Contact during exam
Name: Brynjulf Owren (93021641)
Sensur: 21.12.2009

EXAM IN NUMERICAL LINEAR ALGEBRA (TMA4205)

Monday November 30, 2009
Time: 09:00–13:00

Aids: Category A, All printed and hand written aids allowed. All calculators allowed.

Problem 1 Given the matrix

$$A = \frac{1}{21} \cdot \begin{bmatrix} -9 & 32 & -62 \\ -72 & 67 & -34 \\ -18 & 106 & 2 \end{bmatrix}.$$

a) Fill in $\mu_i, \nu_i,\ i = 1, 2, 3$ and σ_3 such that the product

$$A = \begin{bmatrix} 1/3 & -2/3 & \mu_1 \\ 2/3 & -1/3 & \mu_2 \\ 2/3 & 2/3 & \mu_3 \end{bmatrix} \begin{bmatrix} 7 & & \\ & 3 & \\ & & \sigma_3 \end{bmatrix} \begin{bmatrix} -3/7 & 6/7 & -2/7 \\ 2/7 & 3/7 & 6/7 \\ \nu_1 & \nu_2 & \nu_3 \end{bmatrix}$$

is a singular value decomposition of A.

Answer:

$$A = \frac{1}{21} \cdot \begin{bmatrix} 1 & -2 & 2 \\ 2 & -1 & -2 \\ 2 & 2 & 1 \end{bmatrix} \begin{bmatrix} 7 & & \\ & 3 & \\ & & 2 \end{bmatrix} \begin{bmatrix} -3 & 6 & -2 \\ 2 & 3 & 6 \\ 6 & 2 & -3 \end{bmatrix}$$

b) We define the set of matrices

$$\mathcal{M} = \{a_1 b_1^T + a_2 b_2^T,\ a_1, b_1, a_2, b_2 \in \mathbb{R}^3\}$$

Determine

$$\tilde{A} = \arg\min_{B \in \mathcal{M}} \|A - B\|_2$$

where A is the matrix defined above.

Answer: We note that $\mathcal{M}$ is precisely the set of matrices of rank at most 2. Therefore the best approximation in $\|\cdot\|_2$ is given as $\tilde{A} = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T$

$$\tilde{A} = \frac{1}{21}\begin{bmatrix} -33 & 24 & -50 \\ -48 & 75 & -46 \\ -30 & 102 & 8 \end{bmatrix}$$

Problem 2 Let us define the shift matrix $S \in \mathbb{R}^{n\times n}$ as

$$S = \begin{bmatrix} & & & & 1 \\ 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \end{bmatrix} \tag{1}$$

the matrix with 1 on the subdiagonal and upper right corner and 0 elsewhere. The effect of applying this matrix to a vector is that all components are shifted one position down and the last component is shifted to the first. Clearly, S is orthogonal and so $S^{-1} = S^T$ and solving problems $Sx = b$ is trivial. Nevertheless, we shall use this linear system as a test case for Krylov subspace methods.

a) Prove that the eigenvalues of S are the nth roots of unity, i.e.

$$\lambda_k = \mathrm{e}^{\frac{2ik\pi}{n}}, \quad k = 1, \ldots, n,\ (\mathrm{i} = \sqrt{-1}).$$

Answer: Various arguments could be used here, but a simple constructive one is the following: Writing the eigenvalue equation $Sx = \lambda x$ on component form, we get

$$\lambda x_1 = x_n, \quad x_{j-1} = \lambda x_j,\ j = 2, \ldots, n$$

so that we get $x_n = \lambda^n x_n$. And $x_n = 0$ would yield all other $x_i = 0$ as well. To have an eigenvector we therefore need $\lambda^n = 1$ so the eigenvalues are the nth roots of unity as given.

b) Let $v_1 = e_1 = [1, 0, \ldots, 0]^T \in \mathbb{R}^n$ and for each $m = 1, \ldots, n$ derive explicitly the matrices V_m and H_m from the Arnoldi algorithm, such that the columns of V_m form an orthonormal basis for $\mathcal{K}(S, e_1)$.

Answer: The Arnoldi algorithm can be easily computed by an induction argument. We claim that $v_j = e_j$ (a guess motivated by running the first step of the algorithm). True by assumption for $j = 1$. Suppose $v_i = e_i,\ 1 \le i \le j$, we compute $w_j = Sv_j = e_{j+1}$ if $j < n$ and $w_n = Se_n = e_1$. Thus, $h_{ij} = \langle w_j, v_i \rangle = \langle e_{j+1}, e_i \rangle = 0,\ 1 \le i \le j < n$. The algorithm now subtracts all components $h_{ij}v_i$ from w_j and this has no effect, we still have $w_j = e_{j+1}$. Then $h_{j+1,j} = \|w_j\|_2 = 1$ and $v_{j+1} = w_j / h_{j+1,j} = e_{j+1}$ so the induction works. We conclude that V_m is, for $m \le n$ the the $n \times m$ matrix whose first m rows is the identity matrix and the last $n - m$ rows are zeros. The matrix H_m is the upper left $m \times m$ part of the matrix S.

c) Suppose that we use the GMRES method to solve the linear system $Sx = b$. We assume that an initial approximation x_0 has been chosen such that $r_0 = b - Sx_0 = e_1$. Compute all approximations x_m, $m = 1, \ldots, n$. Show how each residual r_m can be expressed as $r_m = p_m(S)r_0$ for some polynomial $p_m(z)$ of degree at most m, and determine each $p_m(z)$ for $m = 1, \ldots, n$. Comment on why the usual convergence analysis presented in the book and lectures fails in this case. Discuss in particular what happens in the very last iteration ($m = n$).

Answer: We have $\beta = \|r_0\|_2 = 1$. Note that $\bar{H}_m,\ m < n$ is the upper left $m{+}1 \times m$ submatrix of S, the $(m{+}1) \times m$-matrix with 1's on the subdiagonal. We have $x_m = x_0 + V_m y$ where $x_0 = S^T(b - e_1),\ e_1 \in \mathbb{R}^n$, and b was not given. Here $y \in \mathbb{R}^m$ is the vector which solved the LS problem

$$\arg\min_{y \in \mathbb{R}^m} \|\beta e_1 - \bar{H}_m y\|_2, \quad e_1 \in \mathbb{R}^{m+1}$$

When $m < n$ we find that $\|\beta e_1 - \bar{H}_m y\|_2 = 1 + y^T y$, so clearly the minimum is achieved for $y = 0$. In other words, $x_m = x_0$ for every $m = 1, \ldots, n-1$. For $m = n$ $\bar{H}_m$ looks a little different, its last column equals $e_1 \in \mathrm{R}^{n+1}$. We compute

$$\|\beta e_1 - \bar{H}_n y\|^2 = (1 - y_n)^2 + \sum_{k=1}^{n-1} y_k^2$$

so the minimum must be at $y = e_n$, the nth canonical unit vector in $\mathbb{R}^n$. Therefore $x_n = x_0 + V_n e_n = x_0 + v_n = x_0 + e_n = S^T(b - e_1) + e_n = S^T b = S^{-1}b$ since $S^T e_1 = e_n$. We may compute $r_m = b - Sx_m = b - S(x_0 + V_m y) = r_0 - SV_m y$. For $m < n$ we have $y = 0$ and therefore $p_m(z) \equiv 1,\ m < n$. But for $m = n$, we get $r_n = r_0 - SV_n e_n = e_1 - Se_n = 0$ so that $p_n(z) \equiv 0$ would fulfill $r_n = p_n(S)r_0$, however, we should require $p_n(0) = 1$. What happens instead is that we get the polynomial $p_n(z) = 1 - z^n$ which works because $S^n = I$. In the book it is assumed that the eigenvalues of the matrix (S) can be located in an ellipsis which is separated from the origin, this can not be done with the our S since the eigenvalues are uniformly distributed on the unit circle. If we try to adapt the strategy from the book to our case anyway, we would consider

$$\min_{p \in \tilde{\mathbb{P}}_m} \max_{z \in \mathbb{S}^1} |p(z)|$$

where $\tilde{\mathbb{P}}_m$ is the set of polynomials of degree at most m taking the value 1 at $z = 0$, and $\mathbb{S}^1$ is the unit circle. But by the maximum principle, we know that the maximum value over the closed unit disk of such polynomials must be attained on the unit circle so that the inner maximum above must be at least 1, and decrease of the residual cannot be achieved. In fact, the only polynomial that takes the max value 1 is $p_m(z) \equiv 1$. What happens when $m = n$? Still, no polynomials can have smaller max value on the unit circle than 1, however everything is lost in the estimate

$$\min_{p \in \tilde{\mathbb{P}}_n} \max_{\lambda \in \sigma(S)} |p(\lambda)| \le \min_{p \in \tilde{\mathbb{P}}_n} \max_{z \in \mathbb{S}^1} |p(z)|$$

since the spectrum of S exactly coincides with the zeros of $p_n(z) = 1 - z^n$, the continuous max is in fact equal to 2.

d) What happens if we replace GMRES by the full orthogonalization method (FOM).

Answer: FOM counts on computing $H_m^{-1}(\beta e_1)$, but H_m is singular so the method breaks down in the first step.

Problem 3 We now consider the matrix $A = I + \theta S$, $|\theta| < 1$, where S is the shift matrix defined by (1). You may need the result in the appendix (see below) for this problem.

a) Argue that there exists a diagonal matrix $\Lambda = \Lambda(\theta) \in \mathbb{C}^{n \times n}$ and a unitary matrix $X \in \mathbb{C}^{n \times n}$, not depending on θ, such that $A = X \Lambda X^H$.

Answer: S is unitary (orthogonal) and therefore normal and unitarily diagonalizable. We therefore have $S = XDX^H$ for a unitary X and a diagonal D (both X and D are complex). Then

$$A = I + \theta X D X^H = X(I + \theta D)X^H = X\Lambda(\theta)X^H, \qquad \Lambda = I + \theta D.$$

b) We look at solving the equation $Ax = b$, again by GMRES. Derive an estimate for the convergence of the residual after m iterations of the form

$$\|r_m\|_2 \le \epsilon^{(m)}(\theta)\, \|r_0\|_2, \tag{2}$$

that is, determine $\epsilon^{(m)}(\theta)$.

Answer: We know from the previous question that $A = X\Lambda X^H$ with $X^H X = I$ so that the condition number $\kappa_2(X) = 1$. Furthermore, we shall need the fact that all the eigenvalues of A are located on the circle $C(1, \theta)$ which is easily seen from the first question in the previous problem and the fact that $\sigma(I + \theta S) = 1 + \theta\sigma(S)$. As in Proposition 6.32 in Saad, we realize that

$$\|r_m\|_2 \le \epsilon^{(m)} \|r_0\|_2$$

Here $\epsilon^{(m)}$ could be taken as

$$\min_{p \in \tilde{\mathbb{P}}_m} \max_{\lambda \in \sigma(A)} |p(\lambda)| \le \min_{p \in \tilde{\mathbb{P}}_m} \max_{z \in C(1,\theta)} |p(z)| =: \epsilon^{(m)}$$

We now invoke Zarantonello's lemma as given in appendix, to conclude that

$$\epsilon^{(m)} = \theta^m.$$

c) Suppose we us a preconditioner, $B^{-1} = I - \theta S$ and consider the system

$$B^{-1}Ax = B^{-1}b$$

Find the corresponding convergence estimate as in (2) obtained by replacing A by $B^{-1}A$.

Answer: We compute $B^{-1}A = (I - \theta S)(I + \theta S) = I - \theta^2 S^2$. The eigenvalues of this matrix are of the form $\lambda_k = 1 - \theta^2 \exp(4ik\pi/n),\ k = 1, \ldots, n$, so they are located on the circle $C(1, \theta^2)$, and it follows that we can take

$$\epsilon^{(m)} = \theta^{2m}$$

Problem 4 Given an arbitrary 2×2 real symmetric matrix written in the form

$$A = \begin{bmatrix} w+z & \varepsilon \\ \varepsilon & z \end{bmatrix}.$$

a) Perform the following shifted QR step: $A - zI = QR$, $\bar{A} = RQ + zI$. Show that

$$\bar{A} = \begin{bmatrix} \bar{w}+\bar{z} & \bar{\varepsilon} \\ \bar{\varepsilon} & \bar{z} \end{bmatrix}, \qquad \bar{z} = z - \frac{\varepsilon^2 w}{w^2+\varepsilon^2}, \quad \bar{w} = w + 2\frac{\varepsilon^2 w}{w^2+\varepsilon^2}, \quad \bar{\varepsilon} = \frac{\varepsilon^3}{w^2+\varepsilon^2}.$$

Answer:

$$A - zI = \begin{bmatrix} z & \varepsilon \\ \varepsilon & 0 \end{bmatrix} = \begin{bmatrix} w/\alpha & \varepsilon/\alpha \\ \varepsilon/\alpha & -w/\alpha \end{bmatrix} \cdot \begin{bmatrix} \alpha & \varepsilon w/\alpha \\ 0 & \varepsilon^2/\alpha \end{bmatrix}, \quad \alpha = \sqrt{w^2+\varepsilon^2}$$

So we form

$$RQ + zI = \begin{bmatrix} w + z + \varepsilon^2 w/\alpha^2 & \varepsilon^3/\alpha^2 \\ \varepsilon^3/\alpha^2 & z - \varepsilon^2 w/\alpha^2 \end{bmatrix}$$

Then

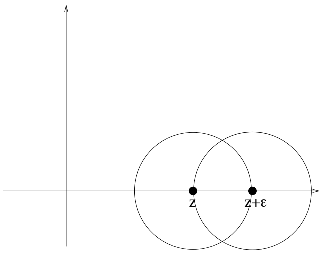
$$\bar{z} = z - \varepsilon^2 w/\alpha^2, \quad \bar{w} + \bar{z} = \bar{w} + z - \varepsilon^2 w/\alpha^2 = w + z + \varepsilon^2 w/\alpha^2$$

so that $\bar{w} = w + 2\varepsilon^2 w/\alpha^2$ etc.

b) What does the result in the previous question tell you about the convergence of the QR-iteration for this type of matrix? What happens to the convergence rate if $w \leq \varepsilon$? Draw the Gerschgorin disks for A in the case that $w = \varepsilon$ and comment on how this result compare to what you know in general about the convergence of the QR-iteration.

Answer: The interesting quantity is $\bar{\varepsilon} = \mathcal{O}(\varepsilon^3)$ unless w is small. So generally we have cubic convergence.

If $|w| \leq \varepsilon$ then $\varepsilon/2 \leq \bar{\varepsilon} \leq \varepsilon$ so the convergence in this first step is slow. However, cubic convergence will be recovered at some point. The only possibility that the eigenvalues coincide is that $\varepsilon = w = 0$ and then $A = zI$ so convergence is immediate. Since the matrix is symmetric, the eigenvalues are real and so they are located between $z - \varepsilon$ and $z + 2\varepsilon$.



Appendix. A version of Zarantonello's lemma (Saad, Lemma 6.26). Let $C(c,\rho)$ be a circle centered at c with radius ρ where $\rho < |c|$. Then

$$\min_{p\in\mathbb{P}_m,\ p(0)=1}\ \max_{z\in C(c,\rho)} |p(z)| = \left(\frac{\rho}{|c|}\right)^m$$