Solution 8 – MATH-250 Advanced Numerical Analysis I

The exercise sheet is divided into two sections: quiz and exercises. The quiz will be discussed in the beginning of the lecture on Thursday, April 17. The exercises marked with ($\star$) are graded homework. The exercises marked with **(Python)** are implementation based and can be solved in the Jupyter notebooks which are available on Moodle/Noto. **The deadline for submitting your solutions to the homework is Friday, May 2 at 10h15.**

Quiz

a) If A is not invertible then A does not have an LU factorization (without pivoting).

☐ True ■ False

b) If A is invertible then Algorithm 4.13 in the lecture notes does not fail, that is, it always finds nonzero pivot elements $|a_{ik}|$ (in exact arithmetic) and produces an LU factorization with pivoting for A.

■ True ☐ False

c) The norm defined by $\|A\|_{\max} := \max_{ij} |a_{ij}|$ is a matrix norm but it is not submultiplactive.

■ True ☐ False

d) On the vector space of square symmetric matrices, $\text{trace}(A) = a_{11} + \cdots + a_{nn}$ is a matrix norm.

☐ True ■ False

e) Given a diagonal matrix $A = \text{diag}(a_{11}, \ldots, a_{nn})$, which of the following statements is *wrong*?

☐ $\|A\|_F = \|\mathbf{d}\|_2$ for the vector $\mathbf{d} = [a_{11}, a_{22}, \ldots, a_{nn}]$

☐ $\|A\|_2 = \|\mathbf{d}\|_\infty$

■ $\|A\|_1 = \|\mathbf{d}\|_1$

Solution.

(a) Choose $A = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$, clearly, A is not invertible, but $A = LU$ with

$$L = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad U = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

(b) We prove this by induction. For the first column, if there are only zero elements, then A is already singular. Let us now examine the inductive step $k \mapsto k+1$.

If the $k+1$-th step produces only zero elements in the pivot selection, then equivalently the matrix $\hat{A} = A[k+1:, k+1:]$ in Python index notation has to contain a leading column equal to zero. Thus, $\hat{A}$ has to be singular, which in turn implies that the original A was already singular. This concludes the inductive proof.

(c) Obviously, the norm $\|\cdot\|_{\max}$ fulfills all criteria to be a matrix norm. On the other hand, we can choose $A = B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ to see that $\|AB\|_{\max} = 2$, but $\|A\|_{\max} = \|B\|_{\max} = 1$.

(d) The trace does not fulfill the positivity requirement.

(e) We have that $\|A\|_1 = \max_j \sum_i |a_{ij}| = \max_j |a_{jj}|$ by definition, but $\|\mathbf{d}\|_1 = \sum_i |\mathbf{d}_i|$.

Exercises

Problem 1.

(a) Show that for $\mathbf{x} \in \mathbb{R}^n$

(i) $\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{n}\|\mathbf{x}\|_2$

(ii) $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_1 \leq n\|\mathbf{x}\|_\infty$

(iii) $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \sqrt{n}\|\mathbf{x}\|_\infty$

In addition, show that the bounds are tight.

(b) For $A \in \mathbb{R}^{n\times n}$ and $p \geq 1$, the matrix p-norm of A is defined as

$$\|A\|_p = \sup_{\mathbf{x}\neq 0} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p}$$

Show that

(i) $\frac{1}{\sqrt{n}}\|A\|_2 \leq \|A\|_1 \leq \sqrt{n}\|A\|_2$

(ii) $\frac{1}{\sqrt{n}}\|A\|_\infty \leq \|A\|_2 \leq \sqrt{n}\|A\|_\infty$

(c) For $A \in \mathbb{R}^{n\times n}$ the Frobenius norm of A is defined as

$$\|A\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2}$$

Show that $\|A\|_F = \sqrt{\operatorname{tr}(A^TA)} = \sqrt{\operatorname{tr}(AA^T)}$ where $\operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii}$.

(d) Show that for $A, B \in \mathbb{R}^{n\times n}$ and $p \geq 1$ we have

(i) $\|AB\|_p \leq \|A\|_p\|B\|_p$ using the definition of the matrix p-norm.

(ii) $\|AB\|_F \leq \|A\|_F\|B\|_F$

(e) Show that for $A \in \mathbb{R}^{n\times n}$

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

(ii) $\|A\|_\infty = \max_{1\le i\le n} \sum_{j=1}^{n} |a_{ij}|$

Solution.

(a) (i)

$$\begin{aligned}\|\mathbf{x}\|_2^2 &= \sum_{i=1}^{n} x_i^2 \le \sum_{i=1}^{n} x_i^2 + 2\sum_{i\ne j} |x_i||x_j| \\ &= \left(\sum_{i=1}^{n} |x_i|\right)^2 \\ &= \|\mathbf{x}\|_1^2\end{aligned}$$

Hence, $\|\mathbf{x}\|_2 \le \|\mathbf{x}\|_1$. Equality is achieved by letting $\mathbf{x} = \mathbf{e}_1$, where $\mathbf{e}_1$ is the first canonical vector.

Let $\mathbf{e}$ be the vector with all ones. Let $|\mathbf{x}|$ be the vector that results from taking the elementwise absolute value of $\mathbf{x}$. Then using the Cauchy-Schwarz inequality we get

$$\begin{aligned}\|\mathbf{x}\|_1 &= \langle |\mathbf{x}|, \mathbf{e}\rangle \\ &\le \||\mathbf{x}|\|_2 \|e\|_2 = \sqrt{n}\|\mathbf{x}\|_2\end{aligned}$$

Hence, $\|\mathbf{x}\|_1 \le \sqrt{n}\|\mathbf{x}\|_2$. Equality is achieved by letting $\mathbf{x} = \mathbf{e}$.

(ii) Suppose $|x_{j^*}| = \|\mathbf{x}\|_\infty$. Hence,

$$\|\mathbf{x}\|_\infty = |x_{j^*}| \le \sum_{i=1}^{n} |x_i| \le n|x_{j^*}| = n\|\mathbf{x}\|_\infty$$

If $\mathbf{x} = \mathbf{e}_1$ we have $\|\mathbf{x}\|_\infty = \|\mathbf{x}\|_1$. If $\mathbf{x} = \mathbf{e}$ we have $\|\mathbf{x}\|_1 = n\|\mathbf{x}\|_\infty$.

(iii) Suppose $|x_{j^*}| = \|\mathbf{x}\|_\infty$. Hence,

$$\|\mathbf{x}\|_\infty = |x_{j^*}| = \sqrt{x_{j^*}^2} \le \sqrt{\sum_{i=1}^{n} x_i^2} \le \sqrt{n x_{j^*}^2} = \sqrt{n}\|\mathbf{x}\|_\infty$$

If $\mathbf{x} = \mathbf{e}_1$ we have $\|\mathbf{x}\|_\infty = \|\mathbf{x}\|_2$. If $\mathbf{x} = \mathbf{e}$ we have $\|\mathbf{x}\|_1 = \sqrt{n}\|\mathbf{x}\|_\infty$.

(b) (i) Let $\mathbf{x} \in \mathbb{R}^n$

$$\begin{aligned}\frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2} &\le \frac{\|A\mathbf{x}\|_1}{\|\mathbf{x}\|_2} \\ &\le \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_1/\sqrt{n}} \\ &= \sqrt{n}\frac{\|A\mathbf{x}\|_1}{\|\mathbf{x}\|_1}\end{aligned}$$

Taking supremum and yields $\|A\|_2 \leq \sqrt{n}\|A\|_1 \Leftrightarrow \frac{1}{\sqrt{n}}\|A\|_2 \leq \|A\|_1$.

Similarly,

$$\begin{aligned}\frac{\|A\mathbf{x}\|_1}{\|\mathbf{x}\|_1} &\leq \sqrt{n}\frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_1} \\ &\leq \sqrt{n}\frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2}\end{aligned}$$

Taking supremum and yields $\|A\|_1 \leq \sqrt{n}\|A\|_2$.

(ii) Let $\mathbf{x} \in \mathbb{R}^n$

$$\begin{aligned}\frac{\|A\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty} &\leq \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_\infty} \\ &\leq \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2/\sqrt{n}} \\ &= \sqrt{n}\frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2}\end{aligned}$$

Taking supremum and yields $\|A\|_\infty \leq \sqrt{n}\|A\|_2 \Leftrightarrow \frac{1}{\sqrt{n}}\|A\|_\infty \leq \|A\|_2$.

Similarly,

$$\begin{aligned}\frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2} &\leq \sqrt{n}\frac{\|A\mathbf{x}\|_\infty}{\|\mathbf{x}\|_2} \\ &\leq \sqrt{n}\frac{\|A\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty}\end{aligned}$$

Taking supremum and yields $\|A\|_2 \leq \sqrt{n}\|A\|_\infty$.

(c) Note $(A^TA)_{ii} = \sum_{j=1}^{n} a_{ji}a_{ji} = \sum_{j=1}^{n} a_{ji}^2$

$$\text{trace}(A^TA) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ji}^2 = \|A\|_F^2$$

and $\|A\|_F^2 = \text{trace}(AA^T)$ follows from $\|A^T\|_F = \|A\|_F$.

(d) (i) One can see that $\forall \mathbf{x} \in \mathbb{R}^n$ we have $\|A\mathbf{x}\|_p \leq \|A\|_p\|\mathbf{x}\|_p$. Hence, if $\mathbf{x} \in \mathbb{R}^n$

$$\|AB\mathbf{x}\|_p = \|A(B\mathbf{x})\|_p \leq \|A\|_p\|B\mathbf{x}\|_p \leq \|A\|_p\|B\|_p\|\mathbf{x}\|_p$$

Hence,

$$\forall \mathbf{x} \in \mathbb{R}^n \quad \frac{\|AB\mathbf{x}\|_p}{\|\mathbf{x}\|_p} \leq \|A\|_p\|B\|_p$$

Taking supremum yields the result $\|AB\|_p \leq \|A\|_p\|B\|_p$.

(ii) Let $c_{ij} = (AB)_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$. Abusing the Matlab notation, let $A(i,:)$ denote the i:th row of A and $B(:,j)$ the j:th column of B. Then, $c_{ij} = \langle A(i,:), B(:,j)\rangle \leq \|A(i,:)\|_2 \|B(:,j)\|_2$ by the Cauchy-Schwarz inequality. Hence,

$$\begin{aligned}
\|AB\|_F^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}^2 \\
&\leq \sum_{i=1}^{n}\sum_{j=1}^{n} \|A(i,:)\|_2^2 \|B(:,j)\|_2^2 \\
&= \sum_{i=1}^{n} \|A(i,:)\|_2^2 \sum_{j=1}^{n} \|B(:,j)\|_2^2 \\
&= \|A\|_F^2 \|B\|_F^2
\end{aligned}$$

(e) (i) Again, we will abuse the Matlab notation to denote $A(:,j)$ to be the j:th column of A. Let $\mathbf{x} \in \mathbb{R}^n$. Then,

$$\|A\mathbf{x}\|_1 = \|\sum_{j=1}^{n} x_j A(:,j)\|_1 \leq \sum_{j=1}^{n} |x_j| \|A(:,j)\|_1 \leq \|\mathbf{x}\|_1 \max_{1\leq j\leq n} \|A(:,j)\|_1 \tag{1}$$

Hence,

$$\|A\|_1 \leq \max_{1\leq j\leq n} \|A(:,j)\|_1$$

Let $\|A(:,i)\|_2 = \max_{1\leq j\leq n} \|A(:,j)\|_1$. Then, letting $\mathbf{x} = \mathbf{e}_i$ will attain the bound in (1)

(ii) Let $\mathbf{x} \in \mathbb{R}^n$. Then,

$$\|Ax\|_\infty = \max_{1\leq i\leq n} \left|\sum_{j=1}^{n} a_{ij}x_j\right| \leq \|x\|_\infty \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}| \tag{2}$$

Hence,

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

Let k be such that $\sum_{j=1}^{n} |a_{kj}| = \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|$. Letting $\mathbf{x} \in \mathbb{R}^n$ be such that

$$\mathbf{x}_j = \begin{cases} 1 & \text{if } a_{kj} \geq 0 \\ -1 & \text{if } a_{kj} < 0 \end{cases}$$

then the upper bound in (2) is attained.

Problem 2. Consider the matrix $A \in \mathbb{R}^{10\times 10}$ and the vector $\mathbf{b} \in \mathbb{R}^{10}$ given below

$$A = \begin{pmatrix} 1 & 10 & & & 0 \\ & 1 & 10 & & \\ & & \ddots & \ddots & \\ & & & 1 & 10 \\ 0 & & & & 1 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} 9 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

(a) Solve $A\mathbf{x} = \mathbf{b}$ for $\mathbf{x}$ exactly.

(b) Solve the perturbed system $A\widehat{\mathbf{x}}^{(1)} = \mathbf{b} + \Delta\mathbf{b}$ where $\Delta\mathbf{b} = 10^{-8}\mathbf{e}_3$, where

$$\mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

(c) Compute the relative error

$$\frac{\|\mathbf{x} - \widehat{\mathbf{x}}^{(1)}\|_\infty}{\|\mathbf{x}\|_\infty}$$

(d) Now consider the perturbed system $(A + \Delta A)\widehat{\mathbf{x}}^{(2)} = \mathbf{b}$ where $\Delta A = \varepsilon I_{10}$. Using results about sensitivity of linear systems, what is the maximum value of ε so that the relative error

$$\frac{\|\mathbf{x} - \widehat{\mathbf{x}}^{(2)}\|_\infty}{\|\mathbf{x}\|_\infty}$$

is guaranteed to be less than 10^{-3}? Use Python to compute `np.linalg.cond(A, np.inf)` and/or `np.linalg.norm(np.linalg.inv(A), np.inf)`, if necessary.

(e) Let $\varepsilon = 10^{-6}$ and solve the system $(A + \Delta A)\widehat{\mathbf{x}}^{(2)} = \mathbf{b}$ in Python and compute the relative error.

Solution.

(a) One can either see directly or obtain the solution via backward substition that the solution is

$$\mathbf{x} = \begin{pmatrix} -1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

(b) Backward substitution gives the solution

$$\widehat{\mathbf{x}}_i^{(1)} = \begin{cases} -1 + 10^{-6} & i = 1 \\ 1 - 10^{-7} & i = 2 \\ 10^{-8} & i = 3 \\ 0 & i \geq 4 \end{cases}$$

(c) The relative error is given by $\frac{\|\mathbf{x} - \widehat{\mathbf{x}}^{(1)}\|_\infty}{\|\mathbf{x}\|_\infty}$. Hence, the relative error is

$$\frac{10^{-6}}{1} = 10^{-6}$$

(d) Using Theorem 4.23 we get that the relative error is is bounded as follows

$$\frac{\|\mathbf{x}-\widehat{\mathbf{x}}^{(2)}\|_\infty}{\|\mathbf{x}\|_\infty} \leq \frac{\kappa_\infty(A)}{1-\|A^{-1}\Delta A\|_\infty} \cdot \frac{\|\Delta A\|_\infty}{\|A\|_\infty}.$$

To guarantee a relative error less than 10^{-3}, we bound the right hand side by 10^{-3} and uses the special form of ΔA to obtain

$$\frac{\kappa_\infty(A)}{1-\varepsilon\|A^{-1}\|_\infty} \cdot \frac{\varepsilon}{\|A\|_\infty} \leq 10^{-3},$$

Now to obtain ε:

$$\varepsilon \leq \frac{10^{-3}}{(1+10^{-3})\|A^{-1}\|_\infty} \approx 0.9 \cdot 10^{-12}.$$

(e) Available in the Jupyter notebook `serie08-sol.ipynb` on Moodle.

Problem 3. A matrix A is strictly diagonally dominant by rows if

$$|a_{ii}| > \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}|, \quad i=1,\dots,n.$$

By using the Neumann series that you have seen in Proposition 4.22, show that a strictly diagonally dominant matrix is non-singular.

Hint: Without loss of generality, assume that the diagonal entries of A all equal 1 by a suitable scaling of the rows of A. Now recall what you know about some matrix norms, such as the operator norms induced by $\|\cdot\|_1$ and $\|\cdot\|_\infty$.

Solution. We first note that since A is strictly diagonally dominant by rows we must have $a_{ii} \neq 0 \quad \forall i = 1,\cdots,n$ because

$$|a_{ii}| > \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}| \geq 0, \quad i=1,\dots,n. \tag{3}$$

Hence, we may scale each row in A so that the diagonal entries are all 1. This is equivalent to

$$A \leftarrow DA, \quad D = \begin{pmatrix} \frac{1}{a_{11}} & & & \\ & \frac{1}{a_{22}} & & \\ & & \ddots & \\ & & & \frac{1}{a_{nn}} \end{pmatrix}$$

A non-zero scaling along the rows does not change whether A is non-singular or not. Now let T be such that $A = I_n - T$. Then, we know since A is strictly diagonally dominant that T is 0 along its diagonal and

$$\sum_{j=1}^{n} |t_{ij}| = \sum_{\substack{j=1 \\ j\neq i}}^{n} |t_{ij}| = \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}| < |a_{ii}| = 1$$

which implies

$$\|T\|_\infty = \max_{i=1,\cdots,n} \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}| < 1$$

Then, by the Neumann series

$$\begin{aligned}\|A^{-1}\|_\infty &= \|I_n - T\|_\infty \\ &\left\|\sum_{k=0}^{\infty} T^k\right\|_\infty \\ &\leq \sum_{k=0}^{\infty} \|T^k\|_\infty \\ &\leq \sum_{k=0}^{\infty} \|T\|_\infty^k < \infty\end{aligned}$$

because $\|T\|_\infty < 1$. Hence, $\|A^{-1}\|_\infty < \infty$ and therefore A is non-singular, as required.