis the same? If not, which one is the desired orthonormal basis? Report what you have found.

Algorithm 2: Modified Gram-Schmidt algorithm

Input : A collection of linearly independent vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$.

1 **Initilization:**

2 *Complete your algorithm here...*

Output: $\mathbf{q}_1, \ldots, \mathbf{q}_n$

---

4) **Programming part:** In this part, you are required to code both the **classical Gram-Schmidt** and **the modified Gram-Schmidt** algorithms. For $\epsilon = 1e-4$ and $\epsilon = 1e-9$ in sub-problem 2), give the outputs of two algorithms and calculate $\|\mathbf{Q}^T\mathbf{Q} - \mathbf{I}\|_F$, where $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3]$.

Remarks:

- Coding languages are not restricted, but do not use built-in function such as `qr`.
- When handing in your homework in gradescope, package all your codes into your_student_id+hw3_code.zip and upload. In the package, you also need to include a file named README.txt/md to clearly identify the function of each file.
- Make sure that your codes can run and are consistent with your solutions.

Solution.

1) 1) According to Algorithm

$\tilde{\mathbf{q}}_1 = \mathbf{a_1} = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}^T$.

$\mathbf{q}_1 = \tilde{\mathbf{q}}_1 / \|\tilde{\mathbf{q}}_1\|_2 = \frac{1}{\sqrt{30}} \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}^T$

$\tilde{\mathbf{q}}_2 = \mathbf{a}_2 - \left(\mathbf{q}_1^T \mathbf{a}_2\right) \mathbf{q}_1 = \begin{bmatrix} 2 & 3 & 4 & 5 \end{bmatrix}^T - \frac{1}{3} \begin{bmatrix} 4 & 8 & 12 & 16 \end{bmatrix}^T = \frac{1}{3} \begin{bmatrix} 2 & 1 & 0 & -1 \end{bmatrix}^T$

$\mathbf{q}_2 = \tilde{\mathbf{q}}_2 / \|\tilde{\mathbf{q}}_2\|_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} 2 & 1 & 0 & -1 \end{bmatrix}^T$

$$\tilde{\mathbf{q_3}} = \mathbf{a_3} - \left(\mathbf{q_1}^T \mathbf{a_3}\right) \mathbf{q_1} - \left(\mathbf{q_2}^T \mathbf{a_3}\right) \mathbf{q_2} = 0,$$

So $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4)$ are not linearly independent. We can get the set of orthonormal basis from a_1, a_2, a_4

$\tilde{\mathbf{q_4}} = \mathbf{a_4} - \left(\mathbf{q_1}^T \mathbf{a_4}\right) \mathbf{q_1} - \left(\mathbf{q_2}^T \mathbf{a_4}\right) \mathbf{q_2} = \frac{1}{5} \begin{bmatrix} 2 & -1 & -4 & 3 \end{bmatrix}^T, \quad \mathbf{q_4} = \frac{\bar{\mathbf{q}}_4}{\|\bar{\mathbf{q}}_4\|} = \frac{\sqrt{30}}{30} \begin{bmatrix} 2 & -1 & -4 & 3 \end{bmatrix}^T$

We can get it easily that $\mathbf{q_1}^T \mathbf{q_2} = \mathbf{q_1}^T \mathbf{q_4} = \mathbf{q_2}^T \mathbf{q_4} = 0$, so set of the basis $\mathbf{q_1}, \mathbf{q_2}, \mathbf{q_4}$ is orthonormal basis.

$$\mathbf{Q} = \begin{bmatrix} \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{6}} & \frac{2}{\sqrt{30}} \\ \frac{2}{\sqrt{30}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{30}} \\ \frac{3}{\sqrt{30}} & 0 & -\frac{4}{\sqrt{30}} \\ \frac{1}{\sqrt{30}} & -\frac{1}{\sqrt{6}} & \frac{3}{\sqrt{30}} \end{bmatrix}$$

we can verify the orthnormality of the basis via

$$Q^T Q = I$$

2) (1) projection of vector:

$(q_1^T a_2) q_1$ is the orthogonal projection of vector a_2 onto the vector q_1, so $\tilde{q}_2 = a_2 - (q_1^T a_2) q_1$ indicates a vector that a_2 eliminated q_1 direction. by parity of reasoning, $\tilde{\mathbf{q}}_i = \mathbf{a}_i - \sum_{j=1}^{i-1} \text{proj}_{\mathbf{q}}(\mathbf{a}_i)$ indicates a vector that a_i eliminated $q_1, q_2 \cdots q_{i-1}$ directions.

(2) projection onto subspace:

$$\mathbf{P} = \mathbf{U}\mathbf{U}^T, \quad \mathbf{U} = |\mathbf{u}_1| \cdots |\mathbf{u}_k|$$

Then we have

$$\tilde{\mathbf{q}}_k = (\mathbf{I} - \mathbf{Q}_k \mathbf{Q}_k^T) \mathbf{a}_k$$

where

$$\mathbf{Q}_k = [\mathbf{q}_1| \cdots |\mathbf{q}_{k-1}]$$

First unit the vector a_1, get vector q_1, then we have a orthogonal projector matrix $\mathbf{P}_1 = \mathbf{Q}_1 \mathbf{Q}_1^T, \mathbf{Q}_1 = [q_1]$ $\tilde{q}_2 = a_2 - (q_1^T a_2) q_1 = a_2 - \mathbf{Q}_1 \mathbf{Q}_1^T a_2$ indicates a vector that a_2 eliminated P_1 direction. by parity of reasoning, $\tilde{\mathbf{q}}_i = \mathbf{a}_i - \sum_{j=1}^{i-1} (\mathbf{q}_j^T \mathbf{a}_i) \mathbf{q}_j = \mathbf{a}_i - \mathbf{Q}_i \mathbf{Q}_i^T \mathbf{A}_i$ indicates a vector that a_i eliminated $\mathbf{P}_i$ direction. $(\mathbf{A}_i = [a_1, a_2 \cdots a_i])$, the obtained $\{\tilde{\mathbf{q}}_1, \ldots, \tilde{\mathbf{q}}_n\}$ are orthogonal and $\{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$ are orthonomal.

3) classical Gram-Schimidt

$$\tilde{q}_1 = a_1, \quad q_1 = \frac{\tilde{q}_1}{\|\tilde{q}_1\|} = \begin{bmatrix} 1 & \epsilon & \epsilon \end{bmatrix}^T$$
$$\tilde{q}_2 = a_2 - (q_1^T a_2) q_1 = \begin{bmatrix} 0 & 0 & -\epsilon \end{bmatrix}^T, \quad q_2 = \frac{\tilde{q}_2}{\|\tilde{q}_2\|} = \begin{bmatrix} 0 & 0 & -1 \end{bmatrix}^T$$
$$\tilde{q}_3 = a_3 - (q_1^T a_3) q_1 - (q_2^T a_3) q_2 = \begin{bmatrix} 0 & -\epsilon & -\epsilon \end{bmatrix}^T, \quad q_3 = \frac{\tilde{q}_3}{\|\tilde{q}_3\|} = \begin{bmatrix} 0 & -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}^T$$

modified Gram-Schimidt

$$\tilde{q}_1 = a_1, \quad q_1 = \frac{\tilde{q}_1}{\|\tilde{q}_1\|} = \begin{bmatrix} 1 & \epsilon & \epsilon \end{bmatrix}^T$$

$$\tilde{q}_2 = a_2 - (q_1^T a_2) q_1 = \begin{bmatrix} 0 & 0 & -\epsilon \end{bmatrix}^T, \quad q_2 = \frac{\tilde{q}_2}{\|\tilde{q}_2\|} = \begin{bmatrix} 0 & 0 & -1 \end{bmatrix}^T$$

$$\tilde{q}_3^{(1)} = a_3 - (q_1^T a_3) q_1 = \begin{bmatrix} 0 & -\epsilon & 0 \end{bmatrix}$$

$$\tilde{q}_3^{(2)} = \tilde{q}_3^{(1)} - (q_2^T \tilde{q}_3^{(1)}) q_2 = \begin{bmatrix} 0 & -\epsilon & 0 \end{bmatrix}$$

$$\tilde{q}_3 = \tilde{q}_3^{(2)} = \begin{bmatrix} 0 & -\epsilon & 0 \end{bmatrix}^T, \quad q_3 = \frac{\tilde{q}_3}{\|\tilde{q}_3\|} = \begin{bmatrix} 0 & -1 & 0 \end{bmatrix}^T$$

Therefore,classical Gram-Schmidt returns

$$\mathbf{q}_1 = \begin{bmatrix} 1 \\ \epsilon \\ \epsilon \end{bmatrix}, \quad \mathbf{q}_2 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}, \quad \mathbf{q}_3 = \begin{bmatrix} 0 \\ -1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$$

which is not a set of orthonomal basis since q_2 and q_3 are not orthogonal. And the modified Gram-Schmidt algorithm returns,

$$\mathbf{q}_1 = \begin{bmatrix} 1 \\ \epsilon \\ \epsilon \end{bmatrix}, \quad \mathbf{q}_2 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}, \quad \mathbf{q}_3 = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$$

The two sets of basis are different.The result of classical Gram-Schimidt, q_2 and q_3 are not orthogonal.However, the set of basis derived by modified Gram-Schimidt algorithm is the desired orthogonal basis.

4) For $\epsilon = 1e - 4$

$$q_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \text{ (CGS and MGS)}, \mathbf{q}_2 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \text{ (CGS and MGS)}, \mathbf{q}_3 = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} \text{ (CGS and MGS)}.$$

$$\|Q^T Q - I\|_F < 1e - 7 (CGS)$$

$$\|Q^T Q - I\|_F < 1e - 10 (MGS)\|.$$

For $\epsilon = 1e - 9$

$$\mathbf{q}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \text{ (CGS)}, \mathbf{q}_2 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \text{ (CGS)}, \mathbf{q}_3 = \frac{\sqrt{2}}{2} \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix} \text{ (CGS)}.$$

$$\|\mathbf{Q}^T \mathbf{Q} - \mathbf{I}\|_F = 1 (CGS)$$

$$\mathbf{q}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \text{ (MGS)}, \mathbf{q}_2 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \text{ (MGS)}, \mathbf{q}_3 = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} \text{ (MGS)}$$

$$(\mathbf{Q}^T \mathbf{Q} - \mathbf{I}\|_F < 1e - 7 (\mathbf{MGS})$$

Algorithm 3: Modified Gram-Schmidt algorithm

Input : A collection of linearly independent vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$.

1 **Initilization:** $\tilde{\mathbf{q}}_1 = \mathbf{a}_1, \textbf{for} \quad i = 1, ..., n$

2 for $i = 1, \ldots, n$ do

3      $\mathbf{q}_i = \tilde{\mathbf{q}}_i / \tilde{\mathbf{q}}_i\|_2$

4   for $k = i + 1, \ldots, n$ do

5          $\tilde{\mathbf{q}}_k = \tilde{\mathbf{q}}_k - \left(\mathbf{q}_i^T \tilde{\mathbf{q}}_k \right) \mathbf{q}_i$

6   end

7 end

Output: $\mathbf{q}_1, \ldots, \mathbf{q}_n$

IV. SOLVING LS VIA QR FACTORIZATION AND NORMAL EQUATION

Problem 4 [Understanding the influence of the condition number to the solution.]. (4 points + 5 points + 4 points + 4 points + 3 points points)

Consider such two LS problems:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 \tag{1}$$

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x} - (\mathbf{b} + \delta\mathbf{b})\|_2^2$$

with $\mathbf{A} \in \mathbb{R}^{m\times n}$. For $\mathbf{b} = \begin{bmatrix} 1 & 3/2 & 3 & 6 \end{bmatrix}^T$ and $\delta\mathbf{b} = \begin{bmatrix} 1/10 & 0 & 0 & 0 \end{bmatrix}^T$,

1) Computing solution to the problem (1) via QR decomposition when

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 5 \\ 3 & 4 & 7 \\ 4 & 5 & 11 \end{bmatrix}.$$

2) For a full-rank matrix $\mathbf{A}$, consider the equation $\mathbf{A}\mathbf{x} = \mathbf{b}$, after adding some noise $\delta\mathbf{b}$ to $\mathbf{b}$, we have $\mathbf{A}(\mathbf{x}+\delta\mathbf{x}) = \mathbf{b} + \delta\mathbf{b}$, and then proof

$$\frac{1}{\|\mathbf{A}\|\|\mathbf{A}^\dagger\|}\frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|} \le \frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \le \|\mathbf{A}\|\|\mathbf{A}^\dagger\|\frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|},$$

and give it a plain interpretation.

3) Computing the solutions to the two LS problems via the normal equation $\mathbf{A}^T\mathbf{A}\mathbf{x}_{LS} = \mathbf{A}^T\mathbf{b}$ when

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \\ 1 & 1 & 0 \end{bmatrix}.$$

4) Computing the solutions to the two LS problems via the normal equation $\mathbf{A}^T\mathbf{A}\mathbf{x}_{LS} = \mathbf{A}^T\mathbf{b}$ when

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix}.$$

5) Compare the 2-norm condition number $\|\mathbf{A}\|\|\mathbf{A}^\dagger\|$ for $\mathbf{A}$ in 3) and 4) and the influence on the solution to problem (1) resulted by the additional noise $\delta\mathbf{b}$.

Hint: Show the influence on the solution by $\frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|}$.

Remarks: You can use MATLAB for some matrix computations (deviation is expected) in 3), 4), 5). Do not use decimals in your answers, fraction and n-th roots of numbers are accepted.

Solution.

1) 1) $\mathbf{A} = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}$, we have

$$q_1 = \frac{a_1}{\|a_1\|} = \frac{\sqrt{30}}{30} \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}^T$$

$$\bar{q}_2 = a_2 - \left(q_1^T a_2\right) q_1 = \frac{1}{3} \begin{bmatrix} 2 & 1 & 0 & -1 \end{bmatrix}^T$$

$$q_2 = \frac{\bar{q}_2}{\|\bar{q}_2\|} = \frac{\sqrt{6}}{6} \begin{bmatrix} 2 & 1 & 0 & -1 \end{bmatrix}^T$$

$$\bar{q}_3 = a_3 - \left(q_1^T a_3\right) q_1 - \left(q_2^T a_3\right) q_1 = \frac{1}{5} \begin{bmatrix} 2 & -1 & -4 & 3 \end{bmatrix}^T$$

$$q_3 = \frac{\bar{q}_3}{\|\bar{q}_3\|} = \frac{\sqrt{30}}{30} \begin{bmatrix} 2 & -1 & -4 & 3 \end{bmatrix}^T$$

$$r_{11} = \|\overline{a_1}\| = \frac{\sqrt{30}}{30}$$

$$r_{22} = \|\overline{q_2}\| = \frac{\sqrt{6}}{3}$$

$$r_{33} = \|\overline{q_3}\| = \frac{\sqrt{30}}{5}$$

$$r_{12} = q_1^T a_2 = \sqrt{30}$$

$$r_{13} = q_1^T a_3 = \frac{13\sqrt{30}}{5}$$

$$r_{23} = q_2^T a_3 = 0$$

$$\mathbf{Q} = \begin{bmatrix} \frac{\sqrt{30}}{30} & \frac{2\sqrt{6}}{6} & \frac{2\sqrt{30}}{300} \\ \frac{2\sqrt{30}}{30} & \frac{\sqrt{6}}{6} & \frac{-\sqrt{30}}{30} \\ \frac{3\sqrt{30}}{30} & 0 & -\frac{4\sqrt{30}}{30} \\ \frac{4\sqrt{30}}{30} & -\frac{\sqrt{6}}{6} & \frac{3\sqrt{30}}{30} \end{bmatrix}$$

$$\mathbf{R} = \begin{bmatrix} \frac{\sqrt{30}}{30} & \frac{4\sqrt{30}}{3} & \frac{1.3\sqrt{30}}{5} \\ 0 & \frac{\sqrt{6}}{3} & 0 \\ 0 & 0 & \frac{\sqrt{30}}{5} \end{bmatrix}$$

$$\mathbf{Ax} = \mathbf{b} \rightarrow \mathbf{QRx} = \mathbf{b} \rightarrow \mathbf{Rx} = \mathbf{Q}^T b$$

$$\begin{bmatrix} \sqrt{30} & \frac{40}{\sqrt{30}} & \frac{78}{\sqrt{30}} \\ 0 & \frac{2}{\sqrt{6}} & 0 \\ 0 & 0 & \frac{6}{\sqrt{30}} \end{bmatrix} \cdot \mathbf{x}_{LS} = \begin{bmatrix} \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & \frac{3}{\sqrt{30}} & \frac{4}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{30}} & -\frac{1}{\sqrt{30}} & -\frac{4}{\sqrt{30}} & \frac{3}{\sqrt{30}} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 3/2 \\ 3 \\ 6 \end{bmatrix} = \begin{bmatrix} \frac{37}{\sqrt{30}} \\ -\frac{5}{2\sqrt{6}} \\ \frac{13}{2\sqrt{30}} \end{bmatrix}$$

solve the equation, we have $\mathbf{x}_{LS} = \begin{bmatrix} \frac{1}{12} \\ -\frac{5}{4} \\ \frac{13}{12} \end{bmatrix}$

2) since $x + \delta x = \mathbf{A}^\dagger (b + \delta b)$ and $x = \mathbf{A}^\dagger b$, we have $\delta x = \mathbf{A}^\dagger \delta b$,

so $\|\delta x\| = \|\mathbf{A}^\dagger \delta b\| \leq \|\mathbf{A}^\dagger\| \|\delta b\|$.

since $\mathbf{A}x = b$, we have $\frac{1}{\|x\|} \leq \frac{\|b\|}{\|x\|}$,

so $\frac{\|\delta x\|}{\|x\|} \leq \|\mathbf{A}\| \|\mathbf{A}^\dagger\| \frac{\|\delta b\|}{\|\mathbf{b}\|}$.

Similarly, $\mathbf{A}(x + \delta x) = b + \delta b$, we have $\delta b = \mathbf{A}$,

so $\|\delta b\| \leq \|\mathbf{A}\| \|\delta x\|.x = \mathbf{A}^\dagger b$,

we have $\|x\| \leq \|\mathbf{A}^\dagger\| \|b\|$, so $\frac{1}{\|\mathbf{A}\| \|\mathbf{A}'\|} \frac{\|\delta \mathbf{b}\|}{\|\mathbf{b}\|} \leq \frac{\|\delta \mathbf{x}\|}{\|\mathbf{x}\|}$

Above all

$$\frac{1}{\|\mathbf{A}\|\,\|\mathbf{A}^\dagger\|}\frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|} \le \frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \le \|\mathbf{A}\|\,\|\mathbf{A}^\dagger\|\frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|},$$

interpretation:

When we think about the relative error of x, If b changes, the sensitivity of the least square solution depends on $\|\mathbf{A}\|\,\|\mathbf{A}^\dagger\|$ And if $\|\mathbf{A}\|\,\|\mathbf{A}^\dagger\|$ is large ,though b changes little, the error of x may change large.

3)

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 15 & 16 & 15 \\ 16 & 18 & 17 \\ 15 & 17 & 17 \end{bmatrix} \quad \mathbf{A}^T\mathbf{b} = \begin{bmatrix} 19 \\ 20 \\ 14 \end{bmatrix}$$

$$\mathbf{A}^T(\mathbf{b}+\delta\mathbf{b}) = \begin{bmatrix} \frac{191}{10} \\ \frac{101}{5} \\ \frac{71}{5} \end{bmatrix}$$

Solve the cquation $\mathbf{x}_{LS1} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}$,we have

$$\mathbf{x}_{Ls1} = \begin{bmatrix} -1 & 4/13 & 6/13 & 0 \\ 1 & -4/13 & -6/13 & 1 \\ 0 & 2/13 & 3/13 & -1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 3/2 \\ 3 \\ 6 \end{bmatrix} = \begin{bmatrix} 11/13 \\ 67/13 \\ -66/13 \end{bmatrix}$$

Solve the cquation $\hat{\mathbf{x}}_{LS2} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T(\mathbf{b}+\delta\mathbf{b})$,we have

$$\hat{\mathbf{x}}_{\mathbf{LS2}} = \begin{bmatrix} -1 & 4/13 & 6/13 & 0 \\ 1 & -4/13 & -6/13 & 1 \\ 0 & 2/13 & 3/13 & -1 \end{bmatrix} \cdot \begin{bmatrix} 11/10 \\ 3/2 \\ 3 \\ 6 \end{bmatrix} = \begin{bmatrix} 97/130 \\ 683/130 \\ -66/13 \end{bmatrix}$$

4)

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 4 & 10 & 30 \\ 10 & 30 & 100 \\ 30 & 100 & 357 \end{bmatrix} \quad \mathbf{A}^T\mathbf{b} = \begin{bmatrix} \frac{23}{2} \\ 37 \\ 130 \end{bmatrix}$$

Solve the equation $\mathbf{x}_{LS1} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}$ we have

$$\mathbf{x}_{LS1} = \begin{bmatrix} 9/4 & -3/4 & -5/4 & 3/4 \\ -31/20 & 23/20 & 27/20 & -19/20 \\ 1/4 & -1/4 & -1/4 & 1/4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 3/2 \\ 3 \\ 6 \end{bmatrix} = \begin{bmatrix} 15/8 \\ -59/40 \\ 5/8 \end{bmatrix}$$

Solve the cquation $\hat{\mathbf{x}}_{LS2} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T(\mathbf{b} + \delta\mathbf{b})$, we have

$$\mathbf{x}_{Ls2} = \hat{\mathbf{x}}_{LS} = \begin{bmatrix} 9/4 & -3/4 & -5/4 & 3/4 \\ -31/20 & 23/20 & 27/20 & -19/20 \\ 1/4 & -1/4 & -1/4 & 1/4 \end{bmatrix} \cdot \begin{bmatrix} 11/10 \\ 3/2 \\ 3 \\ 6 \end{bmatrix} = \begin{bmatrix} 21/10 \\ -163/100 \\ 13/20 \end{bmatrix}$$

5)

$$A^\dagger = \left(A^TA\right)^{-1}A^T$$

$$k_1 = \|\mathbf{A}\|\,\|\mathbf{A}^\dagger\| = \frac{5610}{421}$$

$$k_2 = \|\mathbf{A}\|\,\|\mathbf{A}^\dagger\| = \frac{2653}{36}$$

$$\frac{\|\hat{\mathbf{x}}_{LS} - \mathbf{x}_{LS}\|_2}{\|\mathbf{x}_{LS}\|_2} = \frac{125}{6438}$$

$$\frac{\|\hat{\mathbf{x}}_{LS} - \mathbf{x}_{LS}\|_2}{\|\mathbf{x}_{LS}\|_2} = \frac{1231}{11065}$$

V. Underdetermined System

Problem 5 [Solving Underdetermined System by QR]. (10 points + 5 points)

Consider the following underdetermined system $\mathbf{A}\mathbf{x} = \mathbf{b}$ with $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $m < n$. Let

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 2 & 0 \\ 0 & -2 & 2 & 1 \\ 2 & 5 & 6 & 1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix},$$

1) Use Householder reflection to give the full QR decomposition of tall $\mathbf{A}^T$, i.e., $\mathbf{A}^T = \mathbf{Q}\mathbf{R}$ with $\mathbf{Q}$ being a square matrix with orthonormal columns.

2) Give one possible solution via QR decomposition of $\mathbf{A}^T$, write down your solution using $\mathbf{b}$.

Solution.

1)

$$\mathbf{A^T} = \begin{bmatrix} 1 & 0 & 2 \\ 2 & -2 & 5 \\ 2 & 2 & 6 \\ 0 & 1 & 1 \end{bmatrix} \mathbf{e_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \mathbf{u_1} = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 0 \end{bmatrix} - 3\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \\ 2 \\ 0 \end{bmatrix} \quad P_1 = \mathbf{I} - 2\frac{u_1 u_1^T}{u_1^T u_1} = \frac{1}{3}\begin{bmatrix} 1 & 2 & 2 & 0 \\ 2 & 1 & -2 & 0 \\ 2 & -2 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$P_1 \mathbf{A^T} = \begin{bmatrix} 3 & 0 & 8 \\ 0 & -2 & -1 \\ 0 & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix} \mathbf{e_2} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \mathbf{u_2} = \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix} - 3\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -5 \\ 2 \\ 1 \end{bmatrix}$$

$$P_2^* = \mathbf{I} - 2\frac{u_2 u_2^T}{u_2^T u_2} = \frac{1}{3}\begin{bmatrix} 2 & -2 & -1 \\ -2 & -1 & -2 \\ -1 & -2 & 2 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 1 & 0 \\ 0 & P_2^* \end{bmatrix}$$

$$P_2 P_1 \mathbf{A^T} = \begin{bmatrix} 3 & 0 & 8 \\ 0 & 3 & 1 \\ 0 & 0 & \frac{4}{5} \\ 0 & 0 & \frac{3}{5} \end{bmatrix} \mathbf{e_2} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \mathbf{u_3} = \begin{bmatrix} -\frac{4}{5} \\ \frac{3}{5} \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -\frac{9}{5} \\ \frac{3}{5} \end{bmatrix}$$

$$P_3^* = \mathbf{I} - 2\frac{u_3 u_3^T}{u_3^T u_3}, \quad P_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & P_3^* \end{bmatrix}$$

$$P_3 P_2 P_1 \mathbf{A^T} = \begin{bmatrix} 3 & 0 & 8 \\ 0 & 3 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{A}^T = P_1 P_2 P_3 \mathbf{R} = \mathbf{Q}\mathbf{R} = \begin{bmatrix} \frac{1}{3} & 0 & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} & 0 \\ \frac{2}{3} & \frac{2}{3} & 0 & -\frac{1}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \end{bmatrix} \times \begin{bmatrix} 3 & 0 & 8 \\ 0 & 3 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

2) Solving Underdetermined Systems by QR For $A \in R^{m \times n}$ with $m < n$ and $\text{rank}(A) = m$, we have

$$\mathbf{A}^T = \mathbf{QR} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix} = \mathbf{Q}_1 \mathbf{R}_1 + \mathbf{Q}_2 \mathbf{0}$$

$$\mathbf{Ax} = \mathbf{R}_1^\top \mathbf{Q}_1^T \mathbf{x} + \mathbf{0}^T \mathbf{Q}_2^T \mathbf{x} = \mathbf{b}$$

which indicates

$$\mathbf{Q}_1^T \mathbf{x} = \mathbf{R}_1^{-T} \mathbf{b}$$

and $Q_2^T x$ can be anything I set it to be 0

$$\begin{bmatrix} \mathbf{Q}_1^\top \mathbf{x} \\ \mathbf{Q}_2^\top \mathbf{x} \end{bmatrix} = \mathbf{Q}^T \mathbf{x} = \begin{bmatrix} \mathbf{R}_1^{-T} \mathbf{b} \\ \mathbf{0} \end{bmatrix}$$

hence, one possible solution is

$$R_1^{-T} b = \begin{bmatrix} 3 & 0 & 8 \\ 0 & 3 & 1 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \qquad x = \begin{bmatrix} \frac{1}{3} b_1 - \frac{8}{3} b_3 \\ \frac{1}{3} b_2 - \frac{1}{3} b_3 \\ b_3 \\ 0 \end{bmatrix}$$

$$x = Q \begin{bmatrix} R_1^{-T} b \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & 0 & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} & 0 \\ \frac{2}{3} & \frac{2}{3} & 0 & -\frac{1}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \end{bmatrix} \times \begin{bmatrix} \frac{1}{3} & 0 & -\frac{8}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{9} & 0 & -\frac{8}{9} & -\frac{2}{3} \\ \frac{2}{9} & -\frac{2}{9} & -2 & \frac{1}{3} \\ \frac{2}{9} & \frac{2}{9} & \frac{14}{9} & 0 \\ 0 & \frac{1}{9} & \frac{1}{9} & -\frac{2}{3} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ 0 \end{bmatrix}$$

So x above is one possible solution.

VI. Solving LS via Projection

Problem 6. (Bonus question, 6 points + 4 points)

Consider the Least Square (LS) problem:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2 \tag{2}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ ($m > n$) may not be full rank. Denote

$$X_{\mathrm{LS}} = \left\{\mathbf{x} \in \mathbb{R}^n | \mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{y}\right\}$$

as the set of all solutions to (2), and

$$\mathbf{x}_{\mathrm{LS}} = \mathbf{A}^\dagger\mathbf{y}$$

where $\mathbf{A}^\dagger \in \mathbb{R}^{n\times m}$ is the *pseudo inverse of* $\mathbf{A}$ satisfies the following properties:

1) $\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}$.
2) $\mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\dagger$.
3) $(\mathbf{A}\mathbf{A}^\dagger)^T = \mathbf{A}\mathbf{A}^\dagger$.
4) $(\mathbf{A}^\dagger\mathbf{A})^T = \mathbf{A}^\dagger\mathbf{A}$.

Answer the following questions:

1) Prove that $\mathbf{x}_{\mathrm{LS}}$ is a solution to (2) and is of minimum 2-norm in X_{LS}, that is

$$\mathbf{x}_{\mathrm{LS}} = \arg\min_{\mathbf{x}\in X_{\mathrm{LS}}} \|\mathbf{x}\|_2 \ .$$

Hint. Notice that the orthogonal projection onto $\mathcal{N}(A)$ is given by

$$\mathbf{\Pi}_{\mathcal{N}(A)} = \mathbf{I} - \mathbf{A}^\dagger\mathbf{A}$$

2) Prove that $X_{\mathrm{LS}} = \{\mathbf{x}_{\mathrm{LS}}\}$ if and only if $\mathrm{rank}(\mathbf{A}) = n$.

Solution.

1) $\because x_{LS} = \mathbf{A}^\dagger y$

$$\mathbf{A}^T\mathbf{A}\mathbf{x}_{LS} = \mathbf{A}^T\mathbf{A}\mathbf{A}^\dagger y$$

$\because \left(\mathbf{A}\mathbf{A}^\dagger\right)^T = \mathbf{A}\mathbf{A}^\dagger$

$$\mathbf{A}^T\mathbf{A}\mathbf{A}^\dagger y = \mathbf{A}^T\left(\mathbf{A}\mathbf{A}^\dagger\right)^T y$$

$$\mathbf{A}^T\left(\mathbf{A}\mathbf{A}^\dagger\right)^T y = \mathbf{A}^T\left(\mathbf{A}^\dagger\right)^T\mathbf{A}^T y = \left(\mathbf{A}\mathbf{A}^\dagger\mathbf{A}\right)^T y$$

$\because \mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}$

$$\left(\mathbf{A}\mathbf{A}^\dagger\mathbf{A}\right)^T y = \mathbf{A}^T y$$

We have $x_{LS} \in X_{LS}$, so x_{LS} is a solution to the LS problem.

denote $x_{\mathrm{Null}} = \left\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}^T\mathbf{A}\mathbf{x} = 0\right\}$ so $x_{\mathrm{Null}} \in \mathcal{N}(\mathbf{A})$ and any $x \in \mathbf{x}_{\mathrm{LS}}$ can be represented by the addition of $\mathbf{x_L s}$ and $\mathbf{x}_{\mathrm{Null}}$

The 2 -norm in X_{LS} is equal to $\|x_{Null} - (-x_{LS})\|_2$. when x_{Null} is the vector which is x_{LS} projects onto the subspace $\mathcal{N}(A)$ the 2 -norm in X_{LS} is minimum. since the orthogonal projection onto $\mathcal{N}(\mathbf{A})$ is given by

$$\mathbf{\Pi}_{\mathcal{N}(A)} = \mathbf{I} - \mathbf{A}^\dagger\mathbf{A}$$

$$\mathbf{x}_{Null} = \left(\mathbf{I} - \mathbf{A}^\dagger\mathbf{A}\right)\mathbf{x_Ls} = \mathbf{x_Ls} - \mathbf{A}^\dagger\mathbf{A}\mathbf{x_Ls} = \mathbf{x_Ls} - \mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger\mathbf{y} = \mathbf{x_Ls} - \mathbf{A}^\dagger\mathbf{y} = 0$$

So $\mathbf{x}_{L.S} = \arg\min_{\mathbf{x} \in X_{La}} \|\mathbf{x}\|_2$

2) If $X_{LS} = \{x_{LS}\}$, the solution of equation $A^TAX = A^Ty$ is unique. So matrix A^TA is column full rank. since $\mathcal{R}(\mathbf{A}) = \mathcal{R}\left(\mathbf{A^TA}\right)$, we have rank $(\mathbf{A}) = n$ If $\text{rank}(\mathbf{A}) = n, \mathbf{A^TA}$ is invertible, the solution of equation $\mathbf{A}^T\mathbf{A}X = \mathbf{A}^Ty$ just is $\left(A^TA^{-1}\mathbf{A}^T\right)$, so the solution is unique, it means that $X_{LS} = \{x_{LS}\}$ Above all, $X_{LS} = \{x_{LS}\}$ if and only if rank $(\mathbf{A}) = n$