MA2501 - Assignment 2

Andreas B. Berg

28.02.2022

Problem 1

Part 1

Show that the function $f(x) = (x+1)(x-1)/3$ has a unique fixed point in the interval $[-1, 1]$. What can you say about the interval $[3, 4]$?

The definition of a fixed point is a point where $f(x) = x$ for a function f. This is equivalent with $f(x) - x = 0$. Observe that

$$\begin{aligned} f(x) - x &= \frac{(x+1)(x-1)}{3} - x \\ &= \frac{x^2 - 1}{3} - x \\ &= \frac{x^2 - 3x - 1}{3} \end{aligned}$$

Let $g(x) = f(x) - x$. Note that

$$\begin{aligned} g(-1) &= \frac{1 + 3 - 1}{3} > 0 \\ g(1) &= \frac{1 - 3 - 1}{3} < 0 \end{aligned}$$

As f (and therefore also g) is continuous, $g(-1) > 0$ and $g(1) < 0$, we have from the Intermediate Value Theorem that g has at least one zero on the interval $[-1, 1]$. This means that $f(x) = x$ for at least one $x \in [-1, 1]$, meaning that f has a unique fixed point in the interval.

For the interval $[3, 4]$, observe that

$$\begin{aligned} g(3) &= \frac{9 - 9 - 1}{3} < 0 \\ g(4) &= \frac{16 - 12 - 1}{3} > 0 \end{aligned}$$

g is still continuous, meaning that the Intermediate Value Theorem gives us the same result here - g has at least one zero on the interval $[3, 4]$, meaning that f has a unique fixed point in the interval.

Part 2

Compute the spectral radius of the matrices

$$T_1 = \begin{pmatrix} 0 & 1/2 & -1/2 \\ -1 & 0 & -1 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

$$T_2 = \begin{pmatrix} 0 & 1/2 & -1/2 \\ 0 & -1/2 & -1/2 \\ 0 & 0 & -1/2 \end{pmatrix}$$

The spectral radius is defined as the largest absolute value of its eigenvalues. We therefore start by finding the eigenvalues of the matrices.

$$\begin{aligned} |T_1 - \lambda I| &= \begin{vmatrix} -\lambda & 1/2 & -1/2 \\ -1 & -\lambda & -1 \\ 1/2 & 1/2 & -\lambda \end{vmatrix} \\ &= -\lambda(\lambda^2 + 1/2) - 1/2(\lambda + 1/2) - 1/2(-1/2 + \lambda/2) \\ &= -\lambda^3 - \lambda/2 - \lambda/2 - 1/4 + 1/4 - \lambda/4 \\ &= -\lambda^3 - \frac{5}{4}\lambda \\ &= -\lambda(\lambda^2 + \frac{5}{4}) \end{aligned}$$

The roots of the characteristic polynomal, and thus the eigenvalues of T_1 are 0 and $\pm i\sqrt{\frac{5}{4}}$. The spectral radius of T_1 is given as

$$\begin{aligned} \rho(T_1) &= \max\left(|0|, \left|i\sqrt{\frac{5}{4}}\right|, \left|-i\sqrt{\frac{5}{4}}\right|\right) \\ &= \max\left(0, \sqrt{\frac{5}{4}}, \sqrt{\frac{5}{4}}\right) \\ &= \sqrt{\frac{5}{4}} \end{aligned}$$

We do the same for T_2:

$$\begin{aligned} |T_2 - \lambda I| &= \begin{vmatrix} -\lambda & 1/2 & -1/2 \\ 0 & -1/2 - \lambda & -1/2 \\ 0 & 0 & -1/2 - \lambda \end{vmatrix} \\ &= -\lambda(-1/2 - \lambda)(-1/2 - \lambda) - 0 + 0 \\ &= -\lambda(1/2 + \lambda)(1/2 + \lambda) \end{aligned}$$

The roots of the characteristic polynomal, and thus the eigenvalues of T_1 are 0 and $\pm\frac{1}{2}$. The spectral radius of T_2 is given as

$$\begin{aligned} \rho(T_2) &= \max\left(|0|, \left|\frac{1}{2}\right|, \left|-\frac{1}{2}\right|\right) \\ &= \max\left(0, \frac{1}{2}, \frac{1}{2}\right) \\ &= \frac{1}{2} \end{aligned}$$

Part 3

Show that for any matrix $A \in \mathbb{R}^{nxn}$

$$\|A\|_F := \left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2}$$

defines a matrix norm (the so-called Frobenius norm.) Use the Cauchy-Schwarz inequality to show that for any matrix $A \in \mathbb{R}^{nxn}$ and any vector $x \in \mathbb{R}^n$

$$\|Ax\|_2 \leq \|A\|_F\|x\|_2.$$

We need to show four properties that must be satisfied for $\|A\|_F$ to be considered a matrix norm:

1. $\|A\|_F \geq 0$
2. $\|A\|_F = 0 \Leftrightarrow A = 0_{n,n}$
3. $\|\alpha A\|_F = |\alpha|\|A\|_F$ for scalar α
4. $\|A+B\|_F \leq \|A\|_F + \|B\|_F$ for other matrix B

1. $\|A\|_F \geq 0$

A squared scalar is always ≥ 0. This gives the following results:

$$\begin{aligned}
|a_{ij}|^2 &\geq 0 \\
\Rightarrow \left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\right) &\geq 0 \\
\Rightarrow \left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2} &\geq 0 \\
\Rightarrow \|A\|_F &\geq 0
\end{aligned}$$

2. $\|A\|_F = 0 \Leftrightarrow A = 0_{n,n}$

We know from (1) that $|a_{ij}|^2 \geq 0$. This means that $\left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\right) \geq |a_{ij}|^2$.

Assume $\|A\|_F = 0$. Then

$$\begin{aligned}
\left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2} &= 0 \\
\Rightarrow \sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2 &= 0 \\
\Rightarrow |a_{ij}|^2 &\leq 0
\end{aligned}$$

As we know from (1) that $|a_{ij}|^2 \geq 0$. This means that

$$\begin{aligned}
|a_{ij}|^2 &= 0 \\
\Rightarrow |a_{ij}| &= 0 \\
\Rightarrow A &= 0_{n,n}
\end{aligned}$$

3. $\|\alpha A\|_F = |\alpha|\|A\|_F$ for scalar α

Observe that

$$\begin{aligned}
\|\alpha A\|_F &= \left(\sum_{i=1}^{n}\sum_{j=1}^{n}|\alpha a_{ij}|^2\right)^{1/2} \\
&= \left(\sum_{i=1}^{n}\sum_{j=1}^{n}(|a_{ij}||\alpha|)^2\right)^{1/2} \\
&= \left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2|\alpha|^2\right)^{1/2} \\
&= \left(|\alpha|^2\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2} \\
&= |\alpha|\left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2} \\
&= |\alpha|\|A\|_F
\end{aligned}$$

4. $\|A+B\|_F \leq \|A\|_F + \|B\|_F$ for other matrix B

Observe that

$$\begin{aligned}
\|A+B\|_F &= \left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}+b_{ij}|^2\right)^{1/2} \\
&\leq \left(\sum_{i=1}^{n}\sum_{j=1}^{n}(|a_{ij}|+|b_{ij}|)^2\right)^{1/2} \\
&= \left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2 + 2|a_{ij}||b_{ij}| + |b_{ij}|^2\right)^{1/2}
\end{aligned}$$

Because $|a_{ij}| \geq 0$ and $|b_{ij}| \geq 0$, $2|a_{ij}||b_{ij}| \geq 0$, and we have

$$\begin{aligned}
\|A+B\|_F &\leq \left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2 + 2|a_{ij}||b_{ij}| + |b_{ij}|^2\right)^{1/2} \\
&\leq \left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2 + |b_{ij}|^2\right)^{1/2} \\
&= \left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2 + \sum_{i=1}^{n}\sum_{j=1}^{n}|b_{ij}|^2\right)^{1/2} \\
&\leq \left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2} + \left(\sum_{i=1}^{n}\sum_{j=1}^{n}|b_{ij}|^2\right)^{1/2} \\
&= \|A\|_F + \|B\|_F
\end{aligned}$$

Use the Cauchy-Schwarz inequality to show that for any matrix $A \in \mathbb{R}^{nxn}$ and any vector $x \in \mathbb{R}^n$

$$\|Ax\|_2 \leq \|A\|_F \|x\|_2.$$

Note that we can write $\|Ax\|_2$ as

$$\|Ax\|_2 = \sum_{i=1}^{n} \left| \sum_{j=1}^{n} a_{ij} h_j \right|^2$$

Using the Cauchy-Schwarz inequality, observe that

$$\begin{aligned}
\|Ax\|_2 &= \sum_{i=1}^{n} \left| \sum_{j=1}^{n} a_{ij} h_j \right|^2 \\
&\leq \sum_{i=1}^{n} \left\{ \left(\sum_{j=1}^{n} |a_{ij}|^2 \right) \left(\sum_{j=1}^{n} |h_j|^2 \right) \right\} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}|^2 \|x\|_2 \\
&= \|A\|_F \|x\|_2
\end{aligned}$$

Problem 2

Consider the system

$$\begin{aligned} 10x_1 - x_2 + 2x_3 &= 6 \\ -x_1 + 11x_2 - x_3 + 3x_4 &= 25 \\ 2x_1 - x_2 - 10x_3 - x_4 &= -11 \\ 3x_2 - x_3 + x_4 &= 15 \end{aligned}$$

Find its exact solution. Write down the Jacobi iterative method and generate the first 3 entries, $x^{(1)}$, $x^{(2)}$, $x^{(3)}$ in the sequence of approximations $\{x^{(n)}\}_{n>0}$, $x^{(0)} = (0,0,0,0)^T$. Repeat the approximation using the Gauss-Seidel iterative method.

We start by finding the exact solution of the matrix form of the system:

$$\begin{bmatrix} 10 & -1 & 2 & 0 \\ -1 & 11 & -1 & 3 \\ 2 & -1 & -10 & -1 \\ 0 & 3 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 6 \\ 25 \\ -11 \\ 15 \end{bmatrix}$$

$$\begin{bmatrix} -1 & 11 & -1 & 3 & 25 \\ 10 & -1 & 2 & 0 & 6 \\ 2 & -1 & -10 & -1 & -11 \\ 0 & 3 & -1 & 1 & 15 \end{bmatrix}$$

$$\begin{bmatrix} 1 & -11 & 1 & -3 & -25 \\ 0 & 109 & -8 & 30 & 256 \\ 0 & 21 & -12 & 5 & 39 \\ 0 & 3 & -1 & 1 & 15 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & -8/3 & 2/3 & 30 \\ 0 & 1 & -1/3 & 1/3 & 5 \\ 0 & 0 & 85/3 & -19/3 & -289 \\ 0 & 0 & -5 & -2 & -66 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 26/15 & 326/5 \\ 0 & 1 & 0 & 7/15 & 47/5 \\ 0 & 0 & 1 & 2/5 & 66/5 \\ 0 & 0 & 0 & -53/3 & -663 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 8/53 \\ 0 & 1 & 0 & 0 & -430/53 \\ 0 & 0 & 1 & 0 & -96/53 \\ 0 & 0 & 0 & 1 & 1989/53 \end{bmatrix}$$

Which gives the exact solution

$$\begin{aligned} x_1 &= 8/53 \\ x_2 &= -430/53 \\ x_3 &= -96/53 \\ x_4 &= 1989/53 \end{aligned}$$

Jacobi iterative method

Let b be the target vector $(6, 25, -11, 15)$, D be the diagonal of the matrix form of the system and $L + U =$ matrix form $-D$. With the Jacobi iterative method, we have that

$$x^{(k)} = D^{-1}\left(b - (L+U)x^{(k-1)}\right)$$
$$x_i^{(k)} = \frac{1}{a_{ii}}\left(b_i - \sum_{j\neq i} a_{ij}x_j^{(k-1)}\right)$$

Remember that $x^{(0)} = (0,0,0,0)^T$.

$$x^{(1)} = \begin{bmatrix} \frac{1}{10}(6 - (-1*0 + 2*0 + 0*0)) \\ \frac{1}{11}(25 - (-1*0 - 1*0 + 3*0)) \\ \frac{1}{-10}(-11 - (2*0 - 1*0 - 1*0)) \\ 1(15 - (0*0 + 3*0 - 1*0)) \end{bmatrix} = \begin{bmatrix} 6/10 \\ 25/11 \\ 11/10 \\ 15 \end{bmatrix}$$

$$x^{(2)} = \begin{bmatrix} \frac{1}{10}(6 - (-1*25/11 + 2*11/10 + 0*15)) \\ \frac{1}{11}(25 - (-1*6/10 - 1*11/10 + 3*15)) \\ \frac{1}{-10}(-11 - (2*6/10 - 1*25/11 - 1*15)) \\ 1(15 - (0*6/10 + 3*25/11 - 1*11/10)) \end{bmatrix} = \begin{bmatrix} 334/550 \\ -183/110 \\ -279/550 \\ 1021/110 \end{bmatrix}$$

$$x^{(3)} = \begin{bmatrix} \frac{1}{10}(6 - (-1*-183/110 + 2*-279/550 + 0*1021/110)) \\ \frac{1}{11}(25 - (-1*334/550 - 1*-279/550 + 3*1021/110)) \\ \frac{1}{-10}(-11 - (2*334/550 - 1*-183/110 - 1*1021/110)) \\ 1(15 - (0*334/550 + 3*-183/110 - 1*1021/110)) \end{bmatrix} = \begin{bmatrix} 2943/5500 \\ -151/605 \\ 632/1375 \\ 322/11 \end{bmatrix}$$

Gauss-Seidel iterative method

With the Gauss-Seidel iterative method, we have that

$$x_i^{(k)} = \frac{1}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k)} - \sum_{j=i+1}^{n} a_{ij}x_j^{k-1}\right)$$

Remember that $x^{(0)} = (0,0,0,0)^T$. Note that the Gauss-Seidel method computes each element $x_i^{(k)}$ one at a time, using the previously calculated elements for later calculations with the same k

We can compute the general formula for $x^{(k)}$ ahead of time:

$$x^{(k)} = \begin{bmatrix} \frac{1}{10}\left(6 - (0) - (-1*x_2^{(k-1)} + 2*x_3^{(k-1)} + 0*x_4^{(k-1)}))\right) \\ \frac{1}{11}\left(25 - (-1*x_1^{(k)}) - (-1*x_3^{(k-1)} + 3*x_4^{(k-1)})\right) \\ \frac{1}{-10}\left(-11 - (2*x_1^{(k)} - 1*x_2^{(k)}) - (1*x_4^{(k-1)})\right) \\ 1\left(15 - (0*x_1^{(k)} + 3*x_2^{(k)} - 1*x_3^{(k)})\right) \end{bmatrix} = \begin{bmatrix} \frac{1}{10}\left(6 + x_2^{(k-1)} - 2*x_3^{(k-1)}\right) \\ \frac{1}{11}\left(25 + x_1^{(k)} + x_3^{(k-1)} - 3*x_4^{(k-1)}\right) \\ \frac{1}{-10}\left(-11 - 2*x_1^{(k)} + x_2^{(k)} + x_4^{(k-1)}\right) \\ \left(15 - 3*x_2^{(k)} + x_3^{(k)}\right) \end{bmatrix}$$

$$x^{(1)} = \begin{bmatrix} \frac{1}{10}(6 + 0 - 2*0) \\ \frac{1}{11}(25 + 6/10 + 0 - 3*0) \\ \frac{1}{-10}(-11 - 2*6/10 + 128/55 + 0) \\ (15 - 3*128/55 + 543/550) \end{bmatrix} = \begin{bmatrix} 6/10 \\ 128/55 \\ 543/550 \\ 4953/550 \end{bmatrix}$$

$$x^{(2)} = \begin{bmatrix} \frac{1}{10}\left(6 + 128/55 - 2 * 543/550\right) \\ \frac{1}{11}\left(25 + 1747/2750 + 543/550 - 3 * 4953/550\right) \\ \frac{1}{-10}\left(-11 - 2 * 1747/2750 + 1181/6050 + 4953/550\right) \\ \left(15 - 3 * 1181/6050 + 0.306988\right) \end{bmatrix} = \begin{bmatrix} 1747/2750 \\ 1181/6050 \\ 0.306988 \\ 14.721368 \end{bmatrix}$$

$$x^{(1)} = \begin{bmatrix} \frac{1}{10}\left(6 + 1181/6050 - 2 * 0.306988\right) \\ \frac{1}{11}\left(25 + 0.558123 + 0.306988 - 3 * 14.721368\right) \\ \frac{1}{-10}\left(-11 - 2 * 0.558123 - 1.663545 + 14.721368\right) \\ \left(15 - 3 * -1.663545 - 0.094158\right) \end{bmatrix} = \begin{bmatrix} 0.558123 \\ -1.663545 \\ -0.094158 \\ 19.896477 \end{bmatrix}$$

Problem 3

Looking at the system

$$
\begin{aligned}
2x_1 - x_2 + x_3 &= -1 \\
2x_1 + 2x_2 + 2x_3 &= 4 \\
-x_1 - x_2 + 2x_3 &= -5
\end{aligned}
$$

and using the results from Problem 1(2), what can you say about applying the Jacobi iterative method and the Gauss–Seidel iterative method, both for initial value $x^{(0)} = (0,0,0)^T$*?*

We start by writing this system in matrix form, $Ax = b$:

$$
\begin{bmatrix} 2 & -1 & 1 \\ 2 & 2 & 2 \\ -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 4 \\ -5 \end{bmatrix}
$$

The standard convergence condition for any iterative method is that the spectral radius of the iteration matrix is less than 1:

$$
\rho(D^{-1}(L+U)) < 1
$$

where D is the diagonal of A, L is the lower triangular matrix and U is the upper triangular matrix of A.

Let T_1 be the matrix from Problem 1(2). Observe that $A = -2T_1 + 2I = 2(I - T_1)$. This is the same as saying

$$
\begin{aligned}
L + U &= -2T_1 \\
D = 2I &\Rightarrow D^{-1} = I/2
\end{aligned}
$$

We thus have

$$
\begin{aligned}
\rho(D^{-1}(L+U)) &= \rho(I/2(-2T_1)) \\
&= \rho(-T_1)
\end{aligned}
$$

As the spectral radius looks at the absolute value of eigenvalues of a matrix and $eigenvalues_{-T_1} = -eigenvalues_{T_1}$, we have that $\rho(-T_1) = \rho(T_1)$. Combine this with what we found in Problem 1(2), and we see that

$$
\rho(D^{-1}(L+U)) = \rho(-T_1) = \rho(T_1) = \sqrt{\frac{5}{4}} > 1
$$

As the standard convergence condition is not satisfied, we see that neither the Jacobi iterative method nor the Gauss-Seidel iterative method will converge.

Problem 4

Use LU factorisation to find the solution of the system (provide all details):

$$
\begin{aligned}
x_1 + x_2 + 3x_4 &= 8 \\
2x_1 + x_2 - x_3 + x_4 &= 7 \\
3x_1 - x_2 - x_3 + 2x_4 &= 14 \\
-x_1 + 2x_2 + 3x_3 - x_4 &= -7
\end{aligned}
$$

Let

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

Then we want to find triangular matrices L, U such that $A = LU$. We have that

$$
A = \begin{bmatrix} 1 & 1 & 0 & 3 \\ 2 & 1 & -1 & 1 \\ 3 & -1 & -1 & 2 \\ -1 & 2 & 3 & -1 \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 & 0 \\ l_{21} & l_{22} & 0 & 0 \\ l_{31} & l_{32} & l_{33} & 0 \\ l_{41} & l_{42} & l_{43} & l_{44} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ 0 & u_{22} & u_{23} & u_{24} \\ 0 & 0 & u_{33} & u_{34} \\ 0 & 0 & 0 & u_{44} \end{bmatrix}
$$

We can solve this as a system of equations:

$$
\begin{aligned}
l_{11} * u_{11} &= 1 \\
l_{11} * u_{12} &= 1 \\
l_{11} * u_{13} &= 0 \\
l_{11} * u_{14} &= 3 \\
l_{21} * u_{11} &= 2 \\
l_{21} * u_{12} + l_{22} * u_{22} &= 1 \\
l_{21} * u_{13} + l_{22} * u_{23} &= -1 \\
l_{21} * u_{14} + l_{22} * u_{24} &= 1 \\
l_{31} * u_{11} &= 3 \\
l_{31} * u_{12} + l_{32} * u_{22} &= -1 \\
l_{31} * u_{13} + l_{32} * u_{23} + l_{33} * u_{33} &= -1 \\
l_{31} * u_{14} + l_{32} * u_{24} + l_{33} * u_{34} &= 2 \\
l_{41} * u_{11} &= -1 \\
l_{41} * u_{12} + l_{42} * u_{22} &= 2 \\
l_{41} * u_{13} + l_{42} * u_{23} + l_{43} * u_{33} &= 3 \\
l_{41} * u_{14} + l_{42} * u_{24} + l_{43} * u_{34} + l_{44} * u_{44} &= -1
\end{aligned}
$$

To solve this, we need some limitations on the elements. We require L to be a unit triangular matrix. We can solve the system, and see that

$$\begin{aligned}
l_{11} = l_{22} = l_{33} = l_{44} &= 1 \\
u_{11} &= 1 \\
u_{12} &= 1 \\
u_{13} &= 0 \\
u_{14} &= 3 \\
l_{21} &= 2 \\
u_{22} &= 1 - 2 = -1 \\
u_{23} &= -1 \\
u_{24} &= 1 - 6 = -5 \\
l_{31} &= 3 \\
l_{32} &= (-1 - 3)/-1 = 4 \\
u_{33} &= -1 + 4 = 3 \\
u_{34} &= 2 - 9 + 20 = 13 \\
l_{41} &= -1 \\
l_{42} &= (2 + 1)/-1 = -3 \\
l_{43} &= (3 - 3)/3 = 0 \\
u_{44} &= -1 + 3 - 15 = -13
\end{aligned}$$

We then have

$$A = \begin{bmatrix} 1 & 1 & 0 & 3 \\ 2 & 1 & -1 & 1 \\ 3 & -1 & -1 & 2 \\ -1 & 2 & 3 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ -1 & -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{bmatrix} = LU$$

To solve the system given in the assignment, observe that $Ax = LUx$. We first calculate y such that $Ly = b$, then we calculate x such that $Ux = y$, which means that $Ax = LUx = Ly = b$.

$$Ly = b$$
$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ -1 & -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 8 \\ 7 \\ 14 \\ -7 \end{bmatrix}$$

We can convert this to a set of equations, and find that

$$\begin{aligned}
y_1 &= 8 \\
2y_1 + y_2 = 7 \Rightarrow y_2 &= 7 - 16 = -9 \\
3y_1 + 4y_2 + y_3 = 14 \Rightarrow y_3 &= 14 - 24 + 36 = 26 \\
-y_1 - 3y_2 + y_4 = -7 \Rightarrow y_4 &= -7 + 8 - 27 = -26
\end{aligned}$$

We then find

$$Ux = y$$
$$\begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 8 \\ -9 \\ 26 \\ -26 \end{bmatrix}$$

We convert this to a set of equations, and find that

$$\begin{aligned}
-13x_4 = -26 &\Rightarrow x_4 = -26/-13 = 2 \\
3x_3 + 13x_4 = 26 &\Rightarrow x_3 = (26-26)/3 = 0 \\
-x_2 - x_3 - 5x_4 = -9 &\Rightarrow x_2 = (-9+0+10)/-1 = -1 \\
x_1 + x_2 + 3x_4 = 8 &\Rightarrow x_1 = 8+1-6 = 3
\end{aligned}$$

We then have a solution to the system of equations above:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ 0 \\ 2 \end{bmatrix}$$

Problem 5

Formulate the problem of finding a straight line $y = x_1 + tx_2$ fitting the following points in the (t, y)-plane

$$\begin{gathered}
(1, 1.4501) \\
(2, 1.7311) \\
(3, 3.1068) \\
(4, 3.9860) \\
(5, 5.3913)
\end{gathered}$$

as a least squares problem and solve it (find x_1, x_2).

We define $f(t) = x_1 + tx_2$ to the function for our line, and define r_i as the residuals for each i, i.e.

$$r_i = y_i - f(t_i) = y_i - x_1 - t_i x_2 \qquad \forall i \in [1, 5]$$

$$S = \sum_{i=1}^{5} r_i^2 = \sum_{i=1}^{5} (y_i - f(t_i))^2 = \sum_{i=1}^{5} (y_i - x_1 - t_i x_2)^2$$

To find the straight line fitting the points, we want to minimize S, i.e. $\min_{x_1, x_2} S$.

Note that we want to find values of x_1, x_2 that minimizes the error. This is, by definition, in the points where

$$\frac{\partial S}{\partial x_1} = \frac{\partial S}{\partial x_2} = 0$$

Observe that

$$\begin{aligned}
\frac{\partial S}{\partial x_1} &= \sum_{i=1}^{5} 2(y_i - x_1 - t_i x_2)(-1) \\
\frac{\partial S}{\partial x_1} &= \sum_{i=1}^{5} 2(y_i - x_1 - t_i x_2)(-t_i)
\end{aligned}$$

Setting these equal to 0 gives us

$$\sum_{i=1}^{5}(y_i - x_1 - t_i x_2) = 0$$
$$\sum_{i=1}^{5}(y_i - x_1 - t_i x_2)(t_i) = 0$$

We can rewrite this as

$$x_1 \left(\sum_{i=1}^{5} 1\right) + x_2 \left(\sum_{i=1}^{5} t_i\right) = \left(\sum_{i=1}^{5} y_i\right)$$
$$x_1 \left(\sum_{i=1}^{5} t_i\right) + x_2 \left(\sum_{i=1}^{5} t_i^2\right) = \left(\sum_{i=1}^{5} t_i y_i\right)$$

We can compute these sums by hand, easily

$$\sum_{i=1}^{5} 1 = 5$$
$$\sum_{i=1}^{5} t_i = 1 + 2 + 3 + 4 + 5 = 15$$
$$\sum_{i=1}^{5} y_i = 1.4501 + 1.7311 + 3.1068 + 3.9860 + 5.3913 = 15.6653$$
$$\sum_{i=1}^{5} t_i^2 = 1 + 4 + 9 + 16 + 25 = 55$$
$$\sum_{i=1}^{5} t_i y_i = 1.4501 + 2 * 1.7311 + 3 * 3.1068 + 4 * 3.9860 + 5 * 5.3913 = 57.1332$$

We then have a system of two equations

$$x_1 * 5 + x_2 * 15 = 15.6653$$
$$x_1 * 15 + x_2 * 55 = 57.1332$$

We simply solve this by hand:

$$\begin{aligned} x_1 * 5 &= 15.6653 - x_2 * 15 \\ 3 * (15.6653 - x_2 * 15) + x_2 * 55 &= 57.1332 \\ 49.9959 - 45x_2 + 55x_2 &= 57.1332 \\ 10x_2 &= 7.1373 \\ x_2 &= 0.71373 \\ x_1 &= 15.6653 - x_2 * 15 = 15.6653 - 10.70595 = 4.95935 \end{aligned}$$

We then have the best-fitting straight line for the problem given as

$$y = 4.95935 + 0.71373t$$

Problem 6

Süli-Mayers: Ex. 1.10, 2.7, 2.14, 2.15, 5.1, 5.2

Exercise 1.10

Write the secant iteration in the form

$$x_{k+1} = \frac{x_k f(x_{k-1}) - x_{k-1} f(x_k)}{f(x_{k-1}) - f(x_k)}$$

Supposing that f has a continuous second derivative in the neighbourhood of the solution ξ of $f(x) = 0$, and that $f'(\xi) > 0$ and $f''(\xi) > 0$, define

$$\varphi(x_k, x_{k-1}) = \frac{x_{k+1} - \xi}{(x_k - \xi)(x_{k-1} - \xi)}$$

where x_{k+1} has been expressed in terms of x_k and x_{k-1}. Find an expression for

$$\psi(x_{k-1}) = \lim_{x_k \to \xi} \varphi(x_k, x_{k-1})$$

and determine $\lim_{x_{k-1} \to \xi} \psi(x_{k-1})$.

Observe that

$$\begin{aligned}
\psi(x_{k-1}) &= \lim_{x_k \to \xi} \varphi(x_k, x_{k-1}) \\
&= \lim_{x_k \to \xi} \frac{x_{k+1} - \xi}{(x_k - \xi)(x_{k-1} - \xi)} \\
&= \lim_{x_k \to \xi} \frac{\frac{x_k f(x_{k-1}) - x_{k-1} f(x_k)}{f(x_{k-1}) - f(x_k)} - \xi}{(x_k - \xi)(x_{k-1} - \xi)} \\
&= \lim_{x_k \to \xi} \frac{\frac{x_k f(x_{k-1}) - x_{k-1} f(x_k) - \xi(f(x_{k-1}) - f(x_k))}{f(x_{k-1}) - f(x_k)}}{(x_k - \xi)(x_{k-1} - \xi)} \\
&= \lim_{x_k \to \xi} \frac{x_k f(x_{k-1}) - x_{k-1} f(x_k) - \xi(f(x_{k-1}) - f(x_k))}{(x_k - \xi)(x_{k-1} - \xi)(f(x_{k-1}) - f(x_k))} \\
&= \lim_{x_k \to \xi} \frac{f(x_{k-1})(x_k - \xi) - f(x_k)(x_{k-1} - \xi)}{(x_k - \xi)(x_{k-1} - \xi)(f(x_{k-1}) - f(x_k))}
\end{aligned}$$

This is a 0/0-case, so we can use L'Hôpital's rule to get further:

$$\begin{aligned}
\psi(x_{k-1}) &= \lim_{x_k \to \xi} \frac{f(x_{k-1})(x_k - \xi) - f(x_k)(x_{k-1} - \xi)}{(x_k - \xi)(x_{k-1} - \xi)(f(x_{k-1}) - f(x_k))} \\
&= \lim_{x_k \to \xi} \frac{\frac{d}{dx_k} f(x_{k-1})(x_k - \xi) - \frac{d}{dx_k} f(x_k)(x_{k-1} - \xi)}{\frac{d}{dx_k}(x_k - \xi)(x_{k-1} - \xi)(f(x_{k-1}) - f(x_k))} \\
&= \lim_{x_k \to \xi} \frac{f(x_{k-1}) - f'(x_k)(x_{k-1} - \xi)}{(x_{k-1} - \xi)(f(x_{k-1}) - f(x_k)) - (x_k - \xi)(x_{k-1} - \xi) f'(x_k))} \\
&= \lim_{x_k \to \xi} \frac{f(x_{k-1}) - f'(x_k)(x_{k-1} - \xi)}{(x_{k-1} - \xi)(f(x_{k-1}) - f(x_k)) - (x_k - \xi)(x_{k-1} - \xi) f'(x_k))} \\
&= \frac{f(x_{k-1}) - f'(\xi)(x_{k-1} - \xi)}{(x_{k-1} - \xi) f(x_{k-1})}
\end{aligned}$$

We can then determine

$$\lim_{x_{k-1}\to\xi}\psi(x_{k-1}) = \lim_{x_{k-1}\to\xi}\frac{f(x_{k-1}) - f'(\xi)(x_{k-1}-\xi)}{(x_{k-1}-\xi)f(x_{k-1})}$$

This is, once more, a 0/0-case, and we once more apply L'Hôpitals method

$$\begin{aligned}
\lim_{x_{k-1}\to\xi}\psi(x_{k-1}) &= \lim_{x_{k-1}\to\xi}\frac{f(x_{k-1}) - f'(\xi)(x_{k-1}-\xi)}{(x_{k-1}-\xi)f(x_{k-1})} \\
&= \lim_{x_{k-1}\to\xi}\frac{\frac{d}{dx_{k-1}}f(x_{k-1}) - \frac{d}{dx_{k-1}}f'(\xi)(x_{k-1}-\xi)}{\frac{d}{dx_{k-1}}(x_{k-1}-\xi)f(x_{k-1})} \\
&= \lim_{x_{k-1}\to\xi}\frac{f'(x_{k-1}) - f'(\xi)}{f(x_{k-1}) + (x_{k-1}-\xi)f'(x_{k-1})}
\end{aligned}$$

This is, once more, a 0/0-case, and we once more apply L'Hôpitals method

$$\begin{aligned}
\lim_{x_{k-1}\to\xi}\psi(x_{k-1}) &= \lim_{x_{k-1}\to\xi}\frac{f'(x_{k-1}) - f'(\xi)}{f(x_{k-1}) + (x_{k-1}-\xi)f'(x_{k-1})} \\
&= \lim_{x_{k-1}\to\xi}\frac{\frac{d}{dx_{k-1}}f'(x_{k-1}) - \frac{d}{dx_{k-1}}f'(\xi)}{\frac{d}{dx_{k-1}}f(x_{k-1}) + \frac{d}{dx_{k-1}}(x_{k-1}-\xi)f'(x_{k-1})} \\
&= \lim_{x_{k-1}\to\xi}\frac{f''(x_{k-1})}{f'(x_{k-1}) + f'(x_{k-1}) + (x_{k-1}-\xi)f''(x_{k-1})} \\
&= \frac{f''(\xi)}{f'(\xi)+f'(\xi)} \\
&= \frac{f''(\xi)}{2f'(\xi)}
\end{aligned}$$

Deduce that

$$\lim_{x_k,x_{k-1}\to\xi}\varphi(x_k,x_{k-1}) = f''(\xi)/2f'(\xi)$$

This follows from our previous calculation:

$$\begin{aligned}
\lim_{x_k,x_{k-1}\to\xi}\varphi(x_k,x_{k-1}) &= \lim_{x_{k-1}\to\xi}\lim_{x_k\to\xi}\varphi(x_k,x_{k-1}) \\
&= \lim_{x_k\to\xi}\psi(x_{k-1}) \\
&= \frac{f''(\xi)}{2f'(\xi)}
\end{aligned}$$

Now assume that

$$\lim_{k\to\infty}\frac{|x_{k+1}-\xi|}{|x_k-\xi|^q} = A$$

Show that $q - 1 - 1/q = 0$, and hence that $q = \frac{1}{2}(1+\sqrt{5})$.

If $q - 1 - 1/q = 0$, then we have $q^2 - q - 1 = 0$. The ABC-rule then gives us

$$q = \frac{1\pm\sqrt{5}}{2}$$

and, since $q > 0$, it must be that

$$q = \frac{1}{2}(1 + \sqrt{5})$$

Deduce finally that

$$\lim_{k\to\infty} \frac{|x_{k+1} - \xi|}{|x_k - \xi|^q} = \left(\frac{f''(\xi)}{2f'(\xi)}\right)^{q/(1+q)}$$

I ran out of time for this exercise, after struggling with the above. I am therefore unable to show that $q - 1 - q/1 = 0$ and the final deduction. I am sorry for that!

Exercise 2.7

Suppose that for a matrix $A \in \mathbb{R}^{nxn}$,

$$\sum_{i=1}^{n} |a_{ij}| \leq C, \; j = 1, ..., n$$

Show that, for any vector $x \in \mathbb{R}^n$,

$$\sum_{i=1}^{n} |(Ax)_i| \leq C \|x\|_1$$

Note that

$$(Ax)_i = \sum_{k=1}^{n} a_{ik} x_k$$

It follows from this, the triangle inequality and the definition of the 1-norm that

$$\begin{aligned}
\sum_{i=1}^{n} |(Ax)_i| &= \sum_{i=1}^{n} \left| \sum_{k=1}^{n} a_{ik} x_k \right| \\
&\leq \sum_{i=1}^{n} \sum_{k=1}^{n} |a_{ik}||x_k| \\
&= \sum_{k=1}^{n} \sum_{i=1}^{n} |a_{ik}||x_k| \\
&\leq \sum_{k=1}^{n} C|x_k| \\
&= C \sum_{k=1}^{n} |x_k| \\
&= C\|x\|_1
\end{aligned}$$

Find a nonzero vector x for which equality can be achieved, and deduce that

$$\|A\|_1 = \max_{j=1}^{n} \sum_{i=1}^{n} |a_{ij}|$$

Let

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}, x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Observe that

$$\sum_{i=1}^{n} |a_{ij}| = 3 \; \forall j$$
$$\Rightarrow C = 3$$
$$\|x\|_1 = \sum_{i=1}^{2} |x_i| = 1 + 1 = 2$$
$$Ax = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$$

It is easy to see that

$$\sum_{i=1}^{n} |(Ax)_i| = 1 + 5 = 6 = 3 * 2 = C\|x\|_1$$

Exercise 2.14

Suppose that $A \in \mathbb{R}^{nxn}$ is a nonsingular matrix, and $b \in \mathbb{R}^n_*$. Given that $Ax = b$ and $A(x + \delta x) = b + \delta b$, Theorem 2.11 states that

$$\frac{\|\delta x\|}{\|x\|} \leq \kappa(A)\frac{\|\delta b\|}{\|b\|}$$

By considering the eigenvectors of $A^T A$, show how to find vectors b and δb for which equality is attained, when using the two-norm.

We want to find vectors b and δb such that

$$\frac{\|\delta x\|_2}{\|x\|_2} = \kappa(A)\frac{\|\delta b\|_2}{\|b\|_2}$$

Note that theorem 2.11 comes from multiplying two inequalities:

$$\|b\|_2 \leq \|A\|_2 \|x\|_2$$
$$\|\delta x\|_2 \leq \left\|A^{-1}\right\|_2 \|\delta b\|_2$$

In order to end up with an equality, we want to find vectors b and δb that makes the latter two inequalities equalities. I.e., we want to find vectors such that

$$\|b\|_2 = \|A\|_2 \|x\|_2$$
$$\|\delta b\|_2 = \frac{\|\delta x\|_2}{\|A^{-1}\|_2}$$

As the 2-norm of a matrix A is the largest singular value of the matrix $A^T A$, use the eigenvectors of $A^T A$ to find the eigenvalues of $A^T A$, by applying the formula $(A^T A)x = \lambda x$, where x is the eigenvectors and λ is the eigenvalues. Then, $\|A\|_2 = \sqrt{\lambda_{\max}}$. For the 2-norm of the inverse matrix, note that $\left\|A^{-1}\right\|_2 = 1/\sqrt{\lambda_{\min}}$.

As such, we achieve equality by letting

$$b = \sqrt{\lambda_{max}} * x$$
$$\delta b = \sqrt{\lambda_{\min}} * \delta x$$

Exercise 2.15

Find the QR factorization of the matrix

$$A = \begin{bmatrix} 9 & -6 \\ 12 & -8 \\ 0 & 20 \end{bmatrix}$$

We use the Gram-Schmidt process to find the QR factorization. We first find the projection matrix U, where $u_1 = a_1$, $u_2 = a_2 - proj_{u_1} a_2$ for column a_i in A, and $proj_u a = \frac{u \cdot a}{u \cdot u} u$.

$$\begin{aligned}
U &= [u_1, u_2] \\
u_1 &= a_1 = (9, 12, 0)^T \\
u_2 &= a_2 - proj_{u_1} a_2 = (-6, -8, 20)^T - \frac{u_1 \cdot a_2}{u_1 \cdot u_1} u_1 \\
&= (-6, -8, 20)^T - \frac{9 * (-6) + 12 * (-8)}{9^2 + 12^2} (9, 12, 0)^T \\
&= (-6, -8, 20)^T - \frac{-150}{225} (9, 12, 0)^T \\
&= (-6, -8, 20)^T - (-6, -8, 0)^T \\
&= (0, 0, 20)^T
\end{aligned}$$

We then compute $Q = [\frac{u_1}{\|u_1\|}, \frac{u_2}{\|u_2\|}]$

$$\begin{aligned}
Q &= [\frac{u_1}{\|u_1\|}, \frac{u_2}{\|u_2\|}] \\
\frac{u_1}{\|u_1\|} &= \frac{(9, 12, 0)^T}{\sqrt{9^2 + 12^2}} \\
&= (\frac{3}{5}, \frac{4}{5}, 0)^T \\
\frac{u_2}{\|u_2\|} &= \frac{(0, 020)^T}{\sqrt{20^2}} \\
&= (0, 0, 1)^T \\
Q &= \begin{bmatrix} 3/5 & 0 \\ 4/5 & 0 \\ 0 & 1 \end{bmatrix}
\end{aligned}$$

We use the fact that $Q^T Q = I$ to find $R = Q^T Q R = Q^T A$

$$\begin{aligned}
R &= Q^T A \\
&= \begin{bmatrix} 3/5 & 4/5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 9 & -6 \\ 12 & -8 \\ 0 & 20 \end{bmatrix} \\
&= \begin{bmatrix} 15 & -10 \\ 0 & 20 \end{bmatrix}
\end{aligned}$$

We thus have

$$A = QR = \begin{bmatrix} 3/5 & 0 \\ 4/5 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 15 & -10 \\ 0 & 20 \end{bmatrix}$$

Find the least squares solution of the system of linear equations

$$\begin{aligned} 9x - 6y &= 300 \\ 12x - 8y &= 600 \\ 20y &= 900 \end{aligned}$$

To solve this, we know that the optimal solution can be found as

$$\hat{x} = R_1^{-1}(Q_1^T b)$$

where $R_1 = R$ and $Q_1 = Q$ (in our case) and $b = (300, 600, 900)^T$. Note that

$$R^{-1} = \begin{bmatrix} 1/15 & 1/30 \\ 0 & 1/20 \end{bmatrix}$$

We can then compute

$$\begin{aligned} \hat{x} &= R^{-1}(Q^T b) \\ &= \begin{bmatrix} 1/15 & 1/30 \\ 0 & 1/20 \end{bmatrix} \left(\begin{bmatrix} 3/5 & 4/5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 300 \\ 600 \\ 900 \end{bmatrix} \right) \\ &= \begin{bmatrix} 1/15 & 1/30 \\ 0 & 1/20 \end{bmatrix} \begin{bmatrix} 660 \\ 900 \end{bmatrix} \\ &= \begin{bmatrix} 74 \\ 45 \end{bmatrix} \end{aligned}$$

The least squares solution of the system is given as

$$\begin{aligned} x &= 74 \\ y &= 45 \end{aligned}$$

Exercise 5.1

Give a proof of Lemma 5.3

H_k is created by an original vector v^0 with length k. Assume now that you create a new Household matrix based on a new vector v^1 with length n, created by combining a 0-vector with length $n-k$ with the original v^0, i.e.

$$v^1 = (0, ..., 0, v_1^0, ..., v_k^0)^T$$

As $\frac{2}{(v^1)^T v^1} v^1 (v^1)^T$ is a matrix with 0 in the first $n-k$ rows and columns, and the rest being equal to $\frac{2}{(v^0)^T v^0} v^0 (v^0)^T$, it is easy to see that the resulting Household matrix $H_n = I - \frac{2}{(v^1)^T v^1} v^1 (v^1)^T$ is equal to the described matrix H.

Exercise 5.2

Use Householder matrices to transform the matrix A to tridiagonal form.

I am sorry to say that I ran out of time before being able to finish this exercise. While I do think it would be relatively straight forward to solve, I simply don't have the time to do this before handing in. I apologize, and hope the other 19 pages of hard work is enough to get this assignment approved!