Exercise 5.1 *(Computing)***.** Given

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ -3 & -6 & -11 & -15 \\ 0 & 4 & 3 & 2 \\ 2 & 0 & 4 & 9 \end{pmatrix}.$$

(a) Determine one (or the unique - see below -) LU decomposition of A with row permutations and partial pivoting. That is, determine permutation matrices P_i as well as matrices L and U, so that

$$PA = LU$$

holds for $P = P_1P_2P_3$. For the sake of simplicity, apply the partial pivoting strategy only in the case of a zero pivot element.

Suggested solution:

1. Step: Permutation vector: $p = (1)$

$$P_1A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ -3 & -6 & -11 & -15 \\ 0 & 4 & 3 & 2 \\ 2 & 0 & 4 & 9 \end{pmatrix}$$

Bem: With the column pivot strategy, the first row would be swapped with the row containing the element with the largest absolute value in the first column. That would be the second row here, since $|-3| > 2 > 1$. (As indicated in the task statement, we step will be skipped here.)

2. Step: (Elimination of the 1^{st} column with L_1.)

$$P_1A = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ -3 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 2 & 0 & 0 & 1 \end{pmatrix}}_{L_1^{-1}} \underbrace{\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & -2 & -3 \\ 0 & 4 & 3 & 2 \\ 0 & -4 & -2 & 1 \end{pmatrix}}_{L_1P_1A}$$

3. Step: (Determine P_2). Permutation vector $p = (1, 3)$, i.e. swap rows 2 and 3.

$$P_2P_1A = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -3 & 0 & 1 & 0 \\ 2 & 0 & 0 & 1 \end{pmatrix}}_{P_2L_1^{-1}P_2} \underbrace{\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 4 & 3 & 2 \\ 0 & 0 & -2 & -3 \\ 0 & -4 & -2 & 1 \end{pmatrix}}_{P_2L_1P_1}$$

4. Step: (Elimination of the 2^{nd} column with L_2.)

$$P_2P_1A = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -3 & 0 & 1 & 0 \\ 2 & -1 & 0 & 1 \end{pmatrix}}_{P_2L_1^{-1}P_2L_2^{-1}} \underbrace{\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 4 & 3 & 2 \\ 0 & 0 & -2 & -3 \\ 0 & 0 & 1 & 3 \end{pmatrix}}_{L_2P_2L_1P_1}$$

5. Step: (Determine P_3). Permutation vector $p = (1, 3, 3)$, i.e. "swap rows 3 and 3" (nothing to do here).

$$P_3 P_2 P_1 A = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -3 & 0 & 1 & 0 \\ 2 & -1 & 0 & 1 \end{pmatrix}}_{P_3 P_2 L_1^{-1} P_2 L_2^{-1} P_3} \underbrace{\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 4 & 3 & 2 \\ 0 & 0 & -2 & -3 \\ 0 & 0 & 1 & 3 \end{pmatrix}}_{P_3 L_2 P_2 L_1 P_1}$$

6. Step: (Elimination of the 3rd column with L_3.)

$$P_3 P_2 P_1 A = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -3 & 0 & 1 & 0 \\ 2 & -1 & -1/2 & 1 \end{pmatrix}}_{P_3 P_2 L_1^{-1} P_2 L_2^{-1} P_3 L_3^{-1}} \underbrace{\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 4 & 3 & 2 \\ 0 & 0 & -2 & -3 \\ 0 & 0 & 0 & 3/2 \end{pmatrix}}_{L_3 P_3 L_2 P_2 L_1 P_1}$$

Since $P_1 = P_3 = E$, it follows that $P = P_3 P_2 P_1 = P_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ and

$$PA = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 4 & 3 & 2 \\ -3 & -6 & -11 & -15 \\ 2 & 0 & 4 & 9 \end{pmatrix}.$$

Solution in short form: If one writes the entries of the Frobenius matrix in the place of the zeros, which arise during the elimination process, the following solution scheme results (cf. the detailed scheme above):

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ -3 & -6 & -11 & -15 \\ 0 & 4 & 3 & 2 \\ 2 & 0 & 4 & 9 \end{pmatrix} \overset{2.}{\rightsquigarrow} \begin{pmatrix} 1 & 2 & 3 & 4 \\ -3 & 0 & -2 & -3 \\ 0 & 4 & 3 & 2 \\ 2 & -4 & -2 & 1 \end{pmatrix} \overset{3.}{\rightsquigarrow} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 4 & 3 & 2 \\ -3 & 0 & -2 & -3 \\ 2 & -4 & -2 & 1 \end{pmatrix}$$

$$\overset{4.}{\rightsquigarrow} \begin{pmatrix} 2 & 1 & 1 & 1 \\ 0 & 4 & 3 & 2 \\ -3 & 0 & -2 & -3 \\ 2 & -1 & 1 & 3 \end{pmatrix} \overset{6.}{\rightsquigarrow} \begin{pmatrix} 2 & 1 & 1 & 1 \\ 0 & 3 & 3 & 3 \\ -3 & 0 & -2 & -3 \\ 2 & -1 & -1/2 & 3/2 \end{pmatrix}.$$

(b) For $b = (-6, 26, 1, -20)^T$ solve the system of linear equations $Ax = b$ using (a) and forward/backward substitution.

Suggested solution: We have

$$Ax = b \quad \Leftrightarrow \quad PAx = Pb \quad \overset{PA=LU}{\Leftrightarrow} \quad L\,\underbrace{Ux}_{=:y} = P\begin{pmatrix} -6 \\ 26 \\ 1 \\ -20 \end{pmatrix} = \begin{pmatrix} -6 \\ 1 \\ 26 \\ -20 \end{pmatrix}.$$

1. Step: Solve $Ly = Pb$ by forward substitution:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -3 & 0 & 1 & 0 \\ 2 & -1 & -1/2 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} -6 \\ 1 \\ 26 \\ -20 \end{pmatrix} \Leftrightarrow \begin{cases} y_1 = -6 \\ y_2 = 1 \\ y_3 = 26 + 3\cdot(-6) = 8 \\ y_4 = -20 - 2\cdot(-6) + 1 + \dfrac{1}{2}\cdot 8 = -3 \end{cases}$$

2. Step: Solve $Rx = y$ by backward substitution:

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 4 & 3 & 2 \\ 0 & 0 & -2 & -3 \\ 0 & 0 & 0 & 3/2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -6 \\ 1 \\ 8 \\ -3 \end{pmatrix} \Leftrightarrow \begin{cases} x_4 = -2 \\ x_3 = -\frac{1}{2}(8 + 3\cdot(-2)) = -1 \\ x_2 = \frac{1}{4}(1 - 3\cdot(-1) - 2\cdot(-2)) = 2 \\ x_1 = -6 - 2\cdot 2 - 3\cdot(-1) - 4\cdot(-2) = 1 \end{cases} \Leftrightarrow x = \begin{pmatrix} 1 \\ 2 \\ -1 \\ -2 \end{pmatrix}.$$

Checking that x really solves $Ax = b$:

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ -3 & -6 & -11 & -15 \\ 0 & 4 & 3 & 2 \\ 2 & 0 & 4 & 9 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ -1 \\ -2 \end{pmatrix} = \begin{pmatrix} -6 \\ 26 \\ 1 \\ -20 \end{pmatrix} \quad \checkmark$$

Exercise 5.2 *(Computing; Octave)***.** The so-called Hilbert matrices of the form

$$A = (a_{ij}) \in \mathbb{R}^{n\times n}, \quad a_{ij} = \frac{1}{i+j-1}$$

are ill-conditioned. Consider the system of linear equations (SLE) $Ax = b$, where $b = (b_j) \in \mathbb{R}^n$ with $b_j = \frac{1}{j+1}$. You can use the solution routine `A\b` included in Matlab/Octave for the solution of this SLE respectively.

(a) Determine a numerical solution to this SLE for matrix dimension $n = 5$: Once the solution x for unperturbed b as above and once $\tilde{x}$ with perturbation of b by assuming $b_1 = 0.51$ (2% perturbation in the first component). Calculate the relative error

$$\frac{\|\tilde{x} - x\|_2}{\|x\|_2}$$

of the solution $\tilde{x}$ of the perturbed system (measured in the Euclidean norm).

Suggested solution: The exact solution of $Ax = b$, where

$$A = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \frac{1}{8} \\ \frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \frac{1}{8} & \frac{1}{9} \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{3} \\ \frac{1}{4} \\ \frac{1}{5} \\ \frac{1}{6} \end{pmatrix},$$

is $e_2 \in \mathbb{R}^5$. Using the built-in routine `A\b` (or the LU decomposition of sheet 4) , we obtain for the solution of the unperturbed problem $Ax = b$:

$$\texttt{A\textbackslash b} = \begin{pmatrix} -0.000000000000000 \\ 1.000000000000004 \\ -0.000000000000011 \\ 0.000000000000007 \\ 0 \end{pmatrix} \quad \text{and with LU decomposition:} \quad x_{\text{calc}} = \begin{pmatrix} 1.260103132949553e{-14} \\ 9.999999999997596e{-01} \\ 1.049438314025922e{-12} \\ -1.598351081117220e{-12} \\ 7.868705687023838e{-13} \end{pmatrix}.$$

Using the built-in routine `A\b` (or the LU decomposition of sheet 4) , we obtain for the solution of the perturbed problem $A\tilde{x} = \tilde{b}$:

$$\texttt{A\textbackslash btilde} = \begin{pmatrix} 2.499999999992926e{-01} \\ -1.999999999987002e{+00} \\ 1.049999999994466e{+01} \\ -1.399999999991723e{+01} \\ 6.299999999959805e{+00} \end{pmatrix} \quad \text{and with LU decomp.:} \quad \tilde{x}_{\text{calc}} = \begin{pmatrix} 2.499999999998515e{-01} \\ -1.999999999997368e{+00} \\ 1.049999999998907e{+01} \\ -1.399999999998394e{+01} \\ 6.299999999992306e{+00} \end{pmatrix}.$$

From this we obtain for the relative error

$$\texttt{norm(xtilde-x)/norm(x)=18.842} \quad \text{i.e.} \quad \frac{\|\tilde{x}-x\|_2}{\|x\|_2} = 18.842$$

and for the relative error in the input data

$$\texttt{norm(btilde-b)/norm(b)=0.014266} \quad \text{i.e.} \quad \frac{\|\tilde{b}-b\|_2}{\|b\|_2} = 0.014266.$$

That is, we have an error amplification by a factor of $\frac{18.842}{0.014266} \approx 1321$. Note that $\mathrm{cond}_2(A) = 4.77 \cdot 10^5$, i.e. A is ill-conditioned.

(b) Calculate the solution x_{num} of $Ax = b$ with b as above (without perturbation) for matrix dimension $n = 20$. Compare x_{num} with the exact solution $x = (0, 1, 0, \ldots, 0)^T$ and calculate $\|x_{\mathrm{num}} - x\|_2$ and $\|x_{\mathrm{num}} - x\|_\infty$. Determine with Matlab/Octave also the condition $\mathrm{cond}_2(A)$ of the matrix A.

Suggested Solution: For Octave's built-in routine `A\b`, the result is (split into 5 columns):

$$\texttt{A\textbackslash b} = x_{\mathrm{num}} \begin{pmatrix} x_1 \cdots\cdots x_5 \\ \vdots \\ x_{16} \cdots\cdots x_{20} \end{pmatrix} = \begin{pmatrix} 5.4785e{-08} & 9.9999e{-01} & 3.1815e{-04} & -5.2499e{-03} & 4.6631e{-02} \\ -2.4818e{-01} & 8.3848e{-01} & -1.8557e{+00} & 2.7470e{+00} & -2.7845e{+00} \\ 1.8807e{+00} & -2.9324e{-01} & -1.2043e{+00} & 6.0960e{-01} & 1.9387e{+00} \\ -2.1990e{+00} & -1.0100e{+00} & 3.0568e{+00} & -1.9395e{+00} & 4.2129e{-01} \end{pmatrix}$$

For the relative errors for the numerically determined solution we obtain

$$\|x_{\mathrm{num}} - x\|_2 = 6.9175 \quad \text{and} \quad \|x_{\mathrm{num}} - x\|_\infty = 3.0568.$$

Using the (self-made) LU decomposition `[L,U]=ludecomposition(A)` we obtain

$$\texttt{U\textbackslash(L\textbackslash b)} = x_{\mathrm{lu}} = \begin{pmatrix} -6.5865e{-08} & 1.0000e{+00} & -3.8392e{-04} & 6.1669e{-03} & -5.1614e{-02} \\ 2.4451e{-01} & -6.5453e{-01} & 8.4378e{-01} & 3.3464e{-03} & -1.0253e{+00} \\ -7.1362e{-01} & 4.7370e{+00} & -5.5756e{+00} & 2.2417e{+00} & 4.5524e{-01} \\ -2.6226e{+00} & 6.2721e{+00} & -7.3153e{+00} & 3.9936e{+00} & -8.3841e{-01} \end{pmatrix}$$

and therefore

$$\|x_{\mathrm{lu}} - x\|_2 = 13.338 \quad \text{and} \quad \|x_{\mathrm{lu}} - x\|_\infty = 7.3153.$$

Note that the condition number of A (wrt the 2-norm) is is unacceptably high:

```
cond(A)=6.8070e+18.
```

Exercise 5.3 *(Eigenvalues and matrix norms)***.** Given $A \in \mathbb{R}^{n\times n}$. Show that

(a) $\sigma(A^2) = \{\lambda^2 \mid \lambda \in \sigma(A)\}$

Suggested Solution: For $\mu \in \mathbb{C}$ let $\pm\lambda \in \mathbb{C}$ be the square roots of μ. By the multiplication theorem for determinants it follows that

$$|A - \lambda E| \cdot |A + \lambda E| = |A^2 - \lambda^2 E| = |A^2 - \mu E|.$$

From this we can deduce:

$$\begin{aligned} & \mu \in \sigma(A^2) \\ \Leftrightarrow \quad & \lambda \in \sigma(A) \text{ or } -\lambda \in \sigma(A) \end{aligned} \tag{1}$$

Equation (1) implies that

$$\mu = \lambda^2 = (-\lambda)^2 \in \{\lambda^2 \mid \lambda \in \sigma(A)\}.$$

Conversely, $\mu \in \{\lambda^2 \mid \lambda \in \sigma(A)\}$ means that one of the two roots $\pm\lambda$ of μ is an eigenvalue of A, i.e. (1) holds true.

(b) $\|A\|_F \leq \sqrt{n}\|A\|_2$. Is this inequality sharp?

Suggested Solution: Let $j_0 \in \arg\max_j \sum_i a_{ij}^2$. Then

$$\|A\|_F^2 = \sum_j \sum_i a_{ij}^2 \leq n \sum_i a_{ij_0}^2 = n\|Ae_{j_0}\|_2^2$$

implies

$$\|A\|_F \leq \sqrt{n}\|Ae_{j_0}\|_2 \leq \sqrt{n} \max_{\|x\|_2=1} \|Ax\|_2 = \sqrt{n}\|A\|_2.$$

Exercise 5.4 *(Eigenvalues, condition numbers; computation in (c))*. Let $A \in \mathbb{R}^{n\times n}$ be symmetric and $\lambda_{\max}(A)$ be the largest and $\lambda_{\min}(A)$ be the smallest absolute eigenvalue of A.

(a) If A is regular and $\sigma(A) = \{\lambda_1, \ldots, \lambda_n\}$ (where the numbers λ_i do not necessarily have to be different) show that the spectrum of A^{-1} is given by

$$\sigma(A^{-1}) = \left\{\frac{1}{\lambda_1}, \ldots, \frac{1}{\lambda_n}\right\}.$$

What is the relationship between the respective eigenvectors of A and A^{-1}?

Suggested Solution: By multiplication with A^{-1} from the left it follows

$$Av = \lambda v \quad \Leftrightarrow \quad A^{-1}v = \frac{1}{\lambda}v.$$

(b) Let A be regular and symmetric. Show that

$$\operatorname{cond}_2(A) = \frac{|\lambda_{\max}(A)|}{|\lambda_{\min}(A)|}.$$

Suggested Solution: Since

$$\operatorname{cond}_2(A) = \|A\|_2\|A^{-1}\|_2$$

we show that

$$\|A\|_2 = |\lambda_{\max}(A)| \quad \text{and} \quad \|A^{-1}\|_2 = |\lambda_{\min}(A)|^{-1}.$$

<u>Step 1:</u> From

$$\begin{aligned} & |\lambda_{\max}(A)| \geq |\lambda| && \forall \lambda \in \sigma(A) \\ \Leftrightarrow \quad & |\lambda_{\max}(A)^2| \geq |\lambda^2| && \forall \lambda \in \sigma(A) \\ \Leftrightarrow \quad & |\underbrace{\lambda_{\max}(A)^2}_{\in\sigma(A^2)}| \geq |\mu| && \forall \mu \in \sigma(A^2) \end{aligned}$$

it follows that

$$\boxed{\lambda_{\max}(A^2) = \lambda_{\max}(A)^2.} \tag{2}$$

<u>Step 2:</u> Using (2) we obtain for all symmetric $A \in \mathbb{R}^{n\times n}$

$$\|A\|_2^2 = \rho(A^TA) = \rho(A^2) = \lambda_{\max}(A^2) = \lambda_{\max}(A)^2, \tag{3}$$

i.e. the first part of the assertion.

<u>Step 3:</u> For all regular $A \in \mathbb{R}^{n\times n}$ it follows from

$$|\lambda_{\min}(A)| \leq |\lambda| \qquad \forall \lambda \in \sigma(A)$$

$$\begin{aligned}
\Leftrightarrow \qquad & \left|\frac{1}{\lambda}\right| \leq \left|\frac{1}{\lambda_{\min}(A)}\right| \quad \forall \lambda \in \sigma(A) \\
\Leftrightarrow \qquad & |\mu| \leq \Bigg|\underbrace{\frac{1}{\lambda_{\min}(A)}}_{\in\sigma(A^{-1})}\Bigg| \quad \forall \mu \in \sigma(A^{-1})
\end{aligned}$$

that

$$\boxed{\lambda_{\max}(A^{-1}) = \lambda_{\min}^{-1}(A).} \tag{4}$$

Step 4: If A is regular and symmetric, so is A^{-1}:

$$\left(A^{-1}\right)^T = \left(A^T\right)^{-1} = A^{-1}.$$

Thus we may substite A for A^{-1} in (3) to get

$$\|A^{-1}\|_2^2 \overset{(3)}{=} \lambda_{\max}(A^{-1})^2 \overset{(4)}{=} \lambda_{\min}^{-2}(A),$$

proving the second part of the assertion.

(c) Determine the value $\operatorname{cond}_2(A)$ of the matrix $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 4 \\ 0 & 4 & 2 \end{pmatrix}$.

We determine the eigenvalues of A:

$$\begin{aligned}
\chi_A(\lambda) &= \det(A - \lambda E) \\
&= (1-\lambda)\det\begin{pmatrix} 2-\lambda & 4 \\ 4 & 2-\lambda \end{pmatrix} \\
&= (1-\lambda)[(2-\lambda)^2 - 16] \\
&= (1-\lambda)[4 - 4\lambda + \lambda^2 - 16] \\
&= (1-\lambda)[\lambda^2 - 4\lambda - 12] \\
&\Rightarrow \lambda_1 = 1 \qquad \text{und} \qquad \lambda_{2,3} = \frac{4 \pm \sqrt{16+48}}{2} = \frac{4 \pm 8}{2} \\
&\Rightarrow \lambda_1 = 1 \quad , \quad \lambda_2 = 6 \quad , \quad \lambda_3 = -2 \\
&\Rightarrow \operatorname{cond}_2(A) = \frac{|\lambda_{\max}|}{|\lambda_{\min}|} = \frac{6}{1} = 6.
\end{aligned}$$