Solutions to Problems of Numerical Methods I.

Csaba J. Hegedüs

ELTE, Faculty of Informatics
Budapest, 2015 November

A jegyzet az ELTE Informatikai Kar 2015 évi Jegyzetpályázatának támogatásával készült
Lektorálta: Dr. Abaffy József, Óbudai Egyetem

Problems in Chapter 1

1.1. Let the set of machine numbers be $M(5,-4,4)$. Identify the special machine numbers! Map the following numbers: 1/50, 0.37, 3.67, 7.2, 21.78 into this set!

$\varepsilon_0 = 1/2 \cdot 2^{k^-} = 2^{-5} = 1/32 = 0.03125$, $\quad M_\infty = (1-2^{-t})2^{k^+} = (1-2^{-5})2^4 = 16 - 1/2 = 15.5$,

$\varepsilon_M = 2^{-t} = 2^{-5} = 1/32 = 0.03125$,

$\mathrm{fl}(1/50) = \mathrm{fl}(0.02) = .00000 \cdot 2^0$, $1/50 < \varepsilon_0$, $\mathrm{fl}(0.37) = .10100 \cdot 2^{-1}$, $\mathrm{fl}(3.67) = .11101 \cdot 2^2$,

$\mathrm{fl}(21.78) = \infty$, because $21.78 > M_\infty$.

1.2. How should we convert 10.87 into a ternary number of base 3?

Division by 3 for the integer part and multiplication by 3 for the fraction part:

$10_3 = 101$ and $0.87_3 = .2121\ldots$

1.3. How the machine epsilon is modified, if chopping is applied instead of rounding?

It will be twice as much because now the last digit is uncertain.

Problems in Chapter 2

2.1. Show that for all induced norm $\|I\| = 1$ holds. May the Frobenius norm be an induced norm?

Apply definition: $\max_{\|x\|=1} \|Ix\| = 1$. If $I \in \mathbb{R}^{n,n}$ then $\|I\|_F = n$ such that it may not be an induced norm.

2.2. If A is invertible then $\|x\|_A = \|Ax\|$ is also a vector norm.

We have to check norm conditions: 1) $\|x\|_A = \|Ax\| = 0$ only if $\|x\| = 0$. It is needed here that $Ax = 0$ only if $x = 0$. If A has an inverse, the only solution for the first equation is $A^{-1}Ax = x = 0$. 2) $\|\lambda x\|_A = \|\lambda Ax\| = |\lambda| \|Ax\|$. 3) The triangle inequality is also inherited from the first vector norm: $\|A(x+y)\| \le \|Ax\| + \|Ay\|$.

2.3. A matrix condition number may not be less then 1 for induced norms.

$1 = \|I\| = \|AA^{-1}\| \le \|A\| \|A^{-1}\| = \mathrm{cond}(A)$.

2.4. Using the 2-norm, the condition number of orthogonal or unitary matrices is equal to 1.

For unitary matrices $U^H U = 1$ such that $\lambda_{\max}(U^H U) = \lambda_{\max}(I) = 1$. The case for orthogonal matrices is similar.

2.5. Prove: $\left\|ab^T\right\|_1 = \|a\|_1 \|b\|_\infty$. $\left\|ab^T\right\|_\infty = \|a\|_\infty \|b\|_1$. $\left\|ab^T\right\|_2 = \|a\|_2 \|b\|_2$.

The 1-norm is the column norm: $\left\|ab^T\right\|_1 = \max_j \|a\|_1 |b_j| = \|a\|_1 \|b\|_\infty$. The ∞-norm is the row norm: $\left\|ab^T\right\|_\infty = \max_i |a_i| \|b\|_1 = \|a\|_\infty \|b\|_1$.

First solution for the 2-norm: $\left\|ab^T\right\|_2 = \max_{\|x\|_2=1} \left\|ab^T x\right\| = \|a\|_2 \max_{\|x\|_2=1} \left\|b^T x\right\|$ and it is maximal if vectors x and b are in the same direction, then $x = b/\|b\|_2$. In that case the result is $\|a\|_2 \|b\|_2$.

Second solution. We exploit the fact that the nonzero eigenvalues of bb^T and $b^T b$ are equal: $\left\|ab^T\right\|_2^2 = \rho(ba^T ab^T) = \|a\|_2^2 \rho(bb^T) = \|a\|_2^2 \rho(b^T b) = \|a\|_2^2 \|b\|_2^2$, where ρ is the spektral radius.

2.6. $U^T U = I$ (orthogonal) $\rightarrow \|AU\|_2 = \|A\|_2$.

$\|AU\|_2^2 = \rho\left(U^T A^T AU\right) = \rho\left(A^T AUU^T\right) = \rho\left(A^T A\right) = \|A\|_2^2$, because the eigenvalues are unchanged if interchanging matrices in a product, see the Remark after the spectral norm in the text.

2.7. $\big|\|A\| - \|B\|\big| \le \|A \pm B\|$.

$\|A + B - B\| \le \|A + B\| + \|B\| \quad \rightarrow \|A\| - \|B\| \le \|A + B\|$ Interchanging A and B gives $\|B\| - \|A\| \le \|A + B\|$. Combining the two gives the statement for $A + B$. Changing B to $-B$ gives the result for $A - B$.

2.8. $A = \begin{bmatrix} 2 & -3 & 1 \\ -4 & -2 & 1 \end{bmatrix}$, $\|A\|_1 = ?$ $\|A\|_\infty = ?$ $\|A\|_2 = ?$

$\|A\|_1 = \max\{6, 5, 2\} = 6$. $\|A\|_2 = \max\{6, 7\} = 7$.

$$\|A\|_2 = \lambda_{\max}^{1/2}(A^T A) = \lambda_{\max}^{1/2}(AA^T) = \lambda_{\max}^{1/2}\left(\begin{bmatrix} 14 & 1 \\ 1 & 21 \end{bmatrix}\right) = \left((35 + \sqrt{53})/2\right)^{1/2} .$$

2.9. $\|A\|_2 \le \sqrt{\|A\|_1 \|A\|_\infty}$.

The 2-norm can be given by the spectral radius of $A^T A$ and using the Theorem on the spectral radius, it is less than any norm of $A^T A$:

$\|A\|_2^2 = \rho(A^T A)\{=\text{spectral radius}\} \le \left\|A^T A\right\|_\infty \le \|A\|_1 \|A\|_\infty$.

2.10. Check the inequality $\|Ax\|_2 \le \|A\|_F \|x\|_2$. (It is consistent with the 2-norm.)

Observe that the Frobenius norm of vectors – as special matrices – is equal to the 2-norm, therefore it is the consequence of the fourth norm condition for matrix norms.

2.11. If $A = A^T$ then $\|A\|_2 = \rho(A)$ = spectral radius, that is, the spectral norm is the minimal norm for symmetric matrices. ($\|\,.\,\|_2 =$ spectral norm.)

$\|A\|_2^2 = \rho(A^T A)\{=\text{spectral radius}\} = \rho(A^2) = \lambda_{\max}(A^2) = \lambda_{\max}^2(A)$. Absolute sign is not needed, because the eigenvalues are squared.

2.12. If $A = A^T$ then $\|A\|_2 \le \|A\|_p$, $p = 1, \infty$.

It is the consequence of the previous problem. For symmetric matrices the 2-norm is minimal.

2.13. $U^T U = I$ (orthogonal) $\rightarrow \|AU\|_F = \|A\|_F$.

We use the identity: $\text{trace}(AB) = \text{trace}(BA)$:

$$\|AU\|_F^2 = \text{trace}\|U^T A^T AU\|_F = \text{trace}\|A^T AUU^T\|_F = \text{trace}\|A^T A\|_F .$$

2.14. $\|AB\|_2 = \|BA\|_2$ if $A = A^T$ and $B = B^T$.

See the Remark after the spectral norm in the text. It follows that matrices can be cyclically permuted if there are more matrices in a product such that the eigenvalue is unchanged:

$$\|AB\|_2^2 = \lambda_{\max}\left(B^T A^T AB\right) = \lambda_{\max}\left(A^2 B^2\right)$$
$$\|BA\|_2^2 = \lambda_{\max}\left(A^T B^T BA\right) = \lambda_{\max}\left(A^2 B^2\right)$$

2.15. $\text{cond}_2(A^T A) = \text{cond}_2^2(A)$.

$$\begin{aligned}\text{cond}_2(A^T A) &= \lambda_{\max}^{1/2}\left(A^T AA^T A\right)\lambda_{\max}^{1/2}\left(A^{-1}A^{-T}A^{-1}A^{-T}\right)\\ &= \lambda_{\max}\left(A^T A\right)\lambda_{\max}\left(A^{-1}A^{-T}\right) = \|A\|_2^2\|A^{-1}\|_2^2\\ &= \text{cond}_2^2(A)\end{aligned}$$

2.16. $\|PA\|_p = \|A\|_p = \|AP\|_p$, $p = 1, 2, \infty$, where P is a permutation matrix.

Matrix P is orthogonal such that the result comes from Problem 2.6 for $p = 2$. For the column norm ($p = 1$) there is no change in $\|AP\|_1$ because only the columns are permuted. In

$\|PA\|_1$ rows are permuted that can not change the 1-norm of a column, therefore the column of maximal norm should be the same. Similar argument applies for the infinity norm.

Problems in Chapter 3

3.1. Perform a dyadic multiplication with two vectors. Explain that it should have rank 1. Which method is simpler to multiply by a dyad? *a)* Form $A = ab^T$ then compute Ax. *b)* Compute $b^T x$ first and then multiply vector a with that scalar.
Every column of the resulting matrix is a scalar multiple of a therefore the columns are linearly dependent. *a)* Forming $A = ab^T$ needs n^2 flops, further computing Ax needs $2n^2$ additional flops, the operation count is $3n^2$ flops. *b)* Computing $b^T x$ needs $2n-1$ flops and multiplying vector a with a scalar still requires n flops, altogether $3n-1$ flops. Clearly, this latter one is simpler.

3.2. Consider the permutation matrix $\Pi = [e_2, e_4, e_3, e_1]$. Check that its transpose gives the inverse. Prove this fact in general! How can we store this permutation matrix in a vector? Check

$$\begin{bmatrix} e_2^T \\ e_4^T \\ e_3^T \\ e_1^T \end{bmatrix} [e_2, e_4, e_3, e_1] = I_4 \ ,$$

where I_4 is the identity of order 4. The general case is similar:

$$\begin{bmatrix} e_{i_1}^T \\ e_{i_2}^T \\ \vdots \\ e_{i_n}^T \end{bmatrix} [e_{i_1}, e_{i_1}, \ldots, e_{i_n}] = I_n$$

For storing a permutation matrix in a vector, chose a vector $a = \begin{bmatrix} 1 & 2 & \ldots & n \end{bmatrix}$ and store Πa.

3.3. Check: $F^{-1}a = e_i$ of (3.3).

$$F^{-1}a = a - \frac{(a - e_i)e_i^T a}{e_i^T a} = a - a + e_i = e_i \ .$$

3.4. With the aid of formula (3.5), show that the determinant of a matrix will not change if a scalar multiple of a column is added to another column of the matrix. Apply the theorem on the determinant of the product of two matrices!
Take the determinant in (3.5): $\left|A(I + \alpha e_i e_k^T)\right| = |A|$ because the second multiplier is a special triangular matrix having 1's in the diagonal such that its determinant is 1.

3.5. Form the rank-1 sum of ADB, where $D = [d_i \delta_{ij}]$ is a diagonal matrix, (only the diagonal elements are nonzero).

$$ADB = \sum_{i=1}^{n} Ae_i e_i^T DB = \sum_{i=1}^{n} Ae_i e_i^T d_i B = \sum_{i=1}^{n} d_i \cdot Ae_i e_i^T B .$$

3.6. Applying the scalar product and the dyadic product forms of matrix multiplication, show that $\mathrm{tr}(AB) = \mathrm{tr}(BA),\ A, B^T \in \mathbb{R}^{m,n}$!
Observe that for vectors a, b $tr(ab^T) = b^T a$ holds. Then

$$tr\left(AB\right) = \sum_{(i)} tr\left(Ae_i e_i^T B\right) = \sum_{(i)} tr\left(e_i^T BAe_i\right) = tr\left(BA\right).$$

3.7. Let matrix A be invertible. Give the expansion of vector x in terms of the columns of A.

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

3.8. Collect the vectors of a biorthogonal system into matrices $A = [a_1, a_2, \ldots, a_n]$ and $B = [b_1, b_2, \ldots, b_n]$. Check that $A^T B$ is a diagonal matrix! How can we give vector x as a linear combination of vectors a_i ? And how can we give the expansion with the aid of vectors b_i ?
Applying definition of biorthogonality gives $A^T B = D = \left[a_i^T b_j\right] = \left[\alpha_i \delta_{ij}\right] = D$. Therefore the inverse of A is $D^{-1}B^T$ and $x = A(D^{-1}B^T x)$. The inverse of B is $D^{-1}A^T$, hence the second answer is: $x = B\left(D^{-1}A^T x\right)$.

3.9. Check: if P is a projection, then $I - P$ is also a projection.

$$\left(I - P\right)\left(I - P\right) = I - 2P + P^2 = I - P .$$

3.10. A plane has normal vector s and its defining equation is $s^T x = \sigma$. Introduce the projection $P = I - ss^T / s^T s$. Show that for all vectors y the operation $Py + \sigma s / s^T s$ produces a vector in the plane.
It is enough to check the statement:

$$s^T Py + \sigma s / s^T s = s^T \left(I - ss^T / s^T s\right) y + \sigma s^T s / s^T s = 0 + \sigma .$$

3.11. Show with the previous matrix P : $Py \perp s$, in other words Py is perpendicular to s . Give the vector that connects $Py + \sigma s / s^T s$ and y !
$s^T P = 0$ therefore the statement follows for any y . The connecting vector is parallel to the normal vector:

$$y - \left(Py + \sigma s / s^T s\right) = y - y + \sigma ss^T y / s^T s = \sigma s \frac{s^T y}{s^T s} .$$

3.12. Show that the *backward identity* $J = [e_n e_{n-1} \ldots e_1]$, where the columns of the unit matrix are given in reverse order, is involutory. What projection will it define for $n = 2, 3$?

In fact, J is a symmetric permutation matrix, therefore it is involutory. The projections defined are:

$$\left(\begin{bmatrix}1 & \\ & 1\end{bmatrix}-\begin{bmatrix} & 1\\ 1 & \end{bmatrix}\right)\frac{1}{2}=\begin{bmatrix}1/2 & -1/2\\ -1/2 & 1/2\end{bmatrix} \text{ and } \begin{bmatrix}1/2 & & -1/2\\ & 0 & \\ -1/2 & & 1/2\end{bmatrix}.$$

3.13. Show that matrix $I-2(x-y)(x-y)^T/(x-y)^T(x-y)$ will reflect vectors x and y into each other, if they are different and have the same length: $x^Tx=y^Ty$.
Assume $x^Tx=y^Ty=\sigma^2$. Then

$$\left(I-2\frac{(x-y)(x-y)^T}{(x-y)^T(x-y)}\right)x=x-2(x-y)\frac{(x-y)^Tx}{2\sigma^2-2y^Tx}=x-(x-y)\frac{\sigma^2-y^Tx}{\sigma^2-y^Tx}=y.$$

The other reflection can be checked similarly.

3.14. We have the possibility to reflect vector x with the previous matrix into vector $y=\pm\sigma e_1$, where $\sigma^2=x^Tx$. How should we choose the sign of σ to avoid cancellation error in the denominator?
Assume $y=\sigma e_1$ such that the sign is attached to sigma. From the previous problem the denominator now is $2\sigma(\sigma-e_i^Tx)$. There will be no cancellation for $\text{sign}(\sigma)=-\text{sign}(e_i^Tx)$.

3.15. Introduce $F=I+UV^T$, where the unit matrix is modified by the $n\times l$ matrices U and V, that is, they have $l<n$ columns. If F is invertible, show that $F^{-1}=I-U(I_l+V^TU)^{-1}V^T$ (*Sherman-Morrison-Woodbury formula*) holds, where I_l is a unit matrix of size $l\times l$.
It is enough to check:

$$\begin{aligned}FF^{-1}&=\left(I+UV^T\right)\left(I-U(I_l+V^TU)^{-1}V^T\right)\\&=I+UV^T-U(I_l+V^TU)^{-1}V^T-UV^TU(I_l+V^TU)^{-1}V^T\\&=I+U\left(I-(I_l+V^TU)^{-1}-V^TU(I_l+V^TU)^{-1}\right)V^T\\&=I+U\left(I-\left(I_l+V^TU\right)(I_l+V^TU)^{-1}\right)V^T=I\end{aligned}$$

Problems in Chapter 4

4.1. Using *LU*-decomposition, solve the following linear system:

$$\begin{bmatrix}2 & 2 & 3\\ 4 & 3 & 7\\ 6 & 7 & 5\end{bmatrix}x=\begin{bmatrix}1\\ 5\\ -3\end{bmatrix}.$$

$$\begin{bmatrix} 2 & 2 & 3 & 1 \\ 4 & 3 & 7 & 5 \\ 6 & 7 & 5 & -3 \end{bmatrix} \to \begin{bmatrix} \boxed{2} & 2 & 3 & 1 \\ 2 & -1 & 1 & 3 \\ 3 & 1 & -4 & -6 \end{bmatrix} \to \begin{bmatrix} 2 & 2 & 3 & 1 \\ 2 & \boxed{-1} & 1 & 3 \\ 3 & -1 & -3 & -3 \end{bmatrix},$$

$$L = \begin{bmatrix} 1 & & \\ 2 & 1 & \\ 3 & -1 & 1 \end{bmatrix}, \quad U = \begin{bmatrix} 2 & 2 & 3 \\ & -1 & 1 \\ & & -3 \end{bmatrix}, \quad x = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

4.2. Find the operation count for Ax, LUx, $U^{-1}L^{-1}x$. The factorizations given in Section 3.11 may be applied for the last case.
All of them needs $2n^2$ flops, the last one requires still n divisions.

4.3. Using Problem 3.15, show that the matrix of (4.6) can be inverted by taking the negative of the block 21. Similar result for the upper triangular case can be found by transposition.
We look for the inverse of the matrix in (4.6): $L^{-1} = I - \left(AE_1A_{11}^{-1} - E_1\right)E_1^T = I - \begin{pmatrix} 0 \\ A_{21}A^{-1} \end{pmatrix} E_1^T$ such that using the Sherman-Morrison-Woodbury formula. The matrix in the middle will be a unit matrix because of $E_1^T \begin{pmatrix} 0 \\ A_{21}A^{-1} \end{pmatrix} = 0$ and it remains only to change sign.

4.4. Let L_{11} be a lower triangular matrix, which is complemented by a block row $\begin{bmatrix} L_{21} & L_{22} \end{bmatrix}$ to a larger lower triangular matrix. Assuming that the diagonal blocks are invertible, apply the partitioned inverse to get

$$\begin{bmatrix} L_{11} & \\ L_{21} & L_{22} \end{bmatrix}^{-1} = \begin{bmatrix} L_{11}^{-1} & \\ -L_{22}^{-1}L_{21}L_{11}^{-1} & L_{22}^{-1} \end{bmatrix}.$$

Applying the partitioned inverse formula (4.9):

$$L^{-1} = \begin{bmatrix} L_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ I_2 \end{bmatrix} \left(L \middle| L_{11}\right)^{-1} \begin{bmatrix} -L_{21}L_{11}^{-1} & I_2 \end{bmatrix}$$

$$= \begin{bmatrix} L_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ -L_{22}^{-1}L_{21}L_{11}^{-1} & L_{22}^{-1} \end{bmatrix}$$

4.5. By using the block partitioned form, check the determinant identity $|A| = |A_{11}|\left|\left(A \middle| A_{11}\right)\right|$.
It can be seen from the block decomposition formula:

$$A = \begin{bmatrix} I_1 & 0 \\ A_{21}A_{11}^{-1} & I_2 \end{bmatrix} \begin{bmatrix} A_{11} & \\ & \left(A \middle| A_{11}\right) \end{bmatrix} \begin{bmatrix} I_1 & A_{11}^{-1}A_{12} \\ 0 & I_2 \end{bmatrix},$$

4.6. With the aid of the previous problem, check the identities

$$\left\| \begin{bmatrix} 1 & -b^T \\ a & I \end{bmatrix} \right\| = 1 + b^T a = \left| I + ab^T \right|.$$

Compare it with the approach given in *Example E3.3*!
We use the formula of Problem 4.5. Chosing for A_{11} the left upper element gives the last determinant. Taking the Schur complement $(A \mid A_{22})$, where A_{22} is the identity matrix will lead to the formula in the middle. As seen, this approach is much simpler.

4.7. What is the dominant term in the operation count of the Gauss-Jordan factorization?
We have to do roughly $2n(n-k)$ flops in the k-th step, see Theorem T3.1. Summing up for $k = 1, \ldots, n-1$ yields for the dominant term n^3 flops.

4.8. $A = \begin{bmatrix} 2 & -2 & 1 & 0 \\ 4 & -1 & 3 & -1 \\ -2 & -1 & 0 & 2 \\ 6 & -3 & 2 & 2 \end{bmatrix}$, $b = \begin{bmatrix} -3 \\ -2 \\ -2 \\ 0 \end{bmatrix}$, $A = LU$, $Ax = b$. $L = ?$, $U = ?$ $x = ?$

$$\begin{bmatrix} 2 & -2 & 1 & 0 & -3 \\ 4 & -1 & 3 & -1 & -2 \\ -2 & -1 & 0 & 2 & -2 \\ 6 & -3 & 2 & 2 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} \boxed{2} & -2 & 1 & 0 & -3 \\ 2 & 3 & 1 & -1 & 4 \\ -1 & -3 & 1 & 2 & -5 \\ 3 & 3 & -1 & 2 & 9 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & -2 & 1 & 0 & -3 \\ 2 & \boxed{3} & 1 & -1 & 4 \\ -1 & -1 & 2 & 1 & -1 \\ 3 & 1 & -2 & 3 & 5 \end{bmatrix} \rightarrow$$

$$\rightarrow \begin{bmatrix} 2 & -2 & 1 & 0 & -3 \\ 2 & 3 & 1 & -1 & 4 \\ -1 & -1 & \boxed{2} & 1 & -1 \\ 3 & 1 & -1 & 4 & 4 \end{bmatrix} \rightarrow L = \begin{bmatrix} 1 & & & \\ 2 & 1 & & \\ -1 & -1 & 1 & \\ 3 & 1 & -1 & 1 \end{bmatrix}, \quad U = \begin{bmatrix} 2 & -2 & 1 & 0 \\ & 3 & 1 & -1 \\ & & 2 & 1 \\ & & & 4 \end{bmatrix}, \quad x = \begin{bmatrix} 1 \\ 2 \\ -1 \\ 1 \end{bmatrix}.$$

Problems in Chapter 5

5.1. We have the Cholesky-decomposition $A = LL^T$. Give the operation count for computing $x^T Ax$ if matrix A is used in the computation! How can we decrease the number of operations if $x^T LL^T x$ is used?

To compute Ax requires $2n^2$ flops and the remaining scalar product $2n$ flops. But the computation of $L^T x$ involves $n(n-1)$ flops plus $2n$ flops for sclar product computation. Therefore the second approach needs half as many operations.

5.2. We can avoid square roots, if we use the form $A = LDL^T$, where L has unit diagonal and D is a diagonal matrix. Elaborate the steps of this decomposition! This method can also be used for indefinite matrices if the pivot elements in D happen to be large enough.

This time we proceed similarly as in *LU*-decomposition. If a step is ready, we divide the row of the pivot with the pivot and save pivot in a diagonal matrix. For symmetric matrices division is not necessary, because the U part is the transpose of *L*.

5.3. Show that the row diagonal dominance is preserved if the matrix is multiplied from the left by a nonsingular diagonal matrix. Also, it is preserved if two rows and columns with the same row and column numbers are interchanged.

A diagonal matrix on the left multiplies rows with a number. If that number is nonzero, the ratio between the entries will be the same such that diagonal dominance is kept.

Interchanging the same two rows and columns will keep the diagonal element in diagonal position, only the order of row elements is changed that will introduce no effect in diagonal dominance.

5.4. Show that for the *LU*-decomposition of essentially diagonally dominant matrices (by row): strict diagonally dominance takes place in the j th step for the *k*-th row, if there was strict diagonal dominance in the j -th row and the element $a_{jk}^{(j)}$, $j<k$ was not zero.

It is enough to check the first step. Strict diagonal dominance in the first row means that the full absolute contribution of the first column element will be less than the first column element, if it is nonzero.

5.5. Show that diagonal dominance by columns is also inherited in *LU*-decomposition.

This time we attach the divisor to columns and consider the contribution of the left out row elements in a column. The only change is that the reasoning is done to columns.

5.6. If we are given a new right vector b, which data should be preserved and which data should be recomputed in both algorithms (fast *LU* and passage)?

Fast *LU*: The second row in (6.5) should be recomputed for b' elements. The off-diagonal and pivot elements are needed in the computation.

Passage: The f_i elements should be recomputed. Still g_i's are neede for the solution.

5.7. Prove that the tridiagonal matrix in (6.2) is positive definite, because it has a LL^T -decomposition.
In fact, diagonal dominance will take place in the actual rows when doing *LU*-decomposition. The second pivot is $2-1/2=3/2$. The third one is $2-2/3=4/3$. Assuming the $n-1$-st pivot is $n/(n-1)$, inductively one gets $2-(n-1)/n=(2n-n+1)/n=(n+1)/n$ for the next pivot, such that Choleski-decomposition is possible.

5.8. $\begin{bmatrix} 4 & -2 & 4 & -4 \\ -2 & 10 & -5 & 5 \\ 4 & -5 & 9 & -3 \\ -4 & 5 & -3 & 22 \end{bmatrix} = LL^T,\ L=?$

$$\begin{bmatrix} 4 & -2 & 4 & -4 \\ -2 & 10 & -5 & 5 \\ 4 & -5 & 9 & -3 \\ -4 & 5 & -3 & 22 \end{bmatrix} \to \begin{bmatrix} \boxed{2} & -1 & 2 & -2 \\ -1 & 9 & -3 & 3 \\ 2 & -3 & 5 & 1 \\ -2 & 3 & 1 & 18 \end{bmatrix} \to \begin{bmatrix} 2 & & & \\ -1 & \boxed{3} & -1 & 1 \\ 2 & -1 & 4 & 2 \\ -2 & 1 & 2 & 17 \end{bmatrix} \to \begin{bmatrix} 2 & & & \\ -1 & 3 & & \\ 2 & -1 & \boxed{2} & 1 \\ -2 & 1 & 1 & 16 \end{bmatrix},$$

$$L = \begin{bmatrix} 2 & & & \\ -1 & 3 & & \\ 2 & -1 & 2 & \\ -2 & 1 & 1 & 4 \end{bmatrix}$$

Problems in Chapter 6

6.1. Prove formula (6.5)!

Multiplying two factors shows, the statement is true for $i = 2$. Taking a next factor will add a new term only in the sum because of orthogonality of the vectors. Inductively, one term will be added for an arbitrary i by the same reasoning.

6.2. Show that $r_{j,i+1}$'s of (6.7) and (6.8) are equal!

It is because the projections for z_j do not make change in the *j*-th orthogonal vector: $q_j^T = q_j^T P_1 P_2 \dots P_{j-1}$.

6.3. Collect the orthogonal vectors into matrix $Q = [q_1 q_2 \dots q_i]$. Derive the formula: $P_i P_{i-1} \dots P_1 = = I - Q\left(Q^T Q\right)^{-1} Q^T$.

Apply the weighted dyadic sum of Sect. 3.4 for the last formula of (6.5). See also: Problem 3.5.

6.4. Let matrix $A \in \mathbb{R}^{m \times n}$ have linearly independent columns. Check that $I - A\left(A^T A\right)^{-1} A^T$ is also a projection and applying it to a vector, the resulting vector will be orthogonal to all columns of A.

It is enough to check the projection condition $P^2 = P$ for $A\left(A^T A\right)^{-1} A^T$ as $I - P$ is also a projection. Really, $z^T \left(I - A\left(A^T A\right)^{-1} A^T\right) A = z^T (A - A) = 0$.

6.5. One can elaborate the variant of GS orthogonalisation, when the q_j's are normed vectors, $\left\|q_j\right\|_2 = 1$. Rewrite formulas for that case!

In this case $Q^TQ = I$ holds such that $Q^TA = R$ from (6.9). All divisors in the projections are 1's and the $i+1$ -st vector is given by

$$r_{i+1,i+1}q_{i+1} = a_{i+1} - \sum_{j=1}^{i} q_j r_{j,i+1}, \quad r_{j,i+1} = q_j^T a_{i+1}, \ ,$$

where $r_{i+1,i+1}$ is found from the condition that q_{i+1} is normalized.

6.6. Having a QR -decomposition of A, how can we solve the linear system $Ax = b$?

First step: form $Q^TAx = Rx = Q^Tb$. Second step: solve $Rx = Q^Tb$ from below for x, because R is upper triangular.

6.7. Make the QR -decomposition of $\begin{bmatrix} 3 & 0 & -1 \\ 3 & 4 & 1 \\ -6 & -4 & 3 \end{bmatrix}$!

We may choose $\begin{bmatrix} 1 & 1 & -2 \end{bmatrix}^T$ for q_1 because all elements of the first column can be divided by 3 and thus the computation is simpler. Now $q_1^Tq_1 = 6$. The projection for the second column results in: $\begin{bmatrix} 0 \\ 4 \\ -4 \end{bmatrix} - \frac{12}{6}\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \\ 0 \end{bmatrix} = 2q_2, \ q_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$. $q_2^Tq_2 = 2$. From the third column $\begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix} - \frac{(-6)}{6}\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} - \frac{2}{2}\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = q_3$ and $\begin{bmatrix} 3 & 0 & -1 \\ 3 & 4 & 1 \\ -6 & -4 & 3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \\ -2 & 0 & 1 \end{bmatrix}\begin{bmatrix} 3 & 2 & -1 \\ & 2 & 1 \\ & & 1 \end{bmatrix}$.

6.8. $A = \begin{pmatrix} 1 & 3 & 4 \\ 2 & 0 & -2 \\ 2 & 3 & 3 \end{pmatrix} = QR$. Give QR -decomposition in modified Gram-Schmidt style!

$$\begin{array}{ccc} 1 & 1 & 2/3 \\ \begin{pmatrix} 1 & 3 & 4 \\ 2 & 0 & -2 \\ 2 & 3 & 3 \end{pmatrix} & \rightarrow & \begin{pmatrix} 2 & 10/3 \\ -2 & -10/3 \\ 1 & 5/3 \end{pmatrix} \\ 9 & 9 & 6 \end{array} \quad \begin{array}{cc} 1 & 5/3 \\ \rightarrow & \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\ 9 & 15 \end{array}. \quad A = \begin{pmatrix} 1 & 2 \\ 2 & -2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 & 2/3 \\ & 1 & 5/3 \end{pmatrix} = QR.$$

Now we calculate the scalar products of the first column with all columns in the initial matrix and the result is shown under the columns. On top of the columns these numbers are divided by the squared norm of the first column (=9). Next the first column is multiplied by the numbers above the second and third columns and the resulting vector is subtracted from the corresponding vectors. In other words, form the dyad by the first column and the row vector on top of the matrix and subtract – similarly to *LU*-decomposition.

Repeat the same procedure for the second matrix that has by one less columns. The third matrix is zero and finally the decomposition is given. Observe, the row vectors above the

intermediate matrices give the nonzero row elements of R. As seen, there are only two linearly independent vectors in this set.

6.9. Let the Arnoldi method is performed so that the orthogonal vectors are normed. Show that if A is symmetric, then the upper Hessenberg matrix H is also symmetric, i.e. tridiagonal.
The matrix form of the recursion is $AQ = QH + h_{i+1,i} q_{i+1} e_i^T$, where $Q = \begin{pmatrix} q_1 & q_2 & \dots & q_i \end{pmatrix}$.
Multiplying from the left by Q^T gives $Q^T AQ = H$. That shows the symmetricity of H for symmetric A and then the Hessenberg matrix should be tridiagonal.

6.10. Let the starting vector x be the sum of three eigenvectors of A having different eigenvalues. How many new vectors can be generated by the Arnoldi method?

Let the starting vector be $x = \alpha_1 u_1 + \alpha_2 u_2 + \alpha_3 u_3$, where α_j's are nonzero scalars and u_j's are the eigenvectors with eigenvalues λ_j. Introduce matrix $U = \begin{bmatrix} \alpha_1 u_1 & \alpha_2 u_2 & \alpha_3 u_3 \end{bmatrix}$. With these the Krylov vectors can be arranged in a matrix as

$$\begin{bmatrix} x & Ax & A^2 x \end{bmatrix} = U \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 \\ 1 & \lambda_2 & \lambda_2^2 \\ 1 & \lambda_3 & \lambda_3^2 \end{bmatrix}.$$

Adding a next column to the transposed Vandermonde matrix would not increase rank, therefore the maximal possible number of linearly independent vectors is 3.

Problems in Chapter 7

7.1. Let A be an upper Hessenberg matrix such that all subdiagonal elements are nonzero. Show that there is only one Jordan block to each eigenvalue.
Consider $\lambda I - H$, where λ is an eigenvalue. Then the rank loss may be no more than 1, because the subdeterminant by deleting first row, last column is nonzero. The rank loss shows the number of Jordan blocks to an eigenvalue.

7.2. Show if the eigenvalues of A are λ_i's, then A^{-1} has eigenvalues $1/\lambda_i$'s.
$Au = \lambda u \;\rightarrow\; A^{-1}u = \lambda^{-1}u$.

7.3. Check: $\|A\|_2 = \sigma_1$. $\qquad \|A\|_F = \left(\sum_{i=1}^{r} \sigma_i^2 \right)^{1/2}$.

Let A have the singular value decomposition $A = V\Sigma U^H$, where U, V are unitary. Then σ_1^2 is the maximal eigenvalue of $A^H A = U\Sigma V^H V\Sigma U^H = U\Sigma^2 U^H$, such that σ_1 gives the 2-norm.
Further $\|A\|_F^2 = \mathrm{tr}(A^H A) = \mathrm{tr}(U\Sigma V^H V\Sigma U^H) = \mathrm{tr}(U\Sigma^2 U^H) = \mathrm{tr}(U^H U\Sigma^2) = \mathrm{tr}(\Sigma^2)$.

7.4. The matrix is diagonally dominant, if the Gershgorin disks do not have zero.

In that case all radii are less than the belonging absolute diagonal element and that is just the condition for diagonal dominance.

7.5. Diagonally dominant matrices are invertible.

Yes, because no Gershgorin disk contains zero eigenvalue.

7.6. The rank of the matrix is at least as large as the number of those Gershgorin disks, which do not contain zero.

These disks belong to diagonally dominant rows. Then it is possible to find a submatrix from these rows that is nonsingular, therefore the rank is at least equals to the number of such rows.

7.7. Gershgorin disks can also be found with respect to columns if using left eigenvectors in the derivation .

Equivalently, consider the transposed matrix.

7.8. By using Gershgorin's theorem and diagonal similarity transform, decide if matrix A is invertible: $A = \begin{bmatrix} 7 & 6 & -3 \\ 1 & 5 & 1 \\ 4 & -2 & 6 \end{bmatrix}$.

Choose $D = \operatorname{diag}\begin{bmatrix} 1 & 2 & 1 \end{bmatrix}$, then DAD^{-1} is diagonally dominant.

7.9. Show that the eigenvalues of a 2×2 matrix A are:

$\lambda_{1,2} = \frac{a_{11}+a_{22}}{2} \pm \sqrt{\left(\frac{a_{11}-a_{22}}{2}\right)^2 + a_{12}a_{21}}$.

It is the root formula of the characteristic polynomial $\lambda^2 - (a_{11}+a_{22}) + a_{11}a_{22} - a_{12}a_{21} = 0$.

Prove that

7.10. $\|U\| \le \|AU\| / m$, where $AU = U\Lambda$.

$U = AU\Lambda^{-1}$ and taking norms gives the result, where m is the smallest absolute eigenvalue.

7.10. $\|U^{-1}\| \le \|U^{-1}A^{-1}\| M$.

$U^{-1} = \Lambda U^{-1}A^{-1}$ and if M is the largest absolute eigenvalue, the result follows by taking norms.

7.12. $\operatorname{cond}(U) \le \operatorname{cond}(AU)\operatorname{cond}(\Lambda)$.

Multiplying the same sides of the previous inequalities give the result.

Problems in Chapter 8

8.1. Let $A = LU$ be a rank-factorisation. What is the orthogonal projection to $\operatorname{Im}(A)$?

$\mathrm{Im}(A)$ is the subspace spanned by the columns of L such that the orthogonal projection to it is given by $LL^+ = L(L^TL)^{-1}L^T$.

8.2. What is the orthogonal projection to the null space of A ? Give the distance of x to $\mathrm{Nul}(A)$ in two-norm!

The null space of A is given by the orthogonal projection $I - AA^+$. Applying the distance theorem: $\mathrm{dist}_2(\mathrm{Nul}(A), x) = \left\| I - (I - AA^+)x \right\|_2 = \left\| AA^+x \right\|_2$

8.3. A line passes points r_0 and r_1 . Give the distance of vector x from this line!

The direction vector of the line is: $d = r_1 - r_0$. Vector $r_1 - x$ is a distance of x from a point of the line such that is has a component parallel to d and an other one, perpendicular to d . The length of the perpendicular component gives the distance from the line:

$\mathrm{dist}_2(line, x) = \left\| \left(I - dd^T / d^Td \right)(r_1 - x) \right\|_2$.

8.4. Show that if a matrix is invertible then its inverse and pseudoinverse are equal.

It comes from the first Penrose condition: $A = AA^+A$ and multiplying by the inverse of A gives: $I = AA^+$.

8.5. $A^T = \begin{bmatrix} 2 & -4 & 6 \\ 0 & 5 & -5 \end{bmatrix}$. Give the orthogonal projection into $\mathrm{Im}(A)$!

$$P = AA^+ = \begin{bmatrix} 2 & 0 \\ -4 & 5 \\ 6 & -5 \end{bmatrix} \left(\begin{bmatrix} 2 & -4 & 6 \\ 0 & 5 & -5 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ -4 & 5 \\ 6 & -5 \end{bmatrix} \right)^{-1} \begin{bmatrix} 2 & -4 & 6 \\ 0 & 5 & -5 \end{bmatrix} .$$

8.6. The row vectors of A^T in Problem 8.5 are the normal vectors of two planes. Give the orthogonal projection into the intersection of the two planes!

Vector x is in the intersection of the two planes if $A^Tx = 0$. The orthogonal projection into the intersection of planes is given by $P = I - A\left(A^TA\right)^{-1}A^T = I - AA^+$ because $A^TPz = 0$ for an arbitrary vector.

8.7. $r^T = [1 \quad -1 \quad 1]$. What is the distance of vector r from the intersection of the two planes in the previous problem?

$$\mathrm{dist}_2(\mathrm{intersection}, r) = \left\| \left(I - P \right) r \right\|_2 = \left\| A\left(A^TA\right)^{-1}A^Tr \right\|_2 .$$

8.8. $A = \begin{bmatrix} 2 & 0 & 3 \\ -4 & 5 & -2 \\ 6 & -5 & 5 \end{bmatrix}$, $\mathrm{rank}(A) = 2$. $A^+ = ?$ (Use *LU*-decomposition!)

LU-decomposition gives: $A=\begin{pmatrix}1 & \\ -2 & 1\\ 3 & -1\end{pmatrix}\begin{pmatrix}2 & 0 & 3\\ & 5 & 4\end{pmatrix}=LU$. It is a rank factorization if both L and U have rank two, such that two columns in L and two rows in U are needed. For the pseudoinverse $A^+=U^+L^+=U^T\left(UU^T\right)^{-1}\left(L^TL\right)^{-1}L^+$ needs to be computed.

8.9. What is the pseudosolution of $Ax=b$ if $b^T=[1\quad -1\quad 1]$ and matrix A comes from Problem 8.8.

8.10. Show $I-A^+A=0$, if the columns of A are linearly independent.

8.11. Derive the relation $(A^+)^T=(A^T)^+$ from the four Pendrose conditions.

8.12. Matrix A has an approximate eigenvector x. Find the belonging approximate eigenvalue λ from the condition that $\|Ax-\lambda x\|_2$ is minimal. Give formula for λ!

Problems in Chapter 9

9.1. Check if the base points are located symmetrically to $x=0$, then $\alpha_j=0,\ \ i=1,2,\ldots$ hold and the polynomials are alternating even and odd functions.

9.2. Find the orthogonal polynomials p_0,p_1,p_2 for the base points $\{-2,-1,0,1,2\}$!

9.3. The Chebyshev polynomials are also orthogonal and they can be generated by the following recursion: $T_0=1,\ \ T_1=x,\ \ T_{n+1}=2xT_n-T_{n-1}$. Although they are not monic now, yet it is the familiar form. Expand $4x^2-3x+2$ with Chebyshev polynomials!

9.4. $P(x)=\sum_{j=0}^{k}(2j+1)T_j(x)$. Give a skillful way of computing the sum at the point x_0!

9.5. Show that $(p_i,p_i)=\mu_0\beta_1\beta_2\ldots\beta_i$, where $\mu_0=(p_0,p_0)\left[=\int_a^b\alpha(x)dx\right]$ is the 0-th moment.

9.6. Show that the principal minors of the tridiagonal matrix

$$\begin{pmatrix}x-\alpha_1 & -\beta_1 & & \\ -\beta_1 & x-\alpha_2 & \ddots & \\ & \ddots & \ddots & -\beta_{n-1}\\ & & -\beta_{n-1} & x-\alpha_n\end{pmatrix}$$

have the same recursions as orthogonal polynomials with parameters α_i and β_i^2.

Problems in Chapter 10

10.1. How should we modify Jacobi iteration, if the matrix is diagonally dominant with respect to columns?

10.2. Show that Theorem 10.3.1 can be reformulated for the case when the matrix is diagonally dominant with respect to columns.

10.3. Elaborate estimate **Hiba! A hivatkozási forrás nem található.** for the GS-iteration! What happens to Jacobi and GS iteration if instead of diagonal dominance we have equality in some equations? And if equality takes place in the last row?

10.4. $A = \begin{pmatrix} 5 & -1 & 2 & 1 \\ -3 & 7 & -2 & 0 \\ 3 & 0 & 5 & -1 \\ 0 & 2 & -4 & 6 \end{pmatrix}$. $\|B_J\|_\infty = ?$ $\|B_{GS}\|_\infty \le ?$

10.5. Applying Theorem 10.3.1 show: $\|A^{-1}\|_\infty \le \max_i \frac{1}{|a_{ii}|(1-\alpha_i-\beta_i)}$, see also **Hiba! A hivatkozási forrás nem található.**, if A is strictly diagonally dominant by rows. How can we modify statement for diagonal dominance with respect to columns?

10.6. Assuming $D+\omega L$ is diagonally dominant by rows, prove $\|B_{GS}(\omega)\|_\infty \le \max_{(j)} \frac{|1-\omega|+\omega\beta_j}{1-\omega\alpha_j}$ by applying Theorem 10.3.1.

10.7. If $\|D^{-1}A_1\| < 1$ holds, then we can derive an inequality similar to that of Theorem 10.3.1 by using (2.15), because of the equality $(A_1+D)^{-1}A_2 = (I+D^{-1}A_1)^{-1}D^{-1}A_2$. Show that $\|(A_1+D)^{-1}A_2\| \le \frac{\|D^{-1}A_2\|}{1-\|D^{-1}A_1\|}$ holds for induced norms. Is that necessary that D be a diagonal matrix? For the matrix of Example 4 which method gives a better estimate?