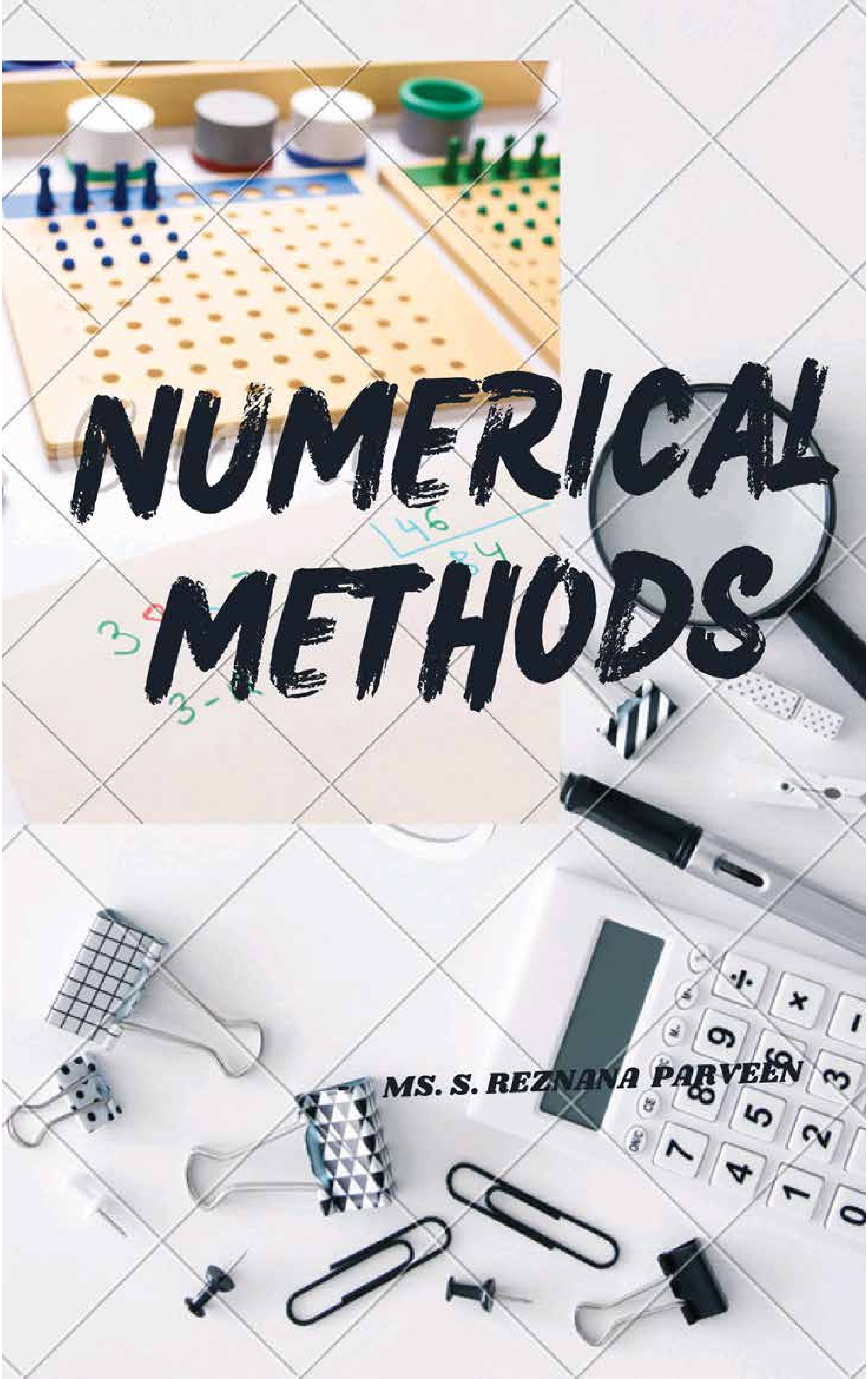


NUMERICAL METHODS

MS. S. REZNANA PARVEEN

UNIT - I

Solution of Algebraic and Transcendental Equations

- Solution of Algebraic and Transcendental Equations
  - Bisection Method
  - Method of False Position
  - The Iteration Method
  - Newton Raphson Method
- Summary
- Solved University Questions (JNTU)
- Objective Type Questions

1.1 Solution of Algebraic and Transcendental Equations

1.1.1 Introduction

A polynomial equation of the form

$$f(x) = p_n(x) = a_0 x^{n-1} + a_1 x^{n-1} + a_2 x^{n-2} + \dots + a_{n-1} x + a_n = 0 \qquad \text{.....(1)}$$

is called an Algebraic equation. For example,

$x^4 - 4x^2 + 5 = 0$, $4x^2 - 5x + 7 = 0$; $2x^3 - 5x^2 + 7x + 5 = 0$ are algebraic equations.

An equation which contains polynomials, trigonometric functions, logarithmic functions, exponential functions etc., is called a Transcendental equation. For example,

$\tan x - e^x = 0$; $\sin x - xe^{2x} = 0$; $x e^x = \cos x$

are transcendental equations.

Finding the roots or zeros of an equation of the form $f(x) = 0$ is an important problem in science and engineering. We assume that $f(x)$ is continuous in the required interval. A root of an equation $f(x) = 0$ is the value of x, say $x = \alpha$ for which $f(\alpha) = 0$. Geometrically, a root of an equation $f(x) = 0$ is the value of x at which the graph of the equation $y = f(x)$ intersects the x – axis (see Fig. 1)

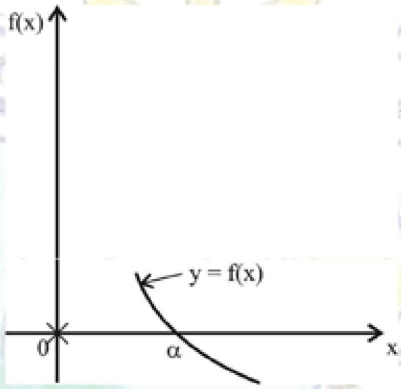


Fig. 1 Geometrical Interpretation of a root of $f(x) = 0$

A number α is a simple root of $f(x) = 0$; if $f(\alpha) = 0$ and $f'(\alpha) \neq 0$. Then, we can write $f(x)$ as,

$$f(x) = (x - \alpha)\, g(x),\ g(\alpha) \neq 0 \qquad \text{.....(2)}$$

A number α is a multiple root of multiplicity m of $f(x) = 0$, if $f(\alpha) = f^1(\alpha) = ... = f^{(m-1)}(\alpha) = 0$ and $f^m(\alpha) = 0$.

Then, $f(x)$ can be writhen as,

$$f(x) = (x - \alpha)^m\, g(x),\ g(\alpha) \neq 0 \qquad \text{.....(3)}$$

A polynomial equation of degree n will have exactly n roots, real or complex, simple or multiple. A transcendental equation may have one root or no root or infinite number of roots depending on the form of $f(x)$.

The methods of finding the roots of $f(x) = 0$ are classified as,

1. Direct Methods
2. Numerical Methods.

Direct methods give the exact values of all the roots in a finite number of steps. Numerical methods are based on the idea of successive approximations. In these methods, we start with one or two initial approximations to the root and obtain a sequence of approximations $x_0, x_1, \ldots x_k$ which in the limit as $k \to \infty$ converge to the exact root $x = a$.

There are no direct methods for solving higher degree algebraic equations or transcendental equations. Such equations can be solved by Numerical methods. In these methods, we first find an interval in which the root lies. If a and b are two numbers such that $f(a)$ and $f(b)$ have opposite signs, then a root of $f(x) = 0$ lies in between a and b. We take a or b or any valve in between a or b as first approximation x_1. This is further improved by numerical methods. Here we discuss few important Numerical methods to find a root of $f(x) = 0$.

1.1.2 Bisection Method

This is a very simple method. Identify two points $x = a$ and $x = b$ such that $f(a)$ and $f(b)$ are having opposite signs. Let $f(a)$ be negative and $f(b)$ be positive. Then there will be a root of $f(x) = 0$ in between a and b.

Let the first approximation be the mid point of the interval (a, b). i.e.

$$x_1 = \frac{(a+b)}{2}$$

If $f(x_1) = 0$, then x_1 is a root, other wise root lies between a and x_1 or x_1 and b according as $f(x_1)$ is positive or negative. Then again we bisect the interval and continue the process until the root is found to desired accuracy. Let $f(x_1)$ is positive, then root lies in between a and x_1 (see fig.2.). The second approximation to the root is given by,

$$x_2 = \frac{(a+x_1)}{2}$$

If $f(x_2)$ is negative, then next approximation is given by

$$x_3 = \frac{(x_2+x_1)}{2}$$

Similarly we can get other approximations. This method is also called Bolzano method.

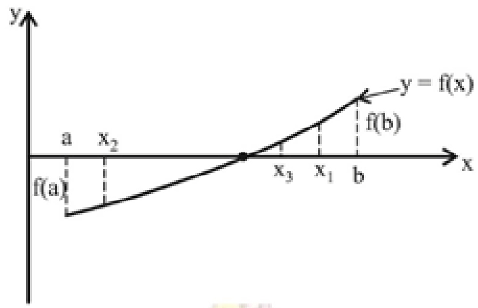


Fig. 2 Bisection Method

***Note*:** The interval width is reduced by a factor of one–half at each step and at the end of the n^{th} step, the new interval will be $[a_n, b_n]$ of length $\frac{|b-a|}{2^n}$. The number of iterations n required to achieve an accuracy $\in$ is given by,

$$n \geq \frac{\log_e\left(\frac{|b-a|}{\in}\right)}{\log_e 2} \qquad .....(4)$$

EXAMPLE 1

Find a real root of the equation $f(x) = x^3 - x - 1 = 0$, using Bisection method.

SOLUTION

First find the interval in which the root lies, by trail and error method.

$f(1) = 1^3 - 1 - 1 = -1$, which is negative

$f(2) = 2^3 - 2 - 1 = 5$, which is positive

$\therefore$ A root of $f(x) = x^3 - x - 1 = 0$ lies in between 1 and 2.

$$\therefore \quad x_1 = \frac{(1+2)}{2} = \frac{3}{2} = 1.5$$

$f(x_1) = f(1.5) = (1.5)^3 - 1.5 - 1 = 0.875$, which is positive.

Hence, the root lies in between 1 and 1.5

$$\therefore \quad x_2 = \frac{(1+1.5)}{2} = 1.25$$

$f(x_2) = f(1.25) = (1.25)^3 - 1.25 - 1 = -0.29$, which is negative.

Hence, the root lies in between 1.25 and 1.5

$$\therefore \qquad x_3 = \frac{(1.25+1.5)}{2} = 1.375$$

Similarly, we get $x_4 = 1.3125$, $x_5 = 1.34375$, $x_6 = 1.328125$ etc.

EXAMPLE 2

Find a root of $f(x) = xe^x - 1 = 0$, using Bisection method, correct to three decimal places.

SOLUTION

$$f(0) = 0.e^0 - 1 = -1 < 0$$

$$f(1) = 1.e^1 - 1 = 1.7183 > 0$$

Hence a root of $f(x) = 0$ lies in between 0 and 1.

$$\therefore \qquad x_1 = \frac{(0+1)}{2} = 0.5$$

$$f(0.5) = 0.5\, e^{0.5} - 1 = -0.1756$$

Hence the root lies in between 0.5 and 1

$$\therefore \qquad x_2 = \frac{(0.5+1)}{2} = 0.75$$

Proceeding like this, we get the sequence of approximations as follows.

$x_3 = 0.625$
$x_4 = 0.5625$
$x_5 = 0.59375$
$x_6 = 0.5781$
$x_7 = 0.5703$
$x_8 = 0.5664$
$x_9 = 0.5684$
$x_{10} = 0.5674$
$x_{11} = 0.5669$
$x_{12} = 0.5672$,
$x_{13} = 0.5671$,

Hence, the required root correct to three decimal places is, $x = 0.567$.

1.1.3 Method of False Position

This is another method to find the roots of $f(x) = 0$. This method is also known as Regular False Method.

In this method, we choose two points a and b such that $f(a)$ and $f(b)$ are of opposite signs. Hence a root lies in between these points. The equation of the chord joining the two points,

$(a, f(a))$ and $(b, f(b))$ is given by

$$\frac{y - f(a)}{x - a} = \frac{f(b) - f(a)}{b - a} \quad .....(5)$$

We replace the part of the curve between the points $[a, f(a)]$ and $[b, f(b)]$ by means of the chord joining these points and we take the point of intersection of the chord with the x axis as an approximation to the root (see Fig.3). The point of intersection is obtained by putting y = 0 in (5), as

$$x = x_1 = \frac{a\,f(b) - b\,f(a)}{f(b) - f(a)} \quad .....(6)$$

x_1 is the first approximation to the root of $f(x) = 0$.

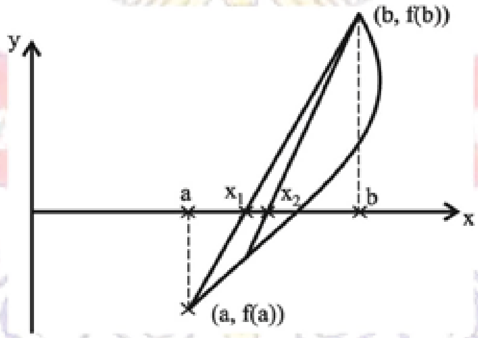


Fig. 3 Method of False Position

If f (x_1) and f (a) are of opposite signs, then the root lies between a and x_1 and we replace b by x_1 in (6) and obtain the next approximation x_2. Otherwise, we replace a by x_1 and generate the next approximation. The procedure is repeated till the root is obtained to the desired accuracy. This method is also called linear interpolation method or chord method.

EXAMPLE 3

Find a real root of the equation $f(x) = x^3 - 2x - 5 = 0$ by method of False position.

SOLUTION

$$f(2) = -1 \text{ and } f(3) = 16$$

Hence the root lies in between 2 and 3.

Take $a = 2$, $b = 3$.

$$\therefore \quad x_1 = \frac{a\,f(b) - b\,f(a)}{f(b) - f(a)}$$

$$= \frac{2(16) - 3(-1)}{16 - (-1)} = \frac{35}{17} = 2.058823529.$$

$$f(x_1) = f(2.058823529) = -0.390799917 < 0.$$

Therefore the root lies between 0.058823529 and 3. Again, using the formula, we get the second approximation as,

$$x_2 = \frac{2.058823529(16) - 3(-0.390799917)}{16 - (-0.390799917)} = 2.08126366$$

Proceeding like this, we get the next approximation as,

$$x_3 = 2.089639211,$$
$$x_4 = 2.092739575,$$
$$x_5 = 2.09388371,$$
$$x_6 = 2.094305452,$$
$$x_7 = 2.094460846$$

EXAMPLE 4

Determine the root of the equation $\cos x - x\,e^x = 0$ by the method of False position.

SOLUTION

$$f(0) = 1 \text{ and } f(1) = -2.177979523$$

$\therefore$ $a = 0$ and $b = 1$. The root lies in between 0 and 1

$$\therefore \quad x_1 = \frac{0(-2.177979523) - 1(1)}{-2.177979523 - 1} = 0.3146653378$$

$$f(x_1) = f(0.314653378) = 0.51986.$$

$\therefore$ The root lies in between 0.314653378 and 1.

Hence,
$$x_2 = \frac{0.3146653378(-2.177979523) - 1(0.51986)}{-2.177979523 - 0.51986} = 0.44673$$

Proceeding like this, we get

$$x_3 = 0.49402,$$
$$x_4 = 0.50995,$$
$$x_5 = 0.51520,$$
$$x_6 = 0.51692,$$

EXAMPLE 5

Determine the smallest positive root of $x - e^{-x} = 0$, correct of three significant figures using Regula False method.

SOLUTION

Here,
$$f(0) = 0 - e^{-0} = -1$$

and $f(1) = 1 - e^{-1} = 0.63212.$

$\therefore$ The smallest positive root lies in between 0 and 1. Here a = 0 and b = 1

$$\therefore \quad x_1 = \frac{0(0.63212) - 1(-1)}{0.63212 + 1} = 0.6127$$

$$f(0.6127) = 0.6127 - e^{-(0.6127)} = 0.0708$$

Hence, the next approximation lies in between 0 and 0.6127. Proceeding like this, we get

$x_2 = 0.57219,0$ $\quad x_3 = 0.5677,$ $\quad x_4 = 0.5672,$ $\quad x_5 = 0.5671,$

Hence, the smallest positive root, which is correct up to three decimal places is,

$x = 0.567$

1.1.4 The Iteration Method

In the previous methods, we have identified the interval in which the root of $f(x) = 0$ lies, we discuss the methods which require one or more starting values of x, which need not necessarily enclose the root of $f(x) = 0$. The iteration method is one such method, which requires one starting value of x.

We can use this method, if we can express $f(x) = 0$, as

$$x = \phi(x) \qquad \text{.....(1)}$$

We can express $f(x) = 0$, in the above form in more than one way also. For example, the equation $x^3 + x^2 - 1 = 0$ can be expressed in the following ways.

$$x = (1+x)^{\frac{-1}{2}}$$

$$x = (1-x^3)^{1/2}$$

$$x = (1-x^2)^{1/3}$$

and so on

Let x_0 be an approximation to the desired root ξ, which we can find graphically or otherwise. Substituting x_0 in right hand side of (1), we get the first approximation as

$$x_1 = \phi(x_0) \qquad \text{.....(2)}$$

The successive approximations are given by

$$x_2 = \phi(x_1)$$

$$x_3 = \phi(x_2) \qquad \text{.....(3)}$$

.

.

.

$$x_n = \phi(x_{n-1})$$

Note: The sequence of approximations $x_0, x_1, x_2 \dots x_n$ given by (3) converges to the root ξ in a interval I, if $|\phi'(x)| < 1$ for all x in I.

EXAMPLE 6

Using the method of iteration find a positive root between 0 and 1 of the equation

$$x\, e^x = 1$$

SOLUTION

The given equation can be writhen as $x = e^{-x}$

$\therefore \quad \phi(x) = e^{-x}$.

Here $|\phi'(x)| < 1$ for $x < 1$

$\therefore$ We can use iterative method

Let $x_0 = 1$

$$\therefore \quad x_1 = e^{-1} = \frac{1}{e} = 0.3678794.$$

$$x_2 = e^{-0.3678794} = 0.6922006.$$

$$x_3 = e^{-0.6922006} = 0.5004735$$

Proceeding like this, we get the required root as $x = 0.5671$.

EXAMPLE 7

Find the root of the equation $2x = \cos x + 3_1$ correct to three decimal places using Iteration method.

SOLUTION

Given equation can be written as

$$x = \frac{(\cos x + 3)}{2}$$

$$|\phi'(x)| = \left|\frac{\sin x}{2}\right| < 1$$

Hence iteration method can be applied

Let $x_0 = \frac{\pi}{2}$

$$\therefore \quad x_1 = \frac{1}{2}\left(\cos\frac{\pi}{2} + 3\right) = 1.5$$

$$x_2 = \frac{1}{2}(\cos 1.5 + 3) = 1.535$$

Similarly,

$x_3 = 1.518$,

$x_4 = 1.526$,

$x_5 = 1.522$,

$x_6 = 1.524$,

$x_7 = 1.523$,

$x_8 = 1.524$.

$\therefore$ The required root is $x = 1.524$

EXAMPLE 8

Find a real root of $2x - \log_{10} x = 7$ by the iteration method

SOLUTION

The given equation can be written as,

$$x = \frac{1}{2}(\log_{10} x + 7)$$

Let $x_0 = 3.8$

$$\therefore \quad x_1 = \frac{1}{2}(\log_{10} 3.8 + 7) = 3.79$$

$$x_2 = \frac{1}{2}(\log_{10} 3.79 + 7) = 3.7893$$

$$x_3 = \frac{1}{2}(\log_{10} 3.7893 + 7) = 3.7893.$$

$\therefore$ $x = 3.7893$ is a root of the given equation which is correct to four significant digits.

1.1.5 Newton Raphson Method

This is another important method. Let x_0 be approximation for the root of $f(x) = 0$. Let $x_1 = x_0 + h$ be the correct root so that $f(x_1) = 0$. Expanding $f(x_1) = f(x_0 + h)$ by Taylor series, we get

$$f(x_1) = f(x_1 + h) = f(x_0) + h\ f'(x_0) + \frac{h^2}{2!}\ f''(x_0) + \ldots\ldots = 0 \qquad \ldots..(1)$$

For small valves of h, neglecting the terms with h^2, h^3 ….. etc,. We get

$$\therefore \quad f(x_0) + h\ f'(x_0) = 0 \qquad \ldots..(2)$$

and $$h = -\frac{f(x_0)}{f^1(x_0)}$$

$\therefore$ $$x_1 = x_0 + h$$
$$= x_0 - \frac{f(x_0)}{f'(x_0)}$$

Proceeding like this, successive approximation $x_2, x_3, \ldots x_{n+1}$ are given by,

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad \ldots\ldots (3)$$

For n = 0, 1, 2, ……

***Note*:**

(i) The approximation x_{n+1} given by (3) converges, provided that the initial approximation x_0 is chosen sufficiently close to root of $f(x) = 0$.

(ii) Convergence of Newton-Raphson method: Newton-Raphson method is similar to iteration method

$$\phi(x) = x - \frac{f(x)}{f'(x)} \quad \ldots..(1)$$

differentiating (1) w.r.t to 'x' and using condition for convergence of iteration method i.e.

$$|\phi'(x)| < 1,$$

We get

$$\left|1 - \frac{f'(x).f'(x) - f(x)f''(x)}{[f'(x)]^2}\right| < 1$$

Simplifying we get condition for convergence of Newton-Raphson method is

$$|f(x).f''(x)| < [f(x)]^2$$

EXAMPLE 9

Find a root of the equation $x^2 - 2x - 5 = 0$ by Newton – Raphson method.

SOLUTION

Here $f(x) = x^3 - 2x - 5$.

$\therefore$ $$f^1(x) = 3x^2 - 2$$

Newton – Raphson method formula is

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

$$\therefore \quad x_{n+1} = x_n - \frac{x_n^3 - 2x_n - 5}{3x_n^2 - 2}, \qquad n = 0, 1, 2, \ldots\ldots (1)$$

Let $x_0 = 2$

$$\therefore \quad f(x_0) = f(2) = 2^3 - 2(2) - 5 = -1$$

and $$f^1(x_0) = f^1(2) = 3(2)^2 - 2 = 10$$

Putting n = 0 in (1), we get

$$x_1 = 2 - \left(\frac{-1}{10}\right) = 2.1$$

$$f(x_1) = f(2.1) = (2.1)^3 - 2(2.1) - 5 = 0.061$$

$$f^1(x_1) = f^1(2.1) = 3(2.1)^2 - 2 = 11.23$$

$$\therefore \quad x_2 = 2.1 - \frac{0.061}{11.23} = 2.094568$$

Similarly, we can calculate x_3, x_4 ……

EXAMPLE 10

Find a root of $x \sin x + \cos x = 0$, using Newton – Raphson method

SOLUTION

$$f(x) = x \sin x + \cos x.$$

$$\therefore \quad f'(x) = \sin x + x \cos x - \sin x = x \cos x$$

The Newton – Raphson method formula is,

$$x_{n+1} = x_n - \frac{x_n \sin x_n + \cos x_n}{x_n \cos x_n}, \qquad n = 0, 1, 2, \ldots.$$

Let $x_0 = \pi = 3.1416.$

$$\therefore \quad x_1 = 3.1416 - \frac{3.1416\sin\pi + \cos\pi}{3.1416\cos\pi} = 2.8233.$$

Similarly,

$x_2 = 2.7986$

$x_3 = 2.7984$

$x_4 = 2.7984$

$\therefore$ $x = 2.7984$ can be taken as a root of the equation $x \sin x + \cos x = 0$.

EXAMPLE 11

Find the smallest positive root of $x - e^{-x} = 0$, using Newton – Raphson method.

SOLUTION

Here $f(x) = x - e^{-x}$

$f^1(x) = 1 + e^{-x}$

$f(0) = -1$ and $f(1) = 0.63212$.

$\therefore$ The smallest positive root of $f(x) = 0$ lies in between 0 and 1.

Let $x_0 = 1$

The Newton – Raphson method formula is,

$$x_{n+1} = x_n - \frac{x_n - e^{-x_n}}{1 + e^{-x_n}}, \text{ n = 0, 1, 2, ……}$$

$f(0) = f(1) = 0.63212$

$f'(0) = f'(1) = 1.3679$

$$\therefore \quad x_1 = x_0 - \frac{x_0 - e^{-x_0}}{1 + e^{-x_0}} = 1 - \frac{0.63212}{1.3679} = 0.5379.$$

$f(0.5379) = -0.0461$

$f'(0.5379) = 1.584.$

$$\therefore \quad x_2 = 0.5379 + \frac{0.0461}{1.584} = 0.567$$

Similarly, $x_3 = 0.56714$

$\therefore$ $x = 0.567$ can be taken as the smallest positive root of $x - e^{-x} = 0$., correct to three decimal places.

***Note*:** A method is said to be of order P or has the rate of convergence P, if P is the largest positive real number for which there exists a finite constant c $\neq$ 0, such that

$$|\in_{K+1}| \leq c|\in_K|^P \qquad \text{….. (A)}$$

Where $\in_K = x_K - \xi$ is the error in the k[th] iterate. C is called Asymptotic Error constant and depends on derivative of $f(x)$ at $x = \xi$. It can be shown easily that the order of convergence of Newton – Raphson method is 2.

Exercise - 1.1

1. Using Bisection method find the smallest positive root of $x^3 - x - 4 = 0$ which is correct to two decimal places.

   [**Ans:** 1.80]

2. Obtain a root correct to three decimal places of $x^3 - 18 = 0$, using Bisection Method.

   [**Ans:** 2.621]

3. Find a root of the equation $xe^x - 1 = 0$ which lies in (0, 1), using Bisection Method.

   [**Ans:** 0.567]

4. Using Method of False position, obtain a root of $x^3 + x^2 + x + 7 = 0$, correct to three decimal places.

   [**Ans:** – 2.105]

5. Find the root of $x^3 - 2x^2 + 3x - 5 = 0$, which lies between 1 and 2, using Regula False method.

   [**Ans:** 1.8438]

6. Compute the real root of $x \log x - 1.2 = 0$, by the Method of False position.

   [**Ans:** 2.740]

7. Find the root of the equation $\cos x - x\, e^x = 0$, correct to four decimal places by Method of False position

   [**Ans:** 0.5178]

8. Using Iteration Method find a real root of the equation $x^3 - x^2 - 1 = 0$.

   [**Ans:** 1.466]

9. Find a real root of $\sin^2 x = x^2 - 1$, using iteration Method.

   [**Ans:** 1.404]

10. Find a root of $\sin x = 10\,(x - 1)$, using Iteration Method.

    [**Ans:** 1.088]

11. Find a real root of $\cot x = e^x$, using Iteration Method.

    [**Ans:** 0.5314]

12. Find a root of $x^4 - x - 10 = 0$ by Newton – Raphson Method.

    [**Ans:** 1.856]

13. Find a real root of $x - \cos x = 0$ by Newton – Raphson Method.

[**Ans:** 0.739]

14. Find a root of $2x - 3 \sin x - 5 = 0$ by Newton – Raphson Method.

[**Ans:** 2.883238]

15. Find a smallest positive root of $\tan x = x$ by Newton – Raphson Method.

[**Ans:** 4.4934]

Summary

Solution of algebraic and transcendental equations

1. The numerical methods to find the roots of $f(x) = 0$

   (i) *Bisection method:* If a function $f(x)$ is continuous between a and b, $f(a)$ and $f(b)$ are of apposite sign then there exists at least one root between a and b. The approximate value of the root between them is $x_0 = \frac{a+b}{2}$

   If $f(x_0) = 0$ then the x_0 is the correct root of $f(x) = 0$. If $f(x_0) \neq 0$, then the root either lies in between $\left(a, \frac{a+b}{2}\right)$ or $\left(\frac{a+b}{2}, b\right)$ depending on whether $f(x_0)$ is negative or positive. Again bisection the interval and repeat same method until the accurate root is obtained.

   (ii) *Method of false position:* (Regula false method): This is another method to find the root of $f(x) = 0$. In this method, we choose two points a and b such that $f(a)$, $f(b)$ are of apposite signs. Hence the root lies in between these points $[a, f(a)]$, $[b, f(b)]$ using equation of the chord joining these points and taking the point of intersection of the chord with the x-axis as an approximate root (using $y = 0$ on x– axis) is $x_1 = \frac{a\,f(b) - b\,f(a)}{f(b) - f(a)}$

   Repeat the same process till the root is obtained to the desired accuracy.

   (iii) *Newton Raphson method:* The successive approximate roots are given by

   $$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \; n = 0, 1, 2 \text{ - - - - -}$$

   provided that the initial approximate root x_0 is choosen sufficiently close to root of $f(x) = 0$

Solved University Questions

1. Find the root of the equation $2x - \log x = 7$ which lies between 3.5 and 4 by Regula–False method. (JNTU 2006)

Solution

Given $f(x) = 2x - \log x_{10} = 7$ …..(1)

Take $x_0 = 3.5, \quad x_1 = 4$

Using Regula Falsi method

$$x_2 = x_0 - \frac{x_1 - x_0}{f(x_1) - f(x)}.f(x_0)$$

$$x_2 = 3.5 - \frac{4 - 3.5}{(0.3979 + 0.5441)}(-0.5441)$$

$$x_2 = 3.7888$$

Now taking $x_0 = 3.7888$ and $x_1 = 4$

$$x_3 = x_0 - \frac{x_1 - x_0}{f(x_1) - f(x_0)}.f(x_0)$$

$$x_3 = 3.7888 - \frac{4 - 3.7888}{0.3988}(-0.0009)$$

$$x_3 = 3.7893$$

The required root is = 3.789

2. Find a real root of $xe^x = 3$ using Regula-Falsi method. (JNTU – 2006)

Solution

Given $f(x) = x\,e^x - 3 = 0$

$f(1) = e - 3 = -0.2817 < 0$

$f(2) = 2e^2 - 3 = 11.778 > 0$

$\therefore$ One root lies between 1 and 2

Now taking $x_0 = 1, x_1 = 2$

Using Regula – Falsi method

$$x_2 = x_0 - \frac{x_1 - x_0}{f(x_1) - f(x_0)}f(x_0)$$

$$\therefore \quad x_2 = \frac{x_0 f(x_1) - x_1 f(x_0)}{f(x_1) - f(x_0)}$$

$$x_2 = \frac{1(11.778) - 2(-0.2817)}{11.778 + 0.2817}$$

$$x_2 = 1.329$$

Now $f(x_2) = f(1.329) = 1.329\ e^{1.329} - 3 = 2.0199 > 0$

$f(1) = -0.2817 < 0$

$\therefore$ The root lies between 1 and 1.329 taking $x_0 = 1$ and $x_2 = 1.329$

$\therefore$ Taking $x_0 = 1$ and $x_2 = 1.329$

$$\therefore \quad x_3 = \frac{x_0 f(x_2) - x_2 f(x_0)}{f(x_2) - f(x_0)}$$

$$= \frac{1(2.0199) + (1.329)(0.2817)}{(2.0199) + (0.2817)}$$

$$= \frac{2.3942}{2.3016} = 1.04$$

Now $f(x^3) = 1.04\ e^{1.04} - 3 = -0.05 < 0$

The root lies between x^2 and x^3

i.e., 1.04 and 1.329 [$f(x_2) > 0$ and $f(x_3) < 0$]

$$\therefore \quad x_4 = \frac{x_2 f(x_3) - x_3 f(x_2)}{f(x_3) - f(x_2)} = \frac{(1.04)(-0.05) - (1.329)(2.0199)}{(-0.05) - (2.0199)}$$

$x_4 = 1.08$ is the approximate root

3. Find a real root of $e^x \sin x = 1$ using Regula – Falsi method (JNTU 2006)

Solution

Given $f(x) = e^x \sin x - 1 = 0$

Consider $x_0 = 2$

$f(x_0) = f(2) = e^2 \sin 2 - 1 = -0.7421 < 0$

$f(x_1) = f(3) = e^3 \sin 3 - 1 = 0.511 > 0$

$\therefore$ The root lies between 2 and 3

Using Regula – Falsi method

$$x_2 = \frac{x_0 f(x_1) - x_1 f(x_0)}{f(x_1) - f(x_0)}$$

$$x_2 = \frac{2(0.511) + 3(0.7421)}{0.511 + 0.7421}$$

$$x_2 = 2.93557$$

$f(x_2) = e^{2.93557} \sin(2.93557) - 1$

$f(x_2) = -0.35538 < 0$

$\therefore$ Root lies between x_2 and x_1

i.e., lies between 2.93557 and 3

$$x_3 = \frac{x_2 f(x_1) - x_1 f(x_2)}{f(x_1) - f(x_2)}$$

$$= \frac{(2.93557)(0.511) - 3(-35538)}{0.511 + 0.35538}$$

$x_3 = 2.96199$

$f(x_3) = e^{2.90199} \sin(2.96199) - 1 = -0.000819 < 0$

$\therefore$ root lies between x_3 and x_1

$$x_4 = \frac{x_3 f(x_1) - x_1 f(x_3)}{f(x_1) - f(x_3)}$$

$$x_4 = \frac{2.96199(0.511) + 3(0.000819)}{0.511 + 0.000819} = 2.9625898$$

$f(x^4) = e^{2.9625898} \sin(2.9625898) - 1$

$f(x^4) = -0.0001898 < 0$

$\therefore$ The root lies between x_4 and x_1

$$x_5 = \frac{x_4 f(x_1) - x_1 f(x_4)}{f(x_1) - f(x_4)}$$

$$= \frac{2.9625898(0.511) + 3(0.0001898)}{0.511 + (0.0001898)}$$

$$x_5 = 2.9626$$

we have

$$x_4 = 2.9625$$

$$x_5 = 2.9626$$

$\therefore$ $x_5 = x_4 = 2.962$

$\therefore$ The root lies between 2 and 3 is 2.962

4. Find a real root of $x\,e^x = 2$ using Regula – Falsi method (JNTU 2007)

Solution

$f(x) = x\,e^x - 2 = 0$

$f(0) = -2 < 0$, $f(1) = \text{i.e.}, -2 = (2.7183)-2$

$f(1) = 0.7183 > 0$

$\therefore$ The root lies between 0 and 1

Considering $x_0 = 0$, $x_1 = 1$

$f(0) = f(x_0) = -2$; $f(1) = f(x_1) = 0.7183$

By Regula – Falsi method

$$x_2 = \frac{x_0 f(x_1) - x_1 f(x_0)}{f(x_1) - f(x_0)}$$

$$x_2 = \frac{0(0.7183) - 1(-2)}{0.7183 - (-2)} = \frac{2}{2.7183}$$

$$x_2 = 0.73575$$

Now $f(x^2) = f(0.73575) = 0.73575\,e^{0.73575} - 2$

$f(x_2) = -0.46445 < 0$

and $f(x_1) = 0.7183 > 0$

$\therefore$ The root x_3 lies between x_1 and x_2

$$x_3 = \frac{x_2 f(x_1) - x_1 f(x_2)}{f(x_1) - f(x_2)}$$

$$x_3 = \frac{(0.73575)(0.7183)}{0.7183 + 0.46445}$$

$$x_3 = \frac{0.52848 + 0.46445}{1.18275}$$

$$x_3 = \frac{0.992939}{1.18275}$$

$$x_3 = 0.83951 \quad f(x^3) = \frac{(0.83951)}{(0.83951)e^{-2}}$$

$f(x_3) = (0.83951)\,e^{0.83951} - 2$

$f(x_3) = -0.056339 < 0$

$\therefore$ One root lies between x_1 and x_3

$$x_4 = \frac{x_3 f(x_1) - x_1 f(x_3)}{f(x_1) - f(x_3)} = \frac{(0.83951)(0.7183) - 1(-0.056339)}{0.7183 + 0.056339}$$

$$x_4 = \frac{0.65935}{0.774639} = 0.851171$$

$f(x_4) = 0.851171\ e0.851171 - 2 = -0.006227 < 0$

Now x_5 lies between x_1 and x_4

$$x_5 = \frac{x_4 f(x_1) - x_1 f(x_4)}{f(x_1) - f(x_4)}$$

$$x_5 = \frac{(0.851171)(0.7183) + (.006227)}{0.7183 + 0.006227}$$

$$x_5 = \frac{0.617623}{0.724527} = 0.85245$$

Now $f(x_5) = 0.85245\ e^{0.85245}\ e^{0.85245} - 2 = -0.0006756 < 0$

$\therefore$ One root lies between x_1 and x_5, (i.e., x_6 lies between x_1 and x_5)

Using Regula – Falsi method

$$x_6 = \frac{(0.85245)(0.7183) + 0.0006756}{0.7183 + 0.0006756}$$

$$x_6 = 0.85260$$

Now $f(x_6) = -0.00006736 < 0$

$\therefore$ One root x_7 lies between x_1 and x_6

By Regula – Falsi method

$$x_7 = \frac{x_6 f(x_1) - x_1 f(x_6)}{f(x_1) - f(x_6)}$$

$$x_7 = \frac{(0.85260)(0.7183) + 0.0006736}{0.7183 + 0.0006736}$$

$$x_7 = 0.85260$$

From $x^6 = 0.85260$ and $x_7 = 0.85260$

$\therefore$ A real root of the given equation is 0.85260

5. Using Newton-Raphson method (a) Find square root of a number (b) Find a reciprocal of a number [JNTU 2008]

Solution

(a) Let n be the number

and $x = \sqrt{n} \Rightarrow x^2 = n$

If $f(x) = x^2 - n = 0$ .....(1)

Then the solution to $f(x) = x^2 - n = 0$ is $x = \sqrt{n}$.

$f^1(x) = 2x$

by Newton Raphson method

$$x_{i+1} = x_i - \frac{f(x_i)}{f^1(x_i)} = x_i - \left(\frac{x_i^2 - n}{2x_i}\right)$$

$$x_{i+1} = \frac{1}{2}\left(x_i + \frac{x}{x_i}\right) \qquad \text{.....(2)}$$

using the above formula the square root of any number 'n' can be found to required accuracy

(b) To find the reciprocal of a number 'n'

$f(x) = \frac{1}{x} - n = 0$ .....(1)

$\therefore$ solution of (1) is $x = \frac{1}{n}$

$f^1(x) = -\frac{1}{x^2}$

Now by Newton-Raphson method, $x_{i+1} = x_i - \left(\frac{f(x_i)}{f^1(x_i)}\right)$

$$x_{i+1} = x_i - \left(\frac{\frac{1}{x_i} - N}{-\frac{1}{x_1^2}}\right)$$

$$x_{i+1} = x_i\,(2 - x_i\, n)$$

using the above formula the reciprocal of a number can be found to required accuracy.

6. Find the reciprocal of 18 using Newton–Raphson method [JNTU 2004]

Solution

The Newton-Raphson method

$$x_{i+1} = x_i\,(2 - x_i\, n) \quad \text{.....(1)}$$

considering the initial approximate value of x as $x_0 = 0.055$ and given $n = 18$

$\therefore \quad x_1 = 0.055\,[2 - (0.055)\,(18)]$

$\therefore \quad x_1 = 0.0555$

$x_2 = 0.0555\,[2 - 0.0555 \times 18]$

$x_2 = (0.0555)\,(1.001)$

$x_2 = 0.0555$

Hence $x_1 = x_2 = 0.0555$

$\therefore$ The reciprocal of 18 is 0.0555

7. Find a real root for $x \tan x + 1 = 0$ using Newton–Raphson method [JNTU 2006]

Solution

Given $f(x) = x \tan x + 1 = 0$

$f^1(x) = x \sec^2 x + \tan x$

$f(2) = 2 \tan 2 + 1 = -3.370079 < 0$

$f(3) = 2 \tan 3 + 1 = -0.572370 > 0$

$\therefore$ The root lies between 2 and 3

Take $\quad x_0 = \dfrac{2+3}{2} = 2.5$ (average of 2 and 3)

By Newton-Raphson method

$$x_{i+1} = x_i - \left(\frac{f(x_i)}{f^1(x_i)}\right)$$

$$x_1 = x_0 \left(\frac{f(x_0)}{f^1(x_0)}\right)$$

$$x_1 = 2.5 - \frac{(-0.86755)}{3.14808}$$

$$x_1 = 2.77558$$

$$x_2 = x_1 - \frac{f(x_i)}{f^1(x_i)};$$

$f(x_1) = -0.06383, \qquad f^1(x_1) = 2.80004$

$$x_2 = 2.77558 - \frac{(-0.06383)}{2.80004}$$

$$x_2 = 2.798$$

$f(x_2) = -0.001080, \qquad f^1(x_2) = 2.7983$

$$x_3 = x_2 - \frac{f(x_2)}{f^1(x_2)} = 2.798 - \frac{[-0.001080]}{2.7983}$$

$$x_3 = 2.798.$$

$\therefore \qquad x_2 = x_3$

$\therefore$ The real root of $x \tan x + 1 = 0$ is 2.798

8. Find a root of $e^x \sin x = 1$ using Newton–Raphson method [JNTU 2006]

Solution

Given $f(x) = e^x \sin x - 1 = 0$

$$f^1(x) = e^x \sec x + ex \cos x$$

Take $\qquad x_1 = 0, x_2 = 1$

$$f(0) = f(x_1) = e^0 \sin 0 - 1 = -1 < 0$$

$$f(1) = f(x_2) = e^1 \sin(1) - 1 = 1.287 > 0$$

The root of the equation lies between 0 and 1

Using Newton-Raphson method

$$x_{i+1} = x_i - \frac{f(x_i)}{f^1(x_i)}$$

Now consider x_0 = average of 0 and 1

$$x_0 = \frac{1+0}{2} = 0.5$$

$$x_0 = 0.5$$

$$f(x_0) = e^{0.5} \sin(0.5) - 1$$

$$f^1(x_0) = e^{0.5} \sin(0.5) + e^{0.5} \cos(0.5) = 2.2373$$

$$x_1 = x_0 - \frac{f(x_0)}{f^1(x_0)} = 0.5 - \frac{(-0.20956)}{2.2373}$$

$x_1 = 0.5936$

$f(x_1) = e^{0.5936} \sin(0.5936) - 1 = 0.0128$

$f^1(x_1) = e^{0.5936} \sin(0.5936) + e^{0.5936} \cos(0.5936) = 2.5136$

$$x_2 = x_1 - \frac{f(x_1)}{f^1(x_1)} = 0.5936 - \frac{(0.0128)}{2.5136}$$

$\therefore \quad x_2 = 0.58854$

similarly $\quad x_3 = x_2 - \dfrac{f(x_1)}{f^1(x_1)}$

$f(x_2) = e^{0.58854} \sin(0.58854) - 1 = 0.0000181$

$f^1(x_2) = e^{0.58854} \sin(0.58854) + e^{0.58854} \cos(0.58854)$

$f(x_2) = 2.4983$

$$\therefore \quad x_3 = 0.58854 - \frac{0.0000181}{2.4983}$$

$x_3 = 0.5885$

$\therefore \quad x_2 - x_3 = 0.5885$

0.5885 is the root of the equation $e^x \sin x - 1 = 0$

9. Find a real root of the equation $xe^x - \cos x = 0$ using Newton-Raphson method [JNTU-2006]

Solution

Given $f(x) = e^x - \cos x = 0$

$f^1(x) = xe^x + e^x + \sin x = (x+1)e^x + \sin x$

Take $\quad f(0) = 0 - \cos 0 = -1 < 0$

$f(1) = e - \cos 1 = 2.1779 > 0$

$\therefore$ The root lies between 0 and 1

Let $\quad x_0 = \dfrac{0+1}{2} = 0.5 \quad$ (average of 0 and 1)

Newton-Raphson method

$$x_{i+1} = x_i - \frac{f(x_i)}{f^1(x_i)}$$

$$x_{i+1} = x_0 - \frac{f(x_0)}{f^1(x_0)} = 0.5 - \frac{(-0.053221)}{(1.715966)}$$

$x_1 = 0.5310$

$f(x_1) = 0.040734, \qquad f^1(x_1) = 3.110063$

$$x_2 = x_1 - \frac{f(x_1)}{f^1(x_1)} = 0.5310 - \frac{0.040734}{3.110064}$$

$\therefore \quad x_2 = 0.5179\,; \quad f(x_2) = 0.0004339, \qquad f^1(x_2) = 3.0428504$

$$x_3 = 0.5179 - \frac{(0.0004339)}{3.0428504}$$

$x_3 = 0.5177$

$\therefore \quad f(x_3) = 0.000001106$

$f(x_3) = 3.04214$

$$x_4 = x_3 - \frac{f(x_3)}{f(x_3)} = 0.5177 - \frac{0.000001106}{3.04212}$$

$x_4 = 0.5177$

$\therefore \quad x_3 = x_4 = 0.5177$

$\therefore$ The root of $xe^x - \cos x = 0$ is 0.5177

10. Find a root of the equation $x^4 - x - 10 = 0$ using Bisection method correct to 2 decimal places. [JNTU 2008]

Solution

Let $f(x) = x^4 - x - 10 = 0$ be the given equation. We observe that f(1) < 0, then f(2) >0. So one root lies between 1 and 2.

$\therefore \quad$ Let $x_0 = 1$, $x_1 = 2$;

Take $x_2 = \frac{x_0 + x_1}{2} = 1.5; \qquad f(1.5) < 0;$

$\therefore \quad$ The root lies between 1.5 and 2

Let us take $x_3 = \frac{1.5 + 2}{2} = 1.75$; we find that $f(1.75) < 0$,

$\therefore \quad$ The root lies between 1.75 and 2

So we take now $x_4 = \frac{1.75 + 1.875}{2} = 1.8125 = 1.81$ can be taken as the root of the given equation.

11. Find a real root of equation $x^3 - x - 11 = 0$ by Bisection method. [JNTU-2007]

Solution

Given equation is $f(x) = x^3 - x - 11 = 0$

We observe that $f(2) = -5 < 0$ and $f(3) = 13 > 0$.

$\therefore$ A root of (1) lies between 2 and 3; take $x_0 = 2$, $x = 3$;

Let $x_2 = \dfrac{x_0 + x_1}{2} = \dfrac{2+3}{2} = 2.5$; Since $f(2.5) > 0$, the root lies between 2 and 2.5

$\therefore$ Taking $x_3 = \dfrac{2+2.5}{2} = 2.25$, we note that $f(2.25) < 0$;

$\therefore$ The root can be taken as lying between 2.25 and 2.5.

$\therefore$ The root = $\dfrac{2.25+2.5}{2} = 2.375$

12. Find a real root of $x^3 - 5x + 3 = 0$ using Bisection method. [JNTU-2007]

Solution

Let $f(x) = x^3 - 5x + 3 = 0$ be the equation given

Since $f(1) = -1 < 0$ and $f(2) = 1 > 0$, a real root lies between 1 and 2.

i.e., $x_0 = 1$, $x_1 = 2$; take $x_2 = \dfrac{1+2}{2} = 1.5$; $f(1.5) = -1.25 < 0$

$\therefore$ The root lies between 1.5 and 2;

$\therefore$ Take $x_3 = \dfrac{1.5+2}{2} = 1.75$

Now $f(1.75) = \left(\dfrac{7}{4}\right)^3 - 5\left(\dfrac{7}{4}\right) + 3 = -\text{ve}$;

$\therefore$ The root lies between 1.75 and 2

Let $x_4 = \dfrac{1.75+2}{2} = 1.875$;

We find that $f(1.875) = (1.875)^3 - 5(1.875) + 3 > 0$

$\therefore$ The root of the given equation lies between 1.75 and 1.875

$\therefore$ The root = $\dfrac{1.75+1.875}{2} = 1.813$

13. Find a real root of the equation $x^3 - 6x - 4 = 0$ by Bisection method [JNTU-2006]

Solution

Here $f(x) = x^3 - 6x - 4$

Take $x_0 = 2, x_1 = 3$; $(\because f(2) < 0, f(3) > 0)$

$x_1 = 2.5; f(x_1) < 0;$ take $x_3 = \dfrac{2.5+3}{2} = 2.75$

$f(2.75) > 0 \Rightarrow x_4 = \dfrac{2.5+2.75}{2} = 2.625$

$f(2.625) < 0 \Rightarrow$ Root lies between 2.625 and 2.75

$\therefore$ Approximately the root will be $= \dfrac{2.625+2.75}{2} = 2.69$

Objective Type Questions

I. Choose correct answer:

1. An example of an algebraic equation is

(1) $\tan x = e^x$ (2) $x = \log x$ (3) $x^3 - 5x + 3 = 0$ (4) None

[*Ans:* (3)]

2. An example of a transcendental equation is

(1) $x^3 - 2x - 10 = 0$ (2) $x^3 e^x = 5$

(3) $x^2 + 11x - 1 = 0$ (4) None

[*Ans:* (2)]

3. In finding a real root of the equation $x^3 - x - 10 = 0$ by bisection, if the root lies between $x_0 = 2$ and $x_1 = 3$, then, $x_2 =$

(1) 2.5 (2) 2.75 (3) 2.60 (4) None

[*Ans:* (1)]

4. If ϕ (a) and ϕ(b) are of opposite signs and the real root of the equation $\phi(x) = 0$ is found by false position method, the first approximation x_1, of the root is

(1) $\dfrac{a\,\phi(b) + b\,\phi(a)}{\phi(b) + \phi(a)}$ (2) $\dfrac{a\,\phi'(b) + b\,\phi'(a)}{\phi(b) + \phi(a)}$

(3) $\dfrac{ab\,\phi(a)\,\phi(b)}{\phi(a) - \phi(b)}$ (4) $\dfrac{a\,\phi(b) - b\,\phi(a)}{\phi(b) - \phi(a)}$

[*Ans:* (4)]

5. The two initial values of the roots of the equation $x^3 - x - 3 = 0$ are

(1) (–1, 0) (2) 1, 2 (3) –2, 1 (4) (1, 0)

[*Ans:* (2)]

6. The iteration method is said to have p^{th} order convergence if for any finite constant $K \neq 0$

(1) $|e_n| \leq K|e_{n-1}|^P$ (2) $|e_n| \leq K|e_{n+1}|^P$

(3) $|e_n + 1| \leq K|e_0|^P$ (4) None

[*Ans:* (1)]

7. Newton-Raphson method formula to find $(n+1)^{th}$ approximation of root of $f(x) = 0$ is

(1) $x_{n+1} = x_n - \frac{f'(x_n)}{f(x_n)}$ (2) $x_{n+1} = \frac{x_n f(x_n)}{f'(x_n)}$

(3) $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$ (4) None

[*Ans:* (3)]

8. In the bisection method e_0 is the initial error and e_n is the error in n^{th} iteration

(1) $\frac{1}{2}$ (2) 1 (3) $\frac{1}{2^n}$ (4) None

[*Ans:* (3)]

9. Which of the following methods has linear rate of convergence

(1) Regular flase (2) Bisection

(3) Newton-Raphson (4) None

[*Ans:* (1)]

10. A non linear equation $x^3 + x^2 - 1 = 0$ is $x = \phi(x)$, then the choice of $\phi(x)$ for which the iteration scheme $x_n = \phi(x_{n-1})$ $x_0 = 1$ converge is $\phi(x)$=

(1) $(1 - x^2)^{1/3}$ (2) $\frac{1}{\sqrt{1+x}}$ (3) $\sqrt{1-x^3}$ (d) None

[*Ans:* (2)]

Finding Square Roots Using Newton's Method

Let $A > 0$ be a positive real number. We want to show that there is a real number x with $x^2 = A$. We already know that for many real numbers, such as $A = 2$, there is no rational number x with this property. Formally, let $fx) := x^2 - A$. We want to solve the equation $f(x) = 0$.

Newton gave a useful general recipe for solving equations of the form $f(x) = 0$. Say we have some approximation x_k to a solution. He showed how to get a better approximation x_{k+1}. It works most of the time if your approximation is close enough to the solution.

Here's the procedure. Go to the point $(x_k, f(x_k))$ and find the tangent line. Its equation is

$$y = f(x_k) + f'(x_k)(x - x_k).$$

The next approximation, x_{k+1}, is where this tangent line crosses the x axis. Thus,

$$0 = f(x_k) + f'(x_k)(x_{k+1} - x_k), \quad \text{that is,} \quad x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}.$$

Applied to compute square roots, so $f(x) := x^2 - A$, this gives

$$x_{k+1} = \frac{1}{2}\left(x_k + \frac{A}{x_k}\right). \tag{1}$$

From this, by simple algebra we find that

$$x_{k+1} - x_k = \frac{1}{2x_k}(A - x_k^2). \tag{2}$$

Pick some x_0 so that $x_0^2 > A$. then equation (2) above shows that subsequent approximations $x_1, x_2, \ldots,$ are monotone decreasing. Equation (2) then shows that the sequence $x_1 \ge x_2 \ge x_3 \ge \ldots,$ is monotone decreasing and non-negative. By the monotone convergence property, it thus converges to some limit x.

I claim that $x^2 = A$. Rewrite (2) as $A - x_k^2 = 2x_k(x_{k+1} - x_k)$ and let $k \to \infty$. Since $x_{k+1} - x_k \to 0$ and x_k is bounded, this is obvious.

We now know that $\sqrt{A}$ exists as a real number. then it is simple to use (1) to verify that

$$x_{k+1} - \sqrt{A} = \frac{1}{2x_k}(x_k - \sqrt{A})^2. \tag{3}$$

Equation (3) measures the error $x_{k+1} - \sqrt{A}$. It shows that the error at the next step is the *square* of the error in the previous step. Thus, if the error at some step is roughly 10^{-6} (so 6 decimal places), then at the next step the error is roughly 10^{-12} (so 12 decimal places).

Example: To 20 decimal places, $\sqrt{7} = 2.6457513110645905905$. Let's see what Newton's method gives with the initial approximation $x_0 = 3$:

$$x_1 = 2.6666666666666666666 \qquad x_2 = 2.6458333333333333333$$
$$x_3 = 2.6457513123359580052 \qquad x_4 = 2.6457513110645905908$$

Remarkable accuracy.

10 NUMERICAL METHODS

Carl Gustav Jacob Jacobi

1 8 0 4 – 1 8 5 1

Carl Gustav Jacob Jacobi was the second son of a successful banker in Potsdam, Germany. After completing his secondary schooling in Potsdam in 1821, he entered the University of Berlin. In 1825, having been granted a doctorate in mathematics, Jacobi served as a lecturer at the University of Berlin. Then he accepted a position in mathematics at the University of Königsberg.

Jacobi's mathematical writings encompassed a wide variety of topics, including elliptic functions, functions of a complex variable, functional determinants (called Jacobians), differential equations, and Abelian functions. Jacobi was the first to apply elliptic functions to the theory of numbers, and he was able to prove a longstanding conjecture by Fermat that every positive integer can be written as the sum of four perfect squares. (For instance, $10 = 1^2 + 1^2 + 2^2 + 2^2$.) He also contributed to several branches of mathematical physics, including dynamics, celestial mechanics, and fluid dynamics.

In spite of his contributions to applied mathematics, Jacobi did not believe that mathematical research needed to be justified by its applicability. He stated that the sole end of science and mathematics is "the honor of the human mind" and that "a question about numbers is worth as much as a question about the system of the world."

Jacobi was such an incessant worker that in 1842 his health failed and he retired to Berlin. By the time of his death in 1851, he had become one of the most famous mathematicians in Europe.

10.1 GAUSSIAN ELIMINATION WITH PARTIAL PIVOTING

In Chapter 1 two methods for solving a system of n linear equations in n variables were discussed. When either of these methods (Gaussian elimination and Gauss-Jordan elimination) is used with a digital computer, the computer introduces a problem that has not yet been discussed—**rounding error.**

Digital computers store real numbers in **floating point form,**

$$\pm M \times 10^k,$$

where k is an integer and the **mantissa** M satisfies the inequality $0.1 \le M < 1$. For instance, the floating point forms of some real numbers are as follows.

Real Number	*Floating Point Form*
527	0.527×10^3
-3.81623	-0.381623×10^1
0.00045	0.45×10^{-3}

The number of decimal places that can be stored in the mantissa depends on the computer. If n places are stored, then it is said that the computer stores ***n* significant digits.** Additional digits are either truncated or rounded off. When a number is **truncated** to n significant digits, all digits after the first n significant digits are simply omitted. For instance, truncated to two significant digits, the number 0.1251 becomes 0.12.

When a number is **rounded** to n significant digits, the last retained digit is increased by one if the discarded portion is greater than half a digit, and the last retained digit is not changed if the discarded portion is less than half a digit. For instance, rounded to two significant digits, 0.1251 becomes 0.13 and 0.1249 becomes 0.12. For the special case in which the discarded portion is precisely half a digit, round so that the last retained digit is even. So, rounded to two significant digits, 0.125 becomes 0.12 and 0.135 becomes 0.14.

Whenever the computer truncates or rounds, a rounding error that can affect subsequent calculations is introduced. The result after rounding or truncating is called the **stored value.**

EXAMPLE 1 ***Finding the Stored Value of Number***

Determine the stored value of each of the following real numbers in a computer that rounds to three significant digits.

(a) 54.7 (b) 0.1134 (c) -8.2256
(d) 0.08335 (e) 0.08345

Solution

Number	*Floating Point Form*	*Stored Value*
(a) 54.7	0.547×10^2	0.547×10^2
(b) 0.1134	0.1134×10^0	0.113×10^0
(c) -8.2256	-0.82256×10^1	-0.823×10^1
(d) 0.08335	0.8335×10^{-1}	0.834×10^{-1}
(e) 0.08345	0.8345×10^{-1}	0.834×10^{-1}

Note in parts (d) and (e) that when the discarded portion of a decimal is precisely half a digit, the number is rounded so that the stored value ends in an even digit.

REMARK: Most computers store numbers in binary form (base two) rather than decimal form (base ten). Because rounding occurs in both systems, however, this discussion will be restricted to the more familiar base ten.

Rounding error tends to propagate as the number of arithmetic operations increases. This phenomenon is illustrated in the following example.

EXAMPLE 2 ***Propagation of Rounding Error***

Evaluate the determinant of the matrix

$$A = \begin{bmatrix} 0.12 & 0.23 \\ 0.12 & 0.12 \end{bmatrix},$$

when rounding each intermediate calculation to two significant digits. Then find the exact solution and compare the two results.

Solution Rounding each intermediate calculation to two significant digits, produces the following.

$$\begin{aligned} |A| &= (0.12)(0.12) - (0.12)(0.23) \\ &= 0.0144 - 0.0276 \\ &\approx 0.014 - 0.028 \qquad \text{Round to two significant digits} \\ &= -0.014 \end{aligned}$$

However, the exact solution is

$$\begin{aligned} |A| &= 0.0144 - 0.0276 \\ &= -0.0132. \end{aligned}$$

So, to two significant digits, the correct solution is -0.013. Note that the rounded solution is not correct to two significant digits, even though each arithmetic operation was performed with two significant digits of accuracy. This is what is meant when it is said that arithmetic operations tend to propagate rounding error.

In Example 2, rounding at the intermediate steps introduced a rounding error of

$$-0.0132 - (-0.014) = 0.0008. \qquad \text{Rounding error}$$

Although this error may seem slight, it represents a **percentage error** of

$$\frac{0.0008}{0.0132} \approx 0.061 = 6.1\%. \qquad \text{Percentage error}$$

In most practical applications, a percentage error of this magnitude would be intolerable. Keep in mind that this particular percentage error arose with only a few arithmetic steps. When the number of arithmetic steps increases, the likelihood of a large percentage error also increases.

TECHNOLOGY NOTE

You can see the effect of rounding on a calculator. For example, the determinant of

$$A = \begin{bmatrix} 3 & 11 \\ 2 & 6 \end{bmatrix}$$

is -4. However, the TI-86 calculates the greatest integer of the determinant of A to be -5: ***int det A*** $= -5$. Do you see what happened?

Gaussian Elimination with Partial Pivoting

For large systems of linear equations, Gaussian elimination can involve hundreds of arithmetic computations, each of which can produce rounding error. The following straightforward example illustrates the potential magnitude of the problem.

EXAMPLE 3 ***Gaussian Elimination and Rounding Error***

Use Gaussian elimination to solve the following system.

$$\begin{aligned} 0.143x_1 + 0.357x_2 + 2.01x_3 &= -5.173 \\ -1.31x_1 + 0.911x_2 + 1.99x_3 &= -5.458 \\ 11.2x_1 - 4.30x_2 - 0.605x_3 &= 4.415 \end{aligned}$$

After *each* intermediate calculation, round the result to three significant digits.

Solution Applying Gaussian elimination to the augmented matrix for this system produces the following.

$$\begin{bmatrix} 0.143 & 0.357 & 2.01 & -5.17 \\ -1.31 & 0.911 & 1.99 & -5.46 \\ 11.2 & -4.30 & -0.605 & 4.42 \end{bmatrix}$$

$$\begin{bmatrix} 1.00 & 2.50 & 14.1 & -36.2 \\ -1.31 & 0.911 & 1.99 & -5.46 \\ 11.2 & -4.30 & -0.605 & 4.42 \end{bmatrix}$$ ← Dividing the first row by 0.143 produces a new first row.

$$\begin{bmatrix} 1.00 & 2.50 & 14.1 & -36.2 \\ 0.00 & 4.19 & 20.5 & -52.9 \\ 11.2 & -4.30 & -0.605 & 4.42 \end{bmatrix}$$ ← Adding 1.31 times the first row to the second row produces a new second row.

$$\begin{bmatrix} 1.00 & 2.50 & 14.1 & -36.2 \\ 0.00 & 4.19 & 20.5 & -52.9 \\ 0.00 & -32.3 & -159. & 409. \end{bmatrix}$$ ← Adding −11.2 times the first row to the third row produces a new third row.

$$\begin{bmatrix} 1.00 & 2.50 & 14.1 & -36.2 \\ 0.00 & 1.00 & 4.89 & -12.6 \\ 0.00 & -32.3 & -159. & 409. \end{bmatrix}$$ ← Dividing the second row by 4.19 produces a new second row.

$$\begin{bmatrix} 1.00 & 2.50 & 14.1 & -36.2 \\ 0.00 & 1.00 & 4.89 & -12.6 \\ 0.00 & 0.00 & -1.00 & 2.00 \end{bmatrix}$$ ← Adding 32.3 times the second row to the third row produces a new third row.

$$\begin{bmatrix} 1.00 & 2.50 & 14.1 & -36.2 \\ 0.00 & 1.00 & 4.89 & -12.6 \\ 0.00 & 0.00 & 1.00 & -2.00 \end{bmatrix}$$ ← Multiplying the third row by −1 produces a new third row.

So $x_3 = -2.00$, and using back-substitution, you can obtain $x_2 = -2.82$ and $x_1 = -0.950$. Try checking this "solution" in the original system of equations to see that it is not correct. (The correct solution is $x_1 = 1$, $x_2 = 2$, and $x_3 = -3$.)

What went wrong with the Gaussian elimination procedure used in Example 3? Clearly, rounding error propagated to such an extent that the final "solution" became hopelessly inaccurate.

Part of the problem is that the original augmented matrix contains entries that differ in orders of magnitude. For instance, the first column of the matrix

$$\begin{bmatrix} 0.143 & 0.357 & 2.01 & -5.17 \\ -1.31 & 0.911 & 1.99 & -5.46 \\ 11.2 & -4.30 & -0.605 & 4.42 \end{bmatrix}$$

has entries that increase roughly by powers of ten as one moves down the column. In subsequent elementary row operations, the first row was multiplied by 1.31 and −11.2 and the second row was multiplied by 32.3. When floating point arithmetic is used, such large row multipliers tend to propagate rounding error. This type of error propagation can be lessened by appropriate row interchanges that produce smaller multipliers. One method for restricting the size of the multipliers is called **Gaussian elimination with partial pivoting.**

Gaussian Elimination with Partial Pivoting

1. Find the entry in the left column with the largest absolute value. This entry is called the **pivot.**
2. Perform a row interchange, if necessary, so that the pivot is in the first row.
3. Divide the first row by the pivot. (This step is unnecessary if the pivot is 1.)
4. Use elementary row operations to reduce the remaining entries in the first column to zero.

The completion of these four steps is called a **pass.** After performing the first pass, ignore the first row and first column and repeat the four steps on the remaining submatrix. Continue this process until the matrix is in row-echelon form.

Example 4 shows what happens when this partial pivoting technique is used on the system of linear equations given in Example 3.

EXAMPLE 4 ***Gaussian Elimination with Partial Pivoting***

Use Gaussian elimination with partial pivoting to solve the system of linear equations given in Example 3. After *each* intermediate calculation, round the result to three significant digits.

Solution As in Example 3, the augmented matrix for this system is

$$\begin{bmatrix} 0.143 & 0.357 & 2.01 & -5.17 \\ -1.31 & 0.911 & 1.99 & -5.46 \\ 11.2 & -4.30 & -0.605 & 4.42 \end{bmatrix}.$$

↑ *Pivot* (11.2)

In the left column 11.2 is the pivot because it is the entry that has the largest absolute value. So, interchange the first and third rows and apply elementary row operations as follows.

$$\begin{bmatrix} 11.2 & -4.30 & -0.605 & 4.42 \\ -1.31 & 0.911 & 1.99 & -5.46 \\ 0.143 & 0.357 & 2.01 & -5.17 \end{bmatrix}$$ ← Interchange the first and third rows.

$$\begin{bmatrix} 1.00 & -0.384 & -0.0540 & 0.395 \\ -1.31 & 0.911 & 1.99 & -5.46 \\ 0.143 & 0.357 & 2.01 & -5.17 \end{bmatrix}$$ ← Dividing the first row by 11.2 produces a new first row.

$$\begin{bmatrix} 1.00 & -0.384 & -0.0540 & 0.395 \\ 0.00 & 0.408 & 1.92 & -4.94 \\ 0.143 & 0.357 & 2.01 & -5.17 \end{bmatrix} \leftarrow$$

Adding 1.31 times the first row to the second row produces a new second row.

$$\begin{bmatrix} 1.00 & -0.384 & -0.0540 & 0.395 \\ 0.00 & 0.408 & 1.92 & -4.94 \\ 0.00 & 0.412 & 2.02 & -5.23 \end{bmatrix} \leftarrow$$

Adding −0.143 times the first row to the third row produces a new third row.

This completes the first pass. For the second pass consider the submatrix formed by deleting the first row and first column. In this matrix the pivot is 0.412, which means that the second and third rows should be interchanged. Then proceed with Gaussian elimination as follows.

Pivot

$$\begin{bmatrix} 1.00 & -0.384 & -0.0540 & 0.395 \\ 0.00 & 0.412 & 2.02 & -5.23 \\ 0.00 & 0.408 & 1.92 & -4.94 \end{bmatrix} \begin{matrix} \\ \leftarrow \\ \leftarrow \end{matrix}$$

Interchange the second and third rows.

$$\begin{bmatrix} 1.00 & -0.384 & -0.0540 & 0.395 \\ 0.00 & 1.00 & 4.90 & -12.7 \\ 0.00 & 0.408 & 1.92 & -4.94 \end{bmatrix} \leftarrow$$

Dividing the second row by 0.412 produces a new second row.

$$\begin{bmatrix} 1.00 & -0.384 & -0.0540 & 0.395 \\ 0.00 & 1.00 & 4.90 & -12.7 \\ 0.00 & 0.00 & -0.0800 & 0.240 \end{bmatrix} \leftarrow$$

Adding −0.408 times the second row to the third row produces a new third row.

This completes the second pass, and you can complete the entire procedure by dividing the third row by -0.0800 as follows.

$$\begin{bmatrix} 1.00 & -0.384 & -0.0540 & 0.395 \\ 0.00 & 1.00 & 4.90 & -12.7 \\ 0.00 & 0.00 & 1.00 & -3.00 \end{bmatrix} \leftarrow$$

Dividing the third row by −0.0800 produces a new third row.

So $x_3 = -3.00$, and back-substitution produces $x_2 = 2.00$ and $x_1 = 1.00$, which agrees with the exact solution of $x_1 = 1$, $x_2 = 2$, and $x_3 = -3$ when rounded to three significant digits.

REMARK: Note that the row multipliers used in Example 4 are 1.31, -0.143, and -0.408, as contrasted with the multipliers of 1.31, 11.2, and 32.3 encountered in Example 3.

The term *partial* in partial pivoting refers to the fact that in each pivot search only entries in the left column of the matrix or submatrix are considered. This search can be extended to include every entry in the coefficient matrix or submatrix; the resulting technique is called **Gaussian elimination with complete pivoting.** Unfortunately, neither complete pivoting nor partial pivoting solves all problems of rounding error. Some systems of linear equations, called **ill-conditioned** systems, are extremely sensitive to numerical errors. For such systems, pivoting is not much help. A common type of system of linear equations that tends to be ill-conditioned is one for which the determinant of the coefficient matrix is nearly zero. The next example illustrates this problem.

EXAMPLE 5 ***An Ill-Conditioned System of Linear Equations***

Use Gaussian elimination to solve the following system of linear equations.

$$x + y = 0$$
$$x + \frac{401}{400}y = 20$$

Round each intermediate calculation to four significant digits.

Solution Using Gaussian elimination with rational arithmetic, you can find the exact solution to be $y = 8000$ and $x = -8000$. But rounding $401/400 = 1.0025$ to four significant digits introduces a large rounding error, as follows.

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1.002 & 20 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0.002 & 20 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1.00 & 10{,}000 \end{bmatrix}$$

So $y = 10{,}000$ and back-substitution produces

$$\begin{aligned} x &= -y \\ &= -10{,}000. \end{aligned}$$

This "solution" represents a percentage error of 25% for both the x-value and the y-value. Note that this error was caused by a rounding error of only 0.0005 (when you rounded 1.0025 to 1.002).

SECTION 10.1 ❑ EXERCISES

In Exercises 1–8, express the real number in floating point form.

1. 4281 **2.** 321.61 **3.** -2.62 **4.** -21.001
5. -0.00121 **6.** 0.00026 **7.** $\frac{1}{8}$ **8.** $16\frac{1}{2}$

In Exercises 9–16, determine the stored value of the real number in a computer that rounds to (a) three significant digits and (b) four significant digits.

9. 331 **10.** 21.4 **11.** -92.646 **12.** 216.964
13. $\frac{7}{16}$ **14.** $\frac{[illegible]}{32}$ **15.** $\frac{[illegible]}{7}$ **16.** $\frac{[illegible]}{3}$

In Exercises 17 and 18, evaluate the determinant of the matrix, rounding each intermediate calculation to three significant digits. Then compare the rounded value with the exact solution.

17. $\begin{bmatrix} 1.24 & 56.00 \\ 66.00 & 1.02 \end{bmatrix}$ **18.** $\begin{bmatrix} 2.12 & 4.22 \\ 1.07 & 2.12 \end{bmatrix}$

In Exercises 19 and 20, use Gaussian elimination to solve the system of linear equations. After each intermediate calculation, round the result to three significant digits. Then compare this solution with the exact solution.

19. $1.21x + 16.7y = 28.8$
$4.66x + 64.4y = 111.0$

20. $14.4x - 17.1y = 31.5$
$81.6x - 97.4y = 179.0$

In Exercises 21–24, use Gaussian elimination without partial pivoting to solve the system of linear equations, rounding to three significant digits after each intermediate calculation. Then use partial pivoting to solve the same system, again rounding to three significant digits after each intermediate calculation. Finally, compare both solutions with the given exact solution.

21. $x + 1.04y = 2.04$
$6x + 6.20y = 12.20$
(Exact: $x = 1, y = 1$)

22. $0.51x + 92.6y = 97.7$
$99.00x - 449.0y = 541.0$
(Exact: $x = 10, y = 1$)

23. $x + 4.01y + 0.00445z = 0.00$
$-x - 4.00y + 0.00600z = 0.21$
$2x - 4.05y + 0.05000z = -0.385$
(Exact: $x = -0.49, y = 0.1, z = 20$)

24. $0.007x + 61.20y + 0.093z = 61.3$
$4.810x - 5.92y + 1.110z = 0.0$
$81.400x + 1.12y + 1.180z = 83.7$
(Exact: $x = 1, y = 1, z = 1$)

In Exercises 25 and 26, use Gaussian elimination to solve the ill-conditioned system of linear equations, rounding each intermediate calculation to three significant digits. Then compare this solution with the given exact solution.

25. $x + y = 2$
$x + \frac{600}{601}y = 20$
(Exact: $x = 10{,}820$, $y = -10{,}818$)

26. $x - \frac{800}{801}y = 10$
$-x + y = 50$
(Exact: $x = 48{,}010$, $y = 48{,}060$)

27. Consider the ill-conditioned systems

$x + y = 2$ and $x + y = 2$
$x + 1.0001y = 2$ $\quad x + 1.0001y = 2.0001.$

Calculate the solution to each system. Notice that although the systems are almost the same, their solutions differ greatly.

28. Repeat Exercise 27 for the systems

$x - y = 0$ and $x - y = 0$
$-1.001x + y = 0.001$ $\quad -1.001x + y = 0.$

29. The **Hilbert matrix** of size $n \times n$ is the $n \times n$ symmetric matrix $H_n = [a_{ij}]$, where $a_{ij} = 1/(i + j - 1)$. As n increases, the Hilbert matrix becomes more and more ill-conditioned. Use Gaussian elimination to solve the following system of linear equations, rounding to two significant digits after each intermediate calculation. Compare this solution with the exact solution ($x_1 = 3, x_2 = -24$, and $x_3 = 30$).

$x_1 + \frac{1}{2}x_2 + \frac{1}{3}x_3 = 1$
$\frac{1}{2}x_1 + \frac{1}{3}x_2 + \frac{1}{4}x_3 = 1$
$\frac{1}{3}x_1 + \frac{1}{4}x_2 + \frac{1}{5}x_3 = 1$

30. Repeat Exercise 29 for $H_4\mathbf{x} = \mathbf{b}$, where $\mathbf{b} = (1, 1, 1, 1)^T$, rounding to four significant digits. Compare this solution with the exact solution ($x_1 = -4, x_2 = 60, x_3 = -180$, and $x_4 = 140$).

31. The inverse of the $n \times n$ Hilbert matrix H_n has integer entries. Use your computer of graphing calculator to calculate the inverses of the Hilbert matrices H_n for $n = 4, 5, 6$, and 7. For what values of n do the inverses appear to be accurate?

10.2 ITERATIVE METHODS FOR SOLVING LINEAR SYSTEMS

As a numerical technique, Gaussian elimination is rather unusual because it is *direct*. That is, a solution is obtained after a single application of Gaussian elimination. Once a "solution" has been obtained, Gaussian elimination offers no method of refinement. The lack of refinements can be a problem because, as the previous section shows, Gaussian elimination is sensitive to rounding error.

Numerical techniques more commonly involve an iterative method. For example, in calculus you probably studied Newton's iterative method for approximating the zeros of a differentiable function. In this section you will look at two iterative methods for approximating the solution of a system of n linear equations in n variables.

The Jacobi Method

The first iterative technique is called the **Jacobi method,** after Carl Gustav Jacob Jacobi (1804–1851). This method makes two assumptions: (1) that the system given by

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned}$$

has a unique solution and (2) that the coefficient matrix A has no zeros on its main diagonal. If any of the diagonal entries $a_{11}, a_{22}, \ldots, a_{nn}$ are zero, then rows or columns must be interchanged to obtain a coefficient matrix that has nonzero entries on the main diagonal.

To begin the Jacobi method, solve the first equation for x_1, the second equation for x_2, and so on, as follows.

$$\begin{aligned} x_1 &= \frac{1}{a_{11}}(b_1 - a_{12}x_2 - a_{13}x_3 - \cdots - a_{1n}x_n) \\ x_2 &= \frac{1}{a_{22}}(b_2 - a_{21}x_1 - a_{23}x_3 - \cdots - a_{2n}x_n) \\ &\vdots \\ x_n &= \frac{1}{a_{nn}}(b_n - a_{n1}x_1 - a_{n2}x_2 - \cdots - a_{n,n-1}x_{n-1}) \end{aligned}$$

Then make an *initial approximation* of the solution,

$(x_1, x_2, x_3, \ldots, x_n)$, Initial approximation

and substitute these values of x_i into the right-hand side of the rewritten equations to obtain the *first approximation*. After this procedure has been completed, one **iteration** has been performed. In the same way, the second approximation is formed by substituting the first approximation's x-values into the right-hand side of the rewritten equations. By repeated iterations, you will form a sequence of approximations that often **converges** to the actual solution. This procedure is illustrated in Example 1.

EXAMPLE 1 ***Applying the Jacobi Method***

Use the Jacobi method to approximate the solution of the following system of linear equations.

$$\begin{aligned} 5x_1 - 2x_2 + 3x_3 &= -1 \\ -3x_1 + 9x_2 + x_3 &= 2 \\ 2x_1 - x_2 - 7x_3 &= 3 \end{aligned}$$

Continue the iterations until two successive approximations are identical when rounded to three significant digits.

Solution To begin, write the system in the form

$$\begin{aligned} x_1 &= -\tfrac{1}{5} + \tfrac{2}{5}x_2 - \tfrac{3}{5}x_3 \\ x_2 &= \tfrac{2}{9} + \tfrac{3}{9}x_1 - \tfrac{1}{9}x_3 \\ x_3 &= -\tfrac{3}{7} + \tfrac{2}{7}x_1 - \tfrac{1}{7}x_2 . \end{aligned}$$

Because you do not know the actual solution, choose

$$x_1 = 0, \quad x_2 = 0, \quad x_3 = 0 \qquad \text{Initial approximation}$$

as a convenient initial approximation. So, the first approximation is

$$\begin{aligned} x_1 &= -\tfrac{1}{5} + \tfrac{2}{5}(0) - \tfrac{3}{5}(0) = -0.200 \\ x_2 &= \tfrac{2}{9} + \tfrac{3}{9}(0) - \tfrac{1}{9}(0) \approx 0.222 \\ x_3 &= -\tfrac{3}{7} + \tfrac{2}{7}(0) - \tfrac{1}{7}(0) \approx -0.429. \end{aligned}$$

Continuing this procedure, you obtain the sequence of approximations shown in Table 10.1.

TABLE 10.1

n	0	1	2	3	4	5	6	7
x_1	0.000	−0.200	0.146	0.192	0.181	0.185	0.186	0.186
x_2	0.000	0.222	0.203	0.328	0.332	0.329	0.331	0.331
x_3	0.000	−0.429	−0.517	−0.416	−0.421	−0.424	−0.423	−0.423

Because the last two columns in Table 10.1 are identical, you can conclude that to three significant digits the solution is

$$x_1 = 0.186, \qquad x_2 = 0.331, \qquad x_3 = -0.423.$$

For the system of linear equations given in Example 1, the Jacobi method is said to **converge.** That is, repeated iterations succeed in producing an approximation that is correct to three significant digits. As is generally true for iterative methods, greater accuracy would require more iterations.

The Gauss-Seidel Method

You will now look at a modification of the Jacobi method called the Gauss-Seidel method, named after Carl Friedrich Gauss (1777–1855) and Philipp L. Seidel (1821–1896). This modification is no more difficult to use than the Jacobi method, and it often requires fewer iterations to produce the same degree of accuracy.

With the Jacobi method, the values of x_i obtained in the nth approximation remain unchanged until the entire $(n + 1)$th approximation has been calculated. With the Gauss-Seidel method, on the other hand, you use the new values of each x_i as soon as they are known. That is, once you have determined x_1 from the first equation, its value is then used in the second equation to obtain the new x_2. Similarly, the new x_1 and x_2 are used in the third equation to obtain the new x_3, and so on. This procedure is demonstrated in Example 2.

EXAMPLE 2 ***Applying the Gauss-Seidel Method***

Use the Gauss-Seidel iteration method to approximate the solution to the system of equations given in Example 1.

Solution The first computation is identical to that given in Example 1. That is, using $(x_1, x_2, x_3) = (0, 0, 0)$ as the initial approximation, you obtain the following new value for x_1.

$$x_1 = -\tfrac{1}{5} + \tfrac{2}{5}(0) - \tfrac{3}{5}(0) = -0.200$$

Now that you have a new value for x_1, however, use it to compute a new value for x_2. That is,

$$x_2 = \tfrac{2}{9} + \tfrac{3}{9}(-0.200) - \tfrac{1}{9}(0) \approx 0.156.$$

Similarly, use $x_1 = -0.200$ and $x_2 = 0.156$ to compute a new value for x_3. That is,

$$x_3 = -\tfrac{3}{7} + \tfrac{2}{7}(-0.200) - \tfrac{1}{7}(0.156) \approx -0.508.$$

So the first approximation is $x_1 = -0.200$, $x_2 = 0.156$, and $x_3 = -0.508$. Continued iterations produce the sequence of approximations shown in Table 10.2.

TABLE 10.2

n	0	1	2	3	4	5
x_1	0.000	−0.200	0.167	0.191	0.186	0.186
x_2	0.000	0.156	0.334	0.333	0.331	0.331
x_3	0.000	−0.508	−0.429	−0.422	−0.423	−0.423

Note that after only five iterations of the Gauss-Seidel method, you achieved the same accuracy as was obtained with seven iterations of the Jacobi method in Example 1.

Neither of the iterative methods presented in this section always converges. That is, it is possible to apply the Jacobi method or the Gauss-Seidel method to a system of linear equations and obtain a divergent sequence of approximations. In such cases, it is said that the method **diverges.**

EXAMPLE 3 ***An Example of Divergence***

Apply the Jacobi method to the system

$$\begin{aligned} x_1 - 5x_2 &= -4 \\ 7x_1 - \ x_2 &= \ \ 6, \end{aligned}$$

using the initial approximation $(x_1, x_2) = (0, 0)$, and show that the method diverges.

Solution As usual, begin by rewriting the given system in the form

$$\begin{aligned} x_1 &= -4 + 5x_2 \\ x_2 &= -6 + 7x_1. \end{aligned}$$

Then the initial approximation (0, 0) produces

$$\begin{aligned} x_1 &= -4 + 5(0) = -4 \\ x_2 &= -6 + 7(0) = -6 \end{aligned}$$

as the first approximation. Repeated iterations produce the sequence of approximations shown in Table 10.3.

TABLE 10.3

n	0	1	2	3	4	5	6	7
x_1	0	−4	−34	−174	−1244	−6124	−42,874	−214,374
x_2	0	−6	−34	−244	−1244	−8574	−42,874	−300,124

For this particular system of linear equations you can determine that the actual solution is $x_1 = 1$ and $x_2 = 1$. So you can see from Table 10.3 that the approximations given by the Jacobi method become progressively *worse* instead of better, and you can conclude that the method diverges.

The problem of divergence in Example 3 is not resolved by using the Gauss-Seidel method rather than the Jacobi method. In fact, for this particular system the Gauss-Seidel method diverges more rapidly, as shown in Table 10.4.

TABLE 10.4

n	0	1	2	3	4	5
x_1	0	−4	−174	−6124	−214,374	−7,503,124
x_2	0	−34	−1224	−42,874	−1,500,624	−52,521,874

With an initial approximation of $(x_1, x_2) = (0, 0)$, neither the Jacobi method nor the Gauss-Seidel method converges to the solution of the system of linear equations given in Example 3. You will now look at a special type of coefficient matrix A, called a **strictly diagonally dominant matrix,** for which it is guaranteed that both methods will converge.

Definition of Strictly Diagonally Dominant Matrix

An $n \times n$ matrix A is **strictly diagonally dominant** if the absolute value of each entry on the main diagonal is greater than the sum of the absolute values of the other entries in the same row. That is,

$$\begin{aligned} |a_{11}| &> |a_{12}| + |a_{13}| + \cdots + |a_{1n}| \\ |a_{22}| &> |a_{21}| + |a_{23}| + \cdots + |a_{2n}| \\ &\vdots \\ |a_{nn}| &> |a_{n1}| + |a_{n2}| + \cdots + |a_{n,n-1}|. \end{aligned}$$

EXAMPLE 4 ***Strictly Diagonally Dominant Matrices***

Which of the following systems of linear equations has a strictly diagonally dominant coefficient matrix?

(a)
$$\begin{aligned} 3x_1 - x_2 &= -4 \\ 2x_1 + 5x_2 &= 2 \end{aligned}$$

(b)
$$\begin{aligned} 4x_1 + 2x_2 - x_3 &= -1 \\ x_1 \qquad + 2x_3 &= -4 \\ 3x_1 - 5x_2 + x_3 &= 3 \end{aligned}$$

Solution (a) The coefficient matrix

$$A = \begin{bmatrix} 3 & -1 \\ 2 & 5 \end{bmatrix}$$

is strictly diagonally dominant because $|3| > |-1|$ and $|5| > |2|$.

(b) The coefficient matrix

$$A = \begin{bmatrix} 4 & 2 & -1 \\ 1 & 0 & 2 \\ 3 & -5 & 1 \end{bmatrix}$$

is not strictly diagonally dominant because the entries in the second and third rows do not conform to the definition. For instance, in the second row $a_{21} = 1$, $a_{22} = 0$, $a_{23} = 2$, and it is not true that $|a_{22}| > |a_{21}| + |a_{23}|$. Interchanging the second and third rows in the original system of linear equations, however, produces the coefficient matrix

$$A' = \begin{bmatrix} 4 & 2 & -1 \\ 3 & -5 & 1 \\ 1 & 0 & 2 \end{bmatrix},$$

and this matrix is strictly diagonally dominant.

The following theorem, which is listed without proof, states that strict diagonal dominance is sufficient for the convergence of either the Jacobi method or the Gauss-Seidel method.

Theorem 10.1
Convergence of the Jacobi and Gauss-Seidel Methods

If A is strictly diagonally dominant, then the system of linear equations given by $A\mathbf{x} = \mathbf{b}$ has a unique solution to which the Jacobi method and the Gauss-Seidel method will converge for any initial approximation.

In Example 3 you looked at a system of linear equations for which the Jacobi and Gauss-Seidel methods diverged. In the following example you can see that by interchanging the rows of the system given in Example 3, you can obtain a coefficient matrix that is strictly diagonally dominant. After this interchange, convergence is assured.

EXAMPLE 5 ***Interchanging Rows to Obtain Convergence***

Interchange the rows of the system

$$\begin{aligned} x_1 - 5x_2 &= -4 \\ 7x_1 - x_2 &= 6 \end{aligned}$$

to obtain one with a strictly diagonally dominant coefficient matrix. Then apply the Gauss-Seidel method to approximate the solution to four significant digits.

Solution Begin by interchanging the two rows of the given system to obtain

$$\begin{aligned} 7x_1 - x_2 &= 6 \\ x_1 - 5x_2 &= -4. \end{aligned}$$

Note that the coefficient matrix of this system is strictly diagonally dominant. Then solve for x_1 and x_2 as follows.

$$\begin{aligned} x_1 &= \tfrac{6}{7} + \tfrac{1}{7}x_2 \\ x_2 &= \tfrac{4}{5} + \tfrac{1}{5}x_1 \end{aligned}$$

Using the initial approximation $(x_1, x_2) = (0, 0)$, you can obtain the sequence of approximations shown in Table 10.5.

TABLE 10.5

n	0	1	2	3	4	5
x_1	0.0000	0.8571	0.9959	0.9999	1.000	1.000
x_2	0.0000	0.9714	0.9992	1.000	1.000	1.000

So you can conclude that the solution is $x_1 = 1$ and $x_2 = 1$.

Do not conclude from Theorem 10.1 that strict diagonal dominance is a necessary condition for convergence of the Jacobi or Gauss-Seidel methods. For instance, the coefficient matrix of the system

$$\begin{aligned} -4x_1 + 5x_2 &= 1 \\ x_1 + 2x_2 &= 3 \end{aligned}$$

is not a strictly diagonally dominant matrix, and yet both methods converge to the solution $x_1 = 1$ and $x_2 = 1$ when you use an initial approximation of $(x_1, x_2) = (0, 0)$. (See Exercises 21–22.)

SECTION 10.2 ❑ EXERCISES

In Exercises 1–4, apply the Jacobi method to the given system of linear equations, using the initial approximation $(x_1, x_2, \ldots, x_n) = (0, 0, \ldots, 0)$. Continue performing iterations until two successive approximations are identical when rounded to three significant digits.

1. $$\begin{aligned} 3x_1 - x_2 &= 2 \\ x_1 + 4x_2 &= 5 \end{aligned}$$

2. $$\begin{aligned} -4x_1 + 2x_2 &= -6 \\ 3x_1 - 5x_2 &= 1 \end{aligned}$$

3. $$\begin{aligned} 2x_1 - x_2 \phantom{{}- 3x_3} &= 2 \\ x_1 - 3x_2 + x_3 &= -2 \\ -x_1 + x_2 - 3x_3 &= -6 \end{aligned}$$

4. $$\begin{aligned} 4x_1 + x_2 + x_3 &= 7 \\ x_1 - 7x_2 + 2x_3 &= -2 \\ 3x_1 \phantom{{}+ x_2} + 4x_3 &= 11 \end{aligned}$$

5. Apply the Gauss-Seidel method to Exercise 1.

6. Apply the Gauss-Seidel method to Exercise 2.

7. Apply the Gauss-Seidel method to Exercise 3.

8. Apply the Gauss-Seidel method to Exercise 4.

In Exercises 9–12, show that the Gauss-Seidel method diverges for the given system using the initial approximation $(x_1, x_2, \ldots, x_n) = (0, 0, \ldots, 0)$.

9. $$\begin{aligned} x_1 - 2x_2 &= -1 \\ 2x_1 + x_2 &= 3 \end{aligned}$$

10. $$\begin{aligned} -x_1 + 4x_2 &= 1 \\ 3x_1 - 2x_2 &= 2 \end{aligned}$$

11. $$\begin{aligned} 2x_1 - 3x_2 \phantom{{}- 10x_3} &= -7 \\ x_1 + 3x_2 - 10x_3 &= 9 \\ 3x_1 \phantom{{}+ 3x_2} + x_3 &= 13 \end{aligned}$$

12. $$\begin{aligned} x_1 + 3x_2 - x_3 &= 5 \\ 3x_1 - x_2 \phantom{{}- x_3} &= 5 \\ x_2 + 2x_3 &= 1 \end{aligned}$$

In Exercises 13–16, determine whether the matrix is strictly diagonally dominant.

13. $$\begin{bmatrix} 2 & 1 \\ 3 & 5 \end{bmatrix}$$

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

15. $$\begin{bmatrix} 12 & 6 & 0 \\ 2 & -3 & 2 \\ 0 & 6 & 13 \end{bmatrix}$$

16. $$\begin{bmatrix} 7 & 5 & -1 \\ 1 & -4 & 1 \\ 0 & 2 & -3 \end{bmatrix}$$

17. Interchange the rows of the system of linear equations in Exercise 9 to obtain a system with a strictly diagonally dominant coefficient matrix. Then apply the Gauss-Seidel method to approximate the solution to two significant digits.

18. Interchange the rows of the system of linear equations in Exercise 10 to obtain a system with a strictly diagonally dominant coefficient matrix. Then apply the Gauss-Seidel method to approximate the solution to two significant digits.

19. Interchange the rows of the system of linear equations in Exercise 11 to obtain a system with a strictly diagonally dominant coefficient matrix. Then apply the Gauss-Seidel method to approximate the solution to two significant digits.

20. Interchange the rows of the system of linear equations in Exercise 12 to obtain a system with a strictly diagonally dominant coefficient matrix. Then apply the Gauss-Seidel method to approximate the solution to two significant digits.

In Exercises 21 and 22, the coefficient matrix of the system of linear equations is not strictly diagonally dominant. Show that the Jacobi and Gauss-Seidel methods converge using an initial approximation of $(x_1, x_2, \ldots, x_n) = (0, 0, \ldots, 0)$.

21. $$\begin{aligned} -4x_1 + 5x_2 &= 1 \\ x_1 + 2x_2 &= 3 \end{aligned}$$

22. $$\begin{aligned} 4x_1 + 2x_2 - 2x_3 &= 0 \\ x_1 - 3x_2 - x_3 &= 7 \\ 3x_1 - x_2 + 4x_3 &= 5 \end{aligned}$$

In Exercises 23 and 24, write a computer program that applies the Gauss-Siedel method to solve the system of linear equations.

23. $$\begin{aligned} 4x_1 + x_2 - x_3 &= 3 \\ x_1 + 6x_2 - 2x_3 + x_4 - x_5 &= -6 \\ x_2 + 5x_3 - x_5 + x_6 &= -5 \\ 2x_2 + 5x_4 - x_5 - x_7 - x_8 &= 0 \\ -x_3 - x_4 + 6x_5 - x_6 - x_8 &= 12 \\ -x_3 - x_5 + 5x_6 &= -12 \\ -x_4 + 4x_7 - x_8 &= -2 \\ -x_4 - x_5 - x_7 + 5x_8 &= 2 \end{aligned}$$

24. $$\begin{aligned} 4x_1 - x_2 - x_3 &= 18 \\ -x_1 + 4x_2 - x_3 - x_4 &= 18 \\ -x_2 + 4x_3 - x_4 - x_5 &= 4 \\ -x_3 + 4x_4 - x_5 - x_6 &= 4 \\ -x_4 + 4x_5 - x_6 - x_7 &= 26 \\ -x_5 + 4x_6 - x_7 - x_8 &= 16 \\ -x_6 + 4x_7 - x_8 &= 10 \\ -x_7 + 4x_8 &= 32 \end{aligned}$$

10.3 POWER METHOD FOR APPROXIMATING EIGENVALUES

In Chapter 7 you saw that the eigenvalues of an $n \times n$ matrix A are obtained by solving its characteristic equation

$$\lambda^n + c_{n-1}\lambda^{n-1} + c_{n-2}\lambda^{n-2} + \cdots + c_0 = 0.$$

For large values of n, polynomial equations like this one are difficult and time-consuming to solve. Moreover, numerical techniques for approximating roots of polynomial equations of high degree are sensitive to rounding errors. In this section you will look at an alternative method for approximating eigenvalues. As presented here, the method can be used only to find the eigenvalue of A that is largest in absolute value—this eigenvalue is called the **dominant eigenvalue** of A. Although this restriction may seem severe, dominant eigenvalues are of primary interest in many physical applications.

Definition of Dominant Eigenvalue and Dominant Eigenvector

Let $\lambda_1, \lambda_2, \ldots,$ and λ_n be the eigenvalues of an $n \times n$ matrix A. λ_1 is called the **dominant eigenvalue** of A if

$$|\lambda_1| > |\lambda_i|, \quad i = 2, \ldots, n.$$

The eigenvectors corresponding to λ_1 are called **dominant eigenvectors** of A.

Not every matrix has a dominant eigenvalue. For instance, the matrix

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

(with eigenvalues of $\lambda_1 = 1$ and $\lambda_2 = -1$) has no dominant eigenvalue. Similarly, the matrix

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(with eigenvalues of $\lambda_1 = 2, \lambda_2 = 2,$ and $\lambda_3 = 1$) has no dominant eigenvalue.

EXAMPLE 1 ***Finding a Dominant Eigenvalue***

Find the dominant eigenvalue and corresponding eigenvectors of the matrix

$$A = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}.$$

Solution From Example 4 of Section 7.1 you know that the characteristic polynomial of A is $\lambda^2 + 3\lambda + 2 = (\lambda + 1)(\lambda + 2)$. So the eigenvalues of A are $\lambda_1 = -1$ and $\lambda_2 = -2$, of which the dominant one is $\lambda_2 = -2$. From the same example you know that the dominant eigenvectors of A (those corresponding to $\lambda_2 = -2$) are of the form

$$\mathbf{x} = t\begin{bmatrix} 3 \\ 1 \end{bmatrix}, \quad t \neq 0.$$

The Power Method

Like the Jacobi and Gauss-Seidel methods, the power method for approximating eigenvalues is iterative. First assume that the matrix A has a dominant eigenvalue with corresponding dominant eigenvectors. Then choose an initial approximation $\mathbf{x}_0$ of one of the dominant eigen*vectors* of A. This initial approximation must be a *nonzero* vector in R^n. Finally, form the sequence given by

$$\begin{aligned}
\mathbf{x}_1 &= A\mathbf{x}_0 \\
\mathbf{x}_2 &= A\mathbf{x}_1 = A(A\mathbf{x}_0) = A^2\mathbf{x}_0 \\
\mathbf{x}_3 &= A\mathbf{x}_2 = A(A^2\mathbf{x}_0) = A^3\mathbf{x}_0 \\
&\vdots \\
\mathbf{x}_k &= A\mathbf{x}_{k-1} = A(A^{k-1}\mathbf{x}_0) = A^k\mathbf{x}_0.
\end{aligned}$$

For large powers of k, and by properly scaling this sequence, you will see that you obtain a good approximation of the dominant eigenvector of A. This procedure is illustrated in Example 2.

EXAMPLE 2 ***Approximating a Dominant Eigenvector by the Power Method***

Complete six iterations of the power method to approximate a dominant eigenvector of

$$A = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}.$$

Solution Begin with an initial nonzero approximation of

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Then obtain the following approximations.

Iteration — *"Scaled" Approximation*

$$\begin{aligned}
\mathbf{x}_1 &= A\mathbf{x}_0 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -10 \\ -4 \end{bmatrix} &\rightarrow\ & -4\begin{bmatrix} 2.50 \\ 1.00 \end{bmatrix} \\
\mathbf{x}_2 &= A\mathbf{x}_1 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} -10 \\ -4 \end{bmatrix} = \begin{bmatrix} 28 \\ 10 \end{bmatrix} &\rightarrow\ & 10\begin{bmatrix} 2.80 \\ 1.00 \end{bmatrix} \\
\mathbf{x}_3 &= A\mathbf{x}_2 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 28 \\ 10 \end{bmatrix} = \begin{bmatrix} -64 \\ -22 \end{bmatrix} &\rightarrow\ & -22\begin{bmatrix} 2.91 \\ 1.00 \end{bmatrix} \\
\mathbf{x}_4 &= A\mathbf{x}_3 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} -64 \\ -22 \end{bmatrix} = \begin{bmatrix} 136 \\ 46 \end{bmatrix} &\rightarrow\ & 46\begin{bmatrix} 2.96 \\ 1.00 \end{bmatrix} \\
\mathbf{x}_5 &= A\mathbf{x}_4 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 136 \\ 46 \end{bmatrix} = \begin{bmatrix} -280 \\ -94 \end{bmatrix} &\rightarrow\ & -94\begin{bmatrix} 2.98 \\ 1.00 \end{bmatrix} \\
\mathbf{x}_6 &= A\mathbf{x}_5 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} -280 \\ -94 \end{bmatrix} = \begin{bmatrix} 568 \\ 190 \end{bmatrix} &\rightarrow\ & 190\begin{bmatrix} 2.99 \\ 1.00 \end{bmatrix}
\end{aligned}$$

Note that the approximations in Example 2 appear to be approaching scalar multiples of

$$\begin{bmatrix} 3 \\ 1 \end{bmatrix},$$

which you know from Example 1 is a dominant eigenvector of the matrix

$$A = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}.$$

In Example 2 the power method was used to approximate a dominant eigenvector of the matrix A. In that example you already knew that the dominant eigenvalue of A was $\lambda = -2$. For the sake of demonstration, however, assume that you do not know the dominant eigenvalue of A. The following theorem provides a formula for determining the eigenvalue corresponding to a given eigenvector. This theorem is credited to the English physicist John William Rayleigh (1842–1919).

Theorem 10.2

Determining an Eigenvalue from an Eigenvector

If $\mathbf{x}$ is an eigenvector of a matrix A, then its corresponding eigenvalue is given by

$$\lambda = \frac{A\mathbf{x} \cdot \mathbf{x}}{\mathbf{x} \cdot \mathbf{x}}.$$

This quotient is called the **Rayleigh quotient.**

Proof Because $\mathbf{x}$ is an eigenvector of A, you know that $A\mathbf{x} = \lambda\mathbf{x}$ and can write

$$\frac{A\mathbf{x} \cdot \mathbf{x}}{\mathbf{x} \cdot \mathbf{x}} = \frac{\lambda\mathbf{x} \cdot \mathbf{x}}{\mathbf{x} \cdot \mathbf{x}} = \frac{\lambda(\mathbf{x} \cdot \mathbf{x})}{\mathbf{x} \cdot \mathbf{x}} = \lambda.$$

In cases for which the power method generates a good approximation of a dominant eigenvector, the Rayleigh quotient provides a correspondingly good approximation of the dominant eigenvalue. The use of the Rayleigh quotient is demonstrated in Example 3.

EXAMPLE 3 ***Approximating a Dominant Eigenvalue***

Use the result of Example 2 to approximate the dominant eigenvalue of the matrix

$$A = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}.$$

Solution After the sixth iteration of the power method in Example 2, obtained

$$\mathbf{x}_6 = \begin{bmatrix} 568 \\ 190 \end{bmatrix} \approx 190 \begin{bmatrix} 2.99 \\ 1.00 \end{bmatrix}$$

With $\mathbf{x} = (2.99, 1)$ as the approximation of a dominant eigenvector of A, use the Rayleigh quotient to obtain an approximation of the dominant eigenvalue of A. First compute the product $A\mathbf{x}$.

$$A\mathbf{x} = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 2.99 \\ 1.00 \end{bmatrix} = \begin{bmatrix} -6.02 \\ -2.01 \end{bmatrix}$$

Then, because

$$A\mathbf{x} \cdot \mathbf{x} = (-6.02)(2.99) + (-2.01)(1) \approx -20.0$$

and

$$\mathbf{x} \cdot \mathbf{x} = (2.99)(2.99) + (1)(1) \approx 9.94,$$

you can compute the Rayleigh quotient to be

$$\lambda = \frac{A\mathbf{x} \cdot \mathbf{x}}{\mathbf{x} \cdot \mathbf{x}} \approx \frac{-20.0}{9.94} \approx -2.01$$

which is a good approximation of the dominant eigenvalue $\lambda = -2$.

From Example 2 you can see that the power method tends to produce approximations with large entries. In practice it is best to "scale down" each approximation before proceeding to the next iteration. One way to accomplish this **scaling** is to determine the component of $A\mathbf{x}_i$ that has the largest absolute value and multiply the vector $A\mathbf{x}_i$ by the reciprocal of this component. The resulting vector will then have components whose absolute values are less than or equal to 1. (Other scaling techniques are possible. For examples, see Exercises 27 and 28.)

EXAMPLE 4 ***The Power Method with Scaling***

Calculate seven iterations of the power method with *scaling* to approximate a dominant eigenvector of the matrix

$$A = \begin{bmatrix} 1 & 2 & 0 \\ -2 & 1 & 2 \\ 1 & 3 & 1 \end{bmatrix}.$$

Use $\mathbf{x}_0 = (1, 1, 1)$ as the initial approximation.

Solution One iteration of the power method produces

$$A\mathbf{x}_0 = \begin{bmatrix} 1 & 2 & 0 \\ -2 & 1 & 2 \\ 1 & 3 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix}$$

and by scaling you obtain the approximation

$$\mathbf{x}_1 = \frac{1}{5}\begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix} = \begin{bmatrix} 0.60 \\ 0.20 \\ 1.00 \end{bmatrix}.$$

A second iteration yields

$$A\mathbf{x}_1 = \begin{bmatrix} 1 & 2 & 0 \\ -2 & 1 & 2 \\ 1 & 3 & 1 \end{bmatrix} \begin{bmatrix} 0.60 \\ 0.20 \\ 1.00 \end{bmatrix} = \begin{bmatrix} 1.00 \\ 1.00 \\ 2.20 \end{bmatrix}$$

and

$$\mathbf{x}_2 = \frac{1}{2.20}\begin{bmatrix} 1.00 \\ 1.00 \\ 2.20 \end{bmatrix} = \begin{bmatrix} 0.45 \\ 0.45 \\ 1.00 \end{bmatrix}.$$

Continuing this process, you obtain the sequence of approximations shown in Table 10.6.

TABLE 10.6

$\mathbf{x}_0$	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$	$\mathbf{x}_6$	$\mathbf{x}_7$
1.00 1.00	0.60 1.00	0.45 1.00	0.48 1.00	0.51 1.00	0.49 1.00	0.50 1.00	0.50 0.50 1.00

From Table 10.6 you can approximate a dominant eigenvector of A to be

$$\mathbf{x} = \begin{bmatrix} 0.50 \\ 0.50 \\ 1.00 \end{bmatrix}.$$

Using the Rayleigh quotient, you can approximate the dominant eigenvalue of A to be $\lambda = 3$. (For this example you can check that the approximations of $\mathbf{x}$ and λ are exact.)

REMARK: Note that the *scaling factors* used to obtain the vectors in Table 10.6,

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$	$\mathbf{x}_6$	$\mathbf{x}_7$
↓	↓	↓	↓	↓	↓	↓
5.00	2.20	2.82	3.13	3.02	2.99	3.00,

are approaching the dominant eigenvalue $\lambda = 3$.

In Example 4 the power method with scaling converges to a dominant eigenvector. The following theorem states that a sufficient condition for convergence of the power method is that the matrix A be diagonalizable (and have a dominant eigenvalue).

Theorem 10.3 Convergence of the Power Method

If A is an $n \times n$ diagonalizable matrix with a dominant eigenvalue, then there exists a nonzero vector $\mathbf{x}_0$ such that the sequence of vectors given by

$$A\mathbf{x}_0,\quad A^2\mathbf{x}_0,\quad A^3\mathbf{x}_0,\quad A^4\mathbf{x}_0,\quad \ldots,\quad A^k\mathbf{x}_0,\quad \ldots$$

approaches a multiple of the dominant eigenvector of A.

Proof Because A is diagonalizable, you know from Theorem 7.5 that it has n linearly independent eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ with corresponding eigenvalues of $\lambda_1, \lambda_2, \ldots, \lambda_n$. Assume that these eigenvalues are ordered so that λ_1 is the dominant eigenvalue (with a corresponding eigenvector of $\mathbf{x}_1$). Because the n eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are linearly independent, they must form a basis for R^n. For the initial approximation $\mathbf{x}_0$, choose a nonzero vector such that the linear combination

$$\mathbf{x}_0 = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_n\mathbf{x}_n$$

has nonzero leading coefficients. (If $c_1 = 0$, the power method may not converge, and a different $\mathbf{x}_0$ must be used as the initial approximation. See Exercises 21 and 22.) Now, multiplying both sides of this equation by A produces

$$\begin{aligned} A\mathbf{x}_0 &= A(c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_n\mathbf{x}_n) \\ &= c_1(A\mathbf{x}_1) + c_2(A\mathbf{x}_2) + \cdots + c_n(A\mathbf{x}_n) \\ &= c_1(\lambda_1\mathbf{x}_1) + c_2(\lambda_2\mathbf{x}_2) + \cdots + c_n(\lambda_n\mathbf{x}_n). \end{aligned}$$

Repeated multiplication of both sides of this equation by A produces

$$A^k\mathbf{x}_0 = c_1(\lambda_1^k\mathbf{x}_1) + c_2(\lambda_2^k\mathbf{x}_2) + \cdots + c_n(\lambda_n^k\mathbf{x}_n),$$

which implies that

$$A^k\mathbf{x}_0 = \lambda_1^k\left[c_1\mathbf{x}_1 + c_2\left(\frac{\lambda_2}{\lambda_1}\right)^k\mathbf{x}_2 + \cdots + c_n\left(\frac{\lambda_n}{\lambda_1}\right)^k\mathbf{x}_n\right].$$

Now, from the original assumption that λ_1 is larger in absolute value than the other eigenvalues it follows that each of the fractions

$$\frac{\lambda_2}{\lambda_1}, \quad \frac{\lambda_3}{\lambda_1}, \quad \ldots, \quad \frac{\lambda_n}{\lambda_1}$$

is less than 1 in absolute value. So each of the factors

$$\left(\frac{\lambda_2}{\lambda_1}\right)^k, \quad \left(\frac{\lambda_3}{\lambda_1}\right)^k, \quad \ldots, \quad \left(\frac{\lambda_n}{\lambda_1}\right)^k$$

must approach 0 as k approaches infinity. This implies that the approximation

$$A^k\mathbf{x}_0 \approx \lambda_1^k c_1\mathbf{x}_1, \quad c_1 \neq 0$$

improves as k increases. Because $\mathbf{x}_1$ is a dominant eigenvector, it follows that any scalar multiple of $\mathbf{x}_1$ is also a dominant eigenvector, so showing that $A^k\mathbf{x}_0$ approaches a multiple of the dominant eigenvector of A.

The proof of Theorem 10.3 provides some insight into the rate of convergence of the power method. That is, if the eigenvalues of A are ordered so that

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_n|,$$

then the power method will converge quickly if $|\lambda_2|/|\lambda_1|$ is small, and slowly if $|\lambda_2|/|\lambda_1|$ is close to 1. This principle is illustrated in Example 5.

EXAMPLE 5 ***The Rate of Convergence of the Power Method***

(a) The matrix

$$A = \begin{bmatrix} 4 & 5 \\ 6 & 5 \end{bmatrix}$$

has eigenvalues of $\lambda_1 = 10$ and $\lambda_2 = -1$. So the ratio $|\lambda_2|/|\lambda_1|$ is 0.1. For this matrix, only four iterations are required to obtain successive approximations that agree when rounded to three significant digits. (See Table 10.7.)

TABLE 10.7

$\mathbf{x}_0$	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
$\begin{bmatrix} 1.000 \\ 1.000 \end{bmatrix}$	$\begin{bmatrix} 0.818 \\ 1.000 \end{bmatrix}$	$\begin{bmatrix} 0.835 \\ 1.000 \end{bmatrix}$	$\begin{bmatrix} 0.833 \\ 1.000 \end{bmatrix}$	$\begin{bmatrix} 0.833 \\ 1.000 \end{bmatrix}$

(b) The matrix

$$A = \begin{bmatrix} -4 & 10 \\ 7 & 5 \end{bmatrix}$$

has eigenvalues of $\lambda_1 = 10$ and $\lambda_2 = -9$. For this matrix, the ratio $|\lambda_2|/|\lambda_1|$ is 0.9, and the power method does not produce successive approximations that agree to three significant digits until sixty-eight iterations have been performed, as shown in Table 10.8.

TABLE 10.8

$\mathbf{x}_0$	$\mathbf{x}_1$	$\mathbf{x}_2$		$\mathbf{x}_{66}$	$\mathbf{x}_{67}$	$\mathbf{x}_{68}$
$\begin{bmatrix} 1.000 \\ 1.000 \end{bmatrix}$	$\begin{bmatrix} 0.500 \\ 1.000 \end{bmatrix}$	$\begin{bmatrix} 0.941 \\ 1.000 \end{bmatrix}$	. . .	$\begin{bmatrix} 0.715 \\ 1.000 \end{bmatrix}$	$\begin{bmatrix} 0.714 \\ 1.000 \end{bmatrix}$	$\begin{bmatrix} 0.714 \\ 1.000 \end{bmatrix}$

In this section you have seen the use of the power method to approximate the *dominant* eigenvalue of a matrix. This method can be modified to approximate other eigenvalues through use of a procedure called **deflation.** Moreover, the power method is only one of several techniques that can be used to approximate the eigenvalues of a matrix. Another popular method is called the **QR algorithm.**

This is the method used in most computer programs and calculators for finding eigenvalues and eigenvectors. The algorithm uses the *QR*–factorization of the matrix, as presented in Chapter 5. Discussions of the deflation method and the *QR* algorithm can be found in most texts on numerical methods.

SECTION 10.3 ❑ EXERCISES

In Exercises 1–6, use the techniques presented in Chapter 7 to find the eigenvalues of the matrix A. If A has a dominant eigenvalue, find a corresponding dominant eigenvector.

1. $A = \begin{bmatrix} 2 & 1 \\ 0 & -4 \end{bmatrix}$ **2.** $A = \begin{bmatrix} -3 & 0 \\ 1 & 3 \end{bmatrix}$

3. $A = \begin{bmatrix} 1 & -5 \\ -3 & -1 \end{bmatrix}$ **4.** $A = \begin{bmatrix} 4 & -5 \\ 2 & -3 \end{bmatrix}$

5. $A = \begin{bmatrix} 2 & 3 & 1 \\ 0 & -1 & 2 \\ 0 & 0 & 3 \end{bmatrix}$ **6.** $A = \begin{bmatrix} -5 & 0 & 0 \\ 3 & 7 & 0 \\ 4 & -2 & 3 \end{bmatrix}$

In Exercises 7–10, use the Rayleigh quotient to compute the eigenvalue λ of A corresponding to the eigenvector $\mathbf{x}$.

7. $A = \begin{bmatrix} 4 & -5 \\ 2 & -3 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$ **8.** $A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$

9. $A = \begin{bmatrix} 1 & 2 & -2 \\ -2 & 5 & -2 \\ -6 & 6 & -3 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$

10. $A = \begin{bmatrix} 3 & 2 & -3 \\ -3 & -4 & 9 \\ -1 & -2 & 5 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}$

In Exercises 11–14, use the power method with scaling to approximate a dominant eigenvector of the matrix A. Start with $\mathbf{x}_0 = (1, 1)$ and calculate five iterations. Then use $\mathbf{x}_5$ to approximate the dominant eigenvalue of A.

11. $A = \begin{bmatrix} 2 & 1 \\ 0 & -7 \end{bmatrix}$ **12.** $A = \begin{bmatrix} -1 & 0 \\ 1 & 6 \end{bmatrix}$

13. $A = \begin{bmatrix} 1 & -4 \\ -2 & 8 \end{bmatrix}$ **14.** $A = \begin{bmatrix} 6 & -3 \\ -2 & 1 \end{bmatrix}$

In Exercises 15–18, use the power method with scaling to approximate a dominant eigenvector of the matrix A. Start with $\mathbf{x}_0 = (1, 1, 1)$ and calculate four iterations. Then use $\mathbf{x}_4$ to approximate the dominant eigenvalue of A.

15. $A = \begin{bmatrix} 3 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 2 & 8 \end{bmatrix}$ **16.** $A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & -7 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

17. $A = \begin{bmatrix} -1 & -6 & 0 \\ 2 & 7 & 0 \\ 1 & 2 & -1 \end{bmatrix}$ **18.** $A = \begin{bmatrix} 0 & 6 & 0 \\ 0 & -4 & 0 \\ 2 & 1 & 1 \end{bmatrix}$

In Exercises 19 and 20, the matrix A does not have a dominant eigenvalue. Apply the power method with scaling, starting with $\mathbf{x}_0 = (1, 1, 1)$, and observe the results of the first four iterations.

19. $A = \begin{bmatrix} 1 & 1 & 0 \\ 3 & -1 & 0 \\ 0 & 0 & -2 \end{bmatrix}$ **20.** $A = \begin{bmatrix} 1 & 2 & -2 \\ -2 & 5 & -2 \\ -6 & 6 & -3 \end{bmatrix}$

21. *Writing* (a) Find the eigenvalues and corresponding eigenvectors of

$$A = \begin{bmatrix} 3 & -1 \\ -2 & 4 \end{bmatrix}.$$

(b) Calculate two iterations of the power method with scaling, starting with $\mathbf{x}_0 = (1, 1)$.

(c) Explain why the method does not seem to converge to a dominant eigenvector.

22. *Writing* Repeat Exercise 21 using $\mathbf{x}_0 = (1, 1, 1)$, for the matrix

$$A = \begin{bmatrix} -3 & 0 & 2 \\ 0 & -1 & 0 \\ 0 & 1 & -2 \end{bmatrix}.$$

23. The matrix

$$A = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}$$

has a dominant eigenvalue of $\lambda = -2$. Observe that $A\mathbf{x} = \lambda\mathbf{x}$ implies that

$$A^{-1}\mathbf{x} = \frac{1}{\lambda}\mathbf{x}.$$

Apply five iterations of the power method (with scaling) on A^{-1} to compute the eigenvalue of A with the smallest magnitude.

24. Repeat Exercise 23 for the matrix

$$A = \begin{bmatrix} 2 & 3 & 1 \\ 0 & -1 & 2 \\ 0 & 0 & 3 \end{bmatrix}.$$

25. (a) Compute the eigenvalues of

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}.$$

(b) Apply four iterations of the power method with scaling to each matrix in part (a), starting with $\mathbf{x}_0 = (-1, 2)$.

(c) Compute the ratios λ_2/λ_1 for A and B. For which do you expect faster convergence?

26. Use the proof of Theorem 10.3 to show that

$$A(A^k\mathbf{x}_0) \approx \lambda_1(A^k\mathbf{x}_0)$$

for large values of k. That is, show that the scale factors obtained in the power method approach the dominant eigenvalue.

In Exercises 27 and 28, apply four iterations of the power method (with scaling) to approximate the dominant eigenvalue of the matrix. After each iteration, scale the approximation by dividing by its length so that the resulting approximation will be a unit vector.

27. $A = \begin{bmatrix} 5 & 6 \\ 4 & 3 \end{bmatrix}$

28. $A = \begin{bmatrix} 7 & -4 & 2 \\ 16 & -9 & 6 \\ 8 & -4 & 5 \end{bmatrix}$

10.4 APPLICATIONS OF NUMERICAL METHODS

Applications of Gaussian Elimination with Pivoting

In Section 2.5 you used least squares regression analysis to find *linear* mathematical models that best fit a set of n points in the plane. This procedure can be extended to cover polynomial models of any degree as follows.

Regression Analysis for Polynomials

The least squares regression polynomial of degree m for the points $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ is given by

$$y = a_m x^m + a_{m-1}x^{m-1} + \cdots + a_2x^2 + a_1x + a_0,$$

where the coefficients are determined by the following system of $m + 1$ linear equations.

$$\begin{aligned}
na_0 + (\Sigma x_i)a_1 + (\Sigma x_i^2)a_2 + \cdots + (\Sigma x_i^m)a_m &= \Sigma y_i \\
(\Sigma x_i)a_0 + (\Sigma x_i^2)a_1 + (\Sigma x_i^3)a_2 + \cdots + (\Sigma x_i^{m+1})a_m &= \Sigma x_i y_i \\
(\Sigma x_i^2)a_0 + (\Sigma x_i^3)a_1 + (\Sigma x_i^4)a_2 + \cdots + (\Sigma x_i^{m+2})a_m &= \Sigma x_i^2 y_i \\
&\ \vdots \\
(\Sigma x_i^m)a_0 + (\Sigma x_i^{m+1})a_1 + (\Sigma x_i^{m+2})a_2 + \cdots + (\Sigma x_i^{2m})a_m &= \Sigma x_i^m y_i
\end{aligned}$$

Note that if $m = 1$ this system of equations reduces to

$$\begin{aligned}
na_0 + (\Sigma x_i)a_1 &= \Sigma y_i \\
(\Sigma x_i)a_0 + (\Sigma x_i^2)a_1 &= \Sigma x_i y_i,
\end{aligned}$$

which has a solution of

$$a_1 = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^2 - (\sum x_i)^2} \quad \text{and} \quad a_0 = \frac{\sum y_i}{n} - a_1\frac{\sum x_i}{n}.$$

Exercise 16 asks you to show that this formula is equivalent to the matrix formula for linear regression that was presented in Section 2.5.

Example 1 illustrates the use of regression analysis to find a second-degree polynomial model.

EXAMPLE 1 ***Least Squares Regression Analysis***

The world population in billions for the years between 1965 and 2000, is shown in Table 10.9. (Source: U.S. Census Bureau)

TABLE 10.9

Year	1965	1970	1975	1980	1985	1990	1995	2000
Population	3.36	3.72	4.10	4.46	4.86	5.28	5.69	6.08

Find the second-degree least squares regression polynomial for these data and use the resulting model to predict the world population for 2005 and 2010.

Solution Begin by letting $x = -4$ represent 1965, $x = -3$ represent 1970, and so on. So the collection of points is given by $\{(-4, 3.36), (-3, 3.72), (-2, 4.10), (-1, 4.46), (0, 4.86), (1, 5.28), (2, 5.69), (3, 6.08)\}$, which yields

$$n = 8, \qquad \sum_{i=1}^{8} x_i = -4, \qquad \sum_{i=1}^{8} x_i^2 = 44, \qquad \sum_{i=1}^{8} x_i^3 = -64,$$

$$\sum_{i=1}^{8} x_i^4 = 452, \qquad \sum_{i=1}^{8} y_i = 37.55, \qquad \sum_{i=1}^{8} x_i y_i = -2.36, \qquad \sum_{i=1}^{8} x_i^2 y_i = 190.86.$$

So the system of linear equations giving the coefficients of the quadratic model $y = a_2x^2 + a_1x + a_0$ is

$$\begin{aligned} 8a_0 - 4a_1 + 44a_2 &= 37.55 \\ -4a_0 + 44a_1 - 64a_2 &= -2.36 \\ 44a_0 - 64a_1 + 452a_2 &= 190.86. \end{aligned}$$

Gaussian elimination with pivoting on the matrix

$$\begin{bmatrix} 8 & -4 & 44 & 37.55 \\ -4 & 44 & -64 & -2.36 \\ 44 & -64 & 452 & 190.86 \end{bmatrix}$$

produces

$$\begin{bmatrix} 1 & -1.4545 & 10.2727 & 4.3377 \\ 0 & 1 & -0.6000 & 0.3926 \\ 0 & 0 & 1 & 0.0045 \end{bmatrix}.$$

So by back substitution you find the solution to be

$$a_2 = 0.0045, \quad a_1 = 0.3953, \quad a_0 = 4.8667,$$

and the regression quadratic is

$$y = 0.0045x^2 + 0.3953x + 4.8667.$$

Figure 10.1

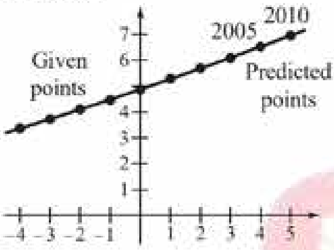


Figure 10.1 compares this model with the given points. To predict the world population for 2005, let $x = 4$, obtaining

$$y = 0.0045(4^2) + 0.3953(4) + 4.8667 = 6.52 \text{ billion.}$$

Similarly, the prediction for 2010 ($x = 5$) is

$$y = 0.0045(5^2) + 0.3953(5) + 4.8667 = 6.96 \text{ billion.}$$

EXAMPLE 2 ***Least Squares Regression Analysis***

Find the third-degree least squares regression polynomial

$$y = a_3x^3 + a_2x^2 + a_1x + a_0$$

for the points

$$\{(0, 0), (1, 2), (2, 3), (3, 2), (4, 1), (5, 2), (6, 4)\}.$$

Solution For this set of points the linear system

$$\begin{aligned} na_0 + (\Sigma x_i)a_1 + (\Sigma x_i^2)a_2 + (\Sigma x_i^3)a_3 &= \Sigma y_i \\ (\Sigma x_i)a_0 + (\Sigma x_i^2)a_1 + (\Sigma x_i^3)a_2 + (\Sigma x_i^4)a_3 &= \Sigma x_i y_i \\ (\Sigma x_i^2)a_0 + (\Sigma x_i^3)a_1 + (\Sigma x_i^4)a_2 + (\Sigma x_i^5)a_3 &= \Sigma x_i^2 y_i \\ (\Sigma x_i^3)a_0 + (\Sigma x_i^4)a_1 + (\Sigma x_i^5)a_2 + (\Sigma x_i^6)a_3 &= \Sigma x_i^3 y_i \end{aligned}$$

becomes

$$\begin{aligned} 7a_0 + \quad 21a_1 + \quad 91a_2 + \quad 441a_3 &= \quad 14 \\ 21a_0 + \quad 91a_1 + \quad 441a_1 + \quad 2275a_3 &= \quad 52 \\ 91a_0 + \quad 441a_1 + \quad 2275a_2 + 12{,}201a_3 &= \quad 242 \\ 441a_0 + 2275a_1 + 12{,}201a_2 + 67{,}171a_3 &= 1258. \end{aligned}$$

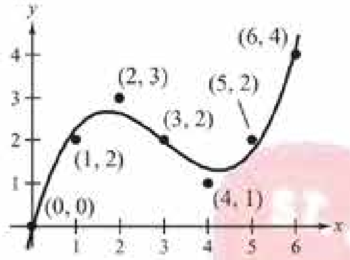
Using Gaussian elimination with pivoting on the matrix

$$\begin{bmatrix} 7 & 21 & 91 & 441 & 14 \\ 21 & 91 & 441 & 2275 & 52 \\ 91 & 441 & 2275 & 12{,}201 & 242 \\ 441 & 2275 & 12{,}201 & 67{,}171 & 1258 \end{bmatrix}$$

produces

$$\begin{bmatrix} 1.0000 & 5.1587 & 27.6667 & 152.3150 & 2.8526 \\ 0.0000 & 1.0000 & 8.5312 & 58.3482 & 0.6183 \\ 0.0000 & 0.0000 & 1.0000 & 9.7714 & 0.1286 \\ 0.0000 & 0.0000 & 0.0000 & 1.0000 & 0.1667 \end{bmatrix},$$

Figure 10.2



which implies that

$$a_3 = 0.1667, \quad a_2 = -1.5000, \quad a_1 = 3.6905, \quad a_0 = -0.0714.$$

So the cubic model is

$$y = 0.1667x^3 - 1.5000x^2 + 3.6905x - 0.0714.$$

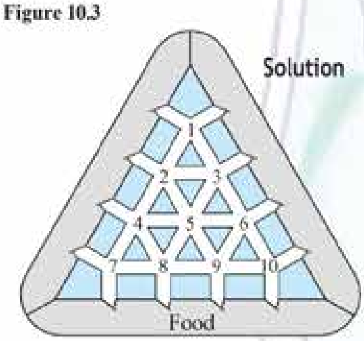
Figure 10.2 compares this model with the given points.

Applications of the Gauss-Seidel Method

EXAMPLE 3 ***An Application to Probability***

Figure 10.3 is a diagram of a maze used in a laboratory experiment. The experiment begins by placing a mouse at one of the ten interior intersections of the maze. Once the mouse emerges in the outer corridor, it cannot return to the maze. When the mouse is at an interior intersection, its choice of paths is assumed to be random. What is the probability that the mouse will emerge in the "food corridor" when it begins at the ith intersection?

Figure 10.3



Solution Let the probability of winning (getting food) by starting at the ith intersection be represented by p_i. Then form a linear equation involving p_i and the probabilities associated with the intersections bordering the ith intersection. For instance, at the first intersection the mouse has a probability of $\frac{1}{4}$ of choosing the upper right path and losing, a probability of $\frac{1}{4}$ of choosing the upper left path and losing, a probability of $\frac{1}{4}$ of choosing the lower right path (at which point it has a probability of p_3 of winning), and a probability of $\frac{1}{4}$ of choosing the lower left path (at which point it has a probability of p_2 of winning). So

$$p_1 = \underbrace{\tfrac{1}{4}(0)}_{\text{Upper right}} + \underbrace{\tfrac{1}{4}(0)}_{\text{Upper left}} + \underbrace{\tfrac{1}{4}p_2}_{\text{Lower left}} + \underbrace{\tfrac{1}{4}p_3}_{\text{Lower right}}.$$

Using similar reasoning, the other nine probabilities can be represented by the following equations.

$$
\begin{aligned}
p_2 &= \tfrac{1}{5}(0) + \tfrac{1}{5}p_1 + \tfrac{1}{5}p_3 + \tfrac{1}{5}p_4 + \tfrac{1}{5}p_5\\
p_3 &= \tfrac{1}{5}(0) + \tfrac{1}{5}p_1 + \tfrac{1}{5}p_2 + \tfrac{1}{5}p_5 + \tfrac{1}{5}p_6\\
p_4 &= \tfrac{1}{5}(0) + \tfrac{1}{5}p_2 + \tfrac{1}{5}p_5 + \tfrac{1}{5}p_7 + \tfrac{1}{5}p_8\\
p_5 &= \tfrac{1}{6}p_2 + \tfrac{1}{6}p_3 + \tfrac{1}{6}p_4 + \tfrac{1}{6}p_6 + \tfrac{1}{6}p_8 + \tfrac{1}{6}p_9\\
p_6 &= \tfrac{1}{5}(0) + \tfrac{1}{5}p_3 + \tfrac{1}{5}p_5 + \tfrac{1}{5}p_9 + \tfrac{1}{5}p_{10}\\
p_7 &= \tfrac{1}{4}(0) + \tfrac{1}{4}(1) + \tfrac{1}{4}p_4 + \tfrac{1}{4}p_8\\
p_8 &= \tfrac{1}{5}(1) + \tfrac{1}{5}p_4 + \tfrac{1}{5}p_5 + \tfrac{1}{5}p_7 + \tfrac{1}{5}p_9\\
p_9 &= \tfrac{1}{5}(1) + \tfrac{1}{5}p_5 + \tfrac{1}{5}p_6 + \tfrac{1}{5}p_8 + \tfrac{1}{5}p_{10}\\
p_{10} &= \tfrac{1}{4}(0) + \tfrac{1}{4}(1) + \tfrac{1}{4}p_6 + \tfrac{1}{4}p_9
\end{aligned}
$$

Rewriting these equations in standard form produces the following system of ten linear equations in ten variables.

$$
\begin{aligned}
4p_1 - p_2 - p_3 &= 0\\
-p_1 + 5p_2 - p_3 - p_4 - p_5 &= 0\\
-p_1 - p_2 + 5p_3 - p_5 - p_6 &= 0\\
- p_2 + 5p_4 - p_5 - p_7 - p_8 &= 0\\
- p_2 - p_3 - p_4 + 6p_5 - p_6 - p_8 - p_9 &= 0\\
- p_3 - p_5 + 5p_6 - p_9 - p_{10} &= 0\\
- p_4 + 4p_7 - p_8 &= 1\\
- p_4 - p_5 - p_7 + 5p_8 - p_9 &= 1\\
- p_5 - p_6 - p_8 + 5p_9 - p_{10} &= 1\\
- p_6 - p_9 + 4p_{10} &= 1
\end{aligned}
$$

The augmented matrix for this system is

$$
\begin{bmatrix}
4 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
-1 & 5 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0\\
-1 & -1 & 5 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0\\
0 & -1 & 0 & 5 & -1 & 0 & -1 & -1 & 0 & 0 & 0\\
0 & -1 & -1 & -1 & 6 & -1 & 0 & -1 & -1 & 0 & 0\\
0 & 0 & -1 & 0 & -1 & 5 & 0 & 0 & -1 & -1 & 0\\
0 & 0 & 0 & -1 & 0 & 0 & 4 & -1 & 0 & 0 & 1\\
0 & 0 & 0 & -1 & -1 & 0 & -1 & 5 & -1 & 0 & 1\\
0 & 0 & 0 & 0 & -1 & -1 & 0 & -1 & 5 & -1 & 1\\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 4 & 1
\end{bmatrix}.
$$

Using the Gauss-Seidel method with an initial approximation of $p_1 = p_2 = \ldots = p_{10} = 0$ produces (after 18 iterations) an approximation of

$$\begin{aligned} p_1 &= 0.090, & p_2 &= 0.180 \\ p_3 &= 0.180, & p_4 &= 0.298 \\ p_5 &= 0.333, & p_6 &= 0.298 \\ p_7 &= 0.455, & p_8 &= 0.522 \\ p_9 &= 0.522, & p_{10} &= 0.455. \end{aligned}$$

The structure of the probability problem described in Example 3 is related to a technique called **finite element analysis,** which is used in many engineering problems.

Note that the matrix developed in Example 3 has mostly zero entries. Such matrices are called **sparse.** For solving systems of equations with sparse coefficient matrices, the Jacobi and Gauss-Seidel methods are much more efficient than Gaussian elimination.

Applications of the Power Method

Section 7.4 introduced the idea of an *age transition matrix* as a model for population growth. Recall that this model was developed by grouping the population into n age classes of equal duration. So for a maximum life span of L years, the age classes are given by the following intervals.

$$\underbrace{\left[0, \frac{L}{n}\right)}_{\text{First age class}}, \quad \underbrace{\left[\frac{L}{n}, \frac{2L}{n}\right)}_{\text{Second age class}}, \ldots, \quad \underbrace{\left[\frac{(n-1)L}{n}, L\right]}_{n\text{th age class}}$$

The number of population members in each age class is then represented by the age distribution vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}. \qquad \begin{matrix} \text{Number in first age class} \\ \text{Number in second age class} \\ \\ \text{Number in } n\text{th age class} \end{matrix}$$

Over a period of L/n years, the *probability* that a member of the ith age class will survive to become a member of the $(i + 1)$th age class is given by p_i, where $0 \le p_i \le 1$, $i = 1, 2, \ldots, n - 1$. The *average number* of offspring produced by a member of the ith age class is given by b_i, where $0 \le b_i$, $i = 1, 2, \ldots, n$. These numbers can be written in matrix form as follows.

$$A = \begin{bmatrix} b_1 & b_2 & b_3 & \cdots & b_{n-1} & b_n \\ p_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & p_2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & p_{n-1} & 0 \end{bmatrix}$$

Multiplying this age transition matrix by the age distribution vector for a given period of time produces the age distribution vector for the next period of time. That is,

$$A\mathbf{x}_i = \mathbf{x}_{i+1}.$$

In Section 7.4 you saw that the growth pattern for a population is *stable* if the same percentage of the total population is in each age class each year. That is,

$$A\mathbf{x}_i = \mathbf{x}_{i+1} = \lambda\mathbf{x}_i.$$

For populations with many age classes, the solution to this eigenvalue problem can be found with the power method, as illustrated in Example 4.

EXAMPLE 4 ***A Population Growth Model***

Assume that a population of human females has the following characteristics.

Age Class (in years)	*Average Number of Female Children During Ten-Year Period*	*Probability of Surviving to Next Age Class*
$0 \le \text{age} < 10$	0.000	0.985
$10 \le \text{age} < 20$	0.174	0.996
$20 \le \text{age} < 30$	0.782	0.994
$30 \le \text{age} < 40$	0.263	0.990
$40 \le \text{age} < 50$	0.022	0.975
$50 \le \text{age} < 60$	0.000	0.940
$60 \le \text{age} < 70$	0.000	0.866
$70 \le \text{age} < 80$	0.000	0.680
$80 \le \text{age} < 90$	0.000	0.361
$90 \le \text{age} < 100$	0.000	0.000

Find a stable age distribution for this population.

Solution The age transition matrix for this population is

$$A = \begin{bmatrix} 0.000 & 0.174 & 0.782 & 0.263 & 0.022 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.985 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.996 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.994 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.990 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.975 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.940 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.866 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.680 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.361 & 0 \end{bmatrix}.$$

To apply the power method with scaling to find an eigenvector for this matrix, use an initial approximation of $\mathbf{x}_0 = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$. The following is an approximation for an eigenvector of A, with the percentage of each age in the total population.

$\mathbf{x} =$

Eigenvector	Age Class	Percentage in Age Class
1.000	$0 \le \text{age} < 10$	15.27
0.925	$10 \le \text{age} < 20$	14.13
0.864	$20 \le \text{age} < 30$	13.20
0.806	$30 \le \text{age} < 40$	12.31
0.749	$40 \le \text{age} < 50$	11.44
0.686	$50 \le \text{age} < 60$	10.48
0.605	$60 \le \text{age} < 70$	9.24
0.492	$70 \le \text{age} < 80$	7.51
0.314	$80 \le \text{age} < 90$	4.80
0.106	$90 \le \text{age} < 100$	1.62

The eigenvalue corresponding to the eigenvector $\mathbf{x}$ in Example 4 is $\lambda \approx 1.065$. That is,

$$A\mathbf{x} = A\begin{bmatrix} 1.000 \\ 0.925 \\ 0.864 \\ 0.806 \\ 0.749 \\ 0.686 \\ 0.605 \\ 0.492 \\ 0.314 \\ 0.106 \end{bmatrix} \approx \begin{bmatrix} 1.065 \\ 0.985 \\ 0.921 \\ 0.859 \\ 0.798 \\ 0.731 \\ 0.645 \\ 0.524 \\ 0.334 \\ 0.113 \end{bmatrix} \approx 1.065\begin{bmatrix} 1.000 \\ 0.925 \\ 0.864 \\ 0.806 \\ 0.749 \\ 0.686 \\ 0.605 \\ 0.492 \\ 0.314 \\ 0.106 \end{bmatrix}.$$

This means that the population in Example 4 increases by 6.5% every ten years.

REMARK: Should you try duplicating the results of Example 4, you would notice that the convergence of the power method for this problem is very slow. The reason is that the dominant eigenvalue of $\lambda \approx 1.065$ is only slightly larger in absolute value than the next largest eigenvalue.

SECTION 10.4 ❑ EXERCISES

Applications of Gaussian Elimination with Pivoting

In Exercises 1–4, find the second-degree least squares regression polynomial for the given data. Then graphically compare the model to the given points.

1. $(-2, 1), (-1, 0), (0, 0), (1, 1), (3, 2)$
2. $(0, 4), (1, 2), (2, -1), (3, 0), (4, 1), (5, 4)$
3. $(-2, 1), (-1, 2), (0, 6), (1, 3), (2, 0), (3, -1)$
4. $(1, 1), (2, 1), (3, 0), (4, -1), (5, -4)$

In Exercises 5–8, find the third-degree least squares regression polynomial for the given data. Then graphically compare the model to the given points.

5. $(0, 0), (1, 2), (2, 4), (3, 1), (4, 0), (5, 1)$
6. $(1, 1), (2, 4), (3, 4), (5, 1), (6, 2)$
7. $(-3, 4), (-1, 1), (0, 0), (1, 2), (2, 5)$
8. $(-7, 2), (-3, 0), (1, -1), (2, 3), (4, 6)$

9. Find the second-degree least squares regression polynomial for the points

$$\left(-\frac{\pi}{2}, 0\right), \left(-\frac{\pi}{3}, \frac{1}{2}\right), (0, 1), \left(\frac{\pi}{3}, \frac{1}{2}\right), \left(\frac{\pi}{2}, 0\right)$$

Then use the results to approximate $\cos(\pi/4)$. Compare the approximation with the exact value.

10. Find the third-degree least squares regression polynomial for the points

$$\left(-\frac{\pi}{4}, -1\right), \left(-\frac{\pi}{3}, -\sqrt{3}\right), (0, 0), \left(\frac{\pi}{3}, \sqrt{3}\right), \left(\frac{\pi}{4}, 1\right).$$

Then use the result to approximate $\tan(\pi/6)$. Compare the approximation with the exact value.

11. The number of minutes a scuba diver can stay at a particular depth without acquiring decompression sickness is shown in the table. (Source: United States Navy's Standard Air Decompression Tables)

Depth (in feet)	35	40	50	60	70
Time (in minutes)	310	200	100	60	50

Depth (in feet)	80	90	100	110
Time (in minutes)	40	30	25	20

(a) Find the least squares regression line for these data.

(b) Find the second-degree least squares regression polynomial for these data.

(c) Sketch the graphs of the models found in parts (a) and (b).

(d) Use the models found in parts (a) and (b) to approximate the maximum number of minutes a diver should stay at a depth of 120 feet. (The value given in the Navy's tables is 15 minutes.)

12. The life expectancy for additional years of life for females in the United States as of 1998 is shown in the table. (Source: U.S. Census Bureau)

Current Age	10	20	30	40
Life Expectancy	70.6	60.8	51.0	41.4

Current Age	50	60	70	80
Life Expectancy	32.0	23.3	15.6	9.1

(a) Find the second-degree least squares regression polynomial for these data.

(b) Use the result of part (a) to predict the life expectancy of a newborn female and a female of age 100 years.

13. Total sales in billions of dollars of cellular phones in the Unites States from 1992 to 1999 are shown in the table. (Source: Electronic Market Data Book).

Year	1992	1993	1994	1995	1996	1997	1998	1999
Sales	1.15	1.26	1.28	2.57	2.66	2.75	2.78	2.81

(a) Find the second degree least squares regression polynomial for the data.

(b) Use the result of part (a) to predict the total cellular phone sales in 2005 and 2010.

(c) Are your predictions from part (b) realistic? Explain.

14. Total new domestic truck unit sales in hundreds of thousands in the United States from 1993 to 2000 are shown in the table. (Source: Ward's Auto info bank)

Year	1993	1994	1995	1996	1997	1998	1999	2000
Trucks	5.29	6.00	6.06	6.48	6.63	7.51	7.92	8.09

(a) Find the second degree least squares regression polynomial for the data.

(b) Use the result of part (a) to predict the total new domestic truck sales in 2005 and 2010.

(c) Are your predictions from part (b) realistic? Explain.

15. Find the least squares regression line for the population data given in Example 1. Then use the model to predict the world population in 2005 and 2010, and compare the results with the predictions obtained in Example 1.

16. Show that the formula for the least squares regression line presented in Section 2.5 is equivalent to the formula presented in this section. That is, if

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad A = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix},$$

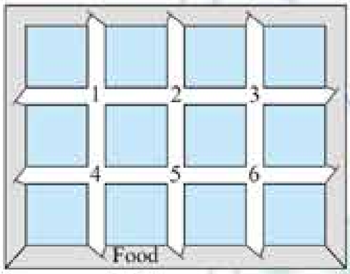
then the matrix equation $A = (X^TX)^{-1}X^TY$ is equivalent to

$$a_1 = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^2 - (\sum x_i)^2} \quad \text{and} \quad a_0 = \frac{\sum y_i}{n} - a_1\frac{\sum x_i}{n}.$$

Applications of the Gauss-Seidel Method

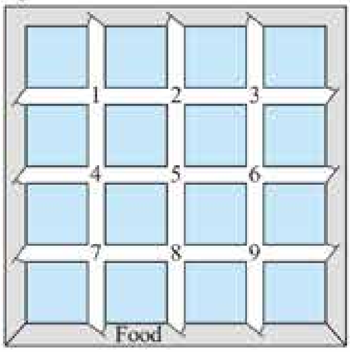
17. Suppose that the experiment in Example 3 is performed with the maze shown in Figure 10.4. Find the probability that the mouse will emerge in the food corridor when it begins in the ith intersection.

Figure 10.4



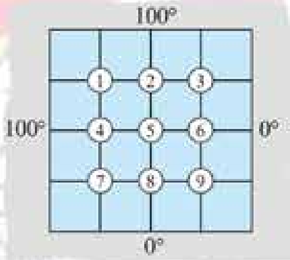
18. Suppose that the experiment in Example 3 is performed with the maze shown in Figure 10.5. Find the probability that the mouse will emerge in the food corridor when it begins in the ith intersection.

Figure 10.5



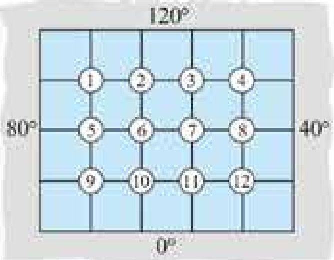
19. A square metal plate has a constant temperature on each of its four boundaries, as shown in Figure 10.6. Use a 4×4 grid to approximate the temperature distribution in the interior of the plate. Assume that the temperature at each interior point is the average of the temperatures at the four closest neighboring points.

Figure 10.6



20. A rectangular metal plate has a constant temperature on each of its four boundaries, as shown in Figure 10.7. Use a 4×5 grid to approximate the temperature distribution in the interior of the plate. Assume that the temperature at each interior point is the average of the temperatures at the four closest neighboring points.

Figure 10.7



Applications of the Power Method

In Exercises 21–24, the matrix represents the age transition matrix for a population. Use the power method with scaling to find a stable age distribution.

21. $A = \begin{bmatrix} 1 & 4 \\ \frac{1}{2} & 0 \end{bmatrix}$

22. $A = \begin{bmatrix} 1 & 2 \\ \frac{1}{4} & 0 \end{bmatrix}$

23. $A = \begin{bmatrix} 0 & 1 & 2 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \end{bmatrix}$

24. $A = \begin{bmatrix} 1 & 2 & 2 \\ \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \end{bmatrix}$

25. In Example 1 of Section 7.4, a laboratory population of rabbits is described. The age transition matrix for the population is

$$A = \begin{bmatrix} 0 & 6 & 8 \\ 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \end{bmatrix}.$$

Find a stable age distribution for this population.

26. A population has the following characteristics.

(a) A total of 75% of the population survives its first year. Of that 75%, 25% survives its second year. The maximum life span is three years.

(b) The average number of offspring for each member of the population is 2 the first year, 4 the second year, and 2 the third year.

Find a stable age distribution for this population. (See Exercise 9, Section 7.4.)

27. Apply the power method to the matrix

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

discussed in Chapter 7 (Fibonacci sequence). Use the power method to approximate the dominant eigenvalue of A. (The dominant eigenvalue is $\lambda = (1 + \sqrt{5})/2$.)

28. ***Writing*** In Example 2 of Section 2.5, the stochastic matrix

$$P = \begin{bmatrix} 0.70 & 0.15 & 0.15 \\ 0.20 & 0.80 & 0.15 \\ 0.10 & 0.05 & 0.70 \end{bmatrix}$$

represents the transition probabilities for a consumer preference model. Use the power method to approximate a dominant eigenvector for this matrix. How does the approximation relate to the steady-state matrix described in the discussion following Example 3 in Section 2.5?

29. In Exercise 9 of Section 2.5, a population of 10,000 is divided into nonsmokers, moderate smokers, and heavy smokers. Use the power method to approximate a dominant eigenvector for this matrix.

MS.S.REZVANA PARVEEN WAS BORN IN 2000 IN MADURAI, TAMIL NADU AND SHE IS CURRENTLY WORKING AS AN ASSITANT PROFESSOR IN THE DEPARTMENT OF MATHEMATICS . ST.JOSEPH'S COLLEGE OF ARTS AND SCIENCE FOR WOMEN,HOSUR. SHE HAS COMPLETED HER M.SC DEGREE IN ST.JOSEPH'S COLLEGE OF ARTS AND SCIENCE FOR WOMEN ,HOSUR.SHE HAS PUBLISHED 5 INTERNATIONAL CONFERENCE PROCEEDINGS. HER AREA OF INTEREST INCLUDING NUMERICAL METHODS

ISBN 978-933402495-1
9 789334 024951