NTNU – Trondheim
Norwegian University of
Science and Technology

Department of Mathematical Sciences

Examination paper for **TMA4205 Numerical Linear Algebra**

Academic contact during examination: Elena Celledoni

Phone: 48238584

Examination date: December 15, 2018

Examination time (from–to): 09:00-13:00

Permitted examination support material: C: Approved calculator.

Language: English

Number of pages: 8

Number of pages enclosed: 3

Checked by:

Informasjon om trykking av eksamensoppgave	
Originalen er:	
1-sidig ☒	**2-sidig** ☐
sort/hvit ☒	**farger** ☐
skal ha flervalgskjema ☐	

Date Signature

This exam set includes an appendix with formulae, algorithms and results useful for the solution of the exam questions.

There are seven subparts in this exam. You need to solve 6 of them to get full score. Problem **5** *is optional, you can solve it instead of one of the other problems, or in addition to the other problems if you have enough time. Problem* **5** *will only contribute positively to the final score.*

Problem 1 Show that the $n \times n$ matrix $n \geq 2$

$$A := \begin{bmatrix} 1 & a & \dots & a \\ a & 1 & \dots & a \\ \vdots & \vdots & \dots & \vdots \\ a & a & \dots & 1 \end{bmatrix}$$

is positive definite if and only if $-\frac{1}{n-1} < a < 1$ and that the Jacobi method converges if and only if $-\frac{1}{n-1} < a < \frac{1}{n-1}$.

Hint Note that $A = I(1-a) + a\,\mathbf{e}\mathbf{e}^T$ where $\mathbf{e} = [1, 1, \dots, 1]^T \in \mathbb{R}^n$, and the eigenvalues of A can be deduced from the eigenvalues of $a\,\mathbf{e}\mathbf{e}^T$ which are easy to find.

Solution We see that A is a linear combination of the identity and of the matrix $a\,\mathbf{e}\mathbf{e}^T$,

$$A = I(1-a) + a\,\mathbf{e}\mathbf{e}^T.$$

We can compute the eigenvalues of A by $Ax = \lambda x$,

$$(1-a)x + a\,\mathbf{e}\mathbf{e}^T x = \lambda x$$

and

$$a\,\mathbf{e}\mathbf{e}^T x = (\lambda - 1 + a)x.$$

The rank one matrix $a\,\mathbf{e}\mathbf{e}^T$ has 0 as an eigenvalue of multiplicity $n-1$ (with $n-1$ linearly independent eigenvectors orthogonal to $\mathbf{e}$), and $n\,a$ as eigenvalue of multiplicity one, with $\mathbf{e}$ as an eigenvector. From this we deduce the eigenvalues of A which are

$$\lambda = 1 + (n-1)a, \qquad \lambda = 1 - a,$$

and with respective multiplicities $n-1$ and 1.

A necessary and sufficient condition for A to be positive definite is that $\lambda > 0$, then we have

$$1 + (n-1)a > 0, \qquad 1 - a > 0,$$

that is

$$-\frac{1}{n-1} < a < 1.$$

A necessary and sufficient condition for the Jacobi method to converge is that the spectral radius of $(I - A)$, so all the eigenvalues of $I - A$ should have absolute value less than 1. By similar reasoning as above, we find that the eigenvalues of $I - A$ are $-(n-1)a$ and a, imposing they have absolute value less than 1 is equivalent to the condition

$$|a|(n-1) < 1 \Leftrightarrow -\frac{1}{n-1} < a < \frac{1}{n-1}.$$

Problem 2

a) Compute one step of the QR-iteration with single shift for the matrix

$$A := \begin{bmatrix} 2 & \varepsilon \\ \epsilon & 1 \end{bmatrix}.$$

Use $\mu_1 = 1$ as shift value in the first iteration. Explain why this is a reasonable choice. Use Givens rotations for the QR-factorization.

b) For $\varepsilon > 0$ small, the eigenvalues of A are $\lambda_1 = 2 + \varepsilon^2 + \mathcal{O}(\varepsilon^4)$ and $\lambda_2 = 1 - \varepsilon^2 + \varepsilon^4 + \mathcal{O}(\varepsilon^6)$. The first iteration of the QR-iteration for A without shift leads to

$$T_1 = \frac{1}{4+\varepsilon^2}\begin{bmatrix} 8 + 5\varepsilon^2 & 2\varepsilon - \varepsilon^3 \\ 2\varepsilon - \varepsilon^3 & 4 - 2\varepsilon^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 + \mathcal{O}(\varepsilon^2) & \mathcal{O}(\varepsilon) \\ \mathcal{O}(\varepsilon) & \lambda_2 + \mathcal{O}(\varepsilon^2) \end{bmatrix}.$$

Compare the result for the first iteration of the QR-iteration with and without shift. What can you say about the the speed of convergence of the two methods? Use Taylor expansion around $\epsilon = 0$.

Solution

a)Because the matrix A is 2×2, it is already in tridiagonal form, $T_0 := A$. We compute the QR-factorization of $T_0 = QR$ using a Givens rotation:

$$\begin{bmatrix} c & s \\ -s & c \end{bmatrix}\begin{bmatrix} 2 & \varepsilon \\ \epsilon & 1 \end{bmatrix} = \frac{1}{\sqrt{4+\varepsilon^2}}\begin{bmatrix} 4+\varepsilon^2 & 3\varepsilon \\ 0 & 2-\varepsilon^2 \end{bmatrix},$$

where $c = \frac{2}{\sqrt{4+\varepsilon^2}}$, $s = \frac{\varepsilon}{\sqrt{4+\varepsilon^2}}$. Next we compute $T_1 = RQ$ and we get

$$T_1 = \frac{1}{4+\varepsilon^2}\begin{bmatrix} 4+\varepsilon^2 & 3\varepsilon \\ 0 & 2-\varepsilon^2 \end{bmatrix}\begin{bmatrix} 2 & -\varepsilon \\ \epsilon & 2 \end{bmatrix} = \frac{1}{4+\varepsilon^2}\begin{bmatrix} 8+5\varepsilon^2 & 2\varepsilon-\varepsilon^3 \\ 2\varepsilon-\varepsilon^3 & 4-2\varepsilon^2 \end{bmatrix},$$

$$T_1 = \begin{bmatrix} \lambda_1 + \mathcal{O}(\varepsilon^2) & \mathcal{O}(\varepsilon) \\ \mathcal{O}(\varepsilon) & \lambda_2 + \mathcal{O}(\varepsilon^2) \end{bmatrix}.$$

To apply one iteration of the QR-iteration with shift, we apply the QR-factorization to $T_0 - \mu I$ where $\mu = A(2,2) = 1$. This value is suitable to use as shift μ because it can be regarded as the best available approximation to an eigenvalue of A (see also Theorem 7.5.1 in Golub and Van Loan). The p-th subdiagonal entry of T_k in the QR-iteration with shift converges to zero with a rate

$$\left|\frac{\lambda_p - \mu}{\lambda_{p+1} - \mu}\right|^k.$$

In our case, we have just 2 eigenvalues: λ_1 and λ_2. If λ_2 is much closer to μ than to λ_1, then the convergence of the $(2,1)$ element to zero is rapid.

In fact we get

$$T_0 - \mu I = \begin{bmatrix} 1 & \varepsilon \\ \epsilon & 0 \end{bmatrix} = \frac{1}{\sqrt{1+\varepsilon^2}}\begin{bmatrix} 1 & -\varepsilon \\ \epsilon & 1 \end{bmatrix}\frac{1}{\sqrt{1+\varepsilon^2}}\begin{bmatrix} 1+\varepsilon^2 & \varepsilon \\ 0 & -\varepsilon^2 \end{bmatrix}$$

and

$$T_1 = \frac{1}{\sqrt{1+\varepsilon^2}}\begin{bmatrix} 1+\varepsilon^2 & \varepsilon \\ 0 & -\varepsilon^2 \end{bmatrix}\frac{1}{\sqrt{1+\varepsilon^2}}\begin{bmatrix} 1 & -\varepsilon \\ \epsilon & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and

$$T_1 = \frac{1}{1+\varepsilon^2}\begin{bmatrix} 2+3\varepsilon^2 & -\epsilon^3 \\ -\varepsilon^3 & 1 \end{bmatrix}.$$

b) From the first iteration, it appears that the QR iteration with shift is converging much faster than the QR iteration without shift. Comparing the results from the two methods, we see that the QR with shift produces a T_1 which is closer to a diagonal matrix (with an error ε^3) compared to the simple QR-iteration (error $\mathcal{O}(\varepsilon)$). The smallest eigenvalue is approximated with an error $\mathcal{O}(\varepsilon^6)$ by the QR iteration with shift and only with an error $\mathcal{O}(\varepsilon^2)$ by the QR-iteration without shift. The same holds for the largest eigenvalue.

Problem 3

We want to solve the saddle point problem

$$\begin{bmatrix} A & B \\ B^T & O \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} b \\ c \end{pmatrix}, \quad A = X^T \Lambda X, \quad B = X^T \Gamma Q,$$

where $A \in \mathbb{R}^{n \times n}$ symmetric and positive definite, $B \in \mathbb{R}^{n \times k}$, $O \in \mathbb{R}^{k \times k}$ the zero matrix with $k \leq n$. Given the eigenvalue-eigenvector decomposition of A, $A = X^T \Lambda X$, with $X \in \mathbb{R}^{n \times n}$ orthogonal and Λ diagonal, it is assumed that B has the form $B = X^T \Gamma Q$, where $\Gamma \in \mathbb{R}^{n \times k}$ is a diagonal matrix with real diagonal elements $\gamma_{i,i}$, and $0.5 \leq \gamma_{i,i}^2 \leq 1$, $(i = 1, \ldots, k)$, and $Q \in \mathbb{R}^{k \times k}$ is orthogonal.

Assume the eigenvalues of A^{-1} are ordered as follows

$$1 \leq \lambda_n(A^{-1}) \leq \lambda_{n-1}(A^{-1}) \leq \cdots \leq \lambda_1(A^{-1}) \leq 4.$$

Eliminate the variable x from the system such that the reformulated system depends only on y and show that

$$B^T A^{-1} B y = B^T A^{-1} b - c.$$

Explain why the CG method can be applied to the resulting system.

For $m = 3$, find an upper bound for

$$\frac{\|y - y_m\|_A}{\|y - y_0\|_A},$$

where y_m is the m-th iterate of the CG method applied to the system for y, and where y_0 is the corresponding initial guess.

Solution

From the equations $Ax + By = b$ and $B^T x = c$ we get $x = A^{-1}b - A^{-1}By$ and so

$$B^T A^{-1} B y = B^T A^{-1} b - c.$$

Using the factorizations of A and B we get

$$Q^T \Gamma^T \Lambda^{-1} \Gamma Q y = Q^T \Gamma^T \Lambda^{-1} X b - c.$$

The product of diagonal matrices $\Gamma^T \Lambda^{-1} \Gamma$ has eigenvalues $\gamma_{i,i}^2 \lambda_i$, for $i = 1, \ldots, k$ all included in the interval $[0.5, 4] \subset \mathbb{R}$. So $Q^T \Gamma^T \Lambda^{-1} \Gamma Q$ is symmetric and positive definite (all eigenvalues are positive) and has (spectral) condition number $\mathcal{K}$ equal to the quotient of its maximum and minimum eigenvalue, so we get the bound

$$\mathcal{K} \leq \frac{\lambda_1(A^{-1})}{\lambda_k(A^{-1})\, 0.5} \leq \frac{4}{0.5} = 8.$$

Using the bound for the relative residual given by the theorem on the convergence of the conjugate gradient method, we obtain

$$\frac{\|y - y_3\|_A}{\|y - y_0\|_A} \leq 2\left(\frac{\sqrt{\mathcal{K}} - 1}{\sqrt{\mathcal{K}} + 1}\right)^3 \leq 0.2179.$$

The last bound is obtained and observing that $\frac{\sqrt{\mathcal{K}}-1}{\sqrt{\mathcal{K}}+1}$ is monotone increasing as a function of $\mathcal{K}$ for $\mathcal{K} > 0$, and replacing $\mathcal{K}$ by the bound 8.

Problem 4 Let

$$A = \begin{bmatrix} 100 & 31 & 17 \\ -100 & 161 & -73 \\ 100 & 119 & -217 \\ 100 & -73 & -161 \end{bmatrix}.$$

a) Find orthogonal matrices U_B and V_B such that $B = U_B^T A V_B$ is a bidiagonal matrix i.e. a matrix with the following sparsity pattern

$$B = \begin{bmatrix} x & x & 0 \\ 0 & x & x \\ 0 & 0 & x \\ 0 & 0 & 0 \end{bmatrix},$$

where x in the matrix denotes a generic non-zero element of B.

Hint Use sign minus in the formula (1) in the appendix to define the Householder transformations. This will keep the calculations relatively simple.

Solution

We use Householder transformations:

$$P_1 := I - 2ww^T, \qquad w := \frac{1}{2}[-1, -1, 1, 1]^T,$$

$P_1 A$ has the first column $200\,\mathbf{e}_1$ and

$$P_1 A = \begin{bmatrix} 200 & -42 & -144 \\ 0 & 88 & -234 \\ 0 & 192 & -56 \\ 0 & 0 & 0 \end{bmatrix},$$

then use

$$Q_1 := I - 2uu^T, \qquad u := \frac{1}{240}[0, -192, -144]^T,$$

and get

$$B := P_1 A Q_1 = \begin{bmatrix} 200 & 150 & 0 \\ 0 & 200 & -150 \\ 0 & 0 & -200 \\ 0 & 0 & 0 \end{bmatrix},$$

so $U_B := P_1$ and $V_B := Q_1$.

b) Give a short explanation[1] of how the bi-diagonalization of A can be used for the efficient computation of the singular values and singular vectors of A.

Solution
The algorithm of Golub and Kahan to compute the singular values of A transforms first A into a bi-diagonal form $B = U_B^T A V_B$ via orthogonal matrices U_B and V_B. The singular values of A coincide with the singular values of B and can be computed applying a QR-iteration method (with shift) or some other method to compute the eigenvalues of the matrix $B^T B$ which is tridiagonal. The QR-iteration applied to tridiagonal matrices, preserves the tridiagonal form throughout the iteration. Once $B^T B = \tilde{V} \Lambda \tilde{V}^T$ is found, the singular value decomposition of B can be deduced with standard methods, $\sigma_i = \sqrt{\lambda_i}$, and $B = \tilde{U} \Sigma \tilde{V}^T$. The singular vectors of A are

$$V := V_B \tilde{V}, \quad U = U_B \tilde{U}.$$

Alternative solution A clever way to explain this is as follows: here B can be diagonalized with two givens rotations: $G_2 G_1 B^T = \Sigma^T$ then

$$B G_1^T G_2^T = U_B^T A V_B G_1^T G_2^T = \Sigma$$

and so

$$A = U_B \Sigma G_2 G_1 V_B^T = U \Sigma V^T,$$

the SVD of A.

[1] Use few sentences. You are not requested to do any calculations, but just to explain how to proceed.

Problem 5 (*This problem is optional. See note at the beginning of the exam text for further explanation.*)
Let a matrix A have the form

$$A = \begin{bmatrix} I & Y \\ 0 & S \end{bmatrix}.$$

Assume that the Full orthogonalisation method (FOM) is used to solve a linear system with the matrix A. Let

$$r_0 = \begin{bmatrix} r_0^{(1)} \\ r_0^{(2)} \end{bmatrix}$$

be the initial residual vector. Assume that the degree of the minimal polynomial of $r_0^{(2)}$ with respect to S (i.e. *grade* of $r_0^{(2)}$ with respect to S) is k. What is the maximum number of iterations that FOM would require to converge for this matrix and the initial residual $r_0^{(2)}$?

If you are not familiar with the FOM you can solve the problem for the GMRES instead.

Hint: Evaluate the sum

$$\sum_{i=0}^{k} \beta_i (A^{i+1} - A^i) r_0$$

where $p(t) = \sum_{i=0}^{k} \beta_i t^i$ is the minimal polynomial of $r_0^{(2)}$ with respect to S.

Solution

The minimal polynomial of $r_0^{(2)}$ with respect to S is k, if and only if

$$p(S)r_0^{(2)} = \sum_{i=0}^{k} \beta_i S^i r_0^{(2)} = 0.$$

We find an expression for A^i.

$$\begin{aligned} A^2 &= \begin{bmatrix} I & Y + YS \\ 0 & S^2 \end{bmatrix} = \begin{bmatrix} I & Y(I+S) \\ 0 & S^2 \end{bmatrix} \\ A^3 &= \begin{bmatrix} I & Y(I+S) + YS^2 \\ 0 & S^3 \end{bmatrix} = \begin{bmatrix} I & Y(I+S+S^2) \\ 0 & S^3 \end{bmatrix} \\ &\vdots \\ A^i &= \begin{bmatrix} I & Y(I+S+\cdots+S^{i-1}) \\ 0 & S^i \end{bmatrix}. \end{aligned}$$

So

$$\begin{aligned} \sum_{i=0}^{k} \beta_i (A^{i+1} - A^i) r_0 &= \sum_{i=0}^{k} \beta_i \begin{bmatrix} 0 & YS^i \\ 0 & S^{i+1} - S^i \end{bmatrix} \begin{bmatrix} r_0^{(1)} \\ r_0^{(2)} \end{bmatrix} \\ &= \begin{bmatrix} 0 & Y \sum_{i=0}^{k} \beta_i S^i r_0^{(2)} \\ 0 & (S-I) \sum_{i=0}^{k} \beta_i S^i r_0^{(2)} \end{bmatrix} = 0. \end{aligned}$$

Since $\sum_{i=0}^{k} \beta_i (A^{i+1} - A^i) r_0$ defines a polynomial of degree $k+1$ in A, the *grade* of r_0 with respect to A must be $k+1$. Using proposition 2 from the appendix, we see that this means that $h_{k+2,k} = 0$ and the residual of FOM $r_{k+1} = b - Ax_{k+1} = -h_{k+2,k+1} v_{k+2} (e_{k+1}^T y_{k+1})$ vanishes. So FOM converges in maximally $k+1$ iterations. Similarly for GMRES, when $h_{k+2,k+1} = 0$ the approximation can be seen to coincide with the one of FOM and the residual is zero in at most $k+1$ iterations. Alternatively one can get to the same conclusion using the minimizing property of GMRES: the GMRES residual $r_m = b - Ax_m$ is minimized in 2-norm among all possible approximations x_m of the form $x_m = x_0 + V_m y_m$.

APPENDIX

This appendix contains useful formulae, algorithms and results to solve the exam questions.

Householder transformations

An Householder transformation is a matrix of the form

$$P = I - 2ww^T,$$

where I is the identity matrix $n \times n$, $w \in \mathbb{R}^n$ with w of 2-norm equal to 1. Given $x \in \mathbb{R}^n$ we can define an Householder transformation such that

$$Px = \gamma\,\mathbf{e}_1, \qquad \gamma \in \mathbb{R}$$

and $\mathbf{e}_1$ the first canonical vector. This can be achieved by taking $w = \tilde{w}/\|\tilde{w}\|_2$ and

$$\tilde{w} = x \pm \|x\|_2\mathbf{e}_1. \tag{1}$$

To solve the exercises, both plus and minus can be used. For the sake of simplicity, it is advisable to choose the sign that gives the simplest calculations.

Givens rotations

A Givens rotation 2×2 is a matrix

$$\begin{bmatrix} c & s \\ -s & c \end{bmatrix}$$

where $c^2 + s^2 = 1$. Given a vector $x = [x_1, x_2]^T$, we can construct a Givens rotation G such that $Gx = \alpha\mathbf{e}_1$ with α a scalar and $\mathbf{e}_1$ the first canonical vector. This can be done by choosing

$$c = \frac{x_1}{\sqrt{x_1^2 + x_2^2}}, \qquad s = \frac{x_2}{\sqrt{x_1^2 + x_2^2}}.$$

QR-iteration

Assume that $A \in \mathbb{R}^{n\times n}$ is symmetric and that $Q_0 \in \mathbb{R}^{n\times n}$ is orthogonal.
$T_0 = Q_0^T A Q_0$
for $k = 1, 2, \dots$ **do**
    $T_k = Q_k R_k$, (QR-factorization)
    $T_k = R_k Q_k$
end for

QR-iteration with shift

Assume that $A \in \mathbb{R}^{n\times n}$ is symmetric and that $Q_0 \in \mathbb{R}^{n\times n}$ is orthogonal, $I \in \mathbb{R}^{n\times n}$ the identity.
$T_0 = Q_0^T A Q_0$
for $k = 1, 2, \dots$ **do**
    Choose μ_k appropriate scalar
    $T_k - \mu_k I = Q_k R_k$, (QR-factorization)
    $T_k = R_k Q_k + \mu_k I$
end for

Convergence of Conjugate Gradient method

We are solving numerically the system

$$Ax = b$$

where A is symmetric and positive definite.

Theorem 1
Let x_m be the approximate solution obtained at the m-th step of the Conjugate Gradient algorithm, and x the exact solution of $Ax = b$, let $\mathcal{K}$ be the spectral[2] condition number $\mathcal{K} = \frac{\lambda_{\max}}{\lambda_{\min}}$ then

$$\|x - x_m\|_A \leq 2 \left(\frac{\sqrt{\mathcal{K}} - 1}{\sqrt{\mathcal{K}} + 1} \right)^m \|x - x_0\|_A.$$

Krylov subspaces

Consider the Krylov subspace

$$K_m(A, v) := \operatorname{span}\{v, Av, \dots, A^{m-1}v\}.$$

[2] The spectral condition number for symmetric positive definite matrices coincides with $\mathcal{K}_2 = \|A\|_2 \|A^{-1}\|_2$, because $\|A\|_2 = \lambda_{\max}$ and $\|A^{-1}\|_2 = \frac{1}{\lambda_{\min}}$.

Recall that the minimal polynomial of a vector v is the nonzero monic polynomial p of lowest degree such that $p(A)v = 0$. The degree of the minimal polynomial of v with respect to A is called the *grade* of v with respect to A. The grade of v does not exceed n.

The Arnoldi algorithm generates an orthonormal basis of the Krylov subspace whose vectors are the columns of the matrix $V_m := [v_1, \ldots, v_m]$, and an upper Hessenberg matrix H_m such that

$$AV_m = V_m H_m + h_{m+1,m} v_{m+1} e_m^T,$$

where e_m is the m-th canonical vector in $\mathbb{R}^m$.

Proposition 2

$$h_{m+1,m} = 0 \Leftrightarrow grade(v_1) = m.$$