SF2524 Matrix Computations for Large-scale Systems
Exam - solution

Aids: None Time: Four hours

Grades: E: 16 points, D: 19 points, C: 22 points, B: 25 points, A: 28 points (out of the possible 35 points, including bonus points from homeworks).

Problem 1 (5p) Consider the linear system of equations $Ax = b$. The min-max bound for GMRES states that

$$\|Ax_n - b\| \leq \alpha \min_{p\in P_n^0} \max_{\lambda\in\lambda(A)} |p(\lambda)|$$

where α is independent of n.

(a) Suppose A is diagonalizable and the eigenvalues of A are contained in a disk of radius $\rho > 0$ centered at $c \in \mathbb{C}$ and $|c| > \rho$. Derive a formula for a constant $\beta < 1$ such that $\|Ax_n - b\| \leq \alpha\beta^n$ for all $n > 0$.

(b) The modified linear system of equations $\tilde{A}z = \tilde{b}$, where $\tilde{A} = \gamma A$ and $\tilde{b} = \gamma b$, has the same solution as $Ax = b$ for any $\gamma \neq 0$, since $x = A^{-1}b = \tilde{A}^{-1}\tilde{b} = z$. Show that GMRES applied to $\tilde{A}z = \tilde{b}$ generates the same sequence of approximations as GMRES applied to $Ax = b$.

Solution:

(a) According to the min-max bound we have, for any $p \in P_n^0$:

$$\|Ax_n - b\| \leq \alpha \max_{\lambda\in\lambda(A)} |p(\lambda)|.$$

It holds in particular for $q(z) = (c-z)^n/c^n$ because $q \in P_n^0 = \{p \text{ polynomial of degree } n : p(0) = 1\}$. Since $|c-\lambda| \leq \rho$ for any eigenvalue λ, we have

$$\|Ax_n - b\| \leq \alpha \max_{\lambda\in\lambda(A)} \frac{|c-\lambda|^n}{|c|^n} \leq \alpha \frac{\rho^n}{|c|^n}$$

and we can select $\beta = \rho/|c|$.

(b) The GMRES approximation is defined as the minimizer of $\|Ax-b\|_2$ over the set $x \in \mathscr{K}_n(A,b)$. We will first show that the Krylov subspace is the same for the pair A, b and $\tilde{A}, \tilde{b}$. From relations of span we have

$$\begin{aligned} \mathscr{K}_n(\tilde{A},\tilde{b}) &= \operatorname{span}(\tilde{b}, \tilde{A}\tilde{b}, \ldots, \tilde{A}^{n-1}\tilde{b}) \\ &= \operatorname{span}(\gamma b, \gamma^2 Ab, \ldots, \gamma^n A^{n-1}b) = \mathscr{K}_n(A,b) \end{aligned}$$

Hence, if x_n is the GMRES-approximation stemming from $Ax = b$ and $\tilde{x}_n$ is the GMRES-approximation stemming from $\tilde{A}x = \tilde{b}$, we have

$$\begin{aligned}
\|Ax_n - b\| &= \min_{x \in \mathscr{K}_n(A,b)} \|Ax - b\| \\
&= \frac{1}{|\gamma|} \min_{x \in \mathscr{K}_n(A,b)} \|\gamma Ax - \gamma b\| \\
&= \frac{1}{|\gamma|} \min_{x \in \mathscr{K}_n(\tilde{A},\tilde{b})} \|\tilde{A}x - \tilde{b}\| = \frac{1}{\gamma}\|\tilde{A}\tilde{x}_n - \tilde{b}\|
\end{aligned}$$

Problem 2 (4p) Consider a diagonalizable matrix $A \in \mathbb{R}^{(2N+1)\times(2N+1)}$ with eigenvalues

$$\begin{aligned}
\lambda_1 &= 1+\varepsilon \\
\lambda_{2k} &= 2+\sin(2k\pi/N), \;\; k = 1,\ldots,N \\
\lambda_{2k+1} &= 2+i\cos(2k\pi/N), \;\; k = 1,\ldots,N
\end{aligned}$$

(a) Suppose $\varepsilon = -0.5$. Let Q_k be the orthogonal matrix generated by k steps of Arnoldi's method. Derive a constant α such that the indicator $\|(I - Q_kQ_k^T)x_1)\|_2$ can be bounded by

$$\|(I - Q_kQ_k^T)x_1)\|_2 < \xi\alpha^k,$$

for some value $\xi > 0$, where x_1 is the eigenvector associated with eigenvalue λ_1. You do not have to specify the constant ξ and you may directly use theorems from the course.

(b) Suppose $N = 100$. To which eigenvalue will the power method converge if $\varepsilon = -2$, $\varepsilon = -0.5$, $\varepsilon = 1$, $\varepsilon = 3$? Assume that the starting vector is such that it has components in all eigenvector directions.

Solution:

(a) For $\varepsilon = -0.5$, $\lambda_1 = 0.5$ and the other eigenvalues $\lambda_2,\ldots,\lambda_N$ are contained in a disk of radius $\rho = 1$ centered at $\lambda = 2$. One can directly use a corollary in the lecture notes, which means that

$$\alpha = \frac{\rho}{|c-\lambda_1|} = \frac{1}{3/2} = 2/3$$

(Alternatively, carry out the proof of the corollary by using the main min-max bound of the Arnoldi method and selecting $p(z) = (c-z)^n/|c-\lambda_1|^n$.)

(b) The power method converges to the eigenvalue largest in magnitude, which will be different depending on ε. Note that for $N = 100$, one eigenvalue is $\lambda = 3$. The power method converges correspondingly to:

$\varepsilon = -2$: $\lambda = 3$

$\varepsilon = -0.5$: $\lambda = 3$

$\varepsilon = 1$: $\lambda = 3$

$\varepsilon = 3$: $\lambda = 4$

Problem 3 (5p) Let $f(A) = A^{1/3}$. Consider the iteration

$$X_{k+1} = \alpha X_k + \beta X_k^{-1} + \gamma X_k^{-2}$$

for $k = 1, \dots$, and $X_0 = A$ where A is symmetric positive definite. Derive constants α, β and γ such that (if the iteration converges) it converges quadratically to $f(A)$. Justify the quadratic convergence.

Solution: Analogous to the Newton-SQRT method we consider Newton's method to compute the cube root of a scalar. Let $g(x) = x^3 - a$. Newton's method is

$$x_{k+1} = x_k - \frac{g(x_k)}{g'(x_k)} = x_k - (x_k^3 - a)/(3x_k^2) = \frac{2}{3}x_k + \frac{a}{3}x_k^{-2}. \tag{1}$$

A matrix function generalization is

$$X_{k+1} = \frac{2}{3}X_k + \frac{1}{3}AX_k^{-2}$$

where $X_0 = A$. Quadratic convergence follows from the fact that the eigenvalues of X_k satisfy the iteration (1).

(More rigorously: Let $A = V^T \Lambda V$ is the Jordan form. V is orthogonal since A is symmetric. From $X_0 = A$ and induction we have $X_k = V^T \Lambda_k V$, where Λ_k is a diagonal matrix with diagonal elements that satisfy (1). Hence, $\|X_k - X_*\|_2 = \|V^T(\Lambda_k - \Lambda_*)V\|_2 = \|\Lambda_k - \Lambda_*\|_2$ which converges as fast as the slowest convergent element in Λ_k.)

Problem 4 (3p) Let $\|z\|_B := \sqrt{z^T Bz}$. Consider the linear system of equations $Ax_* = b$, where $A \in \mathbb{R}^{m\times m}$ is symmetric positive definite.

(a) Let x be an approximation of x_*. Show that $\|x - x_*\|_A = \|Ax - b\|_{A^{-1}}$.

(b) Let e_n be the error in step n of CG. Show that $\|e_{n+1}\|_A \leq \|e_n\|_A$ using the fact that the CG-iterates minimize the error in the $\|\cdot\|_A$-norm over an associated Krylov subspace.

(c) Let e_n be the error in step n of CG. A theorem in this course stated that

$$\frac{\|e_n\|_A}{\|e_0\|_A} \leq \min_{p\in P_n^0} \max_{\lambda\in\lambda(A)} |p(\lambda)|.$$

Suppose all the eigenvalues are explicitly $\lambda_1 = 10$ and $\lambda_k = 2 + 1/k$ for $k = 2, \dots, 100$. What is maximum (worst-case) error $\|e_n\|_A$ after 100 iterations?

Solution:

(a)

$$\begin{aligned}
\|x - x_*\|_A^2 &= (x - x_*)^T A(x - x_*) \\
&= (x - A^{-1}b)^T A(x - A^{-1}b) \\
&= (x - A^{-1}b)^T (Ax - b) \\
&= (Ax - b)^T A^{-1}(Ax - b) \\
&= \|Ax - b\|_{A^{-1}}^2
\end{aligned}$$

(b) The iterates of CG are minimizers of the error with respect to the A-norm over the Krylov subspace of dimension n. Therefore,

$$\|x_{n+1}-x_*\|_A = \min_{x\in\mathscr{K}_{n+1}(A,b)} \|x-x_*\|_A \leq \min_{x\in\mathscr{K}_n(A,b)} \|x-x_*\|_A = \|x_n-x_*\|_A$$

since $\mathscr{K}_n(A,b)\subset\mathscr{K}_{n+1}(A,b)$.

(c) The matrix has 100 eigenvalues, and is therefore of size 100. After 100 iterations, the Krylov subspace will be the same as the dimension of the problem and the Krylov subspace spans the entire $\mathbb{R}^{100}$. The worst-case error is zero.

Problem 5 (3p) Let $Q=(q_1,\ldots,q_n)\in\mathbb{R}^{m\times n}$ be an orthogonal matrix. Suppose $b\in\mathbb{R}$ is such that $b\notin\operatorname{span}(q_1,\ldots,q_m)$.

(a) Derive the Gram-Schmidt procedure by computing explicit formulas for $h=(h_1,\ldots,h_n)\in\mathbb{R}^n$ and $q_{n+1}\in\mathbb{R}^m$ such that $Q^Tq_{n+1}=0$, $\|q_{n+1}\|=1$, $\operatorname{span}(q_1,\ldots,q_{n+1})=\operatorname{span}(q_1,\ldots,q_n,b)$ and

$$b=h_1q_1+\ldots+h_nq_n+\gamma q_{n+1}.$$

Express the procedure using only products of matrices and vectors (no for-loops).

(b) Describe the double Gram-Schmidt procedure (any version). What are the advantages and disadvantages of classical Gram-Schmidt and double Gram-Schmidt?

Solution:

(a) We want to establish $h\in\mathbb{R}^n$, γ and q_{n+1} such that

$$b=Qh+\gamma q_{n+1}$$

where $Q^Tq_{n+1}=0$ and $\|q_{n+1}\|=1$. We can multiply b by Q^T,

$$Q^Tb=Q^TQh+\gamma Q^Tq_{n+1}$$

and find that $h=Q^Tb$. We can now compute w with

$$w:=b-Qh.$$

We can now set

$$\begin{aligned}\gamma &= \|w\|\\ q_{n+1} &= \frac{1}{\gamma}w\end{aligned}$$

which gives a computational formula for h, γ and q_{n+1}.

(b) Double Gram-Schmidt essentially corresponds to carrying out the steps in (a) twice:

$$\begin{aligned} h &= Q^T b \\ w &= b - Qh \\ g &= Q^T w \\ w &= w - Qg \\ h &= h + g \end{aligned}$$

Classical Gram-Schmidt can be carried out with essentially half as many operations. Double Gram-Schmidt is in general less sensitive to round-off errors (better numerical stability).

Problem 6 (3p) Let

$$A = \begin{pmatrix} \alpha & \times & \times \\ \beta & \times & \times \\ \gamma & \times & \times \end{pmatrix}$$

Derive a formula (involving α, β and γ) for an orthogonal matrix Q such that QAQ^T has the structure

$$QAQ^T = \begin{pmatrix} \times & \times & \times \\ \times & \times & \times \\ 0 & \times & \times \end{pmatrix}$$

Solution: The derivation is analogous to the first step in the Hessenberg reduction of the QR-method. We let,

$$Q = \begin{pmatrix} 1 & 0 \\ 0 & P \end{pmatrix}$$

where P is a Householder reflector defined by

$$P = I - \frac{2}{p^T p} p p^T$$

and

$$p = \begin{pmatrix} \beta \\ \gamma \end{pmatrix} - \begin{pmatrix} \sqrt{\beta^2 + \gamma^2} \\ 0 \end{pmatrix}.$$

By construction, P satisfies

$$P \begin{pmatrix} \beta \\ \gamma \end{pmatrix} = \delta e_1$$

Hence,

$$QAQ^T = \begin{pmatrix} \alpha & \times & \times \\ \delta & \times & \times \\ 0 & \times & \times \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & P^T \end{pmatrix} = \begin{pmatrix} \alpha & \times & \times \\ \delta & \times & \times \\ 0 & \times & \times \end{pmatrix}.$$

Problem 7 (6p) A matrix is called symmetric if $A^T = A$ and anti-symmetric if $A^T = -A$. Suppose more generally that $A \in \mathbb{R}^{m \times m}$ satisfies $A^T = \alpha A$ for some value $\alpha \neq 0$. Let $Q_{n+1} = (q_1, \ldots, q_{n+1}) \in \mathbb{R}^{m \times (n+1)}$ and $\underline{H}_n \in \mathbb{R}^{(n+1) \times n}$ correspond to an Arnoldi factorization for A and let $H_n \in \mathbb{R}^{n \times n}$ be the top block of $\underline{H}_n$.

(a) Show that $H_n^T = \alpha H_n$. Specify which elements of H_n will be zero (for any starting vector and any A satisfying $A^T = \alpha A$). Separate between the two cases $\alpha = 1$ and $\alpha \neq 1$.

(b) Show that for any $k > 1$ there exist c_{k-1}, a_k and b_{k+1} such that

$$Aq_k = c_{k-1}q_{k-1} + a_k q_k + b_{k+1}q_{k+1}.$$

(c) Derive a generalization of the Lanczos procedure for matrices satisfying $A^T = \alpha A$ by deriving formulas for the Arnoldi factorization corresponding to Q_{n+2}, $\underline{H}_{n+1}$ expressed in terms of the Arnoldi factorization corresponding to Q_{n+1}, $\underline{H}_n$. The procedure should not be more computationally expensive than one step of the Lanczos procedure.

Solution:

(a) Suppose $AQ_n = Q_{n+1}\underline{H}_n$ is an Arnoldi factorization. Let $H_n \in \mathbb{R}^{n\times n}$ denote the leading (top) submatrix of $\underline{H}_n \in \mathbb{R}^{(n+1)\times n}$. From the orthogonality of Q_n we have $H_n = Q_n^T A Q_n$ and consequently,

$$H_n^T = (Q_n^T A Q_n)^T = Q_n^T A^T Q_n = \alpha Q_n^T A Q_n = \alpha H_n.$$

Two cases:

$\alpha = 1$: Since H_n is a Hessenberg matrix, symmetry implies that $H_n^T = H_n$ is also a Hessenberg matrix. Hence, H_n is a tridiagonal matrix:

$$H_n = \begin{pmatrix} \times & \times & & \\ \times & \ddots & \ddots & \\ & \ddots & \ddots & \times \\ & & \times & \times \end{pmatrix}$$

$\alpha \neq 1$: Since the diagonal is not changed when transposing a matrix, the diagonal elements of H_n must satisfy $h_{i,i} = \alpha h_{i,i}$ and $(1-\alpha)h_{i,i} = 0$. Since $\alpha \neq 1$, we must have $h_{i,i} = 0$ and the structure is

$$H_n = \begin{pmatrix} 0 & \times & & \\ \times & \ddots & \ddots & \\ & \ddots & \ddots & \times \\ & & \times & 0 \end{pmatrix}$$

(Good comment from student: If $\alpha \neq 1$, we have that $A^T = \alpha A$, and $A = (A^T)^T = \alpha A^T = \alpha^2 A$. Hence, if A is not a zero matrix, $\alpha \neq 1$ actually implies that we must have $\alpha = -1$.)

(b) Consider column k (where $k \leq n$) of the Arnoldi factorization $AQ_n = Q_{n+1}\underline{H}_n$:

$$Aq_k = \sum_{i=0}^{k+1} h_{i,k} q_i.$$

From (a) we know that $h_{i,k} = 0$ if $i < k-1$ or $i > k+1$. Hence,

$$Aq_k = \sum_{i=k-1}^{k+1} h_{i,k} q_i = h_{k-1,k}q_{k-1} + h_{k,k}q_k + h_{k+1,k}q_k,$$

which proves the question. In addition, if $\alpha \neq 1$ the term involving $h_{k,k}$ vanishes.

(c) Let $\underline{H}_n$ and Q_{n+1} be given and denote

$$\underline{H}_n = \begin{pmatrix} a_0 & c_0 & & \\ b_1 & \ddots & \ddots & \\ & \ddots & \ddots & c_{n-1} \\ & & b_n & a_n \\ & & & b_{n+1} \end{pmatrix}.$$

We want to compute $\underline{H}_{n+1}$ and Q_{n+2}, which in the notation above can be seen as deriving computational formulas for c_n, a_{n+1} and b_{n+2} and q_{n+2}. We directly compute c_n from symmetry

$$c_n = \alpha b_{n+1}.$$

The rest of the derivation follows the same steps as the derivation of the Lanczos procedure. From the existance result in (b) we know that

$$w = Aq_{n+1} = c_n q_n + a_{n+1} q_{n+1} + b_{n+2} q_{n+2}. \tag{2}$$

By multiplication from the left with q_{n+1}^T, we have

$$q_{n+1}^T w = c_n q_{n+1}^T q_n + a_{n+1} q_{n+1}^T q_{n+1} + b_{n+2} q_{n+1}^T q_{n+2}$$

and from the orthogonality of Q_{n+2} it follows that $q_{n+1}^T q_n = 0$, $q_{n+1}^T q_{n+1} = 1$, $q_{n+1}^T q_{n+2} = 0$. Hence,

$$a_{n+1} = q_{n+1}^T w.$$

Since c_n, a_{n+1} can now be considered known, we can solve (2) for q_{n+2}, yielding $q_{n+2} = \frac{1}{b_{n+2}}(w - c_n q_n - a_{n+1} q_{n+1})$, which is satisfied if we define (what is called the orthogonal complement)

$$w_\perp := w - c_n q_n - a_{n+1} q_{n+1}$$

and subsequently normalize by setting $b_{n+2} := \|w_\perp\|$ and

$$q_{n+2} := w_\perp / b_{n+2}$$

This can be summarized in the following generalization/variation of the Lanczos algorithm:

1. $c_n = \alpha b_{n+1}$
2. $w = Aq_{n+1}$
3. $a_{n+1} = q_{n+1}^T w$
4. $w_\perp = w - c_n q_n - a_{n+1} q_{n+1}$
5. $b_{n+2} := \|w_\perp\|$
6. $q_{n+2} = w_\perp / b_{n+2}$