ECE 133A – Homework #5 – Disc 1B – Angel Jimenez – 604 667 905

12.12 *Least squares placement.* The 2-vectors $p_1, \ldots, p_N$ represent the locations or positions of N objects, for example, factories, warehouses, and stores. The last K of these locations are fixed and given; the goal in a *placement problem* to choose the locations of the first $N-K$ objects. Our choice of the locations is guided by an undirected graph; an edge between two objects means we would like them to be close to each other. In *least squares placement*, we choose the locations $p_1, \ldots, p_{N-K}$ so as to minimize the sum of the squares of the distances between objects connected by an edge,

$$\|p_{i_1} - p_{j_1}\|^2 + \cdots + \|p_{i_L} - p_{j_L}\|^2,$$

where the L edges of the graph are given by $(i_1, j_1), \ldots, (i_L, j_L)$.

(a) Let $\mathcal{D}$ be the Dirichlet energy of the graph, as defined on page 135. Show that the sum of the squared distances between the N objects can be expressed as $\mathcal{D}(u)+\mathcal{D}(v)$, where $u = ((p_1)_1, \ldots, (p_N)_1)$ and $v = ((p_1)_2, \ldots, (p_N)_2)$ are N-vectors containing the first and second coordinates of the objects, respectively.

(b) Express the least squares placement problem as a least squares problem, with variable $x = (u_{1:(N-K)}, v_{1:(N-K)})$. In other words, express the objective above (the sum of squares of the distances across edges) as $\|Ax-b\|^2$, for an appropriate $m \times n$ matrix A and m-vector b. You will find that $m = 2L$. *Hint.* Recall that $\mathcal{D}(y) = \|B^T y\|^2$, where B is the incidence matrix of the graph.

(c) Solve the least squares placement problem for the specific problem with $N = 10$, $K = 4$, $L = 13$, fixed locations

$$p_7 = (0,0), \quad p_8 = (0,1), \quad p_8 = (1,1), \quad p_{10} = (1,0),$$

and edges

$$(1,3), \quad (1,4), \quad (1,7), \quad (2,3), \quad (2,5), \quad (2,8), \quad (2,9),$$

$$(3,4), \quad (3,5), \quad (4,6), \quad (5,6), \quad (6,9), \quad (6,10).$$

Plot the locations, showing the graph edges as lines connecting the locations.

$$D(v) = \|A^T v\|^2 = \sum_{\text{edges}(k,\ell)} (v_\ell - v_k)^2$$

v = potential differences across all edges

a) Show

$D(u) + D(v)$

$u = ((P_1)_1, \ldots, (P_n)_1)$
$v = ((P_1)_2, \ldots, (P_n)_2)$ } n vectors

$D(u) + D(v)$

$$= \sum (u_\ell - u_k)^2 + \sum (v_\ell - v_k)^2$$

$$= \sum \left\| \begin{bmatrix} u_\ell \\ v_\ell \end{bmatrix} - \begin{bmatrix} u_k \\ v_k \end{bmatrix} \right\|^2$$

$D(u) + D(v)$ is the sum of squares of distances, sum is over the L edges (k, ℓ)

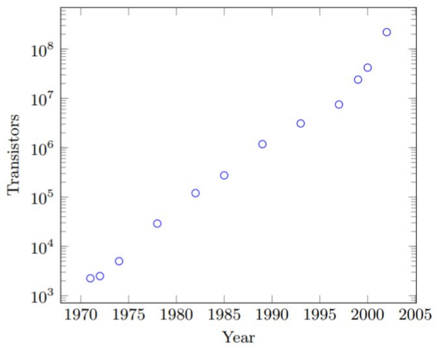
b) $X = \begin{bmatrix} u_{1:(N-K)} \\ v_{1:(N-K)} \end{bmatrix}$

A = contains $-1, 0,$ or 1

Use normal equations?

13.3 *Moore's law.* The figure and table below show the number of transistors N in 13 microprocessors, and the year of their introduction.

	Year	Transistors
1	1971	2,250
2	1972	2,500
3	1974	5,000
4	1978	29,000
5	1982	120,000
6	1985	275,000
7	1989	1,180,000
8	1993	3,100,000
9	1997	7,500,000
10	1999	24,000,000
11	2000	42,000,000
12	2002	220,000,000
13	2003	410,000,000



The plot gives the number of transistors on a logarithmic scale. Find the least squares straight-line fit of the data using the model

$$\log_{10} N \approx \theta_1 + \theta_2(t - 1970),$$

where t is the year and N is the number of transistors. Note that θ_1 is the model's prediction of the log of the number of transistors in 1970, and 10^{θ_2} gives the model's prediction of the fractional increase in number of transistors per year.

(a) Find the coefficients θ_1 and θ_2 that minimize the RMS error on the data, and give the RMS error on the data. Plot the model you find along with the data points.

(b) Use your model to predict the number of transistors in a microprocessor introduced in 2015. Compare the prediction to the IBM Z13 microprocessor, released in 2015, which has around 4×10^9 transistors.

(c) Compare your result with Moore's law, which states that the number of transistors per integrated circuit roughly doubles every one and a half to two years.

The computer scientist and Intel corporation co-founder Gordon Moore formulated the law that bears his name in a magazine article published in 1965.

a) Example from notes

$$f(x) = \|Ax - b\|^2$$

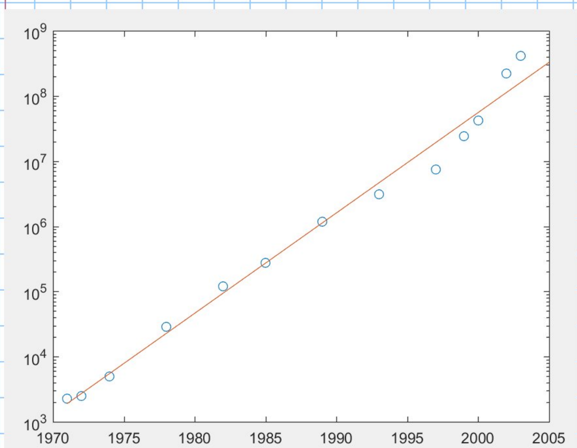
$$x = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \quad \langle 2\times 1 \rangle \qquad A = \begin{bmatrix} 1 & t_1 - 1970 \\ 1 & t_2 - 1970 \\ \vdots & \vdots \\ 1 & t_{13} - 1970 \end{bmatrix}$$

$$b = \begin{bmatrix} \log_{10} N_1 \\ \log_{10} N_2 \\ \vdots \\ \log_{10} N_{13} \end{bmatrix}$$

```
% Homework 5, Problem 2
[t,n] = mooreslaw;
A = [t,n];
B = A(:,2);
C = log10(B);
D = [ones(13,1), A(:,1)-1970];
x = D\C;
disp(x);
E = rms(D*x-C);
disp(E);
```

$$\text{rms} = \frac{\|A\hat{\theta} - b\|}{\sqrt{N}} = 0.2031$$

$$\hat{\theta}_1 = 3.1256$$
$$\hat{\theta}_2 = 0.1540$$




```
% Homework 5, Problem 2
[t,n] = mooreslaw;
A = [t,n];
B = A(:,2);
C = log10(B);
D = [ones(13,1), A(:,1)-1970];
x = D\C;
disp(x);
E = rms(D*x-C);
disp(E);

t1 = 1971;
t2 = 2005;
theta1 = x(1,1);
theta2 = x(2,1);
semilogy(t,n, 'o');
hold on;
semilogy([t1;t2], ...
[ 10^(theta1 + theta2*(t1-1970)); ...
    10^(theta1 + theta2*(t2-1970))], '-');
```

b) Prediction for 2015

$$\log_{10}(N_{15}) \approx \hat{\theta}_1 + \hat{\theta}_2(t_{15} - 1970)$$

Solve for N_{15} $\rightarrow$ $N_{15} \approx 10^{\hat{\theta}_1 + \hat{\theta}_2(t_{15} - 1970)}$

$$= 10^{3.1256 + 0.1540(2015 - 1970)}$$

$$\approx 1.1365799 \times 10^{10}$$

% error = 64% off

getting this code to work was harder than I thought

c) θ_1 – models prediction of log

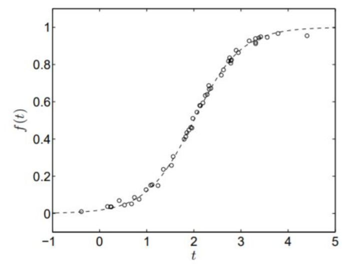
10^{θ_2} gives model's prediction of fractional increase

$$\frac{\log_{10}(2)}{\theta_2} = \frac{\log_{10}(2)}{0.1540} \approx 1.9547 \approx 2, \text{ so its close.}$$

8.3 The figure shows $m = 50$ points (t_i, y_i) as circles. These points are well approximated by a function of the form

$$f(t) = \frac{e^{\alpha t+\beta}}{1+e^{\alpha t+\beta}}.$$

(An example, for two specific values of α and β, is shown in dashed line).



Formulate the following problem as a least squares problem. Find values of the parameters α, β such that

$$\frac{e^{\alpha t_i+\beta}}{1+e^{\alpha t_i+\beta}} \approx y_i, \qquad i = 1, \ldots, m, \tag{15}$$

You can assume that $0 < y_i < 1$ for $i = 1, \ldots, m$.

Clearly state the error function you choose to measure the quality of the fit in (15), and the matrix A and the vector b of the least squares problem. Test your method on the example data in the file `logistic_fit.m`. (The command `[t, y] = logistic_fit;` creates arrays with the points t_i, y_i.)

Hint: $\alpha t_i + \beta \approx h^{-1}(y_i)$

h^{-1} is inverse of $h(x) = \dfrac{e^x}{1+e^x}$

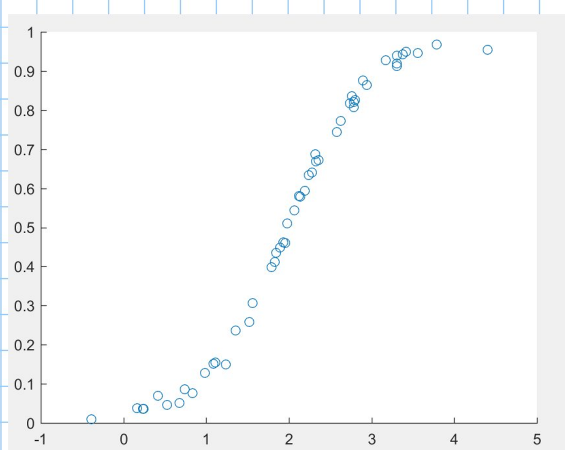
$$y = \frac{e^x}{1+e^x} \longrightarrow h^{-1}(y) = x = \log\left(\frac{y}{1-y}\right) \Rightarrow \alpha t_i + \beta \approx h^{-1}(y_i) = \log\left(\frac{y_i}{1-y_i}\right)$$

$$\hookrightarrow \sum_{i=1}^{m} \left[\alpha t_i + \beta - \log\left(\frac{y_i}{1-y_i}\right)\right]^2$$

$$\therefore \; x = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \qquad A = \begin{bmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_m & 1 \end{bmatrix} \qquad b = \begin{bmatrix} \log(y_1/1-y_1 \\ \log(y_2/1-y_2) \\ \vdots \\ \log(y_m/1-y_n) \end{bmatrix}$$

$$y = \frac{e^x}{1+e^x} \Rightarrow x = \log\left(\frac{y}{1-y}\right)$$

$$\alpha t_i + \beta \approx \log\left(\frac{y_i}{1-y_i}\right) \qquad i = 1, \ldots, m$$



8.8 Let A be an $m \times n$ matrix with linearly independent columns.

(a) Show that the $(m+n) \times (m+n)$ matrix

$$\overset{m+n}{\begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix}}$$

is nonsingular.

(b) Show that the solution $\bar{x}$, $\bar{y}$ of the set of linear equations

$$\begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix}$$

is given by $\bar{x} = b - Ax_{ls}$ and $\bar{y} = x_{ls}$, where x_{ls} is the solution of the least squares problem

$$\text{minimize} \quad \|Ax - b\|^2.$$

$I = \langle m \times m \rangle$

$A = \langle m \times n \rangle$

$0 = \langle n + n \rangle$

$A^T = \langle n \times m \rangle$

A has lin. indep columns

a) to show nonsingular

$$Ax = 0 \Rightarrow \begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0 \rightsquigarrow \begin{matrix} Ix + Ay = 0 \\ A^T x + 0y = 0 \end{matrix} \rightarrow \begin{matrix} x + Ay = 0 \\ A^T x = 0 \end{matrix}$$

$$\hookrightarrow \begin{matrix} x + Ay = 0 \\ A^T x = 0 \end{matrix} \rightsquigarrow -Ay = x \Rightarrow -A^T A y = A^T x \quad \therefore \quad -A^T A y = 0$$

$A^T A$, A has lin. ind. Columns, $\therefore$ $A^T A$ is non singular.

$\therefore$ $y = 0 \rightsquigarrow x = 0$

$\therefore$ $\begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix}$ is also non singular

b) $$\begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix} \Rightarrow \begin{matrix} I\bar{x} + A\bar{y} = b \\ A^T \bar{x} + 0\bar{y} = 0 \end{matrix} \begin{matrix} \rightarrow \\ \rightarrow \end{matrix} \begin{matrix} \bar{x} + A\bar{y} = b \\ A^T \bar{x} = 0 \end{matrix}$$

Solve $A^T x$ first, $A^T(b - Ax_{ls}) = 0 \Rightarrow A^T b - A^T A x_{ls} = 0$

$\hookrightarrow A^T b = A^T A x_{ls}$

$\bar{x} = b - Ax_{ls}$ & $\bar{y} = x_{ls}$

$$\begin{aligned} A^T \bar{x} &= A^T(b - A(A^T A)^{-1} A^T b) \\ &= A^T b - (A^T A)(A^T A)^{-1} A^T b \\ &= A^T b - A^T b = 0 \end{aligned}$$

8.11 *Solving normal equations versus QR factorization.* In this problem we compare the accuracy of the two methods for solving a least squares problem

$$\text{minimize} \quad \|Ax - b\|^2.$$

We take

$$A = \begin{bmatrix} 1 & 1 \\ 10^{-k} & 0 \\ 0 & 10^{-k} \end{bmatrix}, \qquad b = \begin{bmatrix} -10^{-k} \\ 1+10^{-k} \\ 1-10^{-k} \end{bmatrix},$$

for $k = 6$, $k = 7$ and $k = 8$.

(a) Write the normal equations, and solve them analytically (*i.e.*, on paper, without using MATLAB).

(b) Solve the least squares problem in MATLAB, for $k = 6$, $k = 7$ and $k = 8$, using the recommended method `x = A \ b`. This method is based on the QR factorization.

(c) Repeat part (b), using `x = (A'*A) \ (A'*b)`. Compare the results of this method with the results of parts (a) and (b).

Remark. Type `format long` to make MATLAB display more than five digits.

normal Equation

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

a) $A^T A \hat{x} = A^T b$

$$\begin{bmatrix} 1 & 10^{-k} & 0 \\ 1 & 0 & 10^{-k} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 10^{-k} & 0 \\ 0 & 10^{-k} \end{bmatrix} \hat{x} = \begin{bmatrix} 1 & 10^{-k} & 0 \\ 1 & 0 & 10^{-k} \end{bmatrix} \begin{bmatrix} -10^{-k} \\ 1+10^{-k} \\ 1-10^{-k} \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 1\times 1 + 10^{-k}\times 10^{-k} + 0\times 0 & 1\times 1 + 10^{-k}\times 0 + 0\times 10^{-k} \\ 1\times 1 + 0\times 10^{-k} + 10^{-k}\times 0 & 1\times 1 + 0\times 0 + 10^{-k}\times 10^{-k} \end{bmatrix}$$

$$= \begin{bmatrix} 1+10^{-2k} & 1 \\ 1 & 1+10^{-2k} \end{bmatrix} \hat{x} = \begin{bmatrix} 10^{-2k} \\ -10^{-2k} \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 1\times -10^{-k} + 10^{-k}\times(1+10^{-k}) + 0\times(1-10^{-k}) \\ 1\times -10^{-k} + 0\times(1+10^{-k}) + 10^{-k}\times(1-10^{-k}) \end{bmatrix}$$

$$= \begin{bmatrix} -10^{-k} + 10^{-k} + 10^{-2k} \\ -10^{-k} + 0 + 10^{-k} - 10^{-2k} \end{bmatrix}$$

$$= \begin{bmatrix} 10^{-2k} \\ -10^{-2k} \end{bmatrix}$$

$$x_1(1+10^{-2k}) + x_2(1) = 10^{-2k}$$
$$x_1(1) + x_2(1+10^{-2k}) = -10^{-2k}$$

$$\Rightarrow \begin{aligned} x_1 + x_1(10^{-2k}) + x_2 &= 10^{-2k} \\ x_1 + x_2 + x_2 10^{-2k} &= -10^{-2k} \end{aligned} \Bigg\} \Rightarrow \begin{aligned} x_1 + x_2 &= 10^{-2k} - x_1 10^{-2k} \\ x_1 + x_2 &= -10^{-2k} - x_2 10^{-2k} \end{aligned}$$

$$10^{-2k} - x_1 10^{-2k} = -10^{-2k} - x_2 10^{-2k} \Rightarrow 1 - x_1 = -1 - x_2 \Rightarrow 2 = -x_2 + x_1 \quad \therefore \boxed{x_1 = 1,\ x_2 = -1}$$

b)

```
1   % Homework #5, Problem A8.11
2
3   k = 6;
4
5   A = [1 1; 10^(-k) 0; 0 10^(-k)];
6   b = [-10^(-k); 1+10^(-k); 1-10^(-k)];
7
8   format long;
9   x = A\b;
10
11  disp('the value of k is: ');disp(k);
12  disp('corresponding values of x: ');
13  disp(x);
14
```

```
>> Homework5ProblemA811
the value of k is:
     6

corresponding values of x:
   0.999999999860908
  -0.999999999860908
```

```
>> Homework5ProblemA811
the value of k is:
     7

corresponding values of x:
   1.000000001838782
  -1.000000001838781
```

```
>> Homework5ProblemA811
the value of k is:
     8

corresponding values of x:
   1.000000006592782
  -1.000000006592782
```

c)

```
1   % Homework #5, Problem A8.11
2   clear all;
3   k = 6;
4
5   A = [1 1; 10^(-k) 0; 0 10^(-k)];
6   b = [-10^(-k); 1+10^(-k); 1-10^(-k)];
7
8   format long;
9   x = (A'*A) \ (A'*b);
10
11  disp('the value of k is: ');disp(k);
12  disp('corresponding values of x: ');
13  disp(x);
14
```

```
>> Homework5ProblemA811
the value of k is:
     6

corresponding values of x:
   0.999911107241501
  -0.999911107241501
```

```
>> Homework5ProblemA811
the value of k is:
     7

corresponding values of x:
   1.012044860405864
  -1.012044860405863
```

```
Warning: Matrix is singular to working precision.
> In Homework5ProblemA811 (line 9)
the value of k is:
     8

corresponding values of x:
   Inf
  -Inf
```

what does this mean?

8.12 *Least squares updating.* Suppose $\hat{x}$ is the solution of the least squares problem

$$\text{minimize} \quad \|Ax - b\|^2$$

where A is an $m \times n$ matrix with linearly independent columns and b is an m-vector.

(a) Show that the solution of the problem

$$\text{minimize} \quad \|Ay - b\|^2 + (c^T y - d)^2$$

with variable y (where c is an n-vector, and d is a scalar) is given by

$$\hat{y} = \hat{x} + \frac{d - c^T\hat{x}}{1 + c^T(A^TA)^{-1}c}(A^TA)^{-1}c.$$

(b) Describe an efficient method for computing $\hat{x}$ and $\hat{y}$, given A, b, c and d, using the QR factorization of A. Clearly describe the different steps in your algorithm. Give the complexity of each step and the overall complexity. In your total flop count, include all terms that are cubic (n^3, mn^2, m^2n, m^3) and quadratic (m^2, mn, n^2). If you know several methods, give the most efficient one.

a) minimize $\left\| \begin{bmatrix} A \\ c^T \end{bmatrix} y - \begin{bmatrix} b \\ d \end{bmatrix} \right\|^2 = \|Ay - b\|^2 + (c^Ty - d)^2$

↳ Normal Equations form: $(A^TA + cc^T)\hat{y} = A^Tb + dc$

$$\begin{bmatrix} A \\ c^T \end{bmatrix}^T \begin{bmatrix} A \\ c^T \end{bmatrix} \hat{y} = \begin{bmatrix} A \\ c^T \end{bmatrix}^T \begin{bmatrix} b \\ d \end{bmatrix}$$

↳ $(A^TA + cc^T)\hat{x} = A^Tb + (c^T\hat{x})c \Rightarrow (A^TA + cc^T)(A^TA)^{-1}c = c + (c^T(A^TA)^{-1}c)c$

$$\therefore (A^TA + cc^T)\hat{y} = (A^TA + cc^T)\left(\hat{x} + \frac{d - c^T\hat{x}}{1 + c^T(A^TA)^{-1}c}(A^TA)^{-1}c\right)$$

$$= A^Tb + (c^T\hat{x})c + \frac{(d - c^T\hat{x})}{1 + c^T(A^TA)^{-1}c}(1 + c^T(A^TA)^{-1}c)c$$

$$= A^Tb + (c^T\hat{x})c + (d - c^T\hat{x})c$$

$$= A^Tb + dc$$

b) $\hat{x} = R^{-1}Q^Tb$, $\quad \hat{y} = \hat{x} + \dfrac{(d - c^Tx)R^{-1}R^{-T}c}{1 + c^TR^{-1}R^{-T}c}$

- QR factorization $\quad 2mn^2$ flops
- $u = Q^Tb \quad 2mn$ flops
- $\hat{x} = R^{-1}u \quad$ B.W sub $\quad n^2$ flops
- $w = R^{-1}R^{-T}c \quad$ B.W & F.W. sub: $\quad 2n^2$ flops
- $\alpha = \dfrac{(d - c^T\hat{x})}{(1 + c^Tw)} \quad$ ~ $4n$ flops
- $\hat{y} = \hat{x} + \alpha w \quad$ ~ $2n$ flops

$\therefore$ Total flops $\simeq 2mn^2$ flops

8.15 Let A be an $m \times n$ matrix with linearly independent columns, and b an m-vector not in the range of A.

(a) Explain why the QR factorization

$$\left[\begin{array}{cc} A & b \end{array}\right] = QR$$

exists.

(b) Suppose we partition the matrices in the QR factorization of part (a) as

$$Q = \left[\begin{array}{cc} Q_1 & Q_2 \end{array}\right], \qquad R = \left[\begin{array}{cc} R_{11} & R_{12} \\ 0 & R_{22} \end{array}\right],$$

where Q_1 is $m \times n$, Q_2 is $m \times 1$, R_{11} is $n \times n$, R_{12} is $n \times 1$ and R_{22} is a scalar. Show that $x_{\rm ls} = R_{11}^{-1} R_{12}$ is the solution of the least squares problem

$$\text{minimize} \quad \|Ax - b\|^2$$

and that $R_{22} = \|Ax_{\rm ls} - b\|$.

A: $\langle m \times n \rangle$ lin. ind. columns

b: $\langle m \times 1 \rangle$ not in range of A

$\therefore$ b is not a column vector in A

a) $[A \ b] \begin{bmatrix} x \\ y \end{bmatrix} = 0 \Rightarrow Ax + by = 0 \quad \therefore \ x = 0 = y$

if y wasn't zero, $b = \frac{-Ax}{y}$ which means b is in the range of A, contradicting the problem statement. $x = 0$ b/c A has lin. indep columns.

b) $A = Q_1 R_{11} \qquad b = Q_1 R_{12} + Q_2 R_{22}$

Q is orthogonal

ie. $Q_1^T Q_1 = I$

$Q_1^T Q_2 = 0$

R is invertible

$X_{ls} = (A^T A)^{-1} A^T b$

$$\begin{aligned} X &= (A^T A)^{-1} A^T b \\ &= (R_{11}^T Q_1^T Q_1 R_{11})^{-1} R_{11}^T Q_1^T (Q_1 R_{12} + Q_2 R_{22}) \\ &= (R_{11}^T R_{11})^{-1} R_{11}^T R_{12} \\ &= R_{11}^{-1} R_{12} \end{aligned}$$

$$\begin{aligned} \therefore \|A R_{11}^{-1} R_{12} - b\| &= \|Q_1 R_{12} - Q_1 R_{12} + Q_2 R_{22}\| \\ &= \|Q_2 R_{22}\| \\ &= R_{22} \end{aligned}$$