Assignment 3

George Tang

February 2023

1 Question 1

For this question we assume $A \in \mathbb{R}^{m\times n}$ with $m \geq n$. We can apply SVD to A and we obtain $A = U\Sigma V^T$, where $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$, and

$$\Sigma = \begin{bmatrix} \tilde{\Sigma} \\ 0 \end{bmatrix}$$

with $\tilde{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r, 0, \ldots, 0)$, and r the rank of the matrix A. We define the pseudo-inverse of A as

$$A^\dagger := V \begin{bmatrix} \tilde{\Sigma}^{-1} & 0 \end{bmatrix} U^T.$$

where $\tilde{\Sigma}^{-1} \in \mathbb{R}^{n\times n}$

1.1 $AA^\dagger A = A$

Firstly let's look at: $A^\dagger A$: Firstly note:

$$\begin{aligned}
&\Sigma^\dagger\Sigma && (1)\\
=&\begin{bmatrix} \tilde{\Sigma}^{-1} & 0 \end{bmatrix} \begin{bmatrix} \tilde{\Sigma} \\ 0 \end{bmatrix} && (2)\\
=&\tilde{\Sigma}^{-1}\tilde{\Sigma} && (3)\\
=&I_{n\times n} && (4)
\end{aligned}$$

$$\begin{aligned}
&A^\dagger A \\
=&V\Sigma^\dagger U^T U\Sigma V^T \\
=&V\Sigma^\dagger\Sigma V^T && \text{U is an orthogonal matrix}\\
=&VV^T && \text{by (1)}\\
=&I_{n\times n} && \text{V is an orthogonal matrix}
\end{aligned}$$

Remark: I in (4) is not necessary the identity matrix but with r 1s on the diagonal and n-r 0s on the diagonal. However, as we set $r = n$, then I will be the identity matrix in this question. Therefore, It doesn't effect the result $VV^T = I$

Therefore

$$\begin{aligned}
AA^\dagger A =&A \\
=&AI_{n\times n} \\
=&A
\end{aligned}$$

1.2 $(AA^\dagger)^T = AA^\dagger$

Firstly note:

$$\begin{aligned}
&\Sigma\Sigma^\dagger && (5)\\
=&\begin{bmatrix} \tilde{\Sigma} \\ 0 \end{bmatrix} \begin{bmatrix} \tilde{\Sigma}^{-1} & 0 \end{bmatrix} && (6)\\
=&\begin{pmatrix} I_{n\times n} & 0 \\ 0 & 0 \end{pmatrix} && \text{This is a } m\times m \text{ matrix} \quad (7)\\
=&\begin{pmatrix} I_{n\times n} & 0 \\ 0 & 0 \end{pmatrix}^T && (8)
\end{aligned}$$

$$
\begin{aligned}
&(AA^\dagger)^T \\
=&(U\Sigma V^T V\Sigma^\dagger U^T)^T \\
=&(U\Sigma\Sigma^\dagger U^T)^T \\
=&U\Sigma\Sigma^\dagger U^T \\
=&U\Sigma V^T V\Sigma^\dagger U^T \\
=&AA^\dagger
\end{aligned}
$$

1.3 $A^\dagger AA^\dagger = A^\dagger$

We know from 1.1 that $A^\dagger A = I_{n\times n}$
Therefore, the equality holds.

1.4 $(A^\dagger A)^T = (A^\dagger A)$

We know from 1.1 that $A^\dagger A = I_{n\times n}$
And $I_{n\times n}^T = I \implies LHS = RHS$

2 Question 2

From assignment 2, we obtain the following result: $Hx_0 = \alpha\|x_0\|_2 e_1$. That means given a vector x_0, we can try to find a $\tilde{v}$ such that it reflects the vector x_0 to a scalar multiple of e_1.
Here is the outline of how QR factorisation can be done to a matrix A of size $m \times n$:

1. Firstly, find a matrix H_1 such that it sends the first column of A to the form $\begin{pmatrix} \alpha\|x_0\|_2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$
2. Find a matrix H_2 such that it sends the first column of $A(m-1\times n-1)$ to the same form.
3. Continue this procedure until the matrix A is upper triangular.

The matrix H which construct this QR decomposition is like this form: $\begin{pmatrix} 1 & & \\ & 1 & \\ & & H_3 \end{pmatrix} \begin{pmatrix} 1 & \\ & H_2 \end{pmatrix} \begin{pmatrix} H_1 \end{pmatrix}$ We know H will be orthogonal because H_1, H_2, H_3 they are all orthogonal which we already proved in assignment 2. We can get the upper triangular matrix R by compute HA as describe above and will obtain the following form: $\begin{pmatrix} \alpha\|x_0\|_2 & a_{12} & a_{13} \\ 0 & \alpha\|x_1\|_2 & a_23 \\ 0 & 0 & \alpha\|x_2\|_2 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{pmatrix}$

3 Question 3

3.1 Part a

From the linear system given in the question, we can produce two simultaneous linear equation:

$$
\begin{cases} r + A\hat{x} = b \\ A^T r = 0 \end{cases} \tag{9}
$$

$$
\begin{cases} r = b - A\hat{x} \\ A^T r = 0 \end{cases} \tag{10}
$$

Substitute the first equation to the second one for r:

$$
A^T b - A^T A\hat{x} = 0 \tag{11}
$$

Equation (11) is the normal equation by definition, which is used to solve the least square problem. (i.e. by solving (11) we can get the minimum $\hat{x}$.)

3.2 Part 2

Firstly notice:

$$
\begin{aligned}
&A^T A && (12)\\
=&V\Sigma^T U^T U \Sigma V^T && (13)\\
=&V\Sigma^T \Sigma V^T \qquad\qquad \Sigma \in \mathbb{R}^{m\times n} && (14)\\
=&V\tilde{\Sigma}^2 V^T \qquad\qquad \tilde{\Sigma} \in \mathbb{R}^{n\times n} && (15)
\end{aligned}
$$

Remark: The reason why (15) is true is because of the following:

$$
\begin{aligned}
&\Sigma^T \Sigma\\
=&\begin{pmatrix}\tilde{\Sigma} & 0\end{pmatrix}\begin{pmatrix}\tilde{\Sigma}\\ 0\end{pmatrix}\\
=&\tilde{\Sigma}^2
\end{aligned}
$$

$$
\begin{aligned}
A^T A\hat{x} &= A^T b\\
\hat{x} &= \left(A^T A\right)^{-1} A^T b\\
\hat{x} &= V(\tilde{\Sigma}^2)^{-1} V^T A^T b && (V\tilde{\Sigma}^2 V^T)^{-1} = V\tilde{\Sigma}^2 V^T\\
\|\hat{x}\|_2 &= \|V(\tilde{\Sigma}^2)^{-1} V^T A^T b\|_2\\
\|\hat{x}\|_2 &\leq \|V(\tilde{\Sigma}^2)^{-1} V^T A^T\|_2 \|b\|_2 && \text{by submultiplicativity}\\
\|\hat{x}\|_2 &\leq \|(\tilde{\Sigma}^2)^{-1} A^T\|_2 \|b\|_2 && \text{orthogonal matrix doesn't effect the size of 2-norm}\\
\|\hat{x}\|_2 &\leq \sigma_n^{-1}\|b\|_2 && \text{by definition of 2-norm where the biggest is } \sigma_n^{-1}
\end{aligned}
$$

4 Question 4

4.1 Part a

By the question given, we know we have the least square problem:

$$\min_{\overline{a}\in\mathbb{R}^{m+1}} \sum_{i=0}^{n} |\sum_{j=0}^{m} a_j x_i^j - y_i|^2$$

Therefore, we have it in the matrix form:

$$\min_{\overline{a}\in\mathbb{R}^{m+1}} \left\| \begin{pmatrix} 1 & x_0 & x_0^2 & x_0^3 & \cdots & x_1^m\\ 1 & x_1 & x_1^2 & x_1^3 & \cdots & x_2^m\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & x_n & x_n^2 & x_n^3 & \cdots & x_n^m \end{pmatrix} \begin{pmatrix} a_0\\ a_1\\ \vdots\\ a_m \end{pmatrix} - \begin{pmatrix} y_0\\ y_1\\ \vdots\\ y_n \end{pmatrix} \right\|_2 \tag{16}$$

The associated normal equation to (16) is $A^T A\overline{a} = A^T b$

4.2 Part b

Here is the result I get for $\overline{a}$, which is the vectors that satisfy $min_{\overline{a}\in\mathbb{R}^{m+1}}\|A\overline{a} - b\|_2$ for different m,n.

```
array([[ 1.00000000e+00],
       [ 3.48610030e-13],
       [-1.68552036e+01],
       [-5.08748599e-12],
       [ 1.23359729e+02],
       [ 2.21973551e-11],
       [-3.81433824e+02],
       [-3.47313289e-11],
       [ 4.94909502e+02],
       [ 1.71738179e-11],
       [-2.20941742e+02]])
```

Figure 1: m=10,n=10

```
array([[ 9.86177440e-01],
       [-1.82879094e-11],
       [-1.99405121e+01],
       [ 1.00971231e-10],
       [ 2.59603058e+02],
       [ 1.15808518e-09],
       [-2.03527067e+03],
       [-2.16724629e-08],
       [ 9.81576298e+03],
       [ 1.24588723e-07],
       [-3.00269756e+04],
       [-3.63344952e-07],
       [ 5.92445359e+04],
       [ 5.96817699e-07],
       [-7.50393118e+04],
       [-5.56938176e-07],
       [ 5.88646706e+04],
       [ 2.74591912e-07],
       [-2.60104106e+04],
       [-5.53035306e-08],
       [ 4.94638969e+03]])
```

Figure 2: n=80,m=20

Here is the plot of the least square problem:

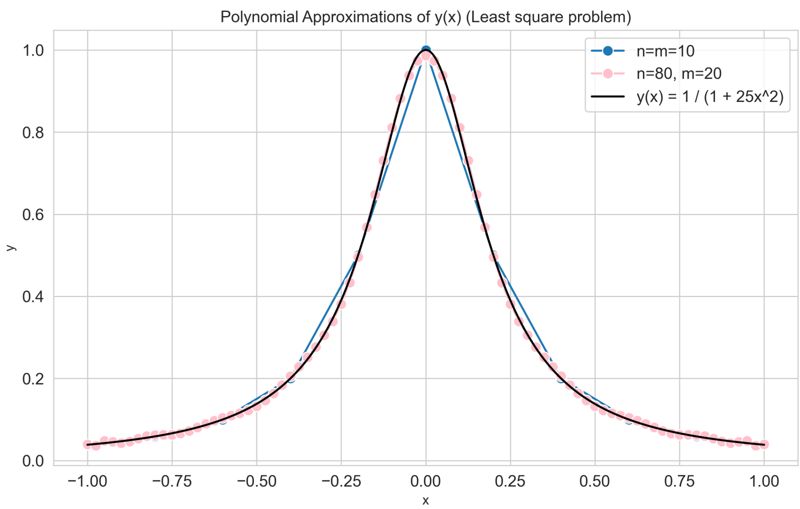

Figure 3: Least square