An Introduction to Numerical Analysis
with MATLAB

Lecture Notes

Mohammad Sabawi

Department of Mathematics
College of Education for Women
Tikrit University

Email: mohammad.sabawi@tu.edu.iq

March 2019
List of Tables

2.1 Bisection Method Solution of Example 9 .......................... 20
2.2 False Position Method Solution of Example 10 .................. 21
2.3 Newton’s Method Solution of Example 11 ......................... 23
2.4 Newton’s Method Solution of Example 12 ........................ 24
2.5 Secant Method Solution of Example 14 ............................ 30
2.6 Fixed Point Method Solution of Example 16 ...................... 33
2.7 Fixed Point Method Solution of Example 16 ...................... 34
2.8 Newton’s and Accelerated Newton’s Methods Solutions of Ex-
ample 17 ................................................................. 36
3.1 Jacobi Iterative Solution of Example 28 ........................... 61
3.2 Gauss-Siedel Iterative Solution of Example 29 .................. 64
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Preface</strong></td>
<td>5</td>
</tr>
<tr>
<td><strong>1 Introduction</strong></td>
<td>6</td>
</tr>
<tr>
<td>1.1 Numerical Analysis: An Introduction</td>
<td>6</td>
</tr>
<tr>
<td>1.2 Numbers Representation in Computer</td>
<td>7</td>
</tr>
<tr>
<td>1.2.1 Floating-Point Numbers</td>
<td>8</td>
</tr>
<tr>
<td>1.3 Errors</td>
<td>9</td>
</tr>
<tr>
<td>1.3.1 Error Analysis</td>
<td>9</td>
</tr>
<tr>
<td>1.3.2 Sources of Error in Numerical Computations</td>
<td>9</td>
</tr>
<tr>
<td>1.3.3 Absolute and Relative Errors</td>
<td>10</td>
</tr>
<tr>
<td>1.3.4 Roundoff and Truncation Errors</td>
<td>11</td>
</tr>
<tr>
<td>1.4 Stable and Unstable Computations: Conditioning</td>
<td>11</td>
</tr>
<tr>
<td>1.5 Convergence and Order of Approximation</td>
<td>13</td>
</tr>
<tr>
<td><strong>Exercises</strong></td>
<td>15</td>
</tr>
<tr>
<td><strong>2 Numerical Solutions of Nonlinear Equations</strong></td>
<td>16</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>16</td>
</tr>
<tr>
<td>2.2 Bisection (Interval Halving) Method</td>
<td>17</td>
</tr>
<tr>
<td>2.3 False-Position Method</td>
<td>20</td>
</tr>
<tr>
<td>2.4 Newton’s Method</td>
<td>22</td>
</tr>
<tr>
<td>2.5 Numerical Solutions of Systems of Nonlinear Equations</td>
<td>25</td>
</tr>
<tr>
<td>2.6 Secant Method</td>
<td>29</td>
</tr>
<tr>
<td>2.7 Fixed Points and Functional Iteration</td>
<td>30</td>
</tr>
<tr>
<td>2.8 Acceleration of Iterative Methods</td>
<td>34</td>
</tr>
<tr>
<td>2.8.1 Modified Newton’s Methods</td>
<td>34</td>
</tr>
<tr>
<td>2.9 Computing Roots of Polynomials</td>
<td>36</td>
</tr>
<tr>
<td><strong>Exercises</strong></td>
<td>38</td>
</tr>
<tr>
<td><strong>3 Solving Systems of Linear Equations</strong></td>
<td>39</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>39</td>
</tr>
<tr>
<td>3.2 Norms of Matrix and Vectors</td>
<td>39</td>
</tr>
<tr>
<td>3.3 Direct Methods</td>
<td>43</td>
</tr>
<tr>
<td>3.3.1 Backward Substitution Method</td>
<td>43</td>
</tr>
<tr>
<td>3.3.2 Forward Substitution Method</td>
<td>45</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Gaussian Elimination Method</td>
</tr>
<tr>
<td>3.3.4</td>
<td>Gauss-Jordan Elimination Method</td>
</tr>
<tr>
<td>3.4</td>
<td>LU and Cholesky Factorisations</td>
</tr>
<tr>
<td>3.5</td>
<td>Iterative Methods</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Jacobi Method</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Gauss-Siedel Method</td>
</tr>
<tr>
<td>Exercises</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Interpolation and Extrapolation</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>4.2</td>
<td>Lagrange Interpolation</td>
</tr>
<tr>
<td>4.3</td>
<td>Newton Interpolation</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Finite Divided Differences</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Newton’s Interpolation Divided Difference Formula</td>
</tr>
<tr>
<td>Bibliography</td>
<td></td>
</tr>
</tbody>
</table>
Preface

The aim of these class notes is to cover the necessary materials in a standard numerical analysis course and it is not intended to add to the plethora of Numerical Analysis texts. We tried our best to write these notes in concise, clear and accessible way, to make them more attractive to the readers. These lecture notes cover the basic and fundamental concepts and principles in numerical analysis and it is not a comprehensive introduction to numerical analysis. We emphasise in these notes on the mathematical principles via explaining them by the aid of numerical software MATLAB. The prerequisite material for this course are a course in Calculus, Linear Algebra and Differential Equations. A basic knowledge in MATLAB is helpful but it is not necessary. There is a glut of numerical software nowadays, among these we chose to use MATLAB because of its wide capabilities in scientific computing. These notes consist of ten chapters and each chapter ends with a set of exercises address the topics covered in each chapter.
Chapter 1
Introduction

1.1 Numerical Analysis: An Introduction

Numerical analysis is a branch of mathematics studies the methods and algorithms which used for solving a variety of problems in different areas of today’s life such as mathematics, physics, engineering, medicine and social and life sciences. The main objective of numerical analysis is investigation finding new mathematical approaches for approximating the underlying problems, and also development of the current algorithms and numerical schemes to make them more efficient and reliable. The advent of computers revolutionise numerical analysis and nowadays with parallel and super computers the numerical computations became more easier compared with the past where solving simple problems take a long time, much effort and require hard work. In principle, numerical analysis mainly focuses on the ideas of stability, convergence, accuracy, consistency and error analysis. In the literature numerical analysis also known as scientific computing, scientific computation, numerics, computational mathematics and numerical mathematics. Numerical analysis can be divided into the following fields:

1. Numerical Solutions of Linear Algebraic Equations.
3. Interpolation and Extrapolation.
5. Numerical Differentiation.

Numerical analysis is dated back to the Babylonians works in approximating the square root of 2. During this long journey of evolution many scientists contributed to its development and progress among these we just name a few such as Lagrange, Gauss, Newton, Euler, Legendre and Simpson.

1.2 Numbers Representation in Computer

Human beings do arithmetic in their daily life using the decimal (base 10) number system. Nowadays, most computers use binary (base 2) number system. We enter the information to computers using the decimal system but computers transform them to the binary system by using the machine language.

Definition 1 (Scientific Notation). Let $k$ be a real number, then $k$ can be written in the following form

$$k = m \times 10^n,$$

where $m$ is any real number and the exponent $n$ is an integer. This notation is called the scientific notation or scientific form and sometimes referred to as standard form.

Example 1. Write the following numbers in scientific notation:

1. 0.00000834.
2. 25.45879.
3. 3400000.
CHAPTER 1. INTRODUCTION

4. 33.
5. 2,300,000,000.

Solution:
1. $0.00000834 = 8.34 \times 10^{-6}$.
2. $25.45879 = 2.545879 \times 10^1$.
3. $3400000 = 3.4 \times 10^6$.
4. $33 = 3.3 \times 10^1$.
5. $2.3 \times 10^9$.

1.2.1 Floating-Point Numbers

In the decimal system any real number $a \neq 0$ can be written in the decimal normalised floating-point form in the following way

$$a = \pm 0.d_1d_2d_3 \cdots d_kd_{k+1}d_{k+2} \cdots \times 10^n, \quad 1 \leq d_1 \leq 9, \quad 0 \leq d_i \leq 9,$$

for each $i = 2, \cdots$, and $n$ is an integer called the exponent ($n$ can be positive, negative or zero). In computers we use a finite number of digits in representing the numbers and we obtain the following form

$$b = \pm 0.d_1d_2d_3 \cdots d_k \times 10^n, \quad 1 \leq d_1 \leq 9, \quad 0 \leq d_i \leq 9,$$

for each $i = 2, \cdots, k$. These numbers are called k-digit decimal machine numbers.

Also, the normalised floating-point decimal representation of the number $a \neq 0$ can be written in other way as

$$a = \pm r \times 10^n, \quad \left(\frac{1}{10} \leq r < 1\right),$$

the number $r$ is called the normalised mantissa.

The floating-point representation in binary number system can be defined by the same way as in the decimal number system. If $a \neq 0$, it can be represented as

$$a = \pm p \times 2^m, \quad \left(\frac{1}{2} \leq p < 1\right),$$

where $p = (0.b_1b_2b_3 \cdots)_2, \quad b_1 = 1$.  

Mohammad Sabawi/Numerical Analysis
1.3 Errors

Occurrence of error is unavoidable in the field of scientific computing. Instead, numerical analysts try to investigate the possible and best ways to minimise the error. The study of the error and how to estimate and minimise it are the fundamental issues in error analysis.

1.3.1 Error Analysis

In numerical analysis we approximate the exact solution of the problem by using numerical method and consequently an error is committed. The numerical error is the difference between the exact solution and the approximate solution.

Definition 2 (Numerical Error). Let $x$ be the exact solution of the underlying problem and $x^*$ its approximate solution, then the error (denoted by $e$) in solving this problem is

$$e = x - x^*.$$

1.3.2 Sources of Error in Numerical Computations

- **Blunders (Gross Errors)** These errors also called humans errors, and are caused by humans mistakes and oversight and can be minimised by taking care during scientific investigations. These errors will add to the total error of the underlying problem and can significantly affect the accuracy of solution.

- **Modelling Errors** These errors arise during the modelling process when scientists ignore effecting factors in the model to simplify the problem. Also, these errors known as formulation errors.

- **Data Uncertainty** These errors are due to the uncertainty of the physical problem data and also known as data errors.

- **Discretisation Errors** Computers represent a function of continuous variable by a number of discrete values. Also, scientists approximate and replace complex continuous problems by discrete ones and this results in discretisation errors.
1.3.3 Absolute and Relative Errors

Definition 3 (Absolute Error). The absolute error $\hat{e}$ of the error $e$ is defined as the absolute value of the error $e$

$$\hat{e} = |x - x^*|.$$  

Definition 4 (Relative Error). The relative error $\tilde{e}$ of the error $e$ is defined as the ratio between the absolute error $\hat{e}$ and the absolute value of the exact solution $x$

$$\tilde{e} = \frac{\hat{e}}{|x|} = \frac{|x - x^*|}{|x|}, x \neq 0.$$  

Example 2. Let $x = 3.141592653589793$ be the value of the constant ratio $\pi$ correct to 15 decimal places and $x^* = 3.14159265$ be an approximation of $x$. Compute the following quantities:

a. The error.

b. The absolute error.

c. The relative error.

Solution:

a. The error

$$e = x - x^* = 3.141592653589793 - 3.14159265 = 3.589792907376932e - 09$$

$$= 3.589792907376932 \times 10^{-9} = 0.00000003589792907376932.$$  

b. The absolute error

$$\hat{e} = |x - x^*| = |3.141592653589793 - 3.14159265| = 3.589792907376932e - 09.$$  

c. The relative error

$$\tilde{e} = \frac{\hat{e}}{|x|} = \frac{|x - x^*|}{|x|} = \frac{3.141592653589793 - 3.14159265}{3.141592653589793}$$

$$= \frac{3.589792907376932e - 09}{3.141592653589793} = 1.142666571770530e - 09.$$
1.3.4 Roundoff and Truncation Errors

Computers represent numbers in finite number of digits and hence some quantities cannot be represented exactly. The error caused by replacing a number \( a \) by its closest machine number is called the \textbf{roundoff error} and the process is called \textbf{correct rounding}.

Truncation errors also sometimes called \textbf{chopping errors} are occurred when chopping an infinite number and replaced it by a finite number or by truncated a series after finite number of terms.

\textbf{Example 3.} Approximate the following decimal numbers to three digits by using rounding and chopping (truncation) rules:

1. \( x_1 = 1.34579 \).
2. \( x_2 = 1.34679 \).
3. \( x_3 = 1.34479 \).
4. \( x_4 = 3.34379 \).
5. \( x_5 = 2.34579 \).

\textbf{Solution:}

(i) **Rounding:**
   
   \begin{align*}
   (a) & \quad x_1 = 1.35. \\
   (b) & \quad x_2 = 1.35. \\
   (c) & \quad x_3 = 1.34. \\
   (d) & \quad x_4 = 3.34. \\
   (e) & \quad x_5 = 2.35. \\
   
   (ii) **Chopping:**
   \end{align*}

1.4 Stable and Unstable Computations: Conditioning

\textit{Stability} is one of the most important characteristics in any efficient and robust numerical scheme.

\textbf{Definition 5} (Numerical Stability). The numerical algorithm is called \textbf{stable} if the final result is relatively not affected by the perturbations during computation process. Otherwise it is called \textbf{unstable}.

The \textit{stability} notion is analogous and closely related to the notion of conditioning.
Definition 6 (Conditioning). **Conditioning** is a measure of how sensitive the output to small changes in the input data. In literature **conditioning** is also called **sensitivity**.

- The problem is called **well-conditioned** or **insensitive** if small changes in the input data lead to small changes in the output data.
- The problem is called **ill-conditioned** or **sensitive** if small changes in the input data lead to big changes in the output data.

Definition 7 (Condition Number of a Function). If \( f \) is a differentiable function at \( x \) in its domain then the **condition number** of \( f \) at \( x \) is

\[
\text{Cond}(f(x)) = \frac{|xf'(x)|}{|f(x)|}, \quad f(x) \neq 0.
\]

Note: Condition number of a function \( f \) at \( x \) in its domain sometimes denoted by \( C_f(x) \).

Definition 8 (Condition Number of a Matrix). If \( A \) is a non-singular \( n \times m \) matrix, the **condition number** of \( A \) is defined by

\[
\text{cond}(A) = \|A\|\|A^{-1}\|,
\]

where

\[
\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|},
\]

and \( x \) is a \( m \times 1 \) column vector.

Definition 9 (Well-Posed Problem). The problem is **well-posed** if satisfies the following three conditions:

a. The solution exists.

b. The solution is unique.

c. The solution depends continuously on problem data.

Otherwise, the problem is called **ill-posed**.

Remark 1. Note that:

1. The problem is **ill-posed** or **sensitive** if \( \text{cond} \gg 1 \).

2. The problem is **well-posed** or **insensitive** if \( \text{cond} < 1 \).
Example 4. Find the condition number of the function \( f(x) = \sqrt{x} \).

**Solution:**

\[
f(x) = \sqrt{x} \implies f'(x) = \frac{1}{2\sqrt{x}}, \quad x \neq 0,
\]

implies that

\[
\text{cond}(f(x)) = \frac{|xf'(x)|}{|f(x)|} = \frac{|\frac{x}{2\sqrt{x}}|}{|\sqrt{x}|} = \frac{1}{2}
\]

This indicates that the small changes in the input data lead to changes in the output data of half size the changes in the input data.

Example 5. Let

\[
A = \begin{bmatrix}
1 & -1 & 1 \\
1 & 0.5 & 3 \\
0.1 & 1 & 0.3
\end{bmatrix},
\]

the inverse of \( A \) can be computed by using MATLAB command `inv(A)` to obtain

\[
A^{-1} = \begin{bmatrix}
4.7500 & -2.1667 & 5.8333 \\
0.5000 & -0.3333 & 1.6667 \\
-3.2500 & 1.8333 & -4.1667
\end{bmatrix}.
\]

Also, the condition number of \( A \) and its inverse can be computed using MATLAB commands `cond(A)` and `cond(inv(A))` to have \( \text{cond}(A) = 37.8704 \) and \( \text{cond}(A^{-1}) = 37.8704 \). We notice that the matrix \( A \) and its inverse have the same condition number.

**Definition 10 (Accuracy).** It is a measure of closeness of the approximate solution to the exact solution.

**Definition 11 (Precision).** It is a measure of closeness of the two or more measurements to each other.

**Remark 2.** Note that the accuracy and precision are different and they are not related. The problem maybe very accurate but imprecise and vice versa.

### 1.5 Convergence and Order of Approximation

Convergence of the numerical solution to the analytical solution is one of the important characteristic in any good and reliable numerical scheme.
CHAPTER 1. INTRODUCTION

**Definition 12** (Convergence of a Sequence). Let \( \{a_n\} \) be an infinite sequence of real numbers. This sequence is said to be **convergent** to a real number \( a \) (has a limit at \( a \)) if, for any \( \epsilon > 0 \) there exists a positive integer \( N(\epsilon) \) such that

\[
|a_n - a| < \epsilon, \text{ whenever } n > N(\epsilon),
\]

Otherwise it is called a **divergent** sequence, \( a \) is called the limit of the sequence \( a_n \). Other commonly used notations for convergence are:

\[
\lim_{n \to \infty} a_n = a \text{ or } a_n \to a \text{ as } n, \text{ or } \lim_{n \to \infty} (a_n - a) = 0,
\]

this means that the sequence \( \{a_n\} \) **converges** to a otherwise it **diverges**.

**Definition 13** (Order of Convergence). Let the sequence \( \{a_n\} \) converges to \( a \) and set \( e_n = a_n - a \) for any \( n > 0 \). If two positive constants \( M \neq 0 \) and \( q > 0 \) exist, such that

\[
\lim_{n \to \infty} \frac{|a_{n+1} - a|}{|a_n - a|^q} = \lim_{n \to \infty} \frac{|e_{n+1}|}{|e_n|^q} = M,
\]

then the sequence \( \{a_n\} \) is to be convergent to \( a \) with the order of convergence \( q \).

If \( q = 1 \), the convergence is called linear.
If \( q = 2 \), the convergence is called quadratic.
If \( q = 3 \), the convergence is called cubic.

Note that the convergence gets more rapid as \( q \) gets larger and larger.

**Example 6.** Consider the sequence \( \{\frac{1}{n}\} \), where \( n \) is a positive integer. Observe that \( \frac{1}{n} \to 0 \) as \( n \to \infty \), it follows that

\[
\lim_{n \to \infty} \frac{1}{n} = 0.
\]

**Definition 14** (Order of Approximation \( O(h^n) \)). The function \( f(h) \) is said to be **big Oh** of the function \( g(h) \), if two real constants \( c \), and \( C \) exist such that

\[
|f(h)| \leq C|g(h)| \text{ whenever } h < c,
\]

and denoted by \( f(h) = O(g(h)) \). The order of approximation is used to determine the rate at which a function converges.

**Example 7.** Consider the functions \( f(x) = x + 1 \) and \( g(x) = x^2 \), where \( x \geq 1 \). Observe that \( x \leq x^2 \) and \( 1 \leq x^2 \) for \( x \geq 1 \), hence \( f(x) = x + 1 \leq 2x^2 = 2g(x) \) for \( x \geq 1 \). Consequently, \( f(x) = O(g(x)) \).
Exercises

Exercise 1. Write the following numbers in scientific form:

1. 23.123.
2. 30,000,000.
3. 0.000001573.
4. 39776444.
5. −345.386443.
6. −23000000.

Exercise 2. Evaluate error, absolute error and relative error of the following values and their approximations:

1. \(x = 1,000,000, \ x^* = 999,999.\)
2. \(y = 0.00012887765, \ y^* = 0.00012897766.\)
3. \(z = 9776.96544, \ z^* = 9775.66544.\)

Exercise 3. Approximate the following numbers to four digits using rounding and chopping:

1. 1.98876.
2. 33.87654.
3. 8.98879.
4. 2.88778.

Exercise 4. Compute the condition number of the following functions:

1. \(f(x) = \cos(x).\)
2. \(f(x) = \cos^{-1}(x).\)
Chapter 2
Numerical Solutions of Nonlinear Equations

2.1 Introduction
Nonlinear algebraic equations are widely spread in science and engineering and therefore their solutions are important scientific applications. There are a glut of numerical methods for solving these equations, and in these lecture notes, we study the most commonly used ones such as bisection, secant and Newton methods. Locating positions of roots of nonlinear equation is a topic of great importance in numerical mathematical analysis. The problem under consideration maybe has a root or has no root at all.

Definition 15 (Zero of a Function). Let $f$ be a real or complex valued function of a real or complex variable $x$. A real or complex number $r$ satisfies $f(r) = 0$ is called zero of $f$ or also called a root of equation $f(x) = 0$.

For example, the function $f(x) = x^2 - 5x + 6 = (x - 2)(x - 3)$ has two real zeros $r_1 = 2$ and $r_2 = 3$, whereas the corresponding equation $x^2 - 5x + 6 = (x - 2)(x - 3) = 0$ has two real roots $r_1 = 2$ and $r_2 = 3$.

Definition 16 (Order of a Zero). Let $f$ and its derivatives $f', f'', \ldots, f^{(M)}$ are continuous and defined on an interval about the zero $x = r$. The function $f$ or the equation $f(x) = 0$ is said to be has a zero or a root of order $M \geq 1$ at $x = r$ if and only if

$$f(r) = 0, \ f'(r) = 0, \ f''(r) = 0, \ldots, \ f^{(M-1)}(r) = 0, \ f^{(M)} \neq 0.$$
CHAPTER 2. NUMERICAL SOLUTIONS OF NONLINEAR EQUATIONS

If \( M = 1 \) then \( r \) is called a simple zero or a simple root, and if \( M > 1 \) it is called a multiple zero or a multiple root. A zero (root) of order \( M = 2 \) is called a double zero (root), and so on. Also, the zero (root) of order \( M \) is called a zero (root) of multiplicity \( M \).

Lemma 1. If the function \( f \) has a zero \( r \) of multiplicity \( M \), then there exists a continuous function \( h \) such that
\[
f(x) = (x - r)^M h(x), \quad \lim_{x \to r} h(x) \neq 0.
\]

Theorem 5 (Simple Zero Theorem). Assume that \( f \in C^1[a, b] \). Then, \( f \) has a simple zero at \( r \in (a, b) \) if and only if \( f(r) = 0 \) and \( f'(r) \neq 0 \).

Example 8. Show that the function \( f(x) = e^{2x} - x^2 - 2x - 1 \) has a zero of multiplicity 2 (double zero) at \( x = 0 \).

Solution:
\[
f(x) = e^{2x} - x^2 - 2x - 1, \quad f'(x) = 2e^{2x} - 2x - 2, \quad \text{and} \quad f''(x) = 4e^{2x} - 2.
\]
Hence,
\[
f(0) = e^0 - 0 - 0 - 1 = 0, \quad f'(0) = 2e^0 - 0 - 2 = 0, \quad \text{and} \quad f''(0) = 4e^0 - 2 = 2 
eq 0,
\]
so, this implies that \( f \) has a double zero at \( x = 0 \).

2.2 Bisection (Interval Halving) Method

It is a bracketing method used to find a zero of a continuous function \( f \) on the interval \([a, b]\) where \( a \) and \( b \) are real numbers, i.e. find \( x \) such that \( f(x) = 0 \). This method based on the Intermediate Value Theorem and also known as Bolzano or binary search method. This method requires \( f(a) \) and \( f(b) \) have opposite signs since \( f \) is continuous its graph is unbroken and according to the Intermediate Value Theorem it has a root \( x = r \) somewhere in the interval. The first step in the solution process is to compute the midpoint \( c = \frac{a+b}{2} \) of the interval \([a, b]\) and to proceed we consider the three cases:

1. If \( f(a)f(c) < 0 \) then \( r \) lies in \([a, c]\).
2. If \( f(c)f(b) < 0 \) then \( r \) lies in \([c, b]\).
3. If \( f(c) = 0 \) then the root is \( c = r \).
Chapter 2. Numerical Solutions of Nonlinear Equations

To begin, set $a_1 = a$ and $b_1 = b$ and let $c_1 = \frac{a_1 + b_1}{2}$ be the midpoint of the interval $[a_1, b_1] = [a, b]$.

- If $f(c_1) = 0$ then the root is $r = c_1$.
- If $f(c_1) \neq 0$ then either $f(a_1)f(c_1) < 0$ or $f(c_1)f(b_1) < 0$.
  
  (i). If $f(a_1)f(c_1) < 0$ then $r$ lies in $[a_1, c_1]$, and set $a_2 = a_1$ and $b_2 = c_1$, i.e. $[a_2, b_2] = [a_1, c_1]$.
  
  (ii). If $f(c_1)f(b_1) < 0$ then $r$ lies in $[c_1, b_1]$, and set $a_2 = c_1$ and $b_2 = b_1$, i.e. $[a_2, b_2] = [c_1, b_1]$.
  
  (iii). Compute $c_2 = \frac{a_2 + b_2}{2}$ the midpoint of the interval $[a_2, b_2]$.
  
  (iv). Then, we proceed in this way until we reach the $n$th interval $[a_n, b_n]$ and then compute its midpoint $c_n = \frac{a_n + b_n}{2}$.

- Finally, construct the interval $[a_{n+1}, b_{n+1}]$ which brackets the root and its midpoint $c_{n+1} = \frac{a_{n+1} + b_{n+1}}{2}$ will be an approximation to the root $r$.

Remark 3. (a). The interval $[a_{n+1}, b_{n+1}]$ is wide as half as the interval $[a_n, b_n]$ i.e. the width of each interval is as half as the width of the previous interval. Let $\{\ell_n\}$ be a sequence of widths of intervals $[a_n, b_n]$, i.e. $\ell_n = \frac{b_n - a_n}{2}$, $n = 1, 2, \cdots$. Hence $\lim_{n \to \infty} \ell_n = 0$, where $\epsilon$ is the preassigned value of the error (tolerance) i.e.

$$|r_{n+1} - r_n| \leq \epsilon, \ n = 0, 1, \cdots,$$

(b). The sequence of left endpoints $a_n, \ n = 1, 2, \cdots, \ $ is increasing and the sequence $b_n, \ n = 1, 2, \cdots \ $ of right endpoints is decreasing i.e.

$$a_0 \leq a_1 \leq a_2 \leq \cdots \leq a_n \leq \cdots \leq r \leq \cdots \leq b_n \leq \cdots \leq b_2 \leq b_1 \leq b_0.$$

Theorem 6 (Bisection Method Theorem). Let $f \in C[a, b]$ such that $f(a)f(b) < 0$ and that there exists a number $r \in [a, b]$ such that $f(r) = 0$, and $\{c_n\}$ a sequence of the midpoints of intervals $[a_n, b_n]$ constructed by the bisection method, then the error in approximating the root $r$ in the $n$th step is:

$$|e_n| = |r - c_n| \leq \frac{b - a}{2^{n+1}}, \ n = 0, 1, \cdots.$$

Hence, the sequence $\{c_n\}$ is convergent and its limit is the root $r$, i.e.

$$\lim_{n\to\infty} c_n = r.$$
CHAPTER 2. NUMERICAL SOLUTIONS OF NONLINEAR EQUATIONS

Proof. For Proof see the References or have a look in any standard numerical analysis text. \hfill \Box

Remark 4. Remark

- The number $N$ of repeated bisections required to compute the $n$th approximation (midpoint) $c_n$ of the root $r$ is:

\[
N = \text{int}\left(\frac{\ln(b-a) - \ln(c)}{\ln(2)}\right).
\]

- The width of the $n$th interval is:

\[
|b_n - a_n| = \frac{|b_0 - a_0|}{2^n}.
\]

Example 9. (a) Use bisection method to show that $f(x) = x \sin(x) - 1 = 0$ has a real root in $[0.5, 1.5]$. Compute eleven approximations (i.e., use $n = 10$) to the root.

(b) Evaluate the number of computations $N$ required to ensure that the error is less than the preassigned value (error bound) $\epsilon = 0.001$.

Solution:

(a) We start with initial interval $[a_0, b_0] = [0.5, 1.5]$ and compute $f(0.5) = -0.76028723$ and $f(1.5) = 0.49624248$. We notice that $f(a_0)$ and $f(b_0)$ have opposite signs and hence, there is a root in the interval $[0.5, 1.5]$. Compute the midpoint $c_0 = \frac{a_0 + b_0}{2} = \frac{0.5 + 1.5}{2} = 1$ and $f(1) = -0.15852902$. The function changes sign on $[c_0, b_0] = [1, 1.5]$, so, we set $[a_1, b_1] = [c_0, b_0] = [1, 1.5]$, and compute the midpoint $c_1 = \frac{a_1 + b_1}{2} = \frac{1 + 1.5}{2} = 1.25$ and $f(1.25) = 0.18623077$. Hence, the root lies in the interval $[a_1, c_1] = [1, 1.25]$. Set $[a_2, b_2] = [a_1, c_1] = [1, 1.25]$ and continue until we compute $c_{10} = 1.11376953125$. The details are explained in Table 2.1.
CHAPTER 2. NUMERICAL SOLUTIONS OF NONLINEAR EQUATIONS

Table 2.1: Bisection Method Solution of Example 9

<table>
<thead>
<tr>
<th>n</th>
<th>Left Endpoint (a_n)</th>
<th>Midpoint (c_n)</th>
<th>Right Endpoint (b_n)</th>
<th>Function Value (f(c_n))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>1.5</td>
<td>-0.15852902</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.25</td>
<td>1.5</td>
<td>0.18623077</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1.125</td>
<td>1.25</td>
<td>0.01505104</td>
</tr>
<tr>
<td>3</td>
<td>1.0625</td>
<td>1.09375</td>
<td>1.125</td>
<td>-0.07182663</td>
</tr>
<tr>
<td>4</td>
<td>1.09375</td>
<td>1.109375</td>
<td>1.125</td>
<td>-0.00664277</td>
</tr>
<tr>
<td>5</td>
<td>1.109375</td>
<td>1.1171875</td>
<td>1.125</td>
<td>0.00420803</td>
</tr>
<tr>
<td>6</td>
<td>1.109375</td>
<td>1.11328125</td>
<td>1.1171875</td>
<td>-0.00121649</td>
</tr>
<tr>
<td>7</td>
<td>1.11328125</td>
<td>1.115234375</td>
<td>1.1171875</td>
<td>0.00149600</td>
</tr>
<tr>
<td>8</td>
<td>1.11328125</td>
<td>1.1142578125</td>
<td>1.115234375</td>
<td>0.00013981</td>
</tr>
<tr>
<td>9</td>
<td>1.11328125</td>
<td>1.11376953125</td>
<td>1.1142578125</td>
<td>-0.00053832</td>
</tr>
<tr>
<td>10</td>
<td>1.11328125</td>
<td>1.11376953125</td>
<td>1.1142578125</td>
<td></td>
</tr>
</tbody>
</table>

(b)

\[
N = \int \left( \frac{\ln(1.5 - 0.5) - \ln(0.001)}{\ln(2)} \right) = \int \left( \frac{\ln(1) - \ln(0.001)}{\ln(2)} \right) = \int \left( \frac{0 - (-6.90775528)}{0.69314718} \right) = \int \left( \frac{6.90775528}{0.69314718} \right) = \int (9.96578429) = 10.
\]

2.3 False-Position Method

It also known as regula falsi method, it is similar to the bisection method in requiring that \(f(a)\) and \(f(b)\) have opposite signs. This method uses the abscissa of the point \((c, 0)\) at which the secant line called it SL joining the points \((a, f(a))\) and \((b, f(b))\) cross the x-axis instead of using the midpoint of the interval as approximation of the zero of the function \(f\) as in the bisection method. To evaluate \(c\), we need to compute the slope of line SL between the two points \((a, f(a))\) and \((b, f(b))\):

\[
m = \frac{f(b) - f(a)}{b - a}.
\]

Now, compute the slope of line SL between the two points \((c, f(c)) = (c, 0)\) and \((b, f(b))\):

\[
m = \frac{f(b) - f(c)}{b - c} = \frac{f(b) - 0}{b - c} = \frac{f(b)}{b - c}.
\]
By equating the two slopes, we obtain
\[
\frac{f(b) - f(a)}{b - a} = \frac{f(b) - f(c)}{b - c} \quad \Rightarrow \quad c = b - \frac{f(b)(b - a)}{f(b) - f(a)}.
\]

Now, we have the same possibilities as in the bisection method:

- If \( f(c_0) = 0 \) then the root is \( r = c_0 \).
- If \( f(c_0) \neq 0 \) then either \( f(a_0)f(c_0) < 0 \) or \( f(c_0)f(b_0) < 0 \).
  
  (i). If \( f(a_0)f(c_0) < 0 \) then \( r \) lies in \([a_0, c_0]\), and set \( a_1 = a_0 \) and \( b_1 = c_0 \), i.e. \([a_1, b_1] = [a_0, c_0]\).
  
  (ii). If \( f(c_0)f(b_0) < 0 \) then \( r \) lies in \([c_0, b_0]\), and set \( a_1 = c_0 \) and \( b_1 = b_0 \), i.e. \([a_1, b_1] = [c_0, b_0]\).
  
  (iii). Compute \( c_1 = b_1 - \frac{f(b_1)(b_1-a_1)}{f(b_1)-f(a_1)} \).
  
  (iv). Then, we proceed in this way until we reach the \( n \)th interval \([a_n, b_n]\) and then compute \( c_n = b_n - \frac{f(b_n)(b_n-a_n)}{f(b_n)-f(a_n)} \).

Example 10. Show that \( f(x) = 2x^3 - x^2 + x - 1 = 0 \) has at least one root in \([0, 1]\).

Solution:
Since \( f(0) = -1 \) and \( f(1) = 1 \), then Intermediate Value Theorem implies that this continuous function has a root in \([0, 1]\). Set \([a_0, b_0] = [0, 1]\) and compute \( c_0 = b_0 - \frac{f(b_0)(b_0-a_0)}{f(b_0)-f(a_0)} = f(1) - \frac{f(1)(1-0)}{f(1)-f(0)} = 1 - \frac{1}{1} = 0.5 \), also compute \( f(c_0) = f(0.5) = -0.5 \). Hence, the root lies in \([c_0, b_0]\) we squeeze from the left and set \( a_1 = c_0 = 0.5 \) and \( b_1 = b_0 = 1 \), to have \([a_1, b_1] = [0.5, 1]\). Now, compute the new approximation to the root \( c_1 = b_1 - \frac{f(b_1)(b_1-a_1)}{f(b_1)-f(a_1)} = f(1) - \frac{f(1)(1-0.5)}{f(1)-f(0.5)} = 1 - \frac{1}{1} = 2/3 \approx 0.66666667 \), \( f(c_1) = -0.18518519 \). The function has opposite signs on the interval \([c_1, b_1]\), set \( a_2 = c_1 = 0.6666667 \) and \( b_2 = b_1 = 1 \), so we have \([a_2, b_2] = [0.6666667, 1]\). Continue this we and stop at \( c_7 = 0.73895443 \).

<table>
<thead>
<tr>
<th>( n )</th>
<th>( a_n )</th>
<th>( c_n )</th>
<th>( b_n )</th>
<th>( f(c_n) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>-0.5</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>0.6666667</td>
<td>1</td>
<td>-0.18518519</td>
</tr>
<tr>
<td>2</td>
<td>0.6666667</td>
<td>0.7187500</td>
<td>1</td>
<td>-0.05523681</td>
</tr>
<tr>
<td>3</td>
<td>0.7187500</td>
<td>0.73347215</td>
<td>1</td>
<td>-0.01532051</td>
</tr>
<tr>
<td>4</td>
<td>0.73347215</td>
<td>0.73749388</td>
<td>1</td>
<td>-0.00416160</td>
</tr>
<tr>
<td>5</td>
<td>0.73749388</td>
<td>0.73858180</td>
<td>1</td>
<td>-0.00112399</td>
</tr>
<tr>
<td>6</td>
<td>0.73858180</td>
<td>0.73887530</td>
<td>1</td>
<td>-0.00030311</td>
</tr>
<tr>
<td>7</td>
<td>0.73887530</td>
<td>0.73895443</td>
<td>1</td>
<td>-0.00008171</td>
</tr>
</tbody>
</table>
2.4 Newton’s Method

Newton’s method is also known as Newton-Raphson method is one of the most powerful and efficient numerical methods for root-finding problems. It is well-known and popular method and there are several variants and extensions of this method. There are more than one approach for deriving this method such as the graphical technique and Taylor series technique, and here we will use the latter. Let \( f, f', f'' \) are continuous functions on the interval \([a,b]\) (i.e. \( f \in C^2[a,b] \)). Let \( r_0 \in [a,b] \) be an approximation to the zero \( r \) of the function \( f \) such that \( f'(r_0) \neq 0 \) and \( r_0 \) is “sufficiently close to \( r \) i.e. \( |r - r_0| \) is relatively small”. Let’s start with first Taylor polynomial of \( f(x) \) expanded about the initial approximation \( r_0 \) and compute it at \( x = r \):

\[
f(r) = f(r_0) + (r - r_0)f'(r_0) + \frac{(r - r_0)^2}{2}f''(\xi(r)),
\]

where \( \xi(r) \) lies between \( r_0 \) and \( r \). Using the fact that \( f(r) = 0 \), this leads to

\[
0 = f(r_0) + (r - r_0)f'(r_0) + \frac{(r - r_0)^2}{2}f''(\xi(r)).
\]

Since \( |r - r_0| \) is small then \( (r - r_0)^2 \) is much smaller, so we can neglect the third term in Taylor’s expansion which contains this term (quadratic power term) to have

\[
0 = f(r_0) + (r - r_0)f'(r_0).
\]

Solving for \( r \) yields

\[
r = r_0 - \frac{f(r_0)}{f'(r_0)}.
\]

To, proceed, set \( r = r_1 \) in the Newton’s formula to compute \( r_1 \) by using the known value \( r_0 \)

\[
r_1 = r_0 - \frac{f(r_0)}{f'(r_0)},
\]

and then we compute \( r_2 \) using the known value \( r_1 \)

\[
r_2 = r_1 - \frac{f(r_1)}{f'(r_1)},
\]

and by following the same fashion, we compute \( r_3, r_4 \) and so on. The general or \( n \)th form of Newton’s method is:

\[
r_n = r_{n-1} - \frac{f(r_{n-1})}{f'(r_{n-1})}, \quad n = 1, 2, \ldots.
\]
CHAPTER 2. NUMERICAL SOLUTIONS OF NONLINEAR EQUATIONS

This is called Newton’s formula or Newton-Raphson formula.

Example 11. Use Newton’s method to find the positive root accurate to within $10^{-5}$ for $f(x) = 3x - e^x = 0$. Start with the initial guess $r_0 = 1.5$.

Solution: Start by finding the derivative of $f(x)$:

$$f(x) = 3x - e^x, \quad f'(x) = 3 - e^x, \quad r_0 = 1.5, \quad f(r_0) = 0.01831093, \quad f'(r_0) = -1.48168907,$$

so, the Newton-Raphson iteration formula for this problem is:

$$r_n = r_{n-1} - \frac{f(r_{n-1})}{f'(r_{n-1})} = r_{n-1} - \frac{3r_{n-1} - e^{r_{n-1}}}{3 - e^{r_{n-1}}}, \quad n = 1, 2, \ldots ,$$

Computing $r_1$ by using the known value $r_0$,

$$r_1 = r_0 - \frac{f(r_0)}{f'(r_0)} = r_0 - \frac{3r_0 - e^{r_0}}{3 - e^{r_0}} = 1.5 - \frac{3(1.5) - e^{1.5}}{3 - e^{1.5}} = 1.51235815.$$

Now, compute $f(r_1)$ and $f'(r_1)$,

$$f(r_1) = 3r_1 - e^{r_1} = -0.00034364, \quad f'(r_1) = 3 - e^{r_1} = -1.53741808.$$

Next, we compute $r_2$,

$$r_2 = r_1 - \frac{f(r_1)}{f'(r_1)} = r_1 - \frac{3r_1 - e^{r_1}}{3 - e^{r_1}} = 1.51213463, \quad f(r_2) = -1.1e-07, \quad f'(r_2) = -1.53640399.$$

A summary of the computations is given in Table 2.3.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$r_n$</th>
<th>$f(r_n)$</th>
<th>$f'(r_n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.50000000</td>
<td>0.01831093</td>
<td>-1.48168907</td>
</tr>
<tr>
<td>1</td>
<td>1.51235815</td>
<td>-0.00034364</td>
<td>-1.53741808</td>
</tr>
<tr>
<td>2</td>
<td>1.51213463</td>
<td>-0.00000011</td>
<td>-1.53640399</td>
</tr>
<tr>
<td>3</td>
<td>1.51213455</td>
<td>-0.00000000</td>
<td>-1.53640365</td>
</tr>
<tr>
<td>4</td>
<td>1.51213455</td>
<td>0.00000000</td>
<td>-1.53640365</td>
</tr>
</tbody>
</table>

Table 2.3: Newton’s Method Solution of Example 11

Newton’s Method for Finding the $n$th Roots

We start with square roots. Let $B > 0$ a real number and $r_0$ be an initial approximation to $\sqrt{B}$. Our goal is to find a square root of a number $B$. Let $x = \sqrt{B}$, so $x^2 = B$, which implies that $x^2 - B = 0$, define $f(x) = x^2 - B = 0$.  

Mohammad Sabawi/Numerical Analysis
Note that this equation has two roots \( x = \pm \sqrt{B} \). Now, find the derivative of \( f \), \( f'(x) = 2x \) and use the Newton’s fixed point formula

\[
x = g(x) = x - \frac{f(x)}{f'(x)} = x - \frac{x^2 - B}{2x} = \frac{x^2 + B}{2x} = \frac{x + \frac{B}{x}}{2}.
\]

Now, using Newton’s iteration formula

\[
r_{n+1} = \frac{r_n + \frac{B}{r_n}}{2}, \quad n = 0, 1, \cdots.
\]

The sequence of iterations \( \{r_n\}_{n=0}^{\infty} \) converges to \( \sqrt{B} \). Note that in computing the square root of \( B \), we do not need to evaluate \( f \) and \( f' \) and this makes the calculations easier and faster since we just need the values of the iterates \( r_n, n = 0, 1, \cdots \).

**Example 12.** Use Newton’s square-root algorithm to find \( \sqrt{3} \), use \( r_0 = 1 \).

**Solution:** Starting with \( r_0 = 1 \) when \( n = 0 \), we have

\[
r_1 = \frac{r_0 + \frac{3}{r_0}}{2} = \frac{1 + 3}{2} = 2.
\]

For \( n = 1 \),

\[
r_2 = \frac{r_1 + \frac{3}{r_1}}{2} = \frac{2 + \frac{3}{2}}{2} = 1.75.
\]

\( n = 2, r_3 = \frac{r_2 + \frac{3}{r_2}}{2} = \frac{1.75 + \frac{3}{1.75}}{2} = 1.732142857142857 \).

\( n = 3, r_4 = \frac{r_3 + \frac{3}{r_3}}{2} = 1.732050810014727 \).

A summary of results is given in Table 2.4

<table>
<thead>
<tr>
<th>( n )</th>
<th>( r_n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>1.75</td>
</tr>
<tr>
<td>3</td>
<td>1.732142857142857</td>
</tr>
<tr>
<td>4</td>
<td>1.732050810014727</td>
</tr>
<tr>
<td>5</td>
<td>1.732050807568877</td>
</tr>
<tr>
<td>6</td>
<td>1.732050807568877</td>
</tr>
</tbody>
</table>

**Table 2.4:** Newton’s Method Solution of Example 12
2.5 Numerical Solutions of Systems of Nonlinear Equations

Some phenomena in nature are modelled by systems of $N$ nonlinear equations in $N$ unknowns. These systems can be handled firstly by linearising them and then solving them in repeated way. Newton’s method for a single nonlinear equation follows the same approach and can be easily extended for solving a system of nonlinear equations.

The general form of a system of $N$ nonlinear equations in $N$ unknowns $x_i$ is:

\[
\begin{align*}
    f_1(x_1, x_2, \cdots, x_N) &= 0 \\
    f_2(x_1, x_2, \cdots, x_N) &= 0 \\
    & \vdots \\
    f_N(x_1, x_2, \cdots, x_N) &= 0.
\end{align*}
\]

Using vector notation, this system can be written in this concise form

\[
F(X) = 0,
\]

where

\[
F = [f_1, f_2, \cdots, f_N]^T, \\
X = [x_1, x_2, \cdots, x_N]^T.
\]

Newton’s formula for a single nonlinear equation can be extended to a system of nonlinear equations in the following form

\[
X^{(k+1)} = X^{(k)} - [F'(X^{(k)})]^{-1}F(X^{(k)}),
\]

where $F'(X^{(k)})$ is the Jacobian matrix which will be defined below. It contains the partial derivatives of $F$ evaluated at $X^{(k)} = [x_1^{(k)}, x_2^{(k)}, \cdots, x_N^{(k)}]^T$. The above-mentioned formula is similar to the Newton’s formula for a single nonlinear equation except that the derivative appeared in the numerator as an inverse of the Jacobian matrix. In practice, the inverse will not be computed since this is impractical because its computational cost and instead we will solve a related linear system.
The method will explained by solving a system of three nonlinear equations

\[
\begin{align*}
    f_1(x_1, x_2, x_3) &= 0 \\
    f_2(x_1, x_2, x_3) &= 0 \\
    f_3(x_1, x_2, x_3) &= 0.
\end{align*}
\]

The Taylor series expansion in three variables \(x_1, x_2, x_3\):

\[
f_i(x_1 + h_1, x_2 + h_2, x_3 + h_3) = f_i(x_1, x_2, x_3) + h_1 \frac{\partial f_i}{\partial x_1} + h_2 \frac{\partial f_i}{\partial x_2} + h_3 \frac{\partial f_i}{\partial x_3} + \cdots,
\]

where the partial derivatives are evaluated at the point \((x_1, x_2, x_3)\). We consider just the linear terms in step sizes \(h_i\) for \(i = 1, 2, 3\). Assume that we have in vector notation

\[
0 \approx F(X^{(0)} + H^{(0)}) \approx F(X^{(0)}) + F'(X^{(0)})H^{(0)},
\]

where \(F'(X^{(0)})\) is the Jacobian matrix at the initial guess \(X^{(0)} = (x_1^0, x_2^0, x_3^0)\),

\[
F'(X^{(0)}) = \begin{bmatrix}
\frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} \\
\frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} \\
\frac{\partial f_3}{\partial x_1} & \frac{\partial f_3}{\partial x_2} & \frac{\partial f_3}{\partial x_3}
\end{bmatrix},
\]

where the partial derivatives are evaluated as follows:

\[
\frac{\partial f_i}{\partial x_j} = \frac{\partial f_i(X^{(0)})}{\partial x_j}, i, j = 1, 2, 3.
\]

If the Jacobian matrix \(F'(X^{(0)})\) is non singular i.e its inverse is existed, then solving for \(H\), we find

\[
H^{(0)} = -[F'(X^{(0)})]^{-1}F(X^{(0)}).
\]

The next iteration after correction \(X^{(1)} = X^{(0)} + H^{(0)}\) is closer to the root than \(X^{(0)}\). Hence, Newton’s formula for the first iteration is

\[
X^{(1)} = X^{(0)} - [F'(X^{(0)})]^{-1}F(X^{(0))}.
\]
Consequently, the general form of Newton’s method for solving the nonlinear system is:

\[ X^{(k+1)} = X^{(k)} - [F'(X^{(k)})]^{-1}F(X^{(k)}), k = 0, 1, \ldots . \]

To avoid computing the inverse of the Jacobian matrix at each iteration, we instead resort to solving the Jacobian linear systems

\[ [F'(X^{(k)})]H^{(k)} = -F(X^{(k)}), k = 0, 1, \ldots , \]

Hence, the next Newton’s iteration is computed using the formula

\[ X^{(k+1)} = X^{(k)} + H^{(k)}, k = 0, 1, \ldots . \]

**Example 13.** Solve the following nonlinear system using Newton’s method. Start with the initial guess \( X^{(0)} = (x_1^0 = 1, x_2^0 = 0, x_3^0 = 0) \). The exact solution to this system is \( X = (x_1 = 0, x_2 = 1, x_3 = 1) \).

\[
\begin{align*}
x_1 + x_2 + x_3 &= 2 \\
x_1^2 + x_2^2 + x_3^2 &= 2 \\
e^{x_1} + x_1x_2 - x_1x_3 &= 1.
\end{align*}
\]

**Solution:** We compute the Jacobian matrix

\[
F'(X) = \begin{bmatrix}
\frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} \\
\frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} \\
\frac{\partial f_3}{\partial x_1} & \frac{\partial f_3}{\partial x_2} & \frac{\partial f_3}{\partial x_3}
\end{bmatrix} = \begin{bmatrix}
1 & 1 & 1 \\
2x_1 & 2x_2 & 2x_3 \\
e^{x_1} + x_2 - x_3 & x_1 & -x_1
\end{bmatrix}.
\]

The Jacobian matrix at the initial guess \( X^{(0)} \)

\[
F'(X^{(0)}) = \begin{bmatrix}
\frac{\partial f_1}{\partial x_1}^{(0)} & \frac{\partial f_1}{\partial x_2}^{(0)} & \frac{\partial f_1}{\partial x_3}^{(0)} \\
\frac{\partial f_2}{\partial x_1}^{(0)} & \frac{\partial f_2}{\partial x_2}^{(0)} & \frac{\partial f_2}{\partial x_3}^{(0)} \\
\frac{\partial f_3}{\partial x_1}^{(0)} & \frac{\partial f_3}{\partial x_2}^{(0)} & \frac{\partial f_3}{\partial x_3}^{(0)}
\end{bmatrix} = \begin{bmatrix}
1 & 1 & 1 \\
2x_1^{(0)} & 2x_2^{(0)} & 2x_3^{(0)} \\
e^{x_1^{(0)}} + x_2^{(0)} - x_3^{(0)} & x_1^{(0)} & -x_1^{(0)}
\end{bmatrix} = \begin{bmatrix}
1 & 1 & 1 \\
2 & 0 & 0 \\
e^1 & 1 & -1
\end{bmatrix}.
\]

Now, solving the Jacobian linear system for \( H^{(0)} \)

\[ [F'(X^{(0)})]H^{(0)} = -F(X^{(0)}), \]
we get
\[
\begin{bmatrix}
1 & 1 & 1 \\
2 & 0 & 0 \\
e^1 & 1 & -1
\end{bmatrix}
\begin{bmatrix}
h_1^0 \\
h_2^0 \\
h_3^0
\end{bmatrix} = 
\begin{bmatrix}
-1 \\
-1 \\
e^1 - 1
\end{bmatrix},
\]
implies that
\[
H^{(0)} = 
\begin{bmatrix}
0.5000 \\
-1.2887 \\
1.7887
\end{bmatrix}.
\]
Hence,
\[
X^{(1)} = X^{(0)} + H^{(0)} = 
\begin{bmatrix}
1 \\
0 \\
0
\end{bmatrix} + 
\begin{bmatrix}
0.5000 \\
-1.2887 \\
1.7887
\end{bmatrix} = 
\begin{bmatrix}
1.5000 \\
-1.2887 \\
1.7887
\end{bmatrix}.
\]
Now, compute the Jacobian matrix for \(X^{(1)}\).
\[
F'(X^{(1)}) = 
\begin{bmatrix}
1 & 2x_1^1 & 1 \\
e^{e^1} + x_2^{(1)} - x_3^{(1)} - x_1^{(1)} & 2x_2^1 & 2x_3^1
\end{bmatrix} = 
\begin{bmatrix}
1.0000 & 1.0000 & 1.0000 \\
3.0000 & -2.5774 & 3.5774 \\
1.4043 & 1.5000 & -1.5000
\end{bmatrix}.
\]
\[
F(X^{(1)}) = 
\begin{bmatrix}
f_1(x_1, x_2, x_3) \\
f_2(x_1, x_2, x_3) \\
-f_3(x_1, x_2, x_3)
\end{bmatrix} = 
\begin{bmatrix}
0 \\
5.1102 \\
-1.1344
\end{bmatrix}.
\]
Solving for \(H^{(1)}\), we have
\[
\begin{bmatrix}
1.0000 & 1.0000 & 1.0000 \\
3.0000 & -2.5774 & 3.5774 \\
1.4043 & 1.5000 & -1.5000
\end{bmatrix}
\begin{bmatrix}
h_1^1 \\
h_2^1 \\
h_3^1
\end{bmatrix} = 
\begin{bmatrix}
0 \\
5.1102 \\
-1.1344
\end{bmatrix},
\]
implies
\[
H^{(1)} = 
\begin{bmatrix}
-0.5172 \\
0.8788 \\
-0.3616
\end{bmatrix}.
\]
CHAPTER 2. NUMERICAL SOLUTIONS OF NONLINEAR EQUATIONS

The next approximation is

\[ X^{(2)} = X^{(1)} + H^{(1)} = \begin{bmatrix} 1.5000 \\ -1.2887 \\ 1.7887 \end{bmatrix} + \begin{bmatrix} -0.5172 \\ 0.8788 \\ -0.3616 \end{bmatrix} = \begin{bmatrix} 0.9828 \\ -0.4099 \\ 1.4271 \end{bmatrix}. \]

Continuing as before, until we reach the required results.

2.6 Secant Method

Newton’s method is a very powerful and efficient technique for solving root-finding problems but one of the drawbacks of the method is the need of derivative evaluations of \( f \) at the approximations \( r_n, n \geq 0 \), and this is not a trivial task. To avoid this we introduce secant method which is a variation of Newton’s method. Secant method is similar to the false position method but it differs in the way of choosing the succeeding terms. We start with two initial points \( (r_0, f(r_0)) \) and \( (r_1, f(r_1)) \) near the point \( (r, 0) \), where \( r \) is the root of equation \( f(x) = 0 \). Define the point \( (r_2, 0) \) to be the point of intersection of the secant line joining the points \( (r_0, f(r_0)) \) and \( (r_1, f(r_1)) \) with the \( x \)-axis. Geometrically, the abscissa of the point of intersection \( r_2 \) is closer to the root \( r \) than to either \( r_0 \) and \( r_1 \).

The slope of the secant line relating these three points \( (r_0, f(r_0)), (r_1, f(r_1)) \) and \( (r_2, f(r_2)) \) is:

\[ m = \frac{f(r_1) - f(r_0)}{r_1 - r_0} \quad \text{and} \quad m = \frac{f(r_2) - f(r_1)}{r_2 - r_1} = \frac{0 - f(r_1)}{r_2 - r_1} = \frac{-f(r_1)}{r_2 - r_1}. \]

Equating the two values of the slope, we have

\[ \frac{f(r_1) - f(r_0)}{r_1 - r_0} = \frac{-f(r_1)}{r_2 - r_1}. \]

Solving slope’s equation for \( r_2 \), we obtain

\[ r_2 = r_1 - \frac{f(r_1)(r_1 - r_0)}{f(r_1) - f(r_0)}. \]

So, the general form of the secant method is:

\[ r_{n+2} = r_{n+1} - \frac{f(r_{n+1})(r_{n+1} - r_n)}{f(r_{n+1}) - f(r_n)}, \quad n = 0, 1, \ldots. \]
Example 14. Find the root of equation $x - \cos(x) = 0$ using the secant method and the two initial guesses $r_0 = 0.5$ and $r_1 = 0.6$.

Solution: To compute the first approximation $r_2$, we need to compute $f(r_0)$ and $f(r_1)$

$$f(r_0) = f(0.5) = 0.5 - \cos(0.5) = -0.377582560000000,$$
$$f(r_1) = f(0.6) = 0.6 - \cos(0.6) = -0.225335610000000.$$

So,

$$r_2 = r_1 - \frac{f(r_1)(r_1 - r_0)}{f(r_1) - f(r_0)} = 0.6 - \frac{f(0.6)(0.6 - 0.5)}{f(0.6) - f(0.5)} = 0.748006655882730,$$
$$f(r_2) = r_2 - \cos(r_2) = 0.014960500949714.$$

Now, we compute the next approximation $r_3$,

$$r_3 = r_2 - \frac{f(r_2)(r_2 - r_1)}{f(r_2) - f(r_1)} = 0.738791967963291, f(r_3) = -0.000490613128583.$$

Continuing until satisfying the required accuracy. A summary of the calculations is given in Table 2.5.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$r_n$</th>
<th>$f(r_n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.500000000000000</td>
<td>-0.377582560000000</td>
</tr>
<tr>
<td>1</td>
<td>0.600000000000000</td>
<td>-0.225335610000000</td>
</tr>
<tr>
<td>2</td>
<td>0.748006655882730</td>
<td>0.014960500949714</td>
</tr>
<tr>
<td>3</td>
<td>0.738791967963291</td>
<td>-0.000490613128583</td>
</tr>
<tr>
<td>4</td>
<td>0.739084558312839</td>
<td>-0.000000962163319</td>
</tr>
<tr>
<td>5</td>
<td>0.739085133252381</td>
<td>0.000000000062293</td>
</tr>
<tr>
<td>6</td>
<td>0.739085133215161</td>
<td>0.000000000000000</td>
</tr>
<tr>
<td>7</td>
<td>0.739085133215161</td>
<td>0.000000000000000</td>
</tr>
</tbody>
</table>

Table 2.5: Secant Method Solution of Example 14

2.7 Fixed Points and Functional Iteration

Iteration is a fundamental concept in computer sciences and numerical analysis and is used for solving a wide variety of problems. There is a strong connection between root-finding problems and fixed point problems, and in this section, we use fixed point problems to solve root-finding problems.
Definition 17 (Fixed Point). The number \( p \) is called a **fixed point** of the function \( g \) if \( p = g(p) \).

We start by transforming the root-finding problem \( f(x) = 0 \) to a fixed point problem \( x = g(x) \) by algebraic manipulations. There are more than one way of rearranging \( f(x) = 0 \) into an equivalent form \( x = g(x) \). Note that if \( r \) is a zero of the function \( f \) (i.e. \( r \) is a root of the equation \( f(r) = 0 \)) then \( r = g(r) \) i.e. \( r \) is a fixed point of the function \( g \). Conversely, if \( g \) has a fixed point at \( r \) then the function \( f(x) = x - g(x) \) has a zero at \( r \). Geometrically, the fixed points of a function \( y = g(x) \) are the points of intersection of its curve with the straight line \( y = x \).

Example 15. Find the fixed points of the function \( g(x) = 2 - x^2 \) and verify that they are the solutions to the equation \( f(x) = x - g(x) = 0 \).

Solution: The fixed points of \( g \) are the points satisfying the fixed point equation \( x = g(x) \), so intersect the graph of \( y = g(x) \) with the graph of the straight line \( y = x \)

\[
x = g(x) = 2 - x^2,
\]

which implies that

\[
-x^2 - x + 2 = -(x^2 + x - 2) = -(x - 1)(x + 2) = 0.
\]

So, either \( x - 1 = 0 \) implies \( x = 1 \) or \( x + 2 = 0 \) implies \( x = -2 \). Hence, the fixed points are \( x = 1 \) and \( x = -2 \). We notice that these fixed points are the same the zeros of \( f(x) = x - g(x) = -(x^2 + x - 2) = -(x - 1)(x + 2) = 0 \).

Definition 18 (Fixed Point Iteration). The iteration \( r_{n+1} = g(r_n) \), \( n = 0, 1, \ldots \), obtained by using fixed point formula \( r = g(r) \) is called a **fixed point iteration** or **functional iteration**.

Theorem 7. Assume that \( g \in C[a, b] \) such that

1. If \( g(x) \in [a, b] \) for all \( x \in [a, b] \), then \( g \) has a at least one fixed point \( r \) in \( [a, b] \).

2. If also, \( g'(x) \) existed and defined on \((a, b)\) and there exists a positive constant \( K < 1 \) such that \( |g'(x)| \leq K < 1 \), for all \( x \in (a, b) \), then \( g \) has a unique fixed point \( r \) in \([a, b] \).

Theorem 8 (Fixed Point Theorem). Assume that \( g \in C[a, b] \) such that

1. \( g(x) \in [a, b] \) for all \( x \in [a, b] \), then \( g \) has a fixed point \( r \) in \([a, b] \).
CHAPTER 2. NUMERICAL SOLUTIONS OF NONLINEAR EQUATIONS

2. \( g'(x) \) existed and defined over \((a,b)\) and there exists a positive constant \( K < 1 \) such that \( |g'(x)| \leq K < 1 \), for all \( x \in (a,b) \), then \( g \) has a unique fixed point \( r \) in \([a,b]\).

3. If \( g \) satisfies the conditions (1) and (2), then for any number \( r_0 \) in \([a,b]\) the sequence \( \{r_n\}_{n=0}^{\infty} \) of iterations generated by fixed point iteration \( r_{n+1} = g(r_n) \), \( n = 0,1,\ldots \), converges to the unique fixed point \( r \) in \([a,b]\).

Corollary 1. If \( g \) satisfies the hypotheses of Fixed Point Theorem, then the error bounds for approximating \( r \) using \( r_n \) are given by

\[
|r - r_n| \leq K^n \max |r - r_0|,
\]

and

\[
|r - r_n| \leq \frac{K^n}{1 - K} \max |r_1 - r_0|, \text{ for all } n \geq 1.
\]

Example 16. Use the fixed point method to find the zero of the function \( f(x) = x^3 - 3x^2 + 2 \) in \([0,2]\), start with \( r_0 = 1.5 \).

Solution: There are many possibilities to write \( f(x) = 0 \) as a fixed point form \( x = g(x) \) using mathematical manipulations.

(1) \( x = g_1(x) = x - x^3 + 3x^2 - 2 \).

(2) \( x = g_2(x) = (\frac{x^3 + 2}{3})^{1/2} \).

(3) \( x = g_3(x) = -(\frac{x^2 + 2}{3})^{1/2} \).

(4) \( x = g_4(x) = (\frac{2}{3-x})^{1/2} \).

(5) \( x = g_5(x) = \frac{-2}{x(x-3)} \).

(6) \( x = g_6(x) = \left(3x^2 - 2\right)^{1/3} \).

(7) \( x = g_7(x) = \left(3x - \frac{3}{2}\right)^{1/2} \).

For example, to obtain \( g_1(x) \) just add \( x \) to both sides of the equation \(-f(x) = 0\) and this is the simplest way to write the problem as a fixed point form

\[-f(x) = 0, \quad -x^3 + 3x^2 - 2 = 0, \quad \text{so} \quad x = x - x^3 + 3x^2 - 2 = g_1(x).\]

Also, \( g_2(x) \) and \( g_3(x) \) can be obtained as follows:

\[x^3 - 3x^2 + 2 = 0, \quad \text{so} \quad 3x^2 = x^3 + 2, \quad \text{and} \quad x^2 = \frac{x^3 + 2}{3}.\]
implies that

\[ x = \pm \left( \frac{x^3 + 2}{3} \right)^{1/2}, \quad \text{so} \quad g_2(x) = \left( \frac{x^3 + 2}{3} \right)^{1/2}, \quad \text{and} \quad g_3(x) = -\left( \frac{x^3 + 2}{3} \right)^{1/2}. \]

Note that it is important to check that the fixed point of each derived function \(g\) is a solution to the problem \(f(x) = 0\). For example, because the solution is positive and lies between 0 and 2, so we choose the positive function \(g_2(x)\), since the negative function \(g_3(x)\) is not a choice here. The results are outlined in Tables 2.6 and 2.7 below.

<table>
<thead>
<tr>
<th>(n)</th>
<th>(g_1(x))</th>
<th>(g_2(x))</th>
<th>(g_3(x))</th>
<th>(g_4(x))</th>
<th>(g_5(x))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>1</td>
<td>2.875</td>
<td>1.33853153</td>
<td>-1.33853153</td>
<td>1.15470054</td>
<td>0.88888889</td>
</tr>
<tr>
<td>2</td>
<td>1.90820313</td>
<td>1.21081272</td>
<td>0 - 0.36i</td>
<td>1.04107393</td>
<td>1.06578947</td>
</tr>
<tr>
<td>3</td>
<td>3.88369668</td>
<td>1.12177435</td>
<td>1.01042940</td>
<td>0.97018561</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-11.44518863</td>
<td>1.06639827</td>
<td>1.00261759</td>
<td>1.01559099</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.8788e+03</td>
<td>1.03484519</td>
<td>1.00065504</td>
<td>0.99238449</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-6.6210e+09</td>
<td>1.01787695</td>
<td>1.00016380</td>
<td>1.00385153</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2.9025e+29</td>
<td>1.000905819</td>
<td>1.00004095</td>
<td>0.99808533</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1.00455985</td>
<td>1.00001024</td>
<td>1.00096009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1.00228772</td>
<td>1.00000256</td>
<td>0.99952065</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1.00114582</td>
<td>1.00000644</td>
<td>1.00023985</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1.00057340</td>
<td>1.0000016</td>
<td>0.99988012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1.00000004</td>
<td>1.000005995</td>
<td>0.99997003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>1.00000001</td>
<td>1.00001499</td>
<td>0.99999251</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1.00000000</td>
<td>1.00000375</td>
<td>0.99999813</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>1.00000000</td>
<td>0.99999813</td>
<td>1.00000094</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>1.00000000</td>
<td>0.99999953</td>
<td>1.00000024</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>1.00000000</td>
<td>0.99999988</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.6: Fixed Point Method Solution of Example 16
2.8 Acceleration of Iterative Methods

The linear convergence a sequence \( \{r_n\} \) to the limit \( r \) such as the sequences of the fixed point iterations can be accelerated by using some techniques such as Aitken’s \( \Delta^2 \) method (Aitken’s acceleration) and Steffensen’s method. For more details see references [4, 15] and the references therein.

2.8.1 Modified Newton’s Methods

Newton’s method is a fixed point method since it can be written as

\[
x = g(x) = x - \frac{f(x)}{f'(x)}.
\]
and in iterative way
\[ r_n = g(r_{n-1}) = r_{n-1} - \frac{f(r_{n-1})}{f'(r_{n-1})}, \quad n = 1, 2, \ldots, \]
and this is called Newton-Raphson iteration formula or simply Newton’s iteration. The convergence of Newton’s method can be modified to accelerate its rate of convergence at the root \( x = r \) of order \( M > 1 \)
\[ r_n = r_{n-1} - \frac{f(r_{n-1}) f'(r_{n-1})}{(f'(r_{n-1}))^2 - f(r_{n-1}) f''(r_{n-1})}, \quad n = 1, 2, \ldots. \]
This formula is called a modified Newton’s method.

Also, Newton’s method can be accelerated in an another way.

**Theorem 9 (Acceleration of Newton’s Iteration).** Assume that Newton’s method produces a linearly convergent sequence to the root \( x = r \) of order \( M > 1 \). Then Newton’s iteration formula
\[ r_n = r_{n-1} - \frac{M f(r_{n-1})}{f'(r_{n-1})}, \quad n = 1, 2, \ldots, \]
produces a quadratically convergent sequence \( \{r_n\}_{n=0}^{\infty} \) to the root \( x = r \).

**Example 17.** Show that \( r = 1 \) is a double zero (double root) of \( f(x) = -x^3 + 3x - 2 = 0 \). Start with \( r_0 = 1.25 \) as an initial guess of \( r \) and compare the performance of Newton’s method and accelerated Newton’s method for solving \( f(x) = 0 \).

**Solution :** Since \( r = 1 \) is a double root then \( M = 2 \), so the accelerated Newton’s method becomes
\[ r_n = r_{n-1} - \frac{2 f(r_{n-1})}{f'(r_{n-1})} = r_{n-1} - \frac{2(-r^3_{n-1} + 3r_{n-1} - 2)}{-3r^2_{n-1} + 3}, \quad n = 1, 2, \ldots, \]
or
\[ r_n = r_{n-1} - \frac{-2r^3_{n-1} + 6r_{n-1} - 4}{-3r^2_{n-1} + 3}, \quad n = 1, 2, \ldots. \]
Start by computing \( r_1 \)
\[ r_1 = r_0 - \frac{-2r_0^3 + 6r_0 - 4}{-3r_0^2 + 3} = 1.25 - \frac{-2(1.25)^3 + 6(1.25) - 4}{-3(1.25)^2 + 3} = 1.00925926. \]
Table 2.8 compares the performance of both methods.
CHAPTER 2. NUMERICAL SOLUTIONS OF NONLINEAR EQUATIONS

Table 2.8: Newton’s and Accelerated Newton’s Methods Solutions of Example 17

<table>
<thead>
<tr>
<th>n</th>
<th>Newton’s Method</th>
<th>Accelerated Newton’s Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.25</td>
<td>1.25</td>
</tr>
<tr>
<td>1</td>
<td>1.12962963</td>
<td>1.00925926</td>
</tr>
<tr>
<td>2</td>
<td>1.06612990</td>
<td>1.00001422</td>
</tr>
<tr>
<td>3</td>
<td>1.03341772</td>
<td>1.00001422</td>
</tr>
<tr>
<td>4</td>
<td>1.01680039</td>
<td>1.00000000</td>
</tr>
<tr>
<td>5</td>
<td>1.00842352</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1.00421765</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1.00211030</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1.00105552</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1.00052785</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1.00026395</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1.00013198</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1.00006599</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>1.00003300</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1.00001650</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>1.00000825</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>1.00000413</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>1.00000207</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>1.00000104</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>1.00000052</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>1.00000026</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>1.00000013</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>1.00000007</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>1.00000004</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>1.00000002</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>1.00000001</td>
<td></td>
</tr>
</tbody>
</table>

2.9 Computing Roots of Polynomials

Computing roots of polynomials has important applications in different areas of mathematics and other sciences.

**Definition 19** (nth Degree Polynomial). A polynomial of degree $n$ has the general form

\[ P(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0, \]
where the coefficients $a_i$, $i = 0, 1, \ldots, n$, are real numbers (constants) and $a_n \neq 0$. The $n$th degree polynomial $P(x)$ is sometimes referred to as $P_n(x)$, and also named algebraic polynomial.

Note that the zero function $P(x) = 0$ is a polynomial but has no degree. There are several techniques for finding zeros of polynomials in the literature such as Müller method, Laguerre’s method, Bairstow method, Brent’s method and Jenkins-Traub method, and these methods are beyond the scope of this lecture notes and interested readers can see the references.
Exercises

Exercise 10. Solve Example 10 using the bisection method and compare the solution with false position method’s solution of the same problem.

Exercise 11. Repeat solving Example 9 using the false position method and compare the results with the solution of the bisection method for the same problem.

Exercise 12. Find the solution to the equation \( e^x - x - 1 = 0 \) accurate to six decimal places (i.e. \( \epsilon = 0.0000001 \)) using Newton’s and modified Newton’s methods. Start with \( r_0 = 0.6 \). Compare the results of both methods.

Exercise 13. Use the secant method to find the solution accurate to within \( 10^{-5} \) to the following problem \( x \sin(x) - 1 = 0, \ 0 \leq x \leq 2 \).

Exercise 14. Use the fixed point method to locate the root of \( f(x) = x - e^{-x} = 0 \), start with an initial guess of \( x = 0.1 \).

Exercise 15. Let \( f(x) = x^2 - 5 \) and \( r_0 = 1.5 \). Use bisection, false position, secant, fixed point, Newton’s and modified Newton’s methods to find \( r_7 \) the approximation to the positive root \( r = \sqrt{5} \).

Exercise 16. Use modified (accelerated) Newton’s method to solve the equation \( x^2 - 3x - 1 = 0 \) in the interval \([-1, 1]\).
Chapter 3

Solving Systems of Linear Equations

3.1 Introduction

Many phenomena and relationships in nature and real life applications are linear, meaning that results and their causes are proportional to each other. Solving linear algebraic equations is a topic of great importance in numerical analysis and other scientific disciplines such as engineering and physics. Solutions to many problems reduced to solve a system of linear equations. For example, in finite element analysis a solution of a partial differential equation is reduced to solve a system of linear equations.

3.2 Norms of Matrix and Vectors

In error and convergence analyses we need a measure to determine the distance (difference) between the exact solution and approximate solution or to determine the differences between consecutive approximations.

Definition 20 (Vector Norm). A vector norm is a real-valued function $\| \cdot \| : \mathbb{R}^n \rightarrow \mathbb{R}$ satisfies the following conditions:

(i) $\| x \| \geq 0$ for all $x \in \mathbb{R}^n$.

(ii) $\| x \| = 0$ if and only if $x = 0$ for all $x \in \mathbb{R}^n$.

(iii) $\| \alpha x \| = |\alpha| \| x \|$ for all $\alpha \in \mathbb{R}$ and $x \in \mathbb{R}^n$.

(iv) $\| x + y \| \leq \| x \| + \| y \|$ for all $x, y \in \mathbb{R}^n$ (Triangle Inequality).
CHAPTER 3. SOLVING SYSTEMS OF LINEAR EQUATIONS

Definition 21 (\(l_1\) Vector Norm). Let \(x = (x_1, x_2, \ldots, x_n)\). Then the \(l_1\) norm for the vector \(x\) is defined by

\[
\|x\|_1 = \sum_{i=1}^{n} |x_i|.
\]

Definition 22 (Euclidean Vector Norm). Let \(x = (x_1, x_2, \ldots, x_n)\). Then the Euclidean norm (\(l_2\) norm) for the vector \(x\) is defined by

\[
\|x\|_2 = \left( \sum_{i=1}^{n} x_i^2 \right)^{1/2}.
\]

Definition 23 (Maximum Vector Norm). Let \(x = (x_1, x_2, \ldots, x_n)\). Then the maximum norm (\(l_\infty\) norm) for the vector \(x\) is defined by

\[
\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|.
\]

Remark 5. Note that when \(n = 1\) both norms reduce to the absolute value function of real numbers.

Example 18. Determine the \(l_1\) norm, \(l_2\) norm and \(l_\infty\) norm of the vector \(x = (1, 0, -1, 2, 3)\).

Solution: The required norms of vector \(x = (1, 0, -1, 2, 3)\) in \(\mathbb{R}^5\) are:

\[
\|x\|_1 = \sum_{i=1}^{5} |x_i| = |1| + |0| + |-1| + |2| + |3| = 7,
\]

\[
\|x\|_2 = \left( \sum_{i=1}^{5} x_i^2 \right)^{1/2} = \left( x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 \right)^{1/2}
\]

\[
= \left( (1)^2 + (0)^2 + (-1)^2 + (2)^2 + (3)^2 \right)^{1/2} = \left( 15 \right)^{1/2},
\]

and

\[
\|x\|_\infty = \max_{1 \leq i \leq 5} |x_i| = \max\{|1|, |0|, |-1|, |2|, |3|\}
\]

\[
= \max\{|1|, |0|, |-1|, |2|, |3|\} = 3.
\]
CHAPTER 3. SOLVING SYSTEMS OF LINEAR EQUATIONS

Definition 24 (Matrix Norm). A matrix norm is a real-valued function \( \| \cdot \| : \mathbb{R}^{n \times m} \to \mathbb{R} \) satisfies the following conditions:

(i) \( \| A \| \geq 0 \) for all \( A \in \mathbb{R}^{n \times m} \).

(ii) \( \| A \| = 0 \) if and only if \( A = 0 \) for all \( A \in \mathbb{R}^{n \times m} \).

(iii) \( \| \alpha A \| = |\alpha| \| A \| \) for all \( \alpha \in \mathbb{R} \) and \( A \in \mathbb{R}^{n \times m} \).

(iv) \( \| A + B \| \leq \| A \| + \| B \| \) for all \( A, B \in \mathbb{R}^{n \times m} \) (Triangle Inequality).

If matrix norm is related to a vector norm, then we have two additional properties:

(v) \( \| AB \| \leq \| A \| \| B \| \) for all \( A, B \in \mathbb{R}^{n \times m} \).

(vi) \( \| Ax \| \leq \| A \| \| x \| \) for all \( A \in \mathbb{R}^{n \times m} \) and \( x \in \mathbb{R}^{n} \).

We give here some equivalent definitions of the matrix norm particularly when matrix norm is related to the vector norm.

Definition 25 (Subordinate Matrix Norm). Let \( A \) is a \( n \times n \) matrix and \( x \in \mathbb{R}^{n} \), then the subordinate matrix norm is defined by

\[ \| A \| = \sup \{ \| Ax \| : x \in \mathbb{R}^{n} \text{ and } \| x \| = 1 \} \]

or, alternatively
\[ \| A \| = \max_{\| x \| = 1} \| Ax \| . \]

Definition 26 (Natural Matrix Norm). Let \( A \) is a \( n \times n \) matrix and for any \( z \neq 0 \), and \( x = \frac{z}{\| z \|} \) is the unit vector. Then the natural / reduced matrix norm is defined by

\[ \| A \| = \max_{z \neq 0} \| A z \| / \| z \| , \]

or, alternatively
\[ \| A \| = \max_{z \neq 0} \frac{\| A z \|}{\| z \|} . \]

Definition 27 (l_1 Matrix Norm). Let \( A \) is a \( n \times n \) matrix and \( x = (x_1, x_2, \cdots, x_n)' \). Then the \( l_1 \) matrix norm is defined by

\[ \| A \|_1 = \max_{\| x \|_1 = 1} \| Ax \|_1 = \max_{1 \leq i \leq n} \sum_{i=1}^{n} |a_{ij}| . \]
CHAPTER 3. SOLVING SYSTEMS OF LINEAR EQUATIONS

Definition 28 (Spectral Matrix Norm). Let $A$ be a $n \times n$ matrix and $x = (x_1, x_2, \cdots, x_n)'$. Then the spectral / $l_2$-matrix norm is defined by

$$\|A\|_2 = \max_{\|x\|_2 = 1} \|Ax\|_2 = \max_{1 \leq i \leq n} \sqrt{|\sigma_{\text{max}}|},$$

where $\sigma_i$ are the eigenvalues of $A^T A$, which are called the singular values of $A$ and the largest eigenvalue in absolute value ($|\sigma_{\text{max}}|$) is called the spectral radius of $A$.

Definition 29 ($l_\infty$ Matrix Norm). Let $A$ be a $n \times n$ matrix and $x = (x_1, x_2, \cdots, x_n)'$. Then the $l_\infty$ (maximum) matrix norm is defined by

$$\|A\|_\infty = \max_{\|x\|_\infty = 1} \|Ax\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|.$$

Remark 6. Note that $\|I\| = 1$.

Example 19. Determine $\|A\|_\infty$ for the matrix

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

Solution: For $i = 1$, we have

$$\sum_{j=1}^3 |a_{1j}| = |a_{11}| + |a_{12}| + |a_{13}| = |1| + |-1| + |2| = 4,$$

and for $i = 2$, we obtain

$$\sum_{j=1}^3 |a_{2j}| = |a_{21}| + |a_{22}| + |a_{23}| = |0| + |5| + |3| = 8,$$

for $i = 3$, we get

$$\sum_{j=1}^3 |a_{3j}| = |a_{31}| + |a_{32}| + |a_{33}| = |-1| + |6| + |-4| = 11.$$

Consequently,

$$\|A\|_\infty = \max_{1 \leq i \leq 3} \sum_{j=1}^3 |a_{ij}| = \max\{4, 8, 11\} = 11.$$
3.3 Direct Methods

Direct methods are techniques used for solving and obtaining the exact solutions (in theory) of linear algebraic equations in a finite number of steps. The main widely used direct methods are Gaussian elimination method and Gauss-Jordan method.

Consider the following linear system of dimension \( n \times (n + 1) \)

\[
\begin{align*}
 a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
 a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
 &\vdots \\
 a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n.
\end{align*}
\]

This system can be written in concise form by using matrix notation as \( AX = B \) as follows:

\[
\begin{bmatrix}
 a_{11} & a_{12} & \cdots & a_{1n} \\
 a_{21} & a_{22} & \cdots & a_{2n} \\
 \vdots & \vdots & \ddots & \vdots \\
 a_{n1} & a_{n2} & \cdots & a_{nn}
\end{bmatrix}
\begin{bmatrix}
 x_1 \\
 x_2 \\
 \vdots \\
 x_n
\end{bmatrix}
= 
\begin{bmatrix}
 b_1 \\
 b_2 \\
 \vdots \\
 b_n
\end{bmatrix},
\]

where \( A_{n \times n} \) is square matrix and is called a coefficient matrix, \( B_{n \times 1} \) is a column vector known as the right hand side vector and \( X_{n \times 1} \) is a column vector known as unknowns vector. Also, this system can be written as

\[
[A|B] = 
\begin{bmatrix}
 a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\
 a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\
 \vdots & \vdots & \ddots & \vdots & \vdots \\
 a_{n1} & a_{n2} & \cdots & a_{nn} & b_n
\end{bmatrix},
\]

where \([A|B]\) is called the augmented matrix.

3.3.1 Backward Substitution Method

Backward substitution also called back substitution is an algorithm or technique used for solving upper-triangular systems which are systems such that their coefficient matrices are upper-triangular matrices. Assume that we have the following upper-triangular system
CHAPTER 3. SOLVING SYSTEMS OF LINEAR EQUATIONS

\[a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n-1}x_{n-1} + a_{1n}x_n = b_1\]
\[a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n-1}x_{n-1} + a_{2n}x_n = b_2\]
\[a_{33}x_3 + \cdots + a_{3n-1}x_{n-1} + a_{3n}x_n = b_3\]
\[\vdots\]
\[a_{n-1n-1}x_{n-1} + a_{n-1n}x_n = b_{n-1}\]
\[a_{nn}x_n = b_n.\]

To find a solution to this system we follow the following steps provided that \(x_{rr} \neq 0\), \(r = 1, 2, \ldots, n\):

1. Solve the last \((n)th\) equation for \(x_n\):
   \[x_n = \frac{b_n}{a_{nn}}.\]

2. Substitute \(x_n\) in the next-to-last \(((n-1)th)\) equation and solve it for \(x_{n-1}\):
   \[x_{n-1} = \frac{b_{n-1} - a_{n-1n}x_n}{a_{n-1n-1}}.\]

3. Now, \(x_n\) and \(x_{n-1}\) are known and can be used to find \(x_{n-2}\):
   \[x_{n-2} = \frac{b_{n-2} - a_{n-2n-1}x_{n-1} - a_{n-1n}x_n}{a_{n-2n-2}}.\]

4. Continuing in this way until we arrive at the general step:
   \[x_r = \frac{b_r - \sum_{j=r+1}^{n} a_{rj}x_j}{a_{rr}}, \quad r = n-1, n-2, \ldots, 1.\]

Example 20. Solve the following linear system using back substitution method

\[3x_1 + 2x_2 - x_3 + x_4 = 10\]
\[x_2 - x_3 + 2x_4 = 9\]
\[3x_3 - x_4 = 1\]
\[3x_4 = 6\]
Solution: Solve the last equation for $x_4$ to obtain

$$x_4 = \frac{6}{3} = 2.$$  

Substitute $x_4 = 2$ in the third equation, we have

$$x_3 = \frac{1 + x_4}{3} = \frac{1 + 2}{3} = \frac{3}{3} = 1.$$  

Now, use values $x_3 = 1$ and $x_4 = 2$ in the second equation to find $x_2$

$$x_2 = 9 + x_3 - 2x_4 = 9 + 1 - 4 = 6.$$  

Finally, solve the first equation for $x_1$ yields

$$x_1 = \frac{10 - 2x_2 + x_3 - x_4}{3} = \frac{10 - 12 + 1 - 2}{3} = \frac{-3}{3} = -1.$$  

3.3.2 Forward Substitution Method

Forward substitution is an algorithm or technique used for solving lower-triangular systems which are systems such that their coefficient matrices are lower-triangular matrices.

$$a_{11}x_1 = b_1$$
$$a_{21}x_1 + a_{22}x_2 = b_2$$
$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3$$
$$\vdots$$
$$a_{n-1,1}x_1 + a_{n-1,2}x_2 + a_{n-1,3}x_3 + \cdots + a_{n-1,n-1}x_{n-1} = b_{n-1}$$
$$a_{n1}x_1 + a_{n2}x_2 + a_{n3}x_3 + \cdots + a_{n,n-1}x_{n-1} + a_{nn}x_n = b_n.$$

To find a solution to this system we follow the following steps provided that $x_{rr} \neq 0$, $r = 1, 2, \cdots, n$:

(1) Solve the first (1st) equation for $x_1$:

$$x_1 = \frac{b_1}{a_{11}}.$$

(2) Substitute $x_1$ in the second equation (2nd) equation and solve it for $x_2$:

$$x_2 = \frac{b_2 - a_{21}x_1}{a_{22}}.$$
(3) Now, \( x_1 \) and \( x_2 \) are known and can be used to find \( x_3 \):
\[
x_3 = \frac{b_3 - a_{31}x_1 - a_{32}x_2}{a_{33}}.
\]

(4) Continuing in this way until we arrive at the general step:
\[
x_r = \frac{b_r - \sum_{j=1}^{r-1} a_{rj}x_j}{a_{rr}}, \quad r = 2, 3, \ldots, n.
\]

Example 21. Use the forward substitution method for solving the following linear system
\[
\begin{align*}
4x_1 &= 8 \\
2x_1 + x_2 &= -1 \\
x_1 - x_2 + 5x_3 &= 0.5 \\
0.1x_1 + 2x_2 - x_3 + 2x_4 &= 12
\end{align*}
\]

Solution: Solving the first equation for \( x_1 \) yields
\[
x_1 = \frac{8}{4} = 2.
\]

Using the value of \( x_1 \) to find \( x_2 \)
\[
x_2 = \frac{-1 - 2x_1}{2} = \frac{-1 - 2(2)}{2} = -2.5.
\]

Use \( x_1 \) and \( x_2 \) to find \( x_3 \)
\[
x_3 = \frac{0.5 - x_1 + x_2}{5} = \frac{0.5 - 2 - 2.5}{5} = \frac{-4}{5} = -0.8.
\]

Finally, solve for \( x_4 \) to have
\[
x_4 = \frac{12 - 0.1x_1 - 2x_2 + x_3}{2} = \frac{12 - 0.1(2) - 2(2.5) - 0.8}{2} = \frac{6}{2} = 3.
\]

3.3.3 Gaussian Elimination Method

Gaussian elimination method is also known as Gauss elimination method or simply elimination method. It is a direct method used for solving a system of linear algebraic equations. In this method we transform the linear system to an equivalent upper or lower triangular system and then solve it by backward or forward substitution. The process of transforming the linear system to an equivalent upper or lower triangular system is called triangularisation.

46 Mohammad Sabawi/Numerical Analysis
CHAPTER 3. SOLVING SYSTEMS OF LINEAR EQUATIONS

Definition 30 (Equivalent Systems). Two linear algebraic systems of dimension $n \times n$ is said to be they are equivalent if they have the same solution sets.

Definition 31 (Elementary Transformations). The following operations performed on a linear system transform it to an equivalent system:

- **Interchanges**: Changing the order of any two equations in the system.
- **Replacement**: Any equation of the system can be replaced by itself and a nonzero multiple of any other equation in the system.
- **Scaling**: Multiplying any equation in the system by a nonzero real constant.

Definition 32 (Elementary Row Operations). The following operations performed on a linear system transform it to an equivalent system:

- **Interchanges**: Changing the order of any two rows in the matrix.
- **Replacement**: Any row in the matrix can be replaced by its sum and a nonzero multiple of any other row in the matrix.
- **Scaling**: Multiplying any row in the matrix by a nonzero real constant.

Definition 33 (Pivoting). The process of interchanging the rows of the coefficient matrix $A$ containing the element (entry) $a_{kk}$ and using it to eliminate elements $a_{rk}$, $r = k+1, k+2, \cdots n$ is called pivoting process. The element $a_{kk}$ is called pivotal element and the $k$th row is called pivotal row.

Example 22. Write the following linear system in the augmented form and then solve it by using Gauss elimination method.

\[
\begin{align*}
2x_1 + 2x_2 - x_3 + 3x_4 &= 12 \\
2x_1 + x_2 + x_3 + x_4 &= 10 \\
-3x_1 - x_2 + 4x_3 + x_4 &= 2 \\
x_1 + x_2 - x_3 + 3x_4 &= 6
\end{align*}
\]

Solution: The augmented matrix is

\[
\begin{bmatrix}
1 & 2 & -1 & 4 & 12 \\
2 & 1 & 1 & 1 & 10 \\
-3 & -1 & 4 & 1 & 2 \\
1 & 1 & -1 & 3 & 6
\end{bmatrix}
\]
CHAPTER 3. SOLVING SYSTEMS OF LINEAR EQUATIONS

The first row is the pivotal row, so the pivotal element is \(a_{11} = 1\) and is used to eliminate the first column below the diagonal. We will denote by \(m_r\) to the multiples of the row 1 subtracted from row \(r\) for \(r = 2, 3, 4\). Multiplying the first row by \(m_{21} = -2\) and add it to the second row to have

\[
\begin{bmatrix}
1 & 2 & -1 & 4 & 12 \\
0 & -3 & 3 & -7 & -14 \\
-3 & -1 & 4 & 1 & 2 \\
1 & 1 & -1 & 3 & 6
\end{bmatrix}.
\]

Now, multiply the first row by \(m_{31} = 3\) and add it to the third row to obtain

\[
\begin{bmatrix}
1 & 2 & -1 & 4 & 12 \\
0 & -3 & 3 & -7 & -14 \\
0 & 5 & 1 & 13 & 38 \\
1 & 1 & -1 & 3 & 6
\end{bmatrix}.
\]

Multiplying the first row by \(m_{41} = -1\) and adding it to the fourth row yields

\[
\begin{bmatrix}
1 & 2 & -1 & 4 & 12 \\
0 & -3 & 3 & -7 & -14 \\
0 & 5 & 1 & 13 & 38 \\
0 & -1 & 0 & -1 & -6
\end{bmatrix}.
\]

Now, the pivotal row is the second row and the pivotal element is \(a_{22} = -3\). Multiply the second row by \(m_{32} = \frac{2}{3}\) to have

\[
\begin{bmatrix}
1 & 2 & -1 & 4 & 12 \\
0 & -3 & 3 & -7 & -14 \\
0 & 0 & 6 & 4/3 & 44/3 \\
0 & -1 & 0 & -1 & -6
\end{bmatrix}.
\]

Multiply the the second row by \(m_{42} = \frac{1}{3}\) and add it to the fourth row to obtain

\[
\begin{bmatrix}
1 & 2 & -1 & 4 & 12 \\
0 & -3 & 3 & -7 & -14 \\
0 & 0 & 6 & 4/3 & 44/3 \\
0 & 0 & -1 & 4/3 & -4/3
\end{bmatrix}.
\]

Now, the pivotal row is the third row and the third element is \(a_{33} = 6\). Finally, multiply the third row by \(m_{43} = \frac{1}{6}\) to the fourth row to have

\[
\begin{bmatrix}
1 & 2 & -1 & 4 & 12 \\
0 & -3 & 3 & -7 & -14 \\
0 & 0 & 6 & 4/3 & 44/3 \\
0 & 0 & -1 & 4/3 & -4/3
\end{bmatrix}.
\]
CHAPTER 3. SOLVING SYSTEMS OF LINEAR EQUATIONS

\[
\begin{bmatrix}
1 & 2 & -1 & 4 & | & 12 \\
0 & -3 & 3 & -7 & | & -14 \\
0 & 0 & 6 & 4/3 & | & 44/3 \\
0 & 0 & 0 & 14/9 & | & 10/9 \\
\end{bmatrix}.
\]

Now, note that the coefficient matrix is transformed into an upper triangular matrix and can be solved by backward substitution method. Firstly, we from the last row we compute

\[x_4 = \frac{10/9}{14/9} = \frac{5}{7}.
\]

Use the third row to solve for \(x_3\)

\[x_3 = \frac{44/3 - 4/3(5/7)}{6} = \frac{288/21}{6} = \frac{16}{7}.
\]

Now, solve the second equation for \(x_2\)

\[x_2 = \frac{-14 - 3x_3 + 7x_4}{-3} = \frac{-14 - 3(16/7) + 7(5/7)}{-3} = \frac{111}{21} = \frac{37}{7}.
\]

Finally, solve the first equation for \(x_1\)

\[x_1 = 12 - 2x_2 + x_3 - 4x_4 = 12 - 2(37/7) + 16/7 - 4(5/7) = \frac{6}{7}.
\]

**Example 23.** Solve the following linear system using Gauss elimination method by using forward substitution technique

\[
\begin{align*}
x_1 + 2x_2 + x_3 + 4x_4 &= 13 \\
2x_1 + 0x_2 + 4x_3 + 3x_4 &= 28 \\
4x_1 + 2x_2 + 2x_3 + x_4 &= 20 \\
-3x_1 + x_2 + 3x_3 + 2x_4 &= 6
\end{align*}
\]

**Solution:** We start our solution strategy by transforming this square system to equivalent lower-triangular system and then solve it by using forward substitution method. Write the system in augmented matrix form

\[
\begin{bmatrix}
1 & 2 & 1 & 4 & | & 13 \\
2 & 0 & 4 & 3 & | & 28 \\
4 & 2 & 2 & 1 & | & 20 \\
-3 & 2 & 1 & 3 & | & 6
\end{bmatrix}.
\]
Note that now the pivotal row is the fourth row and the pivotal element is \( a_{44} = 2 \). Multiply the fourth row by the multiple \( m_{14} = -2 \) and add it to the first row to have

\[
\begin{bmatrix}
7 & 0 & -5 & 0 & 1 \\
2 & 0 & 4 & 3 & 28 \\
4 & 2 & 2 & 1 & 20 \\
-3 & 1 & 3 & 2 & 6 \\
\end{bmatrix}.
\]

Multiply the fourth row by \( m_{24} = -\frac{3}{2} \) and add it to the second row to obtain

\[
\begin{bmatrix}
7 & 0 & -5 & 0 & 1 \\
13/2 & -3/2 & -1/2 & 0 & 19 \\
4 & 2 & 2 & 1 & 20 \\
-3 & 1 & 3 & 2 & 6 \\
\end{bmatrix}.
\]

Now multiply the fourth equation by \( m_{34} = \frac{-1}{2} \) and add it to the third row to have

\[
\begin{bmatrix}
7 & 0 & -5 & 0 & 1 \\
13/2 & -3/2 & -1/2 & 0 & 19 \\
11/2 & 3/2 & 1/2 & 0 & 17 \\
-3 & 1 & 3 & 2 & 6 \\
\end{bmatrix}.
\]

The pivotal row now is the third row and the pivotal element is \( a_{33} = 1/2 \). Add the third row to the second row (i.e. multiply it by \( m_{23} = 1 \)) to get

\[
\begin{bmatrix}
7 & 0 & -5 & 0 & 1 \\
12 & 0 & 0 & 0 & 36 \\
11/2 & 3/2 & 1/2 & 0 & 17 \\
-3 & 1 & 3 & 2 & 6 \\
\end{bmatrix}.
\]

Now

\[
\begin{bmatrix}
12 & 0 & 0 & 0 & 36 \\
11/2 & 3/2 & 1/2 & 0 & 17 \\
7 & 0 & -5 & 0 & 1 \\
-3 & 1 & 3 & 2 & 6 \\
\end{bmatrix}.
\]

The pivotal row (third row) is used to eliminate elements in the second row and the pivotal element is \( a_{33} = -5 \). Multiply the third row by \( m_{23} = \frac{4}{10} \) to have
Chapter 3. Solving Systems of Linear Equations

Now, use forward substitution to solve the lower-triangular matrix. Solve the first equation for $x_1$

$$x_1 = \frac{36}{12} = 3.$$  

Use the equation to find $x_2$

$$x_2 = \frac{171/10 - (31/5)3}{3/2} = -1.$$  

Now, solve the third equation for $x_3$

$$x_3 = \frac{1 - 7(3)}{-5} = 4.$$  

Finally, solve the fourth equation for $x_4$

$$x_4 = \frac{6 - (-3)(3) - 1(-1) - 3(4)}{2} = 2.$$  

3.3.4 Gauss-Jordan Elimination Method

In this method instead of transforming the coefficient matrix into upper or lower triangular system, we transform the coefficient matrix into diagonal (in particular identity) matrix using elementary row operations.

Example 24. Solve the following linear system using Gauss-Jordan elimination method

$$3x_1 + 4x_2 + 3x_3 = 10$$  
$$x_1 + 5x_2 - x_3 = 7$$  
$$6x_1 + 3x_2 + 7x_3 = 15$$  

Solution: Express the system in augmented matrix form
CHAPTER 3. SOLVING SYSTEMS OF LINEAR EQUATIONS

\[
\begin{bmatrix}
3 & 4 & 3 & | & 10 \\
1 & 5 & -1 & | & 7 \\
6 & 3 & 7 & | & 15 \\
\end{bmatrix}.
\]

The pivot row is the first row and the pivot element is \(a_{11} = 3\). Multiply it by \(m_{11} = 1/3\) to get

\[
\begin{bmatrix}
1 & 4/3 & 1 & | & 10/3 \\
1 & 5 & -1 & | & 7 \\
6 & 3 & 7 & | & 15 \\
\end{bmatrix}.
\]

Subtract the second equation from the first (i.e. multiply it by \(m_{21} = -1\)) and multiply the first equation by \(m_{31} = -6\) and add it to the third equation to have

\[
\begin{bmatrix}
1 & 4/3 & 1 & | & 10/3 \\
0 & -11/3 & 2 & | & -11/3 \\
0 & -5 & 1 & | & -5 \\
\end{bmatrix}.
\]

Now, the pivot row is the second row and the pivot element \(a_{22} = -11/3\). Multiply it by \(m_{22} = -3/11\) to have

\[
\begin{bmatrix}
1 & 4/3 & 1 & | & 10/3 \\
0 & 1 & -6/11 & | & 1 \\
0 & -5 & 1 & | & -5 \\
\end{bmatrix}.
\]

Multiply the first and third rows by \(m_{12} = -4/3\) and \(m_{32} = 5\) to obtain

\[
\begin{bmatrix}
1 & 0 & 19/11 & | & 2 \\
0 & 1 & -6/11 & | & 1 \\
0 & 0 & -19/11 & | & 0 \\
\end{bmatrix}.
\]

The pivot element now is third row and the pivot element is \(a_{33} = -19/11\). Multiply it by \(m_{33} = -11/19\) to get

\[
\begin{bmatrix}
1 & 0 & 19/11 & | & 2 \\
0 & 1 & -6/11 & | & 1 \\
0 & 0 & 1 & | & 0 \\
\end{bmatrix}.
\]

Finally, multiply the third row by \(m_{13} = -19/11\) and \(m_{23} = 6/11\) and add it to the first and second rows to have

\[
\begin{bmatrix}
1 & 0 & 0 & | & 2 \\
0 & 1 & 0 & | & 1 \\
0 & 0 & 1 & | & 0 \\
\end{bmatrix}.
\]
Hence, we have \( x_1 = 2 \), \( x_2 = 1 \) and \( x_3 = 0 \).

**Example 25.** Solve the following linear system using Gauss-Jordan elimination method

\[
\begin{align*}
-2x_1 + x_2 + 5x_3 &= 15 \\
4x_1 - 8x_2 + x_3 &= -21 \\
4x_1 - x_2 + x_3 &= 7
\end{align*}
\]

**Solution:** Write the system in augmented matrix form

\[
\begin{bmatrix}
-2 & 1 & 5 & | & 15 \\
4 & -8 & 1 & | & -21 \\
4 & -1 & 1 & | & 7
\end{bmatrix}.
\]

Multiply the first row by \( m_{21} = m_{31} = -2 \) and it to the second and third rows respectively, to obtain

\[
\begin{bmatrix}
-2 & 1 & 5 & | & 15 \\
0 & -6 & 11 & | & 9 \\
0 & 1 & 11 & | & 37
\end{bmatrix}.
\]

Now, multiply the second row by \( m_{12} = m_{32} = \frac{1}{6} \) and it to the first and third rows respectively, to have

\[
\begin{bmatrix}
-2 & 0 & 41/6 & | & 33/2 \\
0 & -6 & 11 & | & 9 \\
0 & 0 & 77/6 & | & 77/2
\end{bmatrix}.
\]

Finally, multiply the third row by \( m_{13} = \frac{-41}{77} \) and \( m_{32} = \frac{-6}{7} \) and it to the first and third rows respectively, to obtain

\[
\begin{bmatrix}
-2 & 0 & 0 & | & -4 \\
0 & -6 & 0 & | & -24 \\
0 & 0 & 77/6 & | & 77/2
\end{bmatrix},
\]

implies that

\[
x_1 = \frac{-4}{-2} = 2, \quad x_2 = \frac{-24}{-6} = 4 \quad \text{and} \quad x_3 = \frac{77/2}{77/6} = 3.
\]
CHAPTER 3. SOLVING SYSTEMS OF LINEAR EQUATIONS

3.4 **LU and Cholesky Factorisations**

In this section we will discuss the triangular factorisations of matrices.

**Definition 34 (Positive Definite Matrix).** Let $A_{n \times n}$ be symmetric real matrix and $x \in \mathbb{R}^n$ a nonzero vector. Then, $A$ is said to be **positive definite matrix** if $A = A^T$ and $x^T Ax > 0$ for any $x$.

**Remark 7.** Note that the matrix $A$ is nonsingular by definition.

**Definition 35 (Triangular Factorisation).** Assume that $A$ is a nonsingular matrix. It said to be $A$ has a **triangular factorisation** or **triangular decomposition** if it can be factorised as a product of unit lower-triangular matrix $L$ and an upper triangular matrix $U$:

$$A = LU.$$ 

or in matrix form

$$
\begin{bmatrix}
a_{11} & a_{12} & a_{13} \\
a_{21} & a_{22} & a_{23} \\
a_{31} & a_{32} & a_{33}
\end{bmatrix}
= 
\begin{bmatrix}
1 & 0 & 0 \\
l_{21} & 1 & 0 \\
l_{31} & l_{32} & 1
\end{bmatrix}
\begin{bmatrix}
u_{11} & u_{12} & u_{13} \\
u_{22} & u_{23} & \\
u_{33}
\end{bmatrix}.
$$

Note that since $A$ is nonsingular matrix this implies that $u_{rr} \neq 0$ for all $r$ and this is called **Doolittle factorisation**.

Also, $A$ can be expressed as a product of lower-triangular matrix $L$ and unit upper triangular matrix $U$:

$$
\begin{bmatrix}
a_{11} & a_{12} & a_{13} \\
a_{21} & a_{22} & a_{23} \\
a_{31} & a_{32} & a_{33}
\end{bmatrix}
= 
\begin{bmatrix}
l_{11} & 0 & 0 \\
l_{21} & l_{22} & 0 \\
l_{31} & l_{32} & l_{33}
\end{bmatrix}
\begin{bmatrix}
1 & u_{12} & u_{13} \\
0 & 1 & u_{23} \\
0 & 0 & 1
\end{bmatrix},
$$

and this is called **Crout factorisation**.

To solve the linear system $AX = B$ using $LU$ factorisation, we do the following two steps:

1. Using forward substitution to solve the the lower-triangular linear system $LY = B$ for $Y$.

2. Using backward substitution to solve the upper-triangular linear system $UX = Y$ for $X$. 

54 Mohammad Sabawi/Numerical Analysis
Example 26. Solve the following linear system using LU (Doolittle) decomposition

\[
\begin{align*}
2x_1 - 3x_2 + x_3 &= 2 \\
x_1 + x_2 - x_3 &= -1 \\
-x_1 + x_2 - x_3 &= 0
\end{align*}
\]

Solution: Express the system in matrix form

\[
\begin{bmatrix}
2 & -3 & 1 \\
1 & 1 & -1 \\
-1 & 1 & -1
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix}
= 
\begin{bmatrix}
2 \\
-1 \\
0
\end{bmatrix}
\]

Factor A as follows:

\[
\begin{bmatrix}
2 & -3 & 1 \\
1 & 1 & -1 \\
-1 & 1 & -1
\end{bmatrix}
= 
\begin{bmatrix}
1 & 0 & 0 \\
l_{21} & 1 & 0 \\
l_{31} & l_{32} & 1
\end{bmatrix}
\begin{bmatrix}
u_{11} & u_{12} & u_{13} \\
u_{22} & u_{23} \\
u_{33}
\end{bmatrix}
\]

Find the values of the entries of matrices \( L \) and \( U \). From the first column we have

\[
2 = l_{21}u_{11} \implies u_{11} = 2,
\]

and

\[
1 = l_{21}u_{11} = l_{21}2 \implies l_{21} = 0.5,
\]

finally

\[
-1 = l_{31}u_{11} = l_{31}2 \implies l_{31} = -0.5.
\]

In the second column, we have

\[
-3 = l_{21}u_{12} \implies u_{12} = -3,
\]

and

\[
1 = l_{21}u_{12} + u_{22} = -1.5 + u_{22} \implies u_{22} = 2.5,
\]

so

\[
1 = l_{31}u_{12} + l_{32}u_{22} = (-0.5)(-3) + l_{32}(2.5) \implies l_{32} = -0.2.
\]
Finally, in the third column we have

\[ 1 = 1u_{13} \implies u_{13} = 1, \]

and

\[ -1 = l_{21}u_{13} + 1u_{23} = 0.5 + u_{23} \implies u_{23} = -1.5, \]

finally,

\[ -1 = l_{31}u_{13} + l_{32}u_{23} + 1u_{33} = -0.5(1) + (-0.2)(-1.5) + u_{33} \implies u_{33} = -0.8. \]

Now, we have the \( LU \) factorisation

\[
A = \begin{bmatrix} 2 & -3 & 1 \\ 1 & 1 & -1 \\ -1 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 1 & 0 \\ -0.5 & -0.2 & 1 \end{bmatrix} \begin{bmatrix} 2 & -3 & 1 \\ 0 & 2.5 & -1.5 \\ 0 & 0 & -0.8 \end{bmatrix} = LU.
\]

Now, we have the following lower-triangular linear system \( LY = B \) for \( Y \)

\[
\begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 1 & 0 \\ -0.5 & -0.2 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}.
\]

Write the system in augmented matrix form

\[
\begin{bmatrix} 1 & 0 & 0 & 2 \\ 0.5 & 1 & 0 & -1 \\ -0.5 & -0.2 & 1 & 0 \end{bmatrix}.
\]

Solve this system by forward substitution to have

\[ y_1 = 2, \quad y_2 = -1 - 0.5(y_1) = -1 - 0.5(2) = -2, \]

and
\[ y_3 = 0 + 0.5(y_1) + 0.2(y_2) = 0.5(2) + 0.2(-2) = 0.6. \]

Now, we have the following upper-triangular linear system \( UX = Y \)

\[
\begin{bmatrix}
2 & -3 & 1 \\
0 & 2.5 & -1.5 \\
0 & 0 & -0.8
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix}
=
\begin{bmatrix}
2 \\
-2 \\
0.6
\end{bmatrix}.
\]

Express the system in augmented matrix form

\[
\begin{bmatrix}
2 & -3 & 1 & 2 \\
0 & 2.5 & -1.5 & -2 \\
0 & 0 & -0.8 & 0.6
\end{bmatrix}.
\]

Finally, use the values of \( Y \) to solve the upper-triangular linear system \( UX = Y \) by back substitution to have

\[
x_3 = \frac{0.6}{-0.8} = \frac{-3}{4}, \quad x_2 = \frac{-2 + 1.5(x_3)}{2.5} = \frac{-2 + 1.5(-3/4)}{2.5} = -5/4,
\]

and

\[
x_1 = \frac{2 + 3(x_2) - 1(x_3)}{2} = \frac{2 + 3(-5/4) - (-3/4)}{2} = -1/2.
\]

**Definition 36** (Cholesky Factorisation). Let \( A \) be a real, symmetric and positive definite matrix. Then, it can be **factored** or **decomposed** in a unique way \( A = LL^T \), in which \( L \) is a lower-triangular matrix with a positive diagonal, and is termed **Cholesky factorisation**.

\[
\begin{bmatrix}
a_{11} & a_{12} & a_{13} \\
a_{21} & a_{22} & a_{23} \\
a_{31} & a_{32} & a_{33}
\end{bmatrix}
=
\begin{bmatrix}
l_{11} & 0 & 0 \\
l_{21} & l_{22} & 0 \\
l_{31} & l_{32} & l_{33}
\end{bmatrix}
\begin{bmatrix}
l_{11} & l_{21} & l_{31} \\
l_{21} & l_{22} & l_{32} \\
l_{31} & l_{32} & l_{33}
\end{bmatrix}.
\]

**Example 27.** (a) Determine the Cholesky decomposition of the matrix

\[
A = \begin{bmatrix}
2 & 1 & 0 \\
1 & 2 & 1 \\
0 & 1 & 2
\end{bmatrix}
\]
(b) Then, use the decomposition from part (a) to solve the linear system

\[
\begin{align*}
2x_1 + x_2 + 0x_3 &= -1 \\
x_1 + 2x_2 + x_3 &= -4 \\
0x_1 + x_2 + 2x_3 &= 2
\end{align*}
\]

Solution: Factor \( A \) as a product \( LL^T \) as follows:

\[
\begin{bmatrix}
2 & 1 & 0 \\
1 & 2 & 1 \\
0 & 1 & 2
\end{bmatrix}
= \begin{bmatrix}
l_{11} & 0 & 0 \\
l_{21} & l_{22} & 0 \\
l_{31} & l_{32} & l_{33}
\end{bmatrix}
\begin{bmatrix}
l_{11} & l_{21} & l_{31} \\
l_{21} & l_{22} & l_{32} \\
0 & 0 & l_{33}
\end{bmatrix}.
\]

From the first column we obtain

\[
2 = l_{11}^2 \implies l_{11} = \sqrt{2},
\]

\[
1 = l_{21}l_{11} + l_{22}(0) \implies 1 = l_{21}\sqrt{2} \implies l_{21} = \frac{1}{\sqrt{2}},
\]

\[
0 = l_{31}l_{11} + l_{32}(0) + l_{33}(0) \implies 0 = l_{31}\sqrt{2} \implies l_{31} = 0.
\]

Now, from the second column we have

\[
2 = l_{21}^2 + l_{22}^2 \implies 2 = \frac{1}{2} + l_{22}^2 \implies l_{22} = \sqrt{3}.
\]

\[
1 = l_{31}l_{21} + l_{32}l_{22} + l_{33}(0) \implies 1 = 0 + l_{32}\sqrt{3} \implies l_{32} = \frac{\sqrt{3}}{3}.
\]

Finally, from the third column we get

\[
2 = l_{31}^2 + l_{32}^2 + l_{33}^2 \implies 2 = 0 + \frac{2}{3} + l_{33}^2 \implies l_{33} = \sqrt{\frac{1}{3}} = \frac{2}{\sqrt{3}}.
\]
3.5 Iterative Methods

Direct methods are more efficient in solving linear systems of small dimensions in less computational cost than iterative methods. For large linear systems in particular for sparse linear systems iterative methods are more efficient for solving linear systems in terms of computational cost and effort compared to direct methods. In this section we will study the most common and basic iterative methods for solving linear algebraic systems which are Jacobi method and Gauss-Siedel method.

3.5.1 Jacobi Method

The general form of Jacobi iterative method for solving the \(i\)th equation in the linear system \(AX = B\) for unknown \(x_i, i = 1, \cdots, n\) is:

\[
x_i^k = \sum_{j=1}^{n} \left( - \frac{a_{ij} x_j^{k-1}}{a_{ii}} \right) + \frac{b_i}{a_{ii}}, \quad j \neq i, \quad a_{ii} \neq 0, \text{ for } i = 1, \cdots, n, \quad k = 1, \cdots, n.
\]

It is also known as Jacobi iterative process or Jacobi iterative technique

Example 28. Solve the following linear system using Jacobi iterative method

\[
\begin{align*}
2x_1 + x_2 + x_3 &= 0 \\
x_1 + 3x_2 + x_3 &= 0.5 \\
x_1 + x_2 + 2.5x_3 &= 0
\end{align*}
\]

Solution: These equations can be written in the form

\[
\begin{align*}
x_1 &= \frac{-x_2 - x_3}{2}, \\
x_2 &= \frac{0.5 - x_1 - x_3}{3}, \\
x_3 &= \frac{-x_1 - x_2}{2.5}.
\end{align*}
\]

Writing these equations in iterative form

\[
x_1^{k+1} = \frac{-x_2^k - x_3^k}{2},
\]
Let us start with initial guess \( P_0 = (x_0^1, x_0^2, x_0^3) = (0, 0.1, -0.1) \). Substituting these values in the right-hand side of each equation in above to find the new iterations

\[
\begin{align*}
x_1^1 &= \frac{-x_0^2 - x_0^0}{2} = \frac{-0.1 - (-0.1)}{2} = -0.1 + 0.1 = 0, \\
x_2^1 &= 0.5 - x_1^0 - x_3^0 = 0.5 - 0 - (-0.1) = 0.2, \\
x_3^1 &= \frac{-x_1^0 - x_2^0}{2.5} = -0 - 0.1 = -0.04.
\end{align*}
\]

Now, the new point \( P_1 = (x_1^1, x_2^1, x_3^1) = (0, 0.2, -0.04) \) is used in the Jacobi iterative form to find the next approximation \( P_2 \)

\[
\begin{align*}
x_1^2 &= \frac{-x_2^1 - x_3^1}{2} = \frac{-0.2 + 0.04}{2} = -0.16, \\
x_2^2 &= \frac{0.5 - x_1^1 - x_3^1}{3} = \frac{0.5 + 0.04}{3} = 0.54, \\
x_3^2 &= \frac{-x_1^1 - x_2^1}{2.5} = -0.2 = -0.08.
\end{align*}
\]

The new point \( P_2 = (x_1^2, x_2^2, x_3^2) = (-0.08, 0.18, -0.08) \) is closer to the solution than \( P_0 \) and \( P_1 \) and is used to find \( P_3 \)

\[
\begin{align*}
x_1^3 &= \frac{-x_2^2 - x_3^2}{2} = \frac{-0.18 + 0.08}{2} = -0.1 = -0.05, \\
x_2^3 &= \frac{0.5 - x_1^2 - x_3^2}{3} = \frac{0.5 + 0.08 + 0.08}{3} = 0.66, \\
x_3^3 &= \frac{-x_1^2 - x_2^2}{2.5} = \frac{0.08 - 0.18}{2.5} = -0.1 = -0.04.
\end{align*}
\]
CHAPTER 3. SOLVING SYSTEMS OF LINEAR EQUATIONS

This Jacobi iteration process generates a sequence of points \( \{P_n\} = \{(x_1^n, x_2^n, x_3^n)\} \) that converges to the solution \((x_1, x_2, x_3) = (-3/38, 4/19, -1/19) = (-0.078947368421053, 0.210526315789474, -0.052631578947368)\). The outline of the results is given in the Table 3.1.

<table>
<thead>
<tr>
<th>n</th>
<th>( x_1^n )</th>
<th>( x_2^n )</th>
<th>( x_3^n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.000000000000000</td>
<td>0.100000000000000</td>
<td>-0.100000000000000</td>
</tr>
<tr>
<td>1</td>
<td>0.000000000000000</td>
<td>0.200000000000000</td>
<td>-0.040000000000000</td>
</tr>
<tr>
<td>2</td>
<td>-0.080000000000000</td>
<td>0.180000000000000</td>
<td>-0.080000000000000</td>
</tr>
<tr>
<td>3</td>
<td>-0.050000000000000</td>
<td>0.220000000000000</td>
<td>-0.040000000000000</td>
</tr>
<tr>
<td>4</td>
<td>-0.090000000000000</td>
<td>0.196666666666667</td>
<td>-0.066000000000000</td>
</tr>
<tr>
<td>5</td>
<td>-0.064333333333333</td>
<td>0.219333333333333</td>
<td>-0.042666666666667</td>
</tr>
<tr>
<td>6</td>
<td>-0.083333333333333</td>
<td>0.202333333333333</td>
<td>-0.062000000000000</td>
</tr>
<tr>
<td>7</td>
<td>-0.070166666666667</td>
<td>0.216777777777778</td>
<td>-0.045600000000000</td>
</tr>
<tr>
<td>8</td>
<td>-0.085588888888889</td>
<td>0.205255555555556</td>
<td>-0.058644444444444</td>
</tr>
<tr>
<td>9</td>
<td>-0.073055555555556</td>
<td>0.214744444444444</td>
<td>-0.047866666666667</td>
</tr>
<tr>
<td>10</td>
<td>-0.083438888888889</td>
<td>0.207057407407407</td>
<td>-0.056755555555556</td>
</tr>
<tr>
<td>11</td>
<td>-0.075240925925926</td>
<td>0.213381481481481</td>
<td>-0.049470470470470</td>
</tr>
<tr>
<td>12</td>
<td>-0.081945370370370</td>
<td>0.208229444444444</td>
<td>-0.055238888888889</td>
</tr>
<tr>
<td>13</td>
<td>-0.076492777777778</td>
<td>0.212394753086820</td>
<td>-0.050513629629630</td>
</tr>
<tr>
<td>14</td>
<td>-0.080940561728395</td>
<td>0.209029691358023</td>
<td>-0.054359790123457</td>
</tr>
<tr>
<td>15</td>
<td>-0.077321589506173</td>
<td>0.211766739506173</td>
<td>-0.051224962962963</td>
</tr>
<tr>
<td>16</td>
<td>-0.080270910493287</td>
<td>0.209515517489712</td>
<td>-0.053778077777778</td>
</tr>
<tr>
<td>17</td>
<td>-0.077867198559675</td>
<td>0.211349662757202</td>
<td>-0.051697842798354</td>
</tr>
<tr>
<td>18</td>
<td>-0.079825909979424</td>
<td>0.209825520884774</td>
<td>-0.053392377160494</td>
</tr>
<tr>
<td>19</td>
<td>-0.078231571862140</td>
<td>0.211072623799732</td>
<td>-0.0520184362140</td>
</tr>
<tr>
<td>20</td>
<td>-0.079530459008916</td>
<td>0.210081138741427</td>
<td>-0.053136476207133</td>
</tr>
</tbody>
</table>

Table 3.1: Jacobi Iterative Solution of Example 28

3.5.2 Gauss-Seidel Method

An improvement of Jacobi method can be made by using the recent values \( x_i^k, i, k = 1, \ldots, n \), in the calculations once their values are obtained. This improvement is called \textit{Gauss-Seidel iterative method} and its general form for solving the \( i \text{th} \) equation in the linear system \( AX = B \) for unknown \( x_i, i = 1, \ldots, n \) is:
\[ x_i^k = \sum_{j=1}^{i-1} \left( - \frac{a_{ij} x_j^{k-1}}{a_{ii}} \right) + \sum_{j=i+1}^{n} \left( - \frac{a_{ij} x_j^{k-1}}{a_{ii}} \right) + \frac{b_i}{a_{ii}}, \ j \neq i, \ a_{ii} \neq 0, \]

for \( i = 1, \ldots, n, \) and \( k = 1, \ldots, n. \)

It is also known as Gauss-Sedel iterative process or Gauss-Sedel iterative technique

**Example 29.** Solve the following linear system using Gauss-Siedel iterative method

\[
\begin{align*}
2x_1 - 4x_2 + x_3 &= -1 \\
x_1 + x_2 + 6x_3 &= 1 \\
3x_1 + 3x_2 + 5x_3 &= 4
\end{align*}
\]

**Solution:** Rearrange the system in above such that the coefficient matrix is strictly diagonally dominant

\[
\begin{align*}
3x_1 + 3x_2 + 5x_3 &= 4 \\
2x_1 - 4x_2 + x_3 &= -1 \\
x_1 + x_2 + 6x_3 &= 1
\end{align*}
\]

These equations can be written in the form

\[
\begin{align*}
x_1 &= \frac{4 - 3x_2 - 5x_3}{3}, \\
x_2 &= \frac{-1 - 2x_1 - x_3}{-4} = \frac{1 + 2x_1 + x_3}{4}, \\
x_3 &= \frac{1 - x_1 - x_2}{6}.
\end{align*}
\]

This suggests the following Gauss-Siedel iterative process

\[
x_1^{n+1} = \frac{4 - 3x_2^n - 5x_3^n}{3},
\]
CHAPTER 3. SOLVING SYSTEMS OF LINEAR EQUATIONS

\[
\begin{align*}
x_2^{n+1} &= \frac{1 + 2x_1^{n+1} + x_3^n}{4}, \\
x_3^{n+1} &= \frac{1 - x_1^{n+1} - x_2^{n+1}}{6}.
\end{align*}
\]

We start with initial guess \( P_0 = (x_1^0, x_2^0, x_3^0) = (1, 0.1, -1). \) Substitute \( x_2^0 = 0.1 \) and \( x_3^0 = -1 \) in the first equation and have

\[
x_1^1 = \frac{4 - 3x_2^0 - 5x_3^0}{3} = \frac{4 - 3(0.1) - 5(-1)}{3} = \frac{8.7}{3} = 2.9.
\]

Then, substitute the new value \( x_1^1 = 2.9 \) and \( x_3^0 = -1 \) into the second equation to obtain

\[
x_2^1 = \frac{1 + 2x_1^1 + x_3^0}{4} = \frac{1 + 2(2.9) + (-1)}{4} = 1.45.
\]

Finally, substitute the new values \( x_1^1 = 2.9 \) and \( x_2^1 = 1.45 \) in the third equation and get

\[
x_3^1 = \frac{1 - x_1^1 - x_2^1}{6} = \frac{1 - 2.9 - 1.45}{6} = -\frac{3.35}{6} = -0.558333333333333.
\]

Now, we have the new point \( P_1 = (x_1^1, x_2^1, x_3^1) = (2.9, 1.45, -0.558333333333333) \) is used to find the next approximation \( P_2. \)

Substitute \( x_2^1 = 1.45 \) and \( x_3^1 = -0.558333333333333 \) in the first equation and get

\[
x_1^2 = \frac{4 - 3x_2^1 - 5x_3^1}{3} = \frac{4 - 3(1.45) - 5(-0.558333333333333)}{3} = \frac{2.416666666666666}{3} = 0.813888888888889.
\]

Then, substitute the new value \( x_2^1 = 0.813888888888889 \) and \( x_3^1 = -0.558333333333333 \) into the second equation to obtain

\[
x_2^2 = \frac{1 + 2x_1^2 + x_3^1}{4} = \frac{1 + 2(0.813888888888889) + (-0.558333333333333)}{4} = \frac{2.069444444444445}{4} = 0.51736111111111.
\]
Finally, substitute the new values $x_1^1 = 0.813888888888889$ and $x_2^1 = 0.517361111111111$ in the third equation and get

$$x_3^2 = \frac{1 - x_1^2 - x_2^2}{6} = \frac{1 - 0.813888888888889 - 0.517361111111111}{6} = -0.331250000000000$$

This iteration process generates a sequence of points $\{P_n\} = \{(x_1^n, x_2^n, x_3^n)\}$ that converges to the solution $(x_1, x_2, x_3) = (32/39, 25/39, -1/13) = (0.82051280512820, 0.641025641025641, -0.076923076923077)$. The results are given in the Table 3.2.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$x_1^n$</th>
<th>$x_2^n$</th>
<th>$x_3^n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.000000000000000</td>
<td>0.100000000000000</td>
<td>-1.000000000000000</td>
</tr>
<tr>
<td>1</td>
<td>2.900000000000000</td>
<td>1.450000000000000</td>
<td>-0.558333333333333</td>
</tr>
<tr>
<td>2</td>
<td>0.813888888888889</td>
<td>0.517361111111111</td>
<td>-0.055208333333333</td>
</tr>
<tr>
<td>3</td>
<td>0.907986111111111</td>
<td>0.690190972222222</td>
<td>-0.099696180555556</td>
</tr>
<tr>
<td>4</td>
<td>0.809302662037037</td>
<td>0.629727858796300</td>
<td>-0.073171657986111</td>
</tr>
<tr>
<td>5</td>
<td>0.825588107638889</td>
<td>0.644486490885417</td>
<td>-0.078340883608218</td>
</tr>
<tr>
<td>6</td>
<td>0.819414981794946</td>
<td>0.640122269995419</td>
<td>-0.076589541965061</td>
</tr>
<tr>
<td>7</td>
<td>0.820860299946349</td>
<td>0.641282764481909</td>
<td>-0.077023844071376</td>
</tr>
<tr>
<td>8</td>
<td>0.820423642303718</td>
<td>0.640955860134015</td>
<td>-0.076896583739622</td>
</tr>
<tr>
<td>9</td>
<td>0.820538446098689</td>
<td>0.641045077114439</td>
<td>-0.07693087202188</td>
</tr>
<tr>
<td>10</td>
<td>0.820505901555874</td>
<td>0.641020303977390</td>
<td>-0.076921034255544</td>
</tr>
<tr>
<td>11</td>
<td>0.820514753115183</td>
<td>0.641027117993706</td>
<td>-0.076923645184815</td>
</tr>
<tr>
<td>12</td>
<td>0.82051290647653</td>
<td>0.64102534027623</td>
<td>-0.076922920779213</td>
</tr>
<tr>
<td>13</td>
<td>0.820512967271065</td>
<td>0.641025753440729</td>
<td>-0.076923120118632</td>
</tr>
<tr>
<td>14</td>
<td>0.820512780090325</td>
<td>0.641025610015504</td>
<td>-0.076923065017638</td>
</tr>
<tr>
<td>15</td>
<td>0.820512831680559</td>
<td>0.641025649585870</td>
<td>-0.07692308211072</td>
</tr>
<tr>
<td>16</td>
<td>0.820512817432583</td>
<td>0.641025638663523</td>
<td>-0.076923076016018</td>
</tr>
<tr>
<td>17</td>
<td>0.820512821363173</td>
<td>0.641025641677582</td>
<td>-0.076923077173459</td>
</tr>
<tr>
<td>18</td>
<td>0.820512820278183</td>
<td>0.641025640845727</td>
<td>-0.076923076539385</td>
</tr>
<tr>
<td>19</td>
<td>0.820512820577581</td>
<td>0.641025641075294</td>
<td>-0.076923076942146</td>
</tr>
<tr>
<td>20</td>
<td>0.820512820494949</td>
<td>0.641025641019388</td>
<td>-0.076923076917814</td>
</tr>
</tbody>
</table>

Table 3.2: Gauss-Siedel Iterative Solution of Example 29
Exercises

Exercise 17. Solve Example 22 Using Gauss elimination with forward substitution method. Compare the solution with solution of the same example.

Exercise 18. Solve Example 23 Using Gauss elimination with backward substitution method. Compare the solution with solution of the same example.

Exercise 19. Repeat Example 28 with Gauss-Siedel iteration. Compute five iterations and compare them with Jacobi iterations in the same example.

Exercise 20. Redo Example 29 with Jacobi iteration. Compute five iterations and compare them with Gauss-Siedel iterations in the same example.

Exercise 21. Use Gauss elimination with backward substitution method and three-digit rounding arithmetic to solve the following linear system

\[
\begin{align*}
x_1 + 3x_2 + 2x_3 &= 5 \\
x_1 + 2x_2 - 3x_3 &= -2 \\
x_1 + 5x_2 + 3x_3 &= 10 \\
\end{align*}
\]

Exercise 22. (a) Determine the \( LU \) factorisation for matrix \( A \) in the linear system \( AX = B \), where

\[
A = \begin{bmatrix}
-1 & 1 & -2 \\
2 & -1 & 1 \\
-4 & 1 & -2
\end{bmatrix}
\quad \text{and} \quad
B = \begin{bmatrix}
2 \\
1 \\
4
\end{bmatrix}.
\]

(b) Then use the factorisation to solve the system

\[
\begin{align*}
-x_1 + x_2 - 2x_3 &= 2 \\
2x_1 - x_2 + x_3 &= 1 \\
-4x_1 + x_2 - 2x_3 &= 4 \\
\end{align*}
\]

Exercise 23. Solve the following linear system using Gauss-Jordan elimination method

\[
\begin{align*}
-4x_1 - x_2 - 2x_3 &= -9 \\
-x_1 - x_2 + 3x_3 &= 9 \\
-2x_1 - 4x_2 + x_3 &= 5 \\
\end{align*}
\]
Chapter 4
Interpolation and Extrapolation

4.1 Introduction

In applied sciences and engineering, scientists and engineers often collect a number of data points of different scientific phenomena via experimentation and sampling. In many cases they need to estimate (interpolate) a function at a point its functional value is not in the range of the collected data. Interpolation is a branch of numerical analysis studies the methods and techniques of estimating an unknown value of a function at an intermediate value of its independent variable. Also, interpolation is used to replace a complicated function by a simpler one.

4.2 Lagrange Interpolation

Suppose that we would like to interpolate an arbitrary function $f$ at a set of limited points $x_0, x_1, \cdots, x_n$. These $n+1$ points are known as interpolation nodes in interpolation theory. Firstly, we need to introduce a system of $n+1$ special polynomials of degree $n$ known as interpolating polynomials or cardinal polynomials. These polynomials are denoted by $\ell_0, \ell_1, \cdots, \ell_n$ and defined using Kronecker delta notation as follows

$$\ell_i(x_j) = \delta_{ij} = \begin{cases} 
1 & \text{if } i = j, \\
0 & \text{if } i \neq j.
\end{cases}$$
Then, we can interpolate the function $f$ by a polynomial $P_n$ of degree $n$ defined by

$$P_n(x) = \sum_{i=0}^{n} \ell_i(x)f(x_i),$$

this polynomial is called **Lagrange polynomial** or **Lagrange form of the interpolation polynomial**, and it is a linear combination of the cardinal polynomials $\ell_i, i = 0, 1, \ldots, n$. Moreover, it coincides with the function $f$ at the the nodes $x_j, j = 0, 1, \ldots, n$, namely

$$P_n(x_j) = \sum_{i=0}^{n} \ell_i(x_j)f(x_j) = \ell_j(x_j)f(x_j) = f(x_j).$$

The interpolating polynomials can be expressed as a product of $n$ linear factors

$$\ell_i(x) = \prod_{j \neq i}^{n} \frac{(x-x_j)}{(x_i-x_j)}, \quad i = 0, 1, \ldots,$$

i.e.

$$\ell_i(x) = \frac{(x-x_0)}{(x_i-x_0)} \frac{(x-x_1)}{(x_i-x_1)} \ldots \frac{(x-x_{i-1})}{(x_i-x_{i-1})} \frac{(x-x_{i+1})}{(x_i-x_{i+1})} \ldots \frac{(x-x_n)}{(x_i-x_n)}.$$  

**Example 30.** Determine the linear Lagrange polynomial that passes through the points $(1, 5)$ and $(4, 2)$ and use it to interpolate the linear function at $x = 3$.

**Solution:** Writing out the cardinal polynomials

$$\ell_0(x) = \frac{(x-x_1)}{(x_0-x_1)} = \frac{(x-4)}{(1-4)} = -\frac{1}{3} (x-4),$$

and

$$\ell_1(x) = \frac{(x-x_0)}{(x_1-x_0)} = \frac{(x-1)}{(4-1)} = \frac{1}{3} (x-1).$$

Hence, the Lagrange polynomial is

$$P_1(x) = \sum_{i=0}^{1} \ell_i(x)f(x_i) = \ell_0(x)f(x_0) + \ell_1(x)f(x_1) =$$

$$-\frac{1}{3} (x-4)(5) + \frac{1}{3} (x-1)(2) = -x + 6.$$
CHAPTER 4. INTERPOLATION AND EXTRAPOLATION

So,

\[ P_1(3) = -(3) + 6 = 3. \]

Note that

\[ P_1(1) = -(1) + 6 = 5 = f(1), \text{ and } P_1(4) = -(4) + 6 = 2 = f(4). \]

Example 31. Find the Lagrange polynomial that interpolates the following data

<table>
<thead>
<tr>
<th>x</th>
<th>1</th>
<th>2</th>
<th>2.5</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>f(x)</td>
<td>0</td>
<td>5</td>
<td>6.5</td>
<td>7</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

Solution: The cardinal polynomials are:

\[
\ell_0(x) = \frac{(x - x_1)(x - x_2)(x - x_3)(x - x_4)(x - x_5)}{(x_0 - x_1)(x_0 - x_2)(x_0 - x_3)(x_0 - x_4)(x_0 - x_5)}
= \frac{(x - 2)(x - 2.5)(x - 3)(x - 4)(x - 5)}{(1 - 2)(1 - 2.5)(1 - 3)(1 - 4)(1 - 5)}
= -\frac{1}{36}x^5 + \frac{11}{24}x^4 - \frac{53}{18}x^3 + \frac{221}{24}x^2 - \frac{505}{36}x + \frac{25}{3},
\]

\[
\ell_1(x) = \frac{(x - x_0)(x - x_2)(x - x_3)(x - x_4)(x - x_5)}{(x_1 - x_0)(x_1 - x_2)(x_1 - x_3)(x_1 - x_4)(x_1 - x_5)}
= \frac{(x - 1)(x - 2.5)(x - 3)(x - 4)(x - 5)}{(2 - 1)(2 - 2.5)(2 - 3)(2 - 4)(2 - 5)}
= \frac{1}{3}x^5 - \frac{31}{6}x^4 + \frac{61}{2}x^3 - \frac{509}{6}x^2 + \frac{655}{6}x - 50,
\]

\[
\ell_2(x) = \frac{(x - x_0)(x - x_1)(x - x_3)(x - x_4)(x - x_5)}{(x_2 - x_0)(x_2 - x_1)(x_2 - x_3)(x_2 - x_4)(x_2 - x_5)}
= \frac{(x - 1)(x - 2)(x - 3)(x - 4)(x - 5)}{(2.5 - 1)(2.5 - 2)(2.5 - 3)(2.5 - 4)(2.5 - 5)}
= -\frac{5000}{7031}x^5 + \frac{75000}{7031}x^4 - \frac{425000}{7031}x^3 + \frac{1125000}{7031}x^2 - \frac{1370000}{7031}x + \frac{600000}{7031},
\]
CHAPTER 4. INTERPOLATION AND EXTRAPOLATION

\[ \ell_3(x) = \frac{(x - x_0)(x - x_1)(x - x_2)(x - x_4)(x - x_5)}{(x_3 - x_0)(x_3 - x_1)(x_3 - x_2)(x_3 - x_4)(x_3 - x_5)} \]
\[ = \frac{(x - 1)(x - 2)(x - 2.5)(x - 4)(x - 5)}{(3 - 1)(3 - 2)(3 - 2.5)(3 - 4)(3 - 5)} \]
\[ = \frac{1}{2}x^5 - \frac{29}{4}x^4 + \frac{79}{2}x^3 - \frac{401}{4}x^2 + \frac{235}{2}x - 50, \]

\[ \ell_4(x) = \frac{(x - x_0)(x - x_1)(x - x_2)(x - x_3)(x - x_5)}{(x_4 - x_0)(x_4 - x_1)(x_4 - x_2)(x_4 - x_3)(x_4 - x_5)} \]
\[ = \frac{(x - 1)(x - 2)(x - 2.5)(x - 3)(x - 5)}{(4 - 1)(4 - 2)(4 - 2.5)(4 - 3)(4 - 5)} \]
\[ = -\frac{1}{9}x^5 + \frac{3}{2}x^4 - \frac{137}{18}x^3 + \frac{109}{6}x^2 - \frac{365}{18}x + \frac{25}{3}, \]

\[ \ell_5(x) = \frac{(x - x_0)(x - x_1)(x - x_2)(x - x_3)(x - x_4)}{(x_5 - x_0)(x_5 - x_1)(x_5 - x_2)(x_5 - x_3)(x_5 - x_4)} \]
\[ = \frac{(x - 1)(x - 2)(x - 2.5)(x - 3)(x - 4)}{(5 - 1)(5 - 2)(5 - 2.5)(5 - 3)(5 - 4)} \]
\[ = \frac{1}{60}x^5 - \frac{5}{24}x^4 + \frac{55}{24}x^3 - \frac{55}{6}x^2 + \frac{149}{60}x - 1. \]

Hence, the Lagrange polynomial is
CHAPTER 4. INTERPOLATION AND EXTRAPOLATION

\[ P_5(x) = \sum_{i=0}^{6} \ell_i(x)f(x_i) = \ell_i(x)f(x_i) + \ell_i(x)f(x_i) + \ell_i(x)f(x_i) + \ell_i(x)f(x_i) = \]

\[ -\frac{1}{36}x^5 + \frac{11}{24}x^4 - \frac{53}{18}x^3 + \frac{221}{24}x^2 - \frac{505}{36}x + \frac{25}{3} \quad (0) + \]

\[ \frac{1}{3}x^5 - \frac{31}{6}x^4 + \frac{61}{2}x^3 - \frac{509}{6}x^2 + \frac{655}{6}x - 50 \quad (5) + \]

\[ -\frac{5000}{7031}x^5 + \frac{75000}{7031}x^4 - \frac{425000}{7031}x^3 + \frac{1125000}{7031}x^2 - \frac{1370000}{7031}x + \frac{600000}{7031} \quad (6.5) + \]

\[ \frac{1}{2}x^5 - \frac{29}{4}x^4 + \frac{79}{2}x^3 - \frac{401}{4}x^2 + \frac{235}{2}x - 50 \quad (7) + \]

\[ -\frac{1}{9}x^5 + \frac{3}{2}x^4 - \frac{137}{18}x^3 + \frac{109}{6}x^2 - \frac{365}{18}x + \frac{25}{3} \quad (3) + \]

\[ \frac{1}{60}x^5 - \frac{5}{24}x^4 + x^3 - \frac{55}{24}x^2 + \frac{149}{60}x - 1 \quad (1). \]

After some mathematical manipulation, we have

\[ P_5(x) = -\frac{8}{316395}x^5 + \frac{8}{21093}x^4 - \frac{2722272566677}{1266637395197952}x^3 + \frac{200167100491}{3518437208832}x^2 \]

\[ -\frac{10969157106929}{1583296743997440}x - \frac{187476506320011}{8796093022208}. \]

Note that the Lagrange interpolant is used to interpolate a function at a set of non-equally spaced points.

4.3 Newton Interpolation

Newton interpolation is used to interpolate a function at a set of given equally spaced points \( x_0, x_1, \ldots, x_n \). Before we start we need to define the finite divided differences.
4.3.1 Finite Divided Differences

The first finite divided difference of the function \( f \) is in general given by

\[
f[x_i, x_j] = \frac{f(x_j) - f(x_i)}{x_j - x_i}.
\]

The second finite divided difference is the difference between the two divided difference, is represented by

\[
f[x_i, x_j, x_k] = \frac{f[x_j, x_k] - f[x_i, x_j]}{x_k - x_i}.
\]

Likewise, the \( n \)th finite divided difference is expressed by

\[
f[x_0, x_1, \cdots, x_{n-1}, x_n] = \frac{f[x_1, \cdots, x_{n-1}, x_n] - f[x_0, x_1, \cdots, x_{n-1}]}{x_n - x_0}.
\]

Note that the zero-order difference is defined as

\[
f[x_i] = f(x_i) = f_i.
\]

Also, observe that

\[
f[x_i, x_j] = \frac{f(x_j) - f(x_i)}{x_j - x_i} = \frac{f(x_i) - f(x_j)}{x_i - x_j} = f[x_j, x_i].
\]

The divided differences is summarised in the divided difference table given below:

<table>
<thead>
<tr>
<th>( x_i )</th>
<th>( f_i )</th>
<th>( f[x_i, x_{i+1}] )</th>
<th>( f[x_i, x_{i+1}, x_{i+2}] )</th>
<th>( f[x_i, x_{i+1}, x_{i+2}, x_{i+3}] )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_0 )</td>
<td>( f_0 )</td>
<td>( f[x_0, x_1] )</td>
<td>( f[x_0, x_1, x_2] )</td>
<td>( f[x_0, x_1, x_2, x_3] )</td>
</tr>
<tr>
<td>( x_1 )</td>
<td>( f_1 )</td>
<td>( f[x_1, x_2] )</td>
<td>( f[x_1, x_2, x_3] )</td>
<td>( f[x_1, x_2, x_3, x_4] )</td>
</tr>
<tr>
<td>( x_2 )</td>
<td>( f_2 )</td>
<td>( f[x_2, x_3] )</td>
<td>( f[x_2, x_3, x_4] )</td>
<td></td>
</tr>
<tr>
<td>( x_3 )</td>
<td>( f_3 )</td>
<td>( f[x_3, x_4] )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( x_4 )</td>
<td>( f_4 )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Example 32. Compute the divided differences of the following data

\[
\begin{align*}
x_0 & = 0.5 \\
x_1 & = 1.0 \\
x_2 & = 1.5 \\
x_3 & = 2.0 \\
x_4 & = 2.5 \\
f_0 & = 1.0 \\
f_1 & = 2.0 \\
f_2 & = 3.0 \\
f_3 & = 4.0 \\
f_4 & = 5.0
\end{align*}
\]
Solution: Using the standard notation the first finite divided differences are:

\[
\begin{align*}
    f[x_0, x_1] &= \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{3 - 1.1250}{1 - 0.5} = \frac{1.8750}{0.5} = 3.7500. \\
    f[x_1, x_2] &= \frac{f(x_2) - f(x_1)}{x_2 - x_1} = \frac{7.3750 - 3}{1.5 - 1} = \frac{4.3750}{0.5} = 8.7500. \\
    f[x_2, x_3] &= \frac{f(x_3) - f(x_2)}{x_3 - x_2} = \frac{15 - 7.3750}{2 - 1.5} = \frac{7.6250}{0.5} = 15.2500. \\
    f[x_3, x_4] &= \frac{f(x_4) - f(x_3)}{x_4 - x_3} = \frac{26.6250 - 15}{2.5 - 2} = \frac{11.6250}{0.5} = 23.2500.
\end{align*}
\]

Now, using the computed first divided differences, we compute the second divided differences

\[
\begin{align*}
    f[x_0, x_1, x_2] &= \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0} = \frac{8.7500 - 3.7500}{1.5 - 0.5} = \frac{5}{1} = 5.000. \\
    f[x_0, x_1, x_2] &= \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0} = \frac{8.7500 - 3.7500}{1.5 - 0.5} = \frac{5}{1} = 5.000. \\
    f[x_1, x_2, x_3] &= \frac{f[x_2, x_3] - f[x_1, x_2]}{x_3 - x_1} = \frac{15.2500 - 8.7500}{2 - 1} = \frac{6.5000}{1} = 6.5000. \\
    f[x_2, x_3, x_4] &= \frac{f[x_3, x_4] - f[x_2, x_3]}{x_4 - x_2} = \frac{23.2500 - 15.2500}{2.5 - 1.5} = \frac{8.0000}{1} = 8.0000.
\end{align*}
\]

Finally, we compute the third divided differences using the computed second divided differences

\[
\begin{align*}
    f[x_0, x_1, x_2, x_3] &= \frac{f[x_1, x_2, x_3] - f[x_0, x_1, x_2]}{x_3 - x_0} = \frac{6.5000 - 5.0000}{2 - 0.5} = \frac{1.5000}{1.5000} = 1.0000. \\
    f[x_1, x_2, x_3, x_4] &= \frac{f[x_2, x_3, x_4] - f[x_1, x_2, x_3]}{x_4 - x_1} = \frac{8.0000 - 6.5000}{2.5 - 1} = \frac{1.5000}{1.5000} = 1.0000.
\end{align*}
\]

The results are outlined in the following table

<table>
<thead>
<tr>
<th>(x)</th>
<th>0.5000</th>
<th>1.000</th>
<th>1.500</th>
<th>2.000</th>
<th>2.500</th>
</tr>
</thead>
<tbody>
<tr>
<td>(f(x))</td>
<td>1.1250</td>
<td>3.000</td>
<td>7.3750</td>
<td>15.0000</td>
<td>26.6250</td>
</tr>
</tbody>
</table>
4.3.2 Newton’s Interpolation Divided Difference Formula

The general form of Newton’s interpolation polynomial of order $n$ for $n + 1$ data points is

$$P_n(x) = d_0 + d_1(x - x_0) + d_2(x - x_0)(x - x_1) + d_3(x - x_0)(x - x_1)(x - x_2) + \cdots + d_n(x - x_0)(x - x_1)(x - x_2)\cdots(x - x_{n-1}),$$

where

$$d_0 = f[x_0],$$

$$d_1 = f[x_0, x_1],$$

$$d_2 = f[x_0, x_1, x_2],$$

$$d_3 = f[x_0, x_1, x_2, x_3],$$

$$\vdots$$

$$d_n = f[x_0, x_1, \cdots, x_n].$$

Example 33. Use the data from Example 32 to construct Newton’s interpolation divided difference formula, and use it to evaluate $f(0)$, $f(3)$, and $f(3.25)$.  

<table>
<thead>
<tr>
<th>$x_i$</th>
<th>$f_i$</th>
<th>$f[x_i, x_{i+1}]$</th>
<th>$f[x_i, x_{i+1}, x_{i+2}]$</th>
<th>$f[x_i, x_{i+1}, x_{i+2}, x_{i+3}]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5000</td>
<td>1.1250</td>
<td>3.7500</td>
<td>5.000</td>
<td>1.0000</td>
</tr>
<tr>
<td>1.000</td>
<td>3.000</td>
<td>8.7500</td>
<td>6.5000</td>
<td>1.0000</td>
</tr>
<tr>
<td>1.5000</td>
<td>7.3750</td>
<td>15.2500</td>
<td>8.0000</td>
<td></td>
</tr>
<tr>
<td>2.000</td>
<td>15.000</td>
<td>23.2500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5000</td>
<td>26.6250</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 4. INTERPOLATION AND EXTRAPOLATION

**Solution:** The Newton’s polynomial of third order for the data in the table above is

\[ P_3(x) = d_0 + d_1(x - x_0) + d_2(x - x_0)(x - x_1) + d_3(x - x_0)(x - x_1)(x - x_2) = 1.1250 + 3.75(x - 0.5) + 5(x - 0.5)(x - 1) + (x - 0.5)(x - 1)(x - 1.5). \]

After some mathematical manipulation, we have

\[ P_3(x) = x^3 + 2x^2 - x + 1. \]

Hence,

\[ f(0) = P_3(0) = 1, \quad f(3) = P_3(3) = 43, \quad f(3.25) = P_3(3.25) = 53.2031. \]
Bibliography


