Section exam 2 Numerical Analysis: Linear Algebra — Solutions

Oct. 15, 2026. 75 minutes.

Instructions:

- ***This is a closed-books exam.*** No books or calculators are allowed. However, you may bring one hand-written, single-sided formula sheet.
- Try to answer the questions within the space given. If you need extra pages, attach them and indicate clearly which question they address.
- If a question does not specifically ask for a justification, the correct answer alone earns full credit. You are welcome to include a brief explanation, which may help you receive partial credit (except for Question 1, where only the answer will be graded).
- Use your time strategically. You may wish to save problems 2(c), 4(d), and 5 for last – they do not provide many points in relation to the amount of work required.
- Unless otherwise indicated, $\|\mathbf{x}\|$ denotes the Euclidean norm, $\|\mathbf{A}\|$ the corresponding operator norm, and $\|\mathbf{A}\|_{\mathrm{F}}$ the Frobenius norm. The notation $\mathbf{A}^*$ denotes conjugate transpose. We use $\kappa(\mathbf{A})$ to denote the Euclidean norm matrix condition number, so that $\kappa(\mathbf{A}) = \sigma_{\max}(\mathbf{A})/\sigma_{\min}(\mathbf{A})$.

Question 1: (30p) For every part of this question, give only your answers. Derivations or explanations are not required, and will not be graded.

(a) (10p) Consider a hypothetical 26-bit binary floating-point format, with one sign bit, five exponent bits, and twenty stored fraction bits. Assume the format uses the same exponent conventions as IEEE floating-point arithmetic.

(i) (3p) Specify, as a power of two, the unit roundoff.

With an implicit leading bit, the significand has $p = 21$ binary digits. Hence the unit roundoff (rounding to nearest) is

$$\boxed{\epsilon_{\text{mach}} \approx 2^{-21}}.$$

(ii) (4p) Specify, as a power of two, the magnitude of the smallest and largest positive numbers that you can accurately represent in this format.

The IEEE exponent bias is 15; the normal exponent range is $-14, \ldots, 15$, so

$$\boxed{f_{\min} \approx 2^{-14}}, \qquad \boxed{f_{\max} \approx 2^{16}}.$$

(This was graded quite generously, you did not need the exact numbers listed.)

(iii) (3p) Approximately how many decimal significant digits can you represent?

Since $2^{10} = 1024$, you get about three correct decimal digits for every 10 binary bits. Since we have 20 binary bits here, we expect slightly in excess of $\boxed{\text{6 correct digits}}$.

(If you had a calculator, you would simply evaluate $21 \times \log_{10} 2 \approx 6.32$.)

(b) (5p) Let $\mathbf{A} \in \mathbb{R}^{3\times 3}$ be a matrix. You know that

$$\sup_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\| = 4, \qquad \inf_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\| = 1, \qquad \|\mathbf{A}\|_{\text{F}} = \sqrt{21}.$$

Specify the three singular values of $\mathbf{A}$, in decreasing order.

The sup statement tells us that $\sigma_1 = 4$, and the inf statements says that $\sigma_3 = 1$. Since $\sigma_1^2 + \sigma_2^2 + \sigma_3^2 = \|\mathbf{A}\|_{\text{F}}^2 = 21$, we find $\sigma_2^2 = 21 - 16 - 1 = 4$. Thus

$$\boxed{\sigma_1(\mathbf{A}) = 4, \qquad \sigma_2(\mathbf{A}) = 2, \qquad \sigma_3(\mathbf{A}) = 1}.$$

(c) (15p) The following matrices satisfy $\mathbf{A} = \mathbf{Q}\mathbf{R}$, where $\mathbf{Q}$ is orthogonal and $\mathbf{R}$ is upper triangular. Each cross represents one missing scalar. Specify these missing entries.

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 2 & 4 \\ 1 & 0 & 0 & 2 \\ 1 & 2 & 0 & 0 \\ 1 & 0 & \times & 2 \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & 1/2 & -1/2 \\ 1/2 & 1/2 & -1/2 & -1/2 \\ 1/2 & -1/2 & -1/2 & \times \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} 2 & 2 & 0 & \times \\ 0 & 2 & 2 & 0 \\ 0 & 0 & 2 & \times \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

Since $\mathbf{A} = \mathbf{Q}\mathbf{R}$, we find $\mathbf{A}(4,3) = \mathbf{Q}(4,:)\mathbf{R}(:,3) = -2$.

Since $\mathbf{Q}$ is unitary, its columns are orthogonal. Using that $\mathbf{q}_1 \cdot \mathbf{q}_4 = 0$, we obtain $1/4 - 1/4 - 1/4 + (1/2)\mathbf{Q}(4,4) = 0$, which gives $\mathbf{Q}(4,4) = 1/2$.

Since $\mathbf{R} = \mathbf{Q}^*\mathbf{A}$ we find that $\mathbf{R}(1,4) = \mathbf{q}_1^*\mathbf{a}_4 = 4$ and $\mathbf{R}(3,4) = \mathbf{q}_3^*\mathbf{a}_4 = 2$.

Question 2: (15p) Consider two matrices $\mathbf{A}_1, \mathbf{A}_2 \in \mathbb{R}^{4\times 3}$ with reduced QR factorizations $\mathbf{A}_1 = \mathbf{Q}\mathbf{R}_1$ and $\mathbf{A}_2 = \mathbf{Q}\mathbf{R}_2$, where

$$\mathbf{Q} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & -1 \end{bmatrix}.$$

The columns of $\mathbf{Q}$ are orthonormal. The matrices and their triangular factors are

$$\mathbf{A}_1 = \begin{bmatrix} 3 & 2 & 1 \\ 3 & -2 & 1 \\ 3 & 2 & -1 \\ 3 & -2 & -1 \end{bmatrix}, \quad \mathbf{R}_1 = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad \mathbf{A}_2 = \frac{1}{2}\begin{bmatrix} 1 & 2 & 2 \\ 1 & 0 & 2 \\ 1 & 2 & -2 \\ 1 & 0 & -2 \end{bmatrix}, \quad \mathbf{R}_2 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

No proofs or derivations are required.

(a) (6p) Specify the singular values of $\mathbf{A}_1$, in decreasing order.

(b) (4p) Specify $\kappa(\mathbf{A}_1)$ and $\kappa(\mathbf{A}_1^*\mathbf{A}_1)$.

(c) (5p) Specify the singular values of $\mathbf{A}_2$, in decreasing order.

Solution: Recall from the homework problem that in a tall QR factorization $\mathbf{A} = \mathbf{Q}\mathbf{R}$, the singular values of $\mathbf{A}$ and $\mathbf{R}$ are identical. (If you did not remember this, then you could verify that $\mathbf{A}^*\mathbf{A} = \mathbf{R}^*\mathbf{R}$.)

(a) Since $\mathbf{R}_1$ is diagonal, its singular values are just its diagonal values. So

$$\sigma_1(\mathbf{A}_1) = 6, \qquad \sigma_2(\mathbf{A}_1) = 4, \qquad \sigma_3(\mathbf{A}_1) = 2.$$

(b) We find

$$\kappa(\mathbf{A}_1) = \frac{\sigma_{\max}}{\sigma_{\min}} = \frac{6}{2} = 3, \qquad \kappa(\mathbf{A}_1^*\mathbf{A}_1) = (\kappa(\mathbf{A}_1))^2 = 9.$$

(c) Again it suffices to use $\mathbf{R}_2$. We obtain

$$\mathbf{R}_2^*\mathbf{R}_2 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix},$$

and set out to compute the eigenvalues of $\mathbf{R}_2^*\mathbf{R}_2$. We immediately see that $\lambda = 4$ is one eigenvalue. For the remaining two, we analyze the upper 2×2 block. Setting the characteristic polynomial to zero we get

$$0 = (\lambda - 1)(\lambda - 2) - 1 = \lambda^2 - 3\lambda + 1.$$

Therefore

$$\lambda_\pm = \frac{3 \pm \sqrt{9-4}}{2} = \frac{3 \pm \sqrt{5}}{2}.$$

The singular values of $\mathbf{R}_2$ (and hence $\mathbf{A}_2$) are the square roots of the eigenvalues of $\mathbf{R}_2^*\mathbf{R}_2$. So

$$\sigma_1(\mathbf{A}_2) = 2, \quad \sigma_2(\mathbf{A}_2) = \sqrt{\frac{3+\sqrt{5}}{2}}, \quad \sigma_3(\mathbf{A}_2) = \sqrt{\frac{3-\sqrt{5}}{2}}.$$

Question 3: (20p) Let $\mathbf{A} \in \mathbb{C}^{m\times m}$ be nonsingular, with SVD

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*, \qquad \mathbf{\Sigma} = \operatorname{diag}(\sigma_1, \ldots, \sigma_m), \quad \sigma_1 \geq \cdots \geq \sigma_m > 0.$$

Write $\mathbf{u}_j$ and $\mathbf{v}_j$ for the columns of $\mathbf{U}$ and $\mathbf{V}$. Consider the self adjoint block matrix

$$\mathbf{H} = \begin{bmatrix} \mathbf{0} & \mathbf{A}^* \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \in \mathbb{C}^{2m\times 2m}.$$

(a) (10p) For each $j = 1, \ldots, m$, find two eigenvectors of $\mathbf{H}$ in terms of $\mathbf{u}_j$ and $\mathbf{v}_j$, and verify their eigenvalue equations.

The singular-vector relations are $\mathbf{A}\mathbf{v}_j = \sigma_j \mathbf{u}_j$ and $\mathbf{A}^*\mathbf{u}_j = \sigma_j \mathbf{v}_j$. Hence

$$\mathbf{H}\begin{bmatrix} \mathbf{v}_j \\ \mathbf{u}_j \end{bmatrix} = \begin{bmatrix} \mathbf{A}^*\mathbf{u}_j \\ \mathbf{A}\mathbf{v}_j \end{bmatrix} = \sigma_j \begin{bmatrix} \mathbf{v}_j \\ \mathbf{u}_j \end{bmatrix},$$

$$\mathbf{H}\begin{bmatrix} \mathbf{v}_j \\ -\mathbf{u}_j \end{bmatrix} = \begin{bmatrix} -\mathbf{A}^*\mathbf{u}_j \\ \mathbf{A}\mathbf{v}_j \end{bmatrix} = -\sigma_j \begin{bmatrix} \mathbf{v}_j \\ -\mathbf{u}_j \end{bmatrix}.$$

Thus $[\mathbf{v}_j; \mathbf{u}_j]$ and $[\mathbf{v}_j; -\mathbf{u}_j]$ are eigenvectors with eigenvalues $+\sigma_j$ and $-\sigma_j$, respectively.

(b) (6p) List all $2m$ eigenvalues of $\mathbf{H}$, including multiplicities.

$$\boxed{\sigma_1, \ldots, \sigma_m, \ -\sigma_1, \ldots, -\sigma_m}.$$

These are all $2m$ eigenvalues, counting repetitions among the singular values. The eigenvectors from part (a) form an orthogonal basis of $\mathbb{C}^{2m}$.

(c) (4p) Express $\|\mathbf{H}\|$ and $\|\mathbf{H}\|_{\mathrm{F}}$ in terms of the singular values of $\mathbf{A}$.

Since $\mathbf{H}$ is self adjoint, its singular values are the absolute values of its eigenvalues. Thus

$$\boxed{\|\mathbf{H}\| = \sigma_1, \qquad \|\mathbf{H}\|_{\mathrm{F}} = \sqrt{2\sum_{j=1}^{m} \sigma_j^2} = \sqrt{2}\,\|\mathbf{A}\|_{\mathrm{F}}}.$$

Question 4: (20p) Set $t = 10^{-6}$. Consider an equation $\mathbf{Ax} = \mathbf{b}$ with a proposed approximate solution $\widehat{\mathbf{x}}$, where

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & t \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 1 \\ t \end{bmatrix}, \qquad \widehat{\mathbf{x}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

(a) (5p) Find the exact solution $\mathbf{x}_\star = \mathbf{A}^{-1}\mathbf{b}$.

The second row requires $tx_2 = t$, so $x_2 = 1$; the first gives $x_1 + x_2 = 1$. Therefore

$$\boxed{\mathbf{x}_\star = \begin{bmatrix} 0 \\ 1 \end{bmatrix}}.$$

(b) (5p) Compute the relative residual $\|\mathbf{b} - \mathbf{A}\widehat{\mathbf{x}}\|/\|\mathbf{b}\|$.

Since $\mathbf{A}\widehat{\mathbf{x}} = [1, 0]^T$, we have $\mathbf{b} - \mathbf{A}\widehat{\mathbf{x}} = [0, t]^T$, and thus

$$\boxed{\frac{\|\mathbf{b} - \mathbf{A}\widehat{\mathbf{x}}\|}{\|\mathbf{b}\|} = \frac{t}{\sqrt{1+t^2}} \approx 10^{-6}}.$$

(c) (5p) Compute the relative forward error $\|\mathbf{x}_\star - \widehat{\mathbf{x}}\|/\|\mathbf{x}_\star\|$.

$$\boxed{\frac{\|\mathbf{x}_\star - \widehat{\mathbf{x}}\|}{\|\mathbf{x}_\star\|} = \frac{\|[-1, 1]^T\|}{\|[0, 1]^T\|} = \sqrt{2}}.$$

(d) (5p) Now regard t as a variable small positive real number. Prove that as $t \searrow 0$, it is the case that $t\kappa(\mathbf{A}) \to C$ for some constant C. (So that $\kappa(A) \approx C/t$ for small t.) Specify C.

We first compute the singular values of $\mathbf{A}$ through the eigenvalues of $\mathbf{A}^*\mathbf{A}$. We find that

$$\mathbf{A}^*\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1+t^2 \end{bmatrix}.$$

Setting the characteristic polynomial to zero we get

$$0 = (\lambda - 1)(\lambda - 1 - t^2) - 1 = \lambda^2 - (2 + t^2)\lambda + t^2.$$

The roots are

$$\lambda_\pm = 1 + \frac{1}{2}t^2 \pm \sqrt{1 + t^2 + (1/4)t^4 - t^2} = 1 + \frac{1}{2}t^2 \pm \sqrt{1 + (1/4)t^4}.$$

We find that $\lambda_+ = 2 + O(t^2)$ and that

$$\lambda_- = 1 + \frac{1}{2}t^2 - (1 + O(t^4)) = \frac{1}{2}t^2 + O(t^4).$$

Consequently

$$\frac{\lambda_+}{\lambda_-} = \frac{2 + O(t^2)}{\frac{1}{2}t^2 + O(t^4)} = \frac{4}{t^2}(1 + O(t^2))$$

so $\kappa(\mathbf{A}) = \sqrt{\lambda_+/\lambda_-} = (2/t) + O(t)$.

Faster solution: Observe that $\mathbf{A}^*\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1+t^2 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ as $t \to 0$. It follows that $\lambda_{\max}(\mathbf{A}^*\mathbf{A}) = \sigma_{\max}(\mathbf{A})^2 \to 2$. Since $\det \mathbf{A} = t > 0$ and $\sigma_{\max}\sigma_{\min} = |\det \mathbf{A}| = t$, it follows that

$$t\kappa(\mathbf{A}) = t\frac{\sigma_{\max}}{\sigma_{\min}} = t\frac{\sigma_{\max}^2}{\sigma_{\min}\sigma_{\max}} = t\frac{\sigma_{\max}^2}{t} = \sigma_{\max}^2 \longrightarrow \boxed{C = 2}.$$

Question 5: (15p) Let $\mathbf{A} \in \mathbb{R}^{m\times m}$ be nonsingular, and let $\mathbf{B}$ be an approximate inverse that satisfies

$$\|\mathbf{B} - \mathbf{A}^{-1}\| \le 10^{-6}\|\mathbf{A}^{-1}\|.$$

Given a vector $\mathbf{b}$, we attempt to solve $\mathbf{Ax} = \mathbf{b}$ through an iteration that starts with $\mathbf{x}_1 = \mathbf{Bb}$, and then proceeds through

$$\mathbf{r}_k = \mathbf{b} - \mathbf{Ax}_k, \qquad \mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{Br}_k, \qquad k = 1, 2, 3, \ldots$$

For this question, ignore all rounding errors beyond the initial inaccuracy of $\mathbf{B}$. We observe the following relative residual norms:

k	1	2	3	4
$\|\mathbf{r}_k\|/\|\mathbf{b}\|$	10^{-3}	10^{-5}	10^{-7}	10^{-9}

(a) (4p) Show that $\mathbf{r}_{k+1} = \mathbf{M}\mathbf{r}_k$ for some matrix $\mathbf{M}$. Specify $\mathbf{M}$.

We find that

$$\mathbf{r}_{k+1} = \mathbf{b} - \mathbf{Ax}_{k+1} = \mathbf{b} - \mathbf{A}(\mathbf{x}_k + \mathbf{Br}_k) = \mathbf{r}_k - \mathbf{ABr}_k = (\mathbf{I} - \mathbf{AB})\mathbf{r}_k.$$

Therefore: $\boxed{\mathbf{M} = \mathbf{I} - \mathbf{AB}}$

(b) (7p) What can you tell about the condition number of $\mathbf{A}$, based on the observed norms $\|\mathbf{r}_k\|$? Explain your reasoning.

The assumption on $\mathbf{B}$ implies

$$\|\mathbf{M}\| = \|\mathbf{A}\,(\mathbf{A}^{-1} - \mathbf{B})\| \le \|\mathbf{A}\|\,\|\mathbf{A}^{-1} - \mathbf{B}\| \le \|\mathbf{A}\|\,10^{-6}\,\|\mathbf{A}^{-1}\| = 10^{-6}\kappa(\mathbf{A}).$$

But the residual data show $\|\mathbf{r}_{k+1}\|/\|\mathbf{r}_k\| = 10^{-2}$, so $\|\mathbf{M}\| \ge 10^{-2}$. Consequently

$$\kappa(\mathbf{A}) \ge 10^{6}\|\mathbf{M}\| \ge 10^{6}\,10^{-2} = 10^{4},$$

(Note: The information given is not sufficient to provide an upper bound on $\kappa(\mathbf{A})$.)

(c) (4p) If the observed factor-of-100 reduction continues, for what k would you expect $\|\mathbf{r}_k\|/\|\mathbf{b}\|$ to be less than 10^{-14}? At this point, what would your expectation be in regards to $\|\mathbf{x}_\star - \mathbf{x}_k\|/\|\mathbf{x}_\star\|$?

Continuing the observed pattern, we would expect

k	1	2	3	4	5	6	7
$\|\mathbf{r}_k\|/\|\mathbf{b}\|$	10^{-3}	10^{-5}	10^{-7}	10^{-9}	10^{-11}	10^{-13}	10^{-15}

So: $\boxed{k = 7}$

For the forward error, the standard estimate is

$$\frac{\|\mathbf{x}_\star - \mathbf{x}_k\|}{\|\mathbf{x}_\star\|} \le \kappa(\mathbf{A})\frac{\|\mathbf{r}_k\|}{\|\mathbf{b}\|}.$$

Since in this case, we do *not* have an upper bound of $\kappa(\mathbf{A})$, we cannot reliably bound $\frac{\|\mathbf{x}_\star - \mathbf{x}_k\|}{\|\mathbf{x}_\star\|}$ at all.

However, if you surmise from the observed data that $\kappa(\mathbf{A}) \approx 10^4$, then the forward error could be as large as

$$\frac{\|\mathbf{x}_\star - \mathbf{x}_k\|}{\|\mathbf{x}_\star\|} \approx \kappa(\mathbf{A})\frac{\|\mathbf{r}_k\|}{\|\mathbf{b}\|} \approx 10^4 \times 10^{-15} = 10^{-11}.$$

This is a heuristic only, but would produce partial credit.