CHAPTER 20 Numeric Linear Algebra

SECTION 20.1. Linear Systems: Gauss Elimination, page 841

Purpose. To explain the Gauss elimination, which is a solution method for linear systems of equations by systematic elimination (reduction to triangular form).

Main Content, Important Concepts

Gauss elimination, back substitution

Pivot equation, pivot, choice of pivot

Operations count, order [e.g., $O(n^3)$]

Comments on Content

This section is independent of Chap. 7 on matrices (in particular, independent of Sec. 7.3, where the Gauss elimination is also considered).

Gauss's method and its variants (Sec. 20.2) are the most important solution methods for those systems (with matrices that do not have too many zeros).

The Gauss–Jordan method (Sec. 20.2) is less practical because it requires more operations than the Gauss elimination.

Cramer's rule (Sec. 7.7) would be totally impractical in numeric work, even for systems of modest size.

SOLUTIONS TO PROBLEM SET 20.1, page 848

2.
$$x_1 = t_1$$
 arbitrary, $x_2 = (25/42)t_1$

4.
$$x_1 = 0$$
, $x_2 = -3$

6. $x_1 = (30.60 + 15.48x_2)/25.38$, x_2 arbitrary. Remember from Sec. 7.1 that one also writes $x_2 = t_1$ = first arbitrary unknown and thus $x_1 = (30.60 + 15.48t_1)/25.38$, $x_2 = t_1$ (arbitrary).

Hence the system has no solution.

10. Gauss reduction to triangular form gives

$$\begin{bmatrix} 4 & 4 & 2 & 0 \\ 0 & -4 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Back substitution gives

$$x_1 = -\frac{5}{8}t_1$$
, $x_2 = \frac{1}{8}t_1$, $x_3 = t_1$ arbitrary.

Note that rank A = 2.

12. Gauss reduction to triangular form gives

$$\begin{bmatrix} 5 & 3 & 1 & 2 \\ 0 & -4 & 8 & -3 \\ 0 & 0 & 0 & 5 \end{bmatrix}.$$

This shows that the system has no solution.

14. Gauss reduction to triangular form gives

$$\begin{bmatrix} -47 & 4 & -7 & -118 \\ 0 & -\frac{65}{47} & -\frac{39}{47} & -\frac{221}{47} \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Back substitution gives the solution

$$x_1 = t_1$$
 arbitrary, $x_2 = 3t_1 - 5$, $x_3 = -5t_1 + 14$.

16. Gauss reduction to triangular form gives

$$\begin{bmatrix} 3.2 & 1.6 & 0 & 0 & -0.8 \\ 0 & 2.4 & -4.8 & 3.6 & -39 \\ 0 & 0 & 3.6 & 2.4 & 10.2 \\ 0 & 0 & 0 & 2.93333 & -7.33333 \end{bmatrix}.$$

Back substitution now gives

$$x_1 = 1.5$$
, $x_2 = -3.5$, $x_3 = 4.5$, $x = -2.5$.

- **18. Team Project.** (a) (i) $a \ne 1$ to make $D = a 1 \ne 0$; (ii) a = 1, b = 3; (iii) $a = 1, b \ne 3$.
 - **(b)** $x_1 = \frac{1}{2}(3x_3 1), x_2 = \frac{1}{2}(-5x_3 + 7), x_3$ arbitrary is the solution of the system. The second system has no solution.
 - (c) det A = 0 can change to det $A \neq 0$ because of roundoff.
 - (d) $(1-1/\epsilon)x_2=2-1/\epsilon$ eventually becomes $x_2/\epsilon\approx 1/\epsilon$, $x_2=1$, $x_1=(1-x_2)/\epsilon\approx 0$. The exact solution is $x_1=1/(1-\epsilon)$, $x_2=(1-2\epsilon)/(1-\epsilon)$. We obtain it if we take $x_1+x_2=2$ as the pivot equation.
 - (e) The exact solution is $x_1 = 1$, $x_2 = -4$. The 3-digit calculation gives $x_2 = -4.5$, $x_1 = 1.27$ without pivoting and $x_2 = -6$, $x_1 = 2.08$ with pivoting. This shows that 3S is simply not enough. The 4-digit calculation give $x_2 = -4.095$, $x_1 = 1.051$ without pivoting and the exact result $x_2 = -4$, $x_1 = 1$ with pivoting.

SECTION 20.2. Linear Systems: LU-Factorization, Matrix Inversion, page 849

Purpose. To discuss Doolittle's, Crout's, and Cholesky's methods, three methods for solving linear systems that are based on the idea of writing the coefficient matrix as a product of two triangular matrices ("LU-factorization"). Furthermore, we discuss matrix inversion by the Gauss–Jordan elimination.

Main Content, Important Concepts

Doolittle's and Crout's methods for arbitrary square matrices

Cholesky's method for positive definite symmetric matrices

Numerical matrix inversion

Short Courses. Doolittle's method and the Gauss–Jordan elimination.

Comment on Content

L suggests "lower triangular" and U "upper triangular." For Doolittle's method, these are the same as the matrix of the multipliers and of the triangular system in the Gauss elimination.

The point is that in the present methods, one solves one equation at a time, no systems.

SOLUTIONS TO PROBLEM SET 20.2, page 854

2.
$$\begin{bmatrix} 1 & 0 \\ 1.5 & 1 \end{bmatrix} \begin{bmatrix} 2 & 9 \\ 0 & -18.5 \end{bmatrix}, \quad x_1 = -4$$
4.
$$\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ \frac{1}{2} & \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 2 \\ 0 & 3 & 3 \\ 0 & 0 & -\frac{9}{2} \end{bmatrix}, \quad x_2 = 4$$
6. **Team Project.** (a) The formula for the entries of $\mathbf{L} = [l_{jk}]$ and $\mathbf{U} = [u_{jk}]$ are

$$l_{j1} = a_{j1} j = 1, \dots, n$$

$$u_{1k} = \frac{a_{1k}}{l_{11}} k = 2, \dots, n$$

$$l_{jk} = a_{jk} - \sum_{s=1}^{k-1} l_{js} u_{sk} j = k, \dots, n; \quad k \ge 2$$

$$u_{jk} = \frac{1}{l_{jj}} \left(a_{jk} - \sum_{s=1}^{j-1} l_{js} u_{sk} \right) k = j+1, \dots, n; \quad j \ge 2.$$

(b) The factorization and the solutions are

$$\begin{bmatrix} 3 & 0 & 0 \\ 18 & -6 & 0 \\ 9 & -54 & -3 \end{bmatrix} \begin{bmatrix} 1 & 3 & 2 \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} \frac{23}{15} \\ \frac{1}{15} \\ \frac{2}{5} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} -\frac{1}{15} \\ \frac{4}{15} \\ \frac{2}{5} \end{bmatrix}.$$

(c) The three factorizations are

$$\begin{bmatrix} 1 & 0 & 0 \\ -4 & 1 & 0 \\ 2 & \frac{4}{3} & 1 \end{bmatrix} \begin{bmatrix} 1 & -4 & 2 \\ 0 & 9 & 12 \\ 0 & 0 & 4 \end{bmatrix}$$
(Doolittle)

$$\begin{bmatrix} 1 & 0 & 0 \\ -4 & 9 & 0 \\ 2 & 12 & 4 \end{bmatrix} \begin{bmatrix} 1 & -4 & 2 \\ 0 & 1 & \frac{4}{3} \\ 0 & 0 & 1 \end{bmatrix}$$
 (Crout)
$$\begin{bmatrix} 1 & 0 & 0 \\ -4 & 3 & 0 \\ 2 & 4 & 2 \end{bmatrix} \begin{bmatrix} 1 & -4 & 2 \\ 0 & 3 & 4 \\ 0 & 0 & 2 \end{bmatrix}$$
 (Cholesky).

(d) For fixing the notation, for n = 4 we have

For fixing the notation, for
$$n = 4$$
 we have
$$\mathbf{A} = \mathbf{L}\mathbf{U} = \begin{bmatrix} a_1 & c_1 & 0 & 0 \\ b_2 & a_2 & c_2 & 0 \\ 0 & b_3 & a_3 & c_3 \\ 0 & 0 & b_4 & a_4 \end{bmatrix} = \begin{bmatrix} \alpha_1 & 0 & 0 & 0 \\ b_2 & \alpha_2 & 0 & 0 \\ 0 & b_3 & \alpha_3 & 0 \\ 0 & 0 & b_4 & \alpha_4 \end{bmatrix} \begin{bmatrix} 1 & \gamma_1 & 0 & 0 \\ 0 & 1 & \gamma_2 & 0 \\ 0 & 0 & 1 & \gamma_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
here

where

$$\alpha_1 = a_1,$$
 $\alpha_j = a_j - b_j \gamma_{j-1},$ $j = 2, \dots, n$
 $\gamma_1 = c_1/\alpha_1,$ $\gamma_j = c_j/\alpha_j,$ $j = 2, \dots, n-1$

(e) If A is symmetric.

8.
$$\begin{bmatrix} 2 & 0 & 0 \\ 3 & 5 & 0 \\ 4 & 8 & 7 \end{bmatrix} \begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 8 \\ 0 & 0 & 7 \end{bmatrix}, x_1 = 8$$
10.
$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & \frac{1}{2} & \frac{1}{2}\sqrt{3} \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2}\sqrt{3} \end{bmatrix}, x_2 = \frac{3}{4}$$

$$x_3 = 1$$
12.
$$\begin{bmatrix} 2 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 2 & 1 & 1 & 0 \\ 0 & 2 & 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 & 2 & 0 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix} = 6$$
16.
$$\frac{1}{9} \begin{bmatrix} 2 & -2 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & -2 \end{bmatrix} = \frac{1}{9} \mathbf{A}^{\mathsf{T}}. \text{ Hence } \frac{1}{3} \mathbf{A} \text{ is orthogonal.}$$

18.
$$150$$
 150 -300 -300 -50 100

Note that det $\mathbf{A} = 0.000016$, det $\mathbf{A}^{-1} = 1/\det \mathbf{A} = 62,500$. Furthermore, the condition number of \mathbf{A} (Sec 20.4) will be 0.25 \cdot 1450 = 362.5, showing that \mathbf{A} is ill-conditioned.

20. det A = 0 as given, but rounding makes det $A \neq 0$ and may completely change the situation with respect to existence of solutions of linear systems, a point to be watched for when using a CAS. In the present case we get (a) -0.00000035, (b) -0.00001998, (c) -0.00028189, (d) 0.002012, (e) 0.0038.

SECTION 20.3 Linear Systems: Solution by Iteration, page 855

Purpose. To familiarize the student with the idea of solving linear systems by iteration, to explain in what situations that is practical, and to discuss the most important method (Gauss–Seidel iteration) and its convergence.

Main Content, Important Concepts

Distinction between direct and indirect methods

Gauss-Seidel iteration, its convergence, its range of applicability

Matrix norms

Jacobi iteration

Short Courses. Gauss–Seidel iteration only.

Comments on Content

The Jacobi iteration appeals by its simplicity but is of limited practical value.

A word on the frequently occurring sparse matrices may be good. For instance, we have about 99.5% zeros in solving the Laplace equation in two dimensions by using a 1000×1000 grid and the usual five-point pattern (Sec. 21.4).

SOLUTIONS TO PROBLEM SET 20.3, page 860

2. The eigenvalues of I - A are 0.5, 0.5, -1. Here A is $\frac{1}{2}$ times the coefficient matrix of the given system; thus,

$$\mathbf{I} - \mathbf{A} = \begin{bmatrix} 0 & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & 0 \end{bmatrix}.$$

4. The exact solution is $x = \begin{bmatrix} 3 & -9 & 6 \end{bmatrix}^T$ is reached at Step 8 rather quickly owing to the fact that the spectral radius of **C** is 0.125, hence rather small; here

$$\mathbf{C} = \begin{bmatrix} 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{16} & \frac{1}{4} \\ 0 & \frac{1}{64} & \frac{1}{16} \end{bmatrix}.$$

322

6. Row 1 becomes Row 3, Row 2 becomes Row 1, and Row 3 becomes Row 2. The exact solution is $x_1 = 0.5$, $x_2 = -2.5$, $x_3 = 4.0$. The choice of $\mathbf{x_0}$ is of much less influence than one would expect. We obtain for $\mathbf{x_0} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^\mathsf{T}$

For
$$\mathbf{x_0} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^\mathsf{T}$$

$$1, \begin{bmatrix} -0.200000, -1.88333, 3.91190 \end{bmatrix}$$

$$2, \begin{bmatrix} 0.376666, -2.46477, 3.99497 \end{bmatrix}$$

$$3, \begin{bmatrix} 0.492954, -2.49798, 3.99971 \end{bmatrix}$$

$$4, \begin{bmatrix} 0.499596, -2.49988, 3.99998 \end{bmatrix}$$

$$5, \begin{bmatrix} 0.499976, -2.50000, 4.00000 \end{bmatrix}$$

for
$$\mathbf{x_0} = [10 \quad 10 \quad 10]^T$$

$$1, [-2, -3.08333, 4.08333]$$

$$2, [0.616666, -2.53333, 4.00476]$$

$$3, [0.506666, -2.50190, 4.00027]$$

$$4, [0.500380, -2.50010, 4.00001]$$

$$5, [0.500020, -2.50000, 4.00000]$$

$$6, [0.500000, -2.50000, 4.00000]$$

and for
$$\mathbf{x_0} = [100 \quad 100 \quad 100]^T$$

$$1, [-20, -15.0833, 5.79761]$$

$$2, [3.01666, -3.21905, 4.10271]$$

$$3, [0.643810, -2.54108, 4.00587]$$

$$4, [0.508216, -2.50235, 4.00034]$$

$$5, [0.500470, -2.50013, 4.00001]$$

$$6, [0.500026, -2.50000, 4.00000]$$

The spectral radius of C (and its only nonzero eigenvalue) is $\frac{2}{35} = 0.057$.

- **8.** The exact solution is 2, 0, 1. Step 10 gives $[2.00144 -0.00221311 \ 0.999779]^T$. The spectral radius $(\frac{2}{3})^{3/2} = 0.544331$ of **C** is relatively large.
- 10. Interchange the first equation and the last equation. Then the exact solution -2.5, 2, 4.5 is reached at Step 11, the spectral radius of

$$\mathbf{C} = \begin{bmatrix} 0 & -\frac{1}{4} & -\frac{1}{8} \\ 0 & \frac{1}{24} & -\frac{5}{16} \\ 0 & \frac{1}{5} & \frac{1}{10} \end{bmatrix}$$

being $1/\sqrt{15} = 0.258199$. (The eigenvalues are complex conjugates and the third eigenvalue is 0, as always for the present \mathbb{C} .)

12. In (a) we obtain

$$\mathbf{C} = -(\mathbf{I} + \mathbf{L})^{-1}\mathbf{U}$$

$$= -\begin{bmatrix} 1 & 0 & 0 \\ -0.1 & 1 & 0 \\ -0.09 & -0.1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0.1 & 0.1 \\ 0 & 0 & 0.1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & -0.100 & -0.100 \\ 0 & 0.010 & -0.090 \\ 0 & 0.009 & 0.019 \end{bmatrix}$$

and $\|\mathbf{C}\| = 0.2 < 1$ by (11), which implies convergence by (8). In (b) we have

$$\begin{bmatrix} 1 & 1 & 10 \\ 10 & 1 & 1 \\ 1 & 10 & 1 \end{bmatrix} = (\mathbf{I} + \mathbf{L}) + \mathbf{U}$$
$$= \begin{bmatrix} 1 & 0 & 0 \\ 10 & 1 & 0 \\ 1 & 10 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 10 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

From this we compute

$$\mathbf{C} = -(\mathbf{I} + \mathbf{L})^{-1}\mathbf{U} = -\begin{bmatrix} 1 & 0 & 0 \\ -10 & 1 & 0 \\ 99 & -10 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 10 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$
$$= -\begin{bmatrix} 0 & 1 & 10 \\ 0 & -10 & -99 \\ 0 & 99 & 980 \end{bmatrix}$$
$$= \begin{bmatrix} 0 & -1 & -10 \\ 0 & 10 & 99 \\ 0 & -99 & -980 \end{bmatrix}.$$

Developing the characteristic determinant of C by its first column, we obtain

$$\begin{vmatrix} -\lambda & -\lambda & 99 \\ -99 & -\lambda - 980 \end{vmatrix} = -\lambda(\lambda^2 + 970\lambda + 1)$$

which shows that one of the eigenvalues is greater than 1 in absolute value, so that we have divergence. In fact, $\lambda = -0, 0.001$, and -970, approximately.

Step 5 of the Gauss-Seidel iteration gives the better result

$$[2.99969 -9.00015 5.99996]^{\mathsf{T}}$$
. Exact: $[3 -9 6]^{\mathsf{T}}$.

16.
$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$
, $\begin{bmatrix} -1.8125 \\ 2.58333 \\ 1.7 \end{bmatrix}$, $\begin{bmatrix} -2.29583 \\ 2.81875 \\ 3.95 \end{bmatrix}$, $\begin{bmatrix} -2.63594 \\ 2.14931 \\ 4.33667 \end{bmatrix}$, $\begin{bmatrix} -2.51691 \\ 2.07710 \\ 4.60875 \end{bmatrix}$, $\begin{bmatrix} 1.96657 \\ 4.51353 \end{bmatrix}$

Step 5 of the Gauss-Seidel iteration gives the more accurate result

$$[-2.49475 \quad 1.99981 \quad 4.49580]^{\mathsf{T}}$$
. Exact: $[-2.5 \quad 2 \quad 4.5]^{\mathsf{T}}$.

18. $\sqrt{151} = 12.29$, 13 (column sum norm), 11 (row sum norm)

20.
$$\sqrt{18k^2} = 4.24|k|, 4|k|, 4|k|$$

SECTION 20.4. Linear Systems: III-Conditioning, Norms, page 861

Purpose. To discuss ill-conditioning quantitatively in terms of norms, leading to the condition number and its role in judging the effect of inaccuracies on solutions.

Main Content, Important Concepts

Ill-conditioning, well-conditioning

Symptoms of ill-conditioning

Residual

Vector norms

Matrix norms

Condition number

Bounds for effect of inaccuracies of coefficients on solutions

Comment on Content

Reference [E9] in App. 1 gives some help when A^{-1} , needed in $\kappa(A)$, is unknown (as is the case in practice).

SOLUTIONS TO PROBLEM SET 20.4, page 868

2. 13, 9, 8, $\left[\frac{1}{2} - \frac{1}{8}\right]$ 1]. Note that the l_2 -norm of a vector with integer components will generally not be an integer. **4.** $4k + k^2 + k^3$, $\sqrt{k^6 + k^4 + 16k^2}$, k^3 , $[1/k \ 4/k^2 \ 1]$

4.
$$4k + k^2 + k^3$$
. $\sqrt{k^6 + k^4 + 16k^2}$. k^3 . $[1/k, 4/k^2, 1]$

6. 1, 1. The given vector is a unit vector.

8. Square the double inequality to be proved, that is,

$$\|\mathbf{x}\|_{\infty}^2 \le \|\mathbf{x}\|_2^2 \le \|\mathbf{x}\|_1^2$$

326

and write this inequality in terms of components, obtaining

$$(\max_{j} |x_{j}|)^{2} \le \sum_{j=1}^{n} x_{j}^{2} \le \left(\sum_{j=1}^{n} |x_{j}|\right)^{2}.$$

Now perform the square on the right.

- **10.** $\|\mathbf{A}\|_1 = 6.3$, $\|\mathbf{A}^{-1}\|_1 = \frac{220}{51} = 4.3137$, so that $\kappa = 27.176$.
- **12.** $\kappa = \|\mathbf{A}\|_1 \cdot \|\mathbf{A}^{-1}\|_1 = 13 \cdot 13 = 169$. The matrix is ill-conditioned.
- **14.** $\kappa = \frac{51}{50} \cdot \frac{5100}{4999} = 1.0406$. For the unit matrix we would have $\kappa = 1$, and the problem shows the effect of small off-diagonal entries, as they may occur in numeric diagonalization.
- **16.** A is the 4×4 Hilbert matrix times 21. Its inverse is

$$\mathbf{A}^{-1} = \frac{1}{21} \begin{bmatrix} 16 & -120 & 240 & -140 \\ -120 & 1200 & -2700 & 1680 \\ 240 & -2700 & 6480 & -4200 \\ -140 & 1680 & -4200 & 2800 \end{bmatrix}.$$

This gives the condition number (in both norms, since A and A^{-1} are symmetric)

$$\kappa = 43.75 \cdot \frac{13,620}{21} = 28,375.000$$

showing that **A** is very ill-conditioned.

18. The product of the two matrices equals

$$\mathbf{AB} = \begin{bmatrix} 4.7 & 10.8 \\ 2.0 & 7.2 \end{bmatrix}.$$

Hence for the l_1 -norm (column "sum" norm) we have (12) in the form

$$18.0 < 5.0 \cdot 6.3 = 31.5$$

and for the l_{∞} -norm (row "sum" norm)

$$15.5 < 4.0 \cdot 6.6 = 26.4$$
.

20. The systems have the solutions

$$\begin{bmatrix} 1 & 1 \end{bmatrix}^\mathsf{T}$$

and (4S-values)

$$[0.8455 \quad 1.273]^{\mathsf{T}}.$$

Hence a change of 0.2% in **b** has produced changes of 16 and 27% in the components of the solution. The matrix is ill-conditioned; the condition number is (4S-value)

$$\kappa(\mathbf{A}) = 4.7 \cdot 42.73 = 200.8.$$

22. By (12), $1 = \|\mathbf{I}\| = \|\mathbf{A}\mathbf{A}^{-1}\| \le \|\mathbf{A}\| \cdot \|\mathbf{A}^{-1}\| = \kappa(\mathbf{A})$. For the Frobenius norm, $\sqrt{n} = \|\mathbf{I}\| \le \kappa(\mathbf{A})$.

24. Team Project. (a) Formula (18a) is obtained from

$$\max |x_j| \le \Sigma |x_k| = \|\mathbf{x}\|_1 \le n \max |x_j| = n \|\mathbf{x}\|_{\infty}.$$

Equation (18b) follows from (18a) by division by n.

(b) To get the first inequality in (19a), consider the square of both sides and then take square roots on both sides. The second inequality in (19a) follows by means of the Cauchy–Schwarz inequality and a little trick worth remembering,

$$\sum |x_j| = \sum 1 \cdot |x_j| \le \sqrt{\sum 1^2} \sqrt{\sum |x_j|^2} = \sqrt{n} \|\mathbf{x}\|_2.$$

To get (19b), divide (19a) by \sqrt{n} .

(c) Let $x \neq 0$. Set x = ||x||y. Then ||y|| = ||x||/||x|| = 1. Also,

$$\mathbf{A}\mathbf{x} = \mathbf{A}(\|\mathbf{x}\|\mathbf{y}) = \|\mathbf{x}\|\,\mathbf{A}\mathbf{y}$$

since $\|\mathbf{x}\|$ is a number. Hence $\|\mathbf{A}\mathbf{x}\|/\|\mathbf{x}\| = \|\mathbf{A}\mathbf{y}\|$, and in (9), instead of taking the maximum over all $\mathbf{x} \neq \mathbf{0}$, since $\|\mathbf{y}\| = 1$ we only take the maximum over all \mathbf{y} of norm 1. Write \mathbf{x} for \mathbf{y} to get (10) from this.

(d) These "axioms of a norm" follow from (3), which are the axioms of a vector norm.

SECTION 20.5. Least Squares Method, page 869

Purpose. To explain Gauss's least squares method of "best fit" of straight lines to given data $(x_0, y_0), \dots, (x_n, y_n)$ and its extension to best fit of quadratic polynomials, etc.

Main Content, Important Concepts

Least squares method

Normal equations (4) for straight lines

Normal equations (8) for quadratic polynomials

Short Courses. Discuss the linear case only.

Comment. Normal equations are often *ill-conditioned*, so that results may be sensitive to roundoff. For another (theoretically much more complicated) method, see Ref. [E5], p. 201.

SOLUTIONS TO PROBLEM SET 20.5, page 872

- 2. 2.676 1.216x. Note the change of the slope.
- **4.** Hook's law F = ks gives the spring modulus k = F/s; for the present data,

$$s(F) = 0.03349 + 0.3139F$$
, hence $k = 1/0.3139 = 3.186$.

- **6.** U = -5.20 + 53.4i. Estimate: $R = 53.4\Omega$. Note that the line does not pass through the origin, as it should. This is typical.
- 8. $2.955 1.159x + 0.932x^2$
- **10.** $v = 2.29 0.433t + 0.105t^2$

12.
$$b_0 n + b_1 \sum x_j + b_2 \sum x_j^2 + b_3 \sum x_j^3 = \sum y_j$$

 $b_0 \sum x_j + b_1 \sum x_j^2 + b_2 \sum x_j^3 + b_3 \sum x_j^4 = \sum x_j y_j$
 $b_0 \sum x_j^2 + b_1 \sum x_j^3 + b_2 \sum x_j^4 + b_3 \sum x_j^5 = \sum x_j^2 y_j$
 $b_0 \sum x_j^3 + b_1 \sum x_j^4 + b_2 \sum x_j^5 + b_3 \sum x_j^6 = \sum x_j^3 y_j$

14. Team Project. (a) We substitute $F_m(x)$ into the integral and perform the square. This gives

$$||f - F_m||^2 = \int_a^b f^2 dx - 2 \sum_{j=0}^m a_j \int_a^b fy_j dx + \sum_{j=0}^m \sum_{k=0}^m a_j a_k \int_a^b y_j y_k dx.$$

This is a quadratic function in the coefficients. We take the partial derivative with respect to any one of them, call it a_l , and equate this derivative to zero. This gives

$$0 - 2 \int_{a}^{b} f y_{l} dx + 2 \sum_{i=0}^{m} a_{j} \int_{a}^{b} y_{j} y_{l} dx = 0.$$

Dividing by 2 and taking the first integral to the right gives the system of normal equations, with $l = 0, \dots, m$.

(b) In the case of a polynomial we have

$$\int_{a}^{b} y_{j} y_{l} dx = \int_{a}^{b} x^{j+l} dx$$

which can be readily integrated. In particular, if a = 0 and b = 1, integration from 0 to 1 gives 1/(j + l + 1), and we obtain the **Hilbert matrix** as the coefficient matrix.

SECTION 20.6. Matrix Eigenvalue Problems: Introduction, page 873

Purpose. This section is a collection of concepts and a handful of theorems on matrix eigenvalues and eigenvectors that are frequently needed in numerics; some of them will be discussed in the remaining sections of the chapter and others can be found in more advanced or more specialized books listed in Part E of App. 1.

The section frees both the instructor and the student from the task of locating these matters in Chaps. 7 and 8, which contain much more material and should be consulted only if problems on one or another matter are wanted (depending on the background of the student) or if a proof might be of interest.

SECTION 20.7. Inclusion of Matrix Eigenvalues, page 876

Purpose. To discuss theorems that give approximate values and error bounds of eigenvalues of general (square) matrices (Theorems 1, 2, 4, Example 2) and of special matrices (Theorem 6).

Main Content, Important Concepts

Gerschgorin's theorem (Theorem 1)

Sharpened Gerschgorin's theorem (Theorem 2)

Gerschgorin's theorem improved by similarity (Example 2)

Strict diagonal dominance (Theorem 3)

Schur's inequality (Theorem 4), normal matrices

Perron's theorem (Theorem 5)

Collatz's theorem (Theorem 6)

Short Courses. Discuss Theorems 1 and 6.

Comments on Content

It is important to emphasize that one must always make sure whether or not a thoerem applies to a given matrix. Some theorems apply to any real or complex square matrices whatsoever, whereas others are restricted to certain classes of matrices.

The exciting Gerschgorin theorem was one of the early theorems on numerics for eigenvalues; it appeared in the *Bull. Acad. Sciences de l'URSS* (Classe mathém, 7-e série, Leningrad, 1931, p. 749), and shortly thereafter in the German *Zeitschrift für angewandte Mathematik und Mechanik*.

SOLUTIONS TO PROBLEM SET 20.7, page 881

- 2. 5, 8, 9; radii $2 \cdot 10^{-2}$. Estimates of this kind can be useful when a matrix has been diagonalized numerically and some very small nonzero off-diagonal entries are left.
- **4.** 1, 4, 12; radii 1, 3, 4. Since the matrix is symmetric, this gives the two overlapping intervals $0 \le \lambda \le 2$ and $1 \le \lambda \le 7$ and the separate interval $0 \le \lambda \le 16$. The actual spectrum is (4S-values)

$$\{0.8786, 3.047, 13.07\},\$$

again illustrating Theorem 2; cf. Example 1.

- **6.** 10, 6, 3; radii 0.3, 0.1, 0.2; actually (4S) 10.01, 6.000, 2.994
- **8.** T with $t_{11} = t_{22} = 1$, $t_{33} = 34$ gives

$$\mathbf{T}^{-1}\mathbf{A}\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{34} \end{bmatrix} \begin{bmatrix} 10 & 0.1 & -0.2 \\ 0.1 & 6 & 0 \\ -0.2 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 34 \end{bmatrix}$$
$$= \begin{bmatrix} 10 & 0.1 & -6.8 \\ 0.1 & 6 & 0 \\ -\frac{0.2}{34} & 0 & 3 \end{bmatrix}$$

Note that the disk with center 3 is still disjoint from that with center 10. Indeed,

$$10 - 6.9 = 3.1$$
 whereas $3 + \frac{0.2}{34} = 3.00588$.

The next integer, 35, is already too large and leads to overlapping,

$$3 + \frac{0.2}{35} = 3.057 > 10 - 7.1 = 2.9.$$

10. An example is

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

The eigenvalues are -1 and 1, so that the entire spectrum lies on the circle. A similar looking 3×3 matrix or 4×4 matrix, etc., can be constructed with some or all of its eigenvalues on the circle.

- **12.** $\sqrt{181} = 13.45$
- **14.** $\sqrt{145.1} = 12.05$
- **16.** $\sqrt{83} = 9.110$

- **18.** Proofs follow readily from the definitions of these classes of matrices. This *normality* is of interest, since normal matrices have important properties; in particular, they have an orthonormal set of *n* eigenvectors, where *n* is the size of the matrix. See [B3], vol. 1, pp. 268–274.
- **20.** This is a *continuity proof.* Let $S = D_1 \cup D_2 \cup \cdots \cup \mathbf{p}$ thout restriction, where D_j is the Gerschgorin disk with center a_{jj} . We write $\mathbf{A} = \mathbf{B} + \mathbf{C}$, where $\mathbf{B} = \mathrm{diag}(a_{jj})$ is the diagonal matrix with the main diagonal of \mathbf{A} as its diagonal. We now consider

$$\mathbf{A}_t = \mathbf{B} + t\mathbf{C} \qquad \text{for } 0 \le t \le 1.$$

Then $A_0 = B$ and $A_1 = A$. Now by algebra, the roots of the characteristic polynomial $f_t(\lambda)$ of A_t (that is, the eigenvalues of A_t) depend continuously on the coefficients of $f_t(\lambda)$, which in turn depend continuously on t. For t = 0 the eigenvalues are a_{11}, \dots, a_{nn} . If we let t increase continuously from 0 to 1, the eigenvalues move continuously and, by Theorem 1, for each t, lie in the Gerschgorin disks with centers a_{ij} and radii

$$tr_j$$
 where $r_j = \sum_{k \neq j} |a_{jk}|$.

Since at the end, S is disjoint from the other disks, the assertion follows.

SECTION 20.8. Power Method for Eigenvalues, page 882

Purpose. Explanation of the power method for determining approximations and error bounds for eigenvalues of real symmetric matrices.

Main Content, Important Concepts

Iteration process of the power method

Rayleigh quotient (the approximate value)

Improvement of convergence by a spectral shift

Scaling (for eigenvectors)

Short Courses. Omit spectral shift.

Comments on Content

The method is simple but converges slowly, in general.

Symmetry of the matrix is essential to the validity of the error bound (2). The method as such can be applied to more general matrices.

SOLUTIONS TO PROBLEM SET 20.8, page 884

2. The Rayleigh quotients for the first five steps are (4S-values)

$$q = 0, 6, 7.846, 7.990, 7.999.$$

These approximate the larger eigenvalue 8, the other one being -2, which is included in the first two of the five resulting inclusion intervals.

$$[q - \delta, q + \delta] = [-4, 4]$$

$$[2, 10]$$

$$[6.615, 9.077]$$

$$[7.678, 8.302]$$

$$[7.921, 8.078]$$

where

$$\delta = 4, 4, 1.231, 0.312, 0.078.$$

Note that the first interval includes only -2.

4. Computation gives (4S-values)

\overline{q}	$\pm \delta$	$q - \delta$	$q + \delta$	Error
1.333	2.175	-0.841	3.508	5.867
6.574	1.742	4.832	8.316	0.626
7.158	0.474	6.684	7.632	0.042
7.197	0.119	7.078	7.317	0.003
7.1998	0.02983	7.170	7.230	0.0002

Note that the error is much less than the error bound. This is typical. Since the bound is optimal, it simply means that the present case is not the worst possible.

6. $q = 12.333, 12.962, 12.998; |\epsilon| \le 2.944, 0.614, 0.142$. The spectrum is

$$\{-1, 3, 13\}.$$

8. $q = 10.5000, 11.1303, 11.1831; |\epsilon| \le 2.9580, 1.3689, 0.9637$. The spectrum is (4S-values)

$$\sigma = \{-7.956, 0.4426, 4.283, 11.232\}.$$

10. Let $\mathbf{z}_1, \dots, \mathbf{z}_n$ be orthonormal eigenvectors corresponding to the eigenvalues $\lambda_1, \dots, \lambda_n$, respectively. Then for any initial vector \mathbf{x}_0 we have (summations from 1 to n)

$$\mathbf{x}_0 = \sum c_j \mathbf{z}_j$$

$$\mathbf{x}_1 = \sum c_j \lambda_j \mathbf{z}_j$$

$$\mathbf{x}_s = \sum c_j \lambda_j^s \mathbf{z}_j$$

$$\mathbf{x}_{s+1} = \sum c_j \lambda_j^{s+1} \mathbf{z}_j$$

and, using the last step, for the Rayleigh quotient

$$q = \frac{\mathbf{x}_{s+1}^{\mathsf{T}} \mathbf{x}_{s}}{\mathbf{x}_{s}^{\mathsf{T}} \mathbf{x}_{s}} = \frac{\sum c_{j}^{2} \lambda_{j}^{2s+1}}{\sum c_{j}^{2} \lambda_{j}^{2s}} \approx \lambda_{m}$$

where λ_m is of maximum absolute value, and the quality of the approximation increases with increasing number of steps s. Here we have to assume that $c_m \neq 0$, that is, that not just by chance we picked an \mathbf{x}_0 orthogonal to the eigenvector \mathbf{z}_m of λ_m . The chance that this happens is practically zero, but should it occur, then rounding will bring in a component in the direction of \mathbf{z}_m and one should eventually expect good approximations, although perhaps only after a great number of steps.

12. CAS Experiment. (a) We obtain

16, 41.2, 34.64, 32.888, 32.317, 32.116, 32.043, 32.0158, 32.0059, 32.0022,

etc. The spectrum is {32, 12, 8}. Eigenvectors are $[3 \ 6 \ -7]^T$, $[1 \ 0 \ -1]^T$, $[3 \ -2 \ 1]^T$.

(b) For instance, for A - 10I we get

$$q = 6$$
, 22.5, 22.166, 22.007, 22.00137,

etc. Whereas for **A** the ratio of eigenvalues is 32/12, for **A** -10**I** we have the spectrum $\{22, 2, -2\}$, hence the ratio 22/2. This explains the improvement of convergence.

(d) The eigenvalues are $\lambda = \pm 1$. Corresponding eigenvectors are

$$\mathbf{z}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \mathbf{z}_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

and we have chosen x_0 as

$$\mathbf{x}_0 = \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \mathbf{z}_1 + \mathbf{z}_2,$$

so that

$$\mathbf{x}_1 = \mathbf{z}_1 - \mathbf{z}_2 = [1 \ 3]^T$$

 $\mathbf{x}_2 = \mathbf{z}_1 + \mathbf{z}_2 = [3 \ -1]^T$

etc. From this,

$$\mathbf{x_0}^\mathsf{T}\mathbf{x_1} = 0$$

and for the error bound we get

$$\delta = \sqrt{\frac{{\mathbf{x_1}}^{\mathsf{T}} {\mathbf{x_1}}}{{\mathbf{x_0}}^{\mathsf{T}} {\mathbf{x_0}}} - q^2} = \sqrt{\frac{10}{10} - 0} = 1,$$

and similarly in all further steps. This illustrates that our error bound is best possible in general.

(e) A in (a) provides an example,

q	16	41.2	34.64	32.888	32.317
δ	59	5.5	1.45	0.557	0.225

etc. Also $\mathbf{A} - 10\mathbf{I}$ dose,

$$\frac{q}{\delta}$$
 6 22.5 22.166 22.007 22.00137 $\frac{q}{\delta}$ 59 3.0 0.072 0.023 0.00032

Further examples can easily be found. For instance, the matrix

$$\begin{bmatrix} -2 & 12 \\ 1 & -1 \end{bmatrix}$$

has the spectrum $\{-5, 2\}$, but we obtain, with $\mathbf{x_0} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}^\mathsf{T}$,

$$\frac{q}{\delta}$$
 5 -7 -4.3208 -5.2969 δ 5 2 0.37736 0.2006

etc.

SECTION 20.9. Tridiagonalization and QR-Factorization, page 885

Purpose. Explanation of an optimal method for determining the whole spectrum of a real symmetric matrix by first reducing the matrix to a tridiagonal matrix with the same spectrum and then applying the QR-method, an iteration in which each step consists of a factorization (5) and a multiplication.

Comment on Content

The n-2 Householder steps ($n \times n$ the size of the matrix) correspond to similarity transformations; hence the spectrum is preserved. The same holds for QR. But we can perform any number of QR steps, depending on the desired accuracy.

SOLUTIONS TO PROBLEM SET 20.9, page 893

2. We have to do n-2=1 Householder step. We obtain $S=\sqrt{2}$ and

$$\mathbf{v}_1 = [0 \quad \frac{1}{2}\sqrt{2 + \sqrt{2}} \quad 1/(\sqrt{2}\sqrt{2 + \sqrt{2}})]^{\mathsf{T}}.$$

This gives

$$\mathbf{P}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1/\sqrt{2} & -1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1 - 1/(2 + \sqrt{2}) \end{bmatrix}.$$

From this and the given matrix we obtain the tridiagonal matrix

$$\mathbf{B} = \mathbf{P_1} \mathbf{A} \mathbf{P_1} = \begin{bmatrix} 0 & -\sqrt{2} & 0 \\ -\sqrt{2} & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Hence -1 is an eigenvalue. The other eigenvalues, -1 and 2, are now obtained by solving the remaining quadratic characteristic equation.

4. We have to do two Householder steps because n-2=2.

Step1. We obtain $S_1 = 4.2426$ and

$$\mathbf{v_1} = [0 \quad 0.98560 \quad 0.11957 \quad 0.11957]^\mathsf{T}.$$

This gives

$$\mathbf{P_1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -0.94281 & -0.23570 & -0.23570 \\ 0 & -0.23570 & 0.97140 & -0.02860 \\ 0 & -0.23570 & -0.02860 & 0.97140 \end{bmatrix}$$

From this and the given matrix follows

$$\mathbf{A}_1 = \begin{bmatrix} 5 & -4.2426 & 0 & 0 \\ -4.2426 & 6 & -1 & -1 \\ 0 & -1 & 3.5 & 1.5 \\ 0 & -1 & 1.5 & 3.5 \end{bmatrix}.$$

Step 2. We obtain $S_2 = 1.4142$ and

$$\mathbf{v_2} = [0 \quad 0 \quad 0.38268 \quad -0.92388]^\mathsf{T}.$$

This gives

$$\mathbf{P_2} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.70711 & 0.70711 \\ 0 & 0 & 0.70711 & -0.70711 \end{vmatrix}$$

From this and the given matrix we obtain the tridiagonal matrix

$$\mathbf{B} = \mathbf{A}_2 = \begin{bmatrix} 5 & -4.2426 & 0 & 0 \\ -4.2426 & 6 & -1.4142 & 0 \\ 0 & -1.4142 & 5 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

6. The matrices **B** in the first three steps are

$$\begin{bmatrix} 1.2925 & -0.2942 & 0 \\ -0.2942 & 0.8402 & 0.1260 \\ 0 & 0.1260 & 0.2072 \end{bmatrix}$$
$$\begin{bmatrix} 1.3976 & -0.1697 & 0 \\ -0.1697 & 0.7608 & 0.0303 \\ 0 & 0.0303 & 0.1817 \end{bmatrix}$$
$$\begin{bmatrix} 1.4289 & -0.0886 & 0 \\ -0.0886 & 0.7310 & 0.0074 \\ 0 & 0.0074 & 0.1801 \end{bmatrix}$$

The eigenvalues are 1.44, 0.72, 0.18. The calculations were done with 8S and then rounded in the results to 4D. We see that the approximations of the eigenvalues are somewhat better than can be expected by looking at the size of the off-diagonal entries. Similarly for Probs. 7–9.

8. The matrices \mathbf{B} in the first three steps are

$$\begin{bmatrix} 14.2005 & -0.0088 & 0 \\ -0.0088 & -6.3052 & 0.0074 \\ 0 & 0.0074 & 2.1048 \end{bmatrix}$$

The eigenvalues are (5S)

$$14.200, -6.3052, 2.1048.$$

Note that the off-diagonal entries of the matrix **B** are small and that the approximations of the eigenvalues are more exact than in the other problems.

SOLUTIONS TO CHAPTER 20 REVIEW QUESTIONS AND PROBLEMS, page 893

14.
$$[4 t_1 \frac{1}{2}t_1]^{\mathsf{T}}$$

16.
$$[3 3t_1 + 2 t_1]^T$$

18.
$$\begin{bmatrix} 10 & -10 & -10 \\ -2.9924 & 3.1859 & 2.9578 \\ -5.6669 & 5.9641 & 5.9710 \end{bmatrix}$$
20.
$$\frac{1}{226} \begin{bmatrix} 48 & -8 & -6 \\ -8 & 39 & 1 \\ -6 & 1 & 29 \end{bmatrix}$$

$$\mathbf{20.} \ \, \frac{1}{226} \begin{bmatrix} 48 & -8 & -6 \\ -8 & 39 & 1 \\ -6 & 1 & 29 \end{bmatrix}$$

22. Reorder to get convergence.

Equation 1 becomes 2 2 becomes 3 3 becomes 1.

The iteration gives

$$\begin{bmatrix} -1.5067 \\ 8.1753 \\ -1.0846 \end{bmatrix}, \begin{bmatrix} -2.0193 \\ 7.9925 \\ -0.9967 \end{bmatrix}, \begin{bmatrix} -1.9992 \\ 8.0003 \\ -1.0001 \end{bmatrix}$$

The solution is $x_1 = -2, x_2 = 8, x_3 = -1$

24. 12.0,
$$\sqrt{71.50} = 8.4558$$
, 8.1

28. 152

30. 9.1

32.
$$(0.3 + 4.3 + 2.8)(10 + 10 + 10) = 222$$

34.
$$1.3 + 0.7x$$

36. Centers 2.0, 4.4, 2.8; radii 3.4, 2.1, 4.6, respectively. The eigenvalues are (4S-values) $0.05313, 4.573 \pm 2.514i.$

38. Centers 5, 6, 8; radii 2, 1, 1, respectively. Eigenvalues (4S-values) 4.186, 6.471, 8.343.

Note that the first disk includes the second and touches the third, the point x = 7 being common to all three disks. The third disk includes one eigenvalue and so does the second disk. The first disk includes the first two eigenvalues.

40. Computation gives

n	1	2	3	4
\overline{q}	76.67	92.46	95.81	96.46
δ	28.96	14.62	6.58	2.80

The spectrum is {2.635, 40.77, 96.60} (4S-values).