Test 2 Solution

Solve the following problems (6 course points each). Present a brief motivation of your method of solution.

1. Consider a computer satisfying the floating point arithmetic axiom $x \circledast y = (x * y)(1+\epsilon)$ for all $x, y \in \mathbb{F} \subset \mathbb{R}$ (set of real floating point numbers), with machine epsilon denoted by ϵ, $\circledast$ a floating point operation, $*$ the corresponding real number operation. Also consider construction of the Newton interpolating poynomial

$$p_n(x) = y_0 + [y_1, y_0](x - x_0) + \cdots + [y_n, y_{n-1}, ..., y_0](x - x_0)...(x - x_{n-1}) = a_0 + a_1 x + \cdots + a_n x^n$$

of data $\mathcal{D} = \{(x_k, y_k), x_k = kh, k = 0, ..., n\}$, $h = 1/n$, $n \in \mathbb{N}_+$, with divided differences defined by $[y_k] = y_k$,

$$[y_k, y_{k-1}, ..., y_{k-l}] = \frac{[y_k, y_{k-1}, ..., y_{k-l+1}] - [y_{k-1}, y_{k-2}, ..., y_{k-l}]}{x_k - x_{k-l}} = \frac{[y_k, y_{k-1}, ..., y_{k-l+1}] - [y_{k-1}, y_{k-2}, ..., y_{k-l}]}{lh}.$$

   a) What is the condition number of the problem $\mathcal{D} \to^f a_n$?

   b) Estimate the error δa_n produced by error δy_j in j^{th} data measurement, i.e., $\tilde{y_j} = y_j + \delta y_j$.

   c) What is the condition number of the problem $h \to^g a_n$?

   d) Is the evaluation of $g(h) = a_n(h)$ well-conditioned, ill-conditioned or ill-posed? Consider limiting values of the sampling step size h.

   Solution.

   a) Note that $a_n = [y_n, y_{n-1}, ..., y_0] = f(\mathcal{D})$, and is linear in $\boldsymbol{y}$, hence $a_n = \boldsymbol{F}\boldsymbol{y}$, with $\boldsymbol{F} \in \mathbb{R}^{1\times(n+1)}$ the matrix encoding the linear mapping f. The condition number of the problem is $\kappa_f = \kappa(\boldsymbol{F})$. (Full credit, the question asks: "do you recognize linear dependence and recall the condition number of matrix-vector multiplication?").

   Note. The standard procedure to find the matrix encoding a linear mapping is to apply the mapping to the standard basis vectors $\boldsymbol{e}_0, \boldsymbol{e}_1, ..., \boldsymbol{e}_n$, i.e.,

$$\boldsymbol{F} = [\ f(\boldsymbol{e}_0)\ \ f(\boldsymbol{e}_1)\ \ \ldots\ \ \boldsymbol{f}(\boldsymbol{e}_n)\ ].$$

   Let $a_{n,i} = f(\boldsymbol{e}_i)$. When $\boldsymbol{y} = \boldsymbol{e}_i$, $p_{n,i}(x)$ is the interpolating polynomial with roots $x_j = jh$, for $j = 0, ..., n$ but $j \neq i$, hence has form

$$p_{n,i}(x) = a_{n,i} \prod_{j=0, i\neq j}^{n} (x - jh).$$

   Evaluation at x_i gives

$$p_{n,i}(ih) = a_{n,i} h^n \prod_{j=0, i\neq j}^{n} (i - j) = 1 \Rightarrow a_{n,i} = \left[h^n \prod_{j=0, i\neq j}^{n} (i - j) \right]^{-1} = n^n \left[\prod_{j=0, i\neq j}^{n} (i - j) \right]^{-1}.$$

   The smallest value of the product is $(-1)^{n/2}(n/2)^2$, and arises at $i = n/2$, hence

$$\max |a_{n,i}| = 4n^{n-2} = \|\boldsymbol{F}\|.$$

   b) $|\delta a_n| \leqslant \kappa_f\, |\delta y_j| \leqslant \kappa(\boldsymbol{F})\, |\delta y_j|$.

   c) From divided difference recurrence formula, deduce

$$a_n = A\frac{n^n}{n!} = G(n),$$

   with A some coefficient, and $a_n = g(h) = G(1/h)$ differentiable, hence condition number is given by derivative

$$\kappa_g = \frac{|g'|}{|g|/h},\ g'(h) = -\frac{1}{h^2} G'\left(\frac{1}{h}\right),\ G'(n) = A\frac{\mathrm{d}}{\mathrm{d}n}\left(\frac{n^n}{n!}\right).$$

(Full credit, questions asks: "do you recognize nonlinear dependence and condition number evaluation through Jacobian?")

Notes.

i. The coefficient A is known from divided difference calculus

$$A = \Delta^n y_0 = \begin{pmatrix} n \\ 0 \end{pmatrix} y_n - \begin{pmatrix} n \\ 1 \end{pmatrix} y_{n-1} + \begin{pmatrix} n \\ 2 \end{pmatrix} y_{n-2} - \cdots, \begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{k!(n-k)!}.$$

ii. $n^n/n! > 1$, and can be estimated using Stirling's formula (for large n)

$$\ln n! \cong n \ln n - n \Rightarrow \ln \frac{n^n}{n!} = n \Rightarrow \frac{n!}{n^n} \cong e^n.$$

This gives

$$G'(n) = Ae^n = Ae^{1/h}, \kappa_g = \frac{|A|\, e^{1/h}}{h^2}.$$

d) As $h \to 0$, $e^{1/h} h^{-2}$ rapidly increases and the problem is ill-conditioned.

2. Let $\boldsymbol{A} \in \mathbb{R}^{m \times m}$ denote the matrix obtained by second-order accurate, centered finite difference approximation of the Helmholtz equation $\nabla^2 u = -k^2 u$ in $(0,1) \times (0,1)$ with periodic boundary conditions $u(x+p, y+q) = u(x,y)$, $p, q \in \mathbb{Z}$, $h = 1/n$, $n \in \mathbb{N}_+$, $m = n^2$, leading to the eigenvalue problem $\boldsymbol{A}\boldsymbol{u} = -h^2 k^2 \boldsymbol{u}$.

$$(\nabla^2 u)_{ij} \cong \frac{u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 4u_{i,j}}{h^2} \Rightarrow \boldsymbol{A} = \mathrm{diag}([0, ..., 1, ..., 1, -4, 1, ..., 1, ...0])$$

a) Present an algorithm to reduce $\boldsymbol{A}$ to symmetric Hessenberg form $\boldsymbol{H}$ using Householder reflectors that preserves eigenvalues of $\boldsymbol{A}$.

b) Is the algorithm accurate in floating point arithmetic?

c) Is the algorithm forward stable?

d) Is the algorithm backward stable?

Provide either an analysis or a qualitative motivation using established theorems for answers to (b)-(d).

Solution.

a) Eigenvalues are preserved by similarity transformations $\boldsymbol{A} \sim \boldsymbol{T}\boldsymbol{A}\boldsymbol{T}^{-1}$, or $\boldsymbol{A} \sim \boldsymbol{Q}\boldsymbol{A}\boldsymbol{Q}^T$ for $\boldsymbol{A}$ with real elements. Start from the standard similarity reduction to Hessenberg form through Householder reflectors.

Input: $\boldsymbol{A} \in \mathbb{R}^{m \times m}$
Output:
for $k = 1$ to $m-2$
  $\boldsymbol{x} = A_{k+1:m,k} \in \mathbb{R}^{m-k}$
  $\boldsymbol{v}_k = \mathrm{sign}(x_1) \|\boldsymbol{x}\|_2 \boldsymbol{e}_1 + \boldsymbol{x}$
  $\boldsymbol{v}_k = \boldsymbol{v}_k / \|\boldsymbol{v}_k\|$
  $A_{k+1:m,k:m} = A_{k+1:m,k:m} - 2\boldsymbol{v}_k (\boldsymbol{v}_k^T A_{k+1:m,k:m})$
  $A_{1:m,k+1:m} = A_{1:m,k+1:m} - 2 (A_{1:m,k+1:m} \boldsymbol{v}_k) \boldsymbol{v}_k^T$

Note that $\boldsymbol{A}$ is symmetric and banded with semi-bandwidth n, and eliminating redundant multiplications with zero gives

Input: $\boldsymbol{A} \in \mathbb{R}^{m \times m}$
Output:
for $k = 1$ to $m-2$
  $p = \min(k+n, m)$
  $x_{1:n} = A_{k+1:p,k}$
  $\boldsymbol{v}_k = \mathrm{sign}(x_1) \|\boldsymbol{x}\|_2 \boldsymbol{e}_1 + \boldsymbol{x}$ (or $\boldsymbol{v}_k = \boldsymbol{e}_1 - \boldsymbol{x}$)

$$\boldsymbol{v}_k = \boldsymbol{v}_k / \|\boldsymbol{v}_k\|$$
$$A_{k+1:p,k:p} = A_{k+1:p,k:p} - 2\boldsymbol{v}_k\,(\boldsymbol{v}_k^T A_{k+1:p,k:p})$$
$$A_{1:m,k+1:p} = A_{1:m,k+1:p} - 2\,(A_{k+1:m,k:m}\,\boldsymbol{v}_k)\boldsymbol{v}_k^T$$

b) An algorithm $\tilde{f}$ to solve problem f is accurate if the relative error is of order machine epsilon ϵ

$$\frac{\|\tilde{f}(x) - f(x)\|}{\|f(x)\|} = \mathcal{O}(\epsilon).$$

The problem is to reduce $\boldsymbol{A}$ to Hessenberg form through an orthogonal similarity transformation

$$\boldsymbol{A} \xrightarrow{f} (\boldsymbol{Q}, \boldsymbol{H}), \boldsymbol{A} = \boldsymbol{Q}\boldsymbol{H}\boldsymbol{Q}^T.$$

The algorithm constructs an approximation $\boldsymbol{Q}$ through Householder reflectors, $\tilde{\boldsymbol{Q}} = \boldsymbol{Q}_1\boldsymbol{Q}_2 \cdot \boldsymbol{Q}_{m-2}$, and is accurate if it is backward stable, with error estimate (cf. Theorem 15.1)

$$\frac{\|\tilde{f}(x) - f(x)\|}{\|f(x)\|} = \mathcal{O}(\kappa(x)\epsilon),$$

where κ is the condition number of the problem.

c) Forward stability is defined as

$$\frac{\|\tilde{f}(x) - f(\tilde{x})\|}{\|f(\tilde{x})\|} = \mathcal{O}(\epsilon) \text{ for some } \tilde{x} \text{ such that } \frac{\|\tilde{x} - x\|}{\|x\|} = \mathcal{O}(\epsilon).$$

Note that $\boldsymbol{A}$ is symmetric hence unitarily diagonalizable, and the singular values of $\boldsymbol{A}$ are the absolute values of the eigenvalues $\sigma_i = |\lambda_i|$, with $|\lambda_1| \geqslant \cdots \geqslant |\lambda_m|$. The condition number of $\boldsymbol{A}$ is $\kappa(\boldsymbol{A}) = |\lambda_1/\lambda_m|$. From $\boldsymbol{A}\boldsymbol{u} = -h^2k^2\,\boldsymbol{u}$ note that $k = 0$, $\boldsymbol{u} = 1$ is a solution of the eigenvalue problem, hence $\lambda_m = 0$, implying $\kappa(\boldsymbol{A}) \to \infty$, the problem is ill-conditioned and the algorithm $\boldsymbol{A} \xrightarrow{\tilde{f}} (\tilde{\boldsymbol{Q}}, \tilde{\boldsymbol{H}})$ is not forward stable

d) Backward stability is defined as

$$\tilde{f}(x) = f(\tilde{x}) \quad \text{for some } \tilde{x} \text{ such that } \frac{\|\tilde{x} - x\|}{\|x\|} = \mathcal{O}(\epsilon).$$

As in Householder triangularization, reduction to symmetric Hessenberg form is backward stable in the sense that

$$\tilde{\boldsymbol{Q}}\,\tilde{\boldsymbol{H}}\,\tilde{\boldsymbol{Q}}^T = \boldsymbol{Q}\boldsymbol{H}\boldsymbol{Q}^T = \tilde{\boldsymbol{A}}, \text{for some } \tilde{\boldsymbol{A}} \text{ such that } \|\tilde{\boldsymbol{A}} - \boldsymbol{A}\| = \mathcal{O}(\epsilon).$$