Unit 4: Linear Least Squares

Revisit the 4 fundamental subspaces

- The column space:
  $$C(A) = \{y \mid Ax = y\}$$
- The null space:
  $$N(A) = \{x \mid Ax = 0\}$$
- The row space:
  $$R(A) = \{y \mid y^H = x^H A\}$$
- The left null space:
  $$N(A^H) = \{x \mid x^H A = 0\}$$

Orthogonal subspaces:

Two subspaces $S, T \subset \mathbb{C}^n$ are orthogonal if any two arbitrary vectors $x \in S$, $y \in T$ are orthogonal:
$$x^H y = 0$$

4.2.1.1

$$R(A) \perp N(A)$$

4.2.1.2

$$C(A) \perp N(A^H)$$

4.2.1.3

These are equivalent statements for all $\{s_0, \ldots s_{r-1}\}$ be the basis for subspace $S \in \mathbb{C}^n$ and $\{t_0, \ldots t_{k-1}\}$ be the basis for subspace $T \in \mathbb{C}^n$.

1. Subspaces S and T are orthogonal
2. The vectors in $\{s_0, \ldots s_{r-1}\}$ are orthogonal to the vectors in $\{t_0, \ldots, t_{k-1}\}$
3. $s_i^H t_j = 0$ for all $0 \leq i < r$ and $0 \leq j < k$
4. $(s_0 | s_1 | \ldots | s_{r-1})^H (t_0 | t_1 | \ldots | t_{k-1}) = 0$, the zero matrix of appropriate size

4.2.1.4

Let $A \in \mathbb{C}^{m \times n}$, any $x \in \mathbb{C}^n$ can be written as:
$$x = x_r + x_n$$
where $x_r \in R(A)$ and $x_n \in N(A)$ and $x_r^H x_n = 0$

The method of normal equations

Given $A \in \mathbb{C}^{m \times n}$, $b \in \mathbb{C}^m$ find $\hat{x} \in \mathbb{C}^n$ such that
$$\| b - A\hat{x} \|_2 = \min_{x \in \mathbb{C}^n} \| b - Ax \|_2$$
We can solve this by finding $\hat{x}$ that satisfy:
$$A^H A \hat{x} = A^H b$$
If A has linearly independent columns: $\hat{x} = (A^H A)^{-1} A^H b$

4.2.2.2 (Left) pseudo inverse

Let $A \in \mathbb{C}^{m\times n}$ have linearly independent columns:

$$A^{\dagger} = (A^H A)^{-1} A^H$$

is it (left) pseudo inverse

Hw 4.2.2.1

Let $A \in \mathbb{C}^{m\times m}$ be nonsingular, then: $A^{-1} = A^{\dagger}$

Hw 4.2.2.2

Let $A \in \mathbb{C}^{m\times n}$ have linearly independent columns.

$AA^{\dagger} = I$ <SOMETIMES> only when A is square

Solving the normal equations:

To solve $A^H A \hat{x} = A^H b$, we can instead solve $B\hat{x} = y$ with $B = A^H A$ and $y = A^H b$

- Step 1: compute $B = A^H A$ (mn^2 flops)
- Step 2: compute $y = A^H b$ ($2mn$ flops)
- Step 3: compute Cholesky factorization $B = LL^H$ ($\frac{1}{3}n^3$ flops)
- Step 4: solve $Lz = y$ and $L^H \hat{x} = z$

Conditioning of linear least square

Given $A \in \mathbb{C}^{m\times n}$ with linearly independent columns and $b \in \mathbb{C}^m$, the condition of the linear least squares problem

$$\| b - A\hat{x} \|_2 = \min_x \| b - Ax \|_2$$

and the perturbed problem

$$\|(b + \delta b) - A(\hat{x} + \delta\hat{x})\|_2 = \min_x \|(b+\delta b) - A(x + \delta x)\|_2$$

is determined by:

$$\frac{\|\delta\hat{x}\|_2}{\|\hat{x}\|_2} \leq \frac{1}{\cos(\theta)} \frac{\sigma_0}{\sigma_{n-1}} \frac{\|\delta b\|_2}{\|b\|_2}$$

Condition number of matrix with linearly independent columns

$$\kappa_2(A) = \|A\|_2 \|A^{\dagger}\|_2 = \frac{\sigma_0}{\sigma_{n-1}}$$

Solving using normal equations could be bad

Hw 4.2.5.1

$$\kappa_2(A^H A) = [\kappa_2(A)]^2$$

SOLUTION VIA SVD

The SVD and the 4 fundamental spaces

Theorem 4.3.1.1

Given $A \in \mathbb{C}^{m \times n}$, $A = U_L \Sigma_{TL} V_L^H$ is reduced SVD and $A = (U_L \mid U_R) \left(\begin{array}{c|c} \Sigma_{TL} & 0 \\ \hline 0 & 0 \end{array} \right) (V_L \mid V_R)^H$ is SVD:

- $\mathcal{C}(A) = \mathcal{C}(U_L)$
- $\mathcal{N}(A) = \mathcal{C}(V_R)$
- $\mathcal{R}(A) = \mathcal{C}(A^H) = \mathcal{C}(V_L)$
- $\mathcal{N}(A^H) = \mathcal{C}(U_R)$

Hw 4.3.1.1

$\mathcal{R}(A) = \mathcal{C}(A^H) = \mathcal{C}(V_L)$

Hw 4.3.1.3

$r = \text{rank}(A) = \dim(\mathcal{C}(A)) = \dim(\mathcal{C}(U_L))$

$r =$ ~~rank~~ $\dim(\mathcal{R}(A)) = \dim(\mathcal{C}(V_L))$

$n - r = \dim(\mathcal{N}(A)) = \dim(\mathcal{C}(V_R))$

$m - r = \dim(\mathcal{N}(A^H)) = \dim(\mathcal{C}(U_R))$

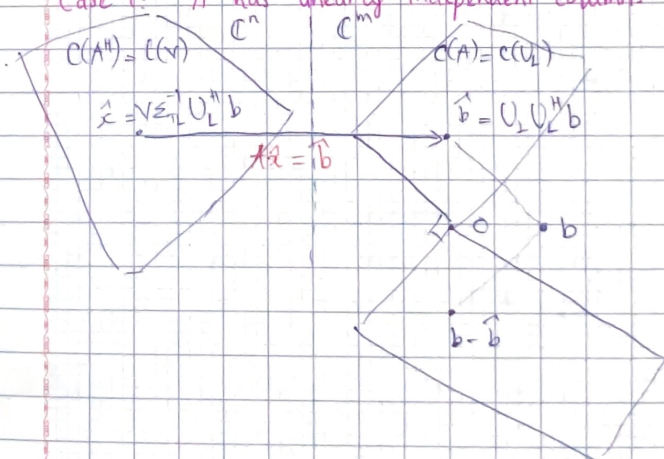
Hw 4.3.1.4

Any vector x can be written as:

$x = x_r + x_n$

with $x_r \in \mathcal{C}(V_L)$ and $x_n \in \mathcal{C}(V_R)$

Case 1: A has linearly independent columns



Cost of solving LLS problem via SVD

- Compute the reduced SVD: $A = U_L \Sigma_{TL} V^H$
  $O(mn^2)$ with a large constant
- Compute $\hat{x} = V \Sigma_{TL}^{-1} U_L^H b$
  The cost of this is approximately:

<continue>

- Form $y_T = U_L^H b$ (2mn flops)
- Scale the individual of y_T by dividing by the corresponding singular values (n flops). Overwriting $y_T = \Sigma_{TL}^{-1} y_T$. The cost of this step is negligible
- Compute $\hat{x} = V y_T$ ($2n^2$ flops)

Case 2: General case (A doesn't have linearly independent columns)

We try to solve $\| b - A\hat{x} \|_2 = \min_x \| b - Ax \|_2$ with no assumptions about the relative size of m and n. We essentially try to find $\hat{x}$ that minimise $\| b - Ax \|_2$ that satisfy:

$$A\hat{x} = \hat{b}$$

$$\Leftrightarrow (U_L | U_R) \left(\begin{array}{c|c} \Sigma_{TL} & 0 \\ \hline 0 & 0 \end{array} \right) (V_L | V_R)^H \hat{x} = U_L U_L^H \hat{b}$$

$$\Leftrightarrow \Sigma_{TL} V_L^H \hat{x} = U_L^H \hat{b}$$

$$\Leftrightarrow \Sigma_{TL} V_L^H (V_L z_T + V_R z_B) = U_L^H \hat{b}$$

$$\Leftrightarrow \Sigma_{TL} z_T = U_L^H b$$

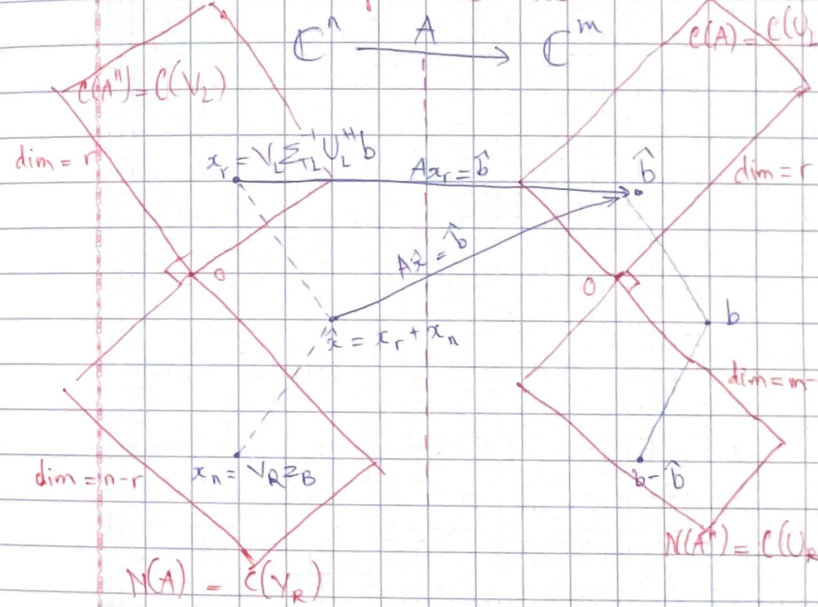
So, we can see that: $x_R = V_L z_T = V_L \Sigma_{TL}^{-1} U_L^H b$

And the general solution: $\hat{x} = V_L \Sigma_{TL}^{-1} U_L^H b + V_R z_B$

Giấy độ trắng tự nhiên không hại mắt

fahasa

Revisit the subspace cheat sheet



Thm 4.3.3.1

$\hat{x} = V_L \Sigma_{TL}^{-1} U_L^H b$ is the solution to the LLS problem. And x^* satisfies:

$$\| b - Ax^* \|_2 = \min_x \| b - Ax \|_2$$

Then:

$$\| \hat{x} \|_2 \leq \| x^* \|_2$$

Giấy độ trắng tự nhiên không hại mắt

fahasa

SOLUTION VIA QR FACTORIZATION

Case 1: A has linearly independent columns

Theorem 4.4.1.1

Assume $A \in \mathbb{C}^{m\times n}$ has linearly independent columns $A = QR$ with $Q \in \mathbb{C}^{m\times n}$ and upper triangular matrix $R \in \mathbb{C}^{n\times n}$. Then:

$$\|b - A\hat{x}\|_2 = \min_x \|b - Ax\|_2$$

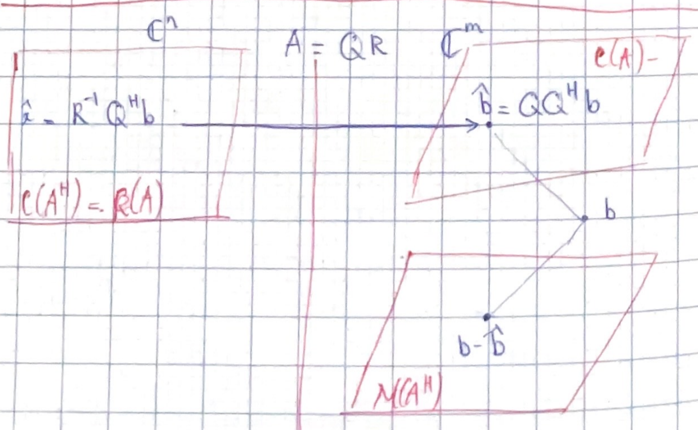
is solved by the unique solution of

$$R\hat{x} = Q^H b$$

Hw 4.4.1.1

Given $\hat{x} = (A^H A)^{-1} A^H b$ and $A = QR$ is the factorizati[on]

Then: $\hat{x} = R^{-1} Q^H b$



Giấy độ trắng tự nhiên không hại mắt

fahasa

Cost of solving via QR factorization

Given $A \in \mathbb{C}^{m\times n}$, the steps required to solve LLS problem via QR factorization are:

- Factoring $A = QR$ ($2mn^2$ flops)
- Compute $y = Q^H b$ ($2mn$ flops)
- Solve $R\hat{x} = y$ (n^2 flops)

Total: $2mn^2 + 2mn + n^2$ flops

Solve via Householder QR factorization

Intuition

We know that: $H_{n-1} \cdots H_1 H_0 A = R$

And to solve LLS problem $\|b - A\hat{x}\|_2 = \min_x \|b - Ax\|_2$

We are actually solving: $Ax \approx b$

Which can also be seen as:

$$H_{n-1} \cdots H_1 H_0 A x \approx H_{n-1} \cdots H_1 H_0 b$$

Formal definition

To solve: $\|b - A\hat{x}\|_2 = \min_x \|b - Ax\|_2$

We use: ~~$\tilde{y} := H_{n-1} \cdots H_1 H_0 y$~~ $\tilde{b} := \underbrace{H_{n-1} \cdots H_0}_{Q^H} b$

And then solve $\hat{x}$ by: $R_T \hat{x} = b_T$ with $b = \begin{pmatrix} b_T \\ b_B \end{pmatrix}$

Giấy độ trắng tự nhiên không hại mắt

fahasa

Cost of solving via Householder QR factorization

- Factoring $A = QR$ via Householder $(2mn^2 - \frac{2}{3}n^3)$
- Applying Q as a sequence of Householder $(4mn - 2n^2)$
- Solve $R_T \hat{x} = y_T$ (n^2)

Total: $2mn^2 - \frac{2}{3}n^3 + 4mn - 2n^2 + n^2$

$\approx 2mn^2 - \frac{2}{3}n^3$

Case 2: A has linearly dependent columns

Intuition:

Permute columns in A so that the first r columns are linearly independent and then solve $\hat{x}$ on the permuted matrix

Theorem 4.4.4.1

Assume $A \in \mathbb{C}^{m \times n}$ and $r = \text{rank}(A)$. Then there exists a permutation vector $p \in \mathbb{C}^n$, orthonormal matrix $Q_L \in \mathbb{C}^{m \times r}$, upper triangular matrix $R_{TL} \in \mathbb{C}^{r \times r}$ and $R_{TR} \in \mathbb{C}^{r \times (n-r)}$ such that

$$A P(p)^T = Q_L (R_{TL} \mid R_{TR})$$

Solving LLS with A linearly dependent via ~~Househo~~ QR factorization

Find $\hat{x}$ such that $\| b - A\hat{x} \|_2 = \min_x \| b - Ax \|_2$ when $\text{rank}(A) < n$:

- ~~Replace $AP(p)^T = Q_L(R_{TL} \mid R_{TR})$~~
- With $AP(p)^T = Q_L(R_{TL} \mid R_{TR})$ we can reach:

$$(R_{TL} \mid R_{TR}) P(p) \hat{x} = Q_L^H b$$

- Substitute $z = \begin{pmatrix} z_T \\ z_B \end{pmatrix} = P(p)\hat{x}$, we have:

$$(R_{TL} \mid R_{TR}) \begin{pmatrix} z_T \\ z_B \end{pmatrix} = Q_L^H b$$

$$\Leftrightarrow \quad R_{TL} z_T = Q_L^H b - R_{TR} z_B$$

- The solution $\hat{x}$ is given by:

$$\hat{x} = P(p)^T \begin{pmatrix} R_{TL}^{-1} (Q_L^H b - R_{TR} z_B) \\ \hline z_B \end{pmatrix}$$

Cost of solving via Householder QR factorization

- Factoring $A = QR$ via Householder $(2mn^2 - \frac{2}{3}n^3)$
- Applying Q as a sequence of Householder $(4mn - 2n^2)$
- Solve $R_T \hat{x} = y_T$ (n^2)

Total: $2mn^2 - \frac{2}{3}n^3 + 4mn - 2n^2 + n^2$

$\approx 2mn^2 - \frac{2}{3}n^3$

Case 2: A has linearly dependent columns

Intuition:

Permute columns in A so that the first r columns are linearly independent and then solve $\hat{x}$ on the permuted matrix

Theorem 4.4.4.1

Assume $A \in \mathbb{C}^{m\times n}$ and $r = \text{rank}(A)$. Then there exists a permutation vector $p \in \mathbb{C}^n$, orthonormal matrix $Q_L \in \mathbb{C}^{m\times r}$, upper triangular matrix $R_{TL} \in \mathbb{C}^{r\times r}$ and $R_{TR} \in \mathbb{C}^{r\times(n-r)}$ such that

$$AP(p)^T = Q_L(R_{TL} \mid R_{TR})$$

Solving LLS with A linearly dependent via ~~Househo~~ QR factorization

Find $\hat{x}$ such that $\| b - A\hat{x}\|_2 = \min_x \| b - Ax\|_2$ when $\text{rank}(A) < n$:

- ~~Replace $AP(p)^T = Q_L(R_{TL} \mid R_{TR})$~~
- With $AP(p)^T = Q_L(R_{TL} \mid R_{TR})$ we can reach:

$$(R_{TL} \mid R_{TR})\, P(p)\,\hat{x} = Q_L^H b$$

- Substitute $z = \begin{pmatrix} z_T \\ z_B \end{pmatrix} = P(p)\hat{x}$, we have:

$$(R_{TL} \mid R_{TR}) \begin{pmatrix} z_T \\ z_B \end{pmatrix} = Q_L^H b$$

$$\Leftrightarrow \quad R_{TL} z_T = Q_L^H b - R_{TR} z_B$$

- The solution $\hat{x}$ is given by:

$$\hat{x} = P(p)^T \begin{pmatrix} R_{TL}^{-1}(Q_L^H b - R_{TR} z_B) \\ \hline z_B \end{pmatrix}$$