

1. Ex. 2.1, part (b), p.42, *Shampine, Allen & Preuss*.
2. Ex. 2.5, part (b), p.48, *Shampine, Allen & Preuss*.
3. Consider the linear system

$$\begin{aligned} 6x_1 + 2x_2 + 2x_3 &= -4 \\ 2x_1 + \frac{2}{3}x_2 + \frac{1}{3}x_3 &= 2 \\ x_1 + 2x_2 - x_3 &= 0 \end{aligned}$$

- (a) Verify that its solution is

$$x_1 = 5.2 \quad x_2 = -7.6 \quad x_3 = -10.0$$

- (b) Using four digit floating-point decimal arithmetic with rounding, solve the system by Gaussian elimination without pivoting.
- (c) Repeat part (b) using partial pivoting.

In performing the arithmetic operations, remember to round to four significant digits after each operation, just as would be done on a computer.

What do you conclude ?

4. The Hilbert matrix of order  $n$ ,  $H_n$  is defined by

$$(H_n)_{i,j} = \frac{1}{i+j-1}, \quad i = 1, \dots, n, \quad j = 1, \dots, n.$$

$H_n$  is nonsingular. However, as  $n$  increases, the condition number of  $H_n$  increases rapidly.  $H_n$  is a library function in MATLAB, `hilb(n)`. Let  $n = 10$ ,  $\mathbf{x} = \text{ones}(10, 1)$  and  $\mathbf{b} = H_{10}\mathbf{x}$ . Now use the backslash operator to solve the system  $H_n\mathbf{x} = \mathbf{b}$ , obtaining  $\mathbf{x}^*$ . Since we know  $\mathbf{x}$  exactly, we can compute  $\mathbf{e} = \mathbf{x} - \mathbf{x}^*$ , the error, and  $\mathbf{r} = \mathbf{b} - H_{10}\mathbf{x}^*$ , the residual. Compute these quantities and also `cond(A)` (a MATLAB function). Show that the two basic principles of solving linear systems by G.E./P.P. in floating point arithmetic hold. How many correct digits does  $\mathbf{x}^*$  have ? Repeat with  $n = 11, 12, \dots$ . Stop when some component of  $\mathbf{x}^*$  has no correct digits.

5. Define the matrix  $A_n$  of order  $n$  by

$$A_n = \begin{bmatrix} 1 & -1 & -1 & -1 & \cdots & -1 \\ 0 & 1 & -1 & -1 & \cdots & -1 \\ & & & \ddots & & \\ \vdots & & & & & 1 & -1 \\ 0 & \cdots & & & 0 & 1 \end{bmatrix}$$

- (a) Find the inverse of  $A_n$  explicitly.

Hint: Find the inverse of  $A_6$  by using MATLAB. Then use the result to “guess” the inverse of  $A_n$  in general.

- (b) Calculate  $\text{cond}(A_n)$  in the  $\infty$ -norm.  
 (c) With  $\mathbf{b} = [-n + 2, -n + 3, \dots, -1, 0, 1]^T$ , the solution of  $A_n \mathbf{x} = \mathbf{b}$  is  $\mathbf{x} = [1, 1, \dots, 1]^T$ . Perturb  $\mathbf{b}$  to  $\hat{\mathbf{b}} = \mathbf{b} + [0, \dots, 0, \epsilon]^T$ . Solve for  $\hat{\mathbf{x}}$  in  $A_n \hat{\mathbf{x}} = \hat{\mathbf{b}}$ . Show that these values of  $\mathbf{b}, \hat{\mathbf{b}}, \mathbf{x}, \hat{\mathbf{x}}$  satisfy the fundamental inequality,

$$\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\|\mathbf{x}\|} \leq \text{cond}(A_n) \frac{\|\mathbf{b} - \hat{\mathbf{b}}\|}{\|\mathbf{b}\|}.$$

(Use the  $\infty$ -norm.)

Hint  $\hat{\mathbf{x}} = \mathbf{x} + A_n^{-1}(\hat{\mathbf{b}} - \mathbf{b})$ .

6. Suppose  $\mathbf{x}$  satisfies  $A\mathbf{x} = \mathbf{b}$  and  $\mathbf{x} + \Delta\mathbf{x}$  satisfies  $(A + \Delta A)(\mathbf{x} + \Delta\mathbf{x}) = \mathbf{b} + \Delta\mathbf{b}$ . Then we have the *condition number inequality*: If  $\rho = \|A^{-1}\| \cdot \|\Delta A\| < 1$

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\text{cond}(A)}{1 - \rho} \left( \frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} \right). \quad (1)$$

Consider the linear system  $A\mathbf{x} = \mathbf{b}$  where

$$A = \begin{pmatrix} .5055 & .6412 & .8035 \\ .1693 & .0162 & .6978 \\ .5280 & .8369 & .4617 \end{pmatrix}, \quad \mathbf{b} = \begin{bmatrix} .4939 \\ .4175 \\ .2923 \end{bmatrix}$$

- (a) Solve  $A\mathbf{x} = \mathbf{b}$  using the backslash operator.  
 (b) Use equation (1) to answer the following question: If each entry in  $A$  and  $\mathbf{b}$  might have an error of  $\pm .00005$ , how reliable is  $\mathbf{x}$ ? Use the  $\infty$ -norm.  
 (c) Let

$$\Delta A = .0001 * \text{rand}(3) - .00005 * \text{ones}(3), \quad \Delta \mathbf{b} = .0001 * \text{rand}(3, 1) - .00005 * \text{ones}(3, 1).$$

Solve  $(A + \Delta A)(\mathbf{x} + \Delta\mathbf{x}) = \mathbf{b} + \Delta\mathbf{b}$  to get  $\mathbf{x} + \Delta\mathbf{x}$ . Calculate  $\|\Delta\mathbf{x}\|/\|\mathbf{x}\|$ . Is this consistent with (b)? What is the relative change in each  $x_i$ ?

7.

- (a) Let  $A$  be an  $n \times n$  matrix and  $\mathbf{x} \in \mathbf{R}^n$ . How many *flops* does it take to form the product  $A\mathbf{x}$ ?  
 (b) Let  $A$  and  $B$  be  $n \times n$  matrices. How many *flops* does it take to form the product  $AB$ ?  
 (c) In light of the results of (a) and (b), from the standpoint of efficiency, how should one compute  $A^k \mathbf{x}$  for  $k$  a positive integer  $k > 1$ ?

8. Construct a tridiagonal solver along the lines outlined on pp.68-69 of *Shampine, Allen & Preuss*. Your solver should be a function M-file with four vectors as input and the solution as the output. Test your solver on the problem  $A\mathbf{x} = \mathbf{b}$  where  $A$  is the  $9 \times 9$  tridiagonal matrix with 2's on the main diagonal and  $-1$ 's on the super- and subdiagonals and  $\mathbf{b}(j) = .01, j = 1, \dots, 9$ . As a check, the answer should be  $\mathbf{x}(j) = \frac{1}{2} \left( \frac{j}{10} \right) \left( 1 - \frac{j}{10} \right)$ .