Fall 2022 Numerical Analysis MS/PhD Qualifying Examination

Instructions

- *Please write your code number (not your name) on each work sheet.*
- *Always provide concise answers with justification.*
- *Each problem needs to start on a new sheet of paper.*
- *There are five problems in total, and each problem is worth 20 points.*

1. (20 points) Let $A \in \mathbb{R}^{m\times n}$ for $m > n$, and recall that the first singular value of A is defined as

$$\sigma_1(A) = \max_{x\in\mathbb{R}^n,\, \|x\|_2=1} \|Ax\|_2.$$

(a) Give a similar definition for $\sigma_n(A)$.

(b) With $y \in \mathbb{R}^m$, define $\bar{A} = [A, y] \in \mathbb{R}^{m\times(n+1)}$. Show that

$$\sigma_1(\bar{A}) \geq \sigma_1(A) \qquad \text{and} \qquad \sigma_{n+1}(\bar{A}) \leq \sigma_n(A).$$

2. (20 points) Let us use the normalized power method to compute an approximate eigenvector

$$v_{k+1} = \frac{Av_k}{\|Av_k\|},$$

where $\|v_0\| = 1, v_0 \in \mathbb{R}^m$ and the matrix $A \in \mathbb{R}^{m\times m}$ is symmetric $A = A^T$.

(a) Let $|\lambda_1| = |\lambda_2| > |\lambda_3| \geq \cdots \geq |\lambda_m|$ be the m eigenvalues of A with corresponding eigenvectors $q_1, q_2, \ldots, q_m$. Also let the initial vector v_0 satisfy $q_1^* v_0 \neq 0$ and $q_2^* v_0 \neq 0$.
What vector(s) will v_k converge to? Justify your answer.

(b) Next, let $q_1^* v^0 \neq 0$ but $q_2^* v^0 = 0$.
What vector(s) will v_k converge to in exact arithmetic? Justify your answer.
What vector(s) will v_k converge to in floating point arithmetic? Justify your answer.

(c) Derive the rate of convergence of the Rayleigh quotient $r(v_k) = v_k^T A v_k$ to λ_1 when $|\lambda_1| > |\lambda_i|$, for all $i > 1$.

3. (20 points) This problem considers the solution to the following boundary value problem (BVP):

$$\begin{aligned} -p''(x) + p(x) &= g(x), \qquad x \in (0,1) \\ p(0) = p(1) &= 0 \end{aligned} \tag{1}$$

where $p''(x)$ denotes the second derivative of the function $p(x)$ with respect to x.

(a) Given a smooth function f, its second derivative may be approximated by the following finite difference formula:

$$f''(x) \approx \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}$$

where h is some small positive real number. Show that if f has bounded derivatives, then the error in this approximation satisfies,

$$\left| f''(x) - \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} \right| = O(h^2)$$

(b) Now we use the approximation from part (a) for the solution of BVP (1). Given a positive integer N, let $h = 1/(N+1)$, and $x_i = ih$ for $i = 0, 1, \cdots, N+1$. Let $g_i = g(x_i)$ and the approximations to the solution at the points x_i be $p_i \approx p(x_i)$. Then, using the finite difference approximation to the second derivative, we obtain the following set of equations:

(2)
$$\frac{-p_{i+1} + 2p_i - p_{i-1}}{h^2} + p_i = g_i, \quad i = 1, 2, \cdots, N$$

where $p_0 = p_{N+1} = 0$ (by using the boundary values from (1)). This set of equations may be written as a linear system $Ax = b$, where $x = [p_1, p_2, \cdots, p_{N-1}]^T$. Write out the matrix A and vector b for this linear system.

(c) Show that the matrix A in part (b) is symmetric positive definite. You may make an assumption on h if needed, but specify that assumption clearly.

(d) Name two numerical algorithms which may **only** be used to solve symmetric positive definite linear systems (that is, linear systems $Bx = b$ such that B is symmetric positive definite). For large N (say in the billions), which of these two methods would you recommend and why?

4. (20 points) Let ϵ denote machine precision (or unit roundoff), $\|\cdot\|$ the infinity norm, $\kappa(\cdot)$ the infinity norm condition number, and I the identity matrix. Consider the solution of

$$Ax = b$$

using Gaussian Elimination with partial pivoting (GEPP), where $A \in \mathbb{R}^{n\times n}$ is an invertible matrix, and x and b are vectors in $\mathbb{R}^n$. Assume the matrix A is such that no pivoting is needed during GEPP. GEPP first computes the LU-factorization of A in floating point arithmetic, which is equivalent to computing a unit lower triangle matrix $\hat{L}$ and an upper triangular matrix $\hat{U}$ such that

$$A + E = \hat{L}\hat{U}$$

where $\|E\| \le O(\epsilon)\|\hat{L}\|\|\hat{U}\|$.

Then we solve $\mathrm{fl}\left(\hat{L}\hat{y} = b\right)$, where $\mathrm{fl}(\cdot)$ indicates that the arithmetic is carried out in floating point (that is, $\hat{y}$ is the numerically computed solution using floating point arithmetic to $\hat{L}y = b$). The solution $\hat{y}$ satisfies

$$(\hat{L} + F)\hat{y} = b$$

for some $F \in \mathbb{R}^{n\times n}$. Then, we carry out back-substitution in floating point: $\mathrm{fl}\left(\hat{U}\hat{x} = \hat{y}\right)$ to get the computed solution $\hat{x}$. The solution $\hat{x}$ satisfies

$$(\hat{U} + G)\hat{x} = \hat{y}$$

for some $G \in \mathbb{R}^{n\times n}$. Since forward and backward substitution are backward stable, we have $\|F\| = O(\epsilon)\|\hat{L}\|$ and $G = O(\epsilon)\|\hat{U}\|$.

Assume that $\|\hat{U}\| = c\|A\|$ for some small constant c, and that A is well conditioned (and hence $\kappa(A) = O(1) \ll 1/\epsilon$). $O(\epsilon)$ may have a factor of n inside. This is fine, that is, treat n as a constant and so $n \in O(1)$.

(a) Show that $\|\hat{L}\| \leq n$ (and hence $\|L\| = O(1)$ since we consider n as a constant).

(b) Show that the computed solution $\hat{x}$ satisfies $(A + H)\hat{x} = b$ such that $H = O(\epsilon)\|A\|$. This shows that GEPP is backward stable with the given assumptions. (Hint: Start with $(\hat{L} + F)\hat{y} = b$).

(c) We have a result which states: If square matrix X satisfies $\|X\| < 1$, then $I - X$ is invertible and $\|(I - X)^{-1}\| \leq (1/(1 - \|X\|)$. Use this result to show that $(A + H)$ is invertible.

(d) Show that the error in the solution satisfies $x - \hat{x} = -(I + A^{-1}H)^{-1}A^{-1}Hx$, and the relative error satisfies the bound:

$$\frac{\|x - \hat{x}\|}{\|x\|} \leq \frac{\kappa(A)}{1 - \kappa(A)\frac{\|H\|}{\|A\|}} \left(\frac{\|H\|}{\|A\|}\right)$$

(e) Argue that if A is well-conditioned, then the relative error satisfies $\frac{\|x - \hat{x}\|}{\|x\|} = O(\epsilon)$.

5. (20pt) Let $A \in \mathbb{R}^{n\times n}$ be symmetric positive definite (SPD) and consider using the Conjugate Gradients (CG) algorithm to solve the linear system $Ax = b$. CG is well-known to minimize the A-norm of the error $x^* - x$ over the Krylov space $x_0 + \mathcal{K}_k(A, b)$, where $\mathcal{K}_k(A, b) = \text{span}(A^0b, Ab, A^2b, \ldots, A^{k-1}b)$, x_0 is the initial guess to the linear system, and x^* is the exact solution. More precisely, we can say

$$x_k = \min_{x\in x_0+\mathcal{K}_k(A,b)} \|x^* - x\|_A.$$

Remember that $\|y\|_A = \sqrt{y^TAy}$ is the definition of the A-norm.

Next, let $B \in \mathbb{R}^{n\times n}$ be a general nonsymmetric matrix whose inverse exists. We want to use CG to solve a linear system

$$By = f,$$

where $y, f \in \mathbb{R}^n$ and y_0 is some initial guess. To consider CG, we turn our attention to the matrices B^TB and BB^T.

(a) Show that B^TB and BB^T are both SPD. Thus, we can apply CG to the two systems $B^TBy = B^Tf$ and $BB^Tz = f$.

(b) Let k iterations of CG be applied to $B^TBy = B^Tf$ with initial guess y_0, yielding the approximate solution y_k. Show that the 2-norm of the residual $r = f - By_k$ is minimized over the Krylov space $y_0 + \mathcal{K}_k(B^TB, B^Tf)$.

(c) Let k iterations of CG be applied to $BB^Tz = f$ with initial guess z_0, yielding the approximate solution z_k. Show that the resulting approximate solution to the original system $y_k = B^Tz_k$ minimizes the 2-norm of the error $\|y^* - y_k\|$ over the Krylov space $z_k \in z_0 + \mathcal{K}_k(BB^T, f)$. The quantity y^* is the exact solution to $By = f$.